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

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

Figures

Table of Contents

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 CO_2 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 N_2 to NH_3. 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 O_2 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; Euzeby, 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* (Euzeby, 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*, Oscillatorials, *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 bacterlogical 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 alkinetes 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. Mycrocystin 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 is 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 *Cylinderospermopsis 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 that 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/cm^2 to treat pathogens such as *Crytosporidium 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 saxitons. 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 cooper sulphates is environmentally controversial, cooper 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	LD_{50} (i.p. in mouse)	Toxin Producing Organism	Effect	Symptoms
Microcystins	45-1000µg/kg (Yoshida et al., 1997)	*Microcystis, Anabaena, Planktothrix, Nostoc, Hepalosiphon*	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)	*Nodularia spumigena*	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)	*Cylindrospermopsis, Aphanizomenon, Umezakia*	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)	*Anabaena, Aphanizomenon flos aquae, Lyngbya, Cylindrospermopsis, Planktothrix*	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)	*Anabaena flos aquae, Anabaems lemmermanii, Anabaena, Planktothrix, Aphanizomenon, Cylindrospermopsis*	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	*Aphanizomenon. flos aquae*	Acetylcholinesterase inhibitor. (Chorus et al., 2000)	Respiratory paralysis Hypersalivation, muciod nasal discharge, tremors, diarrhea, paresis, death. (Chorus et al., 2000)

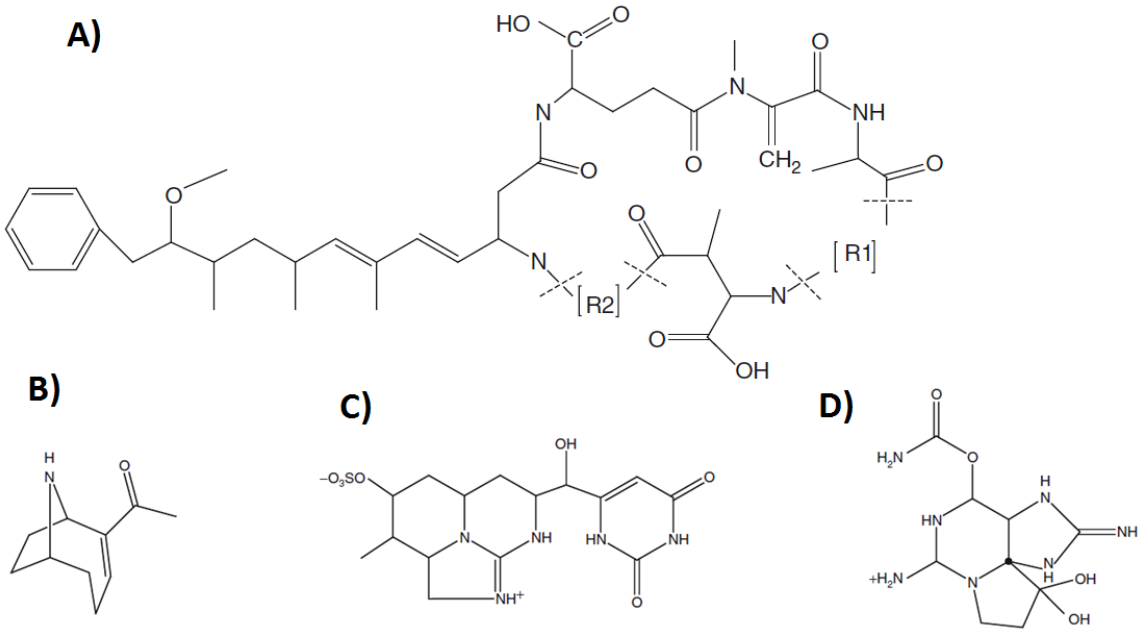


Figure 1 - General structures of A) Microcystin B) Anatoxin-a C) Saxitoxin and D) Cylindrospermopsin (Pelaez et al., 2010).

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 lemmerannii 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 aeruginsoa 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 micorocystin production is likely from *Planktothrix* species present in the lower, deeper portions of blooms dominated by *Microcystis aeruginosa (Rinta-Kanto 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	*Microcystis aeruginosa* (Brittain et al., 2000) (Vanderploeg et al., 2001) (Rinta-Kanto and Wilhelm, 2006) (Watson et al., 2008)	Microcystin
Nostocales	*Cylindrospermopsis raciborskii (Conroy et al., 2007)*	Cylindrospermopsin Saxitoxin Anatoxin-a
	Aphanizomenon flos-aquae (Sawyer et al., 1968)	Anatoxin-a Saxitox Cylindrospermopsin
	Anabaena lemmermannii	Anatoxin-a(s) Anatoxin-a Microcystins
	Anabaena circinalisi	Anatoxin-a Anatoxin-a(s) Microcystins
	Anabaena sp. (Watson et al., 2008)	Anatoxin-a(s) Anatoxin-a Microcystins
	Anabaena flos-aquae	Anatoxin-a Anatoxin-a(s) Microcystins
Oscillatoriales	*Pseudanabaena contorta spec. nova* (Kling et al., 2012)	Anatoxin-a
	Planktothrix aghardii (Makarewicz et al., 1999) (Rinta-Kanto and Wilhelm, 2006)	Microcystins Anatoxin-a
	Oscillatoria 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**				
Anabaena	*Anabaena*	*Aphanothece*	*Gloeotrichia*	*Nostoc*	*Rhabdoderma lineare*
Anabaena lemmermanii	*Anabaena affinis*	*Aphanothece clathrata*	*Gloeotrichia echinulata*	*Oscillatoria*	*Romeria*
Anabaena flos-aquae	*Anabaena aphanizomenoides*	*Aphanothece clathrata* var. *brevis*	*Gomphosphaeria*	*Oscillatoria Agardhii* var. *isothrix*	*Scytonema*
Anabaena planctonica	*Anabaena augustunalis* ?	*Aphanothece nidulans*	*Gomphosphaeria aponina*	*Oscillatoria geminata*	*Spirulina*
Anabaena spiroides	*Anabaena circinalis*	*Aphanothece stagnina*	*Gomphosphaeria lacu*stris	*Oscillatoria limnetica*	*Spirulina laxa*
Aphanizomenon	*Anabaena constricta*	*Arthrospira*	*Johannesbaptisti a*	*Oscillatoria limnetica* var. *acicularis*	*Synechococcus*
Aphanizomenon flos-aquae	*Anabaena cylindrica*	*Chroococcus*	*Limnothrix*	*Oscillatoria planctonica*	*Synechocystis*
Aphanothece	*Anabaena flos-aquae*	*Chroococcus dispersus*	*Lyngbya*	*Oscillatoria redeckei*	*Tolypothrix*
Gloeotrichia	*Anabaena lemmermanii*	*Chroococcus dispersus* var. *minor*	*Lyngbya birgei*	*Oscillatoria* sp. "A"	*Trichodesmium*
Microcystis	*Anabaena limnetica*	*Chroococcus limneticus*	*Lyngbya contorta*	*Oscillatoria* sp. "B"	*Wollea*
Microcystis aeruginosa	*Anabaena planctonica*	*Chroococcus plantonicus*	*Lyngbya epiphytica*	*Oscillatoria splendida = Geitlerinema*	*Woronichinia klingae*
Oscillatoria	*Anabaena sphaerica*	*Chroococcus prescottii*	*Lyngbya limnetica*	*Oscillatoria utermoehlii*	*Woronichinia naegeliana* (=*Coelosphaeriu m naegelianum*)
Lyngbya	*Anabaena spiroides*	*Chroococcus varians* ?	*Lyngbya vacuolifera*	*Phormidium*	
Planktothrix	*Anabaena spiroides* var. *crassa*	*Coelomoron*	*Marsoniella*	*Phormidium antarcticum*	
Planktolyngbya	*Anabaena wisconsinense*	*Coelosphaerium*	*Merismopedia*	*Phormidium mucicola*	
Scytonema	*Anabaenopsis*	*Coelosphaerium kuetzingianum*	*Merismopedia elegans*	*Planktolyngbya*	
Woronichinia naegeliana (=*Coelosphaerium naegelianum*)	*Aphanizomenon*	*Cyanarcus*	*Merismopedia glauca*	*Planktothrix*	
Nostoc	*Aphanizomenon flos-aquae*	*Cyanococcus*	*Merismopedia minima* ?	*Planktothrix agardhii* (=*Oscillatoria agardhii*)	
Gloeocapsa	*Aphanizomenon flos-aquae* var. *gracile*	*Cyanodictyon*	*Merismopedia tenuissima*	*Plectonema*	
Chroococcus	*Aphanizomenon gracile*	Cyanophyceae colonial	*Microcystis*	*Pseudanabaena*	
Coelosphaerium	*Aphanizomenon issatschenkoi*	Cyanophyceae filamentous	*Microcystis aeruginosa*	*Pseudanabaena constricta*	
Limnothrix	*Aphanizomenon issatschenkoi* var. *gracilis*	Cyanophyceae mixed	*Microcystis botrys*	*Pseudanabaena contorta*	
Phormidium	*Aphanizomenon schindleri*	*Cylindrospermum*	*Microcystis flos-aquae*	*Pseudanabaena galeata*	
	Aphanocapsa	*Dactylococcopsis*	*Microcystis incerta*	*Radiocystis*	
	Aphanocapsa delicatissima	*Eucapsis*	*Microcystis nov*acekii	*Radiocystis geminata*	
	Aphanocapsa elachista	*Gloeocapsa*	*Microcystis pulverea*	*Raphidiopsis*	
	Aphanocapsa elachista var. *conferta*	*Gloeothece*	*Microcystis viridis*	*Rhabdoderma*	
	Aphanocapsa holsatica	*Gloeothece linearis* var. *composita*	*Microcystis wesenbergii*	*Rhabdoderma gorskii*	

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; Euzeby, 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 differenciates *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, tricome 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.

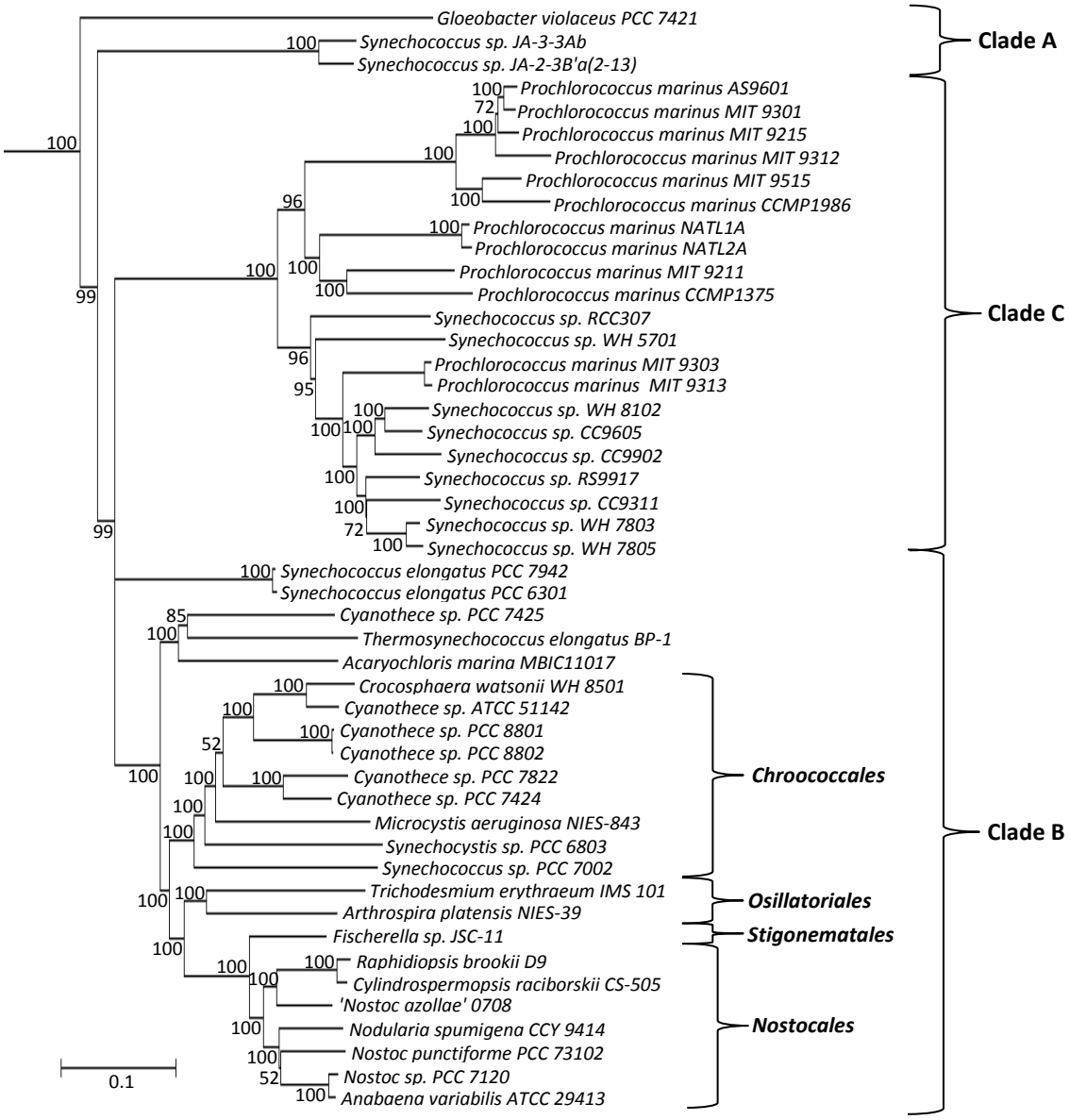


Figure 3: A maximum-likelihood consensus tree based on 45 concatenated sequences for 48 species of Cyanobacteria. The tree was rooted using sequences for *Staphylococcus aureus subsp. aureus Mu50* and *Bacillus subtilis str. 168*. Numbers located at nodes indicate the bootstrap values out of 100. Major 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 memebers are those which have sequenced genomes available. ^{N} and ^{S} denot 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
Acaryochloris marina MBIC11017	CP000828.1	B	8.36	8383	(Swingley et al., 2008b)
Anabaena variabilis ATCC 29413	CP000117.1	B	7.11	5710	DOE JGI [b]
Arthrospira platensis NIES-39[f]	AP011615.1	B	6.79	6630	(Fujisawa et al., 2010)
Crocosphaera watsonii WH 8501[a]	AADV00000000.2	B	6.24	5958	DOE JGI [b]
Cyanothece sp. ATCC 51142	CP000806.1	B	5.46	5304	(Welsh et al., 2008)
Cyanothece sp. PCC 7424[f]	CP001291.1	B	6.55	5710	DOE JGI [b]
Cyanothece sp. PCC 7425[f]	CP001344.1	B	5.79	5327	DOE JGI [b]
Cyanothece sp. PCC 7822[f]	CP002198.1	B	7.84	6642	DOE JGI [b]
Cyanothece sp. PCC 8801	CP001287.1	B	4.79	4367	DOE JGI [b]
Cyanothece sp. PCC 8802[f]	CP001701.1	B	4.8	4444	DOE JGI [b]
Cylindrospermopsis raciborskii CS-505[a,f]	ACYA00000000.1	B	3.88	3449	FLI Jena [c]
Fischerella sp. JSC-11[a,f]	AGIZ00000000.1	B	5.38	4627	DOE JGI [b]
Gloeobacter violaceus PCC 7421	BA000045.2	A	4.66	4430	(Nakamura et al., 2003)
Microcystis aeruginosa NIES-843	AP009552.1	B	5.84	6312	(Kaneko et al., 2007)
Nodularia spumigena CCY 9414[a]	AAVW00000000.1	B	5.32	4860	GBM Foundation [d]
'Nostoc azollae' 0708[f]	CP002059.1	B	5.49	3,651	(Ran et al., 2010)
Nostoc punctiforme PCC 73102	CP001037.1	B	9.06	6689	DOE JGI [b]
Nostoc sp. PCC 7120	BA000019.2	B	7.21	6129	(Kaneko et al., 2001)
Prochlorococcus marinus AS 9601	CP000551.1	C	1.67	1920	GBM Foundation [d]
Prochlorococcus marinus CCMP1375	AE017126.1	C	1.75	1883	(Dufresne et al., 2003)
Prochlorococcus marinus CCMP1986	BX548174.1	C	1.66	1717	(Rocap et al., 2003)
Prochlorococcus marinus MIT 9211	CP000878.1	C	1.69	1854	GBM Foundation [d]
Prochlorococcus marinus MIT 9215	CP000825.1	C	1.74	1982	DOE JGI [b]
Prochlorococcus marinus MIT 9301	CP000576.1	C	1.64	1906	GBM Foundation [d]
Prochlorococcus marinus MIT 9303	CP000554.1	C	2.68	2997	GBM Foundation [d]
Prochlorococcus marinus MIT 9312	CP000111.1	C	1.71	1810	DOE JGI [b]
Prochlorococcus marinus MIT 9313	BX548175.1	C	2.41	2269	DOE JGI [b]
Prochlorococcus marinus MIT 9515	CP000552.1	C	1.7	1905	GBM Foundation [d]
Prochlorococcus marinus NATL1A	CP000553.1	C	1.86	2193	GBM Foundation [d]
Prochlorococcus marinus NATL2A	CP000095.2	C	1.84	2162	DOE JGI [b]
Raphidiopsis brookii D9[f]	CP000095.2	B	3.19	3007	DOE JGI [b]
Synechococcus elongatus PCC 6301	AP008231.1	B	2.7	2523	(Sugita et al., 2007)
Synechococcus elongatus PCC 7942	CP000100.1	B	2.74	2662	DOE JGI [b]
Synechococcus sp. CC9311	CP000435.1	C	2.61	2892	(Palenik et al., 2006)
Synechococcus sp. CC9605	CP000110.1	C	2.51	2645	DOE JGI [b]
Synechococcus sp. CC9902	CP000097.1	C	2.23	2306	DOE JGI [b]
Synechococcus sp. JA-2-3B'a(2-13)	CP000240.1	A	3.05	2862	TIGR [e]
Synechococcus sp. JA-3-3Ab	CP000239.1	A	2.93	2760	TIGR [e]
Synechococcus sp. PCC 7002	CP000951.1	B	3.41	3187	Penn. State University
Synechococcus sp. RCC307	CT978603.1	C	2.22	2534	Institut Pasteur
Synechococcus sp. RS9917[a]	AANP00000000	C	2.58	2770	GBM Foundation [d]
Synechococcus sp. WH 5701[a]	AANO00000000	C	3.04	3346	GBM Foundation [d]
Synechococcus sp. WH 7803	CT971583.1	C	2.37	2533	Institut Pasteur
Synechococcus sp. WH 7805[a]	AAOK00000000	C	2.62	2883	GBM Foundation [d]
Synechococcus sp. WH 8102	BX548020.1	C	2.43	2519	(Palenik et al., 2003)
Synechocystis sp. PCC 6803	BA000022.2	B	3.95	3575	(Kaneko et al., 1996)
Thermosynechococcus elongatus BP-1	BA000039.2	B	2.59	2476	(Nakamura et al., 2002)
Trichodesmium erythraeum 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*, *Crocosphaera*, *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 *Nostocalean* 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 thymidylyltransferase (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 *R. Brookii D9*, *C. raciborskii CS-505* and *'N. azollae' 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 *R. Brookii D9* and *C. raciborskii CS-505*				
O-succinylbenzoyl-CoA synthetase	282901079[C]	Fig # 6B	7 aa ins	278-321
Prolipoprotein diacylglyceryl 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 diacylglyceryl 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 *N. sp. PCC 7120, N. spumigena, N. punctiforme PCC 73102* and *A. variabilis ATCC 29413*				
Glucose-1-phosphate thymidylyltransferase	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 *N. sp. PCC 7120, N. punctiforme PCC 73102* and *A. variabilis ATCC 29413*				
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 *N. sp. PCC 7120* and *A. variabilis ATCC 29413*				
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 *Nostocales/Stigonematales*			
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 *Nostocales*			
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 N. PCC 7120, A. variabilis ATCC 2913, N. punctiforme PCC 73102 & N. spumigenaCCY9414			
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 N. PCC 7120, A. variabilis ATCC 2913 & N. punctiforme PCC 73102			
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 N. PCC 7120 & A. variabilis ATCC 2913			
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; Euzeby, 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 *Gloeobacteriphycidae,* 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).

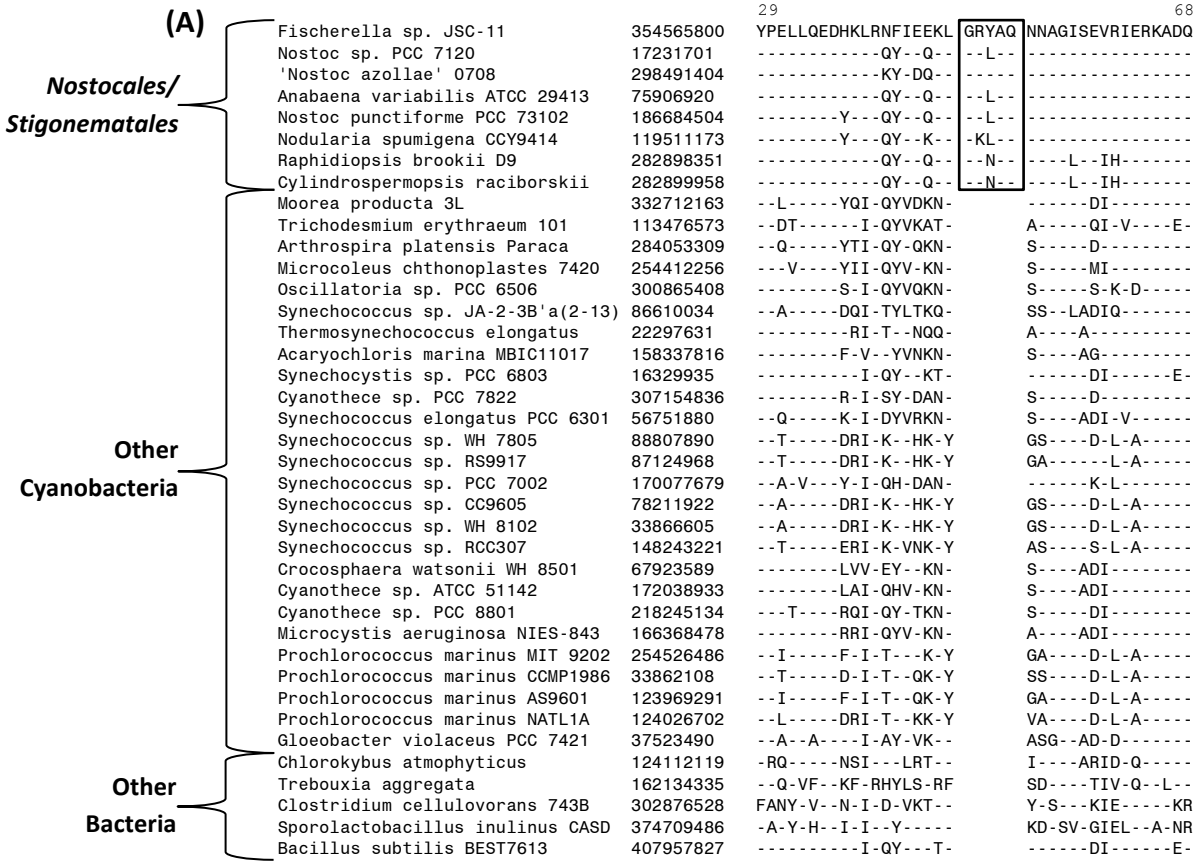
4.4 Conserved Signature Indels Specific to Individual Members of Sequenced *Nostocales*/*Stigonematales*

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 a 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.


(A)
Nostocales/
Stigonematales
Other
Cyanobacteria
Other
Bacteria
29
68
Fischerella sp. JSC-11 354565800 YPELLQEDHKLRNFIEEKL 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-- ----L--IH-------
Cylindrospermopsis raciborskii 282899958 ----------QY--Q-- --N-- ----L--IH-------
Moorea producta 3L 332712163 --L-----YQI-QYVDKN- ------DI--------
Trichodesmium erythraeum 101 113476573 --DT------I-QYVKAT- A-----QI-V----E-
Arthrospira platensis Paraca 284053309 --Q-----YTI-QY-QKN- S-----D---------
Microcoleus chthonoplastes 7420 254412256 ---V----YII-QYV-KN- S-----MI--------
Oscillatoria sp. PCC 6506 300865408 --------S-I-QYVQKN- S-----S-K-D-----
Synechococcus sp. JA-2-3B'a(2-13) 86610034 --A-----DQI-TYLTKQ- SS--LADIQ-------
Thermosynechococcus elongatus 22297631 ---------RI-T--NQQ- A----A----------
Acaryochloris marina MBIC11017 158337816 --------F-V--YVNKN- S----AG---------
Synechocystis sp. PCC 6803 16329935 ----------I-QY--KT- ------DI------E-
Cyanothece sp. PCC 7822 307154836 --------R-I-SY-DAN- S-----D---------
Synechococcus elongatus PCC 6301 56751880 --Q-----K-I-DYVRKN- S----ADI-V------
Synechococcus sp. WH 7805 88807890 --T-----DRI-K--HK-Y GS----D-L-A-----
Synechococcus sp. RS9917 87124968 --T-----DRI-K--HK-Y GA------L-A-----
Synechococcus sp. PCC 7002 170077679 --A-V---Y-I-QH-DAN- ------K-L-------
Synechococcus sp. CC9605 78211922 --A-----DRI-K--HK-Y GS----D-L-A-----
Synechococcus sp. WH 8102 33866605 --A-----DRI-K--HK-Y GS----D-L-A-----
Synechococcus sp. RCC307 148243221 --T-----ERI-K-VNK-Y AS----S-L-A-----
Crocosphaera watsonii WH 8501 67923589 --------LVV-EY--KN- S----ADI--------
Cyanothece sp. ATCC 51142 172038933 --------LAI-QHV-KN- S----ADI--------
Cyanothece sp. PCC 8801 218245134 ---T----RQI-QY-TKN- S-----DI--------
Microcystis aeruginosa NIES-843 166368478 --------RRI-QYV-KN- A----ADI--------
Prochlorococcus marinus MIT 9202 254526486 --I-----F-I-T---K-Y GA----D-L-A-----
Prochlorococcus marinus CCMP1986 33862108 --T-----D-I-T--QK-Y SS----D-L-A-----
Prochlorococcus marinus AS9601 123969291 --I-----F-I-T--QK-Y GA----D-L-A-----
Prochlorococcus marinus NATL1A 124026702 --L-----DRI-T--KK-Y VA----D-L-A-----
Gloeobacter violaceus PCC 7421 37523490 --A--A----I-AY-VK-- ASG--AD-D-------
Chlorokybus atmophyticus 124112119 -RQ-----NSI---LRT-- I----ARID-Q-----
Trebouxia aggregata 162134335 --Q-VF--KF-RHYLS-RF SD----TIV-Q--L--
Clostridium cellulovorans 743B 302876528 FANY-V--N-I-D-VKT-- Y-S---KIE-----KR
Sporolactobacillus inulinus CASD 374709486 -A-Y-H--I-I--Y----- KD-SV-GIEL--A-NR
Bacillus subtilis BEST7613 407957827 ----------I-QY---T- ------DI------E-

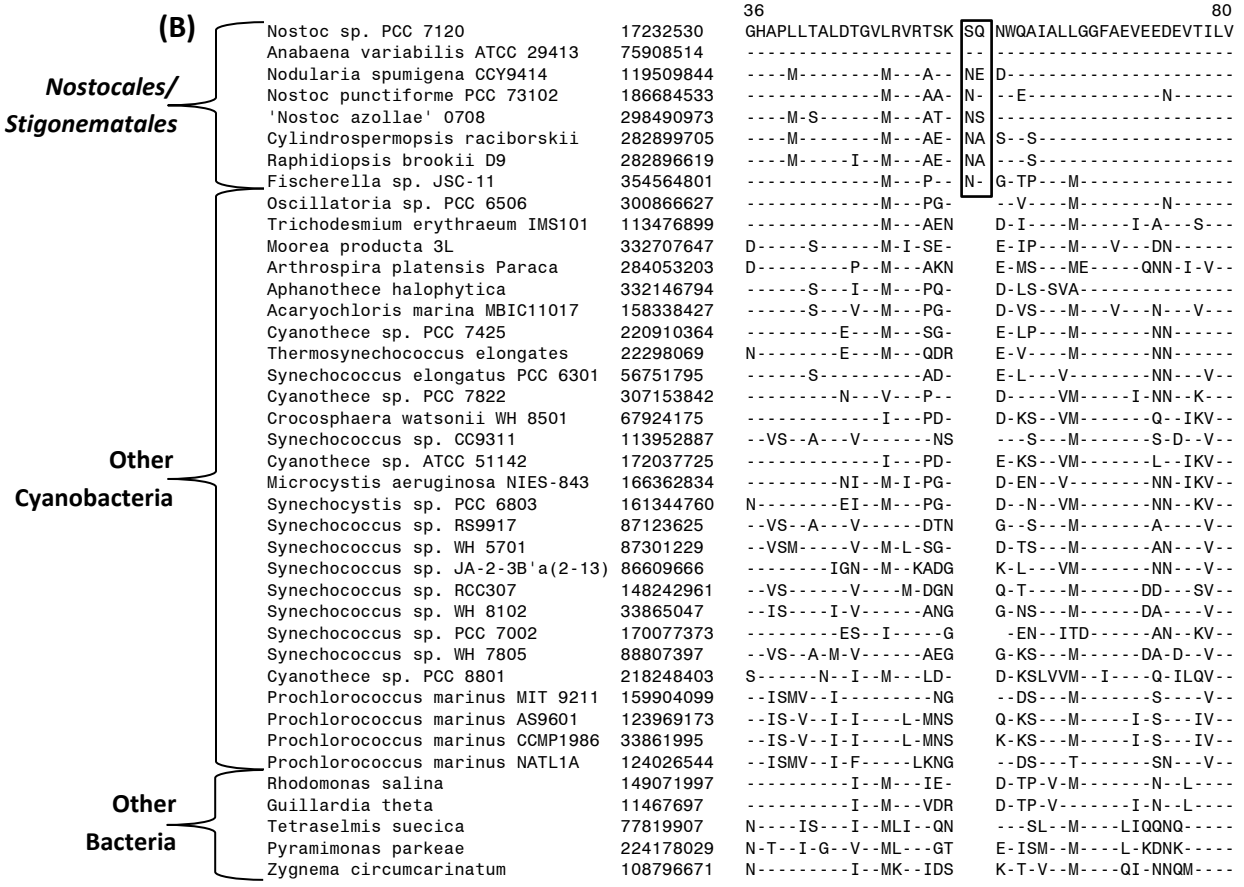
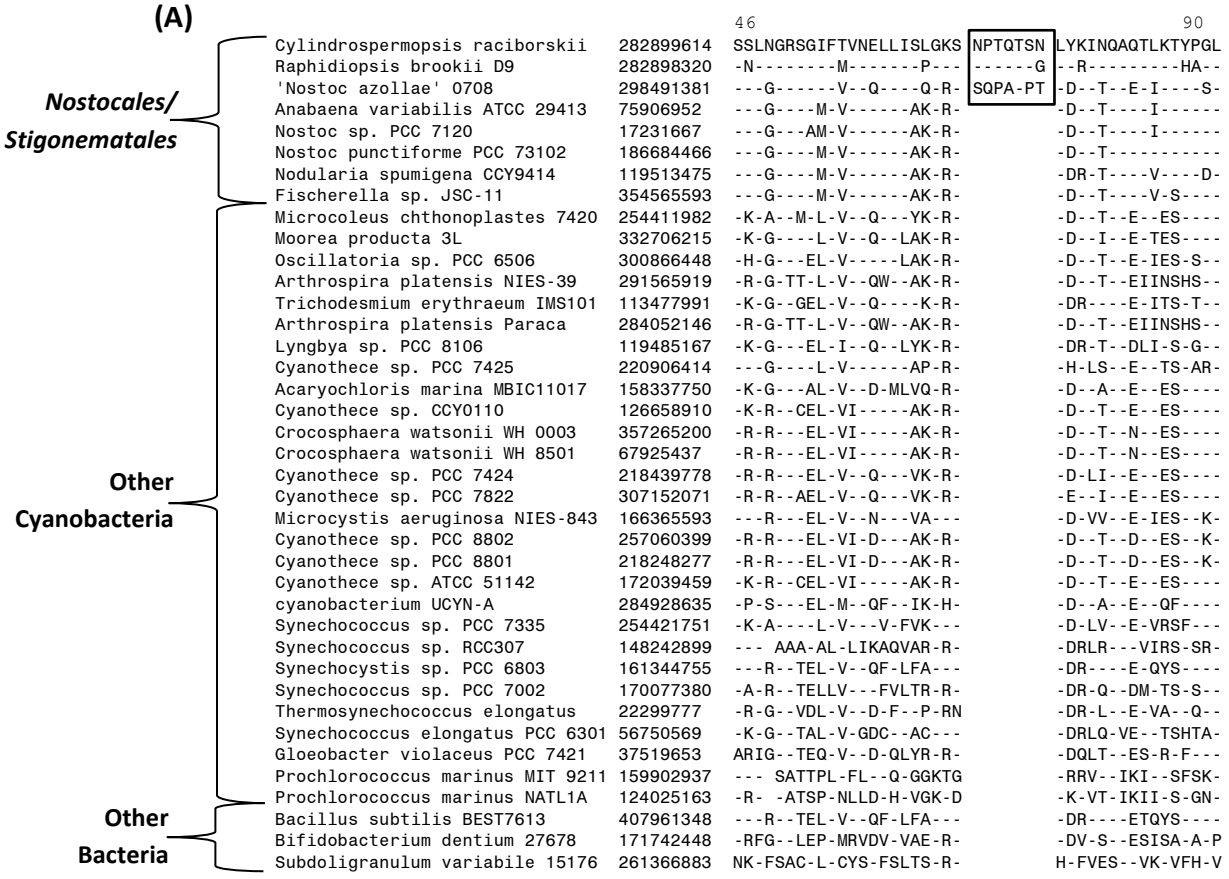


Figure 5: Partial sequence alignments for the proteins (**A**) 30S ribosomal protein S3, showing a 5 aa insertion and (**B**) ATP synthase epsilon subunit, showing a 2 aa insertion, both flanked by conserved regions that are specifically present in all sequenced *Nostocales/Stigonematales* members. Dashes 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
Other
Cyanobacteria
Other
Bacteria
Cylindrospermopsis raciborskii 282899614 SSLNGRSGIFTVNELLISLGKS NPTQTSN LYKINQAQTLKTYPGL
Raphidiopsis brookii D9 282898320 -N--------M-------P--- ------G --R---------HA--
'Nostoc azollae' 0708 298491381 ---G------V--Q----Q-R- SQPA-PT -D--T--E-I----S-
Anabaena variabilis ATCC 29413 75906952 ---G----M-V------AK-R- -D--T----I------
Nostoc sp. PCC 7120 17231667 ---G---AM-V------AK-R- -D--T----I------
Nostoc punctiforme PCC 73102 186684466 ---G----M-V------AK-R- -D--T----------
Nodularia spumigena CCY9414 119513475 ---G----M-V------AK-R- -DR-T----V----D-
Fischerella sp. JSC-11 354565593 ---G----M-V------AK-R- -D--T----V-S----
Microcoleus chthonoplastes 7420 254411982 -K-A--M-L-V--Q---YK-R- -D--T--E--ES----
Moorea producta 3L 332706215 -K-G----L-V--Q--LAK-R- -D--I--E-TES----
Oscillatoria sp. PCC 6506 300866448 -H-G---EL-V-----LAK-R- -D--T--E-IES-S--
Arthrospira platensis NIES-39 291565919 -R-G-TT-L-V--QW--AK-R- -D--T--EIINSHS--
Trichodesmium erythraeum IMS101 113477991 -K-G--GEL-V--Q----K-R- -DR----E-ITS-T--
Arthrospira platensis Paraca 284052146 -R-G-TT-L-V--QW--AK-R- -D--T--EIINSHS--
Lyngbya sp. PCC 8106 119485167 -K-G---EL-I--Q--LYK-R- -DR-T--DLI-S-G--
Cyanothece sp. PCC 7425 220906414 ---G----L-V------AP-R- -H-LS--E--TS-AR-
Acaryochloris marina MBIC11017 158337750 -K-G---AL-V--D-MLVQ-R- -D--A--E--ES----
Cyanothece sp. CCY0110 126658910 -K-R--CEL-VI-----AK-R- -D--T--E--ES----
Crocosphaera watsonii WH 0003 357265200 -R-R---EL-VI-----AK-R- -D--T--N--ES----
Crocosphaera watsonii WH 8501 67925437 -R-R---EL-VI-----AK-R- -D--T--N--ES----
Cyanothece sp. PCC 7424 218439778 -R-R---EL-V--Q---VK-R- -D-LI--E--ES----
Cyanothece sp. PCC 7822 307152071 -R-R--AEL-V--Q---VK-R- -E--I--E--ES----
Microcystis aeruginosa NIES-843 166365593 ---R---EL-V--N---VA--- -D-VV--E-IES--K-
Cyanothece sp. PCC 8802 257060399 -R-R---EL-VI-D---AK-R- -D--T--D--ES--K-
Cyanothece sp. PCC 8801 218248277 -R-R---EL-VI-D---AK-R- -D--T--D--ES--K-
Cyanothece sp. ATCC 51142 172039459 -K-R--CEL-VI-----AK-R- -D--T--E--ES----
cyanobacterium UCYN-A 284928635 -P-S---EL-M--QF--IK-H- -D--A--E--QF----
Synechococcus sp. PCC 7335 254421751 -K-A----L-V---V-FVK--- -D-LV--E-VRSF---
Synechococcus sp. RCC307 148242899 --- AAA-AL-LIKAQVAR-R- -DRLR---VIRS-SR-
Synechocystis sp. PCC 6803 161344755 ---R--TEL-V--QF-LFA--- -DR----E-QYS----
Synechococcus sp. PCC 7002 170077380 -A-R--TELLV---FVLTR-R- -DR-Q--DM-TS-S--
Thermosynechococcus elongatus 22299777 -R-G--VDL-V--D-F--P-RN -DR-L--E-VA--Q--
Synechococcus elongatus PCC 6301 56750569 -K-G--TAL-V-GDC--AC--- -DRLQ-VE--TSHTA-
Gloeobacter violaceus PCC 7421 37519653 ARIG--TEQ-V--D-QLYR-R- -DQLT--ES-R-F---
Prochlorococcus marinus MIT 9211 159902937 --- SATTPL-FL--Q-GGKTG -RRV--IKI--SFSK-
Prochlorococcus marinus NATL1A 124025163 -R- -ATSP-NLLD-H-VGK-D -K-VT-IKII-S-GN-
Bacillus subtilis BEST7613 407961348 ---R--TEL-V--QF-LFA--- -DR----ETQYS----
Bifidobacterium dentium 27678 171742448 -RFG--LEP-MRVDV-VAE-R- -DV-S--ESISA-A-P
Subdoligranulum variabile 15176 261366883 NK-FSAC-L-CYS-FSLTS-R- H-FVES--VK-VFH-V

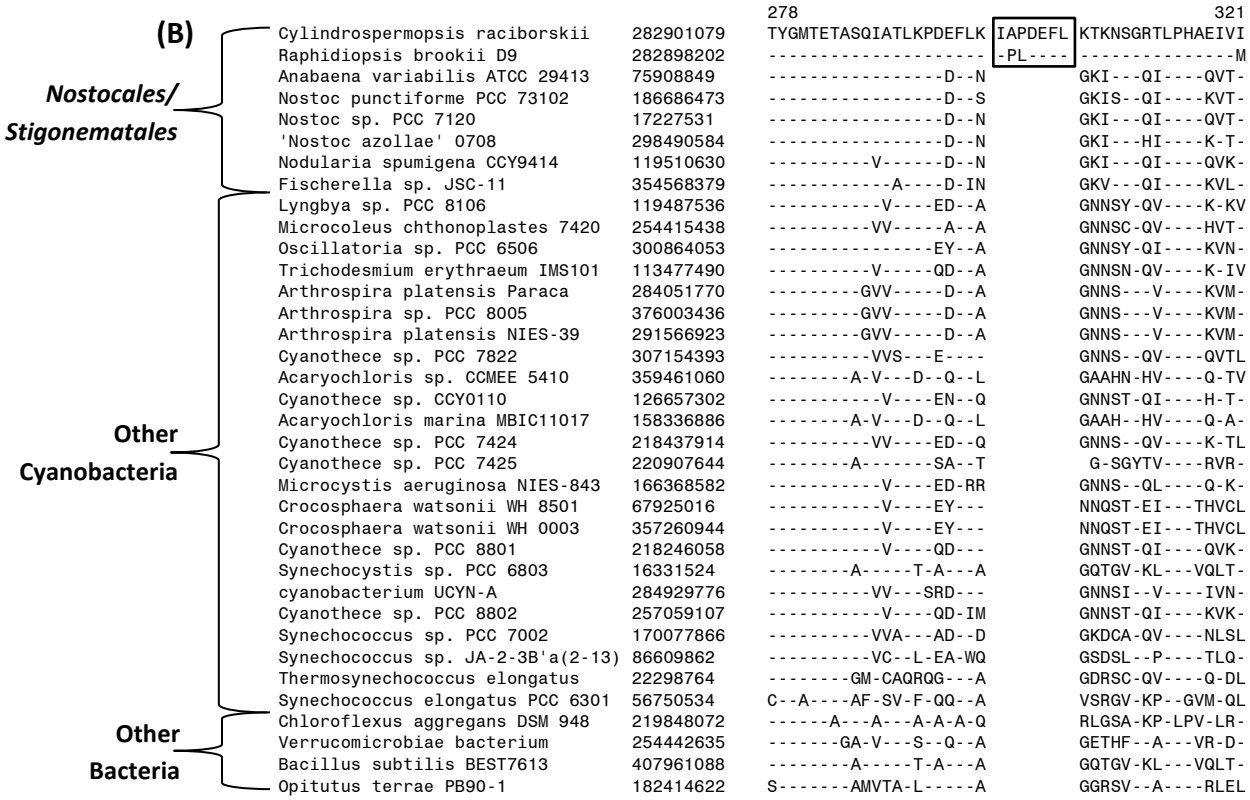
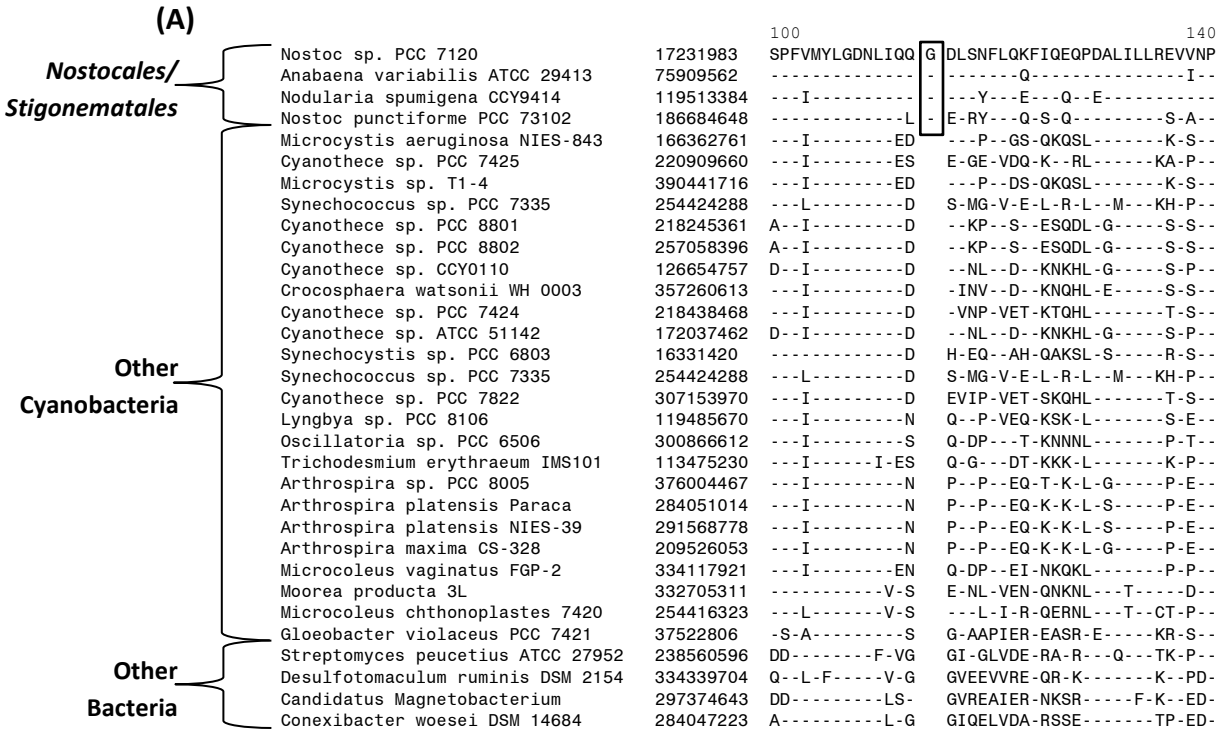


Figure 6: Partial sequence alignments for the proteins (**A**) Recombination protein O, showing a 7 aa insertion and (**B**) O-succinylbenzoyl-CoA synthetase, showing a 7 aa insertion, both flanked by conserved regions that are specifically present in all members which represent a subclade/close affinity 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)
Nostocales/
Stigonematales
Other
Cyanobacteria
Other
Bacteria
100
140
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--
Cyanothece sp. CCY0110 126654757 D--I---------D --NL--D--KNKHL-G-----S-P--
Crocosphaera 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--
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 GVEEVVRE-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-

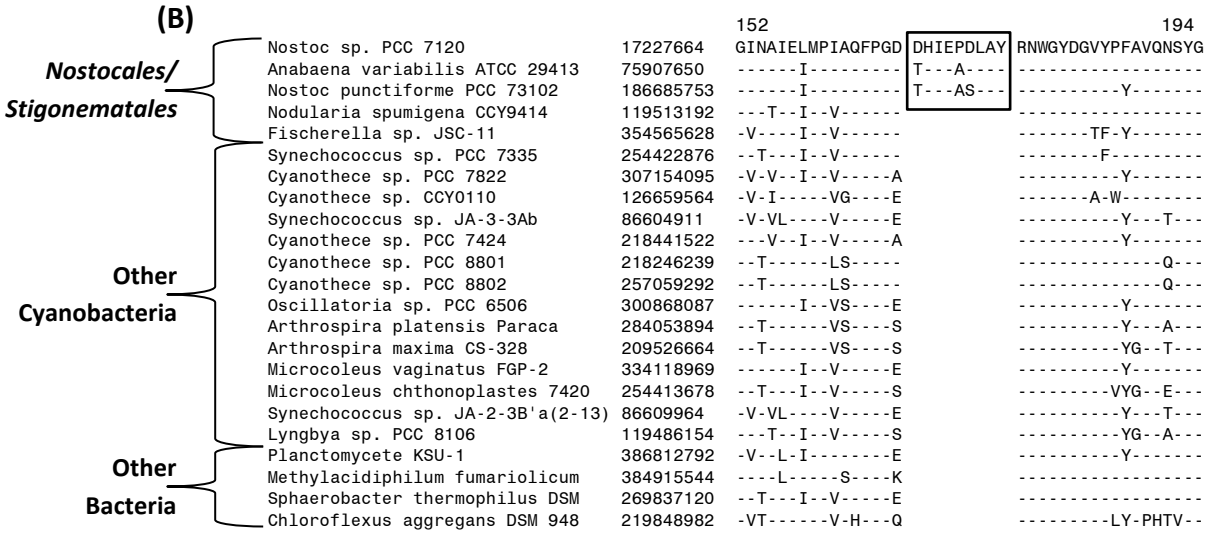


Figure 7: Partial sequence alignments for the proteins (**A**) Glucose-1-phosphate thymidylyltransferase, showing a 1 aa insertion and (**B**) Alpha-amylase, showing a 9 aa insertion, both flanked by conserved regions that are specifically present in all members belonging to two consecutive subclades 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.

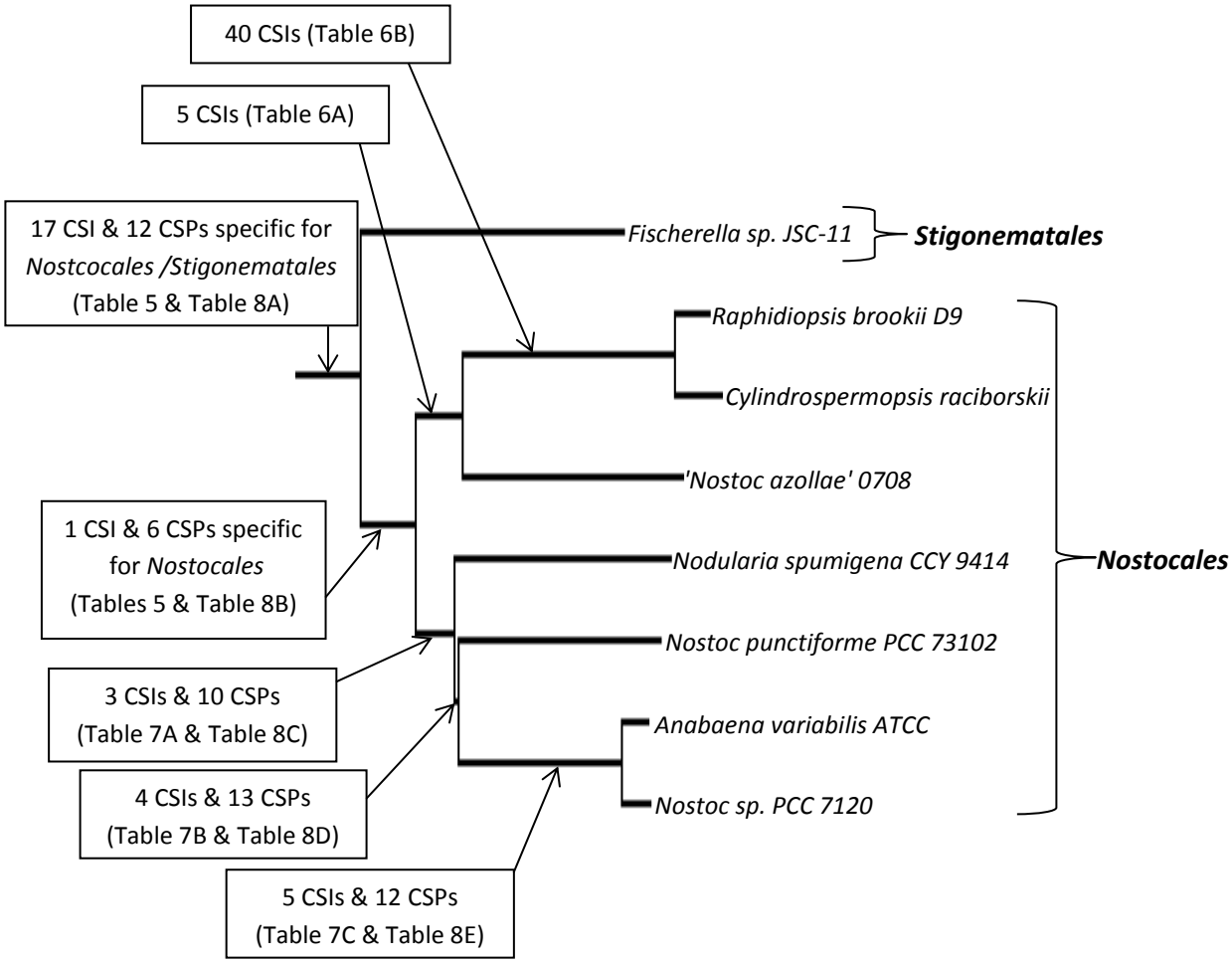


Figure 8: A summary diagram showing the distribution of identified CSIs and CSPs specific to various clades within *Nostocales/Stigonematales*. The identified clades supported by both phylogenetic studies as well as various molecular markers are indicated. CSIs specific for only one sequenced specifies 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* then *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 SystematicBacteriology. (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 Asscociation *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). Cyanophycea. 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 archaebacteria, 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., Hirosawa,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). Cylindrospermopsin: a potent hepatotoxin from the blue-green alga Cylindrospermopsis raciborskii. Journal of the American Chemical Society *114*, 7941-7942.

Padisak,J. (1997). Cylindrospermopsis 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., Brahamsha,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 Gl+Âckner,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 S *2*, 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., Sistrom,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 thealth 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 relationshipsamong the cyanobacteria based on 16S rRNA sequences. In Bergey'sManual 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 (Desikachary, 1959; Stanier et al., 1978; Rippka 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; Euzeby, 1997; Wilmotte, 2004)
- Cyanobacteria are most notorious for their formation of blooms and production of toxins.

16S rRNA

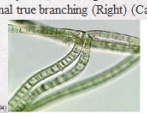
- Currently, the use of 16S rRNA sequence data is one of the most widely implemented approaches to understanding microbial taxonomy (Embley and Stackebrandt, 1997; Rossello-Mora and Amann, 2001)
- Cyanobacterial phylogeny based on 16S rRNA show little correlation with the morphological and biochemical 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; Euzeby, 1997)
  - 3 classes - Chroobacteria, Hormogoneae and Gloeobacteria
  - 6 orders - *Chroococcales*, *Pleurocapsales*, *Oscillatorials*, *Nostocales*, *Stigonematales* and *Gloeobacterales*
  - 5 genera - *Prochlorothrix*, *Prochloron*, *Prochlorococcus*, *Planktothricoides* and *Halospirulina*
- Cyanobacteria taxonomy has several contradictions with phylogeny
  - Extensive polyphyletic branching of the orders *Chroococcales* and *Oscillatorriales* 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) (Castenholz, 2001)
  - Form hormogonia 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) (Castenholz, 2001)
  - Form hormogonia and are branched (Cavalier-Smith, 2002)

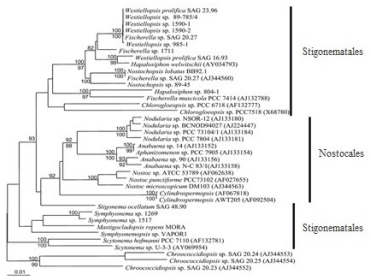


16S rRNA

- 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

16S rRNA Tree

(Brun et al., 2000)



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 & Mathews 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

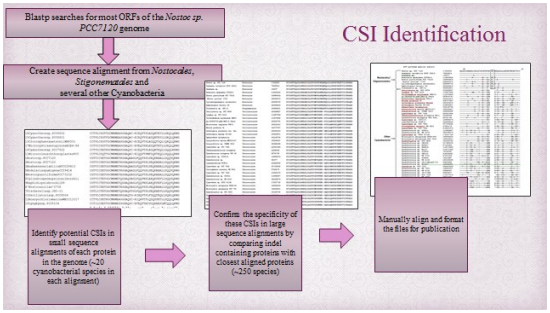
16S rRNA Trees

- Phylogenetic analysis was also performed on a 16S rRNA alignment of sequences:
  - From at least one good quality representative from every genera (except *Riveria*) 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

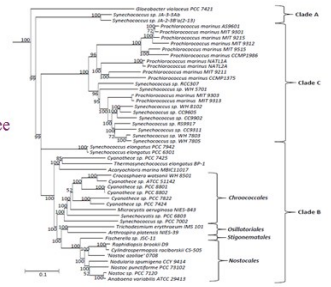
- Dr. Gupta's earlier work has identified a number of CSPs which were shown to be specific for the species of *Nostocales*
- New sequence information for additional *Nostocales/Stigonematales* has now become available
  - Blastp searches were conducted on all identified proteins from his earlier publication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

CSI Identification

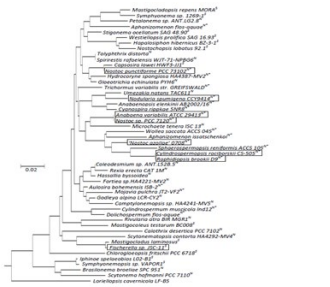


Results

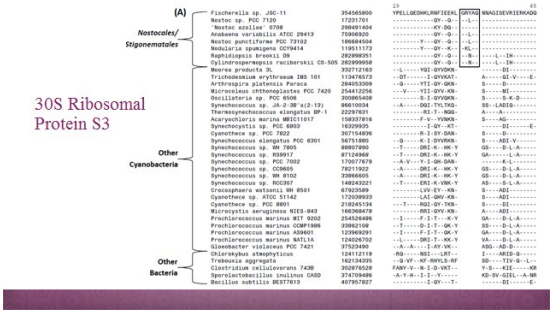
Maximum-Likelihood Tree



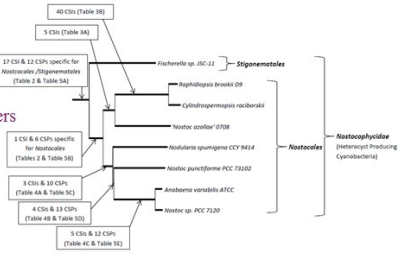
16S rRNA Tree



30S Ribosomal Protein S3



Summary of Molecular Markers



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 intermixed 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:
  - *Cylindrospermopsis raciborskii CS-505*
  - *Raphidiopsis 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 *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

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 CCY9414*
  - *Nostoc punctiforme PCC73102*
  - *Nostoc sp. PCC 7120*
  - *Anabaena variabilis ATCC2914*
- Within this subclade, a more recent divergence among the members is observed:
  - *Nostoc punctiforme PCC 73102*
  - *Nostoc sp. PCC 7120*
  - *Anabaena variabilis ATCC 29413*
- A close affinity is observed between *Nostoc sp. PCC 7120* and *Anabaena variabilis ATCC 29413* 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

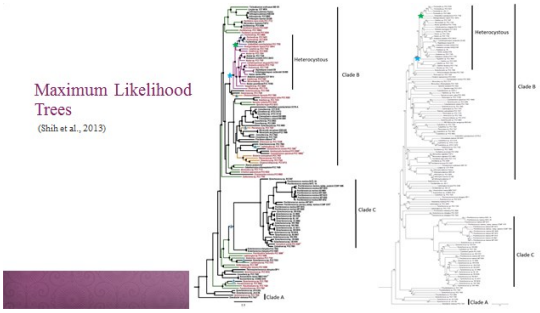
New Data

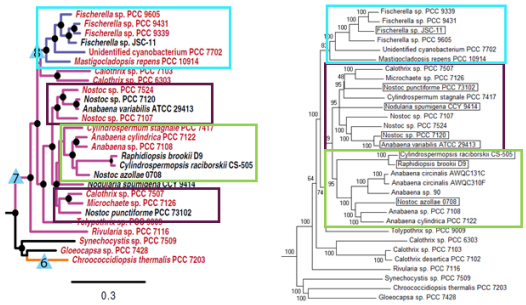
Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing (2013)

Patrick M. Shih, Dongying Wu, Amel Latifi, Seth D. Axen, David P. Fewer, Emmanuel Talla, Alexandra Calteau, Fei Cai, Nicole Tandeau de Marsac, Rosmarie Rippka, Michael Herdman, Kaarina Sivonen, Therese Coursin, Thierry Laurent, Lynne Goodwin, Matt Nolan, Karen W. Davenport, Cliff S. Han, Edward M. Rubin, Jonathan A. Eisen, Tanja Woyke, Muriel Gugger and Cheryl A. Kerfeld

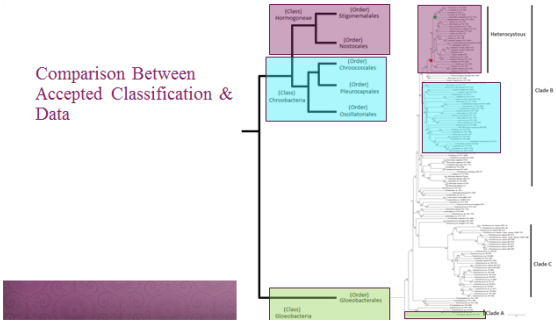
Maximum Likelihood Trees

(Shih et al., 2013)

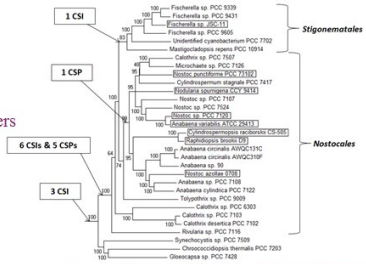




Comparison Between Accepted Classification & Data



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 intermixed relationship
  - Forming a monophyletic group that represents the sole heterocyst differentiating Cyanobacteria

References

- Belcher,A., Holland,H.D., Wang,P.L., Rumble,D.III., Stein,H.J., Hannah,J.L., Coetzee,L.L., and Beukes,N.J. (2004). Dating the rise of atmospheric oxygen. Nature *427*, 117-120.
- Desikachary,T.V. (1959). Cyanophyta, Indian Council of Agricultural Research, monographs on Algae. New Delhi, India.
- 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.
- Stanier,R.Y., Sistrom,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.
- Wilmotte,A. (2004). Molecular evolution and taxonomy of the cyanobacteria. Springer), pp. 1-25.
- 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.
- Embley,T.M. and Stackebrandt,E. (1997). Species in practice: exploring uncultured prokaryote diversity in natural samples. SYSTEMATICS ASSOCIATION SPECIAL VOLUME *54*, 61-82.
- 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.

References

- Rossello-Mora,R. and Amann,R. (2001). The species concept for prokaryotes. FEMS microbiology reviews *25*, 39-67.
- Castenholz,R.W. (2001). Phylum BX. Cyanobacteria. Springer).
- 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.
- Brun, Y. V., & Shimkets, L. J. (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.
- Gugger,M.F. and Hoffmann,L. (2004). Polyphyly of true branching cyanobacteria (*Stigonematales*). International Journal of Systematic and Evolutionary Microbiology *54*, 349-357.
- Patrick M. Shih, Dongying Wu, Amel Latifi, Seth D. Axen, David P. Fewer, Emmanuel Talla, Alexandra Calteau, Fei Cai, Nicole Tandeau de Marsac, Rosmarie Rippka, Michael Herdman, Kaarina Sivonen, Therese Coursin, Thierry Laurent, Lynne Goodwin, Matt Nolan, Karen W. Davenport, Cliff S. Han, Edward M. Rubin, Jonathan A. Eisen, Tanja Woyke, Muriel Gugger 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 cytidylyltransferase	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

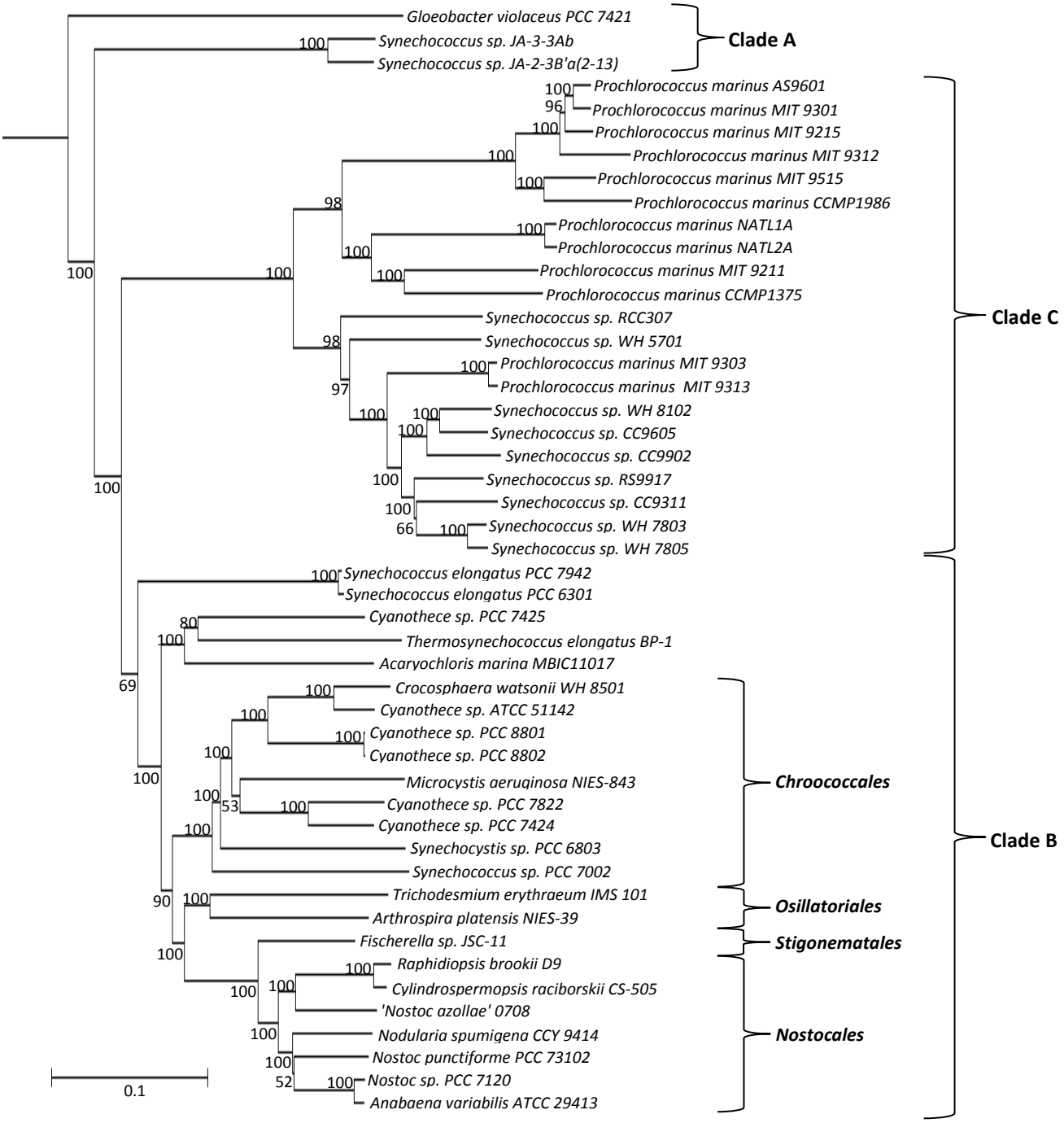
Protein lengths are associated with Nostoc sp. PCC 7120

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

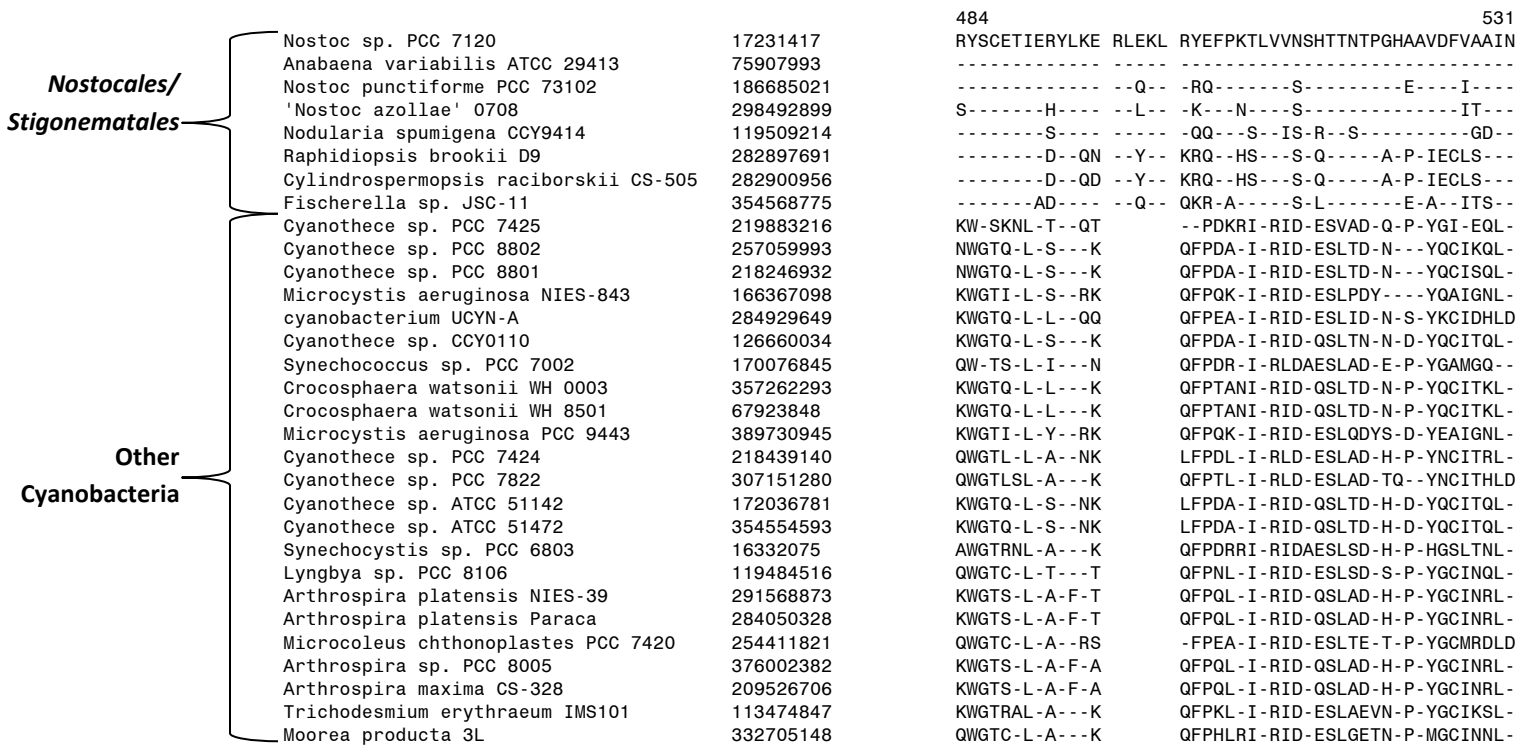
Twin-arginine translocation pathway signal	119511267	S. Fig # 104	10 aa ins	209-258
Hypothetical protein N9414_02991	119511526	S. Fig # 105	4 aa ins	213-267
Photosystem II chlorophyll-binding protein CP47	119509033	S. Fig # 106	1 aa ins	64-98
HAD-superfamily hydrolase subfamily IA, variant 3	119510625	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



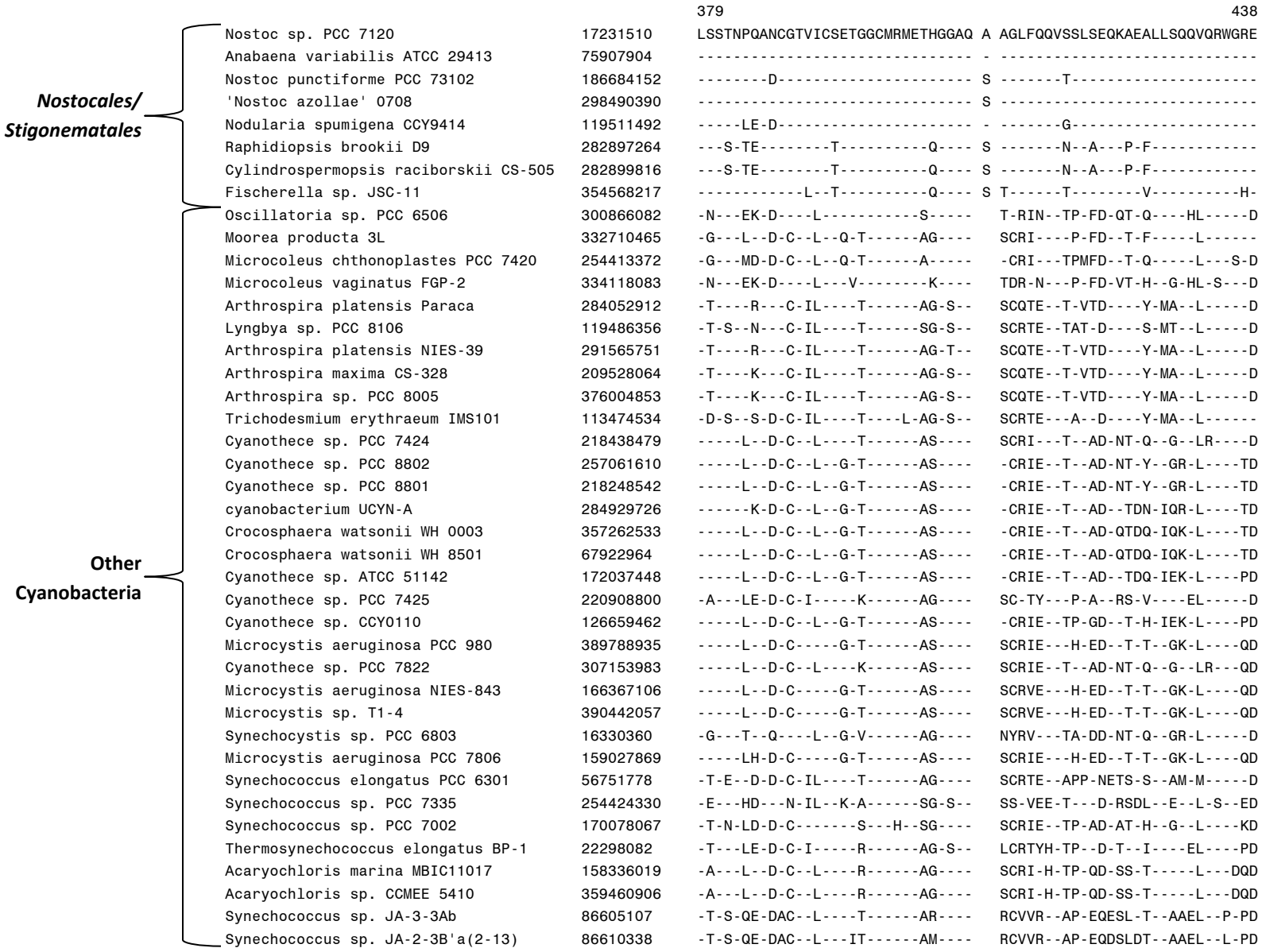
Supplementary Figure 1

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.



