Genome-scale characterisation of symbiotic fitness determinants of *Rhizobium leguminosarum* using INSeq



Rachel Wheatley

University College

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Trinity Term 2018

Abstract

Genome-scale characterisation of symbiotic fitness determinants of *Rhizobium leguminosarum* using INSeq

Rachel Wheatley, University College
Submitted for the degree of DPhil., Trinity Term 2018

Mariner-based insertion sequencing (INSeq) has been used to characterise symbiotic fitness determinants in *Rhizobium leguminosarum* bv. *viciae* 3841 (Rlv3841) on a whole genome-scale. An INSeq database has been constructed that defines the mutational phenotypes for 7316 genes (99.7% of the genome) in Rlv3841 across nine *in vitro* and *in planta* INSeq conditions, including growth on glucose or succinate at both 21% and 1% O_2, growth in the pea rhizosphere, attachment to pea roots and pea nodulation (collaboration). A total of 463 genes were identified to be required for the successful nodulation of pea, and a genetic map was generated identifying the specific stages of symbiosis these genes are required in. There was identification of 55 genes required specifically for growth in the rhizosphere and 101 genes required specifically for root attachment. Analysis of the central carbon metabolism pathways indicated presence of a methylglyoxal pathway in Rlv3841 and suggested that *in planta* bacteroids might synthesise glycogen for carbon storage via gluconeogenesis of the plant-provided TCA-cycle intermediates. Only one gene, RL2393 (*glnB*), was found to be uniquely essential for growth under 1% O_2 with succinate. RL2393 (*glnB*) encodes the nitrogen regulatory protein (PII). Mutation of *glnB* effectively generates a constitutively activated glutamine synthesis response that increases the removal of 2-ketoglutarate from the TCA-cycle to combine with ammonium to form glutamine. This removal of 2-ketoglutarate from the TCA-cycle becomes problematic under low O_2, replenishing 2-ketoglutarate requires an additional turn of the TCA-cycle and the generation of NADH, NADPH and $FADH_2$ reductant molecules whose reoxidation is limited by O_2-availability. Ammonium assimilation similarly relies on the removal of 2-ketoglutarate. A reoxidation block model was proposed, in which the provision of dicarboxylates under low O_2 prevents removal of 2-ketoglutarate from the TCA-cycle and the assimilation of ammonium. This reoxidation block model provides a novel driving force behind ammonium export in symbioses, alongside explaining why legumes always provide TCA-cycle dicarboxylates to fuel N_2-fixation in the low O_2 environment of the nodules.

Acknowledgements

First and foremost, I would like to thank my supervisor Professor Phil Poole for the opportunity to work on this project and for all his guidance, help and support. I would also like to thank my second supervisor, Professor Gail Preston, for her input into my project and support with all the assessments over the course of my DPhil. I am extremely grateful for the mentorship of Dr Vinoy Ramachandran, who has assisted and advised me throughout the entire length of my project and provided an invaluable amount of support. I am also very thankful to Dr Carmen Sanchez Canizares and Dr Barney Geddes for all their help, advice and encouragement.

I would like to thank all members of the Poole group who I have worked with over the last four years, for all the help they have given me, the rich research environment they have provided and the warmth and friendship that is shared within the group. Isabel Webb, Francesco Pini, Marcela Mendoza-Suarez, Andrzej Tkacz, Joshua Roworth, Kyle Grant, Nick Crang, Jack Parsons, Alison East, Bea Jorrin, Khushboo Borah, and everyone else who joined later, thank you! I would also like to thank Helen Prescott, Tom King, Gem Toes, Rob Bryant and all other staff in the Department of Plant Sciences, without whom none of the research would be possible.

I am very grateful to my parents, for always supporting and encouraging me in science and education. I would also like to thank everyone who has made my time at Oxford infinitely more enjoyable over these last four years: to all my friends, the Oxford University Judo Club, the Oxford University Amateur Boxing Club and all my work colleagues at University College, thank you. I would finally like to thank the Natural Environmental Research Council for funding this DPhil project, and the grants I have received from the Microbiology Society, St Hugh's College and University College that have allowed me to attend conferences throughout my project.

Abbreviations

ABC	ATP-binding cassette
ADP	Adenosine diphosphate
Amp	Ampicillin
Amp^r	Ampicillin resistance
ATP	Adenosine triphosphate
Cfu	Colony-forming units
CoA	Coenzyme A
DNA	Deoxyribonucleic acid
Dpi	Days post inoculation
ED	Entner-Doudoroff
EPS	Exopolysaccharide
ES	Essential
FADH	Flavine adanine dinucleotide
$FADH_2$	Dihydroflavine-adenine dinucleotide
GA	Growth-advantaged
gDNA	Genomic DNA
GD	Growth-defective
Gm	Gentamycin
Gm^r	Gentamycin resistance
GOGAT	Glutamate oxoglutarate amidotransferase
GS	Glutamine synthetase
INSeq	Insertion sequencing
Km	Kanamycin

Km^r	Kanamycin resistance
LB	Luria Berani broth
LPS	Lipopolysaccharide
MS	Mass spectrometry
N_2	Nitrogen
NAD^+	Nicotinamide adenine dinucleotide (oxidised form)
NADH	Nicotinamide adenine dinucleotide (reduced form)
NADP	Nicotinamide adenine dinucleotide phosphate
NADPH	Dihydronicotinamide adenine dinucleotide phosphate
NE	Growth-neutral
NH_3	Ammonia
NH_4	Ammonium
Nm	Neomycin
Nm^r	Neomycin resistance
O_2	Oxygen
OD	Optical density
PCR	Polymerase chain reaction
P_i	Inorganic phosphate
PPP	Pentose phosphate pathway
Rlv3841	*Rhizobium leguminosarum* bv. *viciae* 3841
RNA	Ribonucleic acid
SEM	Standard error of the mean
Str	Streptinomycin
Str^r	Streptinomycin resistance

TCA-cycle	Tricarboxylic acid cycle
T_m	Melting temperature
TY	Tryptone yeast
UMS	Universal Minimal Salts
VMM	Vincents minimal media
X-gal	5-bromo-4-chloro-indolyl- β-D-galactoside

Table of contents

Abstract 3

Acknowledgements 5

Abbreviations 7

Publications 17

Introduction

1.1 The Nitrogen Crisis 20

1.1.1 The importance of nitrogen in the biosphere 20

1.1.2 Managing the nitrogen cycle 20

1.2 Nature's nitrogen fixers: a sustainable method of nitrogen fertilisation 24

1.2.1 Biological nitrogen fixation 24

1.2.2 *Rhizobium*-legume symbioses 25

1.3 Global perspectives for biological nitrogen fixation research 29

1.3.1 Enhancing symbiotic nitrogen fixation efficiency 31

1.4 Rhizobial taxonomy 33

1.4.1 *Rhizobium leguminosarum* biovar *viciae* 3841 34

1.5 Lifestyle changes in *Rhizobium*-legume symbioses 37

1.5.1 Growth in the rhizosphere 39

1.5.2 Root attachment and colonisation 40

1.5.3 Infection thread formation 45

1.5.4 Nodule organogenesis and bacteroid formation 46

1.6 Functional genomic approaches for understanding rhizobial genomes 48

1.6.1 Genetic approaches 51

1.6.2 Transcriptomic approaches 53

1.6.3 Proteomic approaches 54

1.6.4 Metabolomic approaches 55

1.7 Insertion sequencing as a method of high-throughput genetic screening 56

1.7.1 *Mariner*-based insertion sequencing 58

1.7.2 Methodology of insertion sequencing studies 59

1.7.3 Statistical models for analysing insertion sequencing data 62

1.8 Research objectives 65

1.8.1 Chapter 3 67

1.8.2 Chapter 4 68

1.8.3 Chapter 5 69

Materials and methods

2.1 Bacterial strains, plasmids and primers 71

2.1.1 Strains 71

2.1.2 Plasmids 75

2.1.3 Primers 77

2.2 Media and antibiotics 89

2.2.1 Media 89

2.2.2 Antibiotics 91

2.3 Growth conditions and assays 93

2.3.1 Growth conditions 93

2.3.2 Generation time of rhizobial strains 93

2.3.3 Plant growth conditions and inoculation 93

2.3.4 Seed sterilisation and germination of *P. sativum* 95

2.3.5 Inoculation onto *P. sativum* seedling roots 95

2.3.6 Rhizosphere and root attachment assays with *P. sativum* 95

2.3.7 Determining cellular NAD/NADH ratio 96

2.3.8 Measuring mean oxygen consumption rate 97

2.4 Molecular techniques 98

2.4.1 Isolation of genomic DNA 98

2.4.2 PCR amplification 98

2.4.3 Gel electrophoresis 99

2.4.4 Restriction digest and DNA ligation 99

2.5 Cloning techniques 100

2.5.1 Transformation 100

2.5.2 Conjugation from *E. coli* to *R. leguminosarum* 100

2.5.3 Mutagenesis by pK19mob-integration 101

2.6 Insertion sequencing 101

2.6.1 *Mariner* library construction 101
2.6.2 Library preparation and sequencing 102
2.6.3 Transposon insertion analysis with a four-state HMM 104
2.6.4 Transposon insertion analysis with a calculation of gene fitness values 104

2.7 RNA sequencing 106
2.7.1 RNA extractions from root attached rhizobia 106
2.7.2 Isolation of mRNA and library preparation for sequencing 107
2.7.3 RNA sequencing analysis 108

2.8 Bioinformatics, data handling and statistical analysis 109

The role of O_2 in the growth of *Rhizobium leguminosarum on glucose and succinate*

3.1 Abstract 111

3.2 Introduction 112

3.3 Results and discussion 115
3.3.1 Mean generation time of Rlv3841 115
3.3.2 INSeq experimental design 116
3.3.3 HMM genome analysis 117
3.3.4 Literature validation of INSeq predictions 118
3.3.5 Experimental validation of INSeq predictions 122
3.3.6 Central Metabolism 124
3.3.6.1 The Entner-Doudoroff pathway is needed for optimal growth on both glucose and succinate 124
3.3.6.2 The TCA-cycle is essential for growth on glucose and succinate 125
3.3.7 Mutations that specifically reduced growth on glucose at both 1% and 21% O_2 128
3.3.8 Mutations that specifically reduced growth on succinate at both 21% and 1% O_2 133
3.3.9 Mutations that reduced growth at 21% O_2 on both glucose and succinate 136
3.3.10 Mutations that reduced growth at 1% O_2 on both glucose and succinate 137
3.3.11 Mutations that specifically reduced growth on succinate at 1% O_2 139
3.3.12 A plasmid specific toxin-antitoxin system on pRL10 140

3.4 Conclusion 141

Genome-scale characterisation of the fitness determinants in the early stage of the *Rhizobium leguminosarum* symbiosis

4.1 Abstract 144

4.2 Introduction 145

4.3 Results and discussion 148

4.3.1 Colonisation assays of Rlv3841 grown in the pea rhizosphere and attached to roots 148

4.3.2 INSeq experimental design 150

4.3.2.1 Rhizosphere and root attachment 150

4.3.2.2 Nodulation and nodule recovery 151

4.3.3 HMM genome analysis 152

4.3.4 Construction of an INSeq database in Rlv3841 153

4.3.5 Core fitness requirements in Rlv3841 154

4.3.5.1 Functional dissection of the core fitness requirements in Rlv3841 156

4.3.6 Classification of fitness determinants 160

4.3.7 Genes required for growth in the rhizosphere 161

4.3.7.1 Functional dissection of genes required for growth in the rhizosphere 162

4.3.8 Genes required for root attachment 166

4.3.8.1 Functional dissection of genes required for attaching to roots 167

4.3.9 Analysis of central carbon metabolism 172

4.3.10 Mutation of genes to improve rhizobial fitness 177

4.3.11 Transcriptional regulators involved in root attachment 179

4.3.11.1 Genomic context of the transcriptional regulators involved in root attachment 182

4.3.12 RNA sequencing to characterise root attachment transcriptional regulatory networks 186

4.3.12.1 RNA sequencing data analysis 187

4.3.12.2 Identification of common factors in a root attachment defective transcriptome 189

4.3.12.3 Putative regulator targets identified through genomic proximity and expression level changes 191

4.3.13 Mutation of a putative LysR family transcriptional regulator on pRL8 prevents root attachment 197

4.4 Conclusion 200

Low O_2 prevents rhizobia from assimilation ammonium and induces amino acid auxotrophy

5.1 Abstract 208

5.2 Introduction 209

5.3 Results 212

5.3.1 RL2392 (*glnB*) is essential for growth on dicarboxylates under low O_2 212

5.3.2 Growth on a dicarboxylic acid and ammonium under low O_2 induces glutamate auxotrophy and retards growth 215

5.3.3 Ammonium assimilation and dicarboxylate metabolism increases cellular reductant ratios and O_2 consumption rate 219

5.3.4 Low O_2 limits ammonium assimilation across rhizobial genera 223

5.4 Discussion 224

5.5 Revised model of symbiosis 230

5.6 Conclusion 232

General discussion

6.1 Overview of results 236

6.2 Evaluation of INSeq as a technique for whole-genome screening applications 240

6.2.1 The Hidden Markov Model as a method of data analysis 241

6.2.2 *Mariner*-based transposon mutagenesis 242

6.2.3 Evaluation of INSeq in this thesis 243

6.2.4 INSeq experiment parameters 244

6.2.5 INSeq analyses of microbial competition 245

6.2.6 Other uses of INSeq 247

6.3 Limitations and future directions 247

6.3.1 Scope 247

6.3.2 INSeq analyses 249

6.3.3 RNA sequencing analyses 252

6.3.4 Reoxidation block model 253

6.4 Concluding remarks 255

Bibliography 257

Appendix A. Supplementary material for chapter 3 and chapter 4 285

Appendix B. Research article publication 363

Appendix C. Review article publication 379

Publications

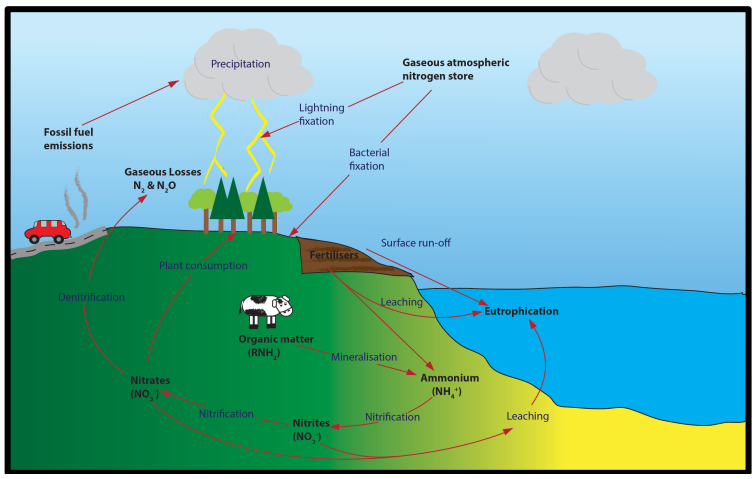
Wheatley, Rachel M., et al. "Role of O_2 in the growth of *Rhizobium leguminosarum* bv. *viciae* 3841 on glucose and succinate." *Journal of bacteriology* 199.1 (2017): e00572-16. (Appendix B)

Wheatley, Rachel M., and Philip S. Poole. "Mechanisms of bacterial attachment to roots." *FEMS microbiology reviews* (2018). (Appendix C)

Wheatley, Rachel M. “A problem for the 21st Century: tackling the Nitrogen Crisis” The Phenotype Journal, Issue 25, Michaelmas Term 2016. (https://issuu.com/phenotypejournal/docs/mt16_main_web)

Chapter 1

Introduction



1.1 The Nitrogen Crisis

1.1.1 The importance of nitrogen in the biosphere

Nitrogen is a building block of life: it is a constituent of nucleic acids, amino acids and ATP. Stable dinitrogen (N_2) forms 78% of the Earth's atmosphere but is biologically unavailable in this form. Insufficient nitrogen availability is a limiting factor on crop growth. Since the dawn of agriculture, increasing crop yields has been a major drive, and a necessity to feed the constantly growing world-population. Two hundred years ago our global population was less than one billion, now, it sits at almost eight billion, and by 2100, the United Nations estimates this population will continue to rise to over 11 billion [1].

An increasing population puts an increased strain on resources; energy, land, and most significantly, food. Population growth is not the only strain on global food security; climate change is affecting the growth of stable food crops [2], countries becoming more developed are seeing a transition to more meat-based and resource-intensive diets [3], and yield is being further affected by environmental pollution, soil erosion and biodiversity loss. Increasing crop yields is certainly an important area to be managed with regards to food security. An increased yield allows for a higher output from the same input, inputs including limited resources such as energy and land. Nitrogen, as an essential element for plant biomass, is therefore clearly an important piece in this problem.

1.1.2 Managing the nitrogen cycle

The nitrogen cycle is the biogeochemical cycle of nitrogen forms circulating through the atmosphere and terrestrial ecosystems (Fig. 1-1) [4]. Plants require biologically active

compounds of nitrogen for growth, such as nitrates and ammonia [5]. Unreactive atmospheric N_2 can be converted into reactive forms through a number of biological and anthropogenic processes. These biological processes include the fixation of N_2 by lightning and by N_2-fixing bacteria in the soil (Fig. 1-1). The anthropogenic formation of reactive nitrogen has caused significant disruptions to the global nitrogen cycle: more than doubling the annual transfer of N_2 into reactive forms [6]. A new era, known as the Anthropocene [7], has arisen following the industrial revolution, in which human actions are now the main driver of global environmental change [8]. Agriculture is a significant contributor to production of reactive nitrogen forms through its utilisation of the Haber-Bosch process. The Haber-Bosch process catalyses the conversion of dinitrogen into ammonia and is used to produce nitrogen fertiliser to apply to crops as a means of increasing crop yields [6]. The Haber-Bosch process is estimated to use approximately 1-2% of the world's energy supply due to its demands for high temperature and pressures [9, 10]. This manufacture of nitrogen-based fertiliser has allowed food production to roughly keep pace with population growth over the last 50 years and is responsible for feeding around 40% of the world's population [11].

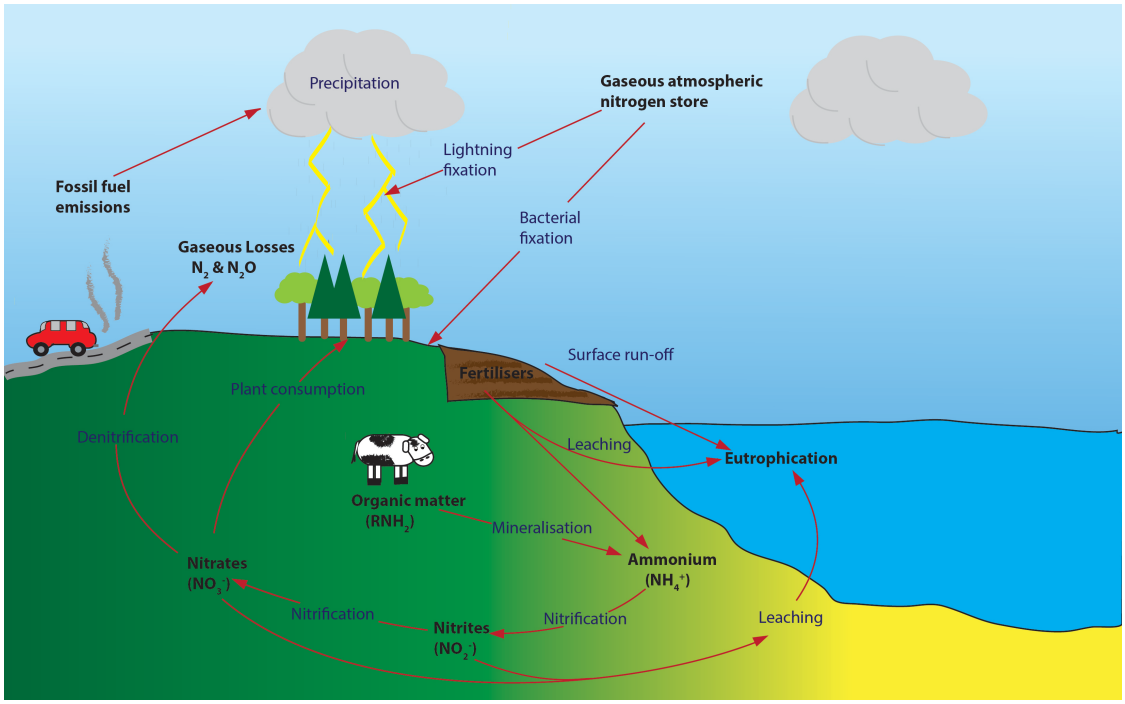


Figure 1-1. Schematic representation of the nitrogen cycle. Figure from "A problem for the 21st Century: tackling the Nitrogen Crisis" by Rachel Wheatley, The Phenotype Journal, Issue 25, 2016.

Adding nitrogen fertiliser enhances plant growth, however, there is a 'critical load' to the nitrogen that plants can absorb, after which excess will leach into the environment. Usually only 30 to 50% of the inorganic nitrogen fertiliser applied is used by the crop, and the rest is lost by volatilisation, denitrification, or leaching into groundwater [12]. Over its long lifetime in the environment (approximately 120 years) nitrogen can change forms and detrimentally affect a myriad of systems, before eventually denitrification releases it back into the atmosphere [13, 14]. The problems associated with the accumulation of reactive nitrogen into the environment are substantial and manifold, spreading across both human health and ecosystem impacts [15]. Overloading nitrogen in the environment drives biodiversity loss, forest dieback, contributes to global warming and causes human health problems [16]. Nitrogen entering water systems

promotes algal blooms, which harm water quality and cause low-oxygen (O_2) zones and death of aquatic organisms. Oxides of nitrogen (NO_x), such as nitric oxide (NO) and nitrogen dioxide (NO_2), are particularly problematic. Their breakdown in the stratosphere acts as a catalyst for ozone destruction and as a greenhouse gas they are currently the third largest contributor to global warming. Not only a major smog contributor, reactive oxides of nitrogen have also been shown to decrease human life expectancy in the areas most polluted by them. Fossil fuel combustion is another major anthropogenic driver of reactive nitrogen species generation and is a significant source of oxides of nitrogen [6, 13-15].

To address the problems presented by this new Anthropocene era a framework has been designed to define 'planetary boundaries'; the safe operating spaces associated with the planet's biophysical systems [17]. Within a certain range, Earth's complex systems generally respond smoothly to slight changing pressures. The planetary boundaries identify these threshold levels after which too much strain in excess of this range may cause system reactions to be significantly amplified with deleterious or potentially even disastrous consequences. Three of the nine identified Earth-system processes have already exceeded their planetary boundaries; one of these being anthropogenic interference with the nitrogen cycle [17]. There needs to be changes to the anthropogenic production and use of reactive nitrogen species in order to return to this safe operating space and to reduce the deleterious impacts of excess reactive nitrogen species across the ecosystem and on human health. This is a problem that will require very careful management considering the billions of people globally reliant on nitrogen-based fertiliser to produce food. Finding more environmentally friendly alternatives to increase

crop yields, and ensuring these are accessible to less developed global regions, are of utmost importance in tackling this problem.

1.2 Nature's nitrogen fixers: a sustainable method of nitrogen fertilisation

1.2.1 Biological nitrogen fixation

Biological N_2-fixation carried out by bacteria provides ~65% of the biosphere's available nitrogen [18], and is one of the key biological processes within the nitrogen cycle (Fig. 1-1). The single greatest contribution to this process comes from the symbiosis formed between soil-dwelling bacteria of the *Rhizobiaceae* family and plants of the legume family [6, 14, 15]. Rhizobia engaging in symbiosis with legumes provide biologically useable nitrogen in the form of ammonium (NH_4^+) to the plant and receive carbon sources as an energy supply in return. This effectively functions as a bio-fertiliser for the legume host, increasing plant growth without the need for exogenously applied chemical fertiliser. This form of biological N_2-fixation inputs approximately 40 million tonnes of nitrogen into agricultural systems on an annual basis [19]. By providing nitrogen directly into plant transport systems, biological N_2-fixation avoids the detrimental ecosystem impacts caused by nitrogen fertiliser run-off and leaching into the environment.

Dating back to the middle ages, crop rotations have been used to benefit soil quality and crop yields. These rotations commonly involve cereal crops, such as maize, wheat and barley, being alternated with various leguminous crops, such as beans, peas, clover and alfalfa [20]. By using legumes in crop rotation, the biologically available forms of nitrogen in the soil are increased, and this both increases crop yields and decreases the need for chemical nitrogen fertiliser application on subsequent crop cycles [21]. Biological N_2-fixation, as achieved by these *Rhizobium*-legume symbioses, provides a

sustainable method of nitrogen fertilisation without the costs of large energy expenditure, greenhouse gas production and adverse environmental effects from fertiliser leaching [21, 22].

1.2.2 *Rhizobium*-legume symbioses

To carry out biological N_2-fixation, rhizobia must first engage in symbiosis, which requires attachment and colonisation of the legume roots [23]. Legume roots secrete exudates, such as flavonoids, that can serve as chemoattractants and help facilitate rhizobial chemotaxis towards the roots. Rhizobia are able to attach to root hairs on legume roots and induce root hair curling and internalisation into the root. Rhizobia can then progress towards the root cortex along an infection thread. In the root cortex, a series of differentiation processes take place, resulting in the formation of bacteroid cells housed in protective root structures called nodules. It is within nodules that bacteroids carry out biological N_2-fixation. Two main types of nodules are characterised: indeterminate and determinate [24]. Indeterminate nodules are formed on the majority of legume species across temperate and tropic regions. They are characterised by persistent meristematic activity throughout the nodule's life that results in elongated cylindrical nodules with distinct zones of development [24]. This means a distinct zone of N_2-fixation will be present within the nodule. Determinate nodules are formed on a more restricted range of only tropical legumes and are characterised by a loss of meristematic activity shortly after formation. This results in spherical nodules, which, unlike indeterminate nodules, do not possess distinct developmental zones [25].

The bacteroid differentiation process involves the induced transcription and translation of a metalloenzyme complex known as nitrogenase, which is the enzyme that catalyses

biological N_2-fixation [26]. Nitrogenase catalyses the conversion of stable atmospheric N_2 into biologically usable NH_3 in the reaction described below [26]:

$$N_2 + 16\ ATP + 8\ H^+ + 8\ e^- \rightarrow 2\ NH_3 + 16\ ADP + 16\ P_i + H_2$$

The molybdenum (Mo)-nitrogenase is the best characterised member of this enzyme family. Mo-nitrogenase is comprised of two iron-protein dimers and one iron-molybdenum protein heterodimer (Fig. 1-2) [27]. Each iron-protein dimer contains an iron-sulphur (Fe_4S_4) cluster and an ATP binding site (Fig. 1-2) [28]. The iron-molybdenum protein is an $\alpha_2\beta_2$-tetramer that contains two complex metalloclusters per αβ-subunit pair: the 'P-cluster' (Fe_8S_7) and the 'M-cluster' ($MoFe_7S_9C$-homocitrate) (Fig. 1-2) [28]. The M-cluster serves as the active site for N_2 substrate reduction, and the iron-protein dimer acts as an electron donor for this site in the iron-molybdenum protein.

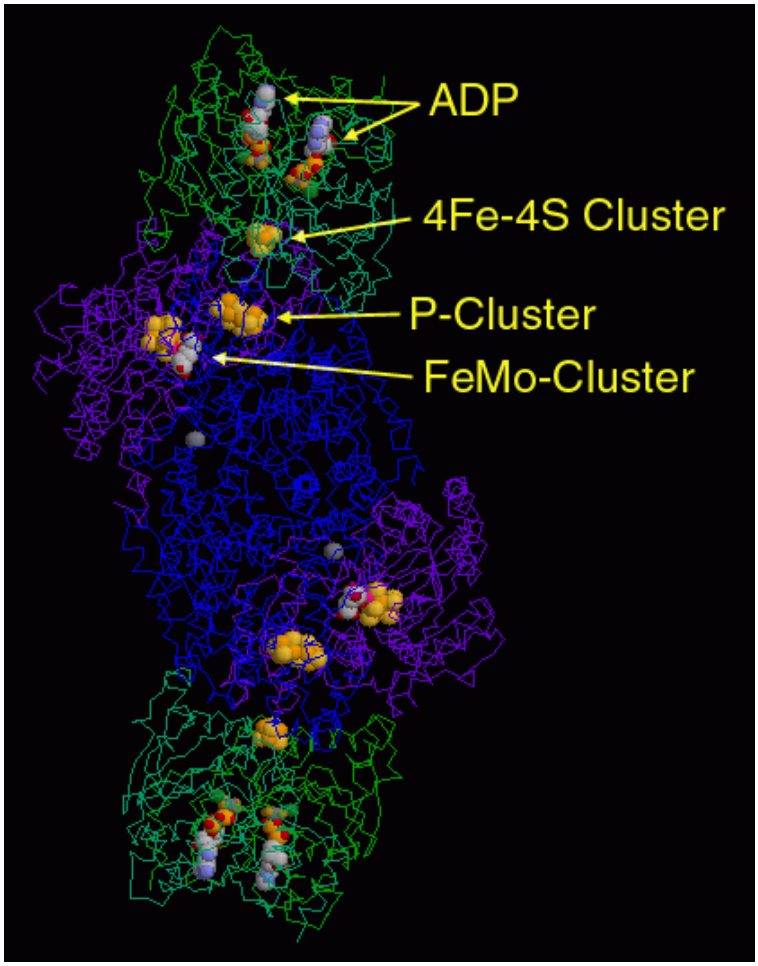


Figure 1-2. Structure of the molybdenum (Mo)-nitrogenase indicating the iron-protein dimers in green and the iron-molybdenum protein heterodimer in blue and purple. The flow of electrons for N_2-fixation progresses inwards from the ATP-binding site, to the iron-sulphur cluster of the iron-protein, to the P-cluster and finally the FeMo/'M'-cluster of the iron-molybdenum heterodimer. Figure used from Goodsell, 2002 [https://pdb101.rcsb.org/motm/26].

Oxidation of iron in the iron-sulphur clusters of nitrogenase inactivates the enzyme. As a result, N_2-fixation can only be catalysed in low O_2 environments, and so a micro-aerobic environment must be maintained in the root nodules [29]. In order to do this, an O_2-carrying hemoprotein 'leghaemoglobin' is produced by the plant host to buffer the concentration of free O_2 [30]. There are also physical barriers to O_2 diffusion in the

form of the nodule outer layers and consumption of O_2 by mitochondria of the plant, both of which contribute to a lower O_2 concentration within the nodules [19]. This low O_2 environment stimulates the onset of N_2-fixation [27], and the transcription of genes required for N_2-fixation ('*nif*' and '*fix'* genes) is regulated by O_2-status in the cell [31].

Biological N_2-fixation is a highly reductive reaction and consumes large quantities of ATP [29]. In exchange for ammonia, the legume host provides plant photosynthates, primarily in the form of dicarboxylic acids. Bacteroids are able to utilise these for energy production and generation of the ATP needed to fuel fixation [32]. These dicarboxylic acids, of which malate and succinate form significant proportions, are imported into the bacteroid via a high affinity dicarboxylic acid transport system [33-35]. Carbon sources can be metabolised via three major routes within a cell: the Entner-Doudoroff (ED) pathway (a variant of glycolysis that is used in *Rhizobium*), the tricarboxylic acid (TCA) cycle and the pentose phosphate pathway [36]. The plant-provided dicarboxylates, succinate and malate, feed into the TCA-cycle, which produces ATP in addition to reductant molecules as it cycles between carbon compounds.

Ammonia is suggested to be transported out of the bacteroids in its neutral lipophilic form via simple diffusion [27]. This is believed to occur across either the lipid bilayer, or through protein channels that are non-specific to ammonia, because the genes encoding ammonia transporters are found to be downregulated in bacteroids [19]. It is possible other specific ammonia transporters exist, but these are yet to be identified. Once out of the bacteroids, ammonia is suggested to become protonated ($NH_3 + H^+ \rightarrow NH_4^+$) in the acidic environment of the symbiosome space between the bacteroid and the plant. Two potential modes of transport across the symbiosome membrane have been

identified: an NH_3 channel [37], and a cation channel that can transport K^+, Na^+ and NH_4^+ [27]. Once in the plant cell, ammonia will become rapidly assimilated into organic form and can be transported throughout the plant via the xylem or phloem.

Symbiotic nutrient exchange is not limited to nitrogen and carbon, and the integration of plant and bacterial metabolism also includes the reliance of rhizobia on plants for the provision of components required for nitrogenase. Bacteroids are unable to produce their own homocitrate [38], an essential component of the Iron-Molybdenum cofactor of nitrogenase, and so rely on its transfer from the plant to enable N_2-fixation. Iron and sulphur are also essential components of nitrogenases that are similarly transferred across the symbiosome membrane and provided by the plant.

1.3 Global perspectives for biological nitrogen fixation research

There are two major global aims in the biological N_2-fixation field: to improve the existing symbiosis and to extend biological N_2-fixation into non-legume crops. This is with the goal of increasing crop yields, to improve global food security, and decreasing chemical nitrogen fertiliser use, to reduce the environmental and health problems associated with fertiliser production and application. In working towards these goals, it has become apparent there are a number of limiting factors that first need to be addressed; most significantly, a better understanding of the genetic requirements to successfully engage in symbiosis and the technology to facilitate successful transfer and expression of these components in other species.

Forty years have passed since the emergence of the recombinant DNA era and the original drive to transfer N_2-fixing capabilities into non-leguminous plants [39]. There have been some significant advances in this period, and while this dream has not been realised, the advent and advancement of synthetic biological and functional genomic approaches hold promise for the future. It has been suggested that the transfer of N_2-fixation to non-leguminous crops could be achieved by two major approaches: to introduce N_2-fixing abilities directly into the plant or to enable the plant to establish N_2-fixing relationships with bacteria [27, 40]. Engineering N_2-fixation with bacteria could potentially be achieved by transferring the components required for N_2-fixation into existing endophytic colonisers of the non-leguminous plants. As these bacteria already possess the ability to colonise the roots, plant associated N_2-fixation could occur without the need to engineer additional colonisation pathways [27, 40]. Or, an alternative strategy could be to take existing N_2-fixing bacteria and engineer them to recognise, colonise and establish N_2-fixing relationships with new non-leguminous plant hosts [27, 40].

A common factor to the diverse range of microorganisms with the ability to carry out biological N_2-fixation is catalysis by the O_2-sensitive metalloenzyme nitrogenase [41]. In terms of transferring N_2-fixing abilities to a new host, whether it be prokaryotic or eukaryotic in origin, it must be demonstrated that a minimum set of genetic components for N_2-fixation can be transferred to produce active nitrogenase components, and that the new host is able to provide the reducing power and energy necessary for sustaining nitrogenase catalysis [41]. Although extensive genetic and biochemical studies have been able to identify a common core set of genes/gene products required for functional nitrogenase biosynthesis [42], the complexity of this process and the sensitivity of

nitrogenase to O_2 present a significant challenge to its successful transfer to new host environments [27]. A simple model has successfully been developed in *E. coli* in which the transfer of 10 proteins is required to sustain N_2-fixation [43], and it has recently been reported that certain electron-transport chains of plant origin can be recruited to provide the reducing equivalents necessary to drive N_2-fixation [44].

Major limiting factors to engineering symbiotic N_2-fixation in plant cells exist in the limited number of well-characterised promoter elements in cereals, problems with the construction of large multigene synthetic cassettes and difficulties in engineering nitrogenase biosynthesis and stability in a microaerobic environment [27, 40]. In light of these difficulties, engineering N_2-fixation in association with bacteria might provide a more fruitful alternative. This would likely require the manipulation of both partners to exchange appropriate signals to establish recognition and colonisation. In order to achieve this, there is first the need for a better understanding of the complete process of symbiosis. While we have obtained a good understanding of the genes needed specifically for N_2-fixation (i.e. nitrogenase biosynthesis genes), the regulatory control of this process is much more poorly characterised, as are the genes needed for full symbiotic development, i.e. the genes involved in root attachment and colonisation prior to fixation. Research needs to be expanded across the whole process of N_2-fixation, not just the biological reaction, if we want to engineer new symbioses [27].

1.3.1 Enhancing symbiotic nitrogen fixation efficiency

Research into enhancing N_2-fixation efficiency in existing *Rhizobium*-legume symbioses is important for increasing the yields of currently used legume crops. Environmental and agricultural practices affect N_2-fixation efficiency, for instance,

sufficient phosphate fertilisation is important for optimizing this process. N_2-fixation rates are also influenced by temperature, light, moisture, soil salinity and soil acidity [45]. It is therefore important to understand the best soil-management practices to tackle these limiting factors, and how to implement them effectively. Additionally, if the optimum environmental conditions for particular *Rhizobium*-legume symbioses are understood, these can be taken into consideration to help ensure the best crops for the soil are planted in that region. Genetic techniques are starting to elucidate the mechanisms and systems associated with tolerance to these stress conditions [45], which could be useful for engineering or screening for more tolerant rhizobial strains.

Another important piece of research is to identify the most effective rhizobial strains that nodulate different legume species, and this can be useful for optimising rhizobial inoculum technologies. Legumes are not always colonised by the most effective fixers, instead by the strains which are most competitive in nodulation. The inoculation of crops with superior rhizobial strains, with regards to both fixing effectively and nodulating competitively, would improve crop productivity. Other aims in this field include identifying the best legumes to use in crop rotations, determining which *Rhizobium*-legume symbioses leave the soil most enriched with nitrogen, and better understanding how interactions in the microbiome can affect symbiosis efficiencies. Different bacteria in the microbiome have been identified with either stimulatory or inhibitory effects on nodulation, and better understanding these microbial interactions will surely aid manipulating them for agricultural benefit [45].

A longstanding area of research has been to see if the efficiency of symbiosis and N_2-fixation can be improved by genetic manipulation [46]. With the advent of full genome

sequencing facilities, and the large number of fully sequenced rhizobial species genomes now available, this has been an area of increased interest and significantly increased promise. A common thread to many of these global research ideas is the need for a better understanding of both the plant and the bacterial genome, and how these genetic components work together to engage successfully in symbiosis. This area is receiving significantly more attention with the development of improved genetic screening techniques and new genomic technologies, such as CRISPR-Cas9 for targeted genome studies [47].

1.4 Rhizobial taxonomy

Rhizobiaceae is a large and diverse family of gram-negative proteobacteria that consists of five principle genera: *Rhizobium, Bradyrhizobium, Azorhizobium, Mesorhizobium* and *Sinorhizobium* [48]. As of January 2016, 98 species have been described, and another eight genera exist in addition to these five major ones [49]. Rhizobial genomes are large and multipartite, consisting of a circular chromosome alongside a set of plasmids [50, 51]. Genome size and diversity are related to bacterial lifestyle. The large and diverse genomes of *Rhizobiaceae* relates to the heterogenous soil environment they inhabit, where energy sources are limited but diverse, and they are presented with many different substrates and possible hazards [52, 53]. A large genome encoding many capabilities endows a long-term selective advantage [52]. The genes responsible for symbiotic interactions with legumes, such as the nodulation (*nod*) and nitrogen-fixation (*nif-fix*) genes, are either carried by large 'symbiosis' plasmids or incorporated into the chromosome as symbiotic islands [50]. Symbiosis plasmids allow the symbiotic N_2-fixation function to be conferred via conjugation and horizonal transfer [51, 54].

Variation in plasmid number and size is not just species-specific, but strains of the same species can also show variation [53].

Rhizobium leguminosarum is a species within the genus *Rhizobium* and has distinct symbiotic variants (symbiovars). As a species, *R. leguminosarum* has been the subject of over a thousand publications [52]. *Rhizobium leguminosarum* biovar *viciae* strain 3841 (Rlv3841) has been chosen as the model strain for focus in this project. Symbiovar viciae forms root nodules on *Viciae* legumes, such as pea (*Pisum sativum*) and vetch (*Vicia cracca*) [55, 56]. Rlv3841 was chosen as the strain in this project due to the volume of transcriptomic and physiological knowledge available in the Poole lab [36, 55, 57, 58], availability of a fully sequenced genome [52], simple and well-defined growth conditions and its engagement in symbiosis with agriculturally important *Viciae* family legumes such as pea (*Pisum sativum*).

1.4.1 *Rhizobium leguminosarum* biovar *viciae* 3841

Rlv3841 is a spontaneous streptomycin- resistant mutant originating from strain 300. The genome of Rlv3841 has been sequenced [52]. This 7.75 Mb genome consists of a large, circular 5.08 Mb chromosome and six plasmids: pRL12 (0.87 Mb), pRL11 (0.68 Mb), pRL10 (0.49 Mb), pRL9 (0.35 Mb), pRL8 (0.15 Mb) and pRL7 (0.15 Mb) (Fig. 1-3). Analysis of the sequenced genome led to the prediction of 7263 protein encoding genes, about 65% of which lie on the chromosome [52]. The essential protein-encoding genes are largely chromosomal, however, there are some exceptions to this, which generally lie on the larger plasmids. The six plasmids all share the same putative replication systems based on *repABC* genes. RepA and RepB are thought to encode the

components for a partitioning system required for plasmid stability, and RepC is believed to be needed for plasmid replication [59].

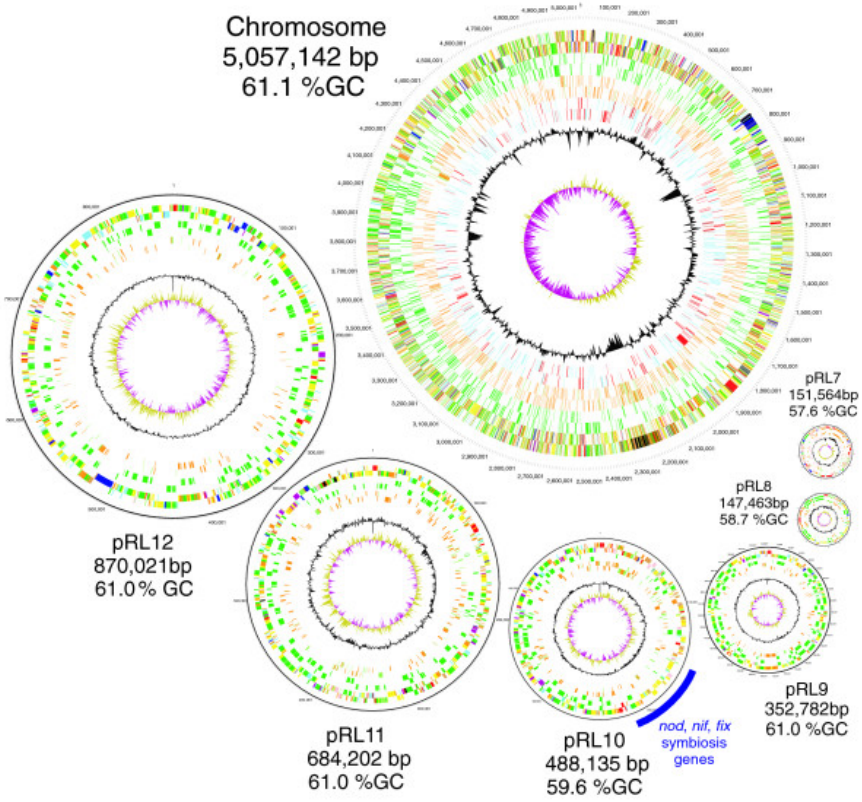


Figure 1-3. The chromosome and six plasmids of Rlv3841, figure used from Young et al. 2006 [52]. Plasmids are shown at same relative scale, chromosome at one-fourth of that relative scale. Functional classifications of genes are indicated as: membrane proteins (bright green), conserved and unconserved hypotheticals (brown conserved, pale green unconserved), phage and transposons (pink, shown for pRL7 only), and (for the chromosome only) DNA transcription/restriction/helicases (red) and transcriptional regulators (blue). Inner circles indicate deviations in G+C content (black) and G-C skew (olive/maroon). Bp; base pairs. %GC; percentage of guanine and cytosine nucleotides.

The chromosome and all plasmids except pRL7 encode a remarkably similar mix of functional classes [52]. The two smallest plasmids, pRL7 and pRL8, are transferable by conjugation, and the majority of the genes on pRL7 are apparently foreign or of unknown function. Plasmid pRL10 is regarded as the 'symbiosis plasmid' and encodes a large number of genes involved in establishing symbiosis and N_2-fixation. Clustered on pRL10 lie host plant nodulation (*nod*) genes and essential N_2-fixation (*nif* and *fix*) genes, including the *nifHDKEN* cluster, encoding nitrogenase biosynthesis components, and the *fixABCX* cluster, encoding proteins for electron transfer to nitrogenase [52]. Plasmid pRL8 has been more recently described as the 'rhizosphere plasmid' and encodes a large number of genes specifically up-regulated in the pea rhizosphere [57]. The two largest plasmids, pRL11 and pRL12, share features closest to the chromosome. pRL11 is the most chromosome-like plasmid in terms of gene composition, and pRL12 has a similar nucleotide composition to the chromosome [52].

In terms of its relevance as a model organism to other *Rhizobium* species, the genome of Rlv3841 can be split into two components. There is a 'core', mostly chromosomal component, which has a consistent phylogeny and is shared with related organisms, and an 'accessory' component, which is more individual and located on the plasmids and chromosomal islands [56]. At least two-thirds of genes within the genome are considered accessory [52], because they are not universally shared even among closely related species. While accessory genes are more frequent on the plasmids, they can also be found located in genomic islands across the chromosome [52]. Better characterising and understanding the functions of these accessory genes plays an important role in improving understanding of the bacterial adaptation and specialisation of Rlv3841. In

fact, around one-third of the genes within the genome of Rlv3841 remain of novel and uncharacterised function [52].

1.5 Lifestyle changes in *Rhizobium*-legume symbioses

Rhizobia are able to exist both as free-living soil microorganisms and as root endophytic N_2-fixing symbionts. In the transition between these two states the bacterium must progress through a number of different lifestyle changes: from navigating the rhizosphere, to attaching to roots, root internalisation and the growth in plant derived infection threads and finally, terminal differentiation into a N_2-fixing bacteroid (Fig. 1-4). These dramatic lifestyle and developmental changes are underpinned by large and diverse genomes, and are under the influence of complex regulatory pathways, responding to signals of both host plant and bacterial origin. Rhizobia must be well-adapted to thrive at each of these life stages in order to maximise success in competing for colonisation and nodulation. Transitioning between these life stages involves significant metabolic adaptations. As free-living soil microorganisms, rhizobia exist as saprophytes adopting an oligotrophic lifestyle. Subsequent to infection thread colonisation rhizobia are engulfed by plant cells where they differentiate into endosymbiotic bacteroids, receiving plant-derived nutrients and an abundant supply of carbon directly [60].

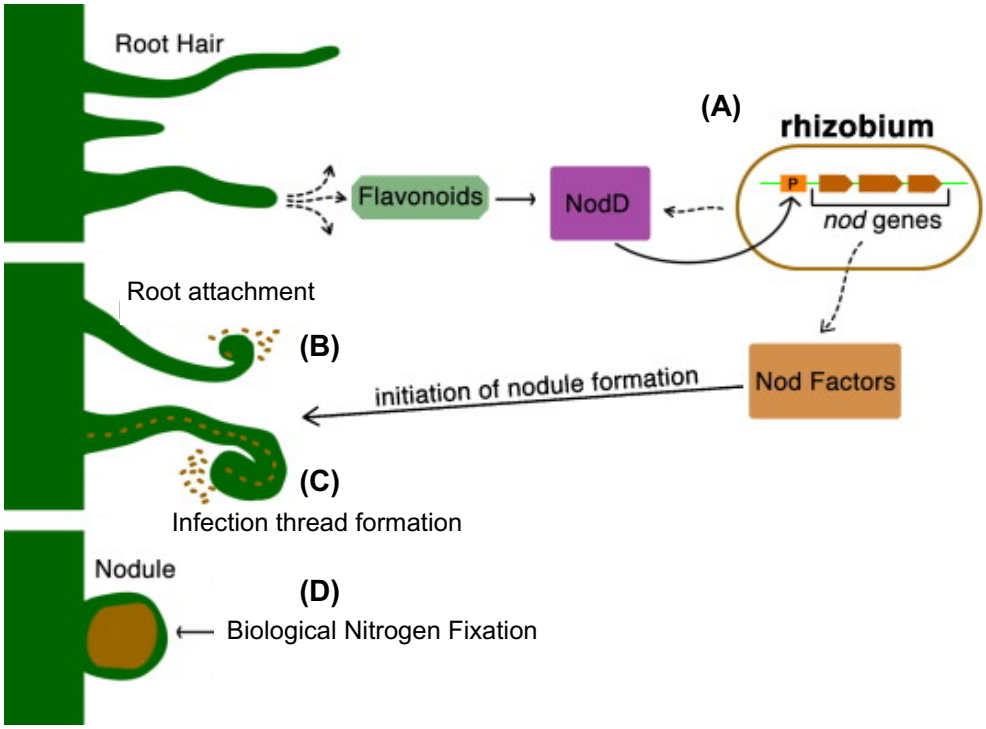


Figure 1-4. Lifestyle changes underpinning *Rhizobium*-legume symbioses. (A) Growth in the rhizosphere: rhizobia are free-living soil bacteria and can respond to symbiosis-initiating signals such as flavonoids. Flavonoids are detected by the NodD transcriptional regulator, and induce the transcription of Nod Factors. (B) Root attachment: attachment to the root hair can induce root hair curling, which entraps the attached *Rhizobium* between plant cell walls to form an infection pocket. (C) Infection thread formation: the plant cell plasma membrane can invaginate to form an infection thread, which is a long narrow passage that allows the bacteria to progress down towards the base of the root hair cell through continuous division at the leading edge. Infection threads deliver rhizobia to the root cortex. (D) Nodule formation: the lateral organ nodule is generated in response to bacterial symbiosis signals, i.e. Nod Factors. Rhizobial cells undergo a series of terminal differentiation processes to become bacteroids, specialised N_2-fixing units. Figure adapted from Larango et al. (2014) [61].

1.5.1 Growth in the rhizosphere

Rhizobia exist saprophytically in the soil and must be able to compete with the microbial community for nutrients and space [60]. Rhizobia must often be able to survive harsh soil conditions and compete with a diverse microbial community, with up to 10^4 bacterial species and 10^9 bacterial cells per g of soil, before establishing symbiosis [60]. The rhizosphere is defined as the zone of soil immediately surrounding plant roots [62]. Plants can release up to a fifth of their photosynthate via their roots [63], and so the rhizosphere can be a nutritionally-rich zone under the strong influence of root exudates. Growth in the rhizosphere represents a timepoint immediately prior to root attachment and potential symbiotic internalisation, and is the last stage rhizobia are motile microorganisms in the soil.

While motile in the soil, the initiation of symbiosis begins with a complex molecular dialogue in response to detection of legume root exudates, and the first step in this process is the release of flavonoids by the plant roots into the surrounding soil (Fig. 1-4). Flavonoids are cyclic plant secondary metabolites with different chemical modifications that lead to a spectrum of compounds conveying host-specificity. These compounds act both as chemoattractants and as inducers for rhizobial nodulation 'nod' factor signalling. Nod factor signalling is initiated following the perception of plant-released flavonoids by a positively acting transcription factor, encoded by the nodD gene (Fig. 1-4) [64]. Nod factors are a family of related lipochitooligosaccharides that share a similar chitin-like N-acetylglucosamine oligosaccharide backbone, but have different patterns of glycosylation and sulphation that determine recognition between different legumes [65]. Nod factors are secreted by the soil-dwelling rhizobia and perceived in

the host by Nod factor receptors (NFRs), which trigger a Ca_2^+-signalling pathway that activates transcription of the legume genes involved in root nodule development [65].

Rhizobia must be well-adapted to thrive in this rhizosphere environment, and this will require a host of defensive machinery to combat the oxidative environment generated by the release of reactive oxygen species from plant roots and the competition with other microbes in the soil [66]. In terms of competing with other microbes for root exudates and root proximity, this will require good nutrient scavenging capabilities, defensive machinery to any microbe produced toxin and good motility for moving towards the roots. They must also be well-adapted to deal with the general environmental stresses they may encounter, such as changing temperatures, salinity, pH and osmotic changes.

1.5.2 Root attachment and colonisation

Root attachment is not only an essential step for symbiosis, but also carries the significant benefit of anchoring bacteria in the nutrient-rich environment of the rhizosphere. Rhizobia attach specifically to the tips of developing root hairs, a region distinct from the epidermal surface of the root. This root hair binding represents the first physical contact between prospective symbiotic partners [24, 67]. Rhizobia exhibit a biphasic mechanism of root attachment: a 'primary attachment' phase followed by a 'secondary attachment' phase (Fig. 1-5). This biphasic mechanism of root attachment is observed across other agriculturally important rhizobacteria, including in *Pseudomonas*, *Salmonella*, *Azospirillum* and *Agrobacterium*. Primary attachment is generally characterised by the weak, reversible and non-specific binding of bacteria to the root surface, and secondary attachment sees a switch to a stronger, more specific binding mode with the synthesis of extracellular fibrils aiding further bacterial accumulation and

aggregation (Fig. 1-5). This tight secondary attachment is important for allowing infection thread formation when rhizobia becomes entrapped between root hair cell walls after nodulation signalling induces curling of the root hair [24].

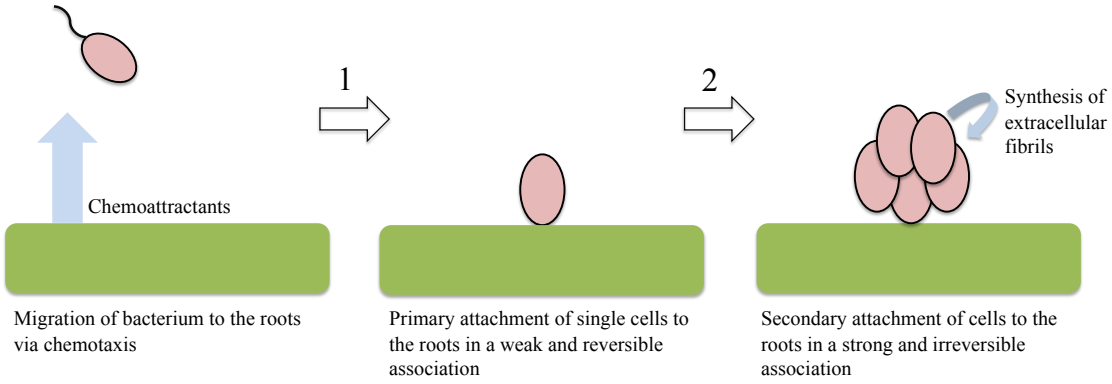


Figure 1-5. A biphasic model of root attachment. Bacteria migrate towards the roots via chemotaxis in response to root exuded chemoattractants. Root association is initially dictated by physiochemical and electrostatic attractions between the surface molecules of the root and bacterial cell envelope. Stronger microbe-specific primary attachment factors may then bind to the root surface and will dictate the overall mechanics of primary attachment. Strong and irreversible binding between the bacterium and the root surface occurs in the secondary attachment phase: there is synthesis of extracellular fibrils, cellulose and microbe-specific secondary attachment factors and the formation of bacterial aggregates. Figure published in own review article: Wheatley, Rachel M., and Philip S. Poole. "Mechanisms of bacterial attachment to roots." *FEMS microbiology reviews* (2018) (Appendix C) [68].

In *R. leguminosarum*, the primary attachment mechanism is dependent on soil pH. In acidic soil conditions, the bacterial surface polysaccharide glucomannan binds root-hair expressed lectin (Fig. 1-6) [69]. Lectins function as recognition molecules across a wide range of biological systems and mediate cell-cell interactions in a reversible and highly

specific manner [70, 71]. However, root lectins are solubilised under alkaline conditions and under these soil pHs glucomannan is no longer able to bind [69]. The mechanism of attachment under alkaline and neutral pH has been suggested to involve an extracellular rhizobial protein named 'rhicadhesin' (Fig. 1-6).

Rhicadhesin is suggested to be a 14-kDa binding protein that mediates attachment in a calcium-dependant manner, and low calcium conditions produce a significant reduction in rhizobial attachment to pea root hairs [72, 73]. Similarly to the dispersal of glucomannan under alkaline conditions, rhicadhesin is also believed to be dispersed under acidic conditions. However, neither the gene encoding rhicadhesin nor the protein itself have yet been identified, 30 years on from the original findings. It is possible that there are multiple copies of the gene, which might explain difficulties in its identification [74] but the role that rhicadhesin plays in attachment remains unclear [23, 74]. Rhicadhesin was originally defined by its ability to inhibit attachment, and while this certainly supports a role for such a protein in the stability and structure of the cell outer surface, this does not necessary support a direct role in attachment [74]. More evidence is needed to better define a role for a rhicadhesin, and this requires identification of the encoding gene(s). The mechanism that facilitates attachment under acidic conditions remains unclear.

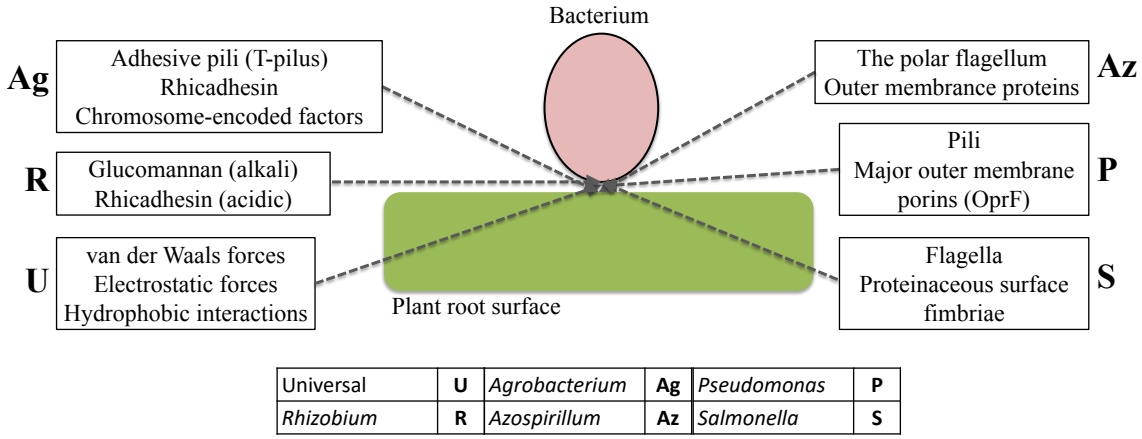
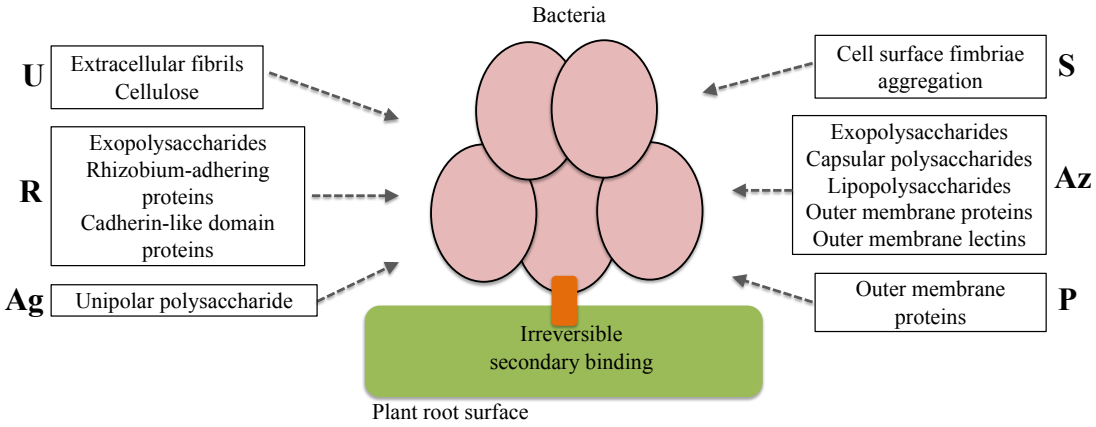


Figure 1-6. Comparison of the molecular mechanisms of primary attachment allowing the weak and reversible binding of a single cell to the root in selected agriculturally important rhizobacteria. Universal binding forces and microbe-specific attachment factors are indicated: U; Universal, R; *Rhizobium*, Ag; *Agrobacterium*, Az; *Azospirillum*, P; *Pseudomonas*, S; *Salmonella*. Figure published in own review article: Wheatley, Rachel M., and Philip S. Poole. "Mechanisms of bacterial attachment to roots." *FEMS microbiology reviews* (2018) (Appendix C) [68].

Secondary attachment generally involves extracellular fibrils aiding the formation of bacterial aggregates as a 'cap' on the root hair (Fig. 1-7). While the structure of the cap can be likened to a biofilm, the attachment process leading to the cap is very distinct from that of biofilm formation on an inert surface. A number of different secondary attachment factors have been identified, including cellulose fibrils, polysaccharides and secreted proteins (Fig. 1-7). These secreted proteins commonly contain *Rhizobium*-adhering or Cadherin-like domains such as the *Rhizobium*-adhering protein (Rap) family [75]. In *R. leguminosarum*, cellulose fibrils help facilitate tight adherence between rhizobial cells [75], but are not essential for symbiosis. Rlv3841 mutants deficient in cellulose fibril production are still able to form nodules [72], and it is believed likely

that cellulose fibril formation helps to facilitate the colonisation of fast-growing root hairs but is not strictly essential for the symbiosis [69].



Universal	**U**	*Agrobacterium*	**Ag**	*Pseudomonas*	**P**
Rhizobium	**R**	*Azospirillum*	**Az**	*Salmonella*	**S**

Figure 1-7. Comparison of the molecular mechanisms of secondary attachment allowing stronger and irreversible binding to the root in selected agriculturally important rhizobacteria. Secondary attachment is characterised by the synthesis of cellulose and other extracellular fibrils aiding the accumulation of bacterial aggregates. Universal attachment factors and microbe-specific attachment factors are indicated: U; Universal, R; *Rhizobium*, Ag; *Agrobacterium*, Az; *Azospirillum*, P; *Pseudomonas*, S; *Salmonella*. Figure published in own review article: Wheatley, Rachel M., and Philip S. Poole. "Mechanisms of bacterial attachment to roots." *FEMS microbiology reviews* (2018) (Appendix C) [68].

Exopolysaccharides (EPS) form a large component of the cell surface and are believed to play a significant role in secondary attachment. EPS is formed of carbohydrate polymers demonstrating highly variable composition between strains and species. The EPS present in Rlv3841 is known as acidic EPS and it is formed of octasaccharide

repeating units of D-glucose, D-galactose, and D-glucuronic acid residues in a 5:1:2 molar ratio. It is believed that acidic EPS is required for attaching to root hairs but is not essential for attaching at the boundaries of root epidermal cells. It is logical that the differing structural physiologies of the root hair and the root epidermis leave them receptive to different attachment mechanisms, and it is the attachment to root hairs that is important for the initiation of symbiosis.

1.5.3 Infection thread formation

Following successful attachment at the root hair, rhizobial nod factor signalling induces a Ca_2^+ influx at the root hair tip [19]. This nod factor signalling interrupts apical root hair growth and stimulates the root hair to grow back on itself. Root hair curling entraps the attached *Rhizobium* between plant cell walls to form an infection pocket (Fig. 1-4) [65]. The direction of root hair curling is dependent on the location of nod factor detection [76]. From this infection pocket site, the plant cell plasma membrane can invaginate to form an infection thread, which is a long narrow passage that allows the bacteria to progress down towards the base of the root hair cell through continuous division at the leading edge (Fig. 1-4). Division of rhizobia at the growing infection thread tip provides a forwards force for infection thread progression [77], and the plant cytoskeleton plays a key role in developing the infection thread and directing its inwards growth [78].

An infection thread will traverse the length of a cell, to release bacteria in the extracellular space, from which additional infection threads may form through underlying cells. This is an intracellular mode of infection, and through this, the infection thread will transport rhizobia through the root epidermis and cortical cells and

into the developing nodule. Infection thread formation is usually a clonal event [24], selecting one *Rhizobium* and allowing it to multiply to high numbers within the thread. This presents a highly beneficial biological event, providing one successful bacterium with a protected selective environment and the opportunity to multiply without competitors.

Rhizobial species specificity is a key checkpoint for infection thread initiation; infection by non- symbionts is unproductive and pathogen infection is a risk. In addition to Nod-factor specificity, rhizobial surface polysaccharides play a crucial role in this. It has been suggested that specificity in *Rhizobium*-legume infection thread initiation can be conveyed through the specificity of EPS in suppressing plant defence responses [79]. EPS receptors have been identified with a specificity role in *Rhizobium*-legume symbioses, an example being the EPS receptor 3 (EPR3) in *Lotus japonicus* [80]. EPR3 has extracellular chitin-binding protein related LysM domains that recognise EPS specificity, and an intracellular kinase domain that transduces these binding signals [81]. Both wild-type host rhizobial EPS and a functional EPR3 are required for sustained infection-thread initiation in *Lotus japonicus* [80].

1.5.4 Nodule organogenesis and bacteroid formation

Infection threads deliver rhizobia to the root cortex. Here the host plant cells, in response to bacterial symbiosis signals, have generated the lateral organ nodule. Rhizobial cells will be internalised by nodule cells and undergo a series of differentiation processes, the culmination of which includes the production of nitrogenase and other accessory proteins that allow N_2-fixation. This differentiation results in bacteroid cells: specialised N_2-fixing units programmed to fix nitrogen and unable to assimilate ammonia [19].

Bacteroids are encompassed by a plant-derived symbiosome membrane, across which the exchanges of symbiosis take place: fixed nitrogen in the form of ammonium is provided to the plant and carbon sources, as a supply of energy to fuel this fixation, are provided to the bacteria. *P. sativum* nodulating rhizobia, such as Rlv3841, form indeterminate nodules on the roots; these are elongated and cylindrical with distinct zones of development and N_2-fixation [24]. *L. japonicus* nodulating rhizobia, on the other hand, form determinate nodules; characterised by a loss of meristematic activity shortly after formation resulting in spherical nodules lacking distinct development zones [25].

Beyond the exchange of carbon and nitrogen, the integration of plant and bacterial metabolism extends to the dependency of bacteria on their plant hosts for the provision of other metabolic substrates and elements also important for N_2-fixation. The nodules provide a protected and microaerobic environment for the bacteria to reside in, and give the rhizobia a significant advantage in being able to multiply to high numbers. The transcription of many of the genes important for nitrogenase biosynthesis is activated by low O_2, as opposed to the nitrogen status of the cell. Leghaemoglobin is an O_2 carrier protein that buffers the concentration of O_2 in the nodules, alongside being responsible for causing the pink colouration of N_2-fixing nodules. Nodule senescence is visually identified by a colour shift from pink to green, with the green nodule colouration associated with degradation of the leghaemoglobin heme group [82]. After nodule senescence, bacteria from both determinate and indeterminate nodules are able to recolonise the soil, increasing the rhizobial soil population, and these released bacteria can then reinitiate the cycle of legume infection [83]. Bacteroids from determinate nodules are able to do this via de-differentiation into their free-living state [84].

Bacteroids from indeterminate nodules are unable to de-differentiate, but undifferentiated rhizobia, such as those recently released from the infection thread, are present in the nodule and are capable of colonising the soil [83, 85].

The irreversible changes that bacteroids in indeterminate nodules undergo include huge morphological changes, genome amplification and the loss of reproduction ability [86]. Plant factors known as nodule-specific cysteine-rich (NCR) peptides are believed to be responsible for orchestrating these terminal differentiation changes observed in indeterminate nodules. NCR peptides resemble antimicrobial peptides, and so demonstrate an adoption of immune system effectors in controlling endosymbiotic bacteria. Obstruction of NCR transport or expression correlates with absence of terminal bacteroid differentiation [87, 88]. NCR peptides have mostly been investigated in legumes of the Invert Repeat Lacking Clade (IRLC), such as *Medicago spp*., although they are also present in other groups such as plants of the genus *Aeschynomene,* belonging to the more ancient Dalbergioid lineage [89, 90]. This is likely an example of convergent evolution, bacteroid differentiation evolved independently in the Dalbergioid and IRLC legumes, but has resulted in very similar mechanisms [89, 90].

1.6 Functional genomic approaches for understanding rhizobial genomes

A better understanding of how genotype links to phenotype in *Rhizobium*-legume symbioses is important in working towards improving both the existing symbioses and also extending the symbiosis into non-legume crops. Functional genomic approaches for improving this understanding of rhizobial genomes delve into the genetic, transcriptomic, proteomic and metabolomic elements that link genotype and biological phenotype together (Fig. 1-8) [91, 92]. The sequencing and public availability of

complete genomes for numerous rhizobial and legume species opens them up to a multitude of functional genomic approaches. From the fully sequenced genomes, gene boundaries and putative gene functions can be predicted. However, this does not identify which genes are actually utilised in growth, survival and important biological processes.

Knowledge of gene function has lagged increasingly behind gene discovery, and as a result, limits our understanding of the genetic basis of important microbial phenotypes [93]. This is where functional genomic approaches come into play: to decipher which genes code for functionally important proteins within our biological systems of interest and how their products are utilised in producing biological phenotypes. Four broad categories of functional genomic approaches are described: genetic approaches, at the DNA level, transcriptomic approaches, at the RNA level, proteomic approaches, at the protein level, and metabolomic approaches, at the metabolite level (Fig. 1-8) [91, 92].

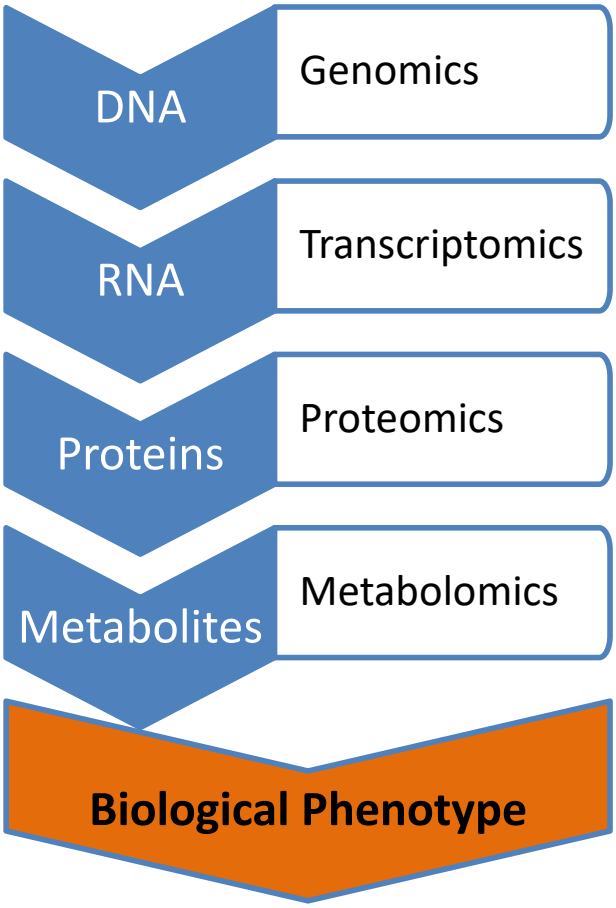


Figure 1-8. The central dogma of biology showing the flow of information from DNA to biological phenotype. Associated with each stage is the corresponding systems biology tool, from genomics to metabolomics.

At the DNA level, genetics approaches, traditionally loss-of-function studies, discern gene roles by studying changes at the DNA level which result in a change of gene function (Fig. 1-8). For example, gene function can be predicted from the phenotypic effects of loss-of-function gene mutations. Transcriptomic approaches quantify gene expression levels as a result of quantifying gene transcripts (RNA) (Fig. 1-8). This can be used to investigate gene expression levels in response to defined changes [91], and define the set of genes being actively transcribed under a certain set of conditions.

Proteomic technologies focus on the large-scale study of proteins encoded by the genome, and allow for an investigation of expression changes directly at protein level (Fig. 1-8) [91]. Metabolomics is the measurement of metabolites within a system, and is used to investigate the influence of metabolic fluxes and enzyme activity on small molecules (Fig. 1-8) [91]. While these approaches allow us to answer different questions, one of the big challenges of functional genomics is to integrate genetics with transcriptomic, proteomic and metabolomic information to provide a better understanding of the cellular biology and genome of an organism as a whole (Fig. 1-8).

Numerous factors need to be considered when integrating these approaches. For example, different functional genomic approaches may not always produce results that agree with each other, and understanding the reasons behind this can also improve our understanding of the genome and its link to functional cellular biology. Discrepancies between genetic, transcriptomic and proteomic studies may indicate the presence of genetic redundancy, protein degradation, or transcriptional regulation.

1.6.1 Genetic approaches

Despite the emergence of genome sequencing and bioinformatic analysis, there are large gaps in our knowledge of gene function and pathway connections [92]. Genetic approaches aim to deduce gene function from genome manipulation at the DNA level, and this has classically been approached through 'loss-of-function' single gene mutation studies. The C_4-dicarboxylic acid transport (Dct) system in *Rhizobium* was initially characterised through this approach by investigating the phenotype effects, in particular the reductions in C_4-dicarboxylic acid transport into the cell, of mutations to Dct system genes [34, 94-96]. For example, mutations to *dctA* were observed to generate bacteroid

strains unable to transport C_4-dicarboxylates and which displayed a N_2-fixation defective phenotype [94]. A number of different methods to induce gene mutations exist, a common technique being to insert a non-coding DNA construct that disrupts the functional coding sequence of the gene. However, single 'loss-of-function' studies are very time and labour expensive, especially if there is not a defined target, making them a more impractical solution for genome screening.

One of the new key techniques underpinning large volumes of recent genetics research is the construction and genome screening of large mutant libraries [92]. Various methods to construct mutant libraries exist, including chemical mutagenesis, ultra-violet mutagenesis, and one of the most popular systems, using transposable elements for insertion mutagenesis. Generating insertion mutant libraries by random transposon mutagenesis is a less labour intensive and more time-effective alternative that allows the study of more poorly characterised genomes [97]. Transposition events are used to disrupt the genome in a high-throughput manner gaining coverage of many genes, then genomic DNA flanking the site of transposon insertion can be used to identify the site of mutation [92]. Coupling this technique with next generation sequencing allows for high-throughput genome screening, which has revolutionised and revitalised functional genetic-based approaches [92].

The Tn5 transposon has arguably been used most frequently for insertion sequencing studies in the *Rhizobiaeceae* family [97-99]. Tn5 transposon mutagenesis conveys the advantages of a relatively high frequency of transposition with almost no insertion bias and the generation of stable insertion mutants [100-102]. Screening the Tn5 insertion libraries under conditions of interest, such as growth on a particular carbon source,

allows identification of genes whose disruption via transposition is detrimental for growth in those conditions. Transposon insertion libraries provide powerful methodologies for exploring gene function on a whole genome scale, but do have several drawbacks. The transposon mutagenesis-dense random coverage of the genome means there is no simple way to target a subset of genes, library composition can be biased against low fitness mutants after initial construction and there is selective pressure for compromised strains to accumulate second site suppressor mutations.

1.6.2 Transcriptomic approaches

Transcriptomic approaches focus on the RNA level and quantify the expression levels of genes in a comparative or quantitative manner [91]. This can identify the genes being actively transcribed under a set of conditions and can be very informative as often numerous genes within a genome can encode putative proteins of a similar predicted function. The two most commonly used transcriptomic techniques are DNA microarrays and RNA sequencing (RNA-seq) [103, 104]. Microarrays are a comparative technique to investigate the transcriptome through assessing which genes are increased or decreased in expression between conditions. They screen based on the hybridisation of cDNAs, synthesised to compliment the transcribed mRNA sequence, to specifically designed probes. Probes are designed from the genome sequence or from predicted open reading frames and are presented on an array chip for potential hybridisation with target sequences in the sample. This means the technique is limited to measuring the transcription of known sequences, and only the sequences targeted for on the probes will be quantified.

Microarrays have been successfully used in numerous comparative studies in *Rhizobium*, including in Rlv3841 specifically [55, 57]. A wide variety of comparative transcriptomic studies have been carried out, including comparing growth on different metabolites, such as glucose, pyruvate, acetate and succinate, to better understand central metabolism pathways, and to investigate general and plant-specific adaptations in rhizosphere colonisation with host and non-host legume species [57]. RNA-seq, published more recently in 2009, succeeds microarrays and allows for the quantification of all RNA transcripts present in the cell, not just those specifically probed for [104]. RNA-seq provides a high-throughput method to analyse the whole transcriptome of a target organism, and as a result, has largely replaced microarrays in transcriptomic analyses. RNA-Seq has successfully been used to quantify the transcriptome of numerous bacterial species [105], including strains of *Rhizobium leguminosarum* [106].

1.6.3 Proteomic approaches

Proteomics encapsulates the study of proteomes and their functions. Shortfalls in the ability of bioinformatics to predict the function of genes has illustrated the need for protein analysis [107], and proteomic studies can quantify expression changes in proteins, as well as allowing investigation into their subcellular localisation, interaction partners and posttranslational modifications. It is not possible to predict posttranslational modifications through genetic or transcriptomic studies, but these adaptations can profoundly affect the function of a protein [107]. Similar to the transcriptome, the proteome is not a set constant; it differs between cells and changes over time. While to some extent the proteome will reflect the underlying transcriptome, protein activity is modulated by many additional factors on top of the expression level of the encoding gene.

Traditionally the isolation and study of proteins within a complex biological sample has been carried out using 2-D Gel Electrophoresis techniques. 2-D Gel Electrophoresis separates and purifies proteins based on their molecular weights and isoelectric points. The relative abundances of separated proteins can then be compared using visualisation techniques such as protein-staining on a western blot. Protein constituents can be analysed in more detail using techniques such as mass-spectrometry that will characterise the peptide components of the protein [91].

Proteomic studies have since been advanced through liquid chromatography combined with tandem mass spectrometry, a technique also known as 'shotgun proteomics'. This is effectively a more sensitive method for the purification and peptide analysis of proteins, and one that also allows for the identification of harder-to-purify hydrophobic membrane proteins. Proteomics has been successfully used to investigate the rhizobial protein profile in several N_2-fixing symbioses. For example, the symbiosis between *Sinorhizobium meliloti* and *Medicago* has been probed using a 2-D Gel Electrophoresis comparative proteomics approach. This was able to show the metabolic shift that rhizobia undergo when entering into symbiosis is reflected by an upregulation of nitrogen-fixation proteins, amino acid ABC-transporters, inorganic ions, and stress-related proteins such as heat-shock proteins and chaperone proteins [108-110].

1.6.4 Metabolomic approaches

The objective of a metabolomic study is to measure the presence and abundance of metabolites within a system. Metabolites are defined as small molecule intermediates and products of metabolism [91]. Metabolic profiling builds on transcriptomic and

proteomic analyses by giving an instantaneous snapshot of the physiology of the cell, and effectively provides a direct functional readout of the cells physiological state. The metabolome represents the complete set of metabolites within a cell, and analysis of this under different conditions then allows comparison of the different metabolic phenotypes that the cell can adopt. Metabolomics has already been used to analyse the metabolic profiles of rhizobial cells engaged in symbiosis with legumes, in order to investigate how their metabolism changes through differentiation and in response to the requirements to fuel N_2-fixation [111, 112]. Increases in the amount of plant provided C4-dicarboxylate metabolites in the bacteroid cells and changes to the levels of several amino acids have been quantified in this way [111].

Mass spectrometry and nuclear magnetic resonance (NMR) are commonly used analytical approaches for metabolomic analysis, and effectively allow the identification and quantification of metabolite structure. Two types of approaches can be taken with these analyses: targeted metabolomics and untargeted metabolomics. Targeted metabolomics focuses on the measurement of a defined and typically small number of chemically characterised metabolites, and is used to measure known variables within a system [91]. Untargeted metabolomics on the other hand aims to comprehensively and non-selectively analyse all molecules within a sample. This untargeted approach is useful for the discovery of new metabolites or the analysis of less well characterised systems [113].

1.7 Insertion sequencing as a method of high-throughput genetic screening

A focus within this project has been to better characterise rhizobial gene function starting from this first level of functional genomics, the DNA-level. Transposon mutagenesis is

a technique of gene disruption and allows gene function to be investigated directly from this DNA-level. High-throughput genome screening of large transposon insertion libraries using next-generation sequencing technologies makes it possible to characterise putative gene functions on a whole-genome scale. This mutagenesis technique provides the ability to generate large and unbiased mutant libraries covering both characterised and uncharacterised regions of the genome. In 2009, four research groups independently published transposon sequencing methods for this purpose [93]: transposon sequencing (Tn-seq) [114], insertion sequencing (INSeq) [115], high-throughput insertion tracking by deep sequencing (HITS) [116], and transposon-directed insertion site sequencing (TraDIS) [117].

Tn-seq and INSeq are highly similar, and HITS and TraDIS are more similar to each other than to Tn-seq and INSeq [93]. Both Tn-seq and INSeq make use of a type IIS restriction site adapted transposon, such as the MmeI-adapted *Mariner* transposon [114, 115]. MmeI is a type IIS restriction endonuclease and cuts downstream of the *Mariner* transposon to capture a 16bp 'transposon tag' of the flanking genomic DNA to the site of insertion [93, 114, 115]. This 16bp of genomic DNA can be used to map the site of insertion to the genome following sequencing and identify which gene has been disrupted in each bacterium. HITS and TraDIS use transposons that lack type IIS restriction sites, and to generate transposon end fragments for genome mapping DNA is sheared randomly [116, 117]. This generates transposon tag mapping fragments of variable length. HITS and TraDIS carry the advantages of being amenable for any transposon or insertional element, whereas INSeq and Tn-seq must utilise a transposon adapted with IIS restriction sites, but do generate a final sequencing product of a precise length [93].

1.7.1 *Mariner*-based insertion sequencing

Mariner-based insertion sequencing has successfully been used for a range of different applications in various microbial species, including to characterise gene requirements in *Porphyromonas gingivalis* [118], to characterise antibiotic resistance in *Pseudomonas aeruginosa* [119] and to characterise cholesterol utilisation in *Mycobacterium tuberculosis* [120]. A MmeI-adapted *Mariner* transposon sequencing vector, pSAM_RI, that is functional in members of *Rhizobiaceae*, including Rlv3841, has recently been described and characterised [98]. It has so far been used to identify the growth requirements of Rlv3841 for growth in rich media [98]. *Mariner* class transposons carry the benefit of inserting specifically at thymine-adenine 'TA' motifs. This defined insertion specificity differs from random insertion transposons, such as the Tn5 class, in that it allows transposition events to be modelled *in silico*, which carries huge benefits in enabling robust statistical analysis and the prediction of gene essentiality. Analysis no longer relies entirely on comparison of an input and output pool of mutants but can incorporate statistical inference from the defined positions of insertion sites in order to quantify if there is an under- or over- represented number of insertions in a gene [97, 98].

It has been argued that the requirements for a specific insertion motif may limit insertion density in certain genomic regions, which could affect downstream analysis and potentially bias results [97]. This can be addressed with a prior assessment of the target microbial G+C content, the total number of potential insertion sites, and recognition of any genes lacking insertion sites or with especially low numbers. TA motifs are found abundantly in Rlv3841 [121, 122]; with a total of 140, 056 potential insertion sites in the genome, equating to an average of 15 insertion sites per gene. There are only 21

genes that contain no TA motifs [98], and so this transposon mutagenesis system should allow for the construction of highly saturated mutant libraries. The Mme1-adapted *Mariner* transposon cassette is carried on a suicide delivery vector [98], so a mutant library should be constructed in which each bacterium contains a single genome insertion. These principles have both been validated in Rlv3841 [98].

1.7.2 Methodology of insertion sequencing studies

All four transposon sequencing methods effectively work on the principle that large libraries of transposon mutants are subjected to conditions of interest, these conditions function as a selection pressure in which deleterious insertion mutants fall out of the population, and high-throughput sequencing can be used to quantify the frequencies of mutants in the population subsequent to this selection phase (Fig. 1-9). Prior to sequencing, genomic DNA must be extracted from the mutant libraries and undergo library preparation to make it amenable to sequencing. It is important to note that only regions that accumulated insertions during library construction will be able to be assayed for fitness effects in further analysis [97]. To minimise the impacts of this, library construction should be optimised to ensure an input library of a sufficiently high insertion density is created. Pooling multiple transposon mutagenesis conjugations and constructing libraries on optimal growth conditions to minimise negative selection effects should help achieve this.

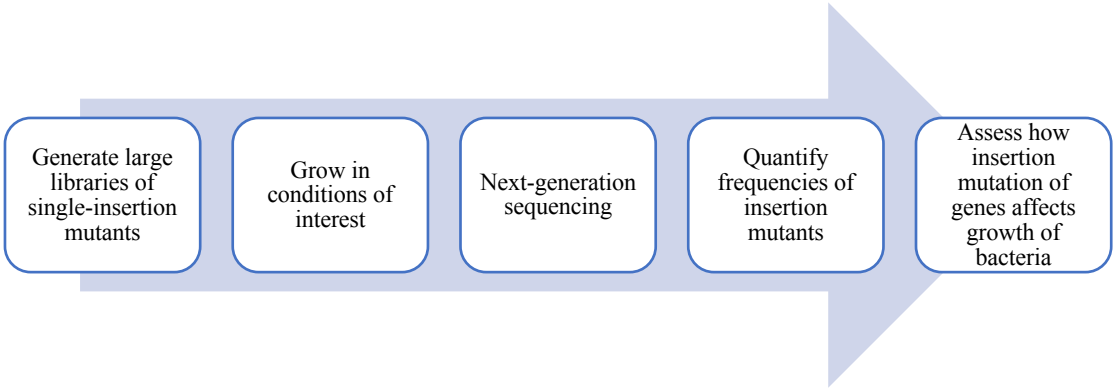


Figure 1-9. Workflow of transposon sequencing methods.

Changes to the frequencies of transposon mutants in the population can effectively be related to the contribution of the mutated gene to fitness under those conditions. Following sequencing, quantifying the frequencies of insertion mutants will allow determination of how insertion mutation of a gene affects the growth of the bacterium (Fig. 1-9). Quantification of mutant numbers and insertion locations in the population will also allow fitness values to be quantified under those conditions (Fig. 1-9). The *Rhizobium* utilisable INSeq protocol for constructing *Mariner* insertion libraries and sequencing (Fig. 1-10) is described by Perry and Yost (2014) [98], and is based on the INSeq protocol described by Goodman et al. (2011) [123]. A range of different statistical models are available to analyse this insertion sequencing data. As a limitation it should be noted that individual mutants cannot be recovered from insertion library pools for further characterisation [93]. Following insertion sequencing analysis, targeted deletions must be constructed to better investigate these loss-of-function gene phenotypes, thus gene function leads must be followed up with other, often lower-throughput approaches [93].

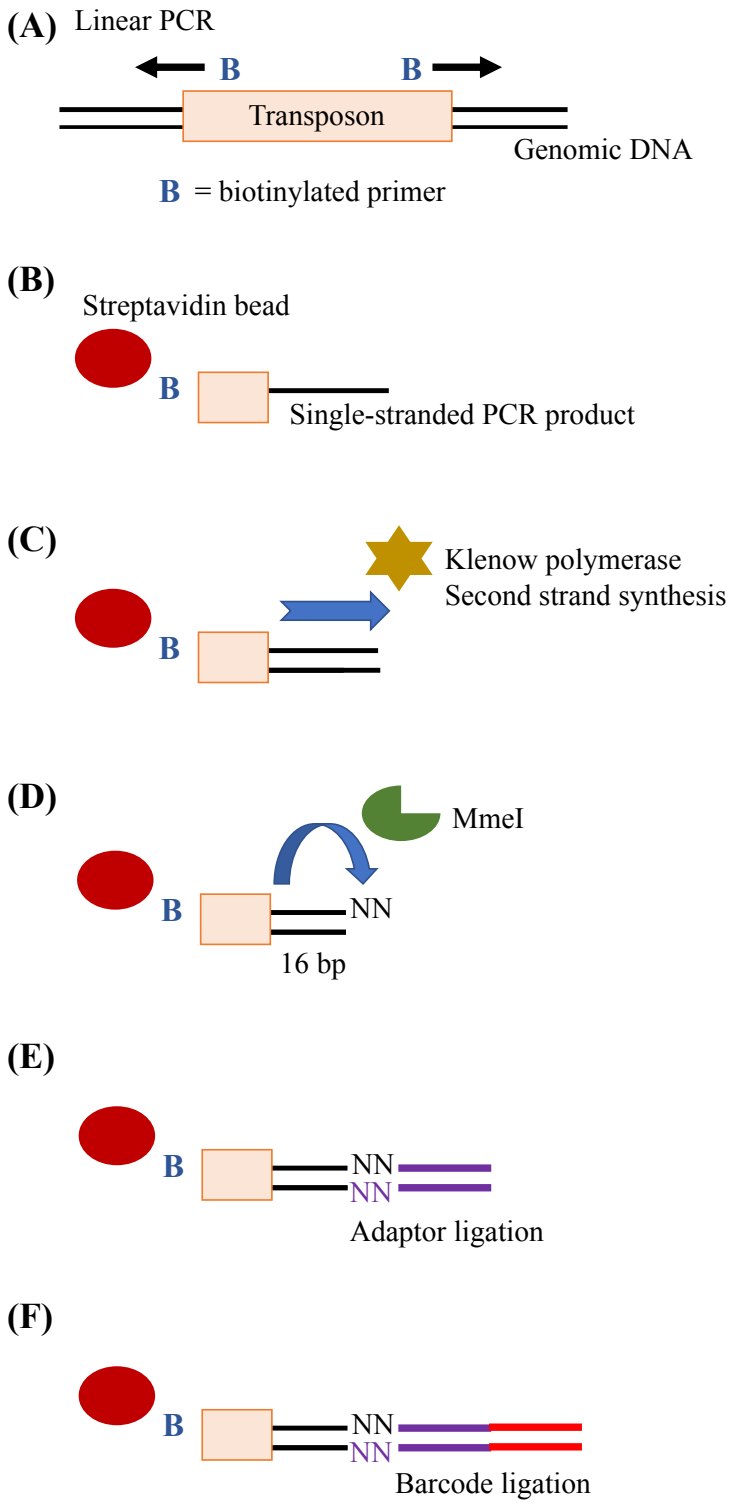


Figure 1-10. Schematic representation of the insertion sequencing library preparation protocol to process genomic DNA samples for sequencing, as described by Goodman et al (2011) and Perry and Yost (2014) [98, 123]. Genomic DNA is extracted from transposon mutant libraries. (A) Linear PCR is performed with a biotinylated primer to

amplify the transposon with adjacent chromosomal DNA. The primer is 5'-biotinylated and encodes a spacer unit to reduce steric hindrance during subsequent enzymatic steps. (B) Biotinylation allows the single stranded PCR product to be purified by binding to streptavidin beads. (C) Second strand DNA synthesis is carried out with Klenow polymerase and hexanucleotide primers. (D) Bound DNA fragments are digested with MmeI, a type IIS restriction endonuclease that cuts 16bp downstream of the *Mariner* transposon to capture a 16bp 'transposon tag' of site of transposon insertion. MmeI leaves a two-nucleotide overhang. (E) Adaptors are ligated to the two-nucleotide overhang (shown in purple). (F) Sequencing barcodes are ligated to the adaptors (shown in red), these are used to distinguish between samples. A final PCR amplification is carried out and samples are removed from the streptavidin beads. The final product is around 187bp and can be sent for sequencing. For convenience, only one side of the transposon is illustrated, in reality, both sides are amplified, sequenced and used for mapping of transposon insertion location.

1.7.3 Statistical models for analysing insertion sequencing data

Statistical models are applied to insertion sequencing data to provide a method to analyse how bacterial growth is affected by gene disruption. Different statistical models offer different ways to categorise or quantify the effects of gene disruption. Models generally rely on a number of assumptions, including that the abundance of mutants in the library reflects relative fitness, that transposon insertions occur randomly at candidate insertion sites (i.e. that there is no preferential insertion bias into the genome), and that read-counts obtained from sequencing proportionally reflect abundance of mutants in the library [124].

Many statistical models for assessing gene essentiality focus assignment to strict 'essential' or 'non-essential' categories. “ESSENTIALS” is an example of such a model [125], and assigns genes a ‘yes’ or ‘no’ output as to whether or not they are essential in analysis [125]. Gene essentiality is predicted by comparing measured number of reads per gene against the expected number of reads per gene, which is calculated based on the number of insertion sites per gene, the sequencing depth, and the mutant library size [125]. Essentiality boundaries are then calculated based on fold-change underrepresentation in the library.

Bayesian models of predicting gene essentiality also exist [126]. Bayesian inference effectively uses a rolling model of essentiality per site prediction, such that it updates the probability for an essentiality hypothesis sequentially as more insertion sites within a gene included in the analysis [126]. In basic principle, the Bayesian framework model estimates the probability of essentiality for each gene based on the statistical significance of the longest gene region lacking insertions [126]. To put simply: is a stretch of insertion sites lacking insertions a statistically significant indication of gene essentiality? This has the added advantage of being able to further resolve essential domains within a gene.

While this categorisation of essential vs. non-essential genes can be very powerful, this limited binary view does not accurately represent the more subtle ways in which bacterial growth can be affected by gene disruption. Statistical models that expand on this binary view are also available [97, 124, 127]. A four-state Hidden Markov Model (HMM) is an example of one of these, and builds on this limitation by assigning genes to one of four phenotype classifications: ‘growth-essential’, ‘growth-neutral’, ‘growth-

advantage' and 'growth-defective' [127]. As with previous models, genes that cannot tolerate disruption are classified as growth-essential, and genes whose disruption does not significantly affect organism fitness are classified as growth-neutral. However, regions of the genome with significantly higher or lower transposon read counts can now also be classified. Genes with a lower than average read count of transposon insertions are classified as "growth-defective" genes, as their insertion disruption results in impaired bacterial growth and a lower representation of clones in the library. Genes that have a higher than average read count are classified as "growth-advantaged" genes, as their disruption results in improved bacterial growth and a higher representation of clones in the library [127].

Within the HMM, the sequenced genome is viewed as an alternating sequence of essential and non-essential regions, in which a sequence of observed values (i.e. insertion site insertions) is explained by an underlying state sequence (i.e. "essentiality" of each region) [127]. The HMM incorporates information from read counts at individual TA sites to infer the probability distribution over states and infer the most likely state sequence, and there is the conditional probability of a state dependant on the previous neighbouring site. This sequential-dependence of the model affords a 'smoothing' of the read-count data that helps disambiguate the interpretation of each site and benefits correct state assignment. For example, insertion sites in non-essential regions with no insertions (i.e. because they were absent from the library in construction) are tolerated because neighbouring sites have insertions [127]. State assignment of genes relies on boundaries set for likelihood functions for read counts. In simplified form, growth-essential genes will generally have zero or very few insertions, and growth-neutral genes will sit within a mean parameter of reads that do not excel past the

boundaries for growth-defective or growth-advantage. For the growth-advantage state, the boundary is set at over five times the mean read count for a statistically significant consecutive stretch of insertion sites. For the growth-defective state, the boundary is set at less than $1/100^{th}$ of the read counts for a statistically significant consecutive stretch of insertion sites.

This HMM technique has been successfully and accurately used in the analysis of transposon insertion sequencing data in numerous studies, including in *Vibrio cholerae* [128], *Mycobacterium tuberculosis* [129], *Haemophilus influenzae* [127], and most significantly, in the initial study of *Mariner* transposon insertion mutagenesis in *Rhizobium leguminosarum* [98]. In addition to the advantage of distinguishing between four different mutant phenotype states, the model is not limited by annotated gene boundaries but is capable of identifying independent regulatory regions with a fitness value. The HMM averages over neighbouring 'TA' sites, in a statistically rigorous manner, and does not require the discarding of insertion reads in the C- and N- termini of genes, which are commonly discarded due to the uncertainty regarding entire gene disruption caused by insertions at these termini [116].

1.8 Research objectives

Our knowledge of gene function has increasingly lagged behind gene discovery, and despite the emergence of genome sequencing and bioinformatic analysis, there are large gaps in our understanding of gene function and pathway connections [92]. Understanding the genetic basis of *Rhizobium*-legume symbioses is key in working towards the global research goals of improving the existing symbiosis and extending biological N_2-fixation into non-legume crops. To these ends, the aim within this project

was to use *Mariner*-based insertion sequencing (INSeq) to characterise symbiotic fitness determinants in Rlv3841 on a whole genome-scale (Fig. 1-11).

A *Rhizobiaceae*-compatible *Mariner* transposon vector has recently been constructed that opens up the possibility of INSeq screening applications in Rlv3841 [98]. Adopting this newly published INSeq methodology should enable the robust and powerful genetic screening of Rlv3841 across symbiosis-relevant growth conditions, and this can help improve our understanding of the genetic basis of legume symbioses from the bacterial perspective (Fig. 1-11). Rlv3841 was chosen as the model strain in this project, due to the volume of transcriptomic and physiological knowledge available in the Poole lab [36, 55, 57, 58], the availability of a fully sequenced genome [52], simple and well-defined growth conditions and its symbiosis with an agriculturally important legume (pea; *Pisum sativum*).

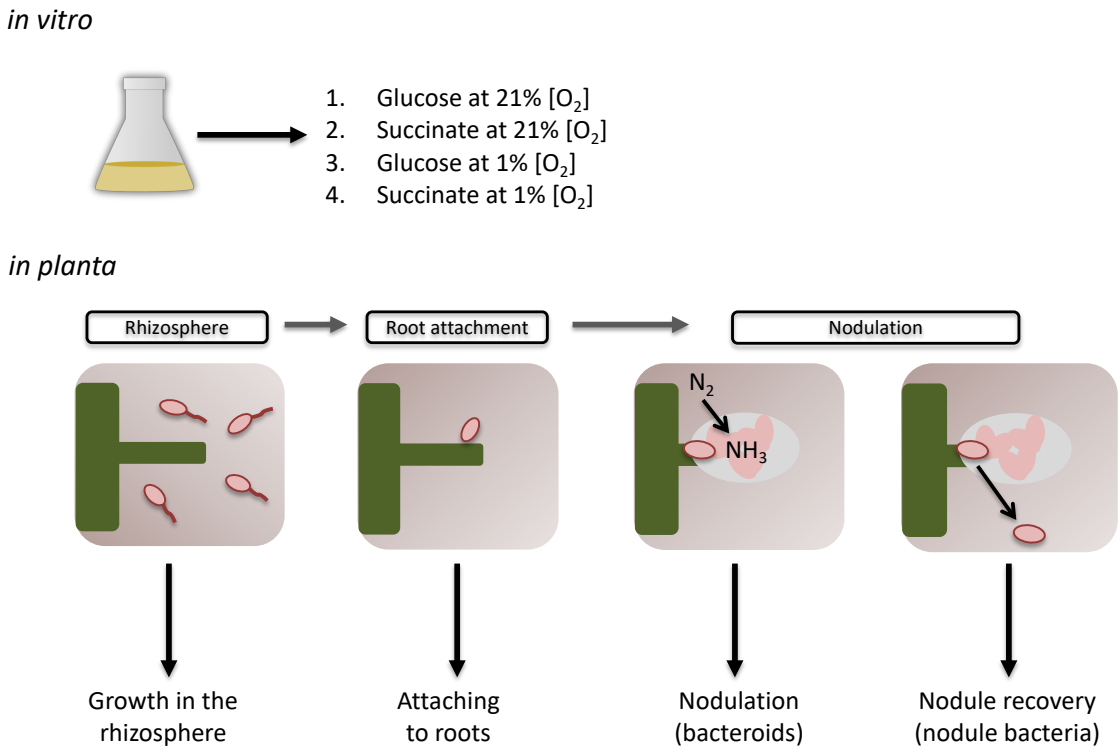


Figure 1-11. Outline of INSeq experiments to characterise symbiotic fitness determinants in Rlv3841. INSeq can be used to assess symbiotic fitness determinants *in vitro*: characterising growth at low O_2, relevant to the microaerobic nodules environment, and growth on C_4-dicarboxylates, relevant to the provision of these carbon sources by the legume host to fuel nitrogen-fixation (chapter 3). INSeq can be used to assess symbiotic fitness determinants *in planta*: characterising growth in the rhizosphere, attaching to roots, nodulation (bacteroids) and nodule recovery (nodule bacteria) (chapter 4).

1.8.1 Chapter 3

The INSeq experiment described in chapter 3 was devised with two primary research objectives: to optimise and validate the INSeq methodology, which constituted the first use of the rhizobial *Mariner*-based transposon vector following the initial publication [98], and to investigate the role of O_2 in the growth of Rlv3841 on glucose and succinate.

Understanding gene fitness at low O_2 is important for understanding adaption to the microaerobic environment of the nodules, and understanding gene fitness for growth on a C_4-dicarboxylate (succinate) is important for understanding the metabolism of C_4-dicarboxylates, which are provided by the legume host to fuel nitrogen-fixation. These variables could be investigated *in vitro* with free-living microbial culture conditions, to provide a simple and easy-to-control methodology to optimise and validate this new INSeq technique before moving onto more complex *in planta* conditions. The accuracy of INSeq fitness predictions will need to be validated through comparison against previously validated literature and through targeted loss-of-function genetic based approaches *in vivo*.

1.8.2 Chapter 4

The research objective within chapter 4 was to utilise INSeq for a more complex *in planta* application: the genome-scale characterisation of fitness determinants in early stage *Rhizobium*-legume symbioses, growth of Rlv3841 in the pea rhizosphere and attaching to pea roots. This was carried out in collaboration with an INSeq analysis of the fitness determinants in later stage symbiosis, nodulation (bacteroids) and nodule recovery (nodule bacteria). The aim within this collaborative research project was to use INSeq to characterise the symbiotic fitness determinants in Rlv3841 across four consecutive symbiotic lifestyles: 1) growth in the rhizosphere, 2) attaching to roots, 3) nodulation (bacteroids), and 4) nodule recovery (nodule bacteria). Within this study I aimed to characterise the core fitness requirements in Rlv3841, the gene requirements for growth in the rhizosphere, the gene requirements for attaching to roots, and to analyse the central carbon metabolism requirements across all four stages of symbiosis. Analysing central carbon metabolism requirements should be insightful for

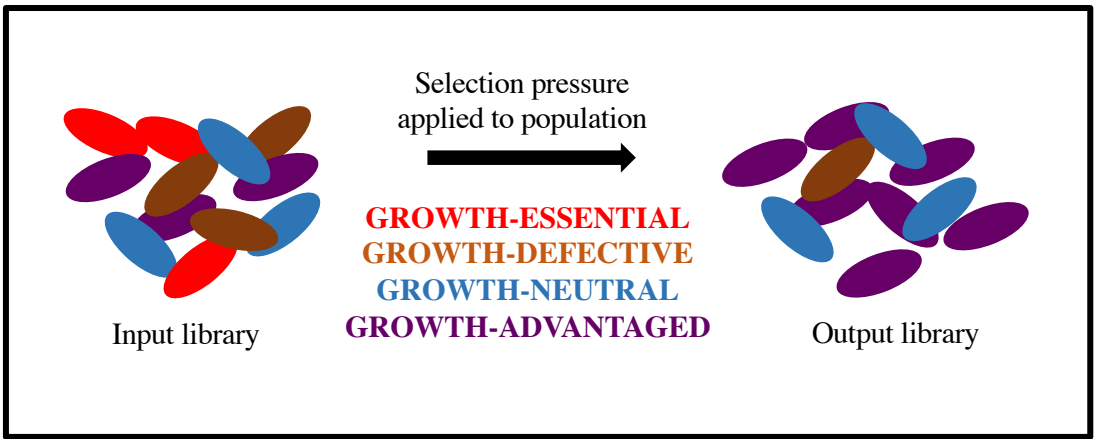
understanding the metabolic adaptations Rlv3841 undergoes in engaging in symbiosis, and might also reveal information on which carbon sources Rlv3841 is being provided with, which metabolic steps are the most important and where metabolism needs to route carbon flux for fitness in symbiosis.

1.8.3 Chapter 5

Three fundamental requirements within the *Rhizobium*-legume symbiosis are 1) the provision of a carbon source to fuel nitrogen-fixation, which is primarily in the form of dicarboxylic acids, 2) the shut-down of ammonium assimilation in bacteroids so that fixed nitrogen is provided to the legume host, and finally, 3) low O_2 in the nodules to enable functioning of O_2-sensitive nitrogenase. Analysis of the metabolic fitness determinants across symbiosis-relevant conditions (chapter 3 and chapter 4) led to the generation of two hypotheses: 1) ammonia assimilation is limited by low O_2 in rhizobia, and 2) generation of a low O_2 nodule environment alongside the provision of a dicarboxylic acid carbon source restricts ammonia assimilation in bacteroids. The final research objective within chapter 5 was to investigate these hypotheses and understand the implications this has for N_2-fixing symbioses.

Chapter 2

Materials and methods



2.1 Bacterial strains, plasmids and primers

2.1.1 Strains

The bacterial strains used in this thesis are listed in Table 2-1. All strains were stocked in 15% glycerol and flash-frozen in liquid nitrogen for storage at -80 ºC.

Strain	Description	Reference
	Rhizobium leguminosarum	
3841	*Rhizobium leguminosarum* biovar *viciae* 3841 (Rlv3841); Str^r derivative of strain 300. Nodulates *Viciae* family legumes such as pea (*Pisum sativum*) and Vetch (*Vicia cracca*)	[54]
OPS1031	Rlv3841 *RL2393*::pK19mob (oxp2109, oxp2110); Str^r Nm^r. RL2393 (*glnB*): putative nitrogen regulatory protein PII	This work
OPS1032	Rlv3841 *pRL80032*::pK19mob (oxp1072, oxp1073); Str^r Nm^r. pRL80032: putative LysR family transcriptional regulator	This work
OPS1033	Rlv3841 *pRL90149*::pK19mob (oxp1074, oxp1075); Str^r Nm^r. pRL90149: putative LuxR/GerE family transcriptional regulator	This work

OPS1034	Rlv3841 *pRL100057*::pK19mob (oxp1076, oxp1077); Str^r Nm^r. pRL100057: putative AsnC family transcriptional regulator	This work
OPS1035	*Rlv3841 pRL100172::pK19mob* (oxp1080, oxp1081)*; Str^r Nm^r*. pRL100172 (*rhiR*): putative transcriptional regulatory protein controlling rhi gene expression	This work
OPS1036	Rlv3841 *pRL120160*::pK19mob (oxp1082, oxp1083); Str^r Nm^r. pRL120160: putative DeoR family transcriptional regulator	This work
OPS1037	Rlv3841 *pRL120692*::pK19mob (oxp1084, oxp1085); Str^r Nm^r. pRL120692: putative ROK family transcriptional regulator	This work
OPS1038	Rlv3841 *RL0256*::pK19mob (oxp1086, oxp1087); Str^r Nm^r. RL0256: putative XRE family (HipB) transcriptional regulator	This work
OPS1039	Rlv3841 *RL0561*::pK19mob (oxp1088, oxp1089); Str^r Nm^r. RL0561: putative AraC family transcriptional regulator	This work

OPS1040	Rlv3841 *RL1040*::pK19mob (oxp1090, oxp1091); Str^r Nm^r. RL1040: putative LysR family transcriptional regulator	This work
OPS1041	Rlv3841 *RL2776*::pK19mob (oxp1094, oxp1095); Str^r Nm^r. RL2776: putative AsnC family transcriptional regulator	This work
OPS1042	Rlv3841 *RL2828*::pK19mob (oxp1096, oxp1097); Str^r Nm^r. RL2828: putative XRE family (HipB) family transcriptional regulator	This work
OPS1043	Rlv3841 *RL3828*::pK19mob (oxp1098, oxp1099); Str^r Nm^r. RL3828: putative FNR/CRP family transcriptional regulator	This work
RU4039	Rlv3841 *RL0037*::pK19mob; Str^r Nm^r. RL0037 (*pckA*): putative phosphoenolpyruvate carboxykinase	[130]
RU4302	Rlv3841 *RL1172*::pK19mob; Str^r Nm^r. RL1172: putative transmembrane protein	[130]

RU4285	Rlv3841 *RL4283*::pK19mob; Str^r Nm^r. RL4283 (*ptsP*): putative phosphoenolpyruvate phosphotransferase	[130]
RU4284	Rlv3841 *RL4284*::pK19mob; Str^r Nm^r. RL4284 (*aspC*): putative aspartokinase	[130]
RU4258	Rlv3841 *RL3424*::pK19mob; Str^r Nm^r. RL3424 (*dctA*): putative C4-dicarboxylate transport protein	[130]
	E. coli	
DH5alpha	Competent E. coli strain for use in transformations, carrying the following mutations; *F-deoR endA1 recA1 relA1 gyrA96 hsdR17(rk-mk+) supE44 thi-1 - phoA Δ(lacZYA-argF) U169 Φ80lacZΔM15 λ.*	Bioline
pSAM_Rl	SM10λ*pir* carrying pSAM_Rl *Mariner* transposon vector; Neo^r, Amp^r, Kan^r.	[98]
	Rhizobium etli	
CFN42	*Rhizobium etli* biovar *phaseoli* CFN42. Predominantly nodulates bean plants (*Phaseolus vulgaris*)	[131]
	Mesorhizobium loti	

MAFF303099	*Mesorhizobium loti* biovar *loti* MAFF303099. Nodulates *Lotus* species.	[132]
	Sinorhizobium meliloti	
2011	*Sinorhizobium meliloti* 2011; Str^r derivative of strain SU47. Nodulates plants from the genera *Medicago*, *Melilotus* and *Trigonella*.	[133]
	Sinorhizobium fredii	
HH103	*Sinorhizobium fredii* HH103. Nodulates soybean (e.g. *Glycine max*)	[134]

Table 2-1. *Rhizobium* and *E. coli* strains used in this work. Str^r; Streptomycin resistance, Nm^r; Neomycin resistance, Amp^r; Ampicillin resistance, Kan^r; Kanamycin resistance. Gene amplification primer numbers are included for the pK19mob strains constructed in Rlv3841.

2.1.2 Plasmids

The plasmids used in this thesis are listed in Table 2-2. The plasmid map for pSAM_Rl is shown in Fig. 2-1 [98].

Plasmid	Description	Reference
pK19mob	Mobilisable vector used for integration mutagenesis; Km^r Nm^r	[135]
pRK2013	Helper plasmid for tripartental conjugation; Km^r	[136]
pSAM_Rl	*Mariner* transposon vector; Amp^r Kan^r	[98]

Table 2-2. Plasmids used in this work. Nm^r; Neomycin resistance, Amp^r; Ampicillin resistance, Kan^r; Kanamycin reistance.

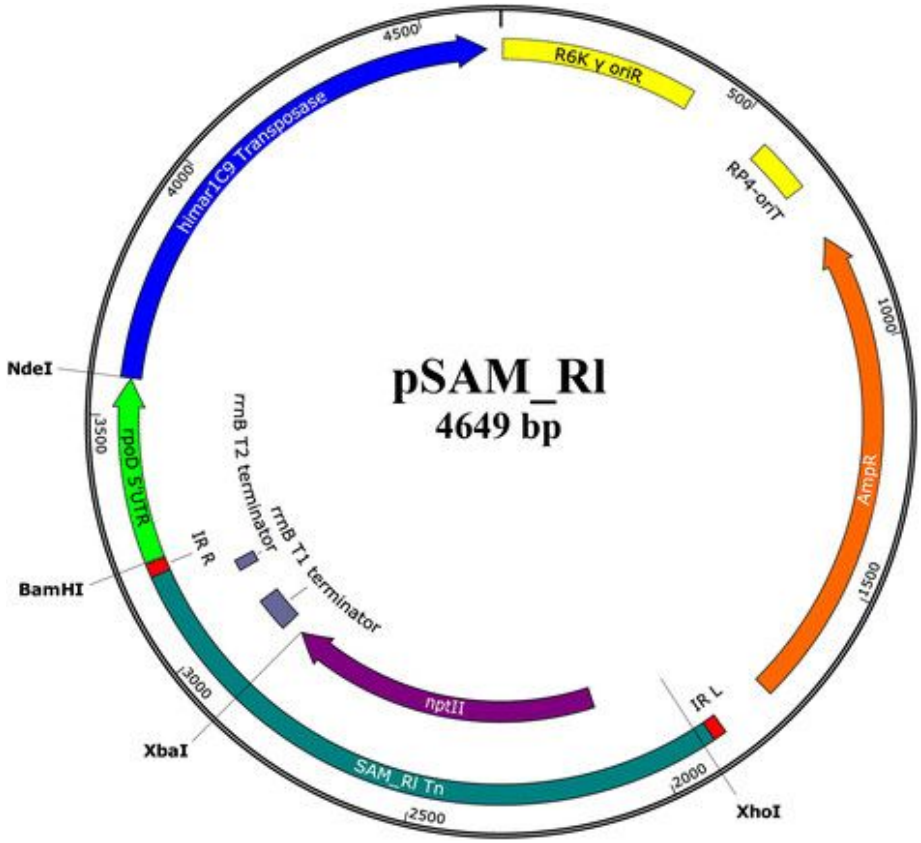


Figure 2-1. Plasmid map for *Mariner* transposon vector pSAM_Rl. Antibiotic markers (ampicillin, *AmpR*; neomycin/kanamycin, *nptII*), origin of replication (R6K γ oriR), origin of transfer (RP4-oriT), transposase (himar1C9), tranposase promoter (*rpoD* 5'UTR), MmeI-adapted *Mariner* inverse repeats (IR_R, IR_L), transposon borne Rho-independant terminator (*rrnB* T1, *rrn* B T2). Figure is reproduced from Perry and Yost (2014) [98]. Restriction enzymes used for cloning are indicated [98].

2.1.3 Primers

The primers used in this thesis are listed in Table 2-3.

Name	Sequence 5' → 3'	Description	Chapter
INSeq_Adpt_Top	AGATCGGAAGAGCGTCGTGTAGGGAA	INSeq Adapter (Top) [123]	3 & 4
INSeq_Adpt_Bottom	TTCCCTACACGACGCTCTTCCGATCTNN	INSep Adapter (Bottom) [123]	3 & 4
Ion Torrent BioSAM	/BiotinTEG/CGGTTCGCTTGCTGTCCATAAAACC	Ion Torrent BioSAM with 5' Biotin TEG [98]	3 & 4
IT_A_FP_1	CCATCTCATCCCTGCGTGTCTCCGACTCAG CTAAGGTAACGATATAAAACCGCCCAGTCTACTCGAGGG	Ion express forwards barcode primer 1 [98]. Barcode sequence is underlined.	3 & 4
IT_A_FP_2	CCATCTCATCCCTGCGTGTCTCCGACTCAG TAAGGAGAACGATATAAAACCGCCCAGTCTACTCGAGGG	Ion express forwards barcode primer 2 [98].	3 & 4

IT_A_FP_3	CCATCTCATCCCTGCGTGTCT CCGACTCAG AAGAGGATTC GATATAAAACCGCCCAGTCT ACTCGAGGG	Ion express forwards barcode primer 3 [98].	3 & 4
IT_A_FP_4	CCATCTCATCCCTGCGTGTCT CCGACTCAG TACCAAGATC GATATAAAACCGCCCAGTCT ACTCGAGGG	Ion express forwards barcode primer 4	3 & 4
IT_A_FP_5	CCATCTCATCCCTGCGTGTCT CCGACTCAG CAGAAGGAAC GATATAAAACCGCCCAGTCT ACTCGAGGG	Ion express forwards barcode primer 5	3 & 4
IT_A_FP_6	CCATCTCATCCCTGCGTGTCT CCGACTCAG CTGCAAGTTC GATATAAAACCGCCCAGTCT ACTCGAGGG	Ion express forwards barcode primer 6	3 & 4
IT_A_FP_7	CCATCTCATCCCTGCGTGTCT CCGACTCAG TTCGTGATTC GATATAAAACCGCCCAGTCT ACTCGAGGG	Ion express forwards barcode primer 7	3 & 4

IT_A_FP_8	CCATCTCATCCCTGCGTGTCT CCGACTCAG TTCCGATAAC GATATAAAACCGCCCAGTCT ACTCGAGGG	Ion express forwards barcode primer 8	3 & 4
IT_A_FP_9	CCATCTCATCCCTGCGTGTCT CCGACTCAG TGAGCGGAAC GATATAAAACCGCCCAGTCT ACTCGAGGG	Ion express forwards barcode primer 9	3 & 4
IT_A_FP_10	CCATCTCATCCCTGCGTGTCT CCGACTCAG CTGACCGAAC GATATAAAACCGCCCAGTCT ACTCGAGGG	Ion express forwards barcode primer 10	3 & 4
IT_A_FP_11	CCATCTCATCCCTGCGTGTCT CCGACTCAG TCCTCGAATC GATATAAAACCGCCCAGTCT ACTCGAGGG	Ion express forwards barcode primer 11	3 & 4
IT_A_FP_12	CCATCTCATCCCTGCGTGTCT CCGACTCAG TAGGTGGTTC GATATAAAACCGCCCAGTCT ACTCGAGGG	Ion express forwards barcode primer 12	3 & 4

IT_trP1_FP	CCTCTCTATGGGCAGTCGGT GATTTCCCTACACGACGCTCT TCCGATCT	Universal reverse barcode primer [98]	3 & 4
M12 Top	CTGTCCGTTCCGACTACCCTC CCGAC	M12 adapter [123]	3 & 4
M12 Bottom	GTCGGGAGGGTAGTCGGAAC GGACAG	M12 adapter [123]	3 & 4
M13 uni (-21)	TGTAAAACGACGGCCAGT	Sequencing/mapping primer for pK19mob containing vectors	3 & 4
M13 rev (-29)	CAGGAAACAGCTATGACC	Sequencing/mapping primer for pK19mob containing vectors	3 & 4
oxp0328	GGCTCAGTCGAAAGACTGGG C	pSAM_Rl-1st round Mariner Transposon mapping primer	3 & 4

oxp0329	GGCCACGCGTCGACTAGTCA NNNNNNNNNNCGATC	ArbRL1-2 First Round Primer for Arbitrary Primed PCR	3 & 4
oxp0330	GGCCACGCGTCGACTAGTCA	Arb2 Second Round Primer for Arbitrary Primed PCR	3 & 4
oxp0340	GCTAGACTGGGCGGTTTTAT G	pSAM_Rl mapping forwards primer	3 & 4
oxp0341	CTGCAGGTAGAAACGCAAAA AG	pSAM_Rl mapping reverse primer	3 & 4
oxp1072	TGATTACGCCAAGCTAATCC GATATTTTCTCGAGATTGCC	Amplification primer (forwards) for *pRL80032*:: pK19mob	4

oxp1073	GCAGGCATGCAAGCTTCGAT AGAAGCTTGCTGATTATCTG	Amplification primer (reverse) for *pRL80032*:: pK19mob	4
oxp1074	TGATTACGCCAAGCTCTCAG GCAGGAGAACGTAGC	Amplification primer (forwards) for *pRL90149*:: pK19mob	4
oxp1075	GCAGGCATGCAAGCTCTGCC CGAAAGAAAAAGATCGT	Amplification primer (reverse) for *pRL90149*:: pK19mob	4
oxp1076	TGATTACGCCAAGCTAGATA TTATCGTGCATGGAAATCAC C	Amplification primer (forwards) for *pRL100057*::pK19 mob	4
oxp1077	GCAGGCATGCAAGCTGAGTT CAGTGCTCTTCTTAATGGT	Amplification primer (reverse) for *pRL100057*:: pK19mob	4

oxp1080	TGATTACGCCAAGCTTCTGG TGTTCGATTTCCTGTCA	Amplification primer (forwards) for *pRL100172*:: pK19mob	4
oxp1081	GCAGGCATGCAAGCTTTTCG GCAATCATCTGAGGTGT	Amplification primer (reverse) for *pRL100172*:: pK19mob	4
oxp1082	TGATTACGCCAAGCTCTCGA AGAACTGGCCAAGCA	Amplification primer (forwards) for *pRL120160*:: pK19mob	4
oxp1083	GCAGGCATGCAAGCTTCCTC AAGCTTTTCAAGAAAATCTG C	Amplification primer (reverse) for *pRL120160*:: pK19mob	4
oxp1084	GCAGGCATGCAAGCTTCGAT CAGCATTCTCGCCAG	Amplification primer (reverse) for *pRL120692*:: pK19mob	4

oxp1085	TGATTACGCCAAGCTAAGTC TCCGGGGATCGAGC	Amplification primer (forwards) for *pRL120692*:: pK19mob	4
oxp1086	TGATTACGCCAAGCTTGGAA AATGGAACCGGAACTC	Amplification primer (forwards) for *RL0256*:: pK19mob	4
oxp1087	GCAGGCATGCAAGCTGCCGC GATTGAGCACGAC	Amplification primer (reverse) for *RL0256*:: pK19mob	4
oxp1088	TGATTACGCCAAGCTGAGGG AAAATCACCGGAAGAC	Amplification primer (forwards) for *RL0561*::pK19mob	4
oxp1089	GCAGGCATGCAAGCTCGGCT TCTGCGATTTCTGAA	Amplification primer (reverse) for *RL0561*::pK19mob	4

oxp1090	TGATTACGCCAAGCTGCAGC TTGAAGCCTTCGC	Amplification primer (forwards) for *RL1040*::pK19mob	4
oxp1091	GCAGGCATGCAAGCTAGATG GCGCCGAACAGGAAG	Amplification primer (reverse) for *RL1040*::pK19mob	4
oxp1094	TGATTACGCCAAGCTGCCTC GACCGAAAAATACTGC	Amplification primer (forwards) for *RL2776*::pK19mob	4
oxp1095	GCAGGCATGCAAGCTCCTCG TTGGACTTGGCATTG	Amplification primer (reverse) for *RL2776*::pK19mob	4
oxp1096	TGATTACGCCAAGCTTCTGA GCTTCATCGAGAGCG	Amplification primer (forwards) for *RL2828*::pK19mob	4

oxp1097	GCAGGCATGCAAGCTAGCCG GCCTTGGTTGAAAG	Amplification primer (reverse) for *RL2828*::pK19mob	4
oxp1098	TGATTACGCCAAGCTCAACA GAGCTAGAGGAAACACAG	Amplification primer (forwards) for *RL3828*::pK19mob	4
oxp1099	GCAGGCATGCAAGCTGCCTT CCGATTGAGAATACGG	Amplification primer (reverse) for *RL3828*::pK19mob	4
oxp1223	ATCCGCATTATCTCACTGGA	*pRL80032*:: pK19mob mapping primer	4
oxp1224	CACCGACAGGACTAATATTT TTCT	*pRL90149*:: pK19mob mapping primer	4
oxp1226	AAGTCGATTTGCTGCTTGGC	*pRL100057*:: pK19mob	4

		mapping primer	
oxp1227	TGCTGATCATGGACGAGACG	*pRL120160*:: pK19mob mapping primer	4
oxp1228	GTTCTCAAGTCCGCTGCCC	*pRL120692*:: pK19mob mapping primer	4
oxp1229	TGTGATAGGCGGTGGCAAG	*RL0256*::pK19mob mapping primer	4
oxp1230	GAAACAAAGTCGTTCGATCA TGC	*RL0561*::pK19mob mapping primer	4
oxp1231	ACCCAAATACAAGGCAACCG	*pRL100172*:: pK19mob mapping primer	4
oxp1232	GCTCATCGCATCTCTCCGAG	*RL1040*::pK19mob mapping primer	4
oxp1234	CATGGCTTGGCGTTGACAG	*RL2776*::pK19mob mapping primer	4

oxp1235	CGCCTTCCCGCATGATCG	*RL2828*::pK19mob mapping primer	4
oxp1236	GAATTCCAGAAATCACAGCCA	*RL3828*::pK19mob mapping primer	4
oxp2109	TGATTACGCCAAGCTGCGATCATTAAGCCTTTCAAG	Amplification primer (forwards) for *RL2393*::pK19mob	5
oxp2110	GCAGGCATGCAAGCTCGGATGCGGATAACCTCTTC	Amplification primer (reverse) for *RL2393*::pK19mob	5
oxp2111	AGACCGGATCGAGGATGTC	*RL2393*::pK19mob mapping primer	5

Table 2-3. Primers used in this work. Barcode sequences in Ion Express Barcodes are shown underlined.

2.2 Media and antibiotics

2.2.1 Media

Tryptone-yeast (TY) media was used as a rich-media growth source for *Rhizobium* strains. This consisted of: 3 g L^{-1} yeast extract, 5 g L^{-1} NaCl, 5 g L^{-1} tryptone [137]. For solid medium, 17.5 g L^{-1} agar was added before autoclaving.

Universal minimal salts (UMS) media was used a minimal-media growth source for *Rhizobium* strains. Preparation for one litre of media is as follows: 0.5 g L^{-1} $MgSO_4 \bullet 7H_2O$, 0.2 g L^{-1} NaCl, 4.19 g L^{-1} MOPS, 0.5mM K_2HPO_4, and 1 mL of trace elements (0.375 g L^{-1} EDTA-Na_2, 016g L^{-1} $ZnSO_4 \bullet 7H_2O$, 0.2 g L^{-1} $NaMoO_4$, 0.25 g L^{-1} H_3BO_3, 0.2g L^{-1} $MnSO_4 \bullet 4H_2O$, 0.02 g L^{-1} $CuSO_4 \bullet 5H_2O$, 1 g L^{-1} $CoCl_2 \bullet 6H_2O$). pH was adjusted to 7 before autoclaving. Media was supplemented after autoclaving with 1 mL of calcium stock solution (75 g L^{-1} $CaCl_2 \bullet 2H_2O$), 1 mL of iron stock solution (12 g L-1 $FeSO_4 \bullet 7H_2O$ dissolved in 1M HCl), and 1 ml of vitamin stock solution (1 g L^{-1} Thiamine hydrochloride, 2 g L^{-1} D-Pantothenic acid calcium salt, 100 mg L^{-1} Biotin). UMS was then supplemented with variable carbon and nitrogen source additions, depending on the experimental requirements (Table 2-4).

Carbon/nitrogen supplementation	Concentration in UMS	Stock
Acetate	50 mM	Sodium Acetate ($C_2H_3NaO_2$)
Ammonium	10 mM	Ammonium chloride (NH_4Cl)
Arabinose	20 mM	L-(+)-Arabinose ($C_5H_{10}O_5$)
Glucose	10 mM	D-(+)-Glucose ($C_6H_{12}O_6$)
Glutamate	10 mM	L-glutamic acid ($C_5H_9NO_4$)
Glutamine	10 mM	L-glutamine ($C_5H_{10}N_2O_3$)
Malate	20 mM	DL-Malic acid ($C_4H_6O_5$)
Pyruvate	30 mM	Sodium pyruvate ($C_3H_3NaO_3$)
Succinate	20 mM	Sodium succinate dibasic hexahydrate ($C_4H_4Na_2O_4 \bullet 6H_2O$)

Table 2-4. Carbon and nitrogen sources added to UMS and the appropriate concentrations, all stocks were resuspended in MiliQ water and filter sterilised (Millex® Filter Unit, 33mm). All stocks supplied by Sigma Aldrich.

Luria-Bertani broth (LB) was used for the growth of *E. coli* strains. This consisted of: 10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 5 g L^{-1} NaCl. For solid medium, 14 g L^{-1} agar was added before autoclaving.

2.2.2 Antibiotics

Antibiotics were added to media in concentrations shown in Table 2-5 to provide desired strain selection. Rlv3841 is a spontaneous streptomycin resistant mutant of UK field isolate 300 and streptomycin (500 µg mL-1) was used for standard selection [54]. For *E. coli* transformation where screening was carried out via blue-white colony screening, X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) was added to LB agar at 40 μg mL^{-1}.

Antibiotic	Antibiotic suspended in	Rlv3841	*E. coli*
Ampicillin (Amp)	Water	-	100 μg mL^{-1}
Gentamycin (Gent)	HEPES	20 μg mL^{-1}	10 μg mL^{-1}
Kanamycin (Kan)	Water	50 μg mL^{-1}	20 μg mL^{-1}
Neomycin (Neo)	HEPES	80 μg mL^{-1}	20 μg mL^{-1}
Spectinomysin (Spec)	HEPES	100 μg mL^{-1}	50 μg mL^{-1}
Streptomycin (Str)	HEPES	500 μg mL^{-1}	25 μg mL^{-1}
Tetracycline (Tet)	Water	5 μg mL^{-1}	10 μg mL^{-1}

Table 2-5. Working concentrations of antibiotics used in this thesis. 10 mM HEPES pH 7.0 buffer supplied by Thermo Fisher Scientific.

2.3 Growth conditions and assays

2.3.1 Growth conditions

All *Rhizobium* species used in this thesis were grown at 28 °C in TY or UMS media. Liquid cultures were aerated in a 28 °C shaker at 200-225rpm. *E. coli* strains were grown at 37ºC in LB media, and liquid cultures were aerated in a 37 °C shaker at 150-200rpm. For conjugations, *E. coli* strains were aerated at a lower speed of 100rpm, to reduce damage to fragile conjugative pili.

2.3.2 Generation time of rhizobial strains

To assess growth, 1 x 10^6 cells were grown in 50 mL of the appropriate media (in 250 mL conical flasks) at 28 ºC in an orbital shaker set at 200 rpm. The media used was either TY or UMS supplemented with 1000 X stock vitamins and the desired carbon/nitrogen sources (Table 2-4). The cultures were grown at either atmospheric conditions (21% O_2) or at 1% O_2 or 0.5% O_2 in an oxygen cabinet (Belle Technology). Optical density measurements (OD_{600nm}) were taken at four-hour time intervals until growth research stationary phase, measurements were taken at minimum for biological triplicates. The mean generation time (MGT) was calculated as the number of hours it takes the population to double while in exponential growth phase.

2.3.3 Plant growth conditions and inoculation

Pea (*Pisum sativum*) was used in this thesis as the symbiotic host plant of Rlv3841. Boiling tubes (100 mL size, DURAN ®) containing fine vermiculite (Sinclair Pro, 1.0 – 3.0 mm) supplemented with nitrogen-free rooting solution were prepared to germinate seedlings in; tubes were three-quarter filled with fine vermiculite, 20 mL nitrogen-free

rooting solution was added (Table 2-6) and tubes were then topped with foam bungs, wrapped in aluminium foil and autoclaved.

Chemical	Concentration
$CaCl_2 \bullet 2H_20$	2.67 mM
KCl	276 μM
$MgSO_4 \bullet 7H_2O$	2.13 mM
Fe EDTA	26.67 μM
H_3BO_3	93.33 μM
$MnCl_2 \bullet 4H_2O$	24 μM
$ZnCl_2$	2.13 μM
$Na_2MoO_4 \bullet 2H_2O$	1.33 μM
$CuSO_4 \bullet 5H_2O$	0.8 μM
KH_2PO_4	1.33 g/L
Na_2HPO_4	1.52 g/L

Table 2-6. Recipe to make nitrogen free rooting solution. Salts (KH_2PO_4 and Na_2HPO_4) were first dissolved into deionised water, and then the remaining solution chemicals were added.

2.3.4 Seed sterilisation and germination of *P. sativum*

P. sativum seeds were surface sterilised by immersion in 95% ethanol for 30 seconds in a sterilised flask, rinsed with sterile distilled water and then immersed in 2% bleach (sodium hypochlorite, NaClO) for five minutes. These seeds were then rinsed 10 times in sterile distilled water, transferred to a fresh sterile flask, then rinsed a further 10 times with sterile distilled water. Wrinkled, non-bloated seeds were selected for transfer into vermiculite, and placed 1-2 cm under the surface. Seed transfer was done in the Laminer Flow Workstation (Astec Microflow) under sterile conditions. One seed was planted per boiling tube, and tubes were wrapped in aluminum foil for two days, after which the foil was removed, and the plants exposed to the light. Seedlings were germinated in a controlled growth chamber, set at 23^{o}C and 16:8 photoperiod.

2.3.5 Inoculation onto *P. sativum* seedling roots

Inoculation was always carried out on seven-day old seedlings. Single colonies of the desired *Rhizobium* strain were selected from a TY agar plate (with appropriate antibiotics), streaked onto a slope of TY (10 mL of TY set at a 30^{O} angle in a 25 mL universal tube), and allowed to grow for three days prior to inoculation. Slopes were washed in UMS, diluted to the desired cell density, and inoculated into these seven-day old pea seedlings.

2.3.6 Rhizosphere and root attachment assays with *P. sativum*

Seven-day old pea seedlings were inoculated with a population of 10^5 rhizobial cells, and the population in the rhizosphere and attached to the roots was assayed at 1 day(s) post inoculation (dpi), 2 dpi, 3 dpi, 5 dpi and 7 dpi for five biological replicates. To assay the rhizosphere population, the vermiculite surrounding the roots was separated and 10

mL phosphate-buffered saline (PBS) was added to the boiling tube. The PBS and rhizosphere vermiculite was vortexed vigorously for one minute and filtered through four layers of muslin cloth into a sterile 50 mL falcon tube. This falcon tube was spun at 1000 rpm for one minute to pellet the large vermiculite, and the supernatant was serial diluted and plated onto TY agar plates supplemented with streptomycin to select for Rlv3841. The rhizosphere population (CFU; colony forming units) was calculated from the colonies counted at the appropriate serial dilution. To assay the root attached population, the pea seedling was removed from the boil tube and the roots snipped from the seed and transferred to a 50 mL falcon tube. The root was vortexed vigorously for one minute with 20 mL PBS and then transferred to a sterile pestle and mortar. The root was ground by the pestle and mortar and filtered through four layers of muslin cloth into a sterile 50 mL falcon tube. Again, this falcon tube was spun at 1000 rpm for one minute, and the supernatant was serial diluted and plated onto TY agar plates supplemented with streptomycin to select for Rlv3841. The root attached population (CFU; colony forming units) was calculated from the colonies counted at the appropriate serial dilution. The mean rhizosphere and root attached populations (CFU) at 1 dpi, 2 dpi, 3 dpi, 5 dpi and 7 dpi were plotted on GraphPad Prism 7.

2.3.7 Determining cellular NAD/NADH ratio

The NAD/NADH-Glo^{TM} Assay kit (Promega) was used for measuring NAD:NADH ratios, following the manufacturer's instructions. The linear range and specificity of the NAD/NADH-Glo^{TM} Assay was first calculated using individual purified nicotinamide adenine dinucleotides; NADH (Sigma Cat N6660) and NAD^+ (Sigma Cat N8035). A concentration range of 0 to 1000 nM [dinucleotide] was assayed; 50 µL samples at each dinucleotide concentration were incubated with 50 µL of NAD/NADH-Glo^{TM} Detection

Reagent in a white 96-well luminometer plate. Three different incubation times were assayed; 30 minutes, 40 minutes and 50 minutes. Luminescense was measured using the GloMax Multi Detection System (Promega). To measure NADH and NAD^+ in cell samples, cells were first lysed in base solution (0.2 M NaOH) + 1% DTAB (100 µL total volume), then split into two 50 μL portions, to measure NADH and NAD^+ individually. To measure NADH, lysed samples were heated at 60 ^oC for 15 minutes, incubated at room temperature (approx. 23 ^oC) for 10 minutes, 50 μL HCl/Trizma® was added, and then samples were incubated with a 1:1 volume ratio of NAD/NADH-Glo^{TM} Detection Reagent for 50 minutes before luminescence was quantified on a GloMax Multi Detection System (Promega). To measure NAD+, lysed samples were mixed with 25 μL of 0.4 M HCl, heated at 60 ^oC for 15 minutes, incubated at room temperature (approx. 23 ^oC) for 10 minutes, 25 μL of Trizma ® base was added, and then samples were incubated with a 1:1 volume ratio of NAD/NADH-Glo^{TM} Detection Reagent for 50 minutes before luminescence was quantified on the GloMax Multi Detection System (Promega). A standard curve to establish the linear range with cells was first carried out, to finalise the cell sample size and incubation time to use in the final assays. Cells were grown in 50 mL UMS (in 250 mL conical flasks) containing 50 µL 1000X stock vitamins. Either glucose (10 mM) or succinate (20 mM) was added as a carbon source, and either ammonium chloride (10 mM) or glutamate (10 mM) added as a nitrogen source (Table 2-4).

2.3.8 Measuring mean oxygen consumption rate

The OxySense® 325i oxygen analyzer system was used to measure the partial pressure of oxygen within sealed cultures of Rlv3841 provided with variable nitrogen and carbon sources. Cells were grown in UMS supplemented with 1000X vitamins and the

appropriate carbon/nitrogen sources (Table 2-4), to an OD_{600nm} of approximately 0.150, and 22 mL of this culture transferred to 22 mL universal glass tubes equipped with an oxidot sensor and stirred on a magnetic stirrer plate to equalise the diffusion of oxygen. Oxygen measurements were taken every 15 seconds until oxygen levels were depleted by the culture to below 5800 parts per billion (ppb). Mean oxygen consumption (µmol L^{-1} min^{-1} OD^{-1}) was calculated from six biological replicates across a standardised change in oxygen partial pressure.

2.4 Molecular techniques

2.4.1 Isolation of genomic DNA

Genomic DNA was extracted from *R. leguminosarum* using DNeasy Blood and Tissue Kit (Qiagen), following the manufacturers protocol. A maximum of 2 x 10^9 cells were used per isolation. Genomic DNA was then diluted to 1/100 with sterile water for use in PCR amplifications of gene fragments. Plasmid DNA was extracted using a GeneJET Plasmid Minirep Kit (Thermo Fisher Scientific), following the manufacturers protocol.

2.4.2 PCR amplification

Primers for PCR amplification were designed using Geneious R9 [138], and synthesised by Eurofins MWG Operon. Mapping or screening reactions carried out using GoTaq ® Green Master Mix (Promega). Phusion ® High-fidelity PCR Master Mix (Thermo Fisher Scientific) was used for the error-free amplification of DNA products, i.e. for use in downstream applications such as cloning. Geneious R9 determined primer melting temperature (T_m), and annealing temperatures were calculated as 5 ^oC below the T_m. A Verti® Applied Biosystems thermocycler was used for all reactions, and thermocycler conditions were set according to the manufactures instructions for polymerases:

denaturation temperature of 95 ºC, annealing temperature set according to primer T_m, extension temperature of 72 ºC (GoTaq®: 60 s per 1 kb of DNA for amplification, Phusion ®: 30 s per 1 kb of DNA for amplication). PCR products were purified using the DNA Clean and Concentrator Kit (Zymo research), following the manufacturers protocol.

When carrying out PCR mapping or screening reactions, alkaline polyethylene glycol (PEG) was used for quick and easy DNA extraction [139]. To extract DNA, a large single colony was resuspending in 20 μL alkaline PEG, heated at 60 ºC for 5 minutes, and 1 μL of this extract would be used in a PCR reaction per 10 μL total reaction volume.

2.4.3 Gel electrophoresis

Gel electrophoresis was used to separate PCR products and restriction digests. Gels were made from 0.9% agarose (Sigma Aldrich) dissolved in TAE buffer (400 mM Tris acetate, 1 mM EDTA), with Sybr® Safe (Invitrogen) added at a 1:10000 ratio for DNA staining. GeneRuler 1 kb or 100 bp (Thermo Fisher Scientific) was used a marker ladder, and electrophoresis was run at 100 mV for 25-40 minutes, dependent on fragment size for resolution. Gels were visualised using a GelDoc EZ System (BioRad).

2.4.4 Restriction digest and DNA ligation

Restriction digests were carried out using restriction endonucleases (NED, Roche or Thermo Fisher Scientific) and their respective buffers, following manufacturers instructions. DNA ligation was performed using T4 DNA Ligase and 10X T4 DNA Ligase Buffer (Thermo Fisher Scientific), following manufacturers instructions for ligation time and ligation temperature.

2.5 Cloning techniques

2.5.1 Transformation

Chemically competent *E. coli* DH5a cells stored at -80ºC were used for transformations. Prior to transformation, cells were thawed on ice for 15-20 minutes, before the transformation target DNA was added to the cells in a 1:10 μL ratio, mixed gently, and incubated on ice together for another 30 minutes. Cells were heat-shocked at 42 ºC for 60 seconds, to facilitate the uptake of DNA, and incubated on ice for a further two minutes. Cells were then suspended in SOC medium and aerated in a shaker at 37 ºC for 1 hour, and subsequently plated on LB agar for overnight growth containing the appropriate antibiotics to select for transformants. Colonies were PCR screened, and sequenced if necessary, to confirm these transformants.

2.5.2 Conjugation from *E. coli* to *R. leguminosarum*

E. coli DH5a was used to facilitate triparental mating to conjugate integration plasmids from *E. coli* into Rlv3841 with the assistance of helper plasmid pRK2013. Recipient Rlv3841 was purified from a single colony and grown on a TY slope for three days prior to conjugation. E. coli strains were grown overnight in LB, then subcultured on the morning of conjugation to reach exponential growth phase. *E. coli* cultures were pelleted and washed with TY medium three times to remove antibiotic traces, the *Rhizobium* slope was washed with 4 mL TY to resuspend cells. Recipient rhizobia, donor E. coli and helper *E. coli* were mixed in a 2:2:1 ratio, pelleted via centrifuge, and resuspended in 30 μL SOC medium (rich recovery medium; 2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose). This was transferred onto a sterile nitrocellulose filter on a TY plate, and incubated overnight at 28 ºC. The next morning, a sterile loop was used to transfer bacteria from the

nitrocellulose filter onto a TY plate containing the appropriate antibiotics to select for correct rhizobial transformants.

2.5.3 Mutagenesis by pK19mob-integration

Integration mutants were generated using pK19mob [135]. An internal fragment of the gene of choice was cloned inside the pK19 mob vector using the InFusion®HD Cloning Kit (ClonTech), and transformed into chemically competent DH5α cells. Correct colonies were confirmed using restriction digest and sequencing (Eurofins MWG Operon) with M13 uni (-21) and M13 rev (-29) primers (Table 2-3). The correct pK19mob plasmid was conjugated into Rlv3841. Conjugants were selected on TY containing streptomycin and neomycin; streptomycin to select for Rlv3841, and neomycin to select for pK19mob strains. Genomic DNA was made from conjugants and used to confirm the correct insertion.

2.6 Insertion sequencing

The protocol for insertion sequencing was used from Perry and Yost (2014) and Goodman et al (2011) [98, 123].

2.6.1 *Mariner* library construction

Donor *Mariner* transposon *E. coli* cells were grown in LB broth culture overnight and recipient rhizobial cells were grown on a TY agar slope [98]. Cultures were subjected to three cycles of pelleting at 6,000 rpm for 5 minutes, followed by resuspension in TY medium. Donor and recipient cells were pooled in equal ratios, pelleted at 6,000 rpm for 5 minutes and resuspended in 30 µl TY medium. Cell suspensions were spotted onto nitrocellulose filters on TY agar plates, incubated at 27 ºC overnight, and then the

bacteria from the filters were resuspended in UMS medium plus 15% glycerol. Enumeration of total transposon insertion Rlv3841 was performed on 500 µg/mL streptomycin plus 50 µg/mL neomycin TY agar. Independent pools of transposon mutants were generated by plating three libraries on 500 µg/mL streptomycin plus 50 µg/mL neomycin TY agar. Approximately 300,000 CFU were plated on each 245 x 245 x 25 mm plate and allowed to grow for approximately 12-18 hours until pinprick colonies were visible. Colonies were scrapped off plates resuspended in UMS. One input library was formed from the pooling of six independent conjugations of donor *Mariner* transposon *E. coli* cells and rhizobial recipients.

Output libraries were collected following growth of the input library in the desired experimental condition, and DNA was extracted from cells using a Qiagen DNeasy Blood & Tissue Kit for gram-negative bacteria according to the manufacturer's protocol with the following modifications; 2 µl RNase A (100 mg/mL) was added in parallel to proteinase K, and DNA was eluted in 100 µl of MiliQ water. To confirm successful rhizobial DNA extraction, samples were run under electrophoresis on an agarose gel alongside a GeneRuler 1 kb DNA Ladder (Thermo Fisher Scientific). DNA was quantified with a NanoDrop spectrophotometer (Thermo Fisher Scientific). For the rhizosphere and root attached INSeq samples, an additional step of growth on TY for 12 hours was used following output library extraction, to decrease plant DNA contaminations and increase bacterial DNA for gDNA extraction.

2.6.2 Library preparation and sequencing

Library preparation was carried out independently for each of the mutant libraries. Transposon-tags were prepared for DNA sequencing using a modified version of the

INSeq method to make the process compatible with the Ion Proton System sequencing platform, this method is outlined in Wheatley et al (2017) [130]. Linear PCR products were amplified using Ion Torrent Proton compatible BioSAM primers with an annealing temperature of 58.6 ºC and 500 ng of template DNA. Primer and adaptor sequences are detailed in (Table 2-3). Linear PCR products were purified (GeneJET PCR purification kit) and eluted in 50 µl MiliQ water. The biotinylated linear PCR products were bound for affinity capture with Pierce Streptavidin Magnetic Beads (Thermo Fisher Scientific) and the enzymatic library preparation steps were performed with the substitution of Klenow (New England Biolabs), Random Primer 6 (New England Biolabs) and T4 DNA Ligase (New England Biolabs). A custom INSeq library adapter was used in the adapter ligation step and then final PCR amplification used sequencing template-incorporated fusion primers designed to be compatible with Ion Amplicon Library Preparation.

Libraries were barcoded using different forward fusion primers which gave different IonXpress barcode sequences for downstream sequence separation. An Ion Torrent Proton System reverse sequencing target trP1 was used in conjunction with these forward barcode primers. This final sequencing template was 187bp in length and was gel purified from the PCR products with a GeneJET gel purification kit (Thermo Fisher Scientific). DNA to be used for sequencing was analyzed with a Bioanalyzer High Sensitivity DNA Chip (Agilent Technologies) to ensure its quality and molarity.

Libraries were diluted to 100 pM and pooled. The libraries (100 pM DNA in 25 µl) underwent automated template preparation using an Ion Chef (Thermo Fisher Scientific) and the Ion PI Chip Kit V3 (Thermo Fisher Scientific). Sequencing was performed on an Ion Proton System (Thermo Fisher Scientific).

2.6.3 Transposon insertion analysis with a four-state HMM

Sequencing reads were uploaded onto a Linux server for analysis as previously described [98, 123]. In short; 'cutadapt' was used for quality trimming and removal of adapter sequences and the resulting 15-16bp tn-tags were checked for a leading TA motif with a custom- written Perl script [140]. These tags were mapped to the Rlv3841 reference genome using Bowtie short read aligner [141], and grouped into Rlv3841 replicons. The files were converted to a .wig format by a custom-written Perl script for further analysis using the Tn-HMM Python module [127]. This first calculates the HMM state of each individual TA site, and then determines the state of all TA sites within gene boundaries to assign each gene a state as a whole. HMM analysis assigned genes to one of four classification states: growth-essential, growth-defective, growth-neutral, or growth-advantaged. The Tn-HMM Python module was used to quantify the number of genes assigned to each state. Gene annotations were obtained from a lab-curated annotation file.

2.6.4 Transposon insertion analysis with a calculation of gene fitness values

Fitness values were calculated for genes from the HMM outputs to give a quantification of mutants retrieved for that gene within the population. These fitness values can be standardised per million library reads to compare libraries across different sequencing runs. The calculation for fitness values has been devised in this thesis. The equation to calculate fitness value is given below:

Fitness value = (potential insertion sites x insertion density) x mean read count

Potential insertion sites is defined as the number of 'TA' insertion motifs within the length of the gene, insertion density is defined as the proportion of potential 'TA' insertion sites with at least one insertion or more, and mean read count is defined as the mean number of reads for insertion sites that contain at least one insertion. For an example gene (Fig. 2-2), "gene E", this can be calculated as:

Fitness value = (**4** *potential insertion sites* x **0.75** *insertion density*) x **2** *mean read count*

Fitness value = 6

The fitness value for gene E is calculated as six, which equates to the number of insertion mutants found for gene E, and can be calculated from these HMM outputs.

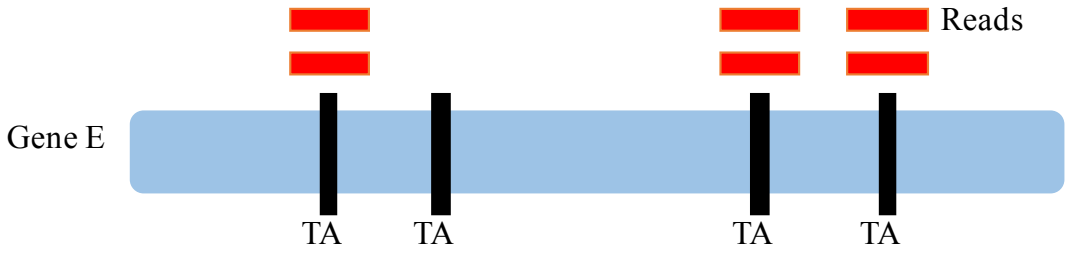


Figure 2-2. Schematic representation of HMM outputs for example gene, "Gene E", used to calculate fitness values. Gene E contains four potential *Mariner* insertion sites, represented by the four 'TA' motifs, has an insertion density of 0.75, represented by reads counted at three out of the four potential insertion sites, and has a mean read count of two, represented by the reads in red.

2.7 RNA sequencing

2.7.1 RNA extractions from root attached rhizobia

Pea seedlings were germinated for 7 days in sterile boiling tubes containing fine vermiculite and nitrogen-free rooting solution. Each seedling was inoculated with 10^5 bacteria of wild-type or relevant mutant, which were subsequently isolated from the roots at 5 days post inoculation. Twenty seedlings were inoculated per strain. All work surfaces, equipment and gloves were sterilised with 70% ethanol followed by RNAaseZap™ (Invitrogen). A face mask was worn to limit airborne RNAase contamination. Pea roots were removed from boiling tubes and snipped off at the seed. Roots were dipped in RNA later (700 g L^{-1} ammonium sulphate, 25 mM sodium citrate, 10 mM EDTA, filter sterilised , 0.22 μM) and transferred to 50 mL Falcon tubes (5 roots per falcon tube) and vortexed for 5 minutes in 20 mL of RNA Later.

Tubes were spun at 1000 rpm for 1 minute to pellet remaining vermiculite and large plant matter. The supernatants were transferred to sterile 36 mL Nalgene centrifuge tubes and centrifuged at 10,000 rpm for 10 minutes at 4 ºC to collect bacteria. RNA isolation was carried out using a RNeasy plus Mini Kit (Qiagen), following the manufacturers instructions. In brief, cells were lysed using a ribolyser (FastPrep FP120, Thermo Fisher Scientific), passed through gDNA eliminator spin columns to remove genomic DNA, and samples were then passed through the RNeasy spin columns. RNeasy spin columns bound total RNA to the column membrane and allowed contaminants to be efficiently washed away. RNA was quantified using an Experion™ Automated Electrophoresis Station (Bio-Rad) and the Experion™ RNA High Sensitivity analysis kit (Bio-Rad).

2.7.2 Isolation of mRNA and library preparation for sequencing

Total RNA samples underwent a series of purification steps to isolate mRNA and convert this to cDNA for the sequencing reaction (Fig. 2-5). RNA concentrations were quantified after each step using an Experion™ Automated Electrophoresis Station (Bio-Rad) and the Experion™ RNA High Sensitivity analysis kit (Bio-Rad). DNA was depleted from samples using TURBO DNA-*free*™ Kit (Ambion) (1; Fig. 2-5), following manufacturer's instructions. Plant mRNA was depleted from samples using Oligo d(T)25 magnetic beads (NEB) that bind the poly A tail of eukaryotic mRNA (2; Fig. 2-4). rRNA depletion was carried out using a Ribo-Zero® rRNA Removal Kit (illumina ®) (3; Fig. 2-5), and then an Ion Total RNA-Seq Kit v2 (Thermo Fisher Scientific) and an Ion Xpress™ RNA-Seq Barcode 01-16 Kit (Thermo Fisher Scientific) was used to synthesise and barcode cDNA libraries from the purified mRNA (4; Fig. 2-5). This yielded 12 barcoded samples: wild-type Rlv3841 and the 11 regulator mutants.

The concentrations and qualities of these cDNA libraries were assessed on an Agilent 2100 Bioanalyzer System and the High Sensitivity DNA Kit (Agilent). Library samples were diluted to 75 pM and pooled for final library preparation for sequencing. Library preparation was fully automated using the Ion Chef System (Thermo Fisher Scientific) and the Ion PI Chip Kit V3 (Thermo Fisher Scientific). Reagents and sample concentrations were used according to manufacturer's instructions. Sequencing was performed using an Ion Proton System (Thermo Fisher Scientific). A run plan was created on an Ion PGM Server (Thermo Fisher Scientific).

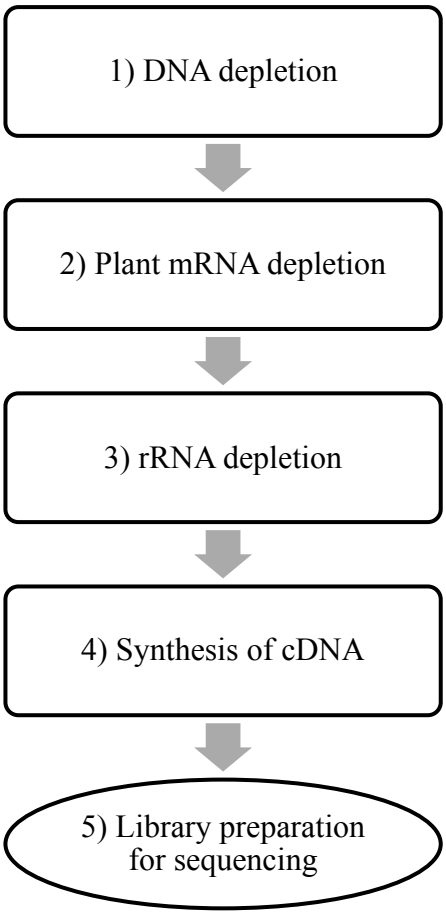


Figure 2-5. Workflow of RNA sample preparation for sequencing.

2.7.3 RNA sequencing analysis

Reads from sequencing on the Ion Proton System (Thermo Fisher Scientific) were uploaded onto a Linux server for analysis and separated by barcode to their corresponding strain. 'Cutadapt' was used for adaptor and barcode trimming. mRNA sequence reads were mapped using protein translation to the Rlv3841 genome and the degree of gene expression quantified using EDGE-pro [142]. All genome coordinates of tRNA/rRNA sequences were filtered in this analysis to remove any tRNA/rRNA reads that escaped the *in vivo* clean up.

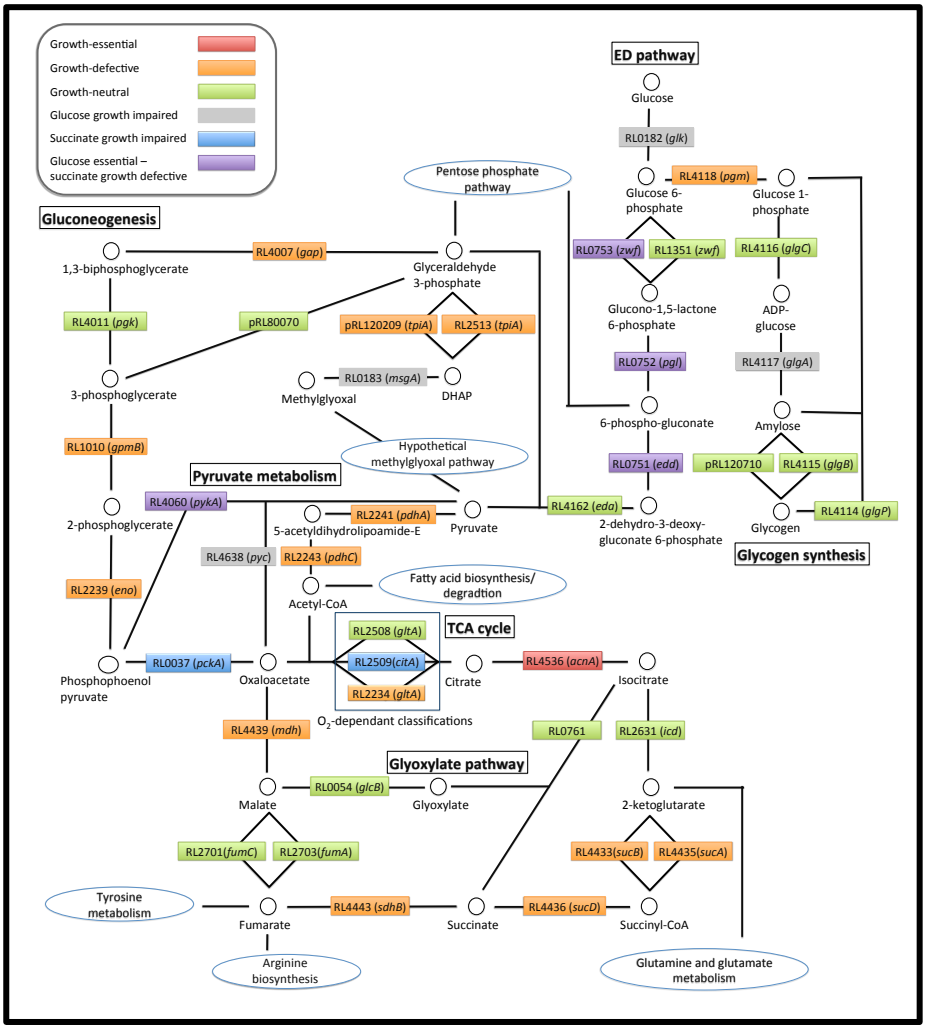
2.8 Bioinformatics, data handling and statistical analysis

The Rlv3841 genome annotation was acquired from a lab-curated genome annotation file or Rhizobase (http://genome.annotation.jp/RhizoBase). Global nucleotide and protein alignments were carried out using BLASTn and BLASTp (NCBI) respectively [143]. Geneious R9 was used to carry out local alignments [138]. Protein-protein interaction networks were predicted using STRING (https://string-db.org/) [144]. Venn-diagram comparison of common factors in multiple gene lists was carried out using Venny 2.1 (http://bioinfogp.cnb.csic.es/tools/venny/).

Data handling was predominantly carried out in MS Excel. Graphs were generated using GraphPad Prism 7. Statistical tests were carried out in either MS Excel or GraphPad Prism 7.

Chapter 3

The role of O_2 in the growth of *Rhizobium leguminosarum* on glucose and succinate



This work has been published in Wheatley et al. (2017) [130] (Appendix B).

3.1 Abstract

Insertion sequencing (INSeq) analysis of *Rhizobium leguminosarum* bv. *viciae* 3841 (Rlv3841) grown on glucose or succinate at both 21% and 1% O_2 was used to understand how O_2 concentration alters metabolism. Identification of the factors needed for growth on different carbon sources and O_2 levels is fundamental to a better understanding of the cell physiology and core metabolism of this bacterium, which adapts to a variety of different carbon sources and O_2 tensions during growth in soil and N_2 fixation in symbiosis with legumes. Two transcriptional regulators were required for growth on glucose (pRL120207 (*eryD*) and RL0547 (*phoB*)), five on succinate (pRL100388, RL1641, RL1642, RL3427 and RL4524 (*ecfL*)) and three on 1% O_2 (pRL110072, RL0545 (*phoU*) and RL4042). A novel toxin-antitoxin system was identified that could be important for generation of new plasmid-less rhizobial strains. Rlv3841 appears to use the methylglyoxal pathway alongside the ED pathway and TCA-cycle for optimal growth on glucose. Surprisingly the ED pathway was required for growth on succinate, suggesting sugars made by gluconeogenesis must undergo recycling. Altered amino acid metabolism was specifically needed for growth on glucose, including RL2082 (*gatB*) and pRL120419 (*opaA,* omega-amino acid:pyruvate transaminase). Growth on succinate specifically required enzymes of nucleobase synthesis including ribose-phosphate pyrophosphokinase (RL3468 (*prs*)) and a cytosine deaminase (pRL90208 (*codA*)). Succinate growth was particularly dependent on cell surface factors, including the PrsD-PrsE type I secretion system and UDP-galactose production. Only RL2393 (*glnB*, nitrogen regulatory protein PII) was specifically essential for growth on succinate at 1% O_2, conditions similar to those experienced by N_2-fixing bacteroids. Glutamate synthesis is constitutively activated in *glnB* mutants, suggesting consumption of 2-

ketoglutarate may increase flux through the TCA-cycle, leading to excess reductant that cannot be reoxidised at 1% O_2 and cell death.

3.2 Introduction

Rhizobia are α-proteobacteria able to form symbioses with legumes, where they induce development of root nodules. Nodules provide an environment with both a low O_2 concentration and the correct nutrients to enable differentiation of rhizobia into nitrogen (N_2)-fixing bacteroids [29]. Bacteroids fix atmospheric N_2 into ammonium (NH_3), which is secreted to the plant host, and are, in turn, provided with carbon that they utilise as an energy source [145]. To establish an effective N_2-fixing root nodule, a highly specific exchange of signals between plant and bacteria is required. Rhizobia attach to root hairs and grow down plant-derived infection threads into the root cortex [29, 146, 147]. During release from infection threads, rhizobia become surrounded by a plant-derived symbiosome membrane, creating a low O_2 environment for N_2 fixation to occur [148]. Nitrogenase, the enzyme responsible for N_2 fixation, requires at least sixteen molecules of magnesium-ATP (MgATP) per molecule of N_2 fixed [$N_2 + 8H^+ + 16MgATP + 8e^- \rightarrow 2NH_3 + H_2 + 16MgADP + 16Pi$] [149]. NH_3 is not assimilated by bacteroids, instead it is exported to the plant where it is incorporated into amino acids which are transported to growing shoots [29].

Symbiotic N_2 fixation relies on precise and efficient integration of bacterial and plant metabolism. The plant provides dicarboxylic acids, predominately malate and succinate, and these are taken up by bacteroids via the dicarboxylic transport system (Dct) that is essential for N_2 fixation [18, 32, 150]. Furthermore, either NAD^+ malic enzyme (diphosphopyridine nucleotide-dependent malic enzyme; Dme) or the combined

activities of Dme, PEP carboxykinase and pyruvate kinase are essential for dicarboxylate metabolism and N_2 fixation in bacteroids [150, 151].

Rhizobium leguminosarum bv. *viciae* strain 3841 (Rlv3841) induces root nodules on *viciae* legumes, such as agriculturally important pea (*Pisum sativum*), vetch (*Vicia cracca*) and lentil (*Lens culinaris*) [52, 55]. The genome of Rlv3841 is 7.75 Mb consisting of a circular chromosome (4788 genes) and six plasmids: pRL12 (790 genes), pRL11 (635 genes), pRL10 (461 genes), pRL9 (313 genes), pRL8 (141 genes) and pRL7 (189 genes) [52]. Microarray, transcriptomic and bioinformatics studies carried out on *Rhizobium* have indicated many gene functions. However, a high proportion (25%) remain uncharacterised across the genome, and while transcriptional studies allow identification of induced genes, they do not necessarily identify those important for growth in a particular environment [55, 57].

Mariner transposon insertion sequencing (INSeq) is a powerful and robust technique for the study of gene fitness at the genome scale. Libraries of insertion mutants are analysed by high-throughput sequencing to assess the effect of mutation of a particular gene on growth and survival of the bacterium [98]. *Mariner* transposons insert at thymine-adenine (TA) motifs found abundantly in Rlv3841 [121, 122], and INSeq sequencing data can be coupled with a Hidden Markov Model (HMM) for analysis (Fig. 3-1) [127]. The HMM assigns genes to one of four classification states based on the fitness of mutants inferred from their population frequency in the sequenced library: growth-essential (cannot tolerate insertion), growth-defective (insertion impairs growth), growth-neutral (insertion has a neutral impact on growth) or growth-advantaged (insertion enhances growth) (Fig. 3-1).

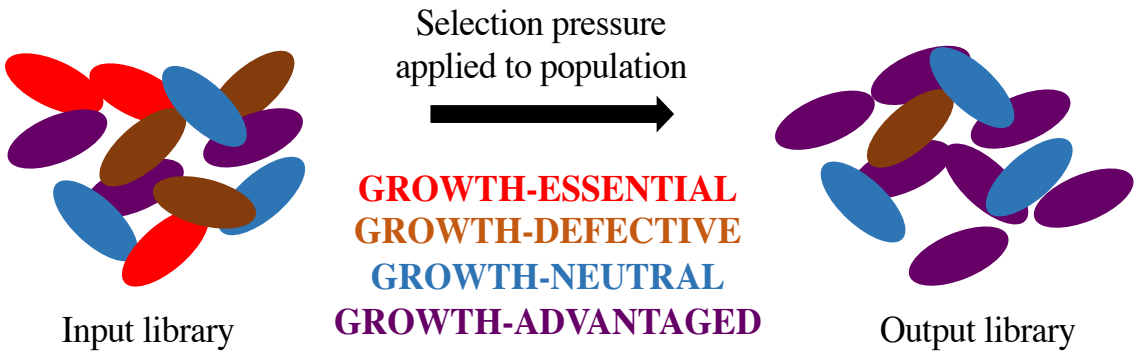


Figure 3-1. HMM analysis of INSeq experiment. Mutant cells are represented as ovals in the input and output libraries. Colour indicates HMM classification of mutant following selection pressure; red: growth-essential genes that cannot tolerate insertion and mutants are lost from the population, brown: growth-defective genes whose insertion impairs growth and mutants reduce in frequency in the population, blue: growth-neutral genes whose insertion has a neutral impact on growth and mutants do not change in frequency, and purple: growth-advantaged genes whose insertion enhances growth and mutants increase in frequency in the population.

C_4-dicarboxylates that fuel N_2 fixation in the low O_2 environment of bacteroids are catabolised by the O_2-dependent TCA-cycle. Mutational studies of enzymes in the TCA-cycle show it is required for N_2 fixation in *R. leguminosarum* and *Sinorhizobium meliloti* [152-155], although the entire cycle may not always be essential, as is the case in soybean bacteroids [156, 157]. Comparing catabolism of succinate with a sugar, such as glucose, at different O_2 tensions, is important in understanding how succinate catabolism is affected by low O_2 environments and how C_4-dicarboxylic acids fuel N_2 fixation. Here I describe the use of a *Rhizobiaceae*-compatible MmeI-adapted mariner transposon sequencing vector, pSAM_RI [98], in an INSeq genetic screen of Rlv3841 growth under

four different conditions: glucose (10 mM) at 21% O_2, succinate (20 mM) at 21% O_2, glucose (10 mM) at 1% O_2, and succinate (20 mM) at 1% O_2. This work constitutes the first use of the *Rhizobiaceae*-compatible *Mariner*-based transposon vector, pSAM_RI [98], following its initial publication. A major aim within this work was to optimise and validate this new INSeq methodology within the Poole laboratory, and to those ends, I designed the initial INSeq experiment in free-living microbial culture conditions; to give simple and easy-to-control *in vitro* variables, before moving onto investigating more complex *in planta* environments later in the project. Analysis of the growth-requirements for the chosen conditions, particularly growth on succinate at low O_2, is relevant to understanding bacteroid metabolism in legume nodules.

3.3 Results and discussion

3.3.1 Mean generation time of Rlv3841

The generation time of Rlv3841 was 3.7 h (± 0.1) grown on glucose at 21% O_2, 4 h (± 0.1) on succinate at 21% O_2, 3.8 h (± 0.2) on glucose at 1% O_2, and 6.2 h (± 0.3) on succinate at 1% O_2. The combination of succinate and 1% O_2 imposed the greatest growth restriction on Rlv3841. This is consistent with succinate being catabolised by the TCA-cycle that generates large amounts of NADH and $FADH_2$ which must then be reoxidised by the electron transport chain. For growth on succinate at 1% O_2, electron transport may start to limit growth as reoxidisation of the TCA-cycle generated reductant is reliant on the presence of molecular O_2 to act as a final electron acceptor.

3.3.2 INSeq experimental design

Three independent pools of 2 × 10^7 transposon mutants derived from mutant library construction on rich Tryptophan Yeast (TY) media were grown for 14 generations under four different growth conditions: glucose at 21% O_2, succinate at 21% O_2, glucose at 1% O_2, and succinate at 1% O_2 (Fig. 3-2). Library preparation was carried out independently for each of the 12 output mutant libraries and from sequencing a total of 78 million barcoded sequencing reads were obtained.

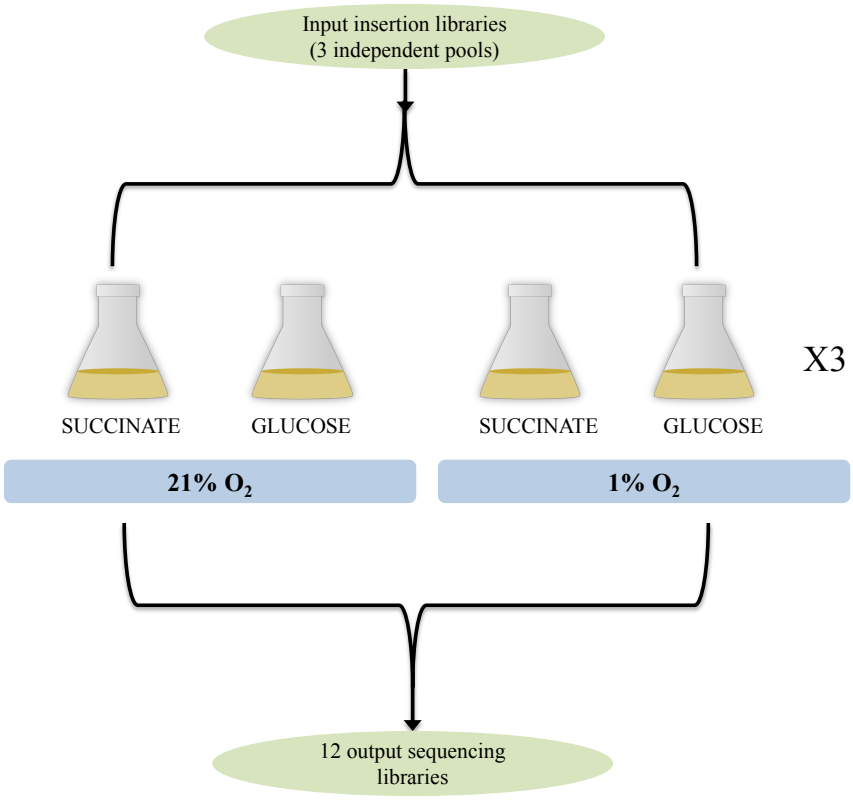


Figure 3-2. INSeq experiment design consisting of four different growth conditions: 21% O_2 + succinate, 21% O_2 + glucose, 1% O_2 + succinate and 1% O_2 + glucose. Three replicates at each condition produced 12 output libraries for sequencing. Cultures of 2 × 10^7 transposon mutants were inoculated into 500 mL UMS (in 2.5 L conical flasks) containing 100 mM ammonium chloride, 500 µl 1000 X stock vitamins and the appropriate succinate (20 mM) or glucose (10 mM) carbon source at either atmospheric

conditions (21% O_2) or at 1% O_2 in an oxygen cabinet at 200 rpm (Belle Technology).

3.3.3 HMM genome analysis

Sequencing data coupled with HMM analysis enabled the phenotypes of mutants in 7316 genes (99.7% of the genome) to be classified into one of four categories (growth-essential, growth-defective, growth-neutral, or growth-advantaged) [130]. No data are available for 21 genes of Rlv3841 which lack TA motifs [98]. Across the four different growth conditions, mutants of 307 genes (4.2% of the genome) resulted in classification as growth-essential, 900 as growth-defective (12.3%), 6072 as (83.3%) growth-neutral, and 7 as growth-advantaged (0.1%) (Appendix A; Table 1). For 288 genes of unknown function mutants were growth-essential, growth-defective or growth-advantaged in phenotype.

Under all four conditions mutants in 124 genes (core genes) were found to be growth-essential (Fig. 3-3). These encode well-characterised essential functions, including translation, transcription, ATP synthesis, membrane transport, and metabolite biosynthesis. A total of 605 genes were identified whose mutants were core growth-defective (Fig. 3-3), encoding functions including transport, protein chaperones, DNA repair, metabolism and transcriptional regulators. Included in this category are 127 encoding hypothetical proteins of unknown function. An average of 7 mutants were classified as growth-advantaged in each condition, and these genes were distributed on pRL7, pRL9, pRL10 and the chromosome (Appendix A; Table 1). Mutation of a pseudogene on pRL10 (pRL100096) was growth-advantaged specifically on succinate as a carbon source. Pseudogenes are generally considered to be non-functional genomic sequences, however, it is possible for them to encode partial activity, sRNAs or to affect

the mRNA stability of closely-related functional genes [158]. No mutants were growth-advantaged in all growth conditions, consistent with natural selection removing any generally deleterious genes from the population.

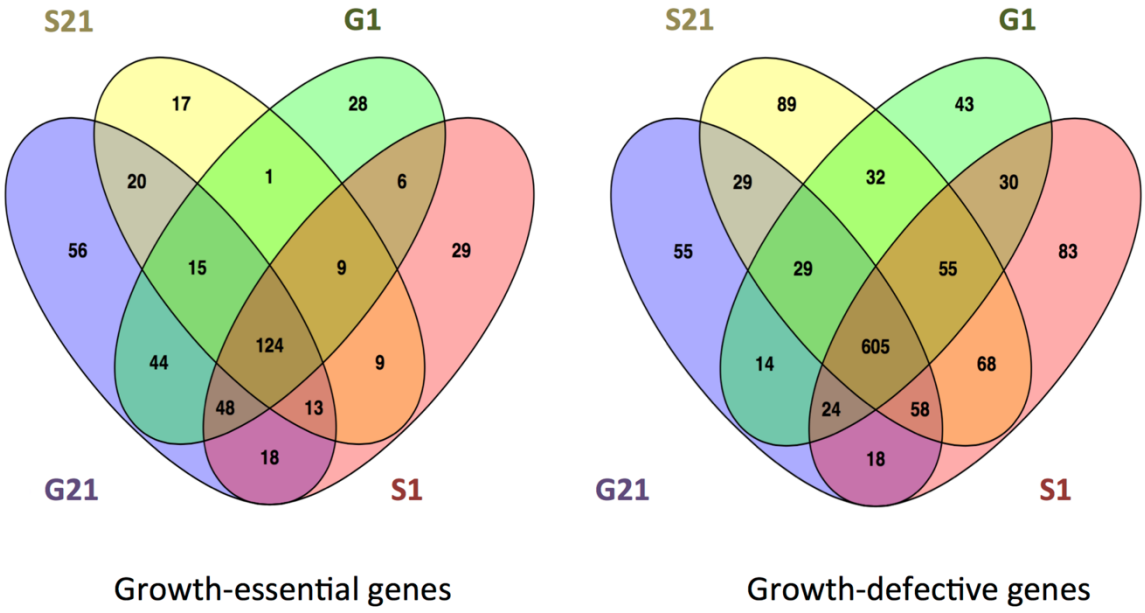


Figure 3-3. Venn diagram showing the number of gene mutants classified as growth-essential and growth-defective under different growth conditions: G21; glucose at 21% O_2, S21; succinate at 21% O_2, G1; glucose at 1% O_2, S1; succinate at 1% O_2.

3.3.4 Literature validation of INSeq predictions

As previously discussed, INSeq identified mutants in 124 genes to be growth-essential under all four conditions (Fig. 3-3). An initial validation of results was given in that these genes encode well-characterised essential functions. For example, twenty-nine of these genes encode ribosomal proteins (30S and 50S); essential for protein synthesis [159]. Another proportion of these genes encode ATP synthase chain components; essential for a functional ATP synthase complex that creates ATP as an energy storage molecule [160]. A putative DNA helicase (RL1551; *dnaC*), a putative DNA polymerase

III (RL0134; *dnaX*), and a putative single-stranded-DNA-specific exonuclease (RL2099; *recJ*) were also all classified as essential. DNA helicases are essential during DNA replication where they function to separate double-stranded DNA [159], DNA polymerases are similarly essential for DNA replication, and DNA polymerase III is involved in the synthesis of DNA from the single template strand [161]. Single-stranded-DNA-specific exonucleases, such as recombinational repair (RecJ), are multifunction enzymes and have been implicated with essential cellular roles across the key processes of repair, recombination and replication [162, 163]. Finally, the plasmid replication 'rep' systems, formed by the *repABC* modules, were also present within this essential group of genes [164]. RepABC systems constitute the rhizobial plasmid maintenance systems, and contain all the elements necessary for plasmid active segregation/partition (*repAB*) and replication (*repC*) within this one module unit [164, 165]. All plasmids in Rlv3841 were identified to contain a number of genes with a growth-essential or growth-defective mutation (Appendix A; Table 1), so it is logical that the replication systems to maintain the plasmids in the cell would also be required.

In addition to identification of these well-characterised essential functions, the mutation of genes in Rlv3841 previously reported in the literature was assessed to help validate INSeq classifications obtained in this study (Table 3-1). Mutations to Rlv3841 genes classified as growth-essential, growth-defective and growth-neutral were all identified to have previously been reported in directed mutational studies (Table 3-1). For the growth-neutral genes listed, relevant growth rates were only reported for RL0940 (*mntH*), so this validation is in the form of a non-lethal phenotype of mutation and no reported growth defects in relevant conditions.

Gene	INSeq phenotype	Literature validation	Reference
RL2578 (*sufS*): putative aminotransferase involved in iron-sulphur cluster biogenesis **RL2582 (*sufB*):** putative iron-sulphur cluster formation subunit protein **RL2580 (*sufC*):** putative ATP-binding component of ABC transporter	Growth-essential (all conditions)	Unable to mutate – lethal phenotype	[166]
RL3424- RL3426 (*dctABD*): C_4-dicarboxylate transport protein system	Growth-defective on succinate, growth-neutral on glucose	Transport of succinate into the cell occurs via the Dct C_4-dicarboxylate transport system in Rlv3841(DctABD).	[35, 94]
pRL120206 (*eryC*): putative erythrulose 4-phosphate dehydrogenase	Growth-neutral (all conditions)	Non-lethal phenotype of mutant – inability to grow on erythritol as sole carbon source	[167]

pRL120201 (*eryF*): putative permease component of ABC transporter CUT2 erythritol transporter **pRL120202 (*eryE*):** putative ATP-binding component of ABC transporter CUT2 for erythritol **pRL120204 (*eryA*):** putative erythritol kinase	Growth-neutral (all conditions)	Non-lethal phenotype of mutant	
RL0101 (gabD2): *putative succinate-semialdehyde dehydrogenase [NAD(P)+]* **RL0102 (*gabT*):** putative 4-aminobutyrate aminotransferase	Growth-neutral (all conditions)	Non-lethal phenotype of mutant	[58]
RL0940 (*mntH*): putative manganese transport protein	Growth-neutral (all conditions)	Mutant has a comparable growth rate to wild-type (in media not limited in Mn^{2+})	[168]
RL2204 (*aapJ*): putative SBP of ABC transporter PAAT (general L-amino acid, Aap system)	Growth-neutral (all conditions)	Non-lethal phenotype of mutant	[169]
RL2975-RL2977:	Growth-neutral (all conditions)	Non-lethal phenotype of mutant	[170]

putative ABC transporter system operon	

Table 3-1. Rlv3841 gene mutant classifications previously reported in the literature and their INSeq classifications identified in this study.

3.3.5 Experimental validation of INSeq predictions

To further validate the accuracy of INSeq gene fitness predictions, mutants in five genes, identified by INSeq as growth-defective on succinate, were selected for experimental validation in this thesis. The *in vivo* mean generation time (MGT) of wild-type Rlv3841 on succinate at 21% O_2 (4.0 h) was compared with that of strains mutated by pK19mob-integration in RL0037 (*pckA*), RL1172, RL3424 (*dctA*), RL4283 (*ptsP*), and RL4284 (*ask*). Each mutant strain had a significantly longer MGT than wild-type Rlv3841 (Table 3-2). The output data from HMM analysis (number of insertion sites, insertion density and mean read count) was analysed for these genes and a fitness value was calculated for each by multiplying insertion density by the number of TA sites, and then by mean read count (Table 3-2) (Insertion density is the proportion of TA sites containing an insertion in the mutant population, and mean read count is the average number of insertions at TA sites containing at least one insertion) (chapter 2; materials and methods).

For four out of the five mutants there was an inverse relationship between fitness value and MGT (Table 3-2); the higher the fitness value, the shorter the MGT. Mutation to RL3424 (*dctA*) was an exception to this, having a reasonably high fitness value (13.28) yet one of the longest MGTs (16.0 h).

Gene	TA sites	Insertion density	Mean read count	Fitness value	MGT (h)
RL0037 (*pckA*)	22	0.09	1.67	3.34	16.0
RL1172	10	0.23	7.42	17.29	5.0
RL3424 (*dctA*)	17	0.27	2.85	13.28	16.0
RL4283 (*ptsP*)	25	0.28	2.15	15.05	5.2
RL4284 (*ask*)	24	0.21	1.08	5.39	6.4

Table 3-2. HMM analysis of five INSeq-identified genes, mutation of which leads to a growth-defective phenotype, and the mean generation time (MGT) on succinate at 21% O_2 of mutants. Insertion density is defined as the proportion of TA insertion sites containing an insertion in the mutant population relative to the HMM predicted value. Mean read count is defined as the mean number of insertion mutants found at insertion sites that contained at least one insertion in the mutant population. Fitness value is a measure of total gene mutants in the population. Mean Generation Time (MGT) is calculated as the time (h) it takes the population to double during exponential growth phase, measured in triplicate. RL0037 (*pckA*) encodes a putative phosphoenolpyruvate carboxykinase, RL1172 encodes a putative transmembrane protein, RL3424 (*dctA*) encodes a putative C_4-dicarboxylate transport protein, RL4283 (*ptsP*) encodes a putative phosphoenolpyruvate phosphotransferase, and RL4284 (*ask*) encodes a putative aspartokinase.

3.3.6 Central Metabolism

3.3.6.1 The Entner-Doudoroff pathway is needed for optimal growth on both glucose and succinate

The Entner-Doudoroff (ED) pathway was shown to be essential for growth on glucose, as expected [171], but surprisingly its mutation caused growth deficiency on succinate (Fig. 3-4 and Appendix A; Table 2). Most free-living rhizobia utilise the ED pathway for the catabolism of glucose to pyruvate [171, 172], with the pathway converting glucose-6-phosphate into pyruvate in a 1:2 mole ratio and generating one molecule of ATP [172]. The requirement for the ED pathway for growth on glucose suggests that sugars made by gluconeogenesis may, at least partly, cycle back through the ED pathway (Fig. 3-4, Table 3-2). Carbon cycling has been reported in a number of other bacterial species including *Sinorhizobium meliloti* [171-173]. RL0753 and RL1315 encode putative glucose-6-phosphate-1-dehydrogenases (Fig. 3-4 and Appendix A; Table 2), with these proteins sharing 43% amino acid identity [143]. However, the essential classification of RL0753 suggests it is the "real" *zwf* with RL1315 unable to compensate. RL0753 was also significantly transcriptionally up-regulated on glucose versus (vs) succinate ($p < 0.05$), while expression of RL1315 was unchanged [55]. Finally, RL0753 is in an operon with genes encoding enzymes catalysing subsequent steps in the ED pathway (RL0751-3). Mutations of all three genes in this cluster are growth-essential on glucose and growth-defective on succinate (Appendix A; Table 2).

Intriguingly, mutation of RL4162 (*eda*) encoding the putative aldolase enzyme catalysing formation of pyruvate and glyceraldehyde-3-phosphate resulted in a growth-neutral classification across all conditions (Fig. 3-4 and Appendix A; Table 2). RL4162 was up-regulated on glucose (vs succinate) [55], but its growth-neutral classification

suggests an alternative aldolase is present [143]. Sugars are also catabolised by the pentose phosphate pathway (PPP) in Rlv3841 [174]. Enzymes catalysing formation of 6-phospho-gluconate from glucose-6-phosphate were essential, but these reactions are shared by both the ED pathway and the PPP (Fig. 3-4 and Appendix A; Table 2) [174]. There were no essential enzymes unique to the oxidative branch of the PPP.

3.3.6.2 The TCA-cycle is essential for growth on glucose and succinate

The chromosome of Rlv3841 contains all the genes for a functional TCA-cycle [52], and most genes encoding TCA-cycle enzymes showed the same INSeq classification on both glucose and succinate (Fig. 3-4 and Appendix A; Table 2). Aconitase (encoded by RL4536, *acnA*) was growth-essential across all four growth conditions (Fig. 3-4 and Appendix A; Table 2), consistent with the TCA-cycle being essential for growth on both glucose and succinate. Koziol et al (2009) reported that stable *acnA Sinorhizobium* mutants are only possible in a citrate synthase (*gltA*) null background suggesting intracellular accumulation of citrate is toxic [175]. Fumarate hydratase (RL2701 and RL2703) and isocitrate dehydrogenase (RL2631, *icd*) catalysing the formation of malate from fumarate and 2-ketoglutarate from isocitrate, respectively, were growth-neutral (Appendix A; Table 2). While RL2701 and RL2703 indicate redundancy of fumarate hydratase, there is only a single gene with homology to isocitrate dehydrogenase (RL2631) [143]. In both *Bradyrhizobium japonicum* and *S. meliloti*, mutation of isocitrate dehydrogenase has no severe effects on growth, and mutants nodulated their respective host plants [176, 177]. In *S. meliloti,* isocitrate dehydrogenase is essential for N_2-fixation [176], however, it is not needed in *B. japonicum* [177]. When the gene encoding isocitrate dehydrogenase is mutated, the pathway of alternative flux is unknown [176, 177].

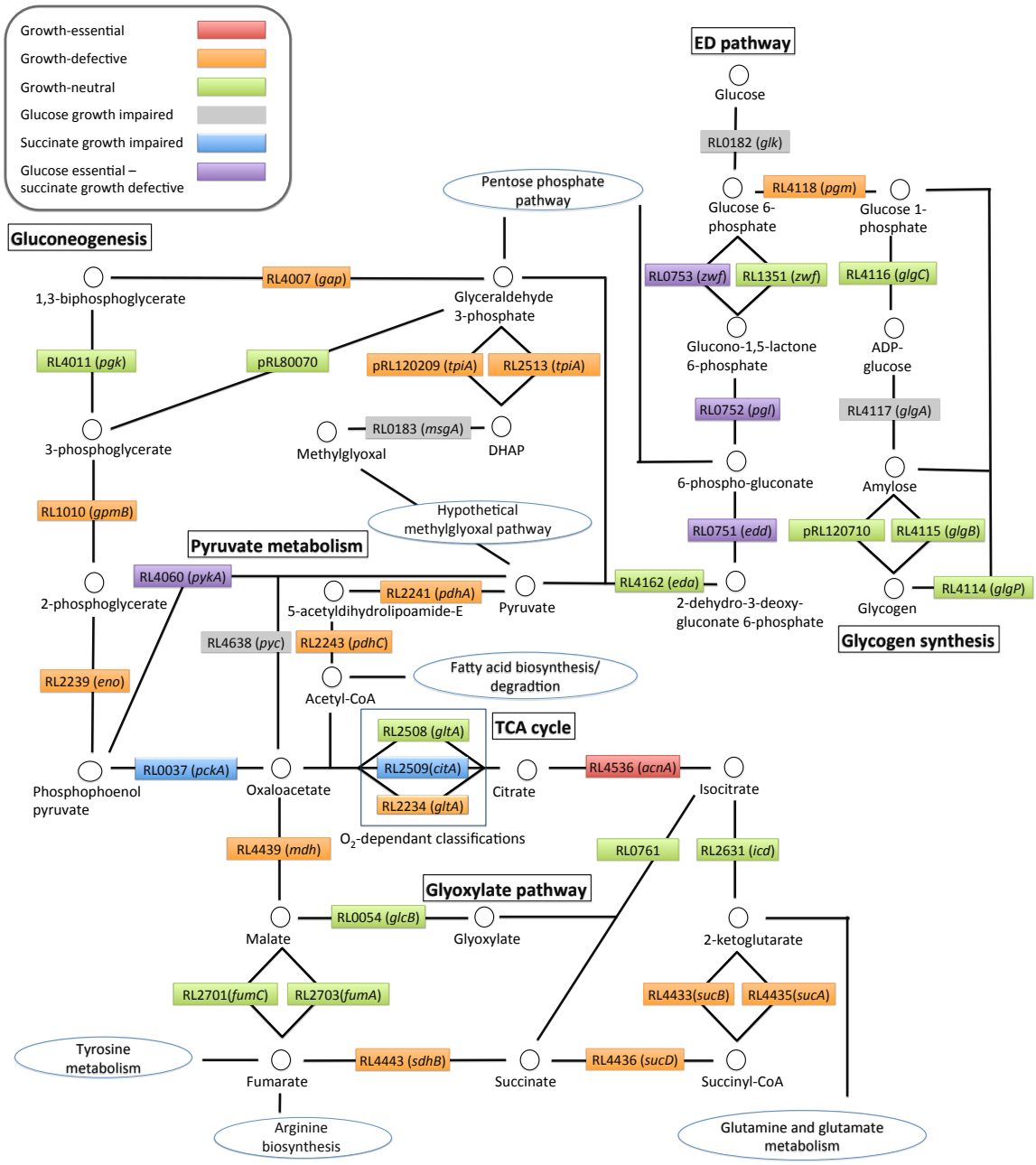


Figure 3-4. Central metabolic pathway of Rlv3841 showing the metabolism of glucose and succinate. Candidate genes for enzymes performing the catalytic steps are shown in coloured boxes according to their INSeq mutant classification; red (growth-essential (ES) under all growth conditions), orange (growth-defective (GD) under all growth conditions), green (growth-neutral (NE) under all growth conditions), grey (growth-impaired on glucose, i.e. growth-essential or growth-defective specifically on glucose), blue (growth-impaired on succinate, i.e. growth-essential or growth-defective

specifically on succinate) and purple (growth-essential (ES) on glucose and growth-defective (GD) on succinate).

Three candidate citrate synthase genes are present in Rlv3841; RL2508, RL2509 and RL2234 [52] (Appendix A; Table 2). On succinate at 21% O_2, mutation of either RL2234 or RL2509 caused a growth-defective phenotype. While on glucose at both 21% O_2 and 1% O_2 only RL2234 was required. Optimum growth on a TCA-cycle intermediate (succinate) requires greater TCA-cycle enzyme activity compared to growth on a sugar such as glucose. This may explain why activity of two citrate synthases is needed when growing on succinate while one is sufficient for growth on glucose. However, at 1% O_2 with succinate, mutation of any of the three-candidate citrate synthases was growth-neutral (Fig. 3-4 and Appendix A; Table 2). The combination of 1% O_2 and succinate results in the longest MGT for Rlv3841 and suggests TCA-cycle activity is severely reduced, minimizing the requirement for citrate synthase isozymes and, presumably, activity (Fig. 3-4 and Appendix A; Table 2). The TCA-cycle generates large amounts of reductant, posing reoxidisation problems at such low O_2 availability. It is therefore likely there is reduced flux through the TCA-cycle at 1% O_2 and bacterial growth is slowed (MGT of 3.7 h on glucose at 1% O_2 vs 6.2 h on succinate at 1% O_2). Reduced flux through the TCA-cycle at low O_2 is strikingly similar to metabolism occurring in bacteroids, particularly of soybean, where mutation of genes encoding even crucial TCA-cycle enzymes does not prevent N_2-fixation [176, 177].

3.3.7 Mutations that specifically reduced growth on glucose at both 1% and 21% O_2

Thirteen genes were required specifically for growth on glucose (i.e. mutation of these genes results in a growth-essential or growth-defective phenotype at both 21% and 1% O_2) (Appendix A; Table 3), and encoded functions span regulation, the cell envelope, small molecule metabolism, macromolecule metabolism and cell processes (Fig. 3-5). The largest proportion are involved in small molecule metabolism (Fig. 3-5), and a number of these genes have known enzymatic functions in central carbohydrate metabolism; RL0182 (*glk*) encodes a glucokinase needed for the first step in glucose activation to glucose-6-phosphate (Fig. 3-4), RL4117 (*glgA*) encodes a glycogen synthase and RL4638 encodes a pyruvate carboxylase (Fig. 3-4). Glycogen synthesis is required for growth on glucose [171], while pyruvate carboxylase is an essential anaplerotic function converting pyruvate to oxaloacetate, for further metabolism in the TCA-cycle [178].

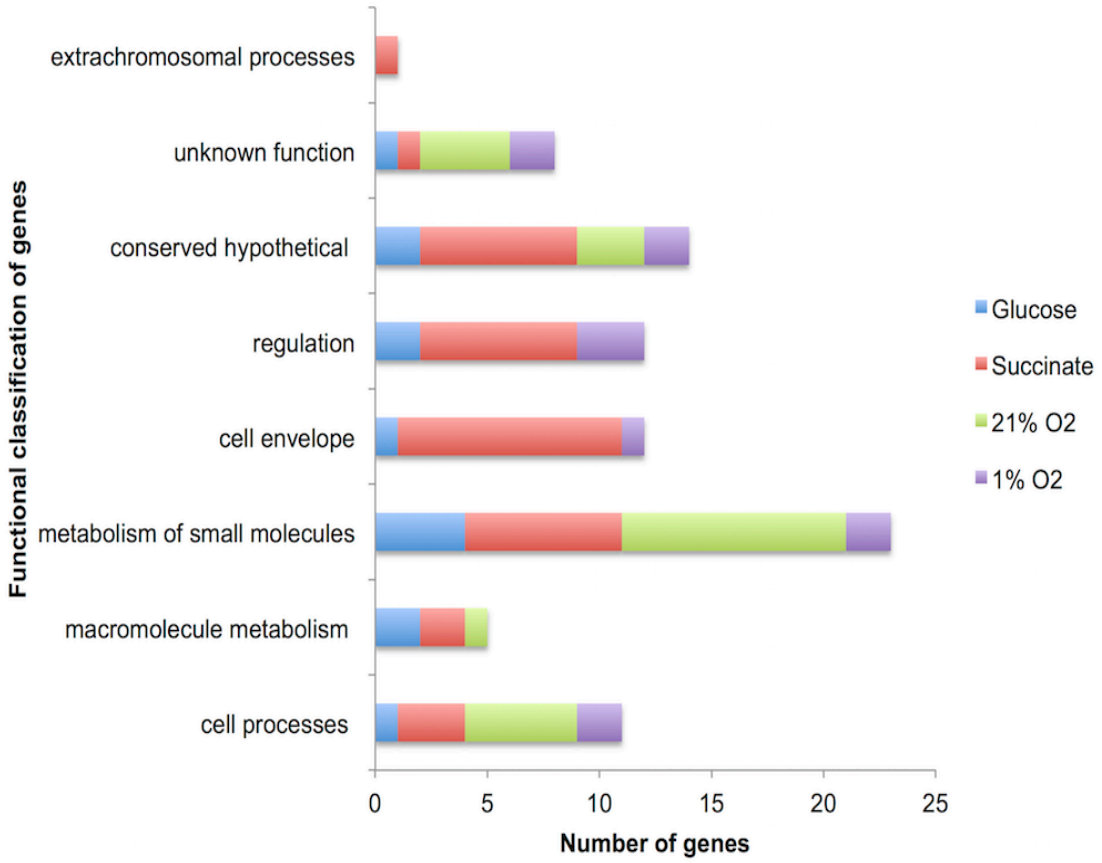


Figure 3-5. Function distribution of the INSeq-identified genes required specifically for growth on glucose (at either 21% or 1% O_2), succinate (at either 21% or 1% O_2), 21% O_2 (on either glucose or succinate), and 1% O_2 (on either glucose or succinate) (listed in Table 3, Table 4, Table 5 and Table 6). Functional classification is based on Riley codes [52].

Growth on glucose required the transcriptional regulators RL0547 (*phoB*) and pRL120207 (*eryD*). pRL120207 (*eryD*) encodes a repressor of the erythritol catabolic operon [179], suggesting unregulated expression of the erythritol operon interferes with glucose catabolism. The requirement of pRL120207 (*eryD*) for growth on glucose was not observed in *S. meliloti* [180, 181]. However, this difference highlights how single mutational studies are not necessarily comparable to INSeq, where mutants are observed

in a competition with thousands of other strains. This can allow resolution of growth defects that may not be evident when the mutant is the sole strain present.

RL0547 (*phoB*) encodes the response regulator of the phosphate-limitation (PhoR-PhoB) two-component regulatory system [182]. Glucose is metabolised via sugar phosphate intermediates and so phosphate and glucose metabolism are linked, which is not the case for dicarboxylate metabolism. Thus the requirement for PhoB for growth on glucose suggests integration with the Pho regulon may be due to the synthesis of phosphorylated intermediates. [183, 184].

Rlv3841 required the methylglyoxal pathway alongside the ED pathway for optimal growth on glucose. The methylglyoxal pathway is present in some prokaryotic species, converting dihydroxyacetone phosphate (DHAP) derived from glycolysis or the ED pathway into methylglyoxal, and then into pyruvate [185]. This pathway is activated by increased concentrations of glucose-phosphates in *E. coli* [186]. It bypasses the 3-phosphoglyceraldehyde dehydrogenase reaction carried out by RL4007 (*gap*) (3-phosphoglyceraldehyde + NAD^+ + P_i = 1,3-diphosphoglyceric acid + NADH) which consumes inorganic phosphate. This suggests the pathway reduces sugar phosphate concentrations and increases inorganic phosphate without generating ATP [186].

I propose that the methylglyoxal pathway may run in parallel to the ED pathway in Rlv3841 to consume excess glucose-6-phosphate and generate pyruvate (Table 3-3), consequently reducing the concentration of phosphorylated intermediates generated in glucose catabolism. This concept is supported by the clustering of methylglyoxal synthase (RL0183) and glucokinase (RL0182). While it is possible polarity effects from

mutating the gene encoding methylglyoxal synthase may contribute some of the defects to glucose catabolism; the ED and methylglyoxal pathways are clearly linked. A functional methylglyoxal pathway has not previously been identified in rhizobia, however, the committed enzyme in the methylglyoxal pathway, methylglyoxal synthase (RL0183), was required for growth specifically on glucose (Fig. 3-4 and Appendix A; Table 2). Microarray data shows that expression of RL0183 is 2-fold induced on glucose compared to succinate-grown cells ($p < 0.01$) [57]. Bioinformatic analysis indicates that Rlv3841 possesses all necessary genes for a functional methylglyoxal pathway [143, 182] (Table 3-3), and these enzymes are well-integrated into general central metabolism. Mutations in many of the genes encoding enzymes of this pathway were growth-defective on glucose (Table 3-3). The final two steps of the methylglyoxal pathway carried out by methylglyoxal reductase and lactate dehydrogenase have multiple potential candidate genes, which could explain why their mutation is growth-neutral (Table 3-3).

Compound	Enzyme(s)	INSeq classification
Glucose		
↓	Glucokinase (RL0182)	ES
Glucose 6-phosphate		
↓	Via pentose phosphate pathway Via ED pathway	ES
Glyceraldehyde 3-phosphate		
↓	Triosephosphate isomerase (pRL120209, RL2513)	GD
Dihydroxyacetone phosphate		
↓	Methylglyoxal synthase (RL0183)	GD
Methylglyoxal		
↓	Methylglyoxal reductase (pRL120760, RL3064, RL3243)	NE
Lactate		
↓	Lactate dehydrogenase (pRL120231, pRL120601, RL0444, RL3578)	NE
Pyruvate		

Table 3-3. Proposed Rlv3841 methylglyoxal metabolic pathway. INSeq classifications on glucose given as ES; growth-essential, GD; growth-defective, NE; growth-neutral.

Two further plasmid genes were required for growth on glucose; pRL110565 encoding a putative transmembrane protein, and pRL120419 (*opaA*) encoding an omega-amino acid:pyruvate transaminase [187]. The transmembrane protein possesses a conserved MatE domain, which primarily function in the export of metabolic and xenobiotic organic cations [143]. OpaA synthesizes beta-alanine by the alanine-dependent transamination of pyruvate [58], which is a key intermediate in glucose catabolism. This may help regulate pyruvate levels by buffering with a non-proteinogenic amino acid.

3.3.8 Mutations that specifically reduced growth on succinate at both 21% and 1% O_2

C_4-dicarboxylates, including succinate, are provided to bacteroids by the plant host during symbiosis to fuel N_2-fixation [18]. Thirty-eight genes were required specifically for growth on succinate (i.e. mutants are growth-essential or growth-defective on succinate at both 21% and 1% O_2) (Appendix A; Table 4). Functional classification of these genes shows the largest proportion encode cell envelope related proteins (Fig. 3-5). These included genes encoding proteins within the PrsD-PrsE type I secretion system (RL3657), genes involved in exopolysaccharide (EPS) synthesis (RL4658) and genes encoding transmembrane proteins (RL0423, RL1391, RL1392). Bioinformatic analysis indicates RL3654-RL3657 all have potential functional roles in the cell envelope [143]. RL3654 encodes a polysaccharide biosynthesis protein, RL3655 a glycosyltransferase, RL3656 a lipase and RL3657 (*prsE*) encodes the permease component of the PrsDE Type I secretion system. PrsDE exports the EPS-glycanases PlyA and PlyB, which are responsible for correct cleavage of EPS [188, 189]. PrsDE is also responsible for export of rhizobial adhesions and other proteins to the outer surface.

While mutation of the ATP-binding component of this transporter (RL3658, PrsD) had no significant effect on fitness, microarray data shows RL3658 (*prsD)* was significantly up-regulated on succinate vs glucose ($p > 0.05$) [55]. However, three genes (RL0072, RL0181 and RL0623) encode homologs with 33-38% identity to PrsD, suggesting functional redundancy [143, 189].

In addition, of the four genes RL1390-3 required for growth on succinate, three have predicted functional roles in the cell envelope, with both RL1391 and RL1392 encoding putative transmembrane proteins, and RL1393, a putative peptidoglycan biosynthesis protein (Table 5). RL4658 (*exoB*) encodes UDP-glucose-4-epimerase required for synthesis of UDP-galactose [190], which is incorporated into the repeating units of EPS and other polysaccharides [191]. The high number of cell envelope genes required for growth on succinate suggests that restructuring of the cell surface, involving the PrsD-PrsE type I secretion system and UDP-galactose production in particular, is required to enable growth on organic acids.

Genes encoding five transcriptional regulators were required for growth on succinate: pRL100388, RL1641, RL1642, RL3427, and RL4524 (*ecfL*) (Appendix A; Table 4). pRL100388 encodes a putative LacI-family transcriptional regulator with homology to a gluconate utilisation system transcriptional repressor found in many prokaryotes [143, 192]. RL1641 and RL1642 encode two-component sensor/regulator units belonging to the MerR family of transcriptional regulators. RL3427 encodes an AsnC-family transcriptional regulator, which could possibly be involved in regulation of genes encoding the dicarboxylic acid transport (Dct) system located adjacently. As expected, mutants in the dicarboxylic acid transport system (*dctABD*) (RL3424-6) were growth-

defective on succinate (Appendix A; Table 4). Mutations in the *dct* system have been shown to impair uptake of succinate into cells leading to growth arrest [33, 35, 94, 193]. RL4524 (*ecfL*) encodes an extracytoplasmic function (ECF) sigma factor [194]. ECF sigma factors regulate transcription in response to extracellular stimuli [195]. Their activity is commonly under the regulation of an anti-sigma factor (ASF) but RL4524 is one of only four Rlv3841 ECF genes not adjacent to an ASF gene. RL4524 expression increased in the presence of pea root exudate, presumably due to an unknown compound [194], and was significantly up-regulated in the pea rhizosphere (vs Rlv3841 grown on glucose/ammonia) [57]. From results shown here, it may be that EcfL (RL4524) responds to C_4-dicarboxylic acids present in pea root exudates.

Mutation of two genes involved in nucleobase synthesis, pRL90208 and RL3468, were growth-deficient and growth-essential, respectively, on succinate. pRL90208 (*codA*) encodes a putative cytosine deaminase that functions in pyrimidine metabolism [182]. Both CodA and the protein product of RL0891 are suggested to possess deaminase activity utilised in conversion of cytosine to uracil, and 5-methylcytosine to thymine [182]. INSeq analysis showed RL0891 to be to be non-essential, suggesting its product is unable to compensate for the activity of CodA, whose gene insertion mutation was growth-defective on succinate. RL3468 (*prs*) encodes a ribose-phosphate pyrophosphokinase catalysing biosynthesis of 5-phosphoribosyl diphosphate (PRPP) from ribose 5-phosphate. PRPP is utilised in both purine and pyrimidine biosynthesis, with the first committed step to purine biosynthesis utilizing PRPP as a substrate [196]. This suggests that, for optimal growth on succinate, there is a requirement for increased nucleotide recovery or synthesis compared to growth on glucose. RL0037 (*pckA*) encodes phosphoenolpyruvate carboxykinase that converts oxaloacetate into

phosphoenolpyruvate which is essential for gluconeogenesis (Fig. 3-4) [150]. RL2990 (*ubiA*) encodes a 4-hydroxybenzoate octaprenyltransferase required for ubiquinone biosynthesis and electron transport. Cells growing on a TCA-cycle intermediate are likely to have a greater requirement for reoxidation of $NADH/FADH_2$ by the electron transport chain. RL1561 (*pabC*) was also found to be required, and the gene encodes an enzyme in the shikimate pathway also needed for ubiquinone synthesis. This agrees with a greater need for ubiquinone in succinate grown cells.

3.3.9 Mutations that reduced growth at 21% O_2 on both glucose and succinate

Twenty-three genes were required for optimal growth at 21% O_2 on both carbon sources (i.e. mutants are growth-essential or growth-defective on both glucose and succinate at 21% O_2) (Appendix A; Table 5), and encoded functions spanning across small molecule metabolism, macromolecule metabolism, and cell processes (Fig. 3-5). The largest proportion are involved in small molecule metabolism (Fig. 3-5), and eleven genes are predicted to encode enzymes utilizing oxygen: cytochrome c oxidase cytochrome aa3 subunits (RL1021, RL1022), a deoxygenase (RL0802), a hydroperoxide resistance protein (RL2927), a peroxiredoxin (RL2440), three oxidoreductases (RL0847; an inosine-5'-monophosphate oxidoreductase, RL3834; an ErfK oxidoreductase, and RL4186), a mitochondrial respiratory chain complex assembly factor (RL0920) and two adenylate cyclases (RL2441, RL2926). None of the proteins encoded by these genes are required at 1% O_2 presumably as their function is to protect against oxygen toxicity.

Other genes required for optimal growth on 21% O_2 include RL0921, predicted to encode a CorA magnesium transporter, and RL0922, a potassium uptake protein [143]. Cationic transporters are electrochemical potential-driven and activity is dependent on

the oxidation state of the metal ion species [197]. Insertion mutation of the genes encoding these enzymes did not impair growth at 1% O_2, perhaps because transport of ions in their more reduced state is not limiting growth or because other transporters play a greater role at low O_2. Rlv3841 possesses an MgtE magnesium transporter channel (RL1461 (*mgtE*)) that is essential for growth when both magnesium concentration is limiting and pH is low, which is also essential for N_2 fixation on specific legumes [198]. CorA may be the primary transporter of magnesium at 21% O_2 and another transporter, such as MgtE, may play a greater role at decreased O_2.

RL3597 encodes a putative RNA helicase protein, involved in messenger RNA biogenesis, and RL1548 (*radA*) encodes a putative DNA repair protein [143, 199]. The observation that their mutants are growth-defective is presumably due to their preventative role in oxidative damage to nucleic acids.

3.3.10 Mutations that reduced growth at 1% O_2 on both glucose and succinate

Twelve genes were required for optimal growth at 1% O_2 on both succinate and glucose (i.e. mutants are growth-essential or growth-defective on both glucose and succinate at 1% O_2) (Appendix A; Table 6), and encoded functions spanning across regulation, the cell envelope, small molecule metabolism and cell processes (Fig. 3-5). The largest proportion are involved in regulation, and three potential low O_2 growth regulators were identified: pRL110072, RL0546 (*phoU*) and RL4042. pRL110072 encodes a putative GntR family transcriptional regulator and RL4042 encodes a putative AsrR family transcriptional regulator.

RL0546 (*phoU*) is annotated to encode a putative phosphate transport regulatory protein. While the role of PhoU in *R. leguminosarum* has not been investigated, it has been well-studied in *E. coli,* where the protein forms part of the phosphate-sensing pathway that negatively regulates the Pho regulon [200]. Mutants deficient in *phoU* show constitutive expression of Pho regulon genes [200-202], encoding proteins involved in phosphate metabolism, energy production, membrane transport, flagella synthesis and transcription factors, including a stringent response protein [203]. However, it is not known how similar the mechanism of PhoU regulation is between *E. coli* and Rlv3841. PhoU may be involved in growth regulation under low O_2 through a role in the coordination of phosphate and energy metabolism. The generation of ATP will be dependent on the transport of phosphate in the cell, and energy metabolism is O_2-dependant.

While RL2585 is annotated as NodX, the presence of which enables a few strains of *Rhizobium* to nodulate Afghanistan peas [143], the protein encoded by RL2585 shows no significant identity to the characterised NodX (UniProtKB - P08888) [143]. However, RL2585 possesses a conserved acyltransferase domain and that might be required for modification of a surface-exposed sugar. RL4693 encodes a putative polysaccharide deacetylase protein, RL4694 a putative dinucleoside polyphosphate hydrolase, and RL4695 a putative bacterioferritin. RL1506 encodes RelA, a central stringent response protein controlling the synthesis of (p)ppGpp required for adaptation to stress conditions. RelA coordinates responses to the abundance of nitrogen and carbon in *Rhizobium* [204]. In addition to these roles, RelA is suggested to regulate the stress response to O_2-limitation in *B. subtilis* [205], with a *relA -/-* mutant strain unable to synthesize the heat shock proteins induced in response to O_2-limitation in the wild-type [205]. The exact mechanism of this regulation is unclear, and does not appear to have

been studied in any other bacterial species. However, the requirement of RL1506 (*relA*) for growth specifically at low O_2 suggests RelA also regulates a stress response resulting from O_2-limitation in Rlv3841.

3.3.11 Mutations that specifically reduced growth on succinate at 1% O_2

There were a number of mutations resulting in a growth-essential or growth-defective phenotype specific to one condition, but were growth-neutral across all other conditions. Fifty-two mutations were specific to glucose at 21% O_2, 41 specific to succinate at 21% O_2, 29 specific to glucose at 1% O_2, and 65 specific to succinate at 1% O_2. The combination of 1% O_2 and succinate is most relevant to symbiotic conditions with 64 genes classified as causing a growth-defective phenotype upon insertion mutation and a single growth-essential gene, RL2393 (*glnB*). The MGT of Rlv3841 on succinate at 1% O_2 was almost double that of the other three growth conditions (6.2 h compared to 3.7 h, 3.8 h and 4 h) (Fig. 3-1). Under stress Rlv3841 may be particularly susceptible to mutations altering growth, possibly explaining the higher number of mutations that specifically effected growth under this condition.

The sole gene identified as growth-essential on succinate at 1% O_2, RL2393 (*glnB*), encodes the nitrogen regulatory protein, PII. The PII protein acts as an intracellular signal transmitter and regulates activity of the nitrogen regulatory NtrB/NtrC two component system that regulates nitrogen-responsive proteins, including glutamine synthetase. Null mutations in *glnB* cause a constitutively activated nitrogen-regulated response in *R. leguminosarum*, promoting increased synthesis of glutamine [206, 207]. It is intriguing that only on succinate at 1% O_2, conditions similar to that experienced by bacteroids, is the activity of RL2393 (*glnB*) essential. GlnB is down-regulated during

bacteroid differentiation [208], and microarray data shows expression of *glnB* is increasingly down-regulated from 7 to 28 days in pea bacteroids, suggesting that GlnB is only required prior to symbiotic N_2 fixation [57]. Constitutive activation of N_2 assimilation in a RL2393 (*glnB*) mutant may cause too much 2-ketoglutarate, used for glutamate synthesis, to be removed from the TCA-cycle. Depletion of 2-ketoglutarate would force a turn of the TCA-cycle to replace it and lead to the production of extra reductant that must then be reoxidised. In a 1% O_2 environment, reoxidation of NADH via the electron transport chain may be limited by the O_2 available.

3.3.12 A plasmid specific toxin-antitoxin system on pRL10

Mutation of genes in the putative operon pRL100005-10 were growth-essential or growth-defective across all four conditions and the operon is predicted to encode a putative toxin-antitoxin system [143]. The operon consists of two candidate toxin genes (pRL100005 and pRL00009) and three candidate antitoxin or transcriptional regulator component genes (pRL00006, pRL00007, and pRL100010) (Dataset S1; (ref my paper) [143]. The two candidate toxin genes encode PIN-domain proteins, which commonly confer toxicity due to ribonuclease activity and require co-expression of anti-toxin components to ensure cell viability [209]. Finding a novel toxin-antitoxin system on pRL10 accords with the ability to cure Rlv3841 of each plasmids, with the exception of pRL10 [210]. Consideration of this novel toxin-antitoxin system is important for if trying to generate new plasmid-less rhizobial strains.

3.4 Conclusion

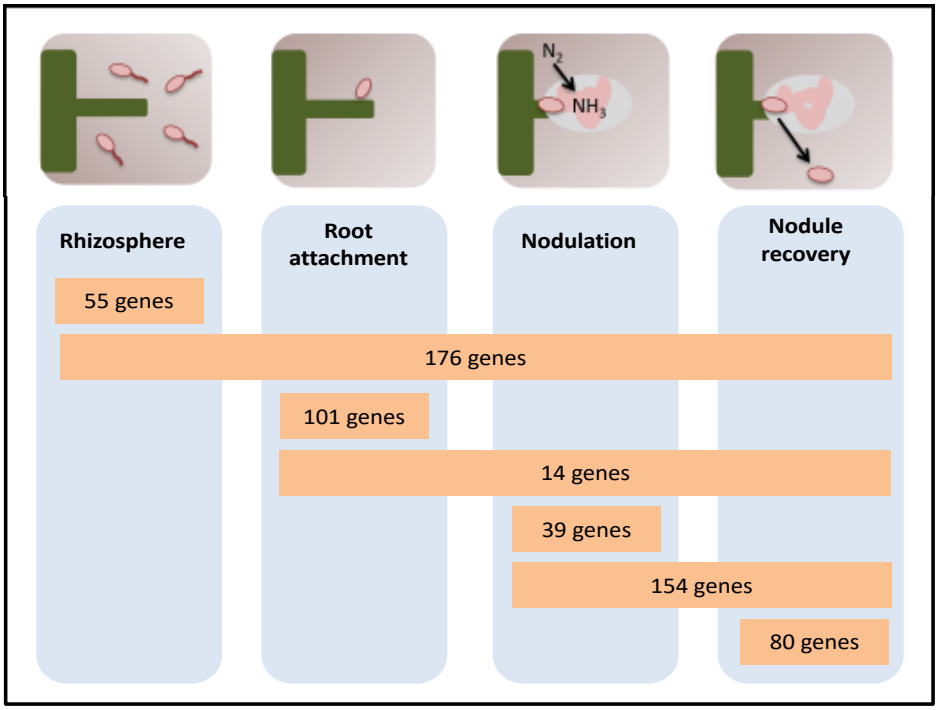
In summary, an INSeq screen of Rlv3841 grown on glucose and succinate at both 21% and 1% O_2 enabled identification of factors needed for growth on different carbon sources and O_2 levels. Use of HMM analysis assigned gene phenotype classifications to 7316 genes in the genome of Rlv3841 across four different growth conditions. Analysis of central carbon metabolism pathways allowed roles in glucose and succinate metabolism to be distinguished and the ability of pathways to compensate for single enzyme mutations to be phenotypically quantified. A hypothetical methylglyoxal pathway is proposed in Rlv3841, and enzymes of this pathway were found to be required for growth on glucose. One of the most dramatic observations is that only mutation of RL2393 (*glnB*) caused a growth-essential phenotype on succinate at 1% O_2. Since glnB mutants of *R. leguminosarum* are constitutively activated for N_2 assimilation this suggests removal of 2-ketoglutarate from the TCA-cycle by glutamate synthesis is highly deleterious when C_4-dicarboxylates are the carbon source in a low O2 environment, conditions mimicking those of N_2-fixing bacteroids in symbiosis with legumes.

This investigation has allowed optimisation and validation of the *Rhizobiaceae*-compatible INSeq methodology with simple *in vitro* experimental conditions. The accuracy of INSeq gene fitness predictions have been validated with a comparison against well-characterised essential gene functions, gene mutational phenotypes previously reported in the literature and directed mutational studies carried out in this work. Mutants in five genes, identified by INSeq as growth-defective on succinate, were selected for experimental validation. The *in vivo* MGT of these strains vs. wild-type Rlv3841 on succinate at 21% O_2 supporting the INSeq growth-defective predictions

(Table 3-2). Furthermore, the power of INSeq in resolving fitness differences between genes was also demonstrated by the relationship between INSeq-calculated fitness values and *in vivo* MGT of the mutant. The success of the INSeq screen in microbial culture provides confidence for moving forwards onto investigating more complex *in planta* conditions.

Chapter 4

Genome-scale characterisation of the fitness determinants in the early stage of the *Rhizobium leguminosarum* symbiosis



4.1 Abstract

Insertion sequencing (INSeq) of *Rhizobium leguminosarum* bv. *viciae* 3841 (Rlv3841) grown in the pea rhizosphere and attached to roots was used to characterise the fitness determinants in the early stages of symbiosis. This project was done in collaboration with Li Li and Vinoy Ramachandran who performed an INSeq analysis of the later stages of symbiosis; characterising the fitness determinants for nodulation (bacteroids) and recovery from nodules (nodule bacteria). An INSeq database characterising the mutational phenotypes for 7316 genes in the genome of Rlv3841 across nine different *in vitro* and *in planta* INSeq conditions screened over the course of this project was made. Comparison across this database allowed identification of 395 genes that form the core fitness requirements in Rlv3841, mutation of which is growth impairing regardless of condition. A total of 463 genes were identified that are required to successfully nodulate pea: 176 genes required from the rhizosphere to nodulation, 14 genes required from root attachment to nodulation, 39 genes required specifically for bacteroid development, 154 genes required in both bacteroids and nodule bacteria, and 80 genes required specifically in nodule bacteria. In addition to these genes, there was identification of 55 genes required specifically for growth in the rhizosphere and 101 genes required specifically for root attachment. Directed mutations were made to eleven transcriptional regulators identified as being required for root attachment. RNA sequencing of these strains alongside wild-type Rlv3841 was carried out with the aim of identifying their directly regulated targets and mapping the transcriptional regulatory pathways required for this highly coordinated process of root attachment.

4.2 Introduction

Rhizobia experience a number of different lifestyle changes in working towards engaging in symbiosis; from free-living microbes growing in the rhizosphere, to attaching to the roots, to colonizing infection threads and finally terminally differentiating into N_2-fixing symbionts. The rhizosphere is the zone of soil immediately surrounding plant roots [62], and is under strong influence of root exudates and the secreted products of bacterial metabolism and degradation. Rhizobia must be able to thrive in this environment, defending against the oxidative root environment and metabolising carboniferous root exudates as an energy source [63, 66], and migrate via chemotaxis towards the roots.

A complex molecular dialogue takes place prior to attachment, in which rhizobia can respond to symbiosis initiating signals, such as flavonoids, to initiate migration [23]. Root attachment is the first physical interaction between prospective symbiotic partners, and is not only an essential step for symbiosis, but also carries the significant benefit of anchoring bacteria in this nutrient-rich rhizosphere environment [68]. To initiate symbiosis, rhizobia must attach specifically to the tips of developing root hairs [29]. Attachment can also occur at the surface epidermis of the root [23, 68, 211], although this is not known to lead to nodulation.

A common biphasic mechanism of attachment exists in agriculturally important rhizobacteria, including in Rlv3841 [68]. Biphasic root attachment is comprised of a primary attachment phase, characterised by a weak, reversible and non-specific binding of bacteria to the root surface, followed by a secondary attachment phase, characterised by a switch to a stronger, more specific binding mode, and the synthesis of extracellular

fibrils [68]. Two soil pH-dependent primary attachment mechanisms have been characterised in Rlv3841. In acidic soil, attachment is mediated by rhizobial surface polysaccharide glucomannan binding to root hair expressed lectin [23, 69]. Under alkaline conditions, root lectins will be solubilised, and this method of attachment utilizing glucomannan no longer functions [23, 69]. The mechanism of attachment under alkaline and neutral pH is suggested to involve an extracellular rhizobial protein 'rhicadhesin', although this remains unproven [72, 73]. Thirty-years on from the original suggestion of rhicadhesin-mediated attachment, neither the gene encoding rhicadhesin nor the protein itself has been identified [23, 68, 72, 73]. Exopolysaccharides (EPS) are known to play an important secondary attachment role in *R. leguminosarum,* aiding the formation of rhizobial aggregates as a 'cap' on the root hair [23]. Other secondary attachment factors in *Rhizobium* are suggested to include cellulose fibrils, lipopolysaccharides and secreted proteins [23, 212]. The mechanism of attachment under alkaline and neutral pH is unclear and much remains to be understood about how primary and secondary attachment are coordinated in *Rhizobium*.

Following successful attachment at the tip of a developing root hair, nodulation signalling induces root hair curling, causing the attached *Rhizobium* to become entrapped between the two root hair cell wells [29]. From this site, the plant cell plasma membrane can invaginate to form an infection thread, which allows the rhizobia to traverse the length of the cell. Division of rhizobia at the growing infection thread tip generates the forwards force for infection thread progression [77]. Infection threads deliver rhizobia to the root cortex, where the host plant cells have generated the lateral organ nodule. Rhizobia will be internalised by nodule cells and undergo a series of terminal differentiation processes, the culmination of which is the generation of

bacteroid cells, and the production of nitrogenase and other accessory proteins that allow N_2-fixation. Pea (*Pisum sativum*) nodulating rhizobia, such as Rlv3841, form indeterminate nodules on the roots. Indeterminate nodules are elongated and cylindrical with distinct zones of development and N_2-fixation [24]. These nodules contain a gradient of rhizobial differentiation processes and both bacteroids and undifferentiated nodule bacteria will be present in this space.

Two major global aims in the field are to improve the existing symbiosis and to extend the symbiosis into non-legume crops. In working towards these aims it has become apparent that there are a number of limiting factors that first need to be addressed, most significantly, to gain a better understanding of the genetic requirements to successfully engage in symbiosis. While nodulation and N_2-fixation have received a great deal of attention [30, 31, 145, 150, 166, 207], how rhizobia thrive in the rhizosphere and attach to roots has received less focus. Previous research has identified a number of individual factors involved in root attachment and rhizosphere colonisation [23, 55, 57], but much remains unclear about the molecular mechanisms governing attachment and the genetic arsenal required to successfully navigate the rhizosphere and compete for the root. Through the application of *Mariner* transposon insertion sequencing (INSeq) to characterise the fitness determinants required for Rlv3841 to engage in symbiosis with pea (*Pisum sativum*), I aim to improve on this understanding of the genetic basis of symbiosis.

Here I describe the use of INSeq in the whole genome-scale characterisation of the fitness determinants in early stage symbiosis [98]: growth in the rhizosphere and attaching to roots. This work has been carried out in collaboration with an INSeq

analysis of the fitness determinants in later stage symbiosis: for nodulation (bacteroids) and recovery from the nodules (nodule bacteria) (with Li Li and Vinoy Ramachandran, Poole Lab). The INSeq investigation of growth in the rhizosphere and root attachment is entirely my own, so shall be the focus of discussion in this thesis. In the INSeq analysis of nodulation and nodule recovery, I contributed to the experimental design and carried out the insertion library preparation and sequencing. However, the mammoth task of retrieving insertion mutants from the nodules of 1,500 plants (picking ~150, 000 nodules!) was carried out by Li Li, a visiting researcher to the Poole Lab. The results of this second INSeq analysis will be included for defining fitness determinants and to give a global perspective of symbiosis, with a focus on central metabolism pathways in particular.

4.3 Results and discussion

4.3.1 Colonisation assays of Rlv3841 grown in the pea rhizosphere and attached to roots

Colonisation assays of Rlv3841 grown in the pea rhizosphere and attached to roots were carried out to determine the optimal time point to extract insertion libraries for INSeq analysis (chapter 2, materials and methods). An input inoculum of 10^5 cells per plant was chosen, as this would allow at least 500-fold coverage of the approximately 7300 genes in the genome while keeping the number of plants below 100 (Table 4-1).

Input inoculum (cfu)	Plants needed for 500-fold coverage of 7300 genes in genome
10^4	365
10^5	36.5
10^6	3.65

Table 4-1. Number of plants needed for 500-fold coverage of 7300 genes in genome based on input inoculum size (cfu; colony-forming units).

The population in the rhizosphere reached saturating capacity around 5 days post inoculation (dpi) at $2.13 \times 10^8 \pm 2.2 \times 10^7$ colony-forming units (cfu) (Fig. 4-1). The root attached population also stabilised around 5 dpi at $1.42 \times 10^7 \pm 2.8 \times 10^6$ cfu (Fig. 4-1). Five dpi was therefore chosen as the optimal time point to extract insertion libraries from the rhizosphere and roots. It is the first point of maximal population saturation, important for obtaining sufficient DNA to give good coverage of the mutant library, and will allow resolution of mutant phenotypes over of a period of selection while still ensuring early-stage symbiosis is a focus.

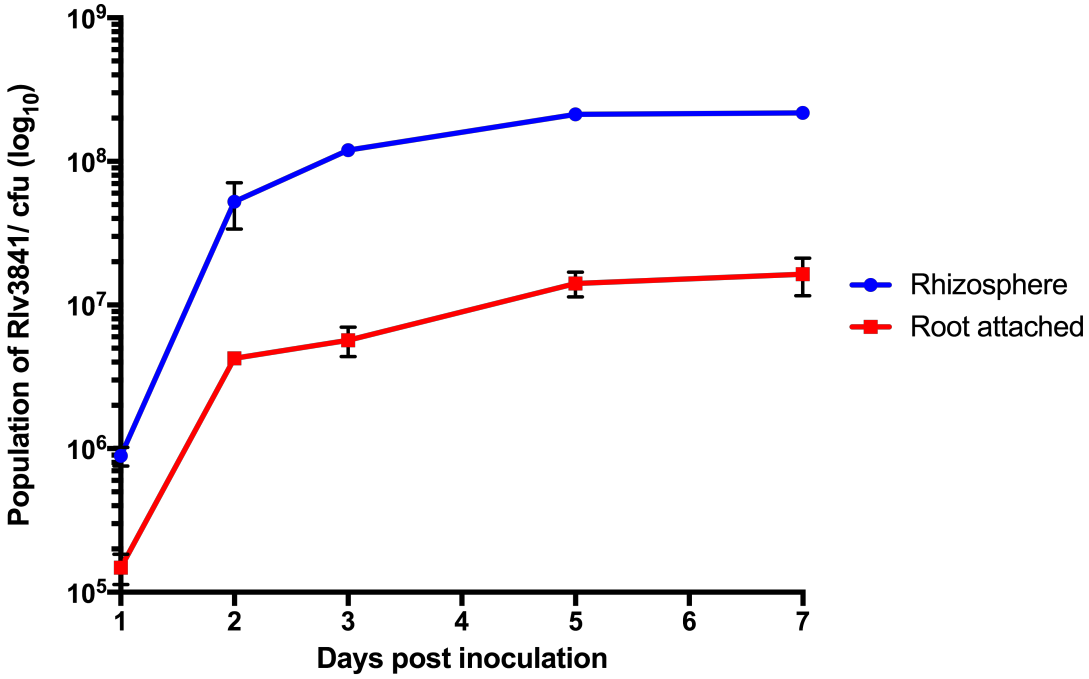


Figure 4-1. Mean population of Rlv3841 in the rhizosphere and root attached over 7 days post inoculation from a 10^5 inoculum (cfu; colony-forming units) calculated from 5 independent biological replicates. Error bars show standard error of the mean.

4.3.2 INSeq experimental design

4.3.2.1 Rhizosphere and root attachment

An INSeq analysis of growth in the rhizosphere and root attachment was designed to enable fitness determinants to be characterised in these important early symbiosis stages. A 10^5 population of insertion mutants was inoculated across 50 seven-day-old pea seedlings, giving a total input inoculum of 5 x10^6 cells (Fig. 4-2). This number equates to 36-fold coverage of the 140,056 potential 'TA' insertion sites in Rlv3841, and over 600-fold coverage of the ~7300 genes in the genome. The pH of this environment was measured to be neutral (pH 7.0). Insertion mutants were retrieved from the roots and the rhizosphere at 5 dpi, samples across the 50 plants were pooled, and three libraries

underwent barcoding, *Mariner* library preparation and sequencing: 1) the input library, 2) the rhizosphere library and 3) the root attached library (Fig. 4-2).

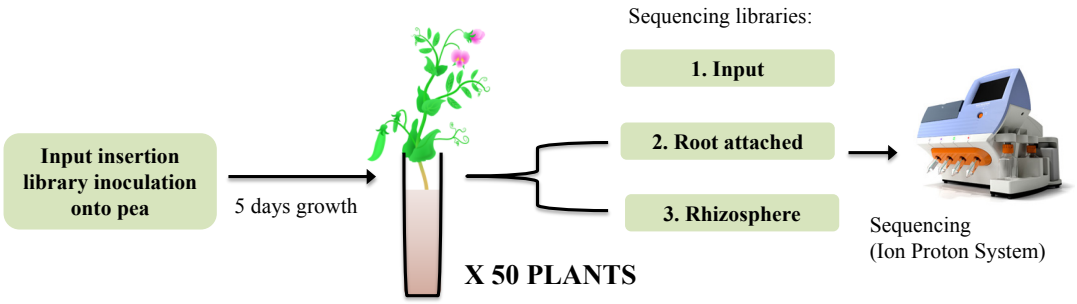


Figure 4-2. INSeq experiment design consisting of 10^5 *Mariner* insertion mutants (generated from six independent library pools) inoculated onto seven-day old pea seedlings and retrieved from the rhizosphere and root attached fractions at 5 days post inoculation for a total of 50 plants following the protocol described for the rhizosphere and root attachment assays (chapter 2, materials and methods). Extractions were spun at 1000 rpm to pellet large plant matter and vermiculite. The supernatant was then grown on Tryptophan Yeast (TY) for 12 hours to decrease plant DNA contamination and increase bacterial DNA for gDNA extraction. After 12 hours, cultures were spun at 4500 rpm for 20 minutes to pellet transposon mutant cells, and these samples underwent DNA extraction, library preparation and sequencing, and transposon insertion analysis with a four-state HMM (chapter 2, materials and methods).

4.3.2.2 Nodulation and nodule recovery

The INSeq analysis on the Rlv3841 mutant library forming nodules on peas was designed to enable fitness determinants to be characterised from two distinct

populations: nodulation (bacteroids) and nodule recovery (nodule bacteria). For this study, each pea seedling was inoculated with a population of 10^6 *Mariner* transposon mutants and allowed to grow for 28 dpi. At 28 dpi, a total of ~150,000 nodules were harvested from 1500 pea plants (performed by Li Li). Two pools were separated from this: DNA was extracted directly from the nodules, to represent nodulation (i.e. bacteroids), and the nodule extracts were then regrown in TY media for 12 hours and DNA was extracted from this sample to represent nodule recovery (i.e. nodule bacteria). The nodule bacteria retrieved from nodule recovery will represent the non-terminally differentiated bacteria present in indeterminate pea nodules. This will include bacteria that have recently been released from infection threads.

4.3.3 HMM genome analysis

INSeq sequencing data analysis was carried out using a Hidden Markov Model (HMM). The HMM assigns genes to one of four classification states based on the fitness of mutants inferred from their population frequency in the sequenced library: growth essential (cannot tolerate insertion), growth defective (insertion impairs growth), growth neutral (insertion has a neutral impact on growth), or growth advantaged (insertion enhances growth) [127]. A total of 63 million barcoded sequencing reads were obtained across the three libraries, and HMM phenotypes were successfully assigned to 7316 genes in the genome of Rlv3841 (99.7% of the genome).

Mariner library construction was successful in producing a highly saturated mutant library for use in analysis, the input pool was sequenced to have an insertion density of 80% (Table 4-2). Insertion density is the percentage of insertion sites ('TA' motifs) that contain at least one insertion. The higher the insertion density in the input pool, the better

the coverage of insertion sites across the genome and the higher the quality that mutant library is. In transposon mutagenesis screens, insertion densities below 25% are generally considered insufficient for robust statistical analysis and the confident prediction of gene essentiality [93, 97, 123, 126]. It was important to design an INSeq experiment in which sufficiently large input libraries could be inoculated onto the pea seedlings and still allowed to grow for a period of selection (to give resolution between mutant fitness defects), and in which sufficiently large mutant libraries could also be retrieved from the roots and rhizosphere post library selection. The insertion densities of the rhizosphere (65%) and root attached (58%) libraries provide initial validation that the INSeq analysis had appropriate experiment parameters.

Library	Insertion density
Input	80%
Rhizosphere	65%
Root attached	58%

Table 4-2. Insertion densities in the input, rhizosphere and root attached libraries given as the percentage of insertion sites that contain at least one insertion.

4.3.4 Construction of an INSeq database in Rlv3841

An INSeq database was constructed that defines mutational phenotypes for 7316 genes in the genome of Rlv3841 across the nine different *in vitro* and *in planta* analyses carried out over the course of this project: growth on succinate at 21% O_2, growth on glucose at 21% O_2, growth on succinate at 1% O_2, growth on glucose at 1% O_2, an input library, growth in the pea rhizosphere, attachment to pea roots, nodulation (i.e. bacteroids) and recovery from nodules (i.e. nodule bacteria) (collaboration). The INSeq datasets

obtained for the growth of Rlv3841 on TY and on Vincent's Minimal Media (VMM) were also added to this database [98, 213]. The Rlv3841 INSeq database details gene locus, gene name, gene description, number of 'TA' motifs and HMM phenotypes across a total of eleven INSeq libraries and can be used to filter by gene phenotype across conditions (Fig. 4-3).

Gene Id	Desig	Description	No of TA sit	21% Oxygen + 10mM Glu	21% Oxygen + 20mM Succ	1% Oxygen + 10mM Glu	1% Oxygen + 20mM Succ	Input	Rhizosph ere (pea	Root Attached (pea)
RL0748	aglA1	putative alpha-glucosidase	33	NE	NE	NE	NE	NE	NE	NE
RL0749	aglK	putative ATP-binding component of ABC transporter CUT1 alpha-glucoside-transporter	21	NE	NE	NE	NE	NE	NE	NE
RL0750		conserved hypothetical protein	8	ES	GD	ES	GD	NE	NE	NE
RL0751	edd	putative phosphogluconate dehydratase	11	ES	GD	ES	GD	NE	GD	ES
RL0752	pgl	putative 6-phosphogluconolactonase	5	ES	GD	ES	GD	GD	GD	ES
RL0753	zwf1	putative glucose-6-phosphate 1-dehydrogenase	22	ES	GD	ES	GD	GD	GD	ES
RL0754	ordL	putative oxidoreductase	23	NE	GD	NE	GD	NE	NE	NE
RL0755		putative glutamine synthetase	25	NE	NE	NE	NE	NE	NE	NE
RL0756		conserved hypothetical protein	20	NE	NE	NE	NE	NE	NE	NE
RL0757	mcpZ	putative methyl-accepting chemotaxis protein McpZ	19	NE	NE	NE	NE	NE	NE	NE

Figure 4-3. Screenshot of the Rlv3841 INSeq database. HMM mutational phenotypes for genes are given as: ES (growth-essential); GD (growth-defective); NE (growth-neutral); GA (growth-advantaged). Only seven library conditions are shown to allow Fig. to fit within page dimensions.

4.3.5 Core fitness requirements in Rlv3841

The INSeq database was analysed to identify the core fitness requirements in Rlv3841. There were 395 genes that gave growth defective or growth essential phenotypes universally across all eleven INSeq libraries contained in the database (Table 4-3)

(Appendix A; Table 7). These 395 genes are defined as the core fitness requirements in Rlv3841, as regardless of the condition, their mutation always leads to impairment in growth (Appendix A; Table 7).

Library name	**Construction media**	**Selection conditions**	**Number of ES/GD genes**	**Library source**
Input	Standard (TY)	None	541	This work
TY	Standard (TY)	TY	722	Perry and Yost [98]
VMM	VMM	VMM	845	Perry and Yost [213]
UMS/glucose/ NH_4 at 21% O_2	Standard (TY)	UMS/glucose/ NH_4 at 21% O_2	1168	This work [130]
UMS/glucose/ NH_4 at 1% O_2	Standard (TY)	UMS/glucose/ NH_4 at 1% O_2	1100	This work [130]
UMS/succinate/ NH_4 at 21% O_2	Standard (TY)	UMS/succinate/ NH_4 at 21% O_2	1166	This work [130]
UMS/succinate/ NH_4 at 1% O_2	Standard (TY)	UMS/succinate/ NH_4 at 1% O_2	1192	This work [130]
Rhizosphere	Standard (TY)	Rhizosphere	904	This work
Root attached	Standard (TY)	Root attachment	911	This work
Bacteroid	Standard (TY)	Nodule bacteroids	1043	Li/Ramachandran /Wheatley
Nodule recovery	Standard (TY)	Nodule recovery	1131	Li/Ramachandran /Wheatley
Shared ES/GD fitness requirements:			**395**	

Table 4-3. Description of the eleven INSeq libraries used to define universal growth-essential (ES) and growth-defective (GD) genes as the core fitness requirements in Rlv3841. A total of 395 genes were growth-essential or growth-defective upon mutation,

and thus required for optimal fitness, across all eleven INSeq libraries. Construction media refers to the media the library was initially screened on following transposon conjugation, selection conditions refer to the growth environment the library was subjected to prior to DNA extraction and library preparation for sequencing.

The VMM library is a useful inclusion in this comparison due to its construction on VMM: a minimal medium supplemented with mannitol with a different composition to TY. Construction and growth of a library on VMM will help distinguish against genes required specifically for growth on TY. This equates to 66 genes in the input library that did not show this same growth impairment (i.e. growth-defective or growth-essential upon mutation) on VMM. While this may initially suggest that VMM might be a better media choice for library preparation, this is not the case. Selection on VMM resulted in a total of 845 genes showing growth-defective or growth-essential phenotypes (Table 4-3) [98, 213], whereas the comparative experiment on TY identified only 722 genes with growth-defective or growth-essential phenotypes (Table 4-3). This is logical as VMM is a minimal medium and TY is a complex rich medium, although TY evidently does causes growth limitations of its own, most of which are likely caused by a limitation of key metals [213]. To minimalise initial growth defects in the input library, it may be best practice to construct insertion libraries using a media that is part TY and part minimal medium supplemented with a range of extra carbon and nitrogen sources [213].

4.3.5.1 Functional dissection of the core fitness requirements in Rlv3841

The majority of the 395 INSeq-identified genes that form the core fitness requirements are chromosomal: 369 out of 395 (Appendix A; Table 7). Notable plasmid genes include the cobalamin '*cob*' metabolism genes on pRL11 and the replication '*rep*' system genes

present on all six plasmids. Genes within the core fitness requirements encoded functions spanning cell processes (50 genes), macromolecule metabolism (82 genes), metabolism of small molecules (133 genes), the cell envelope (78 genes), regulation (15 genes) and 37 genes which encode either unknown, unclassified or conserved hypothetical protein products (Fig. 4-4).

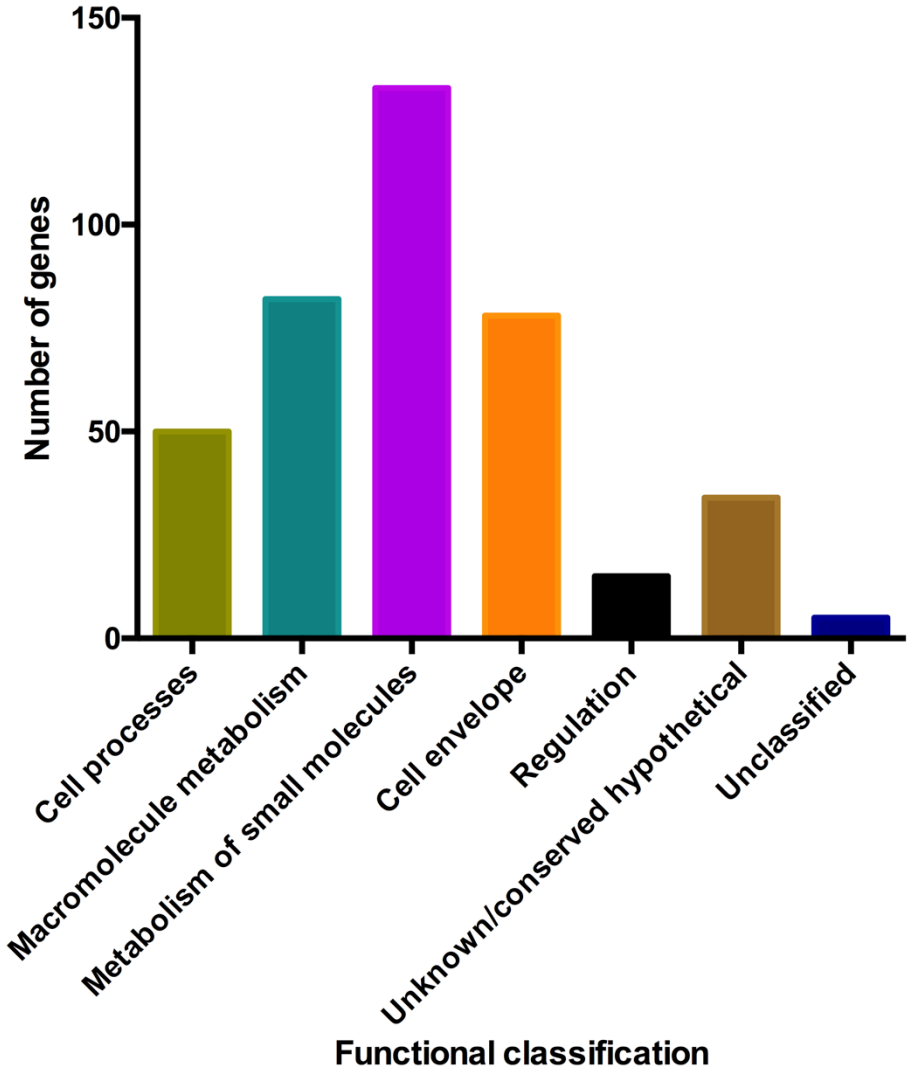


Figure 4-4. Functional distribution of the INSeq-identified core fitness requirements. Functional classification is based on Riley codes [52].

A number of protein secretion systems were identified as core fitness requirements within the functional classification of cell processes, including: the general secretion (Sec) pathway encoded by RL0680 (*secDF2*), RL1759 (*secE*), RL1794 (*secY*), RL2055 (*secDF1*) and RL4298 (*secA*); the twin-arginine translocation (Tat) pathway encoded by RL2046-RL2048 (*tatA-tatC*); and the more poorly characterised Tol system, encoded by RL3969 (*tolB*), RL3971 (*tolA*) and RL3973 (*tolQ*). The Tol system is suggested to be involved in stability of the outer membrane in *E. coli* [214, 215], and to help facilitate the translocation of group A colicins [216], which are bacterial protein toxins. RL3972 (*exbD*), encoding a putative component of the Ton system, was also identified as a core fitness requirement. The Ton system is proposed to function as an energised high-affinity iron (III), heme and vitamin B_{12} transporter and is highly conserved across gram-negative bacteria [217, 218]. Other core fitness requirements within this category of cell processes included genes encoding cell division proteins, chromosomal partitioning proteins (*parA* and *parB*), chaperonins, and ABC transporter systems, including the PhoT ABC transporter system encoded on RL0542-RL0544.

Core fitness requirements within the functional classification of macromolecule metabolism included a number of tRNA synthetases, the plasmid replication systems, and RNA and DNA polymerase components. Within the functional classification of cell envelope components, there were a significant number of genes involved in the construction of cell surface EPS and lipopolysaccharide (LPS), along with a number of less well-characterised putative transmembrane protein encoding genes. Linked to this cell envelope group, RL2037 (*exoR*), encoding the ExoR negative regulator of EPS production, was also identified as a fitness requirement. Correct EPS processing is

important for attachment in *Rhizobium* [219] and, presumably, regulation of EPS production is important for general cell fitness as well.

The functional classifications of the largest group of core fitness requirements, the 133 genes involved in small molecule metabolism, could be further split down into 10 more specific functional classifications (Fig. 4-5). Of note, there were 27 genes involved in amino acid biosynthesis, these amino acids being histidine, lysine, methionine, proline, serine, threonine, tryptophan, arginine and tyrosine. There were 28 genes involved in the biosynthesis of cofactors, including 9 of the *cob* genes involved in cobalamin metabolism. Central intermediary metabolism included genes with predicted functions such as kinases, polyphosphatases and alcohol dehydrogenases. Genes for aerobic respiration were important, with 13 genes required for forming the NADH-quinone oxidoreductase complex and, as would be expected, there was also a requirement for genes involved in ATP synthesis and ATP-proton motive force.

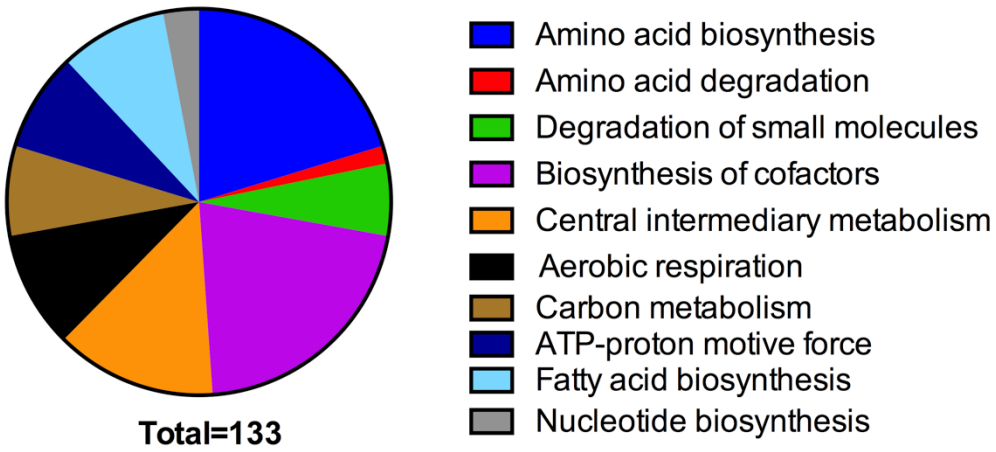


Figure 4-5. Functional distribution breakdown of the 133 genes categorised in the metabolism of small molecules and identified in the core fitness requirements. Functional classification is based on Riley codes [52].

4.3.6 Classification of fitness determinants

The INSeq classifications across the input library, growth in the rhizosphere, attachment to roots, nodulation and nodule recovery were combined to classify the fitness determinant groups for discussion in this thesis. Four broad groups of fitness determinants were defined: 1) rhizosphere; 2) root attachment; 3) nodulation (bacteroids) and 4) nodule recovery (nodule bacteria) (Fig. 4-6). A total of 463 genes were identified to be required for successful nodulation: 176 genes required from the rhizosphere through to nodulation, 14 genes required from root attachment through to nodulation, 39 genes required specifically for bacteroid development, 154 genes required in both bacteroids and nodule bacteria, and 80 genes required specifically in nodule bacteria (Fig. 4-6). There was identification of 55 genes that are required specifically for growth in the rhizosphere (Fig. 4-6), mutation of which does not impair

root attachment or nodulation, and 101 genes that are required specifically for root attachment (Fig. 4-6), mutation of which does not impair nodulation.

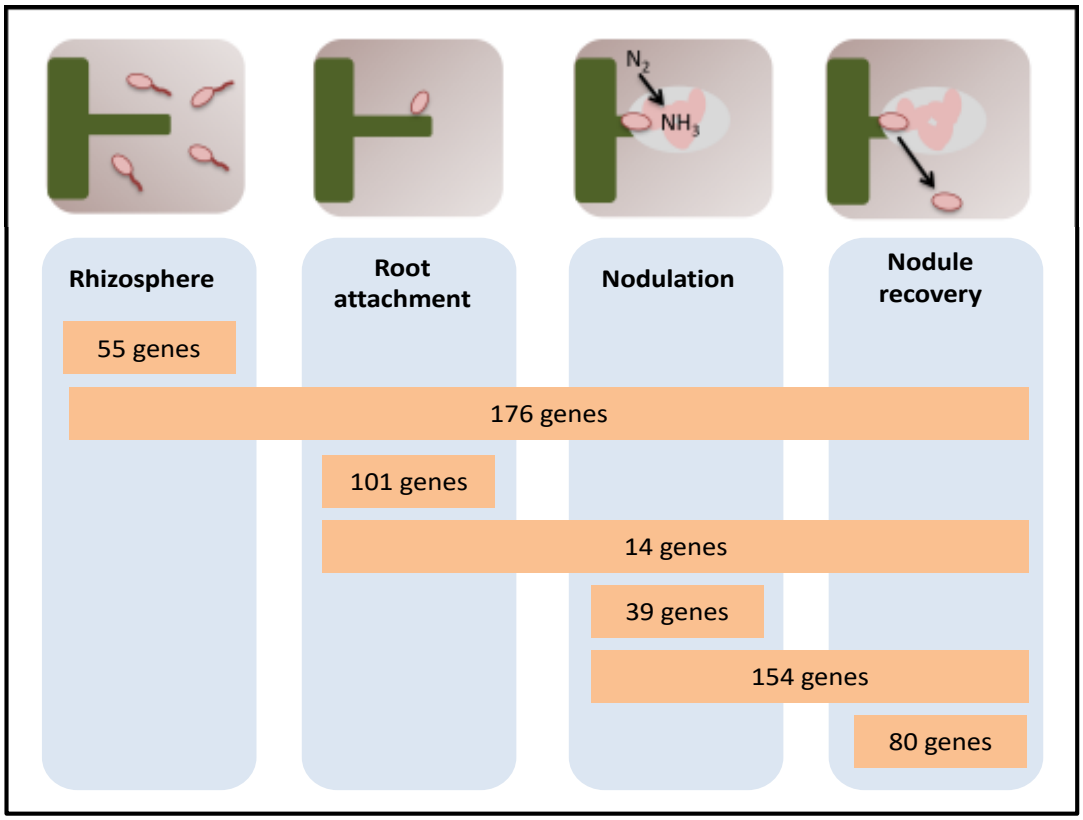


Figure 4-6. Genetic map of the gene requirements for engaging in symbiosis with pea identifying the specific stages of symbiosis these genes are required in. Orange boxes show the number of genes identified with a growth-essential or growth-defective HMM phenotype across each symbiosis stage. All identified genes had a growth-neutral phenotype upon mutation in the input library.

4.3.7 Genes required for growth in the rhizosphere

A total of 231 genes were identified to be required for growth in the rhizosphere, these can be split into two groups; 'rhizosphere-specific' genes and 'rhizosphere-progressive' genes (Table 4-4) (Appendix A; Table 8). The rhizosphere-specific genes are the 55

genes whose mutation was growth-essential or growth-defective specific to the rhizosphere, but did not impair subsequent stages of root attachment or nodulation (Table 4-4). The rhizosphere-progressive genes are the 176 genes whose mutation was initially growth-essential or growth-defective in the rhizosphere, and this phenotype then progressed over the subsequent stages of attachment and nodulation (Table 4-4). For the rhizosphere-progressive genes, it is possible that the effect of mutation is direct to the rhizosphere, but it is so severely impairing that the mutants remain at a significant competitive disadvantage. Or it is possible that the gene plays a more global role in symbiosis and is involved in multiple stages of colonisation.

Classification	**Number of genes**	**INSeq library**				
		Input	Rhizosphere	Root attached	Nodulation	Nodule recovery
Rhizosphere-specific	55	NE	ES/GD	NE	NE	NE
Rhizosphere-progressive	176	NE	ES/GD	ES/GD	ES/GD	ES/GD

Table 4-4. The classification of rhizosphere-specific and rhizosphere-progressive gene requirements based on HMM phenotypes across the stages of symbiosis. HMM phenotypes are given as NE; growth-neutral and ES/GD; growth-essential/growth-defective.

4.3.7.1 Functional dissection of genes required for growth in the rhizosphere

The rhizosphere-specific and rhizosphere-progressive genes are combined for a functional dissection of the global gene requirements for growth in the rhizosphere (Appendix A; Table 8). A host of defensive machinery was required for growth in the rhizosphere, including several genes predicted to be involved in responses to oxidative

stress: pRL100156 (*fdxB1*), a putative ferredoxin; RL1370 (*msrB1*), a putative methionine sulfoxide reductase; RL0482A, a putative oxidoreductase; RL0824, another putative oxidoreductase; RL1041, a putative oxidoreductase/monooxygenase; RL0802, a putative deoxygenase; and RL0025, a putative thioredoxin. Plant roots can release reactive oxygen species (ROS) in the form of hydrogen peroxide, hydroperoxyl radicals, hydroxyl radicals, superoxide, nitric oxide, and phytoalexins from their roots [66, 220]. It is likely these genes are therefore required for combating the oxidative stress generated by plant roots.

Other stress-related or defence-related genes were identified in the form of: RL2964 (*cspA*), encoding a cold shock response protein; RL1716 and RL1093, encoding putative beta-lactamase family proteins; RL1506 (*relA*), encoding a putative stringent response protein; and RL1716 (*phr*), encoding a putative deoxyribodipyrimidine photo-lyase. Deoxyribodipyrimidine photo-lyases are enzymes that are involved in catalysing the repair of UV radiation-induced DNA damage. Identification of RL1506 (*relA*) as a requirement for successful symbiosis across all four stages adds to the increasing volume of evidence supporting the role of RelA as a key global regulator involved in a broad network of regulatory systems [130, 204, 205], as initially hypothesised in chapter 3. The requirement for stress-related and defence-related genes in the rhizosphere is consistent with previous transcriptomic analysis of Rlv3841 [57]. Stress-related and defence-related genes were a highly induced gene group in the pea rhizosphere, and upregulation of the thioredoxin, methionine sulfoxide reductase and cold shock response protein encoding genes specifically identified in this study were also observed [57].

The nitrogen regulatory NtrB/NtrC two component system (RL2256-RL2257) controls the expression of the nitrogen-regulated genes in response to nitrogen limitation and was identified as a requirement in this group of rhizosphere-progressive genes. It seems logical that the regulation of genes responding to nitrogen limitation was important initially in the rhizosphere, because pea seedlings were germinated in nitrogen-free rooting solution to help promote rhizobial association with the plants to satisfy this requirement. RL1621 (*ribG*) and RL1622 (*ribC*) encoding putative riboflavin biosynthesis proteins were also identified in the group of rhizosphere-progressive genes; growth-defective upon mutation in the rhizosphere and root attached libraries, and growth-essential in bacteroids and for nodule recovery (Appendix A; Table 8). Riboflavins are cofactor precursors for flavoproteins, which are essential for numerous cellular processes including biosynthesis, DNA repair and energy production. The availability of riboflavin is known to influence rhizobial survival in the rhizosphere and their capacity to colonise roots. A *R. leguminosarum* mutant of RibN, a putative riboflavin transporter protein, has previously been shown to have a lower nodule occupancy in pea plant assays [221], and this supports these results indicating riboflavin biosynthesis is important for growth in the rhizosphere and the subsequent stages of colonisation and nodulation too.

A number of transporters were required for growth in the rhizosphere; those with predicted cargos included: a ferric cation transporter (RL4584, *fbpA*), a CUT1 carbohydrate ABC transporter (pRL9), a cytochrome bd-related ABC transporter (pRL9), a putative O-antigen transporter (RL0821), a NiT nitrate/nitrite/cyanate ABC transporter (RL4400, RL4401) and a cytochrome c biogenesis export ABC transporter

(RL4537/*ccmA*, RL4538/*ccmB*). The predicted cargos of these transporters are presumably related to the need to scavenge from the environment.

A large proportion of the genes required for growth in the rhizosphere encoded functions classified in the metabolism of small molecules, 81 genes out of the 231 total (Fig. 4-7). This likely represents the significant adaptation required in going from a rich TY media to an *in planta* environment. These genes were predicted to encode proteins including tryptophan synthase (RL0021/*trpB*, RL0022/*trpA*), N-acetylglucosaminyltransferase NodC (pRL100187), cobalamin metabolism enzymes (pRL110632/*cobF*, RL2824/*cobA*, RL2836/*cobQ*), thiamine-phosphate pyrophosphorylase (RL4040/*thiE2*), purine biosynthesis enzymes (RL4044/*purE*, RL4045/*purK,* RL4722/*purH*) and, rather unusually, the nitrogenase molybdenum-iron protein NifD (pRL100161) (Appendix A; Table 8).

A significant proportion of the rhizosphere-required cell envelope genes are involved in cell surface polysaccharide production, EPS in particular. Three of the polysaccharide synthesis (*pss*) genes were included in this group (RL3648/*pssJ*, RL3651/*pssG*, RL3663/*pssP*) along with genes encoding a UDP-glucose-epimerase (RL0813), a UDP-glucose dehydrogenase (RL3667), a GDP-mannose dehydratase (RL0825) and a sugar synthetase (RL0826). Mutation to these genes all resulted in growth impaired phenotypes that were carried through subsequent stages of root attachment and nodulation too. EPS is a major constituent of the cell surface and has previously been reported to play a role in secondary phase root attachment in *R. leguminosarum* [23, 191]. These results indicate correct EPS biosynthesis and processing are also required for optimal fitness in the rhizosphere. As EPS biosynthesis and processing genes were

also identified as important core fitness requirements (Appendix A; Table 7), this is unsurprising.

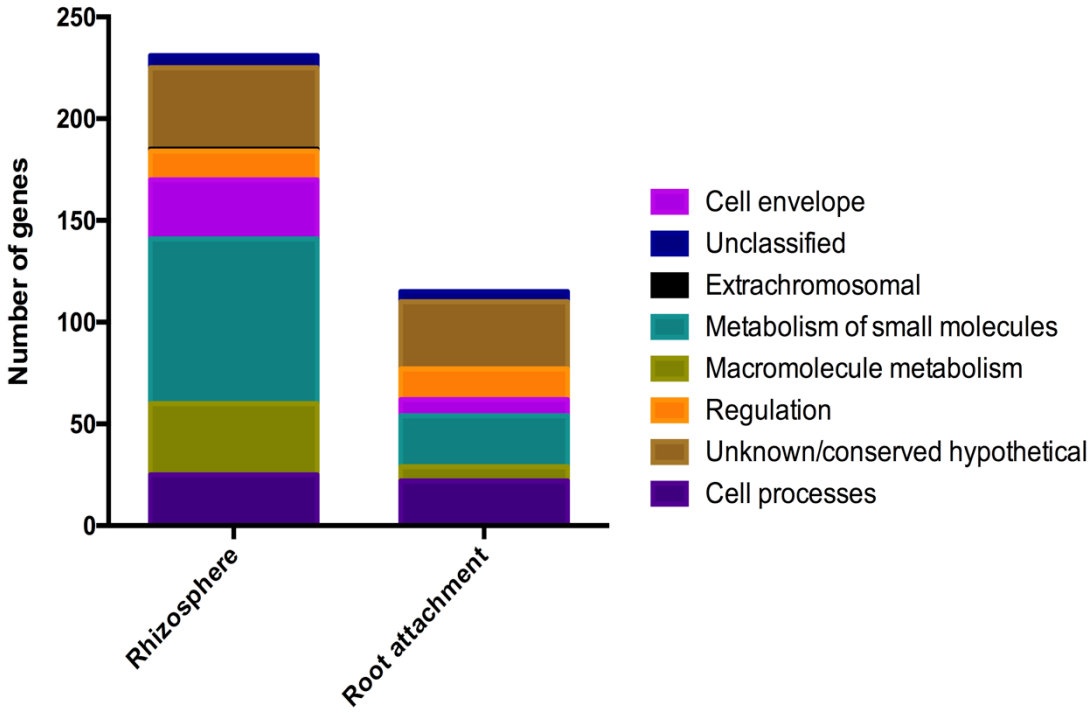


Figure 4-7. Functional distribution of the INSeq-identified rhizosphere and root attachment genes. Functional classification is based on Riley codes [52].

4.3.8 Genes required for root attachment

A total of 115 genes were identified to be required for root attachment, these can be split into two groups; 'attachment-specific' genes and 'attachment-progressive' genes (Table 4-5) (Appendix A; Table 9). The attachment-specific genes are the 101 genes whose mutation was growth-essential or growth-defective specific to root attachment, but did not impair growth in the rhizosphere nor nodulation (Table 4-5) (Appendix A; Table 9). The attachment-progressive genes are the 14 genes whose mutation was initially growth-essential or growth-defective in the rhizosphere, and this phenotype then progressed over the subsequent stages of nodulation too (Table 4-5) (Appendix A; Table 9). It is

likely that the attachment-progressive genes are on the root hair infection pathway needed for nodulation, and the larger group of attachment-specific genes seems to indicate a root attachment phenotype that is distinct from nodulation, and is perhaps more associated with attachment to the root epidermal surface. A significant proportion of the genes required for root attachment were found on pRL10, the 'symbiosis' plasmid in Rlv3841. Genes on pRL10 account for 17% of the total gene requirements for root attachment (Appendix A; Table 9), yet pRL10 only accounts for ~ 6.5% of genes within the genome.

Classification	**Number of genes**	**INSeq library**				
		Input	Rhizosphere	Root attached	Nodulation	Nodule recovery
Attachment-specific	101	NE	NE	ES/GD	NE	NE
Attachment-progressive	14	NE	NE	ES/GD	ES/GD	ES/GD

Table 4-5. The classification of attachment-specific and attachment-progressive gene requirements based on HMM phenotypes across the stages of symbiosis. HMM phenotypes are given as NE (growth-neutral) and ES/GD (growth-essential/growth-defective).

4.3.8.1 Functional dissection of genes required for attaching to roots

The attachment-specific and attachment-progressive genes are combined for a functional dissection of the global gene requirements for attaching to roots. While these two gene groups are suggested to potentially represent different attachment phenotypes,

i.e. attachment to the root hairs with the potential to lead to nodulation vs. attachment to the epidermal surface, the retrieval phase of mutants in this study did not distinguish between different root zones but collected mutants from the whole root surface, so it is not possible to confidently assign different root attachment zone phenotypes post-analysis.

A number of the identified attachment genes on pRL10 have already been linked to roles in root attachment. For example, the *rhi* genes, named on account of them being strongly rhizosphere-induced (rhi) [57], have previously been suggested to play a role in colonisation and nodulation, and mutation to pRL100164 (*rhiI*), pRL100170 (*rhiB*), pRL100171 (*rhiC*), and pRL100172 (*rhiR*), was found to be classified as growth-defective for root attachment in INSeq. Previously, single gene mutation experiments had been unable to reveal a phenotype for these genes (Downie personal communication), and their identification in INSeq highlights the power of this technique in resolving fitness differences between strains as a result of the strong competitive pressure induced by the number of "wild-type"-like strains that will also be present.

Other INSeq-identified attachment genes with a previously suggested role in attachment included: pRL100309 (*cadB*), encoding the predicted attachment cadherin protein CadB [212]; pRL110564, encoding a putative tight adherence protein [55, 57]; and RL2702 encoding a putative rhizobium adhering (rap) protein [23]. Cadherins have been well-characterised to mediate cell-cell adhesion in metazoan organisms [222]. Cadherins have only more recently been identified as present in bacterial genomes [223], and cadherin domain-containing proteins are still considered to be relatively rare [224]. This identification of a role for CadB in attachment suggests their adhesive function still

stands [224, 225], and that CadB plays a role in root attachment in Rlv3841. This could be either through mediating interactions with the root surface directly, or through mediating rhizobial cell-cell adhesion and promoting the secondary attachment of subsequent rhizobia to the site of initial adhesion.

Previous research has identified cellulose synthase CelA (RL1646) and glucomannan (RL1661; *gmsA*) with attachment-related roles in Rlv3841 [85]. A cellulose synthase CelA mutant (RL1646) has been reported to attach and form normal biofilms *in vitro*, but be unable to form a biofilm on root hairs, although attachment can still occur [85]. Cellulose-dependant biofilm formation on root hairs was not deemed to be required for nodulation; the *celA* RL1646 mutant competed equivalently with wild-type Rlv3841 for nodule infection [85]. The INSeq results are found to be consistent with this; mutation to RL1646 (*celA*) was growth-neutral across growth in the rhizosphere, attaching to roots, nodulation and nodule recovery. RL1647 (*celB*) and RL1648 (*celC*) were similarly growth-neutral upon mutation across all conditions.

A glucomannan (RL1661; *gmsA*) mutant has previously been shown to be defective in attachment and biofilm formation on root hairs under acidic conditions [85]. Glucomannan mediates attachment by binding to root hair expressed lectin [23, 69], however, under alkaline or neutral pH conditions, root lectins are solubilised and this method of glucomannan-mediated attachment will no longer be able to function [23, 69]. The pH of the environment in this study was neutral (pH 7.0), and consistent with this, the glucomannan-encoding gene (RL1661; *gmsA*) was characterised as growth-neutral upon mutation for root attachment. The mechanism of attachment under alkaline and neutral pH is suggested to involve an extracellular rhizobial protein ‘rhicadhesin’,

although neither the gene encoding rhicadhesin nor the protein itself has been identified [23, 68, 72, 73]. No obvious candidates for a rhicadhesin-mediated method of attachment under these neutral pH conditions were identified in this study.

In this INSeq analysis the genes encoding SoxD2 and SoxB2, sarcosine oxidase subunits, were found to be required specifically in the root attached library (Appendix A; Table 9). Sarcosine is a common soil metabolite and it appears metabolism of sarcosine is important for optimal fitness in attachment. RL2931 (*mcpN*) that encodes the putative methyl-accepting chemotaxis protein McpN was also an attachment-specific requirement. Motility enables microbes to migrate to the roots, and McpN is likely involved in the chemotaxis of Rlv3841 under these conditions. A number of different attachment-specific transporter systems were identified, including a NiT nitrate/nitrite/cyanate ABC transporter encoded on pRL12, a MZT manganese/zinc/iron transporter system on the chromosome, a QAT proline/glycine ABC transporter and a putative transmembrane ion efflux system protein encoded by RL0682. For some of the identified transporter systems the transported substrate remains unknown, such as the predicted ABC transporter encoded by pRL100127-pRL100129.

The number of genes involved in small molecule metabolism required in the root attached library was significantly lower than in the rhizosphere library (Fig. 4-6). This may reflect the fewer metabolic adaptations needed in progressing from growth in the rhizosphere to attaching to roots compared to the previous adaptation of growth on rich TY media to transfer into the rhizosphere. The percentage of genes aligning to each functional classification group was compared across growth in the rhizosphere, attaching to roots and the core fitness requirements (Table 4-6). As would be expected,

the core fitness requirements are significantly better characterised; less than 10% of the core fitness requirement genes encode uncharacterised proteins, compared to 19.9% and 33.0% in the rhizosphere and root attachment respectively (Table 4-6). It is quite striking that around one third of the genes required specifically for root attachment remain uncharacterised. Root attachment contained the highest proportion of genes involved in regulation; 13% of the total (Table 4-6). Regulatory elements form a less important component of the core genome (3.8%), but are clearly important in coordinating root attachment, which involves a molecular dialogue with the plant.

Around 20% of the core fitness requirements had a functional association to the cell envelope, and the proportion of cell envelope associated genes decreased in the rhizosphere and root attachment gene categories (Table 4-6). EPS is a major constituent of the cell surface and has previously been associated with a role in secondary attachment in *Rhizobium*. EPS is suggested to aid the formation of rhizobial aggregates as a 'cap' on the root hair [23]. However, no genes predicted to encode EPS biosynthesis or processing proteins were found within the group of root attachment specific fitness requirements (Appendix A; Table 9). A significant number of EPS-encoding genes were instead found to be required as core fitness components or for multiple stages of rhizosphere colonisation and nodulation. Although EPS plays a role in attachment [23], the importance of EPS is not specific to attachment and it is also a requirement for general cell fitness.

Another constituent of the cell surface is LPS, and mutation to RL2664 (*lpxH*) which encodes a putative UDP-2,3-diacylglucosamino hydrolase was found to be growth-defective specifically in the root attached library. This enzyme functions in the

biosynthetic pathway of lipid A, a constituent of LPS [226]. LPS is well-characterised with a role in attachment across numerous species of agriculturally important rhizobacteria [68], including in *R. leguminosarum* [23, 68, 227, 228].

Functional classification	**Percentage of genes in functional classification (%)**		
	Core genome	Rhizosphere	Root attachment
Cell processes	12.6	10.8	19.1
Macromolecule metabolism	20.7	15.2	6.1
Metabolism of small molecules	33.6	35.1	21.7
Cell envelope	19.7	12.6	7.0
Regulation	3.8	6.1	13.0
Extrachromosomal	0.0	0.4	0.0
Unknown/conserved hypothetical/unclassified	9.6	19.9	33.0

Table 4-6. Percentage of total gene requirements belonging to each Riley code based functional classification in the core genome (396 genes total), rhizosphere requirements (231 genes total) and root attachment requirements (115 genes total).

4.3.9 Analysis of central carbon metabolism

The mutational phenotypes of genes within the central carbon metabolism pathways of Rlv3841 were analysed across the input, rhizosphere, root attached and bacteroid libraries (Fig. 4-8). The mutations of RL0182, encoding glucokinase, and RL0183, encoding methylglyoxal synthase, were growth-defective specifically in the rhizosphere.

This suggests that the efficient metabolism of glucose (or metabolically related sugars) and phosphorylated intermediates of glucose are important for optimum fitness under these conditions, but are not required in the later stages of colonisation. As there were no carbon sources provided in the root solution, any carbon must be coming from the plant. This likely relates to the type of carboniferous root exudates providing an energy source for Rlv3841 in the rhizosphere, and suggests that glucose-related sugars may be significant among these.

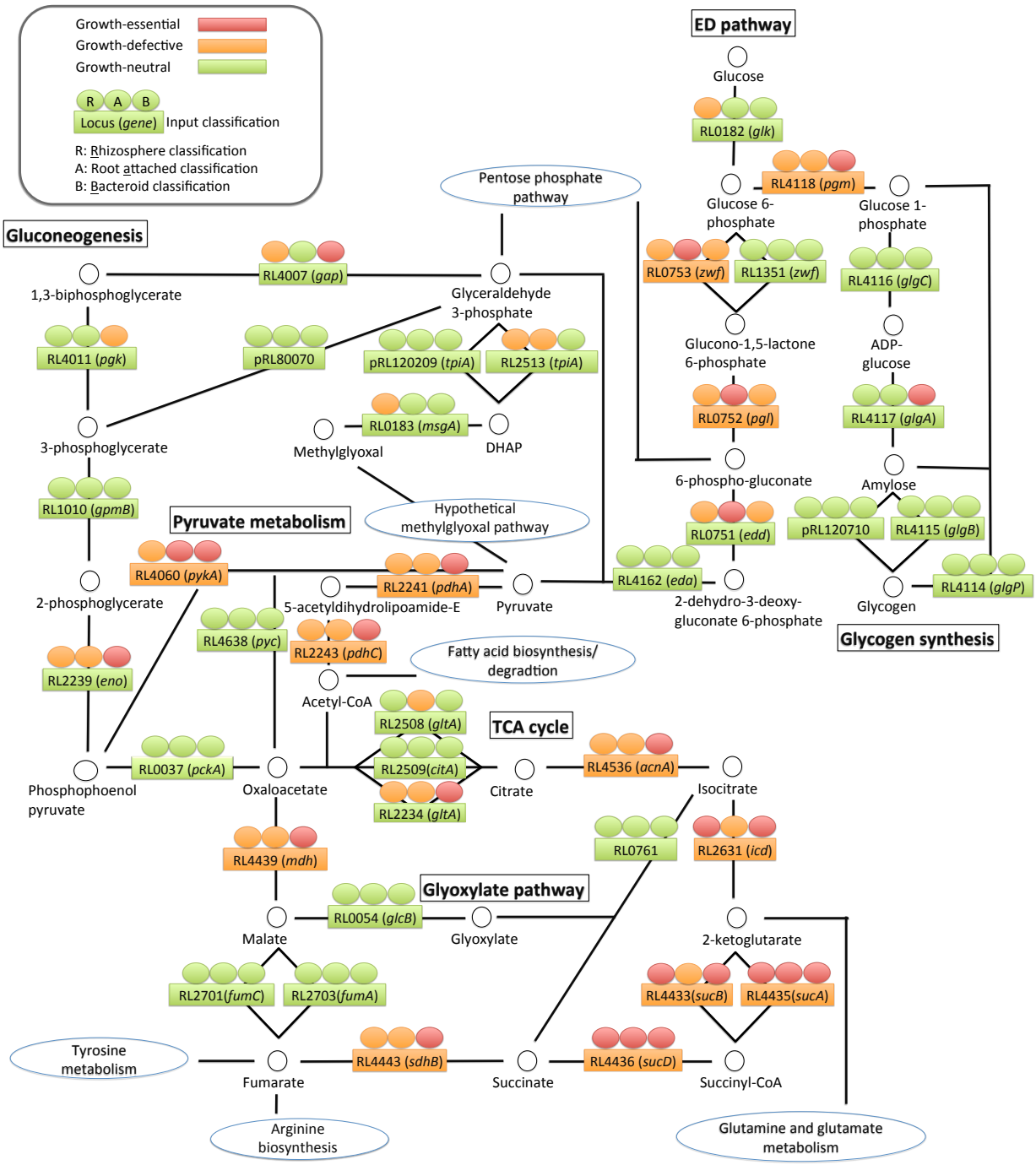


Figure 4-8. Central carbon metabolic pathways of Rlv3841. Candidate genes for enzymes performing the catalytic steps are shown in coloured boxes according to their INSeq mutant classification in the input library, with three circles above indicating the rhizosphere classification, root attached classification and bacteroid classification, from left to right, respective. Classifications are given as: growth-essential (red), growth-defective (orange) and growth-neutral (green).

The core components of the ED pathway (encoded on RL0751-RL0753) were required synonymously across the rhizosphere, root attachment and bacteroid populations (Fig. 4-8). As suggested in the previous INSeq screen, RL0753 is the "real" *zwf [130]*, and RL1315 is unable to compensate for its function. Intriguingly, RL4162 (*eda*) showed a growth-neutral classification upon mutation across all four libraries, yet is located in a pathway catalysing a key link between genes whose mutation causes significant growth-impairment. This suggests that an alternative aldolase may be present in the genome of Rlv3841 that can compensate for the function of *eda*. The genome of Rlv3841 contains 42 putative aldolase enzymes annotated, however, alignment of the amino acid sequence of RL4162 (*eda*) yielded no significant similarities and was unsuccessful in identifying a more likely potential candidate [143].

Glycogen synthesis was comparatively more important in bacteroids, with the genes encoding phosphoglucomutase (RL4118) and glycogen synthase (RL4117) giving uniquely growth-essential phenotypes upon mutation in the bacteroid library (Fig. 5-6). The presence of glycogen in bacteroids was first observed by Dixon (1967), and while glycogen is suggested to function as a carbon storage compound in pea bacteroids [229], the exact role glycogen accumulation or degradation plays regarding nitrogen-fixation remains to elucidated. Mutations to the glycogen synthase (RL4117) gene in *R. leguminosarum* have been made, and the strain is still able to nodulate plants [229], but these nodules show abnormally large accumulations of starch [229]. The INSeq results indicate glycogen synthetase mutants are severely decreased in the bacteroid library, showing that glycogen synthetase mutants are at a significant competitive disadvantage in comparison to other strains. This suggests that the ability to synthesise glycogen confers a fitness advantage to bacteroids that cannot be compensated for by the synthesis

of alternative carbon storage compounds. It may be that glycogen synthesis affects competitiveness for nodules, and as a result the glycogen synthetase mutant forms fewer nodules on the roots and this is the cause behind its decreased frequency in the bacteroid population. Or, perhaps more likely, glycogen synthesis is important for optimal fitness in the nodules (i.e. efficiency of N_2-fixation), and as a result the glycogen synthetase mutant is unable to replicate as well in the nodule and forms smaller nodules containing fewer cells, and this is the cause behind its decreased frequency in the bacteroid population.

Gluconeogenesis was similarly more important for fitness in bacteroids than early stage symbiosis, with three key gluconeogenesis enzyme-encoding genes (RL2239, RL4007, RL4011) showing decreased fitness in the bacteroid library compared to in the rhizosphere and root attachment (Fig. 4-8). From these results it seems plausible that glycogen in bacteroids might be made from sugars synthesised by gluconeogenesis. From this, a metabolic hypothesis was generated: bacteroids synthesise glycogen from plant-provided TCA-cycle intermediates that undergo gluconeogenesis to generate glycogenic substrates. Synthesising glycogen via gluconeogenesis, a pathway that generates a net oxidisation of NADH to NAD^+ [182], would confer an advantage with redox potential and energy balance in the cell at a time where bacteroids are carrying out the highly reductive process of N_2-fixation.

In the TCA-cycle, Rlv3841 possesses three putative citrate synthase genes: RL2508 (*gltA*), RL2509 (*citA*) and RL2234 (*gltA*). Mutation to RL2234 (*gltA*) was growth-defective across all three output libraries. This agrees with the previous INSeq screen, where RL2234 (*gltA*) was the sole citrate synthase with a growth-defective phenotype

across all conditions. Mutation to RL2508 (*gltA*) was growth-defective specifically in the root-attached library. These results indicate that the other citrate synthase genes are unable to compensate for the activity of RL2234 (*gltA*); RL2234 is potentially responsible for the majority of citrate synthase activity in Rlv3841. This knowledge could be useful in synthetic biology, where the minimal genetic requirements for a certain pathway need to be identified to ease their transfer and efficiency in other systems.

One apparent outlier between this screen and the previous analysis (Chapter 3) is the growth impaired phenotype on mutation to isocitrate dehydrogenase (RL2631), which was growth-neutral upon mutation across all four of the UMS and succinate or glucose conditions [130]. This difference in phenotype may be because the symbiosis INSeq screens pose a more competitive and difficult-to-colonise environment, which may help resolve more fitness defects in the population. Indeed, rhizobial strains mutated in isocitrate dehydrogenase have been shown to colonise and nodulate the rooting system when present as the sole inoculated strain [177]. In this INSeq screen, mutants are instead competing with thousands of "wild-type"-like mutated strains (i.e. growth neutral mutants), in an environment that possesses significantly more colonisation difficulties than a large liquid culture provided with plentiful carbon and nitrogen.

4.3.10 Mutation of genes to improve rhizobial fitness

There has been a large focus on genes that decrease fitness upon mutation in the INSeq discussions in this thesis. However, HMM analysis also allows the identification of genes that increase fitness upon mutation, those assigned as growth-advantaged [127]. These are genes that have significantly more insertions along their length, with the

classification boundary set at five times the mean read count [127]. A total of 237 genes were growth-advantaged upon mutation in the input library. Of these 237 genes, only 49 were also growth-advantaged in the rhizosphere, and only a further 32 of these were also subsequently growth-advantaged in root attachment. No genes were growth-advantaged upon mutation across the input, rhizosphere, root attached, nodulation and nodule recovery screens; consistent with natural selection removing any generally deleterious genes from the population.

Growth-advantaged phenotypes may represent genes that have a metabolic cost, such as biosynthesis of a secreted toxin, which are not necessary for growth under the condition given [127]. A number of predicted transporter components were identified with growth-advantaged phenotypes. Genes encoding a substrate-binding component (RL4553) and a permease component (RL4554) of a polar amino acid uptake transporter (PAAT family) were growth-advantaged across five INSeq libraries (input, rhizosphere, root attached, VMM, TY) [98, 213]. Of the 32 growth-advantaged genes in the input, rhizosphere and root attached libraries, three of these encoded a carbohydrate CUT1 ABC transporter (RL1825, RL1826, RL1827) and three a peptide PepT ABC transporter (RL4413, RL4414, RL4415). In some cases, mutation to a transporter can be advantageous because the transporter imports a toxic compound, but as this phenotype was evident in growth media, this seems less likely. Alternatively, the biosynthesis or functioning of these transporters may be metabolically costly, and suggests that they are not needed under the conditions described.

4.3.11 Transcriptional regulators involved in root attachment

Eleven transcriptional regulators were identified as being required specifically for attachment in the INSeq analysis (Table 4-7, Table 4-8). Root attachment had the requirement for the highest percentage of genes with a predicted function in regulation (Table 4-6): 13.0% of the fitness determinants in root attachment compared to 6.1% and 3.8% in the rhizosphere and input library, respectively. This is consistent with root attachment being a highly coordinated process, and that responding to chemotaxis-inducing root exudates and components of the root surface requires signal perception and a regulated response. Root attachment is not only an essential step for symbiosis, but also a key biological process in numerous other agriculturally important plant-microbe interactions [68]. Better understanding the transcriptomic regulation of root attachment will surely improve our understanding of how rhizobia are able to respond to the root and successfully engage in symbiosis, alongside hopefully the understanding of this mechanism across other plant-microbe interactions. To these ends, my aim was to map the transcriptional regulatory pathways required for attachment by identifying the groups of genes directly regulated by these transcriptional regulators (i.e. their regulons) using RNA sequencing [104].

Locus	Description	TA sites
pRL80032	putative LysR family transcriptional regulator	24
pRL90149	putative LuxR/GerE family transcriptional regulator	21
pRL100057	putative AsnC family transcriptional regulator	15
pRL100172	putative transcriptional regulatory protein controlling rhi gene expression RhiR	20
pRL120160	putative DeoR family transcriptional regulator (repressor)	15
pRL120692	putative ROK family transcriptional regulator	13
RL0256	putative XRE family (HipB) transcriptional regulator with cupin2 domain	7
RL0561	putative AraC family transcriptional regulator (activator)	23
RL1040	putative LysR family transcriptional regulator	12
RL2828	putative XRE family (HipB) family transcriptional regulator	7
RL3828	putative FNR/CRP family transcriptional regulator	16

Table 4-7. The locus, description and number of 'TA' insertion sites within the eleven transcriptional regulators required for root attachment selected for downstream analysis.

Locus	HMM Classification								
	In	Rh	Ro	No	Re	G21	S21	G1	S1
pRL80032	NE	NE	GD	NE	NE	NE	NE	NE	NE
pRL90149	NE	NE	GD	NE	NE	ES	GD	ES	ES
pRL100057	NE	NE	GD	NE	NE	GD	GD	GD	GD
pRL100172	NE	NE	GD	NE	NE	NE	GD	GD	GD
pRL120160	NE	NE	ES	NE	NE	NE	NE	NE	NE
pRL120692	NE	NE	GD	NE	NE	NE	NE	NE	NE
RL0256	NE	NE	GD	NE	NE	NE	NE	NE	NE
RL0561	NE	NE	GD	NE	NE	NE	NE	NE	NE
RL1040	NE	NE	GD	NE	NE	NE	NE	NE	NE
RL2828	NE	NE	GD	ES	ES	NE	NE	NE	NE
RL3828	NE	NE	GD	NE	NE	GD	GD	GD	GD

Table 4-8. The HMM classifications of the eleven transcriptional regulators in the nine INSeq experiments carried out as part of this project: the input, rhizosphere, root attached, nodulation and nodule recovery libraries detailed in this chapter, and the four UMS-grown libraries detailed in chapter 3 [130]. In: input, Rh: rhizosphere, Ro: root attachment, No: nodulation, Re: nodule recovery, G21: glucose at 21% O_2, S21: succinate at 21% O_2, G1: glucose at 1% O_2, S1: succinate at 1% O_2. HMM classifications are given as NE: growth-neutral, GD: growth-defective, ES: growth-essential.

4.3.11.1 Genomic context of the transcriptional regulators involved in root attachment

Many transcriptional regulators are located adjacent to the genes they regulate, and genes located upstream and downstream of the 11 transcriptional regulators were screened for this possibility (Appendix A; Table 10). To identify putative regulator candidates, genes located in genomic proximity of the transcriptional regulators were screened for changes in INSeq mutational phenotype and for functional predictions with a clear link to attachment (i.e. encoding an attachment-related protein or a motility-related protein). All the regulators were identified on the basis of a growth-essential or growth-defective phenotype in attachment. As a result, genes in proximity to a regulator that have growth-defective or growth-essential phenotypes in attachment identify themselves as putative regulator targets and suggest that the regulator may be functioning as an activator. Genes in proximity to a regulator that have growth-advantaged phenotypes in attachment also identify themselves as putative regulator targets, but suggest that the regulator may function as a repressor.

The regulon of pRL100172 (*rhiR*) served as a positive control for this regulator target screening method: RhiR is known to activate expression of the other *rhi* genes it is located adjacent to, *rhiABC*. All of the *rhiABC* genes showed identical growth-defective phenotypes to *rhiR* upon mutation in root attachment. While this was the case with RhiR, it must also be considered that mutations to single genes may cause no effect, but mutation to a regulator may have synergistic effects and a mutational defect is only seen when the regulator is mutated.

Putative regulator targets of the eleven transcriptional regulators were identified by screening their adjacent genomic regions (Table 4-9). It seems possible that both the putative LuxR/GerE family transcriptional regulator (pRL90149) and the putative FNR/CRP family transcriptional regulator (RL3828) could be involved in the regulation of polysaccharide biosynthesis and processing (Table 4-9). In addition, three of the transcriptional regulators (pRL80032, pRL120692, RL0561) were located adjacent to putative methyl-accepting chemotaxis proteins (Table 4-9), that have a logical role in motility.

Transcriptional regulator	Putative regulator targets		INSeq classification				
	Locus	Description	In	Rh	Ro	No	Re
pRL80032	pRL80031 (mcpS)	putative methyl-accepting chemotaxis chemoreceptor	NE	NE	NE	NE	NE
pRL90149	pRL90141	hypothetical protein	NE	GD	GD	GD	GD
	pRL90143	putative transposase	NE	GD	GD	GD	GD
	pRL90144	putative exopolysaccharide biosynthesis-related protein	NE	NE	NE	NE	NE
	pRL90146	putative exopolysaccharide biosynthesis UDP-galactose-lipid carrier transferase	NE	NE	NE	NE	NE
	pRL90156	putative transmembrane polysaccharide biosynthesis protein	NE	NE	NE	NE	NE
	pRL90158	putative exopolysaccharide production protein	NE	NE	NE	NE	NE
pRL100057	pRL100053	putative transmembrane protein	NE	NE	ES	GD	GD
	pRL100055	putative acetolactate synthase subunit	NE	GD	GA	NE	ES
	pRL100056	putative aromatic amino acid aminotransferase	NE	NE	GD	NE	NE
pRL120692	pRL120683 (mcpB)	putative methyl accepting chemotaxis protein	NE	NE	NE	NE	NE

RL0256	RL0252	putative nodulation protein	NE	NE	NE	NE	NE
	RL0254 (lepA)	putative GTP-binding protein	NE	GD	GD	ES	ES
	RL0255	putative HemeO family (TenA) transcriptional regulator	NE	GA	GA	NE	NE
RL0561	RL0558	putative transmembrane cellulose synthase catalytic subunit [UDP-forming]	NE	NE	NE	NE	NE
	RL0559	putative UDP-glucose 4-epimerase	NE	NE	NE	NE	NE
	RL0564 (mcpE)	putative methyl-accepting chemotaxis protein	NE	NE	NE	NE	NE
RL1040	RL1047	putative ATP-binding component of ABC transporter MZT	NE	NE	GD	NE	NE
	RL1048	putative permease component of ABC transporter MZT	NE	NE	GD	NE	NE
RL2828	RL2829 (cobO)	putative cob(I)yrinic acid a,c-diamide adenosyltransferase	GD	ES	GD	ES	ES
	RL2830 (cobN)	putative cobalamin synthesis protein	GD	ES	GD	ES	ES
	RL2831 (cobW)	putative cobalamin synthesis protein	GD	GD	GD	GD	GD
	RL2832 (cobP)	putative bifunctional adenosylcobalamin biosynthesis protein	GD	GD	GD	GD	GD
	RL2836 (cobQ)	putative cobyric acid synthase	NE	GD	GD	ES	ES

	RL2837	putative SBP of ABC transporter QAT? orphan proline/glycine betaine	NE	GD	GD	ES	ES
RL3828	RL3820	putative exopolysaccharide production protein	NE	NE	NE	ES	ES
	RL3821	putative tyrosine-protein kinase involved in exopolysaccharide biosynthesis	NE	NE	NE	NE	NE
	RL3822	putative glycosyl transferase	NE	NE	NE	NE	NE
	RL3823	putative exopolysaccharide production protein	NE	NE	NE	NE	NE
	RL3827	conserved hypothetical protein	GA	NE	GD	NE	NE
	RL3829 (exoY)	putative exopolysaccharide production regulator	NE	GD	NE	NE	GA
	RL3831	putative O-acetyl transferase	NE	GA	GA	NE	NE

Table 4-9. Putative regulator targets of the INSeq-identified root attachment transcriptional regulators identified on the basis of genomic proximity, predicted function and INSeq phenotype. In: input, Rh: rhizosphere, Ro: root attachment, No: nodulation, Re: nodule recovery. HMM classifications are given as NE: growth-neutral, GD: growth-defective, ES: growth-essential, GA: growth-advantaged.

4.3.12 RNA sequencing to characterise root attachment transcriptional regulatory networks

pK19mob insertion mutants of these eleven transcriptional regulator genes were made: OPS1032 (*pRL80032*::pK19mob), OPS1033 (*pRL90149*::pK19mob), OPS1034 (*pRL100057*::pK19mob), OPS1035 (*pRL100172*::pK19mob), OPS1036

(*pRL120160*::pK19mob), OPS1037 (*pRL120692*::pK19mob), OPS1038 (*RL0256*::pK19mob), OPS1039 (*RL0561*::pK19mob), OPS1040 (*RL1040*::pK19mob), OPS1042 (*RL2828*::pK19mob), OPS1043 (*RL3828*::pK19mob). RNA sequencing of these strains alongside wild-type Rlv3841 attaching to pea roots was carried out to try and characterise their regulator targets [104]. RNA sequencing was carried out on strains attaching to roots in order to sequence their transcriptome under conditions most relevant to the requirements for root attachment. Although these regulatory genes were characterised as impairing root attachment when mutated according to INSeq analysis (Table 4-7), it was considered that these mutations would likely not completely obliterate root attachment (i.e. some cells would still attach), and that in single-strain inocula for RNA sequencing these attachment defects would likely be significantly less extreme. The method for RNA sequencing is detailed in chapter 2; materials and methods.

4.3.12.1 RNA sequencing data analysis

A total of 81,784,050 usable reads were obtained from sequencing. Reads were separated by strain and all genome coordination of tRNA/rRNA sequences were filtered to remove any reads that escaped the *in vivo* clean up. mRNA sequence reads were mapped using protein translation to the Rlv3841 genome and the degree of gene expression quantified using EDGE-pro [142]. The RNA sequencing data for the 11 mutants were filtered for genes strongly downregulated (at least a 2.5-fold decrease in expression) and strongly upregulated (at least a 2-fold increase in expression). These boundaries represent the average top 1% upregulated and downregulated genes within these datasets.

Unfortunately, upon inspection of the RNA sequencing datasets, the data was not as clear to analyse as hoped. Within the top 1% upregulated and downregulated gene groups, genes were generally in a steady gradient of expression changes, and single standout candidates could not be identified. It appeared a lot of changes in expression were a consequence of lower fitness in root attachment as opposed to a potential causation. For example, a large number of ribosomal proteins showed decreases in expression across many of the regulator mutants. This is more likely to signal halted growth in the mutant as opposed to suggesting putative directly regulated candidates.

There were observed to be a number of common genes in the upregulated and downregulated groups across the defective root attachment transcriptomes of the regulators. RNA sequencing data analysis was unable to confidently predict strong directly regulated candidates as genes with significantly greater expression changes to the rest, instead, genes were generally in a steady gradient of expression changes. As a result, RNA sequencing data analysis was split into two parts: 1) identification of common genes (present in two or more groups) across RNA sequencing datasets (common genes identified in datasets using the multiple list comparer tool in molbiotools: http://www.molbiotools.com/listcompare.html) and 2) analysis of the expression level changes of putative regulator targets previously identified on a basis of genomic proximity (Table 4-9). Characterisation of common genes within this top 1% of upregulated and downregulated genes between regulator mutants was carried out with the aim of characterising common factors in the root attachment defective transcriptome.

4.3.12.2 Identification of common factors in a root attachment defective transcriptome

A total of 233 common downregulated and 98 common upregulated genes were identified within this top 1% of upregulated and downregulated genes between regulator mutants (molbiotools). Common genes were defined as those present in two or more regulator groups. Indeed, within the common downregulated group, numerous genes suggesting the strain had stopped growing were identified, including the large number of ribosomal proteins, present in between five and two common groups (RL1790, RL1792, RL1799, RL4624, RL0267, RL1060, RL1673, RL1765, RL1779, RL1782, RL1786, RL1789, RL1791, RL4017 and RL4676). A number of factors in this common root attachment defective transcriptome remain to be characterised: unknown hypothetical proteins accounted for 19 out of the 48 genes present in 3 or more highly upregulated or downregulated RNA sequencing datasets.

Within the group of common downregulated genes, 52 out of these 233 genes also showed a growth-essential or growth-defective phenotype in root attachment. Some of these genes encoded proteins with a predicted role in chemotaxis: RL0692, RL0693 and RL4386, encoding putative chemotaxis related proteins; pRL120068, encoding a methyl-accepting chemotaxis sensory transducer; RL0725, encoding a putative chemotaxis MotD protein; and RL1083, encoding a putative disulphide bond formation protein. Disulphide bonds can function universally across surface interactions to form rapid and reversible bonding interactions [230, 231], and it is logical that downregulation of this gene might be detrimental to attachment.

The other 181 common downregulated genes gave a growth-neutral or growth-advantaged phenotype in root attachment upon mutation. One of these growth-advantaged genes is RL0718, encoding a putative flagellin A protein. Flagellin proteins are involved in flagella-related processes, including in flagella biosynthesis, motility regulation and surface interactions. This protein might function in flagella motility regulation, such that misregulated motility (i.e. increased or decreased) may be beneficial to root attachment under these conditions. Flagella biosynthesis and motility is also known to have a high energetic cost. Interestingly, pRL120065, encoding a putative symbiosis-related calsymin, was also within this common downregulated group with a growth-neutral phenotype in root attachment upon mutation. The protein product of pRL120065 has a very high homology (72.8% amino acid identity) to the symbiosis-related calsymin found in *Rhizobium etli* CFN42 [143]. In *R. etli*, this calmodulin-like protein has already been shown to be exclusively expressed during colonisation and infection [232], and is believed to play a role in the molecular dialogue between prospective symbiotic partners, through acting as a calcium binding receptor. While this calsymin protein may be utilised during colonisation in Rlv3841, it does not appear to cause a fitness defect as a result of single mutation.

A number of defence or stress related genes appeared in the common downregulated genes; there is a logical link between the decreased expression of defence or stress related proteins and a lower fitness in root attachment: the rhizosphere and root pose an oxidative environment that requires sufficient bacterial defence- and stress-related machinery to navigate. These genes included RL0595, encoding a cold-shock DNA-binding protein; pRL110148, encoding a putative 2-hydroxy-3-oxopropionate reductase; pRL110539, encoding a putative oxidoreductase; pRL90212, encoding a

putative peptide methionine sulfoxide reductase; RL0277, encoding a putative oxidoreductase; RL0384, encoding a putative small heat shock protein; RL0394, encoding a putative phosphate starvation-induced protein; RL2391, encoding a heat shock protein HspQ; RL3255, encoding a dihydrofolate reductase; and RL3440, encoding a putative general stress protein.

Finally, within the group of common upregulated genes, two are worth highlighting with an especially clear link to attachment defects: RL0081, encoding a putative polysaccharide degradation protein, and RL3537, encoding the nodulation repressor NolR. Polysaccharides are a major cell surface constituent and play important roles in facilitating attachment. It is logical that an upregulation to polysaccharide degradation could have negative effect effects on fitness in attachment. Similarly, upregulation to a nodulation repressor can also logically be linked to negative fitness effects. The number of significantly downregulated and upregulated common genes between regulator mutants suggest an underlying common root attachment defective transcriptome in these strains and highlight how vast and interconnecting transcriptional regulatory networks can be. It is likely there are a number of additional regulators linking the common regulator effects between these mutants, and this network of indirect effects is responsible for the common factors observed.

4.3.12.3 Putative regulator targets identified through genomic proximity and expression level changes

Initial putative regulator targets were previously suggested on a basis of genomic proximity, INSeq phenotypes and functional prediction (Table 4-9). Putative regulator targets were then redefined from the RNA expression level changes of the upstream and

downstream regions of these regulators (Table 4-10); these regulator targets likely fall into an operon. Expression level changes are shown as the differential expression between regulator mutant and wild-type Rlv3841 (log_2 fold-change). Putative regulator targets were identified on the basis of genomic proximity to the regulator locus and identification of expression changes indicating significant upregulation (greater than 1.5 log_2 fold-change) or downregulation (greater than -1.5 log_2 fold-change) (Table 4-10) [104].

The majority of the putative regulator targets (Table 4-9) were not carried forward in the second screen for target identification (Table 4-10). For the regulator encoded by pRL80032, genes pRL80020-pRL80025 all had expression changes of -2.7 log_2 fold-change or greater in the mutant and seem to be the strongest candidates. For the regulator encoded by pRL90149, the polysaccharide biosynthesis genes were no longer indicated as putative regulator targets, and pRL90143 encoding a putative transposase with a 2.46 log_2 fold-change in the mutant seemed instead to be the strongest candidate. For the regulator encoded by pRL100057, pRL100053, encoding a putative transmembrane protein, was extremely upregulated (a 5.78 log_2 fold-change) in the mutant, and this was one of the genes that had previously been identified by INSeq and proximity analysis (Table 4-9).

A number of genes were identified as putative regulator targets of the pRL120160 encoded regulator (Table 4-10), and these genes demonstrated a common theme of putative transporter system encoded proteins. Only one gene was identified as a putative regulator target of pRL120692 (Table 4-10), pRL120701, and with only an upregulation of 1.6 log_2 fold-change in the mutant this gene was not deemed a particularly confident

candidate. An absence of convincing candidates was also observed for the regulators encoded by RL0256 and RL1040 (Table 4-10).

For the RL0561 encoding regulator, the strongest expression change was observed for RL0556, predicted to encode a conserved hypothetical exported protein, which was -3.31 $\log_2$ fold-change downregulated. For RL2828, a number of cobalamin-related protein encoding genes had previously been suggested as putative regulator targets (Table 4-9), and one of these genes (RL2832, *cobP*) was also significantly upregulated in the RNA sequencing analysis (Table 4-10). Finally, for the regulator encoded by RL3828, it had previously been suggested that the regulon contains a number of polysaccharide biosynthesis and processing related genes (Table 4-9). The RNA sequencing analysis was in support of this; five of the adjacent polysaccharide-related genes were significantly downregulated in the mutant.

Regulator	Putative regulator target		Differential expression between regulator mutant and wild-type Rlv3841 (log_2 fold-change)
	Locus	**Description**	
pRL80032	pRL80020	conserved hypothetical protein	-3.73
	pRL80022	conserved hypothetical protein	-3.88
	pRL80023	putative carbon monoxide dehydrogenase subunit M	-3.11
	pRL80024	putative iron-sulphur cluster carbon monoxide dehydrogenase subunit S	-3.34
	pRL80025	putative dehydrogenase/reductase subunit L	-2.73
	pRL80029	putative ATP-binding component of ABC transporter HAAT	-1.94
	pRL80036	conserved hypothetical protein	-2.75
	pRL80041	putative histidinol dehydrogenase	-1.95
pRL90149	pRL90143	putative transposase	2.46
	pRL90157	hypothetical protein	1.58

pRL100057	pRL100051	hypothetical protein	1.71
	pRL100053	putative transmembrane protein	5.78
pRL120160	pRL120148	hypothetical protein	2.19
	pRL120157	conserved hypothetical protein	-2.00
	pRL120159	putative amidohydrolase	1.58
	pRL120163	putative permease component of ABC transporter PepT	1.87
	pRL120164	putative SBP of ABC transporter PepT	1.74
	pRL120167	putative MFS family transporter	2.43
pRL120692	pRL120701	putative transmembrane magnesium and cobalt transport protein, CorA family	-1.62
RL0256	RL0264	conserved hypothetical protein	1.60
RL0561	RL0552	hypothetical protein	1.85
	RL0555	putative deoxycytidine triphosphate deaminase	1.95

	RL0556	conserved hypothetical exported protein	-3.31
RL1040	RL1036	conserved hypothetical protein	-1.56
	RL1049	putative SBP of ABC transporter MZT	1.72
RL2828	RL2820	hypothetical protein	-1.54
	RL2832 (cobP)	putative bifunctional adenosylcobalamin biosynthesis protein	1.89
RL3828	RL3817	putative glycosyltransferase	-3.43
	RL3818	putative glycosyl transferase	-1.83
	RL3822	putative glycosyl transferase	-1.81
	RL3825	putative transmembrane protein	2.57
	RL3829 (exoY)	putative exopolysaccharide production regulator	-1.54
	RL3831	putative O-acetyl transferase	-1.51
	RL3835	conserved hypothetical exported protein	1.74

Table 4-10. Putative regulator targets of the transcriptional regulators identified on the basis of RNA sequencing differential expression changes between regulator mutant

and wild-type Rlv3841(log_2 fold-change) indicating significant upregulation (greater than 1.5 log_2 fold-change) or downregulation (greater than -1.5 log_2 fold-change) of genes in genomic proximity of regulator locus [104].

4.3.13 Mutation of a putative LysR family transcriptional regulator on pRL8 prevents root attachment

The putative LysR family transcriptional regulator encoded by pRL80032 was chosen as the candidate with which to provide an example of downstream validation and characterisation. A strain mutated in pRL80032 (OPS1032) was observed to have a strikingly different physical phenotype to the wild-type Rlv3841 strain. OPS1032 formed large, mucoid, amorphous colonies on agar plates, and flocculated in liquid culture (minimal media). These altered physical properties seem indicative of cell surface changes. LysR family transcriptional regulators are regarded as global transcriptional regulators and can act as activators or repressors. NodD is another LysR family transcriptional regulator and is well characterised in *R. leguminosarum.* The observation that disruption of a gene encoding a transcriptional regulator on pRL8 causes significant impairment to root attachment is consistent with transcriptome analyses describing pRL8 as the pea-specific rhizosphere plasmid in Rlv3841 [57].

The INSeq experimental set-up poses a highly competitive environment, which improves the resolution of defects in root attachment by competition of these mutants with a vast fold higher population of strains with “wild-type” attachment properties. Mutation of pRL80032 impaired competitive attachment in the INSeq environment and OPS1032 was severely defective in root attachment when inoculated as a single strain onto roots (Table 4-11). In single strain root attachment assays, the root attached

population of OPS1032 grew to 180 ± 15 cfu per plant at 5 dpi, a 5 orders of magnitude reduction from the $5.4x10^7 \pm 8.5x10^6$ wild-type Rlv3841 cfu able to attach in the same assay (Table 4-11).

Strain	**Root attached population (CFU) per pea seedling at 5dpi (± S.E.M.)**
Rlv3841 (wild-type)	**5.4×10^7** (± 8.5×10^6)
OPS1032	**180** (± 15)

Table 4-11. The mean population of cells attached to pea seedling roots at 5 dpi for OPS1032 and the wild-type Rlv3841 strain (cfu; colony-forming units) calculated from five independent biological replicates. The standard error of the mean (± S.E.M.) is shown in brackets.

The growth-defective phenotype of mutation of pRL80032 is specific to root attachment, as predicted by INSeq (Table 4-8). OPS1032 colonised the rhizosphere at a similar capacity to wild-type Rlv3841 (Fig. 4-9). In rich TY media, OPS1032 also has a similar mean generation time (MGT) to the wild-type Rlv3841 strain (Fig. 4-10): 2.9h ± 0.05 compared to 2.7h ± 0.08. These results validate the INSeq predictions of mutation of pRL80032: growth-defective in the root attachment, growth-neutral in the rhizosphere and growth-neutral in selection on TY (input library). This directed mutational characterisation of pRL80032 (OPS1032) exemplifies the pathway of downstream INSeq target validation. A more extensive characterisation and validation of INSeq identified phenotypes is being carried forward as the project for another PhD student.

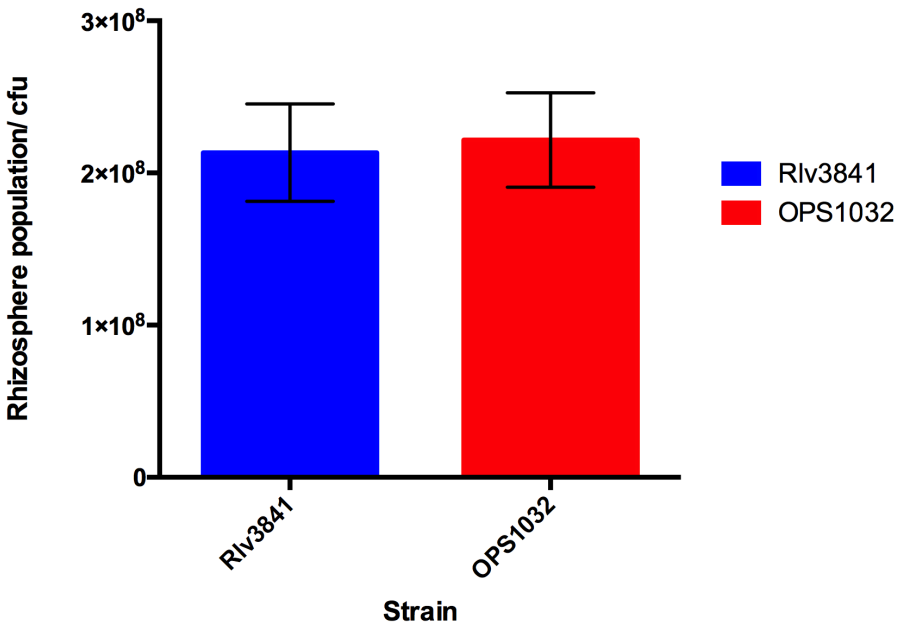


Figure 4-9. The mean population of cells (cfu) in the pea rhizosphere at 5 dpi for OPS1032 and the wild-type Rlv3841 strain calculated from five independent biological replicates. Error bars show standard error of the mean.

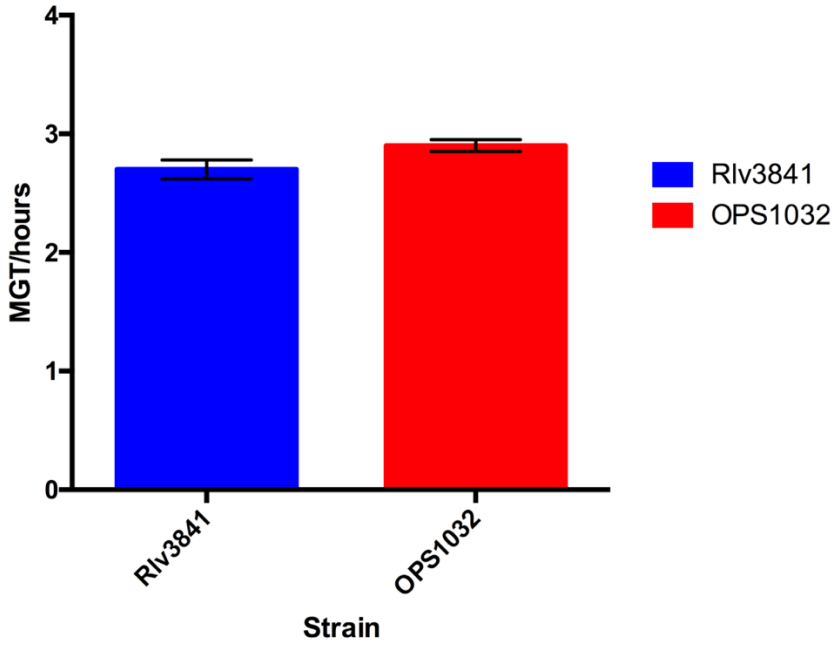


Figure 4-10. The mean generation time (MGT) of Rlv3841 and OPS1032 growing in TY cultures calculated from five independent biological replicates. Error bars show standard error of the mean.

4.4 Conclusion

In summary, an INSeq analysis of Rlv3841 engaging in symbiosis with pea has enabled identification of 463 genes required for successful nodulation; 176 genes are required from the rhizosphere through to nodulation ('rhizosphere-progressive' genes), 14 genes are required from root attachment to nodulation ('attachment-progressive' genes), 39 genes are required specifically for bacteroid development, 154 genes are required in both bacteroids and nodule bacteria, and 80 genes are required specifically in nodule bacteria (Fig. 4-5). This number of 463 genes is significantly higher than has previously been suggested in previous genomic studies, and highlights the power of INSeq in resolving competitive defects between mutant strains. Focusing specifically on early-stage symbiosis, there was the identification of 55 genes that are required specifically for growth in the rhizosphere (Fig. 4-5), mutation of which does not impair root attachment or nodulation, and 101 genes that are required specifically for root attachment (Fig. 4-5), mutation of which does not impair nodulation.

This study did not distinguish between attachment to root hairs (required for nodulation) and attachment to the root epidermal surface. However, it is likely that the attachment genes that are progressive with regards to bacteroid formation (Appendix A, Table 9) are on the root hair infection pathway needed for nodulation. This attachment-progressive group contained only 14 genes (Appendix A, Table 9), and included in this group were three putative transmembrane proteins (pRL100053, RL0817, RL4183), a putative exopolyphosphatase (RL1600; *ppx*) and a putative UDP-glucuronate 5'-epimerase (RL3677; *lspL*). The larger group of attachment-specific genes (101 genes; Appendix A, Table 9) seem to indicate a root attachment phenotype that is not directly linked to nodulation, and likely represents attachment to the root epidermal surface.

There is a need for further work to distinguish between attachment to the root hairs and the epidermal surface, and this would help in understanding how attachment with the potential to lead to nodulation differs from other types of attachment. For this work, rhizobia attached to different zones of the root would need to distinguished, i.e. the retrieval of rhizobia attached specifically to root hairs.

Growth in the rhizosphere was found to require a host of defence and stress-related machinery. A number of transporter systems were identified with predicted cargos including ferric cations, carbohydrates, and nitrate/nitrite/cyanate. Two groups of rhizosphere genes were identified: rhizosphere-specific genes (55 genes) and rhizosphere-progressive genes (176 genes) (Appendix A, Table 9). In considering how these two groups have arisen, it can be hypothesised that there are (at least) two growth phases in the rhizosphere. It is likely growth strategies need to change over the period of colonisation, for example, in response to changing root exudates, an increased rhizobial population in the environment and saturation of attachment sites on the root.

It can be hypothesised that there is an early growth phase, coupled to attachment and colonisation, and a secondary growth phase, associated with more long-term survival in the rhizosphere. An early growth burst may be covering the root in bacteria such that there are no subsequent sites for easy colonisation, and so a secondary growth phase is required that involves a different strategy and focuses on the ability to grow more slowly and steadily in the rhizosphere and to metabolise root exudates effectively. Two different growth phases could explain why two different populations of mutants are observed; the rhizosphere-progressive genes are important for the early growth phase coupled to attachment and colonisation, and the rhizosphere-specific genes are

important for the secondary growth phase and more long-term survival in the rhizosphere. Changes to root exudation likely also sculpt differential fitness in the rhizosphere. Root exudation may be altered by the plant over the period of colonisation, and if a new metabolite is secreted after 2-3 dpi, mutants defective in its catabolism will be disadvantaged for long term rhizosphere growth (i.e. rhizosphere-specific) but have not been disadvantaged in early growth and an initial potential to lead to attachment (i.e. rhizosphere-progressive). Further work is needed to characterise growth strategies in the rhizosphere over a period of colonisation. This could be achieved by an INSeq analysis of mutants retrieved from the rhizosphere over a wider range of sampling timepoints, i.e. at 1 dpi, 3 dpi, 5 dpi and 7 dpi instead of just at 5 dpi.

A significant proportion of the genes required for root attachment were found on pRL10, the 'symbiosis' plasmid in Rlv3841. There were 15 genes on pRL10 in the attachment-specific group and 2 genes on pRL10 in the attachment-progressive group; both types of attachment phenotypes require 'symbiosis' plasmid genes. Root attachment is a highly coordinated process: responding to chemotaxis-inducing root exudates and components of the root surface requires signal perception and a regulated response. Consistent with this, 13.0% of the genes identified as being required for root attachment were predicted to have a regulatory function. Targeted mutations were made to eleven INSeq-identified root attachment transcriptional regulators. RNA sequencing of these regulator mutants was carried out to try and characterise their directly regulated targets, with the overall aim of mapping the transcriptional regulatory pathways required for root attachment. Unfortunately, upon inspection of the RNA sequencing datasets, the data was not as clear to analyse as had been hoped. Within the top 1% upregulated and downregulated gene groups, genes were generally in a steady gradient of expression

changes, and single standout candidates could not be identified. It appeared a lot of changes in expression were a consequence of lower fitness in root attachment as opposed to a potential causation.

Putative regulator targets were predicted on a basis of genomic proximity combined with INSeq data analysis, RNA sequencing data analysis and protein predictions (Table 4-9, Table 4-10). The RNA sequencing data was also analysed for common factors across the root attachment defective transcriptomes. Approximately 25% of the shared significantly downregulated genes were genes that had also been classified in INSeq with a growth-defective or growth-essential phenotype upon mutation for root attachment. These included genes with functions in chemotaxis, motility and adherence. There seemed to be a significant amount of noise generated in the RNA sequencing datasets that made it difficult to confidently predict direct targets of the regulators. This was presumably the result of a combination of effects, including the impacts of lower fitness in root attachment and halted growth, and the interconnection of regulatory networks and indirect regulator effects. Chromatin immunoprecipitation (ChIP) sequencing would provide an alternative way to investigate transcriptional regulator targets, potentially more accurately, as it combines chromatin immunoprecipitation with DNA sequencing to map global binding sites [233]. This would identify DNA sequences directly interacting with the regulators, but is significantly more expensive and time-consuming.

Analysis of central carbon metabolism pathways identified the core components of the ED pathway and the TCA-cycles to be required synonymously across growth in the rhizosphere, root attachment and nodulation. Metabolism of glucose and phosphorylated

intermediates via the methylglyoxal pathway was comparatively more important for growth in the rhizosphere, presumably relating to the requirements for efficient metabolism of these compounds in the root exudates. Glycogen synthesis and gluconeogenesis were comparatively more important in bacteroids. It was hypothesised that bacteroids might synthesise glycogen from plant provided TCA-cycle intermediates that undergo gluconeogenesis to generate glycogenic substrates. Synthesising glycogen via gluconeogenesis, a pathway that results in the net oxidisation of NADH to NAD^+ and the hydrolysis of ATP, would assist with redox and energy storage at a time when bacteroids are carrying out the highly reductive and energy dependent process of N_2-fixation. It is possible that glycogen synthesis might act as a balancing or overflow pathway to consume excess reductant or ATP if these accumulate to levels that inhibit N_2-fixation. This analysis highlights the use of INSeq in generating metabolic hypothesises, and metabolic flux analyses and carbon labelling experiments could provide a way to test this hypothesis in the cell.

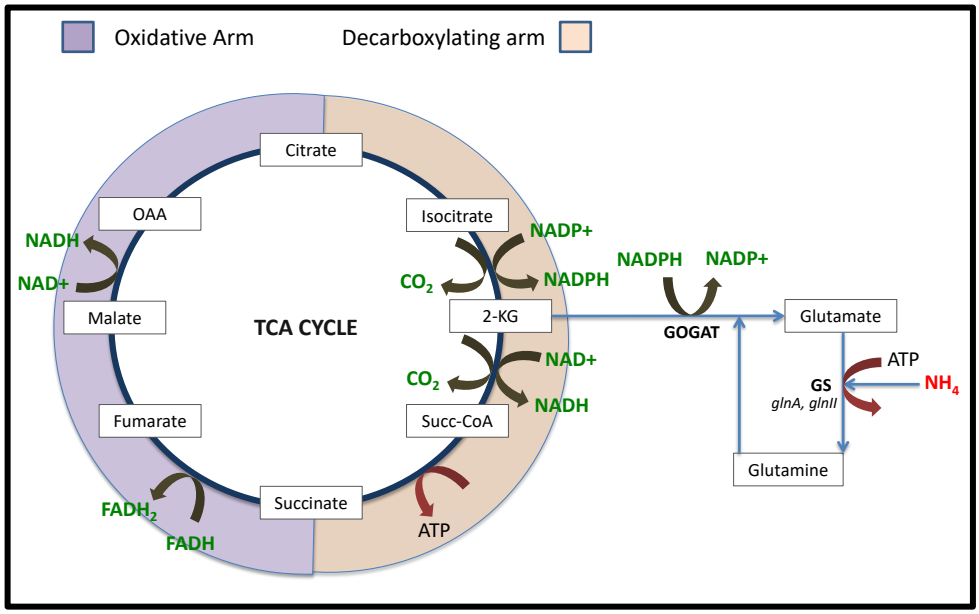
This study was the first application of INSeq *in planta* conditions, and required careful experimental design and adaptations in order to retrieve sufficiently large and saturated mutant pools from the rhizosphere and roots. In terms of the limitations of this work, there were deviations from field conditions for the interaction of *R. leguminosarum* with its host pea in the use of fine vermiculite and sterile conditions with no other competing microbes. A future direction to this study would be to investigate growth in the rhizosphere and attaching to roots in the presence of other competing microbes, or in a non-sterile soil environment. This would provide a better representation of in field conditions, and would also be a highly insightful investigation into the rhizobial fitness determinants of microbial competition for growth substrates, space, root proximity and

surface attachment. Careful experimental design would be required to ensure isolation of the correct rhizobial insertion library DNA, and to deal with presence of in-soil contaminants.

Unfortunately, further target characterisation from the dataset generated in this project was not possible in the time frame. However, these datasets are being carried forward in five additional projects to date. The characterisation of pRL80032 provides an example of the pipeline for INSeq validations to be carried out. Time permitting, I would have liked to have generated luciferase promoter fusions of the INSeq-identified attachment transcriptional regulators. These could have been used to characterise the temporal and spatial activation of these transcriptional regulators on the roots by quantification of luciferase expression in the promoter-fusion strains. It would have been interesting to investigate whether these attachment defects are pea-specific, and this could be achieved by attachment assays on a range of other host legume, non-host legume and non-legume plants.

Chapter 5

Low O_2 prevents rhizobia from assimilating ammonium and induces amino acid auxotrophy



5.1 Abstract

Rhizobium-legume symbioses make one of the largest contributions to biologically available nitrogen. Axiomatic to the evolution of these symbioses is the secretion of ammonium to the plant rather than its assimilation by the N_2-fixing rhizobia. Their secretion of ammonium is unique to the nodule environment and is always fuelled by the legume providing a dicarboxylate carbon source. Here I show that providing rhizobia with dicarboxylic acids in a low O_2 environment restricts their growth via ammonium assimilation, but that growth can be restored by providing glutamate as a nitrogen source. This is because ammonium assimilation relies on the removal of 2-ketoglutarate from the TCA-cycle to form glutamate, forcing a turn of the TCA-cycle with the generation of NADH, NADPH and $FADH_2$, whose reoxidation is limited by low O_2-availability. A reoxidation block model is proposed, in which the provision of dicarboxylates under low O_2 prevents removal of 2-ketoglutarate from the TCA-cycle and the assimilation of ammonium. This simple reoxidation block suggests a powerful selection pressure for the evolution of ammonium secretion for dicarboxylate exchange and the development of mutualism.

5.2 Introduction

While the atmosphere consists of 78% N_2 this is inert, making biologically available nitrogen (e.g. ammonium and nitrate) one of the principal factors limiting biological productivity on earth. Biological N_2-fixation carried out by bacteria provides ~65% of the biosphere's available nitrogen [18], with the single greatest contribution coming from symbioses formed between soil-dwelling rhizobia and plants of the legume family [6, 14, 15]. To engage in symbiosis, rhizobia attach to and colonise legume roots, becoming internalised and induce the development of root nodules in which they are housed. In nodules rhizobia become surrounded by a plant-derived symbiosome membrane and undergo differentiation into N_2-fixing bacteroids [65]. Bacteroids fix atmospheric nitrogen (N_2) into ammonium (NH_4^+), which is secreted to the legume host. N_2-fixation is an energetically expensive process requiring at least 16 molecules of ATP per N_2 reduced and this is fuelled in bacteroids by plant derived C_4-dicarboxylates feeding into the TCA-cycle [29]. Three fundamental requirements of these symbioses are therefore; the provision of a carbon source from the legume, a microaerobic environment for the O_2-sensitive nitrogenase that catalyses N_2-fixation and the shutdown of ammonium assimilation in bacteroids, to prevent their consumption of the fixed nitrogen [60, 234].

Symbiotic N_2-fixation relies on a precise and efficient integration of bacterial and plant metabolism. Carbon and nitrogen metabolism in the *Rhizobium*-legume symbioses has been reviewed several times in recent years [29, 36, 234, 235]. A paradox appears to exist within the symbiosis: the C_4-dicarboxylates that fuel N_2 fixation in the low-O_2 environment of bacteroids are catabolised by the O_2-dependent tricarboxylic acid (TCA) cycle [36, 234]. This raises the question of why carbon sources are being

provided that feed into an O_2-intensive metabolic cycle? Turning of the TCA-cycle generates NADH, NADPH and $FADH_2$ reductant molecules that require reoxidation via the electron transport chain. And more specifically, is there a reason why a less O_2-intensive energy supply has not been selected within the symbioses? Mutational studies of enzymes in the TCA-cycle show certain reactions to be required for N_2 fixation in *Rhizobium leguminosarum* and *Sinorhizobium meliloti* [153, 155, 236], however, the entire cycle may not be essential. This is observed in *Bradyrhizobium japonicum* bacteroids, where the decarboxylating arm (citrate to succinyl-CoA) can be inactivated by mutation [156, 157, 177, 237]. Many prokaryotes, those with an anaerobic lifestyle in particular, use an incomplete TCA-cycle whose reactions aim at the synthesis of essential compounds and at the prevention of the build-up of reducing equivalents [238, 239].

Axiomatic to the evolution of *Rhizobium*-legume symbioses is the shut-down of ammonium assimilation in bacteroids and secretion to the plant. Ammonium assimilation depends on the removal of 2-ketoglutarate from the TCA-cycle as the carbon backbone for glutamate synthesis (Fig. 5-1) [182, 240-242]. In this chapter I present data suggesting that the removal of 2-ketoglutarate from the TCA-cycle may become problematic under low O_2. Turning of the cycle to replenish the 2-ketoglutarate is an O_2-intensive process, generating NADH, NADPH and $FADH_2$ that require O_2 for reoxidation (Fig. 5-1) [240, 243]. Reoxidation occurs through the step-by-step transfer of electrons via the electron transport chain to the final electron acceptor, O_2, in a reaction that forms H_2O and regenerates oxidised NAD and NADP [244, 245]. The provision of C_4-dicarboxylates that feed into the TCA-cycle further exacerbates cellular

O_2-requirements, and is suggested to be key in restricting the removal of 2-ketoglutarate from the TCA-cycle for the assimilation of ammonium.

In this chapter I propose a reoxidation block model, in which the provision of dicarboxylates under low O_2 prevents removal of 2-ketoglutarate from the TCA-cycle and the assimilation of ammonium. This reoxidation block suggests a novel driving force behind ammonium export in symbioses, alongside potentially explaining why legumes always provide TCA-cycle dicarboxylates to fuel N_2-fixation in the low O_2 environment of the nodules.

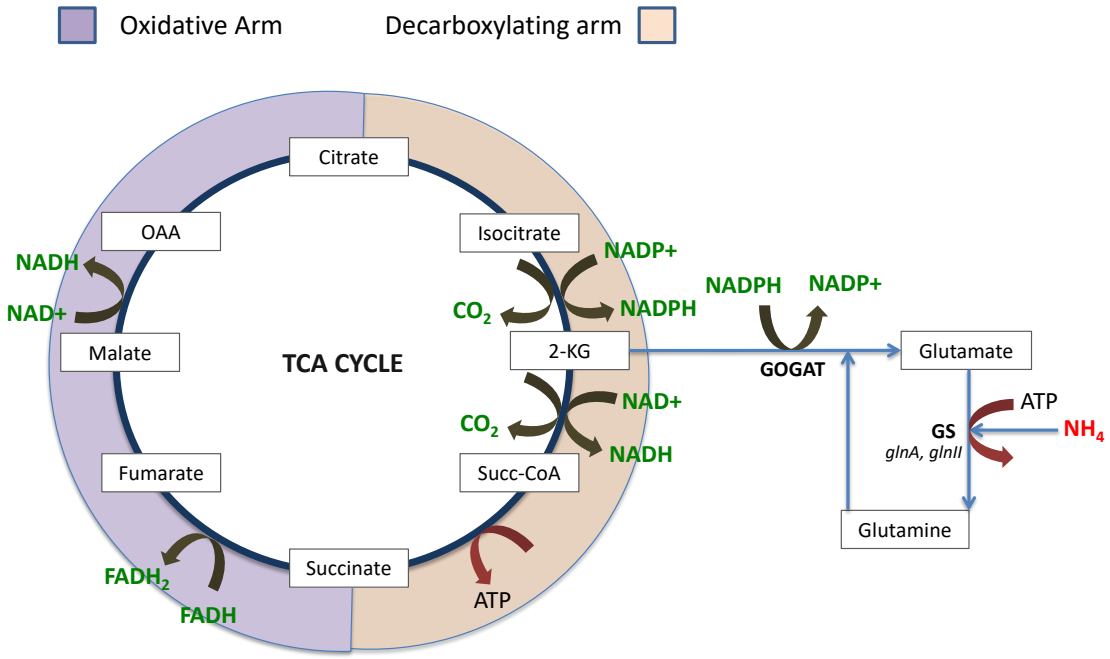


Figure 5-1. The TCA-cycle and ammonium assimilation pathway in *R. leguminosarum*. The TCA-cycle is divided into two stages: 1) the decarboxylating arm (shown in orange), involving the decarboxylating reactions converting citrate to succinyl-CoA and 2) the oxidative arm (shown in purple), involving the successive oxidation of succinate to fumarate, fumarate to malate and malate to oxaloacetate. 2-ketoglutarate (2-KG) can

be used a substrate to generate glutamate, through the activity of glutamine oxoglutarate aminotransferase (GOGAT), glutamate can then be used a substrate to generate glutamine, through the activity of glutamine synthetase (GS), and requiring the incorporation of ammonium (NH_4).

5.3 Results

5.3.1 RL2392 (*glnB*) is essential for growth on dicarboxylates under low O_2

Insertion sequencing (INSeq) analysis of *Rhizobium leguminosarum* bv. *viciae* 3841 (Rlv3841) was used to compare catabolism of the TCA-cycle intermediate succinate and the sugar glucose at both 21% and 1% O_2 (chapter 3) [130]. The combination of 1% O_2 and succinate provides conditions most similar to those experienced by bacteroids and only one gene, RL2393 (*glnB*) was uniquely required for growth under these conditions [130]. RL2393 (*glnB*) encodes the nitrogen regulatory protein (PII) that is a component of the bacterial nitrogen regulation (Ntr) system and an essential regulator of ammonium assimilation [208]. A uridylyltransferase (UTAse) responds to the cellular nitrogen status by uridylylating or deuridylylating the PII protein in conditions of nitrogen-limitation or nitrogen-excess, respectively [246-248]. The PII protein acts as an intracellular signal transmitter and regulates activity of the Ntr system proteins (NtrB and NtrC) and in turn the transcription of glutamine synthetase (GS) [246]. Glutamine synthetase catalyses the reaction that converts glutamate and ammonium to form glutamine (Fig. 5-1) [246, 247], and Rlv3841 has two GS isoenzymes, GSI (*glnA*) and GSII (*glnII*) [249, 250].

Under conditions of nitrogen excess UTAse deuridylylates and activates PII in Rlv3841 [246]. Deuridylylated PII promotes the dephosphorylation and inactivation of NtrC,

which effectively functions to repress transcriptional initiation and accumulation of GS. Under conditions of nitrogen limitation, UTAse uridylylates and changes the ability of PII (PII-UMP) to interact with different targets [246]. The resulting phosphorylation activation of NtrC promotes GS transcriptional activation and accumulation [246]. Mutation of *glnB* in *R. leguminosarum* has been shown to produce a constitutively activated 'nitrogen limitation' response [246]. If null mutation to *glnB* prevents the production of a functional PII, UTAse will no longer be able to activate this PII in response to changing nitrogen conditions and it will lose its ability to repress the transcriptional initiation and accumulation of GS via this mechanism. *R. leguminosarum glnB* mutants have been shown to exhibit constitutive expression of the *glnII* and *glnBA* genes [246]. Mutants also exhibit higher levels of glutamine synthetase (GSII) activity when grown on ammonia, although levels of glutamine synthetase (GSI) remain unchanged [246, 247].

A constitutively activated glutamine synthesis response will require increased removal of 2-ketoglutarate from the TCA-cycle to combine with ammonium to ultimately form glutamine (Fig. 5-1). Replenishing 2-ketoglutarate will necessitate an additional turn of the TCA-cycle, which will generate reductant molecules requiring reoxidation via the electron transport chain (Fig. 5-1). This led to proposal of a reoxidation block model caused by limiting O_2: where the removal of 2-ketoglutarate from the TCA-cycle to assimilate ammonium into amino acids becomes problematic for the bacterium under low O_2. That a reoxidation block limits the removal of 2-ketoglutarate from the TCA-cycle would explain the lethal phenotype of a *glnB* insertion mutation under these low O_2 conditions [208]. Although it should be noted that other models are possible to explain this lethal phenotype, including glutamine toxicity, for example.

A key prediction of this reoxidation block model is that growth at low O_2 on succinate will be rescued by provision of glutamate rather than ammonium as a nitrogen source. The provision of glutamate should diminish the necessity to remove 2-ketoglutarate from the TCA-cycle for nitrogen assimilation. To validate INSeq results and test the reoxidation block model a strain mutated in *glnB* (OPS1031) was isolated in the Rlv3841 background. OPS1031 (*glnB*) was unable to grow using a dicarboxylic acid carbon source and ammonium at 1% O_2, but as predicted by the model, growth at low O_2 was rescued by the provision of glutamate (Table 5-1). Growth of OPS1031 at low O_2 was similarly rescued by the provision of glucose instead of succinate, suggesting growth on TCA-cycle intermediates is a key component of this reoxidation block model. Remarkably at 21% O_2, OPS1031 actually grew faster than the wild-type Rlv3841 on a dicarboxylic acid carbon source and ammonium (Mean Generation Time (MGT) 3.0h ±0.05 vs. 4.0h ±0.1)) (Table 5-1). This suggests that when O_2 is not limited, glutamine synthesis is a growth rate limiting step for rhizobia [246, 247]. Consistent with this, the MGT at 21% O_2 for wild-type cells utilizing succinate was fastest when glutamine was provided as a nitrogen source instead of ammonium or glutamate (3.5h ±0.2 on glutamine vs. 4.0h ±0.1 on ammonium and 4.5h ±0.1 on glutamate) (Table 5-1).

Strain	Carbon source	Nitrogen source	O_2 concentration (%)	MGT (h) (± S.E.M.)
Rlv3841	Succinate	Ammonium	21	4.0 ± 0.1
	Succinate	Glutamate	21	4.5 ± 0.1
	Succinate	Glutamine	21	3.5 ± 0.2
OSP1031	Succinate	Ammonium	21	3.0 ± 0.05
			1	N/G
	Succinate	Glutamate	21	4.6 ± 0.2
			1	4.9 ± 0.2
	Glucose	Ammonium	21	3.0 ± 0.04
			1	3.8 ± 0.3

Table 5-1. The Mean Generation Time (MGT) of strains grown on varied carbon and nitrogen sources at 21% O_2 or 1% O_2 calculated from three independent biological replicates (± S.E.M; standard error of the mean). MGT was calculated as the mean doubling time for cells in exponential growth phase as taken from OD_{600} measurements. Carbon sources were provided as succinate (20 mM) or glucose (10 mM), nitrogen sources were provided as ammonium (10 mM), glutamate (10 mM) or glutamine (10 mM). N/G = No Growth.

5.3.2 Growth on a dicarboxylic acid and ammonium under low O_2 induces glutamate auxotrophy and retards growth

Ammonium assimilation relies on the removal of 2-ketoglutarate from the TCA-cycle to form glutamate (Fig. 5-1). If a reoxidation block limits the removal of 2-ketoglutarate from the TCA-cycle under low O_2, it was hypothesised that decreasing O_2 would hinder the growth rate of Rlv3841 assimilating ammonium. This hypothesis was proven to be

correct: the growth rate of Rlv3841 was most drastically decreased under low O_2 when ammonium was utilised in combination with succinate (MGT 4h ±0.1 at 21% O_2 vs. 6.6h ±0.4 at 1% O_2 vs. 12.2h ±0.6 at 0.5% O_2) (Fig. 5-2, Table 5-2). As predicted by the reoxidation model, low O_2 growth was rescued by the provision of glutamate (Fig. 5-2, Table 5-2). Glutamate restores growth at low O_2 both when provided as an alternative to ammonium and when provided in combination (Fig. 5-2, Table 5-2). The provision of succinate in combination with ammonium also appears to be key: these growth defects were not observed on glucose (Fig. 5-2, Table 5-2). The improvements to growth are possibly due to glutamate providing a source of carbon, although cells grew very poorly or not at all on glutamate as a carbon substrate (Table 5-2). Overall these growth rate experiments suggest that ammonium assimilation becomes problematic under low O_2 in combination with succinate, and that this growth retardation can be rescued by the provision of glutamate (Fig. 5-2, Table 5-2).

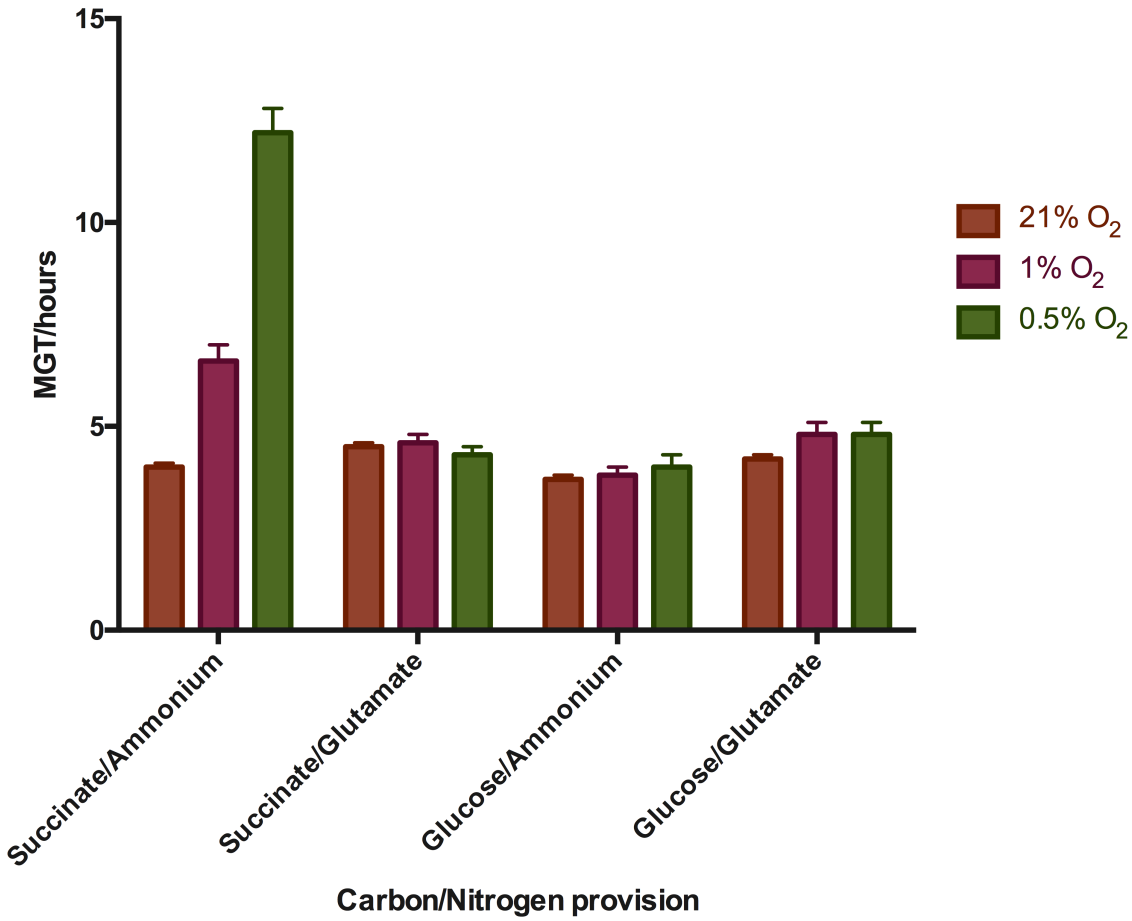


Figure 5-2. Mean generation time (MGT) of Rlv3841 cells utilising succinate or glucose and ammonium or glutamate at 21% O_2, 1% O_2 and 0.5% O_2 calculated from three independent biological replicates (error bars show standard error of the mean).

Carbon source	Nitrogen source	O_2 concentration (%)	MGT (h) (± S.E.M.)
Succinate	Ammonium	21	4.0 ± 0.1
		1	6.6 ± 0.4
		0.5	12.2 ± 0.6
Succinate	Glutamate	21	4.5 ± 0.1
		1	4.6 ± 0.2
		0.5	4.3 ± 0.2
Glucose	Ammonium	21	3.7 ± 0.1
		1	3.8 ± 0.2
		0.5	4 ± 0.3
Glucose	Glutamate	21	4.2 ± 0.1
		1	4.8 ± 0.3
		0.5	4.8 ± 0.3
Succinate	Ammonium and Glutamate	21	3.0 ± 0.02
		1	3.2 ± 0.1
Glutamate	Glutamate	21	N/G

Table 5-2. The Mean Generation Time (MGT) of Rlv3841 grown on varied carbon and nitrogen sources at 21% O_2, 1% O_2 and 0.5% O_2 calculated from three independent biological replicates (± S.E.M; standard error of the mean). MGT was calculated as the mean doubling time for cells in exponential growth phase as taken from OD_{600} measurements. Carbon sources were provided as succinate (20 mM) or glucose (10 mM), nitrogen sources were provided as ammonium (10 mM) or glutamate (10 mM). N/G = No Growth.

5.3.3 Ammonium assimilation and dicarboxylate metabolism increases cellular reductant ratios and O_2 consumption rate

The rationale behind the reoxidation block model is that ammonium assimilation becomes problematic under low O_2 in combination with dicarboxylic acids because both these pathways of dicarboxylate metabolism and ammonium assimilation require a significant generation of reductant molecules that need O_2 for reoxidation. While the generation of glutamate via GOGAT consumes one molecule of NADPH, overall ammonia assimilation will increase flux through a cycle that produces much more than one NADH/NADPH/$FADH_2$ reductant molecule. Therefore, it was hypothesised that growth on a dicarboxylate and ammonium would generate a higher cellular reductant ratio ($NADH:NAD^+$) and require a higher O_2 consumption rate than growth on a non-TCA-cycle intermediate and glutamate. This hypothesis was proven to be correct.

Cellular $NADH:NAD^+$ ratio was determined using the NAD/NADH-Glo™ Assay kit (Promega), optimised for use in Rlv3841 through a standard curve to establish the linear range of cells and an incubation period (chapter 2; materials and methods). The $NADH:NAD^+$ ratio was measured in early exponential growth phase. Mean O_2 consumption (μmol L^{-1} min^{-1} OD^{-1}) was calculated using an OxySense® 325i oxygen analyser system to measure the partial pressure of O_2 depletion within sealed cultures of Rlv3841 provided with variable nitrogen and carbon sources (chapter 2; materials and methods). Consumption rate was determined between 18% and 15% O_2.

Cells utilising a TCA-cycle intermediate (succinate) had a significantly higher reductant ratio ($NADH:NAD^+$) than cells utilising a non TCA-cycle intermediate (glucose) (Fig. 5-3, Table 5-3). Succinate-grown cells utilising ammonium as a nitrogen source had a

significantly higher reductant ratio than succinate-grown cells utilising glutamate as a nitrogen source (Table 5-3). Cellular reductant ratio correlated with O_2 consumption rate (Fig. 5-3), cells grown on succinate consumed O_2 at approximately twice the rate as cells grown on glucose, and this correlated with their significantly higher NADH:NAD^+ ratio. That succinate grown cells consume O_2 at a higher rate than glucose grown cells has previously been shown by Glenn and Dilworth (1981) [251]. This correlation between O_2 consumption and reductant ratio on TCA-cycle vs. non-TCA-cycle intermediates suggests O_2 consumption can be used as a proxy to monitor metabolic activity of the TCA-cycle. The catabolism of TCA-cycle intermediates succinate and malate gave significantly higher cellular O_2 consumption rates than the catabolism of the sugar, glucose and non-TCA-cycle monocarboxylate pyruvate (Fig. 5-4). In fact, cellular O_2 consumption rates could discern the difference between growing on ammonium as opposed to growing on glutamate, as a higher O_2 consumption rate on ammonium grown cells compared to glutamate grown cells was observed across all carbon sources tested (Fig. 5-4).

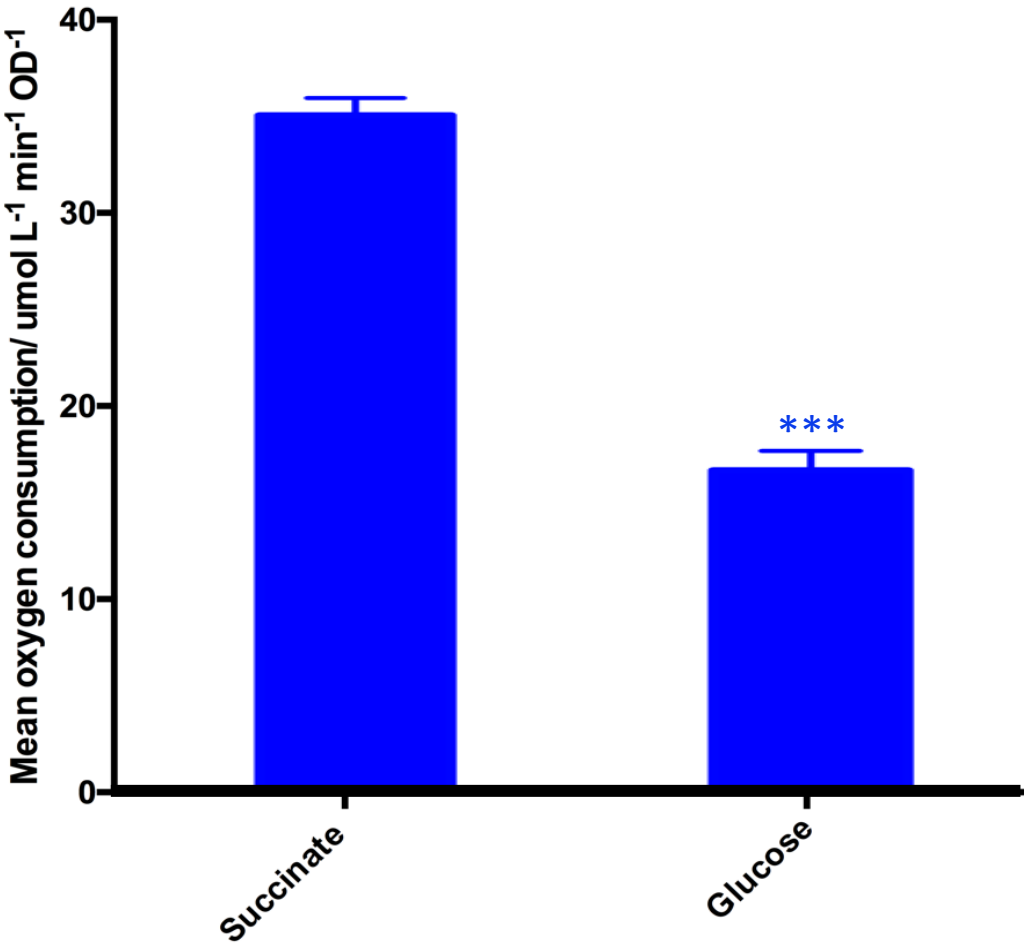


Figure 5-3. Mean O_2 consumption for growth on succinate vs. glucose calculated from five independent biological replicates, rate determined from respiration between 18% and 15% O_2. The difference between mean O_2 consumption is considered to be extremely statistically significant (unpaired t test, p value < 0.005). Error bars show standard error of the mean.

Carbon source	Nitrogen source	Ratio of NADH:NAD^+ (± S.E.M)
TCA-cycle intermediate (succinate)	Ammonium	1.17 ± 0.04
	Glutamate	1.06 ± 0.03
Non-TCA-cycle intermediate (glucose)	Ammonium	0.84 ± 0.05
	Glutamate	0.81 ± 0.04

Table 5-3. The ratio of NADH:NAD^+ in cells grown on succinate or glucose and ammonium or glutamate, calculated from four biological replicates (± S.E.M; standard error of the mean). Cultures for NADH:NAD^+ determination kept at atmospheric O_2. The difference between succinate and glucose is statistically significant (unpaired t test, p value < 0.005), the difference between ammonium and glutamate on succinate is statistically significant (unpaired t test, p value < 0.05), the difference between ammonium and glutamate on glucose is not considered to be statistically significant (unpaired t test).

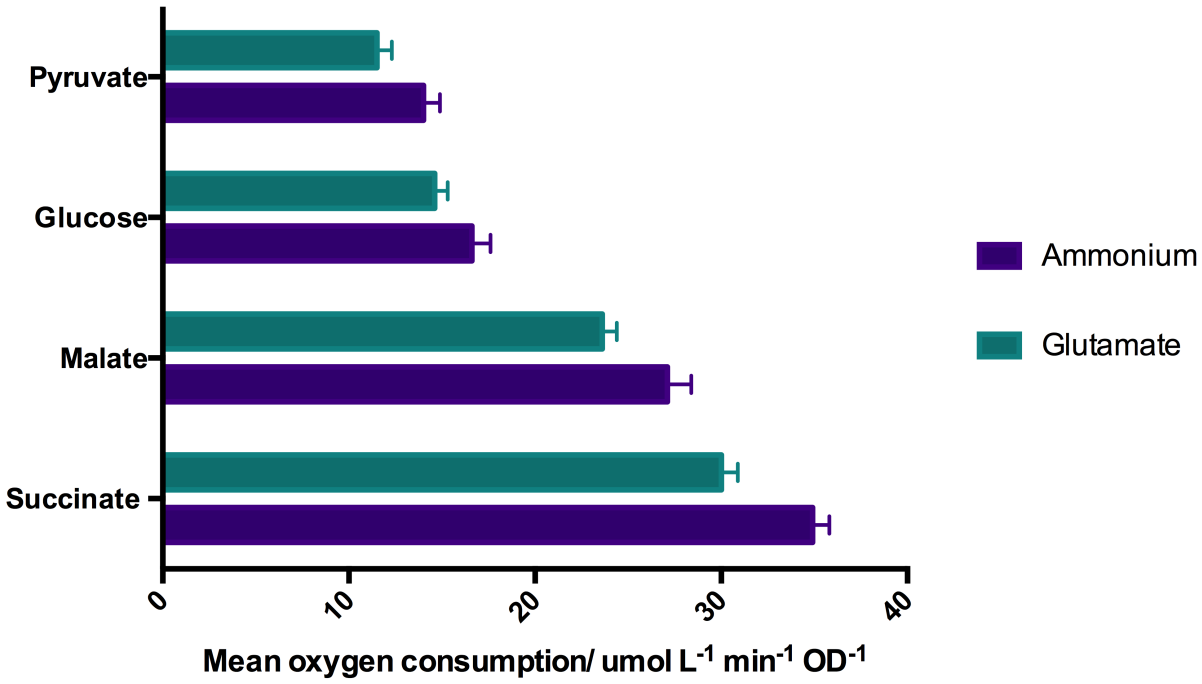


Figure 5-4. Mean O_2 consumption rate (µmol L^{-1} min^{-1} per unit OD) of Rlv3841 cells growing on different carbon sources (succinate, malate, pyruvate, glucose) and on ammonium or glutamate, calculated from six independent biological replicates (error bars show standard error of the mean). Rate determined from respiration between 18% and 15% O_2.

5.3.4 Low O_2 limits ammonium assimilation across rhizobial genera

This reoxidation block model encapsulates basic metabolic principles and holds significance for *Rhizobium*-legume symbioses on the whole. As a result, this hypothesis was tested in a range of other species. Metabolism can be highly variable between rhizobial genera [36, 252, 253], but decreasing O_2 caused severe growth retardations in cells utilising ammonium and a TCA-cycle intermediate across all species tested; *Sinorhizobium meliloti*, *Rhizobium etli*, *Rhizobium leguminosarum* and *Mesorhizobium loti* (Fig. 5-5). These growth defects were synonymously rescued by the provision of glutamate (Fig. 5-5).

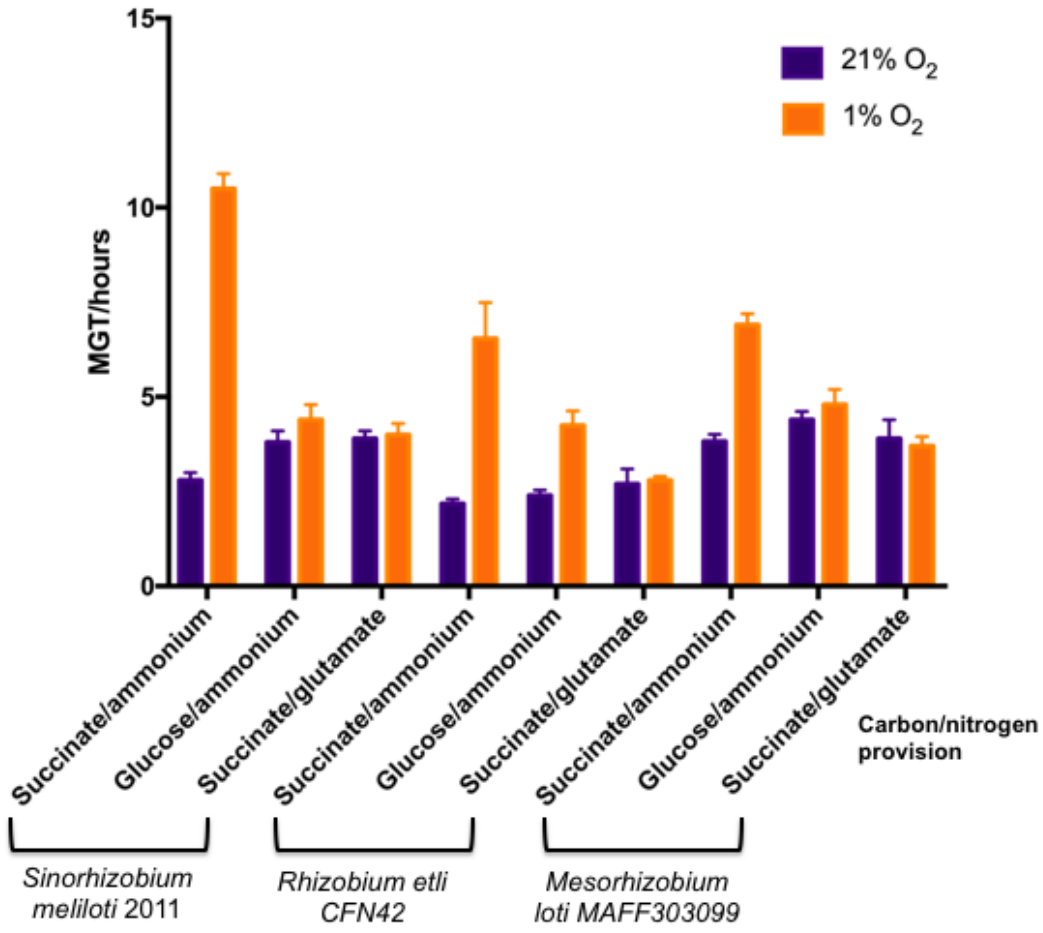


Figure 5-5. Mean generation time (MGT) of rhizobial cells utilising succinate or glucose and ammonium or glutamate at 21% O_2 and 1% O_2 calculated from three independent biological replicates (error bars show standard error of the mean).

5.4 Discussion

The fuelling of N_2-fixation in the low O_2 environment of legume nodules via the provision of dicarboxylates by the plant is widely accepted [36, 178, 254], however, it is surprising how unclear our understanding of the integration between plant and bacterial metabolism remains. A number of questions remain outstanding in the field: how is ammonium assimilation prevented in bacteroids? Why do legume hosts provide C_4-dicarboxylates that feed into the O_2-intensive TCA-cycle in the microaerobic

environment of nodules? Is there a reason why a less O_2-intensive energy supply has not been selected within symbioses? Here I propose a reoxidation block model that provides an answer to these three questions. The combination of low O_2 and the provision of dicarboxylates imposes a reoxidation block that helps limit ammonium assimilation in bacteroids. This reoxidation block would provide a novel driving force behind ammonium export in symbioses, alongside explaining why the provision of C_4-dicarboxylates to fuel nitrogen-fixation in a low O_2 environment is so conserved across nodule symbioses [29, 34, 255, 256]. It remains unclear exactly how ammonium is transported into the plant, but of utmost importance to the symbiosis is that assimilation does not occur in bacteroids. A large concentration gradient between the bacteroid and plant has been suggested to drive the passive diffusion of ammonium, and it has also been shown that ammonium may move through aquaporins, such as Nodulin 26 in soybean bacteroids [257], or specific ammonium channels [234].

A fundamental aspect of the reoxidation block model is that low O_2 limits ammonium assimilation in rhizobia because it relies on the removal of 2-ketoglutarate from the TCA-cycle to form glutamate. Replenishment of 2-ketoglutarate requires the generation of NADH, NADPH and $FADH_2$ (Fig. 5-1), whose reoxidation is limited by low O_2-availability. Mean O_2 consumption and NADH:NAD^+ ratio were proposed to function as a good proxy to monitor TCA-cycle activity (Fig. 5-1, Fig. 5-3, Fig. 5-4, Table 5-3); cells grown on a TCA-cycle intermediate (succinate) consumed O_2 at approximately twice the rate as cells grown on glucose (a sugar whose primary route of metabolism is the Entner-Doudoroff (ED) pathway [130, 172, 258]). The TCA-cycle is an O_2-intensive pathway whose activity generates large amounts of reductant (4 molecules reduced per

turn vs. 2 molecules in the ED pathway). This higher O_2 consumption rate correlated with a significantly higher NADH:NAD^+ ratio (Fig. 5-3).

Mean O_2 consumption rate and NADH:NAD^+ ratio was compared for cells utilising ammonium and cells utilising glutamate (Fig. 5-4, Table 5-3). The higher O_2 consumption rate and NADH:NAD^+ ratio required for cells utilising ammonium supported an increased requirement for TCA-cycle activity in these cells; ammonium assimilation requires removal of 2-ketoglutarate from the TCA-cycle, replenishment requires an additional cycle turn. The provision of glutamate, however, should diminish the necessity to remove 2-ketoglutarate from the TCA-cycle for nitrogen assimilation, and a lower O_2 consumption rate and NADH:NAD^+ ratio was demonstrated to be required to sustain growth (Fig. 5-4, Table 5-3). The provision of dicarboxylates in combination with ammonium is key, dicarboxylate metabolism exacerbates cellular O_2-requirements. This was demonstrated by the significantly higher O_2 consumption rates in cells grown on succinate or malate compared to cells grown on glucose or pyruvate (Fig. 5-4).

Following this reasoning, it is logical to predict that growth on the carbon and nitrogen sources with the highest O_2-requirements (i.e. succinate and ammonium) would be the most restricted by decreasing O_2-availability. This was observed to be the case, the MGT of cells utilising succinate and ammonium was severely retarded under low O_2 (Fig. 5-2), and growth defects were rescued by the provision of glutamate (Table 5-2). Metabolism can be highly variable between rhizobial genera [252, 253], but decreasing O_2 causes severe growth retardations in cells utilising ammonium and a TCA-cycle intermediate across *Sinorhizobium meliloti*, *Rhizobium etli*, *Rhizobium leguminosarum*

and *Mesorhizobium loti* (Fig. 5-5). These growth defects were all observable at 1% O_2 (Fig. 5-5), and it was demonstrated in *Rhizobium leguminosarum* that growth defects on succinate and ammonium were even more exacerbated at 0.5% O_2 (Fig. 5-2). Growth defects were rescued by the provision of glutamate. The combination of low O_2 and the metabolism of dicarboxylates is suggested to be essential for restricting ammonium assimilation (Fig. 5-2, Fig. 5-5). These growth defects were not observed for cells utilising glucose under low O_2, presumably because growth on glucose requires a lower O_2 consumption rate to be sustained (Fig. 5-3, Fig. 5-4) and so a reoxidation block is not imposed on the cell.

This reoxidation block model was proposed from observations in free-living *Rhizobium* cells. However, the model is supported by significant volumes of research from nodule bacteroids over a number of decades. That rhizobia become glutamate auxotrophs under low O_2 and that glutamate auxotrophy prevents the ammonium assimilation pathway (Fig. 5-1) is consistent with both metabolomic profiling studies in *Rhizobium leguminosarum* pea bacteroids and mutational studies to the GS/GOGAT ammonium assimilation pathway [207, 243]. Glutamate levels are 20-fold down in bacteroids (compared to free-living cells) [243], and GS/GOGAT activity is low and can be mutated without impairing nitrogen-fixation in mature bacteroids [207]. Metabolic flux analyses have been carried out in nodule bacteroids of *Azorhizobium caulinodans* [259], and show that nodule bacteroids severely restrict flux through the decarboxylating arm of the TCA-cycle. The TCA-cycle can be divided into two stages (Fig. 5-1): 1) the decarboxylating arm, which links to the ammonium assimilation pathway, and 2) the oxidative arm, which involves the successive oxidation reactions of the dicarboxylates succinate and malate that constitute the primary sources of carbon provided by the

legume. Evidently, there are perturbations to classical TCA-cycle functioning in nodule bacteroids [259]. This reduced flux through the decarboxylating arm supports that there are limitations in removing 2-ketoglutarate from the TCA-cycle in bacteroids.

That use of the TCA-cycle in low O_2 conditions is problematic due to the build-up of reductant molecules has been well-characterised in numerous species of anaerobic bacteria, where an incomplete TCA-cycle is used whose reactions aim at the synthesis of essential compounds and at the prevention of the build-up of reducing equivalents [238, 239]. An incomplete or rerouted TCA-cycle is thought to occur in *Bradyrhizobium japonicum* [156, 157], where enzymes encoding the decarboxylating arm (aconitase, isocitrate dehydrogenase and 2-ketoglutarate dehydrogenase) can all be inactivated by mutation and the mutants are still able to form effective N_2-fixing symbioses with host plant soybean [156, 157, 177, 237]. The aconitase, isocitrate dehydrogenase and 2-ketoglutarate dehydrogenase mutant strains were all glutamate auxotrophs and although the formation of soybean nodules was delayed the mutants could still fix nitrogen at a comparable rate to the wild-type [177]. This again supports the hypothesis that the decarboxylating arm of the TCA-cycle, the route that leads to ammonium assimilation, is not a necessity for N_2-fixation.

If bacteroids are under redox stress and the removal of 2-ketoglutarate from the TCA-cycle is limited, what other effects might you expect to see in the cell? Primarily, this would presumably affect the input of acetyl-CoA into the TCA-cycle. Apart from oxidation in the TCA-cycle, the other major metabolic fate of acetyl-CoA is lipogenesis [243]; most significantly, the production of fatty acids and the lipid-like polymer poly-3-hydroxybutyrate (PHB) [243]. A common observation in bacteroids is increased lipid

synthesis and the accumulation of PHB. PHB is observed to accumulate to significant levels in bacteroids of *Rhizobium leguminosarum* [243], *Rhizobium etli* [260], and *Bradyrhizobium japonicum* [261], amongst others. In pea bacteroids, a recent study showed levels of long-chain fatty acids were significantly increased compared to in free-living cells, alongside the considerable diversion of acetyl-CoA into PHB production [243]. The authors suggested this must indicate that the entry of acetyl-CoA into the TCA-cycle is being limited [243]. Lipid synthesis in free-living cells is similarly shown to undergo severe changes in response to low O_2 in *Bradyrhizobium* [262]. Lowering O_2 tensions resulted in significant increases in lipid accumulation and changes in lipid composition [262]. It is possible that a high affinity oxidise could overcome the problems of a low O_2 concentration in bacteroids, but this would require a sufficiently high flux in order to do so. That bacteroids are lipogenic, producing PHB and lipid, indicates that O_2 flux is likely insufficient for a high affinity oxidise to overcome. The provision of TCA-cycle dicarboxylates exacerbates cellular O_2-requirements and is a key component of the reoxidation block model. TCA-cycle dicarboxylates are conserved as the carbon source provision in all known nodule symbioses [36, 234, 255, 256, 263]. This is not only *Rhizobium*-legume symbioses, but probably also in symbioses between actinorhizal angiosperm plants and *Frankia* actinomycetes [255, 256].

An important question that remains is that: if a reoxidation block is preventing ammonium assimilation, how are bacteroids satisfying their nitrogen requirements? This is suggested to be by provision of nitrogenous amino acids from the legume, in a phenomenon known as amino acid cycling [18, 145, 148, 241, 264]. Amino acid cycling was first observed in pea nodules, where amino acid transport was found to be required for effective N_2-fixation [18, 32]. It was proposed that the provision of nitrogenous

amino acids to the bacteroid from the legume would allow the bacteroid to satisfy its nitrogen requirements without the need to assimilate fixed ammonium [18]. Here, I extend this to propose that bacteroids are actually unable to assimilate their own ammonium due to the combination of low O_2 and dicarboxylate metabolism imposing a reoxidation block on the removal of 2-ketoglutarate from the TCA-cycle. This provision of nitrogenous amino acids from the legume supplies the bacteroids with the nitrogen they need and allows the symbiosis to carry on functioning.

5.5 Revised model of symbiosis

A revised model of symbiosis can be proposed from the hypothesises of this reoxidation block model (Fig. 5-6). In this model, microaerobic nodule conditions mean there is a significantly reduced availability of O_2 to act as a final electron sink in the electron transport chain (1; Fig. 5-6), thus restricting the ability of the electron transport chain to reoxidise TCA-cycle generated reductant (2; Fig. 5-6). This is suggested to impose a reoxidation block on the removal of 2-ketoglutarate from the TCA-cycle to generate glutamate (3; Fig. 5-6), and glutamate levels have been reported to be 20-fold reduced in pea bacteroids [243]. Glutamate is required as a carbon backbone in the ammonium assimilation pathway; bacteroids are restricted in their ability to assimilate ammonium. This build-up of ammonium would help produce a large concentration gradient between the bacteroid and the plant, which is suggested to help drive the diffusion of ammonium out of the bacteroid cell [29, 149]. In lieu of ammonium assimilation, nitrogenous amino acids are provided to the bacteroid by the plant to satisfy their nitrogen requirements [18]. Dicarboxylates (i.e. malate and succinate) are provided to the bacteroid by the plant host to fuel nitrogen-fixation. Malate and succinate feed into the oxidative arm of the TCA-cycle, and there is significantly reduced flux through the decarboxylating arm

[259]. Reduced flux through the decarboxylating arm of the TCA-cycle is suggested to be due to this reoxidation block imposed by limited O_2-availability. This is a hypothetical model of symbiosis based on observations in free-living cells along with previous research and proposed models in nodule bacteroids [18, 29, 207, 229, 243, 259]. Further work is needed to fully validate this model.

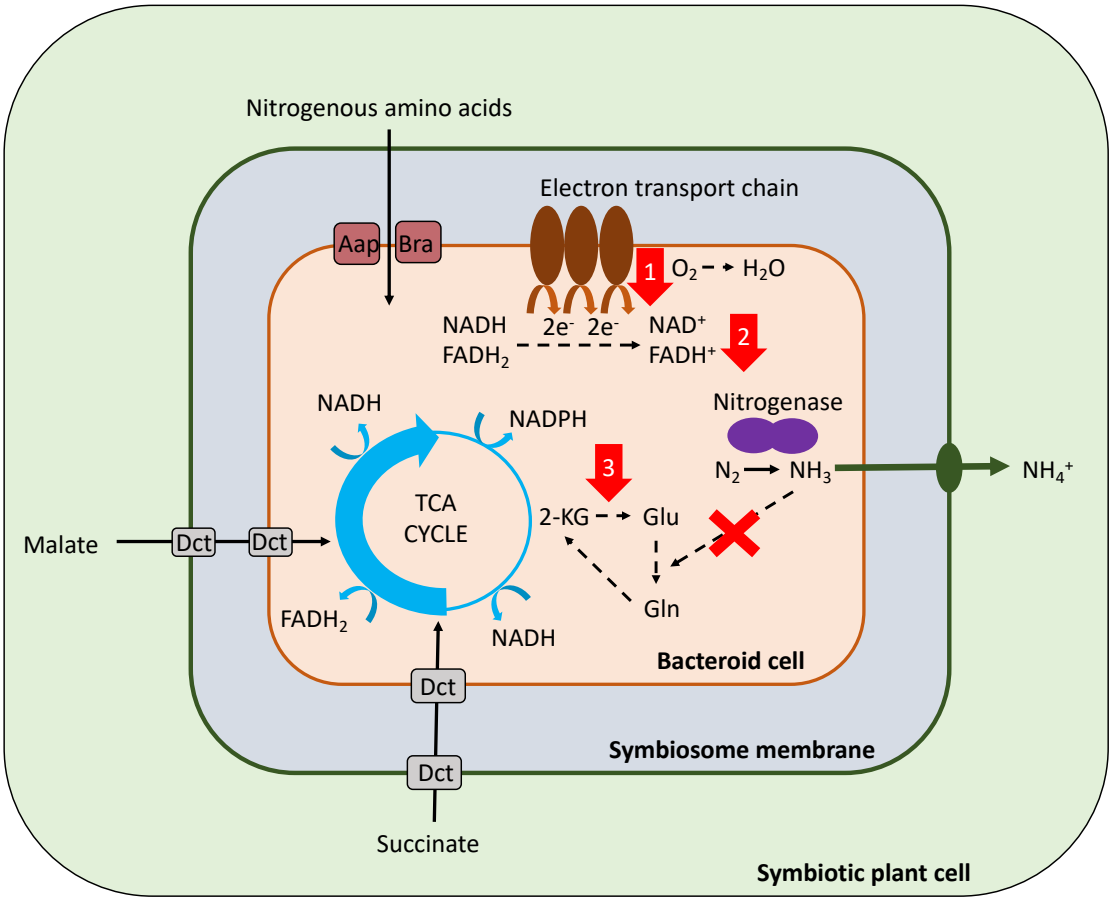


Figure 5-6. A revised model of symbiosis based on the proposed reoxidation block model. Malate and succinate are provided as carbon sources by the plant and are transported via the dicarboxylate transporter (Dct) system into the bacteroid cell [29], here they feed into the TCA-cycle. Reduced O_2-availability (↓1) limits the ability of the electron transport chain, located on the bacteroid membrane, to reoxidise reductant molecules generated by the TCA-cycle (↓2). This reoxidation block restricts the removal

of 2-ketoglutarate (2-KG) from the TCA-cycle to form glutamate (Glu) (↓3). Ammonium assimilation utilises glutamate as a carbon backbone for assimilation to generate glutamine (Gln), and as a result, there is a reoxidation block on the assimilation of ammonium. To satisfy the nitrogen requirements of the bacteroid, nitrogenous amino acids are provided by the plant and transported via the general amino acid permease (Aap) and the branched-chain amino acid permease (Bra) into the cell [18].

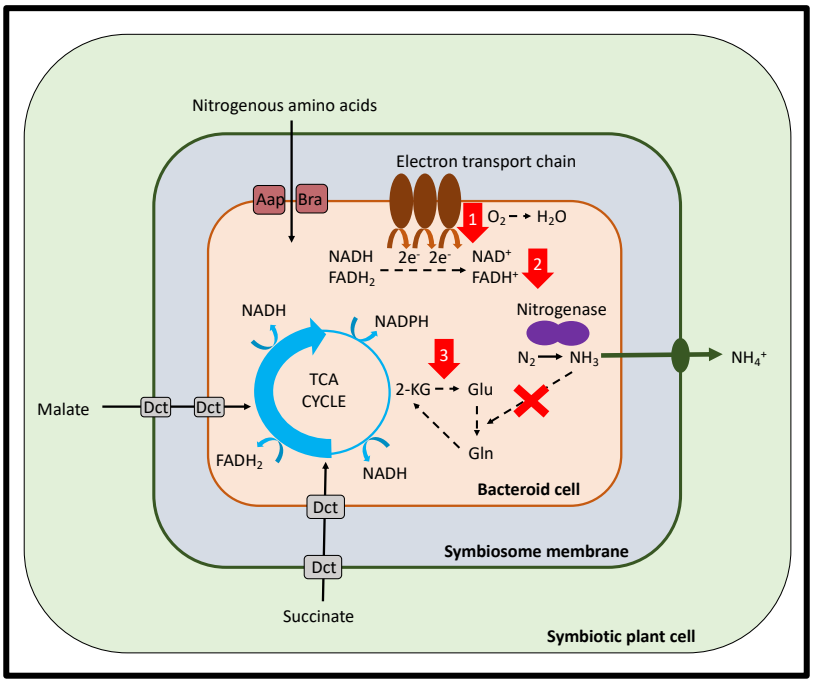
5.6 Conclusion

That ammonium assimilation becomes limiting for rhizobia under low O_2 has a critical implication for *Rhizobium*-legume symbioses. By fuelling N_2-fixation with the provision of dicarboxylates, legumes place an O_2 stranglehold on bacteroids that they precisely control by the synthesis of leghaemoglobin and thus prevent ammonium assimilation. This model suggests a novel driving-force behind ammonium export in symbioses, alongside explaining why legumes always provide TCA-cycle dicarboxylates to fuel N_2-fixation in the low O_2 environment of the nodules. The low O_2 nodule requirement in the prevention of ammonium assimilation might also explain why free-living N_2-fixing *Azorhizobium caulinodans* has a maximum rate of N_2-fixation at 3% O_2 [265], yet O_2 concentration buffered by leghaemoglobin in nodules is demonstrated to be lower than 0.00005% [254, 266-269]. While O_2 concentration cannot be directly correlated to flux, direct metabolic flux measurements confirm a very high rate of TCA-cycle flux in *Azorhizobium* at 3% O_2 compared to a much lower rate in bacteroids [259]. This is consistent with a much higher rate of O_2 flux in free-living *Azorhizobium* fixing N_2 at 3% O_2 than in N_2-fixing bacteroids in nodules.

There have been many discussions on the balance between *Rhizobium* and legume control in the evolution of symbioses [18, 270-273]. This model supports a more legume-dominated evolutionary view of the exchange, in which the imports and exports of the bacteroid are effectively driven by the legume host. However, despite this reoxidation block on ammonium assimilation, the prokaryotic partner is certainly provided with a desirable ‘working environment’; protective housing, the opportunity to multiply to high numbers and reliable sustenance in the form of a carbon source and nitrogenous amino acid supply.

Chapter 6

General discussion



6.1 Overview of results

The *Rhizobiaceae*-compatible *Mariner* transposon sequencing vector, pSAM_Rl [98], was published as a genome screening tool in 2014. This opened up the possibility of carrying out high-throughput *Mariner* transposon insertion sequencing (INSeq) screens to assess gene fitness on a whole-genome scale in *Rhizobium leguminosarum* bv. *viciae* 3841 (Rlv3841). The *Mariner* transposon (pSAM_Rl) was validated to generate highly saturated mutant libraries in Rlv3841, have no bias of insertion and enable gene fitness to be accurately predicted [98]. My aim in the work described in this thesis was to use this *Mariner* transposon INSeq methodology in Rlv3841 to characterise symbiosis-relevant fitness determinants.

In chapter 3, I describe the use of this technique in identifying the genetic requirements for growth at low O_2 compared to atmospheric O_2, and growth on a C_4-dicarboxylate (succinate) compared to a sugar (glucose). This allowed optimisation and validation of the newly published methodology with simple *in vitro* conditions. These conditions also held relevance to the environment within the nodules, where plant-provided C_4-dicarboxylates are used to fuel nitrogen-fixation under low O_2. Analysis of the carbon metabolism requirements under these conditions indicated that Rlv3841 appears to require use of a methylglyoxal pathway alongside the ED pathway and TCA-cycle for optimal growth on glucose. Surprisingly, the ED pathway was also required for optimal growth on succinate, suggesting sugars made by gluconeogenesis must undergo recycling. Within the central carbon metabolism pathways, the only O_2-dependent gene classifications were observed within the TCA-cycle. The TCA-cycle generates a large number of reductant molecules through its oxidation of carbon compounds.

In chapter 4, I describe the novel application of INSeq in an *in planta* environment to analyse the fitness determinants in early-stage symbioses; growth in the pea rhizosphere and attachment to pea roots. This was carried out in collaboration with an INSeq analysis of the fitness determinants for nodulation and recovery from the nodules (with Li Li and Vinoy Ramachandran). A total of 463 genes were identified to be required for the successful nodulation of pea; 176 genes are required from the rhizosphere through to nodulation ('rhizosphere-progressive' genes), 14 genes are required from root attachment to nodulation ('attachment-progressive' genes), 39 genes are required specifically for bacteroid development, 154 genes are required for both bacteroid development and recovery of nodule bacteria, and 80 genes are required specifically for recovery of nodule bacteria. In addition, 55 genes were required specifically for growth in the rhizosphere, mutation of which did not impair root attachment or nodulation, and 101 genes were required specifically for root attachment, mutation of which did not impair nodulation. Understanding the genetic requirements for successfully engaging in symbiosis will likely prove useful in working towards global research goals of improving the existing symbiosis and extending the symbiosis into non-legume crops. This study was able to identify not only genes essential for symbiosis (i.e. growth-essential genes), but also genes required for optimal fitness in engaging in symbiosis (i.e. growth-defective genes), and to identify which stage they are required in (chapter 4; fig. 4-6).

Analysis of the carbon metabolism pathways allowed characterisation of the carbon metabolism requirements for engaging in symbiosis with pea. This included identification of the requirement for the efficient metabolism of glucose-related sugars specifically in the rhizosphere and generation of the hypothesis that bacteroids might

synthesise glycogen via gluconeogenesis of TCA-cycle intermediates provided by the plant. I was able to develop an INSeq methodology suitable for *in planta* applications, where the quantity of contaminating plant material had initially caused problems in library preparation (detailed in chapter 2; materials and methods). In relation to the root attachment research, I have also been able to publish a review article, with the proposal that a common biphasic mechanism of root attachment exists in agriculturally important soil bacteria (Wheatley, Rachel M., and Philip S. Poole. "Mechanisms of bacterial attachment to roots." *FEMS Microbiology Reviews* 2018) (Appendix C) [68].

Across both INSeq chapters (chapter 3 and chapter 4), analyses were successful in characterising mutational phenotypes for 99.7% of genes in the genome of Rlv3841 across the nine different symbiosis-relevant *in vitro* and *in planta* conditions screened: an input library (selected on TY), growth on succinate at 21% O_2, growth on glucose at 21% O_2, growth on succinate at 1% O_2, growth on glucose at 1% O_2, growth in the pea rhizosphere, attachment to pea roots, pea nodulation (i.e. bacteroids) and recovery from pea nodules (i.e. nodule bacteria). INSeq gene fitness predictions were validated with directed mutational studies and comparison against previously-characterised mutational phenotypes in the literature. These datasets have been used to produce an INSeq database in Rlv3841. The *in vitro* dataset has been published [130], and the *in planta* dataset is currently in preparation for publishing. These INSeq analyses have helped improve the genome annotation of Rlv3841, and have been able to characterise growth-essential, growth-defective, or growth-advantaged mutational phenotypes for 578 previously uncharacterised genes encoding unknown hypothetical proteins in Rlv3841. Comparison across the INSeq database allowed identification of 395 genes that form the core fitness requirements in Rlv3841, mutation of which impairs growth regardless of

condition. A functional dissection of these core fitness requirements indicated the requirement for a number of protein secretion systems including the general secretion (Sec) pathway, the twin-arginine translocation (Tat) pathway, and the more poorly characterised Tol and Ton systems.

Analysis of the INSeq results identified that only one gene, RL2393 (*glnB*), was uniquely essential for growth under 1% O_2 with succinate. RL2393 (*glnB*) encodes the nitrogen regulatory protein (PII) that is a component of the bacterial nitrogen regulation (Ntr) system and an essential regulator of ammonium assimilation [208]. Mutation of *glnB* effectively generates a constitutively activated 'nitrogen-limitation' response via the Ntr system in the cell [246, 247]. This will promote increased removal of 2-ketoglutarate from the TCA-cycle to combine with ammonium to ultimately form glutamine in response to this nitrogen-limitation signalling. That the removal of 2-ketoglutarate from the TCA-cycle becomes problematic under low O_2 has significant implications for ammonium assimilation. Ammonium assimilation similarly relies on the removal of 2-ketoglutarate from the TCA-cycle to form glutamate as a carbon backbone. To replenish the 2-ketoglutarate removed from the TCA-cycle, an additional turn of the cycle will be required with the generation of NADH, NADPH and $FADH_2$, reductant molecules that require reoxidisation. This led to generation of a hypothesis: low O_2 limits ammonium assimilation in rhizobia. In chapter 5, I describe the investigation of this hypothesis, and the implications it has for symbioses. I conclude the chapter with proposing a revised model of symbiosis based on the reoxidation block model (chapter 5; Fig. 5-6).

The work in this thesis was carried out in free-living cells, but a review of the literature was able to provide an abundance of bacteroid-specific data. For example, glutamate levels are demonstrated to be 20-fold down in pea bacteroids compared to free-living cells [243], consistent with a reoxidation block preventing removal of 2-ketoglutarate from the TCA-cycle to generate glutamate. This work is being combined for publication with the metabolic flux analyses of Khushboo Borah [259], that show nodule bacteroids of *Azorhizobium caulinodans* severely restrict flux through the decarboxylating arm of the TCA-cycle [259]. This is consistent with mutational studies in *Brazyrhizobium japonicum,* that demonstrate the decarboxylating arm of the TCA-cycle is not required for effective nitrogen-fixation in nodules [157, 177, 237]. Perhaps most significantly, this reoxidation block model might also explain why free-living N_2-fixing *Azorhizobium caulinodans* has a maximum rate of N_2-fixation at 3% O_2 [265], yet O_2 concentration buffered by leghaemoglobin in nodules has been demonstrated to be lower than 0.00005% [254, 266-269]. While O_2 concentration cannot be directly correlated to flux, direct metabolic flux measurements confirm a very high rate of TCA-cycle flux in *Azorhizobium* at 3% O_2 compared to a much lower rate in bacteroids [259].This is consistent with a much higher rate of O_2 flux in free-living *Azorhizobium* fixing N_2 at 3% O_2 than in N_2-fixing bacteroids in nodules.

6.2 Evaluation of INSeq as a technique for whole-genome screening applications

INSeq is a method of transposon mutagenesis coupled with next-generation sequencing that has been successfully used in a range of genome screening applications across a diverse range of microbial species, including to characterise gene requirements in *Porphyromonas gingivalis* [118], antibiotic resistance in *Pseudomonas aeruginosa* [119] and cholesterol utilisation in *Mycobacterium tuberculosis* [120]. Transposon

mutagenesis achieves gene disruption and inactivation in single locations across the genome, and allows gene function to be investigated directly at the DNA-level. Mutant libraries are subjected to a period of negative selection, in which changes to the frequency of mutants within the library will occur based on the contribution to fitness of the mutated gene to growth under those conditions. The frequencies of mutants within the population can then be quantified by next-generation sequencing.

6.2.1 The Hidden Markov Model as a method of data analysis

A variety of different statistical models are available for INSeq data analysis and offer different ways to categorise or quantify the effects of gene disruption. For the analysis in this project, a four-state Hidden Markov Model was used [127]. The HMM allows the identification of regions which are significantly over- or under-represented in the library, and to these ends, can classify 'growth-defective' and 'growth-advantaged' mutations to genes in addition to the traditional 'essential' or 'non-essential' binary identifications. HMM analysis of sequencing data has been successfully applied and validated in a number of other INSeq analyses, including in *Vibrio cholerae* [128], *Mycobacterium tuberculosis* [129], *Haemophilus influenzae* [127], and also the initial INSeq analysis of *Rhizobium leguminosarum* [98]. The parameters of the HMM are dynamically adjusted to the attributes of the dataset in such a way that the model tends to remain within a state (despite a few sites that may not fit) until enough evidence accumulates to justify a transition to another state [274]. As such, the ability to detect essential regions in the genome is highly dependent on the quality of the transposon-insertion sequencing dataset. As with most statistical models, the HMM also relies on a number of assumptions, including that the abundance of mutants in the library reflects relative fitness, that transposon insertions occur randomly at candidate insertion sites (i.e. that

there is no preferential insertion bias into the genome), and that read-counts obtained from sequencing proportionally reflect abundance of mutants in the library [124].

6.2.2 *Mariner*-based transposon mutagenesis

If a gene analysed in a transposon mutagenesis screen is found to contain no transposon insertions, this could either be indicative of essentiality of the gene or because the gene did not accumulate transposon insertions in initial library construction. Use of a *Mariner*-based transposon vector conveys the advantage of having a defined insertion site at thymine-adenine 'TA' motifs, which allows prediction of how many possible insertion events there are along the length of a gene. Combined with the observed insertion density at previous sites, this allows prediction of a significantly significant stretch of empty insertion sites to indicate gene essentiality as opposed to absence of this insertion event in library construction. In effect, a defined insertion site allows transposition events to be modelled *in silico*. This aids robust statistical analysis and the prediction of gene essentiality without the need to sequence both the input and the output library as previously standard in random transposon mutagenesis screens.

Transposon mutagenesis with a defined insertion site is not without limitations. Use of the *Mariner* transposon means that only regions of the genome containing TA insertions can be mutated. This will limit the power of the technique in screening GC-rich regions, or organisms with a high GC-content genome, although this is not considered a limitation for screening in Rlv3841, where the mean number of 'TA' insertion sites per gene is 15 and a relatively consistent GC-content is maintained across the replicons. It should also be noted that the *Rhizobiaceae*-compatible *Mariner* transposon vector (pSAM_Rl) is only suitable for use in Rlv3841 because the strain does not contain

ampicillin (*AmpR*) and neomycin/kanamycin (*nptII*) resistance markers integrated in the genome. The *Mariner* transposon vector will not be suitable for use in strains that do, because this antibiotic resistance constitutes the method of selection for transposon mutants following transposon conjugation into the strain.

6.2.3 Evaluation of INSeq in this thesis

In this thesis I have been able to demonstrate that this *Mariner*-based INSeq methodology can be used to accurately predict gene fitness in Rlv3841 across a range of *in vitro* and *in planta* conditions. INSeq gene fitness predictions have been validated both with directed mutational studies and a comparison against previously characterised mutational phenotypes in the literature. INSeq was able to successfully identify previously characterised growth requirements for a certain environment, i.e. the genes encoding the dicarboxylate transporter (*dct*) proteins were required for growth on succinate but not on glucose, along with being able to identify many novel growth requirements in the environment too. A calculation of gene fitness value has been formulated (chapter 2, materials and methods), which builds on the resolution initially allowed by the four-state HMM. The gene fitness value was shown to correlate with the *in vivo* growth rate of the mutant (chapter 3), although further work is needed to validate the extent of this relationship (only five mutants were tested). Over the course of the project I was able to adapt the insertion library construction methodology to increase the saturation and quality of the libraries created through increasing the number of conjugations and size of the mutant selection screens. I was also able to develop a methodology suitable for *in planta* applications, which had to be adapted to solve the problems of contaminating plant DNA in the DNA extractions. This involved a series

of sequential centrifugal spin speeds, to help separate out bacterial material from plant, and a subsequent selection phase for the desired bacteria.

A limitation of the original INSeq methodology is that the presence of contaminating DNA interferes with the DNA extraction and downstream analysis. This poses problems for analyses involving plants or other bacterial strains. A newly published random bar code transposon-site sequencing (RB-TnSeq) methodology helps solve these problems and would be more suitable for studies involving plants or other strains [275]. RB-TnSeq creates a mutant library using random DNA barcodes incorporated into transposon cassettes such that each mutant can be identified by a specific barcode. The DNA barcodes associated with each insertion only need to be mapped once using TnSeq [275]. Mutant fitness profiling can subsequently be carried out in a very high-throughput manner by isolating barcodes via a PCR reaction, and sequencing the DNA barcodes to quantify mutant abundance [275]. This carries advantages over the original INSeq methodology in allowing the use of multiple strains and the ease of mutant library barcode isolation via PCR, and should perhaps be considered for future directions as an alternative to INSeq. However, while RB-TnSeq should enable more high-throughput mutant screening, the construction of a barcoded transposon library in *Rhizobium* may take some time in initial adaptation and optimisation.

6.2.4 INSeq experiment parameters

When planning an INSeq analysis, the inoculation density, the number of generations of growth, and the ability to recover mutant populations, all need to be carefully assessed to ensure sufficient complexity is retained post-selection for sound statistical analysis [213]. These parameters were set through a combination of mathematical predictions

and growth assays with wild-type Rlv3841 for the work carried out in this thesis. Size of the input population is important to ensure sufficient coverage of the mutant library; however, this needs to be balanced with scope for adequate growth and negative selection of the library, which is important for resolving fitness differences between mutants. For the first INSeq analysis (chapter 3), a large culture volume (500 mL) was used to ensure a sufficiently large input population could be inoculated and still undergo a substantial number of generations of growth.

For the *in planta* INSeq analysis (chapter 4), growth assays with varying inoculum sizes and days-post-inoculation mutant retrieval steps were carried out to set the parameters for input inoculum, the period of negative selection and to calculate the number of mutants that could then be recovered post-selection for analysis. These parameters also had to be balanced with time, labour and expense of the experiment. An input inoculum of 10^5 cells per plant was chosen, as this would allow an at least 500-fold coverage of the approximately 7300 genes in the genome while keeping the plant number below 100. The library retrieval time point of 5 dpi was chosen as this was the first point of maximal population saturation in the wild-type Rlv3841 colonisation assays. This should allow an adequate period of selection to resolve mutant fitness effects while also ensuring early-stage symbiosis is a focus and the population of mutants retrieved is maximised. The highly saturated mutant libraries sequenced from the rhizosphere (65% insertion density) and roots (58% insertion density) highlighted the suitability of the experiment parameters chosen. The input library had an insertion density of 80%, indicating a high-quality mutant library with good coverage of the genome had initially been constructed.

6.2.5 INSeq analyses of microbial competition

The *in planta* INSeq analysis (chapter 4) highlighted the power of INSeq in resolving competitive differences between strains. The INSeq experiments are effectively thousands of different 'strains' competing against one another for growth in a particular environment. The majority of these mutant strains will have 'wild-type'-like properties (i.e. 81% of the genes in the root attachment analysis had growth-neutral phenotypes upon mutation). These thousands of insertion mutants with wild-type-like attaching properties competing for root attachment will help resolve fitness differences in even marginally fitness reduced mutants. This will enable the identification of genes with a role in competitive root attachment that cannot be identified in single-strain attachment assays.

Understanding the genes important for competitive colonisation is very important for understanding how microbes can colonise the root in real field conditions, where there will be the presence of many other microbes competing for root proximity, carbon from root exudates and adherence. Competitive fitness is equally as important to possessing the optimal genetic machinery for attachment; if a strain cannot compete with others for proximity to the root, it will not even gain the opportunity to attach.

This power of INSeq in resolving competitive effects was observed for the '*rhi*' genes on pRL10. These genes had previously been suspected with a role in colonisation due to their significantly upregulated transcriptional response in the rhizosphere, but were unable to be characterised with a defect in attachment in single-strain attachment assays (personal communication, Downie). Their growth-defective phenotype identification upon mutation in root attachment indicates they do indeed have a role in fitness for root attachment, but that this role can only be resolved through competition with other strains.

Consideration of this use of INSeq in resolving fitness differences in competition highlights the potential of this technique in future research applications to better understand more field-relevant traits. The next step to build on this would be to include different microbe species, to characterise the differences in intra- and inter- species competition, and resolve further fitness differences in attachment.

6.2.6 Other uses of INSeq

The work described in this thesis has focused on using INSeq as a technique for whole-genome screening and the functional dissection of bacterial genes. This is not the only type of investigation INSeq can be used in. Another idea from this project was that INSeq could be used to identify novel metabolic pathways from the screening of metabolic mutants. For example, to identify where flux from the TCA-cycle is rerouted when a key enzyme such as isocitrate dehydrogenase is mutated. This could be achieved by carrying out an INSeq analysis of wild-type Rlv3841 and the TCA-cycle enzyme mutant growing on a TCA-cycle intermediate such as succinate. The results of this screen could be used to identify which genes in the mutant are now required for growth that were not required previously in the wild-type, and this would suggest a path of metabolic re-routing.

6.3 Limitations and future directions

6.3.1 Scope

In terms of the scope of this thesis, the experimental work carried out has used Rlv3841 as a model organism for dissecting rhizobial gene requirements. Like with the use of any model organism, the applicability of this work to other rhizobial genera will vary,

especially as phylogenetic distance increases. Rlv3841 was chosen as the strain for use in this thesis due to the volume of transcriptomic and physiological knowledge available in the Poole lab [36, 55, 57, 58], the availability of a fully sequenced genome [52], simple and well-defined growth conditions and its symbiosis with an agriculturally important legume (pea; *Pisum sativum*). INSeq is a genetics-based approach, and the genome of Rlv3841 can be split into two components: a 'core', most chromosomal component, which has a consistent phylogeny and is shared with related organisms, and accounts for approximately one third of the genome, and an 'accessory' component, which is more individual and located on the plasmids and chromosomal islands [56]. The work in this thesis has focused on characterising key biological processes and lifestyle adaptations. The fundamental aspects of these key biological processes are likely relatively well conserved across rhizobial genera, and so understanding the genes and molecular mechanisms behind these in Rlv3841 could be used to help understand how these key processes function across a broader range of species.

Metabolic analyses have been a major focus of the work described in this thesis. The ED pathway, TCA-cycle, gluconeogenesis and pentose phosphate pathways are widely used carbon metabolism pathways across *Rhizobium* [276, 277], and indeed across the majority of well-characterised bacterial species, excluding the ED pathway whose set of reactions are more commonly achieved via glycolysis [278]. The identification of a putative methylglyoxal pathway required for the optimal metabolism of glucose is possibly conserved across other rhizobial genera. The methylglyoxal synthase in Rlv3841 (RL0183) has a high shared amino acid identity with predicted methylglyoxal synthases across a broad range of strains, including in *Rhizobium etli* CFN42 (RHE_CH00173; 96% identity), *Sinorhizobium meliloti* 1021 (SMc02834; 83.5%

identity), *Azospirillum* sp. B510 (AZLd01460; 61.1% identity), *Azorhizobium caulinodans* ORS571 (AZC_4555; 51.7% identity) and *Klebsiella pneomoniae* Kp342 (GKPORF_B5423; 50.4% identity) [143].

Experimental work branching out from Rlv3841 was carried out in chapter 5, where the proposal that low O_2 prevents ammonium assimilation was investigated across *Rhizobium leguminosarum, Sinorhizobium meliloti*, *Rhizobium etli*, and *Mesorhizobium loti.* The reoxidation block model of ammonium assimilation holds significance for *Rhizobium*-legume symbioses on a global scale, and considering that actinorhizal angiosperms are also suggested to provide dicarboxylates to fuel nitrogen-fixation in actinomycetes [255, 256], possibility across other nodule symbioses as well. More work is needed to fully validate the scope of this model, and these limitations and future directions are discussed in more detail below.

6.3.2 INSeq analyses

Reflecting back on limitations of the INSeq experiments, using an O_2 tension lower than 1% in the first INSeq screen (chapter 3) could have helped identify a greater number of O_2-regulated gene requirements and have been more relevant to the microaerobic environment of the nodules (where O_2 concentration is demonstrated to be lower than 0.00005% [254, 266-269]). With the facilities that were available, 0.1% O_2 was the lowest O_2-tension that the microaerobic chamber could accurately be lowered to. An INSeq screen carried out at 0.1% O_2 with a dicarboxylate and amino acid(s) would perhaps better mimic the nodule environment in the future.

There were deviations to field conditions in the *in planta* INSeq analysis (chapter 4) in the use of sterilised vermiculite instead of soil, and the absence of other competing microbe species. It should be noted that the absence of other contaminating microbes was not confirmed experimentally, but that the vermiculite was deemed suitable for use following washing and sterilisation in an autoclave. Using non-sterile field soil would make the study more applicable to natural field conditions, but it would be more difficult to quantify and assess the variables present in the soil samples used, and require a much more complex library retrieval step to separate rhizobial insertion library DNA from contaminating plant and different microbe DNA present. As this was the first study of this nature, this early-stage symbiosis screen was carried out with controlled variables to allow the INSeq to initially be optimised in this *in planta* application. The fundamental components of a competitive colonisation screen were still applicable in this screen: insertion mutants still had to thrive and compete in the oxidative rhizosphere environment, and cells still had to migrate and attach to the root surface, while competing against one another to do so.

A future direction from this work would be to analyse growth in the rhizosphere and root attachment in more field applicable conditions, and investigate the impacts of inter-species competition with other microbes in the environment. RB-TnSeq might be a more suitable method for this type of analysis [275]. The research into root attachment could also have been expanded on by retrieving mutants from a larger selection of dpi time-points to better characterise the period of colonisation. It would have been insightful to investigate both acidic and alkali soil conditions around the roots. Attachment is known to be a multi-phase and pH-dependent process in *R. leguminosarum* [23, 68, 69], so this

would be useful in gaining a more global understanding of the mechanism of attachment and the influence of pH on this mechanism from a genetic perspective.

A major global aim in the field is to transfer N_2-fixing abilities to non-legume crops. To these ends, it would be useful to extend the INSeq characterisation of root attachment in Rlv3841 to investigate the genetic requirements for attaching to a non-host legume, such as soybean, and to a non-legume crop, such as barley. By including a non-host legume such as soybean in this analysis alongside a cereal crop, this would help distinguish between legume-specific and cereal-specific attachment factors, alongside improving understanding of the range of host-specificity. This INSeq analysis of root attachment across non-host legumes and cereals would be especially useful to carry out in the presence of other competing microbes. It is likely the case that in single strain inocula Rlv3841 is able to colonise a range of different roots to a reasonable extent, but understanding the fitness factors against more proficient colonisers of the root is especially important in understanding how colonisation capabilities could be extended across a range of hosts. With the inclusion of other strains of competing microbes, RB-TnSeq might be a more suitable method for this screening [275].

There were a number of additional directions I would have liked to explore that were unfortunately not within the scope of this thesis due to time restraints and prioritising research avenues. The INSeq database generated in this project contains a wealth of information and there were many gene groups within this dataset insightful downstream analyses could have been carried out on. A significant category of these is the growth-advantaged genes, one idea being to use this group to try and characterise a mutant in

Rlv3841 that demonstrates enhanced attachment or colonisation of the rhizosphere. The overall aim of this would be to characterise a mutant displaying enhanced symbiotic properties, i.e. enhanced nodulation, as this could potentially represent a strain that could deliver agricultural improvements in symbiosis, or at the very least, outline a potential pathway to obtain one. How successful following this avenue would be remains to be seen, natural selection tends to select for any generally advantageous traits in the population, so even if a strain demonstrating enhanced symbiotic properties could be isolated, it would likely be disadvantaged in some faculty. However, it might work well with *Rhizobium* inoculant strains where long-term survival in the soil is not necessarily an issue as strains can be reinoculated with each replanting of seed.

6.3.3 RNA sequencing analyses

As a follow on from the INSeq analysis, the INSeq-identified root attachment transcriptional regulators were chosen as candidates to map the transcriptional regulatory pathways required for root attachment. RNA sequencing of these regulator mutants was carried out to try and characterise the direct regulator targets. Unfortunately, the RNA sequencing datasets were not clear to analyse and achieving this end goal was not possible with the strategy chosen. Within the top 1% upregulated and downregulated gene groups, genes were generally in a steady gradient of expression changes, and single standout candidates could not be identified. It appeared a lot of changes in expression were a consequence of lower fitness in root attachment as opposed to a potential causation. Putative regulator targets were predicted on a basis of genomic proximity combined with INSeq data analysis, RNA sequencing data analysis and protein predictions (Chapter 4; Table 4-9, Table 4-10). However, these putative targets were not predicted with a great deal of confidence, as the RNA sequencing data was

affected by a significant amount of noise that made it difficult to identify convincing targets. This was presumably a result of a combination of effects, including the impacts of lower fitness in root attachment and halted growth, and the interconnection of regulatory networks and indirect regulator effects.

On reflection, there were a number of problems with this experimental strategy regardless. In terms of the regulators chosen, although they were identified with a growth-impaired phenotype (i.e. growth-defective or growth-essential) in root attachment when mutated, it is entirely possible that the genes have no direct role in attachment in the wild-type, but produce a phenotype that interferes with attachment when mutated. The RNA sequencing analyses were carried out on strains attaching in a single-strain inoculum, so did not replicate the INSeq attachment analysis in terms of a competitive environment with the presence of 'wild-type'-like attaching strains. An alternative strategy could have been to use chromatin immunoprecipitation (ChIP) sequencing [233]. ChIP sequencing can identify DNA sequences directly interacting with a target, and so would limit the noise generated by indirect targets and other indirect effects of regulator mutation.

6.3.4 Reoxidation block model

A major limitation of the reoxidation block model proposed in chapter 5 is that the experimental work was only carried out in free-living cells, and ammonium assimilation was not measured directly, but growth on ammonium was used as a proxy. Further work is needed to more fully investigate this model. For publication this work is being combined with metabolic flux mapping in *Azorhizobium caulinodans*, which

demonstrates that nodule bacteroids severely restrict flux through the decarboxylating arm of the TCA-cycle, consistent with a reoxidation block model [259].

In terms of any additional free-living experiments that could have been carried out, the work could have been strengthened by measuring NADPH/NADP and $FADH_2$/FADH ratios in addition to NADH/NAD^+ ratios for reductant quantification in the cell, expanding reductant ratio measurements into malate-grown cells, and investigating whether there is a preferential change for nitrogen-utilisation under decreasing O_2-tensions. Succinate was investigated as the major dicarboxylate in this work, and it would be useful to carry out growth rate and O_2 consumption assays on an expanded selection of dicarboxylates. Another experiment that would be useful to validate this model would be to construct a strain with higher glutamine synthetase (GSII) activity, and investigate the phenotype of this strain under low O_2 and in comparison to the *glnB* mutant.

Metabolic flux analysis could be very revealing if conducted on Rlv3841 growing at different O_2-tensions with succinate as a carbon source. This might reveal changes in the balance between the oxidative and decarboxylating arm of the TCA-cycle. A key question to answer is what happens to carbon balance if the decarboxylating arm is reduced as this implies a build-up of acetyl-CoA. Metabolic flux analysis should help determine if this is used for PHB or lipid biosynthesis. Looking to the future, perhaps expanding on this improved understanding of the relationship between dicarboxylate provision, low O_2 and ammonium assimilation could be manipulated for use in a nitrogen-fixation engineering application.

6.4 Concluding remarks

In this thesis I describe the use of INSeq as a high-throughput genome screening technique to characterise symbiosis-relevant fitness determinants in Rlv3841 across a range of *in vitro* and *in planta* conditions, achieving the aims described in the introductory chapter (chapter 1). This has enabled the identification of the gene requirements for growth at low O_2, growth on a C_4-dicarboxylate, growth in the pea rhizosphere and attachment to pea roots. I have been able to define 395 genes that form the core fitness requirements in Rlv3841 and improved genome annotations in the strain. Central metabolism pathways have formed a focal part of the analyses carried out in this thesis. I have been able to map how central metabolism pathway requirements change across a varied carbon source, varied O_2-tension and in engaging in symbiosis with pea. These analyses have allowed me to generate novel metabolic hypotheses.

Investigation of the hypothesis that low O_2 limits ammonium assimilation in rhizobia led to the proposal of a reoxidation block model, in which the provision of dicarboxylates under low O_2 is suggested to limit the removal of 2-ketoglutarate from the TCA-cycle and the assimilation of ammonium. Removal of 2-ketoglutarate from the TCA-cycle is suggested to become problematic under low O_2 because replenishing 2-ketoglutarate requires an additional turn of the TCA-cycle and the generation of NADH, NADPH and $FADH_2$, reductant molecules whose reoxidation is limited by O_2-availability. The provision of dicarboxylates is proposed to be key in exacerbating cellular O_2-requirements and imposing this reoxidation block in the cell. This reoxidation block model requires experimentation and validation in bacteroid cells.

Bibliography

1. Gerland, P., A.E. Raftery, H. Ševčíková, N. Li, D. Gu, T. Spoorenberg, L. Alkema, B.K. Fosdick, J. Chunn, and N. Lalic, *World population stabilization unlikely this century.* Science, 2014. **346**(6206): p. 234-237.
2. Wheeler, T. and J. Von Braun, *Climate change impacts on global food security.* Science, 2013. **341**(6145): p. 508-513.
3. Wiedmann, T.O., H. Schandl, M. Lenzen, D. Moran, S. Suh, J. West, and K. Kanemoto, *The material footprint of nations.* Proceedings of the National Academy of Sciences, 2013: p. 201220362.
4. Gruber, N. and J.N. Galloway, *An Earth-system perspective of the global nitrogen cycle.* Nature, 2008. **451**(7176): p. 293.
5. Hayatsu, M., K. Tago, and M. Saito, *Various players in the nitrogen cycle: diversity and functions of the microorganisms involved in nitrification and denitrification.* Soil Science and Plant Nutrition, 2008. **54**(1): p. 33-45.
6. Fowler, D., M. Coyle, U. Skiba, M.A. Sutton, J.N. Cape, S. Reis, L.J. Sheppard, A. Jenkins, B. Grizzetti, and J.N. Galloway, *The global nitrogen cycle in the twenty-first century.* Philosophical Transactions of the Royal Society, 2013. **368**(1621): p. 20130164.
7. Crutzen, P.J., *Geology of mankind.* Nature, 2002. **415**(6867): p. 23-23.
8. Steffen, W., P.J. Crutzen, and J.R. McNeill, *The Anthropocene: are humans now overwhelming the great forces of nature.* AMBIO: A Journal of the Human Environment, 2007. **36**(8): p. 614-621.
9. Smil, V., *Global energy: the latest infatuations.* American Scientist, 2011. **99**(3): p. 212-219.
10. Scheibel, M.G. and S. Schneider, *New Insights into the Biological and Synthetic Fixation of Nitrogen.* Angewandte Chemie International Edition, 2012. **51**(19): p. 4529-4531.
11. Erisman, J.W., M.A. Sutton, J. Galloway, Z. Klimont, and W. Winiwarter, *How a century of ammonia synthesis changed the world.* Nature Geoscience, 2008. **1**(10): p. 636.
12. Danso, S. and D. Eskew, *Enhancing biological nitrogen fixation.* IAEA Bulletin, 1984. **26**(2): p. 29-34.

13. Kinzig, A.P. and R.H. Socolow, *Human impacts on the nitrogen cycle.* Physics Today, 1994. **47**(11): p. 24-31.
14. Fields, S., *Global nitrogen: cycling out of control.* Environmental Health Perspectives, 2004. **112**(10): p. A556.
15. Galloway, J.N., A.R. Townsend, J.W. Erisman, M. Bekunda, Z. Cai, J.R. Freney, L.A. Martinelli, S.P. Seitzinger, and M.A. Sutton, *Transformation of the nitrogen cycle: recent trends, questions, and potential solutions.* Science, 2008. **320**(5878): p. 889-892.
16. Erisman, J.W., J.N. Galloway, S. Seitzinger, A. Bleeker, N.B. Dise, A.R. Petrescu, A.M. Leach, and W. de Vries, *Consequences of human modification of the global nitrogen cycle.* Philosophical Transactions of the Royal Society, 2013. **368**(1621): p. 20130116.
17. Rockström, J., W. Steffen, K. Noone, Å. Persson, F.S. Chapin, E.F. Lambin, T.M. Lenton, M. Scheffer, C. Folke, and H.J. Schellnhuber, *A safe operating space for humanity.* Nature, 2009. **461**(7263): p. 472-475.
18. Lodwig, E.M., Hosie, A. H., Bourdès, A., Findlay, K., Allaway, D., Karunakaran, R., Downie, J. A., and Poole, P. S. , *Amino-acid cycling drives nitrogen fixation in the legume–Rhizobium symbiosis.* Nature, 2003. **422**(6933): p. 722-726.
19. Udvardi, M. and P.S. Poole, *Transport and Metabolism in Legume-Rhizobia Symbioses*. Annual Review of Plant Biology, 2013. 64(1): p. 781-805.
20. Gan, Y., C. Hamel, J.T. O'Donovan, H. Cutforth, R.P. Zentner, C.A. Campbell, Y. Niu, and L. Poppy, *Diversifying crop rotations with pulses enhances system productivity.* Scientific Reports, 2015. **5**: p. 14625.
21. Stagnari, F., A. Maggio, A. Galieni, and M. Pisante, *Multiple benefits of legumes for agriculture sustainability: an overview.* Chemical and Biological Technologies in Agriculture, 2017. **4**(1): p. 2.
22. Jensen, E.S. and H. Hauggaard-Nielsen, *How can increased use of biological N 2 fixation in agriculture benefit the environment?* Plant and Soil, 2003. **252**(1): p. 177-186.
23. Downie, J.A., *The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots.* FEMS Microbiology Reviews, 2010. **34**(2): p. 150-170.

24. Gage, D.J., *Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes.* Microbiology and Molecular Biology Reviews, 2004. **68**(2): p. 280-300.
25. Oke, V. and S.R. Long, *Bacteroid formation in the Rhizobium–legume symbiosis.* Current Opinion in Microbiology, 1999. **2**(6): p. 641-646.
26. Hoffman, B.M., D. Lukoyanov, Z.-Y. Yang, D.R. Dean, and L.C. Seefeldt, *Mechanism of nitrogen fixation by nitrogenase: the next stage.* Chemical Reviews, 2014. **114**(8): p. 4041-4062.
27. Mus, F., M.B. Crook, K. Garcia, A.G. Costas, B.A. Geddes, E.D. Kouri, P. Paramasivan, M.H. Ryu, G.E.D. Oldroyd, P.S. Poole, M.K. Udvardi, C.A. Voigt, J.M. Ane, and J.W. Peters, *Symbiotic Nitrogen Fixation and the Challenges to Its Extension to Nonlegumes.* Applied and Environmental Microbiology, 2016. **82**(13): p. 3698-3710.
28. Hu, Y. and M.W. Ribbe, *Nitrogenase assembly.* Biochimica et Biophysica Acta (BBA)-Bioenergetics, 2013. **1827**(8-9): p. 1112-1122.
29. Udvardi, M., and Poole, P. S. , *Transport and metabolism in legume-rhizobia symbioses.* Annual Review of Plant Biology, 2013. **64**: p. 781-805.
30. Ott, T., J.T. van Dongen, C. Gu, L. Krusell, G. Desbrosses, H. Vigeolas, V. Bock, T. Czechowski, P. Geigenberger, and M.K. Udvardi, *Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development.* Current Biology, 2005. **15**(6): p. 531-535.
31. Fischer, H.-M., *Genetic regulation of nitrogen fixation in rhizobia.* Microbiological Reviews, 1994. **58**(3): p. 352-386.
32. White, J., Prell, J., James, E. K., and Poole, P., *Nutrient sharing between symbionts.* Plant Physiology, 2007. **144**(2): p. 604-614.
33. Glenn A.R., B.N.J., *Succinate resistant mutants of Rhizobium leguminosarum.* Microbiology, 1981. **126**(1): p. 237-241.
34. Finan, T., J. Wood, and D. Jordan, *Symbiotic properties of C4-dicarboxylic acid transport mutants of Rhizobium leguminosarum.* Journal of Bacteriology, 1983. **154**(3): p. 1403-1413.

35. Ronson C.W. and Downie J.A. , *Molecular cloning and genetic organization of C4-dicarboxylate transport genes from Rhizobium leguminosarum.* Journal of Bacteriology, 1984. **160**: p. 903–909.
36. Poole, P. and D. Allaway, *Carbon and nitrogen metabolism in Rhizobium.* Advances in Microbial Physiology, 2000. **43**: p. 117-163.
37. Niemietz, C.M. and S.D. Tyerman, *Channel-mediated permeation of ammonia gas through the peribacteroid membrane of soybean nodules.* FEBS letters, 2000. **465**(2-3): p. 110-114.
38. Clarke, V.C., P.C. Loughlin, D.A. Day, and P. Smith, *Transport processes of the legume symbiosome membrane.* Frontiers in Plant Science, 2014. **5**: p. 699.
39. Hardy, R. and U. Havelka, *Nitrogen fixation research: a key to world food?* Science, 1975. **188**(4188): p. 633-643.
40. Geddes, B.A., M.-H. Ryu, F. Mus, A.G. Costas, J.W. Peters, C.A. Voigt, and P. Poole, *Use of plant colonizing bacteria as chassis for transfer of N 2-fixation to cereals.* Current Opinion in Biotechnology, 2015. **32**: p. 216-222.
41. Vicente, E.J. and D.R. Dean, *Keeping the nitrogen-fixation dream alive.* Proceedings of the National Academy of Sciences, 2017. **114**(12): p. 3009-3011.
42. Rubio, L.M. and P.W. Ludden, *Biosynthesis of the iron-molybdenum cofactor of nitrogenase.* Annual Review Microbiology, 2008. **62**: p. 93-111.
43. Yang, J., X. Xie, X. Wang, R. Dixon, and Y.-P. Wang, *Reconstruction and minimal gene requirements for the alternative iron-only nitrogenase in Escherichia coli.* Proceedings of the National Academy of Sciences, 2014. **111**(35): p. E3718-E3725.
44. Yang, J., X. Xie, M. Yang, R. Dixon, and Y.-P. Wang, *Modular electron-transport chains from eukaryotic organelles function to support nitrogenase activity.* Proceedings of the National Academy of Sciences, 2017: p. 201620058.
45. Al-Falih, A.M., *Factors affecting the efficiency of symbiotic nitrogen fixation by Rhizobium.* Pakistan Journal of Biological Sciences, 2002. **5**(11): p. 1277-1293.

46. Beringer, J., T. Bisseling, and T. LaRue, *Improving symbiotic nitrogen fixation through the genetic manipulation of Rhizobium and legume host plants*, in *World crops: Cool season food legumes*. 1988, Springer. p. 691-702.
47. Wang, L., L. Wang, Y. Zhou, and D. Duanmu, *Use of CRISPR/Cas9 for Symbiotic Nitrogen Fixation Research in Legumes.* Progress in Molecular Biology and Translational Science, 2017.
48. Alves, L.M.C., J.A.M. de Souza, A. de Mello Varani, and E.G. de Macedo Lemos, *The family rhizobiaceae*, in *The Prokaryotes*. 2014, Springer. p. 419-437.
49. Weir, B.S. *The current taxonomy of rhizobia*. 2016 January, 2016 12 June, 2018].
50. Mazur, A., G. Stasiak, J. Wielbo, A. Kubik-Komar, M. Marek-Kozaczuk, and A. Skorupska, *Intragenomic diversity of Rhizobium leguminosarum bv. trifolii clover nodule isolates.* BMC Microbiology, 2011. **11**(1): p. 123.
51. Pérez Carrascal, O.M., D. VanInsberghe, S. Juárez, M.F. Polz, P. Vinuesa, and V. González, *Population genomics of the symbiotic plasmids of sympatric nitrogen-fixing Rhizobium species associated with Phaseolus vulgaris.* Environmental Microbiology, 2016. **18**(8): p. 2660-2676.
52. Young, J.P.W., Crossman, L. C., Johnston, A. W., Thomson, N. R., Ghazoui, Z. F., Hull, K. H., and Parkhill, J. , *The genome of Rhizobium leguminosarum has recognizable core and accessory components.* Genome Biology, 2006. **7**(4): p. 34.
53. González, V., J.L. Acosta, R.I. Santamaría, P. Bustos, J.L. Fernández, I.L.H. González, R. Díaz, M. Flores, R. Palacios, and J. Mora, *Conserved symbiotic plasmid DNA sequences in the multireplicon pangenomic structure of Rhizobium etli.* Applied and Environmental Microbiology, 2010. **76**(5): p. 1604-1614.
54. Johnston, A.W.B., and Beringer, J. E. , *Identification of the Rhizobium strains in pea root nodules using genetic markers.* Microbiology, 1975. **87**(2): p. 343-350.
55. Karunakaran, R., Ramachandran, V. K., Seaman, J. C., East, A. K., Mouhsine, B., Mauchline, T. H., Prell, J., Skeffington, A., and Poole, P. S. , *Transcriptomic analysis of Rhizobium leguminosarum biovar viciae in*

symbiosis with host plants Pisum sativum and Vicia cracca. Journal of Bacteriology, 2009. **191**(12): p. 4002-4014.

56. Kumar, N., G. Lad, E. Giuntini, M.E. Kaye, P. Udomwong, N.J. Shamsani, J.P.W. Young, and X. Bailly, *Bacterial genospecies that are not ecologically coherent: population genomics of Rhizobium leguminosarum.* Open Biology, 2015. **5**(1): p. 140133.
57. Ramachandran, V.K., East, A. K., Karunakaran, R., Downie, J. A., and Poole, P. S. , *Adaptation of Rhizobium leguminosarum to pea, alfalfa and sugar beet rhizospheres investigated by comparative transcriptomics.* Genome Biology, 2011. **12**(10): p. 106.
58. Prell, J., Bourdes, A., Karunakaran, R., Lopez-Gomez, M., and Poole, P. , *Pathway of γ-aminobutyrate metabolism in Rhizobium leguminosarum 3841 and its role in symbiosis.* Journal of Bacteriology, 2009. **191**(7): p. 2177-2186.
59. Ramírez-Romero, M.A., N. Soberón, A. Pérez-Oseguera, J. Téllez-Sosa, and M.A. Cevallos, *Structural elements required for replication and incompatibility of the Rhizobium etli symbiotic plasmid.* Journal of Bacteriology, 2000. **182**(11): p. 3117-3124.
60. Poole, P., V. Ramachandran, and J. Terpolilli, *Rhizobia: from saprophytes to endosymbionts.* Nature Reviews Microbiology, 2018.
61. Laranjo, M., A. Alexandre, and S. Oliveira, *Legume growth-promoting rhizobia: an overview on the Mesorhizobium genus.* Microbiological Research, 2014. **169**(1): p. 2-17.
62. Walker, T.S., Bais, H. P., Grotewold, E., and Vivanco, J. M. , *Root exudation and rhizosphere biology.* Plant Physiology, 2003. **132**(1): p. 44-51.
63. Estabrook, E.M., and Yoder, J. I. , *Plant-plant communications: rhizosphere signaling between parasitic angiosperms and their hosts.* Plant Physiology, 1998. **116**(1): p. 1-7.
64. Theunis, M., H. Kobayashi, W.J. Broughton, and E. Prinsen, *Flavonoids, NodD1, NodD2, and nod-box NB15 modulate expression of the y4wEFG locus that is required for indole-3-acetic acid synthesis in Rhizobium sp. strain NGR234.* Molecular Plant-Microbe Interactions, 2004. **17**(10): p. 1153-1161.

65. Oldroyd, G.E., J.D. Murray, P.S. Poole, and J.A. Downie, *The rules of engagement in the legume-rhizobial symbiosis.* Annual Review of Genetics, 2011. **45**: p. 119-144.
66. Bhattacharyya, C., U. Bakshi, I. Mallick, S. Mukherji, B. Bera, and A. Ghosh, *Genome-guided insights into the plant growth promotion capabilities of the physiologically versatile Bacillus aryabhattai strain AB211.* Frontiers in Microbiology, 2017. **8**: p. 411.
67. Smit, G., J. Kijne, and B. Lugtenberg, *Correlation between extracellular fibrils and attachment of Rhizobium leguminosarum to pea root hair tips.* Journal of Bacteriology, 1986. **168**(2): p. 821-827.
68. Wheatley, R.M. and P.S. Poole, *Mechanisms of bacterial attachment to roots.* FEMS Microbiology Reviews, 2018.
69. Laus, M.C., T.J. Logman, G.E. Lamers, A.A. Van Brussel, R.W. Carlson, and J.W. Kijne, *A novel polar surface polysaccharide from Rhizobium leguminosarum binds host plant lectin.* Molecular Microbiology, 2006. **59**(6): p. 1704-1713.
70. Sharon, N. and H. Lis, *History of lectins: from hemagglutinins to biological recognition molecules.* Glycobiology, 2004. **14**(11): p. 53R-62R.
71. Lagarda-Diaz, I., A.M. Guzman-Partida, and L. Vazquez-Moreno, *Legume Lectins: Proteins with Diverse Applications.* International Journal of Molecular Sciences, 2017. **18**(6): p. 1242.
72. Smit, G., J. Kijne, and B. Lugtenberg, *Involvement of both cellulose fibrils and a Ca2+-dependent adhesin in the attachment of Rhizobium leguminosarum to pea root hair tips.* Journal of Bacteriology, 1987. **169**(9): p. 4294-4301.
73. Smit, G., T. Logman, M. Boerrigter, J. Kijne, and B. Lugtenberg, *Purification and partial characterization of the Rhizobium leguminosarum biovar viciae Ca2+-dependent adhesin, which mediates the first step in attachment of cells of the family Rhizobiaceae to plant root hair tips.* Journal of Bacteriology, 1989. **171**(7): p. 4054-4062.
74. Matthyssee, A.G., *Attachment of Agrobacterium to plant surfaces.* Frontiers in Plant Science, 2014. **5**(252).
75. Poole, P., *Shining a light on the dark world of plant root–microbe interactions.* Proceedings of the National Academy of Sciences, 2017: p. 201703800.

76. Esseling, J.J., F.G. Lhuissier, and A.M.C. Emons, *Nod factor-induced root hair curling: continuous polar growth towards the point of nod factor application.* Plant Physiology, 2003. **132**(4): p. 1982-1988.
77. Fournier, J., A.C. Timmers, B.J. Sieberer, A. Jauneau, M. Chabaud, and D.G. Barker, *Mechanism of infection thread elongation in root hairs of Medicago truncatula and dynamic interplay with associated rhizobial colonization.* Plant Physiology, 2008. **148**(4): p. 1985-1995.
78. Brewin, N.J., *Plant cell wall remodelling in the Rhizobium–legume symbiosis.* Critical Reviews in Plant Sciences, 2004. **23**(4): p. 293-316.
79. Jones, K.M., N. Sharopova, D.P. Lohar, J.Q. Zhang, K.A. VandenBosch, and G.C. Walker, *Differential response of the plant Medicago truncatula to its symbiont Sinorhizobium meliloti or an exopolysaccharide-deficient mutant.* Proceedings of the National Academy of Sciences, 2008. **105**(2): p. 704-709.
80. Kawaharada, Y., S. Kelly, M.W. Nielsen, C.T. Hjuler, K. Gysel, A. Muszyński, R. Carlson, M.B. Thygesen, N. Sandal, and M. Asmussen, *Receptor-mediated exopolysaccharide perception controls bacterial infection.* Nature, 2015. **523**(7560): p. 308.
81. Marczak, M., A. Mazur, P. Koper, K. Żebracki, and A. Skorupska, *Synthesis of Rhizobial Exopolysaccharides and Their Importance for Symbiosis with Legume Plants.* Genes, 2017. **8**(12): p. 360.
82. Roponen, I., *The effect of darkness on the leghemoglobin content and amino acid levels in the root nodules of pea plants.* Physiologia Plantarum, 1970. **23**(3): p. 452-460.
83. Paau, A.S., C.B. Bloch, and W.J. Brill, *Developmental fate of Rhizobium meliloti bacteroids in alfalfa nodules.* Journal of Bacteriology, 1980. **143**(3): p. 1480-1490.
84. Müller, J., T. Boller, and A. Wiemken, *Trehalose becomes the most abundant non-structural carbohydrate during senescence of soybean nodules.* Journal of Experimental Botany, 2001. **52**(358): p. 943-947.
85. Williams, A., A. Wilkinson, M. Krehenbrink, D.M. Russo, A. Zorreguieta, and J.A. Downie, *Glucomannan-mediated attachment of Rhizobium leguminosarum to pea root hairs is required for competitive nodule infection.* Journal of Bacteriology, 2008. **190**(13): p. 4706-4715.

86. Kondorosi, E., F. Roudier, and E. Gendreau, *Plant cell-size control: growing by ploidy?* Current Opinion in Plant Biology, 2000. **3**(6): p. 488-492.
87. Maunoury, N., M. Redondo-Nieto, M. Bourcy, W. Van de Velde, B. Alunni, P. Laporte, P. Durand, N. Agier, L. Marisa, and D. Vaubert, *Differentiation of symbiotic cells and endosymbionts in Medicago truncatula nodulation are coupled to two transcriptome-switches.* PLoS One, 2010. **5**(3): p. e9519.
88. Van de Velde, W., G. Zehirov, A. Szatmari, M. Debreczeny, H. Ishihara, Z. Kevei, A. Farkas, K. Mikulass, A. Nagy, and H. Tiricz, *Plant peptides govern terminal differentiation of bacteria in symbiosis.* Science, 2010. **327**(5969): p. 1122-1126.
89. Montiel, J., J.A. Downie, A. Farkas, P. Bihari, R. Herczeg, B. Bálint, P. Mergaert, A. Kereszt, and É. Kondorosi, *Morphotype of bacteroids in different legumes correlates with the number and type of symbiotic NCR peptides.* Proceedings of the National Academy of Sciences, 2017: p. 201704217.
90. Czernic, P., D. Gully, F. Cartieaux, L. Moulin, I. Guefrachi, D. Patrel, O. Pierre, J. Fardoux, C. Chaintreuil, and P. Nguyen, *Convergent evolution of endosymbiont differentiation in dalbergioid and inverted repeat-lacking clade legumes mediated by nodule-specific cysteine-rich peptides.* Plant Physiology, 2015. **169**(2): p. 1254-1265.
91. Lardi, M. and G. Pessi, *Functional Genomics Approaches to Studying Symbioses between Legumes and Nitrogen-Fixing Rhizobia.* High Throughput, 2018. **7**(2): p. 15.
92. Gray, A.N., B.-M. Koo, A.L. Shiver, J.M. Peters, H. Osadnik, and C.A. Gross, *High-throughput bacterial functional genomics in the sequencing era.* Current Opinion in Microbiology, 2015. **27**: p. 86-95.
93. Van Opijnen, T. and A. Camilli, *Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms.* Nature Reviews Microbiology, 2013. **11**(7): p. 435.
94. Reid, C.J., and Poole, P. S. , *Roles of DctA and DctB in signal detection by the dicarboxylic acid transport system of Rhizobium leguminosarum.* Journal of Bacteriology, 1998. **180**(10): p. 2660-2669.

95. Finan, T.M., Oresnik, I. , and Bottacin, A. , *Mutants of Rhizobium meliloti defective in succinate metabolism.* Journal of Bacteriology, 1988. **170**(8): p. 3396-3403.
96. Engelke, T., M. Jagadish, and A. Pühler, *Biochemical and genetical analysis of Rhizobium meliloti mutants defective in C4-dicarboxylate transport.* Microbiology, 1987. **133**(11): p. 3019-3029.
97. Barquist, L., Boinett, C. J., and Cain, A. K., *Approaches to querying bacterial genomes with transposon-insertion sequencing.* RNA biology, 2013. **10**(7): p. 1161-1169.
98. Perry, B.J., and Yost, C. K. , *Construction of a mariner-based transposon vector for use in insertion sequence mutagenesis in selected members of the Rhizobiaceae.* BMC Microbiology, 2014. **14**(1): p. 298.
99. Zhao, R., L.X. Liu, Y.Z. Zhang, J. Jiao, W.J. Cui, B. Zhang, X.L. Wang, M.L. Li, Y. Chen, and Z.Q. Xiong, *Adaptive evolution of rhizobial symbiotic compatibility mediated by co-evolved insertion sequences.* The ISME journal, 2017. **12**(1): p. 101.
100. Beringer, J., J. Beynon, A. Buchanan-Wollaston, and A. Johnston, *Transfer of the drug-resistance transposon Tn5 to Rhizobium.* Nature, 1978. **276**(5688): p. 633.
101. Reeve, W.G., R.P. Tiwari, P.S. Worsley, M.J. Dilworth, A.R. Glenn, and J.G. Howieson, *Constructs for insertional mutagenesis, transcriptional signal localization and gene regulation studies in root nodule and other bacteria.* Microbiology, 1999. **145**(6): p. 1307-1316.
102. Wilson, K.J., *Molecular techniques for the study of rhizobial ecology in the field.* Soil Biology and Biochemistry, 1995. **27**(4-5): p. 501-514.
103. Schena, M., D. Shalon, R.W. Davis, and P.O. Brown, *Quantitative monitoring of gene expression patterns with a complementary DNA microarray.* Science, 1995. **270**(5235): p. 467-470.
104. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nature Reviews Genetics, 2009. **10**(1): p. 57.
105. Su, Z., P.P. Łabaj, S. Li, J. Thierry-Mieg, D. Thierry-Mieg, W. Shi, C. Wang, G.P. Schroth, R.A. Setterquist, and J.F. Thompson, *A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by*

the Sequencing Quality Control Consortium. Nature Biotechnology, 2014. **32**(9): p. 903.

106. Rachwał, K., E. Matczyńska, and M. Janczarek, *Transcriptome profiling of a Rhizobium leguminosarum bv. trifolii rosR mutant reveals the role of the transcriptional regulator RosR in motility, synthesis of cell-surface components, and other cellular processes.* BMC Genomics, 2015. **16**(1): p. 1111.
107. Graves, P.R. and T.A. Haystead, *Molecular biologist's guide to proteomics.* Microbiology and Molecular Biology Reviews, 2002. **66**(1): p. 39-63.
108. Natera, S.H., N. Guerreiro, and M.A. Djordjevic, *Proteome analysis of differentially displayed proteins as a tool for the investigation of symbiosis.* Molecular Plant-Microbe Interactions, 2000. **13**(9): p. 995-1009.
109. Djordjevic, M.A., *Sinorhizobium meliloti metabolism in the root nodule: a proteomic perspective.* Proteomics, 2004. **4**(7): p. 1859-1872.
110. Djordjevic, M.A., H.C. Chen, S. Natera, G. Van Noorden, C. Menzel, S. Taylor, C. Renard, O. Geiger, and G.F. Weiller, *A global analysis of protein expression profiles in Sinorhizobium meliloti: discovery of new genes for nodule occupancy and stress adaptation.* Molecular Plant-Microbe Interactions, 2003. **16**(6): p. 508-524.
111. Lardi, M., V. Murset, H.-M. Fischer, S. Mesa, C.H. Ahrens, N. Zamboni, and G. Pessi, *Metabolomic profiling of Bradyrhizobium diazoefficiens-induced root nodules reveals both host plant-specific and developmental signatures.* International Journal of Molecular Sciences, 2016. **17**(6): p. 815.
112. Barsch, A., V. Tellström, T. Patschkowski, H. Küster, and K. Niehaus, *Metabolite profiles of nodulated alfalfa plants indicate that distinct stages of nodule organogenesis are accompanied by global physiological adaptations.* Molecular Plant-Microbe Interactions, 2006. **19**(9): p. 998-1013.
113. Scalbert, A., L. Brennan, O. Fiehn, T. Hankemeier, B.S. Kristal, B. van Ommen, E. Pujos-Guillot, E. Verheij, D. Wishart, and S. Wopereis, *Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research.* Metabolomics, 2009. **5**(4): p. 435.

114. Van Opijnen, T., K.L. Bodi, and A. Camilli, *Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms.* Nature Methods, 2009. **6**(10): p. 767.

115. Goodman, A.L., N.P. McNulty, Y. Zhao, D. Leip, R.D. Mitra, C.A. Lozupone, R. Knight, and J.I. Gordon, *Identifying genetic determinants needed to establish a human gut symbiont in its habitat.* Cell Host & Microbe, 2009. **6**(3): p. 279-289.

116. Gawronski, J.D., S.M. Wong, G. Giannoukos, D.V. Ward, and B.J. Akerley, *Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung.* Proceedings of the National Academy of Sciences, 2009. **106**(38): p. 16422-16427.

117. Langridge, G.C., M.-D. Phan, D.J. Turner, T.T. Perkins, L. Parts, J. Haase, I. Charles, D.J. Maskell, S.E. Peters, and G. Dougan, *Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants.* Genome Research, 2009.

118. Klein, B.A., E.L. Tenorio, D.W. Lazinski, A. Camilli, M.J. Duncan, and L.T. Hu, *Identification of essential genes of the periodontal pathogen Porphyromonas gingivalis.* BMC Genomics, 2012. **13**(1): p. 578.

119. Gallagher, L.A., J. Shendure, and C. Manoil, *Genome-scale identification of resistance functions in Pseudomonas aeruginosa using Tn-seq.* MBio, 2011. **2**(1): p. e00315-10.

120. Griffin, J.E., J.D. Gawronski, M.A. DeJesus, T.R. Ioerger, B.J. Akerley, and C.M. Sassetti, *High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism.* PLoS Pathogens, 2011. **7**(9): p. e1002251.

121. Picardeau, M., *Transposition of fly mariner elements into bacteria as a genetic tool for mutagenesis.* Genetica, 2010. **138**(5): p. 551-558.

122. Plasterk, R.H., Izsvák, Z., and Ivics, Z. , *Resident aliens: the Tc1/mariner superfamily of transposable elements.* Trends in Genetics, 1999. **15**(8): p. 326-332.

123. Goodman, A.L., Wu, M., and Gordon, J. I., *Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries.* Nature protocols, 2011. **6**(12): p. 1969-1980.

124. DeJesus, M.A., S. Nambi, C.M. Smith, R.E. Baker, C.M. Sassetti, and T.R. Ioerger, *Statistical analysis of genetic interactions in Tn-Seq data.* Nucleic Acids Research, 2017. **45**(11): p. e93-e93.
125. Zomer, A., P. Burghout, H.J. Bootsma, P.W. Hermans, and S.A. van Hijum, *ESSENTIALS: software for rapid analysis of high throughput transposon insertion sequencing data.* PloS One, 2012. **7**(8): p. e43012.
126. DeJesus, M.A., Y.J. Zhang, C.M. Sassetti, E.J. Rubin, J.C. Sacchettini, and T.R. Ioerger, *Bayesian analysis of gene essentiality based on sequencing of transposon insertion libraries.* Bioinformatics, 2013. **29**(6): p. 695-703.
127. DeJesus, M.A., and Ioerger, T. R. , *A Hidden Markov Model for identifying essential and growth-defect regions in bacterial genomes from transposon insertion sequencing data.* BMC Bioinformatics, 2013. **14**(1): p. 303.
128. Chao, M.C., J.R. Pritchard, Y.J. Zhang, E.J. Rubin, J. Livny, B.M. Davis, and M.K. Waldor, *High-resolution definition of the Vibrio cholerae essential gene set with hidden Markov model–based analyses of transposon-insertion sequencing data.* Nucleic acids research, 2013. **41**(19): p. 9033-9048.
129. Long, J.E., M. DeJesus, D. Ward, R.E. Baker, T. Ioerger, and C.M. Sassetti, *Identifying essential genes in Mycobacterium tuberculosis by global phenotypic profiling*, in *Gene Essentiality*. 2015, Springer. p. 79-95.
130. Wheatley, R.M., V.K. Ramachandran, B.A. Geddes, B.J. Perry, C.K. Yost, and P.S. Poole, *Role of O2 in the Growth of Rhizobium leguminosarum bv. viciae 3841 on Glucose and Succinate.* Journal of Bacteriology, 2017. **199**(1): p. e00572-16.
131. Segovia, L., J.P.W. Young, and E. Martínez-Romero, *Reclassification of American Rhizobium leguminosarum biovar phaseoli type I strains as Rhizobium etli sp. nov.* International Journal of Systematic and Evolutionary Microbiology, 1993. **43**(2): p. 374-377.
132. Saeki, K. and H. Kouchi, *The Lotus symbiont, Mesorhizobium loti: molecular genetic techniques and application.* Journal of Plant Research, 2000. **113**(4): p. 457-465.
133. Casse, F., C. Boucher, J. Julliot, M. Michel, and J. Denarie, *Identification and characterization of large plasmids in Rhizobium meliloti using agarose gel electrophoresis.* Microbiology, 1979. **113**(2): p. 229-242.

134. Dowdle, S.F. and B.B. Bohlool, *Predominance of fast-growing Rhizobium japonicum in a soybean field in the People's Republic of China.* Applied and Environmental Microbiology, 1985. **50**(5): p. 1171-1176.
135. Schäfer, A., A. Tauch, W. Jäger, J. Kalinowski, G. Thierbach, and A. Pühler, *Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum.* Gene, 1994. **145**(1): p. 69-73.
136. Ditta, G., S. Stanfield, D. Corbin, and D.R. Helinski, *Broad host range DNA cloning system for gram-negative bacteria: construction of a gene bank of Rhizobium meliloti.* Proceedings of the National Academy of Sciences, 1980. **77**(12): p. 7347-7351.
137. Beringer, J.E., *R factor transfer in Rhizobium leguminosarum.* Microbiology, 1974. **84**(1): p. 188-198.
138. Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, and C. Duran, *Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.* Bioinformatics, 2012. **28**(12): p. 1647-1649.
139. Chomczynski, P. and M. Rymaszewski, *Alkaline polyethylene glycol-based method for direct PCR from bacteria, eukaryotic tissue samples, and whole blood.* Biotechniques, 2006. **40**(4): p. 454.
140. Martin, M., *Cutadapt removes adapter sequences from high-throughput sequencing reads.* EMBnet. Journal, 2011. **17**(1): p. 10.
141. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. , *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biology, 2009. **10**(3): p. 25.
142. Magoc, T., D. Wood, and S.L. Salzberg', *EDGE-pro: estimated degree of gene expression in prokaryotic genomes.* Evolutionary Bioinformatics, 2013. **9**: p. EBO. S11250.
143. Gish, W., and States, D. J. , *Identification of protein coding regions by database similarity search.* Nature Genetics, 1993. **3**(3): p. 266-272.
144. Szklarczyk, D., A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, and K.P. Tsafou, *STRING v10:*

protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Research, 2014. **43**(D1): p. D447-D452.

145. Prell, J., White, J. P., Bourdes, A., Bunnewell, S., Bongaerts, R. J., and Poole, P. S. , *Legumes regulate Rhizobium bacteroid development and persistence by the supply of branched-chain amino acids.* Proceedings of the National Academy of Sciences, 2009. **106**(30): p. 12477-12482.
146. Terpolilli, J.J., Hood, G. A., and Poole, P. S. , *What determines the efficiency of N2-fixing Rhizobium-legume symbioses?* Advances in Microbial Physiology, 2012. **60**: p. 326.
147. Oldroyd, G.E., and Downie, J. A. , *Calcium, kinases and nodulation signalling in legumes.* Nature Reviews Molecular Cell Biology, 2004. **5**(7): p. 566-576.
148. Day, D.A., Kaiser, B. N., Thomson, R., Udvardi, M. K., Moreau, S., and Puppo, A. , *Nutrient transport across symbiotic membranes from legume nodules.* Functional Plant Biology, 2001. **28**(7): p. 669-676.
149. Lodwig, E., and Poole, P. S., *Metabolism of Rhizobium bacteroids.* Critical Reviews in Plant Sciences, 2003. **22**: p. 37–78.
150. Mulley, G., Lopez-Gomez, M., Zhang, Y., Terpolilli, J., Prell, J., Finan, T., and Poole, P. , *Pyruvate is synthesized by two pathways in pea bacteroids with different efficiencies for nitrogen fixation.* Journal of Bacteriology, 2010. **192**(19): p. 4944-4953.
151. Zhang, Y., Aono, T., Poole, P., and Finan, T. M. , *NAD (P)+-malic enzyme mutants of Sinorhizobium sp. strain NGR234, but not Azorhizobium caulinodans ORS571, maintain symbiotic N2 fixation capabilities.* Applied and Environmental Microbiology, 2012. **78**(8): p. 2803-2812.
152. Walshaw, D.L., Wilkinson, A., Mundy, M., Smith, M., and Poole, P. S., *Regulation of the TCA cycle and the general amino acid permease by overflow metabolism in Rhizobium leguminosarum.* Microbiology, 1997. **143**(7): p. 2209-2221.
153. Dunn, M.F., *Tricarboxylic acid cycle and anaplerotic enzymes in rhizobia.* FEMS Microbiology Reviews, 1998. **22**(2): p. 105-123.
154. Driscoll, B.T., and Finan, T. M. , *NAD+-dependent malic enzyme of Rhizobium meliloti is required for symbiotic nitrogen fixation.* Molecular microbiology, 1993. **7**(6): p. 865-873.

155. Gardiol, A., Arias, A., Cervenansky, C., and Martinez-Drets, G., *Succinate dehydrogenase mutant of Rhizobium meliloti.* Journal of Bacteriology, 1982. **151**(3): p. 1621-1623.

156. Green, L.S., and Emerich, D. W. , *The formation of nitrogen-fixing bacteroids is delayed but not abolished in soybean infected by an [alpha]-ketoglutarate dehydrogenase-deficient mutant of Bradyrhizobium japonicum.* Plant Physiology, 1997. **114**(4): p. 1359-1368.

157. Green, L.S., and Emerich, D. W. , *Is Bacteroid α-Ketoglutarate Dehydrogenase Needed for Nitrogen Fixation?* In Highlights of Nitrogen Fixation Research, Springer US, 1999: p. 37-40.

158. Pei, B., Sisu, C., Frankish, A., Howald, C., Habegger, L., Mu, X.J., Harte, R., Balasubramanian, S., Tanzer, A., Diekhans, M. and Reymond, A., *The GENCODE pseudogene resource.* Genome Biology, 2012. **13**(9): p. 1.

159. Alberts, B., D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Essential cell biology*. 2013: Garland Science.

160. Alberts, B., A. Johnson, J. Lewis, P. Walter, M. Raff, and K. Roberts, *Molecular Biology of the Cell 4th Edition: International Student Edition*. 2002, Routledge.

161. Glover, B.P. and C.S. McHenry, *The DNA polymerase III holoenzyme: an asymmetric dimeric replicative complex with leading and lagging strand polymerases.* Cell, 2001. **105**(7): p. 925-934.

162. Desai, N.A. and V. Shankar, *Single-strand-specific nucleases.* FEMS Microbiology Reviews, 2003. **26**(5): p. 457-491.

163. Viswanathan, M. and S.T. Lovett, *Single-strand DNA-specific exonucleases in Escherichia coli: roles in repair and mutation avoidance.* Genetics, 1998. **149**(1): p. 7-16.

164. Mazur, A. and P. Koper, *Rhizobial plasmids—replication, structure and biological role.* Open Life Sciences, 2012. **7**(4): p. 571-586.

165. Pérez-Segura, G., Á. Pérez-Oseguera, and M.A. Cevallos, *The repAC replication system of the Rhizobium leguminosarum pRL7 plasmid is functional: implications regarding the origin and evolution of repABC plasmids.* Plasmid, 2013. **69**(1): p. 49-57.

166. Hood, G., *Physiological response of Rhizobium leguminosarum during bacteroid development*. 2013, University of East Anglia.
167. Yost, C.K., A.M. Rath, T.C. Noel, and M.F. Hynes, *Characterization of genes involved in erythritol catabolism in Rhizobium leguminosarum bv. viciae.* Microbiology, 2006. **152**(7): p. 2061-2074.
168. Hood, G., V. Ramachandran, A.K. East, J.A. Downie, and P.S. Poole, *Manganese transport is essential for N2-fixation by Rhizobium leguminosarum in bacteroids from galegoid but not phaseoloid nodules.* Environmental Microbiology, 2017. **19**(7): p. 2715-2726.
169. Walshaw, D.L. and P.S. Poole, *The general l-amino acid permease of Rhizobium leguminosarum is an ABC uptake system that also influences efflux of solutes.* Molecular Microbiology, 1996. **21**(6): p. 1239-1252.
170. Vanderlinde, E.M., J.J. Harrison, A. Muszyński, R.W. Carlson, R.J. Turner, and C.K. Yost, *Identification of a novel ABC transporter required for desiccation tolerance, and biofilm formation in Rhizobium leguminosarum bv. viciae 3841.* FEMS Microbiology Ecology, 2010. **71**(3): p. 327-340.
171. Geddes, B.A., and Oresnik, I. J., *Physiology, genetics, and biochemistry of carbon metabolism in the alphaproteobacterium Sinorhizobium meliloti.* Canadian Journal of Microbiology, 2014. **60**(8): p. 491-507.
172. Fuhrer, T., Fischer, E., and Sauer, U. , *Experimental identification and quantification of glucose metabolism in seven bacterial species.* Journal of Bacteriology, 2005. **187**(5): p. 1581-1590.
173. Portais, J.C., Tavernier, P., Gosselin, I., and Barbotin, J. N. , *Cyclic organization of the carbohydrate metabolism in Sinorhizobium meliloti.* European Journal of Biochemistry, 1999. **265**(1): p. 473-480.
174. McKay, I.A., Dilworth, M. J., and Glenn, A. R. , *Carbon catabolism in continuous cultures and bacteroids of Rhizobium leguminosarum MNF 3841.* Archives of Microbiology, 1989. **152**(606-610).
175. Koziol, U., Hannibal, L., Rodríguez, M. C., Fabiano, E., Kahn, M. L., and Noya, F. , *Deletion of citrate synthase restores growth of Sinorhizobium meliloti 1021 aconitase mutants.* Journal of Bacteriology, 2009. **191**(24): p. 7581-7586.

176. McDermott, T.R., and Kahn, M. L. , *Cloning and mutagenesis of the Rhizobium meliloti isocitrate dehydrogenase gene.* Journal of Bacteriology, 1992. **174**(14): p. 4790-4797.
177. Shah, R., and Emerich, D. W. , *Isocitrate dehydrogenase of Bradyrhizobium japonicum is not required for symbiotic nitrogen fixation with soybean.* Journal of Bacteriology, 2006. **188**(21): p. 7600-7608.
178. Lodiwg, E.M., *Regulation of Carbon Metabolism in Rhizobium leguminosarum*, in *School of Animal and Microbial Sciences*. 2001, University of Reading: United Kingdom.
179. Yost, C.K., Rath, A. M., Noel, T. C., and Hynes, M. F. , *Characterization of genes involved in erythritol catabolism in Rhizobium leguminosarum bv. viciae.* Microbiology, 2006. **152**(7): p. 2061-2074.
180. Geddes, B.A., Pickering, B. S., Poysti, N. J., Collins, H., Yudistira, H., and Oresnik, I. J., *A locus necessary for the transport and catabolism of erythritol in Sinorhizobium meliloti.* Microbiology, 2010. **156**(10): p. 2970-2981.
181. Geddes, B.A., and Oresnik, I. J. , *Genetic characterization of a complex locus necessary for the transport and catabolism of erythritol, adonitol and L-arabitol in Sinorhizobium meliloti.* Microbiology, 2012. **158**(8): p. 2180-2191.
182. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. , *KEGG as a reference resource for gene and protein annotation.* Nucleic acids Research,, 2016. **44**(1): p. 457-462.
183. Krol, E., and Becker, A. , *Global transcriptional analysis of the phosphate starvation response in Sinorhizobium meliloti strains 1021 and 2011.* Molecular Genetics and Genomics, 2004. **272**(1): p. 1-17.
184. Yuan, Z.C., Zaheer, R., Morton, R., and Finan, T. M., *Genome prediction of PhoB regulated promoters in Sinorhizobium meliloti and twelve proteobacteria.* Nucleic Acids Research, 2006. **34**(9): p. 2686-2697.
185. Inoue, Y., and Kimura, A. , *Glycolytic-methylglyoxal pathway.* Proceedings of the Japan Academy, Series B, 1999. **75**(6): p. 127-132.
186. Weber, J., Kayser, A., and Rinas, U. , *Metabolic flux analysis of Escherichia coli in glucose-limited continuous culture. II. Dynamic response to famine and feast, activation of the methylglyoxal pathway and oscillatory behaviour.* Microbiology, 2005. **151**(3): p. 707-716.

187. Bourdes, A., *Role of Aminotransferases during the symbiosis between Rhizobium leguminosarum and Pisum sativum.* 2008, University of Reading: United Kingdom.

188. Finnie, C., Hartley, N. M., Findlay, K. C., and Downie, J. A. , *The Rhizobium leguminosarum prsDE genes are required for secretion of several proteins, some of which influence nodulation, symbiotic nitrogen fixation and exopolysaccharide modification.* Molecular microbiology, 1997. **25**(1): p. 135-146.

189. Krehenbrink, M., and Downie, J. A. , *Identification of protein secretion systems and novel secreted proteins in Rhizobium leguminosarum bv. viciae.* BMC Genomics, 2008. **9**(1): p. 1.

190. Cremers, H.C., Batley, M., Redmond, J. W., Eydems, L., Breedveld, M. W., Zevehuizen, L. P., and Lugtenberg, B. J. , *Rhizobium leguminosarum exoB mutants are deficient in the synthesis of UDP-glucose 4'-epimerase.* Journal of Biological Chemistry, 1990. **265**(34): p. 21122-21127.

191. Janczarek, M., *Environmental signals and regulatory pathways that influence exopolysaccharide production in rhizobia.* International Journal of Molecular Sciences, 2011. **12**(11): p. 7898-7933.

192. Nentwich, S.S., Brinkrolf, K., Gaigalat, L., Hüser, A. T., Rey, D. A., Mohrbach, T., and Kalinowski, J., *Characterization of the LacI-type transcriptional repressor RbsR controlling ribose transport in Corynebacterium glutamicum ATCC 13032.* Microbiology, 2009. **155**(1): p. 150-164.

193. Batista, S., Patriarca, E. J., Tatè, R., Martínez-Drets, G., and Gill, P. R. , *An alternative succinate (2-oxoglutarate) transport system in Rhizobium tropici is induced in nodules of Phaseolus vulgaris.* Journal of Bacteriology, 2009. **191**(16): p. 5057-5067.

194. Stanger, A., *Analysis of the ECF sigma factors of rhizobium leguminosarum biovar viciae 3841.* 2008, University of East Anglia: United Kingdom.

195. Heimann, J.D., *The extracytoplasmic function (ECF) sigma factors.* Advances in Microbial Physiology, 2002. **46**: p. 47-110.

196. Noel, K.D., Diebold, R. J., Cava, J. R., and Brink, B. A. , *Rhizobial purine and pyrimidine auxotrophs: Nutrient supplementation, genetic analysis, and the*

symbiotic requirement for the novo purine biosynthesis. Archives of Microbiology, 1988. **149**(6): p. 499-506.

197. Lemieux, M.J., *A perspective on the structural studies of inner membrane electrochemical potential-driven transporters.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2008. **1778**(9): p. 1805-1813.
198. Hood, G., Karunakaran, R., Downie, J. A., and Poole, P. , *MgtE From Rhizobium leguminosarum Is a Mg2+ Channel Essential for Growth at Low pH and N2 Fixation on Specific Plants.* Molecular Plant-Microbe Interactions, 2015. **28**(12): p. 1281-1287.
199. Sandler, S.J., Hugenholtz, P., Schleper, C., DeLong, E. F., Pace, N. R., and Clark, A. J. , *Diversity of radA Genes from Cultured and UnculturedArchaea: Comparative Analysis of Putative RadA Proteins and Their Use as a Phylogenetic Marker.* Journal of Bacteriology, 1999. **181**(3): p. 907-915.
200. Santos-Beneit, F., *The Pho regulon: a huge regulatory network in bacteria.* Frontiers in Microbiology, 2015. **6**: p. 402.
201. Wanner, B.L., *Phosphorus assimilation and control of the phosphate regulon. Escherichia coli and Salmonella: cellular and molecular biology, 2nd ed.* Vol. 41. 1996, Washington, DC: ASM Press.
202. Steed, P., M., and Wanner, B., L., *Use of the rep technique for allele replacement to construct mutants with deletions of the pstSCAB-phoU operon: evidence of a new role for the PhoU protein in the phosphate regulon.* Journal of Bacteriology, 1993. **175**(21): p. 6797-809.
203. Li, Y., and Zhang, Y., *PhoU is a persistence switch involved in persister formation and tolerance to multiple antibiotics and stresses in Escherichia coli.* Antimicrobial Agents and Chemotherapy, 2007. **51**(6): p. 2092-2099.
204. Calderón-Flores, A., Du Pont, G., Huerta-Saquero, A., Merchant-Larios, H., Servín-González, L., and Durán, S. , *The stringent response is required for amino acid and nitrate utilization, nod factor regulation, nodulation, and nitrogen fixation in Rhizobium etli.* Journal of Bacteriology, 2005. **187**(15): p. 5075-5083.
205. Hecker, M., Richter, A., Schroeter, A., Wölfel, L., and Mach, F. , *Synthesis of heat shock proteins following amino acid or oxygen limitation in Bacillus*

subtilis relA+ and relA strains. Zeitschrift fur Naturforschung., 1986. **42**(7): p. 941-947.

206. Amar, M., Patriarca, E. J., Manco, G., Bernard, P., Riccio, A., Lamberti, A., and Laccarino, M. , *Regulation of nitrogen metabolism is altered in a glnB mutant strain of Rhizobium leguminosarum.* Molecular Microbiology, 1994. **11**(4): p. 685-693.
207. Mulley, G., White, J. P., Karunakaran, R., Prell, J., Bourdes, A., Bunnewell, S., and Poole, P. S. , *Mutation of GOGAT prevents pea bacteroid formation and N2 fixation by globally downregulating transport of organic nitrogen sources.* Molecular Microbiology, 2011. **80**(1): p. 149-167.
208. Ercolano, R., Mirabella, R., Chiurazzi, M., and Merrick, M. , *The Rhizobium leguminosarum glnB gene is down-regulated during symbiosis.* Molecular and General Genetics, 2001. **264**(5): p. 555-564.
209. Unterholzner, S.J., Poppenberger, B., and Rozhon, W. , *Toxin–antitoxin systems: biology, identification, and application.* Mobile Genetic Elements, 2013. **3**(5): p. e26219.
210. Tambalo, D.D., Yost, C. K., and Hynes, M. F., *Characterization of swarming motility in Rhizobium leguminosarum bv. viciae.* FEMS Microbiology Letters, 2010. **307**(2): p. 165-174.
211. De Troch, P. and J. Vanderleyden, *Surface properties and motility of Rhizobium and Azospirillum in relation to plant root attachment.* Microbial Ecology, 1996. **32**(2): p. 149-169.
212. Frederix, M., A. Edwards, A. Swiderska, A. Stanger, R. Karunakaran, A. Williams, P. Abbruscato, M. Sanchez-Contreras, P.S. Poole, and J.A. Downie, *Mutation of praR in Rhizobium leguminosarum enhances root biofilms, improving nodulation competitiveness by increased expression of attachment proteins.* Molecular Microbiology, 2014. **93**(3): p. 464-478.
213. Perry, B.J., M.S. Akter, and C.K. Yost, *The use of transposon insertion sequencing to interrogate the core functional genome of the legume symbiont Rhizobium leguminosarum.* Frontiers in Microbiology, 2016. **7**: p. 1873.
214. Rigal, A., E. Bouveret, R. Lloubes, C. Lazdunski, and H. Benedetti, *The TolB protein interacts with the porins of Escherichia coli.* Journal of Bacteriology, 1997. **179**(23): p. 7274-7279.

215. Bouveret, E., R. Derouiche, A. Rigal, R. Lloubes, C. Lazdunski, and H. Benedetti, *Peptidoglycan-associated lipoprotein-TolB interaction. A possible key to explaining the formation of contact sites between the inner and outer membranes of Escherichia coli.* Journal of Biological Chemistry, 1995. **270**(19): p. 11071-11077.

216. Lazzaroni, J.-C., J.-F. Dubuisson, and A. Vianney, *The Tol proteins of Escherichia coli and their involvement in the translocation of group A colicins.* Biochimie, 2002. **84**(5-6): p. 391-397.

217. Danese, I., V. Haine, R.-M. Delrue, A. Tibor, P. Lestrate, O. Stevaux, P. Mertens, J.-Y. Paquet, J. Godfroid, and X. De Bolle, *The Ton system, an ABC transporter, and a universally conserved GTPase are involved in iron utilization by Brucella melitensis 16M.* Infection and Immunity, 2004. **72**(10): p. 5783-5790.

218. Tang, K., N. Jiao, K. Liu, Y. Zhang, and S. Li, *Distribution and functions of TonB-dependent transporters in marine bacteria and environments: implications for dissolved organic matter utilization.* PLoS One, 2012. **7**(7): p. e41204.

219. Janczarek, M., K. Rachwał, J. Cieśla, G. Ginalska, and A. Bieganowski, *Production of exopolysaccharide by Rhizobium leguminosarum bv. trifolii and its role in bacterial attachment and surface properties.* Plant and Soil, 2015. **388**(1-2): p. 211-227.

220. Doornbos, R.F., L.C. van Loon, and P.A. Bakker, *Impact of root exudates and plant defense signaling on bacterial communities in the rhizosphere. A review.* Agronomy for Sustainable Development, 2012. **32**(1): p. 227-243.

221. Angulo, V.A.G., H.R. Bonomi, D.M. Posadas, M.I. Serer, A.G. Torres, Á. Zorreguieta, and F.A. Goldbaum, *Identification and characterization of RibN, a novel family of riboflavin transporters from Rhizobium leguminosarum and other proteobacteria.* Journal of Bacteriology, 2013. **195**(20): p. 4611-4619.

222. Maître, J.-L. and C.-P. Heisenberg, *Three functions of cadherins in cell adhesion.* Current Biology, 2013. **23**(14): p. R626-R633.

223. Cao, L., X. Yan, C.W. Borysenko, H.C. Blair, C. Wu, and L. Yu, *CHDL: a cadherin-like domain in Proteobacteria and Cyanobacteria.* FEMS Microbiology Letters, 2005. **251**(2): p. 203-209.

224. Fraiberg, M., I. Borovok, R.M. Weiner, and R. Lamed, *Discovery and characterization of cadherin domains in Saccharophagus degradans 2-40.* Journal of Bacteriology, 2010. **192**(4): p. 1066-1074.

225. Pizarro-Cerdá, J. and P. Cossart, *Bacterial adhesion and entry into host cells.* Cell, 2006. **124**(4): p. 715-727.

226. Richie, D.L., K.T. Takeoka, J. Bojkovic, L.E. Metzger IV, C.M. Rath, W.S. Sawyer, J.-R. Wei, and C.R. Dean, *Toxic accumulation of LPS pathway intermediates underlies the requirement of LpxH for growth of Acinetobacter baumannii ATCC 19606.* PloS One, 2016. **11**(8): p. e0160918.

227. Jofré, E., A. Lagares, and G. Mori, *Disruption of dTDP-rhamnose biosynthesis modifies lipopolysaccharide core, exopolysaccharide production, and root colonization in Azospirillum brasilense.* FEMS Microbiology Letters, 2004. **231**(2): p. 267-275.

228. Smit, G., J. Kijne, and B. Lugtenberg, *Roles of flagella, lipopolysaccharide, and a Ca2+-dependent cell surface protein in attachment of Rhizobium leguminosarum biovar viciae to pea root hair tips.* Journal of Bacteriology, 1989. **171**(1): p. 569-572.

229. Lodwig, E., M. Leonard, S. Marroqui, T. Wheeler, K. Findlay, J. Downie, and P. Poole, *Role of polyhydroxybutyrate and glycogen as carbon storage compounds in pea and bean bacteroids.* Molecular Plant-Microbe Interactions, 2005. **18**(1): p. 67-74.

230. Poortinga, A.T., R. Bos, W. Norde, and H.J. Busscher, *Electric double layer interactions in bacterial adhesion to surfaces.* Surface Science Reports, 2002. **47**(1): p. 1-32.

231. Berne, C., A. Ducret, G.G. Hardy, and Y.V. Brun, *Adhesins involved in attachment to abiotic surfaces by Gram-negative bacteria.* Microbiology Spectrum, 2015. **3**(4).

232. Xi, C., E. Schoeters, J. Vanderleyden, and J. Michiels, *Symbiosis-specific expression of Rhizobium etli casA encoding a secreted calmodulin-related protein.* Proceedings of the National Academy of Sciences, 2000. **97**(20): p. 11114-11119.

233. Robertson, G., M. Hirst, M. Bainbridge, M. Bilenky, Y. Zhao, T. Zeng, G. Euskirchen, B. Bernier, R. Varhol, and A. Delaney, *Genome-wide profiles of*

STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nature Methods, 2007. **4**(8): p. 651.

234. Lodwig, E. and P. Poole, *Metabolism of Rhizobium bacteroids.* Critical Reviews in Plant Sciences, 2003. **22**(1): p. 37-78.

235. Libault, M., *The carbon-nitrogen balance of the nodule and its regulation under elevated carbon dioxide concentration.* BioMed Research International, 2014. **2014**.

236. Walshaw, D.L., Wilkinson, A., Mundy, M., Smith, M., and Poole, P. S. , *Regulation of the TCA cycle and the general amino acid permease by overflow metabolism in Rhizobium leguminosarum.* Microbiology, 1997. **143**(7): p. 2209-2221.

237. Thöny-Meyer, L. and P. Künzler, *The Bradyrhizobium japonicum aconitase gene (acnA) is important for free-living growth but not for an effective root nodule symbiosis.* Journal of Bacteriology, 1996. **178**(21): p. 6166-6172.

238. Fuchs, G., *Alternatives to the Calvin cycle and the Krebs cycle in anaerobic bacteria: pathways with carbonylation chemistry*, in *The molecular basis of bacterial metabolism*. 1990, Springer. p. 13-20.

239. da Costa, C. and E. Galembeck, *The evolution of the Krebs cycle: A promising subject for meaningful learning of biochemistry.* Biochemistry and Molecular Biology Education, 2016. **44**(3): p. 288-296.

240. Dunn, M.F., *Tricarboxylic acid cycle and anaplerotic enzymes in rhizobia.* FEMS Microbiology Reviews, 1998. **22**(2): p. 105-123.

241. Prell, J. and P. Poole, *Metabolic changes of rhizobia in legume nodules.* Trends in Microbiology, 2006. **14**(4): p. 161-168.

242. Walshaw, D.L., A. Wilkinson, M. Mundy, M. Smith, and P.S. Poole, *Regulation of the TCA cycle and the general amino acid permease by overflow metabolism in Rhizobium leguminosarum.* Microbiology, 1997. **143**(7): p. 2209-2221.

243. Terpolilli, J.J., S.K. Masakapalli, R. Karunakaran, I.U. Webb, R. Green, N.J. Watmough, N.J. Kruger, R.G. Ratcliffe, and P.S. Poole, *Lipogenesis and redox balance in nitrogen-fixing pea bacteroids.* Journal of Bacteriology, 2016. **198**(20): p. 2864-2875.

244. Anraku, Y., *Bacterial electron transport chains.* Annual Review of Biochemistry, 1988. **57**(1): p. 101-132.
245. Rabaey, K. and W. Verstraete, *Microbial fuel cells: novel biotechnology for energy generation.* TRENDS in Biotechnology, 2005. **23**(6): p. 291-298.
246. Amar, M., E. Patriarca, G. Manco, P. Bernard, A. Riccio, A. Lamberti, R. Defez, and M. Laccarino, *Regulation of nitrogen metabolism is altered in a glnB mutant strain of Rhizobium leguminosarum.* Molecular Microbiology, 1994. **11**(4): p. 685-693.
247. Patriarca, E., M. Chiurazzi, G. Manco, A. Riccio, A. Lamberti, A. De Paolis, M. Rossi, R. Defez, and M. Iaccarino, *Activation of the Rhizobium leguminosarum glnII gene by NtrC is dependent on upstream DNA sequences.* Molecular and General Genetics MGG, 1992. **234**(3): p. 337-345.
248. Merrick, M. and R. Edwards, *Nitrogen control in bacteria.* Microbiological Reviews, 1995. **59**(4): p. 604-622.
249. Patriarca, E.J., R. Tatè, and M. Iaccarino, *Key role of bacterial NH4+ metabolism in Rhizobium-plant symbiosis.* Microbiology and Molecular Biology Reviews, 2002. **66**(2): p. 203-222.
250. Napolitani, C., L. Mandrich, A. Riccio, A. Lamberti, G. Manco, and E.J. Patriarca, *Mutational analysis of GstI protein, a glutamine synthetase translational inhibitor of Rhizobium leguminosarum.* FEBS Letters, 2004. **558**(1-3): p. 45-51.
251. Glenn, A. and M. Dilworth, *Oxidation of substrates by isolated bacteroids and free-living cells of Rhizobium leguminosarum 3841.* Microbiology, 1981. **126**(1): p. 243-247.
252. Wielbo, J., M. Marek-Kozaczuk, A. Mazur, A. Kubik-Komar, and A. Skorupska, *Genetic and metabolic divergence within a Rhizobium leguminosarum bv. trifolii population recovered from clover nodules.* Applied and Environmental Microbiology, 2010. **76**(14): p. 4593-4600.
253. Prell, J., and Poole, P. , *Metabolic changes of rhizobia in legume nodules.* Trends in Microbiology, 2006. **14**(4): p. 161-168.
254. Downie, J.A., *Legume haemoglobins: symbiotic nitrogen fixation needs bloody nodules.* Current Biology, 2005. **15**(6): p. 196-198.

255. Carro, L., T. Persson, P. Pujic, N. Alloisio, P. Fournier, H. Boubakri, K. Pawlowski, and P. Normand, *Organic acids metabolism in Frankia alni.* Symbiosis, 2016. **70**(1-3): p. 37-48.
256. Jeong, J., S. Suh, C. Guan, Y.-F. Tsay, N. Moran, C.J. Oh, C.S. An, K.N. Demchenko, K. Pawlowski, and Y. Lee, *A nodule-specific dicarboxylate transporter from alder is a member of the peptide transporter family.* Plant Physiology, 2004. **134**(3): p. 969-978.
257. Hwang, J.H., S.R. Ellingson, and D.M. Roberts, *Ammonia permeability of the soybean nodulin 26 channel.* FEBS Letters, 2010. **584**(20): p. 4339-4343.
258. Peekhaus, N., and Conway, T. , *What's for dinner?: Entner-Doudoroff metabolism in Escherichia coli.* Journal of Bacteriology, 1998. **180**(14): p. 3495-3502.
259. Borah, K., *Metabolic flux analysis of the nitrogen-fixing bacterium Azorhizobium caulinodans*, in *Department of Plant Sciences*. 2016, University of Oxford.
260. Cevallos, M.A., S. Encarnacion, A. Leija, Y. Mora, and J. Mora, *Genetic and physiological characterization of a Rhizobium etli mutant strain unable to synthesize poly-beta-hydroxybutyrate.* Journal of Bacteriology, 1996. **178**(6): p. 1646-1654.
261. McDermott, T.R., S.M. Griffith, C.P. Vance, and P.H. Graham, *Carbon metabolism in Bradyrhizobium japonicum bacteroids.* FEMS Microbiology Letters, 1989. **63**(4): p. 327-340.
262. Tang, Y. and R.I. Hollingsworth, *Regulation of Lipid Synthesis inBradyrhizobium japonicum: Low Oxygen Concentrations Trigger Phosphatidylinositol Biosynthesis.* Applied and Environmental Microbiology, 1998. **64**(5): p. 1963-1966.
263. Udvardi, M., and Poole, P. S. , *Transport and metabolism in legume-rhizobia symbioses.* Annual Review of Plant Biology, 2013. **64**: p. 781-805.
264. Day, D., P. Poole, S. Tyerman, and L. Rosendahl, *Ammonia and amino acid transport across symbiotic membranes in nitrogen-fixing legume nodules.* Cellular and Molecular Life Sciences CMLS, 2001. **58**(1): p. 61-71.

265. Dreyfus, B.L., Elmerich, C., and Dommergues, Y. R. , *Free-living Rhizobium strain able to grow on N2 as the sole nitrogen source.* Applied and Environmental Microbiology, 1983. **45**(2): p. 711-713.
266. Hunt, S., S.T. Gaito, and D.B. Layzell, *Model of gas exchange and diffusion in legume nodules.* Planta, 1988. **173**(1): p. 128-141.
267. Hunt, S. and D.B. Layzell, *Gas exchange of legume nodules and the regulation of nitrogenase activity.* Annual Review of Plant Biology, 1993. **44**(1): p. 483-511.
268. MINCHIN, F.R. and J. Witty, *Effects of acetylene and external oxygen concentration on the respiratory quotient (RQ) of nodulated roots of soyabean and white clover.* Journal of Experimental Botany, 1990. **41**(10): p. 1271-1277.
269. Witty, J. and F. Minchin, *Oxygen diffusion in the legume root nodule*, in *Nitrogen Fixation*. 1990, Springer. p. 285-292.
270. Denison, R.F., *Legume sanctions and the evolution of symbiotic cooperation by rhizobia.* The American Naturalist, 2000. **156**(6): p. 567-576.
271. Oono, R., R.F. Denison, and E.T. Kiers, *Controlling the reproductive fate of rhizobia: how universal are legume sanctions?* New Phytologist, 2009. **183**(4): p. 967-979.
272. Wang, Q., J. Liu, and H. Zhu, *Genetic and Molecular Mechanisms Underlying Symbiotic Specificity in Legume-Rhizobium Interactions.* Frontiers in Plant Science, 2018. **9**: p. 313.
273. Cao, Y., M.K. Halane, W. Gassmann, and G. Stacey, *The role of plant innate immunity in the legume-rhizobium symbiosis.* Annual Review of Plant Biology, 2017. **68**: p. 535-561.
274. DeJesus, M.A., C. Ambadipudi, R. Baker, C. Sassetti, and T.R. Ioerger, *TRANSIT-a software tool for Himar1 TnSeq analysis.* PLoS Computational Biology, 2015. **11**(10): p. e1004401.
275. Wetmore, K.M., M.N. Price, R.J. Waters, J.S. Lamson, J. He, C.A. Hoover, M.J. Blow, J. Bristow, G. Butland, and A.P. Arkin, *Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons.* MBio, 2015. **6**(3): p. e00306-15.

276. Iyer, B., M.S. Rajput, R. Jog, E. Joshi, K. Bharwad, and S. Rajkumar, *Organic acid mediated repression of sugar utilization in rhizobia.* Microbiological Research, 2016. **192**: p. 211-220.
277. Stowers, M.D., *Carbon metabolism in Rhizobium species.* Annual Reviews in Microbiology, 1985. **39**(1): p. 89-108.
278. Jurtshuk, P.J., *Bacterial Metabolism (chapter 4) in Medical Microbiology (4th edition)*, ed. B. S. 1996, Galveston, Texas: University of Texas Medical Branch.

Appendix A. Supplementary material for chapter 3 and chapter 4

Overview

Chapter 3: Table 1 – Table 6. Published in Wheatley et al. (2017) [1].

Chapter 4: Table 7 – Table 10.

Table 1. The number of genes assigned to each growth phenotype within each replicon across the four growth conditions[a].

Growth condition	INSeq classification[b]	Number of genes per growth classification[c]						
		pRL7	pRL8	pRL9	pRL10	pRL11	pRL12	Chromosome
G21	ES	6.0 ± 1.2	2.7 ± 1.3	7.7 ± 1.2	10.7 ± 1.20	6.3 ± 2.4	7.0 ± 0.6	307.7 ± 28.1
	NE	142.7 ± 5.8	110.0 ± 5.5	249.7 ± 2.9	405.0 ± 4.00	596.7 ± 1.2	739.3 ± 3.8	3850.3 ± 34.4
	GD	31.3 ± 5.2	29.3 ± 5.8	55.3 ± 3.5	52.7 ± 2.19	39.0 ± 2.1	42.7 ± 3.5	619.3 ± 23.3
	GA	0.0 ± 0.0	0.0 ± 0.0	0.3 ± 0.3	0.7 ± 0.67	0.0 ± 0.0	0.0 ± 0.0	5.7 ± 2.3
S21	ES	4.0 ± 2.3	4.7 ± 1.2	7.7 ± 1.9	8.0 ± 0.58	4.7 ± 1.5	6.0 ± 2.5	219.7 ± 27.5
	NE	138.0 ± 5.0	107.7 ± 6.0	246.7 ± 5.2	413.3 ± 1.46	601.7 ± 5.7	750.0 ± 2.3	3833.0 ± 5.3
	GD	38.0 ± 5.3	29.7 ± 6.12	58.0 ± 5.3	46.3 ± 0.67	35.7 ± 4.3	33.0 ± 1.0	721.0 ± 29.1
	GA	0.0 ± 0.0	0.0 ± 0	0.7 ± 0.3	1.3 ± 0.67	0.0 ± 0.0	0.0 ± 0.0	9.3 ± 3.2
G1	ES	5.7 ± 1.4	2.7 ± 0.9	8.7 ± 0.7	11.0 ± 1.15	8.3 ± 2.9	7.0 ± 1.2	265.3 ± 34.7
	NE	123.0 ± 3.5	108.3 ± 3.7	244.3 ± 6.2	408.7 ± 2.96	601.0 ± 2.3	743.7 ± 2.9	3932.7 ± 12.9
	GD	50.7 ± 4.5	31.0 ± 3.8	59.3 ± 6.3	48.7 ± 2.19	32.7 ± 4.6	38.3 ± 3.7	581.3 ± 34.8
	GA	2.0 ± 0.3	0.0 ± 0	0.7 ± 0.3	0.7 ± 0.67	0.0 ± 0.0	0.0 ± 0.0	3.7 ± 1.9
S1	ES	2.0 ± 1.0	4.3 ± 1.3	9.0 ± 2.3	8.7 ± 1.20	5.3 ± 3.2	9.0 ± 3.0	284.7 ± 57.1
	NE	131.0	94.7	236.3	402.7	596.7	739.3	3849.3

	± 4.0	± 9.0	± 6.4	± 3.93	± 2.4	± 1.8	± 23.3
GD	47.0	43.0	67.3	56.3	40.0	40.7	645.3
	± 3.0	± 8.0	± 6.9	± 3.53	± 1.0	± 2.9	± 32.8
GA	0.0	0.0	0.3	1.3	0.0	0.0	3.7
	± 0.0	± 0.0	± 0.3	± 0.67	± 0.0	± 0.0	± 1.6

[a] G21; glucose at 21% O_2, S21; succinate at 21% O_2, G1; glucose at 1% O_2, S1; succinate at 1% O_2.

[b] ES; growth-essential, NE; growth-neutral, GD; growth-defective, GA; growth-advantaged.

[c] Number of genes is given as the three-replicate average ± the SEM.

Table 2. INSeq classification of Rlv3841-encoded enzymes of glucose and succinate central metabolism under different growth conditions[a].

				INSeq classification[c]			
Locus	Gene	Description	Pathway[b]	G21	S21	G1	S1
pRL80070		Aldehyde dehydrogenase	Gluconeogenesis	NE	NE	NE	NE
pRL120710	*glgB*	1,4-alpha-glucan branching enzyme	Glycogen synthesis	NE	NE	NE	NE
RL0037	*pckA*	Phosphoenolpyruvate carboxykinase	Pyruvate metabolism	NE	GD	NE	GD
RL0054	*glcB*	Malate synthase	Glyoxylate pathway	NE	NE	NE	NE
RL0182	*glk*	Glucokinase	ED pathway	ES	NE	GD	NE
RL0183	*msgA*	Methylglyoxal synthase	Hypothetical methyglyoxal pathway	ES	NE	GD	NE
RL0751	*edd*	Phosphogluconate dehydratase	ED pathway	ES	GD	ES	GD
RL0752	*pgl*	6-phosphogluconolactonase	ED pathway/PPP	ES	GD	ES	GD
RL0753	*zwf*	Glucose-6-phosphate 1-dehydrogenase	ED pathway/PPP	ES	GD	ES	GD
RL0761		Isocitrate lyase	Glyoxylate pathway	NE	NE	NE	NE
RL1010	*gpmB*	Phosphoglycerate mutase	Gluconeogenesis	GD	GD	GD	GD
RL1351	*zwf*	Glucose-6-phosphate 1-dehydrogenase	ED pathway/PPP	NE	NE	NE	NE
RL2234	*gltA*	Citrate synthase	TCA-cycle	GD	GD	GD	NE
RL2239	*eno*	Enolase	Gluconeogenesis	GD	GD	GD	GD
RL2241	*pdhA*	Pyruvate dehydrogenase subunit A	Pyruvate metabolism	GD	GD	GD	GD

RL2243	*pdhC*	Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex (PDC)	Pyruvate metabolism	GD	GD	GD	GD
RL2508	*gltA*	Citrate synthase II	TCA-cycle	NE	NE	NE	NE
RL2509	*citA*	Citrate synthase I	TCA-cycle	NE	GD	NE	NE
RL2513	*tpiA*	Triosephosphate isomerase	Gluconeogenesis/ hypothetical methyglyoxal pathway	NE	NE	NE	NE
RL2631	*icd*	Isocitrate dehydrogenase [NADP]	TCA-cycle	NE	NE	NE	NE
RL2701	*fumC*	Fumarate hydratase class II	TCA-cycle	NE	NE	NE	NE
RL2703	*fumA*	Fumarate hydratase class I, aerobic	TCA-cycle	NE	NE	NE	NE
RL4007	*gap*	Glyceraldehyde-3-phosphate dehydrogenase	Gluconeogenesis	GD	GD	GD	GD
RL4011	*pgk*	Phosphoglycerate kinase	Gluconeogenesis	NE	NE	NE	NE
RL4060	*pykA*	Pyruvate kinase	Pyruvate metabolism	ES	GD	ES	GD
RL4114	*glgP*	Glycogen phosphorylase	Glycogen synthesis	NE	NE	NE	NE
RL4115	*glgB*	1,4-alpha-glucan branching enzyme	Glycogen synthesis	NE	NE	NE	NE
RL4116	*glgC*	Glucose-1-phosphate adenylyltransferase	Glycogen synthesis	NE	NE	NE	NE
RL4117	*glgA*	Glycogen synthase	Glycogen synthesis	GD	NE	GD	NE
RL4118	*pgm*	phosphoglucomutase	Glycogen synthesis	GD	GD	GD	GD

RL4162	*eda*	2-dehydro-3-deoxyphosphogluconate aldolase	ED pathway	NE	NE	NE	NE
RL4433	*sucB*	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase	TCA-cycle	GD	GD	GD	GD
RL4435	*sucA*	2-oxoglutarate dehydrogenase E1 component	TCA-cycle	GD	GD	GD	GD
RL4436	*sucD*	Succinyl-CoA synthetase alpha chain	TCA-cycle	GD	GD	GD	GD
RL4439	*mdh*	Malate dehydrogenase	TCA-cycle	GD	GD	GD	GD
RL4443	*sdhB*	Succinate dehydrogenase iron-sulfur protein	TCA-cycle	GD	GD	GD	GD
RL4536	*acnA*	Aconitate hydratase	TCA-cycle	ES	ES	ES	ES
RL4638		pyruvate carboxylase	Pyruvate metabolism	GD	NE	GD	NE

[a] G21; glucose at 21% O_2, S21; succinate at 21% O_2, G1; glucose at 1% O_2, S1; succinate at 1% O_2.

[b] ED pathway; Entner-Doudoroff pathway, TCA-cycle; Tricarboxylic acid (TCA) cycle, PPP; Pentose Phosphate Pathway.

[c] ES; growth-essential, GD; growth-defective, NE; growth-neutral.

Table 3. INSeq-identified Rlv3841 genes, mutation of which specifically reduces growth on glucose, together with INSeq classification under different growth conditions[a]

			INSeq classification[b]			
Locus	Gene	Description	G21	S21	G1	S1
pRL110211		SBP of ABC transporter, PAAT family	GD	NE	GD	NE
pRL110565		Transmembrane protein	ES	NE	ES	NE
pRL120207	*eryD*	DeoR family transcriptional regulator (repressor)	GD	NE	GD	NE
pRL120419	*opaA*	Omega-amino acid:pyruvate transaminase	GD	NE	GD	NE
RL0182	*glk*	Glucokinase	ES	NE	GD	NE
RL0183	*mgsA*	Methylglyoxal synthase	ES	NE	GD	NE
RL0547	*phoB*	Phosphate regulon transcriptional regulator PhoB	GD	NE	GD	NE
RL1479		Conserved hypothetical protein	ES	NE	ES	NE
RL1754		Conserved hypothetical exported protein	GD	NE	GD	NE
RL2082	*gatB*	Aspartyl/glutamyl-tRNA(asn/gln) amidotransferase subunit B	ES	NE	GD	NE
RL3005		Hypothetical protein	ES	NE	ES	NE
RL4117	*glgA*	Glycogen synthase	GD	NE	GD	NE
RL4638		Pyruvate carboxylase	GD	NE	GD	NE

[a] G21; glucose at 21% O_2, S21; succinate at 21% O_2, G1; glucose at 1% O_2, S1; succinate at 1% O_2.

[b] ES; growth-essential, GD; growth-defective, NE; growth-neutral.

Table 4. INSeq-identified Rlv3841 genes, mutation of which specifically reduces growth on succinate, together with INSeq classification under different growth conditions[a]

			INSeq classification[b]			
Locus	Gene	Description	G21	S21	G1	S1
pRL80122	*traA*	Conjugal transfer protein TraA	NE	GD	NE	GD
pRL90053		O-antigen ligase	NE	GD	NE	GD
pRL90207		Permease component of ABC transporter, CUT2 family	NE	GD	NE	GD
pRL90208	*codA*	Cytosine deaminase	NE	GD	NE	GD
pRL100388		LacI family transcriptional regulator (repressor)	NE	GD	NE	GD
RL0037	*pckA*	Phosphoenolpyruvate carboxykinase	NE	GD	NE	GD
RL0055		Conserved hypothetical protein	NE	ES	NE	ES
RL0423		Transmembrane protein	NE	GD	NE	GD
RL0588		Peptidoglycan binding protein	NE	GD	NE	GD
RL0754	*ordL*	Oxidoreductase	NE	GD	NE	GD
RL0957		Conserved hypothetical protein	NE	GD	NE	GD
RL1388		Conserved hypothetical protein	NE	GD	NE	GD
RL1390		Hypothetical protein	NE	GD	NE	GD
RL1391		Transmembrane protein	NE	GD	NE	GD
RL1392		Transmembrane protein	NE	GD	NE	GD
RL1393	*pbpF*	peptidoglycan biosynthesis/penicillin binding protein	NE	GD	NE	GD
RL1561	*pabC*	Aminodeoxychorismate lyase	NE	GD	NE	GD
RL1641		MerR family transcriptional regulator	NE	GD	NE	GD
RL1642		Two-component sensor/regulator; transcriptional regulator	NE	GD	NE	GD
RL2384	*recG*	ATP-dependent DNA helicase	NE	ES	NE	GD
RL2385		Conserved hypothetical protein	NE	ES	NE	GD
RL2478		Outer membrane protein	NE	GD	NE	GD

RL2990	*ubiA*	Para-hydroxybenzoate--polyprenyltransferase	NE	GD	NE	GD
RL3424	*dctA*	C4-dicarboxylate transport protein	NE	GD	NE	GD
RL3425	*dctB*	Histidine kinase (C4-dicarboxylate transport)	NE	GD	NE	GD
RL3426	*dctD*	Transcriptional regulator C4-dicarboxylate transport (sigma-54)	NE	GD	NE	GD
RL3427		AsnC family transcriptional regulator	NE	GD	NE	GD
RL3465		Conserved hypothetical protein	NE	GD	NE	GD
RL3468	*prs*	Ribose-phosphate pyrophosphokinase	NE	ES	NE	ES
RL3512		Conserved hypothetical protein	NE	GD	NE	GD
RL3654		Polysaccharide biosynthesis protein	NE	GD	NE	GD
RL3655		Glycosyltransferase	NE	GD	NE	GD
RL3656		lipase	NE	GD	NE	GD
RL3657	*prsE*	Permease component of ABC transporter, Export family	NE	GD	NE	GD
RL3668		serine/threonine protein phosphatase	NE	GD	NE	ES
RL4523		Conserved hypothetical exported protein	NE	GD	NE	ES
RL4524	*ecfL*	RNA polymerase ECF sigma factor	NE	GD	NE	ES
RL4658	*exoB*	UDP-glucose 4-epimerase	NE	GD	NE	ES

[a] G21; glucose at 21% O_2, S21; succinate at 21% O_2, G1; glucose at 1% O_2, S1; succinate at 1% O_2.

[b] ES; growth-essential, GD; growth-defective, NE; growth-neutral.

Table 5. INSeq-identified Rlv3841 genes, mutation of which specifically reduces growth at 21% O_2, together with INSeq classification under different growth conditions[a].

			INSeq classification[b]			
Locus	Gene	Description	G21	S21	G1	S1
pRL90042		Pseudogene, conserved hypothetical protein	GD	GD	NE	NE
pRL120796		Hypothetical exported protein	GD	GD	NE	NE
RL0802		Deoxygenase	GD	GD	NE	NE
RL0847	*guaB*	Inosine-5'-monophosphate dehydrogenase	ES	ES	NE	NE
RL0920		ATP-binding mrp family protein	GD	GD	NE	NE
RL0921		Cationic transport protein, CorA family	GD	GD	NE	NE
RL0922	*kup*	Potassium uptake transport system protein	GD	GD	NE	NE
RL1021	*coxB*	Cytochrome c oxidase polypeptide II precursor (cytochrome aa3 subunit 2)	GD	GD	NE	NE
RL1022	*coxA*	Cytochrome c oxidase polypeptide I (cytochrome aa3 subunit 1)	GD	GD	NE	NE
RL1547	*cvpA*	Colicin V production protein	GD	GD	NE	NE
RL1548	*radA*	DNA repair protein RadA homologue	GD	GD	NE	NE
RL2440		Peroxiredoxin	GD	GD	NE	NE
RL2441		Adenylate cyclase	GD	GD	NE	NE
RL2926		Adenylyl cyclase	GD	GD	NE	NE
RL2927	*ohrB*	Organic hydroperoxide resistance protein	GD	GD	NE	NE
RL3180		Hypothetical protein	GD	GD	NE	NE
RL3181		Conserved hypothetical protein	GD	GD	NE	NE
RL3182		Conserved hypothetical protein	GD	GD	NE	NE

RL3507	Phosphoesterase	ES	GD	NE	NE
RL3596	Conserved hypothetical protein	GD	GD	NE	NE
RL3597	DEAD-box ATP-dependent RNA helicase protein	GD	GD	NE	NE
RL3834	ErfK/YbiS/YhnG oxidoreductase	GD	GD	NE	NE
RL4186	Oxidoreductase	GD	GD	NE	NE

[a] G21; glucose at 21% O_2, S21; succinate at 21% O_2, G1; glucose at 1% O_2, S1; succinate at 1% O_2.

[b] ES; growth-essential, GD; growth-defective, NE; growth-neutral.

Table 6. INSeq-identified Rlv3841 genes, mutation of which specifically reduces growth at 1% O_2, together with INSeq classification under different growth conditions[a]

			INSeq classification[b]			
Locus	Gene	Description	G21	S21	G1	S1
pRL80039		Pseudogene	NE	NE	GD	GD
pRL100141		Pseudogene, conserved hypothetical protein	NE	NE	GD	GD
pRL110016		Conserved hypothetical protein	NE	NE	GD	GD
pRL110072		GntR family transcriptional regulator	NE	NE	GD	GD
RL0546	*phoU*	Phosphate uptake regulator PhoU	NE	NE	GD	GD
RL1506	*relA*	Stringent response protein	NE	NE	GD	GD
RL2585	*nodX*	Nodulation protein NodX (probable sugar acetylase)	NE	NE	GD	GD
RL3412		Conserved hypothetical protein	NE	NE	GD	GD
RL4042		AsrR family transcriptional regulator	NE	NE	GD	GD
RL4693		Polysaccharide deacetylase	NE	NE	GD	GD
RL4694		(di)nucleoside polyphosphate hydrolase	NE	NE	GD	GD
RL4695	*bfr*	Bacterioferritin	NE	NE	GD	GD

[a] G21; glucose at 21% O_2, S21; succinate at 21% O_2, G1; glucose at 1% O_2, S1; succinate at 1% O_2.

[b] ES; growth-essential, GD; growth-defective, NE; growth-neutral.

Table 7. INSeq-identified[a] core fitness requirements in Rlv3841. Genes that gave growth defective or growth essential phenotypes upon mutation universally across all INSeq libraries.

			INSeq classification[b]								
Locus	Gene	Description	In	TY	V	G	S	Rh	Ro	No	Re
pRL100001	*repAp10a*	putative replication protein RepA	GD	ES	ES	ES	ES	GD	GD	ES	ES
pRL100002	*repBp10*	putative replication protein RepB	GD	ES	ES	ES	ES	GD	GD	ES	ES
pRL100003	*repCp10*	putative replication protein RepC	GD	ES	ES	ES	ES	GD	GD	ES	ES
pRL100022A		putative transmembrane transporter protein	GD	GD	ES	ES	ES	GD	GD	ES	ES
pRL100111		hypothetical protein	GD	ES	GD	ES	ES	GD	GD	ES	ES
pRL110001	*repAp11*	putative replication protein RepA	GD	ES	ES	GD	GD	GD	GD	GD	ES
pRL110002	*repBp11*	putative replication protein RepB	GD	ES	ES	GD	GD	GD	GD	GD	ES
pRL110003	*repCp11*	putative replication protein RepC	GD	ES	ES	GD	GD	GD	GD	GD	ES
pRL110626	*cobL*	precorrin-6y c(5,15)-methyltransferase	GD	GD	ES	GD	GD	GD	ES	ES	ES
pRL110627	*cobK*	precorrin-6a reductase	GD	GD	ES	GD	GD	GD	ES	ES	ES
pRL110628	*cobJ*	precorrin-3b c(17)-methyltransferase	GD	GD	ES	GD	GD	GD	ES	ES	ES
pRL110629	*cobI*	precorrin-2 c(20)-methyltransferase	GD	GD	ES	GD	GD	GD	GD	ES	ES
pRL110630	*cobH*	putative precorrin-8x methylmutase	GD	GD	ES	GD	GD	GD	GD	ES	ES
pRL110631	*cobG*	putative precorrin-3b synthase	GD	GD	ES	GD	GD	GD	GD	ES	GD

pRL120001	*repAp 12*	putative replication protein RepA	GD	ES	ES	ES	GD	GD	GD	ES	GD
pRL120002	*repBp 12*	putative replication protein RepB	GD	ES	ES	ES	GD	GD	GD	ES	GD
pRL120003	*repCp 12*	putative replication protein RepC	GD	ES	ES	ES	GD	GD	GD	ES	GD
pRL70047E		putative transmembrane protein	GD	ES	ES	GD	GD	GD	GD	ES	ES
pRL70092	*repAp 7c*	putative replication protein RepA	GD	GD	GD	ES	ES	GD	GD	GD	ES
pRL70093	*repBp 7c*	putative replication protein RepB	GD	GD	GD	ES	ES	GD	GD	GD	ES
pRL70094	*repCp 7c*	putative replication protein RepC	GD	GD	GD	ES	ES	ES	GD	GD	ES
pRL80001	*repAp 8*	putative replication protein RepA	GD	ES	ES	ES	ES	GD	GD	ES	ES
pRL80002	*repBp 8*	putative replication protein RepB	GD	ES	ES	ES	ES	GD	GD	ES	ES
pRL80003	*repCp 8*	putative replication protein RepC	GD	ES	ES	ES	ES	GD	GD	ES	ES
pRL90001	*repAp 9*	putative replication protein RepA	GD	ES	ES	ES	ES	GD	GD	ES	ES
pRL90002	*repBp 9*	putative replication protein RepB	GD	ES	ES	ES	ES	GD	GD	ES	ES
pRL90003	*repCp 9*	putative replication protein RepC	GD	ES	ES	ES	ES	ES	GD	ES	ES
RL0005		putative DNA polymerase subunit	GD	ES	ES	GD	GD	GD	ES	ES	ES
RL0012	*gyrB1*	putative DNA gyrase subunit B	GD	ES	ES	GD	ES	GD	GD	ES	ES
RL0023		putative acetyl-CoA carboxylase carboxyl transferase subunit	GD	GD	ES	GD	GD	GD	GD	ES	ES

RL0024	*folC2*	putative FolC bifunctional protein [Includes: Folylpolyglutamate synthase (EC 6.3.2.17) (Folylpoly-gamma-glutamate synthetase) (FPGS); Dihydrofolate synthase].	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL0030	*divL*	putative two-component sensor/regulator; histidine kinase	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL0035	*chvG*	putative two-component sensor/regulator; histidine kinase	GD	ES	GD	GD	GD	ES	GD	ES	ES
RL0040	*coaA*	putative pantothenate kinase	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL0041		putative histidine biosynthesis bifunctional protein	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL0042	*hisF*	putative imadazole glycerol phosphate synthase sununit	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL0043	*hisA*	putative imidazole-4-carboxamide isomerase	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL0046	*hisH*	putative imadazole glycerol phosphate synthase subunit	GD	GD	ES	ES	ES	ES	GD	ES	ES
RL0047		putative transmembrane protein	GD	GD	ES	ES	ES	ES	GD	ES	ES
RL0106	*rpsA*	putative 30S ribosomal protein S1	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL0116	*fabA*	putative 3-hydroxydecanoyl-[acyl-	GD	ES	ES	GD	GD	ES	GD	ES	ES

		carrier-protein] dehydratase (Beta-hydroxydecanoyl thioester dehydrase)									
RL0117	*fabB*	putative 3-oxoacyl-[acyl-carrier-protein] synthase I	GD	ES	ES	GD	GD	ES	GD	ES	ES
RL0118	*fabI1*	putative enoyl-[acyl-carrier-protein] reductase [NADH] (NADH-dependent enoyl-ACP reductase)	GD	ES	ES	GD	GD	ES	GD	ES	ES
RL0120	*pnp*	putative polynribonucleotide nucleotidyltransferase (polynucleotide phosphorylase) (PNPase)	GD	ES	ES	GD	ES	ES	GD	ES	ES
RL0121	*rpsO*	putative 30S ribosomal protein S15	GD	ES	ES	GD	ES	ES	GD	ES	ES
RL0125	*infB*	putative translation initiation factor IF-2	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL0126		conserved hypothetical protein	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL0127	*nusA*	putative transcription elongation protein (N utilisation substance protein A) (L factor)	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL0128		conserved hypothetical protein	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL0134	*dnaX*	putative DNA polymerase III	GD	ES	ES	ES	ES	ES	GD	ES	ES
RL0140	*kdsB*	putative 3-deoxy-manno-octulosonate cytidylyltransferase	GD	ES	ES	ES	ES	GD	ES	ES	ES

		(CMP-KDO synthetase) (CMP-2-keto-3-deoxyoctulosonic acid synthetase)									
RL0151	*dnaJ*	putative chaperone protein	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL0152	*dnaK*	putative heat shock chaperone protein	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL0181		putative ATP-binding:permease (ABC:IMP) component of ABC transporter Export, MsbA-like	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL0269	*pheS*	putative phenylalanyl-tRNA synthetase alpha chain	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL0270	*pheT*	putative phenylalanyl-tRNA synthetase beta chain	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL0315	*guaA*	putative GMP synthase [glutamine hydrolysing]	GD	GD	ES	GD	GD	GD	ES	ES	ES
RL0328		undecaprenyl-diphosphatase (undecaprenyl pyrophosphate phosphatase) (bacitracin resistance protein)	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL0371	*ubiE*	putative ubiquinone/menaquinone biosynthesis methyltransferase	GD	ES	ES	ES	GD	ES	ES	ES	ES
RL0375	*dnaA*	putative chromosomal replication initiator protein	GD	ES	ES	ES	ES	ES	GD	ES	GD

RL0389	*metK*	putative S-adenosylmethionine synthetase	GD	ES	ES	GD	GD	GD	ES	ES	ES
RL0404	*mviN*	putative transmembrane MviN virulence factor homologue	GD	GD	ES	ES	ES	ES	GD	ES	ES
RL0405	*glnD*	putative [protein-PII] uridylyltransferase	GD	GD	ES	ES	ES	ES	GD	ES	ES
RL0411		rRNA methyltransferase	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL0412		conserved hypothetical protein	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL0419		putative transmembrane protein	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL0421		putative ATP-binding component of ABC transporter Unclass	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL0436	*dapE*	putative succinyl-diaminopimelate desuccinylase	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL0437	*dapD*	putative 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL0452	*engB*	putative GTP-binding protein	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL0453	*oxaA*	putative transmembrane protein	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL0454		putative ribonuclease protein component	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL0542		putative SBP of ABC transporter PhoT	GD	GD	GD	ES	ES	GD	GD	ES	ES

RL0543		putative permease component of ABC transporter PhoT	GD	GD	GD	ES	ES	GD	GD	ES	ES
RL0544		putative permease component of ABC transporter PhoT	GD	GD	GD	ES	ES	GD	GD	ES	ES
RL0611	*murA*	putative UDP-N-acetylglucosamine 1-carboxyvinyltransferase (enoylpyruvate transferase) (udp-n-acetylglucosamine enolpyruvyl transferase) (ept)	GD	ES	ES	ES	GD	GD	GD	GD	GD
RL0612		conserved hypothetical protein	GD	GD	ES	ES	GD	GD	GD	GD	GD
RL0613	*hisD1*	putative histidinol dehydrogenase	GD	GD	ES	ES	GD	GD	GD	GD	GD
RL0619		conserved hypothetical protein	GD	GD	GD	GD	GD	GD	GD	GD	ES
RL0620		conserved hypothetical protein	GD	GD	GD	GD	GD	GD	GD	GD	ES
RL0626		conserved hypothetical protein	ES	GD	GD	ES	ES	GD	GD	GD	GD
RL0665		putative transmembrane exopolysaccharide production protein	GD	GD	ES	ES	ES	GD	GD	ES	ES
RL0680	*secDF2*	putative transmembrane export SecD/F family protein SecDF2	GD	GD	ES	ES	ES	ES	GD	ES	ES
RL0681		putative transmembrane protein	GD	GD	ES	ES	ES	ES	GD	ES	ES

RL0752	*pgl*	putative 6-phosphogluconolactonase	GD	ES	ES	ES	GD	GD	ES	GD	ES
RL0753	*zwf1*	putative glucose-6-phosphate 1-dehydrogenase	GD	ES	ES	ES	GD	GD	ES	GD	ES
RL0769	*cbrA*	putative two-component sensor/regulator; histidine kinase	GD	GD	GD	ES	ES	GD	GD	ES	ES
RL0842		putative guanosine-5'-triphosphate,3'-diphosphate pyrophosphatase	GD	GD	GD	ES	GD	ES	GD	GD	GD
RL0843		putative ribosomal RNA methyltransferase/cell division FtsJ protein	GD	GD	GD	ES	GD	ES	GD	GD	GD
RL0877		putative histidyl-tRNA synthetase	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL0878	*hisZ*	putative ATP phosphoribosyltransferase regulatory subunit	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL0879	*hisG*	putative ATP phosphoribosyltransferase	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL0883	*groLch1*	putative 60 kda chaperonin (protein cpn60)	GD	ES	ES	GD	GD	GD	GD	GD	ES
RL0884	*groSch1*	putative 10 kda chaperonin (protein cpn10)	GD	ES	ES	GD	GD	GD	GD	GD	ES
RL0886	*ribF*	putative riboflavin biosynthesis protein	GD	GD	ES	GD	GD	GD	GD	GD	ES
RL0889	*ileS*	putative isoleucyl-tRNA synthetase	GD	ES	ES	GD	GD	GD	GD	ES	ES

RL0924	*atpI*	putative ATP synthase protein I	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL0925	*atpB*	putative ATP synthase A chain	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL0926	*atpE*	putative ATP synthase subunit C	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL0927	*atpG2*	putative ATP synthase B' chain (subunit II)	GD	ES	ES	GD	GD	GD	GD	ES	GD
RL0928	*atpF*	putative ATP synthase B chain precursor	GD	ES	ES	GD	GD	GD	GD	ES	GD
RL0935		putative kinase	GD	GD	ES	ES	ES	GD	GD	ES	ES
RL0936		conserved hypothetical protein (TPR repeat family)	GD	GD	ES	ES	ES	GD	GD	ES	ES
RL0937	*ispB*	putative octaprenyl-diphosphate synthase	GD	GD	ES	ES	ES	GD	GD	ES	ES
RL0943	*glyQ*	putative glycyl-tRNA synthetase alpha chain	GD	ES	ES	ES	GD	GD	GD	GD	GD
RL0973	*dxs*	putative 1-deoxy-d-xylulose-5-phosphate synthase	ES	ES	ES	GD	ES	GD	ES	ES	ES
RL1007	*aroC*	putative chorismate synthase	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL1011	*fabI2*	putative enoyl-[acyl-carrier-protein] reductase [NADH] (NADH-dependent enoyl-ACP reductase)	ES	ES	ES	GD	GD	GD	GD	ES	ES
RL1030	*ispH*	putative 4-hydroxy-3-methylbut-2-enyl diphosphate reductase	GD	GD	ES	ES	GD	GD	GD	ES	ES

RL1031	*thrB*	putative homoserine kinase	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL1059		putative transmembrane GGDEF/EAL sensory box protein	GD	GD	ES	GD	ES	GD	ES	ES	ES
RL1062	*thrC*	putative threonine synthase	GD	ES	ES	ES	ES	ES	ES	ES	ES
RL1075	*ccrM*	Adenine-specific methyltransferase	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL1116		conserved hypothetical protein	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL1432		conserved hypothetical protein	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL1445	*pleC*	putative two-component sensor/regulator; histidine kinase	GD	ES	GD	ES	GD	ES	GD	ES	ES
RL1457		putative electron transfer flavoprotein-ubiquinone oxidoreductase	GD	GD	ES	ES	GD	GD	GD	ES	GD
RL1502	*mosA*	putative rhizopine biosynthesis/dihydropicolinate synthase	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL1510	*sipS*	putative signal peptidase	GD	ES	GD	GD	GD	GD	GD	ES	GD
RL1511	*rnc*	putative ribonuclease III	GD	ES	GD	GD	GD	GD	GD	ES	GD
RL1512		putative GTP and RNA-binding cell cycle regulator	GD	ES	GD	GD	GD	GD	GD	ES	GD
RL1543	*cysS*	putative cysteinyl-tRNA synthetase	ES	ES	ES	ES	GD	GD	GD	ES	ES
RL1551	*dnaC*	putative DNA helicase	GD	ES	ES	ES	ES	GD	GD	ES	ES

RL1557	*fabD*	putative malonyl CoA-acyl carrier protein transacylase	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL1558	*fabG1*	putative 3-oxoacyl-[acyl-carrier-protein] reductase	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL1559	*acpP*	putative acyl carrier protein (fatty acid synthase acyl carrier protein)	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL1566		conserved hypothetical protein	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL1567		putative outer membrane protein involved in envelope biosynthesis	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL1568		putative transmembrane protein	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL1569		conserved hypothetical protein	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL1605	*aspS*	putative aspartyl-tRNA synthetase	GD	ES	ES	GD	GD	ES	GD	ES	ES
RL1609	*parC*	putative topoisomerase IV subunit	GD	ES	ES	GD	ES	GD	GD	ES	ES
RL1616	*hemB*	putative delta-aminolevulinic acid dehydratase	GD	GD	GD	ES	ES	ES	ES	ES	GD
RL1620	*glyA*	putative serine hydroxymethyltransferase	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL1638	*plsX*	putative fatty acid/phospholipid synthesis cluster protein	GD	ES	ES	GD	GD	GD	GD	GD	GD
RL1639	*fabH*	putative 3-oxoacyl-[acyl-carrier-protein] synthase III	GD	ES	ES	GD	GD	GD	GD	GD	GD

RL1672	*rpsI*	putative 30S ribosomal protein S9	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL1673	*rplM*	putative 50S ribosomal protein L13	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL1689	*clpX*	putative ATP-dependent Clp protease ATP-binding subunit ClpX	GD	ES	ES	GD	GD	GD	GD	ES	GD
RL1690	*lon*	putative ATP-dependent protease	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL1701	*nuoB*	putative NADH-quinone oxidoreductase subunit B	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL1702	*nuoC*	putative NADH-quinone oxidoreductase subunit C	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL1703	*nuoD*	putative NADH-quinone oxidoreductase subunit D	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL1704	*nuoE*	putative NADH-quinone oxidoreductase subunit E	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL1705	*nuoF*	putative NADH-quinone oxidoreductase subunit F	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL1706		conserved hypothetical protein in NADH-ubiquinone oxidoreductase region	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL1707	*nuoG*	putative NADH-quinone oxidoreductase subunit G	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL1708	*nuoH*	putative NADH-quinone oxidoreductase subunit H	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL1709	*nuoI*	putative NADH-quinone oxidoreductase subunit I	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL1710	*nuoJ*	putative NADH-quinone oxidoreductase subunit J	GD	ES	ES	ES	GD	GD	GD	ES	ES

RL1711	*nuoK*	putative transmembrane NADH-ubiquinone oxidoreductase subunit K	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL1712	*nuoL*	putative NADH-quinone oxidoreductase chain L	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL1713	*nuoM*	putative NADH-quinone oxidoreductase subunit M	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL1714	*nuoN*	putative NADH-quinone oxidoreductase chain N	GD	ES	ES	ES	GD	GD	GD	ES	GD
RL1715	*birA*	putative biotin transcriptional regulator bifunctional protein	GD	GD	ES	GD	GD	GD	GD	ES	GD
RL1719	*proS*	putative prolyl-tRNA synthetase	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL1720		putative permease component of ABC transporter Unclass	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL1721		putative ATP-binding component of ABC transporter Unclass	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL1723	*dnaE3*	putative DNA polymerase III alpha subunit	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1735	*topA*	putative DNA topoisomerase I	GD	GD	ES	ES	GD	GD	GD	GD	ES
RL1741		putative ribonuclease	GD	ES	ES	ES	ES	GD	GD	GD	ES
RL1743		putative peptidoglycan transglycosylase penicillin-binding protein 1A	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL1744	*prfB*	putative peptide chain release factor 2 (rf-2)	GD	ES	ES	GD	GD	GD	GD	ES	ES

RL1755	*ppnK*	putative inorganic polyphosphate/ATP-NAD kinase	GD	GD	ES	GD	ES	GD	ES	GD	ES
RL1757	*tufB1*	putative elongation factor Tu (EF-Tu)	GD	ES	GD	ES	GD	GD	GD	ES	ES
RL1759	*secE*	putative transmembrane translocase SecE	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL1760	*nusG*	putative transcription antitermination protein	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL1761	*rplK*	putative 50S ribosomal protein L11	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL1762	*rplA*	putative 50S ribosomal protein L1	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL1763		hypothetical protein	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL1764	*rplJ*	putative 50S ribosomal protein L10	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL1765	*rplL*	putative 50S ribosomal protein L7/L12 (L8)	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL1766	*rpoB*	putative DNA-directed RNA polymerase beta chain	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL1767	*rpoC*	putative DNA-directed RNA polymerase beta' chain	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL1769	*rpsL*	putative 30S ribosomal protein S12	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1770	*rpsG*	putative 30S ribosomal protein S7	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1771	*fus*	putative translation elongation factor G, EF-G	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1772	*tufB2*	putative elongation factor Tu (EF-Tu)	GD	ES	ES	ES	ES	GD	GD	ES	ES

RL1773	*rpsJ*	putative 30S ribosomal protein S10	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1774	*rplC*	putative 50S ribosomal protein L3	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1775	*rplD*	putative 50S ribosomal protein L4	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1776	*rplW*	putative 50S ribosomal protein L23	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1777	*rplB*	putative 50S ribosomal protein L2	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1778	*rpsS*	putative 30S ribosomal protein S19	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1779	*rplV*	putative 50S ribosomal protein L22	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1780	*rpsC*	putative 30S ribosomal protein S3	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1781	*rplP*	putative 50S ribosomal protein L16	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1782	*rpmC*	putative 50S ribosomal protein L29	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1783	*rpsQ*	putative 30S ribosomal protein S17	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1784	*rplN*	putative 50S ribosomal protein L14	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1785	*rplX*	putative 50S ribosomal protein L24	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1786	*rplE*	putative 50S ribosomal protein L5	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1787	*rpsN*	putative 30S ribosomal protein S14	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1788	*rpsH*	putative 30S ribosomal protein S8	GD	ES	ES	ES	ES	GD	GD	ES	ES

RL1789	*rplF*	putative 50S ribosomal protein L6	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1790	*rplR*	putative 50S ribosomal protein L18	GD	ES	ES	ES	ES	GD	GD	ES	GD
RL1791	*rpsE*	putative 30S ribosomal protein S5	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1792	*rpmD*	putative 50S ribosomal protein L30	GD	GD	ES	ES	ES	GD	GD	ES	ES
RL1793	*rplO*	putative 50S ribosomal protein L15	GD	GD	ES	ES	ES	GD	GD	ES	ES
RL1794	*secY*	putative preprotein translocase SecY	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1795	*adk1*	putative adenylate kinase	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL1796	*rpsM*	putative 30S ribosomal protein S13 (BS14)	GD	ES	ES	ES	ES	GD	GD	ES	GD
RL1797	*rpsK*	putative 30S ribosomal protein S11	GD	ES	ES	ES	ES	GD	GD	ES	GD
RL1798	*rpoA*	putative DNA-directed RNA polymerase alpha chain	GD	ES	ES	ES	ES	GD	GD	ES	GD
RL1799	*rplQ*	putative 50S ribosomal protein L17	GD	ES	ES	ES	ES	GD	GD	ES	GD
RL2034		conserved hypothetical protein	GD	ES	GD	GD	GD	GD	GD	ES	ES
RL2035	*valS*	putative valyl-tRNA synthetase	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL2037	*exoR*	Sel1 domain-containing protein repeat-containing protein, negative regulator of exopolysaccharide production	ES	ES	GD	GD	GD	GD	GD	GD	ES

RL2041	*argS*	putative arginyl-tRNA synthetase	GD	GD	GD	ES	GD	GD	GD	ES	ES
RL2042		conserved hypothetical protein	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL2046	*tatA*	putative Sec-independent protein translocase protein	GD	GD	GD	GD	GD	GD	GD	ES	ES
RL2047	*tatB*	putative Sec-independent protein translocase protein	GD	GD	GD	GD	GD	GD	GD	ES	ES
RL2048	*tatC*	putative Sec-independent protein translocase protein	GD	GD	GD	GD	GD	GD	GD	ES	ES
RL2049	*serS*	putative seryl-tRNA synthetase	ES	ES	ES	GD	GD	GD	GD	ES	ES
RL2054	*yajC*	putative Sec translocase associated protein	GD	GD	GD	ES	GD	GD	GD	ES	ES
RL2055	*secDF1*	putative protein-export membrane protein SecDF1	GD	GD	GD	ES	GD	GD	GD	ES	ES
RL2056		conserved hypothetical protein	GD	GD	GD	ES	GD	GD	GD	ES	ES
RL2057		putative phytoene synthase	GD	GD	GD	ES	GD	GD	GD	ES	ES
RL2075	*gatC*	putative glutamyl-tRNA(Gln) amidotransferase subunit C	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL2076	*gatA*	putative glutamyl-tRNA(Gln) amidotransferase subunit A	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL2088	*accC1*	putative biotin carboxylase	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL2089	*accB1*	putative biotin carboxyl carrier protein of acetyl-CoA carboxylase	GD	ES	ES	ES	ES	GD	GD	ES	ES

RL2090		putative 3-dehydroquinate dehydratase	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL2096		putative transaminase/aminotransferase	GD	GD	ES	ES	ES	GD	GD	ES	GD
RL2099	*recJ*	putative single-stranded-DNA-specific exonuclease	GD	GD	ES	ES	ES	GD	GD	GD	GD
RL2206		putative salicylate hydroxylase	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL2207		conserved hypothetical protein	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL2221	*rpsB*	putative 30S ribosomal protein S2	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL2222	*tsf*	putative elongation factor Ts (EF-Ts)	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL2223	*pyrH*	putative uridylate kinase	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL2224	*frr*	putative ribosome recycling factor (ribosome releasing factor)	ES	ES	ES	GD	GD	GD	GD	ES	ES
RL2225	*uppS*	putative undecaprenyl pyrophosphate synthetase	ES	ES	ES	GD	GD	GD	GD	ES	ES
RL2226	*cdsA*	putative phosphatidate cytidylyltransferase	ES	ES	ES	GD	GD	GD	GD	ES	ES
RL2228		putative outer membrane protein	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL2229	*lpxD*	putative lipid A biosynthesis UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL2230	*fabZ2*	putative (3R)-hydroxymyristoyl-[acyl	GD	ES	ES	GD	GD	GD	GD	ES	ES

		carrier protein] dehydratase									
RL2231	*lpxA*	putative lipid A biosynthesis acyl-[acyl-carrier-protein]--UDP-N-acetylglucosamine O-acyltransferase	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL2232		conserved hypothetical protein	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL2233	*lpxB*	putative lipid-A-disaccharide synthase	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL2240		putative cell division protein	ES	ES	GD	GD	GD	GD	GD	ES	ES
RL2241	*pdhA*	putative pyruvate dehydrogenase E1 subunit A	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL2242	*pdhB*	putative pyruvate dehydrogenase E1 subunit B	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL2243	*pdhC*	putative dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex subunit E2 (PDC)	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL2244		putative arylesterase	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL2249		putative lipoyl synthase	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL2254	*ispDF*	putative IspD/IspF bi-functional enzyme	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL2282	*ntrX*	two-component sensor/regulator; nitrogen transcriptional regulator NtrX	GD	ES	ES	ES	GD	GD	GD	ES	ES

RL2381		putative bifunctional GlmU protein	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL2401	*gyrA*	putative DNA gyrase subunit A	GD	ES	ES	GD	GD	ES	ES	GD	ES
RL2475	*holB*	putative dna polymerase III, delta subunit	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL2476	*tmk*	putative thymidylate kinase (dtmp kinase)	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL2511	*pyrG*	putative CTP synthase	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL2515	*gyrB2*	putative DNA gyrase subunit B	GD	ES	ES	GD	GD	GD	GD	ES	GD
RL2528	*thrS*	putative threonyl-tRNA synthetase	GD	ES	ES	GD	GD	ES	GD	ES	ES
RL2541		putative penicillin-binding protein/D-alanyl-D-alanine carboxypeptidase	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL2578	*sufS1*	putative aminotransferase involved in iron-sulfur cluster biogenesis	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL2579	*sufD*	putative iron-sulfur cluster assembly protein	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL2580	*sufC*	putative ATP-binding component of ABC transporter Unclass	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL2581		hypothetical protein	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL2582	*sufB*	putative iron-sulfur cluster formation subunit protein	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL2583	*sufS2*	putative cysteine desulfurase, nifS or sufS2?	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL2584		conserved hypothetical protein	GD	ES	ES	ES	GD	GD	GD	ES	ES

RL2624	*rpsD*	putative 30S ribosomal protein S4	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL2636	*alaS*	putative alanyl-tRNA synthetase	ES	GD	ES	ES	ES	GD	GD	ES	ES
RL2640	*cckA*	cell cycle histidine kinase	GD	GD	ES	ES	GD	ES	GD	ES	ES
RL2648		putative 2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL2649	*folB*	putative dihydroneopterin aldolase	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL2650	*folC1*	putative dihydroxypteroate synthase	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL2654		putative transmembrane protein	GD	ES	GD	GD	GD	GD	GD	ES	ES
RL2812	*lpxXL*	putative lipid A acyltransferase	GD	ES	GD	GD	GD	GD	GD	ES	ES
RL2813		putative alcohol dehydrogenase	GD	ES	GD	GD	GD	GD	GD	ES	ES
RL2830	*cobN*	putative cobalamin synthesis protein	GD	GD	ES	ES	ES	ES	GD	ES	ES
RL2831	*cobW*	putative cobalamin synthesis protein	GD	GD	ES	ES	ES	GD	GD	GD	GD
RL2832	*cobP*	putative bifunctional adenosylcobalamin biosynthesis protein	GD	GD	ES	ES	ES	GD	GD	GD	GD
RL3249	*miaA*	putative tRNA delta(2)-isopentenylpyrophosphate transferase	GD	GD	GD	GD	ES	GD	GD	ES	ES
RL3250	*serB*	putative phosphoserine phosphatase	GD	GD	GD	GD	ES	GD	GD	ES	ES

RL3255	*folA*	putative dihydrofolate reductase	GD	GD	ES	GD	GD	GD	GD	ES	GD
RL3256	*thyA*	putative thymidylate synthase	GD	GD	ES	GD	GD	GD	GD	ES	GD
RL3293	*ligA*	putative DNA ligase	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL3296	*comL*	putative competence lipoprotein ComL protein	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL3297	*lpxC*	putative UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL3298	*ftsZ1*	putative cell division protein FtsZ	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL3299	*ftsA*	putative cell division protein FtsA	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL3300	*ftsQ*	putative cell division protein FtsQ	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL3301	*ddlB*	putative D-alanine--D-alanine ligase B	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL3305		putative UDP-N-acetylenolpyruvoylglucosamine reductase	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL3306	*murC*	putative UDP-N-acetylmuramate--L-alanine ligase	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL3307	*murG*	putative UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-	GD	ES	ES	GD	GD	GD	GD	ES	ES

		acetylglucosamine transferase									
RL3308	*ftsW*	putative cell division protein FtsW	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL3309	*murD*	putative UDP-N-acetylmuramoylalanine--D-glutamate ligase	ES	ES	ES	GD	GD	GD	GD	ES	ES
RL3310	*mraY*	putative phospho-N-acetylmuramoyl-pentapeptide-transferase	ES	ES	ES	GD	GD	GD	GD	ES	ES
RL3311	*murF*	putative UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase	ES	ES	ES	GD	GD	GD	GD	ES	ES
RL3312	*murE*	putative UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL3313		putative penicillin binding protein	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL3314		conserved hypothetical protein	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL3315	*mraW*	putative S-adenosyl-methyltransferase	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL3316	*mraZ*	putative MraZ protein	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL3331	*metF*	putative 5,10-methylenetetrahydrofolate reductase	GD	GD	ES	GD	GD	ES	GD	ES	GD
RL3332		putative methyltransferase transcriptional regulator protein, ArsR family	GD	GD	ES	GD	GD	ES	GD	ES	GD

RL3362	*metH*	putative methionine synthase	GD	GD	ES	ES	ES	GD	GD	ES	ES
RL3408	*dnaG*	putative DNA primase	GD	ES	ES	GD	GD	ES	ES	ES	ES
RL3411	*carA*	putative carbamoyl-phosphate synthase small chain	GD	GD	ES	ES	ES	GD	GD	ES	ES
RL3419	*carB*	putative carbamoyl-phosphate synthase large chain	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL3428	*trxB*	putative thioredoxin reductase	GD	ES	ES	GD	GD	ES	GD	ES	GD
RL3463	*lgt*	putative prolipoprotein diacylglyceryl transferase	GD	GD	GD	GD	GD	GD	GD	GD	GD
RL3464		conserved hypothetical protein	GD	GD	GD	GD	GD	GD	GD	GD	GD
RL3494	*hemF*	putative coproporphyrinogen III oxidase, aerobic	GD	GD	ES	GD	GD	ES	ES	ES	ES
RL3496	*papS*	putative poly(A) polymerase	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL3497		putative MutT/Nudix family protein (phosphohydrolases)	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL3553		putative engA GTP-binding protein	GD	GD	ES	ES	GD	GD	GD	ES	GD
RL3554		putative transmembrane protein	GD	GD	ES	ES	GD	GD	GD	ES	GD
RL3643	*pssS*	putative glycosyl transferase	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL3644	*pssR*	putative polysaccharide acetyltransferase protein	GD	GD	ES	GD	GD	GD	GD	ES	ES

RL3645	*pssM*	putative succinoglycan biosynthesis protein	GD	GD	ES	GD	GD	ES	GD	ES	ES
RL3646	*pssL*	putative transmembrane polysaccharide synthesis protein	GD	GD	ES	GD	GD	ES	GD	ES	ES
RL3647	*pssK*	putative polysaccharide polymerisation protein	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL3652	*pssF*	putative glycosyl transferase	GD	GD	GD	GD	GD	GD	GD	ES	ES
RL3653	*pssC*	putative succinoglycan biosynthesis protein	GD	GD	GD	GD	GD	GD	GD	ES	ES
RL3664	*pssN*	putative capsule polysaccharide export protein	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL3706		putative two-component sensor/regulator; histidine kinase, ECF15 gene organisation (histidine kinase, response regulator, asf, ecf)	GD	ES	GD	ES	ES	GD	GD	ES	GD
RL3758	*ctrA*	putative two-component sensor/regulator; transcriptional regulator; cell cycle master regulator	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL3762		conserved hypothetical exported protein	GD	ES	ES	GD	GD	GD	GD	ES	GD
RL3767		putative transmembrane protein	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL3876	*nodT2*	putative exported nodulation protein, TolC homologue	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL3948A		conserved hypothetical protein	GD	ES	ES	ES	ES	GD	GD	ES	ES

RL3957		putative tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase	GD	ES	ES	ES	ES	ES	GD	ES	ES
RL3961	*serC*	putative phosphoserine aminotransferase	GD	GD	ES	ES	ES	ES	GD	ES	ES
RL3965	*hflB*	putative cell division protein FtsH	GD	ES	GD	GD	GD	GD	GD	ES	ES
RL3966	*tilS*	putative tRNA(Ile)-lysidine synthase	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL3967		conserved hypothetical protein	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL3969	*tolB*	putative uptake system component TolB	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL3971	*tolA*	putative uptake system component TolA (outer membrane protein)	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL3972	*exbD*	putative biopolymer transport protein	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL3973	*tolQ*	putative uptake system component TolQ	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL4039		putative polar organelle development protein	GD	ES	GD	ES	GD	GD	GD	ES	ES
RL4060	*pykA*	putative pyruvate kinase	GD	GD	ES	ES	GD	GD	ES	ES	ES
RL4061		conserved hypothetical protein	GD	GD	ES	ES	GD	GD	ES	ES	ES
RL4118	*pgm*	putative phosphoglucomutase	GD	GD	GD	GD	ES	GD	GD	ES	ES
RL4130	*rpsU2*	putative 30S ribosomal protein S21	GD	GD	GD	GD	GD	GD	GD	ES	ES
RL4131		putative TPR repeat family protein	GD	GD	GD	GD	GD	GD	GD	ES	ES

RL4156		putative lysine/ornithine decarboxylase	GD	GD	GD	ES	ES	ES	ES	ES	ES
RL4184	*gltX*	putative glutamyl-tRNA synthetase	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL4185	*lysS*	putative lysyl-tRNA synthetase	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL4278		putative transmembrane peptidase family protein	GD	GD	ES	GD	GD	GD	GD	GD	ES
RL4280		conserved hypothetical protein	GD	ES	GD	GD	GD	GD	GD	ES	ES
RL4282	*prfA*	putative peptide chain release factor 1	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL4284	*aspC*	putative aspartokinase	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL4286	*ubiG*	putative 3-demethylubiquinone-9 3-methyltransferase	GD	GD	ES	GD	GD	ES	GD	ES	ES
RL4298	*secA*	putative preprotein translocase subunit SecA	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL4324		conserved hypothetical exported protein	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL4325	*lysA*	putative diaminopimelate decarboxylase	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL4352	*aroB*	putative 3-dehydroquinate synthase	GD	GD	ES	ES	ES	GD	GD	ES	GD
RL4353	*aroK*	putative shikimate kinase I	GD	GD	ES	ES	ES	GD	GD	ES	GD
RL4405	*atpC*	putative ATP synthase epsilon chain	GD	ES	ES	ES	ES	ES	GD	ES	ES
RL4407	*atpD*	putative ATP synthase beta chain	GD	ES	ES	ES	ES	ES	GD	ES	ES
RL4408	*atpG*	putative ATP synthase gamma chain	GD	ES	ES	ES	ES	ES	GD	ES	ES

RL4409	*atpA*	putative ATP synthase alpha chain	GD	ES	ES	ES	ES	ES	GD	ES	ES
RL4410	*atpH*	putative ATP synthase delta chain	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL4412	*priA*	putative primosomal protein N' (replication factor y)	GD	GD	ES	GD	GD	GD	GD	ES	ES
RL4429	*lpdA*	putative dihydrolipoamide dehydrogenase	GD	GD	GD	GD	GD	GD	GD	GD	GD
RL4433	*sucB*	putative dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase	GD	GD	ES	GD	GD	GD	ES	ES	ES
RL4434		conserved hypothetical protein	GD	GD	ES	GD	GD	GD	ES	ES	ES
RL4435	*sucA*	putative 2-oxoglutarate dehydrogenase E1 component	GD	GD	ES	GD	GD	ES	ES	ES	ES
RL4436	*sucD*	putative succinyl-CoA synthetase alpha chain	GD	GD	GD	GD	GD	ES	ES	ES	ES
RL4437		conserved hypothetical protein	GD	GD	GD	GD	GD	ES	ES	ES	ES
RL4438	*sucC*	putative succinyl-CoA synthetase beta chain	GD	GD	GD	GD	GD	ES	ES	ES	ES
RL4439	*mdh*	putative malate dehydrogenase	GD	GD	GD	GD	GD	GD	GD	ES	ES
RL4441		putative outer membrane lipoprotein omp19	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL4443	*sdhB*	putative succinate dehydrogenase iron-sulfur protein	GD	GD	GD	GD	GD	GD	GD	ES	ES

RL4444	*sdhA*	putative succinate dehydrogenase flavoprotein subunit	GD	GD	GD	GD	GD	GD	GD	ES	ES
RL4494	*gcp*	putative O-sialoglycoprotein endopeptidase	GD	GD	ES	ES	GD	GD	GD	GD	ES
RL4495	*hemC*	putative porphobilinogen deaminase	GD	GD	ES	ES	GD	GD	GD	GD	ES
RL4496		putative uroporphyrinogen-III synthase protein	GD	GD	ES	ES	GD	GD	GD	GD	ES
RL4536	*acnA*	putative aconitate hydratase	GD	ES	ES	ES	ES	GD	GD	ES	ES
RL4543	*ftsY*	putative cell division protein FtsY	GD	ES	GD	ES	ES	ES	ES	ES	ES
RL4544		putative methylase	GD	GD	GD	ES	ES	ES	ES	ES	ES
RL4545	*dapF*	putative diaminopimelate epimerase	GD	GD	GD	ES	ES	ES	ES	ES	ES
RL4547	*ffh*	putative signal recognition particle protein (fifty-four homolog)	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL4548		putative chorismate mutase	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL4549		putative 30S ribosomal protein S16	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL4550	*rimM*	putative 16S rRNA processing protein	GD	GD	GD	ES	GD	GD	GD	ES	ES
RL4551	*trmD*	putative tRNA (guanine-N(1)-)-methyltransferase	GD	GD	GD	ES	GD	GD	GD	ES	ES
RL4563		putative transmembrane DNA translocase	ES	ES	ES	ES	ES	GD	GD	ES	ES

RL4615		putative CarD family transcriptional regulator (activator)	GD	GD	GD	GD	GD	GD	GD	ES	ES
RL4616	*fdxA*	putative ferredoxin II (FdII)	GD	GD	GD	GD	GD	GD	GD	ES	ES
RL4640	*ndvA*	putative ATP-binding:permease (ABC:IMP) component of ABC transporter Export beta-(1-->2) glucan export	GD	ES	GD	ES	ES	GD	GD	ES	GD
RL4644	*ndvB*	putative transmembrane beta (1-->2) glucan biosynthesis protein	GD	GD	GD	ES	ES	GD	GD	ES	ES
RL4681		putative GTP-binding protein	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL4682	*proB*	putative glutamate 5-kinase	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL4683	*proA*	putative gamma-glutamyl phosphate reductase	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL4684	*nadD*	putative nicotinate-nucleotide adenylyltransferase	GD	GD	ES	ES	GD	GD	GD	ES	ES
RL4733		conserved hypothetical protein	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL4734		conserved hypothetical protein	GD	ES	ES	GD	GD	GD	GD	ES	ES
RL4735	*parB*	putative chromosome partitioning protein ParB	GD	ES	ES	ES	GD	GD	GD	ES	ES
RL4736	*parA1*	putative chromosome partitioning protein ParA	GD	ES	ES	ES	GD	GD	GD	ES	ES

[a] In; input, TY; TY [2], V; VMM [3], G; UMS/glucose/NH_4, S; UMS/succinate/NH_4, Rh; rhizosphere, Ro; root attachment, No; nodulation, Re; nodule recovery

[b] ES; growth-essential, GD; growth-defective.

Table 8. INSeq-identified[a] Rlv3841 genes, mutation of which reduces fitness in the rhizosphere, either as a specific effect (rhizosphere specific) or carried over subsequent stages of colonisation too (rhizosphere progressive).

			INSeq classification[b]				
Locus	Gene	Description	In	Rh	Ro	No	Re
Rhizosphere-specific							
pRL100156	*fdxB1*	putative ferredoxin	NE	GD	NE	NE	NE
pRL100241	*gtsR*	putative GntR family transcriptional regulator	NE	GD	NE	NE	NE
pRL110441	*thiD*	putative phosphomethylpyrimidine kinase	NE	GD	NE	NE	NE
pRL110442	*thiE*	putative thiamine-phosphate pyrophosphorylase	NE	GD	NE	NE	NE
pRL110570		hypothetical exported protein	NE	GD	NE	NE	NE
pRL110571		conserved hypothetical protein	NE	GD	NE	NE	NE
pRL120378		putative protein-L-isoaspartate O-methyltransferase	NE	ES	NE	NE	NE
pRL70100		hypothetical protein	NE	GD	NE	NE	NE
pRL70101		conserved hypothetical protein, possible fusion protein	NE	GD	NE	NE	NE
pRL80011		conserved hypothetical protein	NE	GD	NE	NE	NE
pRL90066		putative transmembrane protein	NE	ES	NE	NE	NE
pRL90068		putative transmembrane protein	NE	GD	NE	NE	NE
pRL90093		conserved hypothetical protein	NE	GD	NE	NE	NE
pRL90110		putative ATP-binding component of ABC transporter CUT1	NE	GD	NE	NE	NE
pRL90136		putative glycosyl transferase	NE	GD	NE	NE	NE
pRL90277		putative ATP-binding :permease (ABC:IMP) component of ABC	NE	ES	NE	NE	NE

		transporter Export cytochrome bd-related					
RL0182	*glk*	putative glucokinase	NE	GD	NE	NE	NE
RL0183		putative methylglyoxal synthase	NE	GD	NE	NE	NE
RL0417	*ihfB*	putative integration host factor beta subunit	NE	ES	NE	NE	NE
RL0438		putative lysine decarboxylase family protein	NE	GD	NE	NE	NE
RL0439		putative exported protein with EF-hand calcium-binding domain	NE	GD	NE	NE	NE
RL0482A		putative short-chain dehydrogenase/oxidoreductase	NE	GD	NE	NE	NE
RL1020		putative invasion associated protein	NE	GD	NE	NE	NE
RL1106	*pspA*	putative (phage shock protein A) PspA family regulator by protein-protein interactions	NE	GD	NE	NE	NE
RL1118		hypothetical protein	NE	GD	NE	NE	NE
RL1370	*msrB1*	putative methionine sulfoxide reductase	NE	GD	NE	NE	NE
RL1505		putative polymerase	NE	GD	NE	NE	NE
RL1652		conserved hypothetical protein	NE	GD	NE	NE	NE
RL1653		conserved hypothetical protein	NE	GD	NE	NE	NE
RL1716		putative beta-lactamase family protein	NE	GD	NE	NE	NE
RL1965	*aldR*	putative AsnC family transcriptional regulator	NE	GD	NE	NE	NE
RL1966	*aldA*	putative alanine dehydrogenase	NE	GD	NE	NE	NE
RL1985	*phr*	putative deoxyribodipyrimidine photo-lyase	NE	ES	NE	NE	NE
RL2045	*scpB*	putative chromosome segregation and condensation protein B	NE	GD	NE	NE	NE

RL2062	*tig*	putative chaperone trigger factor	NE	GD	NE	NE	NE
RL2389		conserved hypothetical exported protein	NE	GD	NE	NE	NE
RL2394		putative carbohydrate kinase protein	NE	GD	NE	NE	NE
RL2474		putative transmembrane protein	NE	GD	NE	NE	NE
RL2645		conserved hypothetical protein	NE	GD	NE	NE	NE
RL2646		putative transmembrane protein	NE	GD	NE	NE	NE
RL2647		conserved hypothetical exported protein	NE	GD	NE	NE	NE
RL2964	*cspA3*	putative cold shock protein CspA	NE	GD	NE	NE	NE
RL3011		putative transmembrane transporter protein	NE	GD	NE	NE	NE
RL3181		conserved hypothetical protein	NE	GD	NE	NE	NE
RL3334	*rnsA*	putative ribonuclease I	NE	GD	NE	NE	NE
RL3482		putative MaoC family protein	NE	GD	NE	NE	NE
RL3483		putative adenine phosphoribosyltransferase	NE	GD	NE	NE	NE
RL3556		conserved hypothetical protein	NE	GD	NE	NE	NE
RL3663		putative polysaccharide transport outer membrane protein	NE	GD	NE	NE	NE
RL3669		conserved hypothetical protein	NE	GD	NE	NE	NE
RL3808		putative ROK family transcriptional regulator	NE	GD	NE	NE	NE
RL3877		putative LacI family transcriptional regulator (repressor)	NE	GD	NE	NE	NE
RL4074		conserved hypothetical protein	NE	GD	NE	NE	NE
RL4332		putative transmembrane protein	NE	GD	NE	NE	NE
RL4584		putative SBP of ABC transporter Unclass ferric cations transporter	NE	GD	NE	NE	NE

Rhizosphere-progressive							
pRL100161	*nifD*	nitrogenase molybdenum-iron protein alpha chain NifD	NE	GD	GD	GD	ES
pRL100187	*nodC*	N-acetylglucosaminyltransferase NodC	NE	GD	GD	GD	GD
pRL100386		putative von Willebrand factor type A	NE	GD	GD	GD	ES
pRL110004		conserved hypothetical protein	NE	GD	GD	GD	GD
pRL110046		putative FNR/CRP family transcriptional regulator	NE	GD	GD	ES	ES
pRL110632	*cobF*	precorrin-6a synthase [deacetylating]	NE	GD	GD	ES	GD
pRL120618		putative LacI family transcriptional regulator (repressor)	NE	GD	GD	ES	ES
pRL70055		hypothetical protein	NE	GD	GD	GD	ES
pRL70056		hypothetical protein	NE	GD	GD	GD	ES
pRL90053		putative O-antigen ligase	NE	GD	GD	ES	ES
pRL90141		hypothetical protein	NE	GD	GD	GD	GD
pRL90143		putative transposase	NE	GD	GD	GD	GD
RL0002		putative septum formation protein	NE	GD	ES	ES	ES
RL0020		putative N-(5'-phosphoribosyl)anthranilate	NE	GD	GD	ES	ES
RL0021	*trpB*	putative tryptophan synthase beta subunit	NE	GD	GD	ES	ES
RL0022	*trpA*	putative tryptophan synthase alpha subunit	NE	GD	GD	ES	ES
RL0025		putative thioredoxin	NE	GD	GD	GD	GD
RL0026		putative ATP-dependent UvrD family DNA helicase	NE	GD	GD	GD	GD
RL0027		conserved hypothetical protein	NE	GD	GD	GD	ES

RL0028		putative nucleotidyltransferase protein	NE	GD	GD	GD	ES
RL0031		putative S-adenosyl-L-homocysteine hydrolase	NE	GD	GD	ES	ES
RL0034		putative HPr kinase	NE	ES	GD	ES	ES
RL0052		conserved hypothetical protein	NE	GD	ES	ES	ES
RL0108	*aroA1*	putative 3-phosphoshikimate 1-carboxyvinyltransferase (5-enolpyruvylshikimate-3-phosphate synthase) (EPSP synthase)	NE	GD	GD	ES	ES
RL0139		putative P-protein [includes: chorismate mutase and prephenate dehydrogenase]	NE	GD	ES	ES	ES
RL0179	*gpmA*	putative 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase (Phosphoglyceromutase) (PGAM) (BPG-dependent PGAM) (dPGM)	NE	GD	GD	ES	ES
RL0254	*lepA*	putative GTP-binding protein	NE	GD	GD	ES	ES
RL0327		putative glutathione S-transferase	NE	GD	GD	ES	ES
RL0346	*dut*	putative deoxyuridine 5'triphosphate nucleotidohydrolase	NE	GD	GD	GD	GD
RL0347	*prfC*	putative peptide chain release factor (RF-3)	NE	GD	GD	GD	GD
RL0377	*hemN2*	putative oxygen-independent coprophorphyrinogen III oxidase	NE	GD	GD	GD	ES
RL0378		putative HAM1 family protein	NE	GD	GD	GD	ES
RL0379		putative glyoxalase	NE	GD	GD	GD	ES
RL0380	*rph*	putative ribonuclease	NE	GD	GD	GD	GD
RL0424		putative sigma-54 modulation protein (has ribosome-associated inhibitory protein (raiA) domain)	NE	GD	GD	GD	GD

RL0425	*ptsN*	putative nitrogen regulatory IIA protein	NE	GD	GD	GD	GD
RL0430	*def1*	putative peptide deformylase (polypeptide deformylase)	NE	GD	GD	GD	ES
RL0431		putative plasmid stability protein	NE	GD	GD	GD	ES
RL0432		putative plasmid stability protein	NE	GD	GD	GD	ES
RL0433	*fmt*	putative methionyl-tRNA formyltransferase	NE	GD	GD	GD	ES
RL0434	*truA*	putative tRNA pseudouridine synthase A	NE	GD	GD	GD	ES
RL0549	*gabT2*	putative acetylornithine aminotransferase/succinylornithine transaminase	NE	GD	GD	ES	ES
RL0751	*edd*	putative phosphogluconate dehydratase	NE	GD	ES	GD	ES
RL0811		putative glycosyltransferase	NE	GD	GD	ES	ES
RL0811A		putative glycosyltransferase	NE	GD	GD	ES	ES
RL0813		putative UDP-glucose-4-epimerase (putative dTDP-glucose 4,6-dehydratase)	NE	GD	GD	ES	ES
RL0814		conserved hypothetical protein	NE	GD	GD	ES	ES
RL0815		putative acetyltransferase	NE	GD	GD	ES	ES
RL0816		conserved hypothetical protein	NE	GD	GD	ES	ES
RL0818		putative lipopolysaccharide biosynthesis protein	NE	GD	GD	ES	ES
RL0819		putative imidazole glycerol phosphate synthase subunit	NE	GD	GD	ES	ES
RL0820	*hisH2*	putative imidazole glycerol phosphate synthase subunit (igp synthase glutamine amidotransferase subunit)	NE	GD	GD	ES	ES

RL0821		putative O-antigen transporter	NE	GD	GD	ES	ES
RL0822		putative DegT family aminotransferase	NE	GD	GD	ES	ES
RL0823		putative hexapeptide repeat acetyltransferase	NE	GD	GD	ES	ES
RL0824		putative oxidoreductase	NE	GD	GD	ES	ES
RL0825	*gmd*	putative GDP-mannose 4,6-dehydratase (GDP-d-mannose dehydratase)	NE	GD	GD	ES	ES
RL0826	*fcl*	putative GDP-L-fucose synthetase	NE	GD	GD	ES	ES
RL0855	*gshA*	putative gamma-glutamylcysteine synthetase precursor	NE	GD	GD	ES	ES
RL0885		putative hydrolase	NE	GD	GD	GD	ES
RL0891		putative nitrogen fixation symbiosis related protein	NE	GD	GD	GD	GD
RL0892		putative ribosomal large subunit pseudouridine synthase B	NE	GD	GD	GD	GD
RL0920		putative ATP-binding mrp family protein	NE	GD	GD	GD	GD
RL0921		putative cationic transport protein, CorA family	NE	GD	GD	GD	GD
RL0957		conserved hypothetical protein	NE	GD	ES	ES	ES
RL1002	*bioY*	putative transmembrane biotin biosynthesis protein	NE	GD	GD	ES	ES
RL1003		putative permease component of ABC transporter Unclass	NE	GD	GD	ES	ES
RL1004		putative ATP-binding component of ABC transporter Unclass	NE	GD	GD	ES	ES
RL1021	*ctaC1*	putative cytochrome c oxidase polypeptide II precursor (cytochrome aa3 subunit 2)	NE	GD	GD	ES	ES

RL1022	*ctaD1*	putative cytochrome c oxidase polypeptide I (cytochrome aa3 subunit 1)	NE	GD	GD	ES	ES
RL1023	*ctaB*	putative protoheme IX farnesyltransferase (heme o synthase)	NE	GD	GD	ES	ES
RL1024	*coxF*	putative cox locus protein	NE	GD	GD	ES	ES
RL1379	*rosR*	putative MucR/RosR family transcriptional regulator involved in nodulation competitiveness RosR	NE	GD	GD	GD	ES
RL1434	*feuQ*	putative two-component sensor/regulator; histidine kinase	NE	GD	GD	ES	ES
RL1503		putative RNA binding protein	NE	GD	GD	ES	ES
RL1506	*relA*	putative stringent response protein	NE	GD	GD	ES	ES
RL1523		conserved hypothetical protein	NE	GD	GD	GD	GD
RL1524		conserved hypothetical protein	NE	GD	GD	GD	GD
RL1525		putative aminomethyltransferase	NE	GD	GD	GD	GD
RL1526		conserved hypothetical protein	NE	GD	GD	GD	GD
RL1527		conserved hypothetical protein	NE	GD	GD	GD	GD
RL1532		putative phosphatidylserine synthase	NE	GD	GD	GD	ES
RL1552	*rplI*	putative 50S ribosomal protein L9	NE	GD	GD	GD	GD
RL1553		putative transmembrane protein	NE	GD	GD	GD	GD
RL1571	*pepA*	putative aminopeptidase	NE	GD	GD	ES	ES
RL1572		putative DNA polymerase III subunit	NE	GD	GD	ES	ES
RL1597		putative transmembrane protein	NE	GD	GD	ES	ES
RL1621	*ribG*	putative riboflavin biosynthesis protein	NE	GD	GD	ES	ES
RL1622	*ribC*	putative riboflavin biosynthesis protein	NE	GD	GD	ES	ES

RL1640	*ihfA*	putative integration host factor alpha-subunit	NE	GD	GD	GD	GD
RL1737		putative transmembrane protein	NE	GD	GD	GD	GD
RL1738	*pyrC2*	putative dihydroorotase	NE	GD	GD	GD	GD
RL1739	*pyrB*	putative aspartate carbamoyltransferase	NE	GD	GD	GD	GD
RL1742		putative N-acetylmuramoyl-L-alanine amidase	NE	GD	GD	GD	ES
RL1803	*ilvD1*	putative dihydroxy-acid dehydratase	NE	GD	ES	ES	ES
RL1946		hypothetical protein	NE	GD	GD	ES	ES
RL2052		putative peptidase	NE	GD	GD	ES	ES
RL2097	*hom1*	putative homoserine dehydrogenase	NE	GD	GD	ES	GD
RL2209	*cysE1*	putative serine acetyltransferase	NE	GD	GD	GD	ES
RL2234	*gltA*	putative citrate synthase	NE	GD	GD	ES	ES
RL2239	*eno*	putative enolase	NE	GD	GD	ES	ES
RL2247		putative transmembrane transglycosylase-associated protein, integral membrane protein	NE	GD	GD	ES	ES
RL2255	*dus*	putative tRNA-dihydrouridine synthase (nitrogen regulation protein)	NE	GD	GD	GD	ES
RL2256	*ntrB*	putative two-component sensor/regulator; histidine kinase NtrB	NE	GD	GD	GD	ES
RL2257	*ntrC*	two-component sensor/regulator; nitrogen transcriptional regulator NtrC	NE	GD	GD	GD	ES
RL2303	*ccdA*	putative cytochrome c-type biogenesis protein	NE	ES	GD	ES	ES
RL2382	*glmS*	putative glucosamine--fructose-6-phosphate aminotransferase [isomerizing]	NE	GD	GD	ES	ES

RL2492	*ppiD*	putative peptidyl-prolyl cis-trans isomerase D	NE	GD	GD	ES	ES
RL2493	*trpD*	putative anthranilate phosphoribosyltransferase	NE	GD	GD	ES	ES
RL2494	*trpC*	putative indole-3-glycerol phosphate synthase	NE	GD	GD	ES	ES
RL2542		conserved hypothetical protein	NE	GD	GD	ES	ES
RL2555	*lipB*	putative lipoyl transferase	NE	ES	ES	ES	GD
RL2567	*rLuC*	putative ribosomal large subunit pseudouridine synthase C	NE	GD	GD	ES	GD
RL2594	*deaD*	putative cold-shock DEAD-box protein A (ATP-dependent RNA helicase DEAD)	NE	GD	ES	ES	ES
RL2601	*purB*	putative adenylosuccinate lyase	NE	ES	GD	ES	ES
RL2775	*ropA1*	putative outer membrane porin protein RopA	NE	GD	GD	ES	ES
RL2824	*cobA*	putative uroporphyrin-III C-methyltransferase	NE	GD	GD	ES	ES
RL2836	*cobQ*	putative cobyric acid synthase	NE	GD	GD	ES	ES
RL3205	*ilvC*	putative ketol-acid reductoisomerase	NE	GD	GD	GD	ES
RL3206		putative TetR family transcriptional regulator	NE	GD	GD	GD	ES
RL3245	*ilvI*	putative acetolactate synthase isozyme III large subunit	NE	GD	GD	ES	ES
RL3257		conserved hypothetical protein	NE	GD	GD	ES	GD
RL3259		conserved hypothetical protein	NE	GD	GD	ES	ES
RL3260		conserved hypothetical protein	NE	GD	GD	ES	ES
RL3276	*pcrA*	putative ATP-dependent DNA helicase	NE	GD	GD	ES	ES

RL3333		putative ATP-binding:ATP-binding (ABC:ABC) component of ABC transporter Unclass	NE	GD	GD	ES	GD
RL3422	*greA*	putative transcription elongation factor	NE	ES	GD	ES	ES
RL3423	*lpcC*	putative lipopolysaccharide core biosynthesis protein	NE	ES	GD	ES	ES
RL3485	*petB*	putative cytochrome b	NE	GD	GD	ES	ES
RL3500		conserved hypothetical protein	NE	GD	GD	GD	GD
RL3501		putative transmembrane protein	NE	GD	GD	GD	GD
RL3502		putative transmembrane protein	NE	GD	GD	GD	GD
RL3507		putative phosphoesterase	NE	GD	ES	ES	ES
RL3513	*leuA1*	putative 2-isopropylmalate synthase	NE	GD	GD	ES	ES
RL3521	*trpE*	putative anthralinate synthase	NE	GD	GD	ES	ES
RL3557	*bacA*	putative transmembrane transporter protein required for bacteroid development	NE	GD	GD	ES	ES
RL3648	*pssJ*	putative galactosyl transferase	NE	GD	GD	ES	ES
RL3651	*pssG*	putative glycosyl transferase	NE	GD	GD	ES	ES
RL3662	*pssP*	putative succinoglycan biosynthesis transport protein	NE	GD	ES	ES	ES
RL3667	*rkpK*	putative UDP-glucose 6-dehydrogenase	NE	GD	GD	GD	ES
RL3668		putative serine/threonine protein phosphatase	NE	GD	GD	GD	ES
RL3674	*noeK*	putative phosphomannomutase	NE	GD	GD	ES	ES
RL3764		putative transmembrane protein	NE	GD	GD	ES	ES
RL3768	*purA*	putative adenylosuccinate synthetase	NE	GD	GD	ES	ES
RL3987		conserved hypothetical protein	NE	GD	GD	GD	GD

RL3988		conserved hypothetical protein	NE	GD	GD	GD	GD
RL3989	*ruvA*	putative Holliday junction DNA helicase RuvA	NE	GD	GD	GD	GD
RL3990	*ruvB*	putative Holliday junction DNA helicase RuvB	NE	GD	GD	GD	GD
RL4040	*thiE2*	putative thiamine-phosphate pyrophosphorylase	NE	GD	GD	ES	ES
RL4043		conserved hypothetical protein	NE	GD	GD	ES	ES
RL4044	*purE*	putative phosphoribosylaminoimidazole carboxylase catalytic subunit	NE	GD	GD	ES	ES
RL4045	*purK*	putative phosphoribosylaminoimidazole carboxylase ATPase subunit	NE	GD	GD	ES	ES
RL4066		conserved hypothetical protein	NE	GD	GD	ES	ES
RL4281	*hemK*	putative protein methyltransferase	NE	GD	GD	ES	ES
RL4296	*argJ*	putative arginine biosynthesis bifunctional protein	NE	GD	GD	ES	ES
RL4297		putative foldase/peptidyl-prolyl cis-trans isomerase	NE	GD	GD	ES	ES
RL4337	*tyrC*	putative prephenate dehydrogenase	NE	GD	GD	ES	GD
RL4338	*hisC1*	putative histidinol-phosphate aminotransferase	NE	GD	GD	ES	GD
RL4348	*cobT*	putative aerobic cobaltochelatase	NE	GD	GD	ES	ES
RL4349	*cobS*	putative aerobic cobaltochelatase	NE	GD	GD	ES	ES
RL4399		conserved hypothetical protein YKOF super family	NE	GD	GD	ES	ES
RL4400		putative permease component of ABC transporter NitT	NE	GD	GD	ES	ES
RL4401		putative ATP-binding component of ABC transporter NitT	NE	GD	GD	ES	ES

RL4445	*sdhD*	putative succinate dehydrogenase hydrophobic membrane anchor protein	NE	GD	GD	ES	ES
RL4446	*sdhC*	putative succinate dehydrogenase cytochrome b556 subunit	NE	GD	GD	ES	ES
RL4537	*ccmA*	putative ATP-binding component of ABC transporter Export cytochrome c biogenesis	NE	GD	GD	ES	ES
RL4538	*ccmB*	putative permease component of ABC transporter Export cytochrome c binding export protein	NE	GD	GD	ES	ES
RL4539	*cycZ*	putative permease component of ABC transporter Export heme exporter protein c (cytochrome c-type biogenesis protein)	NE	GD	GD	ES	ES
RL4540	*cycY*	putative thiol:disulfide interchange protein CycY precursor (cytochrome c biogenesis protein)	NE	GD	GD	ES	ES
RL4606	*metA*	putative homoserine O-succinyltransferase	NE	GD	GD	ES	ES
RL4631		putative transmembrane transporter protein	NE	GD	GD	GD	GD
RL4692	*ctpA*	putative carboxy-terminal processing protease precursor	NE	GD	GD	ES	ES
RL4707	*leuB*	putative 3-isopropylmalate dehydrogenase	NE	GD	ES	ES	ES
RL4722	*purH*	putative bifunctional purine biosynthesis protein PurH	NE	GD	ES	GD	ES
RL4731		putative racemase	NE	GD	GD	ES	ES
RL4732	*leuS*	putative leucyl-tRNA synthetase	NE	GD	GD	ES	ES

[a] In; input, Rh; rhizosphere, Ro; root attachment, No; nodulation, Re; nodule recovery

[b] ES; growth-essential, GD; growth-defective, NE; growth-neutral.

Table 9. INSeq-identified[a] Rlv3841 genes, mutation of which reduces fitness in root attachment, either as a specific effect (attachment specific) or carried over subsequent stages of colonisation too (attachment progressive).

			INSeq classification[b]				
Locus	Gene	Description	In	Rh	Ro	No	Re
Attachment-specific							
pRL100011		hypothetical protein	NE	NE	GD	NE	NE
pRL100012		conserved hypothetical protein	NE	NE	GD	NE	NE
pRL100026		conserved hypothetical protein	NE	NE	GD	NE	NE
pRL100056		putative aromatic amino acid aminotransferase	NE	NE	GD	NE	NE
pRL100057		putative AsnC family transcriptional regulator	NE	NE	GD	NE	NE
pRL100127		putative ATP-binding component of ABC transporter Unclass	NE	NE	GD	NE	NE
pRL100128		putative ATP-binding component of ABC transporter Unclass	NE	NE	GD	NE	NE
pRL100129		putative permease component of ABC transporter Unclass	NE	NE	GD	NE	NE
pRL100164	*rhiI*	putative enzyme to synthesise (autoinducer) AHL, RhiI	NE	NE	GD	NE	NE
pRL100170	*rhiB*	rhizosphere-induced protein RhiB	NE	NE	GD	NE	NE
pRL100171	*rhiC*	rhizosphere-induced protein RhiC	NE	NE	GD	NE	NE
pRL100172	*rhiR*	putative transcriptional regulatory protein controlling rhi gene expression RhiR	NE	NE	GD	NE	NE
pRL100179	*nodN*	nodulation protein NodN	NE	NE	GD	NE	NE
pRL100309	*cadB*	cadherin domain-containing calcium-binding glycoprotein, involed in cell to cell contact?	NE	NE	GD	NE	NE

pRL100387		putative gluconolactonase precursor	NE	NE	GD	NE	NE
pRL100428		putative permease component of ABC transporter Unclass	NE	NE	GD	NE	NE
pRL100429		putative short-chain dehydrogenase	NE	NE	GD	NE	NE
pRL110142		hypothetical protein	NE	NE	GD	NE	NE
pRL110169		putative SBP of ABC transporter CUT2	NE	NE	GD	NE	NE
pRL110353		conserved hypothetical protein	NE	NE	GD	NE	NE
pRL110543		conserved hypothetical protein	NE	NE	ES	NE	NE
pRL110544	*minE*	putative cell division topological specificity factor	NE	NE	ES	NE	NE
pRL110545	*minD*	putative septum site-determining protein MinD	NE	NE	ES	NE	NE
pRL110560	*soxB2*	putative sarcosine oxidase beta subunit	NE	NE	GD	NE	NE
pRL110561	*soxD2*	putative sarcosine oxidase delta subunit	NE	NE	GD	NE	NE
pRL110564		putative tight adherence protein	NE	NE	ES	NE	NE
pRL120021		hypothetical protein	NE	NE	ES	NE	NE
pRL120058		hypothetical protein	NE	NE	GD	NE	NE
pRL120160		putative DeoR family transcriptional regulator (repressor)	NE	NE	ES	NE	NE
pRL120306		putative ATP-binding component of ABC transporter NitT	NE	NE	GD	NE	NE
pRL120322	*fhuA2*	outer membrane siderophore receptor precursor	NE	NE	GD	NE	NE
pRL120584		putative epimerase	NE	NE	GD	NE	NE
pRL120656		hypothetical protein	NE	NE	ES	NE	NE
pRL120692		putative ROK family transcriptional regulator	NE	NE	GD	NE	NE

pRL70038	*repAp7b*	putative replication protein RepA	NE	NE	GD	NE	NE
pRL80032		putative LysR family transcriptional regulator	NE	NE	GD	NE	NE
pRL90009		conserved hypothetical protein with CBS domain	NE	NE	ES	NE	NE
pRL90149		putative LuxR/GerE family transcriptional regulator, part of two component response regulator?	NE	NE	GD	NE	NE
pRL90204		putative amidase	NE	NE	GD	NE	NE
pRL90243		putative acyl-CoA dehydrogenase	NE	NE	GD	NE	NE
RL0256		putative XRE family (HipB) transcriptional regulator with cupin2 domian	NE	NE	GD	NE	NE
RL0561		putative AraC family transcriptional regulator (activator)	NE	NE	GD	NE	NE
RL0682		putative transmembrane ion efflux system protein	NE	NE	GD	NE	NE
RL0795A		conserved hypothetical protein	NE	NE	GD	NE	NE
RL0799		putative hexapeptide repeat transferase	NE	NE	GD	NE	NE
RL0802		putative deoxygenase	NE	NE	GD	NE	NE
RL0810A		conserved hypothetical protein	NE	NE	GD	NE	NE
RL0929		hypothetical protein	NE	NE	GD	NE	NE
RL0930		putative ribonuclease HII	NE	NE	GD	NE	NE
RL0963		putative transmembrane/surface protein BA14K-like immuno reactive protein	NE	NE	GD	NE	NE
RL1001		conserved hypothetical protein	NE	NE	GD	NE	NE
RL1040		putative LysR family transcriptional regulator	NE	NE	GD	NE	NE

RL1041		putative oxidoreductase/monooxygenase	NE	NE	GD	NE	NE
RL1047		putative ATP-binding component of ABC transporter MZT	NE	NE	GD	NE	NE
RL1048		putative permease component of ABC transporter MZT	NE	NE	GD	NE	NE
RL1084	*smc*	putative structural maintenance of chromosomes protein	NE	NE	GD	NE	NE
RL1085		hypothetical protein	NE	NE	GD	NE	NE
RL1091		conserved hypothetical protein	NE	NE	GD	NE	NE
RL1092		conserved hypothetical protein	NE	NE	GD	NE	NE
RL1093		putative beta-lactamase family protein	NE	NE	GD	NE	NE
RL1101		hypothetical protein	NE	NE	GD	NE	NE
RL1504		conserved hypothetical protein	NE	NE	GD	NE	NE
RL1631	*napC*	putative cytochrome c type nitrate reductase NapC exported to periplasm	NE	NE	GD	NE	NE
RL1937		hypothetical protein	NE	NE	GD	NE	NE
RL1938		putative phosphatase	NE	NE	GD	NE	NE
RL1939		hypothetical protein	NE	NE	GD	NE	NE
RL2109		conserved hypothetical protein	NE	NE	GD	NE	NE
RL2153		conserved hypothetical protein	NE	NE	GD	NE	NE
RL2284	*hfq*	putative host factor protein	NE	NE	ES	NE	NE
RL2304		conserved hypothetical protein	NE	NE	GD	NE	NE
RL2305		hypothetical protein	NE	NE	GD	NE	NE
RL2507		conserved hypothetical exported protein	NE	NE	GD	NE	NE
RL2508	*gltA2*	putative citrate synthase II	NE	NE	GD	NE	NE

RL2662		putative racemase/isomerase	NE	NE	GD	NE	NE
RL2663		putative permease component of ABC transporter Unclass	NE	NE	GD	NE	NE
RL2664	*lpxH*	putative UDP-2,3-diacylglucosamine hydrolase	NE	NE	GD	NE	NE
RL2692		putative DAHP synthetase protein	NE	NE	GD	NE	NE
RL2702		putative rhizobium adhering-like protein	NE	NE	GD	NE	NE
RL2776		putative AsnC family transcriptional regulator	NE	NE	GD	NE	NE
RL2837		putative SBP of ABC transporter QAT? orphan proline/glycine betaine	NE	NE	GD	NE	NE
RL2930		conserved hypothetical protein	NE	NE	GD	NE	NE
RL2931	*mcpN*	putative methyl-accepting chemotaxis protein McpN	NE	NE	GD	NE	NE
RL2932		putative acyl-CoA hydrolase	NE	NE	GD	NE	NE
RL2933		putative transmembrane protein	NE	NE	GD	NE	NE
RL2934		conserved hypothetical exported protein	NE	NE	GD	NE	NE
RL3198		conserved hypothetical protein	NE	NE	GD	NE	NE
RL3199		putative aminotransferase	NE	NE	GD	NE	NE
RL3213		putative tetracycline resistance protein	NE	NE	GD	NE	NE
RL3320		putative signalling and peptidoglycan binding protein	NE	NE	GD	NE	NE
RL3321		putative DnaJ family chaperone	NE	NE	GD	NE	NE
RL3560	*map2*	putative methionine aminopeptidase	NE	NE	GD	NE	NE
RL3561		putative bacterial luciferase family protein	NE	NE	GD	NE	NE
RL3591		putative two-component sensor/regulator; histidine kinase	NE	NE	GD	NE	NE

RL3828		putative FNR/CRP family transcriptional regulator	NE	NE	GD	NE	NE
RL4087		putative LysR family transcriptional regulator	NE	NE	GD	NE	NE
RL4382		putative filamentous hemagglutinin adherence factor precursor	NE	NE	GD	NE	NE
RL4669		hypothetical protein	NE	NE	GD	NE	NE
RL4670		conserved hypothetical protein	NE	NE	GD	NE	NE
RL4671	*rci*	putative shufflon-specific DNA recombinase	NE	NE	GD	NE	NE
RL4689		conserved hypothetical protein	NE	NE	GD	NE	NE
RL4690		conserved hypothetical protein	NE	NE	GD	NE	NE
Attachment-progressive							
pRL100027		putative restriction modification methylase	NE	NE	GD	ES	ES
pRL100053		putative transmembrane protein	NE	NE	ES	GD	GD
RL0133		conserved hypothetical protein	NE	NE	GD	ES	ES
RL0817		putative transmembrane protein	NE	NE	GD	GD	ES
RL0934	*moaB*	putative molybdenum cofactor biosynthesis protein B	NE	NE	GD	ES	ES
RL1032	*rnhA1*	putative ribonuclease HI	NE	NE	GD	ES	ES
RL1600	*ppx*	putative exopolyphosphatase	NE	NE	GD	ES	ES
RL2291		conserved hypothetical protein	NE	NE	GD	ES	ES
RL2307		conserved hypothetical protein	NE	NE	GD	GD	GD
RL2828		putative XRE family (HipB) family transcriptional regulator	NE	NE	GD	ES	ES
RL3460	*proC1*	putative pyrroline-5-carboxylate reductase	NE	NE	GD	GD	ES
RL3677	*lspL*	putative UDP-glucuronate 5'-epimerase	NE	NE	GD	ES	ES

RL4183	putative transmembrane protein	NE	NE	GD	ES	ES
RL4632	putative DeoR family transcriptional regulator (repressor)	NE	NE	GD	GD	GD

[a] In; input, Rh; rhizosphere, Ro; root attachment, No; nodulation, Re; nodule recovery

[b] ES; growth-essential, GD; growth-defective, NE; growth-neutral.

Table 10. Genomic context of the INSeq-identified transcriptional regulators required in the root attached library. The eleven transcriptional regulators are highlighted in yellow, and shown in the context of their upstream and downstream regions, and the HMM phenotypes[a] of these regions across the five different *in planta* INSeq libraries[b].

			INSeq library classification				
Locus	Gene	Description	In	Rh	Ro	No	Re
pRL80022		conserved hypothetical protein	NE	NE	NE	NE	NE
pRL80023	*coxM*	putative carbon monoxide dehydrogenase subunit M	NE	NE	NE	NE	NE
pRL80024	*coxS*	putative iron-sulphur cluster carbon monoxide dehydrogenase subunit S	NE	NE	NE	NE	NE
pRL80025	*coxL*	putative dehydrogenase/reductase subunit L	NE	NE	NE	NE	NE
pRL80026	*livJ*	putative SBP of ABC transporter HAAT	NE	NE	NE	NE	NE
pRL80027	*livM*	putative permease component of ABC transporter HAAT	NE	NE	NE	NE	NE
pRL80028	*livH*	putative permease component of ABC transporter HAAT	NE	NE	NE	NE	NE
pRL80029	*livG*	putative ATP-binding component of ABC transporter HAAT	NE	NE	NE	NE	NE
pRL80030	*livF*	putative ATP-binding component of ABC transporter HAAT	NE	NE	NE	NE	NE
pRL80031	*mcpS*	putative methyl-accepting chemotaxis chemoreceptor	NE	NE	NE	NE	NE
pRL80032		putative LysR family transcriptional regulator	NE	NE	GD	NE	NE
pRL80033		conserved hypothetical protein	NE	NE	NE	NE	NE

pRL80034	*moaA1*	putative molybdenum cofactor biosynthesis protein A	NE	NE	NE	NE	NE
pRL80035		putative TetR family transcriptional regulator	NE	NE	NE	NE	NE
pRL80036		conserved hypothetical protein	NE	NE	NE	NE	NE
pRL80037		putative short-chain dehydrogenase	NE	NE	NE	NE	NE
pRL80038		putative exported xanthine dehydrogenase/CoxI family protein	NE	NE	NE	NE	NE
pRL80039		pseudogene	NE	NE	NE	NE	NE
pRL80039A		pseudogene	NE	NE	NE	NE	NE
pRL80040		putative LacI family transcriptional regulator (repressor)	NE	NE	NE	NE	NE
pRL80041	*hisD2*	putative histidinol dehydrogenase	NE	NE	NE	NE	GD
pRL80042	*pntA*	putative NAD(P) transhydrogenase subunit A	NE	NE	NE	NE	NE
pRL90139		conserved hypothetical protein	NE	NE	NE	NE	NE
pRL90140		putative nitrous oxidase accessory protein	NE	NE	NE	NE	NE
pRL90141		hypothetical protein	NE	GD	GD	GD	GD
pRL90143		putative transposase	NE	GD	GD	GD	GD
pRL90144		putative exopolysaccharide biosynthesis-related protein	NE	NE	NE	NE	NE
pRL90146		putative exoplysaccharide biosynthesis UDP-galactose-lipid carrier transferase	NE	NE	NE	NE	NE
pRL90147		putative epimerase	NE	NE	NE	NE	NE
pRL90148		hypothetical protein	NE	NE	NE	NE	NE

pRL90149	putative LuxR/GerE family transcriptional regulator, part of two component response regulator?	NE	NE	GD	NE	NE
pRL90150	putative hexapeptide repeat acetyltransferase	NE	NE	NE	NE	NE
pRL90151	putative GFO/IDH/MocA family oxidoreductase	NE	NE	NE	NE	NE
pRL90152	putative glutamine--scyllo-inositol transaminase	NE	NE	NE	NE	NE
pRL90153	putative hexapeptide repeat acetyltransferase	NE	NE	NE	NE	NE
pRL90154	putative NAD-dependent epimerase/dehydratase	NE	NE	NE	NE	NE
pRL90155	putative DegT/DnrJ/EryC1/StrS family aminotransferase	NE	NE	NE	NE	NE
pRL90156	putative transmembrane polysaccharide biosynthesis protein	NE	NE	NE	NE	NE
pRL90157	hypothetical protein	NE	NE	NE	NE	NE
pRL90158	putative exopolysaccharide production protein	NE	NE	NE	NE	NE
pRL90159	conserved hypothetical protein	NE	NE	NE	NE	NE
pRL100047	putative phage integrase protein	NE	NE	NE	ES	ES
pRL100048	putative phage integrase/tyrosine recombinase	NE	NE	GA	NE	NE
pRL100049	putative integrase/recombinase protein	NE	GA	GA	NE	NE
pRL100050	putative integrase/recombinase protein	NE	NE	NE	NE	NE
pRL100051	hypothetical protein	NE	NE	NE	GD	GD

pRL100052	*cspAp10*	putative cold shock protein CspA	NE	NE	NE	GD	GD
pRL100052A		pseudogene	NE	NE	NE	NE	NE
pRL100053		putative transmembrane protein	NE	NE	ES	GD	GD
pRL100053A		hypothetical protein	NE	NE	NE	NE	NE
pRL100054		pseudogene, putative glutamate dehydrogenase	NE	NE	NE	NE	NE
pRL100055		putative acetolactate synthase subunit	NE	GD	GA	NE	ES
pRL100056		putative aromatic amino acid aminotransferase	NE	NE	GD	NE	NE
pRL100057		putative AsnC family transcriptional regulator	NE	NE	GD	NE	NE
pRL100058		putative transmembrane protein	NE	NE	NE	NE	NE
pRL100059		pseudogene, putative aldolase	NE	NE	NE	NE	NE
pRL100060		pseudogene	NE	NE	NE	NE	NE
pRL100061		pseudogene, part of an ATP-binding component of ABC transporter PAAT?or POPT (contiguous genes)	NE	NE	NE	NE	NE
pRL100062		putative permease component of ABC transporter POPT	NE	NE	NE	NE	GD
pRL100063		putative permease component of ABC transporter POPT	NE	NE	NE	NE	NE
pRL100064		conserved hypothetical protein	NE	NE	NE	NE	NE
pRL100065		putative GntR family transcriptional regulator	NE	NE	NE	NE	NE
pRL100066		putative hydrolase	NE	NE	NE	NE	NE
pRL100067		putative racemase/decarboxylase	NE	NE	NE	NE	NE
pRL100162	*nifH*	nitrogenase iron protein NifH	NE	NE	NE	GD	NE

pRL100162A		hypothetical protein	NE	GD	GD		
pRL100163		hypothetical protein with homology to asparagine synthase (glutamine hyrdrolysing) at C-term	NE	GD	GD	NE	NE
pRL100164	*rhiI*	putative enzyme to synthesise (autoinducer) AHL, RhiI	NE	NE	GD	NE	NE
pRL100165		putative phage integrase/recombinase part of insertion sequence	GA	GA	GA	ES	ES
pRL100166		putative transposase for insertion sequence element	GA	NE	NE	ES	ES
pRL100167		putative transposase part of insertion sequence	NE	NE	NE	ES	ES
pRL100168		putative transposase part of insertion sequence	NE	NE	NE	NE	NE
pRL100169	*rhiA*	rhizosphere-induced protein RhiA	NE	NE	NE	NE	NE
pRL100170	*rhiB*	rhizosphere-induced protein RhiB	NE	NE	GD	NE	NE
pRL100171	*rhiC*	rhizosphere-induced protein RhiC	NE	NE	GD	NE	NE
pRL100172	*rhiR*	putative transcriptional regulatory protein controlling rhi gene expression RhiR	NE	NE	GD	NE	NE
pRL100173	*rhiD*	putative outer-membrane immunogenic protein precursor RhiD	NE	NE	NE	NE	NE
pRL100174		hypothetical protein, no known homology	NE	NE	NE	NE	NE
pRL100175	*nodO*	putative nodulation protein NodO	NE	NE	NE	NE	NE

pRL100176		pseudogene, incomplete ATP-binding protein	NE	GD	NE	NE	NE
pRL100177		putative homologue of eukaryotic tubulin	NE	NE	NE	NE	NE
pRL100178	*nodT*	putative nodulation protein NodT, TolC homologue	NE	NE	NE	NE	NE
pRL100179	*nodN*	nodulation protein NodN	NE	NE	GD	NE	NE
pRL100180	*nodM*	glucosamine--fructose-6-phosphate aminotransferase NodM	NE	NE	NE	NE	GD
pRL100181	*nodL*	putative nodulation protein NodL	NE	NE	NE	ES	ES
pRL100182	*nodE*	nodulation protein NodE homology with beta-ketoacyl synthases e.g.FabB	NE	NE	NE	GD	GD
pRL120150		putative urea amidolyase related protein	NE	NE	NE	NE	NE
pRL120151		putative urea amidohydrolyase	NE	NE	NE	NE	NE
pRL120152	*accC3*	putative biotin carboxylase	NE	NE	NE	NE	NE
pRL120153	*accB3*	putative biotin carboxyl carrier protein of acetyl-CoA carboxylase	NE	NE	NE	NE	NE
pRL120154		putative LamB/YcsF family protein	NE	NE	NE	NE	NE
pRL120155		putative LysR family transcriptional regulator	NE	NE	NE	NE	NE
pRL120157		conserved hypothetical protein	NE	NE	NE	NE	NE
pRL120158		conserved hypothetical protein	NE	NE	NE	NE	NE
pRL120159		putative amidohydrolase	NE	NE	NE	NE	NE

pRL120160		putative DeoR family transcriptional regulator (repressor)	NE	NE	ES	NE	NE
pRL120161		putative ATP-binding:ATP-binding (ABC:ABC) component of ABC transporter PepT	NE	GA	NE	NE	NE
pRL120162		putative permease component of ABC transporter PepT	NE	GA	NE	NE	NE
pRL120163		putative permease component of ABC transporter PepT	GA	GA	NE	NE	NE
pRL120164		putative SBP of ABC transporter PepT	NE	GA	NE	NE	NE
pRL120165		pseudogene	NE	NE	NE	NE	NE
pRL120166		putative transmembrane protein	NE	NE	NE	NE	NE
pRL120167		putative MFS family transporter	GA	NE	NE	NE	NE
pRL120168		putative LysR family transcriptional regulator	GA	NE	NE	NE	NE
pRL120169		putative permease component of ABC transporter POPT	GA	NE	NE	NE	NE
pRL120170		putative permease component of ABC transporter POPT	NE	NE	NE	NE	NE
pRL120682		putative rhizopine-binding protein precursor, putative substrate-binding component of ABC transporter CUT2	NE	NE	NE	NE	NE
pRL120683	*mcpB*	putative methyl accepting chemotaxis protein	NE	NE	NE	NE	NE
pRL120684		conserved hypothetical protein, uncharacterised membrane-anchored protein	NE	NE	NE	NE	NE

pRL120685	putative SBP of ABC transporter CUT1	NE	NE	NE	NE	NE
pRL120686	conserved hypothetical protein	NE	NE	NE	NE	NE
pRL120687	putative unsaturated glucuronyl hydrolase catalyses the hydrolytic release of unsaturated glucuronic acids from oligosaccharides	NE	NE	NE	NE	NE
pRL120688	putative non-secreted polygalacturonase (degrades pectin found in plant cell walls)	NE	NE	NE	NE	NE
pRL120689	putative ATP-binding component of ABC transporter CUT1	NE	NE	NE	NE	NE
pRL120690	putative permease component of ABC transporter CUT1	GA	NE	NE	NE	NE
pRL120691	putative permease component of ABC transporter CUT1	GA	NE	NE	NE	NE
pRL120692	putative ROK family transcriptional regulator	NE	NE	GD	NE	NE
pRL120693	putative AraC family transcriptional regulator (activator)	NE	NE	NE	NE	NE
pRL120694	putative LacI family transcriptional regulator (repressor)	NE	NE	NE	GD	GD
pRL120695	putative TetR family transcriptional regulator of RND family efflux transporter	NE	NE	NE	GD	GD
pRL120696	putative RND family efflux transporter (membrane fusion protein)	NE	NE	NE	NE	GD

pRL120697		putative RND family efflux transporter (membrane fusion protein)	NE	NE	NE	NE	NE
pRL120698		putative RND family efflux transporter (inner membrane protein)	NE	NE	NE	NE	NE
pRL120699		conserved hypothetical protein	NE	NE	NE	NE	NE
pRL120700		conserved hypothetical exported protein	NE	NE	NE	NE	NE
pRL120701		putative transmembrane magnesium and cobalt transport protein, CorA family	NE	GA	GA	NE	NE
pRL120702		putative transmembrane protein	NE	NE	NE	NE	NE
RL0246	*ddhA*	putative glucose-1-phosphate cytidylyltransferase	GA	NE	NE	NE	NE
RL0247		hypothetical protein	GA	NE	NE	NE	NE
RL0248		hypothetical protein	GA	NE	NE	NE	NE
RL0249		conserved hypothetical exported protein	NE	NE	NE	NE	NE
RL0250		conserved hypothetical protein	NE	NE	NE	NE	NE
RL0251		putative glutamine amidotransferase	NE	NE	NE	NE	NE
RL0252		putative nodulation protein	NE	NE	NE	NE	NE
RL0253	*gfa*	putative glutathione-dependent formaldehyde-activating enzyme	NE	NE	NE	NE	NE
RL0254	*lepA*	putative GTP-binding protein	NE	GD	GD	ES	ES
RL0255		putative HemeO family (TenA) transcriptional regulator	NE	GA	GA	NE	NE

RL0256		putative XRE family (HipB) transcriptional regulator with cupin2 domian	NE	NE	GD	NE	NE
RL0257		putative acetyltransferase	NE	NE	NE	NE	NE
RL0258		putative haloacid dehydrogenase hydrolase	NE	NE	NE	NE	NE
RL0259		putative acetylpolyamine aminohydrolase	NE	NE	NE	NE	NE
RL0260	*recX*	putative regulator of recombination, RecX protein thought to interact with RecA	NE	NE	NE	NE	NE
RL0261		putative transmembrane protein	NE	NE	NE	NE	NE
RL0262		truncated hemoglobin	NE	NE	NE	NE	NE
RL0263	*mbfA*	putative rubrerythrin (contains ferritin fold) transmembrane protein, related to proteins involved in iron and manganese transport (CCC1-like family)	NE	NE	NE	NE	NE
RL0264		conserved hypothetical protein	NE	NE	NE	NE	NE
RL0265		conserved hypothetical protein	NE	NE	NE	NE	NE
RL0266	*infC*	putative translation initiation factor	NE	GD	GD	NE	NE
RL0551	*hslO*	putative Hsp33-like chaperonin	NE	GD	GD	NE	NE
RL0552		hypothetical protein	NE	NE	NE	NE	NE
RL0553	*apaG*	putative methioine biosynthesis related protein	NE	NE	NE	NE	NE
RL0554	*metZ*	putative O-succinylhomoserine sulfhydrylase	NE	GD	NE	ES	ES
RL0555		putative deoxycytidine triphosphate deaminase	NE	NE	NE	NE	NE

RL0556		conserved hypothetical exported protein	NE	NE	NE	NE	NE
RL0558		putative transmembrane cellulose synthase catalytic subunit [UDP-forming]	NE	NE	NE	NE	NE
RL0559		putative UDP-glucose 4-epimerase	NE	NE	NE	NE	NE
RL0560		putative epimerase	NE	NE	NE	NE	NE
RL0561		putative AraC family transcriptional regulator (activator)	NE	NE	GD	NE	NE
RL0562		conserved hypothetical protein	NE	NE	NE	NE	NE
RL0563		conserved hypothetical protein	NE	NE	NE	NE	NE
RL0564	*mcpE*	putative methyl-accepting chemotaxis protein	NE	NE	NE	NE	NE
RL0565		putative SBP of ABC transporter PAAT (S. mel SBP homologue SMc02219 induced by valine, homoserine, isoleucine)	NE	NE	NE	NE	NE
RL0566		conserved hypothetical protein	NE	NE	NE	NE	NE
RL0567		conserved hypothetical exported protein	NE	NE	NE	NE	NE
RL0568		putative phage-related transcriptional regulator	NE	NE	NE	NE	NE
RL0569		putative two-component sensor/regulator; transcriptional regulator	NE	NE	NE	NE	NE
RL0570		conserved hypothetical exported protein	NE	NE	NE	NE	NE
RL0571		conserved hypothetical protein	NE	NE	NE	ES	ES

RL1030	*ispH*	putative 4-hydroxy-3-methylbut-2-enyl diphosphate reductase	GD	GD	GD	ES	ES
RL1031	*thrB*	putative homoserine kinase	GD	GD	GD	ES	ES
RL1032	*rnhA1*	putative ribonuclease HI	NE	NE	GD	ES	ES
RL1033		conserved hypothetical protein	NE	NE	NE	NE	NE
RL1034		conserved hypothetical protein	NE	NE	NE	NE	NE
RL1035		conserved hypothetical protein	NE	NE	NE	NE	NE
RL1036		conserved hypothetical protein	NE	NE	NE	NE	NE
RL1037		putative thioredoxin family protein	NE	NE	NE	NE	NE
RL1038		conserved hypothetical protein	NE	NE	NE	NE	NE
RL1039		conserved hypothetical protein	NE	NE	NE	NE	NE
RL1040		putative LysR family transcriptional regulator	NE	NE	GD	NE	NE
RL1041		putative oxidoreductase/monooxygenase	NE	NE	GD	NE	NE
RL1042		putative ATP-binding component of ABC transporter PAAT	NE	NE	NE	NE	NE
RL1043		putative SBP of ABC transporter PAAT	NE	NE	NE	NE	NE
RL1044		putative permease component of ABC transporter PAAT	NE	NE	NE	NE	NE
RL1045		putative permease component of ABC transporter PAAT	NE	NE	NE	NE	NE
RL1046		hypothetical protein	NE	NE	NE	NE	NE
RL1047		putative ATP-binding component of ABC transporter MZT	NE	NE	GD	NE	NE
RL1048		putative permease component of ABC transporter MZT	NE	NE	GD	NE	NE
RL1049		putative SBP of ABC transporter MZT	NE	NE	NE	NE	NE

RL1050		conserved hypothetical protein	NE	NE	NE	NE	NE
RL2818	*fnrN*	putative FNR/CRP family transcriptional regulator, 100% id to VF39 FnrN	NE	NE	NE	NE	GA
RL2819		hypothetical protein	GD	GD	NE	NE	GA
RL2820		hypothetical protein	GD	GD	GD	NE	NE
RL2821	*cobD*	putative cobalamin synthesis protein	GD	GD	GD	ES	ES
RL2822	*cobC*	putative cobalamin synthesis protein	GD	GD	GD	ES	ES
RL2823	*cobB*	putative cobalamin synthesis protein	GD	GD	GD	ES	ES
RL2824	*cobA*	putative uroporphyrin-III C-methyltransferase	NE	GD	GD	ES	ES
RL2825		putative inosine-uridine preferring nucleoside hydrolase	NE	NE	NE	NE	GA
RL2826	*cobE*	putative cobalamin synthesis protein	NE	NE	NE	NE	GA
RL2827		conserved hypothetical protein	NE	NE	NE	NE	GA
RL2828		putative XRE family (HipB) family transcriptional regulator	NE	NE	GD	ES	ES
RL2829	*cobO*	putative cob(I)yrinic acid a,c-diamide adenosyltransferase	GD	ES	GD	ES	ES
RL2830	*cobN*	putative cobalamin synthesis protein	GD	ES	GD	ES	ES
RL2831	*cobW*	putative cobalamin synthesis protein	GD	GD	GD	GD	GD
RL2832	*cobP*	putative bifunctional adenosylcobalamin biosynthesis protein	GD	GD	GD	GD	GD

RL2833		putative transmembrane ion channel protein	NE	NE	NE	NE	NE
RL2834	*cyaA10*	putative adenylate/guanylate cyclase	NE	NE	NE	NE	NE
RL2835		putative gluconate 5-dehydrogenase	NE	NE	NE	NE	NE
RL2836	*cobQ*	putative cobyric acid synthase	NE	GD	GD	ES	ES
RL2837		putative SBP of ABC transporter QAT? orphan proline/glycine betaine	NE	NE	GD	NE	NE
RL2838		conserved hypothetical protein	NE	NE	GA	NE	NE
RL3818		putative glycosyl transferase	NE	NE	NE	NE	NE
RL3819		putative FNR/CRP family transcriptional regulator	NE	NE	NE	GA	NE
RL3820		putative exopolysaccharide production protein	NE	NE	NE	ES	ES
RL3821		putative tyrosine-protein kinase involved in exopolysaccharide biosynthesis	NE	NE	NE	NE	NE
RL3822		putative glycosyl transferase	NE	NE	NE	NE	NE
RL3823		putative exopolysacchride production protein	NE	NE	NE	NE	NE
RL3824		putative transmembrane protein	NE	GA	GA	NE	NE
RL3825		putative transmembrane protein	NE	NE	NE	NE	NE
RL3826		putative glycosyltransferase	GA	NE	NE	NE	NE
RL3827		conserved hypothetical protein	GA	NE	GD	NE	NE
RL3828		putative FNR/CRP family transcriptional regulator	NE	NE	GD	NE	NE
RL3829	*exoY*	putative exopolysaccharide production regulator	NE	GD	NE	NE	GA

RL3831		putative O-acetyl transferase	NE	GA	GA	NE	NE
RL3832	*quiA*	putative transmembrane glucose dehydrogenase	GA	NE	NE	NE	NE
RL3833		putative short-chain dehydrogenase/reductase	NE	NE	NE	NE	NE
RL3834		putative ErfK/YbiS/YhnG oxidoreductase	NE	NE	NE	NE	NE
RL3835		conserved hypothetical exported protein	NE	NE	NE	NE	NE
RL3836		putative transmembrane protein	NE	GA	GA	NE	NE
RL3837		conserved hypothetical protein	NE	NE	GA	NE	NE
RL3838		putative transmembrane dehydrogenase/oxidoreductase	NE	NE	NE	NE	NE

[a] ES; growth-essential, GD; growth-defective, NE; growth-neutral, GA; growth-advantaged.

[b] In; input, Rh; rhizosphere, Ro; root attachment, No; nodulation, Re; nodule recovery

References

1. Wheatley, R.M., et al., *Role of O2 in the Growth of Rhizobium leguminosarum bv. viciae 3841 on Glucose and Succinate.* Journal of Bacteriology, 2017. **199**(1): p. e00572-16.
2. Perry, B.J., & Yost, C. K. , *Construction of a mariner-based transposon vector for use in insertion sequence mutagenesis in selected members of the Rhizobiaceae.* BMC microbiology, 2014. **14**(1): p. 298.
3. Perry, B.J., M.S. Akter, and C.K. Yost, *The use of transposon insertion sequencing to interrogate the core functional genome of the legume symbiont Rhizobium leguminosarum.* Frontiers in microbiology, 2016. **7**: p. 1873.

Appendix B. Research article publication

Wheatley, Rachel M., et al. "Role of O_2 in the growth of *Rhizobium leguminosarum* bv. *viciae* 3841 on glucose and succinate." *Journal of bacteriology* 199.1 (2017): e00572-16.

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Bacteriology®

RESEARCH ARTICLE

Role of O_2 in the Growth of *Rhizobium leguminosarum* bv. viciae 3841 on Glucose and Succinate

Rachel M. Wheatley,[a] Vinoy K. Ramachandran,[a] Barney A. Geddes,[a] Benjamin J. Perry,[b] Chris K. Yost,[b] Philip S. Poole[a]

Department of Plant Sciences, University of Oxford, Oxford, United Kingdom[a]; Department of Biology, University of Regina, Regina, SK, Canada[b]

ABSTRACT Insertion sequencing (INSeq) analysis of *Rhizobium leguminosarum* bv. viciae 3841 (Rlv3841) grown on glucose or succinate at both 21% and 1% O_2 was used to understand how O_2 concentration alters metabolism. Two transcriptional regulators were required for growth on glucose (pRL120207 [*eryD*] and RL0547 [*phoB*]), five were required on succinate (pRL100388, RL1641, RL1642, RL3427, and RL4524 [*ecfL*]), and three were required on 1% O_2 (pRL110072, RL0545 [*phoU*], and RL4042). A novel toxin-antitoxin system was identified that could be important for generation of new plasmidless rhizobial strains. Rlv3841 appears to use the methylglyoxal pathway alongside the Entner-Doudoroff (ED) pathway and tricarboxylic acid (TCA) cycle for optimal growth on glucose. Surprisingly, the ED pathway was required for growth on succinate, suggesting that sugars made by gluconeogenesis must undergo recycling. Altered amino acid metabolism was specifically needed for growth on glucose, including RL2082 (*gatB*) and pRL120419 (*opaA*, encoding omega-amino acid:pyruvate transaminase). Growth on succinate specifically required enzymes of nucleobase synthesis, including ribose-phosphate pyrophosphokinase (RL3468 [*prs*]) and a cytosine deaminase (pRL90208 [*codA*]). Succinate growth was particularly dependent on cell surface factors, including the PrsD-PrsE type I secretion system and UDP-galactose production. Only RL2393 (*glnB*, encoding nitrogen regulatory protein PII) was specifically essential for growth on succinate at 1% O_2, conditions similar to those experienced by N_2-fixing bacteroids. Glutamate synthesis is constitutively activated in *glnB* mutants, suggesting that consumption of 2-ketoglutarate may increase flux through the TCA cycle, leading to excess reductant that cannot be reoxidized at 1% O_2 and cell death.

IMPORTANCE *Rhizobium leguminosarum*, a soil bacterium that forms N_2-fixing symbioses with several agriculturally important leguminous plants (including pea, vetch, and lentil), has been widely utilized as a model to study *Rhizobium*-legume symbioses. Insertion sequencing (INSeq) has been used to identify factors needed for its growth on different carbon sources and O_2 levels. Identification of these factors is fundamental to a better understanding of the cell physiology and core metabolism of this bacterium, which adapts to a variety of different carbon sources and O_2 tensions during growth in soil and N_2 fixation in symbiosis with legumes.

KEYWORDS INSeq, nodules, *Rhizobium leguminosarum*, legumes, nitrogen fixation, *Pisum sativum*

Rhizobia are alphaproteobacteria able to form symbioses with legumes, on which they induce development of root nodules. Nodules provide an environment with both a low O_2 concentration and the correct nutrients to enable differentiation of rhizobia into nitrogen (N_2)-fixing bacteroids (1). Bacteroids fix atmospheric N_2 into ammonium (NH_4^+), which is secreted to the plant host, and are, in turn, provided with

Received 20 July 2016 **Accepted** 1 October 2016

Accepted manuscript posted online 17 October 2016

Citation Wheatley RM, Ramachandran VK, Geddes BA, Perry BJ, Yost CK, Poole PS. 2017. Role of O_2 in the growth of *Rhizobium leguminosarum* bv. viciae 3841 on glucose and succinate. J Bacteriol 199:e00572-16. https://doi.org/10.1128/JB.00572-16.

Editor Anke Becker, Philipps-Universität Marburg

Copyright © 2016 American Society for Microbiology. All Rights Reserved.

Address correspondence to Philip S. Poole, philip.poole@plants.ox.ac.uk.

carbon that they utilize as an energy source (2). To establish an effective N_2-fixing root nodule, a highly specific exchange of signals between plant and bacteria is required. Rhizobia attach to root hairs and grow down plant-derived infection threads into the root cortex (1, 3, 4). During release from infection threads, rhizobia become surrounded by a plant-derived symbiosome membrane, creating a low-O_2 environment for N_2 fixation to occur (5). Nitrogenase, the enzyme responsible for N_2 fixation, requires at least 16 molecules of magnesium-ATP (MgATP) per molecule of N_2 fixed ($N_2 + 8H^+ + 16MgATP + 8e^- \rightarrow 2NH_3 + H_2 + 16MgADP + 16P_i$) (6). NH_3 is not assimilated by bacteroids; instead, it is exported to the plant, where it is incorporated into amino acids which are transported to growing shoots (1).

Symbiotic N_2 fixation relies on precise and efficient integration of bacterial and plant metabolism. The plant provides dicarboxylic acids, predominately malate and succinate, and these are taken up by bacteroids via the dicarboxylic transport system (Dct), which is essential for N_2 fixation (7–9). Furthermore, either NAD^+ malic enzyme (diphosphopyridine nucleotide-dependent malic enzyme [Dme]) or the combined activities of Dme, phosphoenolpyruvate (PEP) carboxykinase, and pyruvate kinase are essential for dicarboxylate metabolism and N_2 fixation in bacteroids (9, 10).

Rhizobium leguminosarum bv. viciae strain 3841 (Rlv3841) induces root nodules on *viciae* legumes, such as the agriculturally important pea (*Pisum sativum*), vetch (*Vicia cracca*), and lentil (*Lens culinaris*) (11, 12). The genome of Rlv3841 is 7.75 Mb, consisting of a circular chromosome (4,788 genes) and six plasmids: pRL12 (790 genes), pRL11 (635 genes), pRL10 (461 genes), pRL9 (313 genes), pRL8 (141 genes), and pRL7 (189 genes) (12). Microarray, transcriptomic, and bioinformatics studies carried out on *Rhizobium* have indicated many gene functions. However, a high proportion (25%) remain uncharacterized across the genome, and while transcriptional studies allow identification of induced genes, they do not necessarily identify those important for growth in a particular environment (11, 13).

Mariner transposon insertion sequencing (INSeq) is a powerful and robust technique for the study of gene fitness at the genome scale. Libraries of insertion mutants are analyzed by high-throughput sequencing to assess the effect of mutation of a particular gene on growth and survival of the bacterium (14). Mariner transposons insert at thymine-adenine (TA) motifs found abundantly in Rlv3841 (15, 16), and INSeq sequencing data can be coupled with a hidden Markov model (HMM) for analysis (17). The HMM assigns genes to one of four classification states based on the fitness of mutants inferred from their population frequency in the sequenced library: growth essential (cannot tolerate insertion), growth defective (insertion impairs growth), growth neutral (insertion has a neutral impact on growth), or growth advantaged (insertion enhances growth).

C_4-dicarboxylates that fuel N_2 fixation in the low-O_2 environment of bacteroids are catabolized by the O_2-dependent tricarboxylic acid (TCA) cycle. Mutational studies of enzymes in the TCA cycle show that it is required for N_2 fixation in *R. leguminosarum* and *Sinorhizobium meliloti* (18–21), although the entire cycle may not always be essential, as is the case in soybean bacteroids (22, 23). Comparing catabolism of succinate with a sugar, such as glucose, at different O_2 tensions, is important in understanding how succinate catabolism is affected by low-O_2 environments and how C_4-dicarboxylic acids fuel N_2 fixation. Here we describe the use of a *Rhizobiaceae*-compatible MmeI-adapted mariner transposon sequencing vector, pSAM_RI (14), in an INSeq genetic screen of Rlv3841 growth under four different conditions: glucose (10 mM) at 21% O_2, succinate (20 mM) at 21% O_2, glucose (10 mM) at 1% O_2, and succinate (20 mM) at 1% O_2. Analysis of the growth requirements for these conditions, particularly growth on succinate at low O_2, is relevant to understanding bacteroid metabolism in legume nodules.

RESULTS AND DISCUSSION

Growth of *R. leguminosarum* and mariner library construction analysis. The generation times of Rlv3841 were 3.7 h (±0.1) when grown on glucose at 21% O_2, 4 h

Downloaded from http://jb.asm.org/ on August 28, 2018 by guest

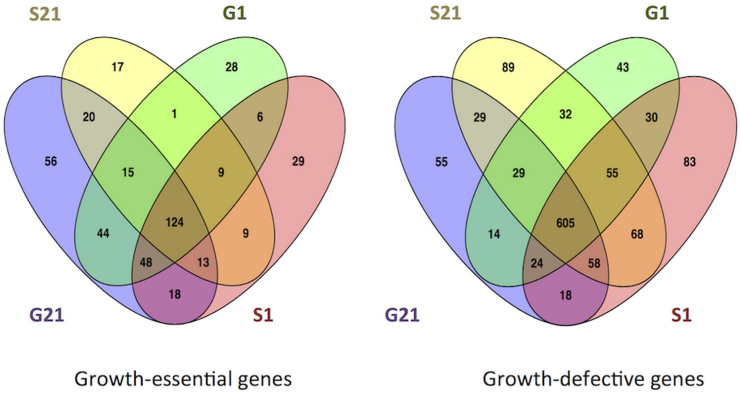


FIG 1 Venn diagram showing number of gene mutants classified as growth essential (ES) and growth defective (GD) under different growth conditions. G21, glucose at 21% O_2; S21, succinate at 21% O_2; G1, glucose at 1% O_2; S1, succinate at 1% O_2.

(±0.1) on succinate at 21% O_2, 3.8 h (±0.2) on glucose at 1% O_2, and 6.2 h (±0.3) on succinate at 1% O_2. The combination of succinate and 1% O_2 imposed the greatest growth restriction on Rlv3841. This is consistent with succinate being catabolized by the TCA cycle, which generates large amounts of NADH and $FADH_2$ which must then be reoxidized by the electron transport chain. For growth on succinate at 1% O_2, electron transport may start to limit growth as reoxidization of the TCA cycle-generated reductant is reliant on the presence of molecular oxygen to act as a final electron acceptor.

From sequencing Rlv3841 INSeq transposon insertion libraries from the four different growth conditions (glucose at 21% O_2, succinate at 21% O_2, glucose at 1% O_2, and succinate at 1% O_2), a total of 78 million barcoded sequencing reads were obtained. Sequencing data coupled with HMM analysis enabled the phenotypes of mutants with changes in 7,316 genes to be classified into one of four categories (growth essential, growth defective, growth neutral, or growth advantaged) (see Data set S1 in the supplemental material). No data are available for 21 genes of Rlv3841 which lack TA motifs (14).

Across the four different growth conditions, mutants of 307 genes (4.2% of the genome) resulted in classification as growth essential, 900 as growth defective (12.3%), 6,072 as (83.3%) growth neutral, and 7 as growth advantaged (0.1%) (Table S3). For 288 genes of unknown function, mutants had a growth-essential, growth-defective, or growth-advantaged phenotype. Under all four conditions, mutants in 124 genes (core genes) were found to be growth essential (Fig. 1). These encode well-characterized essential functions, including translation, transcription, ATP synthesis, membrane transport, and metabolite biosynthesis. A total of 605 genes were identified whose mutants were core growth defective (Fig. 1), encoding functions including transport, protein chaperoning, DNA repair, metabolism, and transcriptional regulation. Included in this category are 127 encoding hypothetical proteins of unknown function. An average of 7 mutants were classified as growth advantaged under each condition (these can be identified in Data set S1), and these genes were distributed on pRL7, pRL9, pRL10, and the chromosome. Mutation of a pseudogene on pRL10 (pRL100096) was growth advantaged specifically on succinate as a carbon source. Pseudogenes are generally considered to be nonfunctional genomic sequences; however, it is possible for them to encode partial activity or effect the mRNA stability of closely related functional genes (24). No mutants were growth advantaged under all growth conditions, consistent with natural selection removing any generally deleterious genes from the population.

Downloaded from http://jb.asm.org/ on August 28, 2018 by guest

TABLE 1 HMM analysis of five INSeq-identified genes, mutation of which leads to a growth-defective phenotype, and mean generation times of mutants on succinate at 21% O_2

Gene	No. of TA sites	Insertion density[a]	Mean read count[b]	Fitness value[c]	MGT[d] (h)
RL0036 (*pckA*)	22	0.09	1.67	3.34	16.0
RL1172	10	0.23	7.42	17.29	5.0
RL3424 (*dctA*)	17	0.27	2.85	13.28	16.0
RL4283 (*ptsP*)	25	0.28	2.15	15.05	5.2
RL4284 (*ask*)	24	0.21	1.08	5.39	6.4

[a]Insertion density is defined as the proportion of TA insertion sites containing an insertion in the mutant population relative to the HMM-predicted value.
[b]Mean read count is defined as the mean number of insertion mutants found at insertion sites that contained at least one insertion in the mutant population.
[c]Fitness value is a measure of total gene mutants in the population.
[d]Mean generation time (MGT) is calculated as the time it takes the population to double during exponential growth phase, measured in triplicate.

To validate the accuracy of INSeq gene fitness predictions, mutants in five genes, identified by INSeq as growth defective on succinate, were selected for experimental validation. The *in vivo* mean generation time (MGT) of wild-type Rlv3841 on succinate at 21% O_2 (4.0 h) was compared with that of strains mutated in RL0037 (*pckA*), RL1172, RL3424 (*dctA*), RL4283 (*ptsP*), and RL4284 (*ask*) (Table S1). Each mutant strain had a significantly longer MGT than wild-type Rlv3841 (Table 1). The output data from HMM analysis (number of insertion sites, insertion density, and mean read count) were analyzed for these genes and a fitness value was calculated for each by multiplying insertion density by the number of TA sites and then by the mean read count (Table 1). (Insertion density is the proportion of TA sites containing an insertion in the mutant population, and mean read count is the average number of insertions at TA sites containing at least one insertion.) There was observed to be an inverse relationship between fitness value and MGT (Table 1); the higher the fitness value, the shorter the MGT. Mutation of RL0037 (*pckA*) led to the lowest fitness value (3.34) and the longest MGT (16.0 h), while mutation of RL1172 gave the highest fitness value (17.29) and shortest MGT (5.0 h). Mutation of RL3424 (*dctA*) was an exception, having a reasonably high fitness value (13.28) yet one of the longest MGTs (16.0 h). This apparent discrepancy may be explained by most TA motifs in RL3424 being in the gene's termini, so mariner insertion may not have fully disrupted gene function (25).

Central metabolism. (i) The ED pathway is needed for optimal growth on both glucose and succinate. The Entner-Doudoroff (ED) pathway was shown to be essential for growth on glucose, as expected (26), but surprisingly, its mutation caused growth deficiency on succinate (Fig. 2 and Table 2). Most free-living rhizobia utilize the ED pathway for the catabolism of glucose to pyruvate (26, 27), with the pathway converting glucose-6-phosphate into pyruvate in a 1:2 molar ratio and generating one molecule of ATP (27). Its requirement for growth on succinate suggests that sugars made by gluconeogenesis may, at least partly, cycle back through the ED pathway (Fig. 2 and Table 2). Carbon cycling has been reported for a number of other bacterial species, including *Sinorhizobium meliloti* (26–28). RL0753 and RL1315 encode putative glucose-6-phosphate-1-dehydrogenases (Fig. 2 and Table 2), with these proteins sharing 43% amino acid identity (29). However, the essential classification of RL0753 suggests that it is the "real" *zwf* with RL1315 unable to compensate. RL0753 was also significantly transcriptionally upregulated on glucose versus succinate ($P < 0.05$), while expression of RL1315 was unchanged (11). Finally, RL0753 is in an operon with genes encoding enzymes catalyzing subsequent steps in the ED pathway (RL0751 to RL0753). Mutations of all three genes in this cluster are growth essential on glucose and growth defective on succinate (Table 2).

Intriguingly, mutation of RL4162 (*eda*), encoding the putative aldolase enzyme catalyzing formation of pyruvate and glyceraldehyde-3-phosphate, resulted in a growth-neutral classification across all conditions (Fig. 2 and Table 3). RL4162 was

Downloaded from http://jb.asm.org/ on August 28, 2018 by guest

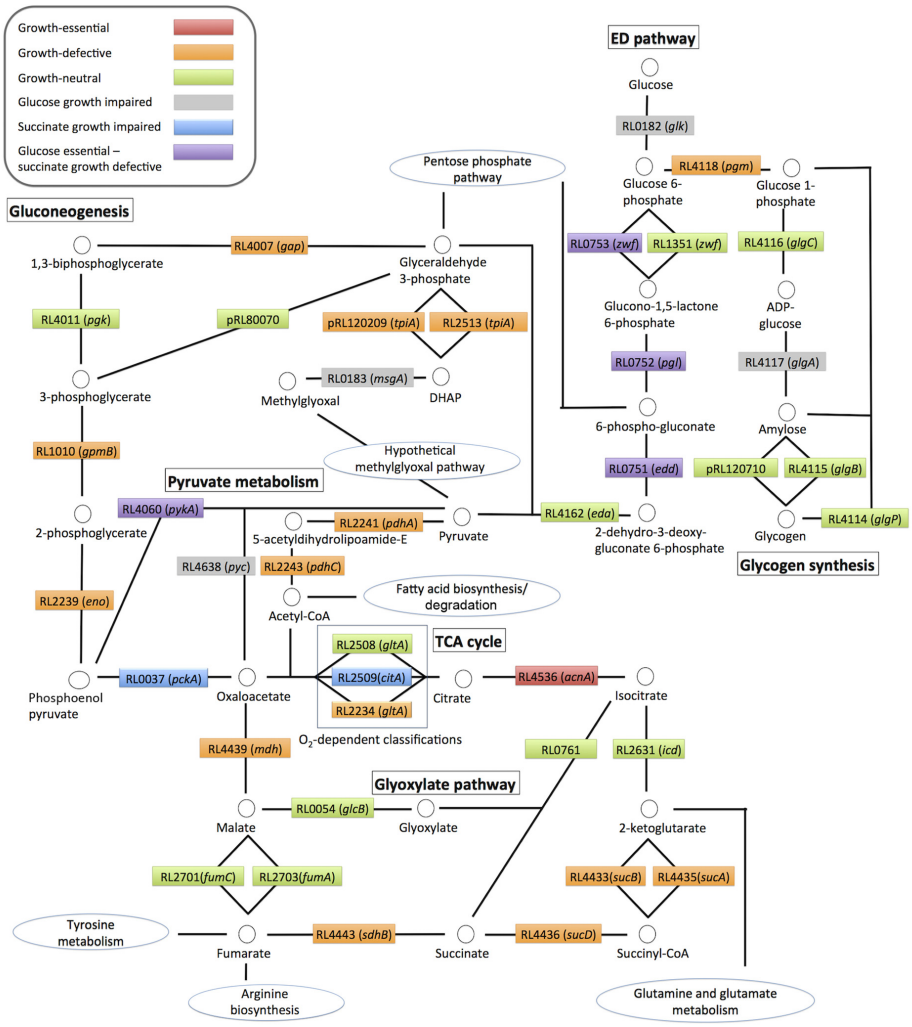


FIG 2 Central metabolic pathway of Rlv3841 showing the metabolism of glucose and succinate. Candidate genes for enzymes performing the catalytic steps are shown in colored boxes according to their INSeq mutant classification: red, growth essential (ES) under all growth conditions; orange, growth defective (GD) under all growth conditions; green, growth neutral (NE) under all growth conditions; gray, growth impaired on glucose, i.e., growth essential or growth defective specifically on glucose; blue, growth impaired on succinate, i.e., growth essential or growth defective specifically on succinate; purple, growth essential (ES) on glucose and growth defective on succinate.

upregulated on glucose (versus succinate) (11), but its growth-neutral classification suggests that an alternative aldolase is present (29). Sugars are also catabolized by the pentose phosphate pathway (PPP) in Rlv3841 (30). Enzymes catalyzing formation of 6-phospho-gluconate from glucose-6-phosphate were essential, but these reactions are shared by both the ED pathway and the PPP (Fig. 2 and Table 3) (30). There were no essential enzymes unique to the oxidative branch of the PPP.

(ii) The TCA cycle is essential for growth on glucose and succinate. The chromosome of Rlv3841 contains all the genes for a functional TCA cycle (12), and most

Downloaded from http://jb.asm.org/ on August 28, 2018 by guest

TABLE 2 INSeq classification of Rlv3841-encoded enzymes of glucose and succinate central metabolism under different growth conditions

Locus	Gene	Product description	Pathway(s)	INSeq classification[a] for:			
				G21	S21	G1	S1
pRL80070		Aldehyde dehydrogenase	Gluconeogenesis	NE	NE	NE	NE
pRL120710	*glgB*	1,4-Alpha-glucan branching enzyme	Glycogen synthesis	NE	NE	NE	NE
RL0037	*pckA*	Phosphoenolpyruvate carboxykinase	Pyruvate metabolism	NE	GD	NE	GD
RL0054	*glcB*	Malate synthase	Glyoxylate pathway	NE	NE	NE	NE
RL0182	*glk*	Glucokinase	ED pathway	ES	NE	GD	NE
RL0183	*msgA*	Methylglyoxal synthase	Hypothetical methyglyoxal pathway	ES	NE	GD	NE
RL0751	*edd*	Phosphogluconate dehydratase	ED pathway	ES	GD	ES	GD
RL0752	*pgl*	6-Phosphogluconolactonase	ED pathway/PPP	ES	GD	ES	GD
RL0753	*zwf*	Glucose-6-phosphate 1-dehydrogenase	ED pathway/PPP	ES	GD	ES	GD
RL0761		Isocitrate lyase	Glyoxylate pathway	NE	NE	NE	NE
RL1010	*gpmB*	Phosphoglycerate mutase	Gluconeogenesis	GD	GD	GD	GD
RL1351	*zwf*	Glucose-6-phosphate 1-dehydrogenase	ED pathway/PPP	NE	NE	NE	NE
RL2234	*gltA*	Citrate synthase	TCA cycle	GD	GD	GD	NE
RL2239	*eno*	Enolase	Gluconeogenesis	GD	GD	GD	GD
RL2241	*pdhA*	Pyruvate dehydrogenase subunit A	Pyruvate metabolism	GD	GD	GD	GD
RL2243	*pdhC*	Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex (PDC)	Pyruvate metabolism	GD	GD	GD	GD
RL2508	*gltA*	Citrate synthase II	TCA cycle	NE	NE	NE	NE
RL2509	*citA*	Citrate synthase I	TCA cycle	NE	GD	NE	NE
RL2513	*tpiA*	Triosephosphate isomerase	Gluconeogenesis/hypothetical methyglyoxal pathway	NE	NE	NE	NE
RL2631	*icd*	Isocitrate dehydrogenase (NADP)	TCA cycle	NE	NE	NE	NE
RL2701	*fumC*	Fumarate hydratase class II	TCA cycle	NE	NE	NE	NE
RL2703	*fumA*	Fumarate hydratase class I, aerobic	TCA cycle	NE	NE	NE	NE
RL4007	*gap*	Glyceraldehyde-3-phosphate dehydrogenase	Gluconeogenesis	GD	GD	GD	GD
RL4011	*pgk*	Phosphoglycerate kinase	Gluconeogenesis	NE	NE	NE	NE
RL4060	*pykA*	Pyruvate kinase	Pyruvate metabolism	ES	GD	ES	GD
RL4114	*glgP*	Glycogen phosphorylase	Glycogen synthesis	NE	NE	NE	NE
RL4115	*glgB*	1,4-Alpha-glucan branching enzyme	Glycogen synthesis	NE	NE	NE	NE
RL4116	*glgC*	Glucose-1-phosphate adenylyltransferase	Glycogen synthesis	NE	NE	NE	NE
RL4117	*glgA*	Glycogen synthase	Glycogen synthesis	GD	NE	GD	NE
RL4118	*pgm*	Phosphoglucomutase	Glycogen synthesis	GD	GD	GD	GD
RL4162	*eda*	2-Dehydro-3-deoxyphosphogluconate aldolase	ED pathway	NE	NE	NE	NE
RL4433	*sucB*	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase	TCA cycle	GD	GD	GD	GD
RL4435	*sucA*	2-Oxoglutarate dehydrogenase E1 component	TCA cycle	GD	GD	GD	GD
RL4436	*sucD*	Succinyl-coenzyme A synthetase alpha chain	TCA cycle	GD	GD	GD	GD
RL4439	*mdh*	Malate dehydrogenase	TCA cycle	GD	GD	GD	GD
RL4443	*sdhB*	Succinate dehydrogenase iron-sulfur protein	TCA cycle	GD	GD	GD	GD
RL4536	*acnA*	Aconitate hydratase	TCA cycle	ES	ES	ES	ES
RL4638		Pyruvate carboxylase	Pyruvate metabolism	GD	NE	GD	NE

[a]G21, glucose at 21% O_2; S21, succinate at 21% O_2; G1, glucose at 1% O_2; S1, succinate at 1% O_2; ES, growth essential; GD, growth defective; NE, growth neutral.

genes encoding TCA cycle enzymes showed the same INSeq classification on both glucose and succinate (Fig. 2 and Table 2). Aconitase (encoded by RL4536 [*acnA*]) was growth essential across all four growth conditions (Fig. 2 and Table 2), consistent with the TCA cycle being essential for growth on both glucose and succinate. Koziol et al. reported that stable *Sinorhizobium acnA* mutants are possible only in a citrate synthase (*gltA*) null background, suggesting that intracellular accumulation of citrate is toxic (31). Fumarate hydratase (RL2701 and RL2703) and isocitrate dehydrogenase (RL2631), catalyzing the formation of malate from fumarate and 2-ketoglutarate from isocitrate, respectively, were growth neutral (Table 2). While RL2701 and RL2703 indicate redundancy of fumarate hydratase, there is only a single gene with homology to isocitrate dehydrogenase (RL2631) (29). In both *Bradyrhizobium japonicum* and *S. meliloti*, mutation of isocitrate dehydrogenase has no severe effects on growth, and mutants nodulated their respective host plants (32, 33). In *S. meliloti*, isocitrate dehydrogenase is essential for N_2 fixation (32); however, it is not needed in *B. japonicum* (33). When isocitrate dehydrogenase was mutated (32, 33), it was unknown if and where flux was rerouted.

Downloaded from http://jb.asm.org/ on August 28, 2018 by guest

TABLE 3 INSeq-identified Rlv3841 genes, mutation of which specifically reduces growth on glucose, together with INSeq classification under different growth conditions

Locus	Gene	Product description	Classification	INSeq classification[a] for:			
				G21	S21	G1	S1
pRL110211		SBP of ABC transporter, PAAT family	Cell processes	GD	NE	GD	NE
pRL110565		Transmembrane protein	Cell envelope	ES	NE	ES	NE
pRL120207	*eryD*	DeoR family transcriptional regulator (repressor)	Regulation	GD	NE	GD	NE
pRL120419	*opaA*	Omega-amino acid:pyruvate transaminase	Metabolism of small molecules	GD	NE	GD	NE
RL0182	*glk*	Glucokinase	Metabolism of small molecules	ES	NE	GD	NE
RL0183	*mgsA*	Methylglyoxal synthase	Metabolism of small molecules	ES	NE	GD	NE
RL0547	*phoB*	Phosphate regulon transcriptional regulator PhoB	Regulation	GD	NE	GD	NE
RL1479		Conserved hypothetical protein	Conserved in *E. coli*	ES	NE	ES	NE
RL1754		Conserved hypothetical exported protein	Conserved in *E. coli*	GD	NE	GD	NE
RL2082	*gatB*	Aspartyl/glutamyl-tRNA$^{Asn/Gln}$ amidotransferase subunit B	Macromolecule metabolism	ES	NE	GD	NE
RL3005		Hypothetical protein	Unknown function, no known homologues	ES	NE	ES	NE
RL4117	*glgA*	Glycogen synthase	Macromolecule metabolism	GD	NE	GD	NE
RL4638		Pyruvate carboxylase	Metabolism of small molecules	GD	NE	GD	NE

[a]G21, glucose at 21% O_2; S21, succinate at 21% O_2; G1, glucose at 1% O_2; S1, succinate at 1% O_2; ES, growth essential; GD, growth defective; NE, growth neutral.

Three candidate citrate synthase genes are present in Rlv3841: RL2508, RL2509, and RL2234 (12) (Table 2). On succinate at 21% O_2, mutation of either RL2234 or RL2509 caused a growth-defective phenotype, while on glucose at both 21% O_2 and 1% O_2, only RL2234 was required. Optimum growth on a TCA cycle intermediate (succinate) requires greater TCA cycle enzyme activity than does growth on a sugar such as glucose. This may explain why activity of two citrate synthases is needed when growing on succinate, while one is sufficient for growth on glucose. However, at 1% O_2 with succinate, mutation of any of the three candidate citrate synthases was growth neutral (Fig. 2 and Table 2). The combination of 1% O_2 and succinate results in the longest MGT for Rlv3841 and suggests that TCA cycle activity is severely reduced, minimizing the requirement for citrate synthase isozymes and, presumably, activity (Fig. 2 and Table 2). The TCA cycle generates large amounts of reductant, posing reoxidization problems at such low O_2 availability. It is therefore likely that there is reduced flux through the TCA cycle at 1% O_2 and bacterial growth is slowed (MGT of 3.7 h on glucose at 1% O_2 versus 6.2 h on succinate at 1% O_2). Reduced flux through the TCA cycle at low O_2 is strikingly similar to metabolism occurring in bacteroids, particularly of soybean, where mutation of genes encoding even crucial TCA cycle enzymes does not prevent N_2 fixation (32, 33).

Mutations that specifically reduced growth on glucose at both 1% and 21% O_2. Thirteen genes were required specifically for growth on glucose (i.e., mutation of these genes results in a growth-essential or growth-defective phenotype at both 21% and 1% O_2) (Table 3) and encoded functions span regulation, the cell envelope, small-molecule metabolism, macromolecule metabolism, and cell processes (Fig. 3). The largest proportion are involved in small-molecule metabolism (Fig. 3), and a number of these genes have known enzymatic functions in central carbohydrate metabolism; RL0182 (*glk*) encodes a glucokinase needed for the first step in glucose activation to glucose-6-phosphate (Fig. 2), RL4117 (*glgA*) encodes a glycogen synthase, and RL4638 encodes a pyruvate carboxylase (Fig. 2). Glycogen synthesis is required for growth on glucose (26), while pyruvate carboxylase has an essential anaplerotic function converting pyruvate to oxaloacetate, for further metabolism in the TCA cycle (34).

Growth on glucose required the transcriptional regulators RL0547 (*phoB*) and pRL120207 (*eryD*). pRL120207 (*eryD*) encodes a repressor of the erythritol catabolic operon (35), suggesting that unregulated expression of the erythritol operon interferes with glucose catabolism. The requirement of pRL120207 (*eryD*) for growth on glucose was not observed in *S. meliloti* (36, 37). However, this difference highlights how single mutational studies are not necessarily comparable to INSeq, in which mutants are

Downloaded from http://jb.asm.org/ on August 28, 2018 by guest

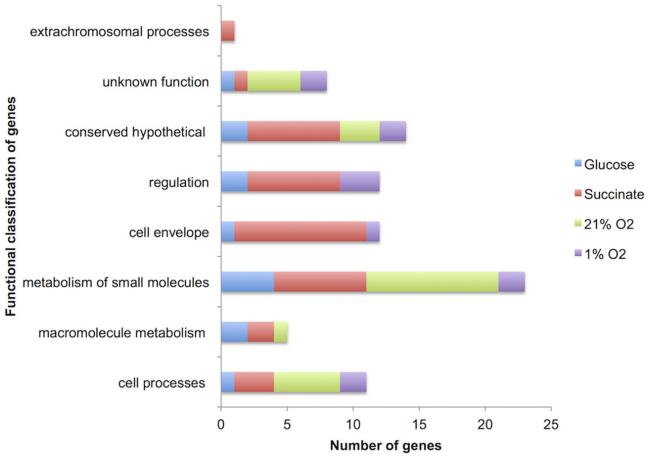


FIG 3 Function distribution of the INSeq-identified genes required specifically for growth on glucose, succinate, 21% O_2, and 1% O_2 (listed in Tables 3 to 6). Functional classification is based on Riley codes (12).

observed in a competition with thousands of other strains. This can allow resolution of growth defects that may not be evident when the mutant is the sole strain present.

RL0547 (*phoB*) encodes the response regulator of the phosphate limitation (PhoR-PhoB) two-component regulatory system (38). Glucose is metabolized via sugar phosphate intermediates and so phosphate and glucose metabolism are linked, which is not the case for dicarboxylate metabolism. Thus, the requirement for PhoB for growth on glucose suggests that integration with the Pho regulon may be due to the synthesis of phosphorylated intermediates (39, 40).

Rlv3841 required the methylglyoxal pathway alongside the ED pathway for optimal growth on glucose. The methylglyoxal pathway is present in some prokaryotic species, converting dihydroxyacetone phosphate (DHAP) derived from glycolysis or the ED pathway into methylglyoxal and then into pyruvate (41). This pathway is activated by increased concentrations of glucose phosphates in *E. coli* (42). This suggests that the pathway reduces sugar phosphate concentrations and increases inorganic phosphate without generating ATP (42). We propose that the methylglyoxal pathway may run in parallel to the ED pathway in Rlv3841 to consume excess glucose-6-phosphate and generate pyruvate (Table 4), consequently reducing the concentration of phosphorylated intermediates generated in glucose catabolism. This concept is supported by the clustering of methylglyoxal synthase (RL0183) and glucokinase (RL0182). While mutating methylglyoxal synthase may be polar on glucokinase and impair glucose catabolism, the ED and methylgloxal pathways are clearly linked. A functional methylglyoxal pathway has not previously been identified in rhizobia; however, the committed enzyme in the methylglyoxal pathway, methylglyoxal synthase (RL0183), was required for growth specifically on glucose (Fig. 2 and Table 4). Microarray data show that expression of RL0183 is 2-fold induced on glucose compared to the value for succinate-grown cells ($P < 0.01$) (13). Bioinformatics analysis indicates that Rlv3841 possesses all necessary genes for a functional methylglyoxal pathway (29, 38) (Table 4) and that the enzymes are well integrated into general central metabolism. Mutations in many of the genes encoding enzymes of this pathway were growth defective on glucose (Table 4). The final two steps of the methylglyoxal pathway have multiple potential candidate genes, which could explain why their mutation is growth neutral (Table 4).

Two further plasmid genes were required for growth on glucose: pRL110565, encoding a putative transmembrane protein, and pRL120419 (*opaA*), encoding an

Downloaded from http://jb.asm.org/ on August 28, 2018 by guest

TABLE 4 Proposed Rlv3841 methylglyoxal metabolic pathway

Compound	Enzyme (gene[s]) or pathways	INSeq classification[a]
Glucose		ES
↓	Glucokinase (RL0182)	
Glucose 6-phosphate		ES
↓	Pentose phosphate and ED pathways	
Glyceraldehyde 3-phosphate		GD
↓	Triosephosphate isomerase (pRL120209, RL2513)	
Dihydroxyacetone phosphate		GD
↓	Methylglyoxal synthase (RL0183)	
Methylglyoxal		NE
↓	Methylglyoxal reductase (pRL120760, RL3064, RL3243)	
Lactate		NE
↓	Lactate dehydrogenase (pRL120231, pRL120601, RL0444, RL3578)	
Pyruvate		

[a]ES, growth essential; GD, growth defective; NE, growth neutral.

omega-amino acid:pyruvate transaminase (43). The transmembrane protein possesses a conserved MatE domain, which primarily functions in the export of metabolic and xenobiotic organic cations (29). OpaA synthesizes beta-alanine by the alanine-dependent transamination of pyruvate (44), which is a key intermediate in glucose catabolism. This may help regulate pyruvate levels by buffering with a nonprotogenic amino acid.

Mutations that specifically reduced growth on succinate at both 21% and 1% O_2. C_4-dicarboxylates, including succinate, are provided to bacteroids by the plant host during symbiosis to fuel N_2 fixation (7). Thirty-eight genes were required specifically for growth on succinate (i.e., mutants are growth essential or growth defective on succinate at both 21% and 1% O_2) (Table 5). Functional classification of these genes shows that the largest proportion encodes cell envelope-related proteins (Fig. 3). These included genes encoding proteins within the PrsD-PrsE type I secretion system (RL3657), genes involved in exopolysaccharide (EPS) synthesis (RL4658), and genes encoding transmembrane proteins (RL0423, RL1391, and RL1392). Bioinformatics analysis indicates that RL3654 to RL3657 all have potential functional roles in the cell envelope (29). RL3654 encodes a polysaccharide biosynthesis protein, RL3655 a glycosyltransferase, RL3656 a lipase, and RL3657 (*prsE*) the permease component of the PrsDE type I secretion system. PrsDE exports the EPS glycanases PlyA and PlyB, which are responsible for correct cleavage of EPS (45, 46). PrsDE is also responsible for export of rhizobial adhesins and other proteins to the outer surface. While mutation of the ATP-binding component of this transporter (RL3658, encoding PrsD) had no significant effect on fitness, microarray data show that RL3658 (*prsD*) was significantly upregulated on succinate versus glucose ($P > 0.05$) (11). However, three genes (RL0072, RL0181, and RL0623) encode homologues with 33 to 38% identity to PrsD, suggesting functional redundancy (29, 46).

In addition, of the four genes, RL1390 to RL1393, required for growth on succinate, three have predicted functional roles in the cell envelope, with both RL1391 and RL1392 encoding putative transmembrane proteins and RL1393 encoding a putative peptidoglycan biosynthesis protein (Table 5). RL4658 (*exoB*) encodes UDP-glucose-4-epimerase, required for synthesis of UDP-galactose (47), which is incorporated into the repeating units of EPS and other polysaccharides (48). The high number of cell envelope genes required for growth on succinate suggests that restructuring of the cell surface, involving the PrsD-PrsE type I secretion system and UDP-galactose production in particular, is required to enable growth on organic acids.

Genes encoding five transcriptional regulators were required for growth on succinate: pRL100388, RL1641, RL1642, RL3427, and RL4524 (*ecfL*) (Table 5). pRL100388

Downloaded from http://jb.asm.org/ on August 28, 2018 by guest

TABLE 5 INSeq-identified Rlv3841 genes, mutation of which specifically reduces growth on succinate, together with INSeq classification under different growth conditions

Locus	Gene	Product description	Classification	INSeq classification[a] for:			
				G21	S21	G1	S1
pRL80122	*traA*	Conjugal transfer protein TraA	Extrachromosomal element	NE	GD	NE	GD
pRL90053		O-antigen ligase	Cell envelope	NE	GD	NE	GD
pRL90207		Permease component of ABC transporter, CUT2 family	Cell processes	NE	GD	NE	GD
pRL90208	*codA*	Cytosine deaminase	Metabolism of small molecules	NE	GD	NE	GD
pRL100388		LacI family transcriptional regulator (repressor)	Regulation	NE	GD	NE	GD
RL0037	*pckA*	Phosphoenolpyruvate carboxykinase	Metabolism of small molecules	NE	GD	NE	GD
RL0055		Conserved hypothetical protein	Conserved in *E. coli*	NE	ES	NE	ES
RL0423		Transmembrane protein	Cell envelope	NE	GD	NE	GD
RL0588		Peptidoglycan binding protein	Cell envelope	NE	GD	NE	GD
RL0754	*ordL*	Oxidoreductase	Metabolism of small molecules	NE	GD	NE	GD
RL0957		Conserved hypothetical protein	Conserved in other organism than *E. coli*	NE	GD	NE	GD
RL1388		Conserved hypothetical protein	Conserved in other organism than *E. coli*	NE	GD	NE	GD
RL1390		Hypothetical protein	Unknown function, no known homologues	NE	GD	NE	GD
RL1391		Transmembrane protein	Cell envelope	NE	GD	NE	GD
RL1392		Transmembrane protein	Cell envelope	NE	GD	NE	GD
RL1393	*pbpF*	Peptidoglycan biosynthesis/penicillin binding protein	Cell envelope	NE	GD	NE	GD
RL1561	*pabC*	Aminodeoxychorismate lyase	Metabolism of small molecules	NE	GD	NE	GD
RL1641		MerR family transcriptional regulator	Regulation	NE	GD	NE	GD
RL1642		Two-component sensor/regulator; transcriptional regulator	Regulation	NE	GD	NE	GD
RL2384	*recG*	ATP-dependent DNA helicase	Macromolecule metabolism	NE	ES	NE	GD
RL2385		Conserved hypothetical protein	Conserved in *E. coli*	NE	ES	NE	GD
RL2478		Outer membrane protein	Cell envelope	NE	GD	NE	GD
RL2990	*ubiA*	*para*-Hydroxybenzoate-polyprenyltransferase	Metabolism of small molecules	NE	GD	NE	GD
RL3424	*dctA*	C_4-dicarboxylate transport protein	Cell processes	NE	GD	NE	GD
RL3425	*dctB*	Histidine kinase (C_4-dicarboxylate transport)	Regulation	NE	GD	NE	GD
RL3426	*dctD*	Transcriptional regulator C_4-dicarboxylate transport (sigma 54)	Regulation	NE	GD	NE	GD
RL3427		AsnC family transcriptional regulator	Regulation	NE	GD	NE	GD
RL3465		Conserved hypothetical protein	Conserved in *E. coli*	NE	GD	NE	GD
RL3468	*prs*	Ribose-phosphate pyrophosphokinase	Metabolism of small molecules	NE	ES	NE	ES
RL3512		Conserved hypothetical protein	Conserved in *E. coli*	NE	GD	NE	GD
RL3654		Polysaccharide biosynthesis protein	Cell envelope	NE	GD	NE	GD
RL3655		Glycosyltransferase	Cell envelope	NE	GD	NE	GD
RL3656		Lipase	Metabolism of small molecules	NE	GD	NE	GD
RL3657	*prsE*	Permease component of ABC transporter, export family	Cell processes	NE	GD	NE	GD
RL3668		Serine/threonine protein phosphatase	Macromolecule metabolism	NE	GD	NE	ES
RL4523		Conserved hypothetical exported protein	Conserved in organisms other than *E. coli*	NE	GD	NE	ES
RL4524	*ecfL*	RNA polymerase ECF sigma factor	Regulation	NE	GD	NE	ES
RL4658	*exoB*	UDP-glucose 4-epimerase	Cell envelope	NE	GD	NE	ES

[a]G21, glucose at 21% O_2; S21, succinate at 21% O_2; G1, glucose at 1% O_2; S1, succinate at 1% O_2; ES, growth essential; GD, growth defective; NE, growth neutral.

encodes a putative LacI family transcriptional regulator with homology to a gluconate utilization system transcriptional repressor found in many prokaryotes (29, 49). RL1641 and RL1642 encode two-component sensor/regulator units belonging to the MerR family of transcriptional regulators. RL3427 encodes an AsnC family transcriptional regulator, which could possibly be involved in regulation of genes encoding the dicarboxylic acid transport (Dct) system located adjacently. As expected, mutants with changes in the dicarboxylic acid transport system (*dctABD*) (RL3424 to RL3426) were growth defective on succinate (Table 5). Mutations in the *dct* system have been shown to impair uptake of succinate into cells, leading to growth arrest (50–53). RL4524 (*ecfL*) encodes an extracytoplasmic function (ECF) sigma factor (54). ECF sigma factors regulate transcription in response to extracellular stimuli (55). Their activity is commonly under the regulation of an anti-sigma factor (ASF), but RL4524 is one of only four Rlv3841 ECF genes not adjacent to an ASF gene. RL4524 expression increased in the

Downloaded from http://jb.asm.org/ on August 28, 2018 by guest

presence of pea root exudate, presumably due to an unknown compound (54), and was significantly upregulated in the pea rhizosphere (versus Rlv3841 grown on glucose and ammonia) (13). From results shown here, it may be that EcfL (RL4524) responds to C_4-dicarboxylic acids present in pea root exudates.

Mutations of two genes involved in nucleobase synthesis, pRL90208 and RL3468, were growth deficient and growth essential, respectively, on succinate. pRL90208 (*codA*) encodes a putative cytosine deaminase that functions in pyrimidine metabolism (38). Both CodA and the protein product of RL0891 are suggested to possess deaminase activity utilized in conversion of cytosine to uracil and of 5-methylcytosine to thymine (38). INSeq analysis showed RL0891 to be to be nonessential, suggesting that its product is unable to compensate for the activity of CodA, whose gene insertion mutation was growth defective on succinate. RL3468 (*prs*) encodes a ribose-phosphate pyrophosphokinase catalyzing biosynthesis of 5-phosphoribosyl diphosphate (PRPP) from ribose 5-phosphate. PRPP is utilized in both purine and pyrimidine biosynthesis, with the first committed step to purine biosynthesis utilizing PRPP as a substrate (56). This suggests that for optimal growth on succinate, there is a requirement for increased nucleotide recovery or synthesis compared to growth on glucose. RL0037 (*pckA*) encodes phosphoenolpyruvate carboxykinase, which converts oxaloacetate into phosphoenolpyruvate, essential for gluconeogenesis (Fig. 2) (9). RL2990 (*ubiA*) encodes a 4-hydroxybenzoate octaprenyltransferase required for ubiquinone biosynthesis and electron transport. Cells growing on a TCA cycle intermediate are likely to have a greater requirement for reoxidation of $NADH/FADH_2$ by the electron transport chain. RL1561 (*pabC*) was also found to be required, and the gene encodes an enzyme in the shikimate pathway also needed for ubiquinone synthesis. This agrees with a greater need for ubiquinone in succinate-grown cells.

Mutations that reduced growth at 21% O_2 on both glucose and succinate. Twenty-three genes were required for optimal growth at 21% O_2 on both carbon sources (i.e., mutants are growth essential or growth defective on both glucose and succinate at 21% O_2) (Table 6) and encoded functions spanning small-molecule metabolism, macromolecule metabolism, and cell processes (Fig. 3). The largest proportion is involved in small-molecule metabolism (Fig. 3), and 11 genes are predicted to encode enzymes utilizing oxygen:cytochrome *c* oxidase subunits (RL1021 and RL1022), a deoxygenase (RL0802), a hydroperoxide resistance protein (RL2927), a peroxiredoxin (RL2440), three oxidoreductases (RL0847 [an IMP oxidoreductase], RL3834 [an ErfK oxidoreductase], and RL4186), a mitochondrial respiratory chain complex assembly factor (RL0920), and two adenylate cyclases (RL2441 and RL2926). None of the proteins encoded by these genes are required at 1% O_2, presumably because their function is to protect against oxygen toxicity.

Other genes required for optimal growth on 21% O_2 include RL0921, predicted to encode a CorA magnesium transporter, and RL0922, encoding a potassium uptake protein (29). Cationic transporters are electrochemical potential driven, and activity is dependent on the oxidation state of the metal ion species (57). Insertion mutation of these enzymes did not impair growth at 1% O_2, perhaps because transport of ions in their more reduced state does not limit growth or because other transporters play a greater role at low O_2. Rlv3841 possesses an MgtE magnesium transporter channel (RL1461 [*mgtE*]) that is essential for growth when both the magnesium concentration is limiting and the pH is low, which is also essential for N_2 fixation on specific legumes (58). CorA may be the primary transporter of magnesium at 21% O_2 and another transporter, such as MgtE, may play a greater role at decreased O_2.

RL3597 encodes a putative RNA helicase protein, involved in mRNA biogenesis, and RL1548 (*radA*) encodes a putative DNA repair protein (29, 59). The observation that their mutants are growth defective is presumably due to their preventative role in oxidative damage to nucleic acids.

Mutations that reduced growth at 1% O_2 on both glucose and succinate. Twelve genes were required for optimal growth at 1% O_2 on both succinate and

Downloaded from http://jb.asm.org/ on August 28, 2018 by guest

TABLE 6 INSeq-identified Rlv3841 genes, mutation of which specifically reduces growth at 21% O_2, together with INSeq classification under different growth conditions

Locus	Gene	Product description	Classification	INSeq classification[a] for:			
				G21	S21	G1	S1
pRL90042		Pseudogene, conserved hypothetical protein	Not classified (cryptic gene)	GD	GD	NE	NE
pRL120796		Hypothetical exported protein	Unknown function, no known homologues	GD	GD	NE	NE
RL0802		Deoxygenase	Metabolism of small molecules	GD	GD	NE	NE
RL0847	*guaB*	IMP dehydrogenase	Metabolism of small molecules	ES	ES	NE	NE
RL0920		ATP-binding Mrp family protein	Metabolism of small molecules	GD	GD	NE	NE
RL0921		Cationic transport protein, CorA family	Cell processes	GD	GD	NE	NE
RL0922	*kup*	Potassium uptake transport system protein	Cell processes	GD	GD	NE	NE
RL1021	*coxB*	Cytochrome *c* oxidase polypeptide II precursor (cytochrome aa_3 subunit 2)	Metabolism of small molecules	GD	GD	NE	NE
RL1022	*coxA*	Cytochrome *c* oxidase polypeptide I (cytochrome aa_3 subunit 1)	Metabolism of small molecules	GD	GD	NE	NE
RL1547	*cvpA*	Colicin V production protein	Cell processes	GD	GD	NE	NE
RL1548	*radA*	DNA repair protein RadA homologue	Macromolecule metabolism	GD	GD	NE	NE
RL2440		Peroxiredoxin	Metabolism of small molecules	GD	GD	NE	NE
RL2441		Adenylate cyclase	Metabolism of small molecules	GD	GD	NE	NE
RL2926		Adenylyl cyclase	Metabolism of small molecules	GD	GD	NE	NE
RL2927	*ohrB*	Organic hydroperoxide resistance protein	Cell processes	GD	GD	NE	NE
RL3180		Hypothetical protein	Unknown function, no known homologues	GD	GD	NE	NE
RL3181		Conserved hypothetical protein	Conserved in organisms other than *E. coli*	GD	GD	NE	NE
RL3182		Conserved hypothetical protein	Conserved in organisms other than *E. coli*	GD	GD	NE	NE
RL3507		Phosphoesterase	Not classified (included assignments)	ES	GD	NE	NE
RL3596		Conserved hypothetical protein	Conserved in organisms other than *E. coli*	GD	GD	NE	NE
RL3597		DEAD box ATP-dependent RNA helicase protein	Cell processes	GD	GD	NE	NE
RL3834		ErfK/YbiS/YhnG oxidoreductase	Metabolism of small molecules	GD	GD	NE	NE
RL4186		Oxidoreductase	Metabolism of small molecules	GD	GD	NE	NE

[a]G21, glucose at 21% O_2; S21, succinate at 21% O_2; G1, glucose at 1% O_2; S1, succinate at 1% O_2; ES, growth essential; GD, growth defective; NE, growth neutral.

glucose (i.e., mutants are growth essential or growth defective on both glucose and succinate at 1% O_2) (Table 7) and encoded functions spanning regulation, the cell envelope, small-molecule metabolism, and cell processes (Fig. 3). The largest proportion is involved in regulation, and three potential low-O_2 growth regulators were identified: pRL110072, RL0546 (*phoU*), and RL4042. pRL110072 encodes a putative GntR family transcriptional regulator, and RL4042 encodes a putative AsrR family transcriptional regulator.

TABLE 7 INSeq-identified Rlv3841 genes, mutation of which specifically reduces growth at 1% O_2, together with INSeq classification under different growth conditions

Locus	Gene	Product description	Classification	INSeq classification[a] for:			
				G21	S21	G1	S1
pRL80039		Pseudogene	Not classified	NE	NE	GD	GD
pRL100141		Pseudogene, conserved hypothetical protein	Not classified	NE	NE	GD	GD
pRL110016		Conserved hypothetical protein	Conserved in organisms other than *E. coli*	NE	NE	GD	GD
pRL110072		GntR family transcriptional regulator	Regulation	NE	NE	GD	GD
RL0546	*phoU*	Phosphate uptake regulator PhoU	Regulation	NE	NE	GD	GD
RL1506	*relA*	Stringent response protein	Cell processes	NE	NE	GD	GD
RL2585	*nodX*	Nodulation protein NodX (probable sugar acetylase)	Metabolism of small molecules	NE	NE	GD	GD
RL3412		Conserved hypothetical protein	Conserved in *E. coli*	NE	NE	GD	GD
RL4042		AsrR family transcriptional regulator	Regulation	NE	NE	GD	GD
RL4693		Polysaccharide deacetylase	Cell envelope	NE	NE	GD	GD
RL4694		(Di)nucleoside polyphosphate hydrolase	Metabolism of small molecules	NE	NE	GD	GD
RL4695	*bfr*	Bacterioferritin	Cell processes	NE	NE	GD	GD

[a]G21, glucose at 21% O_2; S21, succinate at 21% O_2; G1, glucose at 1% O_2; S1, succinate at 1% O_2; ES, growth essential; GD, growth defective; NE, growth neutral.

Downloaded from http://jb.asm.org/ on August 28, 2018 by guest

RL0546 (*phoU*) is annotated to encode a putative phosphate transport regulatory protein. While the role of PhoU in *R. leguminosarum* has not been investigated, it has been well studied for *E. coli*, in which the protein forms part of the phosphate-sensing pathway that negatively regulates the Pho regulon (60). Mutants deficient in *phoU* show constitutive expression of Pho regulon genes (60–62), encoding proteins involved in phosphate metabolism, energy production, membrane transport, flagellum synthesis, and transcription factors, including a stringent response protein (63). However, it is not known how similar the mechanisms of PhoU regulation are between *E. coli* and Rlv3841. PhoU may be involved in growth regulation under low O_2 through a role in the coordination of phosphate and energy metabolism. The generation of ATP is dependent on the transport of phosphate in the cell, and energy metabolism is O_2 dependent.

While RL2585 is annotated as NodX, the presence of which enables a few strains of *Rhizobium* to nodulate Afghanistan peas (29), the protein encoded by RL2585 shows no significant identity to the characterized NodX (UniProtKB accession number P08888) (29). However, RL2585 possesses a conserved acyltransferase domain, and that might be required for modification of a surface-exposed sugar. RL4693 encodes a putative polysaccharide deacetylase protein, RL4694 a putative dinucleoside polyphosphate hydrolase, and RL4695 a putative bacterioferritin. RL1506 encodes RelA, a central stringent response protein controlling the synthesis of (p)ppGpp, required for adaptation to stress conditions. RelA coordinates responses to the abundance of nitrogen and carbon in *Rhizobium* (64). In addition to these roles, RelA is suggested to regulate the stress response to oxygen limitation in *Bacillus subtilis* (65), with a *relA* deletion mutant strain unable to synthesize the heat shock proteins induced in response to oxygen limitation in the wild type (65). The exact mechanism of this regulation is unclear and does not appear to have been studied for any other bacterial species. However, the requirement of RL1506 (*relA*) for growth specifically at low oxygen suggests that RelA also regulates a stress response resulting from oxygen limitation in Rlv3841.

Mutations that specifically reduced growth on succinate at 1% O_2. There were a number of mutations resulting in a growth-essential or growth-defective phenotypes specific to one condition but growth neutral across all other conditions. Fifty-two mutations were specific to glucose at 21% O_2, 41 specific to succinate at 21% O_2, 29 specific to glucose at 1% O_2, and 65 specific to succinate at 1% O_2. The combination of 1% O_2 and succinate is most relevant to symbiotic conditions, with 64 genes classified as causing a growth-defective phenotype upon insertion mutation and a single growth-essential gene, RL2393 (*glnB*). The MGT of Rlv3841 on succinate at 1% O_2 was almost double that of the other three growth conditions (6.2 h compared to 3.7 h, 3.8 h, and 4 h). Under stress, Rlv3841 may be particularly susceptible to mutations altering growth, possibly explaining the higher number of mutations that specifically effected growth under this condition.

The sole gene identified as growth essential on succinate at 1% O_2, RL2393 (*glnB*), encodes the nitrogen regulatory protein PII. This protein forms part of the two-component signaling system that regulates nitrogen-responsive proteins, including glutamine synthetase. Null mutations in *glnB* cause constitutive activation of ammonia assimilation in *R. leguminosarum*, promoting increased synthesis of glutamate and glutamine (66, 67). It is intriguing that only on succinate at 1% O_2, conditions similar to that experienced by bacteroids, is the activity of RL2393 (*glnB*) essential. GlnB is downregulated during bacteroid differentiation (68), and microarray data show that expression of *glnB* is increasingly downregulated from 7 to 28 days in pea bacteroids, suggesting that GlnB is required only prior to symbiotic N_2 fixation (13). Constitutive activation of nitrogen assimilation in an RL2393 (*glnB*) mutant may cause too much 2-ketoglutarate, used for glutamate synthesis, to be removed from the TCA cycle. Depletion of 2-ketoglutarate would force a turn of the TCA cycle to replace it and lead to the production of extra reductant that must then be deoxidized. In a 1% O_2

Downloaded from http://jb.asm.org/ on August 28, 2018 by guest

environment, reoxidation of NADH via the electron transport chain may be limited by the O_2 available.

A plasmid-specific toxin-antitoxin system on pRL10. Mutations to genes in the putative operon pRL100005-10 were growth essential or growth defective across all four conditions, and the operon is predicted to encode a putative toxin-antitoxin system (29). The operon consists of two candidate toxin genes (pRL100005 and pRL00009) and three candidate antitoxin or transcriptional regulator component genes (pRL00006, pRL00007, and pRL100010) (see Data set S1) (29). The two candidate toxin genes encode PIN domain proteins, which commonly confer toxicity due to RNase activity and require coexpression of antitoxin components to ensure cell viability (69). Finding a novel toxin-antitoxin system on pRL10 accords with the ability to cure Rlv3841 of each plasmids, with the exception of pRL10 (70). Consideration of this novel toxin-antitoxin system is important if trying to generate new plasmidless rhizobial strains.

Conclusion. In summary, an INSeq screen of Rlv3841 grown on glucose and succinate at both 21% and 1% O_2 enabled identification of factors needed for growth on different carbon sources and O_2 levels. Use of HMM analysis assigned gene phenotype classifications to 7,316 genes in the genome of Rlv3841 across four different growth conditions. Analysis of central carbon metabolism pathways allowed roles in glucose and succinate metabolism to be distinguished and the ability of pathways to compensate for single enzyme mutations to be phenotypically quantified. A hypothetical methylglyoxal pathway is proposed for Rlv3841, and enzymes of this pathway were found to be required for growth on glucose. One of the most dramatic observations is that only mutation of RL2393 (*glnB*) caused a growth-essential phenotype on succinate at 1% O_2. Since *glnB* mutants of *R. leguminosarum* are constitutively activated for ammonia assimilation, this suggests that removal of 2-ketoglutarate from the TCA cycle by glutamate synthesis is highly deleterious when C_4-dicarboxylates are the carbon source in a low-O_2 environment, conditions mimicking those of N_2-fixing bacteroids in symbiosis with legumes.

MATERIALS AND METHODS

Strains, plasmids, and culture conditions. Rlv3841 strains listed in Table S1 in the supplemental material were grown at 28°C in tryptone-yeast extract (TY) (71) or universal minimal salts (UMS) medium with appropriate carbon and nitrogen sources. UMS medium is derived from AMS (72), with the alterations made being $CaCl_2{\cdot}2H_2O$ (0.51 mM), $CoCl_2{\cdot}6H_2O$ (4.2 μM), EDTA-Na_2 (1 μM), and $FeSO_4{\cdot}7H_2O$ (0.04 mM). All *Escherichia coli* strains (Table S1) were grown at 37°C in Lennox (L) broth or on L agar (73). Antibiotics were used at the following concentrations (in micrograms per milliliter) unless otherwise stated: gentamicin, 20 (10 for *E. coli*); kanamycin, 20; neomycin, 40; spectinomycin, 50; streptomycin, 500; and tetracycline, 2 in UMS and 5 in TY. The donor *E. coli* strain, SM10λ*pir* carrying the mariner INSeq vector pSAM_RI, was supplemented with 50 μg/ml of neomycin to select for the plasmid and 100 μg/ml of ampicillin (Amp) to select for *E. coli*.

Generation time of Rlv3841. To assess growth, 2×10^7 Rlv3841 cells were grown in 500 ml of UMS containing 100 mM ammonium chloride and 500 μl of 1,000× stock vitamins. Either glucose (10 mM) or succinate (20 mM) was added as a carbon source. The cultures were grown at 25°C under either atmospheric conditions (21% O_2) or at 1% O_2 in an oxygen cabinet (Belle Technology). For cultures (biological triplicates), optical density (at 600 nm [OD_{600}]) measurements were taken at 4-h intervals until growth reached stationary phase. The mean generation time (MGT) of Rlv3841 was calculated as the number of hours it took the population to double while in exponential growth phase. To assess *in vivo* growth of growth-defective mutants, wild-type Rlv3841 and pK19-generated gene mutants (listed in Table S1) were grown on TY agar slopes. Growth was measured at 21% O_2 with succinate (20 mM) as used for mutant library DNA extractions, and OD_{600} measurements at 4-h intervals were taken in triplicate.

Mariner library construction. Donor mariner transposon *E. coli* cells were grown in L broth culture overnight, and recipient rhizobial cells were grown on a TY agar slope. Cultures were subjected to three cycles of pelleting at 6,000 rpm for 5 min, followed by resuspension in TY medium. Donor and recipient cells were pooled in equal ratios, pelleted at 6,000 rpm for 5 min, and resuspended in 30 μl of TY medium. Cell suspensions were spotted onto nitrocellulose filters on TY agar plates and incubated at 27°C overnight, and then the bacteria from the filters were resuspended in UMS medium plus 15% glycerol. Enumeration of total transposon insertion Rlv3841 was performed on 500 μg/ml of streptomycin plus 50 μg/ml of neomycin TY agar. Independent pools of transposon mutants were generated by plating three libraries on 500 μg/ml of streptomycin plus 50 μg/ml of neomycin TY agar and pooling of rhizobial mutants.

Cultures were prepared as described above to measure MGT, with the substitution of 2×10^7 cells from the mariner transposon mutant library. At mid-exponential growth (OD_{600} of 0.5), DNA was

Downloaded from http://jb.asm.org/ on August 28, 2018 by guest

extracted from approximately 10^9 cells using a Qiagen DNeasy blood and tissue kit for Gram-negative bacteria according to the manufacturer's protocol, with the following modifications: 2 μl of RNase A (100 mg/ml) was added in parallel to proteinase K, and DNA was eluted in 100 μl of Milli-Q water. DNA extraction for three biological replicates generated 12 different library pools. To confirm successful rhizobial DNA extraction, samples were run under electrophoresis on an agarose gel alongside a GeneRuler 1-kb DNA ladder (Thermo Scientific). DNA was quantified with a NanoDrop spectrophotometer (Thermo Scientific).

Library preparation and sequencing. Three independent pools of 2×10^7 transposon mutants were grown for 14 generations under four different growth conditions: glucose at 21% O_2, succinate at 21% O_2, glucose at 1% O_2, and succinate at 1% O_2. Library preparation was carried out independently for each of the mutant libraries. Transposon tags were prepared for DNA sequencing using a modified version of the INSeq method to make the process compatible with the Ion Proton system sequencing platform (74). Linear PCR products were amplified using Ion Proton-compatible BioSAM primers (14) with an annealing temperature of 58.6°C and 500 ng of template DNA. Primer and adaptor sequences are detailed in Table S2. Linear PCR products were purified (GeneJET PCR purification kit) and eluted in 50 μl of Milli-Q water. The biotinylated linear PCR products were bound for affinity capture with Pierce streptavidin magnetic beads (Thermo Scientific), and the enzymatic library preparation steps were performed with the substitution of Klenow (New England BioLabs), Random Primer 6 (New England BioLabs), and T4 DNA ligase (New England BioLabs). A custom INSeq library adapter was used in the adapter ligation step, and then final PCR amplification used sequencing template-incorporated fusion primers designed to be compatible with Ion amplicon library preparation. The 12 INSeq libraries were barcoded using 12 different forward fusion primers which gave different IonXpress barcode sequences for downstream sequence separation. An Ion Proton system reverse-sequencing target, trP1, was used in conjunction with these forward barcode primers. This final sequencing template was 187 bp in length and was gel purified from the PCR products with a GeneJET gel purification kit (Thermo Scientific). DNA to be used for sequencing was analyzed with a bioanalyzer high-sensitivity DNA chip (Agilent Technologies) to ensure its quality and molarity.

Libraries were diluted to 100 pM and pooled. The libraries (100 pM DNA in 25 μl) underwent automated template preparation using an Ion Chef (Life Technologies). DNA sequencing was performed on an Ion Proton system sequencer.

Transposon insertion analysis with a four-state HMM. Sequencing reads were uploaded onto a Linux server for analysis as previously described (14). In short, "cutadapt" (75) was used for quality trimming and removal of adapter sequences, and the resulting 15- to 16-bp tn-tags were checked for a leading TA motif with a custom-written Perl script. These tags were mapped to the Rlv3841 reference genome using the Bowtie short-read aligner (76) and grouped into Rlv3841 replicons. The files were converted to a .wig format by a custom-written Perl script for further analysis using the Tn-HMM Python module (17). This first calculates the HMM state of each individual TA site and then determines the state of all TA sites within gene boundaries to assign each gene a state as a whole. HMM analysis assigned genes to one of four classification states: growth essential, growth defective, growth neutral, or growth advantaged. The Tn-HMM Python module was used to quantify the number of genes assigned to each state. Gene annotations were from RlegDB (http://rlegdb.jic.ac.uk/).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/JB.00572-16.

TEXT S1, PDF file, 0.08 MB.

DATA SET S2, XLS file, 1.1 MB.

ACKNOWLEDGMENTS

This work was supported by the Natural Environment Research Council (grant number NE/L501530/1) and the Biotechnology and Biological Sciences Research Council (grant numbers BB/K001868/1 and BB/K001868/2).

REFERENCES

1. Udvardi MK, Poole PS. 2013. Transport and metabolism in legume-rhizobia symbioses. Annu Rev Plant Biol 64:781–805. https://doi.org/10.1146/annurev-arplant-050312-120235.
2. Prell J, White JP, Bourdes A, Bunnewell S, Bongaerts RJ, Poole PS. 2009. Legumes regulate Rhizobium bacteroid development and persistence by the supply of branched-chain amino acids. Proc Natl Acad Sci U S A 106:12477–12482. https://doi.org/10.1073/pnas.0903653106.
3. Terpolilli JJ, Hood GA, Poole PS. 2012. What determines the efficiency of N2-fixing Rhizobium-legume symbioses? Adv Microb Physiol 60:326.
4. Oldroyd GE, Downie JA. 2004. Calcium, kinases and nodulation signalling in legumes. Nat Rev Mol Cell Biol 5:566–576. https://doi.org/10.1038/nrm1424.
5. Day DA, Kaiser BN, Thomson R, Udvardi MK, Moreau S, Puppo A. 2001. Nutrient transport across symbiotic membranes from legume nodules. Funct Plant Biol 28:669–676. https://doi.org/10.1071/PP01028.
6. Lodwig EM, Poole PS. 2003. Metabolism of Rhizobium bacteroids. CRC Crit Rev Plant Sci 22:37–78. https://doi.org/10.1080/713610850.
7. Lodwig EM, Hosie AH, Bourdès A, Findlay K, Allaway D, Karunakaran R, Downie JA, Poole PS. 2003. Amino-acid cycling drives nitrogen fixation in the legume-Rhizobium symbiosis. Nature 422:722–726. https://doi.org/10.1038/nature01527.
8. White J, Prell J, James EK, Poole PS. 2007. Nutrient sharing between symbionts. J Plant Physiol 144:604–614. https://doi.org/10.1104/pp.107.097741.
9. Mulley G, Lopez-Gomez M, Zhang Y, Terpolilli J, Prell J, Finan T, Poole PS. 2010. Pyruvate is synthesized by two pathways in pea bacteroids with

Downloaded from http://jb.asm.org/ on August 28, 2018 by guest

Appendix C. Review article publication

Wheatley, Rachel M., and Philip S. Poole. "Mechanisms of bacterial attachment to roots." *FEMS microbiology reviews* (2018).

FEMS Microbiology Reviews, fuy014, 42, 2018, 448–461

doi: 10.1093/femsre/fuy014
Advance Access Publication Date: 17 April 2018
Review Article

REVIEW ARTICLE

Mechanisms of bacterial attachment to roots

Rachel M. Wheatley* and Philip S. Poole

Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK

*Corresponding address: Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK. Tel: +44 (0) 1865 275000; E-mail: rachel.wheatley@plants.ox.ac.uk

One sentence summary: A common biphasic model of root attachment exists in agriculturally important microbial species.
Editor: Bart Thomma

ABSTRACT

The attachment of bacteria to roots constitutes the first physical step in many plant–microbe interactions. These interactions exert both positive and negative influences on agricultural systems depending on whether a growth-promoting, symbiotic or pathogenic relationship transpires. A common biphasic mechanism of root attachment exists across agriculturally important microbial species, including *Rhizobium*, *Agrobacterium*, *Pseudomonas*, *Azospirillum* and *Salmonella*. Attachment studies have revealed how plant–microbe interactions develop, and how to manipulate these relationships for agricultural benefit. Here, we review our current understanding of the molecular mechanisms governing plant–microbe root attachment and draw together a common biphasic model.

Keywords: root attachment; plant–microbe interaction; rhizobacteria; adhesion; molecular mechanism; plants

INTRODUCTION

The rhizosphere is the zone of soil immediately surrounding plant roots and can support a rich and diverse bacterial community (Walker *et al.* 2003). Plants may release up to a fifth of their photosynthate via their roots (Estabrook and Yoder 1998), and so the rhizosphere is under strong influence of root exudates and the secreted products of bacterial metabolism and degradation. Attaching to the rooting system not only enables plant–microbe interactions to develop, but also carries the significant benefit of anchoring bacteria in the nutrient-rich environment of the rhizosphere. Root exudation is a critical driving force shaping the establishment of microbial populations (Grayston and Campbell 1996). Exudates can serve as preferential growth substrates for particular microbes, and in turn, the interactions that develop with these microbes can exert preferential selectivity on plant fitness through their positive or negative growth influences.

Reviewing root attachment in agriculturally important microbial genera, *Rhizobium*, *Agrobacterium*, *Pseudomonas*, *Azospirillum* and *Salmonella*, demonstrates a common biphasic mechanism exists within them. This comprises a primary attachment phase, characterised by a weak, reversible and non-specific binding of bacteria to the root surface. This is followed by a secondary attachment phase, characterised by a switch to a stronger, more specific binding mode, and the synthesis of extracellular fibrils aiding further bacterial accumulation and aggregation. Divisions of this biphasic mechanism are denoted differently throughout the literature. Primary attachment is commonly referred to as a 'settlement', 'adsorption' or 'reversible' phase, and secondary attachment is commonly referred to as a 'residence', 'anchoring' or 'irreversible' phase (Michiels, Croes and Vanderleyden 1991; Rodríguez-Navarro, Dardanelli and Ruíz-Saínz 2007; Matthyssee 2014).

Bacteria exhibiting this biphasic root attachment mechanism encompass a diverse range of plant interactions, as plant-growth promoters, biofertilisers and pathogens. One factor common to the development of these plant–microbe interactions is the initial requirement to attach to plant roots. In some cases, the attached bacteria reside adhered to the outer surface of the root, or in other cases, attachment is a prerequisite for the endophytic or pathogenic colonisation of the inner root. It is important to note that the vast majority of root attachment research has been carried out in plant–microbe interactions involving Gram-negative bacteria (Rodríguez-Navarro, Dardanelli

Received: 29 November 2017; **Accepted:** 14 April 2018
© FEMS 2018. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

rom https://academic.oup.com/femsre/article-abstract/42/4/448/4975273
iversity user
ber 2018

and Ruíz-Saínz 2007; Bogino *et al.* 2013; de Jesus Sousa and Olivares 2016), such as the five well-characterised genera of *Rhizobium, Agrobacterium, Pseudomonas, Azospirillum* and *Salmonella*. Gram-positive bacteria, such as members of the *Bacillus* and *Streptomyces* genus, also engage in agriculturally important plant interactions, and have more recently become a bigger focus of attention in root attachment research, having already been generally well characterised in abiotic biofilm formation processes (Beauregard *et al.* 2013; Viaene *et al.* 2016; Habib *et al.* 2017).

Rhizobium, Agrobacterium and *Azospirillum* are grouped within the class of Alphaproteobacteria. *Rhizobium* species are able to engage in symbiosis with leguminous plants and function as a nitrogen biofertiliser. After attaching to and colonising legume roots, rhizobia become internally housed in protective 'nodule' structures, within which they catalyse biological nitrogen fixation and provide the plant with useable nitrogen sources for growth (Udvardi and Poole 2013). They receive carbon sources primarily in the form of dicarboxylates in return, the metabolism of which helps fuel the energetically expensive process of nitrogen fixation (Udvardi and Poole 2013). Many species of *Agrobacterium* are non-pathogenic and live saprophytically in the soil (Danhorn and Fuqua 2007). However, the best-studied species, *Agrobacterium tumefaciens*, is the causative agent behind crown-gall disease in many flowering plants, and the nature of this pathogenic DNA transmission gives it significant uses as a tool for genetic transformation.

Azospirillum is probably the best-characterised genus of plant growth-promoting rhizobacteria (PGPR) (Steenhoudt and Vanderleyden 2000). *Azospirillum* species associate with numerous agriculturally important plants and positively influence plant growth, crop yields and nitrogen content (Steenhoudt and Vanderleyden 2000). Plant growth promotion is attributed to several mechanisms, including the production of growth phytohormones, such as auxins, cytokyinins and gibberellins, and biological nitrogen fixation. Growth phytohormones commonly improve plant growth by stimulating root hair growth and increasing the root surface area, which improves the acquisition of water and mineral nutrients (Steenhoudt and Vanderleyden 2000). Azospirilla are able to exist as free-living nitrogen-fixing bacteria, but can also attach to roots, and some species can colonise the internal root tissues as endophytes. Intriguingly, *Azospirillum* may represent an evolutionary bridge between water- and soil-dwelling microbes, and a better understanding of how *Azospirillum* colonises plants might therefore also help improve understanding of how microbes transferred from water to soil environments (Orlandini *et al.* 2014).

Pseudomonas and *Salmonella* are both genera of Gammaproteobacteria, they exhibit extensive species diversity and are capable of colonising a wide range of niches (Madigan and Martinko 2005). The best-studied rhizosphere pseudomonads are *Pseudomonas fluorescens, P. syringae* and *P. putida*. From an agricultural perspective, *P. fluorescens* and *P. putida* are valued for their plant growth-promoting properties, including growth phytohormone production, plant pathogen suppression, iron chelation by siderophore production and aiding in bioremediation (Goswami *et al.* 2013). Conversely, strains of *P. syringae* are known for plant pathogenicity and their ability to infect a diverse range of plant species. Following attachment and infection, the pathogenic action of the majority of *P. syringae* strains is to secrete plant toxins and produce ice nucleation proteins, which cause water to freeze and induce ice injury to plant cells. *Salmonella* are commonly known for pathogenicity in animal species, but these microbes can also live saprophytically in soil and infect plants. By infecting plants, *Salmonella* are later able to transfer infection to animals through plant ingestion (Barak *et al.* 2005; Brankatschk *et al.* 2014; Wiedemann *et al.* 2015). There are numerous reports linking *Salmonella* contaminated plant produce to food poisoning in humans (Krtinić, Đurić and Ilić 2010; Wiedemann *et al.* 2015), yet the root attachment mechanism of *Salmonella* remains poorly characterised.

MOLECULAR MECHANISMS GOVERNING PRIMARY ATTACHMENT

Root migration

Migration towards the root via chemotaxis constitutes an attraction phase that generally precedes attachment (Begonia and Kremer 1994). Root-exuded nutrients often serve as chemoattractants. Common chemoattractant signals include phenolics, organic acids, amino acids and phytosiderophores. Flavonoids induce positive chemotaxis in strains of *Rhizobium* (Aguilar *et al.* 1988; Armitage, Gallagher and Johnston 1988; Dharmatilake and Bauer 1992), and malate and benzoates induce positive chemotaxis in strains of *Azospirillum* (Lopez-de-Victoria and Lovell 1993). Flagella and pili are filamentous proteinaceous appendages that extend from the bacterial cell surface. Bacteria typically use flagella or pili as propellers to migrate towards the root and position within sufficient proximity for initial weak attractive forces to develop (Caetano-Anollés *et al.* 1988; Berne *et al.* 2015; Zheng *et al.* 2015). This active propulsion helps overcome any repulsive barriers, such as repulsive electrostatic forces due to the electrostatic charge of the cell envelope (Berne *et al.* 2015). Flagella are significantly longer than pili and are present at fewer numbers per cell, they generally show a distinct pattern of distribution (i.e. polar flagellum), whereas pili are distributed randomly over the cell surface (Van Gerven, Waksman and Remaut 2011). In addition to a role in motility, flagella and pili can also function adhesively in attachment (Vesper 1987; Smit, Kijne and Lugtenberg 1989; Smit *et al.* 1989; Croes *et al.* 1993; Pratt and Kolter 1998; Van Gerven, Waksman and Remaut 2011; Zheng *et al.* 2015).

Fundamental weak binding forces

The universal characteristics of primary attachment are weak, reversible and non-specific binding, allowing association of single cells to the surface. Primary attachment is initially dictated by physiochemical and electrostatic attractions between surface molecules of the root and bacterial cell envelope (Berne *et al.* 2015). A number of fundamental weak binding forces form these preliminary surface associations across all plant–microbe interactions: Van der Waals forces, electrostatic forces and hydrophobic interactions (Van Loosdrecht *et al.* 1987; Berne *et al.* 2015). Van der Waals forces are the ubiquitous attractions that arise from momentary changes in electron density in a molecule (Kendall and Roberts 2015). They are displaced by stronger electrostatic forces, caused by differences in electric charge between molecules (Van Loosdrecht *et al.* 1987; Berne *et al.* 2015). The hydrophobic properties of the microbial cells and root surface contribute the strongest force of these initial associations, such that hydrophobic cells experience attractive forces to hydrophobic surfaces (Kochkodan *et al.* 2008; Giaouris, Chapot-Chartier and Briandet 2009; Krasowska and Sigler 2014).

The bacterial cell surface possesses a net negative charge under the majority of physiological conditions (Poortinga *et al.* 2002). Protrusive components of the cell surface initially function in primary attachment by means of their electrostatic

from https://academic.oup.com/femsre/article-abstract/42/4/448/4975273
niversity user
nber 2018

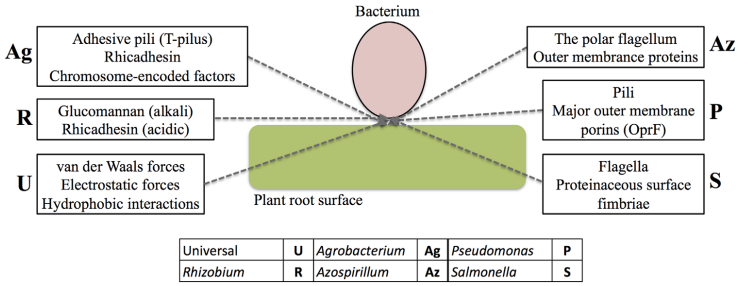


Figure 1. Molecular mechanisms of primary attachment allowing the weak and reversible binding of a single cell to the root. Universal binding forces and microbe-specific attachment factors are indicated: U; Universal, R; *Rhizobium*, Ag; *Agrobacterium*, Az; *Azospirillum*, P; *Pseudomonas*, S; *Salmonella*.

charge and steric properties (Janczarek *et al.* 2015). Cell surface electrostatic properties are modified by contact of the cells with inert surfaces such as soil particles (Bushby 1990), and unsurprisingly, the heterogeneity of electrostatic cell surface distribution also affects repulsion during bacterial adhesion (Poortinga *et al.* 2002). However, electrostatic surface properties are only important in the short-term (Van Loosdrecht *et al.* 1987; Poortinga *et al.* 2002; Bogino *et al.* 2013; Janczarek *et al.* 2015), and contribute relatively weak interactions towards the overall efficiency of primary attachment. This is exemplified well with exopolysaccharides (EPS) coating rhizobial cells. EPS electrostatic charge plays a minor role in overall mechanics of initial physiochemical attachment compared to the much stronger forces of EPS hydrophobicity (Janczarek *et al.* 2015), which in turn plays a minor role comparative to adhesive binding mediated by any microbe-specific binding factors later in the attaching mechanism.

MICROBE-SPECIFIC PRIMARY ATTACHMENT FACTORS

Primary attachment in *Rhizobium*

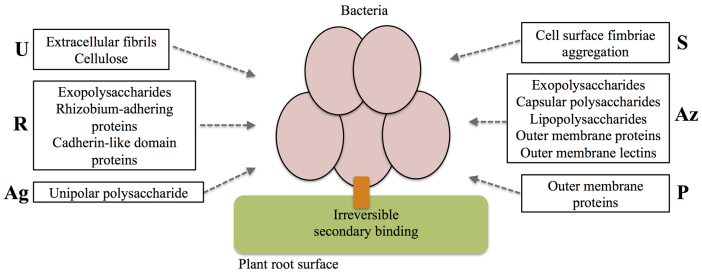
Microbe-specific primary attachment factors, such as outer surface proteins, polysacharrides and flagella, form stronger yet still reversible binding associations with the root (Fig. 1). Within a genus, bacteria can display a range of species-specific and even strain-specific attachment and colonisation mechanisms (Steenhoudt and Vanderleyden 2000; Berne *et al.* 2015). Attachment is affected by a range of soil and root physiochemical properties, such as pH, Ca^{2+} and Mg^{2+} concentrations, water availability and pre-treatment of roots (Rodríguez-Navarro, Dardanelli and Ruíz-Saínz 2007). This is exemplified well in *Rhizobium leguminosarum*, a rhizobial species that forms nitrogen-fixing symbioses with legumes such as pea, vetch, lentil and faba bean. *Rhizobium leguminosarum* utilises primary attachment mechanisms dependent on soil pH (Fig. 1). In acidic soil, the rhizobial surface polysaccharide glucomannan binds root hair expressed lectin (Laus *et al.* 2006). Lectins are carbohydrate-binding proteins, found ubiquitously in nature, that function as recognition molecules in cell–cell interactions across a wide variety of biological systems to mediate binding in a reversible and highly specific manner (Sharon and Lis 2004; Lagarda-Diaz, Guzman-Partida and Vazquez-Moreno 2017). This lectin-mediated mechanism of primary attachment is also used in *Bradyrhizobium japonicum* attaching to soybean roots (Lodeiro and Favelukes 1999; Lodeiro *et al.* 2000). Considering the phylogenetic separation between *B. japonicum* and *R. leguminosarum* and the high lectin content in leguminous species (Lagarda-Diaz, Guzman-Partida and Vazquez-Moreno 2017), this lectin-mediated attachment mechanism is likely utilised in other *Rhizobium*–legume symbioses too.

Under alkaline conditions, root lectins are solubilised, and bacterial polysaccharides such as glucomannan no longer bind (Laus *et al.* 2006). The mechanism of attachment under alkaline and neutral pH might involve an extracellular rhizobial protein 'rhicadhesin', although this is unproven. Rhicadhesin was postulated to mediate attachment between the rhizobial cell and root hair surface in a calcium(Ca^{2+})-dependent manner, and similarly to the dispersal of plant lectin under alkaline pH, rhicadhesin is suggested to disperse under acidic pH. Low Ca^{2+} conditions produce a significant reduction in the attachment of rhizobia to pea root hair surfaces (Smit, Kijne and Lugtenberg 1987), and this is believed to be due to the presence of a Ca^{2+}-dependent attachment protein. Rhicadhesin was reported to be a 14-kDa Ca^{2+}-binding protein that inhibits rhizobial attachment to pea and is secreted by cells growing in a low-[Ca^{2+}] medium (Smit *et al.* 1989). However, 30 years past these original findings, neither the gene encoding rhicadhesin nor the protein itself has been identified. It is possible multiple copies of the gene exist, and this might explain difficulties in its identification (Matthyssee 2014), but the role rhicadhesin plays in attachment remains unclear (Downie 2010; Matthyssee 2014). It can be argued that because the protein has been defined by its ability to inhibit attachment, this certainly supports a role for such a protein in the stability and structure of the cell outer surface, but not necessary a direct role in attachment (Matthyssee 2014). More evidence is needed to better define a role for a rhicadhesin in attachment, and this requires identification of the encoding gene(s). This should now be possible by peptide sequencing of the 14 kDa fraction of rhizobial surface proteins.

Primary attachment in *Agrobacterium*

The importance of characterising root attachment in *A. tumefaciens* is highlighted by the strong correlation between attachment to root cells and susceptibility of the plant to crown gall disease (Hawes and Pueppke 1987), and that direct physical contact via attachment is essential to facilitate DNA transfer into cells (Penalver *et al.* 1996). *Agrobacterium* possess a very closely related polysaccharide to glucomannan in the form of a unipolar polysaccharide (UPP) adhesin. However, a key difference between the two is that UPP mediates an irreversible polar attachment in *A. tumefaciens*, and glucomannan mediates a reversible lectin-receptor attachment in *R. leguminosarum*

om https://academic.oup.com/femsre/article-abstract/42/4/448/4975273
versity user
er 2018



Universal	U	*Agrobacterium*	Ag	*Pseudomonas*	P
Rhizobium	R	*Azospirillum*	Az	*Salmonella*	S

Figure 2. Molecular mechanisms of secondary attachment allowing stronger and irreversible binding to the root, and characterised by the synthesis of cellulose and other extracellular fibrils aiding the accumulation of bacterial aggregates. Universal attachment factors and microbe-specific attachment factors are indicated: U; Universal, R; *Rhizobium*, Ag; *Agrobacterium*, Az; *Azospirillum*, P; *Pseudomonas*, S; *Salmonella*.

(Tomlinson and Fuqua 2009; Matthyssee 2014). The irreversible association mediated by UPP gives it a role in secondary attachment, and primary attachment is facilitated by a combination of surface proteins, molecular adhesins, pili and capsular polysaccharides (Rodríguez-Navarro, Dardanelli and Ruíz-Saínz 2007). The Ca^{2+} adhesion rhicadhesin is also proposed to function in primary attachment in *A. tumefaciens* (Smit *et al.* 1989; Swart *et al.* 1993); however, genetic or biochemical support is yet to be shown and so similarly to rhizobia a defined role remains unclear.

The Ti or 'tumour-inducing' plasmid of *Agrobacterium* contains a large part of the genetic equipment for transferring DNA to plant cell (Christie and Gordon 2014). On this plasmid are virulence '*vir*' genes essential to gene transfer during plant infection, and within this group is the VirB operon that plays a role in primary attachment (Tomlinson and Fuqua 2009). VirB forms the T-pilus that functions as an adhesive pili and initiates physical contact with the root surface (Fig. 1) (Fullner, Lara and Nester 1996; Christie *et al.* 2005; Backert, Fronzes and Waksman 2008). VirB localises at the cell poles and the T-pilus is visualised to extend from these poles (Lai and Kado 2000; Atmakuri *et al.* 2007); this system mediates polar attachment (Tomlinson and Fuqua 2009). However, the T-pilus is not essential for attachment as strains unable to produce this structure or even deficient in the Ti plasmid entirely are still able to attach to plant tissue surfaces in an indistinguishable manner from the wild type (Danhorn *et al.* 2004; Ramey *et al.* 2004; Tomlinson and Fuqua 2009). This points to the presence of attachment elements encoded separately from the Ti plasmid, such as chromosomally encoded polysaccharides and adhesive proteins (Reuhs, Kim and Matthysse 1997). Intriguing, the cryptic plasmid (pAT) of *A. tumefaciens* is not required for attachment or virulence of the strain, but mutation of two genes (attachment 'att' genes *attC* and *attG*) on the plasmid has a dominant negative effect that obliterates primary attachment, causing the bacteria to become avirulent and non-attaching on both carrot and tomato plants (Matthysse, Jaeckel and Jeter 2008). The mechanism by which attachment and virulence is blocked by these mutations is unclear (Matthysse, Jaeckel and Jeter 2008).

Primary attachment in *Azospirillum*

The majority of *Azospirillum* strains can associate with the root surface but only a few strains are able to colonise the inner root as endophytes (Patriquin, Döbereiner and Jain 1983; