



All Theses and Dissertations

2015-12-01

Gene Networks Involved in Competitive Root Colonization and Nodulation in the *Sinorhizobium meliloti*-*Medicago truncatula* Symbiosis

Ryan D. VanYperen
Brigham Young University

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>

 Part of the [Microbiology Commons](#)

BYU ScholarsArchive Citation

VanYperen, Ryan D., "Gene Networks Involved in Competitive Root Colonization and Nodulation in the *Sinorhizobium meliloti*-*Medicago truncatula* Symbiosis" (2015). *All Theses and Dissertations*. 6177.
<https://scholarsarchive.byu.edu/etd/6177>

This Dissertation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Gene Networks Involved in Competitive Root Colonization and Nodulation in the
Sinorhizobium meliloti – *Medicago truncatula*

Symbiosis

Ryan D. VanYperen

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Joel S. Griffitts, Chair
David L. Erickson
Julianne Grose
Jeff Maughan
William R. McCleary

Department of Microbiology and Molecular Biology
Brigham Young University

December 2015

Copyright © 2015 Ryan D. VanYperen

All Rights Reserved

ABSTRACT

Gene Networks Involved in Competitive Root Colonization and Nodulation in the *Sinorhizobium meliloti* – *Medicago truncatula* Symbiosis

Ryan D. VanYperen
Department of Microbiology and Molecular Biology, BYU
Doctor of Philosophy

The rhizobia-legume symbiosis is the most agriculturally significant source of naturally fixed nitrogen, accounting for almost 25% of all biologically available nitrogen. Rhizobia-legume compatibility restrictions impose limits on symbiotic nitrogen fixation. In many cases, the molecular basis for symbiotic compatibility is not fully understood. The signals required for establishing a symbiotic partnership between nitrogen-fixing bacteria (e.g. *Sinorhizobium meliloti*) and leguminous plants (e.g. *Medicago truncatula*) have been partially characterized at the molecular level. The first stage of successful root colonization is competitive occupation of the rhizosphere (which is poorly understood). Here, the bacteria introduce themselves as potential symbiotic partners through the secretion of glycolipid “Nod” factors. In response, the host facilitates a more exclusive mode of colonization by the formation of a root nodule – a new organ capable of hosting dense intracellular populations of symbiotic rhizobia for nitrogen fixation. This dissertation reports the exhaustive identification of *S. meliloti* genes that permit competitive colonization of the *M. truncatula* rhizosphere, and includes a mechanistic study of one particular bacterial signaling pathway that is crucial for both rhizosphere colonization and nodulation.

I have made use of Tn-seq technology, which relies on deep sequencing of large transposon mutant libraries to monitor *S. meliloti* genotypes that increase or decrease in relative abundance after competition in the rhizosphere. This work included the collaborative development of a new computational pipeline for performing Tn-seq analysis. Our analysis implicates a large ensemble of bacterial genes and pathways promoting rhizosphere colonization, provides hints about how the host plant shapes this environment, and opens the door for mechanistic studies about how changes in the rhizosphere are sensed and interpreted by the microbial community. Notable among these sensory pathways is a three-protein signaling system, consisting of FeuQ, FeuP, and FeuN, which are important for both rhizosphere colonization and nodule invasion by *S. meliloti*. The membrane-bound sensor kinase FeuQ can either positively or negatively influence downstream transcription of target genes by modulating the phosphorylation state of the transcriptional activator FeuP. FeuN, a small periplasmic protein, inhibits the positive mode of FeuPQ signaling by its direct interaction with the extracellular region of FeuQ. FeuN is essential for *S. meliloti* viability, underscoring the vital importance of controlling the activity of downstream genes. In summary, I have employed several powerful genetic, genomic, computational, and biochemical approaches to uncover a network of genes and pathways that coordinate root colonization and nodulation functions.

Keywords: rhizosphere colonization, nitrogen fixation, Tn-seq, two-component signaling

ACKNOWLEDGEMENTS

The work reported here would not be possible without the support of my wife and children. I would also like to thank family members who have provided encouragement and support throughout my undergraduate and graduate education: my grandfathers Jeff Horn and Joseph Odom, my aunts Charlotte Huntzinger and Georgia Simmons, my wife's parents Harold and Shauna Bell, my mother Martha Odom and stepfather Mark Odom.

I would like to thank my friend and mentor Dr. Joel Griffiths for helping me develop my ability to think analytically, plan properly controlled experiments, and to speak and write precisely and accurately. The influence Dr. Griffiths has had on my academic and personal development cannot be overstated.

I would also like to thank my excellent committee members for providing important training and critique of my work. I would like to thank William Zundel for providing me with a rare opportunity to develop skills in teaching and coursework preparation as a microbiology course instructor.

I have received valuable advice and insight from post-doctoral fellows Mathew Crook, Skip Price and Stewart Gardner, as well as fellow graduate students Phillip Bennallack, and Clarice Perry.

I would like to thank the BYU undergraduates who have made significant contributions to work on my projects: Taylor Orton, Jessica Meyers, and Kevin Boehme. Lastly, I would like to thank research collaborators, including Graham Walker, Jon Penterman, Marcus Arnold (Massachusetts Institute of Technology), Cara Haney (Harvard University), and Charles Nicolet (University of Southern California).

TABLE OF CONTENTS

Title Page.....	i
ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF FIGURES.....	ix
LIST OF TABLES.....	xii
ABBREVIATIONS	xiv
Chapter 1. Introduction	1
1.1 The importance of nitrogen fixation.....	1
1.2 Symbiotic nitrogen fixation	5
1.2.1 Introduction	5
1.2.2 Initiation of infection.....	7
1.2.3 Root hair colonization and invasion	9
1.2.3.1 Role of exopolysaccharides in root hair invasion	11
1.2.3.2 Role of cyclic β glucans in root hair invasion	11
1.2.3.3 Role of lipopolysaccharides in root hair invasion.....	12
1.2.3.4 Role of type three secretion systems	13
1.2.3.5 Role of microbial signal cascades in root hair invasion.....	13
1.2.4 Nodule tissue development	14
1.2.5 Bacteroid formation and nitrogen fixation	15
1.2.6 Section conclusions.....	15
1.3 Bacterial stimulus recognition: Two-component signaling systems	17

1.3.1 Introduction	17
1.3.2 Sensor kinases.....	19
1.3.2.1 Sensor kinase structural elements.....	19
1.3.2.1.1 SK sensing domain.....	20
1.3.2.1.2 SK HAMP domain	21
1.3.2.1.3 SK DHp and CA Transmitter Domains	22
1.3.2.2 Sensor kinase catalytic activities	23
1.3.2.2.1 Autophosphorylation activity.....	23
1.3.2.2.2 Phosphotransfer activity	23
1.3.2.2.3 Phosphatase activity.....	24
1.3.3 Response regulator proteins.....	26
1.3.3.1 Response regulator half life	26
1.3.3.2 Activation of target genes	27
1.3.4 SK-RR Specificity.....	27
1.3.5 Auxiliary regulator proteins of two-component signaling	29
1.3.5.1 Inhibition of autophosphorylation	29
1.3.5.2 Inhibition of phosphotransfer.....	30
1.3.5.3 Direct stimulation of phosphatase activity	30
1.3.5.4 Periplasmic regulators of TCSs.....	32
1.3.6 Section conclusions.....	35
1.4 Microbial genetics of rhizosphere colonization	36
1.4.1 Introduction	36
1.4.2 Amino acid synthesis genes.....	38
1.4.3 Flagella driven chemotaxis	38
1.4.4 Thiamine, biotin, and riboflavin production.....	39
1.4.5 Lipopolysaccharide (LPS).....	39

1.4.6 Iron binding.....	40
1.4.7 Section conclusions.....	40
1.5 Conclusions	42
 Chapter 2. Genetic analysis of signal integration by the <i>Sinorhizobium meliloti</i> sensor kinase	
FeuQ	44
2.1 Summary	44
2.2 Introduction.....	45
2.3 Materials and Methods	60
2.3.1 Bacteria culture technique.....	60
2.3.2 Plasmid construction.....	60
2.3.3 Spheroplast preparation.....	60
2.3.4 Western blot assay	61
2.3.5 Bacterial adenylate cyclase two-hybrid (BACTH) analysis	61
2.3.6 β -galactosidase assays	61
2.3.7 Random mutagenesis	62
2.3.8 TTT scan of feuQ and feuN.....	62
2.4 Results.....	66
2.4.1 Bacterial two-hybrid analysis indicates a physical interaction between FeuN and FeuQ.....	66
2.4.2 Does the FeuN/Q-interaction cause an increase in FeuQ phosphatase activity?	66
2.4.3 FeuQ mutations lead to both FeuN-insensitive and FeuN-mimicking phenotypes.....	69
2.4.4 TTT scan of mature FeuN peptide shows regions critical for FeuN function	71
2.4.5 Conserved compatibly-charged residues within FeuQ and FeuN are required for FeuN/Q interaction.....	73
2.5 Discussion	77
2.6 Acknowledgements.....	78

2.7 Open questions	79
Chapter 3. Gene networks involved in competitive rhizosphere colonization in the	
<i>Sinorhizobium meliloti</i> – <i>Medicago truncatula</i> symbiosis	80
3.1 Summary	80
3.2 Introduction	82
3.3 Materials and Methods	85
3.3.1 Bacteria culture technique	85
3.3.1.1 BRM	85
3.3.1.2 RSM	85
3.3.2 Tn-seq library construction	86
3.3.3 Tn-seq experimental conditions	86
3.3.3.1 Rich medium control	86
3.3.3.2 Minimal medium control	87
3.3.3.3 Plant inoculation and growth	87
3.3.3.4 Sequence library preparation	87
3.3.3.5 Reproducibility	88
3.3.4 Data analysis	88
3.3.5 Fitness calculations	89
3.3.6 Plasmid and strain construction	90
3.3.7 Competition assays	90
3.4 Results	91
3.4.1 Tn-seq screen for rhizosphere-required genes	91
3.4.2 Summary of Tn-seq data collection and reproducibility of results	94
3.4.3 Essential gene candidates	97
3.4.4 Tn-seq-predicted genes required for root fitness	97

3.4.5 Tn-seq predicted genes required for RSM fitness only	99
3.4.6 Tn-seq predicted genes more important for root than for RSM fitness.....	100
3.4.7 Selection of candidates for Tn-seq prediction verification.....	101
3.4.8 Verification tests for genes predicted to be required for root fitness.....	103
3.4.9 Verification tests for mutations predicted to improve root fitness	106
3.4.10 Additional exploration of genes in the FeuPQ two-component system	108
3.5 Discussion	109
3.6 Acknowledgements.....	111
Chapter 4. Future work and Conclusions	112
4.1 Continuing Tn-seq analysis.....	112
4.2 Future work.....	117
4.3 Conclusions	118
REFERENCES	120
SUPPLEMENTAL FIGURES	135
SUPPLEMENTAL TABLES	138
SUPPLEMENTAL METHODS	197
Tn-seq protocols.....	197
pJG714 sequence	200

LIST OF FIGURES

Figure 1-1 Effects of industrially fixed N ₂	2
Figure 1-2 Required energy investments for symbiotic nitrogen fixation and an illustrated schematic of the dynamics of host-microbe benefit in symbiosis.	5
Figure 1-3 Stages of root infection and nodule formation.....	7
Figure 1-4 Chemical structure of <i>Sinorhizobium meliloti</i> NodSM-IV	8
Figure 1-5 Chemical structures for selected <i>S. meliloti</i> signals.....	10
Figure 1-6 Schematic illustration of a typical two-component signaling system.....	20
Figure 1-7 Crystal Structure of an HK-RR Complex	28
Figure 2-1 The <i>S. meliloti</i> RR FeuP is required for efficient nodulation on alfalfa	47
Figure 2-2 Lack of <i>ndvA</i> expression accounts for the symbiotic phenotype of the <i>feuP</i> mutant .	48
Figure 2-3 <i>feuN</i> encodes a conserved alphaproteobacterial protein	50
Figure 2-4 FeuN acts directly on FeuQ-FeuP in a heterologous system	52
Figure 2-5 FeuN/Q-inhibition of FeuP activity requires the FeuP residue Asp-51.....	53
Figure 2-6 Specific FeuQ alterations diminish the effects of FeuN	55
Figure 2-7 FeuN is a cleaved periplasmic protein	59
Figure 2-8 Schematic model of FeuNPQ function, and evidence for direct interaction between FeuN and FeuQ.....	65

Figure 2-9 FeuN enhances the negative effect of a kinase-deficient FeuQ allele	68
Figure 2-10 Amino acid changes in FeuQ that cause FeuN-insensitive or FeuN-mimicking phenotypes	70
Figure 2-11 Map of TTT scan phenotypes of FeuQ (periplasmic domain) and FeuN	72
Figure 2-12 Conserved, compatibly charged residues in the periplasmic region of FeuQ and in FeuN are involved in target gene regulation.....	75
Figure 2-13 Charge alterations in the FeuQ ETE and FeuN RPKK motifs abolish the FeuN/FeuQ physical interaction.....	76
Figure 3-1 Schematic overview for the Tn-seq screen for effective rhizosphere colonization	91
Figure 3-2 Schematic Illustration of Illumina library preparation steps.....	93
Figure 3-3 Visualization of insertion frequencies for selected genes on BRM, RSM2, and Root Conditions.....	95
Figure 3-4 Tn-seq data reproducibility and phenotype analysis.....	96
Figure 3-5 Competition experiments comparing mutant growth to wild type growth in RSM and <i>M. truncatula</i> root conditions	104
Figure 3-6 Competition experiments comparing mutant growth to wild type growth in RSM and <i>Medicago truncatula</i> root conditions.....	106
Figure 3-7 <i>ndvA</i> over-expression does not restore competitive rhizosphere colonization to Δ <i>feuP</i> mutants.....	109

Figure 4-1 Use of an expanding panel for identification of general and host-specific required genes for rhizosphere colonization	113
Figure 4-2 Tn-seq data reproducibility and phenotype analysis.....	114
Figure S-1 Quantitative measurement of FeuQ TTT scan phenotypes.....	135
Figure S-2 Quantitative measurement of FeuN TTT scan phenotypes compared to wild-type activity.....	136
Figure S-3 Quantitative measurement of alkaline phosphatase activity of Δ ssPhoA fusions to <i>feuN</i> and two <i>feuN</i> -charge reverse variants.	137

LIST OF TABLES

Table 1-1 Summary table of auxiliary TCS regulators.....	34
Table 2-1. Strains and plasmids used in Chapter 2.....	63
Table 3-1 Illumina run and mapping statistics.....	92
Table 3-2 Selected genes to be disrupted for competition tests.....	102
Table 3-3 Summary of genes verified to play a role in competitive root colonization	105
Table 4-1 Summary of Tn-seq data for selected auxotrophs predicted to be able to colonize root systems	116
Table S1. Primers used in this study*	138
Table S2. Random FeuQ mutants*	142
Table S3. Tri-threonine (TTT) substitution mutants for <i>feuQ</i> and <i>feuN</i>	143
Table S4 Tn-seq predicted <i>S. meliloti</i> Essential Gene Candidates	145
Table S5 Tn-seq predicted <i>S. meliloti</i> genes required for growth in RSM and on <i>Medicago truncatula</i> roots.....	169
Table S6 Tn-seq predicted <i>S. meliloti</i> genes required for <i>Medicago truncatula</i> root only.....	172
Table S7 Tn-seq predicted <i>S. meliloti</i> genes required for growth on RSM only	175
Table S8 Tn-seq predicted <i>S. meliloti</i> genes more important for RSM than for roots	177

Table S9 Tn-seq predicted <i>S. meliloti</i> genes required for <i>Medicago truncatula</i> roots but not Rice roots.....	189
Table S10 Tn-seq predicted <i>S. meliloti</i> genes required for Rice roots but not <i>M. truncatula</i>	191
Table S11 Tn-seq predicted <i>S. meliloti</i> genes required for both Rice and <i>M. truncatula</i> roots..	193

ABBREVIATIONS

LPS – lipopolysaccharides
SK – Sensor Histidine Kinase
RR – Response regulator
His - Histidine
DHp – Dimerization and Histidine phosphorylation
HAMP – histidine kinases, adenylyl transferases, methyl accepting proteins and phosphatases
CA – catalytic and ATP-binding
ORF – Open reading frame
MBP – maltose binding protein
AP – ampicillin
Ara – L-arabinose
BACTH – bacterial adenylate cyclase two-hybrid
Km – kanamycin
Tc – tetracycline
TCS – two-component system
TTT – targeted tri-threonine
Nm – Neomycin
DAP – Diaminopimelic acid
N – nitrogen
NH₃ – Ammonium
PAMPs - pathogen-associated molecular patterns
ROS – reactive oxygen species
SPRI - Solid Phase Reversible Immobilization
ATP - Adenosine triphosphate
T3SS – Type 3 secretion system
Nops – nodulation outer proteins
PCR – polymerase chain reaction

Chapter 1. Introduction

1.1 The importance of nitrogen fixation

Although nitrogen (N) is among the most abundant elements on earth, N in its natural triple-bonded N₂ state is biologically unavailable and therefore constitutes an extremely limiting factor for plant growth (Smil, 1999; Vance, 2001). A small subset of plants including legumes can gain access to N through symbiotic association with nitrogen-fixing microbes, which convert atmospheric nitrogen into biologically available ammonium (NH₃). All other plants must obtain N from the soil in the form of decaying organic matter, manure, or commercial fertilizer made with industrially fixed nitrogen.

Commercial nitrogen fertilization by the Haber-Bosch process is the single largest global source of N for plants (Manual, 1998). Industrially fixed nitrogen converts atmospheric N₂ into biologically available NH₃ in a series of directed chemical reactions. The use of chemical fertilizer has facilitated the global expansion of the human population to a level over 3 times what was previously possible, and currently almost 50 percent of the world population is fed by Haber Bosch nitrogen (Figure 1-1) (Erisman, Sutton, Galloway, Klimont, & Winiwarter, 2008).

Unfortunately, the waste products associated with chemical nitrogen fertilization present a number of problems to the environment, and have lead to serious water and atmosphere pollution. Chemical soil fertilization is a relatively inefficient nitrogen delivery system for plants. Up to half of the nitrogen fertilizer applied to crops is not taken up by plants, but leached into the surrounding soil and water (Zahran, 1999). As the use of chemical fertilizer increases, the agricultural gains level off while the toll of pollution and pollution-associated disease continues to rise. The net public benefits of industrial nitrogen fixation begin to rapidly decline.

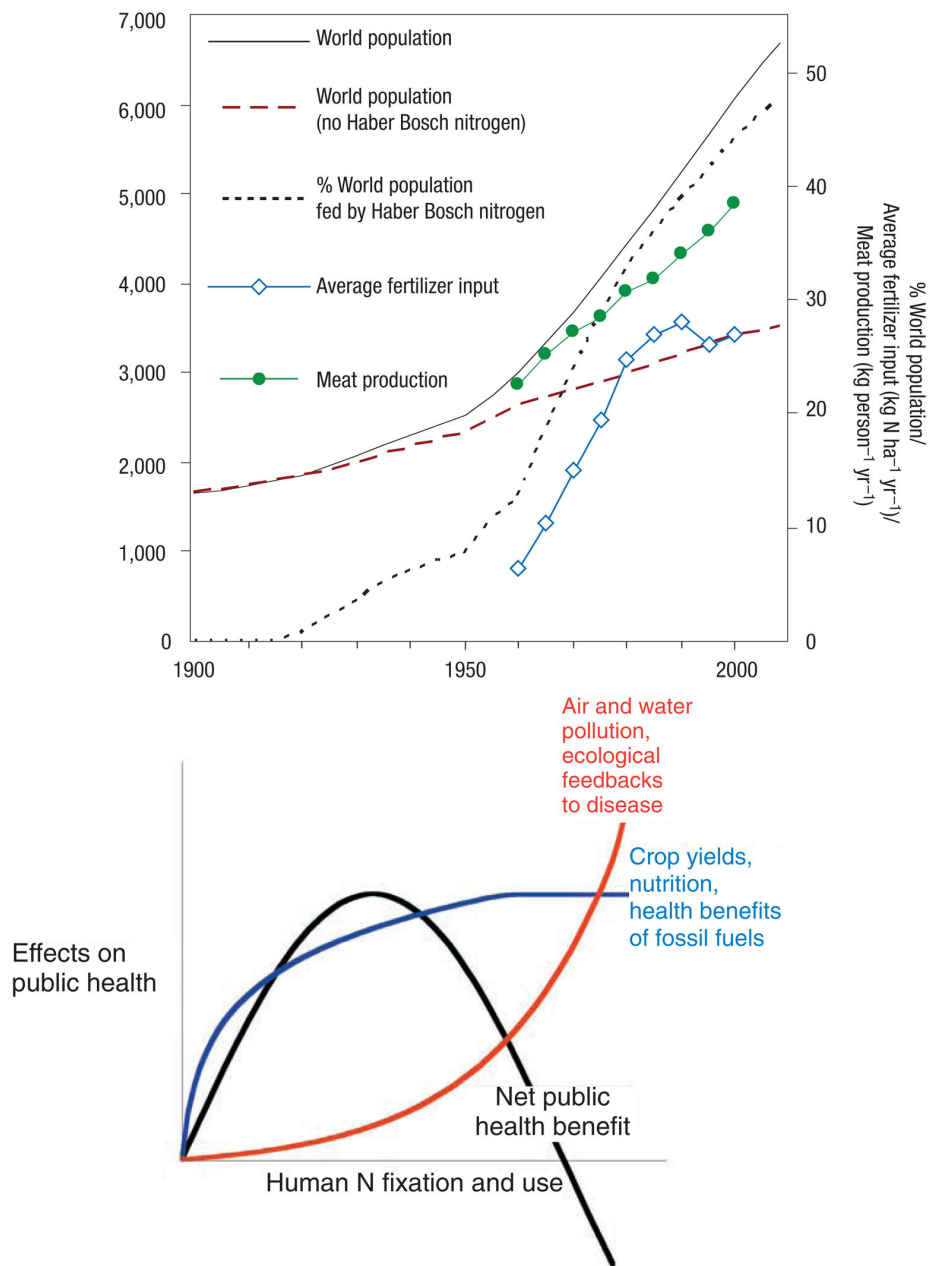


Figure 1-1 Effects of industrially fixed N_2

Top: Trends in human population and nitrogen use throughout the twentieth century. Reprinted with permission from (Erisman et al., 2008)

Bottom: Conceptual model of the overall net public health effects of increasing human fixation and use of atmospheric N_2 . Reprinted with permission from (Townsend et al., 2003)

Symbiotic N₂ fixation is by contrast a much cheaper and more efficient plant delivery system. A small subset of microbes can convert N₂ to biologically available NH₃ using an energy-dependent enzymatic reaction catalyzed by the enzyme nitrogenase. This process can be completed by many genera of cyanobacteria and symbiotic actinomycetes, as well as a number of other free-living and symbiotic bacteria (Smil, 1999). Symbiotic N₂ fixation is the most agriculturally significant form of natural nitrogen fixation. Symbiotic nitrogen fixation occurs between bacteria of the *Rhizobia* family and legumes (Howard & Rees, 1996). The rhizobia-legume symbiosis accounts for up to nearly twenty five percent of all fixed nitrogen (Zahran, 1999). In these systems, the plant has evolved the ability to recruit nitrogen-fixing microbes into plant organs known as nodules, which houses symbiotic bacteria and provides them with the necessary nutrients and conditions for nitrogen fixation.

Rhizobia initiate nodule development with molecular signals that coordinate cooperative infection of a legume host (Oldroyd, 2013). Once the bacterial cells have entered the nodule tissue, the plant induces terminal differentiation of the microbial cells into bacteroids. During differentiation, the bacteria swell and duplicate their genomes many times, transforming into nitrogen-fixing organelle-like bacteroids no longer capable of living independently or reproducing. Though these irreversibly differentiated bacteroids pay a steep price for symbiosis, undifferentiated siblings within the nodule benefit from plant association and are re-released into the soil after nodule senescence.

Not all rhizobia are compatible for symbiotic nitrogen fixation with all legumes. In fact, many rhizobia-legume partnerships are governed by strict requirements that lead to host-range restrictions (Crook et al., 2012; Long, Buikema, & Ausubel, 1982). Researchers have identified some of the basic mechanisms for successful nodule formation, and much is known about the

biological reduction of N₂ by the enzyme nitrogenase. However, many questions about host-specificity and the plant-microbe signals that negotiate a successful symbiotic relationship remain unanswered.

This lack of knowledge restricts the usefulness of symbiotic nitrogen fixation as an alternative to the use of industrial soil fertilizers in spite of the many potential economical and environmental advantages. Additional research is required to more fully understand the principles of host-range restriction and symbiotic compatibility. This dissertation reports the investigation of genes involved in rhizobia-legume root colonization and nodulation in the *S. meliloti* – *M. truncatula* model system. Chapter 2 reports a mechanistic study of a bacterial two-component signaling system shown to be required for successful nodulation. The reported study aims to understand requirements for nodulation at the molecular level. Chapter 3 reports the identification of a large set of genes predicted to be involved in plant root colonization. This study involves a larger scale genomics approach aiming to understand the general requirements for plant infection, identifying targets for mechanistic studies in the future. A number of these genes are verified within this study as determinants of competitive rhizosphere colonization in head to head competition experiments. Chapter 4 of this report includes preliminary data expanding the analysis of rhizosphere colonization genes and outlines ideas for future work.

The following sections of this introduction include a more detailed review of topics important for this study, including symbiotic nitrogen fixation, two-component signaling, and root colonization relevant to the reported findings of this dissertation.

1.2 Symbiotic nitrogen fixation

1.2.1 Introduction

The *Sinorhizobium meliloti* – *Medicago truncatula* symbiosis is an effective and well-characterized model system for studies aimed at identifying the genetic requirements for compatibility in rhizobia-legume symbiotic nitrogen fixation.

Successful symbiotic partnership requires a significant energy investment from both plant and microbe participants. From the microbial perspective, the nitrogenase reaction that catalyzes nitrogen fixation requires 16 ATP and 8 e⁻ per molecule of atmospheric N₂ (Figure 1-2) (Phillips, 1980). Because the symbiotic bacteria rely on the plant for these resources, the plant spends an estimated 8 grams of carbon per gram of nitrogen received to develop nodule tissue and supply the housed microbial community with the nutrients and proper conditions for nitrogen fixation (Phillips, 1980).

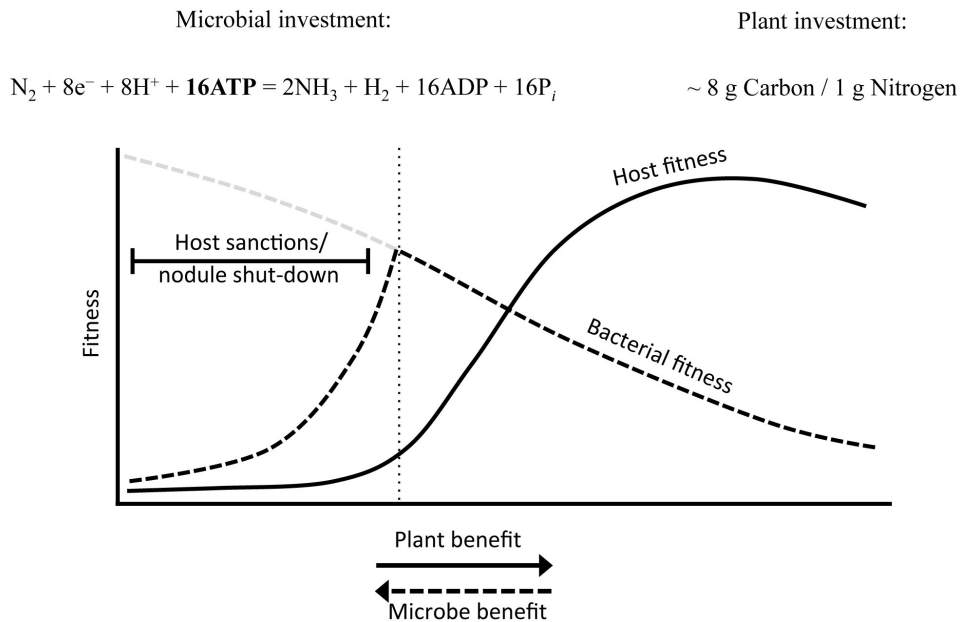


Figure 1-2 Required energy investments for symbiotic nitrogen fixation and an illustrated schematic of the dynamics of host-microbe benefit in symbiosis.

The benefit of symbiotic partnership outweighs the risk of this investment as long as both participants fully complete their responsibilities. Thus, at varying points throughout the process of root invasion and nodulation, the bacteria and plant check one another for compatibility in a continuing negotiation for mutualism throughout the infection process. Miscommunication at important checkpoints for symbiosis results in abortive nodulation. If a strain of rhizobia fails to fix nitrogen for the host plant, the host may respond with “sanctions” that halt nodule development and reduce the fitness of the reproductive clones of that strain (Oono, Anderson, & Denison, 2011).

Characterization of the signals required for symbiotic nitrogen fixation has focused primarily on molecular interactions required for the infection of root tissue and the development of nitrogen fixing nodules (illustrated in figure 1-3). Briefly, successful symbiosis requires a coordinated and cooperative bacterial invasion of root tissue. A plant root hair ensnares a compatible bacterium and undergoes morphological changes resulting in the creation and extension of infection thread tubes that facilitate transportation of the bacteria into a newly developing plant organ known as a nodule. When invading bacteria reach the developing nodule tissue, the bacteria enter the nodule cells by endocytosis, forming a symbiosome. There, the bacteria undergo terminal differentiation into bacteroids. These bacteroids are swollen organelles with multiple copies of the bacterial chromosome, capable of fixing large amounts of nitrogen so long as the nodule provides them with the energy source and oxygen-free conditions in which to do so.

The complex signal interactions between plant and bacteria required to coordinate these stages of infection are regulated both temporally and spatially throughout the stages of nodulation to ensure that each side is fulfilling their obligation for mutualism (Oldroyd, Murray,

Poole, & Downie, 2011). The following sections will describe these stages of infection in greater detail.

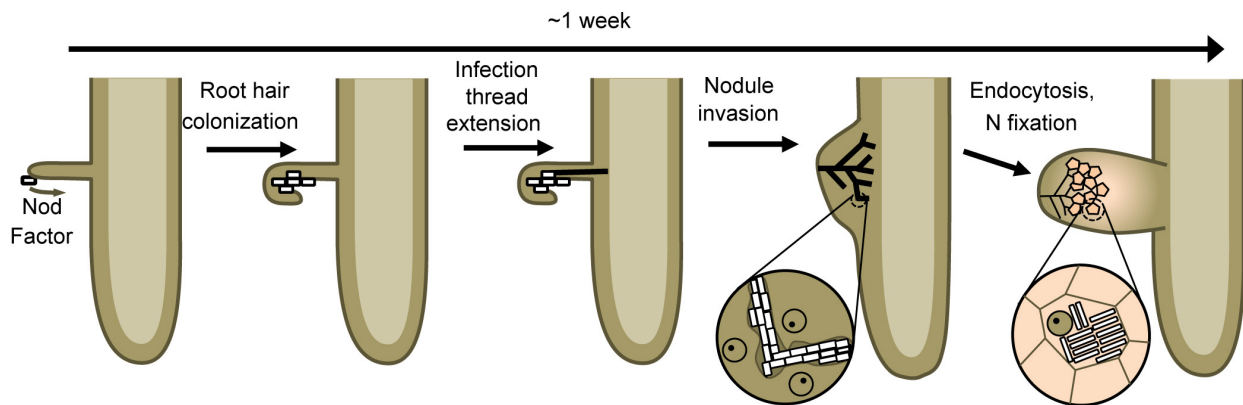


Figure 1-3 Stages of root infection and nodule formation

Illustrated stages of root infection and nodule formation taking place over the course of a week. Cooperative infection is mediated by complex signaling interaction between plant and microbe.

1.2.2 Initiation of infection

Early research in nodule formation led to the discovery that legumes actively seek out symbiotic partners through the release of flavonoids, betaines, and isoflavones into the surrounding soil (reviewed by Gage, 2004). These chemical compounds recruit rhizobia to the area and serve as the first plant introductory signals to a potential symbiotic partner.

In response to these secreted signals, rhizobia initiate the production and secretion of lipochitooligosaccharides known as Nod factors (Perret, Staehelin, & Broughton, 2000). Nod factors are acylated chitin chains that can be decorated with various bacteria-specific functional group modifications. These modifications are identifying signatures that serve as the primary molecular basis for selective host-microbe specificity (Lerouge et al., 1990). Rhizobia that are defective in Nod factor production are not capable of associating with legumes (Long et al., 1982; Meade, Long, Ruvkun, Brown, & Ausubel, 1982), and the secretion of Nod factor by a

compatible bacterium is the initiating step for root invasion, nodulation and symbiotic nitrogen fixation (Dénarié, Debelle, & Promé, 1996; Oldroyd, 2013).

Presently, the ability to recognize a compatible Nod factor and initiate nodule formation leading to symbiotic nitrogen fixation is a defining attribute of legumes (Oldroyd & Dixon, 2014). Interestingly, many non-legumes (including the crop plants tomato and corn), though not capable of participating in symbiotic nitrogen fixation, are able to recognize Nod factors and have been shown to suppress immune response and even stimulate plant growth in response to Nod factor stimulation (Khan, Prithiviraj, & Smith, 2008; Liang et al., 2013; Prithiviraj, Zhou, Souleimanov, Kahn, & Smith, 2003). This has fueled hope that genetic modifications to non-legume crop plants might one day allow them to participate in symbiotic nitrogen fixation.

Legumes exhibit strict specificity in the bacterial Nod factors that they recognize. Some rhizobia species exhibit broad host ranges while others are restricted to compatibility with only a few or even a single species of legume (Denarie, Debelle, & Rosenberg, 1992; Young & Johnston, 1989). The chemical structure for the *S. meliloti* Nod factor NodSM-IV is shown in figure 1-4.

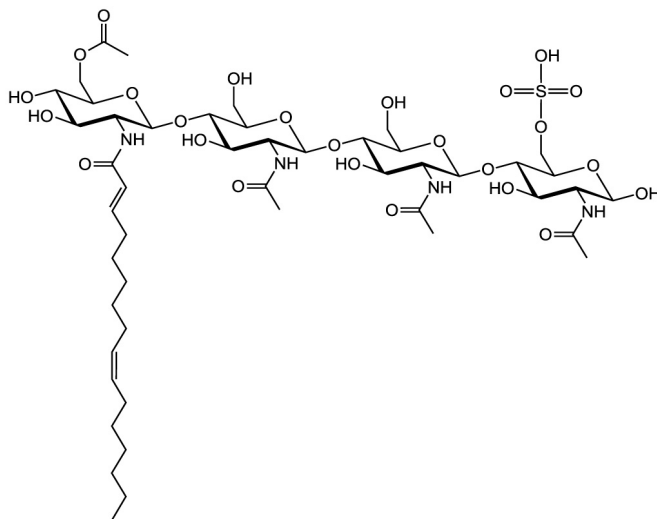


Figure 1-4 Chemical structure of *Sinorhizobium meliloti* NodSM-IV

Rhizobial Nod factors are detected by plant receptors containing LysM domains (Madsen et al., 2003; Radutoiu et al., 2003). The LysM domain is a general peptidoglycan-binding molecule. The potential for variation within this receptor domain likely accounts for the variations in host compatibility observed between a given plant receptor LysM domain and a particular rhizobial Nod factor (Limpens et al., 2003). For *S. meliloti*, the LysM receptor-like kinases *NFR1* and *NFR5* have been shown to be the primary receptors for Nod factor specificity (Oldroyd, 2013).

1.2.3 Root hair colonization and invasion

When a legume has identified a compatible Nod factor, a number of physiological changes begin to take place on the surface of the plant root. Communication signals continue to pass between the plant and bacteria as invasion continues – primarily in the form of secretions and outer membrane signal proteins.

For the plant, auxin, cytokinin, and calcium fluctuations induced by compatible Nod factor initiate nodule organogenesis at the site of infection. Auxin accumulation causes the tip of a root hair to swell and curl around the compatible bacterium, entrapping it in a “shepherd’s crook” (refer to Figure 1-3) (Esseling, Lhuissier, & Emons, 2003). This entrapment leads to calcium fluctuations in the plant tissue that precede the development of an invagination in the root hair, and the formation of tube-like pathways known as infection threads that make their way down to the root tissue (Gibson, Kobayashi, & Walker, 2008; Miwa, Sun, Oldroyd, & Downie, 2006). The invading bacteria travel along these infection threads via cellular division, eventually progressing to the developing nodule at the base of the root hair. Here, bacteria escape the infection threads and are enveloped into the nodule tissue.

The development of infection threads represents another major checkpoint in host-microbe compatibility. In an effort to ensure that only beneficial rhizobia are allowed to infect the root hairs and eventually occupy the developing nodule, the plant continues to monitor secretions and surface molecules throughout the infection process. Communication failures at this stage result in abortive nodulation, characterized by small, white nodules devoid of bacteria or nitrogen-fixing bacteroids.

At this stage, the bacteria secrete a number of signals important for down-regulating the plant immune system and coordinating cooperative infection. Chief among these are bacterial surface polysaccharides, secreted exopolysaccharides, secreted cyclic β glucans, and lipopolysaccharide (LPS) (Gibson et al., 2008). These signals are known determinants of successful nodulation, and the roles of these components will be discussed in further detail below.

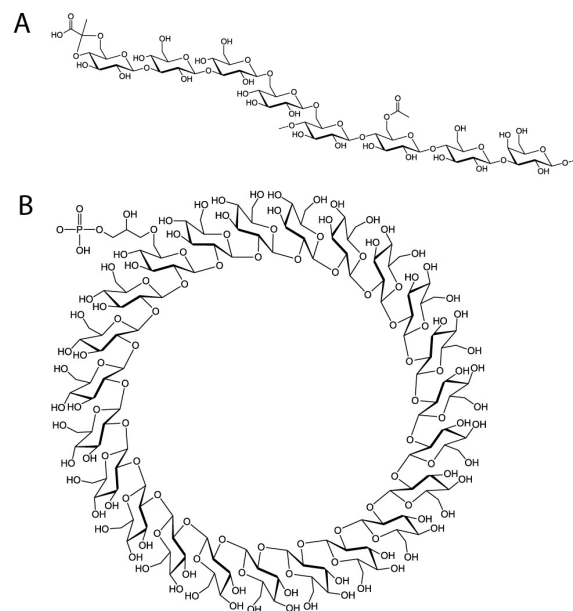


Figure 1-5 Chemical structures for selected *S. meliloti* signals
Shown are examples of secreted signals important for *S. meliloti* symbiosis with *M. truncatula*.
A. The repeating subunit of succinoglycan (EPS I)
B. A cyclic β -(1 \rightarrow 2)-glucan (the degree of substitution with the *sn*-1-phosphoglycerol is variable).

1.2.3.1 Role of exopolysaccharides in root hair invasion

Rhizobia secrete exopolysaccharides (EPS) during the infection process. These exopolysaccharides have been shown to play a role in host-microbe specificity. The best characterized exopolysaccharide important for developing symbiosis is the *S. meliloti* EPS succinoglycan. Succinoglycan is a polysaccharide with an eight sugar repeating unit composed of one galactose and seven glucose residues, with possible acetyl, succinyl, and pyruvyl modifications (See figure 1-5) (Åman, McNeil, Franzén, Darvill, & Albersheim, 1981; L.-X. Wang, Wang, Pellock, & Walker, 1999).

S. meliloti mutants with defects in succinoglycan are capable of inducing nodule formation, but the resulting nodules do not contain bacteria or differentiated bacteroids, and do not facilitate nitrogen fixation (Leigh, Signer, & Walker, 1985). The abortive nodulation appears to be caused by an inability to mediate the formation of infection threads (Pellock, Cheng, & Walker, 2000). Mutations preventing exopolysaccharide production genes range in effect from complete inability to induce infection thread formation to premature abortion of infection threads before the bacteria have reached the base of the root hairs (Cheng & Walker, 1998). Overproduction of exopolysaccharides may also disrupt symbiotic development and prevent colonization of root hairs (Yao et al., 2004). Thus, symbiotic partners must be compatible in both the type and the amount of exopolysaccharides secreted during the invasion process.

1.2.3.2 Role of cyclic β glucans in root hair invasion

Rhizobia also produce cyclic β glucans during plant infection (Spaink, 2000) (see figure 1-5). These cyclized β -1,2 chains of 17-25 glucose residues are important for maintaining symbiotic compatibility during root hair attachment and progression through infection threads

(Dickstein, Bisseling, Reinhold, & Ausubel, 1988; Jones, Kobayashi, Davies, Taga, & Walker, 2007).

Cyclic glucans are produced in response to osmotic conditions and are transported to the periplasm, where they are involved in osmolarity regulation. Cyclic glucans in the periplasmic space are important factors in host-microbe signaling. Disruptions in cyclic glucan production or secretion result in an abortive nodulation phenotype (Griffitts et al., 2008). It is believed that compatible cyclic glucans can facilitate the infection process by mediating host-defense responses and allowing compatible bacteria to survive during the early stages of infection. In soybean, cyclic- β -glucan has been shown to suppress anti-microbial defense responses (Mithöfer, Bhagwat, Feger, & Ebel, 1996).

1.2.3.3 Role of lipopolysaccharides in root hair invasion

In order to protect themselves from pathogenic microbes, plants have evolved sophisticated immune responses to discourage infection. As beneficial microbes enter the plant cell via infection threads and progress into the tissue of the nodule, they must be able to survive within this environment and evade or suppress host immune responses (Toth & Stacey, 2015). Gram-negative bacteria have evolved an outer membrane component known as LPS to help protect themselves from the external environment and avoid chemical attacks and host immune responses. LPS has been shown to be a critical component in nodule invasion (Ernst, Guina, & Miller, 1999). LPS mutants are capable of inducing nodulation and passing through infection threads, but are not able to continue to the stages of bacteroid differentiation and nitrogen fixation in the nodule (Campbell, Reuhs, & Walker, 2002).

1.2.3.4 Role of type three secretion systems

Many bacteria have evolved type three secretion systems (T3SSs) as a means of injecting effector proteins into host cell tissues. These effector proteins facilitate successful microbial infection. This behavior is primarily associated with pathogenic infection. However, T3SSs also appear to play a role in permissive infections, including some examples of symbiotic nitrogen fixation (Deakin & Broughton, 2009). Some of these T3SSs have been shown to play a role in determining host specificity, and mutations in these systems result in abortive nodulation phenotypes (Marie et al., 2003). These T3SSs secrete effector “Nops” (nodulation outer proteins) and are important for maintaining host-microbe compatibility during nodulation. T3SS involvement in symbiotic nitrogen fixation is not wide spread, but appears to be a bacteria genus-specific phenomenon. T3SSs have not been shown to play a role in the *S. meliloti*-*M. truncatula* symbiosis.

1.2.3.5 Role of microbial signal cascades in root hair invasion

Bacteria must recognize and adapt to changes in the environment as the process of nodule colonization progresses. For this purpose, bacteria have adapted two-component signaling cascades. These systems play a major role in controlling the microbial gene expression. A number of two-component signaling systems have been identified as critical signals for continued nodule development and the coordination of nitrogen fixation, including NodVW, FixLJ, and ExoS/ChvI, and FeuPQ (Loh, Garcia, & Stacey, 1997); (David et al., 1988); (Gilles-Gonzalez & Gonzalez, 1993); (Chen, Sabio, & Long, 2008); (Griffitts et al., 2008). Mutations in two-component signaling systems involved in symbiotic compatibility result in abortive nodulation phenotypes.

1.2.4 Nodule tissue development

Nitrogen reduction to biologically available NH_3 occurs within the tissue of the root nodule. Root nodules develop via plant cell division at the nodule primordium – located at the base of an infected root hair (W.-C. Yang et al., 1994). When developing infection threads reach the nodule tissue, the bacteria escape from the infection threads and enter the nodule tissue. Root nodules are divided into two major categories – determinate and indeterminate nodules. Nodule type is determined by the host plant (Newcomb, 1981).

Determinant nodules are circular in shape, and their development ceases somewhat quickly. The infected nodule cells swell as infection progresses, and overall nodule size results from cellular enlargement rather than division. Many bean plants produce determinate nodules (Hirsch, 1992).

Medicago truncatula and other clover, alfalfa, and pea plants produce indeterminate nodules. Unlike determinate nodules, indeterminate nodules are characterized by a persistent developing meristem where cell division continues from the tip as the nodule grows out from the root tissue (Hirsch, 1992). The result is an elongated nodule with a distinct age gradient from the newly developing cells at the tip to the older cells at the base of the root. As the nodule continues to extend, bacterial invasion progresses into the newly developing tissue while bacteroid formation and nitrogen fixation is already under way in the older nodule tissue. For indeterminate nodules, all stages of nodule development are represented within a single nodule.

1.2.5 Bacteroid formation and nitrogen fixation

Once the infection thread has progressed to the root cortex, bacteria escape the infection threads and enter the plant cells. The bacteria terminally differentiate into bacteroids, swelling in size and duplicating their genome to reach up to 24 copies (Mergaert et al., 2006). Bacteroids are enlarged, organelle-like bodies that are optimized for nitrogen fixation and host-microbe nutrient exchange, and are no longer capable of living independently or dividing.

Bacteroid differentiation within the nodule tissue is thought to be driven by plant produced nodule-specific cysteine-rich (NCR) peptides (Van de Velde et al., 2010). Rhizobia strains equipped with enzymes that degrade NCR peptides in some cases do not progress to nitrogen-fixing bacteroids and exhibit increased proliferation within the nodule tissue (Price et al., 2015).

1.2.6 Section conclusions

Researchers are characterizing an increasing number of signals involved in nodule formation and the negotiation of symbiotic nitrogen fixation. The potential for diversity within these signals begins to explain the molecular basis for host-range restrictions and incompatibility in many rhizobia-legume interactions. The chemical signals sent and received throughout the infection process have been extensively reviewed (Jones et al., 2007; Oldroyd, 2013; Oldroyd et al., 2011; Spaink, 2000). For the purposes of this study, I have focused on a few key classes of proteins necessary for symbiotic nitrogen fixation from the microbial perspective. Chapter 2 of this study will further focus attention on microbial receptors required for sensing environmental stimuli, including a dissection of a particular receptor involved in regulating cyclic glucan production.

Studies focusing on nodule development have primarily taken as a given the very first step required for compatibility – soil persistence and competitive colonization of the root surface where Nod factor initiation occurs. Perhaps surprisingly, relatively little is known about the genetic requirements for this competitive root colonization. Chapter 3 of this dissertation explores this question and includes the identification of a number of genes required for competitive root colonization.

1.3 Bacterial stimulus recognition: Two-component signaling systems

1.3.1 Introduction

Bacteria must recognize and adapt to changes in the environment in order to respond to changes in environmental conditions. This requires an ability to sense external stimuli and transmit this signal into the microbial cell, where transcription of appropriate response genes can be coordinated. Bacteria have developed two-component signaling systems (TCSs) that are able to recognize and respond to environmental cues via protein phosphorylation.

Bacteria species rely on the combination of many two-component systems to coordinate the regulation of gene networks required for a robust response to diverse environmental conditions. *E. coli* combines as many as 62 different TCSs (Mizuno, 1997) to detect and react to external stimuli. These systems are critical for interaction with neighboring organisms and for coordinating processes of microbial infection. The previous section on nodule development mentioned several two-component signaling systems involved in the development of symbiotic nitrogen fixation.

Two-component signaling systems are comprised of at least a homodimeric histidine kinase (SK), responsible for sensing a stimulus, and a cognate response regulator (RR) responsible for activating downstream target genes in response to SK activation (J. S. Parkinson, 1993). Common structural and physical properties of the SK and RR determine the specificity and sensitivity of the signaling system, and will be reviewed in detail in the following sections.

Though two-component signaling systems have been extensively studied, and the molecular mechanisms of signaling inside the cytoplasm have been well characterized, there are still a number of questions concerning stimulus recognition and signal transmission across the

membrane into the microbial cell. For a great many sensor kinases, the nature of the stimulus required for activating the system remains mysterious. The mechanism by which external signals are transmitted across the cell membrane into the cytoplasm is also as yet not entirely clear.

In some cases, a two-component signaling system may be further regulated by one or more additional auxiliary proteins, allowing for a more finely tuned response to an external stimulus. These regulator proteins have been found in all cellular compartments, and may modulate signaling by interaction with the SK or the RR. Chapter 2 of this dissertation investigates the regulation of a two-component system by a third periplasmic protein in an effort to better understand how the sensing of external stimuli is converted to produce changes in gene expression within a microbial cell.

1.3.2 Sensor kinases

Sensor kinases are responsible for sensing a stimulus and relaying this information into the cell. An SK is typically made up of a sensing domain, a dimerization domain, and an ATP-binding domain (Stock, Robinson, & Goudreau, 2000). SKs may also contain additional structural elements involved in amplification or transmission of an associated signal.

Sensor kinases must contain a conserved His residue, and may exhibit three different types of activities – autophosphorylation, phosphotransfer to a cognate RR (both dependent on the conserved His residue), and phosphatase activity of a cognate RR. The steady state phosphorylation level of the RR is the result of the combined effects of these three activities (Salazar & Laub, 2015). The three activities of a sensor kinase are described in detail below in section 1.3.2.2

Studies attempting to understand the mechanisms of action for the three sensor kinase activities have in large part focused on how sensor kinase structures give rise to these functions and coordinate transcription of target genes in response to stimulus recognition.

1.3.2.1 Sensor kinase structural elements

SKs show wide structural variability. Some SKs are free-floating or membrane-bound, but the majority of SKs are transmembrane proteins (West & Stock, 2001). Sensor kinases have typical features that may be combined to facilitate the detection of stimuli and the amplification and transmission of this signal into cytoplasmic protein expression. Some of the common structural elements of sensor kinases involved in these functions are illustrated in figure 1-6. These elements will be reviewed in detail in the following sub-sections.

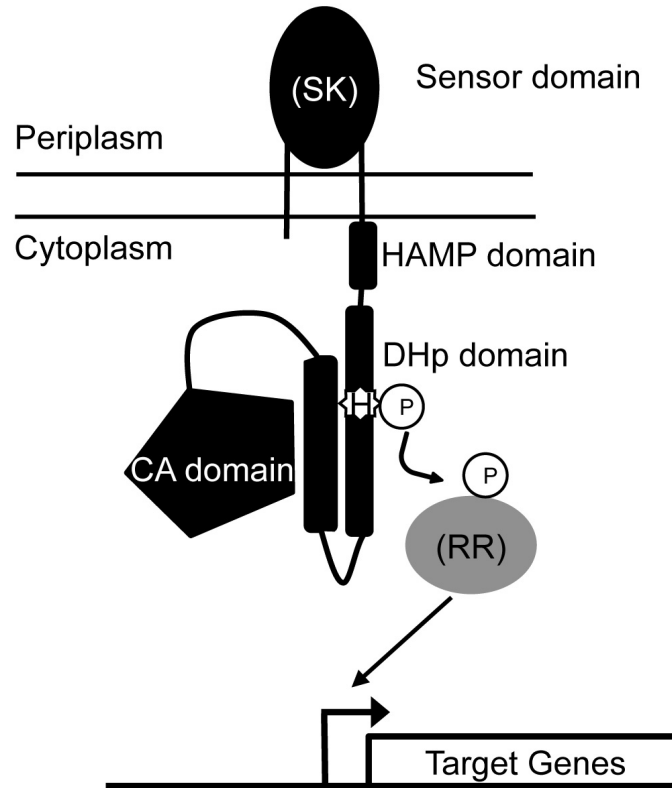


Figure 1-6 Schematic illustration of a typical two-component signaling system. Important structural domains for signal transmission are labeled. Sensor kinases may not have periplasmic sensor domains or HAMP domains, or may contain PAS domains (not shown).

1.3.2.1.1 SK sensing domain

In many cases, an SK sensing domain extends into the periplasm and allows for detection of external stimuli (Stock et al., 2000). While many structural elements of sensor kinases are strongly conserved, the sensing domains are highly variable. This is likely accounted for by the diversity of stimuli detected by different SKs. For many sensor kinases with extra-cytoplasmic sensing domains, the nature of the stimulus and the exact mechanism for stimulus detection and transfer of this signal across the membrane is unclear. Some SK with periplasmic domains may

not actually detect extracellular signals, but rather coordinate responses to internal conditions (Foo, Gao, Zhang, & Kenney, 2015).

1.3.2.1.2 SK HAMP domain

Transmembrane SKs may have a HAMP domain (so designated because of its association with histidine kinases, adenylyl cyclases, methyl accepting proteins, and phosphatases) positioned immediately downstream of the final transmembrane segment. The HAMP domain is a four-helix bundle that is generally responsible for coordinating signal modulation within the cytoplasm, toggling between active and inactive signaling states (Swain & Falke, 2007) (Hulko et al., 2006).

The mechanism by which extracellular signal is relayed from the sensing domain across the membrane to the HAMP domain is not fully understood. However, it is clear that signal transmission can trigger piston-like or rotational movements in the HAMP domain that modulate SK activity (Hulko et al., 2006) (Zhou, Ames, & Parkinson, 2011). Mutational analysis of the HAMP domain has shown that changes potentially over-stabilizing or under-stabilizing the structure of the HAMP domain often result in locked-on (kinase) or locked-off (phosphatase) activities (Ames, Zhou, & Parkinson, 2008). This suggests that functional HAMP domain signal transmission is under biphasic control, dependent on the ability to shift the helices of the domain between tight and loose formations (Manson, 2011; Q. Zhou et al., 2011).

The few residues connecting the HAMP domain to the transmembrane section of an SK, the so-called “control cable” may also play a role in transmitting this signal, possibly by applying physical tension to the HAMP domain that effects the stability of the four helix bundle (Kitanovic, Ames, & Parkinson, 2011).

1.3.2.1.3 SK DHp and CA Transmitter Domains

The cytoplasmic transmitter domains of SKs are involved in dimerization and signal transmission to the RR, and are much more tightly conserved (J. S. Parkinson & Kofoid, 1992; West & Stock, 2001). The cytoplasmic domains can be separated into two major functional regions: the dimerization and histidine phosphorylation (DHp) domain and the catalytic and ATP-binding (CA) domain (Jung, Fried, Behr, & Heermann, 2012).

The DHp domain includes the conserved His residue and is instrumental in dimerization, substrate autophosphorylation, and phosphotransfer. During SK kinase activity, this domain participates in an ATP-dependent autophosphorylation reaction at the conserved His residue. This phosphate can then be transferred to a conserved Asp residue on the cognate RR. The DHp domain is also the site of phosphatase activity. The conserved His residue may play a role in phosphatase activity in some systems (Hsing & Silhavy, 1997), but many systems have shown the phosphatase activity of an SK to be independent of this residue (Freeman, Lilley, & Bassler, 2000; VanYperen, Orton, & Griffiths, 2015). The exact role of the His residue in phosphatase activity is therefore unresolved.

The CA domain binds an ATP molecule and catalyzes autophosphorylation of the SK, positioning the γ -phosphate of ATP for attack by the conserved His residue of the DHp domain (Casino, Rubio, & Marina, 2010). The autophosphorylated SK positions the phosphate for subsequent reaction with the cognate RR. The CA domain also plays a role in phosphatase activity (Hsing, Russo, Bernd, & Silhavy, 1998) (Huynh & Stewart, 2011). Although the phosphatase reaction is not a reverse reaction of the phosphotransfer reaction, evidence has shown that the phosphatase activity requires ATP or ADP (Hsing et al., 1998). It may be that the

CA domain bound to ATP or ADP plays a role in structure stabilization that allows for the alignment of phosphorylated RR in position for the phosphate hydrolysis reaction.

1.3.2.2 Sensor kinase catalytic activities

The structural elements of the SK dictate the dynamics of three potential activities of the SK – autophosphorylation, phosphotransfer to the RR, and phosphatase activity of the RR. The net result of these reactions controls the activation state of the response regulator, and therefore the transcription of downstream target genes. The following sections describe how the functional domains of the SK work together to activate the catalytic activity of the sensor molecule.

1.3.2.2.1 Autophosphorylation activity

After stimulus recognition, the CA domain of the SK binds ATP, positioning the γ -phosphate in place to be transferred to the conserved His residue of the DHp domain (Ashenberg, Keating, & Laub, 2013). The autophosphorylation reaction may be a *trans* reaction (each subunit of the homodimer phosphorylating the DHp of the other subunit; e.g. EnvZ (Y. Yang & Inouye, 1991)) or a *cis* reaction (each subunit of the homodimer phosphorylating its own DHp subunit; e.g. PhoR (Casino, Rubio, & Marina, 2009)).

The autophosphorylation reaction is independent of phosphotransfer and phosphatase activities. In some studies it has been observed that loss of autophosphorylation activity results in a net increase in phosphatase reactions (Hsing et al., 1998).

1.3.2.2.2 Phosphotransfer activity

Phosphotransfer activity refers to the transfer of phosphate from the SK conserved His residue to the conserved Asp residue of the RR, and necessarily depends on autophosphorylation.

This reaction is initiated by the RR and is discussed further in section 1.3.3. For the purposes of this dissertation, the term kinase activity will refer to the combined autophosphorylation and phosphotransfer activities.

1.3.2.2.3 Phosphatase activity

Some sensor kinases have the ability to remove phosphate from their cognate RRs. This provides an additional means of regulation and allows for a quick reset of the signaling system under phosphatase conditions (Grebe & Stock, 1999; Huynh & Stewart, 2011). Though the conserved His residue may be involved in enhancing phosphatase activity (Hsing et al., 1998; Hsing & Silhavy, 1997), the phosphatase reaction is not a reverse reaction of the phosphotransfer reaction. The reaction is likely similar to that facilitated by known phosphatase enzymes (Huynh & Stewart, 2011), and involves a hydrolysis reaction (Casino et al., 2010) in which an activated water molecule in the active site attacks the phosphoryl group of the RR (Pioszak & Ninfa, 2004).

The highly characterized sensor kinase EnvZ has been shown to have phosphatase activity, but the low level of affinity for OmpR ~ P and small level of observed phosphatase reactions has led to conclusions that, at least in the case of EnvZ, the phosphatase activity of the SK is perhaps not biologically relevant (Kenney, 2010; King & Kenney, 2007; Mattison & Kenney, 2002). However, it should be noted that defects in phosphatase activity have been shown to result in aberrant phenotypes (Russo & Silhavy, 1993).

It has been suggested that SK phosphatase activity might be an important regulatory mechanism to prevent cross-talk or incidental phosphorylation of RRs which might otherwise errantly activate their systems (Groban, Clarke, Salis, Miller, & Voigt, 2009; Siryaporn &

Goulian, 2008). In fact, studies have shown that the RRs CheY and PhoB can be readily phosphorylated by acetyl phosphate or non-cognate SKs (McCleary, Stock, & Ninfa, 1993; Wanner, 1992). However, this activation is observed to be dependent upon the absence of the cognate SK. The phosphatase activity of the cognate SK therefore appears to be instrumental in regulating this activation and maintaining appropriate signal levels within the cell (Gao & Stock, 2013). A number of auxiliary regulators of two-component signaling appear to act by stimulating SK phosphatase activity, further suggesting the biological importance for a mechanism for quickly suppressing the production of downstream target genes.

1.3.3 Response regulator proteins

Response regulator proteins are a class of transcriptional activators that typically serve as relay proteins connecting sensor kinase signal recognition with transcription of downstream target genes. The phosphorylation state of the RR determines the level of downstream gene expression. Response regulators catalyze phosphotransfer from the conserved His of an SK to the conserved Asp residue of the RR via nucleophilic attack (Lukat, McCleary, Stock, & Stock, 1992). As mentioned previously, RRs are capable of similarly catalyzing phosphorylation with non-cognate SKs or other small molecules like acetyl-CoA (McCleary et al., 1993) in the absence of the cognate SK. This introduces an interesting question of RR specificity. In contrast to eukaryotic signaling systems, crosstalk between non-cognate SKs and RRs appears relatively rare in nature (M. A. Rowland & Deeds, 2014). While crosstalk between SKs and RRs can be measured *in vitro*, in practice the rate of phosphorylation between an SK and its cognate RR is significantly higher than any possible incidental cross-talk phosphorylation, and therefore is not likely to play a major biological role (West & Stock, 2001).