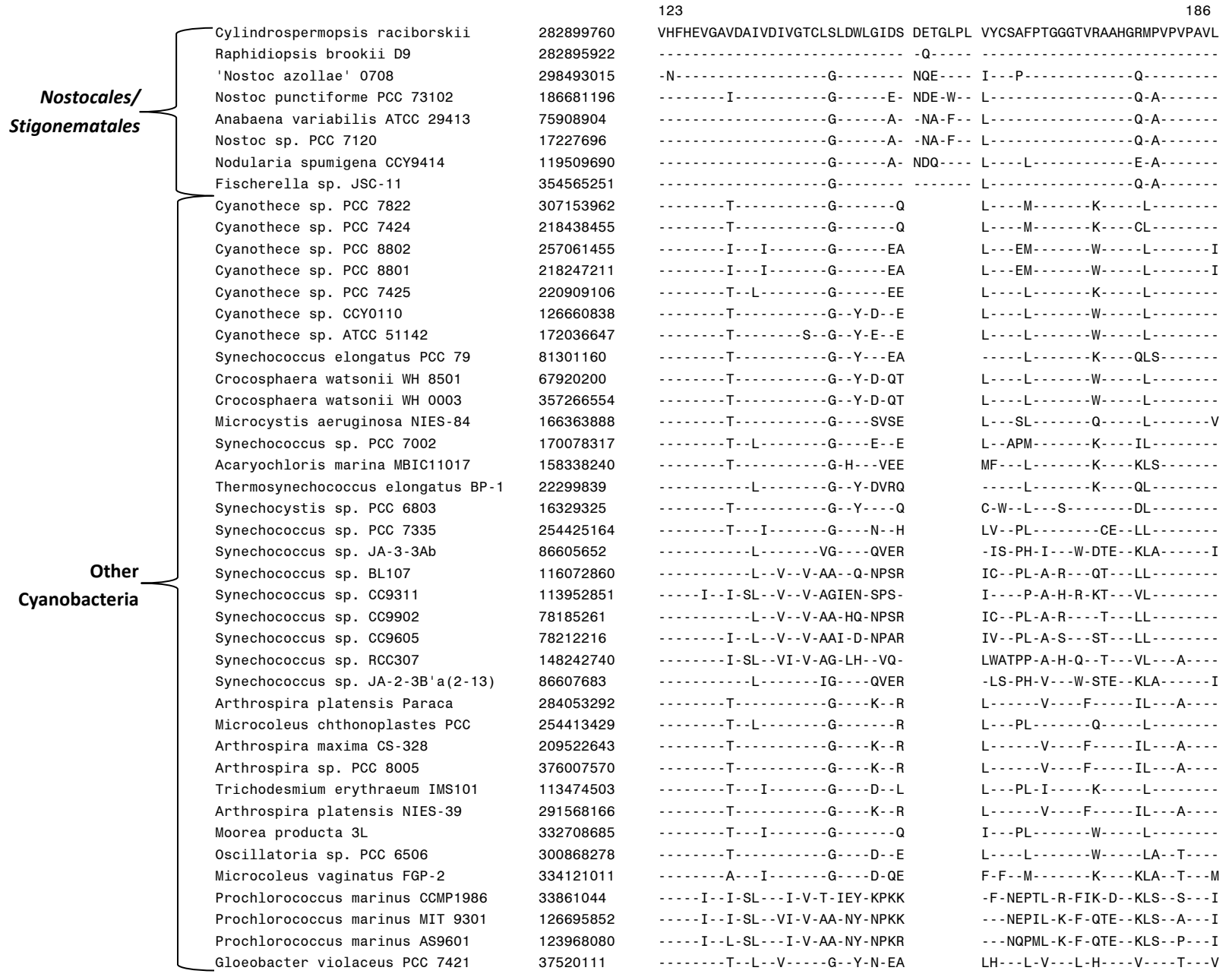
Supplementary Figure 2

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



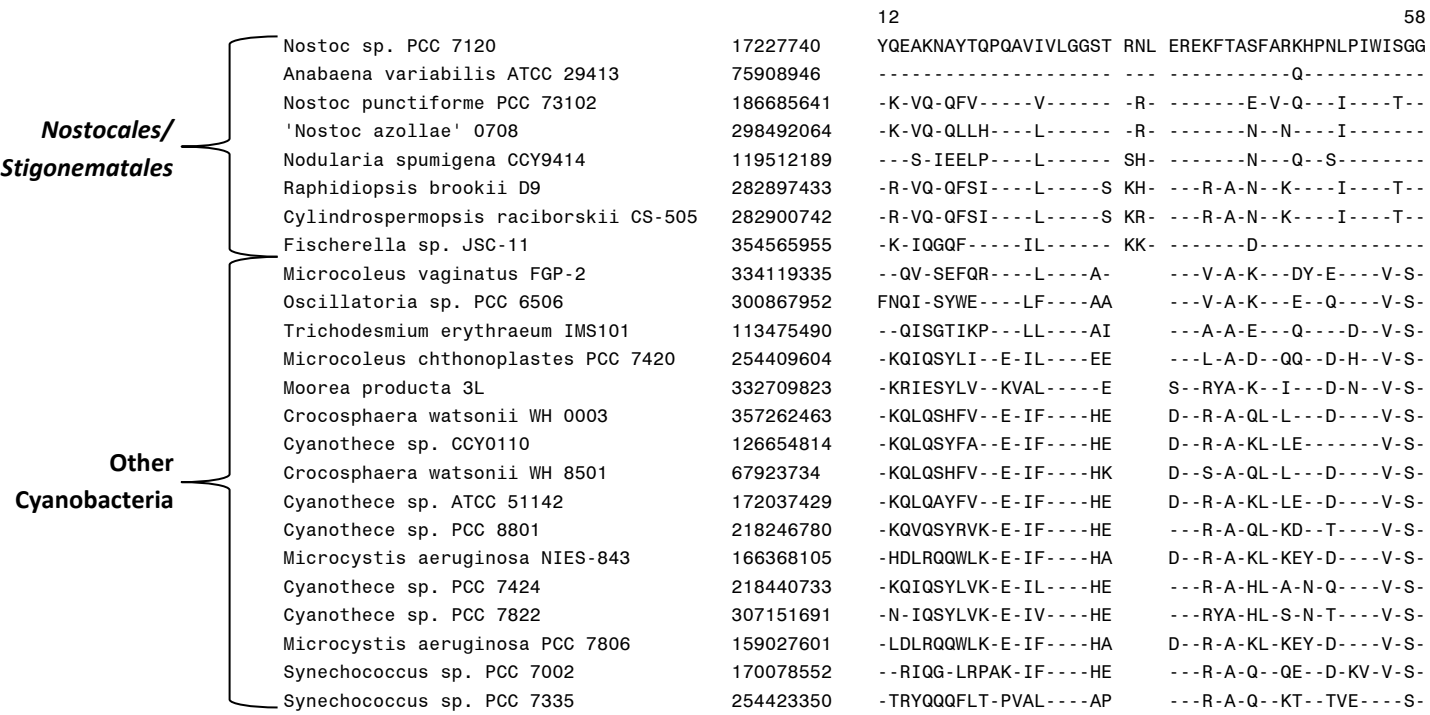
Supplementary Figure 3

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



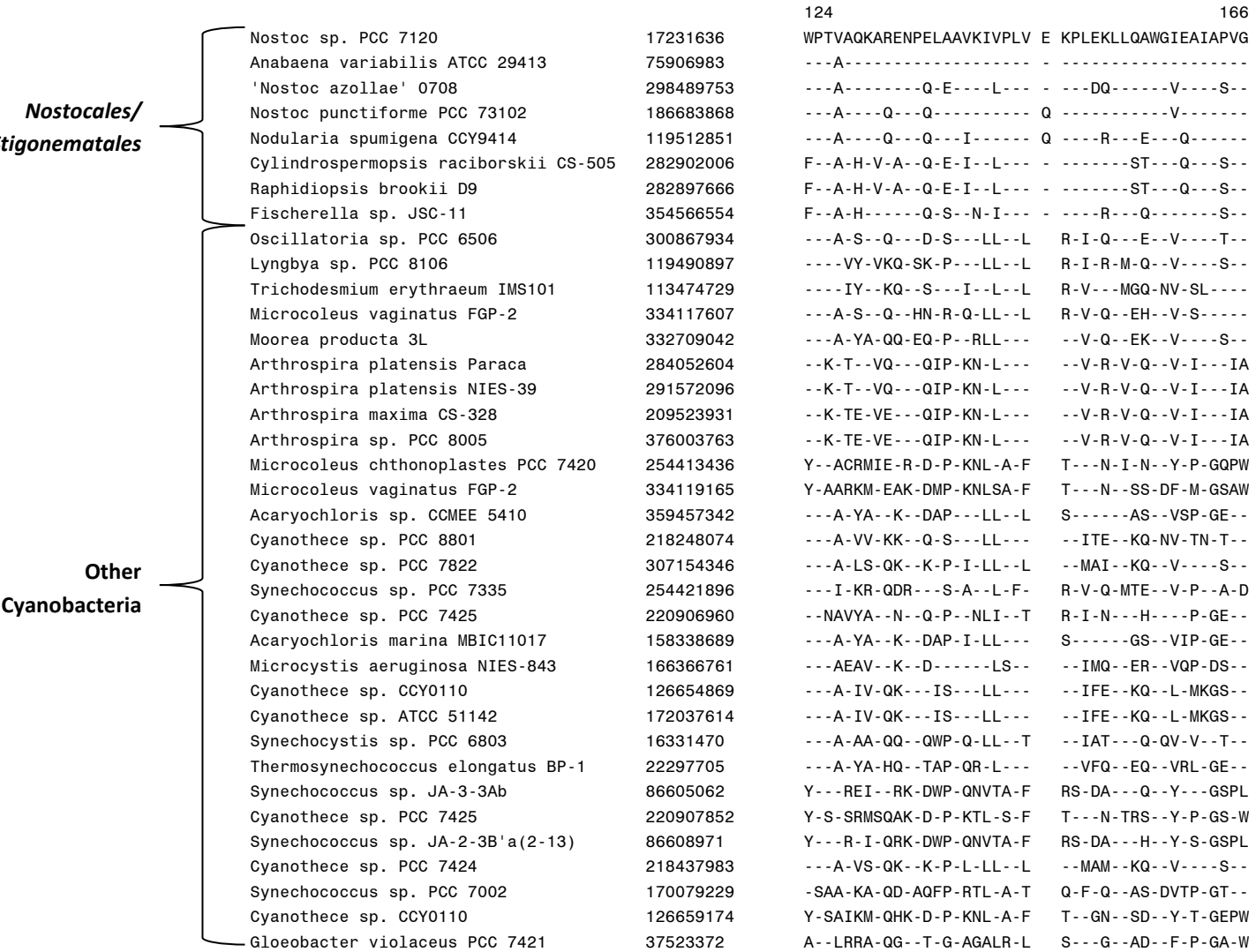
Supplementary Figure 4

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



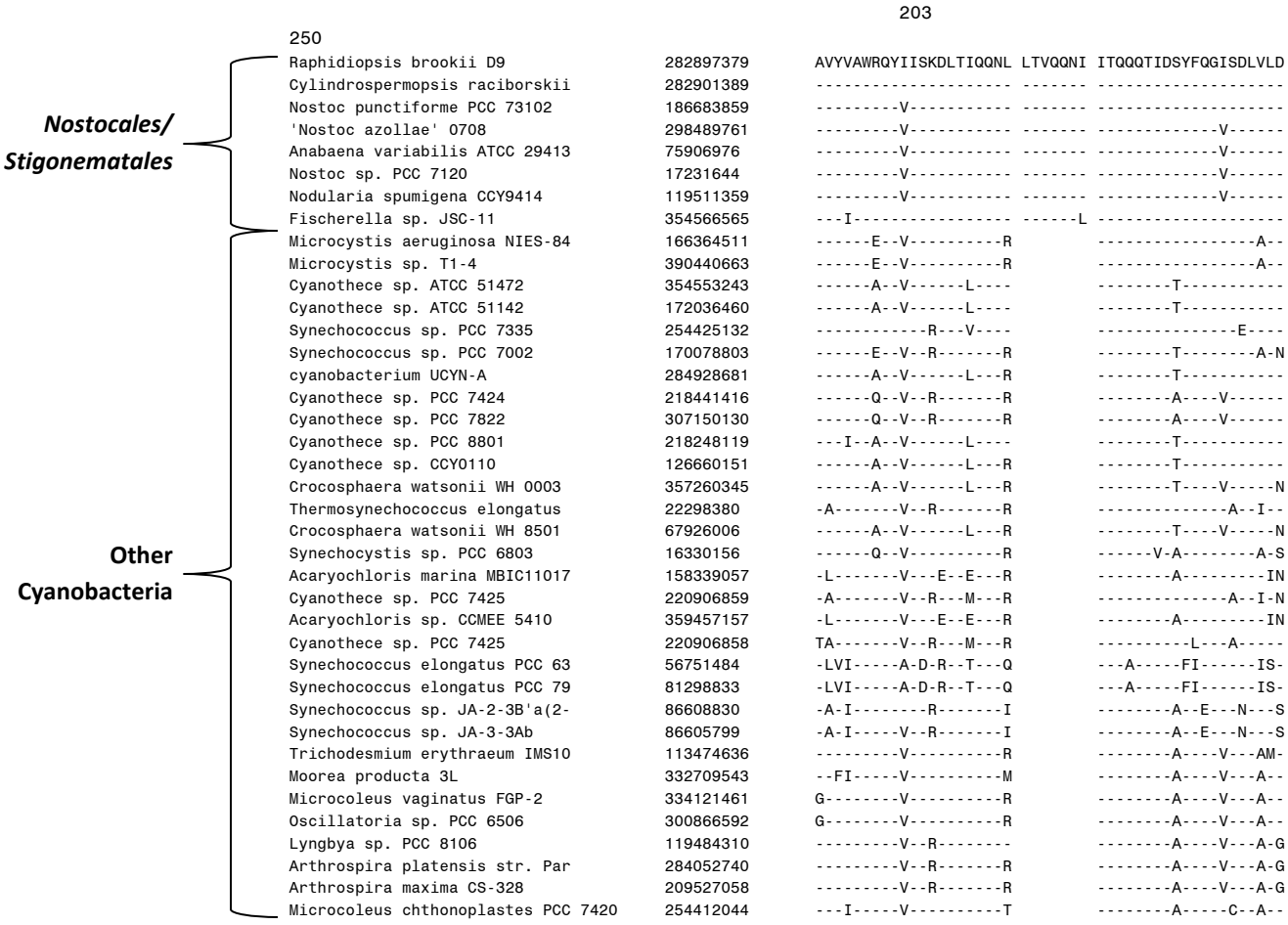
Supplementary Figure 5

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



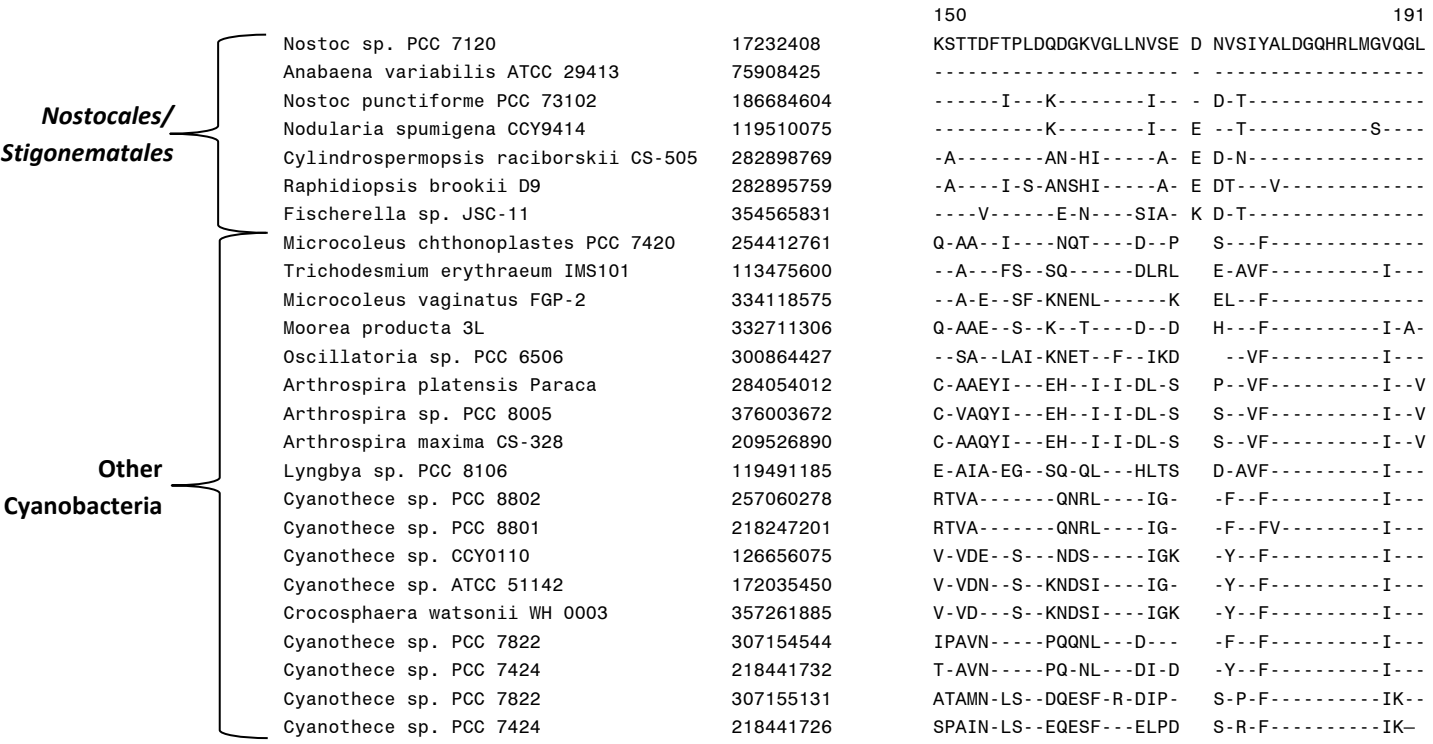
Supplementary Figure 6

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



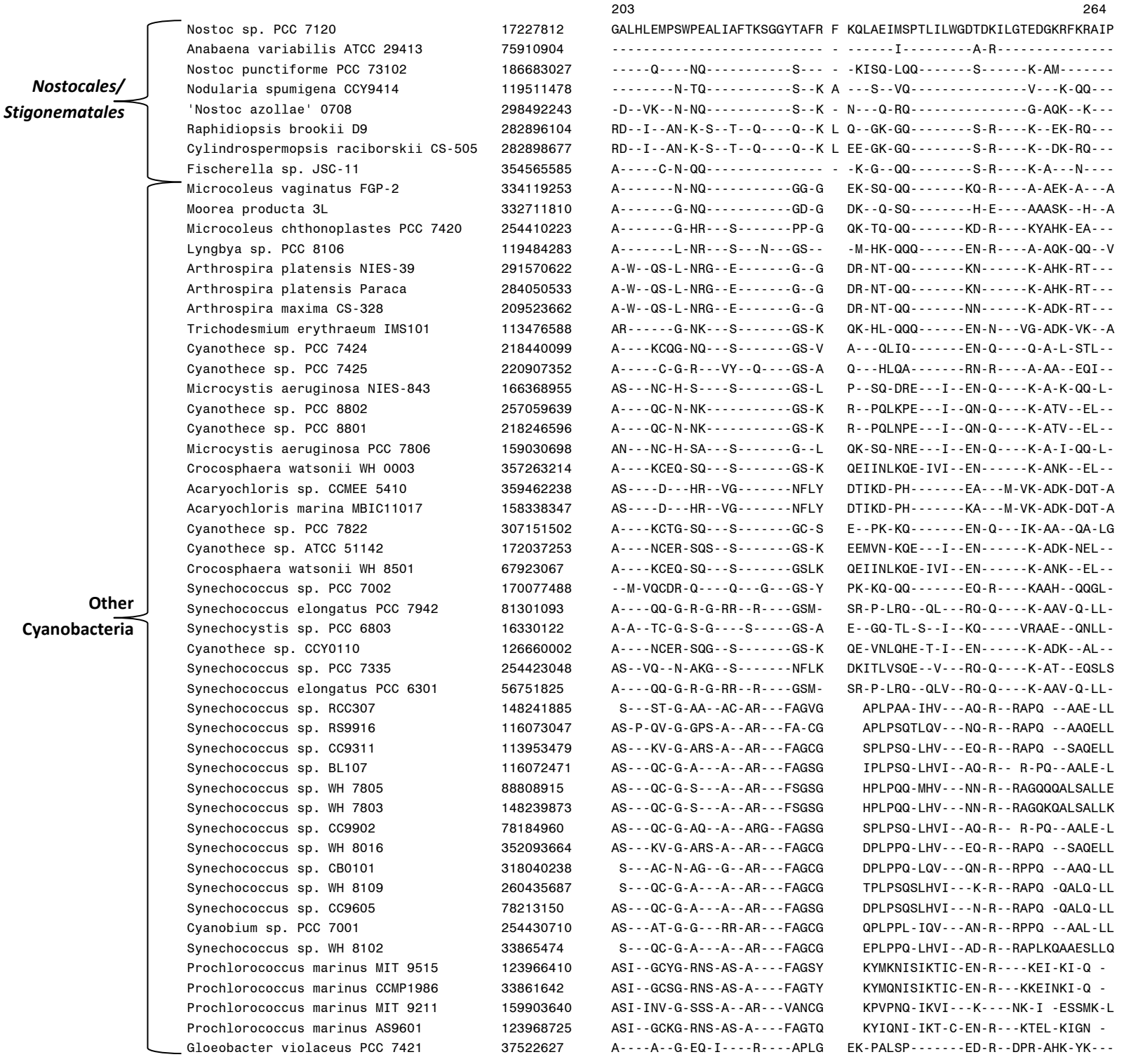
Supplementary Figure 7

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



Supplementary Figure 8

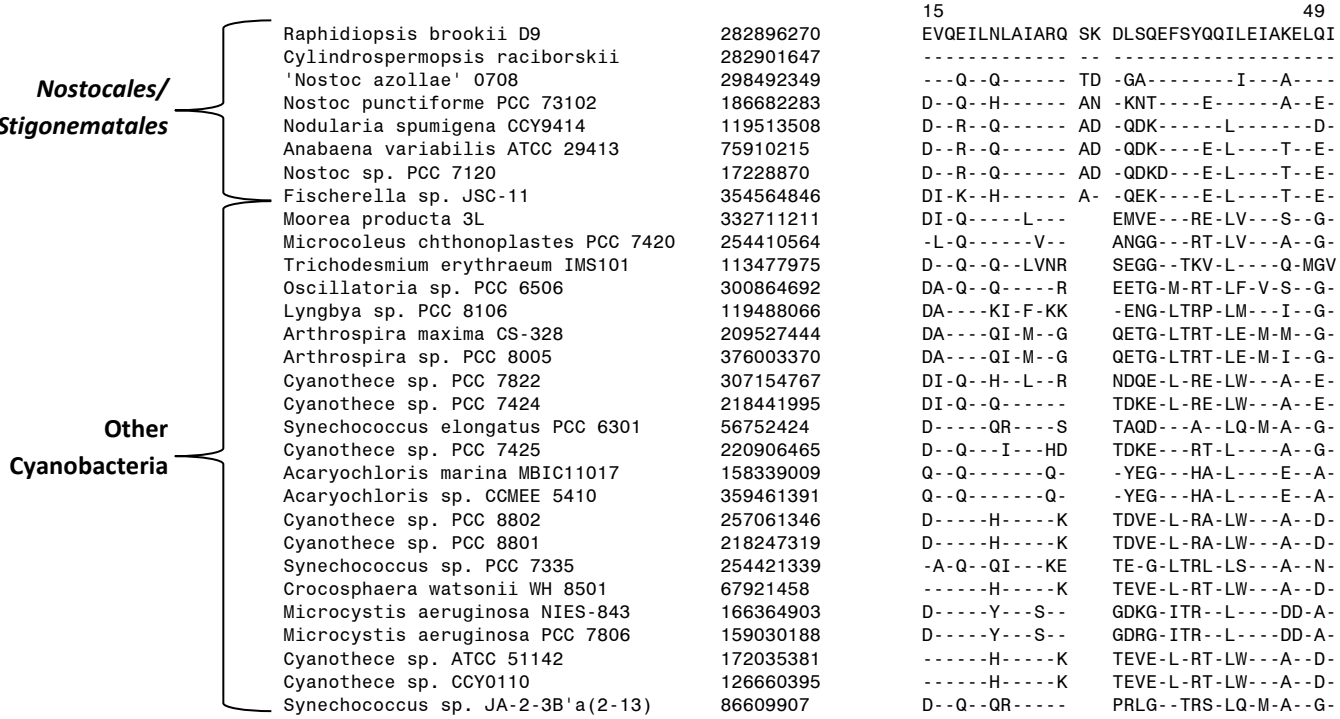
Partial sequence alignment for hypothetical protein Npun_R4490, showing a 1 aa insert that is specific for all *Nostocales* and *Stigonematales*.



		203	264
Nostocales/ Stigonematales	Nostoc sp. PCC 7120	17227812	GALHLEMPSWPEALIAFTKSGGYTAFR F KQLAEIMSPTLILWGDTDKILGTEDGKRFKRAIP
	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----
Other Cyanobacteria	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----KCQG-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-
	Crocosphaera watsonii WH 0003	357263214	A----KCEQ-SQ---S-------GS-K QEIINLKQE-IVI--EN------K-ANK--EL--
	Acaryochloris sp. CCMEE 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--
	Crocosphaera 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-
	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--RAGQQQALSALLE
	Synechococcus sp. WH 7803	148239873	AS---QC-G-S---A--AR---FSGSG HPLPQQ-LHV---NN-R--RAGQKQALSALLK
	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 DPLPPQ-LHV---EQ-R--RAPQ --SAQELL
	Synechococcus sp. CB0101	318040238	S---AC-N-AG--G--AR---FAGCG DPLPPQ-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 EPLPPQ-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--GCSG-RNS-AS-A----FAGTY KYMQNISIKTIC-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*.



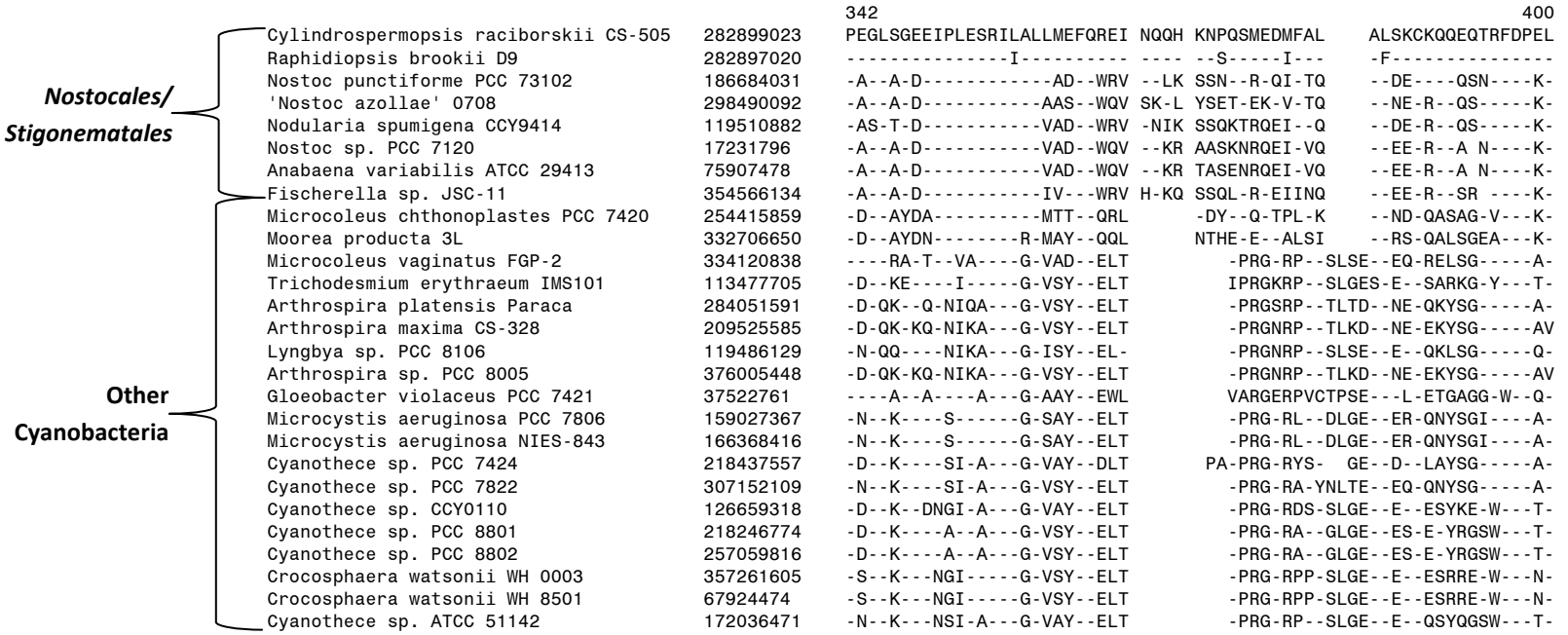
Supplementary Figure 10

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

Group	Organism	ID	126		184
Nostocales/ Stigonematales	Cylindrospermopsis raciborskii cs-505	282900401	GMNLADTHGRGLLRLINPNTVETLVV		EGTLKPANTKVLDTSCIIDGRIELLLETGFLEG
	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---------
Other Cyanobacteria	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------KSFD-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--
	Crocosphaera 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--
	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------RG--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------V------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---

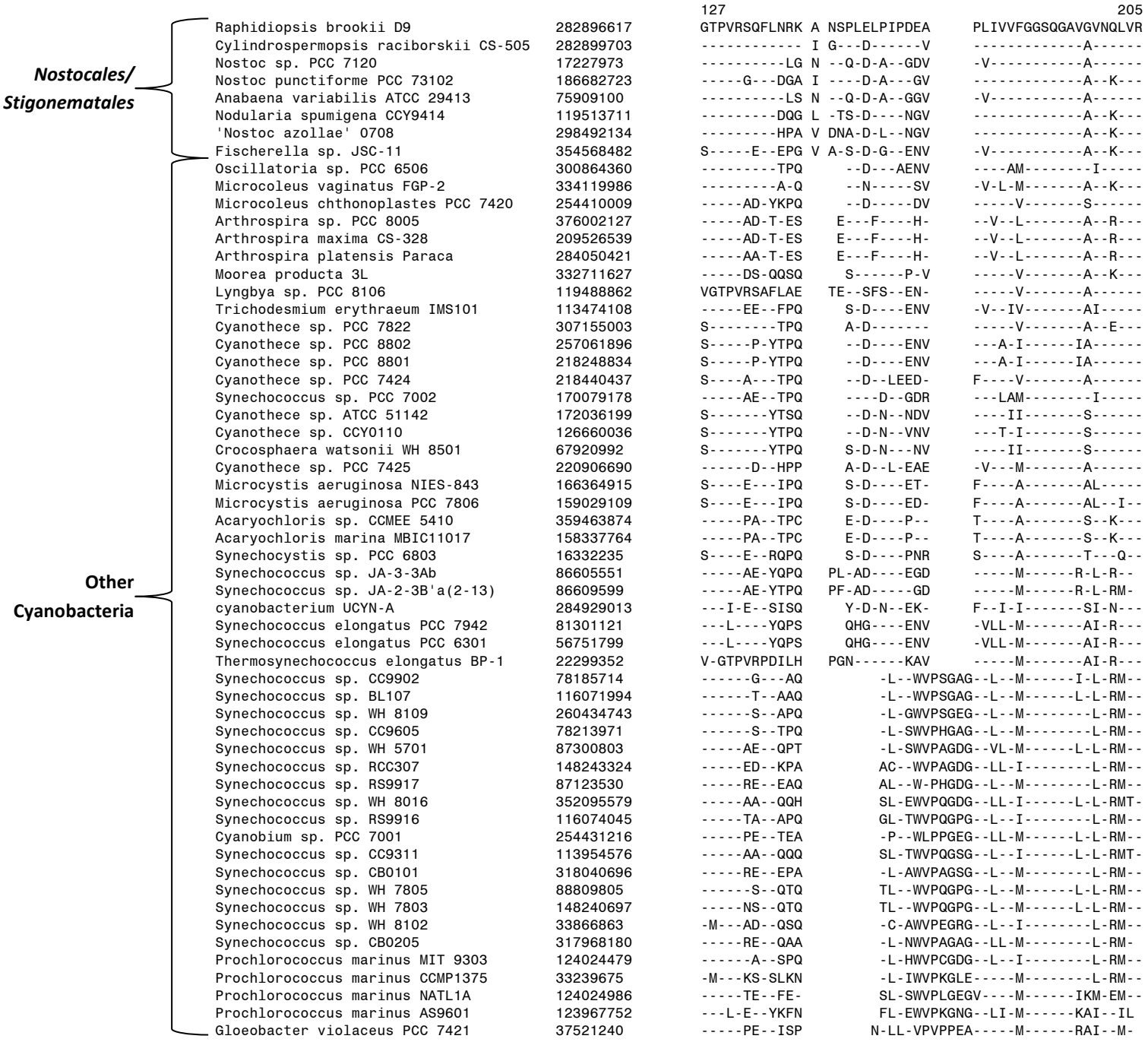
Supplementary Figure 11

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



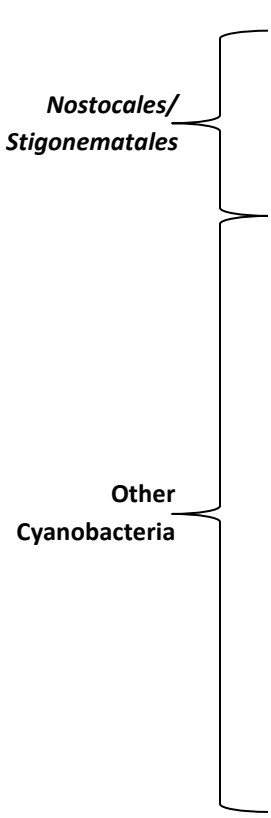
Supplementary Figure 12

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



Supplementary Figure 13

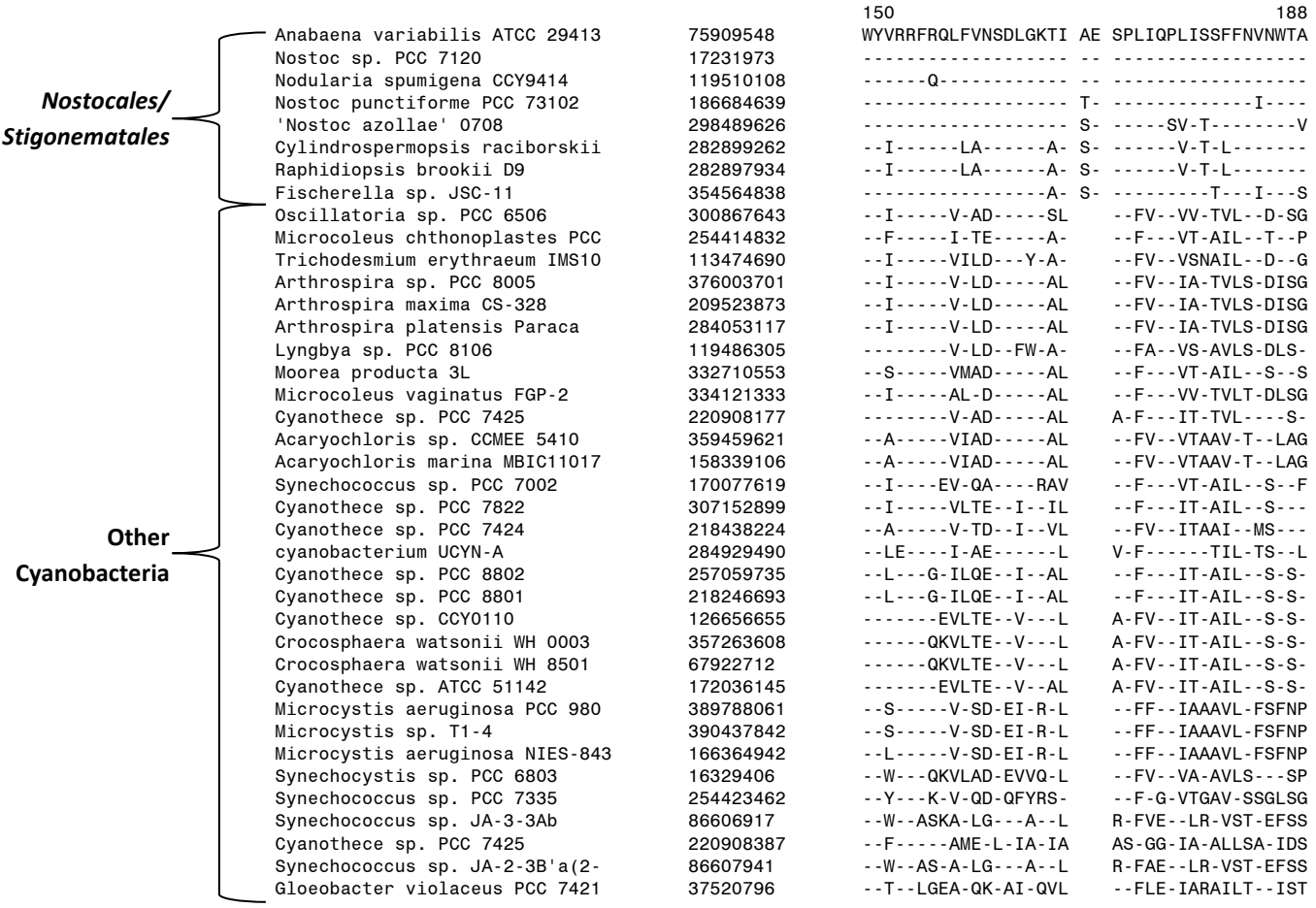
Partial sequence alignment for the protein N-acetylglucosaminyltransferase (MurG), showing a 1 aa insert that is specific for all *Nostocales* and *Stigonematales*.



		177	237
Cylindrospermopsis raciborskii CS-505	282901218	DRQRLDALEKYATL	VADIIRMGVNSESL VRNYPPPGPVLETMNMLGRSATPPAALQLLVDL
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	-YN---II--F--H	GEETSHRSQAM--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-----
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	PEQKHSSAQDIL-AVD--Q--E---E--IS-
Microcystis sp. T1-4	390439312	--I--ES---FILQ	PEQKYPAAMDILSL----Q--E--FE-----
Microcystis aeruginosa NIES-843	166367429	--I--ES---FILQ	PEQKYPAAMDILSL----Q--E--FE-----
Cyanothece sp. PCC 7424	218438216	--P--E----LVLQ	PEQKNQAAQDLLSS----T--E---E--IE-
Cyanothece sp. CCY0110	126660660	--T---S---FVLH	PEQNHRNAQDLLKE--QAT-AE--FE---A-
Cyanothece sp. PCC 8801	218246504	--N--ES--RLILQ	PEQTHRNAQ-LLSEI---Q--E--FE---E-
Cyanothece sp. PCC 8802	257059537	--N--ES--RLILQ	PEQTHRNAQ-LLSEI---Q--E--FE---E-
Cyanothece sp. ATCC 51142	172038259	--T--EL---FVLQ	PEQNHRNAQDLLKE--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	PEQSARAAHDLLE-----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	PEQNHRQA-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	SEDGNNTRAAT-VLTA---Q-SAQ--HD---A-
Synechococcus sp. PCC 7335	254423969	--P--EIV-R--L-	GDESSQKSAAQ-LLTE-NQDKS-E--FDF--A-
Crocosphaera watsonii WH 8501	67924092	--S--ES---FVLQ	PEQSHRNAQDLLKE--HKTI-DS-FE---A-
Gloeobacter violaceus PCC 7421	37521185	--L------R--LW	GDEASDRASAQDVLSQ--HTTRSED-IA---A-

Supplementary Figure 14

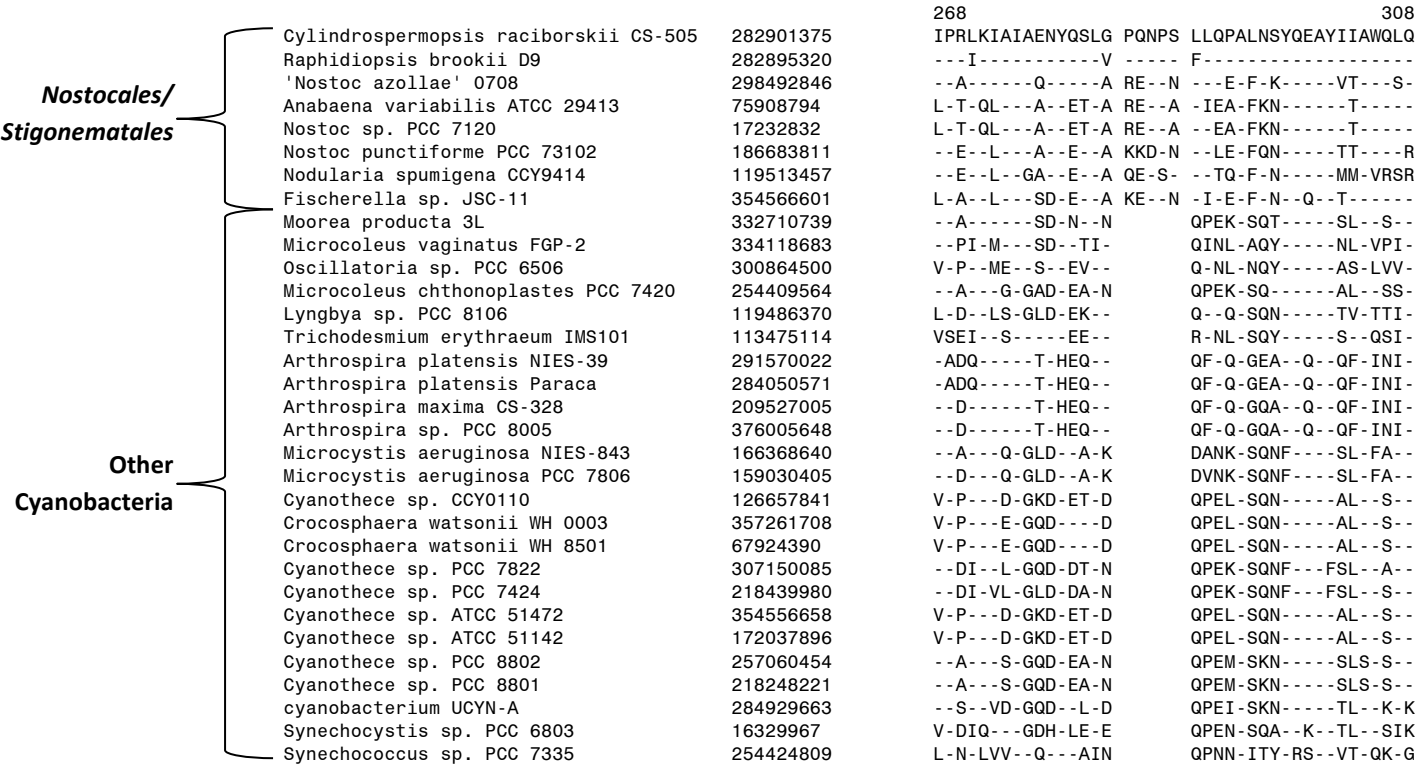
Partial sequence alignment for the protein Ribonuclease II, showing a 14 aa insert that is specific for all *Nostocales* and *Stigonematales*.



		150 188
Anabaena variabilis ATCC 29413	75909548	WYVRRFRQLFVNSDLGKTI AE SPLIQPLISSFFNVNWTA
Nostoc sp. PCC 7120	17231973	---------- -- ----------
Nodularia spumigena CCY9414	119510108	------Q----------- -- ----------
Nostoc punctiforme PCC 73102	186684639	---------- T- ----------I----
'Nostoc azollae' 0708	298489626	---------- S- -----SV-T--------V
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. CCMEE 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-
Crocosphaera watsonii WH 0003	357263608	------QKVLTE--V---L A-FV--IT-AIL--S-S-
Crocosphaera 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

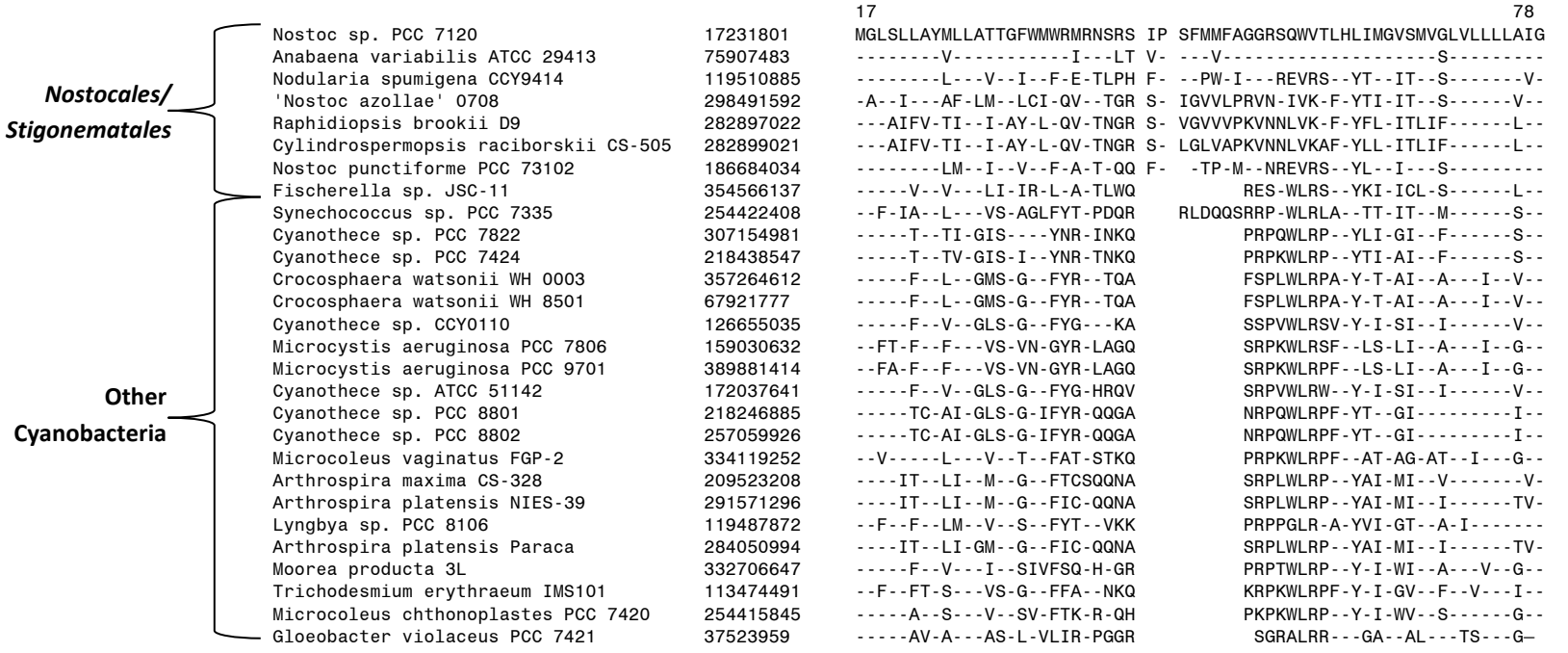
Partial sequence alignment for the protein arsenite efflux ATP-binding protein (ArsA), showing a 2 aa insert that is specific for all *Nostocales* and *Stigonematales*.



		268	308
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-
Arthrospira platensis NIES-39	291570022	-ADQ-----T-HEQ--	QF-Q-GEA--Q--QF-INI-
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--
Crocosphaera watsonii WH 0003	357261708	V-P---E-GQD----D	QPEL-SQN-----AL--S--
Crocosphaera 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

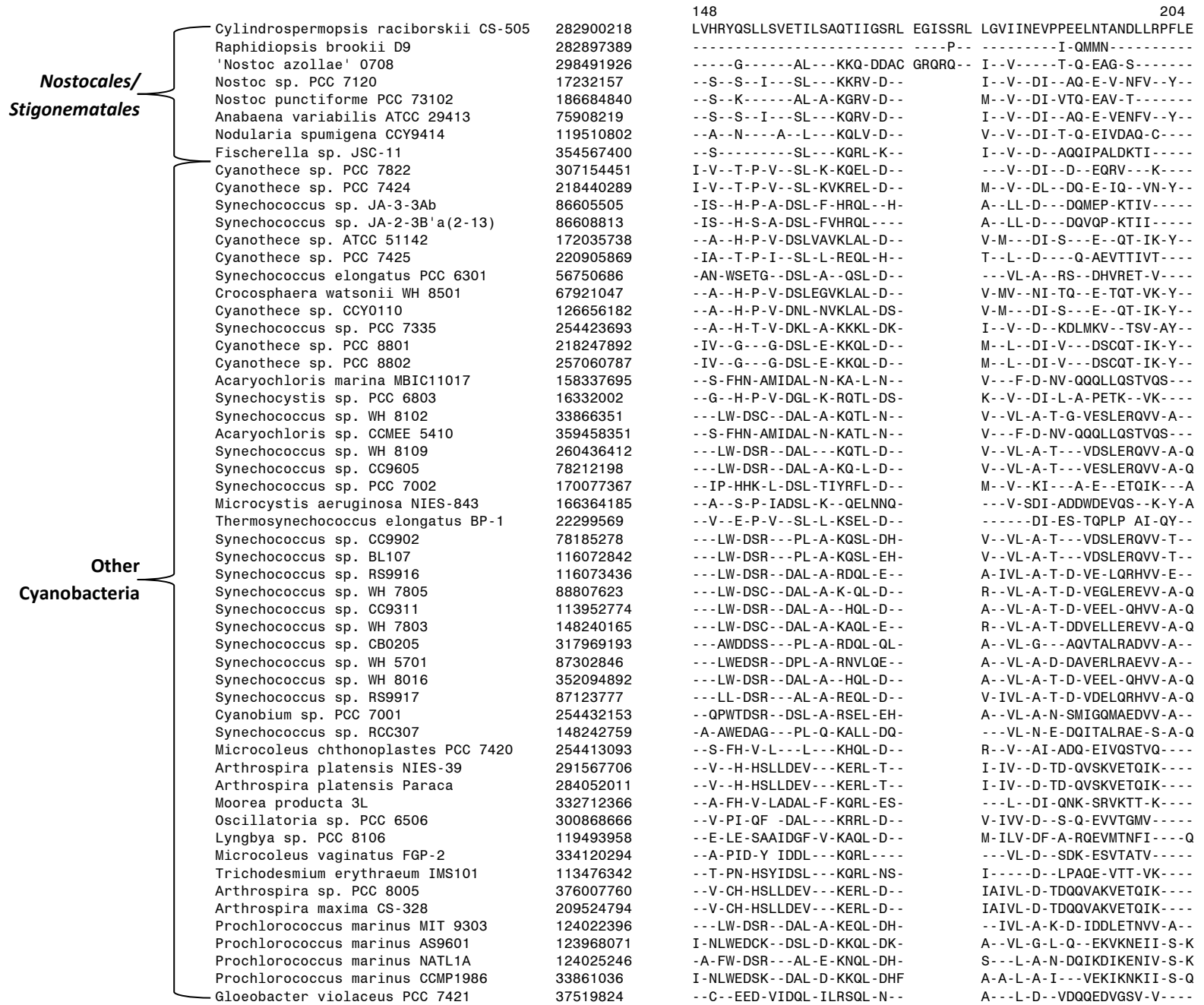
Supplementary Figure 16

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



Supplementary Figure 17

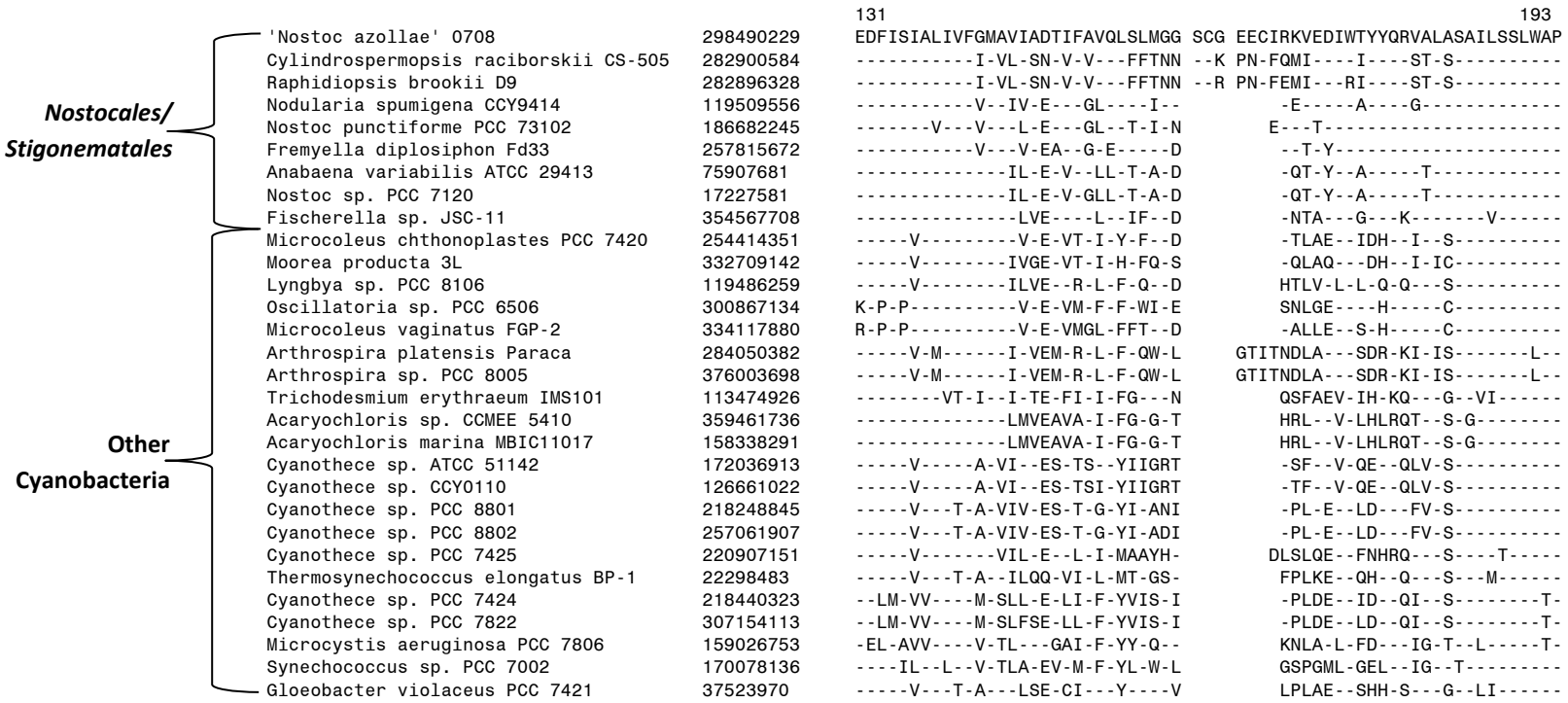
Partial sequence alignment for the hypothetical protein Npun_F3904, showing a 2 aa insert that is specific for all *Nostocales*.



		148	204
Cylindrospermopsis raciborskii CS-505	282900218	LVHRYQSLLSVETILSAQTIIGSRL EGISSRL	LGVIINEVPPEELNTANDLLRPFLE
Raphidiopsis brookii D9	282897389	------------------------- ----P--	---------I-QMMN----------
'Nostoc azollae' 0708	298491926	-----G------AL---KKQ-DDAC GRQRQ--	I--V-----T-Q-EAG-S-------
Nostoc sp. PCC 7120	17232157	--S--S--I---SL---KKRV-D--	I--V--DI--AQ-E-V-NFV--Y--
Nostoc punctiforme PCC 73102	186684840	--S--K------AL-A-KGRV-D--	M--V--DI-VTQ-EAV-T-------
Anabaena variabilis ATCC 29413	75908219	--S--S--I---SL---KQRV-D--	I--V--DI--AQ-E-VENFV--Y--
Nodularia spumigena CCY9414	119510802	--A--N----A--L---KQLV-D--	V--V--DI-T-Q-EIVDAQ-C----
Fischerella sp. JSC-11	354567400	--S---------SL---KQRL-K--	I--V--D--AQQIPALDKTI-----
Cyanothece sp. PCC 7822	307154451	I-V--T-P-V--SL-K-KQEL-D--	---V--DI--D--EQRV---K----
Cyanothece sp. PCC 7424	218440289	I-V--T-P-V--SL-KVKREL-D--	M--V--DL--DQ-E-IQ--VN-Y--
Synechococcus sp. JA-3-3Ab	86605505	-IS--H-P-A-DSL-F-HRQL--H-	A--LL-D---DQMEP-KTIV-----
Synechococcus sp. JA-2-3B'a(2-13)	86608813	-IS--H-S-A-DSL-FVHRQL----	A--LL-D---DQVQP-KTII-----
Cyanothece sp. ATCC 51142	172035738	--A--H-P-V-DSLVAVKLAL-D--	V-M---DI-S---E--QT-IK-Y--
Cyanothece sp. PCC 7425	220905869	-IA--T-P-I--SL-L-REQL-H--	T--L--D----Q-AEVTTIVT----
Synechococcus elongatus PCC 6301	56750686	-AN-WSETG--DSL-A--QSL-D--	---VL-A--RS--DHVRET-V----
Crocosphaera watsonii WH 8501	67921047	--A--H-P-V-DSLEGVKLAL-D--	V-MV--NI-TQ--E-TQT-VK-Y--
Cyanothece sp. CCY0110	126656182	--A--H-P-V-DNL-NVKLAL-DS-	V-M---DI-S---E--QT-IK-Y--
Synechococcus sp. PCC 7335	254423693	--A--H-T-V-DKL-A-KKKL-DK-	I--V--D--KDLMKV--TSV-AY--
Cyanothece sp. PCC 8801	218247892	-IV--G---G-DSL-E-KKQL-D--	M--L--DI-V---DSCQT-IK-Y--
Cyanothece sp. PCC 8802	257060787	-IV--G---G-DSL-E-KKQL-D--	M--L--DI-V---DSCQT-IK-Y--
Acaryochloris marina MBIC11017	158337695	--S-FHN-AMIDAL-N-KA-L-N--	V---F-D-NV-QQQLLQSTVQS---
Synechocystis sp. PCC 6803	16332002	--G--H-P-V-DGL-K-RQTL-DS-	K--V--DI-L-A-PETK--VK----
Synechococcus sp. WH 8102	33866351	---LW-DSC--DAL-A-KQTL-N--	V--VL-A-T-G-VESLERQVV-A--
Acaryochloris sp. CCMEE 5410	359458351	--S-FHN-AMIDAL-N-KATL-N--	V---F-D-NV-QQQLLQSTVQS---
Synechococcus sp. WH 8109	260436412	---LW-DSR--DAL---KQTL-D--	V--VL-A-T---VDSLERQVV-A-Q
Synechococcus sp. CC9605	78212198	---LW-DSR--DAL-A-KQ-L-D--	V--VL-A-T---VESLERQVV-A-Q
Synechococcus sp. PCC 7002	170077367	--IP-HHK-L-DSL-TIYRFL-D--	M--V--KI---A-E--ETQIK---A
Microcystis aeruginosa NIES-843	166364185	--A--S-P-IADSL-K--QELNNQ-	---V-SDI-ADDWDEVQS--K-Y-A
Thermosynechococcus elongatus BP-1	22299569	--V--E-P-V--SL-L-KSEL-D--	------DI-ES-TQPLP AI-QY--
Synechococcus sp. CC9902	78185278	---LW-DSR---PL-A-KQSL-DH-	V--VL-A-T---VDSLERQVV-T--
Synechococcus sp. BL107	116072842	---LW-DSR---PL-A-KQSL-EH-	V--VL-A-T---VDSLERQVV-T--
Synechococcus sp. RS9916	116073436	---LW-DSR--DAL-A-RDQL-E--	A-IVL-A-T-D-VE-LQRHVV-E--
Synechococcus sp. WH 7805	88807623	---LW-DSC--DAL-A-K-QL-D--	R--VL-A-T-D-VEGLEREVV-A-Q
Synechococcus sp. CC9311	113952774	---LW-DSR--DAL-A--HQL-D--	A--VL-A-T-D-VEEL-QHVV-A-Q
Synechococcus sp. WH 7803	148240165	---LW-DSC--DAL-A-KAQL-E--	R--VL-A-T-DDVELLEREVV-A-Q
Synechococcus sp. CB0205	317969193	---AWDDSS---PL-A-RDQL-QL-	A--VL-G---AQVTALRADVV-A--
Synechococcus sp. WH 5701	87302846	---LWEDSR--DPL-A-RNVLQE--	A--VL-A-D-DAVERLRAEVV-A--
Synechococcus sp. WH 8016	352094892	---LW-DSR--DAL-A--HQL-D--	A--VL-A-T-D-VEEL-QHVV-A-Q
Synechococcus sp. RS9917	87123777	---LL-DSR---AL-A-REQL-D--	V-IVL-A-T-D-VDELQRHVV-A-Q
Cyanobium sp. PCC 7001	254432153	--QPWTDSR--DSL-A-RSEL-EH-	A--VL-A-N-SMIGQMAEDVV-A--
Synechococcus sp. RCC307	148242759	-A-AWEDAG---PL-Q-KALL-DQ-	---VL-N-E-DQITALRAE-S-A-Q
Microcoleus chthonoplastes PCC 7420	254413093	--S-FH-V-L---L---KHQL-D--	R--V--AI-ADQ-EIVQSTVQ----
Arthrospira platensis NIES-39	291567706	--V--H-HSLLDEV---KERL-T--	I-IV--D-TD-QVSKVETQIK----
Arthrospira platensis Paraca	284052011	--V--H-HSLLDEV---KERL-T--	I-IV--D-TD-QVSKVETQIK----
Moorea producta 3L	332712366	--A-FH-V-LADAL-F-KQRL-ES-	---L--DI-QNK-SRVKTT-K----
Oscillatoria sp. PCC 6506	300868666	--V-PI-QF -DAL---KRRL-D--	V-IVV-D--S-Q-EVVTGMV-----
Lyngbya sp. PCC 8106	119493958	--E-LE-SAAIDGF-V-KAQL-D--	M-ILV-DF-A-RQEVMTNFI----Q
Microcoleus vaginatus FGP-2	334120294	--A-PID-Y IDDL---KQRL----	---VL-D--SDK-ESVTATV-----
Trichodesmium erythraeum IMS101	113476342	--T-PN-HSYIDSL---KQRL-NS-	I-----D--LPAQE-VTT-VK----
Arthrospira sp. PCC 8005	376007760	--V-CH-HSLLDEV---KERL-D--	IAIVL-D-TDQQVAKVETQIK----
Arthrospira maxima CS-328	209524794	--V-CH-HSLLDEV---KERL-D--	IAIVL-D-TDQQVAKVETQIK----
Prochlorococcus marinus MIT 9303	124022396	---LW-DSR--DAL-A-KEQL-DH-	--IVL-A-K-D-IDDLETNVV-A--
Prochlorococcus marinus AS9601	123968071	I-NLWEDCK--DSL-D-KKQL-DK-	A--VL-G-L-Q--EKVKNEII-S-K
Prochlorococcus marinus NATL1A	124025246	-A-FW-DSR---AL-E-KNQL-DH-	S---L-A-N-DQIKDIKENIV-S-K
Prochlorococcus marinus CCMP1986	33861036	I-NLWEDSK--DAL-D-KKQL-DHF	A-A-L-A-I---VEKIKNKII-S-Q
Gloeobacter violaceus PCC 7421	37519824	--C--EED-VIDQL-ILRSQL-N--	A---L-D--VDQQEDVGSV-V----

Supplementary Figure 18

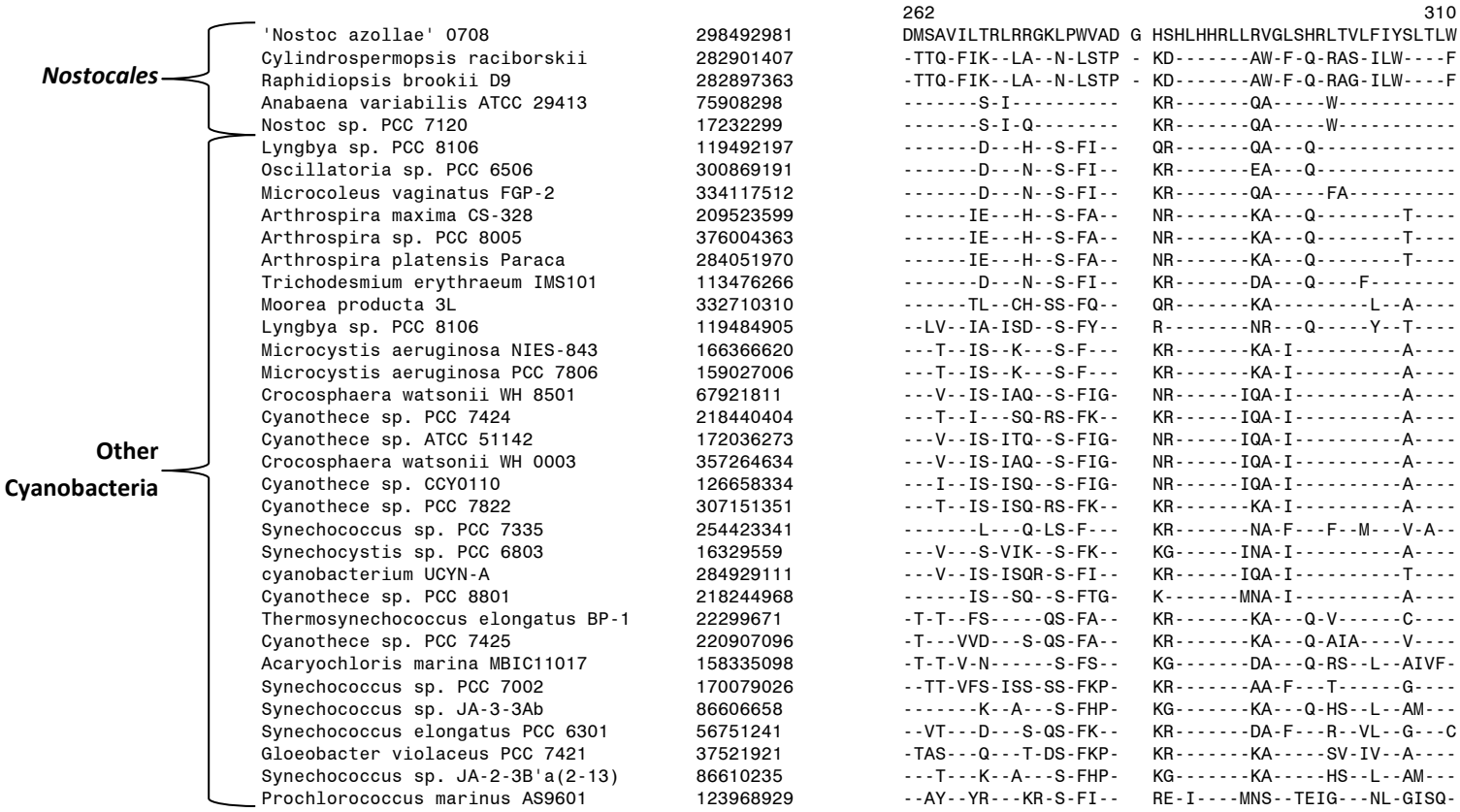
Partial sequence alignment for the hypothetical protein CRC_01609, showing a 7 aa insert that is specific for all members of a *Nostocales* subclade which includes *Cylindrospermopsis raciborskii CS-505*, *Raphidiopsis brookii D9* and *'Nostoc azollae' 0708*.



		131 193
'Nostoc azollae' 0708	298490229	EDFISIALIVFGMAVIADTIFAVQLSLMGG SCG EECIRKVEDIWTYYQRVALASAILSSLWAP
Cylindrospermopsis raciborskii CS-505	282900584	----------I-VL-SN-V-V---FFTNN --K PN-FQMI----I----ST-S----------
Raphidiopsis brookii D9	282896328	----------I-VL-SN-V-V---FFTNN --R PN-FEMI---RI----ST-S----------
Nodularia spumigena CCY9414	119509556	----------V--IV-E---GL----I-- -E-----A----G----------
Nostoc punctiforme PCC 73102	186682245	-------V---V---L-E---GL--T-I-N E---T----------
Fremyella diplosiphon Fd33	257815672	----------V---V-EA--G-E-----D --T-Y----------
Anabaena variabilis ATCC 29413	75907681	--------------IL-E-V--LL-T-A-D -QT-Y--A-----T----------
Nostoc sp. PCC 7120	17227581	--------------IL-E-V-GLL-T-A-D -QT-Y--A-----T----------
Fischerella sp. JSC-11	354567708	---------------LVE----L--IF--D -NTA---G---K-------V------
Microcoleus chthonoplastes PCC 7420	254414351	-----V---------V-E-VT-I-Y-F--D -TLAE--IDH--I--S----------
Moorea producta 3L	332709142	-----V--------IVGE-VT-I-H-FQ-S -QLAQ---DH--I-IC----------
Lyngbya sp. PCC 8106	119486259	-----V--------ILVE--R-L-F-Q--D HTLV-L-L-Q-Q---S----------
Oscillatoria sp. PCC 6506	300867134	K-P-P----------V-E-VM-F-F-WI-E SNLGE----H-----C----------
Microcoleus vaginatus FGP-2	334117880	R-P-P----------V-E-VMGL-FFT--D -ALLE--S-H-----C----------
Arthrospira platensis Paraca	284050382	-----V-M------I-VEM-R-L-F-QW-L GTITNDLA---SDR-KI-IS-------L--
Arthrospira sp. PCC 8005	376003698	-----V-M------I-VEM-R-L-F-QW-L GTITNDLA---SDR-KI-IS-------L--
Trichodesmium erythraeum IMS101	113474926	--------VT-I--I-TE-FI-I-FG---N QSFAEV-IH-KQ---G--VI------
Acaryochloris sp. CCMEE 5410	359461736	--------------LMVEAVA-I-FG-G-T HRL--V-LHLRQT--S-G--------
Acaryochloris marina MBIC11017	158338291	--------------LMVEAVA-I-FG-G-T HRL--V-LHLRQT--S-G--------
Cyanothece sp. ATCC 51142	172036913	-----V-----A-VI--ES-TS--YIIGRT -SF--V-QE--QLV-S----------
Cyanothece sp. CCY0110	126661022	-----V-----A-VI--ES-TSI-YIIGRT -TF--V-QE--QLV-S----------
Cyanothece sp. PCC 8801	218248845	-----V---T-A-VIV-ES-T-G-YI-ANI -PL-E--LD---FV-S----------
Cyanothece sp. PCC 8802	257061907	-----V---T-A-VIV-ES-T-G-YI-ADI -PL-E--LD---FV-S----------
Cyanothece sp. PCC 7425	220907151	-----V-------VIL-E--L-I-MAAYH- DLSLQE--FNHRQ---S----T-----
Thermosynechococcus elongatus BP-1	22298483	-----V---T-A--ILQQ-VI-L-MT-GS- FPLKE--QH--Q---S---M------
Cyanothece sp. PCC 7424	218440323	--LM-VV----M-SLL-E-LI-F-YVIS-I -PLDE--ID--QI--S--------T-
Cyanothece sp. PCC 7822	307154113	--LM-VV----M-SLFSE-LL-F-YVIS-I -PLDE--LD--QI--S--------T-
Microcystis aeruginosa PCC 7806	159026753	-EL-AVV----V-TL---GAI-F-YY-Q-- KNLA-L-FD---IG-T--L-----T-
Synechococcus sp. PCC 7002	170078136	----IL--L--V-TLA-EV-M-F-YL-W-L GSPGML-GEL--IG--T---------
Gloeobacter violaceus PCC 7421	37523970	-----V---T-A---LSE-CI---Y----V LPLAE--SHH-S---G--LI------

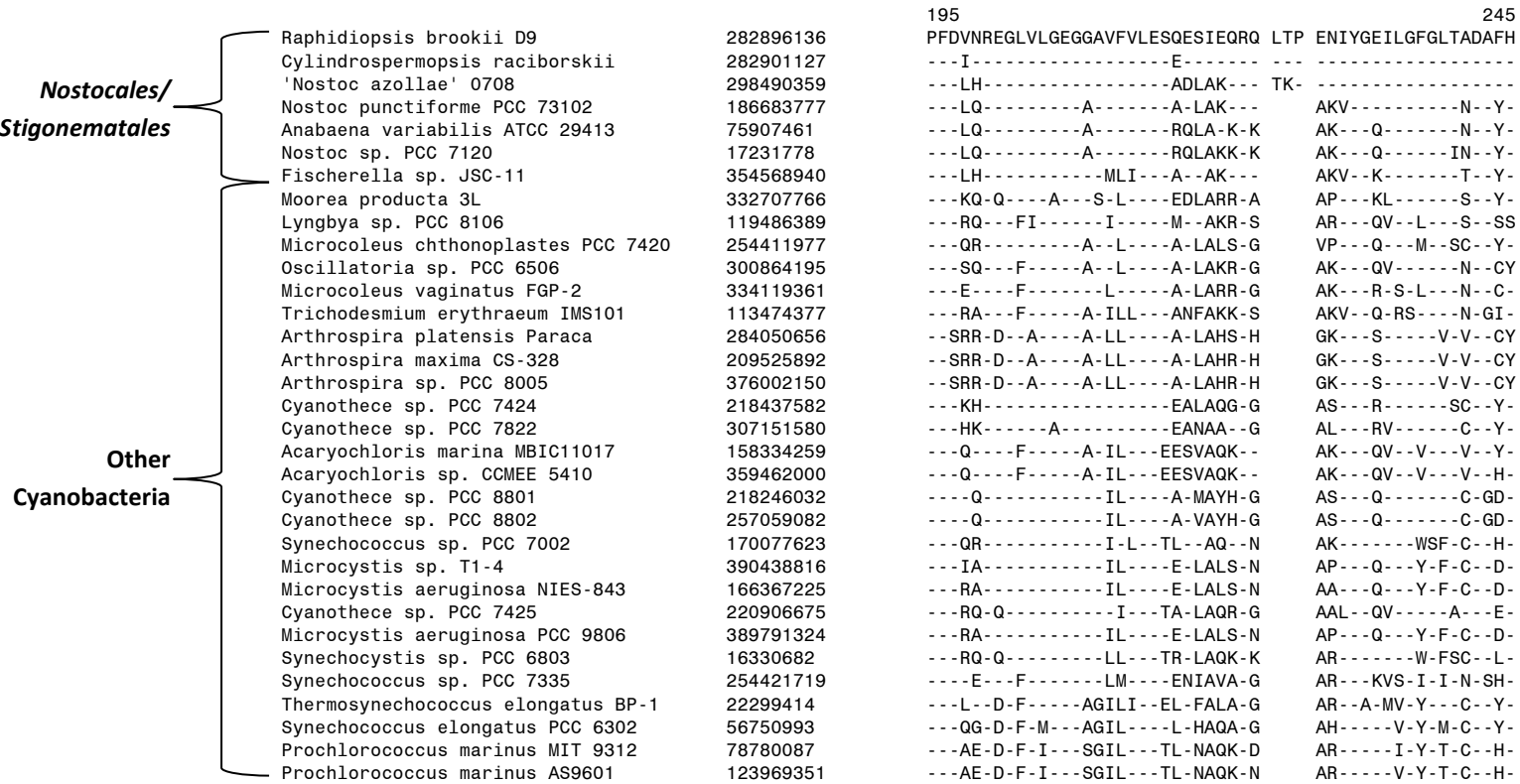
Supplementary Figure 19

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*.



