# Microarray analysis of *Rhizobium leguminosarum* bv. *viciae* 3841 colonization of the rhizosphere.

A thesis submitted for the degree of Doctor of Philosophy

By

Vinoy Kumar Ramachandran

Division of Microbiology
School of Biological Sciences
University of Reading
Reading, UK.

September, 2008

**Dedicated to my parents**

# Declaration

I declare that this is my own account of my research and use of all materials from other sources have been properly acknowledged, this work has not been previously submitted for any degree in another university.

Vinoy Kumar Ramachandran

# Abstract

Establishment of a symbiotic interaction between legume hosts and rhizobia is a very complex and a tightly regulated process. Survival and competition for root colonization in the rhizosphere is a key step for a successful symbiosis. Determining the temporal and spatial gene expression of rhizobia in the rhizosphere is an important task for better understanding symbiotic interactions. Here, I describe a method combining an *in vitro* sterile rhizosphere system, RNA amplification and microarray technology to decipher the gene expression of *Rhizobium leguminosarum* bv. *viciae* strain 3841 (hereafter Rlv3841) during colonization of the rhizosphere. Initially experiments were performed to analyze the changes in the transcriptome of free-living Rlv3841 individually supplemented with pea root-exudates and hesperetin. The transcriptome of Rlv3841 in the pea rhizosphere was explored by different approaches (i) different post inoculation time points on a 7d old pea rhizosphere and (ii) on differently aged pea seedling (rhizosphere). The study was further extended to investigate the changes in the transcriptome of Rlv3841 on three different rhizospheres (pea (host legume), alfalfa (non-host legume) and sugar-beet (non-legume)) to identify the host specific genes. Data analysis revealed induction of many genes coding for Nod signaling, quorum sensing, nutrient acquisition, environmental stress related proteins, rhizosphere induced proteins, and a range of transporters and enzymes involved in various metabolic pathways. Integrative analysis of all these datasets sheds light on the genes specifically induced in pea and legume rhizospheres. A set of 16 randomly selected differentially expressed genes was validated by qRT-PCR. The biological importance of 46 of these up-regulated genes was determined by isolation of mutations in them and comparison against wild-type for their ability to colonize the pea rhizosphere. Out of the 46 mutants screened, 28 showed slight competitive defectiveness in colonizing the pea rhizosphere compared to the wild-type, substantiating the role of the upregulated genes in competitive fitness during pea rhizosphere colonization.

# Acknowledgements

I would like to express my gratitude to Prof. Philip Poole for providing me an opportunity to work in his laboratory. I am greatly indebted to him for his stimulating guidance, constructive criticism and encouragement to try new experiments throughout my study. I would like to thank Dr. Kevin Purdy and Dr. Simon Andrews for critical scientific discussions and for being my second supervisors. Furthermore, I would also like to thank Dr. Sheila McIntyre for her support and encouragement, Dr. Mick Watson for help with statistical design of direct microarray experiment design and Prof. Allan Downie (John Innes Centre) for providing with pea root exudates.

I would like to take this opportunity to express my cordial thanks and sincerest respect to Felix Trust Scholarship for the financial assistance, without which none of this would have been possible.

A special thanks to Karunakaran for all the support and encouragement throughout this project. I record my sincere thanks to Alison for critically reading my thesis in detail. I also thank all the past and present members of the group Mary, Alex, Marc, Michelle, Tim, James, Jon, Jurgen and Jay.

Many thanks to the technical assistance of Jane, Diane and all the members of the CSS, and to Alan and Chaz for the IT support. I also thank Jenny for the assistance with rotatory evaporator. I further extend my thanks to Biocentre and all the central facilities of School of Biological Sciences.

Last but not least, I thank my parents and sisters for all the motivation and encouragement to attain this stage in my life. I extend my heartfelt thanks to my friends Ramseek, Mohan, Aditya, Dhami, Vishnu and Venu for their timely help in numerous occasions during this course.

# Publication and Presentation

Karunakaran R, Ebert K, Harvey S, Leonard ME, Ramachandran V and Poole PS. (2006) Thiamine is synthesized by a salvage pathway in *Rhizobium leguminosarum* bv. *viciae* strain 3841. *Journal of Bacteriology*, **188:** 6661-6668.

Ramachandran V, Karunakaran R and Poole PS. (2007)
Transcriptomic analysis of *Rhizobium leguminosarum* in the pea rhizosphere. Poster presentation at the Rhizosphere 2 - International Conference, INRA, Montpellier, France.

# List of Abbreviations

| | |
|---|---|
| ABC | ATP-binding cassette |
| AMA | Acid minimal agar |
| AMS | Acid minimal salts |
| cfu | Colony forming units |
| CUT1 | Carbohydrate Uptake Transporter-1 Family |
| CUT2 | Carbohydrate Uptake Transporter-2 Family |
| Dct | Dicarboxylate transport system |
| dNTP | 2'-deoxynucleoside 5'-triphosphate |
| EDTA | Ethylenediaminetetraacetic acid |
| EPS | Exopolysaccharide |
| FeT | Ferric Iron Uptake Transporter (FeT) Family |
| Fix | Fixation |
| $Fix^+$ | Nitrogen fixing phenotype |
| $Fix^-$ | Non-nitrogen fixing phenotype |
| GFP | Green fluorescent protein |
| Glc | Glucose |
| GUS | β –glucuronidase |
| HAAT | Hydrophobic amino acid transporter family |
| HEPES | N-[2-hydroxyethyl]piperazine-N'-[2-ethanesulphonic acid] |
| IMP | Integral membrane permease |
| IT | Infection thread |
| kan | Kanamycin |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |
| LA | Luria-Bertani agar |
| LB | Luria-Bertani broth |
| LIMMA | Linear Models for Microarray Data |
| LPS | Lipopolysaccharide |
| MDH | Malate dehydrogenase |
| MIAME | Minimum Information About a Microarray Experiment |

| | |
|---|---|
| MOPS | 3-[N-morpholino] propanesulfonic acid |
| N-free | Nitrogen free |
| neo | Neomycin |
| NitT | Nitrate/Nitrite/Cyanate Uptake Transporter Family |
| Nod | Nodulation |
| milliQ | Nanopure water |
| nys | Nystatin |
| OD | Optical density |
| ORF | Open reading frame |
| PAAT | Polar amino acid transporter family |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| PDH | Pyruvate dehydrogenase |
| PEG | Polyethylene glycol |
| PEP | Phosphoenolpyruvate |
| PEPCK | Phosphoenolpyruvate carboxykinase |
| PepT | Peptide/Opine/Nickel Uptake Transporter Family |
| POPT | Polyamine/Opine/Phosphonate Uptake Transporter family |
| PNPG | ρ -nitrophenyl β -D-galacto pyranoside |
| pSym | Symbiotic plasmid |
| QAT | Quaternary Amine Uptake Transporter Family |
| qRT-PCR | Quantitative reverse transcriptase polymerase chain reaction |
| REST | Relative Expression Software Tool |
| Rlv300 | *Rhizobium leguminosarum* bv. *viciae* 300 |
| Rlv3841 | *Rhizobium leguminosarum* bv. *viciae* 3841 |
| RNA | Ribonucleic acid |
| RNase | Ribonuclease |
| SBP | Solute binding protein |
| SEM | Standard error mean |
| SDS | Sodium dodecyl sulphate |
| str | Streptomycin |

| | |
|---|---|
| TE | Tris-EDTA |
| TAE | Tris acetate EDTA |
| TCA | Tricarboxylic acid |
| tet | Tetracycline |
| TIFF | Tagged Information File Format |
| trim | Trimethioprim |
| TRIS | 2-amino-2-(hydroxymethyl)-1,3-propanediol |
| TY | Tryptone-Yeast media |
| UV | Ultraviolet |
| wt | Wild-type |
| X-Gal | 5-bromo-4-chloro-3-indolyl-β -D-galactoside |

# List of Figures

# List of Tables

# Contents

## Chapter 1: Literature review.

## Chapter 2: Materials and Methods.

## Chapter 3: Transcriptomic analysis of *Rhizobium leguminosarum* 3841 in the pea rhizosphere.

## Chapter 4: Comparative transcriptomics of *Rhizobium leguminosarum* 3841 in three different plant rhizospheres.

# Chapter 1: Literature review

### 1.1.1 Introduction

Bacterial survival, adaptation and signalling with the host environment are the fundamental steps in establishing a successful symbiotic, pathogenic or associative interactions (Puhler *et al*., 2004). Though the contributions of bacteria in these interactions are well studied, the mechanism behind the progressive interactions remains unclear. Of the many mutalistic interactions studied, the *Rhizobium*-legume symbiosis is a particularly important example, contributing about 65% of the total available nitrogen to the biosphere, which is the largest single input into the nitrogen cycle (Batut *et al.*, 2004; Newton, 2000).

The *Rhizobium*-legume symbiosis is a complex process which requires successful completion of many steps. During the growth of rhizobia in the plant rhizosphere, rhizobia are chemotactially attracted to flavonoids and other organic acids released by the plants through root exudates. They respond to these exudates by inducing *nod* genes. The *nod* genes encode approximately 25 proteins required for bacterial synthesis and export of Nod factor. Nod factors are lipochitooligosaccharides, which can stimulate root cell division. Rhizobia attach to the developing root hairs, and penetrate the growing root hairs via infection threads. The invading rhizobia are engulfed into the plant, where they are surrounded by a peribacteroid membrane, and there the bacteria (rhizobia) differentiate into bacteroids that fix nitrogen (Gage, 2004).

The rhizosphere is the soil compartment influenced by living roots (Hiltner, 1904), which serves as a platform for most of the molecular events before nodulation (Walker *et al.*, 2003). Rhizosphere colonization, the first phase of the symbiotic interaction is an essential and a multidimensional task influenced by many factors including nutrient scavenging, response to flavonoids, signal exchange, chemotaxis and competition with other microbes before invasion of roots to form nodules (Gage, 2004). Despite being an important step in establishing a successful symbiotic interaction, rhizosphere colonization is a fairly unexplored area compared to nodulation and bacteroid metabolism. As

rhizosphere colonization is a multidimensional task, a genome level understanding of gene expression of rhizobia in the rhizosphere will help our understanding of the principles behind these symbiotic interactions, and will strengthen our knowledge of the ecological significance of rhizobia in the rhizosphere.

### 1.1.2 Rhizobial taxonomy and genetic structure

Bacteria of genus *Rhizobium* play a very important role in agriculture by inducing nitrogen-fixing nodules on the roots of legumes such as peas, beans, clover and alfalfa. The members of the family are generally called rhizobia; it is a large and diverse family of five principal genera including *Azorhizobium, Bradyrhizobium*, *Mesorhizobium*, *Rhizobium* and *Sinorhizobium* (Table 1.1.1.). The original taxonomic classification of rhizobia was based on phylogenetic observations like growth, phenotype and host range (Long, 1989). The discovery of the root nodule forming methylotrophic bacterium *Methylobacterium nodulans* and the ability of the bacteria like *Burkholderia* and *Ralstonia* belonging to the β-proteobacterial to nodulate legumes (Chen *et al.*, 2003; Moulin *et al.*, 2001) often causes confusion in the taxonomic classification. These new species may be the result of lateral gene transfer of symbiotic genes. However, modern techniques such as 16S ribosomal DNA/RNA analysis and DNA restriction fragment length polymorphisms (RFLP) has refined the taxonomic classification (Graham *et al.*, 1991) (Figure 1.1.1.).

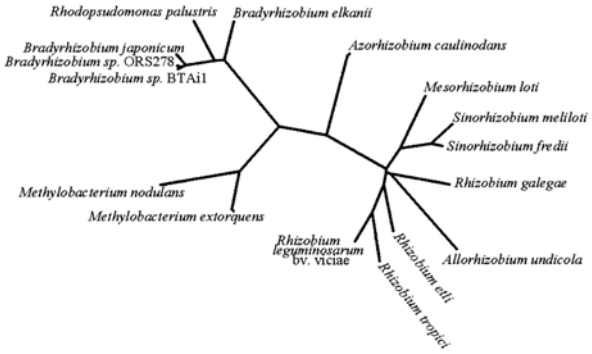



**Figure 1.1.1.** Unrooted phylogenetic tree of 16S rDNA sequences from selected rhizobial strains in the α-proteobacteria. Sequences were aligned using ClustalW and the tree was constructed by using the neighbor-joining method (MacLean *et al*., 2007).

| Species | Genera | Biovar | Host |
|---|---|---|---|
| *Azorhizobium* | *A. caulinodans* | | *Sesbania rostrata* |
| | *A. doebereinerae* | | *Sesbania virgata* |
| *Bradyrhizobium* | *B. elkanii* | | *Glycine max* |
| | *B. japonicum* | | *Glycine max* |
| | *B. liaoningense* | | *Glycine max* |
| | *B. yuanmingense* | | *Lespedeza, Medicago, Melilotus* |
| | *B. canariense* | | *Genisteae* |
| *Mesorhizobium* | *M. amorphae* | | *Amorpha fruticosa* |
| | *M. chacoense* | | *Prosopis alba* |
| | *M. ciceri* | | *Cicer arietinum* |
| | *M. huakuii* | | *Astragalus sinicus, Acacia* |
| | *M. loti* | | *Lotus corniculatus* |
| | *M. mediterraneum* | | *Cicer arietinum* |
| | *M. plurifarium* | | *Acacia senegal, Prosopis juriflora, Leucaena* |
| | *M. septentrionale* | | *Astragalus adsurgens* |
| | *M. temperatum* | | *Astragalus adsurgens* |
| | *M. tianshanense* | | *Glycyrrhiza pallidflora, Swansonia, Glycine, Caragana, Sophora* |
| | *M. albiziae* | | *Albizia kalkora* |
| *Rhizobium* | *R. etli* | | *Phaseolus vulgaris, Mimosa affinis* |
| | *R. galegae* | | *Galega orientalis, G.officinalis* |
| | *R. gallicum* | | *Phaseolus vulgaris, Leucaena, Onobrychis Macroptilium,* |
| | *R. giardini* | | *Phaseolus vulgaris, Leucaena, Macroptilium* |
| | *R. hainanense* | | *Desmodium sinuatum, Stylosanthes, Vigna, Arachis, Centrosema* |
| | *R. huautlense* | | *Sesbania herbacea* |
| | *R. indigoferae* | | *Indigofera* |
| | *R. leguminosarum* | *bv. trifolii* | *Trifolium* |
| | | *bv. viciae* | *Lathyrus, Lens, Pisum,* |

| | | | |
|---|---|---|---|
| | | | *and Vicia* |
| | | *bv. phaseoli* | *Phaseolus vulgaris* |
| | *R. mongolense* | | *Medicago ruthenica, Phaseolus vulgaris* |
| | *R. sullae* | | *Hedysarum coronarium* |
| | *R. tropici* | | *Phaseolus vulgaris, Dalea, Leucaena, Macroptilium,Onobrychis* |
| | *R. yanglingense* | | *Amphicarpaea trisperma, Coronilla varia, Gueldenstaedtia multiflora* |
| | *R. cellulosilyticum* | | *Unknown* |
| | *R. daejeonense* | | *Unknown* |
| | *R. loessense* | | *Astragalus, Lespedeza* |
| | *R. lusitanum* | | *Unknown* |
| | *R. multihospitium* | | *Unknown* |
| | *R. undicola* | | *Unknown* |
| *Sinorhizobium* | *S. abri* | | *Abrus precatorius* |
| | *S. adhaerens* | | *Unknown* |
| | *S. americanus* | | *Acacia spp.* |
| | *S. arboris* | | *Acacia senegal, Prosopis chilensis* |
| | *S. fredii* | | *Glycine max* |
| | *S. indiaense* | | *Sesbania rostrata* |
| | *S. kostiense* | | *Acacia senegal, Prosopis chilensis* |
| | *S. kummerowiae* | | *Kummerowia stipulacea* |
| | *S. medicae* | | *Medicago truncatula, M. polymorpha, M.orbicularis* |
| | *S. meliloti* | | *Medicago, Melilotus, Trigonella* |
| | *S. morelense* | | *Leucaena leucocephala* |
| | *S. sahelense* | | *Acacia, Sesbania* |
| | *S. terangae* | *bv. acaciae* | *Acacia* |
| | | *bv. sesbaniae* | *Sesbania Glycine max* |
| | *S. xinjiangense* | | |

**Table 1.1.1.** Major genera and species of rhizobia and their hosts (adapted from (Weir, 2006; White, 2006; Zakhia & de Lajudie, 2001).

In this genomic era, currently more than 650 bacterial genomes have been sequenced and a large number are being sequenced. The genomes of closely related species are of high interest as comparative genomic approaches help in characterizing and better understanding of a group of bacteria. Among the genera of rhizobia, seven species have been completely sequenced to date; *Mesorhizobium loti* (Kaneko *et al.*, 2000), *Sinorhizobium meliloti* (Barnett *et al.*, 2001; Capela *et al.*, 2001; Finan *et al.*, 2001; Galibert *et al.*, 2001), *Bradyrhizobium japonicum* (Kaneko *et al.*, 2002), *Rhizobium leguminosarum* bv. *viciae* (Young *et al.*, 2006), *Rhizobium etli* (González *et al.*, 2006) and two photosynthetic *Bradyrhizobium* strains (Giraud *et al.*, 2007). Comparative genomic analyses of these species have provided insights about the genomic evolution, structure and the biology behind the symbiotic interactions. The genome architecture of these seven species varies with each other in genome size, structure and complexity (Table 1.1.2.).

| Species | Genomic Architecture (Mb) | Total size (Mb) | No. of ORFs |
|---|---|---|---|
| *B. japonicum* | Chr (9.2) | 9.2 | 8,317 |
| *Bradyrhizobium sp. strain BTAi1* | Chr (8.26), pBTAi1(0.23) | 8.49 | 7,729 |
| *Bradyrhizobium sp. strain ORS278* | Chr (7.5) | 7.5 | 6,752 |
| *M. loti* MAFF | Chr (7.04), pMLa (0.35), pMLb (0.21) | 7.59 | 7,281 |
| *S. meliloti* 1021 | Chr (3.66), pSymB (1.35), pSymA (1.68) | 6.69 | 6,204 |
| *R. leguminosarum* bv. *viciae* 3841 | Chr (5.05), pRL12 (0.87), pRL11 (0.68), pRL10 (0.49), pRL9 (0.36), pRL8 (0.15), pRL7 (0.15) | 7.75 | 7,263 |
| *R. etli* CFN42 | Chr (4.39), p42f (0.65), p42e (0.5), p42d (0.37), p42c (0.25), p42b (0.18), p42a (0.19) | 6.53 | 6,034 |

**Table 1.1.2.** Architecture of rhizobial genomes (adapted from (MacLean *et al.*, 2007)). Chr- chromosome.

Rhizobial genomes consist of single large circular chromosome, and in some species extra chromosomal DNA (plasmids) are present. The size of the

genome and the number of plasmids vary with species, *R. leguminosarum* and *R. etli* have six plasmids, *M. loti* and *S. meliloti* have two plasmids, whereas *Bradyrhizobium* species generally lack plasmids. The reason for the varying size of the genome and the presence of different numbers of plasmids among the rhizobial species is not very clear, but it is proposed that it may be influenced by environmental factors that allow rhizobia to adapt to a flexible life style in complex environments (González *et al.*, 2006). The genes involved in symbiosis are mostly present in the plasmids or in the symbiosis island, which favours transmissibility of these genes to non-symbiotic bacteria by horizontal gene transfer. This may be one of the important reasons to identify novel non-rhizobial nitrogen fixing bacteria, like *Methylobacterium nodulans*. Comparative genomic analysis of the rhizobia shows that they are rich in transporters, regulatory and stress-related systems, which are essential genetic determinants for surviving in the complex environment (Boussau *et al.*, 2004). The genome of the *Bradyrhizobium* sp. strains BTAi1 and ORS278 (Giraud *et al.*, 2007) is rich in genes encoding enzymes involved in metabolic pathway, which enable them to grow as a heterotrophs, autotrophs, prototrophs and also form symbiotic interaction with the plant genus *Aeschynomene* (MacLean *et al*., 2007). Also, another important observation from comparative genomic analysis is that the phylogenetically diverse bacteria show similarities in metabolism, physiology and in the principles behind interactions with the host (Tsolis, 2002). Genomic analysis of *R. leguminosarum* sequence shows more common genes with *S. meliloti* and *M. loti* than with closely related *A. tumefaciens* (Young *et al.*, 2006). In addition to these species discussed, *Agrobacterium vitis S4*, *Agrobacterium radiobacter*, *Rhizobium tropici* and *Sinorhizobium medicae* are being sequenced.

Comparative genomic analysis of all the rhizobial genomes is expected to provide a more elaborate view about the genomic architecture, patterns and evolution. *R. leguminosarum* bv. *viciae* and *R. etli* are very closely related species (Fig 1.1.1), that form symbiotic interactions with specific host plants. The genomic architecture of both the species is quite similar, comprised of a circular chromosome and six large accessory (plasmid) components (González *et al.*,

2006; Young *et al.*, 2006). The comparative analysis also showed that the gene content and gene order is highly conserved in the chromosome, but not in the plasmids, showing that the species have a common evolutionary origin and that the plasmids were acquired during the course of evolution (Crossman *et al.*, 2008; MacLean *et al.*, 2007). The completion of *R. tropici* genome sequencing, a closely related species to *R. etli* and *R. leguminosarum*, is expected to provide more insights about the genomic architecture of *Rhizobium*.

### 1.1.3 Genome of *Rhizobium leguminosarum* bv. *viciae* strain 3841

The work carried out in this thesis is on *Rhizobium leguminosarum* bv. *viciae* strain 3841 (Rlv3841), a spontaneous streptomycin resistant mutant of strain 300. The genome of Rlv3841 was sequenced recently and has a total size of 7.75 Mb with average GC content of 60.86% and a prediction of 7263 protein-encoding genes. The genome consists of one circular chromosome (5.0 Mb), which accounts for 65% of the total protein-encoding genes and six large plasmids (pRL12 (0.87 Mb), pRL11 (0.68 Mb), pRL10 (0.49 Mb), pRL9 (0.35 Mb), pRL8 (0.15 Mb), pRL7 (0.15 Mb)). The plasmid pRL10 is the symbiotic plasmid (pSym) and pRL8 and pRL7 are transferable by conjugation (Young *et al.*, 2006).

Riley functional classification of all the protein-encoding genes, showed a high similarity in distribution of functional classes in the chromosome and all the plasmids, except for pRL7 (Figure 1.1.2) (Young *et al.*, 2006).

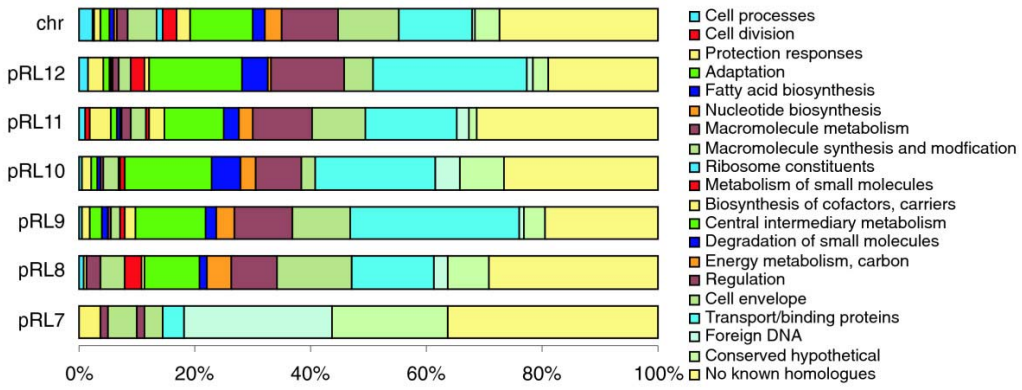



**Figure 1.1.2.** Distribution of functional classes of genes within the replicons. The classes are based on those presented by Riley (Young *et al*., 2006).

The genes coding for core functions are rich in the chromosome compared with the plasmids. The Rlv3841 genome has high proportion of novel or uncharacterized genes in both the chromosome and all the plasmids. The genomes of Rlv3841 and other six rhizobial α-proteobacterial genomes sequenced (Fig 1.1.3), share 648 genes in common.

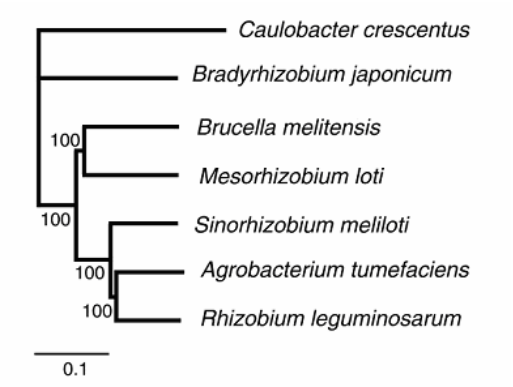



**Figure 1.1.3.** Phylogeny of completely sequenced genomes of selected α-proteobacteria. The phylogeny is based on the concatenated sequences of 648 orthologous proteins. Neighbor-Joining method with % bootstrap support indicated (Young *et al.*, 2006).

Further quartop (quartet of orthologus proteins) analysis, with the three closely related species to Rlv3841 identified 2056 (27% of the total protein coding genes) common genes, with higher GC content than the average and predominantly present on the chromosome. Surprisingly, 546 genes were found to be absent in closely related *A. tumefaceins*, but present in the other three species. Of these 546 genes, 264 had homologs in *Bradyrhizobium*, a distant relative of Rlv3841, which consists mostly of known symbiotic related genes and the remaining genes are hypothesized to have unrecognized roles in symbiosis (Young *et al*., 2006).

Rlv3841 is rich in ABC transporters with 183 complete ABC operons, distributed throughout the genome, but abundant on pRL12, pRL10 and especially on pRL9. In addition, Rlv3841 has a large proportion of genes encoding enzymes involved in general metabolic pathways. The central metabolism is characterized by a complete set of genes for tri-carboxylic acid cycle and glyoxylate pathway in the chromosome. The sugar catabolism in most free-living rhizobia is believed to

proceed via the Entner-Doudoroff or pentose phosphate pathways. Three genes forming an operon encoding enzymes for Entner-Doudoroff pathway and two genes encoding enzymes involved in the oxidative branch of pentose phosphate pathway are present in the chromosome. In addition, a cluster of 13 genes involved in nodulation (*nod*) are present on pRL10, along with the rhizosphere induced genes (*rhiABCR*) (Cubo *et al.*, 1992). The cluster of genes involved in nitrogen fixation, *nifHDKEN*, *nifAB*, *fixABC* and *fixNOQPGHIS* were present around the nodulation genes, making pRL10 a symbiotic plasmid (Young *et al.*, 2006).

Biological nitrogen fixation accounts for up to 65% of the biosphere's total available nitrogen, the majority of which is contributed by symbiotic interaction of bacteria of the genus *Rhizobium* with leguminous plants (Gage, 2004; Newton, 2000). *Rhizobium*-legume symbiosis is a complex process which requires successful completion of many steps including root colonization, signal exchange with plants, nodule formation, differentiation a free-living bacteria into bacteroid and nitrogen fixation (Figure 1.2.1) (Gage, 2004).

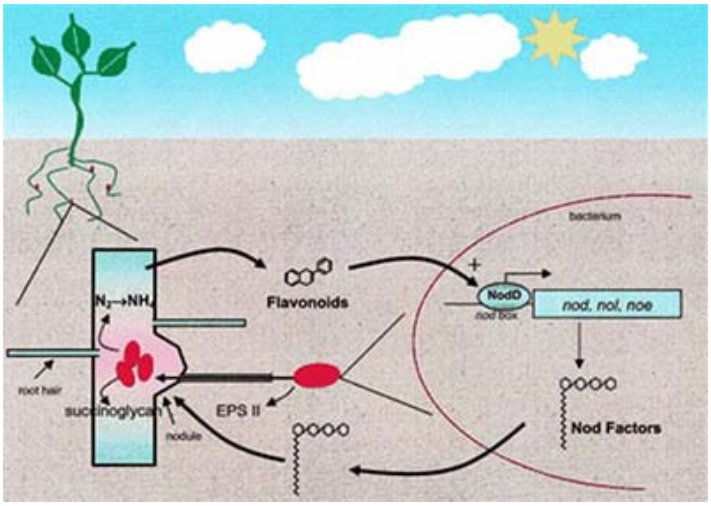



**Figure 1.2.1.** Overview of *Rhizobium*-legume symbiosis (adapted from (Gonzalez & Marketon, 2003)).

### 1.2.1 Nod factor synthesis and perception

The root exudates of leguminous plants contains compounds such as flavonoids and betaines, produced via the isopropanoid pathway, which induces *nod* genes to synthesize Nod factors (Broughton *et al.*, 2000; Perret *et al.*, 2000). The *nod* genes are the host specificity determining factor, as specific Nod factors or lipo-chitooligosaccharides (LCO) will be synthesized by the *nod* genes, when they are induced by flavonoids. In the genome, *nod* genes are found either in chromosome or in symbiotic plasmid. In *R. leguminosarum*, 13 *nod* genes are organized in five operons in the symbiotic plasmid pRL10; *nodABCIJ*, *nodD*, *nodEFL*, *nodNMT* and *nodO* (Downie & Surin, 1990; van Rhijn & Vanderleyden,

1995; Young *et al.*, 2006). NodD, a LysR transcriptional regulator positively regulates the transcription of approximately 25 *nod* genes required for the production and export of Nod factors, in response to flavonoids (Gage, 2004; Schell, 1993).

Nod factors are lipo-chitooligosaccharides, consisting of a backbone of 3-6 beta-1,4 linked N-acetyl-D-glucosamine residues with an n-acyl chain attached to the C-2 position of non-reducing terminal glucosamine (Lerouge *et al*., 1990). Comparison of Nod factors from different rhizobia showed the differences in the structure of the acyl chain and the substitutions in the reducing and non-reducing terminal glucosamine residues. In general, they have a conserved backbone (R3) with the addition of sulphuryl, methyl, carbomyl, acetyl, fucosyl and arabinosyl groups at various position, which determines the host specificity (Denarie *et al.*, 1996; Long, 1996). The *nodABC* genes synthesize the backbone of the Nod factor, NodA (acyl-transferase), NodB (deacetylase) and NodC (N-acetylglucosamine transferase). The *nodABC* genes are conserved through all rhizobium species, as is *nodD*, a regulatory gene for other *nod* genes (Roche *et al.*, 1996; Rohrig *et al.*, 1994) and mutants can be complemented by the homologous genes from any other species. The *nodABC* mutants lacked the production of Nod factors and show a strict $Nod^-$ phenotype on all host plants while *nodEFG* mutants show delayed nodulation (Spaink *et al.*, 1991).

The Nod factor perception occurs in a two step lock and key mechanism. Initial Nod factor recognition and subsequent root curling is a less stringent process, but the initiation of infection threads and activation of plant genes are highly stringent (Ardourel *et al.*, 1994; Limpens *et al.*, 2003; Radutoiu *et al.*, 2003) (Figure 1.2.2). After rhizobia chemotactically move towards the roots and attach to the plant root via adhesins. The adhesins and cellulose fibrils enable rhizobia to attach to the plant root hair (Smit *et al*., 1992). The plant root senses the Nod factors released by the attached rhizobia, and then plant nodulation genes are activated, which make the root hair swell and deform. Subsequently, the root hairs curl and trap the attached rhizobia within the pocket of the curl (Emons & Mulder, 2000; Limpens & Bisseling, 2003). After that, rhizobia within the pocket of the

curl begin to engulf in a cell wall sheath, which subsequently become the "infection thread" (IT). Cell wall sheath is produced by the plant cells in response to the attached rhizobia (Callaham & Torrey, 1981).

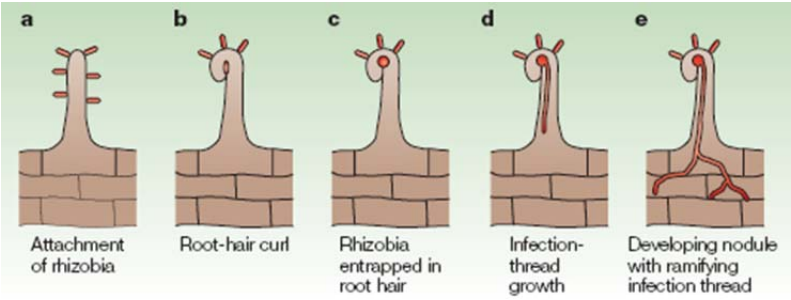



**Figure 1.2.2.** Root hair and nodule infection by rhizobia (Batut *et al.*, 2004)
a) Bacteria attaching to a growing root hair. b) Induced curling of the root hair. c) A few bacterial cells are trapped in the centre of the curled root hair. d) Infection thread (IT) formation. e) IT growth and ramification.

The infection thread grows, and the rhizobia start to divide within them and move towards the nodule, penetrating the plant cells and releasing the rhizobia into the plant cytoplasm. In the plant cytoplasm, the rhizobia are engulfed in a plant-derived membrane, peribacteroid membrane (PBM), which segregates the bacteroids from plant cytosol, with the intervening peribacteroid space (PBS) (Fig 1.2.3.) (Robertson *et al.*, 1978; Whitehead & Day, 1997). Here, the bacteria (rhizobia) differentiate into bacteroids and fix atmospheric nitrogen (Oke & Long, 1999).

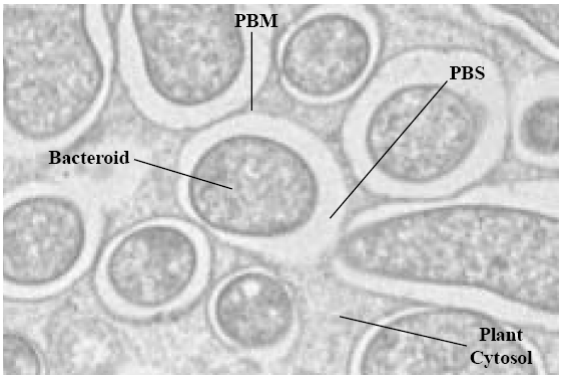



**Figure 1.2.3.** Symbiosome of a pea nodule infected with *R. leguminosarum* bv. *viciae* 3841 (adapted from (Lodwig & Poole, 2003)).

### 1.2.2 Bacteroid differentiation and nitrogen fixation

Bacteroid development was first reported by Beijernick in 1888, when he observed the transformation of rod-shaped cells to Y-shaped cells. The transition of rhizobial cells from free-living cells into nitrogen fixing bacteroids involves an alteration in its developmental pathway. Legumes are capable of forming two different types of nodule, determinate and indeterminate (Oke & Long, 1999). Determinate nodules are characterized by a single bacteroid in a peribacteroid membrane like *Glycine max*, *Phaseolus vulgaris* and *Lotus japonicus* and indeterminate nodules are characterized by many bacteroids in a peribacteroid membrane like *Medicago sativa*, *Medicago truncatula*, *Vicia faba* and *Pisum sativum*. The pattern of nodule development is defined by the host with the influence of plant on bacterial cell division. Mostly genes related to symbiosis and physiological adaptation were identified in differentiating bacteria. Some of the bacterial genes responsible for nutritional uptake were also identified. For example, gene *dctA* codes for the dicarboxylic acids transporter, mutants in this gene results in a fix$^-$ phenotype (Engelke *et al.*, 1989; Yarosh *et al.*, 1989). Also, the disruption of *glmS* encoding for glucosamine synthase leads to early senescence in nodules. Another important gene *bacA*, is essential for bacteroid differentiation, disruption of *bacA* blocks bacteroid differentiation. The exact function of *bacA* is not clear, but it is thought to have two possible functions, (1) BacA is interchangeable with SbmA of *E.coli*, which is involved in transport of peptide antibiotics (Ichige & Walker, 1997), (2) BacA could be involved in cell envelope integrity as *bacA* mutants have altered sensitivity to various chemicals including detergents (Ferguson *et al.*, 2002).

Nitrogen fixation inside the bacteroid occurs via a nitrogenase enzyme complex. There are two main groups of genes in rhizobia that codes for fixation, *nif* and *fix* genes. Nif proteins have homologues in other free-living bacterial nitrogen systems but *fix* genes are specific to rhizobia (Earl *et al.*, 1987; Ruvkun *et al.*, 1980; Ruvkun *et al.*, 1982). The induction of *fix* genes and microaerobic respiration is an important factor during bacteroid differentiation, which must occur before nitrogenase biosynthesis. In *R. leguminosarum*, the *nod*, *nif* and *fix*

genes are all organized in the symbiotic plasmid pRL10 (Young *et al.*, 2006). Nitrogen fixing organisms use the enzyme nitrogenase to catalyse the reductive breakage of strong triple bond between $N_2$ molecules to yield $NH_3$, in an energetically expensive reaction.

$$8H^+ + N_2 + 8e^- + 16ATP \longrightarrow 2NH_3 + H_2 + 16ADP + 16Pi$$

The *nif* genes encode for nitrogenase, a large enzyme composed of two structurally and mechanistically conserved metalloenzymes, dinitrogenase reductase (Fe protein) and dinitrogenase (Mo-Fe protein) (Halbleib & Ludden, 2000). Dinitrogenase reductase is a dimeric iron containing protein (60 kDa) encoded by *nifH*. It acts as an obligate electron donor to dinitrogenase and contains a single 4Fe-4S cluster at the subunit interface as well as two Mg-ATP binding sites, one on each subunit (Georgiadis *et al.*, 1992; Ruvkun *et al.*, 1982; Schindelin *et al.*, 1997). Dinitrogenase is a molybdenum-iron (MoFe), containing protein (220-240 kDa) which forms a $\alpha_2\beta_2$ tetramer of two pairs of two metaloclusters (Chan *et al.*, 1993; Kim & Rees, 1992). Each αβ pair contains one MoFe cofactor, which contains the site of substrate reduction and one P-cluster (Einsle *et al.*, 2002). The *nifD* encodes the α-subunit, which contains the MoFe cofactor, whereas *nifK* encodes the β-subunit and associates with the α-subunit, with a P-cluster present at the αβ interface (Ruvkun *et al.*, 1982; Schindelin *et al.*, 1997). The oxygen labile nature of nitrogenase is due to this surface exposed 4Fe-4S cluster on the dinitrogenase reductase as the enzyme mechanism requires electron transfer from here to the P-cluster of the dinitrogenase and subsequently to the FeMo cofactor, the site of substrate reduction (Dixon & Kahn, 2004).

To allow the nitrogen fixation to occur, a low redox potential reductant molecule such as flavodoxin or ferrodoxin donates electrons to dinitrogen reductase, enabling it to react with Mg $ATP^{2+}$ (Chan *et al.*, 2000). The dinitrogen molecule, binds to the molybdenum ferrous cofactor on the dinitrogenase component of the enzyme and forms a nitrogenase enzyme complex (Chan *et al.*, 2000; Rees, 2000). Electrons from the dinitrogen reductase flow one by one to

dinitrogenase with the concomitant hydrolysis of two ATP molecules, reducing the nitrogen gas and protons.

Though this reaction seems to be simple, it is a tightly regulated process. Rhizobia need more energy in the form of ATP to break the trivalent bond of the inert dinitrogen, for which it needs to respire more by consuming more oxygen. However, the enzyme nitrogenase is an oxygen labile enzyme, which will be denatured in the presence of oxygen. To overcome this delicate situation, symbiotic partners have evolved several methods: 1) an oxygen diffusion barrier in the nodule cortex lowers the absolute oxygen concentration. 2) Synthesis of oxygen binding protein leghaemoglobin to bind the oxygen, which then facilitates the movement of oxygen to the bacteroids for respiration. 3) Plants redirect glycolysis towards malate with subsequent reductive formation of succinate under micro-aerobic conditions. 4) Bacteroids use C4-dicarboxylates rather that mono or disaccharides. 5) Bacteroid ATP production is coupled to a high affinity terminal oxidase (Batut & Boistard, 1994; Day & Copeland, 1991; Hennecke, 1993; Udvardi & Day, 1997).

### 1.2.3 Nutrient exchange and biochemistry of symbiosome

Rhizobia convert the gaseous nitrogen to ammonia and provide this to the plant while, in return the plant supplies energy and a carbon source in the form of dicarboxylates, to fuel the nitrogen fixation process. Bacteroids lack a full glycolytic pathway, indicating that part of their metabolism has been shut during symbiosis (Copeland *et al.*, 1989). Bacteroids have high affinity C4-carboxylate uptake systems (McRae *et al.*, 1989), which strongly suggest the presence of a fully functional tricarboxylic acid (TCA) cycle. The carbon source supplied to the bacteroid to fuel nitrogen fixation is derived from the photosynthate transported to the nodule via the phloem in the form of sucrose (Gordon *et al.*, 1985; Kouchi *et al.*, 1986; Streeter, 1981), which is then converted to glucose and subsequently to phosphoenolpyruvate (PEP) by glycolysis in the nodule. PEP is metabolized to malate by phosphoenolpyruvate carboxylase and malate dehydrogenase (Day & Copeland, 1991; Rosendahl *et al.*, 1990). Dicarboxylates such as malate are

abundant inside nodules and they are transported across the bacteroid membrane (DeVries *et al.*, 1980; Rosendahl *et al.*, 1990). The transport of dicarboxylates is carried out by the dicarboxylate transport system (Dct). The Dct system codes for three proteins; DctA, which codes for the transport protein, and DctB and DctD which activate transcription of *dctA* in the presence dicarboxylates (Glenn *et al.*, 1980). The $C_4$ dicarboxylate transport system was shown to be important for both free-living and bacteroid *R. leguminosarum* bv *trifolli*, and disruption of the Dct system produces ineffective bacteroids (Ronson *et al*., 1981). Dct mutants of *R. leguminosarum* and *S. meliloti* are unable to transport of L-malate, fumarate and succinate, and form ineffective nodules (Engelke *et al.*, 1987; Finan *et al.*, 1981; Finan *et al.*, 1983; Glenn *et al.*, 1980). The Dct system can also transport aspartate at low rates, but it is preferably transported via the Aap & Bra system (Walshaw & Poole, 1996). Inside the bacteroid, the dicarboxylates enter the TCA cycle. The TCA cycle in the bacteroid is tightly regulated to ensure optimal functionality, especially within the low oxygen environment of the bacteroid (Lodwig & Poole, 2003). Malic enzyme (ME) regulates the TCA cycle, by maintaining the availability of acetyl-coA for citrate synthase. Malic enzyme, decarboxylates the malate to form pyruvate, which is then decarboxylated to acetyl-CoA by pyruvate dehydrogenase. Malic enzyme exists as two forms, DME ($NAD^+$ Malic enzyme) and TME ($NADP^+$ Malic enzyme). These two enzymes ensures the correct redox balance of the TCA cycle (Lodwig & Poole, 2003) .

Bacteroids export ammonia and some alanine through the peribacteroid cytoplasm into the plant cytosol. Ammonia movement is thought to occur by diffusion across the membrane. Ammonia is assimilated in plant cytosol by GS-GOGAT (glutamine synthetase-glutamate synthase) forming glutamine and then glutamate (Lodwig & Poole, 2003). Bacteroids cannot assimilate ammonium, they depend on plant derived amino acids for nitrogen, and glutamate is readily converted to aspartate and alanine, when supplied (Rosendahl *et al.*, 1992). Recent research demonstrated the symbiotic and metabolic importance of this nutrient exchange (Lodwig *et al.*, 2003). A double mutant of *R. leguminosarum*, completely blocked in amino acid uptake was unable to provide nitrogen to the

plant. The proposed model shows that amino acid cycling drives nitrogen assimilation by the plant (Fig 1.2.4).

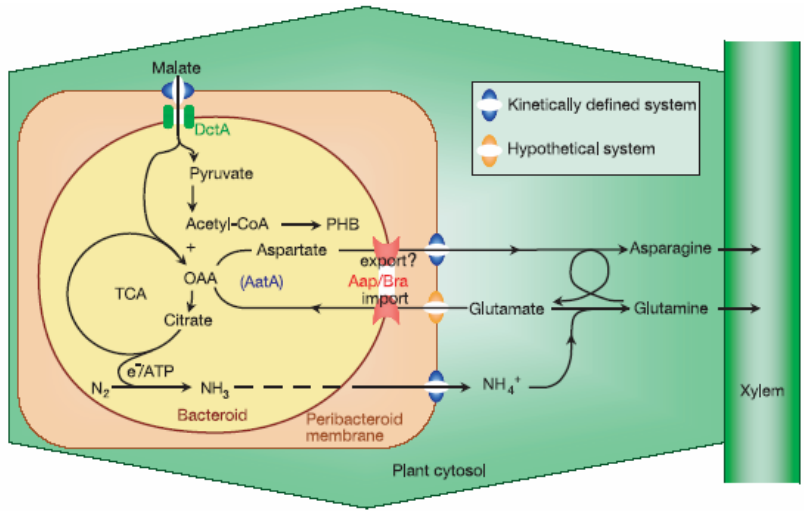



**Figure 1.2.4.** Amino acid cycling in bacteroids of *R. leguminosarum* (adapted from (Lodwig *et al.*, 2003).

Glutamate and dicarboxylates enter the bacteroid. Dicarboxylates drive the TCA cycle, forming oxaloacetate for amino acid cycling and ATP for nitrogen fixation. The glutamate acts as a transamination donor for oxaloacetate to produce aspartate or alanine, which is then secreted to the plant. This aspartate can then be made into asparagine, while glutamate is reconstituted using the ammonia secreted from the bacteroid. This shows that amino acid cycling is important for nitrogen fixation. If the plant does not provide amino acids to the bacteroid, it will not receive ammonium from the bacteroid. This ensures mutualism is dominant for both plant and rhizobia.

### 1.3.1 Overview of Rhizosphere

Lorenz Hiltner (1904) described the rhizosphere as the "soil compartment influenced by the roots". His work emphasized the importance of microorganisms in plant health and nutrition (Hiltner, 1904). The word rhizosphere derived from greek, where "Rhizo" means "roots" and "sphere" means "field of action". After more than 100 years of the extensive research of rhizosphere, the definition of rhizosphere has evolved as "the zone of interaction influenced by living roots, where complex biological and ecological interactions occur between microorganisms and roots" (Hirsch *et al.*, 2003; Lynch, 1987; Lynch, 1990). The area around the rhizosphere can be classified into three zones; 1) endorhizosphere (root tissue including the endodermis and cortical cells), 2) rhizoplane (root surface with the epidermis and mucilaginous layer) and 3) ectorhizopshere (soil immediately adjacent to the roots) (Lynch, 1987; Morgan *et al.*, 2005) (Figure 1.3.2).

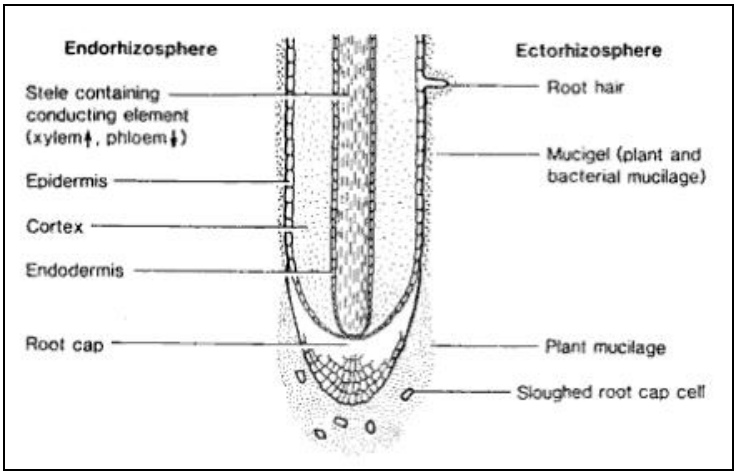



**Figure 1.3.1.** General structure of the rhizosphere (adapted from (Pinton *et al.*, 2001)).

As the influence of the living roots may extend around the roots in the soil, the exact volume of the rhizosphere is difficult to assess. It also varies with respect to plant species, soil type and water content of the soil. In practical terms

the rhizosphere soil is the soil which adheres to the roots, but which can be removed by gentle shaking in water, unlike the bulk soil which does not to adhere to roots (Bolton *et al.*, 1992; Curl & Truelove., 1986).

The rhizosphere is generally characterized as a dynamic environment, rich in diverse carbon and organic compounds, micronutrients and chemoattractants secreted by the plant as root exudates. The composition of root exudate varies with the plant species, making the rhizosphere more specific and favorable to its partner (Sharma *et al.*, 2004). The proportion of the photosynthates released into the roots varies from 30 to 60% depending on the plant species (Morgan *et al.*, 2005). The composition of root exudates also varies with the growth of the plant (Kamilova *et al.*, 2006). The signaling compounds and the chemoattractants form the basis for the early interaction, which is well known as *molecular dialogue* (Denarie & Cullimore, 1993).

The deposition of root exudates makes the rhizospheric soil very different from the bulk soil. Comparison of the community profiles from the rhizosphere and from bulk soil, often show unique fingerprints for the rhizosphere soil from particular plant species (Marschner *et al.*, 2001; Smalla *et al.*, 2001), suggesting the plant-specific rhizosphere effect on the bacterial community. The availability of an array of carbon compounds in the rhizosphere highly favours colonizing bacteria in the host environment. On the other hand, it also invites other microorganisms (*uninvited guest*) as predators or competitors, which subsequently leads to a high density of inter- and intra-species microorganisms in the rhizosphere (Bolton *et al.*, 1992; Curl & Truelove., 1986). Microbial diversity in the rhizosphere is estimated as 10,000 species per gram of soil (Parkinson, 1994), which is approximately 500 times more than the bulk soil. As the rhizosphere is a very general and wide topic, most of the information presented in this review is based on rhizobia, unless otherwise stated.

### 1.3.2 Survival and colonization in the rhizosphere

Rhizosphere, the soil compartment surrounding living roots serves as a platform for most of the molecular events before nodulation (Walker *et al.*, 2003). Rhizosphere colonization is the first and foremost step in establishing a successful symbiotic interaction. In the rhizosphere, rhizobia have to survive acid or saline conditions (Glenn *et al.*, 1999), feed on nutrients, cope with environmental stress, signal exchange and regulate quorum sensing, before entering into the plant roots (Fig 1.3.2.). The different kinds of environmental stress which rhizobia are required to survive to establish a successful symbiosis range from oxidative stress to detoxification of toxic compounds (Munchbach *et al*., 1999).

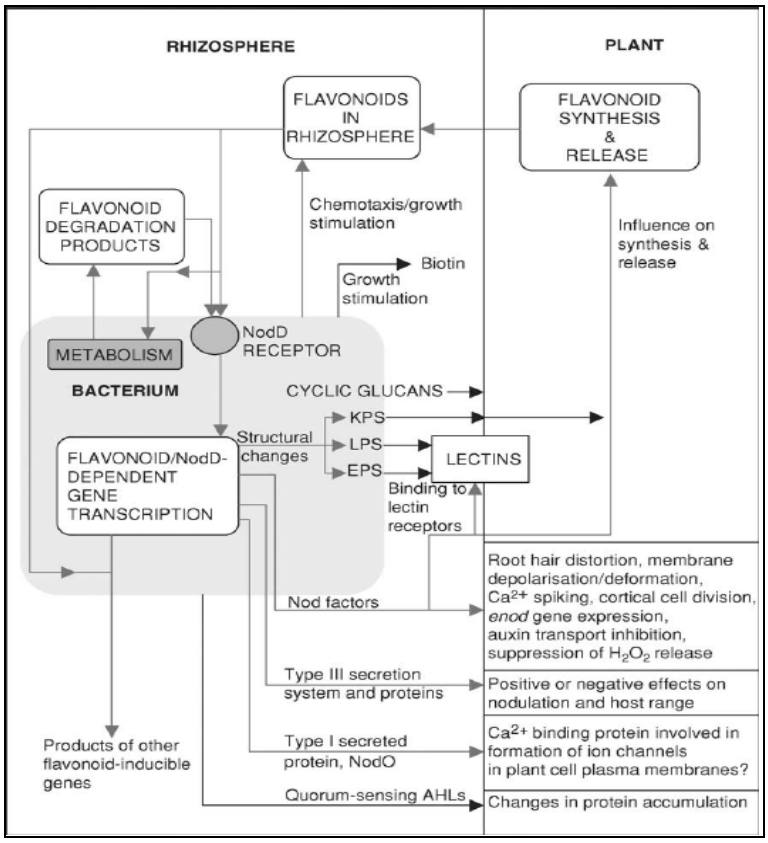



**Figure 1.3.2.** Overview of early signalling events in rhizobia-legume symbiosis. Pathways mediated by flavonoids are shown in grey. AHL, N-acyl homoserine lactones; EPS, extracellular polysaccharides; KPS, capsular polysaccharides; LPS, lipopolysaccharides (Adapted from (Cooper, 2004)).

### 1.3.3 Root exudates

Root exudation into the rhizosphere is one of the extremely important events in nature. Root exudates, makes the rhizosphere a dynamic environment for bacteria. Plant roots continuously synthesize and secrete a diverse array of compounds into the rhizosphere (Bais *et al*., 2006). Root exudates consists of ions, free-oxygen, water, enzymes, chemical substrates, mucilage and diverse array of carbon containing metabolites (Bertin *et al*., 2003). Chemical substrates in the root exudates serve as chemoattractants for microorganisms in the rhizosphere, which initiate the communication and signal transduction between the interacting partners in a symbiotic interaction (Bais *et al.*, 2004). Root exudates contain flavonoids, betaines, aldonic acids, xanthones and other secondary metabolites, which elicit the expression of *nod* genes, one of the important molecular events in symbiosis (Cooper, 2007). The concentration of the flavonoids in the root exudates increases in the presence of compatible rhizobia (symbiotic partner of the plant species) (van Brussel *et al*., 1990). (see section 1.2.2.3 for a detailed review on nodulation)

The components in the root exudates can be broadly classified into two groups, low and high molecular weight ($M_r$) compounds. The low $M_r$ compounds consist of amino acids, organic acids, sugars, phenolics and secondary metabolites, whereas the high $M_r$ compounds consist of mucilage (polysaccharides) and proteins (Bais *et al.*, 2006). Root exudates contains 30% (cereals) to 60% (trees) of the photosynthates, and the proportion varies with the type of soil, nutrient availability, age, physiology and the plant species (Bais *et al.*, 2006; Morgan *et al.*, 2005). Rhizobia are peritrichously flagellated and can react to chemical gradients in the rhizosphere. Plant root exudates contain many substances identified as chemoattractants that promote rhizosphere colonization, which includes carbohydrates, amino acids, carboxylic acids and phenolic compounds (Aguilar *et al.*, 1988; Barbour *et al.*, 1991; Bowra & Dilworth, 1981; Gaworzewska & Carlile, 1982; Gotz *et al.*, 1982; Parke *et al.*, 1985).

The availability of range of diverse nutrients in the rhizosphere attracts inter- and intra-species microorganisms, leading to a highly competitive

environment. The increase in the population of heterotrophic microorganisms will discourage symbiotic bacteria from establishing an interaction, by decreasing the availability of nutrients. Plant species provide a competitive advantage to the interacting bacteria, by imposing selective pressure in the form of sugars or toxic/unusual compounds. The ability to catabolize/degrade the host specific compounds by rhizobia provides competitive fitness during rhizosphere colonization. Comparative growth studies of *R. leguminosarum*, *Burholderia cepacia* and *Pseudomonas fluorescens* on pea root mucilage as sole carbon source, showed that only *R. leguminosarum* can completely utilize the pea root mucilage, as a carbon source (Knee *et al.*, 2001). Another comparative study, on the ability of two nearly isogenic strains to colonize the root of a transgenic plant engineered to produce a novel substrate, which can only be catabolized by one of the strains. When inoculated alone, both the strains colonized both control and transgenic plants equally, whereas when co-inoculated on the transgenic plant the catabolizer strain out-competed the non-catabolizer strain. No differences were observed in the control plants, supporting the concept of host-induced selection of root colonizing bacteria (Savka & Farrand, 1997; Savka *et al.*, 2002).

Quorum sensing is a form of cell-cell communication, mediated by small diffusible molecules called as autoinducers (acyl-homoserine lactone) to co-ordinate the regulation of cell density. Pea (*P. sativum*) root exudates contains several bioactive components that mimic N-acyl-homoserine lactone (AHL) signals, which stimulates AHL-regulated behaviours in some strains, while inhibiting in others (Teplitski *et al.*, 2000; Walker *et al.*, 2003). It is also shown that crude aqueous extracts of plant species exhibit AHL inhibitory activity, which suggest that plants may have a mechanism to interfere with bacteria to regulate cell density in the rhizosphere (Walker *et al.*, 2003). *R. leguminosarum* bv. *viciae* quorum sensing was discovered with the identification of *rhiA*, encoding a 24,000 K Da protein which was highly expressed in stationary phase (Dibb *et al*., 1984). *rhiA* is in the operon (*rhiABC*), located adjacent to genes involved in nodulation and nitrogen fixation on the symbiotic plasmid pRL1JI. The *rhiABC* operon is regulated by RhiR (Economou *et al*., 1989) and found to be strongly induced by

an N-acyl-homoserine lactone (AHL) (Rodelas *et al*., 1999). *rhiA* is expressed in the rhizosphere but not in nitrogen-fixing bacteroids. Its expression can also be negatively regulated by NodD (Cubo *et al*., 1992). Further studies with *rhiA-lacZ* fusion showed a decrease in the expression *rhiA* in the presence of flavonoids. The *rhi* genes are found only in few strains of *R. leguminosarum* bv. *viciae*, but not in other rhizobia, suggesting its role in determining host specificity. The AHL that induce *rhi* genes is $3OH,C_{14:1}$-HSL, a potent growth inhibitor, known as bacteriocin. Bacteriocin was produced by the *cinRI* locus. *cinR* positively regulates *cinI*, and three other AHL-dependent systems; *raiI*/*raiR* for regulation of AHL biosynthesis, *traI*/*traR* for symbiotic plasmid transfer and *rhiI*/*rhiR* for rhizosphere expressed genes (Wisniewski-Dye and Downie, 2002).

### 1.3.4. Metabolism in rhizosphere

Rhizobia have three different phases, starting as free-living rhizobia struggling to survive in the soil (stressful), then in the rhizosphere (nutrient rich and competitive) and finally as bacteroids inside the nodules of the cognate host (comfortable). So, the physiology of the rhizobia differs highly with life styles, which is reflected in the rhizobial genomes. Rhizobial genomes consists of a large proportion (both characterized and putative) of genes coding for transport systems and for enzymes involved in catabolism of carbon compounds, enabling them to lead a metabolically flexible life-style (Young *et al.*, 2006). Rhizobia can grow on a wide variety of sugars, amino acids and phenolic compounds (Stowers, 1985).

Over the past two decades, extensive research has been carried out in characterizing the physiology of rhizobia, mostly focussed on free-living bacteria and bacteroids. Free-living rhizobia use the Entner-Doudoroff (ED) and pentose phosphate pathway (PP) to catabolize sugars, as the Emden-Meyerhoff pathway is absent (Lodwig & Poole, 2003). Recent, metabolic flux ratio (METAFoR) analysis of seven bacterial species including *S. meliloti*, confirmed the use of ED pathway as the major catabolic route to sugar catabolism, and pentose phosphate pathway for biosynthesis (Fuhrer *et al.*, 2005). The use of gluconeogenesis by rhizobia is identified by the characterization of two gluconeogenic enzymes

phophophenolpyruvate carboxykinase and fructose bi-phosphate aldolase, (Lodwig & Poole, 2003). These two enzymes were reported to be repressed by low level of sugars (0.1mM sucrose) in *R. leguminosarum* (McKay *et al.*, 1985), whereas in *S. meliloti* they are induced only by dicarboxylates (Osteras *et al.*, 1995). Rhizobia use tricarboxylic acid (TCA) cycle for oxidation of organic acids and all the enzymes involved in TCA cycle have been detected in rhizobia (McKay *et al.*, 1989). Disruption of the gene encoding the TCA cycle enzyme aconitase in *B. japonicum* resulted in very poor growth under laboratory conditions, and formation of effective nitrogen fixing nodules (Thonymeyer & Kunzler, 1996). A large proportion of genes identified as essential for free-living rhizobia and/or legume bacteroids to be significant were often tested for their biological significance in the rhizosphere by assessing their competitiveness by colonization assay. The ability to utilize or degrade certain sugars, aminoacids, aromatic compounds, organic acids and host-secreted toxic compounds were reported to provide competitive advantage during rhizosphere colonization, which will be discussed in the subsequent section.

### 1.3.5 Rhizosphere competitiveness

Competition between microorganisms determines the outcome of many biological events in nature and yet is a poorly understood process (Bittinger & Handelsman, 2000). Identifying the genetic trait conferring competitive success to bacteria *in vivo* is a cumbersome task, involving a large number of replicated screenings. However, two methods were used to identify the biological significance of a gene in rhizobia; 1) rhizosphere competitiveness and 2) nodule competitiveness. In rhizosphere competition studies, a gene of interest is disrupted (mutant) and challenged against the wild-type (wt) in a sterile *in vitro* host rhizosphere. A rhizosphere colonization index (RCI) is calculated based on the ratio of the percentage of recovered wildtype from the mutant assay over the control assay, from the rhizosphere after a stipulated time period. The biological significance of that particular gene is assessed based on the rhizosphere colonization index. On the other hand, nodule competitiveness is a readily

quantifiable trait by scoring nodule occupancy, which often agrees with rhizobial fitness in symbiosis. Nodule competitiveness is measured by comparing the proportion of rhizobial strains inoculated to the seed or seedlings, with the strains occupying in the nodule.

Legumes root exudates contain a wide variety of carbon sources including non-protein aminoacids, flavonoids and phenolic compounds (Bell, 2003; Lambein *et al.*, 1993). Pea root mucilage analysis showed the presence of a wide variety of sugars, aminoacids and arabinogalactan proteins (Knee *et al.*, 2001). *S. meliloti* α-galactosidase tagged with green fluorescent protein (GFP) was used as biosensors to probe the alfalfa rhizosphere, which revealed the presence of α-galactosides in the alfalfa rhizosphere (Bringhurst *et al.*, 2001). Root exudates also contain many specific host-secreted toxic or unusual compounds, which can only be degraded or catabolized by compatible bacteria. Many compounds were reported to provide a selective competitive advantage to symbiotic bacteria by its plant host. Homoserine, a non-protein aminoacid found in pea rhizosphere can be utilized by *R. leguminosarum* bv. *viciae* as a sole carbon and nitrogen source (Van Egeraat, 1975), whereas other bean strains were often unable to utilize it as a carbon source (Hynes & O'Connell, 1990). Rhizopine (methyl-scyollo-inosamine) is secreted from the nodule of some legume host plants and released into the rhizosphere. The ability to catabolize rhizopines provides a competitive advantage for certain strains of *S. meliloti* and *R. leguminosarum* (Gordon *et al.*, 1996; Heinrich *et al.*, 1999). Mimosine, a toxic compound produced by the tree legume *Leucaena leucocephala*, can only be catabolized by some strains of *Rhizobium spp.,* such as TAL1145, which can nodulate *Leucaena*. *Rhizobium spp.,* TAL1145 can grow on mimosine as sole carbon source, which confers a competitive advantage (Borthakur *et al.*, 2003; Soedarjo & Borthakur, 1996). The ability of some of rhizobia to grow in non-legume rhizosphere has been reported. These strains can catabolize unusual compounds such as calystegines from plants such as *Calystegia sepium*, *Convolvulus arvensis* and *Atropa belladonna* (Tepfer *et al.*, 1988). The ability to catabolize calystegines by *S. meliloti* Rm41, offers a distinct competitive advantage in colonizing calystegine producing plant rhizosphere

(Guntli *et al.*, 1999). *R. leguminosarum* bv. *trifolii* forms a symbiotic interaction with clover plant. *R. leguminosarum* bv. *trifolii* produces trifolitoxin (TFX), an antimicrobial peptide that inhibits members of α-proteobacteria which includes plant symbionts (Robleto *et al.*, 1998). Under laboratory conditions, trifolitoxin increased the nodulation competitiveness in *Rhizobium* and *Sinorhizobium* species (Robleto *et al.*, 1997; Triplett, 1990) and also increases the nodule occupancy under field conditions (Robleto *et al.*, 1998).

Flavonoids are complex aromatic compounds that induce *nod* genes to synthesize Nod factors, which is the basis for communication in symbiosis (see section 1.1.4.2 for Nod factors). Rhizobia posses the ability to degrade flavonoids to monocyclic aromatic compounds such as para-hydroxy benzoate, catechol and protocatechuate (Rao *et al.*, 1991; Rao & Cooper, 1994) and to catabolize succinyl CoA and acetyl CoA, to feed into the tricarboxylic acid cycle via the β-ketoadipate pathway (Harwood & Parales, 1996; Parke & Ornston, 1986; Parke *et al.*, 1991). *Rhizobium loti* NZP2042 and LC22, *Bradyrhizobium sp.* CC814s and CC829 have the ability to degrade the flavonoid quercetin to protocatechuate and phloroglucinol (Rao *et al.*, 1991). *B. japonicum* can degrade catechin to protocatechuate and phloroglucinol carboxylic acid (Hopper & Mahadevan, 1991). The β-ketoadipate pathway is main route for most aromatic compound catabolism, once the complex aromatic compound is degraded to monocyclic aromatic compounds. Some of the flavonoids (e.g. catechin) were reported to be present in the root exudates at toxic concentrations (Bais *et al.*, 2002; Veluri *et al.*, 2004).

Vitamins such as biotin, thiamine, riboflavin and nicotinic acid are essential growth factors for a wide variety of bacteria. However, the precise requirement for these compounds differs between species. A vitamin requirement is one of the important traits which favour rhizosphere colonization of many rhizobia. *S. meliloti*, has a regulatory locus *bioS*, which allow it to respond to biotin signals from the alfalfa plant. Strains with a disruption in biotin biosynthetic genes showed reduced rhizosphere colonization and nodulation competitiveness (Heinz *et al.*, 1999; Streit *et al.*, 1996). In *R. leguminosarum* bv. *viciae*, the *thiMED* operon for thiamine biosynthesis are found to be highly

expressed in the pea rhizosphere (Allaway *et al.*, 2001). Further characterization showed that thiamine is synthesized by a salvage pathway from the precursor compounds found in the pea rhizosphere and mutation in thiamine biosynthetic genes showed a very compromised colonization level when compared to the wild-type in the pea rhizosphere (Karunakaran *et al.*, 2006). In *Mesorhizobium sp* strain R7A, biosynthetic genes for biotin, thiamine and nicotinic acid along with the symbiotic genes are present in the symbiosis island on the chromosome, whereas it is absent in some of the non-symbiotic strains. The transfer of the symbiosis island to the non-symbiotic strains, enable them to biosynthesize the three vitamins (Sullivan et al., 2001). Overall, this observation suggests that the co-evolution of host plant and bacteria is a clear strategy to select compatible rhizobia by providing competitive advantage over *uninvited guests*.

In addition, there are many chemical substrates in the legume root exudates which are reported to confer competitive advantage in rhizosphere colonization and nodule occupancy in wide variety rhizobia, rather than being restricted to selected strains. The ability to catabolize, *myo*-inositol (Fry *et al.*, 2001), rhamnose (Oresnik *et al.*, 1998; Richardson *et al.*, 2004) and erythirtol (Yost *et al.*, 2006) by *R. leguminosarum* and trehalose (Jensen *et al.*, 2005), proline (Jiménez-Zurdo *et al.*, 1995; Jiménez-Zurdo *et al.*, 1997) trigonelline, betaines, carnitine and stachydrine by *S. meliloti* (Boivin *et al.*, 1990; Boivin *et al.*, 1991; Burnet *et al.*, 2000; Goldmann *et al.*, 1991) confers a competitive advantage in colonization and/or in nodulation when isogenic strains with and without the catabolic activity are compared.

Transcriptional regulators also play a major role in the regulation of gene expression in the rhizosphere in response to environmental factors. Mutation in RosR, a transcriptional regulator of *R. etli* CE3 showed a reduced rhizosphere competitiveness and decreased nodule occupancy (Bittinger *et al.*, 1997). Further miniTn*5* based mutagenesis and mass screening of 18,000 mutants, showed that atleast 52 genes were negatively regulated by RosR and a single gene was positively regulated, indicating the influence of RosR on many genes which are

essential for rhizosphere and nodulation competitiveness (Bittinger & Handelsman, 2000).

Rhizobial metabolism varies with respect to nutritional availability, species and phase of life. The physiology of rhizobia during rhizosphere colonization is an unexplored area till date, due to the complex nature of the rhizosphere and lack of experimental techniques. However, with the advances in genome sequencing and microarray technology, a considerable amount of research is focussed on the biology of rhizosphere bacteria, especially on plant growth promoting *Pseudomonas spp.* This research work is focussed on the transcriptome of *R. leguminosarum* bv. *viciae* strain 3841 in the pea rhizosphere, and is expected to provide more insights on the biology of symbiotic bacteria in the rhizosphere.

Rhizosphere colonization is characterized by many complex biological interactions between bacteria and the host plant. From the bacterial point of view, it is a multi-tasking event, where bacteria have to manage stress, nutrient competition and chemotaxis, signal exchange and regulate quorum sensing. However, very little is known about these interactions and the physiology of the bacteria in the rhizosphere. Several strategies have therefore been developed to study the biology of bacteria in the rhizosphere.

Initially, most of the work was focussed on analyzing and assessing the bacterial communities in the rhizosphere. Classical molecular techniques such as denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) were used to study the genetic diversity of bacterial communities in the rhizosphere (Muyzer, 1999). The limitation of these techniques is that they can only detect the presence of microbes, not the metabolically active microbes. Stable isotope probing (SIP), overcame this limitation and enabled to identify the functionally and metabolically active microbes form the rhizosphere (Radajewski *et al.*, 2000). Though SIP is a very promising technique, it is still in developmental stage (Kiely *et al.*, 2006).

To understand more about the biology of bacteria in the rhizosphere, identifying the genes specifically expressed in the rhizosphere and assessing the function of the genes is necessary. Over the past two decades, many techniques have been developed to identify novel genes induced specifically in the rhizosphere which includes reporter gene technology (Hautefort & Hinton, 2000), differential fluorescence induction (DFI) (Allaway *et al.*, 2001), selective capture of transcribed sequences (SCOTS) (Daigle *et al.*, 2002), subtractive hybridization and differential display (Handfield & Levesque, 1999), signature-tagged mutagenesis (STM) (Hensel *et al.*, 1995a), *in vivo* expression technology (IVET) and many flavours of IVET (Mahan *et al.*, 1993; Rediers *et al.*, 2005). Although these techniques were very promising, they are time consuming and permit the identification of only one or a few genes at a time. As rhizosphere colonization is

multi-tasking event, a genome-wide understanding of gene expression will provide more information and help us better understand the complex interactions in the rhizosphere.

The recent advances in genome sequencing and "omic" technologies such as genomics, proteomics and metabolomics, offer a new opportunity for the microbial ecologist to understand the molecular basis underpinning the interactions in the rhizosphere. The combination of molecular-microbiological techniques and the "omics" technology will prove to be a powerful tool to explore rhizosphere bacterial genomes. However, each of these strategies from molecular-microbiological to -omic technology has its own strengths and weaknesses. So, an integrated approach will allow the study of plant-bacterial interactions in rhizosphere as a "system" in complete and holistic manner, in this era of systems biology (Kiely *et al.*, 2006).

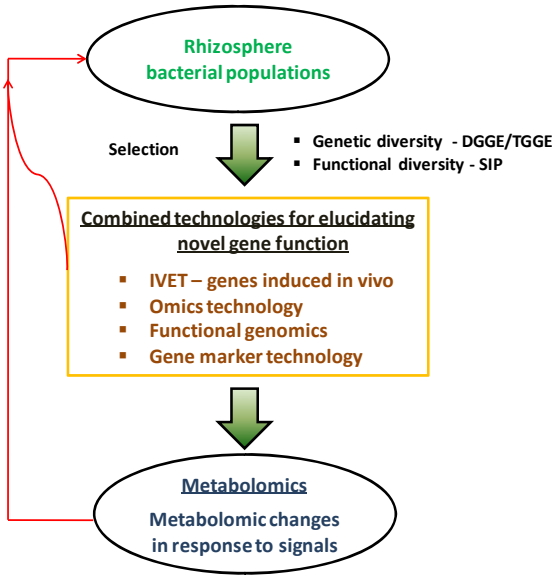



**Figure 1.4.1.** Integrated approach for the analysis of the microbial-host interactions in the rhizosphere (redrawn from (Kiely *et al.*, 2006)).

Another powerful technology to study microbial communities is metagenomics and metatranscriptomics. Jo Handelsman (1998), described metagenomics as the collection of genes sequenced from the (uncultured) samples obtained from the environment (Handelsman *et al.*, 1998). Later, Kevin Chen and Lior Pachter

redefined metagenomics as the application of modern genomic techniques to study microbial communities directly from the natural environment, bypassing the need for isolation and culturing under laboratory conditions (Chen and Pachter, 2005). Metagenomic approaches have been successfully applied in studying many microbial communities from complex environement such as sargasso sea (Venter *et al*., 2004) and acid mine drainage (Tyson *et al*., 2004). Recent rhizosphere metagenomic study, from the rhizosphere of plants adapted to acid mine drainage identified many novel nickel resistance genes (Mirete *et al*., 2007). The advances in genome sequencing technology such as Genome sequencer FLX systems (GS FLX) offers more advantages to study complex environmental samples through metagenomic approach. Its ability to achieve long and accurate sequence reads enables a wide range of applications in genome sequencing such as transcriptome sequencing. Recent, transcriptome sequencing with GS FLX system of *S. meliloti* identified 20 new genes (Mao *et al*., 2008). Transcriptome sequencing with cDNA derived from RNA samples recovered from rhizosphere, may aid in discovery of new genes and improve genome annotation of plant associated bacteria.

### 1.4.1 Culture-independent techniques.

It is well known that the plant species can select the microbial communities in the rhizosphere (Kuske *et al.*, 2002; Smalla *et al.*, 2001). A recent study showed that even the plant varieties can select for genetically distinct microbial populations in the rhizosphere (Kiely *et al.*, 2006). Earlier studies of the rhizosphere aimed at studying the influence of plant signals on the genetic diversity of the microbial communities in the rhizosphere. The two major methods developed to study the genetic diversity in the rhizosphere are, (1) denaturing gradient gel electrophoresis (DGGE), a molecular fingerprinting method that separates PCR amplified products, (2) temperature gradient gel electrophoresis (TGGE), electrophoresis performed with a temperature gradient across the gel (Muyzer, 1999). Both the methods were successful in detecting the microbial communities in the rhizosphere, but they failed to detect metabolically active microbes.

Stable isotope probing (SIP), is a developing technique used to study the functional diversity of microbial communities in the rhizosphere. The basic principle behind SIP is that the microbes are fed with [$^{13}$C] labeled substrates, the "heavy C" will be then incorporated into DNA. The heavy C will then be separated from the "normal C" by density gradient centrifugation (Radajewski *et al.,* 2000). SIP has been successfully applied to study biodegradation of pollutants in the rhizosphere and to explore biogeochemical cycles. Although some of the preliminary data are interesting, it is still at the developmental stage (Kiely *et al*., 2006). Another interesting method is the reverse of SIP principle, which is indirect labelling where the labeled substrates [$^{13}CO_2$] are incorporated into photosynthates. It is well known that 30-60% of the plant photosynthates are released in the root exudates and microbes will feed on it. Tracking the labelled substrates in the microbes will provide more clues to understand plant-microbial interactions (Griffiths *et al.*, 2004). This technique, also at the developmental stage, is expected to provide interesting information on microbial diversity on the rhizosphere.

### 1.4.2 Techniques to identify novel rhizosphere specific genes

Several elegant techniques have been developed to study the expression of microbial genes specifically in the host environment (Lee & Camilli, 2000; Rainey & Preston, 2000). Of these techniques many have been adapted to study bacterial genes specifically induced in rhizosphere (Rediers *et al.*, 2005). These techniques include differential fluorescence induction (DFI) (Allaway *et al.*, 2001), signature-tagged mutagenesis (STM) (Hensel *et al.*, 1995a), *in vivo* expression technology (IVET) and many modifications of IVET (Mahan *et al.*, 1993; Rediers *et al.*, 2005).

#### 1.4.2.1 *In vivo* expression technology

*In vivo* expression technology (IVET) was developed to study genes expressed by *Salmonella* during survival in mouse. IVET was based on the promoter trap technique developed to identify *Xanthomonas campestris* genes

induced during infection of turnips (Osbourn *et al.*, 1987). IVET is a promoter trapping technology, which allows positive selection of microbial genes specifically induced in the host environment (Mahan *et al.*, 1993). The success of IVET relies on two components; first, it requires a conditionally compromised strain, which is mutated in gene encoding essential growth factor (*egf*) and second, it requires a plasmid carrying the promoter trap composed of the promoterless *egf* gene and linked to a reporter gene (*rep*). The bacterial genomic DNA is cloned randomly into the promoter trap and integrated into the chromosome of the *egf* mutant strain. The mutant strains are then introduced into the host environment and the promoters which drive the expression of promoterless *egf* gene in the environment were identified and characterized. Any constitutive promoters were eliminated by screening the recovered bacteria in the general growth medium (Rediers *et al.*, 2005). The selection of the *egf* gene used to date can be broadly classified into auxotrophic, antibiotic and system-specific selection. The auxotrophic and antibiotic selection are very stringent and may be helpful in identifying strongly expressed promoters but miss those of weakly or transiently expressed genes (Hautefort & Hinton, 2000). Many modifications of IVET were developed to overcome these limitations and to be more specific for identifying the host induced genes. A thorough review of IVET and flavours of IVET can be found in (Rediers *et al.*, 2005). IVET has been applied very successfully in identifying many novel genes from rhizosphere bacteria such as rhizobia and many *Pseudomonas spp.*

IVET studies in *P. fluorescens* SBW25 during colonization of the sugarbeet rhizosphere identified 20 genes which code for proteins involved in Type III secretion, oxidative stress and transport (Rainey, 1999). Another IVET study, on colonization of soil by *P. fluorescens* PF0-1 identified many genes involved in nutrient utilization and transport, detoxification and genes involved in metabolic pathway (Silby & Levy, 2004). An IVET based study in *R. leguminosarum* A34 identified 29 genes related to transport, environmental response, metabolic and regulators, specifically induced in the pea rhizosphere (Barr *et al.*, 2008). IVET has also been used to identify genes involved in the early

stages of symbiosis in *S. meliloti*-alfalfa interaction (Oke & Long, 1999). Recently, a modified version of IVET, recombinase based IVET (RIVET) was successfully applied to identify genes induced during the *Rhizobium*-legume symbiosis (Gao & Teplitski, 2008). Another major advance in IVET is SpyVET, which combines suppressor analysis with IVET. In *P. fluorescens* SBW25, SpyVET identified many regulators which regulate the expression of environmentally induced genes (Giddens *et al.*, 2007). Despite these advances, only limited information is available about gene expression in the rhizosphere. The major limitation of the IVET strategy is that it is very time consuming and it can only identify up-regulated genes and not those that are down-regulated.

