5-2021

Hemin Utilization in Rhizobium leguminosarum ATCC 14479

John Lusby
East Tennessee State University

Hemin Utilization in *Rhizobium leguminosarum* ATCC 14479

A thesis

presented to

the faculty of the Department of Biological Sciences

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Science in Biology, Concentration in Microbiology

---

by

John R. Lusby

May 2021

Dr. Ranjan Chakraborty

Dr. Sean Fox

Dr. Bert Lampson

Keywords: *Rhizobium leguminosarum* ATCC 14779, hemin, Hmu operon, iron uptake, HWE sensor kinase

ABSTRACT

Hemin Utilization in *Rhizobium leguminosarum* ATCC 14479

by

John R. Lusby

Rhizobium leguminosarum is a Gram negative, motile, nitrogen-fixing soil bacterium. Due to the scarcity of iron in the soil bacteria have developed a wide range of iron scavenging systems. The two types of iron scavenging systems used are indirect and direct. *In-silico* analysis of the genome identified a unique direct iron scavenging system the *Hmu* operon. This system has been identified in other closely related *rhizobium* species and is believed to be involved in utilizing heme compounds as a sole source of iron. We have attempted to characterize the role of the *Hmu* operon in iron utilization by monitoring the growth of *R. leguminosarum* ATCC 14479 in hemin supplemented media. Growth curves show that it is capable of using hemin as a sole source of iron. The outer membrane profiles were analyzed for the presence of hemin binding proteins.

ACKNOWLEDGEMENTS

I would like to thank my committee members, Dr. Chakraborty, Dr. Fox, and Dr. Lampson for their wisdom and guidance. I would also like to thank Dr. Pritchett for his support in trouble shooting problems that I encountered during my research. I am thankful for Dr. Chakraborty's support and encouragement throughout the completion of my degree. I am grateful for Dr. Lampson's Genetics insight and feedback during sequence analysis. I am indebted to Dr. Fox for his guidance and informing me about the graduate program. I would like to thank Robin Grindstaff for her help in preparation of media and overall guidance in the media kitchen. Thank you to the faculty and staff of the Health Sciences department and my fellow graduate students in their encouragement and support during the completion of the program. Lastly I would like to thank my family and friends for their encouragement and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

Iron and Its Importance

Iron is an essential micronutrient for living organisms including both eukaryotes and prokaryotes, It's required at an intracellular concentration of 10^{-6} to 10^{-7} Molar (M) to maintain homeostasis in bacteria (Weinberg 1974). Iron is required for many cellular functions as it serves as an important cofactor for many crucial enzymes required for a multitude of functions such as DNA synthesis, respiration, and gene regulation (Andrews et al. 2003). Iron deficiencies in bacteria can lead to: defects in DNA/ RNA biosynthesis, cellular morphology, reduced respiration activity, oxidative stress response, Nitrogen fixation, and oxidative phosphorylation (Messenger and Barclay 1983). Inversely an excess of iron in bacteria can lead to the buildup of reactive oxygen species (ROS) which can lead to cell death (Touati et al. 1995).

In humans, excess iron also known as iron overload is mainly caused by hemochromatosis, a hereditary condition which causes an individual to absorb to much iron from their diet. The excess absorption of iron leads to deposition of the excess iron into various tissues and organs, the main regions it is deposited are: the liver, skin, heart, pancreases and joints (Hereditary hemochromatosis 2020). Excess iron buildup in these organs can lead to a multitude of issues ranging from impaired organ function to organ failure due to the buildup of ROSs (Crownover & Covey 2013). Inversely, iron deficiency can cause a vast array of physiological problems as well, ranging from: impaired cognitive performance in children and elderly, pregnancy complications, chronic fatigue, impaired heart function, and increased risk of infections (Camaschella 2015).

Iron Availability

Iron is found in two oxidation states in the environment, the insoluble ferric (Fe III) and soluble ferrous (Fe II); soluble iron is a scarce resource that is found in most soil environments at a concentration from 10^{-9} to 10^{-18} M (Miethke and Marahiel 2007). Most microorganisms require an intracellular iron concentration around 10^{-6}M in order to maintain homeostasis (Miethke and Marahiel 2007). Pathogenic bacteria are also faced with iron scarcity while inside their host. Iron is tightly regulated in eukaryotic cells, due to its homeostatic importance. Mammalian cells keep the free iron pool extremely low to prevent iron overload and to limit pathogen growth (Miethke and Marahiel 2007). Eukaryotic cells reduce the free iron pool by storing it in proteins such as heme compounds and ferritins, keeping free iron concentrations at 10^{-24}M (Miethke and Marahiel 2007). Due to both soil dwelling and pathogenic bacteria being faced with environments in which iron is scarce, they have developed a multitude of complex iron acquisition systems.

Iron Acquisition Systems

Due to the important role iron plays in maintaining homeostasis and it's scarcity in the environment bacteria have developed a wide array of iron scavenging systems to gather both free and complexed forms of iron from their environment. Bacteria can acquire iron from the environment via direct and indirect iron uptake systems (Krewulak and Vogel 2007). Direct systems require the bacteria to make physical contact with a source of iron such as heme, transferrin, lactoferrin, and hemoproteins (Miethke am Marahiel 2007). The drawback to direct uptake systems is each type of iron compound must have its own specific surface receptor and transport system for that compound to be utilized. This high specificity limits the range of

growth to areas where that iron source is available (Miethke and Marahiel 2007). Pathogenic bacteria contain a plethora of direct iron acquisition systems, an example would be *Pseudomonas aeruginosa'*s *phuSTUVW* system. This system binds heme through *PhuR* a specific heme surface receptor. Once bound, the ATP-binding cassette (ABC) encoded by the *phuSTUVW* operon transports and breaks down the heme to release the iron to be used by the cell (Ochsner et al. 2000). Other pathogenic bacteria utilize similar approaches such as the hemoglobin utilization system *hmuRPSTUV* in *Yersinia pestis*, and *Neisseria*'s Transferrin-binding protein A and B (Perkins-Balding et al. 2004; Schwieosow et al. 2018). Due to the limitation of direct iron acquisition systems limiting bacterial growth to locations where the useable iron source is available, bacteria have also developed indirect iron acquisitions systems to help compensate for this limitation.

Indirect iron acquisition systems are more diverse than direct iron acquisition systems, thus broadening the area of colonization. Indirect systems involve the secretion of compounds into the environment to capture free iron or iron containing compounds (Miethke and Marahiel 2007). The *HxuABC* system in *Haemophlis influenzae* is an example of an indirect uptake system. *HxuA* is secreted into the extracellular environment through *HxuB*, once secreted *HxuA* binds a hemoprotein which results in the release of the heme component (Zambolin et al. 2015). Once the heme component is released it is bound by the TonB dependent *2* surface receptor (Zambolin et al. 2016). The other major type of indirect iron acquisition system is a low molecular weight compound known as siderophores. Siderophores are produced by both prokaryotic and eukaryotic cells and are classified into three major categories; which are catecholates, hydroxamates, and carboxylates (Miethke and Marahiel 2007). Siderophores have a high affinity for iron, and when secreted into the environment they bind ferric iron (Miethke

Marahiel 2007). Once the ferric siderophore complex is obtained bacteria are able to bind this complex through siderophore specific receptors (Wilson et al. 2017). Differences arise between Gram positive and Gram negative organisms once this complex is bound. Gram negative bacteria have to transport this complex across two membranes, requiring the help of a periplasmic binding protein to traffic the complex to the inner membrane permease (Wilson et al. 2017). This trafficking process of the ferric siderophore complex from extracellular to intracellular requires energy which is presumably provided by the TonB-ExbB/D complex (Miethke and Marahiel 2007). Gram positive organisms contain only one membrane to transport the ferric siderophore complex, requiring only the siderophore receptor and associated permease (Wilson et al. 2017). Once the ferric siderophore complex is inside the cell the bacteria must release the iron to make it available for use. The release of iron from the siderophore is done by breaking down the siderophore via esterases or reducing the ferric iron to ferrous iron via reductases (Miethke and Marahiel 2007). The combination of both direct and indirect iron acquisition systems give bacteria an evolutionary advantage by increasing the range of environments they can colonize; however, due to the high specificity of direct acquisition systems it is more beneficial for bacteria to utilize indirect acquisition systems.

Genetic Regulation of Iron Acquisition Systems

Due to the importance iron plays in maintaining homeostasis in bacterial growth, they have developed a multitude of iron scavenging systems. These systems are only expressed in iron limited conditions due to the amount of energy they require to scavenge iron from the environment (Miethke and Marahiel 2007). These systems are also tightly regulated due to the buildup of deadly ROS when intracellular iron is in excess (Kadner 2005). One of the most common iron starvation gene regulation mechanisms is known as the ferric uptake regulator

(Fur). The Fur protein is an iron-dependent transcriptional repressor; the system was first found in *Escherichia coli,* but it is also common in other Gram negative bacteria (Kadner 2005). Fur is a Fe-sulfur protein that is a global iron starvation response regulator which binds to a specific DNA sequence called the Fur box. This protein works by using either ferric iron or Manganese^{+2} as a corepressor (Bagg and Neilands 1978; De Lorenzo et al. 1987). When ferric iron is abundant inside the cell it binds to Fur causing a conformational change, this allows for the protein to bind to the Fur box. Once the protein is bound to the Fur box it blocks DNA transcription of the downstream genes involved in iron scavenging systems (Miethke and Marahiel 2007). When bacteria are in iron limited conditions the ferric iron is released from the fur protein resulting a conformational change that leads to the disassociation between Fur and the Fur box, thus allowing for gene expression (Troxell and Hassan 2013).

The Fur Family of proteins also contains the iron-response regulator (Irr), while Irr is in the Fur family of proteins it uses a different mode of action to measure iron levels (Costa et al. 2017). Fur operates by measuring intracellular iron levels, while Irr in *Bradyrhizobium* operates by sensing heme biosynthetic levels (O'Brian 2015).

Another iron gene regulator is the rhizobial iron regulator (RirA) which belongs to the Rrf2 family of transcription regulators, mutations to RirA have shown to alter not only iron-response genes but also alter the expression of over 100 proteins (Chao et al. 2005). DNA and protein sequence homology searches of the RirA protein shows no homology to known iron-response regulators and is unique to *Rhizobiales* order (Chao et al. 2005). The RirA protein is a Fe-sulfur protein that's exact mode of action is not fully understood, but an iron-responsive operator (IRO) motive has been described as a putative DNA binding site for RirA protein (Costa et al. 2017).

RirA is known to repress iron-response genes in iron rich environment, and allows for iron-response gene expression in iron poor environments (Todd et al. 2002).

A newly investigated method of gene regulation in bacterial iron starvation response is the Two-Component Signaling system (TCS) also called histidine sensor kinase system. The TCS work by sensing changes in the environment around the bacteria, once a change is perceived a phosphorylation relay occurs that leads to a change in gene expression allowing the bacteria to adapt to the changing environment (Bijlsma and Groisman, 2003; Cheung and Hendrickson, 2010). This system is comprised of two components, a sensor histidine kinase that is transmembrane, and a cytoplasmic response regulator (Bijlsma and Groisman, 2003; Cheung and Hendrickson, 2010). The sensor histidine kinase detects changes in the environment through the N-terminus leading to autophosphorylation of a conserved histidine residue on the C-terminus (Tiwari et al. 2017). The phosphorylated C-terminus then interacts with the response regulator and the phosphate group is transferred to a conserved aspartate residue on the N-terminus of the regulator. This phosphate transfer causes a conformational change in the regulator's C-terminus (Tiwari et al. 2017). Once the response regulator undergoes a conformational change it interacts with the bacterial DNA allowing for changes in gene regulation (Tiwari et al. 2017). A new TCS in *Xanthomonas campestris* has been shown to be involved with iron homeostasis (Wang et al. 2016). The sensor histidine kinase is called *VgrS*, while the response regulator is called *VgrR*. This system has been shown to not only detect extracellular iron scarcity but also intracellular; when *VgrS/R* was mutated the bacteria showed highly impaired growth compared to the wild type (Wang et al. 2016).

Rhizobium leguminosarum

Rhizobium leguminosarum ATCC 14479 is a Gram negative bacteria which is known for being a soil dwelling, aerobic, motile, nitrogen fixing bacteria that forms a symbiotic relationship with legume plant roots (LeRoux et al. 2016). Rhizobia are present in a free-living form, and they infect legumes once they detect the flavonoids released by plant roots which triggers the infection of the root hairs. (Peters et al. 1986). Upon infecting the root hairs the bacteria express the nodulation factors and begin formation of root nodules (D'Haeze and Holsters 2002). Upon the completion of the root nodule, a portion of the bacterial population invades the nodule. Once intracellular the bacteria differentiate into bacteroides which are capable of converting atmospheric nitrogen into ammonia (Gage 2004). The root nodules provide bacteria with suitable environment to grow by providing protection and nutrients, while bacteria supply the plant with a nitrogen source (Postgate 1998.). The root nodule provides a low oxygen environment by the production of leghemoglobin, which binds molecular oxygen (Brear et al. 2013). The creation of this low oxygen environment is crucial in the nitrogen fixation process due to bacterial oxygen demands for growth and irreversible inactivation of bacterial nitrogenase by molecular oxygen (Postgate 1998). The resulting ammonia formed by a reduction of nitrogen by the bacteria is then assimilated into organic compounds used by the plant (Postgate 1998). The organic-nitrogen from the dead plants is then dispersed in the surrounding soil increasing its fertility, in turn allowing other legumes and plants to thrive (Postgate 1998).

There are three biovars of *R. leguminosarum* each of which form nodules with different plants: biovar *trifolii* nodulates with clovers, biovar *viciae* nodulates with legumes, and biovar *phasoli* specifically nodulates with kidney beans (Young et al. 2006). The three biovars share many similarities from infection process, nodule formation, nitrogen fixation, to genetic

similarities (Ramirez-Bahena 2008). The Strain of *Rhizobium* used in this study is *R. leguminosarum* ATCC 14479, it belongs to the *trifolii* biovar, it's preferred plant symbiote is *Trifolium pretense* also known as red clover (Wright et al. 2013.)

Heme iron uptake and the HmuPSTUV operon

Heme and hemoproteins have been utilized by pathogenic bacteria as a source of iron during infection to deal with the iron limited environment of their host (Contreras et al. 2014). Many different heme uptake systems have been studied in various pathogenic organisms such as the *HmuR* system in *Y. pestis* and the *phuPSTUVW* system in *P. aeruginosa* are a few examples. Heme utilization systems work as a direct iron acquisition system to help bacteria meet iron demands. They work by binding heme compounds to a specific surface receptor and transporting them into the cells to be degraded to release the iron (Ochsner et al. 2000; Perkins-Balding et al. 2004; Schwieosow et al. 2018). The reason so many pathogenic bacteria have developed heme utilization systems is due to it being the most widespread source of available iron in mammals (Contreras et al. 2014). While pathogenic bacteria have been widely studied for heme utilization systems, these systems are not exclusively found in human pathogenic bacteria but also in plant pathogens and plant symbiotic bacteria (Anzaldi and Skaar 2010).

Plant symbiotes have recently had their heme utilization systems investigated. These systems need more investigation to fully understand their role in iron acquisition and in symbiosis. *Bradyrhizobium japonicum* recently had a heme utilization system discovered, the system was labeled as *hmuRTUV*. This system shares similarities to the human pathogen *Y. enterocolitica* heme utilization system (Nienaber et al. 2001). It works by the heme surface receptor(*HmuR*)

binding heme compounds. Once bound the ABC membrane transport system (*HmuTUV*) works to internalize the heme compound for degradation and iron release (Nienaber et al. 2001). *Rhizobium leguminosarum* is another plant symbiote that has recently had a heme utilization system discovered, the system is the *hmuPSTUV* operon (Wexler et al. 2001). This system (Figure 1) has only had putative functions determined by sequence alignments with other known heme utilization genes (Wexler et al. 2001). The predicted functions for the proteins coincide with other heme ABC transport systems, *HmuP* is a predicted heme receptor/ transporter, while *HmuS* is predicted to be a heme degradation protein (Wexler et al. 2001). *HmuTUV* is the predicted ABC transporter component responsible for transporting heme from the periplasm to inside the cell (Wexler et al. 2001). This predicted ABC transporter is composed of a putative periplasmic heme binding protein (*HmuT*), a periplasmic heme permease (*HmuU*), and a heme ATPase component (*HmuV*) (Wexler et al 2001). While the functions of this system is putative, *R. leguminosarum* species are shown to be able to utilize heme compounds such as hemin and leghemoglobin as a sole source of iron (Wexler et al. 2001). Previous work in our lab along with full genome sequencing has confirmed the presence this predicted Heme uptake system in *R. leguminosarum* ATCC 14479

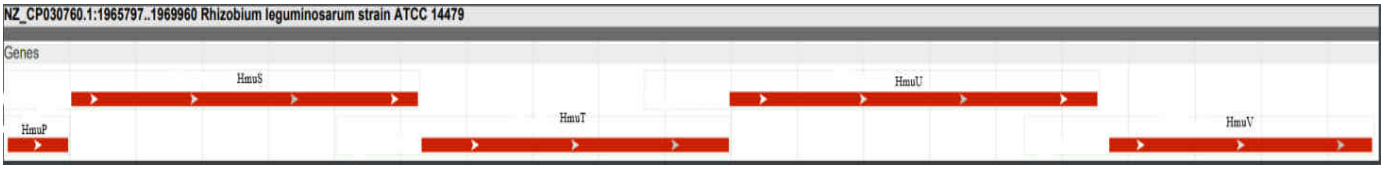


Figure 1. Orientation and layout of *R. leguminosarum* ATCC 14479 *HmuPSTUV* operon

TonB-ExbB-ExbD

Gram Negative bacteria have many outer membrane transporters, these transporters require energy to facilitate transport into the bacteria's inner membrane. The periplasm is the space

between the outer and inner membranes, this space resembles the pH and solute concentration of the external environment of the bacteria and is completely devoid of adenosine tri-phosphate (ATP) (Mogenesn and Otzen, 2005). Since the periplasm lacks the normal means of energy to facilitate transport, bacteria have developed systems to overcome the lack of energy to allow for transfer of molecules across the outer membrane and periplasm to the inner membrane. The TonB-ExbB-ExbD system provides energy across the periplasmic space to help facilitate transport of iron chelators, vitamin B_{12}, nickel chelators, and carbohydrates (Schauer et al. 2008). The energy produced for this transport is obtained by proton motive force (PMF) across the inner membrane, then the energy is transduced through TonB to the TonB dependent transporters (Noinaj et al. 2010).

The TonB complex is mostly studied in *E. coli,* where TonB is a 26kDa protein consisting of three domains. These domains are a cytoplasmic N-terminal domain, a periplasmic C-terminal domain, and a proline rich spacer separating the N- and C-terminal domains (Postle and Larsen, 2007). The N-terminal domain contains a 32-residue transmembrane helix, which anchors the protein to the cytoplasmic membrane and also serves as the interaction site with ExbB/D (Krewulak et al. 2007). The C-terminal domain of TonB interacts with a conserved region on TonB dependent transporter's N-terminal region called the TonB box and allows for energy transfer from TonB to the receptor (Peacock et al. 2007). The Proline rich spacer is located in the periplasmic space and contains a series of proline-glutamine and proline-lysine repeats (Kohler et al. 2010) This series of repeats is believed to provide rigidity which supports the proteins as it extends into the periplasmic space (Krewulak et al. 2007). ExbB and ExbD are in complex with TonB, these cytoplasmic proteins are around 26 kDa for ExbB and 17 kDa for ExbD (Ollis and Postle, 2012). ExbB and ExbD are located in the cytoplasmic membrane and work together using

PMF to create energy that is then transduced to the TonB dependent via TonB to allow for active transport (Ahmer et al. 1995).

Heme/ Hemin/ Hemeglobin and leghemoglobin

Heme (Fig. 2) is described as an iron protoporphyrin IX ring, and is one of the most abundant and widely used metalloporphyrins in biological systems (Poulos 2014). Heme serves many biological functions from electron shuttling, to storage and transport of oxygen when coupled with globulin proteins (Poulos 2014).

