Metabolism and host specificity in the *Rhizobium leguminosarum* species complex

Hui Kailin

Doctor of Philosophy

University of York

Biology

July 2014

Abstract

Forming symbiosis with legumes and fixing nitrogen, rhizobia can be ranked among the most ecologically significant bacteria in the world. Two features of rhizobia viz. metabolism and host specificity are important determinants of their growth and their ability to form symbiosis with plants. Investigations of these two features are common in studies conducted using laboratory based experimental and computer based bioinformatics methods.

Rhizobium is the largest genus of root-nodulating bacteria. Hence, abundant data can be collected by investigating members of this group. 72 isolates of root-nodule bacteria were collected from a field adjacent to Wentworth College, University of York from plants of genus *Vicia* and *Trifolium*; 36 from each. Their metabolic assays showed diversity. Analysis of genome data revealed the presence of five contiguous genes in isolates from *Vicia*, as well as in the reference strain *Rlv*. 3841; but absent in the isolates from *Trifolium*. Limited information was available on the five genes.

Work in this thesis promotes our understanding of *R. leguminosarum* genotype and phenotype. Cross nodulation assays showed that all biovar *viciae* strains retained their ability to form symbiosis plants of genera *Vicia, Pisum* and *Lathyrus.* Metabolic study and *celC* phylogeny showed that clusters of different species - type strains of *R. pisi* and *R. fabae*, and type strains of *R. phaseoli* and *R. etli* were closely related to each other. Studies on distribution of the five biovar-specific genes showed that they were widely distributed in biovar *viciae* strains with one exception – the type strain of *R. fabae* did not have any of five *bvs* genes, which suggested these *bvs* genes were not essential for nodulating *Vicia* plants. Single gene mutation of the five genes in *Rlv*. 3841 and subsequent competition nodulation tests suggested that inactivation of the *bvs4* and *bvs5* genes could lead to decreased competitiveness in mutants as compared to wild type on *Vicia sativa, Vicia faba and Pisum sativum* plants. Complementation of *bvs5* gene restored competitiveness. Enzyme assays for the two genes confirmed the activities described in their annotation viz. aliphatic nitrilase for *bvs4* and sulfite oxidase for *bvs5*. Mutation of *bvs4* results in a loss of *bvs5*-encoded enzyme activity, indicating that these two genes are probably in an operon.

In conclusion, the study shows that relations among rhizobial strains and species can be observed by studying bacterial metabolism. The work hints at a novel competitive mechanism that requires further investigation.

List of contents

List of figures

Chapter 2

Chapter 3

Chapter 4

List of tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Acknowledgements

First of all, I would like to thank my supervisor Professor Peter Young for all his help and support. It has been a pleasure working with him for the past four-years. His dedication and devotion to research and his patience provided with the impetus to complete my research work and this thesis. I would also like to express my gratitude towards my thesis advisory panel members - Dr. James Moir and Dr. Thorunn Helgason for their support, advice and suggestions during the four years of my PhD. Without their help it would be impossible for me to finish my work.

Next, I would like to thank all the previous and current members of the J1 group at the Department of Biology, University of York. Thanks to Shakeel for his friendly and helpful advice when I arrived as a novice in York. Thanks to Nitin and Piyachat for extending their friendship and their patience in helping me solve bioinformatics and statistics problems. A special thanks to Ganesh, his knowledge and dedication to science influenced me positively throughout the four years of my study. In addition, his brotherly affection gave me a warm and welcome feeling while studying in a foreign country.

I take this opportunity to thank Dr. Jürgen Prell for his advice on construction of gene mutants in rhizobia, Dr. Gail Preston for her advice on the study of nitrilases, Dr. Oliver Missa for his help with R, Dr. Euan James for providing the Scottish isolates of rhizobia and Osie Ampomah for the Swedish isolates of rhizobia. A warm thanks to all the staff in Biology Department for their help and support for the past four years.

Finally, I thank my parents for the courage and encouragement provided. They gave me the motivation to finish my studies and move forward to the next stage in my life. Thanks are due to my partner Lu Li, who encouraged me to consider life in an innovative way and has designed the future for us upholding and affirming the belief that the family is a group of people who contribute to a fruitful and meaningful life.

Author's declaration :

I, Kailin Hui, declare that I am the sole author for the work contained in this thesis. I also declare that all the work described in the thesis is the result of my PhD work and that it is original except, as indicated below. The thesis has been written by me and no part of it has been submitted for any other degree to any other institution.

Exception to declaration :

1. Nodulation tests, construction of mutants and complementation strains in *Rlv*. 3841 : This work was carried out with the help of Mr. Ganesh Lad who was also doing similar work for his PhD project. The laboratory work was distributed between the two of us for these experiments.

CHAPTER 1 : GENERAL INTRODUCTION

1.1. Introduction

Bacteria were amongst the first life forms on earth. They are found in habitats like air, soil, water, acidic host spring, radioactive waste and even on spacecraft (Fredrickson *et al.*, 2004). In addition, they may also exist in symbiotic and parasitic relationships with plants and animals. Bacteria constitute a large part of the prokaryotic organisms and exist in a diversity of shapes and size. Most bacterial species exist as single cells with a length of 0.5-5.0 μm, approximately one-tenth the length of a typical eukaryotic cell. The total biomass of bacteria surpasses that of plants and animals put together, since there are about $5x10^{30}$ bacteria on earth (Cleveland and Szostak, 2008, Whitman *et al.*, 1998).

Bacteria play an important part in the recycling of nutrients on earth, e.g. nitrogen-fixing bacteria take part in the nitrogen cycle to biologically fix nitrogen from the atmosphere (Wagner, 2012). In some extreme environments like those found near hydrothermal vents and cold seeps, bacteria were found to use hydrogen sulphide and methane to generate energy and provide for other living forms nearby (Cary and Giovannoni, 1993). Data from studies on bacteria living in other extreme conditions of pH, temperature and high salt concentrations, indicate that bacteria have universally adapted to their habitats (Eisenberg and Wachtel, 1987, Lowe *et al.*, 1993).

As a branch of microbiology, bacteriology is the study of bacteria. However, the characterization of bacteria is complex and far from complete. Only about half of the bacteria that have been described have been grown in the laboratory (Rappé and Giovannoni, 2003). On the other hand, further studies are required on the species that have been identified. Hence, classification of bacteria is a long journey but one that will significantly benefit our knowledge of the whole eco-system. Bacteria can be classified using a number of features. Different morphological, physiological and metabolic phenotypes such as Gram staining, ability to utilize various carbon or nitrogen sources, tolerance to environmental stress and pathogenicity of bacteria can be applied for classification. Common methods used in phenotype analysis to describe new species are substrate utilization and fatty acid analysis (Rashid *et al.*, 2012). The use of Biolog Phenotype Microarrays is now popular in bacterial studies since a number of

substrates commonly used in phenotypic characterisation are readily available on the plates which are easy to use and score the results. Some bacteria require special tests for characterisation. For example, the root-nodule bacteria like rhizobia are often characterised using cross-nodulation tests to get an idea of their host range or host-specificity (Friesen, 2012).

The development of modern methods of genetic analyses has allowed the use of genotype diversity in describing new species. Frequently used methods include DNA-DNA hybridization, DNA profiling and sequence comparison. In addition, phylogenetic analysis based on multi-gene sequences and RFLP data are also useful. Recent studies mostly use 16S-23S rRNA, ITS region and other specific genes like housekeeping genes.

Compared to eukaryotic genomes, the genomes of bacteria are more compact and the number of functional genes can be directly linked to genome size. Some of these genes may be grouped together under a common set of regulatory genes to form operons. The compact genome of bacteria is a result of the presence of very little noncoding DNA (Gregory, 2011, Koonin, 2009). Bacterial genomes also show less variation in size as compared to other living forms (Gregory, 2005). These features of bacteria make them an interesting model in studying the evolution of living forms.

One of the interesting groups of bacteria forms symbiosis with legume plants and is responsible for the fixation of atmospheric nitrogen and making it available to plants – the rhizobia.

1.2. Introduction to rhizobia

Rhizobia are aerobic, motile, Gram-negative, non-sporeforming, rod-shaped bacteria (when cultured on bacteriological media) with a 57-66% percent G+C content, capable of forming nodules on legume (and some non-legume) plants with or without nitrogen fixation (Kuykendall *et al.*, 2006).

1.2.1. Taxonomy of rhizobia

The term rhizobia is now used to refer to all bacteria with the ability to form nodules on the roots of legume plants and fix nitrogen within them. Previously, all rhizobia were included in the genus *Rhizobium* - the type genus for root nodulating and nitrogen fixing bacteria (or closely related bacteria). In the

1984 edition of Bergey's Manual of Systematic Bacteriology (Bergey *et al.*, 1984), *Rhizobium* (together with *Bradyrhizobium*), *Agrobacterium* and *Phyllobacterium* comprised the family *Rhizobiaceae*.

The ability of legume plants to assimilate nitrogen had been noticed by the end of the 19th century. In 1888, Beijerinck, in the Netherlands, first isolated root nodule bacteria, suggested their involvement in atmospheric nitrogen fixation and named them *Bacillus radicicola*. A year later, Frank assigned this bacterium into a new genus *Rhizobium* and described the first species - *Rhizobium leguminosarum* (Frank, 1889). After that, all discovered root–nodule bacteria were referred to as *Rhizobium*. The development of rhizobium taxonomy is dynamic and the original genus *Rhizobium* has now been split into a number of genera.

Rhizobia were considered difficult to classify. At the beginning of 20th century, the definition of rhizobia was based on the ability of the bacteria to nodulate different legume plants. Microbiologists extensively tested and defined cross-inoculation groups by selecting isolates to nodulate diverse legume hosts. In each cross-inoculation group, rhizobial isolates from one host plants were able to form nodules on other plants within the group (Fred *et al.*, 1932). However, this concept was soon abandoned since the taxonomic marker was demonstrated to be unreliable due to abnormalities and complexity of cross-inoculation among different groups (Graham, 1964, Wilson, 1944).

The diversity of morphological, nutritional and metabolic characters was used by bacteriologists from early 1960s in numerical taxonomy studies on rhizobia (Graham, 1964, Mannetje, 1967, Moffett and Colwell, 1968) along with other methods like serological study and simple DNA characterisations (De Ley and Rassel, 1965, Graham, 1963, Vincent, 1970).

Discovery of nodulation genes borne on plasmids (Brewin *et al.*, 1980, Nuti *et al.*, 1979) or on chromosomal islands and their transfer amongst soil bacteria further proved the unreliability of cross-inoculation based taxonomy analysis. These studies also contributed in showing the relatedness between *Rhizobium* and *Agrobacterium* and splitting the rhizobia into fast-growing and slow-growing groups (Graham, 1964). The slow-growing group was later included and described as a new genus *Bradyrhizobium* in the Bergey's Manual.

Since 1980s, the study of rhizobial taxonomy has made rapid progress due to the development of new techniques (e.g. DNA-DNA hybridization and classification using marker genes) and advancements in sequencing technology. The relationship between rhizobial strains is now becoming clear. Currently there are 128 species in 16 genera that constitute rhizobia. Two factors led to a substantial and dramatic increase in the number of rhizobial genera and species – the study of rare or geographically-restricted legume plants and the development and use of new taxonomic markers and methods.

Most members of rhizobia belong to class *Alphaproteobacteria* (120 out of 128). The genera included in class Alphaproteobacteria are : *Allorhizobium*, *Azorhizobium*, *Rhizobium*, *Mesorhizobium*, *Ensifer* (formerly *Sinorhizobium*), *Bradyrhizobium* (Martinez- Romero and Caballero-Mellado 1996; Young and Haukka 1996). Chen et al. (1988), first described the genus *Sinorhizobium*. However, the argument concerning *Sinorhizobium* and genus *Ensifer* (Casida 1982) belonging to a single taxon has been raised recently. *Ensifer* takes priority since it was annotated earlier in both heterotypic synonyms. The renaming of *Sinorhizobium* to *Ensifer* is a question that is not universally agreed on and in literature, the discussion is ongoing (Young, 2010).

Other genera in class Alphaproteobacteria include *Bosea*, *Ochrobactrum*, *Devosia*, *Methylobacterium*, *Microvirga*, *Aminobacter, Phyllobacterium* and *Shinella*, (Garrity *et al.*, 2005, Holmes *et al.*, 1988, Mantelin *et al.*, 2006). Species within these groups e.g. *Devosia neptuniae* (Rivas *et al.*, 2003), *Methylobacterium nodulans* (Sy *et al.*, 2001), *Ochrobactrum lupine* (Trujillo *et al.*, 2005), *Ochrobactrum cytisi* (Zurdo-Pineiro *et al.*, 2007), and *Shinella kummerowiae* (Lin *et al.*, 2008) have been demonstrated to be members of the *Alphaproteobacteria*. Species included in the class *Betaproteobacteria* including *Cupriavidus* (formerly *Ralstonia*) (Chen *et al.*, 2001) and *Burkholderia* (Moulin *et al.*, 2001, Vandamme *et al.*, 2002), also have the ability to fix nitrogen.

1.2.2. Genus *Rhizobium*

1.2.2.1. Introduction to the genus *Rhizobium*

Rhizobium (Gr. *rhiza* = root; *bios* = life) (Frank, 1889), was the first genus described for bacteria from root nodule. For many years this was the only

genus that included all strains of root-nodule bacteria. Today, *Rhizobium* is the type genus of family Rhizobiaceae. The type species of genus *Rhizobium* is *Rhizobium leguminosarum*, first described by Frank (1889).

The general characteristics of *Rhizobium* are as follows. The cells of *Rhizobium* are rods (0.5-1.0 x 1.2-3.0 μm). They are non-sporeforming, Gram negative bacteria and are mobile due to the presence of 1-6 peritrichous flagella. They are aerobic and use oxygen as the terminal electron acceptor in their respiratory chain. Some strains e.g. *R. leguminosarum* often contain metachromatic granules, composed of poly-β-hydroxybutyrate, which can be demonstrated by a staining test (Graham and Parker, 1964).

The structure of the cell wall of *Rhizobium* is similar to that of other Gram-negative bacteria. The peptidoglycan, forming the cell-wall, consists of glutamic acid, alanine, diaminopimelic acid, and amino sugars. The lipopolysaccharide (LPS) cell wall composition is diverse in different strains, but most strains contain 2-keto-3-deoxyoctanoic acid (KDO), uronic acids, glucosamine, glucose, mannose, rhamnose, fucose, and galactose (Carlson *et al.*, 1987). And LPS structures in a number of species have been demonstrated to have an unusually long 27-hydroxyoctacosanoic acid (Jeyaretnam *et al.*, 2002). Strains of *Rhizobium* also have an unusually complex composition of membrane phospholipids, of which phosphatidylcholine and phospholipids may be replaced by membrane lipids without phosphorus under conditions of phosphorus limitation (López-Lara *et al.*, 2003).

The optimal temperature for growth of *Rhizobium* strains is 25-30^{O}C, with the exception of some species can that grow at about 40^{O}C. All strains of *R. leguminosarum* grow between 20-28^{O}C with an upper tolerance limit of 38^{O}C. The *Rhizobium* can grow in a pH range of 4-10; however, the optimal pH for growth is 6–7. The generation time in culture varies from 1.5 to 5.0 hours. On Yeast-Mannitol agar (YMA) their colonies are usually white or beige, circular, convex, semi-translucent or opaque, raised and mucilaginous; usually growing to a diameter of 2 to 4 mm within 3 to 5 days 28^{O}C. Colonies of nearly all species of the genus *Rhizobium* appear white on YMA containing 0.0025% Congo red (Vincent, 1970). The growth

of a *Rhizobium* strain on carbohydrate-containing media is normally accompanied with the production of abundant amounts of extracellular polysaccharide. Broth cultures of *Rhizobium* strains become turbid after 2 or 3 days in aerated or agitated incubation conditions.

Rhizobia are chemoorganotrophic and can utilize a wide range of carbohydrates (except cellulose and starch) and salts of organic acids as a carbon source without the formation of any gas. An acidic reaction will be observed in growth on a mineral-salts medium containing mannitol or other carbohydrates. Most *Rhizobium* strains can use ammonium salts, nitrate, nitrite, and most amino acids as a source of nitrogen. Many strains require one or more growth factors such as biotin, pantothenate, or nicotinic acid for growth. Strains of some species can use vitamin-free casein hydrolysate as the sole source of carbon and well as nitrogen in a simple mineral salts medium. Peptone, casein, starch, chitin, and agar are poorly utilized or not utilized by *Rhizobium* strains (Kuykendall *et al.*, 2005, Schaechter, 2009).

All known species of *Rhizobium* include strains capable of inducing hypertrophism in plants to form root nodules with or without the ability to fix nitrogen. The rhizobia enter the roots of the leguminous host plant in a number of ways including invagination of root-hair cells or through wounds and elicit the formation of nodules on the root where the bacteria will live as intracellular symbionts to fix nitrogen.

1.2.2.2. Introduction to *Rhizobium* strains used in this study

Numerous new strains of *Rhizobium* are identified and described every year. Hence, an introduction to the strains used in this study is necessary to understand their characters and the methods to describe them.

1.2.2.2.1 *Rhizobium leguminosarum* biovar *viciae* 3841

R. leguminosarum biovar *viciae* (Rlv) strain 3841 is a spontaneous streptomycin-resistant mutant of field isolate *R. leguminosarum* 300. The strain was fully sequenced in 2006 by Young *et al.* (2006). The *Rlv*.3841 has a distinct genomic architecture comprising one circular chromosome and several large plasmids. The plasmid profiles in different strains may vary significantly with respect to size, numbers,

and incompatibility groups. *Rlv*.3841 has six large plasmids of which pRL10 is the symbiosis plasmid (pSym), whereas pRL7 and pRL8 are conjugative plasmids (transferable by conjugation) (Johnston *et al.*, 1982) (Table 1.1). The genome information from sequencing was consistent with earlier electrophoretic and genetic data available for this strain from the studies of Hirsch *et al.* (1980).

Table 1.1. Genome statistics for *Rlv*. 3841 (Young *et al.*, 2006)

Replicon	Base pairs	% G+C	Protein encoding genes	% of coding genes	Mean protein length (aa)	rRNA operons	tRNA genes
Chromosome	5,057,142	61.1	4,736	86.3	309	3	52
pRL12	870,021	61.0	790	90.3	335		
pRL11	684,202	61.0	635	87.5	318		
pRL10	488,135	59.6	461	81.7	304		
pRL9	352,782	61.0	313	88.8	337		
pRL8	147,463	58.7	140	83.4	306		
pRL7	151,546	57.6	188	74.6	224		
Total	7,751,309	60.86	7263	86.4	309	3	52

Note: aa, amino acids; *Rlv*. 3841, *Rhizobium leguminosarum* biovar *viciae* 3841.

Plasmid replication and partitioning system based on *repABC* genes is common in Alphaproteobacteria. RepA and RepB are thought to be a partitioning system essential for plasmid stability; whereas RepC is required for plasmid replication. All six plasmids of *Rlv. 3841* have this system. The number of mutually compatible *repABC* plasmids in *Rlv*. 3841 is the largest among any bacterial strains. However, RepA and RepB proteins from each plasmid are highly diverse. In case of RepC, only RepCs from pRL9 and pRL12 are identical suggesting that recombination may have happened since RepC is not essential for plasmid compatibility (Kim, 2012, Young *et al.*, 2006).

The distribution of functional gene classes on plasmids (besides pRL7 and pRL8 which have more than 80% of the genes that are foreign and/or of unknown function) is similar to the chromosome. Genes that encode proteins for core functions are generally found on the chromosome but some core genes are also found on plasmids. For example, the core genes of the *minCDE* operon involved in septum formation and cell division are found on the plasmid pRL11. Other essential genes including the major heat-shock chaperone genes *cpn10*/*cpn60* (*groES*/*groEL*), *cpn60 and* ribosomal protein S21 gene are located on the plasmids pRL12, pRL9 and pRL10, respectively. However, these essential genes on plasmids have chromosomal paralogs, so the copies may serve specialist functions or may simply be functionally redundant (Young *et al.*, 2006).

Rlv.3841 was the first fully sequenced strain of *R. leguminosarum* and serves as a representative strain of the species *R. leguminosarum.* Within a bacterial species, strains can differ with respect to the presence or absence of some genes, especially on plasmids. In *Rlv.3841*, pRL10-borne *hup* genes for the uptake dehydrogenase system are absent; but these have been studied in depth in other *R. leguminosarum* strains (Leyva *et al.*, 1990, Young *et al.*, 2006). *Rlv.3841* has six plasmids of which the symbiosis plasmid is pRL10 (*repC3* group).

The genome profiles of other natural *R. leguminosarum* strains are different in terms of replication groups and sizes. A previously studied plasmid pRL1, which has a 200kb *repC4* plasmid, shows little difference in *nod* and *nif* gene sequences to those located on pRL10. Since the symbiosis regions of both plasmids are very similar, the differences among them are trivial; it is quite possible that horizontal gene transfer may have occurred between ancestors of these rhizobial strains.

The number of genes and replicons in *Rlv.3841* is greater than in most common bacteria. This is also seen in other *R. leguminosarum* biovar *viciae* strains. Large genomes are widely seen in soil bacteria since their capability to encode multiple gene functions bring the bacteria

long term selective advantage in adapting the bacteria to the soil environment since many different substrates and potential hazards exist (Martinez *et al.*, 1990).

Following is a brief introduction of each replicon.

The total genome size of *Rlv*.3841 is 7,751,309 base pairs, of which 65% is contained in a single circular chromosome and the rest in six circular plasmids. Three identical rRNA operons and 52 tRNA genes are present on the chromosome. The chromosome of *Rlv*. 3841 (5.06 Mb) and total plasmid content of *Rlv*.3841 (2.69 Mb) is large (Young *et al.*, 2006).

Most genes identified in *Rlv*.3841 are chromosomal genes and their orthologs can be found in other fully sequenced alphaproteobacterial genomes. Young *et al.* (2006) built a phylogeny based on 648 identified proteins and found the phylogenies to be consistent with the relationships based on 16S rRNA analysis - the closely related strains of *R. leguminosarum* are *A. tumefaciens*, followed by *S. meliloti*, and then *M. loti*.

pRL12 is the largest plasmid and has nucleotide composition similar to that of the chromosome.

.pRL11 is the most chromosome-like plasmid among all six plasmids. This plasmid carries the cell division genes *minCDE*.

pRL10 can be divided into two parts differing in composition. The first part includes about 200 genes which contains the symbiosis genes. The remaining part of pRL10 resembles the larger plasmids with high-GC3 content, with occasional low-GC3 islands. The nodulation genes *nodOTNMLEFDABCIJ* are located on pRL10. The 13 *nod* genes that are known to be involved in nodulation of the host plant are tightly clustered on pRL10 (pRL100175, pRL100178-pRL100189).

Ding *et al.* (2012) proposed that pRL10, pRL11, & pRL12 possessed a type-III rhizobial conjugation system, and none of them were self-transmissible as type-III plasmids lack the Mpf component. This character is different from the other two types : quorum sensing (QS)-

regulated conjugation system (type-I) and the RctA-repressed conjugation system (type-II). They proved that these two conjugation systems are phylogenetically separate, consistent with their characterized transfer regulation.

pRL9 has low GC3, with two high-GC3s regions. ABC transporter coding genes comprise one quarter of all genes on pRL9.

pRL8 largely comprises accessory genome with low GC3 content. A gene cluster (pRL80079-pRL80088) has been identified on pRL8 that is required for homoserine utilization by strains belonging to *R. leguminosarum* bv. *viciae*. These genes also affect competitiveness during rhizobial colonization. The genes are arranged as two divergently expressed operons induced by L-homoserine or pea root exudate, and are expressed on pea roots. (Vanderlinde *et al.*, 2014).

Based on their work, Ramachandran *et al.* (2011) suggested that pRL8 can be considered as a specific plasmid for pea rhizosphere since a number of genes on pRL8 are specifically upregulated in the pea rhizosphere. Previous study by Ramachandran *et al.* (2011) mutating many of the upregulated genes showed reduction in the competitiveness for pea rhizosphere colonization, while two genes specifically upregulated in the pea rhizosphere, on mutation, reduced colonization of pea but not alfalfa rhizosphere.

pRL7 has an assembly of a number of mobile elements, many of them derived from phages and transposable elements. It has many pseudogenes and a short average gene length.

pRL7 contains three *repABC* operons, which regulate the replication/ partitioning systems. Two of them are complete operons having all the three protein–encoding genes (*repA*, *repB* and *repC*). The third operon lacks the *repB* gene. The operons, as a single transcriptional unit, encode all elements required for replication and stable maintenance of the plasmids in dividing cells. Proteins encoded by *repABC* operons have a distinct phylogeny and homologous recombination events are common within these operons suggesting that gene shuffling and gene recombination play a crucial role in the

generation of new replication systems in Alphaproteobacteria. The *repAC* genes are located in an operon and have the ability to maintain replication in a relatively unstable manner. Retaining the operon depends on the replication and stability provided by the *repABC* operons present on pRL7 (Pérez-Segura *et al.*, 2013).

1.2.2.2.2. *Rhizobium leguminosarum* USDA2370^T

Ramirez-Bahena *et al.* (2008) considered USDA 2370 as the type strain of *R. leguminosarum* biovar *viciae.* They analysed the16S rRNA, *recA* and *atpD* genes and the 16S-23S ITS region. The results showed 100% sequence similarity with ATCC10004^T, LMG 14904^T and USDA 2370^T. Hence, they concluded that all the three strains were identical to each other. They were designated as USDA 2370^T since the original deposit in the USDA culture collection corresponding to strain 3Hoq18T was made in that name. Hence, the strain had the preference for retaining the name *R. leguminosarum* and should be considered as the true type strain of this species. It can effectively nodulate *Pisum sativum*; although strains of *R. leguminosarum* have also been reported to nodulate *Trifolium repens* and *Phaseolus vulgaris*. But information about its ability to nodulate other plants is limited. Mutch and Young (2004) demonstrated that *Rlv* (*R. leguminosarum* biovar *viciae*) might have ability to nodulate all *Viciae* species included in the four genera: *Vicia, Lathyrus, Pisum* and *Lens.*

R. *leguminosarum* USDA2370 is often used as a reference strain to compare with other field-isolated rhizobium strains based on methods like RFLP, AFLP, BOX-PCR and specific gene phylogenies. Tian *et al.* (2007) studied 75 *Vicia faba* isolates from different ecotypes of *Vicia faba* in different regions of China. They too selected *R*. *leguminosarum* USDA2370 as one of the reference strains. In AFLP and BOX-PCR studies, *R*. *leguminosarum* USDA2370 was reported as one of the ungrouped strains compared to other strain clusters. In DNA-DNA hybridization, *R*. *leguminosarum* USDA2370 showed only 14% DNA relatedness with a representative *Vicia faba* isolate - CCBAU33202, which was later annotated as type strain of a new species - *Rhizobium fabae*. They concluded from the results of RFLP

of *nodD* and *nodC* genes (earlier studies often used them to investigate the diversity of rhizobia population) that, among all strains *R*. *leguminosarum* USDA2370 is different from other reference strains and *Vicia faba* isolates. The findings of Tian *et al.* (2007) are similar to those of Mutch and Young (2004) who observed that isolates from *Vicia faba* are unique compared to isolates from other *Vicia* plants and *R*. *leguminosarum* USDA2370 can be treated as representative strain. Alvarez-Martinez *et al.* (2009) compared the core and accessory genes of rhizobial strains isolated from Northwest Spain. In their study, USDA 2370 was most similar to wild *Vicia sativa* nodule isolates with the same ARDRA pattern and identical *rrs* sequence, high similarity of 16S-23S ITS sequence, *recA* and *atpD* sequence analysis and *nodC* analysis.

1.2.2.2.3. *Rhizobium pisi* DSM30132^{T}

R.pisi DSM30132 was reclassified as the type strain of a novel species - *Rhizobium pisi* by Ramirez-Bahena *et al.* (2008). It was derived from a strain that had been distributed by ATCC as the type strain of Rhizobium *leguminosarum*, with the strain name ATCC10004, but which was different from the true type strain (LMG 14904^{T} and USDA 2370^{T}), so reclassification was required. They compared 16S-23S intergenic spacer region, *rrs, recA, atpD* gene sequences, conducted DNA-DNA hybridization and phenotypic characterization of DSM30132, USDA 2370, *R. phaseoli* ATCC14482^{T} and *R. etli* CFN42^{T}. The results of DNA-DNA hybridization showed a low similarity value between DSM30132 and the true type strain. Phenotypic tests including bacterial growth in different temperature and pH range, different carbon and nitrogen sources, the production of exoenzymes (on PNP-substrates), and resistance to different antibiotics were also conducted using these strains. Based on both genotypic and phenotypic results, they concluded that the strain was different from other type strains and should be named as the type strain of a novel species - *Rhizobium pisi* DSM30132^{T}.

Rhizobium pisi DSM30132^{T} was described as being isolated from nodules of *Pisum sativum*.

1.2.2.2.4. *Rhizobium fabae* CCBAU33202^T

R. fabae CCBAU 33202^T was characterized by Tian *et al.* (2008) from *Vicia faba* in China along with other 5 strains. The most closely related species to it are *Rhizobium etli* and *Rhizobium leguminosarum*.

Tian *et al.* (2007) found that CCBAU 33202^T and five other *Vicia faba* isolates were grouped in the same ARDRA type and 16S rRNA gene sequence showed that it belonged to *Rhizobium*. CCBAU 33202^T showed more than 98% relatedness with other *Rhizobium* type strains : *Rhizobium etli* CFN42^T (99.5% similarity), *R. leguminosarum* bv. *trifolii* T24 (99.3%), *R. leguminosarum* bv. *viciae* USDA 2370^T (99.1%), *R. leguminosarum* bv. *phaseoli* USDA 2671(99.1%), *Rhizobium rhizogenes* IAM 13570^T (98.1%), *Rhizobium lusitanum* P1-7^T (98.0%), *Rhizobium tropici* type B CIAT 899^T (97.5%) and *R. tropici* type A LMG 9517 (96.8%). Later, they analysed the housekeeping genes (*recA*, *atpD*) and performed DNA-DNA hybridization using six *Vicia faba* isolates and type strains of other species.

Housekeeping gene phylogenies showed that more than 97.9% *recA* and 99.8% *atpD* similarity among six *Vicia faba* isolates, but the similarity between CCBAU 33202^T and type strains of other species was less than 95%. DNA-DNA hybridization results showed more than 70% similarity among six *Vicia faba* isolates but less than 43% value between CCBAU 33202^T and strains from other species. In addition, 16S-IGS RFLP profiles and whole-cell protein SDS-PAGE also grouped these six *Vicia faba* isolates and separated them from other reference strains. In phenotypic tests, six *Vicia faba* isolates were different from *R. etli* CFN42^T and *R. leguminosarum* USDA2370^T in utilization of several carbon sources, reduction of dye, litmus acid coagulation and antibiotic resistance. All the assays above suggested these isolates may not belong to these recognized species but belong to a single new species. Tian *et al.* (2007) proposed that these *Vicia faba* isolates belong to a novel species, *Rhizobium fabae*, and CCBAU33202^T was designated as the type strain.

Analysis of nodulation genes tells us the host range of rhizobial strains. *nodD* sequence analysis of the strains showed that these strains

grouped with *R. leguminosarum* biovar *viciae* strains which can nodulate a variety of *Vicia*, *Lathryus* and *Pisum* plants. Nodulation tests showed that *R. fabae* CCBAU 33202^T could form effective pink nodules on *Pisum sativum* and *Lathyrus* species but did not nodulate *Phaseolus vulgaris*, *Trifolium*, *Glycine max* or *Medicago sativa*. In all, *R. fabae* CCBAU 33202^T shared symbiotic characteristics with *R. leguminosarum* biovar *viciae* strains.

R. fabae CCBAU33202^T and its host *Vicia faba* form an interesting symbiotic pair. In the work published by Mutch and Young (2004), *Vicia faba* was the most discriminating host with respect to isolates from other plants, including other species of *Vicia*. But, its isolates could effectively nodulate other plants. Their studies on *nodDF* genes suggested that nearly all *Vicia faba* isolates belong to the same *nodDF* gene type (type 2 in the paper) which also included the pRL1 plasmid that contains the same symbiosis gene sequences as pRL10 in *Rlv.3841*. Tian *et al.* (2010) have found new strains from *V. faba* which also form a new single group.

1.2.2.2.5. *Rhizobium phaseoli* ATCC14482^T

Rhizobium phaseoli ATCC 14482^T biovar *phaseoli*, described by Ramirez-Bahena *et al.* (2008), was found to have less than 95% similarity in analysis of *recA* and *atpD* when compared to *R. etli* CFN42^T, *R. leguminosarum* USDA2370^T and *R. pisi* DSM30132^T. DNA-DNA hybridization between *Rhizobium phaseoli* ATCC 14482^T and three type strains showed 52%, 51% and 44% relatedness, respectively. Phenotypic tests also supported that *Rhizobium phaseoli* should be retained as a separate genus and the type strain being *R. phaseoli* ATCC 14482^T. This strain was isolated from effective nodules of *Phaseolus vulgaris* and is also reported to form nodules on *Trifolium repens*, but not on *Pisum sativum*.

R. etli CFN42^T is the closest type strain to *Rhizobium phaseoli* ATCC 14482^T. Mulas *et al.* (2011) studied *nodC* genes among rhizobia biovar *phaseoli* strains. *nodC* is one of the nodulation genes which determines the host ranges of rhizobia. Thus, it is highly similar within the same biovar. They found *Rhizobium phaseoli* ATCC 14482^T and *R.*

etli CFN42^T shared the same α allele and a different γ allele which is present in most strains isolated from North Spain. The γ allele was later demonstrated to be the dominant allele in that area. Thus, strains from these two species may be easily mixed up. López-Guerrero *et al.* (2012) studied three *R. etli* strains - CIAT 652, Ch24-10 and CNPAF 512. DNA-DNA Hybridization was carried out for all three strains against *R. etli* CFN 42^T. The results showed that strains CIAT 652, Ch24-10 and CNPAF 512 are at the borderline that distinguishes the different species - from 68% to 75%. However, the results against *Rhizobium phaseoli* ATCC 14482^T showed that the strains were more similar as the values ranged from 84% to 88%. In addition, phylogenies of housekeeping genes including *recA, atpD* and *rpoB* grouped these three strains to *Rhizobium phaseoli* ATCC 14482^T. Genome analysis and ANIm (average nucleotide identity) also show that these three strains form a group with *Rhizobium phaseoli* ATCC 14482^T and are separate from *R. etli* CFN42^T (López-Guerrero *et al.*, 2012).

1.2.2.2.6. *Rhizobium tropici* CIAT899^T

Strain *Rhizobium tropici* CIAT 899^T was described by Martínez-Romero *et al.* (1991) as the type strain of *Rhizobium tropici* . Its reported hosts are *Phaseolus vulgaris, Leucaena esculenta* and *L. leucocephala.*

(Michiels *et al.*, 1998) studied the interaction of different strains with *P. vulgaris* and found isolates from *P. vulgaris* that diverged from other species. Their results showed that *P. vulgaris* could receive signals for nodulation from many rhizobia, though most of the interactions are not effective. Previous study found that some of *P. vulgaris* isolates were identified to be different from *Rhizobium leguminosarum*. Although they had been classified as *Rhizobium leguminosarum* biovar *phaseoli*, in genetic analysis they always formed a single cluster. Therefore *Rhizobium tropici* was proposed to be a species to include the group containing these strains. Strains belonging to this species normally have a wider host range, including *Leucaena* spp. They carry a single copy of *nif* gene, and exhibit low DNA-DNA hybridization with other

species of *Rhizobium*. Also, the genetic distances calculated by multilocus enzyme electrophoresis and by 16s rRNA sequence comparison also support the separation of these strains as a single species.

Rhizobium tropici was previously named *Rhizobium leguminosarum* biovar *phaseoli* (type II strain). The symbiosis plasmids of type II strains promote an effective and completely differentiated symbiotic process in *A. tumefaciens* recipients. These strains have some special characteristics like tolerance to heat or metal ions. The symbiosis plasmids are not lost or cured at 37^{O}C and some of them confer aluminium resistance. *Rhizobium tropici* CIAT 899^{T}, a Type II strain, can grow on LB, on which most *R. leguminosarum* strains show no growth. The strain is also resistant to low pH, heavy metals and to the antibiotics chloramphenicol, spectinomycin, carbenicillin, and streptomycin.

Rhizobium tropici has been recently studied in depth. Since phenotypic and genotypic diversity has been reported in this species, subgroups type A and B were defined within it. Ribeiro *et al.* (2012) reclassified *Rhizobium tropici* type A subgroup into a new species *Rhizobium leucaenae*. Phenotypic assays including carbohydrate utilization, nitrogen sources utilization, and growth conditions (e.g. temperature, pH, medium types and ion tolerance) as well as genotype assays including DNA-DNA hybridization, 16S rRNA and concatenated protein-coding gene (recA+gltA+rpoA+glnII+rpoB) phylogenies separated *Rhizobium tropici* and other species and supported the reclassification.

The adaptation of *Rhizobium tropici* CIAT 899^{T} to different growth conditions attracts studies on its metabolism and makes it a good model to study bacterial response to environmental stress. Riccillo *et al.* (2000) created a library of *Rhizobium tropici* CIAT 899^{T} mutants using Tn5-*luxAB* insertion mutagenesis and found that one of the mutants - CIAT899-13T2 was extremely sensitivity to acid conditions and stopped growing, the gene inactivated by insertion of Tn5-*luxAB* was related to production of glutathione, and the deficiency of

glutathione was confirmed in the mutant by comparing with *Rhizobium tropici* CIAT 899^{T}. CIAT899-13T2 had been shown to be sensitive to weak organic acids, osmotic and oxidative stresses, and the presence of methylglyoxal. Competition assays also demonstrated that this mutant showed less competition ability compared to its parent strain. External addition of glutathione was reported to restore the response of the mutants to near wild type levels. Thus, they concluded glutathione production was important for the adaptation of *Rhizobium tropici* CIAT 899^{T} to extreme environment**s**. Ormeno-Orrillo *et al.* (2012) found there was a highly-conserved symbiosis plasmid in *Rhizobium tropici* CIAT 899^{T}, *Rhizobium* sp. PRF 81 and *R. leucaenae* CFN 299. The host range of these three strains was also similar and they could all form effective nodules and carry out nitrogen fixation on *Phaseolus vulgaris*. This observation revealed the common evolution taking place in rhizobia - in order to have a successful symbiosis with certain host plants, rhizobia may obtain a set of nodulation and nitrogen fixation genes which might have been possessed and spread by rhizobia even when present in different geographic environment. This symbiosis plasmid obtained by these three strains was reported to carry genes for biosynthesis and modulation of plant-hormone levels. In addition, all three of these strains were observed to have three divergent *nodA* genes, which possibly allow them to have a wider host range compared to strains having a single *nodA* gene.

The outstanding abilities of CIAT899^{T} and PRF81 in both competition with other rhizobia and in overcoming environment stress have been studied. It is speculated that these abilities were due to the presence of functionally different genes : *e.g.* a lipid A phosphoethanolamine transferase, a KcsA-like ion channel and genes encoding drug-efflux pumps (key for resistance to some antimicrobial compounds). This study gives a good example of diverse genes in one symbiosis plasmid which are important in rhizobial host range and competitiveness. A study on non-symbiosis plasmid functions was performed by Barreto *et al.* (2012). They generated a collection of CIAT899 cured derivatives in which the CIAT899 were lacking at least one plasmid. A nodule occupation test was carried out on *Phaseolus*

vulgaris and the results showed one of the cryptic plasmid (annotated as plasmid ‘a’ in the paper) was involved in nitrogen fixation, and genes distributed in different plasmids were found likely to effect rhizobial competitiveness. For the study of plasmids in rhizobia, complete genome information is very important for further identification of resident genes functions.

Since *Rhizobium tropici* CIAT 899^{T} has outstanding abilities in terms of growth and nodulation, it was chosen to be tested for promotion of *Phaseolus vulgaris* in low fertility soils by (Hungria *et al.*, 2003). They mixed a number of inoculants (included CIAT899) and N-fertilizer to evaluate the symbiotic effectiveness, assess the effects of reinoculation on rhizobial soil population and on common bean yield and verify the effects of supplying N-fertilizer. Their results showed that there was an obvious increase in the yield of crops under inoculation. This demonstrated that low levels of N fertilizer and inoculation with superior strains can increase yield of this important protein source in Latin America at a low cost.

1.2.2.2.7. *Rhizobium etli* CFN42^{T}

Strain *Rhizobium etli* CFN42^{T} was described by Segovia *et al.* (1993) as the type strain of *Rhizobium etli*. It has only one reported host - *Phaseolus vulgaris.*

After the reclassification of *R. leguminosarum* biovar *phaseoli* type II strains as *R. tropici*, type I strains were also considered to be a new species as they formed a cluster distinct from *R. leguminosarum* biovar *viciae* and biovar *trifolii* strains in a study by Segovia *et al.* (1991). Since then they found lower DNA-DNA hybridization values in type I strains than other two biovars. Similarly, Eardly *et al.* (1992) found that one of the type I strains (Olivia 4) showed distinct sequences mismatches when compared to *R. leguminosarum* USDA2370^{T} in a part of 16S rRNA nucleotides. Hence, Segovia *et al.* (1993) collected numbers of type I strains from America and *R. leguminosarum* biovar *viciae* strains from Mexico and analysed the same region (position 20 to 338) of 16S rRNA. The phylogenetic result from UPGMA method supported their suspicion and they

annotated the type I strains as *R. etli* and *Rhizobium etli* CFN42^T as the type strain.

R. etli strains have been studied for many years; especially CFN42^T, which has been fully sequenced. The genome of CFN42^T contains one chromosome and six plasmids ranging in size from 184kb to 600kb (Brom *et al.*, 1992). One of the six plasmids, p42d, was studied thoroughly as a symbiosis plasmid. The results of the study led to the whole genome sequence of CFN42^T (González *et al.*, 2003). p42d is a circular molecule of 371,255 bp. It has a *repABC* type replication system and has an average GC content of 58.1% (Ramírez-Romero *et al.*, 1997). CFN42^T a good model to study rhizobial metabolism and rhizobial genome (Encarnacion *et al.*, 1995, Romero *et al.*, 1991). Laguerre *et al.* (1994) first developed the PCR-RFLP method for rapid identification of rhizobia, using *R. etli* strains and other species to evaluate the reliability of this method. Results showed the divergent distance between two *R. etli* strains was 0.4% and the results were identical to the 16S rRNA sequence analysis results. This method was used in a later study by Laguerre *et al.* (2001) who studied *nodC* and *nifH* genes in 83 representative rhizobia strains of 23 species. RFLP results showed *R. etli*, with other *Phaseolus vulgaris* isolates, were well grouped with each other regardless of their taxonomic status. Phylogenetic trees based on sequences of symbiosis genes were made with representative strains, and in these trees the isolates from *Phaseolus vulgaris* were close to each other irrespective of their 16S rDNA based classification. Generally, similar *nodC* and *nifH* genes were observed in these strains but exceptions also occurred. These results supported the hypothesis of lateral gene transfers taking place across rhizobial species.

To better understand the symbiosis genes within *R. etli* strains, González *et al.* (2003) studied CFN42^T symbiosis plasmid p42d and chromosomal symbiosis islands for their content, organization and differences. After analysing 359 coding sequences, they found that the nodulation and nitrogen fixation genes were clustered in a 125kb region. A lot of genes related to mobile elements were separately located on the plasmid. They concluded that symbiotic compartments

of rhizobial genomes are mosaic structures that have been frequently tailored by recombination, horizontal transfer and transposition. Based on the studies by González *et al.* (2003), Flores *et al.* (2005) analysed the distribution of single-nucleotide polymorphisms in homologous regions from different *R. etli* strains. Their data indicated that recombination is the main reason that led to the majority of nucleotide substitutions spreading in the population, rather than new mutations. They suggested that the structure of the symbiosis compartment of *R. etli* strains was formed as a result of a mosaic type of evolutionary history. The studies mentioned above suggest that the symbiosis plasmid in *R. etli* carries a number of genes involved in nodulation and gene transfer; often within and sometimes, outside the species.

The plasmids other than the symbiosis plasmid (p42d) in CFN42^T are also interesting. Brom *et al.* (2000) cured a number of plasmids in *Rhizobium etli* CFN42^T. They found that the strains cured of multiple plasmids were approximately 1000 times less competitive than strains cured for single plasmid or the parental strain, suggesting that the genes responsible for competitiveness were distributed between the different plasmids on genome of CFN42^T. González *et al.* (2006) found that the genome of CFN42^T had multiple replicons unlike many of the fully sequenced nitrogen fixing bacteria. Genes on the chromosome are mostly related to functions like cell growth, whereas few essential genes or complete metabolic pathways were found on plasmids. These plasmids are absent in species closely related *R. etli* and suggest that they had an independent evolution or that plasmids in rhizobia are more diverse than the chromosome, especially p42b, p42c, p42e and p42f. Genes on different plasmids were found to work together, which was identical to the results obtained by Brom *et al.* (2000). Landeta *et al.* (2011) focused on the chromosome-like plasmid p42e, which was the only stable plasmid among all the six plasmids. The other five could be removed without affecting growth. The study of the p42e partition system and gene function showed that an ancestor of p42e was probably a typical *repABC* plasmid, and some essential genes were transferred from the chromosome without duplication as a consequence of recombination. This suggested that p42e and its homologous plasmids in other species could be considered as

secondary chromosomes. Harrison *et al.* (2010) made similar observations and suggested the term ‘chromid’ for secondary chromosome.

1.2.2.2.8. *Rhizobium leguminosarum* biovar *trifolii* strains WSM1325 and WSM2304

Rhizobium leguminosarum biovar *trifolii* strain WSM1325 was fully sequenced and reported by Reeve *et al.* (2010a). *Rhizobium leguminosarum* biovar *trifolii* strain WSM2304 was also fully sequenced and described by the same group (Reeve *et al.*, 2010b). Both the strains belong to *R. leguminosarum* which is one of the most widely occurring root-nodule bacteria species. The common host of *Rhizobium* biovar *trifolii* i.e. clover (*Trifolium*) is found on pastures. It includes both annual species (e.g. *T. subterraneum*) and perennial species (e.g. *T. pratense, T. repens* and *T. polymorphum*). Clovers are valuable legume plants in natural and agricultural areas of the world since they have adapted to grow under different environmental conditions (Zohary and Heller, 1984).

R. leguminosarum bv. *trifolii* strain WSM1325 was isolated from an annual clover plant growing near Livadi beach on the Greek Cyclades island of Serifos in 1993. *R. leguminosarum* bv *trifolii* strain WSM2304 was isolated from a nodule recovered from the roots of the perennial clover *Trifolium polymorphum* growing at Glencoe Research Station near Tacuarembó, Uruguay in December 1998 (Howieson *et al.*, 2005). WSM1325 is attractive because it is highly effective in nodule formation and nitrogen fixation on a broad range of annual clovers from the Mediterranean region, whereas WSM2304, a highly effective microsymbiont of a perennial clover of South American origin is of interest since it has a narrow, specialized host range (Howieson *et al.*, 2005). Both strains have a competitive edge for nodulation in acid and infertile soils of Uruguay on their preferred host plants (Yates *et al.*, 2005).

WSM1325 is not only a highly effective inoculant strain for annual *Trifolium* spp., but can also symbiose with key perennial clovers of Mediterranean origin used in farming, such as *T. repens* and *T.*

fragiferum, making it one of the most important inoculants for clover agriculture and production in Australia and Brazil. WSM 1325 cannot form symbiosis with American and African clovers or forms ineffective symbiosis. WSM2304 can fix nitrogen effectively with some perennial American clovers, but is ineffective with the Mediterranean clovers. Both strains are effective microsymbionts under competitive conditions for nodulation since their involvement in nodulation appears to be a host-mediated selection process. Under conditions of competitive nodulation, WSM1325 can preferentially nodulate *T. purpureum* even when outnumbered 100:1 by WSM2304. However, when competition for nodulation is done on *T. polymorphum*, WSM2304 is selected preferentially to form symbiosis even when outnumbered 100:1 by WSM1325 (Yates *et al.*, 2008).

WSM 1325 has 6 replicons, one circular chromosome and five circular plasmids, whereas WSM 2304 has 5 replicons, one circular chromosome and four circular plasmids. 16S rRNA sequence analysis showed that both the strains have identical 16S rRNA sequences which differs by 1bp from the 16S sequence of *R. leguminosarum* bv. *viciae* 3841 (Reeve *et al.*, 2010a, Reeve *et al.*, 2010b).

Rogel *et al.* (2011) compared the gene sequences from the genomes of *R. leguminosarum* bv. *viciae* strain 3841 and *R. leguminosarum* bv. *trifolii* strain WSM1325. The results suggested that the two strains probably belonged to a single species but their different host specificities (*Trifolium* versus *Pisum*) strongly support the concept of different biovars. WSM1325 and WSM2304 have different host specificity since their preferred host plants are different in taxonomy and geography. In principle, they should not belong to the same biovar since only WSM 1325 corresponds to bv. *trifolii*. WSM 2304 and probably deserves to be classified as a distinct biovar. The *nodD* gene sequences of the two strains are divergent with only 79.1% similarity and *nodR* was found only in WSM1325 and not in WSM2304. Schlaman *et al.* (2006) suggests that *nodR* is essential for the existence of fatty acyl moieties in the Nod factors synthesized by *R. leguminosarum* bv. *trifolii*. The difference between these two strains in terms of host range and effectiveness might be a consequence of this

nodR gene which results in the production of different Nod factors. (Yates, 2008).

1.2.3. Rhizobial genome

The genomes of rhizobia encode a large number of proteins that participate in the growth of rhizobia and in the nodulation process. Like most bacteria, the rhizobial genome can be divided into two compartments viz. the chromosomal genome – responsible for growth and reproduction – and the extrachromosomal genome (e.g. plasmid and chromid) which is associated with adapting to external environment (Harrison *et al.*, 2010, Martinez *et al.*, 1990, Mercado-Blanco and Toro, 1996). Most essential genes are present on the chromosomal DNA. Their functions include cell survival and growth. Plasmids and chromids constitute a large part of the rhizobial genome. They play an important role in rhizobial interaction with plants and nodulation competition, especially in the two fast-growing genera - *Rhizobium* and *Sinorhizobium*, but not in *Azorhizobium*, *Bradyrhizobium* and *Mesorhizobium*. In these three species, the genes involved in symbiosis are found on the chromosome. (Hanin et al., 1999, Young, 1996).

Conceptually, the rhizobial genome can also be divided into core genome and accessory genome (Young *et al.*, 2006). The core genome includes the housekeeping genes. These genes are involved in metabolism, information exchange for homeostasis etc. (Feil, 2004). These genes normally have a higher %G+C content and are consistent in codon usage (Hacker and Carniel, 2001, Young *et al.*, 2006). In addition, core genes can be shared within related organisms, mostly within the species. The sharing of core genes can be analysed at different taxonomic levels to identify the core genes at that level of rank in the taxonomic hierarchy (Feil, 2004).

Since genes from the core genome are housekeeping genes, they are highly conserved and are widely used in rhizobial taxonomy in the description of new species. Examples of genes from the core genome used in taxonomy include genes like *recA* and *atpD*. RecA is a 38 kilodalton protein essential for the repair and maintenance of DNA, as concluded by Cox (1991) after analysing the molecular properties of RecA. Cox continued studying RecA and in 1993 demonstrated that it was the central protein mediating a DNA recombination-repair system, later known as homologous recombination (Cox,

1993). Large-scale phylogenies of *recA* in the bacteria are identical to the corresponding small-subunit rRNA (16S, SSU) phylogenies (Eisen, 1995, Young, 1998).

The *atpD* gene encodes one of the subunits of ATP synthase, which is essential for energy production. The enzyme is divided into the two subunits F0 and F1. F1 is composed of the five polypeptides α, β, γ, δ and ε. Subunit β, which is responsible for nucleotide binding and catalysis, is regulated by *atpD* (Walker *et al.*, 1984). *atpD* is also an excellent candidate for phylogenetic analysis since it has a ubiquitous distribution, is functionally consistent and shows sequence conservation in bacteria. Amann *et al.* (1988) and Ludwig *et al.* (1993) demonstrated that the results of phylogenetic relationships between bacteria based on *atpD* gene sequence also corresponded to results based on 16S rRNA sequences.

The use of *recA* and *atpD* in rhizobial classification was first described by Gaunt *et al.* (2001). SSU rRNA was commonly used for rhizobia phylogeny study, but single-gene phylogenies did not properly reflect rhizobial genome evolution, so more loci were required for the analysis. Partial *recA* and *atpD* sequences from 25 type strains of *Alphaproteobacteria* were used for phylogenetic study. Results showed that data from these two genes supported the classification of the species similar to that obtained using the sequence of SSU rRNA. They also pointed out that only housekeeping genes could possibly show phylogenetic agreement since their evolution occurred within the species, whereas the symbiosis genes had a different evolution history, evolving as a result of lateral gene transfer from other species.

Contrary to genes in the core genomes, genes from the accessory genome can be diverse even within same species (Young *et al.*, 2006). They can be removed without affecting the survival of the bacteria under normal conditions (Feil, 2004). However, the removal of accessory genes may impact their adaptation to natural environment since they confer selective advantage to the bacteria. Accessory genes have less G+C% as compared to core genes and have distinct codon usage (Hacker and Carniel, 2001). In *Rhizobium* and *Sinorhizobium*, important accessory genes like nodulation genes (*nod* genes) and nitrogen fixation genes (*nif* and *fix* genes) are normally located on large plasmids, or megaplasmids depending on their sizes (e.g. pRL10 in *Rlv*. 3841

and p42d in CFN42^T). These plasmids are called symbiosis plasmids (pSym) and are unique in different strains. The symbiosis plasmids carry many unknown genes that do not have an obvious role in symbiosis. In some species like *Bradyrhizobium* and *Mesorhizobium*, the symbiosis genes are present on symbiosis islands on chromosomes (Sullivan and Ronson, 1998, Vinuesa *et al.*, 1998). As mentioned before, lateral gene transfer of some accessory genes is common within different strains. This phenomenon results in the expression and inheritance of plant-interactive properties of the plasmid-donor species, and new genetic material can benefit the recipient strain for environmental adaptation or bacterial competition.

nod and *nif* genes are two classes of representative accessory genes of rhizobia. Their functions and involvement in symbiosis and nitrogen fixation will be discussed later. Studies on the *nod* and *nif* genes can help us understand the coevolution of rhizobia with the environment. Haukka *et al.* (1998) collected 52 isolates of rhizobia from *Acacia senegal*, *Prosopis chilensis*, and related leguminous trees growing in Africa and Latin America. Previous 16S rRNA gene phylogenies showed the strains belonged to *Sinorhizobium* and *Mesorhizobium*. The restriction patterns of the *nodA* and *nifH* genes were analysed and was found to reflect their host specificity.

Housekeeping genes and symbiotic genes have raised interest of scientists to study the relationships among rhizobia strains. Perret and Broughton (1998) developed the Targeted PCR Fingerprinting (TPF) technology to discriminate between closely related strains of rhizobia. They amplified *nifH* and *recA* genes from *Rhizobium* NGR234 and *R. fredii* USDA257. Differences were observed in these two closely related strains which were 98% identical at their symbiosis loci. They also suggested that chromosomes and symbiosis plasmids evolved differently, and symbiosis genes were acquired by lateral gene transfer. Results obtained by Wernegreen and Riley (1999) also supported this finding. They compared symbiosis genes *nodB* and *nodC* with housekeeping locus glutamine synthetase II (*GSII*), and a portion of the 16S rRNA gene from rhizobial isolates from different host plants. Results showed each of the loci could separate these strains into their own genera, but the phylogenetic congruence indicated that chromosomal genes were less prone to lateral gene transfer than symbiosis genes. Wu *et al.* (2011) collected 99 rhizobial isolates from nodules of wild soybean in different regions of China

and investigated their diversity. They performed phylogenetic analysis of both core genes (*recA* and *atpD*) and symbiosis gene (*nodC*). Results from core genes studies showed 72 of them belonged to four *Bradyrhizobium* species and one *Sinorhizobium* species. *nodC* phylogeny showed that this gene was vertically transferred and reflected the original geographic conditions of each isolate. A similar study was carried out by Rahi *et al.* (2012). They collected 120 isolates from the nodules of *Pisum sativum* cultivated at 22 different locations of the trans-Himalayan valleys of Lahaul and Spiti in the state of Himachal Pradesh of India. Their results showed that the clustering of isolates reflected their origins from different valleys in data analysed from core and accessory genes; as well as from utilization of carbon substrates and whole-cell fatty acid methyl ester (FAME) production. Rashid *et al.* (2012) confirmed the 16S rRNA analysis of a group of lentil isolates to be *R. leguminosarum* and *R. etli* by analysing phylogenies of housekeeping genes. Analysis of nodulation genes and nodulation tests both supported the relatedness of the strains to *R. leguminosarum* biovar *viciae*.

1.2.4. Rhizobial host plants

The host plants which allow rhizobia to form symbiosis and fix nitrogen belong to the family *Fabaceae* (*Leguminosae*). It is the third largest family of flowering plants with more than 18,000 described species. The family includes herbs, shrubs, trees and vines distributed throughout the world in different geographic conditions, environments and climates, including tropical rain forests. The fruit of a tree from the family *Leguminosae* is technically called a legume or pod. The family is divided into three subfamilies: *Papilionoideae*, *Mimosoideae* and *Caesalpinioideae*. 97%, 90% and 29% respectively of studied species from each of these three subfamilies have been observed to form nodules (Bryan *et al.*, 1997). Some members of the family are among the most valuable crops in different parts of the world. For example : peanuts in Sichuan, China (Zhang *et al.*, 1999), *Pisum sativum* and *Vicia faba* in salt stressed areas (Del Pilar Cordovilla *et al.*, 1999) and cowpea (a known broad host-range leguminous plant) in north-east region of Brazil (Martins *et al.*, 1997).

Some non-legume plants can also form symbiosis with rhizobia. Trinick (1979) examined the non-legume plant *Parasponia andersonii* using light and

electron microscopy. They found nodules on the plant which had a central vascular bundle surrounded by an endophyte-infected zone. However, unlike the nodules in legumes, the rhizobia in *Parasponia* were not released from the infection thread. Op den Camp *et al.* (2012) suggested that the symbiosis between rhizobia and *Parasponia* had evolved independently from legumes. Santi *et al.* (2013) reviewed a number of non-rhizobia and legume symbiosis including actinorhizal plants with *Frankia* and cyanobacterial association with *Poaceae* family, e.g. maize, rice and wheat symbiosis with bacteria. They, along with Charpentier and Oldroyd (2010) and Beatty and Good (2011) indicated that since we have the basic knowledge of symbiosis, genetic basis of nitrogen fixation and possess the methods to transform model legumes for investigation, further studies could focus on artificially generating symbiosis between nitrogen-fixing bacteria and non-legume plants. The creation of such associations could benefit agriculture in many ways, for example, by promoting plant growth and reducing the use of chemical fertilizers.

Rhizobia that belong to the same species but have a different host range are said to belong to different biovars. For example, *R. leguminosarum*, has three biovars - biovar *viciae*, *trifolii* and *phaseoli*. All these strains belong to *R. leguminosarum* because they are share a common core or chromosomal genome. The host range genes are located on plasmids. Host range has been abandoned as the standard to define species for two reasons. First, host range in different biovars may overlap, for example, *R. leguminosarum* USDA2370^{T} and *R. pisi* DSM30132^{T} were reported to form nodules on *Trifolium repens* and *Phaseolus* although they have been assigned to be biovar *viciae*. Second, host range of a strain could be changed by replacing plasmids containing genes for an altered host range. Since the rhizobial plasmids are transferable, rhizobial host ranges are variable.

1.2.5. *Rhizobium*-legume symbiosis

1.2.5.1. Rhizobial infection

The symbiosis process between rhizobia and legume takes place in a number of steps. During the process, a series of communication mediated by chemical signals occurs resulting in the formation of nodules (van Rhijn and Vanderleyden, 1995, Wang *et al.*, 2012). Preinfection events which include rhizobial attraction in the rhizosphere initiate the process.

Rhizobia are attracted by molecules contained in the exudates of plants. For example, strains of *Rhizobium* and *Bradyrhizobium* will attach to legume root by detecting amino acids and dicarboxylic acids. However, the flavonoids secreted by plants act as the main chemical signal to attract rhizobia since it can be absorbed passively across rhizobial membrane (Gnanamanickam, 2007).

Flavonoids received by rhizobia activate the genes coding for the synthesis of Nod factors. Nod factors belong to the family of lipochitooligosaccharides and are essential for the development of symbiosis in most legumes (Oldroyd *et al.*, 2011). Long (1996) indicated that the transcription of *nod* genes is regulated by NodD. These proteins are members of LysR family of transcription regulators and specially activated by flavonoids. They bind to *nod* boxes, conserved DNA motifs located within the promoter regions of *nod* genes (Honma *et al.*, 1990). NodD proteins are flavonoid-specific. Peck *et al.* (2006) found that since homologs of *nod* boxes were found in different rhizobia, some NodD proteins from different rhizobia can bind the same nod gene promoter. In addition, they indicated that although some flavonoids were not essential to activate NodD expression, they could increase the protein binding to *nod* boxes and mediate conformational changes in *nod* gene promoters.

Nod factors, are essential for successful nodulation in most legume-rhizobium symbiosis. The only exception was reported by Giraud *et al.* (2007). They reported two fully sequenced photosynthetic strains - *Bradyrhizobium* BTAi1 and ORS278, in which the typical *nod* genes and Nod factors were not required to establish symbiosis with some legumes. A purine derivative was described by them to play a key role in nodule formation by these two strains.

Structures of different Nod factors share a similar chitin-like *N*-acetyl glucosamine oligosaccharide backbone with a fatty acyl chain at the non-reducing end. However, the variable length of backbone, different size and saturation of fatty acyl chain in addition with other decorations at both ends of the molecules play a decisive role in the specific communication between rhizobia and their host plants (Lerouge *et al.*, 1990). Crucial structures of Nod factors are synthesized by products of the *nodABC*

genes. In general, nodulation genes, especially *nodD* and *nodABC* genes are the determinants of the host range in most rhizobia (Roche *et al.*, 1996, Schlaman *et al.*, 1992). Thus loss or mutations of these genes entirely prevents the establishment of symbiosis. Downie and Surin (1990) found that one *R. leguminosarum* biovar *viciae* mutant lacking *nodFEL*, *nodMNT* and *nodO* region could not nodulate on *Pisum sativum* and *Vicia hirsuta*, but the complementary strains created using either partial *nodFE* or *nodO* genes were found to possess some nodulation ability. Their results suggested either pathway could independently activate nodulation.

When plants perceive Nod factors, a series of plant responses occurs including root hair branching, deformation, curling and cell wall hydrolysis (Oldroyd *et al.*, 2011). Rhizobia then enter into the curling root hairs and cause invagination of plasma membrane of the cells of the root hair of the plant. The plant deposits new cell wall material to synthesize and form an infection thread which cover rhizobia, enabling them to enter the root hairs (Rae *et al.*, 1992).

Two types of nodules may be developed in legumes : determinate and indeterminate nodules. Determinate nodules do not contain a persistent meristem unlike indeterminate nodules (Cohn *et al.*, 1998). During the formation of the infection thread, hypodermis cells of determinate nodules and root cortical cells in indeterminate nodules start to form a nodule primordium (Hadri *et al.*, 1998). When rhizobia reach the primordium, morphogenesis occurs and rod-shaped cells of rhizobia change into pleomorphic endosymbiotic form, called bacteroids (Sprent and Sprent, 1990). The nodules thereafter develop into their nitrogen fixing forms.

Besides the common infection route through the root hair mediated by the flavonoids and the *nod* factors, others modes of infection like crack entry and entry through intact epidermis have also been identified and studied by some groups (Chandler *et al.*, 1982, De Faria *et al.*, 1988). However, the determining partner of rhizobial infection route and nodule formation process is the plant, not the bacteria. A single rhizobial strain can infect plants in different ways. This suggests that the host plants possess the genetic information required for the establishment of symbiosis and the rhizobia have to adapt accordingly (Denarie *et al.*, 1992).

1.2.5.2. Nitrogen fixation

After rhizobia enter the cells of the root and form mature nodules, the functional bacteroids present within the nodules are ready to fix nitrogen in co-operation with the plants (Mylona *et al.*, 1995). Two groups of genes are essential for nitrogen fixation - *nif* and *fix*. *nif* genes are homologous in most nitrogen-fixing bacteria and *fix* genes are required for symbiotic nitrogen fixation. Mutation of these genes will not abolish the formation of nodules, but nitrogen fixation will not take place (Long, 1989). Previously, studies on the nitrogen fixation process concentrated on legumes of commercial value from the subfamily *Papilionoideae* (Elliott *et al.*, 2007).

Nitrogen fixation (*nif*) genes were first identified in *Klebsiella pneumoniae* and were found to be conserved in nitrogen-fixing bacteria (Merrick, 1992, Ruvkun and Ausubel, 1980). For example, the *nifA* gene which acts as regulator to activate other *nif* genes can always be found in nitrogen fixing bacteria. About 10 *nif* genes have been observed in rhizobia but their locations differ depending on the strain. The nitrogenase enzyme complex is encoded by the *nif* genes (Hennecke *et al.*, 1985). The reaction catalysed by nitrogenase is :

$$N_2 + 8H^+ + 8e^- + 16\text{Mg-ATP} \longrightarrow 2NH_3 + H_2 + 16\text{MgADP} + 16\text{Pi}$$

Two enzymes form the nitrogenase enzymatic complex : the first enzyme, dinitrogenase, has a Fe-Mo cofactor and catalyses the main reaction. Dinitrogenase is encoded by *nifD* and *nifK*. High energy electrons are received by dinitrogenase after reduction by the second enzyme – dinitrogenase reductase. This reaction requires the hydrolysis of two ATP molecules to ADP and phosphate for each electron transferred. Dinitrogenase reductase is a homodimeric Fe-protein and is encoded by the *nifH* gene (Dixon and Kahn, 2004). Hadfield and Bulen (1969) found that the electron flux could go through hydrogen production when there is a lack of reducible substrates. Studies carried out by Georgiadis *et al.* (1992) showed that the electron transfer system was similar to other organisms (Kirn and Rees, 1992). Electrons and ATP required for nitrogen fixation are provided by plants from oxidation of dicarboxylic acid and aerobic respiration (Lodwig *et al.*, 2003, Thorneley and Ashby, 1989).

