South Dakota State University
Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange

Electronic Theses and Dissertations

2018

Proteogenomic Study of Nitrogen-Fixing Cyanobacterium Anabaena Cylindrica

Yeyan Qiu
South Dakota State University

Follow this and additional works at: https://openprairie.sdstate.edu/etd

 Part of the Molecular Biology Commons

Recommended Citation
Qiu, Yeyan, "Proteogenomic Study of Nitrogen-Fixing Cyanobacterium Anabaena Cylindrica" (2018). *Electronic Theses and Dissertations*. 2478.
https://openprairie.sdstate.edu/etd/2478

This Dissertation - Open Access is brought to you for free and open access by Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. For more information, please contact michael.biondo@sdstate.edu.

PROTEOGENOMIC STUDY OF NITROGEN-FIXING CYANOBACTERIUM *ANABAENA CYLINDRICA*

BY

YEYAN QIU

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy

Major in Biological Sciences

South Dakota State University

2018

PROTEOGENOMIC STUDY OF NITROGEN-FIXING CYANOBACTERIUM *ANABAENA CYLINDRICA*

YEYAN QIU

This dissertation is approved as a creditable and independent investigation by a candidate for the Doctor of Philosophy in Biological Sciences degree and is acceptable for meeting the dissertation requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Ruanbao Zhou, Ph.D. Date

Thesis Advisor

Volker Brözel, Ph.D. Date

Head, Department of Microbiology

Dean, Graduate School Date

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor Dr. Ruanbao Zhou, who opened a door for me to Molecular Microbiology. Still, remember the first time heard his talk at our college - Anhui Normal University, his enthusiasm, and kindness encouraged my farthest journey to the US from my home and to pursue my Ph.D. in South Dakota State University. Here, I received my molecular biology training and critical thinking in science, more importantly, the passion for science is a treasure for my life.

I would like to thank my committee members Dr. Volker Brözel, Dr. William Gibbons, Dr. Michael Hildreth, Dr. Feng Li and Dr. Igor Sergeev. The Ph.D. journey was impossible without their encouragement, critiques, and generous advice. I wholeheartedly thank Dr. Liping Gu, who always offers me the most practical advice on my experiments, data analysis as well as care for my daily life. I would also like to extend my thanks to the SDSU Functional Genomics Core Facility and the SDSU Biology & Microbiology Department, for offering me the resources I needed to make my research a success.

I would like to dedicate my thesis to my beloved family far away in China and grandma's blessings. I could never have gone this far without their support and love. Last, but not least, I would like to thank Jaimie Gibbons, Jagdeep Singh Sidhu, Nanfang Wang, Aldon Myrlie, Kangming Chen, Shengni Tian, Xinyi Xu, Charles Halfmann, Huilan Zhu, Nate Braselton, Trevor VanDenTop, Matt Mckillop, and James Young for every wonderful day we work and laugh together!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

PROTEOGENOMIC STUDY OF NITROGEN-FIXING CYANOBACTERIUM *ANABAENA CYLINDRICA*

YEYAN QIU

2018

Cyanobacteria are gram-negative photosynthetic bacteria. In some filamentous cyanobacteria such as *Anabaena cylindrica*, every 10^{th} to 20^{th} vegetative cells can differentiate into nitrogen fixing heterocysts. Heterocyst can induce the neighboring cells developing into spore-like akinetes. The specialized cell functions and resourceful networks have contributed to the prosperity of cyanobacteria for over 2 billion years, but the genetic mechanisms for multi-cellular differentiation are barely known, especially for akinete formation.

The heterocysts, akinetes and vegetative cells of *Anabaena cylindrica* were isolated for proteomic study. This study identified a total of 1395 proteins, including 664 proteins from akinetes, 751 proteins from heterocysts, and 1236 proteins from vegetative cells. There were 45 proteins (33 novel proteins) found exclusive in akinetes, 57 heterocyst-specific proteins (33 novel proteins), including *nif* gene products, and 485 proteins exclusively in vegetative cells. HAVe model was proposed that akinetes, unlike the typical spores of bacteria, perform unique biochemical functions that collaborate with both heterocysts and vegetative cells.

Regardless of nitrate availability, some vegetative cells of *Anabaena cylindrica* are programed to differentiate semi-regularly spaced, single heterocysts along filaments. Since heterocysts are non-dividing cells, with the sole known function for solar-powered N_2-

fixation, is it necessary for heterocyst to retain entire genome (≈7.1Mb) from its progenitor vegetative cell? By sequencing the genome of isolated heterocyst, I discovered that at least six DNA elements (≈0.12 Mbp) are deleted from the heterocyst genome during heterocyst development. The six-element deletions restore five genes (*nifH1, nifD, hupL, primase P4*, *acyl_5725* (hypothetical protein) that were interrupted in the genome of vegetative cells. More deletions are expected to be identified in the completed genome of heterocyst, a uniquely solar-powered, oxic N_2-fixing cell. To my best knowledge, this is the first report that (1) different genomes may occur in distinct cell types in a single bacterium; and (2) genome editing is coupled to cellular differentiation and/or cellular function in a multicellular cyanobacterium.

In response to environmental changes, *Anabaena cylindrica* differentiate three cell types, vegetative cells for photosynthesis, heterocysts for nitrogen fixation, and akinetes for stress survival. Cell-surface polysaccharides play important roles in bacterial ecophysiology. In this study, specific cell-surface sugars were discovered in heterocysts, akinetes and vegetative cells of *A. cylindrica* using 20 fluorescein labeled lectins. Both N-acetylglucosamineor-binding lectins WGA and succinylated WGA bound specifically to the vegetative cells. Akinetes bound to three mannose-binding lectins (LCA, PSA, and ConA), and one of the galactose-binding lectins (GSL-I). ConA also bound to heterocyst, and the binding was diminished in the heterocysts of an *all4388* mutant, in which the putative polysaccharide export protein gene *all4388* was disrupted.

Above proteomics, genomics and genetic research greatly added to our understanding of the cell development in *A. cylindrica* and clarify the patterns of gene expression in

heterocysts, akinetes and vegetative cells, which pave the way for further study of *Anabaena cylindrica*.

INTRODUCTION

Cyanobacteria are ancient gram negative photosynthetic bacteria and some of them can fix atmospheric nitrogen. Among these nitrogen-fixing cyanobacteria, a group of filamentous cyanobacteria applies the cell differentiation strategy that progenitor vegetative cells fix the carbon from CO_2, and about 10% of the population are differentiated into specialized N_2 -fixing cells called heterocysts. Vegetative cells and heterocysts are in a reciprocal relationship by exchanging the fixed carbon and nitrogen through the filaments. When the cells are stressed from environmental conditions like cold or desiccation, a third cell type, spore-like akinetes can be formed from vegetative cells in some of these nitrogen-fixing filamentous cyanobacteria. The young akinetes are usually formed at both sides of heterocysts, and they'll fall off from the filaments while maturing.

These photosynthetic, nitrogen-fixing, akinetes forming cyanobacteria encounter several challenges from self-development and environmental factors to thrive, which provide us a unique model to study cell differentiation; the incompatible of nitrogenase activity and the O_2 from photosynthesis and environment; the nutrients exchange between vegetative cells and heterocysts, even more complicated akinetes formation adjacent to heterocysts; the programming of the differentiation that how the cells' fates are determined and the development processes. We have little knowledge of these cell interactions and development so far, with examining the proteomics from these three distinct cell types will deepen our understanding of their unique functions.

On the other hand, heterocyst itself is very important for studying its oxic nitrogen-fixing nature, which is of great potential for future sustainable agriculture production.

Vegetative cells giving birth to heterocysts is a process that is crucial for acquisition of nitrogen fixation, but also the termination of cell division where the gene regulation could occur at any level of the central dogma. Our first-time genomic and transcriptomic analysis combined with proteomics on heterocysts will provide a comprehensive vision for heterocysts functions. As, this fundamental knowledge progress, we hope that our findings help pave the way for a greener biological nitrogen fixation era.

OBJECTIVES

The objectives of this study were to:

I. Investigate the proteomics of purified vegetative cells, heterocysts and akinetes to reveal their differentiation, exclusive functions, network etc.

II. Explore the genomics of heterocysts to further clarify the gene rearrangement including the *nif* genes, which probably associated with other exclusive functions.

III. Discover different polysaccharide among the surface of vegetative cells, heterocysts, and akinetes.

CHAPTER 1: Literature Review

1.1 Cyanobacteria overview

Cyanobacteria, oxyphototrophic eubacteria, are one of the oldest species on the earth that lead to rapidly accumulated atmosphere oxygen levels during the episode of the Great Oxidation Event around 2.5-2.3Ga (Bekker et al., 2004; Schirrmeister, Gugger, & Donoghue, 2015). The facilitation of oxygen on the earth allowed for the evolutionary emergence of the complex life forms on earth.

1.1.1 The classification of cyanobacteria

Cyanobacteria autotrophic lifestyle enabled them to conquer almost every terrestrial and aquatic habitat - marine, fresh water, soil, and even Antarctic rocks, which constitute one of the largest group of gram-negative prokaryotes (Büdel, 2011). Rippaka et al. proposed five subsections (section I, II, III, IV, and V) in cyanobacteria based on a comparative study of 178 cyanobacteria strains (Stanier, Deruelles, Rippka, Herdman, & Waterbury, 1979). These subjects collected from 22 genera were distinguished from their phycological classification, including cell structure, cell division, and development. Subsections I and II are unicellular cyanobacteria. Cyanobacteria in section I are reproduced by binary fission or by budding, for example, *Synechococcus* and *Gloeothece*. Cyanobacteria in section II are given rise to small daughter cells called baeocysts resulted from multiple fission or are reproduced by both multiple fission and binary fission, like *Chroococcidiopsis cyanosphaera*. Subsection III, IV and V are filamentous cyanobacteria. Subsection III are the filamentous cyanobacteria that cell division happens only in one plane, such as *Spirulina* and *PseudAnabaena*. Cyanobacteria from section IV and V can form heterocysts, a specialized cell type that responsible for nitrogen fixation

in the absence of combined nitrogen. The difference is that the cell division in section IV cyanobacteria only in one plane while in more than one plane from section V. The examples of section IV and V are *Anabaena* and *Nostoc*, *Fischerella* and *Chlorogloeopsis*, respectively. This five subsections classification of cyanobacteria has been widely accepted and used so far.

Table 1-1 The five subsections classification in cyanobacteria based on cell morphology and cell replication approaches (Stanier et al., 1979)

<table>
<tr><td>Unicellular; cells single or forming colonial aggregates held together by additional outer cell wall layers</td><td colspan="2">Reproduction by binary fission or by budding</td><td>Section I</td></tr>
<tr><td></td><td colspan="2">Reproduction by multiple fission giving rise to small daughter cells (baeocytes), or by both multiple fission and binary fission</td><td>Section II</td></tr>
<tr><td rowspan="3">Filamentous; a trichome (chain of cells) which grows by intercalary cell division</td><td rowspan="3">Reproduction by random trichome breakage, by formation of hormogonia and (Sections IV and V only) sometimes by germination of akinetes</td><td>Trichome always composed only of vegetative cells</td><td>Division in only one plane
Section III</td></tr>
<tr><td rowspan="2">In the absence of combined nitrogen, trichome contains heterocysts; some also produce akinetes</td><td>Division in only one plane
Section IV</td></tr>
<tr><td>Division in more than one plane
Section V</td></tr>
</table>

1.1.2 The evolutionary and applicable importance of cyanobacteria

The occurrence of free oxygen is probably one of the most important biogeological events that facilitated colorful oxygen-dependent life forms on the earth. This breakthrough resulted from cyanobacteria, which is the pioneer of photosynthesis and spread the phototrophy to eukaryotic lineages.

1.1.2.1 Cyanobacteria - an ancestor of chloroplast

Several phylogenetic studies have established that chloroplast, the primary plastid, exist in photosynthetic eukaryotes arose from cyanobacteria through endosymbiosis

(Giovannoni et al., 1988). The incorporation has greatly modified the earth's ecology and played a crucial and fundamental role in evolution. The origin of chloroplast from cyanobacteria has been testified by multiple approaches, such as the 16S rRNA and protein markers. Due to pitfalls in the reconstruction of ancient evolutionary events, like dilution of phylogenetic landmarks, site saturation, horizontal gene transfer, the debates have been focusing on the closest lineage of cyanobacteria to chloroplast (Rajaniemi, Hrouzek, Rantala, & Hoffmann, 2017; Turner, 1999; Zhaxybayeva, Gogarten, Charlebois, Doolittle, & Papke, 2006). Recently, the identification of core genes enabled a highly resolved cyanobacterial tree and these core sequences were used to track the origin of the chloroplast. Up till now, the chloroplast is thought to be originated among members of one of the major cyanobacterial lineages that contain filamentous nitrogen fixing cyanobacteria (Ochoa de Alda, Esteban, Diago, & Houmard, 2014).

1.1.2.2 Cyanobacterial applications in biofuel production

Cyanobacteria are autotroph photosynthetic microorganism that some can fix nitrogen as well. They convert CO_2 and H_2O to carbohydrates utilizing sunlight. The biofuel is developed aimed to replace traditional fuels, coals by the production of renewable energy. Until the third-generation biofuel, to get final energy stock, additional chemical or biological processes had to be used. The processes including fermentation, gasification are a waste of energy and decrease the efficiency of biofuel production. Cyanobacteria itself inherited all the pathways to energy stock but lacking the pathway to produce specific biofuel products. Only by directing the reactions by inserting necessary genes without disrupting the cell structure and cycle, the bio-products can be harvested continuously as they are secreted from the cyanobacteria cells.

The genetic engineering in cyanobacteria is well studied that several transformation methods could be used to incorporate foreign genes, such as conjugation, natural transformation, and electroporation (J Elhai, Vepritskiy, Muro-Pastor, Flores, & Wolk, 1997). Usually, the transformation of unicellular cyanobacteria is easier due to the natural transformation properties that the DNA can be naturally transported across the cell membrane and naturally transformed. The transformation in some filamentous cyanobacteria is tricky due to the restriction enzymes who can degrade the foreign DNA. Majority of biofuel research has been done in unicellular cyanobacteria, but the high carbon and nitrogen fixation rate in filamentous cyanobacteria have refocused researchers' attention. The foreign genes can be incorporated either into chromosome or plasmids. The insertion in the chromosome is stable but needs to be aware of not interrupting other genes, and the insertion in self-replicated plasmids will have multiple copies that numbers are unknown. Antibiotics resistance can also be introduced to select the successful transformants.

Several biofuels have been successfully produced using cyanobacteria as host, such as ethanol and terpenes (Halfmann, Gu, Gibbons, & Zhou, 2014; Hellier, Al-Haj, Talibi, Purton, & Ladommatos, 2013). The productivity can be improved by modifying the pathways towards the biofuel products and shutting down some other unnecessary pathways. The advancements in cyanobacteria genomics, transcriptomes, proteomics, and metabolomics will deepen the understanding and help improve the biofuel production in the future.

1.2 Activities in cyanobacteria

1.2.1 Photosynthesis

Cyanobacteria are the one of oldest auto phototrophic microorganism that utilize solar energy to support living by photosynthesis carried by vegetative cells. Cyanobacteria have competitive advantages in absorbing a wide range of wavelengths of lights by phycobilisomes and proceeding photosynthesis even at a low CO_2 concentration by carboxysomes.

1.2.1.1 Phycobilisome and photosystems

Cyanobacteria appear blue-greenish color because of phycobilisomes, which are light harvesting antennae of photosystem II (Figure 1-1) (D. Campbell, Hurry, Clarke, & Gustafsson, 1998). The phycobilisome is attached to the photosynthetic membrane surface instead of integrating into membrane as most light-harvesting antenna do. The phycobilisomes absorb wide wavelength range of light between 500-650nm and transfer the solar energy to chlorophyll for photosynthesis (Liu, Chen, Zhang, & Zhou, 2005). There are two types of phycobiliprotein structure: pigmented phycobiliproteins (PBPs) that absorb solar energy and nonpigmented linker polypeptides to stabilize the complex. PBPs were classified into four groups, allophycocyanin (APC), phycoerythrin (PE), phycoerythrocyanin (PEC) and phycocyanin (PC). The composition and organization of the phycobilisome rods and core are variable in different cyanobacteria; the three-cylinder core and six-peripheral rod configuration is common (Glazer, 1978). Phycobilisomes are stable but dynamic structure for cells to adapt different environmental changes. Complementary chromatic adaption, which is the process that composition of phycobilisome alters in response to the prevalent wavelengths of light

exists in some of the filamentous cyanobacteria to utilize available light efficiently. Phycobilisome degrades when the nutrient is limited and its degradation provides a substantial amount of nitrogen; also, when the cells are absorbing light of metabolic arrest to avoid photodamage (Grossman, Schaefer, Chiang, & Collier, 1993).

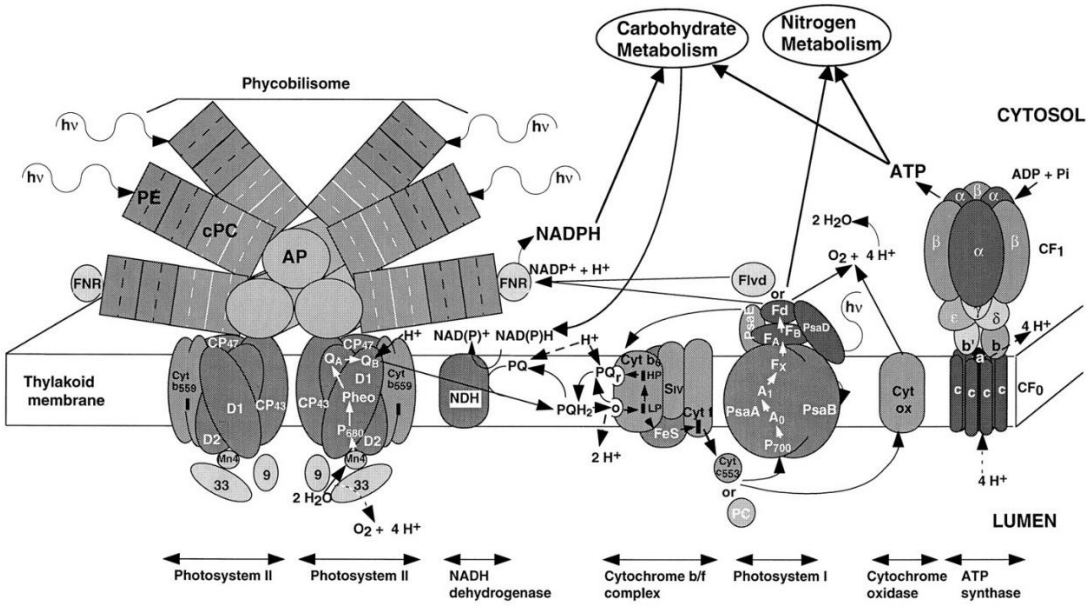

Figure 1-1 Photosystems in cyanobacteria (D. Campbell et al., 1998).

Photosystem I and photosystem II are embedded in thylakoid membrane; photosystem I function as light-driven, cytochrome *c*6: ferredoxin oxidoreductase (Golbeck, 1994), and photosystem II acts as light-driven, water:plastoquinone oxidoreductase (Guskov et al., 2009). Cyanobacteria usually have a lower amount of PSII; The ratio of PSI to PSII is about 3 to 5.8 (Murakami & Fujita, 1988). The conversion of energy from solar to chemical form is the result of two chemical RCs: P680 (chl a-complex in PSII) and P700 (chl a/a'-complex in PSI) acting in tandem to deliver excitation energy and initiates the conversion of energy (Renger & Renger, 2008). Energy rich compounds - NADPH and

ATP that formed by electron transport - drive CO_2 fixation and nitrogen fixation (Tabita, 1994).

1.2.1.2 CO_2 concentration - carboxysome

RuBisCo, Ribulose-1,5-biophosphate carboxylase oxygenase, an enzyme catalyzes the fixation of CO_2 molecule in the first major step of carbon fixation. Two limitations of RuBisCo could affect the efficiency of carbon fixation; its low affinity to CO_2 and the side reaction with O_2 (G Bowes, 1991). Cyanobacteria emerged with increasing O_2 ratio in the air, which posed a challenge to cells how to increase CO_2 within the cells to meet the requirement for RubisCo. Besides, vegetative cells are considered as O_2 tank due to the photosynthesis, which O_2 can also being the substrate for RubisCo and the products are phosphoglycolate and 3-phosphoglycerate (Badger et al., 1998). Phosphoglycolate goes through photorespiration that two molecules of phosphoglycolate are transformed to one molecule of CO_2 and one molecule of 3-phosphoglycerate, which decrease the efficiency of photosynthetic capacity (Gi Bowes, Ogren, & Hageman, 1971).

In cyanobacteria, to improve CO_2 concentration around RuBisCo and prevent the penetration of O_2, CO_2 concentration mechanism (CCM) and its components have been identified. The CCM components at least include two bicarbonate transporters and two CO_2 uptake systems to accumulate HCO_3^- in the cytosol within the cell (Badger & Price, 2017). Carbonic anhydrase in further transform HCO_3^- to CO_2 in the compartment surrounding RubisCo, which is named carboxysome. Two carboxysomes were discovered: α and β carboxysomes, which β carboxysome is more commonly studied. In both carboxysomes, the carboxysomes shell consist of several small polypeptides (8-12 kDa) that are homologous with each other; and two larger polypeptides that do not bear

homology: CsoS2 (80 -90 kDa) and CsoS3 (55 - 65 kDa) in α carboxysome; CcmM (55 - 70 kDa) and CcmN (26 kDa) in β carboxysome (Kaplan & Reinhold, 1999). The shell of carboxysome also prevents the diffusion of CO_2 to maintain higher concertation within the carboxysome, where the mechanisms are not well known yet.

1.2.2 Nitrogen fixing in cyanobacteria

Some cyanobacteria are capable of aerobic nitrogen fixation. Nitrogenase is the enzyme complex that responsible for nitrogen fixation, which is O_2-sensitive (Gallon, 1981). The O_2 generated from photosynthesis along with O_2 from the environment. The aerobic nitrogen fixation and the sensitivity of nitrogenase to O_2 creates a paradox in cyanobacteria. Nitrogenase activities were measured in *Gloeocapsa* and *Anabaena cylindrica*, which were examples of unicellular and filamentous cyanobacteria by acetylene reduction under alternating light-dark cycles revealed their different activity patterns as early as the 1980s (Millineaux, Gallon, & Chaplin, 1981). Recent studies have further shown different strategies carry on in unicellular and filamentous cyanobacteria: temporal and spatial separation of nitrogen fixation and photosynthesis (Stal, 2001).

1.2.2.1 Circadian clock in unicellular cyanobacteria

Briefly, the circadian clock in unicellular cyanobacteria is the temporal separation of photosynthesis and nitrogen fixation that the former occurred in the day and the latter happened in the night to avoid the poisoning of O_2 to nitrogenase. Two marine cyanobacteria strains, *Cyanothece* BH63 and BH68 were first exposed to 12 h light-dark cycle and had shown the aerobic nitrogenase activity being only confined to the dark period (Reddy, Haskell, Sherman, & Sherman, 1993). The circadian rhythms with photosynthesis in *Synechococcus* sp. PCC 7942 was also observed by introducing *Vibrio*

harveyi luciferase structural genes (*luxAB*) downstream of *psbAI* gene promotor, which encodes for photosystem II protein (T. Kondo et al., 1993).

A variety of circadian clock mutants were isolated (Takao Kondo, Tsinoremas, Golden, & Johnson, 1994) and the gene cluster of controlling circadian clock *kaiABC* was discovered by mutational analysis that deletion of any of these three genes would cause arhythmicity of circadian outputs (Ishiura, 1998; Xu, 2000). By gene mapping, the promoters of *kaiABC* were identified that *kaiA* promoter is on the upstream of *kaiA*; and *kaiB* promoter controls *kaiB* and *kaiC* (Ishiura, 1998). *KaiC* is crucial and its expression shows a direct correlation with significant phase shifting; and its ATPase activity and autophosphorylating defines the circadian rhythm (Terauchi et al., 2007). *KaiC* phosphorylation occurs at serine and threomine residues that *KaiA* enhances this process while *KaiB* functions as an attenuator (Kitayama, Iwasaki, Nishiwaki, & Kondo, 2003). The technique of stochastic gene expression advanced measuring gene expression at the single-cell level by a SsrA-tagged fluorescent protein YFP-SsrA to report the periodic activity of *kaiBC* promoter (Figure 1-2) (Chabot, Pedraza, Luitel, & Van Oudenaarden, 2007).

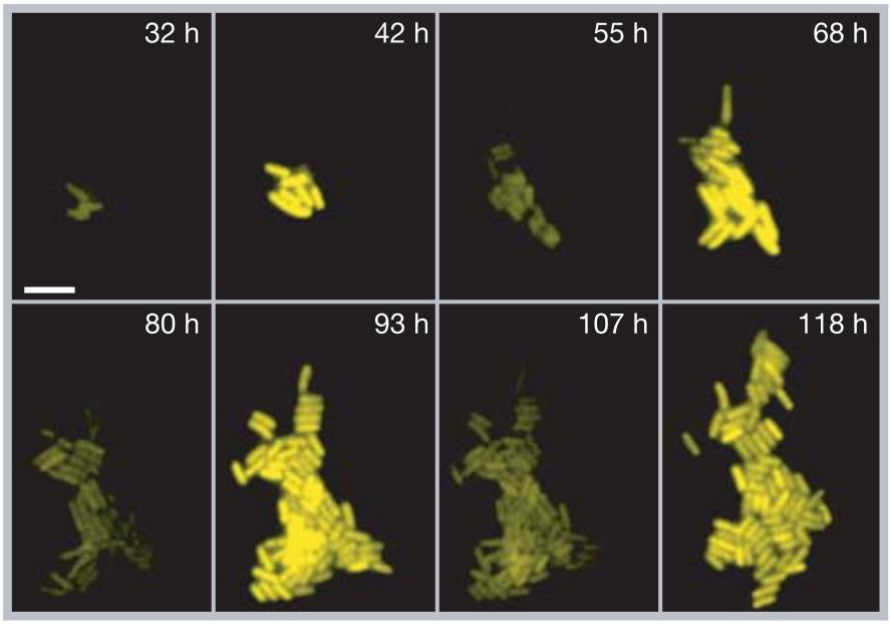

Figure 1-2 Stochastic gene expression of kaiBC to monitor circadian rhythms in single *Synechococcus* PCC 7942 cells using fluorescence microscopy (Chabot et al., 2007).

1.2.2.2 Heterocysts in filamentous cyanobacteria

Heterocysts are specialized cells in filamentous cyanobacteria that can fix nitrogen, which nitrogenase activity had been reported since the 1960s (Fay, Stewart, Walsby, & Fogg, 1968; Stewart, Haystead, & Pearson, 1969). The photosynthesis and nitrogen fixation were spatially separated by located in vegetative cells and heterocysts, respectively. The photosynthesis activity was deactivated in heterocysts primarily due to their loss of photosystem II, where O_2 is produced. The absence of photosystem II was discovered by the low yield of chlorophyll *a* fluorescence, besides, the fluorescence yield did not show any light-induced changes and no Hill-reaction (O_2-releasing reaction) activity was detected (Donze, Haveman, & Schiereck, 1972). Except for the O_2 that generated from photosynthesis, aerobic nitrogen fixing heterocysts developed several mechanisms to exclude the O_2 from the environment and neighboring vegetative cells.

The primary barrier to O_2 is thickening envelope formed in heterocysts outside of normal gram- negative outer membrane (Figure 1-3). The extra envelope in heterocysts consists of inner laminated glycolipid and outer homogeneous polysaccharide layers. The barrier function of these two layers was testified that the mutants lacking either of the heterocysts specific glycolipid or polysaccharide resulted in the failure of aerobic nitrogen fixing in heterocysts, and the restoration of aerobic nitrogen fixation occurred after the complementation of the knock out genes (Murry & Wolk, 1989). At pole that connecting heterocysts and vegetative cells, both layers became thicker to create a narrow cytoplasm channel that allowing the substances exchanges between these two types of cells (Walsby, 2007). Increase respiratory activity was observed in isolated heterocysts with dual functions that to arrest the possible O_2 within the heterocysts and provide ATP for nitrogen fixation (Fay & Walsby, 1966).

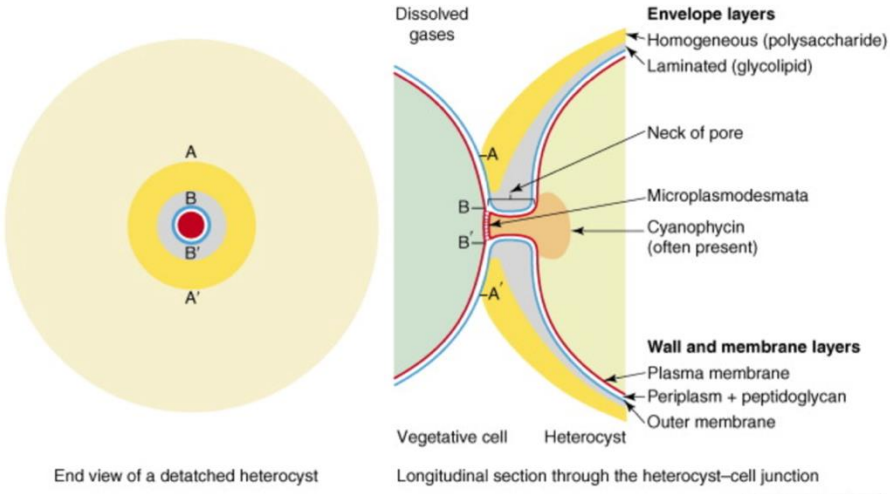

Figure 1-3 The scheme of heterocyst envelope structure (Walsby, 2007).

1.3 Multicellularity in filamentous cyanobacteria

The studies have shown that the multicellularity of cyanobacteria occurred around the GOE period with multiple evidence, including 16S rRNA phylogenetic study

(Schirrmeister et al., 2015) and discovered fossils (Schirrmeister et al., 2015). In further, the multicellular morphotype evolved early in the cyanobacterial lineage, and most of the extent exhibited cyanobacteria with morphological diversity on the earth today - even including most single cell species - were evolved from these ancient multicellular lineages (Bettina E Schirrmeister, Alexandre Antonelli, & Homayoun C Bagheri, 2011). The property of multicellularity could have provided cyanobacteria advantageous conquering to more diverse habitats and facilitated new lineages diversity. The progenitor vegetative cells can differentiate into heterocysts, akinetes, and homogonia but varied from species and different environmental conditions.

1.3.1 Mother cell - vegetative cells

Vegetative cells are the center for carbon fixation, which transformed CO_2 in the air to fixed carbon by utilizing the solar energy, and the mother cells for cell reproduction by cell division. The cells doubling time varies in different species and growth conditions: the typical range in unicellular cyanobacteria (e.g. *Synechococcus elongatus* PCC 7942 and *Synechocystis sp.* strain PCC 6803) is between 7 to 12 h (Mori & Hirschie, 1996); the filamentous cyanobacteria (e.g. *Anabaena* strains) is between 18 to 36 h (Prasanna, Kumar, Sood, Prasanna, & Singh, 2006). The vegetative cells cycle would maintain when the nutrients are sufficient, but heterocysts form when lack of combined nitrogen and akinetes form when lack of phosphate and in response to other stresses.

1.3.2 Nitrogen fixing cells - heterocysts formation and its space pattern

Heterocysts are formed in a beautiful spacing pattern that every 10 to 20 vegetative cells along the filaments when nitrogen-fixing filamentous cyanobacteria are stressed from nitrogen deprivation (C Peter Wolk, Ernst, & Elhai, 1994). The mechanism of patterning

and heterocysts differentiation has been well studied that could be classified into four phases: introduction, patterning, commitment and morphogenesis (Figure 1-4) (Hanlon & Cozy, 2016).

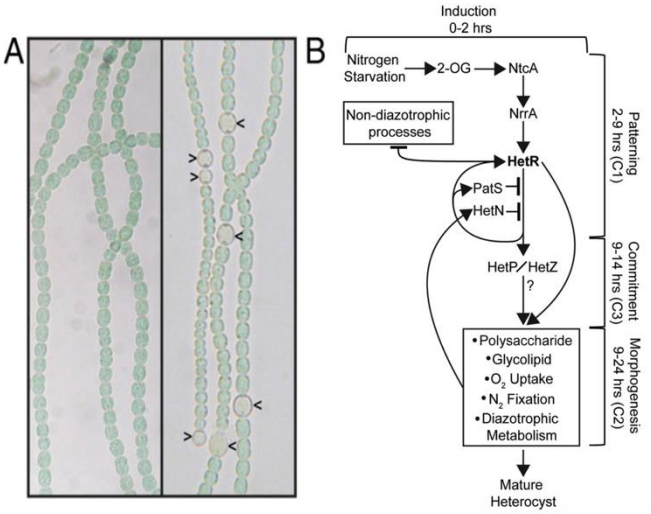

Figure 1-4 Heterocysts differentiation. (A) *Anabaena* PCC 7120 were grown in a combined nitrogen-replete medium that cells were uniform (left), but after nitrogen step-down 24 h, the heterocysts (arrows) were formed about every 10 vegetative cells. (B) Phases of heterocysts differentiation: induction, patterning, commitment, and morphogenesis. The arrows indicate activation and T bars indicate repression (Hanlon & Cozy, 2016).

The metabolite of Krebs cycle 2-oxoglutarate (2-OG), is also called α-ketoglutarate, which serves as carbon skeleton for ammonium assimilation (Vázquez-Bermúdez, Herrero, & Flores, 2000). When cells are starved for the combined nitrogen, 2-OG are accumulated and its accumulation resulted in the complex interactions involved in heterocyst differentiation (Laurent et al., 2005). Two proteins activities respond to the increase of 2-OG levels in the cyanobacteria cells, NtcA and PII (Forchhammer, 2004). NtcA is the main factor; the knockout of *glnB* which encodes for PII would only impair

the heterocysts functions (Laurent et al., 2004). NtcA is a universal transcription factor that controls several downstream genes which are involved in carbon and nitrogen metabolism, including *hetR* and *patS* (Herrero, Muro-Pastor, Valladares, & Flores, 2004). *HetR* is the master positive regulator that functions in heterocyst pattern formation, which is a DNA binding protein and its homodimer by a disulfide bond is required for the DNA binding activity (Black, Cai, & Wolk, 1993; Huang, 2004). HetR binds to the promotor region of *hetR, hepA* and *patS*, suggesting its direct control over these genes (Rivers, Videau, & Callahan, 2014). On the country, *patS* is an inhibitory gene involved in heterocysts development, and studies have shown that the *patS* knockout mutant formed 30% of heterocysts instead of 10% as normal (C. C. Zhang, Laurent, Sakr, Peng, & Bedu, 2006). Some data suggested that the HetR could be the PatS receptor; and HetR DNA-binding activity is inhibited by PatS-5 sequence RGSGR (Huang, Dong, & Zhao, 2004a). Above all, this patterning phase is dynamic that the ratio of HetR and PatS determines the heterocysts development and the ratio is affected by HetR and PatS competitive interactions (I. Y. Khudyakov & Golden, 2004).

The commitment phase is usually irreversible and controlled by a cluster of *hetP* genes and *hetZ* (Hanlon & Cozy, 2016). *Asl1930, alr3234, alr2902* and *hetP* share a functional domain, and have been confirmed that *Asl1930* and *alr3234* delay the commitment; *alr2902* inhibit development while *hetP* push the commitment forward. HetZ has the similar function as *hetP* and both genes are to execute the decision made by *hetR* and *patS* (Patrick Videau et al., 2014).

The morphology and physiology characteristics are distinguished from vegetative cells after the commitment phase, since where heterocysts acquire diazotrophic activity. The

heterocysts are terminal cells that are non-dividing, non-growing cells. Cell division protein, FtsZ, could not be detected in mature heterocysts (C.-c. Zhang, 2000).

