2014

IDENTIFICATION OF MASTIGOCLADUS LAMINOSUS GENES ASSOCIATED WITH ENHANCED NITROGEN FIXATION PERFORMANCE

Patrick R. Hutchins
The University of Montana

IDENTIFICATION OF *MASTIGOCLADUS LAMINOSUS* GENES ASSOCIATED WITH ENHANCED NITROGEN FIXATION PERFORMANCE

By

PATRICK ROSS HUTCHINS

B.S. Marine Science, Coastal Carolina University, Conway, SC, 2010
M.S. Coastal Marine & Wetland Studies, Coastal Carolina University, Conway, SC, 2012

Thesis

presented in partial fulfillment of the requirements
for the degree of

Master of Science
Cellular, Molecular, and Microbial Biology

The University of Montana
Missoula, MT

June 2014

Approved by:

Sandy Ross, Dean of the Graduate School
Graduate School

Dr. Scott Miller, Committee Chair
Division of Biological Science

Dr. Frank Rosenzweig, Committee Member
Division of Biological Sciences

Dr. Cory Cleveland, Committee Member
College of Forestry and Conservation

ABSTRACT

Hutchins, Patrick, M.S., June 2014 Cellular, Molecular, and Microbial Biology

IDENTIFICATION OF *MASTIGOCLADUS LAMINOSUS* GENES ASSOCIATED WITH ENHANCED NITROGEN FIXATION PERFORMANCE

Chairperson: Dr. Scott Miller

Understanding population variation for fitness-related traits is important for our comprehension of evolutionary adaptation and of how populations respond to environmental change. Here, I investigate variation in nitrogen fixation performance for an ecologically-variable population of the cyanobacterium *Mastigocladus laminosus* from White Creek, a nitrogen-limited, geothermally-influenced stream in Yellowstone NP. I next take a population genomics approach to identify candidate loci associated with superior performance. Variation among strains and temperature dependence of the nitrogen fixation process were the most important factors in a linear mixed effects model. Absolute and relative measures of genetic differentiation between strains from the upper quartile of nitrogen fixation performance and the other 75% of strains showed that only a small subset of loci were associated with superior nitrogen fixation. Most notably, the strains that fixed the most nitrogen contained a premature stop codon in a regulatory histidine kinase gene, but this allele was present at low frequency in other strains. Because this nonsense mutation eliminates many important functional sites in the protein, this allele is expected to be non-functional. Both the full-length and the putative null allele, as well as a third recombinant allele, were expressed during nitrogen step-down and in the presence of combined nitrogen. Future studies will investigate whether the nonsense mutation results in transcriptional rewiring that is favorable for nitrogen fixation.

Key Words: cyanobacteria, nitrogen fixation, thermophile, fitness, adaptation, genomics

TABLE OF CONTENTS

ACKNOWLEDGMENTS

Jamie Brusa
Kayli Anderson
Dr. Emiko Sano
Mandy Slate
Angela Stathos
Dr. Elizabeth Crone
Dr. Frank Rosenzweig
Dr. Cory Cleveland
Dr. Scott Miller

LIST OF TABLES

Tables

Figures

Chapter 1

Abstract

Understanding population variation for fitness-related traits is important for our comprehension of evolutionary adaptation and of how populations respond to environmental change. At a nitrogen-limited, geothermally-influenced stream in Yellowstone National Park, the cyanobacterium *Mastigocladus laminosus* fixes abundant nitrogen *in situ*, an important fitness-related trait in nitrogen-limited systems. While extensive work has been done to identify the genes required to perform nitrogen fixation, little is known about the amount or genetic basis of phenotypic variation in nitrogen fixation performance in natural populations. Here, I use standard acetylene reduction assays to quantify the extent of phenotypic variation for nitrogen fixation ability among 23 randomly-selected White Creek *M. laminosus* strains. Variation among strains and temperature dependence of the nitrogen fixation process were the most important factors in a linear mixed effects model. Genome-wide analysis of the assayed strains was next used to identify candidate genes that may contribute to enhanced nitrogen fixation performance. Absolute and relative measures of genetic differentiation between strains from the upper quartile of nitrogen fixation performance and the other 75% of strains showed that only a small subset of loci were associated with superior nitrogen fixation. Most notably, strains that fixed the most nitrogen contained a premature stop codon in a regulatory histidine kinase gene, but this allele was present at low frequency in other strains. Because this nonsense mutation eliminates many important functional sites in the protein, this allele is expected to be non-functional. Expression and functional assays are needed to identify the mechanism through which this putative null allele may confer enhanced nitrogen fixation performance.

Introduction

Understanding population variation for fitness-related traits is important for our comprehension of evolutionary adaptation. As first pointed out by Darwin, heritable variation represents the raw material of evolution by natural selection. Forces that remove variation from a population, such as directional selection and genetic drift, are potentially counterbalanced by the input of mutations, gene flow, and balancing selection. Spatially-varying selection, for instance, is a form of balancing selection whereby spatial heterogeneity in the environment favors alternative genotypes (Hedrick, 2006). The extent of functional variation maintained in a population also has potential implications for both the resilience of ecosystem services in a changing environment (Hughes *et al.*, 1997; Luck *et al.*, 2003) as well as for how populations respond to temporal environmental change, because the rate at which beneficial mutations arise, and subsequently attain high frequencies, is slow in comparison to the speed at which populations can potentially adapt from standing genetic variation (Barrett, Schluter, 2008). Therefore, it is

vital that we understand the extent of standing phenotypic and genetic variation within populations and the mechanisms by which it is maintained.

At White Creek, a nitrogen-limited, geothermally-influenced stream in the Lower Geyser Basin of Yellowstone National Park, a population of the thermophilic, filamentous cyanobacterium, *Mastigocladus* (*Fischerella*) *laminosus*, exhibits tremendous ecological variation for temperature performance along a thermal gradient ranging from 39-54 ºC mean annual temperature (Miller *et al.*, 2009). This strong temperature gradient exists in the presence of little apparent spatial variation in nutrient and light availability (Miller *et al.*, 2009). More than 150 *M. laminosus* strains from five sampling locations spanning their natural range in White Creek have been archived and/or maintained in laboratory culture. *M. laminosus* strains from White Creek tend to grow better under laboratory conditions that mimic the mean temperatures from which they were originally collected, resulting in crossing reaction norms for temperature performance (Miller *et al.*, 2009). Although gene flow along White Creek is generally high throughout much of the genome, upstream and downstream strains of *M. laminosus* are genetically differentiated at specific genomic regions (Wall et al., in press; Miller et al., submitted). However, several other regions of the genome exhibit the signatures of balancing selection in the absence of obvious spatial structure. Questions remain regarding both the functional and the adaptive significance of this variation.

M. laminosus fixes abundant nitrogen *in situ* (Miller *et al.*, 2006), and it is expected that this is an important fitness-related trait in nitrogen-limited systems like White Creek. Biological nitrogen fixation is a globally significant biogeochemical process that many cyanobacteria perform. It is estimated that organisms that fix atmospheric nitrogen (diazotrophs) are responsible for more than half of global nitrogen fixation, in spite of increasing anthropogenic nitrogen fixation since the industrial era (Galloway *et al.*, 2004). Because cyanobacteria also perform oxygenic photosynthesis, they must perform two crucial metabolic processes that are at odds with one another. This is because the enzyme responsible for nitrogen fixation, nitrogenase,

contains a cofactor that is permanently deactivated by oxygen. In order to fix atmospheric nitrogen and perform oxygenic photosynthesis, cyanobacteria must separate these activities in either time or in space. Because photosynthetic oxygen production is light dependent, some cyanobacteria fix nitrogen under dark conditions, when their photosystems are naturally inactive (Berman-Frank *et al.*, 2001). An alternative strategy employed by *M. laminosus* and related cyanobacteria is to spatially separate these biochemically incompatible processes by means of specialized and terminally differentiated nitrogen-fixing cells called heterocysts.

Heterocysts, which are spaced at semi-regular intervals along filaments and typically account for ~5-10% of cells (Kumar *et al.*, 2010), have several important structural and functional differences from vegetative cells that enable nitrogen fixation to occur. Most importantly, the heterocyst creates a micro-oxic environment. The heterocyst's first defense against oxygen poisoning of nitrogenase is a physical barrier to oxygen diffusion in the form of an extracellular heterocyst envelope polysaccharide (HEP) and an underlying heterocyst glycolipid layer (HGL; Kumar *et al.*, 2010). Formation of the HEP layer is one of the earliest morphological changes during differentiation (Kumar *et al.*, 2010), and an intact HEP layer is required for heterocyst function in the presence of oxygen (Huang *et al.*, 2005; Wolk *et al.*, 1988), though it is generally believed that the HGL layer is the primary gas diffusion barrier (Currier *et al.*, 1977). Another measure taken during heterocyst development to enable nitrogenase activity under light conditions is the dismantling of the oxygen-producing photosystem (PS) II (Wolk *et al.*, 1994). An additional consequence of dismantling PSII is that the heterocyst is not able to generate reductant for carbon fixation (Wolk *et al.*, 1994). Consequently, fixed carbon in the form of sucrose is imported from adjacent vegetative cells to provide reducing power for nitrogen fixation (Kumar *et al.*, 2010). PS I, however, remains active, generating much of the ATP required for nitrogen fixation (Ernst *et al.*, 1983). Heterocysts also exhibit increased rates of respiration, the benefit of which is twofold: (1) intracellular oxygen is quickly consumed, which protects nitrogenase and (2) it provides a supplemental source of ATP that is used to power nitrogenase

(Wolk *et al.*, 1994). A return of combined nitrogen during the first 9-12 hours of heterocyst development can reverse the differentiation process , after which the cell is committed to differentiation (Yoon, Golden, 2001). Fixed nitrogen produced by heterocysts rapidly diffuses into adjacent vegetative cells (Popa *et al.*, 2007) via intracellular junctions (Mullineaux *et al.*, 2008) and/or a continuous periplasm (Flores *et al.*, 2006).

Nitrogen fixation has a complex genetic basis in heterocystous cyanobacteria (Wolk, 2000). In addition to nitrogen fixation (*nif*) genes that are common to most diazotrophs, genes involved in heterocyst differentiation are also required. Heterocyst differentiation has been extensively studied in the model cyanobacterium *Anabaena* PCC 7120 and is one of our best understood models of cell differentiation in bacteria (e.g., Kumar *et al.*, 2010). Nitrogen fixation is an energetically expensive process, and the heterocyst envelope is a significant investment, accounting for ~50% of cell dry weight (Dunn, Wolk, 1970); consequently, heterocysts are not produced when a preferred source of nitrogen is available in the environment. Within hours of nitrogen limitation, the master regulator of heterocyst differentiation, *hetR* (Buikema, Haselkorn, 2001), is limited to the semi-regularly spaced 5-10% of cells destined to become heterocysts (Huang *et al.*, 2004). The number of genes estimated to be differentially regulated during heterocyst development is staggering, ranging from just over 1000 in *Anabaena* PCC 7120 (Ehira *et al.*, 2003) to just under 500 in *Nostoc punctiforme* (Campbell *et al.*, 2007). These include between 100-140 "*Fox*" genes that are required for nitrogen fixation in the presence of oxygen (Wolk, 2000). For instance, the development of a heterocyst that is functional in an oxic environment requires the coordinated expression of genes which remodel the cell surface to provide a passive gas diffusion barrier that limits the entry of oxygen (Nicolaisen *et al.*, 2009; see above). Nitrogen fixation (*nif*) genes are expressed late in development, roughly 24 hours after nitrogen deprivation in *Anabaena* PCC 7120 (Ehira *et al.*, 2003). There are at least 18 genes and two excised DNA elements arranged in two separate gene clusters controlled by 4 operons in the *Anabaena* PCC 7120 *nif* regulon (reviewed in Böhm, 1998).

While extensive work has been done to identify the genes required to develop a heterocyst and to perform nitrogen fixation, very little is known about either the amount or the genetic basis of phenotypic variation for this important biogeochemical process in natural populations. Here, I first address the extent of phenotypic variation for nitrogen fixation ability among 23 randomly-selected White Creek *M. laminosus* strains. To assess whether nitrogen fixation co-varies with divergent temperature-specific growth in upstream and downstream sub-populations of *M. laminosus*, nitrogen fixation was tested at both of the approximate temperature extremes of their natural range in White Creek. Because nitrogen fixation requires a significant amount of ATP, the provision of fuel by carbon fixation is likely an important co-occurring process in *M. laminosus* under nitrogen limitation (Kumar *et al.*, 2010). Thus, simultaneous measurements of nitrogen- and carbon-fixation were performed to investigate the expectation that these two crucial metabolic processes are positively correlated in *M. laminosus*.

