Phenomic and genomic diversity of a bacterial species in a local population

Ganesh Raoji Lad

Doctor of Philosophy

University of York

Department of Biology

September 2013

Abstract

Biology aims to understand life and life processes. DNA is the blueprint of life and hence analysis of DNA sequences is expected to explain the way in which life manifests. The genomes of a large number of organisms have been sequenced and more are being sequenced every day. However, just knowing the DNA sequence does not tell us the manner in which it expresses itself to confer phenotypes on an organism. Hence, after the DNA sequence is obtained, it is important to identify the genes in the sequence, their organisation, their complex interactions and their function in the expression of different phenotypes.

Bacteria belonging to the genus *Rhizobium* exhibit immense genotypic and phenotypic diversity. In this thesis, we look at the difference in the metabolic fingerprints of 72 isolates belonging to the genus *Rhizobium* isolated from leguminous plants near Wentworth College, University of York. Sequence data from the 72 isolates was used to correlate the phenotypic diversity to the sequence diversity in order to predict the role of genes involved in metabolism of the substrates investigated. Of the 95 substrates investigated, the ability of the strains to utilize only one substrate viz. γ-hydroxybutyrate (GHB) showed a near absolute correlation to the presence of two genes in the genome data viz. pRL100135 and pRL100136. The mutation of pRL100135 showed a pronounced decrease in the ability of the test organism to use GHB. The mutation of pRL100136 decreased, but did not abolish, the ability to use GHB. Complementation of mutation by introduction of an intact copy cloned into a plasmid under the control of a strong promoter significantly restored the ability of the mutated test bacterium to utilize GHB to significant levels.

The study shows that it is possible to predict the role of genes associated with a phenotype by looking at correlation between the variation in the phenotype in a population and the corresponding gene presence-absence data. However, this approach is both time-consuming and expensive. Of the 95 substrates investigated, only one showed a near-absolute correlation between the appearance of a phenotype and gene presence-absence data. Although an alternate approach to build mutant library and study the effect of mutation in every single gene on a larger array of substrates is possible, that too, would prove expensive. Hence, in conclusion, it can be said that the process of correlating genotype (or specific genes) to phenotypes will probably have to await development of new, cheap, quick and more efficient techniques.

List of contents

List of figures

Chapter 1

Chapter 2

Chapter 3

<u>**Chapter 4**</u>

Chapter 5

List of tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Acknowledgements

First and foremost, I would like to thank my parents, the prime driving force behind my PhD and for whom I set out on this venture. Their constant encouragement and faith in me kept me motivated through my life as a doctoral student. Next, I would like to thank my sisters and my brothers-in-law who stood by me through the thick and thin for the past four years of my life. Special thanks to my 'little' nephew Vyom.

It would have been impossible to complete my PhD without the guidance and ideas provided by Prof. Young and my TAP members Dr. James Moir and Dr. James Chong. Their advice and feedback contributed immensely to the progress of my work. I also take to opportunity to thank Dr. Gail Preston and Aziz Mithani, from the University of Oxford for their help with Rahnuma and KEGG, Dr. Jurgen Prell for his protocol to make insertion mutants and Prof. Christopher Yost and Prof. Daniel Gage for their insights into gene complementation.

Thanks are due also to all the members of the J1 group, the undergraduate students and visiting students who provided me with valuable insight into an educational system far different from that in my country. Special thanks to Ms. Julie Knox, Ms. Anne Walker from the Biology office and Dr. Adrian Harrison and Dr. Dr Betsy Pownall from the Biology Department, University of York, for their support.

I also wish to extend my gratitude to my friends in York viz. Madhuri, Jayant, Chitvan, Harsh, Nishant, Anuja, Priyanka, Joe, Alex, Amanda, Chiara, Waqar and students of the Biology Department who were like a family away from home. I apologise to those whose names have been unintentionally been omitted.

Five people need a special mention. Nantida (Sung) and Terd (Ted), who made me feel at home after I landed in York. Ms. Piyachat Udomwong, a smiling face, ever ready to discuss statistics – thanks!. A special thanks to Shakeel and Kailin – for showering me with all the love, affection, trust and encouragement extended like a brother for the past four years.

Finally, I would like to acknowledge Ms. Radhika Sreedhar, a friend, a neighbour, colleague, a smiling face, whose untimely departure from this world and the events thereafter reminded me of the harsh realities of life and made me a more humble being.

Author's declaration :

I, Ganesh Lad, hereby declare that all the work contained within this thesis is a result of my own work and has been written by myself.

The exceptions to this declaration include :

1. Nodulation test : Ms. Maria Kaye, an undergraduate student, helped me to perform the nodulation test.
2. PCA analysis : The PCA analysis using SPSS was performed by me. Since the scatter-plot was bunched together, the coloured, more resolved output was generated by Dr. Olivier Missa using the R programming language.
3. Heat-map in Python : The instruction set to generate the heat-map for gene-phenotype correlation in Python was written by Mr. Alex Leach.
4. Heat-map in R : The instruction set to generate the heat-map for gene-phenotype correlation in R was written by Ms. Piyachat Udomwong.
5. Making complementation strains for gene mutants : This work was carried out along with Mr. Kailin Hui, who was carrying out similar work for his PhD work. The laboratory work-load was distributed for this part of the work.

CHAPTER 1 : INTRODUCTION

1.1. Introduction :

The genome of most living organisms consists of thousands of genes that interact in many different ways. Since the sequencing of first organism, a bacteriophage, ϕX174 in 1977, the number of organisms that have been sequenced has been increasing rapidly. This has resulted in a huge amount of sequence data being generated and the growth of sequence data is expected in increase in the foreseeable future, especially with the development of fast, cheap and efficient sequencing technologies. Using the sequence data in a meaningful way entails the necessity to find correlations between genes in the genome sequence and their function. Most sequenced genomes feature a large number of genes with unknown function. The rate of decoding this information contained within the sequences has not kept pace with the rate at which sequence data is being generated.

The genes present in an organism are involved in a number of processes including series of reactions - called pathways, which form an essential part of its metabolism and constitute its phenotype. Given the complexity of the interactions, it is a daunting and expensive task to determine the role of genes in reactions of a metabolic pathway and identifying all the genes involved in it. If such genes are identified, then their function can be verified by performing mutation studies.

Most bacteria exhibit only a few physically-recognizable phenotypes, limited largely to cell shape and colony morphology when cultured on solid media, but a wealth of metabolic phenotypes exist, many of which can be scored as colour reactions involving dye-linked substrates incorporated into the culture medium. Phenotypic description has been used to discriminate between bacteria for a long time. With the publication of Bergey's Manual of Determinative Bacteriology in 1923, microbiologists began to systematically describe and define bacterial species based on lists of phenotypes. Because growth phenotypes are involved in fundamental aspects of bacterial physiology and evolution, they remain a cornerstone of microbial taxonomy (Pommerenke et al., 2010).

Phenotypic differences between bacteria were traditionally used in the classification of bacteria. These tests were carried out in the laboratory. However, the behaviour of an organism in nature may be very different from that observed under laboratory conditions. Hence, classification methods based on analysis of conserved core genes, like 16S, (Woese, 1987) have been proposed. However, it is not possible to predict the phenotypic profile of an organism even if its complete genome is known (Brenner et al., 2001, Ludwig and Klenk, 2000). Hence, a polyphasic approach, incorporating inputs from varied sources like sequence data from conserved genes, morphology, physiology, serology, pathology, and chemotaxonomy appears to be a more balanced approach to bacterial taxonomy (Vandamme et al., 1996, Gillis et al., 2005).

Linking the phenotype of a bacterium to its genomic composition and explaining it in terms of interaction between genes is more complicated than in eukaryotes. The polycistronic arrangement of genes on bacterial genomes makes it difficult to study the effect of individual gene mutations on the phenotype. This is further compounded by horizontal gene transfer and the effect of environmental factors. Hence, although more difficult than sequencing, there is now a need to find correlations between genes and functions which can then be extended to other organisms.

Some progress has been made in linking the genome and the metabolism using a systems approach. However, the link between the genotype and the phenotype remains tenuous (Pommerenke et al., 2010). Researchers use software to predict the functions of DNA sequences ("in-silico" biology). However, the only way to be sure what a particular sequence does is to experiment with the organism itself. Once a gene's sequence is known, it can be switched off or silenced. Systematically silencing genes can help reveal genes involved in specific cell processes. Hence, developing high-throughput approaches to assign genotypes to a particular phenotype and vice-versa needs to be given priority.

Most studies carried out to study association between genotype and phenotype in bacteria use a single or few isolates of bacteria to infer the relationship. A major shortcoming of this approach is that such an analysis may introduce bias in the analysis since a large number of metabolic genes may be accessory in nature and hence may not reflect the stable core genetic makeup of the organism under study. Hence, it is essential to use a bacterial system that shows diversity in both, its genotype and phenotype.

Rhizobia, a group of soil bacteria, induce formation of special structures, called nodules, on the roots of their host plants. In these nodules, these bacteria convert atmospheric nitrogen into ammonia – a process referred to as nitrogen fixation. Rhizobia show a remarkable diversity in terms of their metabolism and genomic sequence (Eugenia Marquina et al., 2011, Mazur et al., 2011, Wolde-meskel et al., 2004, Degefu et al., 2013, Pinero et al., 1988). Thus, there exists an opportunity for a systematic and in-depth study of genotypic and phenotypic correlations, make predictions based on association of genes with metabolic pathways, and further, to test the prediction by performing gene mutation studies using bacterial strains of known genetic composition.

The following sections serve as an introduction to rhizobia. Because of the varied aspect of the work included in the thesis, the relevant literature has been cited in the corresponding chapters.

1.2. Introduction to rhizobia :

As mentioned above, Rhizobia are soil bacteria that associate with plants and form specialized structures in which they convert atmospheric nitrogen into ammonia.

1.2.1. Taxonomy of rhizobia :

Beijerinck, in 1888, first isolated bacteria from root nodules and called them *Bacillus radicicola*. It was Frank (Frank, 1889) who placed the bacterium in a separate genus *Rhizobium* with a single type species *Rhizobium leguminosarum*. All legume root-nodule bacteria discovered since were placed in the genus *Rhizobium*. The original genus *Rhizobium* has now been divided into a number of genera. The term rhizobia, in the strictest sense, refers to the members of the genus *Rhizobium*, but is now used to refer to all bacteria capable of nodulating plants that once belonged to the genus *Rhizobium* or are closely related to it.

The taxonomy of rhizobia is dynamic. The earliest attempt of classifying rhizobia was based on their ability to infect specific host plants. This led to the classification of the bacteria based on cross-inoculation groups wherein the isolates from a particular cross-inoculation group were able to nodulate all the hosts included in the group (Fred et al., 1932). However, this was soon abandoned as a reliable taxonomic marker. The study of the different morphological, physiological and biochemical properties of rhizobia led to their differentiation into fast-growing rhizobia (generation time less than 6 hours) and slow-growing rhizobia (generation time greater than 6 hours). The slow growing rhizobia were later separated into a different genus, *Bradyrhizobium* (Jordan, 1982).

The advent of the modern techniques of molecular biology like sequencing of the 16S rRNA, Restriction Fragment Length Polymorphism (RFLP) and Multi-Locus Enzyme Electrophoresis (MLEE) has led to the classification of the rhizobia into 98 species in 13 genera (Weir, 2012).

Taxonomically, all known rhizobia are members of the phylum *Proteobacteria*, a major bacterial phylum, which includes a large number of pathogens. The phylum Proteobacteria consists seven classes of which six are represented by the Greek letters alpha through zeta, and a seventh class called *Acidobacteria* (Boone et al., 2005, Ciccarelli et al., 2006, Yarza et al., 2010, Williams and Kelly, 2013). All the classes within the phylum are monophyletic except *Gammaproteobacteria* (Williams et al., 2010)

Rhizobia form a polyphyletic group with members in two classes of Proteobacteria - the *Alphaproteobacteria* and the *Betaproteobacteria*. Most rhizobia belong to the order *Rhizobiales* of *Alphaproteobacteria*. (Weir, 2012, Vinuesa, 2012, Moulin et al., 2001).

The order *Rhizobiales* incorporates at many genera of nitrogen-fixing bacteria contained in the families *Bradyrhizobiaceae*, *Hyphomicrobiaceae*, *Phyllobacteriaceae*, *Xanthobacteraceae* and *Rhizobiaceae* (Kuykendall and Young, 2005). The family *Rhizobiaceae* includes seven genera – *Agrobacterium, Carbophilus, Chelatobacter, Kaistia, Ensifer*, candidatus *Liberibacter* and the type genus – *Rhizobium*.

1.2.2. The genus *Rhizobium* :

The genus *Rhizobium* was first described by Frank (1889). The bacteria are Gram-negative, aerobic, motile, and rod-shaped when cultured on media. The average %G-C content is 57-66. They induce hypertrophism in plants in the form of root nodules with or without symbiotic nitrogen fixation. Within the nodules the bacteria occur as endosymbionts and exist as pleomorphic forms, termed "bacteroids" that reduce and fix gaseous atmospheric nitrogen into a form utilizable by the host plant (Kuykendall and Young, 2005).

Within the genus *Rhizobium*, different species are distinguished based on their genotypic and phenotypic properties. The type species of *Rhizobium* is *Rhizobium leguminosarum*, which was first described by Frank in 1879 as *Schinzia leguminosarum*. *Rhizobium leguminosarum* (%G-C = 59-63) is amongst the well-studied species of *Rhizobium*. The species has been genetically well characterised. Members belonging to the species *Rhizobium leguminosarum* cannot be further distinguished on any other characters except their host range. Based on this ability to nodulate specific host plants, *Rhizobium leguminosarum* is further divided into three categories which are referred to as biovars of *Rhizobium leguminosarum*: biovar trifolii, biovar phaseoli and biovar *viciae* (Kuykendall and Young, 2005). In *Rhizobium leguminosarum,* the genes responsible for nitrogen fixation, nodulation and host-range determination are present on large symbiotic plasmids. Hence, the classification into biovars can be said to a classification based on the genetic composition of the plasmids (or accessory genome), rather than the chromosomal genome. Since the genes responsible for the host adaptation are present on symbiotic islands, Rogel *et al.* (2011) proposed that the term 'symbiovar' be used instead of 'biovar' to describe functions that more accurately describes the properties conferred by the accessory genomes of bacteria.

1.2.3. The rhizobial genome :

The process of nodule formation involves the participation of a large number of molecules encoded by the rhizobial genome. As with any bacterial genome, the rhizobial genome can be physically divided into two parts – the chromosomal DNA and extrachromosomal DNA, like plasmids (Martinez et al., 1990). Chromosomal DNA houses most of the genes essential for cell survival and replication while plasmids confer the ability on bacterial populations to colonize and compete in natural communities (Mercado-Blanco and Toro, 1996). In the fast growing rhizobia - *Rhizobium* and *Ensifer*, the plasmids play an important role in their interaction with plants while in *Azorhizobium*, *Bradyrhizobium* and *Mesorhizobium*, they generally do not appear to have this function (Hanin et al., 1999). A possible explanation for this might be that in the latter group, the symbiosis genes are mostly found as genomic islands on the chromosome.

The genome can also be conceptually divided into two components : the core or basic genome and the accessory genome. The two components differ in their genetic content suggesting a more constrained genome bias for the basic genome than the accessory genome – a difference that can be used to distinguish between them (Young et al., 2006). The core genome consists of the house keeping genes. These genes are concerned with metabolism and the processing of information resulting in cellular homeostasis (Feil, 2004). The core genes tend to have higher G+C content and are rather consistent in their codon usage (Young et al., 2006, Hacker and Carniel, 2001). The core genome is also the portion of genes which is shared by all members of the same species. This model can also be applied to other taxonomic levels or ranks, resulting in the formation of core genomes all different taxonomic ranks which contain the genes common to all members at that level of rank (Feil, 2004).

The accessory genome consists of genes which are not necessarily present in all members of the species (Young et al., 2006). Edward Feil (2004) describes these genes as being ‘dispensable’ because they can be removed from a bacterium in the laboratory without affecting its survival in the laboratory, but may have a fitness impact in its natural environment. The accessory genes have a lower G+C content and have more differences between them, in terms of codon usage than the core genes (Hacker and Carniel, 2001). These genes can usually belong to one or more of the following groups, genes carried by phages or other mobile elements, genes coding for biochemical pathways (supplementary to those encoded in the core genes) and finally genes concerned with the interaction with the environment (Hacker and Carniel, 2001).

Elements within the accessory genome can be transferred between genomes by transduction, transformation or conjugation. There is evidence that horizontal gene transfer has played an important role in the diversification of certain bacteria. For example, in the case of some bacterial symbionts of *Medicago* plant species, recombination has allowed two different species of bacteria (*Sinorhizobium medicae*, *Sinorhizobium meliloti* bv. meliloti) to interact with the same plant (*M. truncatula*), but also for two variants of the same species (*Sinorhizobium meliloti* bv. meliloti, and *S. meliloti* bv. medicaginis) to have symbiosis with two different plant species (*M. truncatula* and *M. laciniata* respectively), thus facilitating reproductive isolation (Bailly et al., 2007). Thus, having a stable core and a transferable accessory genome allows the exploitation of new environments and the ability to respond to competition without the danger of disrupting essential processes (Feil, 2004).

1.3. The rhizobial host plants :

The host plants that form a symbiotic association with rhizobia belong to the family *Leguminosae* (*Fabaceae*). *Leguminosae* constitutes the third largest family of flowering plants (after *Orchidaceae* and *Asteraceae*) with 730 genera and over 19,400 species. The species of this family are found throughout the world, growing in many different environments and climates. A number of species belonging to the family *Leguminosae* are important agricultural plants (Wojciechowski. F.; Mahn. J.; Jones, 2006). The family is divided into three subfamilies viz. *Caesalpinioideae*, *Mimosoideae* and *Faboideae* (*Papilionoideae*) with effective nodulation being observed in 29%, 90% and 97% of species examined (Bryan et al., 1996).

Besides *Leguminosae*, a member of the elm family, *Parasponia*, can form nodules with root-nodulating bacteria The *Parasponia* nodules have extremely ramifying infection threads from which the bacteria are not released (Trinick, 1979). The non-legume rhizobium symbiosis may have evolved independently from legumes and has done so only recently (Op den Camp et al., 2012).

1.4. Rhizobial ecology :

The rhizobium-plant symbiosis is ecologically and economically important, being responsible for about 50% of the total nitrogen fixed annually by natural processes (Dixon and Wheeler, 1986), providing 25-35% of the world-wide protein intake (Vance, 1998). Nitrogen fixing bacteria are an eco-friendly alternative to the nitrogen-based fertilizers that have detrimental effects on the environment (Newbould, 1989, Kinzig and Socolow, 1994). They are the most important providers of nitrogen in low-input agriculture and natural terrestrial ecosystem. The rhizobial numbers can vary from undetectable to 10^6 rhizobia per gram of soil (Somasegaran and Hoben, 1994).

1.5. Infection of plants by *Rhizobium leguminosarum* :

The symbiosis between rhizobia and legumes is a result of a complex two-way molecular communication that results in the formation of nodules (Franssen et al., 1992, Fisher and Long, 1992). The communication is initiated by the plants that release signal molecules into the rhizosphere. The rhizobia detect the presence of these signal molecules and move up the gradient towards the plants by positive chemotaxis (Caetano-Anolles and Gresshoff, 1991). In turn, the bacteria produce their own signal molecules called Nod factors which are released in the rhizosphere (Heidstra and Ton, 1996).

The Nod factors act on zone II root hairs, which are the primary target cells for root infection, and induce root hair deformation and curling (Heidstra et al., 1994, Heidstra et al., 1997). The rhizobia get entrapped in the curling roots hairs. The plasma membrane of the cell walls of the host plant invaginates and new plant cell wall material is synthesized around it to form a tubular structure, called the infection thread or IT. The bacteria enter the plant root hairs through this infection thread (Rae et al., 1992). The root cortical cells in indeterminate nodules and hypodermis cells in determinate nodules, dedifferentiate and start dividing, to form a nodule primordium (Hadri et al., 1998).

The bacteria travel from the root hair to the nodule primordium through the infection thread. When they reach the nodule primordium, they undergo morphogenesis and become pleomorphic. This pleomorphic form of

Rhizobium, the bacteroid, is the endosymbiotic form of *Rhizobium* (Sprent and Sprent, 1990). The nodules then develop into their nitrogen fixing form. A nodule may develop into either of the two types viz. indeterminate or determinate. Indeterminate nodules contain a persistent meristem. The central tissue of indeterminate nodules is divided into 5 zones : meristem, invasion, development of symbiosomes (bacteroids enclosed by peribacteroid membrane), N fixation & senescence. Determinate nodules do not have a persistent meristem. The cells of the central tissue of determinate nodules are all at a similar developmental stage (Cohn et al., 1998, Rae et al., 1992).

The enzyme responsible for fixing nitrogen, nitrogenase, is oxygen sensitive which poses a problem in the nodule environment. This problem is overcome by the synthesis of a respiratory protein called leghemoglobin. The protein is found in effective nitrogen fixing nodules. It has a high affinity for oxygen and binds to it, making the oxygen concentration low enough to allow nitrogenase to function but high enough to allow bacterial respiration for a few seconds (Bruijn et al., 1989, Bray, 1983).

The life of a nodule, however, is limited (Sprent, 1979). Nodules are shed as the plant comes into flower or fruit (Andreeva et al., 1998), but senescence may be affected by the environment (Sprent and Sprent, 1990).

1.5.1. Plant exudates and signal molecules :

Plants are known to secrete a variety of compounds from their roots including amino acids, sugars, and organic acids. These compounds may function as nutrients for bacteria present in the rhizosphere. They also act as chemoattractants and induce the expression of *nod* genes in rhizobia (Firmin et al., 1986, Peters et al., 1986, Redmond et al., 1986, Spaink et al., 1987). The plants have been shown to regulate the number and type of bacteria that grow in its rhizosphere by regulating the amount and type of signal molecules extruded. Pea roots, for example release an unusual amino acid, homoserine, which is a preferred source of carbon and nitrogen for the pea root symbiont, *Rhizobium leguminosarum* (biovar *viciae*). A large increase in the number of rhizobia on the pea root surface was found to correlate with the liberation of significant amount of homoserine (Egeraat, 1975).

Another group of chemicals that have been implicated in attraction of rhizobia are the phenolics of the flavonoid group (Firmin et al., 1986, Peters et al., 1986, Redmond et al., 1986, Spaink et al., 1987). The identification of flavonoids in root exudates was first reported by Lameta and Jay (1987) for soybean seedlings. A wide variety of flavonoids, all synthesized via the phenylpropanoid pathway, have since been characterized in root exudates (Stafford, 1997). These include chalcones, flavonones, flavones, flavonols, isoflavonoids, coumestans and anthocyanidin, many of which stimulate nodulation but some of which actually inhibit the process (Rolfe, 1988). The interaction between inducer and rhizobia is species specific. For e.g. Different rhizobial species react differently to different inducers (Denarie et al., 1992, Stafford, 1997).

1.5.2. Bacterial genes involved in symbiosis :

A number of genes are involved in the ability of rhizobia to successfully establish a symbiosis. These genes can be grouped into a number of groups based on the role they play. Some of them are directly involved in the process of nitrogen fixations whereas others are involved in other interactions between the host plant and the bacteria that indirectly influence the symbiotic association. The two groups of genes that directly influence nitrogen fixation are the nodulation genes (*nod, nol and noe*) and the nitrogen fixation genes (*nif* and *fix*) (Denarie et al., 1992, van Rhijn and Vanderleyden, 1995). The other groups of genes that are indirectly involved in the establishment of successful symbiosis include the genes involved in exopolysaccharide synthesis (*exo*), rhizosphere expression (*rhi*), bacterial attachment and adhesion and other genes which are active at different stages of infection and nitrogen fixation (Leigh and Walker, 1994, Niner and Hirsch, 1998).

1.5.2.1. Genes involved in nitrogen fixation :

As mentioned above, two groups are genes are involved in nitrogen fixation – the *nif* and *fix* genes, which are usually arranged in operons. The nitrogen fixation (*nif*) genes were first identified in *K. pneumonia* (Dixon et al., 1977) and then found to be conserved in nitrogen fixing organisms (Ruvkun and Ausubel, 1980).

Based on work carried out in *K. pneumonia* and *A. vinelandii* the roles of a number of *nif* genes has been established or proposed. Some gene products are structural components of the nitrogenase catalytic component (*nifH, nifK, nifD, nifJ, nifF*); others synthesize regulatory proteins (*nifA, nifL*), assemble the reactive cluster of nitrogenase (*nifQ, nifB, nifN, nifE, nifV, nifS*) or are involved in enzyme processing (*nifM, nifY*). The functions of some of the *nif* genes are not known (*nifW, nifT, nifZ, nifU*). Homologs for genes encoding these proteins have been found in several organisms. Rhizobia have about 10 *nif* genes whose location differs in different rhizobia and which together encode the enzyme nitrogenase (Iismaa and Watson, 1989). Rhizobia leguminosarum biovar viciae strain 3841 has 8 *nif* genes (nif*N, nifE, nifK, nifD, nifH, nifA, nifB* and *nifS*).

The *fix* genes were first identified in *R. meliloti* because of their close association with the *nif* genes. Mutants in *fix* genes were found to make nodules which did not fix nitrogen (Dusha et al., 1987, Earl et al., 1987). Some *fix* genes function as a membrane-bound cytochrome oxidase (*fixABCX*), *fixGHIS* function as a cation pump, while the functions of most other *fix* genes is not known (Fischer, 1994).

1.5.2.2. Genes involved in nodulation :

Nodulation genes can be defined as genes that play a role in regulation or are co-ordinately regulated by such genes. These constitute the *nod*, the *nol* and the *noe* genes. In rhizobia, these genes are usually located on a large plasmid and are a part of the accessory genome that can be transferred horizontally (Johnston and Beringer, 1977).

The *nod* genes are of fundamental importance to the initiation of symbiosis. They synthesize the Nod factors which act as signal molecules which are essential in the first steps of infection and nodule formation (Martinez et al., 1990, Denarie et al., 1996). A conserved region of DNA sequence, called the *nod* box, allows coordinated regulation of the *nod* genes (Yokota and Hayashi, 2011, Rostas et al., 1986).

Some nodulation genes are conserved in all rhizobia while others are restricted to single species or strains – called 'common' and 'host-specific' nodulation genes respectively. The common *nodABC* genes normally occur as a single operon and have been identified in all rhizobia isolates studied so far (Denarie et al., 1992, Martinez et al., 1990). They encode the enzymes which are responsible for the synthesis of the basic core Nod factor molecule. *nodA* encodes an acyltransferase; *nodB* encodes a chitooligosaccharide deacetylase while *nodC* encodes an N-acetylglucosaminyltransferase.

The *nodD* gene is also found in all rhizobia and is constitutively expressed. The product of *nodD* acts as a transcriptional activator on binding to signal molecules secreted by plants in root exudates (van Rhijn and Vanderleyden, 1995, Rossen et al., 1985, Spaink et al., 1987, Fisher and Long, 1992). The NodD protein, identified as member of the LysR family of transcriptional activators (Henikoff et al., 1988, Schell, 1993, Schlaman et al., 1992), binds to nod box promoters located upstream of the *nod* genes (Schlaman et al., 1998).

The host-specific nodulation genes determine the host plant species that the bacteria are able to nodulate and actively fix nitrogen i.e. they define their host range. (Putnoky and Kondorosi, 1986, van Rhijn and Vanderleyden, 1995, Roche et al., 1996, Denarie et al., 1996, Downie, 1998, Hanin et al., 1999).

1.5.3. Nod factors :

The signal molecules encoded by the *nod* genes are called Nod factors. These molecules interact with the host plant and are the key signals in the host-symbiont recognition (Denarie et al., 1992, Spaink et al., 1991). Nod factors consist of a lipochitooligosaccharide (LCO) backbone of β-1,4-linked *N*-acetyl-D-glucosamine (GlcNAc) varying in length from three to five sugars. The non-reducing end is acylated by a variety of fatty acids. At the reducing terminus there are a variety of possible specific substitutions, all of which are *nod* gene dependent. The host specific Nod proteins modify the basic structure of LCO by adding the various substituents (Streng et al., 2011, Gough, 2003, Carlson et al., 1994, Downie, 1998, Downie, 1991, Hanin et al., 1999, Schultze et al., 1992).

The synthesis of Nod factor probably takes place in the cytosol and the inner membrane of the bacteria since most Nod enzymes are located here (Carlson et al., 1994). *nodI* and *nodJ* are involved in the secretion of the Nod factors, although the other transport systems have been shown to compensate for the absence of these genes (Carlson et al., 1994, Spaink et al., 1995). The *nodIJ* genes are present downstream of *nodC* in most rhizobium species as part of the same operon as *nodABC* (Hanin et al., 1999). Mutations in these genes have a species-specific effect. Mutants in biovar *trifolii* are unable to secrete Nod factors, while mutants in biovar *viciae* are only moderately affected (Downie, 1998). Moreover, mutations of either of these genes decrease the secretion and increase the intra-cellular accumulation of Nod factors. The function of proteins encoded by the nod genes is listed in Table 1.1.

Table 1.1. *nod* genes and the functions of proteins encoded by them (NCBI gene database).

Gene	Protein function
nodE	3-ketoacyl-ACP synthase
nodF	Acyl carrier protein
nodG	3-ketoacyl-ACP reductase
nodH	Sulphotransferase
nodI	ABC transporter, ATP-binding protein
nodJ	ABC transporter, permease protein
nodL	O-acetyltransferase
nodM	Glucosamine-fructose-6-phosphate aminotransferase
nodN	Dehydratase
nodO	RTX toxins and related Ca^{2+}-binding protein
nodP	sulfate adenylyltransferase subunit 2
nodQ	Sulfate adenylyltransferase subunit 1 / Adenylylsulfate kinase
nodS	N-methyl transferase
nodT	Secretion system type I outer membrane efflux pump lipoprotein
nodU	6-O-carbamoyl transferase
nodV	Two-component family sensor
nodW	Two-component family regulator

nodX	Sugar acetylase / O-acetyl transferase
nodY	Transcriptional regulator
nodZ	Chitin oligosaccharide fucosyltransferase

1.5.4. Plant response to Nod factors :

Nod factors induce responses in three different plant tissues – epidermis, cortex and pericycle. But the responses of host plants to Nod factors are diverse and hence it has been suggested that many plant receptors may be involved in these responses (Broghammer et al., 2012, Hanin et al., 1999), The few Nod factor receptors that have been identified so far appear to belong to a specific class of receptor kinases that have LysM domains in their extracellular domains (Popp and Ott, 2011, Geurts et al., 2005).

The Nod factors induce a variety of responses in the host plant. These changes may be classified as morphological changes, physiological changes and changes in gene expression. The morphological changes include seed germination, changes in root hair morphology such as hair deformation, curling, formation of infection thread to allow the entry of bacteria and induction of cortical cell division.

Low concentrations of the Nod factors are sufficient to induce hair deformation, curling and formation of infection thread, whereas relatively higher concentrations are required for the induction of cortical cell division. These changes can be induced by the Nod factors themselves even in the absence of bacteria (Broghammer et al., 2012, Kidaj et al., 2012, Oldroyd et al., 2011, van Brussel et al., 1992, Roche et al., 1991, Spaink et al., 1991, Banfalvi and Kondorosi, 1989, Dudley et al., 1987, Y Yao and Vincent, 1969).

1.6. Metabolism in rhizobia :

According to Kuykendall and Young (2005), the members of the genus *Rhizobium* primarily catabolise glucose through the Entner-Doudoroff pathway (ED) and the pentose phosphate pathway (PP). It is unlikely that the Embden-Meyerhof-Parnas pathway (EMP) operates in *Rhizobium* spp. because activities of fructose-1,6 diphosphate aldolase and 6-

phosphofructokinase are low. The tricarboxylic acid cycle is operative, and the enzymes of the glyoxylate bypass are present. Pyruvate carboxylase is an important anaplerotic enzyme. Although the end products of the EMP and ED pathways is pyruvate, the EMP pathway forms 2 moles of ATP per mole of glucose whereas the ED pathway forms just one making it less efficient than the former. The PP pathway (also called the Hexose Monophosphate Pathways – HMP) is a shunt pathway which produces five-carbon sugar intermediates for nucleotide / nucleic acid synthesis and some amino acids. The three pathways also generate reducing powers; while the EMP pathway forms 2 moles of NADH per mole of glucose, the ED pathway forms one mole each of NADH and NADPH while the PP pathways forms two moles of NADPH. These differences are shown in Figure 1.1.

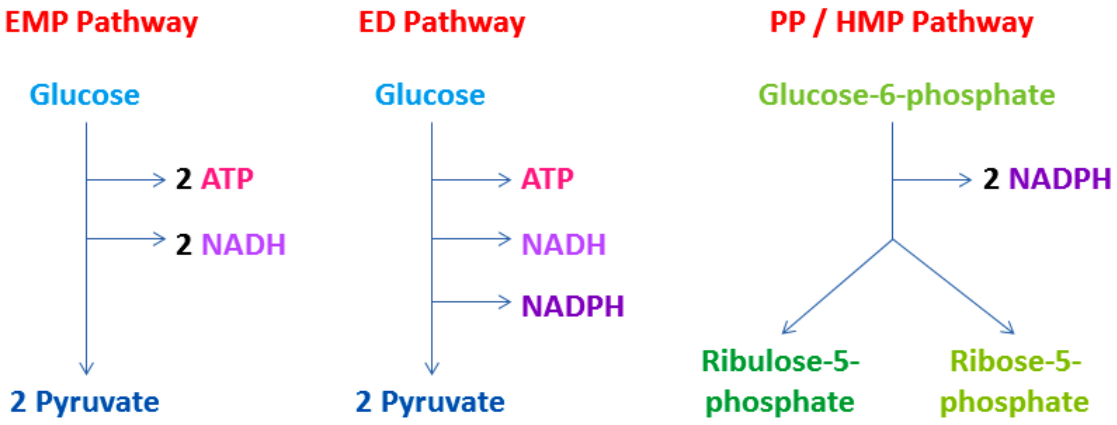


Figure 1.1. The primary pathways of glucose metabolism viz. Embden-Meyerhof-Parnas pathway (EMP), the Entner-Doudoroff pathway (ED) and the pentose phosphate pathway (PP / HMP). The difference in the amount of ATP and reducing equivalents synthesized is shown.

Metabolism in rhizobia has been studied in some depth. Metabolic classification of rhizobia was used to divide the rhizobia into two groups viz. the fast-growing rhizobia (generation time less than 6 hours) and slow-growing rhizobia (generation time greater than 6 hours) (Jordan, 1982). This metabolic difference led to the segregation of slow-growing rhizobia into a separate genus called *Bradyrhizobium* based on Jordan's description. It was soon observed that the fast-growing rhizobia were diverse in terms of their ability to utilize carbon substrates as compared to the slow-growing rhizobia. The slow-growing rhizobia, however, showed more diversity in their ability to utilize aromatic and hydroaromatic compounds (Parke and Ornston, 1984).

The diversity of metabolism in rhizobia has been extensively studied (Fall et al., 2008, Wei et al., 2008, Mierzwa et al., 2009, Ramirez-Bahena et al., 2009, Mehnaz et al., 2010, Djedidi et al., 2011, Rasul et al., 2012, Bianco et al., 2013, Xu et al., 2013). Most of these studies have assessed phenotypic diversity and correlated it to the genomic diversity of the isolates. A second approach to study metabolism has been to study the metabolism of single substrate to try and predict the genes involved in the metabolic process. This includes the study done on rhizopine metabolism by Bahar et al. (1998), myo-inositol metabolism by Fry et al. (2001), gamma-aminobutyrate by Prell et al. (2009), and homoserine metabolism by Vanderlinde et al. (2013), amongst others. Most of these studies use a small number of isolates to establish the correlation. However, small sample sizes may be affected by horizontal gene transfer processes leading to spurious correlations. Hence, it is important to have a relatively larger population size wherein the correlations that are made would be more reliable to make predictions about the correlation between presence / absence of a gene and the phenotype.

1.7. Project background :

Previous work in the lab with bacterial isolates obtained from a small area in a field adjacent to Wentworth College in the University of York campus analysed allelic variation at shared loci in the bacterium *Sinorhizobium medicae* to detect genes that are not in all strains (Bailly et al., 2011). The analysis of genome data revealed the presence of genes for synthesis of rhizobitoxine and genes involved in quorum-sensing. Similar genomic data is now available for other species and biovars of *Rhizobium*.

The genomic data for *R. leguminosarum* shows the presence of genes that are present in one biovar but absent in the other (biovar-specific genes). The analysis of the genome sequences also suggests that bacterial biovars have a part of the genome that is conserved and difference in the biovars arise due to differences in the accessory genes. This is true even when the genome is compared between different strains of the same biovar. Strains may differ by the presence or absence of hundreds of genes that may be chromosomal or extrachromosomal in location and may confer a selective and adaptive advantage to the organism. The accessory genome thus constitutes a reservoir of potential adaptation responding to environmental change.

1.8. Aims of the study :

The project aims to understand the genetic basis of the phenotypic differences in the bacterial population. In this study, we have systematically analysed a large set of isolates of *Rhizobium leguminosarum* for a large set of different phenotypes in order to investigate its relationship to the genome. The study is being carried out with the following working hypothesis : Given that there is phenotypic variation in a bacterial population, it should be possible to explain the variation in terms of its genome.

The aims of the project can be briefly summarized as follows :

a. To investigate the phenotypic variation in bacterial population from a small geographic community.
b. To map the phenotypic data onto the sequence data of the isolates to identify candidate genes showing strong correlation with specific substrate utilization.
c. To mutate selected examples of candidate genes showing correlation with substrate utilization and observe the effect of mutation on the utilization of the substrate.
d. To complement loss of function of the mutated gene by complementation and to observe its effect.

1.9. Thesis overview :

The thesis is divided into the following chapters :

Chapter 1 : General introduction : A general introduction to the thesis, with a review of literature and an outline of the general aims of the project.

Chapter 2 : Validation and study of variation in homoserine catabolism and AHL signal molecule synthesis in Wentworth strains : The strains were validated by checking their ability to nodulate plant hosts. The preliminary investigations were carried out to test phenotypic differences between biovars *trifolii* and *viciae* reported in literature. The two differences tested were the ability of the strains to use homoserine as a carbon source and the ability of strains to synthesize quorum sensing signal molecules.

Chapter 3 : Metabolic fingerprinting of Wentworth strains using Biolog Phenotype Microarray^{TM} plates : In order to carry out a systematic assessment of the phenotypic differences, a metabolic fingerprint of the strains was developed using the Phenotype Microarray technology from Biolog Inc. The data were statistically analysed to identify isolates having similar metabolic patterns.

Chapter 4 : Analysis of phenotype data to identify genes involved in metabolism of substrates in the Biolog GN2 plates : Since the aim of the project is to understand the genetic basis of phenotypic diversity, bioinformatic analysis of the sequence data will be carried out to identify genes that show strong correlation with the ability or inability of a strain to use a particular carbon substrate.

Chapter 5 : pRL100135 and pRL100135 are involved in the utilization of γ-butyrolactone and γ-hydroxybutyrate : In order to investigate the role of genes showing a strong correlation with metabolism, the genes were mutated and the effect of mutation was studied by growing the mutant in the presence of its correlated carbon substrate. The mutation was complemented to check for reversion of loss of function.

Chapter 6 : General discussion : A general commentary on the project and possible future work that can be done to extend the work.

CHAPTER 2 : VALIDATION AND STUDY OF VARIATION IN HOMOSERINE CATABOLISM AND AHL SIGNAL MOLECULE SYNTHESIS IN WENTWORTH STRAINS

2.1. Introduction :

The overall purpose of this study is to obtain phenotypic data for a collection of bacterial strains for which genome sequence is already available, in order to explore the relationship between genotype and phenotype. In this chapter, I will investigate some phenotypic characteristics that are known to be of importance in the ecology of rhizobia and are already known to vary among strains within my study species, viz. homoserine utilisation and AHL signalling. Before performing the phenotypic tests, nodulation tests were performed using the native hosts of the Wentworth strains for strain validation. After confirming their ability to nodulate the host plants, the two phenotypic characteristics were studied using the two biovars to be used in this study.

2.1.1. Nodulation test :

Nodules are produced as a result of complicated signal exchange between rhizobia and the host plant (Spaink, 2000). This signal, which is in the form of chemical molecules, dictates the specificity of the plant-bacterial association and is governed by the regulation of several classes of specific genes. Some bacteria form nodules only with a limited number of legumes and are said to have a narrow host range while others are less discriminating, infecting a large number of host plants, and are hence said to have a broad host range (Heidstra and Ton, 1996).

The process of nodulation is initiated when the host legume releases signal molecules into the rhizosphere. The rhizobia in the rhizosphere respond by positive chemotaxis (Caetano-Anolles and Gresshoff, 1991) and produce the bacterial chemical signals – the Nod factors (Heidstra and Ton, 1996, Geurts and Bisseling, 2002). The Nod factor induces root hair curling and invagination of the plasma membrane of the cells forming a tubular structure (called the infection thread) through which the bacteria enter the plant cells (Rae et al., 1992) and reach the nodule primordium where they differentiate into bacteroids and initiate nitrogen fixation (Sprent and Sprent, 1990).

The nodulation process begins with infection of the root by the bacteria and ends with the formation of mature nodules in which nitrogen is fixed. The process involves a sequence of a number of interactions between the bacteria and the host roots. In effect, the rhizobia and the roots of the prospective host plant establish a dialogue in the form of chemical messages passed between the two partners. (Sprent and Sprent, 1990).

It is important to determine that the isolate being used in the study of rhizobia is a pure culture. A pure culture of a rhizobial isolate should be able to form nodules on legume roots and proves the authenticity the pure culture. This study is referred to as the nodulation test (Somasegaran and Hoben, 2011) and is routinely used not only in strain authentication but also in the describing new species of rhizobia (Guerrouj et al., 2013, Marek-Kozaczuk et al., 2013, Wang et al., Wang et al., 2013, Liu et al., 2012).

2.1.2. Homoserine test :

2.1.2.1. Introduction – Rhizosphere and plant exudates :

The term “rhizosphere”, (Greek : “rhiza” = root, “sphere” = field of influence), introduced by Hiltner (1904), was used to describe the zone of soil immediately adjacent to legume roots that supported a high level of bacterial activity. However the term now includes the region from the root surface, the “rhizoplane”, into the soil from a few millimetres to a few centimetres (Campbell and Greaves 1990).

The rhizosphere contains many organic molecules that include sugars, organic acids, amino acids, proteins, aliphatic and aromatic compounds. The biological, chemical and physical conditions of soil in the rhizosphere are greatly influenced by plant roots. A number of compounds, including secondary metabolites, are secreted by the plants into the rhizosphere. These compounds secreted by the roots of plants are collectively referred to as ‘root exudate’.

2.1.2.2. Composition of root exudates :

Root exudates are compounds released from the root cells. These compounds can be split into five groups viz. diffusates, gases, lysates, mucilage and secretions (Grayston and Campbell, 1996., Lynch 1990).

Diffusates comprise low molecular weight, water-soluble compounds that passively diffuse from the root to the rhizosphere. Sugars, organic acids, and amino acids exuded from the plant roots can be categorised as diffusates.

The main gases exuded from plant roots are ethylene, carbon dioxide, hydrogen cyanide and are usually end products of different metabolic pathways.

Lysates consist of organic materials released into the soil after autolysis of dead cells.

Plant mucilage is composed of polysaccharides and polygalacturonic acids and helps roots penetrate into the soil. If the plant mucilage contains microbial mucilage then it is referred to as “mucigel”.

Secretions contain low or high molecular weight compounds actively secreted by the roots (Bowen and Rovira, 1999, Benizri et al., 2001, Kang and Mills, 2004, Sun and Wang, 2013).

The predominant diffusates in the root exudates are organic acids and amino acids. The organic acids found in root exudates probably play a role in solubilisation of mineral nutrients (Knee et al., 2001, Walker et al., 2003, Prell and Poole, 2006). The amino acids appear to play a less significant role in this process. The proteinaceous amino acids are rarely detected in root exudates whereas the non-proteinaceous amino acids (phytosiderophores) in the rhizosphere may enhance the mobility of plant micronutrients in soil.

The legume rhizosphere is particularly rich in unusual carbon sources, including non-protein amino acids, flavonoids, and phenolic compounds, some of which are toxic to humans and animals (Bell, 2003, Lambein et al., 1993).

The composition of root exudate varies greatly and depends on plant species, physiological conditions of the plant such as plant age, general health of the plant and nutrient status, abiotic conditions such as light, temperature, soil structure, soil aeration, and water content, and microbial activity at the root surface (Hale and Moore, 1980, Campbell and Greaves, 1990). Any type of stress on plant growth may also increase plant root exudation (Curl, 1982).

2.1.2.3. Microbial community of the rhizosphere :

The nutrient-rich rhizosphere supports a wide variety and number of microbes. The composition of the exudate determines the type of organisms residing in the rhizosphere. Hence, the number and variety of organisms in the rhizosphere are related either directly or indirectly to root exudates and vary as per the environmental conditions that influence exudation (Shi et al., 2011, Rovira, 1969, Cook and Baker, 1983, Lundberg et al., 2012). Bolten et al. (1993) suggested that root exudation in plants may have evolved as a means to stimulate an active rhizosphere microflora (Metting, 1993). This argument seems valid since the numbers of microorganisms generally decrease with an increase in the distance from the root surface.

The residents of the rhizosphere can be classified into five groups :

- **Decomposers :** The decomposers include the organisms that are involved in the decomposition and mineralization of organic matter in the rhizosphere, converting it into plant-available forms (van der Heijden et al., 2008).
- **Plant growth-promoting rhizobacteria (PGPR) :** PGPR are specific strains of bacteria in the rhizosphere that enhance seed germination and plant growth. The plant growth promotion may be direct (e.g. secretion of plant growth regulators), indirect (e.g. preventing growth of pathogens by synthesizing antimicrobial compounds), or a combination of both (Whipps, 2001).
- **Mycorrhizal fungi :** Mycorrhizae enhance nutrient solubilisation and absorption in the rhizosphere and expand the volume of soil that the root can explore (Hardie, 1985).
- **Nitrogen-fixing bacteria :** Plants cannot fix atmospheric nitrogen which is one of the main elements limiting plant growth (Chapin, 1980). Rhizobia form nodules on the roots of plants and reduce atmospheric nitrogen to

ammonia, which is a form usable by the plant. The rhizobial-legume symbiosis is normally very specialized. A given rhizobial species nodulates only specific species of plant hosts while legume plants are nodulated by restricted species of rhizobia. Exceptions include Sinorhizobium sp. NGR234, which nodulates over 110 genera of legumes (Pueppke and Broughton, 1999) and *Phaseolus vulgaris* (the common bean), which is nodulated by at least twenty species of rhizobia (Michiels et al., 1998).

- **Pathogens :** Plant pathogens cause plant diseases and include fungi, bacteria, and nematodes. Beneficial interactions of roots with soil micro-organisms are much more frequent than interactions between roots and pathogenic micro-organisms. However, the impact of pathogens on agricultural plants can be enormous, leading to complete destruction of plants and loss of yield (Teng et al., 1984, Katan, 1996).

2.1.2.4. Root exudate and microbial community in the rhizosphere :

Root exudate is an important factor affecting microbial growth in the rhizosphere. The bacterial numbers are many fold greater in the rhizosphere than in the surrounding soil, as is the microbial community. The root exudate can influence the growth and type of organisms in the rhizosphere by selective secretion of those substrates that allow the growth of bacteria that benefit plant growth. This phenomenon in which microbial growth and activity is stimulated due to the effect of plant roots is called as the “rhizosphere effect” (Foster, 1986).

It was suggested many years ago that specific compounds in the roots of different species or cultivars of legumes might affect the growth of specific strains of rhizobia on the roots of that type of plant. Nodulation of legumes is dependent upon the correct combination of symbiotic bacteria and plant species. Competition between symbiotic bacterial strains within the same species also occurs for a particular host plant. This intraspecies competition depends upon a number of factors such as the genes involved in nodulation, abiotic factors, and biotic factors (Shi et al., 2011).

Abiotic environmental factors such as soil type, plant species and the nutritional status of bacteria may influence the rhizosphere community.

The biotic factors include bacteriocins and a variety of compounds synthesized during the symbiotic interaction (Vlassak et al., 1997). The compounds synthesized during symbiotic interaction consists of compounds such as the betaine, trigonelline (Boivin et al., 1990, Goldmann et al., 1991), homoserine (Egeraat, 1975), certain flavonoids (Hartwig et al., 1991, Hartwig and Phillips, 1991), and rhizopines. Colonization of the legume rhizosphere bacteria can be promoted by their ability to utilize these specific carbon substrates exuded by the root. For example, homoserine secreted by pea plants is assimilated by the pea symbiont *R. leguminosarum* bv. *viciae* (Egeraat, 1975, Armitage et al., 1988, Hynes and O'Connell, 1990). Mimosine, found in *Leucaena* plants, is a carbon and nitrogen source for many rhizobium isolates from *Leucaena* (Soedarjo et al., 1994). *Agrobacterium* spp. transform plant cells to make them produce unusual amino acid derivatives, opines, that can be metabolized specifically by strains of *A. tumefaciens* (Scott et al., 1979). Similarly, rhizopines synthesized in nodules are catabolized by other rhizobia in the soil. Rhizopines are usable only by a limited number of *S. meliloti* and *R. leguminosarum* bv. *viciae* strains.

Differential degradation of soil and rhizospheric substances may avoid sympatric bacterial competition. Divergence to avoid competition for nutrients may be a driver of rhizobial speciation and also of intraspecific variation. Natural molecules have not been tested in the laboratory as some of them are unknown, not commercially available or are very expensive. Hence, the spectrum of rhizobial degradative capabilities is largely unknown (Ormeno-Orrillo and Martinez-Romero, 2013).

2.1.2.5. Homoserine and *Rhizobium leguminosarum* :

The initial signalling in the plant-*Rhizobium* symbiotic association involves the host exuding compounds that may selectively stimulate some rhizobial strains. Ayanaba et al. (1986) reported that root exudate treatment increased the nodule occupancy of some, but not all *B. japonicum* strains in their study. Homoserine, a compound excreted from the roots of pea seedlings, stimulated the growth of *R. leguminosarum* bv. viciae, but had little effect on rhizobia belonging to other cross inoculation groups (Egeraat, 1975). Virtanen and Miettinen (1953) first identified homoserine in germinating peas. It is specific to the genus *Pisum*, with little or no homoserine being found in other plants. Homoserine, which is hardly present in dry pea seeds, increased rapidly in the germinating seed. It was synthesized in the cotyledons and transported to the root system. From the 7th day of germination, approximately 70% of ninhydrin-positive compounds in the seedling root consisted of homoserine. At 24th day it was still 50% of the ninhydrin-positive compounds; and decreased thereafter (Egeraat, 1975).

Egeraat (1975) demonstrated that homoserine could be used as a carbon and nitrogen source by pea-nodulating biovar of Rhizobium *leguminosarum* i.e. biovar *viciae* but not by biovars *trifolii* and *phaseoli* or by *S. meliloti*. Thus, *R. leguminosarum* strains found commonly in pea nodules could catabolise homoserine, whereas strains from lentils or faba beans often could not (Hynes and O'Connell, 1990). A large increase in the number of rhizobia on the pea root surface was found to correlate with the liberation of significant amount of homoserine. This difference in the catabolic ability of different strains and species of rhizobia may be important in their adaptation to survive in the rhizospheres of different groups of host and non-host plants. Various plant-produced metabolites can thus play the role of “nutritional mediators”, thereby manipulating the microflora in the rhizosphere by favouring growth of strains likely to be beneficial to the plant.

The effect of homoserine on growth was found to be more around the lateral roots. This means that the homoserine released during the formation of the first lateral roots selectively stimulates the growth of *R. leguminosarum* bv. *viciae* when a mixture of *Rhizobium* strains is present in the surroundings of the young pea root. Even when the rhizobia of other cross-inoculation groups are not inhibited by homoserine, *R. leguminosarum* will accumulate due to its capacity to utilize homoserine as both the C and N-source (Egeraat, 1975).

Johnston (1988) showed that the ability of *R. leguminosarum* to utilize homoserine is encoded on a symbiotic plasmid. The symbiotic plasmid pRLlJI contains genes that allows *R. leguminosarum* bv. *viciae* to utilize homoserine. The genes involved in catabolism of homoserine are characterized in the reference strain of *Rhizobium leguminosarum* biovar *viciae* i.e. *Rlv*. 3841. pRL80071, which encodes a putative homoserine dehydrogenase (which catalyses conversion of homoserine to L-aspartate-semialdehyde), was specifically up-regulated in the pea rhizosphere in a metabolomics analysis performed by Ramachandran et al. (2011). Also elevated specifically in pea rhizospheres were pRL80026-30, which encode proteins belonging to the HAAT (hydrophobic amino acid transporter) ABC family. According to Ramachandran et al. (2011) although this transporter has been annotated as a LIV (leucine, isoleucine, valine) system, it could potentially transport one or more aromatic amino acid(s) or homoserine.

Vanderlinde et al. (2013) reported the identification of a gene cluster on plasmid pRL8JI that is required for homoserine utilization by *R. leguminosarum* bv. *viciae*. The genes were reported to be arranged as two divergently expressed predicted operons that were expressed on pea roots and induced by L-homoserine and pea root exudate. Mutation in pRL80083, a gene in one of the two predicted operons prevented utilization of homoserine as a sole carbon source and obliterated the mutant's ability to nodulate peas and lentils competitively.

2.1.3. Test for synthesis of acylated homoserine lactones (AHLs) :

Bacterial growth and survival depends on their ability to sense and adapt to changes in their environment. They have developed the ability to detect these changes and respond to them. These changes trigger cascades which alter gene expression, affecting behaviour and confer the bacteria the ability to survive in a variety of environments. One of the changing environments encountered by bacteria is cell or population density. Hence, along with the ability to detect 'normal' chemical cues, bacteria have evolved systems to 'detect' or 'count' the number of bacteria in their neighbourhood. This ability of bacteria to detect cell densities is known as "quorum sensing".

Prior to 1994, quorum sensing was commonly referred to as "autoinduction". The term "quorum sensing" was introduced by Dr. Steven Winans in 1994, in a review article on autoinduction in bacteria (Nealson et al., 1970, Fuqua et al., 1994).

Quorum sensing is an effective cell-cell communication system and is mediated through the synthesis of molecules called “autoinducers”, which can be referred to as prokaryotic hormones. Autoinducers are produced by the bacteria and released into the environment at a constant rate. As the population of bacteria increases, the concentration of the autoinducer in the environment increases. When the concentration of the autoinducer reaches a critical level, it triggers a cascade that results in expression of specific target genes. This allows for co-ordinated gene expression of the entire population and makes it behave like a single multicellular organism (Winzer et al., 2002, Waters and Bassler, 2005).

Bacteria can use quorum sensing to communicate both within and between species using species-specific and species-nonspecific autoinducers respectively. The bacteria respond differently to each type of autoinducer. In some conditions these signals allow the bacteria to grow synergistically and make use of complimentary metabolic processes whereas in others it allows the bacteria to detect competing species and restrict their growth or even destroy them. However, the final aim of all quorum sensing system remains the same - i.e., to count one another and regulate gene expression in response to cell number.

2.1.3.1. Quorum sensing in Gram-positive bacteria :

Most Gram-positive quorum sensing bacteria secrete peptide molecule as an autoinducer that is secreted by a dedicated ATP-binding cassette (ABC) transporter. The autoinducer detection and response occurs via a two-component adaptive response circuit (Kleerebezem et al., 1997). The two-component system is made up of a family of homologous proteins that exist in a wide variety of both Gram-negative and Gram-positive bacteria and enable bacteria to adapt to changing environment. The two-component systems relays sensory information by phosphorylation/dephosphorylation cascade consisting of a membrane-bound sensor kinase protein that initiates information transfer by autophosphorylation, and a response regulator protein, which following phosphotransfer from a cognate sensor kinase, typically controls transcription of downstream target gene (Stock et al., 1989, Parkinson, 1995, Novick and Geisinger, 2008).

2.1.3.2. Quorum sensing in Gram-negative bacteria :

Like Gram-positive bacteria, quorum sensing also occurs in Gram-negative bacteria but type of autoinducers, mechanism of secretion, detection of autoinducer, and the pattern of target gene expression is different. While Gram-positive bacteria largely use peptide signals and depend on cell-surface receptors to recognise them, Gram-negative bacteria mostly utilize soluble signals molecules known as 'acyl homoserine lactone' (AHL). Most AHLs are able to pass through the lipid bilayer and are therefore able to interact with cytoplasmic regulatory proteins. Thus, AHLs do not need the phosphorelay cascades that Gram-positive quorum sensing pathways use. A second type of autoinducer molecule used by Gram-negative bacteria is furanosyl borate diester (AI-2) (Federle, 2009). AI-2 is produced by a great variety of bacterial species, but there is discussion about the precise role of AI-2 as a signalling molecule. In addition to the typical Gram-negative QS system, some *Vibrio* species possess an AHL-responsive sensor kinases (e.g. LuxN) as part of a typical two-component signalling system similar to those found in Gram-positive bacteria (Bassler et al., 1994).

AHL-type quorum sensing involves the participation of two genes. The first encodes an AHL-synthase while the other gene acts as a regulatory gene whose activity is modified by binding to the AHL. The basic structure of AHL molecules is similar, consisting of a homoserine lactone (HSL) ring that is linked to an acyl side chain which can vary in type, length and degree of saturation. In addition, the third carbon atom of the lactone ring can carry a substituted hydrogen-, oxo- or hydroxyl- group. This variation, coupled with the ability to produce more than one type of AHLs, provide the bacteria with a mechanism for specificity allowing them to distinguish between inter-species and intra-species signals.

Three known protein families are known to synthesize AHL molecules viz. LuxI, LuxM / AinS / VanM and HdtS / Act. The LuxI-type AHL synthases constitute the largest family of AHL-synthesizing molecules and are found in more than 50 different species – including the α, β and γ-Proteobacteria (Gray and Garey, 2001). They catalyse the ligation of *S*-adenosylmethionine (SAM) with an acylated acyl-carrier protein from lipid metabolism (Parsek et al., 1999). The second family of AHL synthases is found only in *Vibrio* species and includes LuxM from *Vibrio harveyi,* AinS from *Vibrio fischeri* and VanM from *Vibrio anguillarum* (Milton et al., 2001). These proteins shows little sequence similarity with the LuxI-type synthases but seem to use the same reaction mechanism for the synthesis of AHLs (Hanzelka et al., 1999). The third family of AHL-synthases, comprises the HdtS in *Pseudomonas fluorescens* (Laue et al., 2000) and acts in the extreme acidophile *Acidithiobacillus ferrooxidans* (Rivas et al., 2007). HdtS and Act are related to the lysophosphatidic acid acyltransferase protein family, but the enzymatic mechanism of AHL synthesis by these enzymes is unknown.

The AHLs synthesized by the synthases are transported out of the cell to the surrounding environment through the cell membrane mainly by diffusion (Kaplan and Greenberg, 1985). Specialised efflux pumps may be used to transport long chain AHLs (Pearson et al., 1999). The concentration of AHL in the environment is decided by a number of factors. Although the most important deciding factor is the rate of AHL synthesis and diffusion out of the cell, the concentration is also influenced by the rate of AHL degradation. Nonenzymatic degradation occurs at an increased rate at high temperature and an alkaline pH (Byers et al., 2002). Three classes of AHL-degrading enzymes have also been identified : AHL lactonases – inactivate AHLs by hydrolysis of the ester bond of the HSL ring, AHL acylases – hydrolyse the AHL amide bond between the fatty acid and HSL moieties and AHL oxidoreductases – inactivate AHLs by a hydrolysis reaction of the 3-oxo group (Czajkowski and Jafra, 2009, Dong and Zhang, 2005).

The AHL molecules bind to proteins called AHL response regulators. These proteins belong to the LuxR-type response regulators and contain two conserved domains. The N-terminal domain contains a conserved region to which the AHLs bind. This binding leads to dimerization and activation of the regulators (Hanzelka and Greenberg, 1995). The C-terminal domain contains a conserved helix-turn-helix (HTH) motif, which allows activated AHL response regulators to bind to *cis*-acting DNA sequences (called ‘*lux* boxes’) and thus activate DNA transcription. Alternately, some LuxR-type regulators can bind to their target sequences in the absence of AHLs and block transcription. In such cases, the binding of AHLs to the regulatory protein reduces its DNA binding affinity and dissociates it from the DNA, allowing other transcription regulators to activate gene transcription (Horng et al., 2002, Minogue et al., 2002).

2.1.3.3. Quorum sensing in *R. leguminosarum* biovar *viciae* :

Rhizobia and host plants communicate by means of flavonoids and Nod-factors. However, the ability to synthesize *N*-acyl homoserine lactones (AHLs) is common to many plant-associated bacteria, including rhizobia. Quorum sensing in *Rhizobium* has been studied extensively (Downie and González, 2008, González and Marketon, 2003, Sanchez-Contreras et al., 2007, Wisniewski-Dye and Downie, 2002).

Most rhizobial species contain one or more AHL-dependent quorum-sensing systems affecting different aspects of the *Rhizobium*-legume symbiosis including, but not limited to, nodulation efficiency (Yang et al., 2009, Zheng et al., 2006), nodule formation (Zheng et al., 2006), symbiosome development (Daniels et al., 2002), exopolysaccharide production (Marketon and González, 2002), symbiotic plasmid transfer (Danino et al., 2003) and nitrogen fixation (Daniels et al., 2002). However, many rhizobia have been shown to be able to establish effective symbioses after mutation of their quorum sensing genes, indicating that their main role might be to optimization of the symbiosis and not its establishment.

Some plants synthesize AHL-mimicking compounds that could affect rhizobial communication, either positively or negatively, and thus influence the symbiosis (Degrassi et al., 2007, Sanchez-Contreras et al., 2007). Conversely, *Medicago truncatula* can perceive rhizobial AHL signals, inducing changes in gene expression in the plant (Mathesius et al., 2003).

Analysis of AHLs produced by strain A34 of *R. leguminosarum* bv. *viciae* led to the characterization of four LuxI-type AHL synthases (RhiI, CinI, RaiI, and TraI) and five LuxR-type regulators (RhiR, CinR, RaiR, TraR, and BisR). In this strain, the *cinI* and *cinR* genes are located on the chromosome and are on top of a regulatory cascade, inducing the production of RaiI-, RhiI- and TraI-made AHLs (Lithgow et al., 2000, Wisniewski-Dye and Downie, 2002). Mutation of *cinI* or *cinR* affects the expression of the other three AHL synthase genes.

CinR induces *cinI* expression allowing the production of its signal AHL molecule [*N*-(3-hydroxy-7-*cis*-tetradecenoyl)-L-homoserine lactone], which together with CinR activates *cinI* to form a positive feedback loop. The *cinI*-AHL influences the expression of *rhiI*, located on plasmid pRL1JI. RhiR, with the AHLs made by RhiI, induces *rhiI* and the *rhiABC* operon, which is involved with host interaction in the rhizosphere. The *cinI*-AHL activates BisR (a LuxR-type regulator) to induce *traR* and hence *traI* leading to the synthesis of several short chain AHLs which along with TraR allow the expression of the *trb* genes and is responsible for plasmid transfer.

cinI-AHL also affects a fourth group of quorum sensing genes located outside pRL1JI, called the *raiI-raiR* system which in turn synthesize several short acyl chain AHLs. *raiI* and *raiR* expression requires both *expR* (a LuxR-type regulator) and a small gene (*cinS*) to be co-transcribed with *cinI*. The genes regulated by RaiR have not yet been identified (Edwards et al., 2009).