1.3.3.1 Response regulator half life

Two component signal systems have evolved to allow for timely responses in the case of rapid environmental changes. Controlled genes are often maintained at low levels when the system is not activated, and are shut down again quickly when no longer needed. To that end, RRs have evolved auto-phosphatase activity, with short phosphorylation half-lives of minutes to hours (Bourret, 2010) allowing for rapid decrease in downstream gene expression when the signal is no longer active. If needed, the half-life of an RR can be further shortened by activation of SK phosphatase activity, allowing for rapid reset of the system.

1.3.3.2 Activation of target genes

Response regulators typically exist in one of two states – active or inactive. In most cases, RR phosphorylation results in a conformational change to an active state - allowing for interaction with DNA and the activation of downstream target gene transcription (Bourret, 2010; Stock et al., 2000). Non-phosphorylated RRs are typically inactive and do not interact with DNA.

1.3.4 SK-RR Specificity

Figure 1-7 shows the crystal structure of a sample SK-RR interaction. Structural studies suggest that the RR (Shown in Figure 1-7 as an orange ribbon structure) interacts with helix $\alpha 1$ of the DHp domain beneath the conserved His residue (Casino et al., 2009, 2010). This interaction positions the RR optimally for phosphotransfer to the conserved Asp residue. System level studies have shown that SKs will phosphorylate and dephosphorylate their cognate RRs at a much higher level than non-cognate substrates (Skerker, Prasol, Perchuk, Biondi, & Laub, 2005) (Salazar & Laub, 2015), implying some stringency of specificity. Amino acid covariance analysis between SKs and their cognate RRs revealed specific regions of the DHp domain of the SK and a region of the RR predicted to be involved in specificity between cognate pairs (Skerker et al., 2008). This study went on to show that substituting amino acids at these residues was sufficient to re-wire the substrate specificity of the SK EnvZ. Scanning mutation analysis of these regions has further confirmed their role in determining SK specificity (Capra et al., 2010). In these studies, substitutions in the RR OmpR were also sufficient to rewire specificity, losing interaction with EnvZ and gaining compatibility with the non-cognate SK CpxA. It seems clear that cognate pairs have co-evolved an adequately stringent level of specificity to allow for just

the right level of control over downstream target genes allowing for timely and appropriate responses to changes in environmental conditions.

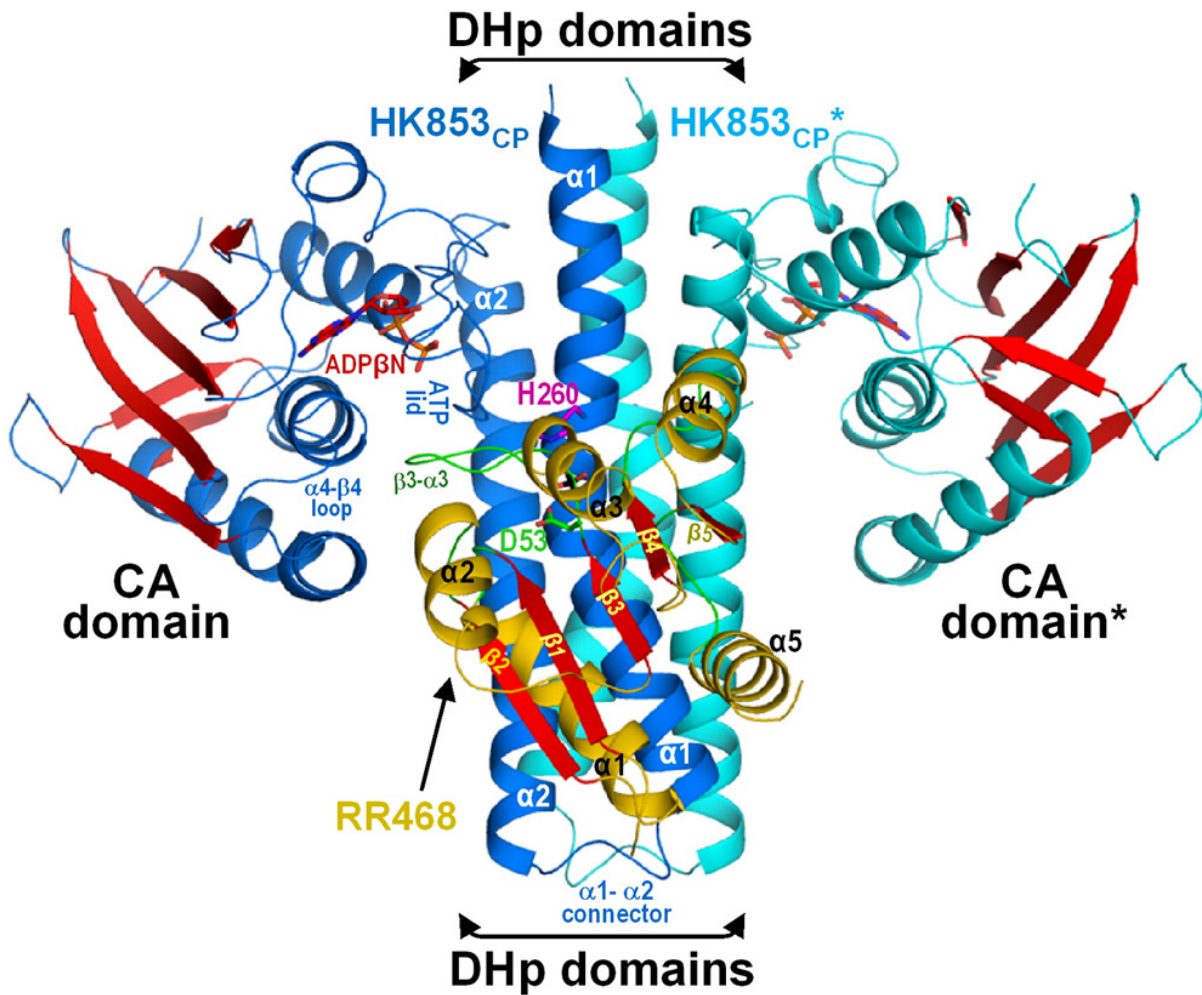


Figure 1-7 Crystal Structure of an HK-RR Complex
 Crystal structure of the cytoplasmic domain of HK853 - RR468 complex with contact residues indicated. The conserved SK histidine and RR aspartate residues are highlighted. Reprinted with permission from (Casino et al., 2009)

1.3.5 Auxiliary regulator proteins of two-component signaling

A relatively new class of auxiliary regulator proteins has been discovered which allows for modification of the signaling state of a TCS to provide an additional level of control over the microbial response to environmental cues. These regulators have been shown to exercise control over TCSs in many different ways, including direct interactions with the RR or modulation of the activity of the SK (Buelow & Raivio, 2010; Mitrophanov & Groisman, 2008).

Chapter two of this dissertation reports the analysis of a novel *Sinorhizobium meliloti* periplasmic regulator protein FeuN on the SK FeuQ. For the purposes of this review, I will therefore focus on regulators of TCSs that act by changing the behavior of a sensor kinase. Representatives of these regulators have been shown to act by modifying one or more of the three SK reactions (autophosphorylation, phosphotransfer, or phosphatase). While some regulator proteins can enhance the signal activation of the SK (Ishii, Eguchi, & Utsumi, 2013), this review will focus on negative regulation of two-component signaling. Table 1-1 at the end of this section provides a summary of these auxiliary regulators.

1.3.5.1 Inhibition of autophosphorylation

Negative regulation of a TCS can be initiated by interference with the autophosphorylation activity of the sensor kinase. If autophosphorylation is selectively inhibited while phosphatase activity remains intact, the net result is an increase in phosphatase activity and a reset of the TCS signaling pathway. One of the earliest identified two-component regulator proteins, the *Bacillus subtilis* KipI, regulates the signaling of the SK kinase A in this manner (L. Wang, Grau, Perego, & Hoch, 1997). KipI regulation of kinase A plays a role in the highly regulated process of *B. subtilis* sporulation, highlighting the importance of these additional layers

of control for mounting an accurate response to the demands of a changing environment. KipI interacts with the ATP-binding region of the CA domain, thereby blocking autophosphorylation reactions without interfering with the residues involved in the phosphotransfer or phosphatase reactions.

1.3.5.2 Inhibition of phosphotransfer

Negative regulation of SK signaling can alternatively be achieved by inhibition of phosphotransfer to the response regulator by the DHp domain. One of the early representatives of this class, the *B. subtilis* protein Sda, works on the previously mentioned SK kinase A to provide yet another layer of control over the delicate regulation of sporulation. Sda binds to the DHp domain to prevent phosphotransfer by steric hindrance (Cunningham & Burkholder, 2009); (Bick et al., 2009). This interaction likely also inhibits or at least has an effect on phosphatase activity due to the close proximity or even overlap of the residues involved in the phosphotransfer and phosphatase reactions.

The *Vibrio cholera* regulator VieB similarly binds to the DHp domain of the SK VieS, preventing its autophosphorylation and phosphotransfer activity, allowing for precision in the control of downstream genes required for biofilm formation and pathogenesis (Mitchell, Ismail, Kenrick, & Camilli, 2015).

1.3.5.3 Direct stimulation of phosphatase activity

Auxiliary proteins have also been shown to directly stimulate the phosphatase activity of a sensor kinase, increasing the rate of dephosphorylation of the response regulator and quickly eliminating TCS signaling. The *Escherichia coli* protein PII is perhaps the most well characterized example of this activity. Direct phosphatase assays have shown that PII interacts

with the SK NRII (NtrB) to increase phosphatase activity on the RR NRI (NtrC) (Pioszak & Ninfa, 2003a, 2003b).

PII interacts with the C-terminal CA domain of NRII, inducing a conformational change that inhibits autophosphorylation and increases phosphatase activity. Mutational analysis suggested this interaction requires the central DHp and PII binding site in one subunit and the ATP-lid of the CA domain of the second subunit of the SK. Interestingly, additional mutational analysis on the RR NRI reveals that NRI plays an active role in the phosphatase reaction, suggesting that phosphatase activity may be catalyzed by response regulators and is a collaborative reaction (Pioszak & Ninfa, 2004).

The *S. meliloti* protein FixT was also demonstrated to stimulate phosphatase activity of a TCS (Garnerone, Cabanes, Foussard, Boistard, & Batut, 1999). The FixL/FixJ TCS was mentioned previously for its involvement in symbiotic nitrogen fixation. Fix T interaction with FixL prevents autophosphorylation and stimulates the phosphatase reaction between FixL and FixJ. The auxiliary FixT protein appears to stimulate phosphatase activity of FixL by interaction with the SK CA domain.

The *Staphylococcus aureus* proteins SaeP and SaeQ were demonstrated to work as a protein complex to negatively regulate the TCS SaeS/SaeR signaling by stimulating the phosphatase reaction of the SK SaeS. Evidence suggests interaction in the cytoplasm between the SK SaeS and the SaePQ complex gives increased access to key amino acids for phosphatase activity. Interestingly, this interaction does not disable autophosphorylation activity, but does result in a net increase in phosphatase activity (Jeong et al., 2012). This regulation allows for increased sensitivity in the regulation of pathogenesis related genes.

1.3.5.4 Periplasmic regulators of TCSs

The previous examples include mechanisms by which cytoplasmic proteins have been shown to negatively regulate SK signaling. The majority of these proteins negatively regulate signaling through interactions with the CA and DHP domains, in which the role for toggling the active state of an SK have already been well established. Interestingly, a number of regulator proteins have been discovered that somehow modulate cytoplasmic SK signaling from inside the periplasmic space, requiring the signal to be transmitted across the cell membrane. Mechanisms for signal disruption by auxiliary regulators from the periplasm are not well understood. These systems provide convenient models for learning more about how the signal is transmitted from the extra-cellular space into the cytoplasm to influence gene transcription.

The *E. coli* two-component system CpxA/CpxR includes the periplasmic regulator protein CpxP (Raivio, Popkin, & Silhavy, 1999). Evidence suggests CpxP negatively regulates CpxA/CpxR signaling through interaction with the periplasmic sensing domain of CpxA (Raivio, Laird, Joly, & Silhavy, 2000). This interaction appears to somehow block autophosphorylation of CpxA (Fleischer, Heermann, Jung, & Hunke, 2007). Structural analysis has highlighted compatibly charged residues of CpxA and CpxP that likely interact to block CpxA signaling. Evidence suggests that the SK CpxA is likely bound to CpxP and downstream genes are off, but environmentally induced displacement of CpxP due to cell membrane stress activates CpxA/CpxR signaling (X. Zhou et al., 2011).

The *E. coli* ZraS/ZraR TCS that regulates genes involved in Zn scavenging is modified by the periplasmic regulator ZraP in the same manner. In this case, zinc-bound ZraP negatively regulates the SK ZraS by interaction with the sensing domain in the periplasm (Petit-Härtlein et

al., 2015). In the absence of Zn, ZraP interaction with ZraS is lost and ZraS/ZraR signaling is activated. ZraS is thought to negatively regulate ZraP from the periplasm in a manner similar to CpxP inhibition of CpxA described previously (Petit-Härtlein et al., 2015).

The *Vibrio harveyi* periplasmic binding protein LuxP represses the kinase activity of the transmembrane sensor kinase LuxQ as part of the regulation that governs quorum sensing. In this particular system, researchers have been able to observe direct interaction between LuxQ and LuxP with crystal structure analysis. LuxP interaction with LuxQ has been shown to induce conformational changes resulting in the asymmetrical alignment of LuxPQ monomers (Neiditch et al., 2006). This misalignment makes autophosphorylation and phosphotransfer impossible, and quickly shuts down transcription of downstream target genes.

Auxiliary regulation of TCS has also been shown to be important in the regulation of genes involved in symbiotic nitrogen fixation. In *S. meliloti*, the periplasmic auxiliary protein ExoR regulates the two-component system ExoS/ChvI (Chen et al., 2008). The ExoS/ChvI TCS plays a role in the regulation of succinoglycan and is required for successful initiation and elongation of infection threads during the process of nodulation. The mechanism of ExoR regulation of ExoS is not understood.

Table 1-1 Summary table of auxiliary TCS regulators

*kinase activity refers to combined autophosphorylation and phosphotransfer activities

Organism	SK	Auxiliary protein	Location of interaction	Mechanism for negative regulation	Citation
<i>B. subtilis</i>	Kinase A	Kipl	cytoplasm	Blocks SK autophosphorylation	(L. Wang, Grau, Perego, & Hoch, 1997)
<i>B. subtilis</i>	Kinase A	Sda	cytoplasm	Blocks SK phosphotransfer to RR	(Cunningham & Burkholder, 2009; Bick et al., 2009)
<i>V. cholera</i>	VicS	VicB	cytoplasm	Blocks SK kinase* activity	(Mitchell, Ismail, Kenrick, & Camilli, 2015)
<i>E. coli</i>	NRII (NtrB)	PII	cytoplasm	Block SK autophosphorylation and stimulates phosphatase activity	(Pioszak & Ninfa, 2003a, 2003b)
<i>S. meliloti</i>	FixL	FixT	cytoplasm	Block SK autophosphorylation and stimulates phosphatase activity	(Gamerone, Cabanes, Foussard, Boistard, & Batut, 1999)
<i>S. aureus</i>	SacS	SacPQ	cytoplasm	Stimulates phosphatase activity	(Jeong et al., 2012)
<i>E. coli</i>	CpxA	CpxP	periplasm	Inhibits SK autophosphorylation	(Fleischer, Heermann, Jung, & Hunke, 2007)
<i>E. coli</i>	ZraS	ZraP	periplasm	Thought to inhibit SK autophosphorylation	(Petit-Hrtlein et al., 2015)
<i>V. harveyi</i>	LuxQ	LuxP	periplasm	Inhibits SK autophosphorylation	(Neiditch et al., 2006)
<i>S. meliloti</i>	ExoS	ExoR	periplasm	unknown	(Chen et al., 2008)
<i>S. meliloti</i>	FcuQ	FcuN	periplasm	Stimulates phosphatase activity	(VanYperen et al., 2015)

1.3.6 Section conclusions

Bacteria have evolved two-component signaling systems to recognize and respond to environmental cues. These systems are carefully regulated to coordinate appropriate changes in gene expression without wasting unnecessary resources in the cell. Sensor kinase proteins recognize environmental stimuli, and amplify and report that signal by phosphorylation of a response regulator protein. The phosphorylated response regulator proteins then coordinate a response by initiating associated gene transcription. Additional auxiliary regulator proteins may influence the activity of the sensor kinase or response regulator, allowing for a carefully organized and precise response to a critical stimulus.

Several two-component signaling systems play a role in the development of nitrogen-fixing nodules in the *S. meliloti* – *M. truncatula* symbiosis. In some cases (e.g. ExoS/ChvI), these TCSs are under additional control by auxiliary proteins. Chapter two of this dissertation reports the examination of yet another *S. meliloti* periplasmic regulator required for infection thread progression. The small periplasmic protein FeuN negatively regulates the FeuQ/FeuP two-component signaling by stimulation of phosphatase activity (VanYperen et al., 2015). The FeuPQN system plays a role in the regulation of cyclic glucan export, which is important for infection thread colonization. Characterization of these systems has provided insight into the molecular determinants of symbiotic compatibility.

1.4 Microbial genetics of rhizosphere colonization

1.4.1 Introduction

The signals required for rhizobia-legume symbiosis begin with microbial secretion of Nod factor after colonization of the legume root systems. These studies have taken rhizobial persistence in the soil and competitive root colonization as a given precursor to Nod factor signaling, yet the genetics of competitive root colonization have not been well characterized. Chapters 3 and 4 of this dissertation report the investigation of genes required for competitive root colonization as the first step in the establishment of rhizobia-legume symbiosis.

The environment surrounding root tissues, including the associated microbial activity is known as the rhizosphere (J. Whipps & Lynch, 1990). The recruitment of microbial communities to the rhizosphere, or “rhizosphere effect” was first described in 1904 (Hiltner, 1904). All plants, including legumes expend a significant amount of resources into the soil as root exudates influencing the properties of the surrounding soil environment. This investment is believed to be an effort to recruit beneficial bacteria to the area and suppress the development of pathogenic communities.

Two major factors determine the diversity and activity of the rhizosphere microbial community – the root plant species and the soil type (Garbeva, Van Elsas, & Van Veen, 2008). The variability in the content of root exudates between different plant species and the differences in soil types (pH, moisture conditions, mineral composition, etc.) in different geographic locations have made identification of requirements for rhizosphere colonization very difficult.

Prior to the newer development of deep sequencing technologies, these difficulties have been further compounded by limitations in the ability to monitor microbial communities within

soil and recover microbial communities back out from the rhizosphere after experimentation (D. Parkinson, Gray, & Williams, 1971) (Kloepper & Beauchamp, 1992) (Lugtenberg, Dekkers, & Bloembergen, 2001). Many of the discovered determinants of competitive rhizosphere colonization have been made in the context of gnotobiotic systems – sterile soil and root systems in which mutant strains are competed against a known competitive rhizosphere colonizer in head to head competition experiments. The inherent limitation of scale in such experiments has been a rate-limiting factor of discovery.

Despite these limitations, researchers have had success in identifying specific microbial genes required for competitive root colonization. Studies on plant growth promoting bacteria (PGPB) (e.g. *Pseudomonas*) and rhizobia have been the most productive in the search for genes involved in competitive root colonization. Some of the emerging themes for rhizosphere competency are summarized in the following sub-sections.

1.4.2 Amino acid synthesis genes

The bioavailability of amino acids can be a limiting factor for rhizosphere colonization. The types and quantities of amino acids available in root exudate and the required levels to support root colonization likely varies depending on the species of both the plant and the bacteria tested. Studies of *Pseudomonas* colonization of tomato roots have shown that amino acid synthesis is required for competitive root colonization (Simons, Permentier, de Weger, Wijffelman, & Lugtenberg, 1997).

Auxotrophs for leucine, arginine, histidine, isoleucine plus valine, and tryptophan were shown to be unable to competitively colonize tomato roots. Interestingly, amino acids are found in tomato secretions into the soil. The researchers concluded that while many amino acids are present in the tomato root exudate, they are not found in sufficient quantities to support *Pseudomonas* colonization. Further tests of *Pseudomonas* on potato root systems has similarly demonstrated that amino acid auxotrophs are not competitive rhizosphere colonizers of potato root systems (Glandorf, 1992). These studies suggest a general requirement for amino acid synthesis genes for competitive root colonization.

1.4.3 Flagella driven chemotaxis

Numerous studies on *Pseudomonas* rhizosphere colonization have established that flagella-driven chemotaxis is also essential for competitive root colonization (Lugtenberg et al., 2001) (de Weert et al., 2002). These studies have highlighted the role of motility in the competitive root colonization of several types of plants, including the previously mentioned tomato and potato root systems. A *Pseudomonas* study on flagellar filament synthesis suggests that competitive root colonization can even be improved by optimizing the motility process

(Capdevila, Martínez-Granero, Sánchez-Contreras, Rivilla, & Martín, 2004). The ability to recognize root exudates and commute to root systems via chemotaxis plays a central role in the enriched recruitment of microbes to the rhizosphere, and is not surprisingly a required feature for competitive root colonization.

1.4.4 Thiamine, biotin, and riboflavin production

Thiamine, biotin, and riboflavin are identified alfalfa root exudates that have been demonstrated to influence the development of rhizosphere communities (Rovira & Harris, 1961). Biotin in particular has been shown to be required for growth by *S. meliloti* (Lowe & Evans, 1962). Studies have shown that in the rhizosphere, availability for these vitamins is limiting, and biotin auxotrophs are not able to effectively compete in the rhizosphere (Streit, Joseph, & Phillips, 1996).

1.4.5 Lipopolysaccharide (LPS)

Plants coordinate defense mechanisms to protect themselves from pathogenic infection. Plants identify harmful microbes by recognition of PAMPs (pathogen-associated molecular patterns). PAMP recognition triggers defense responses including release of reactive oxygen species (ROS) intended to destroy invading microbes (Montesano, Brader, & Palva, 2003).

LPS is a major component of the outer membrane of Gram-negative bacteria and plays a major role in protecting the bacteria from the environment. LPS is a plant-recognized PAMP that elicits strong defense response and the release of ROS to prevent microbial infection. Successful rhizosphere microbes have in some cases adapted LPS structures that suppress defense responses in compatible host plants. For example, *Sinorhizobium meliloti* lipopolysaccharides have been shown to suppress oxidative burst reactions in the host plant *Medicago sativa* (Albus, Baier,

Holst, Pühler, & Niehaus, 2001; Scheidle, Groß, & Niehaus, 2005), allowing for colonization and symbiotic infection. Studies of *Pseudomonas* rhizosphere colonization have also highlighted the important role of LPS in competitive root colonization (Simons et al., 1996).

1.4.6 Iron binding

Iron limitation is often a factor in the competitive colonization of the rhizosphere (J. M. Whipps, 2001), and it appears that the ability to competitively scavenge for iron and other minerals is required for competitive rhizosphere colonization. Studies of rhizobia-legume symbiosis have recognized the importance of the ability to competitively chelate iron for competitive root colonization (C.-H. Yang & Crowley, 2000).

1.4.7 Section conclusions

Despite nearly a century of study, general molecular determinants of rhizosphere colonization are not well understood. Studies on plant growth promoting bacteria and rhizobia have highlighted some emerging themes for root colonization, including the ability to locate and travel to root systems by chemotaxis, the ability to synthesize limited amino acids and vitamins, the ability to competitively scavenge for limited iron and other minerals, and the requirement for LPS to evade or suppress plant immune response.

The requirements for these gene categories may seem somewhat predictable, but requirements for competitive root colonization are not necessarily intuitive. Perhaps surprisingly, studies of *Pseudomonas* colonization of plant roots suggests that the ability to grow on simple root exudate sugars does not play a role in root colonization (Lugtenberg, Kravchenko, & Simons, 1999). It has been proposed that these sugars are not secreted in sufficient amounts to support the growth of microbial communities. This is consistent with similar observations on

amino acid secretions, and reinforces the hypothesis that many plant root exudates have evolved as identifying signatures that recruit bacteria to the area, but are not sufficient in quantities to sustain the growth of these microbial communities.

Efforts to unravel the genetics of rhizosphere colonization have been complicated by the great diversity of plant species, soil types, and microbial species that combine to determine rhizosphere conditions, as well as the difficulty in labeling, recovering, culturing, and analyzing soil microbes. These difficulties have impeded major progress in the identification of determinants for competitive root colonization.

The genetics of competitive root colonization remain a generally mysterious but critical component for the development of symbiotic nitrogen fixation. The development of high-throughput methods for analysis and characterization of complex communities should allow for more rapid progress in the field of rhizosphere genetics. Chapter 3 of this dissertation reports the identification of a number of genes involved in competitive root colonization in the *S. meliloti* – *M. truncatula* model system using Tn-seq analysis – a newly developing technique that combines traditional transposon mutagenesis with deep sequencing to identify gene networks required for a condition of interest. For this study, I have adapted a custom Tn-seq approach to monitor genotypes that are required for root colonization. Chapter 4 begins to expand this analysis with a look at genetic requirements for root colonization between *S. meliloti* and rice. Comparing gene requirements for *S. meliloti* in rhizosphere influenced by *M. truncatula* or rice root systems allows for the identification of both general and host-specific determinants for rhizosphere colonization.

1.5 Conclusions

This dissertation reports an investigation of gene networks involved in competitive rhizosphere colonization and nodulation in the *S. meliloti* – *M. truncatula* symbiosis. Many signals are passed between host and microbe as symbiotic infection progresses. The work presented here examines determinants of host-microbe interactions from two perspectives – a narrowly focused, protein level characterization of a two-component signaling pathway required for nodulation, and a widely focused, genome level identification of gene networks involved in competitive root colonization.

Chapter 2 of this dissertation details the discovery and characterization of a new two-component signaling system required for successful nodulation. This research surveys a molecular contact point for symbiosis at single amino acid resolution to increase understanding of host-microbe signaling and microbial signal transduction during nodulation.

Chapter 3 of this dissertation reports the identification of many genes involved in competitive rhizosphere colonization of *S. meliloti* on *M. truncatula*. Current research characterizing the signals involved in the negotiation of symbiosis has tended to focus on infection steps starting with microbial Nod factor secretion and continuing through infection thread colonization and nodule development. However, before potential symbiotic partners can initiate nodule development with secreted Nod factors, they must first be able to compete for space and establish colonization of the rhizosphere. The genetic requirements for successful rhizosphere colonization are not well understood. This research implicates many previously unsuspected gene networks as critical for *S. meliloti* infection of *M. truncatula* root systems.

Chapter 4 of this work discusses the potential for expanding our analysis by increasing the number of plant and microbe participants in Tn-seq studies. This expanded panel may allow us to identify conserved genes required for general rhizosphere competency, as well as host-specific gene requirements which may yield insight into how specific root systems influence the rhizosphere environment. This research includes preliminary data for rhizosphere colonization requirements for *S. meliloti* on rice root systems.

The combination of a focused mechanistic study of a system required for symbiosis along with a broad survey of gene networks involved in rhizosphere colonization provides a solid foundational perspective on how a large number of complex molecular pathways intersect and combine to dictate the development of host-microbe interactions in the negotiation of symbiotic nitrogen fixation.

Chapter 2. Genetic analysis of signal integration by the *Sinorhizobium meliloti* sensor kinase FeuQ

2.1 Summary

Two-component signaling systems allow bacteria to recognize and respond to diverse environmental stimuli. Auxiliary proteins can provide an additional layer of control to these systems. The *Sinorhizobium meliloti* FeuPQ two-component system is required for symbiotic development and is negatively regulated by the auxiliary small periplasmic protein FeuN. This study explores the mechanistic basis of this regulation. We provide evidence that FeuN directly interacts with the sensor kinase FeuQ. The isolation and characterization of an extensive set of FeuN-insensitive and FeuN-mimicking variants of FeuQ reveal specific FeuQ residues (periplasmic and intracellular) that control the transmission of FeuN-specific signaling information. Similar analysis of the FeuN protein highlights short patches of compatibly charged residues on each protein that likely engage one another, giving rise to the downstream effects on target gene expression. The accumulated evidence suggests that the periplasmic interaction between FeuN and FeuQ introduces an intracellular conformational change in FeuQ, resulting in an increase in its ability to remove phosphate from its cognate response regulator FeuP. These observations underscore the complex manner in which membrane-spanning sensor kinases interface with the extra-cytoplasmic environment and convert that information to changes in intracellular processes.

2.2 Introduction

Bacteria use two-component systems (TCSs) to recognize and respond to a wide variety of external stimuli (Gao & Stock, 2009). Typically, TCSs are comprised of a membrane bound sensor histidine kinase (SK) and its cognate response regulator (RR). For active kinase activity, the SK binds ATP and phosphorylates itself on a conserved histidine residue. The phosphoryl group is then transferred to a conserved aspartate residue in the receiver domain of the cognate RR (Casino et al., 2010; Hoch, 2000; Russo & Silhavy, 1993). The RR often acts as a transcriptional regulator of genes functionally related to the detected stimulus (see Fig. 1-5). The combination of autophosphorylation of the SK and phosphotransfer to the RR will be referred to hereafter as kinase activity. In many cases, the SK can also serve as a phosphatase, facilitating removal of phosphate from the cognate RR (Ninfa & Magasanik, 1986). Depending on the concentration of the external stimulus, an SK can be toggled between activation (kinase-dominant) and inhibition (phosphatase dominant) states.

The manner in which stimuli control the kinase and phosphatase activities of the SK proteins are not well understood and seem to be varied in the broad family of SK proteins. However, biochemical and structural studies of numerous SKs have identified domains that participate in the kinase and phosphatase activities and their control (shown in Fig. 1-5). Many SKs have a periplasmic sensing domain thought to be responsible for stimulus detection. The non-conserved nature of these sensor domains is likely due to the variation in stimuli sensed by different SKs (Pioszak & Ninfa, 2003a). On the cytoplasmic side of the membrane, the SK may have a membrane-proximal HAMP (histidine kinases, adenylyl cyclases, methyl-accepting chemotaxis proteins, and some phosphatases) domain that connects transmembrane sensory inputs to output responses (J. S. Parkinson, 2010). Structural changes in the HAMP domain have

been shown to control the signaling state of SKs (Ferris et al., 2012; J. S. Parkinson, 2010). The cytoplasmic signaling module of an SK is comprised of two additional domains: a dimerization/histidine phosphorylation (DHp) domain and a C-terminal catalytic/ATP-binding (CA) domain (Ferris et al., 2012). Mounting evidence suggests that complex interactions between cytoplasmic HAMP, DHp, and CA domains control the activation or inhibition states of an SK. However, for many SKs, how the periplasmic sensing domain is linked across the membrane to the cytoplasmic control of activation or inhibition states is poorly understood.

Prior to my work on this system, a screen in our lab for *S. meliloti* mutations resulting in abortive nodulation phenotypes demonstrated that a predicted response regulator FeuP is required for successful colonization of infection threads and therefore essential for symbiotic nitrogen fixation (See Figure 2-1) (Griffitts et al., 2008). FeuP is activated by the SK FeuQ.

As a part of that study, microarray analysis revealed that *ndvA*, an ABC transporter required for cyclic glucan export, was strongly controlled by the FeuPQ system. Because cyclic glucan export is known to be required for viable infection threads (Cheng & Walker, 1998) (refer to section 1.2.3.2), it was hypothesized that the symbiotic defect for FeuP deletion was actually a result of eliminated downstream *ndvA* expression. Indeed, independent expression of *ndvA* controlled by the constitutive *P_{trp}* promoter was able to restore a symbiotic phenotype to the Δ *feuP* strain (See Figure 2-2). Within this same study, it was shown that *P_{ndvA}::lacZ* can be used as an effective reporter for FeuPQ signaling. In the course of this study, evidence was also presented that the FeuPQ system is stimulated by low osmolarity, though the actual signal detected by the SK FeuQ in nature is not clear (data not shown).

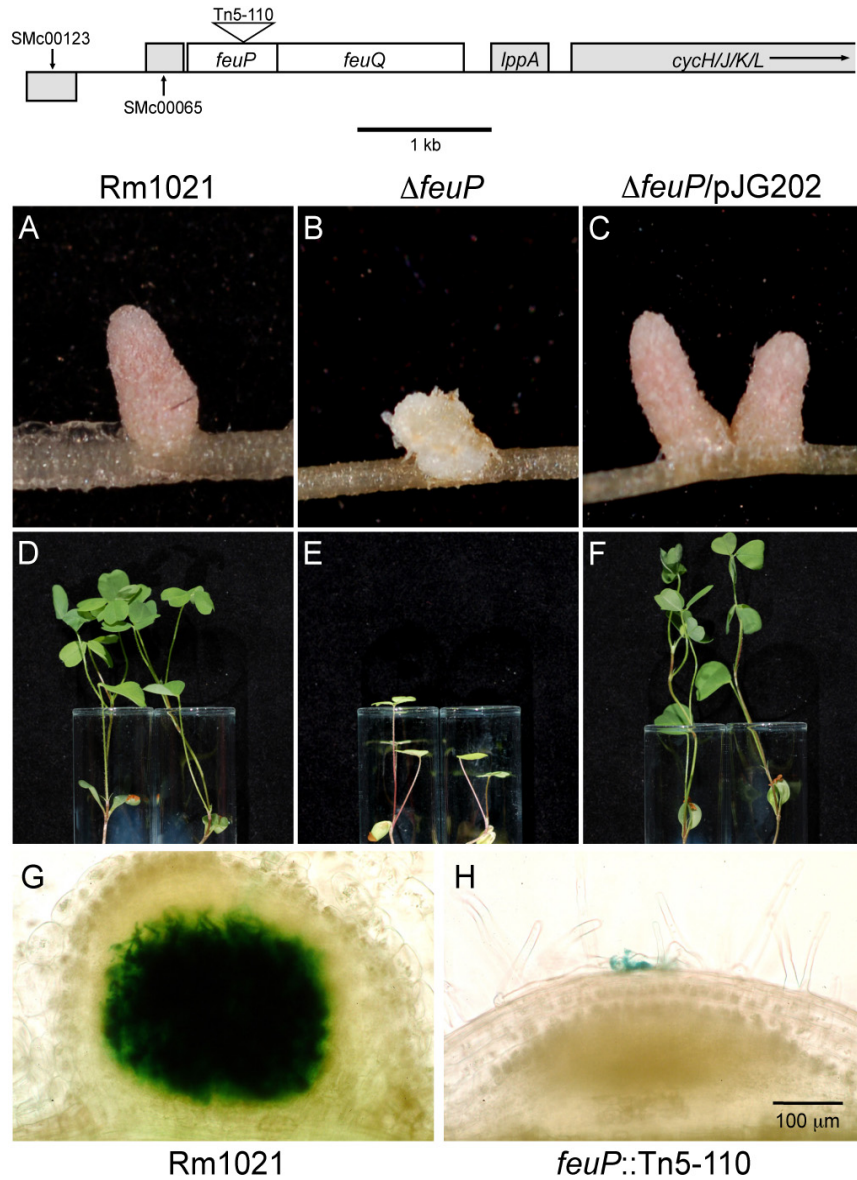


Figure 2-1 The *S. meliloti* RR FeuP is required for efficient nodulation on alfalfa
Top. Arrangement of the FeuPQ TCS genes.

A-C. Representative nodules are shown 30 days after inoculation of alfalfa with the wild-type strain RM1021 (A), the $\Delta feuP$ strain B119 (B), and B119 complemented with the *feuP* expressing plasmid pJG202(C).

D-F. Represented alfalfa shoots grown in the absence of nitrogen for the same strains.

G-H. Microscopic examination of nodule development for wild-type (G) and $\Delta feuP$ strains seven days after inoculation. Figure reprinted with permission from (Griffitts et al., 2008)

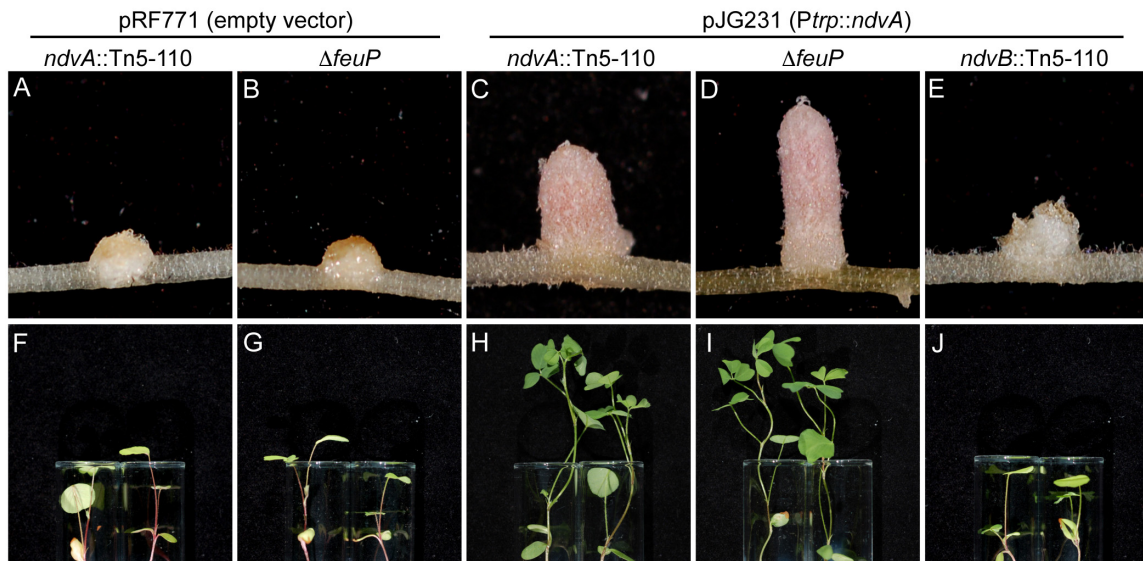


Figure 2-2 Lack of *ndvA* expression accounts for the symbiotic phenotype of the *feuP* mutant
Relative genotype and plasmids harbored by the inoculant bacteria are shown above each pair of images.

A-E. Representative alfalfa nodules 30 days after inoculation.

F-J. Representative alfalfa shoot phenotypes (30 days after inoculation) after growth in nitrogen free medium. Reprinted with permission from (Griffitts et al., 2008)

The FeuPQ regulon consists of at least 15 target genes including *ndvA*. In a recent study a potential FeuP binding site motif was identified along with additional putative gene targets (Schlüter et al., 2013). The FeuPQ regulon is stimulated in certain genetic backgrounds with defects relating to cell polarity and division control (Fields et al., 2012; Pini et al., 2013). Most recently, the FeuPQ regulon has been found to be induced by several cationic antimicrobial peptides including a nodule-specific cysteine-rich (NCR) peptide that is thought to stimulate rhizobial differentiation during symbiosis with host plants (Penterman et al., 2014).

An emerging theme in two-component system research is that of complex regulatory architectures involving protein functions beyond a simple histidine kinase/response regulator pair. An increasing number of small (<200-aa) auxiliary regulator proteins have been discovered which allow for more precise regulation of TCSs. These auxiliary proteins have been identified

in all cellular compartments, and their mechanisms of action vary from one system to another (Buelow & Raivio, 2010). Auxiliary proteins have been shown to link multiple TCSs to one another, to interrupt stimulus detection by the SK, or to modulate SK autophosphorylation, phosphotransfer, or phosphatase activities (Chen et al., 2008; Garnerone et al., 1999; Gerken, Charlson, Cicirelli, Kenney, & Misra, 2009; Jeong et al., 2012; Mitrophanov & Groisman, 2008; Neiditch et al., 2006; Pioszak & Ninfa, 2003b; Szurmant, Bu, Brooks, & Hoch, 2008).

Examination of the genetic region encoding FeuP and FeuQ indicates that a third open reading frame (ORF) designated SMC00065 (called *feuN* hereafter), may be co-transcribed with *feuP* and *feuQ* (Figure 2-3). Interestingly, multiple alphaproteobacterial genera, including the pathogens *Agrobacterium*, *Brucella*, and *Bartonella*, have *feuN*-like ORFs that are conserved and also located adjacent to and upstream of their respective orthologues of *feuP* and *feuQ* (Figure 2-3C). Based on these observations, we hypothesized that *feuN* may be functionally connected to the FeuPQ two-component system.

Attempts to mutate *feuN* were only successful in strains expressing *feuN* on a covering plasmid, suggesting an essential role for *feuN* in *S. meliloti*. Considering that *feuP* and *feuQ* are not essential for cell viability (Griffitts et al., 2008), we tested the possibility that the FeuN protein acts as a negative regulator of signaling through FeuQ and FeuP, such that the lethality associated with the loss of the *feuN* gene arises from over-activity of the FeuPQ signaling system. According to this model, loss of either *feuQ* or *feuP* would suppress the lethal phenotype associated with the loss of *feuN*. As predicted, *feuN* is easily deleted in strains lacking *feuQ* or *feuP* (Griffitts et al., 2008). The mechanism by which FeuPQ over-signaling leads to lethality in *S. meliloti* is still unknown.

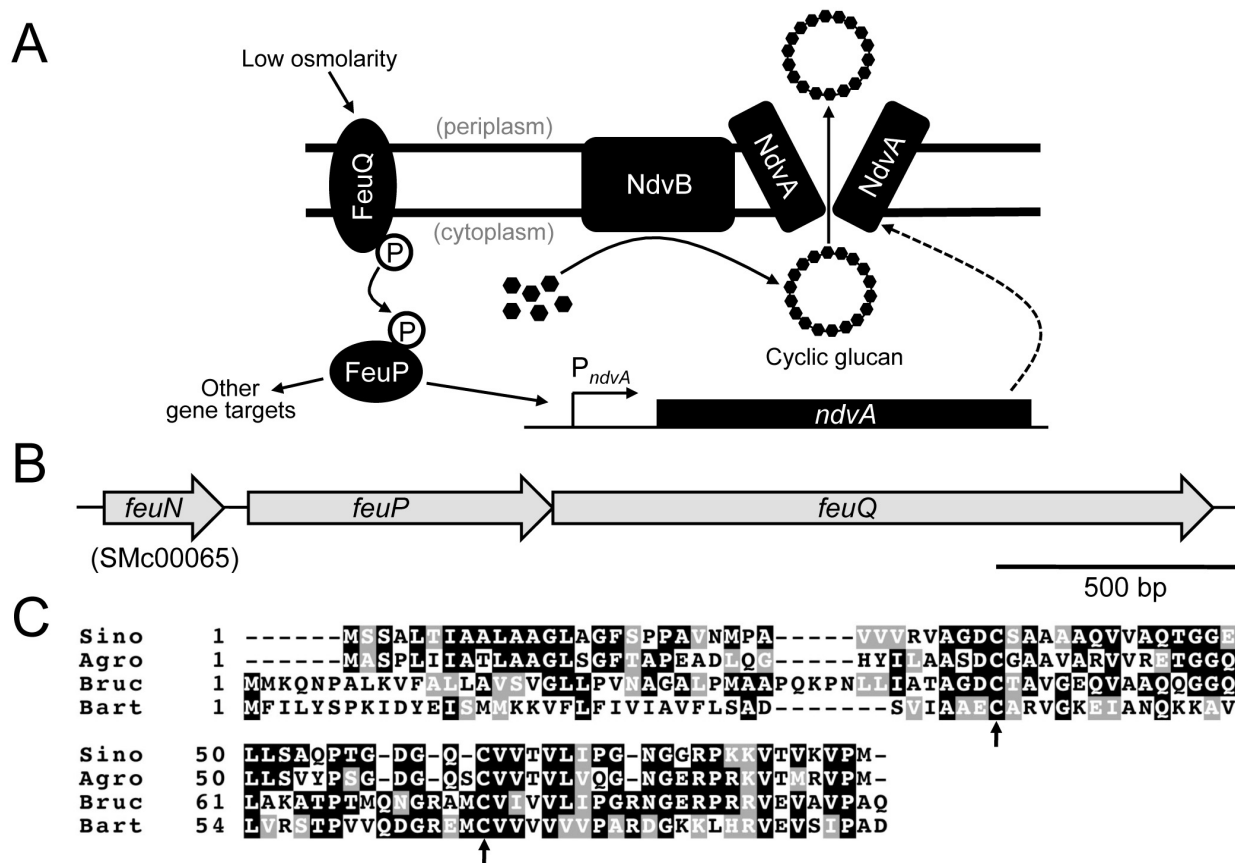


Figure 2-3 *feuN* encodes a conserved alphaproteobacterial protein

A. Model for FeuPQ involvement in cyclic glucan secretion.

B. Genetic position of *feuN* with respect to *feuP* and *feuQ*.

C. Alignment of the FeuN polypeptide from *S. meliloti* (Sino) with orthologues encoded in *Agrobacterium tumefaciens* (Agro), *Brucella melitensis* (Bruc), and *Bartonella quintana* (Bart). Arrows indicate conserved cysteine residues.

The requirement for functional *feuN* in *S. meliloti* is a hindrance to genetic manipulation in this native system. Fortunately, FeuP, FeuQ, and FeuN function heterologously in *E. coli*, allowing for a full genetic analysis using the *PndvA::lacZ* reporter gene (see Figure 2-4). In our plasmid based heterologous system, we observe that FeuPQ signaling activates transcription of the *PndvA::lacZ* reporter, and arabinose-induced expression of FeuN negatively regulates FeuPQ signaling (first 4 bars of Figure 2-4A). This data is confirmed on X-gal plates, where FeuPQ expression results in dark blue colonies, and FeuPQN expression results in white colonies

(Figure 2-4B). The data establish our heterologous system as an effective model to study FeuPQ signal regulation.

The experiments presented in figure 2-4 allow for a number of additional observations important for this study. First, we see that FeuN has no significant negative regulatory effect on FeuP signaling when FeuQ is absent (last 8 bars of Figure 2-4A). Therefore we hypothesize that FeuN negative regulation requires the cooperation of FeuQ to regulate downstream gene expression. We also see that FeuP has measurable activity in the absence of FeuQ, (Figure 2-4A, bars 5-6 and Figure 2-4B, panels B3 and B3) and that all FeuP activity is eliminated by mutation at the conserved FeuP Asp-51 residue (Figure 2-4A bars 7-8). This suggests that FeuP activation is dependent on phosphorylation, and that FeuP can be phosphorylated by other means when the SK FeuQ is absent. This data is consistent with previously discussed studies on TCS cross-talk and incidental RR phosphorylation.

Lastly, it is evident that when acting with FeuN, FeuQ has an active negative regulatory activity on FeuP, decreasing reporter gene activity to a level below that which is observed when FeuQ is absent (Figure 2-4, compare bar 4 to bars 5-6). This phenomenon is also seen on X-gal plates, where FeuP expression in the absence of FeuQ results in an intermediate, light blue phenotype, while expression of FeuPQN results in white colonies (Figure 2-4B, compare panels B3/B4 to B2). The data suggest that FeuQ is a bi-functional sensor kinase, and supports a model in which FeuN activity stimulates FeuQ phosphatase activity on FeuP.

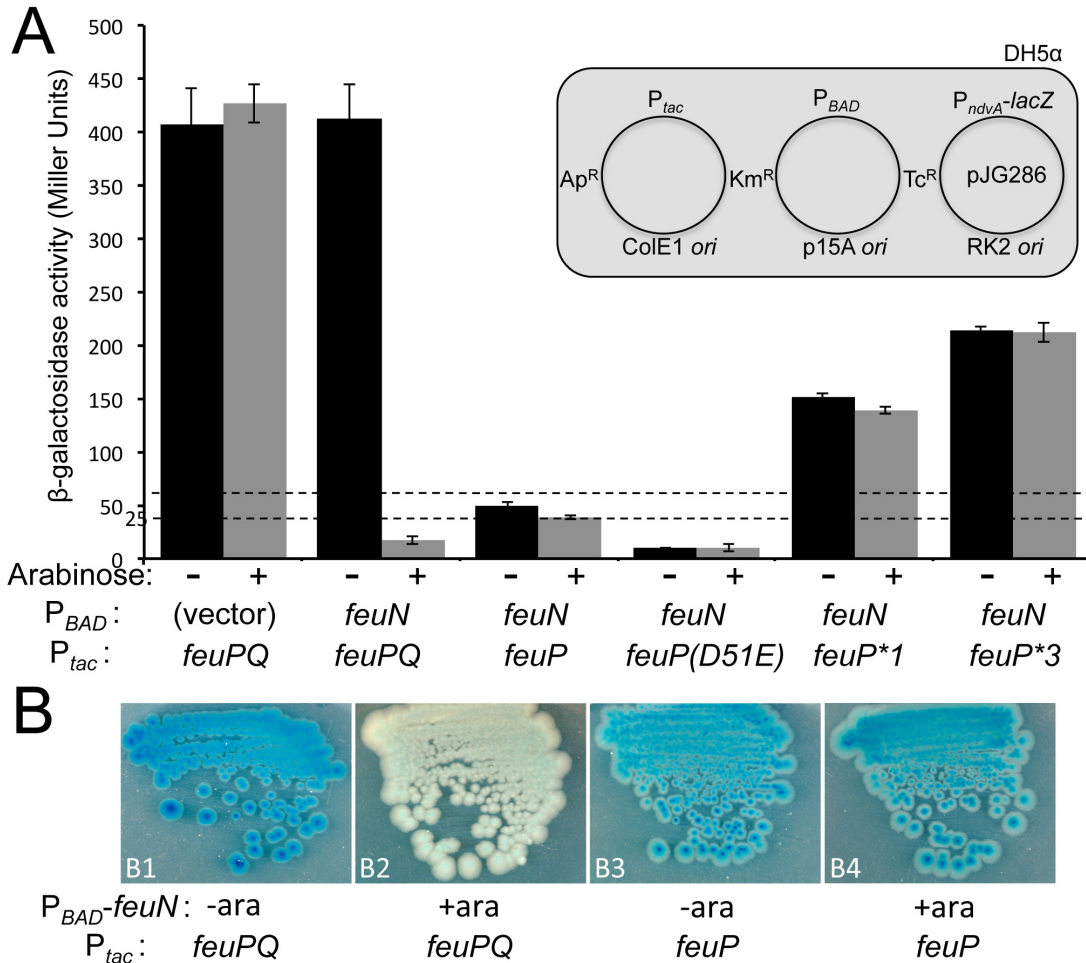


Figure 2-4 FeuN acts directly on FeuQ-FeuP in a heterologous system

A. Inset illustrates the general arrangement of plasmids in tested strains. P_{BAD} expression of plasmids include pJG351 (vector) and pJG355 (*feuN*); P_{tac} expression plasmids include pJG326 (*FeuPQ*), pJG329 (*feuP*) pJG332 [*feuP(D51E)*], pJG406 (*feuP*1*) and pJG408 (*feuP*3*). Grid lines have been added for easier discrimination of lower values.

B. *E. coli* strains were patched onto LB agar supplemented with X-gal and allowed to grow at 30°C for 2 days. All four strains harbor pJG286 and pJG355. The strain showed in B1 and B2 harbor pJG326, while the strain shown in B3 and B4 harbors pJG329. X-gal reaction is shown in the presence (+ara) and absence (-ara) of L-arabinose. Note that IPTG inducer of P_{tac} is not used.

The data suggest that FeuQ activation of FeuP is dependent on the conserved phosphorylation site of FeuP. I wondered if FeuQ negative regulation of FeuP was also dependent on the phosphorylation state of FeuP, or if FeuN/Q inhibition of FeuP might be phosphorylation-independent. To test this idea, constitutively active, non-phosphorylatable alleles of FeuP were co-expressed with FeuQ, in the presence and absence of FeuN (Figure 2-5). The results indicate that FeuN/Q inhibition does not occur with non-phosphorylatable alleles of *feuP*, again consistent with a model in which FeuQ is a bi-functional SK with FeuN either inhibiting its kinase activity or stimulating its phosphatase activity.

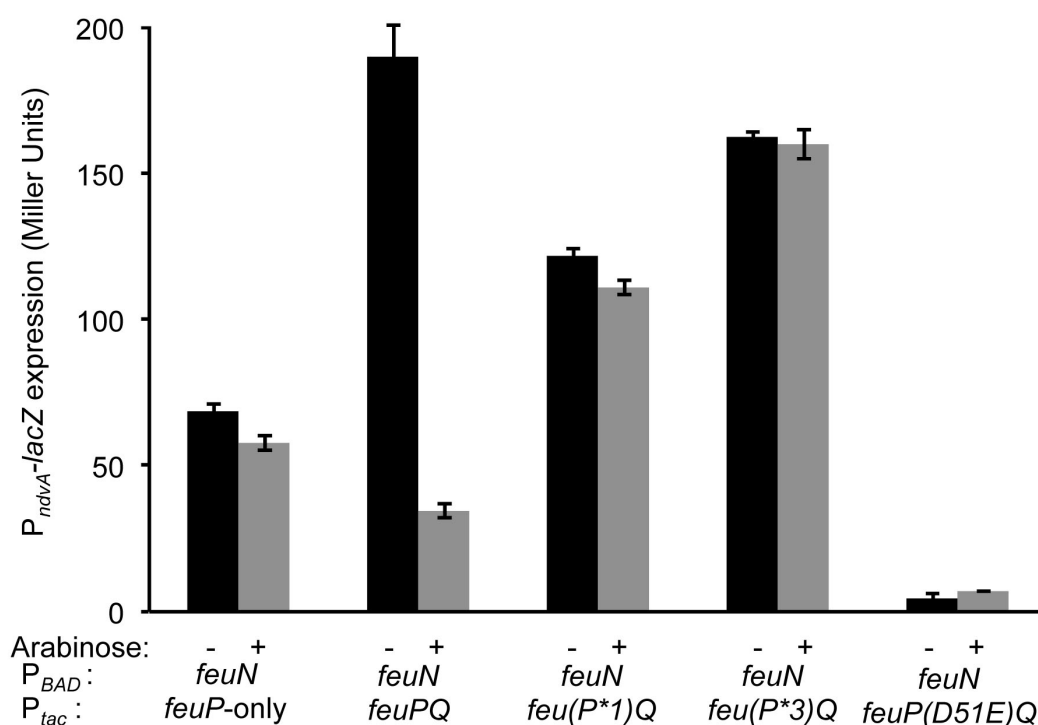


Figure 2-5 FeuN/Q-inhibition of FeuP activity requires the FeuP residue Asp-51.

The P_{BAD} -*feuN* expression plasmid pJG355 is included in all strains. P_{tac} expression plasmids include pJG376 (*feuP*-only), pJG377 (*feuP-feuQ*), pC700 (*feuP*1-feuQ*), plasmid pC701 (*feuP*3-feuQ*), and plasmid pC702 (*feuP(D51E)-feuQ*). *feuP*1* and *feuP*3* alleles are described in the text. In this and all other P_{ndvA} -*lacZ* experiments, the reporter plasmid pJG286 is used, and P_{tac} expression is not induced with IPTG. Error bars represent standard deviation from the mean based on triplicate experiments. Strains were grown in the presence or absence of arabinose as indicated.

Therefore, we observe that when FeuN is absent, FeuQ has an observable positive regulatory function; but when FeuN is present, FeuQ has an observable negative regulatory function. In each case, evidence suggests that this regulation is dependent on the phosphorylation state of FeuP. The precise mechanism for this FeuN negative regulation through FeuQ is unknown.

To gain greater insight into the mechanism of FeuN/Q inhibition, we performed a genetic screen for mutants of FeuQ that caused *ndvA* transcription to become less sensitive to FeuN (Figure 2-6A). FeuQ mutants were transformed into *E. coli* harboring the arabinose-inducible FeuN expression vector and the *PndvA::lacZ* reporter plasmid. Cells were plated on medium supplemented with X-gal and arabinose. Most resulting colonies were white, due to functional inhibition of reporter gene expression in cooperation with FeuN. However, several colonies were dark blue, indicating that one or more mutations in *feuQ* caused the phosphorylation cascade to be insensitive to the inhibitory effect of FeuN. Most FeuQ alterations that give rise to this phenotype cluster in and around its two predicted transmembrane domains. Figure 2-6B shows five such alterations, and a sixth alteration (E346K) occurring in the cytoplasmic kinase/phosphatase domain.

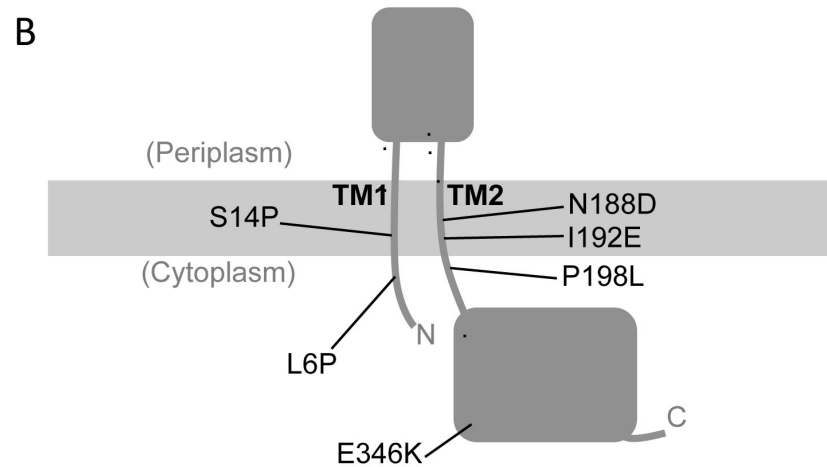
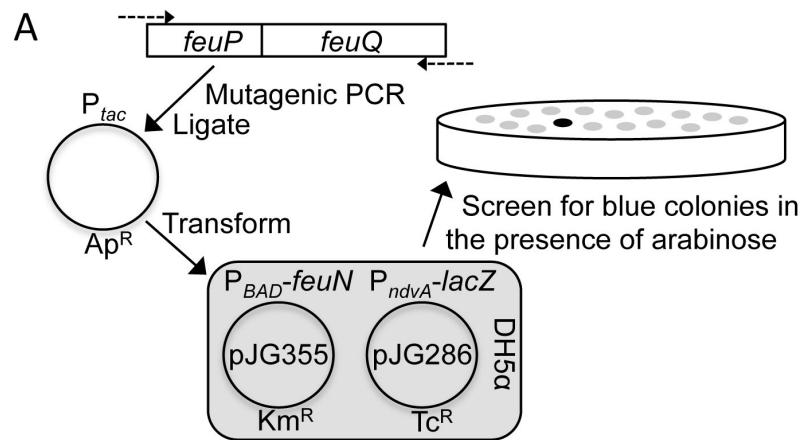


Figure 2-6 Specific FeuQ alterations diminish the effects of FeuN

A. Schematic diagram of the mutant isolation strategy

B. Diagram of the FeuQ polypeptide (dark gray) sitting in the inner membrane (light gray) by way of two transmembrane domains (TM1 and TM2). Amino acid alterations giving rise to the FeuN-insensitivity phenotype are shown.

That several alterations in FeuQ abrogate the action of FeuN in a heterologous host suggests that FeuQ is a direct target of FeuN. Based on sequence similarity with well-characterized sensor kinases (such as the *E. coli* proteins EnvZ and PhoQ discussed previously) and *in silico* analysis of probable transmembrane domains, it is evident that FeuQ is a canonical sensor kinase with two transmembrane segments (TM1: residues 10-32; TM2: residues 174-196), a periplasmic domain (residues 33-173), and a cytoplasmic domain (residues 197-461). A FeuN-FeuQ interaction could therefore occur in either the periplasm or the cytoplasm, or possibly in the membrane.