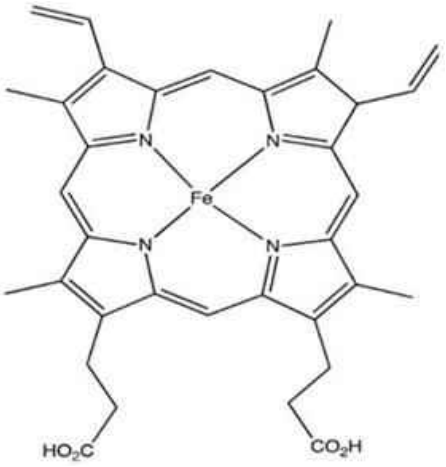


Figure 2. Heme

Hemin is similar to heme but with a small change, Hemin (Fig. 3) is a porphyrin ring IX containing a chloride molecule attached to a central ferric iron molecule (Hans 1941). Hemin is prepared in a lyophilized form from defibrinated blood that is treated with glacial acetic acid and sodium chloride at a temperature of 100 degrees Celsius (Hans 1941). Both heme and hemin are useable sources or iron for bacteria when cells are under iron poor conditions (Noya et al. 1997)

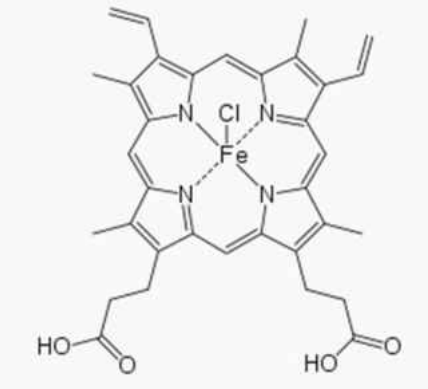


Figure 3. Hemin

Hemoglobin (Fig. 4) is a globular protein in red blood cells of vertebrates that is responsible for transport of oxygen throughout the body. Hemoglobin consists of four tightly packed globular proteins with a combined size of 65 kDa (Anthea et al. 1993). Each globular protein contains one heme molecule which intern contains iron, making the total of four iron molecules in hemoglobin (Anthea et al. 1993). Hemoglobin is a primary target for pathogenic bacteria to satisfy their iron requirements, this is due to hemoglobin being the most wide spread source of iron in the human body (Contreras et al. 2014). Hemoglobin has also been shown to be a source of iron for nonpathogenic bacteria such as Rhizobia bacteria

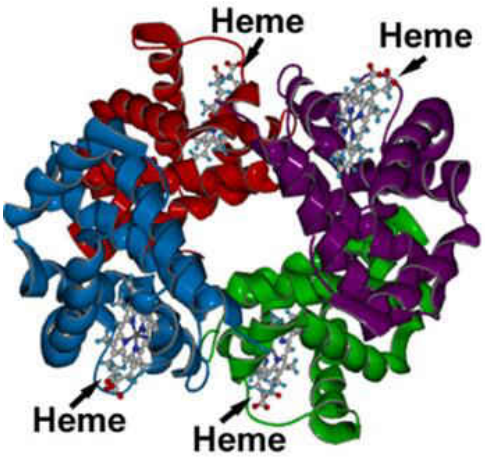


Figure 4. Hemoglobin

Another iron source that is available for both plant pathogens and symbiotes is leghemoglobin (Fig 5.), it is a myoglobin like protein with a size of 16 kilodaltons (kDa) (Becana et. al, 1995). This heme containing protein is found at concentrations of around 3mM in root nodules and plays a role in symbiotic nitrogen fixation by producing a low oxygen environment (Bergersen and Appleby 1981; Becana et al. 1995; Brear et al. 2013). Leghemoglobin consists of a globulin polypeptide produced by the plant and a heme group synthesized by the symbiotic bacteroids (Becana et al. 1995). Once both the heme and globulin components are produced they are assembled in the infected plant cells cytoplasm (Becana et al. 1995). Leghemoglobin is crucial for maintaining the crucial balance of free oxygen levels in the root nodule to allow bacteroid growth and nitrogenase function (Ott et al. 2005). Plant roots that are not infected with symbiotic rhizobium species lack the presence of leghemoglobin, this is due to leghemoglobin is needed for symbiotic nitrogen fixation and not normal plant growth (Ott et al. 2005).

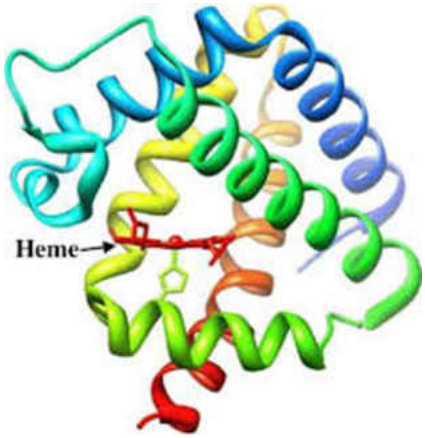


Figure 5. Leghemoglobin

Present Work

Previous work in our lab on *R. leguminosarum* ATCC 14479 has shown that it contains the heme/ hemin uptake system *HmuPSTUV* (Shushant 2017) . This system is believed to give

the bacteria the ability to utilize hemin and heme containing compounds under iron limiting conditions (Wexler et al. 2001). This system in other *Rhizobium* species has been shown to be a TonB dependent transport system (Wexler et al. 2001). The system is an ABC transporter that is composed of a predicted heme receptor (*HmuP*), Heme degrading enzyme (*HmuS*), periplasmic heme binding protein (*HmuT*), periplasmic heme permease (*HmuU*), and a heme ATPase (*HmuV*).

The goal of this work is to investigate the role of the role of the *HmuPSTUV* operon in hemin mediated iron uptake using site directed mutagenesis and other biochemical methods. We hypothesize the mutation of the *HmuPSTUV* operon will disrupt hemin mediated transport. Previous and current work has shown that *R. leguminosarum* ATCC 14479 is capable of using different types of heme compounds as a sole source of iron. The growth of the wild type was found to increase as the concentration of hemin increased, with growth plateauing at 15µM to 30µM. In anerobic conditions The addition of hemin to the media was found to increase cell survivability when compared to the iron free condition.

This Work also includes the investigation of the role of a predicted Two component system role in the iron starvation response. We hypothesize that mutation of this predicted two component system will result in diminished gene expression in iron scavenging systems. *In silico* analysis revealed that this predicted sensor histidine kinase belongs to a HWE family of sensor kinases that are believed to be involved with sensing a wide array of environmental changes.

CHAPTER 2. MATERIALS AND METHODS

Bacterial Strains and Growth

The strains of bacteria used in this study are *R. leguminosarum* ATCC 14479 *trifoli,* which was acquired from American Type Culture Collection; The *E. coli* strains used in this study are NEBα5 and PRK2013.

Table 1. Strains and Plasmids

Strain	Characteristics	Reference/ source
E. coli		
NEBα5	Used for transformation of suicide vector containing knocked out gene of interest	New England Biolabs
DHα5	Used for triparental mating as a helper strain to Kan^R	
R. leguminosarum		
ATCC 14479	Wild-type strain	ATCC
Plasmids		
PRK2013	Helper plasmid used for tri parental conjugation Kan^R	Ditta et al. 1980
pEX18GM	Suicide vector Gm^R	Hoang et al. 1998
pEX18 Δ*HmuU*	pEX18Gm plasmid carrying *HmuU* SOE product with an in-frame deletion of *HmuU* gene	This work

pEX18Δ*HmuV*	pEX18Gm plasmid carrying *HmuV* SOE product with an in-frame deletion of *HmuV* gene	This work
pEX18ΔSK	pEX18Gm plasmid carrying senor histidine kinase SOE product with an in-frame deletion of the sensor kinase gene	This work

Growth Conditions

A variety of media were used for culturing *R. leguminosarum* ATCC 14479 both enriched and minimal, the enriched media used were Congo Red agar (CR) and Yeast Mannitol Broth (YMB). The minimal medium used for iron limited *Rhizobium* culturing is Modified Manhart and Wong broth (MMW) (Manhart and Wong 1979). *Rhizobium* was grown at 30°C under either static conditions or at 250 rpm when grown in broth culture. The *E. coli* strains were grown in LB medium at 37°C under static conditions or at 250 rpm in broth culture.

CR agar is used for culturing and identifying *Rhizobium* species, this is due to the presence of Congo red dye. The dye is only absorbed by fast growing organisms resulting in a red to pink colored colonies, while *Rhizobia* species grow too slow to absorb the dye and colonies maintain a white color (Kneen and larue 1983). CR agar's composition is (W/V) 1% mannitol, 0.05% K_2HPO_4, 0.02% $MgSO_4*7H_2O$, 0.01% NaCl, 0.1% yeast extract, 2.5 x 10^{-5} % Congo red dye. Once the ingredients were combined and water added to the medium, the pH was adjusted to 6.8 by addition of an aqueous HCl solution. Once the pH was adjusted a 1% (W/V) aqueous solution of Congo red dye was added prior to autoclaving.

Yeast mannitol broth was used for culturing *R. leguminosarum* when iron concentration was not of importance. YMB's contains the following (W/V): 1% mannitol, 0.05% K_2HPO_4, 0.02% $MgSO_4*7H_2O$, 0.01% NaCl, and 0.1% yeast extract. Once the ingredients were combined and water added, the medium's pH was adjusted to 6.8 by addition of an aqueous NaOH solution. Once the correct pH was obtained and the ingredients were mixed the solution was autoclaved.

Modified Manhart and Wong (MMW) medium was used for culturing *R. leguminosarum* whenever iron limited conditions were required (Manhart Wong 1979). MMW media has two parts to the media one being the basal media and the other a concentrated vitamin solution (See appendix A). The vitamin solution consists of trace amounts of vitamins and metals it was concentrated at 1000 times. Once the vitamin solution was mixed it was filter sterilized and stored at 4°C and wrapped in aluminum foil to protect from light exposure. The basal media composition is as follows (W/V): 0.0764% K_2HPO_4, 0.1% KH_2PO_4, 0.15% Glutamate, 0.018% $MgSO_4*7H_2O$, 0.013% $CaSO_4$, and 0.6% dextrose. Prior to mixing the ingredients all glassware was washed with concentrated nitric acid for 1 hour in a fume hood to removed residual iron, then the glassware was washed four times with Millipore water to remove residual nitric acid. Once the ingredients were dissolved in water the pH was adjusted to 6.8 by adding an aqueous solution of 5M NaOH. After autoclaving the medium, it was allowed to cool to room temperature before adding 0.001% (V/V) of the filter sterilized concentrated vitamin solution.

Chrome Azurol S is a medium used to test for the production of siderophores and iron chelation. The media contained (W/V): 3.0% MOPS, 0.05% NaCl, 0.03% K_2HPO_4, 0.015% NH_4Cl, 0.5% L-aspargine. The ingredients were mixed in Millipore H_2O and 5M NaOH was

added till the media's pH was 6.8. The CAS indicator dye (Appendix A) and 50% sucrose solution were added post autoclaving a mixed into solution.

The LB broth used for culturing *E. coli* species contained (W/V): 1% tryptone, 0.5% yeast extract, 0.5% NaCl, and 1.5% agar for the plates. Once the ingredients were mixed with water the pH was adjusted to 6.8 by the addition of an aqueous solution of HCl, then autoclaved.

The concentration of antibiotics used in media for both *R. leguminosarum* and *E.coli* strains are as follows:

Table 2. Antibiotic concentrations

Antibiotic	Concentration in media
Nalidixic acid (Na_{15})	15 µg/ml
Penicillin G (Pen_{50})	50 µg/ml
Gentamycin (Gm_{20})	20 µg/ml
Kanamycin (Kan_{35})	35 µg/ml

Hemin supplied growth

As previously stated MMW media was used when iron concentration was being limited and supplemented. The media was made iron free by the addition of the iron chelator 2,2'-dipyridyl to a concentration of 200µM. The iron free media was supplemented with hemin at varying concentrations: 0.0µM, 1.0µM, 1.5µM, 3.0µM, 10µM, 15µM, and 30µM. The hemin supplemented MMW media was inoculated with $1.94*10^7$ CFUs (colony forming units) from a 72 hour grown MMW iron starved culture. For aerobic conditions cultures were grown at 30°C at 250 rpm. For anaerobic conditions the cultures were grown at 30°C under static conditions in

and aerobic jar with a GasPak. Serial dilutions for both aerobic and anaerobic conditions were taken at hours 24, 48, 72, and 96 hours. Dilutions were plated on CR plates and allowed to grow for 72 hours before counting CFUs.

Genomic DNA Extraction

Rhizobium leguminosarum ATCC 14479 was first inoculated onto CR plate from a -80°C stock and grown at 30°C for 72 hours. A single colony was used to inoculate a 5ml YEM broth supplemented with penicillin G to inhibit the growth of possible contaminants. This culture was grown for 48 hours at 30°C on a shaker at 250 rpm. Two milliliters of the culture was spun at 10,000 x g, the supernatant was removed and the pellet was resuspended in 2ml of a sterile 0.85% NaCl solution to remove the exopolysaccharide. Then the genome was collected using the Qiagen DNeasy UltraClean microbial kit, this kit works by lysing the cells by the bead-beating method. The resulting lysate is treated to precipitate out the proteins while leaving the DNA soluble, the proteins are pelleted via centrifugation at >10,000 x g. The resulting lysate is then collected using a column which binds the genomic DNA, where it undergoes a series of ethanol washes. The cleaned DNA is eluted using 10 mM Tris-HCl at a pH of 8.5. The DNA is stored at -80°C for long term storage and -20°C for short term storage.

Splicing by Overlap Extension (SOE)

Knockout constructs for *HmuU, HmuV*, and the Sensor Histidine Kinase were created using the splicing by overlap extension (SOE) method (Figure 6). This method involves the PCR amplification of around 1000 base pairs of both the 3' and 5' flanking regions of the gene of interest. Each of these flanking regions contain a portion of the gene of interest, while the central portion of the gene of interest is removed. Each of these fragments are first amplified separately,

then purified by gel extraction. Once purified the flanking regions are joined together through a subsequent round of PCR due to complementary sequence attached to the primers, resulting in a knockout construct that is indicated by the Δ symbol.

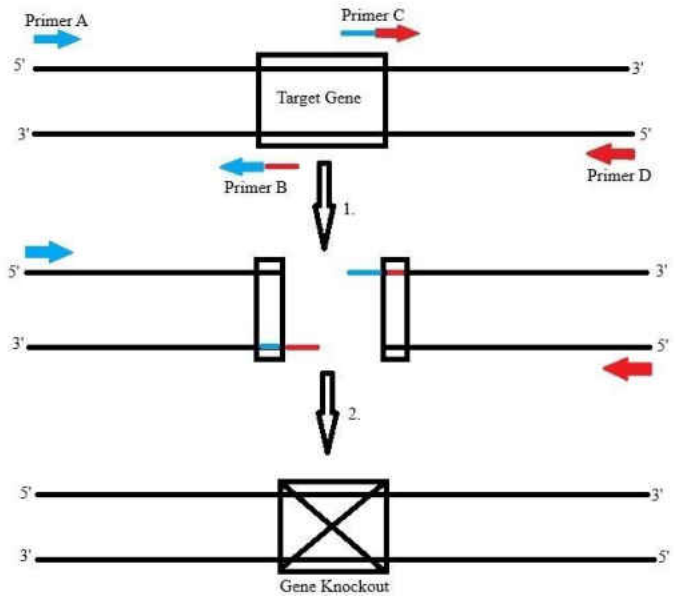


Figure 6. Illustration of SOE method, step one involves producing 5' and 3' fragments through PCR. Step two joins these two fragments through another round of PCR, resulting in deleting the interion portion of the gene.

The construction of the 5' fragment of the *HmuU* knockout used the primers SOEFA (primer A) and SOERHU (Primer B), and SOEFHU (primer C) and SOERA (primer C) were used to amplify the 3' fragment. For the 5' fragment of the *HmuV* knockout used the primers SOEVF and SOEHmuVR and SOEHmuVF and SOEVR for the 3' fragment. For the Sensor Kinase knockout, the 5' fragment was amplified with the primers SOEFB and SOER2SK, and for 3' fragment SOERB and SOEF2SK were used. The Primer sequences and their binding sites are illustrated on a nucleotide map in appendix B.

pEX18 Suicide Vector

The knockout constructs created using SOE were digested by restriction enzymes and ligated into the $pEX18_{GM}$ suicide vector (Figure 7.). This suicide vector contains an origin of replication, gentamycin resistance selectable marker, *SacB* counter selectable marker, and a multiple cloning site (Hoang et al. 1998). The counter selectable *SacB* gene when expressed in the presence of sucrose containing media is lethal to Gram negative cells. The plasmid also contains the lacZα gene, allowing for quick blue-white screening of transformed mutant vectors. Once the knockout constructs were ligated into the restriction digested multiple cloning site (Figure 8), the pEX18 vector was then transformed into the NEBα5 *E. coli* cell line via rubidium chloride heat shock method. Once transformed the cells were plated onto LB_{Gm} agar that contained X-gal (20µg/ml of media) to allow for blue white screening. The cells were allowed to grow overnight at 37°C, white colonies were tested by colony PCR for the presence of the mutant construct and SacB gene. The primers used to test for the mutant construct were designed

to amplify the full SOE crossover product; the amplified products (~2500bp) were purified and sent off for sequencing to confirm the presence of a mutant construct.

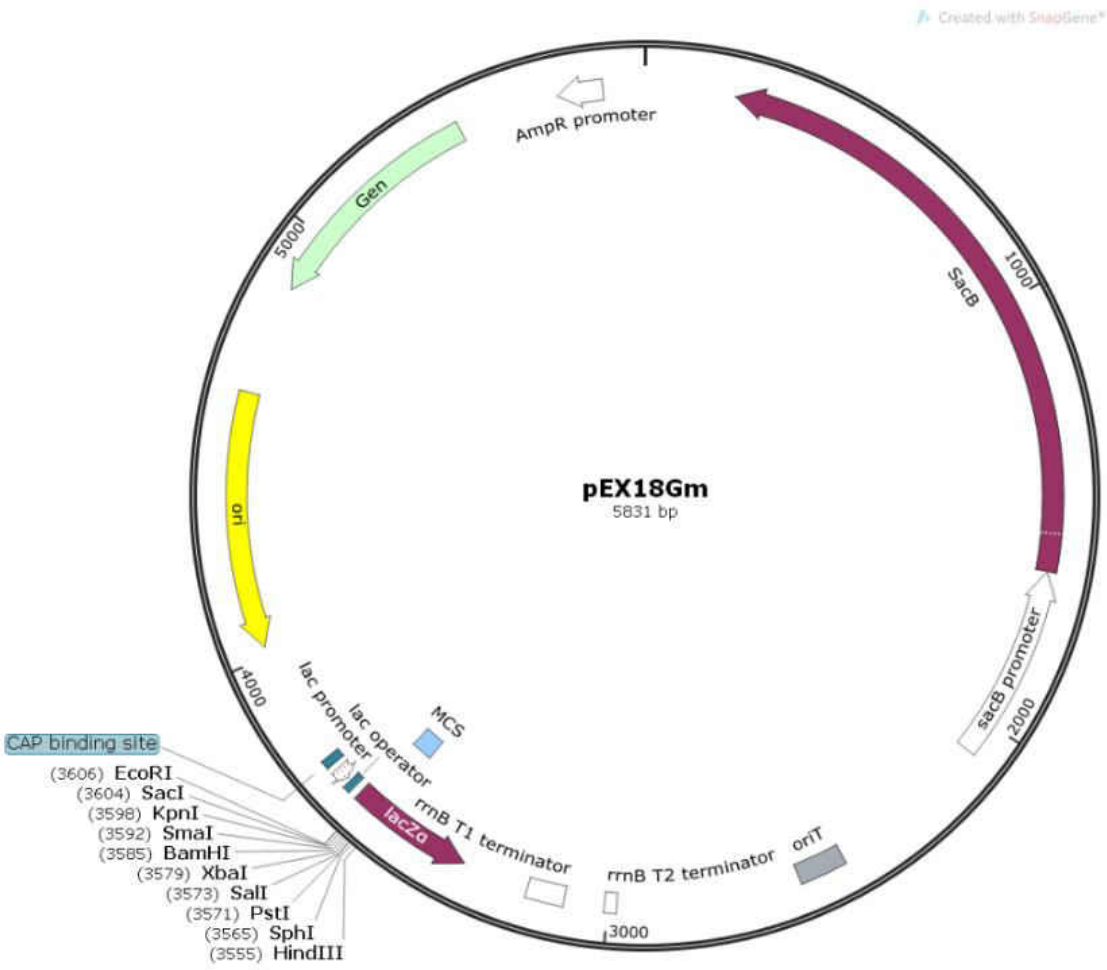


Figure 7. pEX18$_{Gm}$ suicide vector map

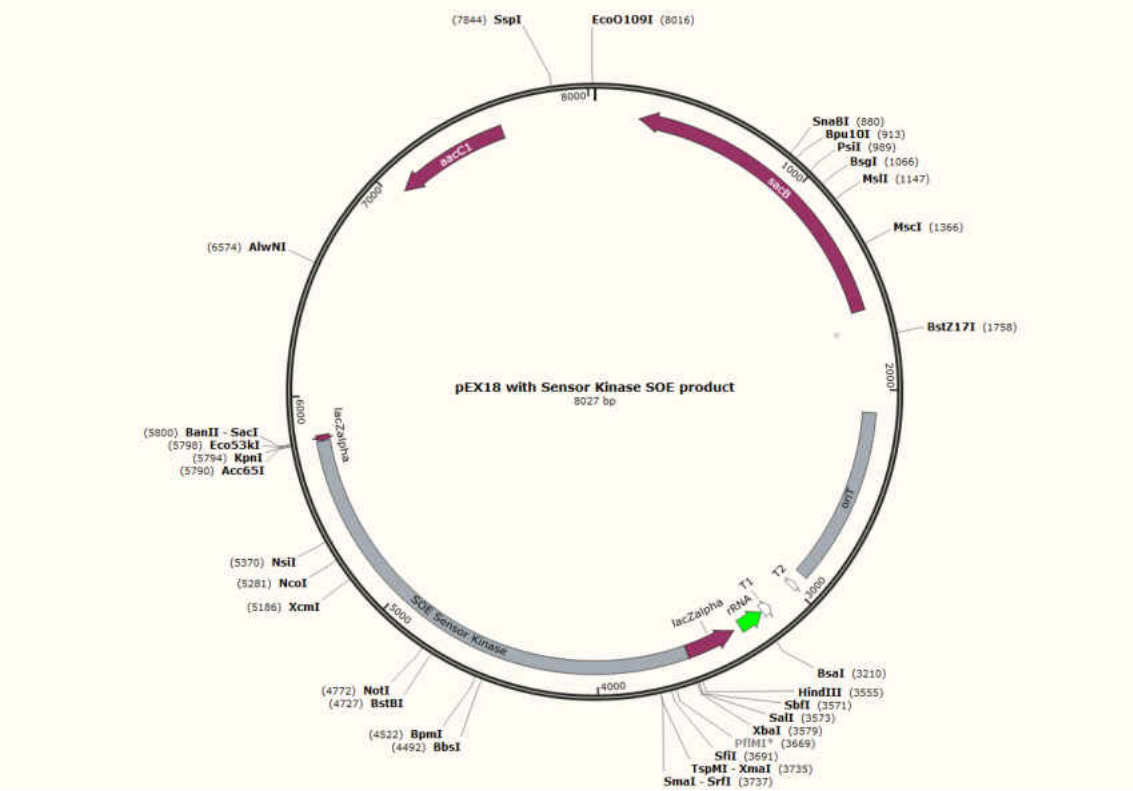


Figure 8. pEX18$_{GM}$ suicide vector with Sensor kinase crossover product

NEBα5 Competent Cells

NEBα5 *E. coli* cells were made competent using the rubidium chloride method. NEBα5 cells were grown in 5ml of LB broth at 37°C at 250 rpm overnight, this fresh culture was used to inoculate 2X YT broth. This culture was allowed to grow until it reached an OD_{600} of 0.6, once this optical density was achieved the cells were collected at 5000x at 4°C for 10 min.. The cells were then resuspended in chilled TFB1 (appendix A) and let sit for 5min. in an ice bath, and centrifuged using the same conditions. The supernatant was removed and the cells were resuspended in chilled TFB2 (appendix A) and incubated for 15-60 min. on wet ice. Cells are then aliquoted out and stored at -80°C until used for heat shock transformation.