1.3.3 Spore-like cells akinetes

Akinetes are spore-like cells in cyanobacteria, which are formed under stress of deficient in phosphate, cold and desiccation (Arizmendi & Serra, 1990). But akinetes are different from traditional spores in several aspects. Compared with the smaller size of spores, usually, akinete size is bigger than vegetative cells, especially akinete size is almost 20 times bigger than vegetative cells in *Anabaena cylindrica.* Akinetes are not entirely resting cells s it typically contains similar proteins, DNA and RNA to vegetative cells. which giving reason to depict that akinetes are cells departing from vegetative cells after cell division, but prior to the stage of chromosome-replication (Fay & Walsby, 1966). The reason of akinete giant size is the conspicuous granulation formed by high concentrations of cyanophycin, which is a nitrogen storage polymer made from an equal ratio of arginine and aspartate (C Peter Wolk et al., 1994). These granules prepare for the germination of akinetes to produce vegetative cells when the environment becomes favorable. Several physical factors increase the akinetes formation frequency, including the temperature fluctuations, the light intensity and concentration of high filterable reactive phosphorus (FRP) (Moore, Donohue, Garnett, Critchley, & Shaw, 2005). Akinetes are usually formed adjacent to of heterocysts, where it is believed that heterocysts induce the neighbor vegetative cells into akinetes. But akinetes and heterocysts share homologues that they have similar polysaccharide composition out layer but vegetative cells do not. Three heterocyst genes have been identified that are also involved in the akinetes formation, such as *hepA*, *devR* and *hetR* (Meeks, Campbell,

Summers, & Wong, 2002). *HepA* is required for the heterocysts polysaccharide layer formation; the cell envelop of heteocysts and akinetes altered in Δ*hepA* mutant of *Anabaena variabilis* (Leganés, 1994). *DevR* is a heterocysts gene that required for the formation of polysaccharide layer, ΔdevR mutant failed to fix nitrogen in aerobic condition but do so in anoxic conditions. DevR did not show essential role to akinetes formation, but increase the akinetes frequency in old culture of *Nostoc sp*. ATCC 29133 with the overexpression of *devR* (Elsie L Campbell, Hagen, Cohen, Summers, & Meeks, 1996). HetR, the mater regulator in heterocyst differentiation, was reported that Δ*hetR* mutant block both the differentiation of heterocysts and akinetes(Leganés, Fernández-Piñas, & Wolk, 1994). An exception was found Δ*hetR* in *N. punctiforme* form large akinete-like and cold resistant cells after the depletion of phosphate (Wong & Meeks, 2002). We do not know much about the mechanisms of akinetes differentiation and formation. One only akinete marker protein AvaK was identified in *Anabaena variabilis* that shows exclusive existence in akinetes (Zhou & Wolk, 2002).

1.4 Omics study in cyanobacteria

The omics study including genomics, transcriptomics, proteomics, and metabolomics provides a comprehensive picture of genetic and regulatory pathways within cells. With the advance of DNA sequencing techniques and related analytic methodology, more cyanobacteria strains were discovered and our knowledge of this group of bacteria has been deepened in multiple angles although challenges still yet to be overcome. In this thesis, genomics and proteomics of *A. cylindrica* are focused on.

1.4.1 Genomics in cyanobacteria

1.4.1.1 The overview of cyanobacteria genome sequencing

The first cyanobacteria genome sequence was reported from *Synechocystis* sp. PCC 6803 in 1996 (Kaneko et al., 1996). With the rapid advancement of next-generation DNA sequence, so far there are over 400 cyanobacteria genome sequences available in public database including NCBI and cyanobase (http://genome.microbedb.jp/cyanobase/). The current completed cyanobacteria genomes account for only 0.6% of available prokaryotic genomes, which is far underrepresented. The main reason is the difficulty to obtain axenic cyanobacteria culture due to the tight associations of cyanobacteria and other symbiont microbes. Multiple chemical and mechanical methods, including washing, filtering, and streaking were employed to exclude other microorganisms but the purifying process is always a time-consuming process and sometimes fail (Waterbury, 2006). Several new sequencing techniques arise, like single-cell genomics and metagenomics in microbial consortia, but still facing methodological challenges and limited access (Davison, Hall, Zare, & Bhaya, 2015). The other challenge is lacking geographical diversity. Current cyanobacteria were deposited under Pasteur Culture Collection (PCC) located at Paris resulted in that majority available cyanobacterial genomes were from European. The consequence is the over-representing of some species, for example, 45 out of 166 genomes in *Prochlorococcus* spp. are belonging to one species - *Prochlorococcus marinus*.

Genomics of the axenic cyanobacteria were routinely acquired from next-generation sequence and sequence assembly. The assembly includes the steps of reading quality control, de novo assembly, scaffolding, assembly statistics and genome annotation.

Various software and servers were developed for assembly purposes, including FastQC, PRINSEQ, SPAdes, CLC and Platanus. The failure of assembly still occurs due to the general reasons that multiple numbers of chromosomes, the occurrence of repeated regions and abundance of mobile genetic elements.

1.4.1.2 The composition of cyanobacteria genome

The cyanobacteria so far sequenced have one or more chromosomes with the sizes ranging from 1.4 to 8.2 Mb, and multiple plasmids up to 12. Polyploidy occurs during exponential growth that the cyanobacteria contain up to 218 chromosomes. Based on the functions, the genome of cyanobacteria can be divided into two categories: core genome and accessory genome. The core genome encodes for indispensable biochemical pathways and the complex protein structures, while the accessory genome is more probably subjected to horizontal gene transfer. As mentioned before, a high abundance of repeated sequences was spotted that increase the difficulty of the genome assembly. The genome content could be changed over time due to adaptation for different conditions or neutral processes. Two adaptation strategies have been revealed from cyanobacteria genomic analyses that genomic expansion through the increase of gene families and genomic reduction through the elimination of genes. The genome size, gene number, and G-C percentage are subjected to change over these adaptations (Larsson, Nylander, & Bergman, 2011).

1.4.1.3 Genomes in differentiated cells

The heterocysts and akinetes are differentiated cells from mother cells - vegetative cells. The heterocysts are terminal nitrogen-fixing cells that no longer divide, and akinetes are spores-like cells that are capable for the germination to generate vegetative cells later.

Some changes of the genomes composition in heterocysts and akinetes varied from vegetative cell have been identified, such as akinete chromosome number is increased up to 450 times; a few *nif* gene elements were deleted in heterocyst genome (Haselkorn, 1992; Sukenik, 2012).

1.4.1.3.1 The genome rearrangements in heterocysts

Heterocysts are nitrogen-fixing cells and the *nif* genes encode for the key enzyme - nitrogenase, which converts atmospheric N_2 into ammonium. But a couple of *nif* genes sequenced from vegetative cell genome were interrupted by some insertion elements that resulted in the disruption of *nif* genes expression and in further nitrogenase synthesis and assembly. In six *Anabaena* and three *Nostoc* strains, an 11 kb element within the *nifD* gene and a 55 kb element within the *fdxN* gene have been confirmed that got deleted in heteocysts using southern analysis (Claudio D Carrasco & Golden, 1995). These rearrangements seem to be exclusive in some heterocystous cyanobacteria that the *nif* genes rearrangements have been confirmed that they do not occur in non heterocystous cyanobacteria, such as *Trichodesmiumm sp.* strain NIBB 1067 (Zehr, Ohki, & Fujita, 1991). These *nif* genes rearrangements have high importance for the acquirement of intact *nif* gene open reading frame and nitrogen fixation activity. But our knowledge of heterocyst genome rearrangement is limited to these couple of *nif* cluster genes, and little known about the heterocyst genomics until this dissertation.

1.4.1.3.2 The multiple genome copies in akinetes

Single akinetes were isolated to study the genome contents compared with vegetative cells in *Aphanizomenon ovalisporum* (Assaf Sukenik, Kaplan-levy, Welch, & Post, 2011). In this species, vegetative cells contain normally 8 genome copies; in contrast,

akinetes have on average of 119 genome copies. Especially some late stage of akinetes have up to 450 genome copies. On average, akinetes have 15-fold higher genome copies than vegetative cells. At the same time, accumulated ribosomes were also observed in akinetes using 16S rRNA probe in *situ* hybridization. The mechanisms of nucleic acid content accumulation in akinetes have not fully understood yet, where inorganic polyphosphate bodies that absent in akinetes but high abundance in vegetative cells were suspected associating with this process. One thing is certain that the multiple genome copies are beneficial for cell division upon germination and probably related to extended survival.

1.4.2 Transcriptomics in cyanobacteria

The most common mechanism for gene expression control is transcriptional regulation. The transcriptome study has advanced our understanding of gene expression and regulation pattern, including sRNA regulators. Genome-wide transcriptional start sites (TSS) map was reported in *Prochlorococcus* and interestingly hundreds of TSS were found within the genes (Voigt et al., 2014). Besides RNA polymerase, transcription factors (TFs) are a family of proteins which initiate and regulate gene transcription. Transcription is initiated by TFs that bind to specific sequences of DNA in the promoter region – transcription factor binding sites (TFBSs). These specific DNA binding sequences are called DNA motifs (Figure 1-5). So far over 27 putative TFs have been reported in cTFbase (Wu et al., 2007), which is a database collection of the TFs in all cyanobacteria. A regulon is a gene group that shares the same regulatory DNA motifs, so usually is controlled by the same regulatory gene that expresses a protein acting as a repressor or activator. OmpR, the archetypal DNA binding protein, binds to different

regulons related with acid resistance in *Escherichia coli* (Quinn, Cameron, & Dorman, 2014). Previous studies in *Cyanothece* 51142 showed a strong functionally associated coregulation in distantly located genes. Also, both the clusters of *nif* and the ribosomal genes are organized as regulons of more than one transcriptional unit (Stockel et al., 2008).

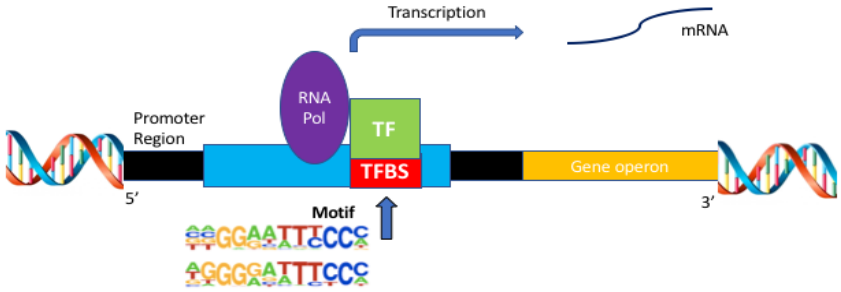

Figure 1-5 Overview of transcription factor binding to the promotor, and association with RNA polymerase to initiate transcription. The TFBSs are conserved DNA motifs among regulons.

Comparative transcriptomes have been widely used to identify the response of cyanobacteria to different environmental factors as well. Cyanobacteria have a high transcriptional expression, especially, antisense transcription from three-quarters of all genes in *Prochlorococcus* has been detected that was substantially higher than other bacteria (Voigt et al., 2014). To study circadian clock, a total of 6766 mRNAs and 1322 proteins have been tested at four-time points during a diurnal light-dark cycle in *Cyanothece* sp. PCC 7822 (Welkie et al., 2014). The strong correlation of antisense gene expression with light was observed and similar robust oscillating expression was found in *Cyanothece* 51142 as well (Stockel et al., 2008). Comparative gene expression of *Nostoc*

puntiforme grown in different medium revealed that the enriched transcriptomes of genes that encoding core metabolic functions that grown in ammonium than steady-state N_2 grown culture, which most up-regulated gene expression are related with heterocyst differentiation and N_2 fixation (E. L. Campbell, Summers, Christman, Martin, & Meeks, 2007). A total of 1,036 and 1,762 transcribed genes were reported during the development of heterocyst without combined nitrogen and in hormogonia that induced by ammonium; commonalities and two distinguished pathways were clarified between these two differentiated structures. Besides, the distinct up-regulation genes that involved with the differentiation of heterocysts, akinetes, and hormogonia were also identified in *Nostoc punciforme* (Christman, Campbell, & Meeks, 2011).

1.4.3 Proteomics in cyanobacteria

Proteomics is a most direct methodology to visualize the metabolic activity within the cells as the proteins are the functional performers. The proteomics techniques mostly used in cyanobacteria are 2DE, two-dimensional gel electrophoresis; nSAF, normalized spectral abundance factor; iTRAQ, isobaric tags for relative and absolute quantitation. As the advance in the proteomics sequence technique, proteomics has been applied to study cyanobacteria in different aspects in the past decade, such as cell structures, protein to protein interactions and the cell's response to different conditions. The cyanobacteria are gram-negative bacteria with their unique characteristics and functions, such as photosynthesis and nitrogen fixation. Proteomics has been identified in different levels of study that helped us advanced our understanding towards this microorganism. Carboxysome is a special structure developed in cyanobacteria to concentrate CO_2 within the confined space. Matthew et. al. isolated the carboxysome and identified its nanoscale

structure and proteomics, which greatly deepened our understanding of this organelle (Faulkner et al., 2017). One fact that circadian clock of cyanobacteria regulates the separation of photosynthesis in the day and nitrogen fixation during the night. The proteomics of *Synechococcus elongatus* PCC 7942 identified from the day and night revealed the multi-protein complexes difference and revealed cell activities during these two phases (Guerreiro et al., 2016). The proteomics tested the response of *Microcystis aeruginosa* to N, P or N and P limitation that N-limitation increased proteins for C metabolism; P limitation reduced C and N assimilation (Yue, Peng, Yin, & Xiao, 2015). Comparative proteomics has also been utilized to study the mechanisms of toxin producing in some types of cyanobacteria by identifying the proteins expression levels in toxin producing and non-toxin producing cyanobacteria (D'Agostino et al., 2016). With the filamentous cyanobacteria, the proteomics among differentiated cells will greatly help us understand their specialized functions which are lacking. Currently, a lot of proteomics databases are available and two of the biggest cyanobacterial database are Cyanobase (Fujisawa et al., 2017) and Cyanobacterial KnowledgeBase (CKB) (Peter et al., 2015). Up to this dissertation, there has been no proteomic study for akinetes although serval proteomic studies have been reported for heterocysts, vegetative cells, carboxysomes.

CHAPTER 2: Unique Proteomes Implicate Functional Specialization across Heterocysts, Akinetes, and Vegetative Cells in *Anabaena cylindrica*

2.1 Abstract

In response to environmental changes, vegetative cells of *Anabaena cylindrica* can differentiate into two other cell types: a heterocyst for oxic N_2-fixation, and an enlarged spore called akinete for stress survival. Akinetes normally differentiate from vegetative cells adjacent to heterocysts. Heterocysts inhibit nearby cells from differentiating into heterocysts but can induce adjacent cells to become akinetes, a rare embryogenetic induction in prokaryotes. The mechanism for a patterned differentiation in *A. cylindrica* has been little studied. Here, we isolated three types of cells from *A. cylindrica* to identify their proteomes using LC-MS/MS.

A total of 1395 proteins were identified, including 664 proteins from akinetes, 751 proteins from heterocysts, and 1236 proteins from vegetative cells. There were 45 proteins (33 novel proteins) found exclusive to akinetes, 57 heterocyst-specific proteins (33 novel proteins), including *nif* gene products, and 485 proteins exclusively in vegetative cells. Our proteomic data suggest that akinetes, unlike the typical spores of bacteria, perform unique biochemical functions that collaborate with both heterocysts and vegetative cells. A HAVe model for collaboration among heterocysts, akinetes, and vegetative cells is proposed to illustrate the metabolic network of cyanophycin and carbohydrates based on the distribution of their biosynthesis-related proteins in three types of cells. Interestingly, cell division proteins, DNA replication proteins, some carboxysomal proteins including RuBisCO and proteins in photosystems I, II were found abundant in heterocysts, the non-dividing cells dedicated exclusively to oxic N_2-fixation.

The identification of the akinete and heterocyst proteomes enables the pursuit of genetic studies into the patterned differentiation of akinetes and heterocysts.

KEYWORDS: cyanobacteria, spores, oxic nitrogen fixation, comparative proteomics, cellular differentiation, HAVe model.

2.2 Introduction

Cyanobacteria are the only prokaryotes capable of oxygenic photosynthesis (Gantt, 2011), and are widely believed to be the ancestors of chloroplasts (Martin et al., 2002). Many cyanobacteria are also capable of photosynthetic fixing atmospheric dinitrogen (N_2) (Kumar, Mella-Herrera, & Golden, 2010). While some cyanobacteria follow a single-cell lifestyle, multicellularity in this group first evolved 2.5 billion years ago (B. E. Schirrmeister, A. Antonelli, & H. C. Bagheri, 2011). Many cyanobacteria are capable of complex biochemical transformations in response to different physicochemical environments. Photosynthesis occurs in light and yields oxygen while N_2 fixation requires a highly reduced environment (Kumar et al., 2010). Unicellular cyanobacteria such as *Cyanothece sp.* ATCC 51142 solve this through a circadian clock to separate photosynthesis and N_2 fixation temporarily into the light and dark periods (Cerveny, Sinetova, Valledor, Sherman, & Nedbal, 2013). Spatial division of labor in multicellular cyanobacteria appears more efficient at energy capture than temporal separation as occurs in unicellular cyanobacteria (Rossetti, Schirrmeister, Bernasconi, & Bagheri, 2010). Some filamentous cyanobacteria can differentiate to form four cell types: photosynthetic vegetative cells, N_2-fixing heterocysts, akinetes, and small motile filaments called hormogonia (Flores & Herrero, 2010; Rippka & Herdman, 1985). Akinetes developed from vegetative cells but are capable of germinating to produce young vegetative cells. Heterocysts develop from vegetative cells to form terminally differentiated, non-dividing cells functionally specialized for oxic N_2-fixation. Heterocysts are formed in filamentous cyanobacteria in response to the depletion of fixed nitrogen (Mitschke, Vioque, Haas, Hess, & Muro-Pastor, 2011). They develop every 10 to 20 cells along the filament

(Kumar et al., 2010), and are larger and more round than vegetative cells. The cell envelope of heterocysts is thicker, and with two additional envelope layers: heterocyst-specific glycolipids (HGL) and an outer polysaccharide layer (HEP). These extra two envelope layers impede the entry of oxygen to protect nitrogenase in the heterocysts (Flores & Herrero, 2010). Heterocysts have diminished levels of pigments, and photosystem II is degraded to shut down O_2-producing reactions (Donze et al., 1972; Thomas, 1970). Thus, the heterocyst creates a micro-oxic environment to house the oxygen-sensitive nitrogenase. However, photosystem I (PS I) is kept intact to generate ATP using light energy for N_2-fixation through cyclic photophosphorylation (Tel-Or & Stewart, 1976; C. P. Wolk & Simon, 1969). Theretofore, nitrogen fixation in heterocysts is a uniquely solar-powered process, which is distinct from N_2-fixation by any other N_2-fixing bacteria. The wall between vegetative cells and heterocysts contains intercellular channels called septosomes, which allow for the exchange of metabolites. Reductants such as sucrose and fixed carbon are obtained from vegetative cells, while heterocysts fix N_2 and provide amino acids to the vegetative cells in a filament (Muro-Pastor & Hess, 2012; Thomas, Meeks, Wolk, Shaffer, & Austin, 1977).

Some cyanobacteria can form akinetes, spore-like cells resistant to desiccation and freezing temperatures, that are able to germinate into new vegetative cells under favorable conditions (Perez, Forchhammer, Salerno, & Maldener, 2015). Unlike endospores of *Bacillus*, akinetes are susceptible to heat and long-term exposure to vacuum (Olsson-Francis, de la Torre, Towner, & Cockell, 2009). Akinetes are larger than vegetative cells (Singh & Montgomery, 2011) and contain large quantities of reserve products, mainly glycogen (Sarma, Ahuja, & Khattar, 2004) and the nitrogen storage

polypeptide polymer cyanophycin (A. Sukenik et al., 2015). Akinetes are enveloped in a thick protective coat (Meeks et al., 2002). They begin to differentiate from vegetative cells during the late exponential phase of growth. Increasing culture density and decreasing light penetration accelerate the formation of akinetes. Intriguingly akinetes normally form adjacent to heterocysts in *Anabaena cylindrica* (Figure 2-1A), implying that these akinetes may play a role in the transportation of N and C between vegetative cells and heterocysts besides their survival role in stress conditions. The significant morphological and metabolic changes observed in heterocysts and akinetes suggest unique phenotypes underpinned by complex regulatory pathways.

Many genes have been reported to be involved in regulating heterocyst differentiation. HetR is a master regulator specifically required for heterocyst differentiation (Buikema & Haselkorn, 1991; Huang, Dong, & Zhao, 2004b; Zhou et al., 1998). Several regulatory genes *nrrA* (Ehira & Ohmori, 2011), *ccbP* (Y. Hu et al., 2011), *hetN (Higa et al., 2012), hetF, patA* (Risser & Callahan, 2008), *patN (Risser, Wong, & Meeks, 2012), patU (Meeks et al., 2002), hetZ (W. Zhang et al., 2007), patS* (H. X. Hu et al., 2015; Yoon & Golden, 1998) and *hetP* (P. Videau et al., 2016) were also found to play very important roles during heterocyst differentiation and its pattern formation. The heterocyst-specific NsiR1 small RNA was recently discovered as an early marker in this process (Muro-Pastor, 2014). Although these genes are clearly involved in the regulation of heterocyst development, their biochemical functions remain to be determined. Unfortunately, the genetic regulation of akinete formation is completely unknown. So far, the only reported akinete-specific protein is AvaK from *Anabaena variabilis* (Zhou & Wolk, 2002). There has been no proteomic study for akinetes to date although a quantitative shotgun

proteomics study of heterocysts was reported for *Anabaena sp.* PCC 7120 (Agrawal et al., 2014; Ow et al., 2008; Panda, Basu, Rajaram, & Kumar Apte, 2014; Pandey, Rai, & Rai, 2012) and *Nostoc punctiforme* (Liang et al., 2012; Sandh, Ramstrom, & Stensjo, 2014).

A. cylindrica can form N_2-fixing heterocysts under both depleted and replete nitrate conditions (Meeks, Wycoff, Chapman, & Enderlin, 1983), which is different from other heterocyst-forming cyanobacteria, such as *Anabaena* sp. strain PCC 7120 (Borthakur & Haselkorn, 1989), *Anabaena variabilis* (Thiel, Lyons, Erker, & Ernst, 1995) and *Nostoc punctiforme* (Summers & Meeks, 1996), whose vegetative cells can differentiate into heterocysts only in response to deprivation of combined nitrogen. Moreover, vegetative cells of *A. cylindrica* can also differentiate into akinetes (arrowheads labeled A), spore-like cells for stress survival. Akinetes (15 ~20 μm in length) are about 10 times larger than vegetative cells, and normally develop adjacent to heterocysts within the same filament (Figure 2-1A), providing a rare opportunity to elucidate what appears to be an embryogenetic induction in a prokaryote (C P Wolk, 1966). Unfortunately, the differentiation of akinetes, heterocysts as well as akinete juxtaposition to heterocysts have not heretofore been studied genetically due to the lack of a genetic transformation method for this organism.

We sought to characterize the phenotype of akinetes of *A. cylindrica* through proteomic analysis, contrasting it to the phenotype of heterocysts and vegetative cells. *A. cylindrica* ATCC 29414 was selected for this study because it differentiates readily into both heterocysts and akinetes in dilute Allen and Arnon medium (AA/8) without combined nitrogen (N. T. Hu, Thiel, Giddings, & Wolk, 1981). Its akinetes are large and readily

separated from heterocysts and vegetative cells. Our proteomic data suggest that akinetes, unlike the typical spores of bacteria, perform unique biochemical functions that collaborate with both heterocysts and vegetative cells.

2.3 Material and Methods

2.3.1 Isolation of akinetes and heterocysts

Isolation of akinetes and heterocysts was based upon the CsCl density gradient centrifugation (C. P. Wolk, 1968) with the following modification. Briefly *A. cylindrica* ATCC 29414 was grown in nitrate free AA/8 medium under continuous light (60 $\mu E/m^2/s$, 150 rpm, 30°C) for 30 days ($OD_{700}\approx0.15$) to allow heterocyst and akinete development. Cultures were harvested (6,400 × *g* 15 min, 4°C), resuspended in ddH_2O, and the vegetative cells were disrupted by passing the suspension through a Nano DeBEE-30 high pressure homogenizer (BEE International) at 4,500 psi and then at 5,000 psi. Akinetes and heterocysts were sedimented (4,000 × *g* 10 min, and 4°C) and washed four times with ddH_2O to remove the vegetative cell debris. There were two distinct layers formed in the last wash pellet. The upper layer was suspended in 1.55 g/mL CsCl density solution, and transferred into in an ultracentrifugation tube. The bottom layer was suspended in 1.45 g/mL CsCl and carefully transferred on-top of the upper layer suspension in the same ultracentrifugation tube. Two distinct fractions were collected from the first CsCl density gradient centrifugation (17,000 × *g*, 60 min, 4°C in a fixed angle MLA-55 rotor, Beckman Coulter), each was sedimented (4000 × *g*, 30 min), and washed with 3x ddH_2O. The heavy fraction was suspended in 1.45 g/ml CsCl solution and re-centrifuged as before. The light fractions from the first and second centrifugations were pooled, suspended in 1.45 g/ml CsCl solution, and re-centrifuged. The supernatant

fraction from this third centrifugation was suspended in 1.3 g/ml CsCl solution and re-centrifuged. The resultant pellets from the second and third centrifugations (containing highly purified akinetes) and pellet from the 4th centrifugation (containing highly purified heterocysts) were washed with ddH_2O as above. The purity of the heterocysts (99.52 ± 0.48%) and akinetes (96.17 ± 0.72%) was examined by differential interference contrast microscopy (AX70 upright, Olympus).

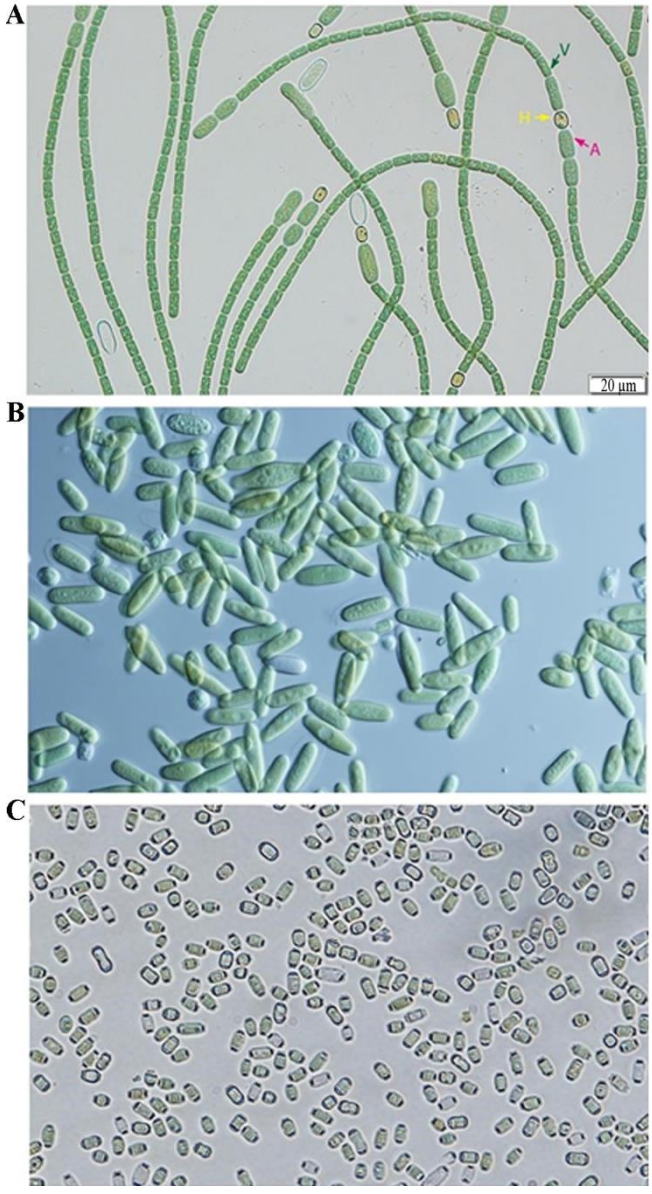

Figure 2-1 *Anabaena cylindrica* ATCC 29414 has three types of cells in filaments (A).

The purity of the isolated akinetes (B) and heterocysts (C) was analyzed by differential interference contrast microscopy. A: Akinetes or developing akinetes, H: Heterocyst, V: Vegetative cell. Scale bar for panel A-C is 20 μm. The purity of the heterocysts and akinetes was 99.52 ± 0.48% and 96.17 ± 0.72%, respectively.

2.3.2 Total protein extraction and SDS-PAGE purification

The purified heterocysts or akinetes were suspended in Phosphate Buffered Saline (PBS, 0.01M, pH7.0) containing 1% N-lauroyl sarcosine and protease Inhibitors [Complete, Mini Protease Inhibitor Cocktail Tablets (Roche) at 1 tablet per 10 mL]. Cells were disrupted on ice by ultra-sonication (Branson digital sonifier 450) for 12 x 5 s bursts with 15 s intervals at an amplitude of 60%. Cell lysates were collected (13,000 × *g* 20 min) respectively. To extract total proteins from the vegetative cells, *A. cylindrica* ATCC 29414 cells grown in 1.0 litter of AA/8N (N. T. Hu et al., 1981) for 8 days (OD700≈0.042) were harvested (6,400 × *g*, 15 min, 4°C) and resuspended in PBS buffer (PBS, 0.01M, pH7.0) containing 1% N-lauroyl sarcosine and protease inhibitors (Complete, Mini Protease Inhibitor Cocktail Tablets (Roche) at 1 tablet per 10 mL). Vegetative-cell lysate was obtained by passing the cell suspension through a Nano DeBEE-30 High-Pressure Homogenizer (BEE International) once at 1,000 psi (only the vegetative cells were disrupted at this pressure), and removing unbroken cells via centrifugation (4,000 × *g*, 10 min, 4°C).

Total proteins from each type of cell lysate were precipitated with 10% trichloroacetic acid (TCA) overnight at 4°C, sedimented (16,000 × *g*, 30 min), the pellets were washed three times with 80% methanol and three times with 80% acetone. The pellets were resuspended in 300 μL of sodium dodecyl sulfate-polyacrylamide gel electrophoresis

(SDS-PAGE) loading buffer containing 1% N-lauroyl sarcosine, boiled for 3 min, clarified by centrifugation at 16,000 × *g* for 20 min at 25°C, and 70 µL was subjected to a 12% SDS-PAGE (Bio-Rad Mini-PROTEAN® Comb, 5-well, 1.0 mm) at 200 mV for approximately 15 min, until all the proteins just entered the resolving gel. The gel was stained with Coomassie Brilliant Blue R-250 for band excision and analysis.

2.3.3 In-gel tryptic digestion and protein identification by LC-MS/MS

The protein gel bands were excised and in-gel tryptic digestion was performed according to Shevchenko (Shevchenko, Wilm, Vorm, & Mann, 1996) with the following modifications. Briefly, gel slices were dehydrated with acetonitrile (ACN) for approximately 5 min incubation and repeated this process until they appear to shrink in size and show a chalk white color. The time required and number of washes vary with gel size and composition. The chalk white color gel was then incubated with 100 mM ammonium bicarbonate (NH_4HCO_3) containing 10 mM dithiothreitol (DTT, pH ≈ 8.0) for 45 min at 56°C, dehydrated again and incubated with 100 mM NH_4HCO_3 containing 50 mM iodoacetamide for 20 min in the dark, and then washed with 100 mM NH_4HCO_3 and dehydrated again. Approximately 50 µL trypsin solution (0.01 µg/µL sequencing grade modified trypsin (Promega, #V5111) in 50 mM NH_4HCO_3) was added to each gel slice so that the gel was completely submerged, and then incubated at 37°C for overnight. The tryptic peptides were extracted with 60% ACN/1% TCA from the gel by water bath sonication (Aquasonic 150T sonicating water bath which puts out 135W. Sonication is done 2 x 20s) and concentrated in a SpeedVac to 2 µL.

For heterocyst and akinete samples, the extracted peptides were re-suspended in 20 µL 2% ACN/0.1% trifluoroacetic acid (TFA), 10 µL was injected by a nanoAcquity Sample

Manager and loaded for 5 min onto a Symmetry C18 peptide trap (5 µm, 180 µm x 20 mm) (Waters) at 4 µL/min in 2% ACN/0.1% Formic Acid. The bound peptides were eluted onto a BH130 C18 column (1.7 µm, 150 µm x 100 mm, Waters) using a nanoAcquity UPLC (Waters) (Buffer A = 99.9% Water/0.1% Formic Acid, Buffer B = 99.9% Acetonitrile/0.1% Formic Acid) with a gradient of 5% B to 30% B over 228 min, ramping to 90% B at 229 min and holding for 1 min, and then ramping back to 5% B at 231 min, and holding for equilibration prior to the next injection for a total run time of 240 min. The eluted peptides were sprayed into an LTQ-FT-ICR Ultra hybrid mass Spectrometer (Thermo Scientific) using an ADVANCE nanospray source (Bruker-Michrom). Survey scans were taken in the FT (25,000 resolution determined at *m/z* 400) and the top five ions in each survey scan were then subjected to automatic low energy collision-induced dissociation (CID) in the LTQ.

For the vegetative-cell sample, 5 µL of the extracted peptide suspension was injected (to the sample loop which is then backflushed using solvent A directly to the column) by EASYnLC and the peptides separated through an Acclaim PepMap RSLC column (0.075 mm x 150 mm C18, Thermo Scientific) with the same gradient as above. The eluted peptides were sprayed into a Q Exactive hybrid quadrupole-Orbitrap mass spectrometer using a Nanospray Flex™ Ion Sources (Thermo Scientific). Survey scans were taken in the Orbi trap (35,000 resolution determined at *m/z* 200) and the top ten ions in each survey scan were then subjected to automatic higher energy collision-induced dissociation (HCD) with fragment spectra acquired at 17,500 resolution (by convention this is a dimensionless measurement).

For protein identification, the resulting MS/MS spectra were converted to peak lists using Mascot Distiller, v2.5.1.0 (www.matrixscience.com) and searched against a protein sequence database containing *A. cylindrica* ATCC 29414 (http://scorpius.ucdavis.edu/gmod/cgi-bin/site/*Anabaena*02?page=gblast), *A. cylindrica* PCC 7122 entries (http://www.ncbi.nlm.nih.gov/genome/?term=*Anabaena*+cylindrica+7122+genome) and common laboratory contaminants downloaded from www.thegpm.org. All searches were performed using the Mascot searching algorithm, v 2.4. The Mascot output was then analyzed using Scaffold, v4.3.4 (www.proteomesoftware.com) to probabilistically validate protein identifications at 1% FDR. The quantification value was calculated using Normalized Total Spectra (For details, see Supplementary Materials). The mass spectrometry proteomics data have been deposited to the ProteomeXchange 213 Consortium via the PRIDE (Vizcaino et al., 2016) partner repository with the dataset identifier PXD006041.

2.4 Results

2.4.1 Proteomic analysis of heterocysts, akinetes, and vegetative cells

To unlock the cellular function of akinetes and the protein network among akinetes, heterocysts, and vegetative cells in *A. cylindrica* ATCC 29414, we performed proteomics through LC-MS/MS. A total of 12616 tryptic peptides were collected and 1426 proteins were identified, including 1395 ORF proteins from *A. cylindrica* ATCC 29414, 14 proteins from common laboratory contaminants, 14 decoy proteins for determination of the false discovery rate, and 3 ORFs (Anacy_0074, Anacy_3940 and Anacy_5216) in *A.*

cylindrica PCC 7122 matched to intergenic regions of *A. cylindrica* ATCC 29414 genome.

Our LC-MS/MS proteomics analysis identified 664 proteins from akinetes, 751 from heterocysts, and 1236 from vegetative cells, with 448 proteins common to all three cell types. There were 45 akinete-specific, 57 heterocyst-specific, and 485 vegetative cell-specific proteins (Figure 2-2). Interestingly, phycocyanin alpha (ORF: 3613) and beta (ORF: 3614) subunits, allophycocyanin beta subunit (ORF: 1908), phycobilisome protein (ORF: 1909), beta subunit of mitochondrial ATP synthase (ORF: 3788), translation elongation factor 1A (EF-1A/EF-Tu, ORF: 5853) and ribulose 1,5-bisphosphate carboxylase (RuBisCO) large (ORF: 6007) and small subunit (ORF: 6009) were among most abundant proteins in all three cell types.

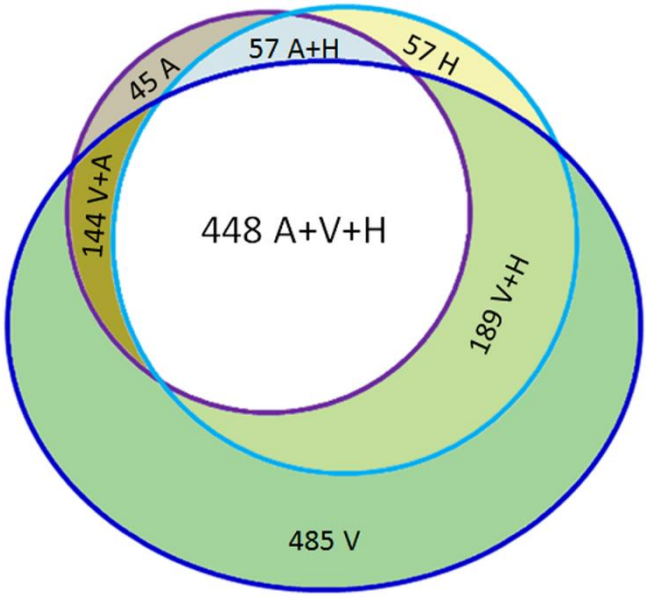

Figure 2-2 Venn analysis showing the proteomic profiles from akinetes (A), heterocysts (H), and vegetative cells (V). **45 A:** 45 proteins were detected exclusively to akinetes; **57 H:** 57 proteins were found to be heterocyst-specific; **485 V:** 485 proteins were found exclusively in vegetative cells; **57 A+H:** 57 proteins were detected in both akinetes and

heterocysts; **144 V+A:** 144 proteins were detected in both vegetative cells and akinetes; **189 V+H:** 189 proteins were detected in both vegetative cells and heterocysts; and **448 A+V+H:** 448 proteins were found to be common to all three cell types.

