IDENTIFICATION OF NOVEL COMPONENT OF RIBOFLAVIN BIOSYNTHETIC PATHWAY IN INVASIVE BACTERIA

by

Omkar Raghunath Kshirsagar

Submitted in partial fulfilment of the requirements

for the degree of Master of Science

at

Dalhousie University

Halifax, Nova Scotia

March 2019

I dedicate this thesis to my all teachers and mentors and

his highness Chhatrapati Shivaji Maharaj

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Riboflavin biosynthetic pathway of *Sinorhizobium meliloti*1021 has a potential alternative enzyme ArfB, essential in riboflavin production for intracellular use. ArfB forms complexes with riboflavin biosynthetic enzymes in *S. meliloti*1021. The objective of this study was to identify the novel enzymes involved in the riboflavin biosynthesis of riboflavin prototrophic revertant strain Rm1021Δ*arfB*xOa *arfB* and analyze the free-living phenotypes and the symbiotic phenotypes. Whole genome sequencing was conducted and several SNPs and indels were identified in the genome of revertants, specifically in genes *phnM*, *SMc00185*, and *argG*. The revertant strains could not form effective symbiosis and exhibited lower growth-rate and flavin secretion compare to the control. Occasionally some revertants exhibited effective symbiosis. Secondary revertants were isolated and sequenced for identification of variations. Genetic modifications were observed on the SMb21442 and Oa *arfB* genes. Characterization of identified genes in future experiments will provide comprehensive understanding of the functions in riboflavin metabolism of α-proteobacteria.

LIST OF ABBREVIATIONS AND SYMBOLS USED

AA	Amino acid
ArfA/GCHIII	GTP cyclohydrolase III
ArfB	Formamide hydrolase
RF Auxotrophy	Bacterial strain require riboflavin supplementation for its growth
BNF	Biological nitrogen fixation
FAD	Flavin adenine dinucleotides
Fix -	Ineffective nitrogen fixation
Fix+	Effective nitrogen fixation
FMN	Flavin mononucleotides
FL	Flavin
GCHII	GTP Cyclohydrolase II
GCHIII	GTP Cyclohydrolase III
Indel	Insertion and deletion
N	Nitrogen
N_2	di-Nitrogen
N_2O	Nitrous oxide
NO_3^-	Nitrate
Oa	*Ochrobactrum anthropi*
RF Prototrophy	Bacterial strain does not require riboflavin supplementation for its growth
RBP	Riboflavin biosynthetic pathway
Revertant/Rev	Bacterial strain accumulated spontaneous mutation resulting riboflavin prototrophy
Rev_1	Rm1021Δ*arfB*xpCPP30 Oa*arfB*_#1.4 rev #1

Rev_5	Rm1021Δ*arfB*xpCPP30 Oa*arfB*_#1.4 rev #5
Rev_6	Rm1021Δ*arfB*xpCPP30 Oa*arfB*_#1.4 rev #6
Rev_8	Rm1021Δ*arfB*xpCPP30 Oa*arfB*_#1.4 rev #8
Rev_10	Rm1021Δ*arfB*xpCPP30 Oa*arfB*_#1.4 rev #10
Rev_5-1	Rm1021Δ*arfB*xpCPP30 Oa*arfB*_#1.4 rev #5-1
Rev_10-1	Rm1021Δ*arfB*xpCPP30 Oa*arfB*_#1.4 rev #10-1
Rev_10-2	Rm1021Δ*arfB*xpCPP30 Oa*arfB*_#1.4 rev #10-2
RF	Riboflavin
RibBA	3,4-Dihydroxy-2-butanone 4-phosphate synthase; GTP cyclohydrolase II
RibD	Bifunctional deaminase-reductase
RibE/C	Riboflavin synthase
RibH1/H2	6,7-dimethyl-8-ribityllumazine synthase
Rm1021	*Sinorhizobium meliloti* 1021
SNPs	Single nucleotide polymorphism
St^r	Streptomycin resistance
Tc^r	Tetracycline resistance
1.4	Rm1021Δ*arfBxpCPP30* Oa*arfB*_#1.4

ACKNOWLEDGEMENTS

First, I would like to thank my supervisor Dr. Svetlana Yurgel for encouraging, inspiring every day and believing in me. Her guidance throughout my research programs has been always valuable and supported me through every difficult situation. Her office door was always open for me any time during my programs, which always encouraged me to success in this research project.

I would like to express my gratitude to my committee members, Dr. Gefu Wang-Pruski and Dr. Balakrishnan Prithiviraj for encouragement, advice and comments to make this research project successful. Thank you for always being there to help me throughout the project. I would like to thank Dr. Wang-Pruski for believing in me and providing opportunity as a teaching assistant and a course marker.

I must acknowledge my lab members Priyanka Rupasinghe, Seun Esan, Melissa Mammoliti, Ashley Dussault, Dandan Wang and Jessica Correa for their friendly support. I would like to thank my friends Pushpsheel, Adil, Nadeem, Shruti, Bhushra, Rosina, Morgan, Chigozie, Mark, Joel for inspiring me every day. My cute Roshaan who always kept me delighted.

I would like to acknowledge love and sacrifice of all my family members they had shown in raising me. I dedicate this work to my late grandmother Gangubai. I wouldn't have been at this stage of my life if my father Shri. Raghunath Kshirsagar, mother Ujwala as my parents, my sisters Snehal, Archana, Priyanka and my brothers-in-law Vijay, Kiran and Rajendra who always had my back. This work was successful due to support of person whom I owe my heart, my wife Trupti. She stood and supported me whenever

there was ups and downs in my life. I love you and will always be grateful for what you have given to me.

Lastly, I thank readers for considering and finding this document useful for their scientific work. I would like to dedicate all my work to scientific community all over the world.

CHAPTER 1: INTRODUCTION

Vitamin B2 (riboflavin; RF), is a water-soluble vitamin and members of vitamin B complex. The vitamin B complex includes vitamins B1 (thiamin), B2 (RF), B3 (niacin), B5 (pantothenic acid), B6 (pyridoxine, pyridoxal and pyridoxamine), B8 (biotin), B9 (folate) and B12 (Cobalamin) (Roje, 2007). The RF is synthesized by plants, and microorganisms, but not by animals. Animals (including humans) depends on nutritional sources for RF (Fischer and Bacher, 2011). Flavin (FL) mononucleotides (FMN) and flavin adenine dinucleotides (FAD) are two cofactors of RF. These cofactors are found to be an important part of various metabolic processes including electron transport chain, biosynthesis of vitamins B6, B12 and B9 (Mansoorabadi *et al*., 2006; Roje, 2007; Yurgel *et al*., 2014). RF is an antioxidant that protects cells from oxidative stresses (Ashoori and Saedisomeolia, 2014). RF also plays an important role in plant-microbe interactions (Yang *et al*., 2002, Yurgel *et al*., 2014).

Nitrogen (N) fixing soil bacteria, commonly known as rhizobia, are capable of fixing atmospheric di-nitrogen (N_2) in association with legume plants (Moreno *et al*., 1990; LeVier *et al*., 2000; Paulsen *et al*., 2002). Biological N fixation (BNF) is an alternative to industrial N fixation. BNF is an inexpensive, sustainable and non-polluting substitute for industrial N. Legumes occupy 11% of agricultural land and contribute 25% of total fixed N (Roy *et al*., 2002). *Rhizobium*-legume symbiosis provides sustainable and ecologically friendly practices. It has potential for reclamation of marginal, saline and degraded lands in agriculture (Coba de la Peña and Pueyo, 2012).

The world population is increasing; in 2015, it was 7.3 billion and is projected to be increased by one billion in the next 15 years (United Nations, Department of Economic and Social Affairs, Population Division, 2015). In addition to the increasing population, the agricultural lands are being converted into non-agricultural uses. Utilization of chemical fertilizers has increased production in agriculture to meet the food needs of the society to prevent hunger. Throughout this time, many farmlands have been transformed into unproductive, salinized land (Fedoroff, 2015). The use of industrial nitrogenous fertilizers in agriculture has been reported to have a harmful impact on the environment, which includes the emission of greenhouse gasses like nitrous oxide (N_2O) and leaching of nitrate (NO_3^-) from chemical N fertilizers into the surrounding environment (Bøckman, 1997). The United States Environment Protection Agency stated that the impact of 1 pound N_2O on global warming is 300 times the impact of 1 pound carbon dioxide. Thus, minimizing the application of chemical nitrogenous fertilizers can play an important role in reducing N_2O emission (EPA, 2016).

Legumes can form a symbiotic association with rhizobia to gain N through BNF. The symbiotic relationship between *Sinorhizobium meliloti* 1021(Rm1021) and alfalfa (*Medicago sativa* L.) is one of the models to study symbiotic N fixation and provides us better understanding of plant-microbe interaction. Additionally, some pathogenic α-proteobacteria, such as *Brucella*, share similar mechanisms of interaction, as rhizobia can establish symbiotic relationships with legumes. Pathogenic α-proteobacteria are responsible for many diseases in animals and plants. For example, *Brucella spp*. is responsible for causing brucellosis in animals, *Liberibacter asiaticus* causes citrus greening in citrus and *Liberibacter solanacearum* causes zebra chip disease in potatoes

(Zambriski *et al*., 2010; Miyata *et al*., 2011; Cooper *et al*., 2014). Understanding intracellular RF metabolism in Rm1021 might give us insight into RF biosynthetic enzymes in phylogenetically closely related pathogenic bacteria. Understanding of enzymes will provide a novel potential approach for the development of new strategies to control pathogenic bacteria. The genomic research revealed that *Brucella* spp. shares similar intracellular activities as rhizobia and evolved from a common ancestor with rhizobia. Moreover, unculturable plant pathogenic species, such as *Candidatus liberibacter,* are phylogenetically closely related to Rm1021 and availability of the complete genome of Rm1021 makes rhizobia a good model organism to study these pathogenic bacteria (Galibert *et al*., 2001; Hartung *et al*., 2011).

It was previously proposed that the RF biosynthesis in Rm1021 involves two interchangeable modules; one for internal bacterial metabolic needs and another for secretion (Yurgel *et al*., 2014). In this study, the focus was on RF biosynthesis of rhizobia in order to identify novel enzymes involved in the RF biosynthesis process in Rm1021.

CHAPTER 2: LITERATURE REVIEW

2.1 *Rhizobium*-legume association

Rhizobia are gram negative, N-fixing soil bacteria. Rhizobia invade root hairs of plants by the formation of tubular infection threads that result in specialized symbiotic organ called nodule. The rhizobia, in association with the legume hosts, symbiotically fix atmospheric N_2 to ammonia by the process of BNF (Oldroyd *et al*., 2008). N is an important element required for plant growth and is required to increase crop yield. N is available through soil application of industrial fertilizers and through BNF (Boyd, 2001). Use of chemical fertilizers accounts for around 25% of total N fixed on earth. Out of 25% chemical fertilizers used in agricultural fields half of them is wasted due to leaching. This causes soil degradation, contamination of ground water and adversely affects the environment. The *Rhizobium*-legume association produces about 70 million tons of fixed N per year (Rajwar *et al*., 2013). The crop rotation with legumes enhances the crop productivity of the non-leguminous crops in the rotation. Rhizobial species has host-specificity in forming symbiotic relationship (Hirsch *et al*., 2001). *Rhizobium*-legume symbiosis helps in reducing the use of N fertilizers, making available N for intercropping, as well as crop rotation. It also helps to enrich soil health, reclaiming acidic and saline soils, diverting the path of agricultural production from intensive agriculture to a sustainable way (Vance, 2001).

2.2 Establishment of nodules on legume hosts by Rm1021

Rm1021 is a symbiotic N fixing α-proteobacteria belonging to the family *Rhizobiaceae*, order *Rhizobales*. It forms a symbiotic relationship with legumes of genera *Medicago, Melilotus, Trigonella*. Legume roots exude host-specific signaling molecules like flavonoid compounds, which act as chemoattractant. The chemoattractants stimulate rhizosphere colonization by attracting rhizobia (Long, 2001). Bacteria synthesize and release lipo-chitooligosaccharides called *nod* factors, induced by plant-produced signaling molecules (Garg *et al*., 2007). *Nod* factors initiate the symbiotic signaling pathway in legumes, which triggers root hair deformation and curling (Oldroyd, 2013). *Nod* factors and exopolysaccharides are required for the formation of infection threads in roots. They also stimulate root cortex cells to initiate cell division to form nodule primordia (Jones *et al*., 2007). The infection threads allow the entry of rhizobia into root cortex. Rhizobia are released into plant cells, where they differentiate into N-fixing bacteroids forming membrane-bound compartments called symbiosomes (**Figure 1**). *Medicago sativa* forms indeterminate type of nodules (Timmers *et al*., 2000). The indeterminate nodules are types of nodules in which cell division occurs in the inner cortex and pericycle region forming different zones with elongated nodule shape (Kohlen *et al.*, 2018). In the case of indeterminate nodules, these bacteroids undergo terminal morphological changes into Y-shaped cells (Oke *et al*., 1999).

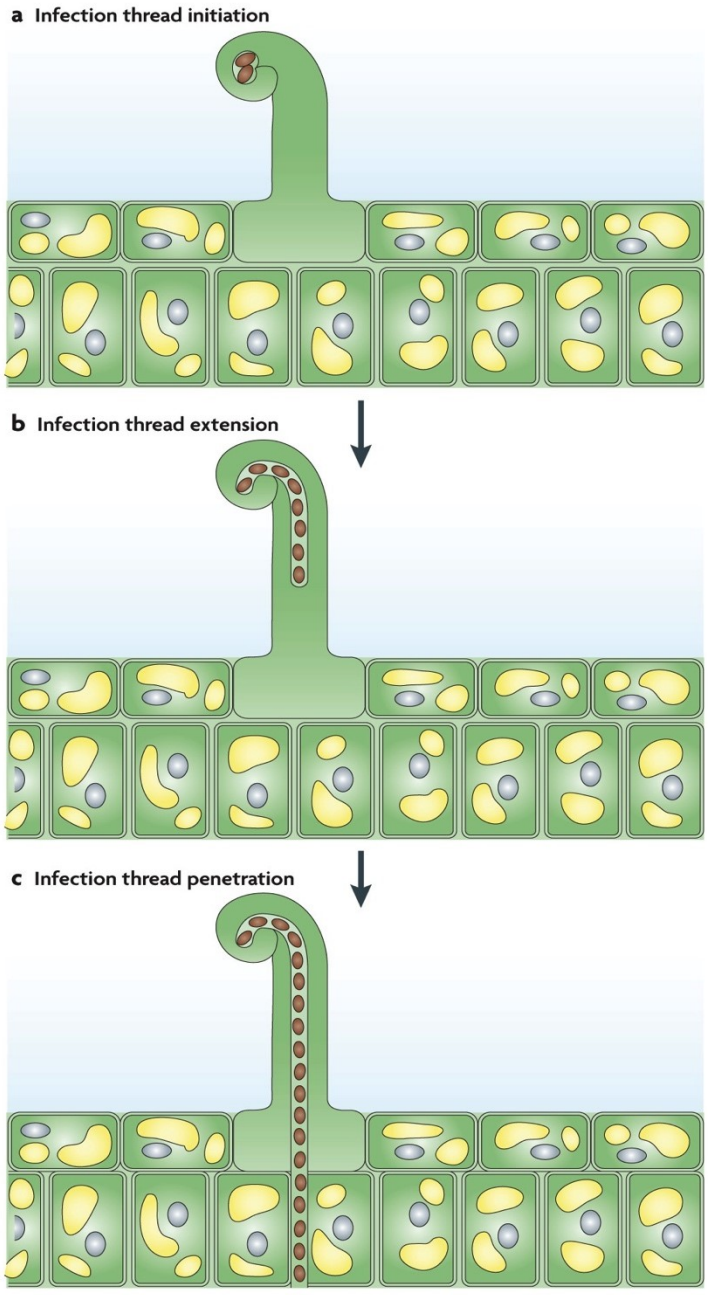


Figure 1. Root hair invasion by Rm1021 (Jones *et al*., 2007).

2.3 Role of flavin in symbiosis

Riboflavin (RF; 7,8-dimethyl-10-ribitylisoalloxazine) or vitamin B2 is a precursor of the cofactors, FMN and FAD. These cofactors play important roles in the electron transport chain, as well as in the metabolism of vitamin B6, B12 and folates, fatty acid oxidation, and DNA repair (Mansoorabadi *et al*., 2006; Roje, 2007). This makes RF a necessary component in cellular metabolism for all living organisms. The RF is not synthesized in humans and animals, but it can be obtained through milk, cereals, grain, eggs, and green vegetables (Edwards, 2014). The RF contribution in *Rhizobium*-legume symbiosis helps in root colonization and improves bacterial competitiveness for nodulation (Yang *et al*., 2002; Yurgel *et al*., 2014). In rhizobia, the enzymes responsible for catalyzing RF biosynthesis are RibBA, RibD, RibE/C, RibH1 and RibH2. In Rm1021, in-frame deletion of the *ribBA* gene affects the RF secretion and results in less ability to fix N_2 (Phillips *et al*.,1999; Yang *et al*., 2002; Yurgel *et al*., 2014). This suggests the importance of RF secretion for bacterial root colonization and nodule formation.

2.4 Riboflavin biosynthetic pathway (RBP)

RBP is almost identical in all higher plants, bacteria and yeast (Bacher *et al*., 2000; Roje, 2007). The only difference among the RF biosynthesis pathways between bacteria and yeast is the inversion of deaminase and reductase reactions encoded by *ribD* gene (**Figure 2**) (Sa *et al*., 2016).