To determine the localization of FeuN, I created translational fusions of FeuN to either green fluorescent protein (GFP) or *E. coli* PhoA. It is well established that the GFP chromophore develops poorly in the bacterial periplasm, and conversely, that PhoA is inactive in the cytoplasm, due to insufficient disulfide bond formation (Hoffman & Wright, 1985) (Feilmeier, Iseminger, Schroeder, Webber, & Phillips, 2000). Using our *E. coli* system, PhoA lacking its native secretion signal (Δ ss-PhoA) was fused to the C-terminus of the full-length FeuN (residues 1-83) or C-terminally truncated FeuN (residues 1-48). In both instances, PhoA was directed to the periplasm, as was evidenced by blue colony color on plates supplemented with X-phos (Figure 2-7A). In a complimentary experiment, I found that a translational fusion of full-length FeuN to GFP did not exhibit fluorescence (Figure 2-7B), again indicative of a periplasmic localization. These experiments suggest that at least a large portion of the FeuN polypeptide is directed to the periplasm, but further evidence was needed to determine whether the N-terminus remains integrated in the inner membrane, or if the polypeptide is processed and released as a soluble periplasmic protein. It is interesting to note that FeuN contains two cysteine residues that are absolutely conserved in homologous proteins found in *Agrobacterium*, *Brucella*, and

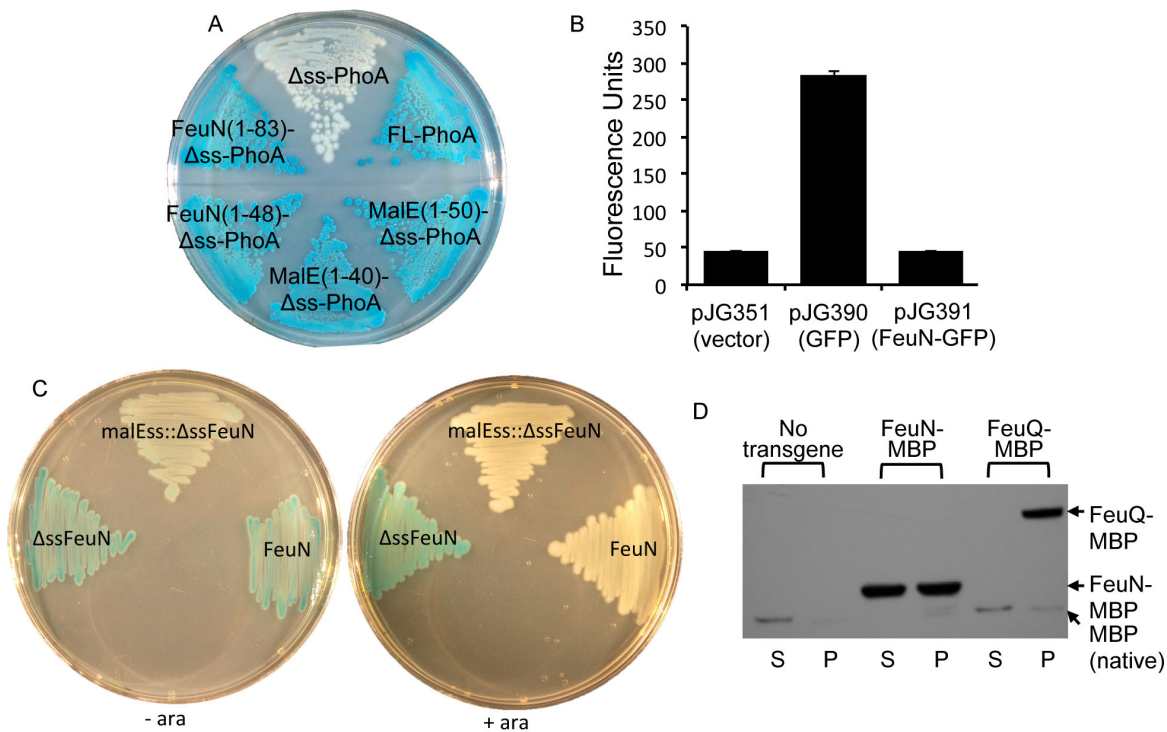
Bartonella (See Figure 2-3C). Screens for non-functional *feuN* have shown both of these cysteine residues to be critical for FeuN function, indicating a possible periplasmic disulfide bond in the FeuN protein.

To further investigate the localization of FeuN, I made use of MalE maltose binding protein (MBP) fusion proteins. MalE is a well characterized cleaved periplasmic protein containing a N-terminal signal sequence that directs translocation into the periplasm and is cleaved at a conserved AXA site (Bassford Jr, 1990; Bedouelle et al., 1980). I noted an RVA site on FeuN that might serve as a potential cleavage site for periplasmic transport (See Figure 2-7 Bottom). According to this hypothesis, the portion of FeuN following this proposed cleavage site, Δ ssFeuN, (FeuN49-83) would be sufficient for FeuQ negative regulation if directed to the periplasm by a canonical periplasmic signal sequence. For the experiment shown in 2-6 C, I tested this hypothesis with a MalE- Δ ssFeuN(49-83) fusion. As predicted, the Δ ssFeuN strain with lacking a signaling sequence was unable to regulate FeuPQ signaling from the cytoplasm. However, when forced into the periplasm by MalE fusion, MalE- Δ ssFeuN was able negatively regulate FeuQ signaling. Interestingly, the more canonical (and presumably more efficient) signal sequence used by MalE- Δ ssFeuN results in significant improvement in negative regulatory activity, showing observable negative regulation of the system even in un-induced, very low-level expression conditions where wild type FeuN fails to achieve an effect. Thus, the hybrid MalE- Δ ssFeuN appears to be a better negative regulator of FeuQ than the wild type FeuN.

In a second experiment, I tested the hypothesis that FeuN includes a signal sequence that directs secretion into the periplasm. MBP lacking its native secretion signal was fused to the C-terminus of FeuN. To provide a negative control, the same MBP fragment was also fused to the

C terminal end of the SK FeuQ. *E. coli* cells harboring these fusions were treated to release outer membrane and periplasmic material from intact cytoplasmic compartments (spheroplasts). Protein samples from supernatants and pellets were probed with anti-MBP antibodies (Figure 2-7C). As expected, the FeuQ-MBP fusion is not released into the supernatant, consistent with its predicted localization to the inner membrane. The FeuN-MBP fusion is strongly detected in the supernatant, consistent with its localization to the periplasm as a cleaved peptide. An equal amount of FeuN-MBP is found in the pellet fraction, possibly as a result of incomplete outer membrane lysis or incomplete secretion of the recombinant protein. Native chromosomally encoded MBP is less abundant, and is enriched in the supernatant fractions as expected for a periplasmic substrate binding protein

Taken together, the evidence is strong that FeuN is a cleaved periplasmic protein that directly interacts with FeuQ in the periplasm or in the inner membrane. It is not clear how a periplasmic protein might interact with an SK to change the cytoplasmic transcription of downstream target genes. In the work reported here, I continue to take advantage of the *E. coli* heterologous system to conduct a detailed genetic dissection of the FeuN/FeuQ regulatory interaction to learn how this signal might be passed across the membrane into the cell. Our data point to a specific periplasmic FeuN/FeuQ binding interface that modulates the state of FeuP in the cytoplasm.



FeuN: MSSALTIAAALAGLAGFSPPAVNMPAVVVVRA*GDSCAAAAQVVAQTGGELLSAQPTGDGQCVVTVLIPNGGRPKKVTVKVP

Figure 2-7 FeuN is a cleaved periplasmic protein

A. Δ ssPhoA fusions to full length PhoA, MalE(1-50), MalE(1-40), FeuN(1-48) (comprised of the predicted signal sequence of FeuN), and FeuN(1-83), full length FeuN. PhoA activity in the periplasm is evident in all cases except Δ ss-PhoA

B. The FeuN-GFP fusion measured for fluorescence compared to GFP and an empty vector control

C. Δ ssFeuN(49-83) was fused to the signal sequence of malE, thereby forcing FeuN into the periplasm. X-gal plates are measuring FeuPQ controlled *ndvA* expression under - ara (FeuN constructs not induced) and + ara (FeuN constructs induced) conditions.

D. *E. coli* cells over-expressing MBP fusion proteins were spheroplasted. Supernatant (S) and pellet (P) were probed with anti-MBP antibody. The migration positions of MBP fusion proteins and native MBP encoded on the *E. coli* chromosomes are indicated with arrows.

Below. Full amino acid sequence of FeuN is shown, with the asterisk denoting the putative signal sequence cleavage site. The gray amino acids represent the FeuN signal sequence, while black amino acids represent the mature FeuN peptide.

2.3 Materials and Methods

2.3.1 Bacteria culture technique

Strains used in this study are shown in Table 2-1. *E. coli* strains were cultured in Luria-Bertani (LB) medium supplemented as appropriate with ampicillin (Ap, 100 $\mu\text{g ml}^{-1}$), kanamycin (Km, 30 $\mu\text{g ml}^{-1}$), tetracycline (Tc, 10 $\mu\text{g ml}^{-1}$), L-arabinose (Ara, 0.3%), and 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-gal, 50 $\mu\text{g ml}^{-1}$). Inducing conditions for the P_{tac} promoter were never used in this study.

2.3.2 Plasmid construction

Plasmids used in this study are listed in Tables 2-1, S2 and S3. Detailed descriptions outlining the construction of these plasmids can be found in the Supplementary Methods section and Tables S2 and S3. All PCR-generated inserts were verified by Sanger sequencing. All primers used for plasmid construction are listed in Tables S1, S2, and S3.

2.3.3 Spheroplast preparation

E. coli DH5 α cells harboring the appropriate expression plasmids were grown overnight in LB. The cells were diluted 1:40 and grown for 3 hours at 30°C. Cells were then induced with 0.5 mM Isopropyl β -D-1-thiogalactopyranoside (IPTG) and allowed to grow for 3 more hours at 30°C. The equivalent of 0.43 mg (dry weight) of cells were pelleted at 16,100 x g and resuspended in 120 μl of 200 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (pH 7.8) buffer. Buffer consisting of 120 μl 200 mM HEPES, 1 M sucrose was then added, followed by 1.2 μl of 100 mM EDTA and 2.9 μl of 5 mg ml^{-1} lysozyme. Cells were osmotically shocked by addition of 244 μl of ddH $_2$ O and left at room temperature for 30 min. Spheroplasts

were centrifuged for 15 min at 16,100 x g. Supernatant was collected and the pellet was re-suspended in 450 μ l of 100 mM HEPES (pH 7.8), 0.25M sucrose. 50- μ l samples were then collected and mixed with 10 μ l of 6xSDS sample buffer (230 mM Tris pH 6.8, 10% glycerol, 1% SDS, 0.6% 2-mercaptoethanol, 0.001% bromophenol blue).

2.3.4 Western blot assay

Spheroplasts were suspended in 6xSDS sample buffer and heated at 100°C for 5 min. Proteins were resolved by SDS-PAGE. Proteins were transferred to PVDF Membrane (BIO-RAD) by wet transfer. Antibodies used were mouse monoclonal anti-MBP (Sigma-Aldrich) and goat HRP-labeled anti-mouse secondary antibody (BD Pharmigen).

2.3.5 Bacterial adenylate cyclase two-hybrid (BACTH) analysis

Two-hybrid screens were performed using the pKT25, pUT18, and pUT18C plasmids from the BACTH kit and the BTH101 indicator strain (EuroMedex). In this split-enzyme system, when the T18 and T25 fragments of *Bordetella pertussis* adenylate cyclase are in contact, they create an active enzyme that produces cyclic AMP, and this can be assayed by monitoring β -galactosidase activity (Karimova, Pidoux, Ullmann, & Ladant, 1998). Control plasmids encoding a leucine zipper (GCN4 from *Saccharomyces cerevisiae*) fused to the T18 and T25 fragments are included in the BACTH kit. Strains were suspended in water, diluted 100-fold, and 1 microliter was plated on LB-Ap/Km/X-gal. Colonies were grown at 30°C for 48 hours.

2.3.6 β -galactosidase assays

For measuring $P_{ndvA-lacZ}$ expression in *E. coli*, bacterial culture and β -galactosidase assays were carried out at 30°C. Stationary phase cultures were diluted 100-fold and grown for 6

hours, at which point assays were carried out according to (Miller, 1972). Activity was also observed qualitatively by growing on plates containing X-gal with or without L-arabinose.

2.3.7 Random mutagenesis

Random mutagenesis PCR was performed using Taq and VentR polymerases (NEB) on eight separate templates in parallel to improve diversity of mutations. The amplification products were digested and ligated downstream of *feuP* in the pJG376 vector. The ligation mixture was transformed into *E. coli* DH5 α /pJG286 followed by plating on LB-Tc/Ap/Km/X-gal. White colonies (indicating FeuN-mimicking phenotype) were re-streaked and re-tested quantitatively with a β -galactosidase test as shown in Figure S1.

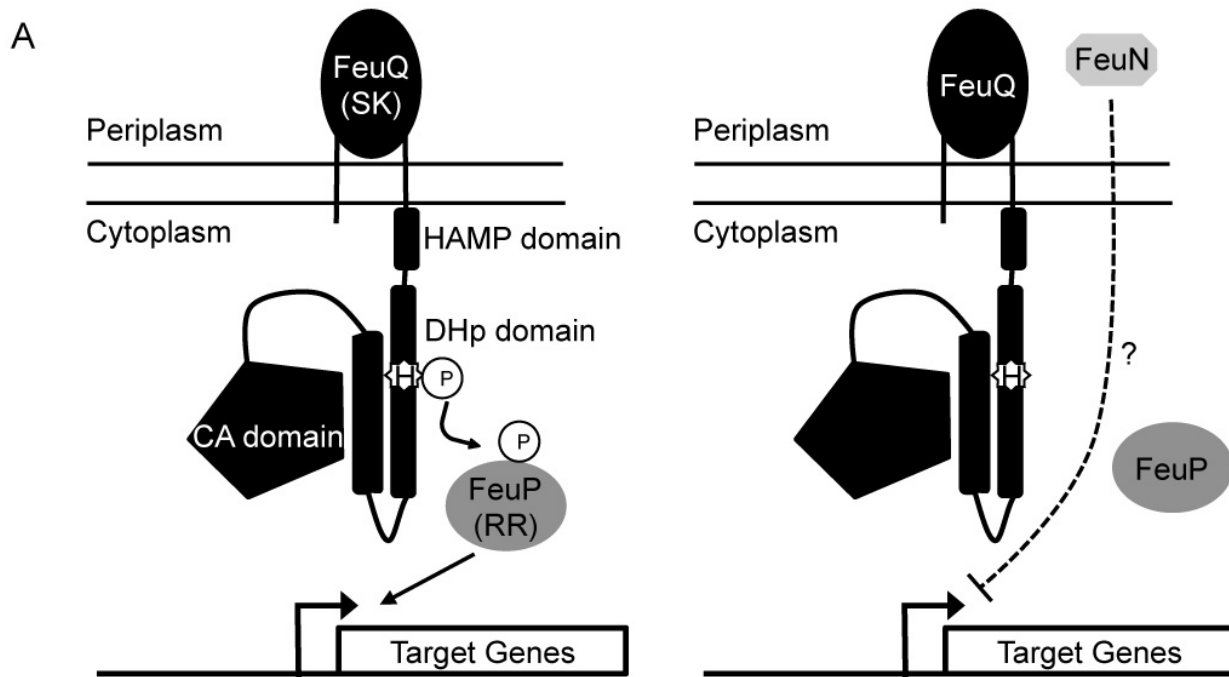
2.3.8 TTT scan of *feuQ* and *feuN*

feuQ TTT substitutions were created using overlap-extension PCR (OE-PCR) across the region of *feuQ* corresponding to periplasmic residues D43-R168. *feuN* TTT substitutions were created using OE-PCR across the region of *feuN* corresponding to the mature peptide (residues S36-K80). Detailed descriptions of the *feuQ* and *feuN* substitutions are provided in Table S3.

Table 2-1. Strains and plasmids used in Chapter 2

Strain	Description*	Source
DH5 α	<i>E. coli</i> cloning and expression strain	(Grant, Jessee)
BTH101	<i>cva-99 araD139 galE15 galK16 rnsL1 (Str^R) hsdR2 mcrA1</i>	EuroMedex
pJG249	P _{lac} expression plasmid (oriColE1: An ^R)	This study
pJG351	P _{BAD} expression plasmid (<i>ori</i> p15A: Km ^R)	(Carlvon, Ryther)
pJG286	P _{ndv4-lacZ} reporter plasmid (Tc ^R)	(Griffitts et al.)
pJG355	pJG351. <i>feuN</i>	(Carlvon et al.)
pJG376	pJG249. <i>feuP</i>	This study
pJG377	pJG376. <i>feuP-feuO</i>	This study
pJG610	pJG376. <i>feuP*1(D51G, T79S, I112V)</i>	This study
pJG611	pJG376. <i>feuP*3(D51E, K87G, K140R, H223O)</i>	This study
pJG620	pJG376. <i>feuP(D51E)</i>	This study
pC700	pJG610 + <i>feuO</i>	This study
pC701	pJG611 + <i>feuO</i>	This study
pC702	pJG620 + <i>feuO</i>	This study
pJG607	pJG376. <i>feuP-feuO(H255A)</i>	This study
pJG644	pJG376. <i>feuP-feuO(G398A)</i>	This study
pJG673	pJG376. <i>feuP-feuN</i>	This study
pJG674	pJG376. <i>feuP-feuN(C61R)</i>	This study
pJG642	pJG351. <i>feuO</i>	This study
pJG643	pJG351. <i>feuO(H255A)</i>	This study
pJG685	pJG376. <i>feuP-feuO</i> (139ETE141→KTT)	This study
pJG686	pJG376. <i>feuP-feuO</i> (139ETE141→TTK)	This study
pJG687	pJG376. <i>feuP-feuO</i> (139ETE141→KTK)	This study
pJG692	pJG351. <i>feuN</i> (73RPKK76→EPKK)	This study
pJG693	pJG351. <i>feuN</i> (73RPKK76→RPKE)	This study
pJG694	pJG351. <i>feuN</i> (73RPKK76→RPEK)	This study
pJG433	pJG351. P _{BAD} - <i>AssphoA</i>	This study
pJG435	pJG351. <i>feuN-AssphoA</i>	This study
pJG695	pJG351. <i>feuN</i> (73RPKK→EPKK)- <i>AssphoA</i> fusion	(Carlvon et al)

pJG696	pJG351. <i>feuN</i> (73RPKK→RPKE)- <i>AssphoA</i> fusion	(Carlvon et al
pKNT25	BACTH plasmid for T25 fragment of adenylate cyclase	EuroMedex
pUT18	BACTH plasmid for T18 fragment of adenylate cyclase (Ap ^R)	EuroMedex
pUT18c	BACTH plasmid for T18 fragment of adenylate cyclase (Ap ^R)	EuroMedex
pKT25-zip	BACTH plasmid for T25 fragment fused to GCN4 leucine	EuroMedex
pUT18C-zip	BATCH plasmid for T18 fragment fused to GCN4 leucine	EuroMedex
pJG699	BACTH plasmid for cleaved FeuN-T25 fusion	This study
pJG700	BACTH plasmid for periplasmic domain of FeuO-T18 fusion	This study
pJG701	BACTH plasmid for T18-periplasmic domain of FeuO fusion	This study
p432-1	pUT18. <i>feuO</i> (139ETE141→TTT) -T18 fusion	This study
p432-2	pUT18. <i>feuO</i> (139ETE141→TTK) -T18 fusion	This study
p432-3	pUT18. <i>feuO</i> (139ETE141→KTK) -T18 fusion	This study
p432-5	pKNT25. <i>feuN</i> (73RPKK76→EPKK) -T25 fusion	This study
p432-6	pKNT25. <i>feuN</i> (73RPKK76→RPKE) -T25 fusion	This study
p432-7	pKNT25. <i>feuN</i> (72GRP74→TTT) -T25 fusion	This study
p432-8	pKNT25. <i>feuN</i> (75KKV77→TTT) -T25 fusion	This study
*Str ^R , streptomycin resistance; Ap ^R , ampicillin resistance; Km ^R , kanamycin resistance; Tc ^R ,		



B

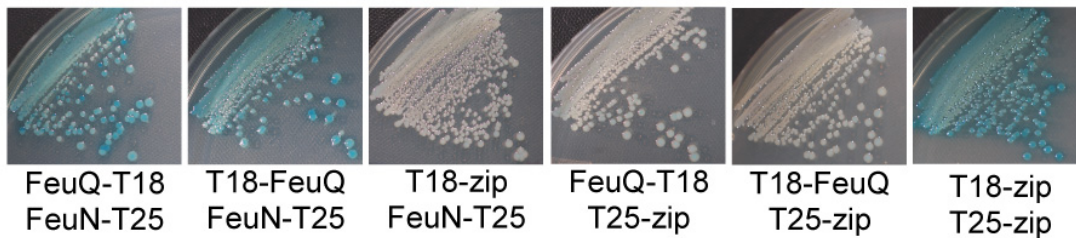


Figure 2-8 Schematic model of FeuNPQ function, and evidence for direct interaction between FeuN and FeuQ

A. Left: current model for FeuP activation by FeuQ via phosphorylation. Right: FeuN inhibits FeuP target gene expression by an unknown mechanism.

B. Bacterial two-hybrid tests: cultures were spread to single colonies on LB plates containing X-gal. The periplasmic domain of FeuQ (R34-R171) and the mature FeuN protein (G33-end) were fused respectively to the T18 and T25 subunits in the BACTH system (described in Methods). T25-zip and T18-zip are controls with the T25 and T18 subunits fused to the yeast homodimerizing GCN4 protein.

2.4 Results

2.4.1 Bacterial two-hybrid analysis indicates a physical interaction between FeuN and FeuQ

To determine whether FeuN and FeuQ interact directly, the mature sequence of the periplasmic protein FeuN (Carlyon et al., 2010), and the periplasmic domain of FeuQ (R34-R171) were tested in a bacterial two-hybrid (BACTH) assay. I fused FeuN to the T25 fragment of *Bordetella pertussis* adenylate cyclase, and the periplasmic domain of FeuQ to the T18 fragment. The BACTH kit includes leucine zipper control fusions to both the T25 and T18 fragments (T18-zip and T25-zip). These T18 and T25 fusions were expressed in various combinations in the *E. coli* indicator strain BTH101 and plated on X-gal (Figure 2-8B). The data indicate a direct interaction between the FeuN-T25 fusion and both the FeuQ-T18 and T18-FeuQ fusions. No interaction was detected in any of the negative control combinations. Our attempts to detect a physical interaction between FeuN and membrane-bound FeuQ using a spheroplast pull-down experiment were unsuccessful, suggesting that the interaction may be transient (data not shown).

2.4.2 Does the FeuN/Q-interaction cause an increase in FeuQ phosphatase activity?

We previously showed that in the presence of FeuN, FeuQ has an observable negative regulatory activity (Carlyon et al., 2010). This negative regulation is presumably due to phosphatase activity. If FeuQ has both kinase and phosphatase activities, I hypothesized that selective elimination of kinase activity by mutation would create dominant-negative variants due to unchecked phosphatase activity. This is consistent with observations in other bifunctional sensor kinases (Atkinson & Ninfa, 1993; Hsing & Silhavy, 1997; Tanaka et al., 1998). To test this hypothesis, I created two FeuQ variants that were predicted to eliminate kinase activity

while potentially leaving phosphatase activity intact. FeuQ (H255A) lacks the critical conserved His residue that is the site of autophosphorylation. An identical mutation in the well-characterized SK EnvZ was shown to eliminate kinase but not phosphatase activity (Hsing & Silhavy, 1997). Another variant, FeuQ (G398A), changes a highly conserved Gly residue previously shown to be critical for ATP binding and autophosphorylation (Tanaka et al., 1998). In the experiment shown in Figure 2-9A, the H255A and G398A variants were each co-expressed with FeuP. As predicted, both FeuQ variants displayed constitutively inhibiting phenotypes. This signaling inhibition is presumably due to intact phosphatase activity dominating the now disabled kinase activity.

To determine if FeuN could directly increase the signal inhibition of our kinase-deficient FeuQ allele, the FeuQ (H255A) variant was measured in the presence and absence of FeuN. For the experiment shown in Figure 2-9B, *feuQ* alleles were expressed from an uninduced P_{BAD} promoter, thus ensuring very low FeuQ concentration in the cell. This is required because FeuQ (H255A) expressed from the uninduced P_{tac} system tends to bring reporter gene activity down to very low levels, obscuring any additional effects that may be brought about by FeuN. From the uninduced P_{tac} promoter, which transcribes at higher levels, I expressed a bicistronic *feuP-feuN* cassette. Where appropriate, the nonfunctional *feuN(C61R)* allele (Carlyon et al., 2010) is used as a negative control. Under these conditions we see that FeuN is indeed able to enhance the already inhibitory effect of FeuQ (H255A). This is consistent with the notion that FeuN directly enhances FeuQ phosphatase activity.

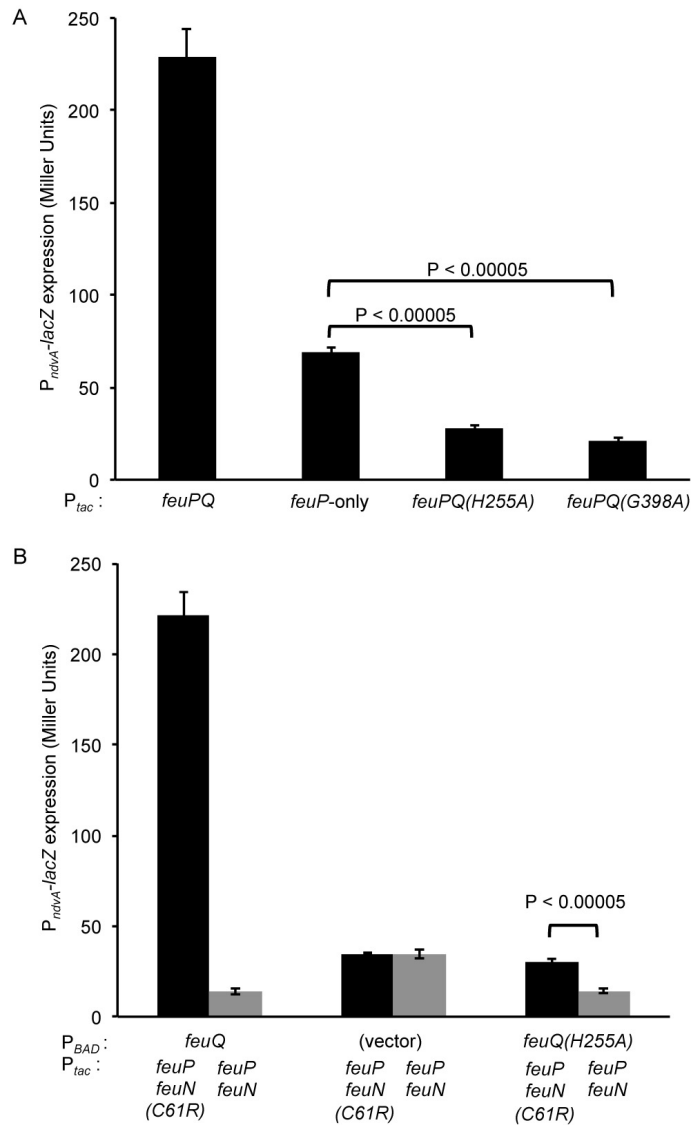


Figure 2-9 FeuN enhances the negative effect of a kinase-deficient FeuQ allele

A. FeuQ alleles deficient in kinase activity negatively regulate FeuP. P_{tac} expression plasmids include pJG377 (*feuP-feuQ*), pJG376 (*feuP*-only), pJG607 (*feuP-feuQ(H255A)*), and pJG644 (*feuP-feuQ(G398A)*). Only basal-level expression from P_{tac} was allowed (no induction).

B. FeuN acts on FeuQ (H255A) despite elimination of kinase activity. P_{tac} bicistronic expression plasmids include pJG674 (*feuP-feuN(C61R)*) and pJG673 (*feuP-feuN*). P_{BAD} expression plasmids include pJG642 (*feuQ*), pJG351 (vector), and pJG643 (*feuQ(H255A)*). In this experiment, only basal-level expression was used for both expression systems (P_{tac} and P_{BAD}). P values in A and B were determined by standard t-tests

2.4.3 FeuQ mutations lead to both FeuN-insensitive and FeuN-mimicking phenotypes

While it seems clear that FeuN causes some shift in FeuQ that stimulates phosphatase activity, the mechanism of this shift is unclear. I performed random mutagenesis of *feuQ* searching for “FeuN-insensitive” (constitutively activating) and “FeuN-mimicking” (constitutively inhibiting) alleles that might reveal domains of the FeuQ protein that are important for potential FeuN-interaction and toggling between kinase and phosphatase activities. I expected to find mutations leading to each phenotype in the HAMP, DHp, and CA domains, as these regions are known to be involved in SK signal integration (Atkinson & Ninfa, 1993; Hsing et al., 1998; Willett & Kirby, 2012). I suspected that mutations might also correspond to the FeuQ periplasmic domain, where FeuQ and FeuN are most likely to directly interact.

To identify FeuN-mimicking alleles of *feuQ*, I created a mutant library of *feuQ* and expressed library members in a strain with wild-type *feuP* and the P_{ndvA} -*lacZ* reporter. I then screened for white colonies on X-gal plates. Conveniently, null alleles of FeuQ appear light blue in this system due to low-level, nonspecific phosphorylation of FeuP, while white colonies correspond only to FeuQ variants with functional negative regulatory activity (the FeuN-mimicking phenotype). As expected, I found FeuN-mimicking alleles clustering to the HAMP, DHp, and CA domains (Figure 2-10). Interestingly, I also found mutations clustering to the transmembrane segments and the N- and C- terminal ends of the periplasmic domain. These mutations highlight residues that potentially mediate the transfer of FeuN binding information into the cytoplasm. The screen for random FeuN-insensitive *feuQ* mutants was described previously (Carlyon et al., 2010). Here the mutant library of *feuQ* was expressed in a strain with wild-type *feuP*, wild-type *feuN*, and the P_{ndvA} -*lacZ* reporter and screened on X-gal plates. Under these conditions, most colonies are white due to FeuN/Q inhibition of the reporter gene, while

dark blue colonies identify FeuN-insensitive FeuQ mutants. The mutations identified in this screen also clustered to amino acids in or flanking the transmembrane domain, in addition to a single mutation in the HAMP domain and a single mutation in the CA domain (see Figure 2-10). These mutations presumably result in either the loss of interaction between FeuQ and FeuN, or serve to lock FeuQ into a constitutively active state downstream of FeuN interaction.

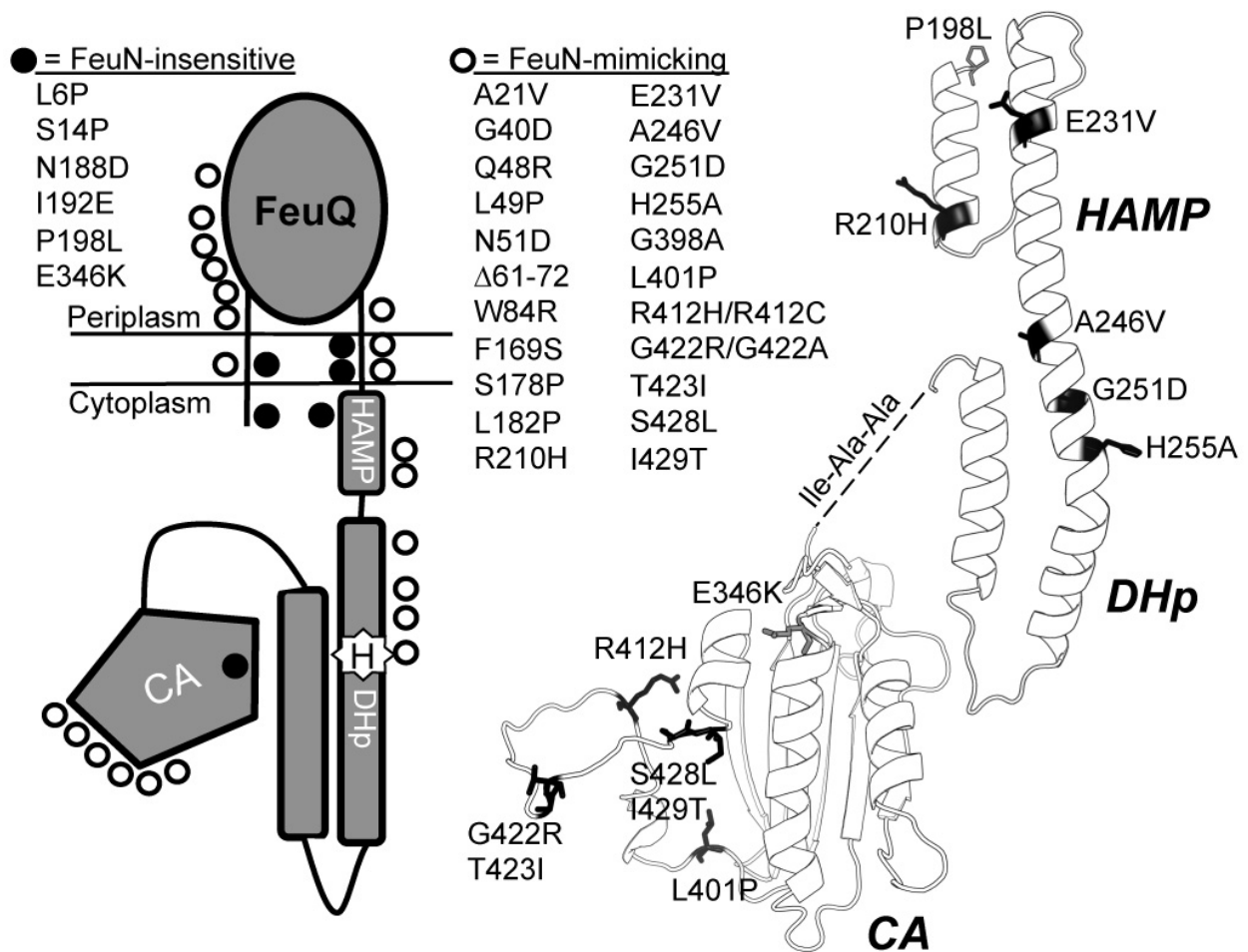


Figure 2-10 Amino acid changes in FeuQ that cause FeuN-insensitive or FeuN-mimicking phenotypes

Left: Schematic map of mutations as they correspond to conserved SK domains. Right: Predicted structure of FeuQ cytoplasmic domains generated by the SWISS-MODEL Workspace (Arnold, Bordoli, Kopp, & Schwede, 2006; Kiefer, Arnold, Künzli, Bordoli, & Schwede, 2009). Selected mutations are mapped to the structure with side chains displayed as points of reference.

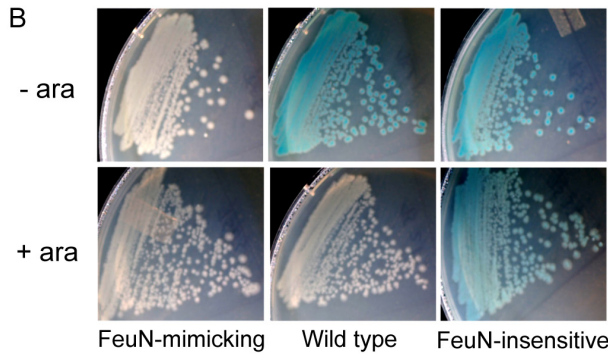
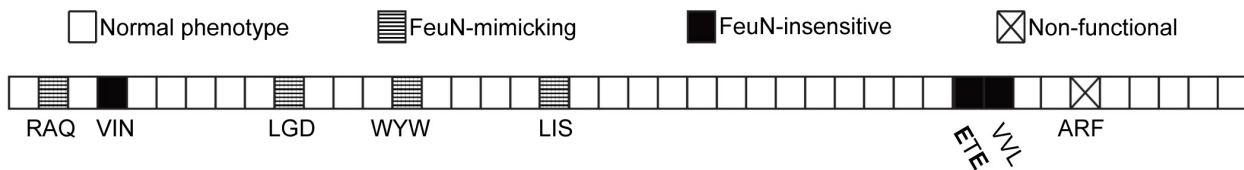
Considering that FeuN is a periplasmic protein, one might expect that FeuN-insensitive behavior might arise from amino acid changes to the FeuQ periplasmic domain. However, our random mutagenesis screen did not identify such a class. It may be that no single nucleotide change results in FeuN-insensitivity. I performed a targeted tri-threonine (TTT) amino acid substitution scan of the FeuQ periplasmic domain to look for FeuN-insensitive mutants. For TTT scanning, consecutive triplets of residues were changed to TTT. TTT substitution variants were made in FeuQ residues 43-168, covering most of the predicted periplasmic domain. The FeuQ variants were co-expressed with FeuP in our *E. coli* system allowing for arabinose-inducible control of FeuN. Each allele was tested both on X-gal plates and by quantitative β -galactosidase assays for signaling in the presence or absence of arabinose. The TTT scan of the FeuQ periplasmic domain revealed substitutions resulting in all four possible FeuQ phenotypes (see Figure 2-11A and quantified data in Fig. S1): wild type (34/42), FeuN-mimicking (4/42), FeuN-insensitive (3/42), and non-functional (1/42). Wild-type alleles show normal reporter gene activation and FeuN-responsiveness. FeuN-mimicking alleles mimic FeuN/Q inhibition even in the absence of FeuN. FeuN-insensitive mutants show reporter gene activation that is only slightly decreased when FeuN is co-expressed (none of the alleles was 100% FeuN-insensitive). Of particular interest were two adjacent TTT mutants at FeuQ residues 139ETE141 and 142VVL144, both of which gave rise to an FeuN-insensitive phenotype. This TTT scanning approach reveals the critical role of the FeuQ periplasmic domain in governing responses to FeuN expression.

2.4.4 TTT scan of mature FeuN peptide shows regions critical for FeuN function

I performed a targeted TTT scan to look for critical FeuN regions potentially involved in its interaction with FeuQ (Figure 2-11C). TTT substitutions were introduced into the mature

FeuN peptide from residues 36-80 (a signal peptide cleavage site is predicted between residues 32 and 33). These FeuN variants were cloned into our *E. coli* system under the control of the arabinose promoter. Strains were assayed with and without arabinose. The FeuN (TTT) phenotypes were characterized by the percent loss of FeuN activity compared to wild type. The majority of TTT substitutions showed less than 5% loss of function. Four alleles showed a 35-60% loss of function, while only three alleles showed loss of function between 68-80% (Figures 2-11C, S2). These three substitutions highlight areas potentially involved in interaction with FeuQ. None of the FeuN (TTT) variants was 100% inactive.

A FeuQ TTT scan of periplasmic domain: allele phenotypes



C FeuN TTT allele phenotypes

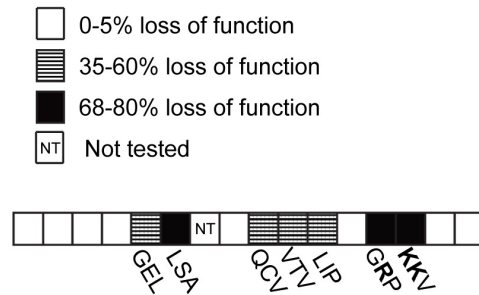


Figure 2-11 Map of TTT scan phenotypes of FeuQ (periplasmic domain) and FeuN

A. Map of phenotypes caused by TTT substitutions across the periplasmic region of FeuQ (D43-R168). Detailed descriptions of these substitutions are given in Table S3. Each box represents three amino acid residues. Where the TTT substitution caused a change in signaling phenotype, the wild-type sequence is annotated. Charged residues important for subsequent studies are shown in bold. Quantitative β -galactosidase activity of each TTT mutant is shown in Fig. S1.

B. Representative photos of FeuN-mimicking, wild-type, and FeuN-insensitive phenotypes when tested on X-gal plates.

C. Map of phenotypes caused by TTT substitutions across FeuN (S36-K80, corresponding to the mature peptide). Detailed descriptions of these substitutions are given in Table S3. Phenotypes are reported as the percent loss of FeuN function compared to the unaltered protein. β -galactosidase activity of each TTT mutant is given in Fig. S2

2.4.5 Conserved compatibly-charged residues within FeuQ and FeuN are required for FeuN/Q interaction

The study of amino acid residues critical for function in both FeuN and the periplasmic region of FeuQ highlighted potential regions of direct interaction between the two proteins. Our attention was drawn to a region of the FeuQ periplasmic domain where a negatively charged patch (139ETE141) leads to FeuN-insensitivity when mutated to TTT (see Figure 2-11A). Interestingly, the FeuN TTT scan revealed a functionally critical positively charged (73RPKK76) region near the C-terminal end of the protein (see Figure 2-11C). In alignments with homologous proteins from other alphaproteobacteria, the FeuQ ETE and FeuN RPKK patches show strong charge conservation (Figure 2-12A). I hypothesized that a direct FeuN/FeuQ interaction may involve these conserved oppositely-charged sequence motifs. According to this model, one would predict that charge-reverse mutations in these regions would enhance the disruption of interaction beyond what was seen in the charge-neutralizing TTT substitutions originally used to discover them. To test this prediction in FeuQ, the single charge-reversal mutants (TTT→KTT) and (TTT→TTK), as well as a double charge-reversed (TTT→KTK) variant were tested in the presence or absence of FeuN (Figure 2-12B). The KTT and TTK variants showed a slight enhancement in FeuN-insensitivity compared to the TTT mutation, consistent with our expectation. These charge-reversal phenotypes associated with the FeuQ 139ETE141 region underscore its importance in periplasmic signal recognition and the transfer of that information into the cytoplasm. Surprisingly, the KTK variant with its double positive charge became strongly FeuN-mimicking, rather than strongly FeuN-insensitive as expected. It may be that introduction of strong positive charge to the FeuQ ETE region is all that is required to induce

phosphatase activity, whether by interaction with the strongly positively charged FeuN RPKK region, or supplied by mutation in the FeuQ KTK mutant.

I then performed the reciprocal experiment on FeuN, creating three single charge-reverse variants at the critical RPKK sequence: FeuN (RPKK→EPKK), (RPKK→RPEK), and (RPKK→RPKE) (Figure 2-12C). The RPEK variant was minimally affected, but the EPKK and RPKE variants were completely non-functional. This is consistent with our prediction that charge-reverse mutants would display greater loss of function than what was observed in our original TTT mutants. To confirm the stability and correct localization of the nonfunctional EPKK and RPKE variants, PhoA fusions to the C terminus of these variants were tested for alkaline phosphatase (AP) activity. Both displayed AP activity similar to that of the PhoA fusion to wild-type FeuN (Figure S3). From these data it appears that the FeuN/FeuQ interaction is strongly influenced by these compatibly charged FeuQ ETE and FeuN RPKK sequences, with the internal lysine residue in the FeuN RPKK sequence playing a minor role.

I selected some of our charge-reversed and TTT-substituted FeuQ and FeuN mutant variants to test in our two-hybrid system to see if these mutants were still capable of physical interaction with their wild-type partner proteins. The periplasmic domains of the FeuQ (ETE→TTT), (ETE→TTK), and (ETE→KTK) variants were fused to the T18 fragment and tested against unaltered FeuN::T25. Conversely, the FeuN (GRP→TTT), (KKV→TTT), (RPKK→EPKK), and (RPKK→RPKE) variants were fused to the T25 fragment and tested against unaltered FeuQ::T18. In all cases, the charge-altering changes abolish the interaction Figure 2-13.

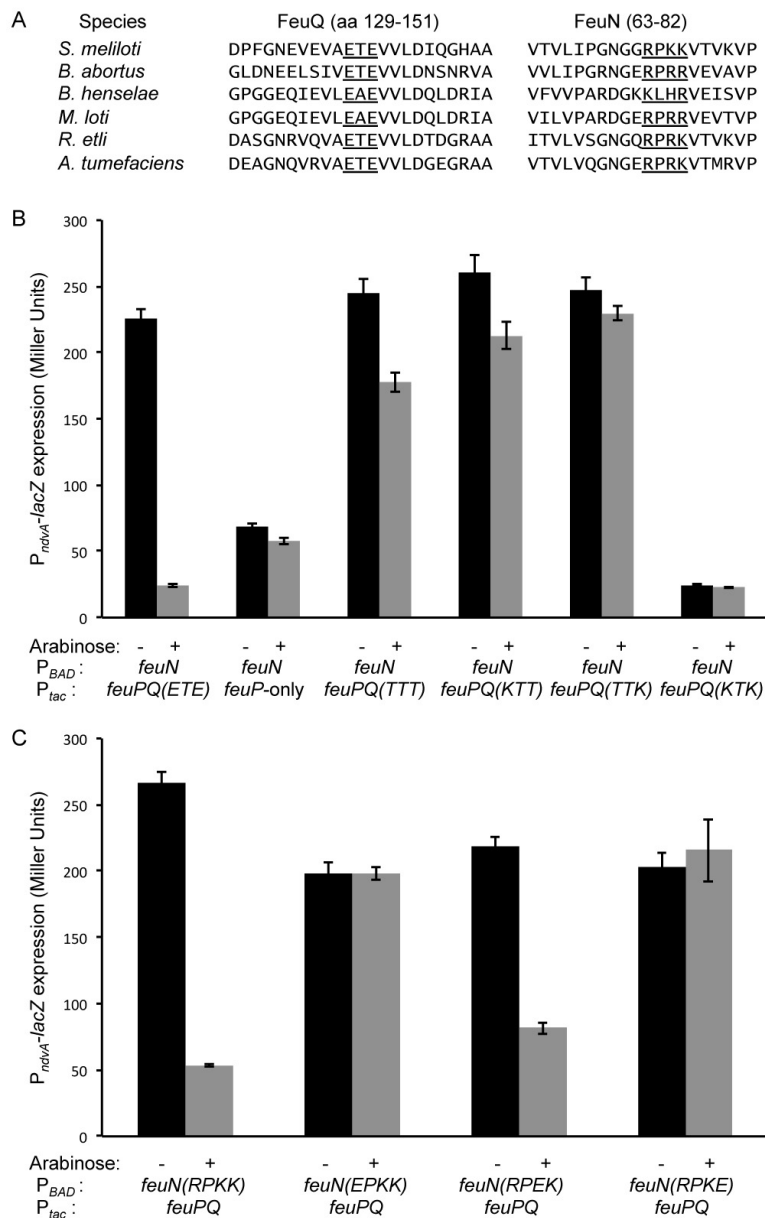


Figure 2-12 Conserved, compatibly charged residues in the periplasmic region of FeuQ and in FeuN are involved in target gene regulation

A. Sequence alignments of the relevant regions of FeuQ and FeuN. Conserved charged regions implicated in the TTT mutant screens are underlined. Alphaproteobacterial species used for this comparison are *Brucella abortus* 2308, *Bartonella henselae* Houston 1, *Mesorhizobium loti* BNC1, *Rhizobium etli* CFN42, and *Agrobacterium tumefaciens* C58.

B. Analysis of the FeuQ ETE motif. P_{tac} expression plasmids used were: pJG377 (wild-type *feuP-feuQ*), pJG376 (*feuP-only*), the original *feuP-feuQ*(ETE→TTT) clone, pJG685 (*feuP-feuQ*(ETE→KTT)), pJG686 (*feuP-feuQ*(ETE→TTK)), and pJG687 (*feuP-feuQ*(ETE→KTK)). The P_{BAD} -*feuN* plasmid pJG355 is included in all strains to monitor responsiveness to FeuN expression. For B and C, only basal-level expression from P_{tac} was allowed.

C. Analysis of the FeuN RPKK motif. The P_{tac} plasmid pJG377 (*feuP-feuQ*) is included in all strains. The P_{BAD} plasmids include pJG355 (*feuN*), pJG692 (*feuN*(RPKK→EPKK)), pJG693 (*feuN*(RPKK→RPKE)), and pJG694 (*feuN*(RPKK→RPEK)).

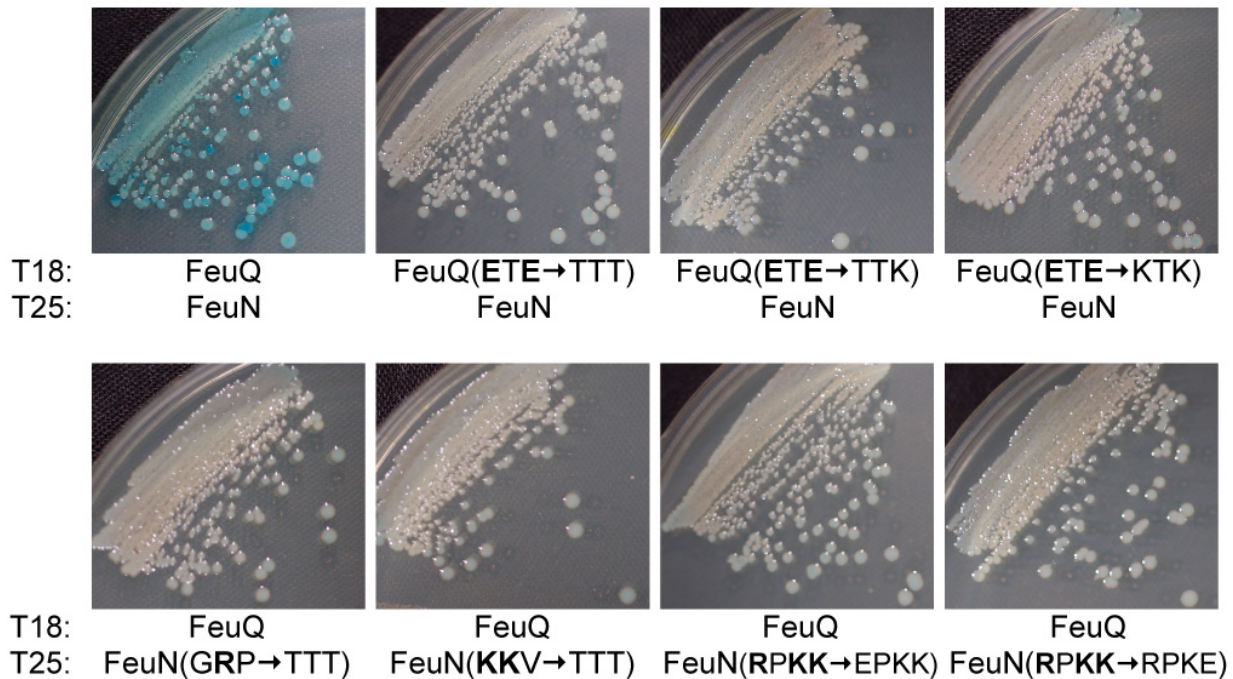


Figure 2-13 Charge alterations in the FeuQ ETE and FeuN RPKK motifs abolish the FeuN/FeuQ physical interaction

Relevant charge-altered variants were tested in the bacterial two-hybrid system. In this experiment, fusion orientations were always FeuQ-T18 and FeuN-T25. The strain expressing wild-type FeuQ-T18 and FeuN-T25 serves as a positive control (see also Fig. 1). Bold letters orient the reader to charged residues investigated in this experiment.

2.5 Discussion

Our data suggest that FeuN interacts directly with the periplasmic domain of FeuQ at a surface mediated in part by electrostatic interactions, and this binding event sends a signal through the plasma membrane that ultimately stimulates FeuQ phosphatase activity. There are several TCS auxiliary proteins (e.g. PII, KipI, LuxP, CpxP, MzrA, Sda) that are known to activate net SK phosphatase activity, but by different means. Some of these proteins appear to directly increase the rate of phosphatase activity, others directly inhibit autokinase and/or phosphotransfer activities, and still others interfere with SK ligand recognition (Gerken & Misra, 2010; Jacques et al., 2011; Neiditch et al., 2006; Pioszak & Ninfa, 2003b; S. L. Rowland et al., 2004; L. Wang et al., 1997; X. Zhou et al., 2011).

Our ability to construct multiple kinase-dead alleles of FeuQ has allowed us to study FeuQ phosphatase activity in isolation and to discern the role of FeuN in stimulating phosphatase activity. I was unable to perform the converse experiment by selectively eliminating phosphatase activity. In some sensor kinases it has been shown that within a conserved E/DxxT motif of the DHP domain, the Thr residue is critical for phosphatase activity (Huynh & Stewart, 2011; Willett & Kirby, 2012). FeuQ shares this conserved motif, but mutations to the conserved Thr did not give a phenotype consistent with eliminated phosphatase activity (data not shown).

To understand how a direct interaction between FeuN and FeuQ in the periplasm might influence cytoplasmic FeuQ functions, I performed both random and targeted mutagenesis on FeuQ, focusing our attention on periplasmic mutations. I discovered critical regions for FeuQ signal modulation that cluster to the N- and C- terminal ends of the periplasmic sensing domain (see Figures 2-9 and 2-10). These clusters may occupy the same three-dimensional space,

creating a single surface for direct interaction with FeuN or other external signals. On closer analysis of this potential interface our attention was drawn to conserved compatibly-charged amino acid residues—a negatively charged ETE sequence on FeuQ and a positively charged RPKK sequence on FeuN. I predicted that if these charged residues were involved in a direct interaction then reversing their charge would disrupt the FeuN/Q interaction due to charge repulsion. In single charge-reverse experiments our prediction proved accurate: charge-reversed mutants showed a dramatic loss of FeuN/Q inhibition in our reporter assay, and showed a loss of interaction in the two-hybrid assay. Interestingly, the charge-reversed variants are not easily corrected by making “compensatory” charge reversals on the partner protein (not shown). This suggests that a specific charge on the appropriate protein (positive on FeuN and negative on FeuQ) is required for a functional regulatory interaction. The fact that the *feuQ* double charge-reverse (ETE→KTK) mutant becomes FeuN-mimicking highlights the possibility that these periplasmic residues mediate a cytoplasmic phosphatase-kinase switch.

I propose that the FeuN/FeuQ interaction causes a conformational cascade that influences the relative rate of phosphorylation and dephosphorylation of FeuP in the cytoplasm. Presumably, many of the residues that mediate this cascade are highlighted in the mutant analysis summarized in this report.

2.6 Acknowledgements

This work was supported by National Institutes of Health Grant 1R15AI082504-01 (to J.S.G.). R.D.V. was supported by a Brigham Young University Graduate Research Fellowship Award. I thank Bill McCleary and Julianne Grose for providing plasmids used in this study.

2.7 Open questions

The discussion notes that charge-reversed variants of FeuQ/FeuN are not easily corrected by making “compensatory” charge reversals on the partner proteins. For these experiments, many charge-reversed FeuQ and FeuN mutants were combined with FeuP and the reporter *ndvA* plasmid and FeuN/Q inhibition was not observed. While it was easily visible that functional regulation was not reconstituted by charge swapping, it is not clear whether these charge-swapped mutant proteins physically interacted with one another. Two-hybrid analysis of charge-swapped mutants might reveal that physical interaction was restored, though functional regulation was not.

This work presents a mostly genetic dissection of the two-component system. Attempts to confirm our findings with biochemical experiments were often complicated by the difficulty of working with the membrane-bound nature of the sensor kinase FeuQ. In many SK studies, this problem has been avoided by working solely with the cytoplasmic portion of the sensor kinase. In such studies, the cytoplasmic domains are able to function and interact with the RR independently from the transmembrane and periplasmic regions. In the case of FeuQ, separation of the cytoplasmic domain from the membrane resulted in non-functional FeuQ – incapable of activating the RR FeuP. To this point, actual phosphotransfer and phosphatase activities of FeuQ on FeuP have not been biochemically observed, but only inferred by the required conservation of the His and Asp residues for each. Biochemical experiments using radio-labeled phosphate to actually track transfer to and from FeuP by FeuQ and FeuQ mutant alleles in the presence and absence of FeuN would complement this genetic analysis with concrete biochemical evidence.

Chapter 3. Gene networks involved in competitive rhizosphere colonization in the *Sinorhizobium meliloti* – *Medicago truncatula* symbiosis

3.1 Summary

The rhizobia-legume symbiosis is the most agriculturally significant form of natural nitrogen fixation, and accounts for almost 25% of all fixed nitrogen. The signals for establishing a symbiotic partnership between nitrogen-fixing bacteria (e.g. *Sinorhizobium meliloti*) and leguminous plants (e.g. *Medicago truncatula*) have been well characterized. Initial host-microbe compatibility is determined by plant recognition of bacteria-secreted glycolipid “Nod” factors. In response, the host facilitates a more exclusive mode of colonization by the formation of a root nodule – a new organ capable of hosting dense intracellular populations of symbiotic rhizobia for nitrogen fixation. Within the nodule, the bacteria differentiate into organelle-like nitrogen-fixing bacteroids. A great many of the signals involved in bacterial invasion of nodule tissue and subsequent bacteroid differentiation and nitrogen fixation have been identified and characterized at the molecular level.

These well-characterized symbiotic events are preceded by rhizosphere colonization, which is somewhat more poorly understood. I have adapted Tn-seq technology to allow for an exhaustive identification of *S. meliloti* genes that permit competitive colonization of the *M. truncatula* rhizosphere. This study combines large-scale transposon mutagenesis with deep sequencing to monitor *S. meliloti* genotypes that increase or decrease in relative abundance after competition in the rhizosphere. The analysis implicates a large ensemble of bacterial genes and pathways promoting rhizosphere colonization, provides hints about how the host plant shapes

this environment, and opens the door for mechanistic studies about how changes in the rhizosphere are sensed and interpreted by the microbial community.

3.2 Introduction

The rhizobia-legume symbiosis is responsible for fixing as much as 70 million tons of N per year, and represents the single largest source of naturally fixed nitrogen in agriculture (Zahran, 1999). An effort to develop more efficient and environmentally responsible methods for crop fertilization has refocused efforts to better understand and utilize symbiotic nitrogen fixation for crop production (Beatty & Good, 2011; Oldroyd & Dixon, 2014).

Studies on rhizobia-legume symbiosis have focused on the exchange of plant-produced flavonoids and bacterium-specific Nod factors as the initiating stages of infection. In response to a compatible Nod factor, a plant root hair curls around and entraps the potential symbiotic partner. The colonizing microbe invades the root tissue by cellular division along a network of developing tubes known as infection threads. Continued invasion of infection threads and penetration into host tissues is dependent upon a number of other compatibility factors, including the production of exopolysaccharides and cyclic glucans. As infection thread colonization progresses, the plant envelops the site of infection with new tissue, creating a new plant organ known as a nodule which houses the symbiotic microbe and provides stable conditions conducive to nitrogen fixation (Jones et al., 2007) (Oldroyd, 2013).

A wealth of information on the signals required for the development of nitrogen fixing nodules has allowed for the creation of mutant rhizobial strains able to more efficiently fix nitrogen for legumes (Maier & Brill, 1978). It is often difficult to establish persistence of these improved rhizobial strains in soils with pre-established microbial communities of ineffective or non-nitrogen fixing strains (Phillips, 1980). This is a major hurdle that must be overcome before

these strains can be used effectively in agriculture. The first requirement for successful nodulation is therefore competitive rhizosphere colonization.

Though soil and rhizosphere bacteria have been studied for over a century (Burriss, 1988), the genetics of competitive rhizosphere colonization remain poorly understood. This is in large part due to the difficulty of studying soil microbial communities. Research has been limited by the inability to grow many soil microbes in culture (D. Parkinson et al., 1971) and the difficulty in recovering microbial samples from the roots and soil (Kloepper & Beauchamp, 1992). Studies on plant growth promoting rhizosphere bacteria including *Pseudomonas* and *S. meliloti* have identified a few types of genes required for competitive root colonization. These genes include motility genes required for chemotaxis to plant root systems (Capdevila et al., 2004) (de Weert et al., 2002), LPS modifications to suppress plant immune response (Scheidle et al., 2005), the ability to competitively scavenge iron (C.-H. Yang & Crowley, 2000), the ability to synthesize required vitamins (e.g. vitamin B, biotin, thiamine) (Lugtenberg et al., 2001) (Streit et al., 1996), and the ability to synthesize amino acids (Simons et al., 1997).

Tn-seq has proven to be a powerful tool in the identification of gene networks involved in microbial fitness under many types of selective conditions ranging from antibiotic resistance (Gallagher, Shendure, & Manoil, 2011), to resistance to heavy metals (Yung et al., 2015), to the identification of genes required for host infection (Dong, Ho, Yoder-Himes, & Mekalanos, 2013) (Fu, Waldor, & Mekalanos, 2013). Tn-seq is a developing technology that combines principles of traditional transposon mutagenesis with deep sequencing to allow for large-scale identification of biological pathways required for growth under a given condition of interest (van Opijnen, Bodi, & Camilli, 2009). Tn-seq allows for simultaneous monitoring of an exhaustive panel of possible genotypes for an organism under selection in a condition of interest. This provides a

method to rapidly identify relevant gene networks and quantify their contributions to microbial fitness.

I have adapted Tn-seq technology to investigate the genetics of competitive rhizosphere colonization in the *Sinorhizobium meliloti* – *Medicago truncatula* symbiosis. Our analysis confirms the role of many genes previously determined to be involved in competitive root colonization, while also implicating a new set of previously unsuspected genes as potential determinants of rhizosphere colonization. Head to head competition assays with a selected panel of mutants provides compelling evidence that many of these genes are in fact involved in competitive rhizosphere colonization.

3.3 Materials and Methods

3.3.1 Bacteria culture technique

Strains were cultured in Luria-Bertani (LB), BRM, or RSM1, or RSM2 medium supplemented as appropriate with kanamycin (Km, 30 $\mu\text{g ml}^{-1}$), tetracycline (Tc, 10 $\mu\text{g ml}^{-1}$), neomycin (Nm, 25 $\mu\text{g ml}^{-1}$) streptomycin (Sm, 50 $\mu\text{g ml}^{-1}$), or diaminopimelic acid (DAP, 50 $\mu\text{g ml}^{-1}$). The recipes for BRM, RSM1, and RSM2 are shown below.