I next build on existing genomic resources available for White Creek *M. laminosus* strains to take a population genomics approach to identify loci associated with superior performance for nitrogen fixation and its temperature dependence. Population genomics approaches are powerful tools that use genome-wide sampling of population genetic variation to detect candidate genes which potentially contribute to population differentiation or phenotypic variation, as evidenced, for example, by outlier levels of genetic differentiation (reviewed by Luikart *et al.*, 2003 and Storz 2005). Although population genomics approaches have transformed the study of adaptation and genetic disease in both model and non-model eukaryotic systems, these methods have only recently been applied to bacteria (e.g. Thomas *et al.*, 2012 and Epstein *et al.*, 2012). Previous genome-wide analysis of genetic differentiation of *M. laminosus* along the White Creek temperature gradient has demonstrated that only a small fraction of White Creek *M. laminosus* loci are highly differentiated between upstream (>50°C) and downstream (<50°C) sites (Wall et al., submitted; Miller et al., submitted). My general approach was to group

strains for which genome data were available into phenotypic classes based on nitrogen fixation performance and use these classes in analyses of genetic differentiation to identify candidate genes that may contribute to high metabolic performance under contrasting temperature regimes. This study provides new insights into the genetic basis of a globally important and biochemically complex metabolic process and on the influence of environment on the maintenance of diversity.

Methods

Nitrogen and carbon fixation assays

Axenic *M. laminosus* filaments were transferred to 125 mL Erlenmeyer flasks with 75 mL of sterile D medium (Castenholz, 1988) and allowed to grow for at least two weeks. Once sufficient biomass accrued in D medium flasks, sub-samples were transferred to flasks with ND medium (D medium without combined nitrogen) to establish steady state growth in the absence of combined nitrogen, as in Miller *et al.* (2006) and Miller *et al.* (2009).White Creek *M. laminosus* strains were grown at the standard maintenance temperature of 50°C in ND medium at a light intensity of 105±5 µE m^{-2} s^{-1} provided by cool white fluorescent bulbs. After two weeks, cultures were split into six sub-lines, with three each of these moved to 37°C and 55°C growth chambers, respectively. Sub-lines were maintained in each incubator in ND medium and with a 12/12 hr light/dark cycle (105±5 µE m^{-2} s^{-1} during the light cycle) for two weeks leading up to the assay. Sub-lines were transferred on days seven and twelve during this acclimation period to ensure that cells were in exponential growth phase on the day of the assay (14 days after cultures were split into sub-lines). For each strain, nitrogen fixation incubation assays were performed two separate times using independent starting cultures.

Sub-samples from each sub-line were homogenized using a tissue grinder and normalized to an OD_{750} of 0.05±0.003. Cultures were homogenized such that large clumps of trichomes (i.e. chains of cells) were broken up, but long chains containing vegetative cells and heterocysts remained intact. Relative nitrogen fixation rates were estimated by the standard acetylene

reduction assay (ARA; Stewart *et al.*, 1967). Because the production of ethylene is proportional to the activity of the nitrogenase enzyme, "nitrogen fixation performance" will be used interchangeably with "normalized ethylene production" throughout this manuscript. Assays were carried out with 10 mL of ND medium in 20 mL crimp-sealed vials with a light and a dark replicate for each sub-line. Samples were incubated for four hours following the addition of 5 mL of acetylene gas (generated by the addition of 5 g of calcium carbide to 100 mL of deionized water) at the beginning of the light cycle of the established light regime. Incubations were terminated by aspirating as much sample headspace as possible (~15 mL) from each incubation vial and injecting it into a pre-evacuated 5 mL crimp vial. Ethylene production was measured using flame-ionization detection gas chromatography (FID-GC) with a Shimadzu GC-2014. Ethylene production measurements were estimated using a standard curve, blank corrected against parallel incubation vials that contained only ND growth medium and normalized to an optical density of 0.050. Optical density was empirically determined to have a linear relationship with cell dry mass for *M. laminosus* samples (Pearson correlation, $R^2 = 0.95$, $p < 0.001$). Microscopic counts of heterocyst frequency were performed for one representative sub-line at each temperature treatment. This was done to ensure that any variability between strains in their ability to form heterocysts was taken into account during data analysis.

Concurrent estimations of carbon fixation by each sub-line were made using ^{14}C-bicarbonate incorporation rates (see Miller et al. 1998). Briefly, incubations were initiated with the addition of 0.2 µCi of ^{14}C-bicarbonate to 3 mL aliquots of each sub-line, carried out for one hour under the same light and temperature conditions as in the acetylene reduction assay above and then terminated with the addition of 200 µL of formalin. To correct for non-biological uptake of radiolabeled carbon, formalin was added to a duplicate aliquot of one of the three sub-lines at each temperature treatment at the start of the incubation. The full 3 mL sample volume was filtered onto a 0.45 µm GN-6 membrane filter (PALL Life Sciences), rinsed first with 3% HCl to remove unincorporated radioisotope, and then rinsed with deionized water. Filters were

then placed into 20 mL scintillation vials and allowed to ventilate in a fume hood for at least one hour before adding 1.5 mL of EcoLite scintillation fluid (ICN). Samples were then read by a Beckman LS6000SE scintillation counter. As with ethylene production, carbon fixation rates were normalized to an optical density of 0.050.

Statistical Analysis

Because of the crossed experimental design and the heteroskedastic nature of the data, even after transformation, I generated a linear mixed effects model using the R "lme4" package (Bates *et al.*, 2014) to understand which factors explained the variation in observed nitrogen fixation performance. Fixed factors of the model were (1) normalized carbon fixation rate, (2) temperature treatment, (3) heterocyst frequency, and all possible interactions. The random effects structure was designed such that the model accounted for variation within incubations and among strains across the two temperature treatments. Other variables in the model were removed via backwards stepwise nested hypothesis testing using the F-test until the lowest Akaike information criterion score was obtained. A post-hoc pseudo-R^2 for linear mixed models (Nakagawa, Schielzeth, 2013) was used to approximate the fit of the model and estimate the amount of variation that could be explained by fixed factors and individual random effects.

Identification of Candidate Genes

Results from the ARA's were used to categorize *M. laminosus* strains into phenotypic classes based on normalized ethylene production within each temperature treatment (37 and 55 ºC datasets) and overall pooled performance (pooled dataset). For each dataset, strains in the upper quartile of mean normalized ethylene production were binned as the "upper" phenotypic class and those below this benchmark categorized as the "lower" phenotypic class. Genomic data for *M. laminosus* strains used in the analysis were obtained previously (Miller *et al.*, submitted). Briefly, paired-end Illumina sequence data were obtained for 20 White Creek strains randomly-selected from the lab strain collection. Draft genomes were assembled *de novo* using Velvet (Zerbino, Birney, 2008). Contigs in these draft genomes were auto-annotated with the RAST server and

saved in GenBank format (Aziz *et al.*, 2008). For each genome, protein-coding genes (CDS) were extracted from the GenBank files with custom Perl scripts to create FASTA-formatted files of all CDS.

Sequential local BLASTn queries of a non-redundant database of CDS were used to build separate FASTA-formatted files of orthologous CDS corresponding to each locus for the two phenotypic classes described above. Only full-length CDS were included, and loci for which fewer than 10 sequences were available were excluded. Custom Perl scripts (Miller *et al.*, submitted) were then used to estimate genome-wide relative (F_{ST}) and absolute (D_{XY}) genetic differentiation of polymorphic loci between phenotypic classes. Though F_{ST} has historically been used to estimate the relative genetic variation between geographically distinct populations (Holsinger, Weir, 2009), F_{ST} may be applied to any pair of defined groups. Here, the groups of interest are based on phenotypic classes rather than geographic location. The resulting distributions of F_{ST} and D_{XY}, respectively, were taken as empirical null distributions for each dataset to avoid assumptions regarding demographic history used by model-based approaches for identifying candidate loci. Vetted outlier loci (top ~1% of the tail, 20 CDS) of both F_{ST} and D_{XY} distributions were further explored by comparing them to available annotated orthologs in the NCBI database.

Results & Discussion

Nitrogen Fixation Activity

Strain means for normalized ethylene production values in the pooled dataset used to develop the model spanned a large range, from 0.35 PPM hr^{-1} in WC434 to 12.36 PPM in WC245 (Table 1.1). There was very little variation in normalized ethylene production in dark treatments and these rates were, on average, 28% of normalized ethylene production in respective light conditions, which is consistent with other studies on various diazotrophs that report light-independent nitrogenase activities at less than half of those under saturating light (e.g. Liengen,

1999; Staal *et al.*, 2001; Fig. 1.1B). Only data from the light treatments were included in the model and in subsequent analyses.

The final linear mixed model used to estimate normalized ethylene production from the pooled dataset (Table 1.2) was a random slope model that included two fixed factors (heterocyst frequency and a carbon fixation by temperature interaction) and two random effects (incubation, and a strain by temperature interaction). The pseudo-R^2 ($R^2_{(c)}$) for this model was 0.74 with fixed factors explaining 21% of the variation ($R^2_{(m)}$), and random effects explaining the remaining 53%. Of the random effects, the strain of *M. laminosus* assayed accounted for 19% of the model variance, temperature accounted for 20% and incubation for 3% (the remaining 11% of variation is the residual for random effects). The strain, and thereby the genomic background, proved to be very influential in determining overall normalized ethylene production, accounting for more than one fourth of the total variation explained by the model.

Heterocyst frequencies were on average 2.4 ± 0.3% (error based on 95% confidence interval). This is lower than the ~5-10% frequency that is typically reported for model heterocyst forming cyanobacteria but comparable to previous results obtained in the lab for *M. laminosus* under these conditions (unpublished data). The correlation between heterocyst frequency and normalized ethylene production was only significant in the 55 ºC dataset (Pearson correlation, R = 0.50, p <0.01), but was weakly positive in the 37 ºC and pooled datasets (R = 0.17 and 0.30, respectively). Mean strain-specific normalized carbon fixation rates for the pooled dataset ranged between 33 and 133 µg C L^{-1} hr^{-1} (data not shown). Normalized carbon fixation rates were generally higher in the 37 ºC dataset than the 55 ºC (averages of 107 ±10 and 33 ±3 µg C L^{-1} hr^{-1}, respectively). There was a highly significantly positive correlation between normalized carbon fixation and normalized ethylene production in the 55 ºC (Pearson correlation, R = 0.73, p <0.01) and pooled datasets (Pearson correlation, R = 0.59, p <0.01). The relationship between normalized carbon fixation and normalized ethylene production in the 37 ºC dataset was positive, but not significant (Pearson correlation, R = 0.50, p >0.05). While the fixed factors described

above cumulatively explained a moderate amount of variation in the model (21%, Table 1.2), strain and temperature effects explained approximately twice as much model variation.

Normalized ethylene production varied widely across temperature treatments and strains (Fig. 1.1, Table 1.1). Out of the 23 strains assayed, 9 had reaction norm slopes that were significantly different from zero in the light treatments (Fig. 1.1A; t-test, $p < 0.05$) and 8 in the dark treatments (Fig. 1.1B). Strains with non-zero slopes usually performed better at the lower temperature than at the higher temperature. This finding is corroborated by field $^{15}N_2$ –uptake experiments with White Creek *M. laminosus* performed by Stewart (1970) and by acetylene reduction assays performed in the field at White Creek (Hutchins and Miller, unpublished). Just one strain (WC344) had higher average nitrogenase activity at 55 ºC than at 37 ºC. There was no correlation between the temperature at which each strain was isolated from White Creek and normalized ethylene production in either temperature-specific or pooled datasets (Pearson correlation, $p > 0.05$; data not shown). The intrinsic temperature dependence of nitrogen fixation performance in the strains assayed here therefore does not appear to be tightly coupled to the location of strain origin along the thermal gradient. In the pooled dataset, strains in the upper phenotypic class were, not surprisingly, often also those that were in the upper class for temperature-specific normalized ethylene production (Table 1.1). The upper classes for 55 ºC and the pooled dataset shared more common strains with each other than either did with those of the 37 ºC group. Strains WC119, WC245, and WC439 were in the upper class for both of the temperature-specific and the pooled datasets.