Heat Shock Transformation of NEBα5

The rubidium chloride competent cells were used for heat shock transformation of the suicide vector containing one of three SOE products. The competent cells were thawed out and 50µl of cells were mixed with 0.3µl of the SOE pEX18 ligated product. This mixture was incubated on ice for 30 min., then they were placed in a 42°C water bath for 30 seconds. Cells were then recovered in 1ml of room temperature LB broth for 1 hour at 37°C at 250rpm. After recovery the cells were plated onto X-gal containing LB_{Gm} agar plates and incubated overnight. White colonies were then used for colony PCR to confirm for presence of the SOE product.

Conjugation

NEBα5 cell lines that were confirmed to contain the suicide vector with a SOE knockout construct with one of the tree genes, were then used for triparental conjugation. Triparental conjugation involves three cell lines, the first is the donor cells containing the plasmid of interest. The second is the helper cells which contain the sex pilus and *tra* genes for plasmid transfer. Third is the recipient cell line which receive the plasmid of interest. The donor cell line is the NEBα5 cells containing the suicide vector with one of three mentioned SOE knockout constructs. The Helper strain *E. coli* PRK2013, is a cell line with the PRK2013 plasmid which carries the sex pilus, *tra* genes, and kanamycin resistance (Ditta et al. 1980). The recipient cell line used is *R. leguminosarum* ATCC 14479. The helper and donor strains were grown overnight at 37°C on LB agar plates containing the required antibiotic to maintain plasmid selection. A couple of colonies were scooped with a inoculating loop and placed into a microcentrifuge tube. These cells were gently resuspended in 1 ml of a 0.85% NaCl solution and then spun at 5000xg, this step was repeated twice to ensure residual antibiotics were removed. Once washed the cells

were resuspended in 200µl of 0.85% NaCl solution. The *R. leguminosarum* cells were grown for 48 hours in YEM broth at 30°C. 2 ml of cells were spun at 10000xg then washed with 0.85% NaCl and re-spun, this step was repeated twice to remove the exopolysaccharide. The washed cells were then combined at a ratio of 1:5:1 (donor: recipient: helper), then spun at 5000xg and resuspended in 50µl of 0.85% NaCl. The mix of cells was then pipetted onto nitrocellulose discs on CR agar plates and incubated at 30°C. The discs containing the cells were taken at 12, 24, 48, and 72 hours and resuspended in 1 ml of 0.85% NaCl solution by vortexing vigorously. Once removed from the disc cells were serial diluted and plated onto CR agar plates containing gentamycin and nalidixic acid and incubated at 30°C until cells appeared. The Gentamycin selects for *R. leguminosarum* cells that acquired the suicide vector containing the SOE construct, while the nalidixic acid selects against the *E. coli* strains. Once colonies appeared they were tested by colony PCR to identify merodiploids. Once identified cells were grown in 5ml of YEM broth at 30°C at 250 rpm for 6 to 48 hours to allow for homologous recombination. Cells were then serial diluted and plated on CR plates containing 5% sucrose (W/V) and incubated at 30°C until colonies appeared. Homologous recombination allows for the cell to either incorporate the mutant allele or expel it. Cells that have undergone homologous recombination contains either the wild type or the mutant allele and no longer contain the pEX18 plasmid. This results in homologous recombinants surviving on the $CR_{suc5\%}$ media since they no longer contain the *sacB* gene. The surviving colonies are tested by colony PCR to identify potential mutants. The potential mutants then have their genome extracted to serve as a template for PCR using primers that amplify the flanking regions of the gene of interest and undergo sequencing to confirm a knockout of a respective gene.

Bioinformatic Tools

NCBI/BLAST program was used to analyze and align DNA and protein sequences. Sequencing data from SOE products were analyzed using Chromas software, once analyzed the DNA sequence was aligned with *R. leguminosarum* ATCC 14479 genome using NCBI/BAST program to confirm. The DNA sequence of the histidine sensor kinase was also analyzed through NCBI/ BLAST to look for homologous sequences in other closely related species to identify potential functions.

DNA sequences of the *Hmu* operon and sensor histidine kinase were analyzed through the EXPASY translation tool to obtain protein sequences. Once the protein sequences were obtained they were used to obtain protein models using SWISS-MODEL program. For the sensor histidine kinase and response regulator models, DNA, and amino acid sequences were compared to *Xanthomonas campestris VgrRS* system.

Outer membrane Protein Extraction

R. leguminosarum ATCC 14479 cells were grown in 1 liter MMW broth containing 200µM 2,2'-dipyridyl. The cells were incubated at 30°C at 250 rpm for 72 hours. Cells were then collected by centrifuging cultures at 10000xg at 4°C, cells are then resuspended in 150ml of chilled 0.85% NaCl solution to remove exopolysaccharide. This NaCl wash procedure is repeated four times, once the exopolysaccharide is removed the cells are suspended in buffer A and sonicated by a Branson digital sonicator at 1 minute pulse and 1 minute pause at a 80% amplitude. Once sonicated the lysed cells were centrifuged for 10 minutes at 10000xg to collect the supernatant. The supernatant was subjected to ultra-centrifugation at 30000xg for 90 minutes at 4°C, the supernatant is removed and saved. The resulting outer membrane pellet is

homogenized in buffer B and subjected to ultra-centrifugation again, this collection and centrifugation process is repeated again for buffer D. The resulting pellet in the final centrifugation step with buffer D is collected and homogenized in 2ml of buffer D and stored at 4°C until analysis of outer membrane proteins by SDS-PAGE.

SDS-PAGE

The outer membrane profiles are analyzed by using Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) (Laemmli 1970). Protein samples were mixed with 2X loading dye and were kept at 95°C for 5 min. for the proteins to denature. Once denatured 20µl of protein were loaded on the SDS-PAGE gel, the gel was ran for 75 min. at 30mA and 120V for one gel and 60mA and 120V for 2 gels. The gel was then stained with Coomassie Brilliant Blue and de-stained with a de-staining solutions.

Hemin-Agarose affinity chromatography

The outer membrane proteins extracted in the 2,2'-dipyridyl condition were used in the hemin-agarose linked chromatography described by Battistoni et al. (1992) and Lee (1992). This is an affinity based chromatography which looks for hemin binding proteins. The outer membranes extracted from 10^{10} cells were resuspended in 500µl of a 50mM Tris-HCL (pH 8.0) and 1M NaCl solution and incubated for 1 hr. at 30°C with 250µl of the hemin-agarose suspension. Nonadherent proteins were removed by incubation for 90 min. at 30°C with a solution containing: 10mM Tris-HCL (pH 7.8), 150mM NaCl, 10mM EDTA, 1% *N*-laurylsarcosine, (W/V) and 0.1% SDS (W/V). The suspension was then centrifuged for 5 min at 5000 xg, the supernatant is removed and used for SDS-PAGE analysis. The hemin linked agarose beads are then washed three times with a 50mM Tris-HCL (pH 8.0), 1M NaCl, 10mM

EDTA,0.75% (W/V) *N*-laurylsarcosine, and 0.075% SDS (W/V) solution. The hemin linked agarose was then washed once with a 50mM Tris-HCL (pH 8.0) and 1M NaCl solution. After each wash the hemin agarose beads were centrifuged for 5 min. at 5000 xg, finally the washed beads were suspended in 200µl of PBS and 200µl of the 2X Laemmeli buffer. This suspension was heated at 100°C for 5 min, the suspension was placed on ice for 1 min. and the agarose was pelleted by centrifugation for 5 min. at 750 xg. The supernatant was placed into a new collection tube and 10-20µl was used for SDS-PAGE analysis.

Detection of Siderophores

During the hemin supplemented MMW media growth experiment, the supernatant of select conditions were tested for siderophore production. The supernatants were tested for siderophore production by incubating the supernatant in Chrome azurol S plates (CAS). These plates contain a dye that is a blueish green when iron is bound, and have an orange color when iron is unbound. If Vicibactin is present, it will strip the iron from the dye producing and orangish/ brown halo. The conditions tested for siderophore production were: 200µM 2'2-dipyridyl, 10µM, 15µM, and 30µM. The supernatants were collected at 24, 48, and 72 hours. Each culture had 150µl collected, then centrifuged at 10,000 xg's at 4°C for 5 min. Then 100µl of the cell free supernatant was loaded into the well of a CAS plate, then this was allowed to incubate at 30° for 48 hours. After incubation any resulting halos were measured.

CHAPTER 3. RESULTS AND DISCUSSIONS

Based on the previous work in our lab and getting the genome of *R. leguminosarum* ATCC 14479, we wanted to explore the effects of hemin as an alternate iron source. Studies in *Bradyrhizobium* and other strains of *R. leguminosarum* have shown that they were capable of utilizing different heme compounds in vitro (Nienaber et al. 2001 & Wexler et al. 2001). Due to the availability of the chromosomal DNA for *R. leguminosarum* ATCC 14479, it was possible to perform *in silico* analysis on genes predicted to be involved with hemin utilization. Once the putative gene cluster was identified, an effort was made to investigate hemin dependent iron utilization as well as the role these putative genes play in hemin uptake and utilization

Hemin Supplied Growth

To determine the effects of hemin as the sole iron source on the growth of *R. leguminosarum* ATCC 14479, an iron depleted culture (containing 2',2'-dipridyl) was used to inoculate varying concentrations of hemin supplemented MMW broth. The growth of the cultures was monitored over 96 hours by plating 10 fold dilutions onto CR agar plates, the colony forming units were counted and log transformed to graph the growth curve (Figure 9). The iron depleted media (containing 2',2'-dipridyl) had the lowest growth as predicted, and a concertation dependent increase in growth was observed in the case of hemin supplemented media. The 1.0µM hemin (gray line) had the lowest growth for hemin supplied media, while 30µM hemin (red line) appeared to have the highest observed growth. The difference between the 15µM and 30µM hemin conditions was minimal (Figure 9 & 10), indicating the cell's iron requirements are being met around these concentrations. There was a full $\log_{10}$ difference between the 2',2'-dipridyl and 30µM hemin growth conditions, this difference in growth and

concentration dependent growth increase shows *R. leguminosarum* is capable of using hemin as a source of iron.

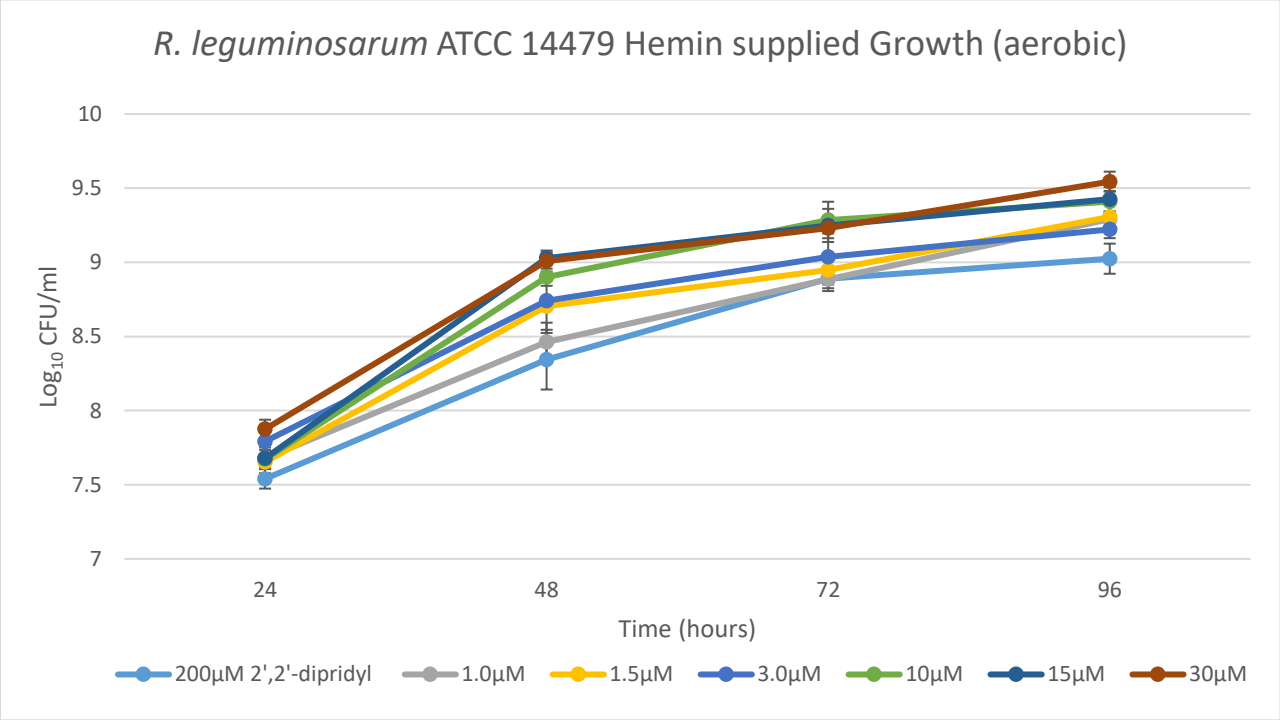


Figure 9. *R. leguminosarum* ATCC 14479 aerobic growth curve of iron free and hemin supplied MMW media, error bars represent standard deviation. (n=3)

This growth study was then repeated but at the start of the 24 hour period the CFUs were measured every four hours up until the 48 hour period. This region was of interest because in the initial aerobic growth study it showed the best dose dependent relationship. The results of this 24-48 hour growth period shows a dose dependent relationship that remains constant over the 24 hour period (Figure 10.). This time point also maintains a full log deference in CFUs when comparing the hemin supplemented to the iron deplete (2',2'-dipridyl) media. Since the difference between the 15µM and 30µM hemin supplemented media growth was minimal, we believe the cells are reaching sufficient intracellular iron levels. The means by which the hemin

is being used could be by the predicted *Hmu* operon or by the siderophore vicibactin striping the iron from hemin. Another potential increase in growth could be due to the natural degradation of the hemin over time releasing the complexed iron, allowing the many indirect iron acquisition systems to uptake the freed iron.

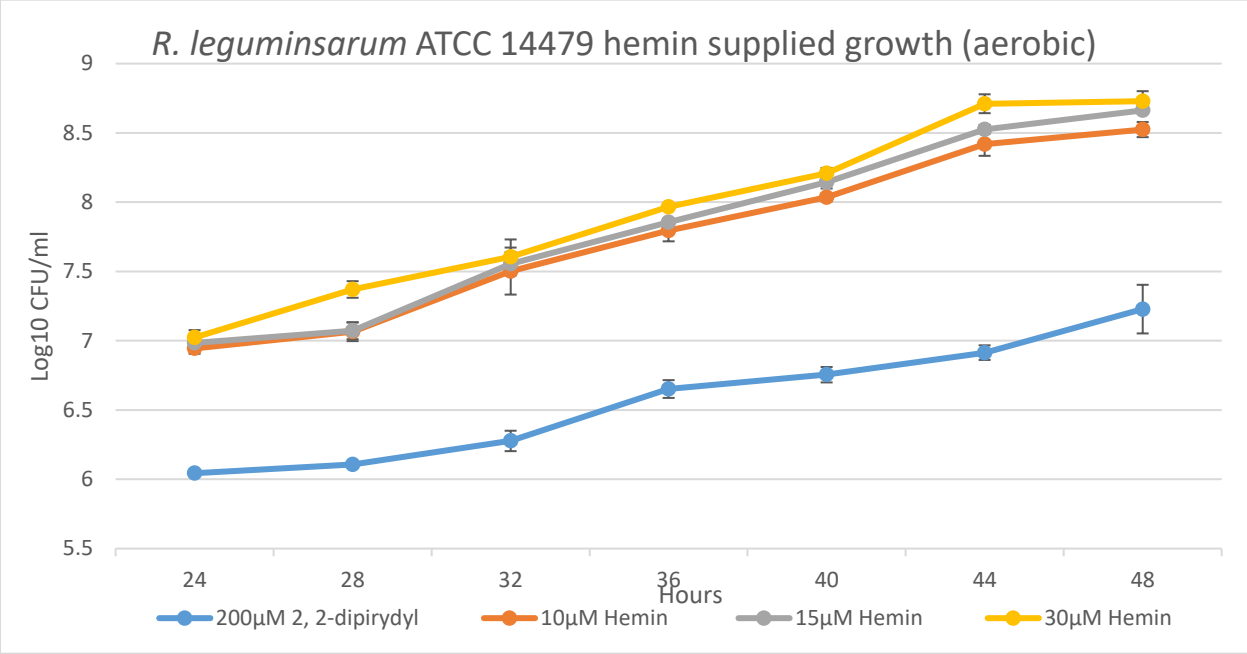


Figure 10. R. leguminosarum ATCC 14479 aerobic growth curve of iron free and hemin supplemented MMW media, error bars represent standard deviation. (N=3)

Since *R. leguminosarum* encounters heme compounds in the low oxygen environment of the root nodule, the effects of hemin supplied media on *R. leguminosarum* survival in an anaerobic environment were measured. The survival of the *R. leguminosarum* was tracked over 96 hours in an anaerobic environment, only the higher concentrations of hemin were tested due to their increased effect on growth in aerobic conditions. The survival was measured in CFUs then log transformed to graph the survival curve (Figure 11). The 2',2'-dipridyl supplemented media (yellow line) had the least impact on maintaining survival of the cell culture, while the

addition of hemin to the media results in drastic increase in cell survival. The highest increase in cell survival is observed at the 72-96 hour mark, at 72 hours all the hemin supplemented media had over a 1.5 $\log_{10}$ increase in cell survival compared to the 2',2'-dipridyl supplemented MMW media. The largest margin of cell survival is when comparing the hemin and 2',2'-dipyridyl supplemented media at the 96 hour point. There is a 2 $\log_{10}$ increase in survival in all the hemin supplemented media, compared to the 2',2'-dipridyl supplemented media. Since *R. leguminosarum* is an aerobic bacteria the cells should fail to grow in an anaerobic environment, this is supported by observing no growth trend in the anerobic environment. The reason for the prolonged cell survival in the hemin supplied media, is most likely due to the ferric iron in hemin being reduced to its ferrous form. This reduction of iron occurs due to the anaerobic environment (Perez-Guzman et al. 2010). The ferrous iron is then released from the hemin molecule, in turn making the iron readily available for cell use.

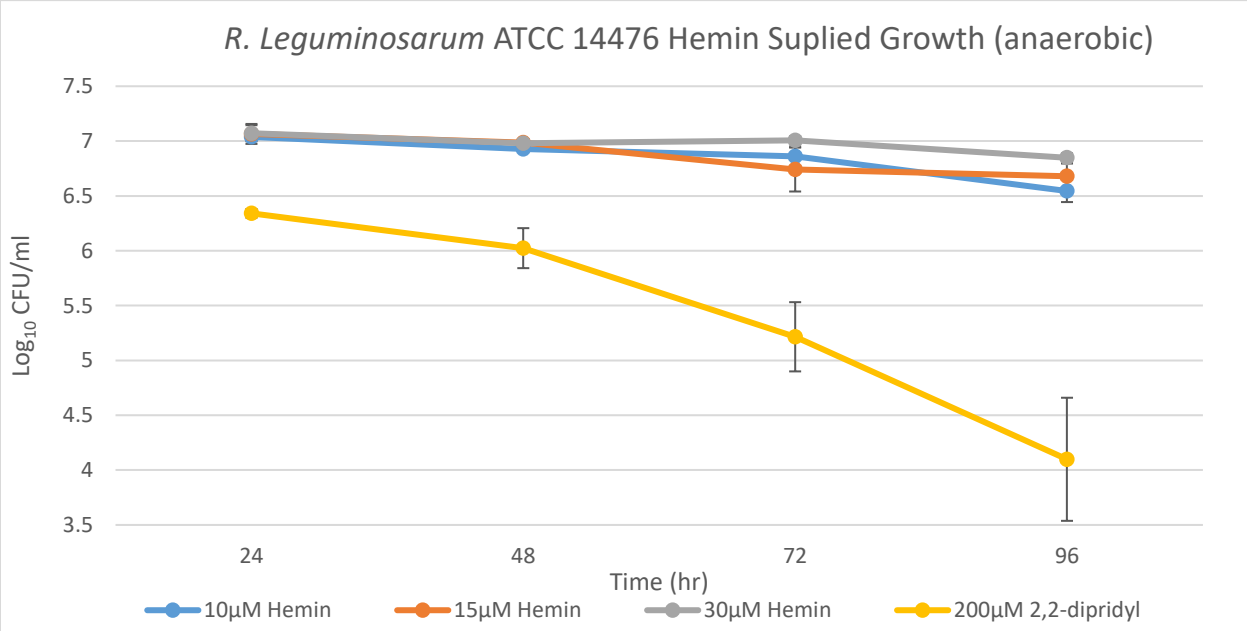


Figure 11. *R. leguminosarum* ATCC 14479 anaerobic survival curve of iron free and hemin supplied MMW media, error bars represent standard deviation (n=3).

Identification of Potential Mutants

After looking at the effects of hemin supplemented media had on the growth of *R. leguminosarum* ATCC 14479, we wanted to determine if the *HmuPSTUV* operon was responsible for hemin utilization. Splicing by Overlap Extension (SOE) was used to generate knockout genes for *HmuU*, *HmuV,* and a sensor histidine kinase. Both *HmuU* and *HmuV* were chosen for mutation due to their predicted function as an ABC transporter for heme compounds, but also due to their close proximity to each other in the operon. The sensor kinase was chosen due to recent research on TCS involvement in sensing iron both intracellularly and extracellularly (Wang et al. 2016). Once SOE knockouts were obtained, the plasmids containing the knockout gene constructs were confirmed by blue-white screening (appendix B), PCR analysis, and finally DNA sequencing. Blue-white screening was used on NEBα5 cells that had been transformed with the ligated pEX18 and SOE Product.