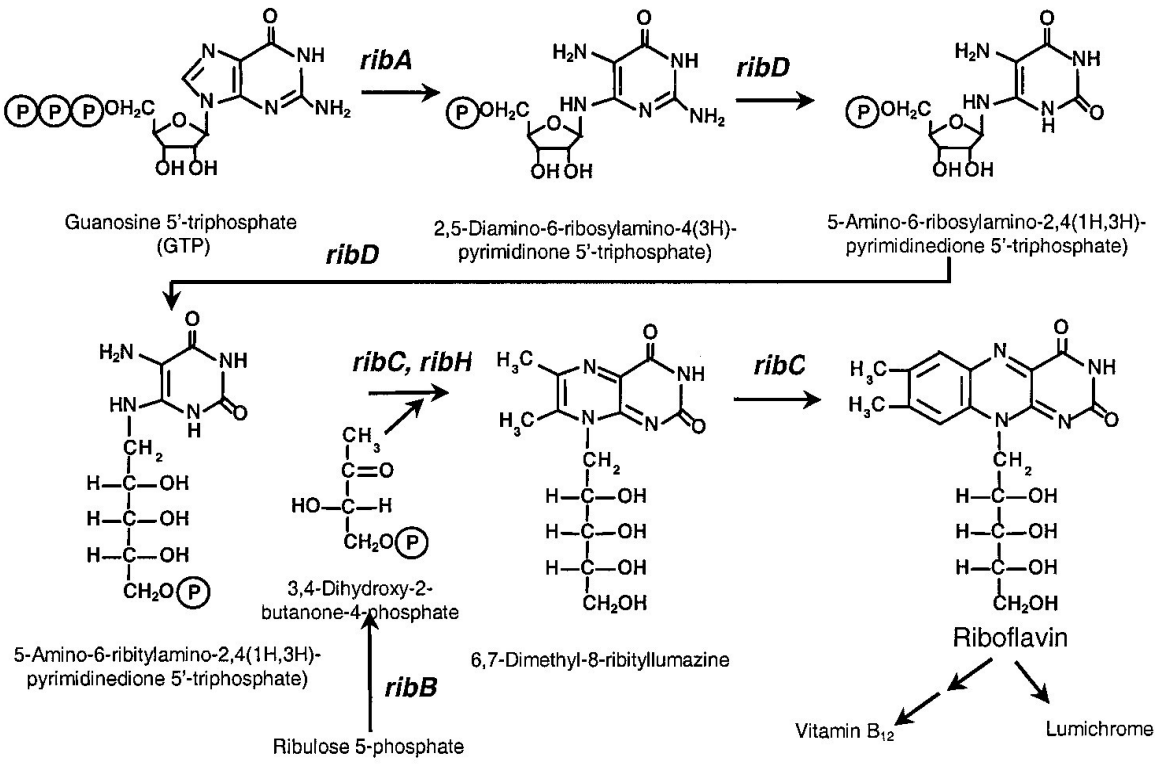


Figure 2. Riboflavin biosynthesis pathway in *E. coli* and the genes (Yang *et al*., 2002).

The RF is synthesized from one GTP and two ribulose 5-phosphate. GTP cyclohydrolase II (GCHII) enzyme encoded by *ribA* gene catalyzes the conversion of GTP into 2,5-Diamino-6-ribosylamino-4(3H)-pyrimidinone 5'-triphosphate. The RibD enzyme has bifunctional deaminase-reductase reactions which proceeds the next two steps of the pathway by deamination of the position 2 amino group and then reduction of the ribosyl side chain (Bacher *et al*., 2000; Yang *et al*., 2002). Bifunctional RibD enzymes converts 2,5-Diamino-6-ribosylamino-4(3H)-pyrimidinone 5'-triphosphate to 5-Amino-6-ribosylamino-2,4(1H,3H)- pyrimidinone 5'-triphosphate in the first step of deaminase and finally, is reduced to 5-Amino-6-ribitylamino-2,4(1H,3H)-pyrimidinone 5'triphosphate (Bacher *et al*., 2000). Further, the reaction of dephosphorylation of 5'-phosphate structure is unknown (Bacher *et al*., 2000). The 6,7-Dimethly-8-

ribityllumazine synthase is encoded by *ribH* and catalyzes the condensation of dephosphorylated 5-Amino-6-ribitlyamino-2,4(1H,3H)-pyrimidinone with 3,4-Dihydroxy 2-butanone-4-phosphate. 3,4-Dihydroxy 2-butanone-4-phosphate synthase (DBPS), encoded by the *ribB* gene, which converts ribulose 5-phosphate into 3,4-Dihydroxy 2-butanone-4-phosphate. Bifunctional protein RibBA exists in certain bacterial species (including Rm1021) with both GCHII and DBPS activities fused into a single enzyme (Galibert *et al*., 2001; Yang *et al*., 2002; Yurgel *et al*., 2014). The RibC, a RF synthase enzyme (also called RibE in Rm1021) catalyzes the dismutation of 6,7-dimethly-8-ribityllumazine to produce RF **(Figure 2).** Lumazine synthase activity is encoded by two homologous genes, *ribH1,* and *ribH2*. In some bacterial genomes, with exception of *ribH*, which is represented by 2 copies of the gene, all other RF biosynthetic enzymes are encoded by a single gene (Galibert *et al.,* 2001; Fischer *et al*., 2011; Ladenstein *et al*., 2013).

The RF biosynthetic enzymes are found to form complexes during RF biosynthesis. It was determined that RF biosynthetic enzymes exist in a multi-enzyme complex, bound together with a non-covalent bond, or fused, forming RF metabolon (Frelin *et al*., 2015). Various research indicated the interactions between bifunctional RibBA with RibH1 and RibH2, as well as interactions of RibE, RibH1 and RibH2 with themselves and each other (Frelin *et al*., 2015; Ladenstein *et al.,* 2013). The protein with N-glucosidase activity is fused with some RBP enzymes, further proving the existence of RF metabolon (Frelin *et al*., 2015). Lumazine synthase (RibH) and RF synthase (RibE) form multi-enzyme complexes during RF biosynthesis (Ladenstein *et al*., 2013). In Rm1021, the genes responsible for catalyzing the initial steps of GCHII (*ribA*) and DHBP

(*ribB*) are found to be a part of a single fused gene *ribBA (*Galibert *et al*., 2001*)*. Additionally, in Rm1021*, ribD* and *ribE* are found to be a part of the single transcriptional unit (Galibert *et al*., 2001). This predicts the interaction between RF enzymes and existence of complex formation in RF biosynthesis. Thus, analysis of complex formation and interaction between RBP enzymes can potentially lead to discovery of novel components of RF enzymes metabolon.

2.5 Riboflavin biosynthesis in Rm1021

a. Role of RibBA

In Rm1021, RF biosynthesis is catalyzed by bifunctional GCHII/DBPS (RibBA) protein, bifunctional deaminase-reductase (RibD) protein, lumazine synthase (RibH), and RF synthase (RibE). RibBA plays an important role in FL secretion, which facilitates rhizobial colonization (Yang *et al*., 2002; Yurgel *et al*., 2014). The *ribBA* gene in Rm1021 colonized root in alfalfa due to elevated level of RF as compared to the control strain Rm1021-WS12 with empty plasmid vector (Yang *et al*., 2002).

In the Rm1021, in-frame deletion mutants of *rib* genes were generated. Surprisingly, Rm1021Δ*ribBA* mutant was able to grow on defined media without RF supplementation (Yurgel *et al*., 2014). While RibD and RibE mutants were RF auxotroph and were not able to grow without RF supplementation. This indicated that RibD and RibE are unique enzymes that catalyze deaminase-reductase and RF synthase reactions in RBP, respectively. However, RF prototrophy of Rm1021Δ*ribBA* was unexpected, as Rm1021 genome contains a single copy of the gene (Yurgel *et al*., 2014). This indicated

the involvement of alternate enzyme to RibBA (Yurgel *et al*., 2014). The Rm1021Δ*ribBA* mutants secreted a smaller amount of RF as compared to its wild type strain but could establish effective (N fixing) symbiosis with alfalfa (Yang *et al*., 2002; Yurgel *et al*., 2014). The alternate pathway modules were proposed to be present in α-proteobacteria for RF biosynthesis; one for intracellular needs of bacteria metabolism itself and the other for secretion to establish symbiosis with the host (Yurgel *et al*., 2014).

b. Novel ArfB

The alternative enzyme to RibBA, ArfB, was identified using Tn*5* mutagenesis of Rm1021Δ*rib*BA and subsequently screened for double mutants for RF auxotrophy. The mutation was identified by genetic analysis and was found to trigger RF auxotrophy, not only in Rm1021Δ*rib*BA but also in the wild type strain Rm1021. This suggested an important role of the mutated gene for RF production.

In bacterial RF biosynthesis pathway, GCHII reaction proceeded in a single step (**Figure 3**, green notation), whereas in the archaeal RBP an initial step of ring cleavage and dephosphorylation of GTP is catalyzed by GTP cyclohydrolase III (GCHIII/ArfA) and formamide hydrolase/ArfB (**Figure 3**, red notation).

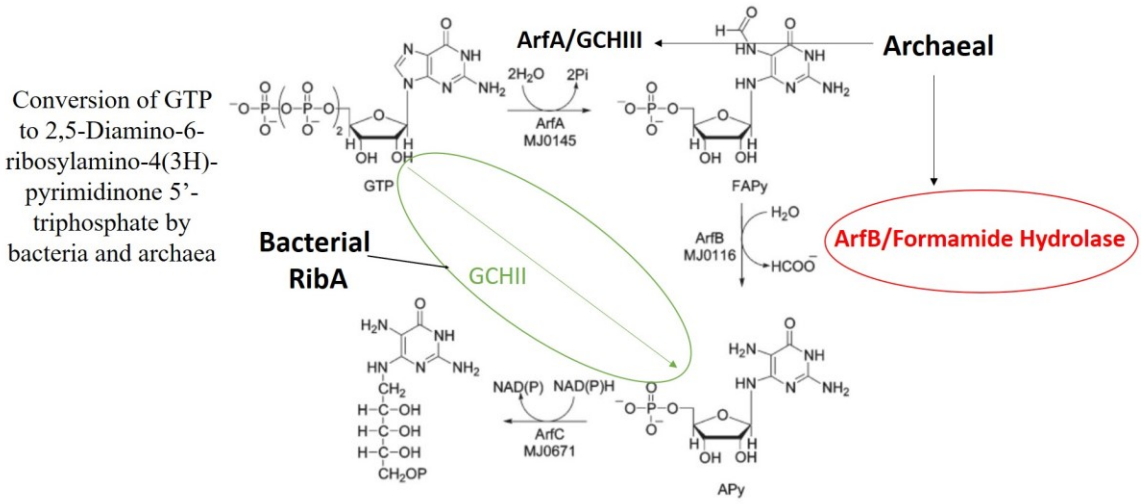


Figure 3. Archaeal riboflavin biosynthesis.

Bacterial RF biosynthesis GCHII (RibA) activity, archaeal RF GCHIII (ArfA) and Formamide Hydrolase (ArfB) activity (modified from Grochowski *et al*., 2009).

The reaction proceeds by dephosphorylation and ring cleavage of GTP, forming an intermediate product 2-amino-5-formlyamino-6-ribosylaminopyrimidin-4(3H) one 5'-monophosphate and deformylation of intermediate into 2,5-diamino-6-ribosylamino-4(3H)-pyrimidine one 5-phosphate and formate. The purified Rm1021 enzyme ArfB showed 2-amino-5-formlyamino-6-ribosylaminopyrimidin-4(3H) one 5'-monophosphate deformylase (formamide hydrolase) activity, which participates in RBP for intracellular needs in Rm1021. This was the first report designating bacterial GCHIII/formamide hydrolase enzymes, which catalyze the reaction of GTP into diamino-ribosylamino-pyrimidinone-phosphate, APy (**Figure 3**). The *arfB* showed no sequence homology with

archaeal formamide hydrolases, and bioinformatics analysis did not identify any Rm1021 proteins with significant homology to archaeal GCHIII. This suggests that GCHIII activity in Rm1021 is catalyzed by another unknown enzyme. The ArfB derived its name as it catalyzes the second step of archaeal RF and 7,8-didemethyl-8-hydroxy-5-deazariboflavin (Folate) biosynthesis (Grochowski *et al*., 2009).

c. ArfB complementation

Phylogenetic analysis of ArfB sequence homologs identified similar proteins in most of the sequenced α-proteobacteria. In *Brucella* and *Liberibacter* species, genes encoding homologs of Rm1021 ArfB protein complemented RF auxotrophy of *Rm1021*Δ*arfB* mutant. In clade IIa of Rm1021 *arfB* sequence (**Figure 4**), homologs from α-proteobacteria such as *Brucella melitensis, and Liberibacter solanacearum* complemented RF auxotrophy. This indicated similar functional enzymes of RBP in *Brucella* and *Liberibacter spp*. In case of *Ochrobactrum anthropi,* two sequence homologs of *arfB* were identified (**Figure 4**, Blue arrow). When the distant copy of *O. anthropi arfB* (Oa *arfB*) was introduced into Rm1021Δ*arfB,* the resulting strain, Rm1021Δ*arfB*xOa *arfB expr* was still RF auxotroph. However, revertants were found in Rm1021Δ*arfB*x Oa *arfB expr*, which could grow in media without RF (Dr. Yurgel's unpublished data). Plasmid overexpressing Oa *arfB* from revertants was isolated and introduced into new Rm1021Δ*arfB*. This did not complement RF auxotrophy (Dr. Yurgel's unpublished data). Further, sequencing of plasmid identified no mutation in the Oa *arfB* gene, suggesting that genetic changes alleviating the RF auxotrophy were

located in the revertants chromosome. Since the existence of multi-protein complexes between RBP enzymes are well documented, these genetic modifications may cause conformational changes in ArfB complex partners, allowing the proteins to form active complexes with exogenous Oa *arfB* necessary for enzymatic activities. Identification of genetic changes in ArfB complex partners can potentially lead to the identification of new enzymes, which might be linked to FL biosynthesis and secretion. The identification of these changes will be helpful for understanding the functions of RBP and its effect on interaction with the host.

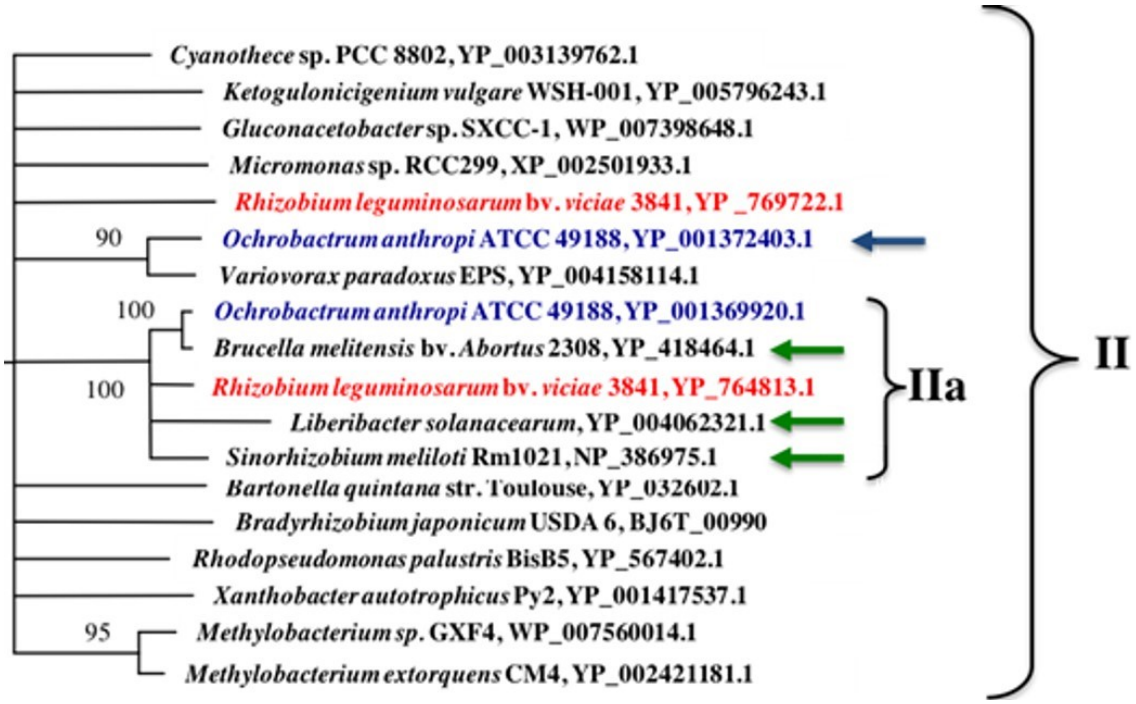


Figure 4. Phylogenetic tree of ArfB sequence homologs (Clade II).

Two copies of ArfB sequence homologs found in the same organism listed with the same color. The green arrow indicates proteins capable of complementing Rm1021 ArfB auxotrophy. Blue arrows indicates two Rm1021 ArfB homologs found in *O. anthropi* (Dr. Yurgel's unpublished data).

2.6 Hypothesis and research objectives

Hypothesis

The research hypothesis of this study is that the alleviation of RF auxotrophy in the Rm1021Δ*arfB*x Oa *arfB expr* strain might be result of one of the two processes:

a. A small set of genetic modification in Oa *arfB*, attenuating its function can improves RF biosynthesis, or
b. A small set of genetic modifications in *S. meliloti* proteins involved in RF metabolism by forming complexes with endogenous ArfB, allowed formation of functional complexes with exogenous Oa *arfB* and therefore improving the efficiency of RF biosynthesis.

Research objective

The objective of this study is to identify the genetic changes that lead to RF prototrophy in *S. meliloti* strain carrying exogenous Oa *arfB* in order to identify novel/alternate enzymes that are involved in RF metabolism in α-proteobacteria.

Specific objectives:

1. To isolate RF prototrophic revertants of *S. meliloti* strain carrying exogenous Oa ArfB;
2. To identify the genetic mutations potentially linked to restoration of the RF prototrophy and RF metabolism;
3. To analyze the effect of the mutations on free living revertants and symbiotic phenotype of the revertants by evaluation of their,
    a. growth properties,
    b. FL secretion, and
    c. the ability to form effective symbiosis the host-plant.

CHAPTER 3: MATERIALS AND METHODS

3.1 Bacterial strains and media

The bacterial strains and plasmids used in the study and their characteristics are listed (**Table 1)**. To complement RF auxotrophy of Rm1021Δ*arfB*, Oa *arfB* gene was cloned into the Tc^r resistant plasmid pCPP30 (**Table 1**). Rm1021 strain carrying empty plasmid vector pCPP30 (Rm1021xpCPP30) was used as a control strain. Rhizobial strains were grown on Minimal mannitol ammonium ($MMNH_4$) media (Somerville and Kahn, 1983) at 28°C. Antibiotics streptomycin (St^r) (200 µg/mL) and tetracycline (Tc^r) (10 µg/mL) were used for strain selections. All genetic manipulations with Rm1021 that contained the *rib* regions were done using $MMNH_4$ media supplemented with 500 µM RF. After verification of RF requirements for growth, strains not requiring RF were routinely grown on media without RF.