2.4.2 Nitrogen fixation in heterocysts

The nif genes encode subunits of nitrogenase for reducing atmospheric N2 to ammonia. Other heterocyst-specific genes encode proteins involved in regulating heterocyst development and N2 fixation, and inactivation of these genes showed diminished or ceased diazotrophic growth in the presence of oxygen due to impaired nitrogenase activity, or forming no or dysfunctional heterocysts (Lechno-Yossef, Fan, Wojciuch, & Wolk, 2011). The above genes are collectively called 'FOX' genes (incapable of N2-fixation in the presence of oxygen) (Lechno-Yossef et al., 2011). LC-MS/MS identified 27 FOX proteins (Appendix Table 1) and 57 heterocyst-specific proteins. Heterocysts had 19 Fox proteins but eleven were also found in akinetes (Appendix Table 1).

2.4.3 Distinct distribution of Photosystem I and II proteins

PS I and PS II are well-known hallmarks primarily associated with the photosynthetic characteristics of vegetative cells**.** Of 20 photosystem proteins identified, all were present in vegetative cells, consistent with vegetative cells bearing both PS I and PS II proteins for fully functional photosynthesis and electron transfer (Appendix Table 1). The protochlorophyllide reductase subunit N catalyzing the penultimate step of chlorophyll biosynthesis (Yamazaki, Nomata, & Fujita, 2006), PS I assembly protein Ycf3 (Wilde, Lunser, Ossenbuhl, Nickelsen, & Borner, 2001), and PS II reaction center Psb28 protein (Dobakova, Sobotka, Tichy, & Komenda, 2009) were unique to vegetative cells of *A. cylindrica*. Notably, there were several PS I and PS II proteins in high abundance in both

akinetes and heterocysts respectively (Appendix Table 1), e.g., PS I P700 apoprotein A2, PS II 44 kDa subunit reaction center protein, and PS II chlorophyll-binding protein CP47, suggesting that PS I and PS II may function partially in both heterocysts and akinetes.

2.4.4 Akinete-specific protein AcaK43 (ORF: 1647) in *A. cylindrica*

The first reported akinete-specific protein was AvaK from *Anabaena variabilis* (Zhou & Wolk, 2002). AvaK homolog AcaK43 (ORF: 1647) was among the top 20 most abundant proteins in *A. cylindrica* akinetes (85.25 counts) and heterocysts (90.73 counts), while only trace amount was detected in vegetative cells, which is consistent with the previous report in *A. variabilis* that AvaK is an akinete marker protein (Zhou & Wolk, 2002). However, our proteomic data showed that AcaK43 was abundant in heterocysts of *A. cylindrica*. Fluorescence of GFP (green fluorescent protein) from P_{acaK43}-*gfp* (promoter of *acaK43* fused to *gfp)* originates primarily in both akinetes and heterocysts of *A. cylindrica* (Zhou et al, unpublished observation). Furthermore, proteins homologous to AapN, Hap, Aet identified as akinete-specifically expressed genes using differential display at mRNA level in *Nostoc punctiforme* (Argueta, Yuksek, Patel, & Summers, 2006) were below the limit of detection in our proteomic study. The presence of AcaK43 in both akinetes and heterocysts suggest that *A. cylindrica* is distinct from *A. variabilis* and *N. punctiforme* although they all are akinete/heterocyst-forming cyanobacteria.

2.4.5 DNA/RNA/protein biosynthesis patterns among akinetes, heterocysts, and vegetative

LC-MS/MS identified a number of proteins related to nucleotide synthesis, DNA packing and repair, RNA and protein synthesis, and cell division within akinetes (Appendix Table 2). Nucleotide synthesis related protein (phosphoribosylformylglycinamidine synthase II)

was found to be akinete-specific. Although DNA polymerases, single-strand binding protein (SSBP) and DEAD/DEAH box helicase domain-containing protein involved in DNA replication were not detected in akinetes, DNA gyrase subunit A and the DNA gyrase modulator Peptidase U62 were only found in akinetes. Akinetes have more chromosome copies per cell than in vegetative cells (A. Sukenik, Kaplan-Levy, Welch, & Post, 2012), so DNA gyrase might play an important role in DNA wrapping and packaging (Gore et al., 2006). Furthermore, DNA gyrase in akinetes may minimize the potential damage caused by light energy to these resting cells (Napoli et al., 2004). Moreover, a similar distribution pattern of DNA replication proteins was seen in heterocysts.

DNA-directed RNA polymerase subunits alpha-, beta, beta′ and gamma were abundant in akinetes. However some proteins required for transcription (like RNA polymerase sigma factor) and translation (like signal recognition particle protein, some tRNA and ribosomal proteins) were only found in heterocysts and vegetative cells (Appendix Table 2), suggesting a very active transcription and translation occurred in heterocysts and vegetative cells. Thirteen out of 25 tRNA synthetases and 23 out of 50 ribosomal proteins were absent in akinetes. Most transcriptional regulators were also undetectable in akinetes. These data imply that akinetes retain a less active transcription machinery and a very weak translational capability.

The proteomics data showed that amidohydrolase 2 (ORF: 3947) and taurine catabolism dioxygenase TauD/TfdA (ORF: 3933) were the third and fifth most abundant akinete-specific proteins, while the other two amidohydrolases were vegetative cell-specific (Appendix Table 2). TauD/TfdA, which can degrade taurine and be a source of sulfur

(Shen et al., 2007), was found in akinetes as well. However, phosphoadenylylsulfate reductase, involved in sulfur (Wang et al., 2004) and pyrimidine metabolism (http://www.genome.jp/kegg-bin/show_pathway?ec00240 +1.8.1.9), was absent in akinetes. Certain enzymes involved in amino acid metabolism were only found in heterocysts and/or vegetative cells, such as arginine (Arg) biosynthetic enzyme, acetylglutamate kinase (Ramon-Maiques, Marina, Gil-Ortiz, Fita, & Rubio, 2002), and lysine biosynthetic enzyme diaminopimelate epimerase (Hor et al., 2013). Interestingly, saccharopine dehydrogenase required for lysine degradation (Serrano, Figueira, Kiyota, Zanata, & Arruda, 2012) was only present in akinetes.

2.4.6 Cell division

Heterocysts, terminally differentiated N_2-fixing cells, do not divide and need not pass DNA information to the next generation, which is consistent with the absence of key DNA replication enzymes (DNA polymerase) in heterocysts. Similarly, no DNA polymerases were detected in akinetes, suggesting that akinetes are not dividing either. Akinetes of *A. cylindrica* store twice as much DNA and 10-fold more protein than vegetative cells (Simon, 1977), preparing them for germination when environmental conditions become favorable. Surprisingly, septum formation protein Maf (Briley, Prepiak, Dias, Hahn, & Dubnau, 2011), participating in cell division in *Bacillus subtilis*, was found to be heterocyst-specific. Cell division protein FtsZ (Bi & Lutkenhaus, 1990) and septum site-determining protein MinD (Maurya, Modi, & Misra, 2016) were more abundant in heterocysts and akinetes than in vegetative cells (Appendix Table 3). We speculated that, instead of involvement in cell division, these proteins might be critical in maintaining septum homeostasis among heterocysts, akinetes, and vegetative cells.

2.4.7 Heterocyst-specific envelope glycolipid and lipopolysaccharide lipid A

Cyanobacterial heterocysts provide a micro-oxic environment to support the oxygen-labile nitrogenase fixing N_2 in an oxic milieu. The heterocyst glycolipid (HGL) layer is an important part of the system for maintaining a micro-oxic environment in heterocysts (Murry & Wolk, 1989). Our proteomics study identified multiple heterocyst-specific proteins required for synthesis, export, and deposition of envelope polysaccharides and glycolipids (Appendix Table 3). For instance, we found polyketide synthase thioester reductase subunit HglB (Fan et al., 2005), an enzyme for synthesizing glycolipid aglycones. DevA and DevB are two components of DevBCA exporter (Fiedler, Arnold, Hannus, & Maldener, 1998) necessary for the formation of the laminated layer of heterocysts (Zhou & Wolk, 2003). Hexapeptide repeat-containing transferase is a sugar transferase which might play a critical role in synthesizing different sugars from the fixed carbon source provided by adjacent vegetative cells (Vaara, 1992). Furthermore, glycosyl transferase (HglT, ORF: 3521) required to glycosylate the glycolipid aglycone (Awai & Wolk, 2007) was present only in heterocysts and akinetes (Appendix Table 3). ORF: 2637 and ORF: 2638, orthologs of LpcC and Omp85 involved in lipopolysaccharide lipid A biosynthesis to form a permeability barrier at the outer membrane (Nicolaisen, Hahn, & Schleiff, 2009), had different distribution, with high abundance of Omp85 in akinetes, supporting the hypothesis that the lipopolysaccharide layer plays an important role in increasing stress tolerance of akinetes in *A. cylindrica*.

2.4.8 S-layer proteins and ATP-binding cassette (ABC) transporter

The *A. cylindrica* genome encodes seven S-layer domain-containing proteins (Supplementary Figure S2) and two S-layer like proteins. S-layer proteins can be self-

assembled to form an array on the surface of the cell (Smarda, Smajs, Komrska, & Krzyzanek, 2002). They have multiple functions, including the maintenance of cell integrity, a permeability barrier, pathogenesis, and immune response (Gerbino, Carasi, Mobili, Serradell, & Gomez-Zavaglia, 2015). Our LC-MS/MS identified all nine S-layer proteins (Appendix Table 3, Figure 2-3). Most S-layer proteins, e.g., all4499 and alr4550 (Oliveira et al., 2015), along with other extracellular proteins, such as FG-GAP repeat-containing protein HesF (Oliveira et asl., 2015) (Appendix Table 3), have been identified as exoproteins. The abundance of these nine S-layer proteins was different, with one S-layer protein (ORF: 1127) unique to akinetes, and two other S-layer proteins (ORF: 5127 and ORF: 2780) absent in vegetative cells (Appendix Table 3), implying unique functionality associated to different cell types.

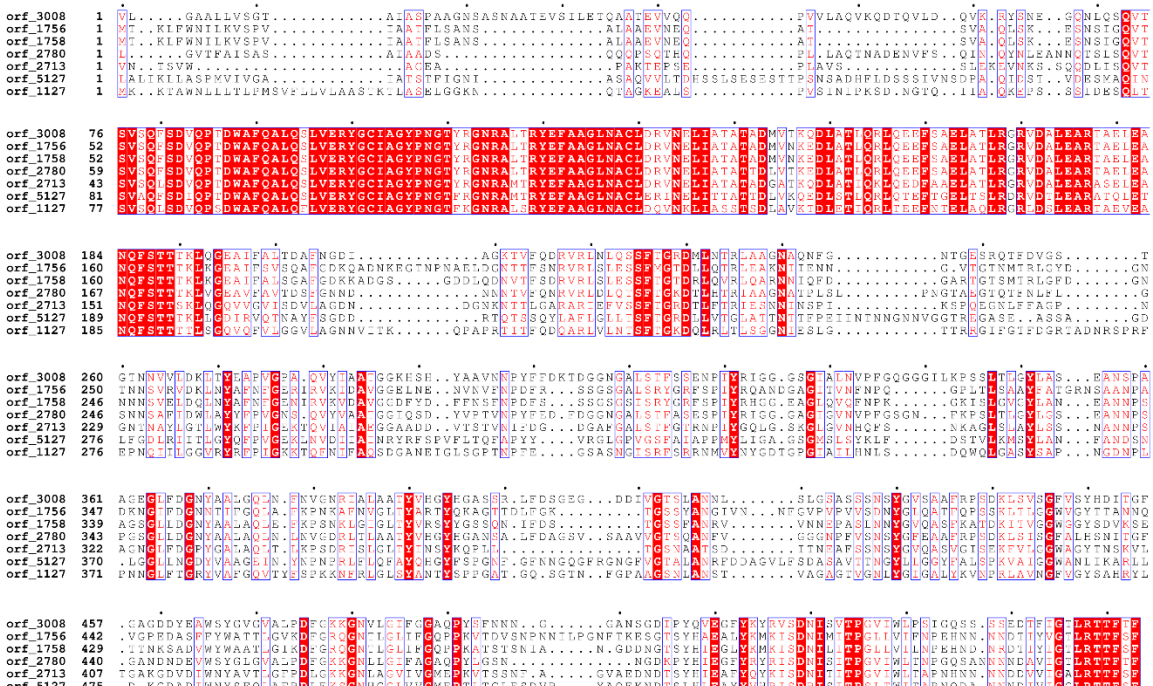

Figure 2-3 The alignment of six S-layer proteins identified in proteomics.

2.4.9 Polysaccharide and Peptidoglycan in cyanobacterial cell wall

We identified a total of 17 proteins involved in peptidoglycan and lipopolysaccharides (LPS) formation, among them, 7, 10, and 16 proteins related to peptidoglycan and lipopolysaccharides were found in akinetes, heterocysts and vegetative cells, respectively (Appendix Table 3). Most cyanobacteria have an additional polysaccharide layer in the cell envelope (Cardemil & Wolk, 1976, 1979). S-layer proteins are anchored to the cell surface through non-covalent interactions with cell surface structures, usually containing LPS (Gandham, Nomellini, & Smit, 2012). UDP-glucose/GDP-mannose dehydrogenase, which takes part in the synthesis of LPSs (Muszynski, Laus, Kijne, & Carlson, 2011) was found in heterocysts and vegetative cells, but absent in akinetes. Notably, orthologs of Alr2887 (ORF: 153) involved in heterocyst-specific glycolipid export and All4388 (ORF: 1651) involved in heterocyst envelope polysaccharide deposition (Maldener, Hannus, & Kammerer, 2003) were shown more abundant in akinetes and heterocysts (Appendix Table 3), suggesting a role in envelope formation of heterocysts and akinetes.

2.4.10 Glycogen serves as a form of energy storage

Glycogen is a multibranched polymer of glucose serving as the major carbon storage in cyanobacteria (Diaz-Troya, Lopez-Maury, Sanchez-Riego, Roldan, & Florencio, 2014). Glycogen biosynthesis is coupled to photosynthesis, and its conversion into glucose in the dark is necessary to maintain cell metabolism. ADP-glucose pyrophosphorylase (AGP) and glycogen synthase are required for synthesis of glycogen. Interestingly, glycogen synthase was highly abundant in akinetes and rare in vegetative cells. Neither akinetes nor heterocysts contained the five proteins involved in glycogen degradation (Table 2-1).

Table 2-1 Proteins involved in metabolism of cyanophycin, glycogen, and sucrose in akinetes, heterocysts, and vegetative cells (enzymes responsible for anabolism are highlighted in grey)

ORF	Annotation	Akinete	Heterocyst	Veget. Cell
		Normalized quantitative value		
Cyanophycin/arginine				
1510	cyanophycin synthetase	2.66	0	0
1272	Cyanophycinase, Serine peptidase, MEROPS family S51	7.99	4.90	0.45
3212	putative cyanophycinase	0	12.26	0
1511	Cyanophycinase	0	9.81	2.26
4256	isoaspartyl dipeptidase, peptidase T2, asparaginase 2	0	0	4.96
2185	acetylglutamate kinase	0	9.81	0.90
1480	N-acetyl-gamma-glutamyl-phosphate reductase	10.66	19.62	5.86
3400	Nitrogen regulatory protein P-II (GlnB, GlnK)	5.33	29.43	5.86
Sucrose/trehalose				
4634	sucrose synthase SuS-B	21.31	0	0
3602	sucrose synthase SuS-A	15.98	0	0
4573	neutral invertase InvB	0	0	0.45

3842	sucrose phosphatase SPP	0	7.36	1.35
238	malto-oligosyltrehalose synthase	5.33	0	0
Glycogen				
1048	glycogen synthase	21.31	12.26	1.35
2144	phosphoglucomutase/phosphomannomutase	13.32	12.26	20.30
3396	phosphoglucomutase/phosphomannomutase alpha/beta/subunit	5.33	2.45	1.80
6292	1,4-alpha-glucan-branching enzyme	2.66	0	0.45
5134	glycogen debranching enzyme GlgX	0	0	0.45
341	Phosphoglycerate/bisphosphoglycerate mutase	0	0	0.45
5891	Phosphoglycerate mutase	0	0	0.45
6104	phosphoglycerate mutase, 2,3-bisphosphoglycerate-independent	0	0	3.61
5419	phosphoglycerate mutase	0	0	0.90

2.4.11 Cyanophycin and β-aspartyl-arginine

Cyanophycin (CpG), or multi-L-arginyl-poly-L-aspartic acid granule polypeptide, is a non-ribosomally produced amino acid polymer composed of an aspartic acid (Asp) backbone and Arg side groups. In heterocysts, nitrogenase converts N_2 to ammonia and then forms glutamine (Gln). Gln can serve as ammonium donor for the synthesis of Asp by aspartate aminotransferase, also known as glutamic oxaloacetic transaminase (Xu et al., 2015). Gln is also the precursor for biosynthesis of Arg and proteins involved in Arg

biosynthesis, and it was found in high amount in heterocysts, such as acetylglutamate kinase (Minaeva, Forchhammer, & Ermilova, 2015; Pang et al., 2015), ArgL (ORF: 1480) (Leganes, Fernandez-Pinas, & Wolk, 1998), and nitrogen regulatory protein P-II GlnB (ORF: 3400) (Llacer et al., 2007; Paz-Yepes, Flores, & Herrero, 2009) (Table 2-1). Asp and Arg are further condensed by cyanophycin synthetase into CpG (Ziegler et al., 1998). This nitrogen storage molecule can be degraded by cyanophycinase (Picossi, Valladares, Flores, & Herrero, 2004) to produce β-aspartyl-arginine. Cyanophycin synthetase was below the limit of detection in heterocysts, but all three putative cyanophycinases in the *A. cylindrica* genome were present in high amounts (Table 2-1), supporting a previous finding that cyanophycinase activity is high in heterocysts (Gupta & Carr, 1981). Asp and Arg can also be transported into akinetes for further condensation into CpG by cyanophycin synthetase (ORF: 1510) and stored. CpG can be degraded by cyanophycinase (ORF: 1272) (Table 2-1) to support growth of other cells in the filament and/or germination in a favorable environment.

2.4.12 Sucrose as a reducing power for N_2 fixation and compatible solute

Sucrose, a universal vehicle of reduced carbon in plants, appears to have a similar role within the diazotrophic cyanobacterial filament (Kolman, Nishi, Perez-Cenci, & Salerno, 2015). Sucrose synthesized by sucrose phosphate synthase and sucrose phosphate phosphatase (SPP) is believed to occur in *Anabaena* strains (Cumino, Curatti, Giarrocco, & Salerno, 2002). Sucrose is then transported into heterocysts (Juttner, 1983) and further hydrolyzed by a specific invertase (InvB) (Lopez-Igual, Flores, & Herrero, 2010; Vargas, Nishi, Giarrocco, & Salerno, 2011). The bidirectional enzyme sucrose synthase SuS-A, on the other hand, exhibited optimal activity at pH 7.5-8.2 in the sucrose-synthesis

direction and at pH 5.9-6.5 in the reverse direction (Porchia, Curatti, & Salerno, 1999). Our proteomic data identified SPP (ORF: 3842) (Cumino, Ekeroth, & Salerno, 2001) present in heterocysts and vegetative cells, but not in akinetes. More strikingly, we detected both SuS-A (ORF: 3602) and SuS-B (ORF: 4634) in high amount, but no invertase in akinetes. Invertase in heterocysts was below the limit of detection in heterocysts. We speculated that the high amount of sucrose synthase present in akinetes might be involved in breaking down sucrose transported from vegetative cells for synthesizing rese0rve glycogen (Perez, Forchhammer, Salerno, & Maldener, 2016), polysaccharides to build akinete envelope, and/or for synthesizing trehalose as an osmoprotectant (Sakamoto et al., 2009) by akinete-specific malto-oligosyltrehalose synthase (ORF: 238) orthologous to All0167 (Higo, Katoh, Ohmori, Ikeuchi, & Ohmori, 2006). Trehalose may play a role in the long-term survival of akinetes under dry conditions.

2.5 Discussion

Some filamentous cyanobacteria can differentiate nitrogen-fixing cells called heterocysts. Normally 2 ~ 10% of vegetative cells develop into heterocysts. In *A. cylindrica*, vegetative cells adjoining heterocysts develop into akinetes (Figure 2-1A). The vegetative cells capture sunlight energy to fix CO_2 and heterocysts carry out solar-powered N_2-fixation. Although akinetes are known as spore-like structures for survival under the unfavorable condition, our proteomic data indicate that akinetes may also play an active role during filamentous growth. Based on the distribution of cyanophycin, glycogen, and sucrose biosynthesis-related proteins, a putative network for fixed nitrogen and carbohydrate among Heterocysts, Akinetes and Vegetative cells, or designated HAVe

model, is proposed for *A. cylindrica* (c). This is the first comprehensive comparison of proteins of akinetes, heterocysts and vegetative cells of *A. cylindrica*. These findings support new insight into the metabolic differences and increase our understanding of the roles played by these three very different but adjacent cells. The distinct distribution of FOX proteins, PS I & II proteins, and AcaK43 in heterocysts, vegetative cells, and akinetes, respectively, is consistent with previous findings, supporting the reliability of our proteomic data. Only the RuBisCO results are inconsistent with the previous observations. We observed high abundance of RuBisCO large and small subunits, and some carboxysomal microcompartment proteins (CcmN, CcmM, ORFs: 2671-2672) in all three cell types (Cameron, Wilson, Bernstein, & Kerfeld, 2013)). However, Cossar et al. reported that RuBisCO protein was undetectable in mature heterocysts of *A. cylindrica* (Cossar et al., 1985). Several lines of evidence from *Anabaena* strain PCC 7120 have shown that promoter activity of RuBisCO was barely detected in heterocysts using P_{rbcLS}-luxAB as a reporter (J. Elhai & Wolk, 1990), and RuBisCO large and small subunit transcripts were not detected in heterocysts by *in situ* hybridization (Madan & Nierzwicki-Bauer, 1993). Whether RuBisCO plays a role in both heterocysts and akinetes of *A. cylindrica* remains to be further investigated.

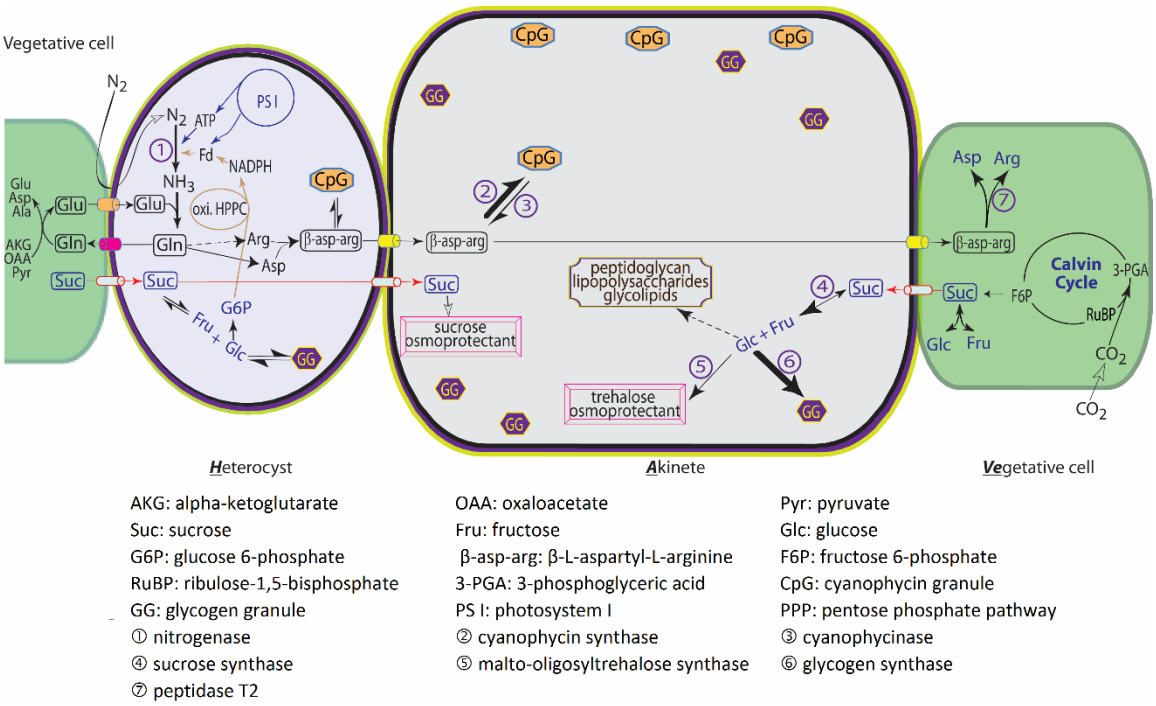

Figure 2-4: The HAVe (Heterocysts, Akinetes and Vegetative cells) model suggesting metabolic networks of cyanophycin and carbohydrates among heterocysts, akinetes, and vegetative cells.

The proteome of vegetative cells confirmed much of what is known about these workhorses. The large complement of PS I and PS II proteins supported active photosynthesis while RuBisCO, carboxysomal proteins, and other enzymes of the Calvin cycle supported carbon fixation. The glucose and fructose produced are likely synthesized into sucrose in vegetative cells and then supplied to the adjoining akinetes and heterocysts as primary energy and carbon source (Figure 2-1A). The HAVe model was supported by the findings of carbohydrate-related proteins in this proteomics study. In vegetative cells, fructose 6-phosphate is generated via the Calvin cycle during photosynthesis, which is then converted to sucrose by sucrose-phosphate synthase (SPS) and sucrose-phosphate phosphatase (SPP) (Cumino et al., 2002). Sucrose can be broken

down by invertase in vegetative cells, or transported to akinetes where sucrose is cleaved into glucose and fructose, serving as building blocks for other carbohydrate biosynthesis, e.g., peptidoglycan, lipopolysaccharide, and glycolipid as envelope materials; glycogen storage molecules; and/or trehalose osmoprotectant. The paucity of FOX proteins along with key components of nitrogenase such as NifD, NifN and NifU not detected supported absence of nitrogen fixation. The FOX protein, Histone-like DNA binding protein Hana was most abundant in vegetative cells, consistent with observations that a strong HanA-GFP fluorescent signal co-localized with DNA in vegetative cells (Lu, Shi, Chen, & Wang, 2014). A HanA mutant exhibited slow growth, altered pigmentation, and inability to differentiate heterocysts (I. Khudyakov & Wolk, 1996). FOX proteins unique to vegetative cells included a trace amount of HepN (Lechno-Yossef, Fan, Ehira, Sato, & Wolk, 2006), InvA, FraG, PrpI, NifU-like, DevR (Elsie L Campbell et al., 1996), and H6L region containing protein (ORF: 2881). Vegetative cells obtain fixed nitrogen from either heterocysts or adjoining akinetes in the form of β-aspartyl-arginine. β-aspartyl-arginine is further degraded into Asp and Arg by isoaspartyl dipeptidase (ORF: 4256) in vegetative cells (Table 2-1, (Burnat, Herrero, & Flores, 2014)). Asp and Arg, in turn, serve as precursors for the biosynthesis of other amino acids and nucleotides, the building blocks for DNA, RNA, and protein biosynthesis. Vegetative cells contained abundant enzymes for nucleotide and amino acid biosynthesis. Forty-nine out of 50 ribosomal proteins and translation factors were found in vegetative cells as well. These data suggest that DNA, RNA, and protein biosynthesis occurs actively in vegetative cells to maintain their cellular function and cell division.

The heterocyst proteome supported what is known about these specialized cells, but also indicated some novel functions. Heterocysts contained all the proteins required for nitrogen fixation, including several proteins absent in akinetes and vegetative cells (Appendix Table 1, S1). These included nitrogenase molybdenum-iron protein NifN (Y. Hu, Fay, Lee, Wiig, & Ribbe, 2010), the Fe-S cluster scaffold protein NifU that facilitates functional expression of nitrogenase in heterocysts (Nomata, Maeda, Isu, Inoue, & Hisabori, 2015), and DevA required for heterocyst maturation (Maldener, Fiedler, Ernst, Fernandez-Pinas, & Wolk, 1994). Nitrogenase iron protein NifH (Mevarech, Rice, & Haselkorn, 1980) had high abundance in heterocysts but was barely detected in vegetative cells and undetectable in akinetes (Appendix Table 1). Thus, the distribution of both Nif and Fox proteins indicated that N_2-fixation only occurred in heterocysts. Ammonia produced by nitrogenase is incorporated into glutamine (Gln), serving as ammonia donor to Asp and Arg. Asp and Arg are condensed by cyanophycin synthetase into cyanophycin in heterocysts (Burnat et al., 2014). The high levels of three cyanophycinases in heterocysts (Table 2-1) indicate that the bulk of fixed nitrogen is then available as β-aspartyl-arginine, a nitrogen vehicle to be transferred intercellularly to be either hydrolyzed into Asp and Arg in the vegetative cells, or condensed into storage cyanophycin granule by cyanophycin synthetase in adjoining akinetes. All but three photosystem proteins occurred in heterocysts. The abundance of several PS I and PS II proteins implied at least partial functioning of PS I and PS II (Appendix Table 1). Generation of oxygen (O_2) through PS II runs counter to the reductive process of nitrogenase. Nitrogenase is very sensitive to oxygen (O_2), so the heterocysts must create a micro-oxic environment. Cytochrome C oxidase subunit II is the last enzyme in the

respiratory electron transport chain. Valladares et al. showed that Cox2 and Cox3 transcription was up-regulated in heterocysts after nitrogen step-down in an NtcA- and HetR-dependent manner, and inactivation of both coxB2 and coxA3 results in the inability of *Anabaena* sp. PCC 7120 to grow diazotrophically under aerobic conditions (Valladares, Herrero, Pils, Schmetterer, & Flores, 2003). Consistent with their observation, CoxB3 was found in akinetes and heterocysts, and CoxB2 was only found in heterocysts (Appendix Table 1). Taken together, as cytochrome C oxidase has high affinity to oxygen, it may play a role of consuming residual oxygen in heterocysts, and keeping nitrogenase in its active state. The high abundance of RuBisCO in heterocysts may play a role in removing residual oxygen by oxidizing Ribulose 1,5-bisphosphate into 3-PGA and 2-phosphoglycolic acid (Eisenhut et al., 2008). Flavodiiron protein Flv3B (ORF: 1739, a homolog of all0178) was identified to be abundant in heterocysts, which may also be responsible for light-induced O_2 uptake in heterocysts to protect nitrogenase activity (Ermakova et al., 2014).

The key DNA replication proteins (DNA polymerases, SSBP) were not detected in heterocysts, consistent with terminal nature of the cells (Appendix Table 2). However, the presence of gyrase and helicase might play an important role in DNA rearrangement observed in heterocysts (Golden, Robinson, & Haselkorn, 1985). Like vegetative cells, heterocysts contained a broad spectrum of proteins involved with transcription and translation. Heterocysts are encased in a thick envelope to supply a micro-oxic environment for the protection of nitrogenase. Proteomic data indicated a number of heterocyst-specific proteins for synthesis, export, and external assembly of envelope polysaccharides and glycolipids (Appendix Table 3).

Heterocysts were also decorated with six of the seven S-layer secretion proteins in Gram-negative species. Secretion relies on specific ATP-biding cassette (ABC) transporters and outer membrane pores (Awram & Smit, 1998; Kawai, Akatsuka, Idei, Shibatani, & Omori, 1998). Heterocysts had several more ABC transporters (14) than did akinetes (7) or vegetative cells (10) (Appendix Table 3). The differential compositions of S-layer proteins and ABC transporters in the three cell types may contribute to the differences in cell envelope structure, including the greater resistance to cell disruption of heterocysts and also akinetes.

Akinetes appear to play a role as nitrogen and carbon storage cum transfer unit in filaments of *A. cylindrica* (Figure 2-1A). With this model fixed carbon enters into akinetes from vegetative cells and is converted to glycogen by glycogen synthase, or into trehalose for osmoprotection during the survival stage. Heterocysts flanked by akinetes on both sides would then obtain carbon for energy via akinetes. Similarly, akinetes receive β-aspartyl-arginine from heterocysts. This dipeptide is then either converted to cyanophycin for temporary or long-term storage or transferred to the adjoining vegetative cells to support the growing chain. Our proteomic data also indicated that akinetes have less active transcriptional and translational machinery. Importantly, proteomic data indicated a cell envelope that was different to those of vegetative cells or heterocysts. Akinetes were decorated with all seven S-layer proteins detected. The suite of peptidoglycan synthesizing machinery and cell wall hydrolases differed (Appendix Table 2), as did the complement of membrane transporters and enzymes involved in polysaccharide structures. It is worthy to note that ORF2780 protein, homologous to carbohydrate-selective porins (OprB), functions as a sugar porin responsible for the

optimal uptake of both fructose and glucose in Nostoc punctiforme ATCC 29133 (Ekman, Picossi, Campbell, Meeks, & Flores, 2013). The distinct distribution of OprB in akinetes and heterocysts at high abundance suggests a role in sugar uptake in these differentiated cells, consistent with the previous observation of carbon movement from vegetative cells to heterocysts of *A. cylindrica* (C. P. Wolk, 1968), which might also be true of carbon movement from vegetative cells to akinetes.

Akinetes have been viewed as spore like cells with the role of species survival under drought conditions. Their location between nitrogen-fixing heterocysts and carbon-fixing vegetative cells, combined with high levels of cyanophycin synthetase, cyanophycinase, sucrose synthetase, and glycogen synthetase suggests a critical role for akinetes during growth of *A. cylindrica* as demonstrated by the HAVe model. The role of the various genes and their regulation, as well as metabolite exchange among akinetes and their adjoining heterocysts and vegetative cells will need to be investigated in future work.

CHAPTER 3: Developmentally Regulated Genome Reduction in Nitrogen-Fixing Heterocysts of *Anabaena cylindrica* ATCC 29414

3.1 Abstract

Regardless of combined nitrogen availability, vegetative cells of *Anabaena cylindrica* differentiate semi-regularly spaced, single heterocysts along filaments. Since heterocysts are non-dividing cells, with the sole function for solar-powered N_2-fixation, is it necessary for heterocyst to retain entire genome (≈7.1Mb) from its progenitor vegetative cell? By sequencing the genome of isolated heterocyst, we discovered that at least six DNA elements (≈0.12 Mbp) are deleted from the heterocyst genome during heterocyst development. The six-element deletions restore five genes (*nifH1, nifD, hupL, primase P4*, *acyl_5725* (hypothetical protein) that were interrupted in the genome of vegetative cells. *A. cylindrica* has two *nifH* genes, *nifH*1 interrupted in vegetative cells and *nifH*2. Like the other four genes, the *nifH1* editing appears to be accomplished during the pro-heterocyst stage or before, while expression of two *nifH* is restricted to mature heterocysts. To our best knowledge, this is the first report that (1) different genomes may occur in distinct cell types in a single bacterium; and (2) genome editing is coupled to cellular differentiation and/or cellular function in a multicellular cyanobacterium.

3.2 Introduction

Nitrogen is one of most important, abundant life elements in bio-macromolecules such as DNA, RNA, and proteins. Although nearly 80% of air is the N_2 gas, most of the organisms are unable to use this form of nitrogen to make these essential bio-macromolecules. Fortunately, some cyanobacteria can photosynthetically fix atmospheric nitrogen gas (N_2) into a form that can be used by other organisms. Through billions of years of evolution, *Anabaena* species has gained the unique capability of using solar energy to reduce atmospheric N_2 to ammonia in specially differentiated N_2-fixing cells called heterocysts (Golden & Yoon, 2003; C. P. Wolk, 1996). The sole function for heterocysts is its solar-powered, oxic N_2-fixation. Thus heterocysts also offer scientists a rare opportunity to unlock the mystery of genome requirements for photosynthetic N_2-fixation. Unlocking genomic secrets of heterocyst would guide scientists to genetically engineering crops (leaves) to make self-fertilizing plants/crops using sunlight and atmospheric N_2 gas, just as heterocysts have done for billions of years.

Regardless of combined nitrogen availability, some vegetative cells of *Anabaena cylindrica* ATCC 29414 (hereafter *A. cylindrica*) can initiate a development program to form heterocysts that are present singly at semi-regular intervals along the filaments (Meeks et al., 1983). Heterocysts are morphologically and biochemically specialized for solar-powered, oxic N_2-fixation. By sequestering nitrogenase within heterocysts, *A. cylindrica* can carry out the two incompatible biochemical processes simultaneously, the O_2-producing photosynthesis and the O_2-labile fixation of N_2. Heterocyst-based N_2-fixation is a uniquely oxic, solar-powered process, which is distinct from anaerobic N_2-fixation by any other bacteria. This provides great potential for application in agriculture

compared to all other N_2-fixing bacteria whose inability to use solar energy, but also requiring anaerobic condition.