Genome-wide Analysis of Loci Associated with Nitrogen Fixation

Genomic data was available for five out of the six strains in the upper class (top quartile of pooled normalized ethylene production; WC119, WC1110, WC245, WC344, and WC439) and for 11 out of the 17 remaining (lower class) strains. The majority of the *M. laminosus* genome exhibited very little differentiation between phenotypic classes for all three datasets but contained distinct outliers in the tails of the distributions (F_{ST} and D_{XY} near zero; Fig. 1.2 and 1.3,

respectively). Candidate loci exhibited the greatest genetic differentiation between classes in the 55 ºC dataset. Nearly half of the candidate genes identified in the results did not have a homolog in the NCBI database with a known function (Fig. 1.4 and 1.5, Tables 1.3 – 1.8). However, for those that did have an identifiable function, the vast majority were involved with carbohydrate, amino acid, or inorganic ion transport/metabolism. The small peaks in the frequencies of F_{ST} values centered on 0.30 – 0.35 for all three datasets are the result of the genetic clustering of a few strains from the lower class (WC1110, WC527, WC538, and WC441) for a sub-set of loci that are not associated with enhanced nitrogen fixation.

At 37 ºC, the most represented cluster of orthologous groups (COG) category among candidates were those with unknown function or general prediction only (Fig. 1.4A and 1.5A, respectively). Those with identifiable functions were most commonly involved with inorganic or amino acid transport/metabolism or cell membrane biogenesis. However, there were several noteworthy candidate genes in the ~1% tail of outlier loci. One of the genes that appeared in tails of both the F_{ST} and D_{XY} distributions was candidate 19-42545. It is annotated as a diguanylate cyclase, an enzyme which is observed in diverse branches of the prokaryotic tree (Galperin, 2004). Diguanylate cyclases catalyze the formation of 3′–5′ cyclic diguanylic acid (c-di-GMP), a secondary messenger protein involved in numerous networks (Hengge, 2009) that leads to the biosynthesis of adhesins and exopolysacharides associated with bacterial biofilm formation (Jenal, 2004). There are two nonsynonymous polymorphisms at this locus: all strains in the upper phenotypic class had a serine rather than an alanine at residue 34 and an aspartic acid instead of an asparagine at residue 42 (the allele fixed in the upper class was present at 42% frequency in the lower class). The highest F_{ST} value belonged to a potassium channel protein gene (candidate 56-42545) orthologous to alr0440 in *Anabaena* PCC 7120. This gene is upregulated during nitrogen step-down and heterocyst development (Ehira *et al.*, 2003), but differentiation between phenotypic classes was manifested by two synonymous polymorphisms, and its role in nitrogen fixation performance is not known (the allele fixed in the upper class was present in 40% of

strains in the lower class). Candidate 131-35450 is annotated as *hupW*, a protease that is involved in the maturation of the uptake hydrogenase (Wang *et al.*, 2012) and is upregulated during heterocyst development (Ehira *et al.*, 2003). The uptake hydrogenase recycles the hydrogen byproduct generated by nitrogenase, providing additional electrons that are used for nitrogen-reduction during fixation (Lindberg *et al.*, 2012). Inactivation of *hupW* results in a malfunctioning uptake hydrogenase and the evolution of excess hydrogen atoms in heterocysts (Lindberg *et al.*, 2012), thus decreasing the reducing power available to the heterocyst (Carrasco *et al.*, 2005). All of the upper class strains at 37 ºC were characterized by a methionine at residue 28, rather than an isoleucine, in the nickel binding site (the allele fixed in the upper class was present at 62% frequency in the lower class). Functional analyses of proteins from these candidate loci are needed to elucidate their effects on nitrogen fixation and fitness in *M. laminosus*.

Though the number of candidate genes encoding proteins with either unknown function or having only a general prediction was also high in the 55 ºC dataset, a large proportion of the genes encoded proteins that are involved with carbohydrate transport and metabolism (Fig. 1.4B and 1.5B for F_{ST} and D_{XY}, respectively). The locus with the greatest F_{ST} value was candidate 28-39736, an adenylylsulfate (APS) kinase. These phosphotransferases catalyze the second reaction of the conversion of inorganic sulfate to 3'-phosphoadenosine 5'-phosphosulfate as part of assimilatory sulfur metabolisms (Renosto *et al.*, 1984). A single nonsynonymous polymorphism between the two classes was present at nucleotide 241, resulting in an aspartic acid in the upper class while the majority of lower class strains contain an asparagine at this position (the allele fixed in the upper class was present at 17% frequency in the lower class). Candidate 1-33964 encodes the hopene-associated glycosyltransferase, *hpnB*. Glycosyltransferases which contain family 2 domains, as is the case with *hpnB*, are generally responsible for transferring nucleotide-diphosphate sugars to polysaccharide and lipid substrates (Perzl *et al.*, 1998). *hpnB* (alr0776) is one of several genes related to heterocyst development that is upregulated by NaCl in the

cyanobacterium, *Anabaena* sp. PCC 7120 (Imashimizu *et al.*, 2005). All of the strains in the upper phenotypic class possessed a nonsynonymous polymorphism that translates to an alanine instead of a glycine at residue 271 (the allele fixed in the upper class was present in 27% of strains in the lower class). The third F_{ST} outlier, candidate 49-34361, is annotated as a cation-transporting ATPase and shows weak homology to all3375 in the *Anabaena* PCC 7120 genome (Kaneko *et al.*, 2001). A nonsynonymous polymorphism between the phenotypic classes resulted in a proline in the upper class and a leucine in the lower class at amino acid 172 (the allele fixed in the upper class was present at 33% frequency in the lower class). Candidate 65-42545 encodes the third subunit of cytochrome oxidase that is most similar to the homologous gene located in the *coxBACI* operon in *Anabaena* PCC 7120. While this gene is mildly upregulated during heterocyst development, it does not appear to be the primary cytochrome oxidase responsible for enhanced respiratory activity within the heterocyst (Jones, Haselkorn, 2002). Furthermore, the interaction between cytochrome-c and cytochrome oxidase occurs on subunits I and II, whereas the third subunit is not involved catalytically (Witt *et al.*, 1998). The differentiation between classes manifested as a synonymous adenine instead of guanine at nucleotide 435 in the gene (the allele fixed in the upper class was present in 36% of strains in the lower class). Still, consumption of intracellular oxygen is crucial for heterocysts, and genetic changes in this gene may potentially contribute to variation in nitrogen fixation among strains of *M. laminosus*. Candidate 93-42545 is homologous to alr4726 in the *Anabaena* PCC 7120 genome, which encodes a protein that has been identified as belonging to the zinc uptake regulator family of sensory kinases. Two synonymous polymorphisms are observed between phenotypic classes: in strains from the upper class a cytosine is present instead of a thymine at nucleotide 252 and thymine rather than a cytosine at 261 (the allele fixed in the upper class was present at 36% frequency in the lower class).

Though in many cases, the polymorphisms that distinguish phenotypic classes at the above candidate loci are synonymous, it does not necessarily mean that these loci are

374 unimportant. While selection may act on codon usage, an alternative possibility is that the actual
375 target of selection is in non-coding DNA that is physically linked to the candidate locus. The
376 analysis screened protein coding regions of the *M. laminosus* genome, but polymorphisms in non-
377 coding regulatory regions that are adjacent to genes may potentially influence nitrogen fixation
378 performance. A sliding window approach along genome contigs could be used to find adjacent,
379 physically linked non-coding regions of the *M. laminosus* genome that may also be differentiated
380 between phenotypic classes and which may be the true target of selection.

381 Within the pooled dataset, the most common functionally identifiable COGs among the
382 F_{ST} candidates were carbohydrate transport/metabolism, amino acid transport/metabolism, and
383 signal transduction proteins. A histidine kinase gene, candidate 167-28586, was an outlier in both
384 the F_{ST} and D_{XY} distributions and had the highest values for each respective metric of any gene
385 for the pooled dataset. Unlike other candidates identified thus far, there appear to be three
386 segregating alleles at this locus: an allele with a nonsense mutation that eliminates 150
387 nucleotides at the 3' end of the gene (fixed in upper class, 36% frequency in lower class); a full-
388 copy allele that differs from the above allele at 39 nucleotide positions, one of which includes the
389 site of the alternative nonsense polymorphism; and an apparently rare (N = 1 in our sample)
390 recombinant allele that is identical to the latter at the 5' end and to the former at the 3' end and
391 therefore contains the nonsense mutation (see 'A Histidine Kinase Candidate' below for more
392 discussion of candidate 167-28586). Candidate 20-24813 is annotated as an enzyme in the
393 cytochrome P450 family. P450s are heme-thiolate proteins that oxidize and degrade a diverse
394 array of substrates and have equally diverse structures throughout all three domains of life
395 (Werck-Reichhart, Feyereisen, 2000). Strains in the upper phenotypic class have a synonymous
396 polymorphism at nucleotide 1317 in the form of a thymine (27% frequency in lower class), rather
397 than a cytosine, as is the case with most strains in the lower class. Candidate 29-33117 encodes
398 the cytochrome-c550 component of PS II. Cytochrome-c-550 is a membrane bound component
399 of the cyanobacterial PS II oxygen-evolving complex (OEC; responsible for the water-splitting

reaction that produces oxygen and provides reducing power for carbon fixation) and is responsible for stabilizing chlorine and calcium-binding to the complex (Roncel *et al.*, 2012). Differentiation between the phenotypic classes was characterized by a single synonymous polymorphism: cytosine in the upper class and thymine in the lower class at nucleotide 471 (the allele fixed in the upper class was present at 50% frequency in the lower class). Candidate 10-32834 was in the upper 1% of both F_{ST} and D_{XY} distributions and is annotated as the nickel-binding alpha subunit of urease, an enzyme that catalyzes the hydrolysis of urea into ammonia and carbon dioxide (Holm, Sander, 1997). There were four nucleotide polymorphisms in the upper class, two of which are adjacent and are nonsynonymous. Strains in the upper class have an alanine at residue 555 (55% frequency in lower class) while some strains in the lower class have an allele with an asparagine at this location. The polymorphism described here does not occur in either a metal binding site or the active site of the protein, though it is possible that it confers a structural modification. Urease is directly involved in assimilatory nitrogen metabolism and the recycling of urea generated by cell metabolism. Chemical analysis of water samples taken along White Creek does suggest that there are occasional pulses of dissolved organic nitrogen in the system (Hutchins, unpublished). Urease's involvement in nitrogen metabolism and the differentiation in the alpha subunit gene between phenotypic classes presented here make this locus an interesting prospect for future investigations.

A Histidine Kinase Candidate

The 167-28586 locus, which appears as a candidate in the pooled dataset and putatively encodes a histidine kinase, is particularly noteworthy. In addition to being an extreme outlier by both metrics of genetic differentiation, 167-28586 also exhibits the molecular evolutionary signatures of long-term balancing selection (Miller et al., submitted). These include an extremely positively skewed value of Tajima's *D* and an excess of polymorphism in the White Creek population. Histidine kinases (HKs) are involved in two-component signal transduction systems (TCSs), the principal means by which bacteria sense and respond to environmental changes (Gao,

Stock, 2009; Wuichet *et al.*, 2010). Prototypical TCSs involve two distinct proteins. A histidine kinase (HK), which usually has a sensory domain that interacts with the intra- or extracellular environment, serves as the input component of the system. The HK then transfers phosphoryl groups to a cognate response regulator (RR) to effect a change in gene expression or, sometimes, protein activity (Galperin, 2010). Once stimulated, a well-conserved ATP binding domain at the C-terminal end of the HK catalyzes the autophosphorylation of a conserved histidine residue. The phosphorylated HK then transfers the His-bound phosphoryl group to an asparagine residue in a highly-conserved receiver domain on the RR.

167-28586 exhibits ~50% amino acid identity with three histidine kinases in the *Anabaena* PCC 7120 genome (alr1551, alr2739, and alr4882). For several reasons, alr4882 appears to be the ortholog in the *Anabaena* PCC 7120 genome. 67-28586 and alr4882 share the same length and domain structure (both lack a sensory domain) as alr4882, which is not the case with other *Anabaena* homologs. Also, local gene order in the region is conserved: both HK167-28586 and alr4882 are downstream of a putatively orthologous annotated gene encoding a protein with a FIST sensory domain that likely serves as the sensory component of this network (alr4881 and the corresponding *M. laminosus* gene are ~56% identical at the amino acid level). While FIST domains are phylogenetically widespread, they are biochemically uncharacterized, though they are predicted to bind small molecules (Borziak, Zhulin, 2007).