The *HmuU* SOE PCR fragments were first checked on an agarose gel (figure 12), to determine if the fragments matched up with their predicted sizes. The fragments were then sequenced to confirm data was obtained for the crossover product, the sequences were aligned with the genome of *R. leguminosarum* ATCC 14479 to confirm the crossover product was indeed the gene of interest. The 5' fragment was sequenced using SOEFA as the primer. The sequence showed a 98% nucleotide identity when aligned to the *R. leguminosarum* ATCC 14479 *Hmu* operon. The 3' fragment for the *HmuU* was sequenced using SOERA as the primer. The sequence showed a 99% nucleotide identity when aligned with *R. leguminosarum* ATCC 14479 *Hmu* operon. The SOE product was then restriction digested and ligated into the pEX18 suicide vector. This ligation product was then transformed into competent NEBα5 cells, the resulting transformant were then screened by blue-white screening on X-gal containing LB_{GM} media.

White colonies were then used for colony PCR to check for the SOE product, if the SOE product was found these cells were then used for triparental mating.

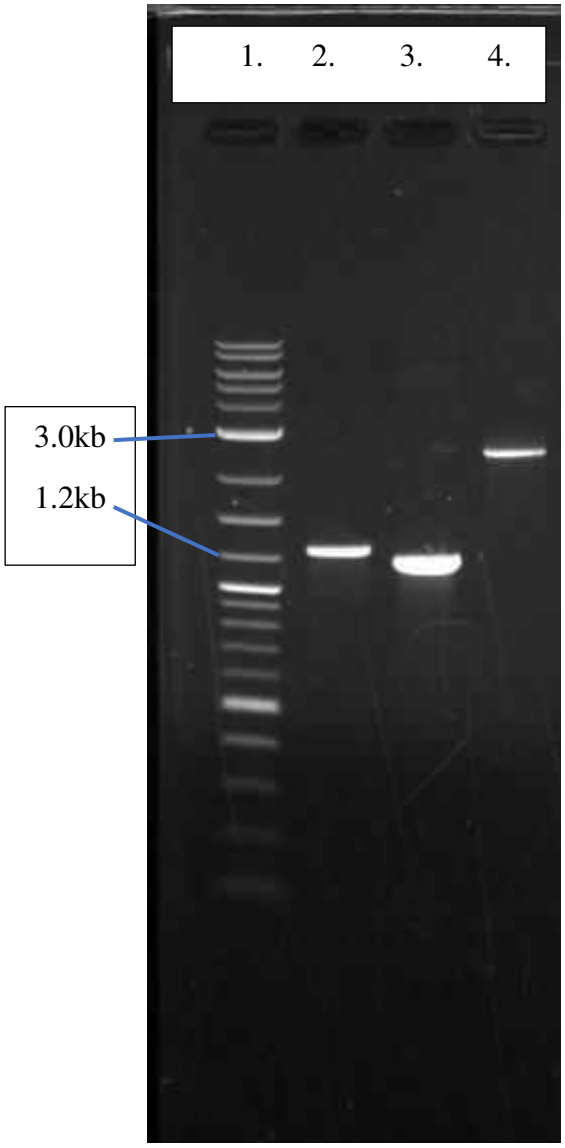


Figure 12. Gel of *Hmu*U SOE PCR products, Lanes: 1. 1kb ladder, 2. 5' fragment, 3. 3' fragment, 4. SOE crossover product. Sequence data for fragments in APPENDIX B

The PCR products for *HmuV* were checked on an agarose gel (Figure 13), then the products were sequenced. the sequence alignment was repeated for the *Hmu*V construct, only the 3' sequence was obtained (appendix B). This could be due to DNA sequencing issues, impurities in plasmid sample or primers. The primer used for the 3' sequence was SOEVR (appendix B), the sequence alignment with *R. leguminosarum* ATCC 14479 *Hmu* operon was a 97% nucleotide identity. The SOE product was then restriction digested and ligated into pEX18 suicide vector, the ligation product was transformed into competent NEBα5 cells. The resulting transformants underwent the same blue-white screening as the *HmuU* construct, and white colonies were confirmed to contain the *HmuV* SOE construct via colony PCR.

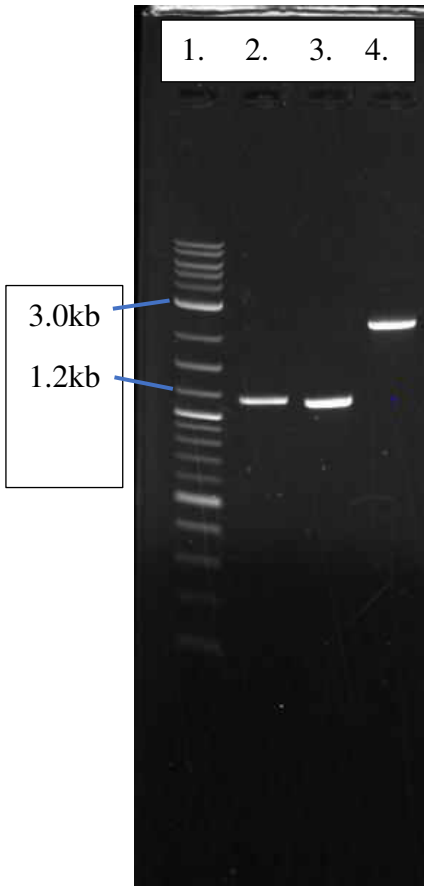


Figure 13. Gel of *Hmu*V SOE PCR products, Lanes: 1. 1kb ladder, 2. 5' fragment, 3. 3' fragment, 4. SOE crossover product. Sequence data for fragments in APPENDIX B

The SOE PCR products for the sensor kinase were first confirmed on an agarose gel (figure 14), then fragments were sequenced (appendix B) to determine if the SOE construct was the intended product. The 5' product was sequenced using the primer SOEF2SK, and SOER2SK for the 3' fragment. The 5' and 3'(appendix B) fragments sequence were aligned with *R. leguminosarum* ATCC 14479 genome. The SOE product underwent the same restriction digestion and ligation as the *Hmu* SOE products. The blue-white screening was repeated, with white colonies undergoing colony PCR to look for the presence of the senor kinase SOE product. Confirmed cell lines were then used for triparental mating.

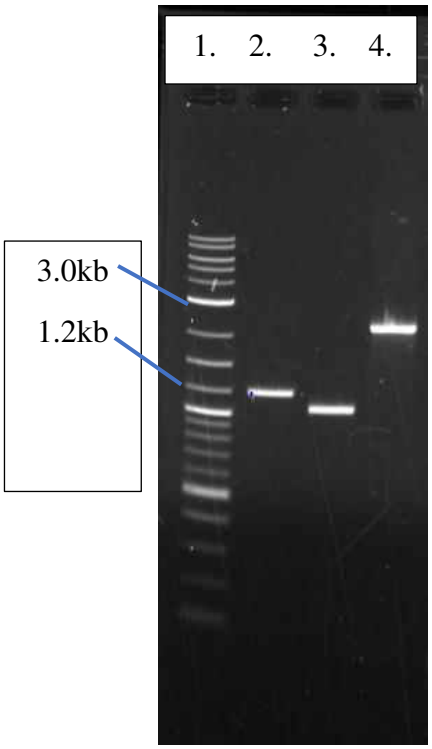


Figure 14. Gel of senor kinase SOE PCR products, Lanes: 1. 1kb ladder, 2. 5' fragment, 3. 3' fragment, 4. SOE crossover product. Sequence data for fragments in APPENDIX B

These confirmed NEBα5 cell lines were used for tri parental mating to obtain a *R. leguminosarum* merodiploids. Only one merodiploid was obtained for the three constructs, The sensor kinase was the only merodiploid obtained. The reason for unsuccessful conjugation could

be due to F pilus being damaged during the mixing stage. Another potential reason could be exopolysaccharide preventing the formation of the sex pilus. One potential merodiploid for the *Hmu*U gene was found out to be a PRK2013 *E. coli* helper strain that had integrated the Δ*Hmu*U pEX18 plasmid into its genome. This was determined by amplifying and sequencing the potential *HmuU* mutant's 16s gene sequence (Appendix B). The Sensor kinase merodiploid had its flanking regions analyzed, upon PCR analysis of the merodiploids flanking regions (Figure 15) showed the SOE product under went illegitimate recombination during the first crossover event. This result from the gel was further supported by having these fragments sequenced and seeing no integration in the genome with the deleted construct. Both the sequences of the -5' and 3' fragments showed no signs of a crossover event occurring in these regions (appendix B). This illegitimate recombination event in the sensor kinase merodiploid is most likely due to another region of homology between the SOE product flanking regions and the genome. The exact site of this recombination is unknown; however, Inverse PCR could be used to determine the integration site.

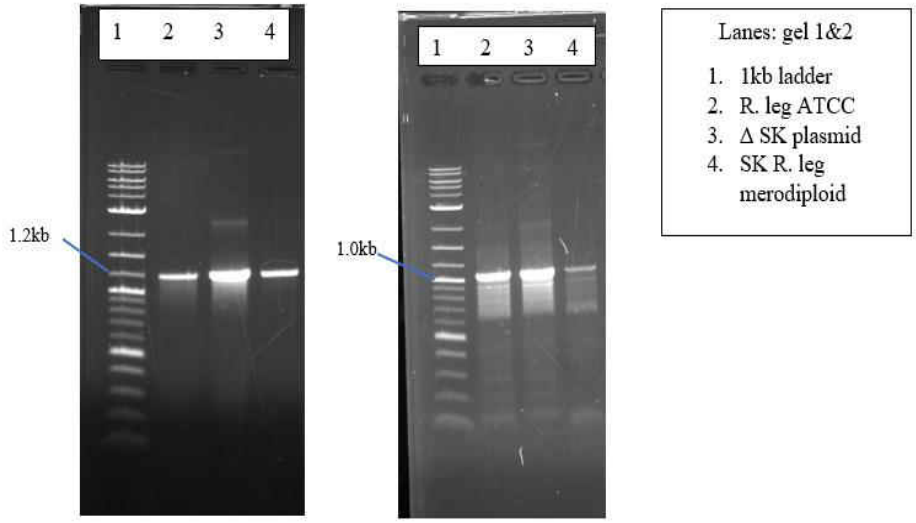


Figure 15. Left gel is a PCR of the 3' flanking region, the right gel is of the 5' flanking regions. 3' primers: SOEFB & SOERSK. 5' primers: SOERB & SOEF2SK

Sensor Histidine Kinase In silico Analysis

During sequencing analysis of the *Hmu* operon and the flanking regions both upstream and downstream, it was found that these flanking regions showed conserved genes. Upstream of the *Hmu* operon is a conserved TonB and downstream is a Two-component system (TCS) (figure 16.) Since this region of DNA had these conserved genes a Blast analysis (Appendix B) was performed on this region to compare sequence identity to related species.

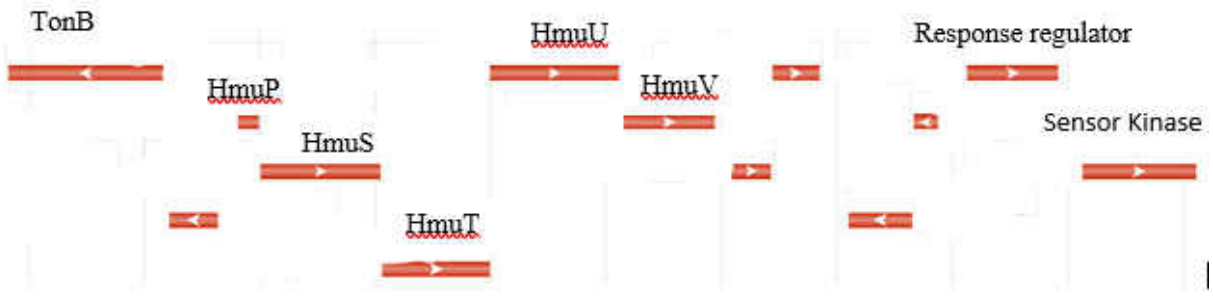


Figure 16. TonB and TCS near Hmu operon in *R. leguminosarum* ATCC 14479 genome. NCBI ref. seq. NZ_CP030760.1 (1,963,826-1,974,106). Unmarked proteins are hypothetical proteins with no assigned function

The Blast analyses showed high similarity to gene sequences in 26 other *R. leguminosarum* strains with high nucleotide identity (>90%). TonB has been shown to be involved in heme utilization, but the involvement of this two component system has not been investigated (Wexler et al. 2001). A newly described TCS in *Xanthomonas campestris* called the *VgrR/S* system, which was shown to be involved in sensing both intracellular and extracellular iron levels (Wang et al. 2016). Since this system was shown to be involved in the iron stress response, we compare the *VgrS* DNA and protein sequences to our uncharacterized sensor kinase. Aligning these protein sequences (Figure 17) shows very little homology except in one location, this location in the *VgrS* gene is the known phosphorylation site. The phosphorylation

site in *VgrS* is the histidine in the 186 amino acid position, and the predicted phosphorylation site in *R. leguminosarum* ATCC 14479 is the histidine in the 134 amino acid position. This region could be confirmed in future studies by site directed mutagenesis targeting this predicted phosphorylation site.

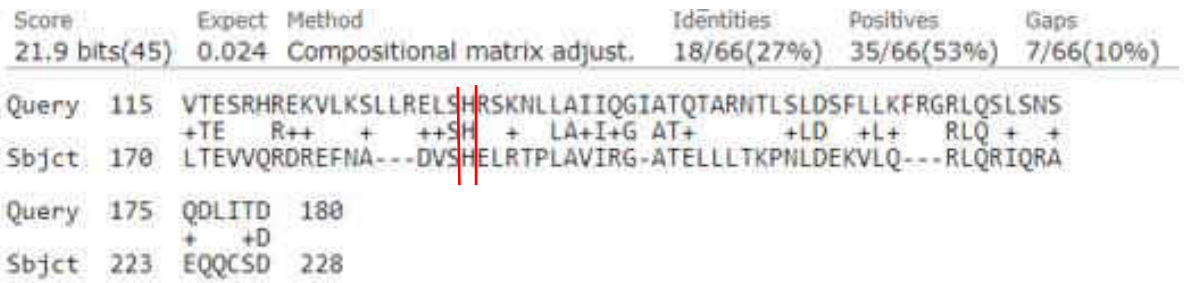


Figure 17. Alignment of *R. leguminosarum* sensor kinase (query) and *Xanthomonas campestris* (subject), the red lines indicate the phosphorylation site

Since the alignment with *VgrS* shows very little homology in the protein alignment, DNA and protein sequences of the sensor kinase was analyzed by NCBI BLAST and compared to other *rhizobium* species. This sensor kinase was found to belong to a new family of sensor kinase called the HWE/ HisKA2 sensor kinases. This family of sensor kinases differs from other sensor kinases because it lacks the F box rather, HWE kinases contain a H-box, N-box and a G1-box. Analysis of the protein sequence showed the *R. leguminosarum* sensor kinase contains these regions that defines the HWE family (figure 18 and appendix B) of the sensor kinases.

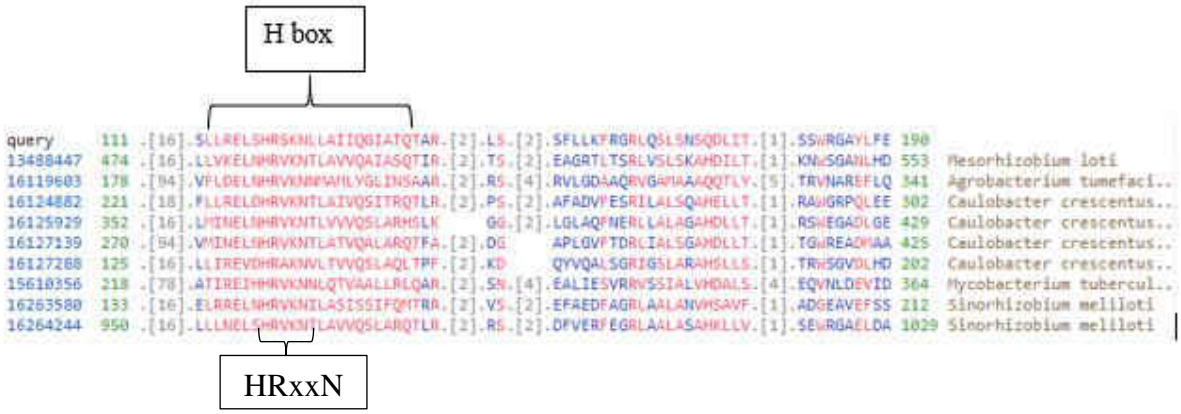


Figure 18. H box in HWE sensor kinase, query sequence is *R. leguminosarum* ATCC14479 sensor kinase (N and alignments in APPENDIX B)

The HWE family of sensor kinases is a newly described family of kinases that has not been studied in-depth. The main mode of study has been *in silico* looking for conserved motifs, These conserved motifs differ from other sensor kinases. The H box (figure 18) contains a conserved histidine residue, this H-box is in *R. leguminosarum*'s sensor kinase. This H-box also matches up with the phosphorylation site predicted by the VgrS alignment. Upon further protein sequence alignments shows the *R. leguminosarum* sensor kinase contains the N-box (appendix B), which is defined by a conserved asparagine residue flanked by an unusual consensus HELATNAXKYGALS (Karniol and Vierstra 2004). The protein sequence alignment also confirmed the presence of the G-box (figure 19), this region is defined by a rich glycine region on the C-terminus of the protein. Since this *R. leguminosarum* sensor kinase contains these motifs it is confirmed to belong to the HWE family of senor kinases, this family has been shown to be active in phosphorelays in α- and γ-proteobacteria (Karniol and Vierstra 2004). This family is believed to be involved in multiple roles of environmental signaling due to the presence of methyltransferase domains (Karniol and Vierstra 2004).

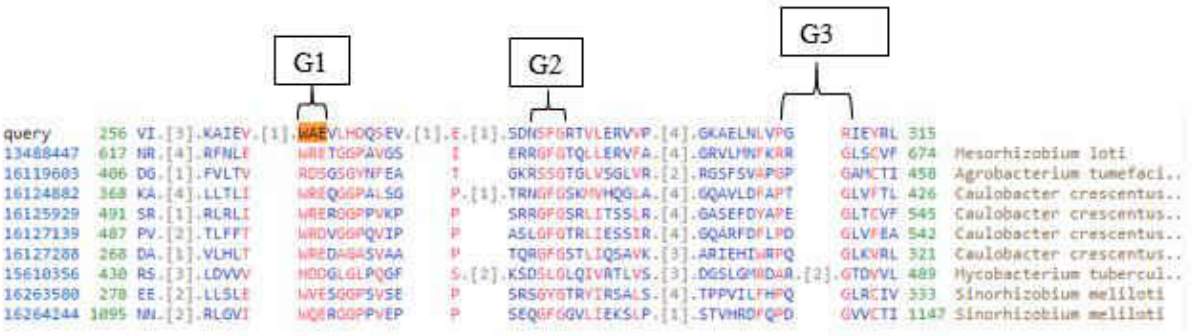


Figure 19. G1-3 box alignment of the *R. leguminosarum*'s sensor kinase with other HWE sensor kinases

The final *in silico* analysis performed was protein modeling, the protein sequence was analyzed through the SWISS-MODEL program. The protein model (figure 20) shows a structure consistent with a sensor kinase. The protein model shows a hydrophobic transmembrane region, a cytoplasmic domain containing α-helices, and sensor domain. With all the available data it is concluded the *R. leguminosarum*'s sensor kinase belongs to the HWE family of sensor kinases, however the exact function could not be confirmed. Future studies involving mutations in the predicted phosphorylation region will be able to elucidate the function of this sensor kinase. Based on the putative role of HWE sensor kinases in sensing a potential wide range of environmental changes, such as iron and flavonoids concentration in the environment. So mutants should be tested for their ability to sense flavonoids for root nodulation and the ability to sense changes in intracellular and extracellular iron levels.

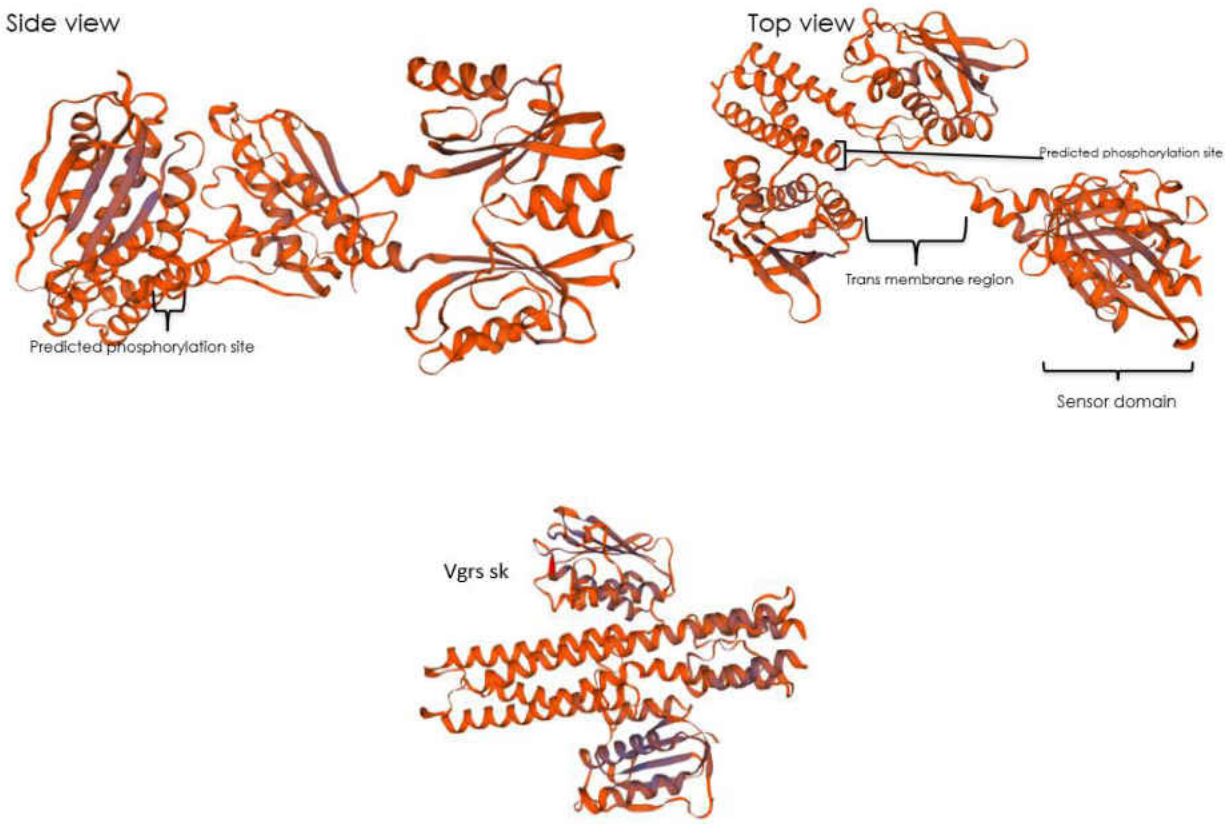


Figure 20. SWISS-MODEL of *R. leguminosarum's* HWE sensor kinase with top and side views, bottom picture is the *X. campestris VgrS* sensor kinase

Hemin Binding Proteins

Since the growth data support that *R. leguminosarum* ATCC 14479 is capable of utilizing heme compounds as a source of iron we wanted to investigate for any heme binding proteins. *In silico* analysis performed to check for the presence of heme binding proteins. This analysis showed that *R. leguminosarum* ATCC 14479 contains a putative heme binding protein that is homologous to *Sinorhizobium meliloti* 242's *ShmR* protein (Amarelle et al. 2008, and Battistoni et al. 2002). This protein was shown to be a TonB dependent heme binding protein, and in *S. meliloti ShmR* expression is directly related to the presence of exogenous heme. A nucleotide and protein blast alignment was performed (Figure 24 and 25) for *R. leguminosarum*'s *ShmR*

homologue. The nucleotide alignment showed a 70% match when aligned with *S. meliloti's ShmR* gene (appendix B). while the Amino acid alignment showed a 65% match (Figure 21).