Supplementary Figure 20

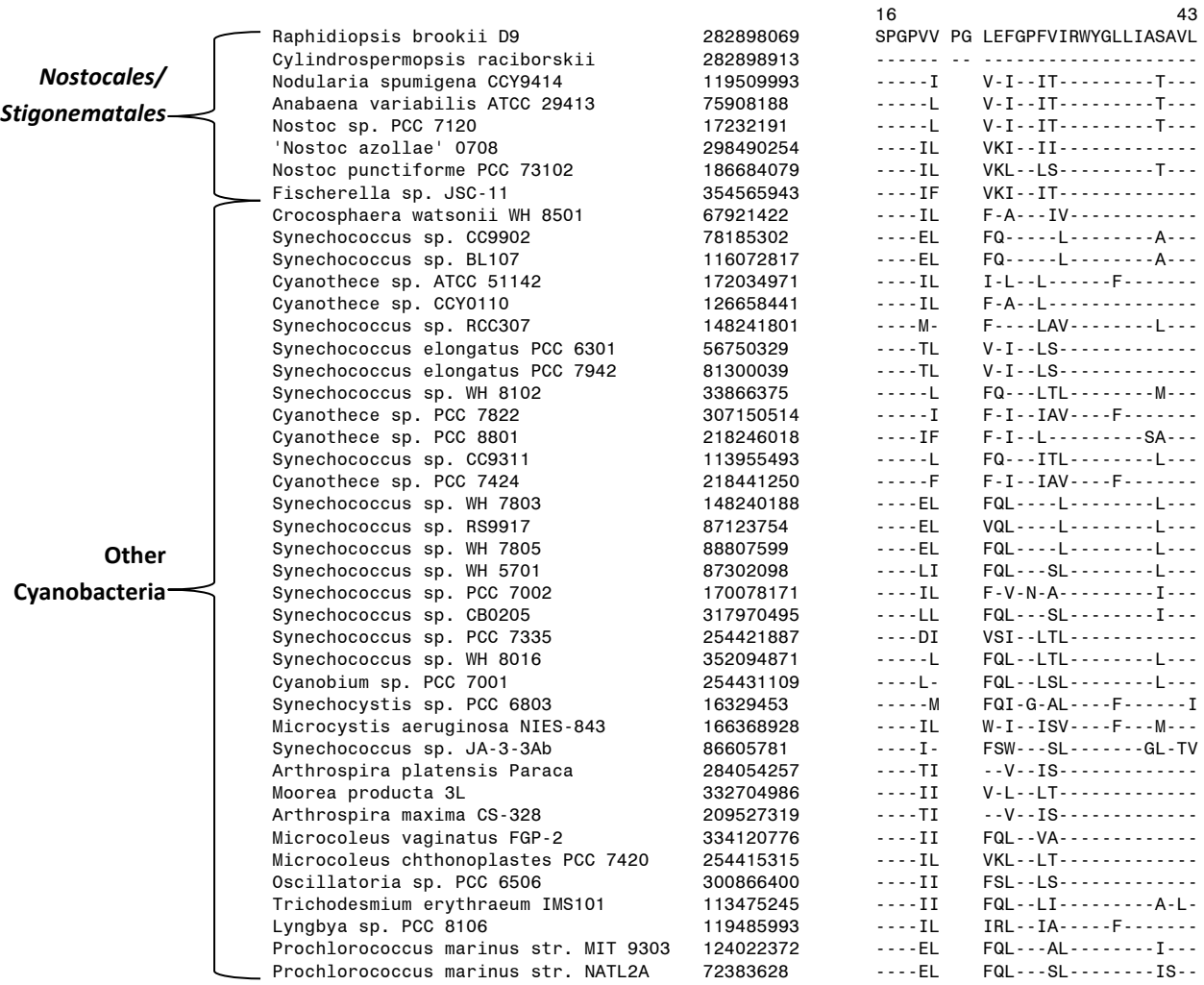
Partial sequence alignment for the family 4 glycosyl transferase, showing a 1 aa insert that is specific for all members of a *Nostocales* subclade which includes *Cylindrospermopsis raciborskii CS-505*, *Raphidiopsis brookii D9* and *'Nostoc azollae' 0708*. A similar insertion was also identified in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



		195 245
Raphidiopsis brookii D9	282896136	PFDVNREGLVLGEGGAVFVLESQESIEQRQ LTP ENIYGEILGFGLTADAFH
Cylindrospermopsis raciborskii	282901127	---I----------E------- --- ----------
'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
Cyanothece sp. PCC 7424	218437582	---KH----------EALAQG-G AS---R------SC--Y-
Cyanothece sp. PCC 7822	307151580	---HK------A----------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

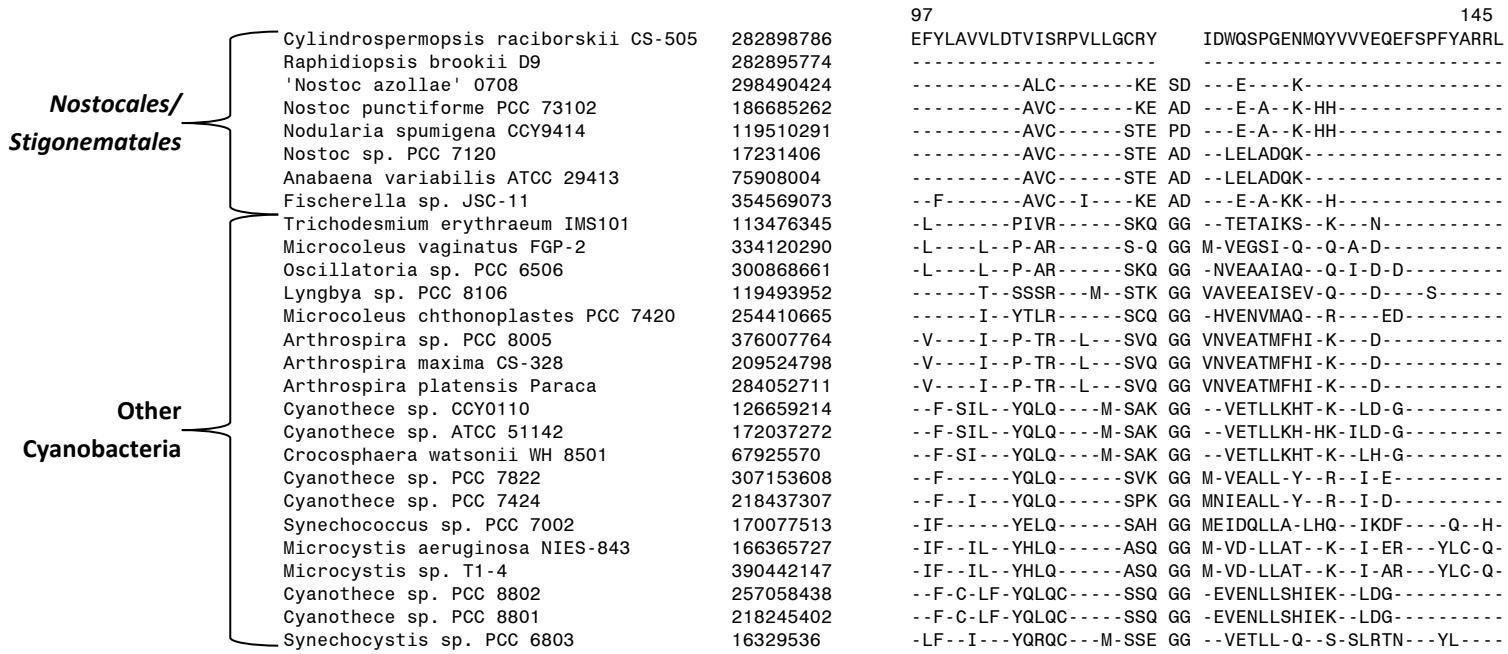
Partial sequence alignment for the protein beta-ketoacyl synthase, showing a 3 aa insert that is specific for all members of a *Nostocales* subclade which includes *Cylindrospermopsis raciborskii CS-505*, *Raphidiopsis brookii D9* and *'Nostoc azollae' 0708*. A similar insertion was also identified in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



		16	43
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-------------
Nostoc punctiforme PCC 73102	186684079	----IL	VKL--LS---------T---
Fischerella sp. JSC-11	354565943	----IF	VKI--IT-------------
Crocosphaera watsonii WH 8501	67921422	----IL	F-A---IV------------
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--------------
Synechococcus sp. RCC307	148241801	----M-	F----LAV--------L---
Synechococcus elongatus PCC 6301	56750329	----TL	V-I--LS-------------
Synechococcus elongatus PCC 7942	81300039	----TL	V-I--LS-------------
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------------
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-------------
Moorea producta 3L	332704986	----II	V-L--LT-------------
Arthrospira maxima CS-328	209527319	----TI	--V--IS-------------
Microcoleus vaginatus FGP-2	334120776	----II	FQL--VA-------------
Microcoleus chthonoplastes PCC 7420	254415315	----IL	VKL--LT-------------
Oscillatoria sp. PCC 6506	300866400	----II	FSL--LS-------------
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

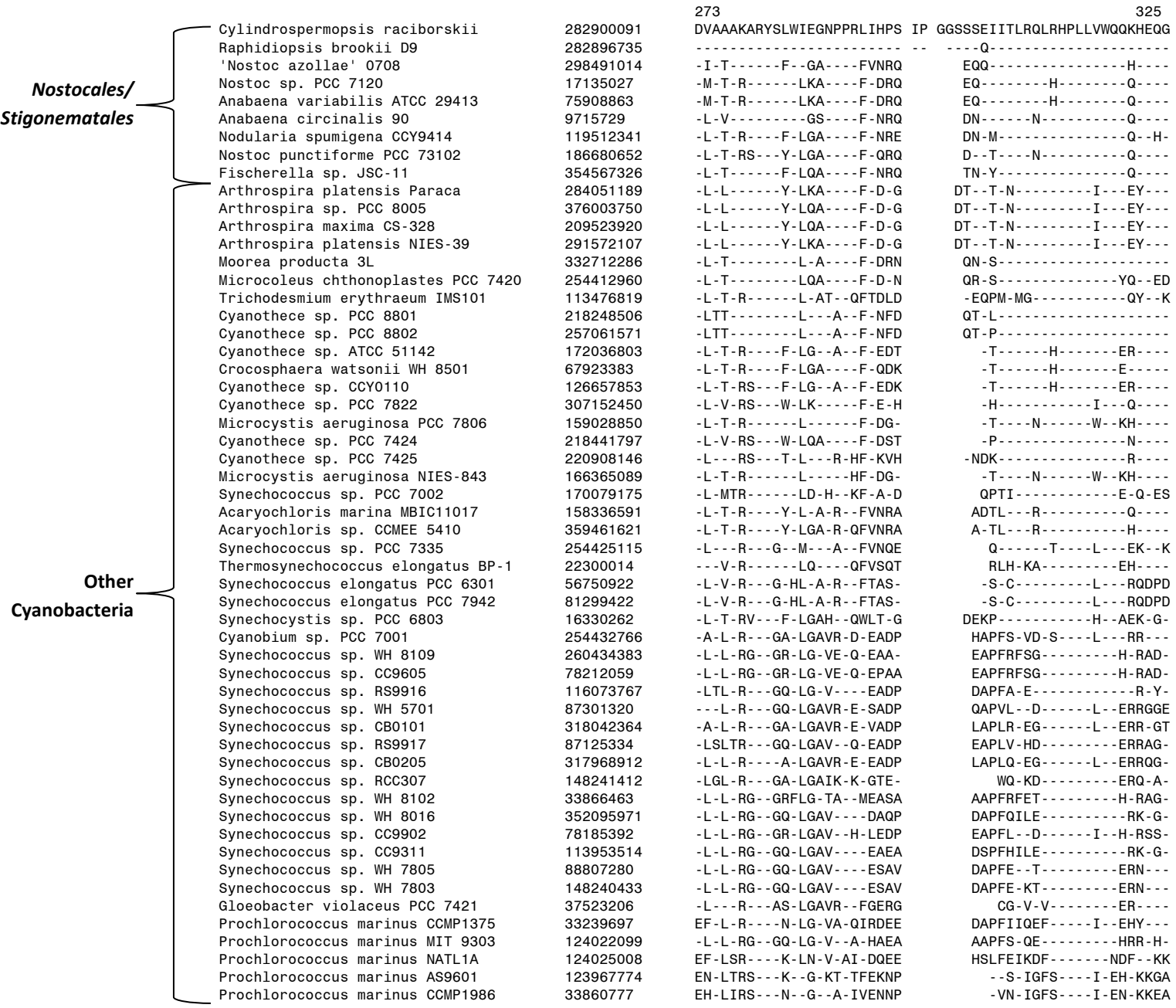
Partial sequence alignment for the prolipoprotein diacylglyceryl transferase, showing a 2 aa insert that is specific for *Cylindrospermopsis raciborskii CS-505 and Raphidiopsis brookii D9.*



		97	145
Cylindrospermopsis raciborskii CS-505	282898786	EFYLAVVLDTVISRPVLLGCRY	IDWQSPGENMQYVVVEQEFSPFYARRL
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----------
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---------
Crocosphaera 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

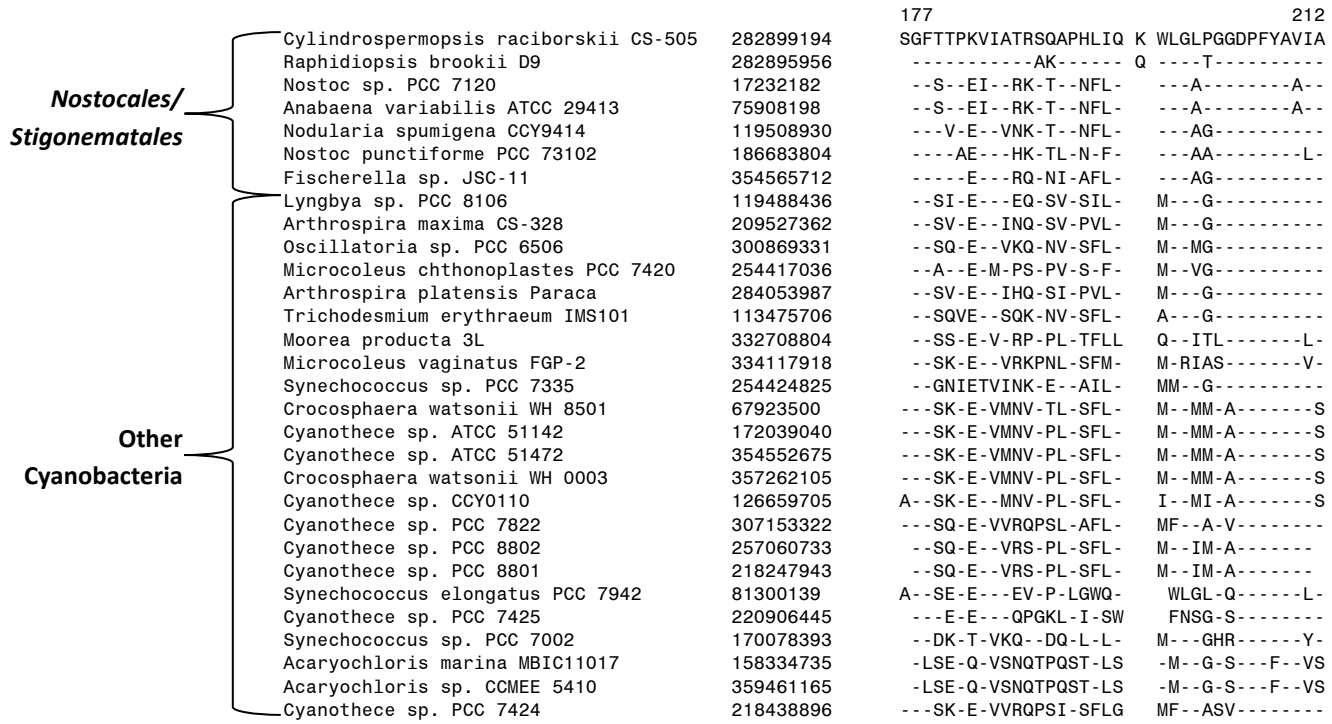
Partial sequence alignment for the protein succinyl-CoA synthetase, showing a 2 aa deletion that is specific for *Cylindrospermopsis raciborskii CS-505 and Raphidiopsis brookii D9.*



		273	325
Cylindrospermopsis raciborskii	282900091	DVAAAKARYSLWIEGNPPRLIHPS IP	GGSSSEIITLRQLRHPLLVWQQKHEQG
Raphidiopsis brookii D9	282896735	------------------------ --	----Q----------
'Nostoc azollae' 0708	298491014	-I-T------F--GA----FVNRQ	EQQ----------H----
Nostoc sp. PCC 7120	17135027	-M-T-R------LKA----F-DRQ	EQ--------H--------Q----
Anabaena variabilis ATCC 29413	75908863	-M-T-R------LKA----F-DRQ	EQ--------H--------Q----
Anabaena circinalis 90	9715729	-L-V---------GS----F-NRQ	DN------N----------Q----
Nodularia spumigena CCY9414	119512341	-L-T-R----F-LGA----F-NRE	DN-M----------Q--H-
Nostoc punctiforme PCC 73102	186680652	-L-T-RS---Y-LGA----F-QRQ	D--T----N----------Q----
Fischerella sp. JSC-11	354567326	-L-T------F-LQA----F-NRQ	TN-Y----------Q----
Arthrospira platensis Paraca	284051189	-L-L------Y-LKA----F-D-G	DT--T-N---------I---EY---
Arthrospira sp. PCC 8005	376003750	-L-L------Y-LQA----F-D-G	DT--T-N---------I---EY---
Arthrospira maxima CS-328	209523920	-L-L------Y-LQA----F-D-G	DT--T-N---------I---EY---
Arthrospira platensis NIES-39	291572107	-L-L------Y-LKA----F-D-G	DT--T-N---------I---EY---
Moorea producta 3L	332712286	-L-T--------L-A----F-DRN	QN-S----------
Microcoleus chthonoplastes PCC 7420	254412960	-L-T--------LQA----F-D-N	QR-S----------YQ--ED
Trichodesmium erythraeum IMS101	113476819	-L-T-R------L-AT--QFTDLD	-EQPM-MG----------QY--K
Cyanothece sp. PCC 8801	218248506	-LTT--------L---A--F-NFD	QT-L----------
Cyanothece sp. PCC 8802	257061571	-LTT--------L---A--F-NFD	QT-P----------
Cyanothece sp. ATCC 51142	172036803	-L-T-R----F-LG--A--F-EDT	-T------H-------ER----
Crocosphaera watsonii WH 8501	67923383	-L-T-R----F-LGA----F-QDK	-T------H-------E-----
Cyanothece sp. CCY0110	126657853	-L-T-RS---F-LG--A--F-EDK	-T------H-------ER----
Cyanothece sp. PCC 7822	307152450	-L-V-RS---W-LK-----F-E-H	-H----------I---Q----
Microcystis aeruginosa PCC 7806	159028850	-L-T-R------L------F-DG-	-T----N------W--KH----
Cyanothece sp. PCC 7424	218441797	-L-V-RS---W-LQA----F-DST	-P---------N----
Cyanothece sp. PCC 7425	220908146	-L---RS---T-L---R-HF-KVH	-NDK----------R----
Microcystis aeruginosa NIES-843	166365089	-L-T-R------L-----HF-DG-	-T----N------W--KH----
Synechococcus sp. PCC 7002	170079175	-L-MTR------LD-H--KF-A-D	QPTI----------E-Q-ES
Acaryochloris marina MBIC11017	158336591	-L-T-R----Y-L-A-R--FVNRA	ADTL---R----------Q----
Acaryochloris sp. CCMEE 5410	359461621	-L-T-R----Y-LGA-R-QFVNRA	A-TL---R----------H----
Synechococcus sp. PCC 7335	254425115	-L---R---G--M---A--FVNQE	Q------T----L---EK--K
Thermosynechococcus elongatus BP-1	22300014	---V-R------LQ----QFVSQT	RLH-KA---------EH----
Synechococcus elongatus PCC 6301	56750922	-L-V-R---G-HL-A-R--FTAS-	-S-C---------L---RQDPD
Synechococcus elongatus PCC 7942	81299422	-L-V-R---G-HL-A-R--FTAS-	-S-C---------L---RQDPD
Synechocystis sp. PCC 6803	16330262	-L-T-RV---F-LGAH--QWLT-G	DEKP----------H--AEK-G-
Cyanobium sp. PCC 7001	254432766	-A-L-R---GA-LGAVR-D-EADP	HAPFS-VD-S----L---RR---
Synechococcus sp. WH 8109	260434383	-L-L-RG--GR-LG-VE-Q-EAA-	EAPFRFSG---------H-RAD-
Synechococcus sp. CC9605	78212059	-L-L-RG--GR-LG-VE-Q-EPAA	EAPFRFSG---------H-RAD-
Synechococcus sp. RS9916	116073767	-LTL-R---GQ-LG-V----EADP	DAPFA-E----------R-Y-
Synechococcus sp. WH 5701	87301320	---L-R---GQ-LGAVR-E-SADP	QAPVL--D------L--ERRGGE
Synechococcus sp. CB0101	318042364	-A-L-R---GA-LGAVR-E-VADP	LAPLR-EG------L--ERR-GT
Synechococcus sp. RS9917	87125334	-LSLTR---GQ-LGAV--Q-EADP	EAPLV-HD---------ERRAG-
Synechococcus sp. CB0205	317968912	-L-L-R----A-LGAVR-E-EADP	LAPLQ-EG------L--ERRQG-
Synechococcus sp. RCC307	148241412	-LGL-R---GA-LGAIK-K-GTE-	WQ-KD---------ERQ-A-
Synechococcus sp. WH 8102	33866463	-L-L-RG--GRFLG-TA--MEASA	AAPFRFET---------H-RAG-
Synechococcus sp. WH 8016	352095971	-L-L-RG--GQ-LGAV----DAQP	DAPFQILE----------RK-G-
Synechococcus sp. CC9902	78185392	-L-L-RG--GR-LGAV--H-LEDP	EAPFL--D------I--H-RSS-
Synechococcus sp. CC9311	113953514	-L-L-RG--GQ-LGAV----EAEA	DSPFHILE----------RK-G-
Synechococcus sp. WH 7805	88807280	-L-L-RG--GQ-LGAV----ESAV	DAPFE--T---------ERN---
Synechococcus sp. WH 7803	148240433	-L-L-RG--GQ-LGAV----ESAV	DAPFE-KT---------ERN---
Gloeobacter violaceus PCC 7421	37523206	-L---R---AS-LGAVR--FGERG	CG-V-V--------ER----
Prochlorococcus marinus CCMP1375	33239697	EF-L-R----N-LG-VA-QIRDEE	DAPFIIQEF-----I--EHY---
Prochlorococcus marinus MIT 9303	124022099	-L-L-RG--GQ-LG-V--A-HAEA	AAPFS-QE---------HRR-H-
Prochlorococcus marinus NATL1A	124025008	EF-LSR----K-LN-V-AI-DQEE	HSLFEIKDF-------NDF--KK
Prochlorococcus marinus AS9601	123967774	EN-LTRS---K--G-KT-TFEKNP	--S-IGFS----I-EH-KKGA
Prochlorococcus marinus CCMP1986	33860777	EH-LIRS---N--G--A-IVENNP	-VN-IGFS----I-EN-KKEA

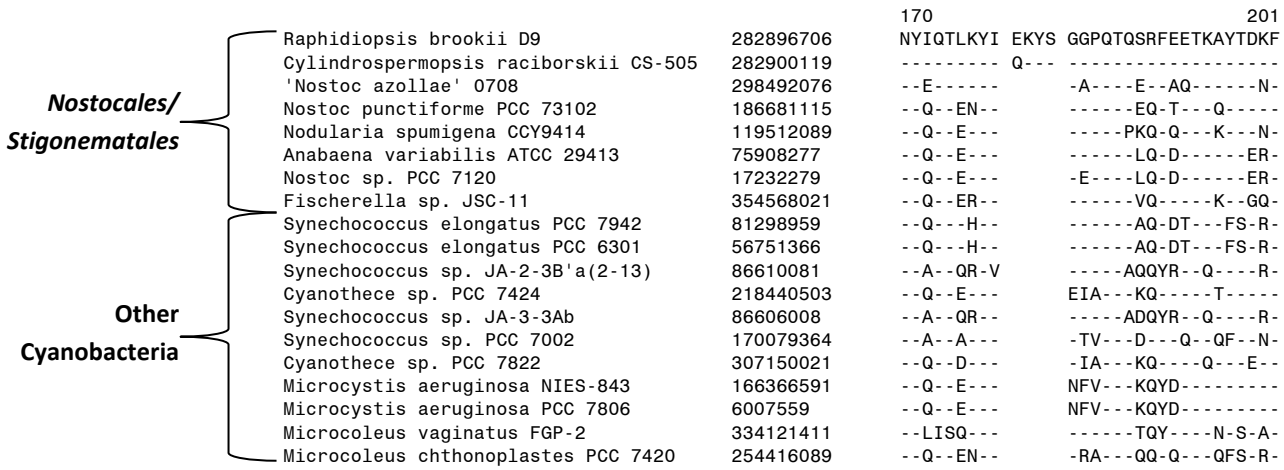
Supplementary Figure 24

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.*



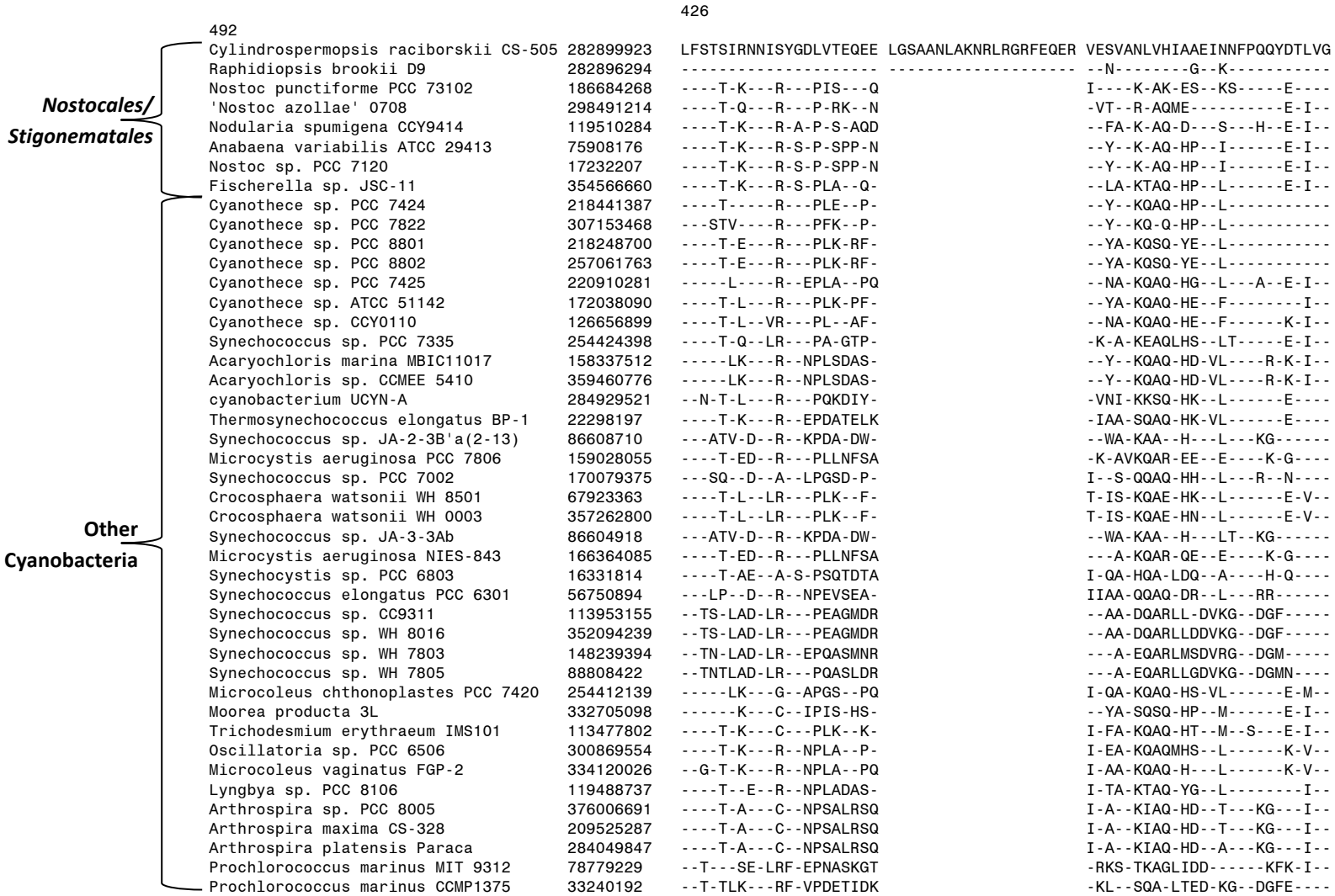
Supplementary Figure 25

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.*



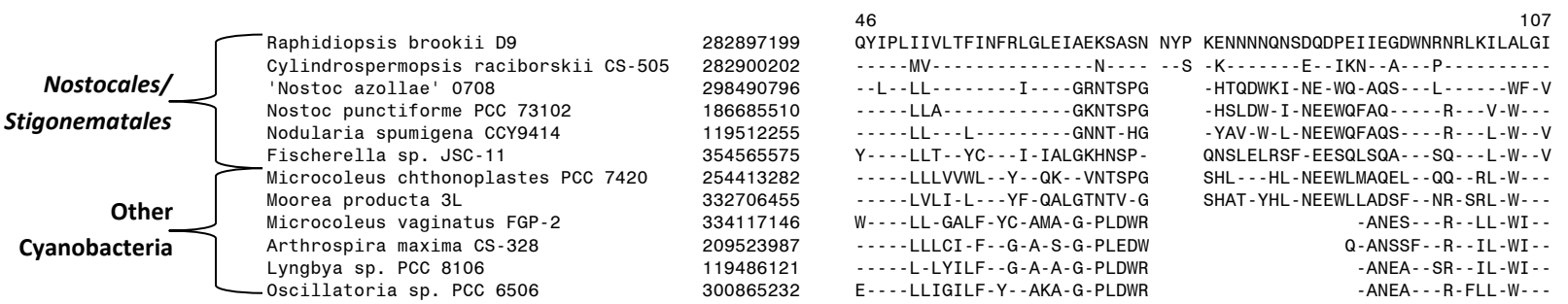
Supplementary Figure 26

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.*



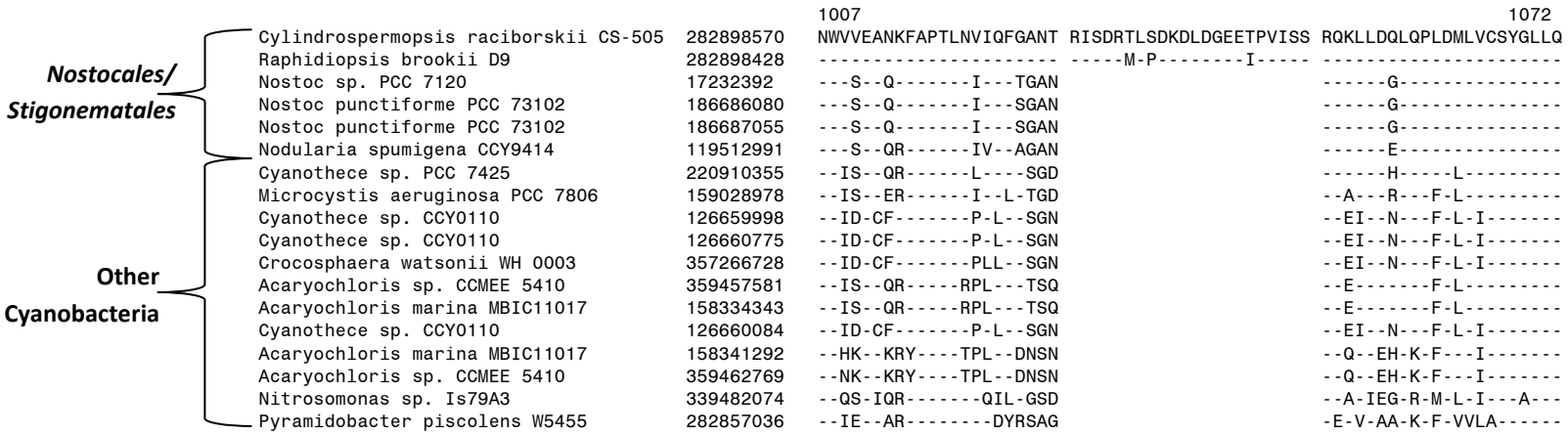
Supplementary Figure 27

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.*



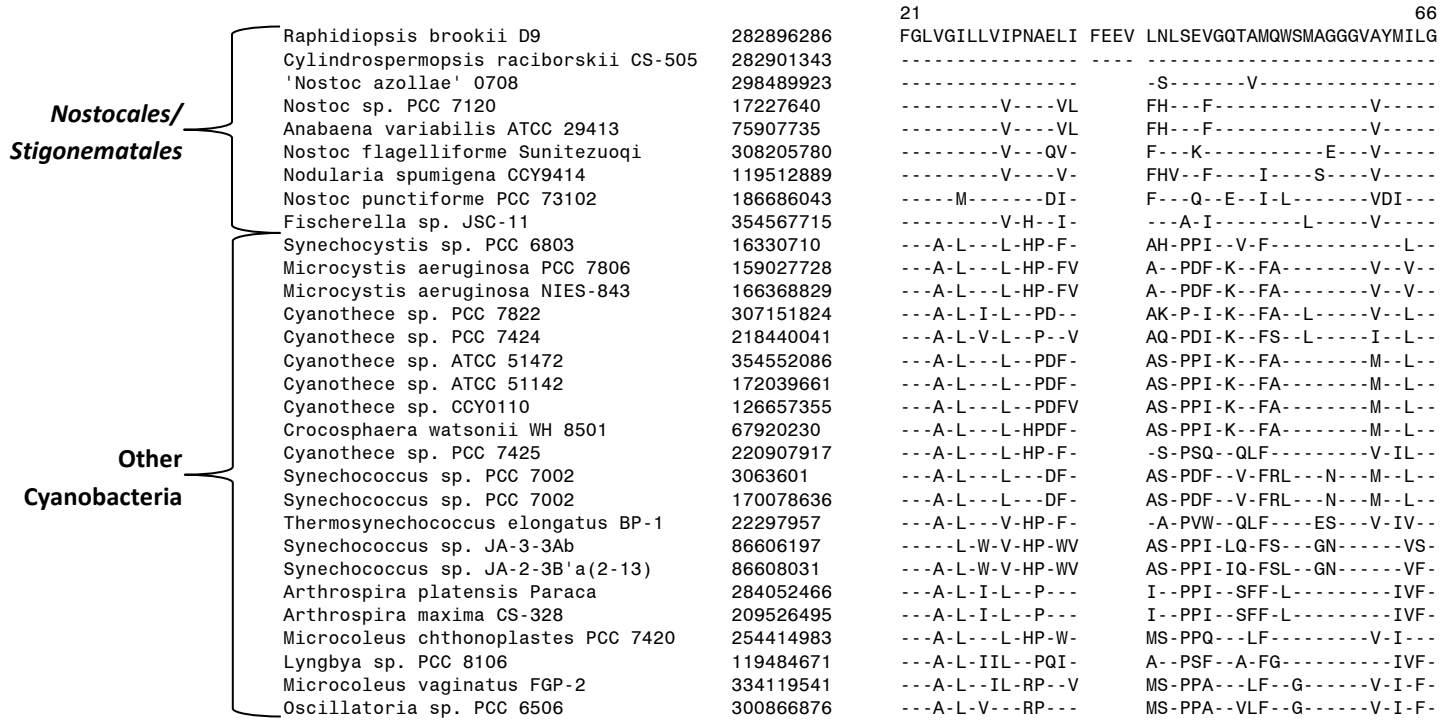
Supplementary Figure 28

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



Supplementary Figure 29

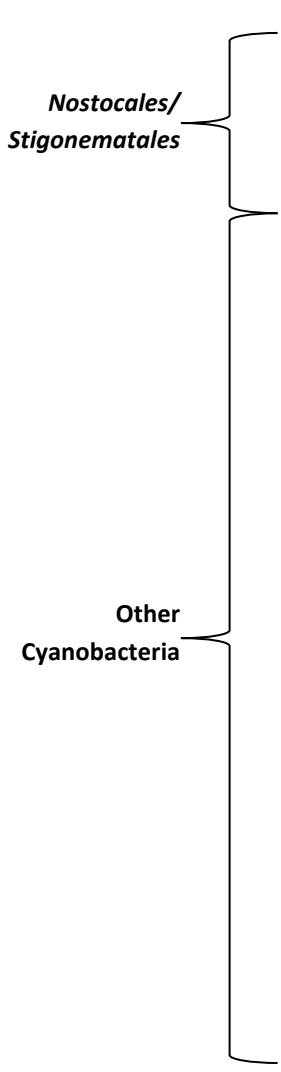
Partial sequence alignment for the protein SNF2-related helicase, showing a 22 aa insert that is specific for *Cylindrospermopsis raciborskii CS-505 and Raphidiopsis brookii D9.* Sequence information for *'Nostoc azollae' 0708* has been removed as it contains an insertion near the insertion of interest.



		21	66
Raphidiopsis brookii D9	282896286	FGLVGILLVIPNAELI	FEEV LNLSEVGQTAMQWSMAGGGVAYMILG
Cylindrospermopsis raciborskii CS-505	282901343	----------	---- ----------
'Nostoc azollae' 0708	298489923	----------	-S-------V----------
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--
Crocosphaera 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--
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

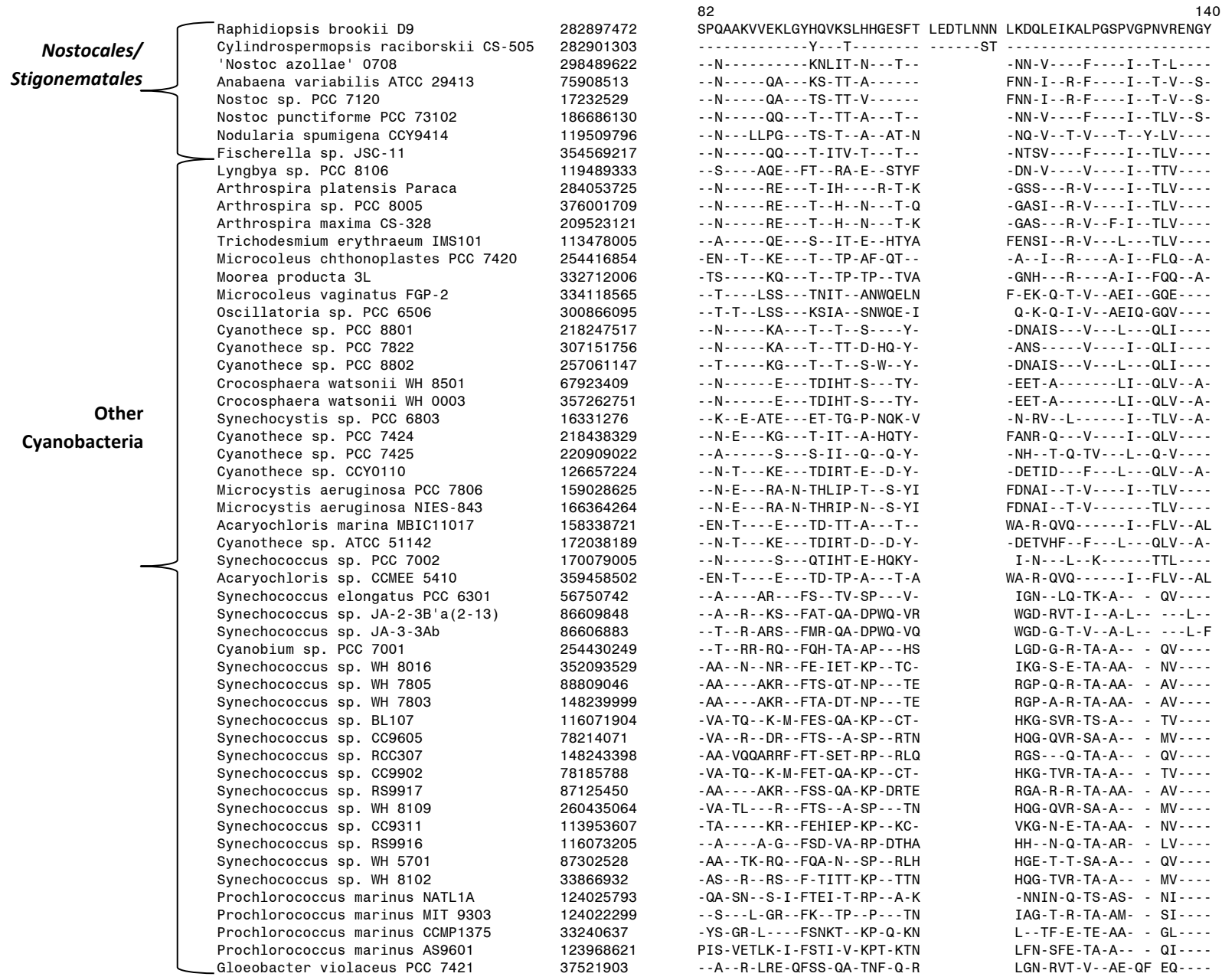
Partial sequence alignment for the hypothetical protein Aazo_0426, showing a 4 aa insert that is specific for *Cylindrospermopsis raciborskii CS-505 and Raphidiopsis brookii D9.*



		79	140
Raphidiopsis brookii D9	282898291	GLACLGLVLAQLFSSQNPFF	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---------
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 sp. PCC 6506	300863785	-ML------SM-VPAKGGI- 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-VPESSKYL 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---
Crocosphaera 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---SFVYPPLT-L- FA	-E-F--VA-----
Cyanothece sp. ATCC 51472	354554405	--L--V---SFVYPPLT-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---SFVYPPLT-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-LPQWGSGC 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

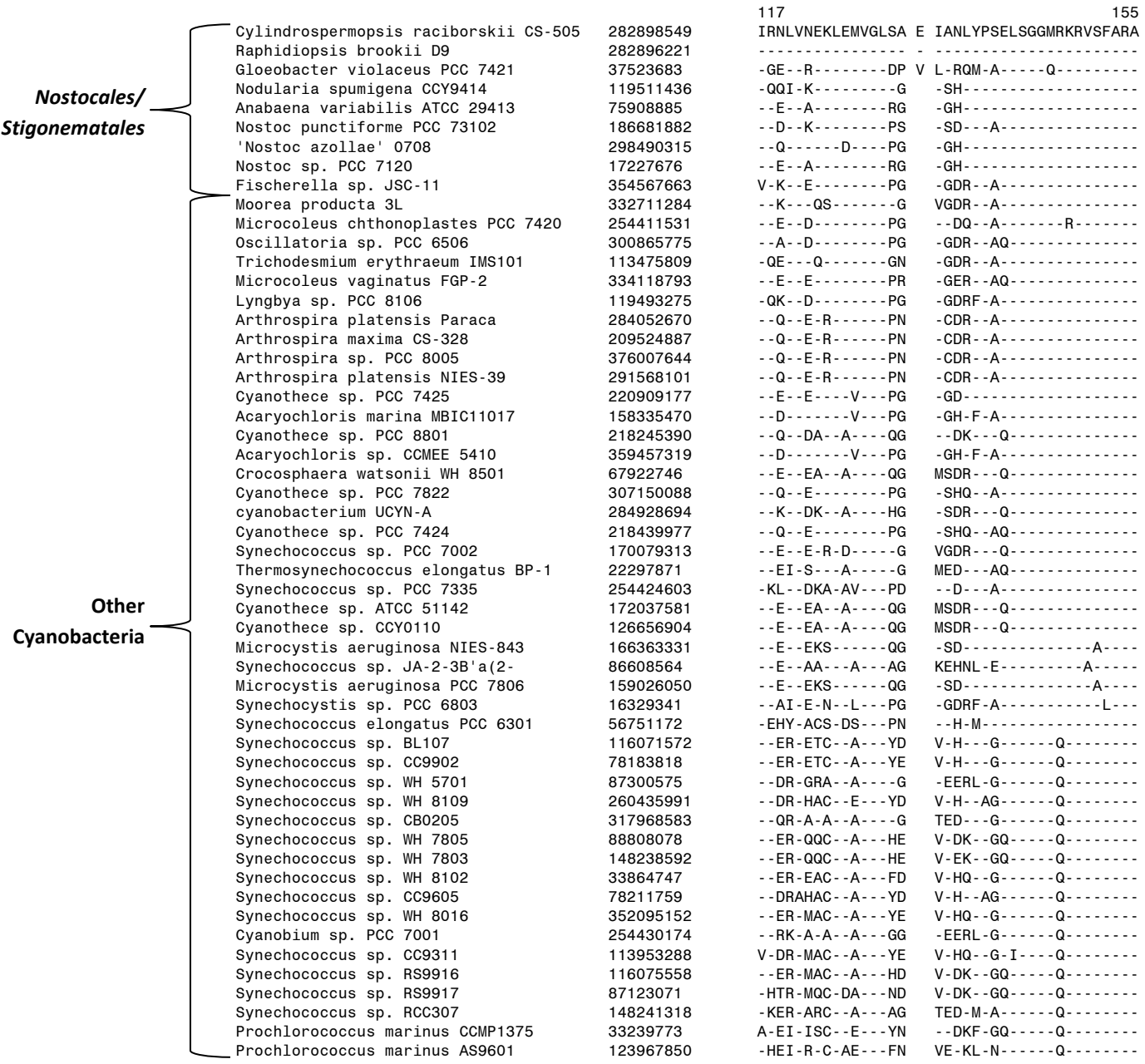
Partial sequence alignment for the hypothetical protein Aazo_2731, showing a 2 aa deletion that is specific for *Cylindrospermopsis raciborskii CS-505 and Raphidiopsis brookii D9.*



		82	140
Raphidiopsis brookii D9	282897472	SPQAAKVVEKLGYHQVKSLHHGESFT LEDTLNNN	LKDQLEIKALPGSPVGPNVRENGY
Cylindrospermopsis raciborskii CS-505	282901303	------------Y---T-------- ------ST	----------
'Nostoc azollae' 0708	298489622	--N----------KNLIT-N---T--	-NN-V----F----I--T-L----
Anabaena variabilis ATCC 29413	75908513	--N-----QA---KS-TT-A------	FNN-I--R-F----I--T-V--S-
Nostoc sp. PCC 7120	17232529	--N-----QA---TS-TT-V------	FNN-I--R-F----I--T-V--S-
Nostoc punctiforme PCC 73102	186686130	--N-----QQ---T--TT-A---T--	-NN-V----F----I--TLV--S-
Nodularia spumigena CCY9414	119509796	--N---LLPG---TS-T--A--AT-N	-NQ-V--T-V---T--Y-LV----
Fischerella sp. JSC-11	354569217	--N-----QQ---T-ITV-T---T--	-NTSV----F----I--TLV----
Lyngbya sp. PCC 8106	119489333	--S----AQE--FT--RA-E--STYF	-DN-V----V----I--TTV----
Arthrospira platensis Paraca	284053725	--N-----RE---T-IH----R-T-K	-GSS---R-V----I--TLV----
Arthrospira sp. PCC 8005	376001709	--N-----RE---T--H--N---T-Q	-GASI--R-V----I--TLV----
Arthrospira maxima CS-328	209523121	--N-----RE---T--H--N---T-K	-GAS---R-V--F-I--TLV----
Trichodesmium erythraeum IMS101	113478005	--A-----QE---S--IT-E--HTYA	FENSI--R-V---L---TLV----
Microcoleus chthonoplastes PCC 7420	254416854	-EN--T--KE---T--TP-AF-QT--	-A--I--R----A-I--FLQ--A-
Moorea producta 3L	332712006	-TS-----KQ---T--TP-TP--TVA	-GNH---R----A-I--FQQ--A-
Microcoleus vaginatus FGP-2	334118565	--T----LSS---TNIT--ANWQELN	F-EK-Q-T-V--AEI--GQE----
Oscillatoria sp. PCC 6506	300866095	--T-T--LSS---KSIA--SNWQE-I	Q-K-Q-I-V--AEIQ-GQV----
Cyanothece sp. PCC 8801	218247517	--N-----KA---T--T--S----Y-	-DNAIS---V---L---QLI----
Cyanothece sp. PCC 7822	307151756	--N-----KA---T--TT-D-HQ-Y-	-ANS-----V----I--QLI----
Cyanothece sp. PCC 8802	257061147	--T-----KG---T--T--S-W--Y-	-DNAIS---V---L---QLI----
Crocosphaera watsonii WH 8501	67923409	--N------E---TDIHT-S---TY-	-EET-A-------LI--QLV--A-
Crocosphaera watsonii WH 0003	357262751	--N------E---TDIHT-S---TY-	-EET-A-------LI--QLV--A-
Synechocystis sp. PCC 6803	16331276	--K--E-ATE---ET-TG-P-NQK-V	-N-RV--L------I--TLV--A-
Cyanothece sp. PCC 7424	218438329	--N-E---KG---T-IT--A-HQTY-	FANR-Q---V----I--QLV----
Cyanothece sp. PCC 7425	220909022	--A------S---S-II--Q--Q-Y-	-NH--T-Q-TV---L--Q-V----
Cyanothece sp. CCY0110	126657224	--N-T---KE---TDIRT-E--D-Y-	-DETID---F---L---QLV--A-
Microcystis aeruginosa PCC 7806	159028625	--N-E---RA-N-THLIP-T--S-YI	FDNAI--T-V----I--TLV----
Microcystis aeruginosa NIES-843	166364264	--N-E---RA-N-THRIP-N--S-YI	FDNAI--T-V-------TLV----
Acaryochloris marina MBIC11017	158338721	-EN-T----E---TD-TT-A---T--	WA-R-QVQ------I--FLV--AL
Cyanothece sp. ATCC 51142	172038189	--N-T---KE---TDIRT-D--D-Y-	-DETVHF--F---L---QLV--A-
Synechococcus sp. PCC 7002	170079005	--N------S---QTIHT-E-HQKY-	I-N---L--K------TTL----
Acaryochloris sp. CCMEE 5410	359458502	-EN-T----E---TD-TP-A---T-A	WA-R-QVQ------I--FLV--AL
Synechococcus elongatus PCC 6301	56750742	--A----AR---FS--TV-SP---V-	IGN--LQ-TK-A-- - QV----
Synechococcus sp. JA-2-3B'a(2-13)	86609848	--A--R--KS--FAT-QA-DPWQ-VR	WGD-RVT-I--A-L-- ---L--
Synechococcus sp. JA-3-3Ab	86606883	--T--R-ARS--FMR-QA-DPWQ-VQ	WGD-G-T-V--A-L-- ---L-F
Cyanobium sp. PCC 7001	254430249	--T--RR-RQ--FQH-TA-AP---HS	LGD-G-R-TA-A-- - QV----
Synechococcus sp. WH 8016	352093529	-AA--N--NR--FE-IET-KP--TC-	IKG-S-E-TA-AA- - NV----
Synechococcus sp. WH 7805	88809046	-AA----AKR--FTS-QT-NP---TE	RGP-Q-R-TA-AA- - AV----
Synechococcus sp. WH 7803	148239999	-AA----AKR--FTA-DT-NP---TE	RGP-A-R-TA-AA- - AV----
Synechococcus sp. BL107	116071904	-VA-TQ--K-M-FES-QA-KP--CT-	HKG-SVR-TS-A-- - TV----
Synechococcus sp. CC9605	78214071	-VA--R--DR--FTS--A-SP--RTN	HQG-QVR-SA-A-- - MV----
Synechococcus sp. RCC307	148243398	-AA-VQQARRF-FT-SET-RP--RLQ	RGS---Q-TA-A-- - QV----
Synechococcus sp. CC9902	78185788	-VA-TQ--K-M-FET-QA-KP--CT-	HKG-TVR-TA-A-- - TV----
Synechococcus sp. RS9917	87125450	-AA----AKR--FSS-QA-KP-DRTE	RGA-R-R-TA-AA- - AV----
Synechococcus sp. WH 8109	260435064	-VA-TL---R--FTS--A-SP---TN	HQG-QVR-SA-A-- - MV----
Synechococcus sp. CC9311	113953607	-TA-----KR--FEHIEP-KP--KC-	VKG-N-E-TA-AA- - NV----
Synechococcus sp. RS9916	116073205	--A----A-G--FSD-VA-RP-DTHA	HH--N-Q-TA-AR- - LV----
Synechococcus sp. WH 5701	87302528	-AA--TK-RQ--FQA-N--SP--RLH	HGE-T-T-SA-A-- - QV----
Synechococcus sp. WH 8102	33866932	-AS--R--RS--F-TITT-KP--TTN	HQG-TVR-TA-A-- - MV----
Prochlorococcus marinus NATL1A	124025793	-QA-SN--S-I-FTEI-T-RP--A-K	-NNIN-Q-TS-AS- - NI----
Prochlorococcus marinus MIT 9303	124022299	--S---L-GR--FK--TP--P---TN	IAG-T-R-TA-AM- - SI----
Prochlorococcus marinus CCMP1375	33240637	-YS-GR-L----FSNKT--KP-Q-KN	L--TF-E-TE-AA- - GL----
Prochlorococcus marinus AS9601	123968621	PIS-VETLK-I-FSTI-V-KPT-KTN	LFN-SFE-TA-A-- - QI----
Gloeobacter violaceus PCC 7421	37521903	--A--R-LRE-QFSS-QA-TNF-Q-R	LGN-RVT-V--AE-QF EQ----

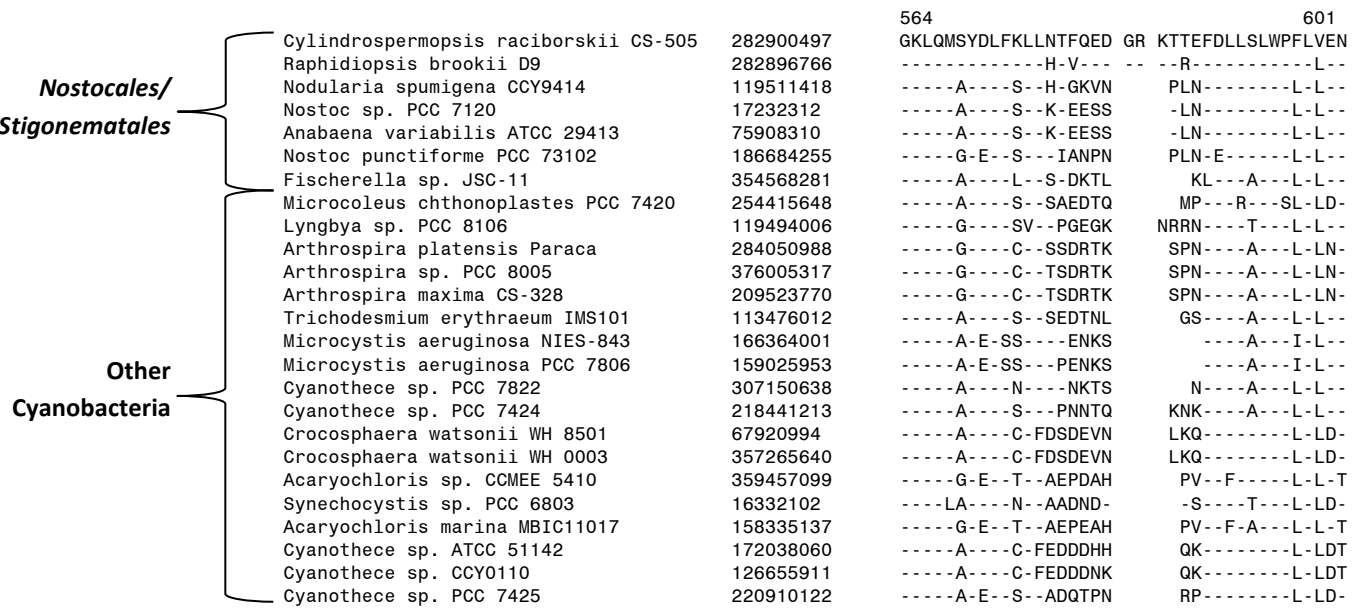
Supplementary Figure 32

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.*



Supplementary Figure 33

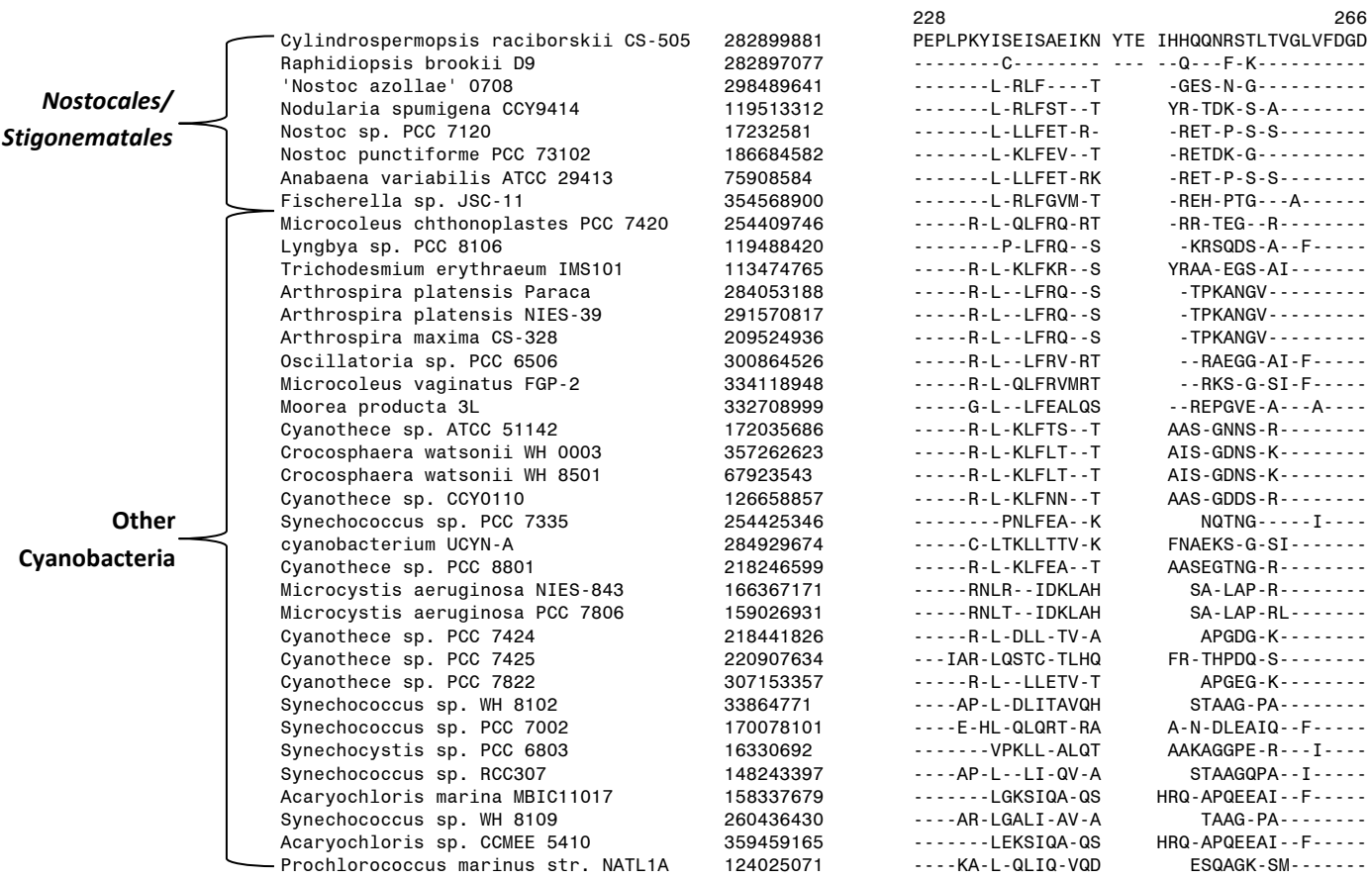
Partial sequence alignment for the ABC transporter-like protein, showing a 1 aa insert that is specific for *Cylindrospermopsis raciborskii CS-505 and Raphidiopsis brookii D9.*



		564 601
Cylindrospermopsis raciborskii CS-505	282900497	GKLQMSYDLFKLLNTFQED GR KTTEFDLLSLWPFLVEN
Raphidiopsis brookii D9	282896766	----------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--PGEGK 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--
Microcystis aeruginosa NIES-843	166364001	-----A-E-SS----ENKS ----A---I-L--
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--
Crocosphaera watsonii WH 8501	67920994	-----A----C-FDSDEVN LKQ--------L-LD-
Crocosphaera 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
Cylindrospermopsis raciborskii CS-505	282899881	PEPLPKYISEISAEIKN	YTE IHHQQNRSTLTVGLVFDGD
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-RLFGVM-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----
Cyanothece sp. ATCC 51142	172035686	-----R-L-KLFTS--T	AAS-GNNS-R--------
Crocosphaera watsonii WH 0003	357262623	-----R-L-KLFLT--T	AIS-GDNS-K--------
Crocosphaera 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-------

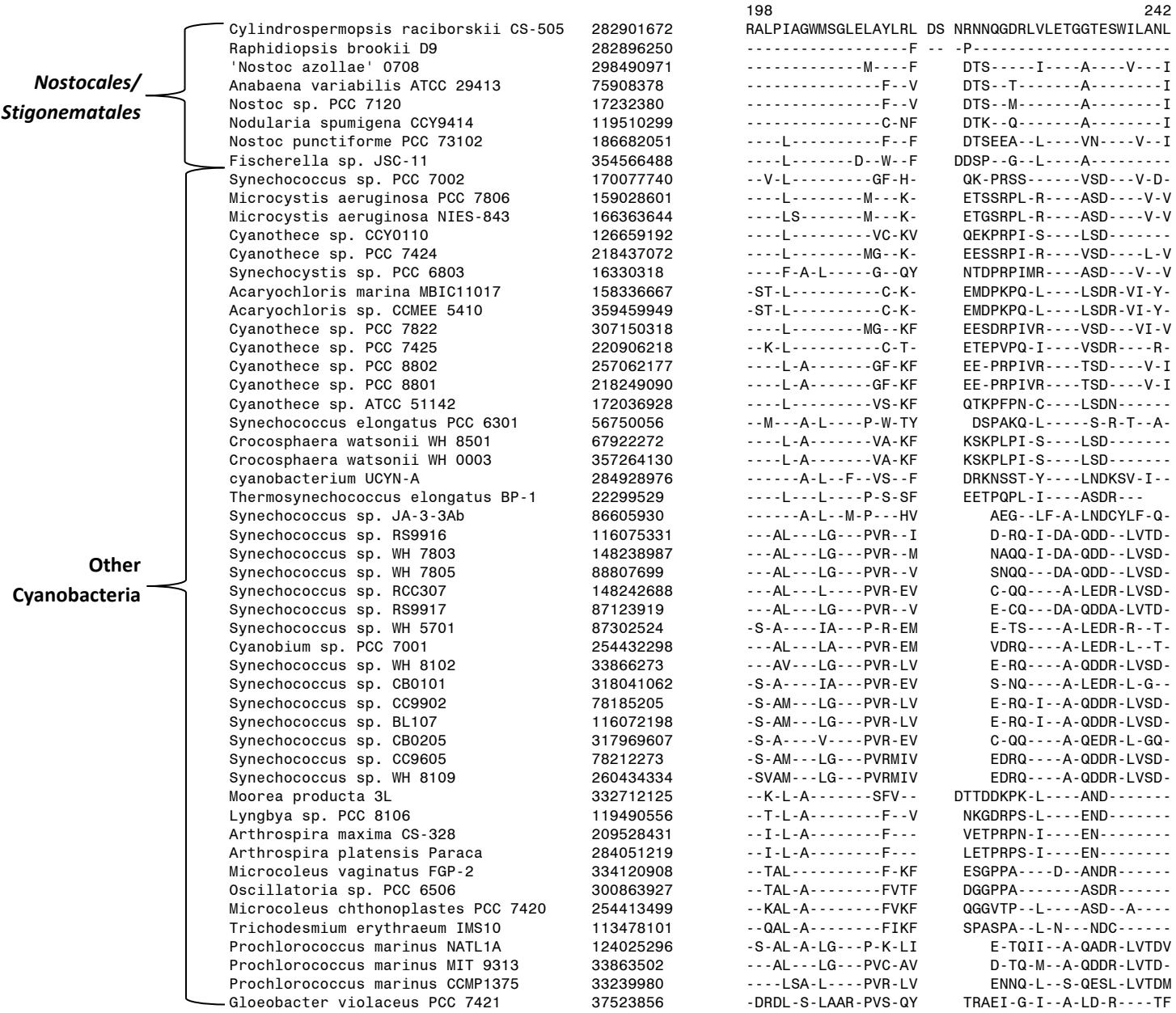
Supplementary Figure 35

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
Nostocales/ Stigonematales	Raphidiopsis brookii D9	282898311	GIDAVKVGVEVDISGG G LPGIIILGLPDSAIQESKERV
	Cylindrospermopsis raciborskii CS-505	282899606	---------- - ----------
	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-------
Other Cyanobacteria	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----
	Trichodesmium erythraeum IMS101	113478131	--E-I-------V--- --KVVVV----A-V-------
	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---------- ----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---
	Crocosphaera watsonii WH 0003	357261924	----------VA-- --A-T-V----T-V---R---
	Synechococcus sp. PCC 7002	170076988	-V--LR---------- --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

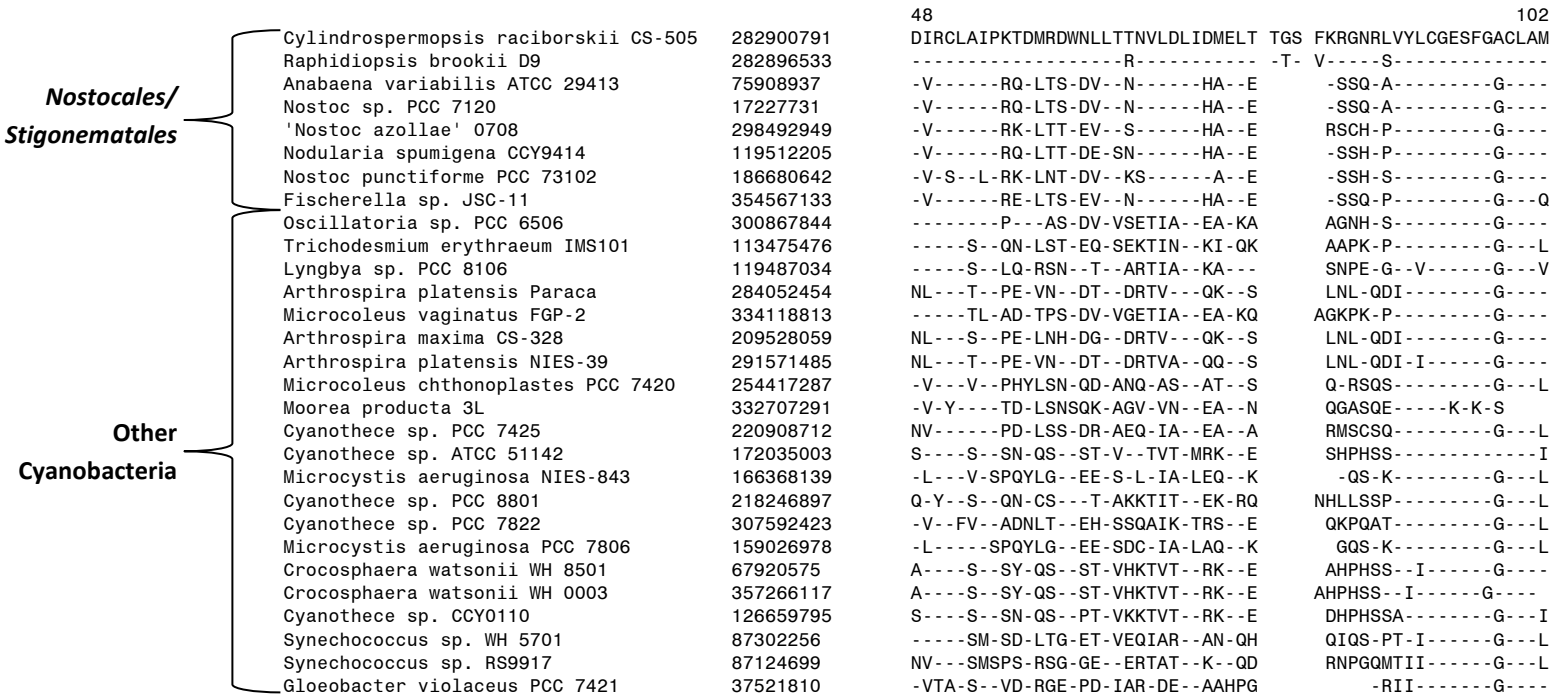
Partial sequence alignment for the Mg chelatase-related protein, showing a 1 aa insert that is specific for *Cylindrospermopsis raciborskii CS-505 and Raphidiopsis brookii D9.*



		198	242
Cylindrospermopsis raciborskii CS-505	282901672	RALPIAGWMSGLELAYLRL DS	NRNNQGDRLVLETGGTESWILANL
Raphidiopsis brookii D9	282896250	----------F -- -P	----------
'Nostoc azollae' 0708	298490971	----------M----F	DTS-----I----A----V---I
Anabaena variabilis ATCC 29413	75908378	----------F--V	DTS--T-------A--------I
Nostoc sp. PCC 7120	17232380	----------F--V	DTS--M-------A--------I
Nodularia spumigena CCY9414	119510299	----------C-NF	DTK--Q-------A--------I
Nostoc punctiforme PCC 73102	186682051	----L----------F--F	DTSEEA--L----VN----V--I
Fischerella sp. JSC-11	354566488	----L-------D--W--F	DDSP--G--L----A---------
Synechococcus sp. PCC 7002	170077740	--V-L---------GF-H-	QK-PRSS------VSD---V-D-
Microcystis aeruginosa PCC 7806	159028601	----L--------M---K-	ETSSRPL-R----ASD----V-V
Microcystis aeruginosa NIES-843	166363644	----LS-------M---K-	ETGSRPL-R----ASD----V-V
Cyanothece sp. CCY0110	126659192	----L---------VC-KV	QEKPRPI-S----LSD-------
Cyanothece sp. PCC 7424	218437072	----L--------MG--K-	EESSRPI-R----VSD----L-V
Synechocystis sp. PCC 6803	16330318	----F-A-L-----G--QY	NTDPRPIMR----ASD---V--V
Acaryochloris marina MBIC11017	158336667	-ST-L----------C-K-	EMDPKPQ-L----LSDR-VI-Y-
Acaryochloris sp. CCMEE 5410	359459949	-ST-L----------C-K-	EMDPKPQ-L----LSDR-VI-Y-
Cyanothece sp. PCC 7822	307150318	----L--------MG--KF	EESDRPIVR----VSD---VI-V
Cyanothece sp. PCC 7425	220906218	--K-L----------C-T-	ETEPVPQ-I----VSDR----R-
Cyanothece sp. PCC 8802	257062177	----L-A-------GF-KF	EE-PRPIVR----TSD----V-I
Cyanothece sp. PCC 8801	218249090	----L-A-------GF-KF	EE-PRPIVR----TSD----V-I
Cyanothece sp. ATCC 51142	172036928	----L---------VS-KF	QTKPFPN-C----LSDN------
Synechococcus elongatus PCC 6301	56750056	--M---A-L----P-W-TY	DSPAKQ-L-----S-R-T--A-
Crocosphaera watsonii WH 8501	67922272	----L-A-------VA-KF	KSKPLPI-S----LSD-------
Crocosphaera watsonii WH 0003	357264130	----L-A-------VA-KF	KSKPLPI-S----LSD-------
cyanobacterium UCYN-A	284928976	------A-L--F--VS--F	DRKNSST-Y----LNDKSV-I--
Thermosynechococcus elongatus BP-1	22299529	----L---L----P-S-SF	EETPQPL-I----ASDR---
Synechococcus sp. JA-3-3Ab	86605930	------A-L--M-P---HV	AEG--LF-A-LNDCYLF-Q-
Synechococcus sp. RS9916	116075331	---AL---LG---PVR--I	D-RQ-I-DA-QDD--LVTD-
Synechococcus sp. WH 7803	148238987	---AL---LG---PVR--M	NAQQ-I-DA-QDD--LVSD-
Synechococcus sp. WH 7805	88807699	---AL---LG---PVR--V	SNQQ---DA-QDD--LVSD-
Synechococcus sp. RCC307	148242688	---AL---L----PVR-EV	C-QQ----A-LEDR-LVSD-
Synechococcus sp. RS9917	87123919	---AL---LG---PVR--V	E-CQ---DA-QDDA-LVTD-
Synechococcus sp. WH 5701	87302524	-S-A----IA---P-R-EM	E-TS----A-LEDR-R--T-
Cyanobium sp. PCC 7001	254432298	---AL---LA---PVR-EM	VDRQ----A-LEDR-L--T-
Synechococcus sp. WH 8102	33866273	---AV---LG---PVR-LV	E-RQ----A-QDDR-LVSD-
Synechococcus sp. CB0101	318041062	-S-A----IA---PVR-EV	S-NQ----A-LEDR-L-G--
Synechococcus sp. CC9902	78185205	-S-AM---LG---PVR-LV	E-RQ-I--A-QDDR-LVSD-
Synechococcus sp. BL107	116072198	-S-AM---LG---PVR-LV	E-RQ-I--A-QDDR-LVSD-
Synechococcus sp. CB0205	317969607	-S-A----V----PVR-EV	C-QQ----A-QEDR-L-GQ-
Synechococcus sp. CC9605	78212273	-S-AM---LG---PVRMIV	EDRQ----A-QDDR-LVSD-
Synechococcus sp. WH 8109	260434334	-SVAM---LG---PVRMIV	EDRQ----A-QDDR-LVSD-
Moorea producta 3L	332712125	--K-L-A-------SFV--	DTTDDKPK-L----AND-------
Lyngbya sp. PCC 8106	119490556	--T-L-A--------F--V	NKGDRPS-L----END-------
Arthrospira maxima CS-328	209528431	--I-L-A--------F---	VETPRPN-I----EN--------
Arthrospira platensis Paraca	284051219	--I-L-A--------F---	LETPRPS-I----EN--------
Microcoleus vaginatus FGP-2	334120908	--TAL----------F-KF	ESGPPA----D--ANDR------
Oscillatoria sp. PCC 6506	300863927	--TAL-A--------FVTF	DGGPPA-------ASDR------
Microcoleus chthonoplastes PCC 7420	254413499	--KAL-A--------FVKF	QGGVTP--L----ASD--A----
Trichodesmium erythraeum IMS101	113478101	--QAL-A--------FIKF	SPASPA--L-N---NDC------
Prochlorococcus marinus NATL1A	124025296	-S-AL-A-LG---P-K-LI	E-TQII--A-QADR-LVTDV
Prochlorococcus marinus MIT 9313	33863502	---AL---LG---PVC-AV	D-TQ-M--A-QDDR-LVTD-
Prochlorococcus marinus CCMP1375	33239980	----LSA-L----PVR-LV	ENNQ-L--S-QESL-LVTDM
Gloeobacter violaceus PCC 7421	37523856	-DRDL-S-LAAR-PVS-QY	TRAEI-G-I--A-LD-R----TF