It is likely that the loss of more than half of the ATP-binding pocket would render HK167-28586 nonfunctional, even if it were expressed. I propose that this would result in a transcriptional "rewiring" that is somehow favorable with respect to nitrogen fixation. Loss of function mutations that alter regulatory networks may be a common mechanism of bacterial adaptation to environmental change (Hottes *et al.*, 2013). "Gain-of-function" mutations require very specific alterations to genes and are rare compared to "loss-of-function" mutations, which can be explored rapidly in a large evolving population (Hottes *et al.*, 2013). However, in order for HK167-28586 to contribute to the observed differences in performance between phenotypic

classes, I expect that the full copy of the allele must, at the very least, be expressed during either nitrogen step-down and/or steady-state growth under nitrogen-limitation (see Chapter 2).

Interestingly, HK167-28586 is located just upstream of what appears to be an orthologous gene to alr0677 in the *Anabaena* PCC 7120 genome. This gene exhibits homology with a site-specific recombinase, XisC, which is required for the excision of ~10.5 kilobases from the *hupL* gene in *Anabaena* PCC 7120, which encodes the large subunit of the uptake hydrogenase (Carrasco *et al.*, 2005). Excision of the *hupL* element is necessary to produce a functional heterocyst-specific [NiFe] uptake hydrogenase, which catalyzes the consumption of hydrogen that is produced as a byproduct of nitrogen fixation (Tamagnini *et al.*, 2002). Two other site-specific recombinases are also required for nitrogen fixation in *Anabaena* PCC 7120 (Böhm, 1998). In order to produce a functional nitrogenase enzyme, two inserted elements in the *nif* operon must be removed: a 55 kilobase element from *fdxN*, which encodes a ferredoxin-like protein, and an 11 kilobase element from *nifD*, which encodes the alpha subunit of the nitrogenase MoFe protein. These elements contain genes encoding the site-specific recombinases required for their own excision: *xisF* and *xisA*, respectively. Though the proximity of Hk167-28586 to a putative site-specific recombinase is intriguing, at this time it is not possible to say what, if any, relationship exists between the two genes and nitrogen fixation performance in *M. laminosus*.

Conclusion

Differences among strains explained a considerable portion of the variation in nitrogen fixation in a mixed effects model. The comparatively low number of loci that were strongly associated with phenotypic variation among strains in nitrogen fixation performance suggests that dissecting the contributions of these genetic factors to variation in this complex trait may be tractable. However, the signatures of selection that we observe in our genome data may be the product of natural selection acting on subtle phenotypic differences that may be difficult to

478 resolve with laboratory experiments. Consequently, it may be challenging to quantify the impact
479 of a locus (i.e., its effect size) on the phenotypic variation for a quantitative trait without large
480 sample sizes. For example, although a positive relationship is estimated between ethylene
481 production and the presence of the premature stop codon in HK167-28586 (Fig. 1.6), the model is
482 not significant for this low sample size (Nagelkerke $R^2 = 0.14$, $p = 0.27$; Nagelkerke, 1991). A
483 much larger sample will be required to obtain an accurate estimate of the contribution of this
484 locus to variation in nitrogen fixation. However, the first step in determining how genetic
485 variation at this locus may act to enhance nitrogen fixation performance is to identify its pattern
486 of expression with respect to nitrogen limitation.

Tables

Table 1.1 Strain means and errors (95% confidence) for normalized ethylene production and summary statistics for all strain means. Underlined values are those which were in the top quartile of their respective columns.

	Normalized Ethylene Production (PPM hr^{-1})		
WC Strain	37ºC	55ºC	Pooled
111	4.92 ±0.60	4.09 ±0.60	4.51 ±0.42
112	4.99 ±0.84	0.41 ±0.03	2.70 ±0.79
114	12.36 ±1.30	2.06 ±0.43	7.21 ±1.66
116	4.78 ±1.13	2.83 ±0.82	3.80 ±0.72
119	7.18 ±0.75	5.49 ±1.32	6.33 ±0.76
1110	6.35 ±1.54	1.80 ±0.45	4.08 ±1.02
213	3.49 ±0.67	2.57 ±1.09	3.07 ±0.60
217	1.35 ±0.22	1.90 ±0.41	1.62 ±0.24
245	8.54 ±0.54	6.71 ±2.37	7.63 ±1.19
246	5.68 ±0.51	2.17 ±1.26	4.27 ±0.78
249	1.94 ±0.23	1.81 ±0.41	1.88 ±0.23
326	8.07 ±1.54	0.68 ±0.33	4.04 ±1.33
344	4.91 ±1.34	9.25 ±1.29	7.08 ±1.09
434	0.30 ±0.04	0.41 ±0.06	0.35 ±0.04
438	5.71 ±0.41	5.49 ±1.00	5.60 ±0.52
439	8.70 ±1.22	5.05 ±0.98	6.88 ±0.92
441	5.58 ±1.25	4.36 ±0.81	5.03 ±0.76
442	4.03 ±0.95	2.47 ±0.68	3.25 ±0.60
527	4.78 ±0.64	2.37 ±0.55	3.58 ±0.54
538	4.79 ±0.50	3.06 ±0.68	3.92 ±0.48
542	5.52 ±0.44	5.58 ±0.90	5.55 ±0.45
558	4.56 ±0.45	2.67 ±0.53	3.61 ±0.43
559	2.91 ±0.47	2.34 ±1.14	2.62 ±0.59
Mean	5.28	3.29	4.29
Minimum	0.30	0.41	0.35
Lower Quartile	4.30	1.98	3.16
Median	4.92	2.57	4.04
Upper Quartile	6.03	4.71	5.57
Maximum	12.36	9.25	7.63

Table 1.2 Mixed effects model summary using R's lme4 package syntax. $R^2_{(m)}$ is the model variation explained by fixed factors and $R^2_{(c)}$ is the total variation explained by the model.

Model	AIC	BIC	Log Likelihood	$R^2_{(m)}$	$R^2_{(c)}$
$ET \sim 1 + H + C \times T + (1|I) + (1+T|S)$	3525.4	3556.7	-1752.7	0.21	0.74

Variables	Definition
C	^{14}C-bicarbonate incorporation rate (μg C hr $^{-1}$)
ET	Normalized ethylene production (PPM hr $^{-1}$)
H	Heterocyst frequency (heterocysts per cell counted)
S	Nominal *M. laminosus* strain ID
T	Temperature treatment (ºC)

489

Table 1.3 Candidate genes associated with variation in nitrogen fixation performance at 37 ºC, the relative genetic differentiation between phenotypic classes (F_{ST}), and corresponding annotations of homologous genes in the NCBI database. Emboldened rows are genes that are both F_{ST} and D_{XY} outliers.

Gene	F_{st}	Annotation
56-42545	0.4836	Potassium channel protein
59-39736	0.4563	Ribosomal protein S12 methylthiotransferase
124-17867	0.4258	Hypothetical Protein
223-48944	0.3992	Cobalt transport protein component
19-42545	**0.3966**	**Diguanylate cyclase**
131-35450	0.3961	Hydrogenase maturation protease hupW
227-48944	0.3886	Hypothetical Protein
44-39685	0.3750	HSP htpX
41-4197	0.3750	Hypothetical Protein
112-40954	0.3684	Hypothetical Protein
24-24749	0.3649	Hypothetical Protein
33-4197	0.3633	Hypothetical Protein
13-30518	0.3633	GCN5 family acetyltransferase
31-39736	**0.3584**	**Hypothetical Protein**
19-24813	0.3514	Hypothetical Protein
237-48944	0.3503	Macrolide ABC transporter/ATP-binding protein
38-47543	0.3468	Hypothetical Protein
7-45	**0.3379**	**Hypothetical Protein**
1-30518	0.3333	Group 1 glycosyl transferase
23-24749	0.3333	Hypothetical Protein

490

Table 1.4 Candidate genes associated with variation in nitrogen fixation performance at 37 ºC, the absolute genetic differentiation between phenotypic classes (D_{XY}), and corresponding annotations of homologous genes in the NCBI database. Emboldened rows are genes that are both F_{ST} and D_{XY} outliers.

Gene	D_{xy}	Annotation
33-9029	0.0057	Hypothetical protein
19-42545	**0.0035**	**Diguanylate cyclase**
11-43020	0.0032	Hypothetical protein
12-14867	0.0031	Hypothetical protein
124-17867	0.0026	SCP-like extracellular protein
20-51983	0.0020	Hypothetical protein
31-39736	**0.0018**	**Hypothetical protein**
17-20539	0.0018	Hypothetical protein
44-43317	0.0018	tRNA(Ile)-lysidine synthase
118-37089	0.0018	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
45-9675	0.0017	Phosphate ABC transporter substrate-binding protein
243-48944	0.0016	Putative Anti-Sigma regulatory factor (Ser/Thr kinase)
68-28680	0.0016	Nitrate ABC transporter, inner membrane subunit
309-48944	0.0015	Hypothetical protein
128-40954	0.0015	Nucleotidyl transferase
44-29888	0.0014	Hypothetical protein
7-45	**0.0014**	**Hypothetical protein**
54-57682	0.0013	Hypothetical protein
69-28680	0.0013	Amino acid ABC transporter substrate-binding protein
228-48944	0.0012	Ferritin, Dps family protein

Table 1.5 Candidate genes associated with variation in nitrogen fixation performance at 55 ºC, the relative genetic differentiation between phenotypic classes (F_{ST}), and corresponding annotations of homologous genes in the NCBI database. Emboldened rows are genes that are both F_{ST} and D_{XY} outliers.

Gene	F_{st}	Annotation
28-39736	0.6954	Adenylyl-sulfate kinase
1-33964	0.5698	Hopene-associated glycosyltransferase HpnB
49-34361	0.5273	Cation-transporting ATPase
65-42545	0.5176	Cytochrome-c oxidase subunit 3
93-42545	0.5150	Membrane-anchored histidine kinase
44-39378	0.4833	Hypothetical Protein
15-39736	0.4625	Teichoic acid-transporting ATPase/ABC transporter
48-34361	0.4611	Cation-transporting ATPase
35-24749	0.4600	Cyclic nucleotide binding
72-4197	0.4563	RNP-1 like binding protein
55-20539	0.4526	Hypothetical Protein
22-24813	0.4328	Hypothetical Protein
29-33117	0.4278	Cytochrome-c 550 psbV
69-4197	0.4172	Hypothetical Protein
7-17867	**0.4103**	**MFS transporter**
4-65273	0.4103	S-adenosylmethionine synthetase
59-13348	0.4082	ArsR family transcriptional regulator
12-39736	0.4074	Hypothetical Protein
77-48944	**0.4045**	**ABC transporter**
29-2411	0.4028	FAD dependent oxidoreductase

491

Table 1.6 Candidate genes associated with variation in nitrogen fixation performance at 55 ºC, the absolute genetic differentiation between phenotypic classes (D_{XY}), and corresponding annotations of homologous genes in the NCBI database. Emboldened rows are genes that are both F_{ST} and D_{XY} outliers.

Gene	D_{xy}	Annotation
77-48944	**0.0081**	**ABC transporter**
10-32834	0.0053	Urease alpha subunit
31-20539	0.0039	Glycosyl transferase family 2
62-37089	0.0038	Chlorophyll A-B binding protein
93-17867	0.0036	Hypothetical protein
19-42545	0.0035	Hypothetical protein
12-17867	0.0033	ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase
80-29888	0.0031	ABC-2 type transporter
92-17867	0.0029	Exodeoxyribonuclease VII small subunit
118-40954	0.0025	Hypothetical protein
102-32982	0.0025	Hydrogenase expression/formation protein HypD
34-39685	0.0024	Glycosyl transferase family 2
76-17867	0.0023	Cobalt transport protein
132-48944	0.0022	Putative ABC-type transport system, permease component
127-48944	0.0021	Basic membrane lipoprotein
84-29888	0.0021	Oxidoreductase domain protein
23-29888	0.0007	FHA domain containing protein
81-29888	0.0020	Teichoic-acid-transporting ATPase
7-17867	**0.0020**	**MFS transporter**
20-51983	0.0020	Hypothetical protein

Table 1.7 Candidate genes associated with variation in pooled nitrogen fixation performance, the relative genetic differentiation (F_{ST}) between phenotypic classes, and corresponding annotations of homologous genes in the NCBI database. Emboldened rows are genes that are both F_{ST} and D_{XY} outliers.