Figure 21. Amino acid alignment of *R. leguminosarum* ATCC 14479 TonB dependent Heme receptor

The final heme receptor *in silico* analysis performed was protein modeling. The amino acid sequence for both *R. leguminosarum*'s putative TonB dependent heme receptor (figure 22), and *S. meliloti ShmR* receptor were modeled using SWISS-EXPASY. The models were built using the crystalized heme receptor *ShuA* from *S. dysinteriea*, the models were then aligned for comparison (figure 23), the green represents similar residues while red is dissimilar residues

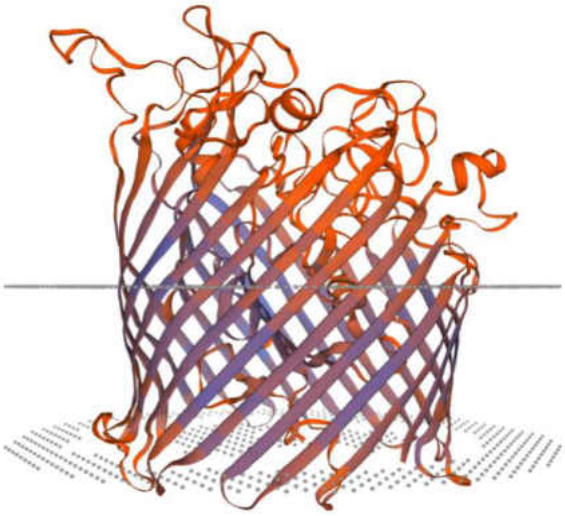


Figure 22. *R. leguminosarum* ATCC 14479 putative TonB dependent heme receptor. Grey rings represent membrane annotation

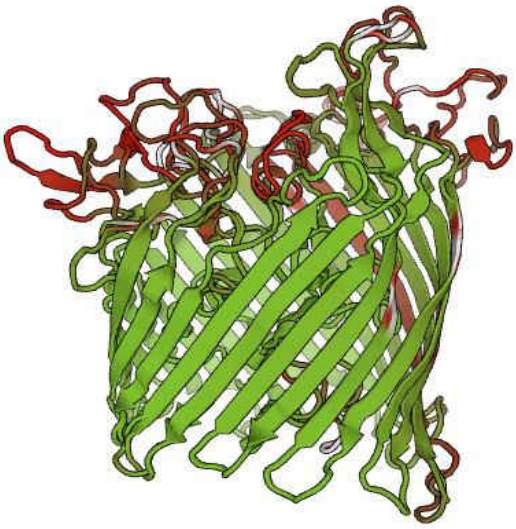


Figure 23. *R. leguminosarum* ATCC 14479 putative TonB dependent heme receptor and *S. meliloti* model alignment. Green = similar residues, red= dissimilar residues, white= extra residues

Detection of Siderophores

While investigating the growth of *R. leguminosarum* ATCC 14479 in hemin supplemented MMW media we also looked for siderophore production. The main siderophore *R. leguminosarum* ATCC 14479 produces is vicibactin (Wright et al. 2013). Vicibactin is a hydroximate siderophore, it is a cyclic siderophore that binds ferric iron by using three hydroximate groups (Wright et al. 2013). We tested for vicibactin production by collecting the supernatants of MMW supplemented with hemin as the sole source. The conditions tested were: 200µM 2',2'-dipyridyl, 10µM hemin, 15µM hemin, and 30µM hemin. The supernatants were collected at 24, 48, and 72 hours and loaded into wells on CAS media. CAS media was used due to the color change when iron is striped from the iron-dye complex. When iron is bound to the dye it has a blue-green color, and when iron is stripped from the dye it produces an orange to brown color. The supernatants of the tested conditions failed to produce halos on CAS media (Appendix B). No siderophore was detected in the conditions tested. This is due to the biphasic relationship between iron concentration and siderophore production (Wright 2010). This biphasic relationship shows that a certain minimum of iron in the media is needed to stimulate siderophore production. Then when the intracellular iron concentration reaches a certain level it represses genes involved in siderophore production (Wright 2010). The reason for no halos being observed indicates that the cells already reached optimum intracellular iron levels in the hemin supplemented media. In the 200µM 2',2'-dipyridyl supplemented media the iron concentration was too low to stimulate siderophore production.

CHAPTER 4. CONCLUSIONS

The Present work was carried out to further confirm the work previously done in the lab reporting the genetic presence of the *HmuPSTUV* operon and the ability of *R. leguminosarum* ATCC 14497 to utilize heme compounds as a sole source of iron (Shushant 2017). The presence of this operon was confirmed once the whole genome for *R. leguminosarum* ATCC 14479 was sequenced. The *HmuPSTUV* operon showed high similarity too operons from other *Rhizobium* species that are shown to be involved in hemoglobin and leghemoglobin utilization.

Attempts were made to mutate select genes in the *Hmu* operon. The genes selected for mutation were *HmuU/V* along with the putative sensor kinase. The plasmids containing the SOE products were confirmed to contain the correct DNA sequence for the desired SOE product. The issues occurred in the conjugation stage, this could be due shearing of the F pilus or exopolysaccharide blocking the formation of the F pilus.

The growth data for *R. leguminosarum* ATCC 14479 in media supplemented with hemin as the sole source of iron show that it is capable of utilizing hemin. In the aerobic condition a dose dependent relationship was observed, the more hemin supplied the more growth was observed. These aerobic growth data support the hypothesis that *R. leguminosarum* ATCC 14479 is capable of utilizing the iron in hemin, but the means by which the iron is being utilized is unconfirmed. *R. leguminosarum*'s hemin supplied growth plateaued between 15µM to 30µM, this plateauing is most likely due to the cells reaching optimum intracellular iron levels. This potential reason for this plateauing is further supported by work done on a homologous system in *S. meloti*. They found that HmuS (hemin degrading factor) is involved with hemin utilization when heme is present at a concentration of 0.5µM-50µM (Amarelle et al. 2016).

Since *Rhizobium* species encounter leghemoglobin in a reduced oxygen environment the growth of *R. leguminosarum* supplemented with hemin in an anerobic environment was investigated. The growth data for anaerobic supplied hemin media show that *R. leguminosarum* ATCC 14479 is incapable of growing in an anaerobic environment. The addition of hemin to the media results in prolonged cell survival, this survival is most likely due to the iron in hemin to be reduced to Fe^{+2}. This reduction of iron results in its release from the hemin molecule, making the iron readily available for uptake by direct iron acquisition systems.

Due to an unavailability of *Hmu* mutants to investigate how the hemin bound iron is being utilized, we looked for the presence of heme binding proteins. *In silico* analysis was performed looking for the presence of heme binding proteins in *R. leguminosarum* ATCC 14479. The results showed a putative TonB dependent heme binding protein, further analysis showed it was homologous to the heme receptor *ShmR* in *S. meliloti*. This receptor was found to be TonB dependent and vital for heme utilization (Amarelle et al. 2008). The *R. leguminosarum ShmR* homolog has a 70% nucleotide and 65% amino acid match compared to *ShmR*. The outer membranes of iron starved *R. leguminosarum* cells were extracted and analyzed for heme binding proteins. Attempts to detect any heme binding proteins in the outer membrane extracts failed, this was most likely due to procedural errors in the outer membrane extraction.

The culture supernatants from the hemin utilization experiment were tested for siderophore production. The hemin supplemented media tested failed to produce siderophores, most likely due to the cells already reaching the level of iron which would repress siderophore production.

Since genomic analysis showed the presence of a conserved sensor kinase in close proximity to the *HmuPSTUV* operon, the senor kinase was investigated via *in silico* analysis. The

sensor kinase had homologous amino acid sequence to *Xanthomonas campestris* sensor kinase *VgrS*, which senses intra- and extracellular iron levels. The protein sequence alignment shows very little homology, but the alignment had some homology around the *VgrS* phosphorylation site. The protein sequence was aligned with sensor kinases in the *Rhizobium* family. These alignments showed the sensor kinase belongs to the newly investigated HWE sensor kinase. This family of proteins is believed to be involved with sensing environmental changes. Future work with mutating the predicted phosphorylation site could elucidate potential functions of this senor kinase in sensing environmental changes. The mutant of this sensor kinase should be tested for its involvement in iron sensing and flavonoid detection since these are two of the major environmental changes that *Rhizobium* species must sense.

REFRENCES

Ahmer BM, Thomas MG, Larsen RA, Postle K. Characterization of exbBD operon of Escherichia coli and the role of ExbB and ExbD in TonB function and stability. 1995; J Bacteriol. 177(16):4742-747.

Amarelle V, Rosconi F, Lárazo-Matinez JM, Buldain G, Noya F, O'Brian MR, Fabiano E. HmuS and HmuQ of Ensifer/ Sinorhizobium meloti degrade heme in vitro and participate in heme metabolism in vivo. 2016; Biometals. 29:333-347.

Anazaldi L, Skaar E. Overcoming the heme paradox: heme toxicity and tolerance in bacterial pathogens. 2010; Infection and Immunity. 78(12):4977-989.

Andrews S, Robinson A, Rodríguez-Quiñones F. Bacterial iron homeostasis. 2003; FEMS

Microbiology Reviews 27:215-237.

Bagg A, Neilands JB. Ferric uptake regulation protein acts as a repressor, employing iron(II) as a cofactor to bind the operator of an iron transport operon in Escherichia coli. 1987; Biochemistry. 26:5471-477.

Battistoni F, Platero R, Duran R, Cervenasky C, Battistoni J, Arias A, Fabiano E. Identification of an iron-regulated, hemin-binding outer membrane protein in Sinorhizobium meliloti. 2002; Appl Environ Microbiol. 68(12):5877-81.

Becana M, Moran JF, Iturbe-Ormaetxe I, Gogorcena Y, Escuredo PR. Structure and function of leghemoglobins. 1995; Estac Exp Aula Dei. 21(3):203-08.

Bergersen FJ, Appleby CA. Leghemoglobin within bacteroid enclosing membrane envelopes from soybean root nodules. 1981; Planta. 152: 534-43.

Bijlsma JJ, Grroisman EA. Making informed decisions: regulatory interactions between two-component systems. 2003; Trends Microbiol. 11:359-66.

Brear EM, Day DA, Smith PMC. Iron: an essential micronutrient for the legume-rhizobium symbiosis. 2013; Front in Plant Sci. 4(359):1-15.

Camaschella C. Iron-deficiency anemia. 2015; N Engl J Med. 372:1832-843.

Contreras H, Chim N, Credali A, Goulding CW. Heme uptake in bacterial pathogens. 2014; Curr Opin Chem Biol. 19:34-41.

Costa D, Amarelle V, Valverde C, O'Brian M, Fabiano E. The Irr and RirA proteins participate in a complex regulatory circuit and act in concert to modulate bacterioferritin expression in Ensifer meliloti. 2017; Appl Environ Microbiol. 83(16)1-14.

Chao TC, Buhrmester J, Hansmeier N, Puhler A, Weidner S. Role of regulatory gene rirA in the transcriptional response of Sinorhizobium meliloti to iron limitation. 2005; Appl Environ Microbiol. 71(10):5969-982.

Cheung J, Hendrickson WA. Sensor domains of two-component regulatory systems. 2010; Curr Opin Microbiol. 13:116-23.

Crownover B, Covey C. 2013. Hereditary hemochromatosis. 2013; Am Fam Physician. 87(3): 183-90.

De Lorenzo V, Wee S, Herrero M, Neilands JB. Operator sequences of aerobactin operon of plasmid ColV-K30 binding the ferric uptake regulation (fur) repressor. 1987; Bacteriol. 169:2624-630.

Ditta G, Stanfield S, Corbin D, Helinski D. Broad host range cloning system for Gram-negative bacteria: Construction of a gene bank of Rhizobium meliloti. 1980. PNAS. 77(12):7347-351.

D'Haeze, W, Holsters M. Nod factor structures, responses and perception during initiation of nodule development. 2002; Glycobiology 12(6):79– 105.

Freeman WM, Walker SJ, Vrana KE. Quantitative RT-PCR: pitfalls and potential. 1999; BioTechniques 26(1): 112-22.

Gage D. Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobium during nodulation of temperate legumes. 2004; Microbiol. Mol Biol Rev 68(2): 280-300.

Hans F. "Hemin". 1941; Org. Synth. 21: 53.

Hereditary hemochromatosis. genetic home reference (GHR) national institutes of health (NIH). Maryland (MD): National library of medicine 2020 [cited 2020 Apr 3]. https://ghr.nlm.nih.gov/condition/hereditary-hemochromatosis#definition

Hoang TT, Karkhoff-Schweizer RR, Kutchma AJ, Schweizer HP. A broad-host-range FlpFRT recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked Pseudomonas aeruginosa mutants. 1998; Gene. 212(1):77-86.

Kadner R. Regulation by iron: RNA rules the rust. 2005; J Bacteriol. 187(20):6870-873.

Karniol B, Vierstra RD. The HWE histidine kinase, a new family of bacterial two-component sensor kinases with potentially diverse role in environmental signaling. 2004; J Bacteriol. 186(2):445-453.

Kneen BE, Larue TA. 1983. Congo red absorption by Rhizobium leguminosarum. 1983; Appl Environ Microbiol. 45(1):340-42.

Kholer SD, Weber A, Howard SP, Welte W, Drescher M. The Proline-rich domain of TonB possesses an extended polyproline II-like conformation of sufficient length to span the periplasm of gram -negative bacteria. 2010; Protein Sci. 19(4)625-30.

Krewulak KD, Vogel HJ. Structural biology of bacterial iron uptake. 2008; Biochim Biophys Acta. 1778(9):1781-804.

Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. 1970; Nature. 227(5259):680-85.

Lee C. Isolation of an outer membrane hemin-binding protein of Haemophilus influenzae type B. 1992; Infection and Immunity. 60(3):810-16.

LeRoux J, Mavengere N, Ellis A. The structure of legume–rhizobium interaction networks and their response to tree invasions. 2016; AoB Plants 8:plw038.

Manhart JR, Wong PP. Nitrase activities of rhizobia and the correlation between nitrate reduction and nitrogen fixation. 1979; Can J microbiol. 25:1169-174.

Miethke M, Marahiel M. Siderophore-based iron acquisition and pathogen control. 2007; Microbiol and Mol Bio Rev. 71(3):413-451.

Messenger AJM , Barclay R. Bacteria, iron and pathogenicity. 1983; Biochem Ed. 11(2):54-63.

Mogensen JE, Otzen DE. Interactions between folding factors and bacterial outer membrane proteins. 2005; Mol Microbiol. 57(2):326-46.

Nienaber A, Hennecke H, Hans-Martin F. Discovery of a haem uptake system in the soil bacterium Bradyrhizobium japonicum. 2001; Mol Microbiol. 41(4):787-800.

Noinaj N, Guillier M, Branard TJ, Buchanan SK. 2010 TonB-dependent transporters: regulation, structure, and function. Annu. Rev. Microbiol. 13(64)43-60.

Noya F, Arias A, Fabiano E. 1997 Heme compounds as iron sources for nonpathogenic Rhizobium bacteria. Journal of Bacteriology. 197:3076-078.

O'Brian MR. Perception and homeostatic control of iron in the rhizobia and related bacteria. 2015; Annu Rev Microbiol. 69:229-45.

Ochsner UA, Johnson Z, Vasil ML. Genetics and regulation of two distinct haemuptake systems, phu and has, in Pseudomonas aeruginosa. 2000; Microbiol 146(1): 185-98.

Ollis AA, Postle K. Identification of functionally important TonB-ExbD periplasmic domain interactions in vivo. 2012; J Bacteriol. 194(12):3078-087.

Ott T, van Dongen J, Gu¨nther C, Krusell L, Desbrosses G, Vigeolas H, Bock V, Czechowski T, Geigenberger P, Udvardi M. Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. 2005; Curr Biol. 15:531-35

Peacock RS, Weljie AM, Howard SP, Price FD, Vogel HJ. The solution structure of the C-terminal domain of TonB and interaction studies with TonB box peptides. 2005; J Mol Biol. 345(5)1185-197.

Perez-Guzamn L, Bogner K, Lower B. Earth's ferrous wheel. 2010; Nature. 3(10):32

Perkins-Balding D, Ratliff-Griffin M, Stojiljkovic I. Iron transporst systems in Neisseria meningitidis. 2004; Microbiol Mol Rev. 68(1):154-71.

Peters N, Frost J, Long S. A plant flavone, luteolin, induces expression of Rhizobium meliloti nodulation genes. 1986; Science 233:977-80.

Postgate J. Nitrogen Fixation (3rd Ed.). 1998; Cambridge University Press. Cambridge, UK. Pg.1-15

Postle K, Larsen RA. TonB-dependent energy transduction between outer and cytoplasmic membranes. 2007; BioMetals. 20(3-4) 453-65.

Poulos L. Heme enzyme structure and function. 2014; Chem. Rev. 114(7)3919-962.

Ramirez-Bahena M.H, Garcia-Fraile P, Peix A, Valverde A, Rivas R, Igual JM. Revision of the taxonomic status of the species Rhizobium leguminosarum (frank 1879) frank 1889AL, Rhizobium phaseoli Dangeard1926AL and Rhizobium trifolii Dangeard 1926AL. R. trifolii is a later synonym of R. leguminosarum. Reclassification of the strain R. leguminosarum DSM 30132(=NCIMB 11478) as Rhizobium pisi sp. nov. 2008; Int J Syst Evol Microbiol 58(11): 2484-490

Schauer K, Rodionov DA, de Reuse H. New substrates for tonb-dependent transport: do we only see the “tip of the iceberg’?. 2008;Trends Biochem Sci. 33:330-38.

Schwiesow L, Metter E, Wei Y, Miller H, Herrera N, Balderas D, Kiley P, Auerbuch V. Control of hmu heme uptake genes in Yersinia pseudoterbculosis in response to iron source. 2018; Front Cell Infect Microbiol. 8(47):1-13.

Khanal, Sushant, "Identification of “fhuA” Like Genes in Rhizobium leguminosarum ATCC 14479 and its Role in Vicibactin Transport and Investigation of Heme Bound Iron Uptake System" (2018). Electronic Theses and Dissertations. Paper 3354. https://dc.etsu.edu/etd/3354

Tiwari S, Jamal SB, Hassan SS, Carvalho PVSD, Almeida S, Barh D, Ghosh P, Silva A, Castro TLP, Azevendo V. Two-Component signaling transduction systems of pathogenic bacteria as targets for antimicrobial therapy. 2017; Front Microbiol. 8(1878):1-7.

Todd JD, Wexler M, Sawers G, Yeoman KH, Poole PS, Johnston A. RirA, an iron-responsive regulator in the symbiotic bacterium Rhizobium leguminosarum. 2002; Microbiology. 148:4059-071.

Touati D, Jacques M, Tardat B, Bouchard L, Despied S. Lethal oxidative damage and mutagenesis are generated by iron in Δfur mutants of Escherichia coli: protective role of superoxide dismutase. 1995; J Bacteriology. 177(9):2305-314.

Troxell B, Hassan H. Transcriptional regulation by ferric uptake regulator (fur) in pathogenic bacteria. 2013; Front. Cell. Infect. Microbiol. 3(59):1-13.

Wang L, Pan Y, Yuan ZH, Zhang H, Peng BY, Wang FF, Qian W. Two-Component signaling system VgrRS directly senses extracytoplasmic and intracellular iron to control bacterial adaptation under iron depleted stress. 2016; PLOS pathogens. 12(12):e1006133.

Weinberg E. Iron and susceptibility to infectious disease. 1974; Science 184(4140):952-56.

Wexler M, Yeoman K, Stevens JB, de Luca NG, Sawers G, Johnston AWB. The Rhizobium leguminosarum tonB gene is required for the uptake of siderophore and haem as a source of iron. 2001; Mol Microbiol. 41(4):801-16.

Wilson B, Bogdan A, Miyazawa M, Hasimotot K, Tsuji Y. 2017 Siderophores in iron metabolism: from mechanism to therapy. 2016; Trends Mol Med. 22(12):1077-090.

Wright W. Isololation and identification of the siderophore “vicibactin” produced by rhizobium leguminosarum ATCC 14479 [Thesis]. 2010. [Johnson City (TN)]:East Tenneessee State University. [accessed 2020] http://dc.etsu.edu/etd/1690/.