2.1.3.4. Estimation of AHL signal molecules :

The signal molecules involved in quorum sensing are produced at very low concentrations. This makes it difficult to detect them by conventional methods. The separation, identification, purification and characterization of these molecules can be achieved by a variety of biophysical and biochemical techniques. However, the detection of these molecules in environment can only be carried out using a bioassay. The bioassays for detecting AHLs use mutants that cannot synthesize their own AHL and hence the response phenotype is expressed only on addition of exogenous AHL. The mutants used generally depend on the use of *lacZ* reporter fusions in *E. coli* or *A. tumefasciens* or on the induction or inhibition of synthesis of the purple pigment violacein in *Chromobacterium* v*iolaceum*.

Chromobacterium violaceum is a Gram-negative, facultatively anaerobic, rod-shaped bacterium living saprophytically in water and soil and is generally considered to be non-pathogenic (Kaufman et al., 1986, Ponte and Jenkins, 1992, Sneath, 1956). *C. violaceum* produces a characteristic purple pigment, violacein, which has antimicrobial characteristics. It acts as a bactericide (Lichstein and Van De Sand, 1946, Durán et al., 2012), a tumoricide (Duran and Menck, 2001, Duran et al., 2007), against soil amoebae, malarial parasites and trypanosomes (Duran et al., 2007, Lopes et al., 2009) and possesses anti-viral and immunomodulatory activities (May et al., 1991, Antonisamy and Ignacimuthu, 2010). Like many other antibiotics produced by bacteria, violacein synthesis is controlled by quorum sensing, and the AHL that regulates this production is N-hexanoyl-L-homoserine lactone (HHL) (McClean et al., 1997).

C. violaceum 026 (CV026) is a *C. violaceum* strain that is deficient in the production of HHL, and hence cannot synthesize violacein. The gene for the AHL synthase is disrupted via Tn5-transposon insertion. The transposon also carries a gene for kanamycin resistance. Therefore, CV026 is cultured on media containing kanamycin. This ensures that surviving bacteria still carries the transposon and thus is incapable of producing AHLs. When grown on plates, CV026 colonies are nearly white in colour whereas wild type *C. violaceum* (*CV* WT) colonies (with a functional HHL/violacein production), have a purple colour. However, CV026 has the ability to synthesize violacein in response to external supply of AHLs, with a subsequent change in colony colour from white to purple. This mutant functions as a simple alternative to the more complex *lux*-based reporter bioassays and is capable of detecting a similar range of AHLs (Càmara et al., 1998, McClean et al., 1997).

AHL compounds with C10 to C14 *N*-acyl chains are unable to induce violacein production. These long-side chains AHLs can be detected by their ability to inhibit violacein production when an activating AHL (e.g. HHL) is included into the assay medium (McClean et al., 1997). In this case, a white halo on a purple background constitutes a positive result. However, this reporter strain cannot detect any of the 3-hydroxy-derivatives and lacks sensitivity to most of the 3-oxo-derivatives (McClean et al., 1997, Cha et al., 1998). It has been reported that violacein synthesis is activated by cyclic dipeptides as well (Holden et al., 1999).

2.2. Materials and methods :

2.2.1. Collection of Wentworth Strains :

The Wentworth isolates or Wentworth strains were isolated from legumes on the University of York campus. The strains were isolated from the said site as a part of a population genomics project at the Department of Biology, University of York, as reported in Bailly *et al*. (2011).

The isolates used in this study were obtained from *Trifolium repens* (*R. leguminosarum* biovar *trifolii*) and *Vicia sativa* (*R. leguminosarum* biovar *viciae*). This study used 36 isolates from *T. repens* (hereafter referred to as the TRX strains) and 36 isolates from *V. sativa* (hereafter referred to as the VSX strains). The strains were sequenced using 454 sequencing technology and the sequence data from the contigs was used for the study.

2.2.2. Nodulation test :

- **Host plants :**

The *R. leguminosarum* bv. *trifolii* isolates were tested for their ability to nodulate their native host plant *Trifolium repens* while the *R*. *leguminosarum* bv. *viciae* isolates were tested for their ability to nodulate their native host plant, *Vicia cracca*.

- **Seed treatment :**

Seeds of *Vicia cracca* and *Trifolium repens* were surface sterilised to remove any bacteria on the coat. They were rinsed briefly in absolute ethanol before soaking in 3% sodium hypochlorite for five minutes. The seeds were then rinsed in seven changes of sterile de-ionised water and left to imbibe in water for four hours. The seeds were then washed in a further seven changes of water, drained and left to germinate in a covered glass beaker with a moist tissue paper at 28^{O}C.

A small number of seeds of each of the two plants were put onto TY agar and incubated to check for efficacy of seed sterilization. None of seeds showed any growth of micro-organisms indicating effective seed sterilization. (Somasegaran and Hoben, 2011).

- **Seed transfer :**

After germination, the seeds were transferred onto prepared agar slants containing Nitrogen-free minimal medium (Fahraeus, 1957). One plant was transferred per container. For the *V. cracca* seeds, 25ml of agar was used inside 50 ml Pyrex borosilicate glass tubes whereas the *T. repens* were grown in 15ml of agar in 30ml Corning polystyrene tubes. A groove was cut on the surface of the slants into which the emerging root of the seed (radicle) was pushed.

- **Growth of test cultures :**

The isolates were tested for purity by repeated subculturing on TY agar and checking macroscopic appearance of the colonies. A loopfull of the pure culture was inoculated into 25 ml of sterile TY broth in 50 ml Corning polypropylene tube and incubated on a rotary shaker at 150 rpm for 24 hours at $28^{O}C$. The culture was then centrifuged at 4000 rpm for 15 minutes at $20^{O}C$. The supernatant was discarded and the pellet washed with two changes of physiological saline. The bacteria were then suspended in physiological saline.

The density of the culture was adjusted to approximately 10^{8} cells/ml using standard McFarland Nephlometer tubes by visual comparison.

- **Seed inoculation :**

100 µl of the bacterial suspension was injected into the groove made on the agar surface to ensure that the bacteria came into contact with the root of the seedling. 4 replicates were used per bacterial strain.

- **Controls :**

Four positive and four negative control replicates were also set up. The negative controls were not inoculated with any rhizobial cultures. The plants included in the negative controls were watered with the N-free minimal liquid medium. The plants included in the positive control were not inoculated with any rhizobial cultures but were watered with liquid minimal medium containing 0.05% KNO_3 to provide them with a readily available source of nitrogen.

- **Incubation :**

The tubes were plugged with cotton wool to reduce the possibility of contamination, and when the plants had grown up to the plug it was replaced with plastic film with a slit in it to allow plant growth. The plants were grown in a CT room in a cycle of 16 hours of light at 28^{o}C and 8 hours of darkness at 18^{o}C and watered with nitrogen free medium every three days.

- **Harvesting :**

After 10 weeks of growth, when the plants had reached maturity, the tubes were half filled with warm water and placed in water baths at 90^{O}C to dissolve the agar. The plants were then carefully pulled out of the gel and rinsed in hot water to remove any traces of agar that remained. The plants were then air dried overnight on paper towels and then weighed.

2.2.3. Homoserine utilization test :

- **Growth of test cultures :**

The isolates were tested for purity by repeated subculturing on TY agar and checking the macroscopic appearance of colonies. A loopfull of the pure culture was inoculated into 25 ml of sterile TY broth in 50 ml Corning polypropylene tube and incubated on a rotary shaker at 150 rpm for 24 hours at 28^{O}C. The culture was then centrifuged at 4000 rpm for 15 minutes at 20^{O}C. The supernatant was discarded and the pellet washed with two changes of physiological saline. The bacterial pellet obtained at the end of the centrifugation was then suspended in homoserine medium. The density of the culture was adjusted to A_{610} = 0.1 read on an ELISA reader (Thermomax, Thermo Scientific) using sterile, uninoculated homoserine medium as blank.

- **Growth medium for homoserine utilization :**

The composition of the medium used for testing homoserine was as described by van Egeraat (1975) . The composition of the medium is as follows : $K_2HPO_4.3H_2O$ – 1 g, $MgSO_4.7H_2O$ – 200 mg, $CaCl_2.H_2O$ – 40 mg, $FeCl_3.6H_2O$ – 0.2 mg, $MnSO_4.H_2O$ – 0.2 mg, $ZnSO_4.7H_2O$ – 0.2 mg,

$CuSO_4.5H_2O$ – 0.02 mg, $CoCl_2.6H_2O$ – 0.002 mg, H_3BO_3 – 0.2 mg, $Na_2MoO_4.2H_2O$ – 0.2 mg, Biotin – 0.05 mg, Thiamine – 0.1 mg, Methionine – 10 mg, Uracil – 10 mg, Homoserine – 1 g, Water – to 1 litre. The pH of the medium was adjusted to 6.8.

The medium was sterilized by autoclaving. 5 ml of the sterile medium was dispensed in 20 ml sterile glass Dram bottles.

- **Inoculation of homoserine medium :**

50 µl of the inoculum (A_{610} = 0.1) was inoculated into the 20 ml Dram bottles containing 5 ml of homoserine utilization medium. The culture was thoroughly mixed with the medium. 200 µl of the inoculated medium was transferred to the wells of an empty microtitre dish (Corning Costar 3595) kept on ice for determining initial absorbance of the test strains in the inoculated medium. 2 replicates were made per strain. When all the Dram bottles were inoculated, the initial absorbance of aliquots was read on an ELISA reader (Thermomax, Thermo Scientific) at 610 nm.

- **Incubation :**

The mouth of the Dram bottles was covered with Parafilm to maintain sterility while allowing gaseous exchange and transferred to 2 litre glass beakers and stacked vertically in layers and incubated on a shaker at 150 rpm for a week.

- **Harvesting :**

The Dram bottles were removed from the shaker after 7 days. 200 µl of the bacterial suspension was transferred to the wells of a microtiter plate on ice. When all the samples were transferred, the absorbance of aliquots was read on an ELISA reader (Thermomax, Thermo Scientific) at 610 nm.

2.2.4. Test for AHL synthesis :

- **Growth of test cultures :**

The isolates were tested for purity by repeated subculturing and loopfull of the pure culture was inoculated into 5 ml of sterile TY broth in a 15 ml

Corning polypropylene tube and incubated on a rotary shaker at 150 rpm for 24 hours at 28^{O}C. The culture was then centrifuged at 4000 rpm for 15 minutes at 20^{O}C. The supernatant was discarded and the pellet washed with two changes of physiological saline. The bacteria were then suspended in sterile TY broth. The density of the culture was adjusted to A_{610} = 0.1 read on an ELISA reader (Thermomax, Thermo Scientific) using sterile, uninoculated TY broth as blank. 1 ml of the density-adjusted culture was added to 9 ml of sterile TY broth in a 15 ml Corning polypropylene tube and incubated on a rotary shaker at 150 rpm for 72 hours at 28^{O}C. 2 ml of the culture was then transferred to a sterile 2 ml Eppendorf tube and centrifuged at 13,000 rpm for 10 minutes. The clear supernatant was transferred to a new sterile 2 ml Eppendorf tube and stored at -20^{O}C to be used to detect presence of AHL.

- **Growth of reporter culture :**

The reporter culture *Chromobacterium violaceum* 026 (CV026) was transferred to sterile LB-kanamycin plate to check for purity. A loopfull of pure culture was inoculated into 25 ml of sterile LB broth containing Kanamycin (50 µg/ml) in a 50 ml tube and incubated at 28^{O}C on a shaker for 24 hours. The culture was centrifuged at 4000 rpm for 15 minutes. The supernatant was discarded and the pellet washed with two changes of physiological saline. The bacteria were then suspended in LB broth. The density of the culture was adjusted to A_{610} = 1.0 read on an ELISA reader (Thermomax, Thermo Scientific) using sterile LB broth as blank.

- **Test proper :** Modified from Blosser and Gray (2000), Hornung et al. (2013) :

To 750 µl of 2X LB broth containing 100 µg/ml Kanamycin, 750 µl of rhizobial culture supernatant was added followed by addition of 15 µl of the reporter culture. The tubes were vortexed briefly. 200 µl of the inoculated medium was transferred to the wells of a microtiter plate on ice. When all the samples were transferred, the initial absorbance of aliquots was read on an ELISA reader (Thermomax, Thermo Scientific) at 660 nm using sterile uninoculated 1X LB broth as blank. The remaining inoculated medium in tubes was then incubated stationary in an upright position at 28^{O}C for 72 hours. 750 µl of sterile TY broth was added to a sterile tube and used as the negative control for measuring absorbance.

After 72 hours, the tubes were vortexed for 30 seconds. 200 µl of the medium was transferred to a fresh sterile Eppendorf tube. 200 µl of sterile 10% SDS was added and mixed with the culture by vortexing for 5 seconds. The culture was then allowed to rest for 5 min. Violacein was then extracted from the lysate by adding 900 µl of water-saturated butanol, vortexing for 5 seconds and centrifuging at 13,000 rpm for 5 minutes in a microcentrifuge. The butanol (upper) phase containing the violacein was collected and its absorbance measured at a wavelength of 590 nm. The optical density of the inoculum measured at 660 nm . The cells' response to additions of spent medium extract was calculated as the ratio of the absorbance of the butanol extract versus the bioassay culture density, multiplied by 1000 : (A_{585} nm/O.D.$_{660}$ nm) x 1000. The values obtained were referred to as violacein units.

2.3. Results :

2.3.1. Grouping strains for analysis :

The reads from the 454 sequencing of Wentworth strains were assembled into contigs. A SplitsTree analysis based on the pairwise divergence between strains based on all reads that mapped to the 3841 genome using the 454 Reference Mapper was constructed by Mr. Ryan Lower from our group. This tree was used for grouping the strains for analysis.

The SplitsTree suggested the possible existence of five cryptic species (labelled 5A through 5E). This observation was used to check for metabolic differences within the five cryptic species. A large number of isolates comprised the cryptic species 5C which was more divergent than the other groups. Hence, another line of investigation was to check for differences between cryptic species 5C and the remaining four groups. In this analysis, the cryptic species 5C was labelled as Cryptic 2B, while the remaining four cryptic species were collectively labelled as Cryptic 2A (See figures 2.1a, 2.1b and 2.1c).

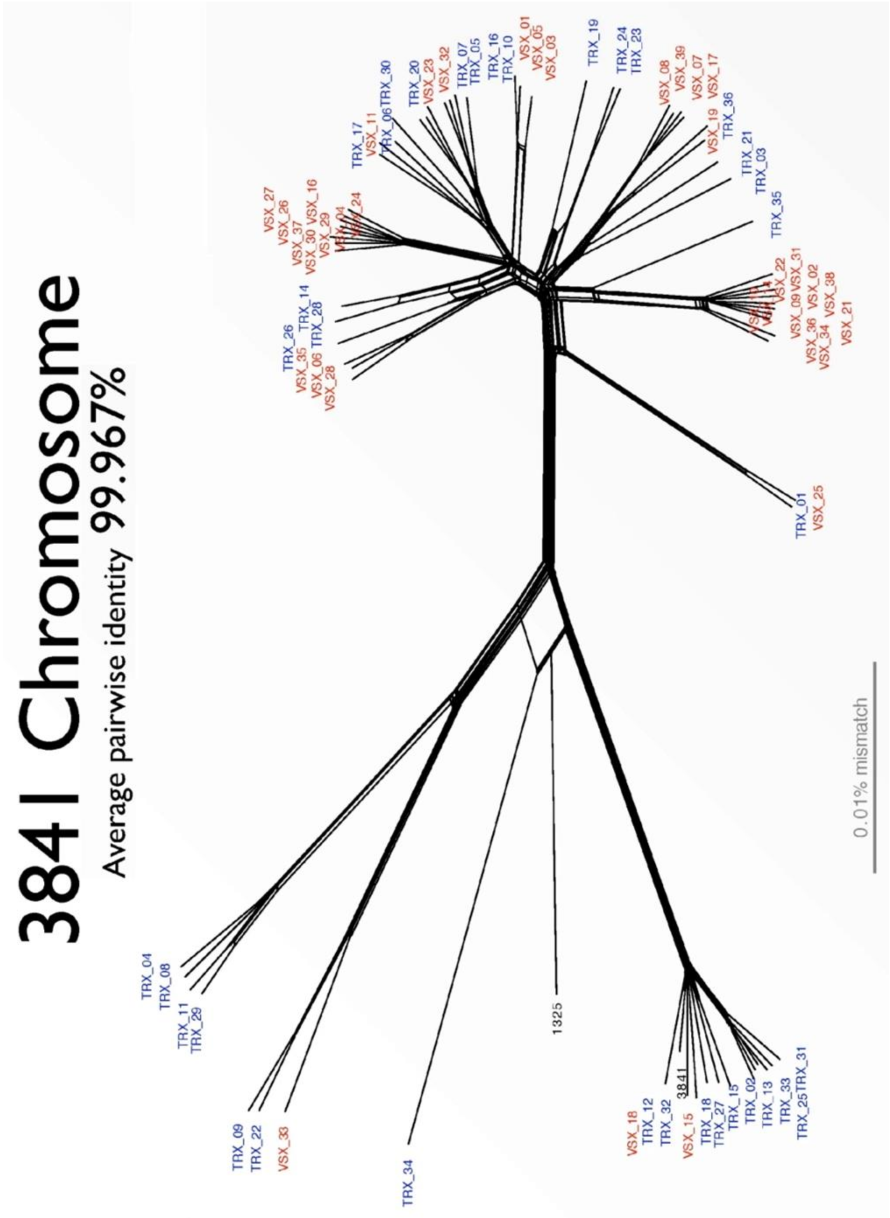
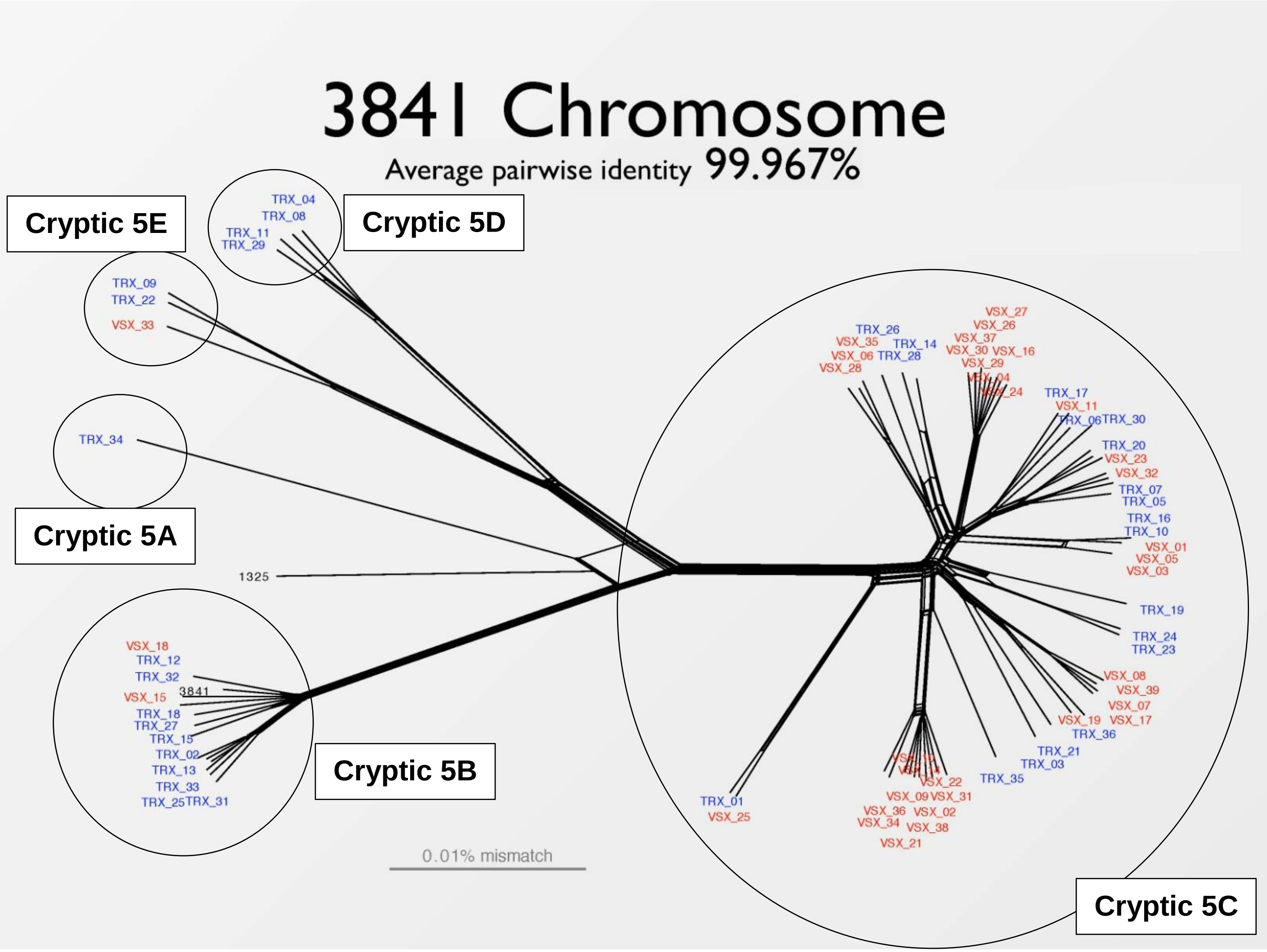


Figure 2.1a. SplitsTree analysis based on the pairwise divergence between strains based on all reads that mapped to the 3841 genome using the 454 Reference Mapper was constructed by Mr. Ryan Lower. The TRX strains are shown in blue text and the viciae strains are shown in red text.



3841 Chromosome

Average pairwise identity 99.967%

Figure 2.1b. The grouping of strains into five cryptic species for use in statistical analysis.

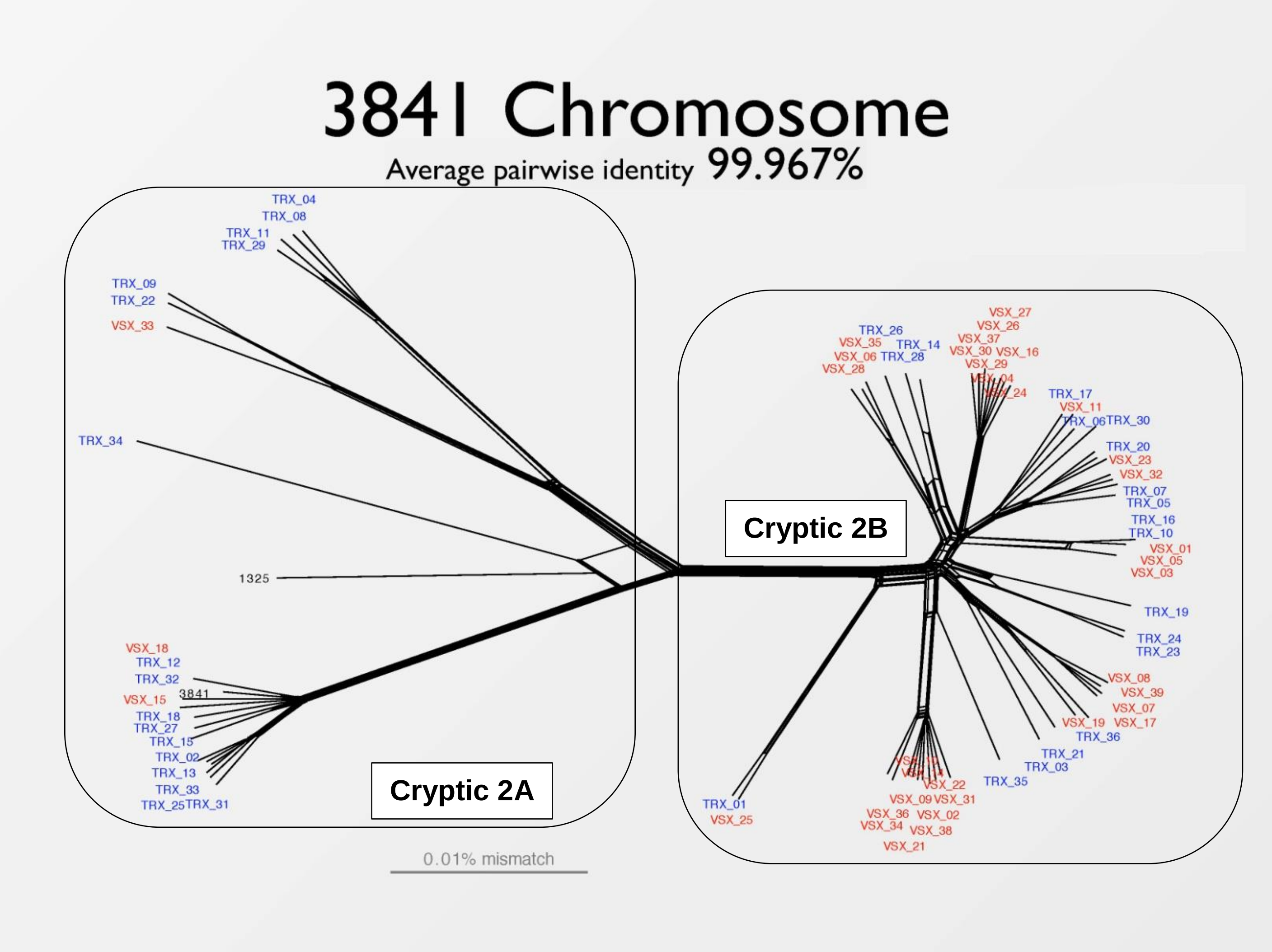


Figure 2.1c. The grouping of strains into two cryptic species for use in statistical analysis.

2.3.2. Nodulation test :

The *Rhizobium leguminosarum* biovar *trifolii* and *viciae* strains were tested for their ability to nodulate their native host plants *Trifolium repens* and *Vicia cracca*. All inoculated plants formed root nodules. The dry weights of the plants from nodulation test are shown in Table 2.1 and Figure 2.2 and were analysed using statistical tests to check if they show any difference in weights amongst the three groups viz. test, positive control and negative control.

Table 2.1. Dry weight data of plants used in nodulation test in grams.

Plant	Strain	Replicate1	Replicate2	Replicate3	Replicate4	Mean
Trifolium repens	TRX01	0.0484	0.0563	0.0316	0.0543	0.0477
	TRX02	0.0492	0.0449	0.0528	0.0554	0.0506
	TRX03	0.0435	0.0363	0.0533	0.0478	0.0452
	TRX04	0.0434	0.0373	0.0365	0.0294	0.0367
	TRX05	0.0683	0.0706	0.0632	0.0633	0.0664
	TRX06	0.0614	0.0593	0.0606	0.0487	0.0575
	TRX07	0.0618	0.0648	0.0265	0.0505	0.0509
	TRX08	0.0695	0.0653	0.0715	0.0605	0.0667
	TRX09	0.0746	0.0599	0.0686	0.0567	0.0650
	TRX10	0.0789	0.0627	0.0603	0.0528	0.0637
	TRX11	0.0767	0.0756	0.0693	0.0737	0.0738
	TRX12	0.0701	0.0812	0.0600	0.0522	0.0659
	TRX13	0.0399	0.0663	0.0447	0.0754	0.0566
	TRX14	0.0684	0.0535	0.0932	0.0862	0.0753
	TRX15	0.0728	0.0850	0.0715	0.0711	0.0751
	TRX16	0.0801	0.0820	0.0561	0.0701	0.0721
	TRX17	0.0631	0.0636	0.0603	0.0760	0.0658
	TRX18	0.0794	0.0932	0.0672	0.0655	0.0763
	TRX19	0.0846	0.0624	0.0610	0.0986	0.0767

	TRX20	0.0776	0.0661	0.0592	0.0637	0.0667
	TRX21	0.0603	0.0572	0.0597	0.0968	0.0685
	TRX22	0.0691	0.0704	0.0497	0.0456	0.0587
	TRX23	0.0643	0.0591	0.0581	0.0428	0.0561
	TRX24	0.0288	0.0633	0.0500	0.0463	0.0471
	TRX25	0.0404	0.0408	0.0550	0.0639	0.0500
	TRX26	0.0137	0.0272	0.0617	0.0470	0.0374
	TRX27	0.0626	0.0323	0.0314	0.0500	0.0441
	TRX28	0.0574	0.0566	0.0301	0.0561	0.0501
	TRX29	0.0309	0.0385	0.0426	0.0471	0.0398
	TRX30	0.0608	0.0474	0.0551	0.0423	0.0514
	TRX31	0.0611	0.0371	0.0316	0.0453	0.0438
	TRX32	0.0392	0.0458	0.0626	0.0549	0.0506
	TRX33	0.0391	0.0425	0.0574	0.0410	0.0450
	TRX34	0.0418	0.0655	0.0606	0.0469	0.0537
	TRX35	0.0793	0.0673	0.1143	0.0721	0.0833
	TRX36	0.0628	0.0393	0.0542	0.0425	0.0497
	TRX+VE	0.0653	0.0537	0.0861	0.0561	0.0653
	TRX-VE	0.0368	0.0364	0.0285	0.0339	0.0339
Vicia cracca	VSX01	0.0579	0.0779	0.0805	0.0988	0.0788
	VSX02	0.1045	0.0983	0.0842	0.1288	0.1039
	VSX03	0.0649	0.0349	0.0411	0.0750	0.0540
	VSX04	0.0825	0.0818	0.0748	0.0865	0.0814
	VSX05	0.0962	0.1113	0.0820	0.0890	0.0946
	VSX06	0.0915	0.0985	0.0340	0.0822	0.0766
	VSX07	0.0903	0.0675	0.1125	0.0814	0.0879
	VSX08	0.0911	0.0430	0.0729	0.1130	0.0800
	VSX09	0.1312	0.0883	0.0902	0.1156	0.1063
	VSX10	0.0695	0.0182	0.1134	0.1108	0.0780
	VSX11	0.0817	0.0730	0.1110	0.0843	0.0875

	VSX14	0.0461	0.0173	0.1004	0.1031	0.0667
	VSX15	0.0484	0.1179	0.0996	0.0929	0.0897
	VSX16	0.0990	0.0909	0.0688	0.1011	0.0900
	VSX17	0.0943	0.0738	0.0875	0.0978	0.0884
	VSX18	0.0601	0.0475	0.0561	0.0763	0.0600
	VSX19	0.0779	0.1067	0.0701	0.0470	0.0754
	VSX21	0.0787	0.0828	0.0847	0.0838	0.0825
	VSX22	0.0148	0.0325	0.0251	0.1239	0.0491
	VSX23	0.0490	0.0371	0.1014	0.0064	0.0485
	VSX24	0.0607	0.0787	0.0515	0.0599	0.0627
	VSX25	0.0605	0.0503	0.0889	0.0863	0.0715
	VSX26	0.0330	0.1431	0.0456	0.1074	0.0823
	VSX27	0.0918	0.1164	0.1425	0.0624	0.1033
	VSX28	0.0881	0.1061	0.0945	0.0798	0.0921
	VSX29	0.1490	0.0421	0.0547	0.1236	0.0924
	VSX30	0.0777	0.0912	0.0445	0.0833	0.0742
	VSX31	0.1182	0.0936	0.0881	0.0697	0.0924
	VSX32	0.0975	0.0637	0.1217	0.0779	0.0902
	VSX33	0.0587	0.0433	0.0648	0.0551	0.0555
	VSX34	0.1172	0.1406	0.0807	0.1409	0.1199
	VSX35	0.1085	0.1078	0.0707	0.1224	0.1024
	VSX36	0.0898	0.0874	0.1355	0.0978	0.1026
	VSX37	0.0870	0.1177	0.0671	0.1088	0.0952
	VSX38	0.0909	0.0555	0.1009	0.0600	0.0768
	VSX39	0.0436	0.0863	0.1093	0.0592	0.0746
	VSX+VE	0.1089	0.1234	0.0973	0.0827	0.1031
	VSX-VE	0.0864	0.1096	0.0716	0.1078	0.0939

Four replicates were used for each strain along with positive (added nitrate) and negative (no inoculation) controls.

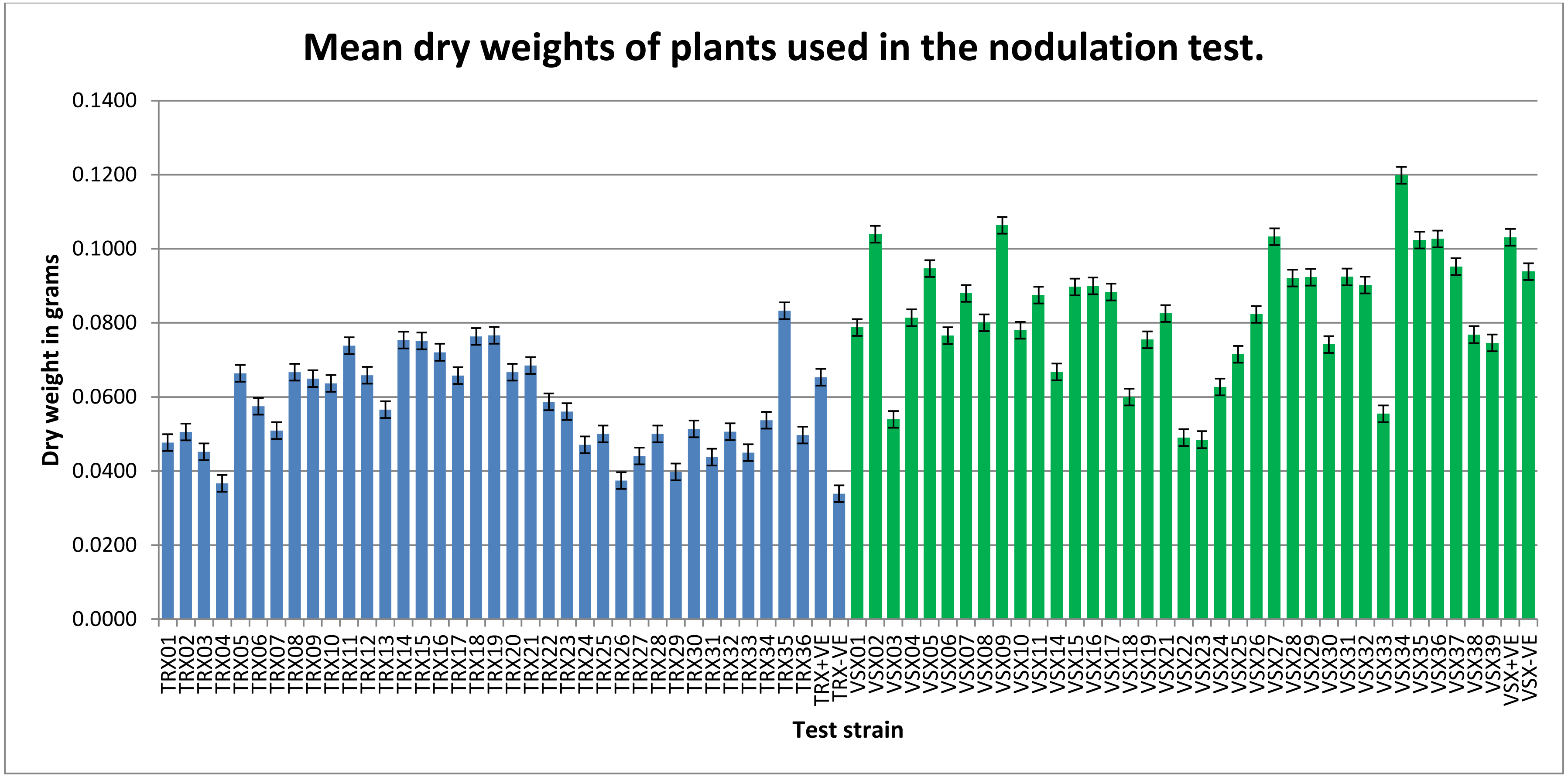


Figure 2.2. Figure showing the mean dry weight of plants for each of the 72 strains and the controls. The error bars represent a standard error of means. Data for *T. repens* is shown in blue while the data for *V. cracca* is shown in green.

2.3.2.1. Analysis of dry weight of *Trifolium repens* :

- **Test of normality :**

The data was tested for normality of distribution in order to decide the test to be used for subsequent analysis. Since the sample size is less than 2000, the Shapiro-Wilk test was used to check the normality of distribution at level of significance (α) = 0.05 using IBM SPSS V.21.

Table 2.2. Tests of Normality (Test plants)

Group	Shapiro-Wilk		
	Statistic	Degrees of freedom	Sig.
TRX test plants	.989	144	.339
Positive control	.869	4	.293
Negative control	.855	4	.242

Results of Shapiro-Wilk test to check normality distribution of *T. repens* dry weight data.

The Shapiro-Wilk test showed that the data for the dry weight of the plants was normally distributed at α = 0.05.

- **Analysis of variance (ANOVA) :**

Since the data for the dry weight of the plants was normally distributed, the data was analysed using one-way ANOVA.

ANOVA is a parametric statistical test to check whether or not the means of several groups are equal. The analysis was carried out to check for difference between the test plants inoculated with the test strains, the uninoculated positive control plants (supplemented with nitrate) and uninoculated negative control plants (not supplemented with nitrate). The analysis was performed using the ‘IBM SPSS Statistics’ software package (Version 21) from IBM Corporation.

The null hypothesis (H_O) for the analysis was : There is no significant difference between the dry weights of the inoculated test plants, the positive and the negative controls.

The alternate hypothesis (H_{α}) for the analysis was : There is a significant difference between the dry weights of the inoculated test plants, the positive and the negative controls.

Table 2.3. ANOVA to test Between-Subjects Effects

Dependent Variable : Weight

Source	Type III Sum of Squares	Degrees of freedom	Mean Square	F	Sig.
Corrected Model	.002[a]	2	.001	4.812	.009
Intercept	.049	1	.049	188.041	.000
Group	.002	2	.001	4.812	.009
Error	.039	149	.00026		
Total	.543	152			
Corrected Total	.041	151			

a. R squared = .061 (Adjusted R Squared = .048)

The results of the ANOVA test showing the difference between the means of the three groups included in the test are show in Figure 2.3 below.

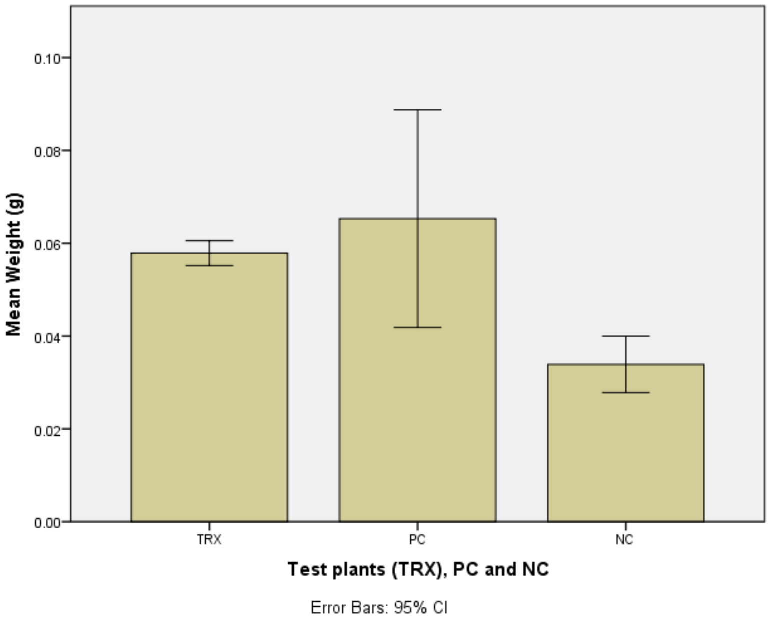


Figure 2.3. Difference in the mean weight of the dry weight of the *T. repens* plants used in the nodulation test as obtained from the ANOVA test.

The results of the analysis indicates that there is a significant difference in the means, *F* (2, *N* = 152) = 4.812, *p* = .009.

Since *p*-value = 0.009 ≤ 0.05 =α, we reject the null hypothesis and accept the alternate hypothesis.

Conclusion in words : At the α=0.05 level of significance, there exists enough evidence to conclude that there is significant difference in the mean test scores among the TRX test strains, positive control and negative control.

- **Post-hoc test : LSD**

The analysis shows that there is a significant difference between the means of. However, it does not indicate which group differs. 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.

LSD was the first statistical pairwise comparison technique developed by Fisher in 1935 and can only be used if the ANOVA F-statistic has a significant value. The main idea of the LSD is to compute the smallest significant difference (i.e. the LSD) between two means as if these means were the only means to be compared (i.e. with a *t*-test) and to declare significant any difference larger than the LSD.

Table 2.4. Multiple Comparisons : LSD

Dependent Variable: Weight

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
TRX	PC	-.0074	.0081	.364	-.0235	.0086
	NC	.0239*	.0081	.004	.0078	.0400
PC	TRX	.0074	.0081	.364	-.0086	.0235
	NC	.0314*	.0113	.007	.0089	.0538
NC	TRX	-.0239*	.0081	.004	-.0400	-.0078
	PC	-.0314*	.0113	.007	-.0538	-.0089

Based on observed means.

The error term is Mean Square (Error) = .00026.

* = The mean difference is significant at that value.

The results of the post-hoc LSD test indicate that there is significant difference between the test group and the negative control and the positive and negative control. There is no significant difference between the positive controls and the plants in the test group.

2.3.2.2. Analysis of dry weight of *Vicia cracca* :

The *Rhizobium leguminosarum* biovar *viciae* strains were tested for their ability to nodulate the host plant, *Vicia cracca*.

- **Test of normality :**

The dry weight data for *Vicia cracca* was tested for normality using the Shapiro-Wilk test using the same parameters used in the analysis of the dry weight data for *T. repens*.

Table 2.5. Tests of Normality (Test plants)

Group	Shapiro-Wilk	
	Degrees of freedom	Sig.
VSX test plants	144	.543
Positive control	4	1.000
Negative control	4	.916

Results of Shapiro-Wilk test to check normality distribution of *V. cracca* dry weight data.

The test showed that dry weight data for *V. cracca* is normally distributed.

- **Analysis of variance (ANOVA) :**

Since the data for the dry weight of the plants was normally distributed, the data was analysed using one-way ANOVA as for the TRX plants.

The null hypothesis (H_O) for the analysis was : There is no significant difference between the dry weights of the inoculated test plants, the positive and the negative controls.

The alternate hypothesis (H_α) for the analysis was : There is a significant difference between the dry weights of the inoculated test plants, the positive and the negative controls.

Table 2.6. ANOVA to test Between-Subjects Effects

Dependent Variable: Weight

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.002[a]	2	.001	1.317	.271
Intercept	.154	1	.154	191.160	.000
Strain3	.002	2	.001	1.317	.271
Error	.120	149	.001		
Total	1.176	152			
Corrected Total	.122	151			

a. R Squared = .017 (Adjusted R Squared = .004)

The results of the ANOVA test showing the difference between the means of the three groups included in the test are show in Figure 2.4 below.

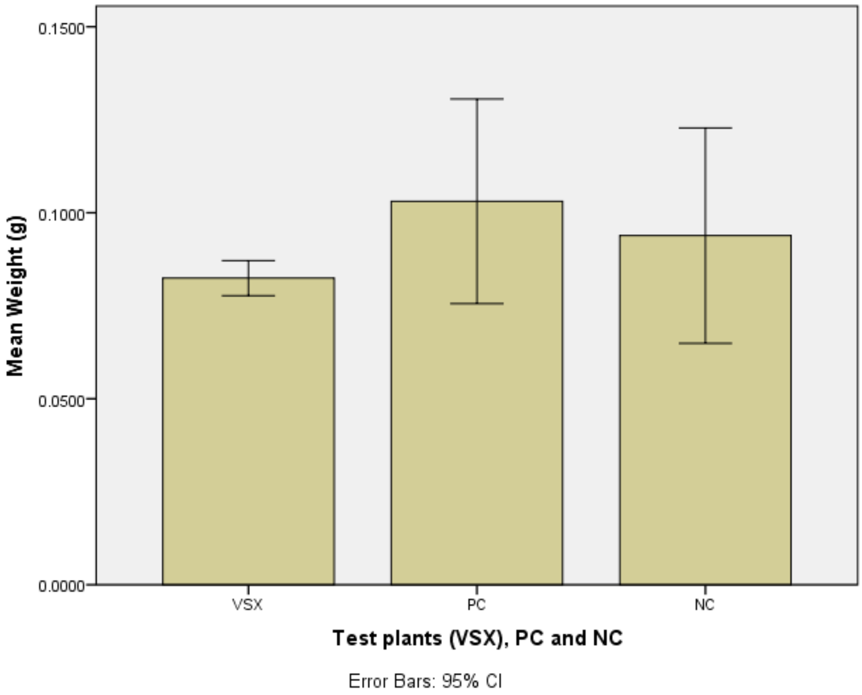


Figure 2.4. Difference in the mean weight of the dry weight of the *V. cracca* plants used in the nodulation test as obtained from the ANOVA test.

The results of the analysis indicates that there is no a significant difference in the means, (ANOVA: F = 1.317, d.f. = 2, 149, *p* = 0.271)

Since *p*-value = 0.271 ≥ 0.05 = α, we do not reject the null hypothesis.

Conclusion in words : At the α = 0.05 level of significance, there exists enough evidence to conclude that there is no difference in the mean test scores among the VSX, positive and negative.

Since there is no significant difference in the mean test scores of the plant weights, no follow-up post-hoc test was conducted.

All the 72 isolates of *R. leguminosarum* formed large and pink effective nodules on the primary and upper lateral roots as described in Nitrogen Fixation for Tropical Agricultural Legumes Project and Food (1984). The statistical tests and the subsequent post-hoc tests employed in the analysis indicate that in *T. repens*, there was a significant difference between the plants included in the negative control and those included in the other two groups (viz. test plants and positive control). Theoretically, all other conditions being constant, a plant with more access to nitrogen should grow better than a plant with lesser access. The plants which were included as negative control showed the least growth which indicates that a lack of nitrogen stunts plant growth. The inoculated plants grew as well as those provided with nitrate, indicating that the symbiosis was effective in meeting their needs for nitrogen.

Contrary to expectation, in the *V. cracca* plants inoculated with the VSX strains, no significant difference was observed between the test plants and the plants included in the positive and negative controls. Moreover, the *V. cracca* plants included in the controls did not show the presence of any nodules indicating absence of cross contamination.

The aim of this study was to validate the strains and identify them as root-nodulating rhizobia. Since all the strains formed effective nodules, the strains could now be said to be validated for use in further studies.

2.3.3. Homoserine test :

The 72 isolates from Wentworth were tested for their ability to utilize homoserine as the sole carbon source. The growth data are shown in Table 2.7 and Figure 2.5.

Table 2.7. Growth data for Homoserine test

Strain	% increase in growth (Rep01)	% increase in growth (Rep02)	Average % increase in growth
TRX01	24.00	64.00	44.00
TRX02	-16.67	0.00	-8.33
TRX03	271.43	300.00	285.71
TRX04	-6.25	-6.25	-6.25
TRX05	83.33	58.33	70.83
TRX06	53.33	46.67	50.00
TRX07	15.38	0.00	7.69
TRX08	40.00	40.00	40.00
TRX09	42.86	35.71	39.29
TRX10	27.27	-9.09	9.09
TRX11	41.18	17.65	29.41
TRX12	106.25	106.25	106.25
TRX13	5.56	22.22	13.89
TRX14	236.36	190.91	213.64
TRX15	42.11	26.32	34.21
TRX16	100.00	111.76	105.88
TRX17	87.50	87.50	87.50
TRX18	93.75	93.75	93.75
TRX19	68.75	68.75	68.75
TRX20	100.00	85.71	92.86
TRX21	236.36	190.91	213.64
TRX22	63.64	18.18	40.91
TRX23	108.33	83.33	95.83
TRX24	53.85	38.46	46.15
TRX25	150.00	90.00	120.00
TRX26	0.00	-23.08	-11.54
TRX27	23.08	0.00	11.54
TRX28	42.86	42.86	42.86
TRX29	0.00	-60.00	-30.00
TRX30	80.00	73.33	76.67

TRX31	6.67	0.00	3.33
TRX32	46.67	40.00	43.33
TRX33	-5.88	5.88	0.00
TRX34	14.29	38.10	26.19
TRX35	50.00	35.71	42.86
TRX36	200.00	200.00	200.00
VSX01	280.00	305.00	292.50
VSX02	69.23	46.15	57.69
VSX03	250.00	208.33	229.17
VSX04	76.47	82.35	79.41
VSX05	350.00	350.00	350.00
VSX06	123.53	129.41	126.47
VSX07	128.57	114.29	121.43
VSX08	706.25	737.50	721.88
VSX09	35.29	35.29	35.29
VSX10	715.38	769.23	742.31
VSX11	204.55	236.36	220.45
VSX14	15.00	30.00	22.50
VSX15	33.33	26.67	30.00
VSX16	178.26	204.35	191.30
VSX17	223.08	207.69	215.38
VSX18	57.89	73.68	65.79
VSX19	593.33	586.67	590.00
VSX21	18.75	25.00	21.88
VSX22	13.33	6.67	10.00
VSX23	373.33	373.33	373.33
VSX24	30.77	92.31	61.54
VSX25	352.94	364.71	358.82
VSX26	90.00	110.00	100.00
VSX27	92.86	85.71	89.29
VSX28	173.33	166.67	170.00
VSX29	35.71	21.43	28.57

VSX30	56.25	56.25	56.25
VSX31	14.29	0.00	7.14
VSX32	184.21	205.26	194.74
VSX33	47.37	68.42	57.89
VSX34	63.64	100.00	81.82
VSX35	93.33	93.33	93.33
VSX36	58.33	33.33	45.83
VSX37	86.67	86.67	86.67
VSX38	31.25	37.50	34.38
VSX39	754.55	718.18	736.36

The table shows the percentage increase in test strains grown in Homoserine medium calculated from difference in the initial and final absorbance of the inoculated medium.

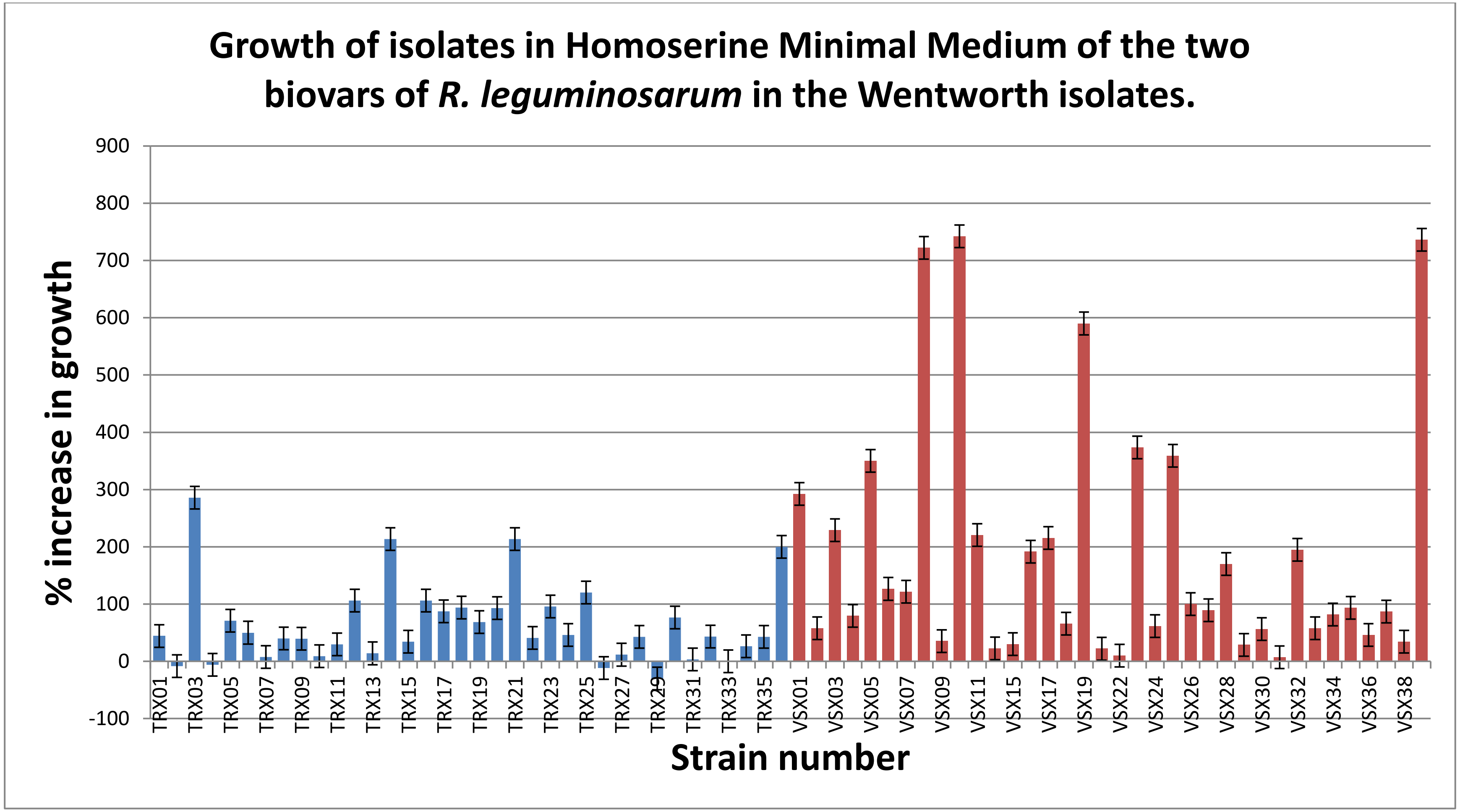


Figure 2.5a. Graphical representation of the Homoserine growth experiment. TRX strains are shown in blue and VSX in red. Error bars indicate standard error of means.

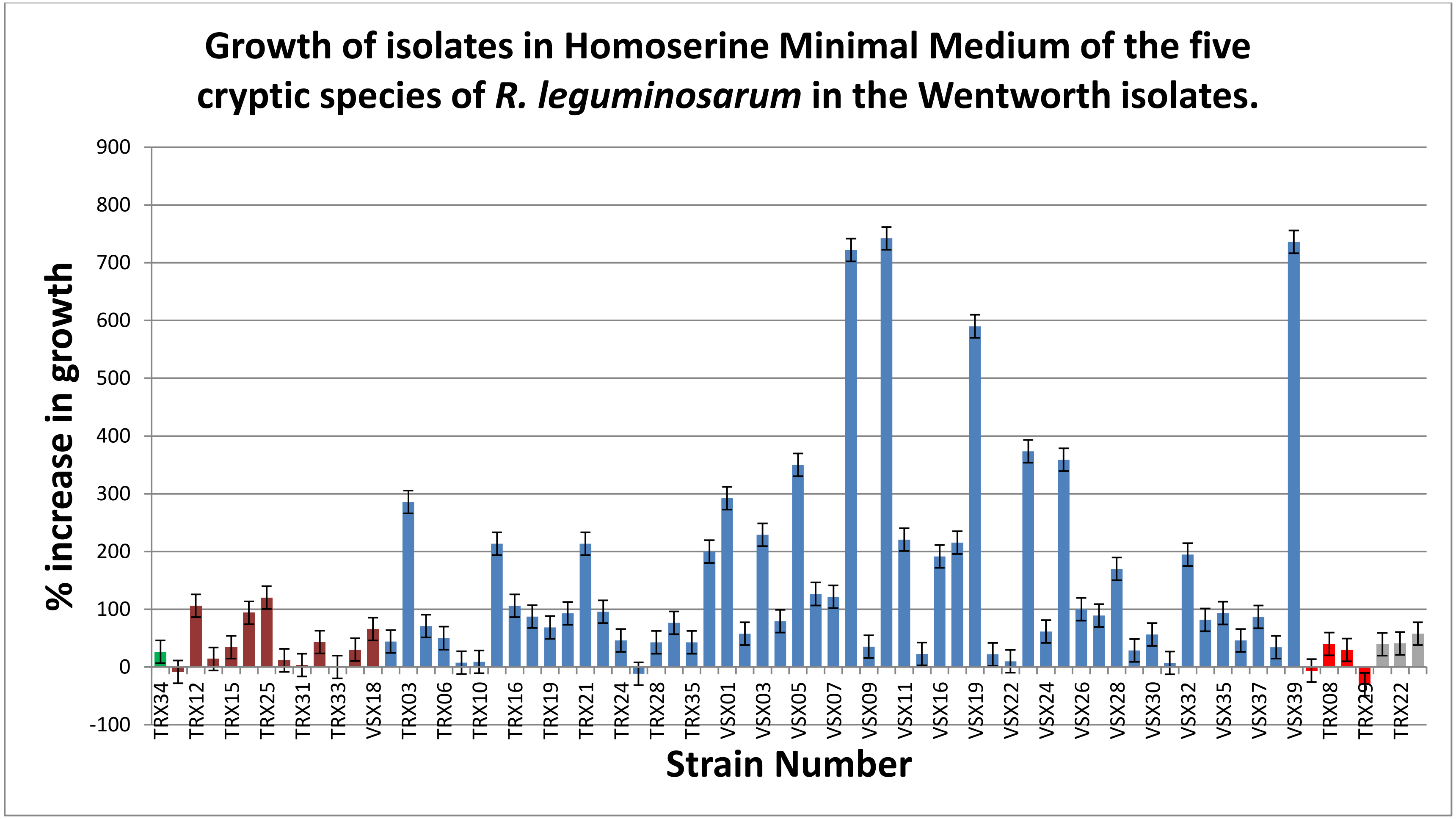


Figure 2.5b. Graphical representation of the Homoserine growth experiment. Strains are arranged as the five cryptic species 5A, 5B, 5C, 5D and 5E shown using green, brown, blue, red and grey colours respectively. Error bars indicate standard error of means.

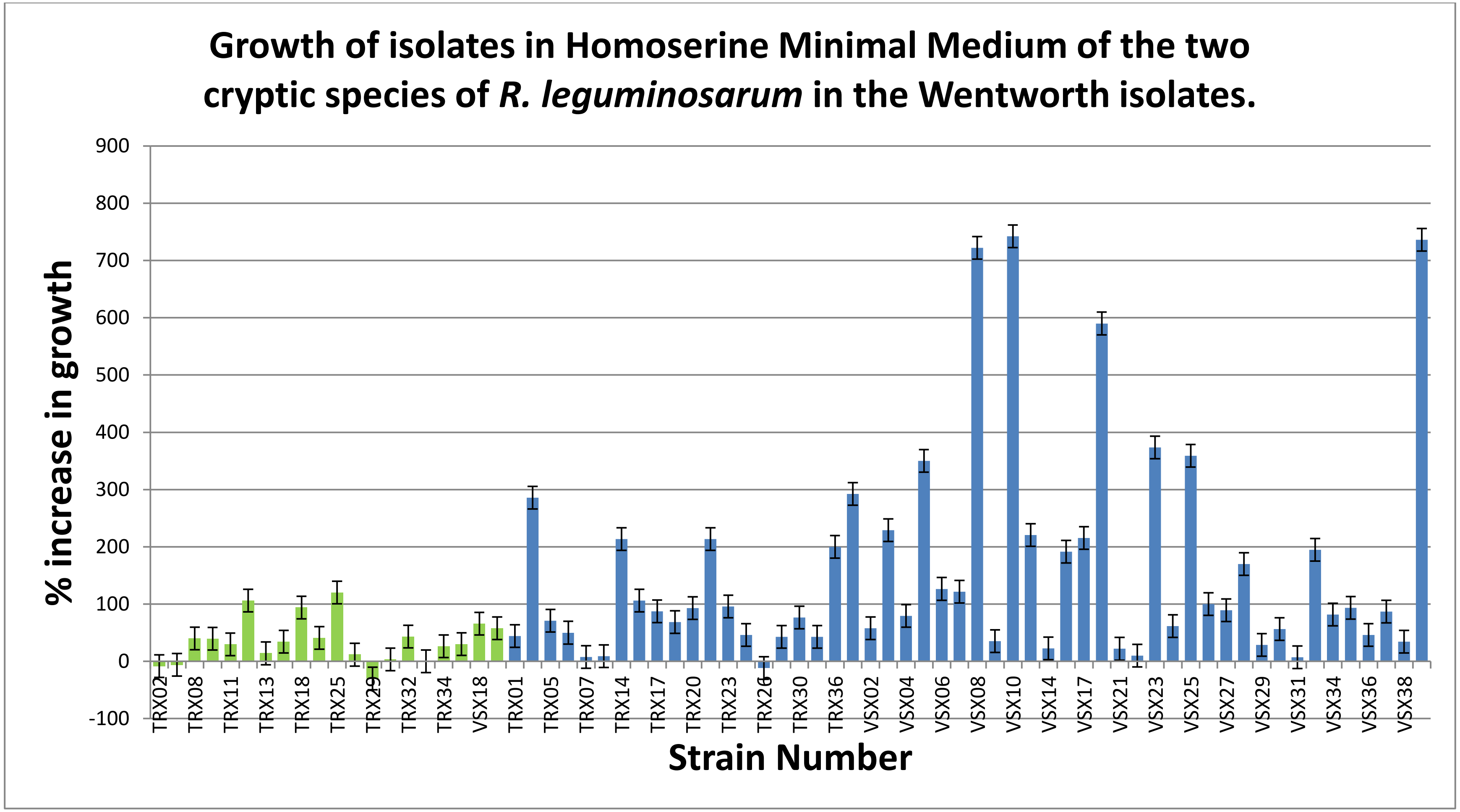


Figure 2.5c. Graphical representation of the Homoserine growth experiment. Strains are arranged as the two cryptic species 2A and 2B shown using green and blue colours respectively. Error bars indicate standard error of means.

2.3.3.1. Test of normality :

The data was tested for normality using the Shapiro-Wilk test with the parameters used for dry weight analysis of plants used in nodulation tests. The test was performed to test the differences between the two biovars, the five cryptic species group and the two cryptic species group.

- **Normality test between biovars :**

The Shapiro-Wilk test (Table 2.8) shows that the homoserine utilization data for the two biovars of *Rhizobium leguminosarum* viz. *trifolii* and *viciae* data is not normally distributed.

Table 2.8. Tests of Normality (Lilliefors Significance Correction)

	Biovar Type	Shapiro-Wilk		
		Statistic	Degrees of freedom	Sig.
Biovar	TRX	.873	72	.000
	VSX	.755	72	.000

Results of Shapiro-Wilk test to check normality distribution of homoserine utilization data.

- **Normality test between the strains in two cryptic species :**

The Shapiro-Wilk test (Table 2.9) shows that the homoserine utilization data for the strains comprising two cryptic species groups 2A and 2B are not normally distributed.

Table 2.9. Tests of Normality (Lilliefors Significance Correction)

	Cryptic2	Shapiro-Wilk		
		Statistic	Degrees of freedom	Sig.
Cryptic species	2A	.960	40	.164
	2B	.739	104	.000

Results of Shapiro-Wilk test to check normality distribution of homoserine utilization data of the strains included in the two cryptic species.

- **Normality test between the strains in five cryptic species :**

The results of the Shapiro-Wilk test (Table 2.10) shows that the homoserine utilization data for the strains comprising five cryptic species groups 5A, 5B, 5C, 5D and 5E was not normally distributed.

Table 2.10. Tests of Normality (Lilliefors Significance Correction)

	Cryptic5	Shapiro-Wilk		
		Statistic	Degrees of freedom	Sig.
Cryptic species	5B	.919	24	.056
	5C	.739	104	.000
	5D	.866	8	.137
	5E	.964	6	.848

Results of Shapiro-Wilk test to check normality distribution of homoserine utilization data of the strains included in the five cryptic species.