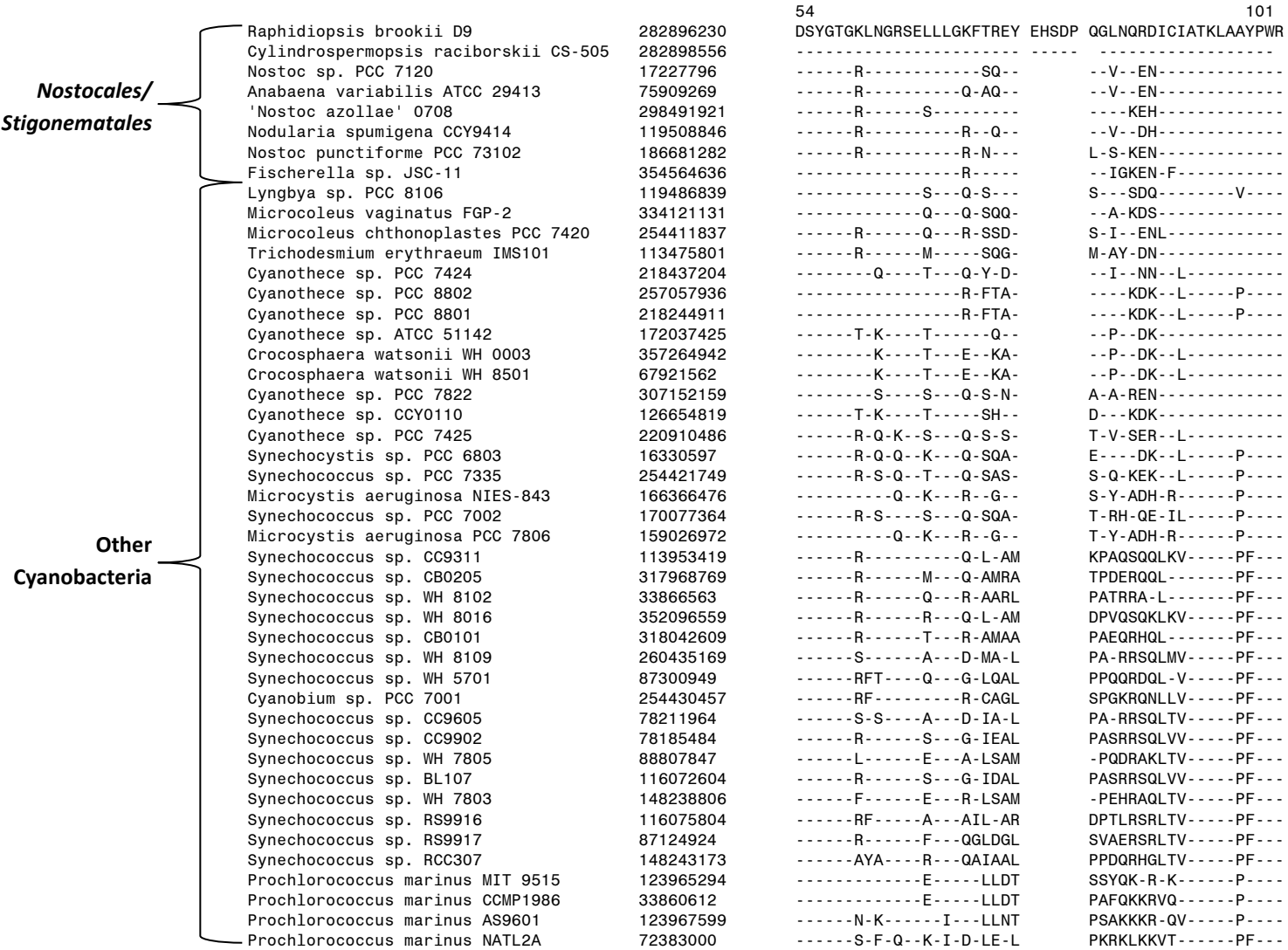
Supplementary Figure 37

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.*



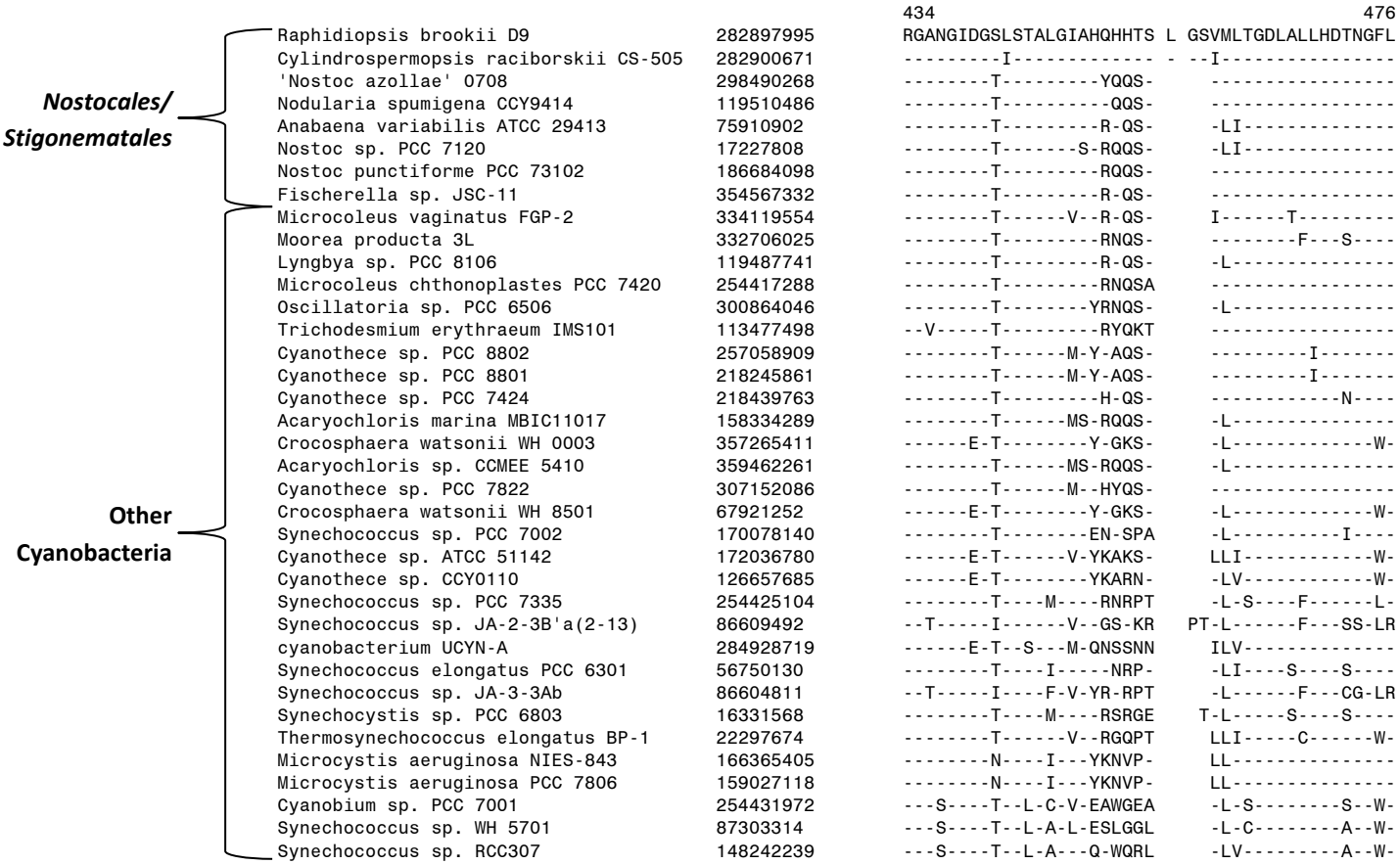
Supplementary Figure 38

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.*



Supplementary Figure 39

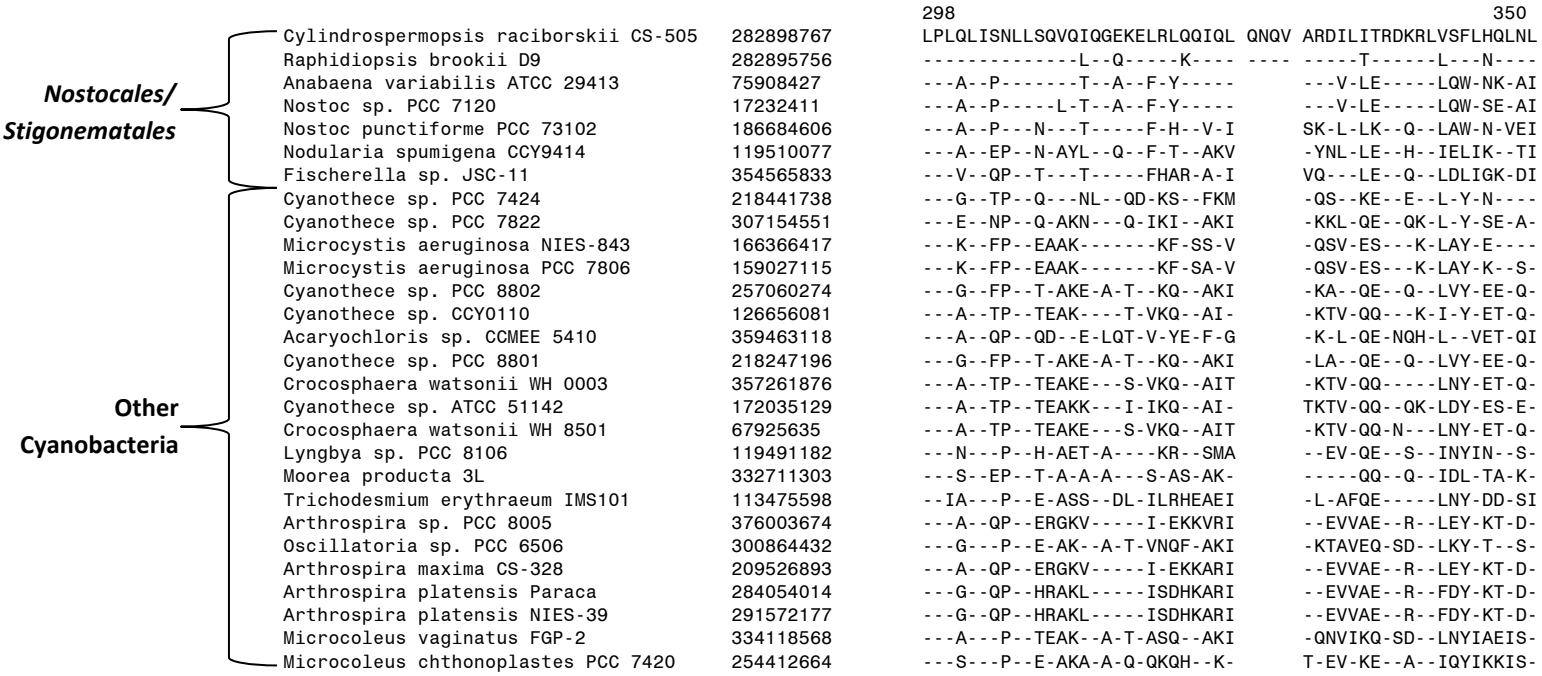
Partial sequence alignment for the protein aldo/keto reductase, showing a 5 aa insert that is specific for *Cylindrospermopsis raciborskii CS-505 and Raphidiopsis brookii D9.*



		434 476
Raphidiopsis brookii D9	282897995	RGANGIDGSLSTALGIAHQHHTS L GSVMLTGDLALLHDTNGFL
Cylindrospermopsis raciborskii CS-505	282900671	---------I---------- - --I----------
'Nostoc azollae' 0708	298490268	--------T---------YQQS- --------
Nodularia spumigena CCY9414	119510486	--------T----------QQS- --------
Anabaena variabilis ATCC 29413	75910902	--------T---------R-QS- -LI----------
Nostoc sp. PCC 7120	17227808	--------T-------S-RQQS- -LI----------
Nostoc punctiforme PCC 73102	186684098	--------T---------RQQS- --------
Fischerella sp. JSC-11	354567332	--------T---------R-QS- --------
Microcoleus vaginatus FGP-2	334119554	--------T------V--R-QS- I------T---------
Moorea producta 3L	332706025	--------T---------RNQS- --------F---S----
Lyngbya sp. PCC 8106	119487741	--------T---------R-QS- -L----------
Microcoleus chthonoplastes PCC 7420	254417288	--------T---------RNQSA --------
Oscillatoria sp. PCC 6506	300864046	--------T--------YRNQS- -L----------
Trichodesmium erythraeum IMS101	113477498	--V-----T---------RYQKT --------
Cyanothece sp. PCC 8802	257058909	--------T------M-Y-AQS- ---------I-------
Cyanothece sp. PCC 8801	218245861	--------T------M-Y-AQS- ---------I-------
Cyanothece sp. PCC 7424	218439763	--------T---------H-QS- ------------N----
Acaryochloris marina MBIC11017	158334289	--------T------MS-RQQS- -L----------
Crocosphaera watsonii WH 0003	357265411	------E-T--------Y-GKS- -L------------W-
Acaryochloris sp. CCMEE 5410	359462261	--------T------MS-RQQS- -L----------
Cyanothece sp. PCC 7822	307152086	--------T------M--HYQS- --------
Crocosphaera watsonii WH 8501	67921252	------E-T--------Y-GKS- -L------------W-
Synechococcus sp. PCC 7002	170078140	--------T--------EN-SPA -L----------I----
Cyanothece sp. ATCC 51142	172036780	------E-T------V-YKAKS- LLI------------W-
Cyanothece sp. CCY0110	126657685	------E-T--------YKARN- -LV------------W-
Synechococcus sp. PCC 7335	254425104	--------T----M----RNRPT -L-S----F------L-
Synechococcus sp. JA-2-3B'a(2-13)	86609492	--T-----I------V--GS-KR PT-L------F---SS-LR
cyanobacterium UCYN-A	284928719	------E-T--S---M-QNSSNN ILV----------
Synechococcus elongatus PCC 6301	56750130	--------T----I-----NRP- -LI----S----S----
Synechococcus sp. JA-3-3Ab	86604811	--T-----I----F-V-YR-RPT -L------F---CG-LR
Synechocystis sp. PCC 6803	16331568	--------T----M----RSRGE T-L-----S----S----
Thermosynechococcus elongatus BP-1	22297674	--------T------V--RGQPT LLI-----C------W-
Microcystis aeruginosa NIES-843	166365405	--------N----I---YKNVP- LL----------
Microcystis aeruginosa PCC 7806	159027118	--------N----I---YKNVP- LL----------
Cyanobium sp. PCC 7001	254431972	---S----T--L-C-V-EAWGEA -L-S--------S--W-
Synechococcus sp. WH 5701	87303314	---S----T--L-A-L-ESLGGL -L-C--------A--W-
Synechococcus sp. RCC307	148242239	---S----T--L-A---Q-WQRL -LV---------A--W-

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
Cylindrospermopsis raciborskii CS-505	282898767	LPLQLISNLLSQVQIQGEKELRLQQIQL QNQV	ARDILITRDKRLVSFLHQLNL
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----
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-
Crocosphaera 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-
Crocosphaera 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-

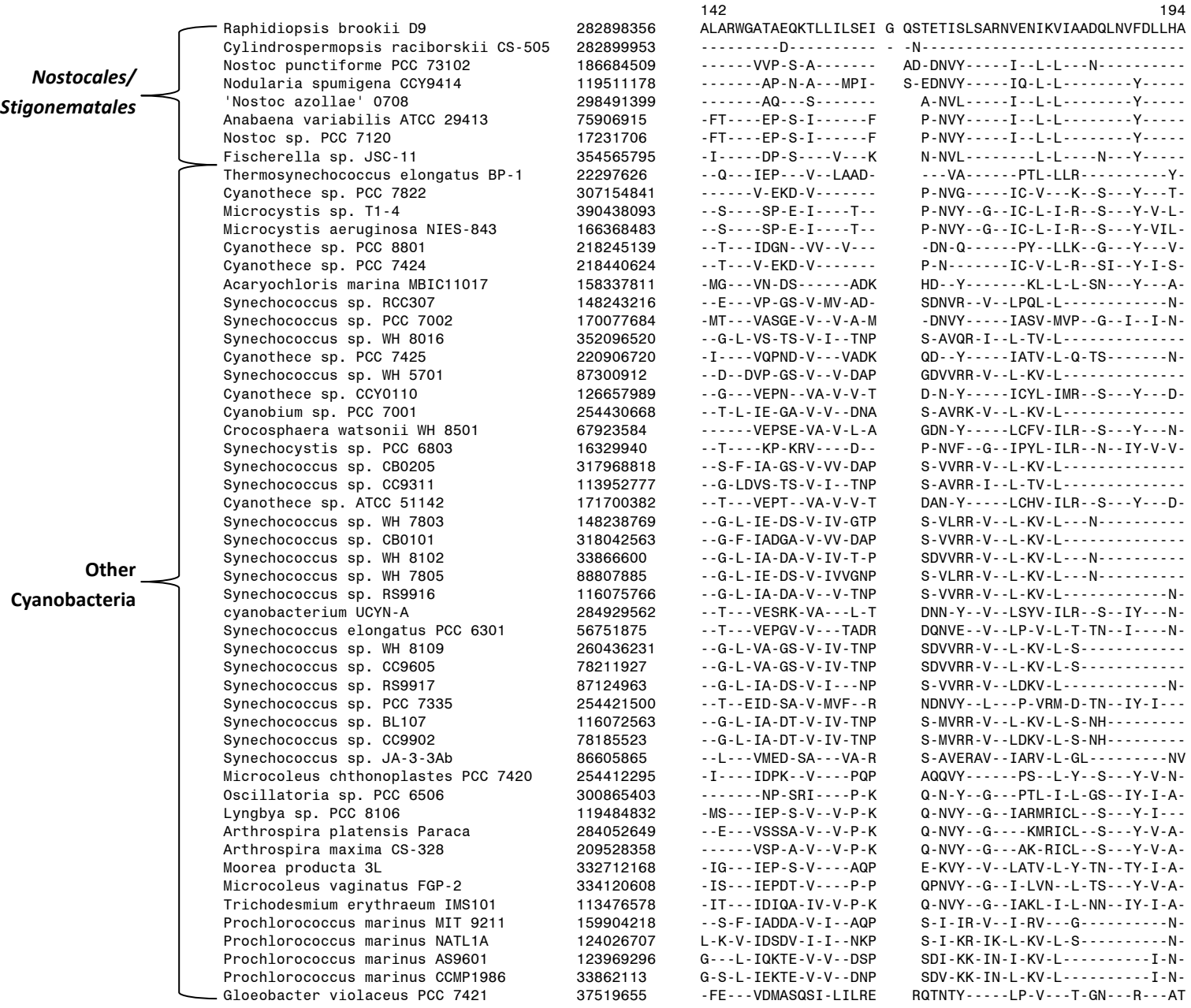
Supplementary Figure 41

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.*

		348	389
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--
Arthrospira platensis S6	206585493	--VR-R---IK---D-FN	RH--LA-FVQ-FRD-MG-I-S
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. CCMEE 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–

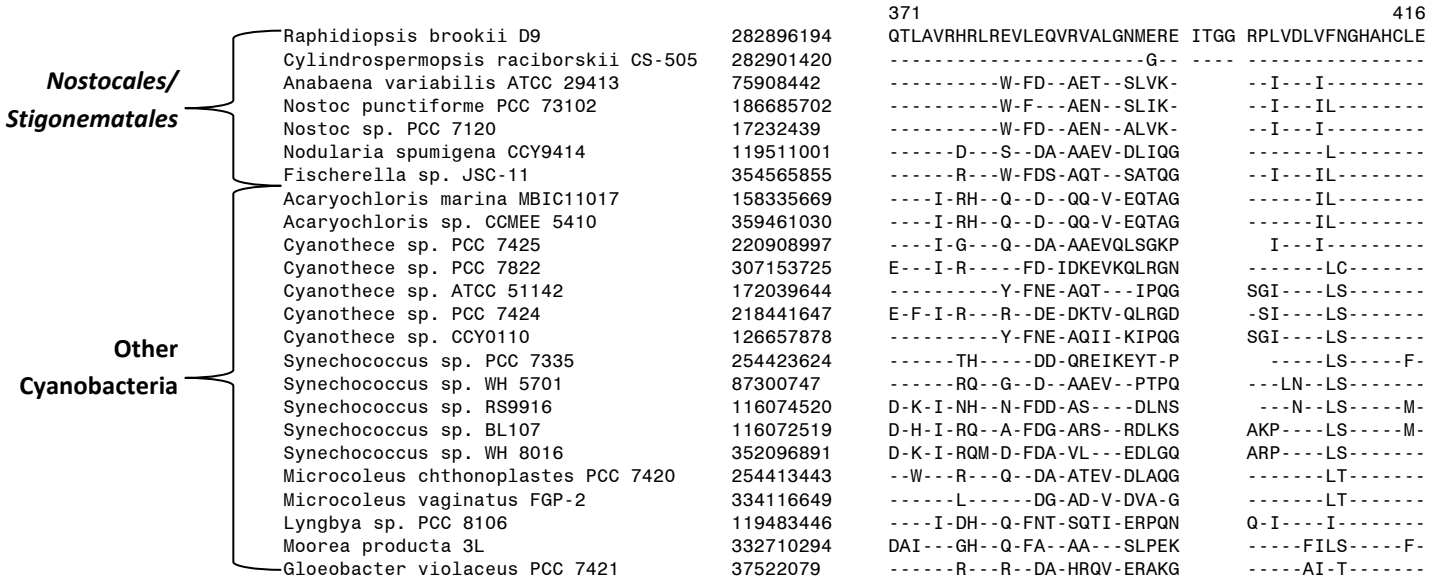
Supplementary Figure 42

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.*



Supplementary Figure 43

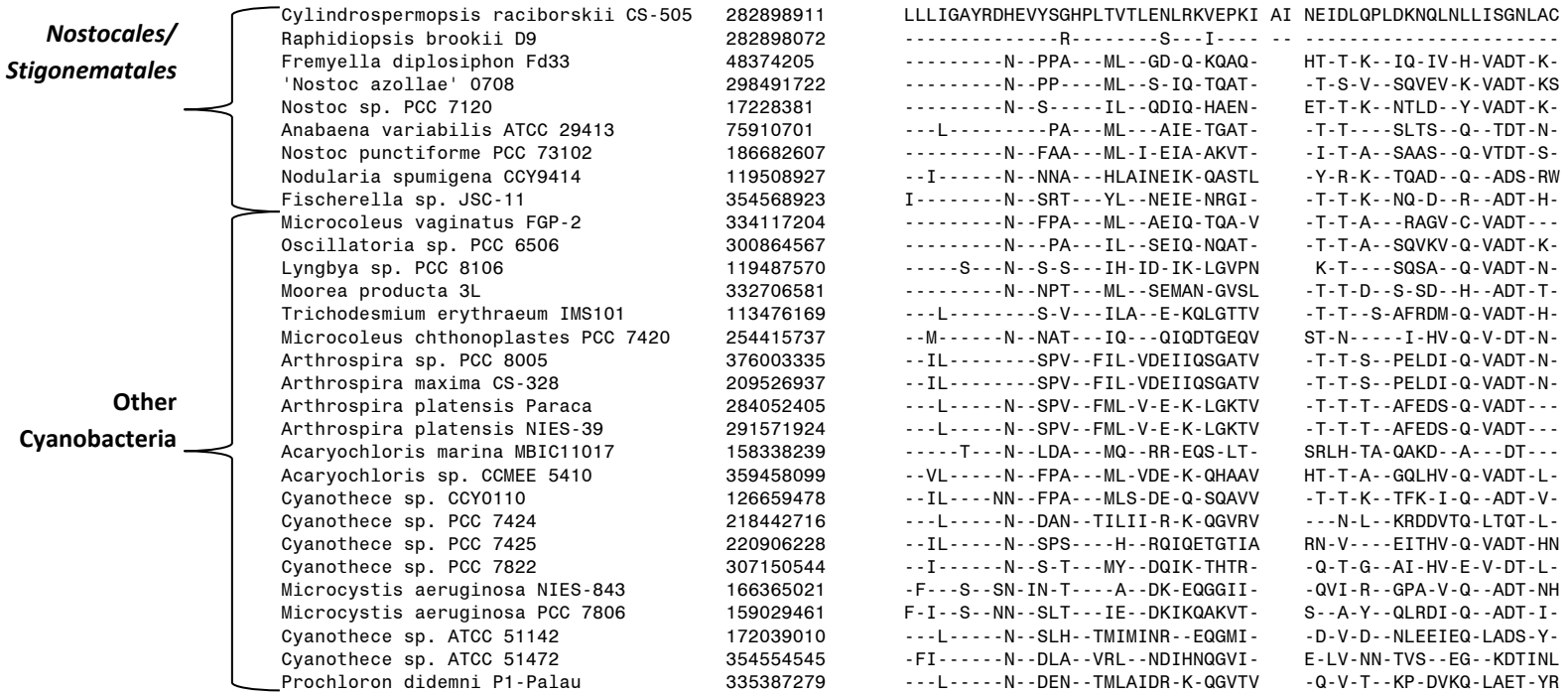
Partial sequence alignment for the ribosomal protein L4/L1e, showing a 1 aa insert that is specific for *Cylindrospermopsis raciborskii CS-505 and Raphidiopsis brookii D9.*



		371	416
Raphidiopsis brookii D9	282896194	QTLAVRHRLREVLEQVRVALGNMERE	ITGG RPLVDLVFNGHAHCLE
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. CCMEE 5410	359461030	----I-RH--Q--D--QQ-V-EQTAG	------IL--------
Cyanothece sp. PCC 7425	220908997	----I-G---Q--DA-AAEVQLSGKP	I---I---------
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-------

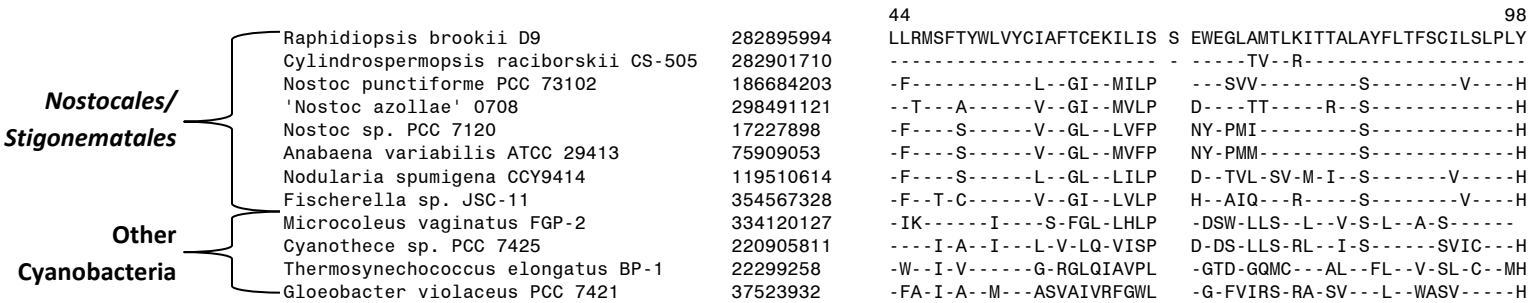
Supplementary Figure 44

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



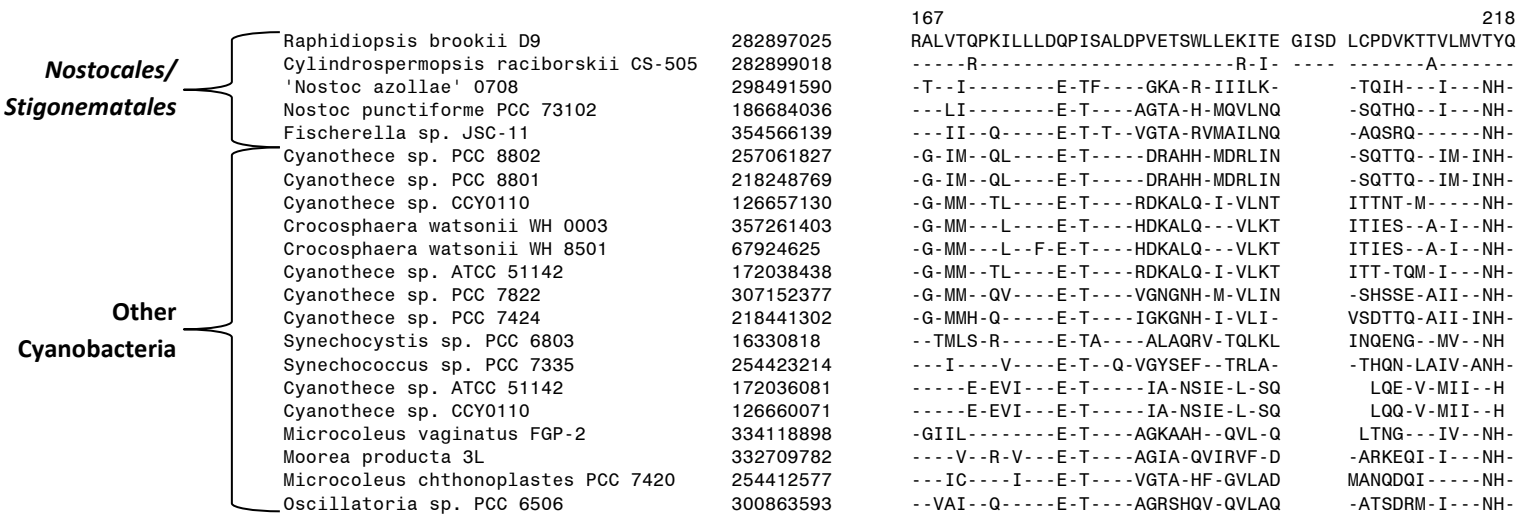
Supplementary Figure 45

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.*



Supplementary Figure 46

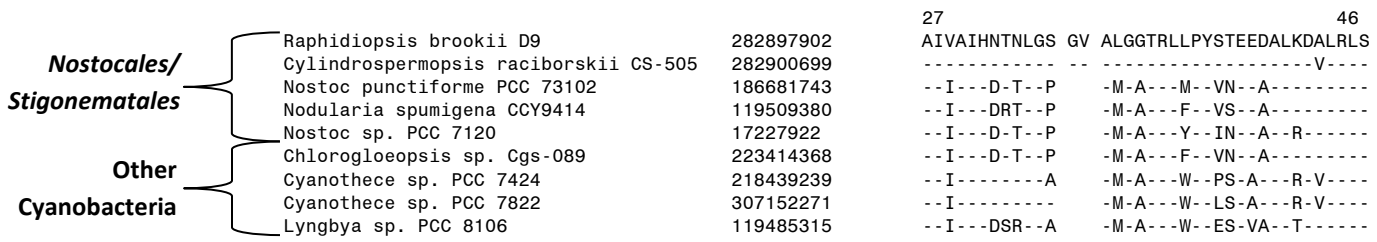
Partial sequence alignment for the hypothetical protein Npun_R4078, showing a 2 aa insert that is specific for *Cylindrospermopsis raciborskii CS-505 and Raphidiopsis brookii D9.*



		167	218
Raphidiopsis brookii D9	282897025	RALVTQPKILLLDQPISALDPVETSWLLEKITE GISD	LCPDVKTTVLMVTYQ
Cylindrospermopsis raciborskii CS-505	282899018	-----R----------R-I- ----	-------A-------
'Nostoc azollae' 0708	298491590	-T--I--------E-TF----GKA-R-IIILK-	-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	-AQSRQ------NH-
Cyanothece sp. PCC 8802	257061827	-G-IM--QL----E-T-----DRAHH-MDRLIN	-SQTTQ--IM-INH-
Cyanothece sp. PCC 8801	218248769	-G-IM--QL----E-T-----DRAHH-MDRLIN	-SQTTQ--IM-INH-
Cyanothece sp. CCY0110	126657130	-G-MM--TL----E-T----RDKALQ-I-VLNT	ITTNT-M-----NH-
Crocosphaera watsonii WH 0003	357261403	-G-MM---L----E-T----HDKALQ---VLKT	ITIES--A-I--NH-
Crocosphaera watsonii WH 8501	67924625	-G-MM---L--F-E-T----HDKALQ---VLKT	ITIES--A-I--NH-
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	LQQ-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-

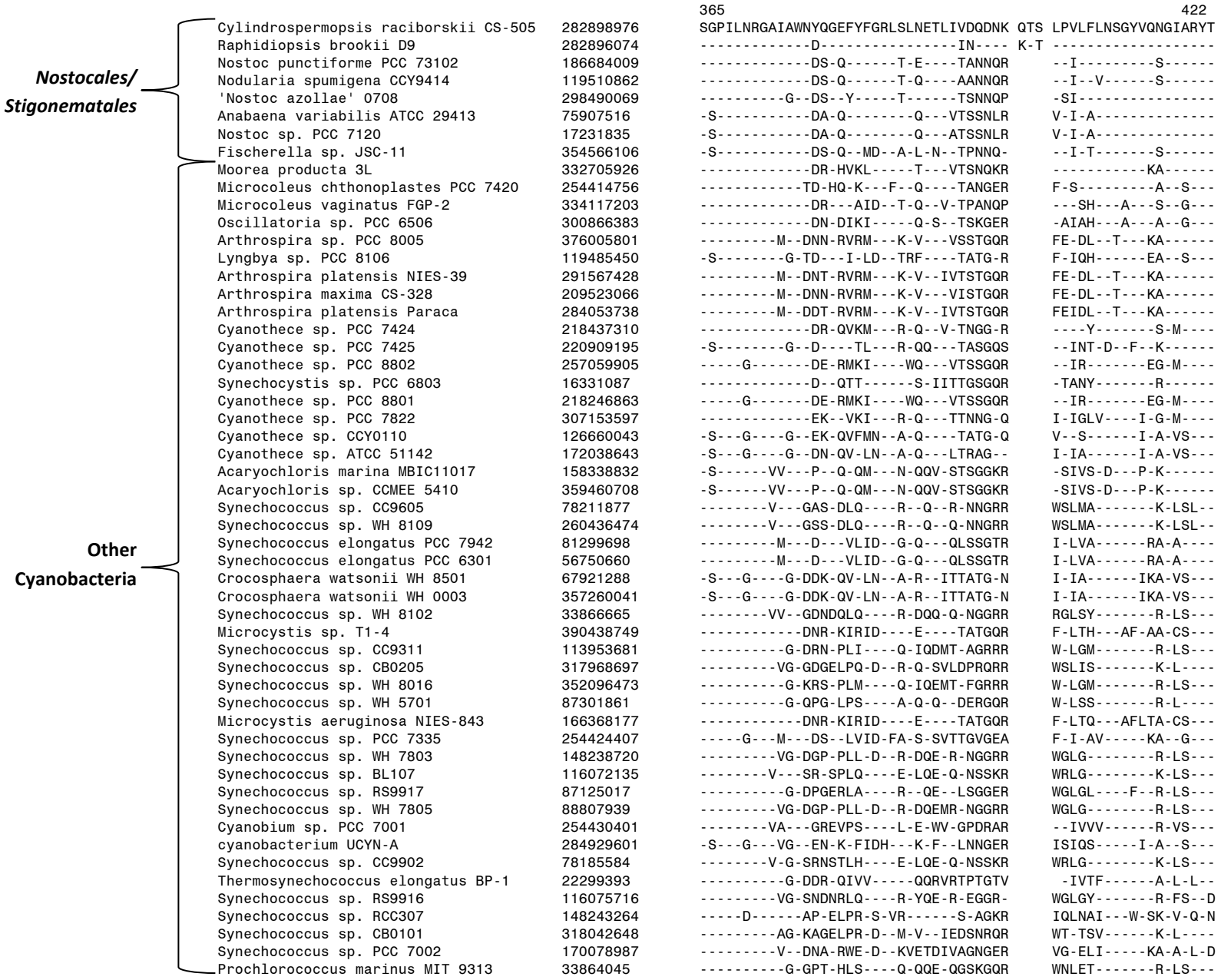
Supplementary Figure 47

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.*



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*.



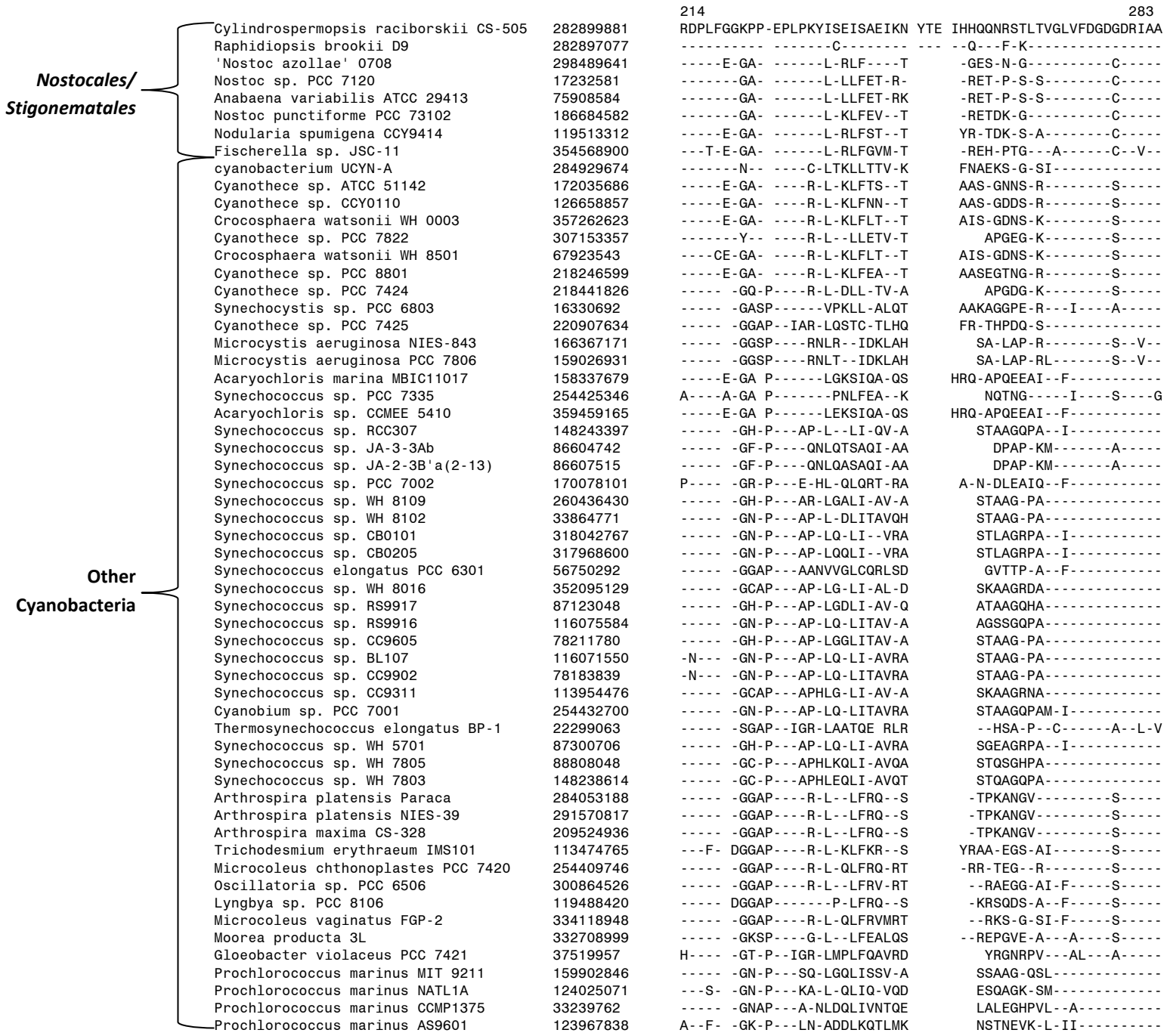
Supplementary Figure 49

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

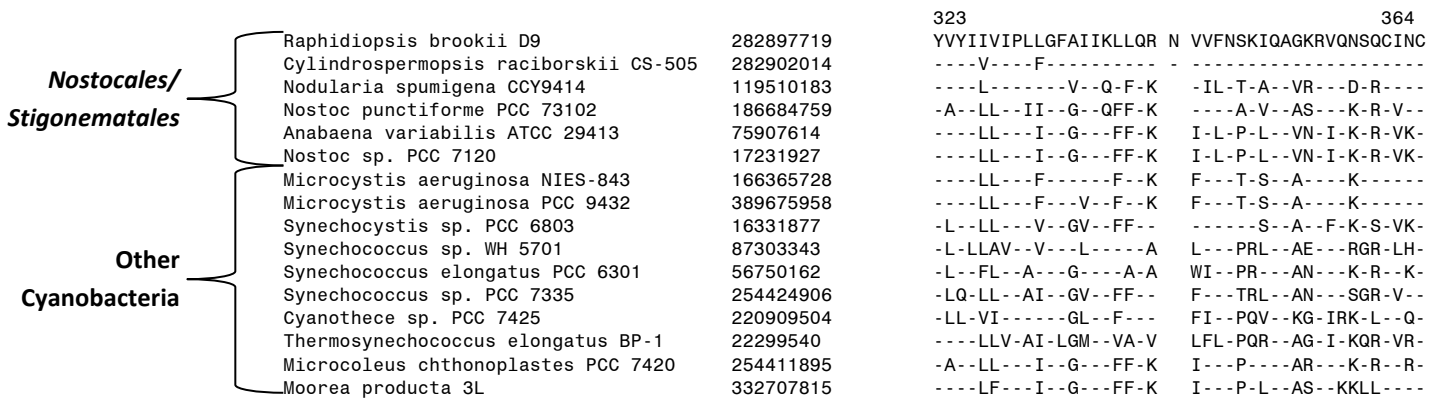
Supplementary Figure 50

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.*



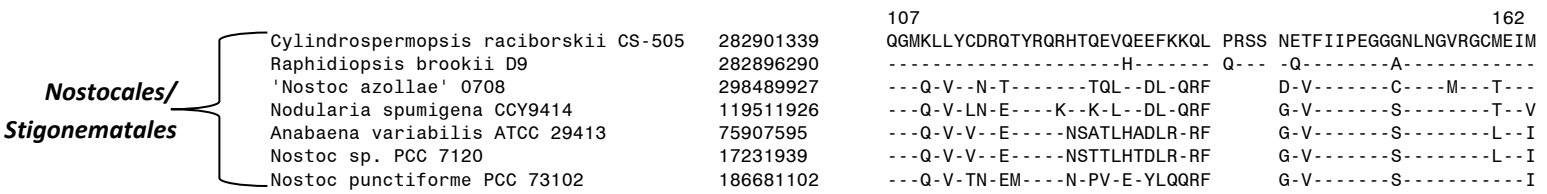
Supplementary Figure 51

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



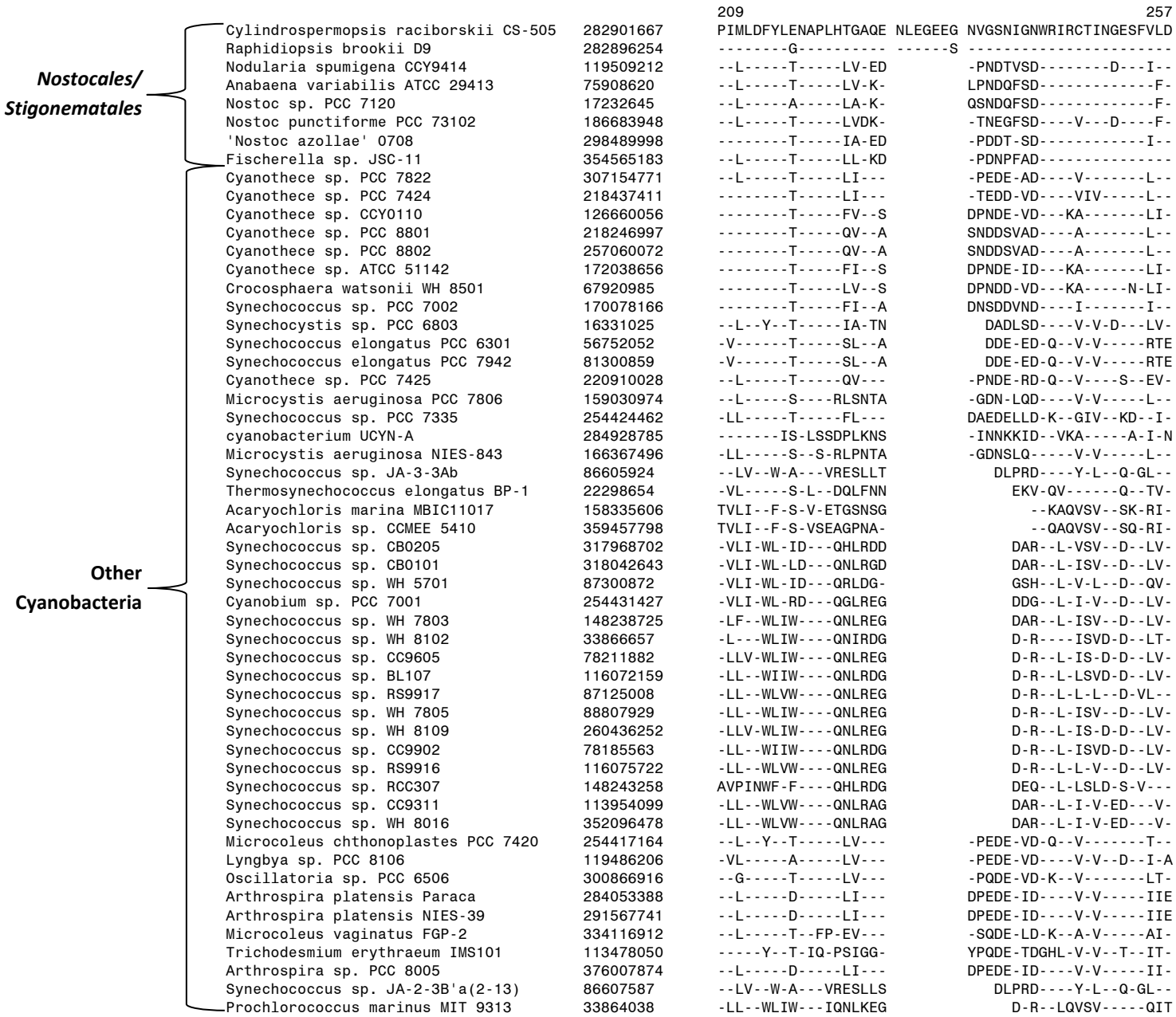
Supplementary Figure 52

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.*



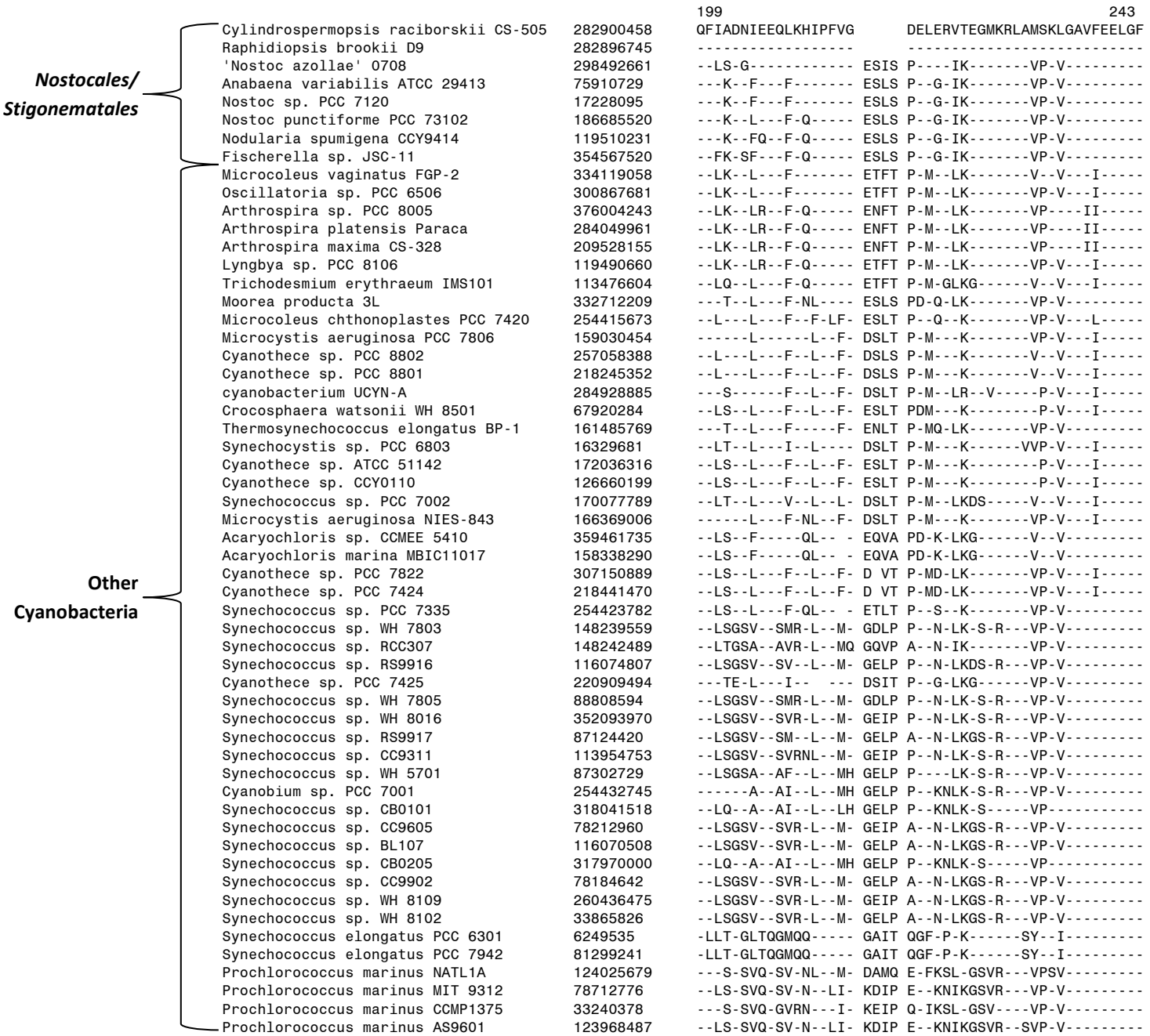
Supplementary Figure 53

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.*



Supplementary Figure 54

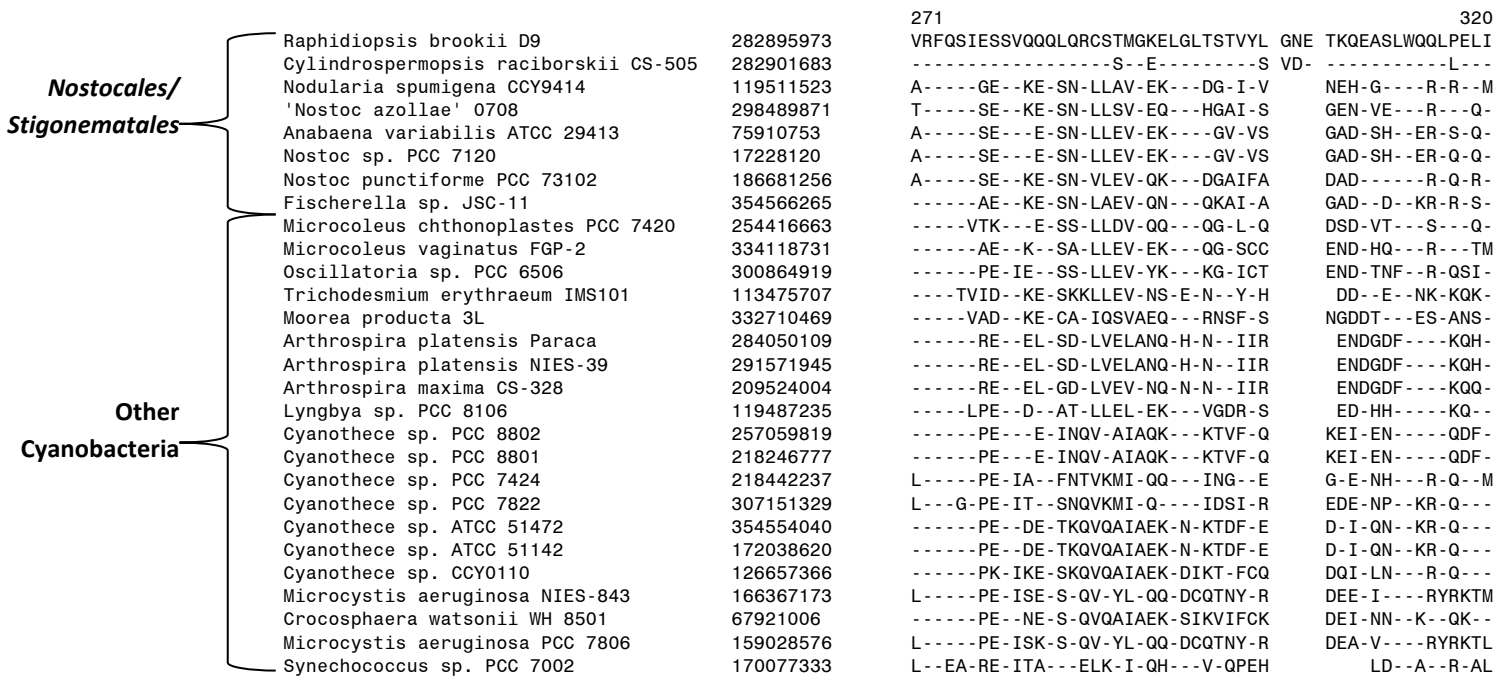
Partial sequence alignment for the hypothetical protein Aazo_0549, showing a 7 aa insert that is specific for *Cylindrospermopsis raciborskii CS-505 and Raphidiopsis brookii D9.*



		199		243
Cylindrospermopsis raciborskii CS-505	282900458	QFIADNIEEQLKHIPFVG		DELERVTEGMKRLAMSKLGAVFEELGF
Raphidiopsis brookii D9	282896745	----------		----------
'Nostoc azollae' 0708	298492661	--LS-G----------	ESIS	P----IK-------VP-V---------
Anabaena variabilis ATCC 29413	75910729	---K--F---F-------	ESLS	P--G-IK-------VP-V---------
Nostoc sp. PCC 7120	17228095	---K--F---F-------	ESLS	P--G-IK-------VP-V---------
Nostoc punctiforme PCC 73102	186685520	---K--L---F-Q-----	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-	DSLT	P-M---K-------VP-V---I-----
Cyanothece sp. PCC 8802	257058388	--L---L---F--L--F-	DSLS	P-M---K-------V--V---I-----
Cyanothece sp. PCC 8801	218245352	--L---L---F--L--F-	DSLS	P-M---K-------V--V---I-----
cyanobacterium UCYN-A	284928885	---S------F--L--F-	DSLT	P-M--LR--V-----P-V---I-----
Crocosphaera watsonii WH 8501	67920284	--LS--L---F--L--F-	ESLT	PDM---K--------P-V---I-----
Thermosynechococcus elongatus BP-1	161485769	---T--L---F-----F-	ENLT	P-MQ-LK-------VP-V---------
Synechocystis sp. PCC 6803	16329681	--LT--L---I--L----	DSLT	P-M---K------VVP-V---I-----
Cyanothece sp. ATCC 51142	172036316	--LS--L---F--L--F-	ESLT	P-M---K--------P-V---I-----
Cyanothece sp. CCY0110	126660199	--LS--L---F--L--F-	ESLT	P-M---K--------P-V---I-----
Synechococcus sp. PCC 7002	170077789	--LT--L---V--L--L-	DSLT	P-M--LKDS-----V--V---I-----
Microcystis aeruginosa NIES-843	166369006	------L---F-NL--F-	DSLT	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-----
Synechococcus sp. PCC 7335	254423782	--LS--L---F-QL-- -	ETLT	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--MQ	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----LK-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---------

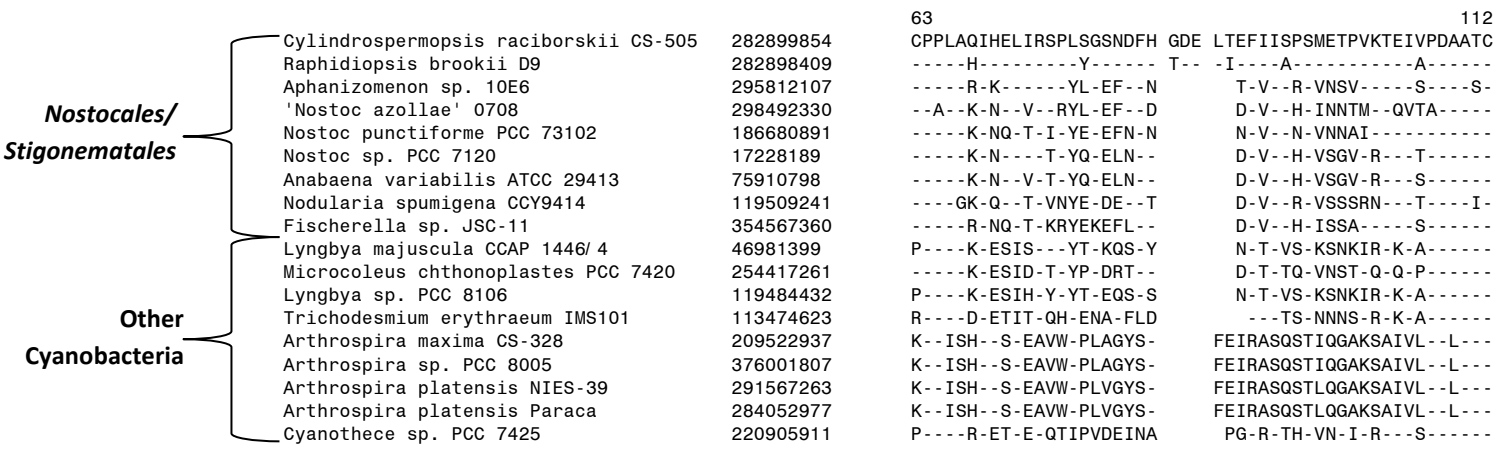
Supplementary Figure 55

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.*



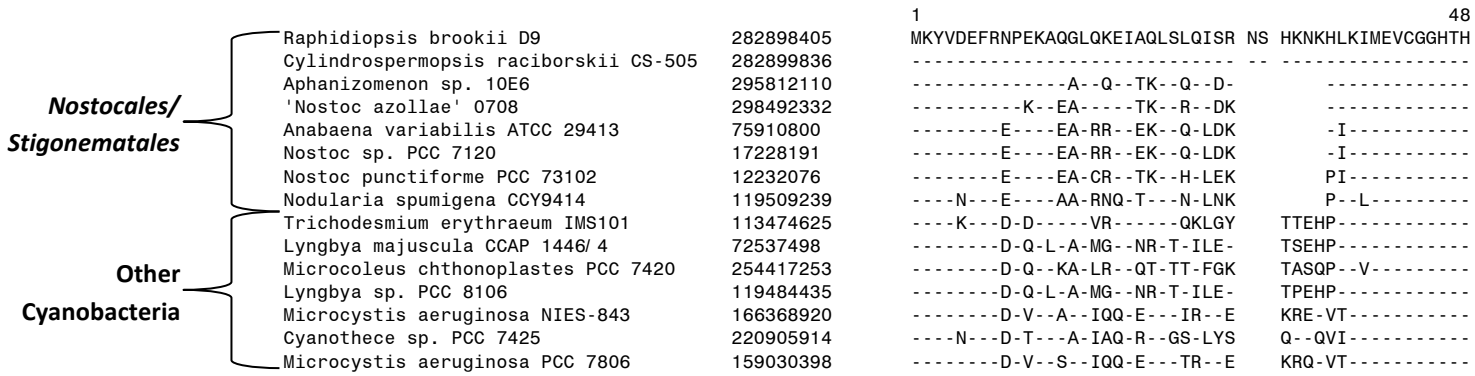
Supplementary Figure 56

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.*



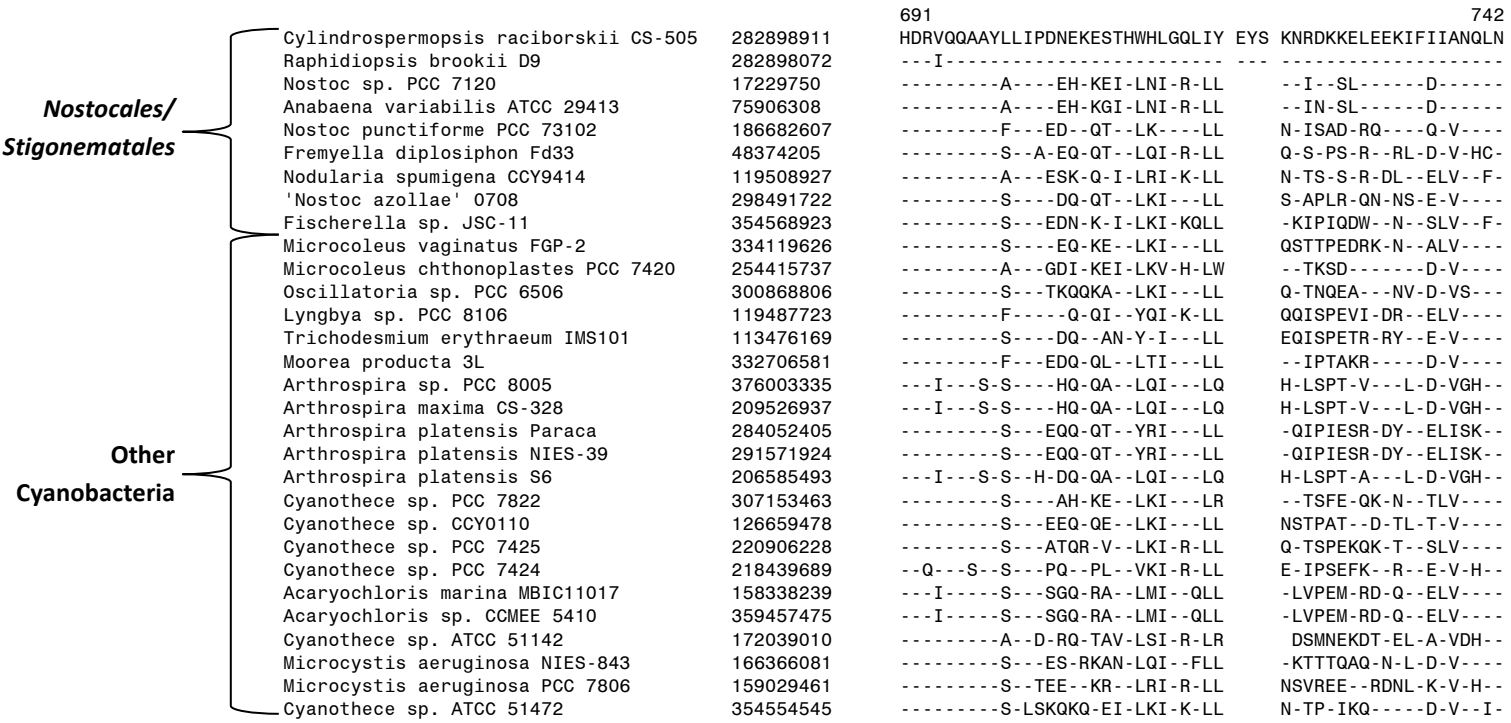
Supplementary Figure 57

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.*



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.*



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.*

Group	Species	ID	217 … 263
Nostocales/ Stigonematales	Raphidiopsis brookii D9	282895647	DFVDGILSGTDRQLYMEEFLLDGD CPFLGQTLQKAKLRSQTGALVLAIRR
	Cylindrospermopsis raciborskii CS-505	282899654	------------------------ -------------------------
	'Nostoc azollae' 0708	298491740	---------S-----------NPA N ---V--S------------------
	Nostoc punctiforme PCC 73102	186685052	-------T-A------------PA F ---V--S----R----S---------
	Anabaena variabilis ATCC 29413	75909935	-------T-A------------PA L ---V--S---------S---------
	Nostoc sp. PCC 7120	17228574	-------T-A------------PA L ---V--S---------S---------
	Nodularia spumigena CCY9414	119509893	-----FI------F--------PA V Y-YV-H--RQ-R----S----I----
	Fischerella sp. JSC-11	354564956	------I---------------PK V --VV-M--GQ-R--A-----I--V--
Other Cyanobacteria	Oscillatoria sp. PCC 6506	300867338	-------T----SF------V-P- S --YS----RD------S---------
	Microcoleus chthonoplastes PCC 7420	254409601	------IT-A--SF------I-PS S --LI-K--RE-H-------------
	Moorea producta 3L	332708976	-------T---SSF--------PE T --CV-K--RE-R-------------
	Microcoleus vaginatus FGP-2	334119086	--L--V-T---SSF------V-PN S -VYV----RN------S------V--
	Cyanothece sp. PCC 7822	307151194	-------T----AF--------P- V --CV--P-RD-R--L-----------
	Crocosphaera watsonii WH 0003	357265619	-------T-S--TF--------PK F --CV----SE-R--L-S---------
	Crocosphaera watsonii WH 8501	67921005	-------T-S--TF--------PK F --CV----SE-R--L-S---------
	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 SSQI----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.*

			57		115
Nostocales/ Stigonematales	Nostoc sp. PCC 7120	17227730	PIRFACHHYMGQVPILREIVTG	Q	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-------
Other Cyanobacteria	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----
	Synechococcus sp. PCC 7335	254421829	-VH-------S---VMSD--DA		------DGPKQ-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-KPLIEL		Y-AY-IKRGSG DL-A-RAAR-A-AHGWA--I-L--
	Cyanothece sp. PCC 8802	257059173	-VAYMAKEELFE--W---GIRL		Y-AY-VKRGSGDRGAIRAAL-A --EGWI--I-L--
	Cyanothece sp. PCC 8801	218246124	-VAYMAKEELFE--W---GIRL		Y-AY-VKRGSGDRGAIRAAL-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----