Unlike vegetative cells, heterocysts are terminally differentiated cells that have two extra O_2-impermeable layers of glycolipids and polysaccharide to exclude the O_2 who inactivates nitrogenase (Murry & Wolk, 1989; Zhou & Wolk, 2003). The heterocysts normally develop and mature within 24 hours, the period that from vegetative cells to morphologically distinguished mature heterocysts is called pro-heterocyst stage. A mature heterocyst is larger, more regular in shape with less granular cytoplasm than a vegetative cell, and it has thickened cell walls and a refractive polar granule at each end of the cell (Walsby, 2007). Cells with these characteristics, but lacking the thickened cell walls and the polar granules, are counted as pro-heterocysts (Adams & Carr, 1989).

Many genes have been identified to be involved in regulating heterocyst differentiation. HetR is a master transcription regulator specifically required for heterocyst differentiation (Buikema & Haselkorn, 1991; Huang et al., 2004b; Zhou et al., 1998). Several other regulatory genes such as *nrrA (Ehira & Ohmori, 2011), ccbP* (Y. Hu et al., 2011), *hetN (Higa et al., 2012), hetF*, *patA* (Risser & Callahan, 2008), *patN (Risser et al., 2012), patU (Meeks et al., 2002), hetZ (Zhang, 2007), patS* (H. X. Hu et al., 2015; Yoon & Golden, 1998), *hepK (Zhou & Wolk, 2003)* and *hetP* (P. Videau et al., 2016) were also found to play critical roles during heterocyst differentiation.

During heterocyst development in *Anabaena* sp. PCC7120, at least three DNA elements (11-kb, 55-kb and 9.4-kb inserted, respectively, within *nifD, fdxN* and *hupL*) are programmed to excise from the heterocyst genome by developmentally regulated site-specific recombination (Golden, Carrasco, Mulligan, Schneider, & Haselkorn, 1988;

Golden et al., 1985; Haselkorn, 1992). Both deletions of the 11-kb and 55-kb elements had been proven to be necessary for the heterocyst-based N_2-fixation, but not required for the differentiation of heterocysts in *Anabaena* sp. PCC7120 (Meeks et al., 2002), while the 9.4kb deletion has no effect on both N_2-fixation and heterocyst formation (C. D. Carrasco, Holliday, Hansel, Lindblad, & Golden, 2005).

The nitrogenase complex is encoded by a group of genes called *nif* genes. Many heterocyst-forming cyanobacteria have *nif* (*ni*trogen *f*ixation) genes (e.g., *nifD*, *nifK*, *nifH*, *fdxN*) interrupted by DNA elements that might be excised during heterocyst development (Hilton, Meeks, & Zehr, 2016). Since heterocysts are terminally differentiated, non-dividing cells, with the sole function for solar-powered N_2-fixation, is it necessary for heterocyst to retain the entire genome (≈7.0Mb) from its progenitor vegetative cell? To answer this question, we isolated heterocysts from *A. cylindrica* and sequenced the genomic DNA from heterocysts and vegetative cells. The genome sequencing of heterocysts and vegetative cells through the Illumina NextSeq technology produced 34,079,110 and 34,052,152 150 bp reads in pairs respectively for a ~799 x coverage of each genome. Assembly of both genomes yielded 1,819 and 566 contigs from heterocyst and vegetative cells respectively with N50 values of 44,251 and 35,464 bp respectively. After analysis of the heterocyst genomic sequence data and confirming it with quantitative PCR, we discovered that at least six DNA elements (≈0.12 Mbp) are deleted from the heterocyst genome during heterocyst development. The five-element deletions restore five genes (*nifH1*, *nifD*, *hupL*, *primase P4*, *acy_5725* (hypothetical protein) that were interrupted in vegetative cells. To our best knowledge, this is the first report that (1) different genomes may occur in distinct cell types in a single bacterium;

and (2) genome editing is coupled to cellular differentiation and/or cellular function in a multicellular cyanobacterium.

3.3 Material and Methods

3.3.1 Isolation and purification of heterocysts

Heterocysts were obtained from *A. cylindrica* ATCC 29414 grown in 5 L of AA/8 medium free of combined nitrogen (N. T. Hu et al., 1981), shaking at 150 rpm under illumination (50-60 $\mu E \cdot m^{-2} \cdot s^{-1}$ at the culture surface) for 7 days to an OD_{700} of 0.03. Cultures were harvested by centrifugation at 6,000 x g for 15 min and re-suspended in 80 mL dd H_2O. Vegetative cells were disrupted by passing suspensions through a Nano DeBEE 30 High-Pressure Homogenizer (BEE International) at 15,000 psi (lb/in^2) three times. The suspensions were centrifuged down at 4,000 x g for 10 min. To separate the debris of vegetative cells from heterocysts, pellets were re-suspended in 1mL ddH_2O, and the suspension was centrifuged at 1,100 x g for 5min. Two layers were formed: bottom green pellet and top lose yellow pellet. The top yellow pellet of vegetative cells debris was discarded and the bottom green heterocysts pellet was washed by re-suspending with 1mL ddH_2O and re-centrifuging at 1,100 x g for 5min. This wash step was repeated 4 times. After each washing step, the heterocyst fraction was checked microscopically in order to ensure that heterocysts were pure. The purified heterocysts were stored at -80°C.

3.3.2 Isolation of genomic DNA

Anabaena cylindrica ATCC 29414 was grown in 50mL AA/8 medium (free of combined nitrogen) and AA/8N (nitrate-containing medium) (N. T. Hu et al., 1981) for 7 days, and OD_{700} were 0.03 and 0.028, respectively. To extract the DNA from vegetative cells, the cultures were centrifuged at 13,000g for l5 min, 500 µL of 10% sucrose buffer (50mM

Tris-HCl, pH 8.0, 10mM EDTA) was to suspend the cell pellets, 50 µL of 125mg/mL lysozyme (Sigma), 150 µL 10% SDS and 10 µL RNase of 10 mg/mL were added. The ≈800 µL suspension was incubated at 37°C for 1 hr. Then total amount of the suspension was measured, equal amount of saturated phenol (pH6.6±0.2) was added, and the reagents were mixed together by vortexing. The suspension was centrifuged at 13,000g at room temperature for 10 min. The top aqueous solution was transferred to a new 1.5mL Eppendorf tube and equal amount of chloroform solution (chloroform: isoamyl aclcohol = 24:1) was added. Vortexed the tube and the suspension was centrifuged at 13,000g at room temperature for 10 min. The top aqueous solution was then transferred to new tube and equal volume of pre-cold isopropanol was added to precipitate total DNA. Some white pellet was obtained after centrifuging at 13,000g at 4°C for 10min. The supernatant was discarded, and the pellet was washed first with 70% ethanol and then 95% ethanol. The white pellet was air-dry for 5 min, and final 30 µL ddH2O was added to dissolve the total DNA.

To break the heterocyst and extract its genomic DNA, the purified 13.1mg (wet weight) of heterocysts stored at-80°C freezer were re-suspended in 500µL 10% sucrose buffer (50mM Tris-HCl, pH 8.0, 10mM EDTA). The suspension was centrifuged at 13,000 x g for 5 min. After removing the supernatant, another 500 µL 10% sucrose buffer was added to suspend the pellets, and 50 µL of 125mg/mL lysozyme was added. The total suspension was incubated at 37°C for 1.5hr. The suspension was sonicated at amplitude 70% for 10s with 0.5s pulse on and 0.5s pulse off. The sonication process was repeated 3 times. TissueLyser II (Qiagen) was further used to break heterocysts at frequency of 30/s for 8 min. The sample was frozen in liquid nitrogen immediately for 2 min and defrosted

at 80°C for 2 min. The TissueLyser, liquid nitrogen freezing, defreezing processes were repeated for another 3 times. Then 150 µL of 10% SDS, 150 µL of 0.5M EDTA (pH=8.0) were added to the suspension and incubated at 80°C for 30 min. 10 µL of RNase (10mg/mL) was added into the tube and incubated at room temperature for 15 min. The heterocyst DNA extraction procedures followed the same saturated-phenol method as described for vegetative cells. Final volume of 15 µL of ddH_2O was used to dissolve DNA. All the DNA was quantified by Qubit 3.0 (Thermo scientific).

3.3.3 RNA isolation

To extract RNA from vegetative cells, *A. cylindrica* was grown in 50mL AA/8 or AA/8N (N. T. Hu et al., 1981) for 7 days, and OD_{700} reaches ≈0.03. These cultures were collected for RNA isolation using RNeasy Mini Kit (Qiagen) with modification. 50 mL cultures were centrifuged down and 450 µL of lysis buffer RLT with 2 µL ß-Mercaptoethanol was added to suspend the culture pellets, then 50 µL 125mg/mL lysozyme was added to the suspension and incubated at 37°C for 30 min. The following RNA extraction was followed the instruction of the RNeasy Mini Kit (Qiagen). To eliminate genomic DNA contamination, an additional DNase treatment was performed according to the RNeasy kit instruction with the RNase-free DNase set (Qiagen). Finally, 30 µL of RNase-free ddH_2O was used to elute the RNA. The purified RNA was quantified using the NanoDrop 2000 spectrophotometer (Thermo Scientific). The RNA sample was stored at -80°C.

To extract RNA from heterocysts, ≈15mg of purified heterocysts (wet weight) stored at -80°C freezer were suspended in 450 µL of buffer RLT with 2 µL ß-mercaptoethanol, 50 µL of 125mg/mL lysozyme was added to the suspension and incubated at 37°C for 30

min, then 100 µL 10% SDS and 10 µL of 30 units DNase I (RNAase-free, Qiagen) were added and incubated at 37°C for 20 min and subsequently at 95°C for 5 min. The suspension was frozen in liquid nitrogen and defrosted at 60°C (before all the suspension was completely thawed). This freezing and thawing process was repeated twice. The following heterocyst RNA extraction procedures were the same as described for vegetative cells. Final volume of 20 µL RNase-free ddH_2O was used to elute heterocysts RNA. All the RNA was quantified by NanoDrop 2000 spectrophotometer (Thermo Scientific).

3.3.4 Genome sequencing

Sequencing libraries for the vegetative and heterocyst DNA samples were produced using a Nextera XT library preparation kit (Illumina) following the protocol described by the manufacturer. Libraries were quantified using a Qubit 3.0 and quality checked with an Agilent Bioanalyzer (Agilent). Equimolar amounts of both libraries were loaded as part of an Illumina NextSeq 500 high output run producing 2x150 bp paired ends reads.

3.3.5 Bioinformatics analysis

Read trimming, assembly and mapping were carried out using CLC Genomics Workbench 10.1.1 (Qiagen). Trimming was carried out using a Q of 20 as cutoff, eliminating any read with any ambiguous nucleotide and removing the 5 and 15 terminal nucleotides in the 3’ and 5’ ends respectively. Assembly of the trimmed reads was accomplished by setting an arbitrary minimum contig length of 3,000 bp and using the automated function to select a word size of 23 and a bubble size of 50; finally, reads were mapped (mismatch cost: 2, insertion cost: 3, deletion cost: 3, minimum length fraction: 0.5 and minimum similarity fraction: 0.9) to the assembly and the results used to correct

the contigs sequences. To detect possible deletions trimmed reads were mapped to the reference genome of *A. cylindrica* PCC 7122 using the large gap read mapper function (mismatch cost: 2, insertion cost: 3, deletion cost: 3, minimum length fraction: 0.9, minimum similarity fraction: 0.95 and randomly assigning those reads mapping in multiple locations).

3.3.6 PCR confirming the edited genes

Primers ZR1676 (0.5 µM) and ZR1677 (0.5 µM) were used to amplify intact *nifH1* using genomic DNA from heterocysts and vegetative cells of *A. cylindrica* ATCC 29414. The 891bp band was extracted using DNA extraction kit (Qiagen), and cloned into pCR2.1-TOPO vector (Invitrogen), and transformed it into TOP10 competent cells. The putative white colonies were verified by colony PCR (cPCR) with primers M13F and M13R. Then cPCR confirmed 3 colonies were grown in LB with Amp (100µg/ml) and Km (50µg/ml) overnight, and the plasmids were extracted and sent for DNA sequencing. The intact *nifD, hupL, primase P4* and *acy_5725*was PCR amplified by Phusion High-Fidelity DNA Polymerase (NEB) with specific primers listed in Table 3-1. The PCR products amplified with genomic DNA from heterocysts and vegetative cells of *A. cylindrica* were purified with Qiagen PCR clean kit for DNA sequencing.

Table 3-1 Primers used in the identification of deletions and gene expressions

Gene	Primer	Sequence
rnpA	ZR1693	TTGGATACTGCCCCTGCAACAAC
(*acy_1194*)	ZR1694	ATTCTGCTGACTTTGGCTTCAC
nifH1	ZR1676	ATGAGTACCGACGCAAATATTAG

	ZR1677	ACTACTTGGTAGCACCTGCGGGC
	ZR1679	GTGGTTTTGCTATGCCTATCCGT
	ZR1680	AAACGTTCAGCCAAGGTTTCGATC
	ZR1681	CGGATGACTGTTAACGAGTAC
nifH2	ZR1470	CCCTTGCTGCTATGGCAGAAATGGG
	ZR1471	CAACACCGGGTTCTGGACCACCAG
nifD	ZR1738	TGGTACAATCAACGAAATGTTGATG
	ZR1739	TACCTTCTCAGCATTTGCTTGG
	ZR1740	GTGAAAGAGAAGTATGTATTCC
	ZR1736	GTGTTTCTAGACGGAAGCTTTAGCC
	ZR1735	ATCCCCCGTTCCTTAGGAGCA
	ZR1737	GCAGTCATTACACCAGGAAGGG
Primase P4	ZR1749	GGTGAGAGAAGCTAAAATCCCTAGAATC
	ZR1748	GAAGGCGTTAGCACCTTCAAACTGG
	ZR1750	TGTCCATGGCCAGTAGCCAAATTG
	ZR1751	CATTCCACCGACGTTCAATTAATTC
	ZR1752	TTATCGCCACATGATCGCCTTCAAC
hupL	ZR1741	CGCTTTGCTGCCTTCACAGG
	ZR1742	CCTAACCAAACTGGTTCTAACC
	ZR1743	TTGAGGGACGTAACGCTGCTG
	ZR1734	GCACTAGACCAGGAATATGCACCTTC
	ZR1733	ATGCCAATTCAAACCTTAGATATTTC
Acy_5725	ZR1744	ATGATAGATTTTCTTAAGGGAAATGAGAG

ZR1745	GGCATTTGGTGGCCTGTAGC
ZR1746	CACTCGATTGGATGACTATAACTGG

3.3.7 Quantitative polymerase chain reaction (qPCR)

To determine the ratios of the edited genome vs unedited genome in heterocysts, vegetative cells grown in AA/8 or AA/8N, two pairs of primers for qPCR for each individual gene are listed in Table 3-1. Ten ng DNA isolated from vegetative cells grown in AA/8, AA/8N and 10 ng DNA isolated from heterocysts were added to 20 µL reaction with 0.2 unites of Phusion High-Fidelity DNA Polymerase (NEB), 1X Phusion buffer, dNTP (0.25 mM), and primers (0.5 µM). Each qPCR reaction had 5 replicates. The RT-PCR program was: 95°C for 10 min, 40 cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 30 s, and dissociation stage of 95°C for 15 s, 55°C for 30 s, 95°C for 15 s.

3.3.8 Reverse transcription qPCR (RT-qPCR)

200 ng of total RNA was reverse-transcribed into cDNA in a 20 µL reaction mixture with specific reverse primers listed in Table 3-1, using the Omniscript RT kit (Qiagen). The ones without reverse transcriptase were used as control. The reverse transcription reactions were incubated at 37°C for 1 hr. Then all the reactions, with and without reverse transcriptase were diluted 100 times for following RT-PCR. Five µL of 1:100 diluted cDNA samples above, Phusion (0.2 unit), Phusion buffer, dNTP (0.25 mM), primers (0.5 µM), 1 x SYBR and 1 x ROX fluorescent dyes (Invitrogen) in total 20 µL reaction. Each reaction was with 5 replicates. The RT-qPCR program was: 95°C for 10 min, 40 cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 30 s, and dissociation stage of 95°C for 15 s, 55°C for 30 s, 95°C for 15 s.

3.3.9 Determining the frequency of heterocysts

A. cylindrica can form heterocysts in both AA/8 (without combined nitrogen) and AA/8N (with combined nitrogen). The same cultures that used for isolation of genomic DNA **(**above**)** were used to determine the frequency of heterocysts using microscopy. Pictures were taken and the total numbers of heteocysts and vegetative cells were counted to determine the heterocysts frequency in both AA/8 and AA/8N growth media.

3.4 Results

3.4.1 Isolation and purification of heterocysts

Approximately 4.46% (Fig. S1 A) of *A. cylindrica* vegetative cells grown in AA/8 medium (free of combined nitrogen) can form heterocysts (Figure 2-1A), and the heterocysts were purified to a purity of 99.52 ± 0.48% (B). Unlike the other heterocyst-forming cyanobacteria, such as *Anabaena* sp. PCC7120, *Anabaena variablis* ATCC29413 and *Nostoc punctiforme* ATCC 29133, *A. cylindrica* can also form heterocysts with a frequency of ≈2.04% (Fig. S1B) when grown with combined nitrogen medium AA/8N.

Sequencing of heterocyst and vegetative cell genomes. The sequencing in the Illumina NextSeq platform of heterocyst and vegetative cells genome libraries produced 34,079,110 and 34,052,152 150 bp reads in pairs respectively for a ~799 x coverage of each genome. Assembly of both genomes yielded 1,819 and 566 contigs from heterocyst and vegetative cells respectively with N50 values of 44,251 and 35,464 bp respectively. The cumulative lengths of both assemblies were 27,543,022 and 8,228,290 bp compared to the 7,062,258 bp of the reference genome of *A. cylindrica* PCC 7122 and associated known plasmids. Mapping reads to the reference genome of *A. cylindrica* PCC 7122

detected three major deletions present in a significant portion of the template molecules. The three major deletions are Δ5,736bp in *nifH*1, Δ15,190bp in *nifD* and Δ20,842 bp in *hupL*. A fourth deletion was found in one of the plasmids (NC_019773) both in the heterocyst and the vegetative genomes.

3.4.2 Five interrupted genes in vegetative cell genome are precisely edited in heterocysts to be intact and functional

We extracted and sequenced both genomic DNA form heterocysts and vegetative cells. By mapping both the short reads and the contigs to the reference genome of *A. cylindrica* PCC 7122, along with specific PCR confirmation, at least six large DNA elements (5736bp *nifH1*-element, 74986bp 5'-*nifD* element, 15190bp 3'-*nifD*-element, 59225bp primase P4-element, 20842bp *hupL*-element, 39998bp hypothetical protein (*acy_5725*)-element present in vegetative cells) are precisely deleted from heterocyst genome (Figure 2-2), thus, the DNA element interrupted five genes (*nifH1*, *nifD*, *primase P4*, *hupL*, *acy_5725*) are restored to be intact and functional in heterocysts.

3.4.3 *nifH1* editing in heterocysts or vegetative cells

Two PCR products (6628bp and 891bp) were amplified with genomic DNA from vegetative cells grown in AA/8 (lane 2 in Figure 3-1B) and AA/8N (lane 4 in Figure 3-1B) with primers ZR1676/ZR1677, while only one product of 891bp was amplified in heterocyst genomic DNA (Figure 3-1B). The three 891bp-products were cloned into pPCR2.1-TOPO vector (Invitrogen) for DNA sequencing. These three 891bp PCR product sequence were confirmed to be identical to the coding region of *nfH1*in the contig 48. In other words, both the heterocyst genomic sequence contig 48 (GenBanK acess #: 2063644) and intact *nifH1* PCR product sequence confirmed that a 5736bp

element with sequence CCGTGAAG at both ends inserted within *nifH1* in vegetative cells was precisely cut out to restore an intact *nifH1* during heterocyst development. Interestingly, the intact *nifH1* was also amplified from the genomic DNA isolated from vegetative cells grown in AA/8 and AA/8N media (Figure 3-1B, lanes 2 &4). Further quantitative PCR (qPCR) with specific primers (Table 3-1) targeting on the edited *nifH1* and total *nifH*1 (both unedited and edited) was performed with genomic DNA isolated from different types of cells. The qPCR data (Figure 3-1C) showed that nearly 100% of interrupted *nifH*1 was edited to be intact in genome of heterocysts, while only 20.10% (AA/8) and 4.87% (AA/8N) of interrupted *nifH*1 was edited to be intact in genome of in vegetative cells that are grown in the medium without combined nitrogen (AA/8) and with combined nitrogen (AA/8N).

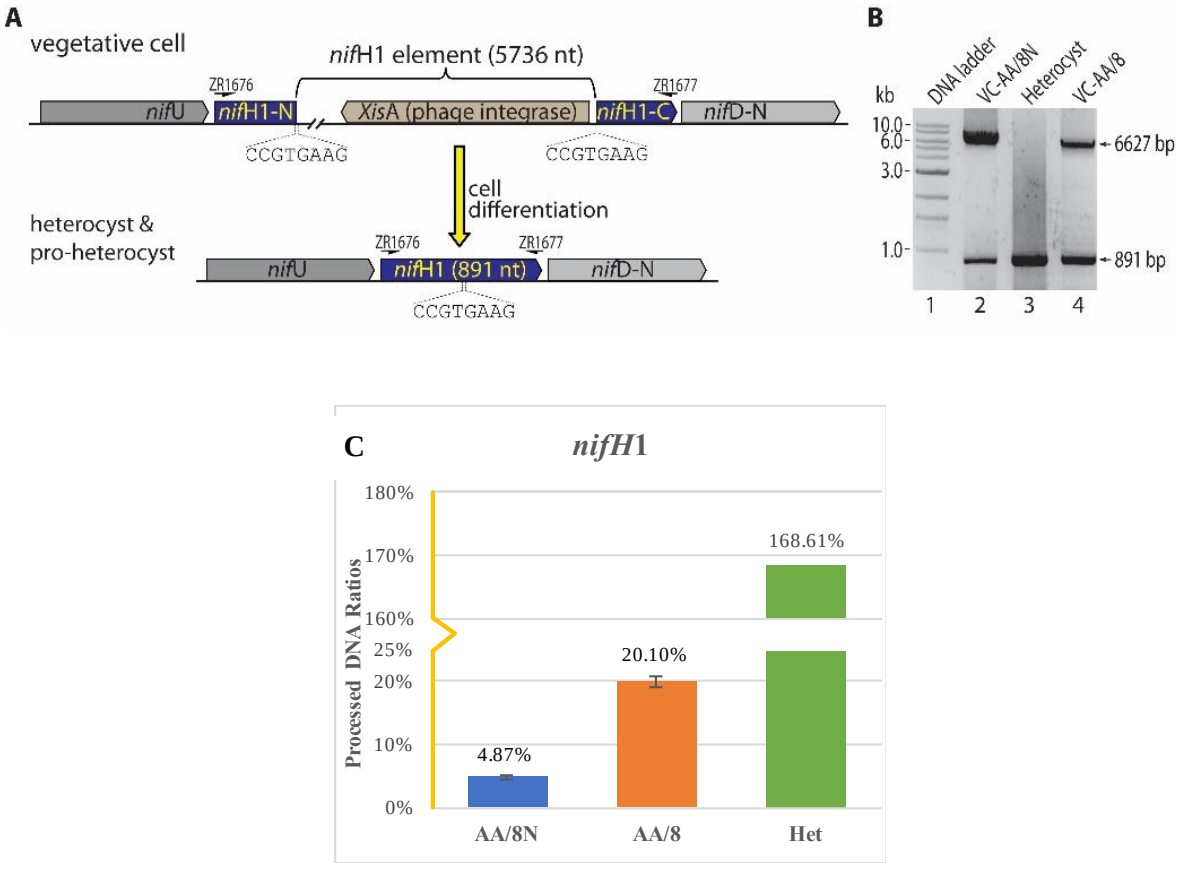

Figure 3-1 *nifH1* editing in heterocyst and vegetative cells. A) DNA edits ratios figures (*nifH1* as one example of five DNA elements deletion). B) Intact *nifH1* were confirmed in heterocysts and partial population of vegetative cells. C) The ratios of edited *nifH1*. The amounts of DNA were measured by RT PCR from DNA harvested from vegetative cells grown without (-N) or with (+N), and heterocysts. The amounts of total *nifH1* DNA and edited *nifH1* DNA were determined from the Ct of *nifH1* total and *nifH1* edited by using primers ZR1681/ZR1677, and ZR1679/ZR1680, respectively. The *nifH1* edited DNA ratios were calculated by 2^[Ct(*nifH1* total)- Ct(*nifH1* edited)]. The error bars indicate standard deviations.

3.4.4 Expression of *nifH1* and *nifH2*

A. cylindrica ATCC29414 has two *nifH* genes, *nifH*1 interrupted in vegetative cells and *nifH*2. There were only 20.10% (AA/8) or 4.87% (AA/8N) of the edited *nifH1* detected by qPCR in the genome of "vegetative cells". These small portions of "vegetative cells" may be responsible for pro-heterocysts or developing heterocysts. Thus, the *nifH1* editing is accomplished during the pro-heterocyst stage or before, while expression of two *nifH* detected by reverse transcription qPCR (RT-qPCR) is restricted to mature heterocysts (Figure 3-2).

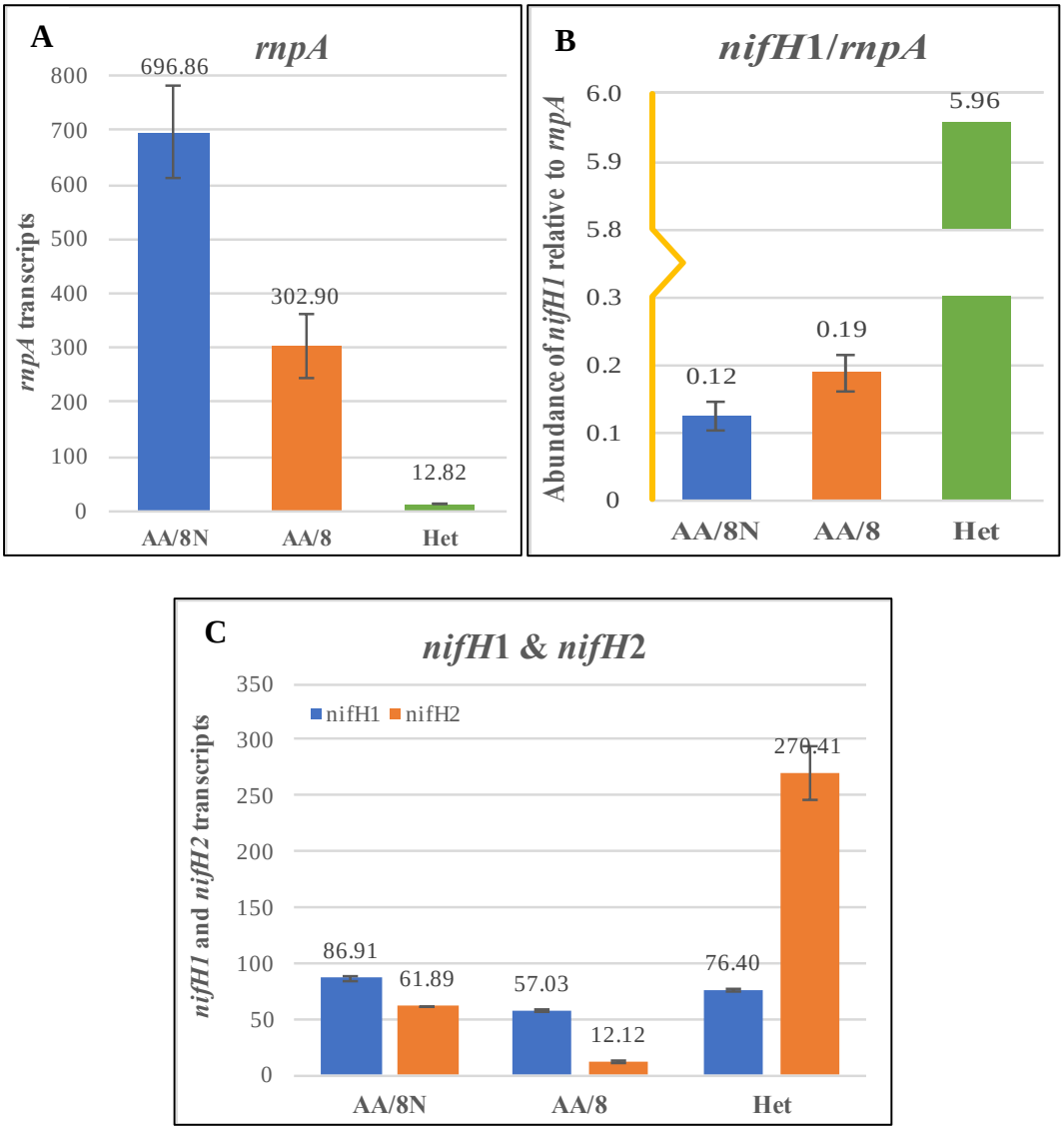

Figure 3-2 Expression of *rnpA (acy_1194)*, *nifH1* and *nifH2* in vegetative cells grown without (-N) or with (+N), and heterocysts. A). The amounts of *rnpA* transcripts were measured by RT-qPCR harvested from vegetative cells grown without (-N) or with (+N), and heterocysts, and calculated as $2^{[Ct(rnpA\ non\text{-}RT) - Ct(rnpA\ RT)]}$ with primers ZR1693/ ZR1694. B) The amounts of *nifH1* and *nifH2* transcripts were calculated as 2^[Ct(*nifH1* non-RT)- Ct(*nifH1* RT)] with primers ZR1679/ZR1680 and ZR1470/ZR147, respectively. C) The abundance of *nifH1* relative to *rnpA*. The relative abundance of

nifH1 to *rnpA* was determined by transcripts (*nifH1*)/transcripts (*rnpA*). The error bars indicate standard deviations.

3.4.5 *nifD* editing in heterocysts or vegetative cells

A 1543bp fragment was amplified with genomic DNA from heterocysts as well as from vegetative cells grown in AA/8 (lane 3 in Figure 3-3) or AA/8N (lane 2 in Figure 3-3) with primers ZR1735, ZR1736 (Figure 3-3). The three 1543bp fragments were purified by gel extraction and sent for DNA sequencing. These three PCR product sequences were confirmed to be identical to the coding region of *nifD* (1503bp). In other words, two DNA elements (74986bp 5'-*nifD* element, 15190bp 3'-*nifD*-element) inserted within *nifD* was precisely removed to restore an intact *nifD* during heterocyst development (Figure 3-3). Further quantitative PCR (qPCR) with specific primers (Table 3-1) targeting on the edited 5'-*nifD, 3'-nifD* and total *nifD* (both unedited and edited) was performed with genomic DNA isolated from different types of cells. The qPCR data (Figure 3-3 B&C) showed that ≈ 6.49% of 5'-*nifD* was edited (a 74986bp 5'-*nifD* element was removed) in heterocysts, while the edited 5'-*nifD* in vegetative cells account for only 7.86×10^{-4} (AA/8) and 3.25×10^{-4} (AA/8N) of total *nifD*. For 3'-*nifD* editing, approximately 70.66% of 3'-*nifD* was edited (a 15190bp 3'-*nifD*-element was removed) in heterocysts, while in vegetative cells the edited 3'-*nifD* accounted for only 4.91×10^{-3} (AA/8) and 4.16×10^{-3} (AA/8N).

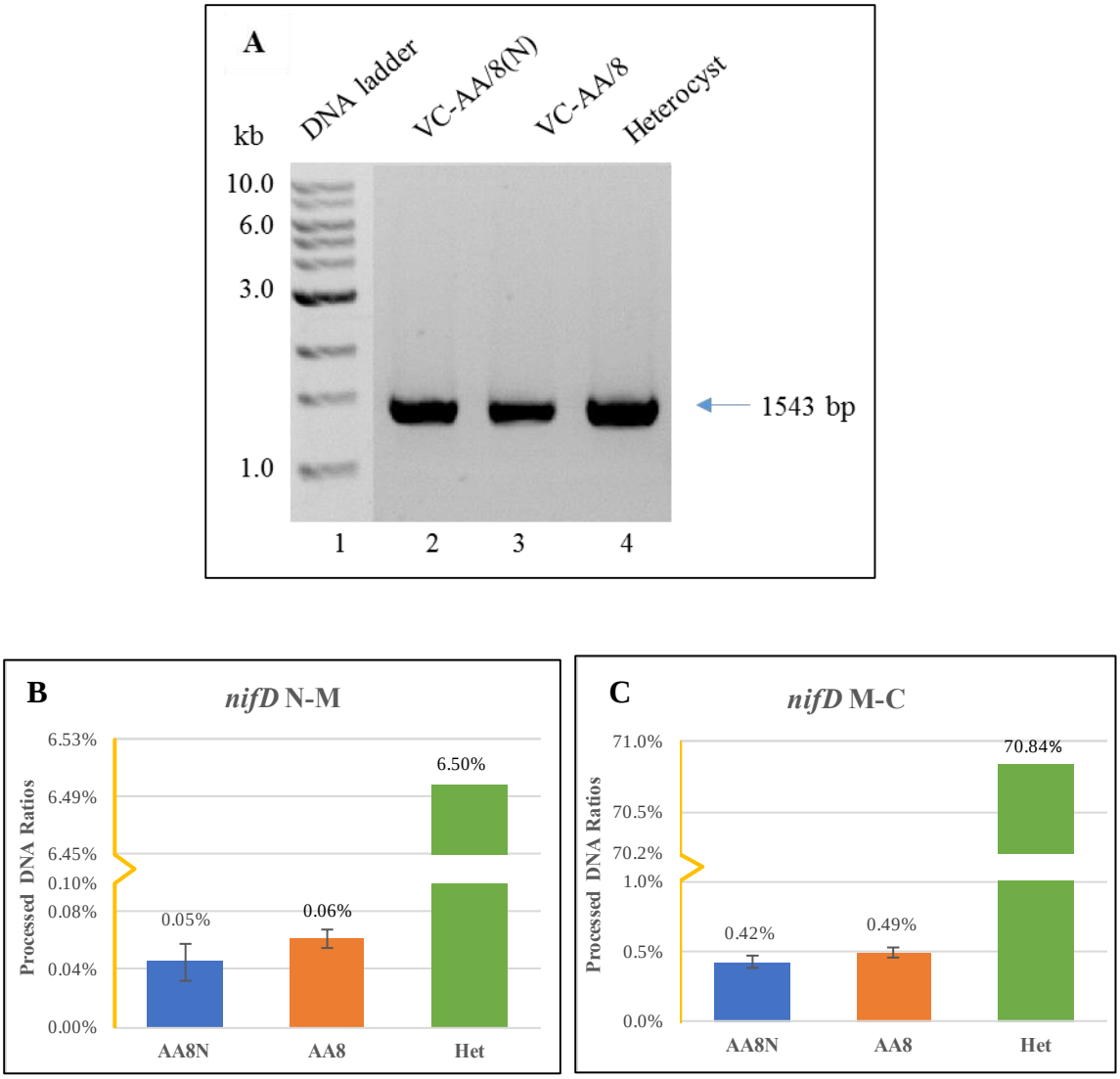

Figure 3-3 *nifD* editing in heterocyst and vegetative cells. **A)** cPCR confirmation of intact *nifD*. **B)** The ratios of edited *nifD* N-M (N-terminus-middle), and *nifD* M-C (middle –C-terminus) to total *nifD* region DNA. The amounts of DNA were measured by RT PCR from DNA harvested from vegetative cells grown without (-N) or with (+N), and heterocysts. **B)** The amounts of total *nifD* DNA and edited *nifD* N-M DNA were determined from the Ct of *nifD* total and *nifD* N-M edited by using primers ZR1735/ ZR1737, and ZR1738/ ZR1739, respectively. The *nifD* N-M edited DNA ratios were calculated by $2^{\wedge[Ct(nifD\ total)-\ Ct(nifD\ N\text{-}M\ edited)]}$. **C)** The amounts of total *nifD* DNA and edited

nifD M-C DNA were determined from the Ct of *nifD* total and *nifD* M-C edited by using primers ZR1735/ ZR1737, and ZR1740 / ZR1736, respectively. The *nifD* M-C edited DNA ratios were calculated by $2^{\wedge[Ct(nifD\ total)-\ Ct(nifD\ M\text{-}C\ edited)]}$.The error bars indicate standard deviations.

3.4.6 Expression of *nifD*

The expression of *nifD* was determined by RT qPCR across vegetative cells and heterocysts. In vegetative cells of AA//8, AA/8N, the transcipts of *nifD* were 48.32 and 43.10, while in heterocysts it was 5111.33 (Figure 3-4).