#### 1.4.2.2 Signature tagged mutagenesis

Signature tagged mutagenesis (STM) is a powerful negative selection method, used to identify the genes specifically required in the host environment. STM combines insertional mutagenesis and *in vivo* negative selection to identify attenuated mutants from a complex pool of mutants (Hensel *et al.*, 1995b; Saenz & Dehio, 2005). STM uses signature tags (short DNA sequences) inserted in each mutants to mark them individually, which allows the identification of attenuated mutants that did not survive in the host environment. The major advantage of STM is that it allows screening of a large number of mutants (library) in one passage through the host. The limitations of STM include those of insertional mutagenesis such as efficient transposon delivery systems, randomness of insertions and the polar and down-stream effects of insertion on the operon. The application of this technique in identifying bacterial genes in the rhizosphere may not identify truly environmentally induced genes, as its performance is controlled by very complex biological interactions in the rhizosphere (Rainey, 1999). STM has its own limitations such as weak or cross-reacting signals, inoculation concentration of mutants pools needs careful optimisation and the fact that loss of tags due to experimental errors can be misinterpreted as attenuation during recovery (Perry, 1999). Despite these limitations, STM has been applied to study many microbial pathogens in the host environment (Saenz & Dehio, 2005).

Recently, an improved STM was developed, taking the advantage of genome sequencing and microarray technology, in which a large library of tagged transposons was constructed and the tags were detected by two-channel microarray, by comparing input and output pools (Pobigaylo *et al.*, 2006). The application of this transposon based STM technique in *S. meliloti*, identified 67 genes, including 38 novel genes, which were specifically required during symbiosis. Mutation of of these genes showed mutants with a range of phenotypes, from attenuated symbiotic performance to reduced symbiotic competitiveness (Pobigaylo *et al.*, 2008). A similar approach is underway with the completion of construction of signature tagged *M. loti* mutants database, which is expected to provide a wealth of information with functional genomics (Shimoda *et al.*, 2008). Although, transposon based STM is focussed primarily on understanding the survival and competitiveness of rhizobia in the host, research is in progress to apply transposon based STM in *R. leguminosarum* bv. *viciae* 3841 to identify the genes involved in survival and colonization of the pea rhizosphere (Poole PS et al., personal communication).

#### 1.4.2.3 Differential Fluorescence Induction

Differential Fluorescence Induction (DFI) is another promoter trapping technique, which uses green fluorescent protein (GFP) reporter fusion to monitor promoter activity. The DFI system is linked with the fluorescence-activated cell sorter (FACS) or an optical trapping (OT) system (Allaway *et al.*, 2001) enables high-throughput screening of gene expression in microorganisms. Optical trapping uses the property of coherent light to produce radiational pressure, sufficient to hold a bacterium within the light beam (Ashkin & Dziedzic, 1987; Ashkin *et al.*, 1987). Continuous screening results in a enriched clone population containing genes specifically induced in environment (Valdivia & Falkow, 1996; Valdivia & Falkow, 1997b). The advantages of DFI are the semiautomated screening of large populations and the ability to alter the sensitivity of the fluorescence to detect changes in expression levels rather than simply identifying tightly regulated promoters. In DFI, the transcriptional fusions are on the plasmid, as single copy of

expression is sufficient to be detected by the FACS. This enables the analysis of the activity of single cells and the results are highly reproducible. The limitations in using DFI are sorting bacteria, interruption of background fluorescence from the particles in the host environment and the difficulty of isolating bacteria from naturally fluorescent environment (Allaway *et al.*, 2001; Valdivia & Falkow, 1996).

DFI was originally developed to identify acid-inducible genes in *S. enterica* serovar *typhimurium* (Valdivia & Falkow, 1996; Valdivia & Falkow, 1997a). Later, DFI was adapted to study host-induced genes in many microbial pathogens including *Mycobacterium marinum* (Barker *et al.*, 1998), *Listeria monocytogenes* (Wilson *et al*., 2001), *Strpetococcus pneumoniae* (Marra *et al.*, 2002). DFI was also used to identify the host specific genes in some plant-associated bacteria such as *R. leguminosarum* (Allaway *et al.*, 2001), *Pseudomonas syringae* (Chang *et al.*, 2005) and *Bacillus cereus* (Dunn *et al.*, 2003). Allaway *et al.,* (2001) identified many genes specifically induced in *R. leguminosarum* when colonizing the pea rhizosphere. With the combination of differential fluorescence induction and optical trapping (DFI-OT) 24 genes were identified, including *thiE*, *putA* and genes involved in homo-spermidine metabolism and methioine transport.

#### 1.4.2.4 Transcriptomics

The transcriptome is the set of all mRNA molecules in the cell. Transcriptomics is a global method of analyzing the transcripts expressed in the cell under various conditions. The most common and powerful tool used in transcriptomics is the microarray; a high-throughput and robust method capable of quantifying all the transcripts of a genome. Microarrays come in two main types; 1) oligonucleotide and 2) complementary DNA (cDNA) microarrays. Microarrays have the option of being performed with single or two colours. Single colour microarrays allow an estimation of the absolute expression of the transcripts, whereas two-colour microarrays estimate the relative expression of experimental sample versus control sample. Two-colour microarrays are generally used in

studying bacterial transcriptomics, where the effect of environmental perturbations, nutrient source or chemical substrates is studied. In a typical two-colour microarray, total RNA samples are extracted from experimental and control samples, and reverse transcribed into cDNA. The cDNA is fluorescently labelled (control - Cy3 and experimental - Cy5) with CyDyes. After labelling, equal amounts of each labelled cDNA are mixed and hybridized onto the microarray slide under controlled conditions. After a stipulated time period, the microarray slides are washed in buffers to remove the unbound dyes and finally, the slides are scanned and the spots on the slide are quantified. The acquired data is then normalized and analyzed with high-end statistical microarray-informatics tools (DeRisi *et al.*, 1997).

The major limitation in performing microarrays with environmental and/or rare samples is the low amounts of available RNA. This limitation is overcome by two main strategies, signal amplification and sample amplification. The first strategy, can be performed with technologies like dendrimer (Stears *et al.*, 2000) or tyramide signal amplification (TSA) (Karsten *et al.*, 2002). In dendrimer technology, cDNA is synthesized using an oligo dT primer with a capture sequence and hybridized on the slide. 3DNA dendrimer molecules with a multitude of fluorescent molecules with sequence complementary to the capture sequence are hybridized to cDNA, bound to a complementary oligo on the slide. In the second strategy, total mRNA from the cell is linearly amplified, by T7 based *in vitro* transcription (IVT), which is referred as the "Eberwine method" (Van Gelder *et al.*, 1990). In this method, the mRNA templates are primed with an oligo (dT) primer and reverse transcribed into cDNA. The RNA-cDNA hybrid is treated with RNAseH, to degrade the RNA. The second strand cDNA is primed by RNA nicking, and the cDNA is synthesized by *E. coli* DNA polymerase and *E. coli* DNA ligase followed by blunt ending with T4 DNA polymerase. The antisense RNA (aRNA) is synthesized by T7 RNA polymerase, which binds to the T7 promoter in the first strand cDNA.

The application to microarrays is very powerful and is enhanced with techniques like linear RNA amplification, which allows use of small quantities of

RNA from the environmental and/or rare samples. A combination of linear RNA amplification and microarray technology even makes it possible to perform microarrays at the single cell level (Nygaard & Hovig, 2006). The completion of many plant associated bacterial genomes provided an opportunity to revisit plant-bacterial interactions on a global scale. Over the past five years, microarrays have been used extensively in studying plant-bacterial interactions. In rhizobia, most of the microarray based analyses were focussed on the identifying novel genes in *S. meliloti* during symbiosis (Ampe *et al.*, 2003; Djordjevic *et al.*, 2003) and on free-living *S. meliloti* grown with the various altered environmental conditions such as osmotic stress (Ruberg et al., 2003), different growth media (Ampe *et al*., 2003) and oxygen limitation (Becker *et al.*, 2004). Microarray analysis *in planta* was further extended when a time course analysis of *S. meliloti-M. sativa* nodules was performed to identify the bacteroid gene expression at various stages of symbiotic interaction (Ampe *et al.*, 2003). Oligonucleotide microarrays of *B. japonicum* were used to analyze the transcription profiles under a variety of conditions, osmotic stress, and free-living cells versus bacteroids (Chang *et al.*, 2007). These microarrays have identified many novel genes specific to symbiosis in rhizobia and a wealth of data to better understand the symbiotic interaction. However, only very little information was known about the survival and competitiveness of rhizobia in the rhizosphere.

The major limitations in performing microarrays on rhizosphere bacteria are the complex nature of the rhizosphere (soil), and the difficulty in isolating the bacterial cells and the low yield of RNA extracted from those cells. So rhizosphere microarrays were performed with bacteria cells recovered from *in vitro* sterile host rhizosphere systems grown under controlled conditions. As an initial step in identifying novel rhizosphere specific genes using microarrays, Mark *et al*. (2005) supplemented aseptically obtained sugarbeet root exudates from two beet cultivars to laboratory media individually and studied the changes in the transcriptome of *P. aeruginosa* PA01 under laboratory conditions. This allowed identification of many genes involved in metabolism, chemotaxis, type III secretion systems and many putative or of unknown function. The major

drawback in this study is that the root exudates were collected in an aseptic environment, without the influence of bacteria, which does not reflect real conditions as the bacterial presence in the rhizosphere largely influences the composition of the secreted root exudates (Cooper, 2004). A recent report, overcame this drawback by performing transcriptomic analysis on *P. putida* recovered from corn rhizosphere (*in vitro*) compared against the exponential, stationary and microcosm grown free-living cells under laboratory conditions. This analysis identified 90 genes to be up-regulated in the rhizosphere and involved in various functions including amino acid uptake, aromatic compound metabolism, cytochrome biosynthesis, chemotaxis and motility, transcriptional regulators and sensor proteins, stress adaptation and detoxification, transporters and efflux pumps, and genes involved in DNA replication and repair mechanisms (Matilla *et al.*, 2007). However, this information only provides partial information on the transcriptome of bacteria in the rhizosphere.

Another important development in microarray is Functional Gene Arrays (FGA). FGA are normal DNA arrays with the oligonuleotide probes for genes of specific interest (carbon or nitrogen cycle) from variety of different organisms. RNA harvested from the soil, reverse transcribed to cDNA and hybridized onto the array to identify the genes induced in the soil (Saleh-Lakha *et al.*, 2005). Schadt *et al*. (2005) described the design of an array targeting soil bacterial genes involved in metal resistance, biodegradation, and carbon, nitrogen and sulfur cycling. The 50-mer oligonucleotides array contained 1622 unique probes and will hybridize to labelled cDNA if the identity between them is <86% (Rhee *et al.*, 2004; Schadt *et al.*, 2005). Although this technique is promising, a lot of optimisation is required to improve the sensitivity of the detection.

#### 1.4.2.5 Proteomics

The proteome, as defined by Marc Wilkins is the protein complement of the genome (Wilkins *et al.*, 1996). Proteomics is the large scale analysis of all the proteins in a cell. In this post-genomic era, with a wealth of data from the annotated genomes, very little is known about the functional significance of the

genes in a global manner. Proteomics is a powerful tool to understand the expression patterns of the genome under various conditions. The comparison of transcriptomic and proteomic data will provide a unique opportunity to study the functional significance of each gene in a genomic context. However, careful optimization of the design and sampling conditions need to be performed to interpret the data together (Hegde *et al.*, 2003). In a proteomic analysis, proteins are extracted from the cell, and then solubilized and separated by 2D gel electrophoresis (2DGE). In 2DGE, the proteins are separated based on isoelectric point in the first dimension and by molecular mass in the second dimension. After staining with Coomassie brilliant blue, the protein spots from the gel are excised and identified using mass spectrometry. Two-dimensional difference gel electrophoresis (2D DIGE) is a new technique where proteins are labelled prior to electrophoresis with fluorescent CyDyes and co-separated on the same 2D gel (Gade *et al.,* 2003).

Proteomics has now been successfully used in many plant symbionts to identify proteins that are specifically expressed during plant-microbial interactions. In rhizobia, proteomic analysis have been successfully used to identify novel proteins induced during symbiosis in *R. leguminosarum*/*P. sativum* (Saalbach *et al.*, 2002), *S. meliloti/M. truncatula* and *M. loti/Lotus japonicus* (Wienkoop & Saalbach, 2003) symbiosis. Comparison of two proteomes will be more interesting, and provide excellent opportunity to identify novel proteins from isogenic strains. The comparison of the proteome of *R. leguminosarum* bv *trifolii* strains ANU843, which forms nitrogen-fixing nodules and ANU794, which forms aberrant nodules, has identified many novel proteins that play role in early stages of nodulation (Morris & Djordjevic, 2001). As of now, proteomics and metabolomics for rhizosphere bacteria have not been reported.

### 1.4.3 Functional genomics of novel genes identified in the rhizosphere

The progression of genome sequencing technology and the rapid accumulation of sequence data has provided a wealth of information to view bacterial cell function in a genomic perspective. The “omic” technologies such as

transcriptomics, proteomics and metabolomics have complemented each other well enabling a "systems" level study. On the other hand, approximately 30-40% of annotated bacterial genes are given only putative or unknown functions. Although sequence comparison and analysis, together with co-regulation by transcriptomics, can identify possible roles of these genes, the exact biological roles remained unanswered. Therefore, functional genomic analysis is required to assess the biological significance of a gene in a global profiling. The two common functional genomic approaches are 1) targeted disruption of the gene to assess the biological role of the gene and 2) reporter gene fusions such as GFP, to study the temporal and spatial expression of the gene.

Targeted mutagenesis has always been a successful method in assessing the biological function of a gene. There are a variety of strategies available for insertional or deletion mutagenesis. The advancement in methods for constructing high-throughput bacterial mutant libraries also provides an opportunity to identify the function of target genes in a relatively short time. In rhizobia, mutagenesis has been widely applied to study the biological function of many novel genes involved in symbiosis, often combined with rhizosphere or nodule competitiveness studies, where mutants were compared against wild-type (already reviewed in detail in section 1.3.5).

Reporter gene technology, either as bioreporters or biosensors, has been extensively used in environmental studies of the rhizosphere to characterize microbial distribution and/or to assess the availability of nutrients (excess/limitation) (Larrainzar *et al.*, 2005). Reporter genes are used to monitor transcription by creating fusions with promoters of genes (Hautefort & Hinton, 2000). In most rhizosphere/soil based studies, the reporter genes are fused with promoters induced by specific compounds, enabling the detection of the presence or absence of the compounds. Commonly used reporter genes include *lacZ* (Labes *et al.*, 1990), *gusA* (Prell *et al.*, 2002; Reeve *et al.*, 1999), *luc* and *lux* (Prosser et al., 1996) and *inaZ* (Miller *et al*., 2000). Apart form these reporter genes, autofluorescent proteins (AFPs) are widely used in studying gene expression in the cell and natural environment (Gage, 2002; Karunakaran *et al.*, 2005; Stuurman

*et al.*, 2000; Xi *et al.*, 2001). Reporter genes can be assayed by their ability to glow or fluoresce. Some reporter genes have the sensitivity to be detected in single bacterial cell (Harms et al., 2001). To use reporter genes in gene expression studies in natural environments, certain criteria should be considered for valid results. The criteria are; 1) detectability in heterogeneous system, 2) distinct from background populations and 3) sensitivity of the reporter gene. The major limitation is that the induction of reporter genes depends on the plant or the host organism (Cardon & Gage, 2006; Wright & Beattie, 2004). In the following section, a brief description of the few reporter genes and their application in studying gene expression in rhizosphere is discussed.

**Luciferase:** One of the more powerful reporter genes is luciferase (*luc* or *luxAB*), which can generate light that can be measured with luminometric devices. Luciferase requires addition of substrates and utilizes cellular energy (ATP) (Kiely *et al.*, 2006). The advantage of using *luxAB*, is that the protein product which emits light has a short half life, ensuring the photon production reflects the real-time of gene expression and this enzyme is not present in bacteria, which rule out the possibility of background luminescence. The disadvantages are that they use energy from the cell and they are highly oxygen dependent, and therefore photo emission is highly correlated with the availability of oxygen. Despite these limitations, *luxAB* has been extensively used in many environmental studies as a bioreporter. In *P. fluorescens*, tripartite *luxAB* reporter system was constructed to report on assimilable carbon, nitrogen and phosphorus status. The quantitative bioluminescence of reference materials and soil water, was in good agreement suggesting the success of the tripartite *lux* system (Standing *et al.*, 2003). Three other bioreporters were reported individually in *P. fluorescens* to detect phosphorous starvation (Kragelund *et al.*, 1995; Kragelund *et al.*, 1997), presence of copper (Tom-Petersen *et al.*, 2001) and nitrate limitation (Darwent *et al.*, 2003) in the soil.

**Green Fluorescent protein (GFP):** The most common auto fluorescent protein (AFP) used is green fluorescent protein (GFP), a monomeric 23 kDa protein, which was isolated from luminous coelenterates of the genus *Aequorea* (Chalfie *et al.*, 1994). The main advantage of GFP is the autofluorescence, which is emitted in response to the exposure of light of a particular wavelength, without the addition of the chemical substrates and cellular energy. This advantage makes GFP a reporter of choice over the other available reporter genes. GFP has been successfully applied to many environmental studies including bioreporter/biosensor studies, biofilm maturation and in plant-bacterial interactions (Larrainzar *et al.*, 2005). For simple experiments, GFP can be tagged to a target gene and monitored *in situ*. To study, complex plant-bacterial interactions, GFP can be combined with microscopy ie, Confocal laser scanning microscopy (CLSM) to monitor the expression in real-time without excessive sample processing (Kiely *et al.*, 2006). To add more advantage to this, variants of GFP were designed with a range of spectral characteristic and improved biochemical properties (Stuurman *et al.*, 2000). The variants allow dual labeling of cells, where two different conditions are tested.

GFP has been used in many plant-bacterial interaction studies. In *R. leguminosarum*, to confirm that the expression of *thiMED* genes was driven by the upstream promoter, a strain with a *thiME*::gfp-UV fusion was inoculated into the vetch rhizosphere, a thiamine limiting environment. After 3-7 days post inoculation, the expression of *gfp* was observed, confirming the expression *thiMED* by the upstream promoter (Karunakaran *et al.*, 2006). Another study, used a *dctAp*::*DsRedT.3* fusion (*dctA* - dicarboxylate transport system) to monitor the expression of dicarboxylate transporter in the infection thread on the vetch roots. In the *Rhizobium*-legume symbiosis, after successful colonization, rhizobia must first attach to root hairs and grow down plant derived infection threads, and bacteria differentiate into bacteroids to fix nitrogen. While dicarboxylates are used by bacteroids to fuel nitrogen fixation they are not the only carbon source available for bacteroid formation (Finan *et al.*, 1981). The expression of *dctAp*::*DsRedT.3* fusion in the infection threads of vetch plants confirms

availability of dicarboxylates in the vetch infection thread (Karunakaran *et al.*, 2005). Gage (2002) infected alfalfa plants with mixtures of *S. meliloti* containing *gfp* (green) and *DsRed* (red) strains to visualize the events during infection thread development. The visualization results show more infection threads with mixed strains than single strains (Gage, 2002).

GFP also has been used widely as a biosensor in detecting the presence or absence of specific compounds in rhizosphere or soil. Two bioreporters were reported individually in *P. fluorescens* designed to detect the presence of 3-chlorobiphenyls in the alfalfa rhizosphere (Boldt *et al.*, 2004) and the presence of toluene in the barley rhizosphere (Casavant *et al.*, 2003). In *S. meliloti*, the galactoside biosensor was used to probe for the presence of galactose and galactosides in the alfalfa rhizosphere (Bringhurst *et al.*, 2001).

Recent high-throughput transportomic study, used a *gfp* reporter gene to construct fusions to all the ATP-binding cassette (ABC) and tripartite ATP-independent periplasmic (TRAP) transport systems in *S. meliloti*, with a total of 498 fusions (300 plasmid and 198 integrated fusions), which were tested with up to 174 biologically active compounds under laboratory conditions. This study identified 76 systems to be induced by one or more compounds including many novel transport systems (Mauchline *et al.*, 2006). Although, this study was not based on the rhizosphere, it is worth mentioning because of its wealth of information regarding the transport system, from which many biosensors can be constructed.

Symbiotic interaction between rhizobia and legume host is a complex and tightly regulated process, which involves successful completion of many steps, including rhizosphere colonization, attachment to roots, infection thread formation and nodulation (Broughton *et al.*, 2000). Legume rhizosphere colonization is one of the first and foremost steps in establishing a symbiotic interaction. Despite being an important step, its complex nature and lack of suitable technology for investigation, kept it as an unexplored area compared to nodulation and bacteroid metabolism. In this post genomic era, this research takes advantage of the completion of Rlv3841 genome sequencing and the advances in microarray technology to decipher the transcriptome of Rlv3841 in the pea rhizosphere.

This research aims to use an *in vitro* sterile pea rhizosphere system as a model rhizosphere to study the transcriptome of Rlv3841. The major limitation in performing microarray analysis in rhizosphere bacteria is the low recovery of total RNA from the rhizosphere samples. This limitation can be overcome by either amplifying RNA or by amplifying the signals from the microarray slide. Therefore, the first step in this research was to optimize a strategy to improve the cell recovery from the rhizosphere system, extract total RNA from recovered cells and amplify RNA to perform microarrays. Control experiments were carried out to study the changes in the transcriptome of free-living Rlv3841 in response to pea root exudates and hesperetin individually under laboratory growth conditions. After optimizing the strategy, microarrays were performed with rhizosphere recovered samples compared to free-living Rlv3841 cells.

In the first section of the project, transcriptome of Rlv3841 in the pea rhizosphere was investigated under two different conditions. In the first condition, microarrays were performed with cells recovered from the 7d old pea rhizosphere after three (1, 3 or 7 days) post inoculation time points to identify all the genes whose expression changes over time. In the second condition, microarrays were performed with cells recovered after one day post inoculation from three differently aged pea rhizospheres (7, 14 or 21 days) to study the effect pea

seedling age on the transcriptome of Rlv3841. The microarray data was then analyzed and then validated by performing qRT-PCR on a set of differentially expressed genes.

The second section of the project was designed to perform comparative rhizosphere transcriptomics with Rlv3841 cells recovered from 7d old pea (cognate host), alfalfa (other-legume host) or sugarbeet (non-legume) rhizospheres after 7 days post inoculation, by two approaches (i) direct and (ii) indirect design. In the direct design, the three rhizosphere samples were compared against each other and in the indirect design, the three rhizosphere samples were compared to common reference (free-living Rlv3841) individually. The comparative data analysis is expected to provide more insights on Rlv3841 pea specific genes.

In the third section, the biological significance of some the upregulated genes in rhizosphere colonization was assessed by performing competitive colonization assay in the pea rhizosphere. Finally, based on the obtained microarray data, transcriptome of the Rlv3841 in the pea rhizosphere was modelled.

In summary the overall objectives were;

1. Optimize RNA extraction, amplification and labelling conditions for microarray analysis of rhizosphere samples.
2. Determine which genes are specifically up-regulated in the pea rhizosphere.
3. Use gene expression as “biosensor” of the rhizosphere environment.
4. Determine if specific genes are switched on in different plant rhizospheres.
5. Determine if up-regulated genes affect colonization ability.

# Chapter 2: Materials and Methods

### 2.1.1 Culture conditions

*R. leguminosarum* strains 3841 (Rlv3841) and 300 (Rlv300) were grown on either agar or in broth at 26ºC. The growth media used was Tryptone Yeast (TY) (Beringer, 1974) or Acid Minimal Salts (AMS) (Poole *et al.*, 1994). AMS media consists of 97 ml of 0.5 mM $K_2HPO_4$, 2 mM $MgSO_4.7H_2O$, 3.4 mM NaCl, 20 mM MOPS buffer; 1 ml of *Rhizobium* solution A (40 mM EDTA-$Na_2$, 0.56 mM $ZnSO_4.7H_2O$, 0.83 mM $NaMoO_4.2H_2O$, 4 mM $H_3BO_3$, 2.3 mM $MnSO_4.4H_2O$, 80 µM $CuSO_4.5H_2O$, 4.2 µM $CoCl_2.6H_2O$) and 1 ml of *Rhizobium* solution B (115.3 mM $CaCl_2$ and 21.7 mM $FeSO_4$), buffered to pH 7 with 1 M NaOH. 1 ml of *Rhizobium* solution C (3 mM Thiamin-HCl, 4.2 mM D-Pantothenic acid calcium salt and 4.1 mM Biotin) was added after autoclaving along with antibiotics, carbon and nitrogen sources.

For extracting total RNA from Rlv3841 grown in the presence of pea root exudates, AMS media was prepared with high concentration stocks of the components to keep the pea root exudates undiluted, with 10µl of 1M $K_2HPO_4$, 40µl of 1 M $MgSO_4.7H_2O$, 63 µl of 1 M NaCl, 200 µl of 2 M MOPS buffer pH 7.0; 20 µl of solution A, 40 µl of solution B, 20 µl of solution C, 200 µl of 1 M $NH_4Cl$ and 18.8 ml of pea root juice to a final volume of 20 ml. AMS was supplemented with 10 mM glucose or 30 mM pyruvate and 10 mM ammonium chloride as carbon and nitrogen sources respectively, unless otherwise stated. *E. coli* strains were grown in Luria-Bertani (LB) (Miller, 1972) agar or broth. Strains were routinely stored at -20ºC for short term use and -80ºC for long term in 15% (v/v) glycerol after snap freezing in liquid nitrogen. Media were supplemented with antibiotics, fungicide and stains at the following concentrations when required (Table 2.1.1).

| **Antibiotic / Fungicide / Stain** | ***E. coli* (µg/ml)** | ***R. leguminosarum* (µg/ml)** |
|---|---|---|
| Streptomycin | - | 500 |
| Neomycin | - | 80 |
| Trimethioprim | - | 10 |
| Kanamycin | 20 | - |
| Nyastatin | - | 50 |
| X-Gal | 40 | - |

**Table 2.1.1.** List of antibiotics, fungicides and stains used.

### 2.1.2 List of Strains

The strains used in this study with their descriptions are listed below.

| Strain | Genotype/ Description | Reference |
|---|---|---|
| *Rhizobium leguminosarum* | | |
| 300 | Wild-type (wt) *R. leguminosarum*, $trim^r$, $str^s$ | (Johnston & Beringer, 1975) |
| 3841 | Wild-type (wt) *R. leguminosarum*, a spontaneous streptomycin-resistant derivative of strain 300, $trim^r$, $str^r$ | (Johnston & Beringer, 1975) |
| RU4222 | Strain 300 containing pRU2082, $trim^r$, $neo^r$ | This work |
| RU4229 | Strain 300 containing pRU2021, $trim^r$, $neo^r$ | This work |
| RU4230 | Strain 300 containing pRU2024, $trim^r$, $neo^r$ | This work |
| RU4231 | Strain 300 containing pRU2025, $trim^r$, $neo^r$ | This work |
| RU4232 | Strain 300 containing pRU2033, $trim^r$, $neo^r$ | This work |
| RU4233 | Strain 300 containing pRU2034, $trim^r$, $neo^r$ | This work |
| RU4234 | Strain 300 containing pRU2042, $trim^r$, $neo^r$ | This work |
| RU4235 | Strain 300 containing pRU2044, $trim^r$, $neo^r$ | This work |
| RU4247 | Strain 300 containing pRU2105, $trim^r$, $neo^r$ | This work |
| RU4248 | Strain 300 containing pRU2128, $trim^r$, $neo^r$ | This work |
| RU4249 | Strain 300 containing pRU2100, $trim^r$, $neo^r$ | This work |
| RU4250 | Strain 300 containing pRU2101, $trim^r$, $neo^r$ | This work |
| RU4251 | Strain 300 containing pRU2130, $trim^r$, $neo^r$ | This work |
| RU4252 | Strain 300 containing pRU2131, $trim^r$, $neo^r$ | This work |
| RU4253 | Strain 300 containing pRU2132, $trim^r$, $neo^r$ | This work |
| RU4254 | Strain 300 containing pRU2102, $trim^r$, $neo^r$ | This work |
| RU4255 | Strain 300 containing pRU2103, $trim^r$, $neo^r$ | This work |
| RU4256 | Strain 300 containing pRU2107, $trim^r$, $neo^r$ | This work |
| RU4257 | Strain 300 containing pRU2108, $trim^r$, $neo^r$ | This work |
| RU4258 | Strain 300 containing pRU2111, $trim^r$, $neo^r$ | This work |
| RU4259 | Strain 300 containing pRU2113, $trim^r$, $neo^r$ | This work |
| RU4260 | Strain 300 containing pRU2115, $trim^r$, $neo^r$ | This work |
| RU4261 | Strain 300 containing pRU2117, $trim^r$, $neo^r$ | This work |
| RU4262 | Strain 300 containing pRU2118, $trim^r$, $neo^r$ | This work |
| RU4263 | Strain 300 containing pRU2120, $trim^r$, $neo^r$ | This work |
| RU4265 | Strain 300 containing pRU2125, $trim^r$, $neo^r$ | This work |
| RU4266 | Strain 300 containing pRU2126, $trim^r$, $neo^r$ | This work |
| RU4267 | Strain 300 containing pRU2109, $trim^r$, $neo^r$ | This work |
| RU4268 | Strain 300 containing pRU2110, $trim^r$, $neo^r$ | This work |
| RU4269 | Strain 300 containing pRU2112, $trim^r$, $neo^r$ | This work |
| RU4270 | Strain 300 containing pRU2133, $trim^r$, $neo^r$ | This work |
| RU4271 | Strain 300 containing pRU2114, $trim^r$, $neo^r$ | This work |
| RU4272 | Strain 300 containing pRU2121, $trim^r$, $neo^r$ | This work |
| RU4273 | Strain 300 containing pRU2123, $trim^r$, $neo^r$ | This work |
| RU4274 | Strain 300 containing pRU2047, $trim^r$, $neo^r$ | This work |

| | | |
|---|---|---|
| RU4295 | Strain 300 containing pRU2116, trim$^r$, neo$^r$ | This work |
| RU4296 | Strain 300 containing pRU2119, trim$^r$, neo$^r$ | This work |
| RU4297 | Strain 300 containing pRU2127, trim$^r$, neo$^r$ | This work |
| RU4298 | Strain 300 containing pRU2038, trim$^r$, neo$^r$ | This work |
| RU4308 | Strain 300 containing pRU2166, trim$^r$, neo$^r$ | This work |
| RU4309 | Strain 300 containing pRU2124, trim$^r$, neo$^r$ | This work |
| RU4310 | Strain 300 containing pRU2167, trim$^r$, neo$^r$ | This work |
| RU4311 | Strain 300 containing pRU2053, trim$^r$, neo$^r$ | This work |
| RU4312 | Strain 300 containing pRU2168, trim$^r$, neo$^r$ | This work |
| RU4317 | Strain 300 containing pRU2129, trim$^r$, neo$^r$ | This work |
| RU4318 | Strain 300 containing pRU2170, trim$^r$, neo$^r$ | This work |
| RU4358 | Strain 300 containing pRU2187, trim$^r$, neo$^r$ | This work |
| RU4360 | Strain 300 containing pRU2056, trim$^r$, neo$^r$ | This work |
| *Escherichia coli* | | |
| DH5α | *Escherichia coli*, nal$^r$ | (Sambrook et al., 1989) |

**Table 2.1.2.** List of strains used in this study.

### 2.1.3 List of Plasmids

The plasmids used in this study with their descriptions are listed below.

| Plasmid | Details | Reference |
|---|---|---|
| pK19mob | used to generate insertion mutants in 300 containing *lacZ*, kan/neo$^r$ | (Schafer *et al.*, 1994) |
| pRK2013 | conjugal helper plasmid, required to allow DH5α to conjugate with 300, kan/neo$^r$ | (Figurski & Helinski, 1979) |
| pRU2021 | p1214 & p1215 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2024 | p1223 & p1224 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2025 | p1226 & p1227 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2033 | p1175 & p1176 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2034 | p1071 & p1072 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2038 | p1097 & p1098 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2042 | p1093 & p1094 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2044 | p1101 & p11102 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2047 | BD_pckA_for & BD_pckA_rev PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2053 | p1545 & p1546 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2056 | p1208 & p1209 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2082 | BD_RL1911_for & BD_RL1911_rev PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |

| | | |
|---|---|---|
| pRU2100 | p1422 & p1423 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |
| pRU2101 | p1431 & p1432 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |
| pRU2102 | p1392 & p1393 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |
| pRU2103 | p1389 & p1390 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |
| pRU2105 | p1095 & p1096 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |
| pRU2107 | p1419 & p1420 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |
| pRU2108 | p1440 & p1441 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |
| pRU2109 | p1470 & p1471 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |
| pRU2110 | p1383 & p1384 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |
| pRU2111 | p1401 & p1402 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |
| pRU2112 | p1416 & p1417 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |
| pRU2113 | p1437 & p1438 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |
| pRU2114 | p1461 & p1462 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |
| pRU2115 | p1410 & p1411 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |
| pRU2116 | p1413 & p1414 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |
| pRU2117 | p1428 & p1429 PCR product from strain | This work |

| | | |
|---|---|---|
| | 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | |
| pRU2118 | p1443 & p1444 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2119 | p1473 & p1474 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2120 | p1395 & p1396 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2121 | p1407 & p1408 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2123 | p1455 & p1456 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2124 | p1464 & p1465 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2125 | p1404 & p1405 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2126 | p1425 & p1426 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2127 | p1449 & p1450 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2128 | p1398 & p1399 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2129 | p1446 & p1447 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2130 | p1478 & p1479 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2131 | p1458 & p1459 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2132 | p1467 & p1468 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, kan$^r$ | This work |
| pRU2133 | p1452 & p1453 PCR product from strain 3841 DNA flanked by 15bp extensions | This work |

| | | |
|---|---|---|
| | homologous to pK19mob ends, $kan^r$ | |
| pRU2166 | p1548 & p1549 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |
| pRU2167 | p1551 & p1552 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |
| pRU2168 | p1557 & p1558 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |
| pRU2170 | p1560 & p1561 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |
| pRU2187 | p1636 & p1637 PCR product from strain 3841 DNA flanked by 15bp extensions homologous to pK19mob ends, $kan^r$ | This work |

**Table 2.1.3.** List of plasmids used in this study.

### 2.2.1 DNA Isolation

Rlv3841 was freshly grown on a TY slope and resuspended in 5 ml of TY broth, spun down and the pellet was resuspended in 1 ml of TY broth. Chromosomal DNA was isolated using the DNeasy tissue kit for isolation of total DNA from animal tissue (Qiagen) or using alkaline polyethylene glycol (PEG) (Chomczynski & Rymaszewski, 2006). Plasmid DNA from *E. coli* cells was isolated and purified by Wizard *Plus* SV Minipreps purification system (Promega) according to the manufacturer's instructions.

### 2.2.2 Agarose Gel Electrophoresis and Staining

Agarose gel was prepared at various concentrations from 0.8% - 1.6% (w/v), in Tris Acetate (400 mM), EDTA (1 mM) buffer to resolve different sized DNA fragment. Samples were loaded in 1:5 ratio with 6x loading buffer (30% (v/v) glycerol, 0.25% (w/v) bromophenol blue), with $1Kb^{+}$ DNA ladder (Invitrogen) to determine size relatively. Gels were run on a Bio-Rad gel electrophoresis system in 90 volts/cm for 40 min and stained in ethidium bromide solution (1 μg/ml) for 20 min and then destained in water for 10 min, and visualized under a UV trans-illuminator.

### 2.2.3 Cloning and Transformation

DNA was digested by restriction endonucleases from Invitrogen, New England Biolabs and Fermentas as per manufacturer's instructions. High-throughput cloning was performed using the In-Fusion dry-down PCR cloning kit (Clontech). The PCR primer pairs that were used to amplify the insert (internal region of gene) had a 15 bp flanking region homologous at each end with the linearization site (HindIII) of the vector (pK19mob), generating a homologous region in the PCR product during amplification. The amplified insert and the HindIII digested pK19 vector was purified and concentrated using a MinElute PCR purification kit (Qiagen). The purified insert and the vector were added to

the tube containing the In-Fusion dry down pellet and mixed well by pipetting. The ligation was carried out in a water-bath at 37ºC for 15 min and subsequently at 50ºC for 15 min, and chilled immediately. The reaction mix was then diluted with 40 µl of 1X TE buffer and 2.5 µl of the diluted reaction mix was incubated with the Fusion-Blue competent cells (Clontech) on ice. After 30 min, the competent cells were then heat-shocked at 42ºC for 45 seconds and chilled. After a min, the cells were then incubated in a rotatory shaker at 37° C for 1 hr with 450 µl of SOC (supplied with competent cells) for an hour. The transformants were selected on LB agar containing kanamycin and X-gal, after incubating at 37° C overnight. The presence of the insert in the vector present in white colonies was confirmed first by colony PCR as described in section 2.2.5 with M13 primer pairs and subsequently by restriction fragment analysis.

### 2.2.4 PCR primers

Primers used for the PCR were designed in Vector NTI version 9 (InforMax), and synthesized by Operon Biotechnologies, GmbH, Germany. A list of PCR primers used in this work is given in the Table 2.2.1**.**

| **Name** | **Sequence 5' – 3'** | **Target** |
|---|---|---|
| pK19A | ATCAGATCTTGATCCCCTGC | Antisense primer used for PCR amplification for region of pK19mob |
| pK19B | GCACGAGGGAGCTTCCAGGG | Sense primer used for PCR amplification for region of pK19mob |
| M13 uni | TGTAAAACGACGGCCAGT | Universal forward sequencing primer |
| M13 rev | CAGGAAACAGCTATGACC | Universal reverse sequencing primer |
| p1383 | TGATTACGCCAAGCTGAACGA CAGATCGGGATCGC | Sense primer for PCR amplification of a internal region of pRL110281 |
| p1384 | GCAGGCATGCAAGCTAAACGC ACCATCGCTCAGCT | Antisense primer for PCR amplification of a internal region of pRL110281 |
| p1385 | CGCCAAGAAACACGACAAGC | Mapping primer for PCR amplification of pRL110281 |

| | | |
|---|---|---|
| p1389 | TGATTACGCCAAGCTGGCTGATGAAGCTCTCGTCC | Sense primer for PCR amplification of a internal region of RL3272 |
| p1390 | GCAGGCATGCAAGCTATATCACCCGCTGCGAGGAA | Antisense primer for PCR amplification of a internal region of RL3272 |
| p1391 | CCCGGAAGATCGCTGCAAAG | Mapping primer for PCR amplification of RL3272 |
| p1392 | TGATTACGCCAAGCTGCCGACGGAAGTGACGATGT | Sense primer for PCR amplification of a internal region of RL1251 |
| p1393 | GCAGGCATGCAAGCTTTCTTCGAGGACCAGGGCAT | Antisense primer for PCR amplification of a internal region of RL1251 |
| p1394 | CGTCAGGCGGGCTTTTCTTG | Mapping primer for PCR amplification of RL1251 |
| p1395 | TGATTACGCCAAGCTGCACCCGTTGCGAAGGCCAG | Sense primer for PCR amplification of a internal region of RL0996 |
| p1396 | GCAGGCATGCAAGCTCTCATGGCTTTCGTGCAGGG | Antisense primer for PCR amplification of a internal region of RL0996 |
| p1397 | GTGCTTCTTGTAGTAGTCGG | Mapping primer for PCR amplification of RL0996 |
| p1398 | TGATTACGCCAAGCTACCAGGGAAGCCCAGCACAG | Sense primer for PCR amplification of a internal region of pRL80026 |
| p1399 | GCAGGCATGCAAGCTGGTGTGATAGTTGACGACCC | Antisense primer for PCR amplification of a internal region of pRL80026 |
| p1400 | AGCGAGGGAAGTGTCGCCAA | Mapping primer for PCR amplification of pRL80026 |
| p1401 | TGATTACGCCAAGCTCTGCGGCACCCTTGGAGCTC | Sense primer for PCR amplification of a internal region of RL3424 |
| p1402 | GCAGGCATGCAAGCTGCATCGTCGGTTTCCTGACC | Antisense primer for PCR amplification of a internal region of RL3424 |
| p1403 | CGGCTGAGTGGAACGAGAAC | Mapping primer for PCR amplification of RL3424 |
| p1404 | TGATTACGCCAAGCTCCGACCGCAACATCGCCATG | Sense primer for PCR amplification of a internal region of RL2418 |
| p1405 | GCAGGCATGCAAGCTACAAAGACGAATTCCGGGCG | Antisense primer for PCR amplification of a internal region of RL2418 |

| | | |
|---|---|---|
| p1406 | CGGATATTGGAGCGTTCGCC | Mapping primer for PCR amplification of RL2418 |
| p1407 | TGATTACGCCAAGCTCTGCCA AGCCACCTGCCGGA | Sense primer for PCR amplification of a internal region of RL0540 |
| p1408 | GCAGGCATGCAAGCTCGATGT GCGATTTCAGCTCC | Antisense primer for PCR amplification of a internal region of RL0540 |
| p1409 | ATTCGTCCACCCATACCGGC | Mapping primer for PCR amplification of RL0540 |
| p1410 | TGATTACGCCAAGCTGCGCCG CTCAACTCCGTCGA | Sense primer for PCR amplification of a internal region of RL4274 |
| p1411 | GCAGGCATGCAAGCTAATAGA CCGGGCAGCATGGC | Antisense primer for PCR amplification of a internal region of RL4274 |
| p1412 | CGTCTGTTGGCTTTCGCTGC | Mapping primer for PCR amplification of RL4274 |
| p1413 | TGATTACGCCAAGCTGCTGCCT CCACGCCATATTC | Sense primer for PCR amplification of a internal region of pRL120500 |
| p1414 | GCAGGCATGCAAGCTAACGGG TCGAACGAATGTCC | Antisense primer for PCR amplification of a internal region of pRL120500 |
| p1415 | ACGCTCCCGCCATCATGAAG | Mapping primer for PCR amplification of pRL120500 |
| p1416 | TGATTACGCCAAGCTGCGACA AGGCGACATCAACT | Sense primer for PCR amplification of a internal region of RL1863 |
| p1417 | GCAGGCATGCAAGCTGTGGCG TCGTAAATATCCTC | Antisense primer for PCR amplification of a internal region of RL1863 |
| p1418 | ATCGGCGGGCATTGTCTCTG | Mapping primer for PCR amplification of RL1863 |
| p1419 | TGATTACGCCAAGCTGCGCCA AGCTGCGGGATGTC | Sense primer for PCR amplification of a internal region of pRL90085 |
| p1420 | GCAGGCATGCAAGCTAGGAGG TAAAGCGCCTCGGT | Antisense primer for PCR amplification of a internal region of pRL90085 |
| p1421 | AGACCGGCCAGCAGCTTGAT | Mapping primer for PCR amplification of pRL90085 |
| p1422 | TGATTACGCCAAGCTGACCGG CTCTGTTCCGCCTT | Sense primer for PCR amplification of a internal region of RL2711 |
| p1423 | GCAGGCATGCAAGCTCAGCAA | Antisense primer for PCR |

| | | |
|---|---|---|
| | GAGTGAACTGGTCT | amplification of a internal region of RL2711 |
| p1424 | ATGTCGTCGAAGCCGATCCG | Mapping primer for PCR amplification of RL2711 |
| p1425 | TGATTACGCCAAGCTGCCGCT GTCGAAAATGTCGC | Sense primer for PCR amplification of a internal region of pRL120479 |
| p1426 | GCAGGCATGCAAGCTATGTCC TTGGGAACCGGCTT | Antisense primer for PCR amplification of a internal region of pRL120479 |
| p1427 | CATGGGCCGGGATATCGAAT | Mapping primer for PCR amplification of pRL120479 |
| p1428 | TGATTACGCCAAGCTCGTTCCA CGCGCCGATAAAG | Sense primer for PCR amplification of a internal region of RL3860 |
| p1429 | GCAGGCATGCAAGCTATGCAA TTCCCGCCGCAATG | Antisense primer for PCR amplification of a internal region of RL3860 |
| p1430 | ATCTCGTCCCTTGGCAGTTT | Mapping primer for PCR amplification of RL3860 |
| p1431 | TGATTACGCCAAGCTGCCCGA GCTTCTGGTGGATG | Sense primer for PCR amplification of a internal region of pRL80023 |
| p1432 | GCAGGCATGCAAGCTGCGATG GTTGCACTACTGCA | Antisense primer for PCR amplification of a internal region of pRL80023 |
| p1433 | AGGAGAAGGTGAAGGCGGCA | Mapping primer for PCR amplification of pRL80023 |
| p1434 | TGATTACGCCAAGCTCTGCGTC CGTTGCCGACAGG | Sense primer for PCR amplification of a internal region of RL4211 |
| p1435 | GCAGGCATGCAAGCTGACCTT ATAGCCGTAGCGCC | Antisense primer for PCR amplification of a internal region of RL4211 |
| p1436 | TCGTGAGCCAGGGAAAGAGA | Mapping primer for PCR amplification of RL4211 |
| p1437 | TGATTACGCCAAGCTGTGCGG AGCACGCTCAGCAG | Sense primer for PCR amplification of a internal region of RL1860 |
| p1438 | GCAGGCATGCAAGCTGCCGCG AATATCTGGATGGC | Antisense primer for PCR amplification of a internal region of RL1860 |
| p1439 | TCCAAGATGTCTCGGCGCTG | Mapping primer for PCR amplification of RL1860 |
| p1440 | TGATTACGCCAAGCTCCTGTCA GATAGATCTCCGG | Sense primer for PCR amplification of a internal |

| | | |
|---|---|---|
| | | region of RL2259 |
| p1441 | GCAGGCATGCAAGCTAAAGTA GTCGAGAGCCCTAG | Antisense primer for PCR amplification of a internal region of RL2259 |
| p1442 | TTCGGCAGAGAGCATCGTCG | Mapping primer for PCR amplification of RL2259 |
| p1443 | TGATTACGCCAAGCTCTTACCG TTCGCCCATGCGC | Sense primer for PCR amplification of a internal region of RL1297 |
| p1444 | GCAGGCATGCAAGCTATTGCC GTGCAGGCGGTAGA | Antisense primer for PCR amplification of a internal region of RL1297 |
| p1445 | GGCGTCATCCAGTGGAAGCA | Mapping primer for PCR amplification of RL1297 |
| p1446 | TGATTACGCCAAGCTGAATGA AATCCCCGCAGCCG | Sense primer for PCR amplification of a internal region of pRL110423 |
| p1447 | GCAGGCATGCAAGCTGAGCAG CAGCGTCAAAGTAT | Antisense primer for PCR amplification of a internal region of pRL110423 |
| p1448 | TCGGCATTGTAATGATCACG | Mapping primer for PCR amplification of pRL110423 |
| p1449 | TGATTACGCCAAGCTGCTTGTT GCCCGTGGAAACC | Sense primer for PCR amplification of a internal region of RL3982 |
| p1450 | GCAGGCATGCAAGCTCGATCA TGGCAAGTCTCAGT | Antisense primer for PCR amplification of a internal region of RL3982 |
| p1451 | CCATCTCGGCCACCTGCACT | Mapping primer for PCR amplification of RL3982 |
| p1452 | TGATTACGCCAAGCTGCTGCCT TGCCGCTTCTGTT | Sense primer for PCR amplification of a internal region of RL0787 |
| p1453 | GCAGGCATGCAAGCTCAGATC GCCTTGATTGACTT | Antisense primer for PCR amplification of a internal region of RL0787 |
| p1454 | TGGAGGAGACCGTCACGACA | Mapping primer for PCR amplification of RL0787 |
| p1455 | TGATTACGCCAAGCTCTAGCT GGGAGAGGGTCTTG | Sense primer for PCR amplification of a internal region of RL0274 |
| p1456 | GCAGGCATGCAAGCTGGAAGA TCTCTTCTACGACA | Antisense primer for PCR amplification of a internal region of RL0274 |
| p1457 | TCAGGGTCTCGCCCGATCTG | Mapping primer for PCR amplification of RL0274 |

| | | |
|---|---|---|
| p1458 | TGATTACGCCAAGCTGGCTTC GGGCCGTCGTCGAA | Sense primer for PCR amplification of a internal region of RL4265 |
| p1459 | GCAGGCATGCAAGCTCTCGCG GTTCACTTGGGCCT | Antisense primer for PCR amplification of a internal region of RL4265 |
| p1460 | TCAGTTGCGGAACGATGGCA | Mapping primer for PCR amplification of RL4265 |
| p1461 | TGATTACGCCAAGCTCCAGCG AGTCAGGCTGAGGT | Sense primer for PCR amplification of a internal region of pRL80021 |
| p1462 | GCAGGCATGCAAGCTGATCTT GCCGTTGACGTTGG | Antisense primer for PCR amplification of a internal region of pRL80021 |
| p1463 | TGGAGCAGATAAAGGAAGCG | Mapping primer for PCR amplification of pRL80021 |
| p1464 | TGATTACGCCAAGCTGAATGC ATCGGTATGTTCGC | Sense primer for PCR amplification of a internal region of pRL90055 |
| p1465 | GCAGGCATGCAAGCTGCAAGG AGACTCTCATGAAA | Antisense primer for PCR amplification of a internal region of pRL90055 |
| p1466 | GCGGGCGTGATGTGAAGGTT | Mapping primer for PCR amplification of pRL90055 |
| p1467 | TGATTACGCCAAGCTGGTCCG TTGTTGTTCGGAGC | Sense primer for PCR amplification of a internal region of pRL80054 |
| p1468 | GCAGGCATGCAAGCTAAAACG CTCACCGGCACATG | Antisense primer for PCR amplification of a internal region of pRL80054 |
| p1469 | TCCAGTTCGCAACTCCGACG | Mapping primer for PCR amplification of pRL80054 |
| p1473 | TGATTACGCCAAGCTGGTACC CGCCATCTTTCGCA | Sense primer for PCR amplification of a internal region of pRL110268 |
| p1474 | GCAGGCATGCAAGCTGACCTG CAAGCGCCTTGAAA | Antisense primer for PCR amplification of a internal region of pRL110268 |
| p1475 | GCGAACTGATCGCCATTCGA | Mapping primer for PCR amplification of pRL110268 |
| p1478 | TGATTACGCCAAGCTCGGTGT CCCGCATTGCCACC | Sense primer for PCR amplification of a internal region of RL2469 |
| p1479 | GCAGGCATGCAAGCTGCGTGC CGGTTGCGAGAATA | Antisense primer for PCR amplification of a internal region of RL2469 |

| | | |
|---|---|---|
| p1480 | GCAAACCCCGCAAGCAACAC | Mapping primer for PCR amplification of RL2469 |
| p1545 | TGATTACGCCAAGCTCCTGAC GCCCGCCGAACACA | Sense primer for PCR amplification of a internal region of pRL100444 |
| p1546 | GCAGGCATGCAAGCTATGTTC CAGGCATCCCTGCG | Antisense primer for PCR amplification of a internal region of pRL100444 |
| p1547 | GGCGGGTTCCTTTGCGGTAA | Mapping primer for PCR amplification of pRL100444 |
| p1548 | TGATTACGCCAAGCTGGCAAT CGGTTTAGCCCGGA | Sense primer for PCR amplification of a internal region of RL1172 |
| p1549 | GCAGGCATGCAAGCTATGAAG CCCGCGATCAGATA | Antisense primer for PCR amplification of a internal region of RL1172 |
| p1550 | CAACGAAATGCCAGGCGAGC | Mapping primer for PCR amplification of RL1172 |
| p1551 | TGATTACGCCAAGCTCAGAAC GCAGCCTTGCTGCT | Sense primer for PCR amplification of a internal region of RL2946 |
| p1552 | GCAGGCATGCAAGCTTTATCG CTCTGGAGCTCGAG | Antisense primer for PCR amplification of a internal region of RL2946 |
| p1553 | CGTGGTGAATGCCTGCGGTA | Mapping primer for PCR amplification of RL2946 |
| p1557 | TGATTACGCCAAGCTGTCGTTC GATAGGACCCGCC | Sense primer for PCR amplification of a internal region of pRL120724 |
| p1558 | GCAGGCATGCAAGCTATCCGA CGCAGATGAACAGC | Antisense primer for PCR amplification of a internal region of pRL120724 |
| p1559 | TATCCCGTCCTGGCTTTGGC | Mapping primer for PCR amplification of pRL120274 |
| p1560 | TGATTACGCCAAGCTGCCAAC GACCCGAACCGCCC | Sense primer for PCR amplification of a internal region of RL3186 |
| p1561 | GCAGGCATGCAAGCTAGCGCT GTTGTCGGTCGGAG | Antisense primer for PCR amplification of a internal region of RL3186 |
| p1562 | CGATGAGGCGGCTGAGCTTT | Mapping primer for PCR amplification of RL3186 |
| p1636 | TGATTACGCCAAGCTGGTTGA AAGTGCAGCGACCG | Sense primer for PCR amplification of a internal region of pRL110443 |
| p1637 | GCAGGCATGCAAGCTCTGTCG | Antisense primer for PCR |

| | | |
|---|---|---|
| | ACGCAATGGATCGA | amplification of a internal region of pRL110443 |
| p1638 | TTGTGATCGGCCTTTGTCGG | Mapping primer for PCR amplification of pRL110443 |
| p1639 | TGATTACGCCAAGCTGGCATTTGATGCAATGACCG | Sense primer for PCR amplification of a internal region of RL1495 |
| p1640 | GCAGGCATGCAAGCTGTGATCTTCGACATCGATTA | Antisense primer for PCR amplification of a internal region of RL1495 |
| p1641 | CTAAGTCTACTGATGTCCGC | Mapping primer for PCR amplification of RL1495 |
| p1208 | TGATTACGCCAAGCTTTGAAGTTCGGCAAGCAT | Sense primer for PCR amplification of a internal region of nifH |
| p1209 | GCAGGCATGCAAGCTAGTCCACTACATCCCAAA | Antisense primer for PCR amplification of a internal region of nifH |
| p1210 | TGTCACCGCCGAAAACGATG | Mapping primer for PCR amplification of nifH |
| p1214 | TGATTACGCCAAGCTGTCGAGCCAGTTGCCGTG | Sense primer for PCR amplification of a internal region of RL0913 |
| p1215 | GCAGGCATGCAAGCTTGTTTGCCACGGCTTTCG | Antisense primer for PCR amplification of a internal region of RL0913 |
| p1216 | GTCGGCTTCGATCCGCATCG | Mapping primer for PCR amplification of RL0913 |
| p1071 | TTTTAAGCTTAGGTGTGGCCATTGTCGC | Sense primer for PCR amplification of a internal region of RL3016 |
| p1072 | TTTTCTAGATAGTCGGGATGCTCGCCA | Antisense primer for PCR amplification of a internal region of RL3016 |
| p1095 | TTTTAAGCTTGCTACCTGATCACCTGGA | Sense primer for PCR amplification of a internal region of pRL120632 |
| p1096 | TTTTCTAGATTGATGTTGGCGATGACG | Antisense primer for PCR amplification of a internal region of pRL120632 |
| p1097 | TTTTAAGCTTAGCCAGAACCAGATCGTC | Sense primer for PCR amplification of a internal region of RL0680 |
| p1098 | TTTTCTAGAGGCCTTCAGCTTGGTCAC | Antisense primer for PCR amplification of a internal |

| | | |
|---|---|---|
| | | region of RL0680 |
| p1101 | TTTTAAGCTTGCTCCTGGGGCG ATATCA | Sense primer for PCR amplification of a internal region of RL1694 |
| p1102 | TTTTCTAGAGACAGCGGCGCA AGGTTG | Antisense primer for PCR amplification of a internal region of RL1694 |
| BD pckA for | TGATTACGCCAAGCTTCGCTGT TCATCCGCAATCT | Sense primer for PCR amplification of a internal region of RL0037 |
| BD pckA rev | GCAGGCATGCAAGCTGTTTTC AGTCAGCGACCCGT | Antisense primer for PCR amplification of a internal region of RL0037 |
| p1093 | TTTTAAGCTTAGCAGTTCAATC CCACGGC | Sense primer for PCR amplification of a internal region of pRL110199 |
| p1094 | TTTTCTAGAATCAGCATCGCAG GCAGC | Antisense primer for PCR amplification of a internal region of pRL110199 |
| p1223 | TGATTACGCCAAGCTGTAGCC GCCGGCTTCGTG | Sense primer for PCR amplification of a internal region of RL3130 |
| p1224 | GCAGGCATGCAAGCTAAGACC GCGCCCGACGAT | Antisense primer for PCR amplification of a internal region of RL3130 |
| p1225 | ATCCACACGTCTCTGCTATC | Mapping primer for PCR amplification of RL3130 |

**Table 2.2.1.** List of PCR primers used in this study.

### 2.2.5 Polymerase Chain Reaction

Genes of interest were amplified from Rlv3841 genomic DNA using GoTaq Green master mix (Promega). PCR was carried out in 25 µl reactions containing GoTaq Green master mix (1X), 50 ng of genomic DNA and 10 pmols of each sense and antisense primer. The PCR was performed in a Px2 thermal cycler (Thermo Hybaid) with the following cycling conditions; 95ºC for 2 min, followed by 30 cycles of 95ºC for 45 seconds, 57ºC for 45 seconds, 72ºC for 3 min and a final extension at 72ºC for 10 min. Colony PCR was carried out in 10 µl reactions containing GoTaq Green master mix (1X), 10 pmols of each M13 universal and reverse primers and pinprick sized colony transferred using a tooth

pick to the PCR mixture. The reaction was performed in the Px2 thermal cycler (Thermo Hybaid) with the following cycling conditions; 95ºC for 10 min, followed by 30 cycles of 95ºC for 1 min, 57ºC for 1 min, 72ºC for 2 min and a final extension at 72ºC for 10 min.

### 2.2.6 Conjugation

The plasmid DNA from *E.coli* was mobilized into Rlv3841 or Rlv300 by the tri-parental conjugation method with the conjugal helper plasmid pRK2013 in *E.coli*. The helper plasmid pRK2013 provides *trans*-acting factors required to facilitate the mobilization and conjugal transfer of the unrelated plasmids (Figurski & Helinski, 1979). *E.coli* containing the donor and helper (pRK2013) plasmids were grown separately in LB broth with antibiotics in an orbital shaker (225 rpm) at 37ºC overnight. 200 µl - 500 µl of cultures were sub-cultured into 10 ml of fresh LB broth with antibiotic at 37ºC in an orbital shaker (100 rpm) for 4 hrs to promote the formation of sex pili. 1 ml of each donor and helper cultures were harvested and washed thrice with TY broth by centrifuging at 6000 rpm, for 20 min and resuspended in 1 ml of fresh TY broth, to remove traces of antibiotics. Freshly grown recipient, Rlv3841 or Rlv300, on a TY slope was washed off with 5 ml of TY broth. 400 µl of donor plasmid, 400 µl of recipient strain and 200 µl of helper plasmid were mixed gently and spun down and resuspended in 30 µl of TY broth, transferred onto a sterile nitrocellulose membrane on a TY agar plate and incubated at 26ºC for overnight. Next day, the mat growth on the nitrocellulose membrane was resuspended in TY broth. An aliquot of the suspension was plated on TY agar containing antibiotics and incubated at 26ºC for 4 days, and the remaining suspension was glycerol stocked as described in section 2.1.1.

### 2.2.7 pK19mob mutagenesis

pK19mob (Schafer *et al.*, 1994) was used to generate integration mutants (Prell *et al.*, 2002). High-throughput cloning was achieved by using the In-Fusion dry down PCR cloning kit (Clontech) as described in section 2.2.3 and the pK19mob derivative was then mobilized into Rlv300 by tri-parental conjugation

as described in section 2.2.6. The conjugation mix was selected on TY agar containing appropriate antibiotics at 26ºC. The trans-conjugants were confirmed by mapping the integrated gene sequence on the chromosome by PCR with mapping primer and pK19 specific primers pK19A/B (Figure 2.1.1).

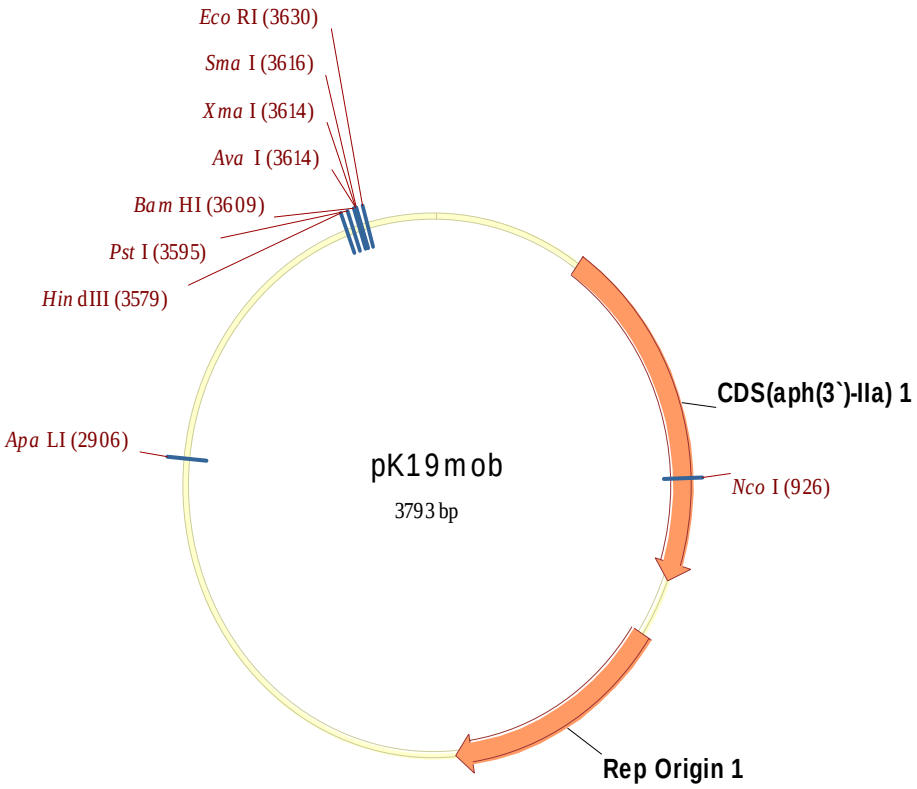



**Figure 2.2.1.** Plasmid map of pK19mob.

### 2.2.8 Isolation of Total RNA

An RNase free working environment was maintained by wiping the work place and the equipments with RNaseZap, an RNase decontamination solution (Ambion). Total RNA was extracted from free living Rlv3841, grown as described in section 2.1.1, by adding 12 ml of culture ($OD_{600nm}$ 0.4-0.6) to 24 ml of RNAprotect, RNA stabilization reagent (Qiagen), vortexed and incubated at room temperature for 5 min. The cells were harvested by centrifugation at 10,000 rpm for 10 min at 4ºC in a Sorvall centrifuge (Rotor-SS34). The cells were then resuspended in 200 µl of 10 mM Tris-Cl pH 8.0 and added to ice-cold RLT buffer (Qiagen) with β-mercaptoethanol (Sigma) in a FastPrep tube containing lysing matrix B (QBIOgene). The cells were lysed in the FastPrep FP120 ribolyser at level 6.5 for 30 seconds (QBIOgene). The tubes were then incubated in ice for 3 min and centrifuged at 13,000 rpm for 3 min at 4ºC. Total RNA was purified using RNeasy mini kit including on-column digestion with RNase free DNaseI (30 Kunitz units) (Qiagen) as per manufacturer's instructions.

For extracting total RNA from rhizosphere bacteria, pea (*P. sativum* cv Avola*)* or alfalfa (*M. sativa)* or sugar-beet (*Beta vulgaris)* seeds were surface sterilized as described in section 2.3.1 and grown in a tube autoclaved with 25 ml of washed medium grade vermiculite (Vermiperl) supplemented with 10 ml of nitrogen free rooting solution at 23ºC with a 16-h/8-h light/dark in a growth room for 7, 14 or 21 days depending upon the experiment. After the initial growth period, seedlings were inoculated with $10^8$ cfu of Rlv3841 grown in AMS as described in section 2.1.1, harvested at $OD_{600nm}$ 0.4-0.6 and washed (4000 rpm, 20 min) thrice and resuspended in sterile nitrogen free rooting solution. The rhizosphere tubes were incubated in the growth room for a further 1, 3 or 7 days post inoculation depending upon the experiment. After the stipulated time period, the bacterial cells were recovered from the rhizosphere by adding 6 ml of sterile water and 12 ml of RNAprotect (Qiagen) after removing the shoots. The mixture was vortexed to ensure the complete mixing of the vermiculite, water and RNAprotect. The mixture was filtered through four layers of sterile muslin cloth and the cloudy filtrate was spun down at 1,000 rpm for 1 min at 4ºC (to pellet the

vermiculite fine particles). The supernatant was removed and further spun down at 10,000 rpm for 10 min at 4ºC in a Sorvall centrifuge (Rotor-SS34) to pellet the bacterial cells.

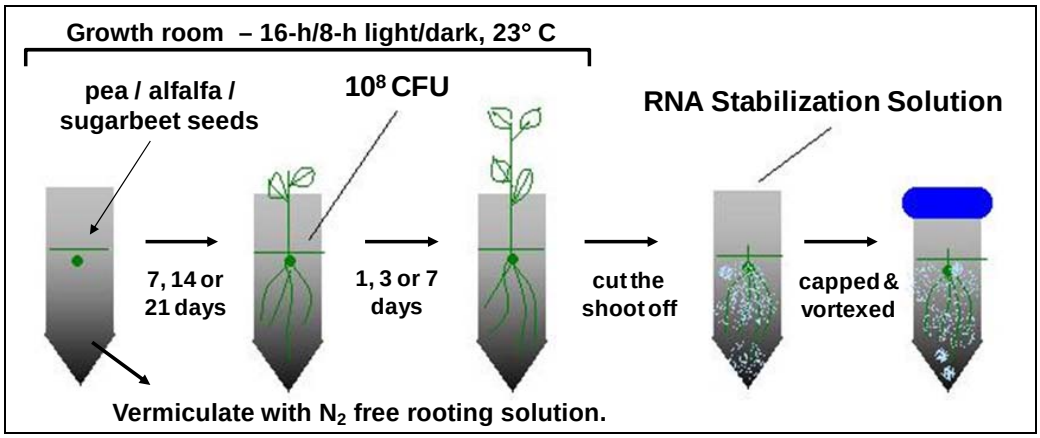



**Figure 2.2.2.** Recovery of bacterial cells from the sterile pea rhizosphere

The cells were then resuspended in 10 mM Tris-Cl pH 8.0 and added to ice-cold RLT buffer (Qiagen) with β-mercaptoethanol (Sigma) in a chilled FastPrep tube containing lysing matrix B (QBIOgene). The cells were lysed in the FastPrep FP120 ribolyser (QBIOgene) at level 6.5 for 30 seconds. The tubes were incubated in ice for 3 min and centrifuged at 13,000 rpm for 3 min at 4ºC. Total RNA was purified using an RNeasy mini kit including on-column digestion with RNase free DNaseI (30 Kunitz units) (Qiagen) as per manufacturer's instructions.

### 2.2.9 Quantification of RNA

The quality and quantity of the isolated total RNA was determined by using an Experion RNA StdSens analysis (25 ng - 500 ng) kit in the Experion microfluidic RNA analyzer (Bio-Rad laboratories). The RNA gel matrix (600 µl) was centrifuged at 4,000 rpm for 10 min in a spin filter, and the filtrate was aliquoted (65 µl) in RNase free microfuge tubes. RNA dye concentrate (1 µl) was added to the RNA gel matrix (65 µl), vortexed and centrifuged at 13,000 rpm for 1 min. 1.5 µl of samples and 1 µl (160 ng/µl) of RNA ladder (supplied with the kit)

were denatured at 70ºC for 2 min, and chilled immediately. The Experion RNA analyzer was warmed up and the electrodes were decontaminated by electrode cleaner and washed with RNase free water. The RNA chip was primed by loading the gel matrix (G) and gel-dye matrix (GS) on the specified wells in the chip and then in the Experion priming station with pressure at B for 30 seconds. The denatured samples and the RNA ladder were loaded onto specified wells in the chip. The samples in the chip were vortexed for one min in the Experion vortex station and read in the Experion RNA analyzer. The quality of the RNA was determined by the intact 16S and 23S rRNA in the electrophoregrams. The quantity of the RNA was mathematically calculated from the fluorescence intensities by the Experion software.

**2.2.10 RNA Amplification:**

The SenseAmp procedure employs a unique RNA amplification technology to amplify nearly identical copies of the original RNA molecules in sense orientation. Most of the commercially available linear RNA amplification procedures amplify RNA in antisense orientation which requires incorporation of modified ribonucleotides, which is often an inefficient method of labeling RNA. In SenseAmp, the cDNA is synthesized from the 3' end of RNA and the T7 amplification from the 3' end of the cDNA (5' end of RNA), whereas in conventional amplification both priming and amplification is done at 3' end of RNA, leading to a 3' bias. Senseamp can amplify as low as 25 ng of total RNA sample, with reliable consistency and it can also efficiently amplify partially degraded RNA samples (Goff *et al.*, 2004). A recent study, successfully used SenseAmp to amplify mRNA isolated from bacteria sampled from nasopharynx, lungs and bloodstream of mice to perform gene expression studies by qRT-PCR (LeMessurier *et al.*, 2006).

The mRNA was amplified using the SenseAmp RNA amplification kit (Genisphere) as described in the Figure 2.2.2. 250 ng of total RNA was reverse transcribed by Superscript II (Invitrogen) using 500 ng of random nonamer reverse transcription primers at 42ºC for 2 hrs. The cDNA was purified with the

MinElute PCR purification kit (Qiagen) and heated at 80ºC for 10 min and chilled for 2 min. The cDNA was tailed in the 3' end with dTTP using Terminal Deoxynucleotidyl Transferase at 37ºC for 3 min. The reaction was stopped by heating at 80ºC for 10 min and a T7 template oligo was annealed to the 3' end of the tailed cDNA at 37ºC for 10 min. Klenow enzyme was added to fill the 3' end of the first strand of the cDNA to produce a double stranded T7 promoter at room temperature for 30 min. The reaction was stopped by heating at 65ºC for 10 min and chilled immediately. Half of the reaction was used for *in vitro* transcription, the strands were re-annealed at 37ºC for 10 min and subsequently incubated with T7 nucleotide mix and T7 enzyme mix at 37ºC for 16 hrs in a heated lid thermal cycler. The amplified RNA was purified by RNeasy mini kit (Qiagen) as per the RNA cleanup protocol. The concentrations of the amplified RNA were determined by Experion microfluidic RNA analyzer as described in section 2.2.9.

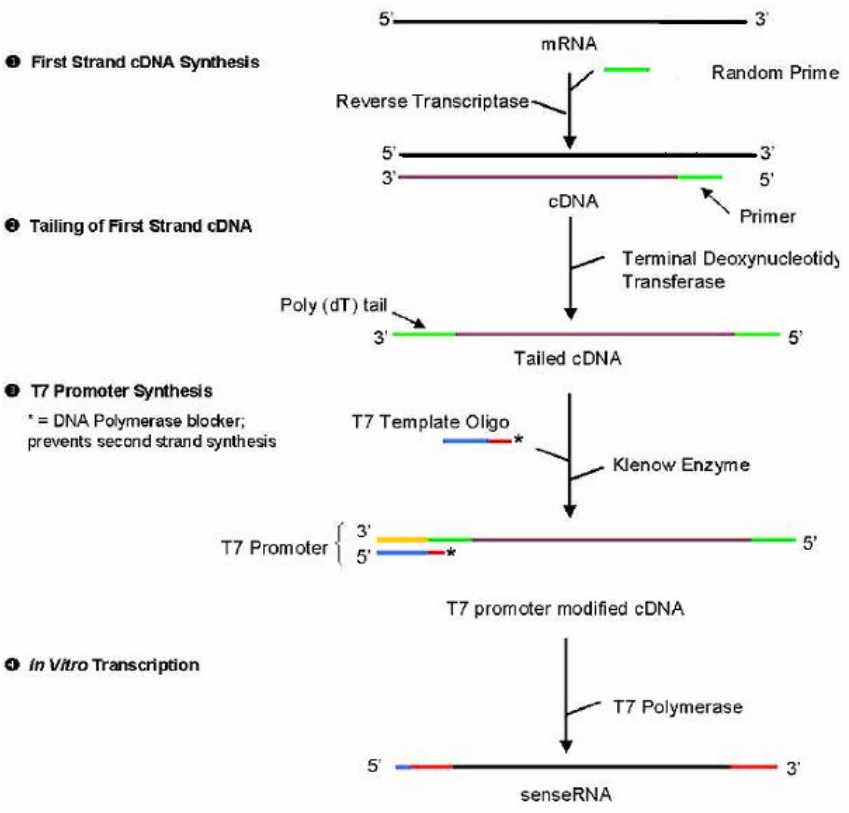



**Figure 2.2.3.** Procedure for RNA amplification using Genisphere SenseAmp (adapted from Genisphere SenseAmp protocol).

### 2.2.11 DNase treatment with Turbo DNaseI

The chromosomal DNA contamination was eliminated by incubating with TURBO DNase (2 U/µl) (Ambion) in a water bath at 37ºC for 30 min. The excess DNase was inactivated and removed by incubating with DNase Inactivation reagent at room temperature for 2 min and centrifuging at 11,000 rpm for 2 min. The supernatant was transferred to a fresh RNase-free Eppendorf tube without disturbing the pellet and stored at -80ºC.

### 2.2.12 Quantitative RT-PCR analysis:

Reverse transcriptase PCR is a more sensitive method for mRNA detection than Northern blot or RNase protection assay. RT-PCR can detect the expression of mRNA from a single cell. qRT-PCR is widely used as a validation tool for microarrays. The detection of the amplified products can be done by many methods including Taqman probe, Molecular Beacon, Scorpions and SYBR Green I. The methods employ Fluorescence Resonance Energy Transfer (FRET), except SYBR Green I. SYBR Green I work basically by the increase in its fluorescence, when it binds to double stranded DNA. The excitation wavelength is 494 nm and emission wavelength is 521 nm. The intensity of its fluorescence increases by 40 times when it binds to dsDNA. The advantage is it is inexpensive and easy to use; the disadvantage is it can bind to any dsDNA including primer dimers. So, it is extremely good with gene specific primers (GSP).

The data analysis can be done by an absolute and relative quantification method. An absolute quantification method is used to quantitate unknown samples by interpolating their quantity from a standard curve, whereas relative quantification method is used to analyse changes in gene expression in a given sample relative to another reference sample. The data can be analysed by two ways in relative quantification method, either by standard curve method or by Comparative $C_T$ Method ($\Delta\Delta C_T$). Comparative $C_T$ Method ($\Delta\Delta C_T$) calculates the fold expression by comparing the target gene with respect to endogenous/reference gene in treated sample and normalises to an untreated sample. The method do not require a standard curve, but both the target gene and

endogenous /reference gene should express equally in the control sample or in nature (Bustin, 2000; Bustin, 2002). The amount of target normalized to an endogenous/reference gene, relative to a calibrator is given by:

**Calculation of ΔCT value: (for both treated and untreated samples)**

ΔCT = CT target - CT reference/endogenous.

**Calculation of ΔΔCT value:**

ΔΔCT = ΔCT test/treated - ΔCT calibrator/untreated sample.