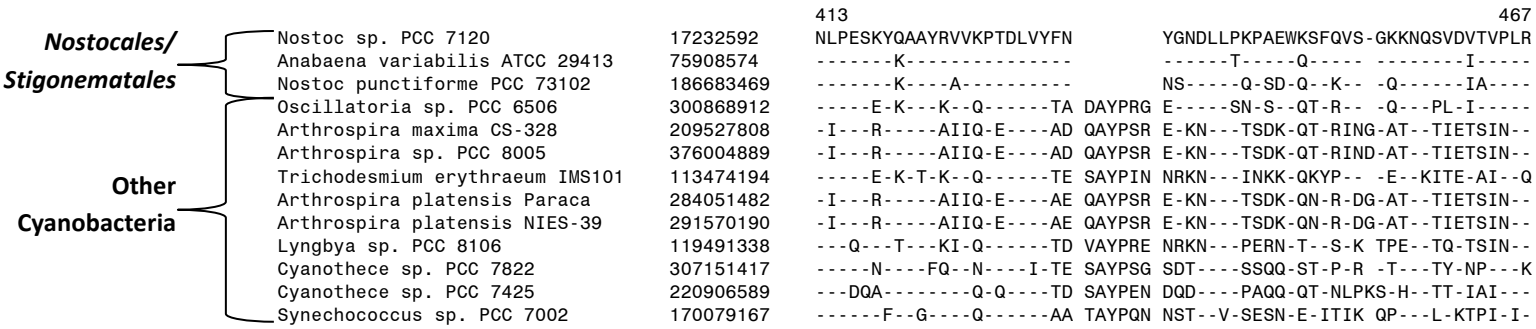
Supplementary Figure 61

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
Nostocales/        Nostoc sp. PCC 7120                    17227750   LIGHSAGGWISRIYLGEKAYAAR GKA TSLWNAHSLVATLVTLGTPH
Stigonematales     Anabaena variabilis ATCC 29413         75908954   ------------------P---- --- -------P------------
                   Nostoc punctiforme PCC 73102           186683066  --------------M---P-MVG SN- P---K--P-----I------
                   Nodularia spumigena CCY9414            119510315  ---------V------DQP-L-P -QV SCC-K------S-I------
Other              Thermosynechococcus elongatus BP-1     22299525   ----------------DQP-DGQ       V-AG-QW-Q--I------
Cyanobacteria      Synechococcus elongatus PCC 6301       56752264   -------------W--DRP-GPT     QQA-QGRDRT-V-IC-----
                   Synechococcus sp. JA-3-3Ab             86606815   -VA-------A-----SVP-GGQ       V-QGREH-S--IS--S--
                   Acaryochloris sp. CCMEE 5410           359459632  IV------------M-AEP-CDQ       I-Q-QPKIH---S-----
                   Acaryochloris marina MBIC11017         158339118  IV------------M-AEP-CDQ       V-Q-QPKIH---S-----
                   Synechococcus sp. JA-2-3B'a(2-13)      86607996   -VA-------A-----SVP-GGQ       V-QGREQ--A-IS--S--
                   Synechococcus elongatus PCC 7942       24251238   ------AADFTNL V-DRP-GPT     KQACQERDRTPV-IC-----
                   Synechococcus sp. RS9916               116074771  -----S--VML-L--SAD-FSG-       CYGGAQFCNR-I---S--
                   Synechococcus sp. WH 8109              260434414  -----S--VML-P--ADQTFLG-       RF-GAA-CNR-I---S--
                   Synechococcus sp. RCC307               148242538  -----S--VML-L--A-RGLEGA       SL-GRQ-CNR-IC--S--
                   Synechococcus sp. WH 8016              352096383  -----S--VML-L--SDDPFEG-       IYAGSKRCDR-----S--
                   Synechococcus sp. WH 7803              148239521  -V---S--VML-L--SDD-FQG-       RYGGSDRCNR-----S--
                   Synechococcus sp. RS9917               87124387   -----S--VML-L--SREPF-G-       VYAGAERCNR-I---S--
                   Cyanobium sp. PCC 7001                 254431182  -----S--IML-LF-DDAPFQG-       RYDGKA-ASA--M--S--
                   Synechococcus sp. WH 5701              87303536   -----S--IML-LL-DDAPFQG-       SYDGKA-AD---M--S--
                   Synechococcus sp. CC9311               113955246  -----S--VML-L--SDEPFDGS       IYAGSKKCDR-----S--
                   Synechococcus sp. WH 7805              88808632   -V---S--VML-L--SDVEFQGC       RYGGSDRCNR-----S--
                   Prochlorococcus marinus MIT 9303       124023266  -----S--VML-L---DQPFSG-       IYSGYKSCNR-I---S--
                   Gloeobacter violaceus PCC 7421         37520555   -V--------A-LL--D-E-GR-       T--YRDR--S-I------
```

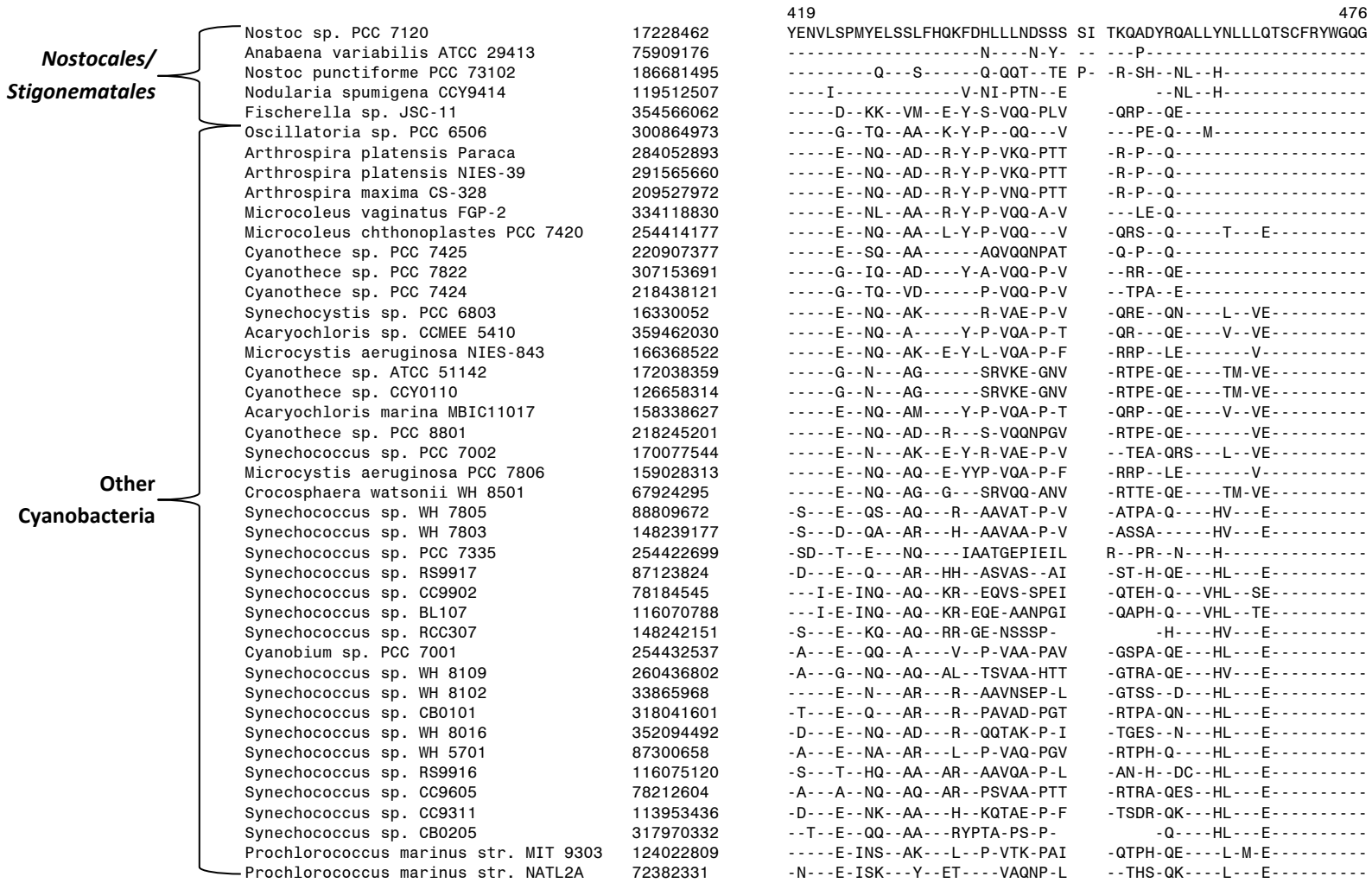
Supplementary Figure 62

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*.



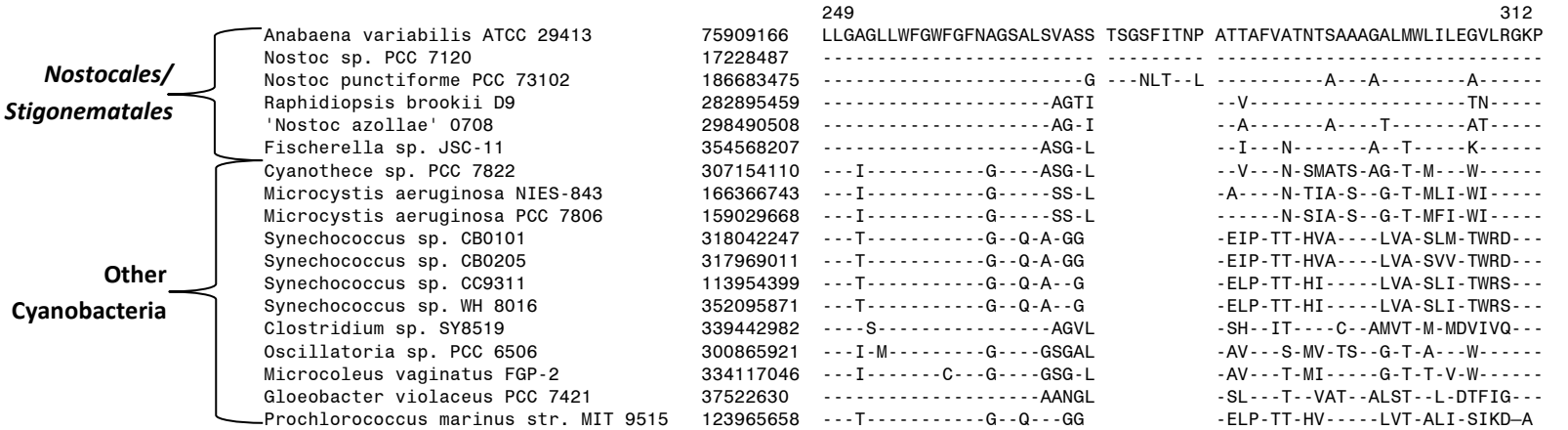
Supplementary Figure 63

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*.



Supplementary Figure 64

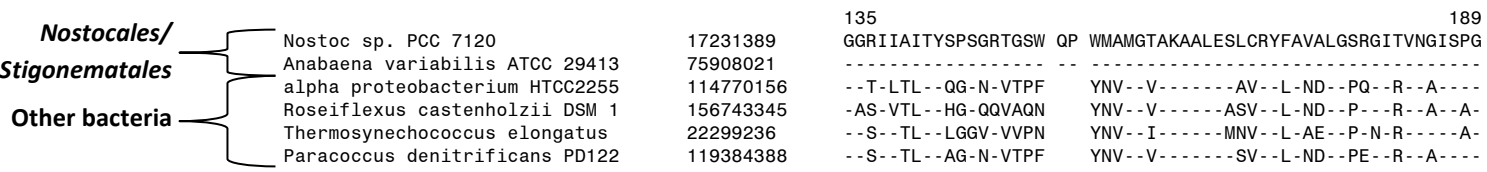
Partial sequence alignment for the hypothetical protein Npun_R1017, showing a 2 aa insert that is specific for all members of a *Nostocales* subclade which includes *Nostoc sp. PCC 7120*, *Anabaena variabilis ATCC 29413* and *Nostoc punctiforme PCC 73102*.



		249		312
Anabaena variabilis ATCC 29413	75909166	LLGAGLLWFGWFGFNAGSALSVASS	TSGSFITNP	ATTAFVATNTSAAAGALMWLILEGVLRGKP
Nostoc sp. PCC 7120	17228487	-------------------------	---------	------------------------------
Nostoc punctiforme PCC 73102	186683475	------------------------G	---NLT--L	----------A---A--------A------
Raphidiopsis brookii D9	282895459	----------------------AGTI		--V--------------------TN-----
'Nostoc azollae' 0708	298490508	----------------------AG-I		--A-------A----T-------AT-----
Fischerella sp. JSC-11	354568207	----------------------ASG-L		--I---N-------A--T-----K------
Cyanothece sp. PCC 7822	307154110	---I-----------G----ASG-L		--V---N-SMATS-AG-T-M---W------
Microcystis aeruginosa NIES-843	166366743	---I-----------G-----SS-L		-A----N-TIA-S--G-T-MLI-WI-----
Microcystis aeruginosa PCC 7806	159029668	---I-----------G-----SS-L		------N-SIA-S--G-T-MFI-WI-----
Synechococcus sp. CB0101	318042247	---T-----------G--Q-A-GG		-EIP-TT-HVA----LVA-SLM-TWRD---
Synechococcus sp. CB0205	317969011	---T-----------G--Q-A-GG		-EIP-TT-HVA----LVA-SVV-TWRD---
Synechococcus sp. CC9311	113954399	---T-----------G--Q-A--G		-ELP-TT-HI-----LVA-SLI-TWRS---
Synechococcus sp. WH 8016	352095871	---T-----------G--Q-A--G		-ELP-TT-HI-----LVA-SLI-TWRS---
Clostridium sp. SY8519	339442982	----S----------------AGVL		-SH--IT----C--AMVT-M-MDVIVQ---
Oscillatoria sp. PCC 6506	300865921	---I-M---------G----GSGAL		-AV---S-MV-TS--G-T-A---W------
Microcoleus vaginatus FGP-2	334117046	---I-------C---G----GSG-L		-AV---T-MI-----G-T-T-V-W------
Gloeobacter violaceus PCC 7421	37522630	----------------------AANGL		-SL---T--VAT--ALST--L-DTFIG---
Prochlorococcus marinus str. MIT 9515	123965658	---T-----------G--Q---GG		-ELP-TT-HV-----LVT-ALI-SIKD—A

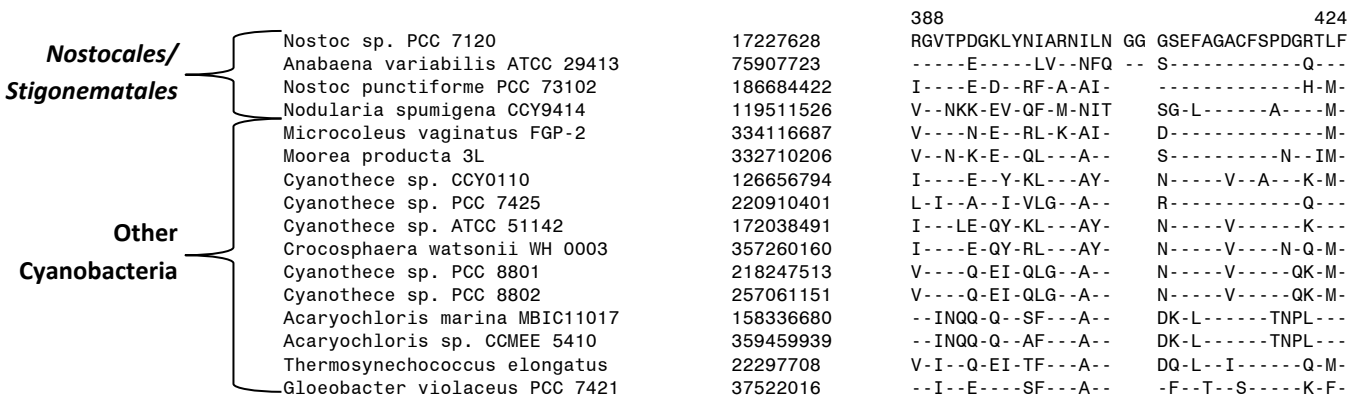
Supplementary Figure 65

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*.



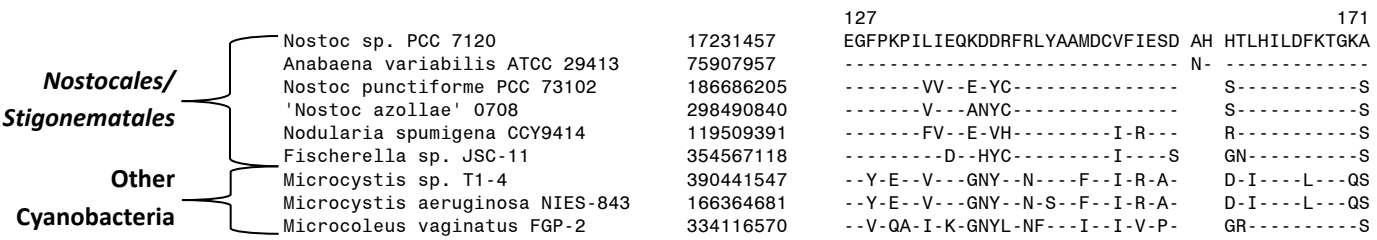
Supplementary Figure 66

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.*



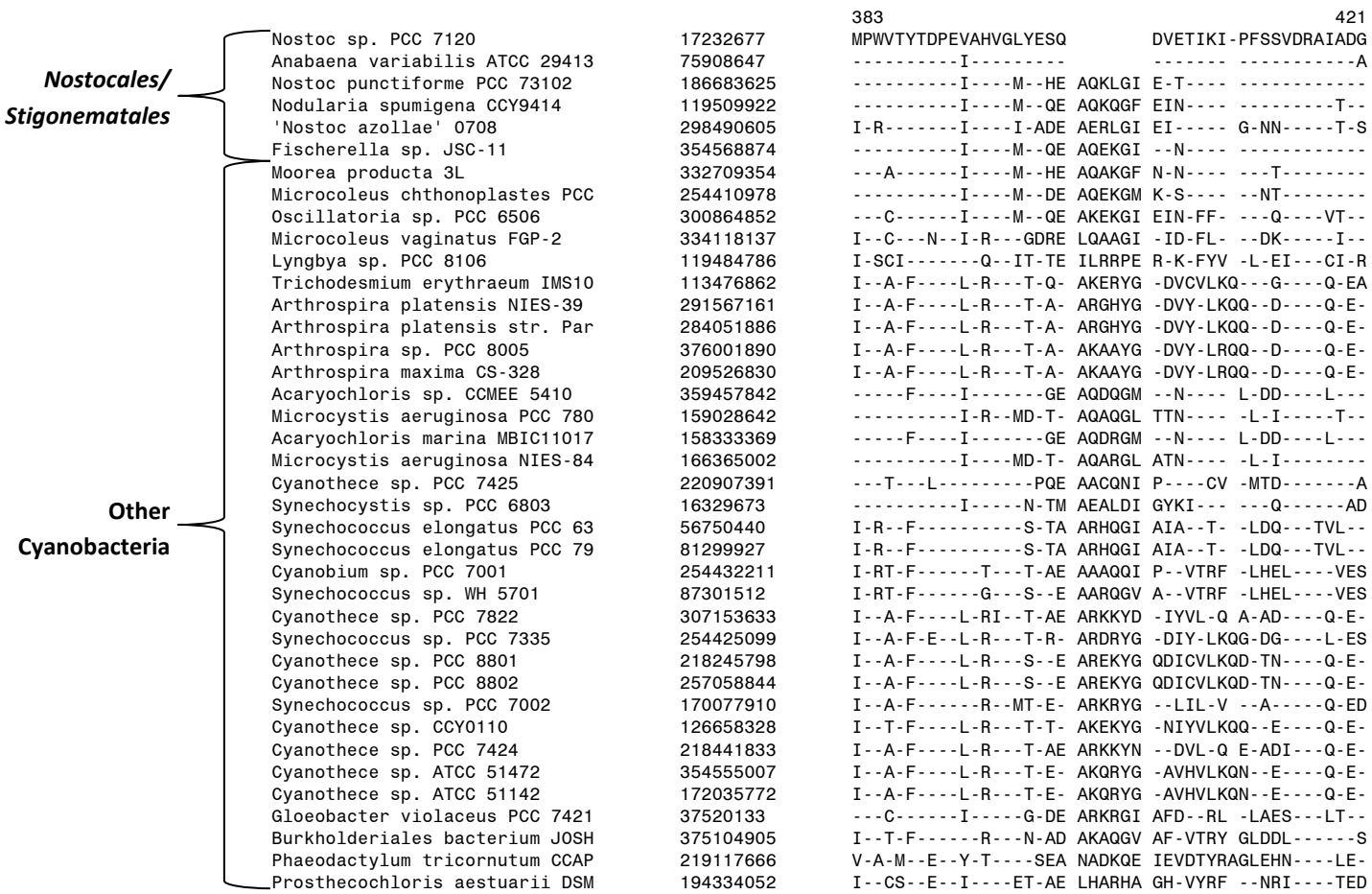
Supplementary Figure 67

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.*



Supplementary Figure 68

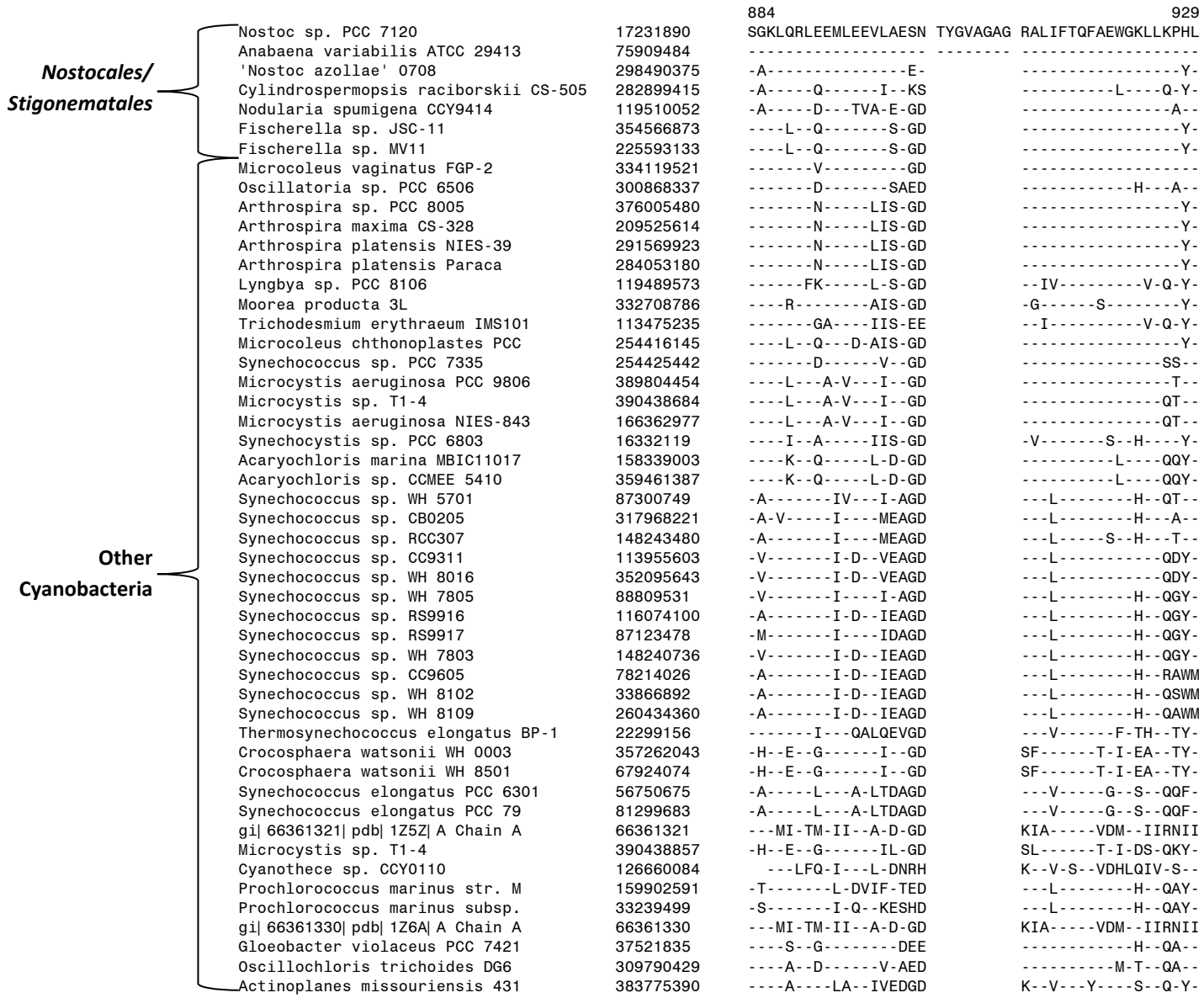
Partial sequence alignment for the hypothetical protein Npun_R6172, showing a 2 aa insert that is specific for *Nostoc sp. PCC 7120 and Anabaena variabilis ATCC 29413.*



		383	421
Nostoc sp. PCC 7120	17232677	MPWVTYTDPEVAHVGLYESQ	DVETIKI-PFSSVDRAIADG
Anabaena variabilis ATCC 29413	75908647	----------I---------	------- -----------A
Nostoc punctiforme PCC 73102	186683625	----------I----M--HE AQKLGI	E-T---- ----------
Nodularia spumigena CCY9414	119509922	----------I----M--QE AQKQGF	EIN---- ---------T--
'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--
Lyngbya 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	-DVY-LKQQ--D----Q-E-
Arthrospira platensis str. Par	284051886	I--A-F----L-R---T-A- ARGHYG	-DVY-LKQQ--D----Q-E-
Arthrospira sp. PCC 8005	376001890	I--A-F----L-R---T-A- AKAAYG	-DVY-LRQQ--D----Q-E-
Arthrospira maxima CS-328	209526830	I--A-F----L-R---T-A- AKAAYG	-DVY-LRQQ--D----Q-E-
Acaryochloris sp. CCMEE 5410	359457842	-----F----I-------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-------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 AAAQQI	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	-NIYVLKQQ--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 AKAQGV	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

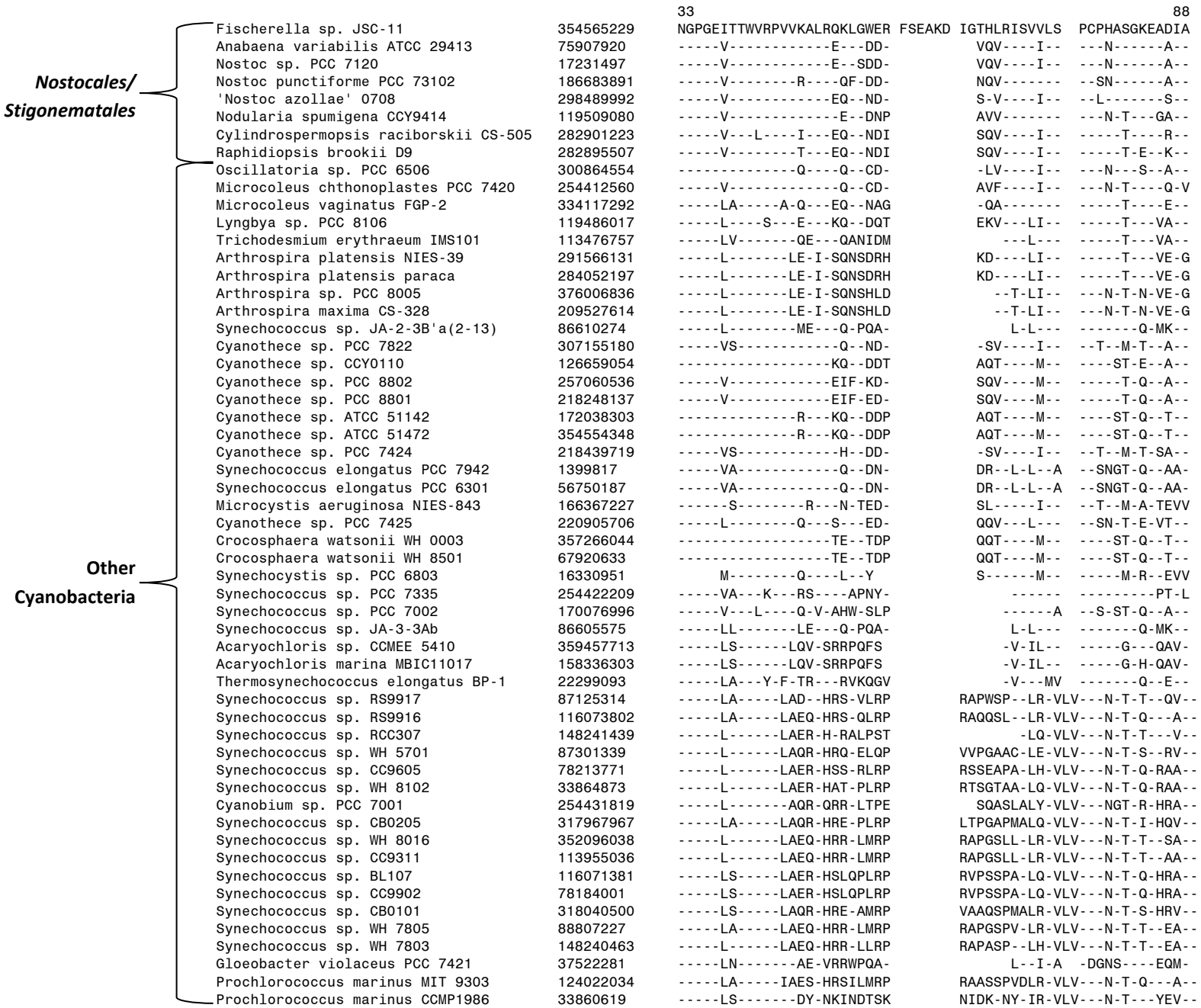
Partial sequence alignment for the protein mercuric reductase, showing a 6 aa deletion that is specific for *Nostoc sp. PCC 7120 and Anabaena variabilis ATCC 29413.*



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

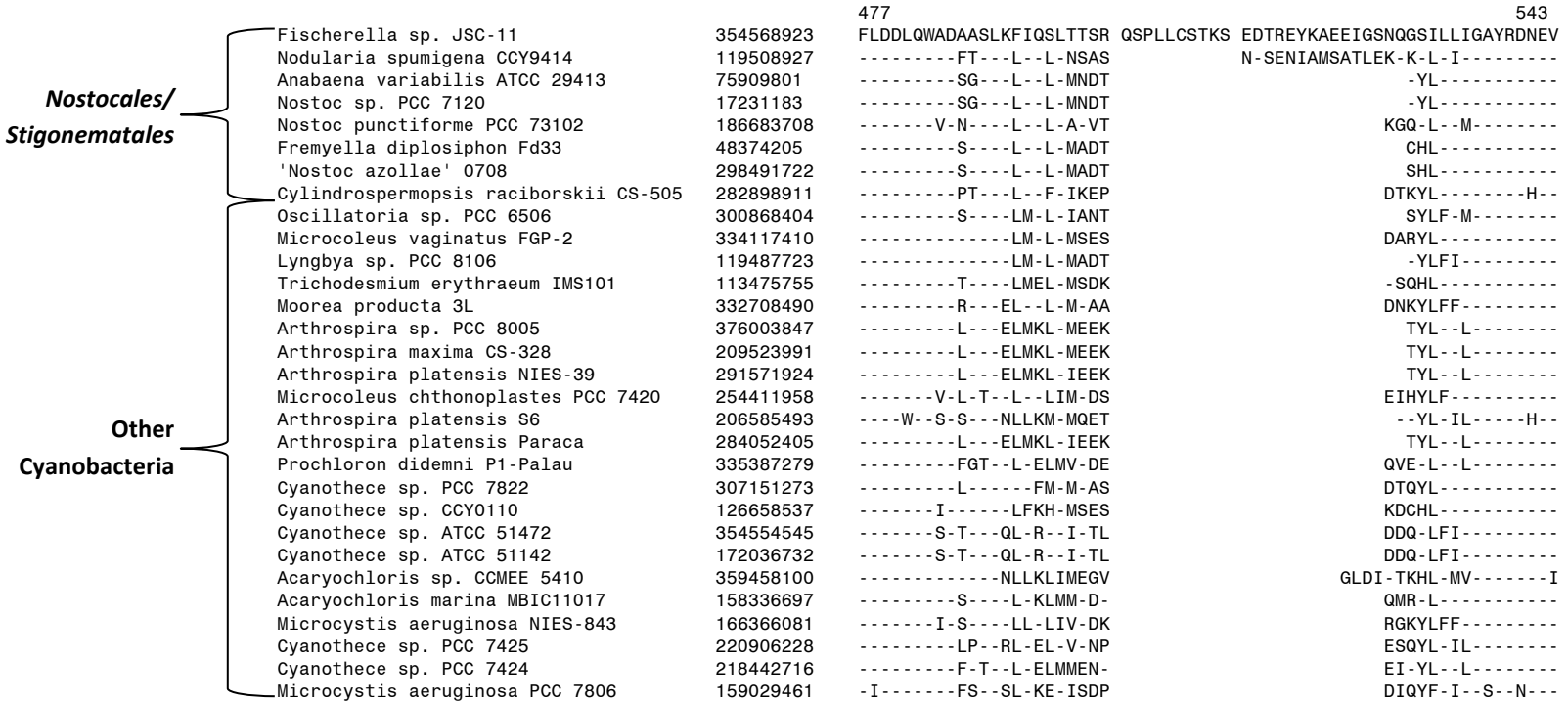
Supplementary Figure 70

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.*



Supplementary Figure 71

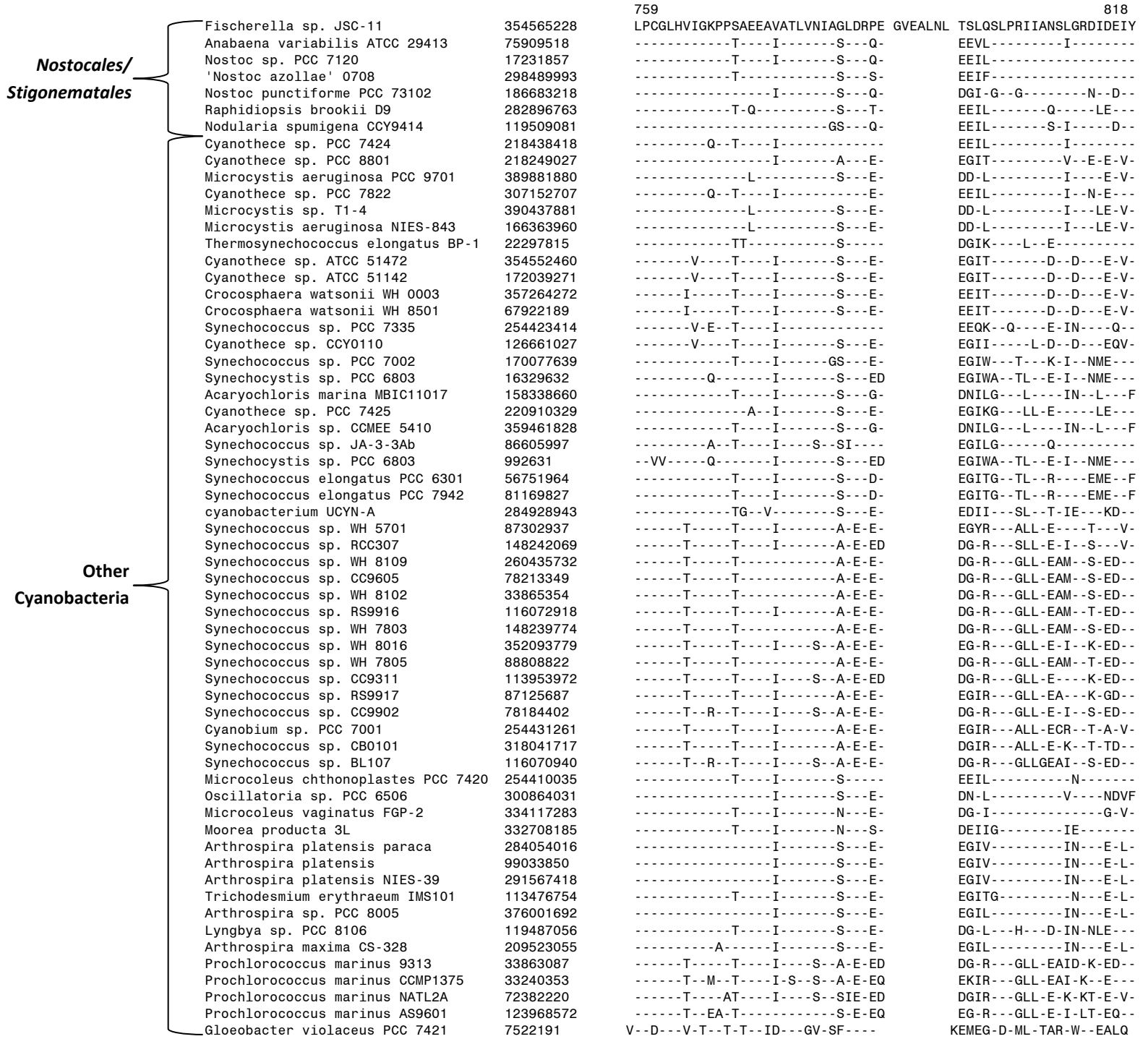
Partial sequence alignment for the protein lipid-A-disaccharide synthase, showing a 6 aa insert that is specific for *Fischerella sp. JSC-11.*



		477	543
Fischerella sp. JSC-11	354568923	FLDDLQWADAASLKFIQSLTTSR	QSPLLCSTKS EDTREYKAEEIGSNQGSILLIGAYRDNEV
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----LM-L-IANT	SYLF-M--------
Microcoleus vaginatus FGP-2	334117410	--------------LM-L-MSES	DARYL----------
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	EIHYLF----------
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---

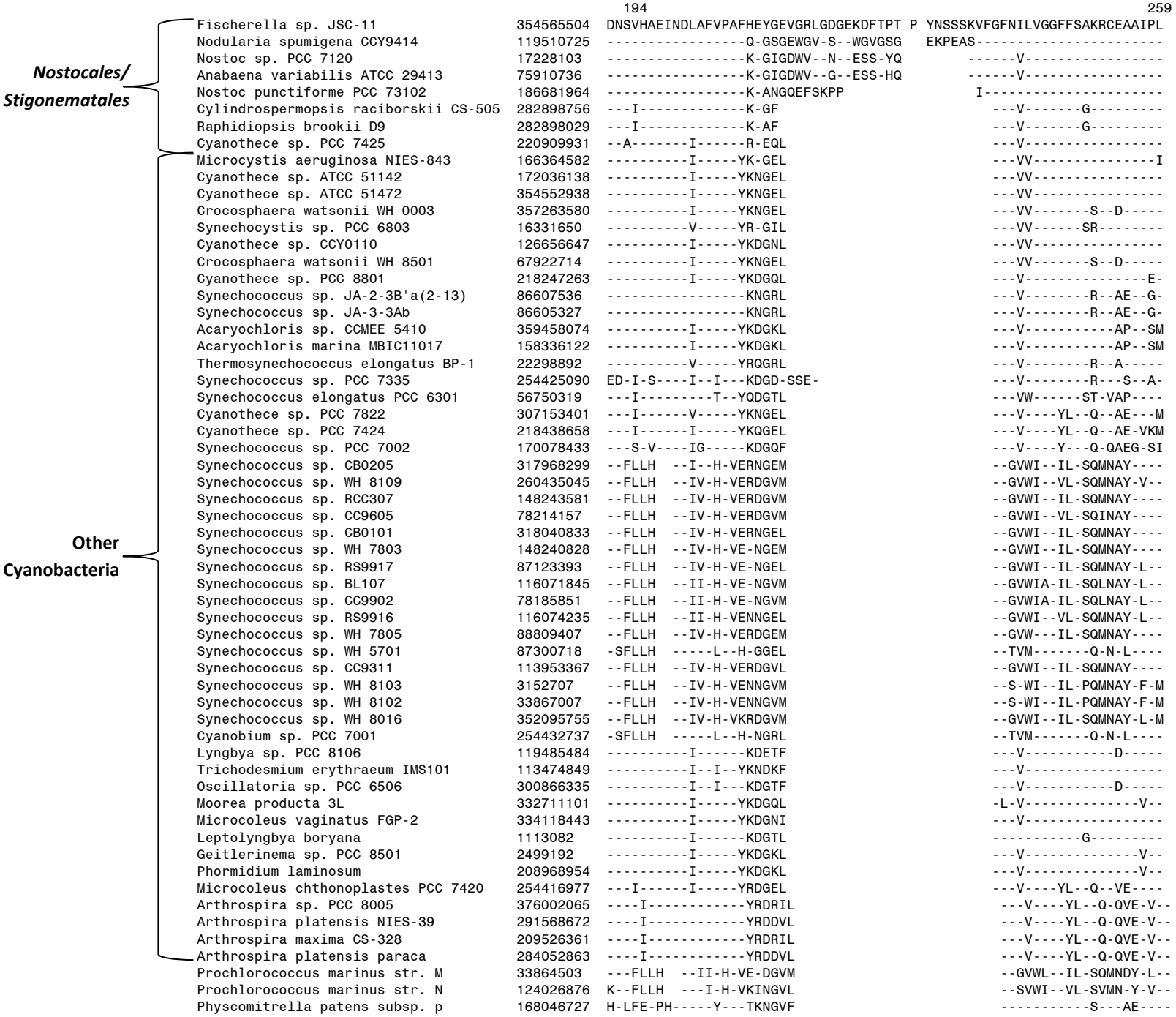
Supplementary Figure 72

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.*



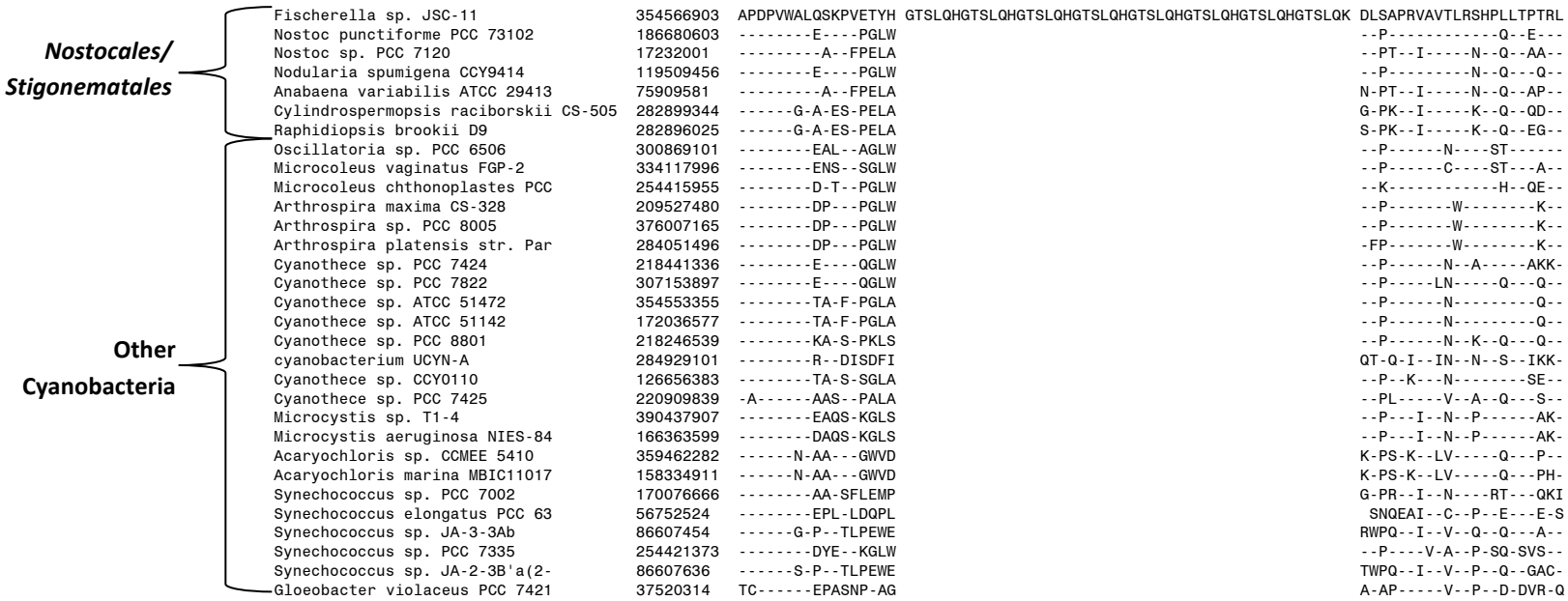
Supplementary Figure 73

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



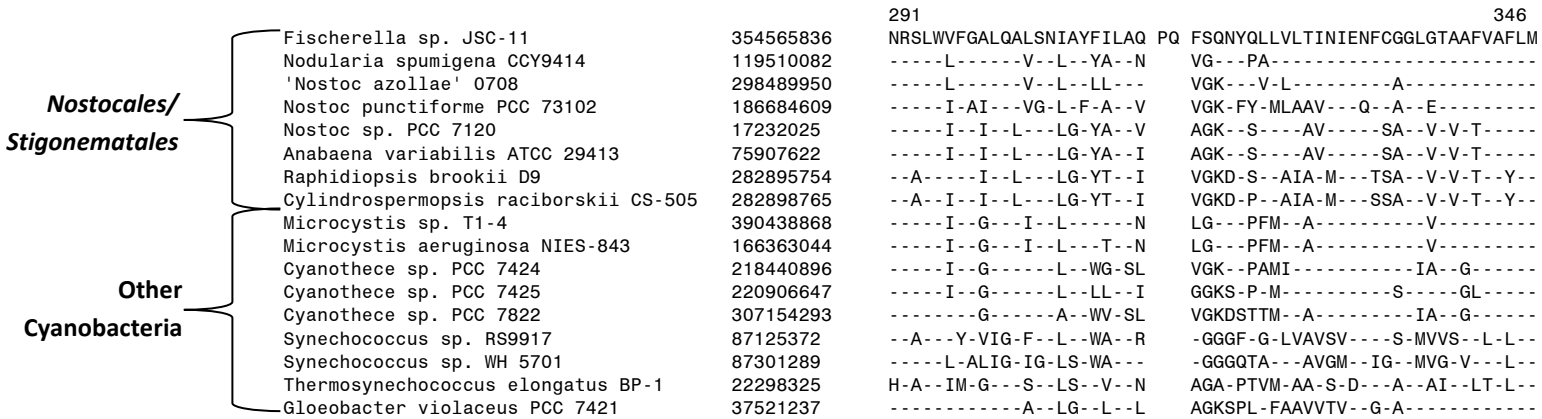
Supplementary Figure 74

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



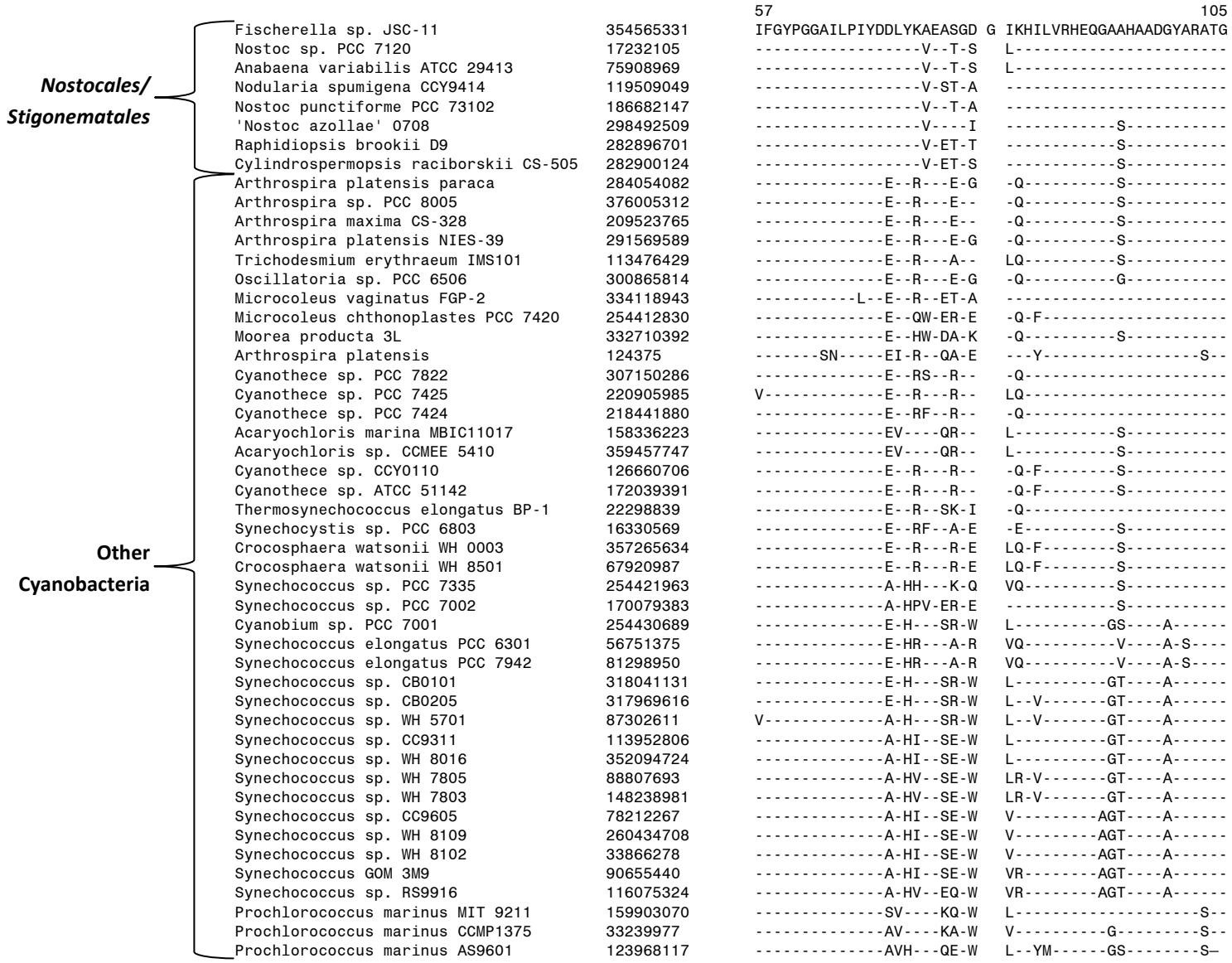
Supplementary Figure 75

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



Supplementary Figure 76

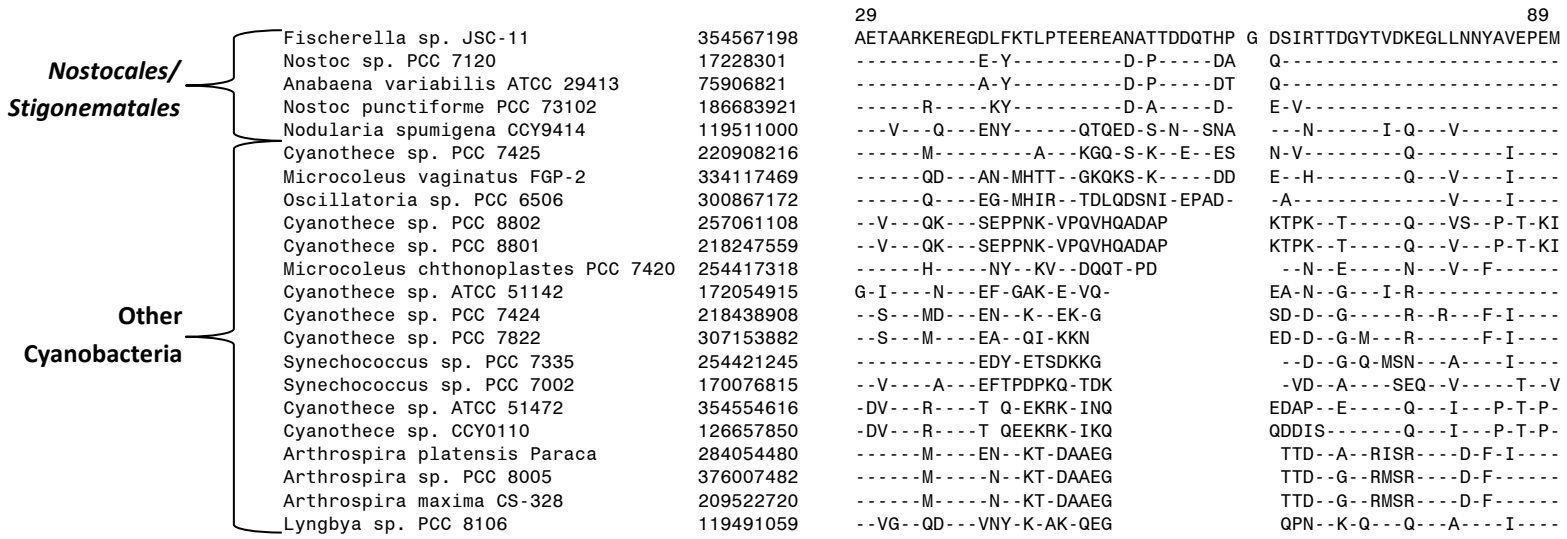
Partial sequence alignment for AmpG-related permease, showing a 2 aa insert that is specific for *Fischerella sp. JSC-11.*



		57	105
Fischerella sp. JSC-11	354565331	IFGYPGGAILPIYDDLYKAEASGD G	IKHILVRHEQGAAHAADGYARATG
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-----EI-R--QA-E	---Y----------S--
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----------
Crocosphaera watsonii WH 0003	357265634	----------E--R---R-E	LQ-F--------S----------
Crocosphaera 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	----------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–

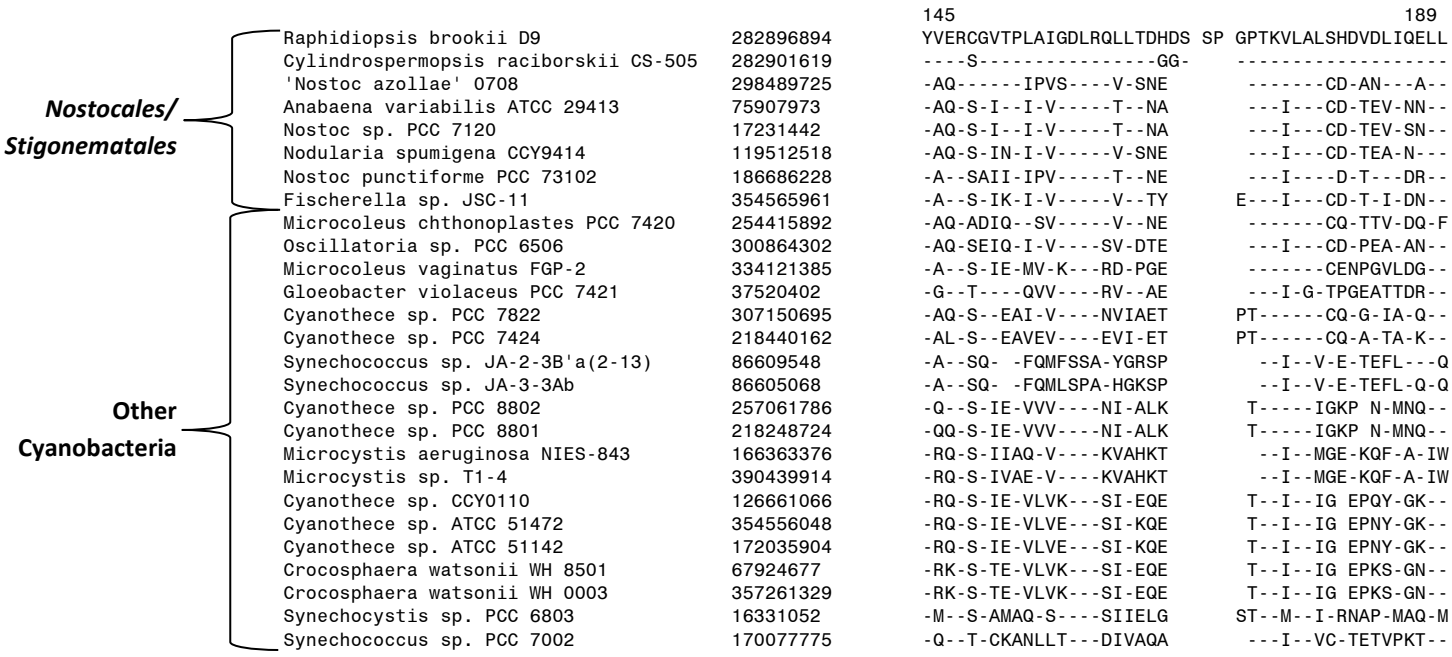
Supplementary Figure 77

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



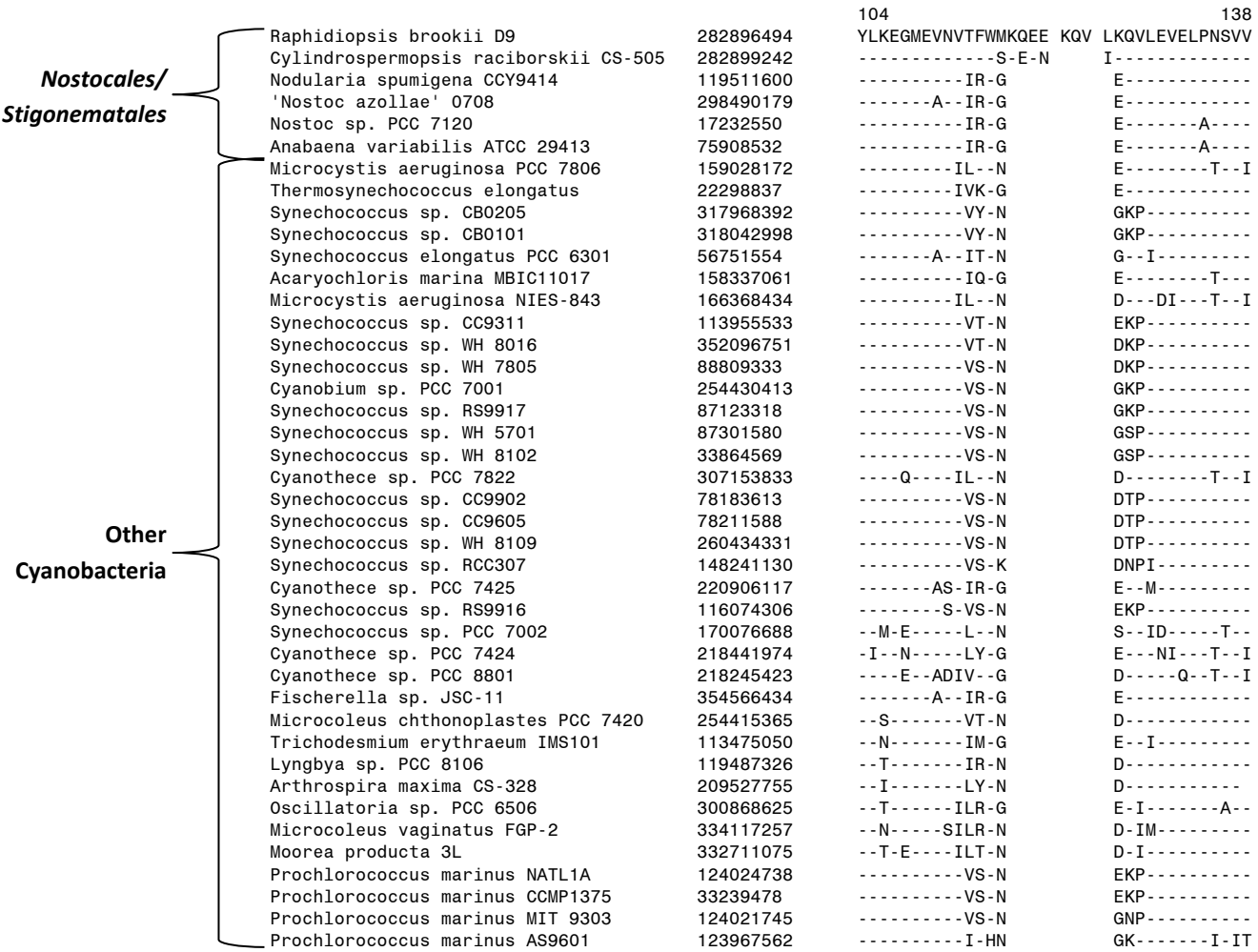
Supplementary Figure 78

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



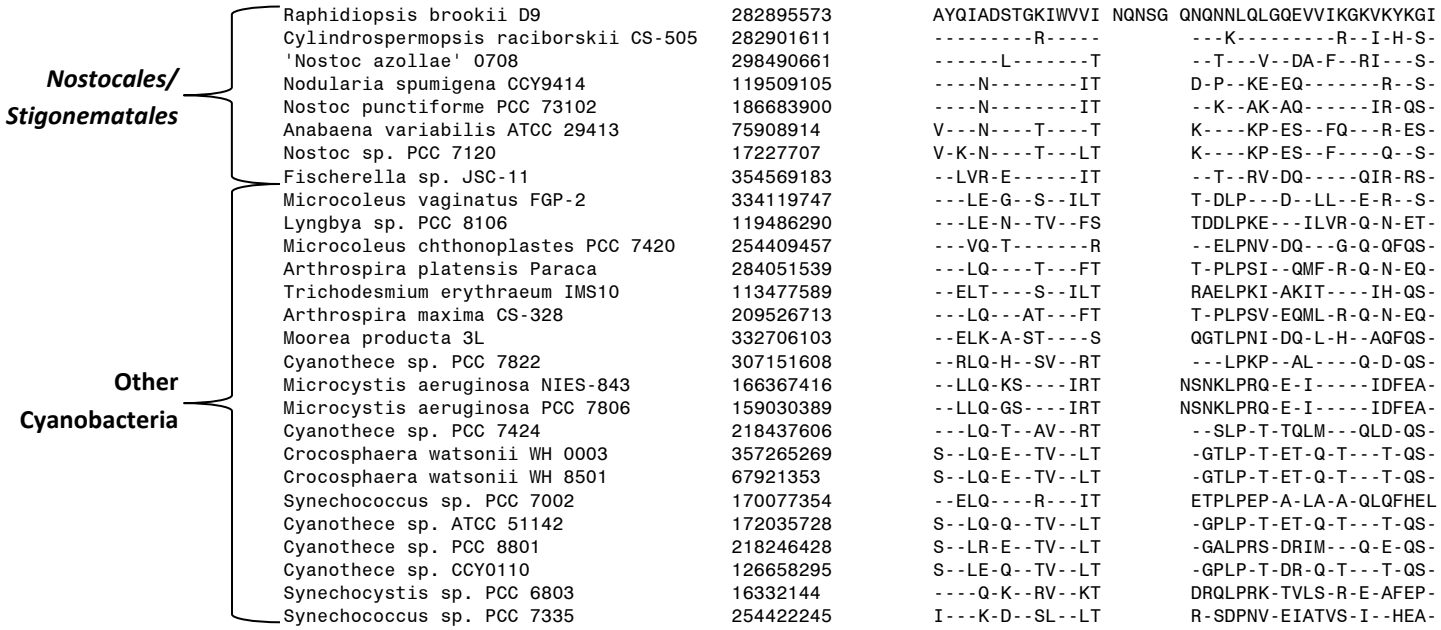
Supplementary Figure 79

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



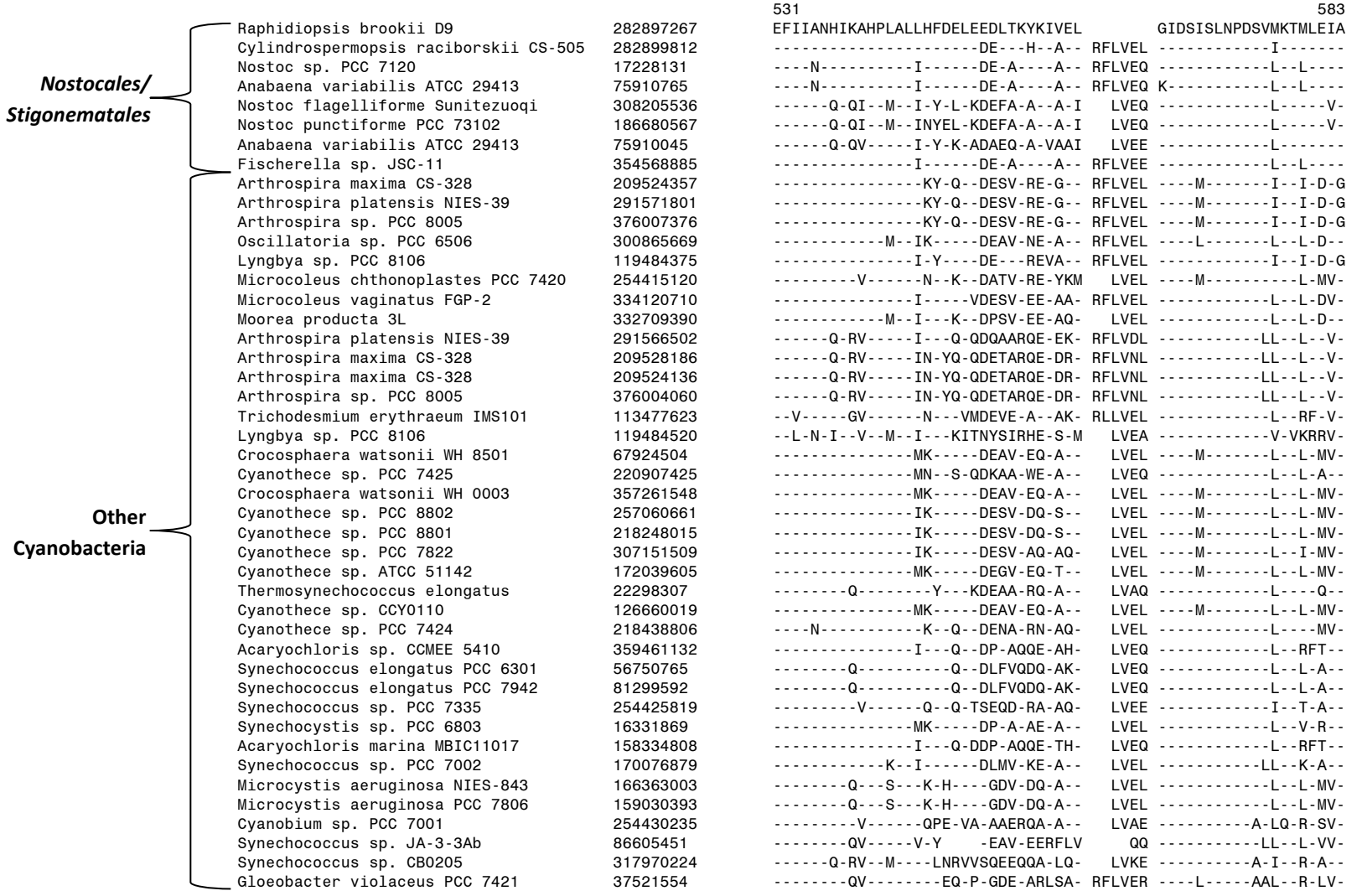
Supplementary Figure 80

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



Supplementary Figure 81

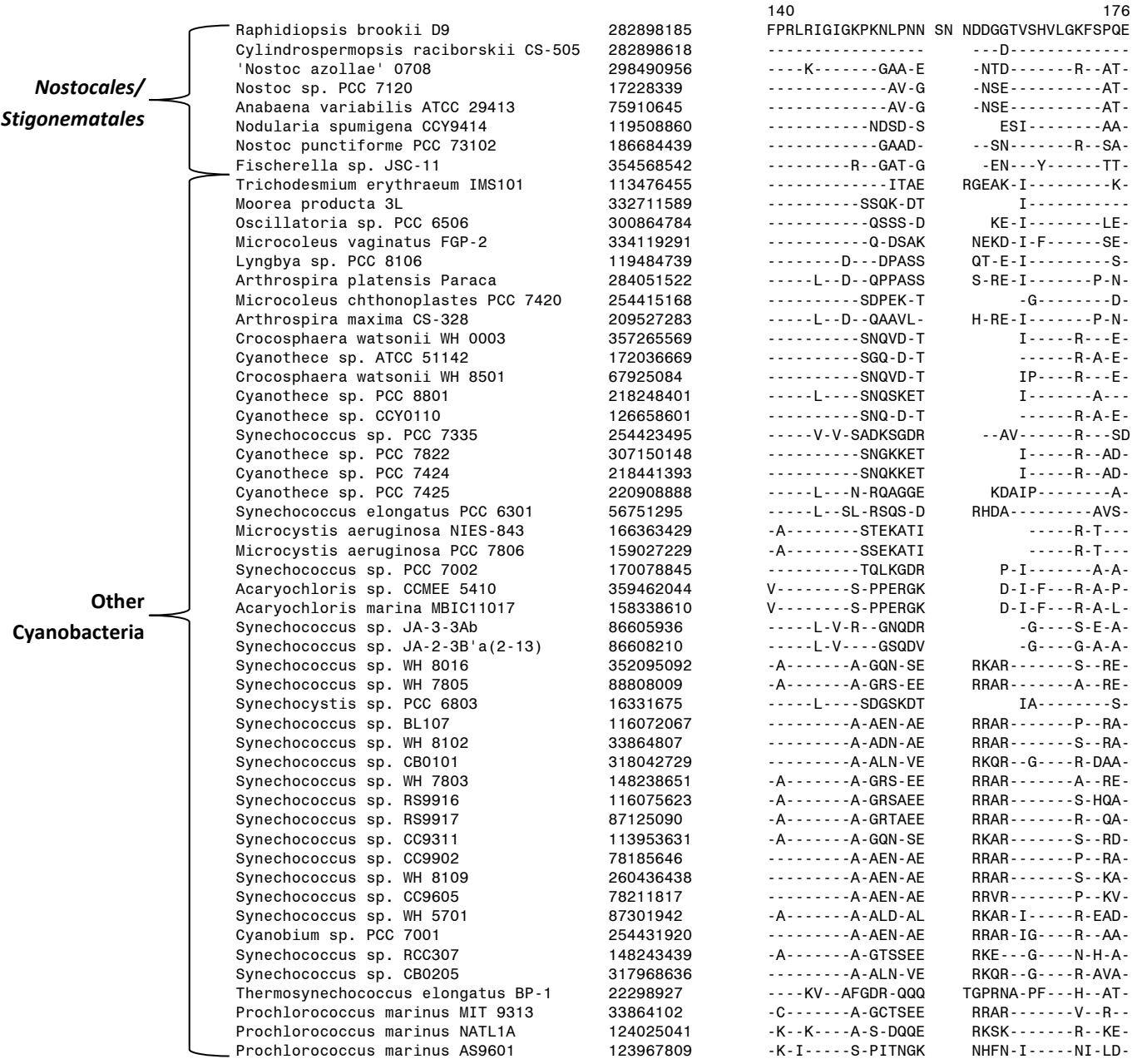
Partial sequence alignment for the hypothetical protein Aazo_1507, showing a 5 aa insert that is specific for *Raphidiopsis brookii D9.*



		531	583
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--INYEL-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	33120710	----------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-QDQAARQE-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-
Crocosphaera 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--
Crocosphaera 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. CCMEE 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

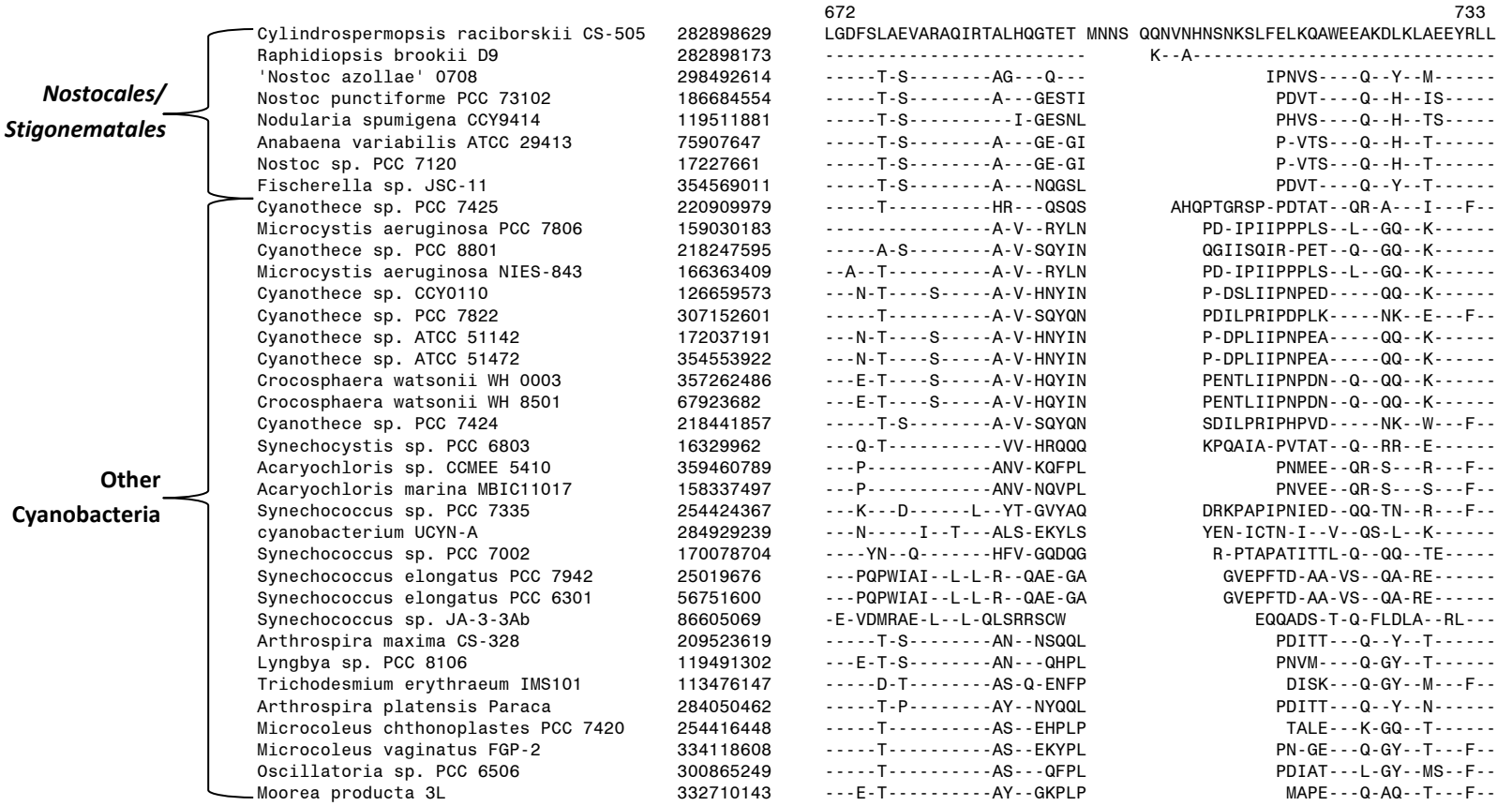
Partial sequence alignment for the protein Phosphoenolpyruvate synthase, showing a 5 aa deletion that is specific for *Raphidiopsis brookii D9.*



		140	176
Raphidiopsis brookii D9	282898185	FPRLRIGIGKPKNLPNN SN	NDDGGTVSHVLGKFSPQE
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	----------NDSD-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-
Crocosphaera watsonii WH 0003	357265569	----------SNQVD-T	I-----R---E-
Cyanothece sp. ATCC 51142	172036669	----------SGQ-D-T	------R-A-E-
Crocosphaera 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 MBIC11017	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-

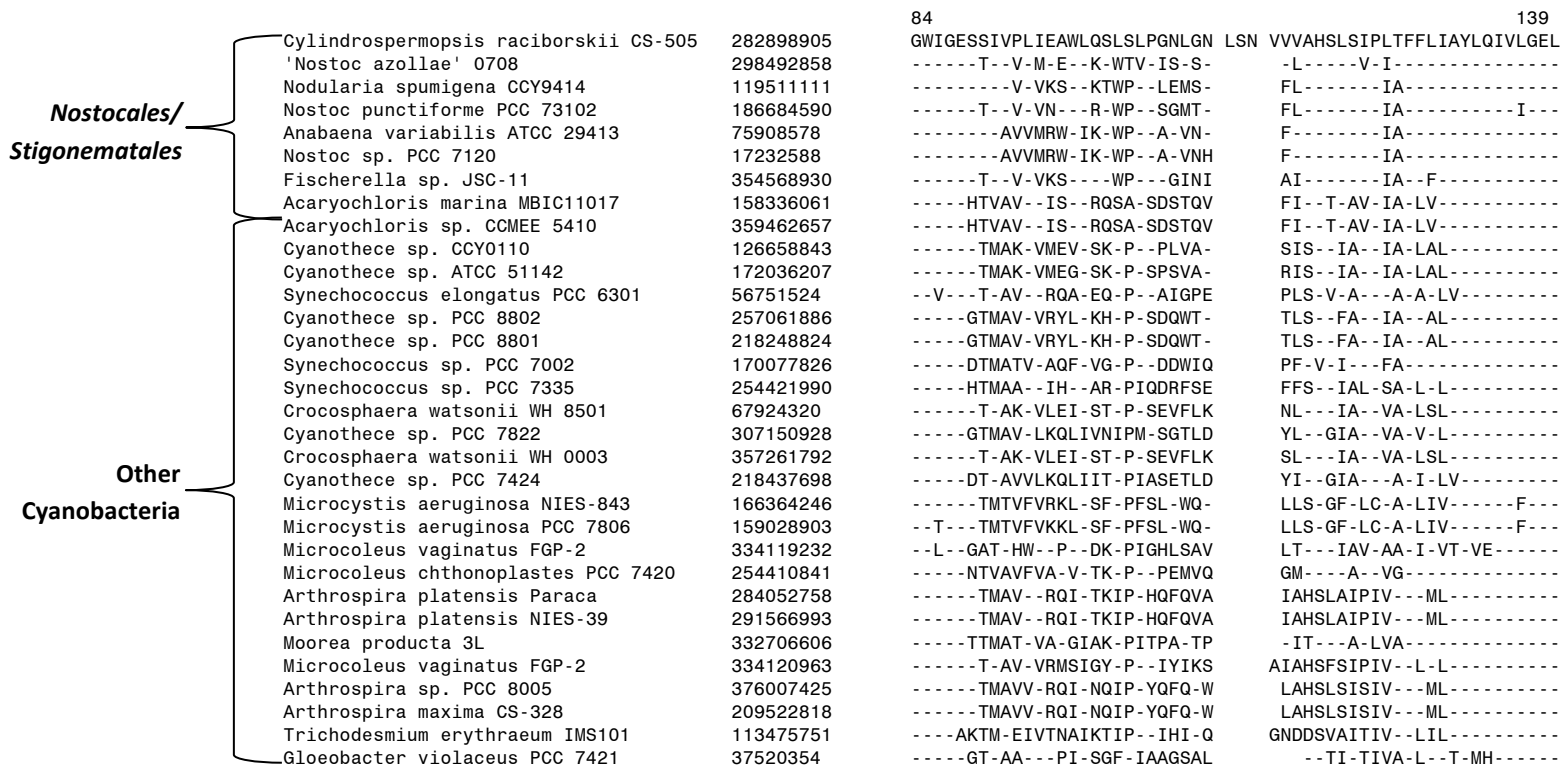
Supplementary Figure 83

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



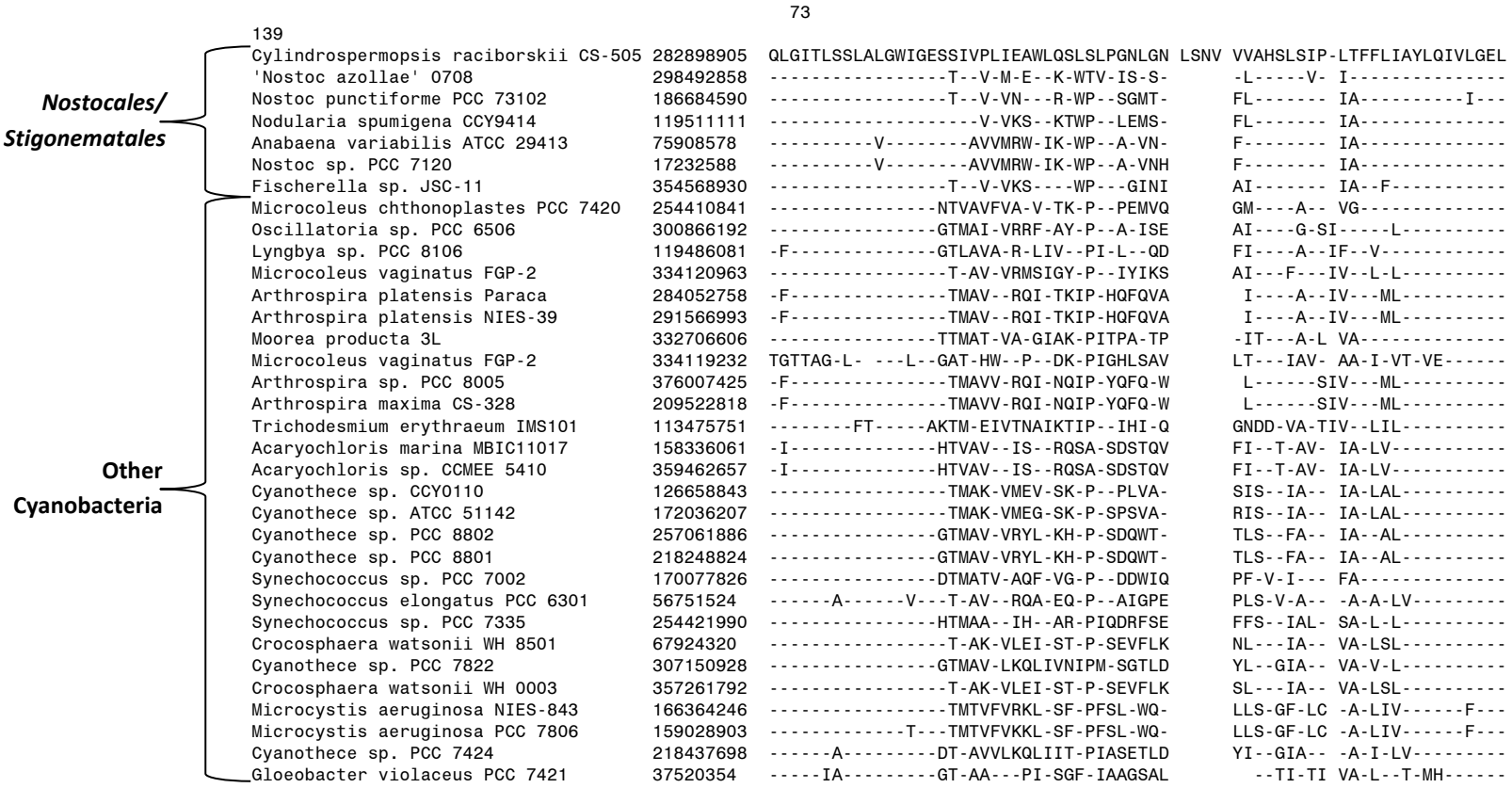
Supplementary Figure 84

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



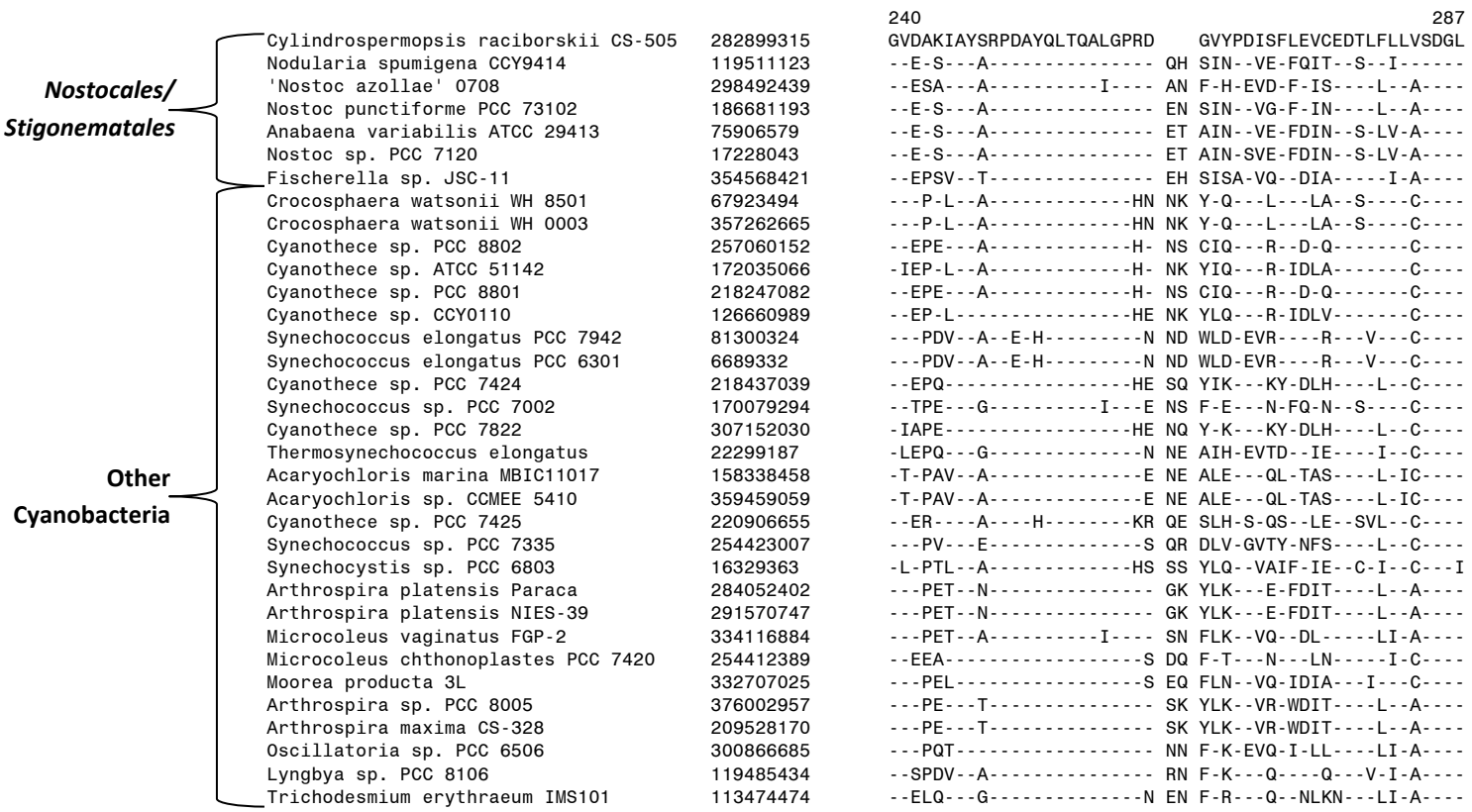
Supplementary Figure 85

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



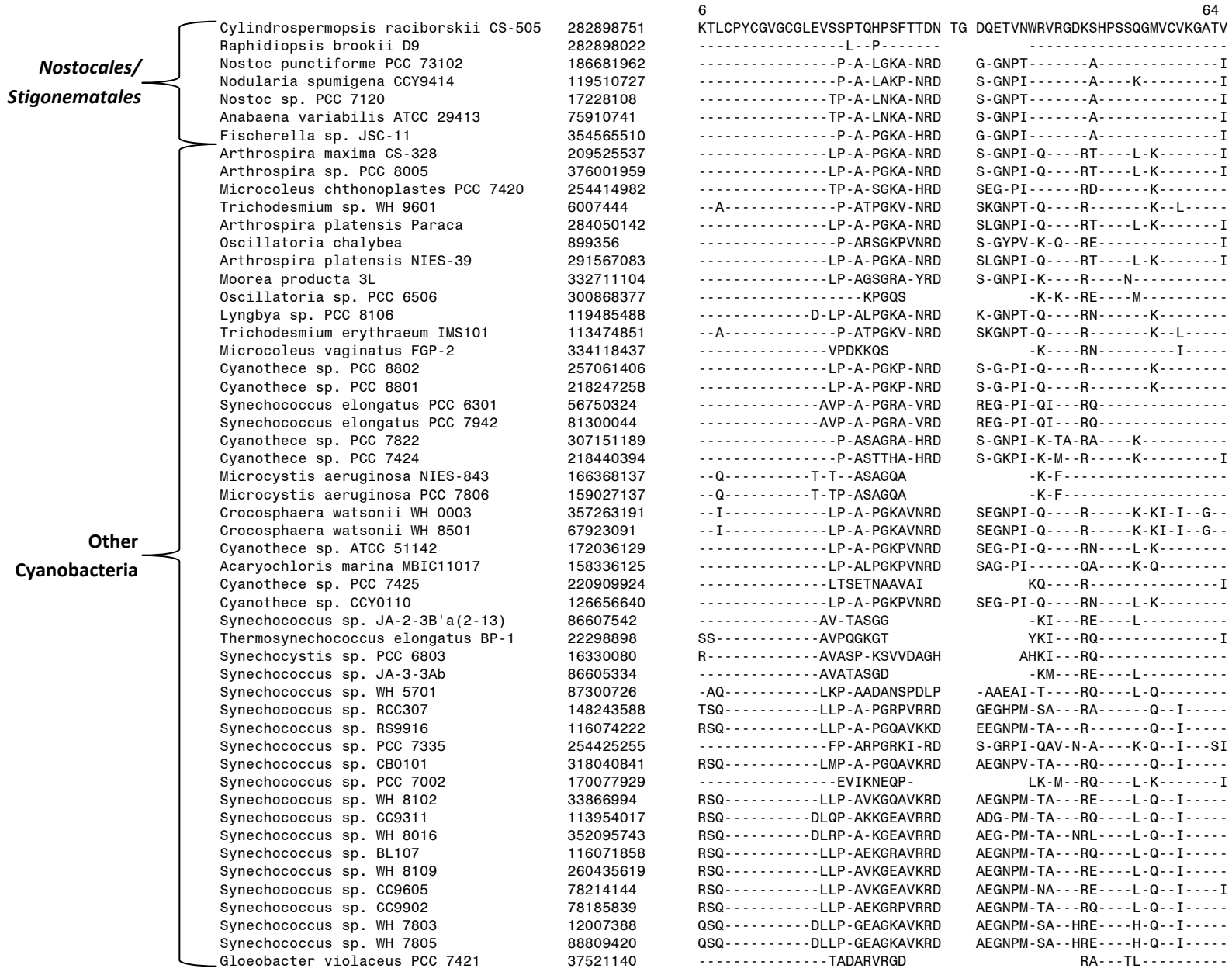