$MMNH_4$ media content per liter:

Mannitol	10.0 g (54.9 mM)
NH_4Cl	0.5 g (9.34 mM)
Agar (for plates, Sigma Aldrich)	15.0 g
Biotin (0.2 mg/mL in 50% EtOH)	1.0 mL
Thiamine (2 mg/mL filter sterilized)	1.0 mL
Min Man Salt I	10 mL
Min Man Salt II	10 mL

Min Man Salt I

K_2HPO_4	100 g/L	574.12 mM
KH_2PO_4	100 g/L	734.8 mM
Na_2SO_4	25 g/L	176 mM
dd-H_2O	970 mL	

Min Man Salt II

FeCl3 $6H_2O$	1.0 g/L	3.7 mM
$CaCl_2.2H_2O$	10.0 g/L	68 mM
$MgCl_2$. $6H_2O$	25.0 g/L	123 mM
ddH_2O	970 mL (Somerville and Kahn, 1983).	

Table 1. Strains and plasmids used in this study.

Strains	Abbreviations	Genotype	Characteristics	Reference
S. meliloti 1021	Rm1021	Wild type	RF prototroph, St^r, Fix+	Galibert *et al*., 2001
Rm1021Δ*arfB*	Rm1021Δ*arfB*	*arfB* deletion in Rm1021	RF auxotroph, St^r, Fix-	Dr. Yurgel's lab
Rm1021Δ*arfB*xpC PP30_Oa *arfB*_#1.4	1.4	Rm1021Δ*arfB* complementation with *O. anthropi arfB* on pCPP30; RF auxotroph	RF auxotroph, Tc^r, Fix-	Dr. Yurgel's lab
Rm1021Δ*arfB*xpC PP30_Oa *arfB*_#1.4_rev#1	Rev_1	Rm1021Δ*arfB*xpC PP30_Oa *arfB*_#1.4 RF prototroph revertant	RF prototroph, Tc^r, Fix-	Dr. Yurgel's lab
Rm1021Δ*arfB*xpC PP30_Oa *arfB*_#1.4_rev#10	Rev_10	Rm1021Δ*arfB*xpC PP30_Oa *arfB*_#1.4 RF prototroph revertant	RF prototroph, Tc^r, Fix-	Dr. Yurgel's lab
Rm1021Δ*arfB*xpC	Rev_5	Rm1021Δ*arfB*xpC	RF prototroph, Tc^r,	This work

PP30_Oa *arfB*_#1.4_rev#5		PP30_Oa *arfB*_#1.4 RF prototroph revertant	Fix-	
Rm1021Δ*arfB*xpC PP30_Oa *arfB*_#1.4_rev#6	Rev_6	Rm1021Δ*arfB*xpC PP30_Oa *arfB*_#1.4 RF prototroph revertant	RF prototroph, Tc^r, Fix-	This work
Rm1021Δ*arfB*xpC PP30_Oa *arfB*_#1.4_rev#8	Rev_8	Rm1021Δ*arfB*xpC PP30_Oa *arfB*_#1.4 RF prototroph revertant	RF prototroph, Tc^r, Fix-	This work
Rm1021Δ*arfB*xpC PP30_Oa *arfB*_#1.4_rev#5-1	Rev_5-1	Rm1021Δ*arfB*xpC PP30_Oa *arfB*_#1.4_rev#5 RF prototroph secondary revertant	RF prototroph, Tc^r, Fix+	This work
Rm1021Δ*arfB*xpC PP30_Oa *arfB*_#1.4_rev#10-	Rev_10-1	Rm1021Δ*arfB*xpC PP30_Oa *arfB*_#1.4_rev#10	RF prototroph, Tc^r, Fix+	This work

1		RF prototroph secondary revertant		
Rm1021Δ*arfB*xpCPP30_Oa *arfB*_#1.4_rev#10-1	Rev_10-2	Rm1021Δ*arfB*xpCPP30_Oa *arfB*_#1.4_rev#10 RF prototroph secondary revertant	RF prototroph, Tc^r, Fix+	This work
Plasmids				
pCPP30	pCPP30	IncP LacZ	Tc^r	Huang *et al.,* 1992
pCPP30xOa *arfB*	pCPP30xOa *arfB*	pCPP30 (*O. anthropi arfB* including a 500-bp promoter region from Rm1021)	Tc^r	Dr. Yurgel's lab

a. Isolation of revertants

The RF auxotrophic parental strain 1.4, containing Oa *arfB* gene on broad host range plasmid pCPP30 (**Table 1**) was used to isolate RF prototrophic revertants. The parental strain 1.4 was streaked on $MMNH_4$ Tc^r plates supplied with RF supplementation. The plates were incubated at 28ºC. After 7 days of incubation, single colonies appeared on the plates. The single colonies were picked and re-streaked on the same medium 3 times to obtain pure culture of RF prototrophic mutants (**Figure 6**). As a result, three RF prototrophic mutants Rev_5, Rev_6, and Rev_8 (**Table 1**) were isolated. Additionally, two previously isolated revertants, Rev_1 and Rev_10 (**Figure 6**), obtained from Dr. Yurgel's lab collection were used in this study.

b. Bacterial growth test

Bacterial growth test/replication test was used for evaluation of bacterial growth phenotype as published in Yurgel *et al*., 2010; Hagberg *et al.*, 2016. For the analysis of bacterial growth, $MMNH_4$ media with and without RF supplementation were used. The cells of RF prototrophic revertants, parental strain 1.4 and control Rm1021xpCPP30 were suspended in $MMNH_4$ broth to $OD_{(600)}$ of 0.5, and successive 10-fold dilutions of the cell suspensions were prepared in a 96-well microplate. The aliquots were then transferred onto the solid medium using a sterile bolt replicator. The replicator was sterilized by dipping in 70% ethanol and briefly holding over a flame of a Bunsen burner. The bolt

replicator was also sterilized between the replication process for individual plate. The replicated plates were then transferred to incubator set at 28ºC for 7-8 days.

c. Flavin characterization

The FL quantification test was performed to evaluate the differences in FL secretion between the wild type strain Rm1021 and RF prototrophic revertant (Yurgel *et al*., 2014). Rm1021xpCPP30 and the mutants were grown on $MMNH_4$ plates at 28ºC for 3-5 days. The cells were picked and transferred into 14 mL Falcon tube with 5 mL $MMNH_4$ broth and incubated on rotary shaker for 48 hr at 28ºC with rotational speed of 250 revolution per minute (rpm). The culture was diluted 20 folds into the 3 mL fresh $MMNH_4$ broth and continued to incubate on rotary shaker at 28ºC with rotational speed of 250 rpm. The fluorescence readings were recorded on 3^{rd}, 5^{th}, and 7^{th} day of the culture period. For every measurement of fluorescence reading, 200 µL cell cultures were transferred to sterile 96-well microplate. The FL secretion was measured using Bio-Tek Synergy H1 Hybrid Multi-Mode Reader and the Gen5 software application. To obtain relative fluorescence measurements for FL, excitation wavelength of 470 nm and emission wavelength of 530 nm were used. The readings were normalized to OD_{600}.

d. Plant test

The plant tests were performed as described previously (Yurgel and Kahn, 2005; Yurgel *et al*., 2007). The seeds of alfalfa (*Medicago sativa* L.) were sterilized using standard protocol. The seeds were scarified and sterilized by concentrated sulfuric acid (H_2SO_4) (95%) for 5 minutes vortexing intermediately and with commercial bleach for 3 minutes, respectively. The seeds were washed in between the treatment and after treatment with sterile H_2O 8 times. After sterilization, the seeds were dried in the laminar hood for 12 h until moisture is eliminated. The dried seeds were stored at 4^oC for use in the experiment. For germination, the dried seeds were spread on water agar plates and kept at 4^oC for 36 h followed by incubation at 30^oC for 24 h.

The magenta boxes (Bio-World, Dublin; OH, U.S.A.) were used for plant test. The following items were used to prepare magenta boxes ready for plant test: pebbles (20 g), vermiculite (25 g), and the N-free plant nutrient solution (PNS, see below) to saturation. The seedlings were transferred in the sterile magenta boxes. The 3 days old seedlings (5 seedlings per box) were inoculated with 10^7 colony forming units of bacterial strains per box. A 1 mL cell suspension was prepared by picking cells from fresh grown strains plates and resuspending them into sterile PNS. The inoculated magenta boxes were placed on plant growth rack for four to five weeks. After that, plants were harvested, and the plant shoot dry mass were measured. The root nodule formation was examined, and pictures were taken. The plant test was performed in three replications. One-way Tukey analysis was performed to differentiate the mean shoot mass between alfalfa inoculated with different strains using Minitab 17.

PNS composition per liter:

$CaSO_4.2H_2O$	0.34 g	1.97 mM
K_2SO_4	1.6 mL	0.8 mM
KH_2PO_4	0.4 mL	0.4 mM
K_2HPO_4	0.6 mL	0.6 mM
$MgSO_4.7H_2O$	0.4 mL	0.4 mM
$CoCl_2.6H_2O$	0.04 mL	0.002 mM
$FeCl_3$	0.4 mL	0.02 mM
M6 micronutrients	0.4 mL	

M6 micronutrients composition per 10 mL:

KCl	37.3 mg	50.03 mM
H_3BO_3	15.5 mg	25.06 mM
$MnSO_4.2H_2O$	0.34 mg	0.18 mM
$ZNSO_4.2H_2O$	0.57 mg	0.28 mM
$CuSO_4.5H_2O$	0.125 mg	0.05 mM
$NaMoO_4.2H_2O$	0.103 mg	0.04 mM

3.2 Isolation of secondary revertants of *S. meliloti* strains from nodules

Secondary revertants were isolated and obtained from the apparently effective root nodules (Fix+) of alfalfa plants of original inoculated revertants of *S. meliloti* strains Rev_5, and Rev_10. The roots with apparently effective pink nodules were washed with water thoroughly to remove the vermiculite particles. The nodules were picked by sterile

razor into a microcentrifuge tube and sterilized using 70% ethanol for 5 minutes, followed by 2.5% sodium hypochlorite for 1 minute. The traces of chemicals were removed by washing the nodules several times with sterile distilled H_2O. The nodules were crushed using sterile a micro-pestle. The aliquots (100 µL) of nodules suspensions were spread on solidified $MMNH_4$ medium and incubated for 48 h at 28°C, until individual colonies appeared. The colonies were re-streaked 3 times on $MMNH_4$ medium plates to obtain pure cultures. The newly isolated strains were used for secondary plant test.

a. DNA isolation and sequencing of RF prototrophic revertants

Bacterial DNA was isolated using high salt DNA isolation protocol as described in Mahuku (2004) for Rev_5, Rev_6, and Rev_8. In this method, proteinase K was used for protein inactivation and high salt concentration was used for polysaccharides precipitation. The bacterial cells were harvested from 24 h old cultures grown in the $MMNH_4$ broth. The cells were suspended in 1 M NaCl. Further, the suspension was vortexed vigorously and centrifuged at 2700 *g* for 10 min. Then cells were washed with sterilized deionized water. The pellets were re-suspended in 500 µL extraction buffer and 50 mg/mL proteinase K (Mahuku, 2004) and the lysates were used for nucleic acid extraction process. Commercially available DNA isolation kit (Qiagen's DNA isolation kit) was also utilized for isolation of bacterial DNA. DNA isolation process was conducted on parental strains 1.4, Rev_1, and Rev_10 . The analysis of DNA quality (OD_{260}/OD_{280}) and quantification was conducted using Synergy take3 plate (**Table 2**). In

general, both protocols of commercially available Qiagen's kit and protocol from Mahuku, 2004, yielded a good quality of DNA. Although, high salt DNA extraction protocol yielded higher DNA concentrations.

b. Whole genome sequencing (WGS)

High-throughput sequencing was used to identify genetic modifications in the genomes of RF prototrophic revertants. The samples of 10 µL DNA of revertants and parental strain 1.4 were sent to Centre for Comparative Genomics and Evolutionary Bioinformatics (CGEB-IMR), Dalhousie University, Halifax, Nova Scotia, Canada for library preparation and sequencing. The sequencing was performed on Illumina MiSeq using 300+300 bp paired-end V3 chemistry according to CGEB-IMR standard protocol. The total number of raw reads obtained after sequencing is shown (**Table 2)**. The DNA quality was confirmed before sequencing followed by library preparation. High-throughput sequencing then produced raw sequences in FASTQ format. In total, 99% of reads were mapped and obtained after sequencing to reference genome.

Table 2. DNA quantity and quality with sequenced results of total raw reads and assembled reads.

Sample	OD_{260}/OD_{280}	ng /µL	Total number of raw reads	Reads used to assemble to reference sequence
1.4	1.81	20.84	1,157,508	1,092,541
Rev_1	2.20	07.10	5,609,508	5,465,805
Rev_5	1.60	57.18	1,474,504	1,462,513
Rev_6	2.01	38.69	1,588,948	1,577,228
Rev_8	1.83	27.85	1,318,774	1,306,701
Rev_10	1.97	10.90	6,005,846	5,926,356
Rev_5-1	2.04	57.94	1,045,804	1,034,990
Rev_10-1	1.70	78.22	965,436	959,399
Rev_10-2	1.77	80.11	1,241,314	1,229,272

c. Geneious mapping and variant calling

Geneious version 11.0.2 (http://www.geneious.com; Kearse *et al*., 2012) was used for mapping of reads to reference genomes and identification of the mutations. The Geneious mapper (**Figure 5**) is fast and highly sensitive software application, which allows to find structural variants. Medium sensitivity was set as the default recommended in Geneious version 11.0.2. Medium sensitivity provides the best option for Nextgen sequencing reads (for more than 100,000 reads). Under this condition, higher sensitivity is improbable to change the result and consumes a long period of time. Medium sensitivity provides good results with enough coverage, which was a case for our

sequencing data. The reads were normalized, and error was corrected using default parameters of Geneious. The sequences were then mapped to reference genome of *Sinorhizobium meliloti* strain Rm1021 that was previously sequenced (Galibert *et al*., 2001). The genome was downloaded from NCBI database and used as reference (https://www.ncbi.nlm.nih.gov/genome/1004?genome_assembly_id=300472). Rm1021 genome contains three replicons: chromosome (3.65 million base pairs; SMc), and two mega plasmids pSymA (1.35 million base pairs; SMa), and pSymB (1.68 million base pairs SMb). The identified variations after mapping were filtered and sorted using screening parameters such as variant frequency (>95%), amino acid (AA) change and coverage (>14). The high confidence variations obtained by the screening are listed (**Table 3)**. After the identifying high confidence genetic variations, the single nucleotide polymorphisms (SNPs) or insertion-deletions (indels) were combined for respective replicons SMc, SMb, and SMa. The combined variations were sorted by location on the genome.

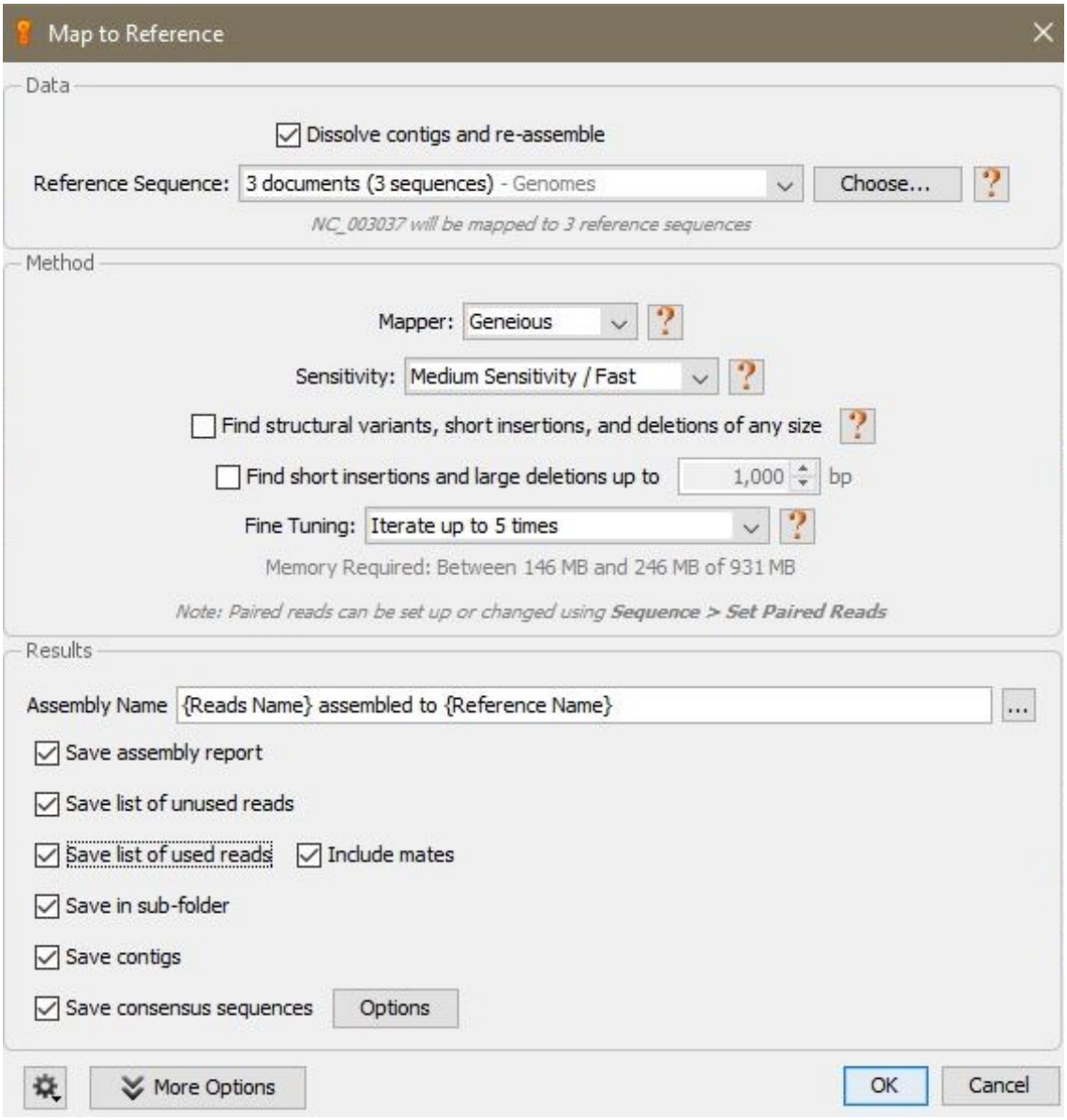


Figure 5. Screenshot of Geneious mapper and variant calling on single nucleotide variants and indels.

The genetic changes common in parental strain 1.4 and RF prototroph revertants were discarded from downstream analysis. Further, the remaining genetic changes were screened for potential changes in AA. After sorting and screening, the selected genetic changes were selected for downstream analysis.

Table 3. Number of initial genetic modifications and high confidence SNPs and indels identified after mapping to reference genomes.

Replicon / Strain	Initials genetic modifications/High confidence genetic modification		
	SMc	**SMb**	**SMa**
1.4	231/96	85/54	108/65
Rev_1	184/87	71/57	98/63
Rev_5	130/61	86/57	94/64
Rev_6	102/49	58/56	98/64
Rev_8	83/58	61/56	98/62
Rev_10	150/89	62/56	85/64
Rev_5-1	105/66	65/57	100/61
Rev_10-1	231/56	73/57	97/62
Rev_10-2	92/54	62/56	80/61

3.3 Identification of variations in *Ochrobactrum anthropi arfB*

The initial hypothesis was that the mutations in Oa *arfB* attenuated the protein function in Rm1021. To test the hypothesis, *arfB* sequences were aligned from each revertant and parental strain 1.4 with the sequence of Oa *arfB*. To perform this test, Oa *arfB* gene sequence (*O. anthropi* ATCC 49188, NC_009668.1-creatininase, located on chromosome 2 between 1278931 to 1279731) was downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov/nuccore/NC_009668.1). The used and unused reads from the RF prototroph revertants and parental strain 1.4 were mapped to Oa *arfB* (creatininase CDS). The mapped Oa *arfB* sequences were extracted from consensus sequence for all

revertants as well as parental strain. The extracted Oa *arfB* sequences were aligned together using MUSCLE alignment. This method provides alignment of large number of sequences with high accuracy and fast speed (Edgar, 2004). This step of analysis was performed to verify the mutations present in Oa *arfB* sequence.

CHAPTER 4. RESULTS

4.1 Identification of phenotypic modifications of RF prototrophic revertants

a. Isolation of RF prototrophic revertants

The revertants were isolated by streaking parental strain 1.4 on media without RF supplementation. The single colonies were picked and re-streaked 3 times to obtain pure cultures. The plates of isolated revertants are shown (**Figure 6)**.

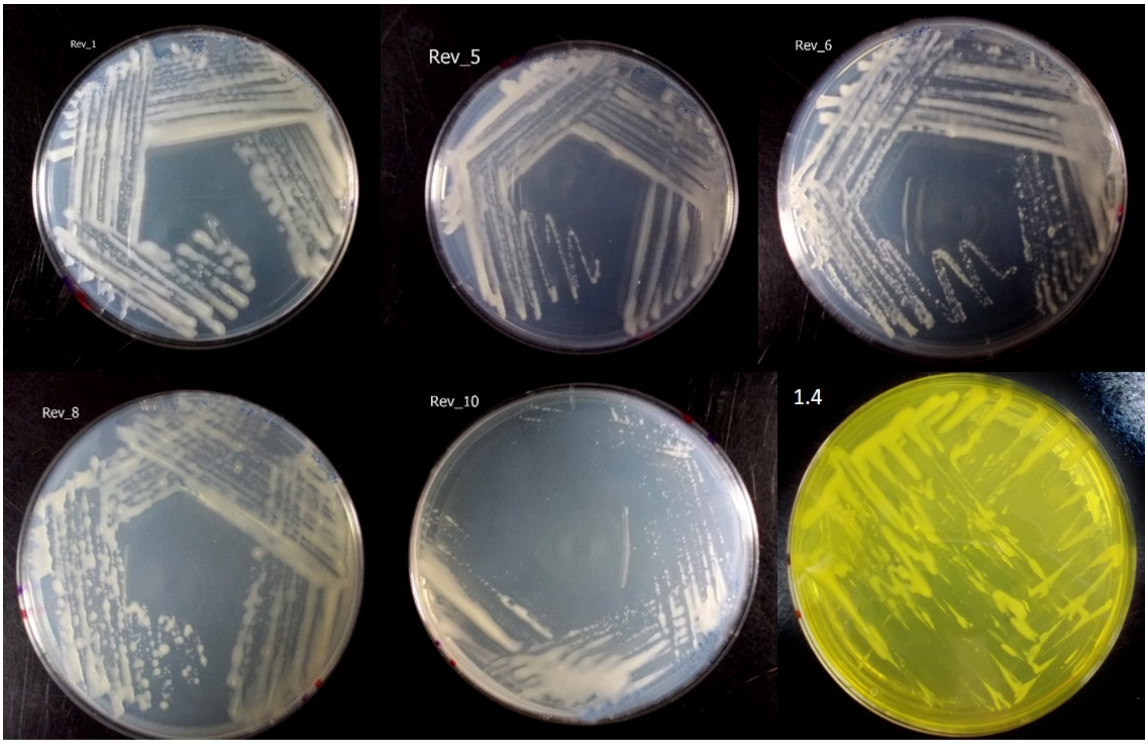


Figure 6. Plates showing growth of revertant strains and Parental strain 1.4

Revertants strains growing on minimal media supplemented without RF except in strain 1.4. The growth on these plates is recorded after pure culture is obtained.

b. Bacterial growth test

The growth of the RF prototrophic revertants were tested on $MMNH_4$ agar plates (**Figure 7**). Parental strain 1.4 and control Rm1021xpCPP30 were used to differentiate the growth of isolated revertants. The growth test indicated that the RF prototrophic revertants could grow without RF supplementation on $MMNH_4$ media plates but grew slower and required more time (7-8 days approximately) compared to the 2-3 days of control Rm1021xpCPP30 strain. The parental strain 1.4 was not able to grow on $MMNH_4$ media plates without RF supplementation, while supplementation with RF restored the growth of 1.4 to a level similar to Rm1021xpCPP30. Interestingly, supplementation of the media with RF did not restore the growth of all of the revertants to the level similar to 1.4 (**Figure 7**).

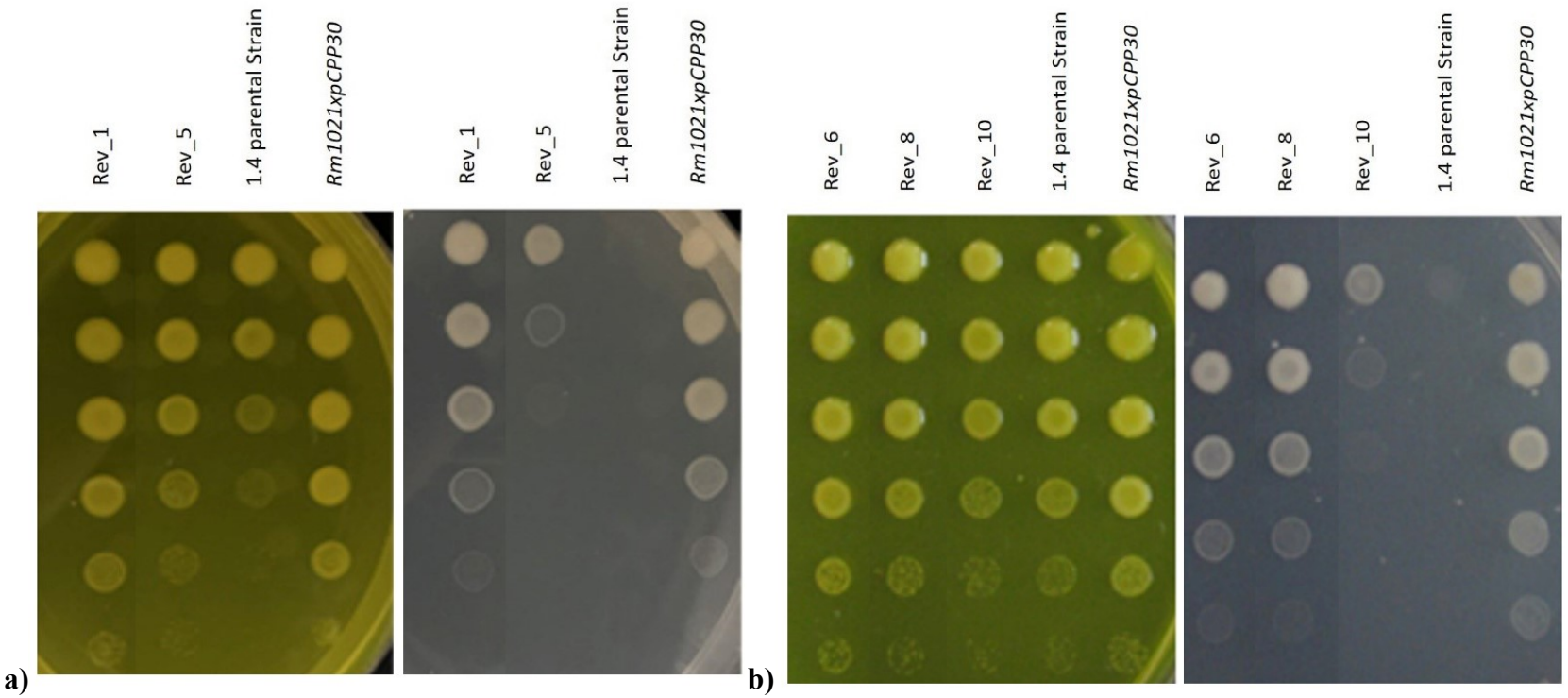


Figure 7. Bacterial replication test

Bacterial growth observed on media plates supplemented with RF (Yellow colored plates) and without RF (white color plates). The growth test result was observed after 7-8 days of replication on plates. The sterile bolt replicator was used to perfom this test. a) The revertant growth test of Rev_1 and Rev_5, including 1.4 and Rm1021xpCPP30. b) The revertant growth test of Rev_6, Rev_8 and Rev_10, including 1.4 and Rm1021xpCPP30.

c. Flavin characterization

The FL characterization in growth media tested the ability of RF prototrophic revertants to secrete RF, FMN and FAD. As observed in the growth test, RF prototrophic revertants can grow without external RF supplementation, which allowed us to measure bacteria-derived FL in the growth media. The FL was measured and potential FL secretion in five selected RF prototrophic revertants (Rev_1, Rev_5, Rev_6, Rev_8 and Rev_10) and control strain Rm1021xpCPP30 was calculated. The relative fluorescence reading was normalized (RFU/OD_{600}). The figures below show the measurement of the relative fluorescence of bacterial growth media (**Figure 8**). The results indicate that all revertants had lower FL secretion compared to Rm1021xpCPP30.

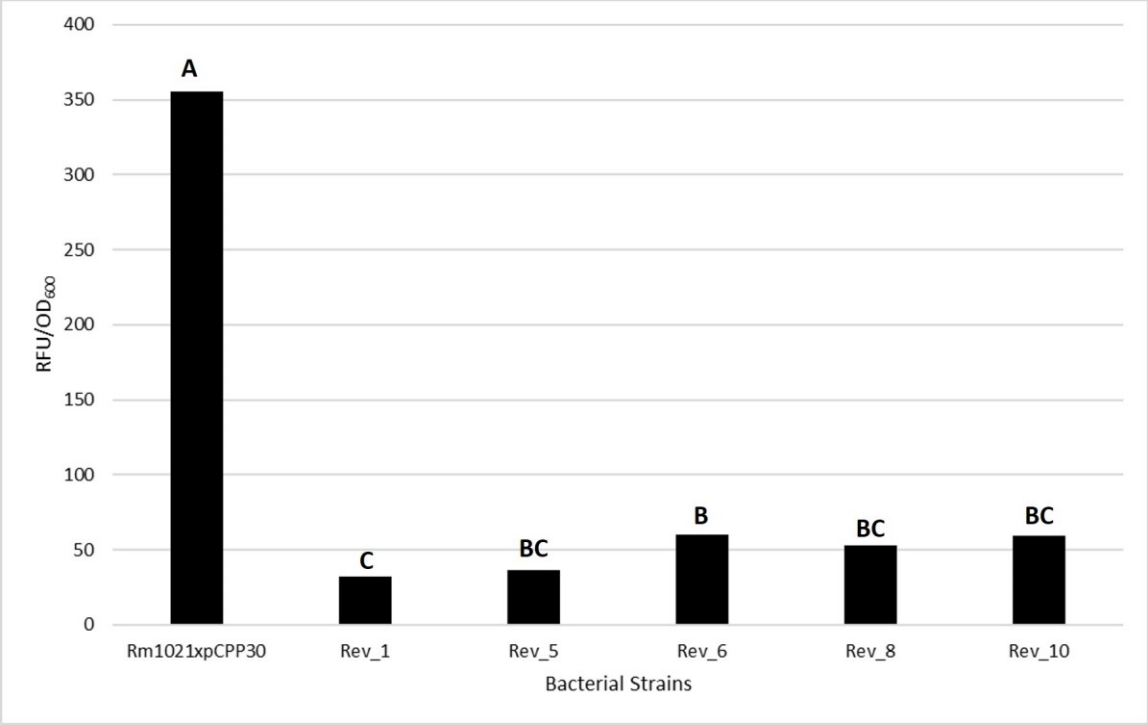


Figure 8. Flavin characterization of RF prototrophic revertants and control strain Rm1021xpCPP30.

The cell cultures were taken on 3rd, 5th, and 7th days of growth. The One-way ANOVA test was used to determine significant ($P<0.05$) difference among the mean relative fluorescence unit per OD_{600} of the revertants and control. (The letters above bar indicates the grouping statistics for mean relative fluorescence unit per OD_{600} with respective strains. The group sharing similar letters have no significant difference between their mean values).

d. Plant test

The symbiotic performance of the RF prototrophic revertants was examined in plants in laboratory conditions. The shoot dry mass of alfalfa plants inoculated with the revertants (Rev_1, Rev_2, Rev_5, Rev_8, and Rev_10), 1.4 and Rm1021xpCPP30 were measured. The phenotype of root nodules formed by each strain was recorded. In general, RF prototrophic revertants exhibited an ineffective symbiosis (Fix-). The mass of the plants inoculated with the RF prototrophic revertants was significantly lower ($P<0.05$)

than the mass of the plants inoculated with the control strain Rm1021xpCPP30. The mass of the plants inoculated with 1.4 was also significantly ($P<0.05$) lower than Rm1021xpCPP30 and was similar to the plants without inoculation (**Figure 9**). Parental strain 1.4 was not able to form nodules (**Figure 10**, i). On the other hand, RF prototrophic revertants produced numerous white ineffective (Fix-) nodules (**Figure 10**, iii), while control strain Rm1021xpCPP30 produced apparently effective pink nodules on the roots of the host-plant (**Figure 10**, ii). The typical nodules form by the strains shown in (**Figure 10)**.

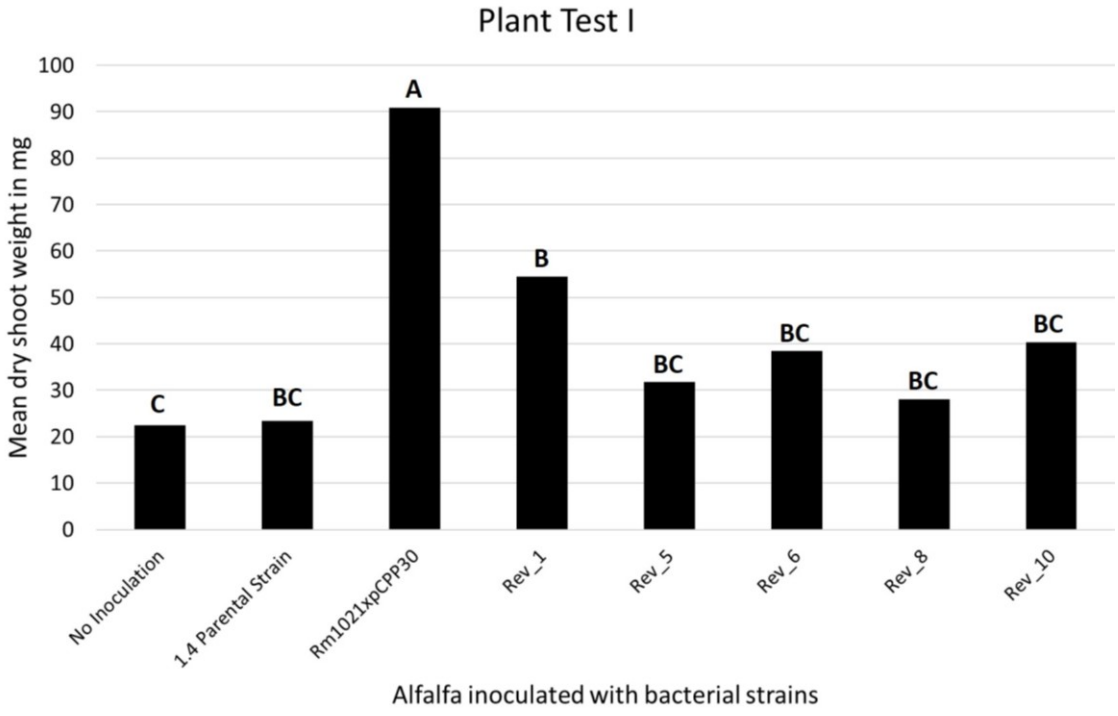
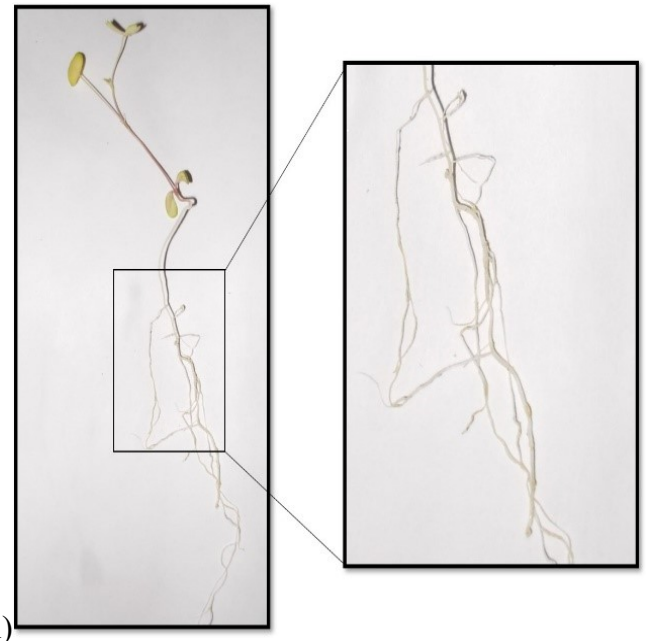


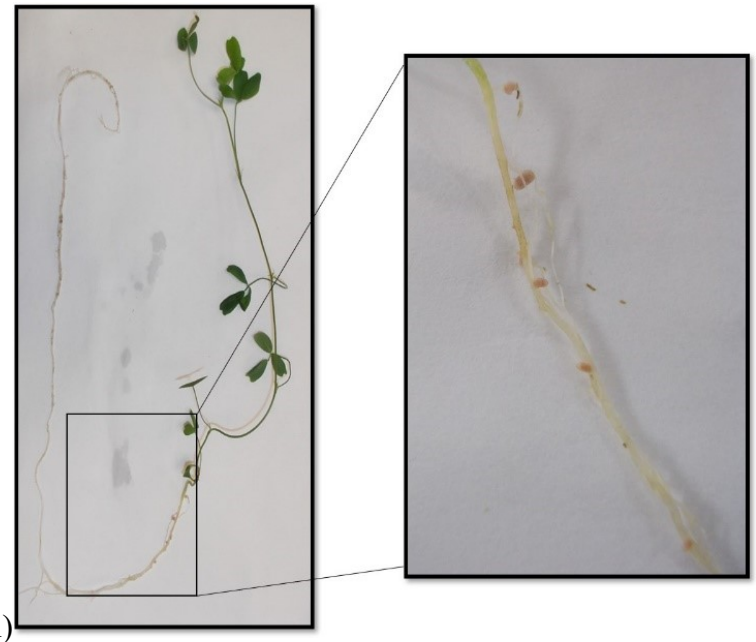
Figure 9. Plant test I with alfalfa (*Medicago sativa* L.) inoculated with RF prototrophic revertants, parental strains, and Rm1021xpCPP30.