3.3.1.1 BRM

BRM is 5g tryptone, 5g yeast extract, 3g NaCl, 0.5g MgSO_4 per liter of sterile H_2O

3.3.1.2 RSM

RSM refers to a defined minimal medium for *S. meliloti* growth. Two similar but different versions of RSM were used in experiments – RSM1 and RSM2. RSM1 differs from RSM2 in two ways – RSM2 includes an addition of 0.1 mM NaCl and contains 1/5 the amount of CaCl_2 .

RSM1 is 10mM KH_2PO_4 (pH 6.5), 1mM $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 1mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 5mM NH_4NO_3 , 50 μM $\text{Na}_2\text{-EDTA} \cdot 2\text{H}_2\text{O}$, 50 μM $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 30 μM H_3BO_3 , 2.5 μM $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, 0.35 μM $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 0.4 μM $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, 0.6 μM CuSO_4 , and 0.1 μM CoCl_2). Sterilized stock solutions were added to sterile water, and bacterial inoculant was added to the medium prior to use. RSM1 was used for plant and RSM head to head competition experiments

RSM2 is 10mM KH_2PO_4 (pH 6.5), 0.1mM NaCl, 0.2 mM $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 1mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 5mM NH_4NO_3 , 50 μM $\text{Na}_2\text{-EDTA} \cdot 2\text{H}_2\text{O}$, 50 μM $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 30 μM H_3BO_3 , 2.5 μM $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, 0.35 μM $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 0.4 μM $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, 0.6 μM CuSO_4 , and 0.1

$\mu\text{M CoCl}_2$). Sterilized stock solutions were added to sterile water, and bacterial inoculant was added to the medium prior to use. RSM2 was used for the Tn-seq screen.

3.3.2 Tn-seq library construction

The *S. meliloti* Tn-seq master library was created by bi-parental mating of *S. meliloti* with *E. coli* strain MFD*pir* harboring pJG714. This *pir*-dependent plasmid contains a Tn5 transposon cassette in which the Km resistance gene points out one side, and the *Salmonella* P_{trp} promoter points out the other side. MFD*pir* is DAP dependent. The pJG714 sequence is given in the supplemental methods section. Mating plates (BRM-DAP) were incubated at 30°C for 6 hours and transposants were then selected on BRM-SmNm with no DAP. The library was plated at a density allowing for the collection of 100,000 independent colonies. The library was collected and frozen down in 1 ml aliquots of BRM-10% glycerol.

3.3.3 Tn-seq experimental conditions

Tn-seq experiments were conducted using an inoculation density of 1×10^9 CFUs of the Tn-seq master library, allowing for 10,000x coverage of the library across each condition.

3.3.3.1 Rich medium control

For each replicate, 10 ml of BRM-SmNm was inoculated with 1×10^9 CFUs of the Tn-seq master library and allowed to grow to saturation at 30°C. 1 ml was pelleted and frozen to for Illumina library preparation (2 biological replicates).

3.3.3.2 Minimal medium control

For each replicate, 500 ml of RSM2-Glucose-SmNm was inoculated with 1×10^9 CFUs of the Tn-seq master library and allowed to grow to saturation. 1 ml of saturated RSM2 was diluted into 10 ml BRM-SmNm. This was grown to saturation at 30°C, and 1 ml was pelleted and frozen for Illumina library preparation (2 biological replicates).

3.3.3.3 Plant inoculation and growth

Medicago truncatula plants were grown on turf (Turface Athletics; Therm-o-Rock West Inc.). Scarified and surface-sterilized seeds were allowed to germinate on Petri plates. For each replicate, 12 seedlings were planted in a Magenta vessel (Sigma). A second Magenta vessel was taped on top to provide a nearly closed system. After two weeks of growth, each Magenta vessel was flood inoculated with 2×10^9 CFUs of the Tn-seq master library added into 100 ml RSM2. After 20 minutes, excess water was removed by suction. Roots were collected 4 days after inoculation. Bacteria were collected from roots by gentle shaking at 30°C in BRM-SmNm (3 biological replicates).

3.3.3.4 Sequence library preparation

Samples were prepped for Illumina sequencing as described in detail in the Supplemental Methods under Tn-seq protocols. Briefly, DNA was extracted using a MoBio PowerSoil DNA Isolation Kit and subjected to enzymatic fragmentation. Fragments were C-tailed and transposon insertions were selectively amplified by PCR using primer tails including the necessary Illumina adapters and sequencing primer-binding sites. Samples were cleaned up by Solid Phase Reversible Immobilization (SPRI) bead size selection to remove fragments smaller than 200 bp. The resulting libraries had an estimated fragment size of 400 bp with a concentration of ~ 40

ng/ μ l. The master library and experimental libraries were sequenced with a 50 cycle, single read sequencing run on an Illumina HiSeq machine.

3.3.3.5 Reproducibility

Reproducibility of results was verified by X-Y plots for biological replicates, with R^2 values no lower than 0.9695.

3.3.4 Data analysis

Data analysis was performed using our custom TnSeq –Pipeline. TnSeq-Pipeline is a lightweight python script that facilitates the analysis of transposon sequence data, from fastq files to tallied transposon counts across a user supplied reference genome. It uses bowtie2 (Langmead, Trapnell, Pop, & Salzberg, 2009) for mapping and currently supports analysis of single end reads and an arbitrary number of experimental conditions; i.e. any number of fasta/fastq files can be run at once. It supports fuzzy transposon matching, reverse complimenting input reads, read normalization, and the removing of transposons located in the edges of genes to ensure only transposon events which lead to functional knockout are included in the analysis. Users configure these settings and more in a configuration file prior to running the program.

A typical TnSeq-Pipeline run consists of three steps: pre-alignment, alignment, and post-alignment/ SAM processing. In the pre-alignment stage fasta/fastq files are processed, transposon sequences are trimmed and sequences not containing the transposon sequence are removed. Reads containing the transposon pass on to the alignment stage. Alignment is done using the program bowtie2 (Langmead et al., 2009) with standard settings. As such this pipeline requires the user to install bowtie2 and create a special indexed reference file for their genome of interest. The third stage involves processing the SAM file output of bowtie2, and using user

supplied .ptt files of the reference genome, produces both a gene level and individual transposon event summary of transposon counts across all supplied chromosomes for all inputs. Tn-Seq Pipeline was able to process 82 million reads in 77 minutes on an iMac with 3.1 GHz Intel Core i5, 4 GB 1333 MHz DDR3.

One of the most powerful features of TnSeq-Pipeline is the IGV (Robinson et al., 2011) (Thorvaldsson, Robinson, & Mesirov, 2012) compatible output which can be loaded into IGV along with the reference genome allowing researchers to visualize differentially abundant transposon counts quickly and intuitively. TnSeq-Pipeline is distributed under the MIT license. Information on downloading, installing and running TnSeq-Pipeline can be found on github (<https://github.com/KBoehme/TnSeq-Pipeline>).

3.3.5 Fitness calculations

Essential gene candidates were selected by normalizing insertion frequencies in rich medium (BRM) for gene length, and identifying gene regions either absent of transposon insertions, or displaying a large under-representation of insertions (average insertions / gene length, with a cutoff of .12, after which point we began to see a more predictable number of insertions per gene length).

RSM2 and root fitness were calculated by comparing the average number of insertions at a location to the average number of insertions at that same location in the rich medium control (BRM) replicates. Significance was calculated using a standard two-tailed t-test.

3.3.6 Plasmid and strain construction

Plasmids and strains were created using standard molecular techniques with enzyme purchased from New England BioLabs and oligonucleotides from Life Technologies. PCR products were generated using Q5® High-Fidelity DNA Polymerase (New England BioLabs).

3.3.7 Competition assays

Plant competitions

M. truncatula plants were grown on turf as described previously. Scarified and surface-sterilized seeds were allowed to germinate on Petri plates. For each replicate, 1 plant was planted and immediately flood inoculated with 7 ml of RSM1 containing a 1:1 ratio of 1.4×10^4 CFUs of both mutant and wild type strains. Plants were grown for 2 weeks and roots were collected and vortexed vigorously in 1 ml LB. CFUs for mutant and wild type were counted by dilution plating. Fitness scores were calculated as follows:

$$\text{Average(Triplicate(Output: mutant/wt)/(Input: mutant/wt))}$$

Standard deviation refers to the deviation between Triplicate scores

RSM competitions

Saturated LB cultures of mutant and wild-type strains from overnight cultures were diluted and combined in a 1:1 ratio. RSM cultures supplemented with .5% glucose as a carbon source were inoculated with 1.4×10^4 CFUs of both mutant and wild-type strains and grown to saturation at 30°C. CFUs for mutants and wild type were counted by dilution plating. Fitness scores were calculated as follows:

$$\text{Average(Triplicate(Output: mutant/wt) / (Input: mutant/wt))}$$

Standard deviation refers to the deviation between Triplicate scores

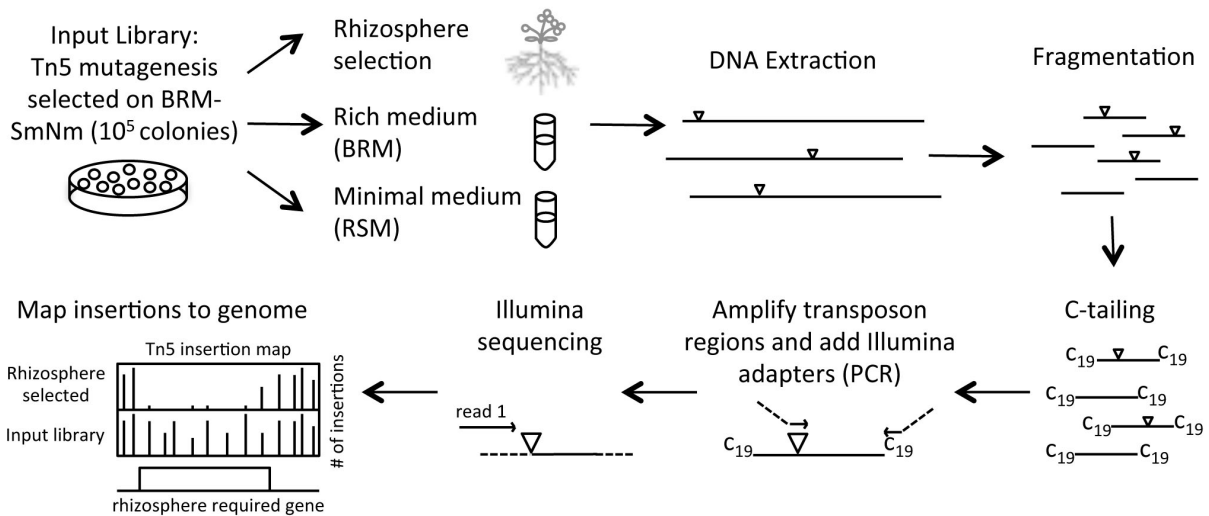


Figure 3-1 Schematic overview for the Tn-seq screen for effective rhizosphere colonization

3.4 Results

3.4.1 Tn-seq screen for rhizosphere-required genes

The experimental design for our Tn-seq screen for effective rhizosphere colonization is shown in Figure 3-1 and described in detail in the materials and methods section. A Tn-seq master library was generated on rich medium (BRM). The master library contained ~ 100,000 independent transposon insertions – an average of 1 insertion for every 86 base pairs. To provide a baseline for subsequent Tn-seq analysis, the library was grown out on BRM and DNA was collected and prepared for Illumina sequencing. To identify *S. meliloti* genes required for competitive colonization of *M. truncatula* roots, the library was selected on *M. truncatula* roots. Many genotypes may fail to colonize root systems due to a general failure to thrive under any restrictive conditions. To account for this, the library was also selected on RSM – a defined minimal medium supplemented with plant nutrients that was used to provide water for the plants (supplemented to .5% glucose to provide a carbon source). This control allows for a more confident determination of root-colonization-specific fitness defects. Genes required for this

condition but not for root conditions may also allow us to deduce how root systems influence the rhizosphere via root exudates.

Tn-seq allows for the simultaneous monitoring of each member of the transposon master library as the population is selected on conditions of interest. For our experiment, the master library was selected on BRM, RSM-glucose, and *Medicago truncatula* roots. After selection, transposon insertions were compared across all conditions and genes were categorized as representatives of five phenotypic categories: neutral genes (not required for any condition), essential genes (required for BRM), RSM-required genes (required on RSM only), RSM + root-required genes (required for both RSM and root), and root-required genes (required for root only)(Figure 3-4B).

Table 3-1 Illumina run and mapping statistics

Library	Number of reads	Number of reads containing transposon (%)	Number of reads mapped to genome (%)	Transposon insertions observed	Normalization coefficient
BRM_1	10,838,607	10,359,304 (96)	9,248,086 (89)	8,080,827	0.94638531427
BRM_2	12,093,164	10,396,796 (86)	8,872,397 (85)	7,743,191	0.98765173169
RSM_1	11,887,624	10,284,261 (87)	8,764,712 (85)	7,647,576	1.0
RSM_2	11,446,447	11,034,201 (96)	9,836,446 (89)	8,587,773	0.89051911362
A17_1	15,413,178	14,858,099 (96)	13,178,427 (89)	11,541,485	0.66261629244
A17_2	15,368,807	12,718,704 (83)	10,775,190 (85)	9,414,457	0.81232258004
A17_3	12,973,589	12,596,249 (97)	11,135,260 (88)	9,749,799	0.78438293958

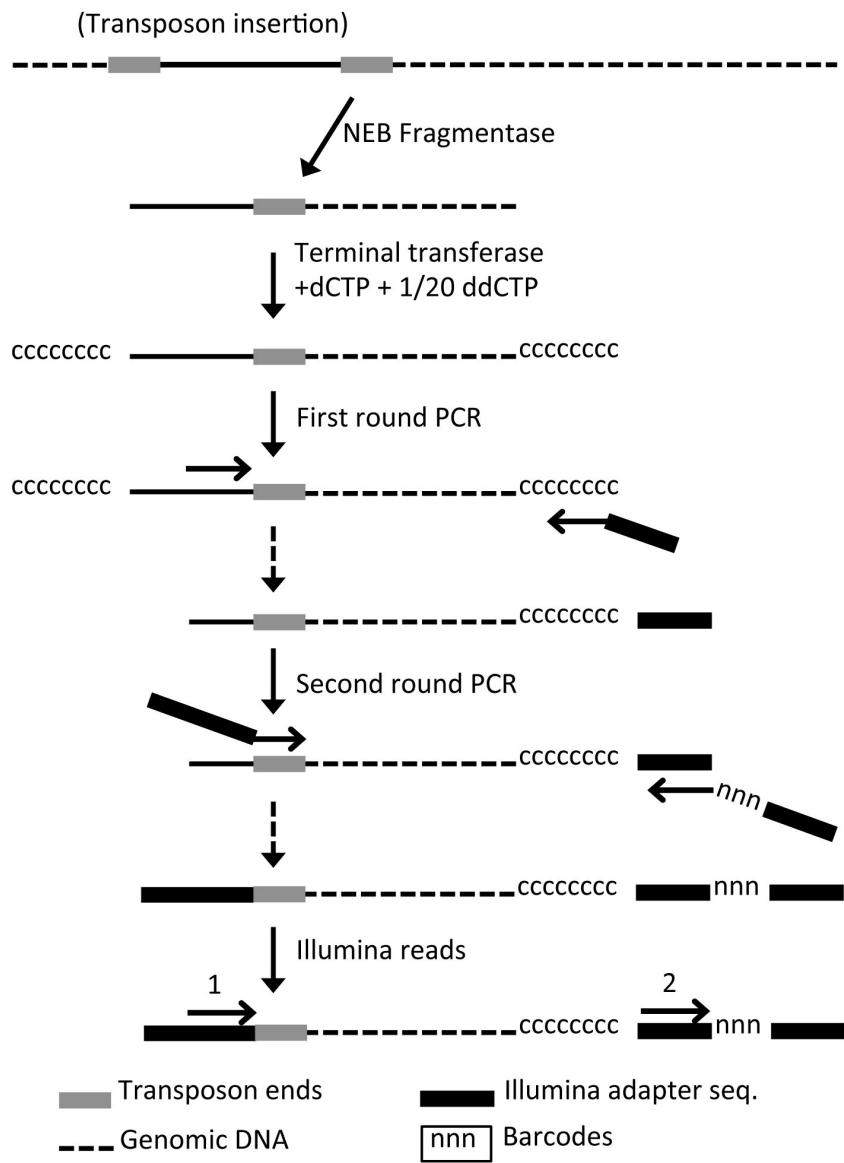


Figure 3-2 Schematic Illustration of Illumina library preparation steps

3.4.2 Summary of Tn-seq data collection and reproducibility of results

DNA was collected and prepared for Illumina sequencing as described in the materials and methods and as illustrated in figure 3-2. Briefly, the DNA was subjected to enzymatic fragmentation and C-tailing. Transposon insertions were amplified in a two-step PCR process, during which the necessary Illumina adapters were added to both the 5' and 3' ends of the sequence. The PCR products were cleaned up, size selected, and submitted for Illumina sequencing.

General statistics for read coverage, mapping efficiency, and insertion density are summarized in Table 3-1. About 10-15 million reads were generated for each sample, with 83-96% of these reads containing the expected transposon sequence. Of these reads, 85-89% mapped to a single location on the *S. meliloti* genome, indicating a high level of accuracy. In total we observe 7.6 – 11.5 million transposon insertions, representing an average of 76 – 115x coverage of our Tn-seq master library across all samples. Insertions are normalized across all conditions. Some sample visualizations of insertion frequencies are shown in Figure 3-3.

Transposon sites were mapped and tabulated using TnSeq-Pipeline – a custom python based pipeline I helped develop in collaboration with others in the lab (described in the materials and methods section). To test the reproducibility of our results, BRM and RSM inoculation, sequencing, and analysis were performed independently in duplicate (BRM_1, BRM2, RSM_1, and RSM_2), and the *M. truncatula* inoculation, sequencing, and analysis were performed in triplicate (A17_1, A17_2, and A17_3). All conditions showed extremely high reproducibility to one another in both insertion counts per gene (Figure 3-4A) and insertion counts per insertion site (data not shown).

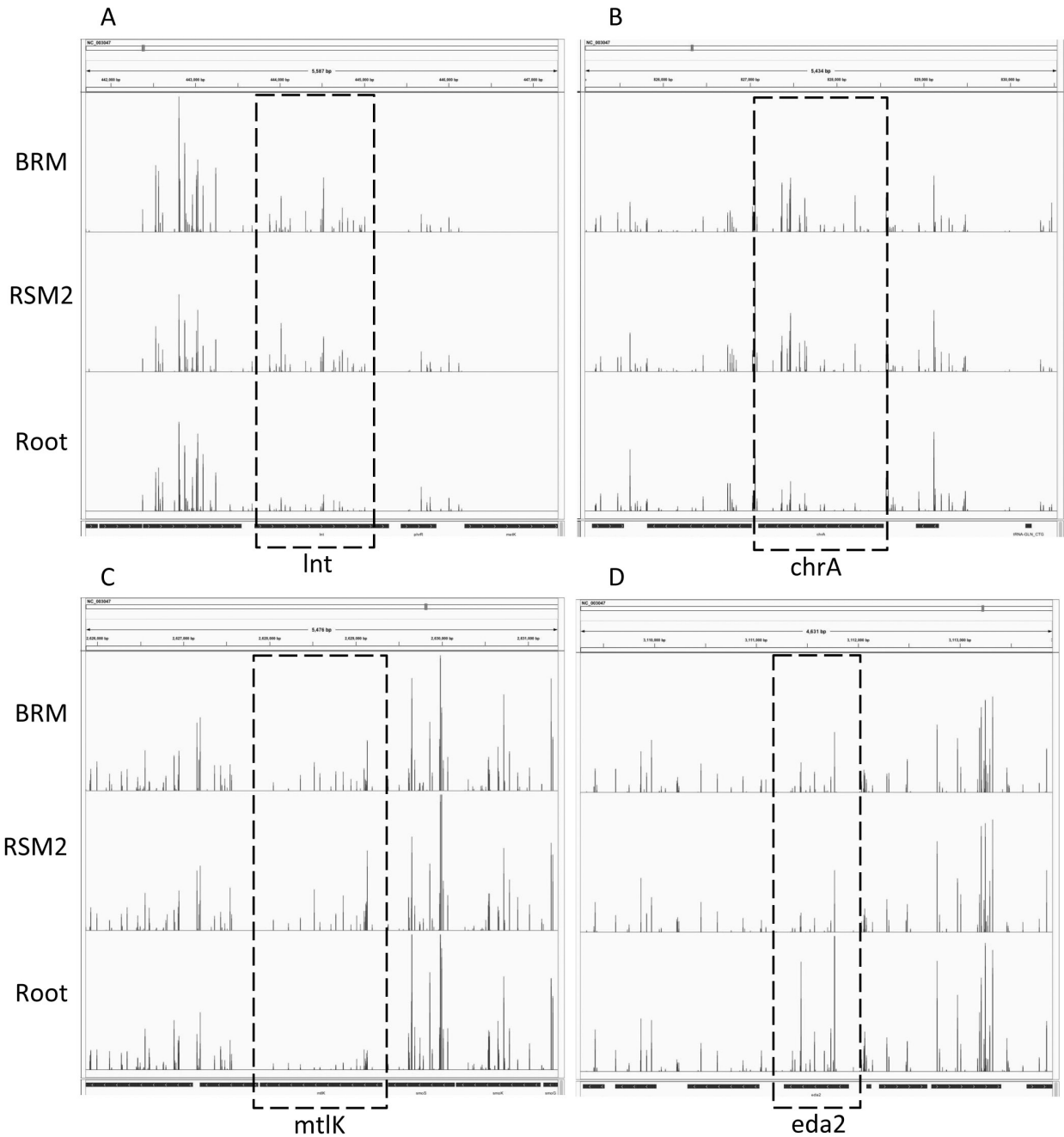


Figure 3-3 Visualization of insertion frequencies for selected genes on BRM, RSM2, and Root Conditions

Total insertion counts are averaged among replicates and visualized using IGV (Robinson et al., 2011) (Thorvaldsson et al., 2012).

A-C. Selected genes have significantly fewer insertions in root conditions than in RSM or BRM, suggesting these genes may be required for rhizosphere colonization

D. Selected gene has significantly more insertions in root conditions than in RSM or BRM, suggesting this gene may be detrimental for rhizosphere colonization

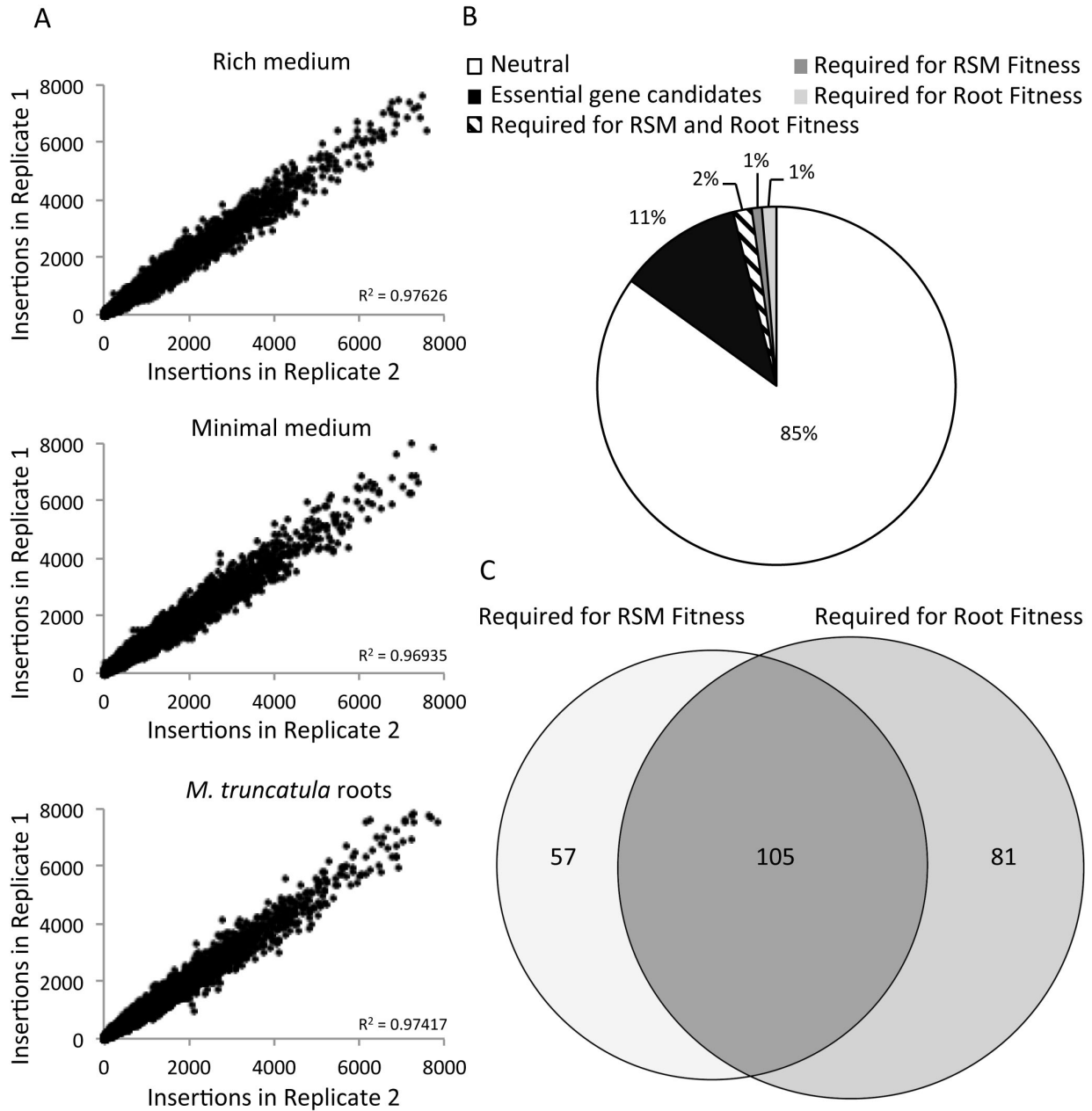


Figure 3-4 Tn-seq data reproducibility and phenotype analysis

A. X-Y plots of insertion data per gene between two independent biological replicates.

B. Pie chart shows percentage of insertions for listed phenotypes. For essential gene candidates, insertion numbers were normalized to gene length. Candidates had either no or very few insertions in the BRM condition (insertions / gene length < 0.12)

RSM and root fitness were determined by comparing normalized insertion numbers to the normalized insertion numbers in the BRM condition. Genes were designated as required if insertions in these genes result in at least a two-fold decrease in fitness (fitness score < 0.5)

C. Venn diagram shows the number of genes conferring poor RSM fitness, poor root fitness, or both. All genes represented in this diagram were well represented in the BRM condition.

Diagram made using BioVenn (Hulsen, de Vlieg, & Alkema, 2008).

3.4.3 Essential gene candidates

The generation of a Tn-seq master library serves to identify a large number of essential gene candidates. I categorized 692 genes as *S. meliloti* essential gene candidates based on a lack of insertions in the BRM condition, normalized for gene size (normalized insertions / base pair > 0.12) (Table S4)(Figure 3-4B). Within this list are many genes previously characterized as essential for bacterial viability, including genes required for DNA replication, RNA transcription, ribosomal subunits, lipid biosynthesis, etc. Also included are many hypothetical genes with uncategorized functions. The original insertion density of the Tn-seq master library (1 insertion / 86 base pairs) likely results in a number of false positives, particularly in smaller genes. However, many of these hypothetical genes are larger than 1,000 base pairs – large enough that insertions would most likely have been found if not for negative selection against mutations within these genes. These larger genes may therefore more confidently be characterized as essential gene candidates from this data set.

3.4.4 Tn-seq-predicted genes required for root fitness

Tn-seq analysis identified 186 genes in which disruptions resulted in at least a two-fold decrease in fitness on *M. truncatula* roots (Tables S5 and S6; Figure 3-4C). Of these, 105 genes are predicted to be required for both minimal medium (RSM) and root colonization (Table S5). This category will be referred to as root + RSM required. While such genes are predicted to play a critical role in rhizosphere colonization, the activity of these genes does not appear to be root-specific. As expected, many of these genes have reported roles in central metabolic processes. Almost 40% of the representatives on this list are involved in amino acid biosynthesis. The list of genes for this category is also enriched for genes encoding enzymes required for glucose

metabolism, including *gap*, *pykA*, *pgK*, *eno*, and *gpmA*. In *S. meliloti*, glucose catabolism is primarily through the Entner-Doudoroff pathway {Fuhrer, 2005 #217}, so the role of these enzymes in RSM and root fitness is presumably gluconeogenesis. The data suggest that glucose production is equally important for RSM growth (which is supplemented with .5% glucose) and for competitive root colonization. Genes involved in nucleotide biosynthesis (including *pur* and *pyr* genes) are also predicted to be required for both minimal medium and root conditions.

Beyond central metabolism, many other less expected biological pathways are predicted to be required for both root and RSM conditions. Among these are genes involved in DNA mismatch and damage repair, including *polA*, *recA*, *ruvA*, and SMc02760, encoding an ATP-dependent nuclease / helicase. Many cytochrome C biogenesis genes are also predicted to be required for both minimal and root conditions, including *ccmA*, *ccmB*, *ccmC*, *ccmG*, *fbcB*, *ctaD*, *cycK*, and *tlpA*. Interestingly, two genes encoding site-specific recombinase (*xerC* and *xerD*) appear to be root + RSM required. The data also predicts the importance of 9 hypothetical genes of unknown function for root and minimal medium colonization.

Tn-seq analysis identifies 81 genes predicted to be required for root colonization but not for growth in minimal medium (Figure 3-4C; Table S6). This category will be referred to as root-only required genes. These genes provide insight into how roots affect rhizosphere growth conditions by secretion of root exudates. Not surprisingly, this list includes a much larger percentage of hypothetical genes (about 32% of all predicted genes). A subset of cytochrome C biogenesis related genes (SMc02897, SMc01800, *ctaG*, *senC*) are predicted to be root-only required. The data also suggests a root-specific role for exopolysaccharide genes *exoX* and *exoB*. *exoX* is a regulator of succinoglycan synthesis, while *exoB* plays a role in exopolysaccharide modification. Interestingly, most other *exo* genes do not appear to be required rhizosphere

colonization. It may be that over-expression of succinoglycan or improper modifications of exopolysaccharides in *exoX* or *exoB* mutants over-stimulate plant immune responses and prevent colonization. The data suggests *lpsB* is also root-only required. *lpsB* is required for LPS core synthesis, confirming previous observations about the importance of LPS in competitive root colonization. It is noteworthy that mutations in many of the predicted root-only required gene networks, including cytochrome C biogenesis, *exo* genes, and LPS synthesis have been shown to disrupt symbiotic nitrogen fixation between *S. meliloti* and *M. truncatula*.

3.4.5 Tn-seq predicted genes required for RSM fitness only

57 genes are predicted to be required for growth in the RSM minimal medium, but not for root colonization (Figure 3-4C; Table S7). This category will be referred to as RSM-only. This interesting category of genes provides important information from which one can deduce which kinds of nutrients might be provided by *M. truncatula* root exudates. Amino acid biosynthesis genes for cysteine, (*cysD*, *cysG*, *cysH*, *cysK2*, *cysN*), aspartate (*aatA*), and arginine (*argD*) are shown to be required for minimal medium, but not for *M. truncatula* root fitness. This suggests that *M. truncatula* root exudates may be sufficient to sustain competitive colonization by *S. meliloti* auxotrophs for these amino acids. This is an interesting contrast to the prevailing notion that all amino acid biosynthetic pathways are required for rhizosphere colonization. In RSM there also appears to be a unique requirement for the high affinity zinc uptake ABC transporter (*znuA*, *znuB*, *znuC*), suggesting that *M. truncatula* root exudates increase the availability of zinc in the rhizosphere.

3.4.6 Tn-seq predicted genes more important for root than for RSM fitness

Fitness cut-offs described above are based on a 2-fold decrease in fitness from BRM to the condition of interest. A notable category of genes that does not satisfy this cut-off, but may still identify rhizosphere-related gene networks includes those that show at least a 30% decrease in fitness in the rhizosphere compared to RSM (Table S8). For example, the gene *thiF* was predicted to have a fitness score of 1.18 in RSM medium, but a fitness score of .65 in the root condition, representing a 53% drop in fitness from RSM to root conditions. Genes in which disruptions cause a drop in fitness from the RSM to the *M. truncatula* rhizosphere conditions may provide insight into how root exudates influence the rhizosphere. This category will be referred to as RSM > root. Within this category are many genes involved in cyclic glucan secretion, including *feuP*, *feuQ* and *ndvA*. Many more *exo* genes are also predicted to be in this RSM > root category. Secretion of cyclic glucans and exopolysaccharides has previously been reported to play a role in symbiotic nitrogen fixation. Another interesting gene predicted to be more important for RSM than for root fitness is *bluB*. *bluB* is necessary for biosynthesis of vitamin B12, suggesting that B12 is less available in the root condition than in the RSM condition. Perhaps competition for resources between the microbial community and the root tissue increases the importance of vitamin B12 biosynthesis in the rhizosphere.

3.4.7 Selection of candidates for Tn-seq prediction verification

To verify the accuracy of my Tn-seq predicted phenotypes, a panel of genes was selected for head to head rhizosphere competition assays against wild *S. meliloti*. Genes selected for testing are representatives of the previously discussed root-required phenotypes: root + RSM, root-only, RSM-only, and RSM > root. In addition to these, I selected a small group of gene candidates for testing that were predicted to be important for rhizosphere colonization based on a preliminary Tn-seq study (preliminary Tn-seq data not shown). These selected genes and their associated Tn-seq insertion data and predicted fitness calculations are summarized in Table 3-2. Reported insertions represent the average number of gene insertions between biological replicates. Mutants with internal fragment disruptions within genes on this panel were tested for competitive colonization in head to head competition against wild type *S. meliloti* in both RSM-glucose and on *M. truncatula* roots as described in the materials and methods.

Table 3-2 Selected genes to be disrupted for competition tests.

Fitness phenotype	GeneID	Gene	Function	Rich Medium	Minimal Medium	Roots	RSM Fitness ^δ	Root Fitness ^δ
Neutral	SMc03003	rhaK	sugar kinase	2636.5	2201	2109.7	0.83	0.80
Required For Growth in RSM and Root Conditions (Fitness <= 0.5)	SMc00010	ctaD	cytochrome C oxidase polypeptide I transmembrane protein	616.5	305.5	250.7	0.50	.41**
	SMc00784	fbpA	iron binding protein	486.5	168.5	239.3	0.35**	.49**
	SMc01770	glyA	serine hydroxymethyltransferase	1788.5	1.5	473.7	0**	.26**
	SMc02569	hisF	imidazole glycerol phosphate synthase subunit HisF	1138	26.5	7.3	0.02**	.01**
	SMc02572	hisH	imidazole glycerol phosphate synthase subunit HisH	951	38.5	2.0	0.04**	.00**
	SMc02725	trpE	anthranilate synthase	2318.5	4	406.7	0**	.18**
	SMc03776	proB1	gamma-glutamyl kinase	4945	764.5	2256.7	0.15**	.46**
	SMc03795	leuD	isopropylmalate isomerase small subunit	305.5	1.5	62.7	0**	.21*
	SMc03823	leuC	isopropylmalate isomerase large subunit	2559.5	51	425.3	0.02**	.17**
	SMc03848	ccmB	heme exporter B (cytochrome C-type biogenesis protein) transmembrane	645.5	310	223.3	0.48	.35*
SMc03979	gap	glyceraldehyde-3-phosphate dehydrogenase	1125	15.5	310.0	0.01*	.28*	
SMc04026	glTD	glutamate synthase	3274.5	9	247.3	0**	.08**	
Required For Growth in Root Conditions Only (Fitness <= 0.5)	SMB21265	redB	glycosyltransferase	3205	2361	1392.0	0.74	.43**
	SMc00175	-	ABC transporter ATP-binding protein	523	337	175.7	0.64	.34**
	SMc00349	lepA	GTP-binding protein LepA	444	379.5	218.3	0.85	.49**
	SMc00419	gshB1	glutathione synthetase	108	86	11.0	0.80	.10**
	SMc00808	chrA	chromate transporter	2105.5	2027	624.0	0.96	.30**
	SMc01041	dusB	NIFR3-like protein	439	373	203.3	0.85	.46**
	SMc01111	Int	apolipoprotein N-acyltransferase	1551.5	1431	422.3	0.92	.27**
	SMc01501	mtkK	mannitol 2-dehydrogenase	1245.5	1625.5	444.7	1.31	.36**
	SMc01595	-	sensor histidine kinase transmembrane protein	698.5	363	38.7	0.52	.06**
	SMc02435	hemK1	methyltransferase	395	237.5	100.7	0.60	.25**
	SMc02520	glpD	glycerol-3-phosphate dehydrogenase	113.5	235	18.0	2.07	.16**
	SMc02641	rkpK	UDP-glucose 6-dehydrogenase	2326	1244.5	500.7	0.54	.22**
SMc03884	ispA	geranyltranstransferase	240.5	337	96.0	1.40	.40**	
More Important For Growth in Root Conditions than in RSM Conditions (Fitness drop <= 0.3)	SMB20290	-	transcriptional regulator	662	843	587.0	1.27	.89**
	SMc00166	bluB	oxidoreductase NAD protein	316	350.5	200.3	1.11	0.63
	SMc00238	moeA	molybdopterin biosynthesis protein	582.5	773	386.3	1.33	.66*
	SMc00458	feuP	transcriptional regulator	311.5	722	398.3	2.32	1.28*
	SMc00951	-	periplasmic solute binding protein	863	1119.5	708.0	1.30	.82*
	SMc02864	thiF	molybdopterin biosynthesis protein MoeB	880	1034.5	569.7	1.18	.65*
	SMc02877	-	sugar isomerase	1407.5	1460.5	711.3	1.04	.51**
	SMc03900	ndvA	cyclic beta-1,2-glucan ABC transporter	683	1488	854.3	2.18	1.25**
SMc04114	pilA1	pilin subunit protein	45.5	63	26.7	1.38	.59*	
Required for Growth in Root Conditions in Preliminary Tn-seq Study	SMc01507	-	sensor histidine kinase	765.5	218	719.0	0.28	0.94
	SMc02361	cycH	cytochrome C-type biogenesis transmembrane protein	857.5	779.5	548.0	0.91	0.64
	SMc02047	gcvI	glycine cleavage system aminomethyltransferase T	1105	860.5	611.3	0.78	0.55
Detrimental to Growth in Root Conditions	SMc03153	eda2	keto-hydroxyglutarate-aldolase/keto-deoxy-phosphogluconate aldolase	831	630.5	2418.0	0.76	2.91
	SMc04134	-	transcriptional regulator	452	422.5	1273.0	0.93	2.82

Reported insertion points are averages between all replicates

^δ Fitness determined by average insertions / average rich medium insertions

* $P < .05$

** $P < .005$

3.4.8 Verification tests for genes predicted to be required for root fitness

The competition tests for genes predicted to be required for our three phenotypes - required for fitness on RSM and root, required for root only, or more important for root than RSM conditions - are reported in figure 3-5. Of the 12 genes tested that were predicted to be required for root and RSM competition by Tn-seq analysis, 7 competed poorly or failed to grow in the head to head competition experiments in both conditions. Most of these accurately predicted phenotypes were for amino acid auxotrophs. Interestingly, two leucine auxotrophs tested competed well in root competitions, suggesting *M. truncatula* root systems were able to sustain these auxotrophs during colonization. A total of 9 of the tested strains for this predicted category failed to competitively colonize *M. truncatula* roots.

I also tested 16 genes that were predicted to be required for root colonization, but not for RSM growth. Of these genes, 5 were confirmed to have root-specific colonization phenotypes (*hemK1*, *rkpK*, *chrA*, *cycH*, and *gcvT*) and 3 performed poorly in both root and RSM competitions (*bshB1*, *lepA*, and SMc01507, encoding an uncharacterized sensor kinase).

I tested 9 genes predicted by Tn-seq data to be more important for root colonization than for RSM growth. Of these, 5 genes were shown to play a major role in rhizosphere colonization (*bluB*, *moeA*, *feuP*, *thiF*, and *ndvA*). This data suggests that a drop of at least 30% in fitness from RSM to root conditions is a useful predictor for colonization phenotypes.

In all, the Tn-seq data accurately predicted a significant decrease in root fitness (at least two-fold less than wild type) for mutations in 21 of 36 genes tested compared to wild type (Table 3-3). Seven additional mutations were accurately predicted to decrease root fitness, though for

these genes the decrease in fitness was much more modest. Taken together, the Tn-seq prediction accuracy rate for root fitness phenotypes fell between 58-75%.

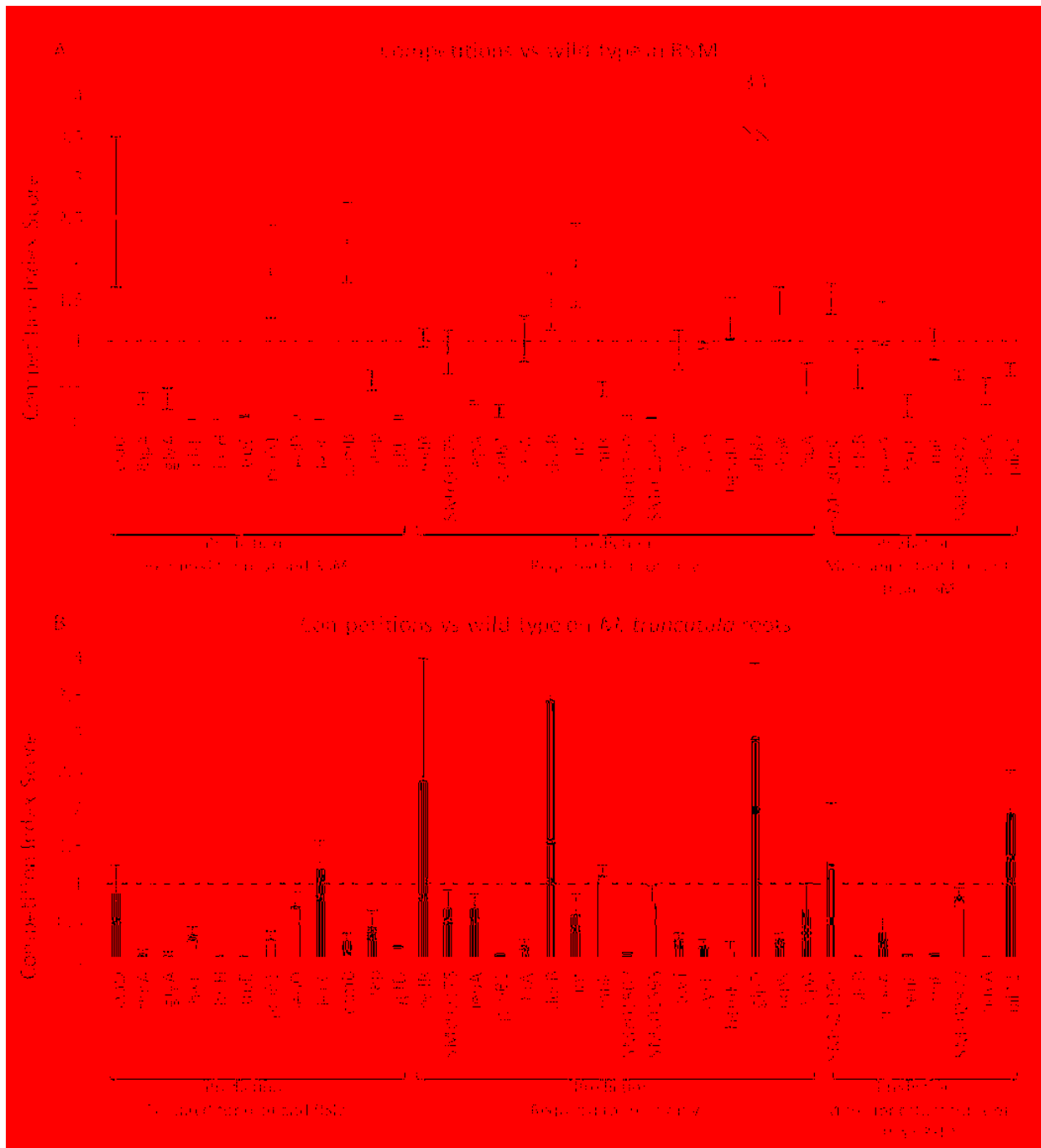


Figure 3-5 Competition experiments comparing mutant growth to wild type growth in RSM and *M. truncatula* root conditions

Competition index scores were computed by dividing observed mutant colonies by observed wild type colonies after competition. The dotted lines are shown as a reference for wild type performance in each condition.

Table 3-3 Summary of genes verified to play a role in competitive root colonization

Gene	Function	Competition index
<i>hisH</i>	imidazole glycerol phosphate synthase subunit HisH	0.000
<i>trpE</i>	anthranilate synthase	0.000
<i>bluB</i>	oxidoreductase NAD protein	0.000
<i>ndvA</i>	cyclic beta-1,2-glucan ABC transporter	0.003
<i>thiF</i>	molybdopterin biosynthesis protein MoeB	0.020
<i>gshB1</i>	glutathione synthetase	0.027
<i>glyA</i>	serine hydroxymethyltransferase	0.043
<i>feuP</i>	transcriptional regulator	0.044
<i>SMc01507</i>	sensor histidine kinase	0.052
<i>fbpA</i>	iron binding protein	0.055
<i>hemK1</i>	methyltransferase	0.110
<i>gltD</i>	glutamate synthase	0.134
<i>chrA</i>	chromate transporter	0.164
<i>cycH</i>	cytochrome C-type biogenesis transmembrane protein	0.182
<i>ccmB</i>	heme exporter B (cytochrome C-type biogenesis protein) transmembrane	0.216
<i>rpkK</i>	UDP-glucose 6-dehydrogenase	0.255
<i>proB1</i>	gamma-glutamyl kinase	0.260
<i>gcvT</i>	glycine cleavage system aminomethyltransferase T	0.264
<i>hisF</i>	imidazole glycerol phosphate synthase subunit HisF	0.312
<i>moeA</i>	molybdopterin biosynthesis protein	0.335
<i>gap</i>	glyceraldehyde-3-phosphate dehydrogenase	0.430

Competition index reports results from head to head competition experiments of disrupted-gene mutants vs. wild type *S. meliloti* on *M. truncatula* root systems.

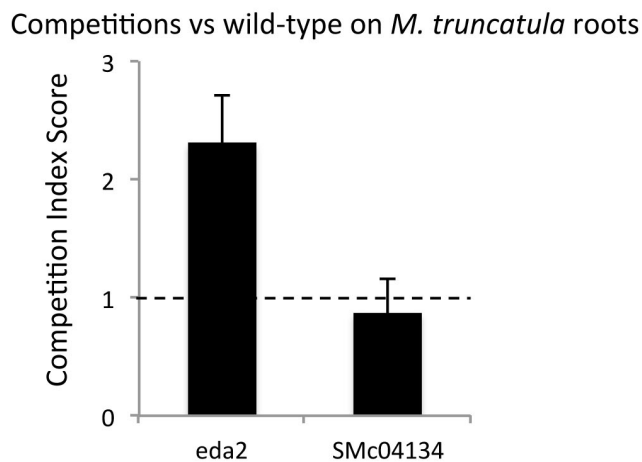
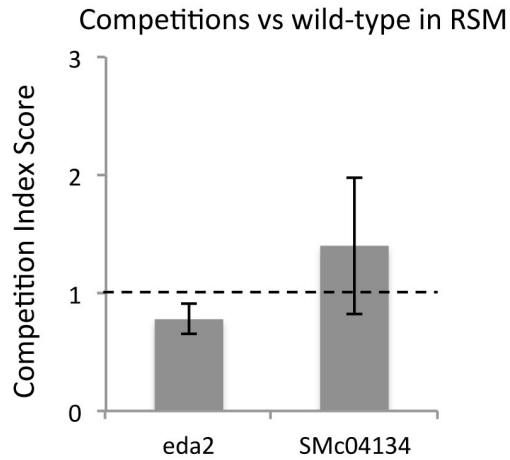


Figure 3-6 Competition experiments comparing mutant growth to wild type growth in RSM and *Medicago truncatula* root conditions. Competition index scores were computed by dividing observed mutant colonies by observed wild type colonies after competition. The dotted lines are shown as a reference for wild type performance in each condition.

3.4.9 Verification tests for mutations predicted to improve root fitness

Two mutations were tested which were predicted by Tn-seq data to increase root fitness compared to wild type (Figure 3-6). In one instance, the Tn-seq data accurately predicted an increase in root fitness for the mutant over wild-type cells (*eda2*). *eda2* encodes a KHG/KDPG aldolase that converts KDPG into pyruvate in the Entner-Doudoroff pathway for glucose degradation (Patil & Dekker, 1992). Interestingly, the *zwf* gene (also required for the Entner-Doudoroff pathway) is also predicted to be detrimental to root fitness (Tn-seq calculated root

fitness for *zwf* insertion mutants = 1.57). In *S. meliloti*, the Entner-Doudoroff pathway is the primary method for glucose catabolism, while the Embden-Meyerhof (glycolysis) pathway is non-functional or extremely limited (Stowers, 1985). Fast growing *S. meliloti* strains have been observed to also utilize the pentose phosphate pathway for glucose catabolism. Disruption of the Entner-Doudoroff pathway therefore likely leads to an increase in pentose phosphate activity. This change would result in increased production of reduced NADPH and pentose sugars, which are precursors for the synthesis of nucleic acid components. Nucleic acid biosynthesis is predicted by the Tn-seq data to play a critical role in rhizosphere colonization. An increased reliance on manufactured nucleic acids may explain why disrupting the Entner-Doudoroff pathway would increase *S. meliloti* fitness in the rhizosphere.

As discussed previously, genes involved in gluconeogenesis are predicted to be required for rhizosphere colonization. Perhaps when the Entner-Doudoroff pathway is intact, accumulated glucose from gluconeogenesis is required for processing through the pentose phosphate pathway.

An insertion disruption in the putative transcriptional regulator SMC04134 was also tested in our competition assay vs. wild-type cells. Although Tn-seq data indicated a nearly 3-fold increase in fitness for insertions within this gene, the competition assays did not confirm this prediction. As discussed previously, it may be that this gene does not play a significant role in root colonization. It may also be that slight differences in conditions between our competition assay and the original Tn-seq experiment result in a difference in phenotypes for this gene disruption.

3.4.10 Additional exploration of genes in the FeuPQ two-component system

Our lab has previously conducted extensive studies of the FeuPQ two-component system, and we have established that FeuPQ signaling plays a role in infection thread development during nodulation in the rhizobia-legume symbiosis (Griffitts et al., 2008). I was surprised to find that the FeuPQ genes and many of their downstream targets also appear play a significant role in competitive rhizosphere colonization. Previous studies established that the abortive nodulation phenotype observed in *feuP* mutants is actually due to aberrant cyclic glucan secretion due to misregulation of *ndvA*, a gene under control of FeuPQ signaling that encodes for a cyclic glucan exporter. The abortive nodule phenotype of *feuP* mutants can be rescued by over-expression of *ndvA* (refer to Figure 2-2).

Over-expression of *ndvA* might similarly restore competitive rhizosphere colonization in *feuP* mutants. The fact that *ndvA* mutants had a root colonization phenotype very similar to *feuP* mutants lends credence to this notion. To learn if *feuP* involvement in competitive rhizosphere colonization is actually due to downstream misregulation of *ndvA*, a Δ *feuP* strain and a Δ *feuP* strain over-expressing *ndvA* from a plasmid (Δ *feuP* *P**trp*::*ndvA*) were both competed against wild type *S. meliloti*. For reference, the neutral control *rhaK*::*neo* (the Nm resistance gene was looped into the end of *rhaK*, reconstituting the gene as part of the loop-in process) was also included in the experiment. Interestingly, over-expression of *ndvA* does not restore the phenotype of Δ *feuP* strains for competitive root colonization as it did for nodulation development (Figure 3-7). This suggests that the phenotype is caused by mechanisms independent of cyclic glucan export, and that the mechanisms for FeuPQ involvement in competitive root colonization and infection thread invasion are not identical.

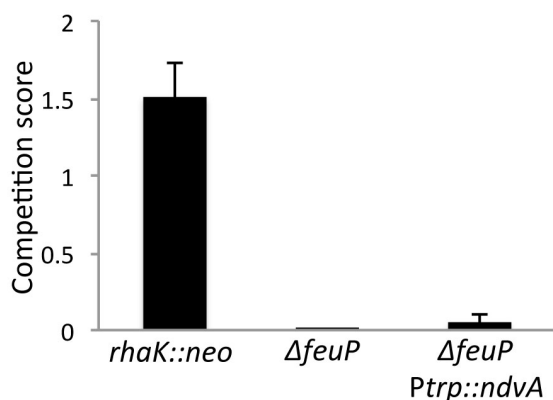


Figure 3-7 *ndvA* over-expression does not restore competitive rhizosphere colonization to $\Delta feuP$ mutants.

The neutral *rhaK::neo*, $\Delta feuP$, and $\Delta feuP$ *Ptrp::ndvA* strains were competed against wild type in our root colonization assay. Error bars indicate variance between three biological replicates.

3.5 Discussion

Tn-seq analysis has implicated a set of *S. meliloti* genes likely to be involved in competitive root colonization. Subsequent head to head competition tests comparing wild type to mutant populations have verified a role in root colonization for many of these genes. These genes provide insight into how plants affect the bacterial community within the rhizosphere and how rhizosphere bacteria adapt to form specialized symbiotic relationships with compatible host plants.

Our analysis confirms the importance of many genes previously reported as determinants of successful root colonization. The Tn-seq data for the rhizosphere condition shows a general depletion of insertions in genes related to amino acid synthesis, biotin synthesis, and LPS synthesis – all consistent with previously reported studies on competitive root colonization done in rhizobia and *Pseudomonas* (Simons et al., 1997) (Streit et al., 1996) (Scheidle et al., 2005).

Though previous studies have shown a defined role for chemotaxis related genes in competitive root colonization (Capdevila et al., 2004), our Tn-seq analysis did not show significant depletion of insertions in genes associated with flagella synthesis and motility. This is likely because our root inoculation density was sufficiently high enough to guarantee proximity to root tissue for every representative of our library, thus removing a chemotaxis or motility requirement. If so, a more dilute inoculation would likely begin to display the importance of motility related genes.

For our competition assays, I selected candidates for testing that exhibited at least a two-fold depletion of insertions in the root condition. In some cases, head to head competitions did not show a fitness defect in the tested mutants, even when three biological replicates of data from our Tn-seq analysis each indicated negative selection against insertions within these genes. A number of other Tn-seq studies have also had similar difficulty in predicting phenotypes accurately with a cutoff threshold of ~ 2-5 fold reduction in fitness (Gallagher et al., 2011) (Yung et al., 2015). It may be that a more stringent requirement for fitness reduction would increase the accuracy of our results. However, this approach would likely also increase the risk of filtering genes out of the analysis that do play a significant role in root colonization.

In some cases, the difficulty of verifying Tn-seq predictions may in part result from a difference in the ratios of deficient mutant to wild type cells in the experimental setup. In our Tn-seq experiment, a deficient mutant represented only one 100,000th of the overall population in the rhizosphere, while in head to head competitions the mutant represented one half of the population. Perhaps in some cases this difference in input ratios is responsible for the difference in phenotypic outcome.

Our analysis has identified 21 *S. meliloti* genes as playing a major role in competitive root colonization of *Medicago truncatula*. Some of these genes fall into predicted categories of genes known to be required for competitive root colonization, while others have not been previously implicated in rhizosphere colonization functions. Our analysis predicted many amino acid synthesis genes, vitamin synthesis genes, and LPS-related genes, lending credibility to Tn-seq as a method for identifying rhizosphere-required genes. Interestingly, our analysis also highlights a number of genes required for *S. meliloti* competitive colonization of *M. truncatula* root systems that have been previously determined to play a critical role for nodulation and symbiotic nitrogen fixation, including cytochrome C biogenesis genes, *exo* genes, two-component signaling systems, and genes involved in cyclic glucan expression. In the case of cyclic glucan related genes, the mechanisms for involvement in root colonization appears to be distinct from those for nodulation. These Tn-seq identified candidate genes are promising leads to help researchers better understand both general and specialized functions required for competitive rhizosphere colonization.

3.6 Acknowledgements





This work was supported by the NSF grant IOS-1054980 CAREER: The molecular basis of abortive symbiosis.

Chapter 4. Future work and Conclusions

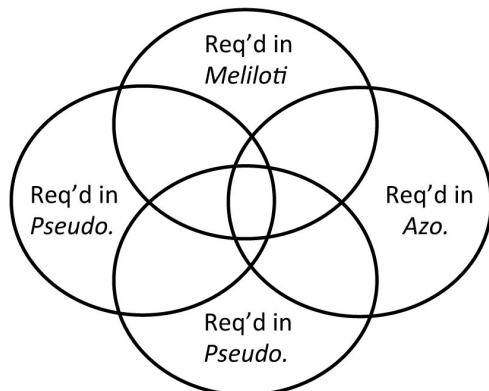
4.1 Continuing Tn-seq analysis

The previous chapter investigating *Sinorhizobium meliloti* colonization of the *M. truncatula* rhizosphere has only scratched the surface in terms of uncovering genes required for general rhizosphere competency. The nature of plant-specific root exudates and the types of the associated microbial communities heavily influence conditions in the rhizosphere. It remains to be seen how root systems from different plant species and the study of different microbial species might influence the requirements for rhizosphere colonization.

Additional Tn-seq experiments looking at *S. meliloti* and other bacterial rhizosphere colonization on a panel of different plants, including rice and tomato would allow us to determine if there is a core overlapping set of genes in common for general rhizosphere colonization. We may also identify non-overlapping gene requirements specific to a given plant host, highlighting specialist functions between specific plant-microbe pairs. Such studies could allow us to deduce how a given plant species shapes the rhizosphere with unique root exudates. Figure 4-1 illustrates the usefulness of an expanding panel of microbes and plants whose determinants of competitive rhizosphere colonization have been tested via Tn-seq analysis.

	 Input: Rich med.	 Output: RSM+Glc	 Output: <i>Medicago</i>	 Output: Rice
<i>Pseudo.</i>	Tn-Seq_01	Tn-Seq_02	Tn-Seq_03	Tn-Seq_04
<i>Bacillus</i>	Tn-Seq_05	Tn-Seq_06	Tn-Seq_07	Tn-Seq_08
<i>Meliloti</i>	Tn-Seq_9	Tn-Seq_10	Tn-Seq_11	Tn-Seq_12
<i>Azo.</i>	Tn-Seq_13	Tn-Seq_14	Tn-Seq_15	Tn-Seq_16

Identification of bacteria-specific colonization genes



Identification of host-specific colonization functions

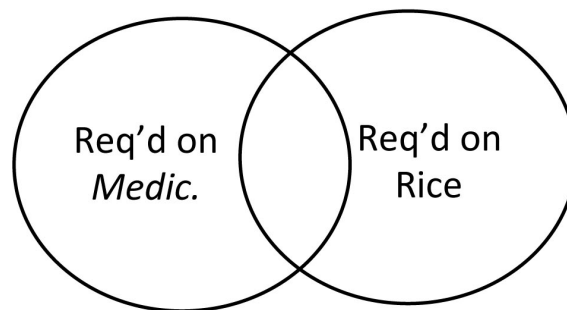


Figure 4-1 Use of an expanding panel for identification of general and host-specific required genes for rhizosphere colonization

Above: Sample panel of Tn-seq tests testing *Pseudomonas*, *Bacillus*, *Meliloti*, and *Azospirillum* strains on BRM, RSM, *Medicago* roots, and Rice roots.

Below: Venn diagrams - Specialized strain and host-specific determinants of rhizosphere colonization will be found in the excluded regions, while general determinants for rhizosphere colonization will be found in the overlapping regions.