2.3.3.2. Kruskal-Wallis test :

The data for homoserine utilization was not normally distributed, hence the data was analysed using the Kruskal-Wallis test.

The test was carried out to test the difference in the ability to utilize homoserine between : the two biovars of *Rhizobium leguminosarum* viz. *trifolii* and *viciae*, the strains in the two cryptic species and the strains that comprise the five cryptic species as suggested by the SplitsTree analysis.

- **Kruskal-Wallis test for homoserine utilization between biovars :**

The null hypothesis (H_O) for the analysis was : There is no significant difference between the ability to utilize homoserine as the sole carbon and nitrogen source between the two biovars of *R. leguminosarum*.

The alternate hypothesis (H_α) for the analysis was : There is a significant difference between the ability to utilize homoserine as sole carbon and nitrogen source between the biovars of *R. leguminosarum*.

Table 2.11. Test Statistics (Kruskal Wallis Test)
Grouping Variable : Biovar

	Growth
Chi-Square	18.590
Degrees of freedom	1
Asymp. Sig.	.000

Results of the Kruskal-Wallis test to check for differences in homoserine utilization between the two biovars of *R. leguminosarum*.

The results of the analysis indicates that there is a significant difference in the medians, $\chi^2(1, N = 144) = 18.590, p = .000$.

Since *p*-value = 0.000 ≤ 0.05 =α, we reject the null hypothesis.

Conclusion in words : At the $\alpha = 0.05$ level of significance, there exists enough evidence to conclude that there is a difference in the median test scores between the two biovars.

- **Kruskal-Wallis test for homoserine utilization between the strains in the two cryptic species :**

The null hypothesis (H_O) for the analysis was : There is no significant difference in the ability to utilize homoserine as the sole carbon source between the strains included in the two cryptic species.

The alternate hypothesis (H_α) for the analysis was : There is a significant difference in the ability to utilize homoserine as the sole carbon source between the strains included in the two cryptic species.

Table 2.12. Test Statistics (Kruskal Wallis Test)
Grouping Variable : cryptic2

	Growth
Chi-Square	27.753
Degrees of freedom	1
Asymp. Sig.	.000

Results of the Kruskal-Wallis test to check for differences in homoserine utilization between strains included in the two cryptic species.

The results of the analysis indicates that there is a significant difference in the medians, $\chi^2(1, N = 144) = 27.753$, $p = .000$.

Since p-value = 0.000 ≤ 0.05 =α, we reject the null hypothesis.

Conclusion in words : At the α = 0.05 level of significance, there exists enough evidence to conclude that there is a difference in the median test scores between the two cryptic species.

- **Kruskal-Wallis test for homoserine utilization between the strains in the five cryptic species :**

The null hypothesis (H_O) for the analysis was : There is no significant difference between the ability to utilize homoserine as the sole carbon source between the strains in the five cryptic species.

The alternate hypothesis (H_α) for the analysis was : There is a significant difference between the ability to utilize homoserine as the sole carbon source between the strains in the five cryptic species.

Table 2.13. Test Statistics (Kruskal Wallis Test)
Grouping Variable: cryptic5

	Growth
Chi-Square	30.069
Degrees of freedom	4
Asymp. Sig.	.000

Results of the Kruskal-Wallis test to check for differences in homoserine utilization between strains included in the five cryptic species.

The results of the analysis indicates that there is a significant difference in the medians, $\chi^2(4, N = 144) = 27.753$, $p = .000$.

Since p-value = 0.000 ≤ 0.05 = α, we reject the null hypothesis.

Conclusion in words : At the α = 0.05 level of significance, there exists enough evidence to conclude that there is a difference in the median test scores among the five cryptic species.

Because the overall test is significant, pairwise comparisons among the five cryptic species should be completed.

2.3.3.3. Post-hoc test : Mann-Whitney Test :

The analysis shows that the mean of at least one group differs from the other groups but does not indicate which group differs. To analyse the pattern of difference between means, the one-way Kruskal–Wallis test was followed by a post-hoc test to carry out specific pairwise comparisons. The Mann-Whitney Test was used as the test of choice.

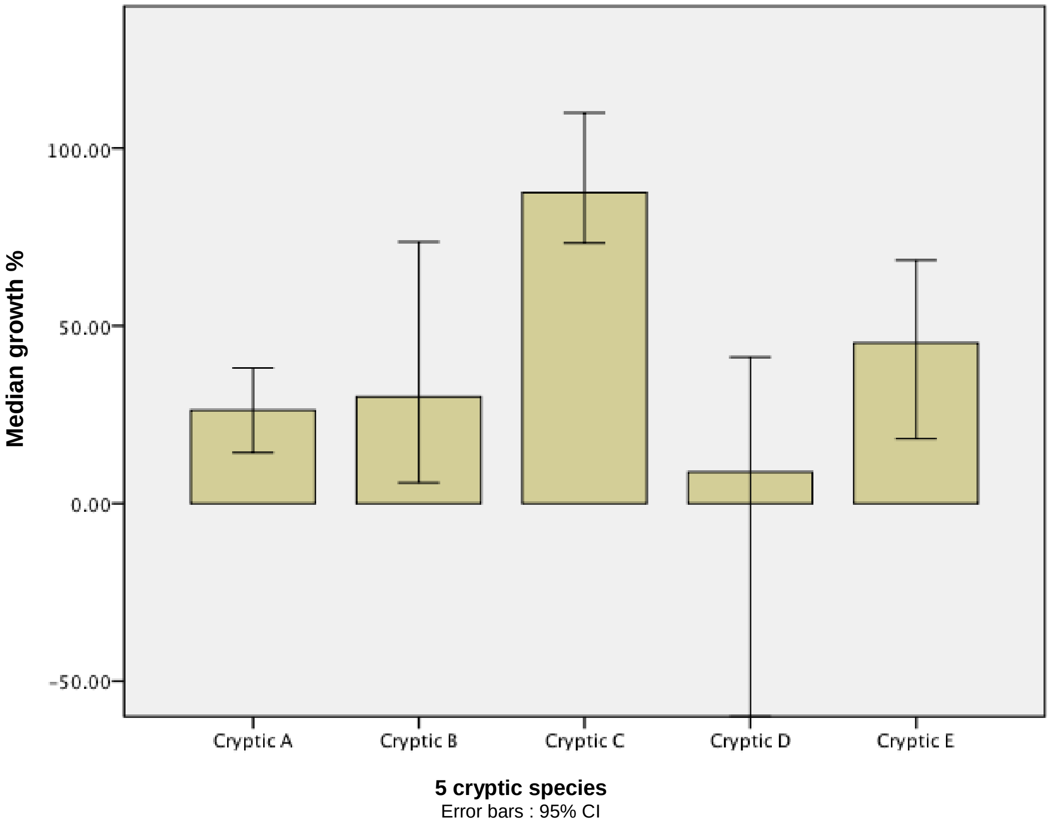


Figure 2.6. Figure showing the results of the Mann-Whitney test for pairwise comparison of difference in homoserine utilization between the five cryptic species.

The results of these tests indicated a significant difference between the cryptic species in their ability to utilize homoserine as a carbon and nitrogen source. The isolates in the cryptic species C are significantly better at utilizing homoserine as compared to those in the other four groups.

The analysis of the difference in the homoserine utilization data indicates that biovar *viciae* strains are, in general, better at homoserine utilization that the strains belonging to biovar *trifolii*. This is also probably the reason why in the analysis of the data as two cryptic species and five cryptic species, the cryptic species 2B (in the two cryptic species dataset) and the cryptic species 5C (in the five cryptic species dataset) show better utilization of homoserine since these the strains included in these groups contain more strains of biovar *viciae* as compared to biovar *trifolii*. This was verified using the Kruskal-Wallis test which showed that there is a significant difference between the biovars *viciae* and *trifolii* (chi-sq. (1, N = 104) = 5.599. p=0.018) included in the group. This is in-line with the observation of Egeraat (1975) who reported this phenomenon.

2.3.4. AHL test :

The 72 isolates from Wentworth were tested for their ability to synthesize AHL quorum sensing signal molecules which were detected using the reporter strain *Chromobacterium violaceum* CV026. The amount of AHL synthesized was calculated in terms of the amount of violacein synthesized, expressed in violacein units. The data for AHL synthesis was not normally distributed, hence the data was analysed using the Kruskal-Wallis test.

The test was carried out to test the difference in the ability to synthesize AHLs between : the two biovars of *Rhizobium leguminosarum* viz. *trifolii* and *viciae*, the strains that comprise the two cryptic species and the strains that comprise the five cryptic species as suggested by the SplitsTree analysis.

Table 2.14. Results of the AHL test (in Violacein Units)

Strain Number	Rep1	Rep2	Average
TRX01	500.00	416.67	458.33
TRX02	354.84	173.91	264.38
TRX03	342.86	1000.00	671.43
TRX04	700.00	1181.82	940.91
TRX05	1250.00	2000.00	1625.00
TRX06	9480.00	12421.05	10950.53
TRX07	12166.67	17555.56	14861.11

TRX08	2320.00	3473.68	2896.84
TRX09	600.00	888.89	744.44
TRX10	1476.19	4133.33	2804.76
TRX11	736.84	1062.50	899.67
TRX12	1870.97	2583.33	2227.15
TRX13	1642.86	2521.74	2082.30
TRX14	2000.00	3375.00	2687.50
TRX15	0.00	1000.00	500.00
TRX16	5833.33	8653.85	7243.59
TRX17	6225.81	7640.00	6932.90
TRX18	444.44	1350.00	897.22
TRX19	1904.76	3062.50	2483.63
TRX20	1142.86	1095.24	1119.05
TRX21	7894.74	11692.31	9793.52
TRX22	6421.05	7071.43	6746.24
TRX23	10333.33	13400.00	11866.67
TRX24	1428.57	904.76	1166.67
TRX25	500.00	578.95	539.47
TRX26	1944.44	2500.00	2222.22
TRX27	1360.00	2300.00	1830.00
TRX28	684.21	3428.57	2056.39
TRX29	1000.00	2529.41	1764.71
TRX30	8142.86	10687.50	9415.18
TRX31	480.00	1050.00	765.00
TRX32	1285.71	2294.12	1789.92
TRX33	1045.45	1705.88	1375.67
TRX34	406.25	500.00	453.13
TRX35	11826.09	14888.89	13357.49
TRX36	1590.91	1888.89	1739.90
VSX01	14615.38	18950.00	16782.69
VSX02	2210.53	3642.86	2926.69
VSX03	17200.00	24333.33	20766.67
VSX04	913.04	1333.33	1123.19

VSX05	1000.00	1500.00	1250.00
VSX06	7272.73	9529.41	8401.07
VSX07	14826.09	20294.12	17560.10
VSX08	4800.00	5950.00	5375.00
VSX09	576.92	1000.00	788.46
VSX10	107.14	227.27	167.21
VSX11	2720.00	4055.56	3387.78
VSX14	12772.73	17235.29	15004.01
VSX15	12681.82	15941.18	14311.50
VSX16	875.00	1823.53	1349.26
VSX17	17714.29	20263.16	18988.72
VSX18	193.55	272.73	233.14
VSX19	2809.52	3882.35	3345.94
VSX21	285.71	1125.00	705.36
VSX22	555.56	733.33	644.44
VSX23	347.83	375.00	361.41
VSX24	250.00	217.39	233.70
VSX25	0.00	95.24	47.62
VSX26	1565.22	2166.67	1865.94
VSX27	10640.00	13300.00	11970.00
VSX28	545.45	1000.00	772.73
VSX29	631.58	1214.29	922.93
VSX30	17476.19	20647.06	19061.62
VSX31	16590.91	23312.50	19951.70
VSX32	2115.38	1250.00	1682.69
VSX33	954.55	562.50	758.52
VSX34	920.00	650.00	785.00
VSX35	9090.91	14411.76	11751.34
VSX36	13666.67	17214.29	15440.48
VSX37	13083.33	18315.79	15699.56
VSX38	15173.91	19555.56	17364.73
VSX39	9742.86	10107.14	9925.00

Results of the AHL signal molecules synthesized by each of the 72 Wentworth isolates. The result in expressed in violacein units as described in the text.

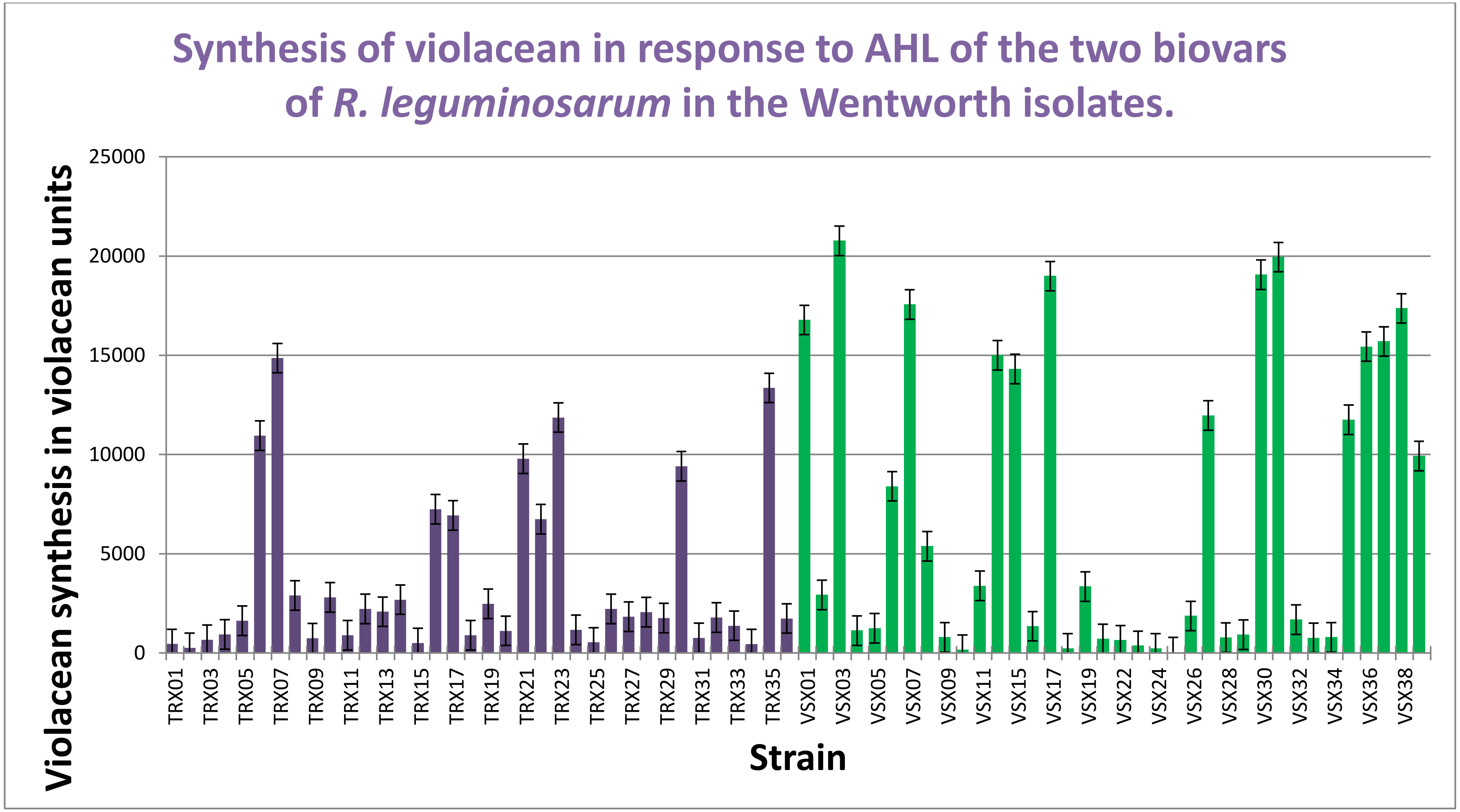


Figure 2.7a. Graphical representation violacein synthesis by *Chromobacterium violaceum* CV026 in response to AHL synthesized by the Wentworth isolates. TRX strains are shown in violet and VSX in green. Error bars indicate standard error of means.

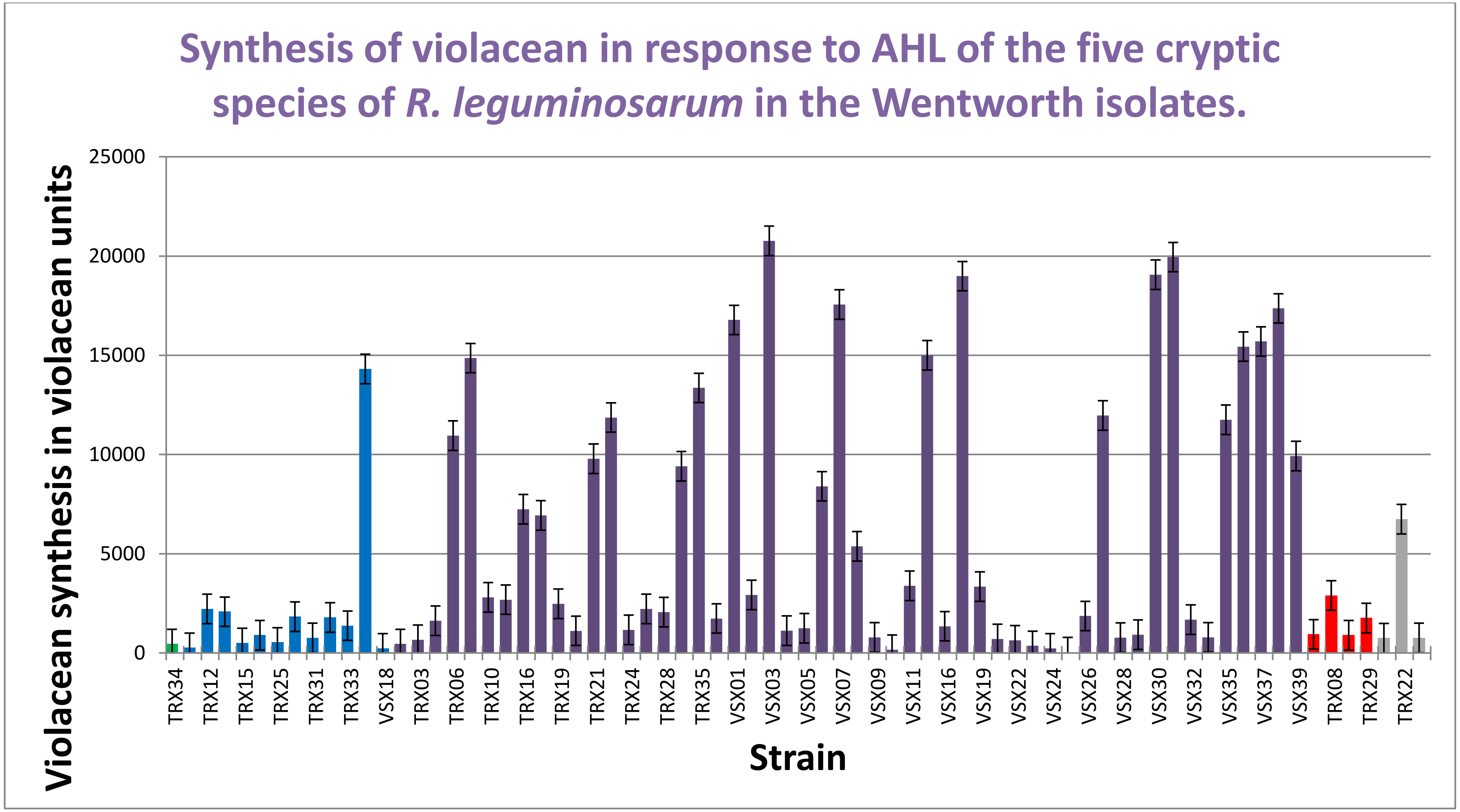


Figure 2.7b. Graphical representation violacein synthesis by *Chromobacterium violaceum* CV026 in response to AHL synthesized by the Wentworth isolates. The strains belonging to the five cryptic species 5A, 5B, 5C, 5D and 5E are shown in green, blue, purple, red and grey colours respectively. Error bars indicate standard error of means.

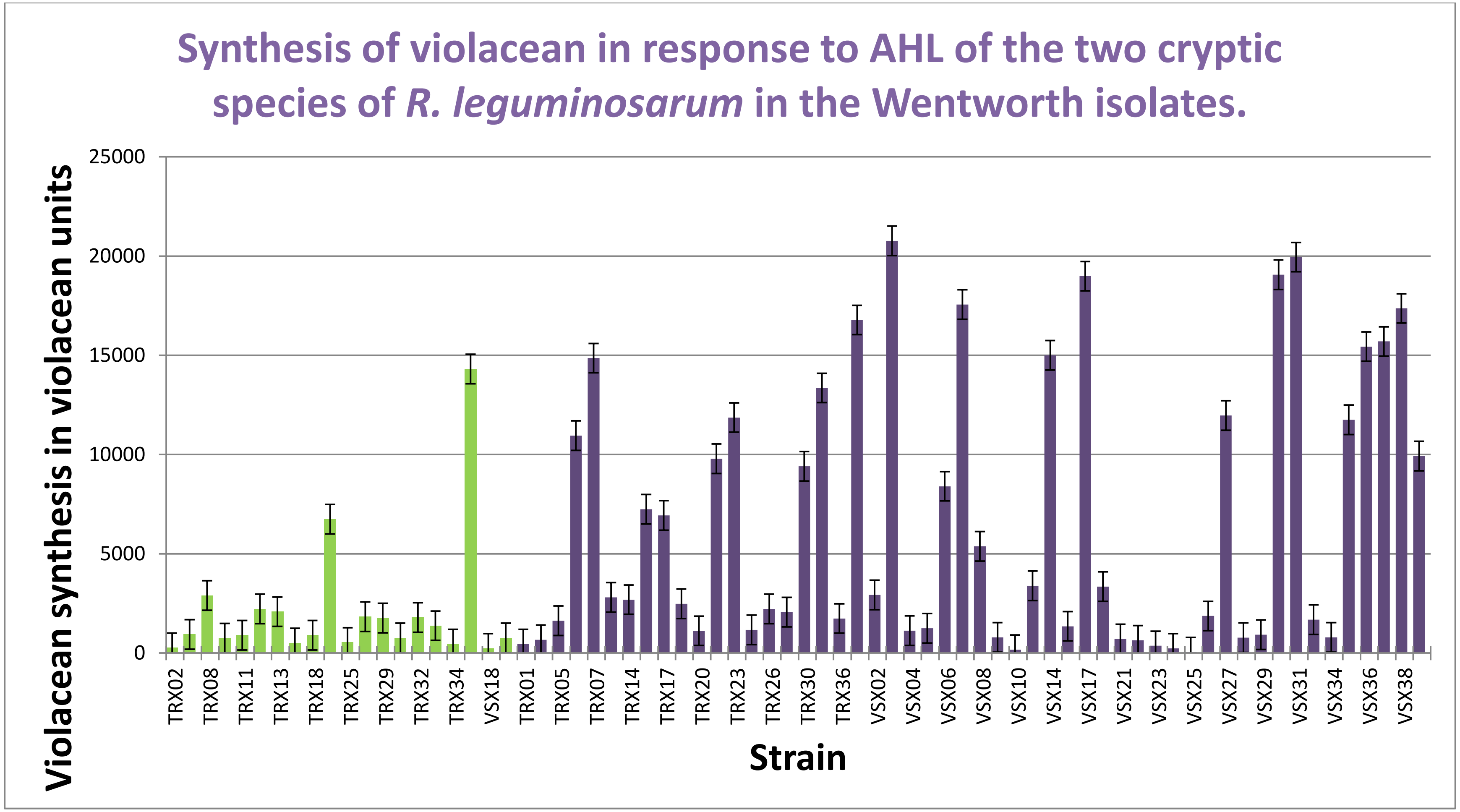


Figure 2.7c. Graphical representation violacein synthesis by *Chromobacterium violaceum* CV026 in response to AHL synthesized by the Wentworth isolates. The strains belonging to the two cryptic species 2A and 2B are shown in green and purple colours respectively. Error bars indicate standard error of means.

2.3.4.1. Test of normality :

The data was tested for normality using the Shapiro-Wilk test with the same parameters as used for dry weight analysis. The test was performed three times to check the differences between the two biovars, the five cryptic species group and the two cryptic species group.

- **Normality test between biovars :**

The Shapiro-Wilk test shows that the AHL synthesis data for the two biovars of *Rhizobium leguminosarum* viz. *trifolii* and *viciae* data is not normally distributed.

Table 2.15. Tests of Normality (Lilliefors Significance Correction)
Grouping = Biovar

	Biovar Type	Shapiro-Wilk		
		Statistic	Degrees of freedom	Sig.
Violacein synthesis	TRX	.753	72	.000
	VSX	.824	72	.000

Results of Shapiro-Wilk test to check normality distribution of AHL synthesis data for the two biovars of *R. leguminosarum.*

- **Normality test between the strains in the two cryptic species :**

Shapiro-Wilk showed that the AHL synthesis data for the strains comprising two cryptic species groups 2A and 2B was not normally distributed.

Table 2.16. Tests of Normality (Lilliefors Significance Correction)
Grouping = 2 cryptic species.

	2 Cryptic species	Shapiro-Wilk		
		Statistic	Degrees of freedom	Sig.
Violacein synthesis	2A	.561	40	.000
	2B	.843	104	.000

Results of Shapiro-Wilk test to check normality distribution of AHL synthesis data for the strains included in the two cryptic species.

- **Normality test between the strains in the five cryptic species :**

Shapiro-Wilk showed that the AHL synthesis data for the strains comprising five cryptic species groups 5A, 5B, 5C, 5D and 5E was not normally distributed.

Table 2.17. Tests of Normality (Lilliefors Significance Correction)
Grouping = 5 cryptic species.

	5 Cryptic species	Shapiro-Wilk		
		Statistic	Degrees of freedom	Sig.
Violacein	Cryptic 5B	.514	24	.000
Violacein	Cryptic 5C	.843	104	.000
Violacein	Cryptic 5D	.849	8	.092
Violacein	Cryptic 5E	.703	6	.007

Results of Shapiro-Wilk test to check normality distribution of AHL synthesis data for the strains included in the five cryptic species.

2.3.4.2. Kruskal-Wallis test :

- **Kruskal-Wallis test for AHL synthesis between biovars :**

The null hypothesis (H_O) for the analysis was : There is no significant difference between the ability to synthesize AHLs between the two biovars of *R. leguminosarum*.

The alternate hypothesis (H_α) for the analysis was : There is a significant difference between the ability to synthesize AHLs between the two biovars of *R. leguminosarum*.

Table 2.18. Test Statistics (Kruskal Wallis Test)
Grouping Variable = Biovar

	Violacein
Chi-Square	2.810
Degrees of freedom	1
Asymp. Sig.	.094

Results of the Kruskal-Wallis test to check for differences in AHL synthesis between the two biovars of *R. leguminosarum*.

The results of the analysis indicates that there is no significant difference in the medians, $\chi^2(1, N = 144) = 2.810$, $p = 0.094$.

Since p-value = 0.094 ≥ 0.05 = α, we do not reject the null hypothesis.

Conclusion in words : At the $\alpha = 0.05$ level of significance, there is no difference in the median test scores between the two biovars.

- **Kruskal-Wallis test for synthesis of AHL signal molecules between strains in the two cryptic species :**

The null hypothesis (H_O) for the analysis was : There is no significant difference between the ability to synthesize AHL signal molecules as between the strains included in the two cryptic species.

The alternate hypothesis (H_α) for the analysis was : There is a significant difference between the ability to synthesize AHL signal molecules between the strains included in the two cryptic species.

Table 2.19. Test Statistics (Kruskal Wallis Test)
Grouping variable = 2 Cryptic species

	Violacein
Chi-Square	14.800
Degrees of freedom	1
Asymp. Sig.	.000

Results of the Kruskal-Wallis test to check for differences in AHL synthesis between strains in the two cryptic species.

The results of the analysis indicates that there is a significant difference in the medians, $\chi^2(1, N = 144) = 14.800$, $p = .000$.

Since p-value = 0.000 ≤ 0.05 = α, we reject the null hypothesis.

Conclusion in words : At the $\alpha = 0.05$ level of significance, there exists enough evidence to conclude that there is a difference in the median test scores between strains in the two cryptic species.

- **Kruskal-Wallis test for AHL signal molecules between the strains in the five cryptic species :**

The null hypothesis (H_O) for the analysis was : There is no significant difference between the ability to synthesize AHL signal molecules between the strains in the five cryptic species.

The alternate hypothesis (H_α) for the analysis was : There is a significant difference between the ability to synthesize AHL signal molecules between the strains in the five cryptic species.

Table 2.20. Test Statistics (Kruskal Wallis Test)
Grouping variable = 5 Cryptic species

	Violacein
Chi-Square	16.296
Degrees of freedom	4
Asymp. Sig.	.003

Results of the Kruskal-Wallis test to check for differences in AHL synthesis between strains in the five cryptic species.

The results of the analysis indicates that there is a significant difference in the medians, $\chi^2(4, N = 144) = 16.296$, $p = .003$.

Since *p*-value = $0.003 \le 0.05 = \alpha$, we reject the null hypothesis.

Conclusion in words : At the $\alpha = 0.05$ level of significance, there exists enough evidence to conclude that there is a significant difference in the median test scores between strains that are included in the five cryptic species.

2.3.4.3. Post-hoc test : Mann-Whitney Test :

In order to analyse the pattern of difference between means, specific pairwise comparisons were carried out using the Mann-Whitney Test.

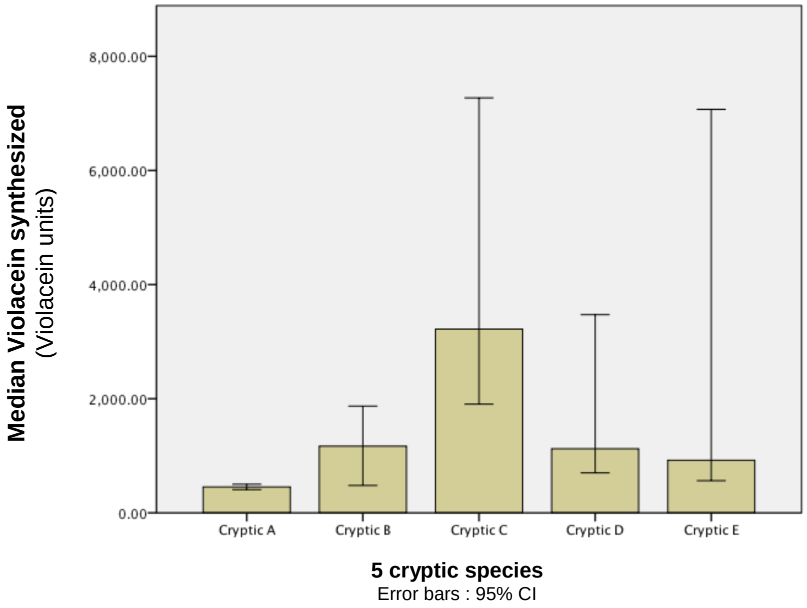


Figure 2.8. Graphical representation of the Mann-Whitney test results for pairwise comparison of difference in AHL synthesis between the five cryptic species.

The results of the Mann-Whitney tests indicate that there is significant difference between the cryptic species 5A and others and between the cryptic species 5B and 5C. Other pairwise comparisons did not yield any significant differences in the amount of violacein made reflecting no significant differences in the amount of AHL made by the species included in the respective cryptic species groups. The strains included in the group 5C/2B did not show a biovar effect. i.e. there was no significant difference in the violacein synthesis between the *trifolii* and *viciae* biovars in that group (chi-sq. (1, N = 104) = 0.077. p = 0.782).

2.4. Discussion :

The first aim of this chapter was to verify that the bacterial strains to be used in the study have retained their ability to nodulate their native plants species. All the 72 isolates of *R. leguminosarum* formed large and pink effective nodules on the primary and upper lateral roots. Study of the normality of distribution of the data revealed that the data for both *T. repens* and *V. cracca* was distributed normally. Using appropriate statistical tests and subsequent post-hoc tests indicated that in *T. repens* there was a significant difference between the plants included in the negative control and those included in the other two groups (viz. test plants and positive control). Theoretically, all other conditions being constant, a plant with more access to nitrogen should grow better than a plant with lower access. The plants which were included as negative control showed the least growth which indicates that a lack of nitrogen stunts plant growth. The inoculated plants grew as well as those provided with nitrate, indicating that the symbiosis was effective in meeting their needs for nitrogen. However, contrary to expectation, the *V. cracca* plants inoculated with the VSX strains showed no difference between the test and either of the control plants. Moreover, *V. cracca* plants included in the controls did not show the presence of any nodules indicating absence of cross contamination. Since the main aim of nodulation tests was to test the nodulation ability of the strains and all the strains formed effective nodules, the now validated strains were ready for use in further studies.

The second aim of this chapter was to check if known phenotypic differences between biovars *viciae* and *trifolii* are also seen in the Wentworth strains. Two phenotypic differences known to vary between the biovars include the ability to use homoserine as a carbon source and the ability to synthesize AHL quorum sensing signal molecules. It has been reported that the biovar *viciae* has better ability to utilize homoserine as well as synthesize AHL quorum sensing signal molecules. The results of the tests showed that, indeed, a greater number of biovar *viciae* isolates were able to use homoserine as a carbon source as well as synthesize AHL quorum sensing signal molecules as compared to *trifolii* isolates. The difference was statistically evaluated and was found to be significant.

The analysis was carried further to check if the ability to utilise homoserine or synthesize AHL signal sensing molecules was related to the position of the strain on the SpiltsTree showing relatedness between species. The analysis showed that the clade containing relatively higher number of biovar *viciae* isolates (2B / 5C) was better in both the tests. The results of the homoserine indicate that there is a difference between the biovars *viciae* and *trifolii*. It also indicates that since the difference, in general, is a difference between biovars, there should be an explanation that can be given in terms of the genes that would be more commonly found in one biovar than the other.

This raises the possibility of exploring phenotypic differences between the isolates and trying to explain the differences at a genetic level in terms of the presence or absence of genes. The system that will be used to study the phenotypes should be robust and at the same time allow the study of a relatively large number of phenotypes with ease. This idea is explored in the next chapter.

CHAPTER 3 : METABOLIC FINGERPRINTING OF WENTWORTH STRAINS USING BIOLOG PHENOTYPE MICROARRAY^{TM} PLATES

3.1. Introduction :

The results of the nodulation test, homoserine utilization test and AHL-synthesis test suggest that a correlation may exist between the genetic makeup of an organism and its phenotype. This correlation is absolute in case of the *nodD* gene which determines the host range of *Rhizobium*. The phylogenetic tree of *nodD* clearly differentiates the *trifolii* strains from the *viciae* strains, which was reflected in the ability of the strains to infect the host plants. This genotype-phenotype relationship was further investigated by the other two tests.

The results obtained from the other two phenotypic tests suggest that there is some difference between biovars *viciae* and *trifolii* in their ability to utilize homoserine and in their ability to synthesize AHL quorum sensing signal molecules. The difference is enough to be statistically validated. The results indicate that a correlation may exist between the genome of an organism its phenotype but it might not be absolute. The results of the nodulation and the other two tests pose interesting research questions. Perhaps the most interesting question is to assess the stringency of the relationship between the genome of an organism and its phenotype and can the genetic composition of an organism be used to predict its phenotype and vice versa. In order to answer these questions a systematic phenotypic and genetic analysis of the Wentworth isolates is required.

The 72 Wentworth isolates were therefore subjected to a systematic phenotypic analysis. The objective of this analysis would be to look for obvious phenotypic differences between the isolates and to trace these differences back to its genome in terms of the presence or absence of genes. The genes that show a high correlation with the ability to utilize a substrate would then become candidates for further analysis. The involvement of these genes in the metabolism of the substrate can then be confirmed in *R. leguminosarum* biovar *viciae* strain 3841 by mutation and complementation analysis. Mutating the gene would allow confirmation of the role of the gene in the metabolism of the substrate. Reversion to wild-type phenotype on complementation would be further confirmation of its role in metabolism.

3.1.1. Classification of bacteria :

The current classification schemes used by most microbiologists rely on phenotypic methods. The 1994 edition of *Bergey's Manual of Determinative Microbiology*, an authoritative text on bacterial classification, states that "the arrangement of the book is based on phenotype and no attempt has been made to offer a natural classification. The arrangement chosen is utilitarian and is intended to aid in the identification of bacteria". Hence, a species is a collection of strains that share many common phenotypic characteristics. Species are assigned to morphologically and biochemically defined genera, which in turn are assigned to families, each of which also has certain morphological, physiological, and biochemical features. The phenotypic classification methods utilize a variety of characteristics for classification including morphology, physiology and a variety of biochemical tests.

The advent of molecular methods like DNA hybridization and sequencing has made it possible to develop classification based on inference of phylogenetic relationships (Palleroni, 2003). Based on these methods, currently different definitions of species exist viz., taxospecies, genomic species and nomenspecies. Taxospecies is defined as a group of bacteria of high mutual similarity, and thus a polythetic phenetic group approximating to a natural taxon. The genospecies is defined as a group of bacteria whose members are capable of exchanging genes while a nomenspecies is a group bearing one binomial name, but not necessarily natural in a phenetic or genetic sense (Sneath, 1972). Taxospecies is based on classification based on phenotype and contains organisms with mutually high phenotypic similarity and form an independent phenotypic cluster. The genospecies contain organisms that as a group show high DNA-DNA similarity (Allsopp et al., 1995).

The word species in bacteria usually means a taxospecies. The cluster of taxospecies is thought to indicate evolutionary relationship in terms of showing the end products of related metabolic processes based on the hypothesis that metabolic functions would be largely conserved in bacteria related by evolutionary descent. Since bacteria reproduce asexually vertical inheritance should predominate. However, some bacteria with highly related phenotypes do not appear to be related phylogenetically. These bacteria that are highly related based on phenotypic characterization but which come from different evolutionary backgrounds are said to be polyphyletic (Spiers et al., 2000).

Phenotypic incongruence is largely the result of gene loss, gain, modification and lateral gene transfer (Forney et al., 2004). Hence, it has been suggested that both the approaches be used to develop a new taxonomic classification system with a concept of phylophenetic species that will include a monophyletic cluster that show a high degree of phenotypic similarity and can be distinguished from other clusters based on a discriminating phenotypic property (Rossello-Mora and Amann, 2001). Such a cluster of species forming a distinct phylogenetic clade is called a phylospecies. The access to a large amount of sequence information has currently made the concept of phylospecies the most widely used species concept by taxonomists. Traditional "polyphasic taxonomy" requires bacterial species to have both phylogenetic coherence and distinctive phenotypic traits, i.e. to be both phylospecies and taxospecies. The classic description by Vandamme et al. (1996) does not have much on phylogeny since there were not many DNA sequences in 1996. More recently, phyogenies of 16S and housekeeping genes have come to dominate species descriptions, so the phylospecies is a reality in practice.

3.1.1.1. Phenotypic classification of bacteria :

Phenotypic classification aims to group or cluster bacteria based on similarities in phenotypes. The phenotypes can be a single-character or a multi-character set of phenotypes. Single-character tests include tests such as staining reaction, motility, pigment production, morphology, etc. These tests form a very small component of the total bacterial phenotype. Hence, they produce a simple, but a relatively inaccurate picture of classification. Yet many bacterial taxa which have unique single-character features like Cyanobacteria, stalked bacteria etc. are strongly supported by more advanced methods of classification. In contrast, multi-character tests study a large number of phenotypes simultaneously and hence better represent the taxonomic landscape. Multi-character tests include tests like carbon-substrate utilization, protein profiling, profiling of metabolic products, antibiotic resistance, phage susceptibility etc.

3.1.1.2. Phylogenetic classification of bacteria :

Phylogenetics studies the evolutionary relationships between organisms. The relationship is represented in the form of tree-like diagrams that depict the evolutionary relationships among molecules, organisms or both (Baxevanis and Ouellette, 2004).

A molecular phylogenetic analysis consists of four steps viz., sequence alignment, determining the substitution model, tree building and tree evaluation. Several areas of the bacterial genome have been used for the inference of phylogenetic relationships. The regions of DNA most commonly used are the 16S rRNA gene, 16S-23S rDNA intergenic spacer (IGS), *atpD* and *recA* (Vinuesa et al., 2005). Their utility as phylogenetic markers arises because they encode essential cellular functions.

3.1.1.3. Genotypic and phenotypic diversity in *Rhizobium* :

The genotypic and phenotypic diversity of the *Rhizobium* group have been studied using a variety of approaches, including, but not limited to enzyme pattern, serological study, plasmid profile, PCR-RFLP fingerprinting, sequence analysis, symbiotic variation, substrate utilization and from various geographical locations around the world (Aoki et al., 2010, Sylla et al., 2002, Rodriguez-Navarro et al., 2004, Nour et al., 1994, Pereira et al., 2002, Ködöböcz et al., 2009, Maatallah et al., 2002, Rai et al., 2012, Mutch and Young, 2004).

The studies show that members of the *Rhizobium* group show a great amount of diversity in terms of genotype and phenotype. Some of these studies have been used to define new species. Genomic analysis has been used to define *Bradyrhizobium liaoningense* (Xu et al., 1995), *Rhizobium tianshanense* (Chen et al., 1995), *Rhizobium hainanense* (Chen et al., 1997) and many others. Similarly phenotypic differences have been offered to define *Rhizobium multihospitium*

(Han et al., 2008) as being able to utilize sodium formate as a sole carbon source, L-threonine and D-threonine as sole nitrogen sources as well as resistance to chloramphenicol (100 µg/ml) and erythromycin (100 µg/ml). Similarly, *Shinella kummerowiae* type strain CCBAU 25048T (Lin et al., 2008) was defined as a species that could be differentiated from the two defined *Shinella* species based on several phenotypic characteristics including the use of citrate and D-ribose as sole carbon sources, the growth at pH 11.0 as well as the fatty acid composition.

3.1.2. The Biolog system for studying bacterial phenotypes :

Bochner and Savageau (1977) described a method for studying the metabolic abilities of bacteria using a tetrazolium dye which Bochner commercialized to manufacture 96-well colorimetry-based metabolic assay microplates through his California-based company, Biolog Inc. (Bochner, 1989).

3.1.2.1. General overview of the Biolog system :

The Biolog system was introduced in 1989 for identification of aerobic Gram-negative bacteria by determining their carbon source utilization profiles, and has now been expanded to included Gram-positive bacteria as well. The Biolog GN MicroPlate (for identification of gram-negative bacteria) and the Biolog GP MicroPlate (for identification of gram-positive bacteria) are 96-well microtitre plates, each well containing a dehydrated mixture of tetrazolium violet, a buffered nutrient medium, a gelling agent and a different carbon source for each well except the control, which does not contain a carbon source. The microwells are hydrated by inoculating a suspension of the test bacterium. The plates are incubated and read at an appropriate time to check for the ability of the bacteria to utilize the carbon source. If the carbon substrate is utilized the dye in the well turns purple, whereas if it not utilized, the dye remains colourless. The formazan dye is clearly visible with the naked eye, producing a visually distinct pattern of purple wells on the microplate for different microbes. The pattern is called the "metabolic fingerprint" of the test bacterium (Bochner, 1989).

3.1.2.2. The chemistry of the Biolog system :

The Biolog system uses redox dyes to colorimetrically measure an increase in the metabolic rate when bacteria are oxidizing a substrate. During oxidation, the electrons generated are normally passed through the electron transport system. Oxidised tetrazolium dyes have a higher affinity for the electrons as compared to the components of the electron transport chain. Hence, in the presence of oxidised tetrazolium salts, the electrons get attracted to the salts and reduce them irreversibly to a highly coloured formazan. As a result, when the cells are incubated in the presence of a substrate that is utilized by them, their respiration increases and the dye is reduced. Conversely, when the cells are incubated in the presence of a substrate that is not utilized by them, their respiration does not increase and the dye is not reduced. The dye incorporated in the Biolog GenII plates is the oxidised form of tetrazolium violet which on reduction becomes a purple formazan.

pH-sensitive dyes, traditionally used for metabolic studies, can be only used to monitor metabolic activities that alter pH, such as sugar catabolism. Tetrazolium is sensitive to any activity that results in the transfer of electrons from NADH to the electron transport chain. Also, the formation of formazan by the reduction of tetrazolium is an irreversible process (Jambor, 1954). This is not the case for pH-sensitive dyes, which show a colour reversion as the acid by-products are oxidized.

The Biolog system can also be used to study and develop metabolic fingerprints of anaerobic bacteria since even anaerobic bacteria have electron transport systems from which the electrons can be diverted to reduce the tetrazolium dye. (Bochner and Savageau 1977; Bochner, 1989).

The basic Biolog identification system uses 95 carbon substrates which are tested for their utilization. However, virtually any substrate, irrespective of its nature or structure, when utilized as a growth substrate, results in the formation of NADH. This means that nearly any compound can be tested in the Biolog system and the response measured is a reflection of its metabolic rate. Even the basic 95 substrate plate can generate 2^{95} (~4 x 10^{28}) positive/negative patterns which can be used to identify bacteria with extremely fine resolution. Additional resolution can be obtained by following the kinetics of colour change (i.e. growth) along with the end-point colour record. Such large number of combination gives great versatility constrained only by the size of the its database.

3.1.2.3. Types of Biolog systems :

As mentioned earlier, the Biolog system was introduced in 1989 for the identification of aerobic Gram-negative bacteria. But since then the technology has been developed and expanded the company now makes a wide variety of assay microplates for identification and characterization of microbes. The microplates used by Biolog have a well diameter of 3 mm instead of the standard 6 mm found in standard 96 well plates and hence the plates must be read in a microplate reader equipped with a narrow light beam. The metabolites on the microplate are arrayed in a manner so as to support microbial identification, as in GEN II and GEN III microplates, or phenotype analysis, as in the extensive number of phenotype microarray (PM) microplates.

- **GEN II MicroPlates :**

There are five different types of GEN II MicroPlates that are currently available from Biolog. The GN2 MicroPlate is designed to identify aerobic Gram-negative bacteria, the GP2 MicroPlate is designed to identify aerobic Gram-positive bacteria. The AN MicroPlate is designed to identify anaerobic bacteria. The last two MicroPlates are designed for yeasts and filamentous fungi are called YT MicroPlate and FF MicroPlate respectively. Each of these plates contains a combination of 95 different sugars, amino acids, nucleic acids, and other metabolites along with a tetrazolium dye.

A special inoculation fluid (IF) is available for each of the above plates to prepare the suspension of the test organism. The metabolic fingerprint patterns developed at the end of the investigation can be compared against a database available for each type of plate. For some plates, it is possible to create user databases by adding patterns produced by new cultures.

- **GEN III MicroPlates :**

  The Biolog GEN III MicroPlates are based on the Phenotype Microarray platform but consist of a single plate using 94 biochemical tests to profile and identify a broad range of Gram-negative and Gram-positive bacteria. The plate includes 94 phenotypic tests (71 carbon utilization tests and 23 chemical sensitivity tests), a positive control and a negative control. All necessary nutrients and biochemicals are prefilled and dried into the 96 wells of the MicroPlate

  However, unlike the GEN II plates, the ‘gelling agent’ and the redox dye is absent in the plates and is included in the inoculation fluid instead. The redox dye is different enabling the use of lower cell densities. The microbial differences are accommodated with three different formulations of inoculating fluid (A, B, and C) and four different protocols. The organisms are identified using Biolog’s Microbial Identification Systems software.

- **Special Purpose MicroPlates :**

  Four different types of Special Purpose MicroPlates are available viz. MT, SF-N2, SF-P2 and ECO MicroPlate. The MT MicroPlate is designed for metabolic capability studies and contains only tetrazolium and a buffered nutrient medium without a carbon source in any of the 96 wells. The user can add various carbon sources to the MicroPlate as per the requirement. Although there is no Biolog data base for the MT MicroPlate, a custom data base can be created by using the Biolog software. The SF-N2 and SF-P2 are designed for metabolic characterization of spore-forming and conidia-forming microbes. They are similar to GN2 and GP3 MicroPlates respectively but do not contain the redox dye. The ECO MicroPlate contains 3 sets of the same 31 carbon sources in a 96 well Biolog MicroPlate and was created for community analysis and microbial ecological studies.

- **The Phenotype Microarray MicroPlates :**

  The Phenotype Microarray (PI) system from Biolog is a high-throughput system that allows the study of thousands of phenotypes at the same time. The system can be used for both Gram-positive and Gram-negative bacteria. The company currently manufactures 20 different PM plates of which the first eight plates study the utilization of carbon, nitrogen, sulphur, phosphorus and utilization of various metabolic intermediates. The ninth and tenth plates study osmotic and ionic response and pH tolerance. The last ten plates are completely dedicated to study of sensitivity to various compounds.

  The PM system is different from the other Biolog systems having multiple inoculation fluids, redox dyes and supplements. The identification is done using Biolog's Microbial Identification Systems software.

3.1.2.4. Studies carried out using the Biolog system :

There have been many studies on the use of the Biolog GN plates for identification and characterisation of bacteria, and many more on the in-depth study and characterisation of bacteria using the Phenotype Microarray system.

Carnahan et al. (1989) used the GN plate to test 20 clinical strains each of *Aeromonas hydrophila*, *A. caviae*, and *A. sobria*. All the 60 *Aeromonas* strains were correctly classified to species. Nine substrates yielded good discriminatory value and seven of these substrates were not previously known to be useful in identifying *Aeromonas* isolates.

Armon et al. (1990) used the Biolog GN panel to study different *Legionella* spp. They tested two strains of *L. pneumophila* (one clinical and one environmental), and one strain each of *L. micdadei*, *L. oakridgensis*, *L. longbeachae*, and *L. gormanii*. The strains varied in their biochemical profiles thus allowing the strains to be distinguished from each other. However, the authors did observe overlap of some biochemical profiles with *Moraxella bovis*.

Mauchline and Keevil (1991) used the Biolog system to identify a saccharolytic *Legionella* spp. They tested single type strains of *L. pneumophila* serogroups 1 through 14 (excluding serogroups 4 and 9) and a single type strain of *L. bozemanii*, *L. dumoffii*, *L. feeleii*, *L. hackeliae*, *L. israelensis*, *L. rubrilucens*, *L. longbeachae*, and *L. micdadei*. They found that each of the *Legionella* isolates has a unique metabolic fingerprint that can be used in its identification. Furthermore, they were able to correctly identify environmental isolates of *Legionella* (previously serotyped for identification) proving that the system is robust for identifying *Legionella* isolates up to the species level.

Currently, the Biolog system is mostly used in assessing the metabolic diversity of isolates than in bacterial identification. In the study of rhizobia too, the Biolog technology has largely been used for studying the phenotypic diversity of isolates. The study of rhizobial isolates from different regions has shown a huge diversity in the phenotypic properties (McInroy et al., 1999, Sylla et al., 2002, Wielbo et al., 2010, Rahi et al., 2012, Mazur et al., 2013).

3.2. Materials and methods :

The results of the preliminary investigation into metabolic differences suggested a need to order to carry out a systematic evaluation of the metabolic differences between the Wentworth isolates. In order to achieve this, metabolic fingerprinting of the strains was performed using the Biolog GN2 Microarray plates from Biolog Inc.

3.2.1. Test strains :

The 72 Wentworth isolates were tested for their ability to utilize the 95 carbon substrates present in Biolog GN2 Microplate Gram-negative identification test panel system.

3.2.2. Growth of test cultures :

The literature on the Biolog website mentions that inoculum from both solid and liquid medium can be used for the Biolog assay. However, while standardising the assay for its use in developing the metabolic fingerprints of the Wentworth strains it was found that the bacteria grown on solid medium (TY agar) gave more consistent and reproducible results as compared to the bacteria grown in liquid medium. Therefore, to obtain the Biolog profiles of the Wentworth strains, the strains were grown on solid TY agar as described below. Moreover, it was observed that the suspension of bacteria in physiological saline gave better results than the Biolog inoculation fluid (IF) and was used in its stead. The preparation and standardization phase was the most important phase of this work.

The isolates were tested for purity by repeated subculturing and checking colony morphology. A heavy inoculum of the pure culture was inoculated onto a sterile TY agar plate and incubated for 24 hours at 28^{O}C. 1 ml of physiological saline was added to the bacterial growth and the culture was scraped with a sterile, thin, flexible inoculation loop. The suspension was pipetted into a sterile 2 ml Eppendorf tube. Additional 1 ml of physiological saline was added to the plate and the plate was swirled. This washing was also transferred to the Eppendorf tube.

About 1.8 ml of the saline was recovered as suspension. The suspension was then vortexed for 15 seconds to break up clumps and then centrifuged at 130000 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 with a density of A_{610} = 0.1 as read on an ELISA reader (Thermomax, Thermo Scientific) using sterile saline as blank.

3.2.3. Inoculation of Biolog plates :

As per the manufacturer’s instructions, 150 µl of the inoculum (A_{610} = 0.1) was added to each of the 96 wells of the Biolog plate. The culture was thoroughly mixed with the substrate and dye in the well by gently pipetting the culture up and down thrice. Two replicate plates were set up per strain. After the inoculation of all the 96 wells was complete, the plate was covered with its lid and sealed on the sides with Parafilm® to prevent the loss of suspending medium from the peripheral wells.

Parafilm® has high oxygen permeability and low permeability to water vapour and hence the presence of Parafilm® would not affect the overall oxygen concentration in the plate and in turn, the growth of bacteria.

3.2.4. Incubation :

The inoculated and sealed plates were incubated for 48 hours at 28^{O}C in an incubator under static conditions.

3.2.5. Reading the plates :

The plates were removed from the incubator at the end of the incubation period and read at 590 nm on an ELISA reader (Thermomax, Thermo Scientific). The absorbance value from the well A1, which did not contain any substrate and hence acted as the blank, was subtracted from all the other wells to get the final absorbance (as compared to the blank).

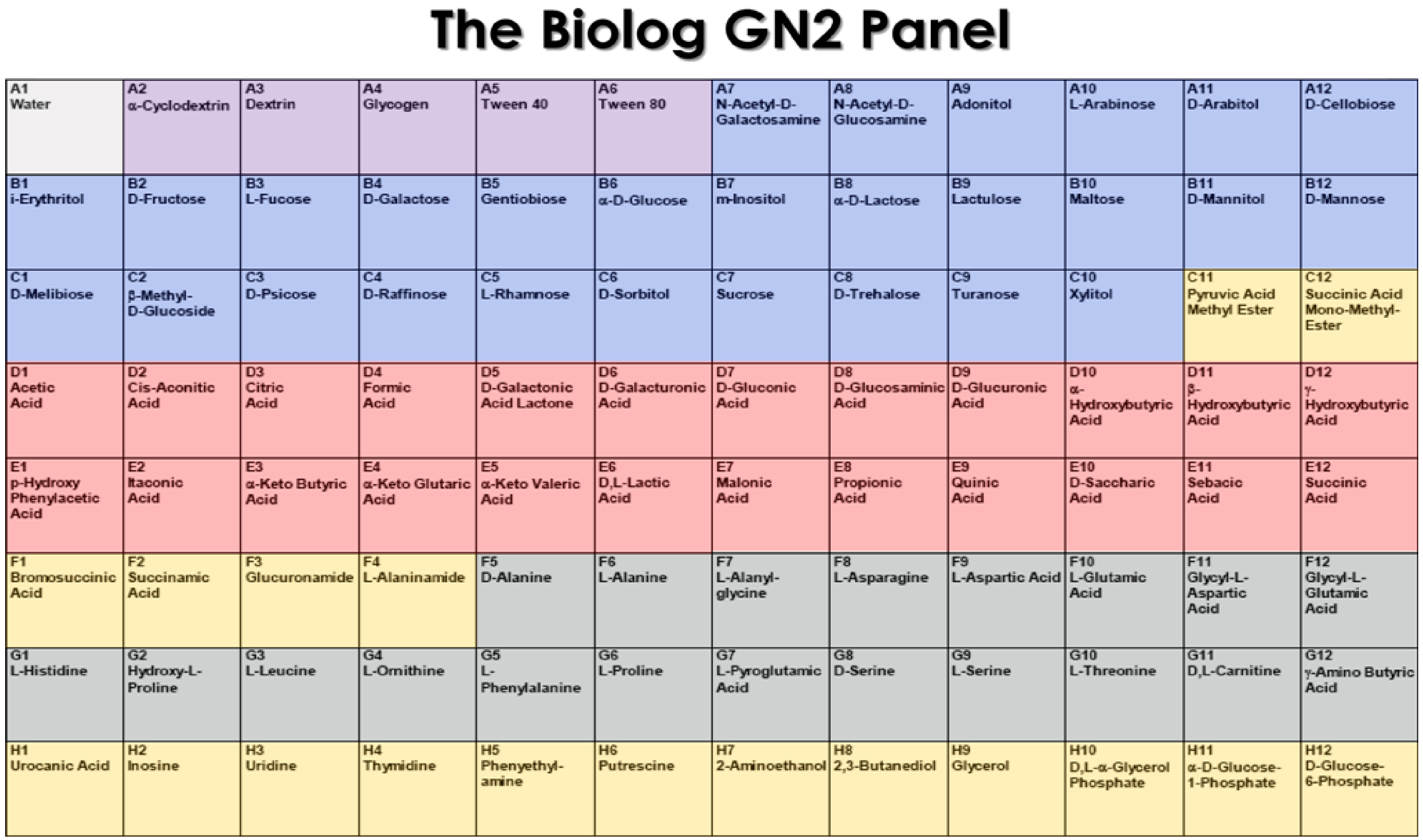
The software used for reading the plates was Softmax (V. 2.35, Molecular Devices Corp.). Since the software did not allow export of results file into a format that could be imported into a spreadsheet application, the results were printed and the absorbance values were manually entered into a Microsoft Excel 2003 spreadsheet.

3.2.6. 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 by the nearest neighbour approach and heatmaps were generated to study the level of substrate utilization.

3.3. Results :

The results of the Biolog metabolic fingerprinting experiments were analysed using multiple approaches. The data was converted into different formats depending on the aim of the analysis and the method used for the analysis. Only the substrates utilized by *Rlv*. 3841 that showed variation in utilization patterns amongst the Wentworth isolates were used for further studies.



The Biolog GN2 Panel

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 Pyruvic Acid Methyl Ester	C12 Succinic Acid Mono-Methyl-Ester
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-Hydroxybutyric Acid	D11 β-Hydroxybutyric Acid	D12 γ-Hydroxybutyric 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 Bromosuccinic 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 α-D-Glucose-1-Phosphate	H12 D-Glucose-6-Phosphate

Figure 3.1. The Biolog GN2 substrate panel. The substrates are colour coded as follows : Purple (polymers), Blue (sugars / sugar derivatives), Red (carboxylic / dicarboxylic acids), Green (amino acid / amino acid derivatives) and Yellow (miscellaneous intermediates of metabolism).

Table 3.1. Mean Absorbance Values for Biolog substrates utilized by *Rhizobium leguminosarum* biovar *viciae* strain 3841 and other strains.

Well ID	TRX01	TRX02	TRX03	TRX04	TRX05	TRX06	TRX07	TRX08	TRX09	TRX10	TRX11	TRX12	TRX13	TRX14	TRX15	TRX16	TRX17	TRX18
A09	0.2345	0.1275	0.4935	0.2630	0.5800	0.5885	0.5335	0.2260	0.0495	0.2625	0.2395	0.8605	0.0920	0.2815	0.4225	0.5430	0.7310	0.8980
B01	0.3485	0.1830	0.4215	0.0315	0.5045	0.5495	0.6365	0.0575	0.1245	0.3630	0.1855	0.8575	0.2475	0.4985	0.6220	0.6345	0.6260	0.9155
B11	0.1725	0.1315	0.5835	0.2555	0.4450	0.4795	0.5890	0.1645	0.1580	0.3030	0.1785	0.9140	0.0470	0.4000	0.4565	0.4670	0.6080	0.9445
C02	0.0540	0.1755	0.3730	0.0225	0.3775	0.2805	0.3945	0.0070	0.0230	0.1310	0.0660	0.6970	0.1335	0.1010	0.4630	0.2490	0.3935	0.9220
C04	0.0545	0.0905	0.3550	0.0875	0.2545	0.3090	0.2955	0.1580	0.1405	0.3185	0.0925	0.4040	0.0710	0.1880	0.3025	0.2655	0.4450	0.7390
C05	0.2885	0.1655	0.5985	0.2310	0.5910	0.5515	0.5605	0.1910	0.1090	0.2190	0.3345	0.9980	0.1690	0.3710	0.6385	0.5600	0.7475	0.9905
C10	0.1130	0.1870	0.3005	0.0830	0.3085	0.4580	0.4075	0.1530	0.1550	0.1800	0.1450	0.6365	0.1980	0.1940	0.3850	0.3545	0.5340	0.8950
D05	0.0595	0.0540	0.3850	0.1345	0.2425	0.3175	0.2055	0.3970	0.1280	0.1760	0.3270	0.6520	-0.0815	0.2730	0.5915	0.3715	0.4165	0.0470
D07	0.1075	0.1280	0.3300	0.0690	0.2085	0.2775	0.1955	0.1385	0.1260	0.1470	0.1805	0.4665	-0.0105	0.1660	0.2925	0.2615	0.3830	0.1675
D08	0.0585	0.0310	0.1635	0.0660	0.1560	0.2055	0.1595	0.0615	0.0105	0.0865	0.0470	0.4715	-0.0760	0.0815	0.2070	0.1745	0.2870	0.0635
D12	0.2510	0.2160	0.5630	0.2305	0.4190	0.4900	0.6000	0.1930	-0.0100	0.0090	0.0175	0.8120	0.2975	0.4520	0.0750	0.5370	0.4850	0.0470
E04	0.0215	0.0815	0.1045	0.0040	0.1695	0.2330	0.1320	0.0785	0.0670	0.1260	0.0810	0.0705	0.1805	0.0975	0.1315	0.1660	0.2320	0.1550
E12	0.2695	0.2255	0.5410	0.2385	0.3285	0.3770	0.4910	0.2930	0.2760	0.3680	0.1640	0.8400	0.3505	0.4235	0.4590	0.5680	0.5365	0.7945
F05	0.0140	0.0465	0.1405	-0.0515	0.0230	0.1605	0.2575	0.0315	-0.0330	0.0275	-0.0115	0.4090	-0.0385	-0.0050	0.0580	0.1465	0.2685	0.5565
F06	0.1910	0.0665	0.4115	0.0190	0.2600	0.3105	0.4000	0.1180	0.0925	0.2200	0.0970	0.3210	0.0315	0.0820	0.0345	0.3875	0.6230	0.7365
G07	0.0290	0.0390	0.0400	-0.0565	0.3890	0.4135	0.6225	0.0025	-0.0240	0.0550	-0.0220	0.6895	-0.0110	0.3930	0.4680	0.4595	0.0195	0.8865
G09	0.0305	0.0675	0.5095	-0.0570	0.0680	0.5135	0.0825	-0.0260	-0.0180	0.0305	-0.0420	0.1475	-0.0710	0.0255	0.0810	0.1390	0.5020	1.1175
H01	0.0695	0.1350	0.1895	0.0465	0.2200	0.2870	0.1815	0.1275	0.0755	0.1250	0.2285	0.3910	0.2395	0.2440	0.2815	0.3955	0.4380	0.7065
H02	0.0015	0.0065	0.0410	-0.0145	0.1390	0.0920	0.1525	0.0405	0.0115	0.0590	0.0370	0.0155	-0.0325	0.0460	0.0110	0.1970	0.3075	0.2940
H03	0.0590	0.0405	0.3765	0.0195	0.4390	0.5015	0.5435	0.1110	0.1070	0.2475	0.1550	0.3385	0.1030	0.2495	0.1260	0.5330	0.7875	0.7875
H04	0.0355	0.0725	0.1460	-0.0050	0.1085	0.2340	0.2315	0.0305	0.0295	0.0435	0.0460	0.1740	0.1095	0.1025	0.2125	0.2640	0.4670	0.5200

Well ID	TRX19	TRX20	TRX21	TRX22	TRX23	TRX24	TRX25	TRX26	TRX27	TRX28	TRX29	TRX30	TRX31	TRX32	TRX33	TRX34	TRX35	TRX36
A09	0.2455	0.5600	0.5300	0.5460	0.7395	0.6385	0.4225	0.6045	0.5460	0.5970	0.5435	0.6380	0.4485	0.4370	0.1115	0.7060	0.7095	0.7545
B01	0.4405	0.6125	0.2140	0.6765	0.7635	0.4525	0.5995	0.5890	0.4380	0.6825	0.4770	0.6035	1.3205	0.6905	0.7300	0.4810	0.7325	0.6880
B11	0.1110	0.3025	0.2575	0.7025	0.8015	0.6370	0.5115	0.5140	0.6095	0.7005	0.5195	0.7360	1.4150	0.6570	0.1000	0.7235	0.7855	0.6555
C02	0.2305	0.2480	0.2135	0.4590	0.3190	0.2545	0.5145	0.2565	0.5820	0.4615	0.3825	0.4340	1.3110	0.6055	0.5985	0.4090	0.7640	0.5060
C04	0.2170	0.3420	0.2840	0.5305	0.2510	0.2100	0.3220	0.1880	0.5065	0.4020	0.4635	0.5640	0.9705	0.4465	0.3385	0.4540	0.4295	0.5370
C05	0.1900	0.4040	0.3985	0.6710	0.5805	0.4625	0.6560	0.5490	0.5145	0.6370	0.4875	0.6035	1.4540	0.7120	1.1265	0.4440	0.9135	0.6970
C10	0.4530	0.5305	0.4655	0.4675	0.4390	0.4110	0.3850	0.6715	0.4245	0.3745	0.3435	0.3195	1.2470	0.5555	0.5700	0.4680	0.6665	0.5190
D05	0.1520	0.3785	0.3715	0.2555	0.4555	0.2030	0.2500	0.3045	0.6870	0.4075	0.4300	0.2390	0.5450	0.6495	0.0035	0.6095	0.6360	0.5485
D07	-0.2630	0.3335	0.2775	0.4140	0.2735	0.1780	0.2945	0.2305	0.6335	0.2815	0.4695	0.2385	0.5210	0.6715	0.0790	0.4420	0.5295	0.4870
D08	0.2065	0.2365	0.2245	0.2365	0.2400	0.1765	0.0655	0.1350	0.4690	0.1665	0.2810	0.1915	0.1895	0.6050	0.0380	0.3125	0.5630	0.3505
D12	0.4540	0.5785	0.6060	0.0280	0.7225	0.5965	0.0440	0.6190	0.0745	0.6355	0.0105	0.5210	0.2560	0.1585	0.9840	0.0220	0.7055	0.6535
E04	0.1785	0.2225	0.1645	0.1515	0.1010	0.0830	0.1015	0.1660	0.1445	0.1240	0.1030	0.1795	0.7240	0.2635	0.5170	0.1010	0.3510	0.2135
E12	0.3960	0.5400	0.6535	0.3925	0.6570	0.5190	0.5280	0.6035	0.4030	0.5490	0.5200	0.6350	1.3505	0.4435	0.6480	0.6445	0.7275	0.5520
F05	0.1470	0.1540	0.0985	0.2440	0.1615	0.1210	0.0695	0.1030	0.2200	0.1795	0.0960	0.1545	0.4220	0.4045	0.0455	0.1380	0.3925	0.1585
F06	0.1550	0.1975	0.2100	0.5000	0.3860	0.2650	0.1935	0.1265	0.3190	0.4015	0.3415	0.3025	0.3590	0.5285	0.1515	0.4050	0.7270	0.5045
G07	-0.0465	0.0945	0.0720	0.0445	-0.1210	-0.0140	0.4840	0.5200	0.3885	0.6510	-0.0100	0.4195	0.1825	0.6315	0.0395	0.3630	0.7215	0.5720
G09	0.0540	0.0760	0.0500	0.4585	0.2045	0.1470	0.3485	0.0375	0.5945	0.5745	0.4130	0.5765	1.0805	0.3465	-0.0115	0.0685	0.8065	0.1975
H01	0.3530	0.3440	0.2630	0.1695	0.1685	0.0715	0.2805	0.1845	0.5120	0.2115	0.2635	0.1315	0.9535	0.4370	0.7235	0.5170	0.4970	0.5115
H02	0.0440	0.0980	0.0155	0.2665	0.1395	0.0860	-0.1615	0.0220	0.1975	0.1280	0.1495	0.2355	0.6440	0.2590	0.1730	0.1675	0.4790	0.1990
H03	0.0845	0.3225	0.2260	0.5235	0.5110	0.4360	0.2770	0.4400	0.3750	0.5610	0.4980	0.4065	0.9525	0.4480	0.3630	0.3615	0.7055	0.4185
H04	0.2310	0.2975	0.2300	0.1660	0.1835	0.1450	0.1745	0.1650	0.2050	0.1745	0.1020	0.1570	0.7105	0.3825	0.5560	0.2370	0.4055	0.3665

Well ID	VSX01	VSX02	VSX03	VSX04	VSX05	VSX06	VSX07	VSX08	VSX09	VSX10	VSX11	VSX14	VSX15	VSX16	VSX17	VSX18	VSX19	VSX21
A09	0.5090	0.3980	0.7120	0.2150	0.9325	0.5955	0.6740	0.3480	0.3625	0.6540	0.1335	0.6615	0.5790	0.7680	0.5005	0.9065	0.4470	0.5500
B01	0.3830	0.3555	0.6540	0.3035	0.8600	0.2595	0.5395	0.4305	0.7815	0.5750	0.3235	0.7080	0.5010	0.8205	0.5855	0.9680	0.3655	0.3485
B11	0.4325	0.3770	0.5375	0.3045	0.8470	0.3630	0.5585	0.3520	0.7050	0.4945	0.1835	0.7205	0.5340	0.8410	0.4520	1.0045	0.3155	0.4375
C02	0.2545	0.1285	0.5500	0.1440	0.7640	0.4900	0.5345	0.1255	0.6025	0.3835	0.0740	0.5655	0.2560	0.6505	0.3445	0.8785	0.3055	0.3450
C04	0.2150	0.1460	0.3965	0.1670	0.3880	0.2485	0.4115	0.1825	0.5780	0.1450	0.0775	0.3615	0.3535	0.4585	0.2620	0.5955	0.2840	0.1910
C05	0.3505	0.4760	0.1535	0.3310	0.9735	0.6015	0.7345	0.2605	0.8250	0.6560	0.1685	0.7500	0.4355	0.8235	0.5225	0.9730	0.4505	0.5240
C10	0.4305	0.1385	0.3550	0.1090	0.5690	0.1845	0.4060	0.2580	0.5655	0.2835	0.1680	0.5625	0.3560	0.5195	0.3515	0.8335	0.5585	0.4385
D05	0.1970	0.4120	0.3150	0.1285	0.5040	0.2665	0.3565	0.2540	0.5375	0.5550	0.2620	0.5915	0.5925	0.4535	0.4120	0.9985	0.6840	0.4205
D07	0.1280	0.1585	0.2180	0.1205	0.3325	0.1175	0.2485	0.1720	0.2595	0.1925	0.1885	0.3840	0.5010	0.4605	0.3600	0.8555	0.6255	0.2520
D08	0.0335	0.0815	0.1880	0.0235	0.2975	-0.0090	0.2165	0.0810	0.2490	0.1175	0.0805	0.2865	0.3235	0.3405	0.2540	0.8700	0.2385	0.1110
D12	0.0265	0.0615	0.4960	0.0435	0.8640	0.0325	0.0370	-0.0025	0.0335	0.0700	0.1500	0.0815	0.4860	0.1445	0.0665	0.0895	0.0575	0.0165
E04	0.1060	0.0325	0.1505	0.0750	0.2630	0.0670	0.1380	0.1200	0.1555	0.0855	0.1080	0.1785	0.1065	0.3910	0.1435	0.1640	0.3325	0.0520
E12	0.3960	0.2965	0.3220	0.2580	0.6325	0.2950	0.5055	0.3715	0.6780	0.4500	0.3060	0.7055	0.4405	0.7085	0.4865	0.8950	0.6570	0.5365
F05	0.1010	0.0060	0.0305	-0.0070	0.1510	-0.0640	0.1980	-0.0225	0.2110	0.0045	-0.0105	0.2635	0.1110	0.2440	0.0485	0.6020	0.1845	0.1115
F06	0.1280	0.0460	0.1155	0.0135	0.2120	0.0700	0.3285	0.2115	0.2675	0.1245	0.0270	0.2785	0.1140	0.4915	0.2875	0.7555	0.4775	0.1590
G07	0.3350	-0.0045	0.2610	0.0145	0.3540	0.4035	0.5290	0.2220	0.8015	0.5440	0.2290	0.6890	0.4265	0.0890	0.4410	0.8205	0.0660	0.1870
G09	0.0600	0.0285	0.0145	0.0185	0.5410	-0.0300	0.5615	0.0200	0.4375	0.0145	-0.0100	0.5905	0.0915	0.5465	0.0875	0.9750	0.1015	0.0340
H01	0.1620	0.0995	0.3060	0.1005	0.3990	0.2420	0.3185	0.3385	0.4305	0.3285	0.3250	0.3065	0.2170	0.5250	0.4195	0.7395	0.4930	0.2680
H02	0.0865	0.0080	0.0160	0.0365	0.2720	0.0395	0.0620	0.0445	0.1140	0.0450	0.0235	0.1195	0.0055	0.3195	0.1395	0.5595	0.0750	0.0135
H03	0.2250	0.1040	0.5255	0.2065	0.5170	0.4685	0.5550	0.1515	0.6450	0.4270	0.0625	0.5350	0.1165	0.7810	0.3675	0.7840	0.2095	0.2600
H04	0.1740	0.0360	0.1900	0.0305	0.3960	0.0665	0.1775	0.0785	0.2125	0.1265	0.0675	0.1760	0.1545	0.0010	0.1645	0.5180	0.3705	0.1500

Well ID	VSX22	VSX23	VSX24	VSX25	VSX26	VSX27	VSX28	VSX29	VSX30	VSX31	VSX32	VSX33	VSX34	VSX35	VSX36	VSX37	VSX38	VSX39	3841
A09	0.4825	0.6920	0.6550	0.4125	0.8575	0.7920	0.6590	0.6930	0.6665	0.6715	0.5860	0.7940	0.7250	0.6630	0.6290	0.8875	0.9315	0.5270	0.1220
B01	0.2695	0.5255	0.6305	0.4495	0.8165	0.7850	0.6310	0.7510	0.6410	0.4810	0.7015	0.9200	0.6240	0.6010	0.5620	0.8080	0.9055	0.5670	0.3815
B11	0.5120	0.7095	0.6200	0.3820	1.0470	0.9085	0.7275	0.7765	0.4395	0.5690	0.6205	0.8715	0.3000	0.6405	0.5055	0.8705	1.0280	0.3705	0.1835
C02	0.2410	0.5640	0.4265	0.2930	0.7845	0.6690	0.5640	0.6105	0.6430	0.5985	0.4735	0.6595	0.3030	0.5155	0.6185	0.7950	0.8565	0.5500	0.1845
C04	0.1975	0.4735	0.4910	0.2910	0.5300	0.4480	0.5530	0.4470	0.4780	0.4335	0.4730	0.5980	0.2605	0.5950	0.4540	0.6060	0.5380	0.3115	0.1760
C05	0.4395	0.6275	0.8815	0.4575	0.8810	0.8330	0.6620	0.7690	0.7245	0.6660	0.6185	0.7880	0.4550	0.6715	0.7055	0.9635	0.9345	0.6025	0.2845
C10	0.2905	0.4785	0.2910	0.2530	0.7195	0.5330	0.5035	0.4505	0.4255	0.3795	0.4640	0.5740	0.6175	0.6085	0.4790	0.7285	0.7575	0.3645	0.1940
D05	0.3370	0.4865	0.2935	0.0260	0.6575	0.5110	0.3570	0.4135	0.4125	0.4945	0.4885	0.6410	0.5620	0.5880	0.5715	0.6915	0.7740	0.4455	0.1160
D07	0.2130	0.3400	0.5765	0.1005	0.6065	0.5410	0.2155	0.4020	0.3975	0.3280	0.4155	0.6435	0.2405	0.2710	0.3170	0.6095	0.4815	0.1675	0.1425
D08	0.1460	0.3185	0.1505	0.0675	0.5160	0.4935	0.1475	0.3235	0.3525	0.2655	0.3010	0.4250	0.1375	0.1760	0.2850	0.5015	0.4710	0.2045	0.0545
D12	0.0140	0.0675	0.0300	0.0110	0.1830	0.2740	0.0670	0.1265	0.1680	0.0610	0.0260	0.8075	0.0020	0.0405	0.0850	0.1630	0.0645	-0.0135	0.2665
E04	0.1355	0.3300	0.0795	0.0115	0.5305	0.4620	0.1485	0.2050	0.3520	0.1910	0.3275	0.0390	0.1750	0.0935	0.1225	0.3995	0.1435	0.0600	0.0915
E12	0.5320	0.5550	0.5870	0.4720	0.9430	0.8640	0.5730	0.7535	0.7390	0.4625	0.5495	0.9050	0.5600	0.5575	0.4830	0.7605	0.8830	0.3715	0.2985
F05	0.1295	0.2420	0.3335	0.0155	0.4395	0.4390	0.2860	0.2530	0.3295	0.1705	0.1625	0.2785	0.2545	0.1280	0.1430	0.2905	0.3480	0.1130	0.1050
F06	0.0980	0.3865	0.4305	0.2790	0.5310	0.3965	0.3100	0.2905	0.3540	0.2760	0.2210	0.6885	0.5095	0.2650	0.4375	0.4695	0.5410	0.3840	0.2510
G07	-0.0045	0.5950	0.0565	-0.0320	0.1590	0.1240	0.7615	0.0625	0.0755	0.0570	0.5990	0.8645	0.4075	0.5980	0.5815	0.0695	0.8760	0.3870	0.3250
G09	0.0245	0.6845	1.0420	0.0200	0.8690	0.8255	0.5480	0.6950	0.7135	0.5625	0.0675	1.0570	0.0240	0.0930	0.5555	0.7425	0.8310	0.3555	0.0780
H01	0.1550	0.2060	0.1570	0.0830	0.4420	0.4085	0.2805	0.3565	0.3140	0.2525	0.4095	0.3565	0.4340	0.5410	0.4555	0.5885	0.6045	0.3810	0.1325
H02	0.0075	0.1735	0.2810	-0.0110	0.3070	0.7205	0.1295	0.1420	0.9295	0.1800	0.1595	0.2005	0.0495	0.0615	0.1555	0.6680	0.1555	0.1390	0.0690
H03	0.2895	0.6600	0.7685	0.3880	0.9585	0.8070	0.6540	0.6660	0.7415	0.5760	0.4530	0.5700	0.3755	0.5505	0.5895	0.8225	0.8675	0.5365	0.0550
H04	0.1495	0.1955	0.2970	0.1015	0.4110	0.3380	0.2685	0.2905	0.2285	0.2270	0.3665	0.2915	0.1655	0.2465	0.2995	0.4295	0.3810	0.2200	0.0715

Average absorbance reading for the Biolog substrates utilized by Rlv 3841 and other strains thereof, used for correlation analysis.

The Biolog GN2 consists of 96 wells that contain 95 different carbon substrates and a control well. Of the 95 carbon substrates, 32 substrates were not utilized by any whereas 19 were utilized by all the 72 Wentworth strains. Of the 44 remaining substrates, 21 substrates were utilized by *Rlv*. 3841. The data from these strains would allow for easier assessment of phenotype-genotype relationship and were hence used for some analysis.

In order to check for patterns in the utilization of the substrates incorporated into the Biolog panel, the absorbance data from the Biolog plates was 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 (Table 3.2). Since the synthesis of mucopolysaccharide may lead to increased absorbance values, visual verification of substrate utilization was done by checking for the presence of reduced tetrazolium dye precipitate at the bottom of the microtitre plate.

The binary data was used to generate **U**nweighted **P**air **G**roup **M**ethod with **A**rithmetic Mean (UPGMA) trees. UPGMA is a agglomerative or hierarchical clustering method based on pairwise similarities of variables, assuming a constant rate of change (Sokal and Michener, 1958). UPGMA analysis first produces a similarity (or dissimilarity) matrix and this matrix is then used to generate a rooted tree or dendrogram in which two clusters that are most similar (or least distant) are combined into a higher-level cluster. This method can therefore be used to study relationships between phenotypes by creating ‘phenetic trees’ or ‘phenograms’ (Legendre and Legendre, 2012).

Two UPGMA trees were created to analyse the substrate utilization patterns. The first tree was created to cluster strains and the second tree to cluster substrates with similar utilization patterns. The two UPGMA trees were used to order the substrates and the strains to generate a matrix ordered horizontally by strains and vertically by substrate. The absorbance data of the 21 substrates was scaled from 0 to 1 and used to generate a heatmap for patterns as shown in the following section.

The data was also used to perform Principal Coordinate Analysis (PCA) to identify strains clustered as a result of similar substrate utilization pattern.

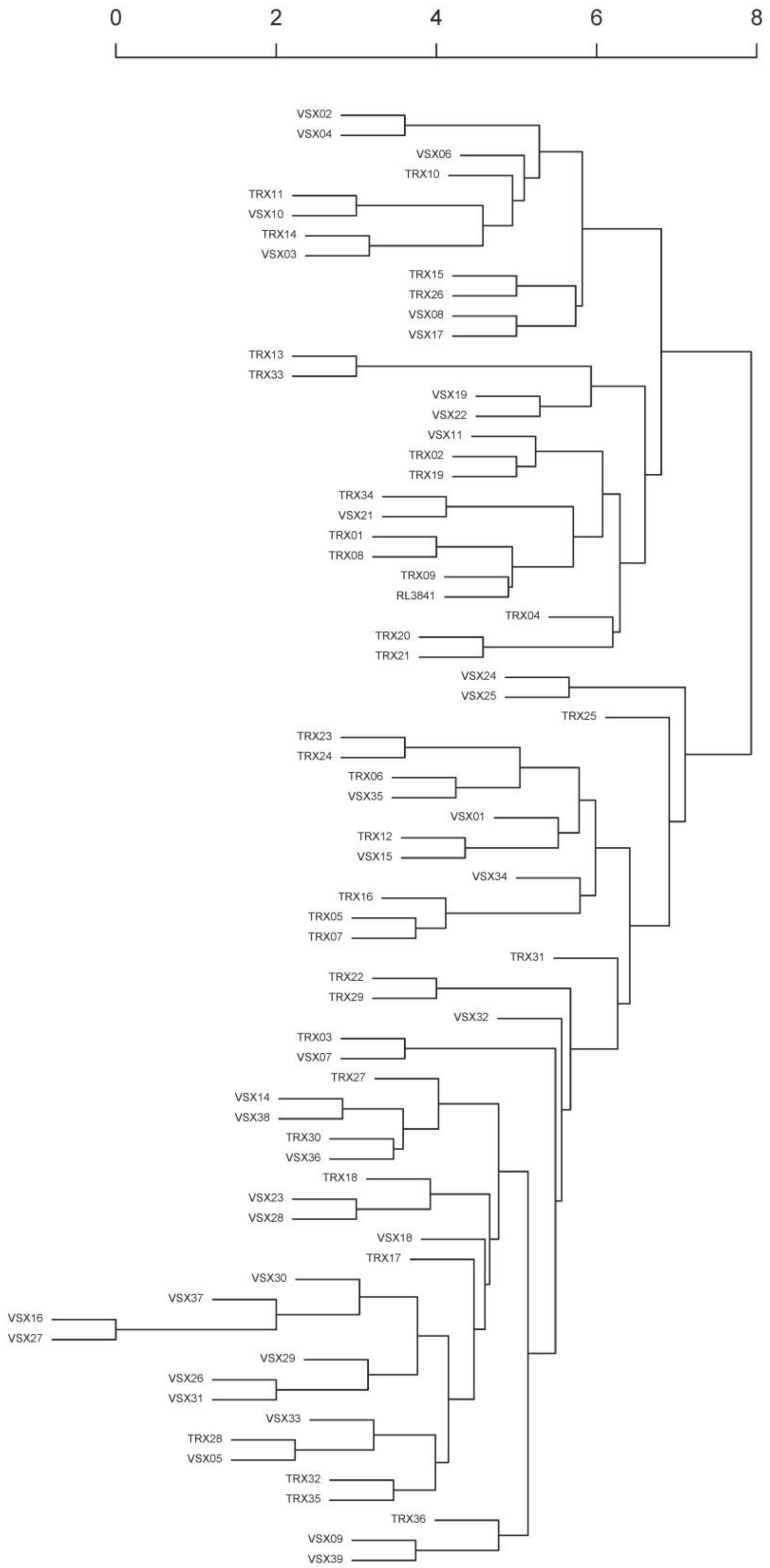


Figure 3.2. Hierarchical clustering of Wentworth strains based on scaled Biolog readings. Distance/Similarity Measure = Euclidean Distance, Cluster Method = Group Average.

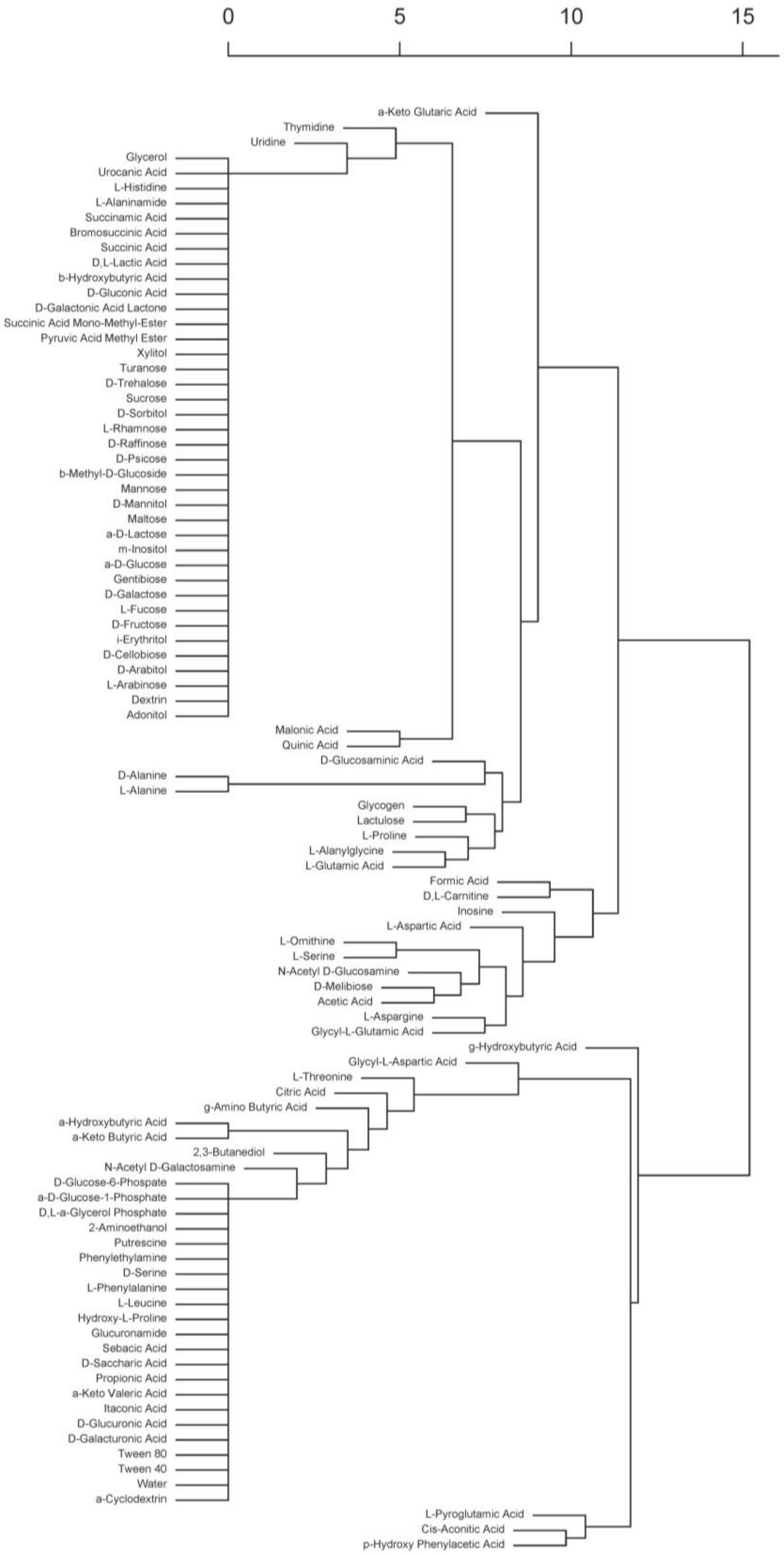


Figure 3.3. Hierarchical clustering of substrates based on scaled Biolog readings. Distance/Similarity Measure = Euclidean Distance, Cluster Method = Group Average.

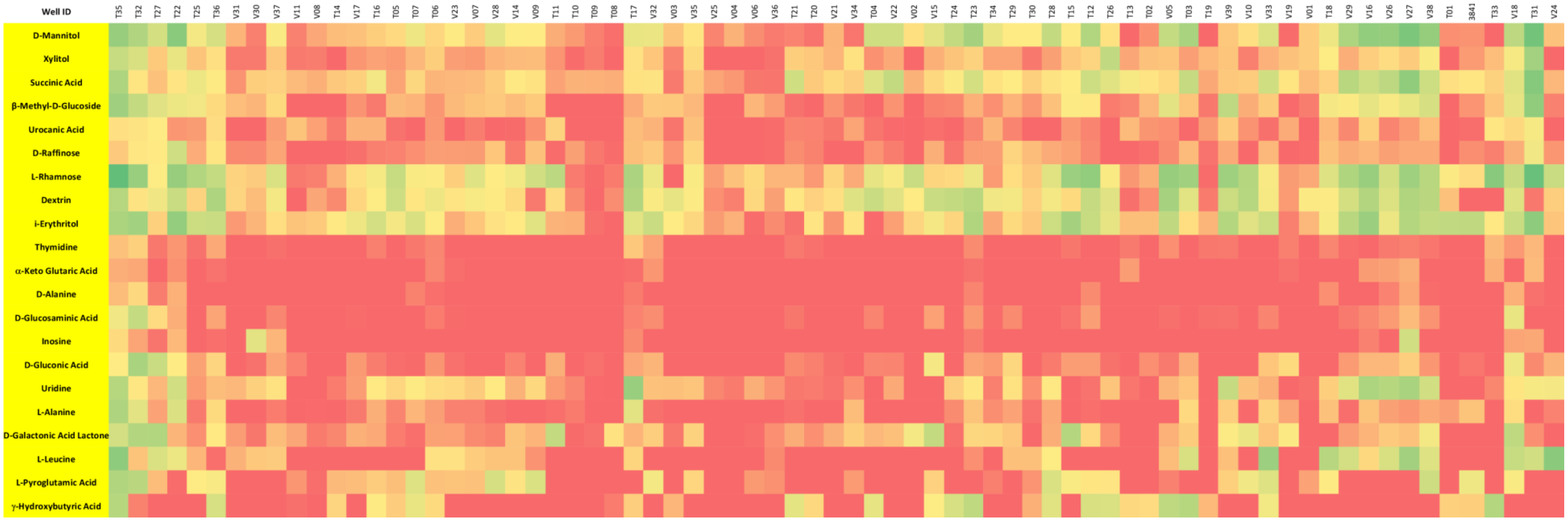


Figure 3.4. Figure showing the relative utilization of the 21 substrate by Wentworth isolates as a heatmap. Green areas denote good utilization with a gradual decrease in utilization till red, which denotes no utilization of the substrate.

Table 3.2. Binary data used for the analysis of substrates utilized by *Rlv*. 3841 and showing variation in utilization by Wentworth isolates.

	B01	C05	E12	A09	B11	H01	H03	C10	C02	H04	D05	D07	C04	F06	F05	E04	D08	G07	H02	G09	D12
TRX01	1	1	1	1	1	1	1	0	0	1	0	0	0	1	1	0	0	0	1	0	1
TRX02	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0	1
TRX03	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1
TRX04	1	1	1	1	1	1	0	0	0	0	1	1	0	1	1	0	1	0	0	0	1
TRX05	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1
TRX06	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
TRX07	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
TRX08	0	1	1	1	1	1	1	0	0	1	1	1	0	1	1	0	0	0	1	0	1
TRX09	1	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	0	0
TRX10	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	1	0	0	0	0	0
TRX11	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0
TRX12	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1
TRX13	1	1	1	1	0	1	1	1	1	1	0	0	0	1	0	1	0	0	0	0	1
TRX14	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	1
TRX15	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	0	0	0
TRX16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1

TRX17	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
TRX18	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	0	1	1	1	0
TRX19	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0	0	1
TRX20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1
TRX21	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1
TRX22	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0
TRX23	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
TRX24	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	0	0	1	1
TRX25	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	1	0
TRX26	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	0	1
TRX27	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
TRX28	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
TRX29	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0
TRX30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
TRX31	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	1	1	0
TRX32	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
TRX33	1	1	1	0	0	1	1	1	1	1	0	0	1	0	0	1	0	0	0	0	1
TRX34	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0
TRX35	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

TRX36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
VSX01	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0
VSX02	1	1	1	1	1	1	1	0	1	1	1	0	1	0	0	0	0	0	1	0	0
VSX03	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1
VSX04	1	1	1	1	1	1	1	0	1	1	0	1	1	0	0	1	0	0	1	0	0
VSX05	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
VSX06	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	0
VSX07	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0
VSX08	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	0	0	0
VSX09	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0
VSX10	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0
VSX11	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0	1	0	1	0	0	0
VSX14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
VSX15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1
VSX16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
VSX17	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0
VSX18	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
VSX19	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
VSX21	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0

VSX22	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
VSX23	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
VSX24	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	0	1	1	0
VSX25	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	0	0	0	0	0
VSX26	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
VSX27	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
VSX28	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
VSX29	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0
VSX30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0
VSX31	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0
VSX32	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
VSX33	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
VSX34	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
VSX35	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0
VSX36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0
VSX37	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
VSX38	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
VSX39	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0

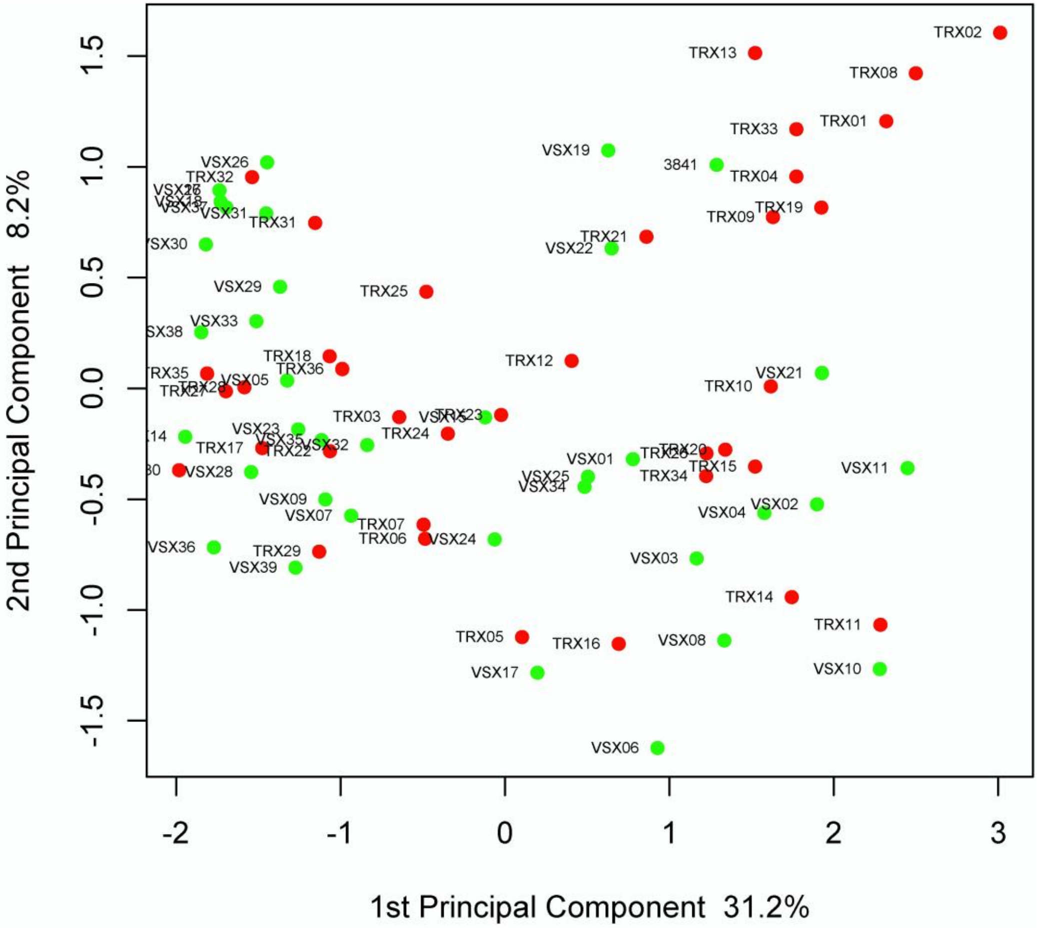


Figure 3.5a. PCA analysis of Biolog data by biovars. The analysis was done observe clustering of strains showing similar metabolic profiles. The strains are widely distributed indicating heterogeneity in substrate utilization. Biovar trifolii plotted in red and biovar viciae in green.

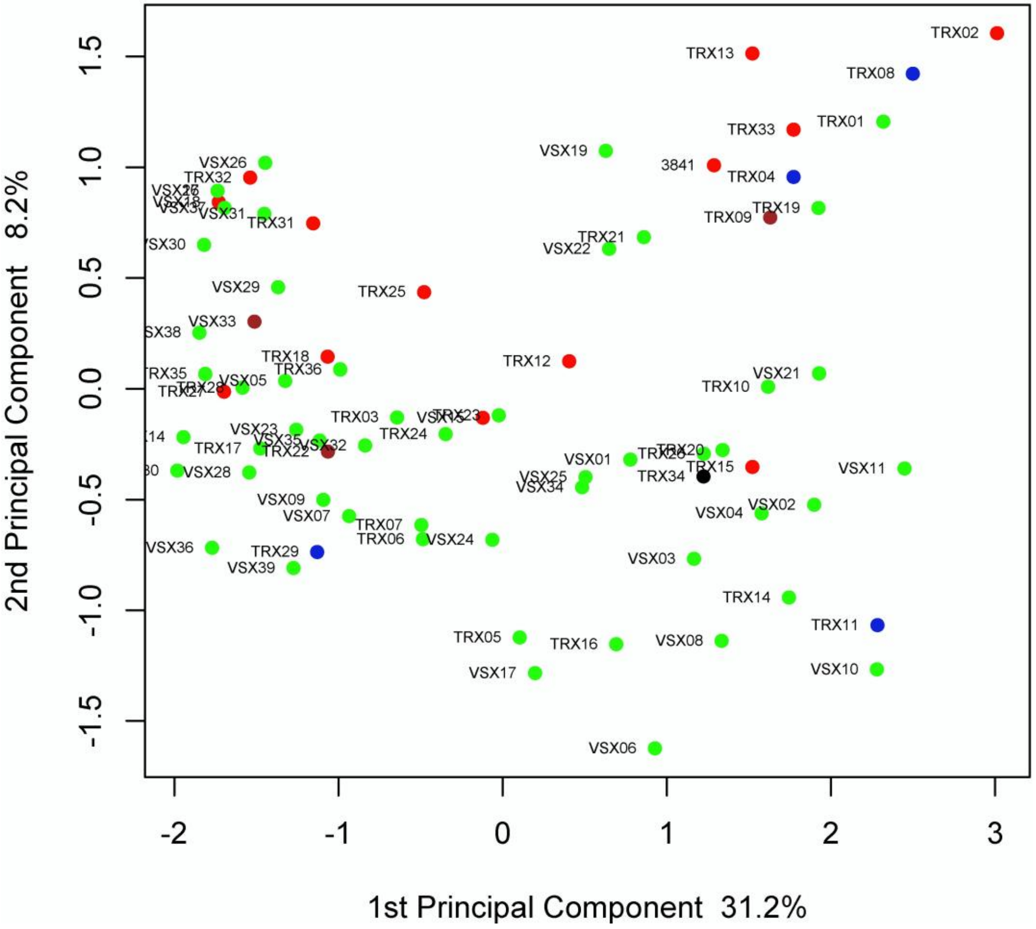


Figure 3.5b. PCA analysis of Biolog data by five cryptic species. The strains are widely distributed indicating heterogeneity in substrate utilization. The strains included in 5A, 5B, 5C, 5D and 5E are shown in black, red, green, blue and brown respectively.

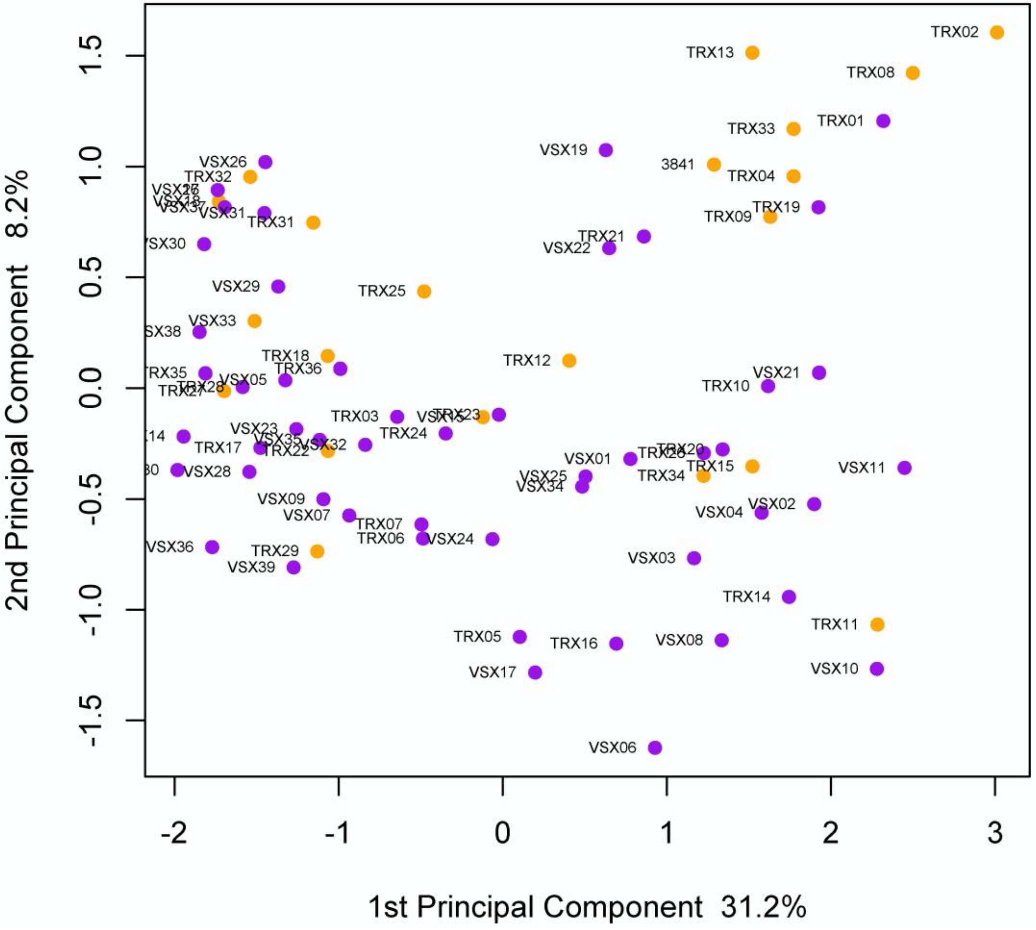


Figure 3.5c. PCA analysis of Biolog data by two cryptic species. The strains are widely distributed indicating heterogeneity in substrate utilization. The strains included in 2A and 2B are shown in orange and purple respectively.

3.4. Discussion :

The aim of this chapter was to study the phenotypic diversity of the Wentworth isolates using a robust method for study. The diversity was studied using the Biolog GN2 system. The Biolog system's unique redox chemistry based system measures bacterial respiration and reports it colorimetrically by reduction of a dye. Based on the utilization / non-utilization of substrates, a pattern or "metabolic fingerprint" was obtained.

The Biolog system is not easily amenable for use with *R. leguminosarum.* The protocol for using the Biolog system with *R. leguminosarum* was standardized and used to study the metabolic differences within the Wentworth isolates. The study of the metabolic pattern or fingerprints shows that the fingerprint obtained for each of the strains was unique. This indicates a high metabolic diversity. This is important given that the isolates belonged to the same species (although two different biovars) and were recovered from a small area of a meter squared. This shows that even bacterial populations belonging to same species and isolated from a very restricted area can show a lot of diversity in terms of its metabolic profile reflected in terms of its phenotype.

Analysis of the utilization pattern by constructing UPGMA trees based on scaled absorbance reading. Whilst the strains (Figure 3.2) did not show any specific clustering pattern in terms of segregating the isolates with respect to the type of biovar or their position in the cryptic species clades in the SplitsTree network tree. The UPGMA tree for substrate (Figure 3.3) however, showed that the isolates were more similar in their ability to use sugars and most mono- and di- carboxylic acids since they were largely utilized. The strains showed diversity in their ability to use amino acids and amino acid derivatives . Most polysaccharides and miscellaneous metabolites (especially the more electronegative metabolites) also did not show a lot of variation since they were mostly not utilized.

Of the 21 substrates that were utilized by *Rlv*. 3841 and that showed a variation in utilization pattern in the Wentworth isolates, at least 6 substrates (H04 - Thymidine, E04 - α-keto glutarate, F05 – D-alanine, D08 – D-Glucos-amine, H02 - Inosine and D07-D-gluconate) were largely unutilized by the Wentworth isolates and can be seen as the central red band in Figure 3.4.

The Principal Components Analysis (PCA) of Biolog data too did not show a very striking pattern that could be said to differentiate the two biovars (Figure 3.5). Although certain regions of the plot appeared to have more of *trifolii* isolates and other regions appeared to have more *viciae* isolates, there was no clear demarcation or boundary separating the two.

The overall impression from the Biolog analysis is that the Wentworth isolates are diverse in their metabolic abilities. It would therefore be interesting to see how the variation can be explained in terms of its gene composition. Hence, the next step in the study would be to study the metabolic pathways of these substrates and then establish a correlation between the utilization of the substrates and presence and absence of genes using various tools and techniques.

CHAPTER 4 : ANALYSIS OF PHENOTYPE DATA TO IDENTIFY GENES INVOLVED IN METABOLISM OF SUBSTRATES IN THE BIOLOG GN2 PLATES

4.1. Introduction :

A metabolic pathway constitutes the conceptual unit of metabolism consisting of an ordered set of interconnected, directed biochemical reactions. A pathway forms a coherent unit with boundaries at high-connectivity substrates that are regulated as a singular unit. The metabolic reactions are catalysed by enzymes encoded by the genome of the organism. Since the sequencing of the first free-living microbe by Fleischmann et al. (1995), more and more organisms are being sequenced and sequence data generated. However, there is now an increased awareness about the paucity of our knowledge with respect to the functions of the genes present in these genome sequences.

Functional analysis of the genome emphasizes gene expression studies (transcriptomics) and protein profiling (proteomics). Most metabolic analysis aim at specific substrates or products that constitute a part of the metabolic pathway. Information from metabolic pathways can be integrated with the other fields to develop a more global view of cell function.

Most sequenced genomes feature a large number of genes with unknown function. Researchers have used software to predict the functions of DNA sequences ("in-silico" biology). A number of databases have been created in recent years to store and organize the ever increasing amount of data on metabolism.

Pathguide - the Pathway Resource List (http://pathguide.org), a meta-database, provides an overview of hundreds of web-accessible biological pathway and network databases including metabolic pathways, signalling pathways, transcription factor targets, gene regulatory networks, genetic interactions, protein-compound interactions, and protein-protein interactions. Pathguide is useful as a starting point for biological pathway analysis and for content aggregation in integrated biological information systems. In September 2013, Pathguide, listed biological pathway resources, contained 133 currently available biological pathway resources related to metabolic data (Bader et al., 2006).

The increasing amount of metabolic data within the databases is increasing making it increasingly difficult to search the data and analyse it manually. Hence it is important to have interfaces that will automatically identify relevant pieces of data. Most metabolic databases provide an application interface which allows the user to find or predict pathways by using data such as the starting and end metabolites or the enzymes participating in it. Based on these two criteria, these databases can broadly be divided into two main types viz. the metabolic pathway databases that organise reactions into pathways and focus on their relationships and enzyme databases that mostly deal with the individual enzymes and their properties.

4.1.1. Metabolic pathway databases :

Two of the largest online databases that focus on metabolic pathways are the Kyoto Encyclopedia of Genes and Genomes (KEGG) PATHWAY database (Kanehisa et al., 2008) and MetaCyc database which is in turn a part of the larger BioCyc Database Collection (Caspi et al., 2012).

4.1.1.1. The KEGG PATHWAY Database :

KEGG PATHWAY is a collection of manually drawn pathway maps derived from textbooks, literature and knowledge from experts. From the manually curated pathways, organism-specific pathways are automatically generated using ortholog identifiers in the KEGG GENE database. The information within the PATHWAY database is divided into seven groups viz. Metabolism, Genetic Information Processing, Environmental Information Processing, Cellular Processes, Organismal Systems, Human Diseases and Drug Development. The database, which was first described by Ogata et al. (1998) now contains over a 150 reference pathways in the Metabolism category alone.

The reference pathways of KEGG are made of enzymatic reactions and chemical compounds constituting the pathway. Every reaction and compound found in KEGG has a unique identifier, starting with "R" for reaction or "C" for compound followed by five digits. The reference pathways reflects the union of the reactions and compounds found across all organisms from which it is then possible to select an organism of interest. The database then highlights the reactions and compounds found in that organism.

4.1.1.2. The MetaCyc Database :

The MetaCyc Database was first described by Karp et al. (2000). It is an independent metabolic pathway database. However, like KEGG, it too contains manually curated information from scientific literature. Currently, MetaCyc contains more than 2042 pathways from more than 2414 different organisms along with information about the compounds and enzymes in these pathways. The pathway also provides information on the literature source of the data. MetaCyc data can be accessed to search for pathways, enzymes, reactions, and metabolites (Caspi et al., 2012).

Both the databases contain similar data on metabolic pathways and their components but they vary in the manner the data is organized and curated. KEGG is based on a set of “reference pathway maps” typically compiled from multiple literature sources, and integrates reactions and pathways found in multiple species. KEGG also contains a smaller type of pathway called a “module.” The “species view” of KEGG reference pathways shows coloured reactions to indicate which enzymes are predicted to be present in a given organism in a reference pathway (Caspi et al., 2012).

The collection of pathways in MetaCyc is analogous to the KEGG reference pathway set, and the organism-specific PGDBs within BioCyc correspond to KEGG species views of its reference pathway maps. KEGG modules are comparable to MetaCyc pathways, whereas KEGG maps are much larger and are comparable to MetaCyc superpathways (Altman et al., 2013).

4.1.1.3. Tools based on pathway databases :

The pathway databases offer tools or interfaces to interact with the database to search and analyse after entering the query. They also offer the possibility of using the database using customised interfaces or tools. Some such tools used in this study are :

- **PathComp :** (http://www.genome.jp/tools/pathcomp/)

  The PathComp utility on the KEGG computes possible reaction paths between a user-selected pair of a substrate and a product - computes reaction pathways that are known to exist in a given organism using the binary relations of substrates and products in known enzymatic reactions. The list of binary relations is generated from the enzyme list that corresponds to the KEGG reference pathways (all known enzymes) or that is found in an annotated genome (organism specific composition of enzymes) (Goto et al., 1997, Ogata et al., 1998, Fujibuchi et al., 1998).

- **PathPred :** (http://www.genome.jp/tools/pathpred/)

  PathPred is a web-based server to predict plausible enzyme-catalysed reaction pathways from a query compound using the information of RDM patterns and chemical structure alignments of substrate-product pairs. This server provides plausible reactions and transformed compounds, and displays all predicted reaction pathways in tree-shaped graph.

- **Rahnuma :** (http://portal.stats.ox.ac.uk:8080/rahnuma)

  Rahnuma is a tool for prediction and analysis of metabolic pathways and comparison of metabolic networks that represents metabolic networks as hypergraphs. Rahnuma computes all possible pathways between two or more metabolites by using constrained depth first traversal of a hypergraph. It also allows pathway based metabolic network comparisons at organism as well as phylogenetic level. Rahnuma is written in Java and uses a MySQL database to store data from KEGG.. (Mithani et al., 2009).

  Rahnuma enables users to perform pathway analysis and comparisons for single organisms or groups of organisms, and in the context of a user-defined phylogeny. Rahnuma also allows identification of differences between networks by comparing predicted pathways between two or more metabolites. Rahnuma computes pathways between individual metabolites or a group of metabolites using depth-first traversal of the hypergraph representing the metabolic network. A pathway is said to exist between any two metabolites if there is a connected sequence of distinct reactions (or hyperedges) between the two metabolites such that the product of one reaction acts as a substrate in the next reaction. Rahnuma is able to predict pathways between metabolites that are not identified by PathComp.

- **iPath (v2) :** (http://pathways.embl.de/iPath2.cgi)

  interactive Pathways Explorer (iPath) is a web-based tool for the visualization, analysis and customization of the various pathways maps. Current version provides three different global overview maps viz. metabolic pathways, regulatory pathways and biosynthesis of secondary metabolites.

  In addition to the KEGG based overview maps, iPath is used for visualization of various species specific, manually created pathway maps.

  iPath provides extensive map customization and data mapping capabilities. Colours, width and opacity of any map element can be customized using various types of data (for example KEGG KOs, COGs or enzyme EC numbers). All maps in iPath can be easily converted to various bitmap and vector graphical formats for easy inclusion in your documents or further processing. (Letunic et al., 2008, Yamada et al., 2011).

4.1.2. Enzyme database : BRaunschweig ENzyme DAtabase (BRENDA):

BRENDA (BRaunschweig ENzyme DAtabase) is an enzyme information system representing one of the most comprehensive enzyme repositories.

BRENDA is an electronic information resource that comprises molecular and biochemical information on enzymes that have been classified by the IUBMB. Every classified enzyme is characterized with respect to its catalysed biochemical reaction. Kinetic properties of the corresponding reactants; i.e., substrates and products are described in detail. BRENDA provides a web-based user interface that allows a convenient and sophisticated access to the data. BRENDA was founded in 1987 at the former German National Research Centre for Biotechnology (now: Helmholtz Centre for Infection Research) in Braunschweig and was originally published as a series of books. From 1996 to 2007, BRENDA was located at the University of Cologne. There, BRENDA developed into a publicly accessible enzyme information system. In 2007, BRENDA returned to Braunschweig. Currently, BRENDA is maintained and further developed at the Department of Bioinformatics and Biochemistry at the TU Braunschweig.

BRENDA contains enzyme-specific data manually extracted from primary scientific literature and additional data derived from automatic information retrieval methods such as text mining.

The database contains more than 40 data fields with enzyme-specific information on more than 4800 EC numbers that are classified according to the IUBMB. The different data fields cover information on the enzyme's nomenclature, reaction and specificity, enzyme structure, isolation and preparation, enzyme stability, kinetic parameters such as Km value and turnover number, occurrence and localization, mutants and engineered enzymes, application of enzymes and ligand-related data. The data originates from almost 85,000 different scientific articles. Each enzyme entry is clearly linked to at least one literature reference, to its source organism, and, where available, to the protein sequence of the enzyme. Furthermore, cross-references to external information resources such as sequence and 3D-structure databases, as well as biomedical ontologies, are provided. (Scheer et al., 2010).

Since 2006, the data in BRENDA is supplemented with information extracted from the scientific literature by a co-occurrence based text mining approach. For this purpose, two text-mining repositories FRENDA (Full Reference ENzyme DAta) and AMENDA (Automatic Mining of ENzyme DAta) were introduced. These text-mining results were derived from the titles and abstracts of all articles in the literature database PubMed. (Chang et al., 2009).

BRENDA provides links to several other databases with a different focus on the enzyme, e.g., metabolic function or enzyme structure. Other links lead to ontological information on the corresponding gene of the enzyme in question. Links to the literature are established with PubMed. BRENDA links to some further databases and repositories. (Schomburg et al., 2002).

The various bioinformatic tools help in identifying genes that may have a very high probability of playing a role in the metabolism of a particular substrate. However, the only way to be sure what a particular sequence does is to experiment with the organism itself. Once a gene's sequence is known, it can be switched off or silenced. Systematically silencing genes in cells will help reveal genes involved in specific cell processes. Using this approach, some progress has been made in linking the genome of a cell and cellular metabolism, yet, the link between the genotype and the phenotype remains tenuous (Pommerenke et al., 2010).

4.2. Materials and methods :

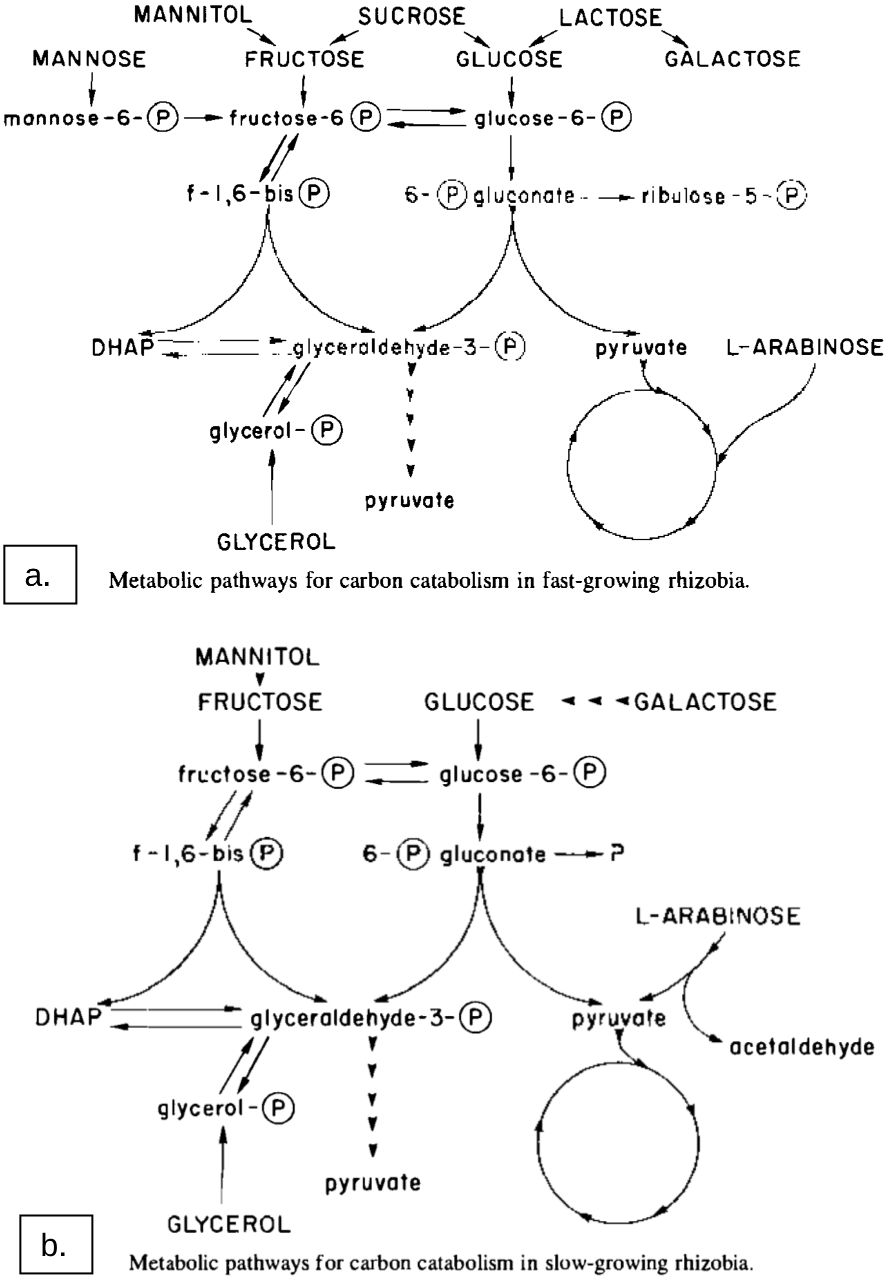
The association between the metabolic differences and differences in the genomic composition of the strains was studied using following approaches :

4.2.1. Bacterial metabolic pathways described in literature :

Some of the metabolic pathways involved in the catabolism of substrates incorporated in the Biolog GN2 plates have been described in literature since long. Many of these pathways have been backed by studies demonstrating that mutations in the genes encoding the enzymes of the pathway result in loss in the ability to utilize the substrate.

For some of the substrates, the catabolic pathways have been known for a long time and excellent information is available in standard textbooks of biochemistry. These were used in obtaining a preliminary idea of the metabolic processes involved in the catabolism of the Biolog substrates. The books used in this study include : Gottschalk (1986), Zubay (1993), Zubay (1998), Metzler and Metzler (2003), Voet and Voet (2011), (White et al., 2011), Gottschalk (2012), Nelson and Cox (2012), Nelson et al. (2013), Voet et al. (2013).

There is some amount of published work specifically on metabolism in *Rhizobium* spp. with reference to certain metabolites. A good amount of useful information is also found in reviews by Stowers (1985) and Poole and Allaway (2000). Relevant references have been referred to at the appropriate places in the chapters. Only the papers that have any relevance to the substrates incorporated into the Biolog GN2 microplate have been referred.



a. Metabolic pathways for carbon catabolism in fast-growing rhizobia.

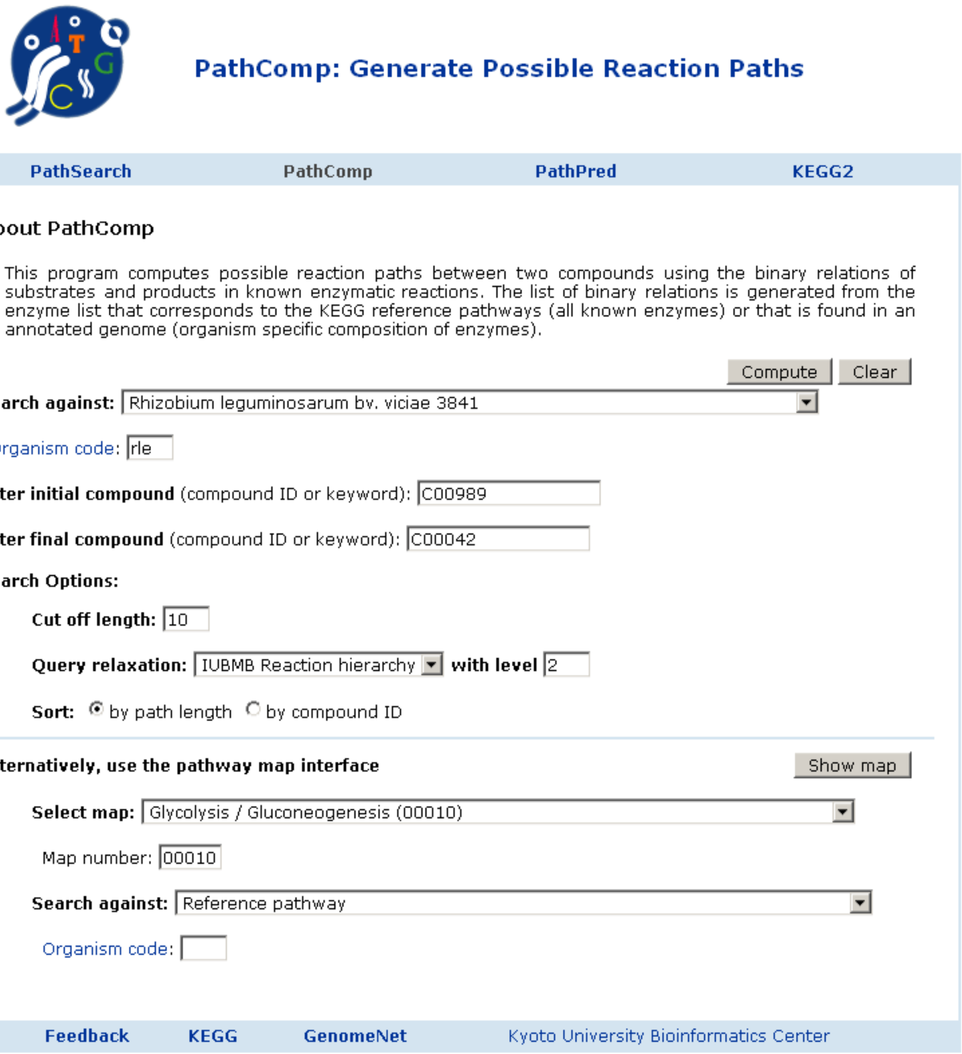
b. Metabolic pathways for carbon catabolism in slow-growing rhizobia.

Figure 4.1. Differences between fast (a.) and slow (b.) growing rhizobia reproduced from Stowers (1985) – an example of metabolism in rhizobia reported in literature.

4.2.2. Use of bioinformatics tools to infer metabolic pathways :

Multiple bioinformatics tools were used to infer metabolic pathways that may be involved in the utilization of substrates incorporated in the Biolog GN2 plates. The protocol for their use is described in this section.

4.2.2.1. **PathComp :** (http://www.genome.jp/tools/pathcomp/)



PathComp: Generate Possible Reaction Paths

About PathComp

This program computes possible reaction paths between two compounds using the binary relations of substrates and products in known enzymatic reactions. The list of binary relations is generated from the enzyme list that corresponds to the KEGG reference pathways (all known enzymes) or that is found in an annotated genome (organism specific composition of enzymes).

Search against: Rhizobium leguminosarum bv. viciae 3841

Organism code: rle

Enter initial compound (compound ID or keyword): C00989

Enter final compound (compound ID or keyword): C00042

Search Options:

Cut off length: 10

Query relaxation: IUBMB Reaction hierarchy **with level** 2

Sort: ● by path length ○ by compound ID

Alternatively, use the pathway map interface

Select map: Glycolysis / Gluconeogenesis (00010)

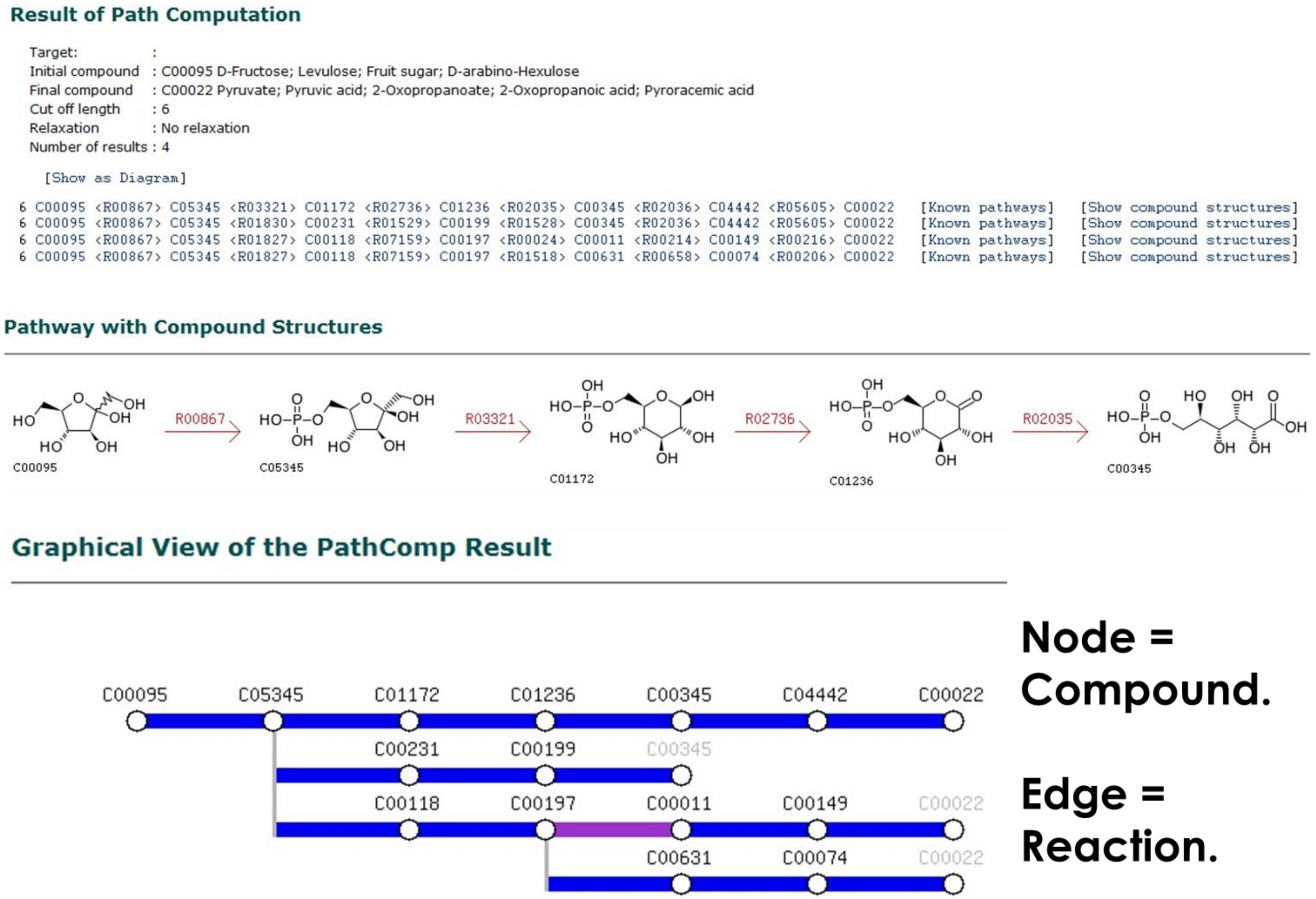
Map number: 00010

Search against: Reference pathway

Organism code:

Figure 4.2. The PathComp page on the KEGG server set up for analysis. The starting compound is γ-hydroxybutyrate (KEGG Compound ID = C00989) and the final compound is succinic acid (KEGG Compound ID = C00042).

PathComp computes possible reaction paths between a user-selected pair of a substrate and a product - computes reaction pathways that are known to exist in a given organism. (Goto et al., 1997, Ogata et al., 1998, Fujibuchi et al., 1998).



Result of Path Computation

```
Target:             :
Initial compound    : C00095 D-Fructose; Levulose; Fruit sugar; D-arabino-Hexulose
Final compound      : C00022 Pyruvate; Pyruvic acid; 2-Oxopropanoate; 2-Oxopropanoic acid; Pyroracemic acid
Cut off length      : 6
Relaxation          : No relaxation
Number of results : 4

  [Show as Diagram]

6 C00095 <R00867> C05345 <R03321> C01172 <R02736> C01236 <R02035> C00345 <R02036> C04442 <R05605> C00022   [Known pathways]   [Show compound structures]
6 C00095 <R00867> C05345 <R01830> C00231 <R01529> C00199 <R01528> C00345 <R02036> C04442 <R05605> C00022   [Known pathways]   [Show compound structures]
6 C00095 <R00867> C05345 <R01827> C00118 <R07159> C00197 <R00024> C00011 <R00214> C00149 <R00216> C00022   [Known pathways]   [Show compound structures]
6 C00095 <R00867> C05345 <R01827> C00118 <R07159> C00197 <R01518> C00631 <R00658> C00074 <R00206> C00022   [Known pathways]   [Show compound structures]
```

Pathway with Compound Structures

Graphical View of the PathComp Result

Figure 4.3. Result of a PathComp analysis performed to check for conversion of D-Fructose (C00095) to Pyruvate (C00022) with a pathway cut-off length of 6. Three views are shown – result of analysis, the pathway with compound structures and a tree showing relation between results.

The PathComp utility requires the identity of the initial and the final compounds that form the metabolic pathway. The utility was observed to work best when the KEGG Compound ID is used as the input method as compared to keywords. Compound ID's were obtained by querying KEGG Compound Database (http://www.kegg.jp/dbget-bin/www_bfind?compound).

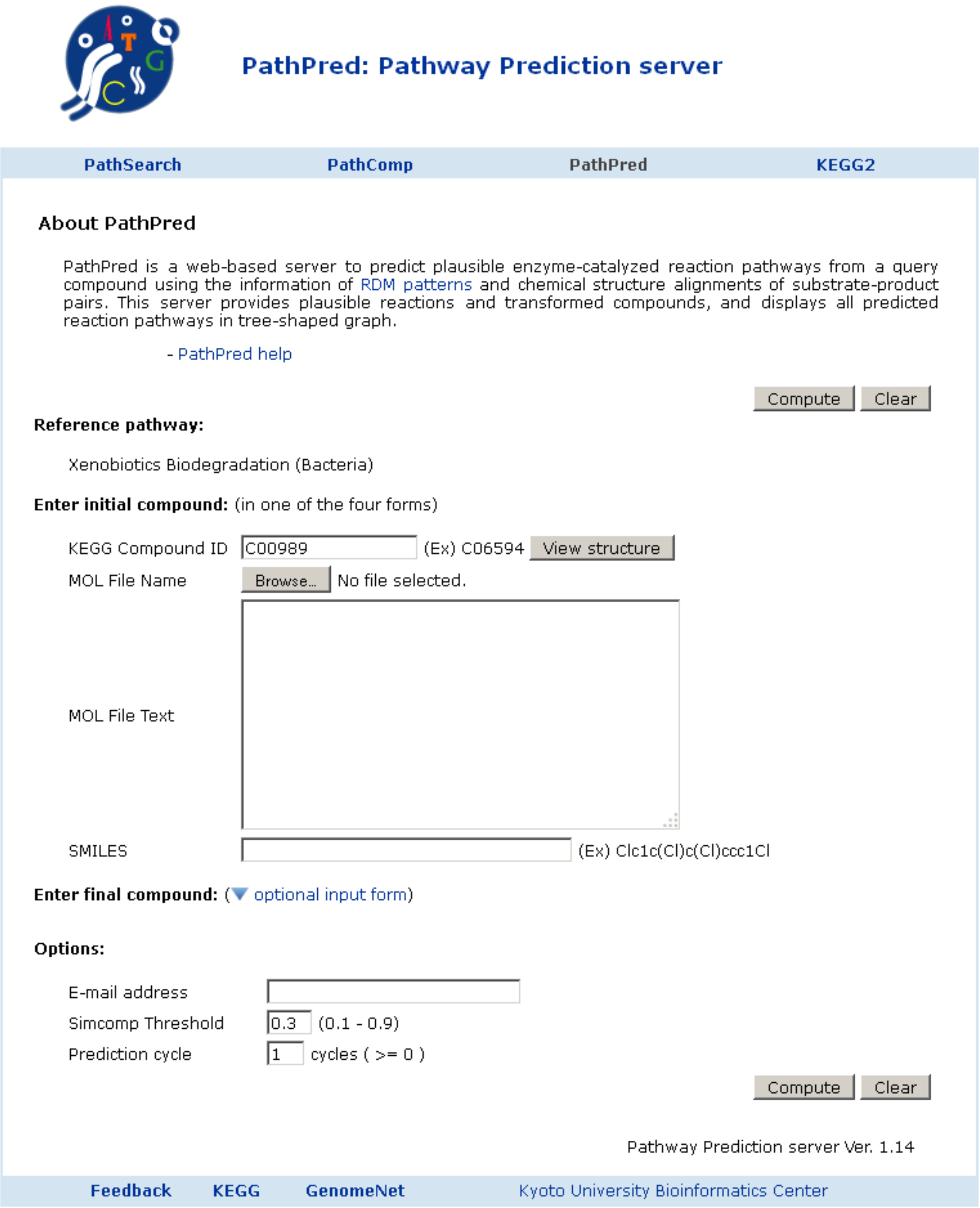
For a sizeable number of Biolog GN2 substrates the metabolic pathways have been investigated the information was used to input the initial and final Compound IDs. For synthetic substrates, the initial Compound ID was used along with the Compound ID of pyruvate as the end-product compound. This was done since it was observed that given the amphibolic nature of the TCA cycle, all the known metabolic pathways terminated in a compound that formed a part of or could be easily transformed into intermediates of the TCA cycle.

The results of the PathComp analysis are shown in Figure 4.3. The graphical view of the analysis shows the manner in which the pathway may branch but yet arrive at the same product i.e. an alternate pathway. The intermediates of the pathway are depicted as nodes with circles with the KEGG COMPOUND number whereas the edges joining the circles constitute the reaction. The edges and the nodes are hyperlinked and can be selected for more in-depth information.

4.2.2.2. **PathPred :** (http://www.genome.jp/tools/pathpred/)

PathPred is another KEGG utility that can predict plausible enzyme-catalysed reaction pathways. It provides plausible reactions and transformed compounds, and displays all predicted reaction pathways in tree-shaped graph (Moriya et al., 2010).

The PathPred program is queried using the KEGG Compound ID, the MDL mol file format or the SMILES representation. For this study, the KEGG Compound IDs were used. The software also provides the option to input the final compound for bi-directional prediction. However, bidirectional prediction gave many irrelevant results and hence only the unidirectional predictions were studied. The results of the PathPred analysis are displayed similar to the PathComp analysis.



About PathPred

PathPred is a web-based server to predict plausible enzyme-catalyzed reaction pathways from a query compound using the information of RDM patterns and chemical structure alignments of substrate-product pairs. This server provides plausible reactions and transformed compounds, and displays all predicted reaction pathways in tree-shaped graph.

- PathPred help

Reference pathway:

Xenobiotics Biodegradation (Bacteria)

Enter initial compound: (in one of the four forms)

KEGG Compound ID: C00989 (Ex) C06594

MOL File Name: No file selected.

MOL File Text:

SMILES: (Ex) Clc1c(Cl)c(Cl)ccc1Cl

Enter final compound: (▼ optional input form)

Options:

E-mail address:

Simcomp Threshold: 0.3 (0.1 - 0.9)

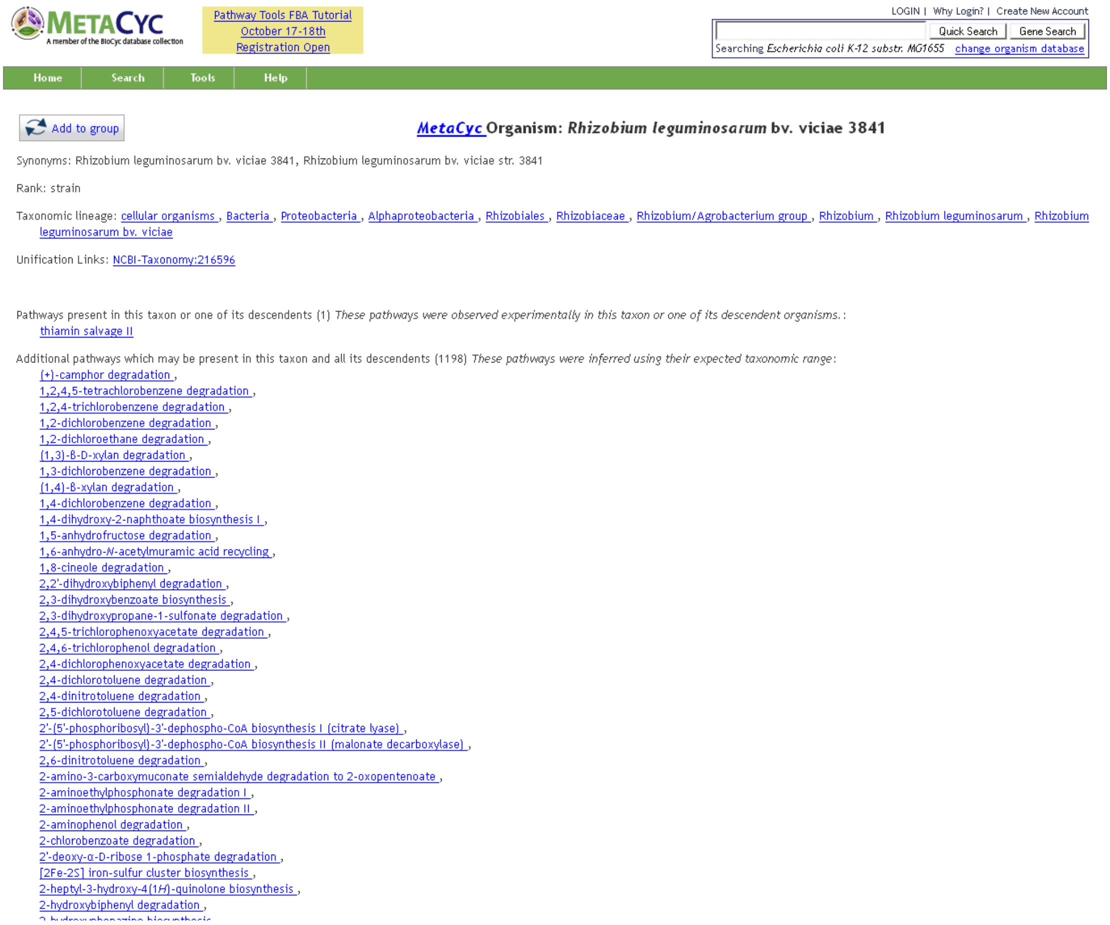
Prediction cycle: 1 cycles (>= 0)

Pathway Prediction server Ver. 1.14

Figure 4.4. The PathPred page on the KEGG server set up for analysis. The starting compound is γ-hydroxybutyrate (KEGG Compound ID = C00989).

4.2.2.3. **MetaCyc :** (http://metacyc.org/)

MetaCyc contains more than 2042 pathways from more than 2414 different organisms, and is curated from the scientific experimental literature (Caspi et al., 2012). The queries to the database can be limited to single species. For this study, the *Rhizobium leguminosarum* bv. *viciae* 3841 database was used (http://ecocyc.org/META/NEW-IMAGE?type=ORGANISM&object =TAX-216596) for all the BIOLOG substrates present in the database..


METACYC
A member of the BioCyc database collection
LOGIN | Why Login? | Create New Account
Quick Search
Gene Search
Home
Search
Tools
Help
Add to group

Pathway Tools FBA Tutorial
October 17-18th
Registration Open

Searching *Escherichia coli K-12 substr. MG1655* change organism database

MetaCyc Organism: *Rhizobium leguminosarum* bv. viciae 3841

Synonyms: Rhizobium leguminosarum bv. viciae 3841, Rhizobium leguminosarum bv. viciae str. 3841

Rank: strain

Taxonomic lineage: cellular organisms , Bacteria , Proteobacteria , Alphaproteobacteria , Rhizobiales , Rhizobiaceae , Rhizobium/Agrobacterium group , Rhizobium , Rhizobium leguminosarum , Rhizobium leguminosarum bv. viciae

Unification Links: NCBI-Taxonomy:216596

Pathways present in this taxon or one of its descendents (1) *These pathways were observed experimentally in this taxon or one of its descendent organisms.*:
thiamin salvage II

Additional pathways which may be present in this taxon and all its descendents (1198) *These pathways were inferred using their expected taxonomic range*:
(+)-camphor degradation ,
1,2,4,5-tetrachlorobenzene degradation ,
1,2,4-trichlorobenzene degradation ,
1,2-dichlorobenzene degradation ,
1,2-dichloroethane degradation ,
(1,3)-ß-D-xylan degradation ,
1,3-dichlorobenzene degradation ,
(1,4)-ß-xylan degradation ,
1,4-dichlorobenzene degradation ,
1,4-dihydroxy-2-naphthoate biosynthesis I ,
1,5-anhydrofructose degradation ,
1,6-anhydro-*N*-acetylmuramic acid recycling ,
1,8-cineole degradation ,
2,2'-dihydroxybiphenyl degradation ,
2,3-dihydroxybenzoate biosynthesis ,
2,3-dihydroxypropane-1-sulfonate degradation ,
2,4,5-trichlorophenoxyacetate degradation ,
2,4,6-trichlorophenol degradation ,
2,4-dichlorophenoxyacetate degradation ,
2,4-dichlorotoluene degradation ,
2,4-dinitrotoluene degradation ,
2,5-dichlorotoluene degradation ,
2'-(5'-phosphoribosyl)-3'-dephospho-CoA biosynthesis I (citrate lyase) ,
2'-(5'-phosphoribosyl)-3'-dephospho-CoA biosynthesis II (malonate decarboxylase) ,
2,6-dinitrotoluene degradation ,
2-amino-3-carboxymuconate semialdehyde degradation to 2-oxopentenoate ,
2-aminoethylphosphonate degradation I ,
2-aminoethylphosphonate degradation II ,
2-aminophenol degradation ,
2-chlorobenzoate degradation ,
2'-deoxy-α-D-ribose 1-phosphate degradation ,
[2Fe-2S] iron-sulfur cluster biosynthesis ,
2-heptyl-3-hydroxy-4(1*H*)-quinolone biosynthesis ,
2-hydroxybiphenyl degradation ,
2-hydroxyphenazine biosynthesis

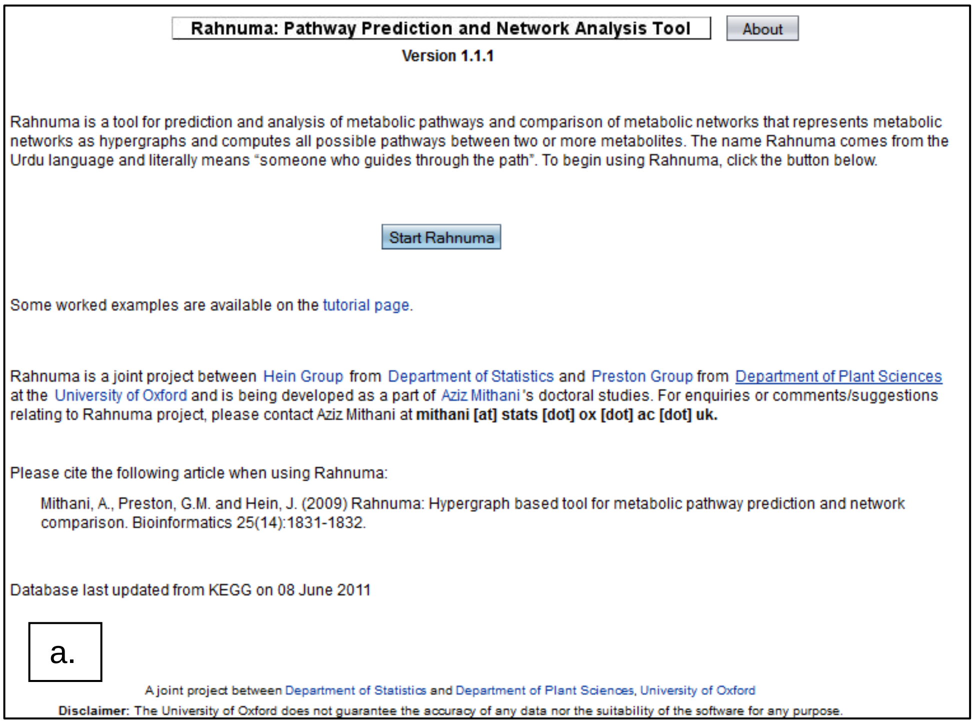
Figure 4.5. The MetaCyc *Rhizobium leguminosarum* bv. *viciae* 3841 database page and the first select substrates whose metabolism can be studied by clicking on the hyperlink contained within the name of substrate.

4.2.2.4. **Rahnuma :** (http://portal.stats.ox.ac.uk:8080/rahnuma)

Rahnuma was used since it claims to predict pathways between metabolites that are not identified by PathComp (Mithani et al., 2009).

The Rahnuma data input page offers various options including emailing links to the results. Three basic types of analysis can be selected from the following options : pathway analysis, comparative analysis and network analysis.

The analysis can be performed on all pathways available in the KEGG database or a specific pathway of interest can be selected. The analysis can be performed on a specific organism or a phylogeny. The results can be exported in different formats.



Rahnuma: Pathway Prediction and Network Analysis Tool

Version 1.1.1

Rahnuma is a tool for prediction and analysis of metabolic pathways and comparison of metabolic networks that represents metabolic networks as hypergraphs and computes all possible pathways between two or more metabolites. The name Rahnuma comes from the Urdu language and literally means "someone who guides through the path". To begin using Rahnuma, click the button below.

Some worked examples are available on the tutorial page.

Rahnuma is a joint project between Hein Group from Department of Statistics and Preston Group from Department of Plant Sciences at the University of Oxford and is being developed as a part of Aziz Mithani's doctoral studies. For enquiries or comments/suggestions relating to Rahnuma project, please contact Aziz Mithani at **mithani [at] stats [dot] ox [dot] ac [dot] uk.**

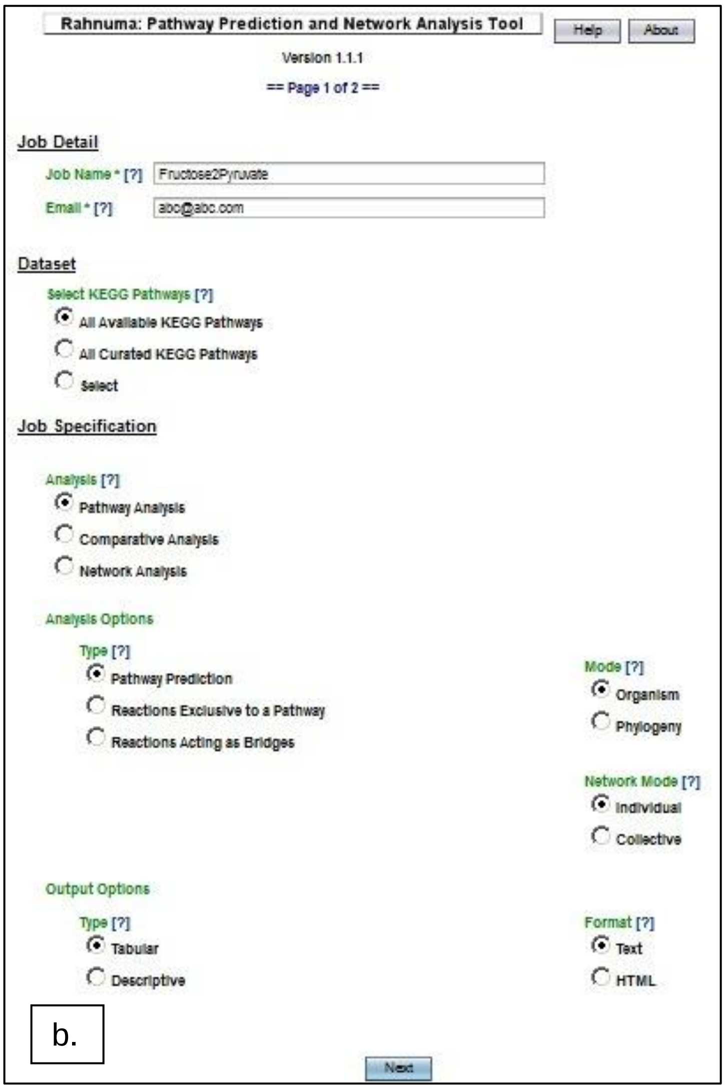
Please cite the following article when using Rahnuma:

> Mithani, A., Preston, G.M. and Hein, J. (2009) Rahnuma: Hypergraph based tool for metabolic pathway prediction and network comparison. Bioinformatics 25(14):1831-1832.

Database last updated from KEGG on 08 June 2011

A joint project between Department of Statistics and Department of Plant Sciences, University of Oxford

Disclaimer: The University of Oxford does not guarantee the accuracy of any data nor the suitability of the software for any purpose.



Rahnuma: Pathway Prediction and Network Analysis Tool

Version 1.1.1

== Page 1 of 2 ==

Job Detail

Job Name * [?] Fructose2Pyruvate

Email * [?] abc@abc.com

Dataset

Select KEGG Pathways [?]

● All Available KEGG Pathways
○ All Curated KEGG Pathways
○ Select

Job Specification

Analysis [?]

● Pathway Analysis
○ Comparative Analysis
○ Network Analysis

Analysis Options

Type [?]

● Pathway Prediction
○ Reactions Exclusive to a Pathway
○ Reactions Acting as Bridges

Mode [?]

● Organism
○ Phylogeny

Network Mode [?]

● Individual
○ Collective

Output Options

Type [?]

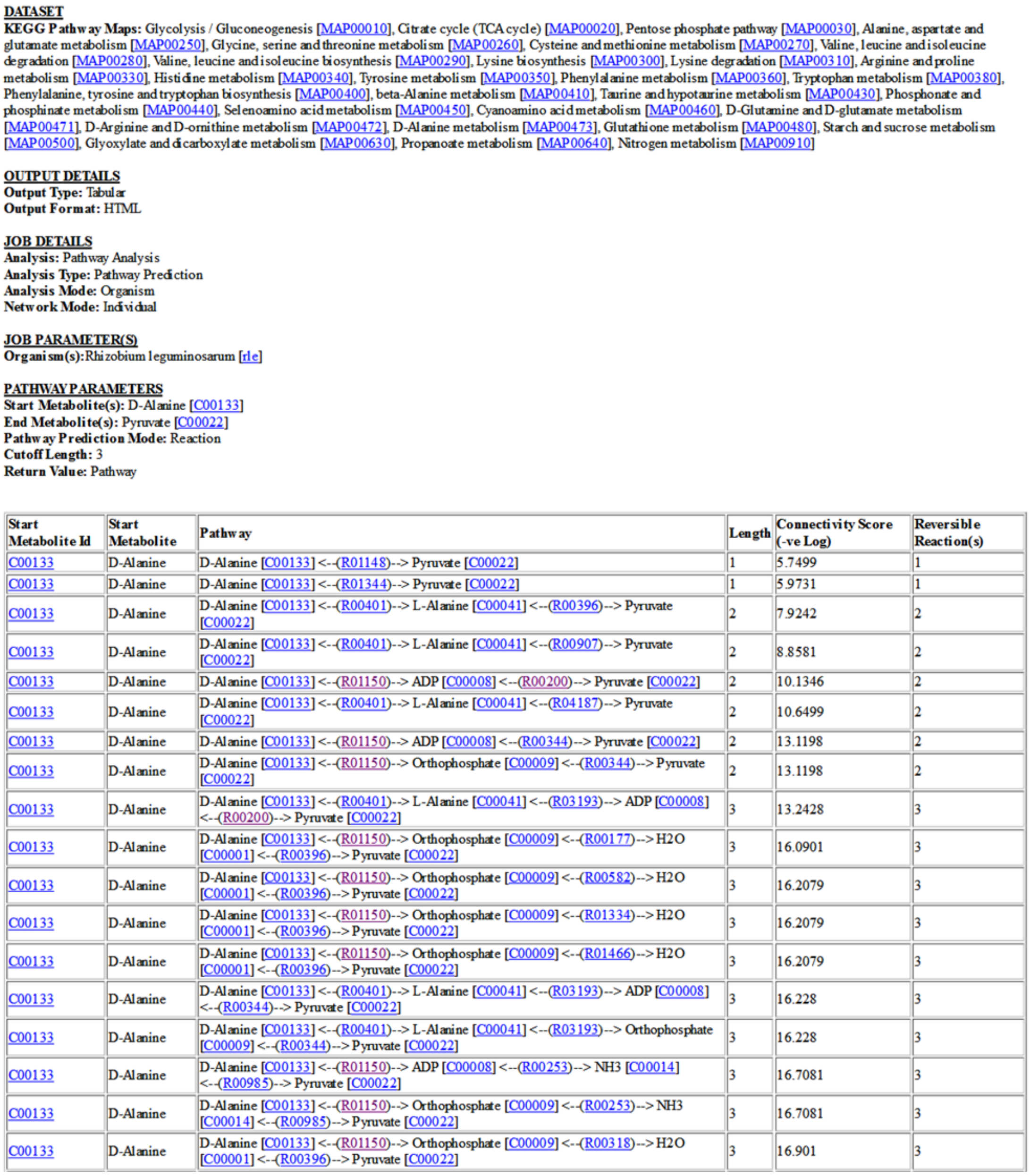
● Tabular
○ Descriptive

Format [?]

● Text
○ HTML

Figure 4.6. (a.) The Rahnuma webpage on the University of Oxford website. Clicking the “Start Rahnuma” button loads the data input page. (b.) The data input page of Rahnuma on the University of Oxford website showing the different options available for analysis. Depending on the selection on the first page (Page 1of1 in the figure), further options are displayed on the second page to select specific test organism, end products and set the pathway cut-off length.

For this study, the analysis of substrates was done on individual substrates and the end products were specified as intermediates of the TCA cycle. The results were exported in a tabular HTML format.



DATASET
KEGG Pathway Maps: Glycolysis / Gluconeogenesis [MAP00010], Citrate cycle (TCA cycle) [MAP00020], Pentose phosphate pathway [MAP00030], Alanine, aspartate and glutamate metabolism [MAP00250], Glycine, serine and threonine metabolism [MAP00260], Cysteine and methionine metabolism [MAP00270], Valine, leucine and isoleucine degradation [MAP00280], Valine, leucine and isoleucine biosynthesis [MAP00290], Lysine biosynthesis [MAP00300], Lysine degradation [MAP00310], Arginine and proline metabolism [MAP00330], Histidine metabolism [MAP00340], Tyrosine metabolism [MAP00350], Phenylalanine metabolism [MAP00360], Tryptophan metabolism [MAP00380], Phenylalanine, tyrosine and tryptophan biosynthesis [MAP00400], beta-Alanine metabolism [MAP00410], Taurine and hypotaurine metabolism [MAP00430], Phosphonate and phosphinate metabolism [MAP00440], Selenoamino acid metabolism [MAP00450], Cyanoamino acid metabolism [MAP00460], D-Glutamine and D-glutamate metabolism [MAP00471], D-Arginine and D-ornithine metabolism [MAP00472], D-Alanine metabolism [MAP00473], Glutathione metabolism [MAP00480], Starch and sucrose metabolism [MAP00500], Glyoxylate and dicarboxylate metabolism [MAP00630], Propanoate metabolism [MAP00640], Nitrogen metabolism [MAP00910]

OUTPUT DETAILS
Output Type: Tabular
Output Format: HTML

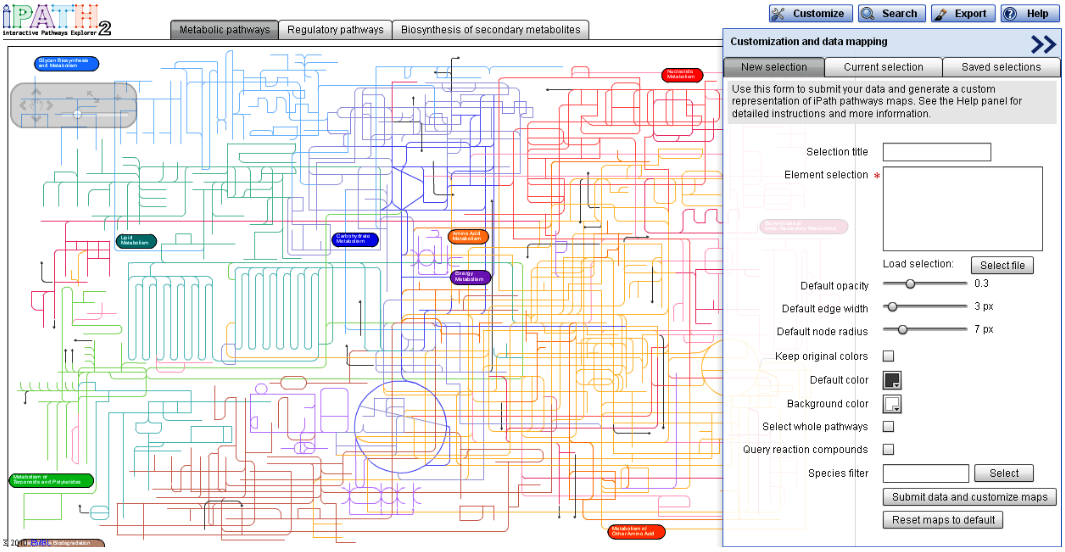
JOB DETAILS
Analysis: Pathway Analysis
Analysis Type: Pathway Prediction
Analysis Mode: Organism
Network Mode: Individual

JOB PARAMETER(S)
Organism(s): Rhizobium leguminosarum [rle]

PATHWAY PARAMETERS
Start Metabolite(s): D-Alanine [C00133]
End Metabolite(s): Pyruvate [C00022]
Pathway Prediction Mode: Reaction
Cutoff Length: 3
Return Value: Pathway

Start Metabolite Id	Start Metabolite	Pathway	Length	Connectivity Score (-ve Log)	Reversible Reaction(s)
C00133	D-Alanine	D-Alanine [C00133] <--(R01148)--> Pyruvate [C00022]	1	5.7499	1
C00133	D-Alanine	D-Alanine [C00133] <--(R01344)--> Pyruvate [C00022]	1	5.9731	1
C00133	D-Alanine	D-Alanine [C00133] <--(R00401)--> L-Alanine [C00041] <--(R00396)--> Pyruvate [C00022]	2	7.9242	2
C00133	D-Alanine	D-Alanine [C00133] <--(R00401)--> L-Alanine [C00041] <--(R00907)--> Pyruvate [C00022]	2	8.8581	2
C00133	D-Alanine	D-Alanine [C00133] <--(R01150)--> ADP [C00008] <--(R00200)--> Pyruvate [C00022]	2	10.1346	2
C00133	D-Alanine	D-Alanine [C00133] <--(R00401)--> L-Alanine [C00041] <--(R04187)--> Pyruvate [C00022]	2	10.6499	2
C00133	D-Alanine	D-Alanine [C00133] <--(R01150)--> ADP [C00008] <--(R00344)--> Pyruvate [C00022]	2	13.1198	2
C00133	D-Alanine	D-Alanine [C00133] <--(R01150)--> Orthophosphate [C00009] <--(R00344)--> Pyruvate [C00022]	2	13.1198	2
C00133	D-Alanine	D-Alanine [C00133] <--(R00401)--> L-Alanine [C00041] <--(R03193)--> ADP [C00008] <--(R00200)--> Pyruvate [C00022]	3	13.2428	3
C00133	D-Alanine	D-Alanine [C00133] <--(R01150)--> Orthophosphate [C00009] <--(R00177)--> H2O [C00001] <--(R00396)--> Pyruvate [C00022]	3	16.0901	3
C00133	D-Alanine	D-Alanine [C00133] <--(R01150)--> Orthophosphate [C00009] <--(R00582)--> H2O [C00001] <--(R00396)--> Pyruvate [C00022]	3	16.2079	3
C00133	D-Alanine	D-Alanine [C00133] <--(R01150)--> Orthophosphate [C00009] <--(R01334)--> H2O [C00001] <--(R00396)--> Pyruvate [C00022]	3	16.2079	3
C00133	D-Alanine	D-Alanine [C00133] <--(R01150)--> Orthophosphate [C00009] <--(R01466)--> H2O [C00001] <--(R00396)--> Pyruvate [C00022]	3	16.2079	3
C00133	D-Alanine	D-Alanine [C00133] <--(R00401)--> L-Alanine [C00041] <--(R03193)--> ADP [C00008] <--(R00344)--> Pyruvate [C00022]	3	16.228	3
C00133	D-Alanine	D-Alanine [C00133] <--(R00401)--> L-Alanine [C00041] <--(R03193)--> Orthophosphate [C00009] <--(R00344)--> Pyruvate [C00022]	3	16.228	3
C00133	D-Alanine	D-Alanine [C00133] <--(R01150)--> ADP [C00008] <--(R00253)--> NH3 [C00014] <--(R00985)--> Pyruvate [C00022]	3	16.7081	3
C00133	D-Alanine	D-Alanine [C00133] <--(R01150)--> Orthophosphate [C00009] <--(R00253)--> NH3 [C00014] <--(R00985)--> Pyruvate [C00022]	3	16.7081	3
C00133	D-Alanine	D-Alanine [C00133] <--(R01150)--> Orthophosphate [C00009] <--(R00318)--> H2O [C00001] <--(R00396)--> Pyruvate [C00022]	3	16.901	3

4.2.2.5. **iPath (v2) :** (http://pathways.embl.de/iPath2.cgi)



Use this form to submit your data and generate a custom representation of iPath pathways maps. See the Help panel for detailed instructions and more information.

Selection title
Element selection *
Load selection: Select file
Default opacity 0.3
Default edge width 3 px
Default node radius 7 px
Keep original colors □
Default color
Background color
Select whole pathways □
Query reaction compounds □
Species filter Select

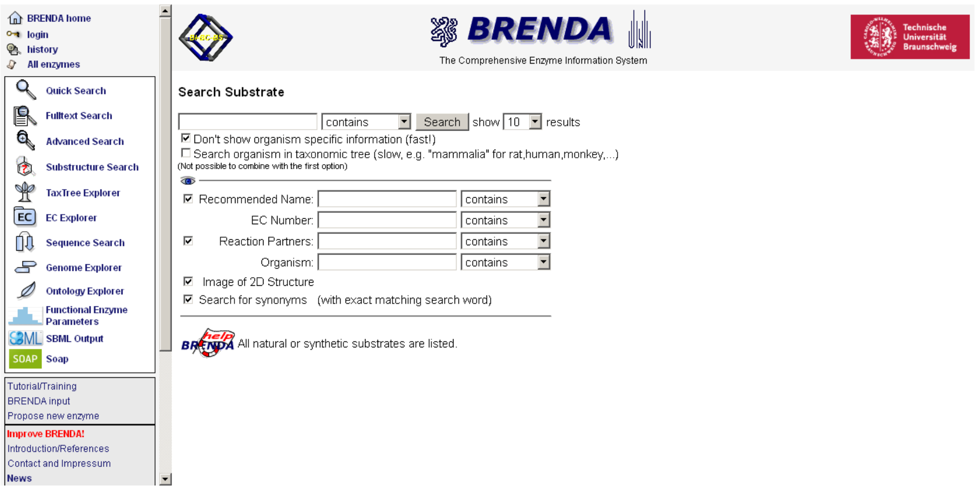
Figure 4.7. The data input page for the iPath v2 web tool. The tab-delimited data is imported using the 'Select File' tab.

interactive Pathways Explorer (iPath) was queried to study metabolic differences between the biovars *viciae* and *trifolii* of *Rhizobium leguminosarum*. (Letunic et al., 2008, Yamada et al., 2011).

The result from Rahnuma showing the metabolic differences between the biovars *trifolii* and *viciae* was used to visualization. The data was input as a tab-delimited file and colours were assigned to the two different species to be shown on the global metabolic map. The maps showing the difference in the metabolism of the two biovars was then exported in a scalable vector graphics (.svg) format for analysis.

4.2.2.6. **BRaunschweig ENzyme DAtabase (BRENDA) :** (http://www.brenda-enzymes.org/)

BRENDA (BRaunschweig ENzyme DAtabase) is one of the most comprehensive enzyme repositories (Scheer et al., 2010). BRENDA was used to investigate the enzymes involved in the reactions shown in KEGG and search for alternate enzymes and reactions that could possibly be catalysed at any given node in the metabolic pathway. This was particularly useful when investigating artificial or synthetic substrates not found in the KEGG database.



Search Substrate

[ ] contains ▾ Search show 10 ▾ results

☑ Don't show organism specific information (fast!)

□ Search organism in taxonomic tree (slow, e.g. "mammalia" for rat,human,monkey,...)
(Not possible to combine with the first option)

☑ Recommended Name: [ ] contains ▾

EC Number: [ ] contains ▾

☑ Reaction Partners: [ ] contains ▾

Organism: [ ] contains ▾

☑ Image of 2D Structure

☑ Search for synonyms (with exact matching search word)

All natural or synthetic substrates are listed.

Figure 4.8. The data input page to search for enzymes involved in the breakdown / utilization of a specific substrate.

4.2.3. Correlation analysis of presence / absence of genes and substrate utilization :

The correlation between the presence / absence of genes and the ability to utilize a particular substrate was investigated using a machine approach.

A logical flowchart was created by me as shown in Figure 4.9. This was converted into a script with the help of two colleagues. Mr. Alex Leach, helped write a script in Python to perform the correlation analysis.

A similar script was written in the programming language R by another colleague Ms. Piyachat Udomwong. The output of the both the scripts listed all the genes all the genes showing strong positive or negative correlation with substrate utilization.

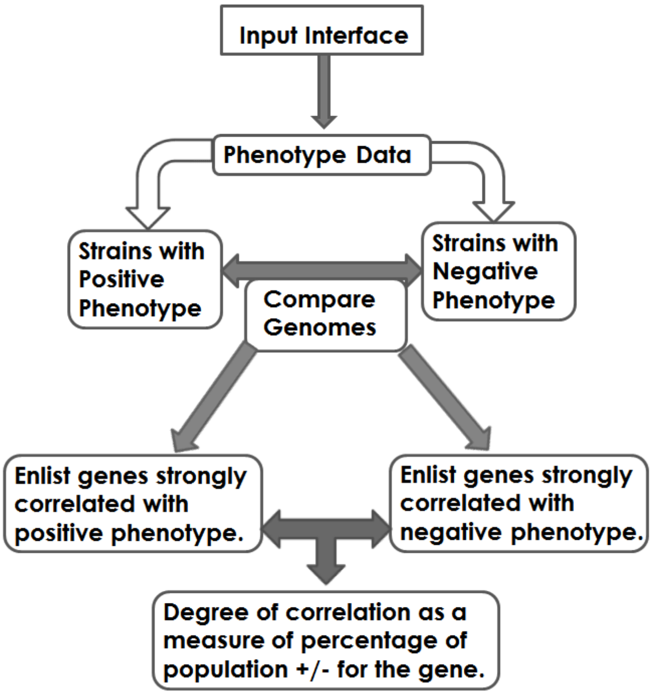

Input Interface
Phenotype Data
Strains with Positive Phenotype
Strains with Negative Phenotype
Compare Genomes
Enlist genes strongly correlated with positive phenotype.
Enlist genes strongly correlated with negative phenotype.
Degree of correlation as a measure of percentage of population +/- for the gene.

Figure 4.9. Logical flow chart to identify genes showing strong positive and negative correlation with substrate utilization.

4.3. Results :

The association between the metabolic differences and differences in the genomic composition of the strains was studied using a variety of approaches. The aim of this analysis was to study the catabolic pathways of metabolism for each of the 95 substrates and arrive at a 'consensus' pathway that could be used for investigating the stringency of correlation between the substrate and the genes encoding the enzymes implicated in the metabolic pathway. The analysis could also potentially identify genes whose absence might be able to explain the inability of strains to use the substrate as a carbon source.

As mentioned earlier, the metabolic pathways for a number of substrates have been known for a long time and have been described in classic biochemistry textbooks. Many of these pathways were discovered by studies demonstrating that mutations in the genes encoding the enzymes of the pathway result in loss of ability to utilize the substrate. These were used as the starting point to investigate the functioning of these pathways in *Rlv*. 3841 using online tools. This resulted in the creation of 95 metabolic pathways, a unique pathway for each of the 95 substrates, verified using multiple approaches. The pathways, along with the name of the enzymes catalysing the reactions of the pathway and the gene accession numbers for the genes encoding those enzymes, were systematically noted down. Of these, the pathways for the 21 substrates that were utilized by *Rlv*. 3841 which showed variation in utilization amongst the Wentworth isolates were investigated further to check for correlation of phenotype to the genotype.

The genotype-phenotype association studies showed a good correlation between a set of genes on the pRL10 plasmid of *Rlv*. 3841 and the ability of a subset of Wentworth isolates to utilize a substrate γ-hydroxybutyrate. Hence, this will be used to demonstrate the steps used in the association studies.

However, since the time the studies have been conducted, Rahnuma has been taken offline and KEGG has now been made a paid site. As a result it is not possible to demonstrate the elucidation of pathway using γ-hydroxybutyrate as the substrate. Screenshot of other substrates have been used to demonstrate the functioning of the web tools.

4.3.1. The consensus pathway for utilization of γ-hydroxybutyrate :

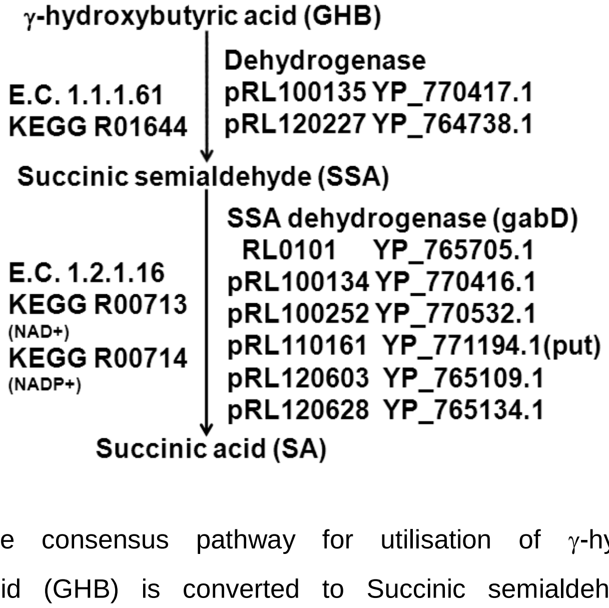
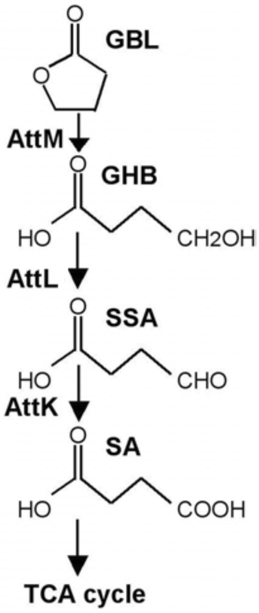


Figure 4.10. The consensus pathway for utilisation of γ-hydroxybutyrate. γ-Hydroxybutyric acid (GHB) is converted to Succinic semialdehyde (SSA) by a dehydrogenase, which in turn in converted to Succinic acid (SA) by another dehydrogenase. The EC enzyme numbers and the KEGG reaction number are on the left of the arrows joining the metabolic intermediates; the gene and protein accession numbers for the enzyme are on the right.

The pathway depicted above shows the consensus pathway for the utilization of γ-hydroxybutyrate by bacteria. The pathway differs from the pathway in mammals where it is converted to a δ-2-hydroxyglutarate by the activity of a transhydrogenase (Struys et al., 2006). In bacteria, it is first converted to succinic semialdehyde by the activity of a dehydrogenase and then metabolised further to succinic acid by the activity of another dehydrogenase. Succinic acid being an intermediate of the TCA cycle is then completely metabolised through the electron transport chain.

4.3.2. Description of the pathway in literature :



The consensus pathway described above has been investigated in the closely related bacterium *Agrobacterium tumefaciens* C58 (Carlier et al., 2004) as shown in Figure 4.11. The study describes the metabolism of γ-butyrolactone (GBL), the precursor of γ-hydroxybutyrate. A lactonase (AttM) encoded by the gene attM converts the ring-structured GHBL to an open-ringed GHB which is converted by the two dehyrogenases encoded by AttL and AttM to successively convert GHB to SSA and SA. The sequence of reaction matches those described in the consensus pathway.

Figure 4.11. Pathway for γ-butyrolactone utilization as described by Carlier et al. (2004).

4.3.3. Verification of the pathway using online pathway tools :

When analysed for its utilization, GBL and GHB were found to be absent in the MetaCyc database. GBL was conspicuous by its absence in KEGG pathways (although it does have a KEGG Compound ID). GHB was found to be present in the map for Butanoate Metabolism (map00650). However, the enzyme converting GHB to SSA (GHB dehydrogenase) was shown to be absent in *Rlv*. 3841. Same was true for *A. tumefaciens* C58. As a result, all three online tools using the KEGG database for pathway elucidation in this study (viz. Rahnuma, PathComp and PathPred) failed to give any results for the utilization of either GBL or GHB.

However, as mentioned above, the utilization of GBL in *A. tumefaciens* C58 has been proved by Carlier et al. (2004) along with the existence of all the necessary genes and enzymes. Hence, the sequence of the genes attK, attL and attM of *A. tumefaciens* C58 was used to perform a BLAST search against the *Rlv*. 3841 genome sequence. The BLAST results returned 6 hits for attK and two each for attL and attM. Of these results, one hit from each of the three sets of results seemed to form a contiguous region like the one found in *A. tumefaciens* C58. This region was found to be present on the plasmid pRL10 and consisted of the genes pRL100134, pRL100135 and pRL100136.

Three different operon prediction tools viz. ProOpDB, MicrobesOnline Operon Predictions and OperonDB all supported the hypothesis that the three genes found on pRL10 indeed constituted an operon with the combination of genes constituting pRL100134:pRL100135 occurring in at least 75 other known bacterial genomes and the combination of genes constituting pRL100135:pRL100136 occurring in at least 5 other known bacterial genomes. The direction of gene transcription was predicted to be in the direction pRL100134→pRL100135→pRL100136 which encode enzymes of the pathway in a bottom-up direction to the reactions of the pathway.

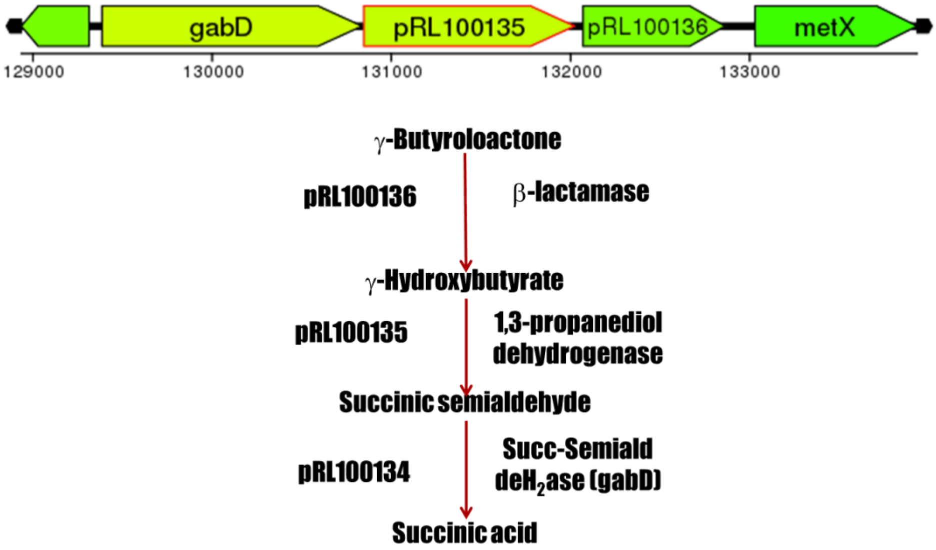


Figure 4.12 : Position of pRL100134 (gabD), pRL100135 and pRL100136 on the plasmid pRL10. Note the direction of transcription relative to the sequence of enzyme action in the pathway.

4.3.4. Correlating the genes with phenotypes :

Since three genes were implicated in the catabolism of GBL, it was now easy to see how well these enzymes correlated with the ability of the bacteria to utilize the GBL and GHB. GHB is a controlled substance and hence its immediate precursor GBL was used for further studies. The conversion of GBL to GHB occurs in a single step reaction catalysed by lactonase. The pattern of utilization of GBL and GHB was exactly similar and hence GBL utilization data was studied in lieu of GHB.

4.3.4.1. Correlation analysis using R :

In order to check the stringency of the correlation between the genes implicated in the utilization of γ-hydroxybutyrate and its actual utilization, a simple flow chart was devised (as described earlier). This was used as a guide by Mr. Alex Leach to write a Python script to check the stringency of correlation. Likewise, a similar script was written in the programming language R to perform similar analysis. For reasons, revealed in the discussions section, the analysis done using R is used in this section.

The correlation analysis between the substrate and the genes was carried out for each replicon. Figure 4.13 shows the correlation plot generated using R for genes present on pRL10 and variation in substrate utilization for 21 substrates in the Wentworth population. The genes showing strong positive correlation are represented by red bars and genes showing strong negative correlation are represented by blue bars. All the genes that showed correlation (positive and negative) with the substrates were studied for their function in different organisms using BLAST search and KEGG pathway tools. Of all the substrates only the genes correlated with utilization of GHB / GBL were found to be involved in a known catabolic pathway.

The genes involved in GHB / GBL utilization are indicated by an arrow and shown magnified on the right. The gene cluster was found to have three genes (pRL100134 - pRL100136) that have been shown to be involved in GHB / GBL utilization in *A. tumefaciens* (Carlier et al., 2004).

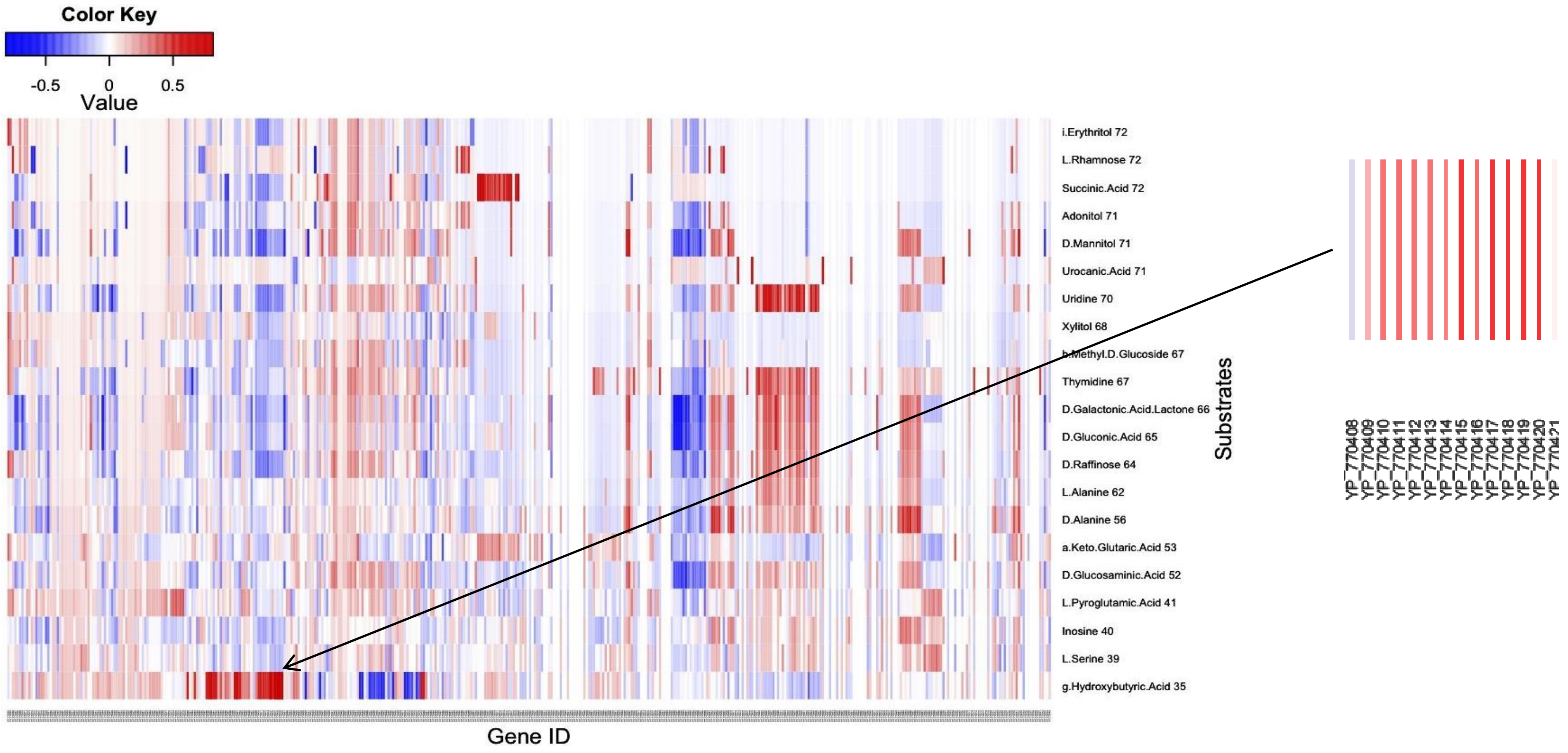


Figure 4.13. Correlation plot for plasmid pRL10 : The figure shows the correlation between genes present on pRL10 and the ability to utilise the 21 substrates investigated. The genes showing strong positive correlation are shown in red and genes showing strong negative correlation are shown in blue. The genes involved in GHB / GBL utilization are indicated by an arrow and shown magnified on the right.

4.3.4.2. Correlation analysis with Excel :

Table 4.1. Correlation between the ability of a strain to utilize GBL and the presence / absence of genes. (See text)

Strain	Util	36	35	34
TRX01	+	1	1	1
TRX02	+	1	1	1
TRX03	+	1	1	1
TRX04	+	1	1	1
TRX05	+	1	1	1
TRX06	+	1	1	1
TRX07	+	1	1	1
TRX08	+	1	1	1
TRX09	-	0	0	0
TRX10	-	1	1	1
TRX11	-	0	0	1
TRX12	+	1	1	1
TRX13	+	1	1	1
TRX14	+	1	1	1
TRX15	-	0	0	0
TRX16	+	1	1	1
TRX17	+	1	1	1
TRX18	-	0	0	0
TRX19	+	1	1	1
TRX20	+	1	1	1
TRX21	+	1	1	1
TRX22	-	0	0	0
TRX23	+	1	1	1
TRX24	+	1	1	1
TRX25	-	0	0	0
TRX26	+	1	1	1
TRX27	-	0	0	0
TRX28	+	1	1	1
TRX29	-	0	0	0
TRX30	+	1	1	1
TRX31	-	0	0	0
TRX32	+	0	0	0
TRX33	+	1	1	1
TRX34	-	0	0	0
TRX35	+	1	1	1
TRX36	+	1	1	1
VSX01	-	0	0	1
VSX02	-	0	0	0
VSX03	+	1	1	1
VSX04	-	0	0	1
VSX05	+	1	1	1
VSX06	-	0	0	1
VSX07	-	0	0	1
VSX08	-	0	0	0
VSX09	-	0	0	1
VSX10	-	0	0	0
VSX11	-	0	0	1
VSX14	-	0	0	0
VSX15	+	1	1	1
VSX16	+	0	0	0
VSX17	-	0	0	0
VSX18	+	0	0	0
VSX19	-	0	0	0
VSX21	-	0	0	0
VSX22	-	0	0	0
VSX23	-	0	0	1
VSX24	-	0	0	0
VSX25	-	0	0	0
VSX26	+	0	0	0
VSX27	+	0	0	0
VSX28	-	0	0	1
VSX29	-	0	0	0
VSX30	-	0	0	0
VSX31	-	0	0	0
VSX32	-	0	0	1
VSX33	+	1	1	1
VSX34	-	0	0	0
VSX35	-	0	0	1
VSX36	-	0	0	1
VSX37	+	0	0	0
VSX38	-	0	0	0
VSX39	-	0	0	0

The adjoining table shows the relationship between the ability of a strain (column = strain) to utilize GBL (column = Util, + = utilization and - = no utilization) and the presence and absence of genes (1 = presence, 0 = absence) suggested to be involved in its metabolism.

Of the 72 Wentworth isolates, 38 strains were unable to utilize GBL as the sole source of carbon, 34 were able to do so. Of the 38 non-utilizing strains, 37 strains did not possess at least two enzymes involved in the utilization of GBL. Only one strain (TRX10) did not show utilization of GBL in any of the replicates although it had all the three enzymes required for its metabolism (shown in blue). GBL is known to pass across the bacterial membranes passively. No transporters have been reported to be involved in the transport of GBL. This would imply that the GBL enters the cells but is not utilized by the bacteria. Since the genes involved in the metabolism of GBL are present, the inability to use GBL might be due to a problem with gene expression as a result of mutation in the structural or functional or genes that form a part of the GBL utilization cluster that could be investigated using microarray analysis.

Six strains (TRX32 and VSX strains 16, 18, 26, 27 and 37) showed utilization of GBL although these strains had neither of three genes required for its utilization (shown in green). The sequence data for these strains has low coverage and hence may not give a positive BLAST result. Alternately, there exists a low possibility of the presence of alternate enzymes or the presence of an entirely different pathway involved in the utilization of GBL.

4.3.4.3. PCR verification of strains with non-correlating results :

Although the sequence data for most strains is of good quality, some strains have low sequence coverage. In order to confirm the absence of genes in the strains that presumably did not have the genes necessary for the utilization of GBL but were still able to use it, the strains, 6 in number viz., TRX32, VSX 16, 18, 26, 27 and 37, were used to perform a colony PCR to confirm the absence of genes. The gene specific PCR primers used in the mutation studies were used for the amplification.

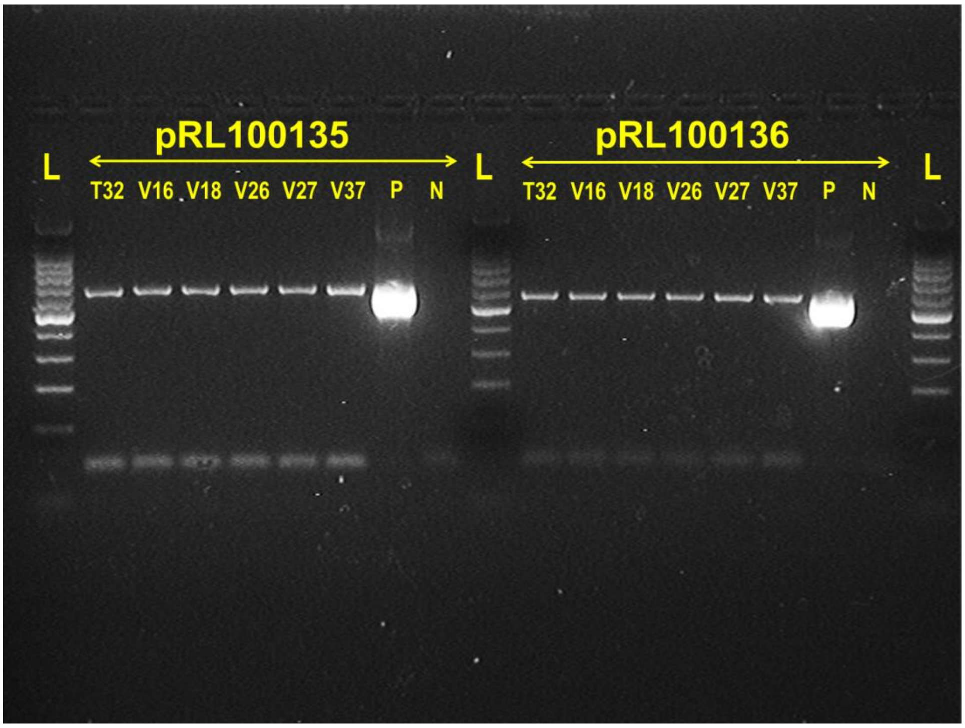


Figure 4.14. Checking for the presence of genes in strains utilizing GBL but not giving a positive BLAST result in sequence data. T and V indicate TRX or VSX strain followed by isolate number, P = positive control *Rlv*. 3841, N = negative control, L= 100 bp ladder from Invitrogen. pRL100135 and pRL100136 indicate gel halves containing PCR amplicons for that gene. Note : For all culture colonies PCR was performed, pure DNA was extracted using FastDNA™ SPIN KIT for positive control.