**Calculation of Fold expression:**

$$2^{-\Delta\Delta CT}$$

Microarray datasets were validated by performing quantitative reverse transcription PCR based on SYBR Green chemistry for 16 randomly selected differentially expressed genes including a reference gene with a QuantiTect SYBR Green RT-PCR (Qiagen) kit. RNA was extracted from laboratory grown cells and cells harvested from 7 day old pea rhizosphere with a 7 day post inoculation. RNA samples were treated with TURBO DNase (Ambion) to eliminate the chromosomal DNA contamination as described in section 2.2.11. The lack of signals in both the no reverse transcriptase and no template controls confirmed the fidelity of the PCR reaction. The qRT-PCR was performed with 200 ng of DNA free total RNA, forward and reverse primers (12.5 pmoles), QuantiTect SYBR green master mix and QuantiTect RT mix in a 25 µl reaction. The cycling conditions were as follows: reverse transcription at 50ºC for 30 min, Taq DNA polymerase activation at 95ºC for 15 min followed by 30 cycles of denaturation at 95ºC for 15 seconds, annealing at 58ºC for 30 seconds and extension at 72ºC for 15 seconds. The data were analyzed and the fold expression was calculated by the relative quantification method (comparative cycle threshold method) (Bustin, 2000; Bustin, 2002). The qRT-PCR primers used in this study are shown in Table 2.2.1.

| Name | Sequence 5'- 3' | Primer binding site |
|---|---|---|
| p1594 | CGATGACGAACACGGCTGGA | Sense primer for qRT-PCR amplification of RL0037 |
| p1595 | GTACCGAAACGCTGCGTTGT | Antisense primer for qRT-PCR amplification of RL0037 |
| p1596 | ATCCTCATGAAGATTGCCCA | Sense primer for qRT-PCR amplification of RL4267 |
| p1597 | CGAAATAACGGAAATGGTCG | Antisense primer for qRT-PCR amplification of RL4267 |
| p1598 | TCGACTACCAGGACGGCATG | Sense primer for qRT-PCR amplification of RL4274 |
| p1599 | CCTGAGATGCCTGTTGCGAG | Antisense primer for qRT-PCR amplification of RL4274 |
| p1600 | GGCACGCCTTCCACACCAAA | Sense primer for qRT-PCR amplification of RL1860 |
| p1601 | TGCTTTCCGATACCGCCTGC | Antisense primer for qRT-PCR amplification of RL1860 |
| p1604 | TGCCCGCTTCTATCTTCAAC | Sense primer for qRT-PCR amplification of RL1251 |
| p1605 | CCAACAACCATGTCATCGAC | Antisense primer for qRT-PCR amplification of RL1251 |
| p1606 | TCTTGGCCGAAAGGTGGAGC | Sense primer for qRT-PCR amplification of pRL80026 |
| p1607 | AGGGCGCAATCAGAAGGTCG | Antisense primer for qRT-PCR amplification of pRL80026 |
| p1608 | AGGGCATTGGCATCACGGAG | Sense primer for qRT-PCR amplification of RL1925 |
| p1609 | AACCACGGGACCGTTCGACA | Antisense primer for qRT-PCR amplification of RL1925 |
| p1610 | GCCGTCAGGGCATTGAGGAA | Sense primer for qRT-PCR amplification of RL3424 |
| p1611 | TTGCCGATGGCGACATTCTG | Antisense primer for qRT-PCR amplification of RL3424 |
| p1612 | AGGCGGACAACGATTTCAAC | Sense primer for qRT-PCR amplification of pRL100187 |
| p1613 | CGAAAGAACGATTGGGTCCT | Antisense primer for qRT-PCR amplification of pRL100187 |
| p1614 | AGACGAAATTATAGGCGAGC | Sense primer for qRT-PCR amplification of pRL100169 |

| | | |
|---|---|---|
| p1615 | AACCTGTTGCCAATGTCTTC | Antisense primer for qRT-PCR amplification of pRL100169 |
| p1616 | CGAAGCCGCCATCAAGAAGG | Sense primer for qRT-PCR amplification of RL3624 |
| p1617 | CCTGCGAGATCAGCGATTCG | Antisense primer for qRT-PCR amplification of RL3624 |
| p1618 | ACGGCGAATCGTTCCAGGAG | Sense primer for qRT-PCR amplification of pRL100451 |
| p1619 | TGACCTGGGTTACGCCTTGG | Antisense primer for qRT-PCR amplification of pRL100451 |
| p1620 | TACCGCAATACGCTACTTTG | Sense primer for qRT-PCR amplification of RL2937 |
| p1621 | AACGGCTATGAGAATGACAG | Antisense primer for qRT-PCR amplification of RL2937 |
| p1622 | TATGCCTCAGCGACGGATGG | Sense primer for qRT-PCR amplification of RL2164 |
| p1623 | GCTGGCATCGGTTCGGAAGA | Antisense primer for qRT-PCR amplification of RL2164 |
| p1624 | CGCTATTCCGCTGGTGCTCA | Sense primer for qRT-PCR amplification of RL0644 |
| p1625 | CGCCGATATAGGTGCCTGCA | Antisense primer for qRT-PCR amplification of RL0644 |
| p1626 | GGTCACATCCATCATCAGGC | Sense primer for qRT-PCR amplification of pRL120336 |
| p1627 | GTTGATCAGGATGTCGACCG | Antisense primer for qRT-PCR amplification of pRL120336 |
| p1634 | ACACTCTGGAAGGCTGGAAG | Sense primer for qRT-PCR amplification of RL0009 |
| p1635 | AGAAGCGGAAGCAGGTCTTC | Antisense primer for qRT-PCR amplification of RL0009 |
| p1476 | CGCCGACCTTGTTGAGAATG | Sense primer for qRT-PCR amplification of RL0996 |
| p1477 | TGACGATGAACCAGGACCTC | Antisense primer for qRT-PCR amplification of RL0996 |
| p827 | CGCCTGCATGCCGTCGATCC | Sense primer for qRT-PCR amplification of pRL110443 |
| p828 | CGCCGCAAATGTCCTGCTCG | Antisense primer for qRT-PCR amplification of pRL110443 |

**Table 2.2.2.** List of primers used for qRT-PCR in this study.

### 2.3.1 Seeds surface sterilization

Pea (*P. sativum* cv Avola) seeds were surface sterilized in 95% (v/v) ethanol for 30 seconds, followed by 5 min in 2% (v/v) sodium hypochlorite (Karunakaran *et al.*, 2006). Alfalfa (*M. sativa*) seeds were surface sterilised in 95% (v/v) ethanol for 45 seconds followed by 15 min in 5.25% (v/v) sodium hypochlorite (Gage *et al.*, 1996). Gaucho treated sugar beet (*B. vulgaris*) seeds were surface sterilised in 70% (v/v) ethanol for 2 min, followed by 2 min in 1.05% (v/v) sodium hypochlorite, the thick gaucho coating was removed before sowing (El-Tarabily, 2004). All the sterilized seeds were washed extensively with sterile distilled water before sowing. The number of seeds sown per tube was 1, 6 and 3 seeds of pea, alfalfa and sugar beet respectively.

### 2.3.2 Nitrogen-free rooting solution

Plants were grown in Nitrogen-free rooting solution containing 100 µM $KH_2PO_4$, 100 µM $Na_2HPO_4$, 1 mM $CaCl_2.2H_2O$, 100 µM KCL, 800 µM $MgSO_4.7H_2O$, 10 µM Fe-EDTA, 35 µM $H_3BO_3$, 9 µM $MnCl_2.4H_2O$, 0.8 µM $ZnCl_2$, 0.5 µM $Na_2MoO_4.2H_2O$, 0.3 µM $CuSO_4.5H_2O$ as described (Allaway et al., 2000). For rhizosphere tube preparation, rooting solution was prepared and added to the vermiculite before autoclaving. For washing the cells, the rooting solution was prepared without salts ($KH_2PO_4$, $Na_2HPO_4$) and stocks of the salts were made separately and added after autoclaving, to prevent the precipitation of salts.

### 2.3.3 Rhizosphere experiments (peas, alfalfa or sugar beet)

Pea (*P. sativum* cv Avola) or alfalfa (*M. sativa)* or sugar-beet (*B. vulgaris)* seeds were surface sterilized as described in section 2.3.1 and grown in a tube autoclaved with 25 ml of washed medium grade vermiculite (Vermiperl) supplemented with 10 ml of nitrogen-free rooting solution as described in section 2.3.2 at 23ºC with a 16-h/8-h light/dark in a growth room for 7, 14 or 21 days depending upon the experiment. After the period, seedlings were inoculated with

$10^8$ cfu of *R. leguminosarum* grown in 50 ml of AMS with 10 mM glucose and 10 mM ammonium chloride as described in section 2.1.1, harvested at $OD_{600\ nm}$ 0.4-0.6 and washed (4000 rpm, 20 min) thrice and resuspended in sterile nitrogen free rooting solution. The rhizosphere tubes were incubated in the growth room for 1, 3 or 7 days post inoculation.

### 2.3.4 Plant growth experiments

Pea seeds were surface sterilised as described in Section 2.3.1 and sown in sterilised two litre pots containing washed medium grade vermiculite (Vermiperl) with 800 ml of nitrogen-free rooting solution as described in section 2.3.2. Three pea seeds were sown and each was inoculated with 1 ml of *R. leguminosarum* strain washed off from a TY slope with 10 ml of AMS. The pots were covered with cling film (to prevent contamination) and incubated in the plant growth room at 23ºC with a 16-h/8-h light/dark period. After germination, the cling-film was pierced to let the germinated seeds grow through, and the germinated seeds were thinned from 3 to 1 and incubated in the growth room for 6 weeks.

### 2.3.5 Rhizosphere competition experiments

Competitive colonization assay (CCA) was performed to determine the biological role of the mutated gene by allowing the mutant strain to compete with the wild-type for colonization of the pea rhizosphere. To take advantage of antibiotic selection during screening, the mutants were made in Rlv300 ($str^s$, parent strain of Rlv3841). During homologous recombination, the internal region of the desired gene cloned in pK19mob was integrated into chromosome along with neomycin marker and other pK19mob sequences, conferring neomycin resistance to Rlv300. Both the strains were resistant to trimethioprim and nystatin (anti-fungal), so the mutant strain (in Rlv300) is $trim^r$, $nys^r$, $neo^r$ and $str^s$ and Rlv3841 (wt) is $trim^r$, $nys^r$, $neo^s$ and $str^r$. The competitive ability in colonizing pea rhizosphere by both the strains Rlv3841 and Rlv300 were confirmed by performing a CCA in five different cfu inoculation ratios.

For the control assays, three replicates of each inoculation ratios were performed. Pea seedlings were grown for 7 days as described in section 2.3.3 and inoculated with Rlv300 and Rlv3841 in the following cfu ratios 1000:0, 0:1000, 1000:1000, 10,000:1000 and 1000:10,000 and incubated as described in section 2.3.3. After 7 days, shoots were cut-off and 30 ml of sterile phosphate-buffered saline (PBS) was added to the roots and vortexed in Multi Reax system (Heidolph) for 30 min at speed 10. After vortexing, 1 ml of the mixture was transferred to a sterile eppendorf tube, under aseptic conditions. From the 1 ml, 660µl of sample was glycerol stocked and 100 µl was used for serial dilution. The sample was serially diluted in sterile phosphate-buffered saline to $10^{-6}$. 50 µl and 100 µl of dilution $10^{-2}$ and $10^{-3}$ were plated on TY medium containing nystatin and trimethioprim. A plate with 100-200 CFU was selected and 100 CFU's were patched onto to TY medium containing nystatin and trimethioprim with or without streptomycin.

For the mutant assays, eight replicates of assays were performed for each mutant (made in Rlv300 as described in sections 2.2.3-2.2.6) over Rlv3841 in the pea rhizosphere. The experiment was performed as described for control assay, but only with one inoculation ratio (10,000:1000 CFU). A control assay was performed with each set of mutant assays, in the same inoculation ratio to ensure that dilution did not introduce any errors. The plant growth, inoculation, recovery and serial dilution was performed as described earlier for control assays, except plating where in control assays, the cfu were patched. For the control assays (Rlv300 v Rlv3841), the serial diluted samples were plated on TY medium containing nystatin and trimethioprim, and subsequently patched on TY medium containing nystatin and trimethioprim with or without streptomycin. For the mutant assays (Rlv3841 v Rlv300 mutant), the 50 µl and 100 µl of the serial diluted samples from $10^{-2}$ and $10^{-3}$ dilution were plated on TY medium with nystatin and trimethioprim containing streptomycin or neomycin. All the plates were incubated at 26ºC for approximately three days.

### 2.4.1 *Rhizobium leguminosarum* oligo array

Unique 70-mer oligonucleotides were custom designed and synthesized for each ORF in the published genome sequence of *R. leguminosarum* strain 3841 (Young *et al.*, 2006) by Operon biotechnologies, Germany. Each oligoarray was spotted with oligonucleotides corresponding to 7344 ORF's from Rlv3841, Lucidea Controls (GE-Amersham) and Pronto buffer as negative controls (Corning). Each oligonucleotide was printed at least twice in a random pattern on UltraGAPS (Gamma Amino Propyl Silane) coated slides (Corning) in a BioRobotics Microgrid TAS arrayer at the Functional Genomics and Proteomics Laboratory, University of Birmingham, UK. After printing, the slides were dried and cross-linked in a UVC500 Ultraviolet Crosslinker (Hoefer) (600mJ/cm$^2$).

### 2.4.2 Indirect CyDye labeling of samples

Total RNA was isolated, amplified and quantified as described in section 2.2.8, 2.2.9 and 2.2.10 respectively and labelled using a Cyscribe Post-Labelling kit (GE Healthcare). For first strand cDNA synthesis, 15 µg of total or amplified RNA, 1 µl of random hexamers, 0.4 µl of 50X dNTP solution (25 mM dGTP, 25 mM dCTP, 25 mM dATP, 5 mM dTTP and 20 mM amino-allyl-UTP in 4:1 ratio), 4 µl of 5x Cyscribe buffer and 1 µl of Cyscript reverse transcriptase were mixed in a reaction volume of 20 µl and incubated at 42°C for 2 hrs. After incubation, the mRNA was degraded by the addition of 2 µl of 2.5 M NaOH to the labeling reaction and incubated at 37°C for 15 min. Then, 10 µl of 2 M HEPES buffer was added and the mixture vortexed. The amino-allyl labelled cDNA was purified by a Cyscribe GFX purification kit (GE Healthcare) according to the manufacturer's instructions. The purified amino-allyl labelled cDNA was labelled by adding Cy3 and Cy5 dyes to control and experimental samples by incubating in the dark for 5 hours at room temperature. After incubation, the unlabelled CyDyes were quenched by the addition of 15 µl of 4 M hydroxylamine hydrochloride and incubated in dark for 15 min. The CyDye labelled cDNAs were by purified with

Cyscribe GFX purification kit according to manufacturer's instructions. The total yield of the labelled cDNA was calculated from spectrophotometer absorbance at 260 nm (cDNA synthesis) 550 nm (Cy3) and 650 nm (Cy5). The entire eluted sample (70µl) was measured in a UV Microcuvette (Fischer: CXA-205-020F). The measured values were converted to picomoles with the formula.

Cy3 = A550 x total eluted volume (µl) x 0.15

Cy5 = A650 x total eluted volume (µl) x 0.25

cDNA = A260 x total eluted volume (µl) x 37

For a typical hybridization reaction, 30pmole of each Cydye labelled cDNA was taken in an amber eppendorf tube and dried to less than 10µl in a DNA Mini Vaccum concentrator, at pressure at 1mm, temp at 37°C, for approx 2 hrs. Store the dried samples at -20°C.

**2.4.3 Microarray slide blocking**

Prior to hybridization, the slides were blocked by succinic anhydride in borate buffered-NMP (1-methyl-2-pyrrolidinone) method. For this, 0.5 g of Succinic anhydride (Sigma) was dissolved in 31.5 ml of 1-Methyl-2-Pyrrolidinone (Sigma) in a 50 ml centrifuge tube (Greiner bio-one). Once the succinic anhydride had dissolved, 3.5 ml of 0.2 M sodium borhydrate pH 8.0 (Sigma) was added to the tube, capped and mixed once by inverting. Immediately, a slide was dipped and plunged up and down 5 times and then incubated at 26ºC in an orbital shaker (100 rpm) for 15 min. After the incubation, the slide was transferred to a 50 ml centrifuge tube containing milliQ water and then incubated in an orbital shaker for 2 min. Finally, the slide was transferred to a 50 ml centrifuge tube containing ethanol and plunged up and down 5 times at room temperature. The slide was then immediately dried at 800 rpm for 5 min, with the

bar-coded side at the bottom and the spotted side facing outwards in the slide holder.

### 2.4.4 Microarray Hybridization in an Advalytix arraybooster

Microarray-based gene expression analysis has become an indispensable tool in large scale gene expression analysis. Microarrays are generally a very lengthy process which involves many steps, where slight change or variation in any of the step might lead to false positives. In conventional microarrays, the hybridization mix is sandwiched between the slide and the coverslip, where the spreading of the mix through the slide is dependent on diffusion. However, inadequate diffusion leads to poor hybridization, which results in a low signal-to-noise ratio. This can be overcome by mixing the hybridization mix. The Arraybooster (Advalytix) employs a novel microagitation technology based on surface acoustic waves (SAW), which mixes the sample during the process, thus increasing the signal-to-noise ratio. This technology takes the advantage of the AdvaCard (Advalytix) which is like a coverslip embedded with mixer chips and contact pads (Figure 2.4.1). The mixer chip contains two interdigital transducers (IDT's) which generates an intense Surface Acoustic Waves (SAW), when a high radio frequency (RF) power is applied. The SAW propagates along the AdvaCard and mixes the samples. The AdvaCard has between1-3 mixer chips depending on the type of the card (Toegl *et al.*, 2003).

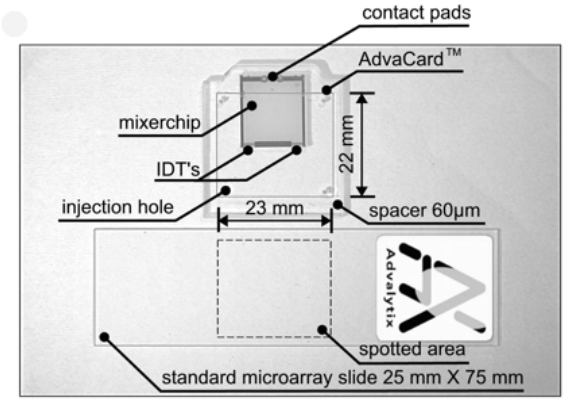



**Figure 2.4.1**. Mixer chip with two interdigital transducer electrodes glued on the AdvaCard (Toegl *et al.*, 2003).

### 2.4.5 Microarray hybridization.

Microarray hybridization was performed in an Arraybooster (Advalytix) controlled by BoosterControl 3.0 software with the mixing power at 27 and pulse/pause ratio of 3:7 at 42ºC for 18 hrs. The hybridization chamber was prewarmed to 42ºC and the reservoirs were filled with 500 µl of Advahum (humidyifing buffer) to maintain the humidity in the chamber. The AdvaCard (AC3C), a microagitation card with three agitation chips, was sandwiched over the slide in the chamber and secured with the fastening bolts as shown in Figure 2.4.2.

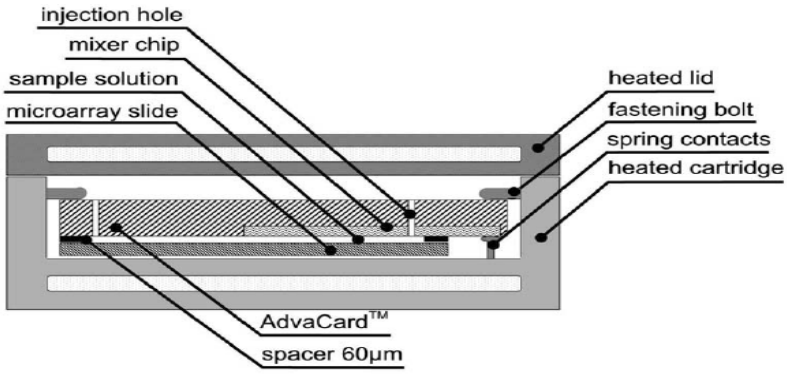



**Figure 2.4.2** Set up of microarray slide in the arraybooster (Toegl *et al.*, 2003).

Equal amounts (30 pmoles) of Cy5 and Cy3 labeled cDNA were mixed and concentrated to 10 µl in an amber tube. 90 µl of hybridization solution containing 25% formamide, 5X SSC, 0.1% SDS and calf thymus DNA (9 µg), was added to the concentrated labelled DNAs and heated at 95°C for 2 min. The mixture was incubated at room temperature for 2 min and centrifuged at 13,000 rpm for 2 min. 95 µl of heated hybridization mixture was loaded into an AdvaCard-slide sandwich by capillary action. After hybridization, AdvaCard-slide sandwiches were placed into pre-warmed buffer I (2X SSC, 0.1% SDS) at 42°C. The slides were removed from sandwiches and washed once in pre-warmed buffer I (2X SSC, 0.1% SDS) for 8 min at 42ºC, twice in pre-warmed buffer II (0.2X SSC, 0.1% SDS) for 4 min at 42°C and two times in buffer III (0.2X SSC) for

4 min and 1 min in buffer IV (0.1X SSC), at 26ºC. Slides were dipped and plunged up and down 2 times in milliQ water and then in isopropanol for 5 seconds each at room temperature before drying by centrifugation at 800 rpm for 5 min keeping the bar-coded side at the bottom and the spotted side facing outwards in the slide holder. Slides were then scanned for fluorescence intensity in a GenePix 4000A (Axon Instruments, USA) microarray scanner at a resolution of 10 μM, and the results recorded in 16-bit multi-image TIFF (Tagged Image File Format) file. Scanned image files were converted to raw data using Bluefuse for Microarrays 2.0, image analyzing software **(**BlueGnome Limited, UK), which employs advanced statistical methods to distinguish signal from background noise. The list of microarrays performed is categorized based on the experiment and shown in Table 2.4.1.

| Rep No | Array Number | Experiment Cy3 | Control Cy5 |
|---|---|---|---|
| **Pea root exudates** | | | |
| 1 | Vinoy25 | Root exudates | Pyruvate and $NH_4^+$ |
| 2 | Vinoy41 | Root exudates | Pyruvate and $NH_4^+$ |
| 3 | Vinoy42 | Pyruvate and $NH_4^+$ | Root exudates |
| **Hesperetin** | | | |
| 1 | Vinoy43 | Hesperetin | Pyruvate and $NH_4^+$ |
| 2 | Vinoy44 | Hesperetin | Pyruvate and $NH_4^+$ |
| 3 | Vinoy45 | Pyruvate and $NH_4^+$ | Hesperetin |
| **7day old pea rhizosphere - 1dpi** | | | |
| 1 | Vinoy26 | Glucose and $NH_4^+$ | 7d Pea 1dpi |
| 2 | Vinoy31 | Glucose and $NH_4^+$ | 7d Pea 1dpi |
| 3 | Vinoy34 | 7d Pea 1dpi | Glucose and $NH_4^+$ |
| 4 | Vinoy55 | Glucose and $NH_4^+$ | 7d Pea 1dpi |
| 5 | Vinoy56 | Glucose and $NH_4^+$ | 7d Pea 1dpi |
| **7day old pea rhizosphere - 3dpi** | | | |
| 1 | Vinoy27 | Glucose and $NH_4^+$ | 7d Pea 3dpi |
| 2 | Vinoy32 | Glucose and $NH_4^+$ | 7d Pea 3dpi |
| 3 | Vinoy35 | 7d Pea 3dpi | Glucose and $NH_4^+$ |
| 4 | Vinoy57 | Glucose and $NH_4^+$ | 7d Pea 3dpi |
| 5 | Vinoy58 | Glucose and $NH_4^+$ | 7d Pea 3dpi |
| **7day old pea rhizosphere - 7dpi** | | | |
| 1 | Vinoy28 | Glucose and $NH_4^+$ | 7d Pea 7dpi |

| | | | |
|---|---|---|---|
| 2 | Vinoy33 | Glucose and $NH_4^+$ | 7d Pea 7dpi |
| 3 | Vinoy36 | 7d Pea 7dpi | Glucose and $NH_4^+$ |
| 4 | Vinoy59 | Glucose and $NH_4^+$ | 7d Pea 7dpi |
| 5 | Vinoy60 | Glucose and $NH_4^+$ | 7d Pea 7dpi |
| **14day old pea rhizosphere - 1dpi** | | | |
| 1 | Vinoy29 | Glucose and $NH_4^+$ | 14d Pea 1dpi |
| 2 | Vinoy37 | Glucose and $NH_4^+$ | 14d Pea 1dpi |
| 3 | Vinoy39 | 14d Pea 1dpi | Glucose and $NH_4^+$ |
| 4 | Vinoy67 | Glucose and $NH_4^+$ | 14d Pea 1dpi |
| **21day old pea rhizosphere - 1dpi** | | | |
| 1 | Vinoy30 | Glucose and $NH_4^+$ | 21d Pea 1dpi |
| 2 | Vinoy38 | Glucose and $NH_4^+$ | 21d Pea 1dpi |
| 3 | Vinoy40 | 21d Pea 1dpi | Glucose and $NH_4^+$ |
| 4 | Vinoy69 | Glucose and $NH_4^+$ | 21d Pea 1dpi |
| **7day old alfalfa rhizosphere - 7dpi** | | | |
| 1 | Vinoy71 | 7d Alfalfa 7dpi | Glucose and $NH_4^+$ |
| 2 | Vinoy72 | Glucose and $NH_4^+$ | 7d Alfalfa 7dpi |
| 3 | Vinoy75 | 7d Alfalfa 7dpi | Glucose and $NH_4^+$ |
| **7day old Sugarbeet rhizosphere - 7dpi** | | | |
| 1 | Vinoy73 | Glucose and $NH_4^+$ | 7d Sugarbeet 7dpi |
| 2 | Vinoy74 | Glucose and $NH_4^+$ | 7d Sugarbeet 7dpi |
| 3 | Vinoy76 | 7d Sugarbeet 7dpi | Glucose and $NH_4^+$ |
| **Comparative rhizosphere (Replicate set 1)** | | | |
| 1 | Vinoy49 | 7d Pea 7dpi | 7d Alfalfa 7dpi |
| 2 | Vinoy50 | 7d Alfalfa 7dpi | 7d Pea 7dpi |
| 3 | Vinoy54 | 7d Sugarbeet 7dpi | 7d Alfalfa 7dpi |
| 4 | Vinoy53 | 7d Alfalfa 7dpi | 7d Sugarbeet 7dpi |
| 5 | Vinoy51 | 7d Sugarbeet 7dpi | 7d Pea 7dpi |
| 6 | Vinoy52 | 7d Pea 7dpi | 7d Sugarbeet 7dpi |
| **(Replicate set 2)** | | | |
| 7 | Vinoy61 | 7d Pea 7dpi | 7d Alfalfa 7dpi |
| 8 | Vinoy621 | 7d Alfalfa 7dpi | 7d Pea 7dpi |
| 9 | Vinoy651 | 7d Alfalfa 7dpi | 7d Sugarbeet 7dpi |
| 10 | Vinoy66 | 7d Sugarbeet 7dpi | 7d Alfalfa 7dpi |
| 11 | Vinoy63 | 7d Pea 7dpi | 7d Sugarbeet 7dpi |
| 12 | Vinoy64 | 7d Sugarbeet 7dpi | 7d Pea 7dpi |

**Table 2.4.1.** List of microarray experiments performed for this study.

### 2.4.6 Microarray data analysis.

Each experiment was performed in 3 or 5 five biological replicates depending on the experiment including a dye swap to minimize the labeling bias. The raw data output from Bluefuse was normalized with Lowess (within array) and quantile (between arrays) method (Smyth & Speed, 2003), and analyzed using the Bioconductor software (www.bioconductor.org) package (Gentleman *et al.*, 2004) LIMMA (Linear Models for Microarray data) (Smyth, 2005), working on the R statistical computing environment (www.r-project.org). The duplicate spots were averaged using the function duplicate correlation (Smyth *et al.*, 2005). The differential gene expression was assessed using linear models and empirical Bayes moderated F statistics (Smyth, 2004). The genes were considered significant if their P values, adjusted for multiple testing corrections by Benjamini and Hochberg false discovery method (Benjamini & Hochberg, 1995) is $\leq 0.01$ and $\log_2$ fold change is $\geq 1$ for upregulated and $\leq -1$ for downregulated genes.

The statistically significant differentially expressed genes were further analyzed and characterized using various bioinformatic tools, webbases and webservers including NCBI's BLAST (Altschul *et al.*, 1990; Altschul & Lipman, 1990; Altschul *et al.*, 1997), ClustalW (Thompson *et al.*, 1994), Rhizodb (Chaudhuri & Pallen, 2006), Rhizobase (http://www.kazusa.or.jp/rhizobase), KEGG (Kanehisa *et al.*, 2008), CDD (Marchler-Bauer *et al.*, 2007), STRING (von Mering *et al.*, 2007), InterPro (Mulder *et al.*, 2007), Pfam (Finn *et al.*, 2006), and PROSITE (Hulo *et al.*, 2008).

The up- and down-regulated gene lists corresponding to each microarray dataset is given at the end of each section or chapter, sorted numerically on gene ID. For ease of finding the gene, the microarray datasets discussed in this thesis, is given on a compact disc accompanying the thesis.

### 2.4.7 Work flow of microarray experiment.

A brief overview of the general strategy followed to perform microarray based gene expression analysis in rhizosphere recovered bacterial cells was shown in Figure 2.4.3. The pea seedlings were grown in the rhizosphere tubes for desired

time period (7, 14 or 21 days old) and inoculated with $10^8$ cells, (harvested at 0.6 $OD_{600}$nm, washed thrice with nitrogen free-rooting solution) and grown for the stipulated time period (1, 3 or 7 days) (section 2.3.3). After the period, the bacterial cells were recovered from the pea rhizosphere tubes by differential centrifugation and the cells were lysed in a ribolyser, and the total RNA was extracted with an RNeasy mini kit (Qiagen) (section 2.2.8). In parallel, free-living Rlv3841 (control) was grown under laboratory conditions in AMS with 10 mM glucose and 10 mM ammonium chloride as sole carbon and nitrogen source (section 2.1.1) and the total RNA extracted (section 2.2.8). The quality and quantity of the extracted total RNA was checked in a Bio-Rad Experion as described in section 2.2.9. The total RNA was amplified using Genisphere SenseAMP as described in section 2.2.10. The integrity and the concentration of the amplified sense oriented total RNA was checked again in Bio-Rad's Experion as described in section 2.2.9. In the case of comparing Rlv3841 cells grown with and without pea root exudates or hesperetin, the total RNA was labeled without amplification.

For microarrays, 10-15μg of both control and experimental amplified sense oriented RNA samples were reverse transcribed and labelled with Cydyes (Cy3 and Cy5) by indirect labeling (section 2.4.2), and hybridized onto a sodium borohydrate-succinic anhydride blocked slide (section 2.4.3) in an arraybooster and washed (section 2.4.4). Images were acquired from the slide using a Genepix 4200A scanner, equipped with Genepix 5.0 pro software, and image analysis was carried out using BlueGnome Bluefuse. The raw data was normalized and subjected to a rigorous statistical analysis in LIMMA (Bioconductor suite of programs) (section 2.4.5) and validated by qRT-PCR (section 2.2.12). The genes were considered significant if their adjusted (Benjamini and Hochberg method) P value is $\leq 0.01$ and the average $\log_2$ fold is $\geq 1$ of 3-5 biological replicates analyzed.

A similar strategy was followed to compare two rhizosphere recovered samples (direct design), whereas in the case of comparing two free-living cells, the RNA samples were not amplified.

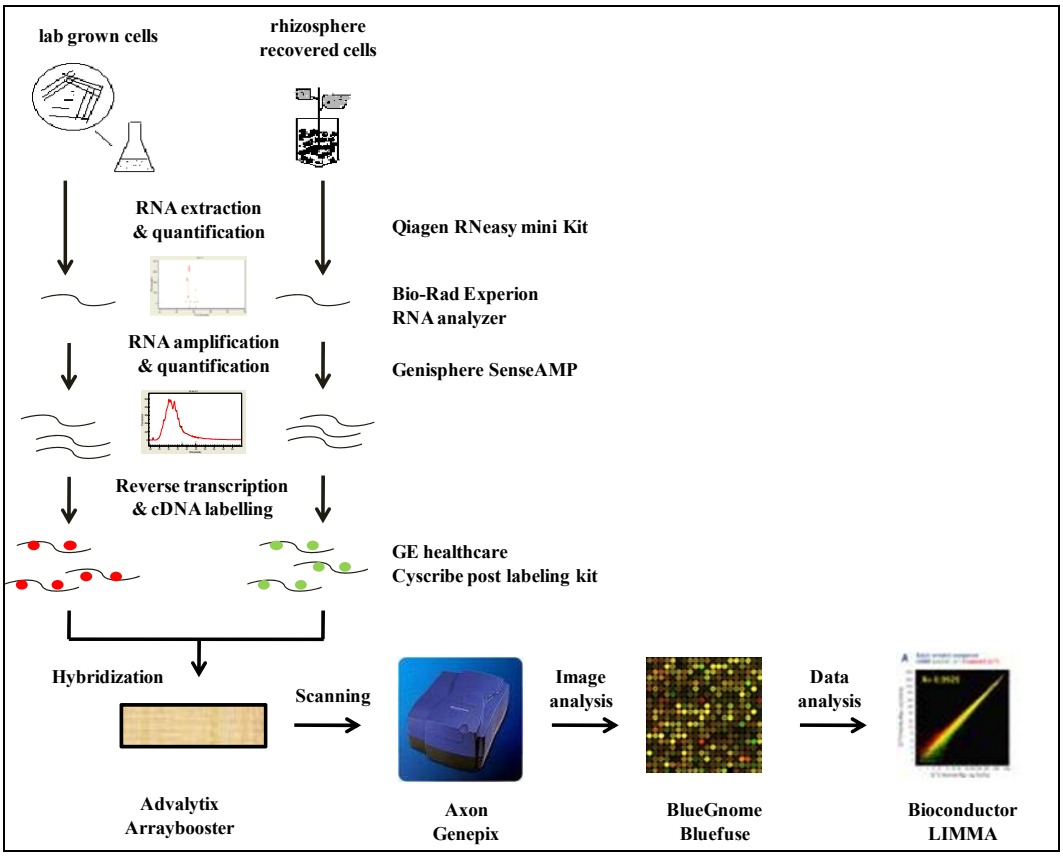



**Figure 2.4.3.** Schematic representation of microarray work flow followed in this study.

# Chapter 3

## Transcriptomic analysis of *Rhizobium leguminosarum* bv. *viciae* strain 3841 in the pea rhizosphere.

As discussed in chapter 1, many techniques have been developed over the past decade to study the expression of bacterial genes in the host environment. Most of these techniques have been successfully used in identifying the genes specifically induced during plant-bacterial interactions, especially during rhizosphere colonization and the *Rhizobium*-legume symbiosis. Although the application of these techniques such as reporter gene technology, DFI-OT (Allaway *et al.,* 2001), IVET (Rainey, 1999), STM (Pobigaylo *et al.,* 2006) and RIVET (Gao & Teplitski, 2008) were very promising, they are very time consuming and identified only one to few genes at a time, thus providing only limited information. As rhizobial symbiosis is a complex process, involving many biological interactions, starting from colonization to nitrogen fixation, a genomic level study of the rhizobia during symbiosis will provide more insights behind these interactions.

In this genomic era, many studies have taken advantage of genome sequences and "omics" technology to explore the principles behind such complex interactions. In rhizobia, all the microarray based transcriptomics and proteomics studies reported so far, have focussed primarily on investigating the gene expression of bacteroids. The complexity of the rhizosphere and the difficulties in performing microarrays with rhizosphere-recovered samples has kept rhizosphere colonization an unexplored area in rhizobia and other plant-associated bacteria.

Mark *et al.,* (2005) first attempted to study the changes in the transcriptome of *P. aeruginosa* PA01 in response to the aseptically obtained sugarbeet root exudates from two beet cultivars under laboratory conditions. This study identified many previously characterized genes and many unknown genes, which may have role in rhizosphere colonization. The major drawback of this study is that the root exudates were collected in an aseptic environment, without the influence of any bacteria, which does not reflect the real conditions as bacterial presence in the rhizosphere largely influences the composition of the secreted root exudates (Cooper, 2004). A recent report overcame this drawback by

performing transcriptomic analysis on *P. putida* recovered from corn rhizosphere (*in vitro*) compared against the exponential, stationary and microcosm-grown free-living cells under laboratory conditions, providing a snapshot of the transcriptome of *P. putida* in the rhizosphere (Matilla *et al.,* 2007). However, as this report identified only 90 genes up-regulated in the rhizosphere, it is unlikely to provide a complete picture of rhizosphere bacteria in the rhizosphere. In this work, we attempt to overcome all the drawbacks and limitations of earlier studies by designing an integrative approach to explore the transcriptome of Rlv3841 in the pea rhizosphere (*in vitro*).

This research work is the first of its kind to investigate the gene expression of rhizobia during rhizosphere colonization of Rlv3841 in the pea rhizosphere. The major objective of this work was to decipher the transcriptome of Rlv3841 in the pea rhizosphere under different conditions, in comparison with laboratory grown cells. As this is the first transcriptomic study of Rlv3841 in the rhizosphere, our initial objective was to investigate the effects of pea root exudates (aseptic) and hesperetin (1μM) (flavonoid) individually on the transcriptome of Rlv3841 under laboratory growth conditions.

The second objective was to perform transcriptomics on Rlv3841 recovered from the *in vitro* pea rhizosphere. For this, two experimental conditions were set up to obtain transcriptomic data; (1) Rlv3841 cells were recovered from 7d old pea (seedling) rhizosphere at 1, 3 and 7 days post-inoculation (2) Rlv3841 cells were recovered from three differently aged pea (seedling) rhizospheres (7, 14 and 21 days old) at 1 day post-inoculation. The microarray was then subjected to rigorous statistical analysis and the data validated with a set of representative differentially expressed genes by quantitative reverse transcriptase (real-time) PCR (qRT-PCR). Finally, based on the data, a model was constructed depicting the important events in the transcriptome of Rlv3841 in the pea rhizosphere.

Henceforth, for clarity, a system of abbreviation is employed. The usual notation for referring to a gene is e.g. *dctA* and its protein is DctA. In this thesis a gene will be referred to as e.g. *pRL100001* or *RL3016* and its protein as pRL100001 or RL3016. Also, the $\log_2$-fold expression values for genes will be

given in square bracket e.g. 5 $\log_2$ fold up- or down-regulation of *pRL100100*, will be denoted as *pRL100100* [5U] and down-regulated as *pRL100100* [-2D], and combined analysis are shown as *pRL100100* [5U,4.5U,6U] for up- and *pRL100102* [-5D,-4.5D,-6D] for down-regulated genes. In addition, for ease of expression, the gene name will often be followed by its expression level e.g., *RL3378* (*cinI*) [2.4U,1.1U,1.0U]. Also, ATP binding cassette-transporter is abbreviated (ABC-T), solute binding protein (SBP), identity (Id) and similarity (Smty). The following abbreviations are used for the microarray experiments:

| **Experiment** | **Abbreviation** |
|---|---|
| 7d old pea rhizosphere | 7dPea |
| 14d old pea rhizosphere | 14dPea |
| 21d old pea rhizosphere | 21dPea |
| 7d old alfalfa rhizosphere | 7dAlf |
| 7d old sugarbeet rhizosphere | 7dSB |

In addition, the length of the incubation of bacteria in the rhizosphere is given as day post inoculation (dpi) as shown below:

| **Experiment** | **Abbreviation** |
|---|---|
| 7d old pea rhizosphere, 1d post-inoculation before harvest | 7dPea-1dpi |
| 7d old pea rhizosphere, 3d post-inoculation before harvest | 7dPea-3dpi |
| 7d old pea rhizosphere, 7d post-inoculation before harvest | 7dPea-7dpi |
| 14d old pea rhizosphere, 1d post-inoculation before harvest | 14dPea-1dpi |
| 21d old pea rhizosphere, 1d post-inoculation before harvest | 21dPea-1dpi |
| 7d old alfalfa rhizosphere, 7d post-inoculation before harvest | 7dAlf-7dpi |
| 7d old sugarbeet rhizosphere, 7d post-inoculation before harvest | 7dSB-7dpi |

Furthermore, the combined data from the microarray datasets are referred as:

| Experiment | Abbreviation |
|---|---|
| 7dPea-1dpi, 7dPea-3dpi and 7dPea-7dpi | 7dPea-137dpi |
| 7dPea-1dpi, 14dPea-1dpi and 21dPea-1dpi | 7-14-21dPea-1dpi |
| 7dPea-1dpi, 7dAlf-7dpi and 7dSB-7dpi | 7dPAS-7dpi |

The expression dataset of 7dPea-137dpi, 7d-14d-21d-1dpi and 7dPAS-7dpi, both up-regulated genes (at least in one time point) and down-regulated genes (at least in one time point) can be found in the supplementary book-let accompanying this thesis. The genes in the table were sorted numerically on the gene identity (Gene ID). This dataset can also be found on the accompanying compact disc in a Microsoft excel format (version 2003).

### 3.2.1 Total RNA and 23S rRNA processing in *R. leguminosarum* 3841

Total RNA was extracted from free-living Rlv3841 cells grown in AMS medium supplemented with glucose and ammonia, and from rhizosphere recovered Rlv3841 cells as described in the section 2.2.8. The quality and the quantity of the extracted total RNA was checked in the microfluidic based RNA analyzer (Experion) as described in section 2.2.9. In contrast with the typical *E. coli* total RNA profile, which shows peaks corresponding to 16S and 23S rRNA, the total RNA profile of free-living and rhizosphere Rlv3841 showed a pattern in which the 23S rRNA peak was reduced. This has been observed in another *R. leguminosarum* strain and is due to processing (Klein *et al.,* 2002). Rlv3841 23S rRNA has two fragmentation sites; the first RNase III cleavage site is located 0.3 kb from the 5' end of 23S rRNA, which, on cleavage, fragments the 23S rRNA into 0.3kb and 2.6kb. The second fragmentation site domain III is located in the middle of the 23S rRNA gene and fragments it from 2.6 kb into two ~1.3kb fragments. The processing is time-dependent, but once the 23S rRNA is completely processed it leads to two sharp concentrated peaks in stationary phase cultures. The schematic representation of the 23S rRNA processing is shown in Figure 3.2.1, adapted from Klein *et al.* (2002).

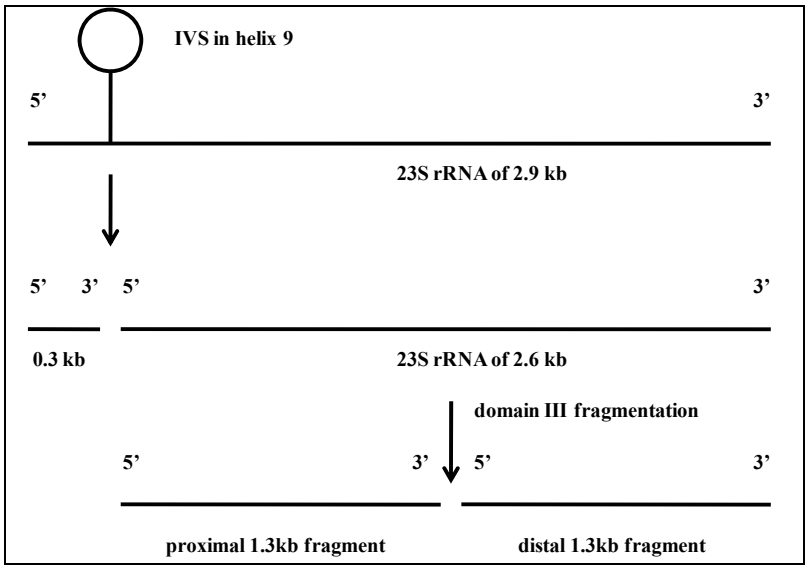
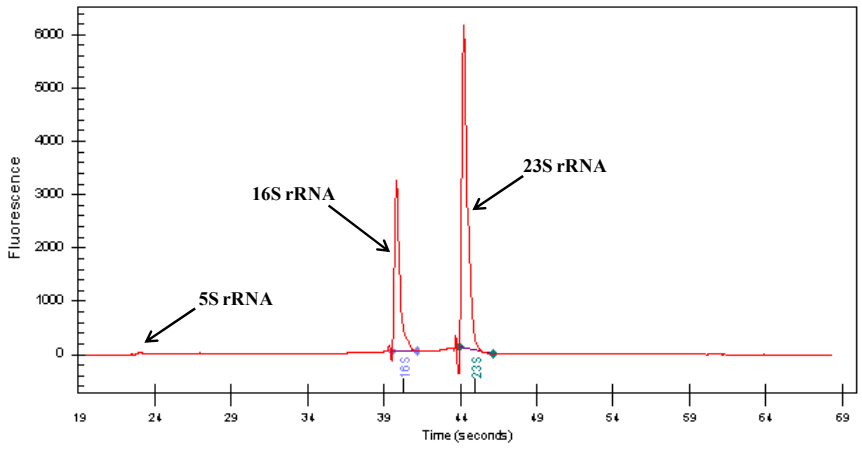
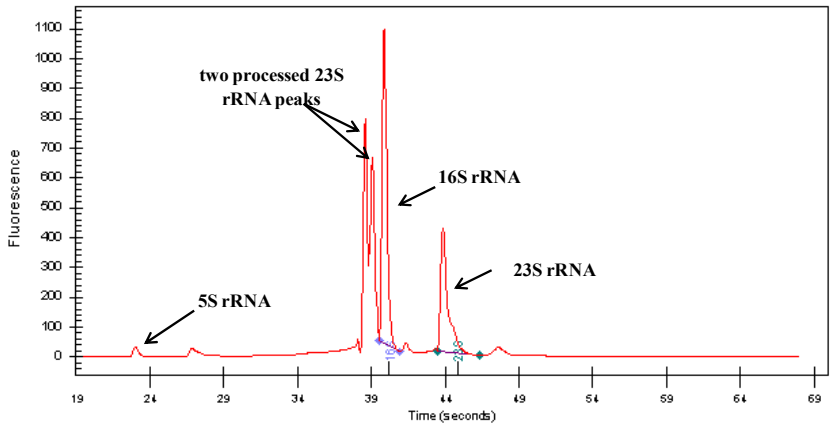



**Figure 3.2.1.** Schematic representation of the 23SrRNA fragmentation in *R. leguminosarum* ATCC 10004$^{T}$ (Klein *et al.,* 2002).

**A.**




**B.**




**C.**

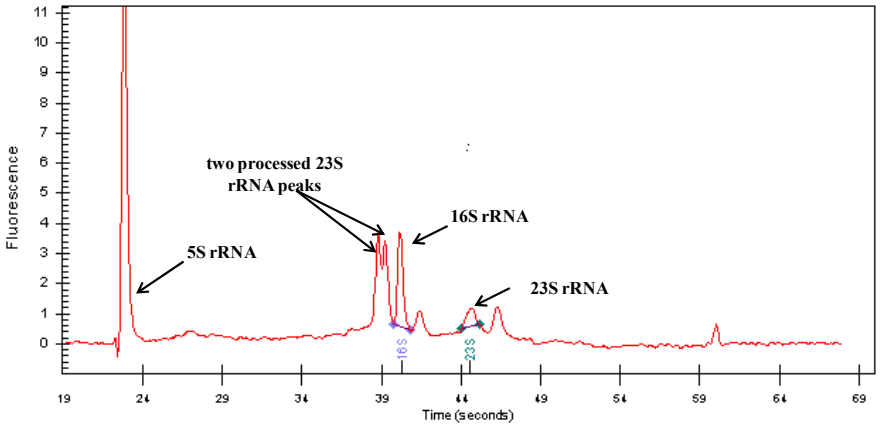



**Figure 3.2.2.** Electrophoregrams of total RNA extracted from (A) *E. coli*, (B) free-living Rlv3841, (C) rhizosphere recovered Rlv3841.

The microfluidic based RNA analyzers like Bio-Rad's Experion and Agilent's Bioanalyzer have a built in function to score the quality and the integrity of the total RNA based on the ratio of 16S and 23S rRNA. Unfortunately, the score will not be meaningful in case of Rlv3841 RNA samples, because of the 23S rRNA fragmentation. Therefore, the quality of the extracted total RNA was checked visually from the electrophoregram to ensure the RNA was free from degradation. Total RNA was amplified using the Genisphere SenseAMP kit as described in section 2.2.10. The integrity and the concentration of the amplified sense oriented total RNA was checked again in a Bio-Rad experion RNA analyzer. A typical electrophoregram of the amplified total RNA is shown in Figure 3.2.3.

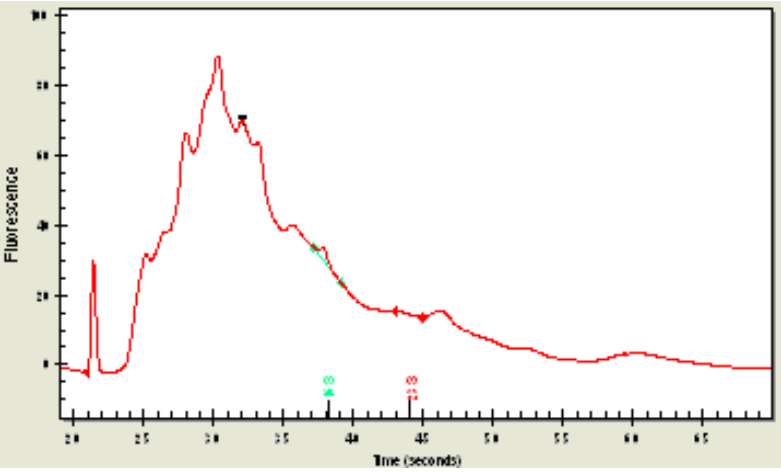



**Figure 3.2.3.** Electrophoregram of amplified sense-oriented total RNA.

### 3.2.2 Optimization of inoculum size and RNA amplification.

In order to perform transcriptomics in Rlv3841 on the pea (*P. sativum* cv Avola) rhizosphere two things need to be considered; (1) recovery of cells from the rhizosphere, (2) low yield of RNA from the recovered samples. Initial total RNA extraction from 12 rhizosphere tubes inoculated with $10^3$ CFU yielded very low concentration (<330ng) of total RNA, which is too low to perform a microarray. Increasing the number of rhizosphere tubes was not feasible, due to the time constraint and difficulties in handling a large number of tubes. From earlier experiments, it was known that the recovery of cells from the rhizosphere

also depends on the inoculum size. Therefore, the amount of inoculum was increased from $10^3$ to $10^8$ CFU. Comparison of RNA extraction from cells recovered from 7d old pea rhizosphere inoculated with two different inoculum sizes ($10^3$ and $10^8$ CFU/ml), showed higher yield of RNA from rhizosphere inoculated with $10^8$ CFU. Although the RNA yield had been increased, it was still not sufficient to perform microarrays.

Another approach to performing microarrays with low amounts of RNA is either to amplify the RNA samples or to amplify the signal from the microarray. Both of these approaches were tried and it was found that to amplify the RNA samples is more reliable and reproducible than amplifying signals from the microarray. For amplifying RNA samples, the MessageAmp II Bacteria RNA amplification kit (Ambion) and SenseAMP RNA amplification kit (Genisphere) were tried and the SenseAMP found to be more reliable and reproducible. For amplifying microarray signals, the 900MPX array detection kit which uses 3DNA dendrimer technology was tried and though it produced good results, RNA amplification was more reproducible. In order to ensure the amplification of the total RNA is linear, a comparative microarray analysis was performed with amplified and unamplified control samples (3841 grown in AMS-10mM glucose and 10mM $NH_4^+$) and it was found the bias was minimal (data not shown).

For all RNA amplification and optimization experiments, RNA was recovered from rhizosphere inoculated with $10^8$ CFU. However, as the effect of the inoculum size on the transcriptome of the colonizing rhizobia was not clear, three replicates of microarrays were performed with RNA extracted from cells recovered 7d post-inoculation from a 7d pea rhizosphere inoculated with either $10^3$ or $10^8$ CFU Rlv3841. It was observed that the bias was minimal (data not shown). Although amplification of RNA from $10^3$ CFU inoculated rhizosphere was sufficient to perform microarrays, $10^8$ CFU was used as a standard inoculum size as it permitted recovery of more total RNA from a manageable number of rhizosphere tubes.

### 3.2.3 Effect of root exudates on free-living *R. leguminosarum* 3841

This first experiment was designed to investigate the changes in the transcriptome of free-living Rlv3841 when grown with aseptic pea root exudate under laboratory conditions. Although it is known that the presence of rhizobia in the rhizosphere influences the composition and concentration of the root exudates (van Brussel *et al.,* 1990), this experiment was performed to get a glimpse of the transcriptional changes in response to flavonoids and other plant derived signals which elicit bacterial gene expression. Rlv3841 cells were grown with and without pea root exudates in growth media containing 30 mM pyruvate and 10 mM ammonia as carbon and nitrogen source as described in section 2.1.1. Total RNA was then isolated from the cells, quantified, amplified and re-quantified as described in section 2.2.8 to 2.2.10. Amplified RNA was then labelled and used in microarrays as described in section 2.4.2-2.4.4. The microarray data were then normalized (Figure 3.2.4) and analysed using LIMMA as described earlier in section 2.4.5. As pea root exudate was supplemented in a minimal media containing pyruvate and ammonia as sole carbon and nitrogen sources, the genes involved in aromatic compound degradation will not be expressed. An attempt was also made to use the pea root exudates as sole carbon and energy source, but the concentration of the pea root exudates did not support sufficient growth of Rlv3841 under laboratory conditions.

Only 33 genes were differentially expressed with 30 up- and 3 down-regulated genes (Table 3.2.1). The up-regulated genes fell into different functional categories including highly induced *nod* operon (Long, 1996): *pRL100175* (*nodO*) [4.8U], *pRL100179* (*nodN*) [4.2U], *pRL100180* (*nodM*) [4.3U], *pRL100181* (*nodL*) [2.4U], *pRL100182* (*nodE*) [4.2U], *pRL100183* (*nodF*) [4.4U], *pRL100185* (*nodA*) [5.0U], *pRL100186* (*nodB*) [6.4U], *pRL100187* (*nodC*) [6.6U], *pRL100188* (*nodI*) [4.9U] and *pRL100189* (*nodJ*) [4.5U]. The, *rhi* genes: *pRL100169* (*rhiA*) [4.2U], *pRL100170* (*rhiB*) [4.5U], *pRL100171* (*rhiC*) [3.0U] (Dibb *et al*., 1984; Economou *et al*., 1989; Cubo *et al*., 1992), and *pRL100164* (*rhiI*) [4.2U] which is involved in quorum sensing (Downie *et al*, personal communication).

A cluster of genes, *RL3346* [1.8U] coding for a glutamine synthase-like protein, *RL3347* [1.6U] coding for permease component of an ABC-T and *RL3348* coding for an N-formylglutamate amidohydrolase family protein were induced. They are probably part of a putative operon from *RL3345-49*. A database similarity search of RL3348 [1.4U] showed 60% identity and 72% similarity with the hypothetical protein (Atu2144) of *A. tumefaciens* C58 (Figure 3.2.5). The protein structure of Atu2144 has been solved recently (PDB Id: 2ODF) and it is an octameric protein with ten sulphate ligands bound at various position. The amino acids in Atu2144, which interacts with the sulphate ligands, were found to be conserved in both RL3348 and Atu2144. In *P. putida*, N-formylglutamate amidohydrolase was reported to be involved in histidine degradation, catalyzing the conversion of N-formyl-L-glutamate to L-glutamate and formate. The function of Atu2144 is not yet determined but taken together both (Atu2144 and RL3378) proteins may play a role in nutrient scavenging in the rhizosphere.

Only three genes were down-regulated, two coding for hypothetical proteins (*RL0610* [-1.3D] & *RL3172* [-1.0D]) and one coding for putative transmembrane protein (*pRL110005* [-1.0D]). The precise functions of these genes are unknown.

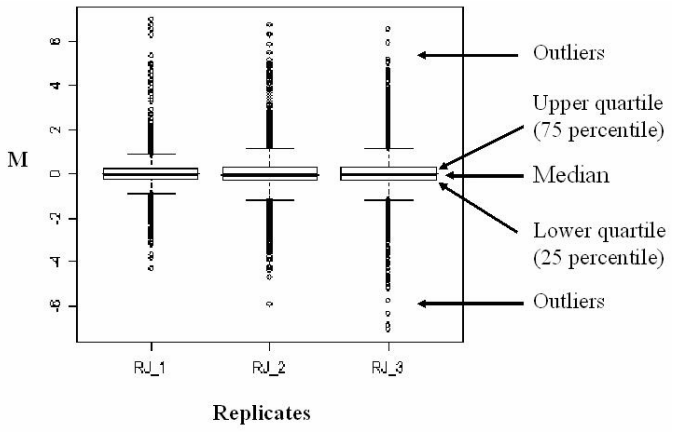



**Figure 3.2.4.** Boxplots showing the differential expression values of the three root exudate biological replicate microarray experiments. M represents the differential expression ($\log_2$ (experiment/control) after quantile normalization. Boxplots provide information on the upper (75%) and lower (25%) quartiles, median, range and extreme values (outliers). In this experiment, RJ_1 and RJ_2 the experimental RNA sample was labelled with cy5 and control RNA sample was labelled with cy3. In RJ_3 the dyes have been swapped.

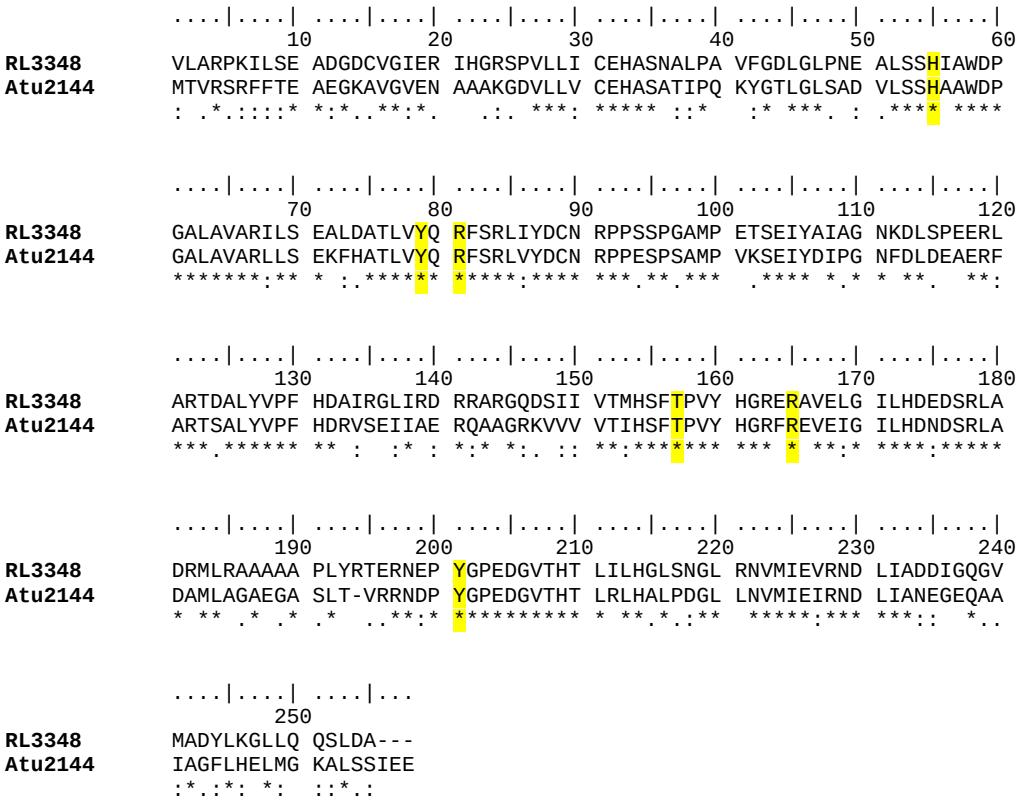



**Figure 3.2.5.** ClustalW 1.4 alignment of RL3348 of *R. leguminosarum* 3841 and Atu2144 of *A. tumefaciens* C58. The shaded residues are conserved and interact with sulphate ion ligands in Atu2144.

* = residues conserved by all.
: = residues conserved according to type (polar, non-polar, aromatic, aliphatic).
. = residues that are semi-conserved.

**Table 3.2.1.** Genes differentially expressed in response to pea root exudates. The genes down-regulated genes are shaded.

| ID | Function | Name | $log_2$ fold | P value |
|---|---|---|---|---|
| **Rhizosphere induced genes** | | | | |
| pRL100175 | putative nodulation protein | *nodO* | 4.8 | 0 |
| pRL100179 | nodulation protein | *nodN* | 4.2 | 0 |
| pRL100180 | glucosamine--fructose-6-phosphate aminotransferase | *nodM* | 4.3 | 0 |
| pRL100181 | putative nodulation protein | *nodL* | 2.4 | 0 |
| pRL100182 | nodulation protein E | *nodE* | 4.2 | 0 |
| pRL100183 | nodulation protein F | *nodF* | 4.5 | 0 |
| pRL100185 | nodulation protein A | *nodA* | 5.0 | 0 |
| pRL100186 | chitooligosaccharide deacetylase | *nodB* | 6.4 | 0 |
| pRL100187 | N-acetylglucosaminyltransferase | *nodC* | 6.6 | 0 |
| pRL100188 | nod factor export ATP-binding protein I | *nodI* | 4.9 | 0 |
| pRL100189 | nodulation protein | *nodJ* | 4.5 | 0 |
| pRL100169 | rhizosphere expressed protein RhiA | *rhiA* | 4.3 | 0 |
| pRL100170 | putative rhizosphere induced protein RhiB | *rhiB* | 4.5 | 0 |
| pRL100171 | putative rhizosphere induced protein | *rhiC* | 3.0 | 0 |
| pRL100164 | putative autoinducer synthesis protein | *rhiI* | 4.2 | 0 |
| **Transport and binding proteins** | | | | |
| pRL120724 | putative transmembrane protein | - | 1.4 | 0.00868 |
| RL2418 | putative substrate-binding component of ABC transporter | - | 3.0 | 0 |
| RL3347 | putative permease component of ABC transporter | - | 1.6 | 0 |
| RL3350 | putative substrate-binding component of ABC transporter | - | 1.0 | 0 |
| RL3702 | putative transmembrane protein | - | 1.6 | 0.00056 |
| RL4195 | putative transmembrane protein | - | 1.7 | 0.00323 |
| pRL110005 | putative transmembrane protein | - | -1.0 | 0.00037 |
| **Metabolism related genes** | | | | |
| RL1911 | putative arylsulfatase | - | 1.8 | 0 |
| RL3346 | putative glutamine synthetase | - | 1.8 | 0 |
| RL3348 | putative N-formylglutamate amidohydrolase family protein | - | 1.4 | 0 |
| **DNA replication repair and modification** | | | | |
| pRL110305 | putative topoisomerase I | - | 1.4 | 0.00041 |
| **Others** | | | | |
| pRL100163 | hypothetical protein | - | 3.4 | 0 |
| pRL100176 | putative insertion sequence, pseudogene | - | 1.7 | 0.00202 |
| RL3868 | conserved hypothetical protein | - | 1.7 | 0.0043 |
| RL1925 | conserved hypothetical protein | - | 1.3 | 0.00095 |
| RL1924 | conserved hypothetical exported protein | - | 1.5 | 3.00E-05 |
| RL0610 | hypothetical exported protein | - | -1.3 | 0.00256 |
| RL3172 | hypothetical protein | - | -1.0 | 0.00021 |

### 3.2.4 Effect of hesperetin on free-living *R. leguminosarum* 3841

Hesperetin is a flavonoid which is reported to induce nod genes in *R. leguminosarum,* cells pre-incubated in flavonoids, like naringenin and hesperetin, showed increased nodulation (Begum *et al.,* 2001). This experiment aimed to study the changes in the transcriptome of free-living Rlv3841 when grown with hesperetin (1μM) under laboratory conditions. Rlv3841 cells were grown with and without hesperetin in growth media containing 30 mM pyruvate and 10 mM ammonia as carbon and nitrogen source as described in section 2.1.1. Total RNA was isolated from the cells, quantified, amplified re-quantified as described in section 2.2.8 to 2.2.10. The amplified RNA was then labelled and microarrays performed as described in section 2.4.2-2.4.4. The microarray data were then normalized (Figure 3.2.6) and analysed using LIMMA as described earlier in section 2.4.5.

In this experiment, 94 genes were found to be differentially expressed with 47 up- and 47 down-regulated genes. The genes fell into different functional categories including a highly induced *nod* operon (Long, 1996): *pRL100175* (*nodO*) [7.1U], *pRL100178* (*nodT*) [2.9U], *pRL100179* (*nodN*) [6.0U], *pRL100180* (*nodM*) [6.2U], *pRL100181* (*nodL*) [5.2U], *pRL100182* (*nodE*) [6.6U], *pRL100183* (*nodF*) [6.4U], *pRL100185* (*nodA*) [6.4U], *pRL100186* (*nodB*) [7.3U], *pRL100187* (*nodC*) [7.2U], *pRL100188* (*nodI*) [5.7U] and *pRL100189* (*nodJ*) [6.0U]. The *rhi* genes; *pRL100169* (*rhiA*) [3.8U], *pRL100170* (*rhiB*) [4.4U], *pRL100171* (*rhiC*) [3.1U] (Dibb *et al*., 1984; Economou *et al*., 1989; Cubo *et al*., 1992) and *pRL100164* (*rhiI*) [4.0U] (Downie *et al*, personal communication) are which is involved in quorum sensing. The expression of *rhi* genes, is in contrast with the earlier reports that flavonoids (hesperetin) repress *rhi* gene expression (Cooper, 2004). We also observed a high induction of *rhiABCR* genes, when Rlv3841 was grown with protocatechuate (3 mM) ([2.7U],[3.0U],[3.0U],[1.5U]) or p-hydroxy benzoate (3 mM) ([0.5],[1.8U],[1.8U],[0.7]) (dataset not shown), suggesting the induction of *rhi* genes may be indirectly influenced by aromatic compound catabolism.

Surprisingly, the genes involved in malonate metabolism *RL0991* (*matB*) [2.4U] and *RL0991* (*matC*) [2.3U] (An & Kim, 1998) were induced. In addition, *copC* [1.2U] and *copA* [1.0U] involved in copper detoxification were also induced. It is essential to note that the copper detoxification genes *copSR* were found to be induced in *P. syringae* in the sugarbeet rhizosphere by IVET studies, indicating a role in copper homeostasis in rhizosphere colonization (Rainey, 1999). Apart from these previously characterized genes, many genes involved in transport, regulation and of unknown functions were induced.

In the down-regulated genes, a large cluster of genes involved in chemotaxis and motility related functions were strongly repressed, which is in contrast to the root exudates dataset. The down-regulation of chemotaxis and motility related genes when grown on hesperetin but not on root exudates, suggests that hesperetin alone, at micromolar concentration, may reduce motility.

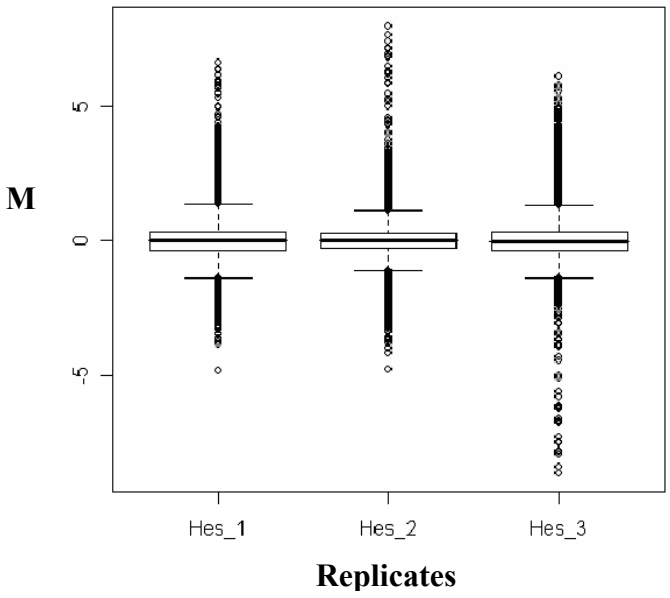



**Figure 3.2.6.** Boxplots showing the differential expression values of the three biological replicate microarray experiments supplemented with 1 µM hesperetin. M represents the differential expression ($\log_2$ (experiment/control)) after quantile normalization. In this experiment, Hes_1 and Hes_2 the experimental RNA sample was labelled with cy5 and control RNA sample was labelled with cy3. Hes_3 is a dye-swap experiment.

**Table 3.2.2.** Genes up-regulated in Rlv3841 in response to hesperetin.

| ID | Function | Name | $log_2$ fold | P value |
|---|---|---|---|---|
| **Rhizosphere induced genes** | | | | |
| pRL100169 | rhizosphere expressed protein RhiA | *rhiA* | 3.8 | 0 |
| pRL100170 | putative rhizosphere induced protein RhiB | *rhiB* | 4.4 | 0 |
| pRL100171 | putative rhizosphere induced protein | *rhiC* | 3.1 | 0 |
| pRL100175 | putative nodulation protein | *nodO* | 7.1 | 0 |
| pRL100178 | putative nodulation protein | *nodT* | 2.9 | 2.00E-05 |
| pRL100179 | nodulation protein | *nodN* | 6.0 | 0 |
| pRL100180 | glucosamine--fructose-6-phosphate aminotransferase | *nodM* | 6.2 | 0 |
| pRL100181 | putative nodulation protein | *nodL* | 5.2 | 0 |
| pRL100182 | nodulation protein E | *nodE* | 6.6 | 0 |
| pRL100183 | nodulation protein F | *nodF* | 6.4 | 0 |
| pRL100185 | nodulation protein A | *nodA* | 6.4 | 0 |
| pRL100186 | chitooligosaccharide deacetylase | *nodB* | 7.3 | 0 |
| pRL100187 | N-acetylglucosaminyltransferase | *nodC* | 7.2 | 0 |
| pRL100188 | nod factor export ATP-binding protein I | *nodI* | 5.7 | 0 |
| pRL100189 | nodulation protein | *nodJ* | 6.0 | 0 |
| pRL100198 | nitrogen fixation FixC protein | *fixC* | 1.2 | 7.00E-05 |
| pRL100156 | putative ferredoxin | *fdxB* | 1.2 | 0.00899 |
| pRL100163 | hypothetical protein | - | 3.6 | 0 |
| pRL100164 | putative autoinducer synthesis protein | *rhiI* | 4.0 | 0 |
| pRL100176 | putative insertion sequence, pseudogene | - | 4.5 | 0 |
| **Metabolism** | | | | |
| pRL90242 | putative nitrilotriacetate monooxygenase | - | 1.5 | 1.00E-05 |
| Malonate metabolism | | | | |
| RL0991 | putative Malonyl CoA synthetase | *matB* | 2.4 | 0 |
| RL0992 | putative transmembrane dicarboxylate carrier protein | *matC* | 2.3 | 2.00E-05 |
| **Copper detoxification** | | | | |
| RL2433 | putative copper oxidase | *copC* | 1.2 | 0.00013 |
| RL2434 | putative copper-containing oxidase subunit | *copA* | 1.0 | 0.00105 |
| RL2435 | putative efflux protein | - | 1.7 | 7.00E-05 |
| **Transport and binding proteins** | | | | |
| pRL100277 | putative permease component of ABC transporter | - | 1.4 | 0.00111 |
| pRL120492 | putative ATP-binding component of ABC transporter | - | 1.4 | 0.00033 |
| pRL90125 | putative permease transporter component | - | 2.1 | 1.00E-05 |
| RL2418 | putative substrate-binding component of ABC transporter | - | 3.2 | 0 |
| RL3359 | putative substrate-binding component of ABC transporter | - | 1.1 | 0.00109 |
| RL4179 | putative transmembrane transporter | - | 1.5 | 1.00E-05 |
| **Others** | | | | |
| pRL100192 | hypothetical protein | - | 1.1 | 0.00017 |
| pRL100193 | conserved hypothetical protein | - | 1.5 | 3.00E-05 |
| pRL100194 | hypothetical protein | - | 1.0 | 0.00233 |
| pRL110016 | conserved hypothetical protein | - | 2.2 | 0.00681 |
| pRL110375 | putative outer membrane porin protein | *ropA* | 1.2 | 1.00E-05 |
| pRL120342 | putative small heat shock protein | *hspD* | 1.1 | 2.00E-05 |
| RL0018 | putative transmembrane protein | - | 1.4 | 2.00E-05 |
| RL2444 | conserved hypothetical protein | - | 2.2 | 1.00E-05 |
| RL3159 | putative beta-lactamase family protein | - | 1.0 | 0.00021 |
| RL3516 | conserved hypothetical protein | - | 1.1 | 3.00E-05 |

| | | | | |
|---|---|---|---|---|
| RL4161 | conserved hypothetical protein | - | 1.3 | 0.00012 |
| RL4181 | putative TetR family transcriptional regulator | - | 1.4 | 0 |
| RL4725 | putative heat-shock peptidase | - | 1.3 | 8.00E-05 |

**Table 3.2.3.** Genes down-regulated in Rlv3841 in response to hesperetin.

| ID | Function | Name | $log_2$ fold | P value |
|---|---|---|---|---|
| **Chemotaxis related genes** | | | | |
| RL0695 | putative flagellar M-ring protein | *fliF* | -1.3 | 1.00E-05 |
| RL0700 | putative flagellar motor switch protein | *fliG* | -1.1 | 0.00013 |
| RL0701 | putative flagellar motor switch protein | *fliN* | -1.3 | 0 |
| RL0702 | putative flagella motor switch protein | *fliM* | -1.3 | 0 |
| RL0703 | putative chemotaxis motility protein | *motA* | -1.3 | 2.00E-05 |
| RL0704 | putative flagella associated protein | *flgF* | -1.2 | 0.00054 |
| RL0705 | putative flagellum-specific ATP synthase | *fliI* | -1.1 | 0.00017 |
| RL0707 | putative flagellar basal-body rod protein | *flgB* | -1.4 | 1.00E-05 |
| RL0710 | putative flagellar basal body rod protein | *flgG* | -1.5 | 0 |
| RL0711 | putative flagellar protein FlgA precursor | *flgA* | -1.1 | 0.00012 |
| RL0712 | putative flagellar P-ring protein precursor | *flgI* | -1.0 | 0.00015 |
| RL0715 | flagellar basal body-associated protein | *fliL* | -1.2 | 0.00841 |
| RL0718 | putative flagellin A protein | *flaA1* | -1.3 | 0 |
| RL0721 | putative flagellin C | *flaC* | -1.6 | 0 |
| RL0723 | putative motility protein | *motB* | -1.3 | 3.00E-05 |
| RL0724 | putative chemotaxis MotC protein precursor (motility protein c) | *motC* | -1.2 | 0.00257 |
| RL0725 | putative chemotaxis MotD protein (motility protein D) | *motD* | -1.2 | 5.00E-04 |
| RL0727 | putative two-component regulator transcriptional regulatory protein | - | -1.5 | 0 |
| RL0728 | flagellar hook protein | *flgE* | -1.3 | 7.00E-05 |
| RL0729 | hook associated protein | *flgK* | -1.5 | 0 |
| RL0730 | putative flagellar hook-associated protein | *flgL* | -1.1 | 1.00E-05 |
| RL0731 | putative flagellar synthesis related protein | *flaF* | -1.1 | 0 |
| RL0733 | putative hook formation protein | *flgD* | -1.3 | 0 |
| RL4029 | putative chemotaxis protein methyltransferase | *cheR* | -1.1 | 0.00137 |
| RL4032 | putative sensory transducer methyl-acccepting chemotaxis protein | - | -1.1 | 2.00E-05 |
| **Metabolism** | | | | |
| RL1966 | putative alanine dehydrogenase | - | -1.6 | 0.00056 |
| **Adaptation** | | | | |
| RL3302 | putative aquaporin Z | *aqpZ* | -1.1 | 0.00033 |
| **Cell envelope** | | | | |
| RL3297 | putative UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase | *lpxC* | -1.7 | 0 |
| **Others** | | | | |
| pRL110238 | hypothetical protein | - | -1.8 | 0.00632 |
| pRL110531 | hypothetical protein | - | -1.6 | 0.00021 |
| RL0239 | conserved hypothetical protein | - | -1.2 | 3.00E-05 |
| RL0241 | conserved hypothetical protein | - | -1.1 | 9.00E-05 |
| RL0717 | putative conserved hypothetical TPR repeat protein | - | -1.4 | 0 |
| RL0722 | conserved hypothetical protein | - | -1.6 | 1.00E-05 |
| RL0726 | conserved hypothetical exported protein | - | -1.2 | 0.00039 |
| RL0852 | conserved hypothetical protein | - | -1.3 | 0 |
| RL1296 | conserved hypothetical protein | - | -1.4 | 1.00E-05 |
| RL1483 | conserved hypothetical protein | - | -1.1 | 7.00E-05 |
| RL1946 | hypothetical protein | - | -1.4 | 0.005 |
| RL2311 | putative oxygenase, pseudogene | - | -1.5 | 0.0028 |
| RL3172 | hypothetical protein | - | -2.1 | 0 |

### 3.2.5 Transcriptomic analysis on different post inoculation time points

In order to obtain a global view of the Rlv3841 transcriptome in the pea rhizosphere, a time-course analysis of cells recovered after one, three and seven days post-inoculation from 7 day-old pea rhizosphere was performed. Pea seedlings were grown for 7 days as described in section 2.3.3 and inoculated with $10^8$ Rlv3841 cells, grown, harvested (0.6 $OD_{600}$nm) and washed as described in section 2.3.3. After the stipulated time period (1, 3 or 7 days), the cells from the rhizosphere were harvested and total RNA extracted and quantified as described in sections 2.2.8 to 2.2.9. In parallel, Rlv3841 cells were grown in AMS media with 10 mM glucose and 10 mM ammonia as carbon and nitrogen source as described in section 2.1.1. Total RNA was then isolated from the cells and quantified as described in sections 2.2.8 to 2.2.9. Equal amounts of both the free-living and rhizosphere Rlv3841 total RNA was amplified, quantified and as described in sections 2.2.10 and 2.2.9. After amplification, both the RNA samples were then labelled with cydyes Cy3 (control - free-living) and Cy5 (experimental - rhizosphere (7dPea-1dpi, 7dPea-3dpi or 7dPea-7dpi)), depending upon the experiment as described in section 2.4.2. Microarray hybridization was performed with the equal amounts of both control and experimental labeled cDNA as described in sections 2.4.3-2.4.4 for five replicates. The microarray data were then normalized (Figure 3.2.7) and subjected to rigorous statistical analysis in LIMMA as described earlier in section 2.4.5. The genes were considered significant if their adjusted (Benjamini and Hochberg method) P value is $\leq 0.01$ and the average $\log_2$-fold is $\geq 1$ of 5 biological replicates analyzed.

The data from the three time points were then analyzed together as a time-course experiment. Genes were considered as significantly differentially expressed when they passed the above mentioned criteria in at least one of the time points. This enables the identification of genes whose expression changes over time. Comparison of the three dpi time points was performed by Venn mapping to identify commonly expressed genes and those which were only expressed at a specific time (Figure 3.2.8.). In total, 1956 genes were differentially expressed (27% of the genome) including 883 (12% of the genome) up- and 1073

(15% of the genome) down-regulated genes. The up-regulated genes are comprised of 18 functional categories; of which 19% relate to metabolism, 14% to transporters, and 9% to cell envelope and regulation respectively (Figure 3.2.9). This pattern of gene expression reflects the need for the bacteria to be metabolically versatile and able to adapt for life in the rhizosphere.