Gene	F_{st}	Annotation
167-28586	**0.5397**	**Histidine Kinase**
20-24813	0.5313	Cytochrome P450
59-17867	**0.5000**	**Hypothetical Protein**
15-39736	0.4813	Teichoic-acid-transporting ATPase/ABC transporter
72-4197	0.4563	RNP-1-like binding protein
77-48944	**0.4378**	**ABC transporter**
7-17867	0.4264	MFS transporter
22-24813	0.4179	Hypothetical Protein
29-33117	0.4175	Cytochrome-c550
12-51983	0.4141	Unknown
108-24813	0.4138	Cyclic nucleotide-binding protein
59-13348	0.4082	ArsR-family transcriptional regulator
12-39736	0.4074	Hypothetical Protein
57-16960	0.4057	Hypothetical Protein
10-32834	**0.4050**	**Urease alpha subunit**
29-2411	0.4028	FAD dependent oxidoreductase
66-15735	0.4000	Hypothetical Protein
62-37089	**0.4000**	**Chlorophyll a-b binding protein**
223-48944	0.3992	Cobalt transport protein component CbiN
19-42545	**0.3966**	**Unknown**

492

Table 1.8 Candidate genes associated with variation in pooled nitrogen fixation performance, the absolute genetic differentiation between phenotypic classes (D_{XY}), and corresponding annotations of homologous genes in the NCBI database. Emboldened rows are genes that are both F_{ST} and D_{XY} outliers.

Gene	D_{xy}	Annotation
167-28586	**0.0161**	**Histidine kinase**
77-48944	**0.0088**	**ABC transporter**
33-9029	0.0057	Hypothetical Protein
10-32834	**0.0056**	**Urease alpha subunit**
93-17867	0.0054	Hypothetical Protein
62-37089	**0.0052**	**Chlorophyll A-B binding protein**
169-28586	0.0039	Hypothetical Protein
31-20539	0.0039	Glycosyl transferase family 2
19-42545	**0.0035**	**Hypothetical Protein**
84-29888	0.0034	Oxidoreductase domain protein
12-17867	0.0033	ABC-type $NO_3^-/SO_2O^-/CHO_3^-$ transport system
11-43020	0.0032	Function Unknown
12-14867	0.0031	Glycosyl transferase, group 1
111-46452	0.0031	HGT; MbtH domain protein
299-48944	0.0030	Putative peptidoglycan binding protein
59-17867	**0.0030**	**Hypothetical Protein**
92-17867	0.0029	Exodeoxyribonuclease 7 small subunit
80-29888	0.0029	Teichoic-acid-transporting ATPase
40-3504	0.0027	Hypothetical Protein
118-40954	0.0026	NAD(P)H-quinone oxidoreductase subunit 4

Figures

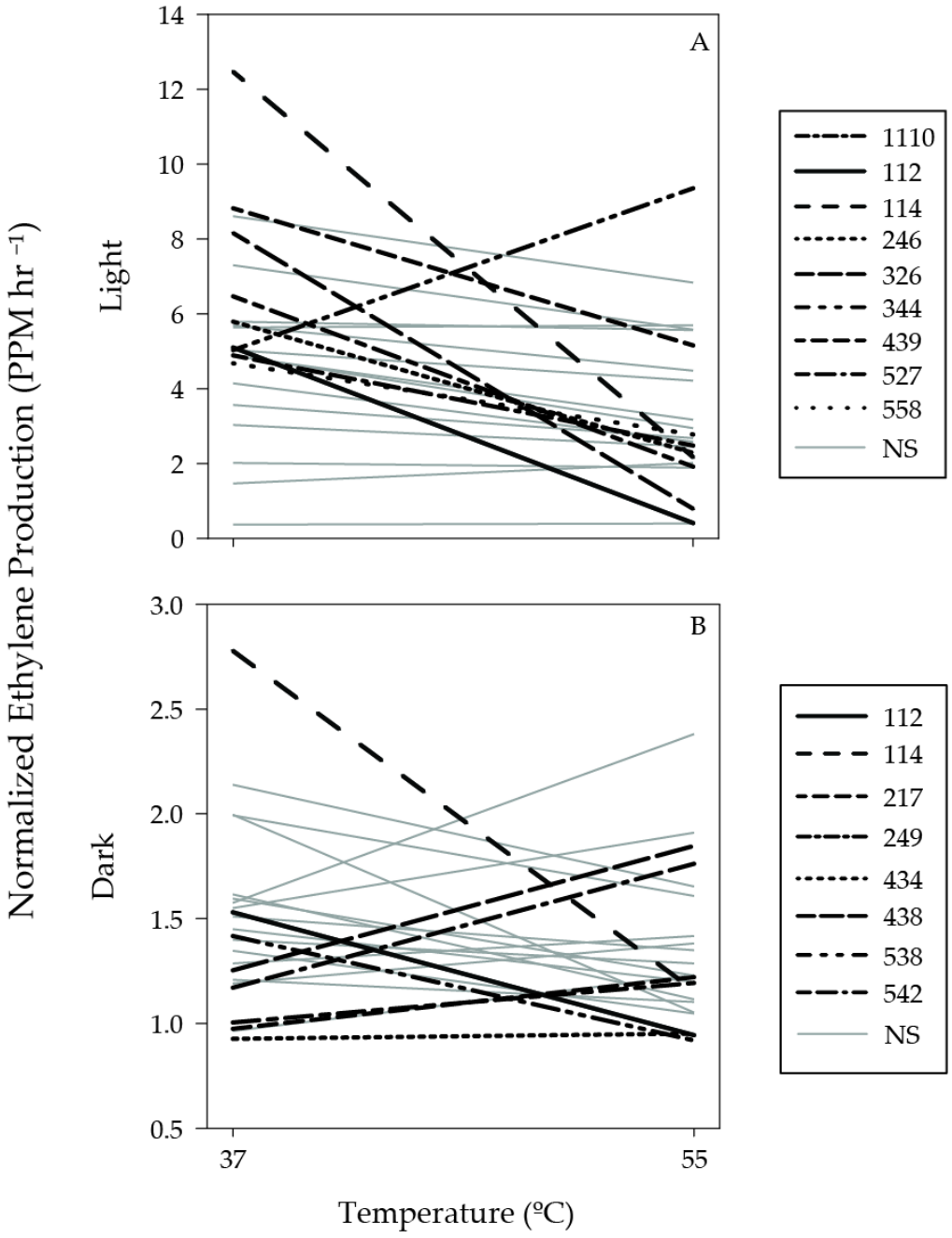

Fig. 1.1 *M. laminosus* reaction norms for normalized ethylene production across temperature treatments. Emboldened lines indicate a slope that is significantly different from zero at the 95% confidence interval (ANOVA, p >0.05). Grey lines represent assayed strains for which slopes were not significantly different from zero (NS). For clarity, error bars are not shown (see Table 1.1 for this information).

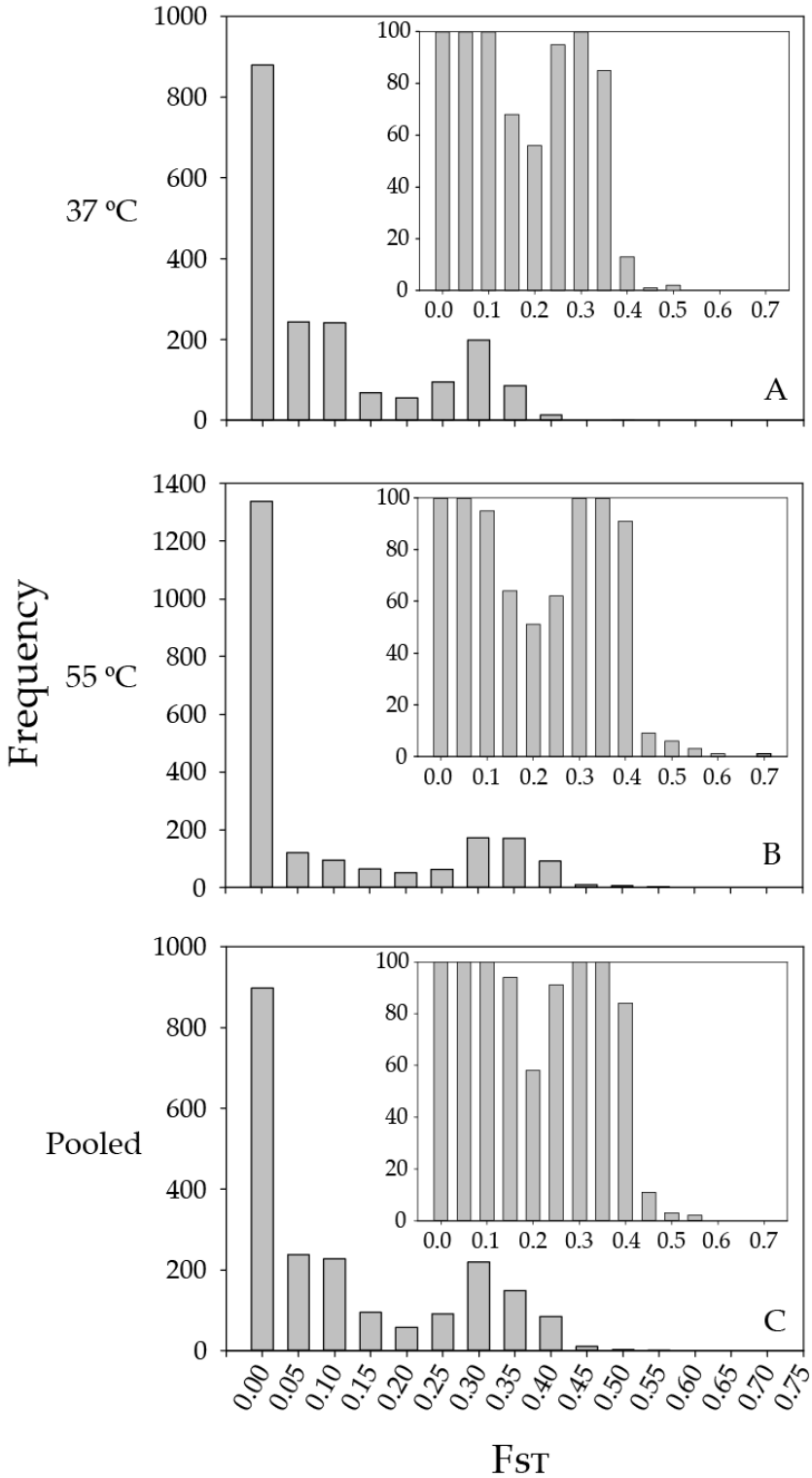

Fig. 1.2 Relative genetic differentiation (F_{ST}) between upper and lower phenotypic classes in the 37 ºC (A), 55 ºC (B), and pooled datasets (C). Insets have re-scaled views of the data in panels A, B, and C to better visualize outlier values.