The Tn-seq analysis of additional plant and bacterial species is already underway. Collaborators at Harvard University have collected data from Tn-seq analysis of *Pseudomonas fluorescens* strain WCS365 on *M. truncatula* roots (so far I have not had access to this data). In our lab, we have generated Tn-seq data looking at *Sinorhizobium meliloti* colonization of Rice roots. The analysis of *S. meliloti* colonization of rice plants is not entirely contrived – interestingly, *S. meliloti* is a great colonizer of rice, producing more CFUs per gram of root tissue on rice than on *Medicago* roots. Preliminary analysis of Tn-seq data comparing root colonization of *S. meliloti* on *M. truncatula* and rice roots reveals required specialist genes for each plant, as well as overlapping genes required for general root colonization (Figure 4-2, right).

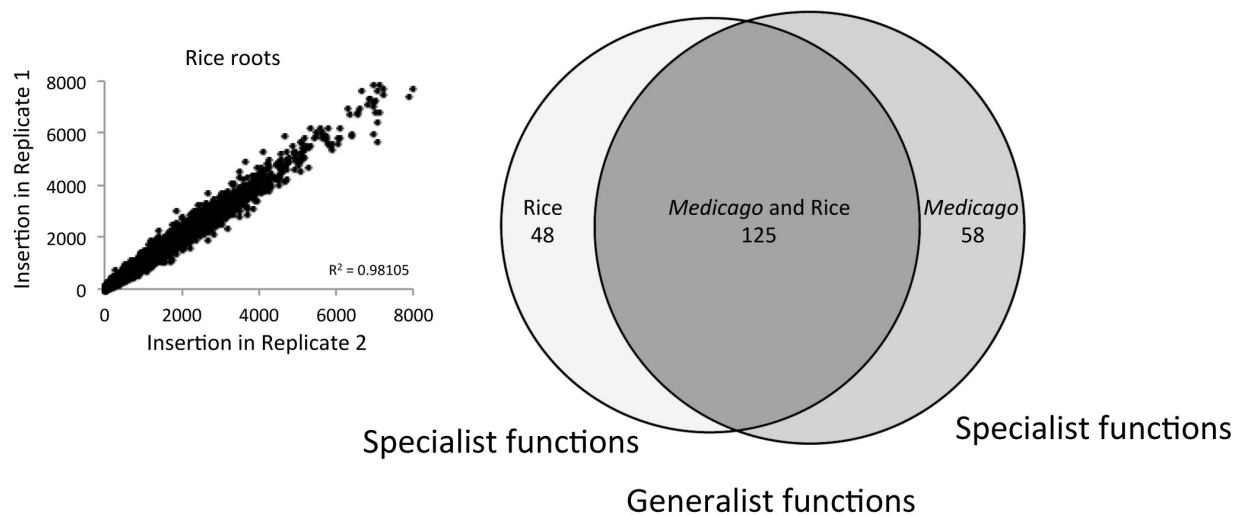


Figure 4-2 Tn-seq data reproducibility and phenotype analysis

A. X-Y plots of insertion data per gene between independent biological replicates.

C. Venn diagram shows the number of genes required for Rice fitness, *Medicago* fitness, or both. All genes represented in this diagram were well represented in the BRM condition. Diagram made using BioVenn (Hulsen et al., 2008).

Conditions for this Tn-seq experiment were very similar to those of the previous *S. meliloti* – *M. truncatula* experiment. Rice was inoculated in three independent biological replicates with the same density of *S. meliloti* Tn-seq master library used in previous experiments. The reproducibility of insertion representation within the three biological replicates extremely high, as was observed in previous experiments (Figure 4-2). Comparing Tn-seq predicted gene requirements between *S. meliloti* colonization of Rice and *S. meliloti* colonization of *M. truncatula* highlights 58 genes predicted to be specifically required for colonization of *Medicago truncatula* root systems (Table S9), 48 genes predicted to be specifically required for colonization of Rice root systems (Table S10), and 125 overlapping genes predicted to be required for general root colonization (Table S11).

Among genes required specifically for colonization of *M. truncatula* but not rice, it is interesting to find genes related to nodulation and symbiotic nitrogen fixation, including *exo* genes (*exoX* and *exoB*) and cytochrome C biogenesis genes (*ccmA*, and *ctaG*). There is also a *M. truncatula*-specific requirement for cobalamine biosynthesis (*cobC*), suggesting that cobalamine may be more scarce in the rhizosphere of *M. truncatula* than for rice. In this list, Tn-seq analysis implicates a number of hypothetical genes of unknown function that appear to be specific to *M. truncatula* root systems. Study of these genes may reveal more information about specialist functions for *S. meliloti* colonization of *M. truncatula* root systems.

For genes required for Rice but not *M. truncatula*, there is an enrichment for genes involved in glucose and sucrose metabolism. This is interesting in light of previously mentioned studies of *Pseudomonas* colonization of tomato roots that have suggested that the ability to metabolize basic root exudate sugars does not influence rhizosphere competitiveness (Lugtenberg et al., 1999). Vitamin B6 and metal/ion binding genes also appear to impact fitness

only in the rice condition. In this condition there appears to be a difference in amino acid biosynthesis requirements. As mentioned in chapter 3, insertions in genes required for biosynthesis of cysteine (*cysN*, *cysD*, *cysH*, *cysG*, and *cysK2*), aspartate (*aatA*), and arginine (*argD*) do not appear to be required for *S. meliloti* colonization of *M. truncatula* root systems. However, these genes appear to play a much more significant role in competitive colonization of rice roots (Table 4-1). The data suggests that *M. truncatula* may provide these amino acids in quantities sufficient to support the competitive colonization of auxotrophs for cysteine, aspartate, and arginine, while rice roots do not.

Comparing both rice and *M. truncatula* fitness to fitness in the minimal RSM control, it appears that both *M. truncatula* and rice root exudates may facilitate proline auxotroph colonization, as *proC* mutants do significantly better on both root systems than in the root-free RSM control condition (Table 4-1).

Table 4-1 Summary of Tn-seq data for selected auxotrophs predicted to be able to colonize root systems

GeneID	Gene	BRM	RSM	<i>M. truncatula</i>	Rice	RSM Fitness	<i>M. truncatula</i> Fitness	Rice Fitness
SMc00090	<i>cysN</i>	1925.5	96.5	1980.7	991.7	0.05	1.03	0.52
SMc00091	<i>cysD</i>	2070	192	2268.3	993.7	0.09	1.10	0.48
SMc00092	<i>cysH</i>	401	4	721.7	282.0	0.01	1.80	0.70
SMc01053	<i>cysG</i>	3256	1205	2860.7	1827.3	0.37	0.88	0.56
SMc01174	<i>cysK2</i>	1754	802.5	977.7	1250.7	0.46	0.56	0.71
SMc01578	<i>aatA</i>	1084.5	383	596.0	198.0	0.35	0.55	0.18
SMc02138	<i>argD</i>	2166	611.5	1119.0	663.7	0.28	0.52	0.31
SMc02677	<i>proC</i>	372.5	165.5	296.3	286.0	0.44	0.80	0.77

4.2 Future work

Tn-seq analysis and subsequent verification tests have revealed a role for many genes in competitive rhizosphere colonization. Data collected in Tn-seq experiments for *S. meliloti* on *Medicago truncatula* and rice suggest overlapping gene functions required for general rhizosphere colonization, as well as individual genes that play a role in host-specific colonization. Of particular interest for continuing study, the data suggest that *S. meliloti* root exudates may provide sufficient cysteine, arginine, and aspartate to support root colonization for auxotrophs of these amino acids, and that root exudate sugar metabolism may determine competitive rhizosphere colonization of rice root systems. Both of these observations are somewhat surprising in light of generally accepted principles of rhizosphere colonization previously discussed. Additional research is required to verify these Tn-seq predicted phenotypes and understand the mechanisms by which predicted genes affect fitness during root colonization.

The Tn-seq analysis implicates many new genes predicted to play a significant role in rhizosphere colonization, including two additional previously uncharacterized sensor kinases that play a role in competitive rhizosphere colonization. Genetic analysis of these systems would include experiments similar to those conducted in the study of the FeuPQN system reported in chapter 2. These sensor molecules may provide key insights toward discovering plant molecules that may play an encouraging or discriminatory role in competitive rhizosphere colonization.

Finally, the Tn-seq studies have highlighted a large number of hypothetical proteins as predicted essential genes, essential genes for growth in minimal medium, or genes required for colonization of *M. truncatula* and/or rice root systems. Further study is needed to determine the role these previously undefined genes may play in survival under the indicated conditions.

This report details an attempt to understand competitive rhizosphere colonization from a microbial perspective. A coupled analysis of plant genes playing a role in influencing rhizosphere colonization would complement this study and bring the issue of competitive rhizosphere colonization into full focus. RNA-seq techniques have been developed that would allow for the identification of genes that are over-expressed during the colonization process. It would be very interesting to see how gene expression changes during the infectious process, particularly in the context of host-microbe compatibility.

4.3 Conclusions

This work has reported the identification and characterization of genes required for successful nodulation and symbiotic nitrogen fixation from two perspectives – a focused, single-amino-acid resolution study of a microbial signaling system required for nodulation and a broad, sweeping search for gene networks required for competitive root colonization prior to the development of symbiotic nitrogen fixation. The combination of large-scale gene-network identification with a detailed mechanistic study lays a foundation upon which the complexity of host-microbe interactions required for symbiotic nitrogen fixation begins to come into focus.

The characterization of the FeuPQN system reported in chapter two has provided genetic evidence for a signal mechanism by which periplasmic information may be relayed into the cytoplasmic compartment of the cell and translated into microbial gene expression. This particular system may provide a convenient model for additional studies aimed at understanding SK periplasmic sensing domains, as FeuN serves as a defined, genetically manipulatable periplasmic signal. Further biochemical analysis of this and similar systems may allow researchers to definitively characterize mechanisms for microbial transmembrane signal

transduction. The established role played by this system in the invasion of infection threads during nodulation provides additional detail to the understanding of the molecular basis for symbiotic nitrogen fixation.

The genes identified in chapters three and four as determinants of competitive rhizosphere colonization contribute to the wealth of information being accumulated in an effort to better understand the processes required for symbiotic nitrogen fixation. Identifying a core set of genes required for competitive persistence in the rhizosphere is a crucial step in the development of microbial solutions to significant problems such as nitrogen fertilization and soil bioremediation. The candidate genes identified by our Tn-seq study can be utilized in mechanistic studies that may help unravel the molecular mechanisms involved in competitive rhizosphere colonization. As Tn-seq and other high throughput methods continue to identify genes involved in an expanding panel of plant and bacterial species, the requirements of both general and host-specific rhizosphere colonization will continue to come into focus.

REFERENCES

- Albus, U., Baier, R., Holst, O., Pühler, A., & Niehaus, K. (2001). Suppression of an elicitor-induced oxidative burst reaction in *Medicago sativa* cell cultures by *Sinorhizobium meliloti* lipopolysaccharides. *New Phytologist*, *151*(3), 597-606.
- Åman, P., McNeil, M., Franzén, L.-E., Darvill, A. G., & Albersheim, P. (1981). Structural elucidation, using HPLC-MS and GLC-MS, of the acidic polysaccharide secreted by *Rhizobium meliloti* strain 1021. *Carbohydrate Research*, *95*(2), 263-282.
- Ames, P., Zhou, Q., & Parkinson, J. S. (2008). Mutational analysis of the connector segment in the HAMP domain of Tsr, the *Escherichia coli* serine chemoreceptor. *Journal of bacteriology*, *190*(20), 6676-6685.
- Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, *22*(2), 195-201.
- Ashenberg, O., Keating, A. E., & Laub, M. T. (2013). Helix bundle loops determine whether histidine kinases autophosphorylate in cis or in trans. *Journal of molecular biology*, *425*(7), 1198-1209.
- Atkinson, M. R., & Ninfa, A. (1993). Mutational analysis of the bacterial signal-transducing protein kinase/phosphatase nitrogen regulator II (NRII or NtrB). *Journal of bacteriology*, *175*(21), 7016-7023.
- Bassford Jr, P. J. (1990). Export of the periplasmic maltose-binding protein of *Escherichia coli*. *Journal of bioenergetics and biomembranes*, *22*(3), 401-439.
- Beatty, P. H., & Good, A. G. (2011). Future prospects for cereals that fix nitrogen. *Science*, *333*(6041), 416-417.
- Bedouelle, H., Bassford, P. J., Fowler, A. V., Zabin, I., Beckwith, J., & Hofnung, M. (1980). Mutations which alter the function of the signal sequence of the maltose binding protein of *Escherichia coli*.
- Bick, M. J., Lamour, V., Rajashankar, K. R., Gordiyenko, Y., Robinson, C. V., & Darst, S. A. (2009). How to switch off a histidine kinase: crystal structure of *Geobacillus stearothermophilus* KinB with the inhibitor Sda. *Journal of molecular biology*, *386*(1), 163-177.
- Bourret, R. B. (2010). Receiver domain structure and function in response regulator proteins. *Current opinion in microbiology*, *13*(2), 142-149.
- Buelow, D. R., & Raivio, T. L. (2010). Three (and more) component regulatory systems—auxiliary regulators of bacterial histidine kinases. *Molecular microbiology*, *75*(3), 547-566.

- Burris, R. (1988). *100 years of discoveries in biological N₂ fixation*. Paper presented at the Nitrogen fixation: hundred years after: proceedings of the 7th International Congress on N [Triple-bond] Nitrogen Fixation, Koln (Cologne), FRG, March 13-20, 1980/edited by H. Bothe, FJ de Bruijn and WE Newton.
- Campbell, G. R., Reuhs, B. L., & Walker, G. C. (2002). Chronic intracellular infection of alfalfa nodules by *Sinorhizobium meliloti* requires correct lipopolysaccharide core. *Proceedings of the National Academy of Sciences*, *99*(6), 3938-3943.
- Capdevila, S., Martínez-Granero, F. M., Sánchez-Contreras, M., Rivilla, R., & Martín, M. (2004). Analysis of *Pseudomonas fluorescens* F113 genes implicated in flagellar filament synthesis and their role in competitive root colonization. *Microbiology*, *150*(11), 3889-3897.
- Capra, E. J., Perchuk, B. S., Lubin, E. A., Ashenberg, O., Skerker, J. M., & Laub, M. T. (2010). Systematic dissection and trajectory-scanning mutagenesis of the molecular interface that ensures specificity of two-component signaling pathways.
- Carlyon, R. E., Ryther, J. L., VanYperen, R. D., & Griffiths, J. S. (2010). FeuN, a novel modulator of two-component signalling identified in *Sinorhizobium meliloti*. *Molecular microbiology*, *77*(1), 170-182.
- Casino, P., Rubio, V., & Marina, A. (2009). Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. *Cell*, *139*(2), 325-336.
- Casino, P., Rubio, V., & Marina, A. (2010). The mechanism of signal transduction by two-component systems. *Current opinion in structural biology*, *20*(6), 763-771.
- Chen, E. J., Sabio, E. A., & Long, S. R. (2008). The periplasmic regulator ExoR inhibits ExoS/ChvI two-component signalling in *Sinorhizobium meliloti*. *Molecular microbiology*, *69*(5), 1290-1303.
- Cheng, H.-P., & Walker, G. C. (1998). Succinoglycan is required for initiation and elongation of infection threads during nodulation of alfalfa by *Rhizobium meliloti*. *Journal of bacteriology*, *180*(19), 5183-5191.
- Crook, M. B., Lindsay, D. P., Biggs, M. B., Bentley, J. S., Price, J. C., Clement, S. C., . . . Griffiths, J. S. (2012). Rhizobial plasmids that cause impaired symbiotic nitrogen fixation and enhanced host invasion. *Molecular Plant-Microbe Interactions*, *25*(8), 1026-1033.
- Cunningham, K. A., & Burkholder, W. F. (2009). The histidine kinase inhibitor Sda binds near the site of autophosphorylation and may sterically hinder autophosphorylation and phosphotransfer to Spo0F. *Molecular microbiology*, *71*(3), 659-677.
- David, M., Daveran, M.-L., Batut, J., Dedieu, A., Domergue, O., Ghai, J., . . . Kahn, D. (1988). Cascade regulation of *nif* gene expression in *Rhizobium meliloti*. *Cell*, *54*(5), 671-683.
- de Weert, S., Vermeiren, H., Mulders, I. H., Kuiper, I., Hendrickx, N., Bloemberg, G. V., . . . Lugtenberg, B. J. (2002). Flagella-driven chemotaxis towards exudate components is an

- important trait for tomato root colonization by *Pseudomonas fluorescens*. *Molecular Plant-Microbe Interactions*, 15(11), 1173-1180.
- Deakin, W. J., & Broughton, W. J. (2009). Symbiotic use of pathogenic strategies: rhizobial protein secretion systems. *Nature Reviews Microbiology*, 7(4), 312-320.
- Dénarié, J., Debelle, F., & Promé, J.-C. (1996). Rhizobium lipo-chitoooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. *Annual review of biochemistry*, 65(1), 503-535.
- Denarie, J., Debelle, F., & Rosenberg, C. (1992). Signaling and host range variation in nodulation. *Annual Reviews in Microbiology*, 46(1), 497-531.
- Dickstein, R., Bisseling, T., Reinhold, V. N., & Ausubel, F. M. (1988). Expression of nodule-specific genes in alfalfa root nodules blocked at an early stage of development. *Genes & development*, 2(6), 677-687.
- Dong, T. G., Ho, B. T., Yoder-Himes, D. R., & Mekalanos, J. J. (2013). Identification of T6SS-dependent effector and immunity proteins by Tn-seq in *Vibrio cholerae*. *Proceedings of the National Academy of Sciences*, 110(7), 2623-2628.
- Erisman, J. W., Sutton, M. A., Galloway, J., Klimont, Z., & Winiwarter, W. (2008). How a century of ammonia synthesis changed the world. *Nature Geoscience*, 1(10), 636-639.
- Ernst, R. K., Guina, T., & Miller, S. I. (1999). How intracellular bacteria survive: surface modifications that promote resistance to host innate immune responses. *Journal of Infectious Diseases*, 179(Supplement 2), S326-S330.
- Esseling, J. J., Lhuissier, F. G., & Emons, A. M. C. (2003). Nod factor-induced root hair curling: continuous polar growth towards the point of nod factor application. *Plant physiology*, 132(4), 1982-1988.
- Feilmeier, B. J., Iseminger, G., Schroeder, D., Webber, H., & Phillips, G. J. (2000). Green fluorescent protein functions as a reporter for protein localization in *Escherichia coli*. *Journal of bacteriology*, 182(14), 4068-4076.
- Ferris, H. U., Dunin-Horkawicz, S., Hornig, N., Hulko, M., Martin, J., Schultz, J. E., . . . Coles, M. (2012). Mechanism of regulation of receptor histidine kinases. *Structure*, 20(1), 56-66.
- Fields, A. T., Navarrete, C. S., Zare, A. Z., Huang, Z., Mostafavi, M., Lewis, J. C., . . . Chen, J. C. (2012). The conserved polarity factor podJ1 impacts multiple cell envelope-associated functions in *Sinorhizobium meliloti*. *Mol Microbiol*, 84(5), 892-920. doi:10.1111/j.1365-2958.2012.08064.x
- Fleischer, R., Heermann, R., Jung, K., & Hunke, S. (2007). Purification, reconstitution, and characterization of the CpxRAP envelope stress system of *Escherichia coli*. *Journal of Biological Chemistry*, 282(12), 8583-8593.

- Foo, Y. H., Gao, Y., Zhang, H., & Kenney, L. J. (2015). Cytoplasmic sensing by the inner membrane histidine kinase EnvZ. *Progress in biophysics and molecular biology*.
- Freeman, J. A., Lilley, B. N., & Bassler, B. L. (2000). A genetic analysis of the functions of LuxN: a two-component hybrid sensor kinase that regulates quorum sensing in *Vibrio harveyi*. *Molecular microbiology*, 35(1), 139-149.
- Fu, Y., Waldor, M. K., & Mekalanos, J. J. (2013). Tn-Seq analysis of *Vibrio cholerae* intestinal colonization reveals a role for T6SS-mediated antibacterial activity in the host. *Cell host & microbe*, 14(6), 652-663.
- Gage, D. J. (2004). Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiology and Molecular Biology Reviews*, 68(2), 280-300.
- Gallagher, L. A., Shendure, J., & Manoil, C. (2011). Genome-scale identification of resistance functions in *Pseudomonas aeruginosa* using Tn-seq. *MBio*, 2(1), e00315-00310.
- Gao, R., & Stock, A. M. (2009). Biological insights from structures of two-component proteins. *Annual review of microbiology*, 63, 133.
- Gao, R., & Stock, A. M. (2013). Probing kinase and phosphatase activities of two-component systems in vivo with concentration-dependent phosphorylation profiling. *Proceedings of the National Academy of Sciences*, 110(2), 672-677.
- Garbeva, P., Van Elsas, J., & Van Veen, J. (2008). Rhizosphere microbial community and its response to plant species and soil history. *Plant and Soil*, 302(1-2), 19-32.
- Garnerone, A.-M., Cabanes, D., Foussard, M., Boistard, P., & Batut, J. (1999). Inhibition of the FixL Sensor Kinase by the FixT Protein in *Sinorhizobium meliloti*. *Journal of Biological Chemistry*, 274(45), 32500-32506.
- Gerken, H., Charlson, E. S., Cicirelli, E. M., Kenney, L. J., & Misra, R. (2009). MzrA: a novel modulator of the EnvZ/OmpR two-component regulon. *Molecular microbiology*, 72(6), 1408-1422.
- Gerken, H., & Misra, R. (2010). MzrA-EnvZ interactions in the periplasm influence the EnvZ/OmpR two-component regulon. *Journal of bacteriology*, 192(23), 6271-6278.
- Gibson, K. E., Kobayashi, H., & Walker, G. C. (2008). Molecular determinants of a symbiotic chronic infection. *Annual review of genetics*, 42, 413.
- Gilles-Gonzalez, M. A., & Gonzalez, G. (1993). Regulation of the kinase activity of heme protein FixL from the two-component system FixL/FixJ of *Rhizobium meliloti*. *Journal of Biological Chemistry*, 268(22), 16293-16297.
- Glandorf, B. (1992). *Root colonization by fluorescent pseudomonads: Wortelkolonisatie door fluorescerende pseudomonas bacteriën*: [SI]: Glandorf.

- Grant, S. G., Jessee, J., Bloom, F. R., & Hanahan, D. (1990). Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proceedings of the National Academy of Sciences*, *87*(12), 4645-4649.
- Grebe, T. W., & Stock, J. B. (1999). The histidine protein kinase superfamily. *Advances in microbial physiology*, *41*, 139-227.
- Griffitts, J. S., Carlyon, R. E., Erickson, J. H., Moulton, J. L., Barnett, M. J., Toman, C. J., & Long, S. R. (2008). A *Sinorhizobium meliloti* osmosensory two-component system required for cyclic glucan export and symbiosis. *Molecular microbiology*, *69*(2), 479-490.
- Groban, E. S., Clarke, E. J., Salis, H. M., Miller, S. M., & Voigt, C. A. (2009). Kinetic buffering of cross talk between bacterial two-component sensors. *Journal of molecular biology*, *390*(3), 380-393.
- Hiltner, L. (1904). Über neuere Erfahrungen und Probleme auf dem Gebiete der Bodenbakteriologie unter besonderer Berücksichtigung der Gründüngung und Brache. *Arbeiten der Deutschen Landwirtschaftlichen Gesellschaft*, *98*, 59-78.
- Hirsch, A. M. (1992). Developmental biology of legume nodulation. *New Phytologist*, *122*(2), 211-237.
- Hoch, J. A. (2000). Two-component and phosphorelay signal transduction. *Current opinion in microbiology*, *3*(2), 165-170.
- Hoffman, C. S., & Wright, A. (1985). Fusions of secreted proteins to alkaline phosphatase: an approach for studying protein secretion. *Proceedings of the National Academy of Sciences*, *82*(15), 5107-5111.
- Howard, J. B., & Rees, D. C. (1996). Structural basis of biological nitrogen fixation. *Chemical Reviews*, *96*(7), 2965-2982.
- Hsing, W., Russo, F. D., Bernd, K. K., & Silhavy, T. J. (1998). Mutations that alter the kinase and phosphatase activities of the two-component sensor EnvZ. *Journal of bacteriology*, *180*(17), 4538-4546.
- Hsing, W., & Silhavy, T. J. (1997). Function of conserved histidine-243 in phosphatase activity of EnvZ, the sensor for porin osmoregulation in *Escherichia coli*. *Journal of bacteriology*, *179*(11), 3729-3735.
- Hulko, M., Berndt, F., Gruber, M., Linder, J. U., Truffault, V., Schultz, A., . . . Coles, M. (2006). The HAMP domain structure implies helix rotation in transmembrane signaling. *Cell*, *126*(5), 929-940.
- Hulsen, T., de Vlieg, J., & Alkema, W. (2008). BioVenn—a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC genomics*, *9*(1), 488.

- Huynh, T. N., & Stewart, V. (2011). Negative control in two-component signal transduction by transmitter phosphatase activity. *Molecular microbiology*, *82*(2), 275-286.
- Ishii, E., Eguchi, Y., & Utsumi, R. (2013). Mechanism of activation of PhoQ/PhoP two-component signal transduction by SafA, an auxiliary protein of PhoQ histidine kinase in *Escherichia coli*. *Bioscience, biotechnology, and biochemistry*, *77*(4), 814-819.
- Jacques, D. A., Langley, D. B., Hynson, R. M., Whitten, A. E., Kwan, A., Guss, J. M., & Trewella, J. (2011). A novel structure of an antikinase and its inhibitor. *Journal of molecular biology*, *405*(1), 214-226.
- Jeong, D. W., Cho, H., Jones, M. B., Shatzkes, K., Sun, F., Ji, Q., . . . Bae, T. (2012). The auxiliary protein complex SaePQ activates the phosphatase activity of sensor kinase SaeS in the SaeRS two-component system of *Staphylococcus aureus*. *Molecular microbiology*, *86*(2), 331-348.
- Jones, K. M., Kobayashi, H., Davies, B. W., Taga, M. E., & Walker, G. C. (2007). How rhizobial symbionts invade plants: the Sinorhizobium–Medicago model. *Nature Reviews Microbiology*, *5*(8), 619-633.
- Jung, K., Fried, L., Behr, S., & Heermann, R. (2012). Histidine kinases and response regulators in networks. *Current opinion in microbiology*, *15*(2), 118-124.
- Karimova, G., Pidoux, J., Ullmann, A., & Ladant, D. (1998). A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proceedings of the National Academy of Sciences*, *95*(10), 5752-5756.
- Kenney, L. J. (2010). How important is the phosphatase activity of sensor kinases? *Current opinion in microbiology*, *13*(2), 168-176.
- Khan, W., Prithiviraj, B., & Smith, D. L. (2008). Nod factor [Nod Bj V (C 18: 1, MeFuc)] and lumichrome enhance photosynthesis and growth of corn and soybean. *Journal of plant physiology*, *165*(13), 1342-1351.
- Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., & Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. *Nucleic acids research*, *37*(suppl 1), D387-D392.
- King, S. T., & Kenney, L. J. (2007). [17]-Application of Fluorescence Resonance Energy Transfer to Examine EnvZ/OmpR Interactions. *Methods in enzymology*, *422*, 352-360.
- Kitanovic, S., Ames, P., & Parkinson, J. S. (2011). Mutational analysis of the control cable that mediates transmembrane signaling in the *Escherichia coli* serine chemoreceptor. *Journal of bacteriology*, *193*(19), 5062-5072.
- Kloepper, J. W., & Beauchamp, C. J. (1992). A review of issues related to measuring colonization of plant roots by bacteria. *Canadian Journal of Microbiology*, *38*(12), 1219-1232.

- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 10(3), R25.
- Leigh, J. A., Signer, E. R., & Walker, G. C. (1985). Exopolysaccharide-deficient mutants of *Rhizobium meliloti* that form ineffective nodules. *Proceedings of the National Academy of Sciences*, 82(18), 6231-6235.
- Lerouge, P., Roche, P., Faucher, C., Maillet, F., Truchet, G., Promé, J. C., & Dénarié, J. (1990). Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. *Nature*, 344(6268), 781-784.
- Liang, Y., Cao, Y., Tanaka, K., Thibivilliers, S., Wan, J., Choi, J., . . . Stacey, G. (2013). Nonlegumes respond to rhizobial Nod factors by suppressing the innate immune response. *Science*, 341(6152), 1384-1387.
- Limpens, E., Franken, C., Smit, P., Willemsse, J., Bisseling, T., & Geurts, R. (2003). LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. *Science*, 302(5645), 630-633.
- Loh, J., Garcia, M., & Stacey, G. (1997). NodV and NodW, a second flavonoid recognition system regulating nod gene expression in *Bradyrhizobium japonicum*. *Journal of bacteriology*, 179(9), 3013-3020.
- Long, S. R., Buikema, W. J., & Ausubel, F. M. (1982). Cloning of *Rhizobium meliloti* nodulation genes by direct complementation of Nod⁻ mutants.
- Lowe, R. H., & Evans, H. J. (1962). CARBON DIOXIDE REQUIREMENT FOR GROWTH OF LEGUME NODULE BACTERIA. *Soil Science*, 94(6), 351-356.
- Lugtenberg, B. J., Dekkers, L., & Bloemberg, G. V. (2001). Molecular determinants of rhizosphere colonization by *Pseudomonas*. *Annual review of phytopathology*, 39(1), 461-490.
- Lugtenberg, B. J., Kravchenko, L. V., & Simons, M. (1999). Tomato seed and root exudate sugars: composition, utilization by *Pseudomonas* biocontrol strains and role in rhizosphere colonization. *Environmental Microbiology*, 1(5), 439-446.
- Lukat, G. S., McCleary, W. R., Stock, A. M., & Stock, J. B. (1992). Phosphorylation of bacterial response regulator proteins by low molecular weight phospho-donors. *Proceedings of the National Academy of Sciences*, 89(2), 718-722.
- Madsen, E. B., Madsen, L. H., Radutoiu, S., Olbryt, M., Rakwalska, M., Szczyglowski, K., . . . Sandal, N. (2003). A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. *Nature*, 425(6958), 637-640.
- Maier, R. J., & Brill, W. J. (1978). Mutant strains of *Rhizobium japonicum* with increased ability to fix nitrogen for soybean. *Science*, 201(4354), 448-450.
- Manson, M. D. (2011). Not too loose, not too tight—just right. Biphasic control of the Tsr HAMP domain. *Molecular microbiology*, 80(3), 573-576.

- Manual, F. (1998). United Nations Industrial Development Organization (UNIDO) and International Fertilizer Development Center (IFDC). *The Netherlands*.
- Marie, C., Deakin, W. J., Viprey, V., Kopcińska, J., Golinowski, W., Krishnan, H. B., . . . Broughton, W. J. (2003). Characterization of Nops, nodulation outer proteins, secreted via the type III secretion system of NGR234. *Molecular Plant-Microbe Interactions*, *16*(9), 743-751.
- Mattison, K., & Kenney, L. J. (2002). Phosphorylation alters the interaction of the response regulator OmpR with its sensor kinase EnvZ. *Journal of Biological Chemistry*, *277*(13), 11143-11148.
- McCleary, W. R., Stock, J., & Ninfa, A. (1993). Is acetyl phosphate a global signal in *Escherichia coli*? *Journal of bacteriology*, *175*(10), 2793.
- Meade, H. M., Long, S. R., Ruvkun, G. B., Brown, S. E., & Ausubel, F. M. (1982). Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn5 mutagenesis. *Journal of bacteriology*, *149*(1), 114-122.
- Mergaert, P., Uchiumi, T., Alunni, B., Evanno, G., Cheron, A., Catrice, O., . . . Kondorosi, A. (2006). Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium-legume symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(13), 5230-5235.
- Miller, J. H. (1972). Experiments in molecular genetics.
- Mitchell, S. L., Ismail, A. M., Kenrick, S. A., & Camilli, A. (2015). The VieB auxiliary protein negatively regulates the VieSA signal transduction system in *Vibrio cholerae*. *BMC microbiology*, *15*(1), 59.
- Mithöfer, A., Bhagwat, A. A., Feger, M., & Ebel, J. (1996). Suppression of fungal β -glucan-induced plant defence in soybean (*Glycine max* L.) by cyclic 1, 3-1, 6- β -glucans from the symbiont *Bradyrhizobium japonicum*. *Planta*, *199*(2), 270-275.
- Mitrophanov, A. Y., & Groisman, E. A. (2008). Signal integration in bacterial two-component regulatory systems. *Genes & development*, *22*(19), 2601-2611.
- Miwa, H., Sun, J., Oldroyd, G. E., & Downie, J. A. (2006). Analysis of Nod-factor-induced calcium signaling in root hairs of symbiotically defective mutants of *Lotus japonicus*. *Molecular Plant-Microbe Interactions*, *19*(8), 914-923.
- Mizuno, T. (1997). Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of *Escherichia coli*. *DNA research*, *4*(2), 161-168.
- Montesano, M., Brader, G., & Palva, E. T. (2003). Pathogen derived elicitors: searching for receptors in plants. *Molecular Plant Pathology*, *4*(1), 73-79.
- Neiditch, M. B., Federle, M. J., Pompeani, A. J., Kelly, R. C., Swem, D. L., Jeffrey, P. D., . . . Hughson, F. M. (2006). Ligand-induced asymmetry in histidine sensor kinase complex regulates quorum sensing. *Cell*, *126*(6), 1095-1108.

- Newcomb, W. (1981). Nodule morphogenesis and differentiation [Rhizobium]. *International Review of Cytology*.
- Ninfa, A. J., & Magasanik, B. (1986). Covalent modification of the glnG product, NRI, by the glnL product, NRII, regulates the transcription of the glnALG operon in Escherichia coli. *Proceedings of the National Academy of Sciences*, 83(16), 5909-5913.
- Oldroyd, G. E. (2013). Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nature Reviews Microbiology*, 11(4), 252-263.
- Oldroyd, G. E., & Dixon, R. (2014). Biotechnological solutions to the nitrogen problem. *Current opinion in biotechnology*, 26, 19-24.
- Oldroyd, G. E., Murray, J. D., Poole, P. S., & Downie, J. A. (2011). The rules of engagement in the legume-rhizobial symbiosis. *Annual review of genetics*, 45, 119-144.
- Oono, R., Anderson, C. G., & Denison, R. F. (2011). Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20102193.
- Parkinson, D., Gray, T. R., & Williams, S. T. (1971). Methods for study-ing the ecology of soil micro-organisms. *Methods for study-ing the ecology of soil micro-organisms*.
- Parkinson, J. S. (1993). Signal transduction schemes of bacteria. *Cell*, 73(5), 857-871.
- Parkinson, J. S. (2010). Signaling mechanisms of HAMP domains in chemoreceptors and sensor kinases. *Annual review of microbiology*, 64, 101-122.
- Parkinson, J. S., & Kofoed, E. C. (1992). Communication modules in bacterial signaling proteins. *Annual review of genetics*, 26(1), 71-112.
- Patil, R. V., & Dekker, E. E. (1992). Cloning, nucleotide sequence, overexpression, and inactivation of the Escherichia coli 2-keto-4-hydroxyglutarate aldolase gene. *Journal of bacteriology*, 174(1), 102-107.
- Pellock, B. J., Cheng, H.-P., & Walker, G. C. (2000). Alfalfa Root Nodule Invasion Efficiency Is Dependent on Sinorhizobium meliloti Polysaccharides. *Journal of bacteriology*, 182(15), 4310-4318.
- Penterman, J., Abo, R. P., De Nisco, N. J., Arnold, M. F., Longhi, R., Zanda, M., & Walker, G. C. (2014). Host plant peptides elicit a transcriptional response to control the Sinorhizobium meliloti cell cycle during symbiosis. *Proceedings of the National Academy of Sciences*, 111(9), 3561-3566.
- Perret, X., Staehelin, C., & Broughton, W. J. (2000). Molecular basis of symbiotic promiscuity. *Microbiology and Molecular Biology Reviews*, 64(1), 180-201.
- Petit-Härtlein, I., Rome, K., de Rosny, E., Molton, F., Duboc, C., Gueguen, E., . . . Covès, J. (2015). Biophysical and physiological characterization of ZraP from Escherichia coli, the

- periplasmic accessory protein of the atypical ZraSR two-component system. *Biochemical Journal*, BJ20150827.
- Phillips, D. A. (1980). Efficiency of symbiotic nitrogen fixation in legumes. *Annual Review of Plant Physiology*, 31(1), 29-49.
- Pini, F., Frage, B., Ferri, L., De Nisco, N. J., Mohapatra, S. S., Taddei, L., . . . Biondi, E. G. (2013). The DivJ, CbrA and PleC system controls DivK phosphorylation and symbiosis in *Sinorhizobium meliloti*. *Mol Microbiol*, 90(1), 54-71. doi:10.1111/mmi.12347
- Pioszak, A. A., & Ninfa, A. J. (2003a). Genetic and biochemical analysis of phosphatase activity of *Escherichia coli* NRII (NtrB) and its regulation by the PII signal transduction protein. *Journal of bacteriology*, 185(4), 1299-1315.
- Pioszak, A. A., & Ninfa, A. J. (2003b). Mechanism of the PII-activated phosphatase activity of *Escherichia coli* NRII (NtrB): how the different domains of NRII collaborate to act as a phosphatase. *Biochemistry*, 42(29), 8885-8899.
- Pioszak, A. A., & Ninfa, A. J. (2004). Mutations altering the N-terminal receiver domain of NRI (NtrC) that prevent dephosphorylation by the NRII-PII complex in *Escherichia coli*. *Journal of bacteriology*, 186(17), 5730-5740.
- Price, P. A., Tanner, H. R., Dillon, B. A., Shabab, M., Walker, G. C., & Griffiths, J. S. (2015). Rhizobial peptidase HrrP cleaves host-encoded signaling peptides and mediates symbiotic compatibility. *Proceedings of the National Academy of Sciences*, 201417797.
- Prithiviraj, B., Zhou, X., Souleimanov, A., Kahn, W., & Smith, D. (2003). A host-specific bacteria-to-plant signal molecule (Nod factor) enhances germination and early growth of diverse crop plants. *Planta*, 216(3), 437-445.
- Radutoiu, S., Madsen, L. H., Madsen, E. B., Felle, H. H., Umehara, Y., Grønlund, M., . . . Sandal, N. (2003). Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature*, 425(6958), 585-592.
- Raivio, T. L., Laird, M. W., Joly, J. C., & Silhavy, T. J. (2000). Tethering of CpxP to the inner membrane prevents spheroplast induction of the cpx envelope stress response. *Molecular microbiology*, 37(5), 1186-1197.
- Raivio, T. L., Popkin, D. L., & Silhavy, T. J. (1999). The Cpx envelope stress response is controlled by amplification and feedback inhibition. *Journal of bacteriology*, 181(17), 5263-5272.
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature biotechnology*, 29(1), 24-26.
- Rovira, A., & Harris, J. (1961). Plant root excretions in relation to the rhizosphere effect. *Plant and Soil*, 14(3), 199-214.

- Rowland, M. A., & Deeds, E. J. (2014). Crosstalk and the evolution of specificity in two-component signaling. *Proceedings of the National Academy of Sciences*, *111*(15), 5550-5555.
- Rowland, S. L., Burkholder, W. F., Cunningham, K. A., Maciejewski, M. W., Grossman, A. D., & King, G. F. (2004). Structure and Mechanism of Action of Sda, an Inhibitor of the Histidine Kinases that Regulate Initiation of Sporulation in *Bacillus subtilis*. *Molecular cell*, *13*(5), 689-701.
- Russo, F. D., & Silhavy, T. J. (1993). The essential tension: opposed reactions in bacterial two-component regulatory systems. *Trends in microbiology*, *1*(8), 306-310.
- Salazar, M. E., & Laub, M. T. (2015). Temporal and evolutionary dynamics of two-component signaling pathways. *Current opinion in microbiology*, *24*, 7-14.
- Scheidle, H., Groß, A., & Niehaus, K. (2005). The lipid A substructure of the *Sinorhizobium meliloti* lipopolysaccharides is sufficient to suppress the oxidative burst in host plants. *New Phytologist*, *165*(2), 559-566.
- Schlüter, J.-P., Reinkensmeier, J., Barnett, M. J., Lang, C., Krol, E., Giegerich, R., . . . Becker, A. (2013). Global mapping of transcription start sites and promoter motifs in the symbiotic α -proteobacterium *Sinorhizobium meliloti* 1021. *BMC genomics*, *14*(1), 156.
- Simons, M., Permentier, H. P., de Weger, L. A., Wijffelman, C. A., & Lugtenberg, B. J. (1997). Amino acid synthesis is necessary for tomato root colonization by *Pseudomonas fluorescens* strain WCS365. *Molecular Plant-Microbe Interactions*, *10*(1), 102-106.
- Simons, M., Van Der Bij, A. J., Brand, I., De Weger, L. A., Wijffelman, C. A., & Lugtenberg, B. J. (1996). Gnotobiotic system for studying rhizosphere colonization by plant growth-promoting *Pseudomonas* bacteria. *MPMI-Molecular Plant Microbe Interactions*, *9*(7), 600-607.
- Siryaporn, A., & Goulian, M. (2008). Cross-talk suppression between the CpxA–CpxR and EnvZ–OmpR two-component systems in *E. coli*. *Molecular microbiology*, *70*(2), 494-506.
- Skerker, J. M., Perchuk, B. S., Siryaporn, A., Lubin, E. A., Ashenberg, O., Goulian, M., & Laub, M. T. (2008). Rewiring the specificity of two-component signal transduction systems. *Cell*, *133*(6), 1043-1054.
- Skerker, J. M., Prasol, M. S., Perchuk, B. S., Biondi, E. G., & Laub, M. T. (2005). Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis.
- Smil, V. (1999). Nitrogen in crop production: An account of global flows. *Global biogeochemical cycles*, *13*(2), 647-662.
- Spaink, H. P. (2000). Root nodulation and infection factors produced by rhizobial bacteria. *Annual Reviews in Microbiology*, *54*(1), 257-288.

- Stock, A. M., Robinson, V. L., & Goudreau, P. N. (2000). Two-component signal transduction. *Annual review of biochemistry*, 69(1), 183-215.
- Stowers, M. D. (1985). Carbon metabolism in Rhizobium species. *Annual Reviews in Microbiology*, 39(1), 89-108.
- Streit, W. R., Joseph, C. M., & Phillips, D. A. (1996). Biotin and other water-soluble vitamins are key growth factors for alfalfa root colonization by Rhizobium meliloti 1021. *Growth*, 1, 1.
- Swain, K. E., & Falke, J. J. (2007). Structure of the conserved HAMP domain in an intact, membrane-bound chemoreceptor: a disulfide mapping study. *Biochemistry*, 46(48), 13684-13695.
- Szurmant, H., Bu, L., Brooks, C. L., & Hoch, J. A. (2008). An essential sensor histidine kinase controlled by transmembrane helix interactions with its auxiliary proteins. *Proceedings of the National Academy of Sciences*, 105(15), 5891-5896.
- Tanaka, T., Saha, S. K., Tomomori, C., Ishima, R., Liu, D., Tong, K. I., . . . Swindells, M. B. (1998). NMR structure of the histidine kinase domain of the E. coli osmosensor EnvZ. *Nature*, 396(6706), 88-92.
- Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2012). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics*, bbs017.
- Toth, K., & Stacey, G. (2015). Does plant immunity have a central role in the legume rhizobium symbiosis? *Name: Frontiers in Plant Science*, 6, 401.
- Townsend, A. R., Howarth, R. W., Bazzaz, F. A., Booth, M. S., Cleveland, C. C., Collinge, S. K., . . . Keeney, D. R. (2003). Human health effects of a changing global nitrogen cycle. *Frontiers in Ecology and the Environment*, 1(5), 240-246.
- Van de Velde, W., Zehirov, G., Szatmari, A., Debreczeny, M., Ishihara, H., Kevei, Z., . . . Tiricz, H. (2010). Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science*, 327(5969), 1122-1126.
- van Opijnen, T., Bodi, K. L., & Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature methods*, 6(10), 767-772.
- Vance, C. P. (2001). Symbiotic nitrogen fixation and phosphorus acquisition. Plant nutrition in a world of declining renewable resources. *Plant physiology*, 127(2), 390-397.
- VanYperen, R. D., Orton, T. S., & Griffiths, J. S. (2015). Genetic analysis of signal integration by the Sinorhizobium meliloti sensor kinase FeuQ. *Microbiology*, 161(Pt 2), 244-253.
- Wang, L., Grau, R., Perego, M., & Hoch, J. A. (1997). A novel histidine kinase inhibitor regulating development in Bacillus subtilis. *Genes & development*, 11(19), 2569-2579.

- Wang, L.-X., Wang, Y., Pellock, B., & Walker, G. C. (1999). Structural Characterization of the Symbiotically Important Low-Molecular-Weight Succinoglycan of *Sinorhizobium meliloti*. *Journal of bacteriology*, *181*(21), 6788-6796.
- Wanner, B. L. (1992). Is cross regulation by phosphorylation of two-component response regulator proteins important in bacteria? *Journal of bacteriology*, *174*(7), 2053.
- West, A. H., & Stock, A. M. (2001). Histidine kinases and response regulator proteins in two-component signaling systems. *Trends in biochemical sciences*, *26*(6), 369-376.
- Whipps, J., & Lynch, J. (1990). Carbon economy. *The rhizosphere.*, 59-97.
- Whipps, J. M. (2001). Microbial interactions and biocontrol in the rhizosphere. *Journal of experimental Botany*, *52*(suppl 1), 487-511.
- Willett, J. W., & Kirby, J. R. (2012). Genetic and biochemical dissection of a HisKA domain identifies residues required exclusively for kinase and phosphatase activities. *PLoS genetics*, *8*(11), e1003084.
- Yang, C.-H., & Crowley, D. E. (2000). Rhizosphere microbial community structure in relation to root location and plant iron nutritional status. *Applied and environmental microbiology*, *66*(1), 345-351.
- Yang, W.-C., de Blank, C., Meskiene, I., Hirt, H., Bakker, J., van Kammen, A., . . . Bisseling, T. (1994). Rhizobium nod factors reactivate the cell cycle during infection and nodule primordium formation, but the cycle is only completed in primordium formation. *The plant cell*, *6*(10), 1415-1426.
- Yang, Y., & Inouye, M. (1991). Intermolecular complementation between two defective mutant signal-transducing receptors of *Escherichia coli*. *Proceedings of the National Academy of Sciences*, *88*(24), 11057-11061.
- Yao, S.-Y., Luo, L., Har, K. J., Becker, A., Rüberg, S., Yu, G.-Q., . . . Cheng, H.-P. (2004). *Sinorhizobium meliloti* ExoR and ExoS proteins regulate both succinoglycan and flagellum production. *Journal of bacteriology*, *186*(18), 6042-6049.
- Young, J., & Johnston, A. (1989). The evolution of specificity in the legume-Rhizobium symbiosis. *Trends in ecology & evolution*, *4*(11), 341-349.
- Yung, M. C., Park, D. M., Overton, K. W., Blow, M. J., Hoover, C. A., Smit, J., . . . Bowman, G. R. (2015). Transposon Mutagenesis Paired with Deep Sequencing of *Caulobacter crescentus* under Uranium Stress Reveals Genes Essential for Detoxification and Stress Tolerance. *Journal of bacteriology*, *197*(19), 3160-3172.
- Zahran, H. H. (1999). Rhizobium-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate. *Microbiology and Molecular Biology Reviews*, *63*(4), 968-989.

- Zhou, Q., Ames, P., & Parkinson, J. S. (2011). Biphasic control logic of HAMP domain signalling in the Escherichia coli serine chemoreceptor. *Molecular microbiology*, 80(3), 596-611.
- Zhou, X., Keller, R., Volkmer, R., Krauss, N., Scheerer, P., & Hunke, S. (2011). Structural basis for two-component system inhibition and pilus sensing by the auxiliary CpxP protein. *Journal of Biological Chemistry*, 286(11), 9805-9814.

Image attributions

Nod factor "NodSm-IV (Ac,C16-2,S)" by Ninjatacoshell - Own work. Licensed under Public Domain via Commons.

SUPPLEMENTAL FIGURES

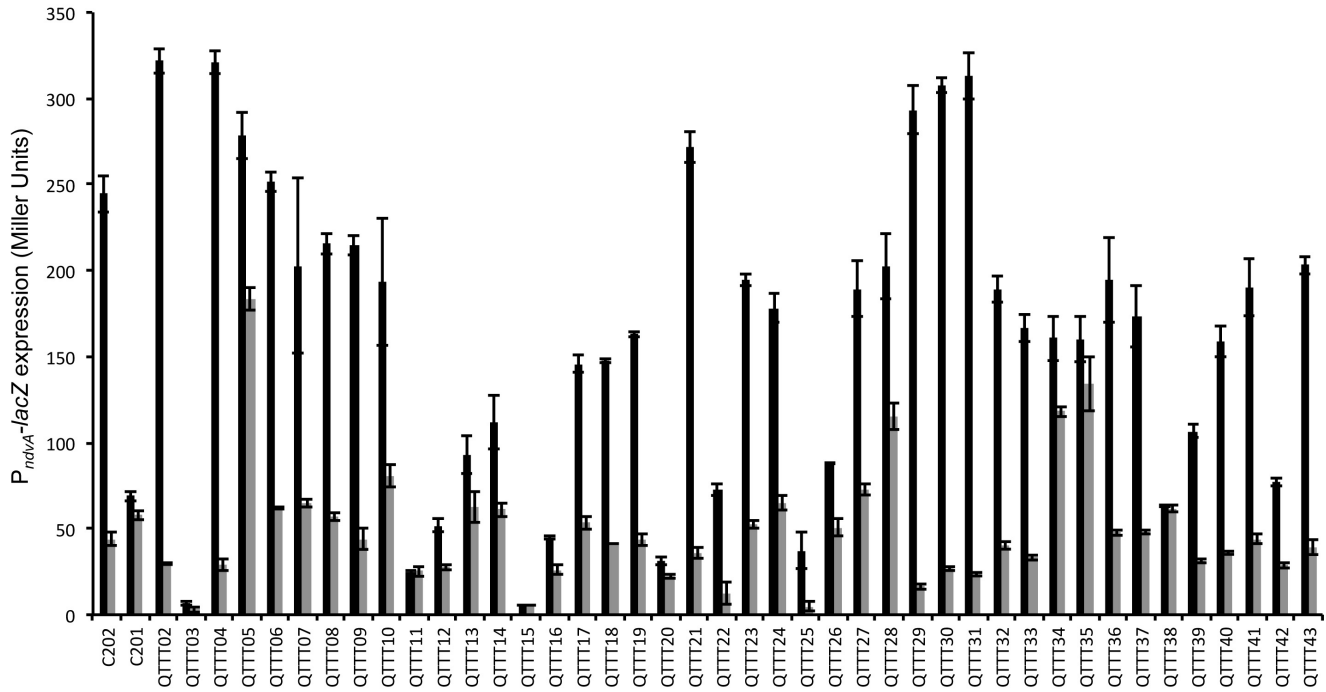


Figure S-1 Quantitative measurement of FeuQ TTT scan phenotypes

Bicistronic *feuP-feuQ*(XXX→TTT) clones (under the control of P_{lac}) are listed in Table S3. The $P_{BAD-feuN}$ plasmid pJG355 is included in all strains. Black bars, $P_{ndvA-lacZ}$ reporter gene activity without arabinose induction of *feuN*; gray bars, reporter gene activity with arabinose induction. P_{lac} is limited to basal level expression.

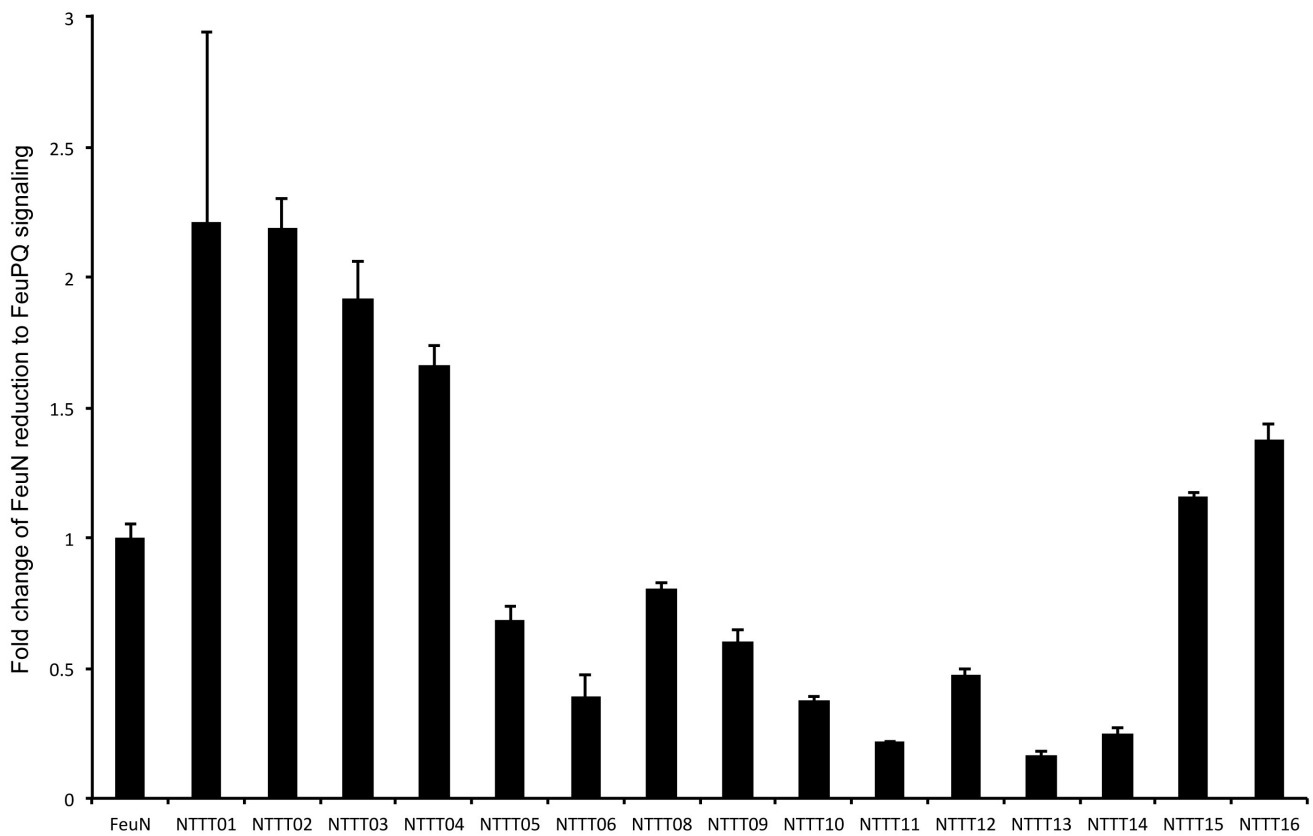


Figure S-2 Quantitative measurement of FeuN TTT scan phenotypes compared to wild-type activity.

feuN(XXX→TTT) clones are listed in Table S3. The TTT alleles were expressed in the presence of *feuP-feuQ* plasmid pJG377 and reporter plasmid pJG286. Activity is given as fold-change compared to wild-type activity. *feuN* was expressed from P_{BAD} under inducing conditions, and P_{tac} expression of *feuP-feuQ* was limited to basal (uninduced) levels.

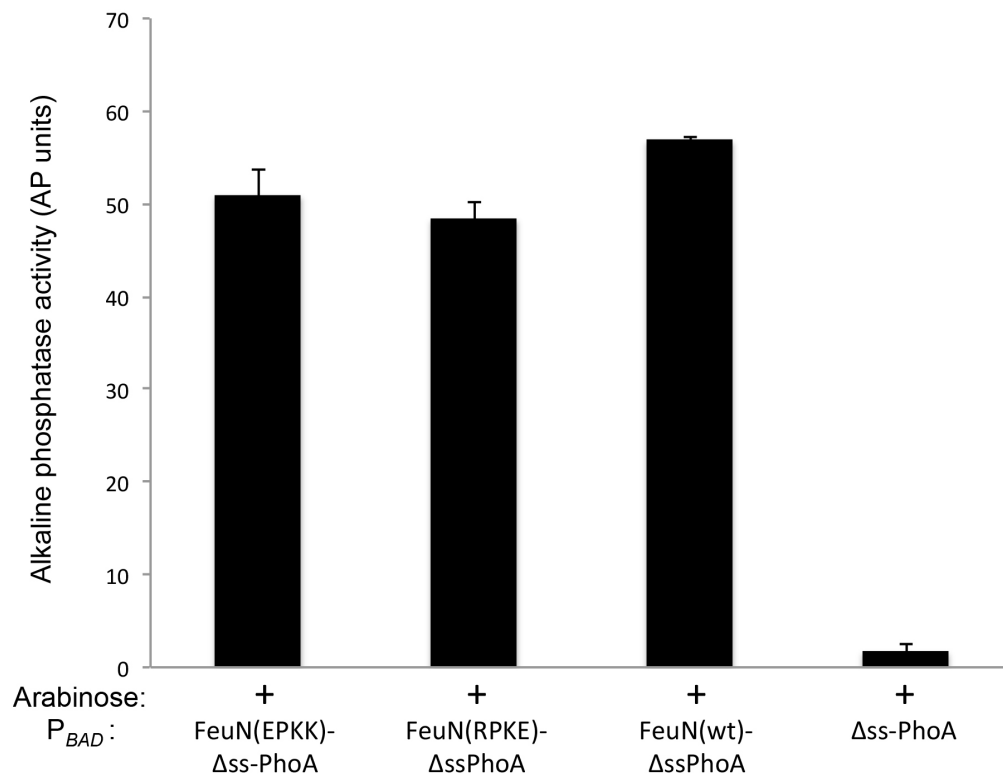


Figure S-3 Quantitative measurement of alkaline phosphatase activity of $\Delta ssPhoA$ fusions to *feuN* and two *feuN*-charge reverse variants.

Cells harboring P_{BAD} plasmids pJG433 ($\Delta ssphoA$), pJG435 (*feuN*- $\Delta ssphoA$), pJG695

(*feuN*(RPKK→EPKK)- $\Delta ssphoA$), or pJG696 (*feuN*(RPKK→RPKE)- $\Delta ssphoA$) were grown in the presence of arabinose and assayed for stability and periplasmic localization by quantitative AP assay.