Supplementary Figure 86

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



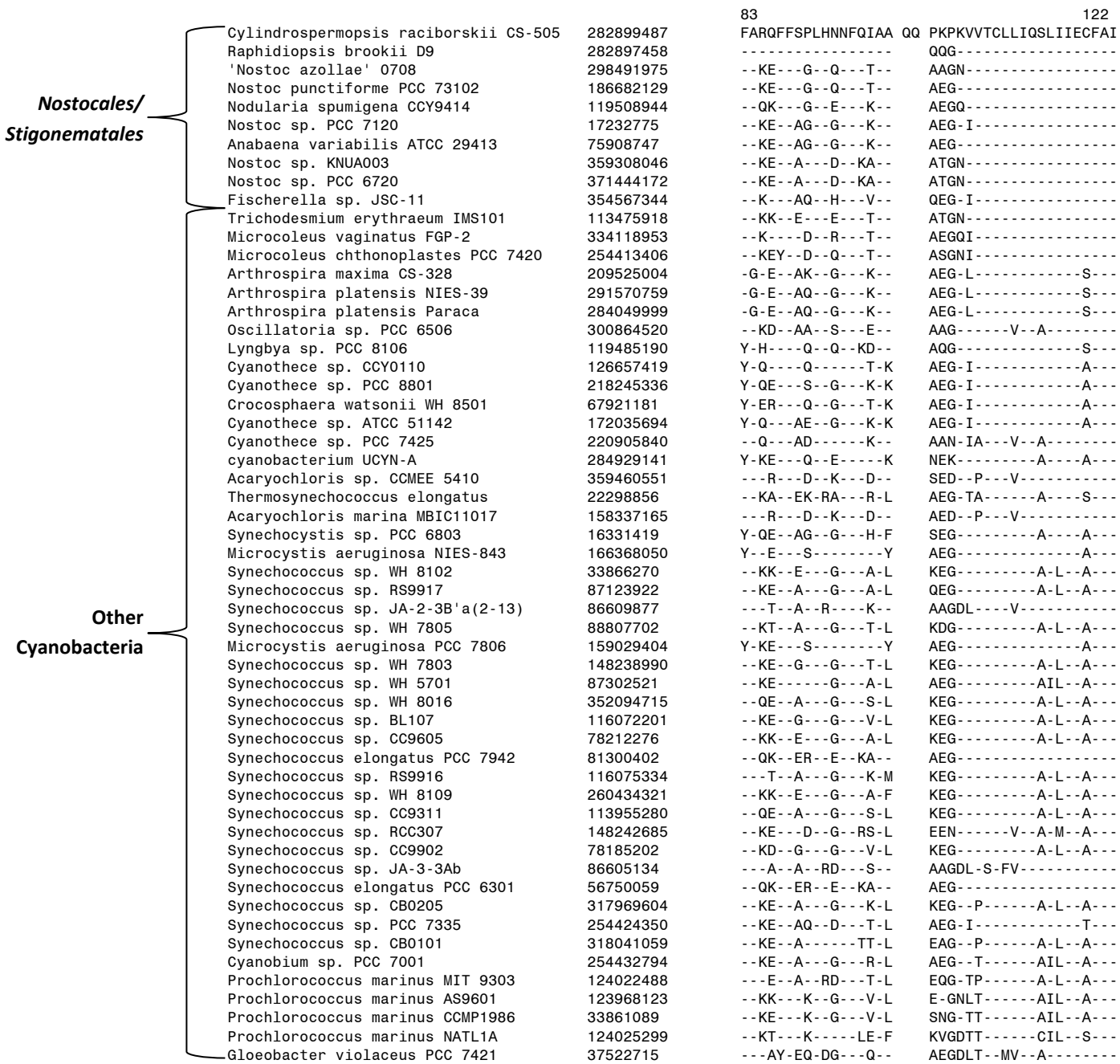
Supplementary Figure 87

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



Supplementary Figure 88

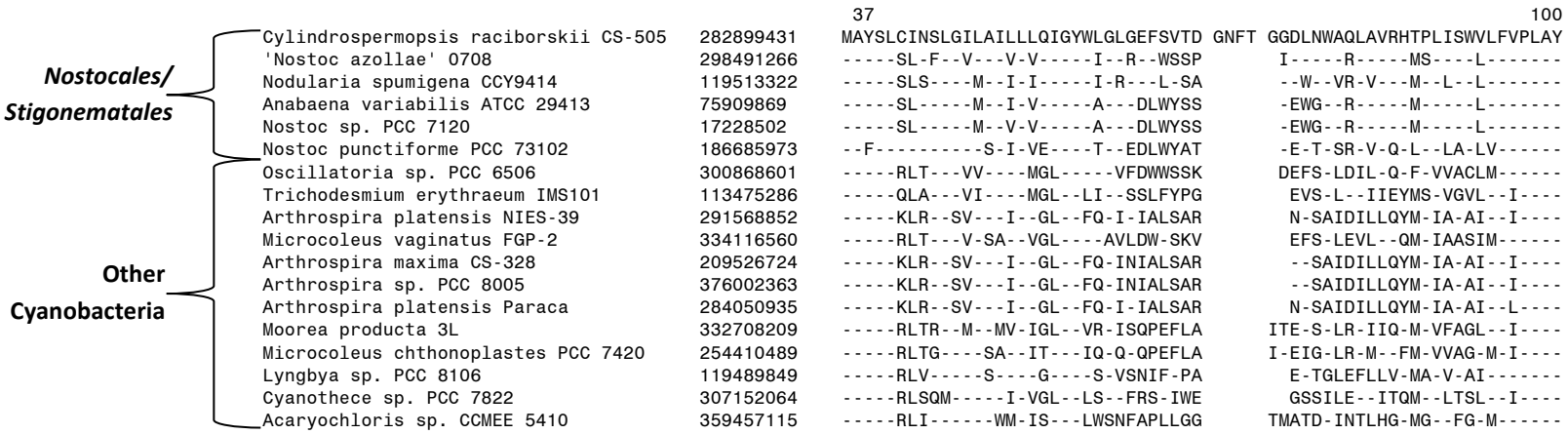
Partial sequence alignment for the protein nitrate reductase, showing a 2 aa insert that is specific for *Cylindrospermopsis raciborskii CS-505.*



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

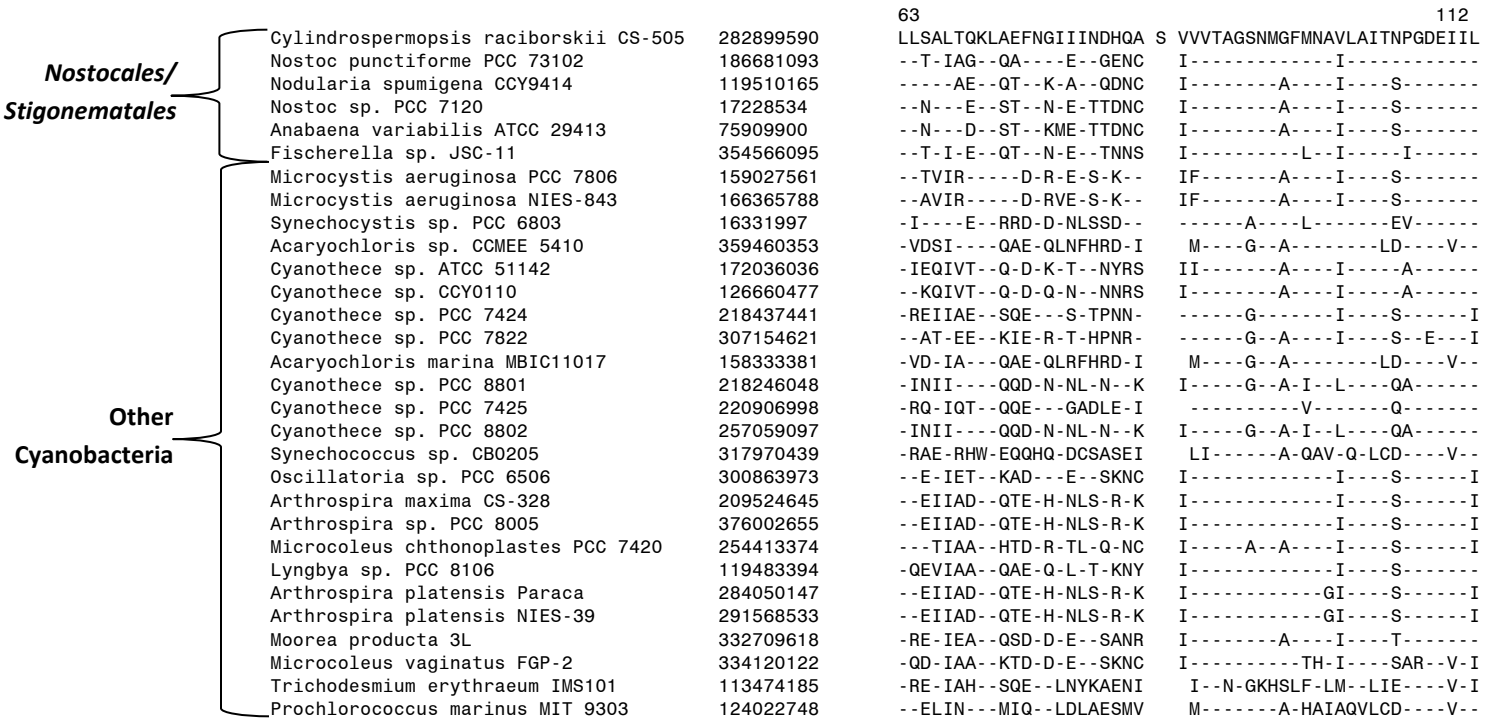
Supplementary Figure 89

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



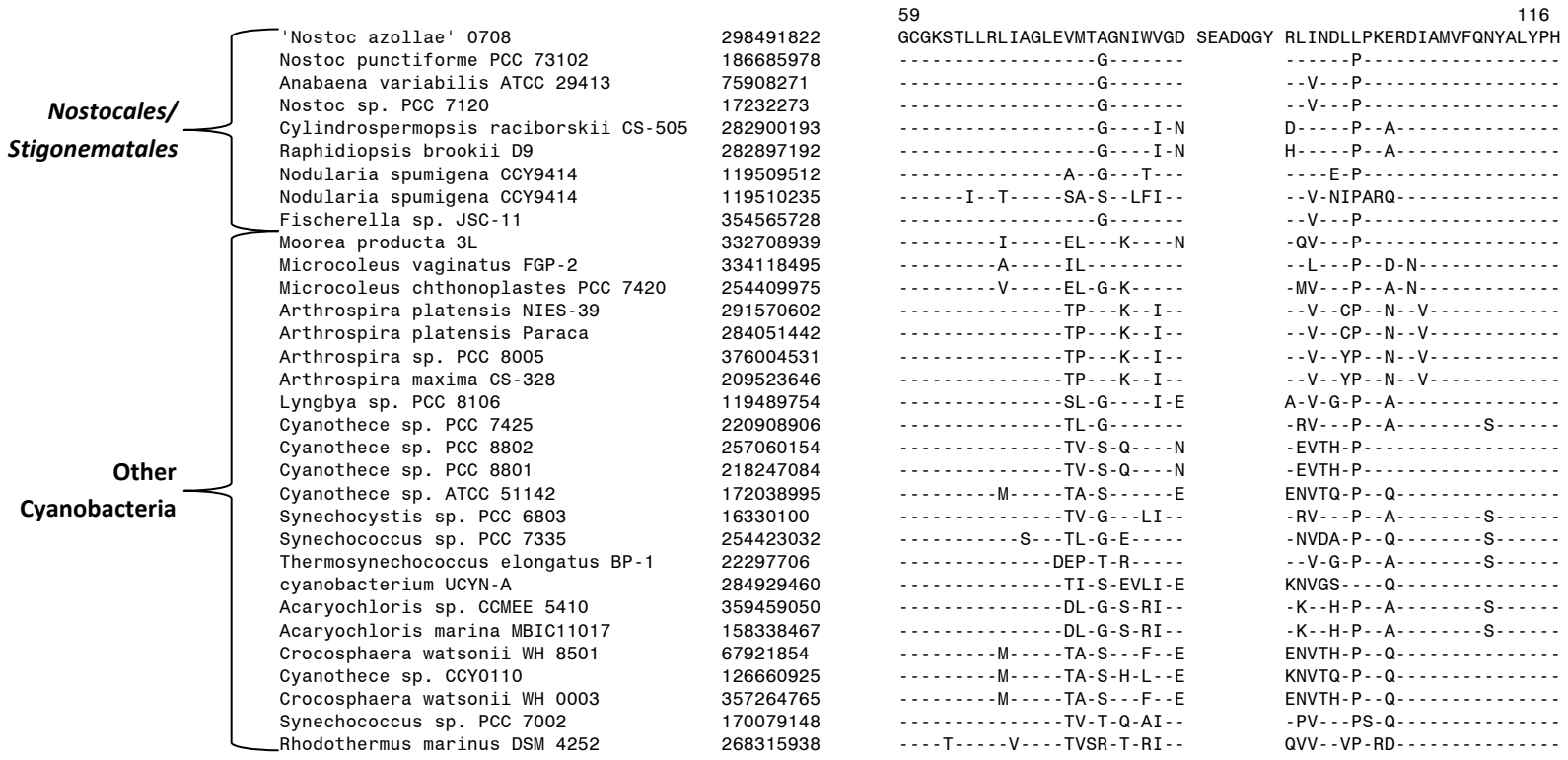
Supplementary Figure 90

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



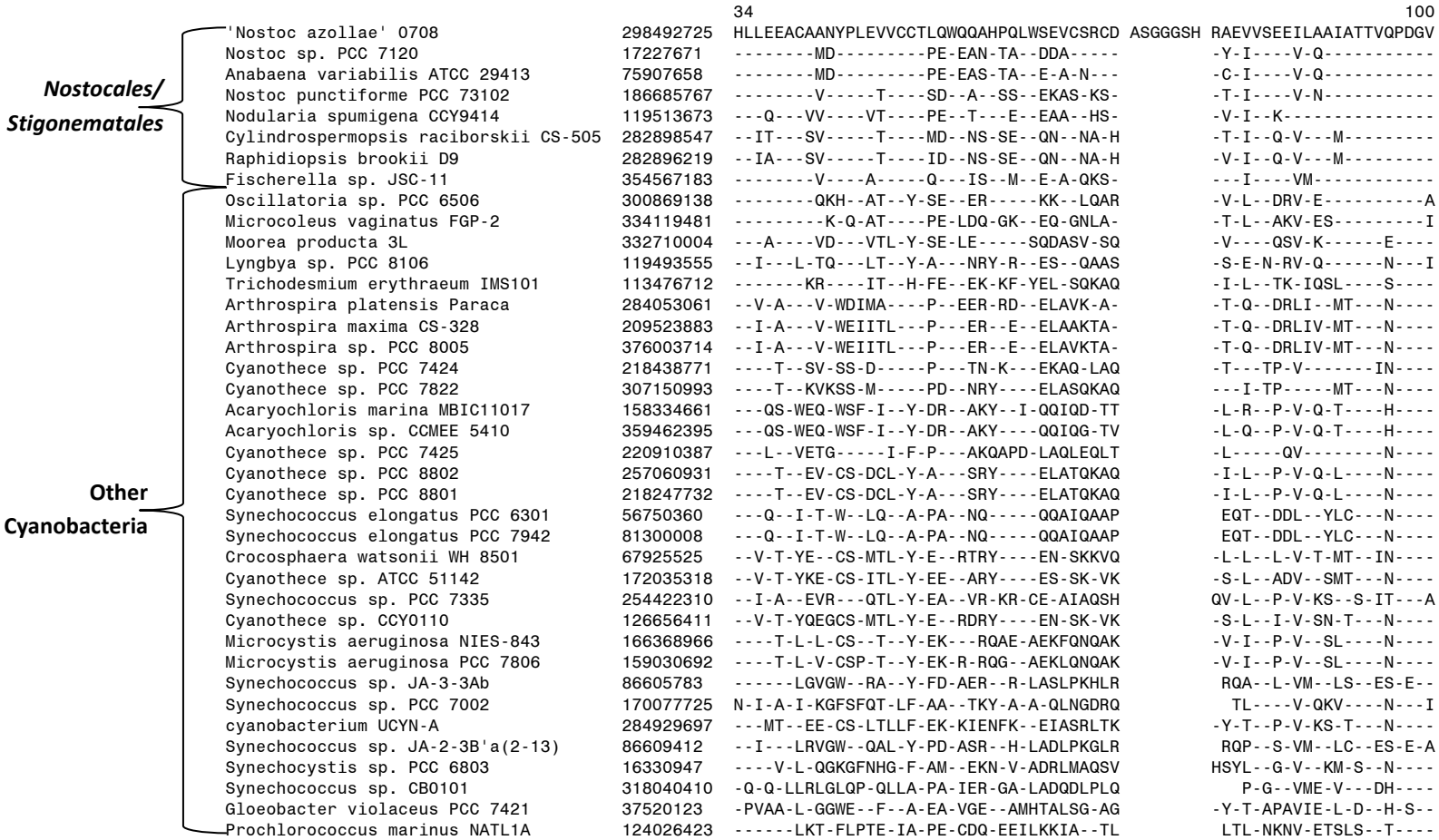
Supplementary Figure 91

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



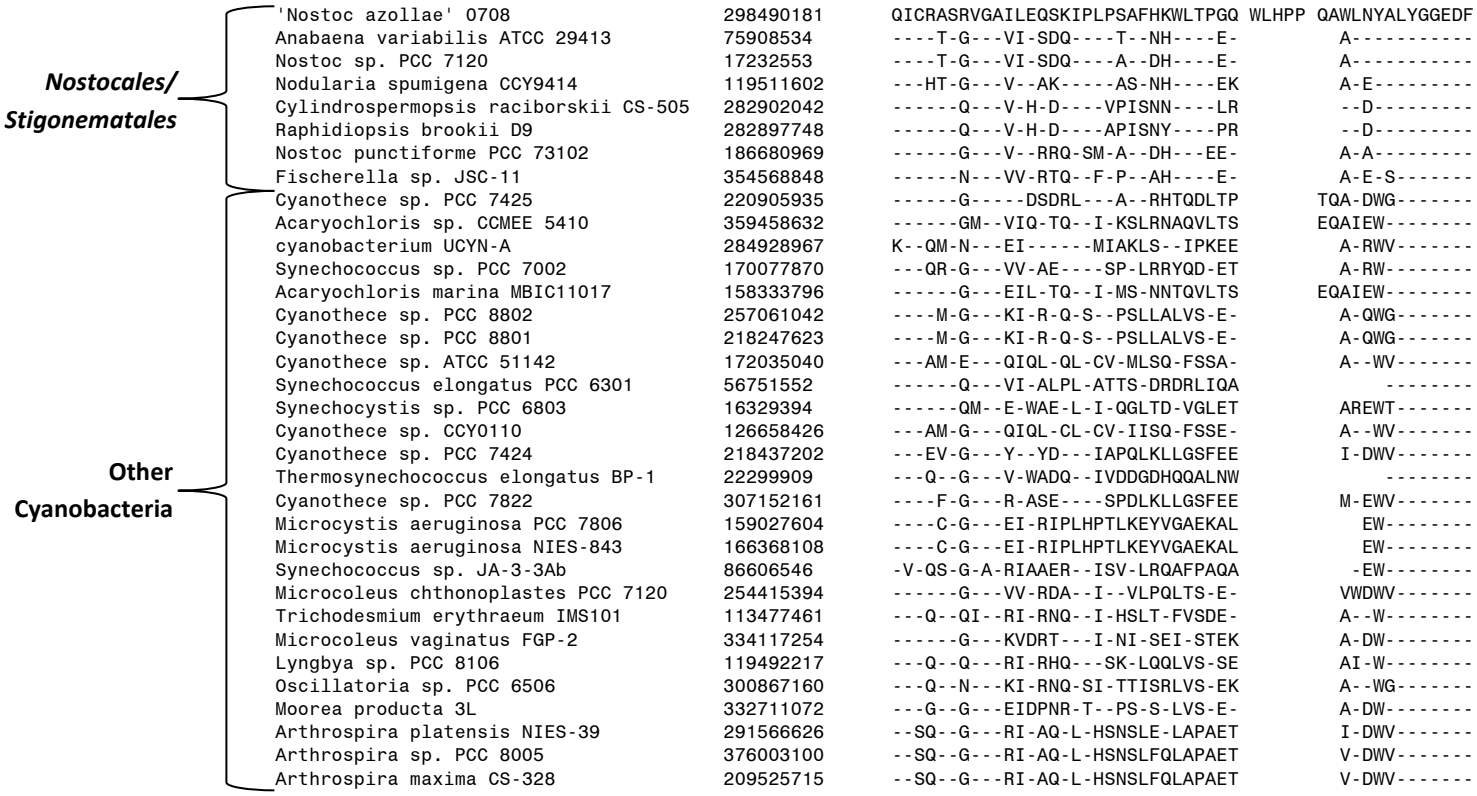
Supplementary Figure 92

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



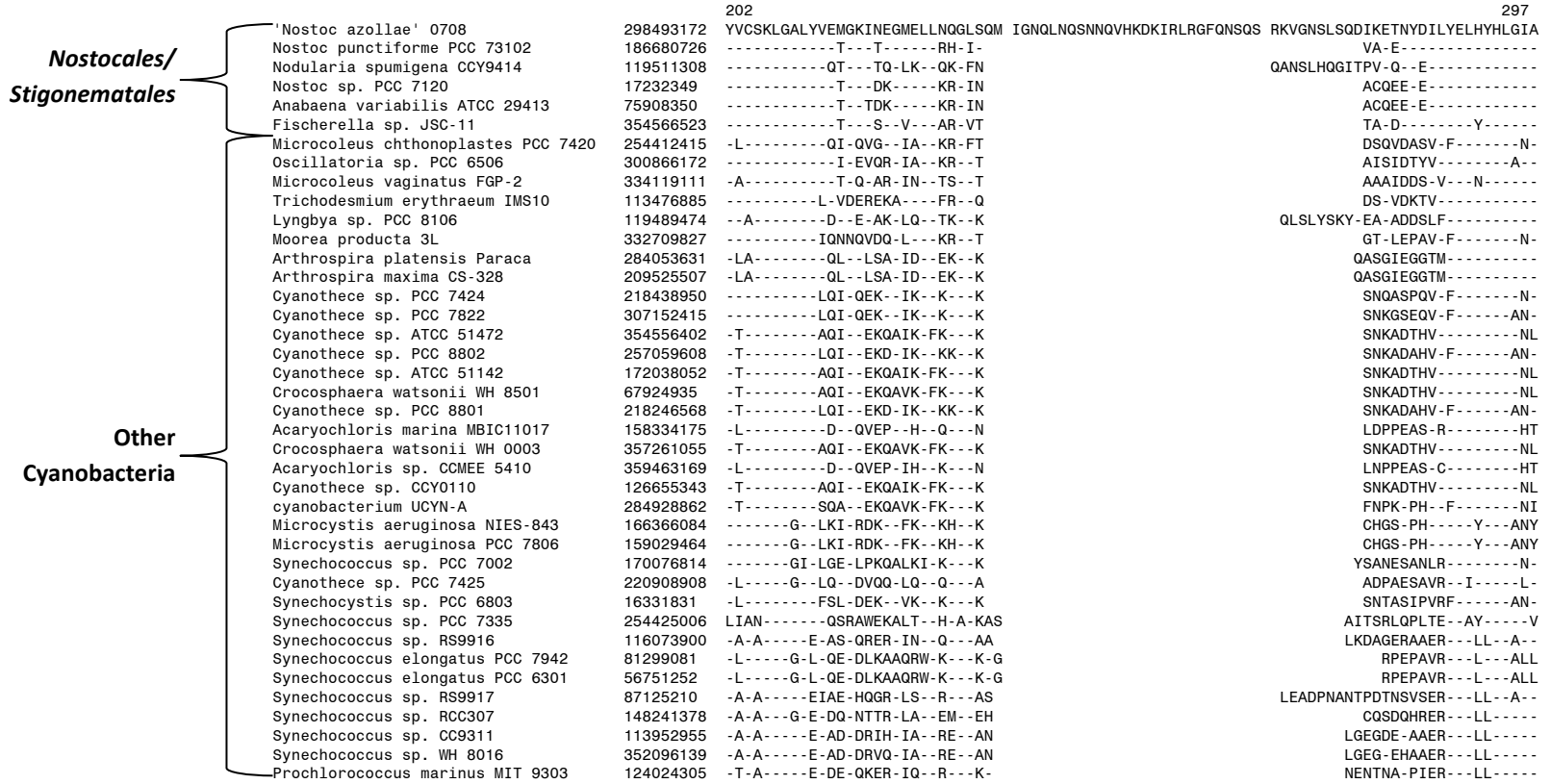
Supplementary Figure 93

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



Supplementary Figure 94

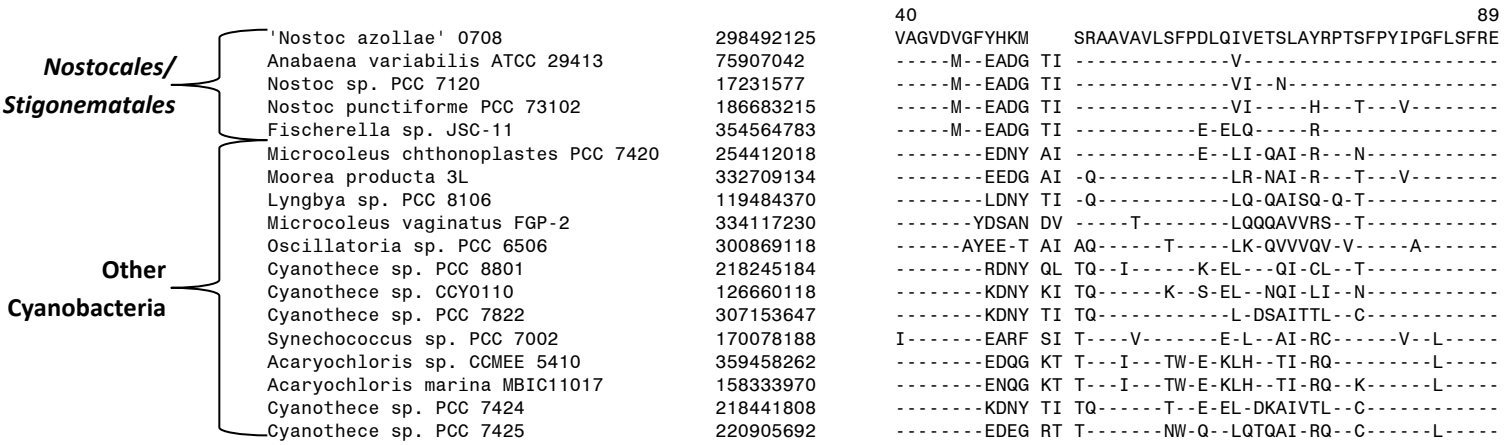
Partial sequence alignment for the protein thiamine-monophosphate kinase, showing a 5 aa insert that is specific for *'Nostoc azollae' 0708.*



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

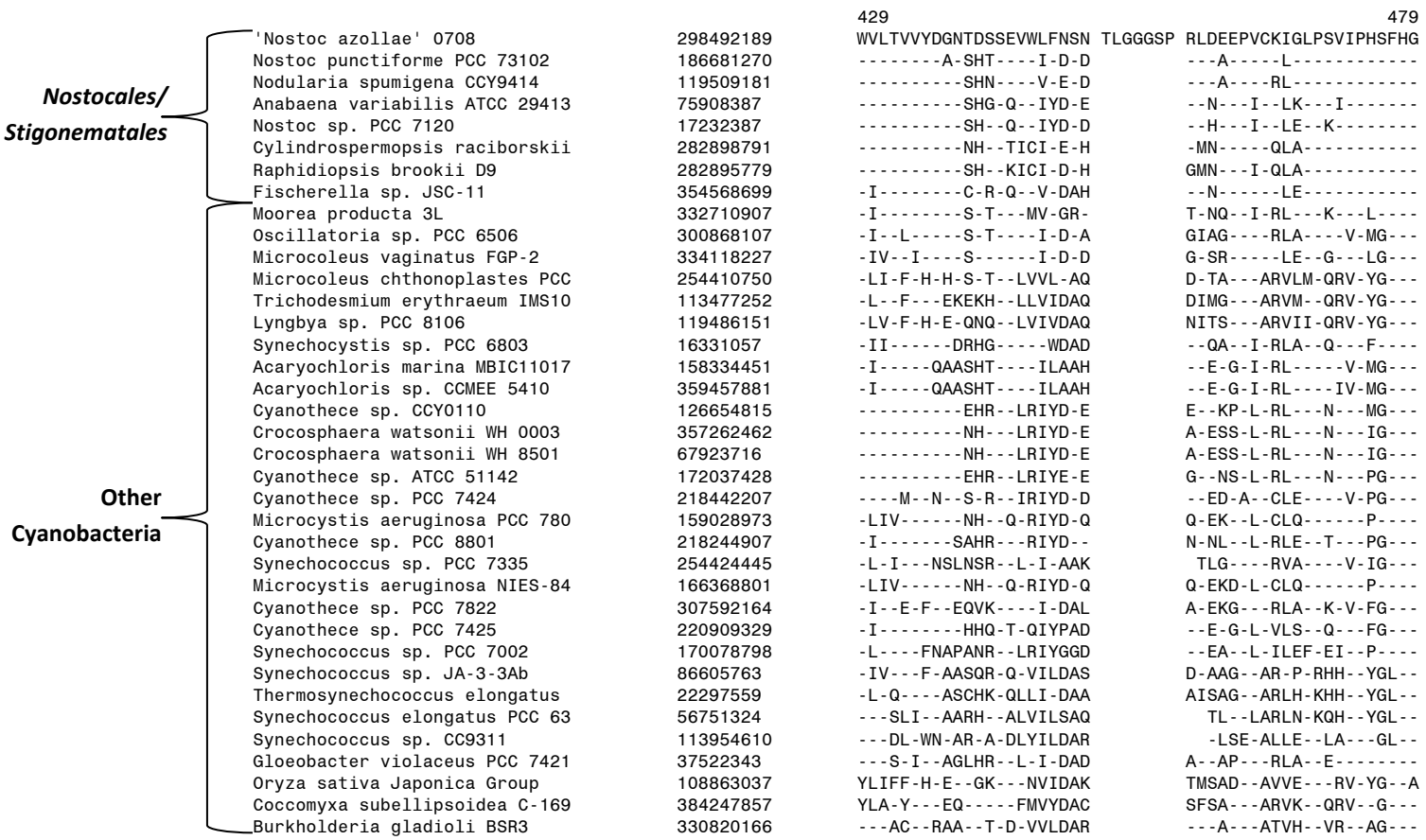
Supplementary Figure 95

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



Supplementary Figure 96

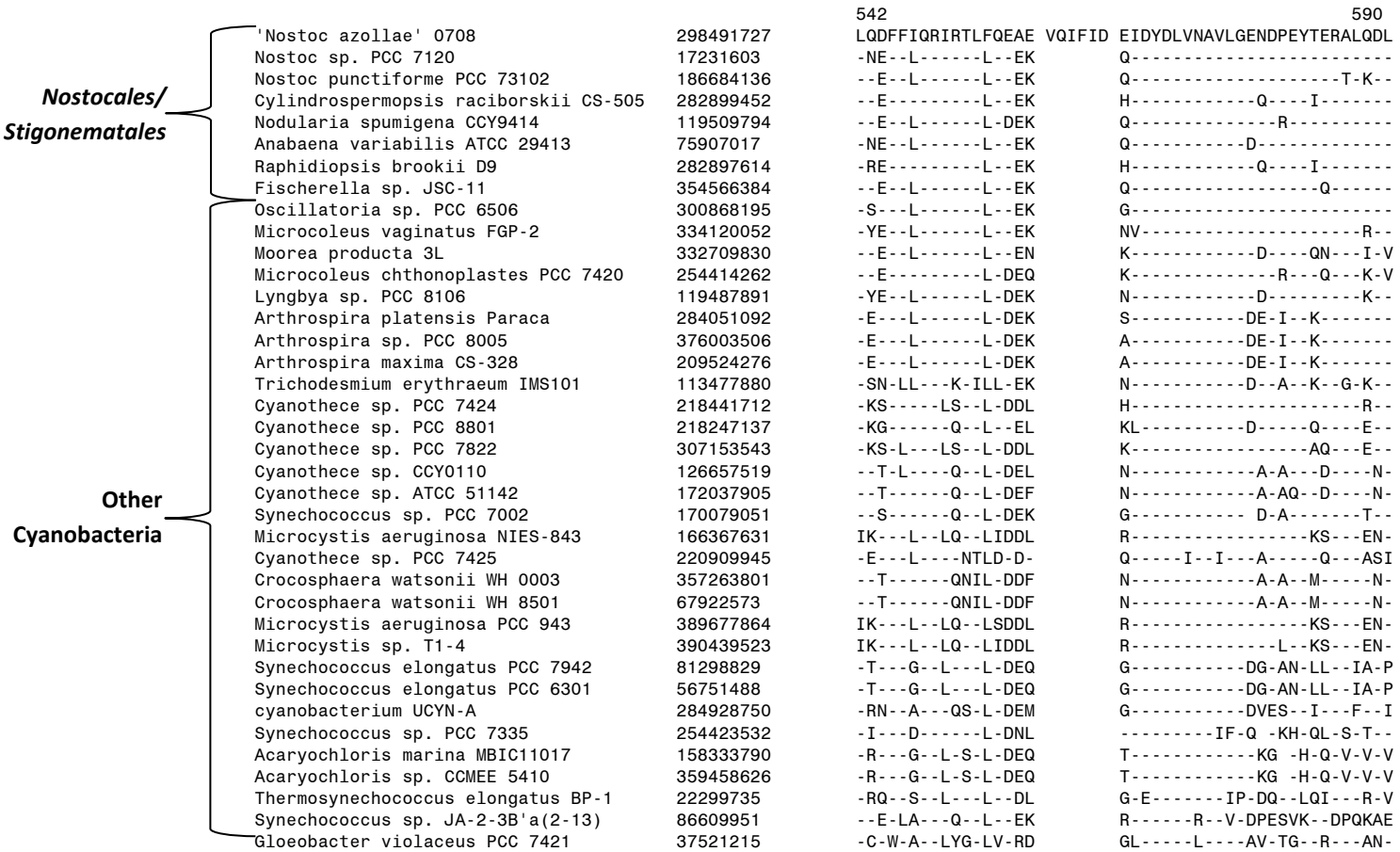
Partial sequence alignment for the protein deoxyribonuclease V, showing a 2 aa deletion that is specific for *'Nostoc azollae' 0708.*



		429	479
'Nostoc azollae' 0708	298492189	WVLTVVYDGNTDSSEVWLFNSN TLGGGSP	RLDEEPVCKIGLPSVIPHSFHG
Nostoc punctiforme PCC 73102	186681270	--------A-SHT----I-D-D	---A-----L----------
Nodularia spumigena CCY9414	119509181	----------SHN----V-E-D	---A----RL----------
Anabaena variabilis ATCC 29413	75908387	----------SHG-Q--IYD-E	--N---I--LK---I-------
Nostoc sp. PCC 7120	17232387	----------SH--Q--IYD-D	--H---I--LE--K--------
Cylindrospermopsis raciborskii	282898791	----------NH--TICI-E-H	-MN-----QLA----------
Raphidiopsis brookii D9	282895779	----------SH--KICI-D-H	GMN---I-QLA----------
Fischerella sp. JSC-11	354568699	-I--------C-R-Q--V-DAH	--N------LE----------
Moorea producta 3L	332710907	-I--------S-T---MV-GR-	T-NQ--I-RL---K---L----
Oscillatoria sp. PCC 6506	300868107	-I--L-----S-T----I-D-A	GIAG----RLA----V-MG---
Microcoleus vaginatus FGP-2	334118227	-IV--I----S------I-D-D	G-SR-----LE--G---LG---
Microcoleus chthonoplastes PCC	254410750	-LI-F-H-H-S-T--LVVL-AQ	D-TA---ARVLM-QRV-YG---
Trichodesmium erythraeum IMS10	113477252	-L--F---EKEKH--LLVIDAQ	DIMG---ARVM--QRV-YG---
Lyngbya sp. PCC 8106	119486151	-LV-F-H-E-QNQ--LVIVDAQ	NITS---ARVII-QRV-YG---
Synechocystis sp. PCC 6803	16331057	-II------DRHG-----WDAD	--QA--I-RLA--Q---F----
Acaryochloris marina MBIC11017	158334451	-I-----QAASHT----ILAAH	--E-G-I-RL-----V-MG---
Acaryochloris sp. CCMEE 5410	359457881	-I-----QAASHT----ILAAH	--E-G-I-RL----IV-MG---
Cyanothece sp. CCY0110	126654815	----------EHR--LRIYD-E	E--KP-L-RL---N---MG---
Crocosphaera watsonii WH 0003	357262462	----------NH---LRIYD-E	A-ESS-L-RL---N---IG---
Crocosphaera watsonii WH 8501	67923716	----------NH---LRIYD-E	A-ESS-L-RL---N---IG---
Cyanothece sp. ATCC 51142	172037428	----------EHR--LRIYE-E	G--NS-L-RL---N---PG---
Cyanothece sp. PCC 7424	218442207	----M--N--S-R--IRIYD-D	--ED-A--CLE----V-PG---
Microcystis aeruginosa PCC 780	159028973	-LIV------NH--Q-RIYD-Q	Q-EK--L-CLQ------P----
Cyanothece sp. PCC 8801	218244907	-I-------SAHR---RIYD--	N-NL--L-RLE--T---PG---
Synechococcus sp. PCC 7335	254424445	-L-I---NSLNSR--L-I-AAK	TLG----RVA----V-IG---
Microcystis aeruginosa NIES-84	166368801	-LIV------NH--Q-RIYD-Q	Q-EKD-L-CLQ------P----
Cyanothece sp. PCC 7822	307592164	-I--E-F--EQVK----I-DAL	A-EKG---RLA--K-V-FG---
Cyanothece sp. PCC 7425	220909329	-I--------HHQ-T-QIYPAD	--E-G-L-VLS--Q---FG---
Synechococcus sp. PCC 7002	170078798	-L----FNAPANR--LRIYGGD	--EA--L-ILEF-EI--P----
Synechococcus sp. JA-3-3Ab	86605763	-IV---F-AASQR-Q-VILDAS	D-AAG--AR-P-RHH--YGL--
Thermosynechococcus elongatus	22297559	-L-Q----ASCHK-QLLI-DAA	AISAG--ARLH-KHH--YGL--
Synechococcus elongatus PCC 63	56751324	---SLI--AARH--ALVILSAQ	TL--LARLN-KQH--YGL--
Synechococcus sp. CC9311	113954610	---DL-WN-AR-A-DLYILDAR	-LSE-ALLE--LA---GL--
Gloeobacter violaceus PCC 7421	37522343	---S-I--AGLHR--L-I-DAD	A--AP---RLA--E--------
Oryza sativa Japonica Group	108863037	YLIFF-H-E--GK---NVIDAK	TMSAD--AVVE---RV-YG--A
Coccomyxa subellipsoidea C-169	384247857	YLA-Y---EQ-----FMVYDAC	SFSA---ARVK--QRV--G---
Burkholderia gladioli BSR3	330820166	---AC--RAA--T-D-VVLDAR	---A---ATVH--VR--AG---

Supplementary Figure 97

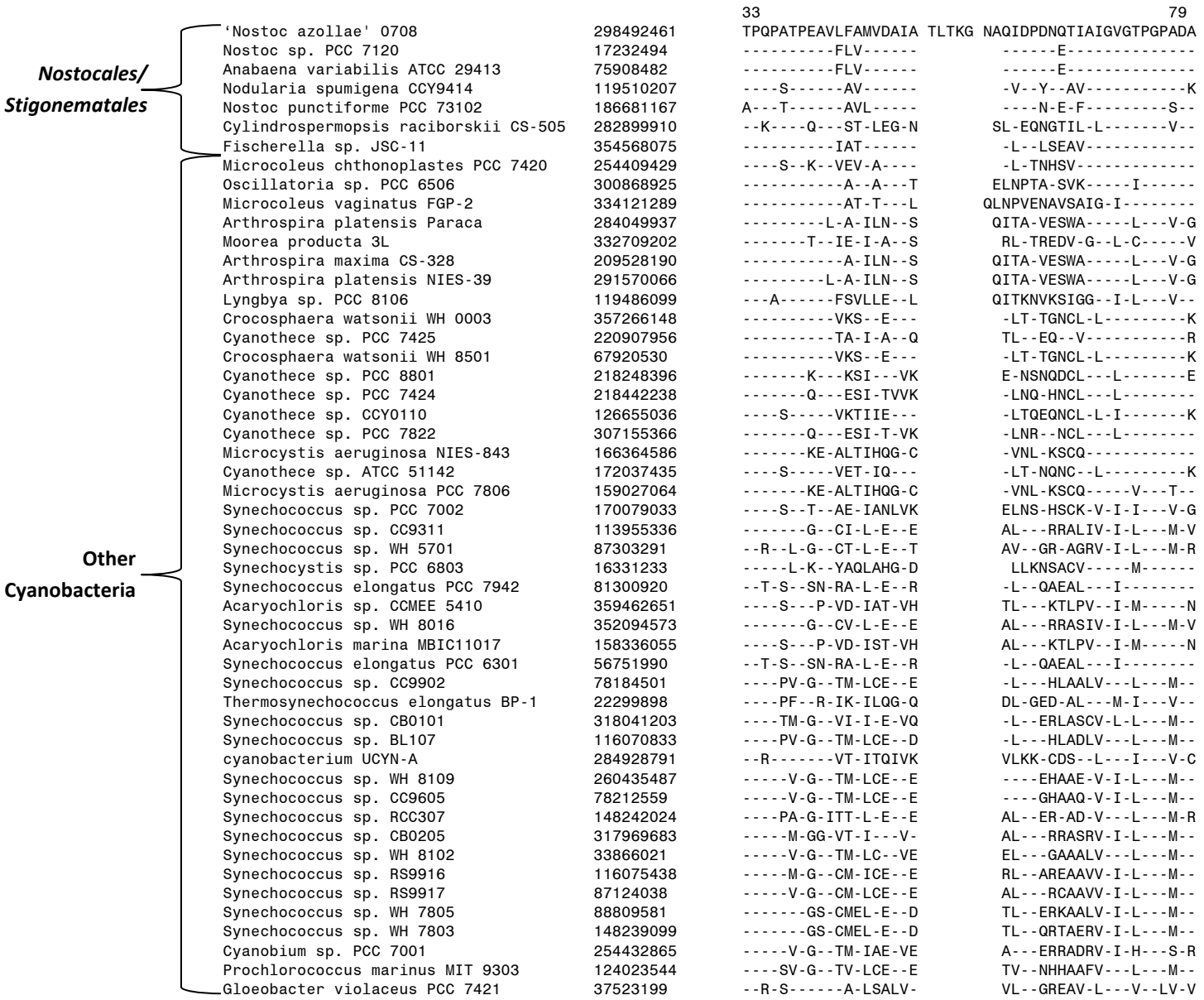
Partial sequence alignment for the protein carotenoid oxygenase, showing a 7 aa insert that is specific for *'Nostoc azollae' 0708*.



		542	590
'Nostoc azollae' 0708	298491727	LQDFFIQRIRTLFQEAE VQIFID	EIDYDLVNAVLGENDPEYTERALQDL
Nostoc sp. PCC 7120	17231603	-NE--L------L--EK	Q------------------------
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-------
Trichodesmium erythraeum IMS101	113477880	-SN-LL---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
Crocosphaera watsonii WH 0003	357263801	--T------QNIL-DDF	N------------A-A--M-----N-
Crocosphaera 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. CCMEE 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
'Nostoc azollae' 0708	298492461	TPQPATPEAVLFAMVDAIA TLTKG	NAQIDPDNQTIAIGVGTPGPADA
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--------
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--
Crocosphaera watsonii WH 0003	357266148	----------VKS--E---	-LT-TGNCL-L---------K
Cyanothece sp. PCC 7425	220907956	----------TA-I-A--Q	TL--EQ--V----------R
Crocosphaera watsonii WH 8501	67920530	----------VKS--E---	-LT-TGNCL-L---------K
Cyanothece sp. PCC 8801	218248396	-------K---KSI---VK	E-NSNQDCL---L-------E
Cyanothece sp. PCC 7424	218442238	-------Q---ESI-TVVK	-LNQ-HNCL---L--------
Cyanothece sp. CCY0110	126655036	----S-----VKTIIE---	-LTQEQNCL-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. ATCC 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--QAEAL---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--QAEAL---I--------
Synechococcus sp. CC9902	78184501	----PV-G--TM-LCE--E	-L---HLAALV---L---M--
Thermosynechococcus elongatus BP-1	22299898	----PF--R-IK-ILQG-Q	DL-GED-AL---M-I---V--
Synechococcus sp. CB0101	318041203	----TM-G--VI-I-E-VQ	-L--ERLASCV-L-L---M--
Synechococcus sp. BL107	116070833	----PV-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	----EHAAE-V-I-L---M--
Synechococcus sp. CC9605	78212559	-----V-G--TM-LCE--E	----GHAAQ-V-I-L---M--
Synechococcus sp. RCC307	148242024	----PA-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--AREAAVV-I-L---M--
Synechococcus sp. RS9917	87124038	-----V-G--CM-LCE--E	AL---RCAAVV-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---ERRADRV-I-H---S-R
Prochlorococcus marinus MIT 9303	124023544	----SV-G--TV-LCE--E	TV--NHHAAFV---L---M--
Gloeobacter violaceus PCC 7421	37523199	--R-S------A-LSALV-	VL--GREAV-L---V--LV-V

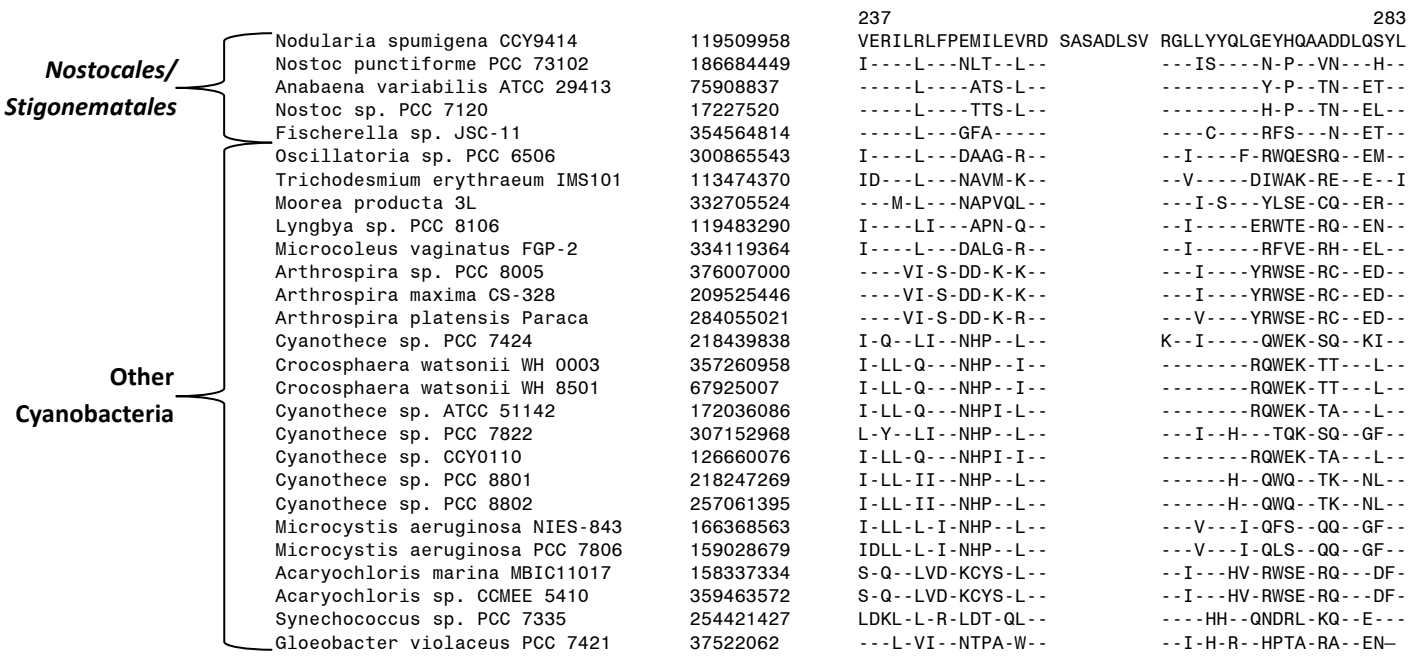
Supplementary Figure 99

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

			335	385
Nostocales/ Stigonematales	'Nostoc azollae' 0708	298490634	GAKVSVGPNFYHVFAAATPD SL	HGLPMQELSLETAENVILIDDLQALHAGV
	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-------GNSPAE	--F---N----------
Other Cyanobacteria	Acaryochloris sp. CCMEE 5410	359463925	----T---H-------S--E	-TY---------F---V-------
	Acaryochloris marina MBIC11017	158335615	----T---H-------S--E	-TY--------VF---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---SV-------
	Microcystis aeruginosa NIES-843	166366110	-S--T-S-H------S-NQP	KSY--------V---S--------
	Cyanothece sp. PCC 7424	218438331	----T-S--Y---VGEE-SG	ESY----V---VY--N--------
	Synechococcus sp. JA-3-3Ab	86606884	----T-AAHY-----G-P-E	STYD-A-----VW--E--------
	Cyanothece sp. ATCC 51142	172035569	----T--SH-----SNSETN	P-YN-DE--------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--------
	Crocosphaera watsonii WH 0003	357263212	----T--SH-----SNSEAN	PQYN-DE-----F--Q--------
	Crocosphaera 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---ISNQQAN	F-CN-NMVK--VF--K--------
	Thermosynechococcus elongatus BP-1	22298624	----T-SGH--D-P-GEGG-	RPDGLVW--E-H------
	Synechococcus sp. WH 8102	33865730	--------DWFV-GTTPGQT	SGQELDHTRE-DTFV--E--S------
	Synechococcus sp. BL107	116070615	--------DWFVLSTSEGCS	SGNNLDQSTEKDTFV--E--S------
	Synechococcus sp. WH 7803	148239436	--------DWFE-HRSEGAT	LPGDHLDHRT-SDSFV--ES-N------
	Synechococcus sp. WH 7805	88808727	--------DWFE-HRSEGCR	PPADHLDHRK-QDSFVM-ES-N------
	Synechococcus sp. CC9902	78184740	--------DWFVLGTSEGCS	SGNTLDQSTEKDTFV--E--S------
	Synechococcus sp. CC9605	78212838	--------DWFV-GSNPQVS	SGNSLDHRTESGPFV--E--S------
	Synechococcus sp. RS9916	116074908	--------DWFE-GTSEGTS	SGTTHLHHAT-T-PFV--ES-N------
	Synechococcus sp. WH 8109	260434581	--------DWFV-GSNPQVS	SGNSLNHRTESGPFV--E--S------
	Synechococcus sp. CC9311	113955164	--------DWFE-SSSEGSS	RPVEHLDHRN-RDTFV--EG-N------
	Synechococcus sp. RCC307	148242573	------S-DWFEIG-TPGAG	CGDAS-NRFQ-DG-VW--S-S------
	Synechococcus sp. CB0101	318041667	--------DWFEIGTTPGS-	-GQAGLNRFKAS-PIVW--S-S------
	Synechococcus sp. WH 5701	87302397	--------HWFEIGP-GSS-	EQTGARG-DRFSQDGIVW--S-S------
	Synechococcus sp. RS9917	87125803	--------DWFE-GTSDGMS	SGNSHLHHARSTDTFV--EG-N------
	Cyanobium sp. PCC 7001	254430985	--------DWFEIGPS-GS-	-GQQGLDRFN-D-PLVW--S-S------
	Synechococcus sp. WH 8016	352093854	--------DWFE-SSSQGLI	SPSPHLDHRS-Q-PFV--ES-S------
	Synechococcus sp. CB0205	317970133	--------DWFEIGPTPGTS	-GQSGLDRFKADQPI-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-DWLV-SKSESEM	DKDESLSIKN-IKEY----E-S-I-S--
	Prochlorococcus marinus 9313	33863043	--------DWFE-SKTEGQN	SANTHLQHSRSK-TFV--EG-N------
	Prochlorococcus marinus CCMP1375	33240446	------S-DWPVIYKTKGET	AKYPDLQFQH-K-EFV-VEN-H-----I

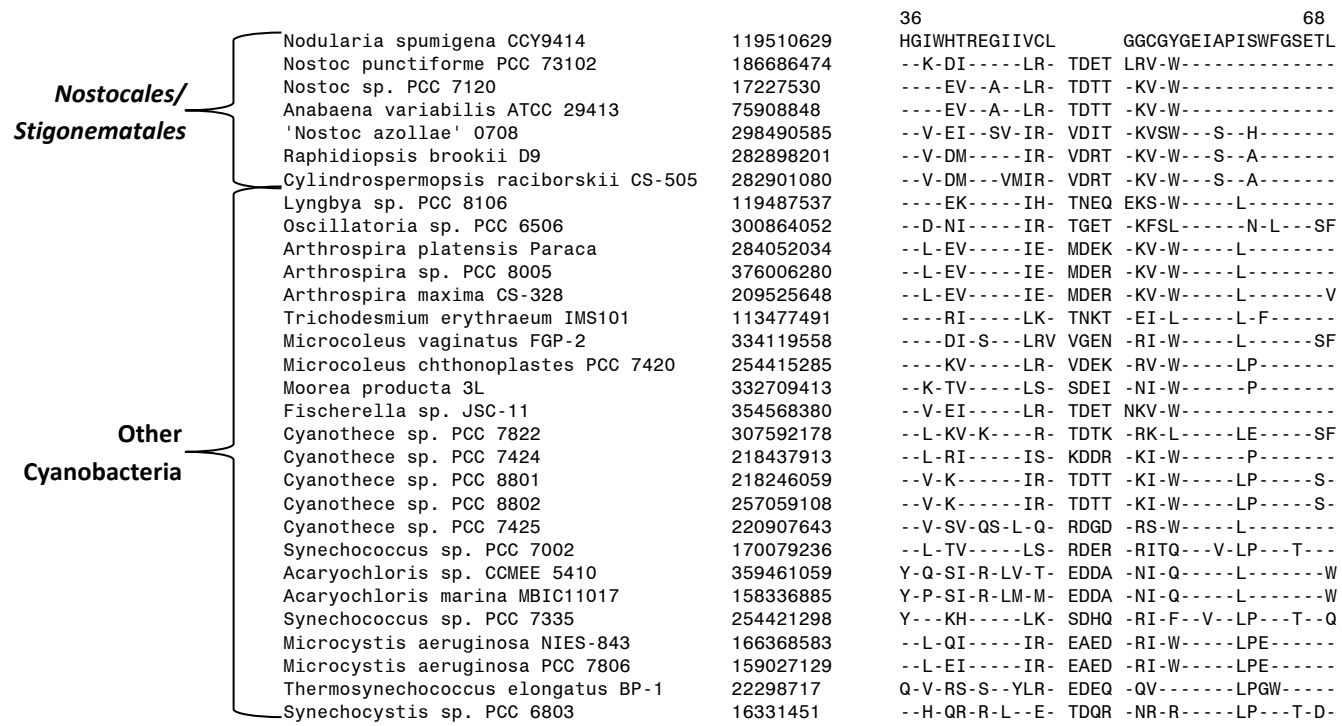
Supplementary Figure 100

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



Supplementary Figure 101

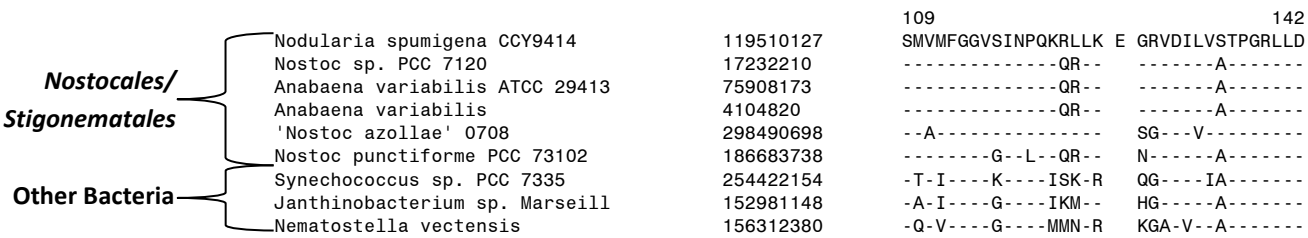
Partial sequence alignment for the hypothetical protein N9414_05744, showing an 8 aa insert that is specific for *Nodularia spumigena CCY9414.*



		36	68
Nodularia spumigena CCY9414	119510629	HGIWHTREGIIVCL	GGCGYGEIAPISWFGSETL
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-------
Fischerella sp. JSC-11	354568380	--V-EI-----LR-	TDET NKV-W----------
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. CCMEE 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-

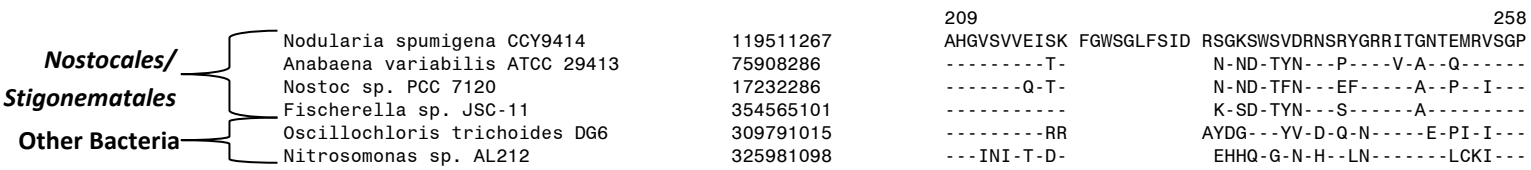
Supplementary Figure 102

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



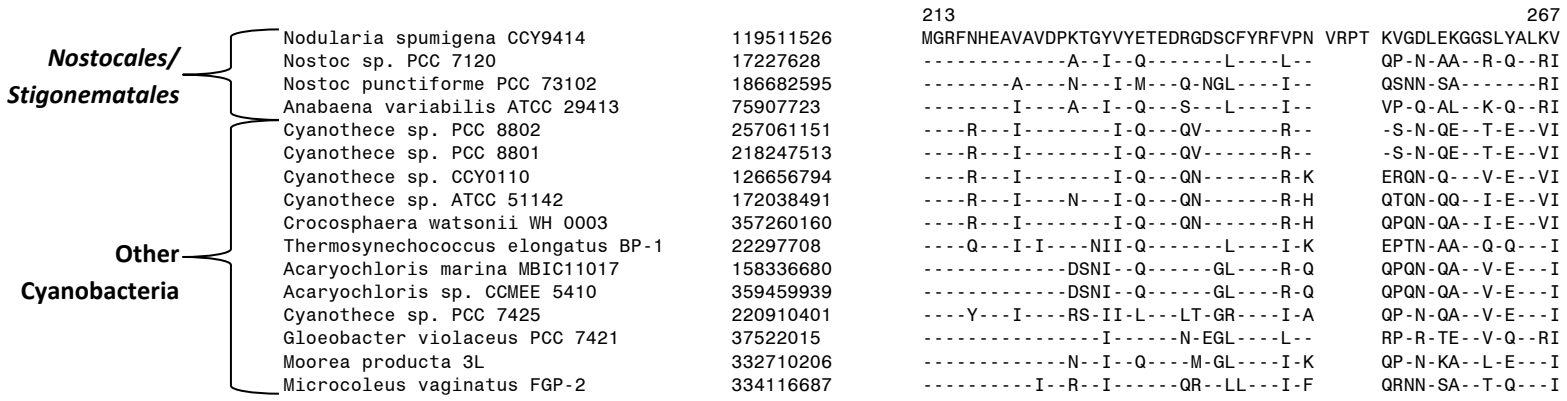
Supplementary Figure 103

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



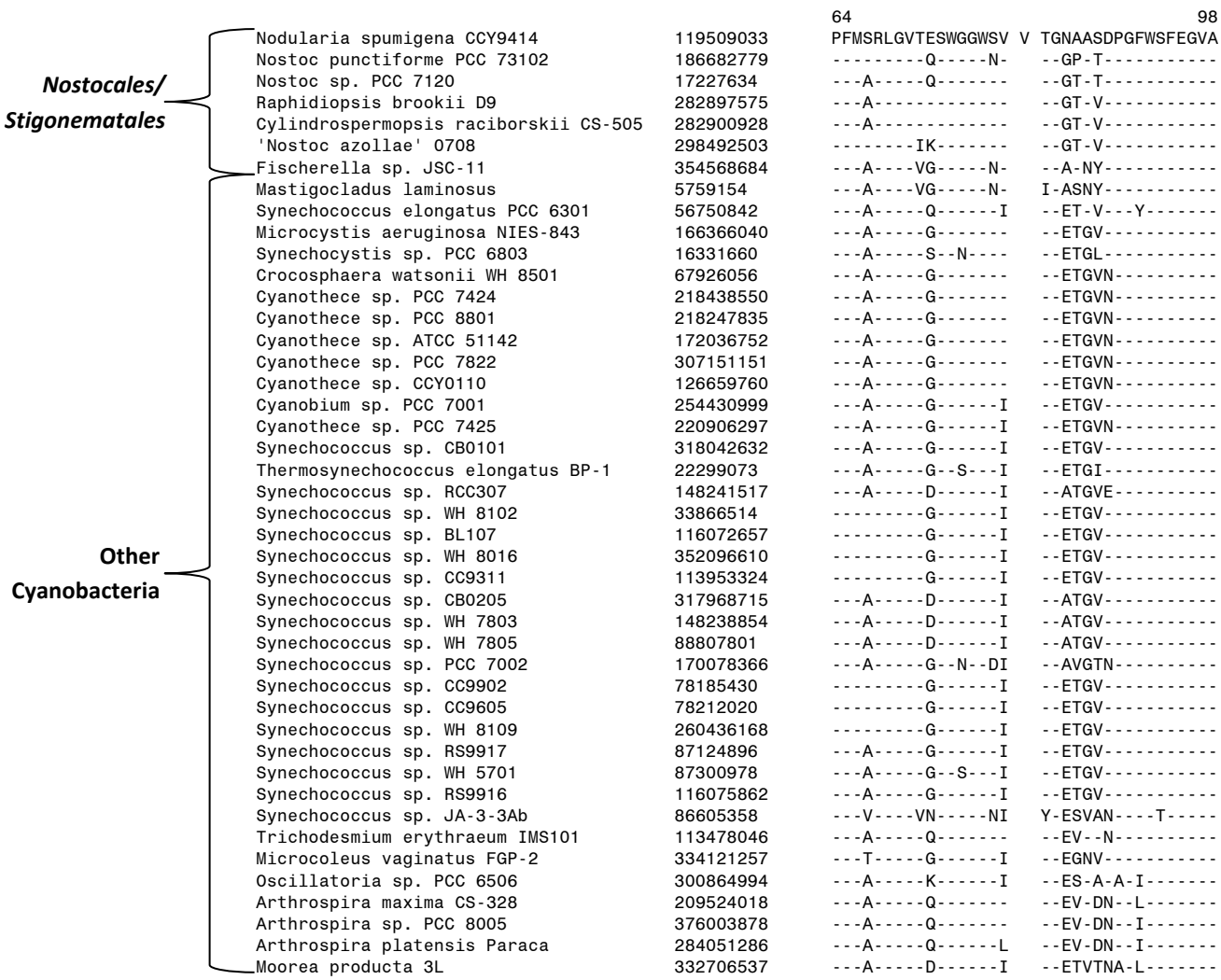
Supplementary Figure 104

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



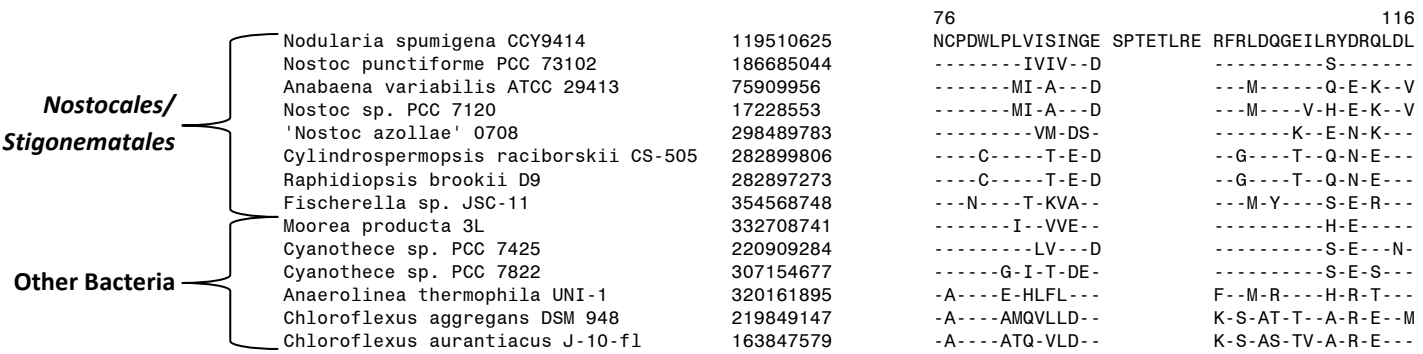
Supplementary Figure 105

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



Supplementary Figure 106