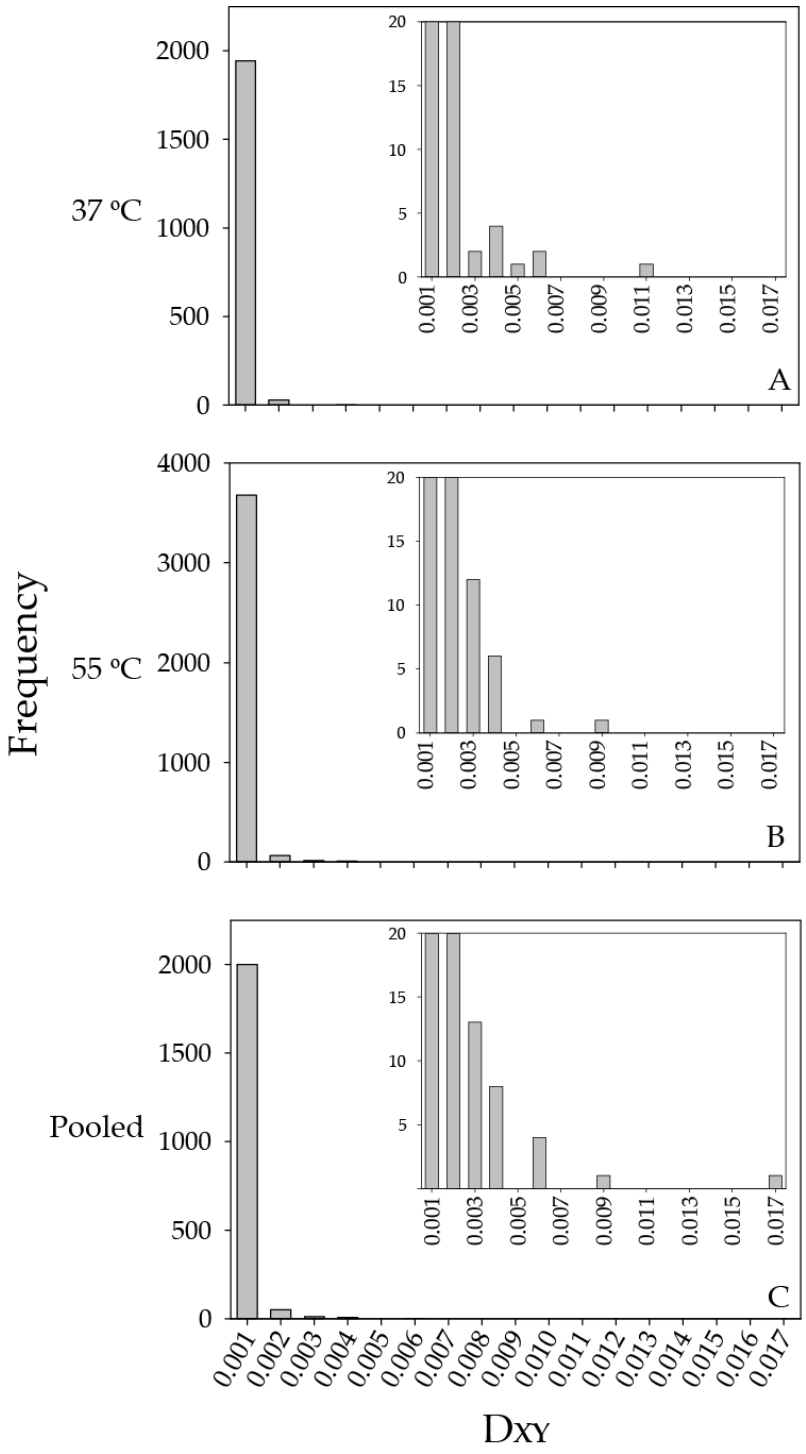

Fig. 1.3 Absolute genetic differentiation (D_{XY}) between upper and lower phenotypic classes in the 37 ºC (A), 55 ºC (B), and pooled datasets (C). Insets have re-scaled views of the data in panels A, B, and C to better visualize outlier values.

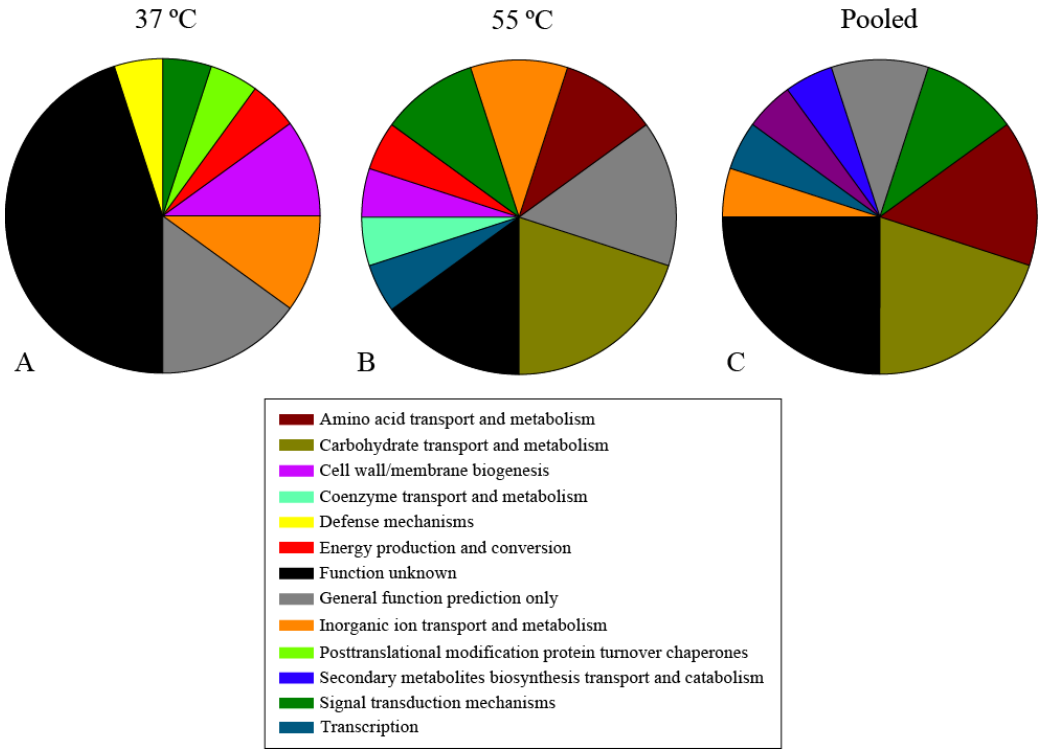

Fig. 1.4 Annotated clusters of orthologous groups (COG) categories for the top 1% of F_{ST} outlier loci in the 37 ºC (A), 55 ºC (B), and pooled datasets (C).

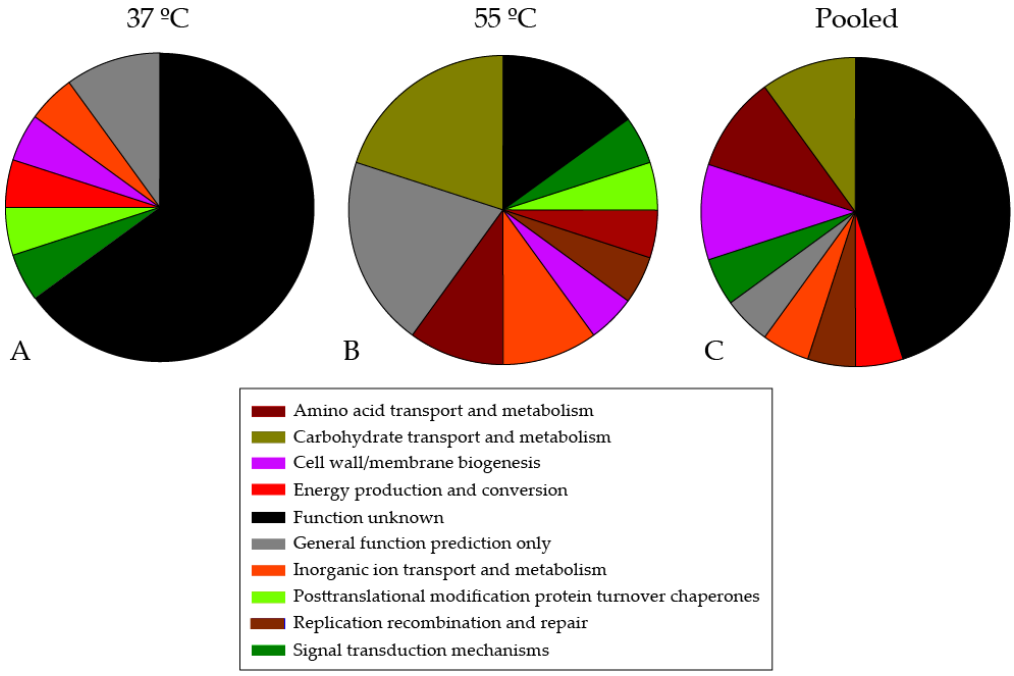

Fig. 1.5 Annotated clusters of orthologous groups (COG) categories for the top 1% of D_{XY} outlier loci in the 37 ºC (A), 55 ºC (B), and pooled datasets (C).

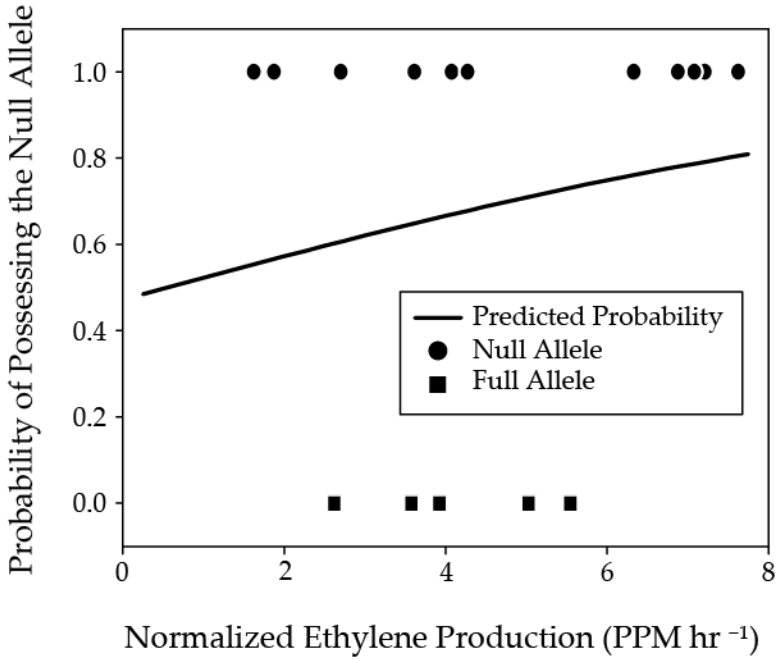

Fig 1.6 General linear model predicting the probability that a White Creek *M. laminosus* strain contains the premature stop codon at the 167-28586 locus based on strain mean normalized ethylene production. Circles are strain means for pooled ethylene production for strains with the premature stop codon and squares are strain means for those which have the full allele.

Chapter 2

Abstract

In chapter 1, I reported that allelic variation at a histidine kinase gene (HK167-28586) was significantly associated with variation in nitrogen fixation performance in *M. laminosus* from a population at White Creek in Yellowstone NP. HK167-28586 also exhibits several molecular evolutionary signatures that suggest that allelic diversity at this locus encodes functionally important variation that has been maintained by some form of balancing selection. For there to be a phenotypic difference between allelic classes, I expect that the expression of the full and functional allele is required during either heterocyst development and/or steady state growth under nitrogen-limitation. Five different strains representing the three different alleles at the HK167-28586 locus that were observed in the White Creek population were tested in a simple expression assay under nitrogen-limitation using a reverse transcription polymerase chain reaction (RT-PCR) approach. Expression of the HK transcript was present at T_0 in all but one strain, and the transcript was not present in any samples at 48 hours after nitrogen step-down. I conclude that gene expression was turned off following heterocyst maturation and the onset of steady-state growth under nitrogen-limitation. More studies will be needed to assign a specific functional role to HK167-28586 and to determine the contribution of allelic variation at this locus to variation in nitrogen fixation.

Introduction

Two-component signal transduction systems (two component systems; TCSs) are the principal means by which bacteria sense and respond to environmental changes (Gao, Stock, 2009; Wuichet *et al.*, 2010). TCSs are involved in a profound suite of critical cellular functions, including, but not limited to, chemotaxis, virulence, symbiosis, and carbon and nitrogen metabolisms (Parkinson, Kofoid, 1992). These signaling pathways can account for a significant proportion of bacterial genomes (up to ~2.5% in the cyanobacterium *Synechocystis sp.*; Mizuno et al. 1996) and have likely been crucial for bacterial adaptation.

Prototypical TCSs involve two separate proteins. A histidine kinase (HK), which usually has a sensory domain that interacts with the intra- or extracellular environment, serves as the input component of the system. The HK then transfers phosphoryl groups to a cognate response regulator (RR) to effect a change in gene expression or, sometimes, protein activity (Galperin, 2010). Once stimulated, a well-conserved ATP binding domain at the C-terminal end of the HK catalyzes the autophosphorylation of a conserved histidine residue. The phosphorylated HK then transfers the His-bound phosphoryl group to an asparagine residue in a well-conserved receiver

domain on the RR. The phosphorylation of the receiver domain changes the structural conformation of a variable effector domain, which carries out the regulatory activity of the pathway. TCSs are often involved with several branching networks, yet operate with astounding fidelity (Laub, Goulian, 2007).