SUPPLEMENTAL TABLES

Table S1. Primers used in this study*

oJG815	CGCTCTAGAGAAGGCAAGGGAAGTGGACGAG
oJG817	CTGCGAAGCTTGTGACCGCCGGATCACGATC
oJG915	ATTAGCGGATCCTACCTGACG
oJG931	GGGATTTTGGTCATGAGATTATC
oJG932	GCTGAAGCTTGGCAGATCTCCTTAATGGCCGTTGCCGGGTGCTTG
oJG934	CAGCAGATCTCACATGGCCATTAGATCGCTCACGG
oJG940	CAACTCTCTACTGTTTCTCCATACC
oJG966	CGCTCTAGAGCCACCAGGCACTTTCACCGTGACTTTC
oRV041	CTCGGATATCGCATGCAAGCACC
oRV042	CAGTCACGACGTTGTAAAACGAC
oRV131	AGATCGGCAACGATCTGATCG
oRV132	AGGGTTTTCCAGTCACGACG
oRV143	GTCGGCAATCTGGCCGCGTCGCTCAAGACGCCG
oRV144	CGGCGTCTTGAGCGACGCGGCCAGATTGCCGAC
oRV149	AAGCTCACATCGCACGAGTTCC
oRV150	CGTTGTAAAACGACGGCCAGTG
oRV164	ACAACCACACGCGCGCAGCTCTACAACGTC
oRV165	TGTGGTTGTCAGCAGATCCTGAAAGCCGC
oRV166	ACAACCACACTCTACAACGTCATCAACTCC
oRV167	TGTGGTTGTCTGCGCGCGCAGCAGATC
oRV168	ACAACCACAGTCATCAACTCCATCTCCGTC
oRV170	ACAACCACATCCATCTCCGTCGACGAGAAGG
oRV171	TGTGGTTGTGTTGATGACGTTGTAGAGCTG
oRV172	ACAACCACAGTCGACGAGAAGGCCGCG
oRV173	TGTGGTTGTGGAGATGGAGTTGATGACGT
oRV175	TGTGGTTGTCTCGTCGACGGAGATGGAG
oRV176	ACAACCACACTCTCCGGGAGCCCGCAG
oRV178	ACAACCACAAGCCCGCAGCTCGGCGACCT
oRV179	TGTGGTTGTCCCGGAGAGCGCGGCCTTC
oRV180	ACAACCACACTCGGCGACCTGCGCTTTTCC
oRV181	TGTGGTTGTCTGCGGGCTCCCGGAGAGC
oRV183	TGTGGTTGTGTGCGCCGAGCTGCGGGCTC
oRV186	ACAACCACACAGACGGGCTGGTACTGGATCG
oRV187	TGTGGTTGTAGGCTGGGAAAAGCGCAGGTC
oRV189	TGTGGTTGTGCCCGTCTGAGGCTGGGAAAAG
oRV191	TGTGGTTGTCCAGTACCAGCCCGTCTGAG
oRV192	ACAACCACACCGATCGGCGAATTCGACACG
oRV194	ACAACCACAGAATTCGACACGCCGCGCTG

oRV196 ACAACCACAACGCCGCCGCTGATATCGAC
oRV198 ACAACCACACTGATATCGACGTCGCTCG
oRV199 TGTGGTTGTCCGGCGGTGTGCAATTCCG
oRV202 ACAACCACAGGCACGGCGAACTGCCGAT
oRV203 TGTGGTTGTGAGCGACGTCGATATCAGCG
oRV204 ACAACCACAAAACCTGCCGATCGTCAGCGT
oRV205 TGTGGTTGTCCCGTGCCGAGCGACGTCG
oRV208 ACAACCACAGTCGACGAGGTCCCCTTCGA
oRV210 ACAACCACAAAGGCCGCGCTCTCCGGGAG
oRV212 ACAACCACAGACATCCGCTACGAGCGGTT
oRV214 ACAACCACATACGAGCGGTTCTACACCAC
oRV215 TGTGGTTGTGCGGATGTCGAAGGGGACCT
oRV217 TGTGGTTGTCCGCTCGTAGCGGATGTGCA
oRV219 TGTGGTTGTGGTGTAGAACCGCTCGTAGC
oRV221 TGTGGTTGTATCCGTGGTGGTGTAGAACC
oRV224 ACAACCACAGAGGTGGCAGAGACCGAGGT
oRV225 TGTGGTTGTGACCTCGTTTCCGAACGGAT
oRV227 TGTGGTTGTTGCCACCTCGACCTCGTTTC
oRV229 TGTGGTTGTCTCGGTCTCTGCCACCTCGA
oRV231 TGTGGTTGTGAGAACGACCTCGGTCTCTG
oRV233 TGTGGTTGTCTGAATGTCGAGAACGACCT
oRV235 TGTGGTTGTGGCGTGGCCCTGAATGTGCA
oRV237 TGTGGTTGTGAAGCGGGCGGCGTGGCCCT
oRV238 ACAACCACAGGCAACCGCGATGTGCTCGA
oRV239 TGTGGTTGTGGCGACGCGGAAGCGGGCGG
oRV241 TGTGGTTGTGCGGTTGCCGGCGACGCGGA
oRV243 TGTGGTTGTGAGCACATCGCGGTTGCCGG
oRV245 TGTGGTTGTATCCGCCTCGAGCACATCGC
oRV249 CAGATCGACGACGACGCGCCGGGCTCGACCCC
oRV250 GGGGTCGAGCCCCGGCGCGTCTCGTCGATCTG
oRV257 AGCTGCGCGCGTGTGGTTGTCTGAAAGCCGCGCTCGGAC
oRV258 ACGGAGATGGATGTGGTTGTGTTGTAGAGCTGCGCGCGCAGCA
oRV259 ACTCCATCTCCACAACCACAAAGGCCGCGCTCTCCGGGAG
oRV260 AGCTGCGGGCTTGTGGTTGTGCGGGCCTTCTCGTCGAC
oRV261 GGAGCCCCGAGACAACCACACTGCGCTTTTCCCAGCCTCAG
oRV262 AGCTCGGCGACACAACCACATCCCAGCCTCAGACGGGCTG
oRV263 CAGCCCGTCTGTGTGGTTGTAAAGCGCAGGTCGCCGAGC
oRV264 TTTCCCAGCCTACAACCACATGGTACTGGATCGTCGAGCC
oRV265 CTCAGACGGGCACAACCACAATCGTCGAGCCGATCGGC
oRV266 GTGTGCAATTCTGTGGTTGTCTCGACGATCCAGTACCAGC
oRV267 AGCGGCGGGGCTTGTGGTTGTGCCGATCGGCTCGACGATC

oRV268 GTCGATATCAGTGTGGTTGTGTCGAATTCGCCGATCGGCTC
oRV269 ACACGCCGCCGACAACCACAACGTCGCTCGGCACGGCGAAAC
oRV270 TTCGCCGTGCCTGTGGTTGTGTCGATATCAGCGCGGCGTGT
oRV271 TCGGCACGGCGACAACCACAATCGTCAGCGTCGACGAGGT
oRV272 ACCTCGTCGACTGTGGTTGTGCGCAGTTTCGCCGTGCCGA
oRV273 AGCGCGGCCTTTGTGGTTGTGGAGATGGAGTTGATGACGT
oRV274 TAGCGGATGTCTGTGGTTGTCTCGTCGACGCTGACGATCG
oRV275 AACCGCTCGTATGTGGTTGTGAAGGGGACCTCGTCGACGC
oRV276 TCGACATCCGCACAACCACATTCTACACCACCACGGATCC
oRV277 GCTACGAGCGGACAACCACAACCACGGATCCGTTCCGAAA
oRV278 GGTTCTACACCACAACCACACCGTTTCGGAAACGAGGTCTGA
oRV279 CCACCACGGATACAACCACAAACGAGGTTCGAGGTGGCAGA
oRV280 TCTGCCACCTCTGTGGTTGTTCCGAACGGATCCGTGGTGG
oRV281 GAAACGAGGTCACAACCACAGAGACCGAGGTCTGTTCTCGA
oRV282 TCGAGGTGGCAACAACCACAGTCGTTCTCGACATTCAGGG
oRV283 CAGAGACCGAGACAACCACAGACATTCAGGGCCACGCCGC
oRV284 AGGTCTGTTCTACAACCACAGGCCACGCCGCCGCTTCCG
oRV285 TCGACATTCAGACAACCACAGCCCGCTTCCGCGTCGCCGG
oRV286 AGGGCCACGCCACAACCACACGCGTCGCCGGCAACCGCGA
oRV287 CCGCCCGCTTACAACCACAGGCAACCGCGATGTGCTCGA
oRV288 TCCGCGTCGCCACAACCACAGATGTGCTCGAGGCGGATAT
oRV289 CCGGCAACCGCACAACCACAGAGGCGGATATCGACCGCTT
oRV290 GCGATGTGCTCACAACCACAATCGACCGCTTACGCGCAA
oRV291 TCGAGGCGGATACAACCACATTCACGCGCAACCTCACGAT
oRV293 CGCTCTAGACCATTAAGGAGATCTCACATG
oRV294 CGCTCTAGACATGGCCATTAGATCGCTCACGG
oRV299 CTCCAAGCTGGACTGTATGCAC
oRV302 ACCGCCAGATCTACCGCCCATAGGCACTTTCACCGTGAC
oRV303 GGCGGTAGATCTGGCGGTAATAATCGAAGAAGGTAAACTGGTAATC
oRV304 GCGAAGCTTACTTGGTGATACGAGTCTGC
oRV305 CGCTCTAGACATGTCTTCAGCATTGACCATAGC
oRV316 ACCGCCAGATCTACCGCCGACTGCGGCCGGCAGTACGAG
oRV317 CAGCAGATCTCACATGTCTTCAGCATTGACCATAGC
oRV318 CACGCACTGTGTGGTTGTGGTTCGGCTGGGCAGAAAGCA
oRV319 CAGCCGACCACAACCACACAGTGCCTGGTGACGGTGCT
oRV322 CTGCGCGGCTGTGGTTGTGCAATCGCCAGCCACGCGGA
oRV323 GGCGATTGCACAACCACAGCCGCGCAGGTGGTCGCC
oRV324 GGCGACCACTGTGGTTGTGGCCGCGCTGCAATCGCCA
oRV325 AGCGCGGCCACAACCACAGTGGTCGCCAGACGGGTG
oRV326 ACCCGTCTGTGTGGTTGTCTGCGCGGCGGCCGCGCT
oRV327 GCCGCGCAGACAACCACACAGACGGGTGGCGAGCTG

oRV328	CAGCTCGCCTGTGGTTGTGGCGACCACCTGCGCGGC
oRV329	GTGGTCGCCACAACCACAGGCGAGCTGCTTTCTGCC
oRV330	GGCAGAAAGTGTGGTTGTACCCGTCTGGGCGACCAC
oRV331	CAGACGGGTACAACCACACTTTCTGCCAGCCGACC
oRV332	GGTCGGCTGTGTGGTTGTCAGCTCGCCACCCGTCTG
oRV333	GGCGAGCTGACAACCACACAGCCGACCGGCGATGGG
oRV336	CACCGTCACTGTGGTTGTCCCATCGCCGGTCGGCTG
oRV337	GGCGATGGGACAACCACAGTGACGGTGCTCATCCCG
oRV338	CGGGATGAGTGTGGTTGTCACGCACTGCCATCGCC
oRV339	CAGTGCCTGACAACCACACTCATCCCGGGCAATGGA
oRV340	TCCATTGCCTGTGGTTGTCACCGTCACCACGCACTG
oRV341	GTGACGGTGACAACCACAGGCAATGGAGGGCGCCCC
oRV342	GGGGCGCCCTGTGGTTGTGGGATGAGCACCGTCAC
oRV343	CTCATCCCGACAACCACAGGGCGCCCCAAGAAAGTC
oRV344	GACTTTCTTTGTGGTTGTTCCATTGCCCGGGATGAG
oRV345	GGCAATGGAACAACCACAAAGAAAGTCACGGTGAAAGTGC
oRV346	TTTCACCGTTGTGGTTGTGGGGCGCCCTCCATTGCC
oRV347	GGGCGCCCCACAACCACAACGGTGAAAGTGCCTATGTAAG
oRV348	CATAGGCACTGTGGTTGTGACTTTCTTGGGGCGCCC
oRV349	AAGAAAGTCACAACCACAGTGCCTATGTAAGCTCTTTCGG
oRV350	TGTGGTCTTTGCCACCTCGACCTCGTTTC
oRV351	TCGAGGTGGCAAAGACCACAGTCGTTCTCGACATTCAGGG
oRV352	CTTGGTTGTTGCCACCTCGACCTCGTTTC
oRV353	TCGAGGTGGCAAACAACCAAGGTCGTTCTCGACATTCAGGG
oRV354	CTTGGTCTTTGCCACCTCGACCTCGTTTC
oRV355	TCGAGGTGGCAAAGACCAAGGTCGTTCTCGACATTCAGGG
oRV365	TCTTGGGTTCCCCTCCATTGCCCGGGATGA
oRV366	TGGAGGGGAACCCAAGAAAGTCACGGTGAA
oRV367	TGACTTTTTCGGGGCGCCCTCCATTGCCCG
oRV368	GCGCCCCGAAAAGTCACGGTGAAAGTGCC
oRV369	CCGTGACTTCCTTGGGGCGCCCTCCATTGC
oRV370	CCCAAGGAAGTCACGGTGAAAGTGCCTAT
oRV375	CGCAAGCTTGGGCGATTGCAGCGCGGCC
oRV376	CGCGGAATTCGCACCGCCCATAGGCACTTTCACCGTGAC
oRV377	CGCAAGCTTGCGGCAGGGTCCGAGCGC
oRV378	CGCGAGCTCGCACCGCCGCGCGTGAAGCGGTCGATATCC
oRV379	CGCCTGCAGCCGGCAGGGTCCGAGCGC
oRV380	GCGGAGCTCTTAGCGCGTGAAGCGGTCGATATCC

*Functions for these primers are detailed in Supporting Methods and Tables S2 and S3.

Table S2. Random FeuQ mutants*

Name	Description	Primers
DNQA04	pJG249 (<i>feuP</i> , <i>feuQ</i> (F169S))	oRV131/oRV132
DNQA08	pJG249 (<i>feuP</i> , <i>feuQ</i> (G422R))	oRV131/oRV132
DNQA09	pJG249 (<i>feuP</i> , <i>feuQ</i> (I429T))	oRV131/oRV132
NQ001	pJG249 (<i>feuP</i> , <i>feuQ</i> (A21V))	oRV131/oRV132
DNQB06	pJG249 (<i>feuP</i> , <i>feuQ</i> (N51D))	oRV131/oRV132
DNQB07	pJG249 (<i>feuP</i> , <i>feuQ</i> (L182P))	oRV131/oRV132
DNQB08	pJG249 (<i>feuP</i> , <i>feuQ</i> (W84R))	oRV131/oRV132
DNQC02	pJG249 (<i>feuP</i> , <i>feuQ</i> (L49P))	oRV131/oRV132
DNQD01	pJG249 (<i>feuP</i> , <i>feuQ</i> (L401P))	oRV131/oRV132
DNQD03	pJG249 (<i>feuP</i> , <i>feuQ</i> (S428L))	oRV131/oRV132
DNQE1	pJG249 (<i>feuP</i> , <i>feuQ</i> (R412H))	oRV131/oRV132
DNQE3	pJG249 (<i>feuP</i> , <i>feuQ</i> (E231V))	oRV131/oRV132
DNQE10	pJG249 (<i>feuP</i> , <i>feuQ</i> (G422A))	oRV131/oRV132
NQ002	pJG249 (<i>feuP</i> , <i>feuQ</i> (Q48R))	oRV131/oRV132
NQ003	pJG249 (<i>feuP</i> , <i>feuQ</i> (G40D))	oRV131/oRV132
NQ004	pJG249 (<i>feuP</i> , <i>feuQ</i> (Q48R))	oRV131/oRV132
NQ006	pJG249 (<i>feuP</i> , <i>feuQ</i> (A246V))	oRV131/oRV132
NQ009	pJG249 (<i>feuP</i> , <i>feuQ</i> (G251D))	oRV131/oRV132
NQ017	pJG249 (<i>feuP</i> , <i>feuQ</i> (S178P))	oRV131/oRV132
NQ022	pJG249 (<i>feuP</i> , <i>feuQ</i> (Δ 61-72))	oRV131/oRV132
NQ023	pJG249 (<i>feuP</i> , <i>feuQ</i> (R210H))	oRV131/oRV132
NQ025	pJG249 (<i>feuP</i> , <i>feuQ</i> (R412C))	oRV131/oRV132
NQ027	pJG249 (<i>feuP</i> , <i>feuQ</i> (T423I))	oRV131/oRV132

*Mutants generated by random mutagenesis using Taq or Vent^R polymerases

Table S3. Tri-threonine (TTT) substitution mutants for *feuQ* and *feuN*

TTT Substitution	Description	1st round primers	2nd round primers
<i>feuQ</i> TTT02	pJG249, <i>feuP</i> , <i>feuQ</i> (43DLL45→TTT)	oRV149/oRV257, oRV164/oRV150	oRV149/oRV150
<i>feuQ</i> TTT03	pJG249, <i>feuP</i> , <i>feuQ</i> (46RAQ48→TTT)	oRV149/oRV165, oRV166/oRV150	oRV149/oRV150
<i>feuQ</i> TTT04	pJG249, <i>feuP</i> , <i>feuQ</i> (49LYN51→TTT)	oRV149/oRV167, oRV168/oRV150	oRV149/oRV150
<i>feuQ</i> TTT05	pJG249, <i>feuP</i> , <i>feuQ</i> (52VIN54→TTT)	oRV149/oRV258, oRV170/oRV150	oRV149/oRV150
<i>feuQ</i> TTT06	pJG249, <i>feuP</i> , <i>feuQ</i> (55SIS57→TTT)	oRV149/oRV171, oRV172/oRV150	oRV149/oRV150
<i>feuQ</i> TTT07	pJG249, <i>feuP</i> , <i>feuQ</i> (58VDE60→TTT)	oRV149/oRV173, oRV259/oRV150	oRV149/oRV150
<i>feuQ</i> TTT08	pJG249, <i>feuP</i> , <i>feuQ</i> (61KAA63→TTT)	oRV149/oRV175, oRV176/oRV150	oRV149/oRV150
<i>feuQ</i> TTT09	pJG249, <i>feuP</i> , <i>feuQ</i> (64LSG66→TTT)	oRV149/oRV260, oRV178/oRV150	oRV149/oRV150
<i>feuQ</i> TTT10	pJG249, <i>feuP</i> , <i>feuQ</i> (67SPQ69→TTT)	oRV149/oRV179, oRV180/oRV150	oRV149/oRV150
<i>feuQ</i> TTT11	pJG249, <i>feuP</i> , <i>feuQ</i> (70LGD72→TTT)	oRV149/oRV181, oRV261/oRV150	oRV149/oRV150
<i>feuQ</i> TTT12	pJG249, <i>feuP</i> , <i>feuQ</i> (73LRF75→TTT)	oRV149/oRV183, oRV262/oRV150	oRV149/oRV150
<i>feuQ</i> TTT13	pJG249, <i>feuP</i> , <i>feuQ</i> (76SQP78→TTT)	oRV149/oRV263, oRV186/oRV150	oRV149/oRV150
<i>feuQ</i> TTT14	pJG249, <i>feuP</i> , <i>feuQ</i> (79QTG8→TTT)	oRV149/oRV187, oRV264/oRV150	oRV149/oRV150
<i>feuQ</i> TTT15	pJG249, <i>feuP</i> , <i>feuQ</i> (82WYW84→TTT)	oRV149/oRV189, oRV265/oRV150	oRV149/oRV150
<i>feuQ</i> TTT16	pJG249, <i>feuP</i> , <i>feuQ</i> (85IVE87→TTT)	oRV149/oRV191, oRV192/oRV150	oRV149/oRV150
<i>feuQ</i> TTT17	pJG249, <i>feuP</i> , <i>feuQ</i> (88PIG90→TTT)	oRV149/oRV266, oRV194/oRV150	oRV149/oRV150
<i>feuQ</i> TTT18	pJG249, <i>feuP</i> , <i>feuQ</i> (91EFD93→TTT)	oRV149/oRV267, oRV196/oRV150	oRV149/oRV150
<i>feuQ</i> TTT19	pJG249, <i>feuP</i> , <i>feuQ</i> (94TPP96→TTT)	oRV149/oRV268, oRV198/oRV150	oRV149/oRV150
<i>feuQ</i> TTT20	pJG249, <i>feuP</i> , <i>feuQ</i> (97LIS99→TTT)	oRV149/oRV199, oRV269/oRV150	oRV149/oRV150
<i>feuQ</i> TTT21	pJG249, <i>feuP</i> , <i>feuQ</i> (100TSL102→TTT)	oRV149/oRV270, oRV202/oRV150	oRV149/oRV150
<i>feuQ</i> TTT22	pJG249, <i>feuP</i> , <i>feuQ</i> (103GTA105→TTT)	oRV149/oRV203, oRV204/oRV150	oRV149/oRV150
<i>feuQ</i> TTT23	pJG249, <i>feuP</i> , <i>feuQ</i> (106KLP108→TTT)	oRV149/oRV205, oRV271/oRV150	oRV149/oRV150
<i>feuQ</i> TTT24	pJG249, <i>feuP</i> , <i>feuQ</i> (109IVS111→TTT)	oRV149/oRV272, oRV208/oRV150	oRV149/oRV150
<i>feuQ</i> TTT25	pJG249, <i>feuP</i> , <i>feuQ</i> (112VDE114→TTT)	oRV149/oRV273, oRV210/oRV150	oRV149/oRV150
<i>feuQ</i> TTT26	pJG249, <i>feuP</i> , <i>feuQ</i> (115VPF117→TTT)	oRV149/oRV274, oRV212/oRV150	oRV149/oRV150
<i>feuQ</i> TTT27	pJG249, <i>feuP</i> , <i>feuQ</i> (118DIR120→TTT)	oRV149/oRV275, oRV214/oRV150	oRV149/oRV150
<i>feuQ</i> TTT28	pJG249, <i>feuP</i> , <i>feuQ</i> (121YER123→TTT)	oRV149/oRV215, oRV276/oRV150	oRV149/oRV150
<i>feuQ</i> TTT29	pJG249, <i>feuP</i> , <i>feuQ</i> (124FYT126→TTT)	oRV149/oRV217, oRV277/oRV150	oRV149/oRV150
<i>feuQ</i> TTT30	pJG249, <i>feuP</i> , <i>feuQ</i> (127TTD129→TTT)	oRV149/oRV219, oRV278/oRV150	oRV149/oRV150
<i>feuQ</i> TTT31	pJG249, <i>feuP</i> , <i>feuQ</i> (130PFG132→TTT)	oRV149/oRV221, oRV279/oRV150	oRV149/oRV150
<i>feuQ</i> TTT32	pJG249, <i>feuP</i> , <i>feuQ</i> (133NEV135→TTT)	oRV149/oRV280, oRV224/oRV150	oRV149/oRV150
<i>feuQ</i> TTT33	pJG249, <i>feuP</i> , <i>feuQ</i> (136EVA138→TTT)	oRV149/oRV225, oRV281/oRV150	oRV149/oRV150
<i>feuQ</i> TTT34	pJG249, <i>feuP</i> , <i>feuQ</i> (139ETE141→TTT)	oRV149/oRV227, oRV282/oRV150	oRV149/oRV150
<i>feuQ</i> TTT35	pJG249, <i>feuP</i> , <i>feuQ</i> (142VVL144→TTT)	oRV149/oRV229, oRV283/oRV150	oRV149/oRV150
<i>feuQ</i> TTT36	pJG249, <i>feuP</i> , <i>feuQ</i> (145DIQ147→TTT)	oRV149/oRV231, oRV284/oRV150	oRV149/oRV150
<i>feuQ</i> TTT37	pJG249, <i>feuP</i> , <i>feuQ</i> (148GHA150→TTT)	oRV149/oRV233, oRV285/oRV150	oRV149/oRV150
<i>feuQ</i> TTT38	pJG249, <i>feuP</i> , <i>feuQ</i> (151ARF153→TTT)	oRV149/oRV235, oRV286/oRV150	oRV149/oRV150
<i>feuQ</i> TTT39	pJG249, <i>feuP</i> , <i>feuQ</i> (154RVA156→TTT)	oRV149/oRV237, oRV287/oRV150	oRV149/oRV150
<i>feuQ</i> TTT40	pJG249, <i>feuP</i> , <i>feuQ</i> (157GNR159→TTT)	oRV149/oRV239, oRV288/oRV150	oRV149/oRV150

<i>feu</i> QTTT41	pJG249, <i>feuP</i> , <i>feuQ</i> (160DVL162→TTT)	oRV149/oRV241, oRV289/oRV150	oRV149/oRV150
<i>feu</i> QTTT42	pJG249, <i>feuP</i> , <i>feuQ</i> (163EAD165→TTT)	oRV149/oRV243, oRV290/oRV150	oRV149/oRV150
<i>feu</i> QTTT43	pJG249, <i>feuP</i> , <i>feuQ</i> (166IDR168→TTT)	oRV149/oRV245, oRV291/oRV150	oRV149/oRV150
<i>feu</i> NTTT01	pJG351, <i>feuN</i> (36SAA38→TTT)	oJG915/oRV322, oRV323/oJG931	oJG915/oJG931
<i>feu</i> NTTT02	pJG351, <i>feuN</i> (39AAQ41→TTT)	oJG915/oRV324, oRV325/oJG931	oJG915/oJG931
<i>feu</i> NTTT03	pJG351, <i>feuN</i> (42VVA44→TTT)	oJG915/oRV326, oRV327/oJG931	oJG915/oJG931
<i>feu</i> NTTT04	pJG351, <i>feuN</i> (45QTG47→TTT)	oJG915/oRV328, oRV329/oJG931	oJG915/oJG931
<i>feu</i> NTTT05	pJG351, <i>feuN</i> (48GEL50→TTT)	oJG915/oRV330, oRV331/oJG931	oJG915/oJG931
<i>feu</i> NTTT06	pJG351, <i>feuN</i> (51LSA53→TTT)	oJG915/oRV332, oRV333/oJG931	oJG915/oJG931
<i>feu</i> NTTT08	pJG351, <i>feuN</i> (57GDG59→TTT)	oJG915/oRV318, oRV319/oJG931	oJG915/oJG931
<i>feu</i> NTTT09	pJG351, <i>feuN</i> (60QCV62→TTT)	oJG915/oRV336, oRV337/oJG931	oJG915/oJG931
<i>feu</i> NTTT10	pJG351, <i>feuN</i> (63VTV65→TTT)	oJG915/oRV338, oRV339/oJG931	oJG915/oJG931
<i>feu</i> NTTT11	pJG351, <i>feuN</i> (66LIP68→TTT)	oJG915/oRV340, oRV341/oJG931	oJG915/oJG931
<i>feu</i> NTTT12	pJG351, <i>feuN</i> (69GNG71→TTT)	oJG915/oRV342, oRV343/oJG931	oJG915/oJG931
<i>feu</i> NTTT13	pJG351, <i>feuN</i> (72GRP74→TTT)	oJG915/oRV344, oRV345/oJG931	oJG915/oJG931
<i>feu</i> NTTT14	pJG351, <i>feuN</i> (75KKV77→TTT)	oJG915/oRV346, oRV347/oJG931	oJG915/oJG931
<i>feu</i> NTTT15	pJG351, <i>feuN</i> (78TVK80→TTT)	oJG915/oRV348, oRV349/oJG931	oJG915/oJG931

Table S4 Tn-seq predicted *S. meliloti* Essential Gene Candidates

GeneID	Gene	Rich insertions	Length	Function	Average Insertions / Base pair
SM_b20044	repC1	0	1311	replication initiation protein RepC	0.00
SM_b20056	-	0	1068	ABC transporter substrate-binding protein	0.00
SM_b20057	-	0	1044	ABC transporter permease	0.00
SM_b20058	-	0	801	ABC transporter ATP-binding protein	0.00
SM_b20995	engA	0	1431	GTP-binding protein EngA	0.00
SM_b21522	minE	0	264	cell division topological specificity factor MinE	0.00
SMa1413	-	0	276	hypothetical protein	0.00
SMa1595	-	0	444	hypothetical protein	0.00
SMa2239	-	0	780	hypothetical protein	0.00
SMa2391	repC2	0	1206	replication initiation protein RepC	0.00
SMa2393	repB2	0	1002	RepB2 replication protein	0.00
SMa2395	repA2	0	1176	RepA2 replication protein	0.00
SMc00005	fabI1	0	819	enoyl-ACP reductase	0.00
SMc00016	ispH	0	1053	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	0.00
SMc00021	ccrM	0	1131	adenine DNA methyltransferase	0.00
SMc00022	-	0	567	hypothetical protein	0.00
SMc00023	-	0	810	hypothetical protein	0.00
SMc00052	phaB2	0	420	monovalent cation/H+ antiporter subunit B	0.00
SMc00053	phaC2	0	378	monovalent cation/H+ antiporter subunit C	0.00
SMc00054	phaD2	0	1566	monovalent cation/H+ antiporter subunit D	0.00
SMc00057	phaG2	0	351	monovalent cation/H+ antiporter subunit G	0.00
SMc00069	pdxH	0	621	pyridoxamine 5'-phosphate oxidase	0.00
SMc00071	ialB	0	531	hypothetical protein	0.00
SMc00077	thrC1	0	1401	threonine synthase	0.00
SMc00118	-	0	423	hypothetical protein	0.00
SMc00149	fumC	0	1422	fumarate hydratase	0.00

SMc00152	rpiA	0	696	ribose-5-phosphate isomerase A	0.00
SMc00153	-	0	543	hypothetical protein	0.00
SMc00161	nadE	0	1683	NAD synthetase	0.00
SMc00180	hemF	0	912	coproporphyrinogen III oxidase	0.00
SMc00190	-	0	6270	hypothetical protein	0.00
SMc00192	fdx	0	321	ferredoxin, 2FE-2S FDI electron transport iron-sulfur protein	0.00
SMc00227	-	0	306	hypothetical protein	0.00
SMc00232	glmU	0	1371	bifunctional N-acetylglucosamine-1-phosphate uridyltransferase/glucosamine-1-phosphate acetyltransferase	0.00
SMc00292	recJ	0	1803	single-stranded-DNA-specific exonuclease	0.00
SMc00302	-	0	381	hypothetical protein	0.00
SMc00316	-	0	1278	2-octaprenyl-6-methoxyphenyl hydroxylase	0.00
SMc00323	rpsO	0	270	30S ribosomal protein S15	0.00
SMc00324	pnp	0	2154	polynucleotide phosphorylase/polyadenylase	0.00
SMc00326	fabI2	0	807	enoyl-ACP reductase	0.00
SMc00327	fabB	0	1221	3-oxoacyl-ACP synthase	0.00
SMc00328	fabA	0	516	3-hydroxydecanoyl-ACP dehydratase	0.00
SMc00333	aroA	0	1368	3-phosphoshikimate 1-carboxyvinyltransferase	0.00
SMc00342	-	0	627	hypothetical protein	0.00
SMc00362	infC	0	432	translation initiation factor IF-3	0.00
SMc00364	rplT	0	405	50S ribosomal protein L20	0.00
SMc00365	pheS	0	1083	phenylalanyl-tRNA synthetase subunit alpha	0.00
SMc00366	pheT	0	2427	phenylalanyl-tRNA synthetase subunit beta	0.00
SMc00394	guaA	0	1563	GMP synthase	0.00
SMc00407	gst4	0	693	glutathione S-transferase	0.00
SMc00408	uppP	0	807	UDP pyrophosphate phosphatase	0.00
SMc00415	dnaN	0	1119	DNA polymerase III subunit beta	0.00
SMc00462	folP	0	852	dihydropteroate synthase antibiotic resistance transferase folate protein	0.00
SMc00463	folB	0	372	dihydroneopterin aldolase DHNA lyase folate biosynthesis protein	0.00
SMc00465	folK	0	525	7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase	0.00

SMc00469	dksA	0	441	DnaK suppressor protein	0.00
SMc00471	-	0	2610	sensor histidine kinase transmembrane protein	0.00
SMc00475	alaS	0	2664	alanyl-tRNA synthetase	0.00
SMc00480	icd	0	1215	isocitrate dehydrogenase	0.00
SMc00485	rpsD	0	618	30S ribosomal protein S4	0.00
SMc00526	tyrS	0	1254	tyrosyl-tRNA synthetase	0.00
SMc00528	-	0	678	hypothetical protein	0.00
SMc00529	nifS	0	1167	pyridoxal-phosphate-dependent aminotransferase	0.00
SMc00530	-	0	1470	cysteine desulfurase activator complex subunit SufB	0.00
SMc00531	-	0	756	ABC transporter ATP-binding protein	0.00
SMc00532	-	0	1278	hypothetical protein	0.00
SMc00533	-	0	1245	pyridoxal-phosphate-dependent aminotransferase	0.00
SMc00551	-	0	699	phosphatidylserine decarboxylase	0.00
SMc00552	pssA	0	870	CDP-diacylglycerol--serine O-phosphatidyltransferase transmembrane protein	0.00
SMc00561	dnaB	0	1500	replicative DNA helicase	0.00
SMc00568	rpsF	0	450	30S ribosomal protein S6	0.00
SMc00572	fabG	0	738	3-ketoacyl-ACP reductase	0.00
SMc00574	fabF	0	1266	3-oxoacyl-ACP synthase	0.00
SMc00577	gmk	0	660	guanylate kinase	0.00
SMc00581	-	0	951	hypothetical protein	0.00
SMc00582	-	0	2346	hypothetical protein	0.00
SMc00583	-	0	1083	hypothetical protein	0.00
SMc00584	-	0	1173	hypothetical protein	0.00
SMc00586	-	0	450	DNA polymerase III subunit chi	0.00
SMc00595	ndk	0	423	nucleoside diphosphate kinase	0.00
SMc00601	pgsA	0	588	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	0.00
SMc00617	-	0	690	hypothetical protein	0.00
SMc00637	glmM	0	1353	phosphoglucosamine mutase	0.00
SMc00646	rpoH1	0	906	RNA polymerase factor sigma-32	0.00

SMc00652	-	0	666	hypothetical protein	0.00
SMc00654	ctrA	0	702	response regulator,controls chromosomal replication initiation protein	0.00
SMc00655	-	0	351	hypothetical protein	0.00
SMc00657	-	0	276	hypothetical protein	0.00
SMc00659	mnmA	0	1197	tRNA-specific 2-thiouridylase MnmA	0.00
SMc00700	cobS	0	996	cobalamin biosynthesis protein	0.00
SMc00704	rpmB	0	291	50S ribosomal protein L28	0.00
SMc00714	-	0	783	1-acyl-SN-glycerol-3-phosphate acyltransferase (PLSC) protein	0.00
SMc00811	ftsJ	0	738	cell division protein	0.00
SMc00851	glyS	0	2166	glycyl-tRNA synthetase subunit beta	0.00
SMc00855	glyQ	0	936	glycyl-tRNA synthetase subunit alpha	0.00
SMc00860	ispB	0	1017	octaprenyl-diphosphate synthase	0.00
SMc00862	ipk	0	906	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	0.00
SMc00868	atpF	0	486	F0F1 ATP synthase subunit B	0.00
SMc00869	atpF2	0	615	F0F1 ATP synthase subunit B'	0.00
SMc00871	atpB	0	753	F0F1 ATP synthase subunit A	0.00
SMc00892	lpxK	0	1041	tetraacyldisaccharide 4'-kinase	0.00
SMc00894	kdtA	0	1314	3-deoxy-D-manno-octulosonic-acid transferase	0.00
SMc00905	-	0	450	deaminase	0.00
SMc00908	ileS	0	2910	isoleucyl-tRNA synthetase	0.00
SMc00909	ribF	0	984	bifunctional riboflavin kinase/FMN adenyltransferase	0.00
SMc00919	hisS	0	1515	histidyl-tRNA synthetase	0.00
SMc00947	glnB	0	339	nitrogen regulatory protein PII	0.00
SMc00950	-	0	630	signal peptide protein	0.00
SMc00972	dxs	0	1938	1-deoxy-D-xylulose-5-phosphate synthase	0.00
SMc00984	-	0	975	hypothetical protein	0.00
SMc00988	ubiA	0	957	4-hydroxybenzoate polyprenyltransferase	0.00
SMc00996	-	0	1503	hypothetical protein	0.00
SMc01005	foIE	0	615	GTP cyclohydrolase I	0.00

SMc01010	thrS	0	1986	threonyl-tRNA synthetase	0.00
SMc01011	-	0	996	hypothetical protein	0.00
SMc01018	parE	0	2061	DNA topoisomerase IV subunit B	0.00
SMc01025	pyrG	0	1629	CTP synthetase	0.00
SMc01027	kdsA	0	843	2-dehydro-3-deoxyphosphooctonate aldolase	0.00
SMc01029	-	0	282	hypothetical protein	0.00
SMc01030	pdhAa	0	1047	pyruvate dehydrogenase alpha2 subunit protein	0.00
SMc01032	pdhB	0	1344	dihydrolipoamide S-acetyltransferase	0.00
SMc01037	lipA	0	969	lipoyl synthase	0.00
SMc01040	ispDF	0	1305	bifunctional 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase/2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	0.00
SMc01045	ntrX	0	1365	nitrogen regulation protein	0.00
SMc01096	dapE	0	1194	succinyl-diaminopimelate desuccinylase	0.00
SMc01100	fmt	0	936	methionyl-tRNA formyltransferase	0.00
SMc01101	def	0	537	peptide deformylase	0.00
SMc01109	metK	0	1281	S-adenosylmethionine synthetase	0.00
SMc01113	-	0	507	hypothetical protein	0.00
SMc01118	-	0	657	hypothetical protein	0.00
SMc01121	trpS	0	1065	tryptophanyl-tRNA synthetase	0.00
SMc01123	mviN	0	1608	virulence factor MviN-like protein	0.00
SMc01129	lspA	0	501	lipoprotein signal peptidase	0.00
SMc01134	ihfB	0	312	integration host factor subunit beta	0.00
SMc01136	-	0	669	hypothetical protein	0.00
SMc01137	-	0	561	hypothetical protein	0.00
SMc01138	-	0	813	ABC transporter ATP-binding protein	0.00
SMc01152	rpsT	0	267	30S ribosomal protein S20	0.00
SMc01156	aarF	0	1575	ubiquinone biosynthesis protein	0.00
SMc01161	dfp	0	1206	bifunctional phosphopantothenoilcysteine decarboxylase/phosphopantothenate synthase	0.00
SMc01167	dnaA	0	1524	chromosomal replication initiation protein	0.00
SMc01189	tmk	0	705	thymidylate kinase	0.00

SMc01190	-	0	1032	DNA polymerase III subunit delta'	0.00
SMc01192	metG	0	1551	methionyl-tRNA synthetase	0.00
SMc01209	coaD	0	492	phosphopantetheine adenylyltransferase	0.00
SMc01224	trxB	0	975	thioredoxin reductase	0.00
SMc01225	-	0	909	transcriptional regulator	0.00
SMc01231	gyrA	0	2796	DNA gyrase subunit A	0.00
SMc01233	ssb	0	525	single-stranded DNA-binding protein	0.00
SMc01237	nrd	0	3792	ribonucleotide-diphosphate reductase subunit alpha	0.00
SMc01283	rplQ	0	426	50S ribosomal protein L17	0.00
SMc01285	rpoA	0	1011	DNA-directed RNA polymerase subunit alpha	0.00
SMc01286	rpsK	0	390	30S ribosomal protein S11	0.00
SMc01287	rpsM	0	369	30S ribosomal protein S13	0.00
SMc01288	adk	0	579	adenylate kinase	0.00
SMc01289	secY	0	1341	preprotein translocase subunit SecY	0.00
SMc01290	rplO	0	471	50S ribosomal protein L15	0.00
SMc01292	rpsE	0	570	30S ribosomal protein S5	0.00
SMc01293	rplR	0	363	50S ribosomal protein L18	0.00
SMc01294	rplF	0	534	50S ribosomal protein L6	0.00
SMc01295	rpsH	0	399	30S ribosomal protein S8	0.00
SMc01296	rpsN	0	306	30S ribosomal protein S14	0.00
SMc01297	rplE	0	558	50S ribosomal protein L5	0.00
SMc01298	rplX	0	312	50S ribosomal protein L24	0.00
SMc01299	rplN	0	369	50S ribosomal protein L14	0.00
SMc01302	rplP	0	414	50S ribosomal protein L16	0.00
SMc01303	rpsC	0	714	30S ribosomal protein S3	0.00
SMc01304	rplV	0	390	50S ribosomal protein L22	0.00
SMc01305	rpsS	0	279	30S ribosomal protein S19	0.00
SMc01306	rplB	0	837	50S ribosomal protein L2	0.00
SMc01307	rplW	0	294	50S ribosomal protein L23	0.00

SMc01308	rplD	0	621	50S ribosomal protein L4	0.00
SMc01309	rplC	0	687	50S ribosomal protein L3	0.00
SMc01310	rpsJ	0	309	30S ribosomal protein S10	0.00
SMc01312	fusA1	0	2100	elongation factor G	0.00
SMc01313	rpsG	0	471	30S ribosomal protein S7	0.00
SMc01314	rpsL	0	372	30S ribosomal protein S12	0.00
SMc01316	rpoC	0	4206	DNA-directed RNA polymerase subunit beta'	0.00
SMc01317	rpoB	0	4143	DNA-directed RNA polymerase subunit beta	0.00
SMc01318	rplL	0	381	50S ribosomal protein L7/L12	0.00
SMc01319	rplJ	0	519	50S ribosomal protein L10	0.00
SMc01320	rplA	0	699	50S ribosomal protein L1	0.00
SMc01321	rplK	0	429	50S ribosomal protein L11	0.00
SMc01322	nusG	0	531	transcription antitermination protein NusG	0.00
SMc01331	ppnK	0	774	inorganic polyphosphate/ATP-NAD kinase	0.00
SMc01333	prfB	0	1029	peptide chain release factor 2	0.00
SMc01334	mrcA1	0	2454	penicillin-binding 1A transmembrane protein	0.00
SMc01344	accB	0	477	acetyl-CoA carboxylase biotin carboxyl carrier protein subunit	0.00
SMc01345	accC	0	1350	acetyl-CoA carboxylase biotin carboxylase subunit	0.00
SMc01348	-	0	402	NADH dehydrogenase	0.00
SMc01350	gatB	0	1551	aspartyl/glutamyl-tRNA amidotransferase subunit B	0.00
SMc01352	gatA	0	1482	aspartyl/glutamyl-tRNA amidotransferase subunit A	0.00
SMc01353	gatC	0	288	aspartyl/glutamyl-tRNA amidotransferase subunit C	0.00
SMc01362	-	0	612	glycerol-3-phosphate acyltransferase PlsY	0.00
SMc01371	celR1	0	372	2-component receiver domain-containing protein	0.00
SMc01375	dnaE	0	3510	DNA polymerase III subunit alpha	0.00
SMc01376	-	0	687	ABC transporter ATP-binding protein	0.00
SMc01407	pdxJ	0	753	pyridoxine 5'-phosphate synthase	0.00
SMc01416	-	0	360	hypothetical protein	0.00
SMc01435	miaA	0	912	tRNA delta(2)-isopentenylpyrophosphate transferase	0.00

SMc01442	folA	0	543	dihydrofolate reductase	0.00
SMc01444	thyA	0	795	thymidylate synthase	0.00
SMc01563	sigA	0	2055	RNA polymerase sigma factor RpoD	0.00
SMc01567	dnaG	0	2004	DNA primase	0.00
SMc01593	-	0	705	transcriptional regulator	0.00
SMc01720	rnpA	0	387	ribonuclease P	0.00
SMc01721	-	0	1788	inner membrane protein translocase component YidC	0.00
SMc01722	engB	0	654	ribosome biogenesis GTP-binding protein YsxC	0.00
SMc01732	dapD	0	891	2,3,4,5-tetrahydropyridine-2,6-carboxylate N-succinyltransferase	0.00
SMc01756	aspS	0	1788	aspartyl-tRNA synthetase	0.00
SMc01761	parC	0	2277	DNA topoisomerase IV subunit A	0.00
SMc01772	ribD	0	1206	riboflavin biosynthesis protein	0.00
SMc01773	ribE	0	621	riboflavin synthase subunit alpha	0.00
SMc01778	nusB	0	483	transcription antitermination protein NusB	0.00
SMc01781	-	0	507	hypothetical protein	0.00
SMc01784	plsX	0	1044	glycerol-3-phosphate acyltransferase PlsX	0.00
SMc01785	fabH	0	972	3-oxoacyl-ACP synthase	0.00
SMc01786	ihfA	0	339	integration host factor subunit alpha	0.00
SMc01803	rpsI	0	468	30S ribosomal protein S9	0.00
SMc01804	rplM	0	465	50S ribosomal protein L13	0.00
SMc01860	ftsI	0	1749	penicillin-binding transmembrane protein	0.00
SMc01861	murE	0	1461	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase	0.00
SMc01862	murF	0	1434	UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6- diaminopimelate--D-alanyl-D-alanine ligase	0.00
SMc01863	mraY	0	1101	phospho-N-acetylmuramoyl-pentapeptide-transferase	0.00
SMc01864	murD	0	1392	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase	0.00
SMc01865	ftsW	0	1155	cell division protein FtsW	0.00
SMc01866	murG	0	1125	UDP-diphospho-muramoylpentapeptide beta-N-acetylglucosaminyltransferase	0.00
SMc01867	murC	0	1416	UDP-N-acetylmuramate--L-alanine ligase	0.00
SMc01868	murB	0	975	UDP-N-acetylenolpyruvoylglucosamine reductase	0.00

SMc01871	ddl	0	927	D-alanine--D-alanine ligase	0.00
SMc01873	ftsA	0	1329	cell division protein	0.00
SMc01874	ftsZ1	0	1773	cell division protein FtsZ	0.00
SMc01875	lpxC	0	969	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	0.00
SMc01876	-	0	867	hypothetical protein	0.00
SMc01878	ligA	0	2154	NAD-dependent DNA ligase LigA	0.00
SMc01904	clpX	0	1278	ATP-dependent protease ATP-binding subunit ClpX	0.00
SMc01905	lon	0	2421	ATP-dependent protease LA protein	0.00
SMc01912	nuoA1	0	366	NADH dehydrogenase subunit A	0.00
SMc01913	nuoB1	0	579	NADH dehydrogenase subunit B	0.00
SMc01914	nuoC1	0	606	NADH dehydrogenase subunit C	0.00
SMc01915	nuoD1	0	1191	NADH dehydrogenase subunit D	0.00
SMc01920	nuoG1	0	2082	NADH dehydrogenase subunit G	0.00
SMc01921	nuoH	0	1044	NADH dehydrogenase subunit H	0.00
SMc01922	nuoI	0	495	NADH dehydrogenase subunit I	0.00
SMc01924	nuoK1	0	309	NADH dehydrogenase subunit K	0.00
SMc01926	nuoM	0	1512	NADH dehydrogenase subunit M	0.00
SMc01928	birA	0	768	biotin--[acetyl-CoA-carboxylase] synthetase	0.00
SMc01934	proS	0	1329	prolyl-tRNA synthetase	0.00
SMc01935	-	0	1227	hypothetical protein	0.00
SMc02060	lppB	0	1539	lipoprotein	0.00
SMc02064	serS	0	1284	seryl-tRNA synthetase	0.00
SMc02065	tatC	0	843	SEC-independent translocase transmembrane protein	0.00
SMc02073	argS	0	1758	arginyl-tRNA synthetase	0.00
SMc02078	exoR	0	807	exopolysaccharide biosynthesis regulatory protein	0.00
SMc02080	valS	0	2844	valyl-tRNA synthetase	0.00
SMc02082	tolC	0	1371	outer membrane secretion protein	0.00
SMc02087	gltA	0	1290	type II citrate synthase	0.00
SMc02089	lpxB	0	1170	lipid-A-disaccharide synthase	0.00

SMc02090	-	0	888	hypothetical protein	0.00
SMc02091	lpxA	0	813	UDP-N-acetylglucosamine acyltransferase	0.00
SMc02092	fabZ	0	465	(3R)-hydroxymyristoyl-ACP dehydratase	0.00
SMc02093	lpxD	0	1065	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	0.00
SMc02094	omp	0	2331	outer membrane transmembrane protein	0.00
SMc02097	uppS	0	744	UDP pyrophosphate synthase	0.00
SMc02098	frr	0	561	ribosome recycling factor	0.00
SMc02099	pyrH	0	723	uridylate kinase	0.00
SMc02100	tsf	0	924	elongation factor Ts	0.00
SMc02101	rpsB	0	768	30S ribosomal protein S2	0.00
SMc02113	cysE	0	828	serine acetyltransferase	0.00
SMc02116	-	0	1164	salicylate hydroxylase	0.00
SMc02122	fpr	0	813	ferredoxin--NADP reductase	0.00
SMc02139	-	0	522	hypothetical protein	0.00
SMc02218	-	0	1098	2'-deoxycytidine 5'-triphosphate deaminase	0.00
SMc02265	secD2	0	2535	preprotein translocase subunit SecD/SecF	0.00
SMc02305	murA	0	1293	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	0.00
SMc02313	-	0	357	hypothetical protein	0.00
SMc02375	-	0	861	hypothetical protein	0.00
SMc02377	etf	0	1665	electron transfer flavoprotein-ubiquinone oxidoreductase	0.00
SMc02386	amn	0	1503	AMP nucleosidase	0.00
SMc02396	-	0	1038	outer membrane protein	0.00
SMc02404	dapA	0	885	dihydrodipicolinate synthase	0.00
SMc02405	smpB	0	480	SsrA-binding protein	0.00
SMc02408	rpoZ	0	408	DNA-directed RNA polymerase subunit omega	0.00
SMc02432	-	0	1941	hypothetical protein	0.00
SMc02436	prfA	0	1083	peptide chain release factor 1	0.00
SMc02440	ubiG	0	747	3-demethylubiquinone-9 3-methyltransferase	0.00
SMc02463	sdhC	0	393	succinate dehydrogenase cytochrome B-556 subunit transmembrane protein	0.00

SMc02464	sdhD	0	381	succinate dehydrogenase membrane anchor subunit protein	0.00
SMc02465	sdhA	0	1842	succinate dehydrogenase flavoprotein subunit	0.00
SMc02466	sdhB	0	780	succinate dehydrogenase iron-sulfur subunit	0.00
SMc02475	-	0	540	outer membrane lipoprotein	0.00
SMc02477	-	0	267	hypothetical protein	0.00
SMc02483	sucB	0	1254	dihydrolipoamide succinyltransferase	0.00
SMc02496	priA	0	2205	primosome assembly protein PriA	0.00
SMc02498	atpH	0	567	F0F1 ATP synthase subunit delta	0.00
SMc02499	atpA	0	1530	F0F1 ATP synthase subunit alpha	0.00
SMc02500	atpG	0	885	F0F1 ATP synthase subunit gamma	0.00
SMc02501	atpD	0	1515	F0F1 ATP synthase subunit beta	0.00
SMc02502	atpC	0	411	F0F1 ATP synthase subunit epsilon	0.00
SMc02506	sitD	0	855	iron transport system membrane ABC transporter protein	0.00
SMc02509	sitA	0	906	iron-binding periplasmic ABC transporter protein	0.00
SMc02551	cysS	0	1401	cysteinyl-tRNA synthetase	0.00
SMc02560	chvI	0	723	transcriptional regulator	0.00
SMc02567	coaA	0	996	pantothenate kinase	0.00
SMc02651	era	0	942	GTP-binding protein Era	0.00
SMc02652	rncS	0	717	ribonuclease III	0.00
SMc02653	lepB	0	744	signal peptidase I transmembrane protein	0.00
SMc02678	-	0	501	hypothetical protein	0.00
SMc02681	lgt	0	843	prolipoprotein diacylglycerol transferase	0.00
SMc02686	prsA	0	933	ribose-phosphate pyrophosphokinase	0.00
SMc02692	rpLY	0	615	50S ribosomal protein L25	0.00
SMc02700	-	0	1260	hypothetical protein	0.00
SMc02712	-	0	903	hypothetical protein	0.00
SMc02756	-	0	2481	sensor histidine kinase	0.00
SMc02761	trxA	0	324	thioredoxin	0.00
SMc02763	folC	0	1344	bifunctional folylpolyglutamate synthase/dihydrofolate synthase	0.00

SMc02764	accD	0	915	acetyl-CoA carboxylase subunit beta	0.00
SMc02782	gyrB	0	2436	DNA gyrase subunit B	0.00
SMc02790	coaE	0	585	dephospho-CoA kinase	0.00
SMc02794	hemE	0	960	uroporphyrinogen decarboxylase	0.00
SMc02795	-	0	543	hypothetical protein	0.00
SMc02796	rho	0	1266	transcription termination factor Rho	0.00
SMc02798	gidA	0	1872	tRNA uridine 5-carboxymethylaminomethyl modification protein GidA	0.00
SMc02800	parA	0	795	chromosome partitioning protein ParA	0.00
SMc02801	parB	0	891	chromosome partitioning protein ParB	0.00
SMc02802	holA	0	1032	DNA polymerase III subunit delta	0.00
SMc02803	-	0	546	hypothetical protein	0.00
SMc02804	leuS	0	2631	leucyl-tRNA synthetase	0.00
SMc02837	dapB	0	819	dihydrodipicolinate reductase	0.00
SMc02878	nagA	0	1161	N-acetylglucosamine-6-phosphate deacetylase	0.00
SMc02898	kdsB	0	753	3-deoxy-manno-octulosonate cytidyltransferase	0.00
SMc02905	dnaX	0	1881	DNA polymerase III subunits gamma and tau	0.00
SMc02913	-	0	693	hypothetical protein	0.00
SMc02914	infB	0	2670	translation initiation factor IF-2	0.00
SMc02940	-	0	1362	hypothetical protein	0.00
SMc02942	pal	0	531	peptidoglycan-associated lipoprotein	0.00
SMc02977	-	0	792	hypothetical protein	0.00
SMc03104	hemA	0	1218	5-aminolevulinate synthase	0.00
SMc03105	dxr	0	1176	1-deoxy-D-xylulose 5-phosphate reductoisomerase	0.00
SMc03172	gltX	0	1458	glutamyl-tRNA synthetase	0.00
SMc03173	lysS	0	1497	lysyl-tRNA synthetase	0.00
SMc03187	cobM	0	759	precorrin-4 C(11)-methyltransferase	0.00
SMc03188	cobL	0	1242	precorrin-6Y C5,15-methyltransferase (decarboxylating) protein	0.00
SMc03190	cobJ	0	765	precorrin-3B C(17)-methyltransferase	0.00
SMc03191	cobI	0	756	precorrin-2 C(20)-methyltransferase	0.00

SMc03192	cobH	0	633	precorrin-8X methylmutase	0.00
SMc03193	cobG	0	1368	cobalamin biosynthesis protein	0.00
SMc03194	cobF	0	771	precorrin 6A synthase	0.00
SMc03231	hemC	0	930	porphobilinogen deaminase	0.00
SMc03232	hemD	0	708	uroporphyrinogen-III synthase	0.00
SMc03239	ppa	0	534	inorganic pyrophosphatase	0.00
SMc03763	-	0	1323	cytosine-specific methyltransferase	0.00
SMc03770	rplU	0	372	50S ribosomal protein L21	0.00
SMc03772	rpmA	0	270	50S ribosomal protein L27	0.00
SMc03775	obgE	0	1131	GTPase ObgE	0.00
SMc03778	nadD	0	588	nicotinic acid mononucleotide adenyltransferase	0.00
SMc03783	ctpA	0	1323	carboxy-terminal processing protease precursor signal peptide protein	0.00
SMc03809	-	0	663	hypothetical protein	0.00
SMc03820	-	0	684	transcriptional regulator	0.00
SMc03832	-	0	504	signal peptide protein	0.00
SMc03846	acnA	0	2691	aconitate hydratase	0.00
SMc03856	dapF	0	921	diaminopimelate epimerase	0.00
SMc03857	ffh	0	1542	signal recognition particle protein	0.00
SMc03859	rpsP	0	375	30S ribosomal protein S16	0.00
SMc03861	trmD	0	723	tRNA (guanine-N(1)-)-methyltransferase	0.00
SMc03867	-	0	651	hypothetical protein	0.00
SMc03874	-	0	570	hypothetical protein	0.00
SMc03875	-	0	339	ferredoxin protein	0.00
SMc03888	ispG	0	1254	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	0.00
SMc03925	pgm	0	1629	phosphoglucomutase	0.00
SMc03956	tolA	0	1071	signal peptide protein	0.00
SMc03957	tolR	0	453	transport transmembrane protein	0.00
SMc03958	tolQ	0	720	transport transmembrane protein	0.00
SMc03991	-	0	1815	ABC transporter ATP-binding protein	0.00

SMc03994	suhB	0	801	inositol monophosphatase	0.00
SMc03995	-	0	1146	hypothetical protein	0.00
SMc04006	-	0	438	hypothetical protein	0.00
SMc04010	-	0	1689	hypothetical protein	0.00
SMc04017	omp10	0	369	outer membrane lipoprotein	0.00
SMc04019	hemH	0	1029	ferrochelataase	0.00
SMc04044	-	0	363	2-component receiver domain-containing protein	0.00
SMc04083	cynT	0	684	carbonic anhydrase	0.00
SMc04198	-	0	690	phage repressor protein	0.00
SMc04214	cobT	0	1017	nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase	0.00
SMc04215	cobS	0	789	cobalamin synthase	0.00
SMc04268	msbB	0	939	lipid A biosynthesis lauroyl acyltransferase	0.00
SMc04279	cobD	0	984	cobalamin biosynthesis protein	0.00
SMc04282	cobB	0	1290	cobyrinic acid a,c-diamide synthase	0.00
SMc04284	cobA	0	840	uroporphyrin-III C-methyltransferase	0.00
SMc04302	cobO	0	645	cob(II)yrinic acid a,c-diamide adenosyltransferase	0.00
SMc04304	cobW	0	1065	cobalamine biosynthesis protein	0.00
SMc04305	cobU	0	528	adenosylcobinamide kinase/adenosylcobinamide-phosphate guanylyltransferase	0.00
SMc04309	cobQ	0	1455	cobyrinic acid synthase	0.00
SMc04410	asd	0	1035	aspartate-semialdehyde dehydrogenase	0.00
SMc04446	chvG	0	1788	histidine kinase sensory transmembrane protein	0.00
SMc04458	secA	0	2712	preprotein translocase subunit SecA	0.00
SMc04459	ftsH	0	1938	metalloprotease transmembrane protein	0.00
SMc04461	tolB	0	1311	translocation protein TolB	0.00
SM_b20204	pqqA	0	96	coenzyme PQQ synthesis protein PqqA	0.00
SM_b20727	-	0	234	hypothetical protein	0.00
SM_b20911	-	0	207	hypothetical protein	0.00
SM_b21672	-	0	168	hypothetical protein	0.00
SM_b21686	-	0	186	hypothetical protein	0.00

SM_b21687	-	0	195	hypothetical protein	0.00
SM_b21689	-	0	183	hypothetical protein	0.00
SMa0126	cspA8	0	210	CspA8 Cold shock family protein	0.00
SMa0285	-	0	225	hypothetical protein	0.00
SMa0625	-	0	162	hypothetical protein	0.00
SMa0811	fdxN	0	195	FdxN ferredoxin	0.00
SMa0833	-	0	141	hypothetical protein	0.00
SMa1032	-	0	246	hypothetical protein	0.00
SMa1063	-	0	252	hypothetical protein	0.00
SMa1361	-	0	183	hypothetical protein	0.00
SMa1700	-	0	210	hypothetical protein	0.00
SMa1766	-	0	252	hypothetical protein	0.00
SMa1767	-	0	168	hypothetical protein	0.00
SMa2335	-	0	204	hypothetical protein	0.00
SMc00065	-	0	252	signal peptide protein	0.00
SMc00193	-	0	168	hypothetical protein	0.00
SMc00363	rplM	0	204	50S ribosomal protein L35	0.00
SMc00567	rpsR	0	249	30S ribosomal protein S18	0.00
SMc00573	acpP	0	237	acyl carrier protein	0.00
SMc00870	atpE	0	228	F0F1 ATP synthase subunit C	0.00
SMc00899	-	0	255	hypothetical protein	0.00
SMc01048	hfq	0	243	RNA-binding protein Hfq	0.00
SMc01291	rplD	0	204	50S ribosomal protein L30	0.00
SMc01300	rpsQ	0	237	30S ribosomal protein S17	0.00
SMc01301	rplC	0	201	50S ribosomal protein L29	0.00
SMc01323	secE	0	201	preprotein translocase subunit SecE	0.00
SMc01357	-	0	156	hypothetical protein	0.00
SMc01369	rplG	0	168	50S ribosomal protein L33	0.00
SMc01505	-	0	168	hypothetical protein	0.00

SMc02051	-	0	144	hypothetical protein	0.00
SMc02067	tatA	0	207	twin arginine translocase A	0.00
SMc02310	infA	0	219	translation initiation factor IF-1	0.00
SMc02319	-	0	189	hypothetical protein	0.00
SMc02718	-	0	138	hypothetical protein	0.00
SMc02900	-	0	138	hypothetical protein	0.00
SMc03284	-	0	180	hypothetical protein	0.00
SMc03881	rpmF	0	186	50S ribosomal protein L32	0.00
SMc03934	rpsU	0	255	30S ribosomal protein S21	0.00
SMc03986	-	0	156	hypothetical protein	0.00
SMc03990	rpmE	0	222	50S ribosomal protein L31	0.00
SMc05019	-	0	153	hypothetical protein	0.00
SMc01925	nuoL	5	1995	NADH dehydrogenase subunit L	0.00
SMc00051	phaA2	6	2376	monovalent cation/H ⁺ antiporter subunit A	0.00
SMc01927	nuoN	4.5	1443	NADH dehydrogenase subunit N	0.00
SMc02836	-	6	1806	ABC transporter ATP-binding protein	0.00
SMc01035	lpdA1	5.5	1446	dihydrolipoamide dehydrogenase	0.00
SMc00007	aroC	5	1098	chorismate synthase	0.00
SMc01277	-	5.5	786	hypothetical protein	0.01
SMc01918	nuoF1	10.5	1305	NADH dehydrogenase I subunit F	0.01
SMc00985	-	16	1443	oxidoreductase	0.01
SMc02682	-	13	1119	hypothetical protein	0.01
SMc02058	-	4.5	333	YAJC protein	0.01
SMc00508	purB	19.5	1308	adenylosuccinate lyase	0.01
SMc02081	-	12	756	hypothetical protein	0.02
SMc01014	-	12.5	711	hypothetical protein	0.02
SMc03238	-	5.5	294	hypothetical protein	0.02
SMA1241	napE	3.5	186	NapE component of periplasmic nitrate reductase	0.02
SMc00696	aroB	22	1134	3-dehydroquinate synthase	0.02