Comparative analysis of all the dpi time points revealed a core set of 354 genes which were induced at all the time points (Figure 3.2.8.). In the core set, 165 genes were induced more than three-fold in at least one of the time points, suggesting the importance of these genes in rhizosphere adaptation. The core set includes a few previously characterized genes reported to play significant role in rhizosphere colonization of Rlv3841 and/or evolutionarily related α-proteobacteria, many genes of probable functions without experimental evidence and a large proportion of genes of unknown function. Apart from those highly expressed, a large proportion of genes had a slight change in their expression (between 1-2 $\log_2$ fold), which may be due to the low concentration of the available nutrients in the rhizosphere. The up-regulated genes were categorized based on the Riley functional classification (Riley, 1993). It is interesting to note that the genes belonging to the functional category cell division and ribosome constituents were highly repressed. This probably reflects that the rhizosphere cultures were compared to exponentially growing cells in laboratory culture. Here, some of the Rlv3841 genes expressed in the pea rhizosphere, which may play an important role in adaptation, survival and colonization in the pea rhizosphere, are highlighted. As previously stated the expression level will be shown for 1, 3 and 7dpi as *Gene Id* (*Gene name*) ([1dpi], [3dpi], [7dpi]), as values of $\log_2$. The combined expression dataset of 7dPea-137dpi, both up- and down-regulated genes at least in one time point were provided in the accompanying compact disc in the Microsoft Excel 2003 format.

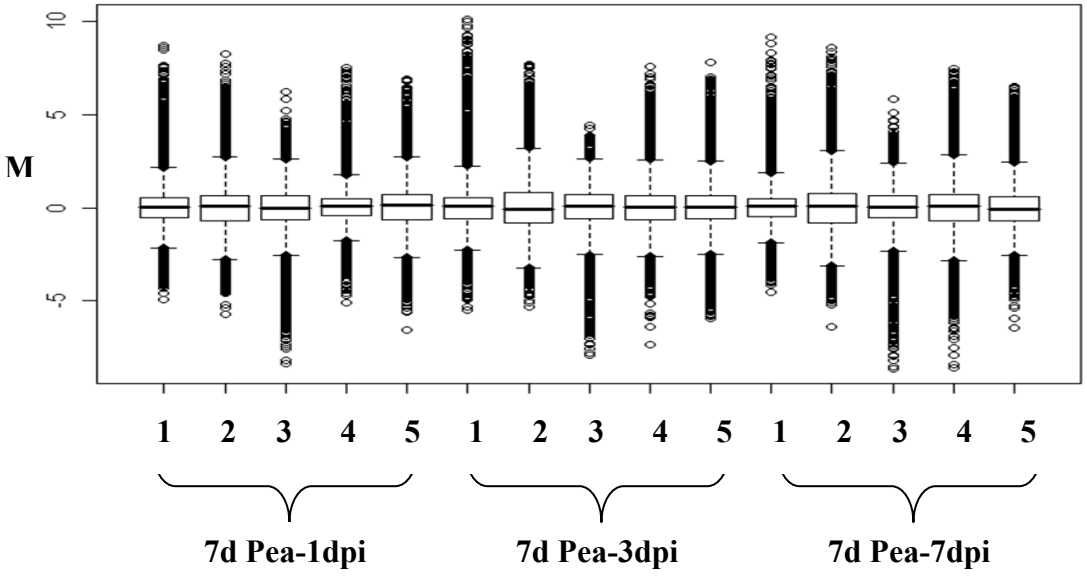



**Figure 3.2.7.** Boxplots showing the differential expression values of the five biological replicates of each (1, 3 or 7dpi) time point pea rhizosphere microarray experiments. M represents the differential expression ($log_2$ (experiment/control) after quantile normalization. In these experiments, the experimental RNA sample was labelled with cy5 and control RNA sample was labelled with cy3 except for the 3rd replicate of each condition, which is a dye swap experiment.

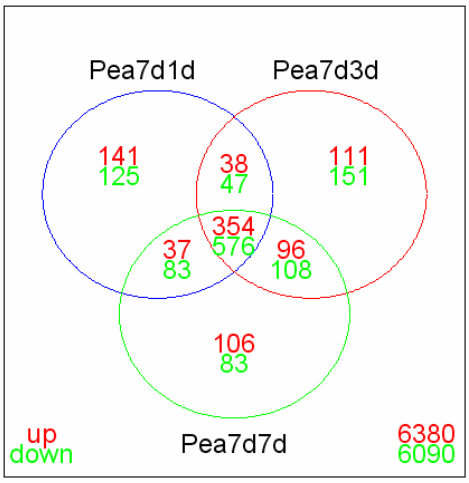



**Figure 3.2.8.** Venn diagram showing the comparison of up- and down-regulated genes of three post-inoculation time-point experiments.

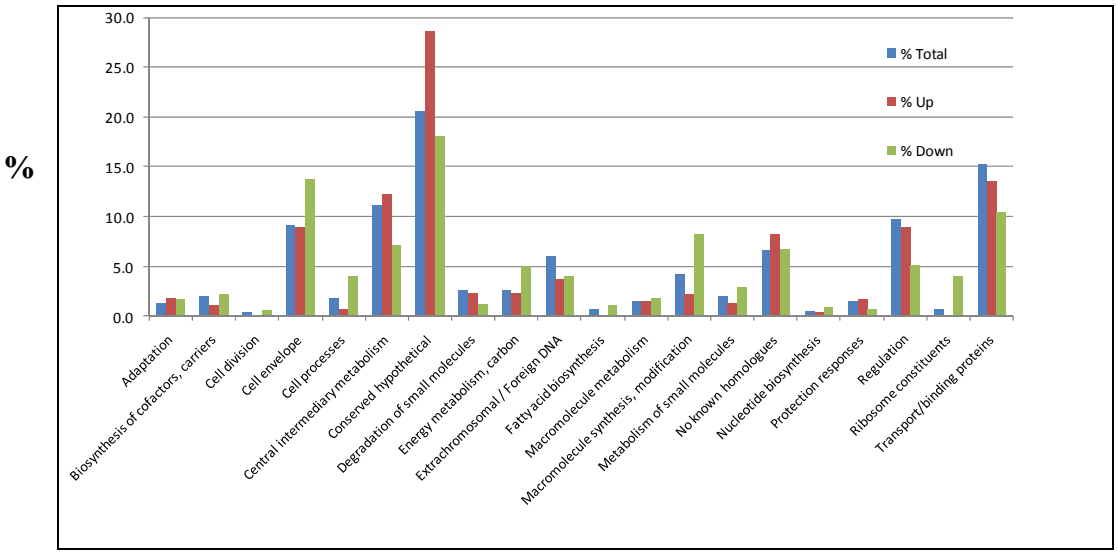



**Figure 3.2.9.** Distribution of functional categories of the genes differentially expressed in at least one of the time points in 7dPea-137dpi experiments, according to Riley classification. The up and down bars represent the percentage of the total up- and down-regulated genes in 7dPea-137dpi combined data and total bar represents the percentage of that functional class in the genome.

### 3.2.5.1 Genes known to be rhizosphere and symbiotic specific

Most of the genes involved in *nod* factor signalling, a well known molecular event in the rhizosphere (Long, 1996) were strongly induced. *pRL100175* (*nodO*) [1.5U,1.5U,1.8U] encodes a nodulation protein, *pRL100180* (*nodM*) [1.8U,1.9U,1.3U] encodes glucosamine fructose-6-phosphate aminotransferase, *pRL100183* (*nodF*) [2.5U,2.6U,1.8U] encodes nodulation protein F, *pRL100185* (*nodA*) [0.8,1.3U,1.6U] encodes nodulation protein A, *pRL100186* (*nodB*) [2.0U,2.3U,2.6U] encodes chitooligosaccharide deacetylase, *pRL100187* (*nodC*) [3.3U,3.3U,3.5U] encodes N-acetylglucosaminyltransferase, *pRL100188* (*nodI*) [1.8U,2.0U,1.8U] encodes nod factor export ATP-binding protein I and *pRL100189* (*nodJ*) [1.4U,2.2U,1.6U] encodes a nodulation protein.

Rhizosphere induced genes (*rhiABCR*), attributed to play an important role in host specificity in the symbiotic interactions (Dibb *et al*., 1984; Economou *et al*., 1989; Cubo *et al*., 1992) was strongly induced. *pRL100169* (*rhiA*) [3.8U,3.8U,3.8U] encodes rhizosphere induced protein RhiA, *pRL100170* (*rhiB*) [4.0U,3.3U,3.0U] encodes rhizosphere induced protein RhiB, *pRL100171* (*rhiC*) [2.0U,1.4U,1.5U] encodes rhizosphere induced protein RhiC and *pRL100172* (*rhiR*) [1.7U,1.0U,0.4] encodes transcriptional regulatory protein. Additionally, *pRL100164* (*rhiI*) [5.1U,5.0U,4.3U] encodes autoinducer synthesis protein which is involved in quorum sensing (Downie *et al.,* personal communication). The detection of the induction of *nod* and *rhi* genes substantiates the success of the strategy developed here.

Quorum sensing is one of the most important types of regulation that bacteria undergo while colonizing a host environment. *RL3378* (*cinI*) [2U,1.1U,1.4U] encodes a protein involved in the production of the *N*-acyl (3-hydroxy-7-*cis*-tetradecenoyl) homoserine lactone (AHL), which plays a key role regulating quorum sensing. The presence of AHL in the rhizosphere was reported to influence nodulation by inducing *rhi* genes expression and also involved in stationary phase adaptation (Lithgow *et al.,* 2000). In addition, *pRL80132* (*traM*) [1.5U,1.1U,0.5U] encodes a transcriptional regulator with high identity (71%) to

TraM (quorum-sensing antiactivator) of *S. meliloti* (Stiens *et al.,* 2007) and to TraM of *A. tumefaciens* (42% Id) (Wang *et al.,* 2006) involved in quorum-sensing.

*pRL120473* (*impC*) [0.6,1.3U,0.9] encodes a virulence-island protein, which was reported to be involved in nitrogen fixation in *R. leguminosarum* strain RBL5523 (Bladergroen *et al.,* 2003). *pRL120065* [1.1U,1.2U,1.1U] is a symbiosis related calsymin, showing high identity (77% Id) to CasA of *R. etli*. Mutants of *casA* in *R. etli* showed partially developed bacteroids, which were significantly affected in symbiotic nitrogen fixation (Xi *et al.,* 2000).

*pRL90019* (*fixK*) [0.7U,1.3U,1.3U] coding for a putative FNR/CRP transcriptional regulator FixK. It is a second copy of FixK located upstream of the *fixNOQP* operon. A mutant of *fixK* showed a reduced rate of nitrogen fixation and *fixK2* showed a $fix^+$ phenotype, whereas the double mutant showed a $fix^-$ phenotype, indicating that FixK2 is important for complete nitrogen fixation (Patschkowski *et al.,* 1996). *RL0532* (*fixI*) [1.0U,1.0U,1.1U] encodes a nitrogen fixation cation ATPase protein, which is located upstream of third *fixNOP* operon. Database similarity search of the amino acid sequence showed moderate identity (<50%) to copper or heavy metal translocating P-type ATPase protein in a wide range of bacteria.

### 3.2.5.2 Genes involved in stress responses and adaptation

In order to survive in the rhizosphere, rhizobia have to cope with different kinds of environmental challenges, ranging from oxidative stress to detoxification of harmful compounds. The induction of many genes which encode general stress related proteins (*RL1366* (*uspF*) [3.3U,2.6U,2.2U], *RL1868* [0.5,1.2U,0.6], *pRL100448* [1.2U,0.6,0.4] and *pRL110421* [1.2U,0.6,1.5U]), putative small heat shock proteins (*pRL120342* (*hspD*) [2.0U,1.9U,2.9U], *RL1883* (*hspF*) [0.5,1.2U,1.9U] and *RL4089* (*ibpA*) [0.4,1.0U,0.9]) (Munchbach *et al.,* 1999) and putative cold shock proteins (*RL3449* (*cspA*) [1.7U,1.3U,0.8] and *RL1298* [0.5,0.8,1.2U]) show that the rhizosphere is a stressful environment. Active functioning of protein secretion systems, including type I secretion systems, is evident from the expression of many genes at 7dPea-1dpi (*RL3911* (*rapB*) [1.2U]

and RL3074 (*rapC*) [1.0U]) which encode type I substrates (Krehenbrink & Downie, 2008). RapBC have been shown to play an important role in root adherence (Ausmees *et al.,* 2001). In addition, *RL0071* (*toaE*) [0.4,1.2U,0.5] encodes a putative HlyD family secretion protein (Krehenbrink & Downie, 2008).

*RL1927* (*autA*) [1.0U,1.0U,0.7] encoding an adhesion autotransporter, classified as Type V secretion system (Krehenbrink & Downie, 2008), was moderately induced at 7dPea-1dpi and 3dpi. The neighbouring genes of unknown function (*RL1924-25* and *RL1928*), which are transcribed in the same direction as *RL1927*, were strongly induced at all time points, indicating that they probably form an operon. It was not possible to predict the function of the neighbouring genes by database similarity searches. The *autA* gene was unique to Rlv3841 and did not have any close homologue in related rhizobia.

Many genes involved in response to oxidative stress were induced. *RL1284* [0.8,1.7U,2.2U] encodes a glutathione-S-transferase generally involved in detoxification of xenobiotics and in a wide variety of biological processes. *RL0025* [-0.2D,0.7,1.1U], *RL2615* [1.2U,0.6,0.9] and *RL4289* [0.4,0.9,1.2U]) and *RL2440* [1.5U,1.2U,1.1U] encoding the antioxidants thioredoxin, glutaredoxin and peroxiredoxin respectively, will alleviate oxidative stress. In addition, *RL2024* (*katE*) [1.7U,0.9,0.8] encodes a monofunctional catalase (HPII). Earlier reports, suggest that the catalase gene was expressed in response to growth in either a poor medium or under oxidative stress (Rediers *et al.,* 2005; Sigaud *et al.,* 1999). Induction of three genes *pRL110014* (*msrA*) [1.9U,2.2U,2.4U], *pRL90212* [1.2U,1.6U,2.1U] and *RL4265* (*msrB*) [1.2U,0.9,1.0U] codes for putative peptide methionine sulfoxide reductase suggests the presence of oxidative stress in the rhizosphere and the ways to cope by Rlv3841.

Rhizobia utilize trehalose in a number of ways either as carbon source, osmoprotectant or stress protector. The expression of *pRL120709* (*treS*) [0.5,0.8,1.2U] encoding trehalose synthase, *pRL110304* (*treX*) [2.2U,0.2,0.7] encoding glycosyl hydrolase (putative glycogen debranching enzyme) and *pRL100466* (*otsB*) [1.3U,0.2,0.4] encoding trehalose phosphatase (Tzvetkov *et al.,* 2003), together with the repression of trehalose transport genes (*thuEFGKAB*),

suggest the biosynthesis of trehalose from maltodextrin. Glycine betaine is also an efficient osmolyte, which can either be taken up from the environment or biosynthesized from choline. The expression of two genes *pRL110429* (*betC*) [1.0U,1.8U,1.5U] and *pRL110491* [1.1U,0.7,0.6] encoding the subunits of putative choline dehydrogenase, involved in glycine betaine biosynthetic pathway, indicates rhizosphere biosynthesis of glycine-betaine from choline (Dominguez-Ferreras *et al.,* 2006; McIntyre *et al.,* 2007; Osteras *et al.,* 1998). These data indicate the biosynthesis of trehalose and choline, probably as a measure to overcome osmotic stress in the rhizosphere.

The rhizobial surface is characterized by many polysaccharides including cyclic glucans, which play a major role in osmotic adaptation. *RL4640* [4.0U,2.7U,2.7U] encodes a putative β-(1→2) glucan export ATP-ABC-T, which has high identity to *ndvA* (76%) of *S. meliloti* 1021, which was shown to be involved in β-(1→2) glucan export from the cell, osmoregulation and in nodule development (Dylan *et al.,* 1990; Stanfield *et al.,* 1988). In addition, *RL4640* was identified by DFI to be induced in pea rhizosphere (Allaway *et al.,* 2001). Moreover, *RL4640* is also homologous to *chvA* (77% Id) of *A. tumefaceins* reported to be involved in glucan export (Cangelosi *et al.,* 2007). Recent STM-based analysis in *S. meliloti* identified a *ndvA* mutant which was hampered in establishing symbiotic interactions with alfalfa (Pobigaylo *et al.,* 2008). *RL0065* [1.2U,1.3U,1.0U] encodes a glycosyl transferase with high identity (86%) to cyclic β (1→2) glucan synthetase (*ndvA*) of *R. leguminosarum* trifolii RtTA1 (Król *et al.,* 2007).

The ability of genes that are involved in detoxification and efflux systems to provide a selective advantage in colonization are often reported in many host-bacterial interactions (Rediers *et al.,* 2005). *pRL90059* (*rmrA*) [2.4U,2.3U,2.4U] encodes a putative type I HlyD-like transporter, which has high identity to RmrA (82%) of *R. etli* CFN42. Studies on *rmrA* of *R. etli* CFN42 showed that it is highly induced in response of bean-released flavonoids and *rmrA* mutants showed high sensitivity to phytoalexins, flavonoids and salicylic acid (Gonzalez-Pasayo & Martinez-Romero, 2000). This suggests a role for *rmrA* in rhizosphere adaptation

by detoxification, which confers a selective advantage in competitiveness and host specificity.

Mimosine, a toxin present in the nodules and root exudates of the tree legume leucaena (*L. leucocephala*) confers a competitive advantage during rhizosphere colonization and nodule occupancy to *Rhizobium* sp. strain TAL1145. The ability to degrade and utilize mimosine is limited to specific rhizobia which colonize leucaena trees. The mimosine transport and catabolic operon are organized as *midABCDR*, where *midABC* codes for an ABC-T system, *midD* for an aminotransferase and *midR*, a regulatory protein which regulates the expression of catabolic genes (Borthakur *et al.,* 2003). Induction was observed of *pRL80060*, *pRL80064* and *pRL80063* which encode components of an ABC-T system with moderate identity (44%, 75% and 79%) with *midABC* of *Rhizobium* sp. strain TAL114.. *pRL80060* (*midA*) [1.7U,3.5U,3.3U] was more strongly induced at all three time points than *pRL80064* (*midB*) [1.5U,2.7U,2.4U] and *pRL80063* (*midC*) [0.7,1.1U,1.3U]. However, *pRL80061* [-1.0D,-1.0D,-1.2D] which has moderate identity (51%) to *midR* of *Rhizobium* sp. strain TAL1145 was repressed. Rlv3841 has genes similar to *midABCR* of *Rhizobium* sp. strain TAL1145, but lacks *midD* which codes for an aminotransferase. This suggests that Rlv3841 may respond to a compound similar to mimosine, but does not need an aminotransferase (i.e. MidD). Alternatively if an aminotransferase is required, it may recruit one from general metabolism.

Pea roots were shown to accumulate high amounts of copper, which is toxic to cells unless it is detoxified. *RL2433* (*copC*) [2.6U,1.6U,1.3U] and *RL2434* (*copA*) [1.0U,0.5,0.4U] were induced in the pea rhizosphere. CopC a is novel multicopper blue protein mostly found in nitrogen fixing symbiotic bacteria, and thought to be involved in periplasmic copper ion transport (Nakamura *et al.,* 2003). Another gene, *pRL110331* [0.6,1.0U,1.3U] encoding a putative copper-transporting P-type ATPase, shows high identity to SMa1013 (*actP*) (73%) encoding copper transport ATPase in *S. meliloti* 1021 (Reeve *et al.,* 2002). Also, it shows moderate identity to PFLU0658 (*cueA*) (45% Id) encoding copper exporter in *P. fluorescens* SBW25 (Zhang & Rainey, 2007). Mutants of *cueA* showed that

copper detoxification is required for rhizosphere competition and survival (Zhang & Rainey, 2007).

*RL4274* [5.8U,6.9U,7.0U] encodes a HlyD transmembrane efflux protein, which showed high identity to RHECIAT_CH0004005 (89%) encoding a putative multidrug efflux protein in *R. etli* CIAT 652. Multidrug efflux proteins are membrane proteins, which pump out toxic compounds from cells. HlyD may be involve in detoxification, which may be one of the host-specific determinants, providing competitive advantage to overcome the selective pressure imposed by the plant in form of toxic compounds, implying its possible role in host specificity.

*R. leguminosarum* bv. *trifolii* produces trifolitoxin (TFX), an antimicrobial peptide that inhibits members of α-proteobacteria, which includes plant symbionts (Robleto *et al.,* 1998). *pRL110604* [1.1U,NE,-0.3D] encodes a putative trifolitoxin related protein. Amino acid sequence analysis showed the presence of a TfuA domain, which is the characteristic feature of genes involved in production of trifolitoxin, suggesting that it might provide selective advantage during rhizosphere colonization.

*pRL120450* (*cpO*) [0.9,2.3U,2.5U] encodes a chloroperoxidase which shows high identity (83%) to a non-heme chloroperoxidase of *A. tumefaciens* C58. In clover, a gene locus (*rwt-1*) was reported to responsible for cultivar specificity and found to interact with *nodM* and non-heme chloroperoxidase of *R. leguminosarum trifolii* ANU794. The precise role of this gene is not known, but may play a role in host specificity (Morris & Djordjevic, 2006).

*pRL90305* [0.4,1.0U,1.4U] encodes a transmembrane MFS family transporter protein with high identity to SMc00990 (*fsr*) (77%) coding for a fosmidomycin resistance protein. RosA of *P. fluorescens* has moderate identity (60%) to *pRL90305* and was reported to be induced in the sugarbeet rhizosphere (Rainey, 1999). *RL1455* [1.1U,0.9,1.4U] encodes an aminoglycoside gentamicin resistance protein. Taken together these observations indicate the selective pressure imposed by the host plant on rhizobia provides a competitive and selective advantage to host-specific rhizobia in the rhizosphere.

### 3.2.5.3 Genes involved in cell process

A large number of genes involved in cell envelope modification indicate a major structural change in Rlv3841 during adaptation to the rhizosphere. A significant number of genes coding for putative transmembrane and outer membrane proteins were induced. Although the specific roles of these genes are unknown, it is clear that it induces an array of membrane proteins to adapt to the rhizosphere. Many genes involved in cell structure modification were often identified by IVET, DFI or STM, and their mutants showed their importance in host interaction (Rediers *et al.,* 2005). *RL0153* [0.9,2.0U,2.3U] encodes a penicillin-binding transpeptidase/transglycosylase protein involved in synthesis of peptidoglycan. Also, *RL1742* [1.9U,1.3U,0.9] encodes an enzyme, N-acetylmuramoyl-L-alanine amidase involved in peptidoglycan synthesis. Orthologues of these genes were identified by IVET and RIVET to be induced in both *M. tuberculosis* and *S. aureus* during interaction with their respective hosts, suggesting a significant role of the bacterial cell envelope in bacterial-host interaction (Rediers *et al.,* 2005).

In addition, *pRL110303* (*cls*) [0.3,1.1U,1.4U] encoding a cardiolipin synthetase, shows high identity to SMc02076 (*cls*) (64% Id) of *S. meliloti* 1021 which is involved in phospholipid biosynthesis (Lopez-Lara *et al.,* 2003). However, RL2226, RL2370 and RL3995 which have high identity with other proteins involved in phospholipid biosynthesis; SMc02096 (CdsA) (60%), SMc00247 (Pcs) (84%) and SMc00611 (MprF) (64%) respectively (Lopez-Lara *et al.,* 2003), were repressed at all time points.

Interestingly, the majority of a large cluster of chemotaxis and motility related genes were repressed in the rhizosphere at all time points, with a few exceptions. These exceptions include *RL0702* (*fliM*) [0.6,1.4U,1.5U] which encodes a flagella motor switch protein with moderate identity to *fliM* (50%) of *S. meliloti* 1021 (reported to be involved in chemotaxis (Sourjik *et al.,* 1998)) and *mcpS* [0.9,1.1U,0.8] encoding an MCP type chemoreceptor. However, chemotaxis genes *fliF* and *flgG*, identified as induced by IVET studies in *P. fluorescens*/sugarbeet (Gal *et al.,* 2003) and *R. leguminosarum* A34/pea

rhizosphere (Barr *et al.*, 2008; Gal *et al.*, 2003), were repressed in this study. In summary, although most of the genes involved in chemotaxis were repressed, a few genes were found to be induced in the rhizosphere implying their essential role in interaction with the host (Rediers *et al.,* 2005). In addition, the apparent repression of these genes in the rhizosphere may reflect comparison with a very motile mid-exponential free-living culture.

*pRL110564* [1.5U,1.5U,1.1U], encoding a putative tight adherence protein, was induced. pRL110564 has high identity to type II secretion system protein E of *R. leguminosarum trifolii* (98%) and *R. etli* CFN42 (88%). However, the exact function of pRL110564 is not clear.

*RL3751* (*pssB*) [3.0U,2.6U,2.2U] has very high identity (99%) to PssB of *R. leguminosarum* VF39. A mutant of *pssB* in *R. leguminosarum* VF39 showed that it is involved in nitrogen fixation (Ivashina *et al.,* 1996). Likewise, *pssB* from *R. leguminosarum trifolii* was reported to play an important role in exopolysaccharide secretions (Janczarek & Skorupska, 2001). *RL0236* [1.1U] encodes a putative endoglucanase (*egl1*) and *RL0237* [1.2U] encodes a putative polysaccharide transport protein (*wzz10*) involved in lipo polysaccharide biosynthesis (*Pss IvB*) (Król *et al.,* 2007), but were induced only at 7dPea-3dpi.

Many genes involved in protein synthesis and degradation were found to be induced during bacterial-host interaction by IVET studies (Rediers *et al.,* 2005). The chaperone are involved in diverse biological process including protein assembly, folding, degradation and transport across membranes (Munchbach *et al.,* 1999). *RL0551* (*hslO*) [1.1U,2.2U,2.2U] encodes a chaperone, *RL4279* (*clpB*) [1.0U,1.2U,1.9U] encodes a chaperone related to heat shock protein and *pRL120643* (*groS*) [1.4U,0.5,1.2U] encodes a 10 KDa chaperone involved in protein folding. Two genes, *RL1690* (*lon*) [0.5,0.7,1.2U] and *RL4570* (*lon*) [1.7U,0.4,0.9] encoding ATP-dependent proteases, were induced. RL1690 has high identity (87%) to Lon protease of *S. meliloti*, which was shown to affect exopolysaccharide synthesis and nitrogen fixation in alfalfa plants (Summers *et al.,* 2000). A Lon protease mutant in *O. carboxidovorans* was also affected in its ability to utilize carbon monoxide as carbon source (Santiago *et al.,* 1999).

In addition, genes encoding for proteolytic enzymes; ATP-dependent proteases (*RL2212 (clpA)* [1.9U,1.4U,1.4U] and *RL2213 (clpS*) [1.4U,1.1U,1.0U]), peptidases (*pRL120625* [2.7U,4.5U,3.8U] and *RL1295* [0.2,0.2,1.9U]) and serine proteases (*RL1251* (*degP*) [2.7U,4.4U,4.6U] and *RL1440* (*degP*) [0.6,1.9U,2.5U]) were induced. ClpA of Rlv3841 shows high identity to ClpA (SMc02109) (84%) of *S. meliloti*, which was identified by STM. A *clpA* mutant in *S. meliloti* showed reduced competitiveness during colonization of the alfalfa rhizosphere (Pobigaylo *et al.,* 2008). Taken together these observations suggest that Lon protease is required for maintenance of the cell and accumulation of unused proteins in the cell will affect normal growth.

**3.2.5.4 Genes involved in regulation**

A total of 139 regulatory genes were differentially expressed, of which 79 were up- and 60 were down- regulated, in at least one time point. Of the up-regulated genes, nineteen genes were induced at all time points. Only a few of these genes have defined functions; *pRL100172* (*rhiR*) [1.7U,1.0U,0.4], *RL3425* (*dctB*) [1.5U,2.1U,1.0U], *RL3829* (*exoY*) [0.3,0.6,1.1U], *RL3766* (*rpoH*) [0.6,1.0U,1.0U], *pRL80132* (*traM*) [1.5U,1.2U,0.6], *pRL110377* (*kdpE*) [1.8U,2.5U,2.4U], while the function of others is unclear.

Two component systems are widely found in prokaryotes and are reported to respond to wide variety of environmental signals in the rhizosphere and/or soil (Ramos-Gonzalez *et al.,* 2005; Silby & Levy, 2004). Recent IVET based studies in *R. leguminosarum* A34 in the pea rhizosphere identified *pRL120793,* a putative repetitive two component sensor histidine-kinase transcriptional regulator to be induced (Barr *et al.,* 2008). Many genes coding for putative two component sensor histidine-kinase transcriptional regulators were induced in the rhizosphere; *RL0030* [1.2U,1.1U,0.9]*, RL0035* [0.7,1.0U,1.1U] and *RL0540* [2.1U,2.5U,2.8U]. *RL0456* [1.4U,1.0U,0.8] encodes a RegM catabolite repressor protein involved in positive regulation of carbohydrate catabolism. Although, the exact function of most these genes are unknown, it strongly suggests the two-component systems mediate many aspects of the response of Rlv3841 to various environmental

signals in the rhizosphere during adaptation. Likewise, only a few of the down-regulated genes have a defined function, including *RL3904* (*pobR*), *pRL120415* (*dadR*), *pRL120319* (*rpoI*), *RL3402* (*rpoD*) and *RL4524* (*rpoE*), while others are unclear. Though the regulatory gene *pobR* (*RL3904*) was repressed, the adjacent gene, *pobA* (*RL3905*), was induced, suggesting the presence of the aromatic substrate in the rhizosphere represses *pobR*. Also, it is interesting to note that *pRL120415* (*dadR*) of *R. leguminosarum* A34 was reported to be induced in the pea rhizosphere (Barr *et al.,* 2008). Interestingly, many genes belonging to particular families of regulators (AraC, AsnC, GntR, LysR and TetR) and genes coding for proteins containing GGDEF domains (*pRL120324* [1.6U,-0.2,0.2], *RL4153* [0.9,1.1U,1.8U] and *RL4717* [1.3U,0.2,0.4]), and GGDEF/EAL domains (*RL0485* [0.7,1.1U,0.6] and *RL1290* [1.3U,0.5,0.6]) were induced. Two GGDEF/EAL containing regulatory proteins (*RL2701* and *RL3958*) were also found to be induced in *R. leguminosarum* A34 in pea rhizosphere by IVET studies (Barr *et al.,* 2008). Although the exact functions of most of these genes are unknown, the GGDEF and EAL domain containing regulatory proteins were reported to play an important role in exopolysaccharide secretion and host-pathogen interactions in a wide variety of bacteria (D'Argenio & Miller, 2004).

#### 3.2.5.5 Genes involved in nutrient acquisition

A total of 241 genes related to transport processes were differentially expressed, of which 119 genes were up- and 122 genes were down-regulated, in at least one time point. Of the up-regulated genes, 37 were induced at all time points, of which 14 genes were induced more than three-fold in at least one time point. Many transporter-related genes were moderately induced between (1-2 $\log_2$-fold), which might be due to the low concentration of the available nutrients in the rhizosphere. The change in induction of these transporters over time, emphasize the fact that the composition of root exudates changes over time. Only a few of the induced transporters are characterized, so to get more insight on the uncharacterized transporters, a similarity search was performed on the *S. meliloti* 1021 transporter genes used in an induction study (Mauchline *et al.,* 2006). Based

on high degree of homology between *S. meliloti* 1021 and Rlv3841 transporters, we re-annotated many genes with the probable function. After re-annotation, it was possible to assign probable functions to 20 genes found to be induced in the rhizosphere (Table. 3.2.4).

| Rlv3841 transport gene ID | Function | Name | *S. meliloti* equivalent gene ID | Identity (%) | Inducer | $log_2$ fold | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | 7dPea 1dpi | 7dPea 3dpi | 7dPea 7dpi |
| pRL100344 | putative nitrate transport component protein | - | SMa0585 | 69 | nitrate | 0.98 | 0.22 | 1.5 |
| pRL100425 | putative ABC transporter component | | SMc03131 | 25 | DL-2 amino adipic acid | | | |
| pRL100426 | putative ABC transporter component | - | | | DL-2 amino adipic acid | 0.41 | 1.02 | 0.651 |
| pRL100427 | putative permease component of ABC transporter | - | | | DL-2 amino adipic acid | 0.30 | 1.24 | 1.411 |
| pRL100440 | putative substrate-binding component of ABC transporter | | SMc04396 | 39 | dextrin | | | |
| pRL100443 | putative ATP-binding component of ABC transporter | - | | | dextrin | 0.71 | 1.07 | 1.03 |
| pRL110207 | putative histidine degradation related amidohydrolase | | SMc00673 | 57 | histidine | | | |
| pRL110212 | putative permease component of ABC transporter | - | | | | 1.00 | 1.63 | 1.515 |
| pRL120047 | putative ATP-binding component of ABC transporter | | Smb20427 | 90 | ecotine | | | |
| pRL120048 | putative substrate-binding component of ABC transporter | | SMb20428 | 85 | ecotine | | | |
| pRL120050 | putative permease component of ABC transporter | *ehuD* | SMb20430 | 78 | ectoine | 0.34 | 2.39 | 0.275 |
| pRL120284 | putative ABC transporter component | - | SMa0203 | 94 | galactose, L-(+) arabinose, glucose, D(-) arabinose, d(+) fucose | 1.25 | 1.19 | 0.686 |
| pRL120593 | putative substrate-binding component of ABC transporter | - | SMa1337 | 90 | CAS | 2.07 | 2.64 | 1.805 |
| pRL90080 | putative periplasmic binding component of ABC transporter | | SMc02154 | 94 | glycerol 3 phosphate, glycerol | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| pRL90182 | putative substrate binding-protein component of transporter | - | | | CAS | 0.63 | 1.04 | 0.67 |
| pRL90185 | putative periplasmic substrate binding component of ABC transporter | | SMc04037 | 81 | CAS | | | |
| pRL90235 | putative permease component of ABC transporter | - | SMb21595 | | arginine | 0.10 | 1.88 | 0.083 |
| RL0091 | putative ATP-binding component of ABC transporter | - | SMc04135 (nodule specific) | 69 | CAS | 1.08 | 0.56 | 0.87 |
| RL0489 | putative substrate-binding component of ABC transporter | - | SMc02171 | 85 | consitutive | 1.03 | 0.96 | 1.08 |
| RL1992 | putative nitrate transport protein | *narK* | | | nitrate | 0.73 | 0.84 | 1.473 |
| RL1993 | putative anionic substrate binding component of ABC transporter | | Smb2114 | 71 | nitrate | | | |
| RL2720 | putative ribose transport system permease protein | *rbsC* | | | dulcitol, tagatose, galactose, sorbose,L-(+) lyxose | 1.48 | 0.52 | 0.841 |
| RL2721 | putative D-ribose-binding protein component of ABC transporter | *rbsB* | SMc21377 | 30 | dulcitol, tagatose, galactose, sorbose,L-(+) lyxose | 1.08 | 0.36 | 0.655 |
| RL3721 | putative permease component of ABC transporter | - | | | | 1.43 | 2.87 | 2.881 |
| RL3723 | putative periplasmic-component of ABC transporter | - | SMc01827 | 69 | uracil, uridine | 1.15 | 0.76 | 0.645 |
| RL3738 | putative ATP-binding component of ABC transporter | - | | | | 0.47 | 0.72 | 1.095 |
| RL3739 | putative permease component of ABC transporter | - | | | | 0.35 | 1.17 | 0.603 |
| RL3741 | putative substrate-binding component of ABC transporter | - | SMb20605 | 83 | nitrate | 1.04 | 1.42 | 1.639 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| RL3804 | putative ATP-binding component of ABC transporter | - | | | | 1.05 | 0.72 | 0.784 |
| RL3805 | putative permease component of ABC transporter protein | - | SMb20416 | | | 0.97 | 2.08 | 2.438 |
| RL3840 | putative substrate-binding protein component of ABC transporter | - | SMb20931 | 91 | raffinose, melibiose, lactose, galactose, lactose, lactulose, meilbiose | 2.09 | 1.24 | 1.276 |
| RL3842 | putative substrate-binding component of ABC transporter | - | | | | 1.37 | 1.06 | 0.877 |
| RL4190 | putative substrate-binding component of ABC transporter | - | SMc01652 | 82 | putrescine/agamatin | 1.22 | 0.60 | 1.27 |
| RL4216 | putative mannitol transmembrane permease component of ABC transporter | *mtlG* | | | dulcitol, sorbitol, mannitol, maltitol | 0.14 | 0.53 | 1.104 |
| RL4217 | putative mannitol transmembrane permease component of ABC transporter | *mtlF* | | | dulcitol, sorbitol, mannitol, maltitol | 0.50 | 0.14 | 1.189 |
| RL4218 | putative mannitol-binding component of ABC transporter | *mtlE* | SMc01496 (sorbitol binding protein) | 86 | dulcitol, sorbitol, mannitol, maltitol | 0.53 | 0.54 | 1.712 |
| RL4655 | putative rhizopine-binding ABC transporter protein | *intA* | SMb20712 | 84 | *myo*-inositol | 0.92 | 1.74 | 1.365 |
| RL4688 | putative molybdenum-pterin binding protein | *mopB* | SMc03196 | 70 | molybdate | 0.81 | 1.01 | 0.82 |

**Table 3.2.4.** Table showing Rlv3841 transport-related genes identical to *S. meliloti* transport-related genes which were found to be induced in the 7dPea-1dpi, 7dPea-3dpi and 7dPea-7dpi rhizosphere. These *S. meliloti* genes were experimentally verified to be induced by the substrate mentioned (Mauchline *et al.*, 2006).

#### 3.2.5.5.1 Sulphur acquisition

*RL2866* [1.1U,0.9,0.7] encodes a putative transmembrane sulphate transporter. *RL2866* along with *RL2865* [4.5U,3.8U,3.1U] and *RL2867* [1.3U,0.9,0.9] encoding proteins of unknown function, may probably form an operon involving in sulphate uptake from the rhizosphere. In addition, *pRL100402* [0.5,1.3U,1.3U] and *RL2769* (*ssuD*) [-0.4,0.1,1.9U] encoding an alkanesulfonate monoxygenases utilizing sulfonates as a sulfur source in the rhizosphere. The components of the sulphate transport systems were reported to be induced in the *R. leguminosarum* A34/pea (Barr *et al.,* 2008) and *P. putida* KT2440/maize rhizosphere (*ssuB*) (Barr *et al.*, 2008; Ramos-Gonzalez *et al.*, 2005).

In addition, *RL1911* [6.4U,4.5U,3.8U] and *RL2756* [4.0U,2.8U,1.7U] encode putative arylsulfatases, while *pRL110429* (*betC*) [1.0U,1.5U,1.8U] encodes a putative choline-sulfatase, which may be involved in the utilization of sulfonates (Kertesz, 2000). Taken together, these observations indicate the possible assimilation of organic sulphur from complex sulfonates. However, Rlv3841 was grown in an *in vitro* pea rhizosphere supplemented with plant growth medium containing sulphate. It implies that sulphonates are utilized in high sulphate-containing medium. So, the further analysis needs to be performed to study the precise roles of these genes.

#### 3.2.5.5.2 Zinc and phosphate limitation

The genes *RL3175-8* encode the Znu ABC uptake system for zinc of the Mzt family in Rlv3841. *RL3177* (*znuC*) [-0.2,-0.1,1.0U], which encodes an ATP binding sub-unit was induced at 7d but the rest of the operon was not induced. *RL3178* (*znuA*), encodes an SBP with high identity to SMc04245 of *S. meliloti* 1021, which was reported to be induced during zinc limitation (Mauchline *et al.,* 2006). Another gene *RL0394* (*phoH*) [1.1U] annotated as phosphate starvation-induced protein was induced at 7dPea-1dpi. Though, the observations indicate that the genes are induced by limiting conditions, it is known that these compounds are present in plant growth medium. However, *znuC* was only expressed only at 7dpi, which may be due to the uptake of zinc by the plants within 6 days causing zinc to

become limiting. These are just possibilities and further analysis needs to be done to identify the reason for the observed induction.

#### 3.2.5.5.3 Nitrogen limitation

Genes (*pRL100344* and *RL1992*) encoding nitrate ABC transporter components (NitT family) were induced. pRL100344 [1.0U,0.2,1.5U] and RL1992 [0.7,0.8,1.5U] show high identity to SMa0585 (69%) and SMb2113 (71%) of *S. meliloti* 1021 respectively, which were reported to be induced by nitrogen limitation in the growth medium (Mauchline *et al.,* 2006). *RL3741* [1.0U,1.4U,1.6U] encoding an SBP of an ABC-T, shows high identity with SMb20605 (HAAT family) (83%) of *S. meliloti* 1021, which was also reported to be induced during growth in nitrogen-limiting medium (Mauchline *et al.,* 2006). It also shares high identity (95%) with RHE_CH03317, annotated as a urea or short chain amide ABC transporter and part of an ABC-T operon *RL3737-41*. The urease gene cluster (structural and accessory genes) is located upstream of this ABC transport system and the region is well conserved in many evolutionary related rhizobia. Additionally, *RL3736* (*ureD*) [0.6,1.1U,1.4U] encodes a urease accessory protein probably involved in the nickel incorporation to the urease enzyme (Toffanin *et al.,* 2002). Taken together these observations suggest that this ABC transport system may be induced by the presence urea in the pea rhizosphere, to facilitate ureolysis. The expression of all these ABC genes suggests nitrogen limitation in the pea rhizosphere.

#### 3.2.5.5.4 Iron Regulation

*RL0777* (*rirA*) [0.6,1.0U,1.1U] encodes the Rhizobium iron regulator. Mutation studies in *rirA* of *R. leguminosarum* revealed that, *rirA* controls at least eight operons involved in synthesis or uptake iron from different iron sources (Todd *et al.,* 2002). Furthermore, mutant studies in *rirA* of *S. meliloti* showed that expression of *rirA* is involved in the repression of iron uptake and siderophore synthesis, but does not affect symbiosis. In addition, *RL3350* [1.3U,1.1U,0.7] encodes an SBP of ABC-T (*RL3350-3*), belonging to the family of ferric iron

uptake (FeT). RL3350 has high identity (72%) to the iron (III) binding periplasmic protein of *B. melitensis* 16M.

**3.2.5.5.5 Quaternary amines (putrescine) uptake**

*RL4190* [1.2U,0.6U,1.3U] encodes an SBP of ABC-T with high identity to SMc01652 (POPT family) (82%) of *S. meliloti* 1021, which was highly induced in the presence of putrescine and agmatine (Mauchline *et al.,* 2006). *RL0765*, [0.3,0.6,1.2U], annotated as a probable putrescine transport system permease protein, was induced at 7dPea-7dpi. *pRL120170* [2.4U,3.9U,3.9U] encodes a permease component of ABC-T and was highly induced in all the time points. pRL120170 has a conserved domain coding for a spermidine/putrescine transport system but there is no close match found in other related rhizobia. In *P. fluorescens*, a gene involved in a putrescine transport system was reported to be induced in the sugarbeet rhizosphere (Gal *et al.,* 2003). Taken together these observations indicate the presence of putrescine in the pea rhizosphere.

**3.2.5.5.6 Nucleotide derivatives uptake**

*RL3721* [1.4U,2.9U,2.9U] and *RL3723* [1.4U,0.8,0.6] encode putative ABC-T components belonging to NitT family of transporters. RL3723 has high identity (69%) to SMc01827 of *S. meliloti* 1021, which was induced by uracil and uridine (Mauchline *et al.,* 2006). These observations indicate the presence of uracil or uridine in the pea rhizosphere.

**3.2.5.5.7 Amino acid uptake**

*pRL120050* [0.3,2.4U,0.3], encoding a permease component of ABC-T, was induced only at 7dPea-3dpi. pRL120050 has high identity to *ehuD* of *R. leguminosarum* bv. *trifolii* (94%) and *S. meliloti* 1021 (78%). In *S. meliloti* 1021, this gene with other genes in the *ehu* gene cluster was reported to be essential for uptake of ectoine/hydroxyectoine and also plays an important role as an osmoprotectant (Jebbar *et al.,* 2005). The adjacent genes *pRL120047-9* encode a complete ABC-T system, that have high identity to SMb20427-9 (*ehuABC*) (90%,

85% and 83% respectively) of *S. meliloti* 1021, but they were not induced in the rhizosphere. SMb20428, was induced by ectoine and classified as a PAAT family transporter (Mauchline *et al.,* 2006).

*pRL120593* [2.1U,2.6U,1.8U] encodes an SBP and *RL0091* [1.1U,0.6,0.9] encodes an ATP binding component of ABC-Ts. pRL120593 and RL0091 show high identity to SMa1337 (90%) and SMc04135 (69%) of *S. meliloti* 1021 respectively, both members of CUT1 family, which were reported to be induced by casamino acids (Mauchline *et al.,* 2006). *pRL90182* [0.6,1.0U,0.7] encodes an SBP from an ABC-T and has high identity to SMc04034 (89%) of *S. meliloti* 1021. *pRL90182-5* probably forms an operon coding for a complete ABC-T system that has high identity to SMc04034-7 in *S. meliloti* 1021 and is located in the same gene order. pRL90185 is classified as member of the PepT family of transporter and has high identity to SMc04037 (81%), which was reported to be induced in the presence of casamino acids (Mauchline *et al.,* 2006). Of these three uptake systems induced by casamino acids, *pRL120593* showed the strongest induction. The induction of these three uptake system suggests the uptake of amino acid or peptide molecule, released by the breakdown of proteins in the rhizosphere.

*pRL80026* [3.2U,4.0U,4.0U], *pRL80027* [0.8,1.5U,1.3U] and *pRL80030* [0.9,1.5U,1.0U] (*livJ, livM* and *livF*), encoding a high-affinity branched amino acid transporter components, were induced. These genes along with *pRL80028* (*livH*) and *pRL80029* (*livG*) probably form an ABC-T operon. Database search of these genes failed to identify any significant homologues even in evolutionarily-close rhizobia, indicating that they may be unique to Rlv3841. It is surprising to see this transport system is present only in Rlv3841, as *liv* transport systems are common in rhizobia. This cluster of genes is located on plasmid pRL8, on which many of the genes are unique to Rlv3841. Taken together, a possible explanation is that this transport system may be mis-annotated as a *liv* transport system, but may transport some other structurally similar amino acids or compound from the rhizosphere, which may be very specific to Rlv3841.

#### 3.2.5.5.8 Sugar uptake

The presence of galactosides in the rhizosphere and their uptake by *S. meliloti* 1021 has been previously reported (Bringhurst *et al.,* 2001; Gage & Long, 1998). Recent research showed that ABC-T SMb20931 (CUT2) of *S. meliloti* 1021, encoding a sugar uptake transporter was induced by both α-galactosides (galactose, raffinose, melibiose) and β-galactosides (lactose and lactulose) (Mauchline *et al.,* 2006). *RL3840* [2.1U,1.2U,1.3U], encoding an SBP of an ABC-T has high identity (91%) to SMb20931 of *S. meliloti* 1021. This indicates the likely presence of galactosides in the pea rhizosphere.

*RL4216* [1.1U], *RL4217* [1.2U] and *RL4218* [1.7U] (*mltGFE*), encoding for mannitol ABC-T, were expressed at 7dPea-7dpi. However, *RL4214* (*mltK*) encoding mannitol dehydrogenase, which catalyzes the conversion of mannitol to fructose, was repressed. RL4218 (MltE) has high identity (86%) to SMc01496, a sorbitol binding protein of *S. meliloti* 1021. A recent study showed that ABC-T SMc01496 (CUT1 family) was induced by $C_6$ polyols (dulcitol, sorbitol, mannitol and maltitol) (Mauchline *et al.,* 2006). *pRL110382* [1.2U], originally annotated as D-xylulose reductase and then reannotated as sorbitol dehydrogenase (*sorD*) (as required during growth in sorbitol), was only induced at 7dPea-7dpi. These observations indicate the possibility that sorbitol may be transported via the mannitol transporter from the pea rhizosphere before being utilized in the cell.

The uptake and utilization of tartrate by *A. vitis* from the grape rhizosphere has been reported (Crouzet & Otten, 1995). *RL0995* [0.8,1.6U,1.6U] (encoding a putative tartrate dehydrogenase), *RL0996* [0.8,2.9U,2.2U] (encoding a putative transmembrane transporter) and *RL0997* [1.4U,1.1U,1.3U] (encoding a putative HTH transcriptional regulator) were induced. These genes form a part of the tartrate transport and catabolic genes *RL0993-7*. These genes are highly conserved and in the same gene order as the tartrate gene cluster of *A. vitis*. RL0993 shows 62% identity with pyruvate kinase (TtuE), RL0994 has 69% identity with hydroxypyruvate reductase (TtuD), RL0995 shows 88% identity with tartrate dehydrogenase (TtuC), RL0996 has 92% identity with tartrate uptake (TtuB) and RL0997, 59% Id with the transcriptional regulator TtuA (Crouzet & Otten, 1995).

A recent study, reannotated *ttuD* as glycerate kinase, which phosphorylates the final product glycerate, enabling it to enter into central carbon metabolism (Yang *et al.,* 2008). These observations indicate the uptake and utilization of tartrate from the pea rhizosphere.

The ability to catabolize certain compounds provides a selective and competitive advantage to rhizobia in the rhizosphere, one such compound is *myo*-inositol (Fry *et al.,* 2001). *RL4655* (*intA*) [1.0U,1.8U,1.4U] codes for SBP of the inositol ABC-T. RL4655 shows high identity to *mocB* (88% Id) of *R. leguminosarum* strain 1a (Bahar *et al.,* 1998) and to SMb20712 (MocB) (84% Id) of *S. meliloti* 1021. *mocB* of *S. meliloti* was shown to be induced by *myo*-inositol (Mauchline *et al.,* 2006).

In addition, *pRL110624* [4.4U,5.7U,5.4U] encoding a protein of unknown function has high identity to SMb20711 (70% Id) and it is located adjacent to *mocB*. Both (pRL110624 and SMb20711) have a conserved IolE domain and are annotated as a conserved hypothetical protein. It also shares moderate identity (49%) with *RSc1245* of *Ralstonia solanacearum* GMI1000, which was annotated as a protein involved in *myo*-inositol catabolism. Interestingly, the protein is coded by the genes upstream of *pRL110624,* pRL110620-23 share high identity with MocF (73%), MocE (88%), MocD (90%) and a putative transcriptional regulator (71%) of *R. leguminosarum* strain 1a. Although this gene cluster was not induced in the rhizosphere, the genes are found as a cluster with a similar gene order in the both organisms. RL1416 [0.5,1.0U,1.1U] is a *myo*-inositol degradation/rhizopine catabolism protein. It shares moderate identity (65%) with RSc1247 (IolI) of *R. solanacearum* GMI1000, which was annotated as gene involved in *myo*-inositol catabolism. In addition, *RL2323* [1.5U,1.9U,1.5U] coding for a putative GFO/IDH/MocA dehydrogenase, which may be involved in rhizopine catabolism, was induced.

Arabinogalactan, which is a complex polysaccharide, is present in large amounts in pea root mucilage and is utilized as a sole carbon source by free-living *R. leguminosarum* 8401 (Knee *et al.,* 2001). *pRL110282* [2.6U,2.3U,2.6U] encodes an enzyme α-N-arabinofuranosidase (*abfA*) which belongs to family 51 of

the glycoside hydrolases (Coutinho, 1999), catalyzing the conversion of arabinan to L-arabinose. This indicates the possible presence of arabinose-containing polysaccharides in the rhizosphere. In addition, *pRL110281* [3.6U,2.8U,2.5U], encoding a putative SBP of an ABC-T, and *pRL110283* [1.9U,2.4U,1.7U], encoding a putative DNA-binding protein flanking *abfA,* probably form an operon. pRL110281 has a DdpA domain which codes for a dipeptide ABC-T, and pRL110283 is very similar to the ArsR family of metalloregulatory transcriptional repressors. However, the exact functions of these genes are unknown. Taken together these data suggest that there may be an active utilization of arabinose from arabinogalactan by Rlv3841 in the pea rhizosphere.

*pRL120284* [1.3U,1.2U,0.7] encodes a ABC transport component (CUT2 family), which shows high identity to SMa0203 (94%), reported to be induced by arabinose and galactose (Mauchline *et al.,* 2006). *RL4229* (*araH*) [1.2U,0.7,1.6U] encodes a L-arabinose transporter permease protein, showing high identity (98%) to *araH* of *R. etli* CFN42. In addition, *RL4212* (*araB*) [1.2U] encodes an enzyme ribulokinase, responsible for catalyzing the conversion of L-ribulose to L-ribulose-5-phosphate (the second step in L-arabinose catabolism), and was induced at 7dPea-7dpi.

Recently, a gene cluster coding for arabinose transport and catabolic genes (*araABCDEF*) was characterized in *S. meliloti*. (Poysti *et al.,* 2007). A database search of this gene cluster in Rlv3841 identified a highly homologous gene cluster (*RL3613-17*) (69%-77% Id), with the same gene order and transcribed in the same orientation, but surprisingly this system was not induced in the pea rhizosphere. Taken together, these observations indicate that the Rlv3841 may have two or more arabinose transport and/or utilization operon, and preference of uptake system may depend on the available nutrients in the rhizosphere.

*RL2720* (*rbsC*) [1.5U,0.5,0.8], encoding an ABC-T permease, and *RL2721* (*rbsB*) [1.1U,0.4,0.7], encoding an SBP of a ribose ABC-T (CUT2 family), were only induced at 7dPea-1dpi. These two genes form an ABC ribose transporter operon with *RL2722* [0.9,0.4,0.4] (*rbsA*), which was induced slightly below the cut-off level at 7dPea-1dpi. Pea seedling exudates have reported to contain ribose

sugars (Roberts *et al.,* 1999; Roberts *et al.,* 2000). However, the slight induction of the ribose uptake genes, only at 7dPea-1dpi, shows weak uptake or possibly indicates a low concentration of ribose in the pea rhizosphere.

*RL0489* (*frcB*) [1.0U,1.0U,1.0U] encodes an SBP and *RL0491* (*frcA*) [1.0U,1.0U,0.7] encodes an ATP-binding component of ABC-T (CUT1 family). RL0489 and RL0491 have high identity to SMc02171 (FrcB) a fructose binding protein and SM02169 (FrcA) an ATP binding cytoplasmic protein respectively. Both are part of the FrcBCAK transporter with a high affinity for fructose (Lambert *et al.,* 2001). The ABC-T operon structure is also conserved in Rlv3841, *RL0489-92* (*frcBCAK*), suggesting the presence of a similar high-affinity fructose uptake system in Rlv3841. Earlier report, showed that pea seedlings exudates consists of high amount of sugars including fructose (Roberts *et al.,* 1999; Roberts *et al.,* 2000). However, slight induction of the genes show weak uptake or low concentration of fructose in the pea rhizosphere.

*RL3804* [1.7U,0.7,0.8] encodes an ATP binding component and *RL3805* [1.0U,2.0U,2.5U] encodes a permease component of ABC-T of the CUT1 family. These genes, together with *RL3806* (permease component) and *RL3807* (SBP) may form a complete ABC transporter operon. However, *RL3806* and *RL3807* were not induced. A database similarity search of RL3804-7 showed high homology with UgpCEAB (70%, 66%, 74%, 63% Id respectively) a sn-glycerol-3-phosphate ABC transport system of *M. loti* MAFF303099. This observation suggests the presence of glycerol-3-phosphate-like compounds in the pea rhizosphere.

*RL3860* [3.2U,4.7U,4.5U] encodes a permease component and *RL3862* [2.7U,1.3U,1.3U] encodes an SBP of an ABC-T belonging to the CUT1 family. These genes together with *RL3859* (ATP binding component) and *RL3861* (permease component) may form a complete ABC transporter operon. However, *RL3859* and *RL3861* were not induced. A database similarity search of the *RL3859-62* showed high identity to many unannotated ABC-Ts in a wide range of bacteria. However, the proteins had moderate identity with those of the maltose transporter *malKGF* (47%, 35% and 26%) and the trehalose/maltose binding

protein (22% Id) of *B. meltensis* 16M respectively. These observations suggest that this ABC-T system may be involved in transport of sugars, probably similar in structure to maltose, from the rhizosphere.

*pRL120200* (*eryG*) [1.3U], encoding an erythritol binding protein, was induced only at 7dPea-7dpi. *eryG* is a part of the erythritol ABC-T operon *pRL120200-202* (*eryEFG*) involved in the uptake of erythritol. Erythritol catabolism was characterized by the catabolic operon *pRL120204-207* (*eryABCD*) which is located downstream of *eryEFG* in *R. leguminosarum* viciae VF39. The erythritol transport and catabolic genes were not induced, but *pRL120203* (*eryH*) [0.9] annotated as putatively involved in erythritol transport or catabolism was slightly induced. Mutants of *eryF* in *R. leguminosarum* viciae VF39 showed slight competitive disadvantage but did not affect the formation of fully-functional nitrogen-fixing nodules (Yost *et al.*, 2006). It is difficult to conclude whether erythritol is present in the pea rhizosphere, but, it suggests that Rlv3841 responds weakly to erythritol, and/or similar compound/s, at a very low concentration in the pea rhizosphere.

**3.2.5.6 Genes involved in aromatic compound degradation**

Rhizobia can utilize a wide range of aromatic compounds as the sole carbon source in the rhizosphere (Stowers, 1985) and the presence of many aromatic compounds in the root exudates has been reported (Cooper *et al*., 2004). *RL3905* (*pobA*) [1.0U,1.2U,1.0U], encoding an enzyme 4-hydroxy benzoate hydroxylase, catalyzes the conversion of 4-hydroxy benzoate to protocatechuate (Wong *et al.,* 1994). Protocatechuate is subsequently catabolized to succinyl CoA and acetyl CoA, via the β-ketoadipate pathway to enter into TCA cycle (Parke *et al.,* 1991) (Figure 3.2.10). Surprisingly, all the genes coding for the enzymes required for the conversion of protocatechuate to succinyl CoA and acetyl CoA were repressed, except *pRL110085* (*pcaB*) (Table 3.2.5). However, *RL3016* (*pcaH2*) [7.5U,7.5U,7.5U] annotated as protocatechuate 3,4 dioxygenase beta subunit was very highly induced. Bioinformatic analysis of RL3016 showed the presence of an intradiol dioxygenase-like conserved domain and a twin arginine

translocation (tat) signal, which also has high degree of similarity to a putative intradiol dioxygenase in *R. leguminosarum trifolii* (93% Id), *R. etli* CFN42 (88% Id) and *Rhizobium sp*. NGR234 (54% Id), but absent in other rhizobia. This shows that this gene may have been acquired during the evolutionary process in which rhizobia diversified.

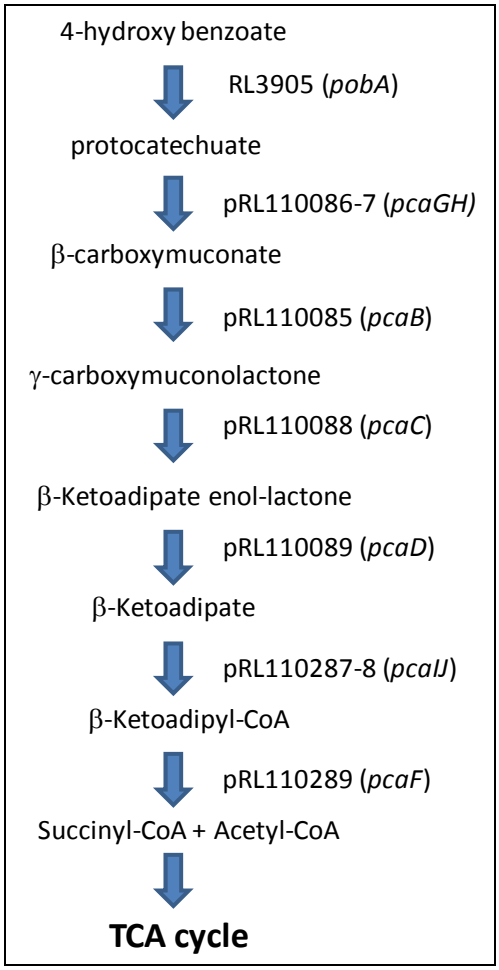



**Figure 3.2.10**. Pathway showing the catabolism of 4-hydroxy benzoate via β-ketoadipate pathway.

*pRL110085* (*pcaB*) [0.7,1.2U,0.7], from the protocatechuate catabolic gene (*pca*) cluster (*pRL110085-90*), encodes a cycloisomerase. pRL110085 has high identity (65%) to PcaB of *S. meliloti* and in other related rhizobia, which were shown to catalyze the second step in the protocatechuate catabolism

(MacLean *et al.,* 2006). The protocatechuate catabolic genes are clustered in Rlv3841 and have synteny to *S. meliloti*.

*pRL110611* [0.5,1.6U,1.5U] encodes a putative flavonol / synthase dioxygenase. pRL110611 shows moderate identity to (40-60% Id) to putative dioxygenases in wide variety of bacteria. KEGG (PATH: rle00362) shows that this protein may involve in the conversion of protocatechuate to hydroxyquinol, which is then converted to succinyl CoA and acetyl CoA via maleylacetate, before entering into TCA cycle.

*pRL110484* [2.5U,1.7U,2.6U] encodes a protein of dienelactone hydrolase family with high identity to dienelactone hydrolases in a wide range of bacteria, including rhizobia. InterPro annotation suggests that it might play an important role in chlorocatechol degradation (InterPro: IPR002925). This gene was also found to be induced when grown on 4-hydroxy benzoate and protocatechuate as the sole carbon source.

Maleylacetate reductase plays an important role in chloroaromatic compound catabolism, catalyzing the conversion of maleylacetate to ketoadipate, the last step in the catabolic pathway. Rlv3841 has two copies of genes *pRL120100* and *pRL120215* encoding maleylacetate reductase (*tftE*). The first gene, *pRL120215* is a part of the gene cluster (*pRL120213-20*) with high identity (53-85% Id) to the γ-resorcylate (2,6-dihydroxybenzoate) catabolic gene cluster in *Rhizobium* sp MTP-10005, but was repressed in the pea rhizosphere. The second gene *pRL120100* (*tftE*) [0.5,1.0U,0.4] encoding maleylacetate reductase was only induced at 7dpea-3dpi, along with *pobA* and *pcaB*. pRL120100 shows moderate identity (52%) to *macA* in *Ralstonia eutropha* 335 which was reported to be induced when grown on 4-flurobenzoate (Seibert *et al.,* 2004). *pRL120100* is part of a large operon involved in aromatic compound uptake (*pRL120095-99*) and catabolism (*pRL120093, 94 and 100*).

In rhizobia, 4-hydroxy benzoate is catabolized to protocatechuate by PobA and then protocatechuate is catabolized to succinyl CoA and acetyl CoA by the Pca proteins via the β-ketoadipate pathway, to feed into TCA cycle. Interestingly, in free-living Rlv3841, *pcaB* was found to be expressed together with *pob* and *pca*

genes clusters when grown on 4-hydroxy benzoate (3 mM), but not induced when grown on protocatechuate (3 mM) as sole carbon source (Table 3.2.5).

This suggests that *pcaB* is involved in catabolism of 4-hydroxy benzoate, but it is not necessary for protocatechuate catabolism. The other genes *RL3016*, *pRL110611* and *pRL120100* were not induced in free-living cells when grown on 4-hydroxy benzoate or protocatechuate as sole carbon source. Also, the genes involved in uptake of 4-hydroxy benzoate and protocatechuate was not induced in the 7d pea rhizosphere (Table 3.2.5).

Taken together, these observations suggest that Rlv3841 responds weakly to aromatic compounds similar to 4-hydroxybenzoate in the pea rhizosphere, which is catabolized by an unknown pathway. However, the concentration of compounds may not be at a level to induce catabolic genes or the catabolic genes may have been repressed by presence of organic acids or readily utilizable compounds.

| ID | Name | Function | PC | 4HB | 7dPea ($log_2$ fold) | | |
|---|---|---|---|---|---|---|---|
| | | | | | 1dpi | 3dpi | 7dpi |
| RL3904 | *pobR* | putative AraC family transcriptional regulator for 4-hydroxybenzoate hydroxylase | 0 | 0 | -0.8 | -1.2 | -1.0 |
| RL3905 | *pobA* | putative -hydroxybenzoate hydroxylase | 0 | **2.4** | **1.0** | **1.2** | **1.0** |
| RL3906 | | putative solute-binding component of ABC transporter | **1.6** | **3.4** | -0.1 | -0.7 | -0.3 |
| RL3907 | | putative permease component of ABC transporter | 0 | **2.4** | 0.3 | 0.3 | 0.0 |
| RL3908 | | putative permease component of ABC transporter | 0.5 | **3.0** | -0.4 | -1.2 | -0.6 |
| RL3909 | | putative ATP-binding component of ABC transporter | -0.02 | **1.5** | 0.3 | 0.2 | 0.3 |
| RL3910 | | putative ATP-binding component of ABC transporter | 0.3 | **1.7** | 0.08 | 0.8 | 0.3 |
| | | | | | | | |
| pRL110085 | *pcaB* | putative cycloisomerase | **0.4** | **2** | **0.7** | **1.2** | **0.7** |
| pRL110086 | *pcaG* | putative protocatechuate 3,4-dioxygenase alpha chain | 1.7 | 2.8 | 0.2 | 0.4 | 0.5 |
| pRL110087 | *pcaH* | putative protocatechuate 3,4-dioxygenase beta chain | 2.6 | 3.8 | -0.3 | -1.0 | -0.8 |
| pRL110088 | *pcaC* | putative carboxymuconolactone decarboxylase | 1.7 | 2.8 | 0.2 | -1.0 | -0.4 |
| pRL110089 | *pcaD* | putative β-ketoadipate enol-lactone hydrolase | 2.6 | 3.7 | 0.0 | -0.6 | -0.5 |
| pRL110090 | *pcaQ* | putative LysR family transcriptional regulator PcaQ | 0.4 | 0 | 0.1 | 0.7 | 0.3 |
| pRL110287 | *pcaI* | putative 3-oxoadipate CoA-transferase subunit A | 1.7 | 2.6 | 0.3 | 0.4 | 0.2 |
| pRL110288 | *pcaJ* | putative 3-oxoadipate CoA-transferase subunit B | 1.8 | 2.9 | -0.1 | -0.3 | -0.4 |
| pRL110289 | *pcaF* | putative beta-ketoadipyl CoA thiolase | 1.3 | 2.3 | 0.4 | 0.6 | 0.2 |
| | | | | | | | |
| pRL110611 | | putative flavonol synthase/dioxygenase | **0.0** | **0.1** | **0.5** | **1.6** | **1.4** |
| pRL110484 | | putative dienelactone hydrolase family protein | **2.8** | **1.7** | **2.6** | **1.6** | **2.7** |

**Table 3.2.5.** Table showing the expression values ($log_2$) of *pob* and *pca* genes differentially expressed in free-living Rlv3841 grown in 3mM of protocatechuate (PC), 3mM 4-hydroxy benzoate (4HB) and rhizosphere recovered Rlv3841; 7dPea-1dpi (1dpi), 7dPea-3dpi (3dpi) and 7dPea-7dpi (7dpi).

### 3.2.5.7 Genes involved in carbon and energy metabolism

With weak induction of many sugar uptake systems, it was expected to see an active carbon metabolism. The dicarboxylate transport system *RL3424* (*dctA*) [2.4U,5.4U,4.8U] and *RL3425* (*dctB*) [1.6U,2.2U,1.1U] was also highly induced, indicating the active transport of dicarboxylates.

The two enzymes, phosphoenolpyruvate carboxykinase (PEPCK) and fructose bisphosphate aldolase were reported to be induced when grown on organic acids as sole carbon source in *R. leguminosarum* MNF3841. PEPCK mutant in *R. leguminosarum* MNF3841 failed to grow on wide range of simple carbon compounds (McKay *et al*., 1985). Rlv3841 genome consists of one gene (*RL0037*) encoding for PEPCK and two genes (*RL4012* and *pRL120196*) annotated as putative fructose-biphosphate aldolase.

RL0037 (PckA) [4.3U,3.0U,2.8U] is a PEPCK, essential for the gluconeogenic pathway was highly induced. *S. meliloti pckA* mutants produced nodules with reduced nitrogen fixing ability (Finan *et al.,* 1991). In addition, transcriptional *gusA* fusion studies of *S. meliloti* showed the expression of *pckA* inside the nodules (Oke & Long, 1999) (Figure 3.2.11.).

RL4012 [2.5U,1.8U,1.5U] is a putative fructose bisphosphate aldolase, which normally catalyzes the breakdown of fructose-1,6-biphosphate to glyceraldehyde-3-phosphate and dihydroxyacetone phosphate in glycolysis. Another enzyme, pRL120196 [-0.6,-0.6,-0.6] annotated as fructose bisphosphate aldolase was not expressed. However, *RL4012* was highly induced in free-living cells grown on pyruvate or succinate as the sole carbon source compared to cells grown on glucose as the sole carbon source (Karunakaran *et al*, personal communication), suggesting that RL4012 is dedicated to the gluconeogenic pathway (Figure 3.2.11.).

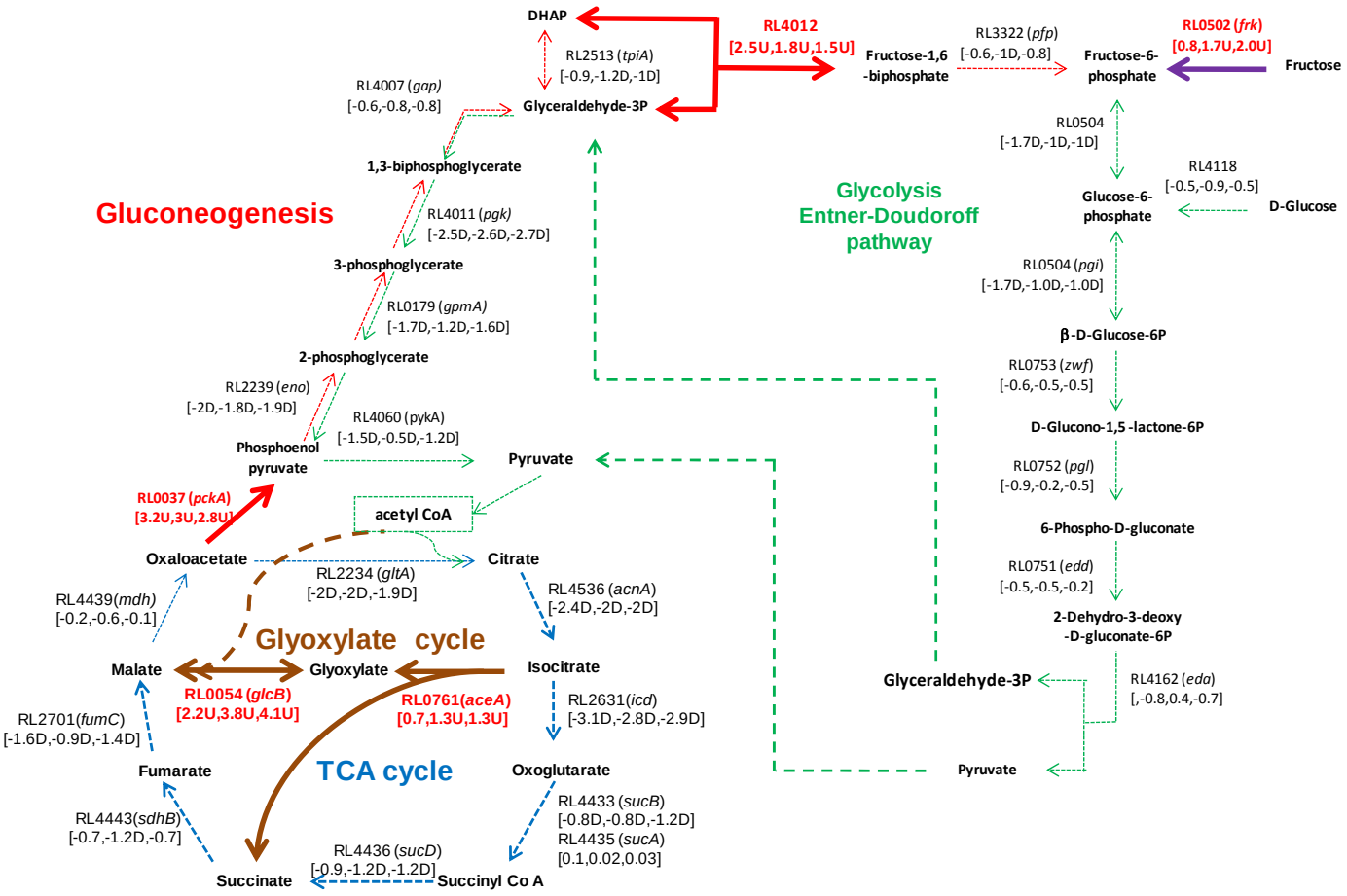



**Figure 3.2.11**. Central carbon metabolic pathway (glycolysis (ED pathway) in green, gluconeogenesis in red, glyoxylate in brown and TCA cycle in blue). Rlv3841 Genes encoding enzymes which were highly expressed in the pea rhizosphere are shown in red.

RL3130 [5.5U,6.5U,6.8U], annotated as a phosphatase, shares high amino acid identity (84%) with a protein annotated as an inositol monophosphatase in *R. leguminosarum* bv. *trifolii* WSM2024 and a wide variety of bacteria. It also shares a moderate identity (35-48%) with many archaeal fructose-1,6-bisphophatases, which catalyze the conversion fructose-1,6,-bisphosphate to fructose-6-phosphate in gluconeogenesis. Based on its moderate identity with archaeal fructose-1,6-bisphophatases, it was thought that this gene may encode for the missing fructose-1,6-phosphatase in gluconeogenesis. However, *RL3130* was not expressed in cells grown on pyruvate or succinate as the sole carbon source compared to cells grown on glucose as the sole carbon source (Karunakaran *et al*, personal communication), which rules out the possibility of its role in gluconeogenesis. A KEGG analysis shows that this gene may be involved in inositol phosphate metabolism (PATH: rle00562).

RL0761 (AceA) [0.7,1.3U,1.3U] is an isocitrate lyase and RL0054 (GlcB) [2.2U,3.8U,4.1U] is malate synthase G, two enzymes involved in glyoxylate pathway and strongly induced in the pea rhizosphere. The enzymes involved in glyoxylate and gluconeogenic pathway were identified by IVET to be induced during many host-bacterial interactions (Rediers *et al.,* 2005) (Figure 3.2.11.).

The uptake of fructose from the pea rhizosphere by Rlv3841 was indicated by the induction of *RL0489* (*frcB*) and *RL0491* (*frcA*). *RL0502* (*frk*) [0.8U,1.7U,2.0U] encodes fructokinase catalyzing the conversion of fructose to fructose-6-phosphate, which subsequently enters into glycolysis (Figure 3.2.11.).

The enzymes transketolase and transaldolase were identified in a number of bacteria during interaction with their host by IVET or DFI-based studies (Rediers *et al.,* 2005), indicating their expression during establishment of an interaction with the host. Three genes involved in non-oxidative pentose phosphate pathway were induced. *RL0066* [2.0U,1.5U,1.0U] encoding a xylulose-5-phosphate phosphoketolase, which catalyzes the conversion of xylulose-5-phosphate to glyceraldehyde-3-phosphate. *RL2718* [1.6U,0.9,0.9] encoding transketolase and *RL2719* [2.9U,3.3U,2.2U] encoding ferrodoxin (transketolase family) were induced. It is notable that the genes involved in ribose transport

systems were also induced at 7dPea-1dpi, showing a correlation between uptake and utilization. This gene cluster is located between the iron and ribose transport system and the whole region is conserved in most of the evolutionarily-related bacteria.

Earlier reports, showed the presence of galactose in pea root mucilage (Knee *et al.,* 2001) and in the alfalfa rhizosphere (Bringhurst *et al.,* 2001). The induction of the galactose-uptake system indicates its presence and utilization by Rlv3841 in the pea rhizosphere (see section 3.2.5.5.8) Two genes involved in the galactose metabolic pathway as described in KEGG (PATH: rle00052) were induced. *pRL120559* [1.4U,1.8U,1.2U] encoding a glycosyl hydrolase, catalyzing the conversion of raffinose to D-fructose and *pRL120249* (*agaZ*) [0.3,1.0U,0.3U] encoding tagatose-6-phosphate kinase, catalyzing the conversion of D-tagatose-6-phosphate to D-tagatose-1,6-biphosphate. The gene coding for the enzyme D-tagtose-1,6-biphosphate aldolase, which catalyzes the conversion of D-tagatose-1,6-biphosphate into glyceraldehyde-3-phosphate, was absent in Rlv3841.

*RL4267* (*acoD*) [4.8U,4.5U,4.6U], encoding a putative acetaldehyde dehydrogenase, was strongly induced and it's amino acid sequence shows high identity to acetaldehyde dehydrogenases in a wide variety of bacteria. It also shows high identity (84%) to chloroacetaldehye dehydrogenase of *Xanthobacter autotropicus,* which was reported to be involved in conversion of chloroacetaldehyde to chloroacetic acid. Subsequently, chloroacetic acid is converted into glycolate before entering glyoxylate metabolism, in the 1,2-dichloroethane degradative pathway (Van Der Ploeg *et al.,* 1994). *RL4046* [2.0U,1.4U,1.7U], encoding a conserved hypothetical exported tetratricopeptide repeat (TPR) protein, and *RL4047* [3.0U,1.0U,1.3U], encoding a putative esterase/hydrolase, were strongly induced and are found in many bacterial genomes. RL4047 shows limited identity (26%) to the putative esterase of *R. leguminosarum* A34, a homologue of RHE_CH01085 of *R. etli* CFN42 involved in aromatic compound metabolism, which was reported to be induced in pea rhizosphere by IVET studies (Barr *et al.,* 2008). A KEGG database search of this gene suggests that it might also be involved in the 1,2-dichloroethane degradative

pathway, catalyzing the first step the conversion of 1,2-dichloroethane to 2-chloroethanol. Additionally, *RL0866* (*glcF*) [0.7U,1.4U,1.0U], coding for a putative glycolate oxidase iron-sulfur subunit which catalyzes the conversion of glycolate to glyoxylate, was also induced. All these genes well fit with the 1,2-dichloroethane degradative pathway as described in KEGG (PATH: rle00631) (Figure.3.2.12). Although, there is no report of the occurrence of 1,2-dichloroethane in the rhizosphere, these observations might indicate the possible presence of 1,2-dichloroethane or similar halogenated compounds.