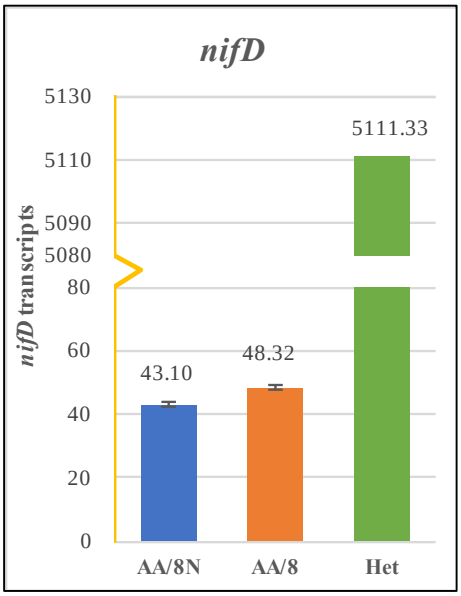

Figure 3-4 Expression of *nifD* in vegetative cells grown without (-N) or with (+N), and heterocysts. The amounts of *nifD* transcripts were measured by RT-qPCR harvested from vegetative cells grown without (-N) or with (+N), and heterocysts, and calculated as $2^{\wedge[Ct(nifD\ non\text{-}RT)-\ Ct(nifD\ RT)]}$ with primers ZR1738/ZR1739. The error bars indicate standard deviations.

3.4.7 *primase P4* editing in heterocysts or vegetative cells

An 1111 bp band was amplified with genomic DNA from vegetative cells grown in AA/8, AA/8N and heterocysts DNA with primers ZR1748, ZR1749 (Figure 3-5). The three 1111 bp bands were purified by gel extraction and sent for sequencing. The three 1111bp fragments were purified by gel extraction and sent for DNA sequencing. These three PCR product sequences were confirmed to be identical to the coding region of *primase P4.* In other words, the 59225bp primase P4-element inserted within *primase P4* was precisely removed to restore an intact primase P4 during heterocyst development (Figure 3-5). Further quantitative PCR (qPCR) with specific primers (Table 3-1) targeting on the edited and total *primase P4* (both unedited and edited) was performed with genomic DNA isolated from different types of cells. The qPCR data (Figure 3-5 B) showed that 67.13% of *primase P4* gene was edited in heterocysts, while the edited *primase P4* gene in vegetative cells accounted for only 2.79 x 10^{-4} (AA/8) and 2.45 x 10^{-4} x 10^{-4} (AA/8N) of total *primase P4* gene.

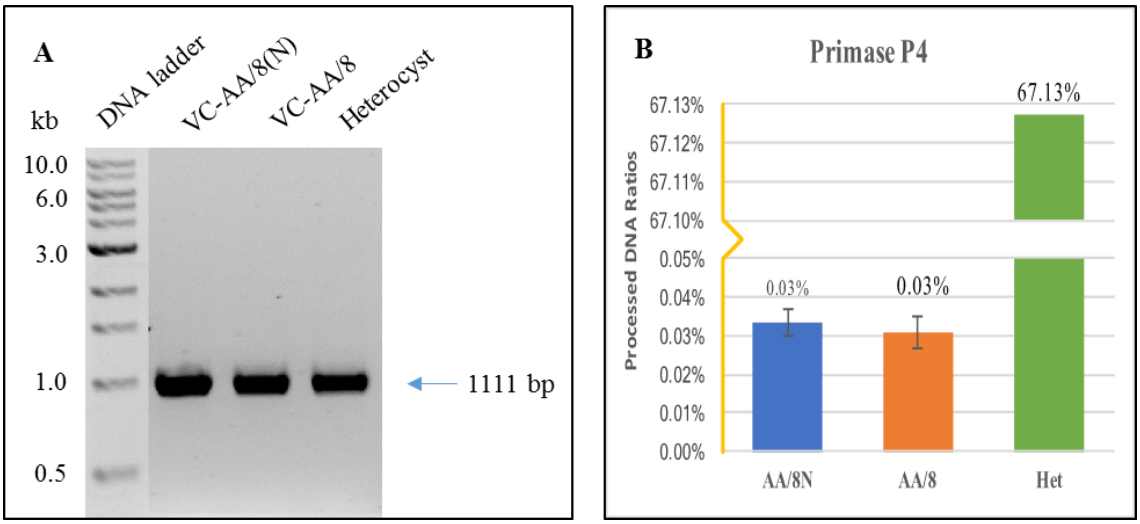

Figure 3-5 *Primase* P4 editing in heterocyst and vegetative cells. A) The PCR confirmation of intact primase P4. B) The ratios of edited primase P4. The amounts of

DNA were measured by RT PCR from DNA harvested from vegetative cells grown without (-N) or with (+N), and heterocysts. The amounts of total primase P4 DNA and edited primase P4 DNA were determined from the Ct of primase P4 total and primase P4 edited by using primers ZR1749 /ZR1750, and, ZR1751/ZR1752, respectively. The primase P4 edited DNA ratios were calculated by $2^{\wedge[Ct(Primse\ P4\ total)-\ Ct(primase\ P4\ edited)]}$. The error bars indicate standard deviations.

3.4.8 Expression of *primase P4*

The expression of primase was determined by RT qPCR across vegetative cells and heterocysts. The transcripts of primase P4 in vegetative cells grown in AA//8 or AA/8N was 3.09 and 4.15, while in heterocysts it was 9.47 (Figure 3-6).

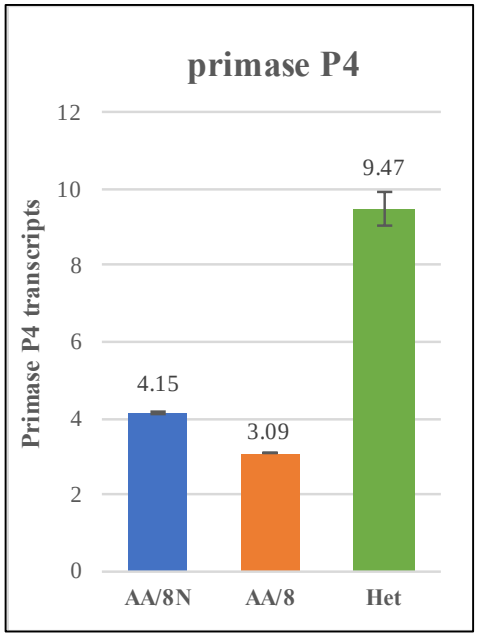

Figure 3-6 Expression of primase P4 in vegetative cells grown without (-N) or with (+N), and heterocysts. The amounts of primase P4 transcripts were measured by RT-qPCR harvested from vegetative cells grown without (-N) or with (+N), and heterocysts,

and calculated as $2^{\wedge[Ct(primase\ P4\ non\text{-}RT)-\ Ct(primase\ P4\ RT)]}$ with primers ZR1751/ZR1752. The error bars indicate standard deviations.

3.4.9 *hupL* editing in heterocysts or vegetative cells

A 1034bp fragment was amplified with genomic DNA from vegetative cells grown in AA/8, AA/8N and heterocysts DNA with primers ZR1743, ZR1734 (Figure 3-7 A). These three 1034bp PCR product sequence were confirmed that a 20,842bp element, inserted within *hupL*, was precisely removed to restore an intact *hupL* during heterocyst development. Further quantitative PCR (qPCR) data (Figure 3-7 B) showed that ≈51% of *hupL* was edited in heterocysts, while the edited *hupL* in vegetative cells accounted for only 4.61% (AA/8) and 2.38% (AA/8N) of total *hupL*.

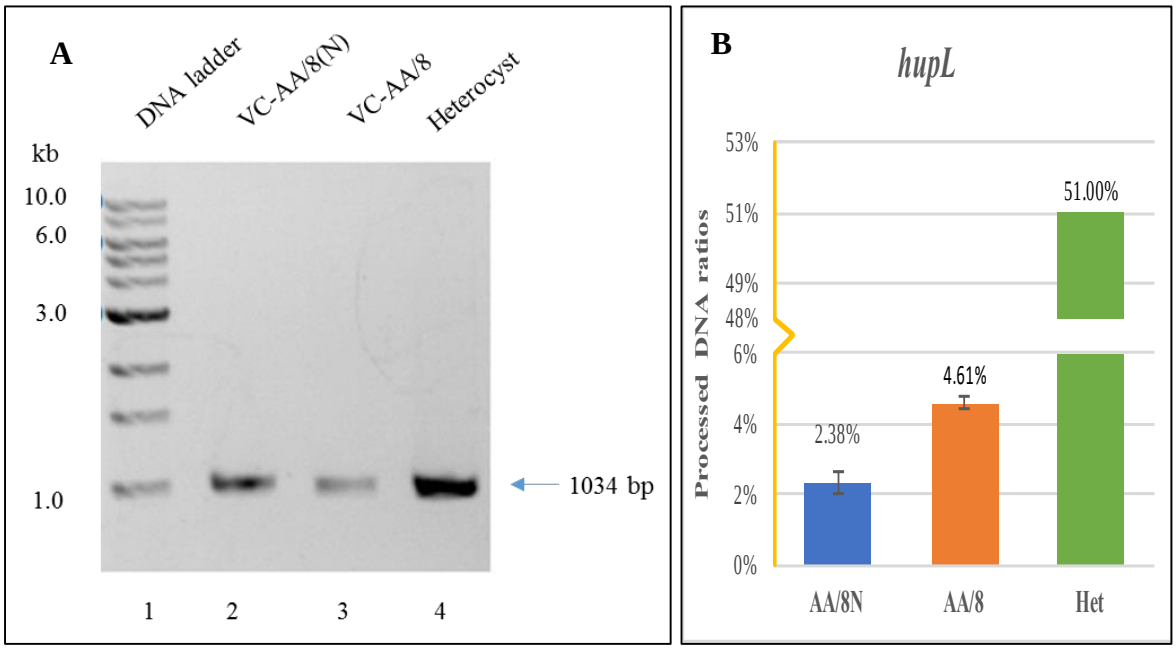

Figure 3-7 *hupL* editing in heterocyst and vegetative cells. A) The PCR confirmation of intact *hupL*. B) The ratios of edited *hupL*. The amounts of DNA were measured by RT PCR from DNA harvested from vegetative cells grown without (-N) or with (+N), and heterocysts. The amounts of total *hupL* DNA and edited *hupL* DNA were determined

from the Ct of *hupL* total and *hupL* edited by using primers ZR1743/ZR1734, and, ZR1741/ ZR1742, respectively. The *hupL* edited DNA ratios were calculated by $2^{\wedge[Ct(hupL\ total)-\ Ct(hupL\ edited)]}$. The error bars indicate standard deviations.

3.4.10 Expression of *hupL*

The expression of *hupL* was determined by RT qPCR across vegetative cells and heterocysts. The transcripts of *hupL* in vegetative cells grown in AA//8 or AA/8N was 77.96 and 157.90, while in heterocysts it was 2695.00.

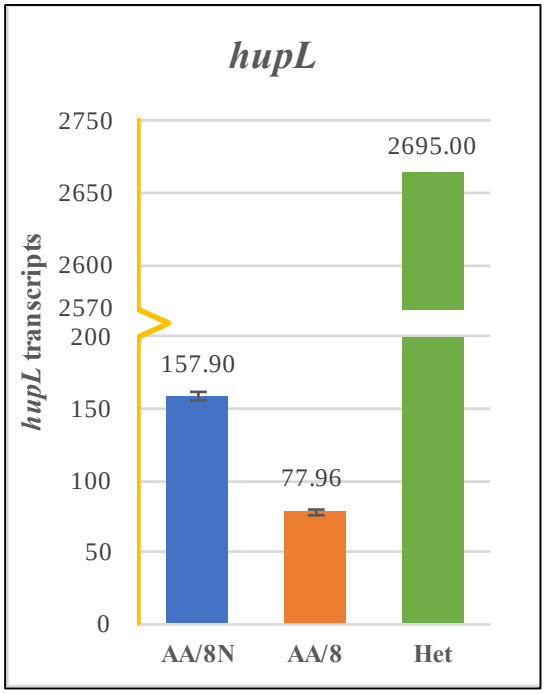

Figure 3-8 Expression of *hupL* in vegetative cells grown without (-N) or with (+N), and heterocysts. The amounts of *hupL* transcripts were measured by RT-qPCR harvested from vegetative cells grown without (-N) or with (+N), and heterocysts, and calculated as $2^{\wedge[Ct(hupL\ non\text{-}RT)-\ Ct(hupL\ RT)]}$ with primers ZR1741/ ZR1742. The error bars indicate standard deviations.

3.4.11 *Acy_5725* editing in heterocysts or vegetative cells

A 744bp fragment was amplified with genomic DNA from vegetative cells grown in AA/8, AA/8N and heterocysts DNA with primers ZR1744, ZR1745 (Figure 3-9). These three 744bp PCR product sequence were confirmed that a 39998bp *acy_5725*-element, inserted within *acy_5725* encoding a hypothetical protein, was precisely removed to restore an intact *acy_5725* during heterocyst development. Further quantitative PCR (qPCR) data showed that 90.76% of *acy_5725* was edited in heterocysts, while the edited *acy_5725* in vegetative cells accounted for 63.49% (AA/8) and 50.73% (AA/8N) of total *acy_5725*.

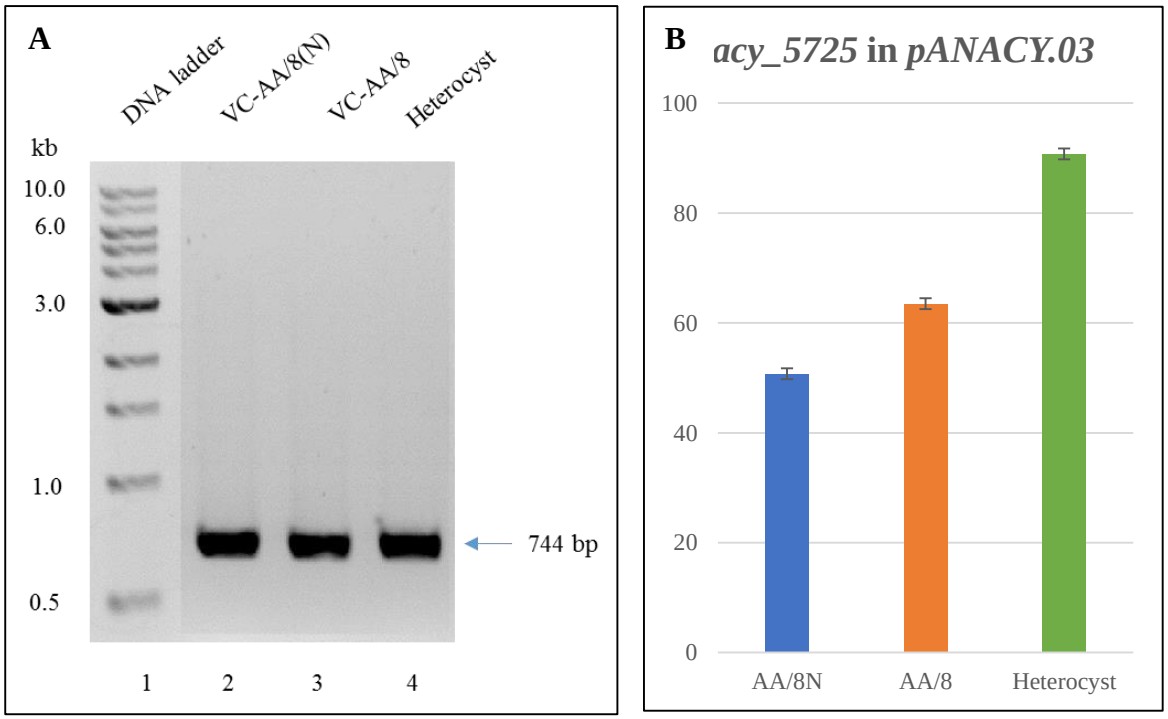

Figure 3-9 acy_5725 editing in heterocyst and vegetative cells A) cPCR confirmation of intact *acy_5725* by amplifying 744 bp using ZR1744/ZR1745. B) The ratios of edited *acy_5725* The amounts of DNA were measured by RT PCR from DNA harvested from vegetative cells grown without (-N) or with (+N), and heterocysts. The amounts of total *acy_5725* DNA and edited *acy_5725* DNA were determined from the Ct of *acy_5725*

total and *hupL* edited by using primers ZR1744/ZR1745, and, ZR1744/ ZR1746, respectively. The *acy_5725* edited DNA ratios were calculated by $2^{\wedge[Ct(acy\_5725\ total)-Ct(acy\_5725\ edited)]}$. The error bars indicate standard deviations.

3.4.12 Expression of *acy_5725*

The expression of *acy_5725* was determined by RT qPCR across vegetative cells and heterocysts. The transcripts of *hupL* in vegetative cells grown in AA/8 or AA/8N was 2.33 and 5.01, while in heterocysts it was 3.37.

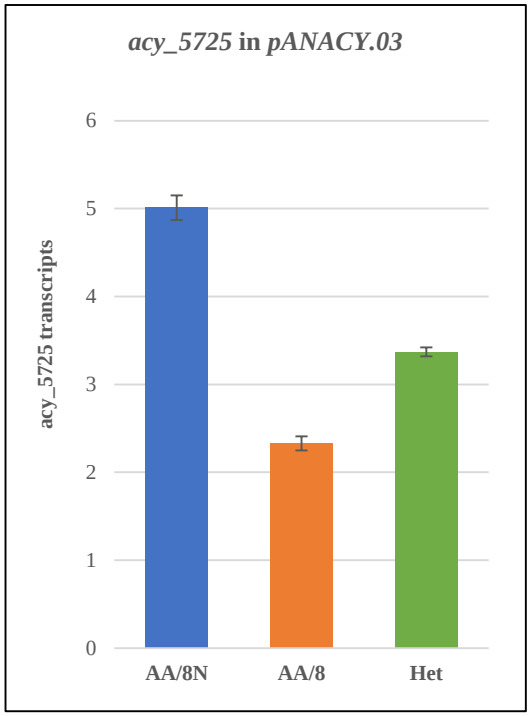

Figure 3-10 Expression of *acy_5725* in vegetative cells grown without (-N) or with (+N), and heterocysts. The amounts of *hupL* transcripts were measured by RT-qPCR harvested from vegetative cells grown without (-N) or with (+N), and heterocysts, and calculated as $2^{\wedge[Ct(acy\_5725\ non\text{-}RT)-\ Ct(acy\_5725\ RT)]}$ with primers ZR1744/ ZR1745. The error bars indicate standard deviations.

3.5 Discussion

3.5.1 *nifH1* editing is accomplished in an early stage of heterocyst development

Heterocyst is a specially differentiated, non-dividing cell with a unique function of solar-powered nitrogen fixation. During the heterocyst development, six large DNA elements (from 5,736bp *nifH1*-element to 74,986bp 5'-*nifD* element) that respectively interrupted five genes (*nifH1, nifD, hupL, primase P4,* and *ac5725*) in the genome of its progenitor vegetative cell were precisely deleted from the heterocyst genome of *A. cylindrica*, thus, the five interrupted genes are restored to be functional in heterocysts. Furthermore, our qPCR data demonstrated that small portion of genomic DNA isolated from vegetative cells also have these six DNA elements removed. For example, nearly 100% of interrupted *nifH*1 was edited in genome of heterocysts (Fig. 2C), while only 21.10% (AA/8) and 4.87% (AA/8N) of interrupted *nifH*1 was edited to be intact in genome of vegetative cells that are grown in the medium without combined nitrogen (AA/8) and with combined nitrogen (AA/8N). The ratios of the edited *nifH1vs*.total *nifH1* in vegetative cells had a good correlation with the heterocyst frequencies (Fig. S1). That is, the higher heterocyst frequency of 4.46% (AA/8) vs. 2.04% (AA/8N, Fig. S1), the higher *nifH1* editing of 21.10% (AA/8) vs.4.87% (AA/8N) in vegetative cells. Therefore, a small fraction of "vegetative cells" here may represent a stage of pro-heterocysts or even earlier stage of heterocyst development. We conclude that *nifH1* editing, like the other four genes, is accomplished in a stage of pro-heterocyst or even earlier stage of heterocyst development.

Among the five interrupted genes, three genes (*hupL, primase P4, and acy_5725*) have no known function for N_2 fixation. Although *nifH* is essential for N_2 fixation, *A.*

cylindrica has two *nifH* genes, *nifH*1 interrupted in vegetative cells and *nifH*2 that is an orphan *nif* gene. *nifH*2 is 1.8 Mbp away from *nif* cluster (*nifB-nifS-nifHD*). Both *nifH1* and *nifH2* are highly expressed in heterocysts, interestingly, the expression of *nifH2* in heterocysts was four times higher than the expression of *nifH1*.

For cyanobacteria, these developmental DNA rearrangements are exclusively seen in developing heterocysts, but not yet seen in vegetative cells. Similarly, in *B. subtilis* the DNA rearrangements are seen in every mother cell but never in the forespore. Both the heterocyst and the mother cell are terminal cells (non-dividing cells), providing no DNA for the next gereration.

3.5.2 Genes in deletion elements

In the five deletion elements, a total of 172 genes (Appendix Table 4) were speculated by comparing with the gene annotation in *A. cylindrica* PCC 7122. There are 87 hypothetical and 10 unknown genes, 14 integrase genes, 5 resolvase genes, 5 transposase genes, 7 ATPase genes, 15 tRNA genes, 14 DNA replication and transcription genes, 1 photosystem genes and 13 other functions genes.

Heteroycsts are terminal, non-dividing specialized nitrogen-fixing cells. The deletions of 15 tRNA genes, 14 DNA replication and transcription genes, 1 photosystem genes probably contribute to the lost or decrease of DNA replication, photosynthesis and transcription levels in heterocysts, which need to be further studied. Since vegetative cells are for photosynthesis and heterocysts are specialized for nitrogen fixation, the gene interruptions could be served as silencing these nitrogen fixation genes in vegetative cells, and re-initiate the nitrogen fixation by removing these elements in heterocysts. At

the same time, some functions can be deactivated in heterocysts by the losing of these elements.

3.5.3 Mechanism of DNA excision

Phage integrases are site-specific recombinases that mediate controlled and precise DNA integration and excision in these DNA elements deletions. Both integration and excision require integrase, the enzyme that mediates the site-specific DNA recombination. Excision also requires a phage-encoded accessory protein, an RDF or a Xis (Fogg, Colloms, Rosser, Stark, & Smith, 2014). In *Anabaena* PCC 7120 genome, three genetic elements: 11,278 kb *nifD*, 59,428 kb *fdxN* and 9,419 kb *hupL* interrupted these three genes (*nifD, FdxN* and *hupL*) (Kaneko et al., 2001). During development of heterocyst they are excised by site-specific recombination mediated by three different excisases (*XisA, XisF*, *XisC*, respectively). Each excisase are encoded within the specific DNA element. It has been reported that the excision of *nifD* element of *Anabaena* PCC 7120 does not occur in a *xisA* mutant strain (Trivedi et al., 2016).

Circularized phage genome can integrate via the phage attachment site attP into the bacterial host attachment site attB. The integration reaction produces the prophage flanked by the new attachment sites, attL and attR, which are hybrids containing half of attP and half of attB. Excision occurs between attL and attR to regenerate attP on the excised phage genome and attB on the host chromosome (Fogg et al., 2014). The direct repeat sequences in the DNA elements of *nifH1*, *nifD* N-M, *nifD* M-C, primase P4, *hupL* are CCGTGAAG, TATACCCTG, TACTCCG, AGTATATG and GCAGTTATATGG, these specific sequences may serve as the recognition sequence for precise excision of each specific DNA element.

3.5.4 Primase P4 restoration

nifD is composed of three fragments (N, M, C) that are interrupted with two DNA elements: N-M 75 kb and M-C 15 kb. Primase locates in N-M 75 kb element and is composed of two fragments that are interrupted with a 60 kb element. Our real-time PCR data suggested that there exists three populations varying among the different edits at the 75kb region: intact *nifD* N-M (75 kb DNA element was deleted), intact primase (60 kb DNA element that inside of 75 kb was deleted), and unedited with 75 kb DNA elements. These three population ratios in heterocysts are 6.5%, 67.1%, 70.7%, respectively. The *nifD* and primase transcription activities are much higher in heterocysts compared to vegetative cells which the genes editing ratios are low. The *nifD* and primase activity were restored to intact gene by removal of the DNA elements. Primase is responsible for building that RNA primer upon the parent strand at the base of the replication fork, besides, primase acts as a molecular brake to prevent the leading strand from outpacing the lagging strand by halting the progression of the replication fork in DNA replication (Jong-Bong et al., 2006). The restored primase P4 gene sequence and the increased transcript activities possibly contribute to slow down the DNA replication, which is consistent with that heterocyst is non-dividing cell with little or no requirement for DNA replication.

CHAPTER 4: Identification of surface polysaccharides in akinetes, heterocysts and vegetative cells of *Anabaena cylindrica* using fluorescein-labeled lectins

4.1 Abstract

In response to environmental changes, *Anabaena cylindrica* differentiate three cell types, vegetative cells for photosynthesis, heterocysts for nitrogen fixation, and akinetes for stress survival. Cell-surface polysaccharides play important roles in bacterial ecophysiology. In this study, specific cell-surface sugars were discovered in heterocysts, akinetes and vegetative cells of *A. cylindrica* using 20 fluorescein labeled lectins. Both N-acetylglucosamineor-binding lectins WGA and succinylated WGA bound specifically to the vegetative cells. Akinetes bound to three mannose-binding lectins (LCA, PSA, and ConA), and one of the galactose-binding lectins (GSL-I). ConA also bound to heterocyst, and the binding was diminished in the heterocysts of an all4388 mutant, in which the putative polysaccharide export protein gene all4388 was disrupted. Identification of distinct cell-surface sugar added our understanding to the role of polysaccharide for each cell type, and the identification may be applicable for fluorescence-activated precision cell sorting for comparative "omics" studies among the three cell types.

KEY WORDS: cyanobacteria, lectins, specific polysaccharides, spore, heterocyst, nitrogen fixation

4.2 Introduction

Cyanobacteria are ancient photosynthetic gram negative bacteria, some of which can also fix nitrogen (Fay 1992; Bergman et al. 1997; Berman-Frank et al. 2003). A matrix of polymeric substances surrounds cyanobacterial cell to form a protective boundary between the bacterial cell and the immediate environment (De Philippis et al. 2001). Extracellular

polymeric substances (EPS) are attached to the cell surface as capsular polysaccharide (CPS) or delivered to the culture medium as released polysaccharides. CPS can appear as a sheath, usually a thin, defined layer loosely covering cells, or a thick layer tightly associated with a single cell, or slime, which surrounds the cells. In addition, the cells are covered by lipopolysaccharides (LPS) anchored in the outer membrane (Kehr and Dittmann 2015) and underneath the CPS. This network of polysaccharides is critical for fundamental function, including cell gliding (Wilde and Mullineaux 2015), uptake of heavy-metals (Volesky and Holan 1995), biosurfactants (Paniagua-Michel Jde et al. 2014), recognition between symbiosis partners (Schussler et al. 1997; Marczak et al. 2017), providing protection against phagocytic predation, desiccation, and lysis by bacteria, viruses (Cheng and Costerton 1975; Dudman 1977) and fungi (Gerphagnon et al. 2013). Cellulose, which is a component of the extracellular matrix of several cyanobacteria of Sections I, III and IV, and consists only of glucose, is among the best-characterized polysaccharides in cyanobacteria (Nobles et al. 2001). The high diversity of monosaccharide building blocks defines the unique properties of cyanobacterial EPS, and clearly sets them apart from other bacteria (Pereira et al. 2009).

Anabaena cylindrica ATCC 29414 (hereafter *A. cylindrica*) is a filamentous cyanobacterium presenting an example of multicellular differentiation process in prokaryotes. Approximately 4-6% of the progenitor photosynthetic vegetative cells can develop into nitrogen-fixing heterocysts (Meeks et al. 1983). Heterocysts inhibit nearby cells from differentiating into heterocysts but can induce adjacent vegetative cells to become akinetes, an embryogenetic-type induction in *A. cylindrica* (Wolk 1966; Hirosawa and Wolk 1979; Zhou and Wolk 2002). The structure of cell wall changes dramatically

when vegetative cells develop into akinetes and heterocysts, which are both specialized cells with thickened cell walls. Morphological changes include the deposition of two additional envelope layers around the heterocyst; an inner “laminated” layer composed of heterocyst glycolipid layer (HGL), and an outer polysaccharide (called heterocyst envelope polysaccharide or HEP) layer (Cardemil and Wolk 1979; Cardemil and Wolk 1981; Zhou and Wolk 2003; Nicolaisen et al. 2009). A cluster genes were found to be required for the synthesis of HEP in *Anabaena* sp. strain PCC 7120 (Huang et al. 2005). The HEP layer is sometimes subdivided into a well-defined homogeneous inner layer and an external fibrous layer (Zhou and Wolk 2003; Kumar et al. 2010). The CPS of mature akinetes is crucial for the stress tolerance during desiccation, freezing and thawing (Tamaru et al. 2005; Perez et al. 2016).

Our previous proteomic study of vegetative cells, heterocysts and akinetes in *A. cylindrica* identified that a polysaccharide export protein coded by *acy1651* had much higher abundance in heterocysts and akinetes than in vegetative cells (Table 3), which suggested that this polysaccharide export protein (Acy1651) may play a role in polysaccharide deposition on different cell types.

Lectins are carbohydrate-binding proteins that can selectively recognize specific carbohydrate structures or sugar moieties (Brooks 2017). In this study, we identified the presence of specific sugar moieties in the surface of vegetative cells, heterocysts and akinetes of *A. cylindrica* by evaluating lectins' specific binding to each cell type. To study the role of Acy1651 (a polysaccharide export protein) in polysaccharide transportation and deposition, we have to examine the heterocyst-specific surface sugar monomers in heterocysts of FQ1595 (Lechno-Yossef et al. 2011), a Δ*all4388-Anabaena* sp. PCC 7120

mutant strain (*all4388* is the homologous gene of *acy1651),* due to lacking genetic tool for *A. cylindrica* to make a Δ*acyl1651*mutant.

4.3 Materials and methods

4.3.1 Lectin bindings to three types of cells of *A. cylindrica*

Culture preparation: *Anabaena cylindrica* ATCC 29414 (hereafter *A. cylindrica)* was grown in 50 mL AA/8 medium free of combined nitrogen (Hu et al. 1981) under continuous light (60 μE/m^2/s, 150 rpm, 30°C) for 10 d, yielding an OD_{700} of 0.149. Microscopy confirmed the formation of both heterocysts and akinetes. The 10-day's *A. cylindrica* culture (1.3 mL) was collected by centrifugation (13,000 x g) and was washed with Phosphate Buffered Saline (PBS) (10 mM, pH7.4) three times and re-suspended in 1 mL of HEPES buffer (10 mM, pH 7.5) ready for lectin-binding assay below.

Lectin binding experiments*:* Three screening kits of fluorescein-conjugated lectins (catalog No. FLK-2100, FLK-3100, FLK-4100) were obtained from Vector Laboratories (Table 1). Because the uncertainty of surface sugar quantity across the three cell types, the initial testing concentration of lectin was 100 μg/mL, and then increased the concentration based on the testing results. The details of finally used concentration of for each lectin were specified in Table 1. Lectins were added to each cell suspension, incubated for 60 min in the dark at room temperature, briefly vortexing the suspension every 15 min. The cells were harvested by centrifugation for 1 min at 13,000 x g. The cell pellets were washed three times with PBS to remove the unbound lectins. All processing was performed in the dark. The final cell pellet was re-suspended in 20 μL of PBS and 5μL of cells applied onto a slide for fluorescence microscopy.

A. cylindrica cells were examined with a BX53 fluorescence microscope (Olympus) under bright field for visualizing the three different cell types along the filaments. Fluorescence from the fluorescein-conjugated lectins was recorded using a GFP-4050A filter (EX466/40, EM525/50, DM495, BrightLine®) with 250 msec exposure time, and autofluorescence was recorded using a TRITC-B filter (EX543/22, EM593/40, DM562, BrightLine®) with 250 msec exposure time.

Table 4-1: Lectins used to visualize extracellular polysaccharides, and final concentrations of each lectin applied

Lectin	Common Abbreviation	Primary sugar specificity	Primary sugar specificity	concentrations (μg/mL)
Concanavalin A	ConA	Mannose	αMan, αGlc	120
Len culinaris lectin	LCA	Mannose	αMan, αGlc	120
Pisum sativum agglutinin	PSA	Mannose	αMan, αGlc	120
Dolichos biflorus agglutinin	DBA	N-Acetylgalatosamine	αGalNac	120
Soybean agglutinin	SBA	N-Acetylgalatosamine	α>βGalNAc	100
Sophora japonica agglutinin	SJA	N-Acetylgalatosamine	βGalNAc	100
Vicia villosa lectin	VVL	N-Acetylgalatosamine	GalNAc	100
Ricinus communis agglutinin I	RCA I	Galactose,	Gal	120

		N-Acetylgalatosamine		
Peanut Agglutinin	PNA	Galactose	Fucα6GlcNAc	120
Griffonia Simplicifolia lectin I	GSL I	Galactose	αGal, αGalNAc	100
Erythrina cristagalli lectin	ECL	Galactose	Galβ4GlcNAc	120
Jacalin	Jacalin	Galactose	Galβ3GalNAc	120
Ulex Europaeus agglutinin I	UEA I	Fucose	αFuc	100
Phaseolus vulgaris Erythroagglutinin	PHA-E	Complex structures	GalβGlcNAcβ2Manα6(GlcNAcβ4)(GLcNAcβ4Manα3)Manβ4	100
Phaseolus vulgaris Leucoagglutinin	PHA-L	Complex structures	Galβ4GlcNAcβ6(GlcNAcβ2Manα3)Manα3	100
Wheat Germ agglutinin	WGA	N-Acetylglucosamine	GlcNAc	160
Wheat Germ agglutinin, succinylated	Succinylated WGA	N-Acetylglucosamine	GlcNAc	160
Griffonia simplicifolia lectin II	GSL II	N-Acetylglucosamine	α or βGlcNAc	120

Datura stramonium lectin	DSL	N-Acetylglucosamine	$(GlcNAc)_{2-4}$	100
Lycopersicon esculentum lectin	LEL	N-Acetylglucosamine	$(GlcNAc)_{2-4}$	100
Solanum tuberosum lectin	STL	N-Acetylglucosamine	$(GlcNAc)_{2-4}$	100

4.3.2 Examination of heterocyst surface sugar in all4388 mutant of *Anabaena* sp. PCC 7120 (hereafter A. PCC 7120)

A. cylindrica A. PCC 7120 is close to *A. cylindrica*, but does not form akinetes. The all4388 is homologous (68% identities) to acy1651 encoding a putative polysaccharide transporter in A. PCC 7120. The all4388-knockout mutant FQ1595 (hereafter Δall4388) was obtained from Dr. Wolk laboratory (Lechno-Yossef et al. 2011). The Δall4388 mutant and wild-type A. PCC 7120 were grown in 25 mL Bg11 with 50□g/mL of neomycin (sigma-aldrich) and Bg11 medium respectively under continuous light (60 µE/m2/s, 150 rpm, 30°C) for 7 d, reaching OD700 0.1. The cultures were harvested by centrifugation for 5 min at 4,000 x g, the cells were washed three times with Bg110 (without combined nitrogen), and re-suspended in 25 mL Bg110 to induce development of heterocysts. After 48h induction, 1.3 mL of culture was collected, washed with 1.0 mL PBS for three times, and suspended in 1 mL of HEPES buffer (10 mM, pH 7.5). The heterocyst signature lectin ConA was added at 120 µg/mL, and the binding activity was observed by fluorescence microscopy as described above.above.

4.4 Results

4.4.1 The specific polysaccharides identified in akinetes, heterocysts and vegetative cells

The *A. cylindrica* appeared green-blue filaments, with yellowish heterocysts (Figure 4-1 column I, Ht). Vegetative cells and akinetes (Ak) displayed strong red auto-fluorescence when viewed with the TRITC-B filter, but not the heterocysts (Figure 4-1 column IV). Heterocysts showed a diminished red auto-fluorescence (Figure 4-1 column IV) due to the absence of phycobilisomes, which involved in light absorbance. Among the 20

fluorescein-conjugated lectins tested (Table 4-1), seven lectins showed some binding to at least one of three cell types (Table 4-2 and Figure 4-1). ConA gave strong signals in heterocysts, akinetes and the terminus of terminal cells in filaments (Figure 4-1A & B, column II &III). PHA-E also bound to the terminus of the terminal cells (Figure 4-1F, column II & III). GSL-I, LCA, and PSA showed a binding pattern similar to ConA for the developing akinetes or akinetes (column II & III in Figure 4-1, A, C, D, E). Succinylated WGA and WGA exclusively bound to vegetative cells (Column II in Figure 4-1, G &H), but this binding was not as intense as the other lectins. The other 13 lectins (Table 4-1) did not show detectable binding to any type of cells of *A. cylindrica*, including the group of lectins that bind to N-acetylgalactosamine, suggesting that few if any traces of N-acetylgalatosamine exist on the cell surface of *A. cylindrica*.