The plants were grown for 4 weeks and dry shoot mass weight was measured and recorded per box (5 alfalfa shoot per box). Each treatment had three replications. The dry shoot mass of three replications for each strain treatment was averaged using Minitab One-way ANOVA. The One-way ANOVA test was used to determine significant ($P<0.05$) difference among the mean weights of the inoculated alfalfa shoot. The letters above bar indicates the grouping statistics for mean dry weight of alfalfa shoot inoculated with respective strains.

i)



ii)



iii)

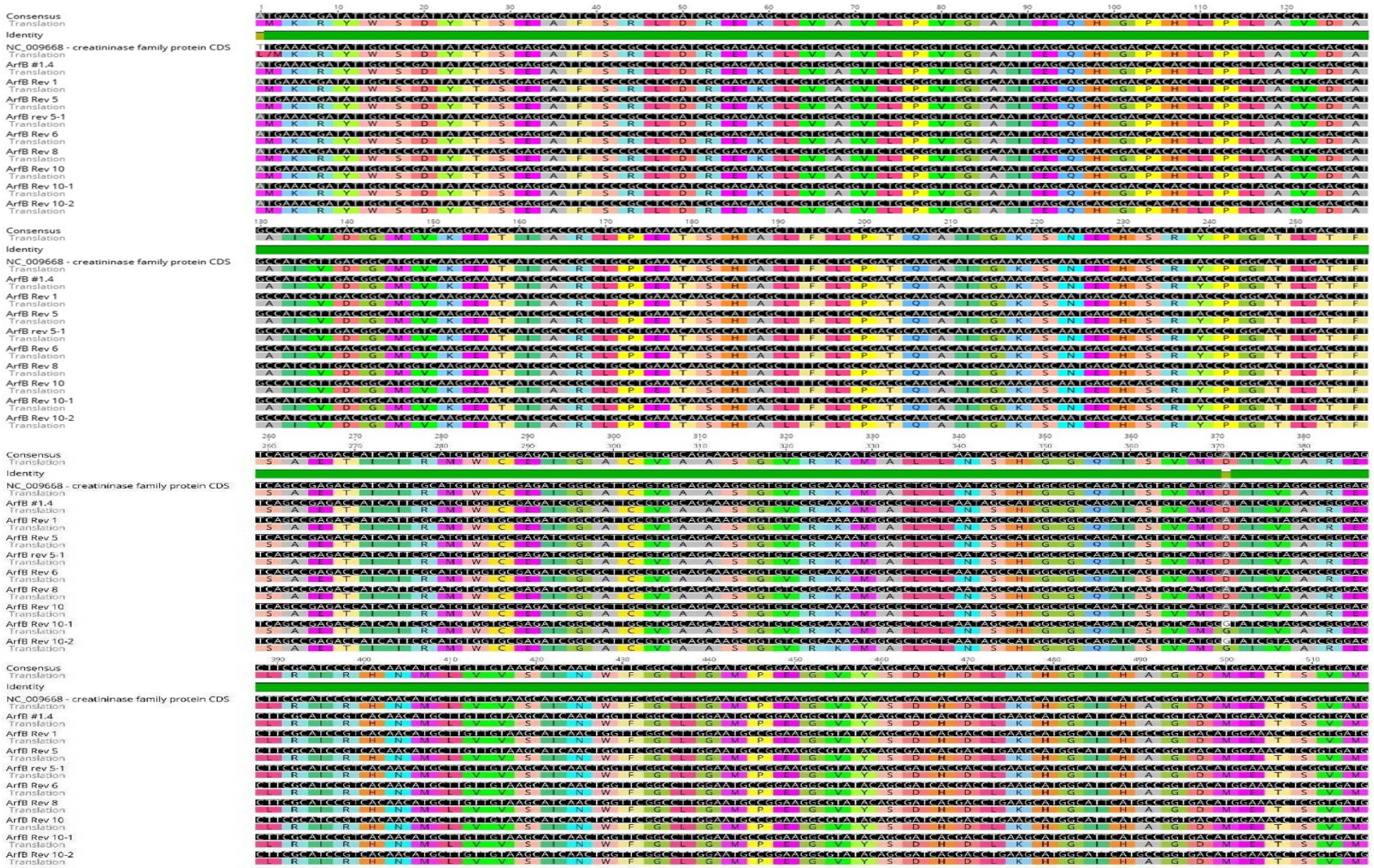
Figure 10. Typical alfalfa phenotype of Plant Test I.

i) Alfalfa inoculated with parental Strain 1.4 was not able to form nodules (Fix-) ii) Alfalfa inoculated with Rm1021xpCPP30 showed pink nodules and apparently better phenotype (Fix+) as compared to alfalfa inoculated with revertant and parental strain 1.4. iii) In the case of alfalfa inoculated with revertants, a high number of white nodules were observed, and phenotypes was distinguished as (Fix-).

4.2 Analysis of genetic modification leading to RF prototrophy

a. Identification of genetic changes in Oa *arfB* sequence of revertants and 1.4 parental strain

First, possibility of mutations in Oa *arfB* that caused the RF prototrophy in the revertants were tested. However, the alignment between original Oa *arfB* and Oa *arfB* sequences from revertants (Rev_1, Rev_5, Rev_6, Rev_8 and Rev_10) and parental strain 1.4 did not identified any genetic modifications in the gene (**Figure 11**). The absence of mutation on Oa *arfB* indicated the possibility of genetic changes in 1.4 genome, leading to RF prototrophy in revertants.


Consensus
Translation
Identity
NC_009668 - creatininase family protein CDS
ArfB #1.4
ArfB Rev 1
ArfB Rev 5
ArfB rev 5-1
ArfB Rev 6
ArfB Rev 8
ArfB Rev 10
ArfB Rev 10-1
ArfB Rev 10-2

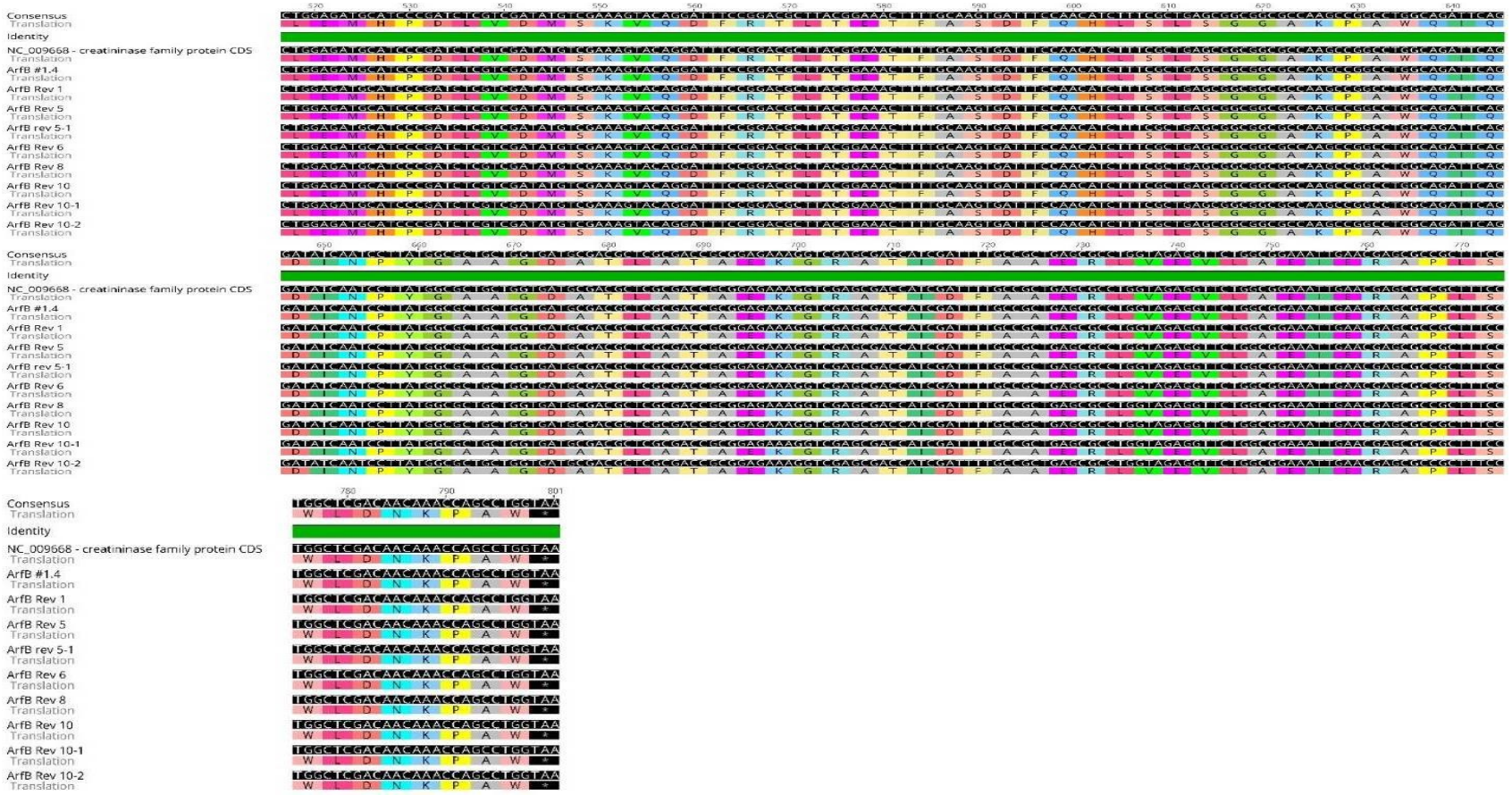


Figure 11. Alignment of consensus Oa *arfB* sequence of revertants and parental strain 1.4 with reference *O. anthropi arfB*. The green color band indicates the 100% identity. The color coding in the alignment is Clustal, generated using Geneious 11.0.2.

b. Identification of potential genetic changes in entire genomes of RF prototroph revertants.

The identification of the genetic changes in Rev_1, Rev_5, Rev_6, Rev_8, Rev_10 was done as described in Materials and Methods. The following mutations were detected in genomes of RF prototroph revertants.

i. SMb21171 (*phnM*)

Variations were identified in coding region of SMb21171 locus in all RF prototroph revertants, while no mutations were detected in this locus in 1.4 genome. The SMb21171 ORF was annotated as *phnM* gene with the length of 1140 nucleotides producing 379 AA long protein (http://iant.toulouse.inra.fr/S.meliloti).

In the Rev_1, the deletion of cytosine resulted in a frameshift. The AA sequence was altered at the position 243 and produced stop codon at the position 270 (**Table 4**). This stop codon was found in the predicted active site of enzymes, which probably resulted in inactivation of the protein (**Figure 12**).

In case of Rev_5 and Rev_8, SNP transversion of cytosine to adenine and guanine to thymine, respectively, resulted in truncation of the proteins at AA positions 151 in Rev_5 and at AA positions 180 in Rev_8. This may also have resulted in inactivation of the protein.

The Rev_10 had SNP transition resulted into an AA change from alanine to valine at residue 81. The valine is C-beta branched AA producing more bulkiness in protein

core as compared to alanine and has difficulty to stable alpha helical conformation (Raman *et al*., 2008).

Rev_6 had insertion of guanine and thymine resulting to frameshift at AA residue 33. This resulted to produce a stop codon at residue 75 changing protein conformation.

Predicted function of PhnM is linked to phosphonate metabolism (Finan *et al*., 2001) The gene ontology of PhnM states that the protein catalyzes hydrolysis of the carbon-nitrogen bond but not the peptide bond; the function characteristic to GCHIII. This predicted function made PhnM a potential candidate for downstream analysis.

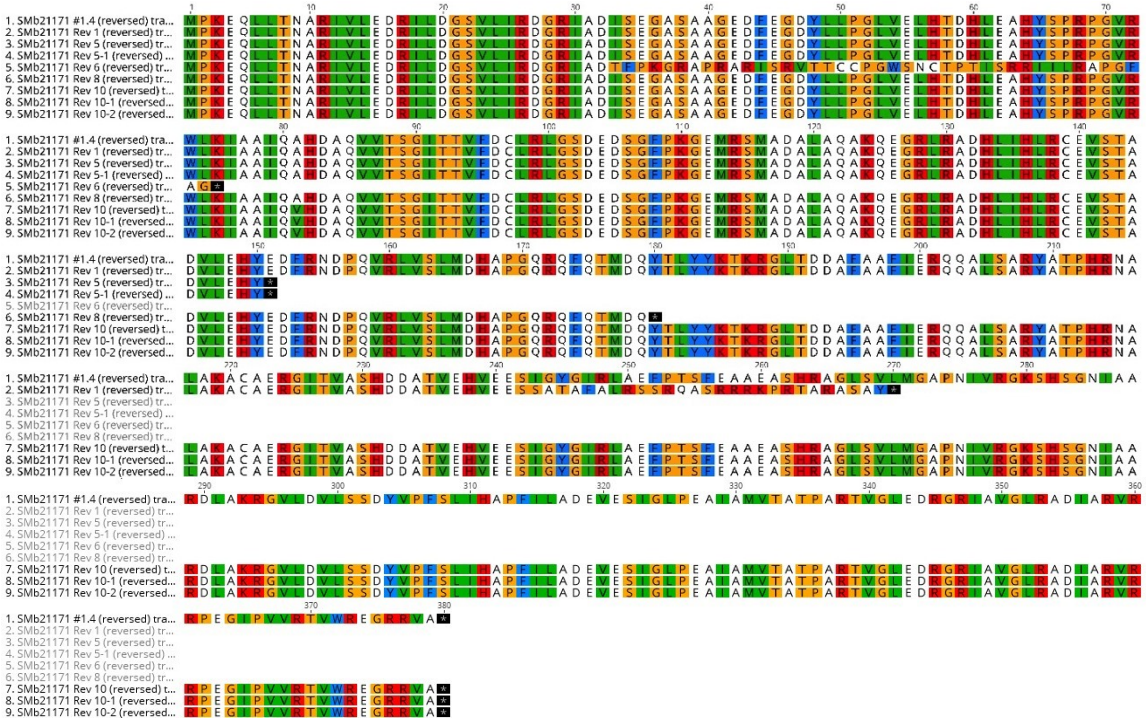


Figure 12. Translated nucleotide alignment of SMb21171(PhnM).

Black colored * indicates stop codon at respective residue.

Table 4. SNPs and Indels obtained on PhnM coding sequence

Minimum	AA Change	Change	Coverage	Protein Effect	Polymorphism Type	Variant Frequency	Variant P-Value (approximate)	Strain
917845	N/A	-C	117	Frame Shift	Deletion	100.00%	6.30E-188	Rev 1
918031	N/A	G -> T	65	Truncation	SNP (transversion)	95.40%	2.70E-188	Rev 8
918120	N/A	C -> A	67	Truncation	SNP (transversion)	100.00%	4.00E-215	Rev 5
918120	N/A	C -> A	38	Truncation	SNP (transversion)	100.00%	2.50E-122	Rev 5-1
918329	A -> V	G -> A	124	Substitution	SNP (transition)	100.00%	0	Rev 10
918329	A -> V	G -> A	14	Substitution	SNP (transition)	100.00%	2.50E-48	Rev 10-1
918329	A -> V	G -> A	41	Substitution	SNP (transition)	100.00%	5.00E-136	Rev 10-2
918477		+GT	64 -> 65	Frame Shift	Insertion	93.80%	1.30E-209	Rev 6

ii. SMc00185 (ABC transporter ATP-binding protein)

Rev_10 had a second mutation in ORF SMc00185, which was annotate as ABC transporter, ATP-binding protein (http://iant.toulouse.inra.fr/S.meliloti) (**Table 5**). The nucleotide changed from thymine to cytosine resulted in the AA changes from tyrosine to cysteine at residue 370 (**Figure 13**). Tyrosine is aromatic partially hydrophobic AA whereas cysteine is more hydrophobic. Tyrosine contains a reactive hydroxyl group making it more interactive with other aromatic AA by stacking.