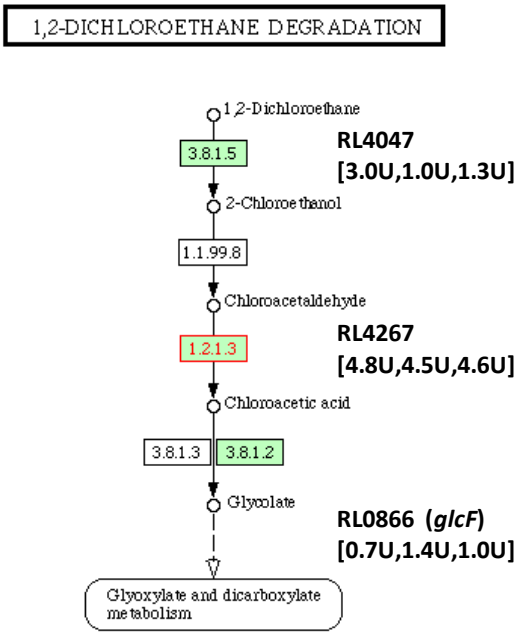



**Figure 3.2.12**. 1,2-dichloroethane degradative pathway (adapted from KEGG - PATH: rle00631).

The expression of genes *coxG* and *coxMSL* coding for carbon monoxide dehydrogenase (CODH), a key enzyme in carbon monoxide (CO) metabolism suggests an aerobic chemolithoautotropic utilization of CO as a carbon source. The utilization of CO as the sole carbon source has been reported in many organisms including *Oligotropha carboxidovorans* (Schubel *et al.,* 1995) and *B. japonicum* (Lorite *et al.,* 2000). The gene cluster involved in CO utilization has been well characterized in *O. carboxidovorans*, which consists of core structural genes *coxMSL* (coding for CODH), flanked by accessory genes (Santiago *et al.,* 1999). The CODH from *O. carboxidovorans* is well studied; it is an $O_2$-stable molybdenum-iron-sulfur-flavin hydroxylase which catalyzes the oxidation of CO to $CO_2$, as described in the equation below (Lorite *et al.,* 2000),

$$CO + H_2O \rightarrow CO_2 + 2e^- + 2H^+$$

In the pea rhizosphere, *pRL80023* (*coxM*) [2.3U,3.3U,3.0U], *pRL80024* [2.4U,3.4U,3.1U] (*coxS*) and *pRL80025* [2.3U,3.3U,2.9U] (*coxL*) which are transcribed in the same direction, were highly induced. The amino acid sequences of these genes show only a moderate identity with *coxMSL* of *O. carboxidovorans*, but have the highly conserved motifs which are characteristic of *coxMSL* (Santiago *et al.,* 1999). The characteristic motifs of *O. carboxidovorans* CoxMSL are as follows; CoxM has three FAD pyrophosphate binding sites ($^{32}$AGGHS$^{36}$, $^{111}$TIGG$^{114}$, $^{193}$Y), CoxS has a type I 2Fe:2S ($^{102}CX_2CX_{31}CNC^{139}$) and a type II 2Fe:2S ($^{42}CX_4CX_2CX_{11}C^{62}$) site, CoxL contains one S-selanylcysteine loop ( $^{388}$VAYRCSFR$^{391}$) and molybdopterin cysteine dinucleotide (MCD) contacting segments ($^{240}$Q, $^{270}$G, $^{528}$QGQGHETY$^{535}$, $^{569}$GSRST$^{573}$,$^{586}$CGTRIN$^{591}$, $^{761}$VGE$^{763}$) (Santiago *et al.,* 1999). Many genes flanking *coxMSL* of unknown function were induced including, *pRL80021* (*coxG*) [3.0U,4.0U,3.9U] and *pRL80038* [1.4U,1.4U,1.5U] showing similarity to *coxI*. The study of a *coxI* mutant of *O. carboxidovorans* suggests that it might be involved in the interaction of CODH with the cytoplasmic membrane. These observations indicate the presence of a CO-utilization gene cluster and flanking genes of unknown function which may

contribute to accessory genes (including *coxI* and *coxG*) as in *O. carboxidovorans*. Any further utilization of the $CO_2$ obtained from the oxidation of CO, is not clear. It is reported that different carboxidotrophic bacteria utilizes the product of CO metabolism in different ways (Lorite *et al.,* 2000).

Formate, a one-carbon compound, is excreted by many plants and bacterial species and accumulates in the environment. Formate dehydrogenase, a key enzyme in the catabolism of formate, catalyzes the conversion of formate to $CO_2$ and $H^+$ with release of electrons. A set of genes *RL4391* (*fdsA)* [1.0U,0.8U,1.0U], *RL4392 (fdsB)* [0.8U,1.0U,0.8U], *RL4393* [3.2U,2.3U,2.6U] *(fdsG*) encoding α, β, γ subunits of NAD-dependent formate dehydrogenase were induced (*fdsG* was highly induced). Database similarity search of FdsABG sequences of Rlv3841 showed high identity to FdsA (77%), FdsB (68%), FdsG (60%) of *R. eutropha* H16, which were shown to be induced in the presence of formate (Oh & Bowien, 1998). Also, FdsABG of Rlv3841 showed high identity to FdsA (SMc04444, 89%), FdsB (SMc02525, 84%) and FdsG (SMc02524, 75%) of *S. meliloti* 1021, respectively. Mutant studies in of *S. meliloti* 1021 showed that *fdsABCDG* operon is essential for formate-dependent autotrophic growth, along with *cbb* operon. It was shown that formate was taken into the cell by diffusion and oxidized to $CO_2$. $CO_2$ is then reduced via Calvin-Benson cycle utilizing RuBisCO and the carbon is converted to glyceraldehyde-3-phosphate, which is subsequently isomerised by triose phosphate isomerase to dihydroxyacetone phosphate. A hexose is synthesized via an aldolase reaction yielding fructose 1,6 bisphosphate (Pickering & Oresnik, 2008). Rlv3841 may oxidize formate to $CO_2$ using formate dehydrogenase, but it lacks Calvin-Benson cycle genes to further reduce $CO_2$ to use as a carbon source. However, *pRL120396* (*cbbL1*) [0.6,0.9,0.6], annotated as encoding ribulose bisphosphate carboxylase (large chain 1) which shows high identity (87%) to SMb20393 (RbcL) (putative ribulose bisphosphate carboxylaseoxygenase) of *S. meliloti* 1021, was not expressed. Rlv3841 may have another mechanism to utilize the $CO_2$ obtained from CO and formate utilization.

Cytochrome c oxidase is the component of respiratory chain that catalyzes the reduction of oxygen to water. The cluster of genes, *RL3041* [3.5U,1.9U,2.2U]

(encoding a putative cytochrome c oxidase subunit IV), *RL3042* (*ctaE2*) [1.7U,1.0U,1.4U] (encoding a putative transmembrane cytochrome oxidase subunit), *RL3043* (*ctaE1*) [1.0U,1.3U,0.7] (encoding a putative cytochrome c oxidase subunit III), *RL3044* (*ctaD*) [1.6U,0.6,0.8] (encoding a putative cytochrome c oxidase subunit I), *RL3045* (*ctaC*) [2.8U,2.6U,2.7U] (encoding a putative cytochrome c oxidase subunit II) and *RL3046* [3.7U,2.8U,2.3U] (encoding a putative exported cytochrome c, SignalP predicts a signal peptide cleavage at residue 23 with a P of 0.998) were highly induced. These genes code for cytochrome c oxidase subunits and transmembrane components (Cta) with RL3046 the putative periplasmic electron donor to Cta. Rlv3841 is known to express $cytaa_3$ as the main cytochrome oxidase in free-living bacteria (Delgado *et al*., 1998). In legume nodules most rhizobia, including Rlv3841, express the *fixNOQP* genes to produce the high affinity $cbb_3$ cytochrome oxidase (Delgado *et al*., 1998). The expression of the Cta cytochrome oxidase in the rhizosphere shows the presence of a novel system and indicates a distinct redox environment.

Though many ABC-T genes involved in uptake of carbon sources were induced in the rhizosphere, in most cases the catabolic genes involved in the utilization of those compounds were not induced. The ABC-T components may be induced by the presence of the respective compound in the rhizosphere but the preference for utilization may be dependent on the rhizobia in the rhizosphere. Induction of the many ABC-T reflects the wide range of compounds available to the bacteria in the pea rhizosphere. Based on the obtained data, a model of the metabolic pathway of Rlv3841 in the rhizosphere is proposed (Figure: 3.2.13).

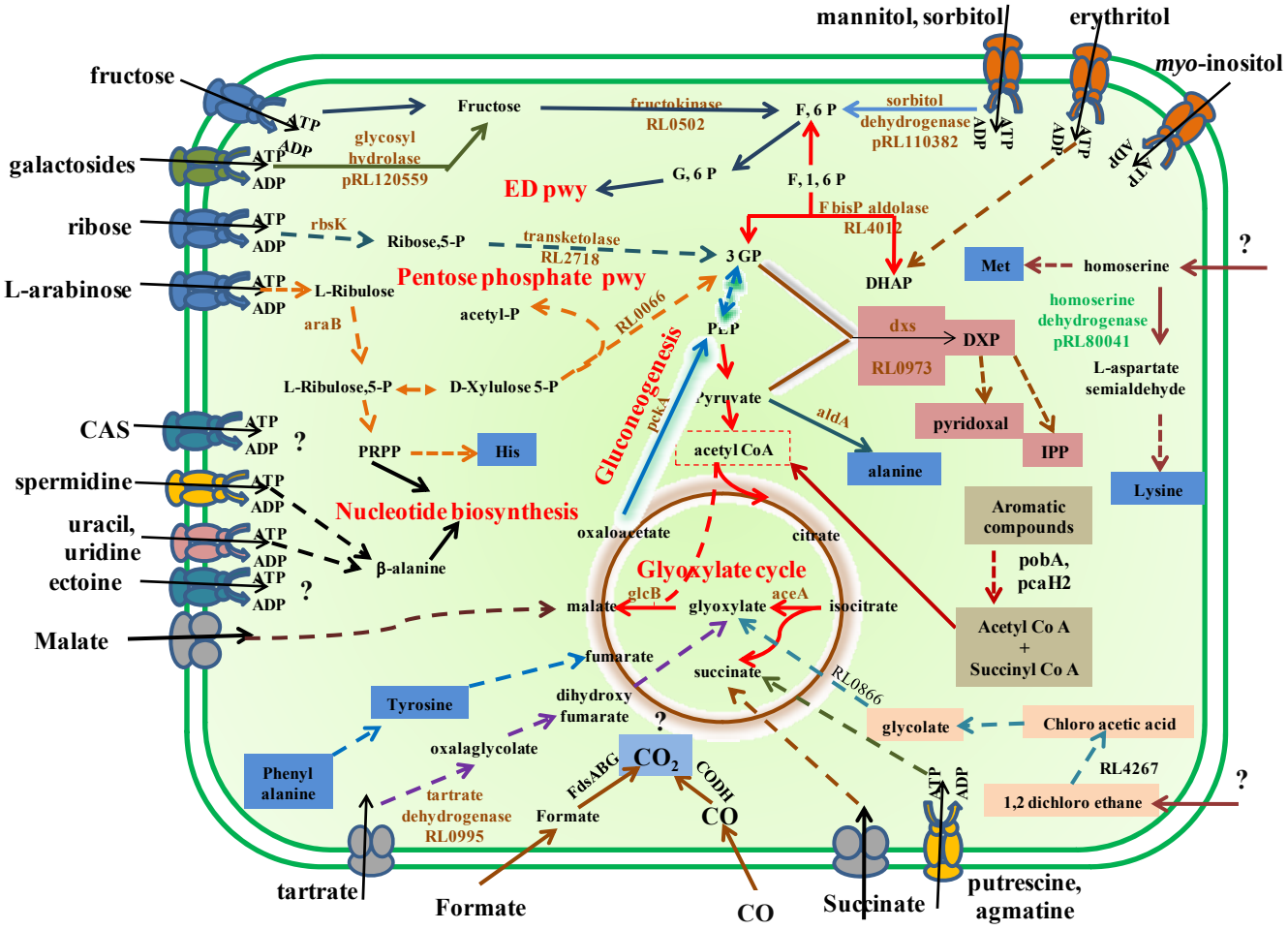



**Figure 3.2.13.** Model of the metabolic pathways in Rlv3841 in the pea rhizosphere based on the 7dPea-137dpi microarray data.

### 3.2.5.8 Genes involved in nucleotide metabolism

Genes involved in nucleotide metabolism include; *RL3708* (*pydA*) [1.1U,1.7U,1.3U], encoding dihydropyrimidine dehydrogenase, *RL3715* [1.1U,0.2,0.2], encoding a TetR transcriptional regulator, *RL3716* (*atcC*) [2.1U,1.5U,1.2U] encoding an N-carbamyl-L-cysteine amidohydrolase and *RL3718* (*dht*) [1.1U,0.0,0.4], encoding D-hydantoinase. *pRL120308* (*codA*) [0.6,1.1U,0.8] encodes cytosine deaminase which catalyzes the conversion of cytosine to uracil and *RL0572* [2.5U,3.0U,3.2U], encoding dihydroorotate dehydrogenase, were induced. Expression of components of the uracil/uridine ABC-T (RL3721 and RL3723) by Rlv3841 in the pea rhizosphere has already discussed section 3.2.5.5.6. KEGG mapping of these enzymes suggest they may be actively involved in two pathways; 1) the conversion of uracil to β-alanine and then to alanine and aspartate metabolism (PATH: rle00410) (Figure 3.2.14), and 2) conversion of β-alanine to uracil and then into pyrimidine metabolism (PATH: rle00240) (Figure 3.2.15).

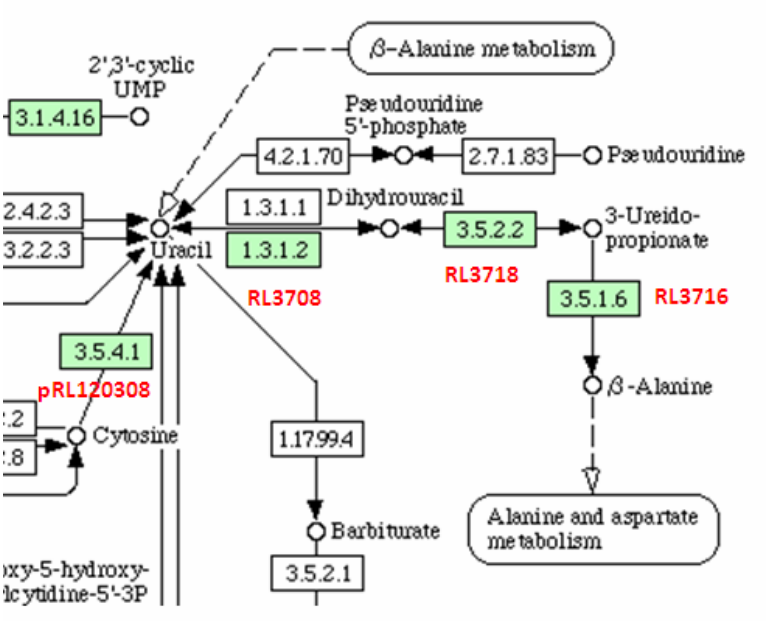



**Figure 3.2.14.** Conversion of uracil to β-alanine to in pyrimidine metabolism.

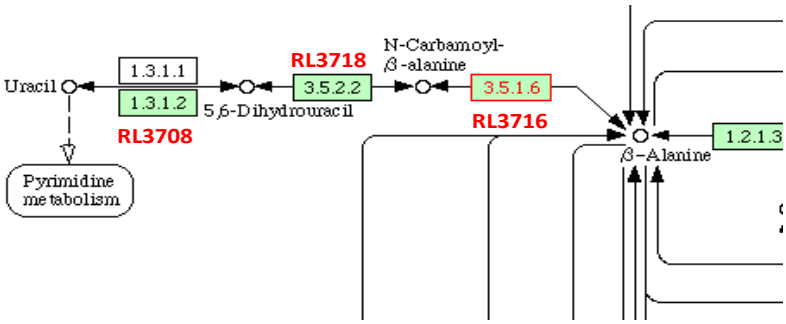



**Figure 3.2.15.** Conversion of β-alanine to uracil in β-alanine metabolism.

Additionally, *pRL120302* (*yagT*) [0.7,1.1U,0.9] encodes the xanthine dehydrogenase YagT iron-sulfur binding subunit and *RL0383* (*iunH*) [0.3,1.1U,0.9] encodes a putative inosine-uridine preferring nucleoside hydrolase involved in nucleotide metabolism.

*RL0638* (*pucH*) [1.5U,1.0U,0.9] encodes a putative allantoinase involved in purine metabolism and was induced. *pRL120027* [4.2U,5.2U,5.0U], annotated as a putative aldolase, was very highly induced. pRL120027 has high identity to deoxyribose-phosphate aldolase of *R. leguminosarum* bv. *trifolii* (99%) and *M. loti* (81%). It also has moderate identity (<40% Id) to fructose-bisphosphate aldolase in many members of the archaea. The enzyme deoxyribose-phosphate aldolase catalyzes the breakdown of deoxyribose-phosphate to glyceraldehyde-3-phosphate and acetaldehyde. It is important to note that another enzyme RL4267 (AcoD) [4.8U,4.4U,4.6U] acetaldehyde dehydrogenase, which catalyzes the conversion of acetaldehyde to acetic acid was also very highly induced. These observations together suggest that these enzymes may play a role in purine metabolism.

Two genes *pRL110557* (*glxB*) [1.2U,1.8U,1.8U], encoding glutamine amidotransferase, and *RL1456* (*purF*) [0.7,1.7U,2.4U], encoding an amidophosphoribosyltransferase, were induced. GlxB and PurF are involved in purine biosynthesis by catalyzing the addition of an amine group to

phosphoribosylpyrophosphate (PRPP) to from phosphoribosylpyrophosphate aminophosphate.

In addition, two genes *RL0395* (*miaB*) [1.2U,0.8,0.2], encoding a putative MiaB protein, and *RL4551* (*trmD*) [1.0U,1,5U,0.9], encoding a tRNA (guanine-N(1)-)-methyltransferase involved in RNA synthesis and modification, were induced. In addition, many genes involved in DNA repair were also induced.

Interestingly, sequence analysis of the gene *pRL110107*, annotated as conserved hypothetical protein, revealed the presence of Ku-like domain, which was reported to be involved in DNA repair (Aravind & Koonin, 2001). Generally, Ku proteins are clustered with ATP-dependent DNA ligase and eukaryotic type primases. Three types of arrangement of this cluster were reported in *M. loti* MAFF303099 (Aravind & Koonin, 2001). A database similarity search with the Ku protein cluster (mlr9623-25) of *M. loti* MAFF303099 against the Rlv3841 genome identified the presence two copies of Ku family proteins (p*RL110107-8* and *RL2141-2*). In contrast to the *M. loti* MAFF303099 arrangement, there is no significant similarity found for the third gene mlr9625 near the Ku family proteins (*pRL110107-8* and *RL2141-2*), but it showed high identity to pRL120229 (80% Id) encoding a DNA ligase family protein, which was also induced in the rhizosphere. The genes *pRL110107* [2.5U,1.3U,1.6U], *RL2141* [1.5U,1.2U,1.5U] and *pRL120229* [1.5U,0.7,1.2U] were induced and are members of the Ku protein cluster. *pRL110108* and *RL2142* were expressed but did not pass the criteria. This observation indicates the presence of a novel arrangement of Ku family proteins and that they are involved in DNA repair mechanisms in Rlv3841 in the pea rhizosphere.

**3.2.5.9 Genes involved in aminoacid metabolism**

Amino acid analysis of pea root mucilage showed the presence of high amounts of Ser, Hyp (hydroxyproline), Gly, Thr, Pro, Ala and Glx (Gln or Glu), moderate amounts of Tyr and Val, lower amounts of Ile, Leu, Phe, Lys and Arg, and the absence of His, Cys and Met (Knee *et al.,* 2001). In this transcriptomics study, genes involved in biosynthesis of the amino acids which are present in

higher amounts in pea root mucilage (i.e. Ser, Hyp Gly, Thr, Pro, Ala and Glx (Gln or Glu)) were not expressed. Only genes involved in the synthesis of amino acids which are present at low concentrations or completely absent from the pea root mucilage were induced.

Homoserine is abundantly found in pea root exudates (Knee *et al.,* 2001). It has also been reported that *R. leguminosarum* can utilize homoserine as a carbon source (Van Egeraat, 1975). Although homoserine is not incorporated into proteins, it serves as a precursor metabolite for lysine biosynthesis (KEGG PATH: rle000300), glycine, serine and threonine metabolism (KEGG PATH: rle000260) and methionine metabolism (KEGG PATH: rle000271). *pRL100137* (*metX*) [0.6,1.5U,0.6] encodes a homoserine-O-acetyltransferase, which catalyzes the conversion of homoserine to O-acetyl homoserine, a precursor metabolite for methionine biosynthesis. *pRL80071* (*hom*) [1.4U,2.3U,2.2U], encoding a homoserine dehydrogenase catalyzing the conversion of L-homoserine to L-aspartate-semialdehyde, an intermediate in lysine biosynthesis and glycine, serine and threonine metabolism. Homoserine dehydrogenase (Hom) may also involve in utilization of homoserine as carbon source by Rlv3841 in the pea rhizosphere.

*pRL120373* [1.1U-0.3-0.5] encodes a putative cystathionine gamma synthase/methionine gamma lyase which catalyzes the conversion of selenomethionine to methaneselenol in selenoamino acid metabolism (KEGG PATH: rle000450).

*RL2756* (usgA) [4.0U,2.8U,1.7U], encoding a phosphoribosyl-anthranilate isomerase subunit of the *usg* family, and *RL3716* (*atcC*) [2.5U,1.5U,1.2U], encoding an N-carbamyl-L-cysteine amidohydrolase involved in tryptophan, lysine, and cysteine biosynthetic pathways, were induced. Additionally, there was induction of *RL0436* (*dapE*) [1.6U,3.5U,3.4U], encoding a succinyl-diaminopimelate desuccinylase also involved in lysine biosynthetic pathway.

Genes coding for enzymes involved in the phenylalanine degradation pathway, *RL1860* (*phhA*) [6.3U,5.3U,4.6U] and *RL1863* (*hpd*) [3.1U,2.1U,1.2U] were induced at all time points. The homogentisate degradation gene cluster

including, *RL1864* (*hmgA*) [1.6U,0.7,0.6] homogentisate 1,2-dioxygenase, *RL1865* (*hmgB*) fumarylactetoacetate hydrolase and *RL1866* (*hmgC*) maleylacetoacteate isomerase and the genes coding for *phhA* and *hpd* are upstream of the *hmg* cluster in Rlv3841. The gene coding for *hmgA* was only induced at 7dPea-1dpi, and the genes coding for *hmgBC* did not pass the criteria. However, another gene *RL2736* (*maiA*) [0.5,1.1U,0.7] coding for putative maleylacetoacetae isomerase was expressed at 7dPea-3dpi. In *S. meliloti* 1021, *hmgA* was shown to be induced by carbon and nitrogen deprivation and also involved in tyrosine degradation (Milcamps & de Bruijn, 1999; Milcamps *et al.,* 2001). In *P. putida*, the phenylalanine and tyrosine degradation pathways utilize the homogentisate degradation pathway, to produce fumarate and acetoacetate (Arias-Barrau *et al.,* 2004). In addition, genes *pRL100303* [1.8U] and *pRL100304* [1.0U], coding for α and β subunit of 2-oxoisovalerate dehydrogenase, involved in valine, leucine and isoleucine degradation (KEGG PATH: rle00280), were expressed only at 7dPea-1dpi. These observations suggest that Rlv3841 may employ a homogentisate like pathway to degrade aromatic aminoacids to use as carbon and nitrogen sources.

#### 3.2.5.10 Genes involved in fatty acids metabolism

In *Rhodobacter sphaeroides*, betaine-lipid biosynthesis is mediated by two genes *btaA* and *btaB*, which are induced during growth in phosphate-limiting conditions. Betaine-lipid resembles phosphatidylcholine and may replace it during phosphate-limiting conditions (Klug & Benning, 2001). Another study, confirmed that these two genes are sufficient for betaine-lipid biosynthesis in bacteria (Riekhof *et al.,* 2005). *RL3327* (*btaB*) [1.0U], encoding putative S-adenosylmethionine:diacylgycerolhomoserine-N-methyltransferase with moderate amino acid identity (53%) with *btaB* of *R. sphaeroides*, was induced. The upstream gene, *RL3326* (*btaA*) encoding a S-adenosylmethionine:diacylglycerol-3-amino-3-carboxypropyl transferase shows moderate identity (48%) with *btaA* of *R. sphaeroides*, and was slightly induced (below 1 $\log_2$-fold) at 7dPea-3dpi. This indicates the expression of the two genes essential for betaine-lipid biosynthesis in

Rlv3841 in the pea rhizosphere. The induction of genes related to phosphate limitation taken together with the genes involved in betaine-lipid synthesis, suggests the adaptation of Rlv3841 to phosphate limitation in the pea rhizosphere.

RL0505 (FadD) [1.2U,1.9U.1.8U], acyl CoA synthetase is involved in fatty acid biosynthesis (KEGG PATH: rle00061). In addition a few genes involved in fatty acid degradation were induced; *pRL120739* [1.1U,1.1U.1.0U] encodes phospholipase, *pRL80050* [0.9,1.5U.1.3U] encodes an exported lipase and *RL1240* [2.2U,0.7.1.1U] encodes a lipase. The lipases are involved in degradation of a variety of fatty acids.

### 3.2.5.11 Genes involved in biosynthesis of vitamins and co-factors

Vitamins and co-factors are essential components for the survival of bacteria in the host environment. Vitamins are either acquired from the host or biosynthesized from precursors found in the host environment. Many genes involved in vitamin biosynthesis have been identified by IVET and mutants were affected in their ability to colonize the rhizosphere or survive (Karunakaran *et al.,* 2006; Rediers *et al.,* 2005). RL3040 [1.9U,0.9,1.0U], an SBP-ABC transporter, has high amino acid identity (40-80%) to a molybdate uptake protein in a wide range of bacteria. The induction of *RL2711* (*moaA*) [2.5U,2.5U,2.7U] and *RL1950* [1.2U,1.7U,1.4U], encoding molybdenum co-factor biosynthesis proteins, and *RL4688* (*mopB*) [0.8,1.0U,0.8], encoding a putative molybdenum-pterin binding protein, was reported to be induced in molybdate-limiting medium (Mauchline *et al.,* 2006). Many enzymes like nitrogenase and carbon monoxide dehydrogenase require molybdenum for their function. Also, several enzymes such as xanthine oxidase, DMSO reductase, sulfite oxidase, and nitrate reductase also require molybdopterin as cofactor for their function.

RL1651 (UbiG) [0.4,0.1,1.1U] and RL1838 (UbiE) [1.2U,0.7,0.4] are involved in ubiquinone biosynthesis (KEGG PATH: rle00130). RL3708 (PydA), RL3715, RL3716 (AtcC) and RL3718 (Dht) D-hydantoinase are involved in the conversion of uracil to β-alanine (discussed in section 3.2.5.8), which is a precursor to pantothenate and CoA biosynthesis. RL1959 (AcpD) [0.8,1.3U,3.0U]

is an acyl carrier protein phosphodiesterase involved in pantothenate and CoA biosynthesis (KEGG PATH: rle00770).

The expression of genes *thiMED* coding for thiamine biosynthesis have been shown to be up-regulated in the rhizosphere (Karunakaran *et al.,* 2006). However, in this study these genes failed to pass the $\log_2$ fold criteria (>1 $\log_2$ fold), although they passed the criteria in another set of experiments (data not shown).

*RL2288* (*cysG2*) [-0.7D,0.7,1.1U], encoding siroheme synthase involved in the synthesis of siroheme, a prosthetic group required for nitrite and sulphite reductases, was induced. CycG2 has high identity with CysG (91%) of *R. etli* and CysG (81%) of *S. meliloti*. A *cysG* mutant in *R. etli* was a cysteine auxotroph, failed to grow on nitrate as sole nitrogen source and showed reduced competitiveness in nodulation. RNA protection analysis showed that in *R. etli* the expression of, *cysG* mRNA was induced (either directly or indirectly) by methionine or sulphate in the medium (Tate *et al.,* 1997). In addition, a *cysG* mutant in *S. meliloti* (identified by STM studies) showed delayed nodulation (Pobigaylo *et al.,* 2008). An earlier report on pea root mucilage analysis showed the absence of the sulphur containing amino acids cysteine and methionine (Knee *et al.,* 2001). Moreover, the induction of genes involved in the biosynthesis of cysteine and methionine adds weight to the absence of cysteine and methionine from the pea rhizosphere. Based on the study of mutants in other species (Tate *et al.,* 1997) (Pobigaylo *et al.,* 2008) and the elevation of *cysG* expression in Rlv3841, it may be attractive to speculate that expression of *cysG* is essential for the survival of Rlv3841 and other rhizobia in the rhizosphere.

RL0973 (Dxs) [0.6,1.8U,1.5U] is a putative 1-deoxy-D-xylulose-5-phosphate synthase, catalyzing the formation of 1-deoxy-D-xylulose-5-phosphate (DXP) from pyruvate and glyceraldehyde-3-phosphate. The DXP serves as a precursor for biosynthesis of isoprenoid, thiamin and pyridoxol (vitamin B6) biosynthesis (Sprenger *et al.,* 1997). Additionally, *RL2979* [0.5,1.0U,0.8] encodes a protein involved in riboflavin biosynthesis, *pRL90162* [0.1,1.0U,1.3U] encodes a protein involved in cobalamin biosynthesis and *pRL120570* [0.9,1.0U,0.7]

encoding GTP cyclohydrolase I involved in folic acid biosynthesis. Expression of these genes indicates the biosynthesis of a wide range of vitamins and co-factors by Rlv3841 during survival in the pea rhizosphere.

### 3.2.4.12 Genes of unknown functions characterized by bioinformatics analysis

Apart from these previously characterized genes, many differentially expressed genes are of unknown function. Most of these up-regulated genes were in gene clusters suggesting that they are part of an operon. Based on sequence analysis it was attempted to predict the possible role of these genes in an environmental context.

RL0786 [1.7U,2.9U,3.6U] is highly similar to chalcone synthase protein of *R. leguminosarum* bv. *trifolii* (98% Id) and *R. etli* CFN42 (90% Id), indicating a possible role of RL0786 in flavonoid biosynthesis (Shaw *et al.,* 2006).

RL0913 [5.4U,5.5U,5.6U] a photosynthetic reaction centre (PRC) family protein has 54% identity to SMc00885 of *S. meliloti* 1021 and to ATU4165 of *A. tumefaciens* C58 and 36% identity to mll3685 of *M. loti* MAFF303099. These proteins were reported to contain a PRC barrel domain involved in a variety of biological functions (Anantharaman & Aravind, 2002).

pRL100445 [1.9U,1.0U,1.2U] a conserved hypothetical protein with high amino acid identity (66% Id) to polyphosphate kinase of *Granulibacter bethesdensis,* CGDNIH1, was induced at all the time points and may participate oxidative phosphorylation.

The development of computational tools to predict the possible function of hypothetical proteins and their interaction with the other proteins is an essential part of understanding bacterial systems in a global manner. STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) is a database of known and predicted protein-protein interactions from wide range of sources (von Mering *et al.,* 2007). In the rhizosphere datasets, a large proportion of genes of unknown function were induced, most of which were in clusters. The STRING database was used to find potential interactions between clusters of hypothetical genes,

which were expressed together. Unfortunately, during the preparation of this thesis the Rlv3841 genome was not integrated into the STRING database, so the genes which had high identity to *R. etli* were used instead. One such gene cluster was *RL3272* [3.7U,2.3U,2.1U], *RL3273* [3.8U,2.4U,1.8U] and *RL3274* [4.4U,2.3U,2.1U], which were transcribed in the same direction, and strongly induced at all time points. The genes are annotated as conserved hypothetical proteins (RL3272-73) and PrkA protein kinase (RL3274). Database similarity search of the amino acid sequences (*RL3272-4*) showed high identity to RHE_CH02815-17 (96%, 94% and 97%) of *R. etli* CFN42 respectively. A STRING database search with the RL3274 (RHE_CH02817) amino acid sequence revealed that this cluster is well conserved in many bacterial genomes. In *E. coli*, *yeaH* and *yeaG,* have high identity with RL3273-74 and are co-expressed. Interestingly, *prkA* (RL3274, 32% Id) and *yhbH* (*RL3273*, 24%Id) mutants in *B. subtilis* was defective in sporulation (Eichenberger *et al.,* 2003). Taken together these observations suggest that these genes may play a role in adaptation to the rhizosphere.

*pRL120730* [3.7U,3.6U,3.6U] codes for a conserved hypothetical protein, a similar gene in *R. leguminosarum* A34 was identified as induced in the pea rhizosphere by IVET studies (Barr *et al.,* 2008). The exact function of this gene is unknown, but it is expected to play a role in rhizosphere adaptation.

*pRL120351* [3.4U,2.4U,2,1U] coding for permease component of an ABC-T, classified as a sugar uptake system. Proteins with high identity to this ABC operon are found in *R. leguminosarum* bv. *trifolii* and *R. etli*, but not in other sequenced rhizobia.

Many genes coding for different classes of enzymes involved in metabolic pathways were induced. Although, specific functions of these genes are unknown, it clearly emphasizes the presence of an active and a complex metabolic life style in the rhizosphere.

### 3.2.6 Effect of pea seedling age on the gene expression of Rlv3841 in the rhizosphere

The main objective of this set of experiments was to study the effect of pea seedling age on the transcriptome of Rlv3841. Recent research on root exudates collected from seed, seedlings and roots of tomato, cucumber and sweet pepper, showed that the amounts of organic acids and sugars present in the root exudates increase with the growth of the plant (Kamilova *et al.*, 2006). To identify the genes which alter with respect to the age of pea seedling, transcriptomic studies were performed with cells recovered one day after inoculation from the rhizospheres of three (7, 14 and 21 days) differently-aged pea seedling compared to laboratory grown Rlv3841.

Pea seedlings were grown for 7, 14 or 21 days as described in section 2.3.3 and inoculated with $10^8$ Rlv3841 cells, grown, harvested (0.6 $OD_{600}$nm) and washed as described in section 2.3.3. After one day of inoculation, the cells from the rhizosphere were harvested and total RNA was extracted and quantified as described in sections 2.2.8 to 2.2.9. In parallel, Rlv3841 cells were grown in AMS media with 10 mM glucose and 10 mM ammonia as carbon and nitrogen source as described in section 2.1.1. Total RNA was then isolated from the cells and quantified as described in sections 2.2.8 to 2.2.9. Equal amounts of both the free-living and rhizosphere Rlv3841 total RNA was amplified, quantified and as described in sections 2.2.10 and 2.2.9. After amplification, both the RNA samples were then labelled with cydyes Cy3 (control - free-living) and Cy5 (experimental - rhizosphere (7dPea-1dpi, 14dPea-1dpi or 21dPea-1dpi)), depending upon the experiment as described in section 2.4.2. Microarray hybridization was performed with the equal amounts of both control and experimental labeled cDNA as described in sections 2.4.3-2.4.4 in five replicates. The obtained microarray data were then normalized (Figure 3.2.16) and subjected to rigorous statistical analysis in LIMMA as described earlier in section 2.4.5. The genes were considered significant if their adjusted (Benjamini and Hochberg method) P value is $\leq 0.01$ and the average $\log_2$ fold is $\geq 1$ of 5 biological replicates analyzed. As previously stated the expression level will be shown for 7d, 14d and 21d as *Gene Id* (*Gene*

*name*) ([7dPea-1dpi], [14dPea-1dpi], [21Pea-1dpi]), as values of $\log_2$. The combined expression dataset of 7d-14d-21dPea-1dpi was provided on the accompanying compact disc in a Microsoft excel format (version 2003).

For a more detailed picture of the transcriptome, the data from the three experiments were analyzed together as single time-course experiment. A gene was considered significant when it passed the above mentioned criteria in at least one of the time points, to include a gene whose expression change over time. Comparison of the three aged pea seedlings was performed by Venn mapping to identify time specific and commonly expressed genes (Figure 3.2.17). In total, 1319, 1175 and 1304 (ie, 18%, 16% and 18% of the genome) genes were differentially expressed in 7dPea-1dpi, 14dPea-1dpi and 21dPea-1dpi rhizosphere experiments. The genes induced on the differently aged rhizosphere were distributed significantly comprising all the functional classes, except cell division, ribosome constituents and fatty acid biosynthesis. Comparative analysis of all the three expression profiles revealed a core set of 285 genes expressed in all rhizospheres, which were also present in the 883 up-regulated gene set identified from the 7dPea-137dpi data. Comparison of both core set of genes, dpi [354] and seedling age [285] identified 254 common genes, which includes genes involved in metabolism, nutrient scavenging, transport systems, regulation and adaptation. They were induced >2 $\log_2$ fold, indicating their significance in rhizosphere adaptation and colonization. Most of the genes which were induced at specific time points were $\leq$ 2 $\log_2$ fold, indicating that these genes respond to subtle changes in the rhizosphereic environment compared with the core set of genes, which may be due to the time-dependent secretion of organic acids and/or sugars in the root exudates. To identify genes which respond specifically to 14d and 21d old pea rhizosphere, the up-regulated gene expression dataset of 14dPea-1dpi and 21dPea-1dpi experiments were compared to the up-regulated combined dataset of 7dPea-137dpi (combined dataset of three experiment; 7dPea-1dpi, 7dPea-3dpi and 7dPea-7dpi). Most of the genes induced at specific time points were $\leq 2$ $\log_2$fold, indicating that these genes respond to subtle changes specific at that time in the rhizosphere compared with the core set of genes. Here, we describe some of the

important and interesting genes which were altered in their expression in 14dPea-1dpi and/or 21dPea-1dpi experiments relative to 7d old pea rhizosphere.

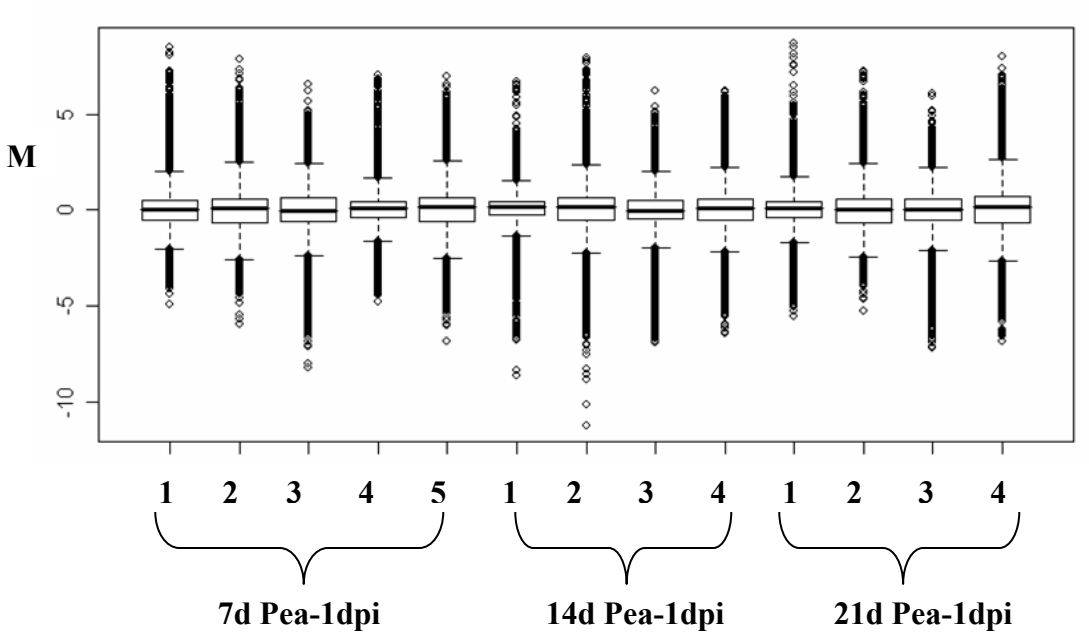



**Figure 3.2.16.** Boxplots showing the differential expression values of the five biological replicates of 7dPea-1dpi and 4 biological replicates of each 14dPea-1dpi and 21dPea-1dpi pea rhizosphere microarray experiments. M represents the differential expression ($log_2$ (experiment/control) after quantile normalization. In these experiments, the experimental RNA sample was labelled with Cy5 and control RNA sample was labelled with Cy3 except for the 3rd replicate of each condition, which is a dye swap experiment.

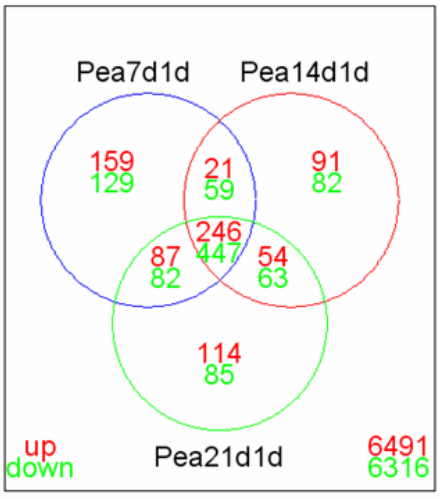



**Figure 3.2.17.** Venn diagram showing the comparison of up- and down-regulated genes of three differently aged pea rhizosphere experiments.

### 3.2.6.1 Genes specific for rhizosphere and symbiosis

Interestingly, the nod genes showed a very low level of induction, except *pRL100186* (*nodB*) [1.9U,0.9,1.2U] and *pRL100187* (*nodC*) [3.1U,0.9,1.2U] ($log_2$ fold ~1) in 14dPea-1dpi and 21dPea-1dpi compared to 7dPea-1dpi, suggesting that secretion of flavonoids essential for nod factor signalling were secreted early and their secretions reduces or stops in the absence of rhizobia during plant growth. In addition, the genes coding for Rhizobium adhering proteins (RapBC) which were reported to play an important role during early interactions showed a low level of expression compared to 7dPea-1dpi. In the past, many reports have emphasized the fact that the presence of rhizobia influences the secretion of flavonoids and other compounds in the rhizosphere (Cooper, 2004).

### 3.2.6.2 Genes differentially expressed in 14d and 21d-1dpi

Twenty five genes were identified to be specifically induced in both 14dPea-1dpi and 21dPea-1dpi experiment, including many genes of unknown function. Some of the genes with putative functions includes, RL0173 (PhnK) [-0.5D,1.5U,1.4U] a putative phosphonate uptake system, pRL120470 (ImpF) [0.4,1.5U,2.0U] a putative protein involved symbiosis, RL0101 (GabD2) [-0.3D,1.4U,1.5U] a succinate semialdehyde dehydrogenase, pRL80132 (TraM) [1.5U,1.3U,1.7U] a putative transcriptional regulator of conjugal transfer protein and pRL80133 (TrbI) [0.6,1.5U,1.6U] a putative conjugal transfer protein. In addition, many regulatory genes including two-component systems and genes related to transport processes were also induced including pRL120047 [3.9U,3.0U,3.0U] an ATP binding component of ABC-T having high amino acid identity to SMa20427 (90%) of *S. meliloti* 1021, which was induced by ectoine (Mauchline *et al.,* 2006).

Twelve genes were repressed in both 14dPea-1dpi and 21dPea-1dpi experiments including RL0542 [-0.5,-1.3D,-1.2D] an SBP of ABC-T having high identity to the phosphate ABC-T of *R. leguminosarum* bv. *trifolii* (95%) and *R. etli* CFN42 (93%), pRL100404 [-0.3,-1.8D,-1.7D] a monocarboxylic acid permease which transports alanine and monocarboxylates (Hosie *et al.,* 2002) and

RL4284 (Ask) [-0.8,-1D,-1.3D] a putative aspartokinase, and a few other genes of unknown function.

### 3.2.6.3 Genes differentially expressed in 14d-1dpi

A total of 128 genes were differentially expressed, including 95 up- and 33 down-regulated genes in the 14d old pea rhizosphere. The differentially expressed dataset contains many genes involved in adaptation, cell envelope, cell processes, metabolism, regulation, transport systems and genes of unknown functions. Interestingly, many transport-related genes were induced, including *RL0167 (phoE*) involved in phosphate uptake along with *RL0169* [0.2,1.3U,0.7] encoding a putative transposase insertion element family, which are part of the operon *RL0165-70*, coding for a phosphate uptake system and induced during phosphate starvation (Bardin *et al.,* 1996). Many genes coding for sugar uptake systems were induced including, *RL1999* [0.4,1.2U,0.7U] encoding a permease component of an ABC-T which has high amino acid identity to *rbsCch3* (91%) annotated as a ribose ABC transporter in *R. etli* CFN42. *RL1826* [-0.3D,1.2U,0.1], encodes a permease component of an ABC-T with high identity to SMc04394 (63%) of *S. meliloti* 1021 which was strongly induced by dextrin (Mauchline *et al.,* 2006). RL3111 [0.1,1.2U,0.4] an SBP ABC-T with high amino acid identity to SMa2305 (73%) of *S. meliloti* 1021 that was induced by tagatose (Mauchline *et al.,* 2006), RL0906 [0.2,1.0U,0.1] a permease component of an ABC transporter with high amino acid identity to Rleg2DRAFT_5795 (92%) a putative monosaccharide transporting ATPase of *R. leguminosarum* bv. *trifolii*, which has a conserved domain (AraH) for ribose/arabinose/galactose/xylose ABC-T systems. Moreover, RL0906 has moderate amino acid identity to SMb20854 (39%) of *S. meliloti* 1021, which was reported to be induced by deoxyribose, as well as to SMb20856 (RL0908 - 30% Id) (Mauchline *et al.,* 2006).

Two genes important for rhizosphere competition were induced, *pRL110412* [0.1,1.1U,0.8] (*rhaT*) coding for an ABC-T component with high homology to *rhaT* (93% Id) of *R. leguminosarum* bv. *trifolii*, which was shown to be involved in rhamnose catabolism and to play a role in nodule occupancy

(Richardson *et al.,* 2004). This gene also shares high homology (67% Id) to SMc02325 (*rhaT*) of *S. meliloti* 1021, which was shown to be induced by rhamnose (Mauchline *et al.,* 2006).

*RL1494* (*iolD*) [-0.2,1.3U,0.7], coding for malonate semialdehyde dehydrogenase, involved in myo-inositol catabolism, was induced. It has been shown previously that the ability to utilize myo-inositol in the rhizosphere confers a competitive advantage to Rlv3841 during rhizosphere colonization (Fry *et al.,* 2001).

*RL2737* (*ohr*) [0.6,1.6U,0.9], encoding organic hydroperoxide resistance protein, was induced. Organic hydroperoxide are toxic molecules secreted by plants in response to microbial infections (Mongkolsuk *et al.,* 1998). There are four organic hydroperoxide resistance proteins in Rlv3841 and this gene was highly induced in the 14d old pea rhizosphere, compared to other pea rhizospheres tested.

*RL1811* (*dapA*) [0.8,1.0U,0.8] codes for dihydrodipicolinate synthase involved in the diaminopimelic acid/lysine biosynthetic pathway (Born & Blanchard, 1999), *RL2231* [0.5,2.3U,0.3] (*lpxA*), involved in lipid A biosynthesis, *pRL110003* (*repC*) [-0.1D,1.4U,0.8], coding for replication protein C, *pRL110626* (*cobL*) [0.1,1.6U,0.9], involved in cobalamin biosynthesis, *RL1257* (*qor*) [0.2,1.0U,0.7] coding for quinone reductase which may involved in chlorocatechol degradation, and *RL0078* (*exoB*) [-0.2D,1.6U,0.4], coding for UDP-glucose-4-epimerase, were induced. *RL0427* (*rbsK*) [0.2,1.8U,0.8], coding for ribokinase, has high homology with *rbsK* (93% Id) of *R.etli* CFN42 and homologous with the PfkB subfamily of carbohydrate and carbohydrate phosphate kinases (Michiels *et al.,* 1998), shows induction.

Thirty three genes were repressed in only the 14dPea-1dpi including many genes coding for putative transmembrane proteins, putative two-component regulatory systems and a few transport-related genes. Interestingly, *RL3557* (*bacA*) coding for a transmembrane transporter required for bacterial development was repressed. Two other genes, *RL2371* coding for putative quinone oxidoreductase

(*qor*) and *RL3841* coding for putative UDP-glucose-4-epimerase were also repressed.

### 3.2.6.4 Genes differentially expressed at 21d-1dpi

A total of 144 genes were differentially expressed, including 98 up- and 46 down-regulated genes in the 21d-old pea rhizosphere experiment. Surprisingly, the two versions of the *fix* operon, encoded on pRL10 and pRL9, were induced, indicating oxygen limitation: *pRL100206A (fixQ)* [0.8,0.2,1.4U], *pRL100207 (fixP)* [0.3,0.1,1.5U], *pRL100209 (fixH)* [0.1,0.5,1.0U] and *pRL90014* (*fixH2)* [0.4,0.8,1.4U], *pRL90015 (fixG2)* [0.8,0.8,1.2U], *pRL90016 (fixP2)* [0.6,0.7,1.8U], *pRL90016A (fixQ2)* [0.6,0.4,1.4U], *pRL90017 (fixO2)* [0.4,0.2,1.5U], *pRL90018 (fixN2*) [0.2,0.2,1.5U]. These genes were reported to play an important role in symbiosis (Schluter *et al.,* 1997). Additionally, *pRL90021* (*azuP*) [0.3,0.9,2.0U] encoding a pseudoazurin protein involved in respiration (Patschkowski *et al.,* 1996) and *pRL90023* (*hemN*) [0.1,0.5,1.2U] encoding a putative oxygen-independent coproporphyrinogen III oxidase involved in heme biosynthesis (Keithly & Nadler, 1983) and thought to be co-expressed with the *fix* cluster of genes, were also induced.

*pRL110010* [0.4,0.8,1.6U], coding for a copper transporting P-type ATPase similar to SMa1012 (*actP*) (88% Id) of *S. meliloti* and involved in copper homeostasis, was induced. *pRL110330* [0.8,0.6,1.3U] (*hmrR*) similar to SMa1015 (*hmrR*) (73% Id) of *S. meliloti* 1021 and coding for a MerR family heavy metal-dependent transcriptional regulator reported to be involved in the regulation of *actP* (Reeve *et al.,* 2002), was also induced.

Two genes, *pRL100300* [0.6,0.4,1.0U], coding for a putative short chain dehydrogenase/reductase family protein, and *pRL100301* [1.0U,0.6,1.2U], encoding a putative acyl-coA thiolase, were induced. In addition to *pRL100302* [1.0U,0.6,1.0U], encoding a putative acyl-CoA dehydrogenase, and *pRL100303* [1.7U,0.3,1.3U], encoding a 2-oxoisovalerate dehydrogenase, and were induced in 7dPea-1dpi. These genes probably form an operon, which may be involved in valine, leucine and isoleucine degradation as described in the KEGG database

(KEGG PATH: rle00280), suggesting the presence of these amino acids in pea root exudates, results in keeping with those reported earlier (Knee *et al.,* 2001).

*pRL120140* [0.8,0.2,1.4U], encoding a putative imidase having high identify with cyclic imide hydrolase (96% Id) of *P. putida* YZ-26 which shows high enzyme activity with succinimide, glutarimide and maleimide (Shi *et al.,* 2007), was induced along with two ABC–T components *pRL120141* and *pRL120142*. These genes probably form an operon. A database similarity search of the later genes showed homology (~60% Id) with putrescine/spermidine transport systems in a wide variety of bacteria. Earlier research in *Blastobacter* sp, showed the presence of the cyclic imide utilization pathway, wherein the cyclic imide compounds are hydrolyzed by imidase to monoamidated dicarboxylates, and subsequently to dicarboxylates by hydrolytic deamidation (Ogawa *et al.,* 1996; Soong *et al.,* 1998) to feed into the TCA cycle. Taken together, these observations indicate the possible presence of a cyclic imide utilization pathway in Rlv3841.

pRL110025 (CaiD) [0.5,0.2,1.3U], a carnitinyl-CoA dehydratase, suggests the presence of carnitine metabolism. From earlier reports, it is known that bacteria can use L-carnitine as an osmoprotectant (Goldmann *et al.,* 1991), as a carbon and nitrogen source under aerobic conditions and as electron acceptor under anaerobic conditions (Rebouche & Seim, 1998). The induction of *fix* genes and *caiD* may be due to an oxygen limitation in the rhizosphere. Moreover, the *fixABCX* genes have been shown to be induced by carnitine in *E. coli*, suggesting a possible role of *fix* operons in carnitine metabolism (Eichler *et al.,* 1995).

*pRL110022* coding for a permease component of ABC-T, located upstream of *pRL110025* was induced in all 7dPea-137dpi experiments and although the genes between them were expressed, they did not pass the criteria. This may be due to a lower concentration of carnitine or the availability of other preferred nutrients in the rhizosphere during early growth of the plant.

pRL110413 (RhaS) [0.5,0.3,1.2U] is an SBP ABC-T showing high homology with RhaS (96% Id) of *R. leguminosarum* bv. *trifolii* shown to be essential for rhamnose uptake and necessary for nodule occupancy (Richardson *et al.,* 2004). It also shares high identity (65% Id) with SMc02324 (RhaS) of *S.*

*meliloti* 1021, which was induced by rhamnose (Mauchline *et al.,* 2006). Earlier research, shows the presence of rhamnose (1%) in the root mucilage of pea (Knee *et al.,* 2001). The genes *rhaTS*, were also induced in 7dPea-137dpi experiments but did not pass the criteria, possibly due the presence of low concentration of rhamnose in the root exudates. Another transport system of the TRAP-T family was expressed; *pRL120400* [0.3,0.7,1.0U] shares high identity with SMb20442 (79% Id) coding for an SBP of *S. meliloti* 1021 and induced by D-(+)-fucose, mannose, L-(-)-fucose (Mauchline *et al.,* 2006). This suggests the presence of pentose monosaccharides in the rhizosphere.

The induction of *pRL90242* [-0.5,-0.1,1.3U] coding for a putative nitrilotriacetate monoxygenase, 49% identity to NmoA, a nitriloacetate monoxygenase of *Chelatobacter heintzii* ATCC 29600, which was reported to catalyze the conversion of nitriloacetate to iminodiacetate, and then to glyoxylate by iminodiacetate dehydrogenase (Xu *et al.,* 1997). This observation along with the up-regulation of glyoxylate enzymes indicates an active nitriloacetate degradation pathway in the rhizosphere.

*RL3095* (*arcB*) [0.2,0.1,1.1U] encodes a putative ornithine cyclodeaminase which catalyzes the conversion of ornithine to L-proline and *RL4385* (*rocD*) [0.9,0.4,1.3U] encodes a putative ornithine aminotransferase which catalyzes the conversion of ornithine to L-glutamate semialdehyde. Both these are involved in arginine and proline metabolism, as described in KEGG database (KEGG PATH: rle00330).

In addition, genes coding for a HlyD efflux pump (RL1329, RL1454 and RL4180), stress related proteins (pRL90047), regulators, transmembrane proteins and conserved hypothetical proteins of unknown functions, were also induced.

### 3.2.7 Genes down-regulated in the pea rhizosphere

Many genes involved, in cell division, including *minCD* and *ftsAZ*, were repressed. Interestingly, a large proportion of genes involved in chemotaxis and motility were also repressed, indicating that Rlv3841 switches off its chemotaxis and motility genes in the rhizosphere. Surprisingly, in free-living cells grown with

root exudates, the motility genes were not down-regulated, whereas they are repressed in free-living cells grown with hesperetin, implying that Rlv3841 needs to sense a rhizobial-induced signal from root exudates, before it represses chemotaxis-related genes. Genes involved in glycolysis, tri-carboxylic acid cycle, nitrogen metabolism, vitamin and cofactor biosynthesis were repressed, but this may be because of the comparison against exponentially growing free-living cells supplemented with glucose and ammonia. A large proportion of genes involved in ribosomal protein synthesis and many regulatory genes were also down-regulated. Interestingly, RL3624 coding for a SBP of an ABC-T was repressed. It showed high amino acid identity with SMb20902 (86% Id) of *S. meliloti* 1021, which was induced by mannose, lyxose, sorbose and glucose (Mauchline *et al.,* 2006). This suggests the absence or presence at only very low concentrations of free sugars in the rhizosphere, which is in contrast to the observation of presence of mannose in the root mucilage (Knee *et al.,* 2001).

### 3.2.8 Real-time quantitative confirmation of microarray data

In order to validate the microarray data, 7d-7dpi RNA samples were used as representative of all rhizosphere samples. Quantitative RT-PCR (qRT-PCR) was performed with amplified total RNA obtained after seven days post-inoculation from a 7d-old pea rhizosphere and from cells grown in AMS (10 mM glucose and 10 mM ammonia) overnight in laboratory conditions. The growth and harvesting conditions were the same as for samples harvested for microarrays. Amplified total RNA was treated with Ambion's Turbo DNase to eliminate genomic DNA contamination. The fidelity of the PCR reaction was confirmed by lack of signals in controls lacking either reverse transcriptase or template.

Fifteen differentially expressed (12 up- and 3 down-regulated) genes, spanning different functional categories and exhibiting various expression levels, including a gene coding for an unknown function, were selected. The gene RL0644, coding for ribitol-2-dehydrogenase was selected as calibrator gene, as it was found to be unaltered in the 7d Pea-7dpi microarray dataset. The qRT-PCR was performed as described earlier and the data was analyzed by comparative $C_t$ method, which is also known as the $2^{-\Delta\Delta Ct}$ method, where $\Delta\Delta Ct = \Delta C_{t,\ sample} - \Delta C_{t,\ reference}$. The $\Delta C_{T,\ sample}$ is the Ct value for any sample normalized to the endogenous housekeeping gene and $\Delta C_{t,reference}$ is the Ct value for the calibrator also normalized to the endogenous housekeeping gene.

The results of the qRT-PCR data analysis (Table 3.2.6. and Figure 3.2.18.) were compared with the corresponding microarray data from the 7d-7dpi expression dataset. Overall, the qRT-PCR data were in good agreement with the microarray data, thus validating the microarray technique used.

| ID | Function | qRT-PCR [a] $log_2$-fold | Microarray [b] $log_2$-fold |
|---|---|---|---|
| **Up-regulated genes** | | | |
| RL4267 | putative acetaldehyde dehydrogenase (*acoD*) | 6.34 +/- 1.49 | 4.55 |
| pRL100169 | rhizosphere expressed protein (*rhiA*) | 4.28 +/- 0.72 | 3.80 |
| RL3424 | putative C4-dicarboxylate transport protein (*dctA*) | 6.47 +/- 0.36 | 4.79 |
| RL1251 | putative serine protease | 4.49 +/- 0.49 | 4.63 |
| RL1860 | putative phenylalanine-4-hydroxylase (*phhA*) | 6.80 +/- 0.75 | 4.58 |
| RL0996 | putative transmembrane transporter | 3.84 +/- 0.71 | 2.17 |
| RL0037 | putative phosphoenolpyruvate carboxykinase (*pckA*) | 4.25 +/- 0.48 | 2.75 |
| pRL80026 | putative branched-chain amino acid ABC transporter binding component (*livH*) | 5.06 +/- 0.68 | 3.96 |
| RL4274 | putative HlyD family transmembrane efflux protein | 5.64 +/- 1.29 | 7.01 |
| pRL100187 | N-acetylglucosaminyltransferase (*nodC*) | 4.57 +/- 0.36 | 3.51 |
| RL1925 | conserved hypothetical protein | 5.26 +/- 0.18 | 5.28 |
| pRL110443 | putative hydroxyethylthiazole kinase (*thiM*) | 1.55 +/- 1.37 | 0.6 |
| **Down-regulated genes** | | | |
| RL2164 | hypothetical protein | -2.16 +/- 0.17 | -2.44 |
| pRL100451 | putative auto aggregation protein | -3.66 +/- 0.58 | -3.23 |
| RL3624 | putative substrate-binding component of ABC transporter | -4.63 +/- 0.9 | -4.47 |

**Table 3.2.6.** Quantitative real time-reverse transcription PCR of genes selected as differentially expressed on the 7d Pea-7dpi rhizosphere, compared to laboratory grown samples. [a] = Number of replicates is n= 3 with standard mean error, [b] = normalized average of 5 replicates.

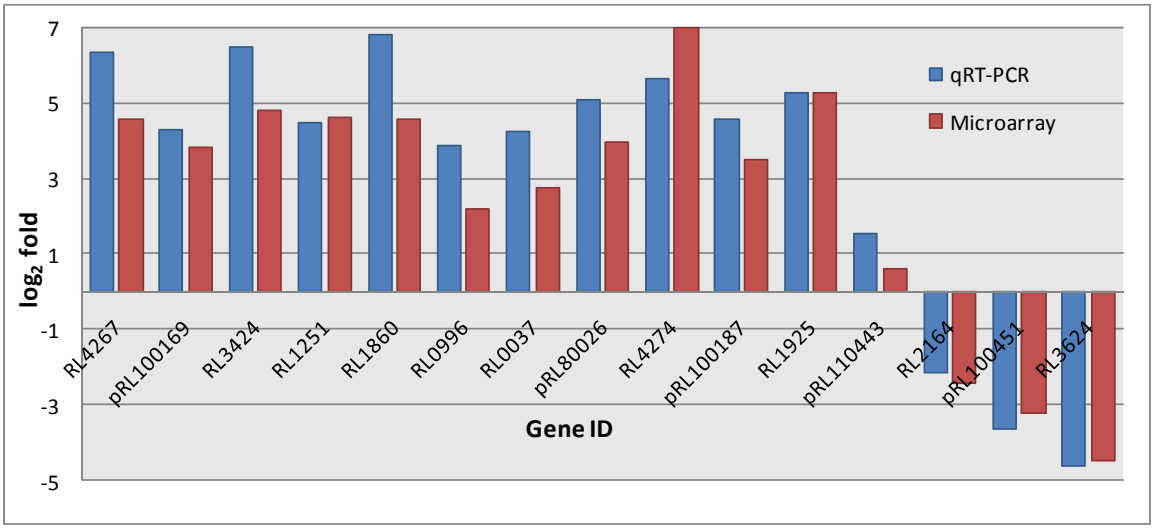



**Figure 3.2.18:** Bar graph showing the comparison of expression values of differentially expressed genes by quantitative real time-reverse transcription PCR and microarray, as described in Table 3.2.6.

### 3.3 Conclusion

This chapter reports on the transcriptome of *R. leguminosarum* bv. *viciae* 3841 in the pea rhizosphere from the microarray dataset obtained from: (1) three different post inoculation time points and (2) three different aged pea (seedlings) rhizospheres. Initial control experiments with pea root exudate and hesperetin individually showed the induction of *nod* and *rhi* genes in laboratory grown cultures, which complemented the rhizosphere datasets. The 7dPea-1dpi experiment showed a considerable amount of genes expressed with just one day incubation, and that there was not a great change in the number of differentially expressed genes in 7dPea-3dpi and 7dPea-7dpi compared to 7dPea-1dpi. The combined analysis of the experiments identified many genes, which would have been missed if analyzed individually. A large percentage of the genes had expression levels < 2 $\log_2$-fold, which may reflect the extremely low concentration of available nutrients found in pea root exudates (per pea seedling).

This work identified many previously characterized rhizosphere and symbiotic specific genes reported for Rlv3841 or for other related rhizobia. Additionally, many unknown and putative transporters were induced, ranging from high induction to those due to subtle changes in the pea rhizosphere. Rlv3841 contains a large proportion of genes of unknown functions, which was reflected in the rhizosphere microarray dataset, confirming that the rhizosphere is an unexplored area. The induction of metabolic genes shows the operation of gluconeogenesis and glyoxylate pathway in the rhizosphere. Genes involved in other metabolic pathways were induced, but no complete pathway was observed.

Though many sugar ABC transporters were induced, the corresponding catabolic genes involved in the utilization of the sugar were not induced. This may be like a catabolite repression where Rlv3841 senses sugar compounds by inducing transporter genes. However, the catabolic genes may be repressed by the unknown catabolite repressors, like dicarboxylates or organic acids. The uptake of dicarboxylates is evident from high induction of *dctA* and the utilization by the genes involved in glyoxlyate and gluconeogenesis pathway. Catabolite repression has not been reported in Rlv3841, but a catabolite repression was reported in free-

living *S. melitoti* (Bringhurst and Gage, 2002). Free-living *S. meliloti* preferentially utilizes succinate and other similar dicarboxylates and exerts a catabolite repression on the other genes required for utilizing secondary carbon sources (Bringhurst and Gage, 2002). Rlv3841 may also preferentially utilize dicarboxylates and other organic acids, which repressed the induction of other catabolic genes, eventhough the genes involved in their uptake are induced. Further, growth experiments with the free-living Rlv3841 in the combination of dicarboxylates and secondary carbon compounds may provide more insights on the catabolite repression in Rlv3841.

Another interesting observation is the apparent utilization of CO and formate by Rlv3841. The genes involved in the utilization of both CO and formate were strongly expressed in the pea rhizosphere. Further, experimental analysis such as growth studies with CO and formate as the sole carbon and energy sources will provide more insight on genes involved in their transport and utilization. Many genes involved in the stress response were expressed, which is to be expected since the rhizosphere is a stressful environment. Genes involved in purine and pyrimidine metabolisms were up-regulated. Another interesting observation is the biosynthesis of betaine-lipid, which is synthesized during growth in phosphate-limiting conditions.

The qRT-PCR analysis validated the quality and integrity of the microarray data. The validated microarray dataset substantiates the integrity of the developed strategy to perform large scale gene expression studies from rhizosphere recovered bacterial samples, which can be adapted to any rhizosphere bacteria. Now, with the transcriptome of Rlv3841-pea rhizosphere established, the next objective is to identify the host-specific genes of Rlv3841. The following chapter reports on experiments performed to identify genes induced specifically in response to peas, as opposed to other plants, in *R. leguminosarum* 3841.

# Chapter 4

## Comparative transcriptomics of *Rhizobium leguminosarum* 3841 in three different plant rhizospheres.

## 4.1 Introduction

From the pea rhizosphere experiments, a number of genes were identified which might play an important role in rhizosphere colonization. The validation by qRT-PCR of genes from the expression datasets substantiates the quality and integrity of the results obtained. With this success, the study was extended to compare the transcriptome of Rlv3841 grown in two other rhizospheres (alfalfa and sugar-beet), in order to identify genes which are specifically expressed in the pea rhizosphere. The dataset from the pea rhizosphere enabled identification of many genes which have already been reported to play an essential role in rhizosphere colonization and/or nitrogen fixation in Rlv3841 and other evolutionarily related rhizobia. In addition, many novel genes which might also play a significant role in symbiosis were identified. Though the general events in rhizobia-legume symbioses are similar, host specificity plays a significant role in establishing a symbiotic interaction with a cognate host. The ability to interact specifically with the cognate host is determined by many host-specific molecular determinants. This is why in some cases, genes responsible for a negative phenotype (competitively defective, $Nod^-$ or $Fix^-$) in a rhizobia-legume interaction, do not always confer a similar phenotype in another *Rhizobium*-legume interaction. A number of host specific genes, which were reported to provide competitive advantage to rhizobia during host rhizosphere colonization were identified.

A host-specific advantage found in some rhizobia is the ability to detoxify, degrade or utilize plant derived toxic compounds such as phytoalexins. The detoxification of bean pterocarpen phytoalexins by *R. etli* (Gonzalez-Pasayo & Martinez-Romero, 2000), mimosine degradation by *Rhizobium* sp. strain TAL1145 (Soedarjo & Borthakur, 1996) and utilization of compounds present in tropici root exudates by *R. tropici* (Rosenblueth *et al.*, 1998) confers a competitive advantage to rhizobia in colonizing the host rhizosphere. In addition, the ability to utilize sugars, aminoacids and organic acids which are abundant in root exudates provide a competitive advantage in rhizosphere colonization and thus nitrogen

fixation. Utilization of *myo*-inositol by *R. leguminosarum* 3841 (Fry *et al.*, 2001), rhamnose by *R. leguminosarum* bv. *trifolii* (Oresnik *et al.*, 1998), erythritol by *R. leguminosarum* VF39 (Yost *et al.*, 2006) and proline by *S. meliloti* (Jiménez-Zurdo *et al.*, 1995) offers a competitive advantage during root colonization, nodule occupancy and sometimes in nitrogen fixation.

Apart from the core set of genes involved in basic cell function, survival and adaptation to a host rhizosphere, a specific set of genes are essential to overcome the selective pressure imposed by the plants through secretion of toxins and specific nutritional compounds through root exudates. Identifying and understanding the functions of these host-specific genes are very important in addition to the basic genes involved in rhizosphere colonization. This will enable us to get a wider picture of the molecular events underpinning the symbiotic interaction in the rhizosphere.

In order to identify the host-specific genes expressed in Rlv3841 during pea rhizosphere colonization, we designed a comparative rhizosphere transcriptomics approach, wherein the transcriptome of Rlv3841 grown in three rhizospheres, pea (cognate), alfalfa (other-legume host) and sugar-beet (non-legume) were compared. Alfalfa and sugar-beet were selected for comparative transcriptomics, as alfalfa is a legume host of *S. meliloti* which is a well characterized and evolutionarily related to Rlv3841 and sugar-beet, a non-legume host for many *Pseudomonas* sp. (PGPR) and their interactions are also well studied.

In this approach, the experiments were designed in two ways (1) indirect (against common reference) and (2) direct (against itself) comparison. In the indirect design, RNA samples extracted from Rlv3841 cells recovered at 7d post-inoculation on 7d old pea (7dPea-7dpi), 7d old alfalfa (7dAlf-7dpi) and 7d old sugar-beet (7dSB-7dpi) rhizospheres were compared against a common reference (3841 grown on AMS - 10 mM glucose and 10 mM ammonia), while in the direct design, they were compared against themselves as described in Figure 4.1.1.A. The 7d post-inoculation time point was chosen to provide sufficient time for Rlv3841 cells to adapt to the non-host rhizospheres. The two designs have their

own advantages, an indirect design provides the expression level of a particular gene in three rhizospheres, whereas the direct design provides expression values with respect to two rhizospheres and moreover it eliminates most genes which have equal or nearly equal expression levels.

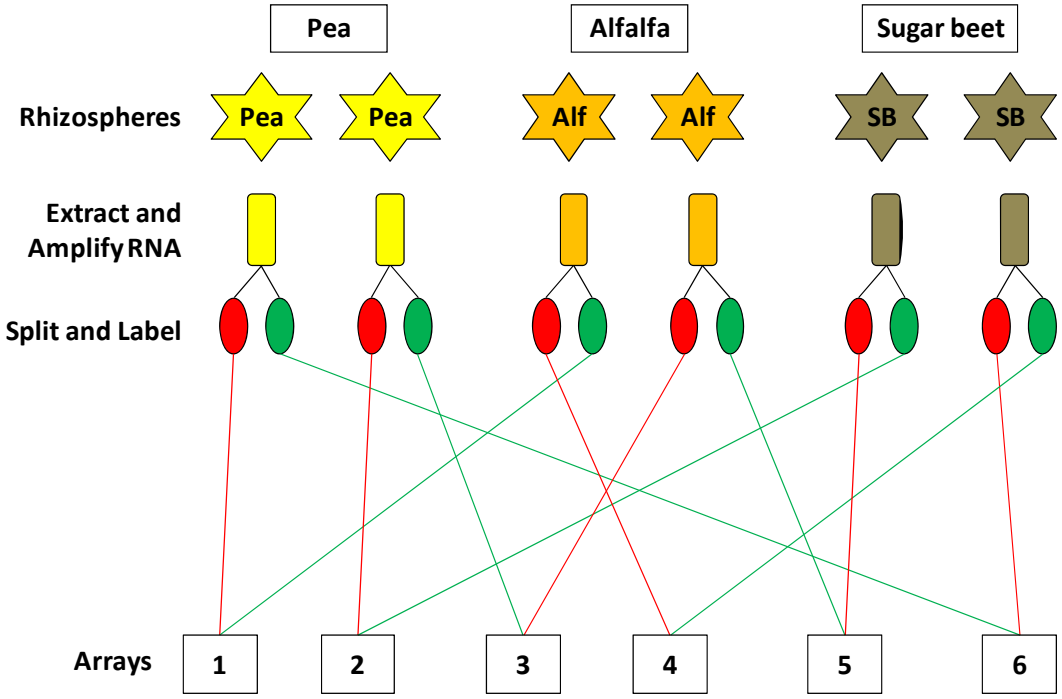
Pea, alfalfa and sugar-beet seeds were grown for 7 days as described in section 2.3.3 and inoculated with $10^8$ Rlv3841 cells, grown, harvested (0.6 $OD_{600}$nm) and washed as described in section 2.3.3. After 7days, the cells from the rhizosphere were harvested and total RNA was extracted and quantified as described in sections 2.2.8 to 2.2.9. In parallel, Rlv3841 cells were grown in AMS media with 10 mM glucose and 10 mM ammonia as sole carbon and nitrogen sources respectively as described in section 2.1.1. Total RNA was then isolated from the cells and quantified as described in sections 2.2.8 to 2.2.9. Equal amounts of both the free-living and rhizosphere Rlv3841 total RNA were amplified, quantified and as described in sections 2.2.10 and 2.2.9.

After amplification, both RNA samples were then labelled with CyDyes, depending upon the experiment as described in section 2.4.2. For the indirect comparison, CyDyes Cy3 (control - free-living) and Cy5 (experimental - rhizosphere (7dPea-7dpi, 7dAlf-7dpi or 7dSB-7dpi)) and for direct comparison, the three rhizosphere samples were labeled as described in Figure 4.1.1.B for each experiment.

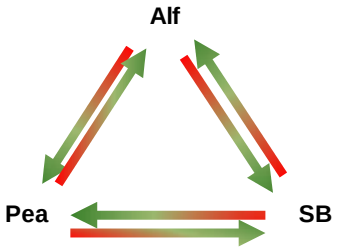
Microarray hybridization was performed with equal amounts of both control and experimental labeled cDNA as described in sections 2.4.3-2.4.4 for both designs. The indirect design was performed with three biological replicates, while the direct design was performed with two sets of two biological replicates. Thus each set of the two-colour direct design was performed with two biological replicate of each rhizosphere sample, so a total of four biological replicates per rhizosphere sample were used. The microarray data obtained was then normalized and subjected to rigorous statistical analysis in LIMMA as described earlier in section 2.4.5. The genes were considered significant if their adjusted (Benjamini and Hochberg method) P value is $\leq 0.01$ and the average $\log_2$ fold is $\geq 1$ of 3-5 biological replicates analyzed. In this chapter, as previously stated the expression

level will be shown for 7dPea-7dpi, 7dAlf-7dpi and 7dSB-7dpi as *Gene Id* (*Gene name*) ([7dPea-7dpi], [7dAlf-7dpi], [7dSB-7dpi]), as values of $\log_2$. The combined expression dataset of 7dPAS-7dpi (indirect comparison), both and up- and down-regulated genes at least in one time point were provided on the accompanying compact disc in the Microsoft Excel 2003 format.

**A.**




**B.**




| Arrays | Cy3 | Cy5 |
|---|---|---|
| 1 | 7d Alfalfa 7dpi | 7d Pea 7dpi |
| 2 | 7d Pea 7dpi | 7d Alfalfa 7dpi |
| 3 | 7d Alfalfa 7dpi | 7d Sugar-beet 7dpi |
| 4 | 7d Sugar-beet 7dpi | 7d Alfalfa 7dpi |
| 5 | 7d Pea 7dpi | 7d Sugar-beet 7dpi |
| 6 | 7d Sugar-beet 7dpi | 7d Pea 7dpi |

**Figure 4.1.1** Experimental design for direct comparison of Rlv3841 grown in three different rhizospheres. (A) A single set for the direct comparison experiment, two biological replicates from each rhizosphere sample were extracted and amplified. From each amplified RNA sample, an equal amount (15µg) of amplified RNA was taken and labelled with Cy3 and Cy5 separately. Equal amounts of labelled cDNA's were used for each microarray experiment as described in the design. A second set was performed before analysis, yielding 4 biological replicates per rhizosphere. (B) The design for direct two-colour experiments and the table showing the combination of rhizosphere microarray experiments corresponding to the numbers shown in the Figure 4.3.1.A.

### 4.2.1 Comparative rhizosphere transcriptomics (Indirect comparison)

The normalized data of five biological replicates of 7dPea-7dpi and three biological replicates of each 7dAlf-7dpi and 7dSB-7dpi rhizosphere experiments are shown as box-plots in Figure 4.2.1. Data analysis of the indirect comparison of the three rhizospheres to a common reference (3841 grown in AMS - 10 mM glucose and 10 mM ammonia) showed a total of 1365, 1408 and 1445 genes differentially expressed in pea, alfalfa and sugar-beet rhizospheres respectively. Comparison of the three rhizospheres datasets using Venn diagrams revealed sets of genes induced and repressed in specific rhizospheres, including 306 genes induced in all the three rhizospheres, 215 specifically induced in the pea rhizosphere, 148 induced in the alfalfa rhizosphere and 169 induced in the sugar-beet rhizosphere (Figure 4.2.2.A). A large number of genes had expression values less than two $\log_2$ fold (Figure 4.2.2.B). The differentially expressed genes in all the three rhizospheres did not show any discrete difference in the distribution of biological functional classification of genes (Figure 4.2.3.A and B).

A global overview of pea, alfalfa and sugar-beet rhizosphere expression datasets show some of the striking differences. Interestingly, the *nod* operon (*nodOMFABCIJ*) coding for nod factor signalling, the characteristic feature of the *Rhizobium*-legume symbiosis, was induced only in pea and alfalfa, but not in the sugar-beet rhizosphere. Surprisingly, the genes in the *rhi* operon; *pRL100169* (*rhiA*) [4U,4.4U,5.5U], *pRL100170* (*rhiB*) [3.2U,4.3U,5.4U], *pRL100171* (*rhiC*) [1.6U,2.9U,3.9U], encoding rhizosphere induced proteins and *RL3378* (*cinI*) [1.5U,2.4U,2.4U] encoding autoinducer synthesis protein, involved in the production of N-acyl homoserine lactone (AHL) were induced in all three rhizospheres. A set of 18 genes from pRL8 plasmid were induced only in the pea rhizosphere but repressed in the alfalfa and sugar-beet rhizospheres. In addition, 42 genes including *nod* genes were found to be induced only in pea and alfalfa, and can be categorized as legume specific genes. Although the exact functions of most of these are unknown, they may play an important role in determining host-

specificity. The Rlv3841 genome contains a large number of transport related genes coding for many uptake and efflux systems (Young *et al*., 2006). Surprisingly, only a portion these genes were found to be induced in the pea rhizosphere experiments, indicating the preference and availability of the nutrients in the rhizosphere. The comparative analysis of all these three rhizospheres, show a specific set of transport related genes induced in each of pea, alfalfa and sugar-beet rhizospheres respectively, apart from the common set of transport related genes induced in all the three rhizospheres. This observation suggests that they may play a significant role in adapting to the host rhizosphere. This may be by gaining a competitive advantage by detoxifying or metabolizing host-specific compounds during rhizosphere colonization or by sustaining a metabolically flexible life-style in the rhizosphere. A total of 788 genes were differentially expressed in all the three rhizospheres, including 306 up- and 482 down-regulated genes. The up-regulated genesets comprised genes involved in cell envelope (13%), transport systems (10%), metabolism (11%), regulation (6%) and a large proportion (47%) of unknown functions. The induction of a large number of transmembrane proteins and ABC-T components irrespective of the specific rhizosphere indicates a major reprogramming of the genetic circuit to adapt to the rhizosphere.

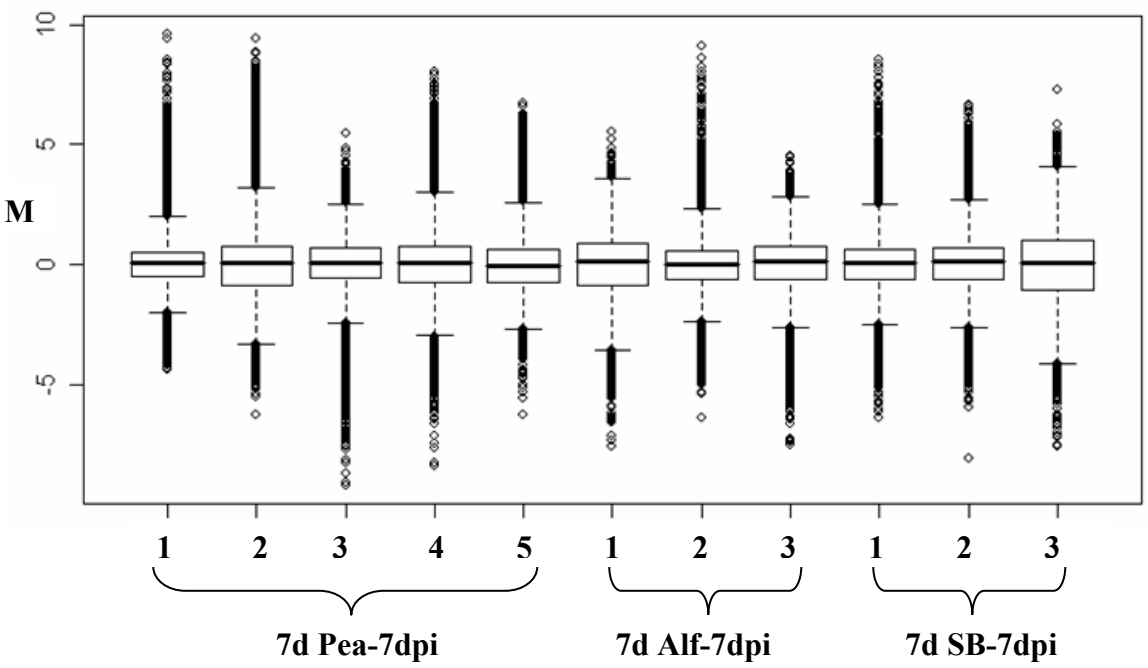



**Figure 4.2.1.** Boxplots showing the differential expression values of the five biological replicates of 7dPea-7dpi pea and 3 biological replicates of each 7dAlf-7dpi for the pea, alfalfa and sugar-beet rhizosphere microarray experiments. M represents the differential expression ($log_2$ (experiment/control)) after quantile normalization. In these experiments, the experimental RNA sample was labelled with Cy5 and control RNA sample was labelled with Cy3 except for the 3rd replicate of each condition, which is a dye-swap experiment.

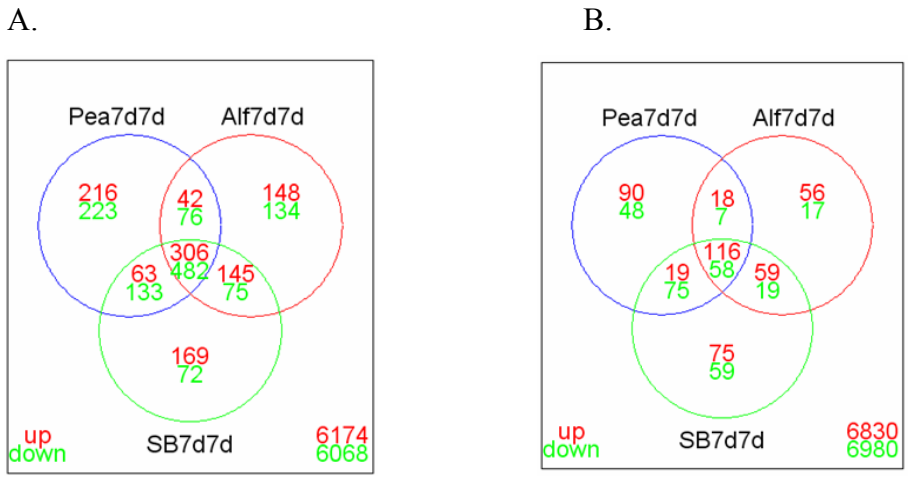
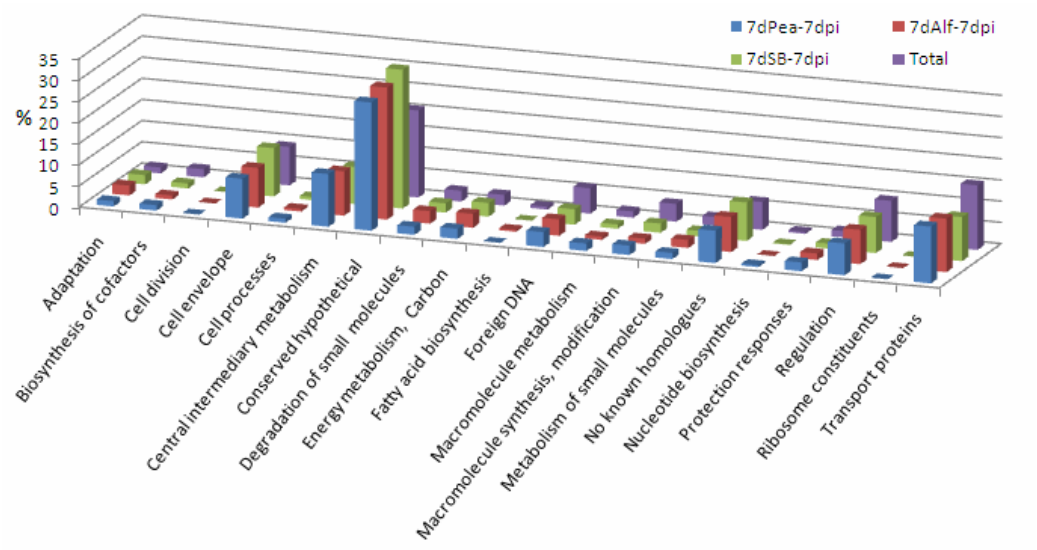



**Figure 4.2.2.** Venn diagrams showing the comparison of differentially regulated genes in indirect comparison of pea, alfalfa and sugar-beet rhizosphere experiments. (A) $log_2$ fold $\geq 1$ and adjusted P value $\leq 0.01$, (B) $log_2$ fold $\geq 2$ and adjusted P value $\leq 0.01$.

A.




B.

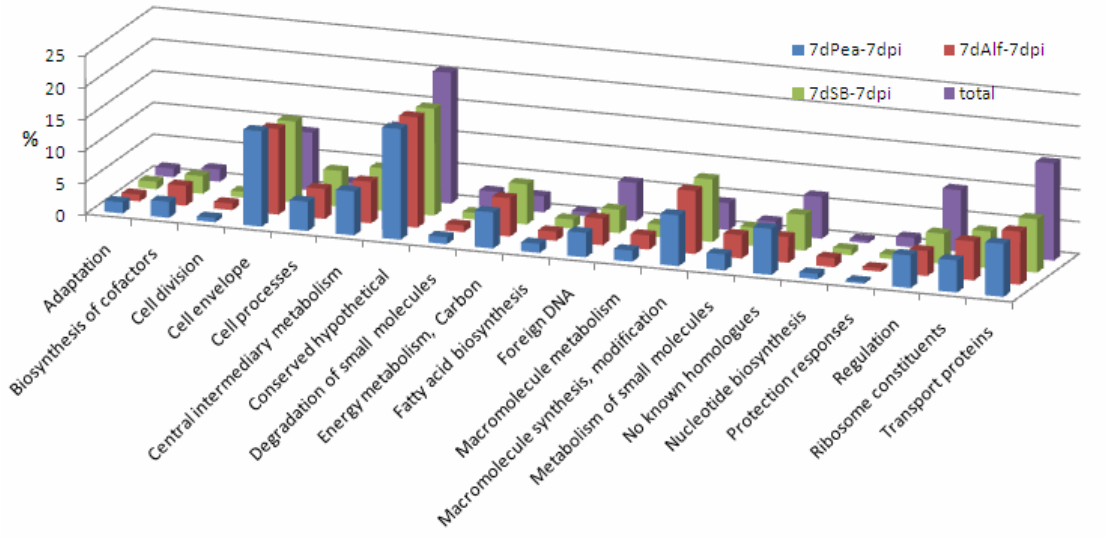



**Figure 4.2.3** Bar charts showing the comparison of percentage of biological functional class among the (A) up-regulated and (B) down-regulated genes obtained from pea, alfalfa and sugar-beet rhizospheres compared to the percentage of the functional class in the genome (total). In indirect comparison experiments, Rlv3841 cells grown in pea, alfalfa and sugar-beet rhizospheres separately compared against a common reference of Rlv3841 cells grown in AMS (10 mM glc, 10 mM $NH_4^+$) under laboratory conditions.

#### 4.2.1.1. Genes specifically induced in the pea rhizosphere

A total of 68 genes had high expression in the pea rhizosphere compared to alfalfa and sugar-beet. Of these up-regulated genes, a large proportion (26%) of the genes was located on pRL8 plasmid (Table 4.2.1.A). The $\log_2$ fold expression values of genes up-regulated on pRL8 are given in Table 4.2.1.

| | Distribution of the genes in the genome (%) | Distribution of the specifically up-regulated genes in the pea rhizosphere (%) of the total up-regulated genes |
|---|---|---|
| Chromosome | 65 | 48.5 |
| pRL12 | 11 | 10.0 |
| pRL11 | 9 | 4.4 |
| pRL10 | 6 | 2.9 |
| pRL9 | 4 | 7.4 |
| **pRL8** | **2** | **26.0** |
| pRL7 | 3 | 0 |

**Table 4.2.1.A.** Distribution of genes specifically induced in the pea rhizosphere relative to the distribution of genes in the genome.

These genes form part of four putative operons *pRL80021-5*, *pRL80026-30*, *pRL80036-40* and *pRL80060-64*. *pRL80021* (*coxG*) and *pRL80023-25* (*coxMSL*) codes for carbon monoxide dehydrogenase, and *pRL80038*, belonging to the cluster *pRL80036-40,* has moderate homology to CoxI family proteins, which may also be involved in carbon monoxide utilization. The operon containing *pRL80038* may be a part of a Cox accessory gene cluster, which generally flanks the *coxMSL* genes.

The next gene cluster *pRL80026-30* is annotated as a high affinity branched chain amino acid uptake transport system (*livJMHGF*). Only *livJ*, *livM*, *livF* was induced in the pea rhizosphere. The amino acid sequences do show only low identity (*livJ* < 30%, *livM* < 31%, *livH* < 33%, *livG* < 42%, *livF* < 46%) with transporters of other bacteria but not to those of related rhizobia, except for *livF* which has 46% identity to (bll6477) the ATP binding component of an ABC-T of *B. japonicum*, suggesting that this transport system may be unique to Rlv3841.

Mimosine (β-3-hydroxy-4 pyridone amino acid) is a toxin produced by leguminous trees (Figure 4.2.4). The last gene cluster, *pRL80060-64* annotated as an amino acid ABC transport system has moderate identity to the mimosine transport system of *Rhizobium sp*. TAL1145.

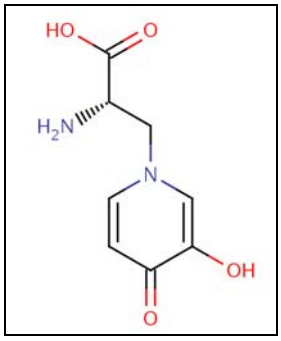



**Figure 4.2.4.** Structure of mimosine.

Homoserine is an abundant aminoacid found in pea root mucilage. *pRL80071* (*hom*) encoding homoserine dehydrogenase catalyzes the conversion of homoserine to L-aspartate-semialdehyde, a precursor in lysine biosynthesis. It has been reported that *R. leguminosarum* can utilize homoserine as a carbon source (Van Egeraat, 1975). The genes involved in the utilization of homoserine are not characterized. Rlv3841 may further utilize L-aspartate semialdehyde, a carbon source, in addition as precursor to methionine and lysine biosynthesis, and glycine, serine and threonine metabolism. The transport of homoserine into the cell has not been identified in Rlv3841. It is known that homoserine can be transported into the cell, through the LIV transport system, which normally transports leucine, isoleucine, valine and threonine. The induction of a high affinity branched amino acid ABC transport system (*livJMF*) specifically in the pea rhizosphere, suggests that it may transport homoserine into the cell. Further qRT-PCR studies of these transporter genes with RNA obtained from free-living Rlv3841 cells grown in homoserine as sole carbon source, is expected to provide more information.

| ID | Function | Name | $log_2$ fold Pea7d 7dpi | P value Pea7d 7dpi | $log_2$ fold Alf7d 7dpi | P value Alf7d 7dpi | $log_2$ fold SB7d 7dpi | P value SB7d 7dpi |
|---|---|---|---|---|---|---|---|---|
| pRL80017 | conserved hypothetical protein, pseudogene | - | 1.4 | 0 | -0.5 | 0.63793 | 0.4 | 0.68407 |
| pRL80021 | putative carbon monoxide dehydrogenase subunit G protein | *coxG* | 4.1 | 0 | 0.0 | 0.99946 | 0.5 | 0.67936 |
| pRL80022 | conserved hypothetical protein | - | 2.6 | 0 | 0.4 | 0.80542 | 0.3 | 0.84278 |
| pRL80023 | putative carbon monoxide dehydrogenase subunit | *coxM* | 3.1 | 0 | 0.5 | 0.59181 | -0.2 | 0.84208 |
| pRL80024 | putative iron-sulphur cluster carbon monooxide dehydrogenase subunit | *coxS* | 3.3 | 0 | 0.4 | 0.58164 | 0.0 | 0.99655 |
| pRL80025 | putative dehydrogenase/reductase | *coxL* | 3.1 | 0 | 0.0 | 0.99634 | 0.3 | 0.74174 |
| pRL80026 | putative branched-chain amino acid ABC transporter binding component | *livJ* | 4.2 | 0 | 0.2 | 0.67562 | 0.3 | 0.62894 |
| pRL80027 | putative branched-chain amino acid ABC transporter permease component | *livM* | 1.4 | 0 | 0.0 | 0.97175 | 0.1 | 0.91368 |
| pRL80030 | putative high-affinity branched-chain amino acid transport ATP-binding protein LivF | *livF* | 1.1 | 0 | -0.4 | 0.46458 | -0.3 | 0.61915 |
| pRL80036 | conserved hypothetical protein | - | 2.7 | 0 | 0.7 | 0.38736 | -0.2 | 0.89792 |
| pRL80037 | putative short-chain | - | 1.4 | 0 | -0.1 | 0.93651 | 0.1 | 0.90398 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | dehydrogenase | | | | | | | |
| pRL80038 | putative exported xanthine dehydrogenase/CoxI family protein | *coxI* | 1.6 | 0 | 0.9 | 0.13627 | 0.6 | 0.41862 |
| pRL80040 | putative HTH transcriptional regulatory protein | - | 1.1 | 1.00E-05 | 0.2 | 0.7677 | 0.6 | 0.26506 |
| pRL80054 | conserved hypothetical protein | - | 2.9 | 0 | 0.4 | 0.34619 | -0.6 | 0.24945 |
| pRL80059 | putative NifS-like cysteine desulfurase/selenocysteine lyase | - | 2.4 | 0 | 0.3 | 0.85584 | 0.1 | 0.93663 |
| pRL80060 | putative exported solute-binding protein | *midA like* | 3.5 | 0 | 0.3 | 0.90701 | 0.2 | 0.91443 |
| pRL80063 | putative ATP-binding component of ABC transporter | *midC like* | 1.4 | 0 | 0.1 | 0.8936 | 0.8 | 0.21339 |
| pRL80064 | putative permease component of ABC transporter | *midB like* | 2.6 | 0 | 0.3 | 0.76128 | 0.4 | 0.71835 |
| pRL80071 | putative Homoserine dehydrogenase | *hom* | 2.4 | 2.00E-05 | -0.5 | 0.79587 | -2.3 | 0.16677 |

**Table 4.2.1.** Genes located on pRL8 which had high induction in pea rhizosphere compared to alfalfa and sugar-beet rhizospheres. The expression values of alfalfa and sugar-beet are not significant, as the adjusted P values were greater than 0.01.

Considering that the expression of these genes is specific to the pea rhizosphere and their possible functions may suggest they may play an important role in determining host specificity. Additionally, many genes involved in other various biological functions were found to be specifically induced in the pea rhizosphere.

*RL0680* (*secDF*) [6.8U,-0.2,0.2] annotated as encoding a transmembrane export protein, was very highly induced in the pea rhizosphere, whereas it was repressed in those of alfalfa and sugar-beet. RL0680 has moderate amino acid identity (41%) with RL2055 (annotated as SecD, but in fact has both SecD and F domains fused into a single polypeptide as originally found in *Bacillus subtilis* (Bolhuis *et al.*, 1998)). In general, SecD and SecF are subunits which regulate the translocation of proteins in the cell. *RL0680* could be an alternative subunit in the Sec export machinery, permitting or promoting export of a sub-set of proteins required under specific conditions.

*RL4274* [7.2U,1.5U,1.5U] annotated as encoding HlyD family transmembrane efflux protein was very highly induced in the pea rhizosphere compared to those of alfalfa and sugar-beet rhizosphere. RL4274 has high identity (89%) to RHECIAT_CH0004005 coding for putative multidrug efflux protein in *R. etli* CIAT 652. Multidrug efflux proteins are membrane proteins, which pump out toxic compounds from cells. HlyD may be involved in detoxification, which may confer selective advantage during rhizosphere colonization.

*pRL120450* (*cpO*) [2.6U,-0.5,-1.5D] coding for chloroperoxidase was induced only in the pea rhizosphere. A homologue of this gene in *R. leguminosarum* bv. *trifolii* ANU794 was reported to be involved in host specificity (Morris & Djordjevic, 2006). Two genes, *RL1440* [2.5U,0.7,0.5] and *RL1251* (*degP*) [4.9U,0.7,0.4D] coding for serine proteases were highly induced in the pea rhizosphere compared to those of alfalfa and sugar-beet. Two other genes, *pRL110199* [7.9U,0.5,0.05] and *pRL120479* [5.9U,0.6,0.4] of unknown function, were very highly expressed in the pea rhizosphere. They are present only in Rlv3841 and not found in any other related rhizobia. Although, the exact functions of these genes are unknown, they may play an essential role in host specificity. *RL0054* [4.3U,0.3,-0.2] encoding malate synthase G (*glcB*) a key

enzyme in glyoxylate cycle was highly induced in the pea rhizosphere compared to alfalfa and sugar-beet rhizospheres, indicating that the glyoxylate cycle is operating only in the pea rhizosphere.

In addition, *pRL90305*, [1.5U,0.08,0.4] was induced only in the pea rhizosphere. The pRL90305 sequence shows high identity (95%) to a fosmidomycin-resistance protein in a wide range of bacteria including *R. etli* and *S. meliloti*. Fosmidomycin (sodium hydrogen 3-(*N*-hydroxyfomamido) propylphosphonate) is a phosphonate antibiotic that specifically inhibits 1-deoxy-D-xylulose 5-phosphate (DXP) reductoisomerase (Dxr), a key enzyme in the non-mevalonate pathway (1-deoxy-D-xylulose-5-phosphate (DOXP) pathway) of steroid biosynthesis (KEGG PATH: rle00100). The presence of the DOXP pathway in Rlv3841 is indicated by the expression of *RL0973* (*dxs*) [1.5U,0.06,0.2], coding for DXP synthase, which catalyzes the first step in the formation of 1-deoxy-D-xylulose-5-phosphate from pyruvate and glyceraldehyde-3-phosphate. The second step is catalyzed by DXP reductoisomerase (Dxr) (RL4372), followed by IspDF (RL2254), IspE (RL0935), IspG (RL4630), and IspH (RL1030) to isopentenyldiphosphate (IPP) (Figure 4.2.5). However, these genes were not induced.

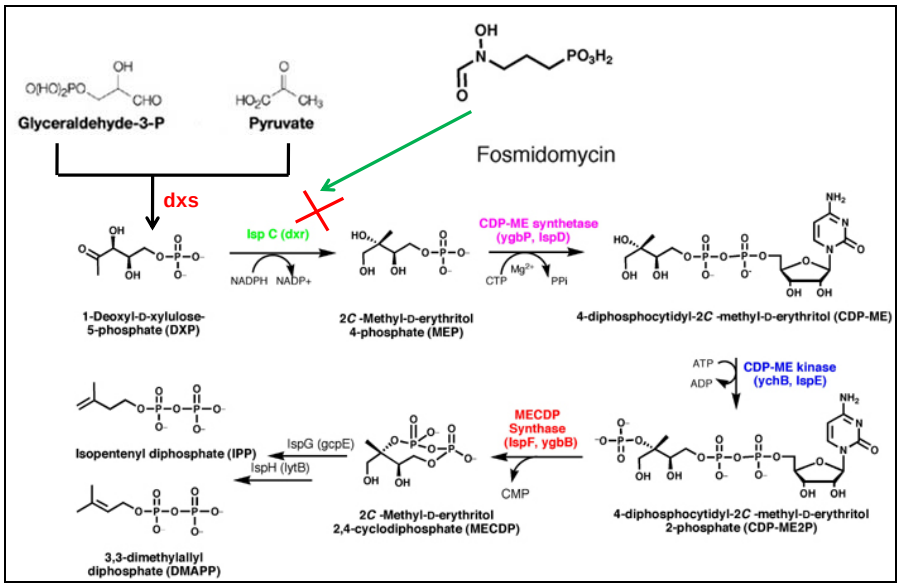



**Figure 4.2.5.** The non-mevalonate pathway for the biosynthesis of isopentenyl diphosphate (IPP).

A possible explanation of these observations, is that pea root exudates may contain fosmidomycin or analogous compounds, thus inhibiting DXP reductoisomerase (Dxr) and stopping the conversion of DXP to 2-C-methylerythritol 4-phosphate (MEP), preventing the formation of IPP. Rlv3841 may gain a selective pressure and competitive advantage during rhizosphere colonization as it will be resistant to fosmidomycin and therefore able to overcome the inhibition of Dxr that it causes.

*RL0786* [3.8U,0.6,0.4], is highly homologous to chalcone synthase-like protein (Type III polyketide synthases) of *R. leguminosarum* bv. *trifolii* (98% Id) and *R. etli* CFN42 (90% Id). In plants, chalcone synthase catalyzes the first step in flavonoid biosynthesis through the phenylpropanoid pathway. *RL0786* was highly induced in the pea rhizosphere, indicating a possible role for flavonoid biosynthesis in host-specific interactions. Although the precise role of chalcone synthase in rhizobia has not been characterized, we can speculate that it may be involved in biosynthesis of host-specific compounds for bacteria to defend themselves against host-derived compounds.

Overall, most of the genes shown to be specifically induced in the pea rhizosphere in these experiments have not been previously reported (Table 4.2.2.). Apart from genes which have been previously characterized and annotated, the pea-specific dataset contains many genes of unknown function. The characterization of these genes will provide more insights in the molecular determinants which determine host specificity during symbiotic interactions.

| ID | Function | Name | $log_2$ fold | | |
|---|---|---|---|---|---|
| | | | 7dPea-7dpi | 7dAlf-7dpi | 7dSB-7dpi |
| pRL100224 | putative nitrilotriacetate monooxygenase component | - | 2.3 | 0.4 | 0.4 |
| pRL100265 | putative permease component of ABC transporter | - | 2.4 | -0.2 | 0.5 |
| pRL110199 | conserved hypothetical protein | - | 7.9 | 0.5 | 0.0 |
| pRL110313 | putative transmembrane cardiolipin synthetase | *cls* | 1.5 | -0.6 | -0.7 |
| pRL110377 | putative Kdp operon transcriptional regulatory protein | *kdpE* | 2.4 | 0.0 | 0.0 |
| pRL120450 | putative chloroperoxidase | *cpO* | 2.6 | -0.6 | -1.6 |
| pRL120479 | conserved hypothetical exported protein | - | 5.9 | 0.6 | 0.4 |
| pRL120480 | putative fimbrial operon related protein | - | 4.0 | 1.1 | 0.1 |
| pRL120584 | putative epimerase | - | 1.8 | 0.5 | 0.3 |
| pRL120585 | putative epimerase | - | 2.7 | 1.3 | 1.1 |
| pRL120624 | hypothetical protein | - | 3.0 | 0.6 | -0.1 |
| pRL120625 | putative peptidase | - | 4.0 | 0.3 | 0.6 |
| pRL80021 | putative carbon monoxide dehydrogenase subunit G protein | *coxG* | 4.1 | 0.0 | 0.5 |
| pRL80022 | conserved hypothetical protein | - | 2.6 | 0.4 | 0.3 |
| pRL80023 | putative carbon monoxide dehydrogenase subunit | *coxM* | 3.1 | 0.5 | -0.2 |
| pRL80024 | putative iron-sulphur cluster carbon monoxide dehydrogenase subunit | *coxS* | 3.3 | 0.4 | 0.0 |
| pRL80025 | putative dehydrogenase/reductase | *coxL* | 3.1 | 0.0 | 0.3 |
| pRL80026 | putative branched-chain amino acid ABC transporter binding component | *livJ* | 4.2 | 0.2 | 0.3 |
| pRL80027 | putative branched-chain amino acid ABC transporter permease component | *livM* | 1.4 | 0.0 | 0.1 |
| pRL80030 | putative high-affinity branched-chain amino acid transport ATP-binding protein | *livF* | 1.1 | -0.4 | -0.3 |
| pRL80036 | conserved hypothetical protein | - | 2.7 | 0.7 | -0.2 |
| pRL80037 | putative short-chain dehydrogenase | - | 1.4 | -0.1 | 0.1 |
| pRL80038 | putative exported xanthine dehydrogenase/CoxI family protein | *coxI* | 1.6 | 0.9 | 0.6 |
| pRL80040 | putative HTH transcriptional regulatory protein | - | 1.1 | 0.2 | 0.6 |
| pRL80054 | conserved hypothetical protein | - | 2.9 | 0.4 | -0.6 |
| pRL80059 | putative NifS-like cysteine desulfurase/selenocysteine lyase | - | 2.4 | 0.3 | 0.1 |
| pRL80060 | putative exported solute-binding protein | - | 3.5 | 0.3 | 0.2 |
| pRL80063 | putative ATP-binding component of ABC transporter | - | 1.4 | 0.1 | 0.8 |
| pRL80064 | putative permease component of ABC transporter | - | 2.6 | 0.3 | 0.4 |
| pRL80071 | putative Homoserine dehydrogenase | *hom* | 2.4 | -0.5 | -2.3 |
| pRL90056 | putative GntR family transcriptional regulator | - | 2.1 | 0.1 | -0.2 |

| | | | | | |
|---|---|---|---|---|---|
| pRL90085 | putative periplasmic substrate binding component of ABC transporter | - | 4.7 | 1.1 | 0.2 |
| pRL90087 | putative AraC family helix-turn-helix transcriptional regulator | - | 1.2 | 0.7 | 0.6 |
| pRL90091 | putative permease component of ABC transporter | - | 2.4 | 0.6 | -0.8 |
| pRL90305 | putative transmembrane MFS family transporter protein | - | 1.5 | 0.1 | 0.4 |
| RL0054 | Malate synthase G | *glcB* | 4.3 | 0.2 | -0.2 |
| RL0551 | chaperonin, heat shock protein | *hslO* | 2.3 | 0.0 | 0.3 |
| RL0572 | putative dihydroorotate dehydrogenase | - | 3.4 | -0.3 | -0.3 |
| RL0680 | putative transmembrane export SecD/F family protein | *secD* | 6.7 | -0.3 | 0.3 |
| RL0755 | putative glutamine synthetase | - | 1.4 | -1.2 | -1.1 |
| RL0786 | conserved hypothetical protein | - | 3.8 | 0.6 | 0.4 |
| RL0830 | hypothetical protein | - | 1.2 | -0.3 | -0.4 |
| RL0836 | putative Insertion sequence ATP-binding protein | - | 1.0 | -0.3 | 0.3 |
| RL0866 | putative glycolate oxidase iron-sulfur subunit | *glcF* | 1.1 | 0.0 | -0.4 |
| RL0959 | putative FAD binding oxidoreductase | - | 1.7 | -0.1 | 0.0 |
| RL0964 | putative cationic transport protein | - | 3.3 | 0.2 | -0.5 |
| RL0973 | putative 1-deoxy-d-xylulose-5-phosphate synthase | *dxs* | 1.6 | 0.1 | -0.2 |
| RL0996 | putative transmembrane transporter | - | 2.3 | 0.5 | 0.3 |
| RL1251 | putative serine protease | *degP* | 4.9 | 0.7 | 0.3 |
| RL1297 | conserved hypothetical protein | - | 3.9 | -1.2 | -1.6 |
| RL1440 | putative serine protease | *degP* | 2.6 | 0.7 | 0.5 |
| RL1546 | putative amidophosphoribosyltransferase | *purF* | 2.4 | -0.4 | -0.7 |
| RL1606 | putative esterase (beta-lactamase family) | - | 3.5 | 0.8 | 0.7 |
| RL1721 | putative ATP-binding component of ABC transporter | - | 1.2 | -0.8 | -0.6 |
| RL2051 | putative protein-L-isoaspartate O-methyltransferase | *pcm* | 1.2 | 0.0 | 0.1 |
| RL2193 | conserved hypothetical protein | - | 1.2 | -0.2 | 0.5 |
| RL2194 | conserved hypothetical protein | - | 2.9 | -0.4 | 1.6 |
| RL2769 | putative alkanesulfonate monooxygenase | *ssuD* | 2.0 | 0.6 | -0.2 |
| RL3345 | putative aldehyde dehydrogenase | - | 3.1 | 0.0 | -0.7 |
| RL3347 | putative permease component of ABC transporter | - | 1.9 | 0.9 | 0.5 |
| RL3348 | putative N-formylglutamate amidohydrolase family protein | - | 1.0 | 0.4 | -0.1 |
| RL3349 | putative transcriptional regulator | - | 1.0 | 0.2 | 0.1 |
| RL3366 | putative flavoprotein | - | 5.0 | 0.8 | 0.5 |
| RL3451 | putative short-chain dehydrogenase/reductase | - | 2.4 | 0.1 | 0.8 |
| RL4186 | putative oxidoreductase | - | 2.8 | 0.0 | -0.1 |
| RL4212 | putative ribulokinase | *araB* | 1.2 | -0.9 | 0.2 |
| RL4244 | putative permease component of ABC transporter | - | 3.5 | 0.7 | 0.3 |

| | | | | | |
|---|---|---|---|---|---|
| RL4274 | putative HlyD family transmembrane efflux protein | - | 7.3 | 1.6 | 1.6 |

**Table 4.2.2.** Genes which are highly induced in the pea rhizosphere compared to those of alfalfa and sugar-beet. The expression values of alfalfa and sugar-beet are not significant, as the adjusted P values were greater than 0.01.

#### 4.2.1.2 Genes induced in both pea and alfalfa rhizospheres (legume specific)

Venn mapping of the three rhizosphere gene expression datasets enabled identification of a set of genes expressed in pea and alfalfa rhizospheres and repressed or nearly repressed in that of sugar-beet. These can be described as legume-specific genes (Table 4.2.3). A close gene-by-gene analysis eliminated those genes whose expression in the sugar-beet rhizosphere was nearly equal or close to the expression in pea and alfalfa rhizospheres. In the final set of legume specific genes, the most prominent set of genes were of those of the *nod* operon (*nodOMFABCIJ*), which showed high expression in the rhizospheres of both pea and alfalfa.

*pRL120200* (*eryG*) [1.4U,1.4U,0.4], encoding an erythritol-binding protein ABC-T was induced only in pea and alfalfa rhizospheres and not in that of sugar-beet. Strains of *R. leguminosarum* VF39 mutated in the erythritol transport genes showed reduced competitiveness in nodulation compared with wild-type, but were not affected in nitrogen fixation (Yost *et al.*, 2006).

pRL120027 [5.3U,1.3U,-0.2], encoding an aldolase, which has high identity to deoxyribose-phosphate aldolase of *R. leguminosarum* bv. *trifolii* and *M. loti* and moderate homology to fructose bisphosphate aldolase in many archaea, was highly induced in pea and alfalfa rhizospheres to that of sugar-beet rhizosphere. In addition many genes, *RL3130* [6.3U,3.3U,0.05] encoding a phosphatase, *pRL120632* [5.4U,2.5U,0.2] encoding a dehydrogenase and *pRL100444* [7.3U,3.2U,0.9] encoding an oxidoreductase were highly induced in the pea and alfalfa rhizospheres compared to that of sugar-beet.

The importance of *nod* genes and their specific role in enabling colonization of legume hosts via nodulation, is well known. Other genes, specifically induced in response to legumes may also play an important role in colonization of legume rhizospheres (pea and alfalfa). Identification of genes specifically expressed in the legume rhizosphere, again proves the success of this strategy.

| ID | Function | Name | $log_2$ fold | | |
|---|---|---|---|---|---|
| | | | 7dPea-7dpi | 7dAlf-7dpi | 7dSB-7dpi |
| pRL100175 | putative nodulation protein | *nodO* | 1.939 | 2.679 | -0.791 |
| pRL100180 | glucosamine-fructose-6-phosphate aminotransferase | *nodM* | 1.364 | 3.164 | -1.565 |
| pRL100183 | nodulation protein F | *nodF* | 1.901 | 3.283 | -0.018 |
| pRL100185 | nodulation protein A | *nodA* | 1.665 | 2.479 | 0.192 |
| pRL100186 | chitooligosaccharide deacetylase | *nodB* | 2.798 | 3.123 | 0.567 |
| pRL100187 | N-acetylglucosaminyltransferase | *nodC* | 3.723 | 4.671 | -0.518 |
| pRL100188 | nod factor export ATP-binding protein I | *nodI* | 1.834 | 2.805 | -1.156 |
| pRL100189 | nodulation protein | *nodJ* | 1.657 | 2.589 | 0.414 |
| pRL100444 | putative oxidoreductase | - | 7.334 | 3.202 | 0.926 |
| pRL110624 | conserved hypothetical protein | - | 5.888 | 2.444 | -0.337 |
| pRL120027 | putative aldolase | - | 5.225 | 1.352 | -0.266 |
| pRL120200 | periplasmic erythritol binding protein | *eryG* | 1.36 | 1.338 | -0.03 |
| pRL120438 | hypothetical protein | - | 3.905 | 3.031 | 0.531 |
| pRL120609 | putative permease component of ABC transporter | - | 1.131 | 2.536 | 0.516 |
| pRL120632 | putative dehydrogenase | - | 5.325 | 2.497 | 0.145 |
| pRL80075 | putative endoribonuclease L-PSP family protein | - | 2.183 | 3.578 | -0.359 |
| pRL80076 | putative aliphatic nitrilase | - | 1.719 | 2.219 | -1.271 |
| RL1489 | putative transmembrane MatE family transporter | - | 1.019 | 1.274 | -0.835 |
| RL1972 | conserved hypothetical protein | - | 1.792 | 2.157 | -0.352 |
| RL3130 | putative phosphatase protein | - | 6.361 | 3.248 | 0.049 |

**Table 4.2.3.** Genes which were expressed in legume (pea and alfalfa) rhizospheres and repressed in the sugar-beet rhizosphere.

#### 4.2.1.3 Genes induced in the alfalfa rhizosphere

It is interesting to study the behaviour of Rlv3841 in the alfalfa rhizosphere, a legume, but not a cognate host. As expected, a large portion of genes up-regulated in the alfalfa rhizosphere overlap with the genes up-regulated in the pea rhizosphere. Analysis of the genes specifically induced in the alfalfa rhizosphere after discarding the genes with expression values close to those genes up-regulated in pea rhizosphere, enable some interesting observations.

The gene cluster encoding malonate uptake and catabolic genes, (*RL0990* (*matA*) [0.4,1.7U,0.2] encoding a malonyl-CoA decarboxylase, *RL0991* (*matB*) [0.3,1.7U,0.1] Malonyl CoA synthetase and *RL0992* (*matC*) [0.5,2.8U,1.0] encoding a transmembrane dicarboxylate carrier protein) were strongly induced, suggesting the presence of malonate, in the alfalfa rhizosphere. RL0992, encoding the malonate transporter was reported to be induced by malonate (Mauchline *et al.*, 2006). As well as being induced in the alfalfa rhizosphere, *RL0992* was also slightly induced in the sugar-beet rhizosphere, but did not pass the P value cut-off. However, it is important to note that this gene cluster was also induced by hesperetin (see Chapter 3).

From earlier reports, it is known that the surfaces of legume roots contain high amounts of arabinose/galactose rich glycoproteins. The pea root mucilage was also reported to contain a large proportion of arabinose sugars (Knee *et al.*, 2001). A recent transportome study in *S. meliloti*, characterized many previously uncharacterized transport systems, of which three transport systems were induced by L-arabinose (*SMb21587*, *SMa0252* and *SMa0203*) (Mauchline *et al.*, 2006). Apart from this, two more systems, *SMb20895* (Poysti *et al.*, 2007) and *SMb20506* (*araH*), are also present in *S. meliloti*. Homologues of all these transport systems are also present in Rlv3841 with very high to moderate homology. The availability of arabinose in the rhizosphere and its preferential use by rhizobia is indicated by the presence of many transports systems specific to arabinose or arabinose-like sugars. In the pea rhizosphere, only components of two arabinose transport systems (*pRL120284* (*SMa0203* - 91% Id with *SMa0203*) and *RL4229* (*SMb20506* - 86% Id with *SMb20506*) were induced. Interestingly, in

the alfalfa rhizosphere it was only *RL4229* [1.6U,4.2U,0.9] encoding ABC-T permease component, that was induced. In addition, *RL0632* (*araA*) [0.1,2.9U,0.2] encoding L-arabinose isomerase catalyzing first step, in the conversion of arabinose to ribulose was induced. Taken together these observations suggest the presence of arabinose or arabinose-like sugars in both pea and alfalfa rhizospheres.

*RL0102* (*gabT*) [0.5,2.8U,0.3] (encoding a putative 4-aminobutyrate aminotransferase) and *RL4385* (*gabT3*) [0.5,1.6U,0.2] (encoding a putative ornithine aminotransferase involved in GABA utilization) were also specifically induced in the alfalfa rhizosphere.

There was specific induction of two ABC-T belonging to the QAT family of transporters: *pRL120515* (*qatW3*) [1.1U,3.2U,1.7U] encoding a permease component of ABC-T and *pRL120516* (*qatX3*) [0.1,1.8U,0.7] encoding an SBP-ABC-T of quaternary amine transporter system. A database similarity search revealed that pRL120516 has high homology (74% Id) with RNGR00282 annotated as glycine-betaine/L-proline binding protein of *Rhizobium sp.* NGR234 and also found in a wide variety of bacteria.

*pRL120517* [0.9,1.3U,1.3U] coding for a putative pyrroline-5-carboxylate reductase, catalyzing the conversion of L-proline to 1-pyroline-5-carboxylate in proline metabolism, and is located downstream of *pRL120516*, which was also induced specifically in the alfalfa rhizosphere. In addition *pRL120053* (*arcB*) [0.5,1.4U,0.4] and *RL3095* (*arcB*) [0.3,2.2U,-0.1] coding for putative ornithine cyclodeaminases which catalyze the conversion of L-ornithine to L-proline and $NH_3$ in proline metabolism, were induced. This is intriguing and suggests that an amino acid found on the biosynthetic pathway leading to proline, e.g. ornithine is found rich in the alfalfa rhizosphere.

*RL0985* [0.6,1.3U,0.4] coding for an SBP-ABC-T having moderate identity to SMb21097 of *S. meliloti* (30% Id and 45% Smty) was induced. In *S. meliloti*, SMb21097, (encoding an SBP of PAAT family) was shown to be induced by citrulline (Mauchline *et al.*, 2006). Earlier reports, showed the presence in *S. meliloti* of ornithine cyclodeaminase activity (Soto *et al.*, 1994), an enzyme which catalyzes the conversion of ornithine to proline. A proline

dehydrogenase (*putA*) mutant of *S. meliloti* showed that proline metabolism is necessary for rhizosphere competitiveness and nodulation in *S. meliloti*-alfalfa symbiosis (Jiménez-Zurdo *et al.*, 1995). Taken together these observations are in line with the earlier reports of the presence of ornithine in the alfalfa rhizosphere and its catabolism via proline formation. This array data suggests that ornithine and proline are specifically important in the alfalfa rhizosphere.

Overall these observations portray the metabolically adaptable nature of Rlv3841, even in non-host rhizospheres. The induction of a specific set of genes involved in transport and catabolism, such as those concerned with malonate, ornithine and proline, solely in the alfalfa rhizosphere provides information about the composition of nutrients in the alfalfa rhizosphere. It also substantiates the integrity and power of this strategy to study the large scale gene expression of rhizosphere bacteria.

### 4.2.2 Comparative rhizosphere transcriptomics (Direct comparison)

In this experiment, the three rhizosphere samples were compared directly to each other without a common reference. This approach is statistically rigorous, very powerful and a perfect choice for comparative transcriptomics. To ensure this approach maximises more consistency and statistical significance, a two-colour direct approach was designed to include both biological and technical replicates. In this two-colour direct comparison, two biological replicates from each rhizosphere sample were extracted and amplified. From each amplified sample, two equal amounts of samples were taken, reverse transcribed and labelled with Cy3 and Cy5 (technical replicates) as described in section 2.3 (Figure 4.1.1.A). For each rhizosphere sample, two of each Cy3- and Cy5-labelled cDNA was obtained and the microarrays were performed as described (Figure 4.1.1.B). Each experiment consists of six microarrays with three dye-swap pairs. Two replicates of the two-colour direct comparisons were performed, which is four biological replicates for each rhizosphere sample (total of 12 microarrays). Microarrays were performed as described in section 2.3 and the data were analyzed by LIMMA using a direct two-color comparison model. Normalization was carried out in LIMMA as described earlier in section 2.3 (Figure 4.2.6). The total number of differentially expressed genes passing the criteria was as shown in the Table 4.2.4.

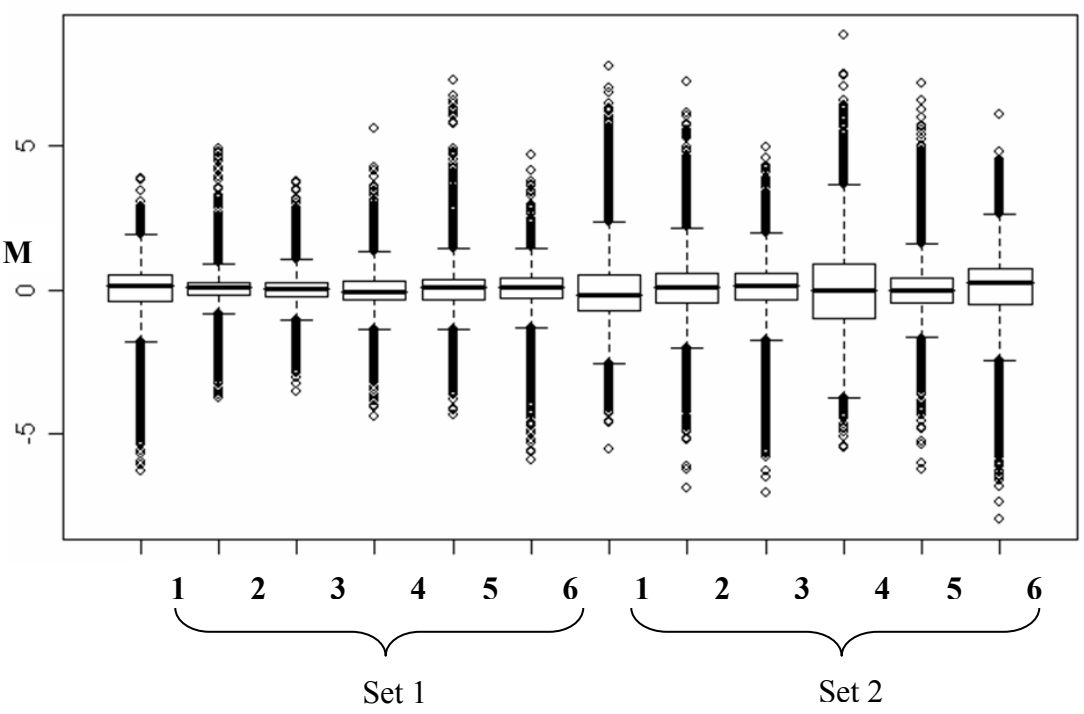



**Figure 4.2.6.** Boxplots showing the differential expression values of the two sets of two-colour direct comparison experiments performed with 7dPea-1dpi, 7dAlf-7dpi and 7dSB-7dpi as shown in Figure 4.1.1. M represents the differential expression ($log_2$ (experiment/control)) after quantile normalization. The labelling and dye-swap were as described in the direct comparison (Figure 4.1.1).

| Experiment | Up-regulated | Down-regulated |
|---|---|---|
| 7d Pea 7dpi Vs 7d Alf 7dpi | 90 | 37 |
| 7d Pea 7dpi Vs 7d SB 7dpi | 129 | 79 |
| 7d Alf 7dpi Vs 7d SB 7dpi | 58 | 44 |

**Table 4.2.4.** The differentially expressed genes from direct comparison two-colour rhizosphere microarrays performed with samples recovered from pea (Pea) alfalfa (Alf), sugar-beet (SB) rhizospheres.

#### 4.2.1.2 Genes differentially expressed in the direct comparison of pea and alfalfa rhizospheres

RNA, extracted from Rlv3841 cells recovered seven days post-inoculation of 7day-old pea and alfalfa rhizospheres was directly compared. After normalization and filtering, 90 genes were found to be induced specifically in the pea rhizosphere (Table 4.2.5) and 44 genes induced specifically in that of alfalfa. Of these 90 genes, 42 were annotated as "conserved hypothetical" and the assigned or possible functions, of the remaining genes has been discussed earlier in the section 3.2.3 and 4.2.1.

| ID | Function | Name | $log_2$ fold |
|---|---|---|---|
| **Biosynthesis of cofactors, carriers** | | | |
| RL1959 | putative acyl carrier protein phosphodiesterase | *acpD* | 1.7 |
| **Cell envelope** | | | |
| pRL120480 | putative fimbrial operon related protein | - | 1.4 |
| RL0469 | putative exported surface protein | - | 2.3 |
| RL1173 | putative transmembrane protein | - | 1.7 |
| RL3234 | putative lipoprotein | - | 1.8 |
| **Metabolism** | | | |
| pRL100444 | putative oxidoreductase | - | 2.9 |
| pRL120027 | putative aldolase | - | 1.8 |
| pRL120632 | putative dehydrogenase | - | 1.5 |
| pRL80021 | putative carbon monoxide dehydrogenase subunit G protein | *coxG* | 2.5 |
| pRL80023 | putative carbon monoxide dehydrogenase subunit | *coxM* | 2.2 |
| pRL80024 | putative iron-sulphur cluster carbon monooxide dehydrogenase subunit | *coxS* | 2.1 |
| pRL80025 | putative dehydrogenase/reductase | *coxL* | 1.9 |
| pRL80037 | putative short-chain dehydrogenase | - | 1.1 |
| RL0959 | putative FAD binding oxidoreductase | - | 1.3 |
| RL1343 | putative short-chain dehydrogenase/reductase | - | 1.0 |
| RL1606 | putative esterase (beta-lactamase family) | - | 1.1 |
| RL3130 | putative phosphatase protein | - | 1.6 |
| RL3345 | putative aldehyde dehydrogenase | - | 1.1 |
| RL3366 | putative flavoprotein | - | 1.9 |
| RL4186 | putative oxidoreductase | - | 1.8 |
| **Macromolecule metabolism** | | | |
| pRL120625 | putative peptidase | - | 2.1 |
| pRL80062 | putative endoribonuclease | - | 1.0 |
| RL1251 | putative serine protease | *degP* | 2.8 |
| RL1295 | putative peptidase | - | 1.7 |
| RL1440 | putative serine protease | *degP* | 1.7 |
| **Metabolism of small molecules** | | | |
| pRL80059 | putative NifS-like cysteine desulfurase/selenocysteine lyase | - | 1.2 |
| pRL80071 | putative Homoserine dehydrogenase | *hom* | 1.6 |
| **Nucleotide biosynthesis** | | | |
| RL1546 | putative amidophosphoribosyltransferase | *purF* | 1.4 |
| RL0572 | putative dihydroorotate dehydrogenase | - | 2.2 |
| **Detoxification** | | | |
| RL2769 | putative alkanesulfonate monooxygenase | *ssuD* | 1.8 |
| RL3016 | putative protocatechuate 3,4-dioxygenase beta chain (3,4-pcd) like protein | *pcaH2* | 1.2 |
| **Transport/binding proteins** | | | |
| pRL80026 | putative branched-chain amino acid ABC transporter binding component | *livJ* | 2.4 |
| pRL80027 | putative branched-chain amino acid ABC transporter permease component | *livM* | 1.0 |
| pRL80030 | putative high-affinity branched-chain amino acid transport ATP-binding protein LivF | *livF* | 1.2 |
| pRL80060 | putative exported solute-binding protein | - | 2.2 |

| | | | |
|---|---|---|---|
| pRL80064 | putative permease component of ABC transporter | - | 2.0 |
| pRL90085 | putative periplasmic substrate binding component of ABC transporter | - | 2.0 |
| RL0680 | putative transmembrane export SecD/F family protein | *secD* | 2.6 |
| RL0964 | putative cationic transport protein | - | 1.1 |
| RL0996 | putative transmembrane transporter | - | 1.0 |
| RL3721 | putative permease component of ABC transporter | - | 1.8 |
| RL3860 | putative permease component of ABC transporter | - | 1.4 |
| RL4244 | putative permease component of ABC transporter | - | 1.1 |
| RL4274 | putative HlyD family transmembrane efflux protein | - | 3.6 |
| **Unknown functions** | | | |
| pRL100148 | putative thiamine biosynthesis protein, pseudogene | *thiC* | 1.5 |
| pRL110199 | conserved hypothetical protein | - | 3.1 |
| pRL110510 | conserved hypothetical protein | - | 1.0 |
| pRL110624 | conserved hypothetical protein | - | 1.1 |
| pRL120479 | conserved hypothetical exported protein | - | 2.9 |
| pRL120624 | hypothetical protein | - | 1.1 |
| pRL80020 | conserved hypothetical protein | - | 1.9 |
| pRL80022 | conserved hypothetical protein | - | 1.9 |
| pRL80036 | conserved hypothetical protein | - | 2.0 |
| pRL80054 | conserved hypothetical protein | - | 1.8 |
| RL0231 | conserved hypothetical protein | - | 1.2 |
| RL0439 | conserved hypothetical exported protein | - | 1.1 |
| RL0447 | conserved hypothetical exported protein | - | 1.5 |
| RL0786 | conserved hypothetical protein | - | 1.3 |
| RL1297 | conserved hypothetical protein | - | 3.4 |
| RL1314 | hypothetical exported protein | - | 1.9 |
| RL1369 | putative pentapeptide repeat protein | - | 1.3 |
| RL1631 | conserved hypothetical exported protein | - | 3.5 |
| RL1925 | conserved hypothetical protein | - | 1.0 |
| RL2554 | hypothetical exported protein | - | 3.6 |
| RL2800 | conserved hypothetical protein | - | 1.1 |
| RL3065 | conserved hypothetical exported protein | - | 1.1 |
| RL3098 | putative glycine-asparagine rich protein | - | 1.3 |
| RL3151 | conserved hypothetical exported protein | - | 1.0 |
| RL3339 | putative ribonuclease-L-PSP family protein | - | 1.1 |
| RL3384 | conserved hypothetical exported protein | - | 1.3 |
| RL3489 | conserved hypothetical protein | - | 1.6 |
| RL3592 | conserved hypothetical exported protein | - | 1.6 |
| RL4562 | conserved hypothetical exported protein | - | 1.0 |
| RL4601 | conserved hypothetical protein | - | 1.4 |

**Table 4.2.5.** Genes up-regulated genes in pea (7dPea-7dpi) and alfalfa (7dAlf-7dpi) rhizospheres by direct comparison.

#### 4.2.1.3. Genes differentially expressed in pea and sugar-beet rhizospheres examined by direct comparison

RNA extracted from Rlv3841 cells recovered after seven days post-inoculation from 7day-old pea and sugar-beet rhizospheres was directly compared. After normalization and filtering, 129 genes were found to be specifically induced in the pea rhizosphere and 79 genes were down-regulated (i.e those 79 genes were induced in the sugar-beet rhizosphere). Genes up-regulated in the pea rhizosphere include those coding for nod factor signalling (*pRL100175-189*), for carbon monoxide dehydrogenase (CODH) (*pRL80021-25*), high affinity amino acid ABC-T components (*pRL80026-7*), mimosine-like compound ABC-T (*pRL80060-4*) and 38 genes of unknown function (Table 4.2.6).

| ID | Function | Name | $log_2$ fold |
|---|---|---|---|
| **Cell envelope** | | | |
| pRL120480 | putative fimbrial operon related protein | - | 1.8 |
| RL0469 | putative exported surface protein | - | 1.4 |
| RL0963 | putative transmembrane/surface protein | - | 1.1 |
| RL1173 | putative transmembrane protein | - | 2.5 |
| RL1499 | putative outer membrane protein | *ropA* | 1.4 |
| RL2969 | putative transmembrane protein | - | 1.2 |
| RL3066 | putative transmembrane protein | - | 1.1 |
| RL3186 | putative transmembrane protein | - | 1.1 |
| RL3234 | putative lipoprotein | - | 1.5 |
| **Cell processes** | | | |
| RL0702 | putative flagella motor switch protein | *fliM* | 1.5 |
| **Metabolism** | | | |
| pRL100067 | putative racemase/decarboxylase | - | 2.4 |
| pRL100175 | putative nodulation protein | *nodO* | 2.3 |
| pRL100180 | glucosamine--fructose-6-phosphate aminotransferase | *nodM* | 2.0 |
| pRL100183 | nodulation protein F | *nodF* | 1.4 |
| pRL100185 | nodulation protein A | *nodA* | 1.2 |
| pRL100186 | chitooligosaccharide deacetylase | *nodB* | 1.8 |
| pRL100187 | N-acetylglucosaminyltransferase | *nodC* | 2.7 |
| pRL100188 | nod factor export ATP-binding protein I | *nodI* | 1.8 |
| pRL100189 | nodulation protein | *nodJ* | 1.3 |
| pRL100226 | putative acetyltransferase | - | 1.1 |
| pRL100444 | putative oxidoreductase | - | 5.2 |
| pRL120027 | putative aldolase | - | 3.5 |
| pRL120585 | putative epimerase | - | 1.2 |
| pRL120632 | putative dehydrogenase | - | 2.8 |
| pRL120789 | putative nodulation protein NolR | *nolR* | 1.3 |
| pRL80021 | putative carbon monoxide dehydrogenase subunit G protein | *cutL* | 2.7 |
| pRL80023 | putative carbon monoxide dehydrogenase subunit | *cutM* | 2.4 |
| pRL80024 | putative iron-sulphur cluster carbon monooxide dehydrogenase subunit | *cutS* | 2.7 |
| pRL80025 | putative dehydrogenase/reductase | - | 2.0 |
| pRL80037 | putative short-chain dehydrogenase | - | 1.2 |
| RL0054 | malate synthase G | *glcB* | 1.0 |
| RL0959 | putative FAD binding oxidoreductase | - | 1.1 |
| RL1343 | putative short-chain dehydrogenase/reductase | - | 1.2 |

| | | | |
|---|---|---|---|
| RL1606 | putative esterase (beta-lactamase family) | - | 1.8 |
| RL1694 | putative aldo-keto reductase | - | 2.4 |
| RL2270 | putative nodulin-related transmembrane protein | - | 1.2 |
| RL2323 | putative GFO/IDH/MocA dehydrogenase | - | 1.1 |
| RL3130 | putative phosphatase protein | - | 4.9 |
| RL3345 | putative aldehyde dehydrogenase | - | 2.0 |
| RL3366 | putative flavoprotein | - | 3.6 |
| RL3451 | putative short-chain dehydrogenase/reductase | - | 1.1 |
| RL4186 | putative oxidoreductase | - | 2.0 |
| **Macromolecule metabolism** | | | |
| pRL120625 | putative peptidase | - | 2.3 |
| pRL80075 | putative endoribonuclease L-PSP family protein | - | 2.5 |
| RL1251 | putative serine protease | *degP* | 2.9 |
| RL1295 | putative peptidase | - | 1.2 |
| RL1440 | putative serine protease | *degP* | 2.0 |
| RL3023 | putative polysaccharidase | - | 2.1 |
| **Macromolecule synthesis, modification** | | | |
| RL2628 | putative rRNA methylase family protein | - | 1.1 |
| RL4260 | putative ribonucleotide reductase protein | *nrdI* | 1.0 |
| pRL80059 | putative NifS-like cysteine desulfurase/selenocysteine lyase | - | 1.2 |
| **Metabolism of small molecules** | | | |
| pRL80071 | putative Homoserine dehydrogenase | *hom* | 1.9 |
| RL1260 | putative sulfate adenylyltransferase | *cysN* | 1.0 |
| RL0572 | putative dihydroorotate dehydrogenase | - | 2.3 |
| **Nucleotide biosynthesis** | | | |
| RL1546 | putative amidophosphoribosyltransferase | *purF* | 2.3 |
| **Protection responses** | | | |
| pRL120450 | putative chloroperoxidase | *cpO* | 1.9 |
| pRL80076 | putative aliphatic nitrilase | - | 2.0 |
| RL2769 | putative alkanesulfonate monooxygenase | *ssuD* | 1.7 |
| RL3016 | putative protocatechuate 3,4-dioxygenase beta chain (3,4-pcd) like protein | *pcaH2* | 2.8 |
| **Regulation** | | | |
| pRL110377 | putative Kdp operon transcriptional regulatory protein | *kdpE* | 1.7 |
| **Ribosome constituents** | | | |
| RL1770 | putative 30S ribosomal protein S7 | *rpsG* | 1.1 |
| RL1797 | putative 30S ribosomal protein S11 | *rpsK* | 1.2 |
| **Transport/binding proteins** | | | |
| pRL100265 | putative permease component of ABC transporter | - | 1.8 |

| | | | |
|---|---|---|---|
| pRL110374 | putative substrate-binding component of ABC transporter | - | 1.4 |
| pRL110381 | putative potassium-transporting ATPase a chain | *kdpA* | 1.0 |
| pRL120170 | putative permease component of ABC transporter | - | 2.1 |
| pRL80026 | putative branched-chain amino acid ABC transporter binding component | *livJ* | 2.7 |
| pRL80027 | putative branched-chain amino acid ABC transporter permease component | *livM* | 1.2 |
| pRL80060 | putative exported solute-binding protein | - | 2.5 |
| pRL80063 | putative ATP-binding component of ABC transporter | - | 1.1 |
| pRL80064 | putative permease component of ABC transporter | - | 1.8 |
| pRL90085 | putative periplasmic substrate binding component of ABC transporter | - | 2.4 |
| pRL90091 | putative permease component of ABC transporter | - | 1.2 |
| pRL90245 | putative periplasmic solute-binding component of ABC transporter | - | 1.2 |
| RL0680 | putative transmembrane export SecD/F family protein | *secD* | 3.6 |
| RL0964 | putative cationic transport protein | - | 1.5 |
| RL0996 | putative transmembrane transporter | - | 1.6 |
| RL1682 | putative substrate-binding component of ABC transporter | - | 1.1 |
| RL1721 | putative ATP-binding component of ABC transporter | - | 1.3 |
| RL3350 | putative substrate-binding component of ABC transporter | - | 1.2 |
| RL3424 | putative C4-dicarboxylate transport protein | *dctA* | 1.3 |
| RL3721 | putative permease component of ABC transporter | - | 1.7 |
| RL3805 | putative permease component of ABC transporter protein | - | 1.1 |
| RL3860 | putative permease component of ABC transporter | - | 2.2 |
| RL4244 | putative permease component of ABC transporter | - | 1.8 |
| RL4274 | putative HlyD family transmembrane efflux protein | - | 4.2 |
| **Unknown function** | | | |
| pRL100148 | putative thiamine biosynthesis protein, pseudogene | *thiC* | 1.9 |
| pRL110068 | conserved hypothetical protein | - | 1.3 |
| pRL110199 | conserved hypothetical protein | - | 4.6 |
| pRL110257A | conserved hypothetical protein | - | 1.0 |
| pRL110510 | conserved hypothetical protein | - | 1.1 |
| pRL110624 | conserved hypothetical protein | - | 3.4 |
| pRL120398 | hypothetical exported protein | - | 1.5 |
| pRL120438 | hypothetical protein | - | 1.5 |
| pRL120479 | conserved hypothetical exported protein | - | 3.8 |
| pRL120624 | hypothetical protein | - | 1.2 |
| pRL80020 | conserved hypothetical protein | - | 3.2 |
| pRL80022 | conserved hypothetical protein | - | 2.0 |
| pRL80036 | conserved hypothetical protein | - | 2.2 |

| | | | |
|---|---|---|---|
| pRL80054 | conserved hypothetical protein | - | 2.2 |
| pRL90194 | putative UPF0261 domain protein | - | 1.6 |
| pRL90322 | conserved hypothetical protein, pseudogene | - | 2.3 |
| RL0439 | conserved hypothetical exported protein | - | 1.1 |
| RL0610 | hypothetical exported protein | - | 1.1 |
| RL0786 | conserved hypothetical protein | - | 1.4 |
| RL0913 | putative PRC family protein | - | 1.3 |
| RL1034 | conserved hypothetical protein | - | 1.0 |
| RL1109 | conserved hypothetical protein | - | 1.1 |
| RL1165 | conserved hypothetical protein | - | 1.6 |
| RL1297 | conserved hypothetical protein | - | 4.5 |
| RL1314 | hypothetical exported protein | - | 1.0 |
| RL1631 | conserved hypothetical exported protein | - | 3.7 |
| RL1855 | conserved hypothetical protein | - | 1.8 |
| RL1925 | conserved hypothetical protein | - | 2.0 |
| RL1972 | conserved hypothetical protein | - | 1.3 |
| RL2194 | conserved hypothetical protein | - | 1.9 |
| RL2469 | conserved hypothetical protein, pseudogene | - | 1.2 |
| RL2554 | hypothetical exported protein | - | 3.0 |
| RL2800 | conserved hypothetical protein | - | 2.0 |
| RL3065 | conserved hypothetical exported protein | - | 1.4 |
| RL3098 | putative glycine-asparagine rich protein | - | 1.4 |
| RL3227 | conserved hypothetical protein | - | 1.4 |
| RL3384 | conserved hypothetical exported protein | - | 1.3 |
| RL3489 | conserved hypothetical protein | - | 1.1 |
| RL3497 | putative NUDIX/MutT family protein | - | 1.9 |
| RL3592 | conserved hypothetical exported protein | - | 1.8 |
| RL3982 | conserved hypothetical protein | - | 1.2 |
| RL4562 | conserved hypothetical exported protein | - | 1.3 |
| RL4601 | conserved hypothetical protein | - | 1.1 |

**Table 4.2.6.** Genes up-regulated in pea (7dPea-7dpi) compared with sugar-beet (7dSB-7dpi) rhizosphere examined by direct comparison.

Comparison of these up-regulated genes with the set of 883 genes expressed at least in one time-point in the 7d Pea-137dpi time-course, analysis identified 11 genes, which were found to be induced only in the direct comparison (Table 4.2.7).

| ID | Function | Name | $log_2$ fold |
|---|---|---|---|
| RL1109 | conserved hypothetical protein | - | 1.1 |
| RL1260 | putative sulfate adenylyltransferase | *cysN* | 1.0 |
| RL1499 | putative outer membrane protein | *ropA* | 1.4 |
| RL1682 | putative substrate-binding component of ABC-T | - | 1.1 |
| RL1770 | putative 30S ribosomal protein S7 | *rpsG* | 1.1 |
| RL1797 | putative 30S ribosomal protein S11 | *rpsK* | 1.2 |
| RL2628 | putative rRNA methylase family protein | - | 1.1 |
| RL4260 | putative ribonucleotide reductase protein | *nrdI* | 1.0 |
| pRL100226 | putative acetyltransferase | - | 1.1 |
| pRL110374 | putative substrate-binding component of ABC-T | - | 1.4 |
| pRL110381 | putative potassium-transporting ATPase a chain | *kdpA* | 1.0 |

**Table 4.2.7.** 11 up-regulated genes only found to be specifically induced by using direct comparison of pea (7dPea-7dpi) and sugar-beet (7dSB-7dpi) rhizosphere samples.

Out of the 11 genes induced, 4 genes were involved in nutrient acquisition. These include *pRL110374* coding for an SBP-ABC-T with high identity (80% Id) to SMb21133 of *S. meliloti*, which was shown to be induced by sulphate limitation (Mauchline *et al.*, 2006). Additionally, *RL12*60 (*cysN*) encoding sulphate adenylyltransferase involved in sulfate metabolism was also induced, suggesting that the pea rhizosphere may be a sulfate limited environment. *RL1682* encodes an SBP-ABC-T, classified in the phosphate uptake (PhoT) family, and has high identity (65%) to XC_2707 coding for a putative phosphate binding protein of *Xanthomonas campestris*. It was induced, as was *pRL110381* (*kdpA*) encoding a potassium transporting ATPase chain A. Two genes coding for ribosomal proteins *rpsG* and *rpsK* were specifically induced in pea rhizosphere compared to that of sugar-beet. However, it should be noted that these genes just pass the $log_2$ fold up-regulation filter. This suggests all these genes are on the borderline of detection by any experimental design.

**Table 4.2.8.** List of genes up-regulated in the pea rhizosphere from the direct comparison of cells grown in pea, alfalfa and sugar-beet rhizospheres (Pea - 7dPea-7dpi, Alf - 7dAlf-7dpi, SB - 7dSB-7dpi).

| ID | Function | Name | $log_2$ fold Alf-SB | P value Alf-SB | $log_2$ fold Pea-SB | P value Pea-SB | $log_2$ fold Alf-Pea | P value Alf-Pea |
|---|---|---|---|---|---|---|---|---|
| pRL100148 | putative thiamine biosynthesis protein, pseudogene | *thiC* | 0.4 | 0.28118 | 2.0 | 0 | -1.6 | 0 |
| pRL100444 | putative oxidoreductase | - | 2.2 | 0 | 5.3 | 0 | -3.0 | 0 |
| pRL110199 | conserved hypothetical protein | - | 1.3 | 0.01058 | 4.7 | 0 | -3.4 | 0 |
| pRL110257A | conserved hypothetical protein | - | -0.2 | 0.91221 | 1.2 | 0.00322 | -1.4 | 0.00082 |
| pRL110377 | putative Kdp operon transcriptional regulatory protein | *kdpE* | 0.7 | 0.04925 | 1.8 | 0 | -1.0 | 0.00101 |
| pRL110510 | conserved hypothetical protein | - | 0.0 | 0.99644 | 1.2 | 6.00E-05 | -1.1 | 0.00042 |
| pRL110624 | conserved hypothetical protein | - | 2.3 | 0 | 3.5 | 0 | -1.2 | 0.00839 |
| pRL120027 | putative aldolase | - | 1.6 | 0 | 3.6 | 0 | -2.0 | 0 |
| pRL120170 | putative permease component of ABC transporter | - | 1.0 | 0.00931 | 2.1 | 0 | -1.1 | 0.00071 |
| pRL120398 | hypothetical exported protein | - | 0.5 | 0.25954 | 1.5 | 0 | -1.0 | 0.00061 |
| pRL120479 | conserved hypothetical exported protein | - | 0.8 | 0.00953 | 4.0 | 0 | -3.1 | 0 |
| pRL120480 | putative fimbrial operon related protein | - | 0.2 | 0.79343 | 1.9 | 0 | -1.6 | 0 |
| pRL120624 | hypothetical protein | - | 0.1 | 0.95777 | 1.3 | 0 | -1.2 | 1.00E-05 |
| pRL120625 | putative peptidase | - | 0.0 | 0.99555 | 2.4 | 0 | -2.3 | 0 |
| pRL120632 | putative dehydrogenase | - | 1.3 | 1.00E-05 | 3.0 | 0 | -1.7 | 0 |
| pRL80020 | conserved hypothetical protein | - | 1.2 | 0.00092 | 3.3 | 0 | -2.0 | 0 |
| pRL80021 | putative carbon monoxide dehydrogenase subunit G protein | *coxL* | 0.0 | 0.99644 | 2.8 | 0 | -2.8 | 0 |
| pRL80022 | conserved hypothetical protein | - | 0.1 | 0.98612 | 2.0 | 0 | -1.9 | 0 |
| pRL80023 | putative carbon monoxide dehydrogenase | *coxS* | 0.1 | 0.98612 | 2.5 | 0 | -2.4 | 0 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | subunit | | | | | | | |
| pRL80024 | putative iron-sulphur cluster carbon monooxide dehydrogenase subunit | *coxM* | 0.5 | 0.53004 | 2.8 | 0 | -2.4 | 0 |
| pRL80025 | putative dehydrogenase/reductase | - | 0.0 | 0.99632 | 2.1 | 0 | -2.1 | 0 |
| pRL80026 | putative branched-chain amino acid ABC transporter binding component | *livJ* | 0.1 | 0.9299 | 2.8 | 0 | -2.7 | 0 |
| pRL80027 | putative branched-chain amino acid ABC transporter permease component | *livM* | 0.2 | 0.88276 | 1.2 | 1.00E-05 | -1.0 | 7.00E-04 |
| pRL80030 | putative high-affinity branched-chain amino acid transport ATP-binding protein LivF | *livF* | -0.3 | 0.70863 | 1.0 | 2.00E-05 | -1.3 | 0 |
| pRL80036 | conserved hypothetical protein | - | -0.1 | 0.99061 | 2.3 | 0 | -2.4 | 0 |
| pRL80037 | putative short-chain dehydrogenase | - | 0.1 | 0.98612 | 1.3 | 5.00E-05 | -1.2 | 0.00038 |
| pRL80054 | conserved hypothetical protein | - | 0.3 | 0.76893 | 2.3 | 0 | -2.0 | 0 |
| pRL80059 | putative NifS-like cysteine desulfurase/selenocysteine lyase | - | -0.1 | 0.98947 | 1.3 | 6.00E-05 | -1.3 | 0.00017 |
| pRL80060 | putative exported solute-binding protein | - | 0.3 | 0.93187 | 2.6 | 0 | -2.3 | 2.00E-05 |
| pRL80063 | putative ATP-binding component of ABC transporter | - | 0.0 | 0.99644 | 1.2 | 0.00568 | -1.2 | 0.01188 |
| pRL80064 | putative permease component of ABC transporter | - | -0.1 | 0.98308 | 1.9 | 0 | -2.0 | 0 |
| pRL80071 | putative Homoserine dehydrogenase | *hom* | 0.3 | 0.8727 | 1.9 | 0 | -1.6 | 1.00E-04 |
| pRL90085 | putative periplasmic substrate binding component of ABC transporter | - | 0.3 | 0.51088 | 2.5 | 0 | -2.1 | 0 |
| pRL90322 | conserved hypothetical protein, pseudogene | - | 1.2 | 6.00E-05 | 2.4 | 0 | -1.2 | 3.00E-05 |
| RL0054 | putative malate synthase | *glcB* | 0.0 | 0.98612 | 1.2 | 2.00E-05 | -1.2 | 1.00E-05 |
| RL0439 | conserved hypothetical exported protein | - | -0.1 | 0.98328 | 1.2 | 2.00E-05 | -1.2 | 3.00E-05 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| RL0469 | putative exported surface protein | - | -1.1 | 0.02391 | 1.5 | 4.00E-05 | -2.5 | 0 |
| RL0572 | putative dihydroorotate dehydrogenase | - | 0.0 | 0.99644 | 2.3 | 0 | -2.4 | 0 |
| RL0680 | putative transmembrane export SecD/F family protein | - | 0.9 | 0.22983 | 3.7 | 0 | -2.8 | 0 |
| RL0702 | putative flagella motor switch protein | *fliM* | 0.5 | 0.2903 | 1.6 | 0 | -1.1 | 0.00093 |
| RL0786 | conserved hypothetical protein | - | 0.0 | 0.99644 | 1.5 | 0 | -1.5 | 0 |
| RL0913 | putative PRC family protein | - | 0.3 | 0.8885 | 1.5 | 0.00021 | -1.2 | 0.00379 |
| RL0959 | putative FAD binding oxidoreductase | - | -0.2 | 0.84068 | 1.1 | 6.00E-05 | -1.4 | 2.00E-05 |
| RL0964 | putative cationic transport protein | - | 0.2 | 0.90203 | 1.6 | 0 | -1.4 | 8.00E-05 |
| RL0996 | putative transmembrane transporter | - | 0.5 | 0.48229 | 1.6 | 1.00E-05 | -1.1 | 0.00828 |
| RL1034 | conserved hypothetical protein | - | 0.1 | 0.95091 | 1.1 | 4.00E-05 | -1.0 | 0.00035 |
| RL1173 | putative transmembrane protein | - | 0.8 | 0.10875 | 2.5 | 0 | -1.7 | 1.00E-05 |
| RL1251 | putative serine protease | *degP* | 0.1 | 0.95177 | 3.0 | 0 | -2.9 | 0 |
| RL1295 | putative peptidase | - | -0.4 | 0.38486 | 1.2 | 0 | -1.6 | 0 |
| RL1297 | conserved hypothetical protein | - | 1.0 | 0.36096 | 4.5 | 0 | -3.5 | 0 |
| RL1314 | hypothetical exported protein | - | -1.0 | 0.02096 | 1.1 | 0.00061 | -2.1 | 0 |
| RL1343 | putative short-chain dehydrogenase/reductase | - | 0.1 | 0.9478 | 1.2 | 0 | -1.1 | 0.00017 |
| RL1440 | putative serine protease | *degP* | 0.1 | 0.98328 | 2.1 | 0 | -1.9 | 2.00E-05 |
| RL1546 | putative amidophosphoribosyltransferase | *purF* | 1.0 | 0.02423 | 2.3 | 0 | -1.3 | 0.00038 |
| RL1606 | putative esterase (beta-lactamase family) | - | 0.7 | 0.07758 | 1.9 | 0 | -1.1 | 0.00034 |
| RL1631 | conserved hypothetical exported protein | - | 0.1 | 0.97688 | 3.8 | 0 | -3.7 | 0 |
| RL1925 | conserved hypothetical protein | - | 0.9 | 0.03054 | 2.0 | 0 | -1.1 | 0.00232 |
| RL2194 | conserved hypothetical protein | - | 1.0 | 0.09117 | 2.0 | 0 | -1.0 | 0.0284 |
| RL2554 | hypothetical exported protein | - | -0.7 | 0.1733 | 3.1 | 0 | -3.8 | 0 |
| RL2769 | putative alkanesulfonate monooxygenase | *ssuD* | 0.0 | 0.99896 | 1.7 | 2.00E- | -1.7 | 0.00017 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | 05 | | |
| RL2800 | conserved hypothetical protein | - | 1.0 | 0.00557 | 2.1 | 0 | -1.1 | 0.00028 |
| RL2969 | putative transmembrane protein | - | -0.1 | 0.98612 | 1.3 | 0.01079 | -1.4 | 0.00507 |
| RL3016 | protocatechuate 3,4-dioxygenase beta chain (3,4-pcd) | *pcaH2* | 1.6 | 2.00E-05 | 2.8 | 0 | -1.2 | 0.00083 |
| RL3065 | conserved hypothetical exported protein | - | 0.3 | 0.74463 | 1.5 | 0 | -1.2 | 5.00E-05 |
| RL3098 | putative glycine-asparagine rich protein | - | 0.3 | 0.8015 | 1.3 | 1.00E-05 | -1.1 | 5.00E-04 |
| RL3130 | putative phosphatase protein | - | 3.2 | 0 | 5.0 | 0 | -1.8 | 0.00018 |
| RL3234 | putative lipoprotein | - | -0.2 | 0.77866 | 1.5 | 0 | -1.7 | 0 |
| RL3345 | putative aldehyde dehydrogenase | - | 0.9 | 0.22661 | 2.0 | 0 | -1.1 | 0.03501 |
| RL3366 | putative flavoprotein | - | 1.7 | 0.00101 | 3.6 | 0 | -1.8 | 8.00E-05 |
| RL3384 | conserved hypothetical exported protein | - | 0.0 | 0.98612 | 1.2 | 0 | -1.2 | 0 |
| RL3489 | conserved hypothetical protein | - | -0.5 | 0.28596 | 1.1 | 4.00E-05 | -1.6 | 0 |
| RL3592 | conserved hypothetical exported protein | - | 0.3 | 0.90186 | 1.8 | 1.00E-05 | -1.5 | 0.00026 |
| RL3721 | putative permease component of ABC transporter | - | -0.1 | 0.98612 | 1.8 | 0 | -1.9 | 0 |
| RL3860 | putative permease component of ABC transporter | - | 0.6 | 0.15379 | 2.4 | 0 | -1.7 | 0 |
| RL3982 | conserved hypothetical protein | - | 0.0 | 0.99681 | 1.3 | 0.00526 | -1.3 | 0.00897 |
| RL4186 | putative oxidoreductase | - | 0.2 | 0.92571 | 2.0 | 0 | -1.8 | 0 |
| RL4244 | putative permease component of ABC transporter | - | 0.8 | 0.01721 | 1.8 | 0 | -1.0 | 0.00038 |
| RL4274 | putative HlyD family transmembrane efflux protein | - | 0.3 | 0.83747 | 4.3 | 0 | -3.9 | 0 |
| RL4562 | conserved hypothetical exported protein | - | 0.2 | 0.66744 | 1.3 | 0 | -1.1 | 0 |
| RL460 | conserved hypothetical protein | - | -0.2 | 0.96476 | 1.1 | 0.00788 | -1.3 | 0.00905 |

**Table 4.2.9.** List of genes up-regulated in pea and alfalfa (legume) rhizospheres from the direct comparison of cells grown in pea, alfalfa and sugar-beet rhizospheres (Pea - 7dPea-7dpi, Alf- 7dAlf-7dpi, SB- 7dSB-7dpi).

| ID | Function | Name | $log_2$ fold Alf-SB | P value Alf-SB | $log_2$ fold Pea-SB | P value Pea-SB | $log_2$ fold Alf-Pea | P value Alf-Pea |
|---|---|---|---|---|---|---|---|---|
| pRL100067 | putative racemase/decarboxylase | - | 2.0 | 0 | 2.5 | 0 | -0.5 | 0.10559 |
| pRL100175 | putative nodulation protein | *nodO* | 2.5 | 0 | 2.2 | 0 | 0.2 | 0.89934 |
| pRL100180 | glucosamine--fructose-6-phosphate aminotransferase | *nodM* | 2.7 | 0 | 1. | 2.00E-05 | 0.7 | 0.2133 |
| pRL100183 | nodulation protein F | *nodF* | 2.0 | 0 | 1.4 | 0 | 0.6 | 0.00028 |
| pRL100185 | nodulation protein A | *nodA* | 1.2 | 0.00281 | 1.2 | 0.00075 | 0.0 | 0.99567 |
| pRL100186 | chitooligosaccharide deacetylase | *nodB* | 1.7 | 0 | 1.7 | 0 | -0.02 | 0.99234 |
| pRL100187 | N-acetylglucosaminyltransferase | *nodC* | 3.0 | 0 | 2.7 | 0 | 0.26 | 0.64634 |
| pRL100188 | nod factor export ATP-binding protein I | *nodI* | 2.1 | 0 | 1.8 | 0 | 0.31 | 0.57001 |
| pRL100189 | nodulation protein | *nodJ* | 1.3 | 0 | 1.3 | 0 | 0.07 | 0.9484 |
| pRL100444 | putative oxidoreductase | - | 2.2 | 0 | 5.2 | 0 | -3.0 | 0 |
| pRL110199 | conserved hypothetical protein | - | 1.3 | 0.01058 | 4.7 | 0 | -3.4 | 0 |
| pRL110281 | putative substrate-binding component of ABC transporter | - | 1.4 | 0 | 1.0 | 4.00E-05 | 0.4 | 0.22203 |
| pRL110624 | conserved hypothetical protein | - | 2.2 | 0 | 3.4 | 0 | -1.1 | 0.00839 |
| pRL120027 | putative aldolase | - | 1.6 | 0 | 3.5 | 0 | -1.9 | 0 |
| pRL120438 | hypothetical protein | - | 2.4 | 0 | 1.7 | 0 | 0.6 | 0.13579 |
| pRL120632 | putative dehydrogenase | - | 1.2 | 1.00E-05 | 2.9 | 0 | -1.6 | 0 |
| pRL80020 | conserved hypothetical protein | - | 1.2 | 0.00092 | 3.2 | 0 | -2.0 | 0 |
| pRL80075 | putative endoribonuclease L-PSP family protein | - | 2.1 | 0.01174 | 2.4 | 0.00102 | -0.2 | 0.95444 |
| pRL90322 | conserved hypothetical protein, pseudogene | - | 1.2 | 6.00E-05 | 2.4 | 0 | -1.2 | 3.00E-05 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| RL1165 | conserved hypothetical protein | - | 1.1 | 0.00068 | 1.6 | 0 | -0.4 | 0.27457 |
| RL1499 | putative outer membrane protein | *ropA* | 2.2 | 0 | 1.3 | 0 | 0.8 | 0.0048 |
| RL1694 | putative aldo-keto reductase | - | 1.8 | 0 | 2.4 | 0 | -0.6 | 0.13869 |
| RL1770 | putative 30S ribosomal protein S7 | *rpsG* | 1.0 | 0.01449 | 1.0 | 0.00126 | -0.06 | 0.98672 |
| RL1855 | conserved hypothetical protein | - | 1.1 | 0.00104 | 1.8 | 0 | -0.6 | 0.11802 |
| RL1972 | conserved hypothetical protein | - | 2.0 | 0 | 1.3 | 0 | 0.7 | 0.00303 |
| RL3016 | protocatechuate 3,4-dioxygenase beta chain (3,4-pcd) | *pcaH2* | 1.6 | 2.00E-05 | 2.8 | 0 | -1.1 | 0.00083 |
| RL3023 | putative polysaccharidase | - | 1.2 | 0.02472 | 2.0 | 0 | -0.8 | 0.1195 |
| RL3130 | putative phosphatase protein | - | 3.1 | 0 | 4.9 | 0 | -1.8 | 0.00018 |
| RL3366 | putative flavoprotein | - | 1.7 | 0.00101 | 3.5 | 0 | -1.8 | 8.00E-05 |
| RL3497 | putative NUDIX/MutT family protein | - | 1.2 | 5.00E-05 | 1.9 | 0 | -0.74 | 0.00955 |

An integrated comparison was performed with all the microarray data obtained from pea and comparative rhizosphere transcriptomics. The basic idea of this comparison is to identify the genes specifically expressed during colonization of the pea rhizosphere by comparing all the data. Although a set of pea-specific genes expressed only in pea rhizosphere were identified by comparative transcriptomics (both direct and indirect comparison), we compared the pea-specific set of genes and the results obtained for other experiments in the pea rhizosphere, both length of time post-inoculation and age of plants.