Table 4-2: Surface polysaccharide identified in all cell types by lectins bindings

Lectins	Polysaccharide	Vegetative cells	Heterocysts	Akinetes	Terminal cells
Con A	Mannose	-	+	+	+
GSL I	Galactose	-	-	+	-
LCA	Mannose	-	-	+	-
PSA	Mannose	-	-	+	-
PHA E	Complex oligosaccharide	-	-	-	+
WGA	N-Acetylglucosamine	+	-	-	-

Succinylated WGA	N-Acetylglucosamine	+	-	-	-

4.4.1.1 Akinete specific surface polysaccharides

Akinetes was bound by the highest number of lectins: ConA, GSL-I, LCA, and PSA (Figure 4-1 A, and C-E, respectively), which implied that akinete cell surfaces had the greatest variety of sugars: both terminal mannose and galactose. GSL-I binding to the akinete surface showed "fuzzy" fluorescence (Figure 4-1, C) compared with the smooth binding surface by ConA (Figure 4-1, A), LCA (Figure 4-1, D) and PSA (Figure 4-1, E). The fact that GSL-I is a galactose-binding lectin, and the other three are mannose-binding lectin, suggesting the structural difference of galactose and mannose on the akinete surface. The mannose-containing polysaccharides are probably associated with the akinete surface tightly while the galactose-containing polysaccharides are more like “microfibers” stretching out of cell wall.

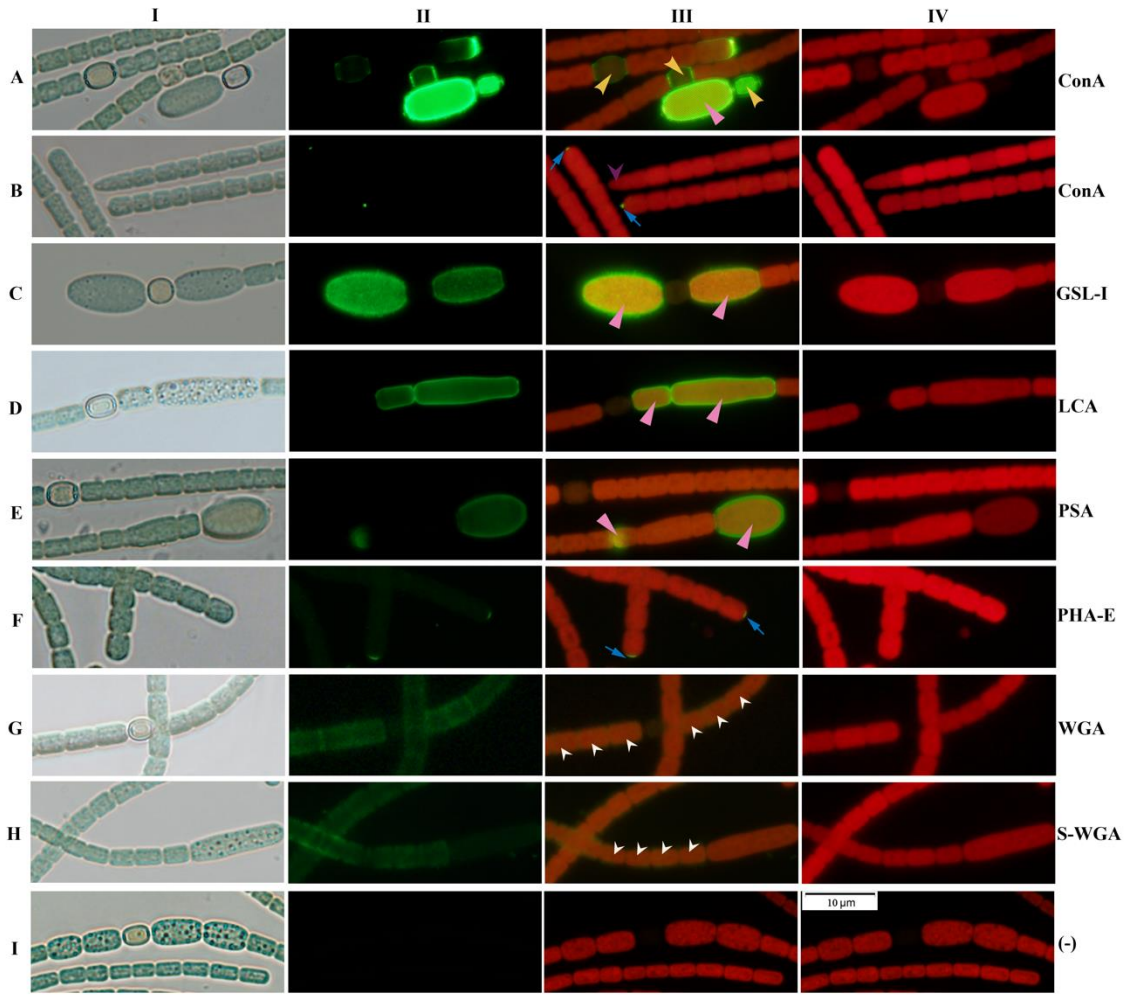

Figure 4-1: Seven fluorescein-conjugated lectins were identified to have specific binding affinity to different cell types of *A. cylindrica*. Photomicrograph of *A. cylindrica* stained with ConA (A & B), GSL I (C), LCA (D), PSA (E), PHA E (F), WGA (G), and succinylated WGA (H) observed under bright field (column I), FITC (column II), FITC-TRITC merge (column III) and TRITC (column IV). The different cell types were labeled as pink (akinetes), yellow (heterocysts), white (vegetative cells), blue (blunt end of terminal cells), and purple (tip end of terminal cells) arrows respectively.

Akinetes are enlarged spore-like cells that differentiated from vegetative cells. In *A. cylindrica*, akinete size is about 10 times larger than vegetative cells, which is different

from bacterial spores. Yet the akinete formation mechanism is barely known. In our lectin research, LCA and PSA showed the affinity to akinetes during akinete development. Figure 4-1D-II revealed the two akinetes both bound with LCA. Figure 4-1E-II shows the PSA bound to a young akinete and one partial of vegetative cells, which potentially developed into an akinete. These results suggested mannose's appearance is an earlier indicator of akinete differentiation.

4.4.1.2 Heterocyst specific surface polysaccharides

Heterocysts bound only to ConA (Fig. 4-1A-II) that prefers to bind to mannose, glucose. Heterocysts are morphologically and biochemically specialized N_2-fixing cells that form an additional two-layer of cell wall with an inner layer of glycolipids and an outer layer of specific polysaccharides to block the environmental O_2 (Zhou and Wolk 2003).

4.4.1.3 Vegetative cell-specific surface polysaccharides

Both WGA and succinylated WGA primarily (if not specifically) bound to filamentous vegetative cells (Figure 4-1 G-H, column II), which suggesting the existence of N-Acetylglucosamine on the surface of vegetative cells. Each filament has two ends: Tapering-end and blunt-end (Figure 4-1 B, column I). Both ConA (Figure 4-1 B, column II & III) and PHA-E (Figure 4-1 F, column II & III) specifically bound to the blunt-end vegetative cells, which suggesting that mannose/glucose and complex oligosaccharides deposited at only the blunt-end vegetative cells, not the tapering-end vegetative cells.

4.4.2 ConA binding activities in A. PCC 7120 and Δall4388 mutant

Our proteomics study detected the high abundance of polysaccharide export protein (Acy1651) in heterocysts and akinetes of *A. cylindrica* 29414 (Table 4-3). To study the role of Acy1651, *Anabaena* PCC 7120 and its Δ*all4388* mutant were examined for ConA-

binding activity. *A.* PCC 7120 is genetically close to *A. cylindrica* 29414 except for no formation of akinetes.

ConA also bound to the heterocysts of *A.* PCC 7120, which suggesting the presence of mannose/glucose-containing polysaccharide on the surface of heterocysts (Figure 4-2A, column II). The similar binding pattern indicated similar polysaccharide composition of EPS in *A.* PCC 7120 and *A. cylindrica* 29414. ConA fluorescence was observed much stronger on the mature heterocyst (yellow arrowed in Figure 4-2A) than the younger heterocyst (yellow arrowed in Figure 4-2B). Different ConA-binding activities were also observed in different stage of heterocyst development (Figure 4-1A, column I & II) even if the morphology of heterocyst was not easily distinguishable (data not shown). The differential ConA-fluorescence from the developing heterocysts suggested the accumulation of mannose/glucose-containing polysaccharides during the development of heterocysts.

Functional heterocysts formed about 24 h after nitrogen step-down in wild type *A.* PCC 7120, while the Δ*all4388* mutant still formed heterocysts, but they were not as distinguishable (yellow arrowed in Figure 4-2C, column I & II) as in wild type, and they cannot fix nitrogen (Ernst et al. 1992; Maldener et al. 2003; Lechno-Yossef et al. 2011). In Δ*all4388*, ConA no longer binds to heterocysts (Figure 4-2C), which suggesting that mannose/glucose-containing polysaccharides are missing from the heterocyst surface due to the absence of this polysaccharide export protein, All4388. It was noticeable in Δ*all4388* mutants that some cell junctions were enriched in mannose/glucose-containing polysaccharides bound by ConA (Figure 4-2 C-II).

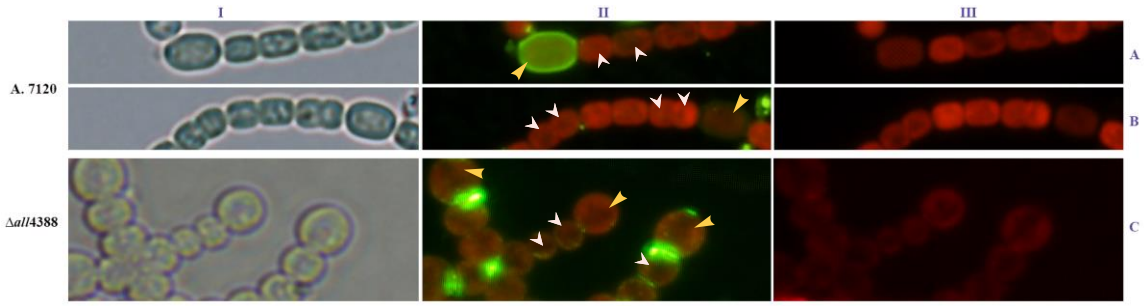

Figure 4-2 Characterization of the surface sugar in *A.* PCC 7120 and Δ*all4388* mutant by ConA staining. Heterocysts in *A.* PCC 7120 was specifically binding by ConA (A-B), and this recognition was diminished in Δ*all4388* mutant heterocysts and strong ConA signal was detected between the junction of cells (C).

Our heterocyst signature ConA was observed attaching to the heterocysts in *A.* PCC 7120, which implied that mannose was a major component of EPS as well. The similar binding pattern indicated similar monosaccharide composition of EPS in *A.* PCC 7120 and *A. cylindria* 29414. ConA binding activity varied with different factors, for example, the growth stages of neighboring vegetative cells and the developmental stage of heterocysts. Comparing the vegetative cells in Figure 4-2A and Figure 4-2B, we could see the vegetative cells in Figure 4-2B were dividing and the ones in Figure 4-2A were not. ConA fluorescence was observed much stronger on heterocysts in Figure 4-2A than in Figure 4-2B. ConA binding was observed in the early stage of heterocysts development when the morphology of heterocyst was not easily distinguishable. However, the mature heterocysts had stronger fluorescence (yellow arrow in Figure 4-2A) compared with young ones (yellow arrow in Figure 4-2B). The differential fluorescence suggested the accumulation of mannose-containing polysaccharides during the development of heterocysts.

Functional heterocysts formed about 24 h after nitrogen step-down in wild type *A.* PCC 7120, while the Δ*all4388* mutant still formed heteocysts, but the shape was not as distinguishable as in wild type, and they cannot fix nitrogen (Lechno-Yossef et al., 2011). In Δ*all4388*, ConA lost binding to heterocysts (Figure 4-2C), which indicated that mannose-containing polysaccharides is missing from the heteocyst EPS due to the absence of this polysaccharide export protein. It was noticeable in Δ*all4388* mutants that some cell junctions were riched in mannose-containing polysaccharides (Figure 4-2C-II)

4.5 Discussion

4.5.1 Specific EPS of akinetes, heterocysts and vegetative cells

The only detectable sugar moiety in vegetative cells was N-acetylglucosamine, which is a derivative of glucose. Glucose is the sole component of cellulose, which is the basic and most common building blocks of polysaccharide in cyanobacteria (Nobles et al. 2001). N-acetylglucosamine was not detectable in heterocysts and akinetes. This result is consistent with our previous proteomic study in *A. cylindrica* 29414 where two enzymes involved in N-acetylglucosamine metabolism: UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase (*acy4163*) and UDP-3-O-acyl N-acetylglucosamine deacetylase (*acy2637*) were only detected in vegetative cells (Qiu et al, unpublished data).

The previous research also indicates that heterocysts and akinetes bear extra polysaccharide layers to adapt their own functions (Zhou and Wolk 2003; Singh and Montgomery 2011; Perez et al. 2016; Perez et al. 2018). Both heterocysts and akinetes differentiated from vegetative cells, heterocysts can induce their neighbor vegetative cells to become akinetes at both sides (Figure 4-1C-I). We found that ConA can bind both

heterocysts and akinetes (Figure 4-1 column II), which is consistent with the presence of both mannose and glucose moieties in the envelope of heterocysts and akinetes (Cardemil and Wolk 1979). Although galactose residues were nearly equal present in the envelope polysaccharide of both heterocyst and akinete (Cardemil and Wolk 1976), the galactose-binding GSL-1 exclusively bound to akinetes, not the heterocysts (Figure 4-1C. II-III). Similarly, the mannose unit is present in both heterocyst and akinetes (Cardemil and Wolk 1979), while two mannose-binding lectins LCA and PSA exclusively binds to akinetes, not the heterocysts (Figure 4-1 D &E, column II). The presence of common ConA-binding polysaccharides in the surface of heterocysts and akinetes suggest that a similar sugar composition in polysaccharide exists between these two cell types, but the distribution of sugar residues in the cell surface are distinguished each other and distinct from their progenitor vegetative cells. The galactose is one of the components in hemicellulose, which could contribute to the properties of akinetes, such as dormancy, being non-motile, and able to geminate (Perez et al. 2018). The greatest number of sugar residues detected in akinetes is also consistent with the finding that akinetes have higher lectin binding activity (Tien et al. 2005).

4.5.2 The directional motility force and the polysaccharide secretion

Recent evidence supported that a Type IV pilus-like nanomotor structure drives the gliding motility and polysaccharide secretion in filamentous cyanobacteria (Khayatan et al. 2015). The structure of specific slime-secreting pores in two cyanobacteria (*Synechocystis* and *Nostoc punctiforme*) were characterized and designated as junctional pore complexes (JPC), which suggested that polysaccharide secretion does not simply provide a suitable surface for gliding, as in other bacteria, but rather generates the directional force for movement

(Wilde and Mullineaux 2015). Our results showed that the secretion of certain polysaccharides (PHA-E, ConA in Figure 4-1, column II & III) occurred at only one terminus of a filament in *A. cylindrica*, which possibly serves as the directional force for filament movement. Based on our observation on motility of 27 *A. cylindrica* filaments, we found that *A. cylindrica* has an average gliding motility with 0.157±0.019 $\mu m.S^{-1}$ and a rotating motility with 0.152±0.022 $\mu m.S^{-1}$ on AAN-1.2% agar plate. However, a young, 9 cell-filament can glide much faster (at 1.219 $\mu m.S^{-1}$) than a long filament. A real time movie showing the motility of *A. cylindrica* on AAN agar plate is provided in supplemental materials (*A. cylindrica* motility video).

4.5.3 The role of Acy1651, a polysaccharide export protein, in oxic nitrogen fixation

The polysaccharide layer is essential for heterocysts to fix dinitrogen in an oxygen-containing milieu. Fox genes (capable of N_2-**f**ixation in the presence of **ox**ygen) are genes required specifically for oxic nitrogen fixation, including the ones that are required for formation of the heterocyst polysaccharide layer (Zhou and Wolk 2003; Huang et al. 2005). Our previous proteomic study discovered that a polysaccharide export protein (Acy1651) is highly expressed in both akinetes and heterocysts, but barely detectable in vegetative cells (Table 4-3). The distribution of the polysaccharide export protein Acy1651 is consistent with the distribution of the ConA-binding polysaccharide in heterocysts (Figure 4-2). Wolk et al and others reported that inactivation of the *acy1651* homologous gene *all4388* in *A.* PCC 7120 did not affect the heterocyst formation, but the mutant heterocyst failed to fix nitrogen aerobically (Ernst et al. 1992; Maldener et al. 2003; Lechno-Yossef et al. 2011). In the Δ*all4388* mutant, the lacking ConA-binding suggest that 1) Acy1651 may involve in the transportation of mannose-containing polysaccharides and in formation

of a polysaccharide layer in heterocysts; and 2) The dysfunction of heterocysts resulted from the missing mannose-containing polysaccharides on their surface that could be a key component for oxygen exclusion and therefore the Δ*all4388* mutant heterocysts lost the aerobic nitrogen fixation function.

4.5.4 Future application of identified signature lectins in vegetative cells, heterocysts and akinetes

We found signature lectin bindings in all three types of cells. ConA for heterocysts and akinetes, GSL-I, LCA and PSA for akinetes, PHA-E for terminal cells, and succinylated WGA and WGA for vegetative cells. Akinetes are enlarged spore-like cells that differentiated from vegetative cells. Akinete of *A. cylindrical* is about 10 times larger than its vegetative cells, which is different from bacterial spores. Yet the akinete formation mechanism is barely known. In our lectin binding assay, LCA showed affinity to developing akinete. Figure 4-1 D-II showed that the two akinetes both bound with LCA, the smaller one was a developing akinete, which may develope into an akinete. These results suggest that mannose's appearance could serve as an earlier molecular marker for akinete differentiation.

The isolation of these three cell types is important since heterocysts are an elegant model system to study the uniquely oxic nitrogen fixation, and akinetes for stress tolerance and germination studies. Traditionally, these type cells had been isolated using CsCl density centrifugation (Wolk 1968), but its purity has some concerns in comparative omics studies. These fluorescein-labeled lectins coupled with flow cytometry may be applicable for fluorescence-activated cell sorting and precision isolation of specific cell type, and further

facilitate the comparative omics studies, such as genomics, transcriptomics, proteomics, metabolomics among these three types of cells.

REFERENCES

Adams, D. G., & Carr, N. G. (1989). Control of heterocyst development in the cyanobacterium *Anabaena* cylindrica. *Microbiology, 135*(4), 839-849.

Agrawal, C., Sen, S., Singh, S., Rai, S., Singh, P. K., Singh, V. K., & Rai, L. C. (2014). Comparative proteomics reveals association of early accumulated proteins in conferring butachlor tolerance in three N(2)-fixing *Anabaena* spp. *J Proteomics, 96*, 271-290. doi:10.1016/j.jprot.2013.11.015

Argueta, C., Yuksek, K., Patel, R., & Summers, M. L. (2006). Identification of Nostoc punctiforme akinete-expressed genes using differential display. *Mol Microbiol, 61*(3), 748-757. doi:10.1111/j.1365-2958.2006.05263.x

Arizmendi, J. M., & Serra, J. L. (1990). Heterocyst and akinete differentiation in cyanobacteria. *New Phytol, 144*, 3-33.

Awai, K., & Wolk, C. P. (2007). Identification of the glycosyl transferase required for synthesis of the principal glycolipid characteristic of heterocysts of *Anabaena* sp. strain PCC 7120. *FEMS Microbiol Lett, 266*(1), 98-102. doi:10.1111/j.1574-6968.2006.00512.x

Awram, P., & Smit, J. (1998). The Caulobacter crescentus paracrystalline S-layer protein is secreted by an ABC transporter (type I) secretion apparatus. *J Bacteriol, 180*(12), 3062-3069.

Badger, M. R., Andrews, T. J., Whitney, S. M., Ludwig, M., Yellowlees, D. C., Leggat, W., & Price, G. D. (1998). The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based CO2-concentrating mechanisms in algae. *Canadian Journal of Botany, 76*, 1052-1071.

Badger, M. R., & Price, G. D. (2017). CO 2 concentrating mechanisms in cyanobacteria : molecular components , their diversity and evolution. *54*, 609-622. doi:10.1093/jxb/erg076

Bergman B, Gallon J, Rai A, Stal L (1997) N2 Fixation by non-heterocystous cyanobacteria. FEMS Microbiology Review 19:139-185 doi: 10.1111/j.1574-6976.1997.tb00296.x

Berman-Frank I, Lundgren P, Falkowski P (2003) Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. Res Microbiol 154:157-164 doi: 10.1016/S0923-2508(03)00029-9

Bekker, A., Holland, H. D., Wang, P.-L., Rumble, D., Stein, H. J., Hannah, J. L., . . . Beukes, N. J. (2004). Dating the rise of atmospheric oxygen. *Nature, 427*, 117-120. doi:10.1038/nature02260

Bi, E., & Lutkenhaus, J. (1990). FtsZ regulates frequency of cell division in Escherichia coli. *J Bacteriol, 172*(5), 2765-2768.

Black, T. A., Cai, Y., & Wolk, C. P. (1993). Spatial expression and autoregulation of hetR, a gene involved in the control of heterocyst development in *Anabaena*. *Molecular microbiology, 9*, 77-84.

Borthakur, D., & Haselkorn, R. (1989). Nucleotide sequence of the gene encoding the 33 kDa water oxidizing polypeptide in *Anabaena* sp. strain PCC 7120 and its expression in *Escherichia coli*. *Plant Molecular Biology Reporter, 13*(4), 427-439.

Bowes, G. (1991). Growth at elevated CO2: photosynthetic responses mediated through Rubisco. *Plant, Cell & Environment, 14*, 795-806.

Bowes, G., Ogren, W. L., & Hageman, R. H. (1971). Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. *Biochemical and biophysical research communications, 45*, 716-722.

Briley, K., Jr., Prepiak, P., Dias, M. J., Hahn, J., & Dubnau, D. (2011). Maf acts downstream of ComGA to arrest cell division in competent cells of B. subtilis. *Mol Microbiol, 81*(1), 23-39. doi:10.1111/j.1365-2958.2011.07695.x

Brooks, S. A. (2017). Lectin Histochemistry: Historical Perspectives, State of the Art, and the Future. *Methods Mol Biol, 1560*, 93-107. doi:10.1007/978-1-4939-6788-9_6

Cyanobacteria: Habitats and Species 11-21 (Springer Berlin Heidelberg 2011).

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

Burnat, M., Herrero, A., & Flores, E. (2014). Compartmentalized cyanophycin metabolism in the diazotrophic filaments of a heterocyst-forming cyanobacterium. *Proc Natl Acad Sci U S A, 111*(10), 3823-3828. doi:10.1073/pnas.1318564111

Cameron, J. C., Wilson, S. C., Bernstein, S. L., & Kerfeld, C. A. (2013). Biogenesis of a bacterial organelle: the carboxysome assembly pathway. *Cell, 155*(5), 1131-1140. doi:10.1016/j.cell.2013.10.044

Campbell, D., Hurry, V., Clarke, A. K., & Gustafsson, P. (1998). Chlorophyll Fluorescence Analysis of Cyanobacterial Photosynthesis and Acclimation. *62*, 667-683.

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

Campbell, E. L., Summers, M. L., Christman, H., Martin, M. E., & Meeks, J. C. (2007). Global gene expression patterns of Nostoc punctiforme in steady-state dinitrogen-grown heterocyst-containing cultures and at single time points during the differentiation of akinetes and hormogonia. *J Bacteriol, 189*(14), 5247-5256. doi:10.1128/JB.00360-07

Cardemil, L., & Wolk, C. P. (1976). The polysaccharides from heterocyst and spore envelopes of a blue-green alga. Methylation analysis and structure of the backbones. *J Biol Chem, 251*(10), 2967-2975.

Cardemil, L., & Wolk, C. P. (1979). The polysaccharides from heterocyst and spore envelopes of a blue-green alga. Structure of the basic repeating unit. *J Biol Chem, 254*(3), 736-741.

Cardemil L, Wolk CP (1981) Isolated heterocysts of *Anabaena* variabilis synthesize envelope polysaccharide. Biochim Biophys Acta 674:265-276

Carrasco, C. D., & Golden, J. W. (1995). Two heterocyst-specific DNA rearrangements of nif operons in *Anabaena* cylindrica and Nostoc sp. strain Mac. *Microbiology, 141*(10), 2479-2487.

Carrasco, C. D., Holliday, S. D., Hansel, A., Lindblad, P., & Golden, J. W. (2005). Heterocyst-specific excision of the *Anabaena* sp. strain PCC 7120 hupL element requires xisC. *J Bacteriol, 187*(17), 6031-6038. doi:10.1128/JB.187.17.6031-6038.2005

Cerveny, J., Sinetova, M. A., Valledor, L., Sherman, L. A., & Nedbal, L. (2013). Ultradian metabolic rhythm in the diazotrophic cyanobacterium Cyanothece sp. ATCC 51142. *Proc Natl Acad Sci U S A, 110*(32), 13210-13215. doi:10.1073/pnas.1301171110

Chabot, J. R., Pedraza, J. M., Luitel, P., & Van Oudenaarden, A. (2007). Stochastic gene expression out-of-steady-state in the cyanobacterial circadian clock. *Nature, 450*, 1249.

Cheng, K., & Costerton, J. (1975). Ultrastructure of cell envelopes of bacteria of the bovine rumen. *Applied microbiology, 29*(6), 841-849.

Christman, H. D., Campbell, E. L., & Meeks, J. C. (2011). Global transcription profiles of the nitrogen stress response resulting in heterocyst or hormogonium development in Nostoc punctiforme. *J Bacteriol, 193*(24), 6874-6886. doi:10.1128/JB.05999-11

Cossar, J. D., Rowell, P., Darling, A. J., Murray, S., Codd, G. A., & Stewart, W. D. P. (1985). Localization of ribulose 1,5-bisphosphate carboxylase/oxygenase in the N I-fixing cyanobacterium *Anabaena* cylindrica. *FEMS Microbiology Letters, 28*, 65-68.

Cumino, A., Curatti, L., Giarrocco, L., & Salerno, G. L. (2002). Sucrose metabolism: *Anabaena* sucrose-phosphate synthase and sucrose-phosphate phosphatase define minimal functional domains shuffled during evolution. *FEBS Lett, 517*(1-3), 19-23.

Cumino, A., Ekeroth, C., & Salerno, G. L. (2001). Sucrose-phosphate phosphatase from *Anabaena* sp. strain PCC 7120: isolation of the protein and gene revealed significant structural differences from the higher-plant enzyme. *Planta, 214*(2), 250-256.

D'Agostino, P. M., Woodhouse, J. N., Makower, A. K., Yeung, A. C. Y., Ongley, S. E., Micallef, M. L., . . . Neilan, B. A. (2016). Advances in genomics, transcriptomics and proteomics of toxin-producing cyanobacteria. *Environmental Microbiology Reports, 8*, 3-13. doi:10.1111/1758-2229.12366

Challenges of metagenomics and single-cell genomics approaches for exploring cyanobacterial diversity, 126 135-146 (2015).

De Philippis, R., Sili, C., Paperi, R., & Vincenzini, M. (2001). Exopolysaccharide-producing cyanobacteria and their possible exploitation: A review. *Journal of Applied Phycology, 13*(4), 293-299. doi:10.1023/a:1017590425924

Diaz-Troya, S., Lopez-Maury, L., Sanchez-Riego, A. M., Roldan, M., & Florencio, F. J. (2014). Redox regulation of glycogen biosynthesis in the cyanobacterium Synechocystis sp. PCC 6803: analysis of the AGP and glycogen synthases. *Mol Plant, 7*(1), 87-100. doi:10.1093/mp/sst137

Dobakova, M., Sobotka, R., Tichy, M., & Komenda, J. (2009). Psb28 protein is involved in the biogenesis of the photosystem II inner antenna CP47 (PsbB) in the cyanobacterium Synechocystis sp. PCC 6803. *Plant Physiol, 149*(2), 1076-1086. doi:10.1104/pp.108.130039

Donze, M., Haveman, J., & Schiereck, P. (1972). Absence of photosystem 2 in heterocysts of the blue-green alga *Anabaena*. *Biochim Biophys Acta, 256*(1), 157-161.

Dudman, W. (1977). The role of surface polysaccharides in natural environments. *Surface carbohydrates of the prokaryotic cell, 4*, 357414.

Ehira, S., & Ohmori, M. (2011). NrrA, a nitrogen-regulated response regulator protein, controls glycogen catabolism in the nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *J Biol Chem, 286*(44), 38109-38114. doi:10.1074/jbc.M111.289124

Eisenhut, M., Ruth, W., Haimovich, M., Bauwe, H., Kaplan, A., & Hagemann, M. (2008). The photorespiratory glycolate metabolism is essential for cyanobacteria and might

have been conveyed endosymbiontically to plants. *Proc Natl Acad Sci U S A, 105*(44), 17199-17204. doi:10.1073/pnas.0807043105

Ekman, M., Picossi, S., Campbell, E. L., Meeks, J. C., & Flores, E. (2013). A Nostoc punctiforme sugar transporter necessary to establish a Cyanobacterium-plant symbiosis. *Plant Physiol, 161*(4), 1984-1992. doi:10.1104/pp.112.213116

Elhai, J., Vepritskiy, A., Muro-Pastor, A., Flores, E., & Wolk, C. (1997). Reduction of conjugal transfer efficiency by three restriction activities of *Anabaena* sp. strain PCC 7120. *J. Bacteriol., 179*, 1998-2005.

Elhai, J., & Wolk, C. P. (1990). Developmental regulation and spatial pattern of expression of the structural genes for nitrogenase in the cyanobacterium *Anabaena*. *EMBO J, 9*(10), 3379-3388.

Ermakova, M., Battchikova, N., Richaud, P., Leino, H., Kosourov, S., Isojarvi, J., . . . Aro, E. M. (2014). Heterocyst-specific flavodiiron protein Flv3B enables oxic diazotrophic growth of the filamentous cyanobacterium *Anabaena* sp. PCC 7120. *Proc Natl Acad Sci U S A, 111*(30), 11205-11210. doi:10.1073/pnas.1407327111

Ernst A, Black T, Cai Y, Panoff JM, Tiwari DN, Wolk CP (1992) Synthesis of nitrogenase in mutants of the cyanobacterium *Anabaena* sp. strain PCC 7120 affected in heterocyst development or metabolism. J Bacteriol 174:6025-6032

Fan, Q., Huang, G., Lechno-Yossef, S., Wolk, C. P., Kaneko, T., & Tabata, S. (2005). Clustered genes required for synthesis and deposition of envelope glycolipids in *Anabaena* sp. strain PCC 7120. *Mol Microbiol, 58*(1), 227-243. doi:10.1111/j.1365-2958.2005.04818.x

Faulkner, M., Rodriguez-Ramos, J., Dykes, G. F., Owen, S. V., Casella, S., Simpson, D. M., . . . Liu, L.-N. (2017). Direct characterization of the native structure and mechanics of cyanobacterial carboxysomes. *Nanoscale, 9*, 10662-10673. doi:10.1039/C7NR02524F

Fay P (1992) Oxygen relations of nitrogen fixation in cyanobacteria. Microbiol Rev 56:340-373

Fay, P., Stewart, W. D. P., Walsby, A. E., & Fogg, G. E. (1968). Is the heterocyst the site of nitrogen fixation in blue-green algae? *Nature, 220*, 810-812.

Fay, P., & Walsby, A. E. (1966). Metabolic activities of isolated heterocysts of the blue-green alga *Anabaena* cylindrica. *Nature, 209*, 94-95.

Fiedler, G., Arnold, M., Hannus, S., & Maldener, I. (1998). The DevBCA exporter is essential for envelope formation in heterocysts of the cyanobacterium *Anabaena* sp. strain PCC 7120. *Mol Microbiol, 27*(6), 1193-1202.

Flores, E., & Herrero, A. (2010). Compartmentalized function through cell differentiation in filamentous cyanobacteria. *Nat Rev Microbiol, 8*(1), 39-50. doi:10.1038/nrmicro2242

Fogg, P. C., Colloms, S., Rosser, S., Stark, M., & Smith, M. C. (2014). New applications for phage integrases. *J Mol Biol, 426*(15), 2703-2716. doi:10.1016/j.jmb.2014.05.014

Forchhammer, K. (2004). Global carbon/nitrogen control by PII signal transduction in cyanobacteria: from signals to targets. *FEMS microbiology reviews, 28*, 319-333.

Fujisawa, T., Narikawa, R., Maeda, S. I., Watanabe, S., Kanesaki, Y., Kobayashi, K., . . . Nakamura, Y. (2017). CyanoBase: A large-scale update on its 20th anniversary. *Nucleic Acids Research, 45*, D551-D554. doi:10.1093/nar/gkw1131

Gallon, J. R. (1981). The oxygen sensitivity of nitrogenase: a problem for biochemists and micro-organisms. *Trends in Biochemical Sciences, 6*, 19-23.

Gandham, L., Nomellini, J. F., & Smit, J. (2012). Evaluating secretion and surface attachment of SapA, an S-layer-associated metalloprotease of Caulobacter crescentus. *Arch Microbiol, 194*(10), 865-877. doi:10.1007/s00203-012-0819-9

Gantt, E. (2011). Oxygenic photosynthesis and the distribution of chloroplasts. *Photosynth Res, 107*(1), 1-6. doi:10.1007/s11120-010-9605-1

Gerbino, E., Carasi, P., Mobili, P., Serradell, M. A., & Gomez-Zavaglia, A. (2015). Role of S-layer proteins in bacteria. *World J Microbiol Biotechnol, 31*(12), 1877-1887. doi:10.1007/s11274-015-1952-9

Gerphagnon M, Latour D, Colombet J, Sime-Ngando T (2013) Fungal parasitism: life cycle, dynamics and impact on cyanobacterial blooms. PLoS One 8:e60894 doi: 10.1371/journal.pone.0060894

Giovannoni, S. J., Turner, S., Olsen, G. J., Barns, S., Lane, D. J., & Pace, N. R. (1988). Evolutionary Relationships among Cyanobacteria and Green Chloroplasts. *170*, 3584-3592.

Glazer, N. (1978). Of the phycobilisomes of. *253*, 8303-8310.

Photosystem I in cyanobacteria 319-360 (Springer 1994).

Golden, J. W., Carrasco, C. D., Mulligan, M. E., Schneider, G. J., & Haselkorn, R. (1988). Deletion of a 55-kilobase-pair DNA element from the chromosome during

heterocyst differentiation of *Anabaena* sp. strain PCC 7120. *J Bacteriol*, *170*(11), 5034-5041.

Golden, J. W., Robinson, S. J., & Haselkorn, R. (1985). Rearrangement of nitrogen fixation genes during heterocyst differentiation in the cyanobacterium *Anabaena*. *Nature*, *314*(6010), 419-423.

Golden, J. W., & Yoon, H. S. (2003). Heterocyst development in *Anabaena*. *Curr Opin Microbiol*, *6*(6), 557-563.

Gore, J., Bryant, Z., Stone, M. D., Nollmann, M., Cozzarelli, N. R., & Bustamante, C. (2006). Mechanochemical analysis of DNA gyrase using rotor bead tracking. *Nature*, *439*(7072), 100-104. doi:10.1038/nature04319

Grossman, A. R., Schaefer, M. R., Chiang, G. G., & Collier, J. L. (1993). The phycobilisome, a light-harvesting complex responsive to environmental conditions. *Microbiological reviews*, *57*, 725-749.

Guerreiro, A. C. L., Penning, R., Raaijmakers, L. M., Axman, I. M., Heck, A. J. R., & Altelaar, A. F. M. (2016). Monitoring light/dark association dynamics of multi-protein complexes in cyanobacteria using size exclusion chromatography-based proteomics. *Journal of Proteomics, 142*, 33-44. doi:10.1016/j.jprot.2016.04.030

Gupta, M., & Carr, N. (1981). Enzyme activities related to cyanophycin metabolism in heterocysts and vegetative cells of *Anabaena* spp. *J Gen Microbiol, 125*(1), 17–23.

Guskov, A., Kern, J., Gabdulkhakov, A., Broser, M., Zouni, A., & Saenger, W. (2009). Cyanobacterial photosystem II at 2.9-Å resolution and the role of quinones, lipids, channels and chloride. *Nature structural & molecular biology, 16*, 334-342.