SMA2263	-	7.5	336	hypothetical protein	0.02
SMA0594	-	23.5	1032	hypothetical protein	0.02
SM_b21254	-	20.5	831	hypothetical protein	0.02
SM_b20302	-	4.5	162	hypothetical protein	0.03
SMc01836	-	9	318	hypothetical protein	0.03
SMc00154	gor	46	1392	glutathione reductase	0.03
SMc01343	aroQ	15	447	3-dehydroquinate dehydratase	0.03
SMc01880	panC	29.5	876	pantoate--beta-alanine ligase	0.03
SMc03967	ruvC	17.5	513	Holliday junction resolvase	0.03
SMc02075	-	11.5	330	hypothetical protein	0.03
SMc00130	-	5	141	hypothetical protein	0.04
SM_b21270	-	19.5	465	transcriptional regulator	0.04
SMc01336	rne	116.5	2775	ribonuclease E protein	0.04
SMc01038	-	20.5	450	hypothetical protein	0.05
SMc02201	-	9.5	198	hypothetical protein	0.05
SMc01215	carB	188.5	3492	carbamoyl phosphate synthase large subunit	0.05
SMc00017	thrB	57.5	981	homoserine kinase	0.06
SMc03059	folD2	57.5	900	bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/ 5,10-methylene-tetrahydrofolate cyclohydrolase	0.06
SMc02438	lysC	89	1275	aspartate kinase	0.07
SMc01569	carA	86.5	1206	carbamoyl phosphate synthase small subunit	0.07
SM_b21175	phoE	72.5	963	phosphate uptake ABC transporter permease	0.08
SMc01881	panB	65	822	3-methyl-2-oxobutanoate hydroxymethyltransferase	0.08
SMc03965	ruvB	85.5	1041	Holliday junction DNA helicase RuvB	0.08
SMc03850	ccmD	15	177	heme exporter D (cytochrome C-type biogenesis protein) transmembrane	0.08
SMc00695	aroK	53	579	shikimate kinase	0.09
SMc02141	phoU	66.5	714	phosphate transporter PhoU	0.09
SM_b20832	rkpS	65	660	cell surface polysaccharide export ABC-2 transporter ATP-binding protein	0.10
SMc00392	-	25	243	hypothetical protein	0.10
SMc00059	-	163	1578	sensor histidine kinase	0.10

SMc02759	-	330.5	3180	hypothetical protein	0.10
SMc00495	purC	80	765	phosphoribosylaminoimidazolesuccinocarboxamide synthase	0.10
SMc00074	-	3.5	2913	transmembrane signal peptide protein	0.00
SMc02797	trmE	2.5	1320	tRNA modification GTPase TrmE	0.00
SMc02163	pgi	6	1626	glucose-6-phosphate isomerase	0.00
SMc03230	gcp	5.5	1083	DNA-binding/iron metalloprotein/AP endonuclease	0.01
SMc02654	acpS	3	420	4'-phosphopantetheinyl transferase	0.01
SMc04009	-	2.5	273	hypothetical protein	0.01
SMc01031	pdhAb	14.5	1383	pyruvate dehydrogenase subunit beta	0.01
SMc00825	gsh1	17	1374	glutamate--cysteine ligase precursor protein	0.01
SMc02083	-	10.5	660	methyltransferase (PCM-like) transmembrane protein	0.02
SMc01906	hrm	6	273	histone-like protein	0.02
SMa0333	-	6	213	hypothetical protein	0.03
SMc02640	lpsL	29.5	1026	UDP-glucuronic acid epimerase	0.03
SMc02066	tatB	19	648	sec-independent translocase	0.03
SMc02481	sucD	28.5	903	succinyl-CoA synthetase subunit alpha	0.03
SMc00785	rirA	15	465	iron-responsive transcriptional regulator	0.03
SMc00357	efp	19.5	570	elongation factor P	0.03
SM_b21668	-	9	228	hypothetical protein	0.04
SMc00043	sodB	25	603	superoxide dismutase Fe protein	0.04
SMc01144	rph	31.5	720	ribonuclease PH	0.04
SMc02911	-	28.5	618	hypothetical protein	0.05
SMc00580	pdxA	48.5	1029	4-hydroxythreonine-4-phosphate dehydrogenase	0.05
SMc02480	sucC	58	1197	succinyl-CoA synthetase subunit beta	0.05
SMc02792	maf	30.5	600	Maf-like protein	0.05
SMc02912	nusA	83.5	1629	transcription elongation factor NusA	0.05
SMa0211	-	10.5	195	hypothetical protein	0.05
SMa1971	-	16	297	hypothetical protein	0.05
SMa1078	-	13	225	hypothetical protein	0.06

SMc01799	-	12.5	189	signal peptide protein	0.07
SMc03833	-	25.5	378	hypothetical protein	0.07
SMa1323	rctA	26	369	Transcription regulator	0.07
SMa0964	-	17	231	hypothetical protein	0.07
SMc02520	glpD	113.5	1512	glycerol-3-phosphate dehydrogenase	0.08
SMc03189	cobK	66.5	801	cobalt-precorrin-6x reductase	0.08
SMc01108	trmB	64	756	tRNA (guanine-N(7)-)-methyltransferase	0.08
SMa0990	-	41	480	hypothetical protein	0.09
SMc01326	tuf	101.5	1176	elongation factor Tu	0.09
SMc01347	-	42	453	hypothetical protein	0.09
SMc01311	tuf	110	1176	elongation factor Tu	0.09
SMc00796	-	16	165	hypothetical protein	0.10
SMa1139	-	27	267	hypothetical protein	0.10
SMc02858	dnaJ	122.5	1140	chaperone protein DnaJ	0.11
SMa0181	cspA7	22.5	204	CspA5 cold shock protein transcriptional regulator	0.11
SMc00419	gshB1	108	948	glutathione synthetase	0.11
SMc00248	ccsA	86.5	744	cytochrome C-type biogenesis protein	0.12
SMc02752	-	1.5	447	hypothetical protein	0.00
SMc02096	cdsA	5.5	834	phosphatidate cytidyltransferase	0.01
SMc02857	dnaK	14.5	1926	molecular chaperone DnaK	0.01
SMc00538	-	3.5	336	hypothetical protein	0.01
SMc00940	-	3	252	hypothetical protein	0.01
SMc00701	cobT	24.5	1896	cobalamin biosynthesis protein	0.01
SMc04437	-	4.5	240	hypothetical protein	0.02
SMc03804	-	25.5	1227	2-octaprenyl-6-methoxyphenyl hydroxylase	0.02
SMa0118	-	11	528	hypothetical protein	0.02
SMc00131	-	11.5	489	transcriptional regulator	0.02
SMc01449	-	9	309	hypothetical protein	0.03
SM_b21176	phoD	29	906	phosphate uptake ABC transporter substrate-binding protein precursor	0.03

SMc00198	-	5	153	hypothetical protein	0.03
SMc04434	rpmH	6	135	50S ribosomal protein L34	0.04
SM_b21174	phoT	90.5	1518	phosphate uptake ABC transporter permease	0.06
SMc04432	-	19	318	hypothetical protein	0.06
SMc02369	-	139	2319	sensor histidine kinase transmembrane protein	0.06
SMc00251	-	20	264	hypothetical protein	0.08
SMA0852	nodF	22	282	acyl carrier protein	0.08
SMc00058	mucR	38.5	432	transcriptional regulator	0.09
SMc02274	rkpU	134.5	1224	capsule polysaccharide exporter protein	0.11
SM_b20906	-	35.5	306	hypothetical protein	0.12
SMc00690	accA	0.5	954	acetyl-CoA carboxylase carboxyltransferase subunit alpha	0.00
SMc02507	sitC	1	858	iron transport system membrane ABC transporter protein	0.00
SMc03744	-	0.5	288	hypothetical protein	0.00
SMc00335	rpsA	3	1707	30S ribosomal protein S1	0.00
SMc02482	sucA	7	2997	2-oxoglutarate dehydrogenase E1 component	0.00
SMc00720	-	1	360	2-component receiver domain-containing protein	0.00
SM_b20084	-	0.5	174	hypothetical protein	0.00
SMc04181	-	1	321	transmembrane protein	0.00
SMc02052	-	0.5	144	hypothetical protein	0.00
SMc01364	topA	14.5	2703	DNA topoisomerase I	0.01
SMc00738	-	3	510	hypothetical protein	0.01
SMc04303	cobN	23	3828	cobaltochelataase subunit CobN	0.01
SMc01183	lexA	4.5	717	LexA repressor	0.01
SM_b20956	exoL	8.5	1212	glucosyltransferase	0.01
SMc02693	pth	9	720	peptidyl-tRNA hydrolase	0.01
SM_b20085	-	2	147	hypothetical protein	0.01
SMc00823	-	5.5	396	hypothetical protein	0.01
SM_b20950	exoT	23	1485	transport protein, Wzx	0.02
SMc03808	ftsK	41	2646	cell division transmembrane protein	0.02

SMc02508	sitB	15	936	iron transport ATP-binding ABC transporter protein	0.02
SMa0872	orf110	5.5	333	hypothetical protein	0.02
SMa1024	-	7.5	423	hypothetical protein	0.02
SMa0341	-	3.5	195	hypothetical protein	0.02
SMc05001	-	5.5	291	hypothetical protein	0.02
SMc00427	prfC	32.5	1584	peptide chain release factor RF-3 protein	0.02
SMc02487	lpdA2	32	1407	dihydrolipoamide dehydrogenase	0.02
SM_b20254	-	6.5	276	hypothetical protein	0.02
SMc00562	-	12	453	transcriptional regulator	0.03
SMc01470	-	4.5	159	signal peptide protein	0.03
SMa1169	-	10	342	hypothetical protein	0.03
SMc04320	rpsU	7	237	30S ribosomal protein S21	0.03
SMa1706	-	12.5	423	hypothetical protein	0.03
SMa1678	-	8.5	282	hypothetical protein	0.03
SMc04408	-	8.5	261	virulence-associated protein	0.03
SMa1319	virB2	10	300	VirB2 type IV secretion protein	0.03
SMc01218	greA	16	477	transcription elongation factor GreA	0.03
SMc00018	rnhA	16	462	ribonuclease H	0.03
SMc01923	nuoJ	21.5	615	NADH dehydrogenase subunit J	0.03
SMc03863	rplS	19	534	50S ribosomal protein L19	0.04
SMc01024	secG	17.5	486	preprotein translocase subunit SecG	0.04
SMc00959	-	6.5	177	hypothetical protein	0.04
SMc03253	-	34.5	843	L-proline 3-hydroxylase	0.04
SMa0355	-	20.5	495	LysR family transcriptional regulator	0.04
SMc02988	-	10	240	hypothetical protein	0.04
SMc02110	clpS	15	354	ATP-dependent Clp protease adaptor protein ClpS	0.04
SMc00619	-	66.5	1524	hypothetical protein	0.04
SMc01141	ptsN	20.5	465	nitrogen regulatory IIA protein	0.04
SMa1079	tspO	24.5	543	TspO/MBR family protein	0.05

SMc00565	rplI	26	576	50S ribosomal protein L9	0.05
SMa1610	-	12	264	transposase	0.05
SMc04085	-	223	4788	hypothetical protein	0.05
SMc01263	-	12.5	267	hypothetical protein	0.05
SMa0806	syrB3	21.5	456	SyrB-like regulator	0.05
SM_b20337	-	29	612	transcriptional regulator	0.05
SM_b20810	-	97	2034	membrane-located cell surface saccharide acetylase	0.05
SMc01427	-	11.5	225	hypothetical protein	0.05
SMa0697	arcC	50	975	carbamate kinase	0.05
SMa0870	nodD1	50	927	NodD1 nod-box dependent transcriptional activator	0.05
SM_b20413	-	12.5	228	hypothetical protein	0.05
SMc04004	-	35	615	signal peptide protein	0.06
SMc04185	-	50	843	glycosyltransferase	0.06
SM_b20909	-	17.5	294	hypothetical protein	0.06
SMc03999	-	10.5	174	hypothetical protein	0.06
SM_b21690	exoW	61.5	960	glucosyltransferase	0.06
SMc04278	acpXL	18.5	288	acyl carrier protein	0.06
SMa0969	-	48.5	750	Response regulator	0.06
SMa1698	syrB1	29.5	456	syrB1 regulator	0.06
SM_b20958	exoM	61.5	930	glucosyltransferase	0.07
SMc01205	-	50	756	amino-acid-binding periplasmic signal peptide protein	0.07
SMc00949	-	22	330	hypothetical protein	0.07
SMa1253	-	18	270	hypothetical protein	0.07
SMa0941	-	26	384	hypothetical protein	0.07
SM_b21126	eccH2	53.5	786	enoyl-CoA hydratase	0.07
SMa0854	nodG	50.5	738	3-ketoacyl-ACP reductase	0.07
SMc00187	fbfF	40	579	ubiquinol-cytochrome C reductase iron-sulfur subunit protein	0.07
SMc03860	rimM	39	564	16S rRNA-processing protein RimM	0.07
SMa2273	-	25	351	hypothetical protein	0.07

SMc02799	gidB	46	642	16S rRNA methyltransferase GidB	0.07
SM_b20553	-	14	195	hypothetical protein	0.07
SMc01044	ntrY	163.5	2262	nitrogen regulation transmembrane protein	0.07
SMa0357	-	20.5	282	hypothetical protein	0.07
SMc03854	ftsY	125.5	1647	cell division protein	0.08
SM_b20831	rkpR/kpsE	73	954	polysaccharide export-associated protein	0.08
SMc04313	-	18	234	hypothetical protein	0.08
SMc03252	proB2	66.5	849	gamma-glutamyl kinase	0.08
SMc01872	ftsQ	74.5	930	cell division transmembrane protein	0.08
SMc01155	ubiE	67	813	ubiquinone/menaquinone biosynthesis methyltransferase	0.08
SMc02048	gcvH	30	363	glycine cleavage system protein H	0.08
SMc01788	-	18	216	hypothetical protein	0.08
SMc00575	-	100.5	1200	hypothetical protein	0.08
SMc00189	fbcC	74.5	876	cytochrome C1 protein	0.09
SMc03244	-	19	222	hypothetical protein	0.09
SMc03949	ntrP	23.5	273	nitrogen regulatory protein	0.09
SMa0726	-	34	393	hypothetical protein	0.09
SMc02109	clpA	219	2517	ATP-dependent Clp protease ATP-binding subunit protein	0.09
SM_b20944	exoQ	116	1308	polysaccharide polymerase, Wzy protein	0.09
SM_b21464	-	69	774	GntR family transcriptional regulator	0.09
SM_b20873	allA	46	513	ureidoglycolate hydrolase	0.09
SMc00600	moaD	23.5	261	molybdopterin MPT converting factor subunit 1	0.09
SM_b20412	-	15	165	hypothetical protein	0.09
SMa5018	-	27	288	hypothetical protein	0.09
SMa2323	-	40	426	hypothetical protein	0.09
SMc01223	-	45	471	transcriptional regulator	0.10
SMc01957	-	30.5	315	hypothetical protein	0.10
SMa1990	-	35	357	hypothetical protein	0.10
SM_b21156	-	56	567	hypothetical protein	0.10

SM_b21501	-	36	363	hypothetical protein	0.10
SMc00703	-	63.5	639	hypothetical protein	0.10
SMa1635	-	46.5	459	hypothetical protein	0.10
SM_b21359	-	46.5	459	LysR family transcriptional regulator	0.10
SMc04436	relE	28.5	279	hypothetical protein	0.10
SMc01724	-	51	495	hypothetical protein	0.10
SMc00176	-	144	1377	hypothetical protein	0.10
SMc01450	-	54.5	516	hypothetical protein	0.11
SMc00320	rbfA	43.5	408	ribosome-binding factor A	0.11
SMa0412	-	34	315	hypothetical protein	0.11
SMa2241	-	135	1248	hypothetical protein	0.11
SMc03988	-	20.5	189	hypothetical protein	0.11
SMc04094	-	76	699	hypothetical protein	0.11
SMa0223	-	86	786	TetR family transcriptional regulator	0.11
SMa1008	-	50	456	hypothetical protein	0.11
SMc03765	-	175	1590	hypothetical protein	0.11
SMa0773	noeA	160.5	1431	NoeA host specific nodulation protein	0.11
SMa1171	-	29.5	261	hypothetical protein	0.11
SMa0667	-	55.5	489	hypothetical protein	0.11
SMa1881	-	67.5	594	hypothetical protein	0.11
SMa2243	-	106.5	927	hypothetical protein	0.11
SMa5007	-	44	381	transcriptional regulator	0.12
SM_b20598	repA3	147	1257	replication protein	0.12
SM_b20552	-	86	732	hypothetical protein	0.12
SM_b20949	exoV	112	951	pyruvyltransferase	0.12
SMc02753	-	47.5	402	PTS system transporter subunit IIA	0.12
SMa2237	-	119	1005	hypothetical protein	0.12
SMc00794	-	94.5	798	two-component response regulator	0.12
*Essential gene candidates have a normalized insertions / gene length < 1.2					

Table S5 Tn-seq predicted *S. meliloti* genes required for growth in RSM and on *Medicago truncatula* roots

GeneID	Gene	Rich	Min	<i>M. truncatula</i>	Function	RSM Fitness	Root Fitness
SMc00711	<i>tyrC</i>	770	0	0.0	cyclohexadienyl dehydrogenase	0.0	0.0
SMc00640	<i>serC</i>	639	0	0.7	phosphoserine aminotransferase	0.0	0.0
SMc02570	<i>hisA</i>	1974.5	4.5	0.7	1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino] imidazole-4-carboxamide isomerase	0.0	0.0
SMc00643	<i>purA</i>	307.5	1	0.0	adenylosuccinate synthetase	0.0	0.0
SMc00918	<i>hisZ</i>	991	0.5	5.0	ATP phosphoribosyltransferase regulatory subunit	0.0	0.0
SMc02307	<i>hisD</i>	3444.5	13.5	16.0	histidinol dehydrogenase	0.0	0.0
SMc00723	<i>lysA</i>	629.5	2.5	4.0	diaminopimelate DAP decarboxylase	0.0	0.0
SMc00917	<i>hisG</i>	1321.5	12	1.7	ATP phosphoribosyltransferase catalytic subunit	0.0	0.0
SMc02137	<i>argF1</i>	1145	1.5	12.3	ornithine carbamoyltransferase	0.0	0.0
SMc02568	<i>hisE</i>	465	0	5.7	phosphoribosyl-ATP pyrophosphatase	0.0	0.0
SMc02791	<i>aroE</i>	204	2.5	0.0	shikimate 5-dehydrogenase	0.0	0.0
SMc01726	<i>argB</i>	1061.5	4	14.0	acetylglutamate kinase	0.0	0.0
SMc01004	<i>hisI</i>	323.5	6.5	0.7	phosphoribosyl-AMP cyclohydrolase	0.0	0.0
SMc01801	<i>argC</i>	1277	3	26.7	N-acetyl-gamma-glutamyl-phosphate reductase	0.0	0.0
SMc02569	<i>hisF</i>	1138	26.5	7.3	imidazole glycerol phosphate synthase subunit HisF	0.0	0.0
SMc03112	<i>metH</i>	4145	31.5	131.0	B12-dependent methionine synthase	0.0	0.0
SMc02572	<i>hisH</i>	951	38.5	2.0	imidazole glycerol phosphate synthase subunit HisH	0.0	0.0
SMc01843	<i>metF</i>	617.5	1	25.3	5,10-methylenetetrahydrofolate reductase	0.0	0.0
SMc02850	<i>polA</i>	366	0.5	15.3	DNA polymerase I	0.0	0.0
SMc03826	<i>argG</i>	325.5	1	13.7	argininosuccinate synthase	0.0	0.0
SMc00155	<i>aroF</i>	264.5	12.5	0.0	DAHP synthetase prtein	0.0	0.0
SMc04026	<i>gluD</i>	3274.5	9	247.3	glutamate synthase	0.0	0.1
SMc03797	<i>metA</i>	594.5	1.5	45.3	homoserine O-succinyltransferase	0.0	0.1
SMc01494	<i>serB</i>	1464.5	1.5	120.3	phosphoserine phosphatase	0.0	0.1
SMc00641	<i>serA</i>	1241	3	102.7	D-3-phosphoglycerate dehydrogenase	0.0	0.1
SMc00293	<i>thrA</i>	1047.5	6.5	86.0	homoserine dehydrogenase	0.0	0.1
SMc01431	<i>ilvI</i>	2942.5	18	273.7	acetolactate synthase 3 catalytic subunit	0.0	0.1
SMc04028	<i>gluB</i>	6980	17.5	732.3	glutamate synthase	0.0	0.1
SMc00993	<i>purD</i>	1121.5	4	120.0	phosphoribosylamine--glycine ligase	0.0	0.1
SMc00615	<i>purM</i>	1144.5	2.5	142.0	phosphoribosylaminoimidazole synthetase	0.0	0.1
SMc04346	<i>ilvC</i>	895.5	2.5	112.0	ketol-acid reductoisomerase	0.0	0.1
SMc00554	<i>purF</i>	2791	4.5	353.7	amidophosphoribosyltransferase	0.0	0.1
SMc04045	<i>ilvD2</i>	3446.5	7.5	475.3	dihydroxy-acid dehydratase	0.0	0.1
SMc02717	<i>leuA1</i>	1377	2.5	192.7	2-isopropylmalate synthase	0.0	0.1
SMc00235	<i>trpD</i>	1257	0.5	178.0	anthranilate phosphoribosyltransferase	0.0	0.1
SMc04001	<i>purE</i>	2677	15	371.0	phosphoribosylaminoimidazole carboxylase catalytic subunit protein	0.0	0.1
SMc04405	<i>leuB</i>	3937	24.5	572.3	3-isopropylmalate dehydrogenase	0.0	0.1
SMc00493	<i>purQ</i>	497.5	0.5	76.3	phosphoribosylformylglycinamide synthase I	0.0	0.2

SMc00488	<i>purL</i>	3540	4	558.3	phosphoribosylformylglycinamide synthase II	0.0	0.2
SMc02755	<i>ahcY</i>	779.5	18.5	109.3	S-adenosyl-L-homocysteine hydrolase	0.0	0.1
SMc02725	<i>trpE</i>	2318.5	4	406.7	anthranilate synthase	0.0	0.2
SMc02767	<i>trpF</i>	712.5	8	121.0	N-(5'-phosphoribosyl)anthranilate isomerase	0.0	0.2
SMc00614	<i>purN</i>	1034.5	16.5	173.7	phosphoribosylglycinamide formyltransferase	0.0	0.2
SMc03823	<i>leuC</i>	2559.5	51	425.3	isopropylmalate isomerase large subunit	0.0	0.2
SMc04088	<i>purH</i>	1687.5	15	301.0	bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase	0.0	0.2
SMc00815	<i>guaB</i>	1410	5	263.7	inosine 5'-monophosphate dehydrogenase	0.0	0.2
SMc02899	<i>pheA</i>	2212	21	422.0	prephenate dehydratase	0.0	0.2
SMc00236	<i>trpC</i>	623	2.5	128.3	Indole-3-glycerol phosphate synthase	0.0	0.2
SMc03795	<i>leuD</i>	305.5	1.5	62.7	isopropylmalate isomerase small subunit	0.0	0.2
SMc00760	<i>recA</i>	251.5	3.5	49.7	recombinase A	0.0	0.2
SMc02217	<i>metZ</i>	1245.5	23	249.3	O-succinylhomoserine sulfhydrylase	0.0	0.2
SMc01770	<i>glyA</i>	1788.5	1.5	473.7	serine hydroxymethyltransferase	0.0	0.3
SMc02766	<i>trpB</i>	4317.5	23.5	1124.3	tryptophan synthase subunit beta	0.0	0.3
SMc02479	<i>mdh</i>	438	73	53.0	malate dehydrogenase	0.2	0.1
SMc02765	<i>trpA</i>	2198.5	2.5	631.0	tryptophan synthase subunit alpha	0.0	0.3
SMc03979	<i>gap</i>	1125	15.5	310.0	glyceraldehyde-3-phosphate dehydrogenase	0.0	0.3
SMc00914	-	903	85.5	176.7	oxidoreductase	0.1	0.2
SMc04002	<i>purK</i>	1545.5	14.5	458.7	phosphoribosylaminoimidazole carboxylase ATPase subunit	0.0	0.3
SMc02574	<i>hisB</i>	2841.5	54.5	826.3	imidazoleglycerol-phosphate dehydratase	0.0	0.3
SMc01360	<i>pyrB</i>	1095.5	0.5	341.0	aspartate carbamoyltransferase catalytic subunit	0.0	0.3
SMc00412	<i>pyrF</i>	357.5	4	107.7	orotidine 5'-phosphate decarboxylase	0.0	0.3
SMc00494	-	293.5	31.5	60.3	phosphoribosylformylglycinamide synthase subunit PurS	0.1	0.2
SMc02165	<i>pyrE</i>	380.5	3	119.7	orotate phosphoribosyltransferase	0.0	0.3
SMc03978	<i>tkt2</i>	425	18.5	127.3	transketolase	0.0	0.3
SMc01361	-	795.5	1	280.0	dihydroorotase	0.0	0.4
SMc02245	<i>pyrD</i>	544.5	2	191.7	dihydroorotate dehydrogenase 2	0.0	0.4
SMc03858	<i>pheAa</i>	117.5	6	38.3	chorismate mutase	0.1	0.3
SMc02166	<i>pyrC</i>	229	0	87.7	dihydroorotase	0.0	0.4
SMc02760	-	1719.5	349.5	332.7	ATP-dependent nuclease/helicase	0.2	0.2
SMc04005	<i>pykA</i>	794	38.5	308.0	pyruvate kinase	0.0	0.4
SMc03777	<i>proA</i>	2951	511	909.7	gamma-glutamyl phosphate reductase	0.2	0.3
SMc03981	<i>pgk</i>	1803	117	754.7	phosphoglycerate kinase	0.1	0.4
SMc01028	<i>eno</i>	739.5	38.5	347.7	phosphopyruvate hydratase	0.1	0.5
SMc02489	<i>xerC</i>	976	216.5	299.7	site-specific tyrosine recombinase XerC	0.2	0.3
SMc02838	<i>gpmA</i>	854.5	102.5	363.7	phosphoglyceromutase	0.1	0.4
SMc02088	-	65	32	7.3	hypothetical protein	0.5	0.1
SMc03776	<i>proBl</i>	4945	764.5	2256.7	gamma-glutamyl kinase	0.2	0.5
SMc00726	<i>tlpA</i>	974.5	332	319.0	thiol:disulfide interchange redox-active center transmembrane protein	0.3	0.3
SMc02363	<i>cycK</i>	1691.5	644.5	491.7	cytochrome C-type biogenesis transmembrane protein	0.4	0.3
SMc01203	-	439	115	180.7	hypothetical protein	0.3	0.4
SMc00691	<i>xerD</i>	1060.5	285.5	438.0	site-specific tyrosine recombinase XerD	0.3	0.4

SMc02707	-	339.5	64	168.7	Putative acetyltransferase	0.2	0.5
SMc02983	-	751.5	330.5	194.7	ornithine, DAP, or arginine decarboxylase	0.4	0.3
SMc03849	<i>ccmC</i>	510	152	205.0	heme exporter C (cytochrome C-type biogenesis protein) transmembrane	0.3	0.4
SMc03973	-	3366.5	865	1512.0	hypothetical protein	0.3	0.4
SMc02068	-	1135.5	366	446.7	ABC transporter ATP-binding protein	0.3	0.4
SMc03851	<i>ccmG</i>	640	231	233.7	thiol:disulfide interchange protein (cytochrome C biogenesis protein)	0.4	0.4
SMc03966	<i>ruvA</i>	237.5	63.5	111.0	Holliday junction DNA helicase RuvA	0.3	0.5
SMc00084	-	109.5	35	45.7	hypothetical protein	0.3	0.4
SMc00838	-	4232.5	1616	1510.0	hypothetical protein	0.4	0.4
SMc02812	-	823	227	396.3	hypothetical protein	0.3	0.5
SMc01428	<i>cspA2</i>	33	13.5	12.0	cold shock transcription regulator protein	0.4	0.4
SMc00450	<i>ctaB</i>	232.5	104	76.3	protoheme IX farnesyltransferase	0.4	0.3
SM_b20563	-	230.5	102.5	81.0	hypothetical protein	0.4	0.4
SMc01933	-	113.5	44	47.0	hypothetical protein	0.4	0.4
SM_b20258	-	1309.5	495	580.0	transcriptional regulator	0.4	0.4
SMc03848	<i>ccmB</i>	645.5	310	223.3	heme exporter B (cytochrome C-type biogenesis protein) transmembrane	0.5	0.3
SMc00784	-	486.5	168.5	239.3	iron binding protein	0.3	0.5
SMc03847	<i>ccmA</i>	349	165	129.3	cytochrome c biogenesis protein CcmA	0.5	0.4
SMc00188	<i>fbcB</i>	245.5	122.5	86.0	cytochrome B transmembrane protein	0.5	0.4
SM_b20804	-	267.5	123	112.3	hypothetical protein	0.5	0.4
SMc02050	<i>tig</i>	265	112.5	122.7	trigger factor	0.4	0.5
SM_b21261	-	6183.5	2751	2815.0	mureinpeptideoligopeptide ABC transporter substrate-binding protein precursor	0.4	0.5
SMc00010	<i>ctaD</i>	616.5	305.5	250.7	cytochrome C oxidase polypeptide I transmembrane protein	0.5	0.4
SMc00522	<i>rhlE1</i>	1176.5	525.5	565.7	ATP-dependent RNA helicase	0.4	0.5

Table S6 Tn-seq predicted *S. meliloti* genes required for *Medicago truncatula* root only

GeneID	Gene	Rich	Min	<i>M. truncatula</i>	Function	Root Fitness	RSM Fitness
SMa5002	-	42.5	33.0	14.0	hypothetical protein	0.33	0.78
SMc00907	-	84.0	49.0	10.8	hypothetical protein	0.09	0.58
SMc00252	-	41.0	34.0	13.2	signal peptide protein	0.46	0.83
SM_b20608	-	49.5	41.5	21.2	ArsR family transcriptional regulator	0.48	0.84
SMc00647	rluD	175.0	111.0	31.7	ribosomal large subunit pseudouridine synthase	0.23	0.63
SM_b20401	-	32.5	20.5	26.0	hypothetical protein	0.38	0.63
SMc02818	-	177.5	147.0	49.3	hypothetical protein	0.49	0.83
SMc01219	lpsB	222.0	180.5	74.7	lipopolysaccharide core biosynthesis mannosyltransferase	0.28	0.81
SMc02897	-	172.0	133.0	63.8	cytochrome C transmembrane protein	0.38	0.77
SMc02659	relA	506.0	317.5	135.7	GTP pyrophosphokinase (ATP:GTP 3'-pyrophosphotransferase) protein	0.41	0.63
SMc02695	-	251.0	191.0	143.0	GTP-dependent nucleic acid-binding protein EngD	0.32	0.76
SMc01260	-	92.0	47.5	57.2	transcriptional regulator	0.38	0.52
SMc01768	-	122.5	109.5	42.7	transcriptional regulator	0.41	0.89
SMc01931	-	92.5	73.5	42.0	hypothetical protein	0.36	0.79
SMc04003	rpmJ	30.5	22.5	18.8	50S ribosomal protein L36	0.13	0.74
SMc00349	lepA	444.0	379.5	111.2	GTP-binding protein LepA	0.49	0.85
SMc02112	-	63.0	58.5	118.7	hypothetical protein	0.30	0.93
SMc03884	ispA	240.5	337.0	57.5	geranyltranstransferase	0.40	1.40
SMa0848	-	116.5	107.5	75.7	hypothetical protein	0.47	0.92
SMc01147	-	339.0	281.0	90.2	coproporphyrinogen III oxidase	0.37	0.83
SMc02636	-	176.0	186.5	104.7	hypothetical protein	0.48	1.06
SMc02558	-	248.5	207.0	98.0	hypothetical protein	0.45	0.83
SMc01613	rpiB	134.0	123.5	84.8	ribose-5-phosphate isomerase B	0.43	0.92
SMc01042	ntrB	363.5	254.5	110.8	nitrogen regulation protein	0.45	0.70
SMc01207	queA	350.5	278.5	142.5	S-adenosylmethionine--tRNA ribosyltransferase-isomerase	0.35	0.79
SM_b20647	-	86.5	84.0	81.0	hypothetical protein	0.47	0.97
SMc01668	-	209.5	148.5	68.0	transcriptional regulator	0.46	0.71
SMa5036	-	178.0	123.0	91.7	hypothetical protein	0.49	0.69
SMc02768	-	274.0	307.0	100.2	hypothetical protein	0.41	1.12

SMc02724	-	341.0	606.0	137.5	hypothetical protein	0.48	1.78
SMc01368	-	445.5	341.0	189.3	transport transmembrane protein	0.48	0.77
SMc02394	-	120.0	65.0	122.8	hypothetical protein	0.25	0.54
SM_b20518	-	284.0	346.0	59.3	hypothetical protein	0.31	1.22
SMc00731	-	232.0	121.0	90.7	hypothetical protein	0.40	0.52
SMc01242	-	416.5	353.5	139.0	signal peptide protein	0.45	0.85
SMc01041	-	439.0	373.0	194.5	NIFR3-like protein	0.46	0.85
SMc01800	-	473.5	244.5	194.2	cytochrome C oxidase assembly transmembrane protein	0.39	0.52
SMc02435	hemK1	395.0	237.5	142.8	methyltransferase	0.25	0.60
SM_b20947	exoX	134.0	132.0	73.8	posttranscriptional regulator, repressor protein	0.35	0.99
SMc01595	-	698.5	363.0	42.8	sensor histidine kinase transmembrane protein	0.06	0.52
SMc01930	-	215.5	168.5	65.0	hypothetical protein	0.42	0.78
SMc00725	argH1	764.5	401.5	119.7	argininosuccinate lyase	0.19	0.53
SMc01355	-	274.0	146.5	92.2	Holliday junction resolvase-like protein	0.13	0.53
SMc02478	-	636.5	448.5	160.2	hypothetical protein	0.45	0.70
SMA2020	-	261.0	132.0	190.7	Transcriptional regulator	0.37	0.51
SMc00177	-	364.0	326.5	137.0	hypothetical protein	0.49	0.90
SMc01945	-	249.0	157.5	139.8	transcriptional regulator	0.41	0.63
SMc02709	-	128.5	70.5	78.0	hypothetical protein	0.41	0.55
SMc00175	-	523.0	337.0	114.3	ABC transporter ATP-binding protein	0.34	0.64
SMc04175	-	455.5	267.0	197.0	transmembrane protein	0.48	0.59
SMc03140	-	549.0	284.5	227.2	transcriptional regulator	0.43	0.52
SMc00012	ctaG	483.5	275.5	228.2	cytochrome C oxidase assembly protein	0.46	0.57
SM_b20808	-	635.5	333.0	266.5	hypothetical protein	0.49	0.52
SMc01501	mtlK	1245.5	1625.5	378.7	mannitol 2-dehydrogenase	0.36	1.31
SMc01783	-	464.5	241.0	308.8	hypothetical protein	0.37	0.52
SMc01471	senC	506.5	310.5	187.7	cytochrome C oxidase assembly factor transmembrane protein	0.40	0.61
SMc02407	-	541.0	393.0	192.0	hypothetical protein	0.34	0.73
SMc04234	csp4	203.5	127.5	141.3	cold shock-like transcription regulator protein	0.50	0.63
SMc01111	lnt	1551.5	1431.0	261.7	apolipoprotein N-acyltransferase	0.27	0.92
SMc04183	-	790.5	763.0	380.7	transmembrane protein	0.43	0.97
SMc04281	cobC	994.0	665.0	351.5	cobalamin biosynthesis protein pyridoxal-phosphate-dependent aminotransferase	0.37	0.67

SM_b21521	-	469.5	307.0	258.3	hypothetical protein	0.33	0.65
SMA0792	-	857.5	605.0	202.7	hypothetical protein	0.29	0.71
SMc00487	-	253.5	129.5	174.8	hypothetical protein	0.38	0.51
SM_b21433	-	899.5	481.0	251.7	methyl-transferase, S-adenosyl-L-methionine (SAM)-MTase	0.45	0.53
SM_b20062	-	287.0	309.0	245.7	hypothetical protein	0.30	1.08
SMc02172	-	1507.5	1400.5	374.0	transcriptional regulator	0.44	0.93
SMc02347	asfB	414.0	355.5	418.3	ferredoxin ASFB iron-sulfur protein	0.42	0.86
SMc00808	chrA	2105.5	2027.0	398.8	chromate transporter	0.30	0.96
SM_b21268	-	1049.0	917.5	549.3	hypothetical protein	0.45	0.87
SMc02461	-	442.0	225.0	293.2	hypothetical protein	0.25	0.51
SM_b20942	exoB	1497.5	755.0	407.2	UDP glucose 4-epimerase	0.47	0.50
SMc02641	rkpK	2326.0	1244.5	601.7	UDP-glucose 6-dehydrogenase	0.22	0.54
SM_b21222	-	2226.5	1837.0	806.8	transcriptional regulator	0.50	0.83
SM_b21253	-	1468.0	1063.5	876.7	hypothetical protein	0.44	0.72
SM_b21266	-	2281.5	1725.5	773.3	hypothetical protein	0.40	0.76
SM_b21269	-	4377.5	3366.0	1447.7	ABC transporter ATPase	0.45	0.77
SMc00674	hutC	1843.0	1053.0	1259.0	histidine utilization repressor transcription regulator protein	0.29	0.57
SM_b21265	redB	3205.0	2361.0	960.5	glycosyltransferase	0.43	0.74
SM_b21264	redA	3160.5	2543.0	1412.3	hypothetical protein	0.45	0.80
SM_b21256	-	5058.5	3562.0	1970.8	nucleotide sugar oxidase	0.50	0.70

Table S7 Tn-seq predicted *S. meliloti* genes required for growth on RSM only

GeneID	Gene	Rich	Min	<i>M. truncatula</i>	Function	RSM Fitness	Root Fitness
SMc00964	-	90.5	0.5	48.0	hypothetical protein	0.01	0.53
SMA0767	fixQ2	32.5	15.0	30.0	FixQ2 nitrogen fixation protein	0.46	0.92
SMc02561	-	117.5	12.0	137.3	hypothetical protein	0.10	1.17
SM_b20806	-	168.5	83.0	110.0	hypothetical protein	0.49	0.65
SM_b21177	phoC	229.0	79.0	194.0	phosphate uptake ABC transporter ATP-binding protein	0.34	0.85
SMc05014	-	127.5	57.5	121.7	recombinase	0.45	0.95
SMc02720	clpP2	193.5	82.0	199.7	ATP-dependent Clp protease proteolytic subunit	0.42	1.03
SMc04243	znuB	272.5	86.0	198.0	high-affinity zinc uptake system membrane ABC transporter protein	0.32	0.73
SMc00013	ctaE	330.0	146.0	191.0	cytochrome C oxidase subunit III transmembrane protein	0.44	0.58
SMc03069	pgl	303.0	69.0	216.0	6-phosphogluconolactonase	0.23	0.71
SMc03070	zwf	644.0	17.5	1009.0	glucose-6-phosphate 1-dehydrogenase	0.03	1.57
SMA2034	-	110.0	14.5	64.3	hypothetical protein	0.13	0.58
SMc02677	proC	369.0	166.5	295.0	pyrroline-5-carboxylate reductase	0.45	0.80
SMc00092	cysH	402.5	4.0	713.3	phosphoadenosine phosphosulfate reductase	0.01	1.77
SMc01507	-	765.5	218.0	719.0	hypothetical protein	0.28	0.94
SMc00294	-	642.5	161.5	530.0	aminotransferase	0.25	0.82
SMc02077	xthA2	420.5	4.0	232.3	exodeoxyribonuclease III protein	0.01	0.55
SMc04244	znuC	476.5	167.0	449.3	high-affinity zinc uptake system ATP-binding ABC transporter protein	0.35	0.94
SMc00414	-	329.0	106.5	327.7	hypothetical protein	0.32	1.00
SMc02123	-	347.5	6.5	354.0	hypothetical protein	0.02	1.02
SMc00661	-	86.5	42.0	57.3	hypothetical protein	0.49	0.66
SMc01877	recN	1210.0	174.0	838.7	DNA repair protein	0.14	0.69
SMc02635	recO	676.0	149.0	414.3	DNA repair protein RecO	0.22	0.61
SMc00963	-	530.0	15.0	519.3	hypothetical protein	0.03	0.98
SMc01578	aatA	1082.0	394.5	592.3	aspartate aminotransferase	0.36	0.55
SMc00602	uvrC	1841.0	868.0	1337.0	excinuclease ABC subunit C	0.47	0.73
SMc03229	gpsA	985.5	131.0	1149.3	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase	0.13	1.17
SMc02124	-	1840.5	85.0	1409.0	nitrite reductase	0.05	0.77
SMc00962	-	764.5	4.0	638.0	ABC transporter ATP-binding protein	0.01	0.83
SMc03878	phbB	820.0	275.0	586.3	acetoacetyl-CoA reductase	0.34	0.72
SMc00856	-	220.0	101.0	152.7	hypothetical protein	0.46	0.69
SMc02908	recR	724.5	218.5	520.7	recombination protein RecR	0.30	0.72
SMc00296	phbC	2263.0	1126.0	1648.3	POLY3-hydroxybutyrate polymerase	0.50	0.73
SMc01054	-	392.0	99.5	382.0	hypothetical protein	0.25	0.97
SMc00090	cysN	1929.0	98.5	1977.7	sulfate adenyltransferase	0.05	1.03
SMc02863	recF	1512.5	364.0	1388.3	recombination protein F	0.24	0.92
SMc04245	znuA	1414.5	408.5	785.3	high-affinity zinc uptake system ABC transporter protein	0.29	0.56
SMc00556	radA	2031.0	334.5	1111.0	DNA repair protein RadA	0.16	0.55
SMc00876	-	1658.0	168.0	923.0	ATP-binding MRP protein	0.10	0.56

SMc01174	cysK2	1753.0	830.0	974.7	cysteine synthase A	0.47	0.56
SMc03895	pyc	5602.5	15.5	3719.0	pyruvate carboxylase	0.00	0.66
SMc00356	genX	1821.0	896.0	1945.0	lysyl-tRNA synthetase	0.49	1.07
SMc00775	fbpB	2953.0	1248.5	1550.3	iron(III) permease	0.42	0.53
SMc02835	glk	1837.5	293.5	1731.7	glucokinase	0.16	0.94
SMc02138	argD	2167.0	634.5	1113.0	acetylornithine transaminase	0.29	0.51
SMc03924	glgA	3027.5	480.5	3167.0	glycogen synthase	0.16	1.05
SMc00091	cysD	2070.5	199.5	2261.0	sulfate adenylyltransferase	0.10	1.09
SMc03068	edd	3994.5	152.0	4368.3	phosphogluconate dehydratase	0.04	1.09
SMc01053	cysG	3252.0	1251.0	2857.3	siroheme synthase	0.38	0.88
SMc03922	glgB1	5391.5	174.0	3822.3	glycogen branching protein	0.03	0.71
SM_b20811	-	846.0	385.5	494.3	hypothetical protein	0.46	0.58
SMc04125	-	2876.0	1429.5	2041.3	ABC transporter permease	0.50	0.71
SMc02234	-	2390.5	1184.0	1410.7	hypothetical protein	0.50	0.59
SM_b20903	-	4136.5	1732.0	4297.3	sugar uptake ABC transporter permease	0.42	1.04
SMc02884	-	3540.5	1605.0	2499.7	lipoprotein	0.45	0.71
SMc04127	-	9775.0	4151.5	7207.0	ABC transporter ATP-binding protein	0.42	0.74
SMc04126	-	8065.0	3887.0	6475.7	ABC transporter permease	0.48	0.80

Table S8 Tn-seq predicted *S. meliloti* genes more important for RSM than for roots

GeneID	Gene	BRM	RSM	Root	Function	RSM Fitness	Root Fitness	RSM - Root Fitness
SMc01852	pfk	188	1836.5	380.67	pyrophosphate--fructose-6-phosphate 1-phosphotransferase	9.77	2.02	7.74
SM_b20624	mtnA	1488	4212.5	981.67	methylthioribose-1-phosphate isomerase	2.83	0.66	2.17
SM_b20945	exoF1	210	823.5	385.67	hypothetical protein	3.92	1.84	2.08
SM_b20948	exoU	286	1011.5	447.67	glucosyltransferase	3.54	1.57	1.97
SMc01023	tpiA	106	867	662.67	triosephosphate isomerase	8.18	6.25	1.93
SMc00334	cmk	165	385	94.00	cytidylate kinase	2.33	0.57	1.76
SM_b20728	-	43	102.5	33.67	hypothetical protein	2.38	0.78	1.60
SM_b21182	-	166.5	572.5	329.33	hypothetical protein	3.44	1.98	1.46
SM_b21181	-	393	794.5	322.00	glutaryl-CoA dehydrogenase	2.02	0.82	1.20
SMA0125	groES	98	232.5	120.33	co-chaperonin GroES	2.37	1.23	1.14
SMc03923	glgC	599.5	1580	900.33	glucose-1-phosphate adenylyltransferase	2.64	1.50	1.13
SMc04431	-	116.5	201.5	71.00	hypothetical protein	1.73	0.61	1.12
SMc01068	-	73	127.5	47.33	hypothetical protein	1.75	0.65	1.10
SMc00520	-	66	181	110.33	hypothetical protein	2.74	1.67	1.07
SM_b20961	exoP	371	727	334.33	protein tyrosine kinase MPA1 family protein	1.96	0.90	1.06
SMc02716	-	67	134.5	63.67	hypothetical protein	2.01	0.95	1.06
SMc00458	feuP	311.5	722	398.33	transcriptional regulator	2.32	1.28	1.04
SMc02106	-	185.5	489.5	300.67	hypothetical protein	2.64	1.62	1.02
SM_b21564	-	120.5	213	94.00	hypothetical protein	1.77	0.78	0.99
SMc00812	-	262.5	430	171.00	hypothetical protein	1.64	0.65	0.99
SM_b20046	repA1	583.5	942	371.67	replication protein A	1.61	0.64	0.98
SMc00496	-	111.5	230.5	121.67	hypothetical protein	2.07	1.09	0.98
SMc01153	-	322.5	790.5	481.00	enoyl-CoA hydratase	2.45	1.49	0.96
SMc02278	-	24.5	45	21.67	hypothetical protein	1.84	0.88	0.95
SMc03900	ndvA	683	1488	854.33	cyclic beta-1,2-glucan ABC transporter	2.18	1.25	0.93
SMA1239	napD	58	103	51.00	NapD component of periplasmic nitrate reductase	1.78	0.88	0.90

SMA1009	-	67	127.5	67.67	Heavy metal binding protein	1.90	1.01	0.89
SM_b20090	-	59	170.5	119.00	hypothetical protein	2.89	2.02	0.87
SM_b21027	-	151.5	257.5	127.33	hypothetical protein	1.70	0.84	0.86
SMc01039	-	117	193.5	99.33	hypothetical protein	1.65	0.85	0.80
SMA0738	cspA6	82.5	158.5	92.33	CspA6 cold shock protein transcriptional regulator	1.92	1.12	0.80
SMc04114	pilA1	45.5	63	26.67	pilin subunit protein	1.38	0.59	0.80
SMc01099	truA	377.5	642	341.00	tRNA pseudouridine synthase A	1.70	0.90	0.80
SMc00129	feuQ	1611.5	3113	1891.33	sensor histidine kinase	1.93	1.17	0.76
SMc00301	-	95.5	137.5	67.00	hypothetical protein	1.44	0.70	0.74
SMc01335	amiC	708	924.5	407.00	N-acetylmuramoyl-L-alanine amidase AMIC precursor transmembrane protein	1.31	0.57	0.73
SMc01412	-	215	424.5	269.33	hypothetical protein	1.97	1.25	0.72
SM_b21008	-	68.5	135	85.67	ArsR family transcriptional regulator	1.97	1.25	0.72
SMc04269	-	60	153	110.00	hypothetical protein	2.55	1.83	0.72
SMc04385	-	389	564.5	291.67	aldehyde dehydrogenase transmembrane protein	1.45	0.75	0.70
SMc02425	-	159.5	248	137.00	transcriptional regulator	1.55	0.86	0.70
SMc01854	-	166.5	222.5	108.67	amidase (AMPD protein)	1.34	0.65	0.68
SMc03242	typA	637.5	835	400.33	GTP-binding protein	1.31	0.63	0.68
SMc00280	-	69.5	98.5	51.67	hypothetical protein	1.42	0.74	0.67
SM_b20458	-	283.5	473.5	285.00	dTDP-glucose 4,6-dehydratase	1.67	1.01	0.66
SMc00238	moeA	582.5	773	386.33	molybdopterin biosynthesis protein	1.33	0.66	0.66
SM_b20313	-	1117	1404.5	678.33	glycerone kinase	1.26	0.61	0.65
SMc02844	-	456.5	572.5	278.00	transcriptional regulator	1.25	0.61	0.65
SMA5023	-	173	244	134.00	hypothetical protein	1.41	0.77	0.64
SM_b20752	-	665	1163.5	742.67	3-hydroxyisobutyryl-CoA hydrolase	1.75	1.12	0.63
SMA1325	-	160.5	275	175.67	hypothetical protein	1.71	1.09	0.62
SMc01533	adeC1	1118	1972	1284.00	adenine deaminase	1.76	1.15	0.62
SMc04093	acsA1	2927	3589.5	1793.33	acetyl-CoA synthetase	1.23	0.61	0.61
SMA5006	-	70	100.5	57.67	hypothetical protein	1.44	0.82	0.61
SMc04382	ndvB	5398	9592.5	6304.33	beta-(1,2)-glucan production associated transmembrane protein	1.78	1.17	0.61

SMc04285	cobE	653	932	539.33	cobalamin biosynthesis protein	1.43	0.83	0.60
SMc01555	-	103.5	120	58.00	hypothetical protein	1.16	0.56	0.60
SMc00055	phaE2	60.5	71.5	36.67	monovalent cation/H ⁺ antiporter subunit E	1.18	0.61	0.58
SMc01515	-	374	743	528.00	hypothetical protein	1.99	1.41	0.57
SM_b21677	-	121.5	149.5	80.00	signal peptide	1.23	0.66	0.57
SM_b20717	-	2346.5	3195.5	1866.67	LacI family transcriptional regulator	1.36	0.80	0.57
SMA0325	-	220.5	284.5	160.33	hypothetical protein	1.29	0.73	0.56
SMc00622	rnd	613	1027.5	682.67	ribonuclease D protein	1.68	1.11	0.56
SMc01748	-	100.5	129.5	73.00	hypothetical protein	1.29	0.73	0.56
SMc01622	-	2128.5	2305.5	1111.33	oxidoreductase	1.08	0.52	0.56
SMA0723	-	110.5	219	157.33	hypothetical protein	1.98	1.42	0.56
SMc04235	-	267	402.5	253.67	tRNA-dihydrouridine synthase A	1.51	0.95	0.56
SMA1136	-	325.5	542	360.67	hypothetical protein	1.67	1.11	0.56
SMA0314	-	121	140	72.67	hypothetical protein	1.16	0.60	0.56
SMA2119	-	343	472.5	282.00	hypothetical protein	1.38	0.82	0.56
SMc00882	-	230	361	233.33	2-dehydro-3-deoxy-6-phosphogalactonate aldolase	1.57	1.01	0.56
SM_b20533	-	92	173	122.00	hypothetical protein	1.88	1.33	0.55
SMc01178	-	261.5	406.5	262.00	multidrug transmembrane resistance signal peptide protein	1.55	1.00	0.55
SMc00781	iolA	568.5	610	296.67	malonic semialdehyde oxidative decarboxylase	1.07	0.52	0.55
SMc00764	-	124	246.5	178.33	hypothetical protein	1.99	1.44	0.55
SMc03782	-	349	381.5	190.00	signal peptide protein	1.09	0.54	0.55
SMc00144	moaA	1048.5	1526.5	951.33	molybdenum cofactor biosynthesis protein A	1.46	0.91	0.55
SMc03989	-	118.5	221	156.00	hypothetical protein	1.86	1.32	0.55
SMc04221	-	324.5	521	343.67	hypothetical protein	1.61	1.06	0.55
SMc04332	ecnB	138.5	224	148.67	entericidin B signal peptide protein	1.62	1.07	0.54
SMA0450	-	187	309.5	208.00	hypothetical protein	1.66	1.11	0.54
SMc01606	-	578.5	784	471.67	permease	1.36	0.82	0.54
SM_b20684	-	574	928	618.33	hypothetical protein	1.62	1.08	0.54
SMc02072	-	374	469	267.67	hypothetical protein	1.25	0.72	0.54

SMA1585	-	225.5	386	264.67	hypothetical protein	1.71	1.17	0.54
SMc00283	-	210.5	268.5	155.33	transcriptional regulator	1.28	0.74	0.54
SMc00123	-	118.5	254	190.67	hypothetical protein	2.14	1.61	0.53
SMc04352	-	513	581	307.67	signal peptide protein	1.13	0.60	0.53
SMc02877	-	1407.5	1460.5	711.33	hypothetical protein	1.04	0.51	0.53
SM_b21197	oppB	757.5	962	560.00	oligopeptide uptake ABC transporter permease	1.27	0.74	0.53
SMc00699	-	644.5	725.5	384.00	hypothetical protein	1.13	0.60	0.53
SMc02864	thiF	880	1034.5	569.67	molybdopterin biosynthesis protein MoeB	1.18	0.65	0.53
SMc02385	-	2666	2949	1551.67	hypothetical protein	1.11	0.58	0.52
SMc02642	-	86.5	161	115.67	signal peptide protein	1.86	1.34	0.52
SMA0951	-	422.5	714.5	493.33	ABC transporter permease	1.69	1.17	0.52
SMA0535	-	30.5	60.5	44.67	hypothetical protein	1.98	1.46	0.52
SMc00698	-	191	290	192.33	transcriptional regulator	1.52	1.01	0.51
SMc04177	-	243.5	357.5	233.00	hypothetical protein	1.47	0.96	0.51
SMc03935	-	1084.5	1452	901.33	hypothetical protein	1.34	0.83	0.51
SM_b20047	-	92.5	117.5	70.67	hypothetical protein	1.27	0.76	0.51
SM_b20154	-	1387	1801	1100.67	transcriptional regulator	1.30	0.79	0.50
SMc00321	truB	1133.5	1194.5	622.67	tRNA pseudouridine synthase B	1.05	0.55	0.50
SM_b20098	-	161.5	181	99.67	hypothetical protein	1.12	0.62	0.50
SM_b20983	-	181.5	227.5	136.33	hypothetical protein	1.25	0.75	0.50
SM_b20560	-	251	408	282.33	hypothetical protein	1.63	1.12	0.50
SMc02978	-	506	529.5	278.00	hypothetical protein	1.05	0.55	0.50
SMc04335	-	51	65	39.67	lipoprotein transmembrane	1.27	0.78	0.50
SM_b21684	-	160.5	204	125.00	hypothetical protein	1.27	0.78	0.49
SMA5033	-	580.5	812	526.33	hypothetical protein	1.40	0.91	0.49
SMc03845	-	150	157.5	84.00	hypothetical protein	1.05	0.56	0.49
SM_b20627	-	310	500.5	348.67	hypothetical protein	1.61	1.12	0.49
SMc02901	-	44.5	54	32.33	hypothetical protein	1.21	0.73	0.49
SMc03883	mtgA	1020	1079	584.00	monofunctional biosynthetic peptidoglycan transglycosylase	1.06	0.57	0.49

SMA5008	-	82.5	131.5	91.67	hypothetical protein	1.59	1.11	0.48
SMc01000	-	578.5	998.5	721.67	hypothetical protein	1.73	1.25	0.48
SMc00951	-	863	1119.5	708.00	hypothetical protein	1.30	0.82	0.48
SM_b20773	-	675	744.5	423.00	GntR family transcriptional regulator	1.10	0.63	0.48
SM_b21381	-	71.5	75	41.00	hypothetical protein	1.05	0.57	0.48
SMc00166	-	316	350.5	200.33	oxidoreductase NAD protein	1.11	0.63	0.48
SM_b21470	-	155.5	284.5	210.67	hypothetical protein	1.83	1.35	0.47
SMc01850	-	280.5	284	151.00	hypothetical protein	1.01	0.54	0.47
SMc01903	clpP	116	156.5	101.67	ATP-dependent Clp protease proteolytic subunit	1.35	0.88	0.47
SMA1562	-	597	1019.5	739.33	Pilus assembly protein	1.71	1.24	0.47
SMc00268	-	1117.5	1484.5	960.33	oxidoreductase	1.33	0.86	0.47
SMc00762	-	1097.5	1851.5	1337.00	glutamine synthetase	1.69	1.22	0.47
SMc02826	-	259.5	370.5	249.00	transcriptional regulator	1.43	0.96	0.47
SMA1160	-	481.5	763.5	539.67	hypothetical protein	1.59	1.12	0.46
SM_b20408	-	1104.5	1403.5	891.33	hypothetical protein	1.27	0.81	0.46
SMc00169	dme	3728.5	5289	3569.67	malic enzyme	1.42	0.96	0.46
SMA0518	-	44	66.5	46.33	hypothetical protein	1.51	1.05	0.46
SMA1727	-	180	315	232.67	alpha/beta hydrolase	1.75	1.29	0.46
SM_b21665	-	173	234	155.67	hypothetical protein	1.35	0.90	0.45
SMc03169	-	1079	1365	878.00	transcriptional regulator	1.27	0.81	0.45
SMc05020	-	30	47.5	34.00	entericidin A precursor	1.58	1.13	0.45
SMc04311	-	534.5	629	388.67	hypothetical protein	1.18	0.73	0.45
SM_b21426	-	1279	1782.5	1209.67	sugar nucleotide oxidoreductase epimerase	1.39	0.95	0.45
SMc00686	-	81.5	114.5	78.00	hypothetical protein	1.40	0.96	0.45
SMc00025	ppdK	2316.5	3285	2249.67	pyruvate phosphate dikinase	1.42	0.97	0.45
SM_b20392	-	558	556.5	307.67	transcriptional regulator NanR	1.00	0.55	0.45
SMc00740	-	323	412	269.00	hypothetical protein	1.28	0.83	0.44
SM_b21467	-	126.5	189	133.00	hypothetical protein	1.49	1.05	0.44
SM_b20559	-	223.5	393	294.33	hypothetical protein	1.76	1.32	0.44

SMc04213	dgkA	159.5	193	123.00	diacylglycerol kinase	1.21	0.77	0.44
SM_b21518	-	249.5	377.5	268.33	hypothetical protein	1.51	1.08	0.44
SMc02173	-	172	301	226.00	hypothetical protein	1.75	1.31	0.44
SMc03053	fliQ	227	288	189.33	flagellar biosynthesis protein FliQ	1.27	0.83	0.43
SMc01609	ribH	813	1016.5	664.33	riboflavin synthase subunit beta	1.25	0.82	0.43
SM_b20061	-	83	114.5	78.67	hypothetical protein	1.38	0.95	0.43
SM_b20418	ugpE	549.5	655	420.00	glycerol-3-phosphate ABC transporter permease	1.19	0.76	0.43
SM_b20209	-	313	446.5	312.67	hypothetical protein	1.43	1.00	0.43
SMc01916	-	138	166.5	107.67	signal peptide protein	1.21	0.78	0.43
SMc03880	-	297.5	393	266.33	hypothetical protein	1.32	0.90	0.43
SMc02845	-	330	306.5	166.67	antibiotic resistance protein	0.93	0.51	0.42
SMA1887	-	413.5	451	276.33	transcriptional regulator	1.09	0.67	0.42
SM_b20367	-	380	581.5	421.67	transcriptional regulator	1.53	1.11	0.42
SMc03012	cheD	1088.5	1182	724.33	chemoreceptor glutamine deamidase CheD	1.09	0.67	0.42
SMc01206	tgt	213	211	121.67	queuine tRNA-ribosyltransferase	0.99	0.57	0.42
SMA1364	-	579	768	525.33	ABC transporter substrate-binding protein	1.33	0.91	0.42
SMc02862	-	365	514	361.67	Pit accessory protein	1.41	0.99	0.42
SMc01607	-	432	472.5	292.33	permease	1.09	0.68	0.42
SMc01216	-	519.5	740.5	524.00	hypothetical protein	1.43	1.01	0.42
SMc02322	-	2309	2503.5	1543.67	short chain dehydrogenase	1.08	0.67	0.42
SMc02515	-	570.5	680.5	445.00	hypothetical protein	1.19	0.78	0.41
SMc00270	-	730.5	898	596.67	transketolase subunit alpha	1.23	0.82	0.41
SMA2265	-	156	245.5	181.33	hypothetical protein	1.57	1.16	0.41
SM_b20166	-	304.5	370	245.33	hypothetical protein	1.22	0.81	0.41
SMc03853	-	236	369.5	273.00	intracellular septation protein A	1.57	1.16	0.41
SM_b20461	-	544.5	865	643.33	hypothetical protein	1.59	1.18	0.41
SMA2309	-	714	981	690.33	ABC transporter permease	1.37	0.97	0.41
SM_b20286	-	816.5	1043.5	711.67	hypothetical protein	1.28	0.87	0.41
SMc01179	-	142.5	187.5	129.67	multidrug transmembrane resistance signal peptide protein	1.32	0.91	0.41