Wright W, Little J, Liu F, Chakraborty R. Isolations and structural identification of the trihydroxamate siderophore vicibactin and its degrative product from Rhizobium leguminosarum ATCC 14479 bv. trifolii. 2013; Biometals. 26(2):271-83.

Young J, Crossman L, Johnston A, Thomas N, Ghazoui Z, Hull K, Wexler M, Curson A, Todd J, Poole P, Mauchline T, East A, Quail M, Churcher C, Arrowsmith C, Cherevach I, Chillingworth T, Clarke K, Cronin A, Davis P, Fraser A, Hance Z, Hauser H, Jagels K, Moule S, Mungall K, Norbertczak M, Rabbinowitsch E, Sanders M, Simmonds M, Whitehead S, Parkhill J. The genome of Rhizobium leguminosarum has recognizable core and accessory components. 2006; Genome Biology 7(4):R34 1-20.

Zambolin S, Clantin B, Chami M, Hoos S, Haouz A, Villeret V, Delepelarie P. Structural basis for haem piracy from host haemopexin by Haemophilus influenzae. 2016; Nat Commun 7:11590.

APPENDICES

Appendix A: Media and Buffers

Luria Broth (LB)

Tryptone	10.0g
Yeast Extract	5.0 g
NaCl	10.0g
ddH_2O	1L

Dissolve ingredients, pH to 7.0 and autoclave
For LB agar plates:
agar: 15.0g

Yeast Extract Mannitol Broth (YEM)/ Congo Red agar (CR)

Mannitol	4.0g
K_2HPO_4	0.2g

$MgSO_4$	0.08g
NaCl	0.04g
Yeast Extract	0.4g
ddH_2O	400ml

dissolve ingredients, adjust pH to 6.8 and autoclave

For CR agar plates:

Agar	12.0g

1.0% Congo Red dye solution 1ml

dissolve ingredients, adjust pH to 6.8 and autoclave

Modified Manhart and Wong (MMW)

Dextrose	6.0g
Glutamate	1.5g
KH_2PO_4	1.0g
K_2HPO_4	0.764g
$MgSO_4$	0.18g
$CaSo_4*2H_2O$	0.13g
ddH_2O	to 1L

Dissolve ingredients, adjust pH to 6.8, and autoclave, post autoclave add 1ml of concentrated vitamin and trace minerals solution per 1L of basal media

Vitamins and Trace Mineral solution (MMW)

H_3BO_3	145.0mg
$CuSO_4*5H_2O$	4.37.0mg
$MnCl_2*4H_2O$	4.3.0mg
$ZnSO_4*7H_2O$	108.0mg
$Na_2MoO_4*2H_2O$	250.0mg
$CoCl_2*6H_2O$	10.0mg
$Na_2EDTA*2H_2O$	250.0mg
Riboflavin	10.0mg
P-aminobenzoic acid	10.0mg
Nicotinic acid	10.0mg
Biotin	12.0mg
Thiamine HCl	40.0mg
Pyridoxine HCl	10.0mg
Calcium Panthenate	50.0mg
Inositol	50.0mg
Vitamin B12	10.0mg
ddH_2O	to 100ml

Mix ingredients, filter sterilize, wrap in aluminum foil and store at 4°C

Chrome Azurol S (CAS)

MOPS	15g
NaCl	0.25g
K_2HPO_4	0.15g

NH_4Cl	0.05g
L-asparagine	2.5g
Agar	7.5g
ddH2O	500ml

pH media to 6.8 with 5M NaOH and autoclave. Then warm 50% sucrose solution and CAS indicator solution in water bath. Add 10 ml of sucrose solution per 500ml of media and add 50ml of CAS indicator solution to basal media and mix while avoiding bubbles.

CAS indicator solution

Iron III solution

$Fe_3Cl * 6H_2O$	27.0mg
12M HCl	83.3µl
ddH_2O	100µl

Solution A

Chrome azurol S dye	60.5mg
ddH_2O	50ml

Dissolve CAS dye is water then add 10ml of iron III Solution

Solution B

HDTMA	72.9mg
ddH_2O	40ml

Dissolve HDTMA in water then add solution B to a mixing solution A. Once added autoclave the CAS indicator dye and store in dark area. Before adding to basal CAS media warm dye and 50% sucrose solution in 50°C water bath

2X YT broth

Bacto Tryptone	16.0g
Yeast Extract	10.0g
NaCl	5.0g
H_2O	1.0L

pH to 7.0 with NaOH and autoclave

Transformation Buffer 1 (100ml)

Rubidium Chloride	1.209g
Manganese Chloride	989.5mg
Potassium Acetate	294mg
Calcium Chloride	147mg
Glycerol	15ml (15% V/V)
ddH_2O	85ml

Adjust pH to 5.8 with 1M acetic acid (do not overshoot) then autoclave and store room temperature

Transformation Buffer 2 (100ml)

MOPS	209.3mg
Rubidium Chloride	120.9mg
Calcium chloride	1.103g
Glycerol	15ml (15% V/V)
ddH_2O	85ml

Adjust pH to 6.5 with KOH (do not overshoot) autoclave and store room temperature

0.025M Hemin Solution

Lyophilized hemin	815.0mg
1.4 M Ammonium hydroxide	50ml

Mix ingredients, heat in 95°C water bath for less than 5 minutes, mix until solution cools down to room temp the filter sterilize and store at 4°C

SDS-PAGE Solutions

Stacking Gel Buffer (pH 6.8)

0.5M Tris
ddH_2O

Resolving Gel buffer (pH 8.8)

1.5M Tris
ddH_2O

Tris-Glycine SDS Buffer (10X)

Tris	0.25M
Glycine	1.92M
Sodium Dodecyl Sulfate (SDS)	1.0% (W/V)
ddH_2O	1.0 L

SDS-PAGE gel

	Stacking	Resolving (12%)
30% Bis-arylamide	0.66ml	3.0ml
Stacking gel buffer	1.26ml	-
Running gel buffer	-	1.88ml
ddH_2O	3ml	2.52ml
10% SDS	50µl	75µl
10% APS	25µl	37.5µl
TEMED	5µl	3.75µl

Coomassie Blue Staining solution

Coomassie R-250	0.1% (V/V)
Methanol	50% (V/V)
Glacial acetic acid	10% (V/V)
ddH_2O	40% (V/V)

2X Laemmli Buffer

SDS	4% (W/V)
Glycerol	20% (V/V)
1M Tris pH 6.8	120mM
Bromophenol Blue	0.02 (W/V)
ddH_2O	

SDS-PAGE Destaining solution

Methanol	50% (v/v)
Glacial acetic acid	10% (v/v)
ddH_2O	40% (v/v)

Appendix B: Supplemental Data

Primers for HmuU SOE, underlined portions bind to *R. leguminosarum*'s genome

Primer name	Primer sequence (5'-3')
SOERHU	CGTGACGTTAGCGTACGAGGACTAAGAATACGCTGGACGGCATGTTGCTGATGACGTCGAGG
SOEFHU	CCGTCCAGCGTATTCTTAGTCCTCGTACGCTAACGTCACGTGATCTTCGCGGATGTCCTGG
SOEFA	CGATGGTACCGACGCATGGCGCGAGATCATGG
SOERA	GCTATCTAGATGGCTCATGGCTGTCTCCATTGG

Primers for HmuU full gene amplification

Primer Name	Primer sequence (5'-3')
HmuUF	CGTCTACCTGCTCGGCTTCG
HmuUR	GATCATCTAAATGTCACTCACGG

Internal HmuU primer, works with SOEFHU to amplify if internal gene sequence is present

Primer Name	Primer sequence (5'-3')
HmuUIntF	ATCCTGCGTATGGCGATCGG

Primers for HmuV SOE, underlined portions bind to *R. leguminosarum's* genome

Primer name	Primer sequence (5'-3')
SOEHmuVF	CCGTCCAGCGTATTCTTAGTCCTCGTACGCTAACGTCACGCATCGTTGAGGTGCTGACG
SOEHmuVR	CCGTCCAGCGTATTCTTAGTCCTCGTACGCTAACGTCACGCATCGTTGAGGTGCTGACG
SOEVR	GCTATCTAGATGCGAGTGCGCATAGAGCGA
SOEVF	CGATGGTACCGAGGATGCCTCGATCCTCG

Primers for HmuV full gene amplification

Primer name	Primer sequence (5'-3')
HmuVF	GGATCCTGCTGAGACAGCG
HmuVR	CGGTTACCTCTGTCGGTACT

Internal HmuV primer, Works with SOEVR to give amplification if internal gene portion is present.

Primer name	Primer sequence (5'-3')
HmuVintF	GCACGATCTCAATCTGACG

Primers for Sensor kinase SOE, underlined portions bind to *R. leguminosarum's* genome

Primer name	Primer sequence (5'-3')
SEOF2SK	CCGTCCAGCGTATTCTTAGTCCTCGTACGCTAACGTCACGATATCGTCTGACCATTCCG
SOER2SK	CGTGACGTTAGCGTACGAGGACTAAGAATACGCTGGACGGGCGTAGAAGATCGACAGC
SOEFB	CGATGGTACCGATTAACCGCTCATTGAAGCC
SOERB	GCTATCTAGATGGTGGCGATGCCGTATTCG

Primers for Sensor kinase full gene amplification

Primer name	Primer sequence (5'-3')
SKF	GTCAACGACATCCTCAAGACG
SKR	GATGGTGATCGCATCGACC

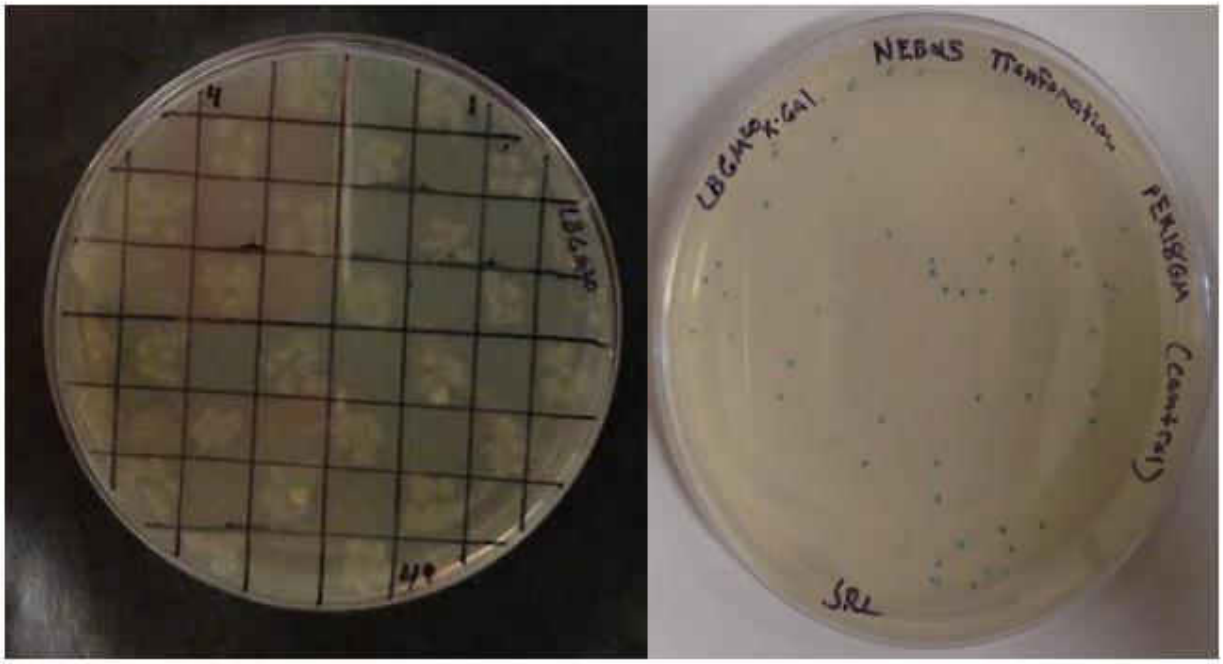
Internal Sensor kinase primer, works with SOERB to amplify if internal gene portion is present.

Primer name	Primer sequence (5'-3')
SKIntF	GTTCTGCATGATCAGTCGG

Primers for RT-PCR, Primers designed by IDT PrimerQuest tool.

Primer Name	Primer sequence (5'-3')
HmuURTF	CGGACATCCGCGAAATCA
HmuURTR	GCCTGTCGTCACCGAAA
HmuPRTF	CGAAAGCGCGGATCTCTT
HmuPRTR	AGAATGAGCTTGCCCTGAC
VbsSRTF	CTTCGAGAGCTTTCCACTGA
VbsSRTR	CGGGAAAGACCGTGTAGTT
16sRTF	TCGGAATTACTGGGCGTAAAG
16sRTR	CTCCAGATCGACAGTATCAAAGG

Blue White Screening of pEX18 NEBα5 transformants



Blue white screening of NEBα5 cells grown of X-gal containing LB agar, left picture is white colonies that contain the pEX18 and *HmuU* SOE product. Right picture NEBα5 cells transformed with a undigested pEX18$_{Gm}$ plasmid (blue positive control)

Blast of *R. leguminosarum* ATCC 14479 regions flanking the Hmu operon (bp 1,963,826-1,974,106 compared to other Rhizobium taxid)

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Rhizobium leguminosarum strain ATCC 14479 chromosome, complete genome	18986	18986	100%	0.0	100.00%	CP030760.1
Rhizobium leguminosarum bv. viciae strain UPM791 chromosome, complete genome	18207	18207	100%	0.0	98.65%	CP025509.1
Rhizobium leguminosarum bv. trifolii strain 23B chromosome, complete genome	18201	18201	100%	0.0	98.60%	CP050085.1
Rhizobium leguminosarum bv. viciae strain BIHB 1217, complete genome	18201	18201	100%	0.0	98.64%	CP022665.1
Rhizobium leguminosarum tonB gene, hmu operon (hmuPV genes) and rpoZ gene	18092	18092	100%	0.0	98.45%	AJ310723.1
Rhizobium leguminosarum bv. viciae strain RCAM0610 chromosome, complete genome	17773	17773	100%	0.0	97.81%	CP050549.1
Rhizobium leguminosarum bv. trifolii strain 4B chromosome, complete genome	16284	16284	89%	0.0	98.47%	CP050103.1
Rhizobium leguminosarum bv. trifolii strain 3B chromosome, complete genome	16284	16284	89%	0.0	98.47%	CP050108.1
Rhizobium leguminosarum strain Vaf10, complete genome	14026	14026	90%	0.0	93.98%	CP016286.1
Rhizobium leguminosarum bv. viciae strain BIHB 1148, complete genome	13815	13815	89%	0.0	93.67%	CP022564.1
Rhizobium leguminosarum bv. trifolii WSM1689, complete genome	9659	14563	99%	0.0	91.13%	CP007045.1
Rhizobium leguminosarum strain A1 chromosome, complete genome	9491	14649	88%	0.0	95.29%	CP049730.1
Rhizobium indicum strain JKLM 12A2 chromosome, complete genome	9485	14520	88%	0.0	95.16%	CP054021.1
Rhizobium leguminosarum bv. trifolii TA1 chromosome, complete genome	9417	14662	88%	0.0	95.08%	CP053205.2
Rhizobium indicum strain JKLM 13E chromosome, complete genome	9371	14419	88%	0.0	94.90%	CP054031.1
Rhizobium leguminosarum strain Norway chromosome, complete genome	9234	14966	87%	0.0	96.31%	CP025012.1
Rhizobium leguminosarum bv. viciae chromosome complete genome, strain 3841	9086	14042	88%	0.0	93.97%	AM236080.1
Rhizobium leguminosarum bv. trifolii strain 22B chromosome, complete genome	9082	14025	88%	0.0	93.98%	CP050091.1
Rhizobium leguminosarum bv. viciae strain RCAM2802 chromosome, complete genome	8981	14578	87%	0.0	95.44%	CP050562.1
Rhizobium leguminosarum bv. trifolii strain 31B chromosome, complete genome	8961	14718	88%	0.0	95.37%	CP050080.1
Rhizobium leguminosarum bv. viciae strain RCAM0626 chromosome, complete genome	8907	14626	88%	0.0	95.20%	CP050555.1
Rhizobium leguminosarum bv. trifolii strain RCAM1365 chromosome, complete genome	8804	14216	89%	0.0	94.56%	CP050514.1
Rhizobium leguminosarum bv. trifolii WSM1325 chromosome, complete genome	8767	14412	88%	0.0	94.72%	CP001622.1
Rhizobium leguminosarum bv. trifolii strain CC275e chromosome, complete genome	8733	14299	88%	0.0	94.67%	CP053439.1
Rhizobium leguminosarum bv. trifolii strain 9B chromosome, complete genome	8722	14282	87%	0.0	94.64%	CP050097.1
Rhizobium leguminosarum bv. viciae 248 chromosome, complete genome	8672	14083	88%	0.0	94.44%	CP048280.1
Rhizobium leguminosarum strain Vaf-108, complete genome	8663	14109	88%	0.0	94.41%	CP018228.1

N-box alignment of *R. leguminosarum* ATCC 14479 sensor kinase compared to other known HWE sensor kinases. N box consensus HELATNAXKYGALS in HWE sensor kinases

16s DNA sequence alignment of E. coli cell that incorporated pEX18*ΔHmuU* plasmid.

Escherichia coli strain 191-a pink 16S ribosomal RNA gene, partial sequence

Sequence ID: MN208066.1 **Length:** 1444 **Number of Matches:** 1

Range 1: 35 to 827 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
1371 bits(742)	0.0	777/794(98%)	10/794(1%)	Plus/Plus

```
Query  1    AGGAAGCAAGCTTGCTTGCTTTCGCTGACGAGTGGCGGACGGGTGAGTAATGTCTGGGAA  60
            ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Sbjct  35   AGGAAGCAAGCTTGCTTGCTTT-GCTGACGAGTGGCGGACGGGTGAGTAATGTCTGGGAA  93

Query  61   ACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATACCGCATAACGTCGCAA  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  94   ACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATACCGCATAACGTCGCAA  153

Query  121  GACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCGGATGTGCCCAGATGGGATTAGCTA  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  154  GACCAAAGAGGGGGACCTTCGGGCCTCTTGCCATCGGATGTGCCCAGATGGGATTAGCTA  213

Query  181  GTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGATGACCAG  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  214  GTAGGTGGGGTAACGGCTCACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGATGACCAG  273

Query  241  CCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGC  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  274  CCACACTGGAACTGAGACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGC  333

Query  301  ACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTTGTA  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  334  ACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTTGTA  393

Query  361  AAGTACTTTCAGCGGGGAGGAAGGGAGTAAAGTTAATACCTTTGCTCATTGACGTTACCC  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  394  AAGTACTTTCAGCGGGGAGGAAGGGAGTAAAGTTAATACCTTTGCTCATTGACGTTACCC  453

Query  421  GCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCG  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  454  GCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCG  513

Query  481  TTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCAGATGTGAAAT  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  514  TTAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCAGATGTGAAAT  573

Query  541  CCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGGGG  600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  574  CCCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGGGG  633

Query  601  GTAGAATTCCAGGTGTAGCGGTGAAATGCGTANAGATCTGGAGGAATACCGGTGGCGAAG  660
            |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Sbjct  634  GTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAG  693

Query  661  GCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAN-ANGAATA  719
            |||||||||||||||||||||||||||||||||||||||||||||||||||  | || ||
Sbjct  694  GCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTA  753

Query  720  GATACCCTGGTANTC-ACGCCGTAA-CAATGTC-ACTTG-AGGTGGT-CCCT-GAGGCGT  773
            |||||||||||| || ||||||||| | ||||| ||||| |||| || |||| |||||||
Sbjct  754  GATACCCTGGTAGTCCACGCCGTAAACGATGTCGACTTGGAGGTTGTGCCCTTGAGGCGT  813

Query  774  G-CTTCCG-AGCTA  785
            | |||||| |||||
Sbjct  814  GGCTTCCGGAGCTA  827
```

Rhizobium leguminosarum strain ATCC 14479 chromosome, complete genome

Sequence ID: NZ_CP030760.1 Length: 4883137 Number of Matches: 105

Range 1: 1967027 to 1967359 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
564 bits(293)	3e-161	329/337(98%)	4/337(1%)	Plus/Plus

```
Query  1        CGGGCAGCAAGCGTGGCCGCATAAGGATCGCAACGATGACGATGCGTAACAATCTGCGCC  60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1967027  CGGGCAGCAAGCGTGGCCGCATAAGGATCGCAACGATGACGATGCGTAACAATCTGCGCC  1967086

Query  61       GGATCCGCCCTTGGGAACTGGCCCTGACGGCGGCCGTCGTGGCGCTGCCGTTGGTCCCGA  120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1967087  GGATCCGCCCTTGGGAACTGGCCCTGACGGCGGCCGTCGTGGCGCTGCCGTTGGTCCCGA  1967146

Query  121      CGGCAGCGCCAATCGGAGGGCTTCGCCTTCGTCCGCGCCGCCCACGCTGACGACAAGAAG  180
                |||||||||||||| | ||||||||||||||||||||||||||||||| ||||||
Sbjct  1967147  CGGCAGCGCCAATCGAA-GGCTTCGCCTTCGTCCGCGCCGCCCACGCTGACGAAAAGAAG  1967205

Query  181      CTCGACACCTCGCGCCTGGTTTCGGTCGGCGGCTACATCACCGAGATCGTCTATGCGCTT  240
                |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Sbjct  1967206  CTCGACACCTCGCGCCTGGTTTCGGTCGGCGGCGACATCACCGAGATCGTCTATGCGCTT  1967265

Query  241      GGCGAAGAAAGCCGGCTGATTGCCCGCGACACCACGAGCATGTATCCGGAGGCGGCGTTG  300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1967266  GGCGAAGAAAGCCGGCTGATTGCCCGCGACACCACGAGCATGTATCCGGAGGCGGCGTTG  1967325

Query  301      GAAGCCTGCCCAACGTCGGTAACATGCCGCGCGCTCT  337
                  ||| |||||||||||||| |||||| |||||||||
Sbjct  1967326  -AAG-CTGCCCAACGTCGGTTACATGC-GCGCGCTCT  1967359
```

5' DNA sequence alignment for the *Hmu*U knockout construct, query line is the Plasmid product and subject line is R. leguminosarum genome.