The comparison of the genes up-regulated in at least one of the conditions in the days post-inoculation (dpi) time-course and differently aged (age) rhizosphere experiments, showed 648 common genes. Likewise, the comparison of pea-specific genes from indirect and direct comparison of the three rhizospheres showed 52 common genes. Finally, comparison of both identified the common genes, a set of 47 genes induced specifically in pea rhizosphere. The schematic representation of the integrated comparison is shown in the Figure 4.3.1. The 47 pea-specific genes, categorized based on their biological functions are shown in the Table 4.3.1. All genes in the pea-specific gene list were discussed earlier in Chapter 3.

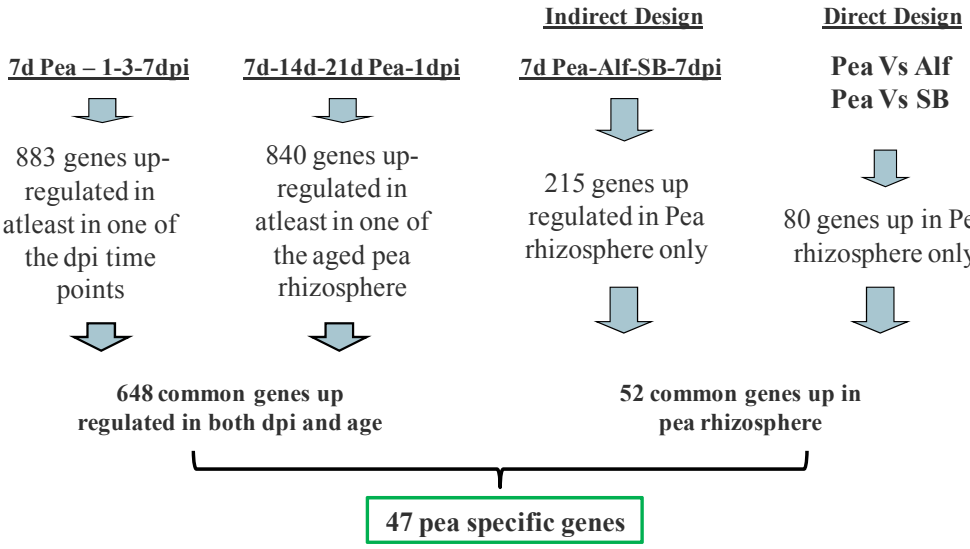



**Figure 4.3.1.** Schematic representation of the integrated comparison of the up-regulated genes from all the microarray expression datasets.

| ID | Function | Name |
|---|---|---|
| **Cell envelope** | | |
| pRL120480 | putative fimbrial operon related protein | - |
| **Central intermediary metabolism** | | |
| pRL80037 | putative short-chain dehydrogenase | - |
| RL0054 | malate synthase G | *glcB* |
| RL1343 | putative short-chain dehydrogenase/reductase | - |
| RL1606 | putative esterase (beta-lactamase family) | - |
| RL3345 | putative aldehyde dehydrogenase | - |
| RL4186 | putative oxidoreductase | - |
| **Energy metabolism, Carbon** | | |
| pRL80021 | putative carbon monoxide dehydrogenase subunit G protein | *coxG* |
| pRL80023 | putative carbon monoxide dehydrogenase subunit | *coxM* |
| pRL80024 | putative iron-sulphur cluster carbon monooxide dehydrogenase subunit | *coxS* |
| pRL80025 | putative dehydrogenase/reductase | *coxL* |
| RL3366 | putative flavoprotein | - |
| **Macromolecule metabolism** | | |
| pRL120625 | putative peptidase | - |
| RL1251 | putative serine protease | *degP* |
| RL1295 | putative peptidase | - |
| RL1440 | putative serine protease | *degP* |
| **Metabolism of small molecules** | | |
| pRL80059 | putative NifS-like cysteine desulfurase/selenocysteine lyase | - |
| pRL80071 | putative Homoserine dehydrogenase | *hom* |
| **Nucleotide biosynthesis** | | |
| RL0572 | putative dihydroorotate dehydrogenase | - |
| **Regulation** | | |
| pRL110377 | putative Kdp operon transcriptional regulatory protein | *kdpE* |
| **Transport/binding proteins** | | |
| pRL120170 | putative permease component of ABC transporter | - |
| pRL80026 | putative branched-chain amino acid ABC transporter binding component | *livJ* |
| pRL80027 | putative branched-chain amino acid ABC transporter permease component | *livM* |
| pRL80030 | putative high-affinity branched-chain amino acid transport ATP-binding protein | *livF* |
| pRL80060 | putative exported solute-binding protein | - |
| pRL80063 | putative ATP-binding component of ABC transporter | - |
| pRL80064 | putative permease component of ABC transporter | - |
| pRL90085 | putative periplasmic substrate binding component of ABC transporter | - |
| RL0680 | putative transmembrane export SecD/F family protein | *secDF* |

| | | |
|---|---|---|
| RL0996 | putative transmembrane transporter | - |
| RL3721 | putative permease component of ABC transporter | - |
| RL3860 | putative permease component of ABC transporter | - |
| RL4244 | putative permease component of ABC transporter | - |
| **Foreign DNA** | | |
| pRL100148 | putative thiamine biosynthesis protein, pseudogene | *thiC* |
| **Unknown function** | | |
| pRL110199 | conserved hypothetical protein | - |
| pRL110510 | conserved hypothetical protein | - |
| pRL120479 | conserved hypothetical exported protein | - |
| pRL120624 | hypothetical protein | - |
| pRL80022 | conserved hypothetical protein | - |
| pRL80036 | conserved hypothetical protein | - |
| pRL80054 | conserved hypothetical protein | - |
| RL0439 | conserved hypothetical exported protein | - |
| RL0786 | conserved hypothetical protein | - |
| RL1297 | conserved hypothetical protein | - |
| RL2800 | conserved hypothetical protein | - |
| RL3592 | conserved hypothetical exported protein | - |
| RL4601 | conserved hypothetical protein | - |

**Table 4.3.1.** 47 pea-specific genes identified from the integrated comparison of all the rhizosphere microarrays.

This chapter reports the comparative rhizosphere transcriptomics of Rlv3841 in the pea, alfalfa or sugar-beet rhizospheres. Earlier the success of this strategy to use transcriptomics to examine rhizosphere bacteria was validated by qRT-PCR (see Chapter 3). In this chapter, this research was further extended to identify the pea-specific genes in Rlv3841, i.e. those specifically induced in the pea rhizosphere.

Many pea-specific genes were identified. They include transport and catabolic genes involved in tartrate metabolism and genes coding for carbon monoxide dehydrogenase (CODH), a key enzyme in carbon monoxide utilization, Also, identified were a high-affinity branched amino acid ABC-T operon (*livJMHGF*) and an ABC-T with high homology to mimosine transporter in *Rhizobium* sp. TAL1455. The high induction of homoserine dehydrogenase fits with earlier reports that homoserine in abundant in pea root exudates and can be utilized as a carbon source. Also, the induction of malate synthase G, a key enzyme in glyoxylate cycle suggests that operation of the glyoxylate cycle in Rlv3841 in the pea rhizosphere. Interestingly, most of the genes specifically induced in the pea rhizosphere are located on a plasmid (pRL8).

Another, specific set of genes were induced in Rlv3841 when grown in alfalfa rhizosphere. These include malonate uptake and catabolic genes. Also, induced were genes involved in the conversion of ornithine to proline. It is reported that the alfalfa rhizosphere is rich in proline and the ability to utilize proline would offer a selective advantage during colonization. Another, observation is the repression of *nod* genes in the sugar-beet rhizosphere, exactly as might be expected. These results give weight to the quality of the data obtained from these microarray analyses and confirm the success of the strategy developed.

Integrated analysis of all the transcriptomic data, showed 47 genes to be specifically induced in the pea rhizosphere. Although many host-specific genes have been identified, the biological functions of most of these genes are not clear. In order to assess biological importance of the up-regulated genes in the pea

rhizosphere, rhizosphere competition studies were performed. The next chapter describes those experiments and the results obtained.

# Chapter 5

# Determination of the biological significance of the genes up-regulated during pea rhizosphere colonization.

## 5.1 Introduction

Plant-bacterial interactions are very complex with the need for timely expression of many genes. However, very little is known about which genes contribute to a successful plant-microbe interaction. Successful colonization of the rhizosphere by rhizobia is a key step in establishing a symbiotic interaction with the host plant. The rhizosphere is complex environment and adaptation to it by rhizobia requires a very specific and orchestrated gene expression in tune with the environment. From our rhizosphere microarray experiments, it is very clear that Rlv3841 has a set of specific genes which are expressed in response to the available nutrients and host-derived signals. The rhizobia adapt to survive and this is apparent even in non-host rhizospheres. Research in the last few decades has identified a number of molecular determinants that, directly or indirectly, contribute to rhizosphere colonization. These participate in a wide range of biological functions including motility, nutrient acquisition, utilization of complex sugars, aromatic compounds and aminoacids, cell maintenance, defence, detoxification and in determining host specificity (Lugtenberg *et al.*, 2001). A competitive advantage can be gained by rhizobia in the rhizosphere by the ability to catabolize sugars present in the root exudates, such as *myo*-inositol (Fry *et al.*, 2001), rhamnose (Oresnik *et al.*, 1998) and erythritol (Yost *et al.*, 2006). Additionally, the ability to detoxify or catabolize certain plant-derived compounds like trigonelline (Boivin *et al.*, 1991), calystegines (Guntli *et al.*, 1999), rhizopines (Heinrich *et al.*, 1999) and aminoacids like proline (Jiménez-Zurdo *et al.*, 1995) and mimosine (Soedarjo & Borthakur, 1998), has been reported to provide a competitive advantage in certain rhizobia during rhizosphere colonization. Recent research from our group showed that a mutant in a gene involved in thiamine biosynthesis, *thiM*, was highly impaired in rhizosphere competition compared with wild-type, (Karunakaran *et al.*, 2006). A thorough review of rhizosphere competitiveness can be found in section 1.3.5.

The main objective of this section of this project is to study the biological significance of the genes up-regulated in the pea rhizosphere, and especially their

role in competitiveness during pea rhizosphere colonization. The most common method used to examine the biological role of a gene, is by disrupting the target gene (creating a mutant) and assessing the competitive fitness of the mutant strain by challenging with the wild-type in the host rhizosphere (*in vitro*). This is referred to as a competitive colonization assay. In a competitive colonization assay, a mutant strain with the gene of interest disrupted, containing a selectable marker (i.e. different from the parental strain) is challenged against the parental strain in equal density (1:1) or in different ratios in the host rhizosphere. After a stipulated time period, the cells are recovered from the host rhizosphere, plated onto agar plated, with appropriate selective agents and the resulting colonies counted. The selective markers can either be a reporter gene or an antibiotic resistance gene. The mutants can be generated in a number of ways, e.g. by transposons mutagenesis, site directed mutagenesis or targeted plasmid integration mutagenesis. In this study, we used the pK19mob plasmid, a broad host range mobilizable vector. It encodes *lacZ* alpha fragment, a kanamycin resistant marker, has multiple cloning site to allow easy cloning of fragments in order to make integration mutants and the bonus of pK19- and M13- specific sequencing primers binding sites (Schafer *et al.*, 1994).

### 5.2.1 Construction of pK19-based integration mutants

In order to assess the biological significance of the up-regulated genes of Rlv3841 during pea rhizosphere colonization, a panel (46) of mutants were made. The selection of genes to be mutated was based on; (i) genes highly induced in all three (pea, alfalfa and sugar beet) rhizospheres, (ii) genes specifically induced in the pea rhizosphere and (iii) highly induced transporters, spanning 12 different functional categories. In addition, *pRL110443* (*thiM*) involved in thiamine biosynthesis in Rlv3841, reported to provide competitive advantage during pea rhizosphere colonization (Karunakaran *et al.*, 2006) was used as a positive control and *pRL100162* (*nifH*), which is not induced in the pea rhizosphere, was used as a negative control. A total of 48 pK19 integration mutants were made in Rlv300 with the general strategy described in section 2.2.7. After screening, putative pK19 integration mutants were checked by PCR using primers specific to a site approximately 500 bp upstream of the forward primer binding site of the amplicon and pK19A/B primers as described in section 2.2.5. The confirmed trans-conjugants (integrated mutant derivatives of Rlv300) with the resistant markers $neo^r$, $trim^r$ and $nys^r$ and Rlv3841 (wt) with $str^r$, $trim^r$ and $nys^r$ were used for competition studies.

Rlv300 ($str^s$, $trim^r$ and $nys^r$), is the parent strain of Rlv3841 ($str^r$, $trim^r$ and $nys^r$) was used for pK19 integration mutagenesis to take advantage of antibiotic selection during screening of rhizosphere-recovered cells in competition assays. The pK19mob plasmid carries a neomycin resistant marker, which confers neomycin resistance to the trans-conjugants. All the mutants had growth indistinguishable from wild-type under laboratory conditions. In order to check the ability of the mutants *in planta* to nodulate and fix nitrogen compared with the wild-type, each mutant was inoculated on pea seeds, in a pot containing vermiculite and nitrogen-free rooting solution with wild-type 3841 as the control, and grown for 28 days in a growth chamber as described in section 2.3.4. After 28 days, the plants and the nodules were examined visually. It was found that all the

mutants formed red nodules indicating their ability to nodulate and form nitrogen-fixing nodules. All plants were green confirming nitrogen fixation compared to uninoculated controls which were yellow.

| Mutant Strain | Plasmid number | ID | Name | Function | Expressed in host rhizosphere* |
|---|---|---|---|---|---|
| **Protection responses** | | | | | |
| RU4233 | pRU2034 | RL3016 | *pcaH2* | protocatechuate 3,4-dioxygenase beta chain (3,4-pcd) | **P,A,SB** |
| **Transport/binding proteins** | | | | | |
| RU4298 | pRU2038 | RL0680 | *secDF* | putative transmembrane export SecD/F family protein | **P** |
| RU4263 | pRU2120 | RL0996 | | putative transmembrane transporter | **P** |
| RU4268 | pRU2110 | pRL110281 | | putative substrate-binding component of ABC transporter | **P,A** |
| RU4265 | pRU2125 | RL2418 | | putative transcriptional regulator | **P,A,SB** |
| RU4258 | pRU2111 | RL3424 | *dctA* | putative C4-dicarboxylate transport protein | **P,A,SB** |
| RU4260 | pRU2115 | RL4274 | | putative HlyD family transmembrane efflux protein | **P,A,SB** |
| RU4261 | pRU2117 | RL3860 | | putative permease component of ABC transporter | **P,A** |
| RU4248 | pRU2128 | pRL80026 | *livJ* | putative branched-chain amino acid ABC transporter binding component | **P** |
| RU4256 | pRU2107 | pRL90085 | | putative periplasmic substrate binding component of ABC transporter | **P,A** |
| RU4295 | pRU2116 | pRL120500 | | putative substrate-binding component of transporter | **P,A,SB** |
| **Macromolecule metabolism** | | | | | |
| RU4254 | pRU2102 | RL1251 | *degP* | putative serine protease | **P** |
| **Macromolecule synthesis, modification** | | | | | |
| RU4252 | pRU2131 | RL4265 | *msrB* | putative peptide methionine sulfoxide reductase | **P,SB** |
| **Biosynthesis of cofactors, carriers** | | | | | |
| RU4358 | pRU2187 | pRL110443 | *thiM* | putative hydroxyethylthiazole kinase | **P,A,SB** |
| RU4249 | pRU2100 | RL2711 | *moaA* | putative molybdenum cofactor biosynthesis protein A | **P,A,SB** |
| **Central intermediary metabolism** | | | | | |
| RU4360 | pRU2056 | pRL100162 | *nifH* | putative nitrogenase iron protein | **-** |
| RU4232 | pRU2033 | RL4267 | *acoD* | putative acetaldehyde dehydrogenase | **P,A,SB** |
| RU4247 | pRU2105 | pRL120632 | | putative dehydrogenase | **P,A** |
| RU4235 | pRU2044 | RL1694 | | putative aldo-keto reductase | **P,A,SB** |
| RU4274 | pRU2047 | RL0037 | *pckA* | putative phosphoenolpyruvate carboxykinase | **P,A,SB** |
| RU4230 | pRU2024 | RL3130 | | putative phosphatase protein | **P,A** |
| RU4311 | pRU2053 | pRL100444 | | putative oxidoreductase | **P,A** |
| RU4222 | pRU2082 | RL1911 | | putative arylsulfatase | **P,A,SB** |
| **Energy metabolism, carbon** | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| RU4231 | pRU2025 | RL3366 | | putative flavoprotein | **P** |
| RU4271 | pRU2114 | pRL80021 | *coxG* | putative carbon monoxide dehydrogenase subunit G protein | **P** |
| RU4250 | pRU2101 | pRL80023 | *coxM* | putative carbon monoxide dehydrogenase subunit | **P** |
| **Degradation of small molecules** | | | | | |
| RU4259 | pRU2113 | RL1860 | *phhA* | putative phenylalanine-4-hydroxylase | **P,A,SB** |
| RU4269 | pRU2112 | RL1863 | | putative 4-hydroxyphenylpyruvate dioxygenase | **P,A,SB** |
| **Cell envelope** | | | | | |
| RU4312 | pRU2168 | pRL120724 | | putative transmembrane protein | **P,A,SB** |
| RU4318 | pRU2170 | RL3186 | | putative transmembrane protein | **P,A,SB** |
| RU4308 | pRU2166 | RL1172 | | putative transmembrane protein | **P,A,SB** |
| **Regulation** | | | | | |
| RU4317 | pRU2129 | pRL110423 | | putative TetR family transcriptional regulator | **P,A,SB** |
| RU4310 | pRU2167 | RL2946 | | putative two component response regulator transcriptional regulatory protein | **P,A,SB** |
| RU4272 | pRU2121 | RL0540 | | putative two-component sensor histidine kinase transcriptional regulatory protein | **P,A,SB** |
| **Unknown function** | | | | | |
| RU4297 | pRU2127 | RL3982 | | conserved hypothetical protein | **P,A,SB** |
| RU4234 | pRU2042 | pRL110199 | | conserved hypothetical protein | **P** |
| RU4257 | pRU2108 | RL2259 | | conserved hypothetical exported protein | **P,A,SB** |
| RU4309 | pRU2124 | pRL90055 | | conserved hypothetical protein | **P,A,SB** |
| RU4255 | pRU2103 | RL3272 | | conserved hypothetical protein | **P,A,SB** |
| RU4266 | pRU2126 | pRL120479 | | conserved hypothetical exported protein | **P,A** |
| RU4273 | pRU2123 | RL0274 | | conserved hypothetical protein | **P,A,SB** |
| RU4267 | pRU2109 | RL1485 | | conserved hypothetical protein | **P,A,SB** |
| RU4262 | pRU2118 | RL1297 | | conserved hypothetical protein | **P** |
| RU4270 | pRU2133 | RL0787 | | conserved hypothetical protein | **P,A,SB** |
| RU4253 | pRU2132 | pRL80054 | | conserved hypothetical protein | **P** |
| RU4296 | pRU2119 | pRL110268 | | conserved hypothetical protein | **P,A,SB** |
| RU4229 | pRU2021 | RL0913 | | putative PRC family protein | **P,A,SB** |
| RU4251 | pRU2130 | RL2469 | | conserved hypothetical protein, pseudogene | **P,A,SB** |

**Table 5.1.1.** Integration mutants, showing corresponding plasmid name, and gene information, categorized by functional category. Positive and negative controls are highlighted in grey shade.
* gene expressed in host rhizosphere: P - pea, A- alfalfa and SB - sugar-beet rhizospheres.

### 5.2.2 Competitive colonization assay

To ensure Rlv300 (strep$^-$) and Rlv3841 (strep$^+$) have equal competence in colonizing the pea rhizosphere, a trial competitive colonization assay was performed. Five different combinations (0:1000, 1000:0, 1000:1000, 1000:10000, 10000:1000 CFU) of both the strains were inoculated onto a 7d old pea rhizosphere, in triplicate, and the cells were recovered after 7 days as described in the section 2.3.6. Each replicate of the rhizosphere-recovered cells were serial diluted and plated on TY medium containing trimethioprim and nystatin. After three days incubation at 26ºC, 100 CFU were patched on TY medium containing trimethioprim, nystatin with and without streptomycin. In the TY medium containing trimethioprim, nystatin all the patched colonies will grow, whereas in TY medium containing trimethioprim, nystatin and streptomycin only Rlv3841 will be able to grow. The patched plates were examined for the growth of strep$^+$ colonies and counted. The results showed that both the strains have equal competence in colonizing the pea rhizosphere without any bias (Figure 5.1.1).

For the competitive colonization assay (CCA) between Rlv3841 and mutant strains, an inoculation ratio of 10000:1000 CFU was inoculated into the rhizosphere of 7d old pea plants. This ratio was selected based on previous results, in which inoculation of a 10-fold excess of *thiM* mutant over the wild-type accounted only for 18% of recovered bacteria (Karunakaran *et al.*, 2006). Moreover, assaying 8 replicates of 48 mutants in 5 combinations is time-consuming and would take time beyond the permitted duration of this project. A preliminary screen was performed with this ratio and was extended with different combinations when mutants, showing highly impaired phenotype in rhizosphere competitiveness, were identified. To ensure that this competitive assay is statistically significant a total of 8 replicates assay were performed for each mutant including positive and negative controls in two batches (3 and 5 replicates). For the ease of understanding, assays between Rlv3841 and a mutant will be referred as a mutant assay and Rlv300 versus Rlv3841 as a control assay. An overall schematic representation of the design of the CCA is shown in Figure 5.1.2.

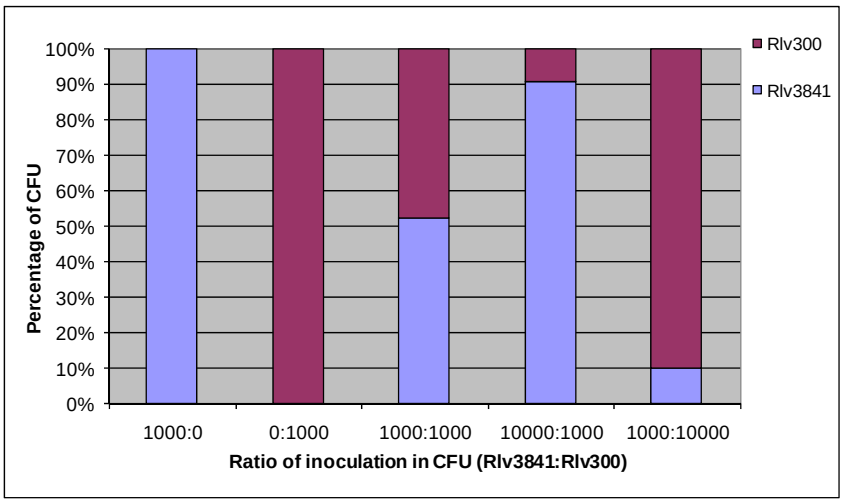



**Figure 5.1.1.** Graph showing the competition of strains Rlv3841 and Rlv300 in different combinations in a pea rhizosphere. The inoculation ratio is given on the x-axis and the average of the percentage of recovered cells from 3 plants on the y-axis.

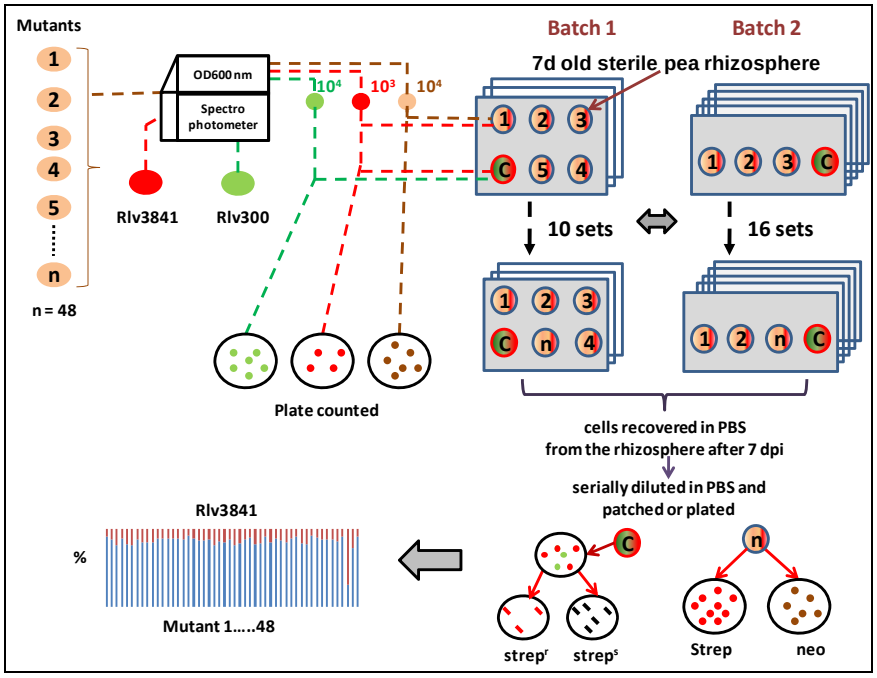



**Figure 5.1.2.** Schematic representation of competitive colonization assay. A total of 48 mutants ($10^4$ CFU) including one positive (*thiM*) and one negative (*nifH*) control were challenged against Rlv3841 ($10^3$ CFU) with an experimental control (Rlv300 ($10^4$ CFU) versus Rlv3841 ($10^3$ CFU)). The cells recovered from the assay were plate counted on plates containing appropriate antibiotics. The experimental control was plated on TY medium without antibiotics and subsequently patched on TY medium, with and without streptomycin. C- control assay, 1,2…n – mutant assay.

In the first batch, 10 sets of experiments were performed, with 5 mutant assays and 1 control assay per set, i.e. each mutant and Rlv300 were co-inoculated with Rlv3841 in the ratio (10000:1000 CFU) onto a 7d old pea rhizosphere. In the second batch, a total of 16 sets were performed in five replicates with 3 mutant assays and 1 control assay per set. To ensure equal dilution, all the strains inoculated in each set were plate counted on TY plates with appropriate antibiotics. The plate count results showed a variation within the acceptable limit, of less than 2%. The data obtained from 8 replicates were averaged and subjected to statistical analysis. The results from the control assay in each set ensured that cell dilutions did not introduce any bias. The set of assays were repeated, when a control assay showed a large difference between the observed and expected ratio in the patch count.

### 5.2.3 Data analysis of competitive colonization assay

Data analysis of all the control assay replicates (78) showed only minimum variation. In theory, a 10-fold excess of inoculation of Rlv300 over Rlv3841 will result in 91% Rlv300 and 9% Rlv3841 in the recovered bacteria. In the control assay, Rlv3841 contributed 9.75% ± 0.37% (means, SEM) with a SD of 2.64% of the total recovered bacterial cells from the pea rhizosphere. The controls in the mutant assays *thiM* (positive) and *nifH* (negative) were recovered from the pea rhizosphere as expected. Only 28% of the *thiM* mutant was recovered from the pea rhizosphere, even after inoculating at a 10-fold excess over Rlv3841, which is in line with the earlier results (Karunakaran *et al.*, 2006). In case of the negative control, 9.73% of the *nifH* mutant was recovered from the pea rhizosphere even after 10-fold excess inoculation, which is not a significant difference from the expected value of 9%. This shows the integrity of the experiment performed and strengthens the reliability of the mutant assays.

In the assay, mutants were considered to be competitively defective when Rlv3841 recovered from the pea rhizosphere accounted for more than 13% of the total viable cells recovered. This cut-off was based on results obtained from the control assays where Rlv3841 accounted for 9.75% ± 2.64% (means, SD) of the

total recovered cells from the pea rhizosphere (t test; $P < 0.05$). Of the 46 mutants assayed, 38 mutants showed reduced competitiveness in colonizing the pea rhizosphere compared to control wild-type strain. To keep this analysis more stringent, the cut-off value was increased to >14%, which excluded ten mutants.

A rhizosphere colonization index (RCI) was calculated, for cells recovered from the rhizosphere after a stipulated time period.

$$\text{RCI} = \frac{\text{the percentage of recovered wild-type from the mutant assay}}{\text{the percentage of recovered wild-type from the control assay}}$$

For example, in a mutant assay where 9.5% wild-type was recovered from the rhizosphere in the mutant assay and 9.4% of wildtype recovered from control assay, the RCI will be 9.5%/9.4% = 1, which shows there is no competitive advantage.

In contrast, an RCI >10 would show that the mutant is severely attenuated and has been completely lost in the rhizosphere. The positive control *thiM* mutant had RCI = 7.33 i.e. (66%/9.4%), showing that the strain is strongly attenuated and severely at a disadvantage in competition with the wild-type strain. Theoretically, RCI cannot go below 1, unless a mutation confers a competitive advantage, although a value slightly below 1 may be due to experimental error.

Based on the percentage of the recovered wild-type (Rlv3841) with respect to the mutants, the competition data were classified as shown in the Table 5.1.2.

| **% Rlv3841 recovered** | **No of mutants** |
|---|---|
| > 20% and < 24% | 5 |
| >17% and < 20% | 12 |
| >14% and < 17% | 11 |
| < 14% | 10 |

**Table 5.1.2.** Mutants were grouped based on the percentage of Rlv3841 recovered from the total viable cells from the pea rhizosphere in the mutant assay.

Competition data for all the mutants, including RCI, grouped according to their competitiveness and classified on the basis of functional category of the mutated gene are shown in the Table 5.1.3.

| Mutant strain | ID | Name | Function | Mutant (%) | Rlv3841 (%) | RCI | 7dPea-7dpi induction $log_2$ fold |
|---|---|---|---|---|---|---|---|
| **> 20% and < 24% Rlv3841 recovered in assay (i.e. least competitive mutants)** | | | | | | | |
| RU4232 | RL4267 | *acoD* | putative acetaldehyde dehydrogenase | 76.29 | 23.71 | 2.43 | 4.55 |
| RU4259 | RL1860 | *phhA* | putative phenylalanine-4-hydroxylase | 77.36 | 22.64 | 2.32 | 4.58 |
| RU4229 | RL0913 | | putative PRC family protein | 77.78 | 22.22 | 2.28 | 5.59 |
| RU4298 | RL0680 | *secD* | putative transmembrane export SecD/F family protein | 78.82 | 21.18 | 2.17 | 6.55 |
| RU4263 | RL0996 | | putative transmembrane transporter | 79.75 | 20.25 | 2.08 | 2.2 |
| **>17% and < 20% Rlv3841 recovered in assay** | | | | | | | |
| RU4297 | RL3982 | | conserved hypothetical protein | 80.05 | 19.95 | 2.05 | 4.76 |
| RU4231 | RL3366 | | putative flavoprotein | 80.62 | 19.38 | 1.99 | 4.84 |
| RU4268 | pRL110281 | | putative substrate-binding component of ABC transporter | 80.68 | 19.32 | 1.98 | 3.3 |
| RU4269 | RL1863 | | putative 4-hydroxyphenylpyruvate dioxygenase | 81.10 | 18.90 | 1.94 | 1.15 |
| RU4247 | pRL120632 | | putative dehydrogenase | 81.28 | 18.72 | 1.92 | 5.1 |
| RU4265 | RL2418 | | putative substrate-binding component of ABC transporter | 81.28 | 18.72 | 1.92 | 1.82 |
| RU4235 | RL1694 | | putative aldo-keto reductase | 81.49 | 18.51 | 1.90 | 4.94 |
| RU4274 | RL0037 | *pckA* | putative phosphoenolpyruvate carboxykinase | 82.13 | 17.87 | 1.83 | 2.8 |
| RU4234 | pRL110199 | | conserved hypothetical protein | 82.25 | 17.75 | 1.82 | 7.6 |
| RU4258 | RL3424 | *dctA* | putative C4-dicarboxylate transport protein | 82.43 | 17.58 | 1.80 | 4.8 |
| RU4271 | pRL80021 | *coxG* | putative carbon monoxide dehydrogenase subunit G protein | 82.87 | 17.13 | 1.76 | 3.9 |
| RU4260 | RL4274 | | putative HlyD family transmembrane efflux protein | 82.83 | 17.17 | 1.76 | 7.0 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **>14% and < 17% Rlv3841 recovered in assay** | | | | | | | |
| RU4261 | RL3860 | | putative permease component of ABC transporter | 83.41 | 16.59 | 1.70 | 4.5 |
| RU4230 | RL3130 | | putative phosphatase protein | 83.67 | 16.33 | 1.67 | 6.2 |
| RU4257 | RL2259 | | conserved hypothetical exported protein | 83.77 | 16.23 | 1.66 | 2.7 |
| RU4248 | pRL80026 | *livJ* | putative branched-chain amino acid ABC transporter binding component | 84.54 | 15.46 | 1.58 | 4.0 |
| RU4256 | pRL90085 | | putative periplasmic substrate binding component of ABC transporter | 84.82 | 15.18 | 1.56 | 4.5 |
| RU4309 | pRL90055 | | conserved hypothetical protein | 84.85 | 15.15 | 1.55 | 1.98 |
| RU4295 | pRL120500 | | putative substrate-binding component of transporter | 85.18 | 14.82 | 1.52 | 1.6 |
| RU4255 | RL3272 | | conserved hypothetical protein | 85.57 | 14.43 | 1.48 | 2.0 |
| RU4266 | pRL120479 | | conserved hypothetical exported protein | 85.88 | 14.12 | 1.45 | 5.6 |
| RU4311 | pRL100444 | | putative oxidoreductase | 85.96 | 14.04 | 1.44 | 7.0 |
| RU4233 | RL3016 | *pcaH2* | protocatechuate 3,4-dioxygenase beta chain (3,4-pcd) | 85.95 | 14.05 | 1.44 | 7.5 |
| **Positive and negative Controls** | | | | | | | |
| RU4358 | pRL110443 | *thiM* | putative hydroxyethylthiazole kinase | 28.59 | 71.41 | 7.33 | |
| RU4360 | pRL100162 | *nifH* | putative nitrogenase iron protein | 90.27 | 9.73 | 1.00 | |

**Table 5.1.3.** Results of the competitive colonization assay for mutant strains classified in groups according to their competitiveness against Rlv3841. Positive and negative controls are highlighted in grey.

In the first group of strongly attenuated mutants, there are five strains which show strongly altered competition compared to wild-type during colonization of pea rhizosphere. The possible role of the protein coded by the gene mutated in each strain is discussed below. For the ease of understanding, the each gene ID is followed by gene annotation and RCI in brackets, further information about each mutant can be seen in the Table 5.1.2.

RU4298, mutated in *RL0680* (*secDF*) (transmembrane export protein, RCI = 2.17), was very highly expressed in the pea rhizosphere [6.8U], but not induced in alfalfa and sugarbeet rhizospheres (Krehenbrink & Downie, 2008). Amino acid sequence analysis of *RL0680* showed moderate identity (41%) with RL2055 (annotated as SecD, but having both SecD and F domains fused into a single polypeptide as found in *Bacillus subtilis* (Bolhuis et al., 1998). In general, SecD and SecF are regulatory subunits which regulate the translocation of proteins in the cell. SecDF may be involved in Sec the export machinery, exporting a sub-set of proteins in response to specific condition. Taken together with the competition data, it suggests that the SecDF may export specific proteins in response to the pea rhizosphere, which provide a competitive advantage.

RU4263, mutated in *RL0996* (putative transmembrane protein, RCI = 2.08), was expressed only in the pea rhizosphere [2.3U]. RL0996 has high identity to the tartrate transporter (TtuB) of *A. vitis* (Crouzet & Otten, 1995). In the pea rhizosphere, the catabolic genes involved in tartrate utilization were also up-regulated along with the tartrate transporter, suggesting the active utilization of tartrate in the pea rhizosphere. The compromised colonization levels show that the ability to utilize tartrate offers a competitive advantage to Rlv3841 in the pea rhizosphere. In *A. vitis*, competition experiments with a tartrate-utilizing strain and its non-utilizing derivative showed that the ability to utilize tartrate offers selective advantage on grapevine (Salomone *et al.*, 1998)

RU4232, mutated in *RL4267* (*acoD*) (acetaldehyde dehydrogenase, RCI = 2.43), was highly expressed in all the three rhizospheres [4.7U,5.7U,4.3U] tested. Aminoacid sequence analysis showed a high identity to acetaldehyde dehydrogenase in a wide variety of bacteria and to chloroacetaldehyde

dehydrogenase (84%) in *Xanthobacter autotropicus* which is involved in the 1,2-dichloroethane degradative pathway (Van Der Ploeg *et al.*, 1994). In *R.leguminosarum*, the metabolic pathway in which this enzyme participates is not clear. KEGG analysis shows that this enzyme may participate in 1,2-dichloroethane degradation pathway (KEGG pathway is shown in Figure 3.2.12 and discussed earlier in the section 3.2.5.7). However, the competition data suggest that the function of AcoD is important for Rlv3841 to survive in the rhizosphere.

RU4259, mutated in *RL1860* (*phhA*) (phenylalanine-4-hydroxylase, RCI = 2.32) was expressed in all the three rhizospheres [4.7U,7.2U,5.3U]. Another strain RU4269, mutated in *RL1863* (4-hydroxyphenylpyruvate dioxygenase, RCI=1.94) expressed in all the three rhizospheres [1.3U,4.0U,1.8U], also showed reduced competitiveness. Both PhhA and Hpd are involved in the phenylalanine and tyrosine degradative pathway (Figure 5.1.3). An earlier reports has shown the presence of phenylalanine and tyrosine in pea root mucilage (Knee *et al.*, 2001). Together that this data indicates that ability to utilize phenylalanine and tyrosine by Rlv3841 in the rhizosphere provides competitive advantage during colonization.

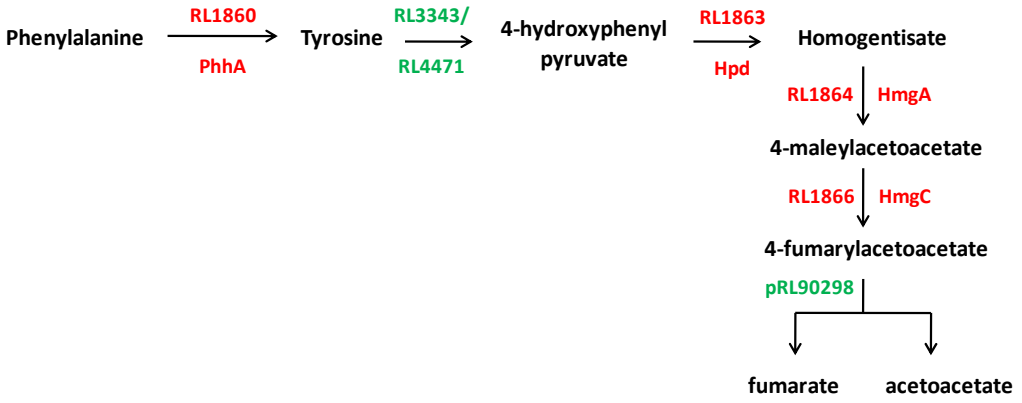



**Figure 5.1.3** Phenylalanine and tyrosine degradation pathway. Genes up-regulated in the pea rhizosphere are marked in red, and those not expressed are shown in green.

RU4229, mutated in *RL0913* (PRC family protein, RCI = 2.28), was expressed in all the three rhizospheres [5.8U,4.0U,4.6U]. PRC family proteins

contain a PRC barrel domain, although the precise function of this protein is not known. However, it has been reported to play a variety of biological functions (Anantharaman & Aravind, 2002). Overall, the mutants in this group suggest that these genes play an important role during rhizosphere colonization.

In the second group (Table 5.1.3.), twelve mutants showed moderately altered competition compared to wildtype during colonization of the pea rhizosphere. The possible role of the protein coded by the gene mutated in each strain is discussed below. RU4231, mutated in *RL3366* (flavoprotein, RCI = 1.99) was highly expressed only in pea rhizosphere [5.03U]. The exact function of this gene in not clear, but may play an important role in host specificity

RU4271, mutated in *pRL80021* (*coxG*) (putative carbon monoxide dehydrogenase subunit G protein, RCI = 1.76), was highly expressed only in the pea rhizosphere [4.07U]. CoxG is a part of the accessory genes flanking *coxMSL* cluster involved in carbon monoxide utilization. Five genes *coxG*, *coxM* [3.2U], *coxS* [3.3U], *coxL* [3.0U] and *coxI* [1.6U] involved in the utilization of carbon monoxide was expressed only in pea rhizosphere indicating that carbon monoxide utilization is important during pea rhizosphere colonization.

RU4234, is mutated in pRL110199 (conserved hypothetical protein, RCI = 1.82), which was highly expressed only in the pea rhizosphere [7.9U]. The precise function of this gene in not clear, further characterization may reveal its role in rhizosphere colonization.

RU4247, mutated in *pRL120632* (putative dehydrogenase, RCI = 1.92), was expressed in pea [5.4U] and alfalfa [3.5U] rhizospheres. Amino acid sequence comparison shows high identity (81%) with RNGR00437 annotated as D-3-phophoglycerate hydrogenase in *Rhizobium* sp. NGR234. In general, D-3-phophoglycerate hydrogenase catalyzes the first committed step in the phosphorylated pathway of L-serine biosynthesis by converting D-3-phosphoglycerate to hydroxypyruvic acid phosphate and subsequently to L-serine.

RU4268, mutated in *pRL110281* (SBP-ABC-T (PepT family), RCI = 1.98), was expressed in all the three rhizospheres [3.5U,3.6U,2.3U]. Amino acid sequence analysis shows high identity (96% to 40%) to an oligopeptide ABC-T in

a wide variety of bacteria. It also showed moderate level identity (23% Id and 39% Smty) to chitin oligosaccharide binding protein (1ZTY_A) of *Vibrio cholera*. The precise function of this gene is not clear, however taken together with its location of this gene next to *pRL110282* which encodes an enzyme α-N-arabinofuranosidase, taken together with the sequence analysis, it suggests that it may be involved in transport of oligosaccharides into the cell. It should be noted that di- and oligosaccharide ABC-T are often classified in the PepT subclass.

RU4274, mutated in *RL0037* (*pckA*) (phosphoenolpyruvate carboxykinase, RCI = 1.83) was expressed in all three rhizospheres [3.0U,4.3U,4.3U]. PckA catalyzes the conversion of oxaloacetate and phosphoenolpyruvate in gluconeogenesis. Transcriptomic data of Rlv3841 in the pea rhizosphere which shows the up-regulation of both the key gluconeogenic enzymes PckA and AceA, confirms the operation of gluconeogenic pathway. Reduced competitiveness of this mutant indicates that gluconeogenesis is one of the important carbon metabolic pathways operated in the rhizosphere. It reinforces the idea that the rhizosphere is likely to be dominated by organic acids, not sugars acting as carbon sources for bacteria.

RU4258, mutated in *RL3424* (*dctA*) (C4-dicarboxylate transport protein, RCI = 1.8) was expressed in all the three rhizospheres [5.0U,6.8U,4.2U]. Dct transport system transports dicarboxylates into the cell. Reduced competitiveness of this mutant indicates that dicarboxylates are one of the most important carbon sources in the rhizosphere.

RU4265, mutated in *RL2418* (SBP-ABC-T (CUT1 family), RCI = 1.92) which was found to be highly expressed in all rhizospheres [2U,3U.3.2U] and in free-living Rlv3841 supplemented with pea root exudates [3U] or hesperetin [3.2U]. The precise function of this gene is not known, but it likely to be involved in sugar uptake.

RU4260, mutated in *RL4274* (HlyD transmembrane efflux protein, RCI = 1.76) was expressed in three rhizospheres. However, it is very highly induced in the pea [7.2U] compared to those of alfalfa [1.5U] and sugarbeet [1.5U]. RL4274 has high identity (89%) to RHECIAT_CH0004005 annotated as putative

multidrug efflux protein in *R. etli* CIAT 652. The HlyD efflux proteins are membrane fusion proteins involved in export variety of compounds across the membrane, the disruption in the export of compounds may have caused the inability to compete in the rhizosphere.

RU4235, mutated in *RL1694* (putative aldo-keto reductase, RCI = 1.9) [5.1U,3.2U,1.6U] and RU4297, mutated in *RL3982* (conserved hypothetical protein, 2.0) [4.9U,3.8U,4.3U] were expressed in all three rhizospheres. The precise functions of these genes are not clear. However, their high expression and the reduced competitiveness of the mutants suggest that they may play an important role during rhizosphere colonization.

In the third group (Table: 5.1.3.), eleven mutants showed slightly altered competition compared to wildtype during colonization of pea rhizosphere. The possible role of the protein coded by the gene mutated in each strain is discussed below. RU4248, mutated in *pRL80026* (*livJ*) (putative branched chain amino acid ABC-T SBP, RCI = 1.58) was expressed only in pea rhizosphere [4.2U]. This gene is a part of a high affinity amino acid ABC-T system (*livJMHGF*), which was found to be expressed only in the pea rhizosphere, although the amino acid specificity of this transport system is not known, it is likely to recognise an aliphatic amino acid.

RU4230, mutated in *RL3130* (putative phosphatase protein, RCI = 1.67) which was expressed in both pea and alfalfa rhizospheres. Amino acid sequence shows high identity (84%) to inositol monophosphatase protein of *R. leguminosarum* bv. *trifolii* WSM2024 and found in wide variety of bacteria. Also, it shares moderate homology (35-48%) with many archaeal fructose-1,6-bisphophatases, which catalyze the conversion of fructose-1,6,-bisphosphate to fructose-6-phosphate in gluconeogenesis. The precise function of this gene is not known, but the reduced competitiveness suggests that it may be involved in this core metabolic pathway.

RU4311, mutated in *pRL100444* (putative oxidoreductase, RCI = 1.44) which was expressed in both pea [7.3U] and alfalfa [3.2U] rhizospheres. The

precise function of both of this gene was not known, but it may be involved in central metabolic reactions.

RU4261, mutated in *RL3860* (putative permease component of ABC-T (CUT1 family), RCI = 1.7) was expressed in both pea [4.8U] and alfalfa [1.5U] rhizospheres. RL3860 has moderate-high identity (44% to 95%) to many sugar ABC-Ts in wide variety of bacteria. It has fairly low identity (34%) to the maltose binding protein (MalG) of *B. melitensis* 16M. This gene along with *RL3859,60-2* may code for a fully functional ABC-T involved in sugar, possibly a maltose like sugar transport, which may be important during rhizosphere colonization.

RU4233, mutated in *RL3016* (*pcaH2*) (protocatechuate 3,4 dioxygenase beta chain, RCI = 1.44) was very highly expressed in all the three rhizospheres. The protocatechuate catabolic genes are located on the plasmid (pRL11) as a gene cluster (*pca*). However, this gene is located in the chromosome as an orphan gene. The precise function of this gene is not known, but may be involved in aromatic compound catabolism.

RU4256, mutated in *pRL90085* (putative periplasmic SBP-ABC-T (CUT2 family), RCI = 1.56), was expressed in both pea [4.8U] and alfalfa [1.0U] rhizospheres. pRL90085 has high identity (65%-95%) to carbohydrate ABC-T in wide variety of bacteria and also shows high identity specifically to the periplasmic ribose binding protein (RbsB) of *P. syringae* pv. tomato str. DC3000. This protein may form a functional ABC-T along with pRL90082-3 a permease component of ABC-T and pRL90084 (RbsA) a ribose ATP binding protein. However, the other genes in thus ABC operon were not expressed in the rhizosphere. The high expression of this gene in the pea rhizosphere and the reduced competitiveness of its mutant suggest that this gene may be involved in the transport of ribose or ribose-like sugars, which is important during rhizosphere colonization.

RU4295, mutated in *pRL120500* (putative SBP-ABC-T, RCI = 1.52) was expressed in all three rhizospheres [1.6U,2.4U,2.0U]. pRL120500 has high identity with a protein Meso_0346, coding for a twin arginine translocation

pathway signal of *Mesorhizobium* sp. BNC1 and moderate identity (45%-73%) to a TRAP-T dicarboxylate transporter in a wide variety of bacteria.

RU4266, mutated in *pRL120479* (conserved hypothetical protein, RCI = 1.45), was highly expressed only in the pea [5.8U] rhizosphere. Interestingly, a database similarity search did not show any significant matches, indicating this gene is unique to Rlv3841. The function of the gene is not known, however the competition data suggests that it may play a role in rhizosphere colonization. Finally three mutant strains, RU4255 mutated in *RL3272* (conserved hypothetical protein, 1.48), RU4309 mutated in *pRL90055* (conserved hypothetical protein, 1.55) and RU4257 mutated in *RL2259* (conserved hypothetical protein, 1.66) showed reduced competitiveness, the precise functions of these genes are not clear.

The genes mutated were selected based on the expression profiles of genes specifically induced in pea, legume or all three rhizospheres, and includes genes involved in transport, regulation, metabolism and performing other functions. The selected set comprised of genes belonging to 12 different functional categories including those annotated as "conserved hypothetical". The competition studies were performed in 8 replicates, to ensure the data is statistically significant. The positive (*thiM* mutant) and negative (*nifH* mutant) controls were as expected and confirmed earlier results, substantiating the integrity of the competition data. Of the 28 mutants, 7 had disruption in genes which were induced only in the pea rhizosphere, 6 had disruption in genes which were induced in both pea and alfalfa (i.e. legume) rhizospheres and 15 had disruption in genes which were induced in all the three rhizospheres.

Two of the pea-specific genes mutated were strongly altered in competition against the wildtype. In the pea-specific set, the competition data support the transcriptomic data (induction of *RL0996*), which suggest the presence of tartrate in the pea rhizosphere (*RL0996*), and that utilization provides selective advantage during rhizosphere colonization. Also, the expression of *RL0680* (*secDF*) in the pea rhizosphere and the reduced competitiveness of SecDF mutant, indicates the functioning of a specific Sec export pathway which might play a role in host-specificity.

Interestingly, mutants of CoxG (pRL80021, putative carbon monoxide dehydrogenase subunit G) showed reduced competitiveness suggesting that utilization of carbon monoxide is important during colonization of the pea rhizosphere. Overall, the rhizosphere competition studies identified that 60% (28/46) of the mutants tested, show altered competitiveness, when compared against wildtype in colonizing the pea rhizosphere. These data further validate the strategy developed in this thesis and reinforce the idea that there is correlation between up-regulated genes and their biological role in the rhizosphere. However, there is not much correlation between the degrees of induction of the up-regulated

genes with the rhizosphere competitive index (RCI). More competition experiments should be performed with more controls (both down-regulated and unchanged genes), to establish the correlation between degree of induction and the rhizosphere competitive index.

In retrospect, it would also be very valuable to test mutants of all the pea-specific genes for competition. However, time did not permit an even larger set of genes to be studied. It would also be valuable to test the mutants of genes specifically induced in the pea rhizosphere, that strongly attenuate competitiveness in the pea rhizosphere, in both alfalfa and sugar-beet rhizospheres. It is predicted that mutation of these genes would not alter the competitiveness of the strains in the rhizospheres of alfalfa or sugar-beet.

# Chapter 6

# Conclusion and future perspectives

In this thesis a strategy to explore the transcriptome of Rlv3841 in sterile rhizospheres has been developed, building on recent advances in microarray and RNA amplification technology and the release of the genome sequence of Rlv3841 (Young *et al*., 2006). The study was extended to include comparative rhizosphere transcriptomics, which is a very powerful approach enabling a direct comparison of cells grown in different rhizospheres, permitting a focus on pea-specific genes.

Overall, the results showed the involvement of many genes involved in adaptation, nutrient uptake, rhizosphere competitiveness and symbiotic interactions, in addition to many novel genes of known and unknown function. The validation of a sub-set of the differentially expressed genes by qRT-PCR substantiates the quality of the data obtained and the success of the strategy developed. There was only a slight (0.5%) change in the transcriptome of free-living Rlv3841 when grown on pea root-exudates, and all the genes induced in response to pea root-exudates were a sub-set of the core genes induced at all the time-points (7d old pea rhizosphere-1,3,7,dpi expression datasets). Likewise, most of the genes induced in response to hesperetin were a subset of the core set, except a few genes involved in malonate uptake and catabolism. In the 7 day old pea rhizosphere experiments, many genes from different functional categories were induced emphasizing that Rlv3841 makes a large number of changes at the level of gene expression in response to life in the pea rhizosphere.

Moreover, the increased expression of uptake systems and corresponding catabolic genes, as shown in transcriptomic data obtained from this work, covers a wide range of sugars, amino acids, organic acids and essential nutrient uptake and utilization portrays the active and versatile metabolic lifestyle of Rlv3841 in the rhizosphere. In some cases the exact compound transported is not clear, but it is possible that it may play a role, maybe an essential one, in colonization. It has been shown that the ability to catabolize certain sugars such as *myo*-inositol (Fry *et al.*, 2001), rhamnose (Oresnik *et al.*, 1998; Richardson *et al.*, 2004) and

erythritol (Yost *et al.*, 2006) by *R. leguminosarum* and trehalose (Jensen *et al.*, 2005), proline (Jiménez-Zurdo *et al.*, 1995; Jiménez-Zurdo *et al.*, 1997), trigonelline, betaines, carnitine and stachydrine (Boivin *et al.*, 1990; Boivin *et al.*, 1991; Burnet *et al.*, 2000; Goldmann *et al.*, 1991) by *S. meliloti* confers a competitive advantage in rhizosphere colonization or nodulation.

There are many bacteria reported to chemolithoautotrophically utilize $H_2$, $CO_2$, CO or formate as sole carbon and energy source. An earlier report, has shown *Bradyrhizobium japonicum* is capable of growing on $H_2$ or $CO_2$ or CO as the sole energy and/or carbon source (Lorite *et al.*, 2000). Recent research, showed that *S. meliloti* SmA818, can utilize formate and bicarbonate as an energy and carbon source respectively (Pickering & Oresnik, 2008). Transcriptomic data obtained from this work showed the induction of gene cluster (*coxMSL*) encoding carbon monoxide dehydrogenase, a key enzyme involved in the utilization of CO and induction of a gene cluster (*fdsABG*), encoding formate dehydrogenase, a key enzyme involved in formate utilization, indicating the active utilization of CO and formate in the pea rhizosphere.

The induction of the tartrate uptake and catabolic gene cluster (*RL0995-7*) in the pea rhizosphere indicates the presence of tartrate in the pea rhizosphere, which is supported by the altered competitiveness of the tartrate transporter (*RL0996*) mutant compared with the wild-type in the pea rhizosphere. This indicates that the ability to utilize the tartrate present in the rhizosphere provides selective advantage during rhizosphere colonization. Also, another interesting observation is the induction of pRL80071 an enzyme homoserine dehydrogenase. It is well known that pea root exudates are abundant in homoserine and *R. leguminosarum* can utilize it as a sole carbon and nitrogen source (Van Egeraat, 1975).

Interestingly, the induction of enzymes involved in gluconeogenesis (RL0054, (PckA, PEPCK) and RL4012 (fructose-1,6-bisphosphate aldolase) and glyoxylate pathway (RL0761 (AceA, isocitrate lyase) and RL0054 (GlcB, malate synthease G)) suggests that the carbon metabolism in the rhizosphere is driven by organic acids, rather than by sugars.

Recent research with a panel of competitive and uncompetitive rhizobial isolates showed that competitive strains were able to utilize wide range of substrates (sugars, amino and organic acids) as carbon and energy sources, whereas uncompetitive strains can utilize only sugars, which suggests that the ability to metabolize a broad range of substrates provides an advantage in nodulation competitiveness (Wielbo *et al.*, 2007).

Based on the expression profiles obtained from the transcriptomic studies of Rlv3841 in the pea rhizosphere, a schematic representation of all the bacterial functions induced in the 7d old pea rhizosphere was modelled (Figure 6.1.1.).

The age comparison showed that Rlv3841 has a core set of genes which respond in the pea rhizosphere at all three time points. Genes in this core set are involved in stress and adaptation, quorum sensing, nutrient acquisition, sugar, organic and amino acids uptake, efflux and gene regulation. However, a large group are of unknown function. Surprisingly, the repression in 14d and 21d old pea rhizospheres of some of the genes involved in nod factor signalling and early adaptation (like *rapBC*) indicates that the secretion of flavonoids in the root exudates is higher in younger seedlings. Overall, though more genes were expressed in bacteria from a 21d old pea rhizosphere compared to those from 7 or 14 d old pea rhizospheres.

The comparative rhizosphere transcriptomics proved to be a powerful way to identify specific genes expressed in response to a rhizosphere and this approach could be extended to any rhizosphere bacteria. The indirect comparative transcriptomics identified a number of genes specifically expressed in pea, alfalfa and sugar beet rhizospheres. Moreover, a set of legume specific genes expressed only in pea and alfalfa was identified, in addition to the core set of genes expressed in all the rhizospheres which are termed as rhizosphere specific genes. Nod factor signalling genes were only up-regulated in the pea and alfalfa rhizospheres (i.e. they are legume-specific genes), validating the strategy to employ comparative rhizosphere transcriptomics. The direct rhizosphere comparison indicated a smaller number of pea-specific genes (80) than the dpi, age and indirect comparative rhizosphere experiments. Together with the qRT-

PCR validation and rhizosphere competition studies, the strategy for studying the transcriptomics of the rhizosphere bacteria in a sterile rhizosphere proved to be successful. It is thought likely that this strategy could be adapted to successfully study any other bacteria which dwell in the rhizosphere and interact with a host plant.

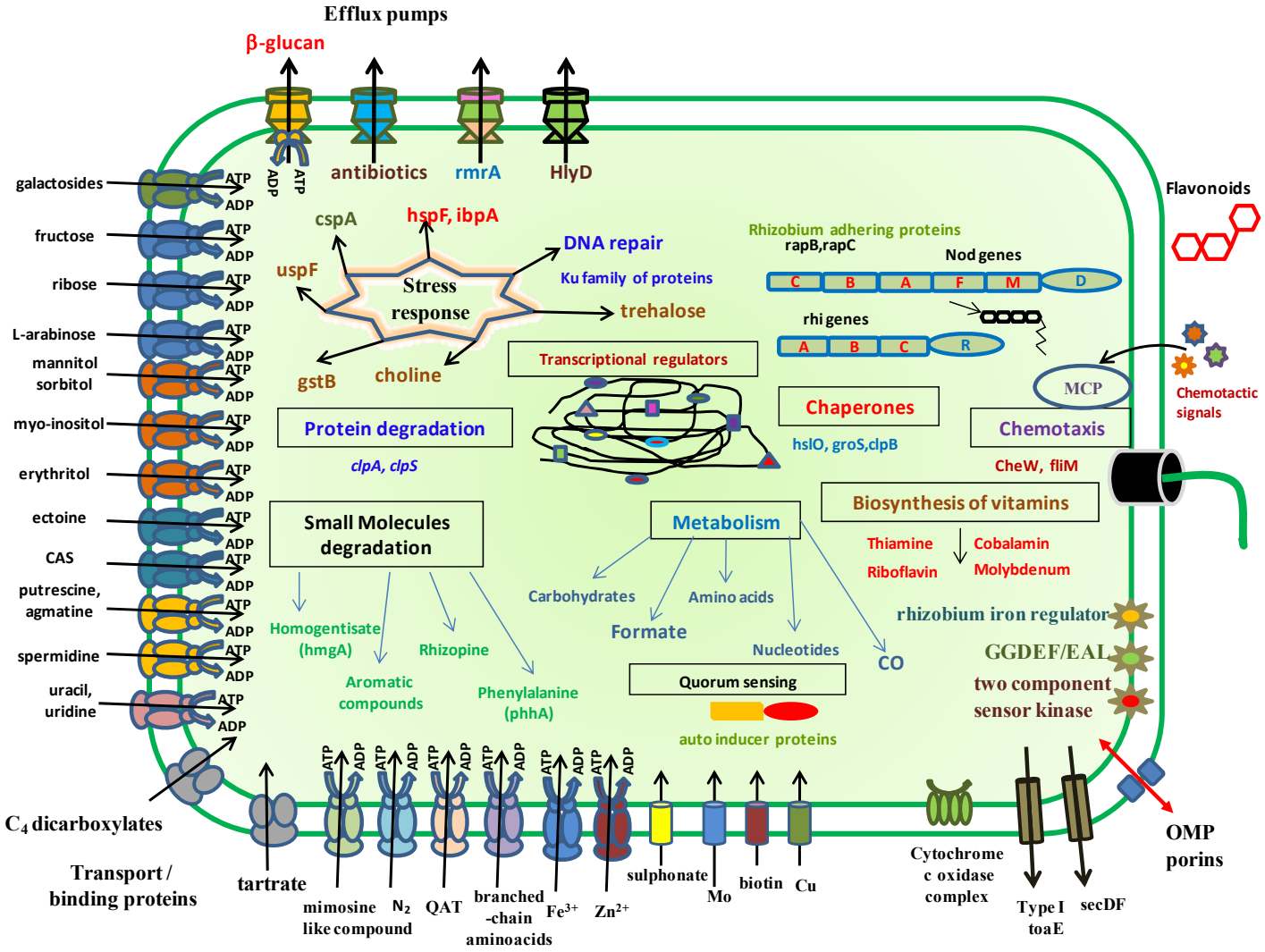



**Figure 6.1.1.** Model of *R. leguminosarum* 3841 transcriptome in the pea rhizosphere**.**

The main objective of this project was to provide a global transcriptomic view of Rlv3841 in the pea rhizosphere and to identify host specific genes expressed during rhizosphere colonization by comparative rhizosphere transcriptomics. This is a first attempt to explore the transcriptomics of Rlv3841 in the pea rhizosphere, combining a sterile rhizosphere system, RNA amplification and microarray technology. This strategy is likely to supersede the existing techniques to study plant-bacterial interactions, both in its power and robustness, and to provide a wealth of information in a global perspective. Over the past decade, many techniques have been developed to study plant-bacterial interactions, but the time-consuming nature of these techniques and their tendency to identify a small number of random genes has always been a limitation.

Though the growth conditions between free-living cells grown on minimal media and cells grown on pea rhizosphere were large, the main objective of this project is to study the effect of root exudates on Rlv3841 in real time. An ideal experiment would be to compare Rlv3841 grown on rhizosphere tubes, with and without pea seedlings. However, Rlv3841 cells failed to grow on the rhizosphere tubes without the pea seedlings, due to absence of carbon source. Another better experiment is the comparison of Rlv3841 cells grown on sugarbeet rhizosphere with cells grown on pea rhizosphere. However, it revealed only genes which had marked difference such as genes involved in nod signalling and other pea specific genes, concealing the information about the common genes involved in survival and adaptation.

The results of this project provide scope for further understanding of rhizobial interactions in the rhizosphere and offer a wide choice of candidate genes for mutational studies. To develop this work further there are several lines of research which could be followed up. Rhizosphere competition studies with the panel of mutants disrupted in pea-specific gene tested in all the three rhizospheres would provide more concrete evidence about the role of the genes in host specificity. High-throughput reporter gene fusion based transportomic studies of

the up-regulated transport systems would provide more information on the composition of the pea root exudates, adding to a comprehensible global picture. Microarray analysis of cells recovered form roots would be very interesting and would be sure to provide more insights about interactions with the host plant. In addition, proteomic based study of the rhizosphere recovered cells would provide more opportunity to explore the interactions, as all transcribed mRNA sequences are not always translated into proteins.

Several studies on small or non-coding RNA's in bacteria showed that they play an important role in regulating gene expression in response to environmental cues like stress and nutrient limitation and etc. Recently research in *S. meliloti*, identified many sRNA, two of which were found to be induced during symbiotic interaction (del Val *et al.*, 2007). A large proportion of genes of unknown function with nucleotide sequence length < 400bp were found among the up-regulated genes in the 7 day old pea rhizosphere, which might be potential candidates for small RNA. Genome-wide identification of small RNA's in Rlv3841 would provide more insights in the gene regulation during rhizosphere colonization and symbiotic interaction with the host plant.

Further characterization of genes, such as those involved in transport and catabolism of tartrate, formate and carbon monoxide would be expected to provide further insight on the information on the carbon and the energy sources available in the rhizosphere. Other genes which could be further characterized include the mimosine-like and the high affinity branched amino acid ABC-T operons.

The rhizosphere is a complex environment where a mixed population will be subjected to intra- and inter-species competition. Several reports have showed that di-, tri- and multi-trophic interactions are necessary for the successful establishment of symbiotic interactions. This transcriptomic strategy can be extended to study di- and tri-trophic interactions with an integrated microarray slide spotted with the genes of two or three interacting organisms. Though this may be a cumbersome task in the beginning, requiring careful selection of oligonucleotides and standardization, it is likely to be the next dimension in

rhizosphere microarrays. It may be possible that this strategy will be possible using next generation high throughput sequencing such as, Solexa or 454 sequencing.

As of now, the information about the regulatory circuits of Rlv3841 in rhizosphere is not characterized on a genomic scale. The temporal and spatial regulation of the genes in the rhizosphere in response to the environmental factors is key process rhizosphere colonization. Though the transcriptomic approaches identify large number of genes, very limited is known about the regulators controlling these rhizosphere-specific genes. A recent report, combined suppressor analysis with IVET (SpyVET) and identified many regulators controlling niche-specific genes in *P. fluorescens* SBW25 (in sugar-beet rhizosphere) (Giddens *et al.*, 2007). Similar technology can be adapted to use with microarrays (SpyArray), which is expected to provide more comprehensive information about the regulatory network of Rlv3841 in the pea rhizosphere.

In the post-genomic systems-biology era, integration of all these data (from transcriptomic, proteomic and metabolomic experiments) will help us to model the symbiotic interactions spanning from rhizosphere colonization to nitrogen fixation. Recent research in *R. etli*, reconstructed the metabolic pathway and modelled nitrogen fixation by integrating all the available experimental data (Resendis-Antonio *et al.*, 2007). The data generated from the strategy developed in this thesis may well complement such approaches in the future, enabling the modelling of the metabolic pathways of Rlv3841.

Overall, the strategy developed here has proved to be a great success in exploring the transcriptomics *R. leguminosarum* in three different (pea, alfalfa and sugar-beet) rhizospheres. It has identified a set of genes specifically induced in the pea rhizosphere and, another set, specifically induced in the rhizospheres of legumes (pea and alfalfa). Construction of mutants has established the influence of many of these genes in the ability of *R. leguminosarum* 3841 to colonise the pea rhizosphere. Furthermore the techniques developed in this work could be extended to study any rhizosphere bacteria.

# References

**Aguilar, J. M. M., Ashby, A. M., Richards, A. J. M., Loake, G. J., Watson, M. D. & Shaw, C. H. (1988).** Chemotaxis of *Rhizobium leguminosarum* biovar *phaseoli* towards flavonoid inducers of the symbiotic nodulation genes. *Journal General Microbiology* **134**, 2741-2746.

**Allaway, D., Lodwig, E., Crompton, L. A., Wood, M., Parsons, T. R., Wheeler, T. & Poole, P. S. (2000).** Identification of alanine dehydrogenase and its role in mixed secretion of ammonium and alanine by pea bacteroids. *Molecular Microbiology* **36**, 508-515.

**Allaway, D., Schofield, N. A., Leonard, M. E., Gilardoni, L., Finan, T. M. & Poole, P. S. (2001).** Use of differential fluorescence induction and optical trapping to isolate environmentally induced genes. *Environmental Microbiology* **3**, 397-406.

**Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990).** Basic local alignment search tool. *Journal Molecular Biology* **215**, 403-410.

**Altschul, S. F. & Lipman, D. J. (1990).** Protein database searches for multiple alignments. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 5509-5513.

**Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997).** Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389-3402.

**Ampe, F., Kiss, E., Sabourdy, F. & Batut, J. (2003).** Transcriptome analysis of *Sinorhizobium meliloti* during symbiosis. *Genome Biology* **4**, R15.

**An, J. H. & Kim, Y. S. (1998).** A gene cluster encoding malonyl-CoA decarboxylase (MatA), malonyl-CoA synthetase (MatB) and a putative dicarboxylate carrier protein (MatC) in *Rhizobium trifolii* - Cloning, sequencing, and expression of the enzymes in *Escherichia coli*. *European Journal of Biochemistry* **257**, 395-402.

**Anantharaman, V. & Aravind, L. (2002).** The PRC-barrel: a widespread, conserved domain shared by photosynthetic reaction center subunits and proteins of RNA metabolism. *Genome Biology* **3**. research0061.1-0061.9

**Aravind, L. & Koonin, E. V. (2001).** Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Research* **11**, 1365-1374.

**Ardourel, M., Demont, N., Debelle, F. D., Maillet, F., Debilly, F., Prome, J. C., Denarie, J. & Truchet, G. (1994).** *Rhizobium meliloti* lipooligosaccharide nodulation factors - different structural requirements for bacterial entry into target

root hair-cells and induction of plant symbiotic developmental responses. *Plant Cell* **6**, 1357-1374.

**Arias-Barrau, E., Olivera, E. R., Luengo, J. M., Fernandez, C., Galan, B., Garcia, J. L., Diaz, E. & Minambres, B. (2004).** The Homogentisate pathway: a central catabolic pathway involved in the degradation of L-phenylalanine, L-tyrosine, and 3-hydroxyphenylacetate in *Pseudomonas putida. Journal of Bacteriology* **186**, 5062-5077.

**Ashkin, A. & Dziedzic, J., M. (1987).** Optical trapping and manipulation of viruses and bacteria. *Science* **235**, 1517-1520.

**Ashkin, A., Dziedzic, J. M. & Yamane, T. (1987).** Optical trapping and manipulation of single cells using infrared laser beams. *Nature* **330**, 769-771.

**Ausmees, N., Jacobsson, K. & Lindberg, M. (2001).** A unipolarly located, cell-surface-associated agglutinin, RapA, belongs to a family of Rhizobium-adhering proteins (Rap) in *Rhizobium leguminosarum* bv. *trifolii*. *Microbiology* **147**, 549 - 559.

**Bahar, M., deMajnik, 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. *Molecular Plant-Microbe Interactions* **11**, 1057-1068.

**Bais, H. P., Walker, T. S., Stermitz, F. R., Hufbauer, R. A. & Vivanco, J. M. (2002).** Enantiomeric-dependent phytotoxic and antimicrobial activity of (+/-)-catechin. A rhizosecreted racemic mixture from spotted knapweed. *Plant Physiology* **128**, 1173-1179.

**Bais, H. P., Park, S. W., Weir, T. L., Callaway, R. M. & Vivanco, J. M. (2004).** How plants communicate using the underground information superhighway. *Trends Plant Science* **9**, 26-32.

**Bais, H. P., Weir, T. L., Perry, L. G., Gilroy, S. & Vivanco, J. M. (2006).** The role of root exudates in rhizosphere interactions with plants and other organisms. *Annual Review Plant Biology* **57**, 233-266.

**Barbour, W. M., Hattermann, D. R. & Stacey, G. (1991).** Chemotaxis of *Bradyrhizobium japonicum* to soybean exudates. *Applied and Environmental Microbiology* **57**, 2635-2639.

**Bardin, S., Dan, S., Osteras, M. & Finan, T. M. (1996).** A phosphate transport system is required for symbiotic nitrogen fixation by *Rhizobium meliloti*. *Journal of Bacteriology* **178**, 4540-4547.

**Barker, L. P., Brooks, D. M. & Small, P. L. (1998).** The identification of *Mycobacterium marinum* genes differentially expressed in macrophage phagosomes using promoter fusions to green fluorescent protein. *Molecular Microbiology* **29**, 1167-1177.

**Barnett, M. J., Fisher, R. F., Jones, T. & other authors (2001).** Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 9883-9888.

**Barr, M., East, A. K., Leonard, M., Mauchline, T. H. & Poole, P. S. (2008).** *In vivo* expression technology (IVET) selection of genes of *Rhizobium leguminosarum* biovar *viciae* A34 expressed in the rhizosphere. *FEMS Microbiology Letters*. **282, 219-227.**

**Batut, J. & Boistard, P. (1994).** Oxygen control in *Rhizobium*. *Antonie Van Leeuwenhoek International Journal Of General And Molecular Microbiology* **66**, 129-150.

**Batut, J., Andersson, S. G. E. & O'Callaghan, D. (2004).** The evolution of chronic infection strategies in the alpha- proteobacteria. *Nature Reviews Microbiology* **2**, 933-945.