In chapter 1, I reported that allelic variation at a HK gene (HK167-28586) is associated with the ability of *M. laminosus* to fix nitrogen. Three alleles ranging in amino acid identity from 96-99% were observed in the White Creek sample. These include: an allele with a nonsense mutation that is expected to eliminate 9 of the 17 predicted ATP binding residues in the encoded protein and therefore is expected to lack autophosphorylation and kinase activities; an allele that differs at 39 nucleotide positions; and an apparently rare recombinant null allele that is identical to the former at the 3' end and to the latter at the 5' end, and therefore contains the nonsense mutation. Because ATP hydrolysis is central to autophosphorylation and subsequent kinase activities, the loss of more than half of the ATP-binding pocket is expected to render the HK nonfunctional for these activities, even if it is expressed. For there to be a phenotypic difference between allelic classes, I further expect that expression of the full and functional allele is required either during heterocyst development and/or steady-state growth under nitrogen limitation. Here, this expectation is tested in a simple expression assay under nitrogen-limitation using a reverse transcription polymerase chain reaction (RT-PCR) approach.

Methods

Culture Conditions and Sample Collection

The assay was designed such that expression of HK167-28586 could be studied during both heterocyst development and subsequent steady-state growth under nitrogen-limitation. Five different strains representing the three different alleles at the HK167-28586 locus that were observed in the White Creek population were used in the experiment to determine whether each allele is transcribed. Strains WC119 and WC344 both have the full copy, while WC527 and

WC538 contain the null allele, and WC249 is the sole representative of the recombinant null allele. *M. laminosus* cells were grown in semi-continuous batch cultures in D medium (Castenholz, 1988) until ~5 mL of cell mass had accumulated for each strain. Just before the expression assay, the cells were washed twice in ND medium (D medium without combined nitrogen) by vortexing, centrifuging, and pouring off the supernatant before adding cells to triplicate flasks containing 250 mL of ND medium. Cultures were maintained at 37 ºC with a 12/12 hour light/dark cycle. The first cell sample was taken ~30 minutes after transfer to ND media (in the last hour of the dark cycle) and serves as the first time-point (T_0).

Approximately 0.5 mL of cell mass was collected at 0, 6, 12, 18, 24, 36, and 48 hours after T_0 using sterile Pasteur pipets and 2 mL microcentrifuge tubes. Samples were immediately immersed in liquid nitrogen and stored at -80 ºC until extraction. A Qiagen RNeasy mini extraction kit was used to isolate RNA according to the manufacturer's instructions. Prior to constructing cDNA from RNA transcripts, the presence and quality of RNA was checked on a NanoDrop spectrometer, and DNA contamination of the RNA prep was screened via PCR using the primers and cycling conditions described below. A Thermo Scientific Maxima First Strand cDNA Synthesis Kit for RT-PCR was used to construct first strand cDNA according to the manufacturer's instructions. First strand synthesis was accompanied by a template-negative control.

RT-PCR

HK cDNA was amplified by touchdown PCR on an MJR PTC-100 thermal cycler. The forward (5'-GGAATCCACCAACTATGG-3') and reverse (5'-CCAGGTGTAGAGTAGCAC-3') primers were designed manually. The resulting amplicon was 1025 bp in length and included the premature stop codon mutation of the putative null alleles. An initial denaturation step at 94 ºC for 3 min was followed by 30 cycles of 1 min at 94 ºC, 30 sec at variable annealing temperatures, and 1 min at 72 ºC. The initial annealing temperature was 54 ºC and decreased every 10 cycles, reaching a final annealing temperature of 50 ºC. A final extension phase at 72

ºC for 3 minutes completed the program. PCRs were run with a template-negative and a positive control. Presence of the HK transcript at each time point was determined for each strain via gel electrophoresis of the cDNA amplicon. TAE gels consisted of 2% agarose and were run for ~15 min at 98 V. Amplified DNA was stained using ethidium bromide and visualized on a UV transilluminator.

Results

Expression of the HK transcript was present at T_0 in all but the strain WC344 samples, which first showed expression after 6 hours (Fig. 2.1). The first strain for which we could not detect the HK transcript was WC527 at 24 hours, though this may be due, in part, to the extremely low biomass left to harvest in this strain at that time point. The HK transcript was not present in any samples at 48 hours, and so appears to have been turned off somewhere between 36-48 hours after nitrogen step-down. Though no quantitative estimates of heterocyst frequencies were made during this experiment, visual inspections of each strain at each time point suggest that all of the experimental strains reached their maximal heterocyst frequencies between 24-36 hours after nitrogen depletion. This timeframe for heterocyst maturation is also corroborated by numerous other studies of heterocyst development in closely related cyanobacteria (Kumar *et al.*, 2010; Wong, Meeks, 2001). In a subsequent expression assay under +N conditions (nitrate as N source), all three alleles of the gene were turned on in representative strains WC119, WC249, and WC344 (data not shown).

Discussion

Our results demonstrate that all three of the HK167-28586 alleles from the gene identified in Chapter 1 are expressed during *M. laminosus* heterocyst development and during +N growth. Although all three alleles are expressed, the alleles containing the premature stop codon are expected to be constitutively "off" because, without half of the ATP-binding sites,

642 autophosphorylation and kinase activities should be effectively nullified. Results from Chapter 1
643 suggest that silencing HK167-28586 prior to steady-state growth under nitrogen limitation
644 contributes to enhanced nitrogen fixation. This presents us with several important questions:
645 what is the regulatory function of HK167-28586 (i.e., its cognate response regulator(s) and the
646 transcriptional network in which it participates)?; is it really a null allele, and what are the
647 regulatory consequences of the elimination of much of the ATP-binding pocket?; and does the
648 nonsense mutation come with a cost under certain environmental conditions?

649 The regulatory role of HK167-28586 cannot be discerned from this study. However, the
650 HK167-28586 exhibits ~50% amino acid identity with three histidine kinases in the *Anabaena*
651 PCC 7120 genome (alr1551, alr2739, and alr4882). For several reasons, alr4882 appears to be the
652 ortholog. The HK gene is the same length as alr4882, which is not the case with other *Anabaena*
653 homologs. Also, local gene order in the region is conserved: both HK167-28586 and alr4882 are
654 downstream of a putatively orthologous annotated gene encoding a protein with a FIST domain
655 (alr4881 and the corresponding *M. laminosus* gene are ~56% identical at the amino acid level).
656 While FIST domains are phylogenetically widespread, they are biochemically uncharacterized,
657 though they are predicted to bind small molecules (Borziak, Zhulin, 2007). In Mella-Herrara *et*
658 *al.* (2011), alr4882 is referred to in unpublished data as a gene that is upregulated during
659 heterocyst development 5-9 hours after nitrogen step-down. Insertional inactivation had no
660 observed negative impact on –N growth, but their observations appear to be qualitative (the
661 standard assay is to identify Fox- mutants by the yellowing of colonies on a plate) and don't
662 speak to the fine-scale fitness effects that may be operating. Gene knock-outs in *Anabaena* PCC
663 7120 and subsequent functional assays for nitrogen fixation may reveal differences in fitness that
664 are too subtle for qualitative assays.

665 We expect that the premature stop codon nullifies the ability of HK167-28586 to function
666 as a kinase. However, there are at least three possible scenarios where HK167-28586 could

continue to function in a TCS. First, the premature stop codon could be "leaky" and allow a full-length HK to be translated often enough to effect a regulatory response. Alternatively, though unlikely, the remaining ATP binding sites that are found upstream of the nonsense mutation may be sufficient to promote ATP-binding and autophosphorylation activity. To determine whether alleles with the premature stop codons have lost the ability to autophosphorylate, heterologously expressed protein can be assayed for autophosphorylation activity (Hastie *et al.*, 2006). Using this approach, the enzyme activity of each allelic variant of HK167-28586 could be compared quantitatively and with high sensitivity. In yet another scenario, the allele may have lost autophosphorylation activity but can still participate in the signal transduction network via phosphatase activity. Many histidine kinases are bifunctional enzymes that can phosphorylate as well as dephosphorylate their cognate response regulators (Alex, Simon, 1994). All of the White Creek *M. laminosus* alleles have an intact phosphotransfer domain, and, in at least one case, this domain alone is sufficient to support phosphatase activity of a histidine kinase (EnvZ; Zhu *et al.*, 2000). Manipulation of the balance of kinase versus phosphatase activities might be an additional possible mechanism by which allelic variation at 167-28586 effects transcriptional rewiring via loss of function (Hottes, 2013).

Understanding the regulatory consequences of each HK167-28586 allele will be crucial in determining how genetic variation affects nitrogen fixation performance. Finding HK167-28586's cognate response regulator(s) would make great strides in this regard. Phosphotransfer profiling (Skerker *et al.*, 2005) uses ATP radiolabelled with ^{32}P and SDS-PAGE to identify phosphotransfer events between purified HKs and RRs. This approach is necessary for identifying "orphans", or HK-RR pairs that are not expressed under the same operon, as is the case here. The differences in the transcriptome among strains with different alleles identified in RNA-seq data would likely yield insight into the downstream regulatory consequences of each HK167-28586 allele and help guide more directed functional assays under varying environmental

conditions. Transcriptomic comparisons between allele variants would help to identify any regulatory networks that have been altered as a result of the nonsense mutation. Furthermore, the existence of a recombinant null allele may enable us to parse the individual regulatory effects of the nonsense mutation and the highly polymorphic region at the 5' end of the gene. It is possible that the null allele results in a fitness trade-off under varying environmental conditions. A transcriptomic approach could also be used to compare differences in gene regulation for each allele under varying environmental conditions. Investigations such as these will inform our understanding of the role that balancing selection may play in the maintenance of the HK167-28586 alleles.

Conclusion

All three HK167-28586 alleles were expressed during nitrogen-replete conditions and during heterocyst development under nitrogen-limitation. Thus, I expect there to be differences in the transcriptional profiles of strains with functional and putative null alleles, respectively. This is the first step in demonstrating a functional basis for the pattern of association with improved nitrogen fixation in *M. laminosus* from Chapter 1. However, more studies will be needed to be able to assign a functional role to HK167-28586 and to determine the underlying cause of variation in nitrogen fixation for each allele.

Figures

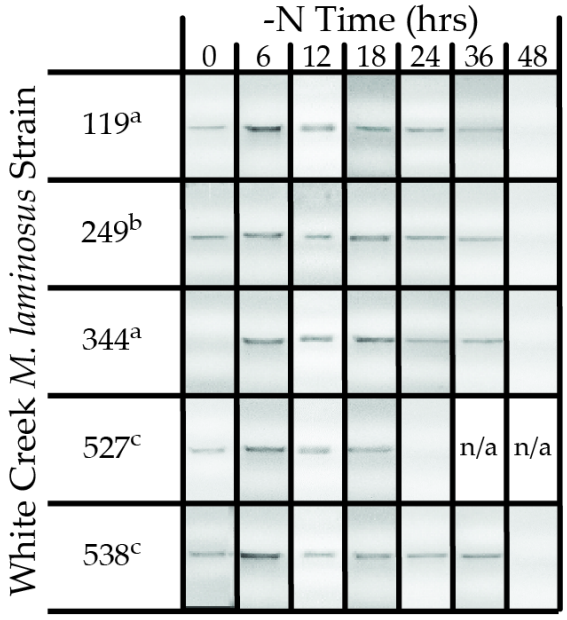

Fig. 2.1 Presence or absence of a HK167-28586 transcript after nitrogen step-down in five *M. laminosus* strains from White Creek. Subscripts next to strain numbers indicate which allele the strain contains ("a" is the null allele, "b" is the recombinant null allele, and "c" is full the copy allele).

Literature Cited

Alex LA, Simon MI (1994) Protein histidine kinases and signal tranduction in prokaryotes and eukaryotes. *Trends in Genetics* **10**, 133-138.

Aziz RK, Bartels D, Best AA, *et al.* (2008) The RAST Server: rapid annotations using subsystems technology. *BMC genomics* **9**, 75.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution* **23**, 38-44.

Bates D, Maechler M, Bolker B, Walker S (2014) lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-6. http://CRAN.R-project.org/package=lme4.

Berman-Frank I, Lundgren P, Chen YB, *et al.* (2001) Segregation of nitrogen fixation and oxygenic photosynthesis in the marine cyanobacterium Trichodesmium. *Science* **294**, 1534-1537.

Böhm H (1998) Regulation of nitrogen fixation in heterocyst-forming cyanobacteria. *Trends in Plant Science* **3**, 346-351.

Borziak K, Zhulin IB (2007) FIST: a sensory domain for diverse signal transduction pathways in prokaryotes and ubiquitin signaling in eukaryotes. *Bioinformatics* **23**, 2518-2521.

Buikema WJ, Haselkorn R (2001) Expression of the Anabaena hetR gene from a copper-regulated promoter leads to heterocyst differentiation under repressing conditions. *Proceedings of the National Academy of Sciences* **98**, 2729-2734.