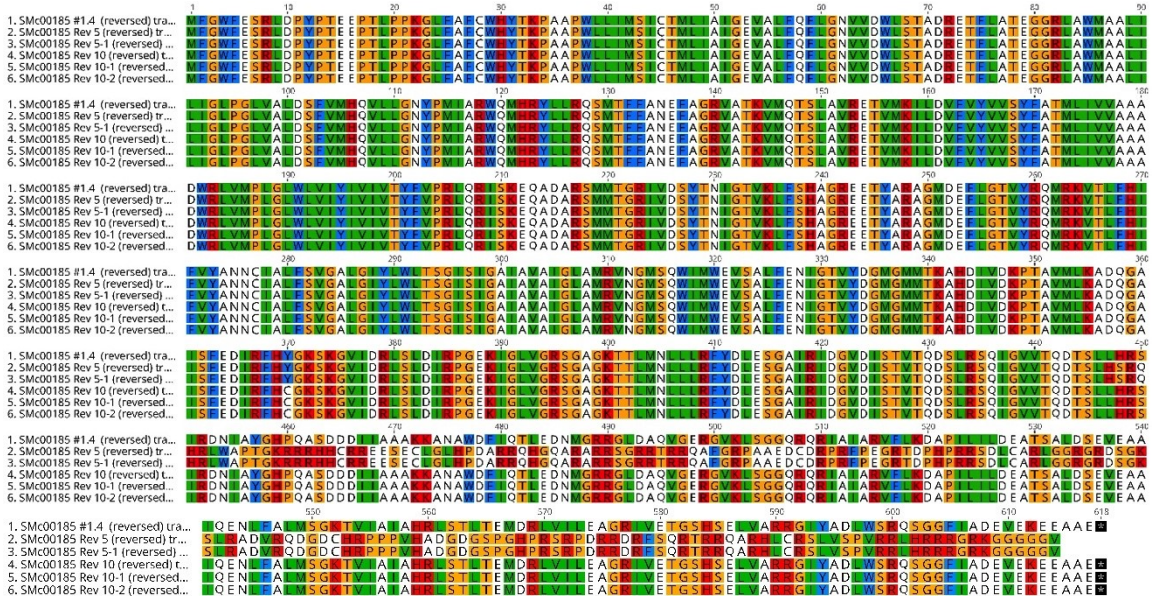


Figure 13. Translated nucleotide alignment of SMc00185(ABC transporter, ATP-binding protein).

Black colored* indicates stop codon at respective residue.

Table 5. SNPs and Indels obtained on ABC Transporter, ATP- binding protein.

Minimum	Amino acid Change	Coverage	Protein Effect	Polymorphism Type	Variant Frequency	Variant P-Value (approximate)	Strain
1978623	Y -> C	150	Substitution	SNP (transition)	100.00%	0	Rev 10
1978623	Y -> C	50	Substitution	SNP (transition)	100.00%	1.00E-180	Rev 10-1
1978623	Y -> C	92	Substitution	SNP (transition)	100.00%	1.6E-313	Rev 10-2

iii. SMc03826 (ArgG)

Additional mutation was identified in the region of SMc03826 in revertant strain Rev_8. SMc03826 codes for argininosuccinate synthase (Soon-Young *et al.*, 2003). It was a nucleotide substitution with SNP transition leading to change from cytosine to thymine resulting in the AA change from valine to isoleucine at residue 224 (**Figure 14**; **Table 6**). The isoleucine is bulkier in nature and C-beta branched AA alike valine.

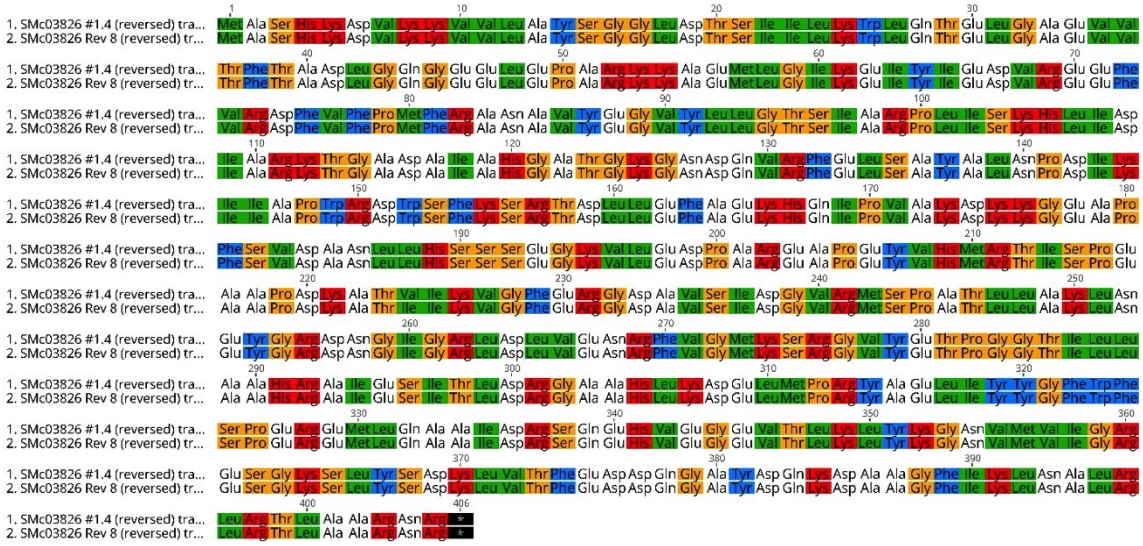


Figure 14. Translated nucleotide alignment of SMc03826 (ArgG).

* indicates stop codon at respective residue.

Table 6. SNPs and Indels obtained on ArgG coding sequence.

Minimum	Amino acid Change	Coverage	Protein Effect	Polymorphism Type	Variant Frequency	Variant P-Value (approximate)	Strain
3493003	V -> I	66	Substitution	SNP (transition)	100.00%	4.00E-225	Rev 8

4.3 Isolation and analysis of secondary revertants from initial RF prototrophic revertants

While in general the revertants exhibited Fix- phenotype, in some cases, the Rev_5 and Rev_10 produced pink nodules, which, based on the improved host-plant growth, might have Fix+ phenotype (**Figure 16**). The structure of these nodules was similar to nodules formed by Rm1021xpCPP30. (**Figure 16**). The alfalfa plants inoculated with Rev_5 and Rev_10 with occasional pink nodules were apparently bigger and more green than the plants without inoculation, suggesting that these nodules could have N fixation activity. The hypothesis was that the presence of secondary mutations in these revertants could further improve the function of RF biosynthesis in the mutants improving their symbiotic performance. To verify this hypothesis, the bacteria from these apparently effective nodules were isolated and tested for their free-living and symbiotic phenotype.

a. Isolation of symbiotically effective RF revertants from Fix+ alfalfa plants

The bacteria were isolated from the symbiotically effective root nodules. The isolated bacteria were streaked on $MMNH_4$ media plates and kept in an incubator for 7-8 days and further streaked at least 3 times on $MMNH_4$ plates supplemented with Tc^r to

obtain a pure culture. The isolated strains of Rev_5-1, Rev_10-1, and Rev_10-2 were further tested for their symbiotic performance with alfalfa host.

b. Evaluation of symbiotic phenotype of secondary RF prototrophic revertants

The symbiotic performance of the secondary revertants was tested in plant tests in laboratory conditions as described above. The shoot dry mass per box was recorded and phenotypic appearance of the plants and nodules were observed and recorded (**Figure 15**). As expected, wild type strain Rm1021xpCPP30 produced effective symbiosis with host-plant forming effective nodules and providing the plant with sufficient N for growth (**Figure 15**; **Figure 16**). The parental strain 1.4 did not form any nodules in the host-plant roots and the plant exhibited N starvation phenotype. The original revertants Rev_5 and Rev_10 produced numerous white nodules, which were apparently ineffective, since this symbiosis did not provide the host-plant with sufficient N for growth. However, in the case of secondary revertants Rev_5-1, Rev_10-1, and Rev_10-2, the effective nodulation with Fix+ phenotypes were observed. The mass of the plants nodulated with secondary revertants Rev_5-1, Rev_10-1, and Rev_10-2 were similar to the mass of the plants nodulated with Rm1021xpCPP30 and the all of the revertants form pink healthy-looking nodules.

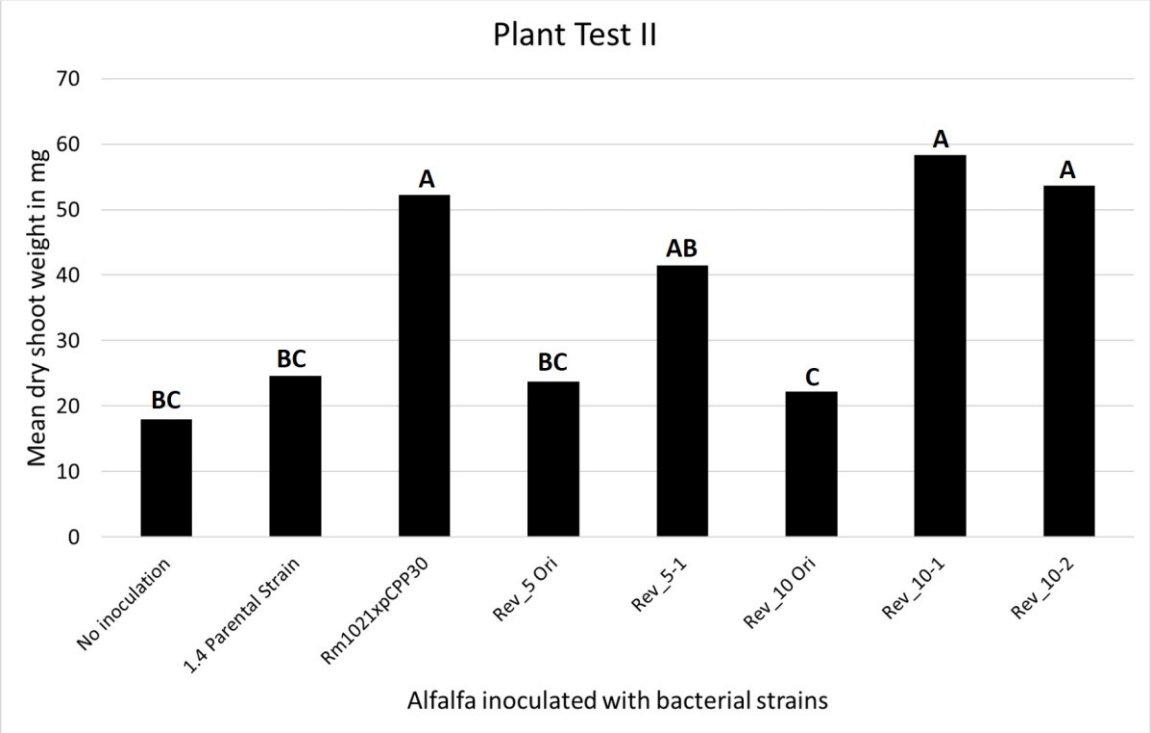
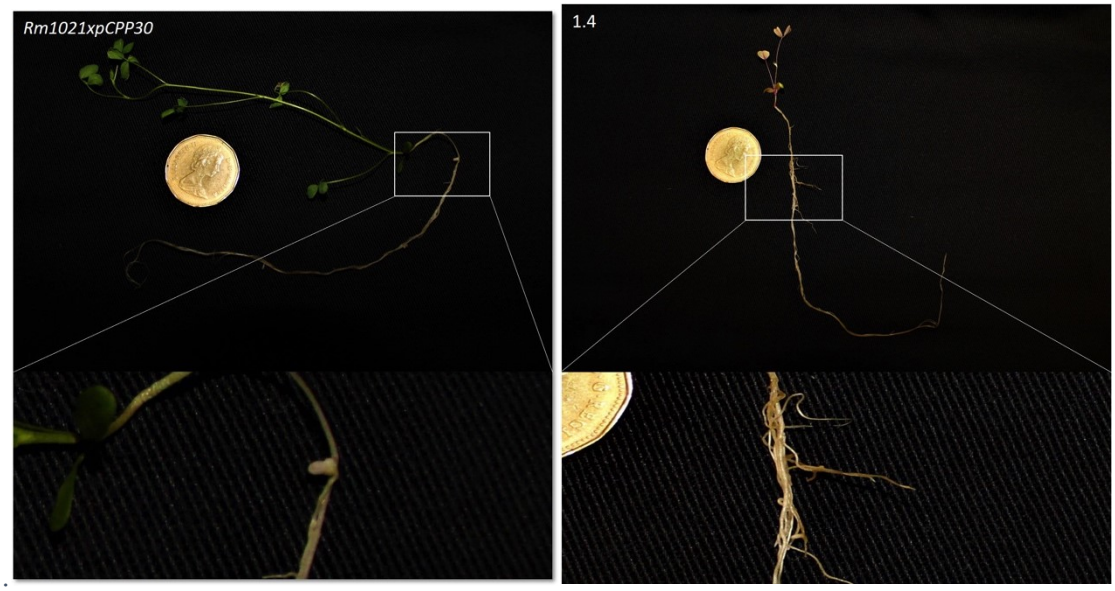
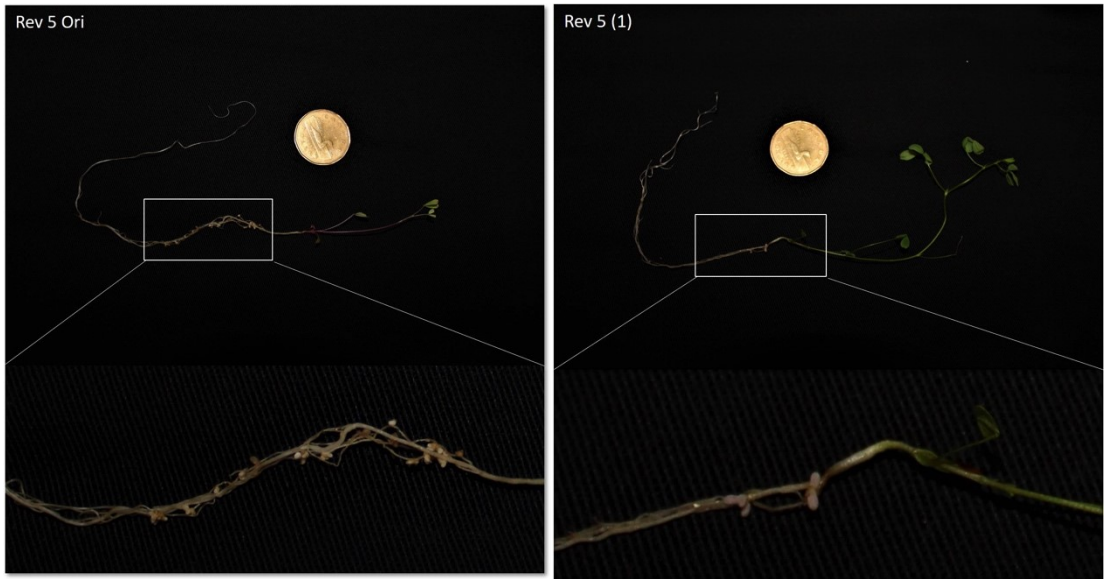


Figure 15. Plant test with alfalfa (*Medicago sativa* L.) inoculated with secondary RF prototrophic revertants, parental strains, and Rm1021xpCPP30.

The plant was grown for 4 weeks and dry shoot weights were measured and recorded per box (5 alfalfa shoot per box). Each treatment had three replications. The dry shoot mass of three replications for each strain treatment was averaged using Minitab One-way ANOVA. The One-way ANOVA test was used to determine significant ($P<0.05$) differences among the mean weights of the inoculated alfalfa shoot. The letters above bar indicates the grouping statistics for mean dry weight of alfalfa shoot inoculated with respective strains.


Rm1021xpCPP30
1.4


Rev 5 Ori
Rev 5 (1)

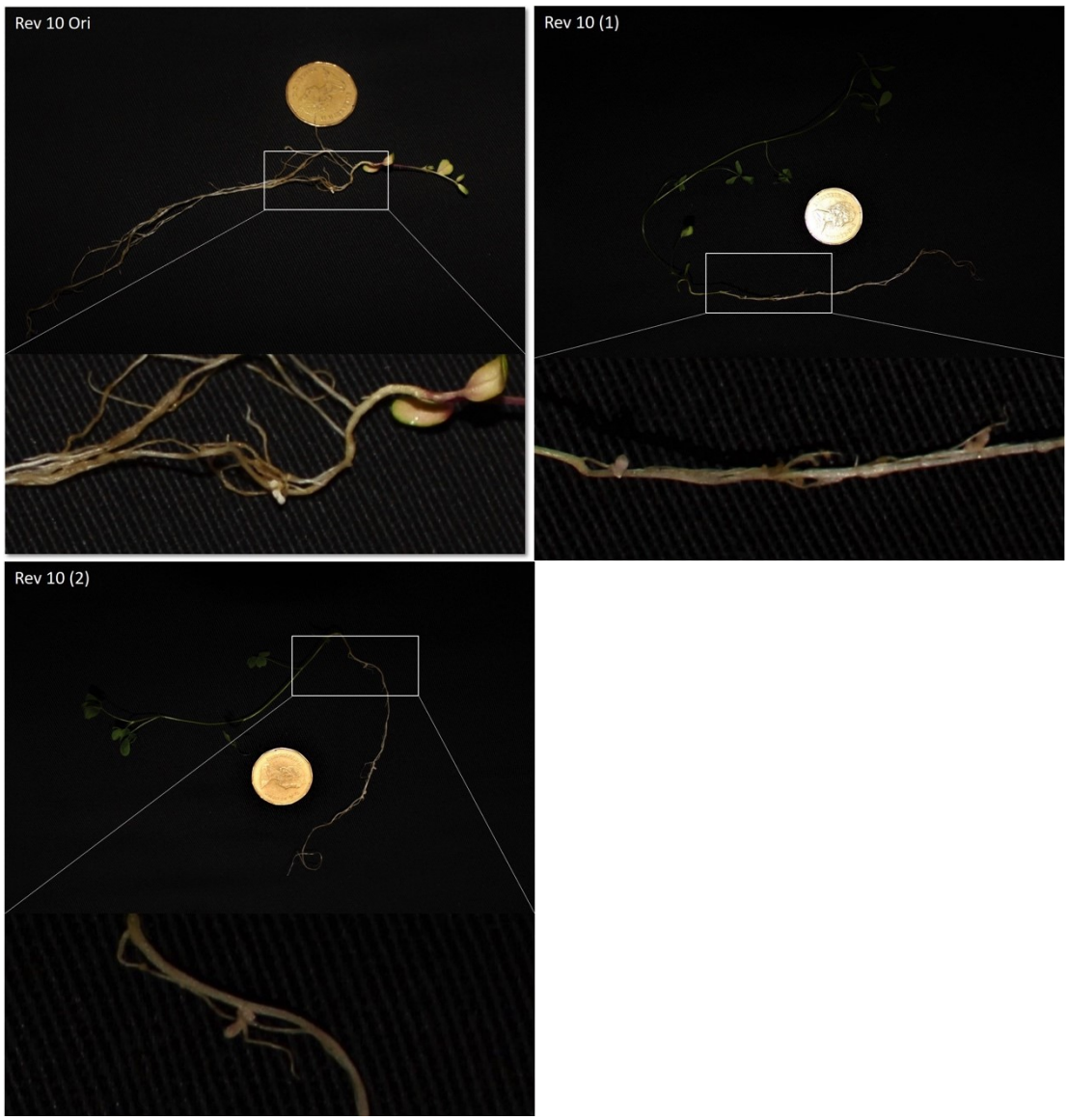


Figure 16. Typical alfalfa phenotype of Plant Test II inoculated by strains of secondary revertants.