Nitrogenase is oxygen sensitive since the oxygen can damage the Fe-Mo cofactor irreversibly (Shah and Brill, 1977). However, a large amount of energy required by nitrogenase for catalysis is derived from oxidation. The bacteria solve this problem by using different strategies. Witty *et al.* (1986) reported that the bacteria have high rates of metabolic activity. This, along with the presence of an oxygen diffusion barrier in the periphery of the nodule in the nodule parenchyma decreases the level of oxygen in the nodule, thus protecting the enzyme. Reguera *et al.* (2010) found that boron, an important element in the rhizobium-legume symbiosis, is involved in the regulation of nodule oxygen concentration. Deficiency of boron leads to the formation of unstable nodules and ineffective nitrogen fixation. The findings also suggest that the expression of nitrogenase is regulated by external oxygen concentration (Dixon and Kahn, 2004).

Enzymes involved in supplying energy and electrons for nitrogen fixation are coded by *fix* genes. *fix* genes were first described in *R. meliloti.* Since that, they were found to form functional clusters with *nif* genes, and mutation of *fix* genes led to a failure to fix nitrogen (David *et al.*, 1987, Renalier *et al.*, 1987). Putnoky *et al.* (1988) found that proteins coded by *fix* genes had a number of functions in the plant-bacterium symbiosis including infection thread formation, release of rhizobia, bacteroid membrane formation and expression and regulation of *nif* genes. *fixNOQP* were found to code cbb_3 cytochrome oxidase, which is related to oxygen metabolism. The *fixGHIS* operon, along with the *fixNOQP* operon adjacent to it, are speculated to mediate the copper insertion into cbb_3. *fixABCX* were also generally present in rhizobia. *fixLJ* and *fixK* were thought to be regulators and were commonly found in the rhizobia (Schaechter, 2009).

Both *nif* and *fix* genes are essential for nitrogen fixation after the formation of bacteroids. While *nif* genes encode proteins of nitrogenase, *fix* genes encode proteins involved in electron transport to nitrogenase. Low oxygen level change the regulation and expression of both genes and the mediatory mechanisms are diverse among the rhizobia strains (Udvardi and Poole, 2013). Mature nodules develop suitable structures for nutrient exchange between rhizobia and legumes.

1.2.6. Ecology of rhizobia

Nitrogen is one of the important chemical elements in any biological system, and is required in the synthesis of amino acids, nucleotides, antibiotics and many other intermediates of metabolism required by the organism. Nitrogen content is a dominating factor contributing to agricultural productivity (Emerich *et al.*, 2009). Nitrogen fertilizers have been used to meet the requirement of agriculture, especially in developing countries like India and China. However, environmental and ecological issues have been raised since the over-utilization of nitrogen fertilizers leads to eutrophication due to fertilizer runoff. In addition, manufacture of nitrogen fertilizer continually contributes to the use of fossil fuel and the release of greenhouse gases. These factors present biological nitrogen fixation based on rhizobia-legume symbiosis as an eco-friendly alternative for the future (Herridge *et al.*, 2008).

Symbiosis between rhizobia and legume has proved to be an important biological process in natural and agricultural environments. The atmospheric inactive dinitrogen gas can be reduced to biologically active ammonia by the symbiosis, which neither plants nor animals are able to perform. Also, biological nitrogen fixation is important in balancing the denitrification events of the nitrogen cycle (Lederberg *et al.*, 2000). Recently, studies have been conducted in Africa and Asia on new rhizobial isolates which are potentially useful for improving local agricultural conditions (Aserse *et al.*, 2012, Kficfik and Cevheri, 2013). Beyhaut *et al.* (2013) indicated that planting of legumes could help prairie restoration with inoculation of suitable rhizobial cultures.

We should realize that long-term legume growth and rhizobium-legume symbiosis can prevent soil nitrogen leakage by using nitrogen from both the symbiotic partners. Also, rhizobia can convert the falling leaves to nitrogen sources for other neighbouring plants (Schaechter, 2009). However, soil and geographic conditions are different in different parts of the world and hence studies are required to deliver rhizobia from lab to farm application.

1.3. Project background

As a part of a population genomics project in the Department of Biology at the University of York, UK, 72 isolates were obtained from a roadside verge adjacent to Wentworth College on the University of York campus. The isolates were

obtained from two leguminous plants from an area of one square meter. Half of the isolates were obtained from *Vicia sativa*, hence called VSX strains. The other half were isolated from *Trifolium repens*, and hence called TRX strains.

In order to investigate the relatedness between the two groups of isolates vis-à-vis known *Rhizobium* strains, the strains were sequenced by 454 pyrosequencing. The sequence data of these strains was analysed and genes in their genomes were identified using the genome data from the sequenced strain *R. leguminosarum* bv*. viciae* 3841 (*Rlv.*3841), as the reference genome. The analysis showed the presence of five genes that were found to be present in all 36 bv*. viciae* strains (i.e. strains isolated from *Vicia sativa*) as well as *Rlv.*3841, but absent in the bv. *trifolii* strains (i.e. strains isolated from *Trifolium repens*). Since these genes were specifically found in bv. *viciae* but not in bv*. trifolii*, they were termed “biovar *viciae* specific” genes, abbreviated as *bvs* genes.

The five *bvs* genes are located on plasmid pRL8 in *Rlv.* 3841 (Young *et al.*, 2006). We have already discussed some of the accessory genes located on pRL8. Previous work on 454 genome data within the research group has shown that these genes may be laterally transferred. Ramachandran *et al.* (2011) considered pRL8 as specific for *Pisum sativum* symbiosis, enabling rhizobia to form nodules on pea, since they found in the pea rhizosphere there was a specific increase in the expression of many genes on pRL8. The bvs genes were not, however, among the genes that they noted.

Complex host-range patterns were observed in biovar *viciae* by Mutch and Young (2004). Furthermore, previous studies on metabolism in different species of rhizobia also found diversity in the utilization of different substrates (Ramirez-Bahena *et al.*, 2008). Host range and metabolic patterns are important features in the identification and definition of unknown isolates. The analysis of new strains using these two criteria will contribute to our knowledge of rhizobia.

1.4. Aims of the project.

The project aims to investigate phenotypic features of isolates from different areas and study the functions of the five biovar *viciae* specific genes. Besides their location on pRL8 and putative annotation, no previous attempt has been made to study the enzymatic function of these genes and their role in the nodulation process. Partial information of genome data is available for some of

these strains in the BLAST database hosted at the University of York (http://blast.york.ac.uk/blast/blast.html). Based on the results from my initial BLAST search, my hypothesis for this project is : the five *bvs* genes are essential or play a role in the nodulation of plants by strains belonging to the biovar *viciae*.

The aims of this project are as follows :

a. Study the phenotypic variation in rhizobial isolates from different hosts.
b. Study the distribution of *bvs* genes in isolates belonging to different biovars.
c. Study the effect of single-gene mutation and the effect of mutation complementation of *bvs* genes under nodulation competition stress.
d. Study the enzymatic activities of *bvs* genes showing altered phenotype on mutation to understand its role in nodulation.

1.5. Outline of thesis.

The following chapters are included in this thesis:

Chapter 2 : Study of host specificity and metabolic differences in the *R. leguminosarum* complex : In this chapter of the thesis rhizobial strains were tested for their nodulation ability on common host plants to validate them and to get an idea of their host range. Moreover, their metabolism patterns were investigated using the BIOLOG GN2 MicroPlates and their clustering, based on metabolic ability, compared with phylogenetic clustering of a taxonomic marker gene to check for similarities and difference in information contained in data analysed from the two very different methods of classification.

Chapter 3 : Distribution and phylogenetic relation between the *bvs* genes in rhizobia : In this chapter of the thesis the distribution of *bvs* genes in rhizobia was studied using PCR and BLAST. In addition, phylogenetic network trees were made using the *bvs* genes sequences by sequencing the PCR amplicons and used for understanding the relationships between the test strains.

Chapter 4 : Mutation and complementation of *bvs* genes and their effects on bacterial competition and plant growth : In this chapter of the thesis ingle gene mutants in the *bvs* genes were isolated and the effects of *bvs* gene mutation on bacteria competition and the growth of the host plant were studied. Cross nodulation tests were carried out using either pure culture or wild type and mutant mixed culture inoculations. Plant dry weights and occupancy ratios were

measured. Complementation of genes in mutants showing altered growth phenotype was carried out and the effects of complementation tested again using mixed culture inoculation. *bvs* genes playing a role in nodulation or nodulation competition were identified. The possibility of *bvs* genes constituting an operon was also investigated.

Chapter 5 : The *bvs4* and *bvs5* genes **encode a nitrilase and sulfite oxidase :** In this chapter we investigated the enzymatic activities of the two *bvs* genes. For *bvs4*, rhizobial growth tests were performed in the presence of nitrile. The ammonia released during the breakdown of nitrile during growth was detected by an ammonia assay. The aim was to check the role of *bvs4* on the ability of the bacteria to grow in the presence of nitrile to produce ammonia. For *bvs5*, sulfite oxidase and nitrate reductase enzyme assays were carried out to clarify the enzymatic function of *bvs5*.

Chapter 6: General discussion: A general discussion and summary of this project and possible future work that can be done to extend this work.

CHAPTER 2 : STUDY OF HOST SPECIFICITY AND METABOLIC DIFFERENCES IN THE *R. leguminosarum* SPECIES COMPLEX.

2.1. Introduction

Contemporary methods of bacterial classification mostly rely on their phenotypic characteristics. The widely accepted authoritative text on bacterial classification, Bergey's Manual of Systematic Bacteriology (2001), states that the criteria used for classification of bacteria are based on the phenotypic behaviour of bacteria in order to assist the microbiologist in identifying the bacteria. Thus, bacterial strains at any taxonomic rank / level in the hierarchy of classification (for example, a species) share a number of common phenotypic characteristics (morphological, physiological or biochemical). Hence, the terms used to define a taxonomic rank / level in microbiology usually include these characteristics.

Rhizobia were first isolated from root nodules and hence all root-nodule bacteria were classified under the genus *Rhizobium*. The early attempts to classify rhizobia were based on their ability to infect host plants. The bacterial isolates that infected identical host plants were included in a single group called a 'cross-inoculation' group (Fred *et al.*, 1932). Although it has now been abandoned as a reliable taxonomic marker, it is still used to investigate the host-specificity or host range of new root-nodule bacteria. Investigations into the growth characteristics of the root-nodule bacteria led to the discovery that the bacteria could be grouped based on their generation times into fast- and slow-growing rhizobia. The fast-growing rhizobia have a generation time of less than 6 hours whereas the slow-growing rhizobia had a generation time of more than 6 hours. The slow-growing rhizobia were later transferred to a separate genus called *Bradyrhizobium* (Jordan, 1982).

Attempts have been made to classify rhizobia using different methods including serology, multilocus enzyme electrophoresis, RFLP fingerprinting, substrate utilization etc. (Rahmani *et al.*, 2011, Vandamme *et al.*, 2002). These methods show that the rhizobia are diverse in terms of their genetic and phenotypic composition. However, these methods have immense promise in their use to investigate the diversity of rhizobia. In some cases they have been used to define new species of rhizobia. For example, analysis of the genome resulted in the description of *R. hainanense* (Chen *et al.*, 1997) whereas phenotypic differences were also used to describe *R. multihospitium* (Han *et al*., 2008).

The aim of this research is to study metabolism and host specificity in the *Rhizobium leguminosarum* species complex. Hence the study will start by looking at the host-specificity of the test strains by performing cross-nodulation tests. The metabolic profiles of the strains will be studied using the Biolog Gram Negative 2 (GN2) MicroPlates from Biolog Inc.

2.1.1. Cross-nodulation tests to study host-specificity / host range :

The information relating to the host plants for the strains to be used in this study was available. The cross-nodulation tests were carried out to achieve two aims. First, the cross-nodulation test would validate the strains, i.e. they would confirm that the strains have retained their ability to form nodules on their native host plant, since the native host plants would be included as one of the test plants. Second, the cross-nodulation test would provide data which would enable us to make informed deductions about the host-range of the test species. Nodulation tests and cross-nodulation tests are regularly used for rhizobial strain validation and in the description of new species or strains of rhizobia (van Rhijn and Vanderleyden, 1995).

Cross-nodulation tests have been used in different studies since described in –depth by Vincent (1970). The studies include microbiological investigation, investigations into early stage nodulation and nodulation genetics (Grassia and Brockwell, 1978, Marshall *et al.*, 1993, Napoli and Hubbell, 1975).

Spaink *et al.* (1989) included cross nodulation test as a main part of their *nodE* study to prove that *nodE* expression was controlled by the nodulation gene promoter which, in turn, was induced by flavonoids. They also proved that this gene was responsible for the difference in host range in the two biovars *R. leguminosarum* bv. *viciae* and bv. *trifolii*. Similarly, Economou *et al.* (1994) studied the effect of complementing *nodO* genes on *nodE* mutants in both biovars. They found that *nodO* gene could trigger *nodE* mutants of *R. leguminosarum* bv. *trifolii* to form nodules on *Vicia hirsuta*. Thus, they concluded that *nodO* could also be a determinant involved in the selection of host in *R. leguminosarum* bv. *viciae*. More recently, Karunakaran *et al.* (2009) studied the development of *Rlv*. 3841 bacteroids and effect of mutation of certain genes on nodulation. *Pisum sativum* and *Vicia cracca* were chosen to be the hosts of nodulation tests. Their data showed some genes involved in *Rlv*. 3841 metabolism and membrane synthesis played an important role in

the development of bacteroids. However, many of the strongly induced genes in both host plants were observed to be absent in other strains of rhizobia, which suggested that diversity of different rhizobia is related to their specific hosts. Van Berkum *et al.* (2010) applied multilocus sequence typing (MLST) to study the structure of megaplasmids in *Medicago* rhizobia, cross nodulation was carried out on *Medicago sativa* plants using a number of rhizobia carrying different *nodC* alleles and corresponding *nifD*. Results showed that failure to form symbiosis or ineffective nitrogen fixation were observed on recombination strains, suggested that loci on both the megaplasmids were essential for effective *Medicago* symbiosis.

Rhizobia isolated from *Vicia faba* are notable since they are distinguishable from isolates obtained from other legumes (van Berkum *et al.*, 1995). The most interesting finding was reported by Mutch and Young (2004). They studied the specificity of rhizobia host range in biovar *viciae* strains. Cross-nodulation tests were conducted among isolates from each of eight plant species in the tribe Vicieae against the other seven plants, and the results showed that only 34% of isolates from the other plants could form nodules on *Vicia faba*, but *Vicia faba* isolates could form nodules on a significant proportion of other host plants. They concluded that this variation could possibly occur due to plasmids transfer.

Cross-nodulation tests, along with other methods, are commonly used in the definition of new species. Graham *et al.* (1991) indicated that a cross nodulation assay was an essential part of describing new species, since host range was an important feature of rhizobia. Zhou *et al.* (2010) described *Mesorhizobium robiniae*, isolated from *Robinia pseudoacacia* in Yangling, China. CCNWYC 115^T was described as the type strain and was tested by cross inoculation on seven common legume plants from the same region which included *Vicia*, common bean and clover plants. Results showed that none of these plants could have symbiosis with this strain. Along with other phenotypic and genotypic results, the cross-nodulation test results were used in describing the new species. New *Bradyrhizobium* species with wider host ranges were discovered by Chang *et al.* (2011). Five isolates from *Arachis hypogaea* (peanuts) and *Lablab purpureus* were found to be new species and cross nodulation tests demonstrated their nodulation ability on these two hosts and *Vigna unguiculata*, but not on other common *Vicia* and clover

plants. Based on their observations, they described the species *Bradyrhizobium lablabi* and CCBAU 23086^T was designated as the type strain. Lin *et al.* (2009) found a set of *Rhizobium* isolates from *Sphaerophysa salsula* in saline soil of north-west China, which they indicated were distinct from other recognizable rhizobia strains with respect to their 16S rRNA sequence. Xu *et al.* (2011) clarified their taxonomic relationship. Their cross nodulation tests carried on the original host *Sphaerophysa salsula* and other common plants of the region showed this new species could only form effective nodules on their original host and *Medicago sativa* but on no other plants. These strains were shown to have novel symbiotic genes and the isolates were described as a new species, *Rhizobium sphaerophysae*. Qin *et al.* (2012) isolated five strains from *Sphaerophysa salsula* in the same region, and their cross nodulation tests showed that five tested strains could only form nodules on their original host but no other plants that they tested. *Rhizobium helanshanense* was described by them. *Mesorhizobium qingshengii* was described by Zheng *et al.* (2013) - four out of five strains in this new species were reported to form nodules on *Astragalus sinicus* and less frequently on *Astragalus adsurgens*, but no other plants that they tested. *Rhizobium qilianshanense* was described by Xu *et al.* (2013) and all strains were found form nodules exclusively on *Oxytropis ochrocephala*. Wang *et al.* (2011) isolated four distinct rhizobia from root nodules of *Phaseolus vulgaris*, *Mimosa pudica* and *Indigofera spicata* in Yunnan. Using cross-nodulation tests, 16S rRNA and housekeeping genes analysis they identified the new type species *Rhizobium vallis* CCBAU 65647^T with wide host range.

Combination of cross nodulation tests with other phenotypic and genotypic analysis is now commonplace in the characterization of new rhizobia isolates and in studying the diversity of these new isolates vis-à-vis other species and biovars. Diouf *et al.* (2010) studied a group of *Mesorhizobium* strains isolated from *Acacia seyal* in different regions of Senegal. Cross inoculation studies of these isolates on their original host and six other woody legumes showed that the isolates had a broad host range and hence had the potential to be developed as inoculants. Woody legume symbioses in Africa were also studied by Degefu *et al.* (2011). Eighteen *Mesorhizobium* isolates were obtained from nodules of *Acacia abyssinica*, *A. senegal*, *A. tortilis* and *Sesbania sesban* in Ethiopia. Based on 16S rRNA, these isolates were divided into three groups. Cross-nodulation tests showed that, regardless of

their genotypic groups, rhizobia-legume symbiosis could be observed using these isolates against other common woody legumes in the area. Degefu *et al.* (2012) indicated these results were in line with their *nodC* analysis. This suggested that the selection of the legume host was dictated to a greater extent by the plasmid borne symbiotic genes compared to the chromosomal genes. In a second project carried out in Southern Ethiopia, Degefu *et al.* (2012) demonstrated using nodulation tests showed their *Ensifer* isolates were able to nodulate their original hosts and other closely related legumes. The studies suggest that these broad host-range isolates were of value in commercial inoculant selection. The legume *Millettia pinnata* was found to have broad symbiosis forming ability similar to rhizobia isolated by Rasul *et al.* (2012) from India. Isolates from this legume were identified to be highly diverse in both genetic analysis (16S rRNA, housekeeping and nodulation genes) and phenotypic studies (substrate utilization). However, nodulation tests showed these isolates could form symbiosis with different types of legumes, including crop and woody legumes. They indicated that isolates from *M. pinnata*, like the African isolates, had significant value in agricultural use.

These broad-host-range rhizobia isolates identified by studies above can not only benefit local agriculture by promoting plant growth, but also can be used as an important participant of ecological restoration process. Beyhaut *et al.* (2013) indicated that legumes were the key component of nitrogen-limited prairie ecosystems. The shortage of legume plants in certain prairie areas could be overcome by inoculating the legumes in the area by the use of suitable rhizobia strains. Their suggestion was to select legumes and study the local soil conditions, and then find out rhizobia strains to inoculate thus allowing them to nodulate their hosts continually.

As is demonstrated in the literature cited above, cross-nodulation tests are an important method in rhizobia investigation and have been used in many studies. The main aim of the test is to identify suitable host plants for rhizobial strains, or to clarify the nodulation ability of test rhizobia on the target plants.

2.1.2. Metabolic profiling using the Biolog GN2 MicroPlates :

The phenotype of a bacterium plays a significant role in the classification of bacteria. Phenotypic characteristics can reflect the diversity among species.

Ramirez-Bahena *et al.* (2008) and Tian *et al.* (2008) have used phenotypic differences to describe new species of *Rhizobium.* Hence, studying the phenotypes of the strains to be used in the study was seen as a logical line of investigation.

The clustering of strains using the phenotype data can be compared to phylogenetic clustering using sequence data of a core gene like *celC* to check for similarities and differences in the clustering patterns from two diverse methods. The *celC* gene phylogenetic tree has been demonstrated to have the ability to distinguish between rhizobial strains. The differences in the clustering patterns give an idea of the relative difference between the genetic and phenotypic behaviour of the strains.

The type strains from different species will be used in a systematic phenotype test. The aim of this test is to look for obvious difference among these type strains in substrate utilization. The phenotypes of the test strains were studied using metabolic fingerprinting technology from Biolog Inc. The GN2 Biolog plates were used in this study to obtain a metabolic fingerprint of the species. The system uses a patented redox chemistry that responds to metabolism i.e. respiration rather than to metabolic by-products. The system works as a universal reporter of metabolism and simplifies the testing process as colour developing chemicals do not have to be added.

The Biolog plate has 96 wells containing a dehydrated mixture of the redox dye tetrazolium violet, a buffered nutrient medium, a gelling agent and a different carbon source in each well except one which serves as a blank (Figure 2.1). The wells are hydrated by inoculating a suspension of the test bacterium. The plates are incubated (usually 24-48 hours) and read to determine substrate utilization. If the carbon substrate is utilized the dye in the well turns purple; if it not utilized, the dye remains colourless. The reduced formazan dye is insoluble and settles at the bottom of the well forming a purple lawn at the bottom which is clearly visible to the naked eye. A visually distinct pattern of purple wells forms for different microbes. The pattern is called the "metabolic fingerprint" of the test bacterium (Bochner, 1989).

A1 Water	A2 α-Cyclodextrin	A3 Dextrin	A4 Glycogen	A5 Tween 40	A6 Tween 80	A7 N-Acetyl-D-galactosamine	A8 N-Acetyl-D-glucosamine	A9 Adonitol	A10 L-Arabinose	A11 D-Arabitol	A12 D-Cellobiose
B1 i-Erythritol	B2 D-Fructose	B3 L-Fucose	B4 D-Galactose	B5 Gentiobiose	B6 α-D-Glucose	B7 m-Inositol	B8 α-D-Lactose	B9 Lactulose	B10 Maltose	B11 D-Mannitol	B12 D-Mannose
C1 D-Melibiose	C2 β-Methyl-D-Glucoside	C3 D-Psicose	C4 D-Raffinose	C5 L-Rhamnose	C6 D-Sorbitol	C7 Sucrose	C8 D-Trehalose	C9 Turanose	C10 Xylitol	C11 Methyl Pyruvate	C12 Mono-Methyl-Succinate
D1 Acetic Acid	D2 Cis-Aconitic Acid	D3 Citric Acid	D4 Formic Acid	D5 D-Galactonic Acid Lactone	D6 D-Galacturonic Acid	D7 D-Gluconic Acid	D8 D-Glucosaminic Acid	D9 D-Glucuronic Acid	D10 α-Hydroxy Butyric Acid	D11 β-Hydroxy Butyric Acid	D12 γ-Hydroxy Butyric Acid
E1 p-Hydroxy Phenylacetic Acid	E2 Itaconic Acid	E3 α-Keto Butyric Acid	E4 α-Keto Glutaric Acid	E5 α-Keto Valeric Acid	E6 D,L-Lactic Acid	E7 Malonic Acid	E8 Propionic Acid	E9 Quinic Acid	E10 D-Saccharic Acid	E11 Sebacic Acid	E12 Succinic Acid
F1 Bromo Succinic Acid	F2 Succinamic Acid	F3 Glucuronamide	F4 L-Alaninamide	F5 D-Alanine	F6 L-Alanine	F7 L-Alanyl-glycine	F8 L-Asparagine	F9 L-Aspartic Acid	F10 L-Glutamic Acid	F11 Glycyl-L-Aspartic Acid	F12 Glycyl-L-Glutamic Acid
G1 L-Histidine	G2 Hydroxy-L-Proline	G3 L-Leucine	G4 L-Ornithine	G5 L-Phenylalanine	G6 L-Proline	G7 L-Pyroglutamic Acid	G8 D-Serine	G9 L-Serine	G10 L-Threonine	G11 D,L-Carnitine	G12 γ-Amino Butyric Acid
H1 Urocanic Acid	H2 Inosine	H3 Uridine	H4 Thymidine	H5 Phenyethylamine	H6 Putrescine	H7 2-Aminoethanol	H8 2,3-Butanediol	H9 Glycerol	H10 D,L-α-Glycerol Phosphate	H11 Glucose-1-Phosphate	H12 Glucose-6-Phosphate

Figure 2.1. Biolog GN2 MicroPlate - Gram Negative Identification Test Panel : Carbon substrates included in the Biolog GN2 plate.

Recent descriptions of new species of rhizobia use both genotypic and phenotypic methods. The convenience of the Biolog GN2 plate makes it an easy and simple method to carry out phenotype studies.

Rhizobium borbori was described by Zhang *et al.* (2011) based on housekeeping genes, DNA-DNA hybridization and metabolic studies using Biolog plates. The type strain DN316^T was compared with other new type strains in the study. *Rhizobium pusense* was described by Panday *et al.* (2011). Biolog GN2 plates were used to compare the substrate utilization pattern of the type strain NRCPB10^T with those of other recognizable closely related type strains. These characters were used to characterise them and the new species were described. Similarly, work by Bibi *et al.* (2012) describing *Rhizobium halophytocola* also compared the Biolog GN2 plate data between the type strain of this species, YC6881^T, and the type strains of other closely related species. Different utilization patterns supported the definition of new species.

Rhizobia which are isolated from unusual environments normally show different substrate utilization patterns. Zhang *et al.* (2012) isolated two rhizobial strains from oil-contaminated soil. 16S rRNA analysis suggested the two strains should form a new species. Biolog GN2 plate data confirmed this hypothesis and thus they were named *Rhizobium petrolearium*. Two rhizobial isolates isolated from beach sands were confirmed to belong to a new species, which was later named *Rhizobium subbaraonis* (Ramana *et al*., 2013), based on evidence that included DNA-DNA hybridization and Biolog GN2 study.

Strains of a new species, *Rhizobium azibense*, were isolated from common bean (*Phaseolus*) by Mnasri *et al.* (2014), and housekeeping genes analysis, DNA-DNA hybridization and Biolog GN2 and GP2 plates were all used to define them. Two species from rice plants, *Rhizobium rhizoryzae* and *Rhizobium straminoryzae* were described by Zhang *et al.* (2014) and Lin *et al.* (2014), respectively, using similar methods including Biolog GN2 plates. Biolog GN2 plates were also applied to compare different rhizobia for in-depth studies.

Marek-Kozaczuk *et al.* (2013) found that *Rhizobium pisi* sv. *trifolii* K3.22, although 16S rRNA and housekeeping genes analysis indicated that it

belonged to *R.pisi* due to similarity with *R. pisi* DSM30132^T, possessed biovar *trifolii*-like nodulation genes and showed differences in Biolog GN2 data with its type strain. This suggests symbiotic plasmids transfer between symbiovars *viciae* and *trifolii*.

Another study on *Rhizobium leguminosarum* bv. *trifolii* was conducted by Stasiak *et al.* (2014). They created *Rhizobium leguminosarum* bv. *trifolii* TA1 derivatives by curing each of its symbiotic plasmids, and Biolog GN2 data showed that all derivatives had metabolic deficiencies. Their results implied that all these symbiotic plasmids played an important role in the growth and metabolism of rhizobia and possible cooperation between plasmids to support the phenotypic and symbiotic properties of rhizobia.

This work studies the metabolic differences between the different species and biovars of rhizobia. The data obtained from the phenotype analysis will be used for clustering the strains using UPGMA to study the similarities and differences in their metabolic patterns of substrate utilization.

Work in this chapter aims to validate the rhizobia strains which would be used for further study. For the nodulation tests, it is expected that all the biovar *viciae* strains and isolates will nodulate plants form the genus *Vicia* and other closely related plants forming pink, healthy, effective nodules. Rhizobial strains from other biovars are not expected to form symbioses with the test host plants. Since strains that are closely related are expected to be metabolically similar, the study using the Biolog GN2 MicroPlate system is predicted to show that the clustering of strains using the metabolic data will closely match their clustering using sequence data.

2.2. Materials and methods

2.2.1. Strains used in this study

A total of fifteen strains were used in this study. Six of the fifteen strains were type strains of different species or biovars of *Rhizobium*. Eight field strains from *Vicia* plants were included in the study. Of the eight field strains, two strains were isolated in Sweden, three strains were isolated in Scotland while the remaining three strains were from a field adjacent to Wentworth College, University of York. The strains were selected to check host specificity and metabolic differences in the population.

The type strains have been described in literature and were maintained as glycerol stocks within the lab. Field strains from Sweden were kindly provided by Osei Ampomah from the Swedish University of Agricultural Sciences, Sweden (OYAVB 169.1 and OYAVB296.5) while the field strains from Scotland were kindly provided by Dr. Euan James of The James Hutton Institute, Scotland (S25, S34 and S36). The three VSX field strains used in the study (VSX11, VSX26 and VSX 36) were available within the research group as isolates obtained for a population genomics project conducted at the University of York. The strains were sequenced using the 454 pyrosequencing technology from Roche. The data from sequencing was used for my study. The strains were used in this study and the relevant information is listed in Table 2.1.

2.2.2. Study of host-specificity / host range using nodulation test

In order to assess the potential host-range of the test strains their host-specificity was studied using the host plants *Vicia sativa, Vicia cracca, Pisum sativum* and *Lathyrus pratensis*.

2.2.2.1. Selection of host plants for the assay

Vicia sativa ssp. *segetalis, Vicia cracca, Pisum sativum* (variety 'Douce Provence') and *Lathyrus pratensis* were selected as the host plants to study the host-specificity or host-range of the test strains. The seeds for the plants used in the nodulation tests for determining host-specificity were purchased from Emorsgate Seeds (Norfolk, UK).

2.2.2.2. Strains used in the assay

All the strains listed in Table 2.1 were included in studying their host-specificity using nodulation test.

Table 2.1. Type strains and Field strains used in this study.

Strain name	Species	Biovar	Strain Type	Host	Described by	Genome data
USDA 2370^T	*Rhizobium leguminosarum*	*viciae*	Type Strain	*P. sativum*	Ramirez-Bahena *et al.* (2008)	Partial
DSM 30132^T	*Rhizobium pisi*	*viciae*	Type Strain	*P. sativum, V. sativa*	Ramirez-Bahena *et al.* (2008)	Partial
CCBAU 33202^T	*Rhizobium fabae*	*viciae*	Type Strain	*V. faba* , *P. sativum* and *Lathyrus* species	Tian *et al.* (2008)	Partial
ATCC 14482^T	*Rhizobium phaseoli*	*phaseoli*	Type Strain	*P. vulgaris*	Ramirez-Bahena *et al.* (2008)	Partial
CIAT 899^T	*Rhizobium tropici*	not available	Type Strain	*P. vulgaris, L. esculenta* and *L. leucocephala*	Martínez-Romero *et al.* (1991)	Fully sequenced (Ormeno-Orrillo *et al.*, 2012)
CFN 42^T	*Rhizobium etli*	*phaseoli*	Type Strain	*P. vulgaris*	Segovia *et al.* (1993)	Fully sequenced (Brom *et al.*, 2000)
Swedish strain OYAVB 169.1	*Rhizobium leguminosarum*	viciae	Field Strain	*V. cracca*	Not available	Partial VCS_3*
Swedish strain OYAVB 296.5	*Rhizobium leguminosarum*	*viciae*	Field Strain	*V. cracca*	Not available	Partial VCS_4*
Scottish strain 25 (S25)	*Rhizobium leguminosarum*	*viciae*	Field Strain	*V. faba*	Not available	Not available
Scottish strain 34 (S34)	*Rhizobium leguminosarum*	*viciae*	Field Strain	*V. faba*	Not available	Not available
Scottish strain 36 (S36)	*Rhizobium leguminosarum*	*viciae*	Field Strain	*V. faba*	Not available	Not available
Wentworth strain VSX11	*Rhizobium leguminosarum*	*viciae*	Field Strain	*Vicia sativa*	Not available	Partial
Wentworth strain VSX26	*Rhizobium leguminosarum*	*viciae*	Field Strain	*Vicia sativa*	Not available	Partial
Wentworth strain VSX36	*Rhizobium leguminosarum*	*viciae*	Field Strain	*Vicia sativa*	Not available	partial

* = sequence data maintained in the University of York BLAST database.

2.2.2.3. Preparation of medium for plant growth

The plants used in the nodulation competition assay were grown using sterile vermiculite as the supporting medium. The bigger plants, viz. *Pisum sativum*, were grown in 250 ml Pyrex borosilicate Erlenmeyer flasks containing 150 ml vermiculite. The smaller plants, viz. *Vicia sativa*, *V. cracca,* and *L. pratensis* were grown in 60 ml Pyrex borosilicate tubes containing 30 ml vermiculite.

Before planting, the flasks and tubes containing vermiculite were plugged with cotton wool and autoclaved at 15 psi in an autoclave 121^{O}C for 20 minutes and allowed to cool. The cooled vermiculite was moistened with sterile Fahraeus N-free liquid medium (70 ml per flask and 14 ml per tube) (Vincent, 1970). The flasks and tubes were covered with sterile film and stored in the cold room until further use.

2.2.2.4. Preparation of seeds for growth

The seeds were washed with multiple rinses of tap water and left to imbibe in tap water for 24 hours. After imbibition the seeds were again rinsed with tap water. The seeds were then surface sterilized with 70% ethanol for 1 minute. The ethanol was removed and the seeds were rinsed with sterile distilled water. The seeds were then treated with 3% hydrogen peroxide for 1 minute. The peroxide was removed and the seeds were washed with seven changes of sterile distilled water.

A small number of seeds were transferred to TY and nutrient agar plates to check for efficiency of sterilization (Burton *et al.*, 1984, Somasegaran and Hoben, 1994, Vincent, 1970). The remaining seeds were for germination as described in section 2.2.2.6.

2.2.2.5. Preparation of test cultures

A loopful of the test cultures from glycerol stock culture was streaked on normal sterile TY agar plate. The plates were incubated overnight at 28^{O}C. A single colony from each plate was transferred to appropriately labelled 50 ml Corning polypropylene tubes containing 25 ml of sterile liquid TY medium. The tubes were vortexed briefly to disperse the cultures and then incubated at 28^{O}C on a rotary shaker at 150 rpm for 24

hours. The cultures were centrifuged at 4000 rpm at 20^{O}C for 10 minutes, washed twice in physiological saline, and resuspended in physiological saline to A_{600} = 0.1 (Burton *et al.*, 1984, Moawad and Bohlool, 1984).

2.2.2.6. Seed germination, inoculation and planting

The sterilized seeds were placed in a sterile beaker containing a sterile moist tissue paper and incubated at room temperature until the seeds germinated and showed presence of a radicle. Seeds with clean and healthy radicles were separated from the non-germinating seeds and seeds with unhealthy radicals. The clean, germinating seeds were washed again with multiple rinses of sterile water in sterile beakers and used for the bacterial inoculation stage in the assay.

The cleaned sterile seeds were then placed in the prepared bacterial cultures for 20 minutes so that the bacteria adsorb on the seed surface. Four seeds were used per test culture (replicates) (Lei *et al.*, 2008). The seeds were then transferred to their respective growth container (flask / tube) in a laminar hood using sterile forceps. The seed was placed such that the radicle was embedded in the vermiculite. The seed was covered with a small amount of sterile vermiculite and the neck of the flasks and tubes were plugged with sterile cotton wool (Vincent, 1970).

Four positive and four negative control replicates were also set up for the assay. The plants included in the positive and negative control were not inoculated with any bacterial culture. However, the plants included in the positive control were watered with sterile Fahraeus medium containing 0.05% potassium nitrate as the nitrogen source (Zézé *et al.*, 2001). The plants included in the negative control were watered regularly with sterile nitrogen-free Fahraeus medium.

2.2.2.7. Plant growth and harvesting

The inoculated flasks and tubes containing the inoculated seeds were transferred to controlled temperature plant growth room. The room had a light cycle of 16 hour at 22^{O}C and dark cycle of 8 hour at 18^{O}C. Initially, the plants were watered using Fahraeus medium every three days; the frequency was increased when the growing plants lost moisture rapidly by transpiration drying out the vermiculite.

The plants were grown for five to seven weeks before harvesting. The time of harvesting was different for each plant since they grow at different rates. The harvesting was done just before the plants started flowering since the nodules start to senesce when flowering begins.

For harvesting, the vermiculite was carefully poured out of the flasks / tubes and the plant removed gently. The vermiculite adhering to the roots was dislodged by immersing the roots in warm water and removing the vermiculite with a brush. The entire plant was then rinsed in two changes of clean water. The washed plants were gently pat-dried using soft tissue paper and observed for presence of nodules on the roots (Moschetti *et al.*, 2005).

2.2.3. Study of metabolic diversity using Biolog GN2 MicroPlates

The metabolic diversity of the test strains was studied using the Biolog GN2 MicroPlates. The plates contain a variety of carbon sources which can be used to assess metabolic diversity of bacteria as described in Section 2.1.2.

2.2.3.1. Strains used in the assay

Only the type strains listed in Tables 2.1 were included in the study of metabolic differences using the Biolog GN2 Microplate system.

2.2.3.2. Growth of test cultures

A loopful of test culture from glycerol stock was streak-plated on a sterile Tryptone-Yeast extract (TY) agar plates and incubated at 28^{O}C for 48 hours. A heavy inoculum from the plate was transferred to a fresh sterile TY agar plate and incubated at 28^{O}C for 24 hours. 1 ml of physiological saline was added to the incubated plate and the bacterial growth was dislodged from the plate and brought into suspension using a thin, sterile, flexible inoculation loop. The suspension was pipetted into a sterile 2 ml Eppendorf tube. Additional 1 ml of physiological saline was added to the TY agar plate and the plate was swirled to collect the bacteria adhering to the plate. This washing, too, was transferred to the Eppendorf tube.

The suspension was vortexed for 15 seconds to break up clumps of bacteria and then centrifuged at 13,000 rpm for 5 minutes on a bench-top

centrifuge. The supernatant was discarded and the pellet washed with two changes of physiological saline. The bacteria were then suspended in 1 ml of physiological saline and vortexed again for 15 seconds. Small aliquots of the culture were added to 20 ml of sterile physiological saline to get an inoculum density of A610 = 0.1 as read on an ELISA reader (Thermomax, Thermo Scientific) using sterile saline as blank (Lad, 2013).

2.2.3.3. Inoculation and incubation of Biolog plates

150 µl of the inoculum was added to each well of the Biolog plate using 8-channel multi-tip pipette. The culture was thoroughly mixed with the dehydrated substrate and dye in the well by gently pipetting the culture up and down to ensure resuspension and even distribution of the components. Two replicates were set up per test strain. After inoculation, the plate was covered with its lid and sealed along edges with Parafilm® to prevent drying of peripheral wells. Parafilm® has high oxygen permeability and low permeability to water vapour and hence would not affect oxygen diffusion into the plate and in turn, the growth of bacteria. The plates were incubated for 48 hours at 28^{O}C in an incubator.

2.2.3.4. Reading the plates

After incubation, the plates were read at 590 nm on an ELISA reader (Thermomax, Thermo Scientific). The absorbance value from the well A1, not containing a carbon substrate but imbibed with the dye functioned as the blank. The absorbance value of A1 was subtracted from all the other wells to obtain the corrected absorbance. The results were printed and the values were manually keyed into a Microsoft Excel 2010 spreadsheet.

2.2.3.5. Analysis

The results of the Biolog metabolic fingerprinting experiments were analysed to look for patterns in the utilization of substrates. The strains and substrates were clustered using UPGMA and Principal Coordinate Analysis (PCoA) using statistiXL and R programming language respectively.

UPGMA is a method of hierarchical clustering that clusters entities by agglomeration based on pairwise similarities of variables, assuming a

constant rate of change (Sokal *et al.*, 1958). The analysis generates a similarity (or dissimilarity) matrix which is then used to generate a rooted tree or dendrogram. Clusters that are most similar (or least distant) are combined into a higher-level cluster. This method is used to study relationships between phenotypes by creating 'phenetic trees' or 'phenograms' (Legendre and Legendre, 2012).

PCoA is also a method to explore and visualize similarities or dissimilarities in data. It starts with a similarity matrix or dissimilarity matrix (i.e. a distance matrix) and assigns each item a location in a low-dimensional space. Applying UPGMA and PCoA to Biolog data will assess the similarity in the phenotypic characteristics of the test strains.

The clustering of strains using the phenotype data can be compared to phylogenetic clustering using sequence data of the *celC* gene to check for similarities and differences in the clustering patterns using the different approaches. The *celC* clustering was performed using clustalX alignment and SplitsTree software.

In order to investigate the genetic basis for variation in the utilization of sugars a third analysis was carried out. The genes and enzymes involved in the utilization of sugars were identified using the KEGG PathComp tool (http://www.genome.jp/tools/pathcomp/). A putative catabolic pathway was created for each sugar substrate. The gene sequences of the enzymes involved in sugar utilization in *R. leguminosarum* biovar *viciae* were downloaded from the NCBI gene database. The gene sequences were used to perform Local-BLAST and NCBI-BLAST to verify the presence of genes in the test strains for which equences were available.

2.3. Results

The results of the investigation into the host-specificity and metabolic diversity of the rhizobium type strains and field strains is presented in this section.

2.3.1. Study of host-specificity / host range using nodulation test

The host-specificity / host-range of strains from different species and biovars of the genus *Rhizobium* were investigated. A nodulation test using the test strains was performed to see if they had the ability to form

nodules on select plants from the genus *Vicia* and the closely related genera *Pisum* and *Lathyrus*.

Table 2.2. Results of study to determine the host-specificity / host range of the test isolates on test host plants in terms of presence (+) or absence (-) of nodules on the plants used in the experiment.

	Vicia sativa	*Vicia cracca*	*Pisum sativum*	*Lathyrus pratensis*
R. leguminosarum USDA 2370	+	+	+	+
R. pisi DSM30132	+	+	+	+
R. fabae CCBAU 33202	+	+	+	+
R. phaseoli ATCC 14482	-	-	-	-
R. tropici CIAT 899	-	-	-	-
R. etli CFN42	-	-	-	-
OYAVB296.5	+	+	+	+
OYAVB169.1	+	+	+	+
S25	+	+	+	+
S34	+	+	+	+
S36	+	+	+	+
VSX11	+	+	+	+
VSX26	+	+	+	+
VSX36	+	+	+	+
R. leguminosarum 3841	+	+	+	+
Positive control	-	-	-	-
Negative control	-	-	-	-

Twelve of the fifteen test strains formed large and pink effective nodules on the primary or upper lateral roots as described in “Legume Inoculants and Their Use” (1984). The three strains that did not form nodules on any of the test plants were *R. phaseoli* ATCC 14482, *R. tropici* CIAT 899 and *R. etli* CFN42. This is unsurprising as the test plants are not their native host plants (Table 2.2).

There is no report in the literature claiming to have isolated strains of these three species from plants of the genus *Vicia, Pisum* or *Lathyrus*.

Similar observations on host-range or host-restriction have been reported before by Mutch and Young (2004) and Zézé *et al.* (2001).

2.3.2 Study of metabolic diversity using Biolog GN2 MicroPlates

The absorbance data of the Biolog plates from the ELISA reader were converted into binary format, with '1' representing the ability of a strain to utilize the substrate and '0' representing the inability of the strain to utilize the substrate.

The utilisation of the substrate was visually verified by checking for the presence of reduced tetrazolium dye precipitate at the bottom of the microtitre plate wells since exopolysaccharide formed by the bacteria increased absorbance values.

Figure 2.2. shows the appearance of the Biolog GN2 Microplate results for the six type strains after incubation for 48 hours and Table 2.3. shows the variation in the utilisation of sugars in the six type strains of *Rhizobium* used in the study.

The results of the Biolog analysis showed that a number of substrates were not utilized by many test strains. For example, from Table 2.3 it is seen that even the utilization of common sugars included in the Biolog GN2 MicroPlate varied considerably. *R. pis*i DSM30132 and *R. fabae* CCBAU33202 can use most sugar substrates in the GN2 plate whereas *R. leguminosarum* USDA2370*, R. tropici* CIAT899 and *R. etli* CFN42 can use several common sugars like D-Fructose and D-Glucose. However, *R. phaseoli* ATCC 14482 is different from others since it cannot utilize any of the common sugars.

The results of *R. phaseoli* ATCC 14482 prompted an investigation into the genetic constitution of the strain to search for an explanation into its inability to utilize the common sugar substrates. Using the KEGG pathway and PathComp tools, putative pathways for sugar catabolism were identified. The genes corresponding to the enzymes involved in these pathways were located in *Rlv*. 3841 and the sequences of these genes were used to identify the presence of homologous genes in *R.phaseoli* ATCC 14482. An example of such a putative pathway for glucose utilization is shown in Figure 2.3.

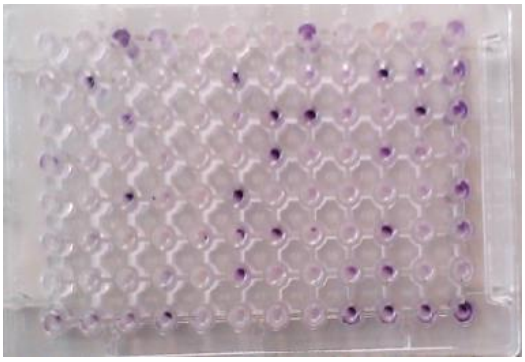
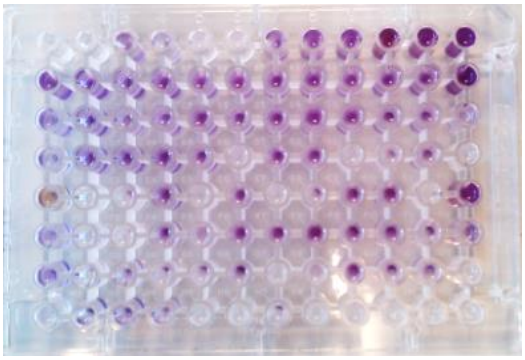
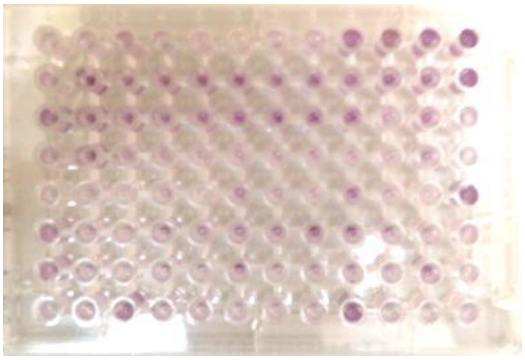
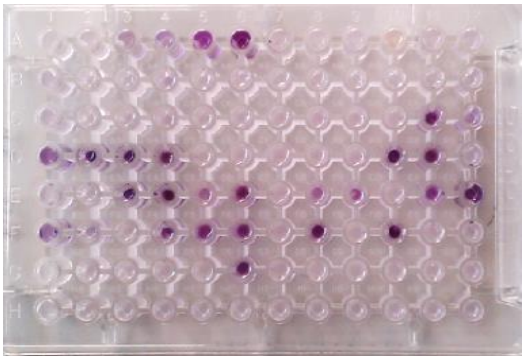
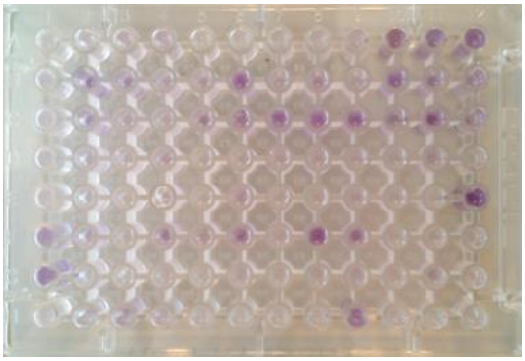
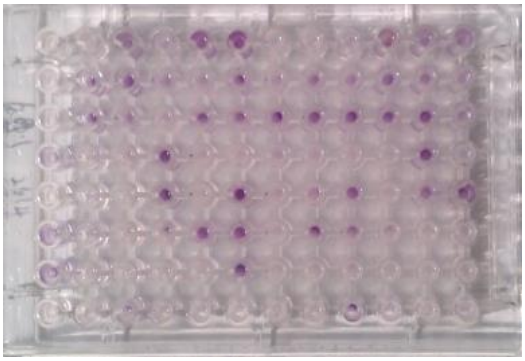
R. leguminosarum	*R. pisi*
	
R. fabae	*R. phaseoli*
	
R. tropici	*R. etli*
	

Figure 2.2. Appearance of Biolog plates after inoculating and culture for 48 hours

Table 2.3. Biolog sugar utilization test results (For substrate legends see Figure 2.1 (Page 42))

	A7	A8	A9	A10	A11	A12	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
R. leg USDA2370	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	1	1	1	0	0	1	0	0	0	1	1	0	0
R. pisi DSM30132	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
R. fabae CCBAU33202	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
R. phaseoli ATCC14482	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R. tropici CIAT899	0	0	0	1	1	1	0	1	1	0	0	1	0	1	0	1	1	1	0	1	0	0	1	1	1	1	1	1
R. etli CFN42	0	0	0	0	1	0	0	1	1	0	0	1	0	0	0	0	1	1	0	1	0	0	1	1	1	1	1	1
Rlv. 3841	0	0	1	1	1	1	1	1	1	0	0	1	0	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1
S25	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	0
S34	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	0	1	1	1	1	1	0
S36	0	0	1	1	1	1	1	1	1	1	0	1	0	1	0	1	1	0	0	1	1	0	1	1	1	1	1	0
OYAVB169.1	0	0	1	0	1	1	0	1	1	0	0	1	0	1	0	0	1	0	0	0	1	0	1	1	1	1	1	0
OYAVB296.5	0	0	1	1	1	1	0	1	1	0	0	1	0	1	0	1	1	0	0	1	1	0	1	1	1	1	1	0
VSX11	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	0	1	1	1	1	1	1
VSX26	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
VSX36	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

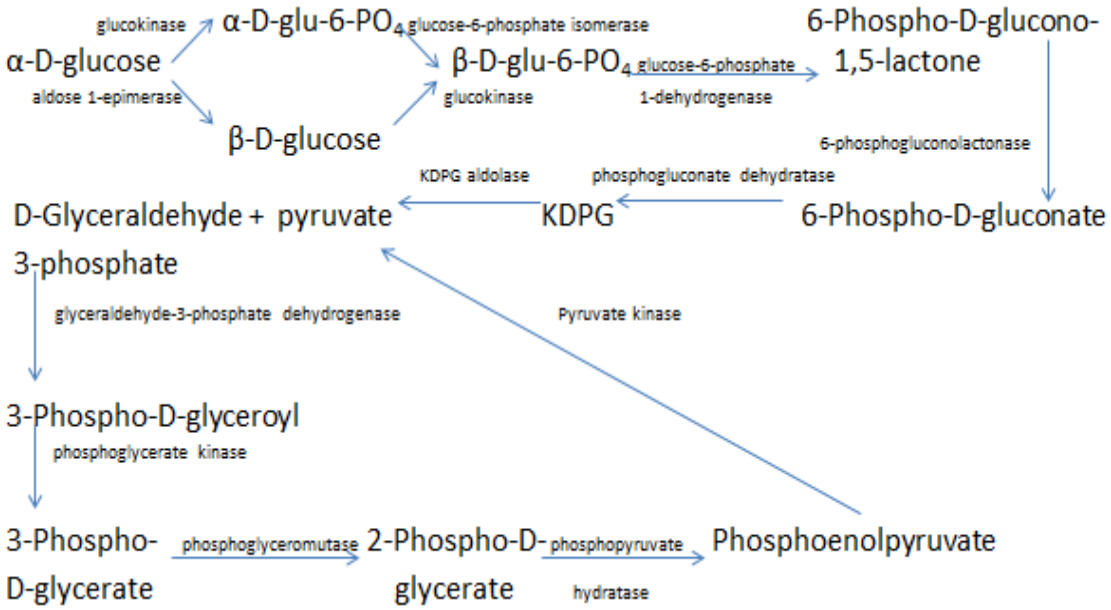


Figure 2.3. Putative pathway for glucose metabolism to pyruvate in *Rlv*. 3841.

The genes encoding the enzymes involved in metabolism of glucose in *Rlv*. 3841 have been characterised. The information about the genes predicted to be involved in glucose utilization in *Rlv*. 3841 according to the putative pathway is listed in Table 2.4. Some of the enzymes are coded by more than one gene. The data was obtained from NCBI-BLASTn search.

The gene sequences of the genes listed in Table 2.4 were used to search for homologous genes in the test strains for which complete or partial genome data was available. It was observed that some strains carried all the genes involved in the metabolic pathway of the substrates. Other strains lacked one or more genes required for catabolism of the substrate. For example, all the genes required for utilization of glucose were present in CFN42, WSM1325, WSM2304, VSX11, VSX26, VSX36, OYAVB169.1 and OYAVB 296.5.

All the strains tested using the Biolog plates could utilize glucose, except *R. phaseoli* ATCC 14482 (Table 2.3). The results of the BLAST search for the presence of glucose utilization genes showed that the fully sequenced strains - *R.etli* CFN42^{T}, *R. leguminosarum* WSM1325 and WSM2304 have genes for encoding all the enzymes of the putative pathway for glucose utilization. The field strains VSX11, VSX26, VSX36 from Wentworth and OYAVB169.1 and OYAVB 296.5 from Sweden were also found to have all the genes for glucose metabolism. Genome data for the three Scottish strains was not available and hence they could not be tested for the presence of genes (Table 2.5).

Table 2.4. List of genes encoding enzymes involved in the putative pathway for glucose utilization along with its relevant details.

Gene/Enzyme Description	Symbol	GeneID	Locus Tag	From	To	Product	EC Number	Reaction Number	Substrate (C Number) KEGG Database	Product (C Number) KEGG Database
glucokinase	glk	4400349	RL0182	217115	218140	YP_765787.1/ YP_764551.1	2.7.1.2	R01786	alpha-D-Glucose (C00267)	alpha-D-Glucose 6-phosphate (C00668)
glucose-6-phosphate isomerase	pgi	4402124	RL0504	541877	543559	YP_766113.1	5.3.1.9	R02739	alpha-D-Glucose 6-phosphate (C00668)	beta-D-Glucose 6-phosphate (C01172)
aldose 1-epimerase	galM	4401803	RL4605	4893131	4894141	YP_770169.1	5.1.3.3	R01602	alpha-D-Glucose (C00267)	beta-D-Glucose (C00221)
glucose-6-phosphate 1-dehydrogenase	zwf	4401982	RL0753	803477	804952	YP_766363.1	1.1.1.49	R02736	beta-D-Glucose 6-phosphate (C01172)	D-Glucono-1,5-lactone 6-phosphate (C01236)
glucose-6-phosphate 1-dehydrogenase	zwf	4401983	RL1315	1378039	1379547	YP_766921.1	1.1.1.49	R02736	beta-D-Glucose 6-phosphate (C01172)	D-Glucono-1,5-lactone 6-phosphate (C01236)
glucose-6-phosphate 1-dehydrogenase	pRL120561	4398348	pRL120561	611772	613238	YP_765066.1	1.1.1.49	R02736	beta-D-Glucose 6-phosphate (C01172)	D-Glucono-1,5-lactone 6-phosphate (C01236)
6-phosphogluconolactonase	pgl	4402126	RL0752	802768	803466	YP_766362.1	3.1.1.31	R02035	D-Glucono-1,5-lactone 6-phosphate (C01236)	6-Phospho-D-gluconate (C00345)
6-phosphogluconate dehydratase	edd	4401542	RL0751	800868	802691	YP_766361.1	4.2.1.12	R02036	6-Phospho-D-gluconate (C00345)	2-Dehydro-3-deoxy-6-phospho-D-gluconate (C04442)
keto-hydroxyglutarate-aldolase/keto-deoxy-phosphogluconate aldolase	eda	4401541	RL4162	4410279	4410917	YP_769737.1	4.1.2.14	R05605	2-Dehydro-3-deoxy-6-phospho-D-gluconate (C04442)	D-Glyceraldehyde 3-phosphate (C00118) pyruvate (C00022)

glyceraldehyde-3-phosphate dehydrogenase	gap	4401804	RL4007	4234629	4235639	YP_769584.1	1.2.1.12	R01061	D-Glyceraldehyde 3-phosphate (C00118)	3-Phospho-D-glyceroyl phosphate (C00236)
phosphoglycerate kinase	pgk	4402125	RL4011	4239030	4240235	YP_769588.1	2.7.2.3	R01512	3-Phospho-D-glyceroyl phosphate (C00236)	3-Phospho-D-glycerate (C00197)
phosphoglyceromutase	gpmA	4402944	RL0179	213683	214318	YP_765784.1	5.4.2.1	R01518	3-Phospho-D-glycerate (C00197)	2-Phospho-D-glycerate (C00631)
phosphoglyceromutase	gpmB	4402945	RL0954	1028003	1028584	YP_766566.1	5.4.2.1	R01518	3-Phospho-D-glycerate (C00197)	2-Phospho-D-glycerate (C00631)
phosphoglyceromutase	gpmB	4402946	RL1010	1093702	1094292	YP_766621.1	5.4.2.1	R01518	3-Phospho-D-glycerate (C00197)	2-Phospho-D-glycerate (C00631)
phosphoglyceromutase	RL2665	4400179	RL2665	2813734	2814483	YP_768249.1	5.4.2.1	R01518	3-Phospho-D-glycerate (C00197)	2-Phospho-D-glycerate (C00631)
phosphoglyceromutase	RL2997	4399156	RL2997	3161952	3162581	YP_768580.1	5.4.2.1	R01518	3-Phospho-D-glycerate (C00197)	2-Phospho-D-glycerate (C00631)
phosphoglyceromutase	RL3898	4403643	RL3898	4126413	4126967	YP_769476.1	5.4.2.1	R01518	3-Phospho-D-glycerate (C00197)	2-Phospho-D-glycerate (C00631)
phosphopyruvate hydratase	eno	4401545	RL2239	2360835	2362109	YP_767833.1	4.2.1.1	R00658	2-Phospho-D-glycerate (C00631)	Phosphoenolpyruvate (C00074)
pyruvate kinase	pykA	4399238	RL4060	4288486	4289925	YP_769637.1	2.7.1.40	R00200 R01138 R00430 R01858 R02320	Phosphoenolpyruvate (C00074)	Pyruvate (C00022)

Table 2.5. Presence of putative enzymes of glucose metabolism in test strains analysed using BLAST and gene sequence of *Rlv*. 3841 as query sequences.

Enzyme	*R. leguminosarum* USDA 2370	*R. pisi* DSM30132	*R. fabae* CCBAU 33202	*R. phaseoli* ATCC 14482
Glucose Utilization	Yes	Yes	Yes	No
glucokinase	+	+	+	+
aldose 1-epimerase	+	+	+	+
glucose-6-phosphate isomerase	+	-	+	-
glucose-6-phosphate 1-dehydrogenase	+	+	+	+
6-phospho gluconolactonase	-	-	-	-
phosphogluconate dehydratase	+	-	+	+
KDPG aldolase	+	-	+	+
glyceraldehyde-3-phosphate dehydrogenase	+	+	-	-
phosphoglycerate kinase	+	+	+	+
phospho-glyceromutase	+	+	+	+
phosphopyruvate hydratase	+	+	+	+
pyruvate kinase	+	-	+	+

Key (+) = gene present, (-) = gene absent

The genes for which partial sequence data is available were found to be lacking some of the genes involved with glucose metabolism. The fact that the strains can still utilize glucose may reflect upon the quality of the

sequence data which could have returned some false negative results. In the case of *R. phaseoli* ATCC 14482 the inability to use glucose as a carbon source could possibly be attributed to the missing genes in its genome. However, the results would have been more convincing if the genome data were more complete with better sequence coverage. *R. leguminosarum* USDA 2370, *R. fabae* CCBAU 33202 and *R. phaseoli* ATCC 14482 had a sequence coverage value of 3.9X while *R. pisi* DSM30132 had a sequence coverage value of 5X

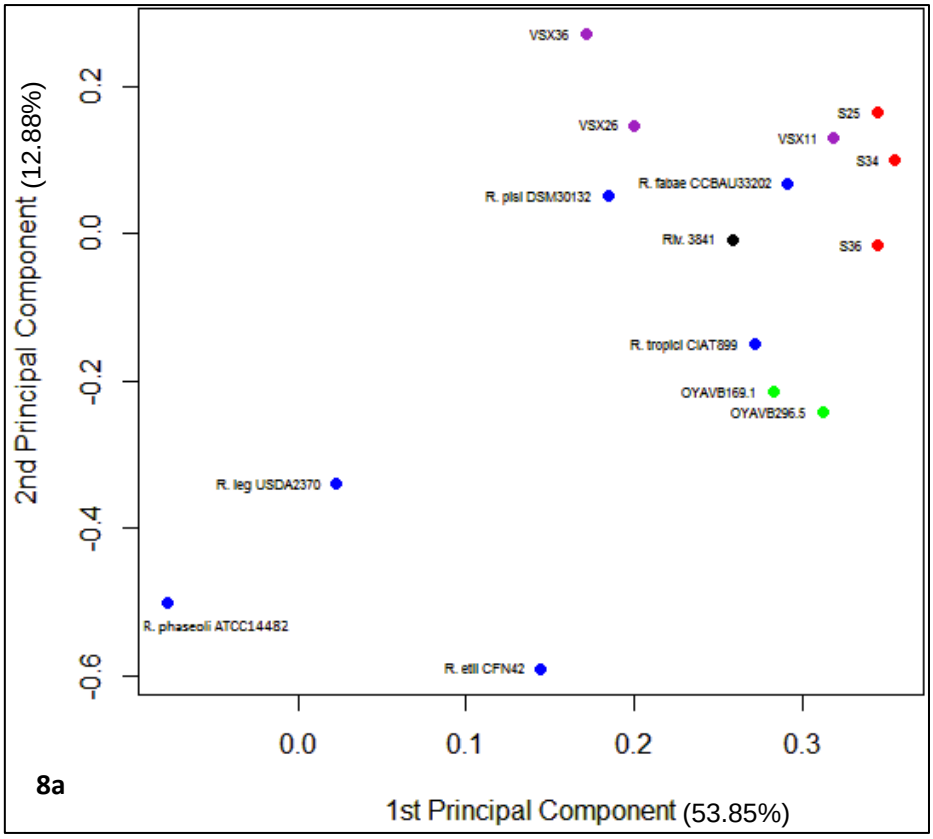
2.3.3 Analysis of Biolog results

The results of the Biolog study were analysed to search for patterns in metabolism of carbon substrates. The analysis was carried out using the data from the entire Biolog GN2 MicroPlate. Since the strains showed a good amount of variation in the ability to utilize sugars, a second investigation was carried out using only the data from the metabolism of sugars.

Two types of analysis were performed viz. Principal Coordinate Analysis and clustering of strains using UPGMA. The results of the analysis were compared with the clustering of strains in a network tree constructed using the *celC* gene sequence from the test strains.

2.3.3.1. Clustering of strains using Principal Co-ordinates Analysis

Figure 2.4. shows the clustering of strains using Principal Co-ordinate Analysis using the Biolog GN2 MicroPlate data. PCoA shows *R.pisi* DSM 30132 and *R.fabae* CCBAU 33202 have more common preference for substrates than others. Moreover, *R.phaseoli* ATCC 14482 and *R.etli* CFN 42 are more similar based on substrates.



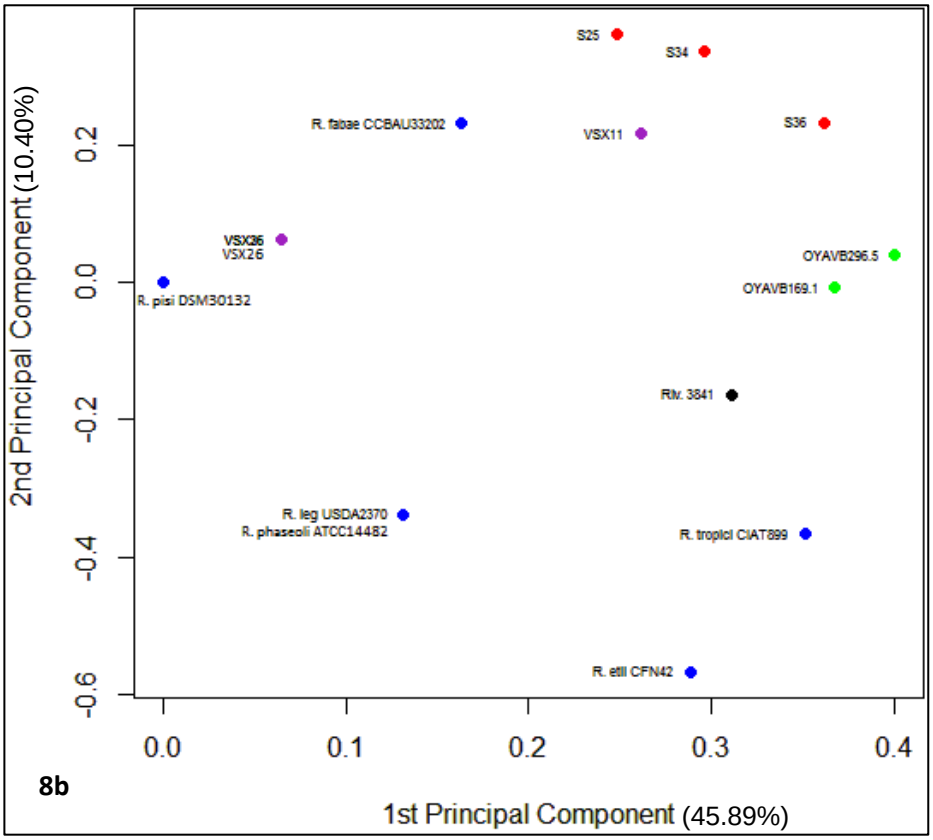


Figure 2.4. Principal Coordinate Analysis of Biolog data for the strains using (a) data from the metabolism of all substrates and (b) data from the metabolism of sugar substrates.

2.3.3.2. Clustering of strains using UPGMA

Figure 2.5. shows the clustering of strains using UPGMA from the Biolog GN2 MicroPlate data.

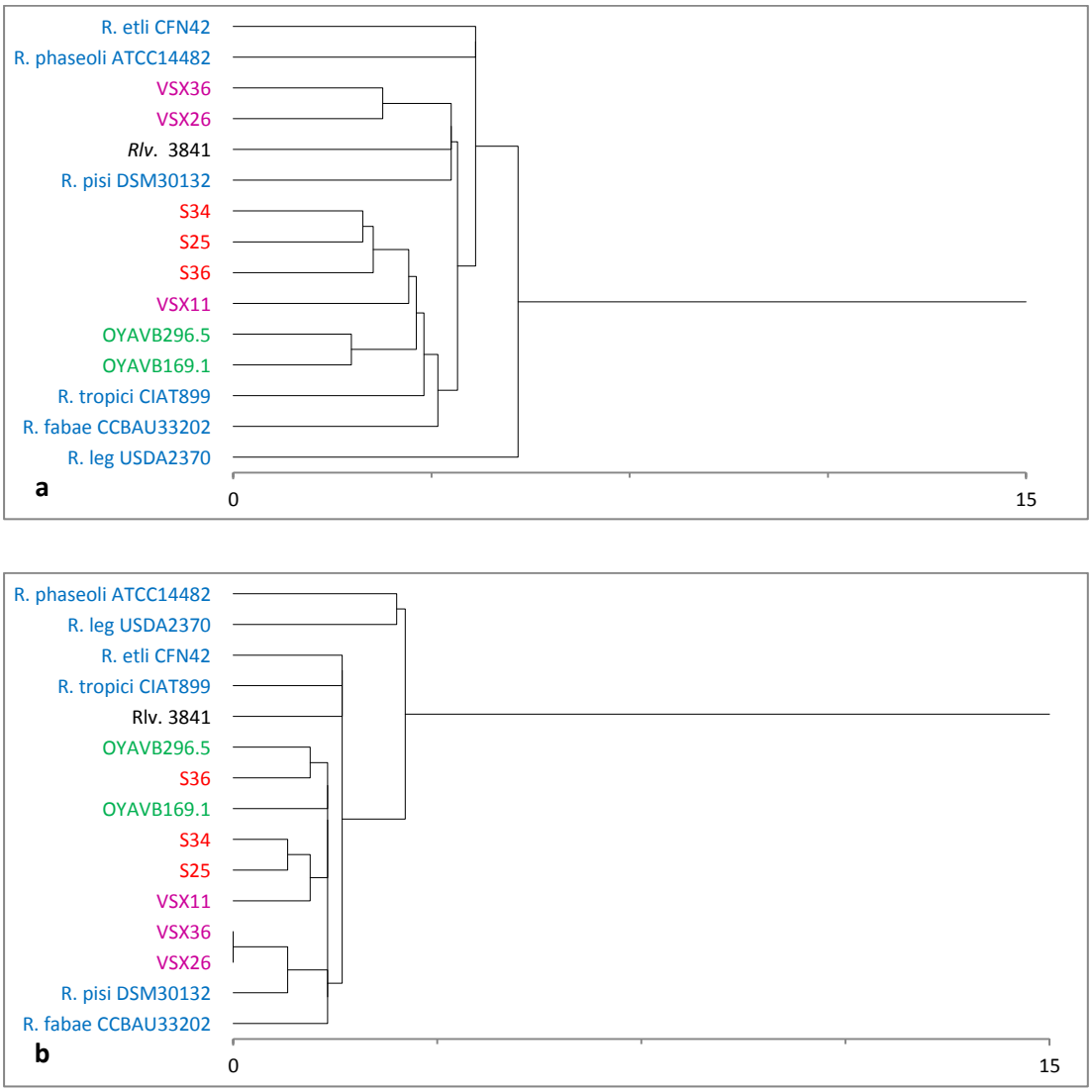


Figure 2.5. UPGMA clustering of strains using Biolog data from (a) data from the metabolism of all substrates and (b) data from the metabolism of sugar substrates.

2.3.3.3. Network tree of the *celC* gene :

Robledo *et al.* (2011) recommended the use of *celC* gene as a new phylogenetic marker for the classification of rhizobia strains. In order to order to see similarities and differences between the clustering of strains using Biolog data with PCoA / UPGMA and phylogenetic clustering, a *celC* network tree was constructed. Comparisons were drawn from the clustering patterns obtained from these analyses.

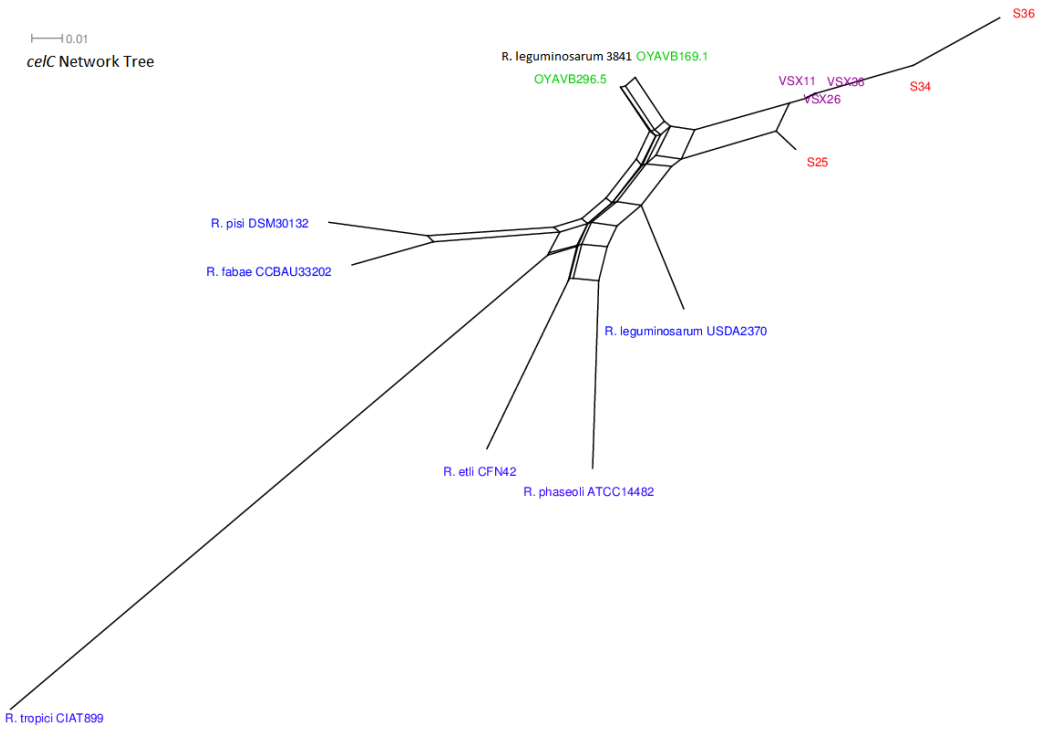


Figure 2.6. Phylogenetic-network tree based on *celC* gene sequences of strains. Type strains (blue) Swedish strains (green), Scottish strains (red), VSX field strains (purple), *Rlv*. 3841 reference strain (black).

On comparing the clustering patterns from the Biolog data and the *celC* data, it was seen that the strains *R. pisi* DSM30132 and *R. fabae* CCBAU 33202 are closely related in both the analyses. Similarly, the strains *R. phaseoli* ATCC 14482 and *R. etli* CFN42 are closer to each other than others in both the analyses.

R. tropici CIAT 899 is distant from any of the other species in sugar metabolism in PCoA analysis and in the *celC* tree. The strain *R. leguminosarum* USDA 2370 strain more closely resembles *R. phaseoli* ATCC 14482 and *R. etli* CFN42 than *R. leguminosarum* biovar *viciae* strain 3841 which, in fact clusters intimately with the Swedish strains.

2.4. Discussion

The strains to be used in the project were studied for their ability to nodulate their native and closely-related host plants. This was essential as it would give an idea of their host-range / host-specificity. This would also help in validating the strains as nodulation is an important criterion (rather, the only criterion) for validation of rhizobium isolates.

Twelve of the fifteen test strains formed large, pink, effective nodules on the primary or upper lateral roots as described in “Legume Inoculants and Their Use” (1984). The three strains that did not form nodules on any of the test plants were *R. phaseoli* ATCC 14482, *R. tropici* CIAT 899 and *R. etli* CFN42. This is not surprising as the test plants are not native host plants for these three strains and hence were not expected to form nodules on the selected host plants. There is no report in literature claiming to isolate any of these three species from plants belonging to the genera *Vicia, Pisum* or *Lathyrus* and hence the results can be said to follow the current trends. The observations are in line with the observations of Mutch and Young (2004) who had performed similar cross-nodulation tests and found that most biovar *viciae* isolates from the wild can form effective nitrogen-fixing nodules on these four plants (Mutch and Young, 2004, Zézé *et al.*, 2001). All the bacterial strains belonging to biovar *viciae* which are to be used in the later part of the study have retained their ability to nodulate their native plant species. The validated isolates were now ready for use in the study.

Using the results of the Biolog data in conjunction with the BLAST search results for the genes of the enzymes involved in the metabolism of sugars it was seen that a number of strains were not able to utilize the sugars because they lacked the key enzymes of the metabolic pathways. Using the example of glucose, it was observed that the four fully sequenced strains *R. etli* CFN42^T, *R. leguminosarum* 3841, *R. leguminosarum* WSM1325 and WSM2304 possessed all the enzymes required for the catabolism of glucose. The BLAST results obtained from the other strains for which the genome data is not complete, it was found that although the strains were able to utilize glucose, they lacked some or many of the enzymes required for metabolising glucose, for example *R. leguminosarum* USDA2370^T, *R. pisi* DSM30132^T and *R. fabae* CCBAU33202^T were expected to have genes coding for the enzyme 6-phosphogluconolactonase, which is an essential enzyme in the pathway from glucose to pyruvate, but in the BLAST results all the above strains were found to be lacking in the gene encoding this enzyme. A similar situation was seen in *R. fabae* CCBAU33202^T for the gene encoding glyceraldehyde-3-phosphate dehydrogenase. The incomplete nature of the sequence data (or rather not being assembled into complete genome in terms of replicons) makes it difficult to explain the results of the Biolog test. It could be possible that the genes are present in the strain but because of poor coverage in those regions of the genome (~3X - 5X), the BLAST search was not able to detect it and returned a false negative result. Alternately, it is also

possible that even though the sequence data is not assembled it is still reliable for BLAST search and that the BLAST results obtained are reliable. In such a scenario, the presence of isozymes (different enzymes catalysing the same reaction) cannot be ruled out. It is also possible that the metabolism of the sugar proceeds through a completely different pathway requiring a completely different set of enzymes.

An interesting strain that stands out from the others in terms of sugar metabolism is *R. phaseoli* ATCC 14482. This strain was unable to metabolise any of the common sugars incorporated in the Biolog GN2 MicroPlate. The strain has also been noted to be poor in metabolizing pyruvate (Ramirez-Bahena *et al.*, 2008). BLAST search for the genes involved in sugar metabolism indicate that some genes required for the metabolism of glucose and other common sugars were missing. For example, 6-phosphogluconolactonase and glyceraldehyde-3-phosphate dehydrogenase, both essential in glucose metabolism, were absent in the BLAST results for *R. phaseoli* ATCC 14482^{T}. Whilst this could be a possible reason for the inability to utilize sugars, the incomplete assembly of the sequence data once again casts doubt on the BLAST results. Hence, it would make an interesting project to study the metabolism of this strain in depth and investigate the abnormal metabolic pattern of this strain.

The analysis of the Biolog results using PCoA and UPGMA to cluster data and comparison of this clustering with the clustering of strains using the phylogenetic network tree obtained using the sequence data of the *celC* gene shows some striking similarities in the clustering. The strains *R. pisi* DSM30132 and *R. fabae* CCBAU 33202 are closely related in both the analyses.

Similarly, the strains *R. phaseoli* ATCC 14482 and *R. etli* CFN42 are closer to each other than others in both the analyses. *R. tropici* CIAT 899 is distant from any of the other species in sugar metabolism in PCoA analysis and in the *celC* tree. The strain *R. leguminosarum* USDA 2370 more closely resembles *R. phaseoli* ATCC 14482 and *R. etli* CFN42 than *R. leguminosarum* biovar *viciae* strain 3841 which, in fact, clusters intimately with the Swedish strains.

To conclude, we can say that the strains used in the study have retained their host specificity. Their metabolic ability is diverse, yet there are patterns in the diversity. The following chapters in the thesis will focus on the differences between two biovars of *R. leguminosarum*, viz. biovar *viciae* and biovar *trifolii*.

CHAPTER 3 : DISTRIBUTION AND PHYLOGENETIC RELATIONSHIPS BETWEEN THE *bvs* GENES IN RHIZOBIA.

3.1. Introduction

The study of the host-specificity and metabolic differences amongst the members of the *R. leguminosarum* species complex in Chapter 2 showed that the members of the group vary considerably in their phenotypes. No clear distinction was seen between the members included in the species complex using results from metabolic fingerprinting or from nodulation tests.

It is known that the ability of rhizobia to nodulate specific host plants is dictated by a set of genes called the *nod* or nodulation genes (located on the plasmid pRL10 in *Rlv*. 3841). The *nod* genes from different species differ in their sequence and these differences can be used to differentiate or identify different symbiovars of the *R. leguminosarum* species complex. Such an approach to link differences in genotype to the ability of the bacteria to establish symbiosis with specific host plants has been used in a study by Mutch and Young (2004) who showed that isolates from *Vicia faba*, which had a distinct *nodD* sequence, could nodulate a large number of other host plants that they tested. The host-range of some wild type strains whose genome sequences are known have been discussed by Rogel *et al.* (2011).

We can speculate that, like the *nod* genes, there may be other genes that are exclusive to a species or a biovar. Hence, a comparative genomics approach could be used to search for differences between the members of the species complex. Such an approach was used in the analysis of genome sequences of strains from two biovars of *R. leguminosarum*. The analysis showed that there were five genes that were found exclusively in the biovar *viciae* isolates. None of the biovar *trifolii* isolates had these genes. Since these genes were specifically found in the biovar *viciae*, they were referred to 'biovar *viciae* specific' genes or *bvs* genes.

In *Rlv*. 3841, the five *bvs* genes are found on the plasmid pRL8, the smallest plasmid in the strain (147,463 bp). Based on findings of sequence similarity, the five genes have been assigned putative functions. The arrangement of the five *bvs* genes on pRL8 and their direction of transcription are shown in Figure 3.1.

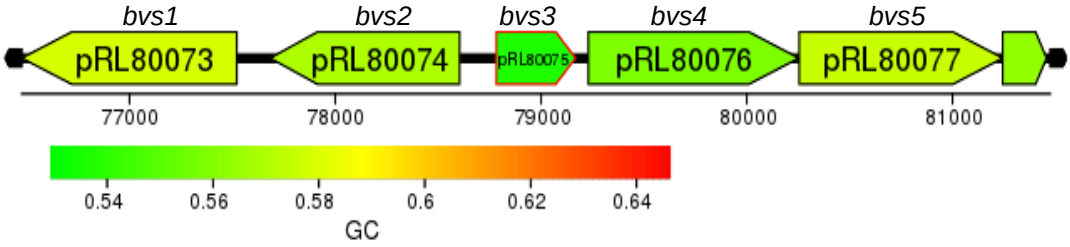


Figure 3.1. The arrangement of *bvs* genes on pRL8 in *Rlv*. 3841.

The putative functions assigned to the five *bvs* genes are as follows :

3.1.1. *bvs1* (pRL80073) : Putative cysteine desulfurase (1206 bp)

The enzyme cysteine desulfurase catalyses transfer of sulfur from L-cysteine to acceptor molecules. The enzyme forms a transient intermediate to form sulfane sulfur from which the sulfur is then transferred to the acceptor molecule. In some organisms like *A. vinelandii*, the enzyme can act on cysteine as well as selenocysteine to form alanine and precipitate elemental sulfur or selenium respectively (Mihara and Esaki, 2002, Mihara *et al.*, 1997).

In some rhizobial strains, cysteine desulfurase is encoded by *nifS* gene and increases the activities of both component proteins of nitrogenase, a complex metalloenzyme that contain iron and sulfur. Thus cysteine desulfurase, the product of *nifS* gene might play a key role in providing those two elements for metallocluster formation (Ribbe *et al.*, 2013, Zheng *et al.*, 1993). Nogales *et al.* (2002) found one mutant of *Rhizobium tropici* CIAT 899 that was less tolerant to saline and hyperosmotic media and the wild type of this strain carried a gene homologous to *nifS.*

There is a *nif*S gene on the R*lv* 3841 chromosome, whereas *bvs*1 is located on pRL8. Two genes may have same function without clearly defined roles. However, it is possible that the *NifS* gene may be functional in metabolism while *bvs*1 may have a role in symbiosis.

3.1.2. *bvs2* (pRL80074) : LysR family transcriptional regulator (921 bp)

The LysR-Type Transcriptional Regulators (LTTRs) constitutes diverse family of oligomeric bacterial transcription factors first described by Henikoff *et al.* (1988). The family is perhaps the largest and the most common family of transcriptional regulators among prokaryotes with over 40,000 potential members (Knapp and Hu, 2010). The conserved structure of this family is an

N-terminal DNA-binding helix–turn–helix motif and a C-terminal co-inducer-binding domain (Cai *et al.*, 2009). Although the structures and functions of this family are conserved, LTTRs are involved in regulation of a wide range of genes *e.g.* virulence, metabolism, quorum sensing and motility (Maddocks and Oyston, 2008). The *bvs2* gene has been annotated as a LysR family transcriptional regulator, but the function has not been experimentally verified and the genes controlled by it are not known.

3.1.3. *bvs3* (pRL80075) : Putative endoribonuclease L-PSP (384 bp)

Putative endoribonuclease L-PSP (liver perchloric acid soluble protein) is an enzyme that belongs to a large family of proteins present in bacteria, archaea, and eukaryotes. Oka *et al.* (1995) first reported the presence of a 136-amino acid perchloric acid-soluble (L-PSP) protein from rat-liver which inhibited protein synthesis. Because of its kinetics, they thought that it inhibited protein synthesis by inhibiting initiation of protein synthesis. Later studies showed that the protein exhibits endoribonuclease activity on single-stranded mRNA. Hence, it is now assumed that the enzyme may be involved in the inhibition of protein synthesis by cleavage of mRNA (Morishita *et al.*, 1999).

Homology studies of the protein shows that it belongs to a family of small proteins with presently unknown function, named YER057c/YjgF/UK114 (Schmiedeknecht *et al.*, 1996). Members of the YjgF family (renamed RidA) are enamine/imine deaminases. They hydrolyse reactive intermediates released by PLP-dependent enzymes, including threonine dehydratase (Lambrecht *et al.*, 2012). In *Bacillus subtilus*, the enzyme has been shown to be involved in the regulation of purine synthesis (Rappu *et al.*, 1999). YjgF also prevents inhibition of transaminase B (IlvE) in *Salmonella enterica* serovar *typhimurium* (Christopherson *et al.*, 2012, Schmitz and Downs, 2004). However, the function of the enzyme in in rhizobia remains unknown.

3.1.4. *bvs4* (pRL80076) : Putative aliphatic nitrilase (1022 bp)

The catalytic function of nitrilase is to convert nitrile compounds to their corresponding carboxylic acids and ammonia. Recently the functions of nitrilases in biological processes have been studied in detail. Their contribution to the interactions between plants and microorganisms are considerable. Howden and Preston (2009) indicated plant-associated

microorganisms may produce nitrilases to convert plant nitriles to carbon and nitrogen sources that they can use. Spaepen *et al.* (2007) studied bacteria that hydrolyse arylacetonitriles to produce auxin phytohormone indole-3-acetic acid, which promote the growth of plants.

In *Rhizobium* the *bvs* 4 gene may encode a nitrilase that influences the symbiosis since these enzymes may relate to interaction between host plants and rhizobia. These enzymes may contribute to defence, detoxification, nitrogen utilization and plant hormone synthesis. Some of them have been shown to increase tolerance of cyanide and nitrile. Sorghum plants store the cyanogenic glycoside dhurrin which is thought to act as a defence compound (Legras *et al.*, 1990). In *Trifolium*, cyanogenesis had been reported by Conn (1979), who found cyanogenic glycosides could be synthesised by white clover and used by plants as defence against herbivores because of its toxicity. Furthermore, Collinge and Hughes (1982) studied the pathway of cyanogenesis in white clover during tissue damage and observed that 2-hydroxy-2-methylpropanenitrile and 2-hydroxy-2-methylbutanenitrile were synthesised by the plant as the aglycone precursors of cyanogenic glucosides, linamarin and lotaustralin. Olsen *et al.* (2008) studied the genes of cyanogenesis and molecular basis of cyanogenic polymorphism in white clover. Their findings suggested that the allotetraploid clover evolved from cyanogenic and acyanogenic diploid progenitors. Similar study in multiple species by Olsen *et al.* (2014) provided evidence for parallel evolution of adaptive biochemical polymorphisms through recurrent gene deletions. *P. fluorescens* SBW25 forms symbiosis with plants and promote its growth. The bacterium produces β-cyano-L-alanine nitrilase that enables it to tolerate toxic concentrations of nitriles (Howden *et al.*, 2009). *R. leguminosarum* 3841 has a chromosomal nitrilase gene in addition to one on pRL8. It would be interesting to test for functional differences between the two nitrilase genes.

3.1.5. *bvs5* (pRL80077) : Putative Mo-binding oxidoreductase (990 bp)

The *bvs5* gene encodes a putative enzyme that belongs to the group of Molybdenum-binding oxidoreductases. More than 50 molybdoenzymes have been identified in bacteria. They are involved in a wide variety of transformation by enzymes of carbon, sulfur and nitrogen metabolism (Mendel and Bittner, 2006). Main members of this family are nitrate reductase

(NR) and sulfite oxidase (SO). SO catalyses the terminal reaction in oxidative degradation of the sulfur-containing amino acids. Assimilatory NRs catalyse the reduction of nitrate to nitrite. Known members of this family possess a single pterin cofactor and a cysteine ligand that catalyse transfer of oxygen to or from a lone pair of electrons on the substrate (Marchler-Bauer *et al.*, 2013).

Amongst the Wentworth isolates, the *bvs* genes were found only in the isolates from *Vicia sativa* i.e. biovar *viciae* strains. This biovar-specific distribution of genes posed a question – are these genes restricted to biovar *viciae* isolates from Wentworth only or are the genes found in biovar *viciae* isolates from other locations as well; while being excluded from other biovars of rhizobia.

The distribution and phylogenetic relatedness of the *bvs* genes from type strains and field strains of *Rhizobium* belonging to different species and biovars is investigated in this part of the work. Based on the BLAST results in Wentworth strains, it is expected that all five *bvs* genes occur widely in biovar *viciae* strains but are rare in other biovars. Phylogenetic trees based on *bvs* genes are expected to show a clustering of rhizobial strains and isolates from same or similar host plants.

3.2. Materials and Methods

3.2.1. Test strains :

Six type strains of *Rhizobium* from different species and biovars, eight field strains of *Rhizobium leguminosarum* biovar *viciae* and one reference strain (*Rlv*. 3841) were used to study the distribution of *bvs* genes. The type strains and field strains used in the study are listed in Tables 2.1 and 2.2 along with their relevant details (Pages 45 and 46 respectively).

3.2.2. Investigating the distribution of *bvs* genes

The distribution and diversity of *bvs* gene homologs in different sets of rhizobial strains was investigated. The sequence of *bvs* genes from the fully sequenced reference strain *Rlv*. 3841 (Young *et al.*, 2006) was used as the query sequence for the investigations. Two approaches were used to detect the presence of the genes in the population viz. BLAST (Basic Local Alignment Search Tool) and PCR (Polymerase Chain Reaction).

3.2.2.1. Basic Local Alignment Search Tool (BLAST)

BLAST is a sophisticated, single most important software package in bioinformatics. It is a powerful, fast, reliable and flexible tool used for searching regions of similarity between sequences. The program was developed by Altschul *et al.* (1990) at NCBI (National Center for Biotechnology Information, part of the National Institutes of Health, a U.S. government-funded center for the curation and presentation of public biological knowledge. The BLAST program can be used to compare nucleotide or protein sequences to sequence databases and calculate the statistical significance of matches. BLAST can also be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

BLAST can be used in five different ways : using the BLAST site on NCBI (http://blast.ncbi.nlm.nih.gov/), as an Application Programming Interface (API) using HTTP-encoded requests directed to the NCBI cgi-bin program (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) on the NCBI web server, as a BLAST standalone application, as the BLAST+ remote service application or using the C++ Application Programming Interface (API). In this study, the NCBI-BLAST and BLAST standalone applications were used.

The BLAST searches can be grouped into two main categories : the basic BLAST search and the specialised BLAST search. There are five types of basic BLAST searches viz. nucleotide BLAST (BLASTn, to search nucleotide database using a nucleotide query set), protein BLAST (BLASTp, to search protein database using a protein query set), blastx (to search protein database using a translated nucleotide query), tblastn (to search translated nucleotide database using a protein query) and tblastx (to search translated nucleotide database using a translated nucleotide query). Based on the algorithm used there are many different sub-types of nucleotide BLAST and protein BLAST searches. There are many specialised BLAST search which can be used for different purposes like finding conserved domains in sequences, sequence alignments, primer designing for PCR etc. (http://blast.ncbi.nlm.nih.gov/Blast.cgi). In this part of the study, the BLASTn and BLASTp algorithms were used for searching the nucleotide and protein databases.

The BLAST tool was used to search for presence of *bvs* genes in the type strains and field strains listed in Tables 2.1 and 2.2. The gene sequence of the five *bvs* genes from *Rlv*. 3841 were used as the query sequence for the nucleotide BLAST (BLASTn) search. The information for the five *bvs* genes, including sequence, size, direction of transcription and putative function were downloaded from *Rlv*.3841 genome data (NCBI) as FASTA files. The genome data of *Rhizobium etli* CFN42^T was also saved as FASTA files from NCBI database (Segovia *et al.*, 1993). Genome data of other rhizobia strains have been sequenced (partially or entirely) were also stored in the University of York BLAST server database (http://blast.york.ac.uk/blast/blast.html). The two other closely related rhizobia that have been fully sequenced are *R. leguminosarum* bv. *trifolii* WSM1325 and *R. leguminosarum* bv *trifolii* WSM2304. Their genome information was obtained from Genbank. The genome data from these two strains was also included in Local-BLAST searches (Reeve *et al.*, 2010a, Reeve *et al.*, 2010b).

Protein-BLAST (BLASTp) was performed (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) using the amino acid sequence of the five *bvs* genes from *Rlv*. 3841 to search for homologous proteins in other species.

3.2.2.2. Polymerase Chain Reaction (PCR)

The PCR technique, developed by Mullis *et al.* (1987) uses a thermostable DNA polymerase (Saiki *et al.*, 1988) to synthesize a complementary strand on a single stranded DNA template. The DNA polymerase cannot synthesize DNA strand *de-novo* and requires a primer for extending its 3'-OH end. Because of this dependency, the PCR can be used to amplify specific regions of DNA whose ends are marked by the binding of sequence-specific primers. The result of this specific amplification is billions of copies or amplicons of the segment of DNA flanked or flanked on both the ends by the primers.

In addition to the BLAST search for the presence of genes in the type and field strains, the presence of genes was also verified using PCR. This was done since the sequence data was not available for all the strains under investigation.

- **Extraction of DNA for PCR :**

  A loopful of glycerol stock culture of the six test cultures, five field cultures and the reference strain (*Rlv*. 3841) was streak-plated onto a sterile TY agar. The plates were incubated at 28^{O}C for 48 hours. A single colony from each plate was transferred onto a new sterile TY agar plate which was incubated at 28^{O}C for 24 hours. A loopful of culture from this plate was transferred to 25 ml of sterile TY broth prewarmed to 28^{O}C in a 50 ml Corning polypropylene tube. The tube was briefly vortexed to disperse the inoculum evenly and then incubated at 28^{O}C on a rotary shaker for 24 hours. After incubation, the culture was centrifuged at 4000 rpm for 15 minutes at 20^{O}C and the resulting bacterial pellet was washed in two changes of physiological saline to remove traces of polysaccharide since it decreased the quality and quantity of the DNA extracted.

  The pellet of bacteria obtained after the final centrifugation was suspended in 200 µl of filter sterilized distilled water. This suspension was used to extract the DNA using the FastDNA® SPIN KIT from MP Biomedicals using the manufacturer's protocol.

  The quality of the DNA was checked by running the DNA on a gel and by running it on the NanoDrop from Thermo Scientific. The absorbance curve was studied and the 230:260 and 260:280 ratios determined to assess the purity of the DNA extracted. The extracted DNA was stored at -20^{O}C until further use.

- **Designing PCR primers for amplification of *bvs* genes**

  Primer pairs for each of the five genes were designed by Primer-BLAST on NCBI (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). Primer-BLAST was developed at NCBI to help users to design primers that are specific to an intended PCR target. It uses Primer3 (http://primer3.wi.mit.edu/) to design PCR primers and then uses BLAST and global alignment algorithm to screen primers against a user-selected database in order to avoid primer pairs (all combinations including forward-reverse primer pair, forward-forward as well as reverse-reverse pairs) that can cause non-specific amplifications (Ye *et al.*, 2012). The primer sequences are listed in Table 3.1.

Table 3.1. Sequence of primers designed to search *bvs* genes using PCR.

Primer name	Sequence (5'→3')	PCR product size	PCR product range on pRL8
bvs1 forward	ATTGACGCAGTGATTGGGTA	1045 bp	76351 - 77395
bvs1 reverse	CCTCAATTTCCTCCTCCGAA		
bvs2 forward	GATCATCTTGACTGGGACGA	667 bp	77923 - 78589
bvs2 reverse	ACATCATGCTGTTTGAGCTG		
bvs3 forward	AGTAGGACCATATTCGCACG	202 bp	78822 - 79023
bvs3 reverse	GGTGAGATAAACGTTGACGG		
bvs4 forward	GAAAAAGCTTGCTCGGTGAT	811 bp	79303 - 80113
bvs4 reverse	GAAATCGTTTCGGCTCAACA		
bvs5 forward	TAAATCGTGGATCACCCCTG	867 bp	80333 - 81199
*bvs*5 reverse	TCAGATACCGGCTGTTCTTC		

- **Composition of PCR mastermix for amplification of *bvs* genes**

  33µl of PCR mastermix was used in the amplification of the *bvs* genes containing 22.575 µl of PCR grade water, 1X FlexiGreen buffer, 0.25mM dNTP, 1.5mM $MgCl_2$, 0.4pM each of forward and reverse primer and 0.025u/µl Gotaq DNA polymerase in 200µl PCR tubes. To 33 µl of the mastermix, 2 µl of the DNA was added to bring the final volume to 35 µl.

  The PCR was carried out with an initial denaturation at 95^{O}C for 3 minutes; followed by 35 rounds of amplification consisting of denaturation at 94^{O}C for 1 minute, primer annealing at 57^{O}C for 1 minute and extension at 72^{O}C for 1 minute. A final extension elongation step was carried out at 72^{O}C for 10 minutes following which the DNA was stored at 4^{O}C.

  The PCR products were checked by electrophoresis. 5 µl of the PCR products was run on a 1.5% agarose gel with 0.1X SYBR® Safe DNA gel stain (Invitrogen). The gel was run at 120V for 30 minutes and visualised over blue light on a Safe Imager™ Blue Light Transilluminator (Life

Technologies Corporation). The gel images were captured using the gel documentation system and software from Ultra-Violet Products Ltd.

The results from the PCR of the five *bvs* genes were compared with results from the Local-BLAST search conducted on the strains for which the genome data were available in the genome database maintained at University of York.

3.2.3. Phylogeny of *bvs* genes

The products of PCR amplification were sent for sequencing to Macrogen Inc. (South Korea) for sequencing using the Sanger Dye-Termination sequencing technology. The sequencing of PCR products was carried out in both the directions.

The sequencing data obtained was used to make network trees using Molecular Evolutionary Genetics Analysis (MEGA) (Version 6) (Tamura *et al.*, 2013) , ClustalX (Version 2.1) (Larkin *et al.*, 2007) and SplitsTree (Version 4) (Huson and Bryant, 2006).

MEGA is a software designed for comparative analysis of homologous gene sequences with emphasis on inferring evolutionary relationships and patterns of DNA and protein evolution. MEGA can be used for the assembly of sequence data sets from files or web-based repositories and includes tools for presenting results in the form of interactive phylogenetic trees and evolutionary distance matrices (Kumar *et al.*, 2008).

ClustalX is the graphical user interface for the ClustalW multiple sequence alignment program for performing multiple sequence / profile alignments and analysing results. The sequence alignment is displayed in a window on the screen with versatile colouring scheme to highlight conserved features in the alignment. The sequences can be moved to change the order of the alignment; to select a subset of sequences to be aligned; or select a sub-range of the alignment to be realigned and inserted back into the original alignment. Alignment quality analysis can be performed and low-scoring segments or exceptional residues can be highlighted (Jeanmougin *et al.*, 1998, Larkin *et al.*, 2007).

SplitsTree4 is an application for computing unrooted phylogenetic networks from molecular sequence data. The program computes a phylogenetic tree or

network from given a given sequence alignment, a distance matrix or a set of trees, using methods such as split decomposition, neighbor-net, consensus network, super networks methods or methods for computing hybridization or simple recombination networks (Huson and Bryant, 2006).

The network trees obtained from the sequence of *bvs* genes was compared to the network tree obtained from the sequence of *celC* gene from the strains studied. The *celC* gene codes for a cellulase which takes part in the infection process of clover by *R. leguminosarum*. It forms part of an operon (celABC) in *R. leguminosarum, R. etli* and *R. radiobacter*. The *celC* gene was studied as phylogenetic marker by Robledo *et al.* (2011) using data from several species. The network tree of the *celC* gene was compared to the network trees obtained from the sequence of the five *bvs* genes.

3.3. Results :

Complete genome sequence information was available for some of the strains. Most strains had incomplete or no sequence data at all (Tables 2.1 and 2.2). To ensure reliability of results, both approaches (viz. BLAST and PCR) were used for all the strains studied. *Rlv*. 3841 was used as the reference strain i.e. the sequence data of the five *bvs* genes from *Rlv*. 3841 was used to perform BLAST searches as well as to design specific primers to amplify the *bvs* genes.

3.3.1 Investigating presence of *bvs* genes using BLAST

3.3.1.1. Investigating presence of *bvs* genes using Local-BLAST

The distribution of the *bvs* genes in the six type strains and five field strains was investigated using BLAST using query sequences from the reference strain *Rlv*. 3841 and the genome data available on the BLAST server at the University of York.

The BLAST hits with e-value (expect value) below 10^{-5} and bit-score above 200 were considered as positive. The bit-score gives a statistical significance of the alignment. The higher the bit score, the more similar the two sequences. The e-value gives a measure of statistical significance. It describes the number of hits one can "expect" to see just by chance when searching a database of a particular size. The results of the investigation are presented in Table 3.2.

Table 3.2. Local BLAST results of 5 *bvs* genes against the genome data of rhizobium strains.

Strain↓ Gene→	*bvs1*	*bvs2*	*bvs3*	*bvs4*	*bvs5*
R. leguminosarum USDA 2370	+	+	-	-	-
R.pisi DSM 30132	+	+	+	+	+
R.fabae CCBAU 33202	-	-	-	-	-
R.phaseoli ATCC 14482	-	-	-	-	-
R.tropici CIAT899	-	-	-	-	-
R.etli CFN 42	-	-	-	-	-
R.leguminosarum bv *trifolii* WSM1325	-	-	-	-	-
R.leguminosarum bv *trifolii* WSM2304	+	+	+	+	+

Key : + = hit, - = no hit

3.3.1.2. Distribution of *bvs* homologous proteins using NCBI-BLAST

The distribution of proteins encoded by the *bvs* genes was studied using the amino acid sequences of the five *bvs* genes. The organism search set was limited to the family *Proteobacteriaceae* to search for homologous proteins in the family. The search was repeated after limiting the organism search set to the genus *Rhizobium* to check for additional species of *Rhizobium* having these genes. The results of NCBI-BLASTp results were screened and the arguments presented are based on the results with an alignment score greater than 200 and e-value less than 10^{-5}. The NCBI-BLASTp results for the *bvs* protein homologs within the family *Proteobacteriaceae* show that :

- *bvs1* homologs are aminotransferase Class V enzymes specifically annotated as a cysteine desulfurase in some strains of *Bradyrhizobium, Sinorhizobium, Pseudomonas* and *Achromobacter*.
- *bvs2* homologs are universally labelled as a LysR family transcriptional regulator or as a transcriptional regulator.

- *bvs3* homologs are also universally labelled as an endoribonuclease L-PSP.
- *bvs4* homologs are widely distributed in the *Proteobacteriaceae* and are labelled as a nitrilase / aliphatic nitrilase / cyanide hydratase.
- *bvs5* is annotated as a molybdenum / molybdopterin binding oxidoreductase or specifically as a sulfite oxidase.

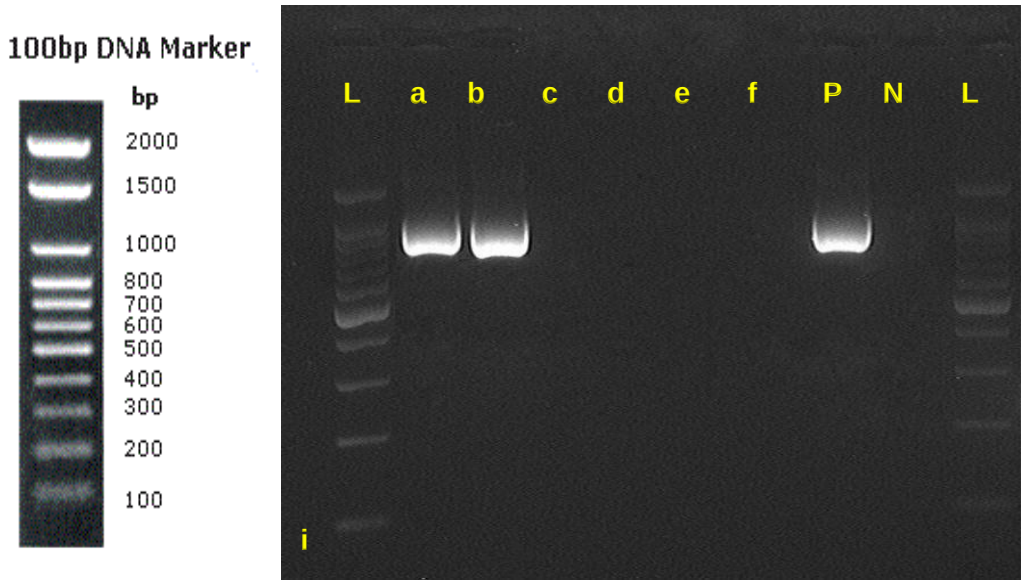
Restricting the search to the genus *Rhizobium* showed that only one strain viz. *R. leguminosarum* bv *trifolii* WSM2304 (besides *Rlv*. 3841), possessed all the *bvs* genes. The annotation of homologous proteins is in line with annotation of *bvs* genes in *Rlv*. 3841 and BLASTp results in *Rhizobium* are in line with local-BLASTn results.

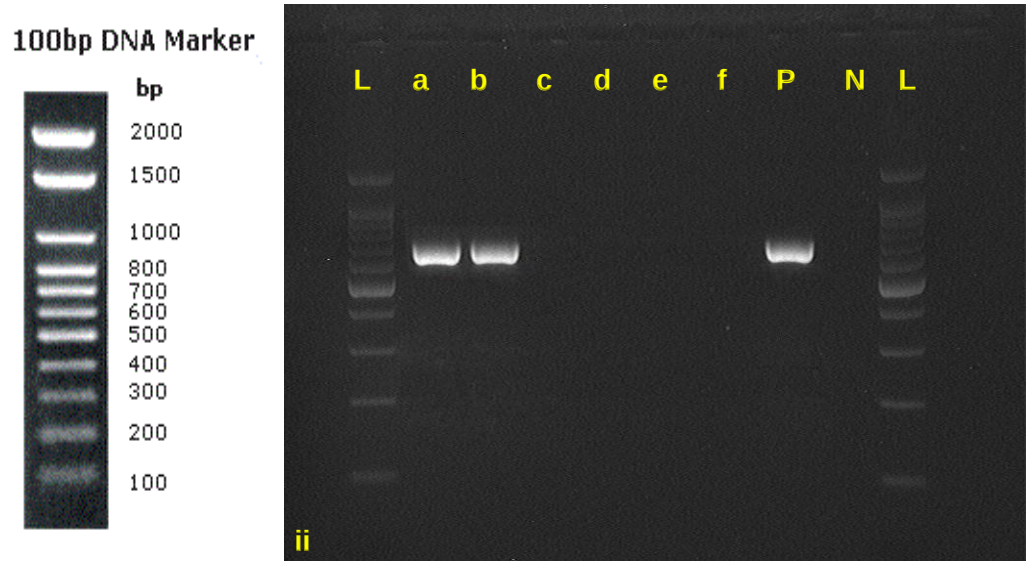
3.3.2 Investigating presence of *bvs* genes using PCR

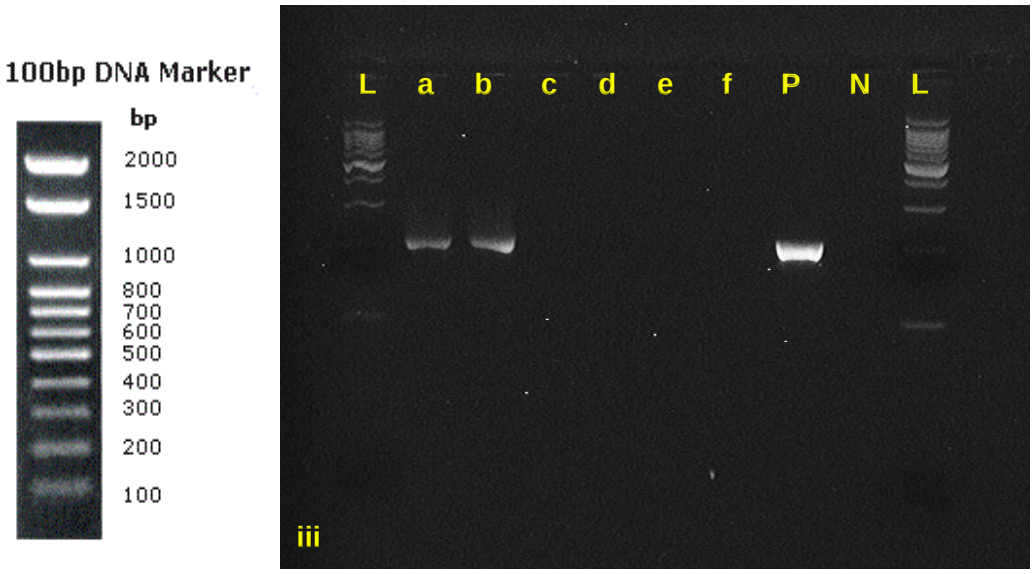
The distribution of the *bvs* genes in the six type strains and five field strains of the *Rhizobium* strains from different biovars were also checked using PCR with primers designed for the *bvs* genes listed in Table 3.1. The results of the investigation are as follows.

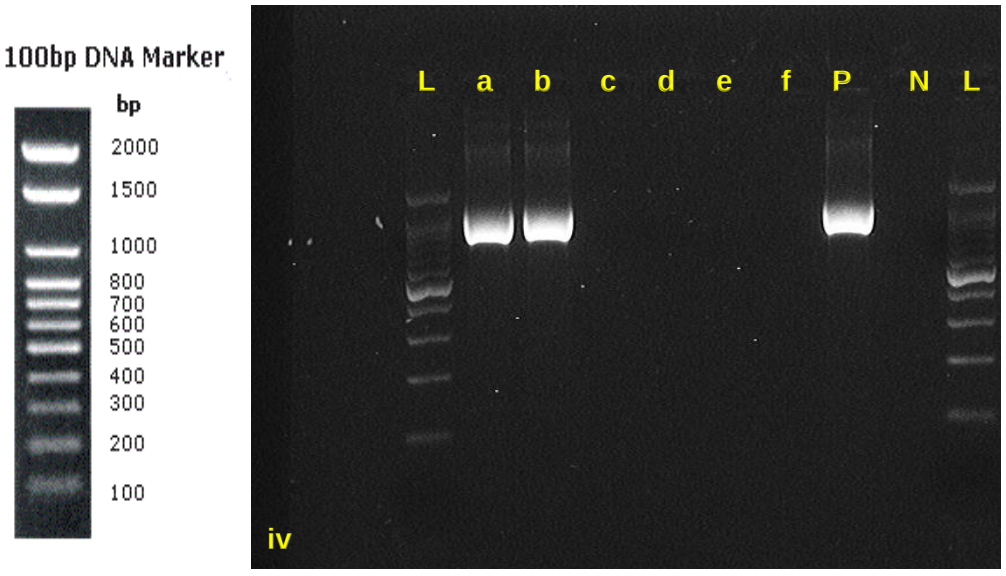
3.3.2.1. Presence of *bvs* genes in the six type strains

The distribution (presence / absence) of the *bvs* genes in the six type strains was carried out using PCR. The gel images of the electrophoretic runs are produced in Figure 3.2. (i to v).




100bp DNA Marker
bp
2000
1500
1000
800
700
600
500
400
300
200
100
L a b c d e f P N L
ii


100bp DNA Marker
bp
2000
1500
1000
800
700
600
500
400
300
200
100
L a b c d e f P N L
iii


100bp DNA Marker
bp
2000
1500
1000
800
700
600
500
400
300
200
100
L a b c d e f P N L
iv

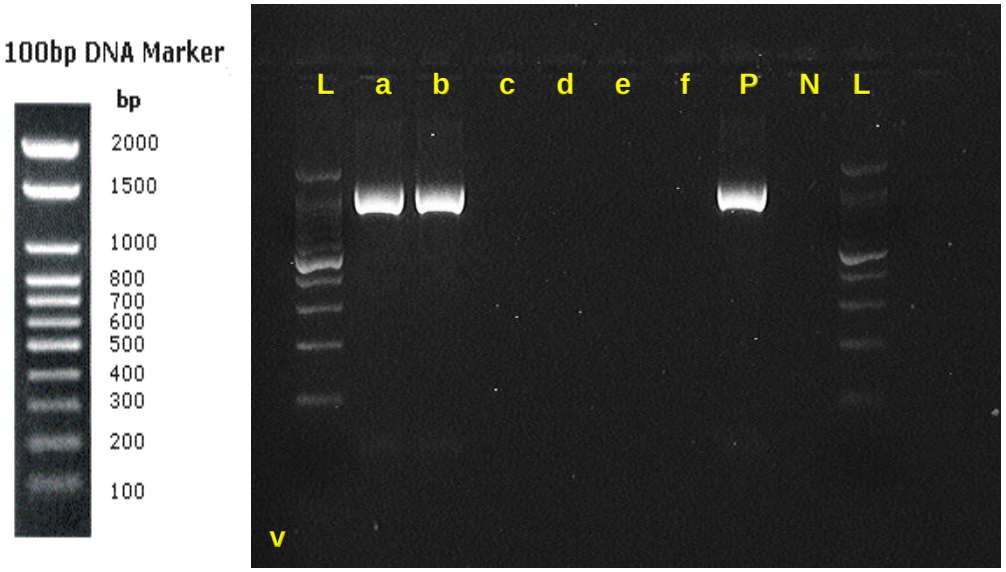
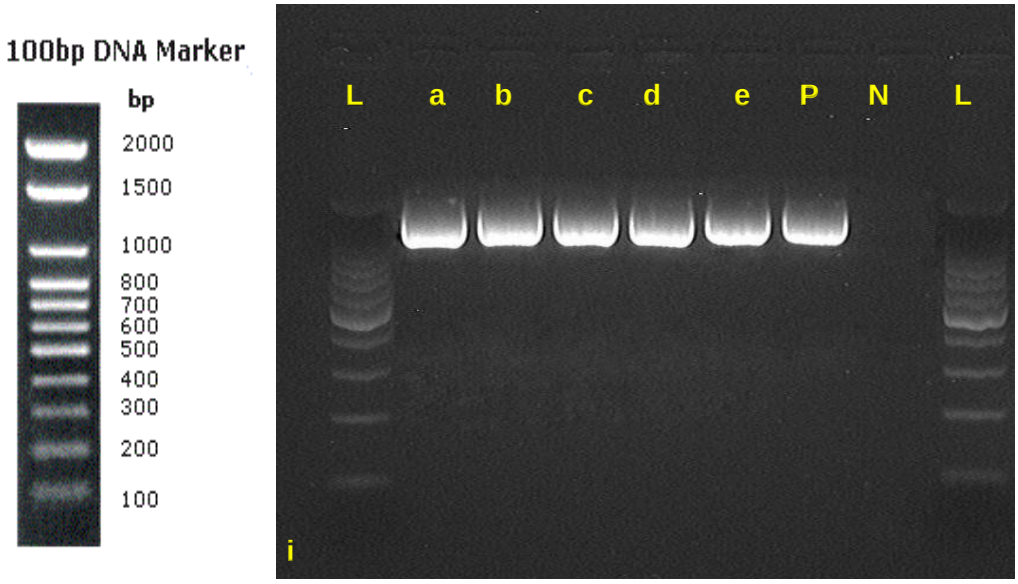
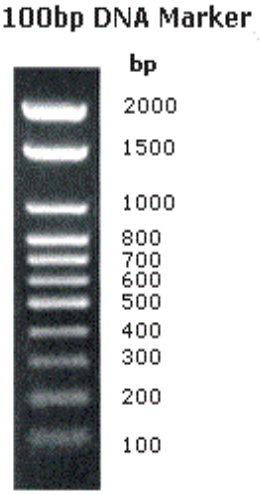


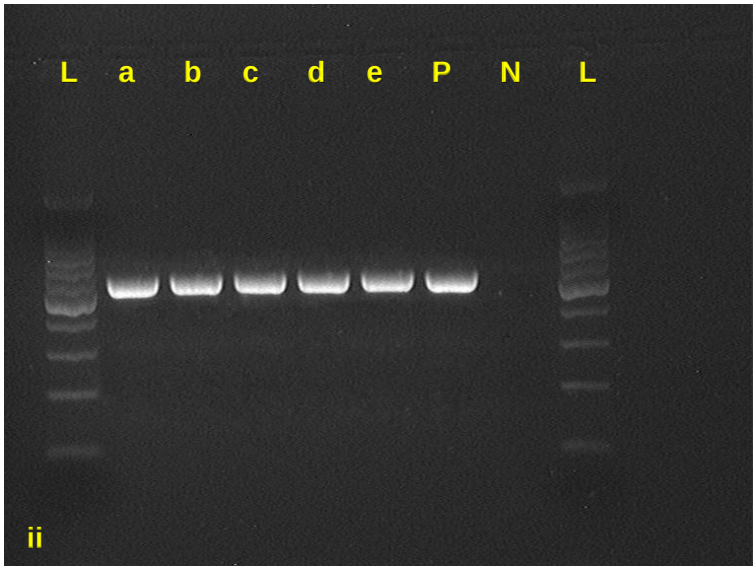
Figure 3.2. Agarose gel electrophoresis of PCR amplicons of *bvs* genes *bvs1-bvs5* correspond to Figures 3.2 i→v. a = *R. leguminosarum* USDA 2370^{T} , b = *R. pisi* DSM 30132^{T}, c = *R. fabae* CCBAU 33202^{T}, d = *R. phaseoli* ATCC 14482^{T}, e = *R. tropici* CIAT 899^{T}, f = *R. etli* CFN 42^{T}, P = positive control (*Rlv*. 3841), N = negative control (water), L = molecular weight marker (100bp DNA Marker).

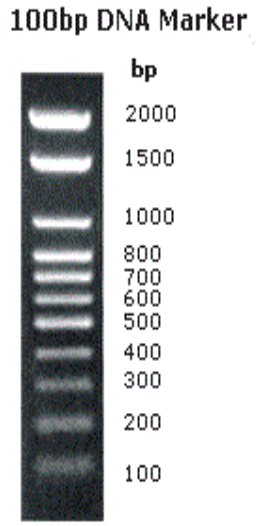
3.3.2.2. Presence of *bvs* genes in the five field strains

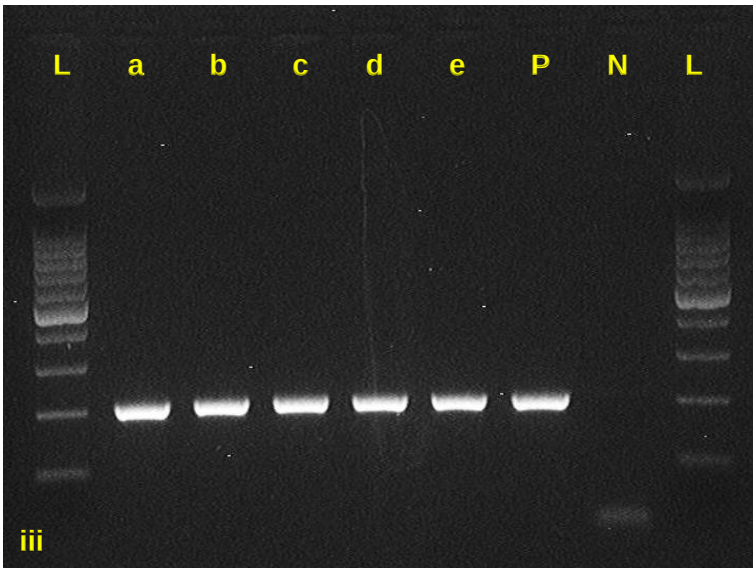
The distribution of the *bvs* genes in the five strains was studied using PCR. 5 µl from each of the amplification product was electrophoresed on agarose gel. The VSX field strains from Wentworth were not included since the presence of *bvs* genes in these strains has been demonstrated. The gel images of the electrophoretic runs are shown in Figure 3.3. (i to v).

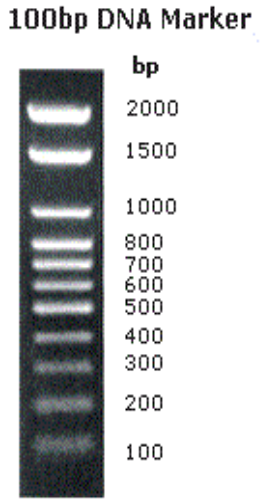


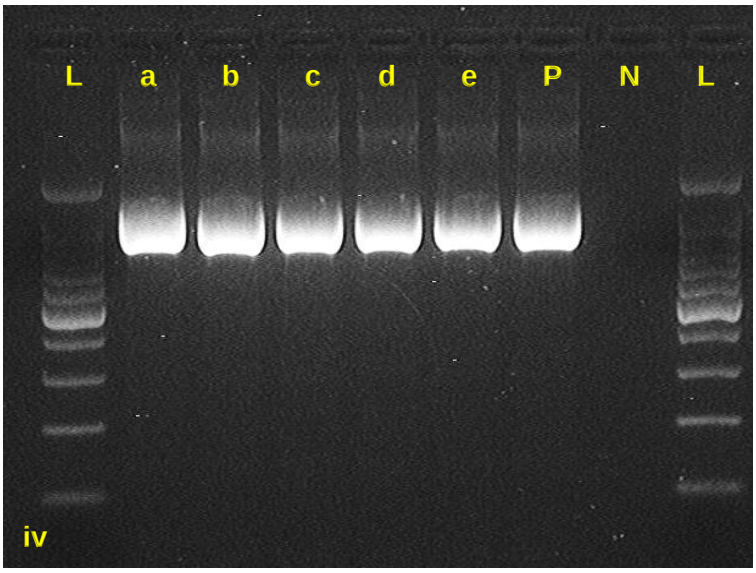

100bp DNA Marker
bp
2000
1500
1000
800
700
600
500
400
300
200
100


L a b c d e P N L
ii


100bp DNA Marker
bp
2000
1500
1000
800
700
600
500
400
300
200
100


L a b c d e P N L
iii


100bp DNA Marker
bp
2000
1500
1000
800
700
600
500
400
300
200
100


L a b c d e P N L
iv

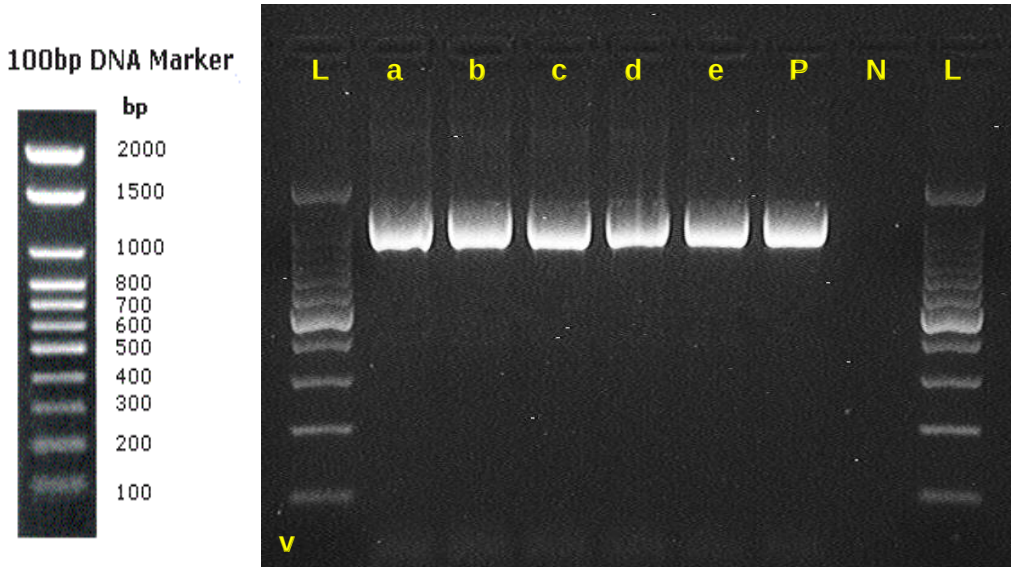


Figure 3.3. Agarose gel electrophoresis of PCR amplicons of *bvs* genes *bvs1-bvs5* correspond to Figures 3.2 i→v. a = Swedish strain OYAVB 169.1, b = Swedish strain OYAVB 296.5, c = Scottish strain 25 (S25), d = Scottish strain 34 (S34), e = Scottish strain 36 (S36), P = positive control (*Rlv*. 3841), N = negative control (water), L = molecular weight marker (100bp DNA Marker).

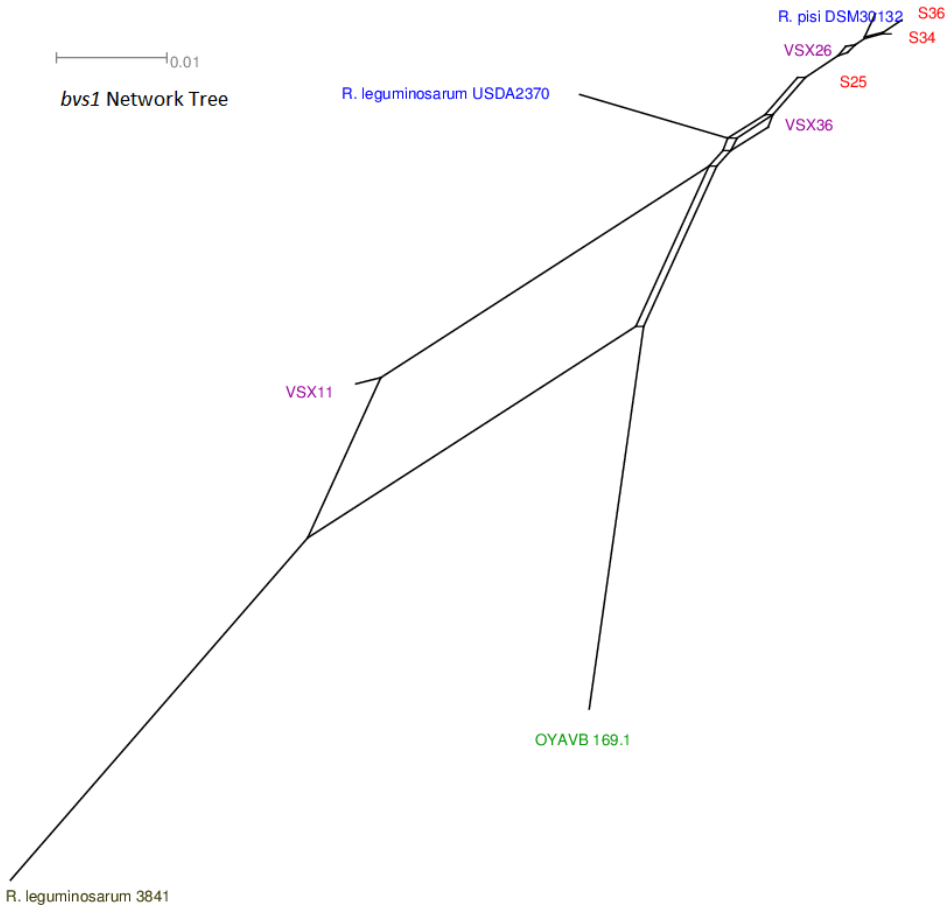
The results of the PCR from the gels of the six type strains and nine wild isolates are tabulated in Table 3.3. The results from the PCR analysis shows that amongst all the strains that belong to the biovar *viciae* (test and field strains), all the strains except *R. fabae* possess all the *bvs* genes.

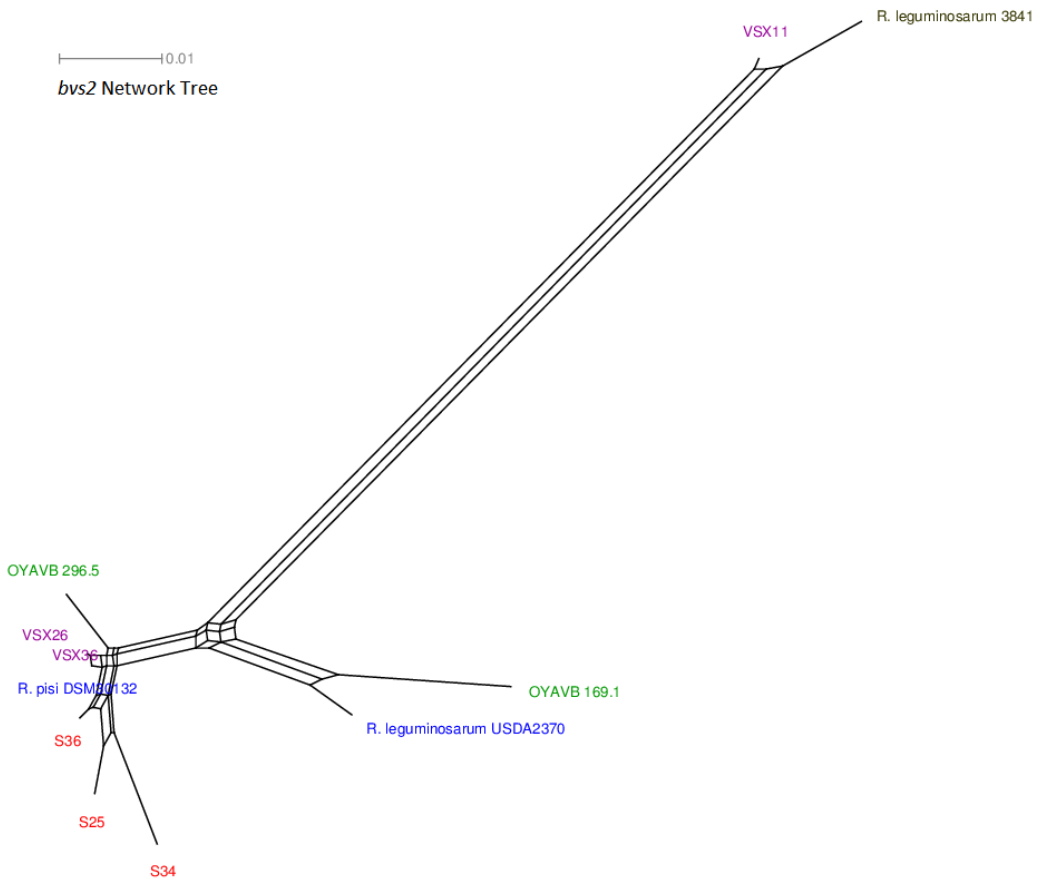
Table 3.3. PCR results of five *bvs* genes in tested rhizobial strains.

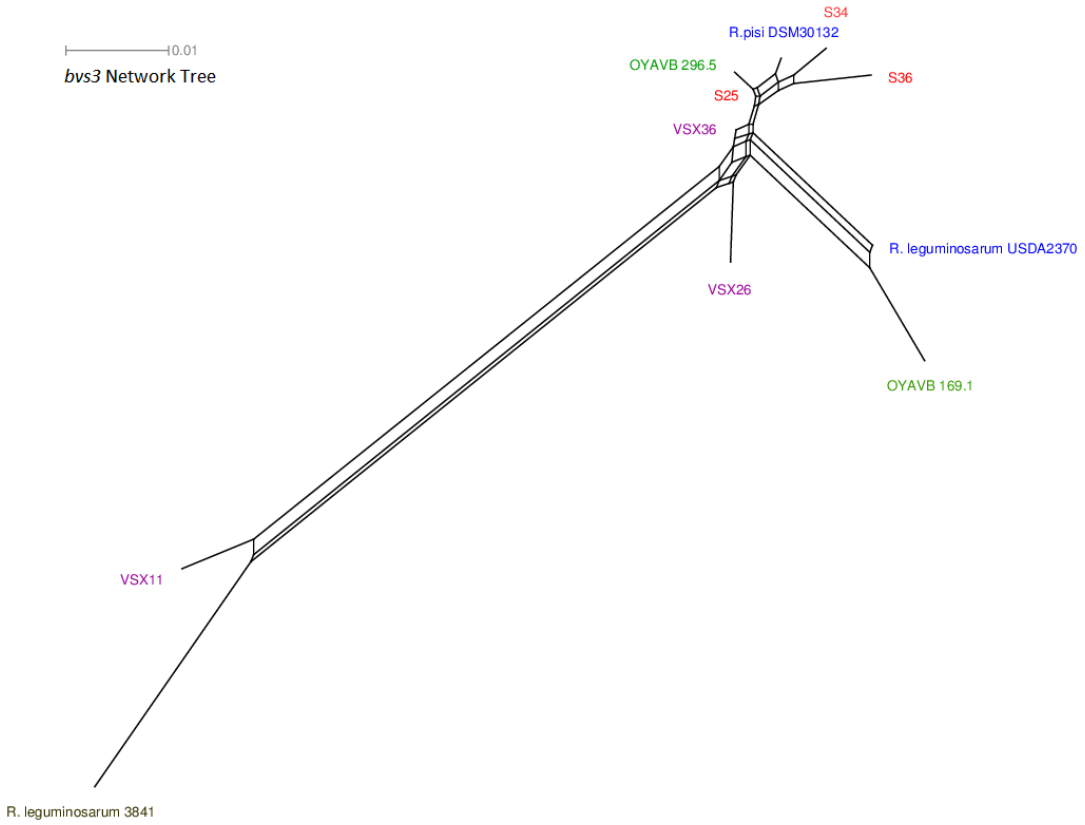
Strain↓ Gene→	Strain type	Biovar	*bvs1*	*bvs2*	*bvs3*	*bvs4*	*bvs5*
R. leguminosarum USDA 2370	Type strain	*Viciae*	+	+	+	+	+
R.pisi DSM 30132	Type strain	*Viciae*	+	+	+	+	+
R.fabae CCBAU 33202	Type strain	*Viciae*	-	-	-	-	-
R.phaseoli ATCC 14482	Type strain	*Phaseoli*	-	-	-	-	-
R.tropici CIAT899	Type strain	*Phaseoli*	-	-	-	-	-
R.etli CFN 42	Type strain	*Phaseoli*	-	-	-	-	-
OYAVB 296.5	Field strain	*Viciae*	+	+	+	+	+
OYAVB169.1	Field strain	*Viciae*	+	+	+	+	+
Scottish strain 25(S25)	Field strain	*Viciae*	+	+	+	+	+
Scottish strain 34(S34)	Field strain	*Viciae*	+	+	+	+	+
Scottish strain 36(S36)	Field strain	*Viciae*	+	+	+	+	+
VSX 11	Field strain	*Viciae*	+	+	+	+	+
VSX 26	Field strain	*Viciae*	+	+	+	+	+
VSX 36	Field strain	*Viciae*	+	+	+	+	+
R.leguminosarum 3841	Field strain	*Viciae*	+	+	+	+	+

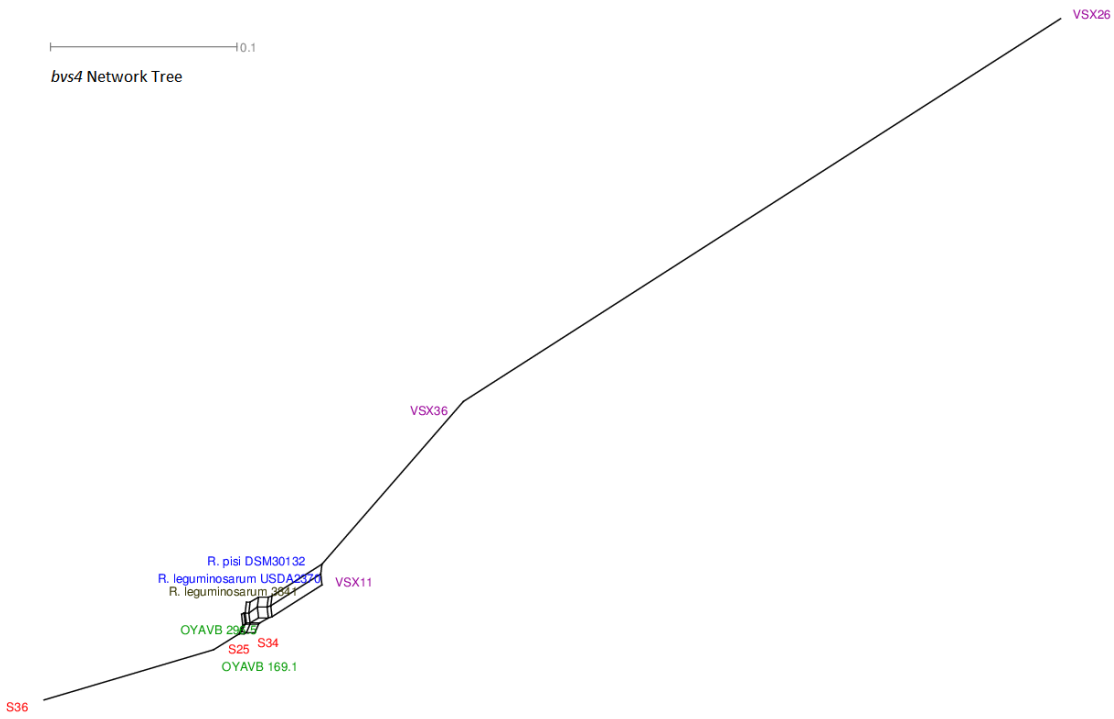
3.3.3 Phylogeny of *bvs* genes

The products of the PCR reactions from the amplification of *bvs* genes from the type strains and field strains were sequenced and the sequences were used for the construction of phylogenetic network trees. The network trees from the sequence of five *bvs* genes are shown in Figure 3.4. The network tree obtained using the sequence of *celC* gene from all the strains included in this study is shown in Figure 3.5. to compare the network trees of *bvs* genes with a reference gene suggested for used in rhizobium phylogenies.


0.01
bvs1 Network Tree
R. pisi DSM30132
S36
S34
VSX26
S25
R. leguminosarum USDA2370
VSX36
VSX11
OYAVB 169.1
R. leguminosarum 3841


0.01
bvs2 Network Tree
VSX11
R. leguminosarum 3841
OYAVB 296.5
VSX26
VSX36
R. pisi DSM30132
S36
S25
S34
OYAVB 169.1
R. leguminosarum USDA2370


0.01
bvs3 Network Tree
S34
R.pisi DSM30132
OYAVB 296.5
S36
S25
VSX36
R. leguminosarum USDA2370
VSX26
OYAVB 169.1
VSX11
R. leguminosarum 3841


0.1
bvs4 Network Tree
VSX26
VSX36
R. pisi DSM30132
R. leguminosarum USDA2370
R. leguminosarum 3841
VSX11
OYAVB 296.5
S34
S25
OYAVB 169.1
S36

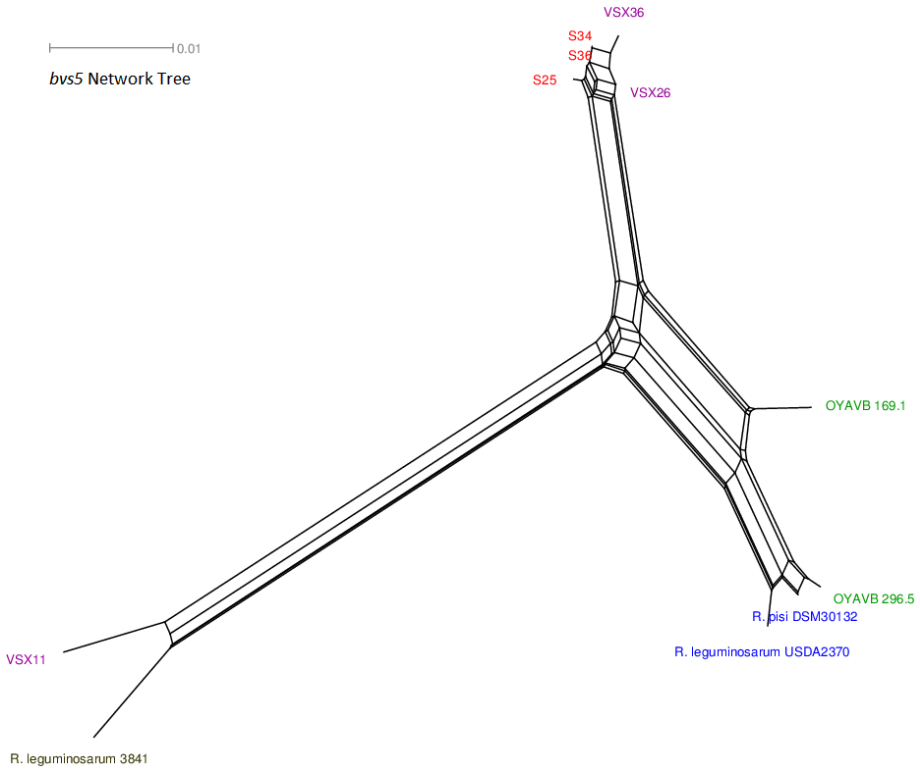


Figure 3.4. Phylogenetic-network trees based on *bvs* gene sequences from test and field strains. (*bvs1-bvs5*). Type strains (blue) Swedish strains (green), Scottish strains (red), VSX field strains (purple), *Rlv*. 3841 reference strain (black).

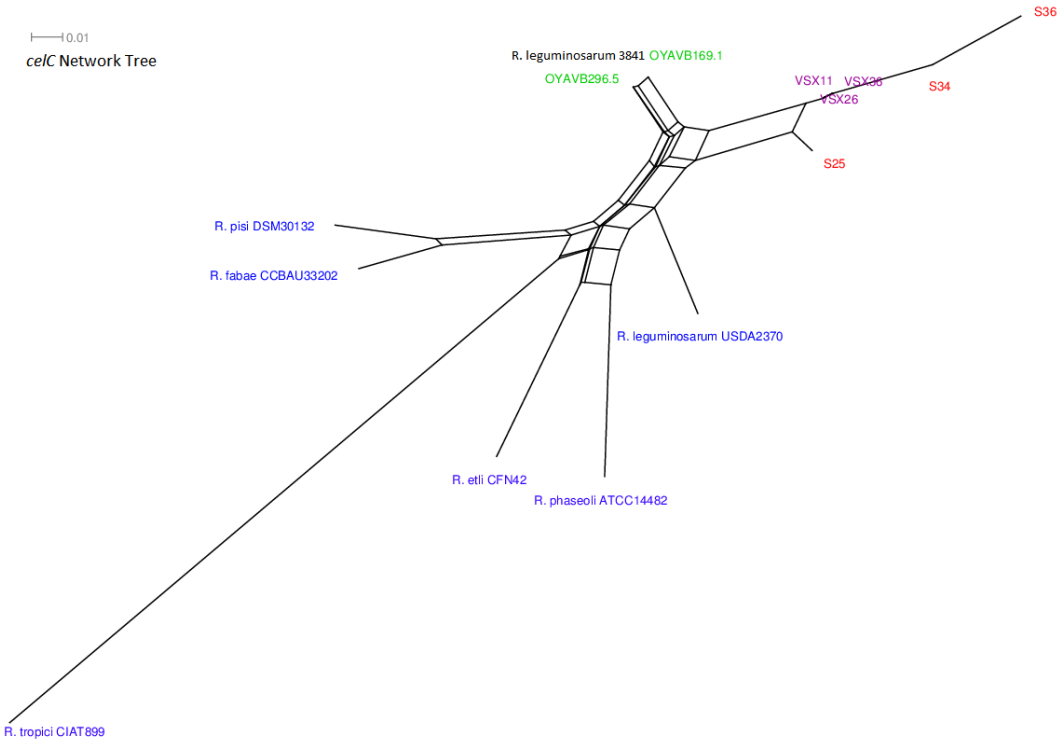


Figure 3.5. Phylogenetic-network tree based on *celC* gene sequences from test and field strains. Type strains (blue) Swedish strains (green), Scottish strains (red), VSX field strains (purple), *Rlv*. 3841 reference strain (black).

3.4. Discussion :

In the previous part of this study, the host-specificity and metabolic differences within the members of the *R. leguminosarum* species complex were studied by carrying out nodulation tests and metabolic fingerprinting using the Biolog GN2 MicroPlate. The results of the two analyses showed considerable variation in host specificity and metabolic profiles. No clear patterns were evident to suggest any differences between the members of the species complex indicating that the members of the *R. leguminosarum* species complex probably are phenotypically different.

Traditionally, the members of the genus *Rhizobium* were further subdivided based on their host range. It was later observed that the ability to nodulate different host plants was dictated by the *nod* genes (Long, 2001). Based on differences in the *nod* gene sequence, it was possible to differentiate between different biovars of *Rhizobium*. Hence, there existed a chance that there might be other genes or sets of genes that tend to occur more frequently in one species or biovar over the other.

Ramachandran *et.al* (2011) considered pRL8 to be a *Pisum sativum* specific symbiosis plasmid that enables rhizobia to form nodules on pea since they found that the expression of many of the genes on pRL8 was specifically increased in pea rhizosphere.

Since the *bvs* genes were found in strains that belonged to biovar *viciae* and since they reside on pRL8 which has been shown to be associated with the pea plant, a native host of *R. leguminosarum* biovar *viciae*, the distribution of the genes in the different species of *Rhizobium* was investigated using the BLAST sequence search tool and PCR. Phylogenetic network trees were constructed using *bvs* gene sequences and compared to the network tree of *celC* gene, a new phylogenetic marker suggested by Robledo *et al.* (2011) for use in taxonomic study of rhizobia.

The results of *bvs* gene distribution in rhizobial strains show that the five *bvs* genes are widely distributed in the rhizobial biovar *viciae* strains. BLASTp results were in line with Local-BLASTn results and *Rlv*. 3841 annotation. Amongst the many type strains studied, the three strains that did not belong to the biovar *viciae* (viz. *R. phaseoli* ATCC14482, *R. tropici* CIAT899 and *R. etli* CFN42) did

not have any of the five *bvs* genes as was seen in the results of BLAST as well as PCR. The results of the analysis are in line with the fact that there is no report of these bacteria nodulating *Vicia* plants. The Local-BLAST searches in *R. leguminosarum* bv. *trifolii* WSM1325 and *R. leguminosarum* bv *trifolii* WSM2304, however, were contrary to the expectation. Both the strains belong to the biovar *trifolii* but *R. leguminosarum* bv *trifolii* WSM2304 shows the presence of all the five *bvs* genes in its genome sequence whereas *R. leguminosarum* bv *trifolii* WSM1325 did not show the presence of any of the *bvs* genes. However, the *bvs* genes of WSM2304 show only 80% similarity to the *bvs* genes in the Wentworth population and the strain is more competitive as compared to other biovar *trifolii* strains (Reeve *et al.*, 2010a, Reeve *et al.*, 2010b).

Amongst the rhizobial biovar *viciae* type strains, *R.pisi* DSM30132 shows the existence of all five *bvs* genes in the results of Local BLAST as well as PCR. However, the host range of DSM 30132 ^{T} includes *V. sativa* and in nodulation tests the nodulation ability of DSM 30132 ^{T} on *V. sativa* was observed to be similar to that of other native symbionts of the legume plant. Since the *bvs* genes are associated with the ability to nodulate *Vicia* plants, the results of PCR amplification and Local-BLAST of *bvs* genes showing that DSM30132 ^{T} has all 5 *bvs* genes is not surprising and is again in line with the observation that is similar to other biovar viciae strains in the study.

Local-BLAST search for the five *bvs* genes in *R. leguminosarum* USDA 2370 shows that three of the five genes are missing. However, PCR amplification shows that all the five *bvs* genes are present in the strains. Nodulation test results show that *R. leguminosarum* USDA 2370 possessed the ability to effectively nodulate *Vicia* plants, suggesting shared similarities with other biovar *viciae* strains. The results of the PCR amplification also support the hypothesis that the existence of *bvs* genes promotes the ability of rhizobia to nodulate *Vicia* legume plants. The genome data available for *R. leguminosarum* USDA 2370 is not complete. Hence, the genes that were not detected in the Local-BLAST searches may be a result of the genome's low genome sequencing coverage.

PCR amplification and Local-BLAST search for the *bvs* genes showed that the genes are absent in *R.fabae* CCBAU33202$^{T.}$ Santillana *et al.* (2008) reported isolates from nodules of *Vicia faba* plants from Peru that were different from *R. pisi* DSM30132 ^{T}. The information on nodulation of *Vicia faba* by other species of

Rhizobium is insufficient. *R. fabae* CCBAU33202^T has been reported to nodulate *Vicia faba.* Mutch and Young (2004) have demonstrated that root-nodule isolates from *Vicia faba* can form effective nodules on other legume plants but isolates from other legumes are much less able to nodulate Vicia faba. Isolates from *Vicia faba* are unique from other rhizobia of biovar *viciae* (Mutch and Young, 2004). Possibly, *R. fabae* CCBAU33202^T has an alternate mechanism to carry out the functions conferred by the *bvs* genes.

All field strains isolated from *Vicia* plants from different places showed the presence of all the five *bvs* genes in PCR study lending strength to the relation between the presence of *bvs* genes and the ability to nodulate *Vicia* plants.

Phylogenetic network trees constructed using *bvs* gene sequence from strains investigated were compared to the phylogenetic network tree of *celC*. It was seen that in *celC* tree the strains from Scotland viz. S25, S34 and S36 belong to one cluster. Similarly, the strains from Sweden OYAVB296.5 and OYAVB169.1 also cluster together. However, the network trees of *bvs* genes did not show this pattern. Instead, in most cases, it clustered the Swedish and Scottish strains together. Also, the network trees of *bvs* genes suggest that, OYAVB296.5 and *R.pisi* DSM30132^T and similarly *Rlv*. 3841 and VSX11 are closely related to each other suggesting that horizontal gene transfer between the strain pairs.

In conclusion, the results of the Local-BLAST and PCR of rhizobial strains suggests that *Rhizobium* biovar *viciae* strains that form a symbiosis with the *Vicia* plants carry a set of genes, the *bvs* genes, that are rare in other biovars. BLAST shows the presence of all five *bvs* genes in *Rhizobium leguminosarum* bv*. trifolii* WSM2304, although it is less similar to the bv. *viciae* strains (80% sequence similarity). *R. fabae* CCBAU 33202^T does not have any of the five genes even though it belongs to biovar *viciae* strains. This suggests that the *bvs* genes, although not crucially important, may have some role to play in the nodulation process. However, it is not known which aspect of nodulation process they influence. Possible avenues include host range selection, nodule growth and nodulation competitiveness. A deeper investigation into the role of the *bvs* genes will form the core of the studies conducted in the remaining part of this study.

CHAPTER 4 : MUTATION AND COMPLEMENTATION OF *bvs* GENES AND THEIR EFFECT ON BACTERIAL COMPETITION AND PLANT GROWTH.

4.1. Introduction

The analysis of sequence data from Wentworth strains, presented in Chapter 3, revealed the presence of five genes consistently present in all the *Rhizobium leguminosarum* biovar *viciae* isolates but absent in the strains belonging to biovar *trifolii* obtained from the same area. Since the genes were specifically found in the biovar *viciae* isolates, they were designated as 'biovar *viciae* specific' or '*bvs*' genes. A detailed description of the five genes and their putative functions has been given in the previous chapter.

The study of distribution of *bvs* gene was extended to cover more members of the *Rhizobium leguminosarum* species complex. The results of the study showed that the five *bvs* genes were widespread amongst the biovar *viciae* species (Table 3.3). Only one biovar *viciae* strain, viz. *Rhizobium fabae* CCBAU 33202^{T} did not possess any of the five genes.

In the sequenced reference strain *Rhizobium leguminosarum* biovar *viciae* 3841, the *bvs* genes are present on the plasmid pRL8, the smallest plasmid of that strain. The plasmid harbours many genes with unknown functions, but it has been speculated that the plasmid may contain genes with some relationship to the rhizobial symbiosis and metabolism (Young *et al.*, 2006). Studies performed using transcriptomics and gene mutation studies have proposed that the plasmid pRL8 harbours genes that are essential for the utilization of the α-amino acid homoserine (Ramachandran *et al.*, 2011, Vanderlinde *et al.*, 2014).

The study of genetic difference between the strains from biovar *viciae* and *trifolii* showed that the five *bvs* genes were present in only the biovar *viciae* strains. Since they are rare in *trifolii* strains, it seems reasonable to assume that the genes may, in some way, be related to or involved in the process of nodulation. This involvement could be direct and affect the host-specificity of the strain; or indirect, altering the nodulation competitiveness of the strains and tilting it in the favour of the strains harbouring these genes. Hence, it is important to investigate the functions of these genes and the role they play in the process of nodulation and / or the host-specificity.

The functions of the *bvs* genes have not been verified experimentally. Because of their biovar-specific distribution, perhaps the most efficient way of studying the functions of these genes would be to mutate the genes and observe the effect of mutation on the symbiosis. The possible effects studied could include the ability of the mutant to nodulate the host plant (i.e. presence / absence of nodules on the root of the plant) and the effect of mutant culture on plant growth itself, which could be measured in terms of the plant dry weight. It is also possible that the mutation of the genes may affect the ability of the mutant to compete with the other strains in the environment to infect the plant and form nodules. Hence, the effects of mutation on the proportion of the wild-type to mutant strains in the nodules of the test plant as compared to that in the surrounding environment was studied, using appropriate genetic markers.

The arrangement of the *bvs* genes on pRL8 and the putative functions they encode have been described in detail in Chapter 3. The analysis of the genome by bioinformatic tools also suggested the existence of at least two operons that could possibly contain the five *bvs* genes. Creating mutants in these genes and studying their effect on different aspects of the symbiosis as mentioned in the above paragraph is an interesting aspect of this study.

Genes arranged in operons in bacteria transcribe a polycistronic mRNA. Mutations in genes contained within an operon exhibit a polarity in mutation wherein a mutation in the gene upstream in the operon affects the expression of the genes downstream in the operon cluster. Complementing these mutants with genes cloned in a suitable plasmid and introducing them into the mutant strains in different combinations would allow us to investigate if the mutations exhibit polarity, which would be a strong indicator of genes being arranged or grouped into operons.

To summarise, the aim of this chapter is to see the effect of *bvs* gene mutations and also to evaluate the effect of gene complementation on phenotype reversion. Thus, mutations will be introduced in each of the five genes in the *bvs* gene set. The effect of these mutations on the ability to nodulate, nodulation competition and plant growth will be studied. The gene which shows a change in nodulation or plant growth phenotype when mutated will be complemented with a working copy of the mutated gene as well as other gene(s) under investigation to check for reversion of phenotype to wild-type and the polar effect of mutation. It is

expected that the mutation of *bvs* genes should not change the host plants of rhizobial strains; since *R. fabae* CCBAU33202^{T} biovar *viciae* does not have any of the *bvs* genes yet it is able to effectively nodulate plants belonging to the genus *Viciae*. However, it is expected that at least one of these five genes should show some effect on rhizobial competition in nodulating *Vicia* plants.

4.2. Materials and methods

The technique to mutate the genes and introduce the mutated construct into *Rlv* 3841 was kindly provided by Dr. Jürgen Prell, and as described in Karunakaran *et al.* (2010). The technique is based on the principle of gene disruption using homologous recombination. The steps involved in the process are shown in Figure 4.1 :

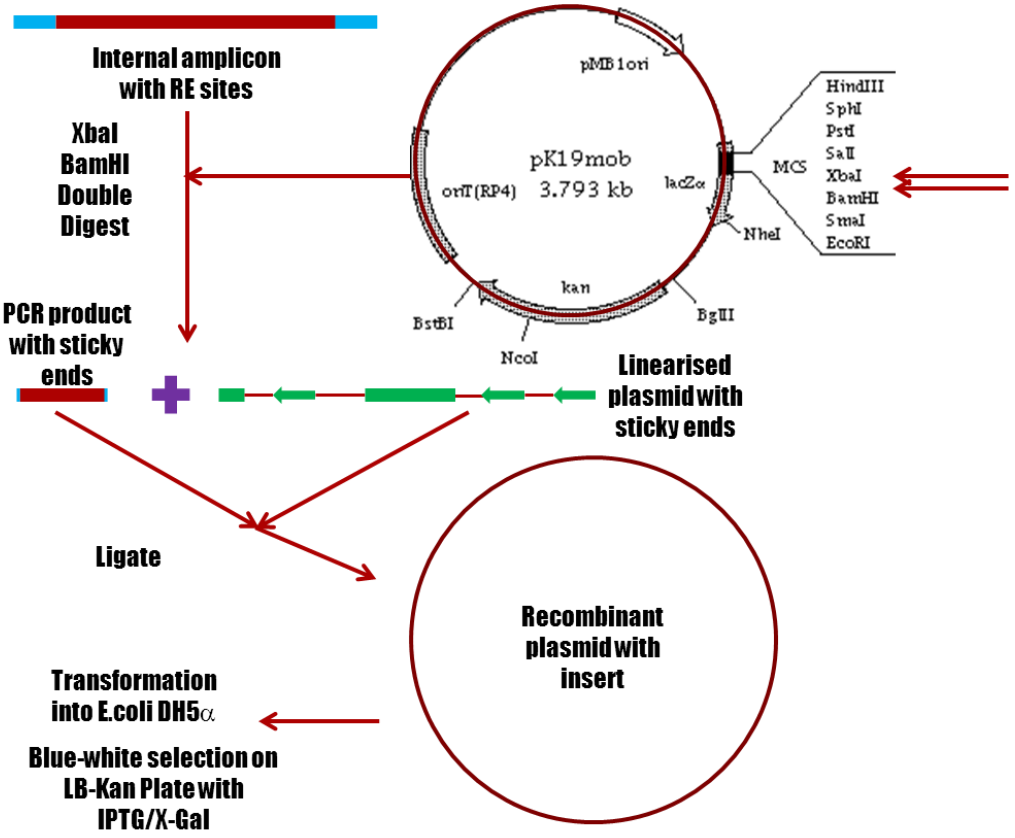
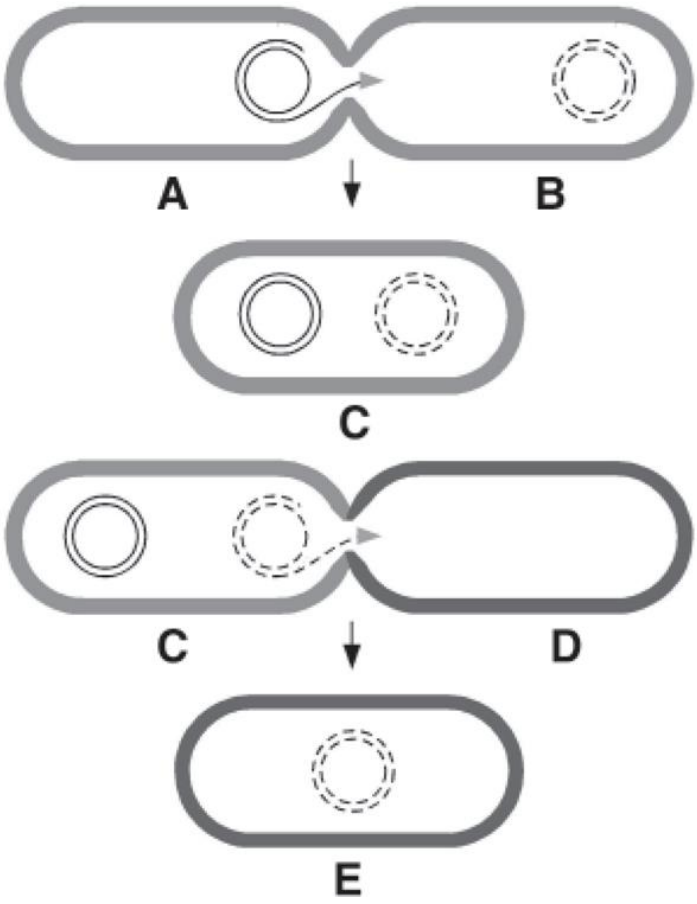


Figure 4.1. Schematic of cloning into pK19mob. The PCR amplicon with RE sites is added to plasmid pK19mob in presence of restriction enzymes. The linearized vector and amplicon, both with sticky ends, ligate to form a composite molecule which is transformed into *E.coli* for blue-white selection (pK19mob map from the cloning vector collection at shigen.nig.ac.jp).

The clones containing the desired insert were conjugated by triparental mating into recipient *Rlv* 3841 using the helper strain pRK2013 (Figure 4.2) and selected on TY agar plates containing Streptomycin and Neomycin. The Streptomycin inhibits the growth of *E.coli* whereas the Neomycin inhibits the growth of *Rlv* 3841 not carrying a copy of the pK19mob. The transconjugant clones were checked by PCR using M13 primers and a chromosomal primer outside the fragment used for the insertion (junction PCR). The clones yielding the desired results were then used for gene complementation studies.



2.

Figure 4.2. Schematic of triparental mating : The *E. coli* helper and donor strains are mixed together with the *Rhizobium* recipient strain. (A) The *E. coli* helper strain transfers the self-transmissible plasmid pRK2013 (solid circles) to *E. coli* donor strain. (B) The *E. coli* donor strain carries an engineered plasmid that is mobilizable but not self-transmissible (dotted circles). (C) The donor strain acquires pRK2013 from the helper strain and now carries both plasmids. (D) Using transfer functions supplied by pRK2013, the donor strain transfers the engineered plasmid to the *Rhizobium* recipient. (E) The engineered plasmid undergoes homologous recombination with the target gene in the *Rhizobium* cells. From Wise *et al.* (2006).

Gene complementation studies were carried out using a modified version of the protocol used by Vanderlinde *et al.* (2014). The method uses the vector pDG71 for complementation (Vanderlinde *et al.*, 2013, Vanderlinde *et al.*, 2009b). The plasmid possesses a strong *trp* promoter and a BamHI site adjacent to it. The proximity of the restriction site to the promoter made it an ideal site for cloning.
The genes were cloned into the plasmid using BamHI adapters and the construct was transformed into chemically competent *Rlv* 3841 cells. The cells were selected on Streptomycin-tetracycline plates. The direction of the insert was checked by PCR before using the strains for any further work.

Nodulation competition tests were used to study the effect of the single gene mutation. The test is designed to create a competition stress by mixing different rhizobial strains in a predetermined ratio and using this mixed inoculum to inoculate the host plants. At the end of the test, the nodules are harvested from the plants and the occupancy ratio of the strains in the nodules can be calculated and compared to the initial ratio. Thus, changes in occupancy ratio can be used to demonstrate an increase or decrease in nodulation competitiveness amongst test strains. Svenning *et al.* (2001) used nodulation competition tests in field experiments to observe nodule occupancy in white clover using three *Rhizobium leguminosarum* bv. *trifolii* strains. Rangin *et al.* (2008) studied 223 *Sinorhizobium* strains of which 13 reference strains were mixed with each other to compare their competitiveness on four test plants.

A marker or reporter gene can be added to one of the competitor strains in the assay to aid in easier identification of the strain. Barreto *et al.* (2012) used the *gfp* gene as a marker in one of their competition tests. In my work, plasmid pK19mob carries a Neomycin resistance gene and pDG71 carries a Tetracycline resistance gene. The wild-type *Rlv*. 3841 does not have resistance to either of these antibiotics. Hence, *Rlv*. 3841 can be identified by its inability to grow on selective medium containing Neomycin and Tetracycline, unlike the engineered strains.
The following section describes the materials and methods used in the construction of gene mutants, gene complemented strains and the effect of mutation on nodulation and nodule competition.

4.2.1. Mutating the genes by plasmid insertion

The method used for constructing mutants has been described by Prell *et al.* (2009). The method makes use of the vector, pK19mob, described by

Schafer *et al.* (1994), that can be easily transferred to a number of Gram positive and Gram negative bacteria to cause gene disruption by homologous recombination (Schafer *et al.*, 1994).

4.2.1.1. Primer design

All primers were designed using the online Primer-BLAST tool on the NCBI website (http://www.ncbi.nlm.nih.gov/ tools/primer-blast/). Two sets of primers were used in the study. The first set of primers, referred to as the “inside primers” were used to amplify a 600 bp region of the gene within the gene to be inactivated. This region was to be used to disrupt the gene later by homologous recombination. Primers were designed with an XbaI restriction-recognition site on the forward primers and a HindIII restriction-recognition site on the reverse primers. The regions that were selected for amplification were checked for presence of XbaI or HindIII site which could affect the creation of plasmid construct for use in gene disruption later (Schafer *et al.*, 1994).

The second sets of primers were called “outside primers” which were designed to amplify portion of the gene longer than the 600 bp region i.e. the primer binding sites were located “outside” the 600 bp region, one primer binding site on either side. These primers were used to confirm the direction of the insert in the transformed cells.

The genes, primer names and sequences are shown in Table 4.1.

Table 4.1. Genes, primer name and primer sequences used in creating and identifying gene mutant and gene complement strains. Lowercase letter show restriction site for enzymes XbaI on forward and HindIII on reverse primers.

Gene	Primer name	Primer sequence 5'→3'	PCR product size	PCR product range
Chromosomal nitrilase	Inside*ChrN*F	ttttctagaATGCCTTCGCCTTCCGTGAG	632 bp	709265 - 709896
	Inside*ChrN*R	tttaagcttTTTTCTTTCGTGCTGCCGCC		
	Outside*ChrN*F	AAGATGCAGCAGCTCGGGTC	755 bp	709171 - 709925
	Outside*ChrN*R	CCAGGGGACGAAGGTTTCGG		
bvs1	Inside*bvs1*F	ttttctagaCAGCGCGCCTTCTATTCCCT	613 bp	76618 – 77230
	Inside*bvs1*R	tttaagcttAGCGCGCTTCGATGTTTTCG		
	Outside*bvs1F*	TCGCCTGCGTGCTGATACTC	907 bp	76571 - 77477
	Outside*bvs1*R	AAGGTCTGAGGGGGATGCGT		
bvs2	Inside *bvs2*F	ttttctagaGACTTCTGCTGACCCCAGCC	268 bp	78165 - 78432
	Inside*bvs2*R	tttaagcttAAAAGATCTGCCTCCCGGCG		
	Outside*bvs2*F	CGCACTCGTCGAAAGGACCA	352 bp	78098 - 78449
	Outside*bvs2*R	TCGGACGCTTCAAGACTGGC		
bvs3	Inside*bvs3*F	ttttctagaTCCAGCACCGGCAAACTTGT	210 bp	78892 - 79101
	Inside*bvs3*R	tttaagcttGATTGTGGTGCGGGCAGGAT		
	Outside*bvs3*F	ATGGGCCGGCAACCTGTTA	314 bp	78846 - 79159
	Outside*bvs3*R	CGGTCTAACGGCAAAACGGC		
bvs4	Inside*bvs4*F	ttttctagaGGTTTTTCCGAGCGCAACCC	577 bp	79528 - 80104
	Inside*bvs4*R	tttaagcttTCGGCTCAACACAGTCGTCG		
	Outside*bvs4*F	GCTGCTCATATTGCCCCCGT	952 bp	79258 - 80209
	Outside*bvs4*R	ATTTTTGGAAGGCCGCGCTG		
bvs5	Inside*bvs5*F	ttttctagaAATGCCACAGCGCGAGTACA	591 bp	80474 - 81064
	Inside*bvs5*R	tttaagcttCATGAGCCTCCACCGTCCAC		
	Outside*bvs5*F	TGGAGATCGACCCCGAGCAT	848 bp	80309 - 81156
	Outside*bvs5*R	AGGCCCCCACGTCCTGTTAT		

4.2.1.2. Preparation of pK19mob

The plasmid pK19mob was provided by Dr. Jürgen Prell and used to create a pK19mob-bearing *E.coli* DH5α glycerol stock culture. A loopful of the glycerol stock culture was transferred to LB agar plate containing 25 µg/ml of Kanamycin (Sambrook, 1989) and incubated overnight at 37^{O}C. A single colony from the plate was transferred to sterile 5ml LB-Kanamycin broth medium in a 15 ml Corning tube and incubated on a rotary shaker at 150 rpm overnight at 37^{O}C. The plasmid was extracted from the overnight culture using the QIAprep Spin MiniPrep Kit from Qiagen using the protocol provided by the manufacturer.

4.2.1.3. DNA preparation :

The method for DNA extraction has been described in Chapter 3.

4.2.1.4. Amplification of internal region of the gene

The primers flanked by the recognition site for the restriction enzyme were used to amplify a 600 bp internal region of the genes which were to be inactivated by disruption. Six PCR reactions were setup, one for each of the *bvs* genes and one for the chromosomal nitrilase gene.

The working concentration of PCR mastermix, PCR conditions and PCR products confirmation method are listed in Chapter 3.

The remaining PCR product (about 30µl) was purified using QIAquick PCR Purification Kit (Qiagen) using the protocol provided by the manufacturer. The elution volume, however, was reduced to 20 µl and the eluate was stored at -20^{O}C until further use.

4.2.1.5. Restriction digestion of pK19mob plasmid and PCR products :

A mixture of the plasmid pK19mob and one of the PCR products was double digested using the restriction enzymes HindIII and BamHI (Fermentas). Six double digests were set up corresponding to PCR products obtained from the five *bvs* genes and chromosomal nitrilase.

In a clean sterile tube, 5 µl of the plasmid pK19mob and 5 µl of PCR product was mixed with 10 µl of 2X Tango Buffer (Fermentas) with restriction enzymes and incubated at 37^{O}C for 1 hour as per the manufacturer's instructions. After incubation, the mixture was cleaned using QIAquick Gel Extraction kit (Qiagen) and eluted in 25 µl of elution buffer. 5 µl of each clean-up product was transferred to a new tube, stored at -20^{O}C as the 'pre-ligation' sample to run on an agarose gel. 17 µl of the leftover elution product was used for ligating the plasmid and the PCR product.

4.2.1.6. Ligation of pK19mob and PCR product

The eluate from the cleaned-up double digest contains the PCR product and the plasmid pK19mob product cleaved at the HindIII and BamHI sites, both of which generate 5' overhangs or 'sticky ends'. The restriction site for the enzymes resided in the MCS region of pK19mob and is separated by a distance of 30 bp whereas the restriction sites in the PCR products resided in the flanking sequences of the primers.

As both the molecules (pK19 and PCR products) possessed complementary sticky ends, they were annealed by adding a ligase in the presence of an appropriate buffer. The DNA ligase formed a stable hybrid by forming a phosphodiester bond. Six ligation reactions were set up corresponding to the restriction digests from the five *bvs* genes and chromosomal nitrilase.

17 µl of the eluate from the double digest mix was used for ligation which was performed in PCR tubes. 2 µl of 10X ligation buffer (New England Biolabs) was added to each tube followed by 1µl of T4 DNA ligase (New England Biolabs). The resulting 20 µl mixture was incubated at 16^{O}C overnight. 5 µl of ligation product was transferred to a new tube and stored at -20^{O}C as 'post-ligation' sample to run on an agarose gel. The remaining ligation product was stored at -20^{O}C for use in transformation of chemically competent *E.coli* DH5α cells.

4.2.1.7. Preparing of chemical competent *E.coli* DH5α cells

The protocol for preparing chemically competent *E.coli* DH5α cells for transformation was kindly provided by Ms. Madhuri Barge from Dr. Daniella Barilla's lab, University of York.

A loopful of *E.coli* DH5α culture from the glycerol stock was streaked on sterile LB agar plate and incubated at 37^{O}C for 48 hours. A single colony from this plate was transferred to a fresh sterile LB agar plate and incubated at 37^{O}C for 24 hours. A single colony the second plate was transferred to 10 ml of sterile LB broth in a 50 ml Corning polypropylene tube and incubated on a rotary shaker overnight at 150 rpm. 300 µl of the overnight culture was transferred to 60 ml sterile LB broth in a 250 ml Erlenmeyer flask and incubated at 37^{O}C on a rotary shaker at 150 rpm until the A_{600} was between 0.4 and 0.6 (~ 4 hours). The flask was cooled on ice for 10 minutes. All operations from this point were carried out in ice-cold conditions using precooled materials and equipment.

RF1 solution contained 4.6M Glycerol, 100mM RbCl, 50mM $MnCl_2$, 30mM K-acetate and 10mM $CaCl_2$, pH 5.8 adjusted with acetic acid. RF2 solution contained 4.6M Glycerol, 10Mm MOPS, 10mM RbCl and 75mM $CaCl_2$, pH 6.8 adjusted with NaOH. Both solutions were stored at 4 ^{O}C.

The culture was transferred to two 50 ml Corning polypropylene tubes (30 ml per tube) and centrifuged at 5000 rpm (~3000g) for 5 minutes in a precooled centrifuge. The supernatant was discarded and each pellet was suspended in 10 ml of ice-cold RF1 solution. The cell suspension was incubated on ice for 15 minutes and then centrifuged at 5000 rpm (~3000g) for 5 minutes in a precooled centrifuge. The supernatant was discarded and each pellet was suspended in 2.4 ml of ice-cold RF2 solution. The cell suspension was incubated on ice for 15 minutes and then 50 µl of the suspension was transferred to sterile pre-cooled 500 µl flip-top tubes. The cells were frozen by dipping the tubes in liquid nitrogen till the cell-suspension turned solid and opaque and transferred immediately to a -80^{O}C freezer for storage until use.

4.2.1.8. Transformation of chemical competent *E.coli* DH5α cells

Six tubes, each containing 50 µl of chemically competent *E.coli* DH5α cells and the tubes containing the ligation preparations for the six genes were thawed on ice. 2 µl of the ligation preparation was added to the cells in appropriately labelled tubes and mixed by pipetting. The tubes were incubated on ice for 30 minutes and then heat-shocked by immersing them in a water bath at 42^{O}C for 90 seconds. The tubes were removed from the water-bath and 250 µl of SOC medium, prewarmed to 37^{O}C, was added to each tube. The tubes were incubated at 37^{O}C for 2 hours without agitation. The entire contents of the tube were plated, as follows, on sterile LB agar plate containing 25 µg/ml Kanamycin and 40 µg/ml X-gal : 100 µl on two plates and 25 µl on two plates. The plates were incubated at 37^{O}C for 48 hours before observing for blue and white coloured colonies of transformants.

Two white colonies from the chromosomal nitrilase transformation reaction and four white colonies from the others were selected for screening and transferred to sterile LB agar slants containing Kanamycin (25 µg/ml). The slants were incubated at 37^{O}C overnight. The slants were used to prepare glycerol stocks and perform colony PCR to confirm the presence of the desired insert in the clones with primers used in amplification of internal region of the gene, but lacking restriction sites.

4.2.1.9. Tri-parental mating for plasmid transfer to *Rlv*.3841

Using the tri-parental mating protocol tri-parental mating was carried out as described below for each of the six plasmid constructs.

A loopful of the donor, recipient and helper cultures were transferred from glycerol stock onto agar media plates containing antibiotics (LB agar with 25 µg/ml of Kanamycin for donor and helper, TY with 500 µg/ml Streptomycin for the recipient). The *E.coli* culture plates (donor and helper) were incubated at 37^{O}C whereas the *Rlv* 3841 culture (recipient) was incubated at 28^{O}C. All cultures were incubated for 48 hours and subcultured on fresh media plates which were then incubated for 24 hours to obtain cultures in vigorous growth phase.

A loopful of culture was transferred to 10 ml of sterile broth medium containing appropriate antibiotics (LB broth with 25 µg/ml of Kanamycin for donor and helper, TY broth with 500 µg/ml Streptomycin for the recipient) and incubated at appropriate temperatures overnight (37^{O}C for donor and helper, 28^{O}C for the recipient) on shaker at 150 rpm.

The recipient Rlv 3841 culture for triparental mating was required in the stationary phase ($A_{600} \approx 1.0$) and required no further treatment. The donor and helper cultures were required in the mid-logarithmic phase. Hence, the donor and helper cultures were centrifuged and resuspended in 1 ml sterile saline. 100 µl of this suspension was transferred to fresh 10 ml of sterile broth and further incubated for 3 hours at 37^{O}C ($A_{600} \sim 0.4$) on a shaker at 150 rpm.

After incubating the donor and helper strains to the required growth phase, 700 µl of recipient culture and 350 µl each of donor and helper cultures were mixed and centrifuged on a bench-top centrifuge at 13,000 rpm (~10400g) for 5 min. The supernatant was discarded and the pellet was vortexed briefly to resuspend the pellet in the residual supernatant. 50 µl of the cell mixture was transferred onto a sterile nitrocellulose membrane filter disk placed on a sterile TY agar plate without antibiotics and incubated upright at 28^{O}C overnight. After incubation, the filter was resuspended in 1 ml sterile TY broth and the resulting bacterial suspension was spread-plated on three sterile TY agar plates containing 500 µg/ml Streptomycin and 80 µg/ml neomycin (10 µl, 100 µl and remaining suspension). The plates were then incubated at 28^{O}C for three days to allow for the growth of recombinants.

12 colonies from each triparental mating growing on the Streptomycin-neomycin TY agar plates were transferred onto sterile TY agar slants containing Streptomycin-neomycin and incubated at 28^{O}C until a good growth was observed. The agar slants were used to prepare glycerol stocks before using the colonies for further analysis.

The 12 clones from each gene mutation preparation were checked for gene insertion by colony PCR using an M13 primer and a

chromosomal primer outside the region of the gene used for insertion into the plasmid pK19mob. The clones showing insertion in the required region were used for further studies.

4.2.2. Nodulation competition assay

The nodulation competition assay was carried out to study the effect of gene mutation on nodulation, competition for nodule formation with the wild-type strain and the effect on plant growth.

4.2.2.1. Selection of host plants for the assay

The native host plants from the family *Fabaceae* viz. *Vicia sativa, Vicia faba* and *Pisum sativum* were selected as the host plants for use in the nodulation competition assay using *Rlv*. 3841 gene mutants.

4.2.2.2. Strains used in the assay

The strains used in the nodulation competition assay were *Rlv*. 3841 gene-disruption mutants in the chromosomal nitrilase and the five *bvs* genes. The wild-type *Rlv*. 3841 was used as the positive control.

4.2.2.3. Preparation of medium for plant growth

The plants used in the nodulation competition assay were grown in sterile vermiculite. The bigger plants, viz. *Pisum sativum* and *Vicia faba* were grown in 250 ml Pyrex borosilicate Erlenmeyer flasks containing 150 ml vermiculite. The smaller plants viz. *Vicia sativa* were grown in 60 ml Pyrex borosilicate tubes containing 30 ml vermiculite. Before planting, the flasks and tubes containing vermiculite were plugged with cotton wool and autoclaved at 15 psi in an autoclave 121^{O}C for 20 minutes and allowed to cool. The cooled vermiculite was moistened with sterile Fahraeus N-free liquid medium (70 ml per flask and 14 ml per tube). The flasks and tubes were covered with sterile film and stored in the cold room until further use.

4.2.2.4. Preparation of seeds for growth

The seeds for *V. sativa* ssp. *segetalis* and *P. sativum* (variety 'Douce Provence') were purchased from Emorsgate Seeds (Norfolk, UK). The

seeds for *V. faba* were purchased from Victoriana Seeds (Kent, UK). The seeds of *V. sativa* and *P. sativum* were washed with multiple rinses of tap water and left to imbibe in tap water for 24 hours. The seeds of *V. faba* were washed with multiple rinses of tap water and soaked in running tap water overnight. After imbibition the seeds were again rinsed with tap water. The seeds were then surface sterilized with 70% ethanol for 1 minute. The ethanol was removed and the seeds were rinsed with sterile distilled water. The seeds were then treated with 3% hydrogen peroxide for 1 minute. The peroxide was removed and the seeds were washed with seven changes of sterile distilled water. A small number of seeds were transferred to TY and nutrient agar plates to check for efficiency of sterilization. The remaining seeds were for germination as described in section 4.2.2.6.

4.2.2.5. Preparation of test cultures

A loopful of the wild-type *Rlv* 3841 glycerol stock culture was streaked on normal TY agar plate. Similarly, a loopful of glycerol stock cultures of the mutants were streaked on TY-Streptomycin-Neomycin agar plates. The plates were incubated overnight at 28^{O}C. A single copy from each plate was transferred to appropriately labelled 50 ml Corning polypropylene tubes containing 25 ml of sterile liquid TY medium supplemented with antibiotics as required. The tubes were vortexed briefly to disperse the cultures and then incubated at 28^{O}C on a rotary shaker at 150 rpm for 24 hours. 100 µl of the culture was then transferred to identically labelled tubes containing identical medium and incubated under identical conditions for 6 hours. The cultures, now in their logarithmic phase of growth, were centrifuged, washed twice in physiological saline, and resuspended in physiological saline to A_{600} = 0.1.

Six test mixtures were prepared for the nodulation competition assay. Each mixture contained one part of the wild-type *Rlv*. 3841 culture and nine parts of one of the mutant cultures (1:9) i.e. the resulting suspension contained 10% wild-type and 90% mutant cell population. The *Rlv*. 3841 culture was used as positive control whereas the plants included in the negative control were not inoculated with any bacteria.

4.2.2.6. Seed germination, inoculation and planting

The sterilized seeds were placed in a sterile beaker containing a sterile moist tissue paper and incubated at room temperature until the seeds germinated and showed presence of a radicle. Seeds with clean and healthy radicles were separated from the non-germinating seeds and seeds with unhealthy radicles. The clean, germinating seeds were washed again with multiple rinses of sterile water in sterile beakers to remove the debris left over from unhealthy seeds and used for the bacterial inoculation stage in the nodulation competition assay.

The cleaned sterile seeds were then placed in the prepared bacterial cultures for 20 minutes so that the bacteria adsorb on the seed surface. Four seeds were used per test mixed culture (replicates). The seeds were then transferred to their respective growth container (flask / tube) in a laminar hood using sterile forceps. The seed was placed such that the radicle was embedded in the vermiculite. The neck of the flasks and tubes were plugged with sterile cotton wool.

Four positive and four negative control replicates were also set up for the assay. The plants included in the positive control were inoculated with the pure culture of wild-type *Rlv*. 3841. The negative control was not inoculated with any bacterial culture. The test plants and the control plants were watered regularly with equal volumes of N-free Fahraeus medium.

4.2.2.7. Plant growth and harvesting

The inoculated flasks and tubes containing the inoculated seeds were transferred to controlled temperature plant growth room. The room had a light cycle of 16 hour at 22^{O}C and dark cycle of 8 hour at 18^{O}C. Initially, the plants were watered using Fahraeus medium every three days; the frequency was increased when the growing plants lost moisture rapidly by transpiration drying out the vermiculite.

The plants were grown for five to seven weeks before harvesting. The time of harvesting was different for each plant since they grow at different rates. The harvesting was done just before the plants started flowering since the nodules start to senesce when flowering begins.

For harvesting, the vermiculite was carefully poured out from the flasks / tubes and the plant was removed gently. The vermiculite adhering to the roots was dislodged by immersing the roots in warm water and removing the vermiculite with a brush. The entire plant was then rinsed in two changes of clean water.

4.2.2.8. Nodules collection and calculation of population ratios

The roots of the cleaned plants were rinsed in 70% alcohol in a laminar flow hood. 100 nodules were picked from every set of plant replicates from each group (except negative control plants which lacked nodules). The nodules were squeezed gently between the tips of a sterile blunt forceps and exudate was streaked gently `on sterile TY agar plates divided into six sectors. The plates were incubated at 28^{O}C for 48 hours. The growth from TY plates was transferred to identical TY-Streptomycin-Neomycin agar plates and incubated at 28^{O}C for 48 hours to determine the antibiotic resistant population within the population isolated from the nodules.

4.2.3. Complementation of genes

Mutations in *bvs4* and *bvs5* were complemented with functional copies of these genes. The complementation of gene function was carried out by using the method described by Vanderlinde *et al.* (2013), with some modifications.

4.2.3.1. Primers design

The primer design tool from NCBI was used to design suitable primers to amplify the entire *bvs4* and *bvs5* genes for cloning. The primers were designed using the following criteria :

- The forward primer binding site was designed to be as close as possible to the start codon of the gene to ensure that after cloning it was close to the promoter in the plasmid vector used for cloning.
- The difference in the Tm of the primers was $\leq 1^{O}$C.
- The primers had low possibility of hairpin formation, self-dimer formation and hetero-dimer formation.
- The primers were specific for *R. leguminosarum* biovar *viciae* 3841.

Table 4.2. Primer sequences for genes used in complementation

Gene	Primer name	Primer sequence 5'→3'	PCR product size	PCR product range
bvs4	Complementary primer *bvs4*F	CTTGTCGAAATGGCTGATGG	1117 bp	79184 - 80300
	Complementary primer *bvs4*R	GGAGAAACCCTTCTTGGTCA		
bvs5	Complementary primer *bvs5*F	CGATATTTTCGACGTCACCG	1133 bp	80142 - 81274
	Complementary primer *bvs5*R	CACCTCCTGATTGCTGTAGA		

Primers were designed within the *trp* promoter of the plasmid pDG71. With the help of these primers and another set of primers within the gene in the opposite direction on the complementary strand, the presence and the direction of the insert after the complementation protocol could be determined. The sequence of these primers was as follows :

*Trp*F primer : TGGGCAAATATACTGAAATAGGTC

*Trp*R primer : GCCCGCTATCAGGAAGTG

4.2.3.2. PCR amplification of genes

The *bvs4* and *bvs5* genes were amplified by PCR using the primers listed in Table 4.2

The PCR was carried out with an initial denaturation at 95^{O}C for 3 minutes; followed by 35 rounds of amplification consisting of denaturation at 95^{O}C for 1 minute, primer annealing at 57^{O}C for 1 minute and extension at 72^{O}C for 4 minutes. A final extension elongation step was carried out at 72^{O}C for 10 minutes following which the DNA was stored at 4^{O}C.

The GoTaq Flexi DNA polymerase normally used for PCR gives 3'A overhangs which are not suitable for adapter ligation. The overhang needs to be digested using a single-strand nuclease with 3'→5' exonuclease activity. Hence, the high-fidelity Pfu DNA polymerase (New England Biolabs) was used for the PCR amplification (using

manufacturer's protocol) since it has 3'→5' proof reading exonuclease activity resulting in PCR products with blunt ends that are better suited for adapter ligation. The PCR product was checked for quality by running on an agarose gel and later purified using QIAquick PCR Purification Kit from Qiagen using the manufacturer's protocol.

4.2.3.3. Preparation of BamHI adapter complex

The genes *bvs4* and *bvs5* were each cloned in the plasmid pDG71 as described by Vanderlinde (2013, 2009a), with modifications. The plasmid has a strong *trp* promoter from *Salmonella* that is also expressed in *Rhizobium* and *Ensifer (Sinorhizobium)* spp. A BamHI site is located close to the *trp* promoter. The proximity of the restriction site makes it an ideal site for cloning. However, the *bvs4* gene has a BamHI site within the amplified region so it was not possible to use primers with BamHI sites as was done for cloning in pK19mob. Hence, the cloning was carried out using adapters with BamHI overhangs and using the adapter-ligated construct for cloning in pDG71 linearized with BamHI and cleaned to remove the enzyme.

The BamHI adapter was designed as an inverted-repeat palindrome (5'TACCGGGATCCCGGTA3'). The single stranded sequence was dimerized by heating the solution to 65^{O}C and then cooling it to 15^{O}C using a step-down program on a thermal cycler, ramping at 1^{O}C/min. The dimerization was confirmed by running an aliquot of the sample before and after the step-down process.

4.2.3.4. Preparation of BamHI adapter

The adapter complex was incubated with BamHI and calf-intestinal phosphatase (CIP) in 'universal' NEB Buffer 3 from New England Biolabs to form dephosphorylated BamHI adapters. The dephosphorylation was performed to ensure that the 'sticky ends' of the adapters do not self-ligate. The adapters were purified using phenol-chloroform-isoamyl alcohol treatment and followed by ethanol precipitation to remove the enzymes and other molecules.

4.2.3.5. Ligation of adapter and PCR product

The dephosphorylated BamHI adapter was ligated to the purified PCR product using T4 polynucleotide kinase and T4 DNA ligase from New England Biolabs using the manufacturer's protocol. The ligation was carried out at room temperature for 30 minutes.

4.2.3.6. Linearization of plasmid pDG71 and BamHI

The plasmid pDG71 was linearized using the restriction enzyme BamHI using the manufacturer's protocol. After linearization, the plasmid was purified using QIAquick Spin Column from Qiagen using the manufacturer's protocol. This was done to remove BamHI which would otherwise cleave the BamHI site within the gene. .

4.2.3.7. Cloning the PCR-adapter construct into linearized pDG71

The linearized and cleaned pDG71 was mixed with the PCR product-adapter complex containing T4 polynucleotide kinase and T4 DNA ligase. The mixture was incubated at 16^{O}C overnight to form a ligation product of the PCR amplicon and pDG71.

4.2.3.8. Inserting the recombinant plasmid into *Rlv*.3841 mutants

The recombinant plasmid was inserted into chemically competent *Rlv*. 3841 mutant cells by chemical transformation using the heat-shock method. Chemically competent cells for *bvs4* / *bvs5* gene mutant were prepared using the protocol described for preparing chemically competent cells for *E.coli* DH5α (Section 4.2.1.7). 5 µl of the recombinant plasmid ligation mixture was used to transform the competent cells by heat-shock using the protocol in section 4.2.1.8.

One tube of chemically competent *bvs4* mutant cells was transformed with the *bvs4* gene while another tube was transformed with the *bvs5* gene. Similar transformation was done for the competent *bvs5* mutant cell tubes. The transformants were selected on Streptomycin-tetracycline plates. The direction of the insert was checked by PCR before using the strains for any further work.

4.2.4. Competitiveness nodulation test of gene complementary strains

The aim of complementing gene mutations was to observe for reversion to wild-type phenotype. In order to assess this reversion, the nodulation competition assay was performed again as described in section 4.2.2. replacing the mutants with the gene complemented strains.

4.3. Results

4.3.1. Creating mutants in chromosomal nitrilase and five *bvs* genes

4.3.1.1. Amplification of central portion of gene by PCR

The central portion of *chrN* and each of the five *bvs* genes were amplified using the designed "inside" primers (section 4.2.1.4) to form PCR products of about 200 bp (*bvs2* and *bvs3*) or 600bp (chromosomal nitrilase, *bvs1*, *bvs4* and *bvs5*). Each PCR product was checked by agarose gel electrophoresis of 5µl as shown in Figure 4.3.

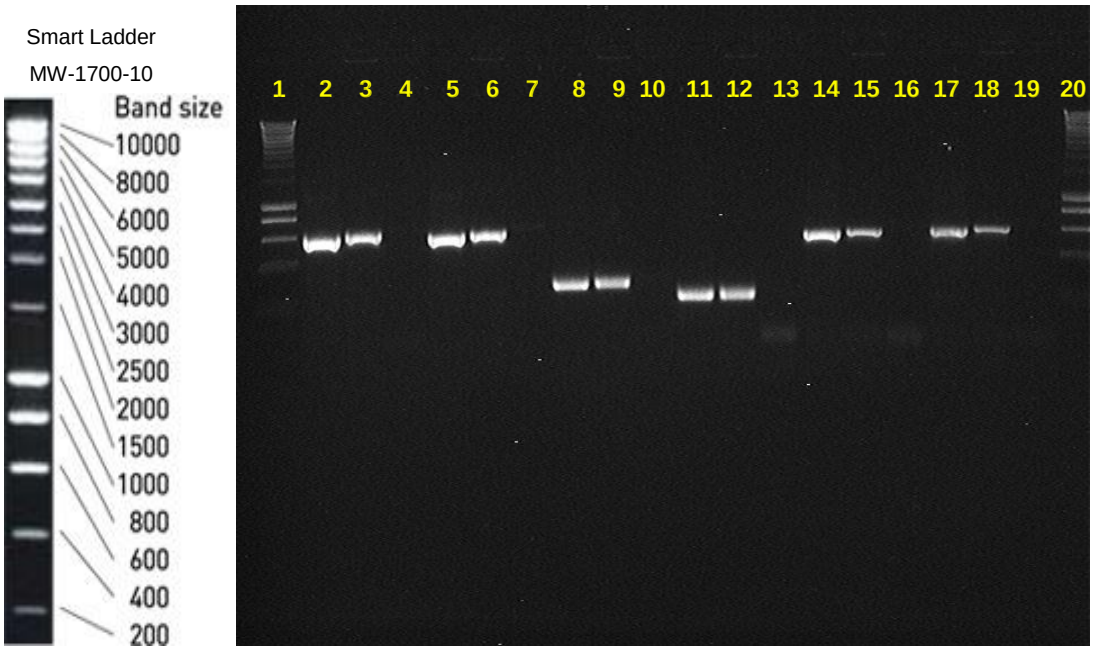


Figure 4.3. Gel picture of the two replicate amplifications of *Rlv*.3841 Chromosomal nitrilase gene (well 2nd and 3rd) and *bvs1* (well 5th and 6th), *bvs2* (well 8th and 9th), *bvs3* (well 11th and 12th), *bvs4* (well14th and 15th) and *bvs5* (well17th and 18th). The 1st and 20th wells contain the DNA molecular weight marker Smart Ladder MW-1700-10 from Eurogentec.

4.3.1.2. Restriction digestion and ligation of PCR product and plasmid

The pre-ligation and post-ligation samples were electrophoresed on an agarose gel to check for a ligation product of higher molecular weight and the simultaneous fading of the bands corresponding to the PCR product and plasmid. The image of the gel after electrophoresis of the pre-ligation and post-ligation samples is shown in Figure 4.4.

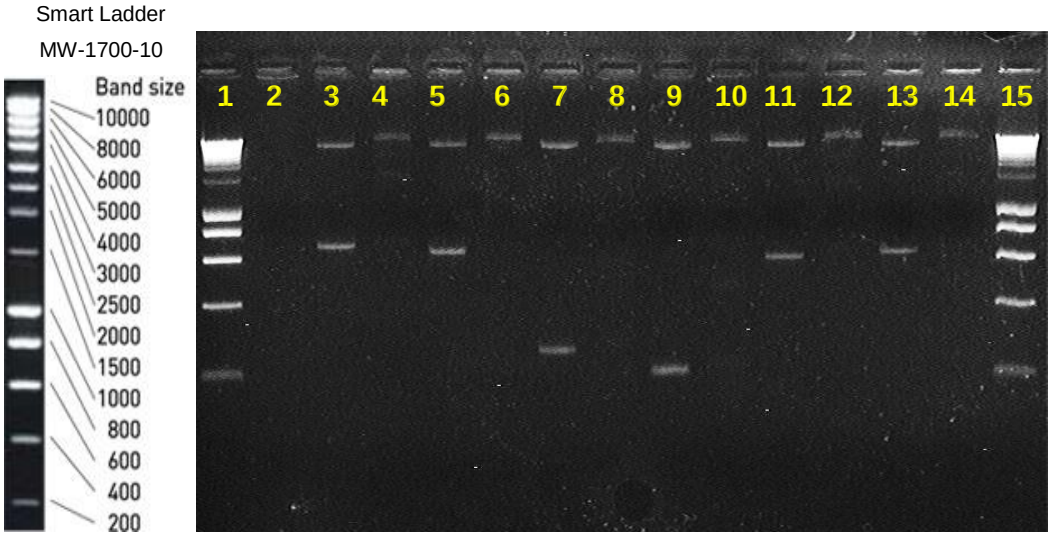


Figure 4.4. Gel electrophoresis of the samples before and after ligation. The 3rd, 5th, 7th, 9th, 11th and 13th wells contain the linearized pK19mob and one of the six PCR products (Chromosomal nitrilase, *bvs1, bvs2, bvs3, bvs4* and *bvs5* respectively) before ligation. The 4th, 6th, 8th, 10th, 12th and 14th wells contain the ligation products. 1st and 15th well contain DNA molecular weight marker Smart Ladder MW-1700-10 from Eurogentec.

4.3.1.3. Screening of transformed DH5α clones for presence of plasmid with insert :

Following transformation of *E.coli* DH5α cells, two white colonies from the chromosomal nitrilase ligation reaction and four white colonies each of the *bvs* gene ligation reaction were screened for the presence of insert by PCR using 'inside' gene primers lacking the RE sites. The PCR products were electrophoresed on a 2% agarose gel containing 1X SybrSafe and visualised over blue light shown in Figure 4.5.

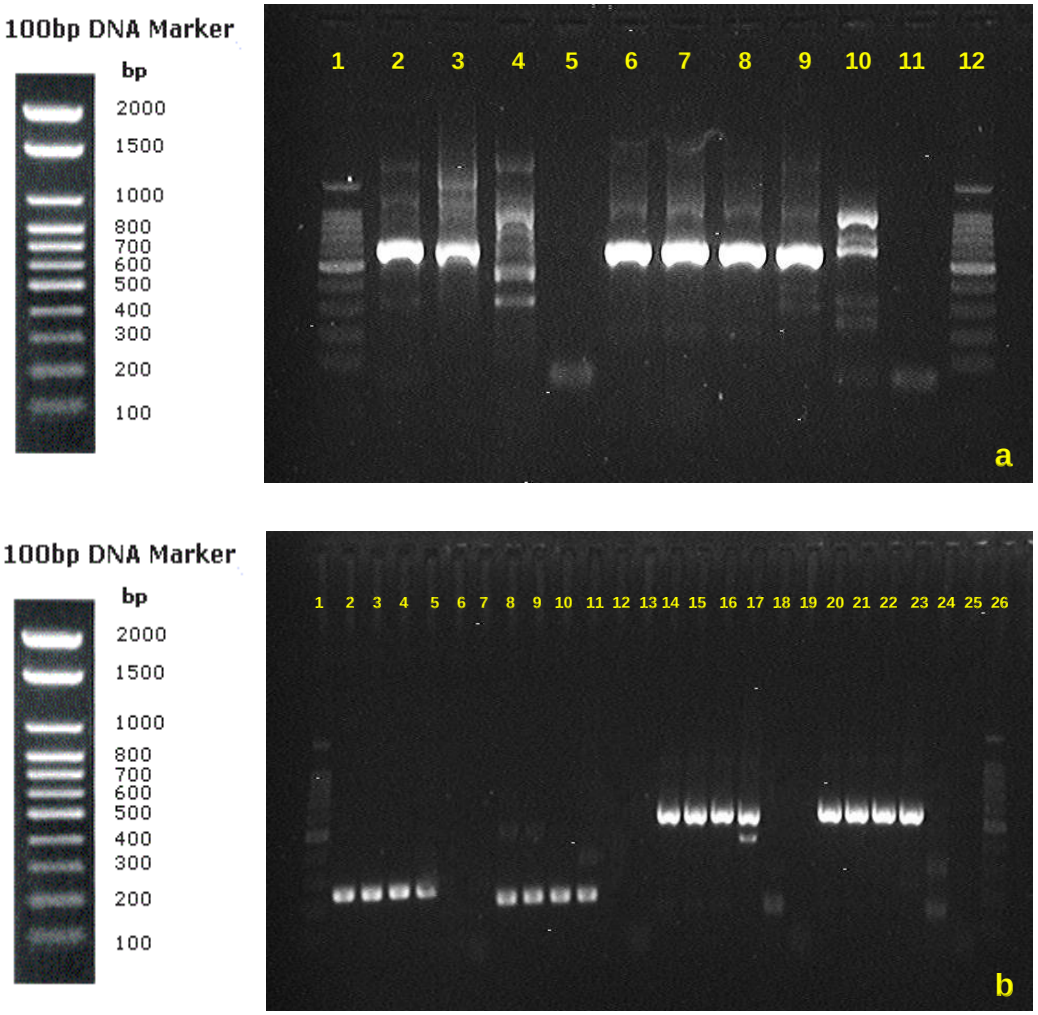


Figure 4.5. Gel electrophoresis of PCR products to check presence of cloned partial gene region in *E.coli* DH5α clones using inside primers. Different numbers of clones were checked. Figure 4.5.a: Chromosomal nitrilase (2nd and 3rd well) and *bvs1* (6th, 7th, 8th and 9th well) clones; Figure 4.5.b: *bvs2* (2nd, 3rd, 4th and 5th well), *bvs3* (8th, 9th, 10th and 11th well), *bvs4* (14th, 15th, 16th and 17th) and *bvs5* (20th, 21st, 22nd and 23rd well) clones. The DNA molecular weight marker used is the 100bp ladder from Invitrogen (Gel a: wells 1 and 12, Gel b: wells 1 and 26).

4.3.1.4. Screening of *Rlv*. 3841 transconjugants for plasmid insertion using PCR :

During triparental mating the plasmid pK19mob was transferred to *Rlv*. 3841 using the mob function of pRK2013. The plasmid integrates into the target gene by homologous recombination, thereby inactivating it. The transconjugants were screened by junction PCR using an M13 primer with a binding site on the plasmid and another primer binding in

the disrupted gene but outside the region used for initial amplification. The results of the gel electrophoresis are shown in Figure 4.6.

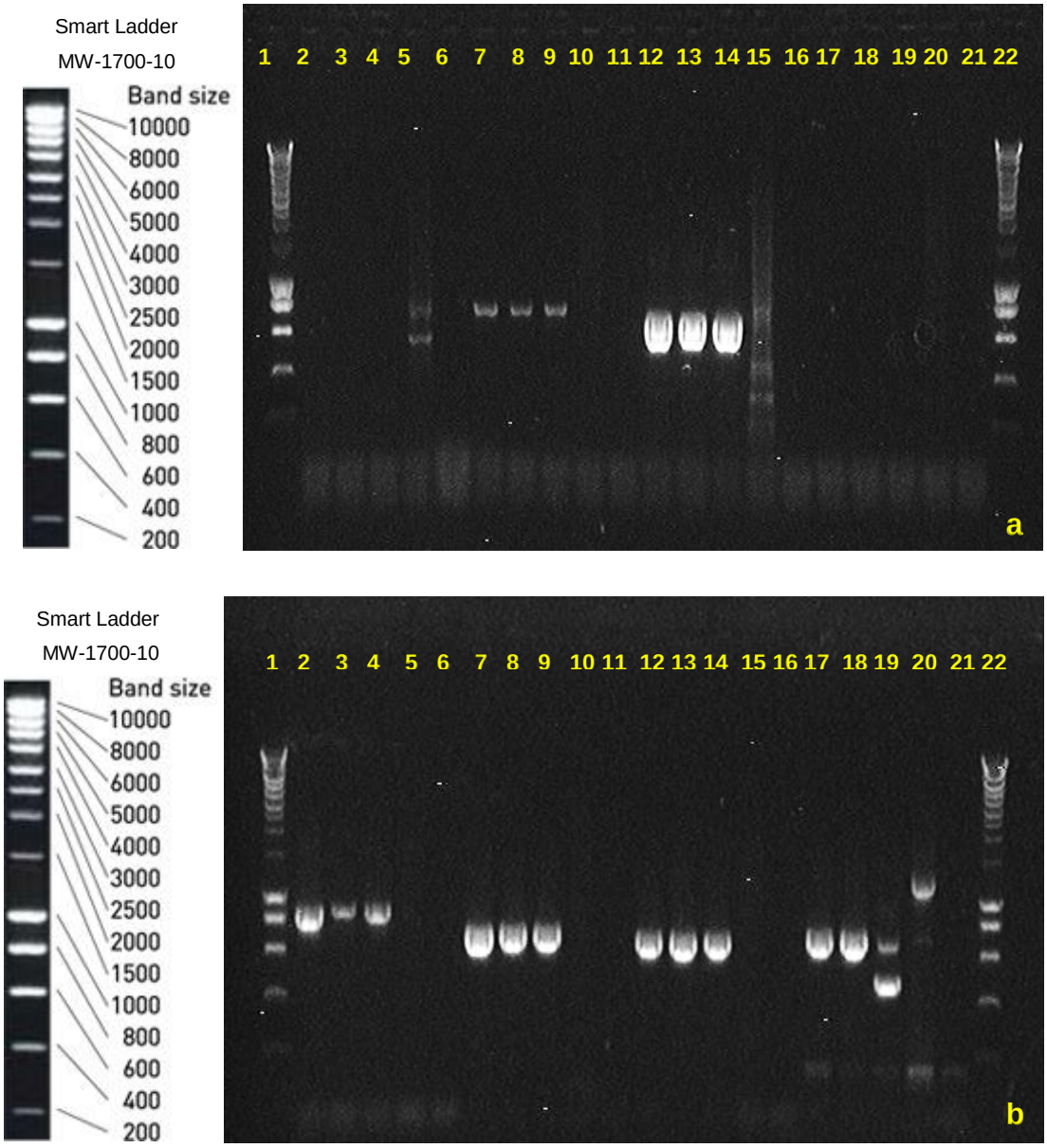


Figure 4.6. Gel electrophoresis of PCR products to check the presence of pK19mob in the disrupted genes of *Rlv.*3841 after tri-parental mating using a forward primer located in the uncloned region of the gene and a M13 reverse primer. Three clones for each gene insertion were checked with negative control. Figure 4.6.a: Chromosomal nitrilase (7th, 8th and 9th well) and *bvs1* (12th, 13th and 14th well) clones; Figure 4.6.b: *bvs2* (2nd, 3rd and 4th well), *bvs3* (7th, 8th and 9th well), *bvs4* (12th, 13th and 14th well) and *bvs5* (17th, 18th and 19th well) clones. 1st and 22nd wells in both figures contain DNA molecular weight marker Smart Ladder MW-1700-10 from Eurogentec.

4.3.2. Complementation of *bvs*4 and *bvs5* genes

4.3.2.1. Whole gene amplification of *bvs4* and *bvs5* genes and cloning into pDG71

For complementation of the *bvs4* and *bvs5* genes, the entire genes were amplified from *Rlv*. 3841 as described in sections 4.2.3.1 and 4.2.3.2. The PCR products were ligated with the adapters and cloned into the plasmid pDG71. The aliquots from each of these steps in the manipulation of DNA are shown in Figure 4.7.

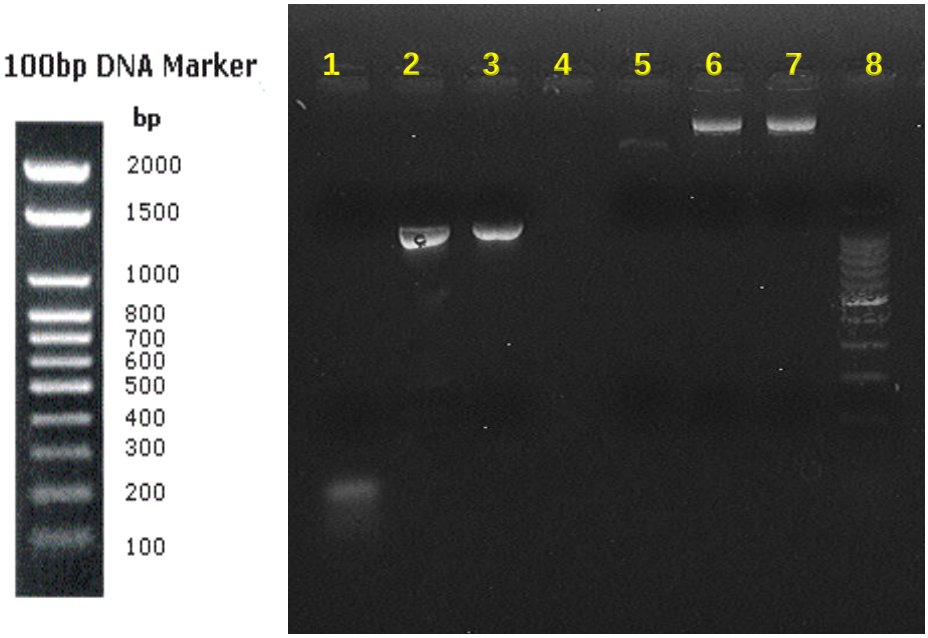


Figure 4.7. Gel electrophoresis of aliquots from DNA manipulations to generate recombinant pDG71 plasmid. 1st well: BamHI complex after ethanol precipitation; 2nd well: *bvs4* whole gene PCR product; 3rd well: *bvs5* whole gene PCR product; 5th well: plasmid pDG71; 6th well: plasmid pDG71 with *bvs4*; 7th well: plasmid pDG71 with *bvs5*; 8th well: DNA molecular weight marker 100bp ladder from Invitrogen.

4.3.2.2. Screening of *Rlv*. 3841 transformants to check for insert

The chemically competent cells of *Rlv*. 3841 mutants were complemented with a copy of functional gene cloned on the plasmid pDG71. The *bvs4* mutants were complemented with a copy of functional *bvs4* gene or *bvs5* gene. Similarly *bvs5* mutants were

complemented with a copy of *bvs5* gene or *bvs4* gene. This was done to check if the genes are organised into an operon.

If the genes are independent then they will not exhibit a polar effect of mutation commonly observed with genes arranged in operons wherein mutation in a gene placed transcriptionally upstream abolishes the functions of genes located downstream to it.

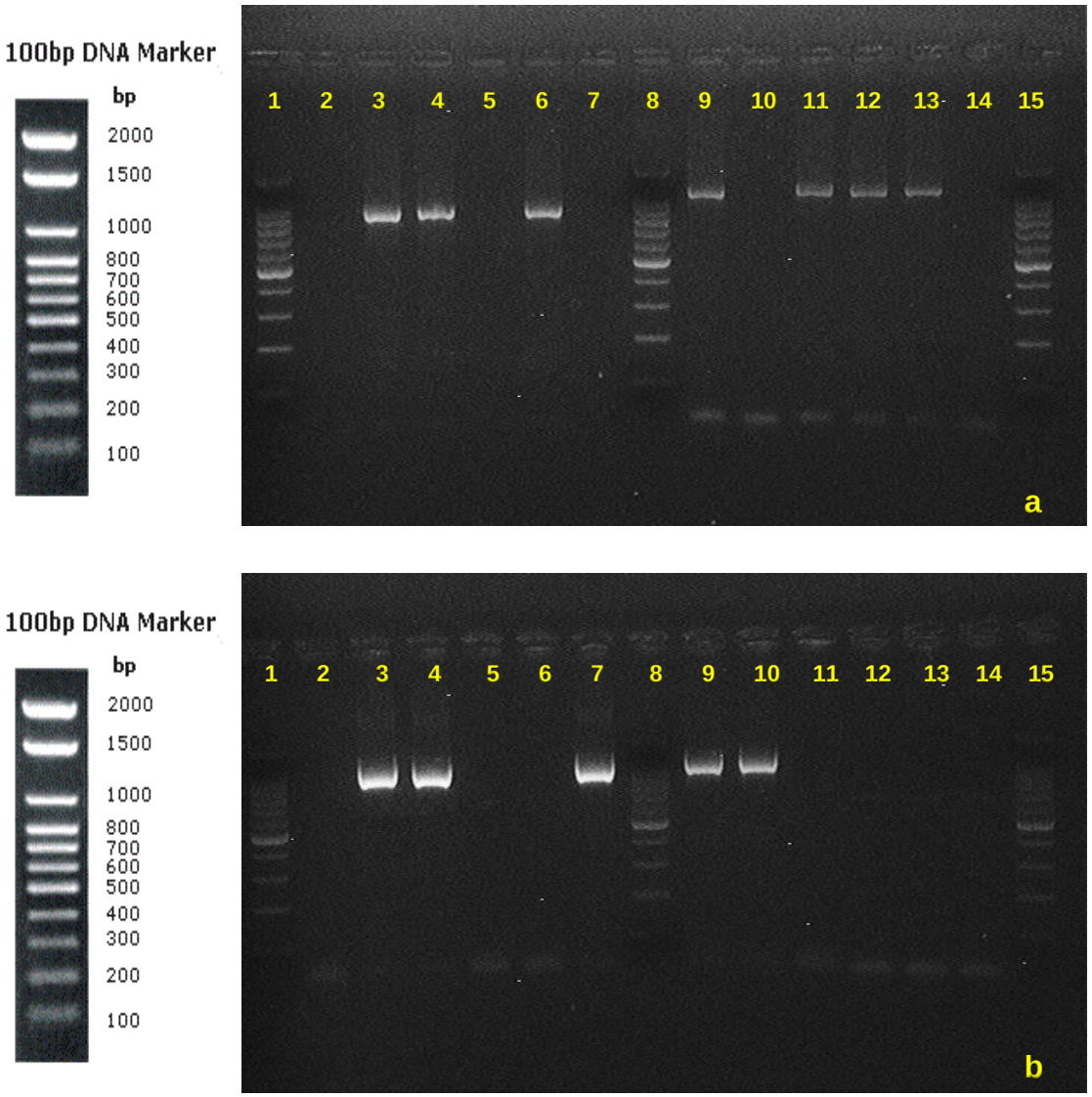


Figure 4.8. Gel electrophoresis to determine the presence and orientation of insert. Colony PCR of six clones of transformants from each gene was done using a forward primer within *trp* promoter in pDG71 and a reverse primer located within the gene. Clones which show bands were inserted in correct direction. Fig 4.8.a: *Rlv*.3841 *bvs4* mutant strains complemented with *bvs4* (2^{nd} to 7^{th} well) or *bvs5* (9^{th} to 14^{th} well); Fig 4.8.b: *Rlv*.3841 *bvs5* mutant strains complemented with *bvs4* (2^{nd} to 7^{th} well) or *bvs5* (9^{th} to 14^{th} well).

4.3.3. Nodulation and Nodulation competition assays using mutants

4.3.3.1. Dry weight of plants inoculated with pure culture of mutant strains and wild-type *Rlv*. 3841.

The nodulation test was performed to study the effect of mutation of chromosomal nitrilase and the five *bvs* genes (independently) on the ability of the mutants to form nodules on the host plants and effect on plant growth as measured by its dry weight.

Table 4.3. Dry weight data for *V. sativa* plants included in nodulation test to study effect of gene mutation on nodulation and plant growth. All measurements are expressed in grams.

Culture	Replicate1	Replicate2	Replicate3	Replicate4	Replicate5	Mean
ChrN mutant	0.0576	0.0862	0.1082	0.0691	0.1007	0.0760
bvs1 mutant	0.0730	0.0832	0.0950	0.0892	0.0850	0.0851
bvs2 mutant	0.1794	0.0649	0.0906	0.0684	0.0702	0.0947
bvs3 mutant	0.0972	0.0556	0.0924	0.0969	0.0596	0.0803
bvs4 mutant	0.0940	0.0509	0.1033	0.0806	0.0427	0.0743
bvs5 mutant	0.0769	0.0820	0.1132	0.0428	0.0820	0.0794
Rlv. 3841	0.0576	0.0862	0.1082	0.0691	0.1007	0.0844
Negative control	0.0302	0.0306	0.0361	0.0334	0.0365	0.0334

The analysis of the data by the Kolmogorov-Smirnov test showed that the plant dry weights were normally distributed. Hence, the data were analysed using the one-way ANOVA test.

Table 4.4. Results of the ANOVA test for *V. sativa* dry weight data.

ANOVA					
V. sativa dry weight					
	Sum of Squares	df	Mean Square	F	Sig.
Between cultures	.012	7	.002	2.681	.026

The results of the analysis indicates that there is a significant difference in the means, F (7, N = 40) = 2.681, p = 0.026.

Since p-value = 0.026 ≤ 0.05 = α, we reject the null hypothesis and accept the alternate hypothesis.

At α = 0.05 level of significance, there exists enough evidence to conclude that there is significant difference in the mean dry weights of the *V. sativa* plants inoculated with different cultures.

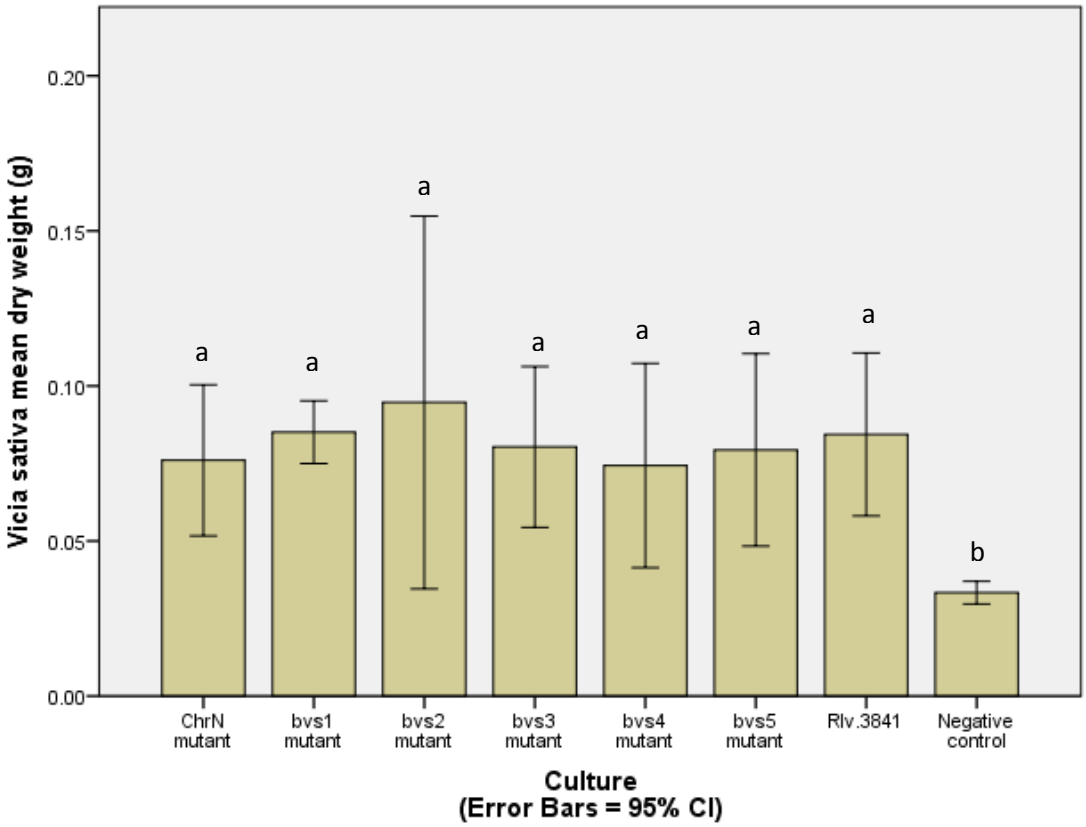


Figure 4.9. Differences in the mean dry weight of *V. sativa* plants used in the nodulation test from ANOVA.

In order to analyse the pattern of difference between means, the one-way ANOVA was followed by a post-hoc LSD (least significant difference) test to carry out specific pairwise comparisons.

The analysis of the dry weight data by LSD showed that amongst all the groups, the only group that differed from other groups was the negative control. There was no difference in the pairwise comparisons of any other groups.

Table 4.5. Dry weight data for *V. faba* plants included in nodulation test to study effect of gene mutation on nodulation and plant growth. All measurements are expressed in grams.

Culture	Replicate1	Replicate2	Replicate3	Mean
ChrN mutant	1.7119	1.6235	1.6607	1.6654
bvs1 mutant	1.2495	1.4715	1.7275	1.4828
bvs2 mutant	1.2036	0.9076	0.6609	0.9240
bvs3 mutant	0.9028	1.2004	1.0040	1.0357
bvs4 mutant	1.5447	1.0970	1.2385	1.2934
bvs5 mutant	1.4405	1.4787	1.9332	1.6175
Rlv. 3841	1.9853	1.5365	1.2641	1.5953
Negative control	1.0518	0.0000	0.5303	0.5274

The analysis of the data by the Kolmogorov-Smirnov test showed that the data of plant dry weights were normally distributed. Hence, the data was analysed using the one-way ANOVA test.

Table 4.6. Results of the ANOVA test for *V. faba* dry weight data.

ANOVA					
V. faba dry weight					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.464	7	.495	5.716	.002

The results of the analysis indicates that there is a significant difference in the means, F (7, N = 24) = 5.716, p = 0.02.

Since p-value = 0.02 ≤ 0.05 = α, we reject the null hypothesis and accept the alternate hypothesis.

At α = 0.05 level of significance, there exists enough evidence to conclude that there is significant difference in the mean dry weights of *V. faba* plants inoculated with different cultures.

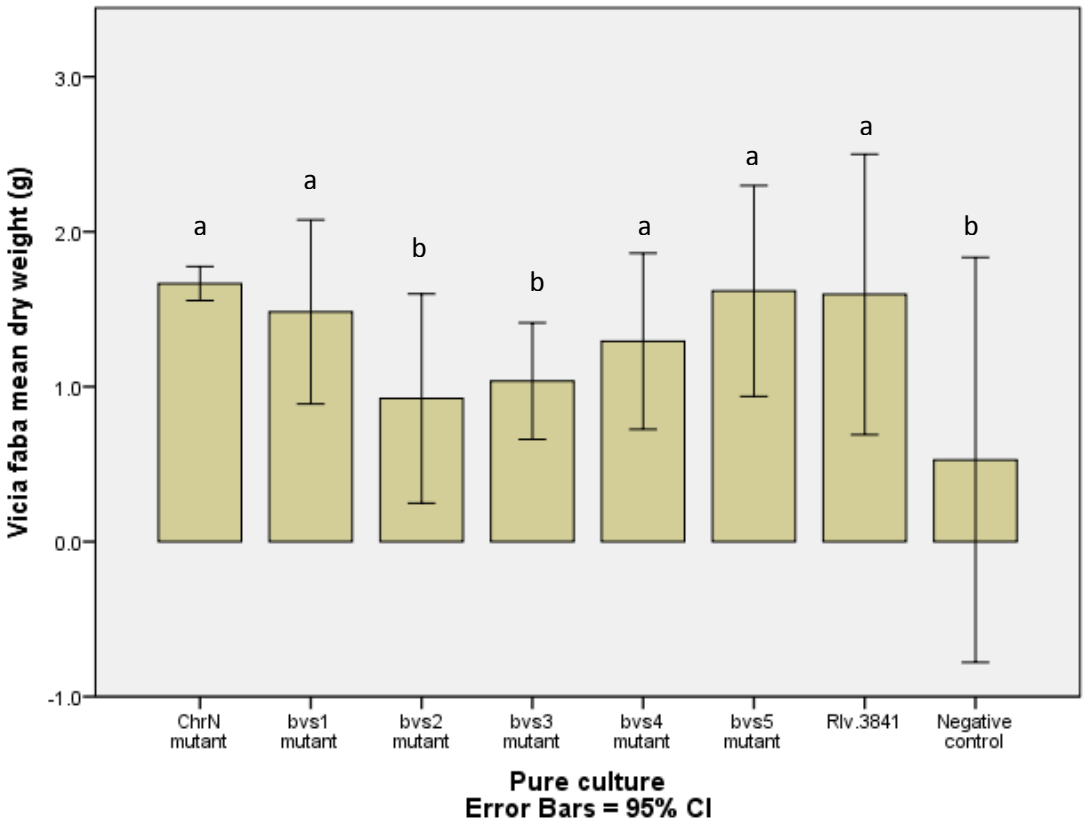


Figure 4.10. Differences in the mean dry weight of *V. faba* plants used in the nodulation test from ANOVA.

In order to analyse the pattern of difference between means, the one-way ANOVA was followed by a post-hoc LSD (least significant difference) test to carry out specific pairwise comparisons.

The analysis of the dry weight data by LSD showed that the means of dry weights of plants from the *bvs2*, *bvs3* mutants and negative control were similar to each other and differed from plants from the other groups. Similarly, the means of the dry weights of plants from ChrN, *bvs1*, *bvs4*, *bvs5* mutants and positive control were similar to each other but different from plants in the other three groups.

The results indicate that the effect of mutating the *bvs2* and *bvs3* is similar to the plants included in the negative control which did not receive any bacterial inoculum i.e. it adversely affected plant growth. The effect of mutating the ChrN, *bvs1*, *bvs4* and *bvs5* genes had no observable effect on plant growth as there was no statistical difference between plants inoculated with bacteria having mutation in these genes and the plants inoculated with *Rlv*. 3841 (positive control).

Table 4.7. Dry weight data for *P. sativum* plants included in nodulation test to study effect of gene mutation on nodulation and plant growth. All measurements are expressed in grams.

Culture	Replicate1	Replicate2	Replicate3	Mean
ChrN mutant	0.3825	0.4750	0.4256	0.4277
bvs1 mutant	0.3498	0.4115	0.4038	0.3884
bvs2 mutant	0.2349	0.4732	0.5846	0.4309
bvs3 mutant	0.5674	0.4365	0.2495	0.4178
bvs4 mutant	0.5126	0.4345	0.6523	0.5331
bvs5 mutant	0.4355	0.3926	0.3516	0.3932
Rlv. 3841	0.4516	0.3964	0.3555	0.4012
Negative control	0.4127	0.3402	0.3217	0.3582

The analysis of the data by the Kolmogorov-Smirnov test showed that the data of plant dry weights were normally distributed. Hence, the data was analysed using the one-way ANOVA test.

Table 4.8. Results of the ANOVA test for *P. sativum* dry weight data.

ANOVA					
P. sativum dry weight					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.057	7	.008	.816	.588

The results of the analysis indicates that there is a significant difference in the means, F (7, N = 24) = 0.816, p = 0.558.

Since p-value = 0.558 ≥ 0.05 = α, we do not reject the null hypothesis.

At α = 0.05 level of significance, there exists enough evidence to conclude that there is no significant difference in the mean dry weights of *P. sativum* plants inoculated with different cultures.

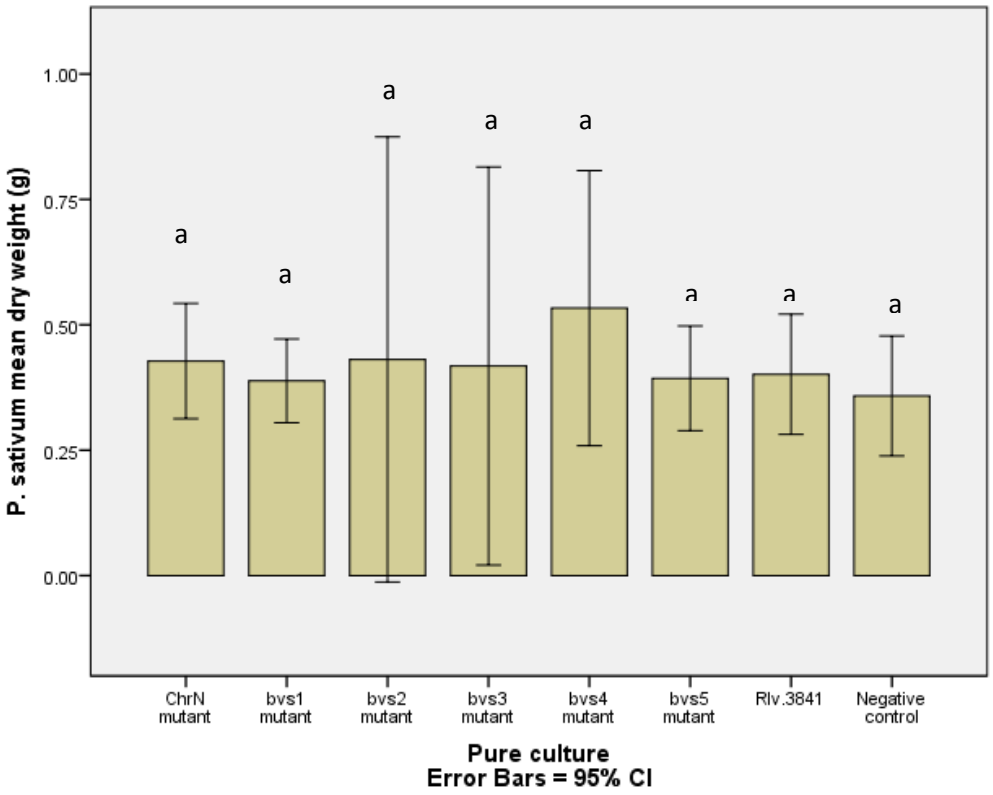


Figure 4.11. Differences in the mean dry weight of *P. sativum* plants used in the nodulation test from ANOVA.

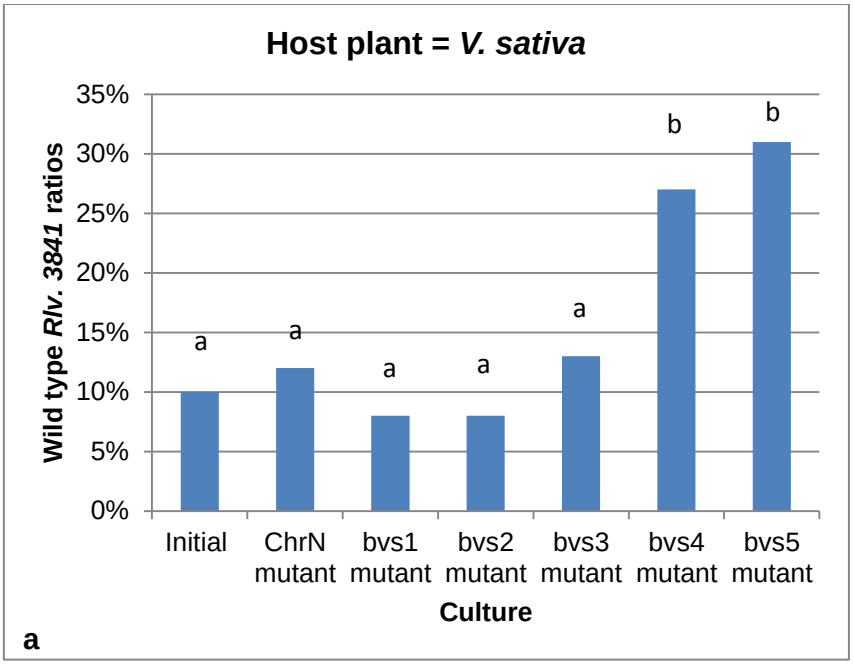
In the pea plants (*P. sativum*) no difference was observed in the growth of any plants used in the test, including negative control. In contrast, the *V. sativa* plants showed difference only between the plants included in the negative control and other groups indicating that while the mutation of the genes did not have any effect on plant growth, the absence of the nodulating bacteria was detrimental to the growth of the plant. The *V. faba* plants differed to a much greater extent. It was observed that plants infected with bacteria having mutation in *bvs2* and *bvs3* showed growth that was statistically similar to that of negative control whereas the plants that were treated with cultures mutated in genes ChrN, *bvs1*, *bvs4* and *bvs5* were not different from plants included in the *Rlv*. 3841 positive control group.

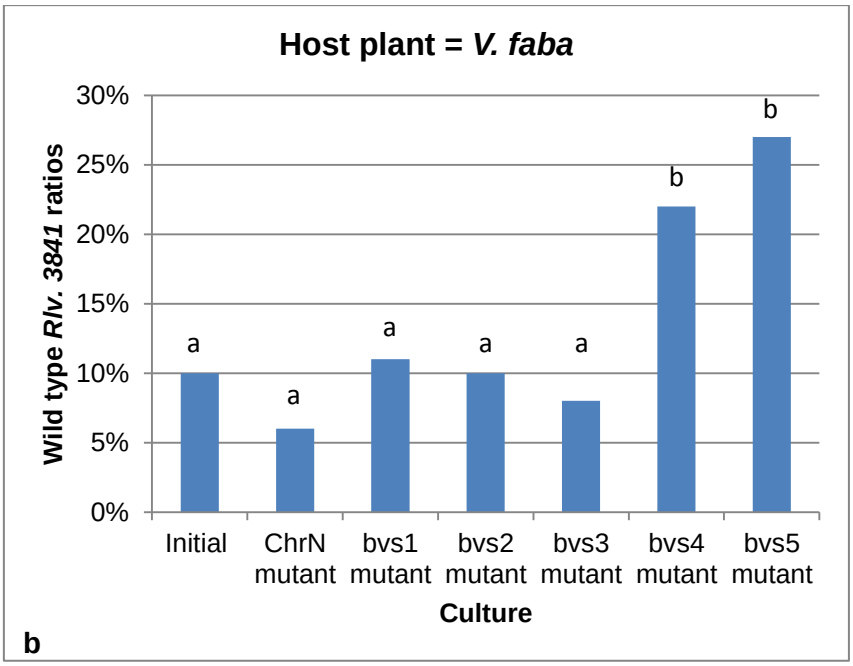
4.3.3.2. Results of nodulation competition assay between wild-type *Rlv*. 3841 and gene mutants

The results of the nodulation test gave a mixed verdict in terms of the ability of gene mutation to affect the growth of plants. The effect of the

mutation on plant growth differed with the host. The same host plants were hence included in the nodulation competition assay and inoculated with a mixed bacterial culture containing wild-type *Rlv*. 3841 and mutants in the ratio of 1:10 (or 10% of the population) to study the effect of mutation on the ability of the strain to compete for nodulation.

The results of the nodulation competition assay carried out on different host plants are shown in Figure 4.12.





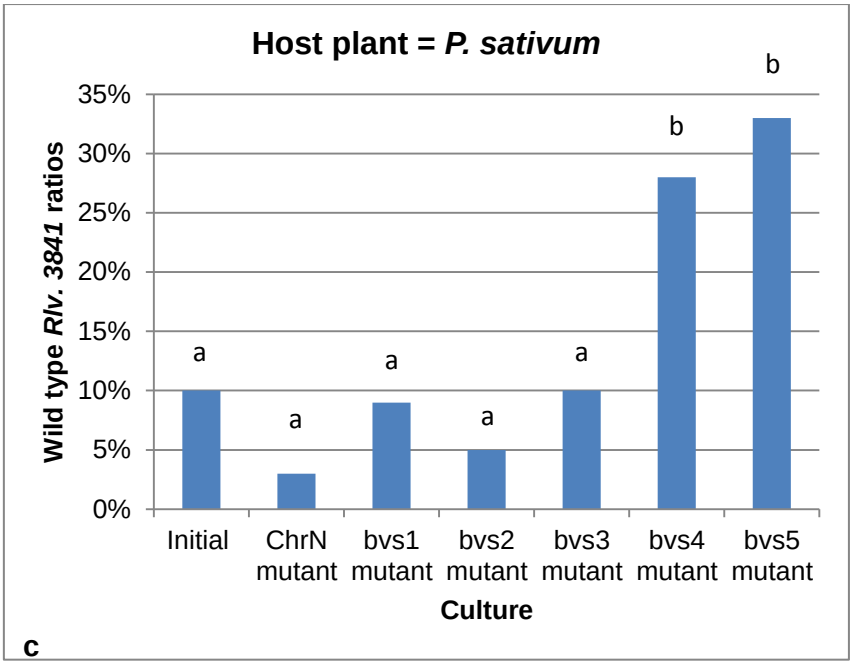


Figure 4.12. Comparison of wild-type *Rlv*. 3841 mutants at the beginning and end of the nodulation competition assay for the host plants *V. sativa, V. faba* and *P. sativum* (Figures 4.12 a, b and c respectively). The initial bar is representational and shows the estimated percentage of wild-type *Rlv*. 3841 population at the beginning of the assay (i.e. 10%). The other bars show the actual percentage of wild-type *Rlv*. 3841 population at the end of the assay when mixed with the cultures indicated at the base of the bar.

The data from the nodulation competition assay was analysed using the chi-squared test to see if there is a significant difference in the initial and final values of the proportion of wild-type *Rlv*. 3841 populations in the nodulation competition assay.

The chi-squared test is a statistical method used to assess the goodness of fit between a set of observed values and those expected theoretically or, in other words, to evaluate how likely it is that any observed difference between the observed and the expected values arose by chance. The chi-squared test here refers to the Pearson's chi-squared test, also known as the chi-squared goodness-of-fit test.

Table 4.9. Chi-squared analysis of wild type *Rlv*.3841 nodulation ratios at the end of the nodulation competition assay in the three host plants using mixed cultures to test whether the proportion of wild-type nodules at the end differ significantly from the expected value of 1 in 10 (Significant results highlighted with a grey background).

***a. Vicia sativa* nodulation competition assay chi-squared test results**			
ChrN mutant	χ-squared = 0.1818	df = 1	p-value = 0.6698
bvs1 mutant	χ-squared = 0.2222	df = 1	p-value = 0.6374
bvs2 mutant	χ-squared = 0.2222	df = 1	p-value = 0.6374
bvs3 mutant	χ-squared = 0.3913	df = 1	p-value = 0.5316
bvs4 mutant	χ-squared = 7.8108	df = 1	p-value = 0.005193
bvs5 mutant	χ-squared = 10.7561	df = 1	p-value = 0.001039

***b. Vicia faba* nodulation competition assay chi-squared test results**			
ChrN mutant	χ-squared = 1	df = 1	p-value = 0.3173
bvs1 mutant	χ-squared = 0.0476	df = 1	p-value = 0.8273
bvs2 mutant	χ-squared = 0	df = 1	p-value = 1
bvs3 mutant	χ-squared = 0.2222	df = 1	p-value = 0.6374
bvs4 mutant	χ-squared = 4.5	df = 1	p-value = 0.03389
bvs5 mutant	χ-squared = 7.8108	df = 1	p-value = 0.005193

***c. Pisum sativum* nodulation competition assay chi-squared test results**			
ChrN mutant	χ-squared = 3.7692	df = 1	p-value = 0.0522
bvs1 mutant	χ-squared = 0.0526	df = 1	p-value = 0.8185
bvs2 mutant	χ-squared = 1.6667	df = 1	p-value = 0.1967
bvs3 mutant	χ-squared = 0	df = 1	p-value = 1
bvs4 mutant	χ-squared = 8.5263	df = 1	p-value = 0.0035
bvs5 mutant	χ-squared = 12.3023	df = 1	p-value = 0.0004524

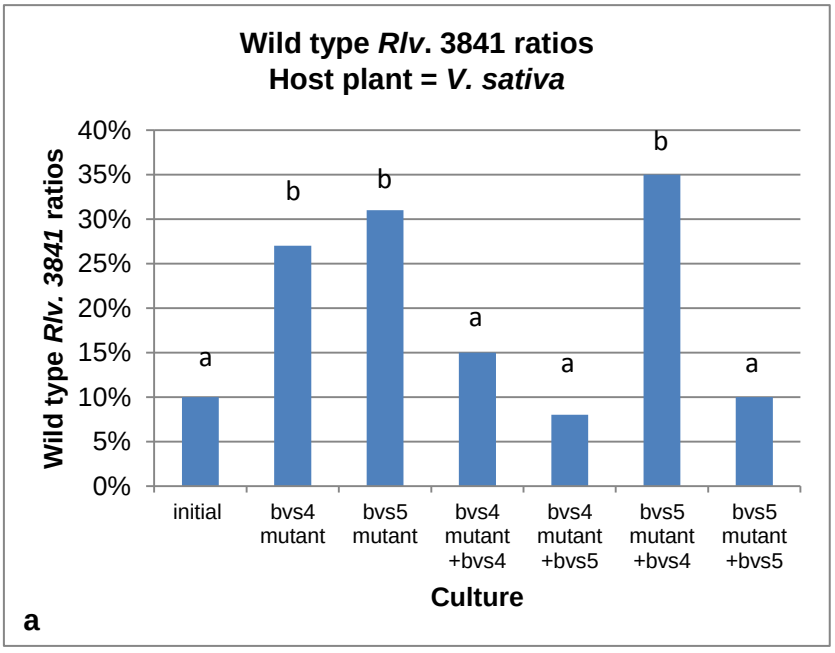
From the results of the chi-squared tests in Table 4.9, it is clear that mutating the ChrN, *bvs1*, *bvs2* and *bvs3* had no effect on the ability of the mutants to compete for nodulation in all the three host plant *V. sativa, V. faba* and *P. sativum* ($p \geq 0.05$). However mutating the *bvs4* and *bvs5* genes adversely affected the ability of the mutants to compete and form nodules on all the three host plants ($p \leq 0.05$).

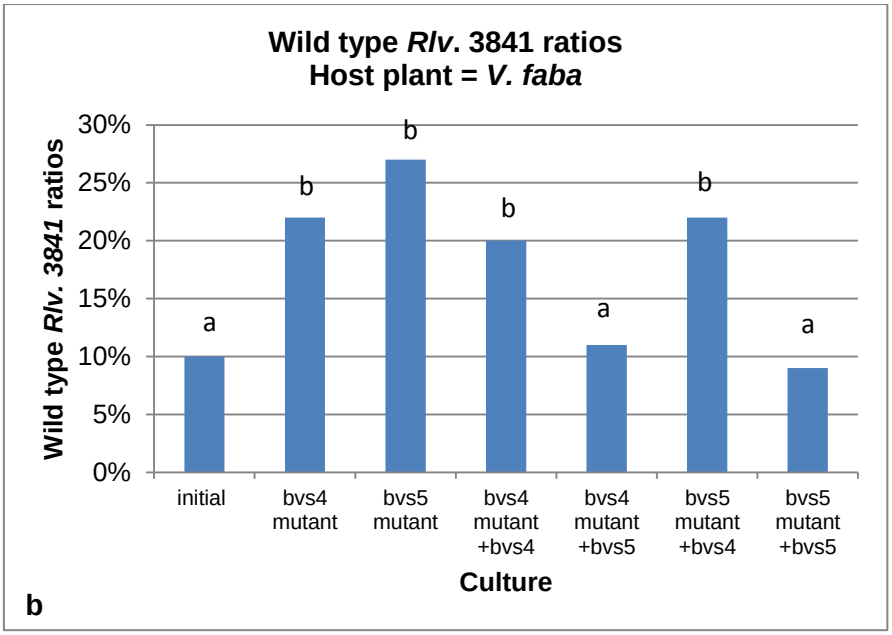
The results from the nodulation tests and from the nodulation competition tests could be used to argue that while the plants exhibited different growth response when inoculated with pure cultures of mutants in the nodulation test, the effect of mutation on the nodulation competition is the same in all the host plants viz. mutating the *bvs4* and *bvs5* genes adversely affects the ability of the bacterium to compete and form nodules on the host plant.

4.3.4. Nodulation competition assay using mutant gene complemented strains

4.3.4.1. Results of nodulation competition assay between wild-type *Rlv*. 3841 and mutant gene complemented strains

The results of the nodulation competition assay are shown in Figure 4.13.





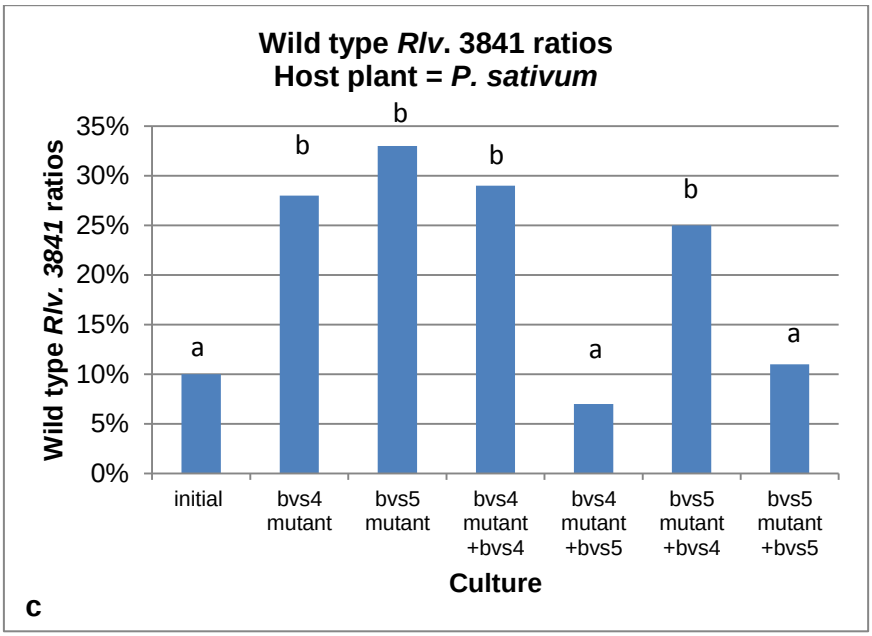


Figure 4.13. Comparison of wild-type *Rlv*. 3841 mutants at the beginning and end of the nodulation competition assay for the host plants *V. sativa, V. faba* and *P. sativum* (Figures 4.13 a, b and c respectively). The initial bar is representational and shows the estimated percentage of wild-type *Rlv*. 3841 population at the beginning of the assay (i.e. 10%). The other bars show the actual percentage of wild-type *Rlv*. 3841 population at the end of the assay when mixed with the cultures indicated at the base of the bar.

The data from the nodulation competition assay was analysed using the chi-squared test to see if there is a significant difference in the initial and final values of the proportion of wild-type *Rlv*. 3841 populations in the nodulation competition assay.

Table 4.10. Chi-squared analysis of wild type *Rlv*.3841 nodulation ratios at the beginning and the end of the nodulation competition assay in the three host plants using mixed cultures (Significant results highlighted with a grey background).

a. Vicia sativa **nodulation competition assay chi-squared test results**			
bvs4 mutant	χ-squared = 7.8108	df = 1	p-value = 0.005193
bvs5 mutant	χ-squared = 10.7561	df = 1	p-value = 0.001039
bvs4 mutant + *bvs4*	χ-squared = 1	df = 1	p-value = 0.3173
bvs4 mutant + *bvs5*	χ-squared = 0.2222	df = 1	p-value = 0.6374
bvs5 mutant + *bvs4*	χ-squared = 13.8889	df = 1	p-value = 0.0001939
bvs5 mutant + *bvs5*	χ-squared = 0.2222	df = 1	p-value = 1

b. Vicia faba **nodulation competition assay chi-squared test results**			
bvs4 mutant	χ-squared = 4.5	df = 1	p-value = 0.03389
bvs5 mutant	χ-squared = 7.8108	df = 1	p-value = 0.005193
bvs4 mutant + *bvs4*	χ-squared = 3.3333	df = 1	p-value = 0.06789
bvs4 mutant + *bvs5*	χ-squared = 0.0476	df = 1	p-value = 0.8273
bvs5 mutant + *bvs4*	χ-squared = 4.5	df = 1	p-value = 0.03389
bvs5 mutant + *bvs5*	χ-squared = 0.0526	df = 1	p-value = 0.8185

c. Pisum sativum **nodulation competition assay chi-squared test results**			
bvs4 mutant	χ-squared = 8.5263	df = 1	p-value = 0.0035
bvs5 mutant	χ-squared = 12.3023	df = 1	p-value = 0.0004524
bvs4 mutant + *bvs4*	χ-squared = 9.2564	df = 1	p-value = 0.002347
bvs4 mutant + *bvs5*	χ-squared = 0.5294	df = 1	p-value = 0.4669
bvs5 mutant + *bvs4*	χ-squared = 6.4286	df = 1	p-value = 0.01123
bvs5 mutant + *bvs5*	χ-squared = 0.0476	df = 1	p-value = 0.8273

From the results of the chi-squared test in Table 4.10, it can be said that in all the three host plants, mutating the *bvs4* and *bvs5* genes adversely affected the nodule occupancy ratio of the mutant ($p \leq 0.05$). It was also observed that when the *bvs4* mutant was complemented with a functional copy of *bvs4* gene, the effect of mutation on the nodulation competition ability was reversed in *V. sativa* and *V. faba* ($p \geq 0.05$) but not in *P. sativum* ($p \leq 0.05$). However, when the *bvs4* mutant was complemented with a copy of *bvs5* gene, the mutant strain regained its ability to compete and form nodules ($p \geq 0.05$) in all the three host plants. When *bvs5* gene mutant was complemented with a copy of *bvs4* gene, the ability to compete for nodulation was not restored in any of the three host plants ($p \leq 0.05$). However, complementing the mutation in *bvs5* with a copy of functional *bvs5* gene significantly restored the ability of the mutants to compete and form nodules on all the three host plants ($p \geq 0.05$).

4.4. Discussion

In the previous chapter it was reasoned that since the *bvs* genes were only found in the strains that belonged to the biovar *viciae* of the *Rhizobium leguminosarum* species complex they may have some role in the nodulation process. The aim of this chapter was to investigate this line of reasoning.

In order to study the effect that *bvs* genes have on nodulation, mutants were isolated for each of the five *bvs* genes and the chromosomal nitrilase gene and used to perform nodulation tests with three of the natural hosts of the biovar *viciae* group viz. *V. sativa, V. faba* and *P. sativum*. The three plants showed different growth reaction (in terms of plant dry weight) to the mutation of the genes. *P. sativum* did not exhibit reaction even to the presence of inoculants whereas *V. sativa* showed difference between the inoculated and uninoculated plants (negative control) but no difference between the mutated and not mutated (positive control) inoculum. The *V. faba* plants infected with bacteria having mutation in *bvs2* and *bvs3* showed growth that was statistically similar to that of negative control whereas the plants that were treated with cultures mutated in genes ChrN, *bvs1*, *bvs4* and *bvs5* were not different from plants included in the *Rlv*. 3841 positive control group.

When the mutants were used to test their ability to compete with the wild-type *Rlv*. 3841 strain, it was observed that mutation in the genes ChrN, *bvs1*, *bvs2* and *bvs3* did not have any effect on the ability of the mutants to compete with the wild-type strain. However, mutating the *bvs4* and *bvs5* genes greatly diminished the ability of these strains to compete and form nodules on all the three host plants.

The results suggest that mutating the *bvs* genes might not have a significant effect on the host plant; the effect would perhaps range from no effect to significant effect depending on the host plant species. What is more striking probably is the effect of mutating the *bvs* genes on nodulation competition.

The effects of mutating the *bvs4* and *bvs5* genes were seen clearly and verified statistically. The nodulation competition assay using the strains complemented with functional genes showed that complementing the mutants with a functional copy of *bvs5* gene significantly restored the nodulation competition ability of the mutant strains irrespective of gene mutated (*bvs4* or *bvs5*). This suggests a greater role for *bvs5* gene in conferring nodulation competition ability on the biovar *viciae* strains. It may also suggest that if the *bvs4* and *bvs5* genes form an operon as suggested by using bioinformatic tools, then the *bvs5* gene is positioned transcriptionally downstream to the *bvs4* gene thus accounting for the polarity of mutation as suggested by the results of complementation analysis.

Previously competition tests have demonstrated that some genes were able to affect rhizobial competitiveness. Triplett (1990) found that *tfx* genes in *R. leguminosarum* bv. *trifolii* T24 play an important role in biosynthesis of trifolitoxin. This compound was found to inhibit nodulation of clover by other *R. leguminosarum* bv. *trifolii* strains. Transformation of this gene to other *Rhizobium* strains could also enhance the competitiveness. Using nodulation competition tests, Miller *et al.* (2007) demonstrated that in *R. leguminosarum* bv. *viciae,* the cluster of genes involved in chemotaxis, *che1*, could be linked to competitiveness. Using the same technique, Williams *et al.* (2008) showed that genes mediating root hair surface attachment in *R. leguminosarum* were important for competition. Eda *et al.* (2011) demonstrated that in *Sinorhizobium meliloti*, gene cluster smeAB, encoding a multidrug efflux pump increased resistance of rhizobia to antibiotics, detergents and plant-derived compounds, affecting competitiveness.

Genes involved in utilization of galactose and trehalose were also demonstrated to affect competitiveness in *Mesorhizobium loti* and *Sinorhizobium meliloti* (Ampomah and Jensen, 2013, Geddes and Oresnik, 2012). In my work, *bvs5* was found to affect rhizobial competitiveness since sulfite oxidase might increase detoxification ability to support formation of symbiosis. The results confirmed the notion that genes which are related to process of nodulation and symbiosis formation are essential for rhizobial competition, as indicated by Pérez-Giménez *et al.* (2011).

The next logical step in the study would be to investigate the biochemical / enzymatic functions of the *bvs4* and *bvs5* genes. In *Rlv*. 3841, the genes are annotated as an aliphatic nitrilase (*bvs4*) and a molybdenum-binding oxidoreductase (*bvs5*) that could either be a sulfite oxidase or nitrate reductase. Since these annotations are based on sequence similarities of the genes with those found in other organisms in which the enzyme activity may have been practically demonstrated, it would be a good idea to use these annotations as the starting point to investigate the biochemical / enzymatic functions of these genes.

CHAPTER 5 : THE *bvs4* AND *bvs5* GENES ENCODE A NITRILASE AND A SULFITE OXIDASE.

5.1. Introduction

The *bvs4* and *bvs5* genes are two of five genes found in *Rlv*. 3841 and all biovar *viciae* isolates from Wentworth College, but none of the biovar *trifolii* isolates from the same area. In *Rhizobium leguminosarum* biovar *viciae* strain 3841, the five genes are located on pRL8, the smallest plasmid. The arrangement of the five *bvs* genes on pRL8 in *Rlv*. 3841 (pRL80073 – pRL80077) is shown below in Figure 5.1 :

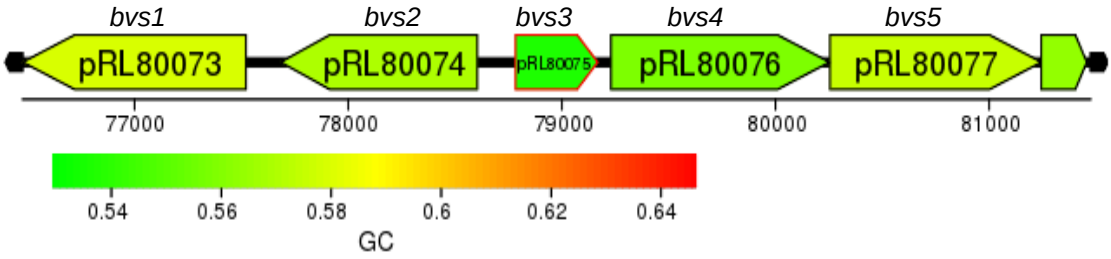


Figure 5.1. The arrangement of *bvs* genes on pRL8 in *Rlv*. 3841 coloured by GC content.

The mutation of *bvs4* and *bvs5* genes and the analysis of the results obtained from competition assays in the previous chapter show that the *bvs4* and *bvs5* genes affect the ability of *Rlv*. 3841 to compete for nodulation of the host plant compared to the native *Rlv*. 3841 isolate. The mutation of other three genes did not demonstrate any significant difference in the ability of the bacteria to nodulate the host plants. This indicates that the products of *bvs4* and *bvs5* expression are required by the bacteria in order to effectively compete for nodulation of the host plant.

The putative functions of the *bvs4* and *bvs5* genes have been postulated using bioinformatic tools but have not been experimentally verified (Young *et al.*, 2006). In the previous chapter, we have determined that the products of *bvs4* and *bvs5* are essential in the symbiotic association by mutating the genes and complementing them and testing the effect on plant growth. It is therefore necessary to investigate the enzymatic functions of these genes in order to comment on their role in the symbiosis. The best point to start, perhaps, is to start by investigating the putative functions of the genes as postulated by the bioinformatic analysis and used in annotating the genome of *Rlv*. 3841.

5.1.1. Putative functions of *bvs4* and *bvs5*

The *bvs4* and *bvs5* genes in the *Rlv*. 3841 genome are annotated as a putative aliphatic nitrilase and a putative molybdenum-binding oxidoreductase respectively. Hence, a short review of the structure and possible functions of the two enzymes is presented here, based on the annotation of the genes on the *Rlv*. 3841 genome.

5.1.1.1. *bvs4* : putative aliphatic nitrilase (pRL80076)

The *bvs4* gene is a 1.33 kb gene located on the plasmid pRL8 of *Rlv*. 3841. The sequence putatively encodes a 340 amino acid containing aliphatic nitrilase similar to the aliphatic nitrilase of *Rhodococcus rhodochrous* (Young *et al.*, 2006).

The nitrilases belong to the nitrilase superfamily (Pace and Brenner, 2001). The enzymes of this superfamily are also called as carbon-nitrogen hydrolases (CN-hydrolases) first described by Bork and Koonin (1994) and belong to the Pfam family PF00795. The C-N hydrolases catalyse the breakdown of non-peptide C-N bonds. Based on sequence identity and catalytic activity, the C-N hydrolases are divided into 13 branches and includes enzymes such as nitrilase, aliphatic amidase, biotinidase, β-ureidopropionase, amongst others (Pace and Brenner, 2001).

Nitrilase, cyanide dihydratase and cyanide hydratase are a group of closely related proteins belonging to the C-N hydrolase family. The proteins of the nitrilase superfamily generally have a conserved glutamic acid, lysine and cysteine (E-K-C) catalytic triad. The role of the amino acid in the catalytic triad has been proposed as follows : cysteine and glutamic acid residues assume the roles of a nucleophile and a general base, respectively, while the lysine residue is involved in the stabilization of the tetrahedral intermediate (Yeom *et al.*, 2008).

The enzymes of the nitrilase branch are multimeric alpha-beta-beta-alpha sandwich proteins and can be differentiated from other members of the superfamily by the presence of a conserved cysteine-tryptophan-glutamic acid (C-W-E) motif positioned at the cysteine residue of the catalytic triad. The cysteine residue is thought to form the active site of the enzyme (Brenner, 2002).

Nitrilases also have a sulfhydryl group that is essential for catalytic activity. Thus, they are also thiol enzymes. Like other thiol enzymes, their activity may be inhibited by thiol binding compounds like silver nitrate and copper sulfate, and enhanced by thiol-reducing agents such as dithiothreitol (Layh *et al.*, 1998, O'Reilly and Turner, 2003, Podar *et al.*, 2005)

These enzymes, although similar in their amino acid composition and protein structure, differ in their catalytic ability. Nitrilases catalyse the hydrolysis of nitriles to carboxylic acid and ammonia. They are produced by a variety of organisms and vary significantly in their specificity (O'Reilly and Turner, 2003). Cyanide hydratase preferentially hydrolyses cyanide to formamide while cyanide dihydratase specifically hydrolyses cyanide to formic acid and ammonia (O'Reilly and Turner, 2003, Singh *et al.*, 2006).

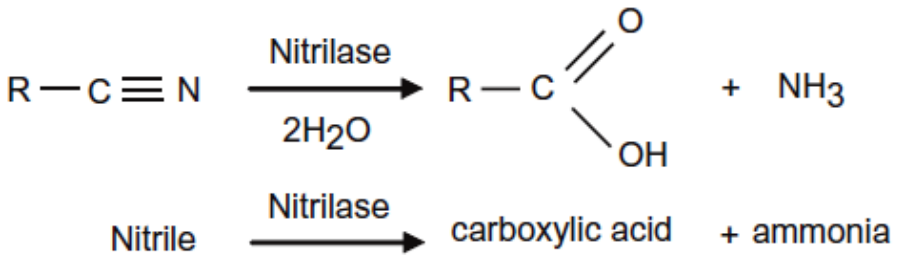


Figure 5.2. The nitrilase reaction. Nitrilase catalyses the hydrolysis of nitriles to the corresponding carboxylic acid plus ammonia.

Nitrilases can be grouped into three categories based on their substrate specificity viz. aliphatic nitrilases (acting on aliphatic nitriles such as acrylonitrile), aromatic and heterocyclic nitrilases (acting on aromatic or heterocyclic nitriles such as benzonitrile) and arylacetonitrilases (acting on arylacetonitriles such as indole-3-acetonitrile and phenylpropio-nitrile) (Brenner, 2002, O'Reilly and Turner, 2003). Some nitrilases are extremely substrate specific, for example, the nitrilase of *Klebsiella pneumoniae* sp. *ozaenae* (McBride *et al.*, 1986), catalyses the hydrolysis of bromoxynil. Others nitrilases, in contrast, have a broad substrate range, such as the nitrilase of *Bacillus pallidus* Dac521, which hydrolyses aromatic, aliphatic and heterocyclic nitriles (Almatawah *et al.*, 1999).

Nitrilases are produced by a wide range of organisms, including bacteria, fungi and plants, but their metabolic role is poorly understood. Nitrilases may play important roles in a number of biological processes.

Nitrilase activity was first detected in plants by Thimann and Mahadevan (1958) and in soil bacteria by Robinson and Hook (1964). Many microbial nitrilases from the nitrilase superfamily have since been identified in organisms isolated from cyanide or nitrile-contaminated land and water. The presence of the enzymes was shown to increase the cyanide and nitrile tolerance of the organisms that produced them suggesting that these enzymes are probably involved in cyanide and nitrile detoxification.

Microbes may also use nitrilase activity for the detoxification and assimilation of nitriles and cyanide present in the plant environment. Nitrile compounds are abundant in the plant environment and since the enzymes are produced by plants and soil bacteria it has been suggested that the activity of the enzymes might play an important role in interaction between plant and bacteria. It has been suggested that the nitrilases might indeed form part of an array of mechanisms that facilitate microbial colonization of plants, with possible roles in plant hormone synthesis, nitrogen utilization, the catabolism of cyanogenic glycosides and glucosinolates and the detoxification of nitriles and cyanide (Howden *et al.*, 2009a, Kiziak *et al.*, 2005, O'Reilly and Turner, 2003).

5.1.1.2. *bvs5* : putative Mo-binding oxidoreductase (pRL80077)

The *bvs5* gene is a 1.28 kb gene located on the plasmid pRL8 of *Rlv*. 3841. The sequence putatively encodes a 329 amino acid oxidoreductase containing a molybdenum cofactor (Moco) binding domain. The gene shows sequence similarity with genes that encode enzymes involved in sulphur metabolism in extremophiles like *Paracoccus denitrificans* and *Thermus thermophilus*.

The molybdenum cofactor is formed as a result of binding of the element molybdenum to a special co-factor molecule, called pterin. The Mo is coordinated to a dithiolene group on the 6-alkyl side chain of the pterin to form molybdopterin (MPT). The metal itself is biologically inactive. The binding results in gain of catalytic activity. The molybdenum atom forms coordination with a cysteine ligand of the protein and catalyses transfer of oxygen to or from a lone pair of electrons on the substrate (Hille *et al.*, 2014, Mendel, 2009, Mendel, 2013, Schwarz, 2005, Schwarz and Mendel, 2006).

More than 50 molybdoenzymes have been identified in bacteria to date. They are involved in a wide variety of transformation by enzymes catalysing diverse key reactions in the global carbon, sulfur and nitrogen metabolism which include the transfer of an oxo group or two electrons to or from the substrate. The Moco is the catalytically active region found at the catalytic site of all molybdoenzymes, except bacterial Mo-nitrogenase where Mo is a constituent of the FeMo-cofactor (Mendel, 2013, Schwarz and Mendel, 2006). Moreover, all Mo-enzymes, except plant sulfite oxidase, need at least one more redox active centre, many of them involving iron in electron transfer (Mendel and Bittner, 2006).

The Molybdenum cofactor is synthesized via an ancient, ubiquitous and highly conserved pathway. The biosynthetic pathway of Moco involves interaction of seven proteins along with iron, ATP, SAM, copper and Mg^{+2}. The synthesized Moco is most likely distributed to the apoproteins of Mo-enzymes by putative Moco-carrier proteins. Since it is a part of many important proteins with enzymatic activities, a deficiency in the biosynthesis of Moco results in a pleiotropic loss of all Mo-enzyme activities and can be lethal for the organism (Mendel, 2013, Mendel and Schwarz, 2002, Schwarz and Mendel, 2006).

The synthesis of Moco can be divided into four steps viz. conversion of GTP into cPMP, synthesis of the molybdopterin dithiolate, adenylylation of molybdopterin dithiolate and finally, insertion of molybdenum to form Moco. Additional modification of Moco leading to its maturation as a result of modifications can be described as the fifth step in Moco synthesis. The modifications result in different catalytic Mo centres. (Schwarz G 2005).

The molybdenum cofactor exists in three basic configurations at the molybdenum atom (Kisker *et al.*, 1997). Hence, the molybdoenzymes are categorized into three families based on the type of cofactor bound, which is mainly classified by the ligands at the molybdenum atom : the xanthine oxidase family, the DMSO reductase family and the sulfite oxidase family (Hille, 1996, Hille *et al.*, 2014).

The xanthine oxidase family contains a sulfurated molybdopterin cytosine dinucleotide cofactor. This final maturation step is catalysed by a Moco-sulfurase enzyme, which mobilizes sulfur from L-cysteine in a pyridoxal

phosphate-dependent manner typical of cysteine desulfurases. Enzymes of the xanthine oxidase family catalyse the oxidation of hypoxanthine to xanthine and further oxidation of xanthine to uric acid. These enzymes play an important role in the catabolism of purines in some species. Members of the xanthine oxidase family include xanthine dehydrogenase and periplasmic aldehyde oxidoreductase.

The DMSO reductase family contains two molybdopterin guanine dinucleotides ligated to one molybdenum atom with additional ligands being an O/S, and a sixth ligand X, which can be a serine, a cysteine, a selenocysteine, an aspartate or a hydroxide and/or water molecule. Members of the DMSO reductase family are found only in prokaryotes and include nitrate reductases, formate dehydrogenases, trimethylamine-N-oxide (TMAO) reductases, DMSO reductase and biotin sulfoxide reductases.

Prokaryotic nitrate reductases are molybdoenzymes that reduce nitrate (NO^{-3}) to nitrite (NO^{-2}). Organisms reduce nitrate for three main reasons : to incorporate nitrogen into biomolecules (assimilatory nitrate reduction); to generate energy for cellular function (respiration, denitrification) or to eliminate energy excess generated by the cell metabolism (dissimilatory nitrate reduction). Most nitrate reductases are mononuclear molybdenum-containing enzymes which have, in addition to molybdenum active site, additional redox cofactors that mediate electron transfer reactions between the electron donor and the electron acceptor (nitrate) (Gonzalez *et al.*, 2006, Hille, 2013, Hille *et al.*, 2014, Tavares *et al.*, 2006).

Prokaryotic nitrate reductases can be classified into three groups (according to different criteria such as: cell localization, protein structure and molecular properties of the catalytic centre and metabolic routes) viz. assimilatory nitrate reductases (Nas), respiratory nitrate reductase (Nar), and periplasmic nitrate reductases (Nap). Nas are cytoplasmic enzymes involved in nitrate assimilation, whereas Nar are membrane-bound enzymes involved exclusively in denitrification. In contrast. Nap does not have a well-defined function (Bittner, 2014, Iobbi-Nivol and Leimkuhler, 2013, Mendel and Bittner, 2006, Schwarz and Mendel, 2006, Tavares *et al.*, 2006).

Assimilatory nitrate reductases catalyse the reduction of nitrate to nitrite which is subsequently converted to NH_4^+ by nitrite reductase. The reduction of nitrate starts in the cytoplasm and the process is used by the cell to incorporate nitrogen into biomolecules. The dissimilatory processes involve the conversion of nitrate into N_2 (respiration) or into ammonia (respiration/ammonification). The intermediates of the process are nitrite, nitric oxide and nitrous oxide – which are amongst the most important global warming compounds. Heterotrophic soil bacteria are involved in this process and a very low concentration of oxygen is required (Carlson and Ingraham, 1983). Respiration is used by the cell to generate the proton motive force (PMF) across the cytoplasmic membrane. Dissimilatory ammonification also starts with the reduction of nitrate to nitrite, but then nitrite is reduced to ammonia. (Gonzalez *et al.*, 2006, Iobbi-Nivol and Leimkuhler, 2013, Mendel and Bittner, 2006). Dissimilatory nitrate reduction to ammonia (DNRA) is less common than the process which produces molecular nitrogen. Many rhizobial strains such as *Bradyrhizobium japonicum* have been reported to have enzymes required for this process (Liu *et al.*, 2003). Rütting *et al.* (2011) concluded that the significance of DNRA in the terrestrial nitrogen cycles might be increasing under the changing climatic conditions and that DNRA could act as an indicator for nitrogen retention in ecosystems.

The sulfite oxidase family is characterized by a di-oxo Moco with an additional protein ligand, which usually is a cysteine. The sulfite oxidase family comprises eukaryotic nitrate reductases (Euk-NR) and sulfite-oxidizing enzymes from animals, plants and bacteria. Sulfite oxidase and nitrate reductase are so related that altering three residues of chicken and human sulfite oxidase decreased sulfite oxidase activity and conferred nitrate reductase activity (Qiu *et al.*, 2012, Tavares *et al.*, 2006).

The sulfite-oxidizing enzymes can be separated into two classes, the sulfite oxidases (SO), found in animals and plants (EC 1.8.3.1), and the sulfite dehydrogenases (SDH), found in bacteria (EC 1.8.2.1). Bacterial enzymes of the SO family appear to be located in the bacterial periplasm, i.e. in the extracellular compartment. (Aguey-Zinsou *et al.*, 2003, Kappler, 2011).

Unlike the prokaryotic nitrate reductase, the sulfite oxidases found in plants and animals use molecular oxygen as an electron acceptor whereas the sulfite dehydrogenases of bacteria use other electron acceptors such as cytochrome c. Sulfite oxidase catalyses the terminal reaction in the oxidative degradation of the sulfur-containing amino acids cysteine and methionine. The enzyme also plays an important role in detoxifying exogenously supplied sulfite and sulfur dioxide. The general reaction catalysed by these enzymes is $SO_3^{-2} + H_2O \rightarrow SO_4^{-2} + 2H^+ + 2e^-$, which in the case of oxygen as the electron acceptor for sulfite oxidase leads to the production of hydrogen peroxide, whereas in the case of sulfite dehydrogenases reduced cytochrome c is produced. (Feng *et al.*, 2007, Kappler, 2011).

Bacteria that have sulphur-oxidizing abilities could substantially benefit the neighbouring ecology by converting sulphur-containing compounds to essential nutrients for plants and other bacteria (Anandham *et al.*, 2007). Sulphur was found to be important in nitrogen fixation in *Vicia faba* and chickpea (Habtegebrial Habtemichial *et al.*, 2007, Togay *et al.*, 2008). In rhizobia, a number of enzymes are linked to sulphur. Fe-S clusters are found in nitrogenase and ferredoxin, which are essential for nitrogen fixation (Tabatabai, 1986). Several rhizobia strains have sulphur-oxidizing enzymes and form symbiosis under stressful conditions (El-Tarabily *et al.*, 2006, Ghosh and Roy, 2006). The evidence suggests that sulphur-oxidizing enzymes may support the rhizobial symbiosis and nodulation processes.

Enzyme assays in this chapter are designed to observe the enzymatic activities of *bvs4* and *bvs5* genes. Based on BLAST the results from NCBI, *bvs4* and *bvs5* might have nitrilase and nitrate reductase/sulphite oxidase activities, respectively. Mutation of *bvs4* genes should abolish the ability of *Rlv*.3841 to grow on media containing nitriles and produce ammonia, these abilities should be restored by complementation of *bvs4* mutant with a functional copy of the *bvs4* gene. Mutation of bvs5 genes should abolish the nitrate reductase/sulphite oxidase activity of *Rlv*. 3841 in a specific assay. This activity should be restored by complementation of *bvs5* mutant with a functional copy of the *bvs5* gene.

5.1.2. Methods for investigating *bvs4* and *bvs5* function

Since the putative enzymatic functions of *bvs4* and *bvs5* have been described in *Rlv* 3841, these functions were tested using the appropriate enzyme and growth assays.

5.1.2.1. Investigating *bvs4* gene function

The *bvs4* gene is annotated as a nitrilase. If the gene encodes a nitrilase, then the strain having the gene such as wild-type *Rlv* 3841 should be able to grow on some of the nitriles associated with plants. This trait should be absent in the *bvs4* mutant but should be present in the strain complemented with a working copy of *bvs4*. In order to ascertain their ability to grow in the presence of nitriles, the wild-type *Rlv*. 3841, the *bvs4* mutant and the complemented strain were grown on solidified medium containing different concentrations of plant-associated nitriles.

The ability to grow in the presence of nitrile does not specifically indicate nitrilase activity. The bacteria growing in the presence of nitriles and producing the enzyme nitrilase would cleave the nitrile into a carboxylic acid and ammonia. In order to confirm that the isolates that grow in the presence of nitriles in the previous half of the experiment were actually cleaving the nitrile, an ammonia assay was set up to detect and quantify the amount of ammonia liberated as a result of enzymatic breakdown of the nitrile.

5.1.2.2. Investigating *bvs5* gene function

The *bvs5* gene annotation suggests that it a molydoenzyme and may either be a nitrate reductase or a sulfite oxidase. Hence the enzyme assays were carried out to investigate enzyme activities for both the enzymes.

As in the tests carried out to verify *bvs4* function, the native *Rlv* 3841 possessing the *bvs5* gene was expected to have the nitrate reductase or sulfite oxidase function. The *bvs5* mutant was expected to lose the enzyme function and the mutant strain complemented with a working copy of the *bvs5* gene was expected to have regained the nitrate reductase / sulfite oxidase activity.

5.2. Materials and methods

bvs4 is annotated as putative aliphatic nitrilase and *bvs5* as putative molybdenum-binding oxidoreductase which could either be a nitrate reductase or sulfite oxidase. Since the enzymatic functions have not been verified experimentally, the following assays were carried out to confirm their functions :

5.2.1. Test for *bvs4* - aliphatic nitrilase

The test for *bvs4* function is divided into two parts. In the first part, the nitrilase growth assay, the cultures (*Rlv* 3841, gene mutants and gene complements) were grown different concentration of three different nitriles to assess the diversity of nitriles that can be used as substrates. In the second part, the ammonia assay, the cultures were tested for their ability to produce ammonia as a result of the breakdown of nitrile into carboxylic acid and ammonia.

5.2.1.1. Nitrile growth assay

The nitrile growth assay was carried out as a quantitative assay to test the ability of the different cultures to grow on a variety of nitriles. The method used here was suggested by Dr. Gail Preston (Howden *et al.*, 2009a, Howden *et al.*, 2009b) and uses Rhizobium Defined Medium, a defined minimal mineral medium for the assay.

5.2.1.1.1. Preparation of reagents

A. *Rhizobium Defined Medium* (Vincent, 1970)

Rhizobium Defined Medium (RDM) is a minimum medium for rhizobia. The recipe of RDM is listed in Appendix A. The list of nitriles and the concentrations used are listed in the following section.

B. *The test nitriles*

The test nitriles were sterilized by filtration and added to sterile, cool, complete RDM medium directly or after dilution in sterile RDM broth. The concentration of nitriles used was as follows : propionitrile (1mM, 2mM, 5mM and 10mM), 3-phenyl propionitrile (1mM, 2mM, 5mM and 10mM) and 3-indole acetonitrile (0.5mM, 1mM and 2mM).

5.2.1.1.2. Procedure

i. The actively growing cultures of mutant and control strains were streaked on sterile RDM medium agar plates containing different concentrations of nitriles.
ii. The plates were incubated at 28^{O}C for a week to observe for evidence of growth.
iii. The nitrile(s) showing growth were used for assessing the ability of the cultures to cleave the nitrile to form carboxylic acid and liberate ammonia.

5.2.1.2. Assay for ammonia production

The hydrolysis of nitriles produces carboxylic acid and ammonia. Ammonia reacts with a-ketoglutaric acid (KGA) and reduced nicotinamide adenine dinucleotide phosphate (NADPH) in the presence of L-glutamate dehydrogenase (GDH) to form L-glutamate and oxidized nicotinamide adenine dinucleotide phosphate (NADP+) as follows:

$$KGA + NH4^{+} + NADPH \xrightarrow{GDH} L\text{-}Glutamate + NADP^{+} + H_2O$$

The decrease in absorbance at 340 nm, due to the oxidation of NADPH, is proportional to the ammonia concentration. L-Glutamate dehydrogenase reacts specifically with ammonia. The Ammonia Assay Kit (Sigma) may be used to determine ammonia concentrations in the range of 0.2–15 mg/ml.

5.2.1.2.1. Preparation of reagents

A. ***The Ammonia Assay Kit :*** (Catalog Number AA0100, Sigma) used in the detection and estimation of ammonia includes :

i. Ammonia Assay Reagent (Catalog Number A0853, Sigma) : The dry reagent contains α-ketoglutaric acid, NADPH, buffers, stabilizers, and nonreactive fillers. The reagent is stored at 2-8^{O}C.
ii. L-Glutamate Dehydrogenase from bovine liver (Catalog Number G2294, Sigma) : The enzyme is supplied as a solution in phosphate buffer, pH 7.4, with 50% glycerol. Store at 2-8^{O}C.
Activity : ≥ 1,200 units/ml.

iii. Ammonia Standard Solution (Catalog Number A0978, Sigma) : The ammonia concentration of the standard solution is ~10 mg/ml (~588 mM as ammonium sulfate).

B. ***The test strains :*** *Rlv*.3841, *Rlv*.3841 Chr. nitrilase , *bvs4* and *bvs5* gene disruption mutants

C. ***Media :*** TY media with appropriate antibiotic, RDM (Rhizobia definition media) with 10mM propionitrile.

The RDM medium contains KNO_3 as a nitrogen source which could potentially serve as a source of nitrogen. Hence, the assay was repeated using RDM medium lacking KNO_3 but supplemented with 10mM propionitrile and RDM medium containing KNO_3 but lacking propionitrile with *Rlv*. 3841 as the positive control, *bvs4* mutant as the test and water as the blank.

5.2.1.2.2. Procedure

A. ***Rhizobial culture suspension :*** The ammonia assay was performed on the wild-type *Rlv* 3841, disruption mutants for the *bvs4, bvs5,* Chr. Nitrilase mutant and the complemented strains. The cultures for the ammonia assay were prepared as follows :

i. Fresh rhizobium cultures were inoculated in 35ml TY in 50ml Corning tubes with appropriate antibiotic and incubated on a rotary shaker incubator at 150rpm at 28°C for 24 hours.

ii. The cultures were then transferred to 35ml RDM media with 10 mM propionitrile and incubated as before for 24 hours.

iii. The cultures were then centrifuged and the supernatant was transferred to clean sterile tubes and used in the assay.

B. ***Ammonia assay***

i. Pipette the solutions into appropriately marked cuvettes as shown in Table 5.1.

Table 5.1. Ammonia assay protocol

Cuvette	Ammonia Assay Reagent (ml)	Ammonia Standard Solution (ml)	Sample (ml)	Water (ml)
Reagent Blank	1.0	-	-	0.1
Test	1.0	-	0.1	-
Standard	1.0	0.05	-	-

ii. Set the spectrophotometer to 340 nm and the absorbance to zero using water as the reference.

iii. Mix the contents in each cuvette and incubate for 5 minutes at 28^{O}C. Measure the absorbance of each solution at 340 nm.

iv. Add 10 µl of L-Glutamate Dehydrogenase solution to each cuvette.

v. Mix the contents of each cuvette and incubate for 5 minutes at 28^{O}C.

vi. Measure the absorbance of each solution at 340 nm.

The response of the assay was checked for linearity between 0-20mM of ammonia using a 20mM solution of ammonia (~1176 mM as ammonium sulfate).

C. ***Calculation :*** Determine Δ_{340} for the Reagent Blank, Test, and Standard. For each: $\Delta_{340} = A_{Initial} - A_{Final}$

Δ (Δ_{340})Test or Standard = Δ_{340} (Test or Standard) – Δ_{340} (Blank) mg of NH_3/ml of original sample

$$= \frac{(A)\,(TV)\,(MW\ of\ Ammonia)\,(F)}{(\varepsilon)\,(d)\,(SV)\,(Conversion\ Factor\ for\ \mu g\ to\ mg)}$$

$$= \frac{(A)\,(TV)\,(17)\,(F)}{(6.22)\,(1)\,(SV)\,(1{,}000)}$$

$$= \frac{(A)\,(TV)\,(F)}{(SV)} \times 0.00273$$

Where,

A = Δ (Δ_{340})Test or Standard,

TV = Total Assay Volume in ml (1.1ml)

SV = Sample Volume in ml (0.1ml),

d = Light path (cm) = 1 cm

MW of Ammonia = 17 g/mole or equivalently 17 mg/mM

F = Dilution Factor from Sample Preparation (=1).

ε = Millimolar Extinction Coefficient for NADPH at 340 nm [mM^{-1} cm^{-1} or equivalently(ml/mM)(1/cm)]

5.2.2. Test for *bvs5* - molybdenum-binding oxidoreductase

The *bvs5* gene is annotated as putative molybdenum-binding oxidoreductase which could either be a nitrate reductase (NR) or sulfite oxidase (SO). Hence, both the enzyme assays were carried out to characterise the gene function.

5.2.2.1. Assay for nitrate reductase

In this assay the nitrate from sodium nitrate is reduced by the enzyme nitrate reductase in the presence of a methyl viologen (1,1'-dimethyl-1,4,4'-bipyridinium dichloride).

Methyl viologen acts as a redox agent, donates an electron for the reduction of nitrate to nitrite and in the process gets oxidised to a colourless product.

$$\text{Methyl Viologen}_{(red)} + \text{Nitrate} \xrightarrow{NR} \text{Methyl Viologen}_{(ox)} + \text{Nitrite} + H_2O$$

Red = reduced, ox = oxidized, NR = nitrate reductase.

The nitrate reductase assay is a colorimetric assay carried out at $30^{O}C$ and at pH = 7. The disappearance of colour is measured at 600 nm using a cuvette with a path length of 1 cm.

5.2.2.1.1. Preparation of reagents

Dissolve all reagents in deionized water which has been boiled for 10 minutes and cooled. After cooling, degas the water to further ensure the removal of oxygen. To dissolve the reagents mix gently to avoid aeration. Never shake any reagent. After preparing the different reagents, store in vials which are tightly capped. Store on ice at 0-5°C. It is best to have a minimum volume of air in these containers.

i. 150 mM Potassium Phosphate Buffer, pH 7.0 at 30°C : Prepare 200 ml in deionized water using 4.08g Potassium Phosphate, Monobasic, Anhydrous. Adjust to pH 7.0 at 30°C with 1 M KOH. (PREPARE FRESH.)
ii. 100 mM Sodium Nitrate Solution (NaNO3) : Prepare 10 ml in deionized water using 0.088g Sodium Nitrate. (PREPARE FRESH.)
iii. 1mM Methyl Viologen Solution (Viologen) : Prepare 10 ml in deionized water using 100mg Methyl Viologen. (PREPARE FRESH.)
iv. 0.8% (w/v) Sodium Hydrosulfite Solution (HS) (Prepare 5 ml in deionized water using 40mg Sodium Hydrosulfite. PREPARE FRESH.)
v. 0.8% (w/v) Sodium Bicarbonate Solution (Bicarb) : Prepare 5 ml in deionized water using 40mg Sodium Bicarbonate.
vi. Sodium Hydrosulfite/Sodium Bicarbonate Solution (Hydro/Bicarb) : Immediately before use, prepare by adding 5 ml of Reagent D and 5 ml of Reagent E to a suitable container. Do not shake.

5.2.2.1.2. Preparation of cell lysate for the enzyme assay :

The nitrate reductase assay was performed on wild-type *Rlv* 3841, gene mutants and complemented strains. *E.coli* DH5α was included as the positive control since it has nitrate reductase activity (Bonnefoy and Demoss, 1994). The cell lysates were prepared as follows :

i. Rhizobium cultures were inoculated in 35 ml TY broth in 50ml Corning tubes with appropriate antibiotics. The *E.coli* DH5α was inoculated in 35 ml LB broth. The cultures were incubated on a rotary shaker incubator at 150rpm at 28°C for 24 hours.

ii. The cultures were then centrifuged, washed and resuspended in 20ml of 100mM potassium phosphate buffer pH 7.6.

iii. The cell suspension (15 ml per conical tube) was sonicated on ice (Bandelin sonoplus, 70 W) in a 50 ml conical tube (1 cycle of 30 sec).

iv. Cell extracts were stored at -80 and used for nitrate reductase and sulfite oxidase assays.

5.2.2.1.3. Procedure

i. Pipette (in millilitres) the reagents from Table 5.2 into suitable containers : (Allow the reagents to flow down the side of the vial. It is essential that all of the reagents be virtually free of oxygen in order to prevent the oxidation of the reduced methyl viologen, the substrate for the reaction. If the substrate is oxidized, no blue colour will appear and no enzymatic reaction will occur).

Table 5.2. Nitrate reductase protocol

Reagent	Test	Blank
Deionized Water	-	0.15
Reagent A (Buffer)	0.90	0.90
Reagent B ($NaNO_3$)	0.15	0.15
Reagent C (Viologen)	0.15	0.15
Reagent J (Enzyme Solution)	0.15	-
Mix by inversion and equilibrate to $30^{O}C$. Then add …		
Reagent F (Hydro/Bicarb)	0.15	0.15

Before the addition of Reagent F (Hydro/Bicarb) gently invert the container. Do not agitate or mix the solution since this may introduce oxygen into the mixture.

ii. Mix by inversion and record the decreasing A_{600} nm for both the Test and Blank using a suitable spectrophotometer.

5.2.2.2. Assay for sulfite oxidase

In this assay the sulfite from sodium sulfite is oxidized to sodium sulfate by sulfite oxidase. During the process, the iron in cytochrome c gets reduced from Fe^{+3} to Fe^{+2} resulting in a decrease in the intensity of colour of the cytochrome C solution which can be monitored at 550nm by an increase in absorbance due to formation of reduced cytochrome c which has an absorption peak at 550 nm.

$$\text{Sulfite} + O_2 + H_2O + \text{Cyt } c_{(Ox)} \xrightarrow{SO} \text{Sulfate} + \text{Cyt } c_{(Red)} + H_2O_2$$

Ox = oxidised, Red = reduced, SO = sulfite oxidase

The sulfite oxidase assay is a colorimetric assay performed at 25^{O}C and pH = 8.5. The decrease in the intensity of colour / increase in absorbance is measured at 550 nm using a cuvette with a path length of 1 cm

5.2.2.2.1. Preparation of reagents

i. 100 mM Tris HCl Buffer, pH 8.5 at 25^{O}C : Prepare 200 ml in deionized water using 3.15g Trizma Base. Adjust to pH 8.5 at 25^{O}C with 1 M HCl.
ii. 33 mM Sodium Sulfite Solution : Prepare 10 ml in Reagent A using 0.042g Sodium Sulfite, Anhydrous.
iii. 2mM Cytochrome c : Prepare 10 ml in Reagent A using 1.25mg Cytochrome c, from Horse Heart.
iv. Cell lysate to test enzyme activity : From nitrate reductase assay. *E.coli* DH5α lysate was used for the positive control.

5.2.2.2.2. Procedure

i. Pipette (in millilitres) the following reagents into suitable cuvettes :

Table 5.3. Sulfite oxidase protocol

Reagent	Test	Blank
Reagent A (Buffer)	2.77	2.77
Reagent B (Sodium Sulfite)	0.03	-
Reagent C (Cytochrome c)	0.10	0.10
Deionized Water	-	0.03
Equilibrate to 25°C. Monitor the A550nm until constant, using a suitably thermostatted spectrophotometer. Then add:		
Reagent D (Enzyme Solution)	0.10	0.10

ii. Immediately mix by inversion and record the increase in A_{550} for approximately 5 minutes.

5.3. Results

5.3.1. Test for *bvs4* - aliphatic nitrilase

5.3.1.1. Nitrile growth assay

The results of the nitrile growth assay shows that native *Rlv* 3841 was capable of growing in presence of propionitrile but does not grow in the presence of 3-phenyl propionitrile or 3-indoleacetonitrile (Table 5.4).

The ability to grow in presence of propionitrile, seen in the form of increase in turbidity of the bacterial suspension, was observed from 1mM to 10mM of phenylpropionitrile. The ability to grow in the presence of phenylpropionitrile was lost when the *bvs4* gene encoding the putative nitrilase is mutated. The ability to grow in presence of propionitrile was regained when the *bvs4* mutant was complemented with a functional copy of the *bvs4* gene. This ability to use nitrilase for detoxification of nitriles and grow in its presence was not restored when the *bvs4* mutant is complemented with a copy of *bvs5* gene indicating that presence of *bvs4* gene was necessary and sufficient to grow in the presence of propionitrile.

Table 5.4. Table showing the growth of cultures on different of nitriles.

	Propio-nitrile 1mM	Propio-nitrile 2mM	Propio-nitrile 5mM	Propio-nitrile 10mM	3-phenyl propio-nitrile	3-indole aceto-nitrile	no nitrile
3841	+	+	+	+	−	−	+
bvs4 mutant	−	−	−	−	−	−	+
bvs5 mutant	+	+	+	+	−	−	+
bvs4 mutant + *bvs4*	+	+	+	+	−	−	+
bvs4 mutant + *bvs5*	−	−	−	−	−	−	+
bvs5 mutant + *bvs4*	+	+	+	+	−	−	+
bvs5 mutant + *bvs5*	+	+	+	+	−	−	+

Key : + = growth, − = no growth.

The loss of ability to grow in presence of phenylpropionitrile when the *bvs4* gene is mutated suggests that *bvs4* gene may be associated with nitrilase activity.

5.3.1.2. Assay for ammonia production

In order to verify the presence of a functional nitrilase, it is necessary to detect either the reduction in nitrile concentration or an increase in the concentration of the reaction products. The ammonia assay was hence used to detect the liberation of the reaction end product, ammonia, from phenylpropionitrile as evidence of nitrilase activity.

The test to determine the linear response range of the assay showed that the response of the assay was linear upto 16mM of ammonia and was in agreement with the value provided by the supplier in the kit.

Table 5.5. The average A_{340} values for reduction in NADPH absorbance for the different test cultures after reaction starting 5 minutes.

Test sample	A_{340}	Ammonia production (ug/ml)
Standard	0.336	10.08
WT3841	0.312	9.36
*bvs4*mut	0.075	2.25
*bvs5*mut	0.305	9.15
*bvs4*mut+*bvs4*	0.341	10.23
*bvs4*mut+*bvs5*	0.062	1.86
*bvs5*mut+*bvs4*	0.350	10.50
*bvs5*mut+*bvs5*	0.321	9.63
ChrNmut	0.276	8.28
Blank	0.000	0.00

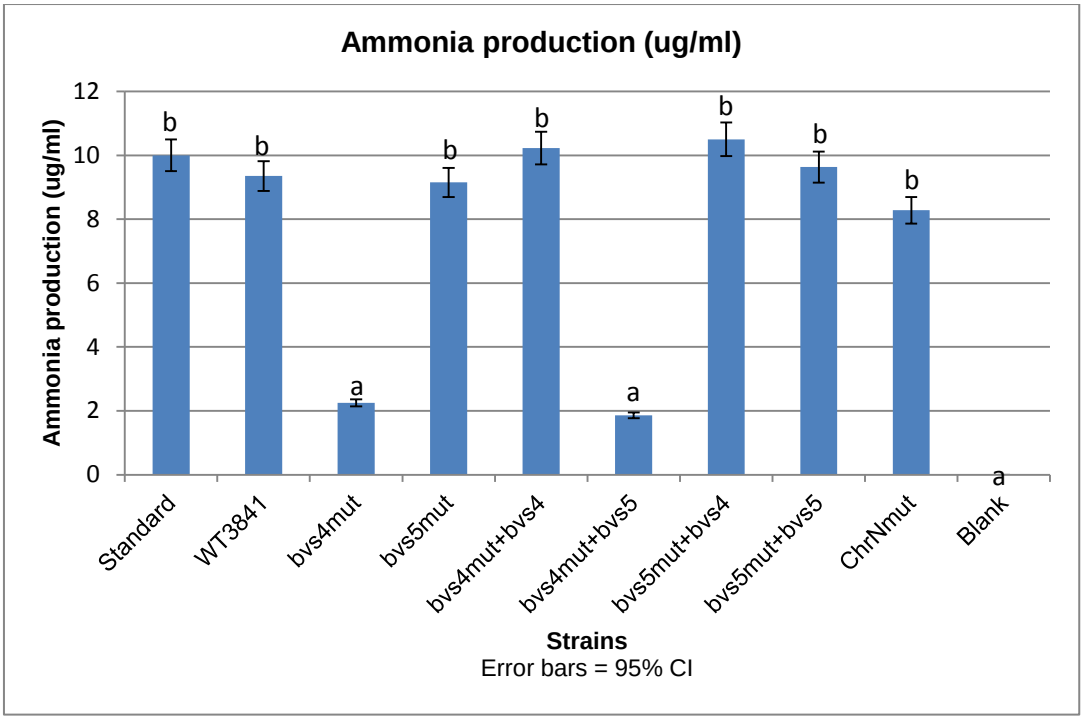


Figure 5.3. Calculated ammonia production from the test strains grown on RDM using phenylpropionitrile as a nitrogen source.

The analysis of data from the ammonia assay by the Kolmogorov–Smirnov test indicated that the data were normally distributed. Hence, the data were analysed using one-way ANOVA (Analysis of Variance), a parametric statistical test to check whether or not the means of groups are equal. The analysis was performed to check for difference in means in the amount of ammonia produced by different cultures as a result of nitrilase activity using the 'IBM SPSS Statistics' software package (Version 21) from IBM Corporation.

Table 5.6. Results of ANOVA to check difference between the amount of ammonia produced by different cultures as a result of nitrilase activity.

ANOVA					
Ammonia production (ug/ml)					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	275.831	8	34.479	72.127	.000

The results of the analysis indicates that there is a significant difference in the means, F (8, N = 27) = 72.127, $p \leq 0.05$.

Since *p*-value ≤ 0.05 = α, we reject the null hypothesis and accept the alternate hypothesis i.e. at α = 0.05 level of significance, there exists enough evidence to conclude that there is significant difference in the mean test scores among the test subjects. In order to analyse the pattern of difference between means, the one-way ANOVA was followed by a post-hoc LSD (least significant difference) test to carry out pairwise comparisons.

The results of the pairwise analysis shows that of the eight cultures tested, the amount of ammonia produced by the two cultures that do not carry a functional copy of *bvs4* differed significantly from those carrying a functional copy of *bvs4*. The loss of *bvs4* severely affected the nitrilase activity of the culture which resulted in decreased ammonia production as quantified by the assay. The ability to cleave propionitrile and form ammonia is regained when the *bvs4* mutant is complemented with a functional copy of the *bvs4* gene. This ability is not restored when the *bvs4* mutant is complemented with a copy of *bvs5* gene (Figure 5.4).

The results of the post-hoc LSD test are shown in the figure below.

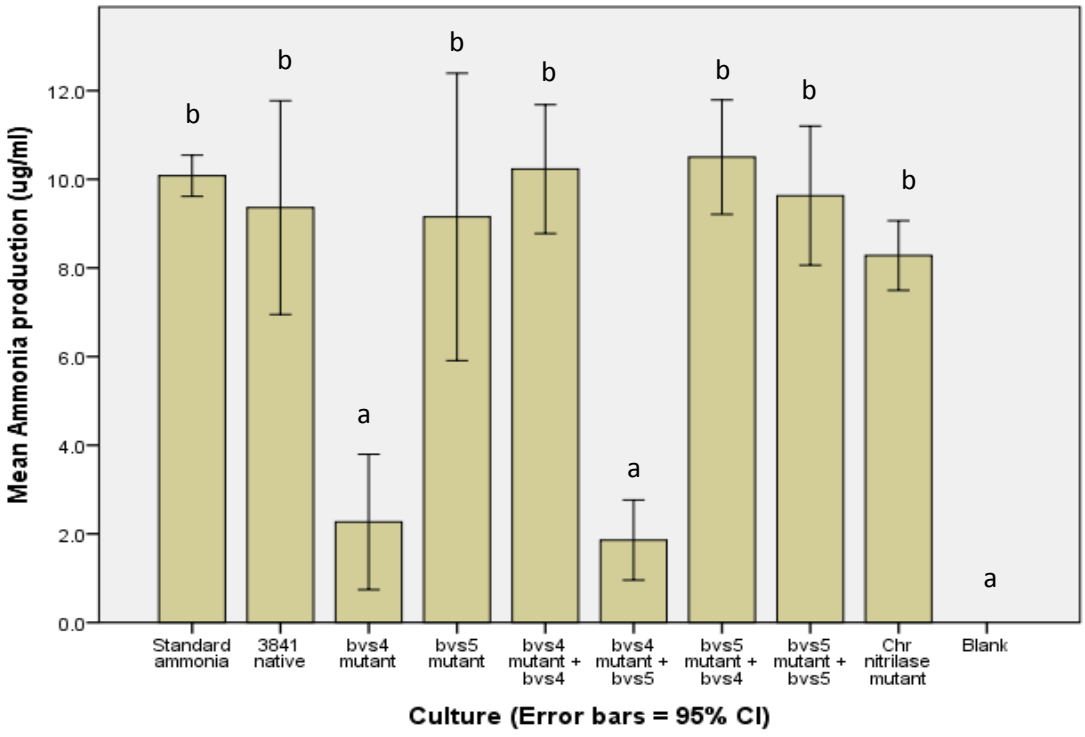


Figure 5.4. Result of the post-hoc LSD test showing pairwise comparison of ammonia estimated in the ammonia assay.

To test the effect of KNO_3 on ammonia production, the assay was repeated with RDM medium lacking KNO_3 but supplemented with 10mM propionitrile and RDM medium containing KNO_3 but lacking propionitrile, using *Rlv*. 3841 as the positive control, *bvs4* mutant as the test and water as the blank. Using the Kolmogorov–Smirnov test, the assay results were found to be normally distributed, so the data were analysed using one-way ANOVA.

Table 5.7. Results of ANOVA to check the effect of KNO_3 on ammonia produced by different cultures as a result of nitrilase activity.

ANOVA					
Ammonia production (ug/ml)					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	289.119	11	26.284	317.147	.000

The results of the analysis indicates that there is a significant difference in the means, F (11, N = 36) = 317.147, $p \leq 0.05$.

Since p-value ≤ 0.05 = α, we reject the null hypothesis and accept the alternate hypothesis i.e. at $\alpha = 0.05$ level of significance, there exists enough evidence to conclude that there is significant difference in the mean test scores among the test subjects. In order to analyse the pattern of difference between means, the one-way ANOVA was followed by a post-hoc LSD (least significant difference) test to carry out pairwise comparisons.

The results from the test show that production of ammonia by *Rlv*. 3841 in the presence of nitrile is significantly different from the synthesis of ammonia by the mutant. Also, the synthesis of ammonia by *Rlv*. 3841 in the presence of KNO_3 is higher than in its absence and is statistically significant. This difference can be attributed to two different phenomena. It is possible that the presence of a ready source of nitrate like KNO_3 stimulated the growth of rhizobia, resulting in higher enzyme activity and hence a greater utilization of the nitrile resulting in more ammonia being produced. There is also a possibility that the nitrate-nitrogen of KNO_3 may have been channelled into the formation of ammonia-nitrogen as a result of bacterial metabolism resulting in an increase in the formation of ammonia.

The results of the post-hoc LSD test are shown in the figure below.

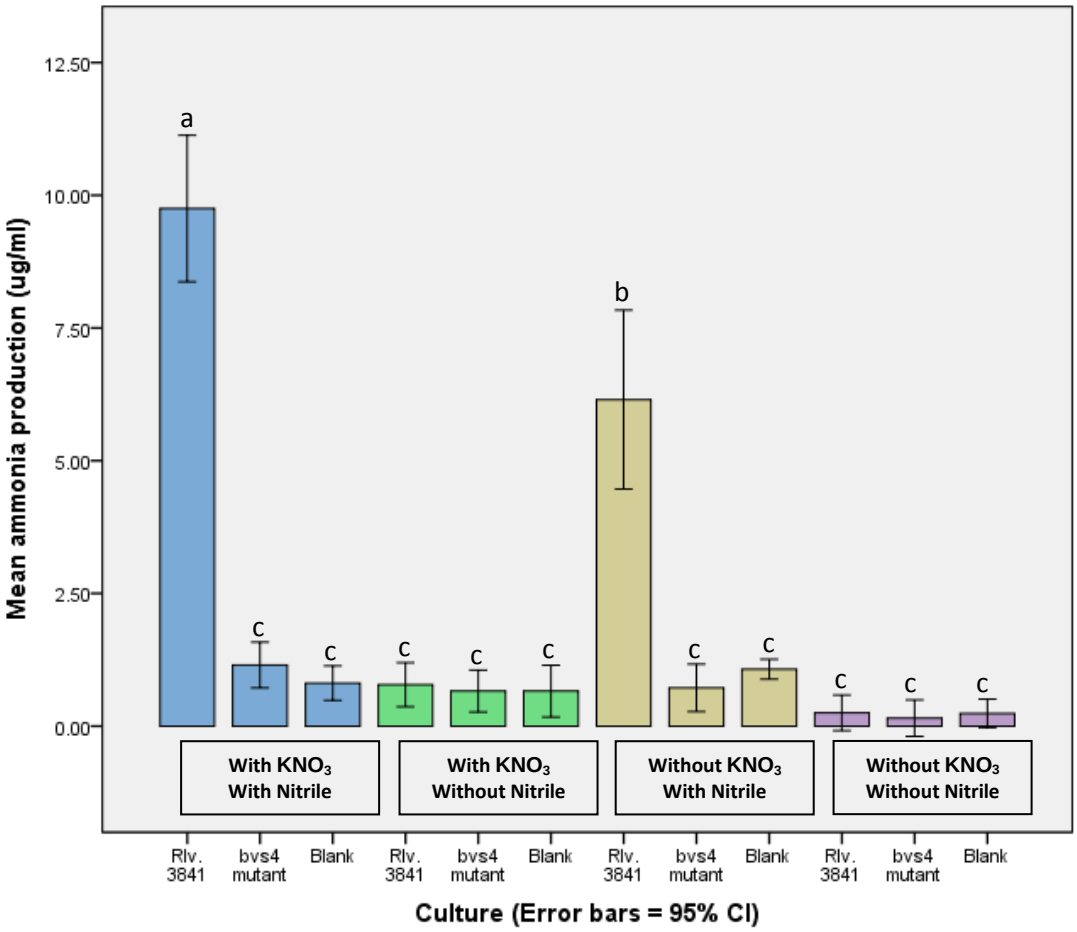


Figure 5.5. Result of the post-hoc LSD test showing pairwise comparisons to check the effect of presence of KNO_3 in the ammonia assay. Histogram bar values with the same letter are not significantly different at p = 0.05 using LSD.

5.3.2. Test for *bvs5* - putative Mo-binding oxidoreductase

5.3.2.1. Nitrate reductase assay

The results of the nitrate reductase test are tabulated in Table 5.7. The rate of reduction of methyl viologen was highest for *E.coli* DH5α, which possesses an active nitrate reductase. Other strains did not demonstrate any significant enzyme activity. The small reduction in the colour of methyl viologen in other cultures might be due to other metabolic process not related to nitrate reductase activity. In *E.coli* DH5α, the reduction of the dye progresses for about 16 minutes and then backtracks probably due to the reoxidation of the dye by atmospheric oxygen. This reason might also explain the plateauing of the rate of decrease in readings in some of the strains when the rate of reduction of methyl viologen would equal the rate of its reoxidation by atmospheric oxygen.

Table 5.8. The average Δ_{A600} values for reduction of methyl viologen for the test cultures at different time intervals in minutes ($A_{initial} - A_{final}$).

Culture↓ / Time (min) →	0	4	8	12	16	20
3841	0	0.001	0.002	0.003	0.003	0.002
bvs4 mutant	0	0	0.002	0.003	0.003	0.004
bvs5 mutant	0	0.001	0.001	0.002	0.003	0.003
bvs4 mutant + *bvs4*	0	0	0.001	0.003	0.003	0.003
bvs4 mutant + *bvs5*	0	0.001	0.002	0.004	0.003	0.004
bvs5 mutant + *bvs4*	0	0.001	0.001	0.003	0.003	0.003
bvs5 mutant + *bvs5*	0	0	0.001	0.002	0.002	0.003
Positive control (*E.coli* DH5α)	0	0.008	0.01	0.01	0.009	0.008
Negative control	0	0	0	0	0	0

Graphical representation of data : nitrate reductase assay :

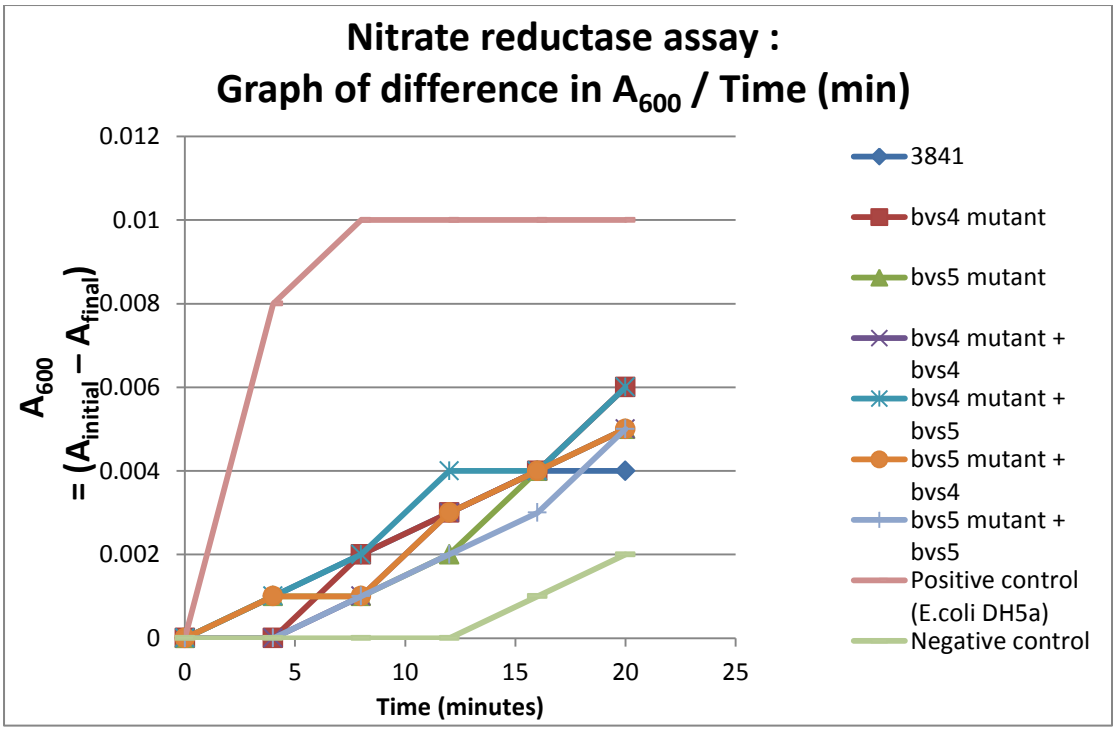


Figure 5.6. Nitrate reductase assay : Plot of A_{600} values for reduction of methyl viologen for test cultures at different time intervals in minutes.

The analysis of data from the nitrate reductase test by the Kolmogorov–Smirnov test indicated that the data were normally distributed. Hence, the data were analysed using ANOVA (Analysis of Variance).

Table 5.9. Results of ANOVA to check differences in the nitrate reductase activities of difference cultures.

ANOVA					
Nitrate reductase activity					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.000	7	.000	10.192	.000

The results of the analysis indicates that there is a significant difference in the means, F (7, N = 24) = 10.192, $p \leq 0.05$.

Since p-value ≤ 0.05 = α, we reject the null hypothesis and accept the alternate hypothesis i.e. at α = 0.05 level of significance, there exists enough evidence to conclude that there is significant difference in the mean test scores among the test subjects.

To analyse the pattern of difference between means, the one-way ANOVA was followed by a post-hoc LSD (least significant difference) test to carry out specific pairwise comparisons.

The results of the pairwise analysis shows that of the eight cultures tested, the positive control, i.e. *E.coli* DH5α, with known nitrate reductase activity, differs significantly from the six test cultures. The means of the six test cultures did not differ significantly from each other (except 3841 and *bvs4* mutant) but differed significantly from the positive control. The results indicate that except for the positive control i.e. *E.coli* DH5α, none of the six test cultures possess nitrate reductase activity.

As mentioned before the small amount of reduction in the colour of methyl viologen in the other cultures could be a result of other metabolic process not related to nitrate reductase activity and the plateauing of the rate of decrease in readings in these strains occurs when the rate of reduction of methyl viologen would equal the rate of its reoxidation by atmospheric oxygen.

The results of the post-hoc LSD test are shown in the figure below.

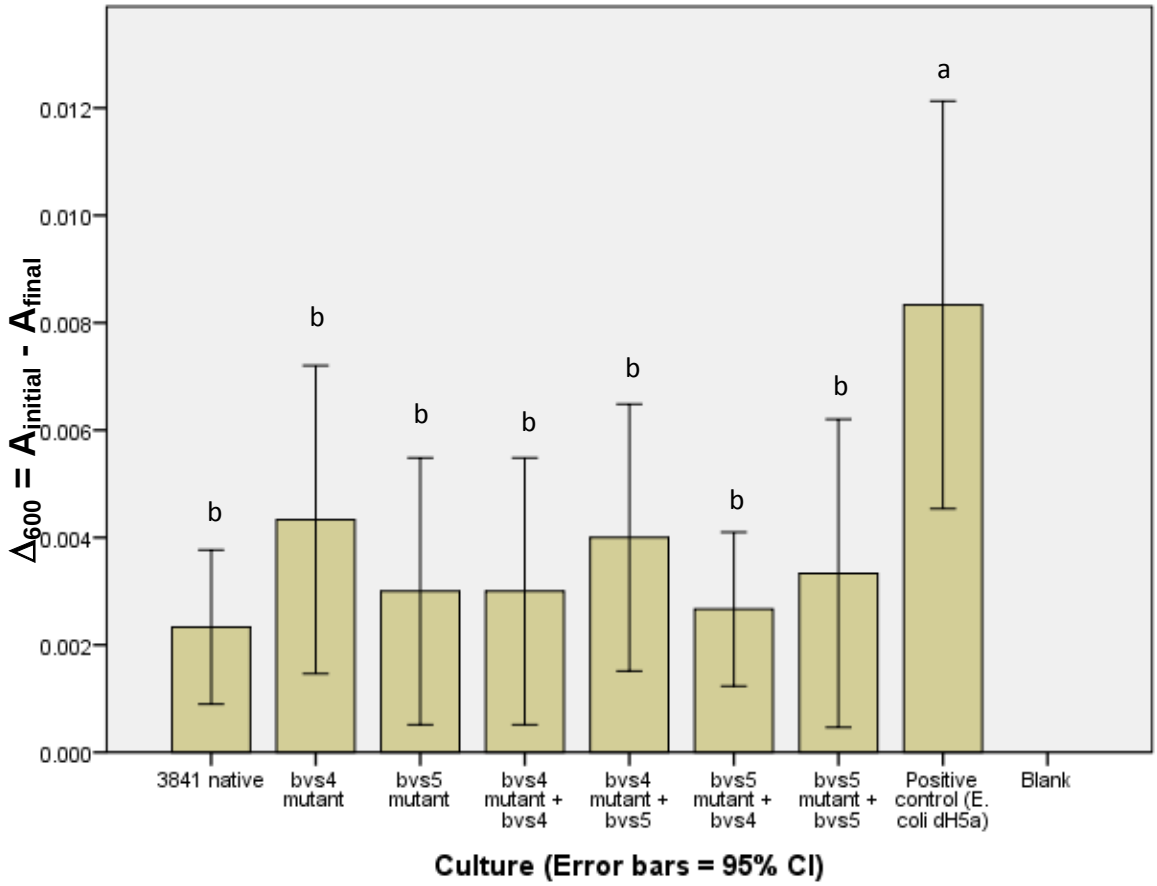


Figure 5.7. Result of the post-hoc LSD test showing pairwise comparison of nitrate reductase activity.

5.3.2.2. Sulfite oxidase assay

The results of the sulfite oxidase assay are tabulated in Table 5.9. It was observed that conversion of the oxidised cytochrome c to its reduced form occurred in all strains carrying a functional *bvs5* gene, including the wild-type *Rlv* 3841 and the positive control *E.coli* DH5α. The negative control and the strains lacking a copy of *bvs4* or carrying a disrupted copy of the gene showed absence of sulfite oxidase activity. The sulfite oxidase activity was regained when the *bvs5* mutant was complemented with a functional copy of the *bvs5* gene. The sulfite oxidase activity was not restored when the *bvs5* mutant is complemented with a copy of *bvs4* gene indicating that presence of *bvs5* gene was necessary and sufficient to exhibit sulfite oxidase activity. Also, the strains carrying the complementary copy of the *bvs5* gene showed enhanced activity of sulfite oxidase, as seen in Figure 5.7.

Table 5.10. The average A_{550} values for reduction of cytochrome c for the test cultures at different time intervals in minutes.

Culture↓ / Time (min) →	0	1	2	3	4	5	6
3841	0	0	0.011	0.144	0.221	0.338	0.453
bvs4 mutant	0	-0.006	-0.005	-0.004	-0.003	-0.003	-0.002
bvs5 mutant	0	0.002	0.002	0.003	0.003	0.003	0.004
bvs4 mutant + *bvs4*	0	-0.002	-0.002	-0.001	-0.001	-0.001	0
bvs4 mutant + *bvs5*	0	0.005	0.073	0.16	0.437	0.688	0.81
bvs5 mutant + *bvs4*	0	0.002	0.002	0.003	0.003	0.003	0.004
bvs5 mutant + *bvs5*	0	0.005	0.006	0.02	0.045	0.786	0.954
Positive control (E.coli DH5α)	0	0.001	0.011	0.138	0.286	0.42	0.564
Negative control	0	0	0	0	0	0	0

Graphical representation of data : Sulfite oxidase assay :

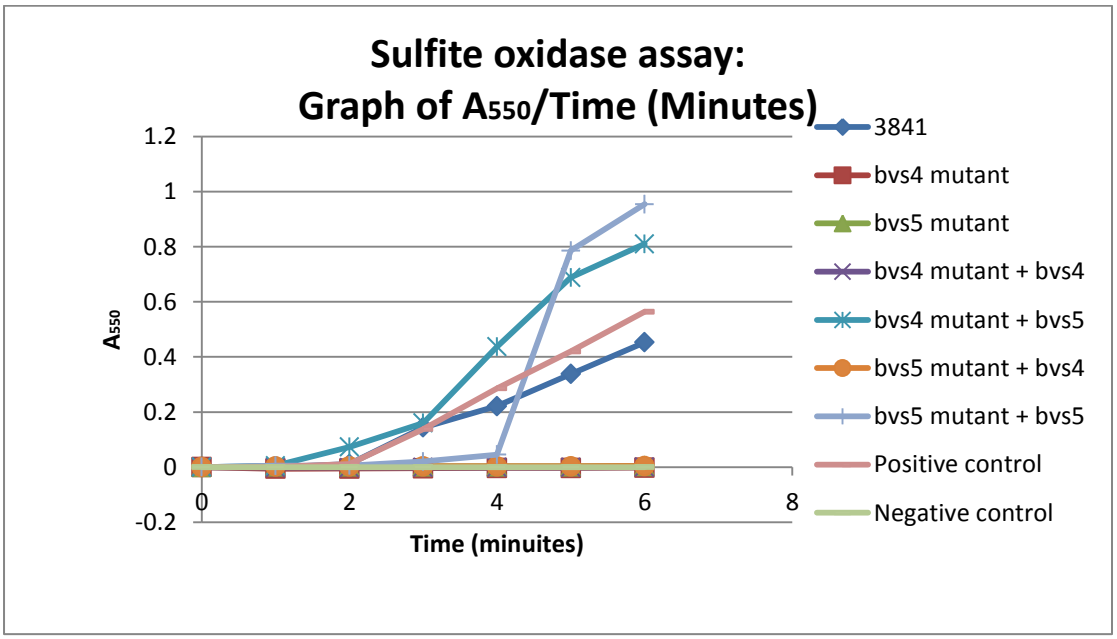


Figure 5.8. Sulfite oxidase assay : Plot of average A_{550} values for reduction of cytochrome c for test cultures at different time intervals in minutes.

Amongst the strains showing sulfite oxidase activity, the strains carrying *bvs5* under the control of a native promoter showed an early onset of sulfite oxidase activity which increases gradually with time suggesting non-cooperative enzyme kinetics. In contrast, the strains carrying the *bvs5* gene under the control of a strong promoter exhibits sigmoidal kinetics typical of enzymes showing co-operative binding. This activity is less pronounced in the *bvs4* mutant carrying *bvs5* complement since it has its native functional *bvs5* gene which can function in a non-cooperative fashion; but can be very clearly observed in the case of *bvs5* mutant carrying a *bvs5* complementary gene in which the only sulfite reductase activity observed comes from the complementary *bvs5* gene and not from the disrupted inactive native *bvs5* gene.

The analysis of data from the sulfite oxidase test by the Kolmogorov–Smirnov test indicated that the data were normally distributed. Hence, the data were analysed using ANOVA (Analysis of Variance).

Table 5.11. Results of ANOVA to check differences in sulfite oxidase activities of difference cultures.

ANOVA					
Sulfite oxidase activity					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.355	7	.479	105.904	.000

The results of the analysis indicates that there is a significant difference in the means, F (7, N = 24) = 105.904, $p \leq 0.05$.

Since p-value ≤ 0.05 = α, we reject the null hypothesis and accept the alternate hypothesis i.e. at α = 0.05 level of significance, there exists enough evidence to conclude that there is significant difference in the mean test scores among the test subjects.

To analyse the pattern of difference between means, the one-way ANOVA was followed by a post-hoc LSD (least significant difference) test to carry out specific pairwise comparisons. The results of the pairwise analysis shows that the strains carrying a functional *bvs5* gene, including the wild-

type *Rlv* 3841 and the positive control *E.coli* DH5α exhibit sulfite oxidase activity. The negative control and the strains lacking a copy of *bvs4* or carrying a disrupted copy of the gene lacked sulfite oxidase activity.

The results of the post-hoc LSD test are shown in the figure below.

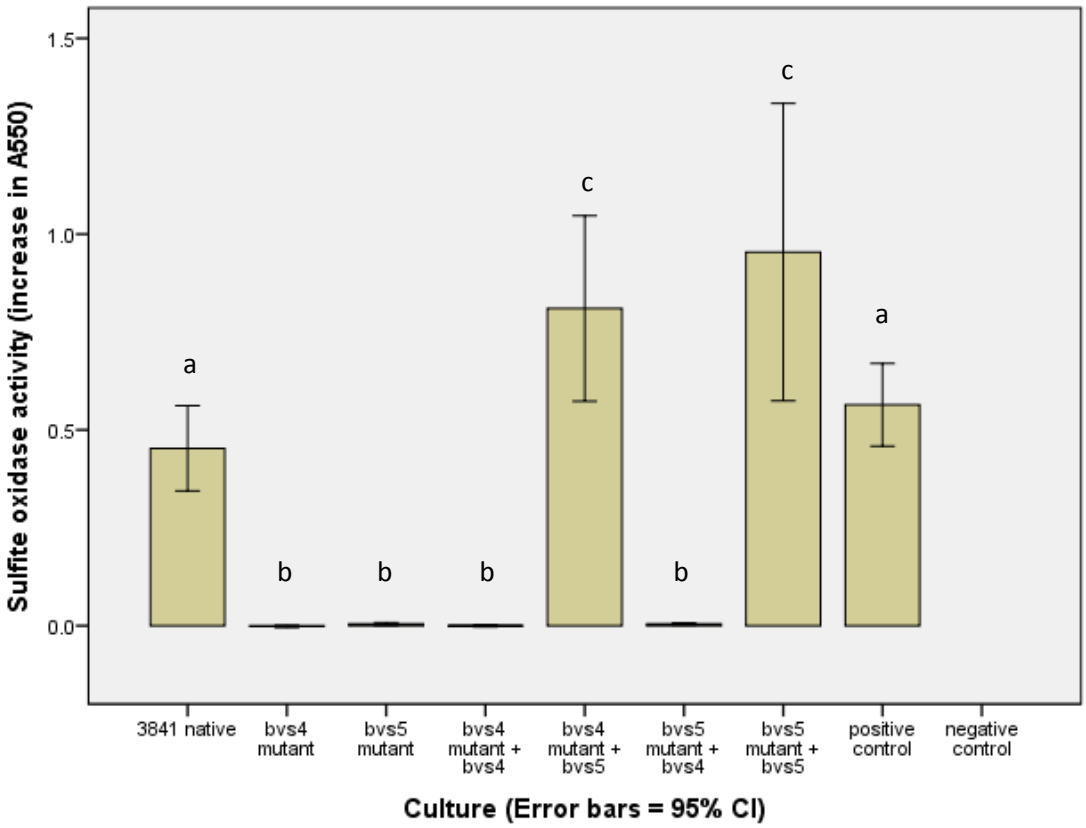


Figure 5.9. Result of the post-hoc LSD test showing pairwise comparison of sulfite oxidase activity.

5.4. Discussion

The results of the enzyme assays give us some insight into the possible functions of the enzymes / proteins encoded by these genes.

The pRL80076 (*bvs4*) gene that is annotated as an aliphatic nitrilase showed evidence of nitrilase activity. Of the three nitriles tested for detoxification viz. 3-Indole acetonitrile, Propionitrile, and Phenylpropionitrile, tested at concentrations of 0.5, 1, 2, 5 and 10mM in a defined mineral medium, the native *Rlv*. 3841 strain was able to grow only on propionitrile at all the concentrations. This indicates that *Rlv*. 3841 is probably able to use propionitrile as a nitrogen source, but not the other two nitriles. The mutation of *bvs4* gene abolished the ability of the

bacterium to grow on propionitrile. Mutation in *bvs5* gene had no effect on the ability of the organism to grow in the presence on propionitrile. Presence of KNO_3 in the RDM medium was found to stimulate ammonia synthesis.

Complementing the *bvs4* mutant with a copy of the gene cloned into a plasmid under the control of a strong promoter restored the ability of the mutant to grow on propionitrile. Complementing the *bvs4* mutant with a complement of the *bvs5* gene had no effect on the mutation. This suggests that if *bvs4* and *bvs5* form an operon then the *bvs4* gene is upstream to the *bvs5* gene, mutation of *bvs4* will abolish *bvs5* transcription but not vice-versa. Such an arrangement would explain the results obtained from mutation and complementation studies.

Native *Rlv* 3841, mutant strains and complemented strains were grown in RDM with 10 mM propionitrile. The supernatants of the cultures were tested for the presence of ammonia which would form as a result of nitrilase action. The results of the ammonia assay were in line with the growth assay. The native *Rlv* 3841 and the strains bearing a copy of the *bvs4* gene introduced using plasmid showed presence of ammonia indicating nitrilase activity. These strains did not show a significant difference between the levels of ammonia produced by nitrilase action. The strains containing disrupted *bvs4* gene or lacking a functional *bvs4* (complemented with *bvs5*) did not show significant production of ammonia. This indicates that the *bvs4* most probably functions as a nitrilase.

The pRL80077 (*bvs5*) gene that is annotated as a molybdenum-binding oxidoreductase contains domains that suggest that the gene could code for either a nitrate reductase or a sulfite oxidase. Hence, the strains were subjected to both the enzyme assays along with *E.coli* DH5α strain as the positive control since it is known to exhibit both the enzyme activities. The results of the nitrate reductase assay show that there is a rapid oxidation of methyl viologen in the setup for *E.coli* DH5α but not in *Rlv*. 3841 or any of the mutant or complemented strains. This suggests that there is no evidence to support the hypothesis that the *bvs5* gene encodes nitrate reductase function.

In contrast, the sulfite oxidase assay showed activity with both the native *Rlv*. 3841 and with *E.coli* DH5α. This activity was abolished in the *bvs5* mutant. The activity was restored when the *bvs5* mutant was complemented with an intact copy of the gene cloned in a plasmid. The complemented strain contained the gene being expressed under the control of a strong promoter. This probably

explains the reason why the complemented strain carrying the copy of *bvs5* gene shows a higher activity as compared to the native *Rlv*. 3841 strain. The highest activity is seen in the *bvs4* mutant complemented with a functional copy of *bvs5*. This strain, therefore, has two copies of the *bvs5* gene. The cloning of *bvs5* gene in plasmid pDG71 under the control of a strong promoter could explain the very high rates of sulfite oxidase activity in the complemented strain.

It was also observed that mutation in *bvs4* also affected *bvs5* (sulfite oxidase) activity which was restored on complementing the mutant with a functional copy of *bvs5*. However, mutations in *bvs5* did not affect the *bvs4* (nitrilase) activity of mutants. This suggests that *bvs4* and *bvs5* may constitute an operon in which *bvs4* is upstream to *bvs5* explaining the polar effect of mutation where *bvs4* mutation affects *bvs5* function and not vice-versa.

From the results and discussion, there is evidence to suggest that the *bvs4* and *bvs5* gene encode the enzymes nitrilase and sulfite oxidase, respectively.

Plants synthesize nitrile compounds as a defence strategy to protect the plant against herbivore and pathogen attack. Since hydrolysis of these compounds results in the production of cyanohydrin, which subsequently decomposes to liberate hydrogen cyanide, this defence strategy is also linked in many cases, with the metabolism of cyanide. The nitrilase could also play a role in the detoxification of certain nitriles to form compounds that may act as plant hormones. The nitriles liberated by plants could also serve as a potential source of nitrogen for bacteria in the rhizosphere due to the presence of nitrilase.

Reduced sulfur compounds are found in soil in the form of inorganic minerals like pyrite or in organic forms such as the sulfur-containing amino acids, cysteine and methionine. Sulfite can cause damage to plant roots, especially in acid soils (Chen *et al.*, 2009). The sulfite anion is highly reactive and can reacts with vital cell components such as DNA (by converting cytosine into uracil), proteins (by reducing disulphide bonds) and lipids (by lipoperoxidation). Hence, the bacteria (and plants) that are exposed to sulfite that is generated internally, as a result of metabolism, or externally, need to be able to detoxify it efficiently. Using sulfite oxidase encoded by *bvs5,* this detoxification can occur via the reduction of the sulfite to elemental sulfur or, more commonly by oxidation to sulphate, which could be used by both plants and rhizobia in biosynthesis of protein, pigment or flavonoids.

CHAPTER 6 : DISCUSSION.

6.1. Introduction

The work described in this thesis was carried out to study the metabolism and host-specificity in the *Rhizobium leguminosarum* species complex. In order to investigate these two criteria a number of strains from within the species complex and from related species were investigated. During the investigation, five genes were found to be strongly correlated with the biovar *viciae* strains, irrespective of site from where the strain originated. Hence, the focus of the work was shifted to these five genes. These five genes that were strongly correlated with the biovar *viciae* strains were called as the 'biovar *viciae* specific' genes, abbreviated as '*bvs*' genes.

The thesis sought to investigate the five *bvs* genes associated with the biovar *viciae* strains of *Rhizobium leguminosarum*. The investigations include the distribution of these genes in isolates from different places, the role of these genes in nodulation and an attempt to deduce the functions of these genes based on laboratory work. The five *bvs* genes are located in a contiguous manner on the smallest plasmid in the reference strain *R. leguminosarum* biovar *viciae* pRL8 with NCBI gene accession numbers pRL80073-pRL80077 (Young *et al.*, 2006).

The association between the *bvs* genes and the biovar *viciae* was derived from the genomic data of 72 *R. leguminosarum* isolates from a field adjacent to Wentworth College in the University of York campus. Of the 72 strains investigated, 36 strains were isolated from *Trifolium repens* plants and belonged to biovar *trifolii* (called the TRX strains) while the remaining 36 strains were isolated from *Vicia sativa* plants and belonged to biovar *viciae* (called the VSX strains). Comparative genomics of the genomic data showed that the five *bvs* genes were found in all the 36 VSX strains but none of the 36 TRX strains; and hence the term '*bvs*' genes.

The *bvs* genes found on pRL8 in *Rlv.* 3841 have been ascribed putative functions based on sequence similarities and homology studies. These annotations can be found on the NCBI server, but there is no experimental evidence confirming these functions. However, being fully sequenced and possessing the *bvs* genes, *Rlv.* 3841 plays an important role as a reference strain for this thesis.

In the 2nd chapter of the thesis, 15 strains from different *Rhizobium* species and biovars were studied for their ability to form nodules on two plants from the genus *Vicia* viz. *Vicia sativa* and *Vicia cracca* as well as plants from the closely related genera *Pisum* (*Pisum sativum*) and *Lathyrus* (*Lathyrus pratensis*). The results of our cross-nodulation tests showed that 12 of the 15 strains belonging to the biovar *viciae* were able to form pink, healthy and effective nodules on all the four test plants.

These results of the nodulation tests correspond to earlier studies investigating the diversity of rhizobium biovar *viciae* strains (Mutch and Young, 2004). The 3 strains that did not belong to biovar *viciae* were not able to form nodules on any of the four host plants and these observations correspond to their description in literature (Martínez-Romero *et al.*, 1991, Ramirez-Bahena *et al.*, 2008, Segovia *et al.*, 1993). This observation points to the host-specificity or the inability of isolates belonging to other species to form nodules on these host plants. However, since the results were in line with the reported observations, the tests also functioned as a means of validating the strains for further use in the study.

In addition to the cross nodulation test, a phenotypic assessment was carried out to investigate the metabolic diversity within the test strains using the Biolog Gram Negative 2 or GN2 MicroPlate system. Recent studies have also used Biolog as an important tool to characterize new bacteria (Kaeppel *et al.*, 2012, Tsubouchi *et al.*, 2013) or to differentiate isolates based on specific criteria as done by Rahi *et al.* (2012). In addition, PCA and UPGMA are commonly used to analysis Biolog substrate utilization data to observe the strain clusters and patterns of substrate utilization (Medina-Pascual *et al.*, 2012, Wong and Lee, 2014). These methods have been used in my studies to differentiate *Rhizobium* isolates.

The pattern resulting from the utilization / non-utilization of substrate is referred to as a metabolic fingerprint and used to identify bacteria. The metabolic fingerprint of each isolate used in the Biolog assay was different and distinct indicating that the strains used in the study are metabolically diverse; although a huge fraction of the test strains were *R. leguminosarum* biovar *viciae*. The PCA analysis and UPGMA clustering of the data also showed that it was difficult to cluster the strains into phenotypically similar groups. Some strains showed inability to utilise the common sugars since they lacked some of the key enzymes of sugar metabolic pathways.

In the 3rd chapter of the thesis, the distribution of *bvs* genes in different species and strains of *Rhizobium* was studied. The study was performed using computer and laboratory-based analysis. The computer studies used the BLASTn and BLASTp NCBI web-tool with NCBI database or a standalone application with a local database. The laboratory studies consisted of using PCR amplification of *bvs* genes using specific primers. The PCR amplicons were sequenced and the sequences were used to infer relationships from phylogenetic trees.

The analysis of data from the BLAST and PCR experiments suggest that the *bvs* genes are widely distributed in the biovar *viciae* strains of rhizobium including *R. leguminosarum* USDA2370^T, *R. pisi* DSM30132^T and all field isolates from *Vicia sativa, Vicia cracca, Vicia faba* and *Pisum sativum*. Type strains belonging to other biovars did not possess any of the five genes. The two exceptions to the rule are *R. fabae* CCBAU33202^T and *R. leguminosarum* bv*. trifolii* WSM2304. *R. fabae* CCBAU33202^T, is a member of biovar *viciae*. It can form a symbiosis with plants belonging to the genus *Vicia* and form effective nodules on their roots. Although it belongs to biovar *viciae* the strain lacks all of the five *bvs* genes in its genome. In contrast, the strain *R. leguminosarum* bv *trifolii* WSM2304 belongs to the biovar *trifolii* yet it shows the presence of all the five bvs genes in its genome sequence. The strain has been reported to have more competitiveness for nodulation of *Trifolium* compared to other biovar *trifolii* strains (Reeve *et al.*, 2010b, Reeve *et al.*, 2010c).

The sequence data obtained from the sequencing of PCR amplicons of the *bvs* genes was used to construct phylogenetic network trees. The relationships inferred from these phylogenetic network trees of *bvs* genes was compared to the relationship between the strains obtained using a phylogenetic network tree constructed using sequence of *celC* gene from the test strains. The *celC* gene was used as it has been recommended by Robledo *et al.* (2011) as a reliable phylogenetic marker for taxonomic studies in rhizobia. Faghire *et al.* (2012) mentioned that *celC* genes were found to exist in different positions : in *Rhizobium* they are present on the chromosome but in *Sinorhizobium* they are found on plasmids. The clustering of the strains in the network trees of *bvs* genes was observed to be different from that of the *celC* network tree. The high degree of correlation between the presence of *bvs* genes and their ability to infect plants belonging to the genus *Vicia* indicate that the five *bvs* genes may directly or indirectly play a role in ability of rhizobia to nodulate *Vicia* plants.

In the 4th chapter of the thesis the effect of *bvs* genes on nodulation was studied. This was done by mutating each of the five *bvs* genes and the chromosomal nitrilase gene using the technique of gene mutation by disruption. The disruption of the gene was achieved by using a plasmid spliced with a significant portion of the target gene facilitating integration into the target gene by homologous recombination. The mutations in the genes were confirmed using PCR. The mutants were then used to perform nodulation tests and nodulation competition tests using native wild-type *Rlv*. 3841 as competitors. The results obtained from the dry weights of the plants in the normal nodulation tests showed that mutating the *bvs* genes had no detrimental effect on the ability of the mutants to nodulate the host plants. The results of the nodulation competition assay, however, told a different story. Mutating the chromosomal nitrilase, *bvs1*, *bvs2* and *bvs3* genes did not affect the ability of the bacteria to compete with the wild-type *Rlv*. 3841 to form nodules on the host plants. The occupancy ratio of wild-type to mutants in the nodules at the end of the experiment did not differ significantly with respect to their ratios at the beginning of the experiment. However, mutating the *bvs4* and *bvs5* genes severely affected the ability of the mutants to compete and form nodules on the host plants which was reflected in a significantly higher wild-type to mutant nodule occupancy ratio at the end of the experiment as compared to the ratio at the beginning of the experiment.

In order to confirm the role of *bvs4* and *bvs5* genes in nodulation competition, the strains carrying mutations in these two genes were complemented with a functional copy of the genes. The mutant received a copy of either *bvs4* or *bvs5* gene irrespective of the mutant gene it carried. Thus four complemented strains were obtained : the *bvs4* mutant complemented with *bvs4* or *bvs5* and the *bvs5* mutant complemented with *bvs4* or *bvs5*. A reversal to wild-type phenotype in terms of nodulation competition would effectively prove the role of the two genes in nodulation competition. The nodulation competition assay was carried out as before using the complemented strains instead of the mutants. The results showed that complementing the mutant strains with a functional copy of *bvs5* gene significantly restored the nodulation competition ability of the mutant strains irrespective of gene mutated (*bvs4* or *bvs5*). Thus, while mutation of the two genes significantly hampered the ability of the strain to compete with the wild-type strains for nodulation, there is probably a greater role played by *bvs5* gene in conferring the nodulation ability in the biovar *viciae* strains. Because of this distinctive feature of the mutation, the organisation of these genes on pRL8 was

studied using bioinformatic tools for operon prediction. The analysis indicated that the *bvs4* and *bvs5* may form an operon and within the operon, the *bvs5* gene is positioned transcriptionally downstream of the *bvs4* gene. This may account for the polar nature of the mutations as suggested by the results of the complementation tests. In order to study gene effects, gene mutation and complementation were widely used in recent studies. Ding *et al.* (2012) carried out a similar competition test. They mutated and complemented genes in *R. leguminosarum* VF39 in order to study the effect of glycerol utilization genes on rhizobial competitiveness.

In the 5th chapter of the thesis, the enzymatic functions of *bvs4* and *bvs5* genes were investigated using enzyme assays. Enzyme assays are the key to study genes involved in rhizobial growth or nodulation and have been used before to investigate the role of enzyme function in the symbiosis between rhizobia and the host plant. For example, Esfahani and Mostajeran (2011) tested enzyme activities of four antioxidant enzymes and demonstrated that two of the four enzymes - peroxidase (POX) and Ascorbate peroxidase (APX) played a role in symbiosis adaptation under drought stress condition. A similar study on other antioxidants was conducted by Bianucci *et al.* (2012). Their aim was to identify enzymes involved in *Bradyrhizobium* tolerance of cadmium, and their findings suggested that glutathione content and related enzymes were contributing to tolerance of cadmium and cellular redox balance.

The gene annotations on the NCBI database were used as the starting point for the investigations. The *bvs4* gene is annotated as an aliphatic nitrilase while the *bvs5* gene is annotated as a molybdenum-binding oxidoreductase. Since the *bvs4* gene was annotated as an aliphatic nitrilase, the first study was designed to test the ability of the wild-type *Rlv*. 3841, *bvs4* / *bvs5* gene mutant and complemented strains to grow on defined medium (RDM) containing different concentration of three different nitriles. Of the three nitriles, two were toxic and did not support the growth of any test culture. Only one aliphatic nitrile, propionitrile, allowed the growth of wild-type *Rlv*. 3841, and was used for further work. The results showed that the strains lacking a copy of a functional *bvs4* gene (*bvs4* mutant strain and *bvs4* mutant complemented with a copy of *bvs5* gene) were unable to grow in the presence of propionitrile. The *bvs4* mutant complemented with the *bvs4* gene and all the other strains were able to grow in the presence of propionitrile.

The ability of the strains to grow in the presence of nitrile was not sufficient to indicate that the nitrile was indeed cleaved and detoxified. In order to verify the breakdown of the nitrile into ammonia and carboxylic acid, an ammonia assay was performed to detect the presence of ammonia formed as a result of the cleavage of the nitrile. The bacteria were grown in liquid RDM media in presence of the nitrile and after incubation, the medium was centrifuged and the clear supernatant used to detect the presence of ammonia. Once again, the results showed that the strains lacking a copy of a functional *bvs4* gene (*bvs4* mutant strain and *bvs4* mutant complemented with a copy of *bvs5* gene) did not show a significant amount of ammonia production indicating that they were unable to metabolise and cleave the propionitrile. The *bvs4* mutant complemented with the *bvs4* gene and all the other strains demonstrated significant a significant level of ammonia production suggesting that they were able to grow in the presence of propionitrile and break it down to form carboxylic acid and ammonia. The inability of the *bvs4* mutant to neither grow in the presence of nitrile nor cleave it to form ammonia and the reversion of mutation on complementation presents compelling evidence that *bvs4* is indeed an aliphatic nitrilase.

The *bvs5* gene was annotated as a molybdenum-binding oxidoreductase containing domains which suggested that the gene could code for either a nitrate reductase or a sulphite oxidase. The BLASTp results from Chapter 3 were strongly in favour of the gene coding for a sulphite oxidase. However, it was decided to test *Rlv*. 3841, *bvs4* / *bvs5* gene mutants and complemented strains for enzymatic activities of both the enzymes. The strain *E.coli* DH5α, which has both the enzyme activities, was used as the reference strain (positive control).

The results of the nitrate reductase assay presented no evidence to support the hypothesis that the *bvs5* gene encodes nitrate reductase function since neither the *Rlv*.3841 wild type nor the mutants / complemented strains showed indication of any nitrate reductase activity. However, since the native *Rlv*. 3841 showed sulphite oxidase activity vis-à-vis *E.coli* DH5α, the mutants and complemented strains were also investigated for sulphite oxidase activity. The results showed that the sulphite oxidase activity was abolished in the *bvs5* mutant. The activity was restored when the *bvs5* mutant was complemented with an intact copy of the *bvs5* gene. Complementation of *bvs5* mutant with *bvs4* gene did not restore enzymatic activity. These observations are sufficient to show that the *bvs5* gene functions as a sulphite oxidase and not as a nitrate reductase.

Mutation in *bvs4* (aliphatic nitrilase) affected *bvs5* (sulphite oxidase) activity which was restored on complementing the mutant with a functional copy of *bvs5*. However, mutations in *bvs5* did not affect the *bvs4* (nitrilase) activity of mutants. This suggests that *bvs4* and *bvs5* may constitute an operon and that they might be transcribed as a polycistronic mRNA in which *bvs4* is upstream to *bvs5* explaining why *bvs4* mutation affects *bvs5* function and not vice-versa. This observation was in line with the predictions of the bioinformatic tools used for operon prediction.

In this part of thesis, we have just reviewed the work described in the previous chapters. We will now discuss the significance of this work in investigating the roles played by genes in bacteria and more specifically in rhizobia and rhizobial symbiosis. The contribution and limitations of this work will also be discussed. Based on the data and knowledge gathered from this work, we will also discuss possible future work which could be carried out in determining enzyme function and annotation and in operon prediction.

6.2. Significance and contributions of this thesis in rhizobia study

The term symbiosis is used to describe interaction between two or more different biological species. The symbiosis between rhizobia and legumes is an endosymbiotic interaction, which means that one partner lives within the tissue of the other, either within the cells or extracellularly.

6.2.1. Symbiosis, gene transfer and bacterial competition

Symbiosis is being increasingly recognized as an important selective force behind evolution, with many species having a long history of interdependent co-evolution. In rhizobia-legume symbiosis, evolutionary processes like gene transfer may have occurred in order to promote plant infection and enhance bacterial competition.

Recent studies appear to confirm the above hypothesis. Lorenz and Wackernagel (1994) reviewed natural gene transfer in bacteria. They mentioned that gene transfer normally happened in communities of bacteria for adapting to environments like water and soil. Soil bacteria like *R. leguminosarum* and *S. meliloti* were found to be capable of gene transfer by natural transformation (Courtois *et al.*, 1988, Selander *et al.*, 1986). Interspecies transformation was also found to happen in rhizobia, e.g. *nif*

genes were observed to be conserved within species but the arrangement of genes in the operon was different (Kaluza *et al.*, 1985).

Gene transfers happen frequently in *Rhizobium* and *Burkholderia* which might ensure higher genetic cohesion and prevent clonal population structure (Souza *et al.*, 1992, Wise *et al.*, 1995). Eisen (2000) reviewed and analyzed complete genome sequences of different bacteria and found that the influence of horizontal gene transfer (HGT) was more common than mutation in acquiring new functions in the bacterial genome and in its evolution. Eisen found that, in rhizobia, megaplasmids were transferred only between closely related strains or species, which suggested relatively narrow evolution occurring within rhizobia. Turner and Young (2000) studied glutamine synthetases in rhizobia and found that this gene was transferred between different rhizobia (*Rhizobium, Mesorhizobium* and *Bradyrhizobium*). This result was cited by Gogarten *et al.* (2002) as evidence of HGT effecting prokaryotic evolution. However, Philippe and Douady (2003) indicated that orthologous genes were less affected by HGT, so the history of microorganisms should be represent by both orthologous genes and genes acquired from other strains. Tian *et al.* (2010) reported lateral gene transfer in isolates from *Vicia faba*, which was thought to be a driver of evolution in rhizobia. Rogel *et al.* (2011) indicated that HGT was the important factor to extend host range of species or genera in rhizobia by transferring symbiotic plasmids or genes.

The plasmid pRL8 of *R. leguminosarum* 3841 is the smallest plasmid in the strain. The genes on the plasmid have attracted less attention unlike the genes on plasmid pRL10 which harbors genes that are important in nodulation, host-range determination, nitrogen fixation etc. However, recent studies have started to shed some light on the importance of the genes located on pRL8. For example, the gene expression analysis by Ramachandran *et al.* (2011) showed that many genes present on pRL8 were specifically up-regulated in the rhizosphere of the pea plant, a plant closely related to the genus *Vicia*.

pRL8 is as a conjugative plasmid and is thus easily transferred between strains of rhizobia. Vanderlinde *et al.* (2014b) identified a gene cluster on pRL8 that was required for homoserine utilization by *Rlv*. 3841. Mutation of

these genes resulted in a decreased ability of the mutant to compete and nodulate pea and lentil plants. L-homoserine is secreted by pea roots and found in the pea rhizosphere. The ability to utilize homoserine is positively correlated with competitiveness for nodulation of pea plants. Since the mutants were not capable of utilizing homoserine they were at a disadvantage over the wild-type strains that could metabolize the homoserine and successfully nodulate the host plant, which clearly showed that these kinds of mutation would lead negative effects on nodulation competition. The same study also noted that transfer of pRL8 by conjugation within the population resulted in the transfer of the ability to utilize homoserine within the population.

The above studies suggest that the transfer of pRL8 is a means of evolutionary adaptation to the rhizosphere of the pea and closely related plants resulting in transfer of the competitiveness for nodulation of these plants. In my work, the *bvs5* gene was demonstrated to be involved in rhizobia nodulation competition for *Vicia* plants and pea. Since it is located on pRL8, like homoserine metabolic relevant gene which were discussed by Vanderlinde *et al.* (2014a), this result suggests that more competitiveness relevant genes could possibly be found on pRL8. Young *et al.* (2006) suggested that pRL8 was conjugative plasmid and a number of genes on it having unknown functions were transferred from other bacteria. Ramachandran *et al.* (2011) studied the expression of genes on pRL8 and found that many genes on it were upregulated in pea rhizosphere. The evidence suggests that the genes on pRL8 might be transferred among bacteria in the rhizoshpere of pea or related plants which might have a similar selection stress for rhizobia.

6.2.2. Use of plasmids in the study of gene inactivation

During the course of this work, we demonstrated the usefulness of the technique of gene inactivation using the plasmid pK19mob for mediating gene insertion and inactivation. Gene inactivation by transformation is commonplace in studying the effects and functions of genes in rhizobia. Schafer *et al.* (1994) introduced the suicide plasmid pK18mob and pK19mob which are widely used. Karunakaran *et al.* (2010) used pK19mob insertion to identify the role of *Rhizobium BacA* genes in bacteroid formation on galegoid, but not phaseoloid legumes. Yurgel *et al.* (2013) created a *Sinorhizobium*

meliloti Rm1021 pSymA deletion mutant library using pK19mob insertion to study the effect of mutating different regions of pSymA on bacteria growth and symbiosis.

The plasmid pDG71 was used to complement the mutations in the inactivated genes using a novel protocol designed by Lad (2013). The success of both the techniques was verified using PCR. The plasmid pDG71 has a strong promoter and the complemented strains showed excellent expression of the cloned genes in terms of enzyme activity. Vanderlinde *et al.* (2009) and Vanderlinde *et al.* (2011) used pDG71 to construct gene complementation strains to study lipopolysaccharide synthesis genes and genes involved in cell envelope development. The protocol for the complementation can be used to clone any gene regardless of the restriction sites contained in it. Also, the protocol used the technique of rubidium chloride mediated chemical transformation (heat-shock method) to transform the plasmid into *Rlv*. 3841, which we believe, is the first time this has been used in *R. leguminosarum*.

Other methods were previously used for introducing genes into rhizobia. Garg *et al.* (1999) introduced electroporation as a transformation method for *R. leguminosarum*, and this was applied by Brito *et al.* (2010) to study the effect of *hupE* on *R. leguminosarum* hydrogenase activity. Calcium chloride, magnesium chloride and rubidium chloride have also been used to perform chemical transformation in other bacteria, but never for rhizobia (Bullerjahn and Benzinger, 1982, Karbarz *et al.*, 2009, Krysciak *et al.*, 2011).

The use of these techniques proved to be useful in unraveling the functions of *bvs4* and *bvs5* genes as they had the following advantages**:** high transformation rate, easily detectable and stable in medium without antibiotic. The study clearly demonstrates the use of these two plasmids in understanding gene function in *Rlv*. 3841. The technique could be used to understand the functions of other genes in *Rlv*. 3841 or could be used in other strains and biovars of rhizobium for the same purpose.

6.2.3. Operon-prediction

Operon prediction using bioinformatic tools was performed as a part of the research work. The results suggest that the *bvs* genes in *Rlv*. 3841 and *R. leguminosarum* WSM2304 may form operons. Based on nodulation

competition tests and enzyme assays using wild-type *Rlv*. 3841, *bvs4* and *bvs5* mutants, complemented and cross-completed strains, we speculate that in *Rlv*. 3841, the *bvs4* and *bvs5* genes may form an operon or part of an operon wherein the former gene is upstream to the latter.

Although further studies are required to completely characterize the operon to locate it promoter, operator and regulatory genes, we consider that the ability to predict an operon and verify it by polarity of gene mutation is of significance.

6.2.4. Demonstration of enzyme activity

Perhaps the most important contribution of this thesis has been the demonstration of enzyme activities of the *bvs4* and *bvs5* genes. Although the genes were annotated based on sequence similarity and homology studies, there was no clear evidence to suggest that the functions ascribed in the annotations were correct. Once again, using the wild-type *Rlv*. 3841, *bvs4* and *bvs4* mutants, complemented and cross-complemented strains the enzymatic activities of the two genes were revealed without leaving any doubt. The *bvs4* can now be annotated as an aliphatic nitrilase (without calling it 'putative') while *bvs5* can be annotated specifically as a sulphite oxidase. This approach using the published putative annotation to confirm gene function by mutation and complementation can be used to determine functions of other genes thus removing the ambiguity in the annotation of genes.

6.2.5. Strain competition : genetic factors and effectiveness of rhizobia

Strain competition is an important factor in assessing rhizobial inoculation effectiveness. Rhizobia can fix nitrogen by forming symbiosis with legumes, which makes them a valuable nitrogen provider in agriculture and natural ecosystem. Some institutions and companies have applied rhizobial inoculants to home-based, small legume fields and large scale in order to promote the growth of crops. Examples could be found in soybean inoculations in Brazil, South Africa, America and Europe (Albareda *et al.*, 2009, Botha *et al.*, 2004, Rumjanek *et al.*, 1993).

Screening effective rhizobial inoculants is the key step in evaluating rhizobia. Previous studies show that native rhizobia strains are better at nitrogen fixation and competition as compared to inoculants (Tian, C.F. and James, E. Personal communication). However, if the soil lacks the native strains

required for the nodulation of legumes, rhizobial inoculants could be used with the crops for better yields. In order to identify good inoculants, field nodulation test and lab-based competition nodulation tests are necessary for determine the superiority of rhizobia strains for use as inoculants.

Recent studies on rhizobial competition have led to the discovery of a number of genes involved in symbiosis and competition. Wang *et al.* (2010) carried out pK19mob insertion mediated gene disruption in *Sinorhizobium meliloti* Rm1021 to create mutants in 14 different genes. Competition tests showed decreased competitiveness in 11 mutants. They concluded that these genes might play roles in rhizobium physiology and interaction between rhizobia and host plants. Another studied conducted by Kohler *et al.* (2010) found that mutating putative inositol catabolic genes of *Sinorhizobium meliloti* strains led to decreased competitiveness and nodule occupancy. Hence they concluded that inositol metabolism and regulation were important for *Sinorhizobium meliloti* symbiosis. Utilization of galactose and glycerol has also been shown to be connected to symbiosis and affecting rhizobial competition (Ding *et al.*, 2012, Geddes and Oresnik, 2012). An interesting study was carried out by Ampomah and Jensen (2013) to show that, unlike the case in *Sinorhizobium meliloti* 1021, mutation of trehalose utilization genes in *Mesorhizobium loti* MAFF303099 had no effect on rhizobial competitiveness, suggesting that trehalose utilization genes were not involved in *Mesorhizobium loti* symbiosis.

Work contained in this thesis is a good example of laboratory experimentation to evaluate competition between two strains for nodule formation in a single plant, especially if one of the competitors is genetically modified. Pérez-Giménez *et al.* (2011) reviewed rhizobial competition and indicated that genes which are related to symbiosis formation would have obvious effect on competition, which could possibly suggest that *bvs* genes might also involve in symbiosis. Work in this thesis suggests that *bvs5* genes might play a role in the symbiosis process. Sulfur is an essential element for both plants and bacteria in synthesis of protein, essential vitamins and co-factors. Rhizobia, which possess sulfur-oxidizing enzymes can promote plant growth when co-inoculated with their hosts (Anandham *et al.*, 2007). Manyani *et al.* (2005) points to the presence of an iron-sulfur subunit in Ni-Fe hydrogenase of *R. leguminosarum* biovar *viciae* UPM791. Although information is limited, sulfur-oxidation is seen in some *Rhizobium* strains. El-Tarabily *et al.* (2006) found

two *Rhizobium* strains that exhibit sulfur-oxidizing activities and could promote maize growth. The product of the *bvs5* gene could be involved directly or indirectly in the symbiotic process by influencing or taking part in these activities.

6.2.6. Study of metabolic diversity of test strains and field strains

The study of metabolic patterns using the Biolog GN2 MicroPlate system showed that the strains used in the study were metabolically similar. The Biolog GN2 Microplate system is a convenient method to assess carbon substrate utilization in rhizobia. Phenotypic features are important in defining new species and recent studies have applied Biolog GN2 plates for characterization of phenotypes. Pulawska *et al.* (2012) isolated three rhizobium strains from cherry plum. 16S rRNA analysis showed that these three isolates were probably identical to each other but DNA-DNA hybridization distinguished them from other species. In the Biolog GN2 results, the novel isolates were all able to metabolise beta-hydroxybutyric acid but not L-fucose. Along with genotypic study results, they defined these isolates as a new species, *Rhizobium skierniewicense*. Silva *et al.* (2014) analysed six rhizobium strains from cowpea in North Brazil. Housekeeping gene analysis showed their closest strain was *Bradyrhizobium huanghuaihaiense* CCBAU 23303^{T}. DNA-DNA hybridization, antibiotic resistance and Biolog metabolic study showed that they were distinct from other *Bradyrhizobium* strains. They defined the isolates as members of a new species - *Bradyrhizobium manausense.*

Physiological studies on rhizobia have also used Biolog GN2 MicroPlates. Florentino *et al.* (2012) studied 25 rhizobia strains isolated from nodules of *Leucaena leucocephala* and *Phaseolus vulgaris* trap plants. They evaluated the ability of these strains to fix nitrogen using carbon sources included in the Biolog GN2 MicroPlate. Their results showed that the inability to use the substrates was directly related to their inability to fix nitrogen and lowered their competitiveness. Similarly, Wielbo *et al.* (2012) assessed the effect of biotic and physical factors in soil on the metabolism using the Biolog GN2 system and competitiveness studies. Their results revealed that a number of physiological traits could affect rhizobial metabolism and competition, and perhaps the more important factors were the environment factors.

The use of Biolog data for clustering is common in recent studies to describe new species (Xu *et al.*, 2011, Zheng *et al.*, 2013) and to investigate the host diversity of new isolates (Degefu *et al.*, 2012, Noisangiam *et al.*, 2012). Beauregard *et al.* (2004) analysed 13 rhizobia isolates from Kura clover (*Trifolium ambiguum* M.B.). The data from the Biolog GN2 MicroPlate was clustered using UPGMA to demonstrate their diversity. The results showed that the similarities between groups varied from 38% to 92% and the results differed from their host specificities and the results of genetic analysis. Some of the isolates showed unique substrate utilization. Wolde-meskel *et al.* (2004a) studied 87 rhizobia strains from native woody legumes in Southern Ethiopia. Biolog GN2 data and clustering analysis showed that 32 slow-growing strains and *Bradyrhizobium* reference strain grouped together; subgroups were also observed. They indicated suggested that the Biolog method had the capability of discriminating bacterial strains. Another South Ethiopian rhizobial diversity study was carried out by Wolde-meskel *et al.* (2004b). Biolog GN2 data analysed using UPGMA separated these isolates into 13 clusters and only 20% of strains were linked to reference strains. Their results suggested that these isolates were formed different metabolic and genetic groups and were less likely to be linked to reference strains. Wielbo *et al.* (2010) came to some interesting conclusions from their studies using Biolog and clustering analysis : rhizobial strains utilizing more of the non-sugar substrates demonstrate more competitiveness. Wielbo concluded that in rhizobial nodule populations, the balance between sugar and non-sugar utilization normally serves as a standard to group different strains.

To summarize, Biolog GN2 metabolic fingerprinting method and clustering analysis are useful in describing new species and in diversity studies. Such studies can not only help group metabolically similar rhizobia together but also help identify unique substrate utilization patterns which may aid their identification. In my study, clustering of rhizobia in Biolog GN2 assay was clearly observed based on this method and was seen to be similar to *celC* phylogenetic network. Studies using this method showed differences in utilization of carbon sources and clusters were observed grouping Scottish strains, Swedish strains, *R. pisi* DSM30132 with *R. fabae* CCBAU33202 and *R. phaseoli* ATCC14482 with *R. etli* CFN42, respectively. In sugar utilization, similar differences and clustering were observed. *R. phaseoli* ATCC14482 did not utilize of the sugars in the Biolog GN2 plate. This work could contribute to

further study of rhizobial metabolism of certain substrates to investigate utilization pathways and genes associated with them.

6.2.7. Shortcomings of this study

However, the study is not without its share of shortcomings. Perhaps, one of the most important shortcomings was the lack of reliable completely assembled genome sequences for all the test strains. Of all the fully sequenced strains only *Rlv*. 3841 and *R. leguminosarum* WSM2304 were found to carry all the five *bvs* genes using BLASTn. The genome data for other strains was incomplete resulting in mismatch between the results of BLAST search for *bvs* genes in the sequence data and the results obtained by PCR amplification of the genes. For example BLAST search for *bvs* genes in the sequence data of *R. leguminosarum* USDA2370^T showed the presence of only *bvs1* and *bvs2* in the strain. However, PCR amplification using primers specific to each of the *bvs* genes showed that all the five genes were present in the strain. The presence of such scenarios leads to confusion and misinterpretation of data and results. However, it would also be prudent to acknowledge that sequencing all the strains of rhizobium might be an expensive task. A global initiative led by Dr. Wayne Reeve of Murdoch University is underway to sequence a large number of rhizobial strains under the Genomics Encyclopedia of Bacteria and Archaea-Root Nodule Bacteria (GEBA-RNB) initiative. Such initiatives will result in more assembled genomes which can then be used reliably for the type of work described in this thesis.

The Genomic Encyclopedia of Bacteria and Archaea (GEBA) project sequenced thousands of bacterial and archaeal genomes. The results of this project have widely benefited scientific investigations such as improving genome annotation, phylogenetic analysis, gene discovery, genome evolution and transfer, microbial classification and establishing relationship between the genotype and phenotype (Reeve, 2013). GEBA-RNB is one branch of the project which aims at sequencing 100 RNB (Root Nodulating Bacteria) strains isolated from different regions of the world. The aim of the project is the study of evolution and mechanisms of symbiotic nitrogen fixation (SNF) by RNB. The results of the project are expected to benefit world agricultural productivity, farming economy and environmental

sustainability. The fully assembled genome sequences of some of the strains used in the project have recently been published (Reeve, 2013, Reeve *et al.*, 2010a, Reeve *et al.*, 2013).

6.3. Possible further study based on data of this thesis

During the course of investigation into the association between the presence of *bvs* genes and the ability of the strain to nodulate *Vicia* or related plants, two strains were exception to the rule that ability to infect *Vicia* or related plants correlates to the presence of *bvs* genes. These two strains were *R. fabae* CCBAU33202^T and *R. leguminosarum* WSM2304.

R. fabae CCBAU33202^T can form effective nodules on common *Vicia* and related plants. Results from *celC* gene phylogeny show that the strain is closely related to *R. pisi* DSM30132^T. *R. fabae* CCBAU33202^T genome was found to be wanting in all the five *bvs* genes in both PCR and local BLAST analysis. Most isolates from *Vicia faba* have a wide host range and possess all of the five *bvs* genes. There is a possibility that the strain might have a different and unrelated nodulation competition mechanism. The opposite scenario was observed in *R. leguminosarum* WSM2304 which is a biovar *trifolii* strain but possesses all the five *bvs* genes on its plasmid pRLG201. Reeve *et al.* (2010b) sequenced this strain and reported that *R. leguminosarum* WSM2304 had a better competitive ability and a wider host range than other rhizobial strains from biovar *trifolii*.

One possible explanation for the behavior of the two strains could be that the genes were horizontally transferred between the biovars, thus providing exogenous genes and improving the ability of the biovar to compete and form nodules on the host plants. *R. leguminosarum* WSM2304 was found to possess all five *bvs* genes and studies by Ramachandran *et al.* (2011) and Vanderlinde *et al.* (2014b) suggest that genes on pRL8 may be involved in determining rhizobial host range competition ability. It would be interesting to see if the genes could also increase the competitiveness of *R. leguminosarum* biovar *trifolii* WSM2304 for nodulation of some clover species as they do in *Rlv*. 3841 for *Vicia* plants.

Enzyme assays to elucidate the functions of *bvs4* and *bvs5* genes have suggested that the *bvs4* and *bvs5* genes may constitute or be a part the functional genes of an operon. Further studies are required for the complete characterization of the operon to locate its promoter, operator and regulatory

genes. In an operon, a single polycistronic mRNA translates a number of functional genes. A possible method to identify an operon could be to reverse transcribe the mRNA and check the sequence of product against the genome sequence to determine the number of genes contained in it. An investigation of the *Mesorhizobium loti* glycogen operon was conducted by Lepek *et al.* (2002) using this method. Besides, application of RT(reverse transcription)-PCR has also been applied to identify operons in *Azoarcus, Xanthomonas* and *Cyanobacterium* (Egener *et al.*, 2001, Nguyen *et al.*, 2012, Weber *et al.*, 2007).

Enzymatic function of *bvs5* has been identified and has been shown to play an important role in nodulation competition. Hence, it would be an interesting project to study the plant exudates, especially from *Vicia* plants, to detect presence of compounds that can be used as substrates by *bvs5* gene product. This will help identify the precise compounds that confer nodulation advantage to the strains carrying this gene.

Recent studies investigating the role of nitrilases in plant-microbial associations have helped understand the possible role of *bvs4* gene in the light of its function as derived from this study. Gong *et al.* (2012) have indicated that nitrilases are widely found in plants and bacteria and the enzyme family is involved in a range of metabolic processes. Spaepen and Vanderleyden (2011) reviewed the pathway on plant auxin production in plant-bacterial symbiosis. They found the presence of specific nitrilases in bacteria that convert indole-3-acetonitrile (IAN) to the plant auxin indole-3-acetic acid (IAA). However, in my work *Rlv*. 3841 did not demonstrate the ability to growth on indole-3-acetonitrile in minimum medium. Study by Mano and Nemoto (2012) and Zhao (2012) also mentioned that this pathway was not the main source of IAA synthesis, and nitrilases of this pathway were normally produced by plants. Ahsan and Stevenson (2014) mentioned that some nitrilases found in bacteria were involved in nitrogen assimilation and other processes. According to Piotrowski *et al.* (2001) and Piotrowski (2008), nitrilases could participate in nitrogen transport mechanisms in which bacterial nitrilases could also play a part in assimilation nitrogen (Howden *et al.*, 2009).

Studies on plant symbiosis and pathogenic bacteria generally support the idea that bacterial hydrolysis of nitriles plays a role of detoxification. Stam *et al.* (1985) mentioned that *Rhizobium* ORS571 could assimilate cyanide but the key enzyme responsible in the assimilation process was nitrogenase. Legras *et al.* (1990)

reviewed different pathways of hydrolysis of nitriles and indicated that nitrilase and nitrile hydratase were commonly involved in the detoxification process. Fallon *et al.* (1997) studied a plant pathogen, *Pseudomonas putida*, and found it could degrade nitriles. Wittstock and Gershenzon (2002) pointed out that nitriles were a kind of toxin produced by plants as a defence against pathogens, whose mechanism was confirmed by Banerjee *et al.* (2002). Feng *et al.* (2008) found that Mesorhizobium spp. F28, isolated from a wastewater treatment system, has a nitrile hydratase (NHase)–amidase (AMase) system to convert nitrile that could possibly be used for nitrile biotransformation. Veselá *et al.* (2010) studied the soil actinobacteria *Rhodococcus rhodochrous* PA-34, *Rhodococcus* sp. NDB 1165 and *Nocardia globerula* NHB-2 grown in the presence of isobutyronitrile and found that the strains exhibited nitrilase activities towards a benzonitrile system.

The studies mentioned above have demonstrated that the nitrile hydrolysis enzymes are widely present on genomes of soil bacteria and serve different roles to adapt to the environment. Based on the studies above, it is possible that *bvs4* may have a role to play in rhizobium-plant defence, nitrogen assimilation and transport or may be involved in the detoxification of toxic nitriles. Further studies could be carried out concentrating on potential substrates of *bvs4* to understand its role in legume plants and possible detoxification processes. Root exudates of the host plants of biovar *viciae* like pea can be analysed to detect the presence of nitriles that can be metabolised by the nitrilase produced by *bvs4* to elucidate its function in the symbiotic process.

Workun *et al.* (2008) observed that the enzyme sulphite oxidase (*bvs5*) contained a Mo-PPT cofactor and a Cys-protein-Mo ligand that played critical role in maintaining the integrity and sustainability of the enzyme. Kappler (2011) indicated that the enzymes could be linked to the detoxification of sulphite produced from amino acid degradation in animals and plants. Mihara and Esaki (2002) indicated that cysteine desulfurase could provide sulphur source for biosynthesis of molybdopterin cofactor, a component of the enzymes belonging to the family of Mo-binding oxidoreductases; of which *bvs5* is a member. Rubio and Ludden (2008) mentioned that cysteine desulfurase could support the synthesis of iron-Mo cofactor in nitrogenase. These studies suggest two possible pathways that *bvs1* could join : synthesis of *bvs5* cofactor or nitrogenase cofactor.

In soil, sulphite can be found in both inorganic and organic forms and it is thought to damage plant roots, especially in acidic conditions (Chen *et al.*, 2009). Sulphite oxidation may occur in both plants and rhizobia for detoxification in the rhizosphere. Other protective functions like preventing DNA, protein and lipid damage may also be performed by these enzymes. In addition, sulfur-metal clusters are needed for nitrogen fixation as a part of the enzyme nitrogenase (Howard and Rees, 2006). Further studies on *bvs5* could investigate sulphite oxidase function in symbiosis, including identifying the *bvs5* substrates in plant secretions or exudates which could tell us the participation of this enzyme in symbiosis. Sugiyama and Yazaki (2012) suggested that some root volatiles from legumes could attract rhizobia from distance. Hence, it is possible that the rhizobial detection and processing of sulfur-containing flavonoids and root volatiles might need the activity of sulfite oxidase. The studies might also be able to suggest if *bvs1* and *bvs5* genes are connected to each other as a part sulfur metabolism.

The function of *bvs4* and *bvs5* genes being elucidated, the structure prediction of *bvs4* and *bvs5* encoded proteins using computer modeling can be done using a variety of software as a part of the future work in order to study the protein-substrate binding. However, as in the case of confirming bioinformatic predictions with laboratory based work, it would be a good idea to confirm the 3-D structure of the crystallized protein using X-ray diffractometry or X-ray crystallography.

Looking at the bigger picture involving all the five *bvs* genes, we can say that the enzymatic activities of *bvs4* and *bvs5* have been confirmed, *bvs1* could possibly be linked to biosynthesis of cofactor required by the *bvs5* coded enzyme. Genes in the families that contain *bvs2* and *bvs3* have been reported as potential regulators of a number of genes (Christopherson *et al.*, 2012, Maddocks and Oyston, 2008). Since they are always found together in different strains, they are likely all involved in the same process. But their functions are unknown and have not been observed to have any obvious effect on rhizobial growth and nodulation. Further studies could be carried out to clarify the organization and functioning of the operon, these studies could possibly be reverse transcription PCR and competitiveness test using more *R. leguminosarum* biovar *viciae* strains.

The study of the metabolic profiles of the test strain revealed an interesting finding. The strain *R. phaseoli* ATCC14482^T lacks the ability to utilize all of the

common sugar substrates included in the Biolog GN2 MicroPlate. Ramirez-Bahena *et al.* (2008) have reported similar findings wherein they found that the strain utilized pyruvate, an important metabolic intermediate, very poorly. This distinct metabolic trait of the strain is worth further investigation and could help shed light on the metabolism in rhizobia. A thorough investigation into the genetic differences between *R. phaseoli* ATCC14482^T and *Rlv*. 3841 may shed some light on the metabolism in rhizobia.

6.4. Final remarks

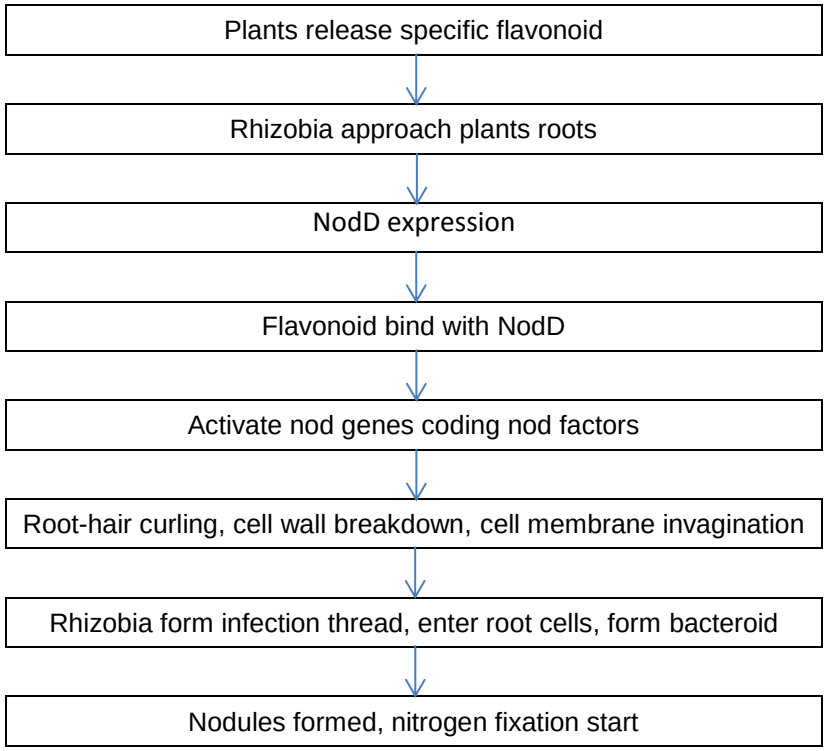


Figure 6.1. Steps involved in the Nodulation process

Symbiosis between rhizobia and legumes involves a number of biochemical reactions, changes and processes including flavonoid synthesis and recognition, nodulation protein synthesis, changes in legume root structures, formation of bacteroids and finally, nitrogen fixation. During this process, relevant genes participate in two ways : by joining the main metabolism for symbiosis - like the nodulation genes, which determine the biovar of rhizobia strains; or by supporting the formation of symbiosis. Genes relevant to rhizobial competitiveness and others that help overcoming plants specific defense barriers are also important in nodulation.

Starting with the discovery of five genes in one biovar of a species but not another in a local population, the study of *bvs* genes was extended to a number of strains from species and biovars from different locations. As more and more genomes are sequenced, more genes will be discovered which may show preponderance in one species or biovar over other. As was observed for the *bvs* genes, such genes may confer traits that provide an adaptive advantage to the strain or species carrying those genes. Further investigations on *bvs* genes should be done to understand in-depth the detailed mechanism of their effects on rhizobium biovar *viciae* strain competitiveness, especially on the pathway which they participate during the formation of symbiosis. In addition, more and better quality sequence data from other *Rhizobium* biovar *viciae* strains, coupled with laboratory based evidence of *bvs* genes and other possible specific gene functions are the need of the day to better understand the relationship between these genes and the environment.

APPENDIX : MEDIA COMPOSITION

1. Tryptone-Yeast extract (TY) medium

Tryptone	5 g
Yeast extract	3 g
$CaCl_2$	0.89 g
Agar (for plates)	15 g
pH	7.2-7.4
Deionized Water	Bring up to 1 litre

2. Fahraeus medium

$CaCl_2$	0.132 g
$MgSO_4.7H_2O$	0.12 g
KH_2PO_4	0.1 g
Na_2HPO_4	0.075 g
Fe-citrate	5 mg
$MnCl_2.4H_2O$	0.07 mg
$CuSO_4.5H_2O$	0.07 mg
$ZnCl_2$	0.07 mg
H_3BO_3	0.07 mg
$Na_2MoO_4.2H_2O$	0.07 mg
pH	7.5
Deionized Water	Bring up to 1 litre

3. Physiological saline

NaCl	9 g
Deionized water	Bring up to 1 Litre

4. Rhizobium Defined Medium (RDF)

Stocks:

RDM A stock (10X) :

KNO_3	6 g
$CaCl_2.2H_2O$	1 g
$MgSO_4.7H_2O$	2.5 g
$FeCl_3.6H_2O$	0.1 g
Deionized Water	Bring up to 1 litre

RDM B stock (10X) :

K_2HPO_4	10 g
KH_2PO_4	10 g
Deionized Water	Bring up to 1 litre

Biotin stock :

Biotin	0.0125 g
Deionized Water	Bring up to 50 ml

Thiamine stock : 10 mg/mL

Thiamine	0.5 g
Deionized Water	Bring up to 50 ml

RDM :

RDM A stock	100 ml
RDM B stock	100 ml
Sucrose	5 g
Thiamine stock	1 ml
Biotin stock	4 ml
Agar (for plate culture)	15 g
Deionized Water	Bring up to 1 litre

LIST OF ABBREVIATIONS

ACC	1-Aminocyclopropane-1-Carboxylate
AFLP	Amplified Fragment Length Polymorphisms
ANOVA	Analysis of Variance
API	Application Programming Interface
APX	Ascorbate Peroxidase
ARDRA	Amplified rDNA (Ribosomal DNA) Restriction Analysis
ATCC	American Type Culture Collection
ATP	Adenosine Triphosphate
BLAST	Basic Local Alignment Search Tool
BLASTn	Basic Local Alignment Search Tool for nucleotide
BLASTp	Basic Local Alignment Search Tool for protein
bvs	biovar *viciae* specific
CCBAU	Culture Collection, Beijing Agricultural University
CFN	Centro de Investigacion sobre Fijacion de Nitrogeno
ChrN	Chromosomal nitrilase
CIAT	Centro Internacional de Agricultura Tropical
CIP	Calf-Intestinal Phosphatase
CN-hydrolases	Carbon-Nitrogen hydrolases
cPMP	Cyclic Pyranopterin Monophosphate
C-W-E motif	Cysteine-Tryptophan-Glutamic acid motif
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
DNRA	Dissimilatory Nitrate Reduction to Ammonia
dNTP	deoxyribonucleic acid
DSM	Deutsche Sammlung von Mikroorganismen und Zellkulturen
E-K-C catalytic triad	Glutamic acid, Lysine and Cysteine catalytic triad
Euk-NR	Eukaryotic Nitrate Reductases
FAME	Fatty Acid Methyl Ester
FASTA	FAST Alignment
GN	Gram Negative
GTP	Guanosine Triphosphate
HGT	Horizontal Gene transfer
IAA	Indole-3-Acetic Acid
IAN	Indole-3-Acetonitrile
ITS	Internal Transcribed Spacer
KDO	2-keto-3-deoxyoctanoic acid
KDPG	2-dehydro-3-deoxy-phosphogluconate aldolase
KEGG	Kyoto Encyclopedia of Genes and Genomes
KGA	a-Ketoglutaric Acid
LB	Luria Broth
LMG	Laboratorium voor Microbiologie
LPS	lipopolysaccharide
L-PSP	liver perchloric acid soluble protein
LSD	Least Significant Difference
LTTRs	LysR-Type Transcriptional Regulators

MCS	Multiple Cloning Site
MEGA	Molecular Evolutionary Genetics Analysis
MLST	Multilocus Sequence Typing
MPT	Molybdopterin
NADP+	oxidized Nicotinamide Adenine Dinucleotide Phosphate
NADPH	reduced Nicotinamide Adenine Dinucleotide Phosphate
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
NR	Nitrate Reductase
NRCPB	National Research Centre on Plant Biotechnology
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
pfu	*Pyrococcus furiosus*
PLP	Pyridoxal 5'-Phosphate
PMF	Proton Motive Force
POX	Peroxidase
pSym	Symbiosis plasmid
QS	Quorum Sensing
RDM	Rhizobia Definition Media
RFLP	Restriction Fragment Length Polymorphism
Rlv	Rhizobium Leguminosarum
RNA	Ribonucleic Acid
rRNA	Ribosomal RNA
RT-PCR	Reverse Transcription PCR
SAM	S-Adenosyl Methionine
SNF	Symbiotic Nitrogen Fixation
SO	Sulfite Oxidases
TMAO	Trimethylamine-N-Oxide
TPF	Targeted PCR Fingerprinting
TY	Tryptone Yeast
UPGMA	Unweighted Pair Group Method with Arithmetic Mean
USDA	United States Department of Agriculture
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

REFERENCES

1. Aguey-Zinsou, K. F., P. V. Bernhardt, U. Kappler and A. G. McEwan (2003). Direct electrochemistry of a bacterial sulfite dehydrogenase. *J Am Chem Soc* 125(2) 530-535.
2. Ahsan, N. and S. E. Stevenson (2014). Proteomic mapping for legume nodule organogenesis. *Proteomics* 14(2-3) 153-154.
3. Albareda, M., D. N. Rodríguez-Navarro and F. J. Temprano (2009). Use of Sinorhizobium (Ensifer) fredii for soybean inoculants in South Spain. *European journal of agronomy* 30(3) 205-211.
4. Almatawah, Q. A., R. Cramp and D. A. Cowan (1999). Characterization of an inducible nitrilase from a thermophilic bacillus. *Extremophiles* 3(4) 283-291.
5. Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). Basic local alignment search tool. *J Mol Biol* 215(3) 403-410.
6. Alvarez-Martinez, E. R., A. Valverde, M. H. Ramirez-Bahena, P. Garcia-Fraile, C. Tejedor, P. F. Mateos, N. Santillana, D. Zuniga, A. Peix and E. Velazquez (2009). The analysis of core and symbiotic genes of rhizobia nodulating Vicia from different continents reveals their common phylogenetic origin and suggests the distribution of Rhizobium leguminosarum strains together with Vicia seeds. *Arch Microbiol* 191(8) 659-668.
7. Amann, R., P. Sostak, W. Ludwig and K. H. Schleifer (1988). Cloning and sequencing of genes encoding the beta subunits of the ATP-synthases from Enterobacter aerogenes and Flavobacterium ferrugineum. *FEMS Microbiology Letters* 50(2) 101-106.
8. Ampomah, O. Y. and J. B. Jensen (2013). The trehalose utilization gene thuA ortholog in Mesorhizobium loti does not influence competitiveness for nodulation on Lotus spp. *World Journal of Microbiology and Biotechnology* 1-6.
9. Anandham, R., R. Sridar, P. Nalayini, S. Poonguzhali and M. Madhaiyan (2007). Potential for plant growth promotion in groundnut (*Arachis hypogaea* L.) cv. ALR-2 by co-inoculation of sulfur-oxidizing bacteria and *Rhizobium*. *Microbiol Res* 162(2) 139-153.
10. Aserse, A. A., L. A. Rasanen, F. Assefa, A. Hailemariam and K. Lindstrom (2012). Phylogeny and genetic diversity of native rhizobia nodulating common bean (Phaseolus vulgaris L.) in Ethiopia. *Syst Appl Microbiol* 35(2) 120-131.
11. Banerjee, A., R. Sharma and U. Banerjee (2002). The nitrile-degrading enzymes: current status and future prospects. *Appl Microbiol Biotechnol* 60(1-2) 33-44.
12. Barreto, E. F., R. Straliotto and J. I. Baldani (2012). Curing of a non-symbiotic plasmid of the Rhizobium tropici strain CIAT 899 affected nodule occupancy and competitiveness of the bacteria in symbiosis with common beans. *European Journal of Soil Biology* 50(0) 91-96.
13. Beatty, P. H. and A. G. Good (2011). Future prospects for cereals that fix nitrogen. *Science* 333(6041) 416-417.
14. Beauregard, M.-S., W. Zheng and P. Seguin (2004). Diversity of Trifolium ambiguum—nodulating rhizobia from the lower Caucasus. *Biology and Fertility of Soils* 40(2) 128-135.

15. Bergey, D. H., N. R. Krieg and J. G. Holt (1984). *Bergey's Manual of Systematic Bacteriology*: Williams & Wilkins.

16. Beyhaut, E., D. L. Larson, D. L. Allan and P. H. Graham (2013). Legumes in prairie restoration: evidence for wide cross-nodulation and improved inoculant delivery. *Plant and Soil* 1-14.

17. Bianucci, E., A. Fabra and S. Castro (2012). Involvement of glutathione and enzymatic defense system against cadmium toxicity in Bradyrhizobium sp. strains (peanut symbionts). *Biometals* 25(1) 23-32.

18. Bibi, F., E. J. Chung, A. Khan, C. O. Jeon and Y. R. Chung (2012). Rhizobium halophytocola sp. nov., isolated from the root of a coastal dune plant. *Int J Syst Evol Microbiol* 62(Pt 8) 1997-2003.

19. Bittner, F. (2014). Molybdenum metabolism in plants and crosstalk to iron. *Front Plant Sci* 5 28.

20. Bochner, B. R. (1989). Sleuthing out bacterial identities. *Nature* 339(6220) 157-158.

21. Bonnefoy, V. and J. A. Demoss (1994). Nitrate reductases in Escherichia coli. *Antonie Van Leeuwenhoek* 66(1-3) 47-56.

22. Bork, P. and E. V. Koonin (1994). A new family of carbon-nitrogen hydrolases. *Protein Sci* 3(8) 1344-1346.

23. Botha, W. J., J. B. Jaftha, J. F. Bloem, J. H. Habig and I. J. Law (2004). Effect of soil bradyrhizobia on the success of soybean inoculant strain CB 1809. *Microbiol Res* 159(3) 219-231.

24. Brenner, C. (2002). Catalysis in the nitrilase superfamily. *Current Opinion in Structural Biology* 12(6) 775-782.

25. Brewin, N., T. DeJong, D. Phillips and A. Johnston (1980). Co-transfer of determinants for hydrogenase activity and nodulation ability in Rhizobium leguminosarum.

26. Brito, B., R.-I. Prieto, E. Cabrera, M.-A. Mandrand-Berthelot, J. Imperial, T. Ruiz-Argüeso and J.-M. Palacios (2010). Rhizobium leguminosarum hupE encodes a nickel transporter required for hydrogenase activity. *J Bacteriol* 192(4) 925-935.

27. Brom, S., A. G. De Los Santos, T. Stepkowsky, M. Flores, G. Dávila, D. Romero and R. Palacios (1992). Different plasmids of Rhizobium leguminosarum bv. phaseoli are required for optimal symbiotic performance. *J Bacteriol* 174(16) 5183-5189.

28. Brom, S., A. Garcia-de los Santos, L. Cervantes, R. Palacios and D. Romero (2000). In Rhizobium etli symbiotic plasmid transfer, nodulation competitivity and cellular growth require interaction among different replicons. *Plasmid* 44(1) 34-43.

29. Bryan, J., G. Berlyn and J. Gordon (1997). Toward a new concept of the evolution of symbiotic nitrogen fixation in the Leguminosae. *In* G. H. Elkan and R. G. Upchurch eds. *Current Issues in Symbiotic Nitrogen Fixation*. Springer Netherlands. 151-159.

30. Bullerjahn, G. S. and R. H. Benzinger (1982). Genetic transformation of Rhizobium leguminosarum by plasmid DNA. *J Bacteriol* 150(1) 421-424.

31. Burton, J. C., U. Nitrogen Fixation for Tropical Agricultural Legumes Project and MIRCEN (1984). *Legume Inoculant Production Manual*: University of Hawaii, Department of Agronomy and Soil Science, College of Tropical Agriculture and Human Resources.

32. Cai, T., W. Cai, J. Zhang, H. Zheng, A. M. Tsou, L. Xiao, Z. Zhong and J. Zhu (2009). Host legume-exuded antimetabolites optimize the symbiotic rhizosphere. *Mol Microbiol* 73(3) 507-517.

33. Carlson, C. A. and J. L. Ingraham (1983). Comparison of denitrification by *Pseudomonas stutzeri*, *Pseudomonas aeruginosa*, and *Paracoccus denitrificans*. *Appl Environ Microbiol* 45(4) 1247-1253.

34. Carlson, R., S. Kalembasa, D. Turowski, P. Pachori and K. Noel (1987). Characterization of the lipopolysaccharide from a Rhizobium phaseoli mutant that is defective in infection thread development. *J Bacteriol* 169(11) 4923-4928.

35. Cary, S. C. and S. J. Giovannoni (1993). Transovarial inheritance of endosymbiotic bacteria in clams inhabiting deep-sea hydrothermal vents and cold seeps. *Proceedings of the National Academy of Sciences* 90(12) 5695-5699.

36. Chandler, M. R., R. Date and R. Roughley (1982). Infection and root-nodule development in Stylosanthes species by Rhizobium. *J Exp Bot* 33(1) 47-57.

37. Chang, Y. L., J. Y. Wang, E. T. Wang, H. C. Liu, X. H. Sui and W. X. Chen (2011). Bradyrhizobium lablabi sp. nov., isolated from effective nodules of Lablab purpureus and Arachis hypogaea. *Int J Syst Evol Microbiol* 61(10) 2496-2502.

38. Charpentier, M. and G. Oldroyd (2010). How close are we to nitrogen-fixing cereals? *Current opinion in plant biology* 13(5) 556-564.

39. Chen, L., C. Ramsier, J. Bigham, B. Slater, D. Kost, Y. B. Lee and W. A. Dick (2009). Oxidation of FGD-CaSO3 and effect on soil chemical properties when applied to the soil surface. *Fuel* 88(7) 1167-1172.

40. Chen, W.-M., S. Laevens, T.-M. Lee, T. Coenye, P. De Vos, M. Mergeay and P. Vandamme (2001). Ralstonia taiwanensis sp. nov., isolated from root nodules of Mimosa species and sputum of a cystic fibrosis patient. *Int J Syst Evol Microbiol* 51(5) 1729-1735.

41. Chen, W. X., Z. Y. Tan, J. L. Gao, Y. Li and E. T. Wang (1997). Rhizobium hainanense sp. nov., isolated from tropical legumes. *Int J Syst Bacteriol* 47(3) 870-873.

42. Christopherson, M. R., J. A. Lambrecht, D. Downs and D. M. Downs (2012). Suppressor analyses identify threonine as a modulator of ridA mutant phenotypes in Salmonella enterica. *PLoS One* 7(8) e43082.

43. Cleveland, C. J. and R. Szostak (2008). Encyclopedia of earth. *Washington, DC: Environmental*.

44. Cohn, J., R. Bradley Day and G. Stacey (1998). Legume nodule organogenesis. *Trends in Plant Science* 3(3) 105-110.

45. Collinge, D. B. and M. Hughes (1982). In vitro characterization of the Ac locus in white clover (Trifolium repens.). *Arch Biochem Biophys* 218(1) 38-45.

46. Conn, E. E. (1979). Biosynthesis of cyanogenic glycosides. *Naturwissenschaften* 66(1) 28-34.

47. Courtois, J., B. Courtois and J. Guillaume (1988). High-frequency transformation of *Rhizobium meliloti*. *J Bacteriol* 170(12) 5925-5927.

48. Cox, M. (1991). The RecA protein as a recombinational repair system. *Mol Microbiol* 5(6) 1295-1299.

49. Cox, M. M. (1993). Problems and Paradigms: Relating biochemistry to biology: How the recombinational repair function of RecA protein is manifested in its molecular properties. *BioEssays* 15(9) 617-623.

50. David, M., O. Domergue, P. Pognonec and D. Kahn (1987). Transcription patterns of Rhizobium meliloti symbiotic plasmid pSym: identification of nifA-independent fix genes. *J Bacteriol* 169(5) 2239-2244.

51. De Faria, S. M., G. T. Hay and J. I. Sprent (1988). Entry of rhizobia into roots of Mimosa scabrella Bentham occurs between epidermal cells. *Journal of General Microbiology* 134(8) 2291-2296.

52. De Ley, J. and A. Rassel (1965). DNA base composition, flagellation and taxonomy of the genus Rhizobium. *Journal of General Microbiology* 41(1) 85-92.

53. Degefu, T., E. Wolde-meskel and Å. Frostegård (2011). Multilocus sequence analyses reveal several unnamed Mesorhizobium genospecies nodulating Acacia species and Sesbania sesban trees in Southern regions of Ethiopia. *Syst Appl Microbiol* 34(3) 216-226.

54. Degefu, T., E. Wolde-meskel and Å. Frostegård (2012). Phylogenetic multilocus sequence analysis identifies seven novel Ensifer genospecies isolated from a less-well-explored biogeographical region in East Africa. *Int J Syst Evol Microbiol* 62(Pt 9) 2286-2295.

55. Del Pilar Cordovilla, M., S. I. Berrido, F. Ligero and C. Lluch (1999). Rhizobium Strain Effects on the Growth and Nitrogen Assimilation in Pisum sativum and Vicia faba Plant Growth under Salt Stress. *Journal of Plant Physiology* 154(1) 127-131.

56. Denarie, J., F. Debelle and C. Rosenberg (1992). Signaling and host range variation in nodulation. *Annual Reviews in Microbiology* 46(1) 497-531.

57. Ding, H., C. B. Yip, B. A. Geddes, I. J. Oresnik and M. F. Hynes (2012). Glycerol utilization by Rhizobium leguminosarum requires an ABC transporter and affects competition for nodulation. *Microbiology* 158(Pt 5) 1369-1378.

58. Diouf, D., D. Fall, C. Chaintreuil, A. Ba, B. Dreyfus, M. Neyra, I. Ndoye and L. Moulin (2010). Phylogenetic analyses of symbiotic genes and characterization of functional traits of Mesorhizobium spp. strains associated with the promiscuous species Acacia seyal Del. *J Appl Microbiol* 108(3) 818-830.

59. Dixon, R. and D. Kahn (2004). Genetic regulation of biological nitrogen fixation. *Nature Reviews Microbiology* 2(8) 621-631.

60. Downie, J. A. and B. P. Surin (1990). Either of two nod gene loci can complement the nodulation defect of a nod deletion mutant of Rhizobium leguminosarum bv viciae. *Molecular and General Genetics MGG* 222(1) 81-86.

61. Eardly, B., J. Young and R. Selander (1992). Phylogenetic position of Rhizobium sp. strain Or 191, a symbiont of both Medicago sativa and Phaseolus vulgaris, based on partial sequences of the 16S rRNA and nifH genes. *Appl Environ Microbiol* 58(6) 1809-1815.

62. Economou, A., A. E. Davies, A. W. B. Johnston and J. A. Downie (1994). The Rhizobium leguminosarum biovar viciae nodO gene can enable a nodE mutant of Rhizobium leguminosarum biovar trifolii to nodulate vetch. *Microbiology* 140(9) 2341-2347.

63. Eda, S., H. Mitsui and K. Minamisawa (2011). Involvement of the smeAB multidrug efflux pump in resistance to plant antimicrobials and contribution to nodulation competitiveness in Sinorhizobium meliloti. *Appl Environ Microbiol* 77(9) 2855-2862.

64. Egener, T., D. E. Martin, A. Sarkar and B. Reinhold-Hurek (2001). Role of a Ferredoxin Gene Cotranscribed with thenifHDK Operon in N2 Fixation and Nitrogenase "Switch-Off" of Azoarcus sp. Strain BH72. *J Bacteriol* 183(12) 3752-3760.

65. Eisen, J. A. (1995). The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. *Journal of molecular evolution* 41(6) 1105-1123.

66. Eisen, J. A. (2000). Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. *Current opinion in genetics & development* 10(6) 606-611.

67. Eisenberg, H. and E. J. Wachtel (1987). Structural studies of halophilic proteins, ribosomes, and organelles of bacteria adapted to extreme salt concentrations. *Annual review of biophysics and biophysical chemistry* 16(1) 69-92.

68. El-Tarabily, K. A., A. A. Soaud, M. E. Saleh and S. Matsumoto (2006a). Isolation and characterisation of sulfur-oxidising bacteria, including strains of Rhizobium, from calcareous sandy soils and their effects on nutrient uptake and growth of maize (*Zea mays* L.). *Crop and Pasture Science* 57(1) 101-111.

69. El-Tarabily, K. A., A. A. Soaud, M. E. Saleh and S. Matsumoto (2006b). Isolation and characterisation of sulfur-oxidising bacteria, including strains of *Rhizobium*, from calcareous sandy soils and their effects on nutrient uptake and growth of maize (*Zea mays* L.). *Australian Journal of Agricultural Research* 57(1) 101-111.

70. Elliott, G. N., W. M. Chen, C. Bontemps, J. H. Chou, J. P. Young, J. I. Sprent and E. K. James (2007). Nodulation of Cyclopia spp. (Leguminosae, Papilionoideae) by Burkholderia tuberum. *Ann Bot* 100(7) 1403-1411.

71. Emerich, D. W., H. B. Krishnan, A. S. o. Agronomy, C. S. S. o. America and S. S. S. o. America (2009). *Nitrogen Fixation in Crop Production*: American Society of Agronomy.

72. Encarnacion, S., M. Dunn, K. Willms and J. Mora (1995). Fermentative and aerobic metabolism in Rhizobium etli. *J Bacteriol* 177(11) 3058-3066.

73. Esfahani, M. and A. Mostajeran (2011). Rhizobial strain involvement in symbiosis efficiency of chickpea–rhizobia under drought stress: plant growth, nitrogen fixation and antioxidant enzyme activities. *Acta Physiologiae Plantarum* 33(4) 1075-1083.

74. Faghire, M., B. Mandri, K. Oufdou, A. Bargaz, C. Ghoulam, M. Ramírez-Bahena, E. Velázquez and A. Peix (2012). Identification at the species and symbiovar levels of strains nodulating Phaseolus vulgaris in saline soils of the Marrakech region (Morocco) and analysis of the otsA gene putatively involved in osmotolerance. *Syst Appl Microbiol* 35(3) 156-164.

75. Fallon, R., B. Stieglitz and I. Turner Jr (1997). A Pseudomonas putida capable of stereoselective hydrolysis of nitriles. *Appl Microbiol Biotechnol* 47(2) 156-161.

76. Feil, E. J. (2004). Small change: keeping pace with microevolution. *Nature Reviews Microbiology* 2(6) 483-495.

77. Feng, C., G. Tollin and J. H. Enemark (2007). Sulfite oxidizing enzymes. *Biochim Biophys Acta* 1774(5) 527-539.

78. Feng, Y.-S., P.-C. Chen, F.-S. Wen, W.-Y. Hsiao and C.-M. Lee (2008). Nitrile hydratase from Mesorhizobium sp. F28 and its potential for nitrile biotransformation. *Process Biochemistry* 43(12) 1391-1397.

79. Florentino, L. A., P. M. D. Jaramillo, K. B. Silva, J. S. d. Silva, S. M. d. Oliveira and F. M. d. S. Moreira (2012). Physiological and symbiotic diversity of Cupriavidus necator strains isolated from nodules of Leguminosae species. *Scientia Agricola* 69(4) 247-258.

80. Flores, M., L. Morales, A. Avila, V. González, P. Bustos, D. García, Y. Mora, X. Guo, J. Collado-Vides and D. Pinero (2005). Diversification of DNA sequences in the symbiotic genome of Rhizobium etli. *J Bacteriol* 187(21) 7185-7192.

81. Frank, B. (1889). *Ueber die Pilzymbiose der Leguminosen*: P. Parey.

82. Fred, E. B., I. L. Baldwin and E. McCoy (1932a). *Root nodule bacteria and leguminous plants*: UW-Madison Libraries Parallel Press.

83. Fred, E. B., I. L. Baldwin and E. McCoy (1932b). *Root Nodule Bacteria and Leguminous Plants*: Parallel Press.

84. Fredrickson, J. K., J. M. Zachara, D. L. Balkwill, D. Kennedy, W. L. Shu-mei, H. M. Kostandarithes, M. J. Daly, M. F. Romine and F. J. Brockman (2004). Geomicrobiology of high-level nuclear waste-contaminated vadose sediments at the Hanford Site, Washington State. *Appl Environ Microbiol* 70(7) 4230-4241.

85. Friesen, M. L. (2012). Widespread fitness alignment in the legume-rhizobium symbiosis. *New Phytol* 194(4) 1096-1111.

86. Garg, B., R. C. Dogra and P. K. Sharma (1999). High-efficiency transformation of Rhizobium leguminosarum by electroporation. *Appl Environ Microbiol* 65(6) 2802-2804.

87. Garrity, G. M., J. A. Bell and T. Lilburn (2005). *Class I. Alphaproteobacteria class. nov*: Springer.

88. Gaunt, M. W., S. L. Turner, L. Rigottier-Gois, S. A. Lloyd-Macgilp and J. P. Young (2001). Phylogenies of atpD and recA support the small subunit rRNA-based classification of rhizobia. *Int J Syst Evol Microbiol* 51(Pt 6) 2037-2048.

89. Geddes, B. A. and I. J. Oresnik (2012). Inability to catabolize galactose leads to increased ability to compete for nodule occupancy in Sinorhizobium meliloti. *J Bacteriol* 194(18) 5044-5053.

90. Georgiadis, M., H. Komiya, P. Chakrabarti, D. Woo, J. Kornuc and D. Rees (1992). Crystallographic structure of the nitrogenase iron protein from Azotobacter vinelandii. *Science* 257(5077) 1653-1659.

91. Ghosh, W. and P. Roy (2006). *Mesorhizobium thiogangeticum* sp. nov., a novel sulfur-oxidizing chemolithoautotroph from rhizosphere soil of an Indian tropical leguminous plant. *Int J Syst Evol Microbiol* 56(1) 91-97.

92. Giraud, E., L. Moulin, D. Vallenet, V. Barbe, E. Cytryn, J.-C. Avarre, M. Jaubert, D. Simon, F. Cartieaux and Y. Prin (2007). Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia. *Science* 316(5829) 1307-1312.

93. Gnanamanickam, S. S. (2007). *Plant-Associated Bacteria*: Springer.

94. Gogarten, J. P., W. F. Doolittle and J. G. Lawrence (2002). Prokaryotic evolution in light of gene transfer. *Molecular Biology and Evolution* 19(12) 2226-2238.

95. Gong, J.-S., Z.-M. Lu, H. Li, J.-S. Shi, Z.-M. Zhou and Z.-H. Xu (2012). Nitrilases in nitrile biocatalysis: recent progress and forthcoming research. *Microb Cell Fact* 11 142.

96. Gonzalez, P. J., C. Correia, I. Moura, C. D. Brondino and J. J. Moura (2006). Bacterial nitrate reductases: Molecular and biological aspects of nitrate reduction. *J Inorg Biochem* 100(5-6) 1015-1023.

97. González, V., P. Bustos, M. A. Ramírez-Romero, A. Medrano-Soto, H. Salgado, I. Hernández-González, J. C. Hernández-Celis, V. Quintero, G. Moreno-Hagelsieb and L. Girard (2003). The mosaic structure of the symbiotic plasmid of Rhizobium etli CFN42 and its relation to other symbiotic genome compartments. *Genome Biol* 4(6) R36.

98. González, V., R. I. Santamaría, P. Bustos, I. Hernández-González, A. Medrano-Soto, G. Moreno-Hagelsieb, S. C. Janga, M. A. Ramírez, V. Jiménez-Jacinto and J. Collado-Vides (2006). The partitioned Rhizobium etli genome: genetic and metabolic redundancy in seven interacting replicons. *Proc Natl Acad Sci U S A* 103(10) 3834-3839.

99. Graham, P. (1963). Antigenic affinities of the root-nodule bacteria of legumes. *Antonie Van Leeuwenhoek* 29(1) 281-291.

100. Graham, P. (1964). The application of computer techniques to the taxonomy of the root-nodule bacteria of legumes. *Journal of General Microbiology* 35(3) 511-517.

101. Graham, P. and C. Parker (1964). Diagnostic features in the characterisation of the root-nodule bacteria of legumes. *Plant and Soil* 20(3) 383-396.

102. Graham, P., M. Sadowsky, H. Keyser, Y. Barnet, R. Bradley, J. Cooper, D. De Ley, B. Jarvis, E. Roslycky and B. Strijdom (1991). Proposed minimal standards for the description of new genera and species of root-and stem-nodulating bacteria. *International Journal of Systematic Bacteriology* 41(4) 582-587.

103. Grassia, A. and J. Brockwell (1978). Enumeration of rhizobia from a plant-infection dilution assay using test plants grown in vermiculite. *Soil Biology and Biochemistry* 10(2) 101-104.

104. Gregory, T. R. (2005). Synergy between sequence and size in large-scale genomics. *Nature Reviews Genetics* 6(9) 699-708.

105. Gregory, T. R. (2011). *The evolution of the genome*: Academic Press.

106. Habtegebrial Habtemichial, K., B. Ram Singh and J. B. Aune (2007). Wheat response to N2 fixed by faba bean (*Vicia faba* L.) as affected by sulfur fertilization and rhizobial inoculation in semi-arid Northern Ethiopia. *Journal of Plant Nutrition and Soil Science* 170(3) 412-418.

107. Hacker, J. and E. Carniel (2001). Ecological fitness, genomic islands and bacterial pathogenicity. *EMBO reports* 2(5) 376-381.

108. Hadfield, K. L. and W. A. Bulen (1969). Adenosine triphosphate requirement of nitrogenase from Azotobacter vinelandii. *Biochemistry* 8(12) 5103-5108.

109. Hadri, A.-E., H. P. Spaink, T. Bisseling and N. J. Brewin (1998). Diversity of root nodulation and rhizobial infection processes. *The Rhizobiaceae*. Springer. 347-360.