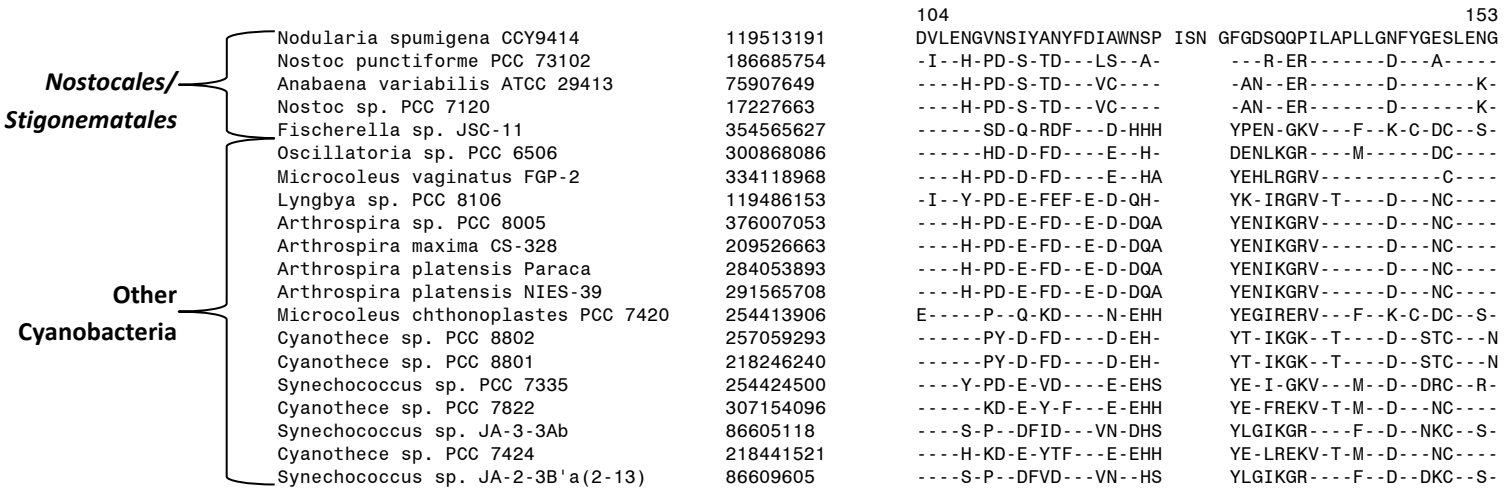
Partial sequence alignment for the protein photosystem II chlorophyll-binding protein CP47, showing a 1 aa insert that is specific for *Nodularia spumigena CCY9414.*



		76	116
Nodularia spumigena CCY9414	119510625	NCPDWLPLVISINGE	SPTETLRE RFRLDQGEILRYDRQLDL
Nostoc punctiforme PCC 73102	186685044	--------IVIV--D	----------S-------
Anabaena variabilis ATCC 29413	75909956	-------MI-A---D	---M------Q-E-K--V
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-
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-fl	163847579	-A----ATQ-VLD--	K-S-AS-TV-A-R-E---

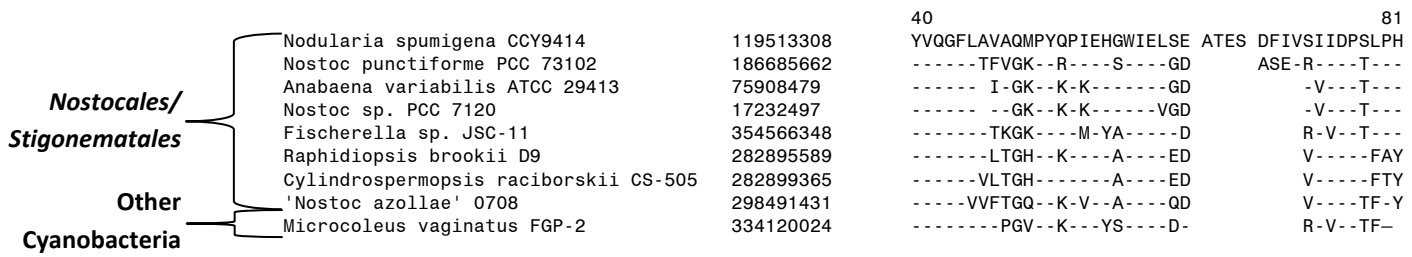
Supplementary Figure 107

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.*



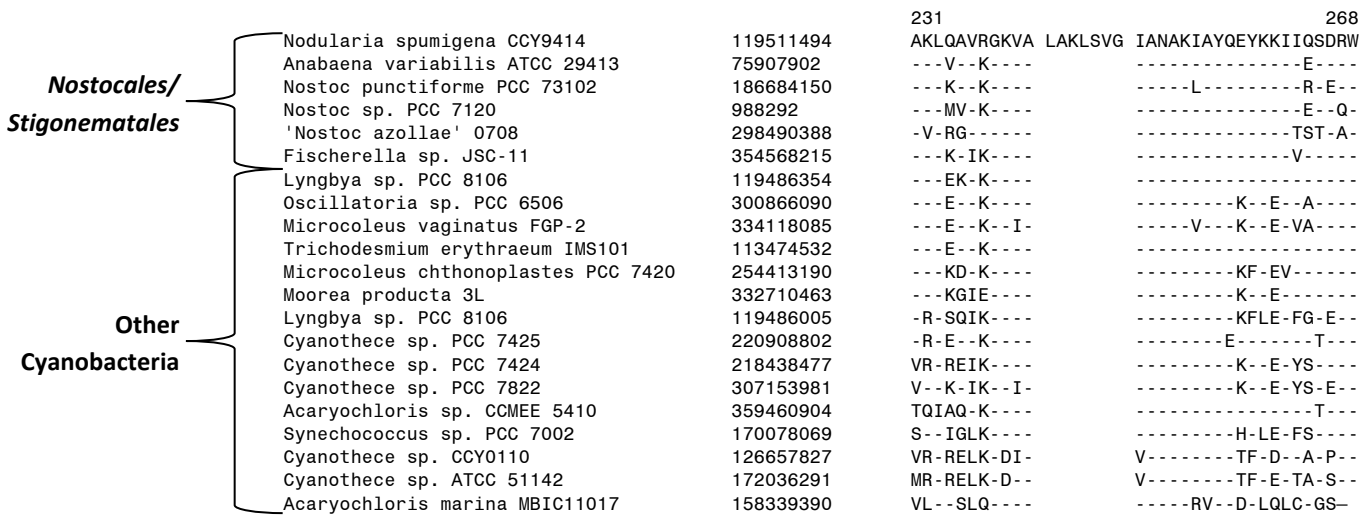
Supplementary Figure 108

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



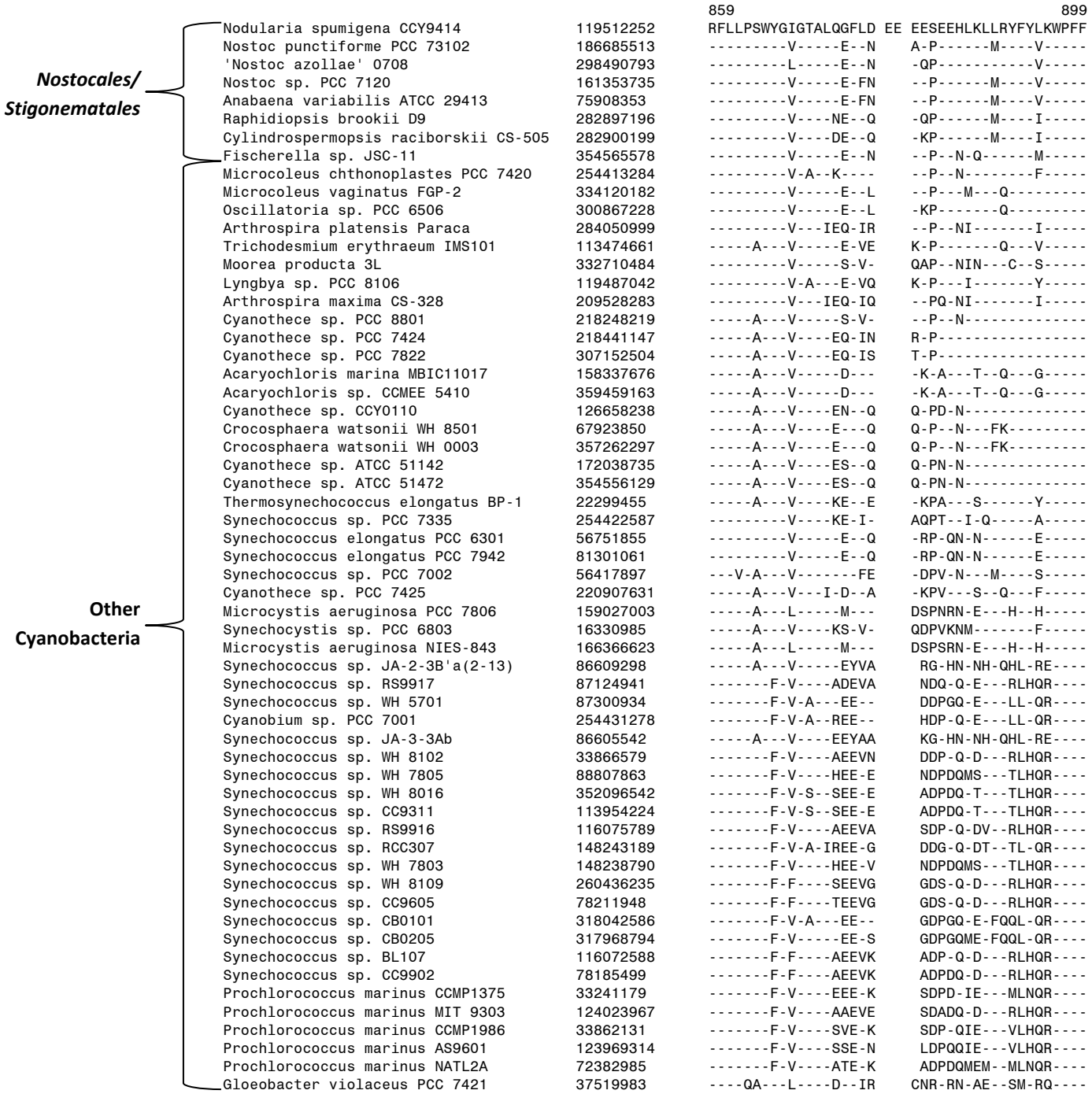
Supplementary Figure 109

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



Supplementary Figure 110

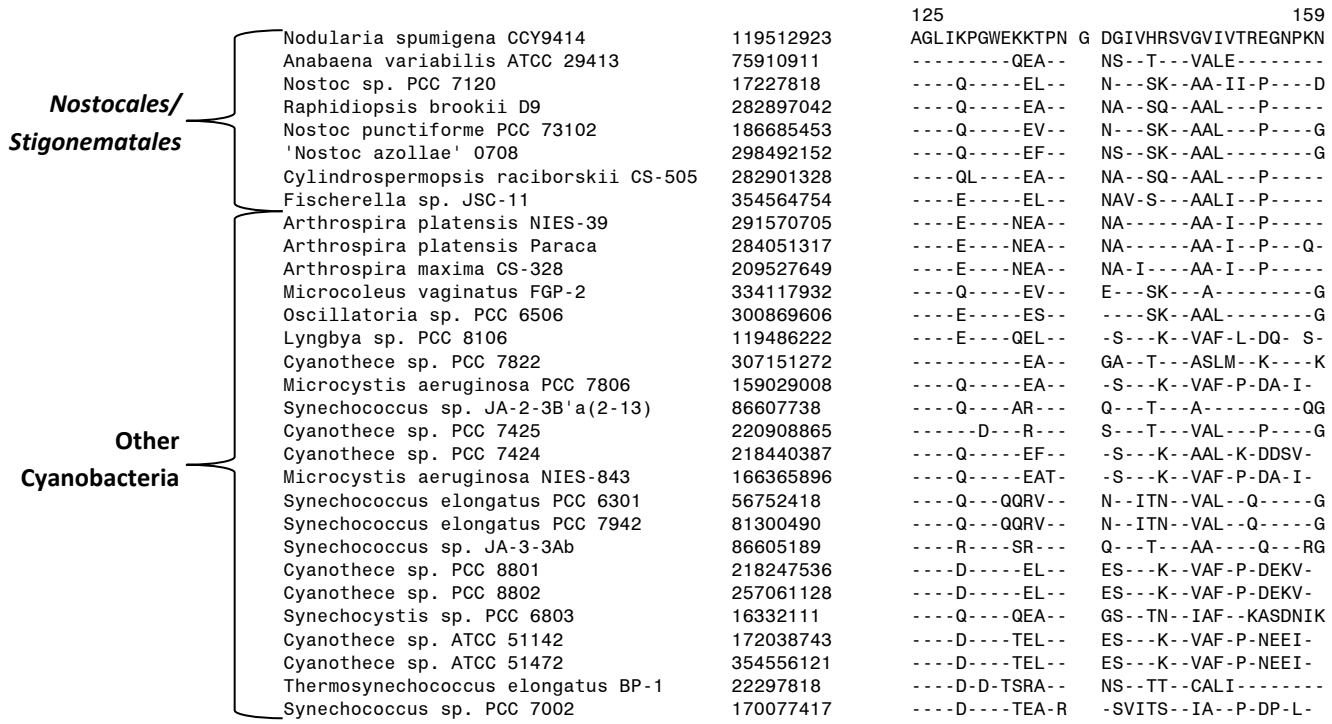
Partial sequence alignment for the protein transaldolase, showing a 7 aa insert that is specific for *Nodularia spumigena CCY9414.*



		859 899
Nodularia spumigena CCY9414	119512252	RFLLPSWYGIGTALQGFLD EE EESEEHLKLLRYFYLKWPFF
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-----
Lyngbya sp. PCC 8106	119487042	---------V-A---E-VQ K-P---I-------Y-----
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--------------
Crocosphaera watsonii WH 8501	67923850	-----A---V----E---Q Q-P--N---FK---------
Crocosphaera 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

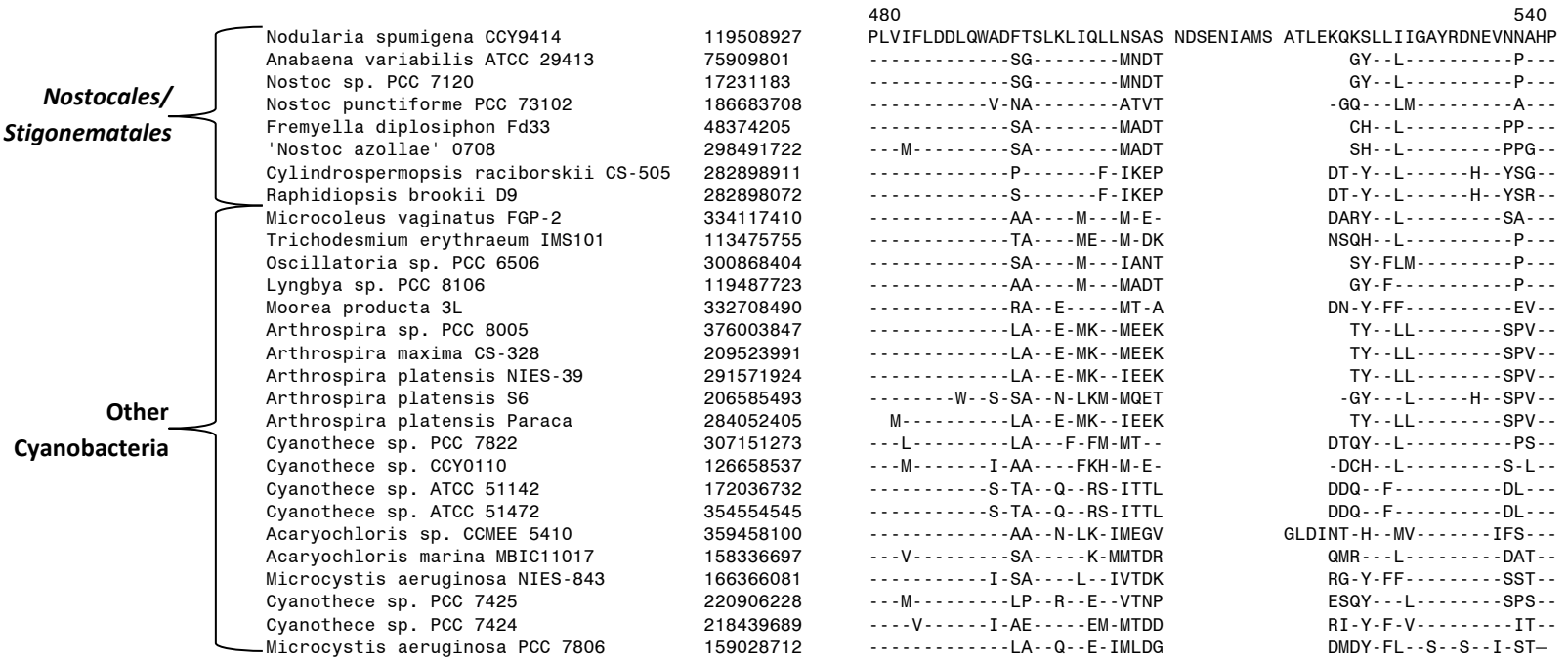
Partial sequence alignment for the protein phosphoenolpyruvate carboxylase, showing a 2 aa insert that is specific for *Nodularia spumigena CCY9414.*



		125	159
Nodularia spumigena CCY9414	119512923	AGLIKPGWEKKTPN G	DGIVHRSVGVIVTREGNPKN
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-
Synechococcus sp. JA-2-3B'a(2-13)	86607738	----Q----AR---	Q---T---A---------QG
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-

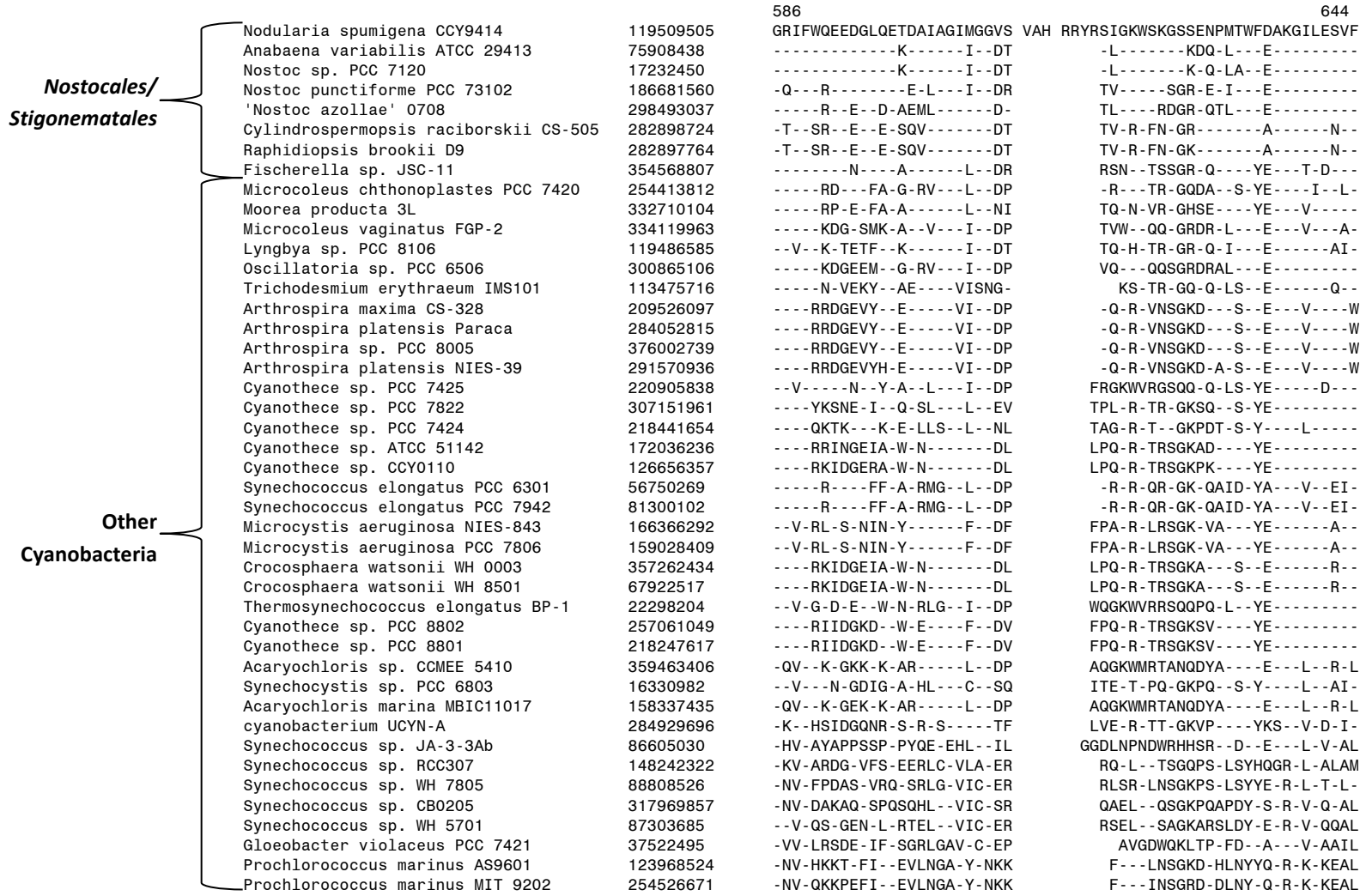
Supplementary Figure 112

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



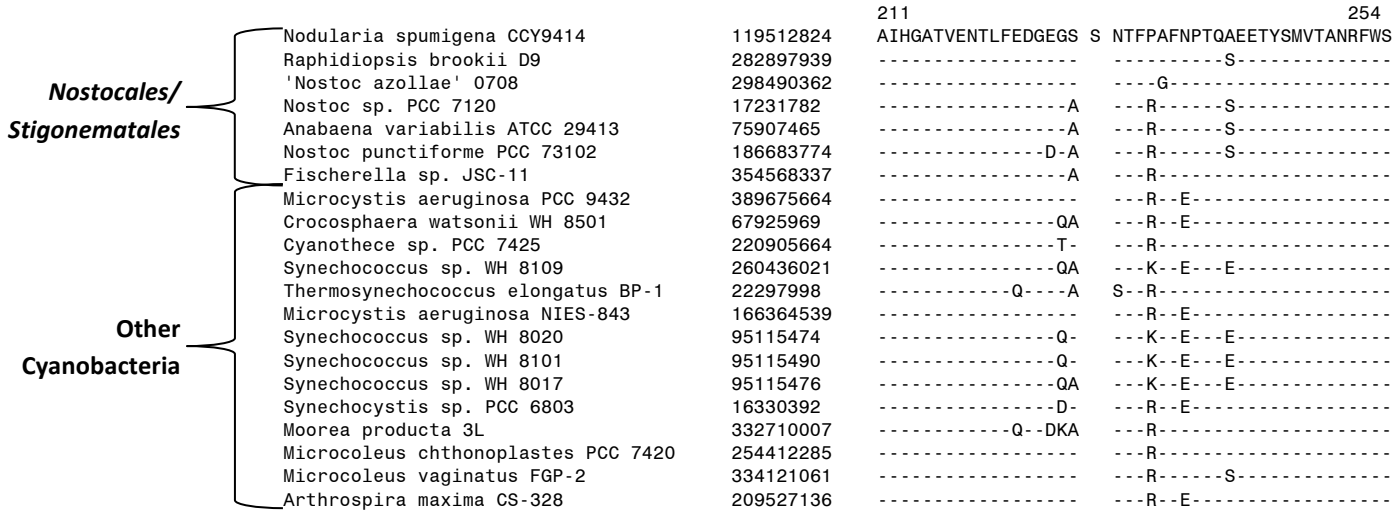
Supplementary Figure 113

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



Supplementary Figure 114

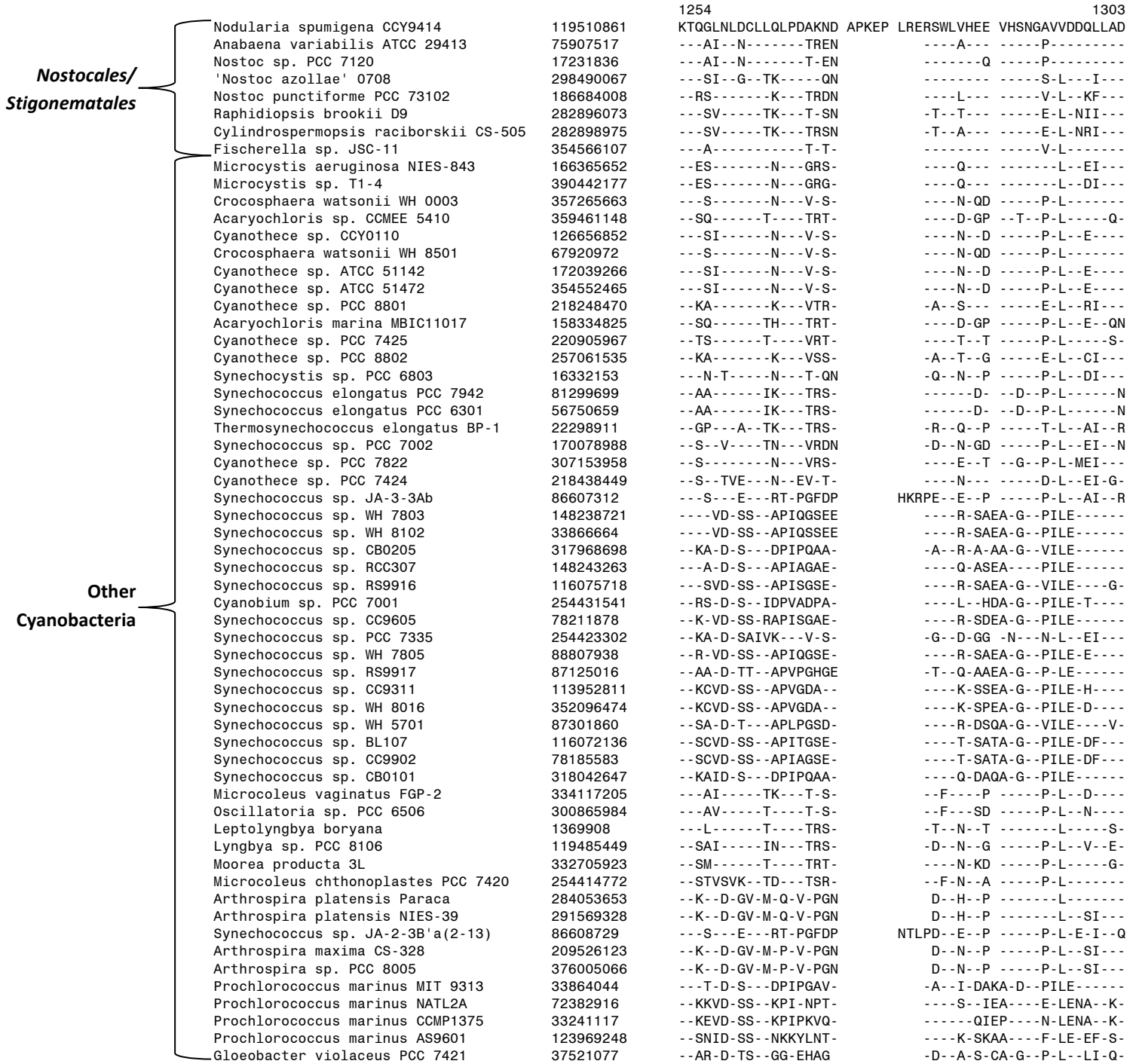
Partial sequence alignment for the protein phenylalanyl-tRNA synthetase, showing a 3 aa insert that is specific for *Nodularia spumigena CCY9414.*



		211	254
Nodularia spumigena CCY9414	119512824	AIHGATVENTLFEDGEGS S	NTFPAFNPTQAEETYSMVTANRFWS
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----------
Crocosphaera watsonii WH 8501	67925969	----------QA	---R--E----------
Cyanothece sp. PCC 7425	220905664	----------T-	---R----------
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

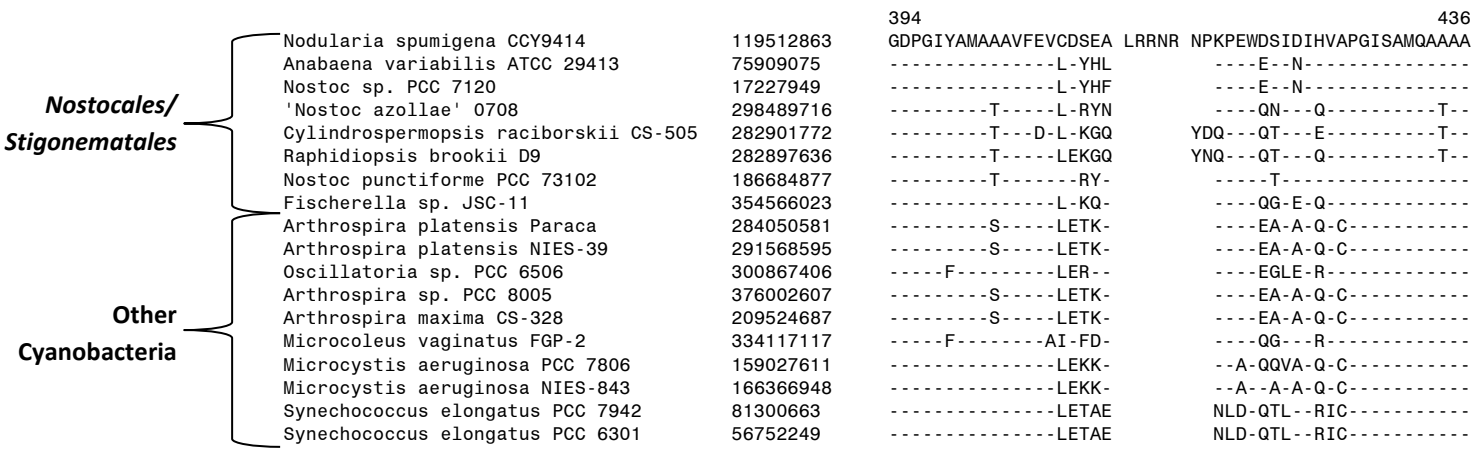
Partial sequence alignment for the protein photosystem II protein D2, showing a 1 aa insert that is specific for *Nodularia spumigena CCY9414.*



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

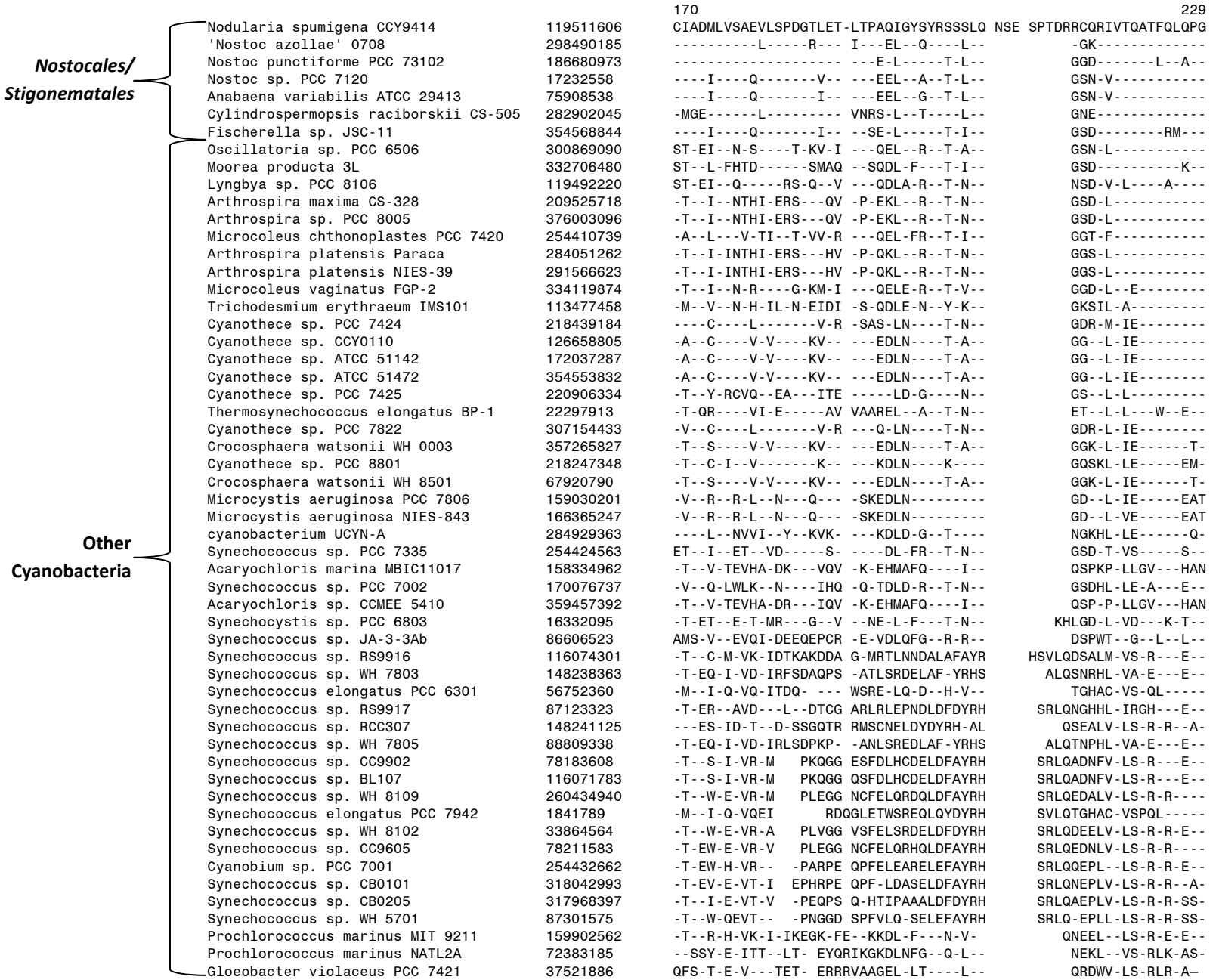
Supplementary Figure 116

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



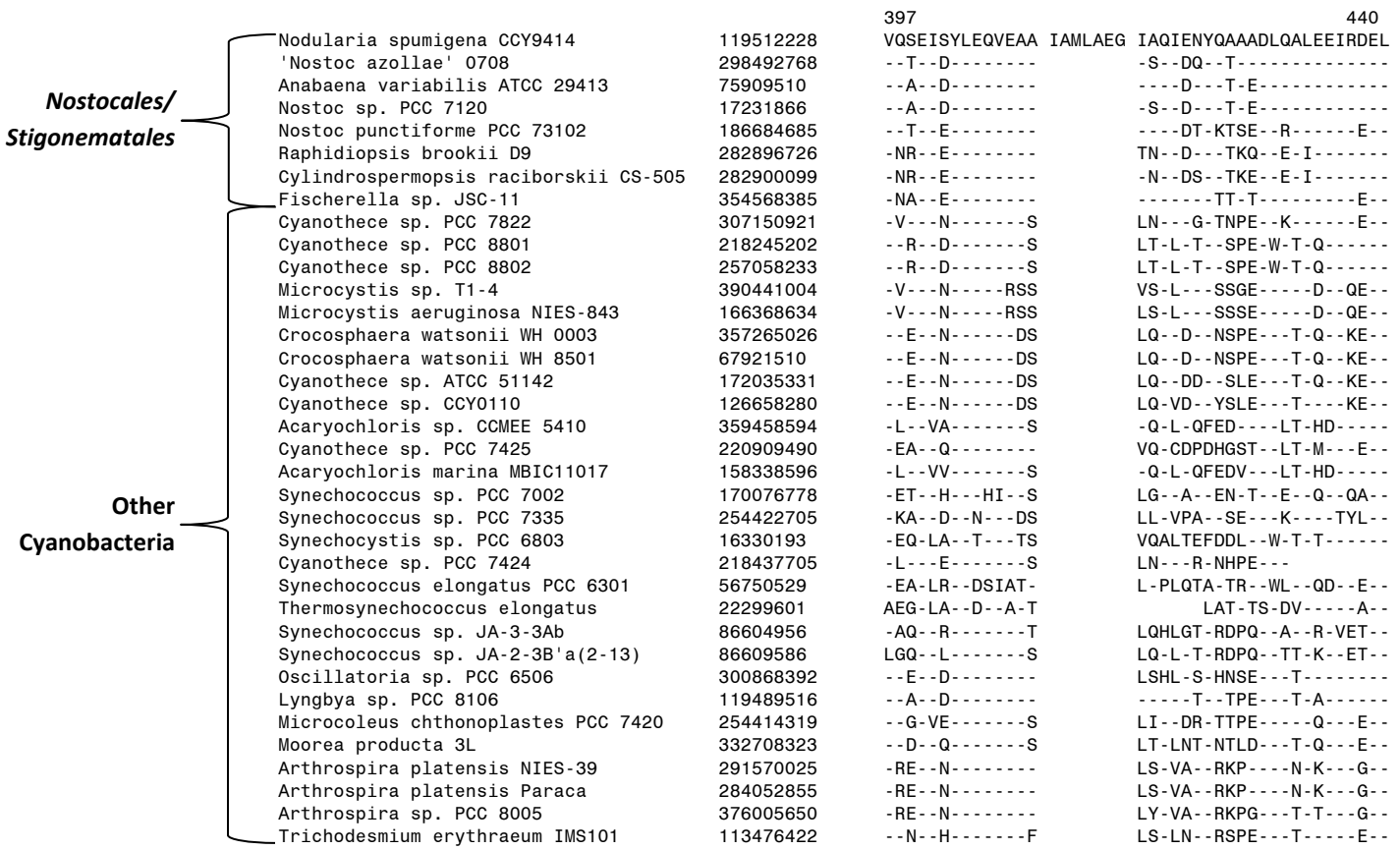
Supplementary Figure 117

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



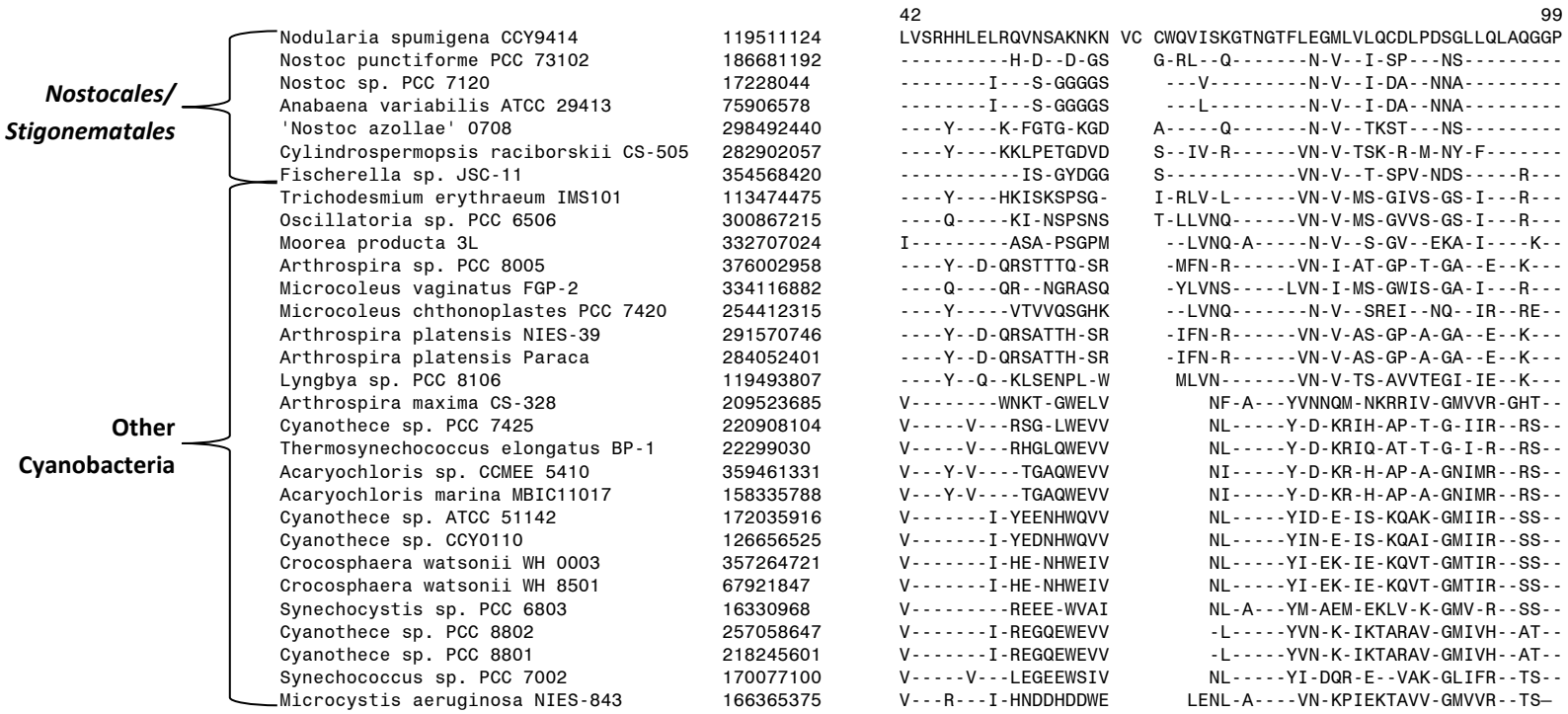
Supplementary Figure 118

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



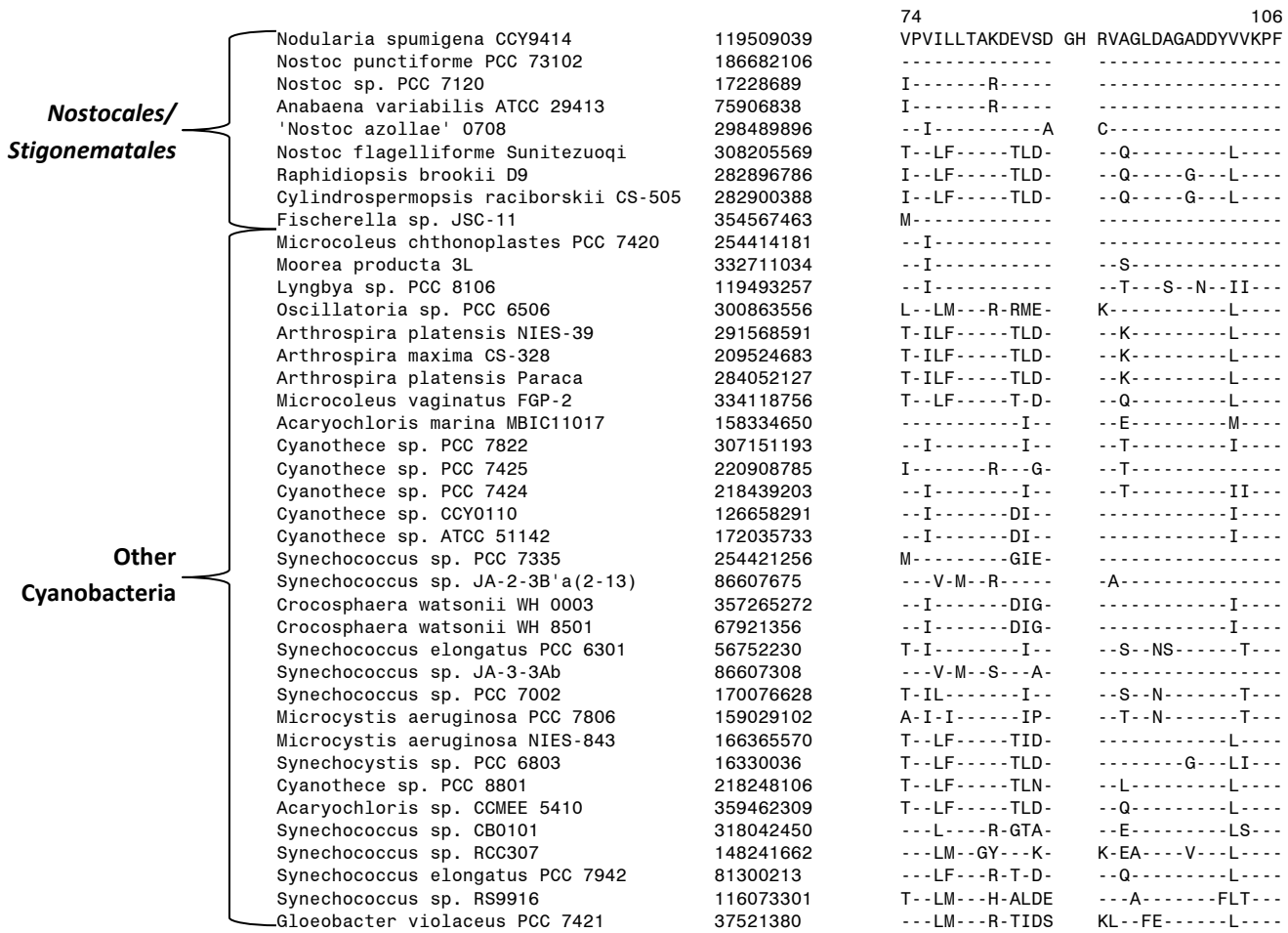
Supplementary Figure 119

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



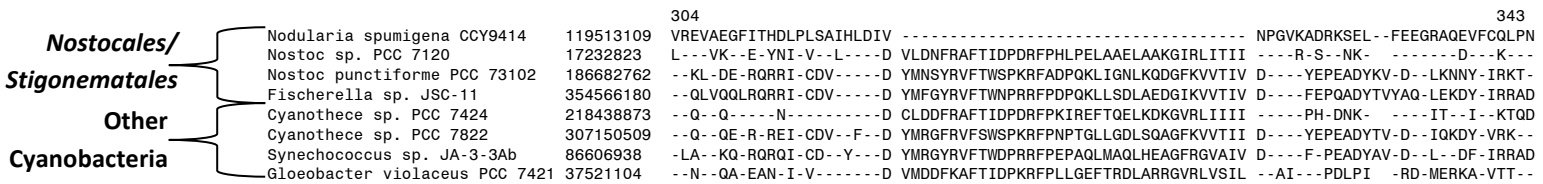
Supplementary Figure 120

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



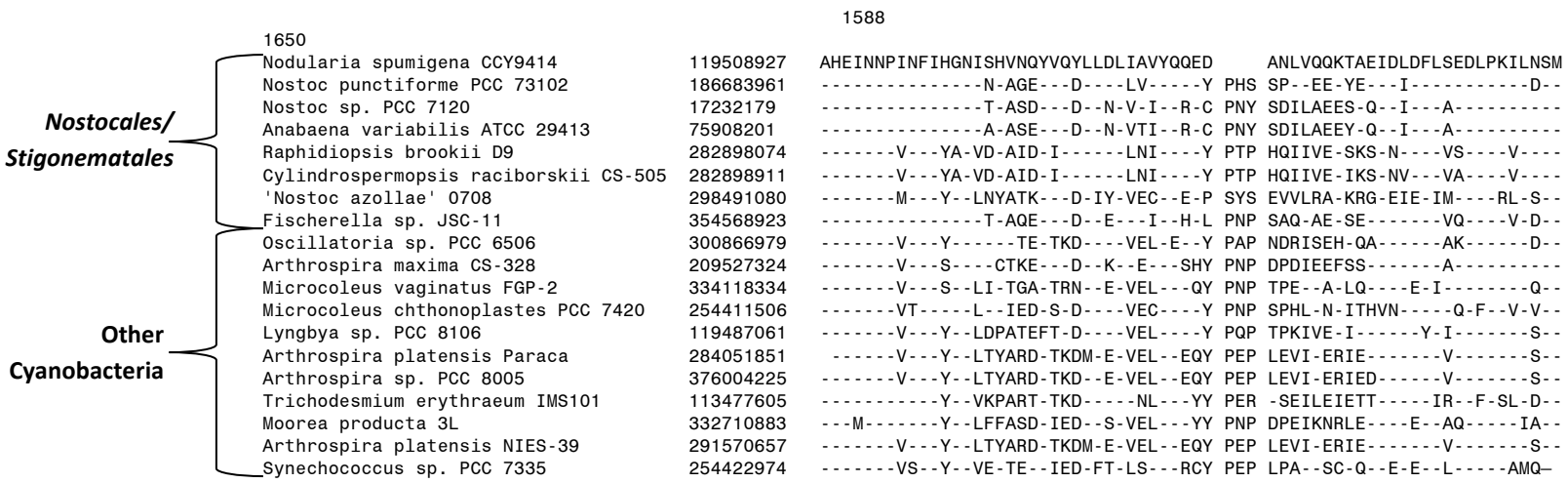
Supplementary Figure 121

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



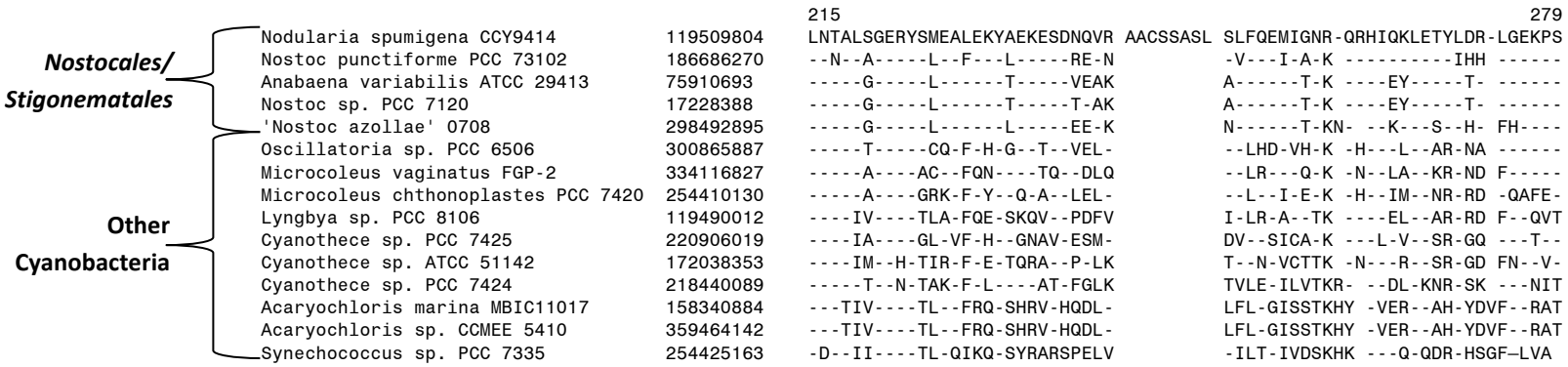
Supplementary Figure 122

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



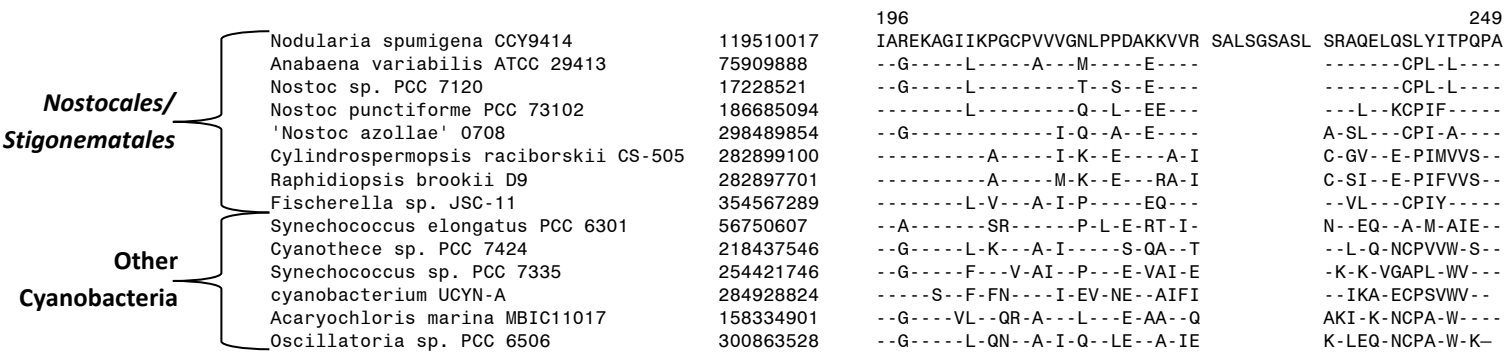
Supplementary Figure 123

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



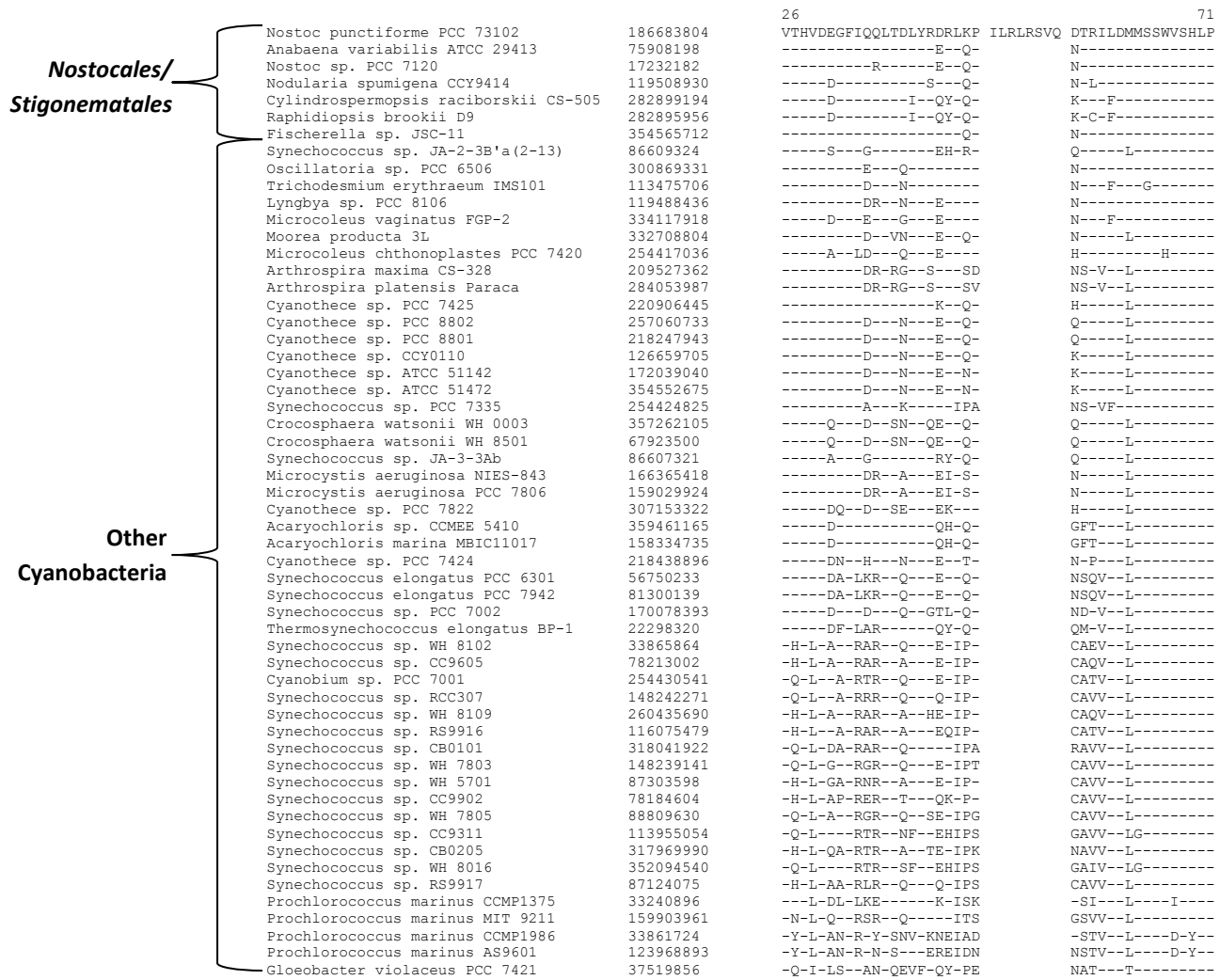
Supplementary Figure 124

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



Supplementary Figure 125

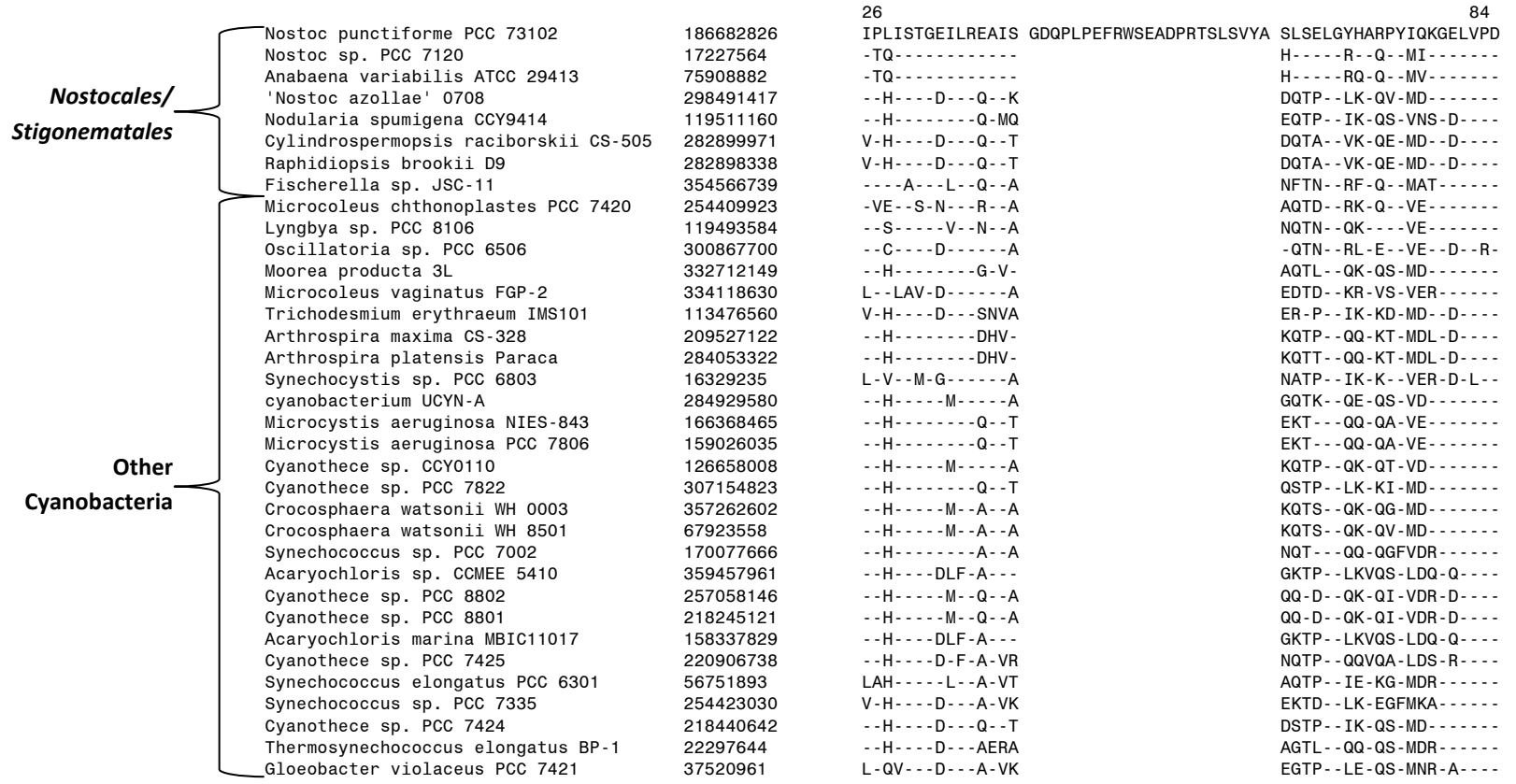
Partial sequence alignment for the protein folylpolyglutamate synthase, showing a 9 aa insert that is specific for *Nodularia spumigena CCY9414.*



		26	71
Nostoc punctiforme PCC 73102	186683804	VTHVDEGFIQQLTDLYRDRLKP	ILRLRSVQ DTRILDMMSSWVSHLP
Anabaena variabilis ATCC 29413	75908198	----------E--Q-	N----------
Nostoc sp. PCC 7120	17232182	----------R------E--Q-	N----------
Nodularia spumigena CCY9414	119508930	-----D----------S---Q-	N-L---------
Cylindrospermopsis raciborskii CS-505	282899194	-----D--------I--QY-Q-	K---F-------
Raphidiopsis brookii D9	282895956	-----D--------I--QY-Q-	K-C-F-------
Fischerella sp. JSC-11	354565712	----------Q-	N----------
Synechococcus sp. JA-2-3B'a(2-13)	86609324	-----S---G-------EH-R-	Q-----L---------
Oscillatoria sp. PCC 6506	300869331	---------E---Q--------	N----------
Trichodesmium erythraeum IMS101	113475706	---------D---N--------	N---F---G-------
Lyngbya sp. PCC 8106	119488436	---------DR--N---E----	N----------
Microcoleus vaginatus FGP-2	334117918	-----D---E---G---E----	N---F-------
Moorea producta 3L	332708804	---------D--VN---E--Q-	N-----L---------
Microcoleus chthonoplastes PCC 7420	254417036	-----A--LD---Q---E----	H---------H-----
Arthrospira maxima CS-328	209527362	---------DR-RG--S---SD	NS-V--L---------
Arthrospira platensis Paraca	284053987	---------DR-RG--S---SV	NS-V--L---------
Cyanothece sp. PCC 7425	220906445	----------K--Q-	H-----L---------
Cyanothece sp. PCC 8802	257060733	---------D---N---E--Q-	Q-----L---------
Cyanothece sp. PCC 8801	218247943	---------D---N---E--Q-	Q-----L---------
Cyanothece sp. CCY0110	126659705	---------D---N---E--Q-	K-----L---------
Cyanothece sp. ATCC 51142	172039040	---------D---N---E--N-	K-----L---------
Cyanothece sp. ATCC 51472	354552675	---------D---N---E--N-	K-----L---------
Synechococcus sp. PCC 7335	254424825	---------A---K-----IPA	NS-VF----------
Crocosphaera watsonii WH 0003	357262105	-----Q---D--SN--QE--Q-	Q-----L---------
Crocosphaera watsonii WH 8501	67923500	-----Q---D--SN--QE--Q-	Q-----L---------
Synechococcus sp. JA-3-3Ab	86607321	-----A---G-------RY-Q-	Q-----L---------
Microcystis aeruginosa NIES-843	166365418	---------DR--A---EI-S-	N-----L---------
Microcystis aeruginosa PCC 7806	159029924	---------DR--A---EI-S-	N-----L---------
Cyanothece sp. PCC 7822	307153322	-----DQ--D--SE---EK---	H-----L---------
Acaryochloris sp. CCMEE 5410	359461165	-----D----------QH-Q-	GFT---L---------
Acaryochloris marina MBIC11017	158334735	-----D----------QH-Q-	GFT---L---------
Cyanothece sp. PCC 7424	218438896	-----DN--H---N---E--T-	N-P---L---------
Synechococcus elongatus PCC 6301	56750233	-----DA-LKR--Q---E--Q-	NSQV--L---------
Synechococcus elongatus PCC 7942	81300139	-----DA-LKR--Q---E--Q-	NSQV--L---------
Synechococcus sp. PCC 7002	170078393	-----D---D---Q--GTL-Q-	ND-V--L---------
Thermosynechococcus elongatus BP-1	22298320	-----DF-LAR------QY-Q-	QM-V--L---------
Synechococcus sp. WH 8102	33865864	-H-L-A--RAR--Q---E-IP-	CAEV--L---------
Synechococcus sp. CC9605	78213002	-H-L-A--RAR--A---E-IP-	CAQV--L---------
Cyanobium sp. PCC 7001	254430541	-Q-L--A-RTR--Q---E-IP-	CATV--L---------
Synechococcus sp. RCC307	148242271	-Q-L--A-RRR--Q---Q-IP-	CAVV--L---------
Synechococcus sp. WH 8109	260435690	-H-L-A--RAR--A--HE-IP-	CAQV--L---------
Synechococcus sp. RS9916	116075479	-H-L--A-RAR--A---EQIP-	CATV--L---------
Synechococcus sp. CB0101	318041922	-Q-L-DA-RAR--Q-----IPA	RAVV--L---------
Synechococcus sp. WH 7803	148239141	-Q-L-G--RGR--Q---E-IPT	CAVV--L---------
Synechococcus sp. WH 5701	87303598	-H-L-GA-RNR--A---E-IP-	CAVV--L---------
Synechococcus sp. CC9902	78184604	-H-L-AP-RER--T---QK-P-	CAVV--L---------
Synechococcus sp. WH 7805	88809630	-Q-L-A--RGR--Q--SE-IPG	CAVV--L---------
Synechococcus sp. CC9311	113955054	-Q-L----RTR--NF--EHIPS	GAVV--LG--------
Synechococcus sp. CB0205	317969990	-H-L-QA-RTR--A--TE-IPK	NAVV--L---------
Synechococcus sp. WH 8016	352094540	-Q-L----RTR--SF--EHIPS	GAIV--LG--------
Synechococcus sp. RS9917	87124075	-H-L-AA-RLR--Q---Q-IPS	CAVV--L---------
Prochlorococcus marinus CCMP1375	33240896	---L-DL-LKE------K-ISK	-SI---L----I----
Prochlorococcus marinus MIT 9211	159903961	-N-L-Q--RSR--Q-----ITS	GSVV--L---------
Prochlorococcus marinus CCMP1986	33861724	-Y-L-AN-R-Y-SNV-KNEIAD	-STV--L----D-Y--
Prochlorococcus marinus AS9601	123968893	-Y-L-AN-R-N-S---EREIDN	NSTV--L----D-Y--
Gloeobacter violaceus PCC 7421	37519856	-Q-I-LS--AN-QEVF-QY-PE	NAT---T---------

Supplementary Figure 126

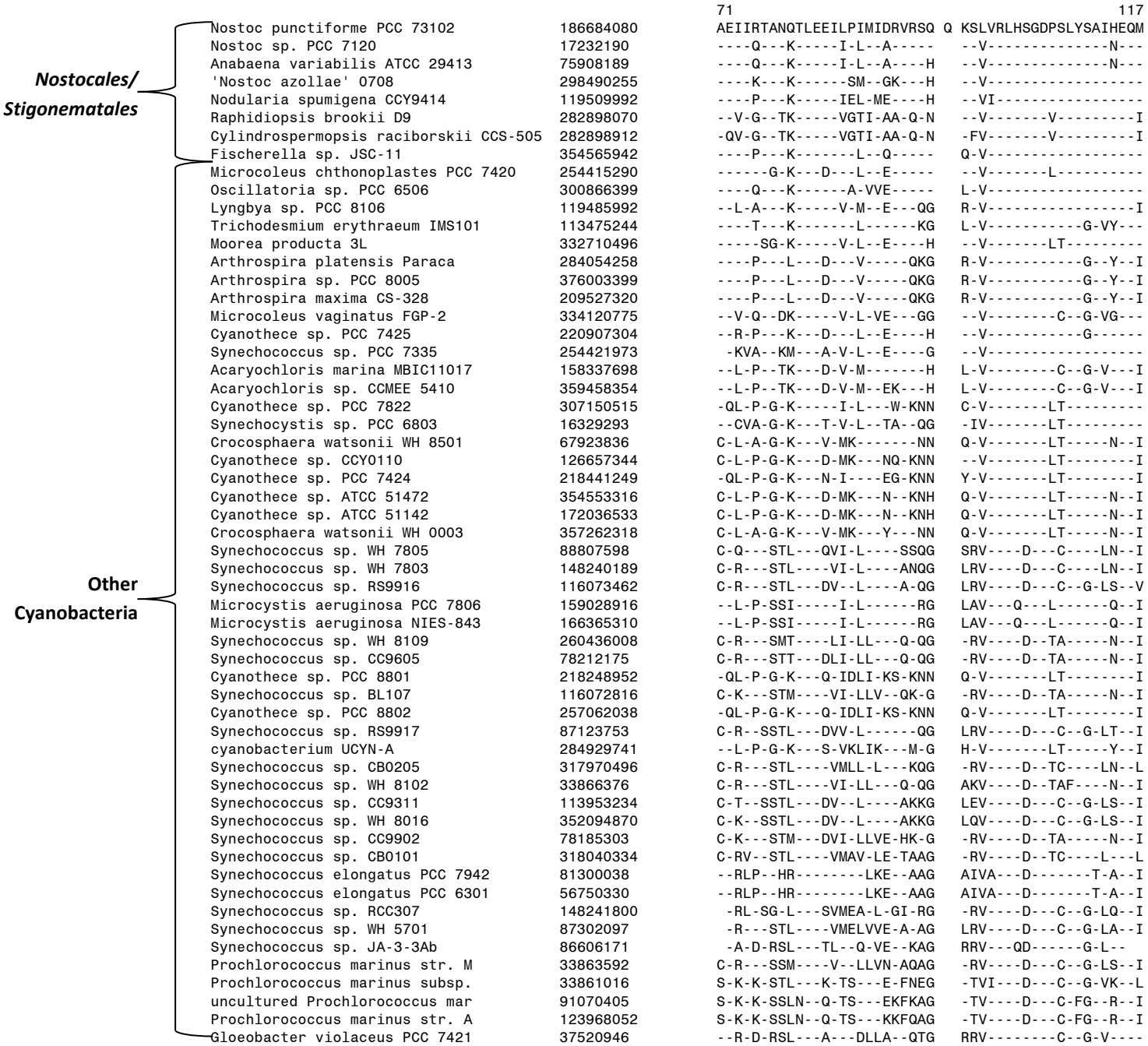
Partial sequence alignment for the protein methyltransferase type 11, showing an 8 aa insert that is specific for *Nostoc punctiforme PCC 73102*.



		26	84
Nostoc punctiforme PCC 73102	186682826	IPLISTGEILREAIS GDQPLPEFRWSEADPRTSLSVYA	SLSELGYHARPYIQKGELVPD
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--------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--RK-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----
Synechocystis sp. PCC 6803	16329235	L-V--M-G------A	NATP--IK-K--VER-D-L--
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-------
Crocosphaera watsonii WH 0003	357262602	--H-----M--A--A	KQTS--QK-QG-MD-------
Crocosphaera 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. CCMEE 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----

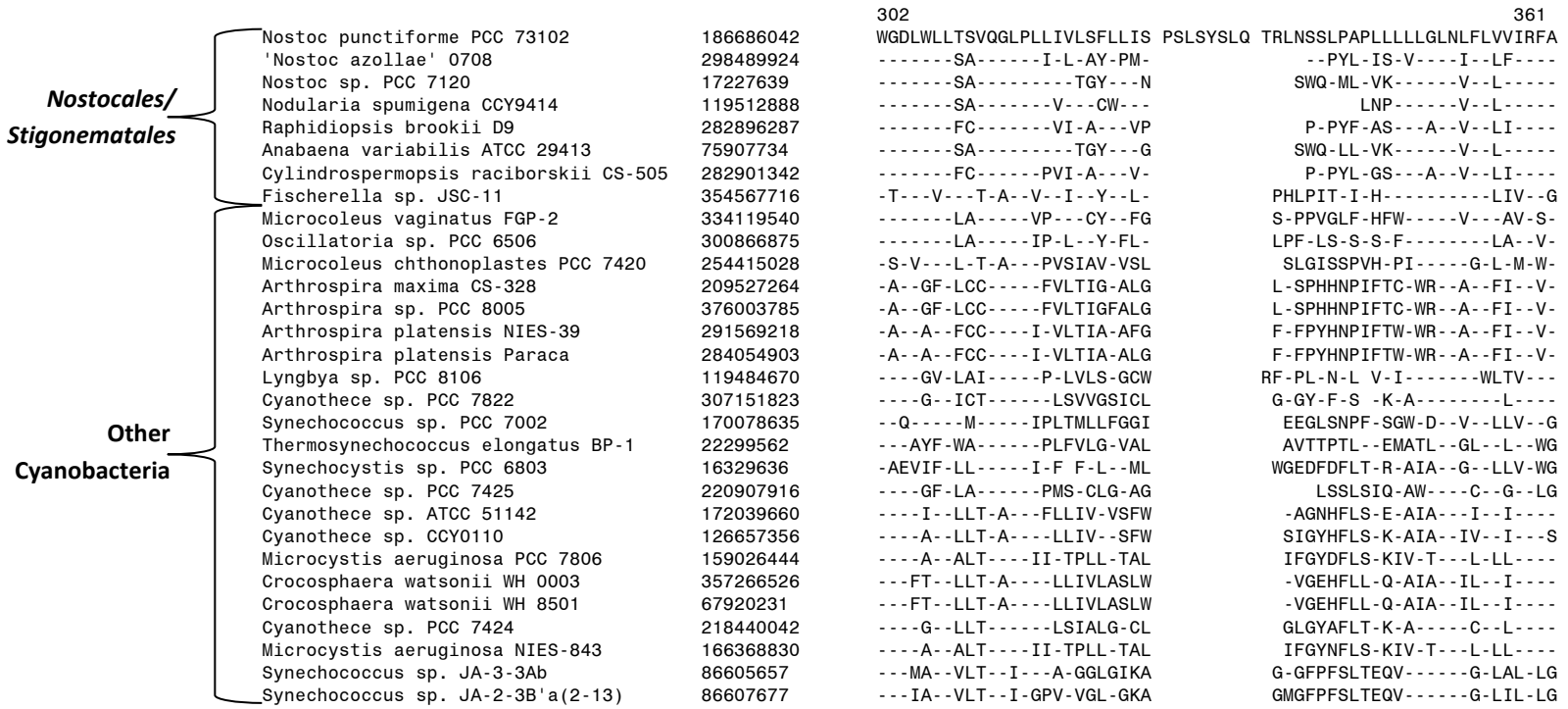
Supplementary Figure 127

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



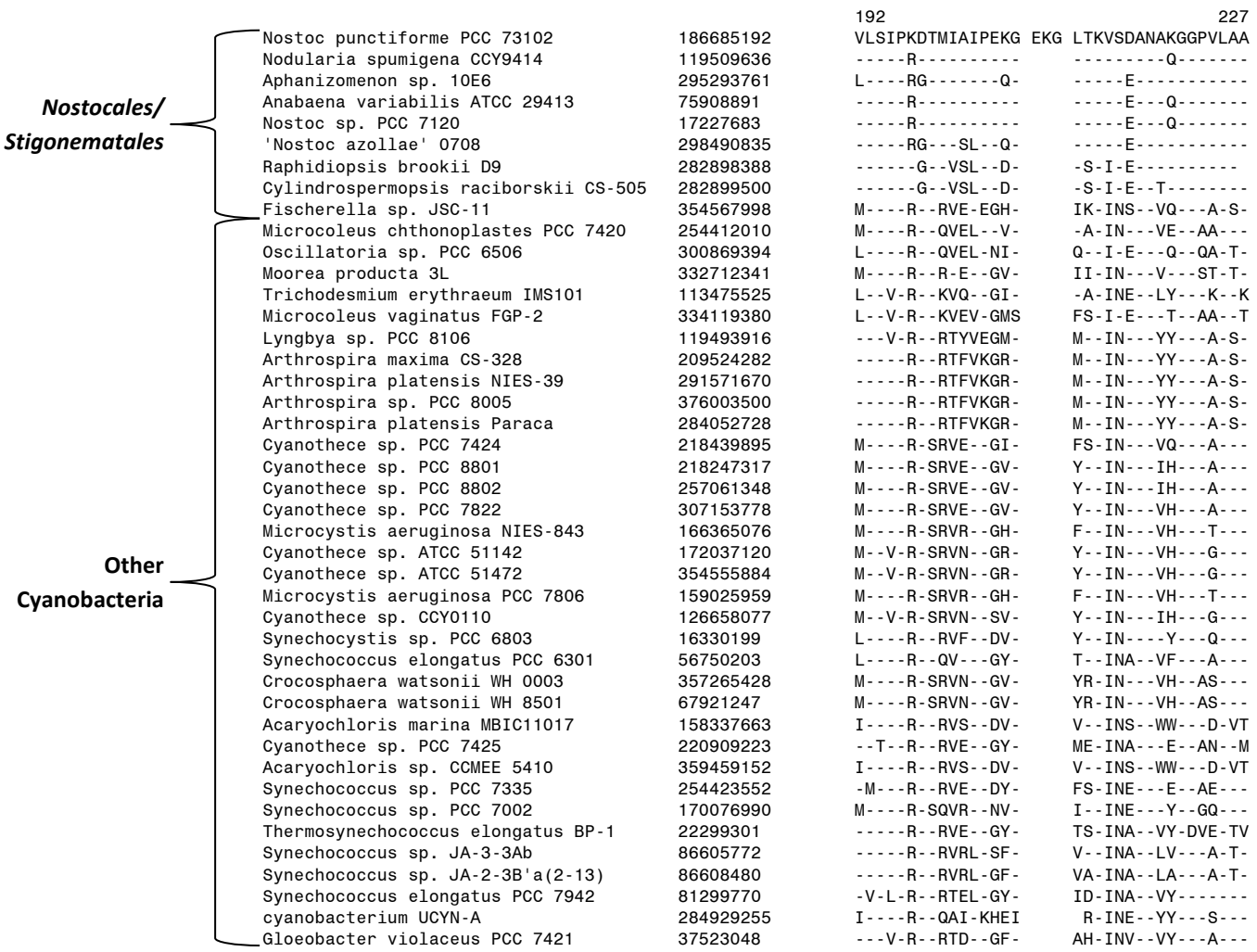
Supplementary Figure 128

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



Supplementary Figure 129

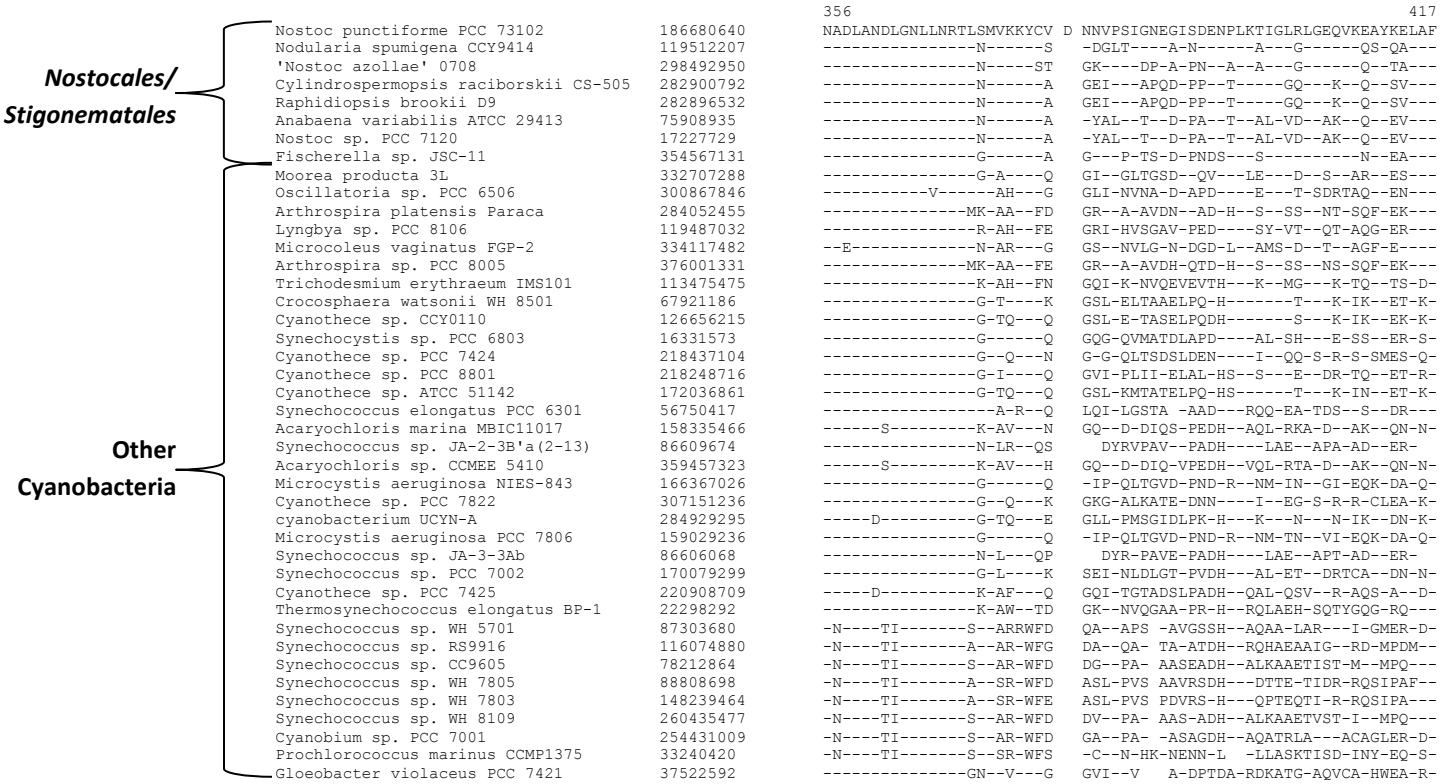
Partial sequence alignment for a glycosyl transferase family protein, showing an 8 aa insert that is specific for *Nostoc punctiforme PCC 73102.*



		192	227
Nostoc punctiforme PCC 73102	186685192	VLSIPKDTMIAIPEKG EKG	LTKVSDANAKGGPVLAA
Nodularia spumigena CCY9414	119509636	-----R----------	---------Q-------
Aphanizomenon sp. 10E6	295293761	L----RG-------Q-	-----E----------
Anabaena variabilis ATCC 29413	75908891	-----R----------	-----E---Q-------
Nostoc sp. PCC 7120	17227683	-----R----------	-----E---Q-------
'Nostoc azollae' 0708	298490835	-----RG---SL--Q-	-----E----------
Raphidiopsis brookii D9	282898388	------G--VSL--D-	-S-I-E----------
Cylindrospermopsis raciborskii CS-505	282899500	------G--VSL--D-	-S-I-E--T--------
Fischerella sp. JSC-11	354567998	M----R--RVE-EGH-	IK-INS--VQ---A-S-
Microcoleus chthonoplastes PCC 7420	254412010	M----R--QVEL--V-	-A-IN---VE--AA---
Oscillatoria sp. PCC 6506	300869394	L----R--QVEL-NI-	Q--I-E---Q--QA-T-
Moorea producta 3L	332712341	M----R--R-E--GV-	II-IN---V---ST-T-
Trichodesmium erythraeum IMS101	113475525	L--V-R--KVQ--GI-	-A-INE--LY---K--K
Microcoleus vaginatus FGP-2	334119380	L--V-R--KVEV-GMS	FS-I-E---T--AA--T
Lyngbya sp. PCC 8106	119493916	---V-R--RTYVEGM-	M--IN---YY---A-S-
Arthrospira maxima CS-328	209524282	-----R--RTFVKGR-	M--IN---YY---A-S-