The result of PCR as seen in the above gels indicates that the central portion of the gene that was amplified by the gene-specific primer was able to amplify pRL100135 and pRL100136 in all the six strains. Although only the central portion of the gene was amplified, there is a very high probability that the entire gene is present in the isolates and was not detected in the BLAST searches; however a very low probability of incomplete gene cannot be discounted since the PCR was not carried out for the entire gene. If the six isolates do have the gene, then of the seven strains that have varying result, six are accounted for leaving the curious case of TRX10.

TRX10 is unique in that it is the only strain that has all the genes required for the utilization of GBL but does not seem to utilize it. In order to confirm the presence of the two genes in the isolate, PCR was carried out using the same gene-specific primers as for the 6 strains above.

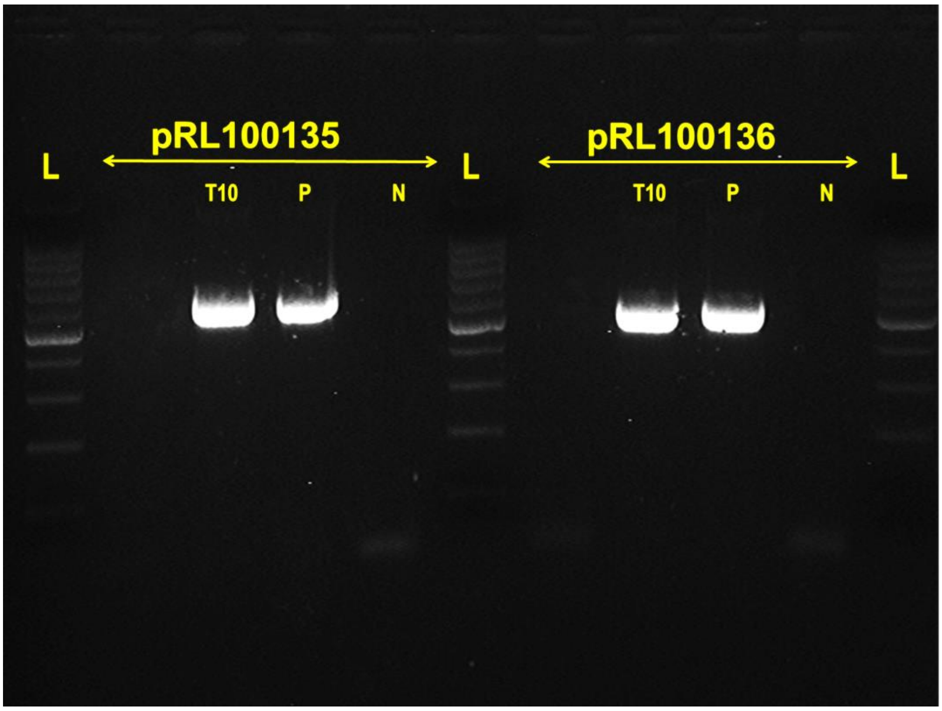


Figure 4.15. Checking for the presence of genes in TRX10. T10 indicates TRX10, P = positive control *Rlv*. 3841, N = negative control, L= 100 bp ladder from Invitrogen. pRL100135 and pRL100136 indicate gel halves containing PCR amplicons for that gene.

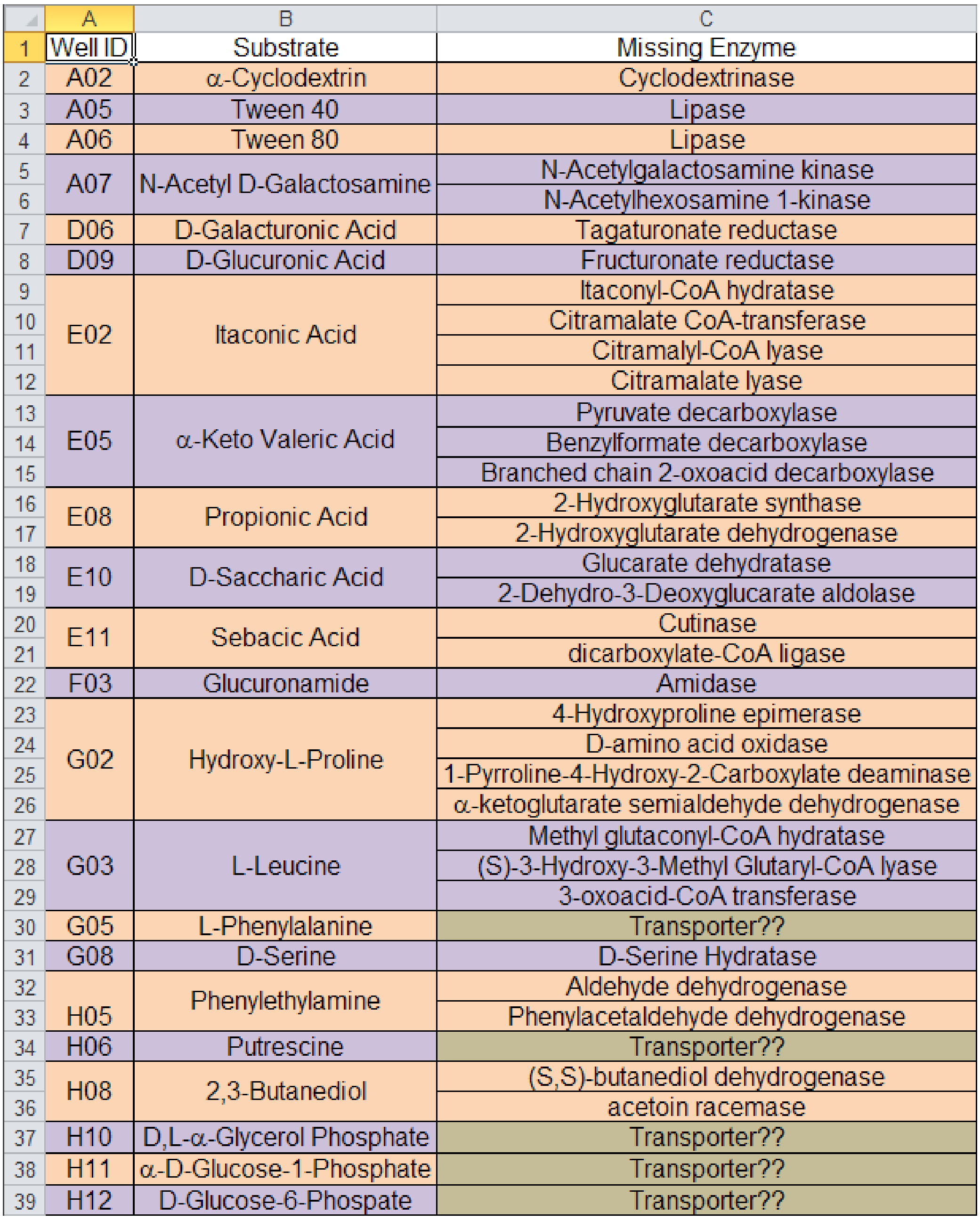
The results show that TRX10 strain does indeed have the two genes but is not capable of utilizing GBL. So, to make sure that the sequence does not have a termination codon internal to it, the sequence data for pRL100135 and pRL100136 from TRX10 was tested using the FindTerm (Solovyev and Salamov, 2010) and Transterm (Jacobs et al., 2009) applications. Both the applications failed to find any termination codon that could probably terminate transcription or translation of the gene. Therefore, there may be a problem with the manner in which these genes are regulated resulting in non-utilization of GBL.

4.3.4.4. Other interesting correlations observed but not pursued for further study :

Investigation of substrates besides GHB also resulted in consensus pathway for each of the substrates. However, most of these substrates were either not utilized by all or most of the strains. If the number of strains using a substrate is few, then using that as a reference to study other strains might result in spurious correlations, Hence, these substrates were not studied further. The list of the substrates and the probable reason why they were not utilized is tabulated in Table 4.2. and some of them discussed below :

- α-cyclodextrin : Cyclodextrin is a closed ring molecule requiring the action of the enzyme cyclodextrinase which opens the ring structure to form maltodextrin which can then be broken down into smaller sugar units. All the strains lacked this enzyme providing a possible explanation to the inability of the strains to use cyclodextrin as a carbon source.
- Tween 40 and Tween 80 : *Rlv*. 3841 and all the Wentworth isolate genomes lacks the lipase involved in the breakdown of the two detergents to their respective fatty acids (palmitic acid for Tween 40 and oleic acid for Tween 80) and polyoxyethylated sorbitol.
- L-Hydroxyproline : Amongst other enzymes, the Wentworth isolates and 3841 lacked the key enzyme of the pathway 4-hydroxyproline epimerase that converts hydroxyproline to proline for further metabolism.
- Many of the substrates included in the Biolog GN2 microarray plate are highly negatively charged. Such substrate molecules need a transporter it across the negatively charged bacterial cell wall. D,L-α-glycerol phosphate, α-D-glucose-1-phosphate and Glucose-6-phosphate were probably not utilized due to the absence of a transport that might be specifically involved in the transport of these substrates across the cell wall.

Table 4.2. List of substrates that showed high correlation to the missing enzymes as gleaned from consensus pathways.



	A	B	C
1	Well ID	Substrate	Missing Enzyme
2	A02	α-Cyclodextrin	Cyclodextrinase
3	A05	Tween 40	Lipase
4	A06	Tween 80	Lipase
5	A07	N-Acetyl D-Galactosamine	N-Acetylgalactosamine kinase
6			N-Acetylhexosamine 1-kinase
7	D06	D-Galacturonic Acid	Tagaturonate reductase
8	D09	D-Glucuronic Acid	Fructuronate reductase
9	E02	Itaconic Acid	Itaconyl-CoA hydratase
10			Citramalate CoA-transferase
11			Citramalyl-CoA lyase
12			Citramalate lyase
13	E05	α-Keto Valeric Acid	Pyruvate decarboxylase
14			Benzylformate decarboxylase
15			Branched chain 2-oxoacid decarboxylase
16	E08	Propionic Acid	2-Hydroxyglutarate synthase
17			2-Hydroxyglutarate dehydrogenase
18	E10	D-Saccharic Acid	Glucarate dehydratase
19			2-Dehydro-3-Deoxyglucarate aldolase
20	E11	Sebacic Acid	Cutinase
21			dicarboxylate-CoA ligase
22	F03	Glucuronamide	Amidase
23	G02	Hydroxy-L-Proline	4-Hydroxyproline epimerase
24			D-amino acid oxidase
25			1-Pyrroline-4-Hydroxy-2-Carboxylate deaminase
26			α-ketoglutarate semialdehyde dehydrogenase
27	G03	L-Leucine	Methyl glutaconyl-CoA hydratase
28			(S)-3-Hydroxy-3-Methyl Glutaryl-CoA lyase
29			3-oxoacid-CoA transferase
30	G05	L-Phenylalanine	Transporter??
31	G08	D-Serine	D-Serine Hydratase
32	H05	Phenylethylamine	Aldehyde dehydrogenase
33			Phenylacetaldehyde dehydrogenase
34	H06	Putrescine	Transporter??
35	H08	2,3-Butanediol	(S,S)-butanediol dehydrogenase
36			acetoin racemase
37	H10	D,L-α-Glycerol Phosphate	Transporter??
38	H11	α-D-Glucose-1-Phosphate	Transporter??
39	H12	D-Glucose-6-Phospate	Transporter??

4.4. Discussion :

The aim of this chapter was to use different tools to obtain correlations between the ability of isolates to utilize a specific substrate and the presence / absence of genes. The metabolic pathways of all the 95 substrates present in the Biolog GN2 plates were studied using a number of metabolic pathway tools and interface software based on these tools to arrive at 'consensus pathways' for each of the substrate.

Some substrates were not used by any of the strains. An attempt has been made to explain the inability of the strains to utilize the substrate in terms of the absence of enzymes based on the information obtained from the consensus pathways. Some of these substrates appear to carry a very high negative charge and hence it might be hypothesized that are probably not utilized due to a lack of a transporter needed to transport the substrate across the negatively charged bacterial cell wall.

On correlating the ability of the strain to use the substrate with the presence or absence of genes, many correlations were observed which are represented as red bars in Figure 4.13. However, when the gene annotations were studied and BLAST analysis was performed to study the function of those genes in other organisms, these correlations were found to be spurious. Only one substrate showed a significant correlation with the gene presence-absence data. This substrate, γ-hydroxybutyrate, was used as a case study to investigate the effect of gene mutation on the ability of the reference strain *R. leguminosarum* 3841, which was used as the test strain. The sequence data of six strains showed that they lacked the enzymes required for the utilization of GBL. PCR amplification with gene specific primers, however, revealed the presence of the genes in these strains.

Only one strain, did not show a correlation between the presence of gene and the ability to utilize GBL viz. TRX10. This strain has all the genes required for utilization of GBL and yet does not do so. No internal stop codons were detected by the tools used indicating that there might be a problem with the manner in which these genes are regulated resulting in non-utilization of GBL.

The role of the genes that showed correlation with GHB in Wentworth population has already been demonstrated in the bacterium *A. tumefaciens* and hence it was thought to be worth investigating. The mutation of these candidate genes and the effect of mutation on the ability of the mutants to use GBL as a carbon substrate will be examined in the next chapter.

CHAPTER 5 : pRL100135 AND pRL100135 ARE INVOLVED IN THE UTILIZATION OF γ-BUTYROLACTONE AND γ-HYDROXYBUTYRATE

5.1. Introduction :

The central dogma of molecular biology advocates linear flow of genetic information from the gene to the protein (function / phenotype) (Crick, 1970) . Hence, disrupting the genetic message would result in the abolition of the function or phenotype. The dogma lies at the heart of the investigation carried out in this part of the thesis.

The results of the analysis carried out in Chapter 4 to identify candidate genes showing strong correlation to substrate utilization showed that three genes located on pRL10 viz. pRL100134 (gabD, succinate-semialdehyde dehydrogenase, NADP$^+$ dependent) (Prell et al., 2009), pRL100135 (1,3-propanediol dehydrogenase) and pRL100136 (beta-lactamase/homoserine lactonase) are strongly correlated with the ability of the strain to utilize GBL as a substrate. In order to test this correlation, the three genes were mutated and the effect of the mutation on the ability to utilize GBL was studied. The three genes and their role in GBL utilization is shown in Figure 5.1.

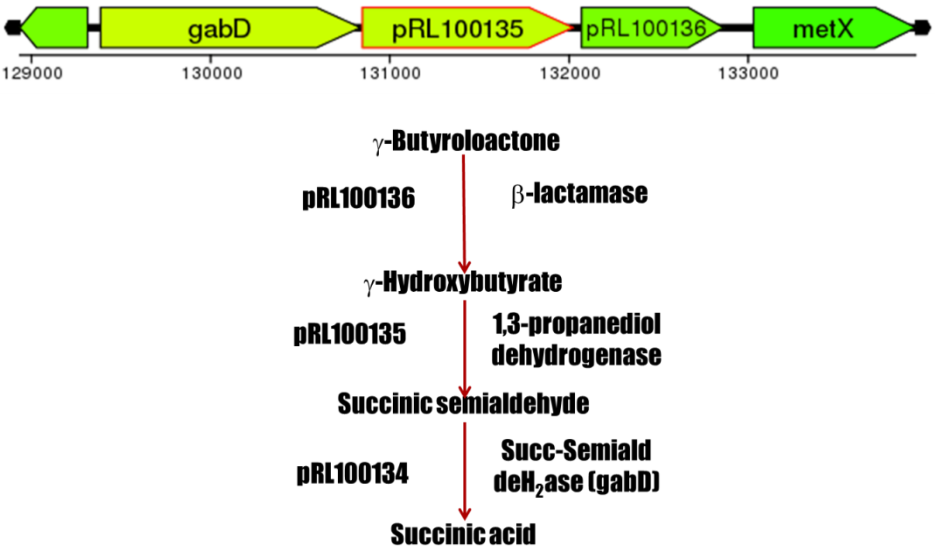


Figure 5.1. Position of pRL100134 (gabD), pRL100135 and pRL100136 on the plasmid pRL10. The pathway from γ-butyrolactone to succinic acid (an intermediate of TCA cycle) is shown with the genes catalysing the steps of the pathway on the left of the reaction and the enzyme encoded by the gene on the right. Note the direction of transcription relative to the sequence of enzyme action in the pathway.

The first enzyme of the pathway is a lactonase that opens the ring structure of GBL and converts it into GHB. The *Rlv*. 3841 genome has two genes that bear close resemblance to the lactonase of *A. tumefaciens* involved in the conversion of GBL to GHB, a copy each located on pRL10 (pRL100136) and pRL11 (pRL110089). The gene pRL110089 is annotated as a (putative) specific 3-ketoadipate enol-lactonase and hence its role in GBL utilization, if any, would be minor thus making pRL100136 unique to the pathway.

The second enzyme of the pathway is GHB-dehydrogenase. A BLAST search for the *A. tumefaciens* analogue of GHB-dehydrogenase in *Rlv* 3841 gives one hit for the gene, located on pRL10 (pRL100135) and annotated as 1,3-propanediol dehydrogenase. A second gene, located on pRL12 (pRL120227), bearing no similarity to the *A. tumefaciens* gene is annotated as GHB-dehydrogenase. However, given that it is not similar in sequence to the *A. tumefaciens* gene, and that the annotation mentions it as putative, it might be an incorrect annotation and was hence not investigated further. Hence pRL100135 (the correct annotation being 4-hydroxybutyrate dehydrogenase or GHB-dehydrogenase), becomes the second unique gene of the pathway.

The third enzyme of the pathway is succinic-semialdehyde dehydrogenase. It is an essential gene converting succinic-semialdehyde derived from various pathways and converting it to succinic acid, to be channelled through the TCA cycle to the electron transport chain. A BLAST search reveals at least 6 copies of the gene present on various replicons. Mutating the gene would affect the metabolism of all substrates channelled through succinic-semialdehyde. Hence this enzyme was not considered for mutation studies.

5.2. Materials and methods :

5.2.1. Mutating the genes by plasmid insertion :

The genes were mutated by plasmid insertion (integration) using the protocol provided by Dr. Jurgen Prell and as described by Karunakaran et al. (2010).

5.2.1.1. Primer design :

Different types of primers were designed, each of which played a different role in the mutation study. All the primers used in this study were designed using the Primer-BLAST tool on the NCBI website : http://www.ncbi.nlm.nih.gov/tools/primer-blast/.

The first set of primers was designed to amplify a 800 bp region within the two genes to be mutated. The primers that amplified the central region of the gene were used for the study. These were labelled as the ‘out’ primers since they annealed to a region of the gene lying outside the mutated region. It was checked that the amplified region contained within the primer sites did not have either a HindIII or BamHI site.

The primers used to obtain the fragment for insertion / integration was designed to amplify a region of 600 bp centred within the 800 bp region amplified by the ‘out’ primers. The primers were designed to be 20 bp in length with a XbaI site on the forward primer and a HindIII site on the reverse primer. These were the internal or ‘in’ primers.

5.2.1.2. pK19mob preparation :

A loopfull of *E.coli* DH5α harbouring plasmid pK19mob was transferred from glycerol stock to LB agar plate containing 25 µg/ml kanamycin (Sambrook et al., 1989). A single colony was transferred to sterile 5 ml LB broth containing 25 µg/ml kanamycin in a 15 ml Corning polypropylene tube and incubated on a rotary shaker at 150 rpm overnight at 37^{O}C. The extraction of pK19mob from the culture was carried out using the QIAprep Spin MiniPrep Kit from Qiagen using the protocol provided by the manufacturer.

5.2.1.3. Preparing DNA for PCR amplification :

A loopfull of *R. leguminosarum* 3841 culture from glycerol stock was transferred onto a sterile TY agar plate and streak plated. The streak-plate was incubated at 28^{O}C for 48 hours. A single colony from the plate was transferred onto a new sterile TY agar plate which was incubated at 28^{O}C for 24 hours. A loopfull of culture from this plate was transferred to 25 ml of sterile TY broth prewarmed to 28^{O}C. The tube was briefly vortexed to disperse the inoculum evenly and then incubated at 28^{O}C on a rotary shaker for 24 hours. The culture was then centrifuged at 4000 rpm for 15 minutes and the resulting bacterial pellet was washed in two changes of physiological saline to remove traces of polysaccharide since it was observed that the polysaccharide decreased the yield of DNA during extraction.

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 gel and by running it on the NanoDrop from Thermo Scientific. The absorbance curve was studied and the 230:260:280 ratio was ratio was determined to assess the purity of the DNA extracted.

5.2.1.4. Amplification of the internal region :

The internal gene fragment to be used for insertion inactivation / gene disruption was amplified using the primers with the restriction sites. The PCR mixture (35 µl) consisted of PCR grade water (21.975 µl), 5X FlexiGreen buffer (7 µl), 25mM dNTP (0.35 µl), 25 mM $MgCl_2$ (2.1 µl) 20 pM primer (0.7 µl each of the forward and reverse 'in' primers with the restriction sites), GoTaq DNA polymerase (0.175 µl). To this mixture was added 2 µl of the kit extracted DNA to bring the final volume to 35 µl.

The PCR program used for amplification started with an initial denaturation at 95^{O}C for 10 minutes. This was followed by 35 amplification cycles; each amplification cycle was made of denaturation time of 95^{O}C for 1 minute, followed by primer annealing at 58^{O}C for 1 minute, primer extension at 72^{O}C for 1 minute. At the end of 35 cycles the a final extension period of 10 minutes at 72^{O}C was added to the program to allow for the extension of incomplete strands.

5µl of the final PCR product was loaded on a 1% agarose gel containing 0.1X SYBR® Safe DNA gel stain (Invitrogen). The gel was run at 100V 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 remaining PCR product (30 µl) was purified using the QIAquick PCR Purification Kit from Qiagen using the manufacturer’s protocol. The elution volume, however, was reduced to 20 µl.

5.2.1.5. Restriction digestion of pK19mob and PCR product :

The pK19mob and the PCR product as a mixture were subjected to a double digestion with the restriction enzymes BamHI and HindIII. 5 µl each of pK19mob plasmid DNA and the PCR product was added to 10 µl of 2X Tango buffer (Fermentas) and the mixture was incubated at 37^{O}C for 60 minutes under static conditions.

At the end of the incubation period, the mixture of plasmid, PCR product and enzymes was cleaned using the QIAquick Gel Extraction Kit (Qiagen). The elution was carried out with 25 µl of elution buffer. 5 µl of the eluate was transferred to a separate tube and labelled as the ‘before ligation sample’ to be run later on an agarose gel. 17 µl of the remaining eluate was used for ligation in the next step.

5.2.1.6. Ligation of the plasmid and PCR product :

The eluate from the double digest contained the plasmid pK19mob cut within the MCS region at the BamHI at HindIII loci. The PCR product too would have been cleaved at the restriction site contained within the primer. Since both the molecules have complementary ‘sticky ends’ they would anneal if kept together in the mixture. However, the formation of a stable hybrid would require the activity of a DNA ligase which would complete the sugar-phosphate backbone by forming a phosphodiester bond.

To carry out ligation, 2 µl of the 10X ligation buffer (New England Biolabs) was added to 17 µl of pK19mob and PCR product. This was followed by addition of 1 µl of T4 DNA ligase (New England Biolabs). The reaction was incubated overnight at 16^{O}C. 5 µl of the reaction mixture was transferred to a separate tube and labelled as the ‘after ligation sample' to be run later on an agarose gel. The remaining reaction mixture was stored at -20^{O}C for later use in transformation of chemically competent *E.coli* DH5α cells.

5.2.1.7. Preparation of chemically competent *E.coli* DH5α cells :

The chemically competent cells for use in transformation protocol were prepared using the following protocol kindly provided by Ms. Madhuri Barge from Dr. Daniella Barilla's Lab, University of York.

A loopfull of *E.coli* DH5α was streak-plated from glycerol stock onto a sterile LB agar plate and incubated at 37^{O}C for 48 hours. A single colony from the plate was streak-plated onto a new sterile LB agar plate and incubated at 37^{O}C for 24 hours. A loopfull of the culture was transferred into 10 ml of sterile LB broth in a 50 ml Corning polypropylene tube and incubated on a rotary shaker at 150 rpm for overnight. 300 µl of the overnight culture was transferred to 60 ml of sterile LB broth in a 250 ml Erlenmeyer flask and incubated on a rotary shaker at 150 rpm till the A_{600} was between 0.4-0.6 (≈4 hours). The flask was then cooled on ice for 10 minutes and all operations from this point were carried out in ice-cold conditions using precooled tubes, centrifugation rotors etc.

The culture was transferred to 2x50 ml Corning polypropylene tubes (30 ml per tube) and centrifuged at 5000 rpm for 5 minutes in a precooled centrifuge. The supernatant was discarded and each of the two pellets was suspended in 10 ml of ice-cold solution RF1 (15% Glycerol, 100 mM RbCl, 50 mM $MnCl_2$, 30 mM K-acetate, 10 mM $CaCl_2$, pH 5.8). The cell suspension was incubated on ice for 15 minutes and then centrifuged at 5000 rpm for 5 minutes in a precooled centrifuge. The supernatant was discarded and each of the two pellets was suspended in 2.4 ml of ice-cold solution RF2 (15% Glycerol, 10 mM MOPS, 10 mM RbCl, 75 mM $CaCl_2$, pH 6.8). The cell suspension was incubated on ice for 15 minutes. The cell suspension was distributed into 50 µl in 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 degree deep freezer for storage until use.

5.2.1.8. Transformation of chemically competent *E.coli* DH5α :

One tube containing 50 µl of chemically competent *E.coli* DH5α cells was thawed on ice. To this 2 µl of the ligation product was added and mixed by pipetting. The mixture was incubated on ice for 30 minutes. The cells were then heat-shocked at 42^{O}C for 90 seconds by immersion in a water bath at 42^{O}C. Upon withdrawal from water-bath, 250 µl of SOC medium prewarmed to 37^{O}C was added to the tube. The tubes were incubated at 37^{O}C for 2 hours under static condition. The entire suspension was then plated on LB agar containing 25 µg/ml Kanamycin and 800 µg X-gal per plate : 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.

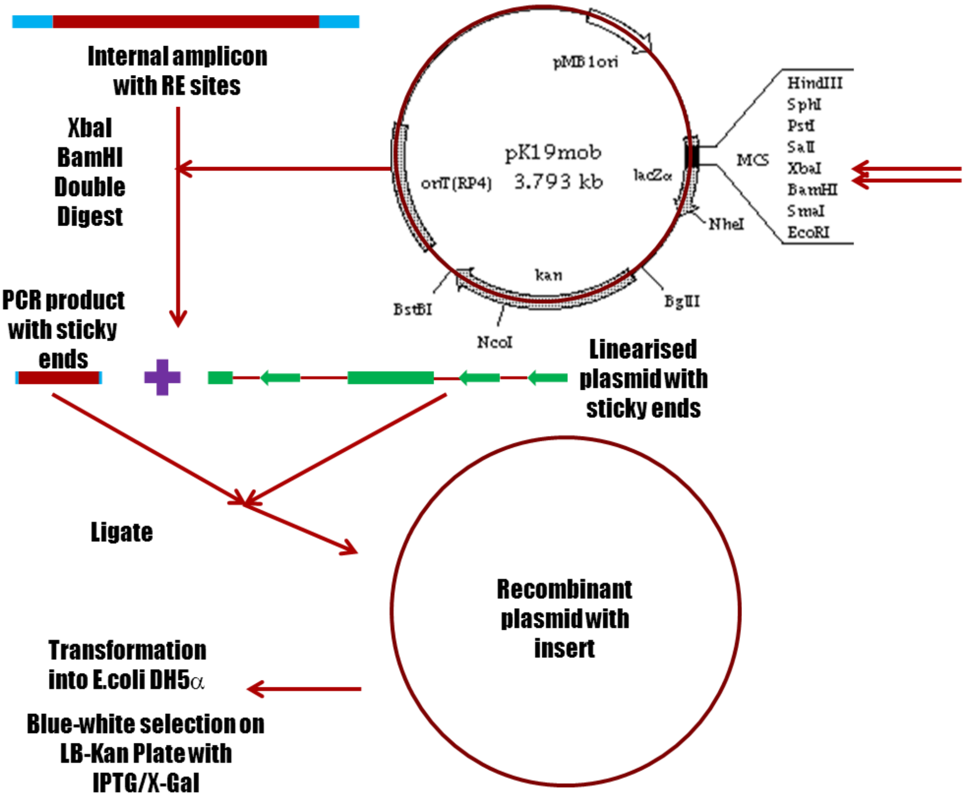

Internal amplicon with RE sites
XbaI BamHI Double Digest
PCR product with sticky ends
pMB1ori
pK19mob
3.793 kb
oriT(RP4)
lacZα
MCS
HindIII
SphI
PstI
SalI
XbaI
BamHI
SmaI
EcoRI
NheI
kan
BstBI
NcoI
BglII
Linearised plasmid with sticky ends
Ligate
Recombinant plasmid with insert
Transformation into E.coli DH5α
Blue-white selection on LB-Kan Plate with IPTG/X-Gal

Figure 5.2. 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.

For screening 12 white colonies were selected for each gene insertion-inactivation setup. The colonies were first transferred onto two LB agar slopes containing 25 µg/ml kanamycin and glycerol stocks were prepared for backup before using the cultures for further analysis. The presence of the insert was confirmed by carrying out PCR using primers similar in sequence to the ones used for amplifying the internal gene region (in 5.2.1.1) but lacking the RE sites.

5.2.1.9. Triparental mating for plasmid transfer to 3841 :

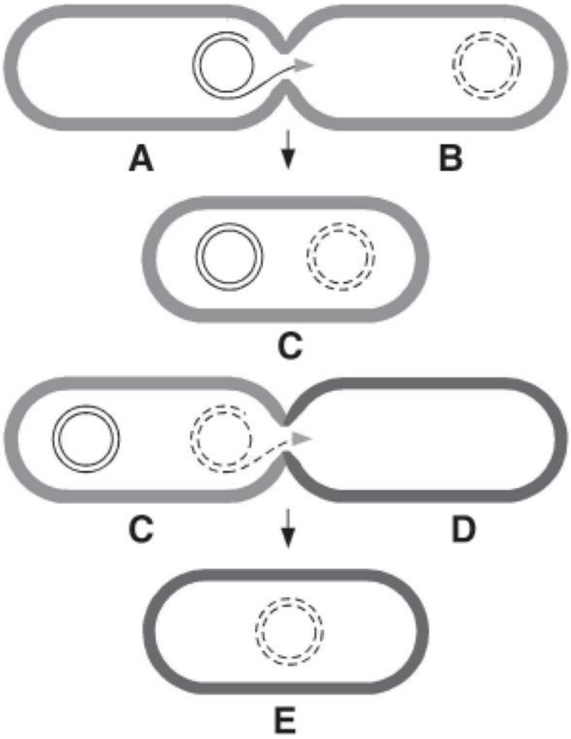


Figure 5.3. Schematic of triparental mating : The *E. coli* helper and donor strains are mixed together with the *Rhizobium* recipient strain. (A) The *E. coli* helper strain transfers the self-transmissible plasmid pRK2013 (solid circles) to *E. coli* donor strain. (B) The *E. coli* donor strain carries 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 and becomes established in the *Rhizobium* cells. From Wise et al. (2006).

In order to carry out gene insertion-inactivation mutation in 3841, it is essential to transfer the recombinant plasmid harbouring a part of the gene into *R. leguminosarum* 3841 (Rlv 3841). To achieve this, the plasmid was transferred from *E.coli* to Rlv 3841 by triparental mating in which the *E.coli* DH5α containing the plasmid construct was the donor, Rlv 3841 was the recipient and *E.coli* DH5α strain harbouring the plasmid pRK2013 was the helper. The triparental mating was carried out as described in the following paragraph.

A loopfull of the donor, recipient and helper cultures were transferred from glycerol stock onto their respective agar media plates containing the respective antibiotics (LB agar with 25 µg/ml of kanamycin for donor and helper, LB with 500 µg/ml streptomycin for the recipient). The donor and helper culture plates were incubated at 37^{O}C whereas the recipient was incubated at 28^{O}C. All cultures were incubated for 48 hours and subcultured on fresh media plates which were further incubated for 24 hours to obtain cultures in vigorous growth phase.

A loopfull of culture was inoculated into 10 ml of their respective liquid growth medium containing the appropriate antibiotics and incubated at appropriate temperatures overnight on shaker at 150 rpm. The recipient Rlv 3841 culture is required in the stationary phase ($A_{600} \approx 1.0$) and required no further treatment. The donor and helper cultures need to be in the mid-logarithmic phase. Hence, the donor and helper cultures were centrifuged and resuspended in 1 ml of sterile saline. 100 µl of this was transferred to fresh broth and further incubated for 3 hours at 37^{O}C ($A_{600} \approx 0.4$) on a shaker at 150 rpm.

After incubation, 700 µl of recipient and 350 µl each of donor and helper was mixed and centrifuged on a bench-top centrifuge for 5 min. The supernatant was discarded and the pellet was vortexed to resuspend the pellet in the residual fluid. The cells were transferred onto a sterile nitrocellulose filter membrane disk placed on a sterile TY agar plate devoid of antibiotics and incubated at 28^{O}C overnight. The following day, the filter was resuspended in 1 ml sterile TY broth and the bacterial suspension in the broth were 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 for recombinants to grow.

12 colonies (for each gene mutation experiment) growing on the streptomycin-neomycin TY agar plates were transferred onto sterile TY agar slopes containing streptomycin-neomycin and glycerol stocks were prepared from the same before using the colonies for further analysis. The 12 clones for each gene were checked for gene inserts by carrying out colony PCR using M13 primers and a chromosomal primer outside the fragment used for the insertion. The clones showing gene insertion within the genomic region were used for growth assays.

5.2.2. Growth curve analysis of wild-type and mutants :

In order to assess the effect of mutation on its ability to grow on γ-butyrolactone, the mutant and the wild type were grown with γ-butyrolactone as the sole carbon source. In order to assess the effect of other variables, the experiment was set up using a combination of different variables as described below in the experimental setup.

5.2.2.1. Growth of test cultures :

A loopfull of *R .leguminosarum* 3841 and the mutant cultures were streak-plated from glycerol stock onto a sterile TY-streptomycin agar plate (for Rlv 3841) and TY-Streptomycin-Neomycin agar plate for the insertion-inactivation mutants and incubated at 28^{O}C for 48 hours. A loopfull of culture was streak-plated onto a new sterile TY agar plate and (with the appropriate antibiotics) and incubated at 28^{O}C for 24 hours. A loopfull of the pure culture was inoculated into 25 ml sterile TY broth (with the appropriate antibiotics) in 50 ml Corning polypropylene tube and incubated on a rotary shaker at 150 rpm for 24 hours at 28^{O}C. The culture was then centrifuged at 4000 rpm for 15 minutes. The supernatant was discarded and the pellet washed with two changes of physiological saline. The bacterial pellet obtained was suspended in Y medium. The density of the culture was adjusted to A_{610} = 1.0 on an ELISA reader (Thermomax, Thermo Scientific) using sterile, uninoculated Y medium as the blank.

10 ml of the above culture was centrifuged. The supernatant was carefully removed to leave behind a pellet+supernatant volume of 1

ml. The tube was vortexed to get a uniform suspension that was 10 times concentrated as compared to the density adjusted culture. This was later used for inoculation of the growth medium.

5.2.2.2. Medium for γ-butyrolactone growth assay :

The Y medium used to study the growth kinetics of the wild type Rlv 3841 and mutants was modified from the original recipe to remove all possible alternate sources of carbon. The composition of the medium used for the assay was as follows : $K_2HPO_4.3H_2O$ – 220 mg, $MgSO_4.7H_2O$ – 100 mg, $CaCl_2.H_2O$ – 220 mg, $FeCl_3.6H_2O$ – 20 mg, Water – to 980 ml. The pH of the medium was adjusted to 6.8.

The medium was sterilized by autoclaving. 6 ml of the sterile medium was dispensed in 20 ml sterile glass Dram bottles.

5.2.2.3. Inoculation of growth assay medium :

Each growth assay was carried out in duplicate. Before inoculating the culture, appropriate antibiotics were added to the Dram tubes. For *R. leguminosarum* 3841, one set of Dram tubes was kept free of antibiotics whereas the other set contained 500 µg/ml of streptomycin. For mutants, the 500 µg/ml of streptomycin and 80 µg/ml of neomycin was added to the medium in order to maintain the selection pressure on the plasmid bearing cultures. 60 µl of filter sterilised γ-butyrolactone was added to each Dram bottle to get a final γ-butyrolactone concentration of 1%.

60 µl of the inoculum (10X A_{610} = 0.1) was inoculated into the 20 ml Dram bottles containing 6 ml of Y medium with antibiotics. The culture was thoroughly mixed with the medium. 200 µl of the inoculated medium was transferred to an empty microtitre dish (Corning Costar 3595) kept on ice for determining initial absorbance of the test strains in the inoculated medium. 2 replicates were made per strain. When all the Dram bottles were inoculated, the initial absorbance of aliquots was read on an ELISA reader (Thermomax, Thermo Scientific) at 610 nm.

5.2.2.4. Incubation :

The mouth of the Dram bottles was covered with Parafilm to maintain sterility while allowing gaseous exchange and transferred to 2 litre glass beakers and stacked vertically in layers to allow for maximum aeration when they were incubated on a shaker at 150 rpm for 72 hours.

5.2.2.5. Harvesting :

During the incubation period, the Dram bottles were removed from the shaker at 12 hour intervals. 200 µl of the medium was transferred to an empty microtitre dish to determine the absorbance of the test strains in the inoculated medium. 2 replicates were made per strain. The absorbance of was read on an ELISA reader (Thermomax, Thermo Scientific) at 610 nm using uninoculated medium as blank and Softmax (V. 2.35, Molecular Devices Corp.).

5.2.2.6. Analysis :

The data from the growth curve were manually keyed into a Microsoft Excel 2003 worksheet. Growth curves were plotted using the same.

5.2.3. Complementation of gene function :

In order to check for reversion of mutant phenotype, gene complementation studies were carried out using a modified version of the protocol used by Vanderlinde et al. (2013).

5.2.3.1. Primer design :

Using the primer design tool on the NCBI website to design suitable primers to amplify the entire genes for cloning. The primers were designed using the following criteria :

- The primer binding site was designed to be as close as possible to the start site of the gene. This would ensure that after cloning, the gene start site was close to the promoter present in the plasmid to be cloned.

- The difference between the T_m of the primers was not more than $1^{O}C$.
- The primers had lowest possible values for hairpin formation, self-dimer formation and hetero-dimer formation.
- The primers were specific for R. leguminosarum biovar viciae 3841.

Another primer was designed within the *trp* promoter which would help in determining the presence and the direction of the insert when used with a normal primer within the gene in the opposite direction on the complementary strand.

5.2.3.2. PCR amplification of the genes :

The genes to be cloned were amplified using the above primers. GoTaq polymerase lacks the 3’→5’ exonuclease activity resulting in 3’ A overhangs. This makes it unsuitable for adapter ligation. Hence, the PCR was carried out using the high-fidelity Pfu DNA polymerase from New England Biolabs using manufacturer’s protocol. This enzyme has 3' to 5' exonuclease proofreading activity, resulting in blunt-ended PCR product suitable for use in adapter ligation. The PCR product was checked by gel electrophoresis and then purified using QIAquick PCR Purification Kit from Qiagen using the manufacturer’s protocol.

5.2.3.3. Making the BamHI adapter complex : Modified from Arneson et al. (2008).

The plasmid pDG71 was selected as the vector to carry out the complementation work (Vanderlinde et al., 2009, Vanderlinde et al., 2013). 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. However, one of the genes to be cloned possessed an internal BamHI site. Therefore it was not possible to use the enzyme for cloning by engineering primers with BamHI site. Hence, it was decided to use link adapters to the PCR products and use the adapter-linked PCR product for cloning into the BamHI site of the vector.

The BamHI adapter was designed to be inverted-repeat palindrome (5'TACCGGGATCCCGGTA3'). The single stranded sequence was to be dimerized by heating the single stranded 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.

5.2.3.4. Preparing the BamHI adapter :

The adapter complex was incubated in presence of BamHI and CIP (calf-intestinal phosphatase) to form dephosphorylated BamHI adapters using enzymes from New England Biolabs. Since the buffers for the enzymes differ in composition, a 'universal' buffer viz. NEB Buffer 3 was used. The dephosphorylation was carried out to ensure that the 'sticky ends' of the adapters do not self-ligate. The adapters were purified using phenol-chloroform-isoamyl alcohol treatment and subsequent ethanol precipitation to remove the enzymes and other non-essential breakdown products like free phosphate.

5.2.3.5. Ligating the adapter to the 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.

5.2.3.6. Linearization of plasmid pDG71 with BamHI :

The plasmid pDG71 was linearized using BamHI. The linearized 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 gene having a BamHI site internal to it from being cloned into pDG71.

5.2.3.7. Cloning the PCR-adapter construct into the linearized pDG71 :

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

5.2.3.8. Inserting the recombinant plasmid into mutant *Rlv.* 3841 :

Chemically competent cells for each *Rlv* 3841 gene mutant were prepared using the same protocol described for preparing chemically competent cells for *E.coli* DH5α.

5 µl of the ligation mixture was used to chemically transform the *Rlv* 3841 cells by heat-shock using the protocol described in 5.2.1.8. The transformants will be selected on Streptomycin-tetracycline plates. The direction of the insert was checked by PCR before using the strains for any further work.

5.2.3.9. Growth studies on the gene complemented strains :

In order to assess the effect of complementation on the ability to grow on γ-butyrolactone, the mutant and the wild type and the strains bearing the plasmid with the full complement of the mutated gene were grown with γ-butyrolactone as the sole carbon source. In order to assess the effect of other variables, the experiment was set up using a combination of different variables as gene mutants (Section 5.2.2).

5.3. Results :

5.3.1. Creating mutants in pRL100135 and pRL10036 :

Using the protocol provided by Dr. Jurgen Prell, insertion mutants were isolated in the genes pRL100135 and pRL100136. The results are indicated below in a stepwise manner.

5.3.1.1. Amplification of central portion of the genes :

Using the 'internal' primers designed for pRL100135 and pRL100136, the central regions of the gene (600 bp) were amplified. The primers were designed to be flanked with restriction sites for HindIII and BamHI. The PCR products were cleaned using the QIAquick PCR Purification Kit from Qiagen using the manufacturer's protocol. The elution volume, however, was reduced to 20 µl. An aliquot of the PCR product was run on agarose gel. The schematic for amplified region and the gel image are shown in Figure 5.4.

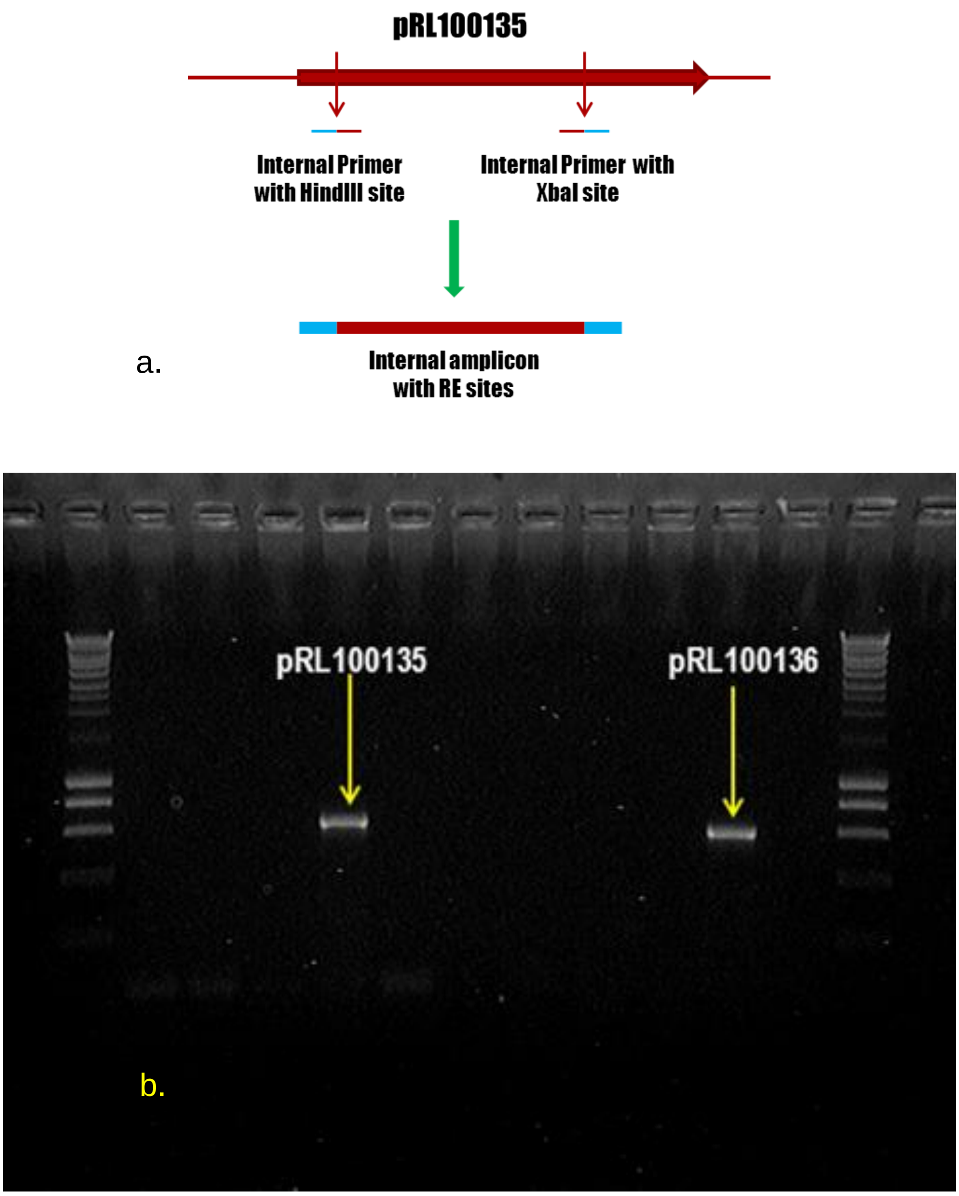

pRL100135
Internal Primer with HindIII site
Internal Primer with XbaI site
a.
Internal amplicon with RE sites
pRL100135
pRL100136
b.

Figure 5.4. Schematic (a) and gel picture (b) of amplification of the core regions of pRL100135 and pRL100136. The 2nd and 14th wells contain the DNA molecular weight marker SmartLadder MW-1700-10 from Eurogentec. The amplification products from the two genes are indicated by labelled arrows.

5.3.1.2. Restriction digestion of pK19mob and PCR product and subsequent ligation :

The cleaned PCR product was added to the plasmid pK19mob and subjected to double restriction digestion with HindIII and BamHI as mentioned in the materials and methods section. This was then purified using QIAquick Gel Extraction Kit (Qiagen). An aliquot was saved for gel electrophoresis as the ‘before ligation sample’. 18µl of the remaining sample was ligated using T4 DNA ligase as mentioned in the materials and methods section. An aliquot was saved for gel electrophoresis as the 'after ligation sample'. The remaining ligation mixture was stored at -20^{O}C for later use in transformation protocol.

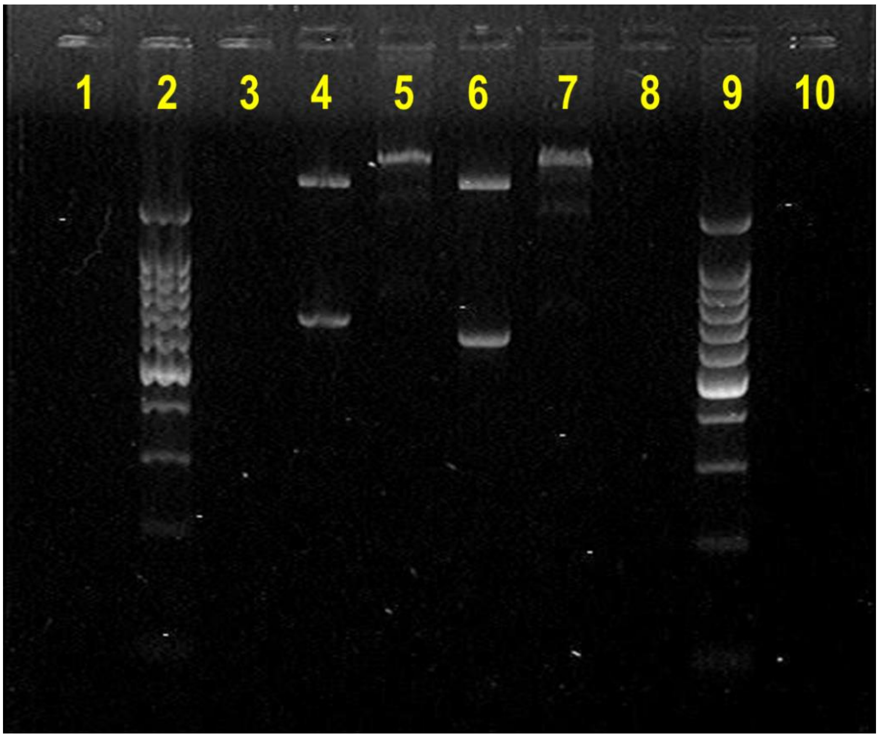


Figure 5.5. Gel electrophoresis of the samples before and after ligation. The 4th well contains the linearized plasmid pK19mob and the pRL100135 before ligation. The 5th well contains the products of the 4th well after ligation. The 6th and 7th wells correspond to before and after ligation products for pRL100136. 2nd and 9th wells contain the molecular weight marker 100 bp ladder from Invitrogen.

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

After transforming the chemically competent *E.coli* DH5α cells with the ligation product, 12 white colonies (from blue-white selection) from the LB-Kanamycin-X-gal were screened for the presence of insert. The presence of the insert was confirmed by carrying out PCR using primers similar in sequence to the ones used for amplifying the internal gene region but lacking the RE sites.

The PCR products were electrophoresed on a 2% agarose gel containing 1X SybrSafe and visualised over blue light.

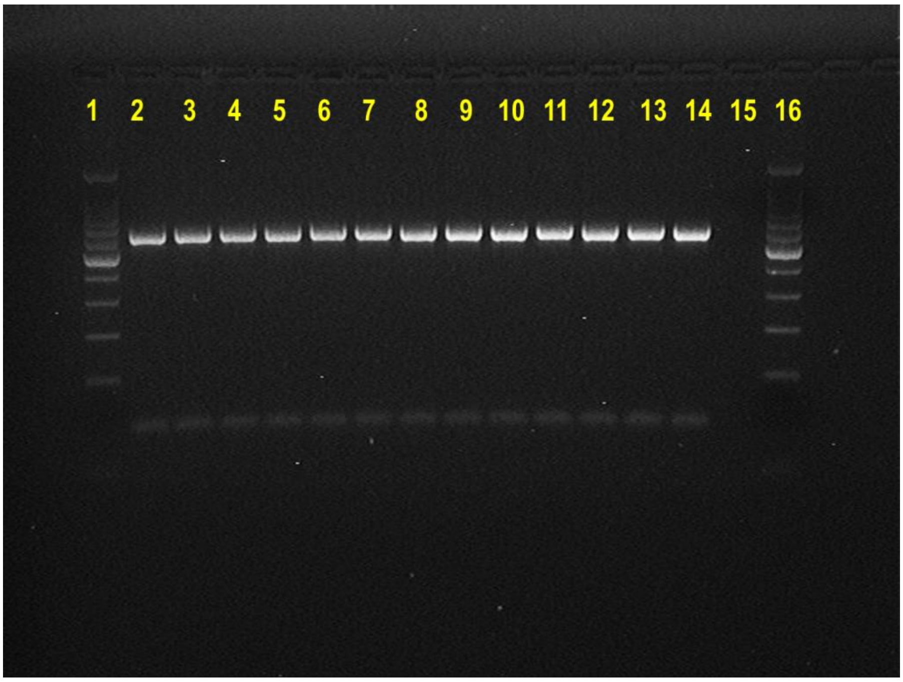


Figure 5.6. Gel electrophoresis of PCR product to check for presence of pK19 mob with pRL100135 in *E.coli* DH5α clones using pRL100135 internal primers. Lanes 1 and 16 – 100 bp ladder from Invitrogen, lanes 2-13 – *E.coli* DH5α clones with pK19mob with pRL100135 insert, lane 14 positive control (*Rlv*. 3841), lane 15 – negative control.

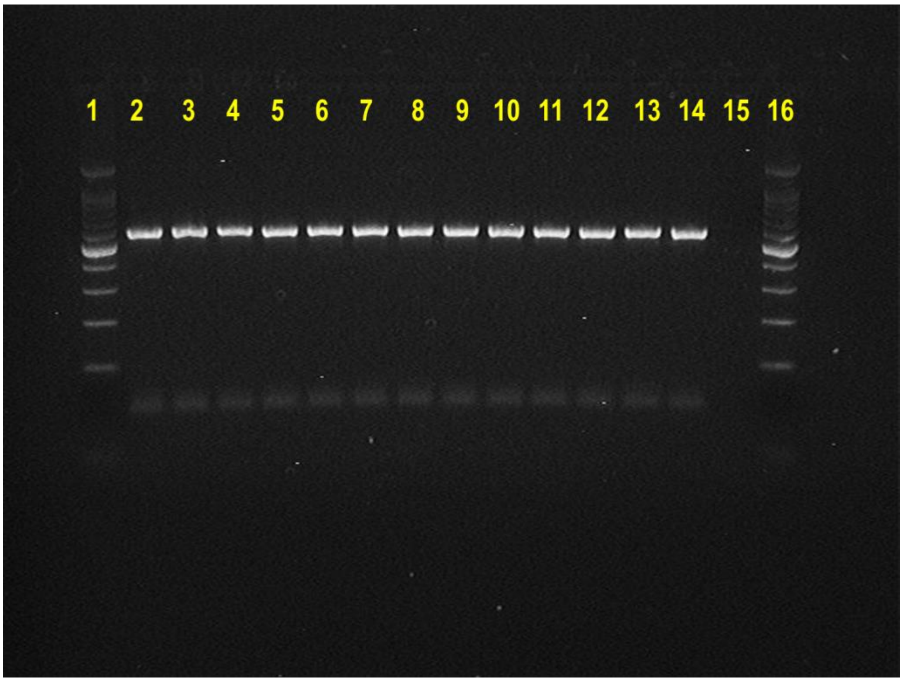


Figure 5.7. Gel electrophoresis of PCR product to check for presence of pK19 mob with pRL100136 in *E.coli* DH5α clones using pRL100136 internal primers. Lanes 1 and 16 – 100 bp ladder from Invitrogen, lanes 2-13 – *E.coli* DH5α clones with pK19mob with pRL100136 insert, lane 14 positive control (*Rlv*. 3841), lane 15 – negative control.

The above gel images show that all the 12 clones screened for each of the two gene construct carried the plasmid along with the inserted gene. One clone from each of the constructs was used to carry out triparental mating with wild type *Rlv*. 3841 in the presence of the *E.coli* helper strain bearing the plasmid pRK2013.

5.3.1.4. Screening of *Rlv*. 3841 transconjugants using PCR :

The plasmid pK19 mob undergoes homologous recombination with a homologous region on the 3841 region and gets inserted into it, inactivating the gene at that site (Schafer et al., 1994). Since the homologous region within the pK19mob is the cloned gene fragment of pRL100135 or pRL100136, the insertion of the plasmid will take place within the gene resulting in inactivation of the functional gene in the wild-type strain.

Twelve *Rlv*. 3841 recombinant (transconjugant) clones from each of the triparental mating experiment, selected on streptomycin-neomycin TY agar plates were screened for plasmid insertion in pRL10 and the direction of the insertion by using combination of PCR primers. The rationale for the PCR is shown in the schematic below using pRL100135 as an example :

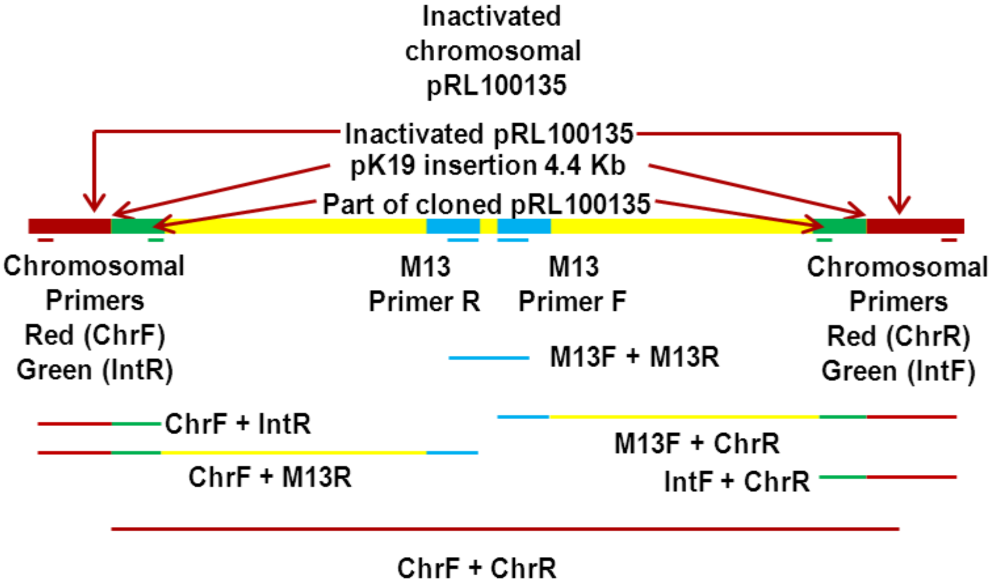


Figure 5.8. Using PCR primers to detect presence and infer direction of insert into the *Rlv*. 3841 mutated gene.

The integration of the partial pRL100135 (green) from the plasmid pK19mob into the pRL10 (referred to as chromosomal for ease) copy of pRL100135 (red) results in the formation of two half-sites that flank the ends of the plasmid region (two green halves at either end) inserted into the pRL10 gene, which in turn gets disrupted and split into two parts (two red ends). Due to the mechanism underlying homologous recombination, the two M13 primer sites (lying halfway across on the construct) move to the central portion of the integrated region resulting in an arrangement of primer sites that is shown in the figure above. The M13 primer sites are represented by blue dashes, the internal primer sites on the insert by green dashes and the chromosomal (pRL10) primer sites by red dashes. The corresponding PCR products are shown by coloured lines.

The colonies obtained on the streptomycin-neomycin TY agar plates were screened for plasmid insertion and the direction of the insertion by using the M13 primer and a primer within the disrupted native gene on pRL10 (chromosomal). The results for the two pairs of primers M13F+chromosomal R (ChrR) and chromosomal F (ChrF)+M13R were the same hence the results for only the first set of primers are shown in the following section. Only the inserts in the right orientation will amplify with these primers.

As a secondary check, a second PCR was performed using the ChrF+ChrR primers. Since the cultures were grown on streptomycin-neomycin TY agar plates, the clones were expected to have the plasmid. However, if the plasmid integration occurred at a site other than the target gene due having high sequence homology, then the M13-chromosomal PCR would not yield any product. This will leave the target gene intact and can be checked by performing a ChrR+ChrF primer pair PCR which would be able to amplify only the native gene. This is possible since the distance between the chromosomal primer sites in the intact gene is far less (800 bp) than the distance of the chromosomal primer sites in a disrupted gene that spans a length of more than 4.4 kb and cannot be amplified in a normal PCR run.

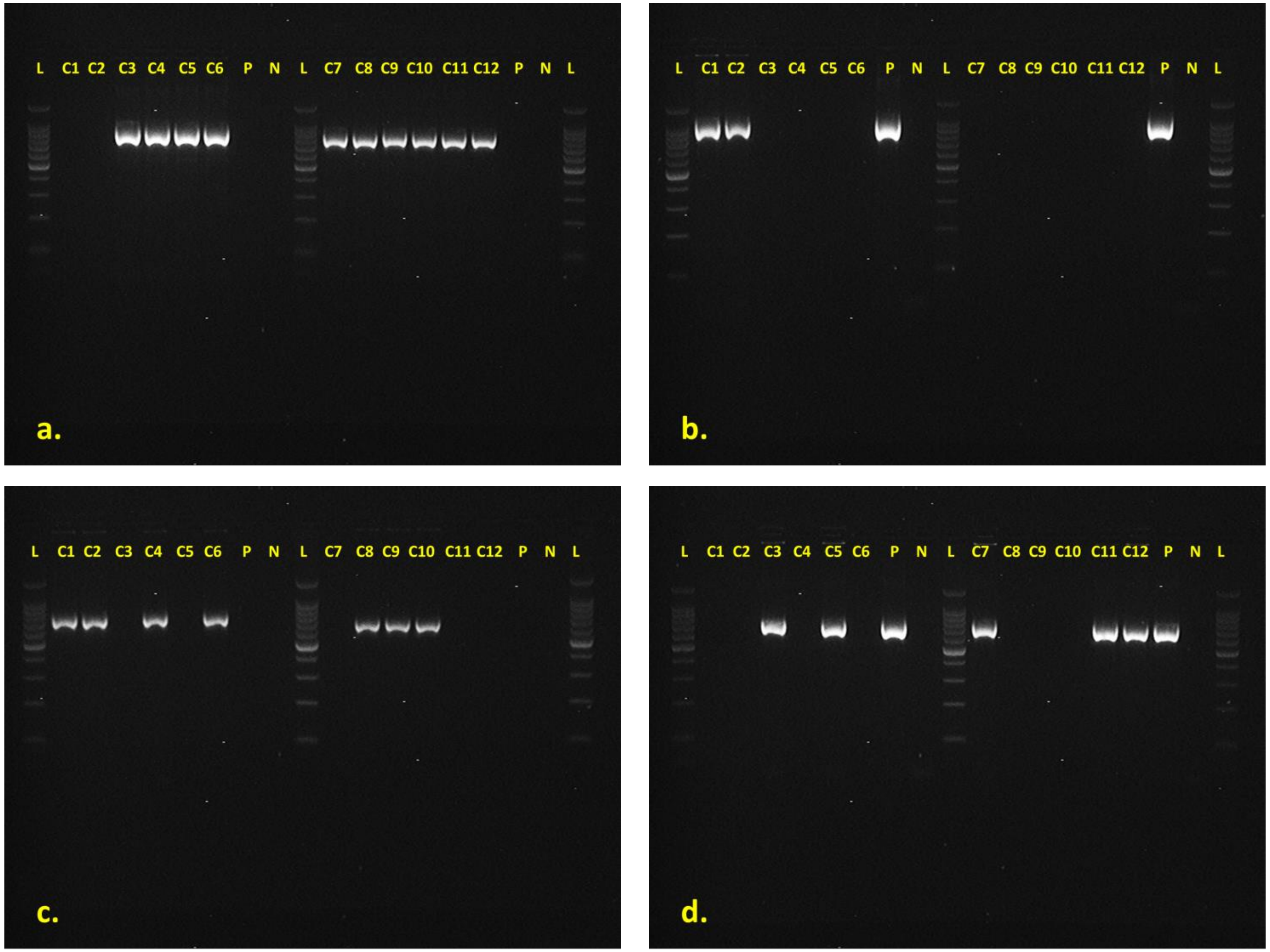

L C1 C2 C3 C4 C5 C6 P N L C7 C8 C9 C10 C11 C12 P N L
a.
L C1 C2 C3 C4 C5 C6 P N L C7 C8 C9 C10 C11 C12 P N L
b.
L C1 C2 C3 C4 C5 C6 P N L C7 C8 C9 C10 C11 C12 P N L
c.
L C1 C2 C3 C4 C5 C6 P N L C7 C8 C9 C10 C11 C12 P N L
d.

Figure 5.9. Gel electrophoresis of PCR products to verify site and direction of inserts. See text for explanation.

Figure 5.9 shows the results of the agarose gel electrophoresis of PCR products obtained from PCR runs performed to detect the site and direction of integration of the pK19mob-gene construct into the *Rlv*. 3841 genome. For all the four gels, the labelling convention is as follows L = DNA MW ladder (SmartLadder MW-1700-10 from Eurogentec for gels a and c and 100 bp ladder from Invitrogen for gels b and d), C1-C12 = test clones, P = *Rlv*. 3841, N = Negative control.

Gels a and b = testing for pRL100135 clones, c and d = testing for pRL100136 clones. Gels a and c = Chromosomal (pRL10) forward primer and M13 reverse primer pair (ChrF+M13R), Gels b and d = Chromosomal (pRL10) forward and chromosomal (pRL10) reverse primer pair (ChrF+ChrR).

The above gel pictures suggest that for pRL100135 clones (gels a and b), C3-12 have the integration of the plasmid in the target gene and in the right orientation. Clones C1 and C2 show amplification with the ChrF+ChrR primer pair (gel b) indicating no disruption of the primer. The *Rlv*. 3841 intact gene also amplifies with the ChrF+ ChrR primer pair (gel b, lanes marked P). The negative control lanes show no non-specific amplification.

For pRL100136 (gels c and d) clones C1, C2, C4, C6, C8-C10 have the integration of the plasmid in the target gene and in the right orientation. Clones C3, C5, C7, C11 and C12 show amplification with the ChrF+ChrR primer pair (gel d) indicating no disruption of the primer. The *Rlv*. 3841 intact gene also amplifies with the ChrF+ ChrR primer pair (gel d, lanes marked P). The negative control lanes show no non-specific amplification.

As a result of the above experiments, 10 clones with a disruption in gene pRL100135 (-pRL100135) and 7 clones with a disruption in gene pRL100136 (-pRL100136) were obtained. These insertion mutants possessed a disrupted gene that was hypothesized to be important in the utilization of GBL. Hence these mutants could now be tested for their ability to utilize GBL as a carbon source and the utilization pattern compared with that of native *Rlv*. 3841.

5.3.2. Growth studies on native *Rlv*. 3841 and insertion mutants :

The mutants in the genes pRL100135 and pRL100136 were tested for their ability to grow on minimal medium supplemented with GBL as the sole source of carbon. The experimental setup is described in the materials and methods. The results of the growth studies are shown here. In the following growth curve figures note the meaning of the labels :

Strain	Strain Description	Growth medium	Antibiotic in medium
*Rlv*3841(N)	Native *Rlv* 3841	Y medium	-
*Rlv*3841(S)	Native *Rlv* 3841	Y medium	Streptomycin
pRL100135(I)	*Rlv* 3841 with pK19mob and intact pRL100135	Y medium	Streptomycin Neomycin
pRL100135(D)	*Rlv* 3841 with disrupted pRL100135	Y medium	Streptomycin Neomycin
pRL100136(I)	*Rlv* 3841 with pK19mob and intact pRL100136	Y medium	Streptomycin Neomycin
pRL100136(D)	*Rlv* 3841 with disrupted pRL100136	Y medium	Streptomycin Neomycin
pRL100135(C)	pRL100135(D) complemented with intact copy of pRL100135	Y medium	Streptomycin Tetracycline
pRL100136(C)	pRL100136(D) complemented with intact copy of pRL100136	Y medium	Streptomycin Tetracycline

5.3.2.1. Growth of wild type and mutants in synthetic medium with GBL as the carbon source :

Table 5.1. Absorbance values at 590 nm for the growth of wild type and target and non-target mutants in synthetic medium with GBL as the carbon source at the end of 0, 12, 24, 36, 48, 60 and 72 hours.

	A_{590} at the following time intervals (in hours)						
Culture↓	0	12	24	36	48	60	72
*Rlv*3841(N)	0.13	0.154	0.198	0.256	0.298	0.328	0.346
*Rlv*3841(S)	0.124	0.136	0.171	0.226	0.275	0.305	0.325
pRL100135(I)	0.102	0.106	0.128	0.164	0.208	0.236	0.246

pRL100135(D)	0.140	0.142	0.146	0.150	0.154	0.159	0.161
pRL100136(I)	0.122	0.126	0.152	0.197	0.244	0.278	0.294
pRL100136(D)	0.129	0.133	0.145	0.16	0.177	0.19	0.199

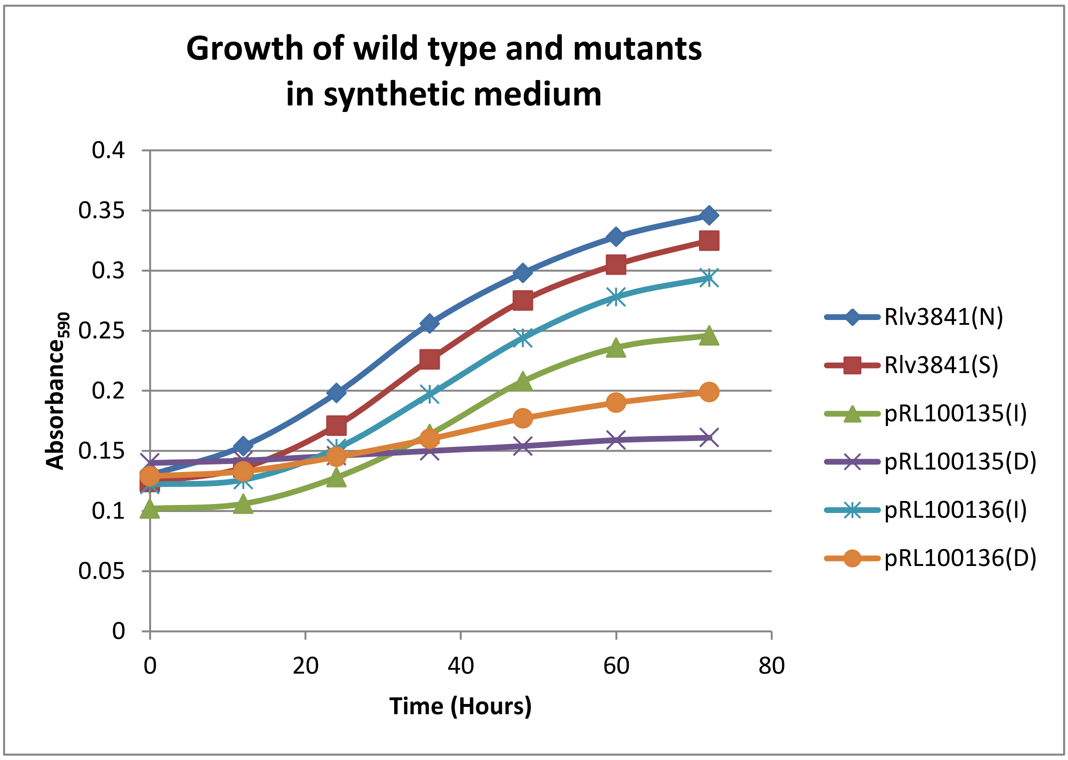


Figure 5.10. Growth curves of native *Rlv* 3841 and constructs carrying insert of pK19 mob integrated at target and non-target site in synthetic medium. Except for *Rlv*3841(N), the other cultures are growth in presence of antibiotics as mentioned in the text.

5.3.2.2. Growth of wild type and mutants in TY medium in absence of antibiotics :

Table 5.2. Absorbance values at 590 nm for the growth of wild type and mutants in TY medium without antibiotics at the end of 72 hours.

	A_{590} at the following time intervals (in hours)						
Culture↓	0	12	24	36	48	60	72
Rlv3841(N)	0.107	0.14	0.23	0.36	0.5	0.635	0.768
pRL100135(I)	0.106	0.13	0.2	0.325	0.465	0.62	0.743
pRL100135(D)	0.105	0.13	0.195	0.32	0.46	0.615	0.739
pRL100136(I)	0.108	0.155	0.26	0.38	0.51	0.64	0.747
pRL100136(D)	0.106	0.13	0.195	0.315	0.445	0.595	0.729

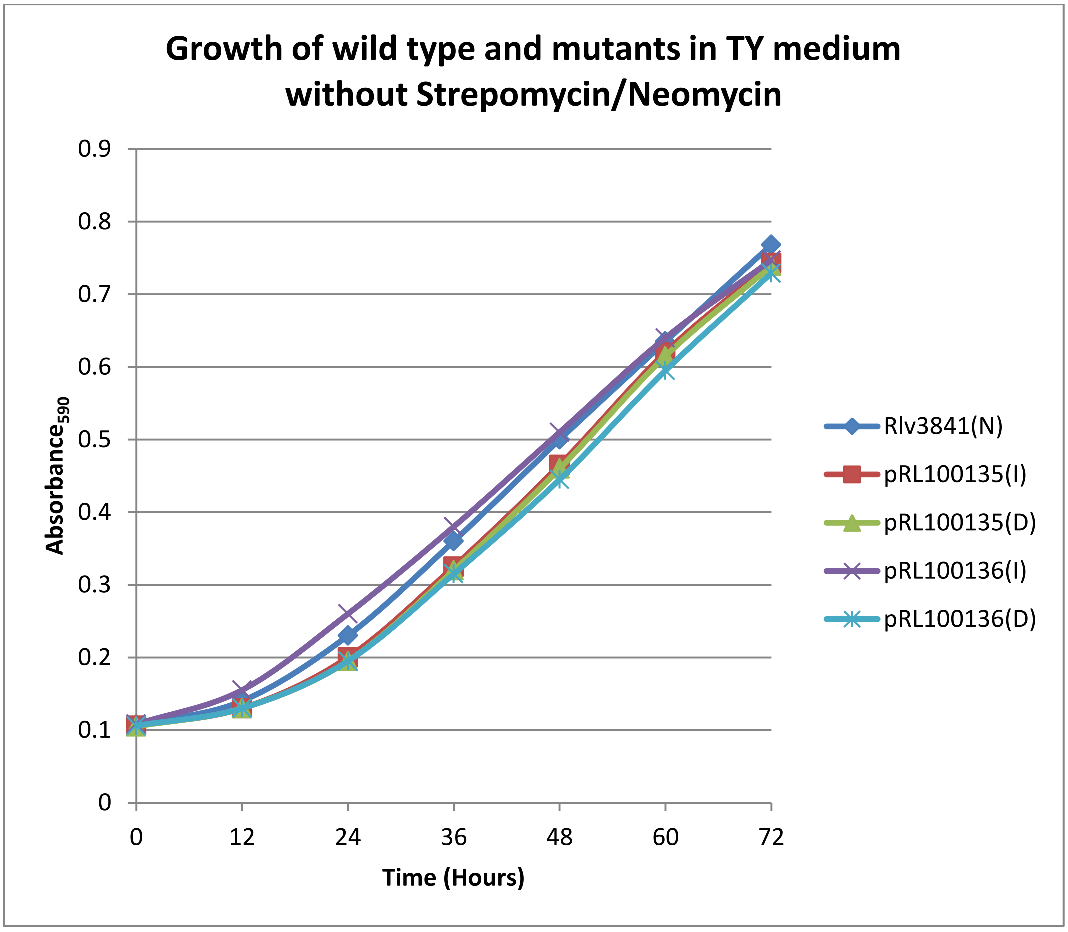


Figure 5.11. Growth curves of native *Rlv* 3841 and constructs carrying insert of pK19 mob integrated at target and non-target site grown in TY medium.

5.3.2.3. Growth of wild type and mutants in TY medium in presence of antibiotics :

Table 5.3. Absorbance values at 590 nm for the growth experiment of wild type and mutants in TY medium with antibiotics at the end of 0, 12, 24, 36, 48, 60 and 72 hours.

	A_{590} at the following time intervals (in hours)						
Culture↓	0	12	24	36	48	60	72
Rlv3841(N)	0.107	0.14	0.23	0.36	0.5	0.635	0.768
Rlv3841(S)	0.105	0.126	0.184	0.311	0.441	0.566	0.695
pRL100135(I)	0.106	0.130	0.190	0.320	0.450	0.580	0.688
pRL100135(D)	0.103	0.130	0.180	0.300	0.430	0.550	0.657
pRL100136(I)	0.106	0.13	0.18	0.3	0.435	0.55	0.675
pRL100136(D)	0.103	0.13	0.18	0.3	0.43	0.55	0.652

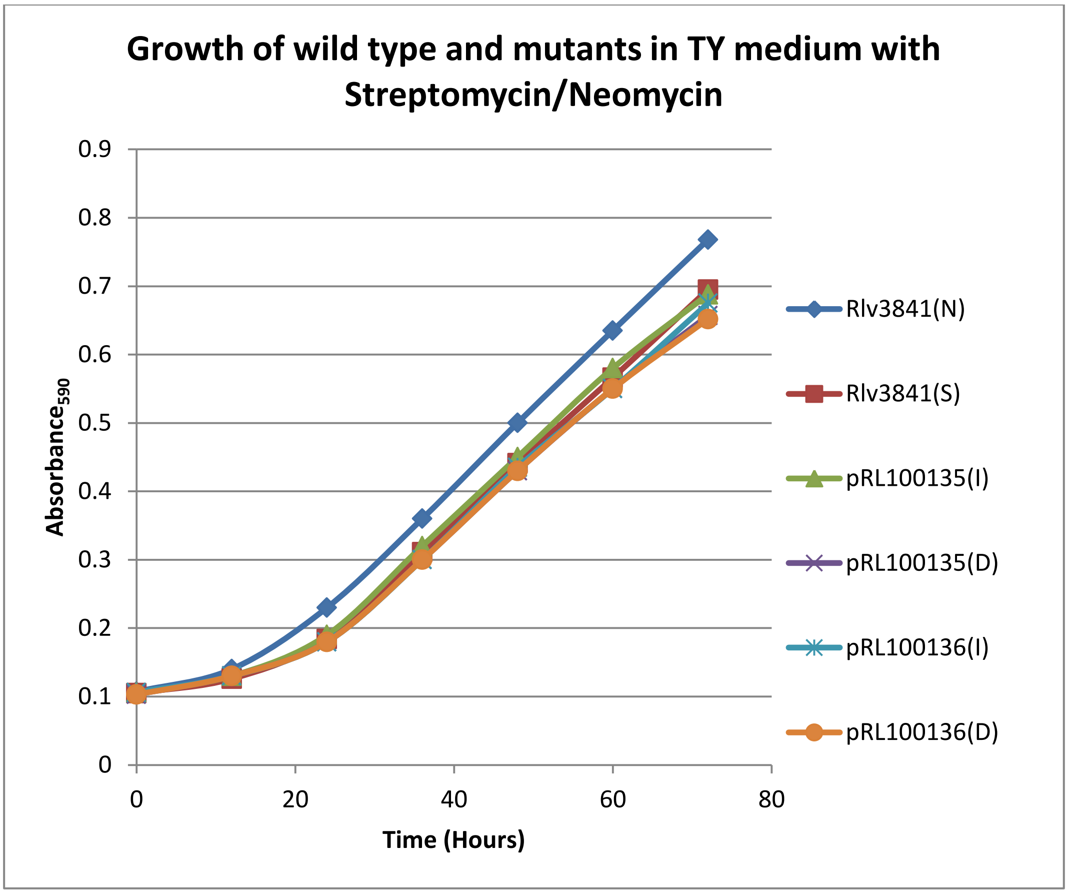


Figure 5.12. Growth curves of native *Rlv* 3841 and constructs carrying insert of pK19 mob integrated at target and non-target site in TY medium. Except for *Rlv*3841(N), the other cultures are growth in presence of antibiotics as mentioned in the text..

The results of the growth experiments demonstrate that mutation in the genes pRL100135 and pRL100136 affect the ability of the strains to grow on GBL. The results of the growth assay in the minimal medium using GBL as the carbon substrate are tabulated in Table 5.4.

The table shows that the native *Rlv*. 3841 grows better in the absence of streptomycin i.e. the presence of streptomycin decreases the growth of the *Rlv*. 3841 strain. The mutants labelled as without insert carried the insert at a site other than the target gene and served as a control to check the effect of antibiotics streptomycin and neomycin when the target get was not mutated. In this case too, it was found that the presence of the antibiotics decreased the final growth yield. Hence, it can be concluded that the presence of antibiotics confers a 'metabolic load' on the growth of the bacterium resulting in a marginal decrease in growth.

Table 5.4. Tabulation of growth kinetic assay results to calculate the doubling time and increase in the biomass in terms of absorbance at 590 nm as compared to the starting absorbance values.

Culture	Replicate	Initial OD	Final OD	Average Initial	Average Final	% increase in growth	Doubling time (hr.)
Rlv3841 (N)	1	0.126	0.335	0.130	0.346	165.47	50.79
	2	0.118	0.325				
	3	0.147	0.378				
Rlv3841 (S)	1	0.125	0.327	0.124	0.325	163.07	48.52
	2	0.122	0.323				
	3	0.124	0.326				
pRL100135 (I)	1	0.097	0.223	0.102	0.246	141.97	50.11
	2	0.102	0.249				
	3	0.106	0.266				
pRL100135 (D)	1	0.124	0.145	0.140	0.161	14.49	333.16
	2	0.129	0.148				
	3	0.168	0.189				
pRL100136 (I)	1	0.125	0.297	0.122	0.294	141.92	50.43
	2	0.119	0.291				
	3	0.121	0.295				
pRL100136 (D)	1	0.138	0.200	0.129	0.199	54.01	105.37
	2	0.116	0.193				
	3	0.133	0.203				

From the table above, it is seen that the growth is most severely affected when the gene pRL100135 is mutated. This gene encodes a dehydrogenase that converts GHB to SSA. A mutation in this gene decreases the growth to 14% of the wild-type. The 14% might be a result of the autolytic products of cell death of a fraction of a population allowing a small amount of growth of the survivors.

The mutation in gene pRL100136 decreases the growth levels achieved by the mutant as compared to wild-type but not to the extent observed for pRL100135 mutant. This may be due to a number of reasons. The first reason lies in the very nature of the substrate GBL. GBL undergoes slow spontaneous conversion to GHB (Ciolino et al., 2001) which can be utilized by the pRL100136 mutants to grow slowly. Alternately, other lactonases (on pRL11) or beta-lactamases may slowly convert GBL to GHB by utilizing GBL as a low-affinity alternate substrate thereby ensuring a slow and steady supply of GHB for the growth of pRL100136 mutants.

5.3.3. Complementation studies on pRL100135 and pRL100136 :

The mutations in the pRL100135 and pRL100136 genes were complemented using a novel method based on the method used by Vanderlinde et al. (2013).. The method for making gene complemented strains is outlined in the materials and methods section. The results of this part of the work are presented below.

5.3.3.1. Whole-gene amplification for complementation :

For complementation to work, the primers were designed to be as close as possible to the start and stop codon of the two genes. This was done so that the start codon was in close proximity to the start codon of the *trp* promoter when the gene was cloned in pDG71 using the BamHI cloning site.

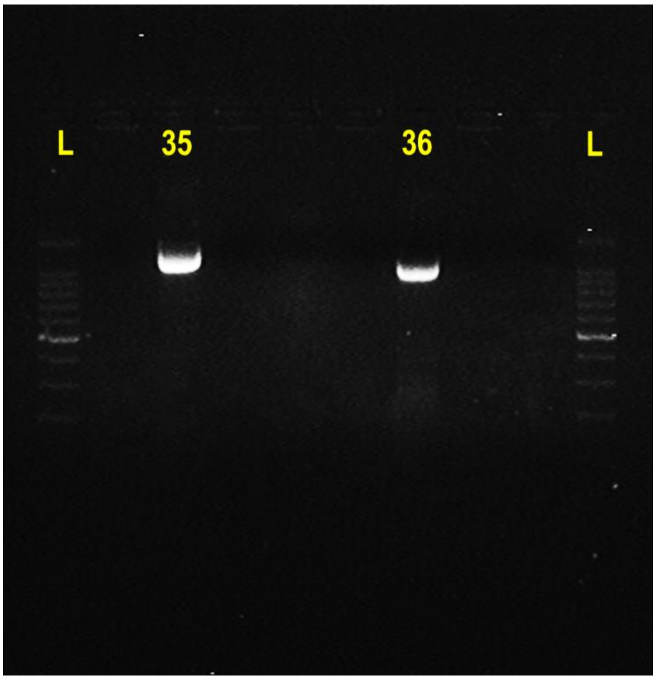


Figure 5.13. Amplification of genes pRL100135 and pRL100136 : Gel electrophoresis of the whole gene amplifications of genes pRL100135 (lane marked 35) and pRL100136 (lane marked 36). The DNA molecular weight markers used is the 100 bp ladder from Invitrogen

The PCR products were purified using QIAquick PCR Purification Kit from Qiagen using the manufacturer’s protocol before using it for ligation.

5.3.3.2. Preparation of adapter complex, adapter and cloning of the construct into pDG71 :

The adapter complex was cut with BamHI, dephosphorylated to prevent self-ligation, mixed with linearized pDG71 in presence of ligase and polynucleotide kinase to get a recombinant molecule as mentioned in the materials and methods.

Aliquots from all the manipulations were saved to analyse on a single gel by electrophoresis, the results of which are presented below.

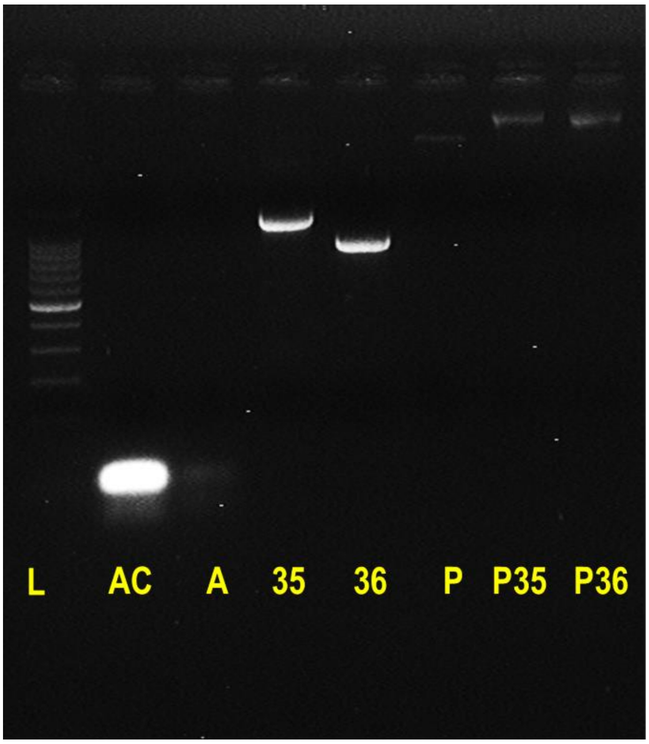


Figure 5.14. Gel electrophoresis of aliquots from DNA manipulations to generate recombinant pDG71 plasmid. Lane L = 100 bp DNA ladder from Invitrogen, Lane AC = purified adapter complex dimer, Lane A = purified and dephosphorylated adapter after digestion with BamHI, Lane 35 = purified pRL100135 whole gene PCR product, Lane 36 = purified pRL100136 whole gene PCR product, Lane P = plasmid pDG71, Lane P35 = plasmid pDG71 carrying gene pRL100135, Lane P36 = plasmid pDG71 the gene pRL100136.

From the above gel image it is clear that a recombinant plasmid was obtained. The recombinant plasmid preparation was used to transform chemically competent *Rlv*. 3841 cells which were selected on Streptomycin-Tetracycline TY agar plates.

5.3.3.3. Screening of *Rlv*. 3841 transformants using PCR :

Colonies of six transformants for each of the two gene preparations were selected to check the direction of the insert within the recombinant plasmid harbored within the bacterial cells. A PCR with a forward primer within the *trp* promoter and an internal gene primer used for making insertion mutants (IntR) were used to check the orientation of insert.

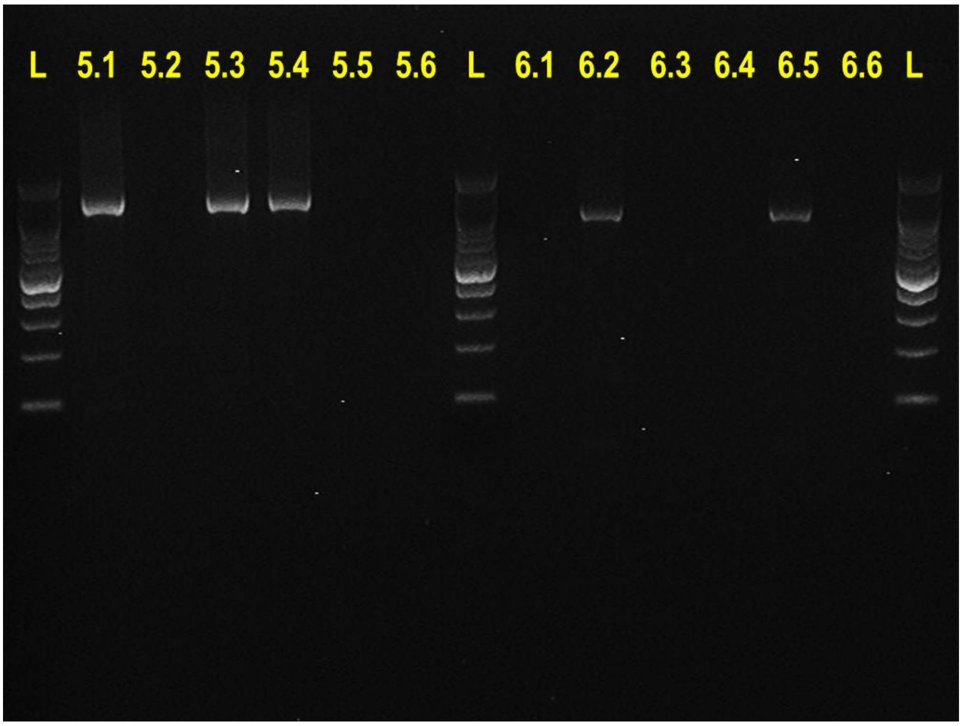


Figure 5.15. Gel electrophoresis to determine orientation of insert : Colony PCR of six colonies each of pRL100135 transformants (5.1 to 5.6) and pRL100136 (6.1 to 6.6) was performed using a forward primer within the *trp* promoter and a reverse primer within the gene. 3 clones of pRL100135 (5.1, 5.3 and 5.4) and 2 clones of pRL100136 (6.2 and 6.5) were found to have the insert in the correct orientation. Lane L = 100 bp DNA ladder from Invitrogen.

The preparation of strains complemented for mutations in the genes pRL100135 and pRL100136 resulted in three complemented clones for the former gene and two for the latter. The complemented strains were tested for their ability to utilize GBL as a carbon source. The native *Rlv*. 3841 strain and the mutants in pRL100135 and pRL100136 were also included in the growth studies.

5.3.4. Growth studies on native *Rlv*. 3841, insertion mutants and gene complemented strains :

The *Rlv*. 3841 deletion mutants complemented with pDG71 bearing a functional copy of the gene was tested for its ability to use GBL as a carbon source. Native *Rlv*. 3841, the strains having non-site-specific insertion and the mutants in the genes pRL100135 and pRL100136 were also included in the test for comparison of growth.

The experimental setup is the same as in Section 5.2.2. The comparison of growth in Y medium and TY medium, however, was not carried out for the complementation strains since the effect of antibiotics on the growth of bacteria has already been demonstrated.

The results of the growth studies are shown here. In the following growth curves note the meaning of the culture labels are as mentioned in section 5.2 and are reproduced below :

Strain	Strain Description	Growth medium	Antibiotic in medium
*Rlv*3841(N)	Native *Rlv* 3841	Y medium	-
*Rlv*3841(S)	Native *Rlv* 3841	Y medium	Streptomycin
pRL100135(I)	*Rlv* 3841 with pK19mob and intact pRL100135	Y medium	Streptomycin Neomycin
pRL100135(D)	*Rlv* 3841 with disrupted pRL100135	Y medium	Streptomycin Neomycin
pRL100136(I)	*Rlv* 3841 with pK19mob and intact pRL100136	Y medium	Streptomycin Neomycin
pRL100136(D)	*Rlv* 3841 with disrupted pRL100136	Y medium	Streptomycin Neomycin
pRL100135(C)	pRL100135(D) complemented with intact copy of pRL100135	Y medium	Streptomycin Tetracycline
pRL100136(C)	pRL100136(D) complemented with intact copy of pRL100136	Y medium	Streptomycin Tetracycline

Table 5.5. Absorbance values at 590 nm and doubling times for the growth of wild type and mutants in synthetic medium with GBL as the carbon source at the end of 72 hours.

Culture	Replicate	Initial OD	Final OD	Average Initial OD	Average Final OD	% increase	Doubling time (hr.)
Rlv3841(N)	1	0.082	0.517	0.092	0.522	464.98	33.62
	2	0.117	0.534				
	3	0.078	0.514				
Rlv3841(S)	1	0.103	0.489	0.095	0.480	403.15	36.50
	2	0.085	0.472				
	3	0.098	0.478				
pRL100135 (I)	1	0.104	0.406	0.108	0.410	278.77	37.83
	2	0.112	0.414				
	3	0.109	0.411				
pRL100135 (D)	1	0.081	0.110	0.090	0.107	18.82	126.86
	2	0.092	0.104				
	3	0.098	0.108				
pRL100136 (I)	1	0.110	0.428	0.114	0.422	270.18	39.12
	2	0.115	0.420				
	3	0.117	0.418				
pRL100136 (D	1	0.118	0.213	0.120	0.217	80.06	82.88
	2	0.122	0.221				
	3	0.121	0.216				
pRL100135 (C)	1	0.094	0.369	0.102	0.373	266.89	37.74
	2	0.106	0.379				
	3	0.105	0.371				
pRL100136 (C)	1	0.085	0.357	0.088	0.352	298.11	35.05
	2	0.091	0.347				
	3	0.089	0.351				

Table 5.6. Absorbance values at 590 nm for the growth of wild type, target and non-target mutants and complemented strains in synthetic medium with GBL as the carbon source at the end of 0, 12, 24, 36, 48, 60 and 72 hours.

	A_{590} at the following time intervals (in hours)						
Culture↓	0	12	24	36	48	60	72
*Rlv*3841(N)	0.092	0.132	0.246	0.355	0.433	0.484	0.522
*Rlv*3841(S)	0.095	0.143	0.248	0.340	0.404	0.453	0.480
pRL100135(I)	0.108	0.136	0.188	0.249	0.313	0.372	0.410
pRL100135(D)	0.090	0.093	0.096	0.100	0.102	0.105	0.107
pRL100136(I)	0.114	0.148	0.200	0.266	0.331	0.387	0.422
pRL100136(D)	0.120	0.131	0.152	0.169	0.186	0.203	0.217
pRL100135(C)	0.102	0.126	0.161	0.210	0.272	0.331	0.373
pRL100136(C)	0.088	0.106	0.142	0.183	0.242	0.302	0.352

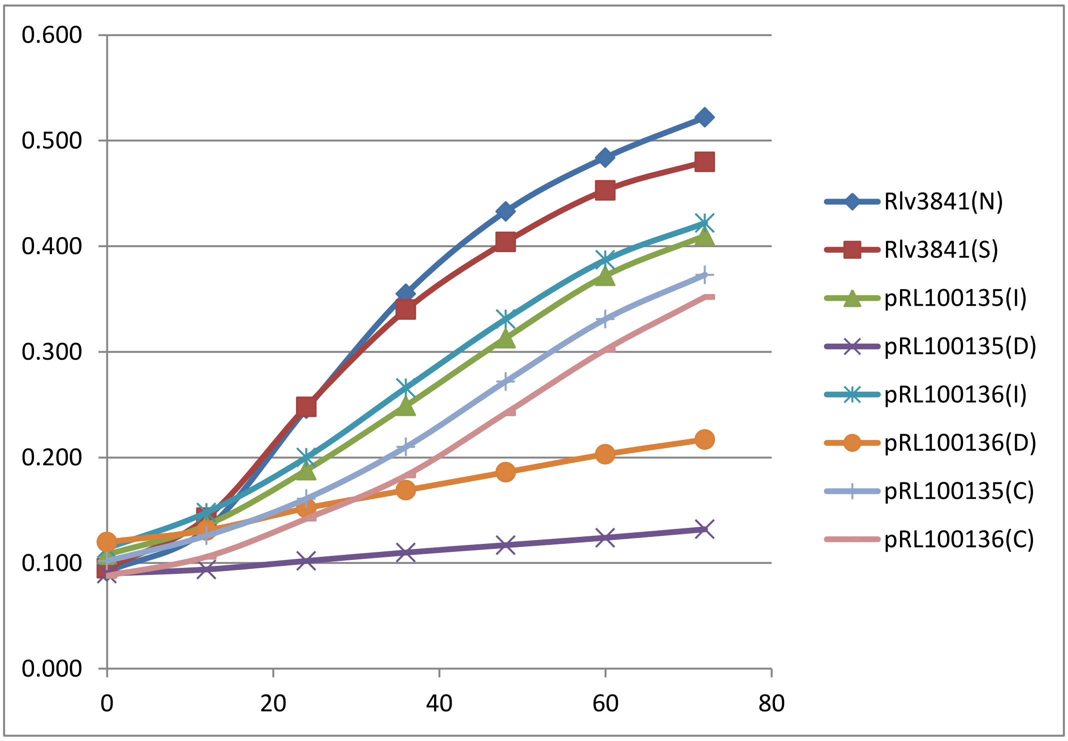


Figure 5.16. Growth curves of native *Rlv* 3841 and constructs carrying insert of pK19 mob integrated at target and non-target site and the complemented strains grown in synthetic medium. Except for *Rlv*3841(N) the other cultures are growth in presence of antibiotics as mentioned in the text..

Figure 5.16 shows that the complemented strains are able to grow better than the mutants. Their growth is comparable to that of mutants carrying the insert at non-target site and not as much as high as the native 3841 strain.

On analysis of the generation time of the cultures, the result of analysis indicates that there is a significant difference in the medians, (Kruskal-Wallis: *N* = 24, d.f. = 7, *p* = 0.002). Follow-up tests (Mann-Whitney) conducted to evaluate pairwise differences among the six cultures indicated that there is no a significant difference of generation time between the pRL100135(I) and pRL100135(C). All other cultures show a significant difference.

The results of the growth studies and the supporting statistics indicate that there is a significant difference between the growths obtained for each of the cultures under the given test conditions. This indicates that the effect of streptomycin on the growth of *Rlv*. 3841 is significant; the difference between the growth of *Rlv*. 3841 and the clones having target and non-target insertions is significant as well. There is no significant difference between the generation times of pRL100135(I) and pRL100135(C). This indicates that on complementing the mutant in pRL100135, the growth of the mutant is reversed to that of a construct carrying pK19mob insert at a non-target site. The most important by far is that there is a significant difference in the growth of the mutants with inserts in target site (5 with insert and 6 with insert) and their respective complemented strains (5 complemented and 6 complemented) which states that the complementation of the gene mutation has been successful.

5.4. Discussion :

The aim of this chapter was to study the role of the genes pRL100135 and pRL100136 on the utilization of GBL by mutation and complementation. The mutation of the genes using a gene inactivation protocol that inserts a plasmid at a site homologous to the portion cloned into the plasmid resulted in two types of mutants – a mutant carrying the insertion at the target site and a mutant carrying the insertion at the non-target site. The non-target site integration would have occurred due to the presence of other regions with sequences closely related to the target site. Moreover, both the types of mutants were used to study the effect of the insertion on the growth of the organism. It was observed that the non-target site insertions showed a decrease in the growth as compared to the native *Rlv*. 3841 culture. This might be due to the presence of the antibiotic that might confer a stress that alters and decreases the growth rate since a similar effect is observed when the native *Rlv*. 3841 strain is grown in the presence of the antibiotic streptomycin.

Mutation of the gene pRL100135 severely affected the ability of the bacteria to grow on GBL. The small amount of growth that is observed might be a due to the presence of the precursors being released from dying cells that allows a slow and steady growth of the survivors. Mutation of the gene pRL100136 markedly decreases the growth of the mutant but does not abolish it completely. The reasons, as stated before, might be attributed to the slow spontaneous conversion to GHB (Ciolino et al., 2001) which can be utilized by the pRL100136 mutants to grow slowly. Or to the presence of other lactonases (on pRL11) or beta-lactamases may slowly convert GBL to GHB by utilizing GBL as an alternate low-affinity substrate thereby ensuring a slow and steady supply of GHB for the growth of pRL100136 mutants.

The complemented strains for genes pRL100135 and pRL100136 showed a significantly better growth than the mutants. This indicated that the genes that were cloned individually into pDG71 and transformed into the mutated *Rlv*. 3841 were functional. The growth, although significantly higher than that of the mutants was also statistically lower than either the native 3841 grown in the presence or absence of streptomycin or the strains carrying the inserts at non-target sites (pRL100135 intact, pRL100136 intact). The possible reasons for this might be the presence of the polyketide antibiotic, tetracycline, in the system. Tetracycline inhibits protein synthesis and it may be possible that a strain that has gained resistance because of introduction of resistance gene during cloning might showed an altered growth phenotype. The other possible explanation could be that the gene cloned under the influence of the *trp* promoter in pDG71 since the *trp* promoter is derived from a *Salmonella* spp. it might not express itself as good as the native 3841 promoter leading results in decreased growth.

CHAPTER 6 : DISCUSSION.

6.1. Introduction :

The aim of this thesis was to look for association between genotypes and phenotypes as suggested by association studies and to verify these associations based on laboratory work. The bacterial isolates used in this study were collected from Wentworth College, University of York, as a part of a population genomics project (Bailly et al., 2011). Of the 72 strains from the Wentworth population, 36 belonged to the biovar *trifolii* and the remaining 36 to biovar *viciae*. Besides these 72 strains, the strain *R. leguminosarum* biovar *viciae* 3841 was included in the study since its genome has been sequenced (Young et al., 2006) and would hence serve as a useful reference organism.

In Chapter 2 of the thesis, after verifying the ability of the isolates to nodulate plants, we tested the isolates to see if the differences between the two biovars of *R. leguminosarum* viz. *trifolii* and *viciae* that are mentioned in literature were reflected in the strains. The differences studied were the ability of the strains to use homoserine as a sole carbon source and the ability of the strains to synthesize AHL quorum sensing signal molecules. The *biovar viciae* were generally more adept at utilizing homoserine as suggested by Egeraat (1975). Similarly, a greater number of strains belonging to biovar *viciae* were also found to synthesizing AHL quorum sensing signal molecules than biovar *trifolii* strains. Both the biovars showed an effect that could be attributed to their distribution as cryptic species – but not entirely. Hence it was necessary to carry out a systematic evaluation of phenotypic diversity using a set of standard tests using a standardized protocol,

In Chapter 3 of the thesis, we investigated the phenotypic diversity of the strains using the Biolog Gram Negative 2 or GN2 system. This redox chemistry based system measures bacterial respiration and reports it colorimetrically by reduction of a dye. Based on the utilization / non-utilization of substrates, a pattern or “metabolic fingerprint” is obtained (Bochner, 1989). The metabolic fingerprints obtained for each of the strains was unique indicating a huge metabolic diversity even though the bacterial isolates belonged to the same species (although two different biovars) and were recovered from a small area of a meter squared. This is in-effect a proof of the huge metabolic diversity of bacteria.

In Chapter 4 of the thesis, an attempt was made to obtain correlations between the ability of isolates to utilize a specific substrate and the presence / absence of genes. The metabolic pathways of all the 95 substrates present in the Biolog GN2 plates were studied using a number of metabolic pathway tools and interface software based on these tools to arrive at 'consensus pathways' for each of the substrate. On correlating the ability of the strain to use the substrate with the presence or absence of genes, only one of the 95 substrates showed a significant correlation with the gene presence-absence data. This substrate, γ-hydroxybutyrate, was used as a case study to investigate the effect of gene mutation on the ability of the reference strain *R. leguminosarum* 3841, which was used as the test strain. The role of the genes that showed correlation with GHB in Wentworth population has already been demonstrated in the bacterium *A. tumefaciens* and hence it was thought to be worth investigating.

In Chapter 5 of the thesis, the genes postulated to be involved in GHB metabolism were mutated and effects of mutation studied. However, since GHB is a regulated substance, its immediate precursor GBL was used as the substrate to study the effect of mutation. The mutation of one of the three genes (pRL100135) that showed a strong correlation with GHB utilization greatly abolished the ability of bacteria to grow on GBL. The complementation of this mutated gene restored the ability of the mutant to utilize GBL to nearly the same level as that of native 3841. This is a proof-of-concept that it is possible to predict associations between genotype and phenotype based on sequence data and phenotype investigation.

In this part of the thesis, we have just reviewed the work done so far and will now discuss why this work is significant in the field of functional genomics. We will also discuss the contribution and limitations of this work. Based on the knowledge gained from this study, we will also discuss lines of research that can further the study of genotype-phenotype associations and deliberate on why such studies carry importance in biology.

6.2. Significance of this thesis in genotype-phenotype studies :

Since the sequencing of ϕX174 in 1977, the number of organisms that have been sequenced has been increasing rapidly. The growth in the sequence data is expected in increase in the foreseeable future, especially with the development of fast, cheap and efficient technologies for sequencing like the 454 from Roche, Illumina (Solexa) sequencing, SOLiD, Single-molecule real-time sequencing and Ion Torrent, (Margulies et al., 2005, Schuster, 2008, Rusk, 2011, Quail et al., 2012). Ongoing developments in sequencing technologies like Oxford Nanopore (Stoddart et al., 2009), Tunnelling currents DNA sequencing (Ohshiro et al., 2012, Massimiliano Di, 2013) hold the promise of delivering the $1000 genome (Service, 2006, Mardis, 2006). To use the sequence data in a meaningful way entails the necessity to find correlations between sequence and its function. The rate of decoding this information has not kept pace with the rate at which sequence data is being generated.

The correlation data between genes and function in humans is easier to assess than bacteria. This may be due to a number of factors such as the ease to assess the effect of a mutation that occurs in the population and looking for the source of the mutation. The search for the mutant gene is also helped by the fact that unlike bacterial genomes, which are polycistronic, human genes are monocistronic so the effect of each gene mutation can be studied separately. As a result a huge number of metabolic disorders and diseases in humans have now been attributed to specific genes.

The problem is more complex in bacteria, which are too small, have phenotypes that are difficult to analyse and have polycistronic genomes arranged in operons which make it difficult to study effect of individual gene mutations. This is further compounded by the horizontal transfer of genetic material and the effect of environmental factors. Hence, although more difficult than sequencing, there is now a need to find correlations between genes and functions which can then be extended to other organisms to test for function. Although some progress has been made in linking the genome and the metabolism using a systems approach, the link between the genotype and the phenotype remains tenuous (Pommerenke et al., 2010).

The genome of most living organisms consists of thousands of genes that interact in many different ways. Some of these genes are involved in series of reactions called pathways that form an essential part of an organism’s metabolism. Given the complexity of the interactions, it is a daunting and expensive task to determine the role of genes in reactions of a metabolic pathway or finding all the genes involved in a pathway. If such genes are found then their function can be verified by performing mutation studies.

Most studies carried out to study association between genotype and phenotype in bacteria use a single or few isolates of bacteria to infer the relationship. A major shortcoming of this approach is that such an analysis may introduce bias in the analysis since a large number of metabolic genes may be accessory in nature and hence may not reflect the stable core genetic makeup of the organism under study.

In this thesis we have used 72 isolates from the same bacterial species viz. *Rhizobium leguminosarum* to study the correlation between genotype and phenotype. Along with the 72 strains a reference strain that has been completely sequenced has also been included in the study. Correlations between genes and a phenotype arising from a population would be more robust than correlations arising from a small population. Using this hypothesis, the study investigated the metabolic patterns of 95 substrates present in a Biolog GN2 plate and correlated them to the genes present in the 72 isolates.

Of the 95 substrates, only those substrates that were utilized by the reference strain (i.e. *Rlv*. 3841) but showed variations in utilization amongst the Wentworth population were used for further analysis. Since it is only possible to investigate those genes that are present in 3841, only those phenotypes that 3841 has were studied, on the assumption that the substrate utilisation will reflect the presence of a gene (though a repressor gene is a logical possibility and would generate a negative correlation).

Of the 21 substrates investigated, only one substrate showed a strong correlation with the presence / absence of a set of genes. The ability of the strains to utilise the substrate γ-hydroxybutyrate showed strong correlation to the presence of two genes present on pRL10 in *Rlv*. 3841.

In order to investigate the correlation, the genes were mutated and the effect of mutation was studied. The gene pRL100135 coding for a dehydrogenase that converts GHB to succinic semialdehyde was found to be crucial in the metabolism of GHB. Mutation of this gene resulted in a near abolition of the organism’s ability to utilize GHB. Complementing the gene with a functional copy on a plasmid under the control of a promoter restored the metabolic ability to a significant level. Hence, the work and this thesis is a proof-in-concept that correlations between phenotypes and genomes are possible.

However, it is important to stress that of the 95 substrates, only one showed a strong correlation with the genetic makeup of the isolates. Hence, implementing this type of analysis for more substrates like those included in the Phenotype Microarray plates for a more genome-wide analysis is not recommended.

A better idea would probably be to make a mutant library for the organism under investigation as done by Pommerenke et al. (2010), but yet again, the cost of making such a library and then investigating the effect of mutation in each gene for each of the 20 microarray plates is a very expensive idea.

6.3. Contribution of this thesis :

The biggest contribution of this thesis is perhaps the elucidation of a pathway for GBL and GHB utilization in *Rlv*. 3841 based on a pathway postulated from observing GHB utilization pattern in 72 isolates, identifying genes strongly correlated to GHB utilization and proving the role of those genes in GHB utilization by mutation and complementation studies.

The involvement of the correlated genes in utilization of GBL in *A. tumefaciens* had already been shown by Carlier et al. (2004). The function of *gabD* (pRL100134) as a succinic semialdehyde dehydrogenase has been demonstrated by Prell et al. (2009) who also commented the the arrangement of the genes pRL100134-pRL100135 in *Rlv* 3841 is similar to the attKLM operon in *Agrobacterium tumefaciens*. The existence similar genes being involved in *Rlv.* 3841 indicates that the pathway might be conserved in a broader group of bacteria.

During the work, it was found that the Wentworth isolates were not as easy to profile using Biolog plates like most other bacteria. The protocols mentioned in this thesis offer a method that is tried and tested and found to work effectively in the metabolic profiling of rhizobial isolates.

The thesis also contributes to the creation of complementation strains by developing a protocol for complementation that can be used for any gene irrespective of the presence of a restriction site contained within it. The protocol developed to create adapters to achieve this is unique in this respect.

6.4. Final remarks :

With the arrival of cheaper techniques for sequencing, study of genotype-phenotype correlations is probably going to be the next ‘big thing’ in biology. But perhaps unravelling this correlation might not be as simple as it looks. This is seen from the work of Sugawara et al. (2013) who have sequenced 48 strains of *Sinorhizobium (Ensifer)* but have not studied metabolic phenotypes. The presence of such large genomic datasets without any information on the phenotypic functions of the genes presents a big opportunity to develop techniques or products that can do sequencing as well as study phenotypes, phenomes and fluxomes at the same time to finally unravel the secrets of the genome.

List of abbreviations

AHL	Acylated Homoserine Lactone
ANOVA	Analysis of Variance
BLAST	Basic Local Alignment Search Tool
bp	base pair
DNA	Deoxyribonucleic acid
ELISA	Enzyme-Linked Immunosorbent Assay
GBL	γ-butyrolactone
GHB	γ-hydroxybutyrate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITS	Internal transcribed spacer
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB agar	Luria-Bertani agar
MLEE	Multi-Locus Enzyme Electrophoresis
NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
NJ	Neighbour-joining
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic acid
rpm	revolutions per minute
SOC	Super Optimal broth with Catabolite repression
TY agar	Tryptone-Yeast agar
UPGMA	Unweighted Pair Group Method with Arithmetic Mean
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

LIST OF THE SEQUENCE OF PRIMERS USED

pRL100135F (in)	CACCGGAAAGCGTTTTGTAT
pRL100135R (in)	AGACCATGCGGAATGTGATA
pRL100136F (in)	CCCTTTTACCTCATCACCCA
pRL100136R (in)	ACAGAGCGAACAGTATCCAC
pRL100135F (out)	ATTTCGAACTGTCCCATCCA
pRL100135R (out)	GAAGGCGGGTTTGTTTCATC
pRL100136F (out)	AAATGCAAGGTGCACAACAT
pRL100136R (out)	ATACTCTGGTGCCTTCTTGA
pRL100135F (res)	ttttctagaCACCGGAAAGCGTTTTGTAT
pRL100135R (res)	tttaagcttAGACCATGCGGAATGTGATA
pRL100136F (res)	ttttctagaCCCTTTTACCTCATCACCCA
pRL100136R (res)	tttaagcttACAGAGCGAACAGTATCCAC

REFERENCES :

1. ALLSOPP, D., COLWELL, R. R. & HAWKSWORTH, D. L. 1995. *Microbial diversity and ecosystem function: proceedings of the IUBS/IUMS Workshop held at Egham, UK, 10-13 August 1993 in support of the IUBS/UNESCO/SCOPE "DIVERSITAS" programme*, CAB International in association with United Nations Environment Programme.
2. ALTMAN, T., TRAVERS, M., KOTHARI, A., CASPI, R. & KARP, P. D. 2013. A systematic comparison of the MetaCyc and KEGG pathway databases. *BMC Bioinformatics,* 14**,** 112.
3. ANDREEVA, I. N., KOZHARINOVA, G. M. & IZMAILOV, S. F. 1998. *SENESCENCE OF LEGUME NODULES,* New York, NY, ETATS-UNIS, Kluwer.
4. ANTONISAMY, P. & IGNACIMUTHU, S. 2010. Immunomodulatory, analgesic and antipyretic effects of violacein isolated from Chromobacterium violaceum. *Phytomedicine,* 17**,** 300-4.
5. AOKI, S., KONDO, T., PREVOST, D., NAKATA, S., KAJITA, T. & ITO, M. 2010. Genotypic and phenotypic diversity of rhizobia isolated from Lathyrus japonicus indigenous to Japan. *Syst Appl Microbiol,* 33**,** 383-97.
6. ARMITAGE, J. P., GALLAGHER, A. & JOHNSTON, A. W. 1988. Comparison of the chemotactic behaviour of Rhizobium leguminosarum with and without the nodulation plasmid. *Mol Microbiol,* 2**,** 743-8.
7. ARMON, R., POTASMAN, I. & GREEN, M. 1990. Biochemical fingerprints of Legionella spp. by the BIOLOG system: presumptive identification of clinical and environmental isolates. *Lett Appl Microbiol,* 11**,** 290-2.
8. ARNESON, N., HUGHES, S., HOULSTON, R. & DONE, S. 2008. Whole-Genome Amplification by Adaptor-Ligation PCR of Randomly Sheared Genomic DNA (PRSG). *CSH Protoc,* 2008**,** pdb prot4922.
9. AYANABA, A., HAUGLAND, R. A., SADOWSKY, M. J., UPCHURCH, R. G., WEILAND, K. D. & ZABLOTOWICZ, R. M. 1986. Rapid Colored-Nodule Assay for Assessing Root Exudate-Enhanced Competitiveness of Bradyrhizobium japonicum. *Appl Environ Microbiol,* 52**,** 847-51.
10. BADER, G. D., CARY, M. P. & SANDER, C. 2006. Pathguide: a pathway resource list. *Nucleic Acids Res,* 34**,** D504-6.
11. BAHAR, M., DE MAJNIK, J., WEXLER, M., FRY, J., POOLE, P. S. & MURPHY, P. J. 1998. A model for the catabolism of rhizopine in Rhizobium leguminosarum involves a ferredoxin oxygenase complex and the inositol degradative pathway. *Mol Plant Microbe Interact,* 11**,** 1057-68.
12. BAILLY, X., GIUNTINI, E., SEXTON, M. C., LOWER, R. P., HARRISON, P. W., KUMAR, N. & YOUNG, J. P. 2011. Population genomics of Sinorhizobium medicae based on low-coverage sequencing of sympatric isolates. *ISME J,* 5**,** 1722-34.
13. BAILLY, X., OLIVIERI, I., BRUNEL, B., CLEYET-MAREL, J.-C. & BÉNA, G. 2007. Horizontal Gene Transfer and Homologous Recombination Drive the Evolution of the Nitrogen-Fixing Symbionts of Medicago Species. *Journal of Bacteriology,* 189**,** 5223-5236.
14. BANFALVI, Z. & KONDOROSI, A. 1989. Production of root hair deformation factors by Rhizobium meliloti nodulation genes in Escherichia coli: HsnD (NodH) is involved in the plant host-specific modification of the NodABC factor. *Plant Molecular Biology,* 13**,** 1-12.
15. BASSLER, B. L., WRIGHT, M. & SILVERMAN, M. R. 1994. Multiple signalling systems controlling expression of luminescence in Vibrio harveyi: sequence and function of genes encoding a second sensory pathway. *Mol Microbiol,* 13**,** 273-86.

16. BAXEVANIS, A. D. & OUELLETTE, B. F. F. 2004. *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, Wiley.
17. BELL, E. A. 2003. Nonprotein amino acids of plants: significance in medicine, nutrition, and agriculture. *J Agric Food Chem,* 51**,** 2854-65.
18. BENIZRI, E., BAUDOIN, E. & GUCKERT, A. 2001. Root Colonization by Inoculated Plant Growth-Promoting Rhizobacteria. *Biocontrol Science and Technology,* 11**,** 557-574.
19. BERGEY, D. H. & HOLT, J. G. 1994. *Bergey's Manual of Determinative Bacteriology*, Williams & Wilkins.
20. BIANCO, L., ANGELINI, J., FABRA, A. & MALPASSI, R. 2013. Diversity and symbiotic effectiveness of indigenous rhizobia-nodulating Adesmia bicolor in soils of Central Argentina. *Curr Microbiol,* 66**,** 174-84.
21. BLOSSER, R. S. & GRAY, K. M. 2000. Extraction of violacein from Chromobacterium violaceum provides a new quantitative bioassay for N-acyl homoserine lactone autoinducers. *J Microbiol Methods,* 40**,** 47-55.
22. BOCHNER, B. R. 1989. Sleuthing out bacterial identities. *Nature,* 339**,** 157-8.
23. BOCHNER, B. R. & SAVAGEAU, M. A. 1977. Generalized indicator plate for genetic, metabolic, and taxonomic studies with microorganisms. *Appl Environ Microbiol,* 33**,** 434-44.
24. BOIVIN, C., CAMUT, S., MALPICA, C. A., TRUCHET, G. & ROSENBERG, C. 1990. Rhizobium meliloti Genes Encoding Catabolism of Trigonelline Are Induced under Symbiotic Conditions. *Plant Cell,* 2**,** 1157-1170.
25. BOLTEN, H. J., FREDRICKSON, J. K. & ELLIOTT, L. F. 1993. *Microbial ecology of the rhizosphere,* New York, Marcel Dekker Incorporated.
26. BOONE, D. R., BRENNER, D. J., STALEY, J. T., DE VOS, P., KRIEG, N. R., GARRITY, G. M. & GOODFELLOW, M. 2005. *Bergey's Manual of Systematic Bacteriology*, Springer London, Limited.
27. BOWEN, G. D. & ROVIRA, A. D. 1999. The Rhizosphere and Its Management To Improve Plant Growth. *In:* DONALD, L. S. (ed.) *Advances in Agronomy.* Academic Press.
28. BRAY, C. M. 1983. *Nitrogen Metabolism in Plants*, Longman Publishing Group.
29. BRENNER, D., STALEY, J. & KRIEG, N. 2001. Classification of Procaryotic Organisms and the Concept of Bacterial Speciation. *In:* BOONE, D. & CASTENHOLZ, R. (eds.) *Bergey's Manual® of Systematic Bacteriology.* Springer New York.
30. BROGHAMMER, A., KRUSELL, L., BLAISE, M., SAUER, J., SULLIVAN, J. T., MAOLANON, N., VINTHER, M., LORENTZEN, A., MADSEN, E. B., JENSEN, K. J., ROEPSTORFF, P., THIRUP, S., RONSON, C. W., THYGESEN, M. B. & STOUGAARD, J. 2012. Legume receptors perceive the rhizobial lipochitin oligosaccharide signal molecules by direct binding. *Proc Natl Acad Sci U S A,* 109**,** 13859-64.
31. BRUIJN, F. J., FELIX, G., GRUNENBERG, B., HOFFMANN, H. J., METZ, B., RATET, P., SIMONS-SCHREIER, A., SZABADOS, L., WELTERS, P. & SCHELL, J. 1989. Regulation of plant genes specifically induced in nitrogen-fixing nodules: role of cis-acting elements and trans-acting factors in leghemoglobin gene expression. *Plant Molecular Biology,* 13**,** 319-325.
32. BRYAN, J., BERLYN, G. & GORDON, J. 1996. Toward a new concept of the evolution of symbiotic nitrogen fixation in the Leguminosae. *Plant and Soil,* 186**,** 151-159.
33. BYERS, J. T., LUCAS, C., SALMOND, G. P. C. & WELCH, M. 2002. Nonenzymatic Turnover of an Erwinia carotovora Quorum-Sensing Signaling Molecule. *Journal of Bacteriology,* 184**,** 1163-1171.
34. CAETANO-ANOLLES, G. & GRESSHOFF, P. M. 1991. Plant Genetic Control of Nodulation. *Annual Review of Microbiology,* 45**,** 345-382.

35. CÀMARA, M., DAYKIN, M. & CHHABRA, S. R. 1998. 6.12 Detection, Purification, and Synthesis of n-acylhomoserine Lactone Quorum Sensing Signal Molecules. *In:* PETER WILLIAMS, J. K. & GEORGE, S. (eds.) *Methods in Microbiology.* Academic Press.
36. CAMPBELL, R. & GREAVES, M. P. 1990. *Anatomy and community structure of the rhizosphere,* New York, John Wiley and Sons.
37. CARLIER, A., CHEVROT, R., DESSAUX, Y. & FAURE, D. 2004. The assimilation of gamma-butyrolactone in Agrobacterium tumefaciens C58 interferes with the accumulation of the N-acyl-homoserine lactone signal. *Mol Plant Microbe Interact,* 17**,** 951-7.
38. CARLSON, R. W., PRICE, N. P. & STACEY, G. 1994. The biosynthesis of rhizobial lipo-oligosaccharide nodulation signal molecules. *Mol Plant Microbe Interact,* 7**,** 684-95.
39. CARNAHAN, A. M., JOSEPH, S. W. & JANDA, J. M. 1989. Species identification of Aeromonas strains based on carbon substrate oxidation profiles. *J Clin Microbiol,* 27**,** 2128-9.
40. CASPI, R., ALTMAN, T., DREHER, K., FULCHER, C. A., SUBHRAVETI, P., KESELER, I. M., KOTHARI, A., KRUMMENACKER, M., LATENDRESSE, M., MUELLER, L. A., ONG, Q., PALEY, S., PUJAR, A., SHEARER, A. G., TRAVERS, M., WEERASINGHE, D., ZHANG, P. & KARP, P. D. 2012. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res,* 40**,** D742-53.
41. CHA, C., GAO, P., CHEN, Y. C., SHAW, P. D. & FARRAND, S. K. 1998. Production of acyl-homoserine lactone quorum-sensing signals by gram-negative plant-associated bacteria. *Mol Plant Microbe Interact,* 11**,** 1119-29.
42. CHANG, A., SCHEER, M., GROTE, A., SCHOMBURG, I. & SCHOMBURG, D. 2009. BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Research,* 37**,** D588-D592.
43. CHAPIN, F. S. 1980. The Mineral Nutrition of Wild Plants. *Annual Review of Ecology and Systematics,* 11**,** 233-260.
44. CHEN, W.-X., TAN, Z.-Y., GAO, J.-L., LI, Y. & WANG, E.-T. 1997. Rhizobium hainanense sp. nov., Isolated from Tropical Legumes. *International Journal of Systematic Bacteriology,* 47**,** 870-873.
45. CHEN, W., WANG, E., WANG, S., LI, Y., CHEN, X. & LI, Y. 1995. Characteristics of Rhizobium tianshanense sp. nov., a moderately and slowly growing root nodule bacterium isolated from an arid saline environment in Xinjiang, People's Republic of China. *Int J Syst Bacteriol,* 45**,** 153-9.
46. CICCARELLI, F. D., DOERKS, T., VON MERING, C., CREEVEY, C. J., SNEL, B. & BORK, P. 2006. Toward Automatic Reconstruction of a Highly Resolved Tree of Life. *Science,* 311**,** 1283-1287.
47. CIOLINO, L. A., MESMER, M. Z., SATZGER, R. D., MACHAL, A. C., MCCAULEY, H. A. & MOHRHAUS, A. S. 2001. The chemical interconversion of GHB and GBL: forensic issues and implications. *J Forensic Sci,* 46**,** 1315-23.
48. COHN, J., BRADLEY DAY, R. & STACEY, G. 1998. Legume nodule organogenesis. *Trends in plant science,* 3**,** 105-110.
49. COOK, R. J. & BAKER, K. F. 1983. *The nature and practice of biological control of plant pathogens*, American Phytopathological Society.
50. CRICK, F. 1970. Central dogma of molecular biology. *Nature,* 227**,** 561-3.
51. CURL, E. A. 1982. The rhizosphere: relation to pathogen behavior and root disease. *Plant Dis.,* 66**,** 625–630.
52. CZAJKOWSKI, R. & JAFRA, S. 2009. Quenching of acyl-homoserine lactone-dependent quorum sensing by enzymatic disruption of signal molecules. *Acta Biochim Pol,* 56**,** 1-16.