Halfmann, C., Gu, L., Gibbons, W., & Zhou, R. (2014). Genetically engineering cyanobacteria to convert CO(2), water, and light into the long-chain hydrocarbon farnesene. *Appl Microbiol Biotechnol, 98*(23), 9869-9877. doi:10.1007/s00253-014-6118-4

Hanlon, S. M. O., & Cozy, L. M. (2016). The heterocyst regulatory protein HetP and its homologs modulate heterocyst commitment in *Anabaena* sp . strain PCC 7120. doi:10.1073/pnas.1610533113

Haselkorn, R. (1992). Developmentally regulated gene rearrangements in prokaryotes. *Annu Rev Genet, 26*, 113-130. doi:10.1146/annurev.ge.26.120192.000553

Hellier, P., Al-Haj, L., Talibi, M., Purton, S., & Ladommatos, N. (2013). Combustion and emissions characterization of terpenes with a view to their biological production in cyanobacteria. *Fuel, 111*, 670-688. doi:10.1016/j.fuel.2013.04.042

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

Higa, K. C., Rajagopalan, R., Risser, D. D., Rivers, O. S., Tom, S. K., Videau, P., & Callahan, S. M. (2012). The RGSGR amino acid motif of the intercellular signalling protein, HetN, is required for patterning of heterocysts in *Anabaena* sp. strain PCC 7120. *Mol Microbiol, 83*(4), 682-693. doi:10.1111/j.1365-2958.2011.07949.x

Higo, A., Katoh, H., Ohmori, K., Ikeuchi, M., & Ohmori, M. (2006). The role of a gene cluster for trehalose metabolism in dehydration tolerance of the filamentous cyanobacterium *Anabaena* sp. PCC 7120. *Microbiology, 152*(Pt 4), 979-987. doi:10.1099/mic.0.28583-0

Hilton, J. A., Meeks, J. C., & Zehr, J. P. (2016). Surveying DNA Elements within Functional Genes of Heterocyst-Forming Cyanobacteria. *PLoS One, 11*(5), e0156034. doi:10.1371/journal.pone.0156034

Hirosawa, T., & Wolk, C. P. (1979). Factors Controlling the Formation of Akinetes Adjacent to Heterocysts in the Cyanobacterium Cylindrospermum licheniforme Kütz. *Microbiology, 114*(2), 423-432. doi:doi:10.1099/00221287-114-2-423

Hor, L., Dobson, R. C., Downton, M. T., Wagner, J., Hutton, C. A., & Perugini, M. A. (2013). Dimerization of bacterial diaminopimelate epimerase is essential for catalysis. *J Biol Chem, 288*(13), 9238-9248. doi:10.1074/jbc.M113.450148

Hu, H. X., Jiang, Y. L., Zhao, M. X., Cai, K., Liu, S., Wen, B., . . . Zhou, C. Z. (2015). Structural insights into HetR-PatS interaction involved in cyanobacterial pattern formation. *Sci Rep, 5*, 16470. doi:10.1038/srep16470

Hu, N. T., Thiel, T., Giddings, T. H., Jr., & Wolk, C. P. (1981). New *Anabaena* and Nostoc cyanophages from sewage settling ponds. *Virology, 114*(1), 236-246.

Hu, Y., Fay, A. W., Lee, C. C., Wiig, J. A., & Ribbe, M. W. (2010). Dual functions of NifEN: insights into the evolution and mechanism of nitrogenase. *Dalton Trans, 39*(12), 2964-2971. doi:10.1039/b922555b

Hu, Y., Zhang, X., Shi, Y., Zhou, Y., Zhang, W., Su, X. D., . . . Jin, C. (2011). Structures of *Anabaena* calcium-binding protein CcbP: insights into Ca2+ signaling during heterocyst differentiation. *J Biol Chem, 286*(14), 12381-12388. doi:10.1074/jbc.M110.201186

Huang G et al. (2005) Clustered genes required for the synthesis of heterocyst envelope polysaccharide in *Anabaena* sp. strain PCC 7120. J Bacteriol 187:1114-1123 doi: 10.1128/JB.187.3.1114-1123.2005

Huang, X., Dong, Y., & Zhao, J. (2004a). HetR homodimer is a DNA-binding protein required for heterocyst differentiation , and the DNA-binding activity is inhibited by PatS. *101*.

Huang, X., Dong, Y., & Zhao, J. (2004b). HetR homodimer is a DNA-binding protein required for heterocyst differentiation, and the DNA-binding activity is inhibited by PatS. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 4848-4853.

Ishiura, M. (1998). Expression of a Gene Cluster kaiABC as a Circadian Feedback Process in Cyanobacteria. *Science, 281*, 1519-1523. doi:10.1126/science.281.5382.1519

Jong-Bong, L., Hite, R. K., Hamdan, S. M., Xie, X. S., Richardson, C. C., & Van Oijen, A. M. (2006). DNA primase acts as a molecular brake in DNA replication. *Nature, 439*(7076), 621.

Juttner, F. (1983). 14C-labeled metabolites in heterocysts and vegetative cells of *Anabaena* cylindrica filaments and their presumptive function as transport vehicles of organic carbon and nitrogen. *J Bacteriol, 155*(2), 628-633.

Kaneko, T., Nakamura, Y., Wolk, C. P., Kuritz, T., Sasamoto, S., Watanabe, A., . . . Tabata, S. (2001). Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res, 8*(5), 205-213; 227-253.

Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., . . . Tabata, 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. doi:10.1093/dnares/3.3.109

Kaplan, A., & Reinhold, L. (1999). CO 2 CONCENTRATING MECHANISMS IN MICROORGANISMS.

Kawai, E., Akatsuka, H., Idei, A., Shibatani, T., & Omori, K. (1998). Serratia marcescens S-layer protein is secreted extracellularly via an ATP-binding cassette exporter, the Lip system. *Mol Microbiol, 27*(5), 941-952.

Kehr, J. C., & Dittmann, E. (2015). Biosynthesis and function of extracellular glycans in cyanobacteria. *Life (Basel), 5*(1), 164-180. doi:10.3390/life5010164

Khayatan, B., Meeks, J. C., & Risser, D. D. (2015). Evidence that a modified type IV pilus-like system powers gliding motility and polysaccharide secretion in filamentous cyanobacteria. *Mol Microbiol, 98*(6), 1021-1036. doi:10.1111/mmi.13205

Khudyakov, I., & Wolk, C. P. (1996). Evidence that the hanA gene coding for HU protein is essential for heterocyst differentiation in, and cyanophage A-4(L) sensitivity of, *Anabaena* sp. strain PCC 7120. *J Bacteriol, 178*(12), 3572-3577.

Khudyakov, I. Y., & Golden, J. W. (2004). Different functions of HetR, a master regulator of heterocyst differentiation in *Anabaena* sp. PCC 7120, can be separated by mutation. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 16040-16045.

Kitayama, Y., Iwasaki, H., Nishiwaki, T., & Kondo, T. (2003). KaiB functions as an attenuator of KaiC phosphorylation in the cyanobacterial circadian clock system. *EMBO Journal, 22*, 2127-2134. doi:10.1093/emboj/cdg212

Kolman, M. A., Nishi, C. N., Perez-Cenci, M., & Salerno, G. L. (2015). Sucrose in cyanobacteria: from a salt-response molecule to play a key role in nitrogen fixation. *Life (Basel), 5*(1), 102-126. doi:10.3390/life5010102

Kondo, T., Strayer, C. A., Kulkarni, R. D., Taylor, W., Ishiura, M., Golden, S. S., & Johnson, C. H. (1993). Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in cyanobacteria. *Proc Natl Acad Sci U S A, 90*(12), 5672-5676.

Kondo, T., Tsinoremas, N. F., Golden, S. S., & Johnson, C. H. (1994). Circadian clock mutants of cyanobacteria. *Science, 266*, 1233.

Kumar, K., Mella-Herrera, R. A., & Golden, J. W. (2010). Cyanobacterial heterocysts. *Cold Spring Harb Perspect Biol, 2*(4), a000315. doi:10.1101/cshperspect.a000315

Larsson, J., Nylander, J. A., & Bergman, B. (2011). Genome fluctuations in cyanobacteria reflect evolutionary, developmental and adaptive traits. *BMC Evolutionary Biology, 11*, 187. doi:10.1186/1471-2148-11-187

Laurent, S., Chen, H., Bédu, S., Ziarelli, F., Peng, L., & Zhang, C.-C. (2005). Nonmetabolizable analogue of 2-oxoglutarate elicits heterocyst differentiation under repressive conditions in *Anabaena* sp. PCC 7120. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 9907-9912.

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

Lechno-Yossef, S., Fan, Q., Ehira, S., Sato, N., & Wolk, C. P. (2006). Mutations in four regulatory genes have interrelated effects on heterocyst maturation in *Anabaena* sp. strain PCC 7120. *J Bacteriol, 188*(21), 7387-7395. doi:10.1128/JB.00974-06

Lechno-Yossef, S., Fan, Q., Wojciuch, E., & Wolk, C. P. (2011). Identification of ten *Anabaena* sp. genes that under aerobic conditions are required for growth on dinitrogen but not for growth on fixed nitrogen. *J Bacteriol, 193*(14), 3482-3489. doi:10.1128/JB.05010-11

Leganés, F. (1994). Genetic evidence that hepA gene is involved in the normal deposition of the envelope of both heterocysts and akinetes in *Anabaena* variabilis ATCC 29413. *FEMS microbiology letters, 123*, 63-67.

Leganes, F., Fernandez-Pinas, F., & Wolk, C. P. (1998). A transposition-induced mutant of Nostoc ellipsosporum implicates an arginine-biosynthetic gene in the formation of cyanophycin granules and of functional heterocysts and akinetes. *Microbiology, 144 (Pt 7)*, 1799-1805. doi:10.1099/00221287-144-7-1799

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

Liang, W., Zhou, Y., Wang, L., You, X., Zhang, Y., Cheng, C. L., & Chen, W. (2012). Ultrastructural, physiological and proteomic analysis of Nostoc flagelliforme in response to dehydration and rehydration. *J Proteomics, 75*(18), 5604-5627. doi:10.1016/j.jprot.2012.07.041

Liu, L.-N., Chen, X.-L., Zhang, Y.-Z., & Zhou, B.-C. (2005). Characterization, structure and function of linker polypeptides in phycobilisomes of cyanobacteria and red

algae: an overview. *Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1708*, 133-142.

Llacer, J. L., Contreras, A., Forchhammer, K., Marco-Marin, C., Gil-Ortiz, F., Maldonado, R., . . . Rubio, V. (2007). The crystal structure of the complex of PII and acetylglutamate kinase reveals how PII controls the storage of nitrogen as arginine. *Proc Natl Acad Sci U S A, 104*(45), 17644-17649. doi:10.1073/pnas.0705987104

Lopez-Igual, R., Flores, E., & Herrero, A. (2010). Inactivation of a heterocyst-specific invertase indicates a principal role of sucrose catabolism in heterocysts of *Anabaena* sp. *J Bacteriol, 192*(20), 5526-5533. doi:10.1128/JB.00776-10

Lu, J. J., Shi, L., Chen, W. L., & Wang, L. (2014). The regulation of HanA during heterocyst development in cyanobacterium *Anabaena* sp. PCC 7120. *World J Microbiol Biotechnol, 30*(10), 2673-2680. doi:10.1007/s11274-014-1691-3

Madan, A. P., & Nierzwicki-Bauer, S. A. (1993). In situ detection of transcripts for ribulose-1,5-bisphosphate carboxylase in cyanobacterial heterocysts. *J Bacteriol, 175*(22), 7301-7306.

Maldener, I., Fiedler, G., Ernst, A., Fernandez-Pinas, F., & Wolk, C. P. (1994). Characterization of devA, a gene required for the maturation of proheterocysts in the cyanobacterium *Anabaena* sp. strain PCC 7120. *J Bacteriol, 176*(24), 7543-7549.

Maldener, I., Hannus, S., & Kammerer, M. (2003). Description of five mutants of the cyanobacterium *Anabaena* sp strain PCC 7120 affected in heterocyst differentiation and identification of the transposon-tagged genes. *FEMS Microbiol Lett, 224*(2), 205-213.

Marczak M, Mazur A, Koper P, Zebracki K, Skorupska A (2017) Synthesis of Rhizobial Exopolysaccharides and Their Importance for Symbiosis with Legume Plants. Genes 8 doi: ARTN 36010.3390/genes8120360

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., . . . Penny, D. (2002). Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci U S A, 99*(19), 12246-12251. doi:10.1073/pnas.182432999

Maurya, G. K., Modi, K., & Misra, H. S. (2016). Divisome and segrosome components of Deinococcus radiodurans interact through cell division regulatory proteins. *Microbiology*. doi:10.1099/mic.0.000330

Meeks, J. C., Campbell, E. L., Summers, M. L., & Wong, F. C. (2002). Cellular differentiation in the cyanobacterium Nostoc punctiforme. *Arch Microbiol, 178*(6), 395-403. doi:10.1007/s00203-002-0476-5

Meeks, J. C., Wycoff, K. L., Chapman, J. S., & Enderlin, C. S. (1983). Regulation of expression of nitrate and dinitrogen assimilation by *Anabaena* species. *Appl Environ Microbiol, 45*(4), 1351-1359.

Mevarech, M., Rice, D., & Haselkorn, R. (1980). Nucleotide sequence of a cyanobacterial nifH gene coding for nitrogenase reductase. *Proc Natl Acad Sci U S A, 77*(11), 6476-6480.

Millineaux, P. M., Gallon, J. R., & Chaplin, A. E. (1981). Acetylene reduction (nitrogen fixation) by cyanobacteria grown under alternating light-dark cycles. *FEMS Microbiology Letters, 10*, 245-247.

Minaeva, E., Forchhammer, K., & Ermilova, E. (2015). Glutamine Assimilation and Feedback Regulation of L-acetyl-N-glutamate Kinase Activity in Chlorella variabilis NC64A Results in Changes in Arginine Pools. *Protist, 166*(5), 493-505. doi:10.1016/j.protis.2015.08.001

Mitschke, J., Vioque, A., Haas, F., Hess, W. R., & Muro-Pastor, A. M. (2011). Dynamics of transcriptional start site selection during nitrogen stress-induced cell differentiation in *Anabaena* sp. PCC7120. *Proc Natl Acad Sci U S A, 108*(50), 20130-20135. doi:10.1073/pnas.1112724108

Moore, D., Donohue, M. O., Garnett, C., Critchley, C., & Shaw, G. (2005). Factors affecting akinete differentiation in Cylindrospermopsis raciborskii (Nostocales , Cyanobacteria). 345-352. doi:10.1111/j.1365-2427.2004.01324.x

Mori, T., & Hirschie, C. (1996). Circadian gating of cell division in cyanobacteria growing with average doubling times of less than 24 hours. *93*, 10183-10188.

Mulkidjanian, A. Y., Koonin, E. V., Makarova, K. S., Mekhedov, S. L., Sorokin, A., Wolf, Y. I., . . . Galperin, M. Y. (2006). The cyanobacterial genome core and the origin of photosynthesis. *Proc Natl Acad Sci U S A, 103*(35), 13126-13131. doi:10.1073/pnas.0605709103

Murakami, A., & Fujita, Y. (1988). Steady state of photosynthesis in cyanobacterial photosynthetic systems before and after regulation of electron transport composition: Overall rate of photosynthesis and PSI/PS II composition. *Plant and cell physiology, 29*, 305-311.

Muro-Pastor, A. M. (2014). The heterocyst-specific NsiR1 small RNA is an early marker of cell differentiation in cyanobacterial filaments. *MBio, 5*(3), e01079-01014. doi:10.1128/mBio.01079-14

Muro-Pastor, A. M., & Hess, W. R. (2012). Heterocyst differentiation: from single mutants to global approaches. *Trends Microbiol, 20*(11), 548-557. doi:10.1016/j.tim.2012.07.005

Murry, M. A., & Wolk, C. P. (1989). Evidence that the barrier to the penetration of oxygen into heterocysts depends upon two layers of the cell envelope. *Archives of Microbiology, 151*, 469-474.

Muszynski, A., Laus, M., Kijne, J. W., & Carlson, R. W. (2011). Structures of the lipopolysaccharides from Rhizobium leguminosarum RBL5523 and its UDP-glucose dehydrogenase mutant (exo5). *Glycobiology, 21*(1), 55-68. doi:10.1093/glycob/cwq131

Napoli, A., Valenti, A., Salerno, V., Nadal, M., Garnier, F., Rossi, M., & Ciaramella, M. (2004). Reverse gyrase recruitment to DNA after UV light irradiation in Sulfolobus solfataricus. *J Biol Chem, 279*(32), 33192-33198. doi:10.1074/jbc.M402619200

Nicolaisen, K., Hahn, A., & Schleiff, E. (2009). The cell wall in heterocyst formation by *Anabaena* sp. PCC 7120. *J Basic Microbiol, 49*(1), 5-24. doi:10.1002/jobm.200800300

Nobles, D. R., Romanovicz, D. K., & Brown, R. M., Jr. (2001). Cellulose in cyanobacteria. Origin of vascular plant cellulose synthase? *Plant Physiol, 127*(2), 529-542.

Nomata, J., Maeda, M., Isu, A., Inoue, K., & Hisabori, T. (2015). Involvement of thioredoxin on the scaffold activity of NifU in heterocyst cells of the diazotrophic

cyanobacterium *Anabaena* sp. strain PCC 7120. *J Biochem, 158*(3), 253-261. doi:10.1093/jb/mvv046

Ochoa de Alda, J. A. G., Esteban, R., Diago, M. L., & Houmard, J. (2014). The plastid ancestor originated among one of the major cyanobacterial lineages. *Nature Communications, 5*, 4937. doi:10.1038/ncomms5937

Oliveira, P., Martins, N. M., Santos, M., Couto, N. A., Wright, P. C., & Tamagnini, P. (2015). The *Anabaena* sp. PCC 7120 Exoproteome: Taking a Peek outside the Box. *Life (Basel), 5*(1), 130-163. doi:10.3390/life5010130

Olsson-Francis, K., de la Torre, R., Towner, M. C., & Cockell, C. S. (2009). Survival of akinetes (resting-state cells of cyanobacteria) in low earth orbit and simulated extraterrestrial conditions. *Orig Life Evol Biosph, 39*(6), 565-579. doi:10.1007/s11084-009-9167-4

Ow, S. Y., Cardona, T., Taton, A., Magnuson, A., Lindblad, P., Stensjo, K., & Wright, P. C. (2008). Quantitative shotgun proteomics of enriched heterocysts from Nostoc sp. PCC 7120 using 8-plex isobaric peptide tags. *J Proteome Res, 7*(4), 1615-1628. doi:10.1021/pr700604v

Panda, B., Basu, B., Rajaram, H., & Kumar Apte, S. (2014). Methyl viologen responsive proteome dynamics of *Anabaena* sp. strain PCC7120. *Proteomics, 14*(16), 1895-1904. doi:10.1002/pmic.201300522

Pandey, S., Rai, R., & Rai, L. C. (2012). Proteomics combines morphological, physiological and biochemical attributes to unravel the survival strategy of *Anabaena* sp. PCC7120 under arsenic stress. *J Proteomics, 75*(3), 921-937. doi:10.1016/j.jprot.2011.10.011

Pang, J. D., Zhuang, B. X., Mai, K., Chen, R. F., Wang, J., & Zhang, L. M. (2015). Click modification of helical amylose by poly(l-lysine) dendrons for non-viral gene delivery. *Mater Sci Eng C Mater Biol Appl, 49*, 485-492. doi:10.1016/j.msec.2015.01.011

Paniagua-Michel Jde, J., Olmos-Soto, J., & Morales-Guerrero, E. R. (2014). Algal and microbial exopolysaccharides: new insights as biosurfactants and bioemulsifiers. *Adv Food Nutr Res, 73*, 221-257. doi:10.1016/B978-0-12-800268-1.00011-1

Paz-Yepes, J., Flores, E., & Herrero, A. (2009). Expression and mutational analysis of the glnB genomic region in the heterocyst-forming Cyanobacterium *Anabaena* sp. strain PCC 7120. *J Bacteriol, 191*(7), 2353-2361. doi:10.1128/JB.01381-08

Pereira, S., Zille, A., Micheletti, E., Moradas-Ferreira, P., De Philippis, R., & Tamagnini, P. (2009). Complexity of cyanobacterial exopolysaccharides: composition, structures, inducing factors and putative genes involved in their biosynthesis and assembly. *FEMS Microbiol Rev, 33*(5), 917-941. doi:10.1111/j.1574-6976.2009.00183.x

Perez, R., Forchhammer, K., Salerno, G., & Maldener, I. (2015). Clear differences in metabolic and morphological adaptations of akinetes of two Nostocales living in different habitats. *Microbiology*. doi:10.1099/mic.0.000230

Perez, R., Forchhammer, K., Salerno, G., & Maldener, I. (2016). Clear differences in metabolic and morphological adaptations of akinetes of two Nostocales living in different habitats. *Microbiology, 162*(2), 214-223. doi:10.1099/mic.0.000230

Peter, A. P., Lakshmanan, K., Mohandass, S., Varadharaj, S., Thilagar, S., Kareem, K. A. A., . . . Lakshmanan, U. (2015). Cyanobacterial knowledgeBase (CKB), a

compendium of cyanobacterial genomes and proteomes. *PLoS One, 10*, 1-12. doi:10.1371/journal.pone.0136262

Picossi, S., Valladares, A., Flores, E., & Herrero, A. (2004). Nitrogen-regulated genes for the metabolism of cyanophycin, a bacterial nitrogen reserve polymer: expression and mutational analysis of two cyanophycin synthetase and cyanophycinase gene clusters in heterocyst-forming cyanobacterium *Anabaena* sp. PCC 7120. *J Biol Chem, 279*(12), 11582-11592. doi:10.1074/jbc.M311518200

Porchia, A. C., Curatti, L., & Salerno, G. L. (1999). Sucrose metabolism in cyanobacteria: sucrose synthase from *Anabaena* sp. strain PCC 7119 is remarkably different from the plant enzymes with respect to substrate affinity and amino-terminal sequence. *Planta, 210*(1), 34-40. doi:10.1007/s004250050651

Prasanna, R., Kumar, R., Sood, A., Prasanna, B. M., & Singh, P. K. (2006). Morphological , physiochemical and molecular characterization of *Anabaena* strains. *161*. doi:10.1016/j.micres.2005.08.001

Quinn, H. J., Cameron, A. D., & Dorman, C. J. (2014). Bacterial regulon evolution: distinct responses and roles for the identical OmpR proteins of Salmonella Typhimurium and Escherichia coli in the acid stress response. *PLoS Genet, 10*(3), e1004215. doi:10.1371/journal.pgen.1004215

Rajaniemi, P., Hrouzek, P., Rantala, A., & Hoffmann, L. (2017). Phylogenetic and morphological evaluation of the genera *Anabaena* , Aphanizomenon , Trichormus and Nostoc (Nostocales , Cyanobacteria). 11-26. doi:10.1099/ijs.0.63276-0

Ramon-Maiques, S., Marina, A., Gil-Ortiz, F., Fita, I., & Rubio, V. (2002). Structure of acetylglutamate kinase, a key enzyme for arginine biosynthesis and a prototype for the amino acid kinase enzyme family, during catalysis. *Structure, 10*(3), 329-342.

Reddy, K. J., Haskell, J. B., Sherman, D. M., & Sherman, L. A. (1993). Unicellular, aerobic nitrogen-fixing cyanobacteria of the genus Cyanothece. *J Bacteriol, 175*(5), 1284-1292.

Renger, G., & Renger, T. (2008). Photosystem II: the machinery of photosynthetic water splitting. *Photosynthesis research, 98*, 53-80.

Rippka, R., & Herdman, M. (1985). Division patterns and cellular differentiation in cyanobacteria. *Ann Inst Pasteur Microbiol, 136A*(1), 33-39.

Risser, D. D., & Callahan, S. M. (2008). HetF and PatA control levels of HetR in *Anabaena* sp. strain PCC 7120. *J Bacteriol, 190*(23), 7645-7654. doi:10.1128/JB.01110-08

Risser, D. D., Wong, F. C., & Meeks, J. C. (2012). Biased inheritance of the protein PatN frees vegetative cells to initiate patterned heterocyst differentiation. *Proc Natl Acad Sci U S A, 109*(38), 15342-15347. doi:10.1073/pnas.1207530109

Rivers, O. S., Videau, P., & Callahan, S. M. (2014). Mutation of sepJ reduces the intercellular signal range of a hetN-dependent paracrine signal , but not of a patS-dependent signal , in the filamentous cyanobacterium. *94*, 1260-1271. doi:10.1111/mmi.12836

Rossetti, V., Schirrmeister, B. E., Bernasconi, M. V., & Bagheri, H. C. (2010). The evolutionary path to terminal differentiation and division of labor in cyanobacteria. *J Theor Biol, 262*(1), 23-34. doi:10.1016/j.jtbi.2009.09.009

Sakamoto, T., Yoshida, T., Arima, H., Hatanaka, Y., Takani, Y., & Tamaru, Y. (2009). Accumulation of trehalose in response to desiccation and salt stress in the terrestrial cyanobacterium Nostoc commune. *Phycological Research, 57*(1), 66-73. doi:10.1111/j.1440-1835.2008.00522.x

Sandh, G., Ramstrom, M., & Stensjo, K. (2014). Analysis of the early heterocyst Cys-proteome in the multicellular cyanobacterium Nostoc punctiforme reveals novel insights into the division of labor within diazotrophic filaments. *BMC Genomics, 15*, 1064. doi:10.1186/1471-2164-15-1064

Sarma, T. A., Ahuja, G., & Khattar, J. I. (2004). Nutrient stress causes akinete differentiation in cyanobacterium *Anabaena* torulosa with concomitant increase in nitrogen reserve substances. *Folia Microbiol (Praha), 49*(5), 557-561.

Schirrmeister, B. E., Antonelli, A., & Bagheri, H. C. (2011). The origin of multicellularity in cyanobacteria. *BMC Evolutionary Biology, 11*, 45. doi:10.1186/1471-2148-11-45

Schirrmeister, B. E., Antonelli, A., & Bagheri, H. C. (2011). The origin of multicellularity in cyanobacteria. *BMC Evol Biol, 11*, 45. doi:10.1186/1471-2148-11-45

Schirrmeister, B. E., Gugger, M., & Donoghue, P. C. J. (2015). Cyanobacteria and the Great Oxidation Event: evidence from genes and fossils. *Palaeontology, 58*, 769-785. doi:10.1111/pala.12178

Schussler, A., Meyer, T., Gehrig, H., & Kluge, M. (1997). Variations of lectin binding sites in extracellular glycoconjugates during the life cycle of Nostoc punctiforme, a potentially endosymbiotic cyanobacterium. *European Journal of Phycology, 32*(3), 233-239. doi:Doi 10.1017/S0967026297001340

Serrano, G. C. D., Figueira, T. R. E. S., Kiyota, E., Zanata, N., & Arruda, P. (2012). Lysine degradation through the saccharopine pathway in bacteria: LKR and SDH in bacteria and its relationship to the plant and animal enzymes. *Febs Letters, 586*(6), 905-911. doi:10.1016/j.febslet.2012.02.023

Shen, G., Balasubramanian, R., Wang, T., Wu, Y., Hoffart, L. M., Krebs, C., . . . Golbeck, J. H. (2007). SufR coordinates two [4Fe-4S]2+, 1+ clusters and functions as a transcriptional repressor of the sufBCDS operon and an autoregulator of sufR in cyanobacteria. *J Biol Chem, 282*(44), 31909-31919. doi:10.1074/jbc.M705554200

Shevchenko, A., Wilm, M., Vorm, O., & Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem, 68*(5), 850-858.

Simon, R. D. (1977). Macromolecular composition of spores from the filamentous cyanobacterium A nabaena cylindrica. *J Bacteriol, 129*(2), 1154-1155.

Singh, S. P., & Montgomery, B. L. (2011). Determining cell shape: adaptive regulation of cyanobacterial cellular differentiation and morphology. *Trends Microbiol, 19*(6), 278-285. doi:10.1016/j.tim.2011.03.001

Smarda, J., Smajs, D., Komrska, J., & Krzyzanek, V. (2002). S-layers on cell walls of cyanobacteria. *Micron, 33*(3), 257-277.

Nitrogen Fixation in Cyanobacteria (John Wiley & Sons, Ltd 2001).

Stanier, R. Y., Deruelles, J., Rippka, R., Herdman, M., & Waterbury, J. B. (1979). Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. *Microbiology, 111*, 1-61. doi:10.1099/00221287-111-1-1

Stewart, W. D. P., Haystead, A., & Pearson, H. W. (1969). Nitrogenase Activity in Heterocysts of Blue–Green Algae. *Nature, 224*, 226-228.

Stockel, J., Welsh, E. A., Liberton, M., Kunnvakkam, R., Aurora, R., & Pakrasi, H. B. (2008). Global transcriptomic analysis of Cyanothece 51142 reveals robust diurnal oscillation of central metabolic processes. *Proc Natl Acad Sci U S A, 105*(16), 6156-6161. doi:10.1073/pnas.0711068105

Sukenik, A., Kaplan-levy, R. N., Welch, J. M., & Post, A. F. (2011). Massive multiplication of genome and ribosomes in dormant cells (akinetes) of Aphanizomenon ovalisporum (Cyanobacteria). *The ISME Journal, 6*, 670-679. doi:10.1038/ismej.2011.128

Sukenik, A., Kaplan-Levy, R. N., Welch, J. M., & Post, A. F. (2012). Massive multiplication of genome and ribosomes in dormant cells (akinetes) of Aphanizomenon ovalisporum (Cyanobacteria). *ISME J, 6*(3), 670-679. doi:10.1038/ismej.2011.128

Sukenik, A., Maldener, I., Delhaye, T., Viner-Mozzini, Y., Sela, D., & Bormans, M. (2015). Carbon assimilation and accumulation of cyanophycin during the development of dormant cells (akinetes) in the cyanobacterium Aphanizomenon ovalisporum. *Front Microbiol, 6*, 1067. doi:10.3389/fmicb.2015.01067

Summers, M. L., & Meeks, J. C. (1996). Transcriptional regulation of zwf, encoding glucose-6-phosphate dehydrogenase, from the cyanobacterium Nostoc punctiforme strain ATCC 29133. *Mol Microbiol, 22*(3), 473-480.

The biochemistry and molecular regulation of carbon dioxide metabolism in cyanobacteria 437-467 (Springer 1994).

Tamaru, Y., Takani, Y., Yoshida, T., & Sakamoto, T. (2005). Crucial role of extracellular polysaccharides in desiccation and freezing tolerance in the terrestrial

cyanobacterium Nostoc commune. *Appl Environ Microbiol, 71*(11), 7327-7333. doi:10.1128/AEM.71.11.7327-7333.2005

Tien CJ, Sigee DC, White KN (2005) Characterization of surface sugars on algal cells with fluorescein isothiocyanate-conjugated lectins. Protoplasma 225:225-233 doi: 10.1007/s00709-005-0092-8

Tel-Or, E., & Stewart, W. D. (1976). Photosynthetic electron transport, ATP synthesis and nitrogenase activity in isolated heterocysts of *Anabaena* cylindrica. *Biochim Biophys Acta, 423*(2), 189-195.

Terauchi, K., Kitayama, Y., Nishiwaki, T., Miwa, K., Murayama, Y., Oyama, T., & Kondo, T. (2007). ATPase activity of KaiC determines the basic timing for circadian clock of cyanobacteria. *Proc.Natl.Acad.Sci.U.S.A, 104*, 16377-16381. doi:0706292104 [pii];10.1073/pnas.0706292104 [doi]

Thiel, T., Lyons, E. M., Erker, J. C., & Ernst, A. (1995). A second nitrogenase in vegetative cells of a heterocyst-forming cyanobacterium. *Proc Natl Acad Sci U S A, 92*(20), 9358-9362.

Thomas, J. (1970). Absence of the pigments of photosystem II of photosynthesis in heterocysts of a blue-green alga. *Nature, 228*(5267), 181-183.

Thomas, J., Meeks, J. C., Wolk, C. P., Shaffer, P. W., & Austin, S. M. (1977). Formation of glutamine from [13n]ammonia, [13n]dinitrogen, and [14C]glutamate by heterocysts isolated from *Anabaena* cylindrica. *J Bacteriol, 129*(3), 1545-1555.

Tien, C. J., Sigee, D. C., & White, K. N. (2005). Characterization of surface sugars on algal cells with fluorescein isothiocyanate-conjugated lectins. *Protoplasma, 225*(3-4), 225-233. doi:10.1007/s00709-005-0092-8

Trivedi, U., Kaushik, S., Kunjadia, P., Saravanan, M., Nagaraja, V., Archana, G., & Nareshkumar, G. (2016). Functional expression and purification of *Anabaena* PCC 7120 XisA protein. *Protein Expr Purif, 118*, 64-69. doi:10.1016/j.pep.2015.09.027

Vaara, M. (1992). Eight bacterial proteins, including UDP-N-acetylglucosamine acyltransferase (LpxA) and three other transferases of Escherichia coli, consist of a six-residue periodicity theme. *FEMS Microbiol Lett, 76*(3), 249-254.

Valladares, A., Herrero, A., Pils, D., Schmetterer, G., & Flores, E. (2003). Cytochrome c oxidase genes required for nitrogenase activity and diazotrophic growth in *Anabaena* sp. PCC 7120. *Mol Microbiol, 47*(5), 1239-1249.

Vargas, W. A., Nishi, C. N., Giarrocco, L. E., & Salerno, G. L. (2011). Differential roles of alkaline/neutral invertases in Nostoc sp. PCC 7120: Inv-B isoform is essential for diazotrophic growth. *Planta, 233*(1), 153-162. doi:10.1007/s00425-010-1288-5

Vázquez-Bermúdez, M. F., Herrero, A., & Flores, E. (2000). Uptake of 2-Oxoglutarate inSynechococcus Strains Transformed with the Escherichia coli kgtP Gene. *J Bacteriol, 182*, 211-215.

Videau, P., Ni, S., Rivers, O. S., Ushijima, B., Feldmann, E. A., Cozy, L. M., . . . Callahan, S. M. (2014). Expanding the Direct HetR Regulon in *Anabaena* sp . Strain PCC 7120. *196*, 1113-1121. doi:10.1128/JB.01372-13

Videau, P., Rivers, O. S., Hurd, K., Ushijima, B., Oshiro, R. T., Ende, R. J., . . . Cozy, L. M. (2016). The heterocyst regulatory protein HetP and its homologs modulate heterocyst commitment in *Anabaena* sp. strain PCC 7120. *Proc Natl Acad Sci U S A.* doi:10.1073/pnas.1610533113

Vizcaino, J. A., Csordas, A., Del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., . . . Hermjakob, H. (2016). 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res, 44*(22), 11033. doi:10.1093/nar/gkw880

Voigt, K., Sharma, C. M., Mitschke, J., Lambrecht, S. J., Voss, B., Hess, W. R., & Steglich, C. (2014). Comparative transcriptomics of two environmentally relevant cyanobacteria reveals unexpected transcriptome diversity. *ISME J, 8*(10), 2056-2068. doi:10.1038/ismej.2014.57

Volesky, B., & Holan, Z. R. (1995). Biosorption of Heavy-Metals. *Biotechnology Progress, 11*(3), 235-250. doi:DOI 10.1021/bp00033a001

Walsby, A. E. (2007). Cyanobacterial heterocysts: terminal pores proposed as sites of gas exchange. *Trends Microbiol, 15*(8), 340-349. doi:10.1016/j.tim.2007.06.007

Wang, T., Shen, G., Balasubramanian, R., McIntosh, L., Bryant, D. A., & Golbeck, J. H. (2004). The sufR gene (sll0088 in Synechocystis sp. strain PCC 6803) functions as a repressor of the sufBCDS operon in iron-sulfur cluster biogenesis in cyanobacteria. *J Bacteriol, 186*(4), 956-967.

The Cyanobacteria—Isolation, Purification and Identification 1053-1073 (2006).

Welkie, D., Zhang, X., Markillie, M. L., Taylor, R., Orr, G., Jacobs, J., . . . Sherman, L. A. (2014). Transcriptomic and proteomic dynamics in the metabolism of a diazotrophic cyanobacterium, Cyanothece sp. PCC 7822 during a diurnal light-dark cycle. *BMC Genomics, 15*, 1185. doi:10.1186/1471-2164-15-1185

Wilde, A., Lunser, K., Ossenbuhl, F., Nickelsen, J., & Borner, T. (2001). Characterization of the cyanobacterial ycf37: mutation decreases the photosystem I content. *Biochem J, 357*(Pt 1), 211-216.

Wilde, A., & Mullineaux, C. W. (2015). Motility in cyanobacteria: polysaccharide tracks and Type IV pilus motors. *Mol Microbiol, 98*(6), 998-1001. doi:10.1111/mmi.13242

Wolk, C. P. (1966). Evidence of a role of heterocysts in the sporulation of a blue-green alga. *American Journal of Botany, 53*(3), 3.

Wolk, C. P. (1968). Movement of carbon from vegetative cells to heterocysts in *Anabaena* cylindrica. *J Bacteriol, 96*(6), 2138-2143.

Wolk, C. P. (1996). Heterocyst formation. *Annu Rev Genet, 30*, 59-78. doi:10.1146/annurev.genet.30.1.59

Heterocyst Metabolism and Development 769-823 (Springer Netherlands 1994).

Wolk, C. P., & Simon, R. D. (1969). Pigments and lipids of heterocysts. *Planta, 86*(1), 92-97. doi:10.1007/BF00385308

Wong, F. C. Y., & Meeks, J. C. (2002). Establishment of a functional symbiosis between the cyanobacterium Nostoc punctiforme and the bryophyte Anthoceros punctatus requires genes involved in nitrogen control and initiation of heterocyst differentiation. *Microbiology, 148*, 315-323.

Wu, J., Zhao, F., Wang, S., Deng, G., Wang, J., Bai, J., . . . Bao, Q. (2007). cTFbase: a database for comparative genomics of transcription factors in cyanobacteria. *BMC Genomics, 8*, 104. doi:10.1186/1471-2164-8-104

Xu, L., Wang, F., Xu, Y., Wang, Y., Zhang, C., Qin, X., . . . Yang, P. (2015). An MRM-based workflow for absolute quantitation of lysine-acetylated metabolic enzymes in mouse liver. *Analyst, 140*(23), 7868-7875. doi:10.1039/c5an01832c

Yamazaki, S., Nomata, J., & Fujita, Y. (2006). Differential operation of dual protochlorophyllide reductases for chlorophyll biosynthesis in response to environmental oxygen levels in the cyanobacterium Leptolyngbya boryana. *Plant Physiol*, *142*(3), 911-922. doi:10.1104/pp.106.086090

Yoon, H.-s., & Golden, J. W. (1998). Heterocyst Pattern Formation Controlled by a Diffusible Peptide. *282*, 935-939.