Arthrospira platensis NIES-39	291571670	-----R--RTFVKGR-	M--IN---YY---A-S-
Arthrospira sp. PCC 8005	376003500	-----R--RTFVKGR-	M--IN---YY---A-S-
Arthrospira platensis Paraca	284052728	-----R--RTFVKGR-	M--IN---YY---A-S-
Cyanothece sp. PCC 7424	218439895	M----R-SRVE--GI-	FS-IN---VQ---A---
Cyanothece sp. PCC 8801	218247317	M----R-SRVE--GV-	Y--IN---IH---A---
Cyanothece sp. PCC 8802	257061348	M----R-SRVE--GV-	Y--IN---IH---A---
Cyanothece sp. PCC 7822	307153778	M----R-SRVE--GV-	Y--IN---VH---A---
Microcystis aeruginosa NIES-843	166365076	M----R-SRVR--GH-	F--IN---VH---T---
Cyanothece sp. ATCC 51142	172037120	M--V-R-SRVN--GR-	Y--IN---VH---G---
Cyanothece sp. ATCC 51472	354555884	M--V-R-SRVN--GR-	Y--IN---VH---G---
Microcystis aeruginosa PCC 7806	159025959	M----R-SRVR--GH-	F--IN---VH---T---
Cyanothece sp. CCY0110	126658077	M--V-R-SRVN--SV-	Y--IN---IH---G---
Synechocystis sp. PCC 6803	16330199	L----R--RVF--DV-	Y--IN----Y---Q---
Synechococcus elongatus PCC 6301	56750203	L----R--QV---GY-	T--INA--VF---A---
Crocosphaera watsonii WH 0003	357265428	M----R-SRVN--GV-	YR-IN---VH--AS---
Crocosphaera watsonii WH 8501	67921247	M----R-SRVN--GV-	YR-IN---VH--AS---
Acaryochloris marina MBIC11017	158337663	I----R--RVS--DV-	V--INS--WW---D-VT
Cyanothece sp. PCC 7425	220909223	--T--R--RVE--GY-	ME-INA---E--AN--M
Acaryochloris sp. CCMEE 5410	359459152	I----R--RVS--DV-	V--INS--WW---D-VT
Synechococcus sp. PCC 7335	254423552	-M---R--RVE--DY-	FS-INE---E--AE---
Synechococcus sp. PCC 7002	170076990	M----R-SQVR--NV-	I--INE---Y--GQ---
Thermosynechococcus elongatus BP-1	22299301	-----R--RVE--GY-	TS-INA--VY-DVE-TV
Synechococcus sp. JA-3-3Ab	86605772	-----R--RVRL-SF-	V--INA--LV---A-T-
Synechococcus sp. JA-2-3B'a(2-13)	86608480	-----R--RVRL-GF-	VA-INA--LA---A-T-
Synechococcus elongatus PCC 7942	81299770	-V-L-R--RTEL-GY-	ID-INA--VY-------
cyanobacterium UCYN-A	284929255	I----R--QAI-KHEI	R-INE--YY---S---
Gloeobacter violaceus PCC 7421	37523048	---V-R--RTD--GF-	AH-INV--VY---A---

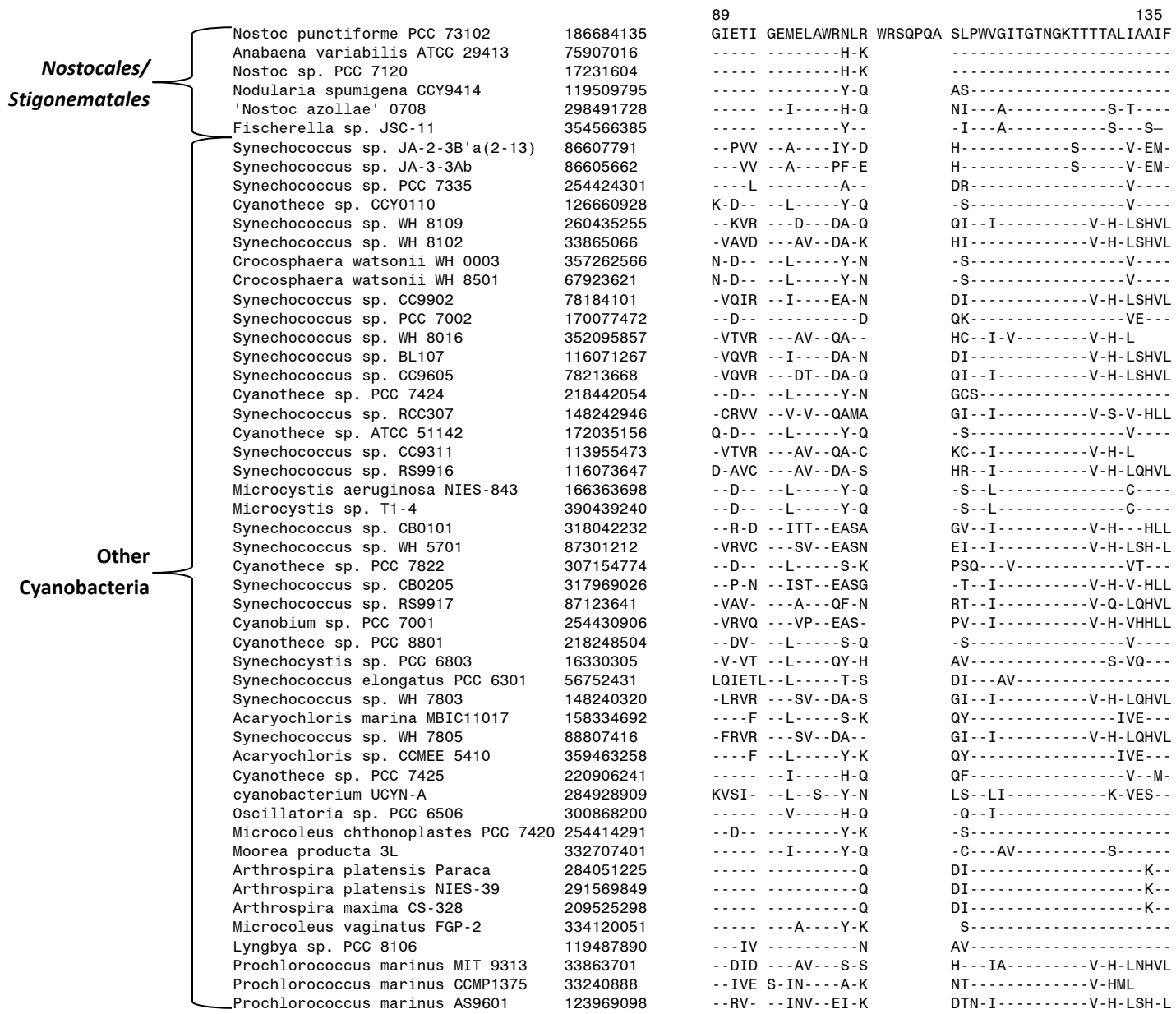
Supplementary Figure 130

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



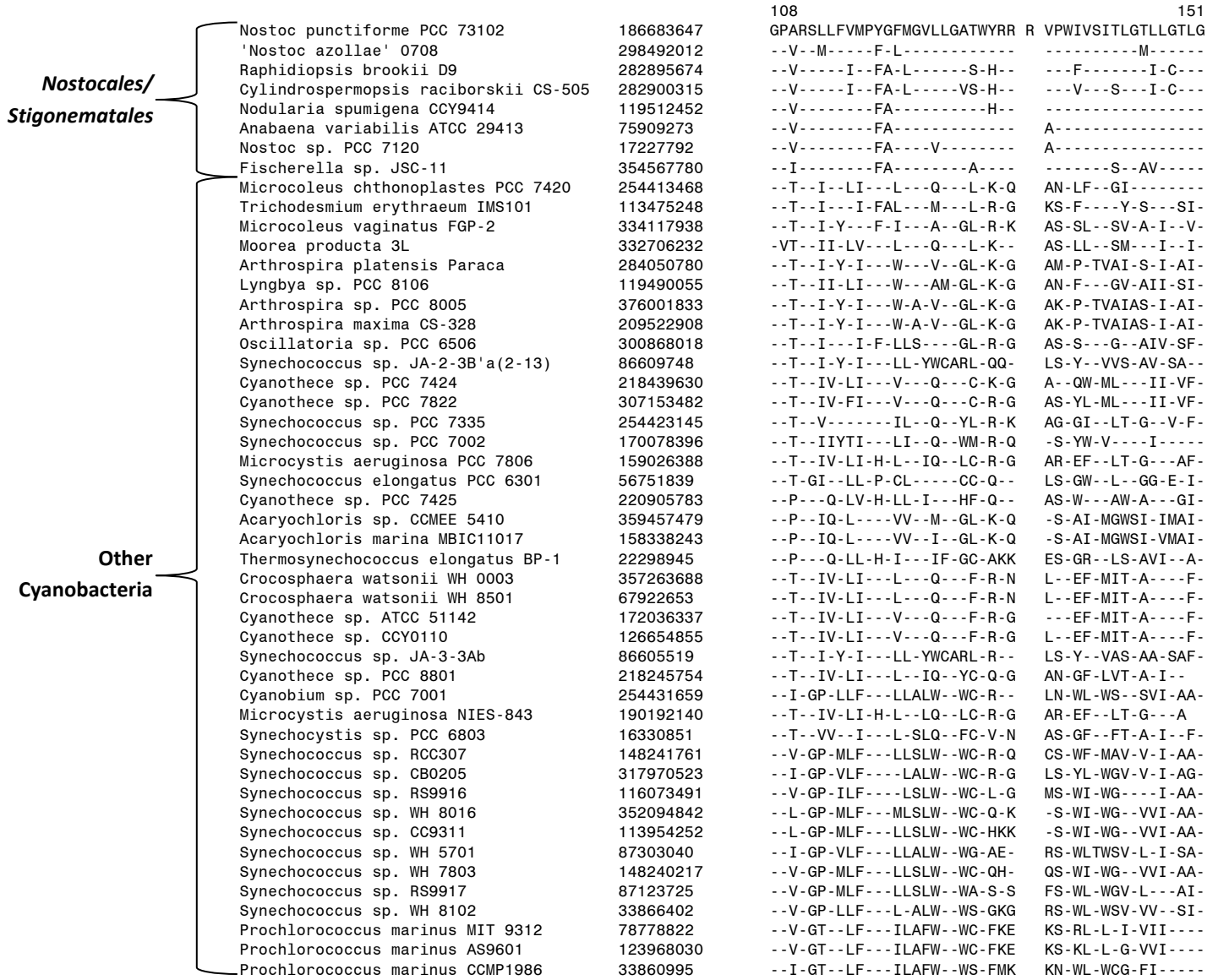
Supplementary Figure 131

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



Supplementary Figure 132

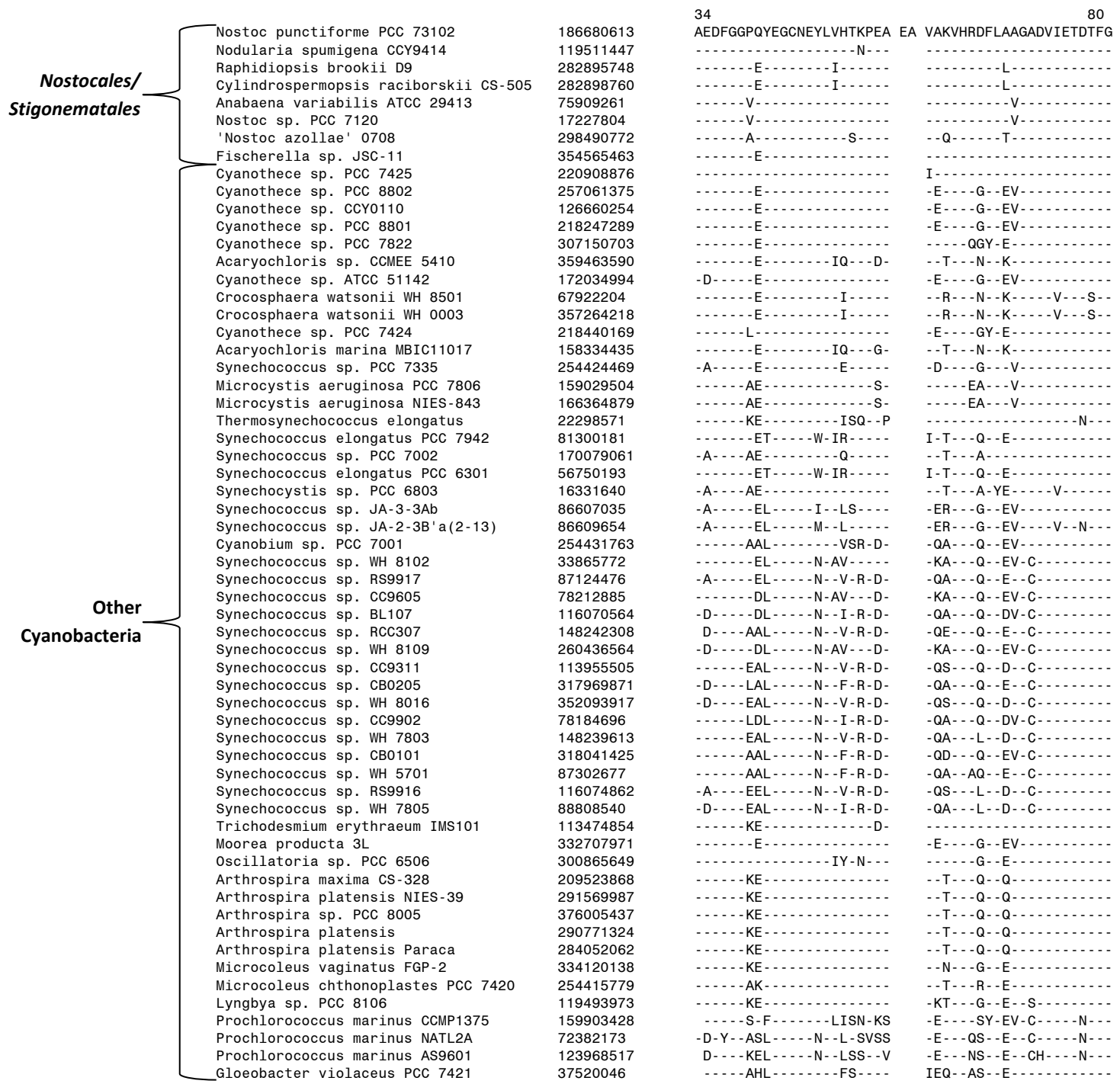
Partial sequence alignment for the protein UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase, showing a 7 aa insert that is specific for *Nostoc punctiforme PCC 73102.*



		108	151
Nostoc punctiforme PCC 73102	186683647	GPARSLLFVMPYGFMGVLLGATWYRR R	VPWIVSITLGTLLGTLG
'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-Q	AN-LF--GI--------
Trichodesmium erythraeum IMS101	113475248	--T--I---I-FAL---M---L-R-G	KS-F----Y-S---SI-
Microcoleus vaginatus FGP-2	334117938	--T--I-Y---F-I---A--GL-R-K	AS-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-G	AM-P-TVAI-S-I-AI-
Lyngbya sp. PCC 8106	119490055	--T--II-LI---W---AM-GL-K-G	AN-F---GV-AII-SI-
Arthrospira sp. PCC 8005	376001833	--T--I-Y-I---W-A-V--GL-K-G	AK-P-TVAIAS-I-AI-
Arthrospira maxima CS-328	209522908	--T--I-Y-I---W-A-V--GL-K-G	AK-P-TVAIAS-I-AI-
Oscillatoria sp. PCC 6506	300868018	--T--I---I-F-LLS----GL-R-G	AS-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-G	A--QW-ML---II-VF-
Cyanothece sp. PCC 7822	307153482	--T--IV-FI---V---Q---C-R-G	AS-YL-ML---II-VF-
Synechococcus sp. PCC 7335	254423145	--T--V-------IL--Q--YL-R-K	AG-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-G	AR-EF--LT-G---AF-
Synechococcus elongatus PCC 6301	56751839	--T-GI--LL-P-CL-----CC-Q--	LS-GW--L--GG-E-I-
Cyanothece sp. PCC 7425	220905783	--P---Q-LV-H-LL-I---HF-Q--	AS-W---AW-A---GI-
Acaryochloris sp. CCMEE 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-AKK	ES-GR--LS-AVI--A-
Crocosphaera watsonii WH 0003	357263688	--T--IV-LI---L---Q---F-R-N	L--EF-MIT-A----F-
Crocosphaera watsonii WH 8501	67922653	--T--IV-LI---L---Q---F-R-N	L--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-G	L--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-G	AN-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-G	AR-EF--LT-G---A
Synechocystis sp. PCC 6803	16330851	--T--VV--I---L-SLQ--FC-V-N	AS-GF--FT-A-I--F-
Synechococcus sp. RCC307	148241761	--V-GP-MLF---LLSLW--WC-R-Q	CS-WF-MAV-V-I-AA-
Synechococcus sp. CB0205	317970523	--I-GP-VLF----LALW--WC-R-G	LS-YL-WGV-V-I-AG-
Synechococcus sp. RS9916	116073491	--V-GP-ILF----LSLW--WC-L-G	MS-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-HKK	-S-WI-WG--VVI-AA-
Synechococcus sp. WH 5701	87303040	--I-GP-VLF---LLALW--WG-AE-	RS-WLTWSV-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-S	FS-WL-WGV-L---AI-
Synechococcus sp. WH 8102	33866402	--V-GP-LLF---L-ALW--WS-GKG	RS-WL-WSV-VV--SI-
Prochlorococcus marinus MIT 9312	78778822	--V-GT--LF---ILAFW--WC-FKE	KS-RL-L-I-VII----
Prochlorococcus marinus AS9601	123968030	--V-GT--LF---ILAFW--WC-FKE	KS-KL-L-G-VVI----
Prochlorococcus marinus CCMP1986	33860995	--I-GT--LF---ILAFW--WS-FMK	KN-WL-WCG-FI-----

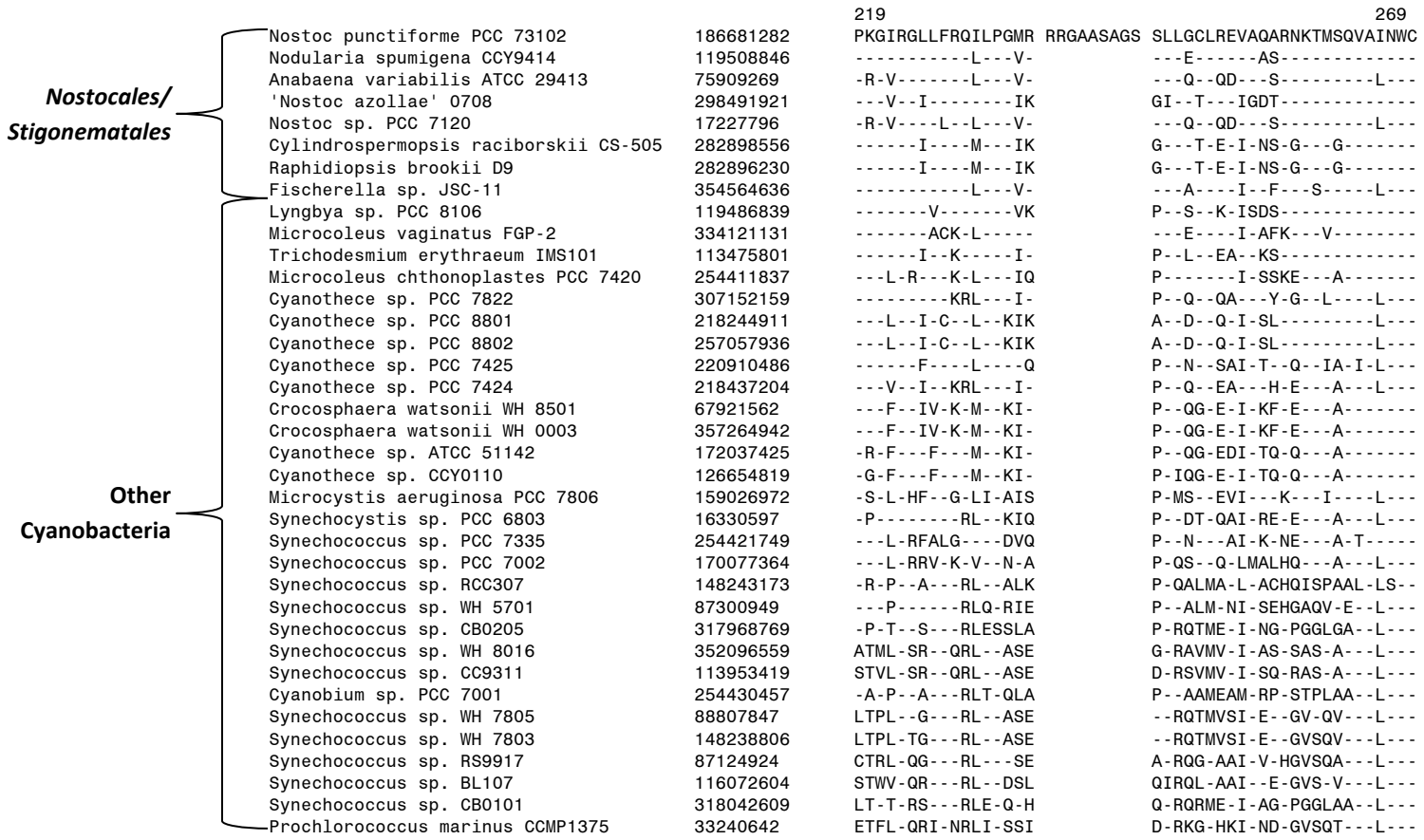
Supplementary Figure 133

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



Supplementary Figure 134

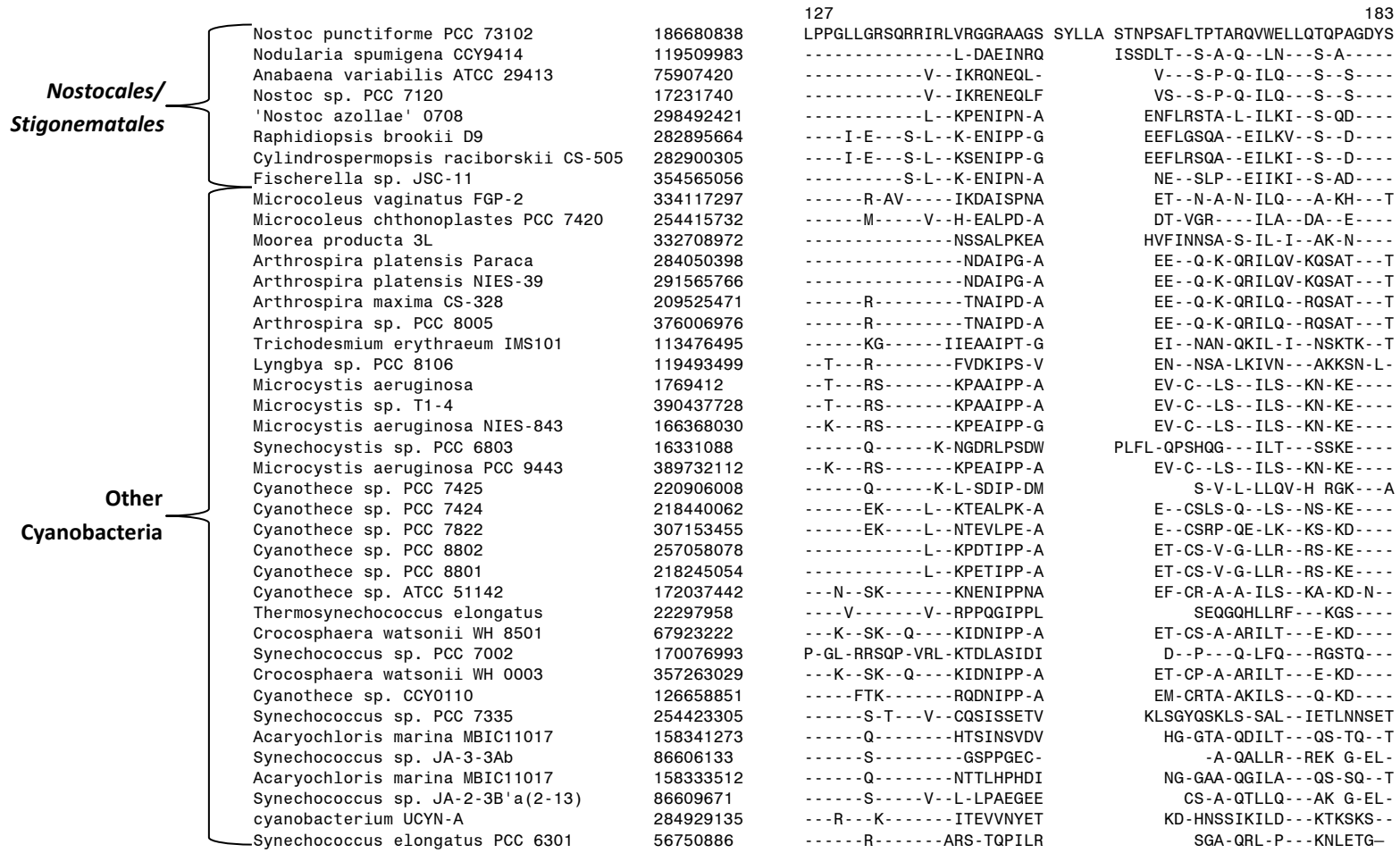
Partial sequence alignment for the protein methionine synthase, showing a 2 aa insert that is specific for *Nostoc punctiforme PCC 73102.*



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

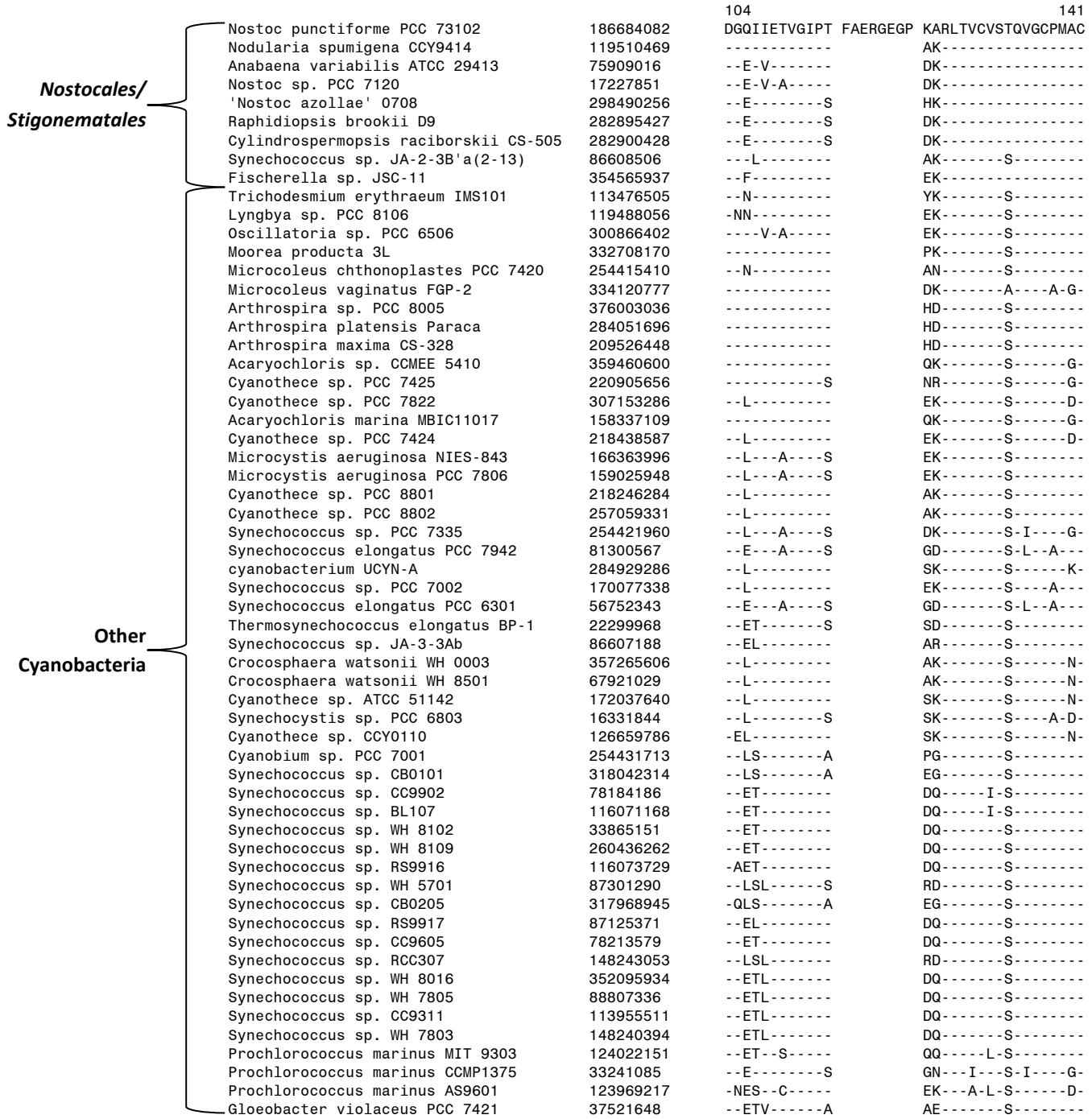
Supplementary Figure 135

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



Supplementary Figure 136

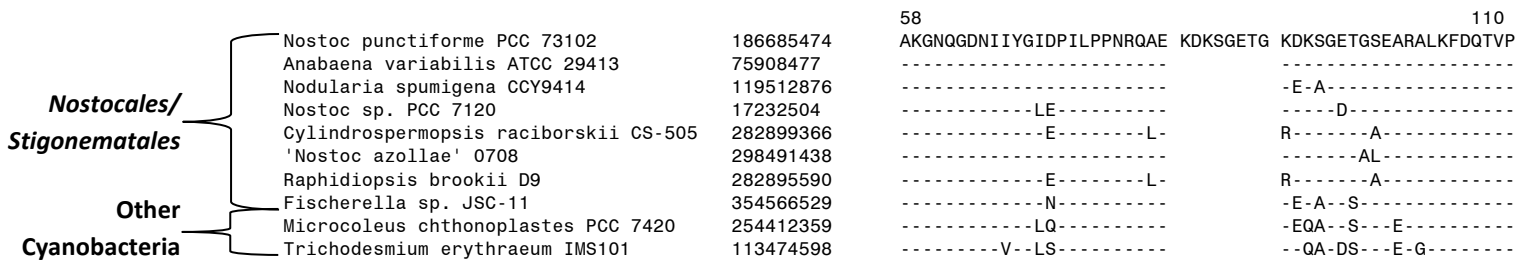
Partial sequence alignment for the primosome assembly protein, showing a 5 aa insert that is specific for *Nostoc punctiforme PCC 73102.*



		104	141
Nostoc punctiforme PCC 73102	186684082	DGQIIETVGIPT FAERGEGP	KARLTVCVSTQVGCPMAC
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--------
Lyngbya 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	-----------S	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-------S	SD-------S--------
Synechococcus sp. JA-3-3Ab	86607188	--EL--------	AR-------S--------
Crocosphaera watsonii WH 0003	357265606	--L---------	AK-------S------N-
Crocosphaera watsonii WH 8501	67921029	--L---------	AK-------S------N-
Cyanothece sp. ATCC 51142	172037640	--L---------	SK-------S------N-
Synechocystis sp. PCC 6803	16331844	--L--------S	SK-------S----A-D-
Cyanothece sp. CCY0110	126659786	-EL---------	SK-------S------N-
Cyanobium sp. PCC 7001	254431713	--LS-------A	PG-------S--------
Synechococcus sp. CB0101	318042314	--LS-------A	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------S	RD-------S--------
Synechococcus sp. CB0205	317968945	-QLS-------A	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--------S	GN---I---S-I----G-
Prochlorococcus marinus AS9601	123969217	-NES--C-----	EK---A-L-S------D-
Gloeobacter violaceus PCC 7421	37521648	--ETV------A	AE-------S--------

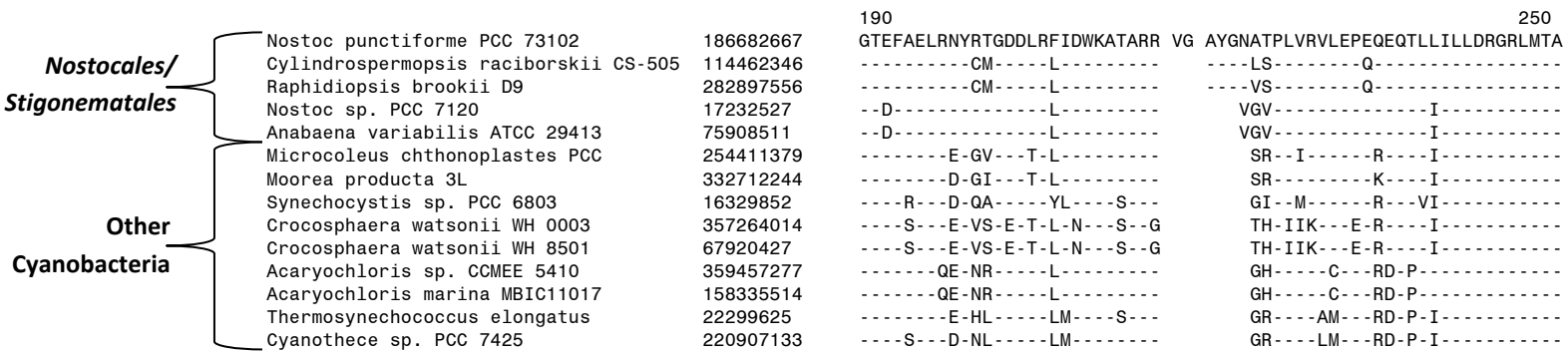
Supplementary Figure 137

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.*



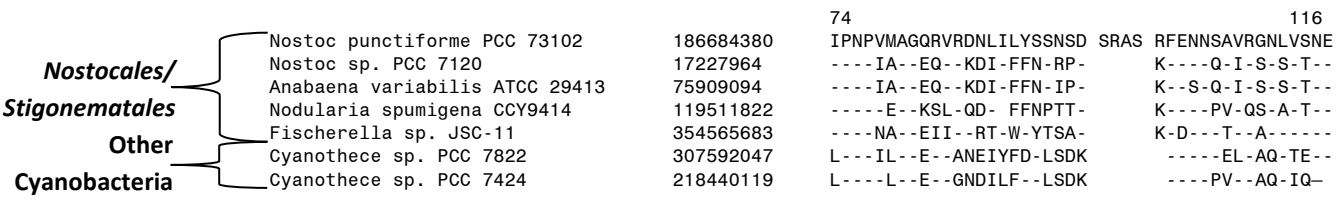
Supplementary Figure 138

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



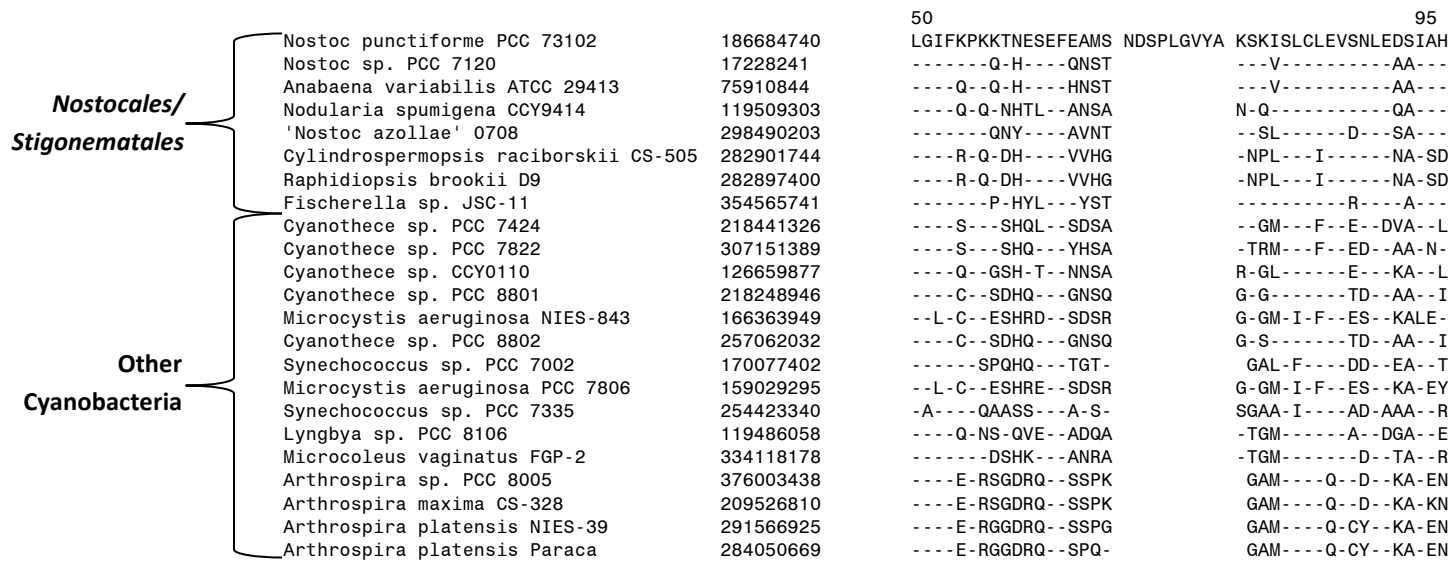
Supplementary Figure 139

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



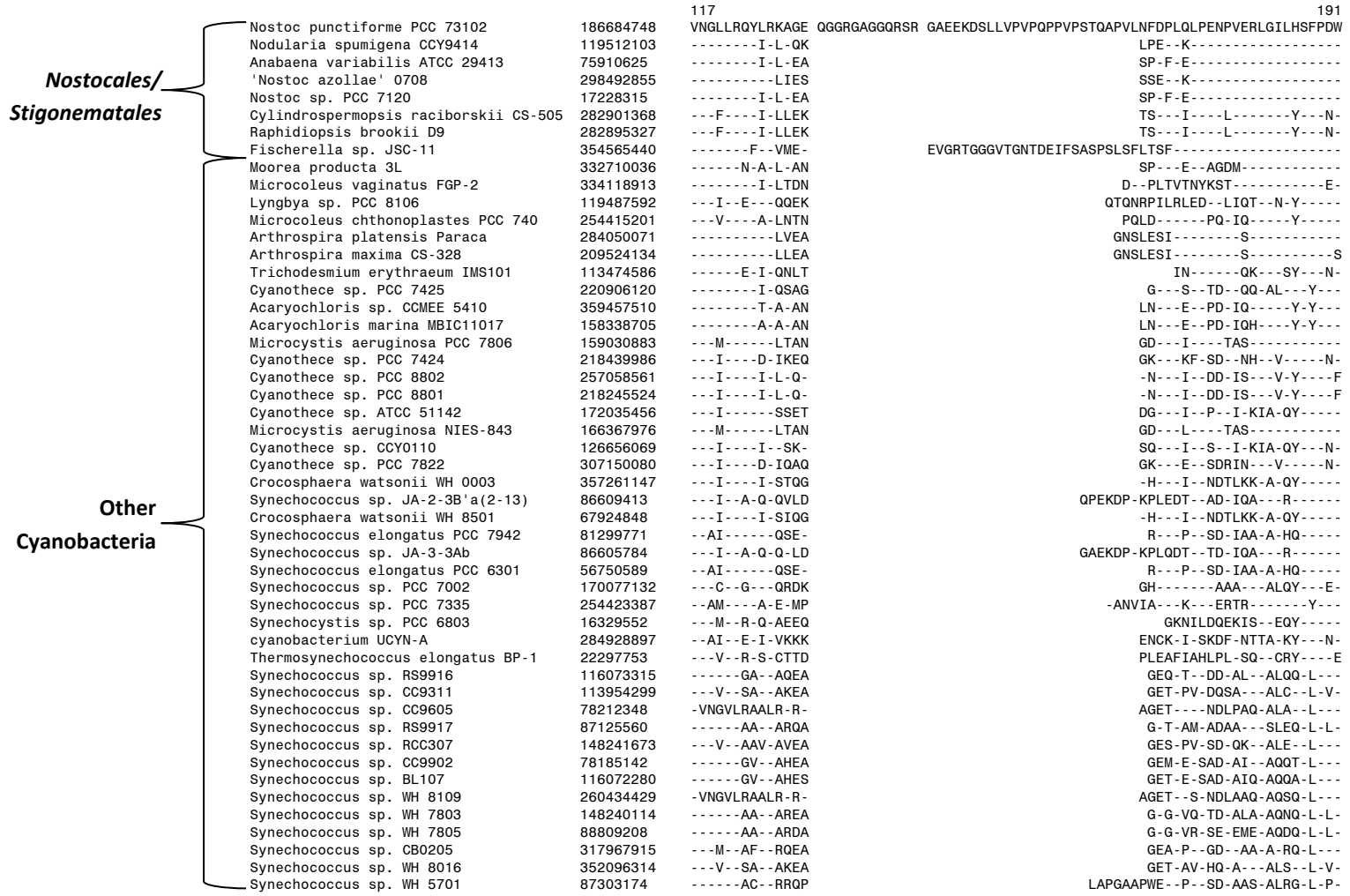
Supplementary Figure 140

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



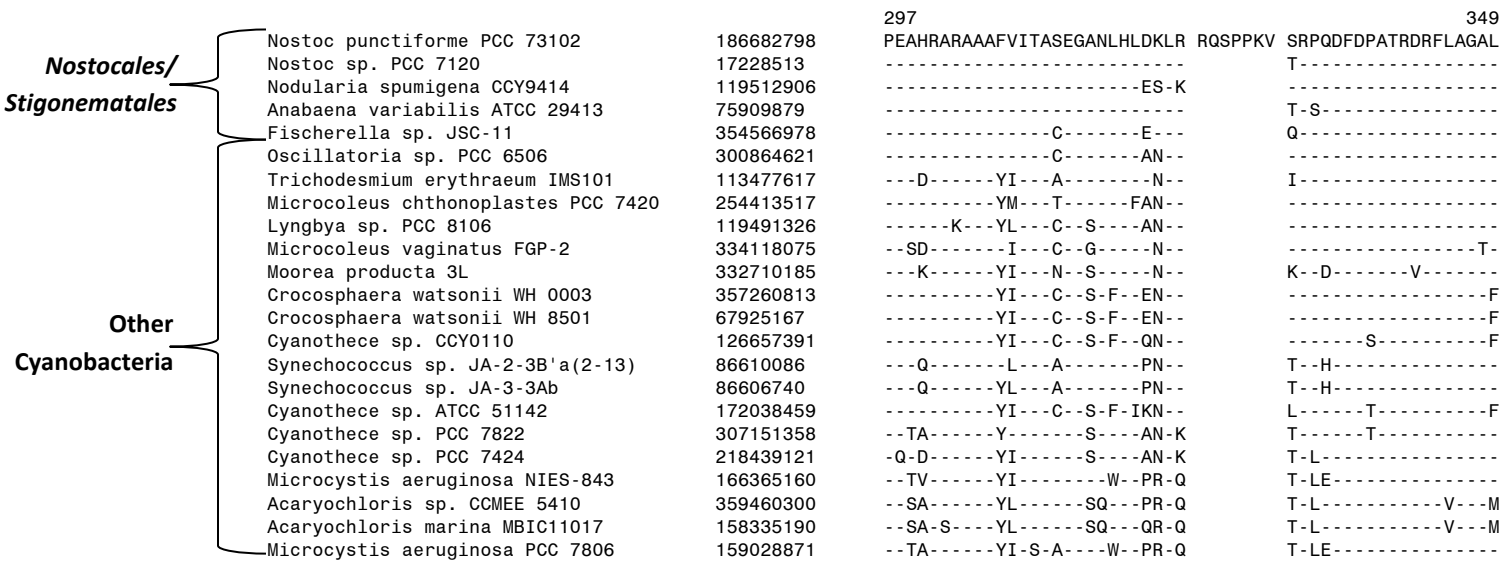
Supplementary Figure 141

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



Supplementary Figure 142

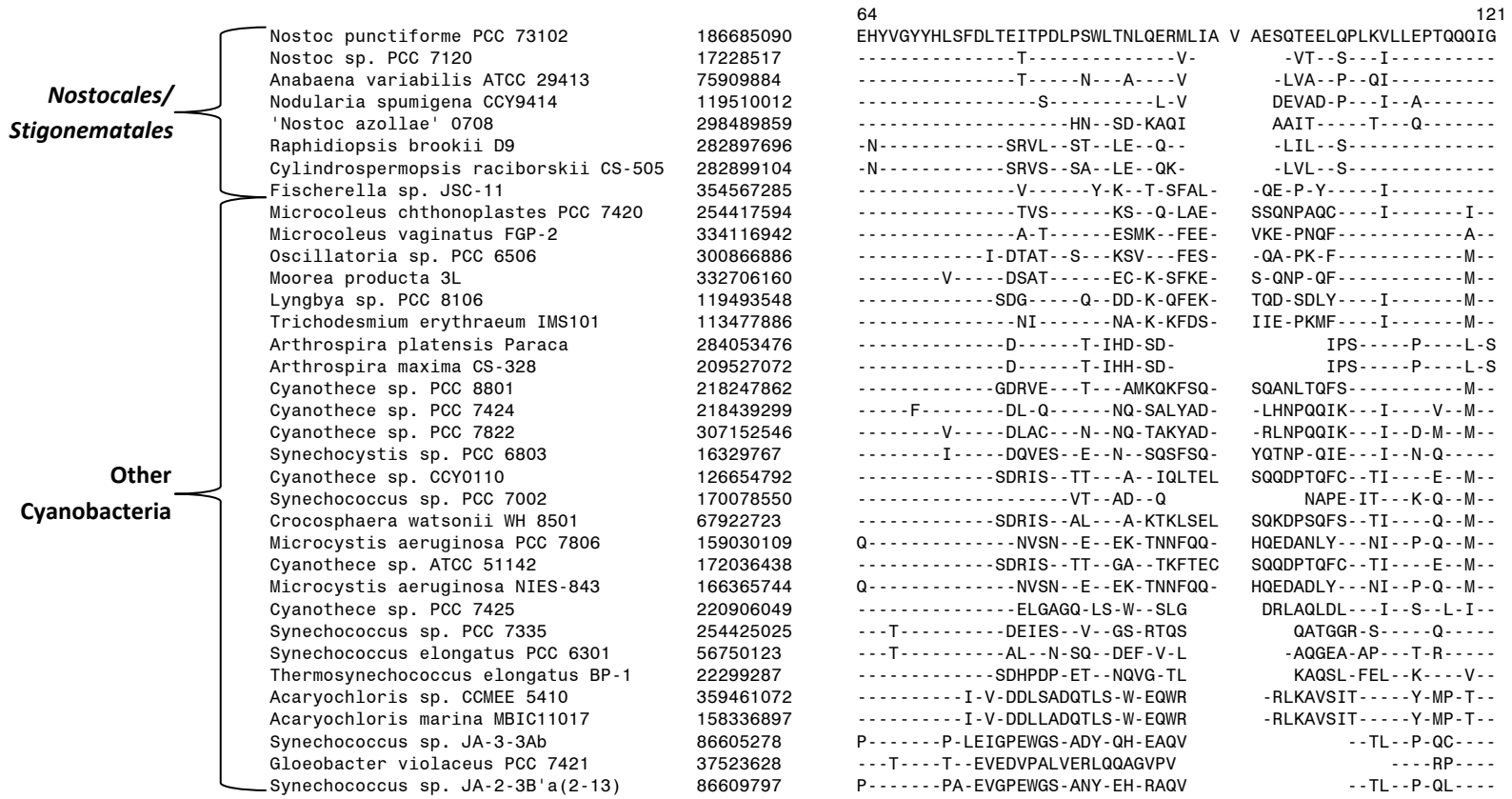
Partial sequence alignment for the protein rRNA SAM-dependent methyltransferase, showing a 12 aa insert that is specific for *Nostoc punctiforme PCC 73102.*



		297	349
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-
Moorea producta 3L	332710185	---K------YI---N--S-----N--	K--D-------V-------
Crocosphaera watsonii WH 0003	357260813	----------YI---C--S-F--EN--	----------F
Crocosphaera 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----------

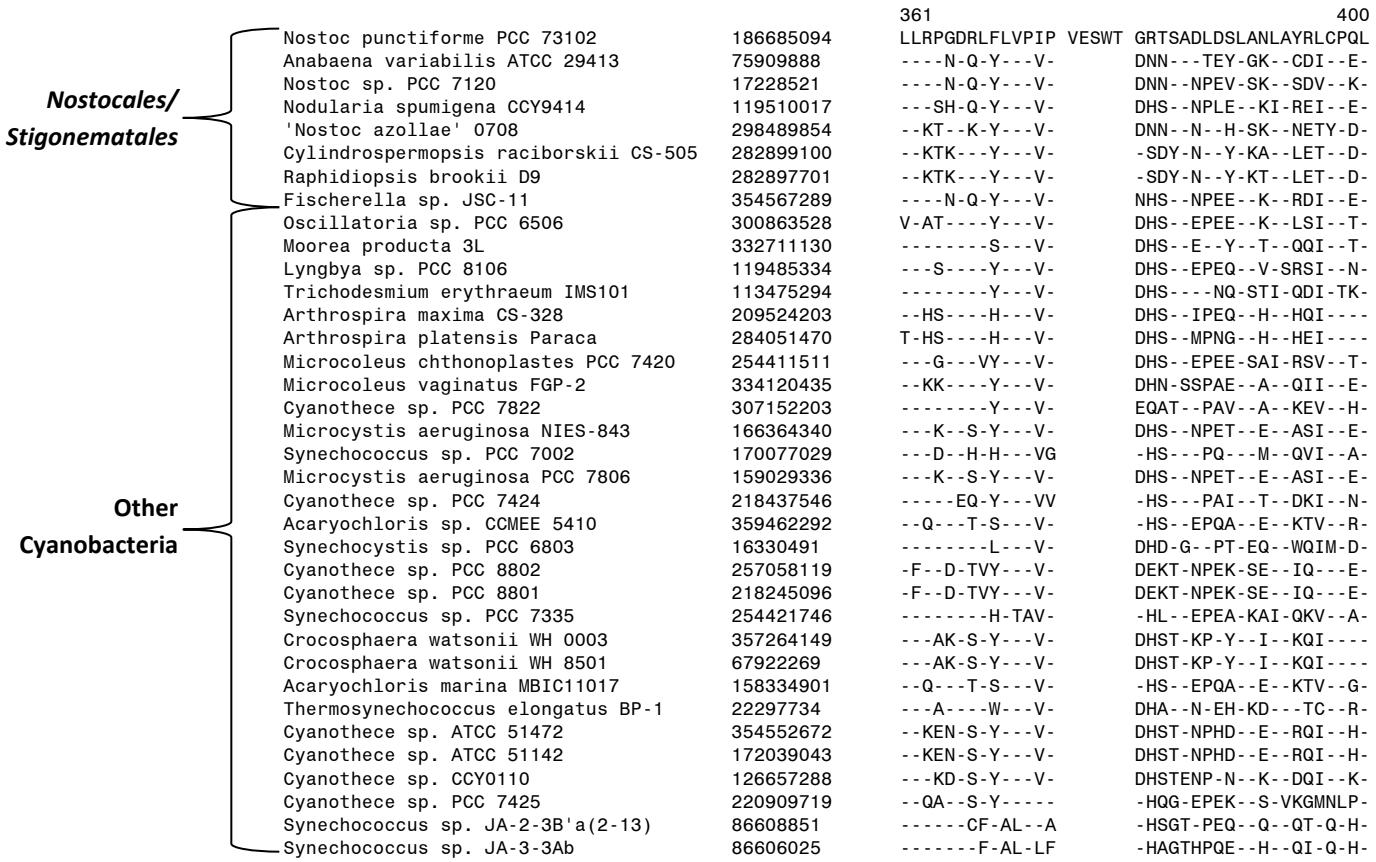
Supplementary Figure 143

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



Supplementary Figure 144

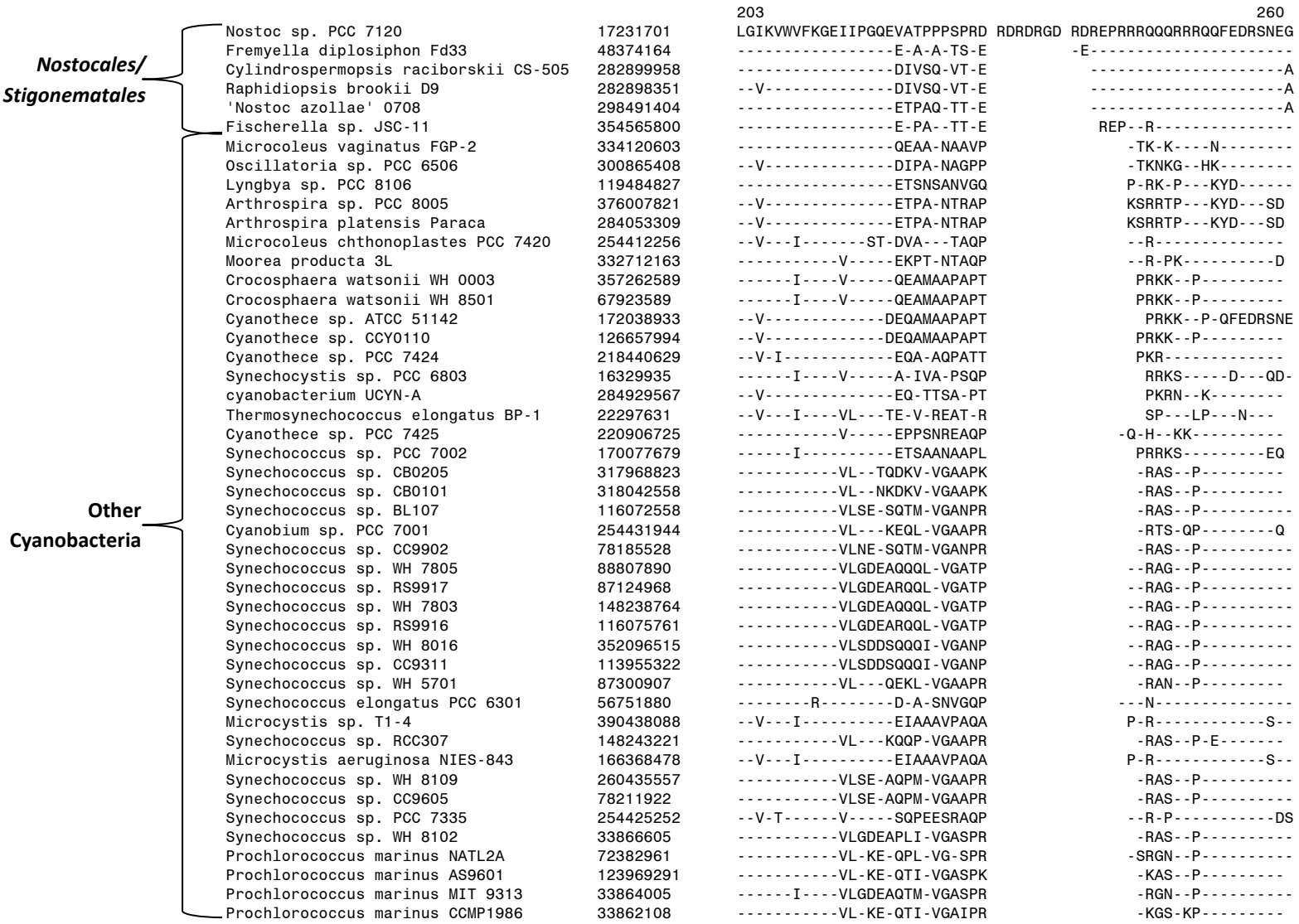
Partial sequence alignment for the hypothetical protein Npun_R4999, showing a 1 aa insert that is specific for *Nostoc punctiforme PCC 73102*.



		361	400
Nostoc punctiforme PCC 73102	186685094	LLRPGDRLFLVPIP VESWT	GRTSADLDSLANLAYRLCPQL
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----
Microcoleus chthonoplastes PCC 7420	254411511	---G---VY---V-	DHS--EPEE-SAI-RSV--T-
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-
Crocosphaera watsonii WH 0003	357264149	---AK-S-Y---V-	DHST-KP-Y--I--KQI----
Crocosphaera 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-

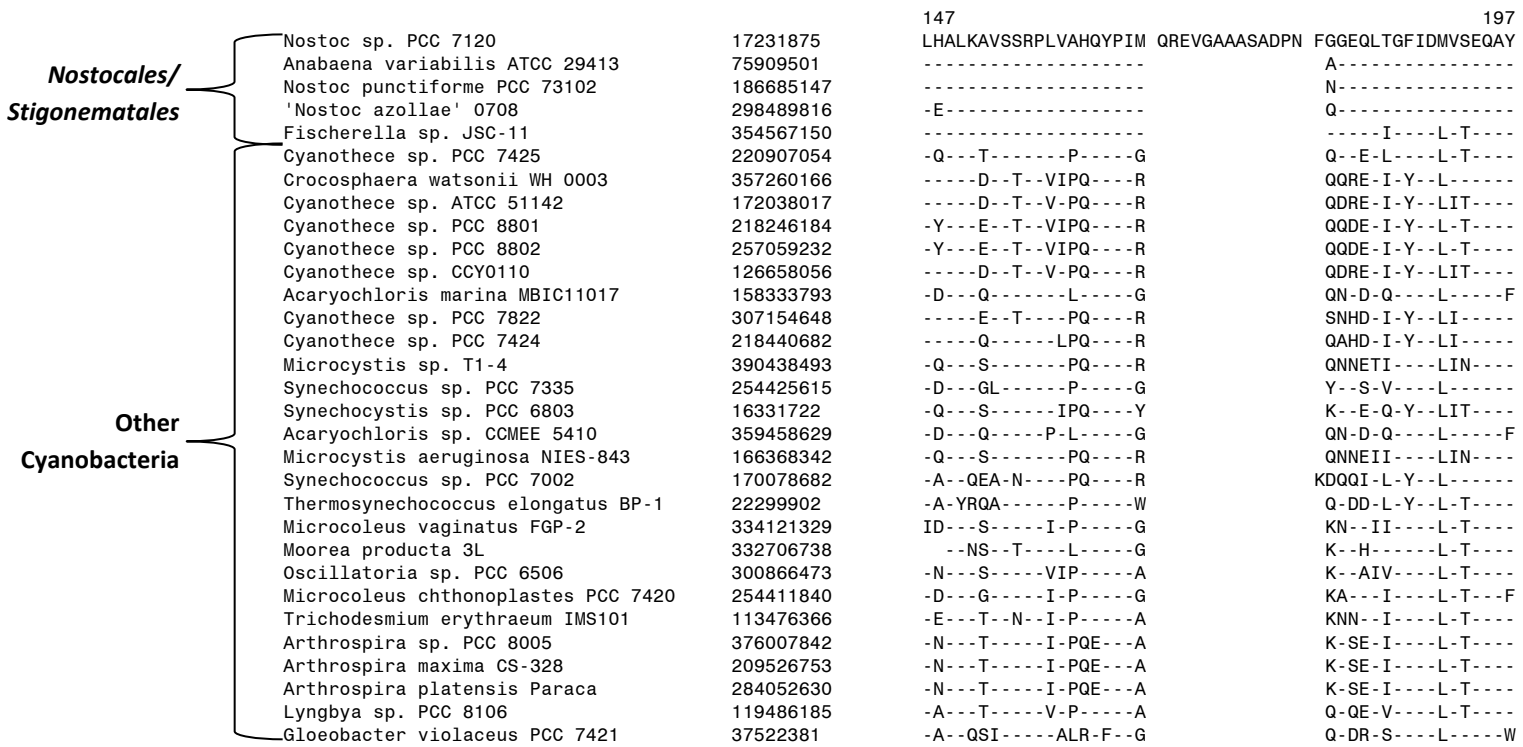
Supplementary Figure 145

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



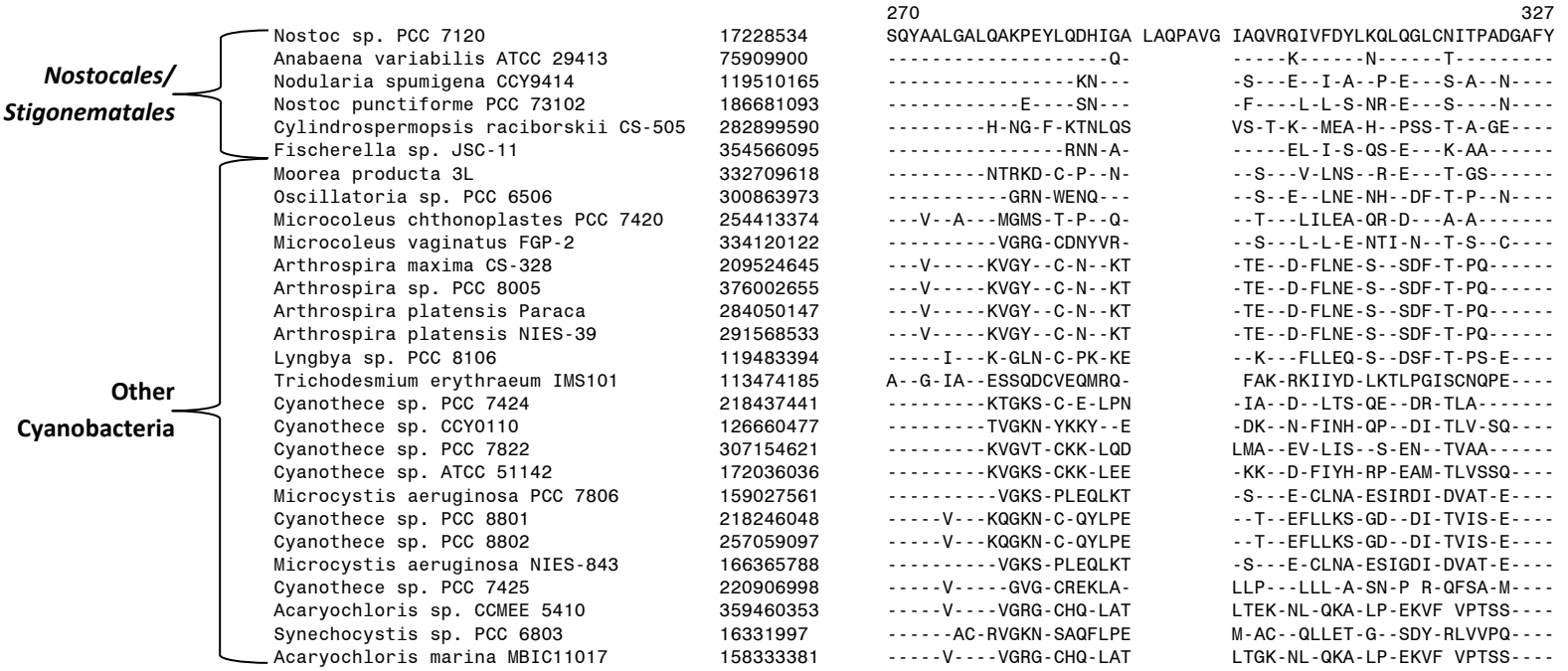
Supplementary Figure 146

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



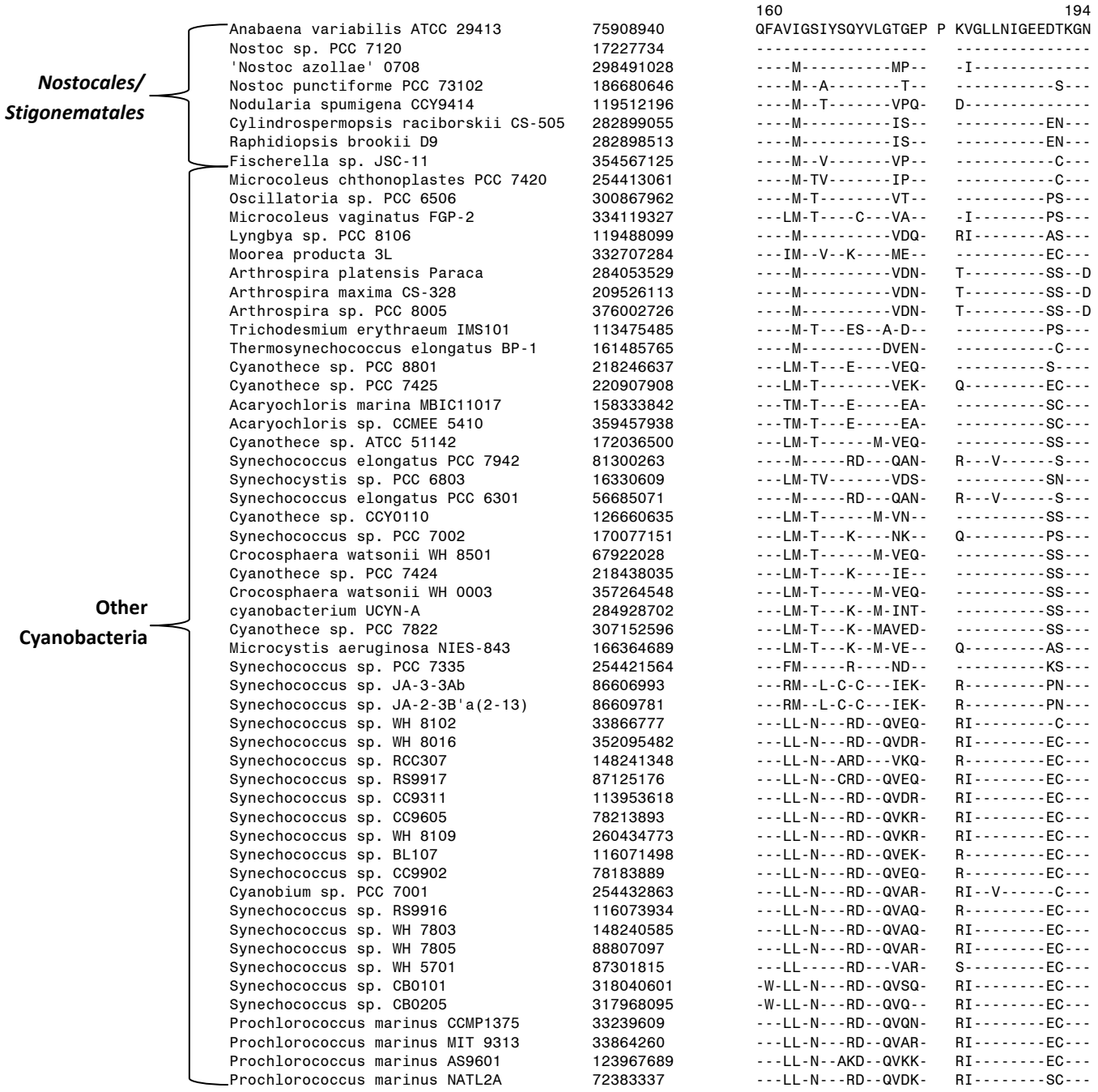
Supplementary Figure 147

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



Supplementary Figure 148

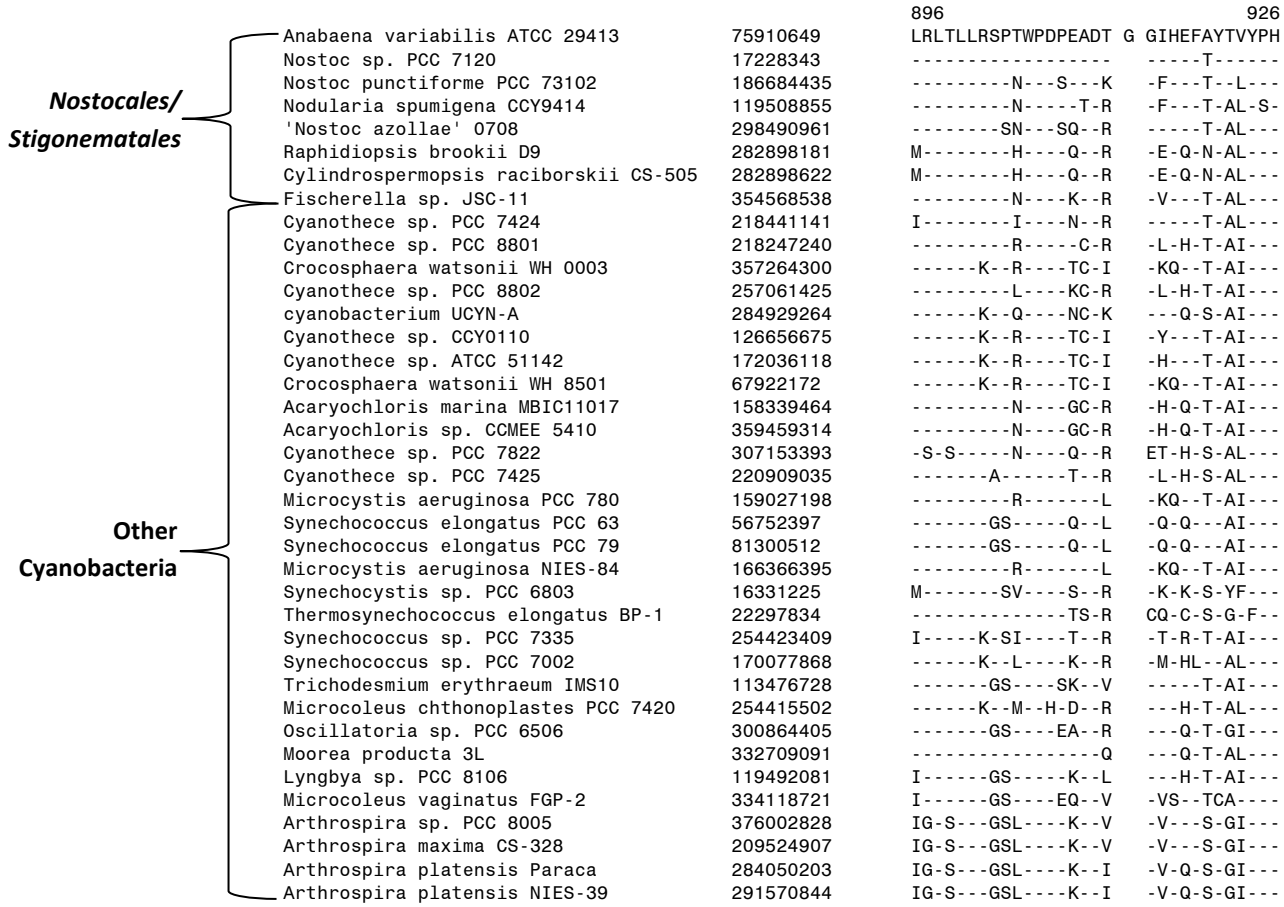
Partial sequence alignment for the protein aspartate aminotransferase, showing a 7 aa insert that is specific for *Nostoc sp. PCC 7120.*



		160	194
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--------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--	----------SS---
Synechococcus sp. PCC 7002	170077151	---LM-T---K----NK--	Q---------PS---
Crocosphaera watsonii WH 8501	67922028	---LM-T------M-VEQ-	----------SS---
Cyanothece sp. PCC 7424	218438035	---LM-T---K----IE--	----------SS---
Crocosphaera watsonii WH 0003	357264548	---LM-T------M-VEQ-	----------SS---
cyanobacterium UCYN-A	284928702	---LM-T---K--M-INT-	----------SS---
Cyanothece sp. PCC 7822	307152596	---LM-T---K--MAVED-	----------SS---
Microcystis aeruginosa NIES-843	166364689	---LM-T---K--M-VE--	Q---------AS---
Synechococcus sp. PCC 7335	254421564	---FM-----R----ND--	----------KS---
Synechococcus sp. JA-3-3Ab	86606993	---RM--L-C-C---IEK-	R---------PN---
Synechococcus sp. JA-2-3B'a(2-13)	86609781	---RM--L-C-C---IEK-	R---------PN---
Synechococcus sp. WH 8102	33866777	---LL-N---RD--QVEQ-	RI---------C---
Synechococcus sp. WH 8016	352095482	---LL-N---RD--QVDR-	RI--------EC---
Synechococcus sp. RCC307	148241348	---LL-N--ARD---VKQ-	R---------EC---
Synechococcus sp. RS9917	87125176	---LL-N--CRD--QVEQ-	RI--------EC---
Synechococcus sp. CC9311	113953618	---LL-N---RD--QVDR-	RI--------EC---
Synechococcus sp. CC9605	78213893	---LL-N---RD--QVKR-	RI--------EC---
Synechococcus sp. WH 8109	260434773	---LL-N---RD--QVKR-	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-----RD---VAR-	S---------EC---
Synechococcus sp. CB0101	318040601	-W-LL-N---RD--QVSQ-	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
Anabaena variabilis ATCC 29413	75910649	LRLTLLRSPTWPDPEADT 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---
Crocosphaera 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---
Cyanothece sp. ATCC 51142	172036118	------K--R----TC-I -H---T-AI---
Crocosphaera 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 IMS101	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.*