From top left to right; Rm1021xpCPP30 was used as control wild type, which showed normal efficient pink root nodules (Fix+). The 1.4 parental strain didn't produce nodules on alfalfa roots. In the pictures of Rev_5 Ori and Rev_10 Ori, showed similar nodulation as described in plant test I with absence of effective pink nodules and presence of high number of white nodules (Fix-). Alfalfa inoculated with Rev_5-1, Rev_10-1 and Rev_10-2 showed pink effective nodules (Fix+).

c. Flavin characterization test for secondary revertants

Secondary revertants were tested for their ability to secrete FL as described in the Material and Methods. The result showed that there was a significant decrease in FL production ($P<0.05$) between Rm1021xpCPP30 and the secondary revertants Rev_5-1, Rev_10-1, and Rev_10-2 (**Figure 17**). There was no difference in FL secretion between original revertant (Rev_5 and Rev_10) and secondary revertants (Rev_5-1, Rev_10-1, Rev_10-2).

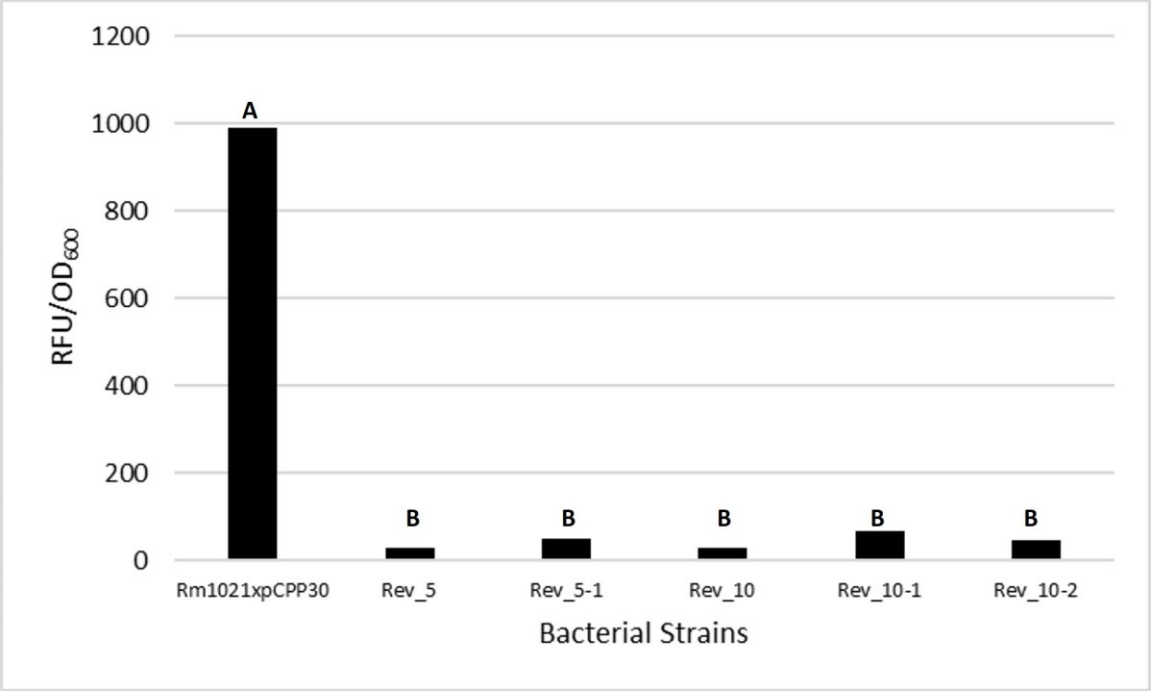


Figure 17. FL characterization of secondary revertants.

The Rm1021xpCPP30 showed significance ($P<0.05$) with higher FL secretion as compared to other revertants including secondary revertants Rev_5-1, Rev_10-1, and Rev_10-2.

4.4 Mutations observed on secondary revertants

To analyze the genetic changes, which lead to improved symbiotic performance of the secondary mutants, the whole genome sequencing was performed. The number of reads obtained after strains sequencing is indicated (**Table 2)**. After mapping the reads, potential mutations were selected based on previous criteria such as high confidence and high variant frequency. The mutations were found only in the genomes of the secondary mutants but not in the original ones were selected.

a. Mutations Observed in Rev_5-1

A new mutation was observed in Rev_5-1 genome, in the gene annotated as hypothetical protein, SMb21442, located on the mega plasmid pSymB. This mutation was present only in Rev_5-1 and not in the original Rev_5. This may have resulted in an accumulation of mutation during the plant host interaction process. The mutation was a deletion of 15 nucleotides with coverage of 26 and variant frequency of 100% (**Figure 18**). Due to low variant frequency and/or low coverage, other mutations were rejected from downstream analysis.

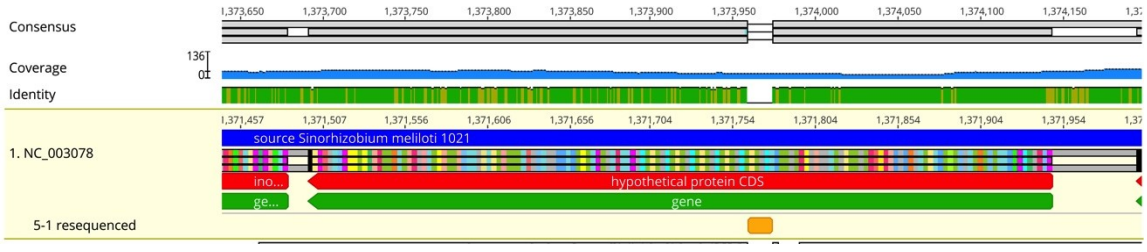


Figure 18. Screenshot of variations observed on SMb21442 (hypothetical protein) present only in Rev_5-1.

b. Mutations observed in Rev_10-1 and Rev_10-2

An identical single new mutation was identified in Rev_10-1 and Rev_10-2 Oa *arfB* coding region (**Figure 19**). The alignment of Oa *arfB* with the original Oa *arfB* sequence identified AA changes from aspartic acid to glycine at the position 371 (**Figure 19**). Aspartic acid is charged amino acid with side chain whereas glycine is hydrophobic amino acid buried in protein core. The change in the amino acid may probably change the protein folding or conformation as glycine is highly conserved within protein family. This mutation was only observed in the secondary revertants but not in the original Rev_10, suggesting the acquisition of the mutations during symbiotic interaction of Rev_10 with host-pant.

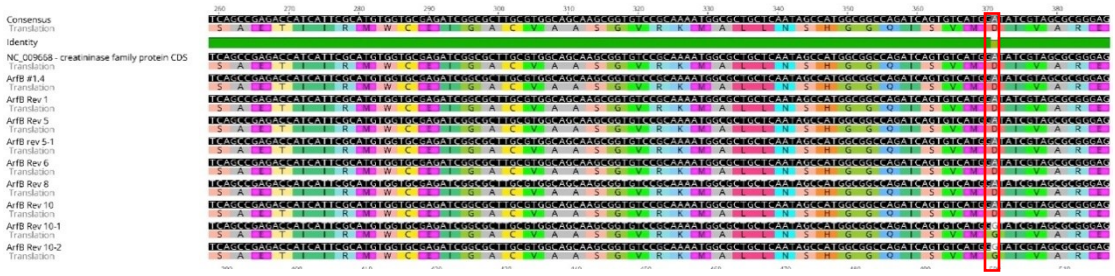


Figure 19. Modified screenshot of *arfB* alignment with Oa *arfB*.
The red outlined rectangle shows the position of variations present in Rev_10-1 and Rev_10-2.

CHAPTER 5: DISCUSSION

The introduction of Oa *arfB* in *S. meliloti* strain missing native *arfB* did not complement RF auxotrophy in the mutants. This indicated that the Oa ArfB was non-functional in the Rm1021 cells. However, in contrast to Rm1021Δ*arfB*, the deletion mutant carrying Oa *arfB* formed revertants capable of growing without RF supplementation, which was probably a result of some level of restoration of RF biosynthesis. The fact that Oa ArfB could gain functional capabilities for RF synthesis, probably by spontaneous mutation(s) in Rm1021Δ*arfB* genome or in Oa *arfB* coding region, suggested that only a minor and small number of modifications were necessary for this process. This in turn indicated a significant similarity between Rm1021 ArfB and Oa ArfB in their function and structure. Previous phylogenetic analysis also indicated significant homology between these proteins. On the other hand, all revertants grew slower than Rm1021 without RF supplementation. This indicates only partial restoration of RF biosynthesis among the revertants needed for its internal metabolic growth.

5.1 Partial restoration of RF prototrophy growth

All the revertants used in the study were isolated independently ensuring independent mutations events. However, if only a single and small difference between Rm1021 and Oa *arfB* affected the function of Oa ArfB in Rm1021 cells, similar effect of these independent mutations would be anticipated on the restoration of Rm1021Δ*arfB* (Oa *arfB)* RF auxotrophy phenotype, resulting in the similar growth of the revertants. However, several revertants with different growth were selected. More specifically, the

Rev_1, Rev_6, Rev_8 were able to grow comparatively faster than Rev_5 and Rev_10. This suggested that the restoration of RF prototrophy in the revertants might have several different mechanisms. For example, different genes could be modified in the case of the revertants. In confirmation to this hypothesis, the Rev_10 had genetic changes in ABC transporter ATP-binding protein, while Rev_8 had genetic alteration in ArgG coding region. In both cases the effect of these mutations resulted in different growth and symbiotic phenotype of the revertants.

RF supplementation restored the growth of Rev_1, Rev_6 and Rev_8 to the level of Rm1021x*pCPP30*, probably providing additional RF supplementation for the growth. On the other hand, the presence of RF in the medium did not restore the growth of revertants Rev_5 and Rev_10 to a similar level to Rm1021x*pCPP30*. One explanation of this phenotype could be that an accumulation of intermediate metabolites produced by RF biosynthetic reactions acting in reverse direction caused RF toxicity to the cells (Frelin *et al*., 2015). However, since *S. meliloti* lack a transporter system, the intake of RF by the cells was through the passive diffusion facilitated by high concentration of RF supplementation in the growth media (Yurgel *et al*., 2014). Consequently, the rate of passive diffusion of the RF might be altered due to changes in cell wall structure in these revertants. Another explanation could be that the mutations in these revertants attenuated the function of Oa ArfB but also affect another unrelated function to RF biosynthesis, resulting in decreased growth.

5.2 Flavin characterization

Similar to the growth phenotype, the FL secretion by the revertants was not restored to the level of Rm1021xpCPP30. It is possible that the level of RF biosynthesis in the revertants was sufficient for the internal needs of bacteria and not sufficient for secretion. On the other hand, based on the fact that ArfB and RibBA form complexes (Dr. Yurgel personal communication) and RibBA was essential for RF secretion (Yurgel *et al.* 2014), it was hypothesized that the FL secretion might require complex formation between ArfB and RibBA. Consequently, the attenuation of Oa ArfB function in the revertants might be sufficient for production of RF to support cell growth, but not sufficient to form fully functional protein complex with RibBA or other potential proteins involved in RF secretion.

5.3 Plant test

In contrast to parental strains not forming nodules, the revertants typically formed nodules when inoculated on alfalfa. This suggested that revertants produced sufficient amounts of RF to be able to nodulate alfalfa. The nodulation process involved infection thread formation by rhizobia into the root hairs of *Medicago sativa* L. (Humann *et al*., 2009). Once infected, symbiosomes are formed by a process of steps involving signaling interaction, formation of infection thread and differentiating bacteroids. These bacteroids fix atmospheric N_2 into ammonia, available for plant growth. However, based on the host-plant growth and nodules phenotype, the nodules formed by revertants were

ineffective. For example, revertants produces white nodules implying the absence of leghemoglobin, a necessary component of N fixation (Becana and Klucas, 1992). Also, the host-plants nodulated with the revertants exhibited N starvation phenotype. This might be a result of FL deficiency in revertants – in free-living conditions they did not grow to the level as compared to Rm1021 without RF supplementation. FL are an essential part of N fixation (Macheroux *et al*., 2011). For example, active FL groups of plants are involved in enzymatic reduction of non-functional ferric leghemoglobin. FL increases rate of reaction by acting as an intermediate electron carrier between NADH and ferric leghemoglobin (Becana and Klucas, 1992). The transport of electrons from dehydrogenase to nitrogenase in bacteroids involves microbial flavoprotein from bacteroids (Wong *et al*., 1971). Additionally, the transcription activation of *nif genes* (encoding for nitrogenase complex) is controlled by a regulatory flavoprotein NifL by modulating NifA activity (Hill *et al*., 1996). In the revertants, insufficient RF biosynthesis might result into deficiency of flavoproteins, which are required for effective Fix+ phenotype.

5.4 Revertants with Fix+ phenotype

Occasionally, the revertants Rev_5 and Rev_10 formed effective N fixing nodules. The root nodules appeared to be similar to Rm1021xpCPP30. The root nodules were pink in color and elongated, indicating presence of bacteroids and leghemoglobin. The two hypotheses were considered to explain this phenomenon. First of all, it was possible that some deviation in the metabolic activities of a small number of *S. meliloti*

cells could help the revertants to overcome RF deficiency and initiate N fixation. In this case, the bacteria isolated from the pink and apparently healthy N fixing nodules would have the symbiotic phenotype similar to the original revertants, producing mostly white and apparently ineffective nodules with occasional apparently effective nodules. This phenotype of bacterial isolated from nodules was observed for Rev_8.

Secondly, during growth and differentiation in plant root tissue, the revertants could accumulate additional mutations, which further attenuated Oa ArfB functions, making the protein function sufficient for establishment of effective symbiosis. In this case, the bacteria isolated from the apparently effective nodules would retain Fix+ phenotype producing N fixing nodules and providing the host-plant with sufficient amount of N for growth. After re-isolation from nodules, several secondary revertants exhibited effective symbiotic phenotype, similar to Rm1021xpCPP30. The shoot dry mass of the alfalfa plants inoculated with secondary revertants Rev_10-1, Rev_10-2 and Rev_5-1 were similar to that of alfalfa inoculated with Rm1021xpCPP30 *(***Figure 15**). The nodules formed by the secondary revertants were pink indicating the presence of leghemoglobin. This further indicated that the secondary mutations restored N fixation capability of the strains, probably by genetic modification in the known/unknown enzymes, which might be linked to the RF biosynthesis. On the other hand, the secondary revertants secreted less FL than Rm1021xpCPP30, confirming the previous report that FL secretion by *S. meliloti* was not enough for N fixation (Yurgel *et al*., 2014). It also suggested that the potential of complex formations between Oa ArfB and RibBA wasn't enough for RF secretion and did not provide the RF for secretion. The identification of the differences between the secondary revertants and primary original revertants was

done by re-isolating secondary revertant rhizobia from the root nodules, isolating the DNA and whole genome sequencing.

5.5 Genetic changes in revertants

The next step was to identify the genetic changes in revertants, which resulted in the partial restoration of RF prototrophy and symbiotic performance of the mutants. Our first hypothesis was that Oa *arfB* accumulated mutations that attenuated its function. The sequencing of Oa *arfB* didn't identify any genetic changes in the coding region of Oa *arfB* present in the initial set of revertants (**Figure 11**). This result led us to further process of sequencing and identification of genetic modification in the entire genomes of the revertants. In this part, the mutations found in the secondary revertants with Fix+ phenotype are discussed.

a. SMb21171 (*phnM*)

The PhnM gene (SMb21171) was annotated as putative phosphonate metabolism protein. Gene ontology provided the molecular function of this protein as hydrolase, which acted on carbon-nitrogen bond but not peptide bond, similar to function of GCHIII (https://iant.toulouse.inra.fr/bacteria/annotation/cgi/rhime.cgi). The mutations in this gene were identified in every revertant, suggesting the PhnM was a strong candidate involved in RF biosynthesis. The mutations identified in the revertants Rev 1, Rev_8, Rev_6 and Rev_5 resulted in the truncated and likely nonfunctional proteins. Based on this, the hypothesis that Rm1021 PhnM could form inactive complex with Oa ArfB preventing its

function, and the truncated PhnM could not form the complex and inactivate Oa ArfB. This implied that there might be another GCHIII present in *S. meliloti* genome.

In the case of the Rev_10, the substitutions of nucleotide might have changed AA in active sites, which may have changed the protein conformation. The AA changes from alanine to valine might have affected the protein conformation as valine is more bulky in nature and forms protein cores as result of hydrophobic properties. This mutated PhnM might form active complex with Oa ArfB. The secondary revertants Rev_10-1 and Rev_10-2 had a single nucleotide substitution, which changed AA in Oa ArfB from aspartic acid to glycine. The aspartic acid is polar charged whereas glycine is aliphatic non-polar AA. Our hypothesis was that the mutation in PhnM in primary revertants attenuated the complex formation between Oa ArfB and PhnM, which resulted in partially active RF biosynthesis. The secondary mutation might have further attenuated Oa ArfB restoring RF biosynthesis and symbiotic phenotype in secondary revertants Rev_10-1 and Rev_10-2.

b. SMc00185

Mutations were identified in the coding sequence of SMc00185. This gene codes for ABC transporter ATP-binding transmembrane protein. The same mutation was observed in Rev_10, and secondary revertants, Rev_10-1, and Rev_10-2. The SMc0185 has molecular function of ATPase along with transmembrane movement of substances. Previously it was shown that *S. meliloti* secret substantial amounts of FL (Yurgel *et al*., 2014). This implied the presence of FL transporter system in the rhizobia. Putative

transporter (SMc00185) identified in our study could carry FL secreting function in α-proteobacteria.

c. ArgG (SMc03826)

The mutation in ArgG was only found in Rev_8. The ArgG codes for argininosuccinate synthase are involved in arginine biosynthetic pathway (Soon-Young *et al.*, 2003). The ArgG gene is important for arginine biosynthesis and involved in the nitrogen fixation. *S. meliloti* arginine biosynthetic mutants, including *argG* mutant is required arginine for growth and induced ineffective white nodules with white, spherical shape. This indicated the importance of arginine in forming effective nitrogen fixing zone and transforming rhizobia into bacteroids (Kumar *et al*., 2003). Further analysis might need to identify the function of ArgG and its involvement in RF biosynthesis.

d. SMb21442

The genetic change in SM21442, a deletion of 15 nucleotides, was observed in secondary revertant Rev_5-1. Mutation were not observed in Rev_5, suggesting that this is a secondary mutation probably involved in attenuation of RF biosynthesis. The accumulation of this mutation might be occurred during the interaction process with the host-plant. The deletion might change protein conformation structure or affect its active site. The function of these protein is yet unknown and might be linked to somehow forming effective N fixing nodules phenotype.