Rhizobium leguminosarum strain ATCC 14479 chromosome, complete genome

Sequence ID: NZ_CP030760.1 **Length:** 4883137 **Number of Matches:** 54

Range 1: 1969571 to 1970061 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
902 bits(469)	0.0	485/492(99%)	1/492(0%)	Plus/Minus

```
Query  1        CGCGCAAGCCCGCCTGTTTCGCACGAGGGTCACCACCGTCCGTTGCATGAATGATATCGC  60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1970061  CGCGCAAGCCCGCCTGTTTCGCACGAGGGTCACCACCGTCCGTTGCATGAATGATATCGC  1970002

Query  61       CCGGCATTGACCCAGATCATGCCAAAAACGTGCCGGACGATCAAAAGTTCCGCCGGAGCA  120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1970001  CCGGCATTGACCCAGATCATGCCAAAAACGTGCCGGACGATCAAAAGTTCCGCCGGAGCA  1969942

Query  121      GTGCTCAGGGGCGGGAAACGGCACTGTGGGCGAGCACGAAGGGCGTGCCGTCGGCCGGCA  180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1969941  GTGCTCAGGGGCGGGAAACGGCACTGTGGGCGAGCACGAAGGGCGTGCCGTCGGCCGGCA  1969882

Query  181      CCTGGTTGATGCGCAGCGCGCAACCGAACACCGAGAGCATCGTTTCGTCCGTCAGCACCT  240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1969881  CCTGGTTGATGCGCAGCGCGCAACCGAACACCGAGAGCATCGTTTCGTCCGTCAGCACCT  1969822

Query  241      CAACGATGCTGCCGGCGGCCGCGAGCCGGCCGGAATTCATCAGCACGATGCGGTCGGCAA  300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1969821  CAACGATGCTGCCGGCGGCCGCGAGCCGGCCGGAATTCATCAGCACGATGCGGTCGGCAA  1969762

Query  301      AGAGCGCCGTCAGATTGAGATCGTGCATGACGGCGATGACACCGCCGCCGCGTTCGCAGA  360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1969761  AGAGCGCCGTCAGATTGAGATCGTGCATGACGGCGATGACACCGCCGCCGCGTTCGCAGA  1969702

Query  361      AATTGCGCGCGAGCGTCATGATGGTCAGCTGGTGGCTGATGTCGAGGCTCGAGACCGGTT  420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1969701  AATTGCGCGCGAGCGTCATGATGGTCAGCTGGTGGCTGATGTCGAGGCTCGAGACCGGTT  1969642

Query  421      CGTCCGAGCAGCAGCCAGCAGGGCTTGCCGTCNACNACGGGCTCGGCGATCTGGCAANAA  480
                ||||  |||||||||||||||||||||||||||  || ||||||||||||||||||  |
Sbjct  1969641  CGTC-GAGCAGCAGCCAGCAGGGCTTGCCGTCGACGACGGGCTCGGCGATCTGGCAGAGA  1969583

Query  481      TTGCGGGCAAGC  492
                ||||||||||||
Sbjct  1969582  ATGCGGGCAAGC  1969571
```

3' DNA sequence alignment for the *Hmu*U knockout construct, query line is the Plasmid product and subject line is R. leguminosarum genome.

Rhizobium leguminosarum strain ATCC 14479 chromosome, complete genome

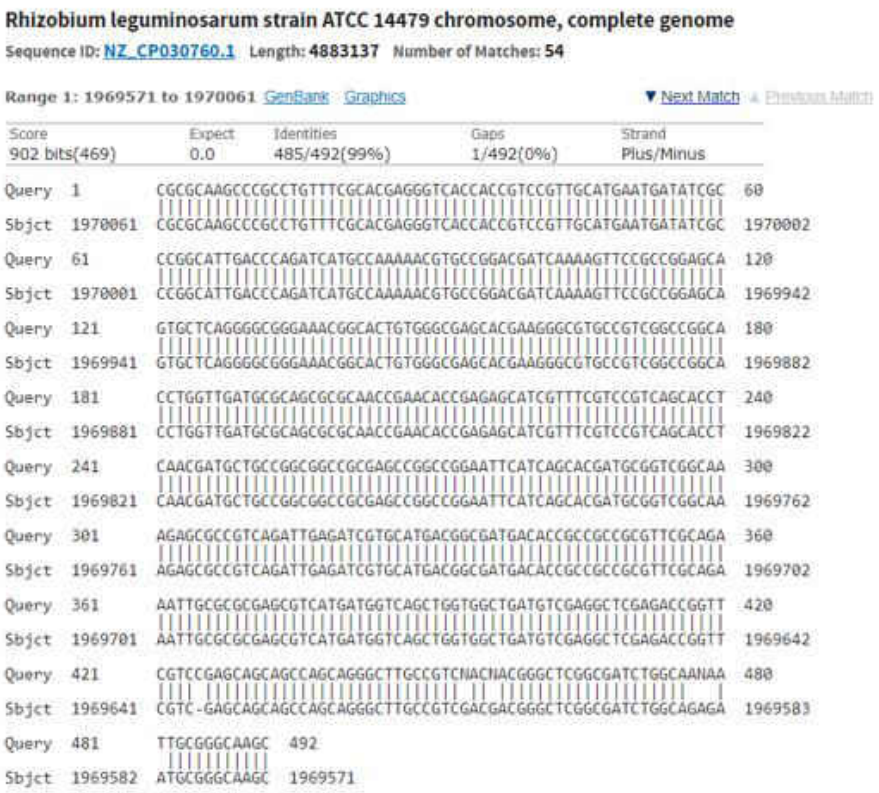
Sequence ID: NZ_CP030760.1 Length: 4883137 Number of Matches: 1

Range 1: 1970226 to 1970840 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
1046 bits(566)	0.0	603/619(97%)	11/619(1%)	Plus/Minus

```
Query  5        TCANATCGTCGCGCTTCGAGCC-ACCAGAATGGGCTGCGACGACTNGCTGCCCGAACACC  63
                |||| |||||||||||||||| ||||| || |||||||||| | ||||||  ||||||
Sbjct  1970840  TCA-ATCGTCGCGCTTCGAGCCCACCAG-ATTGGCTGCGACGA-TTGCTGCCAGAACACC  1970784

Query  64       GGCGGTCAGGAGCGGCTGG-CGCGGATAAAGGCCGAGCCCGACCGT-GCCAGAGA-TCCA  120
                |||||||||||||||||| ||||||| ||||||||||||||||| ||||||| ||||
Sbjct  1970783  GGCGGTCAGGAGCGGCTGGGCGCGGAT-AAGGCCGAGCCCGACCGTTGCCAGAGATTCCA  1970725

Query  121      CCGCCGCACGGCGTACACGGGCGCGGCGCGCTTCCCGGGCGTTGATGATCGACATCGCGA  180
                |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1970724  CCGCCGCACGGCGTTCACGGGCGCGGCGCGCTTCCCGGGCGTTGATGATCGACATCGCGA  1970665

Query  181      CCAGGGCGATGATGGCGATCAGCAGCGCGCAGGCGGCCAGGAAGAGGGCAGCCCCGACTG  240
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1970664  CCAGGGCGATGATGGCGATCAGCAGCGCGCAGGCGGCCAGGAAGAGGGCAGCCCCGACTG  1970605

Query  241      GGCCGTAGACGCCGGCGAGCCAGATGGCGCCGGCGGTGACGGCCAGCGCGTAGGCGGTGA  300
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1970604  GGCCGTAGACGCCGGCGAGCCAGATGGCGCCGGCGGTGACGGCCAGCGCGTAGGCGGTGA  1970545

Query  301      GAAGGAAGAGCACGGCAAGCGCAATGAAGATGCTATTGCGCTTGGCGCGCGCAACGGTCC  360
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1970544  GAAGGAAGAGCACGGCAAGCGCAATGAAGATGCTATTGCGCTTGGCGCGCGCAACGGTCC  1970485

Query  361      GGTGCACGCTTGCGCCCGTCAGCAGGCTGAGGATCGAAAACAGCATCGCGCTTCCTCAGC  420
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1970484  GGTGCACGCTTGCGCCCGTCAGCAGGCTGAGGATCGAAAACAGCATCGCGCTTCCTCAGC  1970425

Query  421      GCCGGGCGAGGAAGGCGATGGCAAGGCCGAAAGCGGCTGCGGCGCCGATGGTTGCTAGCG  480
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1970424  GCCGGGCGAGGAAGGCGATGGCAAGGCCGAAAGCGGCTGCGGCGCCGATGGTTGCTAGCG  1970365

Query  481      GATGCTTGCGTACGGTATCGCGCATTTCCGTCGCGCCGCGCTCATAGCCGTGCTGCAATT  540
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1970364  GATGCTTGCGTACGGTATCGCGCATTTCCGTCGCGCCGCGCTCATAGCCGTGCTGCAATT  1970305

Query  541      CGCGCAGCAGATCCTCGCTGCGGCCAAGCAGTTCTTCGTACCCGGCGCCGGCTTGGCTGC  600
                |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Sbjct  1970304  CGCGCAGCAGATCCTCGCTGCGGCCAAGCAGTTCTTCGTAGCCGGCGCCGGCTTGGCTGC  1970245

Query  601      GGATTT-CTC-CCT-GGTG  616
                |||||| ||| ||| ||||
Sbjct  1970244  GGATTTTCTCGCCTTGGTG  1970226
```

3’ DNA sequence alignment for the *Hmu*V knockout construct, query line is the Plasmid product and subject line is R. leguminosarum genome.



3' DNA sequence alignment for the *Hmu*U knockout construct, query line is the Plasmid product and subject line is R. leguminosarum genome.

Rhizobium leguminosarum strain ATCC 14479 chromosome, complete genome

Sequence ID: NZ_CP030760.1 **Length:** 4883137 **Number of Matches:** 1

Range 1: 1972107 to 1972442 GenBank Graphics ▼ Next Match ▲ Previous

Score	Expect	Identities	Gaps	Strand
560 bits(303)	4e-160	334/347(96%)	11/347(3%)	Plus/Plus

```
Query  1        TATGCAGGAGCCATGAATGACACTTTACTACTCGGATTGCGCCGCACCTTCCTTATCTGC  60
                ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Sbjct  1972107  TATGCAGGAGCCATGAATGACACTTT-CTACTCGGATTGCGCCGCACCTTCCTTATCTGC  1972165

Query  61       GCCGCTATTCCCGCGCCCTTACCGGCACTCAGACTTCGGGCGACGCTTACGTCGCCGCCG  120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1972166  GCCGCTATTCCCGCGCCCTTACCGGCACTCAGACTTCGGGCGACGCTTACGTCGCCGCCG  1972225

Query  121      TTCTCGAAGCCATCATCGCCGATCTGTCGATTTTCCCGGATACGGCGAATGACCGGGTCG  180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1972226  TTCTCGAAGCCATCATCGCCGATCTGTCGATTTTCCCGGATACGGCGAATGACCGGGTCG  1972285

Query  181      CACTCTACAAACTGTTCACGCAACTGTTTGGTTCCACCGCAGTCCAGATTCCAGAGCCGA  240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1972286  CACTCTACAAACTGTTCACGCAACTGTTTGGTTCCACCGCAGTCCAGATTCCAGAGCCGA  1972345

Query  241      CCTCGCCCTATGCCTGGGAGCAACGCGCCCACGCTGAACCTTGCCAAGGTTCCCCGCCGC  300
                ||||||||||||||||||||||||||||| ||||||||||||||||||||||   ||||||
Sbjct  1972346  CCTCGCCCTATGCCTGGGAGCAACGCG-CCACGCTGAACCTTGCCAAGGTTT--CGCCGC  1972402

Query  301      GCGCCCGTCAGGCCCTTTCCTGCTTCGCCCTCCCGTTANAGGAATTT  347
                ||||||||||||| || |||||| |||| ||| || | || |||||
Sbjct  1972403  GCGCCCGTCAGGCC-TT-CCTGCT-CGCC-TCC-GT-AGAG-AATTT  1972442
```

5' DNA sequence alignment for the sensor histidine kinase knockout construct, query line is the Plasmid product and subject line is R. leguminosarum genome. Primer used for sequencing is SOEF2SK

Rhizobium leguminosarum strain ATCC 14479 chromosome, complete genome

Sequence ID: NZ_CP030760.1 **Length:** 4883137 **Number of Matches:** 1

Range 1: 1974788 to 1975038 GenBank Graphics ▼ Next Match ▲ Previous

Score	Expect	Identities	Gaps	Strand
327 bits(177)	3e-90	239/266(90%)	16/266(6%)	Plus/Minus

```
Query  3        GGATATTTCCCTGGCATCATATGATAGTGACGGAATGGCCCTTGACGGCCAGGCGATGGG  62
                ||||  |||||||||||||||| ||||||||||||||||||||||||||||||||||||
Sbjct  1975038  GGATTTTTCCCTGGCATCATA-GATAGTGACGGAATGGCCCTTGACGGCCAGGCGATGGG  1974980

Query  63       CGGCGGCAAGGCCGGCTGGGCCTGCGCCGACAACGGCGATCGTCTTTCCGGTCGGTTCCG  122
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct  1974979  CGGCGGCAAGGCCGGCTGGGCCTGCGCCGACAACGGCGATCGTCTTTCCGGTCGGTTC-G  1974921

Query  123      GCCCCGGGCATAGAACTGCTTGCCCCGGCCTGCCATGGCGGCCATTCCGGGTCCGCATTA  182
                | |||||||||||||||||||  |||||||| |||||||| || | || || |||| |
Sbjct  1974920  G-CCCGGGCATAGAACTGCTTGCC--GGCCTGC-ATGGCGGC-AT-C-GG-TC-GCAT-A  1974871

Query  183      ACCGCCTGCAGGCCGACCGATCTCGACGGGCCGCtccccggccgtgttgcccacccgggc  242
                || || |||||| |||||||||||||||||||| |||||||||| ||| | |||
Sbjct  1974870  AC-GC-TGCAGG-CGACCGATCTCGACGGGCCGCTCCTCGGCCGTGTTGCGCACACAGGC  1974814

Query  243      c-gttcccccagtcctcccgtgggac  267
                | |||| |  ||| || | |||||||
Sbjct  1974813  CTGTTCGCAGAGTTCTTCGGTGGGAC  1974788
```

3' DNA sequence alignment for the sensor histidine kinase knockout construct, query line is the Plasmid product and subject line is R. leguminosarum genome. Primer used for sequencing is SOERB

Potential Sensor Kinase Merodiploid 3’ fragment DNA BLAST alignment (Primer SOEF2SK)

Rhizobium leguminosarum strain ATCC 14479 chromosome, complete genome

Sequence ID: NZ_CP030760.1 **Length:** 4883137 **Number of Matches:** 1

Range 1: 1974087 to 1974678 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
1046 bits(566)	0.0	585/594(98%)	4/594(0%)	Plus/Plus

```
Query  1        AGATCTTCAAGCGAGTGTGAAAGCCCTGGAGACAATCGAGAACGAAAAACAGCCGGTGCG  60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974087  AGATCTTCAAGCGAGTGTGAAAGCCCTGGAGACAATCGAGAACGAAAAACAGCCGGTGCG  1974146

Query  61       AGGGCGGCGCACCGGCTGTCTTCACTCACCGGGAGGGGACCGTGAGTACGTCACGGATAG  120
               |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Sbjct  1974147  AGGGCGGCGCACCGGCTGTCTTCACTCACCGGGAGGGGACCGTGAGTACGTC-CGGATAG  1974205

Query  121      TGGGTTGGGGGCGCGCATGTTGGGTCGATGCGATCACCATCCGGATATCGCAATAACGAC  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974206  TGGGTTGGGGGCGCGCATGTTGGGTCGATGCGATCACCATCCGGATATCGCAATAACGAC  1974265

Query  181      GACATCAAACTTTGGTTCCCTGCCATCCGaaaaaaaTTCGACTTTTTTGAATCTTTTTTC  240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974266  GACATCAAACTTTGGTTCCCTGCCATCCGAAAAAAAATTCGACTTTTTTGAATCTTTTTTC  1974325

Query  241      GGATGCCCTTCCCTCGTCGCGCGTTCAGCCGTTTTACCGATGCAGGCCGAATAAATTCTT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974326  GGATGCCCTTCCCTCGTCGCGCGTTCAGCCGTTTTACCGATGCAGGCCGAATAAATTCTT  1974385

Query  301      CCAGCAAAAACAGCGGATTACTGGCGTGACGGCAAGAGAAAATCGTCAAAATTTTACCGT  360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974386  CCAGCAAAAACAGCGGATTACTGGCGTGACGGCAAGAGAAAATCGTCAAAATTTTACCGT  1974445

Query  361      TTGATAAATTTTTAAACCTGAGTTACGCTTTCCCTCACAGGCCGTTTACCAATAATGAGG  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974446  TTGATAAATTTTTAAACCTGAGTTACGCTTTCCCTCACAGGCCGTTTACCAATAATGAGG  1974505

Query  421      GAACTTCAATGGAACGACTGGAAACTGGGATTCATGCCGGCCGGCTTTCGCCCGCCGAGT  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974506  GAACTTCAATGGAACGACTGGAAACTGGGATTCATGCCGGCCGGCTTTCGCCCGCCGAGT  1974565

Query  481      ATGAGGCTAATTTTTCCGATCTTCATCCGCGCCTCGACAATCACGAGGCGCTGGTCCCCG  540
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Sbjct  1974566  ATGAGGCTAATTTTTCCGATCTTCATCCGCGCCTCGACAATCACGAGGCGCTGGTCGCCG  1974625

Query  541      CCGACCGCTG-TATTCCTGTTATGACCCCCC-TGCATGACCGGCCNGTCCCACC  592
               |||||||||| |||| |||||||||| | || ||||||| ||||| ||||||||
Sbjct  1974626  CCGACCGCTGTTATTTCTGTTATGACGCGCCGTGCATGA-CGGCCTGTCCCACC  1974678
```

Potential Sensor Kinase Merodiploid 3’ fragment DNA BLAST alignment (Primer SOERB)

Rhizobium leguminosarum strain ATCC 14479 chromosome, complete genome

Sequence ID: NZ_CP030760.1 **Length:** 4883137 **Number of Matches:** 1

Range 1: 1974454 to 1975039 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
957 bits(518)	0.0	583/610(96%)	24/610(3%)	Plus/Minus

```
Query  1        CGGATTTATCCCTGGCATCATAGTATAGTGACGGAATGGCCCTTGTACGGCCAGGCGATG  60
               ||||||| ||||||||||||||| |||||||||||||||||||| ||||||||||||||
Sbjct  1975039  CGGATTTTTCCCTGGCATCATAG-ATAGTGACGGAATGGCCCTTG-ACGGCCAGGCGATG  1974982

Query  61       GGCGGCGGCAAGGCCGGCTGGGCCTGCGCCGACAACGGCGATCGTCTTTCCGGTCGGTTC  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974981  GGCGGCGGCAAGGCCGGCTGGGCCTGCGCCGACAACGGCGATCGTCTTTCCGGTCGGTTC  1974922

Query  121      GGCCCGGGCATAGAACTGCTTGCCGGCCTGCATGGCGGCATCGGTCGCATAACGCTGCAG  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974921  GGCCCGGGCATAGAACTGCTTGCCGGCCTGCATGGCGGCATCGGTCGCATAACGCTGCAG  1974862

Query  181      GCGACCGATCTCGACGGGCCGCTCCTCGGCCGTGTTGCGCACACAGGCCTGTTCGCAGAG  240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974861  GCGACCGATCTCGACGGGCCGCTCCTCGGCCGTGTTGCGCACACAGGCCTGTTCGCAGAG  1974802

Query  241      TTCTTCGGTGGGACAGACGCGGGCGCACATGCCGCCGAGGATGTTCTGGTCGAAGATCGT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974801  TTCTTCGGTGGGACAGACGCGGGCGCACATGCCGCCGAGGATGTTCTGGTCGAAGATCGT  1974742

Query  301      CTTTGCCGAGCCGATCGGATTGCCGGTCGAAATCTGGCGGATGAACAGCGGAATGTCGAT  360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974741  CTTTGCCGAGCCGATCGGATTGCCGGTCGAAATCTGGCGGATGAACAGCGGAATGTCGAT  1974682

Query  361      CGAGGTGGGACAGGCCGTCATGCACGGCGCGTCATAACAGAAATAACAGCGGTCGGCGGC  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974681  CGAGGTGGGACAGGCCGTCATGCACGGCGCGTCATAACAGAAATAACAGCGGTCGGCGGC  1974622

Query  421      GACCAGCGCCTCGTGATTGTCGAGGCGCGGATGAAGATCGGAAAAATTAGCCTCATACTC  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974621  GACCAGCGCCTCGTGATTGTCGAGGCGCGGATGAAGATCGGAAAAATTAGCCTCATACTC  1974562

Query  481      GGCGGGCGAAAGCCCGGCCGGCATGAATCCCCAGTTTCCAGTTCGTNCCATTTGAAATTC  540
               ||||||||||||||| |||||||||||||||| |||||||||| ||| ||||| ||| |||
Sbjct  1974561  GGCGGGCGAAAGCC-GGCCGGCATGAATCCC-AGTTTCCAGT-CGTTCCATT-GAAGTTC  1974506

Query  541      CCCCTCATTTATTGGGGTAAACCGGCCCCTGGTGAGGGGGAAAAGCCGTTAAACTTCAAG  600
               ||  ||||| |||||  ||||| ||||  || ||||||  ||| || || || || || | 
Sbjct  1974505  CC--TCATT-ATTGG--TAAAC-GGCC--TG-TGAGGG--AAA-GC-GT-AA-CT-CA-G  1974463

Query  601      GGTTTaaaaa  610
               | ||||||||
Sbjct  1974462  G-TTTAAAAA  1974454
```

Potential Sensor Kinase Merodiploid 5’ fragment DNA BLAST alignment (Primer SOEFB)