SMc01222	lpsC	200.5	633.5	552.33	lipopolysaccharide core biosynthesis glycosyl transferase	3.16	2.75	0.40
SMc00644	-	359	400.5	255.33	hypothetical protein	1.12	0.71	0.40
SMc00455	-	420.5	635	465.33	hypothetical protein	1.51	1.11	0.40
SM_b20561	-	118	115.5	68.00	hypothetical protein	0.98	0.58	0.40
SMc01969	-	811.5	1345	1019.00	hypothetical protein	1.66	1.26	0.40
SMc01489	-	388	526	370.67	signal peptide protein	1.36	0.96	0.40
SM_b20483	-	789.5	929.5	614.33	catabolite repressor protein	1.18	0.78	0.40
SM_b20197	cbbS	177.5	276.5	206.00	ribulose-1,5-bisphosphate carboxylase small subunit protein	1.56	1.16	0.40
SMA0815	nifA	412	575.5	412.00	Fis family transcriptional regulator	1.40	1.00	0.40
SM_b20880	rhIE2	269.5	611.5	505.00	ATP-dependent RNA helicase	2.27	1.87	0.40
SMc03034	flgH	1037.5	1478	1069.00	flagellar basal body L-ring protein	1.42	1.03	0.39
SMA0166	-	107.5	150.5	108.33	hypothetical protein	1.40	1.01	0.39
SMA0841	-	228.5	275.5	186.00	transposase, fragment	1.21	0.81	0.39
SMc00716	-	576.5	576	350.67	hypothetical protein	1.00	0.61	0.39
SMA1162	-	648	792.5	540.00	hypothetical protein	1.22	0.83	0.39
SMc02824	-	2361	2538	1618.67	hypothetical protein	1.07	0.69	0.39
SMA0640	-	54	70	49.00	hypothetical protein	1.30	0.91	0.39
SM_b20551	-	189.5	349	275.33	hypothetical protein	1.84	1.45	0.39
SM_b20591	-	121.5	138.5	91.33	hypothetical protein	1.14	0.75	0.39
SM_b20290	-	662	843	587.00	transcriptional regulator	1.27	0.89	0.39
SMA1256	-	175.5	202.5	134.67	hypothetical protein	1.15	0.77	0.39
SMA0473	-	378	499	353.33	hypothetical protein	1.32	0.93	0.39
SM_b20104	-	535	805	599.00	hypothetical protein	1.50	1.12	0.39
SM_b21597	-	387.5	559.5	410.33	exported oxidoreductase	1.44	1.06	0.38
SMc00998	-	704	687.5	417.33	signal peptide protein	0.98	0.59	0.38
SM_b20226	-	398	569	416.33	hypothetical protein	1.43	1.05	0.38
SM_b20105	-	267.5	289.5	187.33	transcriptional regulator	1.08	0.70	0.38
SMc01851	-	887.5	1171.5	833.00	amino acid efflux protein	1.32	0.94	0.38
SM_b20946	exoY	210	276.5	196.67	galactosyltransferase	1.32	0.94	0.38

SMa0922	-	96.5	144.5	108.00	hypothetical protein	1.50	1.12	0.38
SMa0079	-	275	357	253.00	ABC transporter permease	1.30	0.92	0.38
SMa1033	-	120	223	177.67	hypothetical protein	1.86	1.48	0.38
SM_b20618	thiE	1420.5	1470	934.00	thiamine-phosphate pyrophosphorylase	1.03	0.66	0.38
SM_b21001	-	470	568	391.00	hypothetical protein	1.21	0.83	0.38
SMa2141	purU3	410.5	548	393.67	formyltetrahydrofolate deformylase	1.33	0.96	0.38
SMa1229	fixL	872	1228	901.00	FixL oxygen regulated histidine kinase	1.41	1.03	0.38
SMc02554	-	915.5	974.5	631.67	hypothetical protein	1.06	0.69	0.37
SMc02552	-	128	165.5	117.67	hypothetical protein	1.29	0.92	0.37
SMa0758	-	213.5	359	279.67	hypothetical protein	1.68	1.31	0.37
SM_b21278	adeC2	2335.5	3247	2380.33	adenine deaminase	1.39	1.02	0.37
SMa0405	-	324	370	250.00	transposase, fragment	1.14	0.77	0.37
SM_b21485	-	167	180.5	118.67	hypothetical protein	1.08	0.71	0.37
SMc03786	bfr	127	141	94.00	bacterioferritin	1.11	0.74	0.37
SMc01429	-	413.5	575	422.33	oxidoreductase	1.39	1.02	0.37
SM_b20882	-	499	566	382.00	hypothetical protein	1.13	0.77	0.37
SMc00863	moaB	507.5	683.5	496.67	molybdenum cofactor biosynthesis protein B	1.35	0.98	0.37
SM_b20757	bhbA	2696	2509	1522.33	methylmalonyl-CoA mutase	0.93	0.56	0.37
SM_b20701	-	2044.5	2358.5	1610.67	hypothetical protein	1.15	0.79	0.37
SMc00904	-	1032	1505.5	1128.33	hypothetical protein	1.46	1.09	0.37
SM_b21596	-	325	385.5	267.00	hypothetical protein	1.19	0.82	0.36
SMa0814	nifB	586	956	742.67	NifB FeMo cofactor biosynthesis protein	1.63	1.27	0.36
SM_b20818	mocD	760	1046	769.33	hydrocarbon oxygenase	1.38	1.01	0.36
SMa1158	-	339	373	249.67	hypothetical protein	1.10	0.74	0.36
SM_b20087	-	1235.5	1418.5	969.33	hypothetical protein	1.15	0.78	0.36
SM_b20772	pdxA	802.5	914	622.67	4-hydroxythreonine-4-phosphate dehydrogenase	1.14	0.78	0.36
SM_b20746	-	256.5	323.5	230.67	hypothetical protein	1.26	0.90	0.36
SMa2065	-	78.5	88	59.67	hypothetical protein	1.12	0.76	0.36
SM_b20615	thiC	3517	3993	2728.33	thiamine biosynthesis protein ThiC	1.14	0.78	0.36

SMc02788	secB	195	215	145.00	preprotein translocase subunit SecB	1.10	0.74	0.36
SMa1214	fixQ1	40	65	50.67	FixQ1 nitrogen fixation protein	1.63	1.27	0.36
SMa2008	-	482	605	432.33	transcriptional regulator	1.26	0.90	0.36
SMa0249	-	162	209	151.00	TRAP-type small permease component	1.29	0.93	0.36
SMa2115	gst13	676	953.5	711.67	Gst13 glutathione S-transferase	1.41	1.05	0.36
SM_b21138	-	182	254	189.00	amino acid ABC transporter ATP-binding protein	1.40	1.04	0.36
SMc04199	-	315.5	530	417.33	hypothetical protein	1.68	1.32	0.36
SMa1332	-	330	463.5	345.67	hypothetical protein	1.40	1.05	0.36
SMc01749	-	92.5	117	84.00	replicative DNA helicase	1.26	0.91	0.36
SM_b20172	-	289.5	295.5	192.33	cytochrome c protein	1.02	0.66	0.36
SMc02848	-	95.5	117	83.00	hypothetical protein	1.23	0.87	0.36
SM_b20920	-	335.5	449	329.67	hypothetical protein	1.34	0.98	0.36
SMc01012	-	246.5	273	185.67	hypothetical protein	1.11	0.75	0.35
SMc00041	-	122	174.5	131.33	hypothetical protein	1.43	1.08	0.35
SMc04314	-	97	143.5	109.33	hypothetical protein	1.48	1.13	0.35
SMc04236	-	1879.5	2252	1590.00	glycine-rich cell wall structural transmembrane protein	1.20	0.85	0.35
SMc01140	-	260.5	255	163.33	SIGMA54 modulation protein	0.98	0.63	0.35
SMc00257	ropB2	189.5	187	120.33	outer-membrane protein	0.99	0.64	0.35
SM_b21576	-	151	225	172.00	ArsR family transcriptional regulator	1.49	1.14	0.35
SM_b21312	wgeC	1155.5	1489.5	1084.67	methyltransferase	1.29	0.94	0.35
SMa1170	cycB2	160	287	231.00	hypothetical protein	1.79	1.44	0.35
SM_b20943	exoZ	627	891.5	673.00	acetyltransferase	1.42	1.07	0.35
SMc02861	pit	984	999.5	657.00	phosphate transporter	1.02	0.67	0.35
SMc02787	-	356	593.5	470.00	hypothetical protein	1.67	1.32	0.35
SMc04023	exoN2	914.5	912	596.33	UTP--glucose-1-phosphate uridylyltransferase	1.00	0.65	0.35
SMc00718	-	413.5	625	482.67	hypothetical protein	1.51	1.17	0.34
SMa1043	-	93	91	59.00	hypothetical protein	0.98	0.63	0.34
SM_b21088	-	394.5	437	301.33	hypothetical protein	1.11	0.76	0.34
SM_b20486	-	1109	1399	1018.33	sugar ABC transporter permease	1.26	0.92	0.34

SM_b20955	exoK	320.5	419.5	309.67	endo-beta-1,3-1,4-glycanase	1.31	0.97	0.34
SMc02055	crtB	231.5	209	129.67	phytoene synthase	0.90	0.56	0.34
SMA0567	-	196	345.5	278.33	hypothetical protein	1.76	1.42	0.34
SM_b21539	-	323	356	245.33	aldehyde dehydrogenase	1.10	0.76	0.34
SMA2199	-	198.5	278	210.00	ABC transporter substrate-binding protein	1.40	1.06	0.34
SMA0869	nodA	195.5	219.5	152.67	acyltransferase	1.12	0.78	0.34
SM_b21528	tauC	751	894.5	638.00	taurine uptake ABC transporter permease	1.19	0.85	0.34
SMc00717	-	306.5	292	188.00	ABC transporter ATP-binding protein	0.95	0.61	0.34
SMA0810	fixU	141.5	202	154.00	nitrogen fixation protein FixU	1.43	1.09	0.34
SMA1208	fixS1	240	296	214.67	FixS1 nitrogen fixation protein	1.23	0.89	0.34
SMc01006	-	381	422	293.00	hypothetical protein	1.11	0.77	0.34
SMc01823	-	613	799.5	592.00	ABC transporter ATP-binding protein	1.30	0.97	0.34
SMA1297	-	274	356	264.00	hypothetical protein	1.30	0.96	0.34
SMc01315	-	279.5	353.5	259.67	hypothetical protein	1.26	0.93	0.34
SM_b21092	-	142	203	155.33	acetyltransferase	1.43	1.09	0.34
SMc00861	-	1023	982	638.67	signal peptide protein	0.96	0.62	0.34
SM_b21365	-	333.5	413.5	301.67	hypothetical protein	1.24	0.90	0.34
SM_b20314	-	1043.5	1293	943.67	glycerone kinase	1.24	0.90	0.33
SMc00859	-	126	132	90.00	hypothetical protein	1.05	0.71	0.33
SM_b20050	-	569.5	774	584.33	hypothetical protein	1.36	1.03	0.33
SM_b21059	-	980	1361	1035.00	glucose-1-phosphate cytidyltransferase	1.39	1.06	0.33
SMc00139	-	238	322	243.00	amino acid ABC transporter permease	1.35	1.02	0.33
SMc00945	gpt	392	573	443.00	xanthine-guanine phosphoribosyltransferase	1.46	1.13	0.33
SMA1138	-	340.5	535.5	422.67	hypothetical protein	1.57	1.24	0.33
SMA2267	-	215.5	258	186.67	hypothetical protein	1.20	0.87	0.33
SMA0303	-	545.5	707	526.67	LysR family transcriptional regulator	1.30	0.97	0.33
SMc00369	-	52	45.5	28.33	oxidoreductase	0.88	0.54	0.33
SMA2369	-	1527.5	2125	1621.00	ABC transporter permease	1.39	1.06	0.33
SMA2027	-	378.5	488.5	364.33	LysR family transcriptional regulator	1.29	0.96	0.33

SMc02516	-	1250	1377	967.33	ABC transporter permease	1.10	0.77	0.33
SMc03007	cheA	3794.5	3509	2268.33	chemotaxis protein (sensory transduction histidine kinase)	0.92	0.60	0.33
SMA0838	syrA	90	122	92.67	SyrA protein involved in EPS production	1.36	1.03	0.33
SMc04442	-	232.5	356	280.33	acetyltransferase	1.53	1.21	0.33
SMc01948	livF	1563	1866.5	1359.67	high-affinity branched-chain amino acid ABC transporter ATP-binding protein	1.19	0.87	0.32
SMA0523	-	462	504.5	355.00	ABC transporter substrate-binding protein	1.09	0.77	0.32
SMA0769	fixP2	531.5	741.5	569.67	FixP2 Diheme c-type cytochrome	1.40	1.07	0.32
SMc04338	-	908	1017	723.67	hypothetical protein	1.12	0.80	0.32
SMc01177	-	98	118.5	87.00	hypothetical protein	1.21	0.89	0.32
SMc02758	-	1342	1941	1510.00	nucleotidyl transferase	1.45	1.13	0.32
SMc02723	-	136	169	125.33	7-cyano-7-deazaguanine reductase	1.24	0.92	0.32
SMc02049	gcvP	1277.5	1142.5	732.67	glycine dehydrogenase	0.89	0.57	0.32
SMA1186	nosL	157.5	233.5	183.00	NosL copper chaperone	1.48	1.16	0.32
SMc00171	-	626.5	805	604.67	hypothetical protein	1.28	0.97	0.32
SMA0190	-	182.5	263	204.67	transcriptional regulator	1.44	1.12	0.32
SM_b20527	-	268	324.5	239.00	hypothetical protein	1.21	0.89	0.32
SMA2059	-	103.5	131	98.00	hypothetical protein	1.27	0.95	0.32
SMc04325	-	436	509	370.00	methionine synthase I	1.17	0.85	0.32
SMA0955	atrA	617	639	442.33	GntR family transcriptional regulator	1.04	0.72	0.32
SMA0400	-	2983.5	3271.5	2323.33	Dehydrogenase, Zn-dependent	1.10	0.78	0.32
SMc00599	moaE	545.5	789.5	616.33	molybdopterin MPT converting factor subunit 2	1.45	1.13	0.32
SM_b20323	-	906.5	984	696.33	transcriptional regulator	1.09	0.77	0.32
SM_b20118	-	488.5	723	568.33	hypothetical protein	1.48	1.16	0.32
SMc02155	-	1536	2145	1659.33	hypothetical protein	1.40	1.08	0.32
SM_b21225	-	757	990	750.67	inositol monophosphatase	1.31	0.99	0.32
SMc03005	-	733.5	1116	886.00	hypothetical protein	1.52	1.21	0.31
SMA1570	pilA2	344.5	350	242.00	PilA2 pilus assembly protein	1.02	0.70	0.31
SMA1538	-	1572.5	1835	1342.67	monovalent cation/H+ antiporter subunit D	1.17	0.85	0.31
SMA5015	-	445	647.5	508.33	hypothetical protein	1.46	1.14	0.31

SM_b21676	-	105	160.5	127.67	hypothetical protein	1.53	1.22	0.31
SMc01792	-	1378	1617.5	1187.00	sugar transferase	1.17	0.86	0.31
SMc01858	mraW	321.5	620	519.67	S-adenosyl-methyltransferase MraW	1.93	1.62	0.31
SM_b20287	-	791	892.5	645.67	hypothetical protein	1.13	0.82	0.31
SMc04260	-	490.5	771.5	619.00	transcriptional regulator	1.57	1.26	0.31
SMc00325	-	1565.5	1523.5	1037.00	hypothetical protein	0.97	0.66	0.31
SMc01700	ppiA	860	885	618.00	peptidyl-prolyl cis-trans isomerase A	1.03	0.72	0.31
SMA1334	-	2121	2506	1847.67	hypothetical protein	1.18	0.87	0.31
SMA1210	fixH	330	396	293.67	nitrogen fixation protein FixH	1.20	0.89	0.31
SMc03974	-	665	680.5	474.33	hypothetical protein	1.02	0.71	0.31
SMA1822	-	735.5	1007	779.00	hypothetical protein	1.37	1.06	0.31
SMA1255	-	147.5	217	171.33	hypothetical protein	1.47	1.16	0.31
SM_b21196	oppA	698.5	1177	961.33	oligopeptide uptake ABC transporter substrate-binding protein precursor	1.69	1.38	0.31
SMA1515	-	106.5	157.5	124.67	hypothetical protein	1.48	1.17	0.31
SMA0763	-	193	262	203.00	hypothetical protein	1.36	1.05	0.31
SMA1339	-	744	1007	779.67	ABC transporter permease	1.35	1.05	0.31
SM_b20498	-	799.5	1028.5	785.00	hypothetical protein	1.29	0.98	0.30
SMA0520	-	231.5	319.5	249.00	RpiR family transcriptional regulator	1.38	1.08	0.30
SMc03137	-	155	321.5	274.33	hypothetical protein	2.07	1.77	0.30
SMc01846	-	188.5	267	209.67	hydrolase glycosidase	1.42	1.11	0.30
SM_b20148	-	410.5	502.5	377.67	transcriptional regulator	1.22	0.92	0.30
SM_b20419	ugpC	2749.5	3328.5	2495.33	glycerol-3-phosphate ABC transporter ATP-binding protein	1.21	0.91	0.30
SM_b20149	-	2132.5	2318	1673.67	hypothetical protein	1.09	0.78	0.30
SMA1086	-	376.5	478	364.33	hypothetical protein	1.27	0.97	0.30
SM_b21035	-	47	55.5	41.33	hypothetical protein	1.18	0.88	0.30
SM_b20063	-	187	300	243.67	hypothetical protein	1.60	1.30	0.30
SMc03996	thiE2	286	318.5	232.67	thiamine-phosphate pyrophosphorylase	1.11	0.81	0.30

Table S9 Tn-seq predicted *S. meliloti* genes required for *Medicago truncatula* roots but not Rice roots

Gene	Gene	Average BRM	Average Rice	Average A17	Function	Medicago Fitness	Rice Fitness
SMc01595	-	704.0	354.7	39.7	sensorhistidinekinasetransmembraneprotein	0.06	0.50
SMc04003	rpmJ	31.0	16.3	4.0	50SribosomalproteinL36	0.13	0.53
SMc01111	lnt	1558.5	1284.7	425.3	apolipoproteinN-acyltransferase	0.27	0.82
SMa0792	-	857.5	489.7	255.0	hypotheticalprotein	0.30	0.57
SM_b20518	-	286.0	171.0	89.0	hypotheticalprotein	0.31	0.60
SM_b21521	-	471.0	246.3	153.3	hypotheticalprotein	0.33	0.52
SM_b20563	-	250.0	162.0	83.0	hypotheticalprotein	0.33	0.65
SMa5002	-	42.5	27.0	14.3	hypotheticalprotein	0.34	0.64
SMc00188	fbcb	245.5	157.3	86.7	cytochromeBtransmembraneprotein	0.35	0.64
SM_b20947	exoX	134.0	78.3	47.3	posttranscriptionalregulator,repressorprotein	0.35	0.58
SMc03847	ccmA	356.5	186.0	129.0	cytochromecbiogenesisproteinCcmA	0.36	0.52
SMc01931	-	92.5	73.0	34.0	hypotheticalprotein	0.37	0.79
SMc04281	cobC	989.5	869.7	366.3	cobalaminbiosynthesisproteinpyridoxal-phosphate-dependentaminotransferase	0.37	0.88
SMc00487	-	256.0	150.3	99.7	hypotheticalprotein	0.39	0.59
SMc01260	-	92.0	48.0	36.3	transcriptionalregulator	0.39	0.52
SMc02068	-	1134.5	573.0	450.7	ABCtransporterATP-bindingprotein	0.40	0.51
SMc01800	-	472.0	274.3	187.7	cytochromeCoxidaseassemblytransmembraneprotein	0.40	0.58
SM_b21266	-	2280.5	1312.7	909.3	hypotheticalprotein	0.40	0.58
SMc01768	-	130.5	118.7	52.7	transcriptionalregulator	0.40	0.91
SMc02659	relA	508.0	334.7	206.3	GTPpyrophosphokinase(ATP:GTP3'-pyrophosphotransferase)protein	0.41	0.66
SMc02709	-	131.0	81.3	53.3	hypotheticalprotein	0.41	0.62
SMc01945	-	250.5	135.0	103.7	transcriptionalregulator	0.41	0.54
SMc00084	-	113.5	87.7	47.0	hypotheticalprotein	0.41	0.77
SMc02347	asfB	415.5	261.3	174.3	ferredoxinASFBiron-sulfurprotein	0.42	0.63
SM_b20804	-	267.5	161.0	114.0	hypotheticalprotein	0.43	0.60
SMc04183	-	791.5	610.0	339.0	transmembraneprotein	0.43	0.77
SMc00252	-	42.0	30.7	18.0	signalpeptideprotein	0.43	0.73
SM_b21265	redB	3214.5	1832.0	1400.3	glycosyltransferase	0.44	0.57
SMc01613	rpiB	133.0	116.0	58.0	ribose-5-phosphateisomeraseB	0.44	0.87
SM_b21253	-	1466.0	1124.3	640.0	hypotheticalprotein	0.44	0.77
SMc02172	-	1510.5	851.7	663.7	transcriptionalregulator	0.44	0.56
SM_b20258	-	1312.0	673.0	581.3	transcriptionalregulator	0.44	0.51
SMc01668	-	212.5	148.3	95.7	transcriptionalregulator	0.45	0.70
SMc03973	-	3370.0	1897.0	1518.3	hypotheticalprotein	0.45	0.56
SMc01242	-	416.5	250.3	187.7	signalpeptideprotein	0.45	0.60

SMc01042	ntrB	369.5	289.0	167.0	nitrogenregulationprotein	0.45	0.78
SM_b21268	-	1052.0	688.0	476.0	hypotheticalprotein	0.45	0.65
SM_b21264	redA	3160.5	1733.7	1434.7	hypotheticalprotein	0.45	0.55
SM_b21269	-	4383.0	3031.0	1997.3	ABCtransporterATPase	0.46	0.69
SM_b21261	-	6187.5	3653.3	2825.3	mureinpeptideoligopeptideABCtransportersubstrate-bindingproteinprecursor	0.46	0.59
SMA0848	-	121.5	88.7	55.7	hypotheticalprotein	0.46	0.73
SMc03776	proB1	4940.0	2571.3	2263.3	gamma-glutamylkinase	0.46	0.52
SMc02050	tig	272.5	187.3	125.0	triggerfactor	0.46	0.69
SMc01041	-	450.0	373.0	206.7	NIFR3-likeprotein	0.46	0.83
SMc00012	ctaG	483.0	267.0	223.7	cytochromeCoxidaseassemblyprotein	0.46	0.55
SM_b20942	exoB	1508.0	756.0	704.3	UDPglucose4-epimerase	0.47	0.50
SM_b20647	-	86.0	72.0	40.7	hypotheticalprotein	0.47	0.84
SMc02636	-	177.0	125.0	84.7	hypotheticalprotein	0.48	0.71
SMc04175	-	456.0	234.0	218.3	transmembraneprotein	0.48	0.51
SMc02724	-	342.0	465.7	165.0	hypotheticalprotein	0.48	1.36
SMc00784	-	497.5	284.7	240.7	ironbindingprotein	0.48	0.57
SM_b20608	-	49.5	30.3	24.0	ArsRfamilytranscriptionalregulator	0.48	0.61
SMc00177	-	364.0	280.3	177.7	hypotheticalprotein	0.49	0.77
SMc01368	-	453.0	374.0	222.3	transporttransmembraneprotein	0.49	0.83
SMc02818	-	181.5	127.0	89.3	hypotheticalprotein	0.49	0.70
SMc00349	lepA	451.5	244.7	223.0	GTP-bindingproteinLepA	0.49	0.54
SM_b21256	-	5048.5	3275.0	2511.7	nucleotidesugaroxidase	0.50	0.65
SMA5036	-	178.5	136.0	89.0	hypotheticalprotein	0.50	0.76

Table S10 Tn-seq predicted *S. meliloti* genes required for Rice roots but not *M. truncatula*

Gene	Name	Average BRM	Average A17	Average Rice	Function	Rice Fitness	A17 Fitness
SMc03922	glgB1	5403.0	3842.0	194.7	glycogenbranchingprotein	0.04	0.71
SMc00964	-	92.0	49.3	6.3	hypotheticalprotein	0.07	0.54
SMc03070	zwf	640.0	1018.7	78.0	glucose-6-phosphate1-dehydrogenase	0.12	1.59
SMc00962	-	767.0	639.7	123.0	ABCtransporterATP-bindingprotein	0.16	0.83
SMc03069	pgl	303.0	220.0	50.0	6-phosphogluconolactonase	0.17	0.73
SMc01578	aatA	1084.5	596.0	198.0	aspartateaminotransferase	0.18	0.55
SMc03895	pyc	5591.5	3731.3	1164.7	pyruvatecarboxylase	0.21	0.67
SMc00963	-	530.5	520.3	119.3	hypotheticalprotein	0.22	0.98
SMc00511	rpe	377.0	200.7	87.3	D-ribulose-5-phosphate3-epimerase	0.23	0.53
SMc01140	-	261.0	164.3	67.3	SIGMA54modulationprotein	0.26	0.63
SMc03068	edd	3991.0	4376.3	1033.0	phosphogluconatedehydratase	0.26	1.10
SMc01206	tgt	216.0	121.7	61.0	queuinetRNA-ribosyltransferase	0.28	0.56
SMc02077	xthA2	420.5	235.0	123.0	exodeoxyribonucleaseIIIprotein	0.29	0.56
SMc02164	frk	741.5	386.3	217.0	fructokinase	0.29	0.52
SMc01120	-	587.0	362.0	172.3	hypotheticalprotein	0.29	0.62
SMc02138	argD	2166.0	1119.0	663.7	acetylmithinetransaminase	0.31	0.52
SMc00724	-	24.0	15.3	7.7	hypotheticalprotein	0.32	0.64
SM_b21177	phoC	232.5	195.0	78.0	phosphateuptakeABCtransporterATP-bindingprotein	0.34	0.84
SMc02124	-	1845.0	1417.7	634.7	nitritereductase	0.34	0.77
SMc02322	-	2314.0	1547.7	818.0	shortchaindehydrogenase	0.35	0.67
SMc01124	glnD	1702.0	865.3	609.3	PIIuridylyl-transferase	0.36	0.51
SMc00294	-	645.0	533.3	231.7	aminotransferase	0.36	0.83
SMc02844	-	460.5	277.7	170.3	transcriptionalregulator	0.37	0.60
SMc02861	pit	986.5	659.0	366.0	phosphatetransporter	0.37	0.67
SMc02835	glk	1845.0	1733.7	691.3	glucokinase	0.37	0.94
SM_b20806	-	169.5	110.0	68.3	hypotheticalprotein	0.40	0.65
SMc00876	-	1663.0	927.0	682.7	ATP-bindingMRPprotein	0.41	0.56
SMc01507	-	768.0	724.7	323.3	hypotheticalprotein	0.42	0.94

SMc02049	gcvP	1295.0	737.7	563.7	glycinedehydrogenase	0.44	0.57
SMc01557	-	265.0	135.0	116.0	signalpeptideprotein	0.44	0.51
SMc00321	truB	1139.0	624.3	506.3	tRNApseudouridinesynthaseB	0.44	0.55
SMc03100	-	477.0	312.0	216.0	hypotheticalprotein	0.45	0.65
SM_b21222	-	2227.0	1115.0	1016.3	transcriptionalregulator	0.46	0.50
SMc02234	-	2388.5	1413.3	1100.0	hypotheticalprotein	0.46	0.59
SMc01117	rimI	725.0	366.0	334.0	acetyltransferase	0.46	0.50
SM_b21411	-	64.0	44.0	29.7	hypotheticalprotein	0.46	0.69
SMc03229	gpsA	990.5	1156.7	459.7	NAD(P)H-dependentglycerol-3-phosphatedehydrogenase	0.46	1.17
SMc02123	-	351.5	357.0	166.7	hypotheticalprotein	0.47	1.02
SMc01139	rpoN	1764.0	1959.7	838.7	RNApolymerasefactorsigma-54	0.48	1.11
SMc00091	cysD	2070.0	2268.3	993.7	sulfateadenyltransferase	0.48	1.10
SMc02845	-	329.5	168.7	158.3	antibioticresistanceprotein	0.48	0.51
SMc03883	mtgA	1028.0	587.3	506.7	monofunctionalbiosyntheticpeptidoglycantransglycosylase	0.49	0.57
SMc03924	glgA	3030.0	3175.7	1499.3	glycogensynthase	0.49	1.05
SMc01859	-	66.0	54.7	32.7	hypotheticalprotein	0.49	0.83
SMc02707	-	339.0	169.7	168.0	hypotheticalprotein	0.50	0.50
SMc02047	gcvT	1108.0	614.7	553.0	glycinecleavagesystemaminomethyltransferaseT	0.50	0.55
SMc02561	-	117.5	138.7	58.7	hypotheticalprotein	0.50	1.18
SMc02362	cycJ	181.5	99.0	90.7	cytochromec-typebiogenesisproteinCcmE	0.50	0.55

Table S11 Tn-seq predicted *S. meliloti* genes required for both Rice and *M. truncatula* roots

Gene	Gene	Average BRM	Average Rice	Average A17	Function	Rice Fitness	A17 Fitness
SM_b20062	-	286.0	95.7	85.7	hypotheticalprotein	0.33	0.30
SM_b20401	-	32.0	11.3	13.0	hypotheticalprotein	0.35	0.41
SM_b21433	-	904.5	424.7	406.3	methyl-transferase,S-adenosyl-L-methionine(SAM)-MTase	0.47	0.45
SMa2020	-	261.5	113.3	97.0	Transcriptionalregulator	0.43	0.37
SMc02791	aroE	210.5	2.0	0.0	shikimate5-dehydrogenase	0.01	0.00
SMc02768	-	279.5	110.3	113.3	hypotheticalprotein	0.39	0.41
SMc02767	trpF	714.5	45.3	121.7	N-(5'-phosphoribosyl)anthranilateisomerase	0.06	0.17
SMc02766	trpB	4312.0	364.0	1128.3	tryptophansynthasubunitbeta	0.08	0.26
SMc02765	trpA	2197.5	228.0	632.0	tryptophansynthasubunitalpha	0.10	0.29
SMc02760	-	1734.5	310.7	339.0	ATP-dependentnuclease/helicase	0.18	0.20
SMc02755	ahcY	791.0	51.0	113.3	S-adenosyl-L-homocysteinehydrolase	0.06	0.14
SMc02568	hisE	469.5	2.7	5.3	phosphoribosyl-ATPpyrophosphatase	0.01	0.01
SMc02569	hisF	1143.0	27.0	7.3	imidazoleglycerolphosphatesynthasubunitHisF	0.02	0.01
SMc02570	hisA	1977.0	15.7	0.7	1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamideisomerase	0.01	0.00
SMc02572	hisH	955.5	7.7	2.0	imidazoleglycerolphosphatesynthasubunitHisH	0.01	0.00
SMc02574	hisB	2842.5	107.0	826.7	imidazoleglycerol-phosphatedehydratase	0.04	0.29
SMc02838	gpmA	860.5	141.0	367.3	phosphoglyceromutase	0.16	0.43
SMc02850	polA	381.5	12.3	15.3	DNApolymeraseI	0.03	0.04
SMc02897	-	174.5	45.3	65.0	cytochromeCtransmembraneprotein	0.26	0.37
SMc02899	pheA	2226.0	197.3	427.0	prephenatedehydratase	0.09	0.19
SMc00412	pyrF	359.0	38.7	110.0	orotidine5'-phosphatedecarboxylase	0.11	0.31
SMc01147	-	341.5	165.0	127.0	coproporphyrinogenIIIoxidase	0.48	0.37
SMc01726	argB	1066.0	11.3	15.0	acetylglutamatekinase	0.01	0.01
SMc02166	pyrC	228.5	41.7	89.0	dihydroorotase	0.18	0.39
SMc02165	pyrE	380.5	21.7	122.3	orotatephosphoribosyltransferase	0.06	0.32
SMc02137	argF1	1154.0	5.0	13.0	ornithinecarbamoyltransferase	0.00	0.01
SMc02217	metZ	1252.5	29.0	253.7	O-succinylhomoserinesulfhydylase	0.02	0.20
SMc02245	pyrD	548.0	53.0	192.3	dihydroorotatedehydrogenase2	0.10	0.35
SMc02307	hisD	3441.5	38.3	16.0	histidinoldehydrogenase	0.01	0.00
SMc00838	-	4241.5	1977.7	1512.7	hypotheticalprotein	0.47	0.36
SMc00808	chrA	2108.5	974.7	627.0	chromatetransporter	0.46	0.30
SMc00815	guaB	1411.5	47.7	267.7	inosine5'-monophosphatedehydrogenase	0.03	0.19
SMc00918	hisZ	1002.0	4.0	5.0	ATPphosphoribosyltransferaseregulatorysubunit	0.00	0.00
SMc00917	hisG	1326.0	1.7	1.7	ATPphosphoribosyltransferasecatalyticsubunit	0.00	0.00
SMc00914	-	910.5	139.0	176.7	oxidoreductase	0.15	0.19
SMc00907	-	84.5	41.3	9.0	hypotheticalprotein	0.49	0.11
SMc00993	purD	1122.0	11.7	122.7	phosphoribosylamine--glycineligase	0.01	0.11

SMc00010	ctaD	618.5	259.0	253.7	cytochromeCoxidasepolypeptideItransmembraneprotein	0.42	0.41
SMc00450	ctaB	232.0	92.7	77.3	protohemeIXfarnesyltransferase	0.40	0.33
SMc02363	cycK	1697.5	680.0	498.7	cytochromeC-typebiogenesisistransmembraneprotein	0.40	0.29
SMc02394	-	120.5	51.3	29.7	hypotheticalprotein	0.43	0.25
SMc02461	-	444.0	128.3	113.0	hypotheticalprotein	0.29	0.25
SMc02407	-	540.0	224.0	181.0	hypotheticalprotein	0.41	0.34
SMc02641	rkpK	2339.0	900.7	507.7	UDP-glucose6-dehydrogenase	0.39	0.22
SMc02558	-	251.0	108.3	112.3	hypotheticalprotein	0.43	0.45
SMc00554	purF	2794.0	90.7	356.3	amidophosphoribosyltransferase	0.03	0.13
SMc00614	purN	1037.5	44.3	177.7	phosphoribosylglycinamideformyltransferase	0.04	0.17
SMc00615	purM	1148.5	36.3	143.3	phosphoribosylaminoimidazolesynthetase	0.03	0.12
SMc01770	glyA	1789.0	16.3	477.0	serinehydroxymethyltransferase	0.01	0.27
SMc01783	-	468.5	172.0	172.3	hypotheticalprotein	0.37	0.37
SMc01801	argC	1283.5	29.0	26.7	N-acetyl-gamma-glutamyl-phosphatereductase	0.02	0.02
SMc01930	-	221.5	106.3	93.3	hypotheticalprotein	0.48	0.42
SMc01933	-	114.0	40.7	47.0	hypotheticalprotein	0.36	0.41
SMc01361	-	797.0	93.3	280.7	dihydroorotase	0.12	0.35
SMc01360	pyrB	1099.5	114.3	342.0	aspartatecarbamoyltransferasecatalyticsubunit	0.10	0.31
SMc01355	-	275.5	29.0	37.3	Hollidayjunctionresolvase-likeprotein	0.11	0.14
SMc01004	hisI	327.0	1.7	0.7	phosphoribosyl-AMPcyclohydrolase	0.01	0.00
SMc01028	eno	746.0	176.0	351.0	phosphopyruvatehydratase	0.24	0.47
SMc02112	-	65.5	18.3	19.0	hypotheticalprotein	0.28	0.29
SMc02088	-	66.0	19.3	7.3	hypotheticalprotein	0.29	0.11
SMc01219	lpsB	225.0	95.0	64.0	lipopolysaccharidecorebiosynthesismannosyltransferase	0.42	0.28
SMc01207	queA	353.5	164.0	123.0	S-adenosylmethionine--tRNAribosyltransferase-isomerase	0.46	0.35
SMc01203	-	438.5	207.0	182.3	hypotheticalprotein	0.47	0.42
SMc00235	trpD	1259.5	82.7	178.0	anthranilatephosphoribosyltransferase	0.07	0.14
SMc00236	trpC	625.0	42.7	130.7	Indole-3-glycerolphosphatesynthase	0.07	0.21
SMc00293	thrA	1046.0	51.3	85.7	homoserinedehydrogenase	0.05	0.08
SMc00522	rhIE1	1181.0	477.0	568.7	ATP-dependentRNAhelicase	0.40	0.48
SMc00494	-	292.5	52.7	60.0	phosphoribosylformylglycinamidesynthasesubunitPurS	0.18	0.21
SMc00493	purQ	497.5	26.3	76.7	phosphoribosylformylglycinamidesynthaseI	0.05	0.15
SMc00488	purL	3542.5	178.7	565.3	phosphoribosylformylglycinamidesynthaseII	0.05	0.16
SMc00760	recA	253.0	48.7	52.7	recombinaseA	0.19	0.21
SMc00175	-	522.5	221.0	176.0	ABCtransporterATP-bindingprotein	0.42	0.34
SMc00155	aroF	267.5	16.7	0.0	DAHPSynthetaseprtein	0.06	0.00
SMc04346	ilvC	897.0	43.0	113.7	ketol-acidreductoisomerase	0.05	0.13
SMc01428	cspA2	32.5	15.7	11.7	coldshocktranscriptionregulatorprotein	0.48	0.36
SMc01431	ilvI	2943.0	178.7	275.3	acetolactatesynthase3catalyticsubunit	0.06	0.09
SMc01494	serB	1469.5	26.7	120.3	phosphoserinephosphatase	0.02	0.08
SMc01471	senC	507.5	248.3	202.7	cytochromeCoxidaseassemblyfactortransmembraneprotein	0.49	0.40
SMc01843	metF	625.0	7.3	25.7	5,10-methylenetetrahydrofolatereductase	0.01	0.04
SMc02695	-	251.5	114.3	80.0	GTP-dependentnucleicacid-bindingproteinEngD	0.45	0.32

SMc02717	leuA1	1385.0	81.3	196.3	2-isopropylmalatesynthase	0.06	0.14
SMc02725	trpE	2322.0	153.3	405.7	anthranilatesynthase	0.07	0.17
SMc01501	mtlK	1250.5	470.7	446.0	mannitol2-dehydrogenase	0.38	0.36
SMc02435	hemK1	393.5	93.3	102.3	methyltransferase	0.24	0.26
SMc00731	-	240.0	119.3	92.3	hypotheticalprotein	0.50	0.38
SMc00726	tlpA	971.0	229.3	319.3	thiol:disulfideinterchangeredox-activecentertransmembraneprotein	0.24	0.33
SMc00725	argH1	767.0	341.3	150.7	argininosuccinatelyase	0.45	0.20
SMc00723	lysA	637.5	12.0	4.0	diaminopimelateDAPdecarboxylase	0.02	0.01
SMc00711	tyrC	783.5	4.0	0.0	cyclohexadienyldehydrogenase	0.01	0.00
SMc00691	xerD	1059.5	396.0	442.0	site-specificityrosinerecombinaseXerD	0.37	0.42
SMc00674	hutC	1841.5	650.0	529.7	histidineutilizationrepressortranscriptionregulatorprotein	0.35	0.29
SMc00647	rluD	181.0	43.0	41.0	ribosomallargesubunitpseudouridinesynthase	0.24	0.23
SMc00643	purA	316.0	0.3	0.0	adenylosuccinatesynthetase	0.00	0.00
SMc00641	serA	1243.5	40.7	103.3	D-3-phosphoglyceratedehydrogenase	0.03	0.08
SMc00640	serC	642.0	1.0	0.7	phosphoserineaminotransferase	0.00	0.00
SMc03966	ruvA	245.5	61.3	114.0	HollidayjunctionDNAhelicaseRuvA	0.25	0.46
SMc03978	tkt2	425.5	13.7	127.3	transketolase	0.03	0.30
SMc03979	gap	1129.5	70.3	315.7	glyceraldehyde-3-phosphatedehydrogenase	0.06	0.28
SMc03981	pgk	1808.5	279.7	758.3	phosphoglyceratekinase	0.15	0.42
SMc04001	purE	2678.0	122.0	371.0	phosphoribosylaminoimidazolecarboxylasecatalyticsubunitprotein	0.05	0.14
SMc04002	purK	1545.5	38.0	459.3	phosphoribosylaminoimidazolecarboxylaseATPasesubunit	0.02	0.30
SMc04005	pykA	798.0	173.3	308.7	pyruvatekinase	0.22	0.39
SMc04026	gltD	3276.0	134.0	247.3	glutamatesynthase	0.04	0.08
SMc04028	gltB	7012.5	419.3	740.7	glutamatesynthase	0.06	0.11
SMc04045	ilvD2	3452.5	197.3	475.7	dihydroxy-aciddehydratase	0.06	0.14
SMc02983	-	749.5	116.0	196.0	ornithine,DAP,orargininedecarboxylase	0.15	0.26
SMc03140	-	552.5	245.7	236.3	transcriptionalregulator	0.44	0.43
SMc03112	metH	4176.0	92.0	132.7	B12-dependentmethioninesynthase	0.02	0.03
SMc02489	xerC	987.0	335.3	300.7	site-specificityrosinerecombinaseXerC	0.34	0.30
SMc02479	mdh	438.0	28.3	56.0	malatedehydrogenase	0.06	0.13
SMc02478	-	639.5	265.7	286.0	hypotheticalprotein	0.42	0.45
SMc03777	proA	2950.0	1397.7	917.3	gamma-glutamylphosphatereductase	0.47	0.31
SMc03795	leuD	309.5	35.7	61.7	isopropylmalateisomerasesmallsubunit	0.12	0.20
SMc03797	metA	599.5	20.7	45.3	homoserineO-succinyltransferase	0.03	0.08
SMc03823	leuC	2569.5	185.7	425.7	isopropylmalateisomeraselargesubunit	0.07	0.17
SMc03826	argG	340.5	19.3	15.0	argininosuccinatesynthase	0.06	0.04
SMc03848	ccmB	649.0	280.0	225.3	hemeexporterB(cytochromeC-typebiogenesisprotein)transmembrane	0.43	0.35
SMc03849	ccmC	510.0	185.0	207.3	hemeexporterC(cytochromeC-typebiogenesisprotein)transmembrane	0.36	0.41
SMc03851	ccmG	642.5	260.0	234.0	thiol:disulfideinterchangeprotein(cytochromeCbiogenesisprotein)	0.40	0.36
SMc03858	pheAa	117.0	33.3	39.0	chorismatemutase	0.28	0.33
SMc03884	ispA	252.5	66.7	97.0	geranyltranstransferase	0.26	0.38
SMc04405	leuB	3939.0	220.7	571.3	3-isopropylmalatedehydrogenase	0.06	0.15
SMc04088	purH	1682.5	23.0	303.7	bifunctionalphosphoribosylaminoimidazolecarboxamideformyltransferase/IMPcyclohydrolase	0.01	0.18

SMc02812	-	825.0	130.0	397.0	hypotheticalprotein	0.16	0.48
----------	---	-------	-------	-------	---------------------	------	------

SUPPLEMENTAL METHODS

Tn-seq protocols

A. Generate Complex Transposon Library

Mutagenize strain using pJG714 (which delivers a mini-Tn5 from E. coli delivery strain MFDpir). Set up mating on rich medium containing diaminopimelic acid (DAP; required for MFDpir viability). Allow only a few hours of conjugation time. After selection on rich medium containing kan or neo, scrape library of ~50,000-150,000 colonies into a single Falcon tube, vortex and freeze aliquots. This is the input library. Pass this library through any selective environment (or rich medium as a control), and then grow out in rich medium prior to pelleting and storing at -80.

B. Genomic DNA Extraction

(MoBio Cat# 12255-50)

1. Pellet 1.0-1.5 ml (dep. on cell density) by centrifuging full speed for 1 minute. D/T, then remove all supernatant by centrifuging a second time and removing residual supernatant
2. Resuspend in 300 µl MicroBead solution, transfer to bead tube
3. Add 50 µl MD1 (check precip)
4. Vortex at full speed for 10 minutes using bead beating adapter
5. Using clipped tip, move contents (foam/liquid/beads) to regular microtube
6. Centrifuge at full speed for 3 minutes
7. Move ~290 µl of supernatant to a new tube
8. Add 100 µl MD2 and then vortex for 5 seconds. Ice for 5 minutes
9. Centrifuge at full speed for 3 minutes
10. Move ~370 µl of supernatant to new tube
11. Add 900 µl MD3, vortex for 5 seconds
12. Pass liquid through spin filter, 630 µl at a time (1-min spins), discard flowthrough
13. Wash spin filter with 300 µl of MD4, discard flowthrough
14. Wash spin filter with 500 µl of PE (Qiagen), discard flowthrough
15. Centrifuge at full speed for 1 minute
16. Move filter to new tube, add 55 µl Tris-2.5 (warm)
17. Wait 2 minutes and then centrifuge full speed, 1 minute
18. Check 2-µl sample on a gel. Should be bright, high molecular-weight band.

C. Primers for Illumina Library Construction

The kan-out end of Tn5-714 has the following sequence:

```
ctgaccggctcgacCTCGAGATGTGTATAAGAGACAG  
<1st rnd prim> <--transposon end-->
```

Final library structure should be as follows:

```
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT(XXX)TCGAGATGTGTATAAGAGACAG
```

```
← P5 flowcell anchor → ← Illumina Read 1 primer → ← Transposon end →
```

```
^stagger site
```

```
...FlankSeq-GATCGGAAGAGCACACGTCTGAACTCCAGTCACACAGTATCTCGATGCGCTCTTCTGCTTG
```

```
..<FlankSeq><---Index primer binding site---><INDX><--P7 flowcell anchor-->
```

Indexed side varies:

```
BAR1-GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTATGCCGTCTTCTGCTTG  
BAR2-GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATGCCGTCTTCTGCTTG  
BAR3-GATCGGAAGAGCACACGTCTGAACTCCAGTCACCTAGGCATCTCGTATGCCGTCTTCTGCTTG  
BAR4-GATCGGAAGAGCACACGTCTGAACTCCAGTCACGCAATATCTCGTATGCCGTCTTCTGCTTG  
BAR5-GATCGGAAGAGCACACGTCTGAACTCCAGTCACACAGTATCTCGTATGCCGTCTTCTGCTTG  
BAR6-GATCGGAAGAGCACACGTCTGAACTCCAGTCACAGATCATCTCGTATGCCGTCTTCTGCTTG  
BAR7-GATCGGAAGAGCACACGTCTGAACTCCAGTCACCTAGGCATCTCGTATGCCGTCTTCTGCTTG  
BAR8-GATCGGAAGAGCACACGTCTGAACTCCAGTCACGATCAGATCTCGTATGCCGTCTTCTGCTTG
```

First-round Tn primer (1TN712):

```
CTGACCCGGTTCGAC (14 nt, 46/49)
```

Second-round Tn primers (2TN712X): (51/58)

2TN712a (LB condition)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGAGATGTGTATAAGAGACAG

2TN712b (RSM-G condition)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTaTCGAGATGTGTATAAGAGACAG

2TN712c (A17 condition)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTgaTCGAGATGTGTATAAGAGACAG

2TN712d (Rice condition)

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTcgaTCGAGATGTGTATAAGAGACAG

First-round oligo-G reverse primer (1OLIGOG):

CAGACGTGTGCTCTTCCGATCgggggggggg (11 g's, 44/44)

Second-round reverse primer (2BARX):

CAAGCAGAAGACGGCATAACGAGATnnnnnnGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC (56/63)

2BAR1:

CAAGCAGAAGACGGCATAACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC

2BAR2:

CAAGCAGAAGACGGCATAACGAGATACATCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC

2BAR3:

CAAGCAGAAGACGGCATAACGAGATGCCTAAGTACTGGAGTTCAGACGTGTGCTCTTCCGATC

2BAR4:

CAAGCAGAAGACGGCATAACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC

2BAR5:

CAAGCAGAAGACGGCATAACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC

2BAR6:

CAAGCAGAAGACGGCATAACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC

2BAR7:

CAAGCAGAAGACGGCATAACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC

2BAR8:

CAAGCAGAAGACGGCATAACGAGATCTGATCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC

D. HTML-PCR Library Preparation

Fragmentase digestion (NEB M0348S)

Mix 16 µl genomic DNA + 2 µl 10X fragmentase v.2 buffer
Pre-vortex Fragmentase, place 2 µl in a new tube
Add DNA/buffer mixture to the Fragmentase for rapid mixing
37C (heat block or water bath) for exactly 10 minutes
Add 10 µl 0.25 M EDTA to stop reaction, place on ice

Clean up with size selection (Qiagen 27106)

To the reaction, add exactly 200 µl of 0.3X PB (300PB+700water)
Apply to column, spin 30 sec
Add 200 µl more of 0.3X PB, spin 30 sec, discard flowthrough
Add 650 µl PE, spin 30 sec, discard flowthrough
Spin 1 min to dry the column thoroughly
Elute in 50 µl warm 2.5 mM Tris-8
Check 4 µl on gel; should see broad smear from 500-3000 bp

C-Tailing (TdT is NEB M0315S, ddCTP is Affimetrix 77112 0.5UM)

30 µl cleaned up fragmented DNA
4 µl 10X TdT buffer
4 µl 2.5 mM CoCl₂
2 µl 9.5 mM dCTP/0.5 mM ddCTP mix
(34 of water+ 4 of 100mM dCTP+ 2 of 10mM ddCTP)
0.6 µl TdT enzyme, mix well
37C for 30 min

Clean-up with size selection

(see above; check 7 µl on gel; smear should look identical to frag. product)

First-round PCR (Taq is NEB M026L)

23.5 µl water
4.0 µl ThermoPol buffer
1.2 µl 10 mM dNTPs
0.3 µl Taq
3.0 µl 1/10 diluted (1TN712)
3.0 µl 2/10 diluted (1OLIGOG)
5 µl template DNA from TdT rxn

Program: "Round1"

1. 94 1:00
2. 94 :20
3. 42 :40
4. 70 :20
5. GOTO 2, 25 times
6. 70 1:00
7. 4 00:00
8. END

Clean up with size selection

(see above; check 4 µl on gel)

Second-round PCR

26.2 µl water
4.0 µl ThermoPol buffer
1.2 µl 10 mM dNTPs
0.3 µl Taq
3.0 µl 1/10 (2TN71x; x-a..d)
3.0 µl 1/10 (2BARx; x=1..8)
2.3 µl cleaned up First-round product

Program: "Round2"

1. 94 1:00
2. 94 :20
3. 52 :20
4. 70 :20
5. GOTO 2, 15 times
6. 70 1:00
7. 4 00:00
8. END

Clean up with Gel extraction

Pre-weigh tubes
Load all of Round 2 product in large-comb well (+8 µl dye)
Cut gel slice just above and just below 500bp
Weigh tubes with gel slice
Subtract weight of tube to determine gel weight
Weight of gel in mg constitutes 1 reaction volume
Add 3 reaction volumes Buffer QG to gel slice
Incubate at 37°C until gel is dissolved (vortex every few minutes)
Color should be yellow – if orange or violet consult manual)
Add 1 reaction volume 2-propanol (mix)
Spin down in column (800 µl per spin)
Discard flowthrough
Add 500 µl Buffer QG
Spin and discard flowthrough
Add 750 µl PE Buffer
Sit 5 minutes
Spin 30 seconds and discard flowthrough
Spin 1 Min
Move to new tube
Add 50 µl warm Tris 2.5
Sit 2 minutes
Spin

pJET cloning verification

Blunting

5 µl 2x Reaction Buf
0.5 µl Gel extracted Round 2 PCR product
3 µl H₂O
0.5 µl blunting enzyme
mix
70°C for 5 minutes

Ligation

Add directly to blunting reaction
0.5 µl pJET
0.5 µl T4 DNA Ligase
mix
Room temperature for 5 minutes
Chill on ice for 2 minutes
Add 10 µl to 100 µl DH5α CaCl₂ cells
Ice 5 min
Heat shock 30 seconds
Ice 5 minutes
Add 800 µl LB
Recover at 37°C for 20 minutes
Plate 100 µl on LB-Ap plates
PCR verification
8.3 µl H₂O
1 µl Taq Buffer
0.25 µl dNTPs
0.05 µl Taq
0.2 µl 10 µM pJET1.2 For
0.2 µl 10 µM pJET1.2 REV
Touch single colony with p20 tip and mix in PCR tube

pJG714 sequence

ggtacctcagatcttgatcccctgcgccatcagatccttggcggcaagaaagccatccagtttactttgcagggttcccaaccttcccagag
ggcgccccagctggcaattccggctcgttgcctcgcataaaaccgccagctctagctatcgccatgtaagcccactgcaagctacgtctt
ctctttgcgcttgcgtttcccttgcagatagcccagtagctgacattatccggggcagcaccgtttctgcggactggctttctacgtgtc
cgcttcttttagcagcccttgcgccctgagtgcttgcggcagcgtgaagctttctctgagctgtaacagcctgaccgcaacaacgagagga
tcgagaccatccgctccagattatccggctcctccatgcgcttgcctctcggctcctgctccgggtttccatgccttatggaactcctc
gatccgc
cagcgtatgggtataaatgctgatgacgcgcaaggcttgggctagcactcagccggctcgtggtcagcaacaaccattcaacggggtct
cacccttgggcgggtaaatcctcggccagcaccgcgtgagcgtgatattcccctgttttagcgtgatgcgccactgcgcaggctcaag
ctcgccttgcgggctggctgattttacgtttaccgcgtttatccaccacgcccttttgcggaatgctgatctgatagccaccaactccggtt
gttcttcagatggctgacagatacaaccagactctacgtccttgcgtgggtgcttggagcgcaccacgaagcgtcgttatgcgccagtt
gtcctgcagataagcatgaatateggcttgcggtcacagaccgcaatcacgttctcatcatgctgcccattgcgtaaccggctagttgcgg
ccgctgccagccatttgcactctcctttatccgcacatcggcagggcatccgggcgatccaccactcctgatgcagtaatectacgggtgc
ggaatgtggtggcctcagcaagagaacggagtgaaccaccatccgcgggattatcctgaatagagcccagcttgccaagctcttcgg
cgacttgggtggcgataactcaagaggtggtgctcctaatggccagcagttcgggaaactcctgagccaacttgactgtttgatggcgc
agcctttctgatcgcctcggcagaaacgttgggattgcggataaatcggtaagcgccttctgggcggcttactaccctctgatgagatggt
tattgattaccagaatattttgccaattgggcggcgacgtaaccaagcgggcagtagcggcagggatcaccacgcgcgcccgaagagaa
cacagatttagcccagctcggccgcacgatgaagagcagaagtatcatgaacgttaccatgttaggaggtcacatggaagatcagatcctg
gaaaacgggaaaggttccgttcaggacgctggtacctgtctttatacacatctcaggtcagccgggtcagaagaactcgtcaagaaggc
gatagaaggcgtatgcgctgcgaatcgggagcggcgataaccgtaaagcacgaggaagcgggtcagcccattcggcccaagctcttcagc
aatatcacgggtagccaacgctatgtcctgatagcggctccgccacaccagccggccacagtcgatgaatccagaaaagcggccattttcc
accatgatattcggcaagcaggcatcgccatgggtcacgacgagatcctcgcctcgggcatgcgccttgagcctggcgaacagttcg
gctggcgcgagcccctgatgctcttcgtccagatcctcctgatcgacaagaccggcttccatccagtagctgctcgtcgtcgtgatgcgatgttc
gcttgggtgctgaatgggcaggttagccgatcaagcgtatgcagccggccgattgcatcagccatgatggatactttctcggcaggagcaa

ggtgagatgacaggagatcctgccccggcacttcgccaatagcagccagtccttcccgttcagtgacaacgtcgagcacagctgcgc
aaggaacgcccgtcgtggccagccacgatagccgcgctgcctcgtcctgcagttcattcagggcaccggacaggtcggcttgacaaaa
gaaccgggcgcccctgcgctgacagccggaacacggcggcatcagagcagccgattgtctggtgcccagtcatagccgaatagcctct
ccaccaagcggcgggagaacctgcgtgcaatccatcttgttcaatcatgcgaaacgatcctcatctgtctctgatcagatcttgatccct
gcccacatcagatccttggcggcaagaaagccatccagttactttgcagggcttcccaaccttaccagagggcgccccagctggcaattcc
ggctcgttgctgtgggggatccatcataacgggtccggcaatatactgaaataggtgtgacattatccatcgaactagttaactagtacga
aagttcacatctagatgtgtataagagacagagctccctgctcggggctcattatagcagtttttcgggtatccatccttttcgcacgatatac
aggattttgcaaaagggttcgtgtagacttcttgggtgtatccaacggcgtcagccgggcaggataggtgaagtaggccaccgcgagc
gggtgtccttcttactgtcccttattcgcacctggcgggtgtcaacgggaatcctgctctgcgaggctggccggctaccgcccggcgtaca
gatgagggcaagcggatggctgatgaaaccaagccaaccaggaaggcagccacgatgcgtccggcgtagaggatctgaagatcagc
agttcaacctgttgatagctactaagctctcatgtttcacgtactaagctctcatgtttaacgtactaagctctcatgtttaacgaactaaacc
tcatggctaactactaagctctcatggctaactactaagctctcatgtttcacgtactaagctctcatgtttgaacaataaataataatc
agcaacttaaatagcctctaaggtttaagttttataagaaaaaaagaatatataaggcttttaagcttttaaggtttaacgggtgtggacaaca
agccagggatgtaacgcactgagaagcccttagagcctctcaaagcaatgttgagtgacacaggaacacttaacggctgacatgggagtac
tcaaccaagtcattctgagaatagtgtatcgggcgaccgagtgctcttgcggcgcaatacgggataataaccgcccacatagcagaac
ttaaaagtgtcatcattggaaaacgttctcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgtatgaaccactcgtg
cacccaactgatcttcagcatcttttactttaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaagggaataagg
gcgacacggaaatgtgaatactcatacttctcttttcaatattattgaagcattatcagggttattgtctcatgagcggatacatattgaatgt
atttagaaaaataaacaataggggttcgcgcacatttcccgaaggtccacctgacgcgccctgtagcggcgcaataagcgcggcg
gggtgtgggtgtacgcgcagcgtgaccgctacacttgcagcgccttagcggcctctcttctcgtttcttcccttcttctcgcacgttgc
cggctttcccgtcaagctctaaatcgggggctcccttaggggtccgatttagtctttacggcacctcgacccccaaaaaacttgattagggt
gatggttcacgtagtgggcatcgcctgatagacggttttcggccttgacggtggagtcacgttcttaatagtggactcttgttccaaact
ggaacaactcaaccctatctcggctattctttgattataagggtttgccgattcggcctattggttaaaaaatgagctgatttaacaaaa
atttaacgcgaattttaacaaaatattaacgcttacaatttccattcgcattcaggctgcgcaactgttgggaaggcgatcgggtcgggcct
cttcgctattacgccagctggcgaaagggggatgtgctgcaaggcgattaagttgggtaacgccagggtttcccagtcacgacggtgtaa
acgacggccagtgaattgtaatacgaactactatagggcgaattg