CHAPTER 6: CONCLUSION

The alternate enzyme involved in RF biosynthesis in *S. meliloti* 1021, ArfB, with potential functions as formamide hydrolase was identified previously by Dr. Yurgel's research team. The hypothesis of this project was that an alternate enzyme in the RF biosynthesis pathway, ArfB, could form functional complex with RF biosynthetic enzymes in α-proteobacteria. This study combined a complementation approach and next generation sequencing to uncover the potential proteins involved in complex formation with the ArfB protein. The objective of this project was to identify the genetic changes that lead to RF prototrophy in *S. meliloti* 1021 strain carrying exogenous Oa ArfB and, therefore, identifying novel/alternate enzymes that might be involved with RF metabolism in α-proteobacteria. The RF prototrophic revertants of *S. meliloti* 1021 strain carrying exogenous Oa ArfB were selected. The mutations potentially linked to restoration of the RF prototrophy and RF metabolism were identified in these revertants. To analyze the effect of these mutations on free-living and symbiotic phenotype of the revertants, their growth properties, FL secretion, and the ability to form effective symbiosis with host-plant were tested. As a result, several SNP and Indels in the proteins, potentially, linked to RF metabolism were identified. These includes PhnM (putative hydrolase/GCHIII), SMc00185 (ABC transporter ATP-binding protein), ArgG (probable argininosuccinate synthase).

The identified enzymes may potentially function as alternative enzyme involved in the alternate RF biosynthetic pathway. The slow growth rate of the revertants and their low level of RF secretion suggested partial restoration of RF biosynthesis. The revertants gained the ability to form nodules on host-plant, but exhibited ineffective symbiotic

phenotype, indicating the importance of FL in the N fixation in *Rhizobium*-legume symbiosis. Several secondary revertants were also obtained with improved symbiotic performance, suggesting further improvement of efficient RF production in the strains and leading to significant N fixation. The additional mutations in the secondary revertants were localized in exogenous Oa ArfB and SMb21442 (hypothetical protein). Additionally, our data suggested the presence of another alternative GCHIII in *S. meliloti*. The PhnM have molecular function of putative hydrolase activity acting on carbon -N bonds, but not peptide bonds. This protein PhnM have similar activity as to GCHIII which makes a potential candidate for alternative enzymes in *S. meliloti* 1021.

The future directions of this study are to determine the functions of the identified proteins and their precise role in RF metabolism in *S. meliloti* and other α-proteobacteria. It is also important to test the presence of the alternative enzyme to GCHIII in rhizobia. This research will help us to better understand the alternate enzymes involved in RF biosynthesis. RF is a vital vitamin for metabolic activities in bacteria. By identifying novel enzymes, potential strategies to control pathogenic activities of pathogenic α-proteobacteria such as *Brucella spp.* and *Candidatus Liberibacter spp* may be developed. This research is important to further studies in plant host-microbe interactions and helps scientific researchers to improve N fixation though *Rhizobium*-legume association.

REFERENCES

Ashoori, M. and Saedisomeolia, A. 2014. Riboflavin (vitamin B2) and oxidative stress: a review. *British Journal of Nutrition*, 111: 1985-1991.

Bacher, A., Eberhardt, S., Fischer, M., Kis, K. and Richter, G. 2000. Biosynthesis of vitamin B2 (Riboflavin). *Annu Rev Nutr,* 20: 153-67.

Becana, M. and Klucas, R. V. 1992. Oxidation and reduction of leghemoglobin in root nodules of leguminous plants. *Plant Physiol,* 98: 1217-1221.

Bøckman, O. C. 1997. Fertilizers and biological nitrogen fixation as sources of plant nutrients: Perspectives for future agriculture. *Plant and Soil*, 194: 11-14.

Boyd, S.R. 2001. Nitrogen in future biosphere studies. *Chemical Geology*,176.

CGEB-IMR, Halifax (http://cgeb-imr.ca/)

Coba de la Peña, T. and Pueyo, J. J. 2012. Legumes in the reclaimation of marginal soils, from cultivars and inoculant selection to transgenic approches. *Agron. Sustain. Dev*., 32: 65-91.

Cooper, W. R., Sengoda, V. G. and Munyaneza, J. E. 2014. Localization of '*Candidatus Liberibacter solanacearum*' (Rhizobiales: *Rhizobiaceae*) in *Bactericera cockerelli* (Hemiptera: *Triozidae*). *Annl Entomol Soc Am*, 107(1): 204-210.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res,* 32(5): 1792-1797.

Edwards, A. M. 2014. Structure and general properties of flavins. Flavins and flavoproteins: methods and protocols, Methods in molecular biology. *Spring Sci Business media*, New York, 1146.

EPA. 2016. Overview of Greenhouse Gases. Washington, DC, USA: U.S. *Environmental Protection Agency*.

Fedoroff, N. V. 2015. Food in a future of 10 billion. *Agric & Food Secur*, 4(11): 1-10.

Finan, T. M., Weidner, S., Wong, K., Buhrmester, J., Chain, P., Vorholter, F. J., Hernandez-lucas, I., Becker, A., Cowie, A., Gouzy, J., Golding, B. and Puhler, A. 2001. The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endosymbiont *Sinorhizobium meliloti*. *PNAS*, 98(17): 9889-9894.

Fischer, M. and Bacher, A. 2011. Biosynthesis of vitamin B2: a unique way to assemble a xylene ring. *ChemBioChem*, 12: 670-680.

Frelin, O., Huang, L., Hasnain, G, Jeffryes, J. G., Ziemark, M. J., Rocca, J. R., Wang, B., Rice, J., Roje, S., Yurgel, S, N., Gregory III, J. F., Edison, A. S., Henry, C. S., deCrécy-Lagard, V. and Hanson, A. D. 2015. A directed-overflow and damage-control N-glycosidase in riboflavin biosynthesis. *Biochem J,* 466(1): 137-145.

Galibert, F., Finan, T. M., Long, S. R., Puhler, A., Abola, P., Ampe, F., Barloy-Hubler, F., Barnett, M. J., Becker, A., Boistard, P., Bothe, G., Boutry, M., Bowser, L., Buhrmester, J., Cadieu, E., Capela, D., Chain, P., Cowie, A., Davis, R. W., Dreano, S., Federspiel, N. A., Fisher, R. F., Gloux, S., Godrie, T., Goffeau, A., Golding, B., Gouzy, J., Gurjal, M., Hernandez-Lucas, I., Hong, A., Huizar, L., Hyman, R. W., Jones, T., Kahn, D., Kahn, M. L., Kalman, S., Keating, D. H., Kiss, E., Komp, C., Lalaure, V., Masuy, D., Palm, C., Peck, M. C., Pohl, T. M., Portetelle, D., Purnelle, B., Ramsperger, U., Surzycki, R., Thebault, P., Vandenbol, M., Vorholter, F. J.,

Weidner, S., Wells, D. H., Wong, K., Yeh, K. C., and Batut, J. 2001. The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science,* 293:668-672.

Garg, N., and Geetanjali. 2007. Symbiotic nitrogen fixation in legume nodules: process and signaling. A review. *Agron Sustain Dev,* 27: 59-68.

Grochowski, L. L., Xu, H. and White, R.H. 2009. An iron(II) dependent formamide hydrolase catalyzes the second step in the archaeal biosynthetic pathway to riboflavin and 7,8-didemethyl-8-hydroxy-5-deazariboflavin. *Biochemistry*, 48: 4181-4188.

Hagberg, K. L., Yurgel, S. N., Mulder, M. and Kahn, M. L. 2016. Interaction between nitrogen and phosphate stress responses in *Sinorhizobium meliloti*. *Front. Microbiol*., 7:1928.

Hartung, J. S., Shao, J., Kuykendall, L. D. 2011. Comparison of the '*Ca. Liberibacter asiaticus*' genome adapted for an intracellular lifestyle with other members of the *Rhizobiales*. *PLoS ONE*, 6(8): e23289.

Hill, S., Austin, S., Eydmann, T., Jones, T. and Dixon, R. 1996. A*zotobacter vinelandii* NIFL is a flavoprotein that modulates transcriptional activation of nitrogen-fixation genes via a redox-sensitive switch. *Proc Natl Acad Sci, USA,* 93: 2143-2148.

Hirsch, A. M., Lum, M. R., and Downie, J.A. 2001. What makes the Rhizobia-legume symbiosis so special? *Plant Physiol*, 127.

Huang, H., He, S. Y., Bauer, D. W. and Collmer, A. 1992. The *Pseudomonas syrnngae* pv. syringae 61 hrpH product, an envelope protein required for elicitation of the hypersensitive response in plants. *J Bacteriol,* 174(21): 6878-6885.

Humann, J. L., Ziemkiewicz, H. T., Yurgel, S. N. and Kahn, M. L. 2009. Regulatory and DNA repair genes contribute to the desiccation resistance of *Sinorhizobium meliloti* Rm1021. *Applied and Environmental Microbiology,* 75(2): 446-453.

Jones, K. M., Kobayashi, H., Davies, B. W., Taga, M. E., Walker, G. C. 2007. How rhizobial symbionts invade plants: the *Sinorhizobium– Medicago* mode. *Nat Rev Microbiol,* 5(8): 619-633.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. and Drummond, A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics,* 28(12): 1647-1649.

Kohlen, W., Liang Pin Ng, J., Deinum, E. E., Mathesius, U. 2018. Auxin transport, metabolism, and signalling during nodule initiation: indeterminate and determinate nodules. *Journal of Experimental Botany*, 69(2), 229–244.

Kumar, A., Vij, N. and Randhawa, G. S. 2003. Isolation and symbiotic characterization of transposon Tn5-induced arginine auxotroph of *Sinorhizobium meliloti*. *Indian Journal of Experimental Biology*, 41: 1198-1204.

Ladenstein, R., Fischer, M. and Bacher, A. 2013. The lumazine synthase/riboflavin synthase complex: shapes and functions of a highly variable enzyme system. *FEBS Journal*, 280: 2537–2563.

LeVier, K., Phillips, R. W., Grippe, V. K., Roop I. I, R. M. and Walker, G. C. 2000. Similar requirements of a plant symbiont and a mammalian pathogen for prolonged intracellular survival. *Science*, 287(5462): 2492-2493.

Long, S. R. 2001. Genes and signals in the Rhizobium-legume symbiosis. *Plant Physiol*, 125.

Macheroux, P., Kappes, B. and Ealick, S. E. 2011. Flavogenomics – a genomic and structural view of flavin-dependent proteins. *FEBS*, 278: 2625-2634.

Mahuku, G. S. 2004. A simple extraction method suitable for PCR based analysis of plant, fungal, and bacterial DNA. *Plant Molecular Biology Reporter,*22: 71-81.

Mansoorabadi, S. O., Thibodeaux, C. J. and Liu, H. W. 2006. The diverse roles of flavin coenzymes - nature's most versatile thespians. *J Org Chem*, 72(17): 6329-6342.

Miyata, S., Kato, H., Davis, R., Smith, M. W., Weinert, M. and Iwanami, T. 2011. Asian-common strains of '*Candidatus Liberibacter asiaticus*' are distributed in Northeast India, Papua New Guinea, and Timor-Leste. *J Gen Plant Pathol*, 77: 43-47.

Moreno, E., Stackebrandt, E., Dorsch, M., Wolters, J., Busch, M., and Mayer, H. 1990. *Brucella abortus* 16S rRNA and lipid a reveal a phylogenetic relationship with members of the alpha-2 subdivision of the class proteobacteria. *Journal of Bacteriology*, 172(7): 3569-3576.

NCBI database, *Sinorhizobium meliloti* 1021 genome (https://www.ncbi.nlm.nih.gov/genome/1004?genome_assembly_id=300472) (http://iant.toulouse.inra.fr/S.meliloti) *Ochrobactrum anthropi* (https://www.ncbi.nlm.nih.gov/nuccore/NC_009668.1)

Oke, V. and Long, S. R. 1999. Bacteroid formation in the Rhizobium–legume symbiosis. *Microbiology*, 2: 641-646.

Oldroyd, G. E. and Downie, J.A. 2008. Coordinating nodule morphogenesis with Rhizobial infection in legumes. *Annu Rev Plant Biol,* 59:519-46.

Oldroyd, G. E. 2013. Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nature reviews, Microbiology*, 11: 252-263.

Paulsen, I. T., Seshadri, R., Nelson, K. E., Eisen, J. A., Heidelberg, J. F., Read, T. D., Dodson, R. J., Umayam, L., Brinkac, L. M., Beanan, M. J., Daugherty, S. C., Deboy, R. T., Durkin, S. A., Kolonay, J. F., Madupu, R., Nelson, W. C., Ayodeji, B., Kraul, M., Shetty, J., Malek, J., van Aken, S. E., Riedmuller, S., Tettelin, H., Gill, S. R., White, O., Salzberg, S. L. 2002. The *Brucella* suis Genome Reveals Fundamental Similarities between Animal and Plant Pathogens and Symbionts. *PNAS*, 99(20): 13148-13153.

Phillips, D. A., Joseph, C. M., Martinez-Romero, E., Sanborn, J. R., Volpin, H. 1999. Identification of lumichrome as a *Sinorhizobium* enhancer of alfalfa root respiration and shoot growth. *PNAS*, 96(22): 12275-12280.

Raman, S. S., Vijayaraj, R., Parthasarathi, R. and Subramanian, V. 2008. Helix forming tendency of valine substituted poly-alanine: a molecular dynamics investigation. *J Phys Chem B,* 112, 9100-9104.

Rajwar, A., Sahgal, M. and Johri, B. N. 2013. Legume-Rhizobia symbiosis and interaction in agroecosystems. *Plant Microbe Symbiosis: Fundamentals and advances*, *Springer,* India, 233-265.

Roje, S. 2007. Vitamin B biosynthesis in plants. *Phytochemistry*, 68(14): 1904–1921.

Roy, R. N., Misra, R. V. and Montanez, A. 2002. Decreasing Reliance on Mineral Nitrogen: Yet More Food. *Ambio*, 31(2): 177-183.

Sa, N., Rawat, R., Thornburg, C., Walker, K. D. and Roje, S. 2016. Identification and characterization of the missing phosphatase on the riboflavin biosynthesis pathway in *Arabidopsis thaliana*. *The Plant Journal*, 88: 705-716.

Somerville, J. E. and Kahn, M. L. 1983. Cloning of the glutamine synthetase I gene from Rhizobium meliloti [Nitrogen fixation, nitrogen-fixing bacteria]. *Journal of Bacteriology*, 1: 168-176.

Soon-Young, K., Yim, S., Lee, H. and Lee, M. 2003. Isolation and analysis of the argG gene encoding argininosuccinate synthetase from *Corynebacterium glutamicum*. *J Microbiol Biotechnol,* 13(6): 949-954.

Timmers, A. C. J., Soupène, E., Auriac, M., de Billy, F., Vasse, J., Boistard, P. and Truchet, G. 2000. Saprophytic intracellular rhizobia in alfalfa nodules. *MPMI*, 13(11): 1204-1213.

United Nations, Department of Economic and Social Affairs, Population Division. 2015. World Population Prospects: The 2015 Revision, Key Findings and Advance Tables. New York: Working Paper No. ESA/P/WP.241.

Vance, C. P. 2001. Symbiotic nitrogen fixation and phosphorus acquisition: Plant nutrition in a world of declining renewable resources. *Plant Physiology,* 127 (2).

Vanderlinde, E. M., Harrison, J. J., Muszynski, A., Carlson, R. W., Turner, R. J. and Yost, C. K. 2009. Identification of a novel ABC transporter required for desiccation tolerance,

and biofilm formation in *Rhizobium leguminosarum* bv. viciae 3841. *FEMS Microbiol Eco,* 71: 327-340.

Wong, P., Evans, H. J., Klucas, R. and Russel, S. 1971. Investigations into the pathway of electron transport to the nitrogenase from nodule bacteroids. *Plant and Soil,*525-543.

Yang, G., Bhuvaneswari, T. V., Joseph, C. M., King, M. D. and Phillips, D. A. 2002. Roles for Riboflavin in the *Sinorhizobium*–Alfalfa Association. *MPMI*, 15(5): 456–462.

Yurgel, S. N. and Kahn, M. J. 2005. *Sinorhizobium meliloti dctA* mutants with partial ability to transport dicarboxylic acids. *J Bacteriol,* 187(3):1161-1172.

Yurgel, S. N., Berrocal, J., Wilson, C. and Kahn, M. L. 2007. Pleiotropic effects of mutations that alter the *Sinorhizobium meliloti* cytochrome c respiratory system. *Microbiology*, 153:399-410.

Yurgel, S. N., Rice, J., Mulder, M. and Kahn, M. L. 2010. GlnB/GlnK PII proteins and regulation of the *Sinorhizobium meliloti* Rm1021 nitrogen stress response and symbiotic function. *J Bacteriol*, 192(10):2473-2481.

Yurgel, S. N., Rice, J., Domreis, E., Lynch, J., Sa, N., Qamar, Z., Rajamani, S., Gao, M., Roje, S. and Bauer, W.D. 2014. *Sinorhizobium meliloti* flavin secretion and bacteria-host interaction: role of the bifunctional RibBA protein. *MPMI*, 27(5): 437–445.

Yurgel, S. N., Lynch, J., Rice, J., Adhikari, N. and Roje, S. 2014. Quantification of flavin production by bacteria. *Bio-Protocol*, 4(15): 1-7.

Zambriski, J.A., Saito, M., Nydam, D.V., Reyes-Garay, H.A., Castillo, R., Cepeda, D., Cespedes-Zambrano, M. J., Garcia-Vara, P., Maves, R. C., Solano, M., Torrico, F., Gilman, R. H. 2010. Assessment of *Brucella melitensis* disease burden in lactating

goats in Mizque, Bolivia. *14th International Congress on Infectious Diseases*, 34.020: e163.