**Becker, A., Berges, H., Krol, E. & other authors (2004).** Global changes in gene expression in *Sinorhizobium meliloti 1021* under microoxic and symbiotic conditions. *Molecular Plant-Microbe Interactions* **17**, 292-303.

**Begum, A. A., Leibovitch, S., Migner, P. & Zhang, F. (2001).** Inoculation of pea (*Pisum sativum* L.) by *Rhizobium leguminosarum* bv. *viciae* preincubated with naringenin and hesperetin or application of naringenin and hesperetin directly into soil increased pea nodulation under short season conditions. *Plant and Soil* **237**, 71-80.

**Bell, E. A. (2003).** Nonprotein amino acids of plants: Significance in medicine, nutrition, and agriculture. *Journal of Agricultural and Food Chemistry* **51**, 2854-2865.

**Benjamini, Y. & Hochberg, Y. (1995).** Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of Royal Statistical Society* **57**, 289-300.

**Beringer, J. E. (1974).** R factor transfer in *Rhizobium leguminosarum*. *Journal General Microbiology* **84**, 188-198.

**Bertin, C., Xiaohan, Y. & Weston, L. (2003).** The role of root exudates and allelochemicals in the rhizosphere. *Plant and Soil* **256**, 67–83.

**Bittinger, M. A., Milner, J. L., Saville, B. J. & Handelsman, J. (1997).** Rosr, a determinant of nodulation competitiveness in *Rhizobium etli*. *Molecular Plant-Microbe Interactions* **10**, 180-186.

**Bittinger, M. A. & Handelsman, J. (2000).** Identification of genes in the RosR regulon of *Rhizobium etli*. *Journal of Bacteriology* **182**, 1706-1713.

**Bladergroen, M. R., Badelt, K. & Spaink, H. P. (2003).** Infection-blocking genes of a symbiotic *Rhizobium leguminosarum* strain that are involved in temperature-dependent protein secretion. *Molecular Plant-Microbe Interactions* **16**, 53-64.

**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.

**Boivin, C., Barran, L. R., Malpica, C. A. & Rosenberg, C. (1991).** Genetic analysis of a region of the *Rhizobium meliloti* pSym plasmid specifying catabolism of trigonelline, a secondary metabolite present in legumes. *Journal of Bacteriology* **173**, 2809-2817.

**Boldt, T. S., Sørensen, J., Karlson, U., Molin, S. & Ramos, C. (2004).** Combined use of different Gfp reporters for monitoring single-cell activities of a genetically modified PCB degrader in the rhizosphere of alfalfa. *FEMS Microbiology Ecology* **48**, 139-148.

**Bolhuis, A., Broekhuizen, C. P., Sorokin, A., van Roosmalen, M. L., Venema, G., Bron, S., Quax, W. J. & van Dijl, J. M. (1998).** SecDF of *Bacillus subtilis*, a molecular siamese twin required for the efficient secretion of proteins. *Journal of Biological Chemistry* **273**, 21217-21224.

**Bolton, H., J. K. Fredrickson & Elliot., L. E. (1992).** Microbial ecology of the rhizosphere. *In Soil Microbial Ecology*. *F. Blaine Metting, Jr. (eds.), Marcel Dekker, New York.* p. 27-63

**Born, T. L. & Blanchard, J. S. (1999).** Structure/function studies on enzymes in the diaminopimelate pathway of bacterial cell wall biosynthesis. *Current Opinion in Chemical Biology* **3**, 607-613.

**Borthakur, D., Soedarjo, M., Fox, P. M. & Webb, D. T. (2003).** The mid genes of *Rhizobium* sp strain TAL1145 are required for degradation of mimosine into 3-hydroxy-4-pyridone and are inducible by mimosine. *Microbiology* **149**, 537-546.

**Boussau, B., Karlberg, E. O., Frank, A. C., Legault, B.-A. & Anderson, S. G. E. (2004).** Computational inferences of scenarios for α–proteobacterial genome evolution. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 9722-9727.

**Bowra, B. J. & Dilworth, M. J. (1981).** Motility and chemotaxis towards sugars in *Rhizobium leguminosarum*. *Journal General Microbiology* **126**, 231-235.

**Bringhurst, R. M., Cardon, Z. G. & Gage, D. J. (2001).** Galactosides in the rhizosphere: Utilization by *Sinorhizobium meliloti* and development of a biosensor. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 4540-4545.

**Bringhurst, R. M. & Gage, D. J. (2002).** Control of inducer accumulation plays a key role in succinate-mediated catabolite repression in *Sinorhizobium meliloti*. *Journal of Bacteriology* **184**, 5385-5392.

**Broughton, W. J., Jabbouri, S. & Perret, X. (2000).** Keys to symbiotic harmony. *Journal of Bacteriology* **182**, 5641-5652.

**Burnet, M. W., Goldmann, A., Message, B., Drong, R., El Amrani, A., Loreau, O., Slightom, J. & Tepfer, D. (2000).** The stachydrine catabolism region in *Sinorhizobium meliloti* encodes a multi-enzyme complex similar to the xenobiotic degrading systems in other bacteria. *Gene* **244**, 151-161.

**Bustin, S. A. (2000).** Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology* **25**, 169-193.

**Bustin, S. A. (2002).** Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *Journal of Molecular Endocrinology* **29**, 23-39.

**Callaham, D. A. & Torrey, J. G. (1981).** The structural basis for infection of root hairs of *Trifolium repens* by *Rhizobium*. *Canadian Journal of Botany* **59**, 1647-1664.

**Cangelosi, G. A., Martinetti, G., Leigh, J. A., Lee, C. C., Thienes, C. & Nester, E. W. (2007).** Role of *Agrobacterium tumefaciens* ChvA protein in export of {beta}-1,2-glucan. *Journal of Bacteriology* **189**, 6742-.

**Capela, D., Barloy-Hubler, F., Gouzy, J. & other authors (2001).** Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 9877-9882.

**Cardon, Z. G. & Gage, D. J. (2006).** Resource exchange in the rhizosphere: Molecular tools and the microbial perspective. *Annual Review of Ecology Evolution and Systematics* **37**, 459-488.

**Casavant, N. C., Thompson, D., Beattie, G. A., Phillips, G. J. & Halverson, L. J. (2003).** Use of a site-specific recombination-based biosensor for detecting bioavailable toluene and related compounds on roots. *Environmental Microbiology* **5**, 238-249.

**Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. (1994).** Green fluorescent protein as a marker for gene expression. *Science* **263**, 802-805.

**Chan, J. M., Wu, W., Dean, D. R. & Seefeldt, L. C. (2000).** Construction and characterization of a heterodimeric iron protein: defining roles for adenosine triphosphate in nitrogenase catalysis. *Biochemistry* **39**, 7221-7228.

**Chan, M. K., Kim, J. & Rees, D. C. (1993).** The nitrogenase FeMo-cofactor and P-cluster pair: 2.2 A resolution structures. *Science* **260**, 792-794.

**Chang, J. H., Urbach, J. M., Law, T. F., Arnold, L. W., Hu, A., Gombar, S., Grant, S. R., Ausubel, F. M. & Dangl, J. L. (2005).** A high-throughput, near-saturating screen for type III effector genes from *Pseudomonas syringae*. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 2549-2554.

**Chang, W.-S., Franck, W. L., Cytryn, E., Jeong, S., Joshi, T., Emerich, D. W., Sadowsky, M. J., Xu, D. & Stacey, G. (2007).** An Oligonucleotide microarray resource for transcriptional profiling of *Bradyrhizobium japonicum*. *Molecular Plant-Microbe Interactions* **20**, 1298-1307.

**Chaudhuri, R. R. & Pallen, M. J. (2006).** xBASE, a collection of online databases for bacterial comparative genomics. *Nucleic Acids Res* **34**, D335-337.

**Chen, W. M., James, E. K., Prescott, A. R., Kierans, M. & Sprent, J. I. (2003).** Nodulation of *Mimosa* spp. by the beta-proteobacterium *Ralstonia taiwanensis*. *Molecular Plant-Microbe Interactions* **16**, 1051-1061.

**Chen, K. & Pachter, L. (2005).** Bioinformatics for Whole-Genome Shotgun Sequencing of Microbial Communities. *PLoS Computational Biology* **1**, e24.

**Chomczynski, P. & Rymaszewski, M. (2006).** Alkaline polyethylene glycol-based method for direct PCR from bacteria, eukaryotic tissue samples, and whole blood. *Biotechniques* **40**, 454, 456, 458.

**Cooper, J. E. (2004).** Multiple responses of rhizobia to flavonoids during legume root infection. In *Advances In Botanical Research Incorporating Advances In Plant Pathology, Vol 41*, pp. 1-62.

**Cooper, J. E. (2007).** Early interactions between legumes and rhizobia: disclosing complexity in a molecular dialogue. *Journal of Applied Microbiology* **103**, 1355-1365.

**Copeland, L., Quinnell, R. G. & Day, D. A. (1989).** Malic enzyme activity in bacteroids from soybean nodules. *Journal of General Microbiology* **135**, 2005-2011.

**Coutinho, P. M., and Henrissat, B. (1999).** Carbohydrate-active enzymes: an integrated database approach. *In "Recent Advances in Carbohydrate Bioengineering", HJ Gilbert, G Davies, B Henrissat and B Svensson (Eds), The Royal Society of Chemistry, Cambridge,UK* , p3-12.

**Crossman, L. C., Castillo-Ramirez, S., McAnnula, C. & other authors (2008).** A common genomic framework for a diverse assembly of plasmids in the symbiotic nitrogen fixing bacteria. *PLoS ONE* **3**, e2567.

**Crouzet, P. & Otten, L. (1995).** Sequence and mutational analysis of a tartrate utilization operon from *Agrobacterium vitis*. *Journal of Bacteriology* **177**, 6518-6526.

**Cubo, M. T., Economou, A., Murphy, G., Johnston, A. W. & Downie, J. A. (1992).** Molecular characterization and regulation of the rhizosphere-expressed genes *rhiABCR* that can influence nodulation by *Rhizobium leguminosarum* biovar viciae. *Journal of Bacteriology* **174**, 4026 - 4035.

**Curl, E. A. and B. Truelove. 1986**. The Rhizosphere. *(In: Microbial Interactions: Rhizosphere in Relation to Plant Nutrition and Growth). Springer-Verlag, New York.* p. 167-190.

**D'Argenio, D. A. & Miller, S. I. (2004).** Cyclic di-GMP as a bacterial second messenger. *Microbiology* **150**, 2497-2502.

**Daigle, F., Hou, J. Y. & Clark-Curtiss, J. E. (2002).** Microbial gene expression elucidated by selective capture of transcribed sequences (SCOTS). *Methods Enzymology* **358**, 108-122.

**Darwent, M. J., Paterson, E., McDonald, A. J. & Tomos, A. D. (2003).** Biosensor reporting of root exudation from *Hordeum vulgare* in relation to shoot nitrate concentration. *Journal of Experimental Botany* **54**, 325-334.

**Day, D. A. & Copeland, L. (1991).** Carbon metabolism and compartmentation in nitrogen-fixing legume nodules. *Plant Physiology and Biochemistry* **29**, 185-201.

**del Val, C., Rivas, E., Torres-Quesada, O., Toro, N. & Jimenez-Zurdo, J. I. (2007).** Identification of differentially expressed small non-coding RNAs in the legume endosymbiont *Sinorhizobium meliloti* by comparative genomics. *Molecular Microbiology* **66**, 1080-1091.

**Denarie, J. & Cullimore, J. (1993).** Lipo-oligosaccharide nodulation factors - a minireview new class of signaling molecules mediating recognition and morphogenesis. *Cell* **74**, 951-954.

**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.

**DeRisi, J. L., Iyer, V. R. & Brown, P. O. (1997).** Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680-686.

**DeVries, G. E., in T'Veld, P. & Kijne, J. W. (1980).** Production of organic acids in *Pisum sativum* root nodules as a result of oxygen stress. *Plant Science Letters* **20**, 115-123.

**Dixon, R. & Kahn, D. (2004).** Genetic regulation of biological nitrogen fixation. *Nature Reviews Microbiology* **2**, 621-631.

**Djordjevic, M. A., Chen, H. C., Natera, S., Van Noorden, G., Menzel, C., Taylor, S., Renard, C., Geiger, O. & Weiller, G. F. (2003).** A global analysis of protein expression profiles in *Sinorhizobium meliloti*: Discovery of new genes for nodule occupancy and stress adaptation. *Molecular Plant-Microbe Interactions* **16**, 508-524.

**Dominguez-Ferreras, A., Perez-Arnedo, R., Becker, A., Olivares, J., Soto, M. J. & Sanjuan, J. (2006).** Transcriptome profiling reveals the importance of plasmid pSymB for osmoadaptation of *Sinorhizobium meliloti*. *Journal of Bacteriology* **188**, 7617-7625.

**Downie, J. A. & Surin, B. P. (1990).** Either of two *nod* gene loci can complement the nodulation defect of a nod deletion mutant of *Rhizobium leguminosarum* bv. *viciae*. *Molecular and General Genetics* **222**, 81 - 86.

**Dunn, A. K., Klimowicz, A. K. & Handelsman, J. (2003).** Use of a promoter trap to identify *Bacillus cereus* genes regulated by tomato seed exudate and a rhizosphere resident, *Pseudomonas aureofaciens*. *Applied and Environmental Microbiology* **69**, 1197-1205.

**Dylan, T., Helinski, D. R. & Ditta, G. S. (1990).** Hypoosmotic adaptation in *Rhizobium meliloti* requires beta-(1--2)-glucan. *Journal of Bacteriology* **172**, 1400-1408.

**Earl, C. D., Ronson, C. W. & Ausubel, F. M. (1987).** Genetic and structural-analysis of the *Rhizobium meliloti Rhizobium meliloti fixA, fixB, fixC,* and *fixX* genes. *Journal of Bacteriology* **169**, 1127-1136.

**Economou, A., Hawkins, F. K., Downie, J. A. & Johnston, A. W. (1989).** Transcription of rhiA, a gene on a *Rhizobium leguminosarum* bv. *viciae* Sym plasmid, requires *rhiR* and is repressed by flavanoids that induce nod genes. *Molecular Microbiology* **3**, 87-93.

**Eichenberger, P., Jensen, S. T., Conlon, E. M. & other authors (2003).** The [sigma]E regulon and the identification of additional sporulation genes in *Bacillus subtilis*. *Journal of Molecular Biology* **327**, 945-972.

**Eichler, K., Buchet, A., Bourgis, F., Kleber, H. P. & Mandrandberthelot, M. A. (1995).** The *fix Escherichia coli* region contains 4 genes related to carnitine metabolism. *Journal of Basic Microbiology* **35**, 217-227.

**Einsle, O., Tezcan, F. A., Andrade, S. L., Schmid, B., Yoshida, M., Howard, J. B. & Rees, D. C. (2002).** Nitrogenase MoFe-protein at 1.16 A resolution: a central ligand in the FeMo-cofactor. *Science* **297**, 1696-1700.

**El-Tarabily, K. A. (2004).** Suppression of *Rhizoctonia solani* diseases of sugar beet by antagonistic and plant growth-promoting yeasts. *Journal of Applied Microbiology* **96**, 69-75.

**Emons, A. M. & Mulder, B. M. (2000).** How the deposition of cellulose microfibrils builds cell wall architecture. *Trends Plant Science* **5**, 35-40.

**Engelke, T., Jagadish, M. N. & Pühler, A. (1987).** Biochemical and genetical analysis of *Rhizobium meliloti* mutants defective in $C_4$-dicarboxylate transport. *Journal of General Microbiology* **133**, 3019-3029.

**Engelke, T., Jording, D., Kapp, D. & Pühler, A. (1989).** Identification and sequence analysis of the *Rhizobium meliloti dctA* gene encoding the $C_4$-dicarboxylate carrier. *Journal of Bacteriology* **171**, 5551-5560.

**Ferguson, G. P., Roop, R. M. & Walker, G. C. (2002).** Deficiency of a *Sinorhizobium meliloti bacA* mutant in alfalfa symbiosis correlates with alteration of the cell envelope. *Journal of Bacteriology* **184**, 5625-5632.

**Figurski, D. H. & Helinski, D. R. (1979).** Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans.

*Proceedings of the National Academy of Sciences of the United States of America* **76**, 1648-1652.

**Finan, T. M., Wood, J. M. & Jordan, C. (1981).** Succinate transport in *Rhizobium leguminosarum*. *Journal of Bacteriology* **148**, 193-202.

**Finan, T. M., Wood, J. M. & Jordan, D. C. (1983).** Symbiotic properties of $C_4$-dicarboxylic acid transport mutants of *Rhizobium leguminosarum*. *Journal of Bacteriology* **154**, 1403-1413.

**Finan, T. M., Mcwhinnie, E., Driscoll, B. & Watson, R. J. (1991).** Complex symbiotic phenotypes result from gluconeogenic mutations in *Rhizobium meliloti*. *Molecular Plant-Microbe Interactions* **4**, 386-392.

**Finan, T. M., Weidner, S., Wong, K. & other authors (2001).** The complete sequence of the 1,683-kb pSymB megaplasmid from the $N_2$-fixing endosymbiont *Sinorhizobium meliloti*. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 9889-9894.

**Finn, R. D., Mistry, J., Schuster-Bockler, B. & other authors (2006).** Pfam: clans, web tools and services. *Nucleic Acids Research* **34**, D247-251.

**Fry, J., Wood, M. & Poole, P. S. (2001).** Investigation of myo-inositol catabolism in *Rhizobium leguminosarum* bv. *viciae* and its effect on nodulation competitiveness. *Molecular Plant-Microbe Interactions* **14**, 1016-1025.

**Fuhrer, T., Fischer, E. & Sauer, U. (2005).** Experimental identification and quantification of glucose metabolism in seven bacterial species. *Journal of Bacteriology* **187**, 1581-1590.

**Gade D, Thiermann J, Markowsky D, Rabus R (2003).** Evaluation of two-dimensional difference gel electrophoresis for protein profiling. Soluble proteins of the marine bacterium Pirellula sp. strain 1. *Journal of Molecular Microbiology and Biotechnology*, **5(4)**:240-51.

**Gage, D. J., Bobo, T. & Long, S. R. (1996).** Use of green fluorescent protein to visualize the early events of symbiosis between *Rhizobium meliloti* and alfalfa (*Medicago sativa*). *Journal of Bacteriology* **178**, 7159-7166.

**Gage, D. J. & Long, S. R. (1998).** α-Galactoside uptake in *Rhizobium meliloti*: Isolation and characterization of *agpA*, a gene encoding a periplasmic binding protein required for melibiose and raffinose utilization. *Journal of Bacteriology* **180**, 5739-5748.

**Gage, D. J. (2002).** Analysis of infection thread development using Gfp- and DsRed- expressing *Sinorhizobium meliloti*. *Journal of Bacteriology* **184**, 7042-7046.

**Gage, D. J. (2004).** Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiology and Molecular Biology Reviews* **68**, 280-300.

**Gal, M., Preston, G. M., Massey, R. C., Spiers, A. J. & Rainey, P. B. (2003).** Genes encoding a cellulosic polymer contribute toward the ecological success of *Pseudomonas fluorescens* SBW25 on plant surfaces. *Molecular Ecology* **12**, 3109-3121.

**Galibert, F., Finan, T. M., Long, S. R. & other authors (2001).** The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* **293**, 668 - 672.

**Gao, M. & Teplitski, M. (2008).** RIVET-a tool for *in vivo* analysis of symbiotically relevant gene expression in *Sinorhizobium meliloti*. *Molecular Plant- Microbe Interact* **21**, 162-170.

**Gaworzewska, E. T. & Carlile, M. J. (1982).** Positive chemotaxis of *Rhizobium leguminosarum* and other bacteria towards root exudates from legumes and other plants. *Journal of General Microbiology* **128**, 1179-1188.

**Gentleman, R., Carey, V., Bates, D. & other authors (2004).** Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **5**, R80.

**Georgiadis, M. M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc, J. J. & Rees, D. C. (1992).** Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii*. *Science* **257**, 1653-1659.

**Giddens, S. R., Jackson, R. W., Moon, C. D., Jacobs, M. A., Zhang, X.-X., Gehrig, S. M. & Rainey, P. B. (2007).** Mutational activation of niche-specific genes provides insight into regulatory networks and bacterial function in a complex environment. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 18247-18252.

**Giraud, E., Moulin, L., Vallenet, D. & other authors (2007).** Legumes Symbioses: Absence of nod genes in photosynthetic *Bradyrhizobia*. *Science* **316**, 1307-1312.

**Glenn, A. R., Poole, P. S. & Hudman, J. F. (1980).** Succinate uptake by free-living and bacteroid forms of *Rhizobium leguminosarum*. *Journal of General Microbiology* **119**, 267-271.

**Glenn, A. R., Reeve, W. G., Tiwari, R. P. & other authors (1999).** Acid tolerance in root nodule bacteria. *Novartis Foundation Symposium* **221**, 112-130.

**Goff, L. A., Bowers, J., Schwalm, J., Howerton, K., Getts, R. C. & Hart, R. P. (2004).** Evaluation of sense-strand mRNA amplification by comparative quantitative PCR. *BMC Genomics* **5**, 76.

**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. *Molecular Plant-Microbe Interactions* **4**, 571-578.

**Gonzalez-Pasayo, R. & Martinez-Romero, E. (2000).** Multiresistance genes of *Rhizobium etli* CFN42. *Molecular Plant-Microbe Interactions* **13**, 572-577.

**Gonzalez, J. E. & Marketon, M. M. (2003).** Quorum sensing in nitrogen-fixing rhizobia. *Microbiology and Molecular Biology Reviews* **67**, 574-592.

**González, V., R. I. Santamaría, P. Bustos & other authors (2006).** The partitioned *Rhizobium etli* genome: Genetic and metabolic redundancy in seven interacting replicons. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 3834-3839.

**Gordon, A. J., Ryle, G. J. A., Mitchell, D. F. & Powell, C. E. (1985).** The flux of $^{14}C$-labeled photosynthate through soyabean root nodules during $N_2$ fixation. *Journal of Experimental Botany* **36**, 756-769.

**Gordon, D. M., Ryder, M. H., Heinrich, K. & Murphy, P. J. (1996).** An experimental test of the rhizopine concept in *Rhizobium meliloti*. *Applied and Environmental Microbiology* **62**, 3991-3996.

**Gotz, R., Limmer, N., Ober, K. & Schmitt, R. (1982).** Motility and chemotaxis in 2 strains of *Rhizobium* with complex flagella. *Journal of General Microbiology* **128**, 789-798.

**Graham, P. H., Sadowsky, M. J., Keyser, H. H. & other authors (1991).** Proposed minimal standards for the description of new genera and species of root-nodulating and stem-nodulating bacteria. *International Journal Of Systematic Bacteriology* **41**, 582-587.

**Griffiths, R. I., Manefield, M., Ostle, N., McNamara, N., O'Donnell, A. G., Bailey, M. J. & Whiteley, A. S. (2004).** $^{13}CO_2$ pulse labelling of plants in tandem with stable isotope probing: methodological considerations for examining microbial function in the rhizosphere. *Journal of Microbiology Methods* **58**, 119-129.

**Guntli, D., Heeb, M., MoenneLoccoz, Y. & Defago, G. (1999).** Contribution of calystegine catabolic plasmid to competitive colonization of the rhizosphere of calystegine-producing plants by *Sinorhizobium meliloti* Rm41. *Molecular Ecology* **8**, 855-863.

**Halbleib, C. M. & Ludden, P. W. (2000).** Regulation of biological nitrogen fixation. *Journal of Nutrition* **130**, 1081-1084.

**Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. & Goodman, R. M. (1998).** Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology* **5**, R245-R249.

**Handfield, M. & Levesque, R. C. (1999).** Strategies for isolation of *in vivo* expressed genes from bacteria. *FEMS Microbiology Reviews* **23**, 69-91.

**Harms, G. S., Cognet, L., Lommerse, P. H., Blab, G. A. & Schmidt, T. (2001).** Autofluorescent proteins in single-molecule research: applications to live cell imaging microscopy. *Biophysics Journal* **80**, 2396-2408.

**Harwood, C. S. & Parales, R. E. (1996).** The Beta-ketoadipate pathway and the biology of delf-identity. *Annual Review of Microbiology* **50**, 553-590.

**Hautefort, I. & Hinton, J. C. (2000).** Measurement of bacterial gene expression *in vivo*. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences.* **355**, 601-611.

**Hegde, P. S., White, I. R. & Debouck, C. (2003).** Interplay of transcriptomics and proteomics. *Current Opinon of Biotechnology* **14**, 647-651.

**Heinrich, K., Gordon, D. M., Ryder, M. H. & Murphy, P. J. (1999).** A rhizopine strain of *Sinorhizobium meliloti* remains at a competitive nodulation advantage after an extended period in the soil. *Soil Biology & Biochemistry* **31**, 1063-1065.

**Heinz, E. B., Phillips, D. A. & Streit, W. R. (1999).** BioS, a biotin-induced, stationary-phase, and possible LysR-type regulator in *Sinorhizobium meliloti*. *Molecular Plant-Microbe Interactions* **12**, 803-812.

**Hennecke, H. (1993).** The role of respiration in symbiotic nitrogen fixation. *In New horizons in nitrogen fixation. R. Palacio, Mora. J. & W. E. Newton(Eds).Kluwer Academic Publishers, Dortrecht, Netherlands.* p55-64

**Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D., Dalton, E. & Holden, D. W. (1995).** Simultaneous identification of bacterial virulence genes by negative selection. *Science* **269**, 400-403.

**Hiltner, L. (1904).** Über neuere Erfahrungen und probleme auf dem gebiete der Bodenbakteriologie unter besonderer Berücksichtigung der gründüngung und Brache. *Arbeiten der Deitschen Landwirtschaftlichen Gesellschaft* **98** 59–78.

**Hirsch, A. M., Bauer, W. D., Bird, D. M., Cullimore, J., Tyler, B. & Yoder, J. I. (2003).** Molecular signals and receptors: Controlling rhizosphere interactions between plants and other organisms. *Ecology* **84**, 858-868.

**Hopper, W. & Mahadevan, A. (1991).** Utilization of catechin and its metabolites by *Bradyrhizobium japonicum*. *Applied Microbiology and Biotechnology* **35**, 411-415.

**Hosie, A. H. F., Allaway, D. & Poole, P. S. (2002).** A monocarboxylate permease of *Rhizobium leguminosarum* is the first member of a new subfamily of transporters. *Journal of Bacteriology* **184**, 5436-5448.

**Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., Cuche, B. A., de Castro, E., Lachaize, C., Langendijk-Genevaux, P. S. & Sigrist, C. J. (2008).** The 20 years of PROSITE. *Nucleic Acids Res* **36**, D245-249.

**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.

**Ichige, A. & Walker, G. C. (1997).** Genetic analysis of the *Rhizobium meliloti bacA* gene: functional interchangeability with the *Escherichia coli sbmA* gene and phenotypes of mutants. *Journal of Bacteriology* **179**, 209-216.

**Ivashina, T. V., Sadykov, M. R., Chatuev, B. M., Kanapin, A. A., Shlyapnikov, M. G. & Ksenzenko, V. N. (1996).** *Rhizobium leguminosarum* by *viciae pssB* gene product involved in the symbiotic nitrogen fixation. *Doklady Akademii Nauk* **350**, 712-715.

**Janczarek, M. & Skorupska, A. (2001).** The *Rhizobium leguminosarum* bv. *trifolii* pssB gene product is an inositol monophosphatase that influences exopolysaccharide synthesis. *Archives of Microbiology* **175**, 143-151.

**Jebbar, M., Sohn-Bosser, L., Bremer, E., Bernard, T. & Blanco, C. (2005).** Ectoine-induced proteins in *Sinorhizobium meliloti* include an ectoine ABC-type transporter involved in osmoprotection and ectoine catabolism. *Journal of Bacteriology* **187**, 1293-1304.

**Jensen, J. B., Ampomah, O. Y., Darrah, R., Peters, N. K. & Bhuvaneswari, T. V. (2005).** Role of trehalose transport and utilization in *Sinorhizobium meliloti* - Alfalfa interactions. *Molecular Plant-Microbe Interactions* **18**, 694-702.

**Jiménez-Zurdo, J. I., Van Dillewijn, P., Soto, M. J., de Felipe, M. R., Olivares, J. & Toro, N. (1995).** Characterization of a *Rhizobium meliloti* proline dehydrogenase mutant altered in nodulation efficiency and competitiveness on alfalfa roots. *Molecular Plant-Microbe Interactions* **8**, 492-498.

**Jiménez-Zurdo, J. I., Garcia-Rodríguez, F. M. & Toro, N. (1997).** The *Rhizobium meliloti putA* gene: its role in the establishment of the symbiotic interaction with alfalfa. *Molecular Microbiology* **23**, 85-93.

**Johnston, A. W. B. & Beringer, J. E. (1975).** Identification of the *Rhizobium* strains in pea root nodules using genetic markers. *Journal of General Microbiology* **87**, 343-350.

**Kamilova, F., Kravchenko, L. V., Shaposhnikov, A. I., Azarova, T., Makarova, N. & Lugtenberg, B. (2006).** Organic acids, sugars, and L-tryptophane in exudates of vegetables growing on stonewool and their effects on activities of rhizosphere bacteria. *Mol Plant Microbe Interact* **19**, 250-256.

**Kanehisa, M., Araki, M., Goto, S. & other authors (2008).** KEGG for linking genomes to life and the environment. *Nucleic Acids Research* **36**, D480-484.

**Kaneko, T., Nakamura, Y., Sato, S. & other authors (2000).** Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Research* **7**, 331-338.

**Kaneko, T., Nakamura, Y., Sato, S. & other authors (2002).** Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Research* **9**, 189-197.

**Karsten, S. L., Van Deerlin, V. M., Sabatti, C., Gill, L. H. & Geschwind, D. H. (2002).** An evaluation of tyramide signal amplification and archived fixed and frozen tissue in microarray gene expression analysis. *Nucleic Acids Research* **30**, E4.

**Karunakaran, R., Mauchline, T. H., Hosie, A. H. F. & Poole, P. S. (2005).** A family of promoter probe vectors incorporating autofluorescent and chromogenic reporter proteins for studying gene expression in Gram-negative bacteria. *Microbiology* **151**, 3249-3256.

**Karunakaran, R., Ebert, K., Harvey, S., Leonard, M. E., Ramachandran, V. & Poole, P. S. (2006).** Thiamine is synthesized by a salvage pathway in *Rhizobium leguminosarum* bv. *viciae* strain 3841. *Journal of Bacteriology* **188**, 6661-6668.

**Keithly, J. H. & Nadler, K. D. (1983).** Protoporphyrin formation in *Rhizobium japonicum*. *Journal of Bacteriology* **154**, 838-845.

**Kertesz, M. A. (2000).** Riding the sulfur cycle metabolism of sulfonates and sulfate esters in gram-negative bacteria. *FEMS Microbiol Review* **24**, 135-175.

**Kiely, P. D., Haynes, J. M., Higgins, C. H., Franks, A., Mark, G. L., Morrissey, J. P. & O'Gara, F. (2006).** Exploiting new systems-based strategies to elucidate plant-bacterial interactions in the rhizosphere. *Microbial Ecology* **51**, 257-266.

**Kim, J. & Rees, D. C. (1992).** Crystallographic structure and functional implications of the nitrogenase molybdenum-iron protein from *Azotobacter vinelandii*. *Nature* **360**, 553 – 560.

**Klein, F., Samorski, R., Klug, G. & Evguenieva-Hackenberg, E. (2002).** Atypical processing in domain III of 23S rRNA of *Rhizobium leguminosarum* ATCC 10004(T) at a position homologous to an rRNA fragmentation site in protozoa. *Journal of Bacteriology* **184**, 3176-3185.

**Klug, R. M. & Benning, C. (2001).** Two enzymes of diacylglyceryl-O-4'-(N,N,N,-trimethyl) homoserine biosynthesis are encoded by btaA and btaB in the purple bacterium Rhodobacter sphaeroides. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5910-5915.

**Knee, E. M., Gong, F. C., Gao, M. S., 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. *Molecular Plant-Microbe Interactions* **14**, 775-784.

**Kouchi, H., Akao, S. & Yoneyama, T. (1986).** Respiratory utilization of 13C-labelled photosynthate in nodulated root systems of soybean plants. *Journal of Experimental Botany* **37**, 985-993.

**Kragelund, L., Christoffersen, B., Nybroe, O. & Debruijn, F. J. (1995).** Isolation of lux reporter gene fusions in *Pseudomonas fluorescens* df57 inducible by nitrogen or phosphorus starvation. *FEMS Microbiology Ecology* **17**, 95-106.

**Kragelund, L., Hosbond, C. & Nybroe, O. (1997).** Distribution of metabolic activity and phosphate starvation response of lux-tagged *Pseudomonas fluorescens* reporter bacteria in the barley rhizosphere. *Applied and Environmental Microbiology* **63**, 4920-4928.

**Krehenbrink, M. & Downie, J. A. (2008).** Identification of protein secretion systems and novel secreted proteins in *Rhizobium leguminosarum* bv. *viciae*. *BMC Genomics* **9**, 55.

**Król, J. E., Mazur, A., Marczak, M. & Skorupska, A. (2007).** Syntenic arrangements of the surface polysaccharide biosynthesis genes in *Rhizobium leguminosarum*. *Genomics* **89**, 237-247.

**Kuske, C. R., Ticknor, L. O., Miller, M. E., Dunbar, J. M., Davis, J. A., Barns, S. M. & Belnap, J. (2002).** Comparison of soil bacterial communities in rhizospheres of three plant species and the interspaces in an arid grassland. *Applied and Environmental Microbiology* **68**, 1854-1863.

**Labes, M., Pühler, A. & Simon, R. (1990).** A new family of RSF1010-derived expression and *lac*-fusion broad-host-range vectors for gram-negative bacteria. *Gene* **89**, 37-46.

**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. *Natural Toxins* **1**, 246-249.

**Lambert, A., Osteras, M., Mandon, K., Poggi, M. C. & Le Rudulier, D. (2001).** Fructose uptake in *Sinorhizobium meliloti* is mediated by a high-affinity ATP-binding cassette transport system. *Journal of Bacteriology* **183**, 4709-4717.

**Larrainzar, E., O'Gara, F. & Morrissey, J. P. (2005).** Applications of autofluorescent proteins for in situ studies in microbial ecology. *Annual Review of Microbiology* **59**, 257-277.

**Lee, S. H. & Camilli, A. (2000).** Novel approaches to monitor bacterial gene expression in infected tissue. *Current Opinion In Microbiology* **3**, 97-101.

**LeMessurier, K. S., Ogunniyi, A. D. & Paton, J. C. (2006).** Differential expression of key pneumococcal virulence genes in vivo. *Microbiology* **152**, 305-311.

**Lerouge, P., Roche, P., Faucher, C., Maillet, F., Truchet, G., Prome, J. C., Denarie, J. & CNRS (1990).** Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. *Nature* **1990. 344**, 781-784.

**Limpens, E. & Bisseling, T. (2003).** Signaling in symbiosis. *Current Opinion in Plant Biology* **6**, 343-350.

**Limpens, E., Franken, C., Smit, P., Willemse, J., Bisseling, T. & Geurts, R. (2003).** LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. *Science* **302**, 630-633.

**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. *Molecular Microbiology* **37**, 81-97.

**Lodwig, E. & Poole, P. (2003).** Metabolism of *Rhizobium* bacteroids. *Critical Reviews in Plant Sciences* **22**, 37-78.

**Lodwig, E. M., Hosie, A. H. F., Bourdes, A., Findlay, K., Allaway, D., Karunakaran, R., Downie, J. A. & Poole, P. S. (2003).** Amino-acid cycling drives nitrogen fixation in the legume-*Rhizobium* symbiosis. *Nature* **422**, 722-726.

**Long, S. R. (1989).** *Rhizobium* genetics. *Annual Review of Genetics* **23**, 483-506.

**Long, S. R. (1996).** *Rhizobium* symbiosis - nod factors in perspective. *Plant Cell* **8**, 1885-1898.

**Lopez-Lara, I. M., Sohlenkamp, C. & Geiger, O. (2003).** Membrane lipids in plant-associated bacteria: Their biosyntheses and possible functions. **16**, 567-579.

**Lorite, M. J., Tachil, J., Sanjuan, J., Meyer, O. & Bedmar, E. J. (2000).** Carbon monoxide dehydrogenase activity in *Bradyrhizobium japonicum*. *Applied and Environmental Microbiology* **66**, 1871-1876.

**Lugtenberg, B. J. J., Dekkers, L. & Bloemberg, G. V. (2001).** Molecular determinants of rhizosphere colonization by *Pseudomonas*. *Annual Review of Phytopathology* **39**, 461.

**Lynch, J. M. (1987).** Soil biology – accomplishments and potential. *Soil Science Society of America Journal* **51**, 1409-1412.

**Lynch, J. M. (1990).** Beneficial interactions between micro-organisms and roots. *Biotechnology Advances* **8**, 335-346.

**MacLean, A. A., MacPherson, G., Aneja, P. & Finan, T. M. (2006).** Characterization of the beta-ketoadipate pathway in *Sinorhizobium meliloti*. *Applied and Environmental Microbiology* **72**, 5403-5413.

**MacLean, A. M., Finan, T. M. & Sadowsky, M. J. (2007).** Genomes of the symbiotic nitrogen-fixing bacteria of legumes. *Plant Physiology* **144**, 615-622.

**Mahan, M. J., Slauch, J. M. & Mekalanos, J. J. (1993).** Selection of bacterial virulence genes that are specifically induced in host tissues. *Science* **259**, 686-688.

**Mao, C., Evans, C., Jensen, R. & Sobral, B. (2008).** Identification of new genes in *Sinorhizobium meliloti* using the Genome Sequencer FLX system. *BMC Microbiology* **8**, 72.

**Marchler-Bauer, A., Anderson, J. B., Derbyshire, M. K. & other authors (2007).** CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res* **35**, D237-240.

**Mark, G.L., Dow, J.M., Kiely, P.D., Higgins, H., Haynes, J., Baysse, C., Abbas, A., Foley, T., Franks, A., Morrissey, J., and O'Gara, F. (2005).** Transcriptome profiling of bacterial responses to root exudates identifies genes involved in microbe-plant interactions. *Proceedings of the National Academy of Sciences of the United States of America.* **102**:17454-17459

**Marra, A., Asundi, J., Bartilson, M. & other authors (2002).** Differential fluorescence induction analysis of *Streptococcus pneumoniae* identifies genes involved in pathogenesis. *Infect Immun* **70**, 1422-1433.

**Marschner, P., Yang, C. H. & Lieberei, R. (2001).** Soil and plant specific effects on bacterial community composition in the rhizosphere. *Soil Biology and Chemistry* **33**, 1437-1445.

**Matilla, M. A., Espinosa-Urgel, M., Rodriguez-Herva, J. J., Ramos, J. L. & Ramos-Gonzalez, M. I. (2007).** Genomic analysis reveals the major driving forces of bacterial life in the rhizosphere. *Genome Biology* **8**, R179.

**Mauchline, T. H., Fowler, J. E., East, A. K., Sartor, A. L., Zaheer, R., Hosie, A. H. F., Poole, P. S. & Finan, T. M. (2006).** Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 17933-17938.

**McIntyre, H. J., Davies, H., Hore, T. A., Miller, S. H., Dufour, J. P. & Ronson, C. W. (2007).** Trehalose biosynthesis in *Rhizobium leguminosarum* bv. *trifolii* and its role in desiccation tolerance. *Applied and Environmental Microbiology* **73**, 3984-3992.

**McKay, I., Glenn, A. & Dilworth, M. J. (1985).** Gluconeogenesis in *Rhizobium leguminosarum* MNF3841. *Journal of General Microbiology* **131**, 2067-2073.

**McKay, I. A., Dilworth, M. J. & Glenn, A. R. (1989).** Carbon catabolism in continuous cultures and bacteroids of *Rhizobium leguminosarum* MNF3841. *Archives of Microbiology* **152**, 606-610.

**McRae, D. G., Miller, R. W., Berndt, W. B. & Joy, K. (1989).** Transport of $C_4$-dicarboxylates and amino acids by *Rhizobium meliloti* bacteroids. *Molecular Plant-Microbe Interactions* **2**, 273-278.

**Michiels, J., Verreth, C. & Vanderleyden, J. (1998).** Sequence analysis of the *Rhizobium etli* ribose kinase gene *rbsK* and its phylogenetic position. *DNA Sequence* **9**, 317-321.

**Milcamps, A. & de Bruijn, F. J. (1999).** Identification of a novel nutrient-deprivation-induced *Sinorhizobium meliloti* gene (*hmgA*) involved in the degradation of tyrosine *Microbiology* **145**, 935-947.

**Milcamps, A., Struffi, P. & de Bruijn, F. J. (2001).** The *Sinorhizobium meliloti* nutrient-deprivation-induced tyrosine degradation gene *hmgA* is controlled by a novel member of the *arsR* family of regulatory genes. *Applied and Environmental Microbiology* **67**, 2641-2648.

**Miller, J. H. (1972).** *Experiments in molecular genetics*. New York: Cold Spring Harbor.

**Miller, W. G., Leveau, J. H. & Lindow, S. E. (2000).** Improved *gfp* and *inaZ* broad-host-range promoter probe vectors. *Molecular Plant-Microbe Interactions* **13**, 1243-1250.

**Mirete, S., de Figueras, C. G. & Gonzalez-Pastor, J. E. (2007).** Novel Nickel Resistance Genes from the Rhizosphere Metagenome of Plants Adapted to Acid Mine Drainage. *Applied and environment microbiology* **73**, 6001-6011.

**Mongkolsuk, S., Praituan, W., Loprasert, S., Fuangthong, M. & Chamnongpol, S. (1998).** Identification and characterization of a new organic hydroperoxide resistance (*ohr*) gene with a novel pattern of oxidative stress regulation from *Xanthomonas campestris* pv. *phaseoli. Journal of Bacteriology* **180**, 2636-2643.

**Morgan, J. A., Bending, G. D. & White, P. J. (2005).** Biological costs and benefits to plant-microbe interactions in the rhizosphere. *Journal of Experimental Botany* **56**, 1729-1739.

**Morris, A. C. & Djordjevic, M. A. (2001).** Proteome analysis of cultivar-specific interactions between *Rhizobium leguminosarum* biovar *trifolii* and subterranean clover cultivar Woogenellup. *Electrophoresis* **22**, 586-598.

**Morris, A. C. & Djordjevic, M. A. (2006).** The *Rhizobium leguminosarum* biovar *trifolii* ANU794 induces novel developmental responses on the subterranean clover cultivar woogenellup. *Molecular Plant-Microbe Interactions* **19**, 471-479.

**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.

**Mulder, N. J., Apweiler, R., Attwood, T. K. & other authors (2007).** New developments in the InterPro database. *Nucleic Acids Res* **35**, D224-228.

**Munchbach, M., Nocker, A. & Narberhaus, F. (1999).** Multiple small heat shock proteins in rhizobia. *Journal of Bacteriology* **181**, 83-90.

**Muyzer, G. (1999).** DGGE/TGGE a method for identifying genes from natural ecosystems. *Current Opinion in Microbiology* **2**, 317-322.

**Nakamura, K., Kawabata, T., Yura, K. & Go, N. (2003).** Novel types of two-domain multi-copper oxidases: possible missing links in the evolution. *FEBS Letters* **553**, 239-244.

**Newton, W. E. (2000).** Nitrogen fixation in perspective. In *Nitrogen fixation: from molecules to crop productivity*., *F. O. Pedrosa, M. Hungria, M. G. Yates & W. E. Newton(Eds)Kluwer Publishers. Dortrecht, Netherlands* p. 3-8

**Nygaard, V. & Hovig, E. (2006).** Options available for profiling small samples: a review of sample amplification technology when combined with microarray profiling. *Nucleic Acids Research* **34**, 996-1014.

**Ogawa, J., Soong, C. L., Honda, M. & Shimizu, S. (1996).** Novel metabolic transformation pathway for cyclic imides in *Blastobacter* sp. Strain A17p-4. *Applied and Environmental Microbiology* **62**, 3814-3817.

**Oh, J.-I. & Bowien, B. (1998).** Structural analysis of the *fds* operon encoding the NAD+-linked formate dehydrogenase of *Ralstonia eutropha*. *Journal of Biological Chemistry* **273**, 26349-26360.

**Oke, V. & Long, S. R. (1999).** Bacterial genes induced within the nodule during the *Rhizobium*-legume symbiosis. *Molecular Microbiology* **32**, 837-849.

**Oresnik, I. J., Pacarynuk, L. A., Obrien, S. A. P., Yost, C. K. & Hynes, M. F. (1998).** Plasmid-encoded catabolic genes in *Rhizobium leguminosarum* bv. *trifolii*: Evidence for a plant-inducible rhamnose locus involved in competition for nodulation. *Molecular Plant-Microbe Interactions* **11**, 1175-1185.

**Osbourn, A. F., Barber, C. E. & Daniels, M. J. (1987).** Identification of Plant-induced genes of the bacterial pathogen *Xanthomonas campestris* pathovar *campestris* using a promoter-probe plasmid. *EMBO Journal* **6**, 23-28.

**Osteras, M., Driscoll, B. T. & Finan, T. M. (1995).** Molecular and expression analysis of the *Rhizobium meliloti* phosphoenolpyruvate carboxykinase (*pckA*) gene. *Journal of Bacteriology* **177**, 1452-1460.

**Osteras, M., Boncompagni, E., Vincent, N., Poggi, M. C. & LeRudulier, D. (1998).** Presence of a gene encoding choline sulfatase in *Sinorhizobium meliloti* bet operon: Choline-O-sulfate is metabolized into glycine betaine. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 11394-11399.

**Parke, D., Rivelli, M. & Ornston, L. N. (1985).** Chemotaxis to aromatic and hydroaromatic acids - comparison of *Bradyrhizobium japonicum* and *Rhizobium trifolii*. *Journal of Bacteriology* **163**, 417-422.

**Parke, D. & Ornston, L. N. (1986).** Enzymes of the beta-ketoadipate pathway are inducible in *Rhizobium* and *Agrobacterium* spp and constitutive in *Bradyrhizobium* spp. *Journal of Bacteriology* **165**, 288-292.

**Parke, D., Rynne, F. & Glenn, A. (1991).** Regulation of Phenolic Catabolism in *Rhizobium leguminosarum* biovar *trifolii*. *Journal of Bacteriology* **173**, 5546-5550.

**Parkinson (1994).** Filamentous fungi. *Methods of Soil analysis, part 2 Microbiological and Biochemical properties Soil Science of America, Madison, Wisconsin*, p330-350.

**Patschkowski, T., Schluter, A. & Priefer, U. B. (1996).** *Rhizobium leguminosarum* bv *viciae* contains a 2nd *fnr/fixK*-like gene and an unusual *fixl* homolog. *Molecular Microbiology* **21**, 267-280.

**Perret, X., Staehelin, C. & Broughton, W. J. (2000).** Molecular basis of symbiotic promiscuity. *Microbiology and Molecular Biology Reviews* **64**, 180-+.

**Perry, R. D. (1999).** Signature-tagged mutagenesis and the hunt for virulence factors. *Trends in Microbiology* **7**, 385-389.

**Pfaffl, M. W., Horgan, G. W. & Dempfle, L. (2002).** Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res* **30**, e36.

**Pickering, B. S. & Oresnik, I. J. (2008).** Formate Dependent Autotrophic Growth in Sinorhizobium meliloti. *Journal of Bacteriology*, JB.00757-08.

**Pobigaylo, N., Wetter, D., Szymczak, S., Schiller, U., Kurtz, S., Meyer, F., Nattkemper, T. W. & Becker, A. (2006).** Construction of a large signature-tagged Mini-Tn5 transposon library and its application to mutagenesis of *Sinorhizobium meliloti*. *Applied and Environmental Microbiology* **72**, 4329-4337.

**Pobigaylo, N., Szymczak, S., Nattkemper, T. W. & Becker, A. (2008).** Identification of genes relevant to symbiosis and competitiveness in *Sinorhizobium meliloti* using signature-tagged mutants. *Molecular Plant- Microbe Interaction* **21**, 219-231.

**Poole, P. S., Schofield, N. A., Reid, C. J., Drew, E. M. & Walshaw, D. L. (1994).** Identification of chromosomal genes located downstream of *dctD* that affect the requirement for calcium and the lipopolysaccharide layer of *Rhizobium leguminosarum*. *Microbiology* **140**, 2797-2809.

**Poysti, N. J., Loewen, E. D. M., Wang, Z. & Oresnik, I. J. (2007).** *Sinorhizobium meliloti* pSymB carries genes necessary for arabinose transport and catabolism. *Microbiology* **153**, 727-736.

**Prell, J., Boesten, B., Poole, P. & Priefer, U. B. (2002).** The *Rhizobium leguminosarum* bv. *viciae* VF39 gamma aminobutyrate (GABA) aminotransferase gene (*gabT*) is induced by GABA and highly expressed in bacteroids. *Microbiology* **148**, 615-623.

**Prosser, J. I., Killham, K., Glover, L. A. & Rattray, E. A. S. (1996).** Luminescence-based systems for detection of bacteria in the environment. *Critical Reviews in Biotechnology* **16**, 157-183.

**Puhler, A., Arlat, M., Becker, A., Gottfert, M., Morrissey, J. P. & O'Gara, F. (2004).** What can bacterial genome research teach us about bacteria- plant interactions? *Current Opinion in Plant Biology* **7**, 137-147.

**Radajewski, S., Ineson, P., Parekh, N. R. & Murrell, J. C. (2000).** Stable-isotope probing as a tool in microbial ecology. *Nature* **403**, 646-649.

**Radutoiu, S., Madsen, L. H., Madsen, E. B. & other authors (2003).** Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* **425**, 585-592.

**Rainey, P. B. (1999).** Adaptation of *Pseudomonas fluorescens* to the plant rhizosphere. *Environmental Microbiology* **1**, 243-257.

**Rainey, P. B. & Preston, G. M. (2000).** *In vivo* expression technology strategies: valuable tools for biotechnology. *Current Opinion in Biotechnology* **11**, 440-444.

**Ramos-Gonzalez, M. I., Campos, M. J. & Ramos, J. L. (2005).** Analysis of *Pseudomonas putida* KT2440 gene expression in the maize rhizosphere: *In vitro* expression technology capture and identification of root-activated promoters. *Journal of Bacteriology* **187**, 4033-4041.

**Rao, J. R., Sharma, N. D., Hamilton, J. T. G., Boyd, D. R. & Cooper, J. E. (1991).** Biotransformation of the pentahydroxy flavone quercetin by *Rhizobium loti* and *Bradyrhizobium* Strains (Lotus). *Applied and Environmental Microbiology* **57**, 1563-1565.

**Rao, J. R. & Cooper, J. E. (1994).** Rhizobia catabolize nod gene-inducing flavonoids via C-ring fission mechanisms. *Journal of Bacteriology* **176**, 5409-5413.

**Rebouche, C. J. & Seim, H. (1998).** Carnitine metabolism and its regulation in microorganisms and mammals. *Annual Review of Nutrition* **18**, 39-61.

**Rediers, H., Rainey, P. B., Vanderleyden, J. & De Mot, R. (2005).** Unraveling the secret lives of bacteria: Use of *in vivo* expression technology and differential fluorescence induction promoter traps as tools for exploring niche-specific gene expression. *Microbiology Molecular Biology Review* **69**, 217-261.

**Rees, D. C., Howard, J.B. (2000).** Nitrogenase: standing at the crossroads. *Current opinion in Chemical Biology* **4**, 559-566.

**Reeve, W. G., Tiwari, R. P., Worsley, P. S., Dilworth, M. J., Glenn, A. R. & Howieson, J. G. (1999).** Constructs for insertional mutagenesis, transcriptional signal localization and gene regulation studies in root nodule and other bacteria. *Microbiology* **145**, 1307-1316.

**Reeve, W. G., Tiwari, R. P., Kale, N. B., Dilworth, M. J. & Glenn, A. R. (2002).** ActP controls copper homeostasis in *Rhizobium leguminosarum* bv. *viciae* and *Sinorhizobium meliloti* preventing low pH-induced copper toxicity. *Molecular Microbiology* **43**, 981-991.

**Resendis-Antonio, O., Reed, J. L., EncarnaciÃ³n, S., Collado-Vides, J. & Palsson, B. Ã. (2007).** Metabolic reconstruction and modeling of nitrogen fixation in *Rhizobium etli*. *PLoS Computational Biology* **3**, e192.

**Rhee, S.-K., Liu, X., Wu, L., Chong, S. C., Wan, X. & Zhou, J. (2004).** Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Applied and Environmental Microbiology* **70**, 4303-4317.

**Richardson, J. S., Hynes, M. F. & Oresnik, I. J. (2004).** A genetic locus necessary for rhamnose uptake and catabolism in *Rhizobium leguminosarum* bv. *trifolii*. *Journal of Bacteriology* **186**, 8433-8442.

**Riekhof, W., Andre, C. & Benning, C. (2005).** Two enzymes, BtaA and BtaB, are sufficient for betaine lipid biosynthesis in bacteria. *Archives of biochemistry and biophysics* **441**, 96-105.

**Riley, M. (1993).** Functions of the gene products of *Escherichia coli*. *Microbiology Molecular Biology Review* **57**, 862-952.

**Pinton, R., Varanini, Z., & Nannipieri, P. (2001).** The rhizosphere as a site of biochemical interactions among soil components, plants and microorganisms. *In Rhizosphere : Biochemistry and Organic Substances at the Soil-plant Interface.Pinton, R, Vananini, Z & Nannipieri (Eds) Marcel Dekker Publishers, New York* p1-18.

**Roberts, D. P., Dery, P. D., Yucel, I., Buyer, J., Holtman, M. A. & Kobayashi, D. Y. (1999).** Role of *pfkA* and general carbohydrate catabolism in seed colonization by *Enterobacter cloacae*. *Applied and Environmental Microbiology* **65**, 2513-2519.

**Roberts, D. P., Dery, P. D., Yucel, I. & Buyer, J. S. (2000).** Importance of *pfkA* for rapid growth of *Enterobacter cloacae* during colonization of crop seeds. *Applied and Environmental Microbiology* **66**, 87-91.

**Robertson, J. G., Lyttleton, P., Bullivant, S. & Grayston, G. F. (1978).** Membranes in lupin root nodules. I. The role of Golgi bodies in the biogenesis of infection threads and peribacteroid membranes. *Journal of Cell Science* **30**, 129-149.

**Robleto, E. A., Scupham, A. J. & Triplett, E. W. (1997).** Trifolitoxin production in *Rhizobium etli* strain CE3 increases competitiveness for rhizosphere colonization and root nodulation of *Phaseolus vulgaris* in soil. *Molecular Plant-Microbe Interactions* **10**, 228-233.

**Robleto, E. A., Kmiecik, K., Oplinger, E. S., Nienhuis, J. & Triplett, E. W. (1998).** Trifolitoxin production increases nodulation competitiveness of *Rhizobium etli* CE3 under agricultural conditions. *Applied and Environmental Microbiology* **64**, 2630-2633.

**Roche, P., Maillet, F., Plazanet, C., Debelle, F., Ferro, M., Truchet, G., Prome, J. C. & Denarie, J. (1996).** The common *nodABC* genes of *Rhizobium meliloti* are host-range determinants. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 15305-15310.

**Rohrig, H., Schmidt, J., Wieneke, U., Kondorosi, E., Barlier, I., Schell, J. & John, M. (1994).** Biosynthesis of lipooligosaccharide nodulation factors - *Rhizobium-nodA* protein is involved in n-acylation of the chitooligosaccharide backbone. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 3122-3126.

**Ronson, C. W., Lyttleton, P. & Robertson, J. G. (1981).** $C_4$-dicarboxylate transport mutants of *Rhizobium trifolii* form ineffective nodules on *Trifolium repens*. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 4284-4288.

**Rosenblueth, M., Hynes, M. F. & MartinezRomero, E. (1998).** *Rhizobium tropici teu* genes involved in specific uptake of *Phaseolus vulgaris* bean exudate compounds. *Molecular & General Genetics* **258**, 587-598.

**Rosendahl, L., Vance, C. P. & Pedersen, W. B. (1990).** Products of dark $CO_2$ fixation in pea root nodules support bacteroid metabolism. *Plant Physiology* **93**, 12-19.

**Rosendahl, L., Dilworth, M. J. & Glenn, A. R. (1992).** Exchange of metabolites across the peribacteroid membrane in pea root nodules. *Journal of Plant Physiology* **139**, 635-638.

**Ruberg, S., Tian, Z. X., Krol, E., Linke, B., Meyer, F., Wang, Y., Puhler, A., Weidner, S. & Becker, A. (2003).** Construction and validation of a *Sinorhizobium meliloti* whole genome DNA microarray: genome-wide profiling of osmoadaptive gene expression. *J Biotechnol* **106**, 255-268.

**Ruvkun, G. B., Long, S. R., Meade, H. M. & Ausubel, F. M. (1980).** Molecular-Genetics of Symbiotic Nitrogen-Fixation. *Cold Spring Harbor Symposia On Quantitative Biology* **45**, 492-499.

**Ruvkun, G. B., Sundaresan, V. & Ausubel, F. M. (1982).** Directed transposon Tn5 mutagenesis and complementation analysis of *Rhizobium meliloti* symbiotic nitrogen-fixation genes. *Cell* **29**, 551-559.

**Saalbach, G., Erik, P. & Wienkoop, S. (2002).** Characterisation by proteomics of peribacteroid space and peribacteroid membrane preparations from pea (*Pisum sativum*) symbiosomes. *Proteomics* **2**, 325-337.

**Saeed, A. I., Sharov, V., White, J. & other authors (2003).** TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**, 374-378.

**Saenz, H. L. & Dehio, C. (2005).** Signature-tagged mutagenesis: technical advances in a negative selection method for virulence gene identification. *Curr Opin Microbiol* **8**, 612-619.

**Saleh-Lakha, S., Miller, M., Campbell, R. G., Schneider, K., Elahimanesh, P., Hart, M. M. & Trevors, J. T. (2005).** Microbial gene expression in soil: methods, applications and challenges. *J Microbiol Methods* **63**, 1-19.

**Salomone, J. Y., Szegedi, E., Cobanov, P. & Otten, L. (1998).** Tartrate Utilization Genes Promote Growth of Agrobacterium spp. on Grapevine. *Molecular Plant-Microbe Interactions* **11**, 836-838.

**Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989).** *Molecular Cloning: A Laboratory Manual*, 2 edn. New York: Cold Spring Harbor Laboratory Press.

**Santiago, B., Schübel, U., Egelseer, C. & Meyer, O. (1999).** Sequence analysis, characterization and CO-specific transcription of the *cox* gene cluster on the megaplasmid pHCG3 of Oligotropha carboxidovorans. *Gene* **236**, 115-124.

**Savka, M. A. & Farrand, S. K. (1997).** Modification of rhizobacterial populations by engineering bacterium utilization of a novel plant-produced resource. *Nature Biotechnology* **15**, 363-368.

**Savka, M. A., Dessaux, Y., Oger, P. & Rossbach, S. (2002).** Engineering bacterial competitiveness and persistence in the phytosphere. *Molecular Plant-Microbe Interactions* **15**, 866-874.

**Schadt, C. W., Liebich, J., Chong, S. C., Gentry, T. J. & He, Z. L. (2005).** Design and use of functional gene microarrays (FGAs) for the characterization of microbial communities. *Methods Microbiology* **34**, 331–368.

**Schafer, A., Tauch, A., Jager, W., Kalinowski, J., Thierbach, G. & Puhler, A. (1994).** Small mobilizable multipurpose 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.

**Schell, M. A. (1993).** Molecular biology of the lysr family of transcriptional regulators. *Annual Review of Microbiology* **47597-626**, -626.

**Schindelin, H., Kisker, C., Schlessman, J. L., Howard, J. B. & Rees, D. C. (1997).** Structure of ADP x AIF4(-)-stabilized nitrogenase complex and its implications for signal transduction. *Nature* **387**, 370-376.

**Schluter, A., Patschkowski, T., Quandt, J., Selinger, L. B., Weidner, S., Kramer, M., Zhou, L. M., Hynes, M. F. & Priefer, U. B. (1997).** Functional and regulatory analysis of the two copies of the *fixNOQP* operon of *Rhizobium leguminosarum* strain VF39. *Molecular Plant-Microbe Interactions* **10**, 605-616.

**Schubel, U., Kraut, M., Morsdorf, G. & Meyer, O. (1995).** Molecular characterization of the gene cluster *coxMSL* encoding the molybdenum-containing carbon monoxide dehydrogenase of *Oligotropha carboxidovorans*. *Journal of Bacteriology* **177**, 2197-2203.

**Seibert, V., Thiel, M., Hinner, I.-S. & Schlomann, M. (2004).** Characterization of a gene cluster encoding the maleylacetate reductase from *Ralstonia eutropha* 335T, an enzyme recruited for growth with 4-fluorobenzoate. *Microbiology* **150**, 463-472.

**Sharma, S., Aneja, M. K., Mayer, J., Schloter, M. & Munch, J. C. (2004).** RNA fingerprinting of microbial community in the rhizosphere soil of grain legumes. *FEMS Microbiology Letters* **240**, 181-186.

**Shaw, L. J., Morris, P. & Hooker, J. E. (2006).** Perception and modification of plant flavonoid signals by rhizosphere microorganisms. *Environmental Microbiology* **8**, 1867-1880.

**Shi, Y.-w., Cui, L.-f. & Yuan, J.-m. (2007).** Gene cloning, expression, and substrate specificity of an imidase from the strain *Pseudomonas putida* YZ-26. *Current Microbiology* **55**, 61-64.

**Shimoda, Y., Mitsui, H., Kamimatsuse, H. & other authors (2008).** Construction of signature-tagged mutant Library in *Mesorhizobium loti* as a powerful tool for functional genomics. *DNA Res*, dsn017.

**Sigaud, S., Becquet, V., Frendo, P., Puppo, A. & Herouart, D. (1999).** Differential regulation of two divergent *Sinorhizobium meliloti* genes for HPII-like catalases during free-living growth and protective role of both catalases during symbiosis. *Journal of Bacteriology* **181**, 2634-2639.

**Silby, M. W. & Levy, S. B. (2004).** Use of *in vivo* expression technology to identify genes important in growth and survival of *Pseudomonas fluorescens* Pf0-1 in soil: Discovery of expressed sequences with novel genetic organization. *Journal of Bacteriology* **186**, 7411-7419.

**Smalla, K., Wieland, G., Buchner, A., Zock, A., Parzy, J., Kaiser, S., Roskot, N., Heuer, H. & Berg, G. (2001).** Bulk and rhizosphere soil bacterial communities studied by denaturing gradient gel electrophoresis: plant-dependent enrichment and seasonal shifts revealed. *Applied and Environmental Microbiology* **67**, 4742-4751.

**Smit, G., Swart, S., Lugtenberg, B. J. & Kijne, J. W. (1992).** Molecular mechanisms of attachment of *Rhizobium* bacteria to plant roots. *Molecular Microbiology* **6**, 2897 - 2903.

**Smyth, G. K. & Speed, T. (2003).** Normalization of cDNA microarray data. *Methods* **31**, 265-273.

**Smyth, G. K. (2004).** Linear models and emprical bayes methods for assessing differential expression in microarray experiements. *Statistical Applications in Genetics and Molecular Biology* **3**, Article3.

**Smyth, G. K. (2005).** Limma: linear models for microarray data. *In: Bioinformatics and Computational Biology Solutions using R and Bioconductor* R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds.), Springer, New York, 397-420.

**Smyth, G. K., Michaud, J. & Scott, H. S. (2005).** Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21**, 2067-2075.

**Soedarjo, M. & Borthakur, D. (1996).** Mimosine produced by the tree-legume *Leucaena* provides growth advantages to some *Rhizobium* strains that utilize it as a source of carbon and nitrogen. *Plant and Soil* **186**, 87-92.

**Soedarjo, M. & Borthakur, D. (1998).** Mimosine, a toxin produced by the tree-legume *Leucaena* provides a nodulation competition advantage to mimosine-degrading *Rhizobium* strains. *Soil Biology & Biochemistry* **30**, 1605-1613.

**Soong, C.-L., Ogawa, J., Sukiman, H., Prana, T., Sri Prana, M. & Shimizu, S. (1998).** Distribution of cyclic imide-transforming activity in microorganisms. *FEMS Microbiology Letters* **158**, 51-55.

**Soto, M. J., Vandillewijn, P., Olivares, J. & Toro, N. (1994).** Ornithine cyclodeaminase activity in *Rhizobium meliloti*. *FEMS Microbiology Letters* **119**, 209-213.

**Sourjik, V., Sterr, W., Platzer, J., Bos, I., Haslbeck, M. & Schmitt, R. (1998).** Mapping of 41 chemotaxis, flagellar and motility genes to a single region of the *Sinorhizobium meliloti* chromosome. *Gene* **223**, 283-290.

**Spaink, H. P., Sheeley, D. M., Vanbrussel, A. A. N. & other authors (1991).** A novel highly unsaturated fatty acid moiety of lipo-oligosaccharide signals determines host specificity of *Rhizobium*. *Nature* **354**, 125-130.

**Sprenger, G. A., Schorken, U., Wiegert, T., Grolle, S., de Graaf, A. A., Taylor, S. V., Begley, T. P., Bringer-Meyer, S. & Sahm, H. (1997).** Identification of a thiamin-dependent synthase in *Escherichia coli* required for the formation of the 1-deoxy-D-xylulose 5-phosphate precursor to isoprenoids, thiamin, and pyridoxol. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 12857-12862.

**Standing, D., Meharg, A. A. & Killham, K. (2003).** A tripartite microbial reporter gene system for real-time assays of soil nutrient status. *FEMS Microbiol Lett* **220**, 35-39.

**Stanfield, S. W., Ielpi, L., O'Brochta, D., Helinski, D. R. & Ditta, G. S. (1988).** The *ndvA* gene product of *Rhizobium meliloti* is required for beta-(1-2) glucan production and has homology to the ATP-binding export protein HlyB. *Journal of Bacteriology* **170**, 3523-3530.

**Stears, R. L., Getts, R. C. & Gullans, S. R. (2000).** A novel, sensitive detection system for high-density microarrays using dendrimer technology. *Physiology Genomics* **3**, 93-99.

**Stiens, M., Schneiker, S., hler, A., Schl & ter, A. (2007).** Sequence analysis of the 181-kb accessory plasmid pSmeSM11b, isolated from a dominant *Sinorhizobium meliloti* strain identified during a long-term field release experiment. *FEMS Microbiology Letters* **271**, 297-309.

**Stowers, M. D. (1985).** Carbon metabolism in *Rhizobium* species. *Annual Review of Microbiology* **39**, 89-108.

**Streeter, J. G. (1981).** Seasonal distribution of carbohydrates in nodules and stem exudate from field-grown soya bean-plants. *Annals of Botany* **48**, 441-450.

**Streit, W. R., Joseph, C. M. & Phillips, D. A. (1996).** Biotin and other water-soluble vitamins are key growth-factors for alfalfa root colonization by *Rhizobium meliloti* 1021. *Molecular Plant-Microbe Interactions* **9**, 330-338.

**Stuurman, N., Bras, C. P., Schlaman, H. R. M., Wijfjes, A. H. M., Bloemberg, G. & Spaink, H. P. (2000).** Use of green fluorescent protein color variants expressed on stable broad-host-range vectors to visualize rhizobia interacting with plants. *Molecular Plant-Microbe Interactions* **13**, 1163-1169.

**Sullivan, J. T., Brown, S. D., Yocum, R. R. & Ronson, C. W. (2001).** The *bio* operon on the acquired symbiosis island of *Mesorhizobium* sp strain R7A includes a novel gene involved in pimeloyl-CoA synthesis. *Microbiology* **147**, 1315-1322.

**Summers, M. L., Botero, L. M., Busse, S. C. & McDermott, T. R. (2000).** The *Sinorhizobium meliloti* Lon protease is involved in regulating exopolysaccharide synthesis and is required for nodulation of alfalfa. *Journal of Bacteriology* **182**, 2551-2558.

**Tate, R., Riccio, A., Iaccarino, M. & Patriarca, E. J. (1997).** A cysG mutant strain of *Rhizobium etli* pleiotropically defective in sulfate and nitrate assimilation. *Journal of Bacteriology* **179**, 7343-7350.

**Tepfer, D., Goldmann, A., Pamboukdjian, N., Maille, M., Lepingle, A., Chevalier, D., Denarie, J. & Rosenberg, C. (1988).** A plasmid of *Rhizobium meliloti* 41 encodes catabolism of 2 compounds from root exudate of *Calystegium sepium*. *Journal of Bacteriology* **170**, 1153-1161.

**Teplitski, M., Robinson, J. B. & Bauer, W. D. (2000).** Plants secrete substances that mimic bacterial N-acyl homoserine lactone signal activities and affect population density-dependent behaviors in associated bacteria. *Molecular Plant-Microbe Interactions* **13**, 637-648.

**Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994).** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673-4680.

**Thonymeyer, L. & Kunzler, P. (1996).** The *Bradyrhizobium japonicum* aconitase gene (*acnA*) is important for free-living growth but not for an effective root-nodule symbiosis. *Journal of Bacteriology* **178**, 6166-6172.

**Todd, J. D., Wexler, M., Sawers, G., Yeoman, K. H., Poole, P. S. & Johnston, A. W. B. (2002).** RirA, an iron-responsive regulator in the symbiotic bacterium *Rhizobium leguminosarum*. *Microbiology* **I148**, 4059-4071.

**Toffanin, A., Cadahia, E., Imperial, J., Ruiz-Argüeso, T. & Palacios, J. (2002).** Characterization of the urease gene cluster from *Rhizobium leguminosarum* bv. *viciae*. *Archives of Microbiology* **177**, 290-298.

**Toegl, A., Kirchner, R., Gauer, C. & Wixforth, A. (2003).** Enhancing results of microarray hybridizations through microagitation. *Journal of Biomolecular Techniques* **14**, 197-204.

**Tom-Petersen, A., Hosbond, C. & Nybroe, O. (2001).** Identification of copper-induced genes in *Pseudomonas fluorescens* and use of a reporter strain to monitor bioavailable copper in soil. *FEMS Microbiology Ecology* **38**, 59-67.

**Triplett, E. W. (1990).** Construction of a symbiotically effective strain of Rhizobium leguminosarum bv. trifolii with increased nodulation competitiveness. *Applied and Environmental Microbiology* **56**, 98-103.

**Tsolis, R. M. (2002).** Comparative genome analysis of the alpha-proteobacteria: Relationships between plant and animal host specificity. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 12503-12505.

**Tyson, G. W., Chapman, J., Hugenholtz, P. & other authors (2004).** Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37-43.

**Tzvetkov, M., Klopprogge, C., Zelder, O. & Liebl, W. (2003).** Genetic dissection of trehalose biosynthesis in *Corynebacterium glutamicum*: inactivation of trehalose production leads to impaired growth and an altered cell wall lipid composition. *Microbiology* **149**, 1659-1673.

**Udvardi, M. K. & Day, D. A. (1997).** Metabolite transport across symbiotic membranes of legume nodules. *Annual Review of Plant Physiology and Plant Molecular Biology* **48**, 493-523.

**Valdivia, R. H. & Falkow, S. (1996).** Bacterial genetics by flow cytometry: rapid isolation of *Salmonella typhimurium* acid-inducible promoters by differential fluoresence induction. *Molecular Microbiology* **22**, 367-378.

**Valdivia, R. H. & Falkow, S. (1997a).** Fluorescence-based isolation of bacterial genes expressed within host cells. *Science* **277**, 2007-2011.

**Valdivia, R. H. & Falkow, S. (1997b).** Probing bacterial gene expression within host cells. *Trends in Microbiology* **5**, 360-363.

**van Brussel, A. A., Recourt, K., Pees, E., Spaink, H. P., Tak, T., Wijffelman, C. A., Kijne, J. W. & Lugtenberg, B. J. (1990).** A biovar-specific signal of *Rhizobium leguminosarum* bv. *viciae* induces increased nodulation gene-inducing activity in root exudate of *Vicia sativa* subsp. *nigra*. *Journal of Bacteriology* **172**, 5394-5401.

**Van Der Ploeg, J., Smidt, M. P., Landa, A. S. & Janssen, D. B. (1994).** Identification of chloroacetaldehyde dehydrogenase involved in 1,2-Dichloroethane degradation. *Applied and Environmental Microbiology* **60**, 1599-1605.

**Van Egeraat, A. W. S. M. (1975).** The possible role of homoserine in the development of *Rhizobium leguminosarum* in the rhizosphere of pea seedlings. *Plant Soil* **42**, 380-387.

**Van Gelder, R. N., von Zastrow, M. E., Yool, A., Dement, W. C., Barchas, J. D. & Eberwine, J. H. (1990).** Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proceedings of the National Academy of Sciences of the United States of America* **87**, 1663-1667.

**van Rhijn, P. & Vanderleyden, J. (1995).** The *Rhizobium*-plant symbiosis. *Microbiological Reviews* **59**, 124-142.

**Veluri, R., Weir, T. L., Bais, H. P., Stermitz, F. R. & Vivanco, J. M. (2004).** Phytotoxic and antimicrobial activities of catechin derivatives. *Journal and Agricultre and Food Chemistry* **52**, 1077-1082.

**Venter, J. C., Remington, K., Heidelberg, J. F. & other authors (2004).** Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**, 66-74.

**von Mering, C., Jensen, L. J., Kuhn, M., Chaffron, S., Doerks, T., Kruger, B., Snel, B. & Bork, P. (2007).** STRING 7 - recent developments in the integration and prediction of protein interactions. *Nucleic Acids Research* **35**, D358-362.

**Walker, T. S., Bais, H. P., Grotewold, E. & Vivanco, J. M. (2003).** Root exudation and rhizosphere biology. *Plant Physiology* **132**, 44-51.

**Walshaw, D. L. & Poole, P. S. (1996).** The general L-amino acid permease of *Rhizobium leguminosarum* is an ABC uptake system that influences efflux of solutes. *Molecular Microbiology* **21**, 1239-1252.

**Wang, C., Zhang, H.-B., Chen, G., Chen, L. & Zhang, L.-H. (2006).** Dual Control of Quorum sensing by two TraM-type antiactivators in *Agrobacterium tumefaciens* octopine Strain A6. *Journal of Bacteriology* **188**, 2435-2445.

**Weir, B. S. (2006).** The current taxonomy of rhizobia. *New Zealand rhizobia website* **http://www.rhizobia.co.nz/taxonomy/rhizobia.html**.

**White, J. P. (2006).** Amino Acid Transport and Metabolism by *Rhizobium leguminosarum. Thesis*. submitted to University of Reading, Reading, UK

**Whitehead, L. F. & Day, D. A. (1997).** The peribacteroid membrane. *Physiologia Plantarum* **100**, 30-44.

**Wielbo, J., Marek-Kozaczuk, M., Kubik-Komar, A. & Skorupska, A. (2007).** Increased metabolic potential of *Rhizobium* spp. is associated with bacterial competitiveness. *Canadian Journal of Microbiology* **53**, 957-967.

**Wienkoop, S. & Saalbach, G. (2003).** Proteome analysis. Novel proteins identified at the peribacteroid membrane from *Lotus japonicus* root nodules. *Plant Physiology* **131**, 1080-1090.

**Wilkins, M. R., Sanchez, J. C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F. & Williams, K. L. (1996).** Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology & genetic engineering reviews* **13**, 19-50.

**Wilson, R. L., Tvinnereim, A. R., Jones, B. D. & Harty, J. T. (2001).** Identification of *Listeria monocytogenes* in vivo-induced genes by fluorescence-activated cell sorting. *Infection and Immunity* **69**, 5016-5024.

**Wong, C. M., Dilworth, M. J. & Glenn, A. R. (1994).** Cloning and sequencing show that 4-hydroxybenzoate hydroxylase (*pobA*) is required for uptake of 4-hydroxybenzoate in *Rhizobium leguminosarum. Microbiology* **140**, 2775-2786.

**Wright, C. A. & Beattie, G. A. (2004).** Bacterial species specificity in *proU* osmoinducibility and *nptII* and *lacZ* expression. **8**, 201-208.

**Xi, C., Schoeters, E., Vanderleyden, J. & Michiels, J. (2000).** Symbiosis-specific expression of *Rhizobium etli casA* encoding a secreted calmodulin-related protein. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 11114-11119.

**Xi, C., Dirix, G., Hofkens, J., De Schryver, F. C., Vanderleyden, J. & Michiels, J. (2001).** Use of dual marker transposons to identify new symbiosis genes in *Rhizobium*. *Microbial Ecology* **41**, 325-332.

**Xu, Y., Mortimer, M. W., Fisher, T. S., Kahn, M. L., Brockman, F. J. & Xun, L. (1997).** Cloning, sequencing, and analysis of a gene cluster from *Chelatobacter heintzii* ATCC 29600 encoding nitrilotriacetate monooxygenase and NADH:flavin mononucleotide oxidoreductase. *Journal of Bacteriology* **179**, 1112-1116.

**Yang, C., Rodionov, D. A., Rodionova, I. A., Li, X. & Osterman, A. L. (2008).** Glycerate 2-kinase of *Thermotoga maritima* and genomic reconstruction of related metabolic pathways. *Journal of Bacteriology* **190**, 1773-1782.

**Yarosh, O. K., Charles, T. C. & Finan, T. M. (1989).** Analysis of $C_4$-dicarboxylate transport genes in *Rhizobium meliloti*. *Molecular Microbiology* **3**, 813-823.

**Yost, C. K., Rath, A. M., Noel, T. C. & Hynes, M. F. (2006).** Characterization of genes involved in erythritol catabolism in *Rhizobium leguminosarum* bv. *viciae. Microbiology* **152**, 2061-2074.

**Young, J. P., Crossman, L., Johnston, A. & other authors (2006).** The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biology* **7**, R34.

**Zakhia, F. & de Lajudie, P. (2001).** Taxonomy of rhizobia. *Agronomie* **21**, 569-576.

**Zhang, X. X. & Rainey, P. B. (2007).** The role of a P1-type ATPase from *Pseudomonas fluorescens* SBW25 in copper homeostasis and plant colonization. *Molecular Plant -Microbe Interaction* **20**, 581-588.