Yue, D., Peng, Y., Yin, Q., & Xiao, L. (2015). Proteomic analysis of Microcystis aeruginosa in response to nitrogen and phosphorus starvation. *Journal of Applied Phycology*, *27*, 1195-1204. doi:10.1007/s10811-014-0405-4

Arrangement of nitrogenase structural genes in an aerobic filamentous nonheterocystous cyanobacterium, 173 7055-7058 (1991).

Zhang, C.-c. (2000). Developmental Regulation of the Cell Division Protein FtsZ in *Anabaena* sp . Strain PCC 7120 , a Cyanobacterium Capable of Terminal Differentiation. *182*, 4640-4643.

Zhang, C. C., Laurent, S., Sakr, S., Peng, L., & Bedu, S. (2006). Heterocyst differentiation and pattern formation in cyanobacteria: a chorus of signals. *Mol Microbiol*, *59*(2), 367-375. doi:10.1111/j.1365-2958.2005.04979.x

Zhang, W., Du, Y., Khudyakov, I., Fan, Q., Gao, H., Ning, D., . . . Xu, X. (2007). A gene cluster that regulates both heterocyst differentiation and pattern formation in *Anabaena* sp. strain PCC 7120. *Mol Microbiol*, *66*(6), 1429-1443. doi:10.1111/j.1365-2958.2007.05997.x

Zhaxybayeva, O., Gogarten, J. P., Charlebois, R. L., Doolittle, W. F., & Papke, R. T. (2006). Phylogenetic analyses of cyanobacterial genomes : Quantification of horizontal gene transfer events. 1099-1108. doi:10.1101/gr.5322306.16

Zhou, R., Wei, X., Jiang, N., Li, H., Dong, Y., Hsi, K. L., & Zhao, J. (1998). Evidence that HetR protein is an unusual serine-type protease. *Proc Natl Acad Sci U S A, 95*(9), 4959-4963.

Zhou, R., & Wolk, C. P. (2002). Identification of an akinete marker gene in *Anabaena* variabilis. *J Bacteriol, 184*(9), 2529-2532.

Zhou, R., & Wolk, C. P. (2003). A two-component system mediates developmental regulation of biosynthesis of a heterocyst polysaccharide. *J Biol Chem, 278*(22), 19939-19946. doi:10.1074/jbc.M300577200

Ziegler, K., Diener, A., Herpin, C., Richter, R., Deutzmann, R., & Lockau, W. (1998). Molecular characterization of cyanophycin synthetase, the enzyme catalyzing the biosynthesis of the cyanobacterial reserve material multi-L-arginyl-poly-L-aspartate (cyanophycin). *Eur J Biochem, 254*(1), 154-159.

APPENDIX

Table 1: Distribution of FOX proteins, photosystem I and II proteins, and akinete marker protein in akinetes, heterocysts, and vegetative cells**.**

ORF[#]	Annotation	Akinete	Heterocyst	Veget. Cell
		Normalized quantitative value		
FOX genes				
2715	Histidine kinase HepN	0	0	0.45
4573	Neutral invertase InvB	0	0	0.45
4916	Hypothetical protein Npun_R1723, FraG/SepJ	0	0	0.45
6434	Protein serine/threonine phosphatase PrpJ1	0	0	0.45
222	Nitrogen-fixing NifU-like protein	0	0	0.9
1433	Response regulator receiver protein DevR	0	0	0.9
2881	ParB-like partition protein, HGL region containing	0	0	1.8
3957	Processing proteinase Abp2	0	0	2.71
3514	Polyketide-type polyunsaturated fatty acid synthase PfaA/HglF	0	2.45	0
5142	FHA domain containing protein FraH	0	2.45	0.45
4208	FHA modulated glycosyl transferase/transpeptidase PBP3	2.66	2.45	0.45

3826	Glucose-6-phosphate 1-dehydrogenase Zwl	7.99	2.45	1.35
5437	Nitrogenase molybdenum-iron protein alpha chain NifD	10.66	4.9	0
907	Nitrogenase molybdenum-iron cofactor biosynthesis protein NifN	0	7.36	0
3521	Glycosyl transferase (group 1) HglT	2.66	7.36	0
1865	Hypothetical protein Npun_F0815, Asp-, glu- rich product	0	7.36	5.86
4777	cytochrome c oxidase subunit II coxB2	0	7.36	0
1565	Histidinol dehydrogenase HisD	5.33	9.81	9.02
6242	ABC transporter related DevA	0	12.26	0
6485	Histone-like DNA-binding protein HanA	2.66	12.26	19.4
715	Polyketide synthase thioester reductase subunit HglB	0	14.71	0
6701	Fe-S cluster assembly protein NifU	0	22.07	0
3078	cytochrome c oxidase subunit II coxB3	10.66	22.07	0
911	Nitrogenase FeMo beta subunit protein NifK	10.66	22.07	0.45
153	Outer membrane efflux protein HgdD	23.98	22.07	4.06
3764	Hypothetical protein Npun_R5769 Abp3	37.3	34.33	11.73
3480	Putative transcriptional regulator (Crp/Fnr family) DevH	37.3	44.14	12.18

1651	Polysaccharide export protein	39.96	49.04	2.26
3002	Nitrogenase iron protein NifH	0	61.3	0.45
Photosystems				
3318	Photosystem I assembly BtpA	0	2.45	0.45
5296	Photosystem I assembly protein ycf3	0	0	1.35
1589	Photosystem I assembly Ycf4	2.66	4.9	4.06
4482	Photosystem I P700 apoprotein A2	50.62	17.17	12.18
5755	Photosystem I iron-sulfur center	18.65	26.97	18.04
2498	Photosystem I reaction center protein PsaF, subunit III	13.32	12.26	44.66
2501	Photosystem I reaction center subunit XI	45.29	122.6	59.09
1809	Photosystem I reaction center subunit IV	5.33	29.43	74.88
833	Photosystem I reaction center subunit II	39.96	100.54	163.3
1817	Photosystem II reaction center Psb28 protein	0	0	1.35
6255	Photosystem II oxygen evolving complex protein PsbP	5.33	4.9	4.96
3007	Photosystem II protein D2	13.32	9.81	4.96
83	Photosystem II reaction center protein H	0	2.45	4.96
2046	Photosystem q(b) protein	5.33	9.81	7.22
1649	Photosystem II 44 kDa subunit reaction center protein	47.95	22.07	10.83

5644	Photosystem II chlorophyll-binding protein CP47	47.95	29.43	14.89
3469	Photosystem II 11 kDa protein	21.31	22.07	17.59
3312	Photosystem II oxygen evolving complex protein PsbU	5.33	12.26	31.58
4725	Photosystem II manganese-stabilizing protein PsbO	42.62	71.11	104.66
755	Photochlorophyllide reductase subunit N	0	0	0.45
Akinete marker protein (AcaK43)				
1647	PRC-barrel domain-containing protein AvaK	85.25	90.73	1.35

Table 2: Proteins involved in biosynthesis of DNA, RNA, and protein, as well as cell division in akinetes, heterocysts, and vegetative cells respectively

ORF	**Annotation**	**Akinete**	**Heterocyst**	**Veget. Cell**
		Normalized quantitative value		
DNA replication and repair				
DNA gyrase, helicase, and topoisomerase				
4759	Single-strand binding protein	0	0	3.1577
3320	DEAD/DEAH box helicase domain-containing protein	0	2.45	4.51

5326	Peptidase U62, modulator of DNA gyrase	2.66	0	0
3692	DNA gyrase subunit A	2.66	0	0
2972	Protein splicing (intein) site	5.33	2.45	0.90
5282	DNA topoisomerase I	10.66	0	2.71
5834	Microcin-processing peptidase 2	13.32	2.45	1.35
5835	Peptidase U62 modulator of DNA gyrase	13.32	7.36	5.86
DNA polymerase				
257	Phage SPO1 DNA polymerase-related protein	0	0	0.45
3596	DNA polymerase III, delta subunit	0	0	0.45
Nucleotide biosynthesis				
6322	5-(carboxyamino)imidazole ribonucleotide synthase	0	0	0.45
3851	ATP phosphoribosyltransferase catalytic subunit	0	0	1.35
339	Dihydroorotase	0	2.45	0
4205	Adenine phosphoribosyltransferase	0	2.45	4.51
2020	Uracil phosphoribosyltransferase	0	2.45	2.26
2639	Phosphoribosylaminoimidazole-succinocarboxamide synthase	0	7.36	1.80
6314	Phosphoadenylylsulfate reductase (thioredoxin)	0	7.36	11.28
2240	Phosphoribosyltransferase	0	19.62	5.41

4194	Phosphoribosylformylglycinamidine synthase II	2.66	0	0
2553	Adenylosuccinate synthetase	2.66	0	2.26
3424	Phosphomethylpyrimidine kinase	2.66	0	0.90
6546	Dihydroorotase	2.66	2.45	0
2309	Orotate phosphoribosyltransferase	2.66	2.45	1.80
6261	Bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase	5.33	0	1.80
6333	Adenylosuccinate lyase	7.99	0	4.96
Transcriptional regulation				
RNA polymerase				
608	RNA polymerase, sigma 70 subunit, RpoD subfamily	0	0	0.45
1794	DNA-directed RNA polymerase subunit omega	0	0	3.61
2126	RNA polymerase, sigma subunit, RpsC/SigC	0	0	3.61
1436	RNA polymerase sigma factor	0	0	3.16
1465	RpoD family RNA polymerase sigma factor	0	0	0.45
1436	RNA polymerase sigma factor	0	0	3.16

1571	DNA-directed RNA polymerase gamma chain	5.33	0	4.96
2612	DNA-directed RNA polymerase subunit alpha	15.98	29.43	13.08
1572	DNA-directed RNA polymerase beta' subunit	37.30	2.45	3.16
1569	DNA-directed RNA polymerase subunit beta	39.96	2.45	3.61
Transcriptional regulator				
3282	SOS-response transcriptional repressor, LexA	0	0	0.45
2666	Transcriptional regulator, LysR family	0	0	9.92
5981	Transcriptional regulator, BolA protein family	0	0	1.35
3539	Signal recognition particle protein	0	0	4.511
4517	Two component transcriptional regulator	0	2.45	0.90
1692	Transcriptional regulator, GntR family	0	4.90	0.45
3023	Transcriptional regulator	0	4.90	0.90
3487	Transcriptional regulator, LysR family	0	4.90	3.61
1692	Transcriptional regulator, GntR family	0	4.90	0.45
4132	Anti-sigma-factor antagonist	0	7.36	1.80
6455	Response regulator receiver protein	0	12.26	2.26

1420	Two component LuxR family transcriptional regulator	5.33	4.90	3.16
4375	Putative transcriptional regulator	7.99	0	0.45
1173	Two component Transcriptional regulator, Winged helix family protein	10.66	29.43	5.86
3590	Two component LuxR family transcriptional regulator	26.64	14.71	10.83
557	AbrB family transcriptional regulator	39.96	31.88	29.32
Protein synthesis				
Amino acid synthesis				
524	Amidohydrolase 2	0	0	0.45
1031	Amidohydrolase	0	0	0.45
6700	Aromatic amino acid beta-eliminating lyase/threonine aldolase	0	4.9	0
985	Diaminopimelate epimerase	0	4.90	4.06
2185	Acetylglutamate kinase	0	9.81	0.90
2435	Aspartate-semialdehyde dehydrogenase	0	4.90	1.35
6314	Phosphoadenylylsulfate reductase (thioredoxin)	0	7.36	11.28
2446	Saccharopine dehydrogenase	7.99	0	0
3933	Taurine catabolism dioxygenase TauD/TfdA	23.98	0	0
3947	Amidohydrolase 2	31.97	0	0

tRNA synthetase				
260	Aspartyl/glutamyl-tRNA amidotransferase subunit A	0	0	4.06
5328	Cysteinyl-tRNA synthetase	0	0	0.45
39	Glutamyl-tRNA reductase	0	0	0.45
5530	Isoleucyl-tRNA synthetase	0	0	0.45
2962	Leucyl-tRNA synthetase	0	0	1.35
4316	Methionyl-tRNA synthetase	0	0	1.35
520	Putative histidyl-tRNA synthetase 2	0	0	6.32
1748	Seryl-tRNA synthetase	0	2.45	2.71
1102	Tryptophanyl-tRNA synthetase	0	2.45	0
3220	Tyrosyl-tRNA synthetase	0	2.45	1.35
81	Peptidyl-tRNA hydrolase	0	4.90	4.06
3740	Methionyl-tRNA formyltransferase	0	7.36	4.06
4380	Phenylalanyl-tRNA synthetase alpha subunit	0	14.71	3.16
1850	Glutamyl-tRNA(Gln) amidotransferase, B subunit	2.66	0	5.86
2270	Glycyl-tRNA synthetase alpha chain	2.66	0	0
6531	Histidyl-tRNA synthetase	2.66	7.36	2.71
6635	Arginyl-tRNA synthetase	5.33	0	4.51
2649	Glycyl and Arginyl tRNA synthetase	7.99	0	10.83
4909	Prolyl-tRNA synthetase	7.99	0	4.96

3273	Glutamyl-tRNA synthetase	7.99	2.45	1.80
3210	Lysyl-tRNA synthetase	10.66	0	3.61
495	Phenylalanyl-tRNA synthetase, beta subunit	10.66	0	0
3786	Valyl-tRNA synthetase	10.66	0	2.26
817	Threonyl-tRNA synthetase / Ser-tRNA(Thr) hydrolase	13.32	0	0.45
4277	Asparaginyl-tRNA synthetase	15.98	0	0
Ribosomal protein				
685	50S ribosomal protein L28	0	0	4.06
1006	Ribosomal protein S21	0	0	8.57
1567	Ribosomal protein S20	0	0	10.38
1713	30S ribosomal protein S18	0	0	4.06
1999	Ribosomal protein L27	0	0	4.51
2591	Ribosomal protein S19	0	0	13.98
2594	LSU ribosomal protein L16P	0	0	9.02
2595	Ribosomal protein L29	0	0	6.77
2598	Ribosomal protein L24	0	0	4.96
2603	LSU ribosomal protein L18P	0	0	3.61
2617	50S ribosomal protein L31	0	0	3.61
3791	30S ribosomal protein 3	0	0	2.71
4679	50S ribosomal protein L20	0	0	9.92
5266	Ribosomal protein S15	0	0	3.16

2554	50S ribosomal protein L25	0	2.45	4.96
2592	50S ribosomal protein L22	0	2.45	7.22
2596	30S ribosomal protein S17	0	2.45	4.51
4086	30S ribosomal protein S6	0	2.45	4.06
2613	50S ribosomal protein L17	0	4.90	10.38
2973	50S ribosomal protein L9	0	4.90	11.28
6002	Ribosomal protein L11 methyltransferase	0	4.90	0
2588	LSU ribosomal protein L4P	0	12.26	10.83
6310	Sigma 54 modulation protein/ribosomal protein S30EA	0	19.62	18.50
1549	50S ribosomal protein L7/L12	2.66	0	22.56
2590	50S ribosomal protein L2	2.66	2.45	10.38
2609	SSU ribosomal protein S13P	2.66	2.45	13.08
5950	30S ribosomal protein S12	2.66	2.45	3.16
1543	50S ribosomal protein L19	2.66	4.90	3.61
2615	LSU ribosomal protein L13P	2.66	7.36	10.83
2589	LSU ribosomal protein L23P	2.66	9.81	7.67
2601	50S ribosomal protein L6	2.66	9.81	7.67
5954	Ribosomal protein S10	2.66	12.26	11.73
1548	50S ribosomal protein L10	5.33	4.90	3.61
3538	30S ribosomal protein S16	5.33	4.90	12.18
5685	30S ribosomal protein S14	5.33	4.90	8.12
5951	SSU ribosomal protein S7P	5.33	4.90	20.75

2610	SSU ribosomal protein S11P	5.33	7.36	3.61
2616	30S ribosomal protein S9	5.33	7.36	8.12
2597	LSU ribosomal protein L14P	5.33	12.26	8.57
2586	LSU ribosomal protein L3P	5.33	19.62	7.67
4920	SSU ribosomal protein S4P	7.99	2.45	4.96
2605	50S ribosomal protein L15	10.66	14.71	17.14
1547	50S ribosomal protein L1	10.66	26.97	21.65
2600	30S ribosomal protein S8	13.32	9.81	8.12
2604	30S ribosomal protein S5	13.32	17.17	24.36
2599	50S ribosomal protein L5	13.32	22.07	21.65
1546	50S ribosomal protein L11	15.98	22.07	16.24
2593	SSU ribosomal protein S3P	18.65	14.71	9.47
4068	SSU ribosomal protein S2P	21.31	12.26	10.83
5643	30S ribosomal protein S1	26.64	26.97	4.96
Translation initiation factor				
2608	Bacterial translation initiation factor 1 (bIF-1)	0	0	5.86
5811	Translation initiation factor IF-3	0	0	5.41
3766	Translation initiation factor IF-2	13.32	0	2.71
5952	Translation elongation factor 2 (EF-2/EF-G)	50.62	7.36	12.63
5953	Translation elongation factor 1A (EF-1A/EF-Tu)	175.82	193.72	72.63

ORF[#]: open reading frame # were given by http://scorpius.ucdavis.edu/gmod/cgibin/site/*Anabaena*02?page=gblast

Table 3: Proteins involved in synthesizing and transporting polysaccharides and peptidoglycan in akinetes, heterocysts, and vegetative cells (low similarity of the S-layer domain containing proteins are highlighted in grey)

ORF	**Annotation**	**Akinete**	**Heterocyst**	**Veget. Cell**
		Normalized quantitative value		
Membrane transporter				
S-layer protein				
3288	S-layer domain-containing protein	5.33	2.45	3.61
5127	porin; major outer membrane protein	5.33	4.90	0
1127	S-layer domain-containing protein	37.30	0	0
1753	S-layer domain-containing protein	39.96	24.52	23.46
2713	porin; major outer membrane protein	50.62	90.73	5.86
2780	hypothetical protein all7614	66.60	110.34	0
1756	S-layer region-like	90.57	139.77	14.89
1758	hypothetical protein all4499	191.80	223.14	60.45
3008	hypothetical protein alr4550	359.63	328.58	41.50
ABC transporter				
4203	ABC-type transporter, integral membrane subunit	0	0	0.45

5640	ABC transporter-like	0	0	0.45
241	ABC-type transporter, periplasmic subunit family 3	0	2.45	0
3481	periplasmic phosphate-binding protein of phosphate ABC transporter	0	2.45	1.80
4968	ABC transporter related	0	2.45	0
767	ABC transporter-like	0	4.90	0
2712	ABC-type metal ion transporter, periplasmic subunit	0	4.90	0.45
3485	phosphate ABC-transporter periplasmic phosphate-binding protein	0	4.90	1.35
3524	ABC transporter, phosphate-binding protein	0	9.81	3.61
6242	ABC transporter related	0	12.26	0
6227	nitrate ABC transporter, ATPase subunits C and D	0	17.17	16.69
3042	molybdenum ABC transporter, periplasmic molybdate-binding protein	0	24.52	0
1130	ABC transporter related	2.66	0	0
6476	periplasmic sugar-binding protein of ABC transporter	5.33	17.17	11.28
2319	substrate-binding protein of ABC transporter	7.99	7.36	0
1135	ABC transporter, substrate-binding protein, aliphatic sulphonates	10.66	0	0

5097	ABC transporter ATP-binding protein	10.66	17.17	2.26
4212	phosphate ABC transporter, periplasmic phosphate-binding protein	13.32	22.07	15.34
1132	ABC transporter, substrate-binding protein, aliphatic sulphonates	26.64	0	0
Cell wall and secretion proteins				
Cell division				
668	Cell division transporter substrate-binding protein FtsY	0	0	0.45
2065	Septum formation topological specificity factor MinE	0	0	0.90
6254	Septum formation protein Maf	0	7.36	0
2066	septum site-determining protein MinD	2.66	4.90	1.80
4730	cell division protein FtsZ	10.66	12.26	3.61
Cell wall hydrolase/autolysin and secreted extracellular protein				
2682	secretion protein HlyD family protein	0	0	0.45
1424	secretion protein HlyD	0	0	0.90
4502	Type II secretion system F domain protein	0	0	2.26
4500	type II secretion system protein E	0	0	6.77
2516	cell wall hydrolase/autolysin	0	2.45	0.90
5393	general secretion pathway protein H	0	4.90	78.49
5458	cell wall hydrolase/autolysin	0	9.81	0
3516	Secretion protein HlyD	0	24.52	0

5545	secretion protein HlyD	2.66	7.36	2.26
6481	general secretion pathway protein D	7.99	0	7.22
6360	FG-GAP repeat-containing protein	7.99	58.85	1.35
4111	outer membrane secretion protein Alr0267	37.30	85.82	3.16
Extracellular biomolecules				
Glycolipid				
2637	UDP-3-0-acyl N-acetylglucosamine deacetylase	0	0	0.90
2668	hexapaptide repeat-containing transferase	0	2.45	0
715	polyketide synthase thioester reductase subunit HglB	0	14.713	0
3516	Secretion protein HlyD	0	24.52	0
4208	FHA modulated glycosyl transferase/transpeptidase	2.66	2.45	0.45
3521	glycosyl transferase, group 1	2.66	7.36	0
153	outer membrane efflux protein	23.98	22.07	4.06
2638	surface antigen (D15)	37.30	4.90	2.71
Peptidoglycan				
6325	peptidoglycan binding domain-containing protein	0	0	2.71
6200	N-acetylglucosamine 6-phosphate deacetylase	0	0	0.90

4163	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase	0	0	0.45
2637	UDP-3-0-acyl N-acetylglucosamine deacetylase	0	0	0.90
4461	N-acetylmuramic acid-6-phosphate etherase	0	0	0.45
5297	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase	0	2.45	0.90
4123	UDP-glucose/GDP-mannose dehydrogenase	0	2.45	9.47
1382	peptidoglycan binding domain-containing protein	0	4.90	11.28
5461	peptidoglycan binding domain-containing protein	0	4.90	2.71
715	polyketide synthase thioester reductase subunit HglB	0	14.71	0
5462	UDP-N-acetyl glucosamine-2-epimerase	2.66	2.45	1.35
2312	penicillin-binding protein, transpeptidase	2.66	4.90	1.80
2648	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase	5.33	0	0.90
5403	N-acetylmuramoyl-L-alanine amidase	5.33	0	0.45
3772	peptidoglycan binding domain-containing protein	5.33	14.71	10.38

2518	N-acetylmuramoyl-L-alanine amidase	10.66	14.71	0.45
1651	polysaccharide export protein	39.96	49.04	2.26

Table 5. The 172 Genes in five DNA elements in vegetative cell genome are deleted from heterocyst genome during heterocyst development

Gene No.	**start**	**end**	**Fuction notes**
***nifD* 74986 bp DNA element (2414668--2489653)**			
Anacy_2118	2414799	2416202	integrase family protein
Anacy_2119	2416248	2418260	primase P4
Anacy_2120	2418679	2419248	Resolvase domain protein
Anacy_2121	2419238	2420899	Integrase catalytic region
Anacy_2122	2420889	2421860	AAA ATPase
Anacy_2123	2422388	2422804	hypothetical protein
Anacy_2124	2422902	2423372	hypothetical protein
Anacy_2125	2423714	2424988	hypothetical protein
Anacy_2126	2425425	2427335	plasmid recombination protein
Anacy_2128	2427587	2427808	hypothetical protein
Anacy_2129	2427795	2428217	protein of unknown function DUF132
Anacy_2130	2428598	2428891	hypothetical protein
Anacy_2131	2428965	2429717	hypothetical protein
Anacy_2132	2429913	2430932	integrase family protein
Anacy_2133	2431099	2431497	hypothetical protein

Anacy_2138	2434809	2435492	transposase IS891/IS1136/IS1341 family
Anacy_2139	2435498	2436601	protein of unknown function DUF955
Anacy_2140	2437099	2437557	hypothetical protein
Anacy_2141	2437734	2438585	chromosome partitioning ATPase
Anacy_2142	2439349	2440320	AAA ATPase
Anacy_2143	2440310	2441971	Integrase catalytic region
Anacy_2144	2441961	2442530	Resolvase domain protein
Anacy_2145	2442923	2443345	integrase family protein
Anacy_2147	2443972	2444364	death-on-curing family protein
Anacy_2148	2444361	2444585	addiction module antidote
Anacy_2149	2444882	2446606	von Willebrand factor type D protein
Anacy_2150	2446845	2446970	hypothetical protein
Anacy_2151	2446970	2447260	hypothetical protein
Anacy_2152	2447461	2448189	hypothetical protein
Anacy_2153	2448186	2449142	protein of unknown function DUF1814
Anacy_2155	2451088	2451809	IS1 transposase
Anacy_2156	2455423	2455665	hypothetical protein
Anacy_2157	2455682	2455816	hypothetical protein
Anacy_2158	2456122	2456775	plasmid segregation oscillating ATPase ParF
Anacy_2159	2456762	2457031	hypothetical protein
Anacy_2160	2457157	2457429	hypothetical protein
Anacy_2161	2457433	2457774	hypothetical protein
Anacy_2162	2457913	2458995	Photosystem Q(B) protein

Anacy_2163	2459210	2459602	protein of unknown function DUF1311
Anacy_2164	2459977	2460486	hypothetical protein
Anacy_2165	2461139	2461936	hypothetical protein
Anacy_2166	2461985	2463262	transposase IS204/IS1001/IS1096/IS1165 family protein
Anacy_2167	2463930	2464526	hypothetical protein
Anacy_2168	2464788	2465228	transcriptional regulator, XRE family
Anacy_2169	2465462	2465764	fertility inhibition FinO-like protein
Anacy_2170	2466490	2469597	hypothetical protein
Anacy_2171	2470363	2470899	HNH endonuclease
Anacy_2172	2472087	2472863	HNH endonuclease
Anacy_2173	2474103	2474672	Resolvase domain protein
Anacy_2174	2474662	2476323	Integrase catalytic region
Anacy_R0025	2471106	2471180	tRNA-Asp
Anacy_R0026	2471266	2471341	tRNA-Cys
Anacy_R0027	2471449	2471519	tRNA-Trp
Anacy_R0028	2471525	2471601	tRNA-OTHER
Anacy_R0029	2471789	2471862	tRNA-Phe
Anacy_R0030	2471872	2471946	tRNA-Asn
Anacy_R0031	2472871	2472944	tRNA-Gln
Anacy_R0032	2473093	2473169	tRNA-Pro
Anacy_R0033	2473173	2473246	tRNA-Leu
Anacy_R0034	2473267	2473344	tRNA-Leu

Anacy_R0035	2473350	2473424	tRNA-Lys
Anacy_R0036	2473426	2473499	tRNA-Leu
Anacy_R0037	2473606	2473683	tRNA-OTHER
Anacy_2177	2477480	2478643	hypothetical protein
Anacy_2178	2478915	2479235	hypothetical protein
Anacy_2179	2479969	2480235	DNA-directed RNA polymerase omega subunit family protein-like protein
Anacy_2180	2480396	2480665	hypothetical protein
Anacy_2181	2480713	2480982	hypothetical protein
Anacy_2183	2481650	2482054	hypothetical protein
Anacy_2184	2482049	2482243	hypothetical protein
Anacy_2185	2482361	2482660	hypothetical protein
Anacy_2186	2482664	2482981	hypothetical protein
Anacy_2187	2482981	2483253	hypothetical protein
Anacy_2188	2483231	2483821	hypothetical protein
Anacy_2189	2483862	2484227	hypothetical protein
Anacy_2190	2484413	2485057	hypothetical protein
Anacy_2191	2485059	2485940	hypothetical protein
Anacy_2192	2485941	2488484	Collagen triple helix repeat-containing protein
Anacy_2193	2488481	2489029	hypothetical protein
Anacy_2194	2489029	2489343	hypothetical protein
***nifD* 15190 bp DNA element (2490117--2505306)**			
Anacy_2196	2490592	2491215	hypothetical protein

Anacy_2197	2491391	2492125	protein of unknown function DUF820
Anacy_2198	2492191	2492766	protein of unknown function DUF820
Anacy_2199	2492872	2493387	Peptidylprolyl isomerase
Anacy_2200	2493543	2494121	hypothetical protein
Anacy_2203	2495051	2496595	SPFH domain, Band 7 family protein
Anacy_2204	2497374	2497619	prevent-host-death family protein
Anacy_2205	2497616	2498020	PilT protein domain protein
Anacy_2206	2498084	2498566	hypothetical protein
Anacy_2207	2498646	2499248	hypothetical protein
Anacy_2208	2499261	2499485	hypothetical protein
Anacy_2209	2499485	2499904	hypothetical protein
Anacy_2210	2499897	2500121	hypothetical protein
Anacy_2211	2500236	2501465	hypothetical protein
Anacy_2212	2501827	2503104	transposase IS204/IS1001/IS1096/IS1165 family protein
Anacy_2213	2503101	2503667	hypothetical protein
Anacy_2214	2503768	2505069	phage integrase family protein, XisA
Anacy_R0038	2497013	2497088	tRNA-Lys
Anacy_R0039	2497088	2497163	tRNA-Ile
***hupL* 20842 bp DNA element (2030665--2051506)**			
Anacy_1762	2030785	2032245	integrase family protein
Anacy_1763	2032725	2033351	Curculin domain protein (mannose-binding) lectin

Anacy_1764	2033725	2033949	protein of unknown function DUF433
Anacy_1765	2033933	2034259	hypothetical protein
Anacy_1766	2034565	2034681	hypothetical protein
Anacy_1767	2034784	2035125	hypothetical protein
Anacy_1768	2035130	2035462	RNA-binding domain, S1
Anacy_1769	2035653	2035889	hypothetical protein
Anacy_1771	2036388	2036588	hypothetical protein
Anacy_1772	2036588	2036848	hypothetical protein
Anacy_1773	2037013	2037282	hypothetical protein
Anacy_1774	2037220	2037603	putative transcription regulator with HTH domain
Anacy_1775	2037578	2037886	Protein of unknown function DUF2136
Anacy_1776	2038225	2038542	hypothetical protein
Anacy_1777	2038539	2038886	transcriptional modulator of MazE/toxin, MazF
Anacy_1778	2039031	2039321	hypothetical protein
Anacy_1779	2039311	2039694	hypothetical protein
Anacy_1780	2039894	2040112	hypothetical protein
Anacy_1783	2041171	2043003	peptidase M12A astacin
Anacy_1784	2043073	2043261	hypothetical protein
Anacy_1785	2043402	2044133	uncharacterized protein involved in ubiquinone biosynthesis
Anacy_1786	2044522	2045193	regulatory protein TetR
Anacy_1787	2045395	2045910	hypothetical protein

Anacy_1788	2046109	2046966	hypothetical protein
Anacy_1789	2047091	2048641	AAA ATPase
Anacy_1790	2048885	2049412	hypothetical protein
Anacy_1791	2050113	2050637	protein of unknown function DUF1993-containing protein
Anacy_1792	2050652	2051062	hypothetical protein
***nifH1* 5736 bp DNA element (2407440--2413175)**			
Anacy_2109	2407838	2408236	hypothetical protein
Anacy_2110	2408381	2409601	transposase IS204/IS1001/IS1096/IS1165 family protein
Anacy_2112	2410215	2410493	hypothetical protein
Anacy_2113	2410520	2410843	hypothetical protein
Anacy_2114	2410836	2411255	hypothetical protein
Anacy_2115	2411404	2411646	hypothetical protein
Anacy_2116	2411639	2413066	phage integrase family protein, XisA
***pAnacy03* - 39998 bp DNA element (72642--112639)**			
Anacy_6026	73061	73630	Resolvase domain protein
Anacy_6027	73620	75281	Integrase catalytic region
Anacy_6028	75271	76242	AAA ATPase
Anacy_6029	76474	77523	hypothetical protein
Anacy_6030	77545	78186	Thymidylate kinase
Anacy_6031	78201	79454	hypothetical protein
Anacy_6032	79467	79904	hypothetical protein

Anacy_6033	80079	80360	hypothetical protein
Anacy_6034	80696	81499	hypothetical protein
Anacy_6035	81509	83506	pentapeptide repeat protein
Anacy_6037	84237	85154	hypothetical protein
Anacy_6038	85699	86550	integrase family protein
Anacy_6039	86690	87550	integrase family protein
Anacy_6040	87811	88122	hypothetical protein
Anacy_6041	88178	89593	hypothetical protein
Anacy_6042	89593	89772	hypothetical protein
Anacy_6043	89765	90358	hypothetical protein
Anacy_6044	90429	91103	hypothetical protein
Anacy_6045	91272	92120	hypothetical protein
Anacy_6046	92124	92549	hypothetical protein
Anacy_6047	92914	93090	hypothetical protein
Anacy_6048	93120	93314	hypothetical protein
Anacy_6049	93307	93549	hypothetical protein
Anacy_6050	94121	96385	hypothetical protein
Anacy_6051	96385	98103	Site-specific DNA-methyltransferase (adenine-specific)
Anacy_6052	98100	98360	hypothetical protein
Anacy_6053	98440	99018	protein of unknown function DUF820
Anacy_6054	99095	100378	restriction modification system DNA specificity domain protein

Anacy_6055	100382	100714	hypothetical protein
Anacy_6056	100711	100956	hypothetical protein
Anacy_6058	101187	104363	type I site-specific deoxyribonuclease, HsdR family
Anacy_6059	104528	104857	hypothetical protein
Anacy_6060	105114	105875	Cobyrinic acid ac-diamide synthase
Anacy_6061	105872	106243	hypothetical protein
Anacy_6062	106436	107083	integrase family protein
Anacy_6064	109257	109826	Resolvase domain protein
Anacy_6065	109816	111477	Integrase catalytic region
Anacy_6066	111467	112438	AAA ATPase

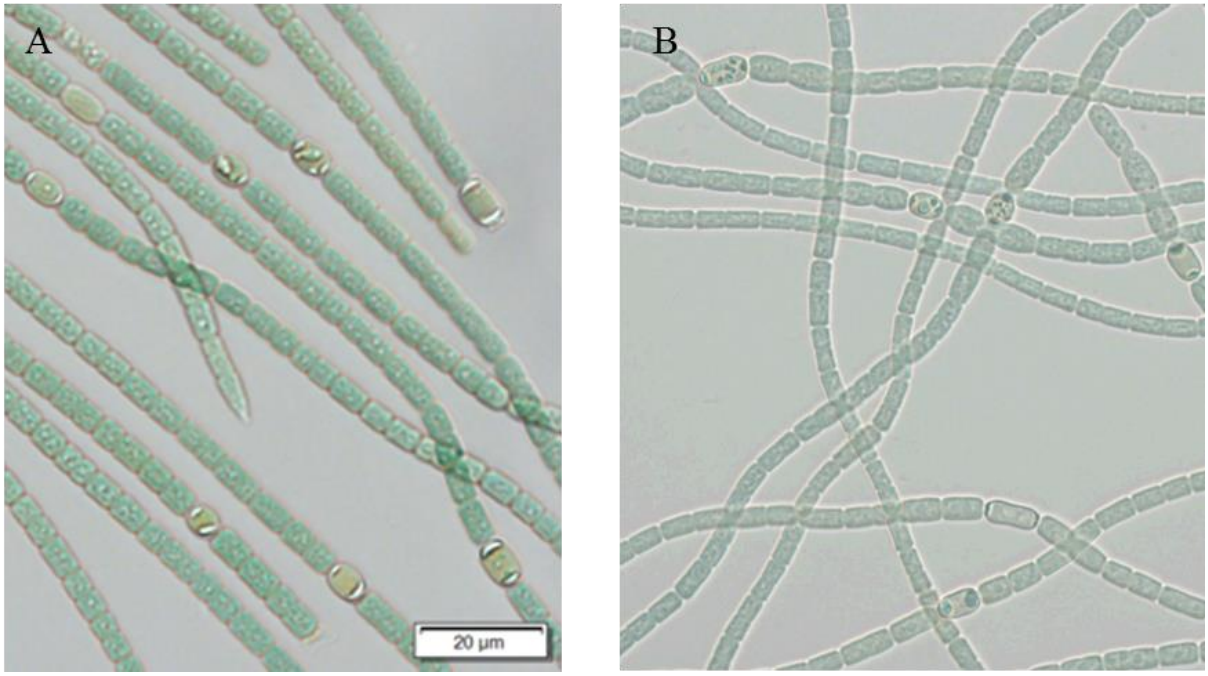

Fig. S1 Heterocysts are formed in both +N and –N cultures. A) Approximately 4.46% of *A. cylindrica* vegetative cells grown in AA/8 medium can form heterocysts. B)

Approximately 2.04% of *A. cylindrica* vegetative cells grown in AA/8N medium can form heterocysts.