53. DANIELS, R., DE VOS, D. E., DESAIR, J., RAEDSCHELDERS, G., LUYTEN, E., ROSEMEYER, V., VERRETH, C., SCHOETERS, E., VANDERLEYDEN, J. & MICHIELS, J. 2002. The cin quorum sensing locus of Rhizobium etli CNPAF512 affects growth and symbiotic nitrogen fixation. *J Biol Chem,* 277**,** 462-8.
54. DANINO, V. E., WILKINSON, A., EDWARDS, A. & DOWNIE, J. A. 2003. Recipient-induced transfer of the symbiotic plasmid pRL1JI in Rhizobium leguminosarum bv. viciae is regulated by a quorum-sensing relay. *Mol Microbiol,* 50**,** 511-25.
55. DEGEFU, T., WOLDE-MESKEL, E. & FROSTEGÅRD, Å. 2013. Phylogenetic diversity of Rhizobium strains nodulating diverse legume species growing in Ethiopia. *Systematic and Applied Microbiology,* 36**,** 272-280.
56. DEGRASSI, G., DEVESCOVI, G., SOLIS, R., STEINDLER, L. & VENTURI, V. 2007. Oryza sativa rice plants contain molecules that activate different quorum-sensing N-acyl homoserine lactone biosensors and are sensitive to the specific AiiA lactonase. *FEMS Microbiol Lett,* 269**,** 213-20.
57. DENARIE, J., DEBELLE, F. & PROME, J.-C. 1996. Rhizobium Lipo-Chitooligosaccharide Nodulation Factors: Signaling Molecules Mediating Recognition and Morphogenesis. *Annual Review of Biochemistry,* 65**,** 503-535.
58. DENARIE, J., DEBELLE, F. & ROSENBERG, C. 1992. Signaling and Host Range Variation in Nodulation. *Annual Review of Microbiology,* 46**,** 497-531.
59. DIXON, R., KENNEDY, C., KONDOROSI, A., KRISHNAPILLAI, V. & MERRICK, M. 1977. Complementation analysis of Klebsiella pneumoniae mutants defective in nitrogen fixation. *Mol Gen Genet,* 157**,** 189-98.
60. DIXON, R. O. D. & WHEELER, C. T. 1986. *Nitrogen fixation in plants*, Blackie.
61. DJEDIDI, S., YOKOYAMA, T., TOMOOKA, N., OHKAMA-OHTSU, N., RISAL, C. P., ABDELLY, C. & SEKIMOTO, H. 2011. Phenotypic and genetic characterization of rhizobia associated with alfalfa in the Hokkaido and Ishigaki regions of Japan. *Syst Appl Microbiol,* 34**,** 453-61.
62. DONG, Y. H. & ZHANG, L. H. 2005. Quorum sensing and quorum-quenching enzymes. *J Microbiol,* 43 Spec No**,** 101-9.
63. DOWNIE, A. 1991. A nod of recognition. *Current Biology,* 1**,** 382-384.
64. DOWNIE, J. A. 1998. Functions of Rhizobial Nodulation Genes. *In:* SPAINK, H., KONDOROSI, A. & HOOYKAAS, P. J. (eds.) *The Rhizobiaceae.* Springer Netherlands.
65. DOWNIE, J. A. & GONZÁLEZ, J. E. 2008. *Cell-to-cell communication in rhizobia: quorum sensing and plant signaling.,* Washington, DC., ASM Press.
66. DUDLEY, M., JACOBS, T. & LONG, S. 1987. Microscopic studies of cell divisions induced in alfalfa roots by Rhizobium meliloti. *Planta,* 171**,** 289-301.
67. DURÁN, M., PONEZI, A., FALJONI-ALARIO, A., TEIXEIRA, M. S., JUSTO, G. & DURÁN, N. 2012. Potential applications of violacein: a microbial pigment. *Medicinal Chemistry Research,* 21**,** 1524-1532.
68. DURAN, N., JUSTO, G. Z., FERREIRA, C. V., MELO, P. S., CORDI, L. & MARTINS, D. 2007. Violacein: properties and biological activities. *Biotechnol Appl Biochem,* 48**,** 127-33.
69. DURAN, N. & MENCK, C. F. 2001. Chromobacterium violaceum: a review of pharmacological and industiral perspectives. *Crit Rev Microbiol,* 27**,** 201-22.
70. DUSHA, I., KOVALENKO, S., BANFALVI, Z. & KONDOROSI, A. 1987. Rhizobium meliloti insertion element ISRm2 and its use for identification of the fixX gene. *Journal of Bacteriology,* 169**,** 1403-1409.
71. EARL, C. D., RONSON, C. W. & AUSUBEL, F. M. 1987. Genetic and structural analysis of the Rhizobium meliloti fixA, fixB, fixC, and fixX genes. *Journal of Bacteriology,* 169**,** 1127-1136.

72. EDWARDS, A., FREDERIX, M., WISNIEWSKI-DYE, F., JONES, J., ZORREGUIETA, A. & DOWNIE, J. A. 2009. The cin and rai quorum-sensing regulatory systems in Rhizobium leguminosarum are coordinated by ExpR and CinS, a small regulatory protein coexpressed with CinI. *J Bacteriol,* 191**,** 3059-67.
73. EGERAAT, A. W. S. M. 1975. The possible role of homoserine in the development of Rhizobium leguminosarum in the rhizosphere of pea seedlings. *Plant and Soil,* 42**,** 381-386.
74. EUGENIA MARQUINA, M., ENRIQUE GONZALEZ, N. & CASTRO, Y. 2011. [Phenotypic and genotypic characterization of twelve rhizobial isolates from different regions of Venezuela]. *Rev Biol Trop,* 59**,** 1017-36.
75. FAHRAEUS, G. 1957. The Infection of Clover Root Hairs by Nodule Bacteria Studied by a Simple Glass Slide Technique. *Journal of General Microbiology,* 16**,** 374-381.
76. FALL, D., DIOUF, D., OURARHI, M., FAYE, A., ABDELMOUNEN, H., NEYRA, M., SYLLA, S. N. & MISSBAH EL IDRISSI, M. 2008. Phenotypic and genotypic characteristics of Acacia senegal (L.) Willd. root-nodulating bacteria isolated from soils in the dryland part of Senegal. *Lett Appl Microbiol,* 47**,** 85-97.
77. FEDERLE, M. J. 2009. Autoinducer-2-based chemical communication in bacteria: complexities of interspecies signaling. *Contrib Microbiol,* 16**,** 18-32.
78. FEIL, E. J. 2004. Small change: keeping pace with microevolution. *Nat Rev Micro,* 2**,** 483-495.
79. FIRMIN, J. L., WILSON, K. E., ROSSEN, L. & JOHNSTON, A. W. B. 1986. Flavonoid activation of nodulation genes in Rhizobium reversed by other compounds present in plants. *Nature,* 324**,** 90-92.
80. FISCHER, H. M. 1994. Genetic regulation of nitrogen fixation in rhizobia. *Microbiol Rev,* 58**,** 352-86.
81. FISHER, R. F. & LONG, S. R. 1992. Rhizobium-plant signal exchange. *Nature,* 357**,** 655-660.
82. FLEISCHMANN, R. D., ADAMS, M. D., WHITE, O., CLAYTON, R. A., KIRKNESS, E. F., KERLAVAGE, A. R., BULT, C. J., TOMB, J. F., DOUGHERTY, B. A., MERRICK, J. M. & ET AL. 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science,* 269**,** 496-512.
83. FORNEY, L. J., ZHOU, X. & BROWN, C. J. 2004. Molecular microbial ecology: land of the one-eyed king. *Curr Opin Microbiol,* 7**,** 210-20.
84. FOSTER, R. C. 1986. The Ultrastructure of the Rhizoplane and Rhizosphere. *Annual Review of Phytopathology,* 24**,** 211-234.
85. FRANK, B. 1889. Ueber die Pilzsymbiose der Leguminosen. *Berichte der Deutschen Botanischen Gesellschaft,* 7**,** 332-346.
86. FRANSSEN, H., VIJN, I., YANG, W. & BISSELING, T. 1992. Developmental aspects of the Rhizobium-legume symbiosis. *Plant Molecular Biology,* 19**,** 89-107.
87. FRED, E. B., BALDWIN, I. L. & MCCOY, E. 1932. *Root Nodule Bacteria and Leguminous Plants*, Parallel Press.
88. FRY, J., WOOD, M. & POOLE, P. S. 2001. Investigation of myo-inositol catabolism in Rhizobium leguminosarum bv. viciae and its effect on nodulation competitiveness. *Mol Plant Microbe Interact,* 14**,** 1016-25.
89. FUJIBUCHI, W., GOTO, S., MIGIMATSU, H., UCHIYAMA, I., OGIWARA, A., AKIYAMA, Y. & KANEHISA, M. 1998. DBGET/LinkDB: an integrated database retrieval system. *Pac Symp Biocomput***,** 683-94.
90. FUQUA, W. C., WINANS, S. C. & GREENBERG, E. P. 1994. Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *J Bacteriol,* 176**,** 269-75.

91. GEURTS, R. & BISSELING, T. 2002. Rhizobium Nod Factor Perception and Signalling. *The Plant Cell Online,* 14**,** S239-S249.
92. GEURTS, R., FEDOROVA, E. & BISSELING, T. 2005. Nod factor signaling genes and their function in the early stages of Rhizobium infection. *Curr Opin Plant Biol,* 8**,** 346-52.
93. GILLIS, M., VANDAMME, P., VOS, P., SWINGS, J. & KERSTERS, K. 2005. Polyphasic Taxonomy. *In:* BRENNER, D., KRIEG, N., STALEY, J. & GARRITY, G. (eds.) *Bergey's Manual® of Systematic Bacteriology.* Springer US.
94. GOLDMANN, A., BOIVIN, C., FLEURY, V., MESSAGE, B., LECOEUR, L., MAILLE, M. & TEPFER, D. 1991. Betaine use by rhizosphere bacteria: genes essential for trigonelline, stachydrine, and carnitine catabolism in Rhizobium meliloti are located on pSym in the symbiotic region. *Mol Plant Microbe Interact,* 4**,** 571-8.
95. GONZÁLEZ, J. E. & MARKETON, M. M. 2003. Quorum Sensing in Nitrogen-Fixing Rhizobia. *Microbiology and Molecular Biology Reviews,* 67**,** 574-592.
96. GOTO, S., BONO, H., OGATA, H., FUJIBUCHI, W., NISHIOKA, T., SATO, K. & KANEHISA, M. 1997. Organizing and computing metabolic pathway data in terms of binary relations. *Pac Symp Biocomput***,** 175-86.
97. GOTTSCHALK, G. 1986. *Bacterial Metabolism*, Springer-Verlag.
98. GOTTSCHALK, G. 2012. *Bacterial Metabolism*, Springer London, Limited.
99. GOUGH, C. 2003. Rhizobium symbiosis: insight into Nod factor receptors. *Curr Biol,* 13**,** R973-5.
100. GRAY, K. M. & GAREY, J. R. 2001. The evolution of bacterial LuxI and LuxR quorum sensing regulators. *Microbiology,* 147**,** 2379-87.
101. GUERROUJ, K., RUÍZ-DÍEZ, B., CHAHBOUNE, R., RAMÍREZ-BAHENA, M.-H., ABDELMOUMEN, H., QUIÑONES, M. A., EL IDRISSI, M. M., VELÁZQUEZ, E., FERNÁNDEZ-PASCUAL, M., BEDMAR, E. J. & PEIX, A. 2013. Definition of a novel symbiovar (sv. retamae) within Bradyrhizobium retamae sp. nov., nodulating Retama sphaerocarpa and Retama monosperma. *Systematic and Applied Microbiology,* 36**,** 218-223.
102. HACKER, J. & CARNIEL, E. 2001. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. *EMBO Rep,* 2**,** 376-81.
103. HADRI, A.-E., SPAINK, H., BISSELING, T. & BREWIN, N. 1998. Diversity of Root Nodulation and Rhizobial Infection Processes. *In:* SPAINK, H., KONDOROSI, A. & HOOYKAAS, P. J. (eds.) *The Rhizobiaceae.* Springer Netherlands.
104. HALE, M. G. & MOORE, L. D. 1980. Factors Affecting Root Exudation Ii: 1970–1978. *In:* BRADY, N. C. (ed.) *Advances in Agronomy.* Academic Press.
105. HAN, T. X., WANG, E. T., WU, L. J., CHEN, W. F., GU, J. G., GU, C. T., TIAN, C. F. & CHEN, W. X. 2008. Rhizobium multihospitium sp. nov., isolated from multiple legume species native of Xinjiang, China. *Int J Syst Evol Microbiol,* 58**,** 1693-9.
106. HANIN, M., SAÏD, J., JOHN, B. W., DOLORES, Q.-V. & RÉMY, F. 1999. Molecular aspects of host-specific nodulation. *Plant-microbe interactions*.
107. HANZELKA, B. L. & GREENBERG, E. P. 1995. Evidence that the N-terminal region of the Vibrio fischeri LuxR protein constitutes an autoinducer-binding domain. *Journal of Bacteriology,* 177**,** 815-7.
108. HANZELKA, B. L., PARSEK, M. R., VAL, D. L., DUNLAP, P. V., CRONAN, J. E., JR. & GREENBERG, E. P. 1999. Acylhomoserine lactone synthase activity of the Vibrio fischeri AinS protein. *J Bacteriol,* 181**,** 5766-70.
109. HARDIE, K. 1985. The Effect of Removal of Extraradical Hyphae on Water Uptake by Vesicular- Arbuscular Mycorrhizal Plants. *New Phytologist,* 101**,** 677-684.

110. HARTWIG, U. A., JOSEPH, C. M. & PHILLIPS, D. A. 1991. Flavonoids Released Naturally from Alfalfa Seeds Enhance Growth Rate of Rhizobium meliloti. *Plant Physiol,* 95**,** 797-803.

111. HARTWIG, U. A. & PHILLIPS, D. A. 1991. Release and Modification of nod-Gene-Inducing Flavonoids from Alfalfa Seeds. *Plant Physiol,* 95**,** 804-7.

112. HEIDSTRA, R., GEURTS, R., FRANSSEN, H., SPAINK, H. P., VAN KAMMEN, A. & BISSELING, T. 1994. Root Hair Deformation Activity of Nodulation Factors and Their Fate on Vicia sativa. *Plant Physiol,* 105**,** 787-797.

113. HEIDSTRA, R. & TON, B. 1996. Nod Factor-Induced Host Responses and Mechanisms of Nod Factor Perception. *New Phytologist,* 133**,** 25-43.

114. HEIDSTRA, R., YANG, W. C., YALCIN, Y., PECK, S., EMONS, A. M., VAN KAMMEN, A. & BISSELING, T. 1997. Ethylene provides positional information on cortical cell division but is not involved in Nod factor-induced root hair tip growth in Rhizobium-legume interaction. *Development,* 124**,** 1781-7.

115. HENIKOFF, S., HAUGHN, G. W., CALVO, J. M. & WALLACE, J. C. 1988. A large family of bacterial activator proteins. *Proceedings of the National Academy of Sciences,* 85**,** 6602-6606.

116. HOLDEN, M. T., RAM CHHABRA, S., DE NYS, R., STEAD, P., BAINTON, N. J., HILL, P. J., MANEFIELD, M., KUMAR, N., LABATTE, M., ENGLAND, D., RICE, S., GIVSKOV, M., SALMOND, G. P., STEWART, G. S., BYCROFT, B. W., KJELLEBERG, S. & WILLIAMS, P. 1999. Quorum-sensing cross talk: isolation and chemical characterization of cyclic dipeptides from Pseudomonas aeruginosa and other gram-negative bacteria. *Mol Microbiol,* 33**,** 1254-66.

117. HORNG, Y. T., DENG, S. C., DAYKIN, M., SOO, P. C., WEI, J. R., LUH, K. T., HO, S. W., SWIFT, S., LAI, H. C. & WILLIAMS, P. 2002. The LuxR family protein SpnR functions as a negative regulator of N-acylhomoserine lactone-dependent quorum sensing in Serratia marcescens. *Mol Microbiol,* 45**,** 1655-71.

118. HORNUNG, C., POEHLEIN, A., HAACK, F. S., SCHMIDT, M., DIERKING, K., POHLEN, A., SCHULENBURG, H., BLOKESCH, M., PLENER, L., JUNG, K., BONGE, A., KROHN-MOLT, I., UTPATEL, C., TIMMERMANN, G., SPIECK, E., POMMERENING-RÖSER, A., BODE, E., BODE, H. B., DANIEL, R., SCHMEISSER, C. & STREIT, W. R. 2013. The *Janthinobacterium* sp. HH01 Genome Encodes a Homologue of the *V. cholerae* CqsA and *L. pneumophila* LqsA Autoinducer Synthases. *PLoS ONE,* 8**,** e55045.

119. HYNES, M. F. & O'CONNELL, M. P. 1990. Host plant effect on competition among strains of Rhizobium leguminosarum. *Canadian Journal of Microbiology,* 36**,** 864-869.

120. IISMAA, S. E. & WATSON, J. M. 1989. The nifA gene product from Rhizobium leguminosarum biovar trifolii lacks the N-terminal domain found in other NifA proteins. *Molecular Microbiology,* 3**,** 943-955.

121. JACOBS, G. H., CHEN, A., STEVENS, S. G., STOCKWELL, P. A., BLACK, M. A., TATE, W. P. & BROWN, C. M. 2009. Transterm: a database to aid the analysis of regulatory sequences in mRNAs. *Nucleic Acids Research,* 37**,** D72-D76.

122. JOHNSTON, A. W. B. 1988. *Genetic factors affecting host range in Rhizobium leguminosarum,* St. Paul, Minn., APS Press.

123. JOHNSTON, A. W. B. & BERINGER, J. E. 1977. Genetic Hydridization of Root-Nodule Bacteria (Rhizobium). *In:* HOLLAENDER, A., BURRIS, R. H., DAY, P. R., HARDY, R. W. F., HELINSKI, D. R., LAMBORG, M. R., OWENS,

L. & VALENTINE, R. C. (eds.) *Genetic Engineering for Nitrogen Fixation.* Springer US.

124. 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**,** 136-139.

125. KANEHISA, M., ARAKI, M., GOTO, S., HATTORI, M., HIRAKAWA, M., ITOH, M., KATAYAMA, T., KAWASHIMA, S., OKUDA, S., TOKIMATSU, T. & YAMANISHI, Y. 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Res,* 36**,** D480-4.

126. KANG, S. & MILLS, A. L. 2004. Soil Bacterial Community Structure Changes Following Disturbance of the Overlying Plant Community. *Soil Science,* 169**,** 55-65.

127. KAPLAN, H. B. & GREENBERG, E. P. 1985. Diffusion of autoinducer is involved in regulation of the Vibrio fischeri luminescence system. *J Bacteriol,* 163**,** 1210-4.

128. KARP, P. D., RILEY, M., SAIER, M., PAULSEN, I. T., PALEY, S. M. & PELLEGRINI-TOOLE, A. 2000. The EcoCyc and MetaCyc databases. *Nucleic Acids Res,* 28**,** 56-9.

129. KARUNAKARAN, R., HAAG, A. F., EAST, A. K., RAMACHANDRAN, V. K., PRELL, J., JAMES, E. K., SCOCCHI, M., FERGUSON, G. P. & POOLE, P. S. 2010. BacA is essential for bacteroid development in nodules of galegoid, but not phaseoloid, legumes. *J Bacteriol,* 192**,** 2920-8.

130. KATAN, J. 1996. *Soil solarization: Integrated control aspects.,* St Paul, APS Press.

131. KAUFMAN, S. C., CERASO, D. & SCHUGURENSKY, A. 1986. First case report from Argentina of fatal septicemia caused by Chromobacterium violaceum. *J Clin Microbiol,* 23**,** 956-8.

132. KIDAJ, D., WIELBO, J. & SKORUPSKA, A. 2012. Nod factors stimulate seed germination and promote growth and nodulation of pea and vetch under competitive conditions. *Microbiol Res,* 167**,** 144-50.

133. KINZIG, A. P. & SOCOLOW, R. H. 1994. Human impacts on the nitrogen cycle. *Journal Name: Physics Today; (United States); Journal Volume: 47:11***,** Medium: X; Size: Pages: 24-31.

134. KLEEREBEZEM, M., QUADRI, L. E., KUIPERS, O. P. & DE VOS, W. M. 1997. Quorum sensing by peptide pheromones and two-component signal-transduction systems in Gram-positive bacteria. *Mol Microbiol,* 24**,** 895-904.

135. KNEE, E. M., GONG, F. C., GAO, M., TEPLITSKI, M., JONES, A. R., FOXWORTHY, A., MORT, A. J. & BAUER, W. D. 2001. Root mucilage from pea and its utilization by rhizosphere bacteria as a sole carbon source. *Mol Plant Microbe Interact,* 14**,** 775-84.

136. KÖDÖBÖCZ, L., HALBRITTER, A., MOGYORÓSSY, T. & KECSKÉS, M. L. 2009. Phenotypic and genotypic diversity of rhizobia in cropping areas under intensive and organic agriculture in Hungary. *European Journal of Soil Biology,* 45**,** 394-399.

137. KUYKENDALL, D. L. & YOUNG, J. M. 2005. *Rhizobium,* New York, Springer.

138. LAMBEIN, F., KHAN, J. K., KUO, Y. H., CAMPBELL, C. G. & BRIGGS, C. J. 1993. Toxins in the seedlings of some varieties of grass pea (Lathyrus sativus). *Nat Toxins,* 1**,** 246-9.

139. LAMETA, A. A. & JAY, M. 1987. Study of soybean and lentil root exudates. *Plant and Soil,* 101**,** 267-272.

140. LAUE, B. E., JIANG, Y., CHHABRA, S. R., JACOB, S., STEWART, G. S. A. B., HARDMAN, A., DOWNIE, J. A., O'GARA, F. & WILLIAMS, P. 2000. The biocontrol strain Pseudomonas fluorescens F113 produces the

Rhizobium small bacteriocin, N-(3-hydroxy-7-cis-tetradecenoyl)homoserine lactone, via HdtS, a putative novel N-acylhomoserine lactone synthase. *Microbiology,* 146**,** 2469-2480.

141. LEGENDRE, P. & LEGENDRE, L. 2012. *Numerical Ecology*, Elsevier Science.

142. LEIGH, J. A. & WALKER, G. C. 1994. Exopolysaccharides of Rhizobium: synthesis, regulation and symbiotic function. *Trends Genet,* 10**,** 63-7.

143. LETUNIC, I., YAMADA, T., KANEHISA, M. & BORK, P. 2008. iPath: interactive exploration of biochemical pathways and networks. *Trends Biochem Sci,* 33**,** 101-3.

144. LICHSTEIN, H. C. & VAN DE SAND, V. F. 1946. The Antibiotic Activity of Violacein, Prodigiosin, and Phthiocol. *J Bacteriol,* 52**,** 145-6.

145. LIN, D. X., WANG, E. T., TANG, H., HAN, T. X., HE, Y. R., GUAN, S. H. & CHEN, W. X. 2008. Shinella kummerowiae sp. nov., a symbiotic bacterium isolated from root nodules of the herbal legume Kummerowia stipulacea. *Int J Syst Evol Microbiol,* 58**,** 1409-13.

146. LITHGOW, J. K., WILKINSON, A., HARDMAN, A., RODELAS, B., WISNIEWSKI-DYE, F., WILLIAMS, P. & DOWNIE, J. A. 2000. The regulatory locus cinRI in Rhizobium leguminosarum controls a network of quorum-sensing loci. *Mol Microbiol,* 37**,** 81-97.

147. LIU, T. Y., LI JR, Y., LIU, X. X., SUI, X. H., ZHANG, X. X., WANG, E. T., CHEN, W. X., CHEN, W. F. & PUŁAWSKA, J. 2012. Rhizobium cauense sp. nov., isolated from root nodules of the herbaceous legume Kummerowia stipulacea grown in campus lawn soil. *Systematic and Applied Microbiology,* 35**,** 415-420.

148. LOPES, S. C., BLANCO, Y. C., JUSTO, G. Z., NOGUEIRA, P. A., RODRIGUES, F. L., GOELNITZ, U., WUNDERLICH, G., FACCHINI, G., BROCCHI, M., DURAN, N. & COSTA, F. T. 2009. Violacein extracted from Chromobacterium violaceum inhibits Plasmodium growth in vitro and in vivo. *Antimicrob Agents Chemother,* 53**,** 2149-52.

149. LUDWIG, W. & KLENK, H.-P. 2000. Overview: A phylogenetic backbone and taxonomic framework for procaryotic systematics. *In:* BOONE, D. R., CASTENHOLZ, R. W. & GARRITY, G. M. (eds.) *Bergey's Manual of Systematic Bacteriology.* New York, N.Y.: Springer-Verlag.

150. LUNDBERG, D. S., LEBEIS, S. L., PAREDES, S. H., YOURSTONE, S., GEHRING, J., MALFATTI, S., TREMBLAY, J., ENGELBREKTSON, A., KUNIN, V., DEL RIO, T. G., EDGAR, R. C., EICKHORST, T., LEY, R. E., HUGENHOLTZ, P., TRINGE, S. G. & DANGL, J. L. 2012. Defining the core Arabidopsis thaliana root microbiome. *Nature,* 488**,** 86-90.

151. MAATALLAH, J., BERRAHO, E. B., MUNOZ, S., SANJUAN, J. & LLUCH, C. 2002. Phenotypic and molecular characterization of chickpea rhizobia isolated from different areas of Morocco. *J Appl Microbiol,* 93**,** 531-40.

152. MARDIS, E. R. 2006. Anticipating the 1,000 dollar genome. *Genome Biol,* 7**,** 112.

153. MAREK-KOZACZUK, M., LESZCZ, A., WIELBO, J., WDOWIAK-WRÓBEL, S. & SKORUPSKA, A. 2013. Rhizobium pisi sv. trifolii K3.22 harboring nod genes of the Rhizobium leguminosarum sv. trifolii cluster. *Systematic and Applied Microbiology,* 36**,** 252-258.

154. MARGULIES, M., EGHOLM, M., ALTMAN, W. E., ATTIYA, S., BADER, J. S., BEMBEN, L. A., BERKA, J., BRAVERMAN, M. S., CHEN, Y. J., CHEN, Z., DEWELL, S. B., DU, L., FIERRO, J. M., GOMES, X. V., GODWIN, B. C., HE, W., HELGESEN, S., HO, C. H., IRZYK, G. P., JANDO, S. C., ALENQUER, M. L., JARVIE, T. P., JIRAGE, K. B., KIM, J. B., KNIGHT, J. R., LANZA, J. R., LEAMON, J. H., LEFKOWITZ, S. M., LEI, M., LI, J., LOHMAN, K. L., LU, H., MAKHIJANI, V. B., MCDADE, K. E., MCKENNA, M. P., MYERS, E. W.,

NICKERSON, E., NOBILE, J. R., PLANT, R., PUC, B. P., RONAN, M. T., ROTH, G. T., SARKIS, G. J., SIMONS, J. F., SIMPSON, J. W., SRINIVASAN, M., TARTARO, K. R., TOMASZ, A., VOGT, K. A., VOLKMER, G. A., WANG, S. H., WANG, Y., WEINER, M. P., YU, P., BEGLEY, R. F. & ROTHBERG, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437**,** 376-80.

155. MARKETON, M. M. & GONZÁLEZ, J. E. 2002. Identification of Two Quorum-Sensing Systems in Sinorhizobium meliloti. *Journal of Bacteriology,* 184**,** 3466-3475.

156. MARTINEZ, E., ROMERO, D. & PALACIOS, R. 1990. The Rhizobium Genome. *Critical Reviews in Plant Sciences,* 9**,** 59-93.

157. MASSIMILIANO DI, V. 2013. Fast DNA sequencing by electrical means inches closer. *Nanotechnology,* 24**,** 342501.

158. MATHESIUS, U., MULDERS, S., GAO, M., TEPLITSKI, M., CAETANO-ANOLLÉS, G., ROLFE, B. G. & BAUER, W. D. 2003. Extensive and specific responses of a eukaryote to bacterial quorum-sensing signals. *Proceedings of the National Academy of Sciences,* 100**,** 1444-1449.

159. MAUCHLINE, W. S. & KEEVIL, C. W. 1991. Development of the BIOLOG substrate utilization system for identification of Legionella spp. *Appl Environ Microbiol,* 57**,** 3345-9.

160. MAY, G., BRUMMER, B. & OTT, H. 1991. *Treatment of prophylaxis of polio and herpes virus infections - comprises admin. of 3-(di:hydro-5-(hydroxy-1H- indolyl-2-oxo-3H-pyrrolidene)-di:hydro-2H-indole*. Germany patent application DE3935066.

161. MAZUR, A., STASIAK, G., WIELBO, J., KOPER, P., KUBIK-KOMAR, A. & SKORUPSKA, A. 2013. Phenotype profiling of Rhizobium leguminosarum bv. trifolii clover nodule isolates reveal their both versatile and specialized metabolic capabilities. *Archives of Microbiology,* 195**,** 255-267.

162. MAZUR, A., STASIAK, G., WIELBO, J., KUBIK-KOMAR, A., MAREK-KOZACZUK, M. & SKORUPSKA, A. 2011. Intragenomic diversity of Rhizobium leguminosarum bv. trifolii clover nodule isolates. *BMC Microbiol,* 11**,** 123.

163. MCCLEAN, K. H., WINSON, M. K., FISH, L., TAYLOR, A., CHHABRA, S. R., CAMARA, M., DAYKIN, M., LAMB, J. H., SWIFT, S., BYCROFT, B. W., STEWART, G. S. & WILLIAMS, P. 1997. Quorum sensing and Chromobacterium violaceum: exploitation of violacein production and inhibition for the detection of N-acylhomoserine lactones. *Microbiology,* 143 (Pt 12)**,** 3703-11.

164. MCINROY, S. G., CAMPBELL, C. D., HAUKKA, K. E., ODEE, D. W., SPRENT, J. I., WANG, W. J., YOUNG, J. P. & SUTHERLAND, J. M. 1999. Characterisation of rhizobia from African acacias and other tropical woody legumes using Biolog and partial 16S rRNA sequencing. *FEMS Microbiol Lett,* 170**,** 111-7.

165. MEHNAZ, S., BAIG, D. N. & LAZAROVITS, G. 2010. Genetic and phenotypic diversity of plant growth promoting rhizobacteria isolated from sugarcane plants growing in pakistan. *J Microbiol Biotechnol,* 20**,** 1614-23.

166. MERCADO-BLANCO, J. & TORO, N. 1996. Plasmids in Rhizobia: The Role of Nonsymbiotic Plasmids. *MPMI,* 9**,** 535-545.

167. METTING, F. B. 1993. *Soil Microbial Ecology: Applications in Agricultural and Environmental Management*, Marcel Dekker Incorporated.

168. METZLER, D. E. & METZLER, C. M. 2003. *Biochemistry: The Chemical Reactions of Living Cells*, Harcourt/Academic Press.

169. MICHIELS, J., DOMBRECHT, B., VERMEIREN, N., XI, C., LUYTEN, E. & VANDERLEYDEN, J. 1998. Phaseolus vulgaris is a non-selective host for nodulation. *FEMS Microbiology Ecology,* 26**,** 193-205.

170. MIERZWA, B., WDOWIAK-WROBEL, S. & MALEK, W. 2009. Phenotypic, genomic and phylogenetic characteristics of rhizobia isolated from root nodules of Robinia pseudoacacia (black locust) growing in Poland and Japan. *Arch Microbiol,* 191**,** 697-710.

171. MILTON, D. L., CHALKER, V. J., KIRKE, D., HARDMAN, A., CAMARA, M. & WILLIAMS, P. 2001. The LuxM homologue VanM from Vibrio anguillarum directs the synthesis of N-(3-hydroxyhexanoyl)homoserine lactone and N-hexanoylhomoserine lactone. *J Bacteriol,* 183**,** 3537-47.

172. MINOGUE, T. D., WEHLAND-VON TREBRA, M., BERNHARD, F. & VON BODMAN, S. B. 2002. The autoregulatory role of EsaR, a quorum-sensing regulator in Pantoea stewartii ssp. stewartii: evidence for a repressor function. *Mol Microbiol,* 44**,** 1625-35.

173. MITHANI, A., PRESTON, G. M. & HEIN, J. 2009. Rahnuma: hypergraph-based tool for metabolic pathway prediction and network comparison. *Bioinformatics,* 25**,** 1831-1832.

174. MORIYA, Y., SHIGEMIZU, D., HATTORI, M., TOKIMATSU, T., KOTERA, M., GOTO, S. & KANEHISA, M. 2010. PathPred: an enzyme-catalyzed metabolic pathway prediction server. *Nucleic Acids Res,* 38**,** W138-43.

175. MOULIN, L., MUNIVE, A., DREYFUS, B. & BOIVIN-MASSON, C. 2001. Nodulation of legumes by members of the [beta]-subclass of Proteobacteria. *Nature,* 411**,** 948-950.

176. MUTCH, L. A. & YOUNG, J. P. 2004. Diversity and specificity of Rhizobium leguminosarum biovar viciae on wild and cultivated legumes. *Mol Ecol,* 13**,** 2435-44.

177. NEALSON, K. H., PLATT, T. & HASTINGS, J. W. 1970. Cellular control of the synthesis and activity of the bacterial luminescent system. *J Bacteriol,* 104**,** 313-22.

178. NELSON, D. D. L., LEHNINGER, A. L. & COX, M. M. 2013. *Lehninger Principles of Biochemistry*, W.H. Freeman.

179. NELSON, D. L. & COX, M. M. 2012. *Lehninger Principles of Biochemistry*, W. H. Freeman.

180. NEWBOULD, P. 1989. The use of nitrogen fertiliser in agriculture. Where do we go practically and ecologically? *In:* CLARHOLM, M. & BERGSTRÖM, L. (eds.) *Ecology of Arable Land — Perspectives and Challenges.* Springer Netherlands.

181. NINER, B. M. & HIRSCH, A. M. 1998. *How many Rhizobium genes, in addition to nod, nif/fix, and exo, are needed for nodule development and function ?,* Philadelphia, PA, ETATS-UNIS, Balaban.

182. NITROGEN FIXATION FOR TROPICAL AGRICULTURAL LEGUMES PROJECT, U. & FOOD 1984. *Legume Inoculants and Their Use: A Pocket Manual*, Food and Agriculture Organization of the United Nations.

183. NOUR, S. M., CLEYET-MAREL, J. C., BECK, D., EFFOSSE, A. & FERNANDEZ, M. P. 1994. Genotypic and phenotypic diversity of Rhizobium isolated from chickpea (Cicer arietinum L.). *Can J Microbiol,* 40**,** 345-54.

184. NOVICK, R. P. & GEISINGER, E. 2008. Quorum sensing in staphylococci. *Annu Rev Genet,* 42**,** 541-64.

185. OGATA, H., GOTO, S., FUJIBUCHI, W. & KANEHISA, M. 1998. Computation with the KEGG pathway database. *Biosystems,* 47**,** 119-28.

186. OHSHIRO, T., MATSUBARA, K., TSUTSUI, M., FURUHASHI, M., TANIGUCHI, M. & KAWAI, T. 2012. Single-Molecule Electrical Random Resequencing of DNA and RNA. *Sci. Rep.,* 2.

187. OLDROYD, G. E., MURRAY, J. D., POOLE, P. S. & DOWNIE, J. A. 2011. The rules of engagement in the legume-rhizobial symbiosis. *Annu Rev Genet,* 45**,** 119-44.

188. OP DEN CAMP, R. H., POLONE, E., FEDOROVA, E., ROELOFSEN, W., SQUARTINI, A., OP DEN CAMP, H. J., BISSELING, T. & GEURTS, R. 2012. Nonlegume Parasponia andersonii deploys a broad rhizobium host range strategy resulting in largely variable symbiotic effectiveness. *Mol Plant Microbe Interact,* 25**,** 954-63.

189. ORMENO-ORRILLO, E. & MARTINEZ-ROMERO, E. 2013. Phenotypic tests in Rhizobium species description: an opinion and (a sympatric speciation) hypothesis. *Syst Appl Microbiol,* 36**,** 145-7.

190. PALLERONI, N. J. 2003. Prokaryote taxonomy of the 20th century and the impact of studies on the genus Pseudomonas: a personal view. *Microbiology,* 149**,** 1-7.

191. PARKE, D. & ORNSTON, L. N. 1984. Nutritional Diversity of Rhizobiaceae Revealed by Auxanography. *Journal of General Microbiology,* 130**,** 1743-1750.

192. PARKINSON, J. S. 1995. *Genetic approaches for signaling pathways and proteins,* Washington D.C., American Society for Microbiology Press.

193. PARSEK, M. R., VAL, D. L., HANZELKA, B. L., CRONAN, J. E., JR. & GREENBERG, E. P. 1999. Acyl homoserine-lactone quorum-sensing signal generation. *Proc Natl Acad Sci U S A,* 96**,** 4360-5.

194. PEARSON, J. P., VAN DELDEN, C. & IGLEWSKI, B. H. 1999. Active Efflux and Diffusion Are Involved in Transport of Pseudomonas aeruginosa Cell-to-Cell Signals. *Journal of Bacteriology,* 181**,** 1203-1210.

195. PEREIRA, E. G., LACERDA, A. M., LIMA, A. S., MOREIRA, F. M. S., CARVALHO, D. & SIQUEIRA, J. O. 2002. Genotypic, Phenotypic and Symbiotic Diversity Amongst Rhizobia Isolates from Phaseolus vulgaris L. Growing in the Amazon Region. *In:* PEDROSA, F., HUNGRIA, M., YATES, G. & NEWTON, W. (eds.) *Nitrogen Fixation: From Molecules to Crop Productivity.* Springer Netherlands.

196. PETERS, N., FROST, J. & LONG, S. 1986. A plant flavone, luteolin, induces expression of Rhizobium meliloti nodulation genes. *Science,* 233**,** 977-980.

197. PINERO, D., MARTINEZ, E. & SELANDER, R. K. 1988. Genetic diversity and relationships among isolates of Rhizobium leguminosarum biovar phaseoli. *Appl Environ Microbiol,* 54**,** 2825-32.

198. POMMERENKE, C., MUSKEN, M., BECKER, T., DOTSCH, A., KLAWONN, F. & HAUSSLER, S. 2010. Global genotype-phenotype correlations in Pseudomonas aeruginosa. *PLoS Pathog,* 6**,** e1001074.

199. PONTE, R. & JENKINS, S. G. 1992. Fatal Chromobacterium violaceum infections associated with exposure to stagnant waters. *Pediatr Infect Dis J,* 11**,** 583-6.

200. POOLE, P. & ALLAWAY, D. 2000. Carbon and nitrogen metabolism in Rhizobium. *Adv Microb Physiol,* 43**,** 117-63.

201. POPP, C. & OTT, T. 2011. Regulation of signal transduction and bacterial infection during root nodule symbiosis. *Curr Opin Plant Biol,* 14**,** 458-67.

202. PRELL, J., BOURDES, A., KARUNAKARAN, R., LOPEZ-GOMEZ, M. & POOLE, P. 2009. Pathway of gamma-aminobutyrate metabolism in Rhizobium leguminosarum 3841 and its role in symbiosis. *J Bacteriol,* 191**,** 2177-86.

203. PRELL, J. & POOLE, P. 2006. Metabolic changes of rhizobia in legume nodules. *Trends Microbiol,* 14**,** 161-8.

204. PUEPPKE, S. G. & BROUGHTON, W. J. 1999. Rhizobium sp. strain NGR234 and R. fredii USDA257 share exceptionally broad, nested host ranges. *Mol Plant Microbe Interact,* 12**,** 293-318.

205. PUTNOKY, P. & KONDOROSI, A. 1986. Two gene clusters of Rhizobium meliloti code for early essential nodulation functions and a third influences nodulation efficiency. *Journal of Bacteriology,* 167**,** 881-887.

206. QUAIL, M. A., SMITH, M., COUPLAND, P., OTTO, T. D., HARRIS, S. R., CONNOR, T. R., BERTONI, A., SWERDLOW, H. P. & GU, Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics,* 13**,** 341.

207. RAE, A. L., BONFANTE-FASOLO, P. & BREWIN, N. J. 1992. Structure and growth of infection threads in the legume symbiosis with Rhizobium leguminosarum. *The Plant Journal,* 2**,** 385-395.

208. RAHI, P., KAPOOR, R., YOUNG, J. P. & GULATI, A. 2012. A genetic discontinuity in root-nodulating bacteria of cultivated pea in the Indian trans-Himalayas. *Mol Ecol,* 21**,** 145-59.

209. RAI, R., DASH, P. K., MOHAPATRA, T. & SINGH, A. 2012. Phenotypic and molecular characterization of indigenous rhizobia nodulating chickpea in India. *Indian J Exp Biol,* 50**,** 340-50.

210. RAMACHANDRAN, V. K., EAST, A. K., KARUNAKARAN, R., DOWNIE, J. A. & POOLE, P. S. 2011. Adaptation of Rhizobium leguminosarum to pea, alfalfa and sugar beet rhizospheres investigated by comparative transcriptomics. *Genome Biol,* 12**,** R106.

211. RAMIREZ-BAHENA, M. H., VELAZQUEZ, E., FERNANDEZ-SANTOS, F., PEIX, A., MARTINEZ-MOLINA, E. & MATEOS, P. F. 2009. Phenotypic, genotypic, and symbiotic diversities in strains nodulating clover in different soils in Spain. *Can J Microbiol,* 55**,** 1207-16.

212. RASUL, A., AMALRAJ, E. L., PRAVEEN KUMAR, G., GROVER, M. & VENKATESWARLU, B. 2012. Characterization of rhizobial isolates nodulating Millettia pinnata in India. *FEMS Microbiol Lett,* 336**,** 148-58.

213. REDMOND, J. W., BATLEY, M., DJORDJEVIC, M. A., INNES, R. W., KUEMPEL, P. L. & ROLFE, B. G. 1986. Flavones induce expression of nodulation genes in Rhizobium. *Nature,* 323**,** 632-635.

214. RIVAS, M., SEEGER, M., JEDLICKI, E. & HOLMES, D. S. 2007. Second acyl homoserine lactone production system in the extreme acidophile Acidithiobacillus ferrooxidans. *Appl Environ Microbiol,* 73**,** 3225-31.

215. ROCHE, P., LEROUGE, P., PONTHUS, C. & PROMÉ, J. C. 1991. Structural determination of bacterial nodulation factors involved in the Rhizobium meliloti-alfalfa symbiosis. *Journal of Biological Chemistry,* 266**,** 10933-10940.

216. ROCHE, P., MAILLET, F., PLAZANET, C., DEBELLÉ, F., FERRO, M., TRUCHET, G., PROMÉ, J.-C. & DÉNARIÉ, J. 1996. The common nodABC genes of Rhizobium meliloti are host-range determinants. *Proceedings of the National Academy of Sciences,* 93**,** 15305-15310.

217. RODRIGUEZ-NAVARRO, D. N., CAMACHO, M., LEIDI, E. O., RIVAS, R. & VELAZQUEZ, E. 2004. Phenotypic and genotypic characterization of rhizobia from diverse geographical origin that nodulate Pachyrhizus species. *Syst Appl Microbiol,* 27**,** 737-45.

218. ROGEL, M. A., ORMENO-ORRILLO, E. & MARTINEZ ROMERO, E. 2011. Symbiovars in rhizobia reflect bacterial adaptation to legumes. *Syst Appl Microbiol,* 34**,** 96-104.

219. ROLFE, B. G. 1988. Flavones and isoflavones as inducing substances of legume nodulation. *Biofactors,* 1**,** 3-10.

220. ROSSELLO-MORA, R. & AMANN, R. 2001. The species concept for prokaryotes. *FEMS Microbiol Rev,* 25**,** 39-67.

221. ROSSEN, L., SHEARMAN, C. A., JOHNSTON, A. W. & DOWNIE, J. A. 1985. The nodD gene of Rhizobium leguminosarum is autoregulatory and in

the presence of plant exudate induces the nodA,B,C genes. *EMBO J,* 4**,** 3369-73.

222. ROSTAS, K., KONDOROSI, E., HORVATH, B., SIMONCSITS, A. & KONDOROSI, A. 1986. Conservation of extended promoter regions of nodulation genes in Rhizobium. *Proceedings of the National Academy of Sciences,* 83**,** 1757-1761.

223. ROVIRA, A. D. 1969. Plant Root Exudates. *Botanical Review,* 35**,** 35-57.

224. RUSK, N. 2011. Torrents of sequence. *Nat Meth,* 8**,** 44-44.

225. RUVKUN, G. B. & AUSUBEL, F. M. 1980. Interspecies homology of nitrogenase genes. *Proc Natl Acad Sci U S A,* 77**,** 191-5.

226. SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. 1989. *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press.

227. SANCHEZ-CONTRERAS, M., BAUER, W. D., GAO, M., ROBINSON, J. B. & ALLAN DOWNIE, J. 2007. Quorum-sensing regulation in rhizobia and its role in symbiotic interactions with legumes. *Philos Trans R Soc Lond B Biol Sci,* 362**,** 1149-63.

228. SANGER, F., AIR, G. M., BARRELL, B. G., BROWN, N. L., COULSON, A. R., FIDDES, C. A., HUTCHISON, C. A., SLOCOMBE, P. M. & SMITH, M. 1977. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature,* 265**,** 687-95.

229. SCHAFER, A., TAUCH, A., JAGER, W., KALINOWSKI, J., THIERBACH, G. & PUHLER, A. 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**,** 69-73.

230. SCHEER, M., GROTE, A., CHANG, A., SCHOMBURG, I., MUNARETTO, C., ROTHER, M., SÖHNGEN, C., STELZER, M., THIELE, J. & SCHOMBURG, D. 2010. BRENDA, the enzyme information system in 2011. *Nucleic Acids Research*.

231. SCHELL, M. A. 1993. Molecular Biology of the LysR Family of Transcriptional Regulators. *Annual Review of Microbiology,* 47**,** 597-626.

232. SCHLAMAN, H., PHILLIPS, D. & KONDOROSI, E. 1998. *Genetic organisation and transcriptional regulation of rhizobial nodulation genes,* Dordrecht, Kluwer Academic Publishers.

233. SCHLAMAN, H. R., OKKER, R. J. & LUGTENBERG, B. J. 1992. Regulation of nodulation gene expression by NodD in rhizobia. *Journal of Bacteriology,* 174**,** 5177-5182.

234. SCHOMBURG, I., CHANG, A. & SCHOMBURG, D. 2002. BRENDA, enzyme data and metabolic information. *Nucleic Acids Research,* 30**,** 47-49.

235. SCHULTZE, M., QUICLET-SIRE, B., KONDOROSI, E., VIRELIZER, H., GLUSHKA, J. N., ENDRE, G., GÉRO, S. D. & KONDOROSI, A. 1992. Rhizobium meliloti produces a family of sulfated lipooligosaccharides exhibiting different degrees of plant host specificity. *Proceedings of the National Academy of Sciences,* 89**,** 192-196.

236. SCHUSTER, S. C. 2008. Next-generation sequencing transforms today's biology. *Nat Methods,* 5**,** 16-8.

237. SCOTT, I. M., FIRMIN, J. L., BUTCHER, D. N., SEARLE, L. M., SOGEKE, A. K., EAGLES, J., MARCH, J. F., SELF, R. & FENWICK, G. R. 1979. Analysis of a range of crown gall and normal plant tissues for Ti plasmid-determined compounds. *Molecular and General Genetics MGG,* 176**,** 57-65.

238. SERVICE, R. F. 2006. Gene sequencing. The race for the $1000 genome. *Science,* 311**,** 1544-6.

239. SHI, S., RICHARDSON, A. E., O'CALLAGHAN, M., DEANGELIS, K. M., JONES, E. E., STEWART, A., FIRESTONE, M. K. & CONDRON, L. M. 2011.

Effects of selected root exudate components on soil bacterial communities. *FEMS Microbiol Ecol,* 77**,** 600-10.

240. SNEATH, P. H. 1956. Cultural and biochemical characteristics of the genus Chromobacterium. *J Gen Microbiol,* 15**,** 70-98.

241. SNEATH, P. H. A. 1972. Chapter II Computer Taxonomy. *In:* NORRIS, J. R. & RIBBONS, D. W. (eds.) *Methods in Microbiology.* Academic Press.

242. SOEDARJO, M., HEMSCHEIDT, T. K. & BORTHAKUR, D. 1994. Mimosine, a Toxin Present in Leguminous Trees (Leucaena spp.), Induces a Mimosine-Degrading Enzyme Activity in Some Rhizobium Strains. *Appl Environ Microbiol,* 60**,** 4268-72.

243. SOKAL, R. R. & MICHENER, C. D. 1958. A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin,* 28**,** 1409-1438.

244. SOLOVYEV, V. & SALAMOV, A. 2010. *Automatic Annotation of Microbial Genomes and Metagenomic Sequences*, Nova Science Publishers, Incorporated.

245. SOMASEGARAN, P. & HOBEN, H. H. J. 1994. *Handbook for Rhizobia: Methods in Legume-Rhizobium Technology*, SPRINGER VERLAG GMBH.

246. SOMASEGARAN, P. & HOBEN, H. J. 2011. *Handbook for Rhizobia: Methods in Legume-Rhizobium Technology*, Springer London, Limited.

247. SPAINK, H. P. 2000. ROOT NODULATION AND INFECTION FACTORS PRODUCED BY RHIZOBIAL BACTERIA. *Annual Review of Microbiology,* 54**,** 257-288.

248. SPAINK, H. P., SHEELEY, D. M., VAN BRUSSEL, A. A. N., GLUSHKA, J., YORK, W. S., TAK, T., GEIGER, O., KENNEDY, E. P., REINHOLD, V. N. & LUGTENBERG, B. J. J. 1991. A novel highly unsaturated fatty acid moiety of lipo-oligosaccharide signals determines host specificity of Rhizobium. *Nature,* 354**,** 125-130.

249. SPAINK, H. P., WIJFFELMAN, C. A., PEES, E., OKKER, R. J. H. & LUGTENBERG, B. J. J. 1987. Rhizobium nodulation gene nodD as a determinant of host specificity. *Nature,* 328**,** 337-340.

250. SPAINK, H. P., WIJFJES, A. H. & LUGTENBERG, B. J. 1995. Rhizobium NodI and NodJ proteins play a role in the efficiency of secretion of lipochitin oligosaccharides. *Journal of Bacteriology,* 177**,** 6276-81.

251. SPIERS, A. J., BUCKLING, A. & RAINEY, P. B. 2000. The causes of Pseudomonas diversity. *Microbiology,* 146 (Pt 10)**,** 2345-50.

252. SPRENT, J. I. 1979. *The biology of nitrogen-fixing organisms*, McGraw-Hill.

253. SPRENT, J. I. & SPRENT, P. 1990. *Nitrogen fixing organisms: pure and applied aspects*, Chapman and Hall.

254. STAFFORD, H. 1997. Roles of flavonoids in symbiotic and defense functions in legume roots. *The Botanical Review,* 63**,** 27-39.

255. STOCK, J. B., NINFA, A. J. & STOCK, A. M. 1989. Protein phosphorylation and regulation of adaptive responses in bacteria. *Microbiol Rev,* 53**,** 450-90.

256. STODDART, D., HERON, A. J., MIKHAILOVA, E., MAGLIA, G. & BAYLEY, H. 2009. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc Natl Acad Sci U S A,* 106**,** 7702-7.

257. STOWERS, M. D. 1985. Carbon Metabolism in Rhizobium Species. *Annual Review of Microbiology,* 39**,** 89-108.

258. STRENG, A., OP DEN CAMP, R., BISSELING, T. & GEURTS, R. 2011. Evolutionary origin of rhizobium Nod factor signaling. *Plant Signal Behav,* 6**,** 1510-4.

259. STRUYS, E. A., VERHOEVEN, N. M., JANSEN, E. E., TEN BRINK, H. J., GUPTA, M., BURLINGAME, T. G., QUANG, L. S., MAHER, T., RINALDO, P., SNEAD, O. C., GOODWIN, A. K., WEERTS, E. M., BROWN, P. R., MURPHY, T. C., PICKLO, M. J., JAKOBS, C. & GIBSON, K. M. 2006. Metabolism of gamma-hydroxybutyrate to d-2-hydroxyglutarate in mammals: further evidence for d-2-hydroxyglutarate transhydrogenase. *Metabolism,* 55**,** 353-8.

260. SUGAWARA, M., EPSTEIN, B., BADGLEY, B., UNNO, T., XU, L., REESE, J., GYANESHWAR, P., DENNY, R., MUDGE, J., BHARTI, A., FARMER, A., MAY, G., WOODWARD, J., MEDIGUE, C., VALLENET, D., LAJUS, A., ROUY, Z., MARTINEZ-VAZ, B., TIFFIN, P., YOUNG, N. & SADOWSKY, M. 2013. Comparative genomics of the core and accessory genomes of 48 Sinorhizobium strains comprising five genospecies. *Genome Biology,* 14**,** R17.

261. SUN, H. & WANG, Y. 2013. Hollow Fiber Liquid-Phase Microextraction with in Situ Derivatization Combined with Gas Chromatography-Mass Spectrometry for the Determination of Root Exudate Phenylamine Compounds in Hot Pepper (Capsicum annuum L.). *J Agric Food Chem,* 61**,** 5494-9.

262. SYLLA, S. N., SAMBA, R. T., NEYRA, M., NDOYE, I., GIRAUD, E., WILLEMS, A., DE LAJUDIE, P. & DREYFUS, B. 2002. Phenotypic and genotypic diversity of rhizobia nodulating Pterocarpus erinaceus and P. lucens in Senegal. *Syst Appl Microbiol,* 25**,** 572-83.

263. TENG, P. S., SHANE, W. W. & MACKENZIE, D. R. 1984. Crop losses due to plant pathogens. *Critical Reviews in Plant Sciences,* 2**,** 21-47.

264. TRINICK, M. J. 1979. Structure of nitrogen-fixing nodules formed by Rhizobium on roots of Parasponia andersonii Planch. *Can J Microbiol,* 25**,** 565-78.

265. VAN BRUSSEL, A. A., BAKHUIZEN, R., VAN SPRONSEN, P. C., SPAINK, H. P., TAK, T., LUGTENBERG, B. J. & KIJNE, J. W. 1992. Induction of pre-infection thread structures in the leguminous host plant by mitogenic lipo-oligosaccharides of Rhizobium. *Science,* 257**,** 70-2.

266. VAN DER HEIJDEN, M. G., BARDGETT, R. D. & VAN STRAALEN, N. M. 2008. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol Lett,* 11**,** 296-310.

267. VAN RHIJN, P. & VANDERLEYDEN, J. 1995. The Rhizobium-plant symbiosis. *Microbiological Reviews,* 59**,** 124-42.

268. VANCE, C. 1998. Legume Symbiotic Nitrogen Fixation: Agronomic Aspects. *In:* SPAINK, H., KONDOROSI, A. & HOOYKAAS, P. J. (eds.) *The Rhizobiaceae.* Springer Netherlands.

269. VANDAMME, P., POT, B., GILLIS, M., DE VOS, P., KERSTERS, K. & SWINGS, J. 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiological Reviews,* 60**,** 407-38.

270. VANDERLINDE, E. M., HYNES, M. F. & YOST, C. K. 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.

271. VANDERLINDE, E. M., MUSZYŃSKI, A., HARRISON, J. J., KOVAL, S. F., FOREMAN, D. L., CERI, H., KANNENBERG, E. L., CARLSON, R. W. & YOST, C. K. 2009. Rhizobium leguminosarum biovar viciae 3841, deficient in 27-hydroxyoctacosanoate-modified lipopolysaccharide, is impaired in desiccation tolerance, biofilm formation and motility. *Microbiology,* 155**,** 3055-3069.

272. VINUESA, P. 2012. *ICSP - Subcommittee on the taxonomy of Rhizobium and Agrobacterium* [Online]. Available: http://edzna.ccg.unam.mx/rhizobial-taxonomy/node/4 [Accessed June 01, 2013.

273. VINUESA, P., SILVA, C., LORITE, M. J., IZAGUIRRE-MAYORAL, M. L., BEDMAR, E. J. & MARTÍNEZ-ROMERO, E. 2005. Molecular systematics of rhizobia based on maximum likelihood and Bayesian phylogenies inferred from rrs, atpD, recA and nifH sequences, and their use in the classification of Sesbania microsymbionts from Venezuelan wetlands. *Systematic and Applied Microbiology,* 28**,** 702-716.

274. VIRTANEN, A. I. & MIETTINEN, J. K. 1953. On the composition of the soluble nitrogen fraction in the pea plant and alder. *Biochim Biophys Acta,* 12**,** 181-7.

275. VLASSAK, K. M., VANDERLEYDEN, J. & GRAHAM, P. H. 1997. Factors Influencing Nodule Occupancy by Inoculant Rhizobia. *Critical Reviews in Plant Sciences,* 16**,** 163-229.

276. VOET, D., PRATT, C. W. & VOET, J. G. 2013. *Principles of Biochemistry*, John Wiley & Sons, Limited.

277. VOET, D. & VOET, J. G. 2011. *Biochemistry*, John Wiley & Sons.

278. WALKER, T. S., BAIS, H. P., GROTEWOLD, E. & VIVANCO, J. M. 2003. Root Exudation and Rhizosphere Biology. *Plant Physiology,* 132**,** 44-51.

279. WANG, R., CHANG, Y. L., ZHENG, W. T., ZHANG, D., ZHANG, X. X., SUI, X. H., WANG, E. T., HU, J. Q., ZHANG, L. Y. & CHEN, W. X. 2013. Bradyrhizobium arachidis sp. nov., isolated from effective nodules of Arachis hypogaea grown in China. *Systematic and Applied Microbiology,* 36**,** 101-105.

280. WANG, Y. C., WANG, F., HOU, B. C., WANG, E. T., CHEN, W. F., SUI, X. H., CHEN, W. X., LI, Y. & ZHANG, Y. B. Proposal of Ensifer psoraleae sp. nov., Ensifer sesbaniae sp. nov., Ensifer morelense comb. nov. and Ensifer americanum comb. nov. *Systematic and Applied Microbiology*.

281. WATERS, C. M. & BASSLER, B. L. 2005. Quorum sensing: cell-to-cell communication in bacteria. *Annu Rev Cell Dev Biol,* 21**,** 319-46.

282. WEI, G. H., ZHANG, Z. X., CHEN, C., CHEN, W. M. & JU, W. T. 2008. Phenotypic and genetic diversity of rhizobia isolated from nodules of the legume genera Astragalus, Lespedeza and Hedysarum in northwestern China. *Microbiol Res,* 163**,** 651-62.

283. WEIR, B. S. 2012. *The current taxonomy of rhizobia* [Online]. Available: http://www.rhizobia.co.nz/taxonomy/rhizobia.html.

284. WHIPPS, J. M. 2001. Microbial interactions and biocontrol in the rhizosphere. *J Exp Bot,* 52**,** 487-511.

285. WHITE, D., DRUMMOND, J. T. & FUQUA, C. 2011. *The Physiology and Biochemistry of Prokaryotes*, Oxford University Press, Incorporated.

286. WIELBO, J., MAREK-KOZACZUK, M., MAZUR, A., KUBIK-KOMAR, A. & SKORUPSKA, A. 2010. Genetic and metabolic divergence within a Rhizobium leguminosarum bv. trifolii population recovered from clover nodules. *Appl Environ Microbiol,* 76**,** 4593-600.

287. WILLIAMS, K. P., GILLESPIE, J. J., SOBRAL, B. W. S., NORDBERG, E. K., SNYDER, E. E., SHALLOM, J. M. & DICKERMAN, A. W. 2010. Phylogeny of Gammaproteobacteria. *Journal of Bacteriology,* 192**,** 2305-2314.

288. WILLIAMS, K. P. & KELLY, D. P. 2013. Proposal for a new Class within the Proteobacteria, the Acidithiobacillia, with the Acidithiobacillales as the type Order. *International Journal of Systematic and Evolutionary Microbiology*.

289. WINZER, K., HARDIE, K. R. & WILLIAMS, P. 2002. Bacterial cell-to-cell communication: sorry, can't talk now - gone to lunch! *Curr Opin Microbiol,* 5**,** 216-22.

290. WISE, A. A., LIU, Z. & BINNS, A. N. 2006. Three Methods for the Introduction of Foreign DNA into Agrobacterium.

291. WISNIEWSKI-DYE, F. & DOWNIE, J. A. 2002. Quorum-sensing in Rhizobium. *Antonie Van Leeuwenhoek,* 81**,** 397-407.

292. WOESE, C. R. 1987. Bacterial evolution. *Microbiological Reviews,* 51**,** 221-271.

293. WOJCIECHOWSKI. F.; MAHN. J.; JONES, B. 2006. *Fabaceae legumes* [Online]. Available: http://tolweb.org/Fabaceae/21093/2006.06.

294. WOLDE-MESKEL, E., TEREFEWORK, Z., LINDSTROM, K. & FROSTEGARD, A. 2004. Metabolic and genomic diversity of rhizobia isolated from field standing native and exotic woody legumes in southern Ethiopia. *Syst Appl Microbiol,* 27**,** 603-11.

295. XU, K. W., PENTTINEN, P., CHEN, Y. X., ZOU, L., ZHOU, T., ZHANG, X., HU, C. & LIU, F. 2013. Polyphasic characterization of rhizobia isolated from Leucaena leucocephala from Panxi, China. *World J Microbiol Biotechnol*.

296. XU, L. M., GE, C., CUI, Z., LI, J. & FAN, H. 1995. Bradyrhizobium liaoningense sp. nov., isolated from the root nodules of soybeans. *Int J Syst Bacteriol,* 45**,** 706-11.

297. Y YAO, P. & VINCENT, J. 1969. Host Specificity In The Root Hair "Curling Factor" of Rhizobium Spp. *Australian Journal of Biological Sciences,* 22**,** 413-424.

298. YAMADA, T., LETUNIC, I., OKUDA, S., KANEHISA, M. & BORK, P. 2011. iPath2.0: interactive pathway explorer. *Nucleic Acids Res,* 39**,** W412-5.

299. YANG, M., SUN, K., ZHOU, L., YANG, R., ZHONG, Z. & ZHU, J. 2009. Functional analysis of three AHL autoinducer synthase genes in Mesorhizobium loti reveals the important role of quorum sensing in symbiotic nodulation. *Can J Microbiol,* 55**,** 210-4.

300. YARZA, P., LUDWIG, W., EUZÉBY, J., AMANN, R., SCHLEIFER, K.-H., GLÖCKNER, F. O. & ROSSELLÓ-MÓRA, R. 2010. Update of the All-Species Living Tree Project based on 16S and 23S rRNA sequence analyses. *Systematic and Applied Microbiology,* 33**,** 291-299.

301. YOKOTA, K. & HAYASHI, M. 2011. Function and evolution of nodulation genes in legumes. *Cellular and Molecular Life Sciences,* 68**,** 1341-1351.

302. YOUNG, J. P., CROSSMAN, L., JOHNSTON, A., THOMSON, N., GHAZOUI, Z., HULL, K., WEXLER, M., CURSON, A., TODD, J., POOLE, P., MAUCHLINE, T., EAST, A., QUAIL, M., CHURCHER, C., ARROWSMITH, C., CHEREVACH, I., CHILLINGWORTH, T., CLARKE, K., CRONIN, A., DAVIS, P., FRASER, A., HANCE, Z., HAUSER, H., JAGELS, K., MOULE, S., MUNGALL, K., NORBERTCZAK, H., RABBINOWITSCH, E., SANDERS, M., SIMMONDS, M., WHITEHEAD, S. & PARKHILL, J. 2006. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biology,* 7**,** R34.

303. ZHENG, H., ZHONG, Z., LAI, X., CHEN, W. X., LI, S. & ZHU, J. 2006. A LuxR/LuxI-type quorum-sensing system in a plant bacterium, Mesorhizobium tianshanense, controls symbiotic nodulation. *J Bacteriol,* 188**,** 1943-9.

304. ZUBAY, G. L. 1993. *Biochemistry*, Wm.C. Brown Publishers.

305. ZUBAY, G. L. 1998. *Biochemistry*, Wm.C. Brown Publishers.