Campbell EL, Summers ML, Christman H, Martin ME, Meeks JC (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. *Journal of Bacteriology* **189**, 5247-5256.

Carrasco CD, Holliday SD, Hansel A, Lindblad P, Golden JW (2005) Heterocyst-specific excision of the Anabaena sp. strain PCC 7120 hupL element requires xisC. *Journal of Bacteriology* **187**, 6031-6038.

Castenholz RW (1988) Culturing methods for cyanobacteria. In: *Methods in Enzymology* (ed. Abelson J), pp. 68-93. Academic Press, San Diego, CA, USA.

Currier TC, Haury JF, Wolk CP (1977) Isolation and preliminary characterization of auxotrophs of a filamentous Cyanobacterium. *Journal of Bacteriology* **129**, 1556-1562.

Dunn JH, Wolk CP (1970) Composition of the Cellular Envelopes of Anabaena cylindrica. *Journal of Bacteriology* **103**, 153-158.

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

Epstein B, Branca A, Mudge J, *et al.* (2012) Population genomics of the facultatively mutualistic bacteria *Sinorhizobium meliloti* and *S. medicae*. *PLoS genetics* **8**, e1002868.

Ernst A, Böhme H, Böger P (1983) Phosphorylation and nitrogenase activity in isolated heterocysts from *Anabaena variabilis* (ATCC 29413). *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **723**, 83-90.

Flores E, Herrero A, Wolk CP, Maldener I (2006) Is the periplasm continuous in filamentous multicellular cyanobacteria? *Trends in Microbiology* **14**, 439-443.

Galloway JN, Dentener FJ, Capone DG, *et al.* (2004) Nitrogen Cycles: Past, Present, and Future. *Biogeochemistry* **70**, 153-226.

Galperin MY (2004) Bacterial signal transduction network in a genomic perspective. *Environmental microbiology* **6**, 552-567.

Galperin MY (2010) Diversity of structure and function of response regulator output domains. *Current Opinion in Microbiology* **13**, 150-159.

Gao R, Stock AM (2009) Biological insights from structures of two-component proteins. *Annual review of microbiology* **63**, 133-154.

767 Hastie CJ, McLauchlan HJ, Cohen P (2006) Assay of protein kinases using radiolabeled ATP: a
768 protocol. *Nature Protocols* **1**, 968-971.
769 Hedrick PW (2006) Genetic Polymorphism in Heterogeneous Environments: The Age of
770 Genomics. *Annual Review of Ecology, Evolution, and Systematics* **37**, 67-93.
771 Hengge R (2009) Principles of c-di-GMP signalling in bacteria. *Nature Reviews Microbiology* **7**,
772 263-273.
773 Holm L, Sander C (1997) An evolutionary treasure: unification of a broad set of amidohydrolases
774 related to urease. *Proteins Structure Function and Genetics* **28**, 72-82.
775 Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining,
776 estimating and interpreting F_{ST}. *Nature Reviews Genetics* **10**, 639-650.
777 Hottes AK, Freddolino PL, Khare A, *et al.* (2013) Bacterial adaptation through loss of function.
778 *PLoS genetics* **9**, e1003617.
779 Huang G, Fan Q, Lechno-Yossef S, *et al.* (2005) Clustered genes required for the synthesis of
780 heterocyst envelope polysaccharide in *Anabaena sp.* strain PCC 7120. *Journal of*
781 *Bacteriology* **187**, 1114-1123.
782 Huang X, Dong Y, Zhao J (2004) HetR homodimer is a DNA-binding protein required for
783 heterocyst differentiation, and the DNA-binding activity is inhibited by PatS.
784 *Proceedings of the National Academy of Sciences of the United States of America* **101**,
785 4848-4853.
786 Hughes JB, Daily GC, Ehrlich PR (1997) Population diversity: its extent and extinction. *Science*
787 **278**, 689-692.
788 Imashimizu M, Yoshimura H, Katoh H, Ehira S, Ohmori M (2005) NaCl enhances cellular cAMP
789 and upregulates genes related to heterocyst development in the cyanobacterium,
790 *Anabaena sp.* strain PCC 7120. *FEMS Microbiology Letters* **252**, 97-103.
791 Jenal U (2004) Cyclic di-guanosine-monophosphate comes of age: a novel secondary messenger
792 involved in modulating cell surface structures in bacteria? *Current Opinion in*
793 *Microbiology* **7**, 185-191.
794 Jones KM, Haselkorn R (2002) Newly Identified Cytochrome c Oxidase Operon in the Nitrogen-
795 Fixing Cyanobacterium *Anabaena sp.* Strain PCC 7120 Specifically Induced in
796 Heterocysts. *Journal of Bacteriology* **184**, 2491-2499.
797 Kaneko T, Nakamura Y, Wolk CP, *et al.* (2001) Complete genomic sequence of the filamentous
798 nitrogen-fixing cyanobacterium *Anabaena sp.* strain PCC 7120. *DNA research* **8**, 205-
799 213.
800 Kumar K, Mella-Herrera RA, Golden JW (2010) Cyanobacterial Heterocysts. *Cold Spring*
801 *Harbor perspectives in biology* **2**, a000315.
802 Laub MT, Goulian M (2007) Specificity in Two-Component Signal Transduction Pathways.
803 *Annual review of genetics* **41**, 121-145.
804 Liengen T (1999) Environmental factors influencing the nitrogen fixation activity of free-living
805 terrestrial cyanobacteria from a high arctic area, Spitsbergen. *Canadian Journal of*
806 *Microbiology* **45**, 573-581.
807 Lindberg P, Devine E, Stensjö K, Lindblad P (2012) HupW Protease Specifically Required for
808 Processing of the Catalytic Subunit of the Uptake Hydrogenase in the Cyanobacterium
809 *Nostoc sp.* Strain PCC 7120. *Applied and Environmental Microbiology* **78**, 273-276.
810 Luck GW, Daily GC, Ehrlich PR (2003) Population diversity and ecosystem services. *Trends in*
811 *Ecology & Evolution* **18**, 331-336.
812 Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of
813 population genomics: from genotyping to genome typing. *Nature Reviews Genetics* **4**,
814 981-994.
815 Mella-Herrera RA, Neunuebel MR, Golden JW (2011) *Anabaena sp.* strain PCC 7120 conR
816 contains a LytR-CpsA-Psr domain, is developmentally regulated, and is essential for
817 diazotrophic growth and heterocyst morphogenesis. *Microbiology* **157**, 617-626.

Miller SR, Carvey D, Kistler C, Pedersen D (2006) Adaptive clinal variation of a microbial population along a natural thermal gradient. *Abstracts of the General Meeting of the American Society for Microbiology* **106**, 587.

Miller SR, Williams C, Strong AL, Carvey D (2009) Ecological Specialization in a Spatially Structured Population of the Thermophilic Cyanobacterium Mastigocladus laminosus. *Applied and Environmental Microbiology* **75**, 729-734.

Miller SR, Wingard CE, Castenholz RW (1998) Effects of Visible Light and UV Radiation on Photosynthesis in a Population of a Hot Spring Cyanobacterium, aSynechococcus sp., Subjected to High-Temperature Stress. *Applied and Environmental Microbiology* **64**, 3893-3899.

Mizuno T, Kaneko T, Tabata S (1996) Compilation of All Genes Encoding Bacterial Two-component Signal Transducers in the Genome of the Cyanobacterium, *Synechocystis sp.* Strain PCC 6803. *DNA research* **3**, 407-414.

Mullineaux CW, Mariscal V, Nenninger A, *et al.* (2008) Mechanism of intercellular molecular exchange in heterocyst-forming cyanobacteria. *The EMBO Journal* **27**, 1299-1308.

Nagelkerke NJ (1991) A note on a general definition of the coefficient of determination. *Biometrika* **78**, 691-692.

Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* **4**, 133-142.

Nicolaisen K, Mariscal V, Bredemeier R, *et al.* (2009) The outer membrane of a heterocyst-forming cyanobacterium is a permeability barrier for uptake of metabolites that are exchanged between cells. *Molecular Microbiology* **74**, 58-70.

Parkinson JS, Kofoid EC (1992) Communication Modules in Bacterial Signaling Proteins. *Annual review of genetics* **26**, 71-112.

Perzl M, Reipen IG, Schmitz S, *et al.* (1998) Cloning of conserved genes from Zymomonas mobilis and Bradyrhizobium japonicum that function in the biosynthesis of hopanoid lipids. *Biochimica et biophysica acta* **1393**, 108-118.

Popa R, Weber PK, Pett-Ridge J, *et al.* (2007) Carbon and nitrogen fixation and metabolite exchange in and between individual cells of Anabaena oscillarioides. *The ISME Journal* **1**, 354-360.

Renosto F, Seubert PA, Segel IH (1984) Adenosine 5'-phosphosulfate kinase from Penicillium chrysogenum. Purification and kinetic characterization. *Journal of Biological Chemistry* **259**, 2113-2123.

Roncel M, Kirilovsky D, Guerrero F, Serrano A, Ortega JM (2012) Photosynthetic cytochrome c550. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1817**, 1152-1163.

Skerker JM, Prasol MS, Perchuk BS, Biondi EG, Laub MT (2005) Two-Component Signal Transduction Pathways Regulating Growth and Cell Cycle Progression in a Bacterium: A System-Level Analysis. *PLoS Biol* **3**, e334.

Staal M, Lintel-Hekkert St, Harren F, Stal L (2001) Nitrogenase activity in cyanobacteria measured by the acetylene reduction assay: a comparison between batch incubation and on-line monitoring. *Environmental microbiology* **3**, 343-351.

Stewart W, Fitzgerald G, Burris R (1967) In situ studies on N_2 fixation using the acetylene reduction technique. *Proceedings of the National Academy of Sciences of the United States of America* **58**, 2071.

Stewart WDP (1970) Nitrogen fixation by blue-green algae in Yellowstone thermal areas*. *Phycologia* **9**, 261-268.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* **14**, 671-688.

Tamagnini P, Axelsson R, Lindberg P, *et al.* (2002) Hydrogenases and hydrogen metabolism of cyanobacteria. *Microbiology and Molecular Biology Reviews* **66**, 1-20.

Thomas JC, Godfrey PA, Feldgarden M, Robinson DA (2012) Candidate Targets of Balancing Selection in the Genome of *Staphylococcus aureus*. *Molecular Biology and Evolution* **29**, 1175-1186.

Wang H, Sivonen K, Rouhiainen L, *et al.* (2012) Genome-derived insights into the biology of the hepatotoxic bloom-forming cyanobacterium *Anabaena sp.* strain 90. *BMC genomics* **13**, 613.

Werck-Reichhart D, Feyereisen R (2000) Cytochromes P450: a success story. *Genome Biology* **1**, reviews3003.3001–reviews3003.3009.

Witt H, Malatesta F, Nicoletti F, Brunori M, Ludwig B (1998) Cytochrome-c- binding site on cytochrome oxidase in Paracoccus denitrificans. *European Journal of Biochemistry* **251**, 367-373.

Wolk CP (2000) Heterocyst Formation in *Anabaena*. In: *Prokaryotic Development* (eds. Brun Y, Shimkets LJ), pp. 83-104. ASM Press, Washington D.C.

Wolk CP, Cai Y, Cardemil L, *et al.* (1988) Isolation and complementation of mutants of Anabaena sp. strain PCC 7120 unable to grow aerobically on dinitrogen. *Journal of Bacteriology* **170**, 1239-1244.

Wolk CP, Ernst A, Elhai J (1994) Heterocyst metabolism and development. In: *The Molecular Biology of Cyanobacteria* (ed. Bryant DA), pp. 769-823. Kluwer Academic Publishers, Dordrecht.

Wong FC, Meeks JC (2001) The hetF gene product is essential to heterocyst differentiation and affects HetR function in the cyanobacterium Nostoc punctiforme. *J Bacteriol* **183**, 2654-2661.

Wuichet K, Cantwell BJ, Zhulin IB (2010) Evolution and phyletic distribution of two-component signal transduction systems. *Current Opinion in Microbiology* **13**, 219-225.

Yoon H-S, Golden JW (2001) PatS and Products of Nitrogen Fixation Control Heterocyst Pattern. *Journal of Bacteriology* **183**, 2605-2613.

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research* **18**, 821-829.

Zhu Y, Qin L, Yoshida T, Inouye M (2000) Phosphatase activity of histidine kinase EnvZ without kinase catalytic domain. *Proceedings of the National Academy of Sciences* **97**, 7808-7813.