110. Han, T. X., E. T. Wang, L. J. Wu, W. F. Chen, J. G. Gu, C. T. Gu, C. F. Tian and W. X. Chen (2008). Rhizobium multihospitium sp. nov., isolated from multiple legume species native of Xinjiang, China. *Int J Syst Evol Microbiol* 58(Pt 7) 1693-1699.

111. Hanin, M., S. Jabbouri, W. Broughton, R. Fellay and D. Quesada-Vincens (1999). Molecular aspects of host-specific nodulation. *Plant-microbe interactions* 4 1-37.

112. Harrison, P. W., R. P. Lower, N. K. Kim and J. P. W. Young (2010). Introducing the bacterial 'chromid': not a chromosome, not a plasmid. *Trends Microbiol* 18(4) 141-148.

113. Haukka, K., K. Lindström and J. P. W. Young (1998). Three Phylogenetic Groups of nodA and nifHGenes in Sinorhizobium and Mesorhizobium Isolates from Leguminous Trees Growing in Africa and Latin America. *Appl Environ Microbiol* 64(2) 419-426.

114. Henikoff, S., G. W. Haughn, J. M. Calvo and J. C. Wallace (1988). A large family of bacterial activator proteins. *Proc Natl Acad Sci U S A* 85(18) 6602-6606.

115. Hennecke, H., K. Kaluza, B. Thöny, M. Fuhrmann, W. Ludwig and E. Stackebrandt (1985). Concurrent evolution of nitrogenase genes and 16S rRNA in Rhizobium species and other nitrogen fixing bacteria. *Arch Microbiol* 142(4) 342-348.

116. Herridge, D. F., M. B. Peoples and R. M. Boddey (2008). Global inputs of biological nitrogen fixation in agricultural systems. *Plant and Soil* 311(1-2) 1-18.

117. Hille, R. (1996). The Mononuclear Molybdenum Enzymes†. *Chemical Reviews* 96(7) 2757-2816.

118. Hille, R. (2013). The molybdenum oxotransferases and related enzymes. *Dalton Transactions* 42(9) 3029-3042.

119. Hille, R., J. Hall and P. Basu (2014). The Mononuclear Molybdenum Enzymes. *Chemical Reviews*.

120. Hirsch, P., M. Van Montagu, A. Johnston, N. Brewin and J. Schell (1980). Physical Identification of Bacteriocinogenic, Nodulation and Other Plasmids in Strains of Rhizobium leguminosarm. *Journal of General Microbiology* 120(2) 403-412.

121. Holmes, B., M. Popoff, M. Kiredjian and K. Kersters (1988). Ochrobactrum anthropi gen. nov., sp. nov. from human clinical specimens and previously known as group Vd. *International Journal of Systematic Bacteriology* 38(4) 406-416.

122. Honma, M., M. Asomaning and F. Ausubel (1990). Rhizobium meliloti nodD genes mediate host-specific activation of nodABC. *J Bacteriol* 172(2) 901-911.

123. Howard, J. B. and D. C. Rees (2006). How many metals does it take to fix N2? A mechanistic overview of biological nitrogen fixation. *Proceedings of the National Academy of Sciences* 103(46) 17088-17093.

124. Howden, A. J., C. J. Harrison and G. M. Preston (2009a). A conserved mechanism for nitrile metabolism in bacteria and plants. *Plant J* 57(2) 243-253.

125. Howden, A. J. and G. M. Preston (2009). Nitrilase enzymes and their role in plant-microbe interactions. *Microb Biotechnol* 2(4) 441-451.

126. Howden, A. J., A. Rico, T. Mentlak, L. Miguet and G. M. Preston (2009b). Pseudomonas syringae pv. syringae B728a hydrolyses indole-3-acetonitrile to the plant hormone indole-3-acetic acid. *Mol Plant Pathol* 10(6) 857-865.

127. Howieson, J., R. Yates, G. O'Hara, M. Ryder and D. Real (2005). The interactions of Rhizobium leguminosarum biovar trifolii in nodulation of annual and perennial Trifolium spp. from diverse centres of origin. *Animal Production Science* 45(3) 199-207.

128. Hungria, M., R. J. Campo and I. C. Mendes (2003). Benefits of inoculation of the common bean (Phaseolus vulgaris) crop with efficient and competitive Rhizobium tropici strains. *Biology and Fertility of Soils* 39(2) 88-93.

129. Huson, D. H. and D. Bryant (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23(2) 254-267.

130. Iobbi-Nivol, C. and S. Leimkuhler (2013). Molybdenum enzymes, their maturation and molybdenum cofactor biosynthesis in Escherichia coli. *Biochim Biophys Acta* 1827(8-9) 1086-1101.

131. Jeanmougin, F., J. D. Thompson, M. Gouy, D. G. Higgins and T. J. Gibson (1998). Multiple sequence alignment with Clustal X. *Trends Biochem Sci* 23(10) 403-405.

132. Jeyaretnam, B., J. Glushka, V. K. Kolli and R. W. Carlson (2002). Characterization of a novel lipid-A from *Rhizobium* species Sin-1. A unique lipid-A structure that is devoid of phosphate and has a glycosyl backbone consisting of glucosamine and 2-aminogluconic acid. *Journal of Biological Chemistry* 277(44) 41802-41810.

133. Johnston, A., G. Hombrecher, N. Brewin and M. Cooper (1982). Two transmissible plasmids in Rhizobium leguminosarum strain 300. *Journal of General Microbiology* 128(1) 85-93.

134. Jordan, D. C. (1982). Notes : Transfer of Rhizobium japonicum Buchanan 1980 to Bradyrhizobium gen. nov., a Genus of Slow-Growing, Root Nodule Bacteria from Leguminous Plants. *International Journal of Systematic Bacteriology* 32(1) 136-139.

135. Kaeppel, E. C., A. Gärdes, S. Seebah, H.-P. Grossart and M. S. Ullrich (2012). Marinobacter adhaerens sp. nov., isolated from marine aggregates formed with the diatom Thalassiosira weissflogii. *Int J Syst Evol Microbiol* 62(1) 124-128.

136. Kaluza, K., M. Hahn and H. Hennecke (1985). Repeated sequences similar to insertion elements clustered around the *nif* region of the *Rhizobium japonicum* genome. *J Bacteriol* 162(2) 535-542.

137. Kappler, U. (2011). Bacterial sulfite-oxidizing enzymes. *Biochim Biophys Acta* 1807(1) 1-10.

138. Karbarz, M. J., D. A. Six and C. R. Raetz (2009). Purification and characterization of the lipid A 1-phosphatase LpxE of Rhizobium leguminosarum. *Journal of Biological Chemistry* 284(1) 414-425.

139. Karunakaran, R., A. F. Haag, A. K. East, V. K. Ramachandran, J. Prell, E. K. James, M. Scocchi, G. P. Ferguson and P. S. Poole (2010). BacA is essential for bacteroid development in nodules of galegoid, but not phaseoloid, legumes. *J Bacteriol* 192(11) 2920-2928.

140. Karunakaran, R., V. K. Ramachandran, J. C. Seaman, A. K. East, B. Mouhsine, T. H. Mauchline, J. Prell, A. Skeffington and P. S. Poole (2009). Transcriptomic analysis of Rhizobium leguminosarum biovar viciae in symbiosis with host plants Pisum sativum and Vicia cracca. *J Bacteriol* 191(12) 4002-4014.

141. Kficfik, Q. and C. Cevheri (2013). Nodulation Study of Natural Forage Legume in Semiarid Region, Turkey.

142. Kim, N. (2012). Investigation into plasmid replication and partitioning systems in proteobacteria.

143. Kirn, J. and D. C. Rees (1992). Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from Azotobacter vinelandii. *Nature* 360(6404) 553-560.

144. Kisker, C., H. Schindelin and D. C. Rees (1997). Molybdenum-cofactor-containing enzymes: structure and mechanism. *Annu Rev Biochem* 66 233-267.

145. Kiziak, C., D. Conradt, A. Stolz, R. Mattes and J. Klein (2005). Nitrilase from Pseudomonas fluorescens EBC191: cloning and heterologous expression of the gene and biochemical characterization of the recombinant enzyme. *Microbiology* 151(Pt 11) 3639-3648.

146. Knapp, G. S. and J. C. Hu (2010). Specificity of the E. coli LysR-type transcriptional regulators. *PLoS One* 5(12) e15189.

147. Kohler, P. R., J. Y. Zheng, E. Schoffers and S. Rossbach (2010). Inositol catabolism, a key pathway in Sinorhizobium meliloti for competitive host nodulation. *Appl Environ Microbiol* 76(24) 7972-7980.

148. Koonin, E. V. (2009). Evolution of genome architecture. *The international journal of biochemistry & cell biology* 41(2) 298-306.

149. Krieg, N. (2001). Identification of Procaryotes. *In* D. Boone, R. Castenholz and G. Garrity eds. *Bergey's Manual® of Systematic Bacteriology*. Springer New York. 33-38.

150. Krysciak, D., C. Schmeisser, S. Preuß, J. Riethausen, M. Quitschau, S. Grond and W. Streit (2011). Involvement of multiple loci in quorum quenching of autoinducer I molecules in the nitrogen-fixing symbiont Rhizobium (Sinorhizobium) sp. strain NGR234. *Appl Environ Microbiol* 77(15) 5089-5099.

151. Kumar, S., M. Nei, J. Dudley and K. Tamura (2008). MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 9(4) 299-306.

152. Kuykendall, L., J. Young, E. Martínez-Romero, A. Kerr and H. Sawada (2005). *Rhizobium* Frank 1889, 338 AL. *Bergey's Manual® of Systematic Bacteriology* 325-340.

153. Kuykendall, L. D., J. M. Young, E. Martínez-Romero, A. Kerr and H. Sawada (2006). *Bergey's Manual of Systematic Bacteriology*. New York: Springer.

154. Lad, G. R. (2013). Phenomic and genomic diversity of a bacterial species in a local population. University of York.

155. Laguerre, G., M.-R. Allard, F. Revoy and N. Amarger (1994). Rapid identification of rhizobia by restriction fragment length polymorphism analysis of PCR-amplified 16S rRNA genes. *Appl Environ Microbiol* 60(1) 56-63.

156. Laguerre, G., S. M. Nour, V. Macheret, J. Sanjuan, P. Drouin and N. Amarger (2001). Classification of rhizobia based on nodC and nifH gene analysis reveals a close phylogenetic relationship among Phaseolus vulgaris symbionts. *Microbiology* 147(Pt 4) 981-993.

157. Lambrecht, J. A., J. M. Flynn and D. M. Downs (2012). Conserved YjgF protein family deaminates reactive enamine/imine intermediates of pyridoxal 5'-phosphate (PLP)-dependent enzyme reactions. *J Biol Chem* 287(5) 3454-3461.

158. Landeta, C., A. Davalos, M. A. Cevallos, O. Geiger, S. Brom and D. Romero (2011). Plasmids with a chromosome-like role in rhizobia. *J Bacteriol* 193(6) 1317-1326.

159. Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21) 2947-2948.

160. Layh, N., J. Parratt and A. Willetts (1998). Characterization and partial purification of an enantioselective arylacetonitrilase from Pseudomonas fluorescens DSM 7155. *Journal of Molecular Catalysis B: Enzymatic* 5(5–6) 467-474.

161. Lederberg, J., M. Alexander and B. R. Bloom (2000). *Encyclopedia of Microbiology*: Academic Press.

162. Legendre, P. and L. Legendre (2012). *Numerical Ecology*: Elsevier.

163. Legras, J., G. Chuzel, A. Arnaud and P. Galzy (1990a). Natural nitriles and their metabolism. *World Journal of Microbiology and Biotechnology* 6(2) 83-108.

164. Legras, J. L., G. Chuzel, A. Arnaud and P. Galzy (1990b). Natural nitriles and their metabolism. *World Journal of Microbiology and Biotechnology* 6(2) 83-108.

165. Lei, X., E. T. Wang, W. F. Chen, X. H. Sui and W. X. Chen (2008). Diverse bacteria isolated from root nodules of wild Vicia species grown in temperate region of China. *Arch Microbiol* 190(6) 657-671.

166. Lepek, V. C., A. L. D'Antuono, P. E. Tomatis, J. E. Ugalde, S. Giambiagi and R. A. Ugalde (2002). Analysis of Mesorhizobium loti glycogen operon: effect of phosphoglucomutase (pgm) and glycogen synthase (glgA) null mutants on nodulation of Lotus tenuis. *Molecular Plant-Microbe Interactions* 15(4) 368-375.

167. Lerouge, P., P. Roche, C. Faucher, F. Maillet, G. Truchet, J. C. Promé and J. Dénarié (1990). Symbiotic host-specificity of Rhizobium meliloti is determined by a sulphated and acylated glucosamine oligosaccharide signal. *Nature* 344(6268) 781-784.

168. Leyva, A., J. Palacios, J. Murillo and T. Ruiz-Argüeso (1990). Genetic organization of the hydrogen uptake (hup) cluster from *Rhizobium leguminosarum*. *J Bacteriol* 172(3) 1647-1655.

169. Lin, D. X., E. T. Wang, H. Tang, T. X. Han, Y. R. He, S. H. Guan and W. X. Chen (2008). Shinella kummerowiae sp. nov., a symbiotic bacterium isolated

from root nodules of the herbal legume Kummerowia stipulacea. *Int J Syst Evol Microbiol* 58(6) 1409-1413.

170. Lin, S.-Y., Y.-H. Hsu, Y.-C. Liu, M.-H. Hung, A. Hameed, W.-A. Lai, W.-S. Yen and C.-C. Young (2014). Rhizobium straminoryzae sp. nov., a novel species isolated from surface of rice-straw in Taiwan. *Int J Syst Evol Microbiol* ijs. 0.062117-062110.

171. Lin, X., X. JiaJie, L. QiaoLi, X. RuiMei and W. GeHong (2009). Genetic diversity in rhizobia isolated from Sphaerophysa salsula in several regions of northwestern China. *Biodiversity Science* 17(1) 69-75.

172. Liu, X., S. M. Tiquia, G. Holguin, L. Wu, S. C. Nold, A. H. Devol, K. Luo, A. V. Palumbo, J. M. Tiedje and J. Zhou (2003). Molecular diversity of denitrifying genes in continental margin sediments within the oxygen-deficient zone off the Pacific coast of Mexico. *Appl Environ Microbiol* 69(6) 3549-3560.

173. Lodwig, E. M., A. H. Hosie, A. Bourdès, K. Findlay, D. Allaway, R. Karunakaran, J. Downie and P. S. Poole (2003). Amino-acid cycling drives nitrogen fixation in the legume–Rhizobium symbiosis. *Nature* 422(6933) 722-726.

174. Long, S. R. (1989). Rhizobium-legume nodulation: life together in the underground. *Cell* 56(2) 203-214.

175. Long, S. R. (1996). Rhizobium symbiosis: nod factors in perspective. *The Plant Cell* 8(10) 1885.

176. Long, S. R. (2001). Genes and signals in the rhizobium-legume symbiosis. *Plant Physiol* 125(1) 69-72.

177. López-Guerrero, M. G., E. Ormeño-Orrillo, E. Velázquez, M. A. Rogel, J. L. Acosta, V. Gónzalez, J. Martínez and E. Martínez-Romero (2012). *Rhizobium etli* taxonomy revised with novel genomic data and analyses. *Syst Appl Microbiol* 35(6) 353-358.

178. López-Lara, I. M., C. Sohlenkamp and O. Geiger (2003). Membrane lipids in plant-associated bacteria: their biosyntheses and possible functions. *Molecular Plant-Microbe Interactions* 16(7) 567-579.

179. Lorenz, M. G. and W. Wackernagel (1994). Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol Rev* 58(3) 563.

180. Lowe, S. E., M. K. Jain and J. G. Zeikus (1993). Biology, ecology, and biotechnological applications of anaerobic bacteria adapted to environmental stresses in temperature, pH, salinity, or substrates. *Microbiol Rev* 57(2) 451.

181. Ludwig, W., J. Neumaier, N. Klugbauer, E. Brockmann, C. Roller, S. Jilg, K. Reetz, I. Schachtner, A. Ludvigsen and M. Bachleitner (1993). Phylogenetic relationships ofBacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase β-subunit genes. *Antonie Van Leeuwenhoek* 64(3-4) 285-305.

182. Maddocks, S. E. and P. C. Oyston (2008). Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology* 154(Pt 12) 3609-3623.

183. Mannetje, L. t. (1967). A re-examination of the taxonomy of the genus Rhizobium and related genera using numerical analysis. *Antonie Van Leeuwenhoek* 33 477-491.

184. Mano, Y. and K. Nemoto (2012). The pathway of auxin biosynthesis in plants. *J Exp Bot* 63(8) 2853-2872.

185. Mantelin, S., M. F. Saux, F. Zakhia, G. Bena, S. Bonneau, H. Jeder, P. de Lajudie and J. C. Cleyet-Marel (2006). Emended description of the genus Phyllobacterium and description of four novel species associated with plant roots: Phyllobacterium bourgognense sp. nov., Phyllobacterium ifriqiyense sp. nov., Phyllobacterium leguminum sp. nov. and Phyllobacterium brassicacearum sp. nov. *Int J Syst Evol Microbiol* 56(Pt 4) 827-839.

186. Manyani, H., L. Rey, J. M. Palacios, J. Imperial and T. Ruiz-Argüeso (2005). Gene products of the *hupGHIJ* operon are involved in maturation of the iron-sulfur subunit of the [NiFe] hydrogenase from *Rhizobium leguminosarum* bv. *viciae*. *J Bacteriol* 187(20) 7018-7026.

187. Marchler-Bauer, A., C. Zheng, F. Chitsaz, M. K. Derbyshire, L. Y. Geer, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, S. Lu, G. H. Marchler, J. S. Song, N. Thanki, R. A. Yamashita, D. Zhang and S. H. Bryant (2013). CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res* 41(Database issue) D348-352.

188. Marek-Kozaczuk, M., A. Leszcz, J. Wielbo, S. Wdowiak-Wróbel and A. Skorupska (2013). Rhizobium pisi sv. trifolii K3.22 harboring nod genes of the Rhizobium leguminosarum sv. trifolii cluster. *Syst Appl Microbiol* 36(4) 252-258.

189. Marshall, D. J., J. D. Bateman and J. Brockwell (1993). Validation of a serial-dilution, plant-infection test for enumerating Rhizobium leguminosarum bv. viciae and its application for counting rhizobia in acid soils. *Soil Biology and Biochemistry* 25(2) 261-268.

190. Martínez-Romero, E., L. Segovia, F. M. Mercante, A. A. Franco, P. Graham and M. A. Pardo (1991). Rhizobium tropici, a novel species nodulating Phaseolus vulgaris L. beans and Leucaena sp. trees. *International Journal of Systematic Bacteriology* 41(3) 417-426.

191. Martinez, E., D. Romero and R. Palacios (1990). The Rhizobium genome. *Critical reviews in plant sciences* 9(1) 59-93.

192. Martins, L. M. V., M. C. P. Neves and N. G. Rumjanek (1997). Growth characteristics and symbiotic efficiency of rhizobia isolated from cowpea nodules of the north-east region of Brazil. *Soil Biology and Biochemistry* 29(5–6) 1005-1010.

193. McBride, K. E., J. W. Kenny and D. M. Stalker (1986). Metabolism of the herbicide bromoxynil by Klebsiella pneumoniae subsp. ozaenae. *Appl Environ Microbiol* 52(2) 325-330.

194. Medina-Pascual, M., S. Valdezate, P. Villalón, N. Garrido, V. Rubio and J. Saéz-Nieto (2012). Identification, molecular characterisation and antimicrobial susceptibility of genomovars of the Burkholderia cepacia complex in Spain. *European journal of clinical microbiology & infectious diseases* 31(12) 3385-3396.

195. Mendel, R. R. (2009). Cell biology of molybdenum. *Biofactors* 35(5) 429-434.

196. Mendel, R. R. (2013). The molybdenum cofactor. *J Biol Chem* 288(19) 13165-13172.

197. Mendel, R. R. and F. Bittner (2006). Cell biology of molybdenum. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1763(7) 621-635.

198. Mendel, R. R. and G. Schwarz (2002). Biosynthesis and molecular biology of the molybdenum cofactor (Moco). *Met Ions Biol Syst* 39 317-368.

199. Mercado-Blanco, J. and N. Toro (1996). Plasmids in rhizobia: the role of nonsymbiotic plasmids. *MPMI-Molecular Plant Microbe Interactions* 9(7) 535-545.

200. Merrick, M. (1992). Regulation of nitrogen fixation genes in free-living and symbiotic bacteria. *Biological nitrogen fixation. Chapman and Hall, New York* 835-876.

201. Michiels, J., B. Dombrecht, N. Vermeiren, C. Xi, E. Luyten and J. Vanderleyden (1998). Phaseolus vulgaris is a non-selective host for nodulation. *FEMS Microbiol Ecol* 26(3) 193-205.

202. Mihara, H. and N. Esaki (2002). Bacterial cysteine desulfurases: their function and mechanisms. *Appl Microbiol Biotechnol* 60(1-2) 12-23.

203. Mihara, H., T. Kurihara, T. Yoshimura, K. Soda and N. Esaki (1997). Cysteine Sulfinate Desulfinase, a NIFS-like Protein ofEscherichia coli with Selenocysteine Lyase and Cysteine Desulfurase Activities: GENE CLONING, PURIFICATION, AND CHARACTERIZATION OF A NOVEL PYRIDOXAL ENZYME. *Journal of Biological Chemistry* 272(36) 22417-22424.

204. Miller, L. D., C. K. Yost, M. F. Hynes and G. Alexandre (2007). The major chemotaxis gene cluster of Rhizobium leguminosarum bv. viciae is essential for competitive nodulation. *Molecular microbiology* 63(2) 348-362.

205. Mnasri, B., T. Y. Liu, S. Saidi, W. F. Chen, W. X. Chen, X. X. Zhang and R. Mhamdi (2014). Rhizobium azibense sp. nov., a nitrogen fixing bacterium isolated from root-nodules of Phaseolus vulgaris. *Int J Syst Evol Microbiol* ijs. 0.058651-058650.

206. Moawad, H. and B. B. Bohlool (1984). Competition Among Rhizobium spp. for Nodulation of Leucaena leucocephala in Two Tropical Soils. *Appl Environ Microbiol* 48(1) 5-9.

207. Moffett, M. L. and R. R. Colwell (1968). Adansonian analysis of the Rhizobiaceae. *Journal of General Microbiology* 51(2) 245-266.

208. Morishita, R., A. Kawagoshi, T. Sawasaki, K. Madin, T. Ogasawara, T. Oka and Y. Endo (1999). Ribonuclease activity of rat liver perchloric acid-soluble protein, a potent inhibitor of protein synthesis. *J Biol Chem* 274(29) 20688-20692.

209. Moschetti, G., A. Peluso, A. Protopapa, M. Anastasio, O. Pepe and R. Defez (2005). Use of nodulation pattern, stress tolerance, nodC gene amplification, RAPD-PCR and RFLP-16S rDNA analysis to discriminate genotypes of Rhizobium leguminosarum biovar viciae. *Syst Appl Microbiol* 28(7) 619-631.

210. Moulin, L., A. Munive, B. Dreyfus and C. Boivin-Masson (2001). Nodulation of legumes by members of the β-subclass of Proteobacteria. *Nature* 411(6840) 948-950.

211. Mulas, D., P. García-Fraile, L. Carro, M.-H. Ramírez-Bahena, P. Casquero, E. Velázquez and F. González-Andrés (2011). Distribution and efficiency of *Rhizobium leguminosarum* strains nodulating *Phaseolus vulgaris* in Northern Spanish soils: Selection of native strains that replace conventional N fertilization. *Soil Biology and Biochemistry* 43(11) 2283-2293.

212. Mullis, K. B., H. A. Erlich, N. Arnheim, G. T. Horn, R. K. Saiki and S. J. Scharf (1987). One of the first Polymerase Chain Reaction (PCR) patents. Google Patents.

213. Mutch, L. A. and J. P. Young (2004). Diversity and specificity of Rhizobium leguminosarum biovar viciae on wild and cultivated legumes. *Mol Ecol* 13(8) 2435-2444.

214. Mylona, P., K. Pawlowski and T. Bisseling (1995). Symbiotic nitrogen fixation. *The Plant Cell* 7(7) 869.

215. Napoli, C. A. and D. H. Hubbell (1975). Ultrastructure of Rhizobium-induced infection threads in clover root hairs. *Appl Microbiol* 30(6) 1003-1009.

216. Nguyen, T. A., J. Brescic, D. J. Vinyard, T. Chandrasekar and G. C. Dismukes (2012). Identification of an oxygenic reaction center psbADC operon in the cyanobacterium Gloeobacter violaceus PCC 7421. *Molecular Biology and Evolution* 29(1) 35-38.

217. Nitrogen Fixation for Tropical Agricultural Legumes Project, U., Food, A. O. o. t. U. N. Fertilizer, P. N. Service, Food, A. O. o. t. U. N. Crop and G. P. Service (1984). *Legume Inoculants and Their Use: A Pocket Manual*: Food and Agriculture Organization of the United Nations.

218. Nogales, J., R. Campos, H. BenAbdelkhalek, J. Olivares, C. Lluch and J. Sanjuan (2002). Rhizobium tropici genes involved in free-living salt tolerance are required for the establishment of efficient nitrogen-fixing symbiosis with Phaseolus vulgaris. *Mol Plant Microbe Interact* 15(3) 225-232.

219. Noisangiam, R., K. Teamtisong, P. Tittabutr, N. Boonkerd, U. Toshiki, K. Minamisawa and N. Teaumroong (2012). Genetic diversity, symbiotic evolution, and proposed infection process of Bradyrhizobium strains isolated from root nodules of Aeschynomene americana L. in Thailand. *Appl Environ Microbiol* 78(17) 6236-6250.

220. Nuti, M., A. Lepidi, R. Prakash, R. Schilperoort and F. Cannon (1979). Evidence for nitrogen fixation (nif) genes on indigenous Rhizobium plasmids.

221. O'Reilly, C. and P. D. Turner (2003). The nitrilase family of CN hydrolysing enzymes - a comparative study. *J Appl Microbiol* 95(6) 1161-1174.

222. Oka, T., H. Tsuji, C. Noda, K. Sakai, Y. M. Hong, I. Suzuki, S. Munoz and Y. Natori (1995). Isolation and characterization of a novel perchloric acid-soluble protein inhibiting cell-free protein synthesis. *J Biol Chem* 270(50) 30060-30067.

223. Oldroyd, G. E., J. D. Murray, P. S. Poole and J. A. Downie (2011). The rules of engagement in the legume-rhizobial symbiosis. *Annual review of genetics* 45 119-144.

224. Olsen, K. M., S.-C. Hsu and L. L. Small (2008). Evidence on the molecular basis of the Ac/ac adaptive cyanogenesis polymorphism in white clover (Trifolium repens L.). *Genetics* 179(1) 517-526.

225. Olsen, K. M., N. J. Kooyers and L. L. Small (2014). Adaptive gains through repeated gene loss: parallel evolution of cyanogenesis polymorphisms in the genus Trifolium (Fabaceae).

226. Op den Camp, R. H., E. Polone, E. Fedorova, W. Roelofsen, A. Squartini, H. J. Op den Camp, T. Bisseling and R. Geurts (2012). Nonlegume Parasponia andersonii deploys a broad rhizobium host range strategy resulting in largely

variable symbiotic effectiveness. *Molecular Plant-Microbe Interactions* 25(7) 954-963.

227. Ormeno-Orrillo, E., P. Menna, L. G. Almeida, F. J. Ollero, M. F. Nicolas, E. Pains Rodrigues, A. Shigueyoshi Nakatani, J. S. Silva Batista, L. M. Oliveira Chueire, R. C. Souza, A. T. Ribeiro Vasconcelos, M. Megias, M. Hungria and E. Martinez-Romero (2012). Genomic basis of broad host range and environmental adaptability of Rhizobium tropici CIAT 899 and Rhizobium sp. PRF 81 which are used in inoculants for common bean (Phaseolus vulgaris L.). *BMC Genomics* 13 735.

228. Pace, H. C. and C. Brenner (2001). The nitrilase superfamily: classification, structure and function. *Genome Biol* 2(1) REVIEWS0001.

229. Panday, D., P. Schumann and S. K. Das (2011). Rhizobium pusense sp. nov., isolated from the rhizosphere of chickpea (Cicer arietinum L.). *Int J Syst Evol Microbiol* 61(11) 2632-2639.

230. Peck, M. C., R. F. Fisher and S. R. Long (2006). Diverse flavonoids stimulate NodD1 binding to nod gene promoters in Sinorhizobium meliloti. *J Bacteriol* 188(15) 5417-5427.

231. Pérez-Giménez, J., J. I. Quelas and A. R. Lodeiro (2011). Competition for nodulation. *Soyean, Physiology and Biochemistry, Rijeka: InTech* 139-166.

232. Pérez-Segura, G., Á. Pérez-Oseguera and M. A. Cevallos (2013). The *repAC* replication system of the Rhizobium leguminosarum pRL7 plasmid is functional: Implications regarding the origin and evolution of *repABC* plasmids. *Plasmid* 69(1) 49-57.

233. Perret, X. and W. J. Broughton (1998). Rapid identification of Rhizobium strains by targeted PCR fingerprinting. *Plant and Soil* 204(1) 21-34.

234. Philippe, H. and C. J. Douady (2003). Horizontal gene transfer and phylogenetics. *Curr Opin Microbiol* 6(5) 498-505.

235. Piotrowski, M. (2008). Primary or secondary? Versatile nitrilases in plant metabolism. *Phytochemistry* 69(15) 2655-2667.

236. Piotrowski, M., S. Schonfelder and E. W. Weiler (2001). The Arabidopsis thaliana isogene NIT4 and its orthologs in tobacco encode beta-cyano-L-alanine hydratase/nitrilase. *J Biol Chem* 276(4) 2616-2621.

237. Podar, M., J. Eads and T. Richardson (2005). Evolution of a microbial nitrilase gene family: a comparative and environmental genomics study. *BMC Evolutionary Biology* 5(1) 1-13.

238. Prell, J., A. Bourdes, R. Karunakaran, M. Lopez-Gomez and P. Poole (2009). Pathway of gamma-aminobutyrate metabolism in Rhizobium leguminosarum 3841 and its role in symbiosis. *J Bacteriol* 191(7) 2177-2186.

239. Pulawska, J., A. Willems and P. Sobiczewski (2012). Rhizobium skierniewicense sp. nov., isolated from tumours on chrysanthemum and cherry plum. *Int J Syst Evol Microbiol* 62(Pt 4) 895-899.

240. Putnoky, P., E. Grosskopf, D. Ha, G. B. Kiss and A. Kondorosi (1988). Rhizobium fix genes mediate at least two communication steps in symbiotic nodule development. *The Journal of cell biology* 106(3) 597-607.

241. Qin, W., Z. S. Deng, L. Xu, N. N. Wang and G. H. Wei (2012). Rhizobium helanshanense sp. nov., a bacterium that nodulates Sphaerophysa salsula (Pall.) DC. in China. *Arch Microbiol* 194(5) 371-378.

242. Qiu, J. A., H. L. Wilson and K. V. Rajagopalan (2012). Structure-based alteration of substrate specificity and catalytic activity of sulfite oxidase from sulfite oxidation to nitrate reduction. *Biochemistry* 51(6) 1134-1147.

243. Rae, A. L., P. Bonfante-Fasolo and N. J. Brewin (1992). Structure and growth of infection threads in the legume symbiosis with Rhizobium leguminosarum. *The Plant Journal* 2(3) 385-395.

244. Rahi, P., R. Kapoor, J. P. Young and A. Gulati (2012). A genetic discontinuity in root-nodulating bacteria of cultivated pea in the Indian trans-Himalayas. *Mol Ecol* 21(1) 145-159.

245. Rahmani, H. A., L. A. Räsänen, M. Afshari and K. Lindström (2011). Genetic diversity and symbiotic effectiveness of rhizobia isolated from root nodules of Phaseolus vulgaris L. grown in soils of Iran. *Applied Soil Ecology* 48(3) 287-293.

246. Ramachandran, V. K., A. K. East, R. Karunakaran, J. A. Downie and P. S. Poole (2011). Adaptation of Rhizobium leguminosarum to pea, alfalfa and sugar beet rhizospheres investigated by comparative transcriptomics. *Genome Biol* 12(10) R106.

247. Ramana, C. V., B. Parag, K. R. Girija, B. R. Ram, V. V. Ramana and C. Sasikala (2013). Rhizobium subbaraonis sp. nov., an endolithic bacterium isolated from beach sand. *Int J Syst Evol Microbiol* 63(Pt 2) 581-585.

248. Ramirez-Bahena, M. H., P. Garcia-Fraile, A. Peix, A. Valverde, R. Rivas, J. M. Igual, P. F. Mateos, E. Martinez-Molina and E. Velazquez (2008). Revision of the taxonomic status of the species Rhizobium leguminosarum (Frank 1879) Frank 1889AL, Rhizobium phaseoli Dangeard 1926AL and Rhizobium trifolii Dangeard 1926AL. R. trifolii is a later synonym of R. leguminosarum. Reclassification of the strain R. leguminosarum DSM 30132 (=NCIMB 11478) as Rhizobium pisi sp. nov. *Int J Syst Evol Microbiol* 58(Pt 11) 2484-2490.

249. Ramírez-Romero, M. A., P. Bustos, L. Girard, O. Rodríguez, M. A. Cevallos and G. Dávila (1997). Sequence, localization and characteristics of the replicator region of the symbiotic plasmid of Rhizobium etli. *Microbiology* 143(8) 2825-2831.

250. Rangin, C., B. Brunel, J.-C. Cleyet-Marel, M.-M. Perrineau and G. Béna (2008). Effects of Medicago truncatula genetic diversity, rhizobial competition, and strain effectiveness on the diversity of a natural Sinorhizobium species community. *Appl Environ Microbiol* 74(18) 5653-5661.

251. Rappé, M. S. and S. J. Giovannoni (2003). The uncultured microbial majority. *Annual Reviews in Microbiology* 57(1) 369-394.

252. Rappu, P., B. S. Shin, H. Zalkin and P. Mäntsälä (1999). A role for a highly conserved protein of unknown function in regulation of Bacillus subtilis purA by the purine repressor. *J Bacteriol* 181(12) 3810-3815.

253. Rashid, M. H., H. Schafer, J. Gonzalez and M. Wink (2012). Genetic diversity of rhizobia nodulating lentil (Lens culinaris) in Bangladesh. *Syst Appl Microbiol* 35(2) 98-109.

254. Rasul, A., E. Amalraj, G. Praveen Kumar, M. Grover and B. Venkateswarlu (2012). Characterization of rhizobial isolates nodulating Millettia pinnata in India. *FEMS Microbiology Letters* 336(2) 148-158.

255. Reeve, W. (2013). Genomics Encyclopedia of Bacteria and Archaea-Root Nodule Bacteria (GEBA-RNB): a resource for microsymbiont genomes (2013 DOE JGI Genomics of Energy and Environment 8th Annual User Meeting). LBNL (Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA (United States)).

256. Reeve, W., P. Chain, G. O'Hara, J. Ardley, K. Nandesena, L. Bräu, R. Tiwari, S. Malfatti, H. Kiss and A. Lapidus (2010a). Complete genome sequence of the Medicago microsymbiont Ensifer (Sinorhizobium) medicae strain WSM419. *Stand Genomic Sci* 2(1) 77.

257. Reeve, W., K. Nandasena, R. Yates, R. Tiwari, G. O'Hara, M. Ninawi, W. Gu, L. Goodwin, C. Detter and R. Tapia (2013). Complete genome sequence of Mesorhizobium australicum type strain (WSM2073 T). *Stand Genomic Sci* 9(2).

258. Reeve, W., G. O'Hara, P. Chain, J. Ardley, L. Brau, K. Nandesena, R. Tiwari, A. Copeland, M. Nolan, C. Han, T. Brettin, M. Land, G. Ovchinikova, N. Ivanova, K. Mavromatis, V. Markowitz, N. Kyrpides, V. Melino, M. Denton, R. Yates and J. Howieson (2010b). Complete genome sequence of Rhizobium leguminosarum bv. trifolii strain WSM1325, an effective microsymbiont of annual Mediterranean clovers. *Stand Genomic Sci* 2(3) 347-356.

259. Reeve, W., G. O'Hara, P. Chain, J. Ardley, L. Brau, K. Nandesena, R. Tiwari, S. Malfatti, H. Kiss, A. Lapidus, A. Copeland, M. Nolan, M. Land, N. Ivanova, K. Mavromatis, V. Markowitz, N. Kyrpides, V. Melino, M. Denton, R. Yates and J. Howieson (2010c). Complete genome sequence of Rhizobium leguminosarum bv trifolii strain WSM2304, an effective microsymbiont of the South American clover Trifolium polymorphum. *Stand Genomic Sci* 2(1) 66-76.

260. Reeve, W., G. O'Hara, P. Chain, J. Ardley, L. Bräu, K. Nandesena, R. Tiwari, S. Malfatti, H. Kiss and A. Lapidus (2010d). Complete genome sequence of Rhizobium leguminosarum bv trifolii strain WSM2304, an effective microsymbiont of the South American clover Trifolium polymorphum. *Stand Genomic Sci* 2(1) 66.

261. Reguera, M., M. Wimmer, P. Bustos, H. E. Goldbach, L. Bolaños and I. Bonilla (2010). Ligands of boron in Pisum sativum nodules are involved in regulation of oxygen concentration and rhizobial infection. *Plant, cell & environment* 33(6) 1039-1048.

262. Renalier, M., J. Batut, J. Ghai, B. Terzaghi, M. Gherardi, M. David, A. Garnerone, J. Vasse, G. Truchet and T. Huguet (1987). A new symbiotic cluster on the pSym megaplasmid of Rhizobium meliloti 2011 carries a functional fix gene repeat and a nod locus. *J Bacteriol* 169(5) 2231-2238.

263. Ribbe, M. W., Y. Hu, K. O. Hodgson and B. Hedman (2013). Biosynthesis of Nitrogenase Metalloclusters. *Chemical Reviews* 114(8) 4063-4080.

264. Ribeiro, R. A., M. A. Rogel, A. Lopez-Lopez, E. Ormeno-Orrillo, F. G. Barcellos, J. Martinez, F. L. Thompson, E. Martinez-Romero and M. Hungria (2012). Reclassification of Rhizobium tropici type A strains as Rhizobium leucaenae sp. nov. *Int J Syst Evol Microbiol* 62(Pt 5) 1179-1184.

265. Riccillo, P. M., C. I. Muglia, F. J. de Bruijn, A. J. Roe, I. R. Booth and O. M. Aguilar (2000). Glutathione is involved in environmental stress responses in Rhizobium tropici, including acid tolerance. *J Bacteriol* 182(6) 1748-1753.

266. Rivas, R., A. Willems, N. S. Subba-Rao, P. F. Mateos, F. B. Dazzo, R. M. Kroppenstedt, E. Martínez-Molina, M. Gillis and E. Velázquez (2003). Description of *Devosia neptuniae* sp. nov. that Nodulates and Fixes Nitrogen in Symbiosis with *Neptunia natans*, an Aquatic Legume from India. *Syst Appl Microbiol* 26(1) 47-53.

267. Robinson, W. G. and R. H. Hook (1964). Ricinine nitrilase I. Reaction product and substrate specificity. *J Biol Chem* 239 4257-4262.

268. Robledo, M., E. Velazquez, M. H. Ramirez-Bahena, P. Garcia-Fraile, A. Perez-Alonso, R. Rivas, E. Martinez-Molina and P. F. Mateos (2011). The celC gene, a new phylogenetic marker useful for taxonomic studies in Rhizobium. *Syst Appl Microbiol* 34(6) 393-399.

269. Roche, P., F. Maillet, C. Plazanet, F. Debellé, M. Ferro, G. Truchet, J.-C. Promé and J. Dénarié (1996). The common nodABC genes of Rhizobium meliloti are host-range determinants. *Proceedings of the National Academy of Sciences* 93(26) 15305-15310.

270. Rogel, M. A., E. Ormeno-Orrillo and E. Martinez Romero (2011). Symbiovars in rhizobia reflect bacterial adaptation to legumes. *Syst Appl Microbiol* 34(2) 96-104.

271. Romero, D., S. Brom, J. Martínez-Salazar, M. d. L. Girard, R. Palacios and G. Dávila (1991). Amplification and deletion of a nod-nif region in the symbiotic plasmid of Rhizobium phaseoli. *J Bacteriol* 173(8) 2435-2441.

272. Rubio, L. M. and P. W. Ludden (2008). Biosynthesis of the iron-molybdenum cofactor of nitrogenase. *Annu. Rev. Microbiol.* 62 93-111.

273. Rumjanek, N. G., R. Dobert, P. Van Berkum and E. Triplett (1993). Common soybean inoculant strains in Brazil are members of Bradyrhizobium elkanii. *Appl Environ Microbiol* 59(12) 4371-4373.

274. Rütting, T., P. Boeckx, C. Müller and L. Klemedtsson (2011). Assessment of the importance of dissimilatory nitrate reduction to ammonium for the terrestrial nitrogen cycle. *Biogeosciences* 8(7) 1779-1791.

275. Ruvkun, G. B. and F. M. Ausubel (1980). Interspecies homology of nitrogenase genes. *Proceedings of the National Academy of Sciences* 77(1) 191-195.

276. Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis and H. A. Erlich (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239(4839) 487-491.

277. Sambrook, J. (1989). *Molecular Cloning: A Laboratory Manual*: Cold Spring Harbor Laboratory Press.

278. Santi, C., D. Bogusz and C. Franche (2013). Biological nitrogen fixation in non-legume plants. *Ann Bot* 111(5) 743-767.

279. Santillana, N., M. H. Ramirez-Bahena, P. Garcia-Fraile, E. Velazquez and D. Zuniga (2008). Phylogenetic diversity based on rrs, atpD, recA genes and 16S-23S intergenic sequence analyses of rhizobial strains isolated from Vicia faba and Pisum sativum in Peru. *Arch Microbiol* 189(3) 239-247.

280. Schaechter, M. (2009). *Encyclopedia of microbiology*: Academic Press.

281. Schafer, A., A. Tauch, W. Jager, J. Kalinowski, G. Thierbach and A. Puhler (1994). Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum. *Gene* 145(1) 69-73.

282. Schlaman, H., R. Okker and B. Lugtenberg (1992). Regulation of nodulation gene expression by NodD in rhizobia. *J Bacteriol* 174(16) 5177.

283. Schlaman, H. R., M. M. Olsthoorn, M. Harteveld, L. Dörner, M. A. Djordjevic, J. E. Thomas-Oates and H. P. Spaink (2006). The production of species-specific highly unsaturated fatty acyl-containing LCOs from Rhizobium leguminosarum bv. trifolii is stringently regulated by nodD and involves the nodRL genes. *Molecular Plant-Microbe Interactions* 19(3) 215-226.

284. Schmiedeknecht, G., C. Kerkhoff, E. Orsó, J. Stöhr, C. Aslanidis, G. M. Nagy, R. Knuechel and G. Schmitz (1996). Isolation and Characterization of a 14.5-kDa Trichloroacetic-Acid-Soluble Translational Inhibitor Protein from Human Monocytes that is Upregulated Upon Cellular Differentiation. *European Journal of Biochemistry* 242(2) 339-351.

285. Schmitz, G. and D. M. Downs (2004). Reduced transaminase B (IlvE) activity caused by the lack of yjgF is dependent on the status of threonine deaminase (IlvA) in Salmonella enterica serovar Typhimurium. *J Bacteriol* 186(3) 803-810.

286. Schwarz, G. (2005). Molybdenum cofactor biosynthesis and deficiency. *Cell Mol Life Sci* 62(23) 2792-2810.

287. Schwarz, G. and R. R. Mendel (2006). Molybdenum cofactor biosynthesis and molybdenum enzymes. *Annu Rev Plant Biol* 57 623-647.

288. Segovia, L., D. Pinero, R. Palacios and E. Martinez-Romero (1991). Genetic structure of a soil population of nonsymbiotic Rhizobium leguminosarum. *Appl Environ Microbiol* 57(2) 426-433.

289. Segovia, L., J. P. W. Young and E. Martínez-Romero (1993). Reclassification of American Rhizobium leguminosarum biovar phaseoli type I strains as Rhizobium etli sp. nov. *International Journal of Systematic Bacteriology* 43(2) 374-377.

290. Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour and T. S. Whittam (1986). Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl Environ Microbiol* 51(5) 873.

291. Shah, V. K. and W. J. Brill (1977). Isolation of an iron-molybdenum cofactor from nitrogenase. *Proceedings of the National Academy of Sciences* 74(8) 3249-3253.

292. Silva, F. V., S. E. De Meyer, J. L. S. de Araujo, T. da Costa Barbé, G. R. Xavier, G. O'Hara, J. Ardley, N. G. Rumjanek, A. Willems and J. É. Zilli (2014). Bradyrhizobium manausense sp. nov., isolated from effective nodules of Vigna unguiculata grown in Brazilian Amazon rainforest soils. *Int J Syst Evol Microbiol* ijs. 0.061259-061250.

293. Singh, R., R. Sharma, N. Tewari and D. S. Rawat (2006). Nitrilase and its application as a 'green' catalyst. *Chem Biodivers* 3(12) 1279-1287.

294. Sokal, R. R., C. D. Michener and U. o. Kansas (1958). *A Statistical Method for Evaluating Systematic Relationships*: University of Kansas.

295. Somasegaran, P. and H. J. Hoben (1994). *Handbook for Rhizobia: Methods in Legume-rhizobium Technology*: U.S. Government Printing Office.

296. Souza, V., T. T. Nguyen, R. R. Hudson, D. Pinero and R. E. Lenski (1992). Hierarchical analysis of linkage disequilibrium in *Rhizobium* populations: evidence for sex? *Proceedings of the National Academy of Sciences* 89(17) 8389-8393.

297. Spaepen, S. and J. Vanderleyden (2011). Auxin and plant-microbe interactions. *Cold Spring Harbor perspectives in biology* 3(4) a001438.

298. Spaepen, S., J. Vanderleyden and R. Remans (2007). Indole-3-acetic acid in microbial and microorganism-plant signaling. *FEMS Microbiol Rev* 31(4) 425-448.

299. Spaink, H. P., J. Weinman, M. A. Djordjevic, C. A. Wijffelman, R. J. Okker and B. J. Lugtenberg (1989). Genetic analysis and cellular localization of the Rhizobium host specificity-determining NodE protein. *Embo j* 8(10) 2811-2818.

300. Sprent, J. I. and P. Sprent (1990). Nitrogen fixing organisms: pure and applied aspects. *Nitrogen fixing organisms: pure and applied aspects.*

301. Stam, H., A. H. Stouthamer and H. W. van Verseveld (1985). Cyanide assimilation in Rhizobium ORS 571: influence of the nitrogenase catalyzed hydrogen production on the efficiency of growth. *Arch Microbiol* 143(2) 196-202.

302. Stasiak, G., A. Mazur, J. Wielbo, M. Marczak, K. Żebracki, P. Koper and A. Skorupska (2014). Functional relationships between plasmids and their significance for metabolism and symbiotic performance of Rhizobium leguminosarum bv. trifolii. *Journal of applied genetics* 1-13.

303. Sugiyama, A. and K. Yazaki (2012). Root exudates of legume plants and their involvement in interactions with soil microbes. *Secretions and exudates in biological systems*. Springer. 27-48.

304. Sullivan, J. T. and C. W. Ronson (1998). Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proceedings of the National Academy of Sciences* 95(9) 5145-5149.

305. Svenning, M., J. Gudmundsson, I.-L. Fagerli and P. Leinonen (2001). Competition for nodule occupancy between introduced strains of Rhizobium leguminosarum biovar trifolii and its influence on plant production. *Ann Bot* 88(suppl 1) 781-787.

306. Sy, A., E. Giraud, P. Jourand, N. Garcia, A. Willems, P. de Lajudie, Y. Prin, M. Neyra, M. Gillis and C. Boivin-Masson (2001). Methylotrophic Methylobacteriumbacteria nodulate and fix nitrogen in symbiosis with legumes. *J Bacteriol* 183(1) 214-220.

307. Tabatabai, M. (1986). Roles and requirements of sulfur in plant nutrition.

308. Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* 30(12) 2725-2729.

309. Tavares, P., A. S. Pereira, J. J. Moura and I. Moura (2006). Metalloenzymes of the denitrification pathway. *J Inorg Biochem* 100(12) 2087-2100.

310. Thimann, K. V. and S. Mahadevan (1958). Enzymatic Hydrolysis of Indoleacetonitrile. *Nature* 181(4621) 1466-1467.

311. Thorneley, R. and G. A. Ashby (1989). Oxidation of nitrogenase iron protein by dioxygen without inactivation could contribute to high respiration rates of Azotobacter species and facilitate nitrogen fixation in other aerobic environments. *Biochem. J* 261 181-187.

312. Tian, C. F., E. T. Wang, T. X. Han, X. H. Sui and W. X. Chen (2007). Genetic diversity of rhizobia associated with Vicia faba in three ecological regions of China. *Arch Microbiol* 188(3) 273-282.

313. Tian, C. F., E. T. Wang, L. J. Wu, T. X. Han, W. F. Chen, C. T. Gu, J. G. Gu and W. X. Chen (2008). Rhizobium fabae sp. nov., a bacterium that nodulates Vicia faba. *Int J Syst Evol Microbiol* 58(Pt 12) 2871-2875.

314. Tian, C. F., J. P. Young, E. T. Wang, S. M. Tamimi and W. X. Chen (2010). Population mixing of Rhizobium leguminosarum bv. viciae nodulating Vicia faba: the role of recombination and lateral gene transfer. *FEMS Microbiol Ecol* 73(3) 563-576.

315. Togay, N., Y. Togay, K. M. Cimrin and M. Turan (2008). Effects of *Rhizobium* inoculation, sulfur and phosphorus applications on yield, yield components and nutrient uptakes in chickpea (*Cicer arietinum* L.). *African Journal of Biotechnology* 7(6).

316. Trinick, M. J. (1979). Structure of nitrogen-fixing nodules formed by Rhizobium on roots of Parasponia andersonii Planch. *Can J Microbiol* 25(5) 565-578.

317. Triplett, E. W. (1990). Construction of a symbiotically effective strain of Rhizobium leguminosarum bv. trifolii with increased nodulation competitiveness. *Appl Environ Microbiol* 56(1) 98-103.

318. Trujillo, M. E., A. Willems, A. Abril, A.-M. Planchuelo, R. Rivas, D. Ludena, P. F. Mateos, E. Martínez-Molina and E. Velázquez (2005). Nodulation of Lupinus albus by Strains of Ochrobactrum lupini sp. nov. *Appl Environ Microbiol* 71(3) 1318-1327.

319. Tsubouchi, T., Y. Shimane, K. Usui, S. Shimamura, K. Mori, T. Hiraki, A. Tame, K. Uematsu, T. Maruyama and Y. Hatada (2013). Brevundimonas abyssalis sp. nov., a dimorphic prosthecate bacterium isolated from deep-subsea floor sediment. *Int J Syst Evol Microbiol* 63(Pt 6) 1987-1994.

320. Turner, S. L. and J. P. W. Young (2000). The glutamine synthetases of rhizobia: phylogenetics and evolutionary implications. *Molecular Biology and Evolution* 17(2) 309-319.

321. Udvardi, M. and P. S. Poole (2013). Transport and metabolism in legume-rhizobia symbioses. *Annu Rev Plant Biol* 64 781-805.

322. van Berkum, P., D. Beyene, F. T. Vera and H. H. Keyser (1995). Variability among Rhizobium Strains Originating from Nodules of Vicia faba. *Appl Environ Microbiol* 61(7) 2649-2653.

323. Van Berkum, P., P. Elia and B. D. Eardly (2010). Application of multilocus sequence typing to study the genetic structure of megaplasmids in Medicago-nodulating rhizobia. *Appl Environ Microbiol* 76(12) 3967-3977.

324. van Rhijn, P. and J. Vanderleyden (1995). The Rhizobium-plant symbiosis. *Microbiol Rev* 59(1) 124-142.

325. Vandamme, P., J. Goris, W.-M. Chen, P. De Vos and A. Willems (2002). *Burkholderia tuberum* sp. nov. and *Burkholderia phymatum* sp. nov., Nodulate the Roots of Tropical Legumes. *Syst Appl Microbiol* 25(4) 507-512.

326. Vanderlinde, E. M., M. F. Hynes and C. K. Yost (2013). Homoserine catabolism by Rhizobium leguminosarum bv. viciae 3841 requires a plasmid-borne gene cluster that also affects competitiveness for nodulation. *Environmental Microbiology* n/a-n/a.

327. Vanderlinde, E. M., M. F. Hynes and C. K. Yost (2014a). Homoserine catabolism by Rhizobium leguminosarum bv. viciae 3841 requires a plasmid-borne gene cluster that also affects competitiveness for nodulation. *Environ Microbiol* 16(1) 205-217.

328. Vanderlinde, E. M., M. F. Hynes and C. K. Yost (2014b). Homoserine catabolism by Rhizobium leguminosarum bv. viciae 3841 requires a plasmid-borne gene cluster that also affects competitiveness for nodulation. *Environ Microbiol* 16(1) 205-217.

329. Vanderlinde, E. M., S. A. Magnus, D. D. Tambalo, S. F. Koval and C. K. Yost (2011). Mutation of a broadly conserved operon (RL3499-RL3502) from Rhizobium leguminosarum Biovar viciae causes defects in cell morphology and envelope integrity. *J Bacteriol* 193(11) 2684-2694.

330. Vanderlinde, E. M., A. Muszynski, J. J. Harrison, S. F. Koval, D. L. Foreman, H. Ceri, E. L. Kannenberg, R. W. Carlson and C. K. Yost (2009a). Rhizobium leguminosarum biovar viciae 3841, deficient in 27-hydroxyoctacosanoate-modified lipopolysaccharide, is impaired in desiccation tolerance, biofilm formation and motility. *Microbiology* 155(Pt 9) 3055-3069.

331. Vanderlinde, E. M., A. Muszyński, J. J. Harrison, S. F. Koval, D. L. Foreman, H. Ceri, E. L. Kannenberg, R. W. Carlson and C. K. Yost (2009b). Rhizobium leguminosarum biovar viciae 3841, deficient in 27-hydroxyoctacosanoate-modified lipopolysaccharide, is impaired in desiccation tolerance, biofilm formation and motility. *Microbiology* 155(9) 3055-3069.

332. Veselá, A., M. Franc, H. Pelantová, D. Kubáč, V. Vejvoda, M. Šulc, T. Bhalla, M. Macková, P. Lovecká and P. Janů (2010). Hydrolysis of benzonitrile herbicides by soil actinobacteria and metabolite toxicity. *Biodegradation* 21(5) 761-770.

333. Vincent, J. M. (1970). *A Manual for the Practical Study of Root-nodule Bacteria*: [Published for the] International Biological Programme [by] Blackwell Scientific.

334. Vinuesa, P., J. L. Rademaker, F. J. de Bruijn and D. Werner (1998). Genotypic Characterization ofBradyrhizobium Strains Nodulating Endemic Woody Legumes of the Canary Islands by PCR-Restriction Fragment Length Polymorphism Analysis of Genes Encoding 16S rRNA (16S rDNA) and 16S-23S rDNA Intergenic Spacers, Repetitive Extragenic Palindromic PCR Genomic Fingerprinting, and Partial 16S rDNA Sequencing. *Appl Environ Microbiol* 64(6) 2096-2104.

335. Wagner, S. (2012). Biological nitrogen fixation. *Nature Education Knowledge* 3(10) 15.

336. Walker, J. E., M. Saraste and N. J. Gay (1984). The UNC operon nucleotide sequence, regulation and structure of ATP-synthase. *Biochimica et Biophysica Acta (BBA)-Reviews on Bioenergetics* 768(2) 164-200.

337. Wang, D., S. Yang, F. Tang and H. Zhu (2012). Symbiosis specificity in the legume–rhizobial mutualism. *Cellular Microbiology* 14(3) 334-342.

338. Wang, F., E. T. Wang, L. J. Wu, X. H. Sui, Y. Li and W. X. Chen (2011). Rhizobium vallis sp. nov., isolated from nodules of three leguminous species. *Int J Syst Evol Microbiol* 61(11) 2582-2588.

339. Wang, Y., J. Xu, A. Chen, Y. Wang, J. Zhu, G. Yu, L. Xu and L. Luo (2010). GGDEF and EAL proteins play different roles in the control of Sinorhizobium meliloti growth, motility, exopolysaccharide production, and competitive nodulation on host alfalfa. *Acta biochimica et biophysica Sinica* 42(6) 410-417.

340. Weber, E., C. Berger, U. Bonas and R. Koebnik (2007). Refinement of the Xanthomonas campestris pv. vesicatoria hrpD and hrpE operon structure. *Molecular Plant-Microbe Interactions* 20(5) 559-567.

341. Wernegreen, J. and M. Riley (1999). Comparison of the evolutionary dynamics of symbiotic and housekeeping loci: a case for the genetic coherence of rhizobial lineages. *Molecular Biology and Evolution* 16(1) 98-113.

342. Whitman, W. B., D. C. Coleman and W. J. Wiebe (1998). Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences* 95(12) 6578-6583.

343. Wielbo, J., D. Kidaj, P. Koper, A. Kubik-Komar and A. Skorupska (2012). The effect of biotic and physical factors on the competitive ability of Rhizobium leguminosarum. *Central European Journal of Biology* 7(1) 13-24.

344. Wielbo, J., M. Marek-Kozaczuk, A. Mazur, A. Kubik-Komar and A. Skorupska (2010). Genetic and metabolic divergence within a Rhizobium leguminosarum bv. trifolii population recovered from clover nodules. *Appl Environ Microbiol* 76(14) 4593-4600.

345. Williams, A., A. Wilkinson, M. Krehenbrink, D. M. Russo, A. Zorreguieta and J. A. Downie (2008). Glucomannan-mediated attachment of Rhizobium leguminosarum to pea root hairs is required for competitive nodule infection. *J Bacteriol* 190(13) 4706-4715.

346. Wilson, J. (1944). Over five hundred reasons for abandoning the cross-inoculation groups of the legumes. *Soil Science* 58(1) 61-70.

347. Wise, A. A., Z. Liu and A. N. Binns (2006). Three Methods for the Introduction of Foreign DNA into Agrobacterium. 43-54.

348. Wise, M. G., L. J. Shimkets and J. V. McArthur (1995). Genetic structure of a lotic population of *Burkolderia* (*Pseudomonas*) *cepacia*. *Appl Environ Microbiol* 61(5) 1791-1798.

349. Wittstock, U. and J. Gershenzon (2002). Constitutive plant toxins and their role in defense against herbivores and pathogens. *Current opinion in plant biology* 5(4) 300-307.

350. Witty, J., F. Minchin, L. Skot and J. Sheehy (1986). Nitrogen fixation and oxygen in legume root nodules. *Oxford surveys of plant molecular and cell biology*.

351. Wolde-meskel, E., Z. Terefework, K. Lindström and Å. Frostegård (2004a). Metabolic and genomic diversity of rhizobia isolated from field standing native and exotic woody legumes in southern Ethiopia. *Syst Appl Microbiol* 27(5) 603-611.

352. Wolde-meskel, E., Z. Terefework, K. Lindström and Å. Frostegård (2004b). Rhizobia nodulating African Acacia spp. and Sesbania sesban trees in southern Ethiopian soils are metabolically and genomically diverse. *Soil Biology and Biochemistry* 36(12) 2013-2025.

353. Wong, B.-T. and D.-J. Lee (2014). Pseudomonas yangmingensis sp. nov., an alkaliphilic denitrifying species isolated from a hot spring. *Journal of bioscience and bioengineering* 117(1) 71-74.

354. Workun, G. J., K. Moquin, R. A. Rothery and J. H. Weiner (2008). Evolutionary persistence of the molybdopyranopterin-containing sulfite oxidase protein fold. *Microbiology and molecular biology reviews* 72(2) 228-248.

355. Wu, L. J., H. Q. Wang, E. T. Wang, W. X. Chen and C. F. Tian (2011). Genetic diversity of nodulating and non-nodulating rhizobia associated with wild soybean (Glycine soja Sieb. & Zucc.) in different ecoregions of China. *FEMS Microbiol Ecol* 76(3) 439-450.

356. Xu, L., J. F. Shi, P. Zhao, W. M. Chen, W. Qin, M. Tang and G. H. Wei (2011). Rhizobium sphaerophysae sp. nov., a novel species isolated from root nodules of Sphaerophysa salsula in China. *Antonie Van Leeuwenhoek* 99(4) 845-854.

357. Xu, L., Y. Zhang, Z. S. Deng, L. Zhao, X. L. Wei and G. H. Wei (2013). Rhizobium qilianshanense sp. nov., a novel species isolated from root nodule of Oxytropis ochrocephala Bunge in China. *Antonie Van Leeuwenhoek* 103(3) 559-565.

358. Yates, R. (2008). Symbiotic interactions of geographically diverse annual and perennial Trifolium spp. with Rhizobium leguminosarum bv. trifolii. Murdoch University.

359. Yates, R., J. Howieson, D. Real, W. Reeve, A. Vivas-Marfisi and G. O'Hara (2005). Evidence of selection for effective nodulation in the Trifolium spp. symbiosis with Rhizobium leguminosarum biovar trifolii. *Animal Production Science* 45(3) 189-198.

360. Yates, R., J. Howieson, W. Reeve, L. Brau, J. Speijers, K. Nandasena, D. Real, E. Sezmis and G. O'hara (2008). Host–strain mediated selection for an effective nitrogen-fixing symbiosis between *Trifolium* spp. and *Rhizobium leguminosarum* biovar *trifolii*. *Soil Biology and Biochemistry* 40(3) 822-833.

361. Ye, J., G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen and T. L. Madden (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13 134.

362. Yeom, S. J., H. J. Kim, J. K. Lee, D. E. Kim and D. K. Oh (2008). An amino acid at position 142 in nitrilase from Rhodococcus rhodochrous ATCC 33278 determines the substrate specificity for aliphatic and aromatic nitriles. *Biochem J* 415(3) 401-407.

363. Young, J. M. (2010). Sinorhizobium versus Ensifer: may a taxonomy subcommittee of the ICSP contradict the Judicial Commission? *Int J Syst Evol Microbiol* 60(7) 1711-1713.

364. Young, J. P., L. C. Crossman, A. W. Johnston, N. R. Thomson, Z. F. Ghazoui, K. H. Hull, M. Wexler, A. R. Curson, J. D. Todd, P. S. Poole, T. H. Mauchline, A. K. East, M. A. Quail, C. Churcher, C. Arrowsmith, I. Cherevach, T. Chillingworth, K. Clarke, A. Cronin, P. Davis, A. Fraser, Z. Hance, H. Hauser, K. Jagels, S. Moule, K. Mungall, H. Norbertczak, E. Rabbinowitsch, M. Sanders, M.

Simmonds, S. Whitehead and J. Parkhill (2006). The genome of Rhizobium leguminosarum has recognizable core and accessory components. *Genome Biol* 7(4) R34.

365. Young, J. P. W. (1996). Phylogeny and taxonomy of rhizobia. *Plant and Soil* 186(1) 45-52.

366. Young, J. P. W. (1998). *Bacterial evolution and the nature of species. In Advances in Molecular Ecology*. Amsterdam: IOS Press.

367. Yurgel, S. N., M. W. Mortimer, J. T. Rice, J. L. Humann and M. L. Kahn (2013). Directed construction and analysis of a Sinorhizobium meliloti pSymA deletion mutant library. *Appl Environ Microbiol* 79(6) 2081-2087.

368. Zézé, A., L. A. Mutch and J. P. Young (2001). Direct amplification of *nodD* from community DNA reveals the genetic diversity of Rhizobium leguminosarum in soil. *Environ Microbiol* 3(6) 363-370.

369. Zhang, G. X., S. Z. Ren, M. Y. Xu, G. Q. Zeng, H. D. Luo, J. L. Chen, Z. Y. Tan and G. P. Sun (2011). Rhizobium borbori sp. nov., aniline-degrading bacteria isolated from activated sludge. *Int J Syst Evol Microbiol* 61(Pt 4) 816-822.

370. Zhang, X.-X., X. Tang, R. A. Sheirdil, L. Sun and X.-T. Ma (2014). Rhizobium rhizoryzae sp. nov., isolated from rice roots. *Int J Syst Evol Microbiol* ijs. 0.056325-056320.

371. Zhang, X., B. Li, H. Wang, X. Sui, X. Ma, Q. Hong and R. Jiang (2012). Rhizobium petrolearium sp. nov., isolated from oil-contaminated soil. *Int J Syst Evol Microbiol* 62(Pt 8) 1871-1876.

372. Zhang, X., G. Nick, S. Kaijalainen, Z. Terefework, L. Paulin, S. W. Tighe, P. H. Graham and K. Lindström (1999). Phylogeny and Diversity of Bradyrhizobium Strains Isolated from the Root Nodules of Peanut (Arachis hypogaea) in Sichuan, China. *Syst Appl Microbiol* 22(3) 378-386.

373. Zhao, Y. (2012). Auxin biosynthesis: a simple two-step pathway converts tryptophan to indole-3-acetic acid in plants. *Molecular plant* 5(2) 334-338.

374. Zheng, L., R. H. White, V. L. Cash, R. F. Jack and D. R. Dean (1993). Cysteine desulfurase activity indicates a role for NIFS in metallocluster biosynthesis. *Proc Natl Acad Sci U S A* 90(7) 2754-2758.

375. Zheng, W. T., Y. Li, R. Wang, X. H. Sui, X. X. Zhang, J. J. Zhang, E. T. Wang and W. X. Chen (2013). Mesorhizobium qingshengii sp. nov., isolated from effective nodules of Astragalus sinicus. *Int J Syst Evol Microbiol* 63(Pt 6) 2002-2007.

376. Zhou, P. F., W. M. Chen and G. H. Wei (2010). Mesorhizobium robiniae sp. nov., isolated from root nodules of Robinia pseudoacacia. *Int J Syst Evol Microbiol* 60(11) 2552-2556.

377. Zohary, M. and D. Heller (1984). *The genus Trifolium*: Israel Academy of Sciences and Humanities.

378. Zurdo-Pineiro, J. L., R. Rivas, M. E. Trujillo, N. Vizcaíno, J. A. Carrasco, M. Chamber, A. Palomares, P. F. Mateos, E. Martínez-Molina and E. Velazquez (2007). Ochrobactrum cytisi sp. nov., isolated from nodules of Cytisus scoparius in Spain. *Int J Syst Evol Microbiol* 57(4) 784-788.