Rhizobium leguminosarum strain ATCC 14479 chromosome, complete genome

Sequence ID: NZ_CP030760.1 **Length:** 4883137 **Number of Matches:** 1

Range 1: 1974087 to 1974678 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
1046 bits(566)	0.0	585/594(98%)	4/594(0%)	Plus/Plus

```
Query  1        AGATCTTCAAGCGAGTGTGAAAGCCCTGGAGACAATCGAGAACGAAAAACAGCCGGTGCG  60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974087  AGATCTTCAAGCGAGTGTGAAAGCCCTGGAGACAATCGAGAACGAAAAACAGCCGGTGCG  1974146

Query  61       AGGGCGGCGCACCGGCTGTCTTCACTCACCGGGAGGGGACCGTGAGTACGTCACGGATAG  120
               |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||| 
Sbjct  1974147  AGGGCGGCGCACCGGCTGTCTTCACTCACCGGGAGGGGACCGTGAGTACGTC-CGGATAG  1974205

Query  121      TGGGTTGGGGGCGCGCATGTTGGGTCGATGCGATCACCATCCGGATATCGCAATAACGAC  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974206  TGGGTTGGGGGCGCGCATGTTGGGTCGATGCGATCACCATCCGGATATCGCAATAACGAC  1974265

Query  181      GACATCAAACTTTGGTTCCCTGCCATCCGaaaaaaaTTCGACTTTTTTGAATCTTTTTTC  240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974266  GACATCAAACTTTGGTTCCCTGCCATCCGAAAAAAATTCGACTTTTTTGAATCTTTTTTC  1974325

Query  241      GGATGCCCTTCCCTCGTCGCGCGTTCAGCCGTTTTACCGATGCAGGCCGAATAAATTCTT  300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974326  GGATGCCCTTCCCTCGTCGCGCGTTCAGCCGTTTTACCGATGCAGGCCGAATAAATTCTT  1974385

Query  301      CCAGCAAAAACAGCGGATTACTGGCGTGACGGCAAGAGAAAATCGTCAAAATTTTACCGT  360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974386  CCAGCAAAAACAGCGGATTACTGGCGTGACGGCAAGAGAAAATCGTCAAAATTTTACCGT  1974445

Query  361      TTGATAAATTTTTAAACCTGAGTTACGCTTTCCCTCACAGGCCGTTTACCAATAATGAGG  420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974446  TTGATAAATTTTTAAACCTGAGTTACGCTTTCCCTCACAGGCCGTTTACCAATAATGAGG  1974505

Query  421      GAACTTCAATGGAACGACTGGAAACTGGGATTCATGCCGGCCGGCTTTCGCCCGCCGAGT  480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1974506  GAACTTCAATGGAACGACTGGAAACTGGGATTCATGCCGGCCGGCTTTCGCCCGCCGAGT  1974565

Query  481      ATGAGGCTAATTTTTCCGATCTTCATCCGCGCCTCGACAATCACGAGGCGCTGGTCCCCG  540
               |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||  
Sbjct  1974566  ATGAGGCTAATTTTTCCGATCTTCATCCGCGCCTCGACAATCACGAGGCGCTGGTCGCCG  1974625

Query  541      CCGACCGCTG-TATTCCTGTTATGACCCCCC-TGCATGACCGGCCNGTCCCACC  592
               |||||||||| |||| |||||||||| | || |||||||| ||||| ||||||||  
Sbjct  1974626  CCGACCGCTGTTATTTCTGTTATGACGCGCCGTGCATGA-CGGCCTGTCCCACC  1974678
```

Potential Sensor Kinase Merodiploid full sensor kinase gene amplification (primer SKF)

Rhizobium leguminosarum strain ATCC 14479 chromosome, complete genome

Sequence ID: NZ_CP030760.1 **Length:** 4883137 **Number of Matches:** 1

Range 1: 1972765 to 1973158 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
640 bits(346)	0.0	383/399(96%)	9/399(2%)	Plus/Plus

```
Query  10       GTGATCTAGCATCACCGCATATCCCGGAACGGCTTCTGACCGGCGAGCGCCCGGAACCGA  69
                |||||||  ||||||||| | ||||||||||||||||||||||||||||||||||||||
Sbjct  1972765  GTGATCT-TCATCACCGC-TTTCCCGGAACGGCTTCTGACCGGCGAGCGCCCGGAACCGA  1972822

Query  70       CTTTCCTTGTCACCAAGCCCTTCAATCCCGACATGGTCAAGGCACTGATCAGCCAAGCGC  129
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1972823  CTTTCCTTGTCACCAAGCCCTTCAATCCCGACATGGTCAAGGCACTGATCAGCCAAGCGC  1972882

Query  130      TTTTCTTCAATGAATCGACCAGAGTAGCCGCCTGAGACGCAATTTTCCGGCCTTCGGAAC  189
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1972883  TTTTCTTCAATGAATCGACCAGAGTAGCCGCCTGAGACGCAATTTTCCGGCCTTCGGAAC  1972942

Query  190      CAAATCGCCAAAGCAGGCGTTCCGGTGCTACCGGGACGCTCCGGTGGCTTTAACGGTTTT  249
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1972943  CAAATCGCCAAAGCAGGCGTTCCGGTGCTACCGGGACGCTCCGGTGGCTTTAACGGTTTT  1973002

Query  250      TTGCAGCGCTTTGTTCTAAAGTTGGCAGGCGGTGTCTGTAAGGGCGCTCCCTTCAGCGAC  309
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1973003  TTGCAGCGCTTTGTTCTAAAGTTGGCAGGCGGTGTCTGTAAGGGCGCTCCCTTCAGCGAC  1973062

Query  310      GTTAAAGATGTCACAAGGAAAGGCGGGCCGCGTGAATACGACTGAACCATTGTTCGGCTT  369
                ||| |||||||||||||||||| |||||||||||||||||||||||||||||| |||| |
Sbjct  1973063  GTTCAAGATGTCACAAGGAAAGGC-GGCCGCGTGAATACGACTGAACCATTGTCCGGCAT  1973121

Query  370      TGTTCTTTCACCTTT-AAGGCAA-GC-G-AATGATGGAA  404
                 |  |||||||||  ||||||| || | ||||||||||
Sbjct  1973122  -GC-CTTTCACCTTCGAAGGCAAAGCCGCAATGATGGAA  1973158
```

Potential Sensor Kinase Merodiploid full sensor kinase gene amplification (primer SKR)

Rhizobium leguminosarum strain ATCC 14479 chromosome, complete genome

Sequence ID: NZ_CP030760.1 **Length:** 4883137 **Number of Matches:** 1

Range 1: 1974107 to 1974231 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
122 bits(66)	8e-29	112/131(85%)	15/131(11%)	Plus/Minus

```
Query  13       GACACAATA-GCGTCG-CCCCAACCCAGCTATACACGGAACGTACTCACGGTACCCCTCC  70
                ||| ||| | ||| || |||||||||| |||| | ||| |||||||||||||| |||||||
Sbjct  1974231  GACCCAACATGCG-CGCCCCCAACCCA-CTAT-C-CGG-ACGTACTCACGGT-CCCCTCC  1974178

Query  71       CGGTGAGTGAAGACAGCCGGTGCGCCGCCATAGCACCGGC-GTTT--CGT-CTCGAT-GT  125
                |||||||||||||||||||||||||||| | |||||||| ||||  ||| ||||||  ||
Sbjct  1974177  CGGTGAGTGAAGACAGCCGGTGCGCCGCCCTCGCACCGGCTGTTTTTCGTTCTCGATTGT  1974118

Query  126      -TC-AGGGCTT  134
                 || |||||||
Sbjct  1974117  CTCCAGGGCTT  1974107
```

R. leguminosarum ATCC 14479 *ShmR* homolog nucleotide alignment with *S. meliloti's ShmR* gene

Sequence ID: Query_59059 Length: 2232 Number of Matches: 1

Range 1: 19 to 2106 Graphics ▼ Next Match ▲ Previous M

Score	Expect	Identities	Gaps	Strand
881 bits(976)	0.0	1476/2115(70%)	51/2115(2%)	Plus/Plus

```
Query  1706821  CGACACCTTGAACGTACGGCCCGGCTCCGAATACCACTCCTGCGGCTGCGACGCCGTCAC  1706880
                |||  ||  ||||||||  ||  |||||  ||  |||||   | ||  |||||||  |   | ||
Sbjct  19       CGAGACTTTGAACGTGCGTCCCGGTTCGGAATAGAAGTCGCGCGGCTGAGCGGACGAT--  76

Query  1706881  CGAGGTCAGGTTGACGTCGCGCACGGCGAGCGCATTGTAATGCTCCTGATCGAAGATGTT  1706940
                 ||||   ||||  ||||  |  ||  |||  |||||||  |||  |||  ||  ||  ||||||||||
Sbjct  77       -GAGGCGAGGTCGACGCCACGTACGCCGAGCGCGTTGAAATACTTCTTGTCGAAGATGTT  135

Query  1706941  GTAGACACCTCCCTGGATCCGCAGACCCTTCAACTGTTCCGGCGTCCACCATGCCGTCAG  1707000
                 |||||  ||   |||   |  ||  ||   |||||  ||    |||  ||||||||||||   |  ||||
Sbjct  136      ATAGACCCCGGCCTCTACCCTCAATCCCTTGAAGCTTTCGGGCGTCCACCACCCGGTCAT  195

Query  1707001  ATCGACGACACCATAACTGGGCGCATCGAATGTGGTCGCAA-CATCGTCGTCAGGCATGG  1707059
                 |||  ||||  |||  ||  |  |||  |||||||||  ||   ||||    |  |   |||||  |||||||
Sbjct  196      ATCCACGAGACCGTAGCCGGGGGCATCGAAGGTT-TCGCTGTCGTTATCGTCCGGCATGG  254

Query  1707060  CTGCCGAAAGCGTCGAGGAAAGATCGAAACCGTAGTTGTCGTTGCTCCAACCGCCGCCGA  1707119
                 |||||||  |  |||||   ||  |  |||||  ||||  ||  |  ||  |  |||   |  ||||||||||
Sbjct  255      CTGCCGAGACCGTCGTCGAGACATCGACACCGAAGGTTTCCTGGCTGTAGCCGCCGCCGA  314

Query  1707120  TGATCGCCTTGAAGGGAGCAACCGATCGCAAATATTTGCCTGTATCCTCGTTGCGGCCAT  1707179
                 ||||||||||||||  ||  ||  ||||   ||||       ||||     ||||||||    |||||
Sbjct  315      TGATCGCCTTGAATGGCGCGACCGTCCGCAGCCGCGTGCCCTCGTCCTCGTTCTTGCCAT  374

Query  1707180  AGGCATAGGCAAGCGACGCATGCAGGTTGATGCCGTTGTTGAACGTCTTGACGGCACTGG  1707239
                 |||  |||||||||||||||||  |||||||  ||||||||  |||  ||  ||  |   ||  ||
Sbjct  375      AGGAGTAGGCAAGCGACGCATGGAGGTTGAAGCCGTTGTCGAAGGTTTTCAACGCGCTCA  434

Query  1707240  CTTCAACACCCGAAATCGTGGCGGCAGACACGTTGGTGTAATTAAATGTCGTGAAGCCGG  1707299
                   ||  |  ||||||  ||   |||          |||||||   ||  ||  |||   |  |||   ||||
Sbjct  435      ACTCGATACCCGAGATTCGGGCCTTGTTGACGTTGGCATATTTGAATTCCCTGATCCCGG  494

Query  1707300  TCGCGTCGACGCTCG--TCACCG-TCTGGATAAAGTTCTGGTAGCGGGTGTGGAAAGCCG  1707356
                 |   ||   |  |  |  ||  |||  |  |   ||  ||  ||  |   |||     ||||||| ||||
Sbjct  495      TGTCGCTGTTGATGGAATCGCCGGTTTCAATGAAATTGTCATAGATATTGTGGAACGCCG  554

Query  1707357  CAATCCGACCGGTAAAATCGCCCGTGTCGAAATTGGCGCCGATTTCGACGCCGCGGCCGA  1707416
                 |  |  |||  |||  |  |   ||  ||  ||||||||||||||  |||||  ||||   ||   ||||
Sbjct  555      CGACCCGGCCGCTCAGCTCACCGGTGTCGAAATTGGCACCGATCTCGAATCCTTTGCCGG  614

Query  1707417  TTTCCGGTTCCAGATCAGGATTGCCAAGCTGGGCATAACGACCCGTCGGGTTGTAGAAGC  1707476
                 |||||||  |  |||||  ||||||||  ||||||||||     ||   ||||||||||||||
Sbjct  615      TTTCCGGCTTGAGATCCGGATTGCCGAGCTGGGCATAGTTTCCGAACGGGTTGTAGAAGC  674

Query  1707477  GGCTGTAGAGCTCGTCCACGGTGGGTGCACGGAAACCGACTGCCAGCTGTGCGTAAATCT  1707536
                 ||||  ||  ||||||||||||||  ||  ||  |||||   |  |  |  |   ||||||  |  |  ||
Sbjct  675      GGCTATAAAGCTCGTCCACGGTCGGCGCCCGGAAGGCCATCGACCATTGTGCGAACAGCT  734

Query  1707537  GCACGTCAGGTGTCAGCTCGTATGTCGCAAGAATCTTCGGCGACAGGCCGGCTTCCGTCC  1707596
                   ||  ||  ||  ||  |   ||  ||  |||||  ||  |  ||||||||||||||    |   |    ||
Sbjct  735      CGACATCCGGCGTAACGTCATAGGTCGCGAGGAGCTTCGGCGACAGGCGAACCCCATCCC  794

Query  1707597  GATCGTTCAGGTCCCCGAAACGGGTAAGC---CCGGTATTGCTTGCGAAACTTCCGCCGG  1707653
                   |  |||||  ||||||||    |  |    |  ||  |||||       ||||  |  |  |    |
Sbjct  795      TGGCCTTCAGATCCCCGAA---GATTCCCGATCCTGTATTCGACTCGAAGCCTGCATTGA  851

Query  1707654  TC-GAGGGATTGTAGTTGAACCAGTCGAAGCGGAAGCCGGGTGTCAGCGCGAAGCCGGTA  1707712
                 |  |||||  |   ||  |  ||||||||||||||  ||  ||  ||  ||  |||||  ||   ||  |
Sbjct  852      GCTGAGGG-TCATATTGGAACCAGTCGAAGCGCAACCCCGGCGTGAGCGCAAACGCGCTG  910
```

```
Query  1707713  TTGCCGATTTCGATCTTGTCTTCGAAGGTCAGCCCAAGATTCTGGCTGTCGACATTGGGC  1707772
Sbjct  911      TCGCCGACCGAAATGCGGTCTTCCAATATGGCGCCGACCGTCATGCTCCTCACGTCGGGG  970

Query  1707773  ACCTCCGCCTGGTTGTTCAGCGCCGGGCACGTTGTGGGCGTCGGGCAAAGGGCCCAGCTA  1707832
Sbjct  971      ACTTCGGACTGGTTGTTGAGCGCCGGGCATGTCGCCGGCGTCGGGCAAACCGCCGAGGTA  1030

Query  1707833  TACTGGCTCCAGCTCGAAACGCCC-ACGTCGAGACCAACGCGAACGGAATGGCTGAGACC  1707891
Sbjct  1031     TACTGGCTCCA-CTCGGAACGGGCGACATCGAGGCCGAAGGTTAGCGAATGATCGAAACC  1089

Query  1707892  GGAATATTCGAAATCCTTCGTCGC-CGTGCCGCTGAAGCCCCAGGTTTCGTTTTCTATCT  1707950
Sbjct  1090     GCCGGTTTCGAAATCCTTGCCTGCGCGT-CCGTTGAAACCCACGCTCTCGTTGCTGATCG  1148

Query  1707951  GGTTGTTACGGCCATAAGCCACATTGGCAGTCGTACGGCCTCTGCTGCCGGCTT--CCTT  1708008
Sbjct  1149     AATTGTCGCGCCCGTAGGGCACATCGGCGATCGTCCGGCCATTCGAGCCGGAGGAGCGTT  1208

Query  1708009  CTTCAAATCCAACCAGTAAAGCGTGGCTCGCGCACTGCTGAAGAATGCGTCGGAAGACTG  1708068
Sbjct  1209     C--CAGATCCTGCCAGTAGAGCGATGCCCTGGCGAAGCTGAAGAAGTCGTCCGACGAGGC  1266

Query  1708069  CGCCTCATAATCGTAATCGAGAGAAACTCGGTCGCGGTCGCGCAATTCGCGGCCGTCATA  1708128
Sbjct  1267     GGCCTCGAAATCATAGTCCAGCGAGACGCGCTTGCGGTCGCGGTCTTCAAAACCGGTATA  1326

Query  1708129  ATTGTCGATCAGGAAGTTGCGGGGCGTAGTGCCGCCCTGGAGCTCGCGCAGGTCGGTCTT  1708188
Sbjct  1327     GTCGCCGATCTTGTACCGGCGCCCCT--GTTCCGCTCTG----------AGATCATTGTC  1374

Query  1708189  GAGATCGCGCCGGAAGCGTTCGGCCGTCAGGCCGATGCGATGACCACCTTCGAGCTCCTG  1708248
Sbjct  1375     GGCGTCGCGGCGGAAGCTTTCCGCCGTCAGGCCGATTCTGTGACCGCCTTCGAGTTCATG  1434

Query  1708249  GCGCAGCTTGAAGAGCAGGTTGTGCTGGTCGAAATCGGCGGGATCCGCCTCGGTACGAGG  1708308
Sbjct  1435     CCGAAACTTGAAGAGCAAGTTGTTCTGGTCGAAATCGGTAGGGTTGGGCTCCGTGCGGG-  1493

Query  1708309  ACGG--CCATAGCTGTC-ATTGTCACCCATATTGTCGCGTTCATGGCCCTTGCGGTAGCC  1708365
Sbjct  1494     -CGGATCCGTAGCCACCGACCGTC-CCCTCGTTATCCCGTTCATTCCCCTTGCGGTAGCT  1551

Query  1708366  GCCCTGAAACAGGATCGACGTGTTGCCGATCTTCTTGGCAGCCGCCGCCGAGTCGGAAAT  1708425
Sbjct  1552     GCCCTGAAAGAGAACGGAGGTCTGGCCGAATCTGTGGGCTCCAGCAACCGAGCCCGCAAT  1611

Query  1708426  GCTGCGGTCTTCGCTGTCGTATGTCGACTTGACGATCGCACCCCAATCCCGGCCTTCGGG  1708485
Sbjct  1612     GCTGTCGTCTTCGCTGTCATAGATGGACCTGAAGATCGCGCCCCAATCCTTTCCGTCGGA  1671

Query  1708486  AATAAGATCCTCCGGCTCGAGCGTATTGAGTACGATGGCGCCGCCGAGCATTCCCGAGCC  1708545
Sbjct  1672     TATGACATCCTCGGGCTCGAGCGTGCGCAGCACCACGGCGCCCCCGAGCATGCCGGAACC  1731

Query  1708546  ACCCTTGCTCGAATCCGCGCCTCGCACAATATCGAGCGAGGAAAGCGAATCGAAATCGAA  1708605
Sbjct  1732     GCCACGGCTCGAATCCGCCCCGCGCACTATGTCGACCACGGAGAGCGAATTGAAATCGAA  1791

Query  1708606  GGTGTCGCCGCCGCCATTGGCGTTGGCCGGCGCGAAAGCGCCTTGGCGCGAACTATTCGA  1708665
Sbjct  1792     CATGTCCCCGCCGCCGTTCGCATTGATCGAGGCAAAGGCACCCTGACGGGCGCTGTTCGA  1851

Query  1708666  AATATACGGGATCGGAATGCCGTCGATGGTGGTCAGGATGCGAGCGCCGGAAAGGCCGCG  1708725
Sbjct  1852     GATATAGGGGATCGGAACGCCGTCGATCGTGGTCACGATACGAGGACCGGAAAGACCGCG  1911

Query  1708726  CAGGTTGAAGCCCGCGTCGCCACGTGAATAGTTCACACCTGCATCGACGCTGCGGCCGAT  1708785
Sbjct  1912     AAGATTGATACCGAAGTCGGCACGCGATGCATTTATCCCTGCGTCGACGCTGCGGCCGAG  1971

Query  1708786  ATCGTCAAAATTGGTGACCTGCTTTTCTTCCAGCGTCTTGGCGGTGATCTCGGTCGCAAG  1708845
Sbjct  1972     ATCGTCGAGATCCGTGACCTGCTTCTCCTCCAGGGTCTTTGCGTCGATCTCGGTCGCAAG  2031
```

```
Query  1708846  CGGCGTGTCGGCGACGCTGCCCGGC---------GCCACGCGTTTGCCCTTGACGACGAT  1708896
                ||||||  ||  ||||  ||  |||   ||          ||||  |||  |  |||  ||   |||||||
Sbjct  2032     CGGCGTATCCGCGATGCCGCCACGCTGCGCACCTGCCAGGCGGTCGCCTTTCGCGACGAT  2091

Query  1708897  CTTCTGCAGGACGGT  1708911
                ||||| |||  |||||
Sbjct  2092     CTTCTTCAGCACGGT  2106
```

R. leguminosarum Siderophore production CAS media assay 24 hour (From left to right: 200µM 2'2'-dipyridyl, 10µM hemin, 15µM hemin, 30µM hemin.

R. leguminosarum Siderophore production CAS media assay 48 hour (From left to right: 200µM 2'2'-dipyridyl, 10µM hemin, 15µM hemin, 30µM hemin.

R. leguminosarum Siderophore production CAS media assay 72 hour (From left to right: 200µM 2'2'-dipyridyl, 10µM hemin, 15µM hemin, 30µM hemin.

CAS media indicator dye control (.5M EDTA)

VITA

JOHN R. LUSBY

Education: Carter High School, Strawberry Plains, TN

A.S. Roane State Community College, Oak Ridge, TN, 2014

B.S. Health Sciences, East Tennessee State University, Johnson City, Tennessee, 2018

M.S. Biology, concentration in Microbiology East Tennessee State University, Johnson City, Tennessee, 2021

Professional Experience: Graduate Assistant, East Tennessee State University, College of Health Sciences, 2017-2019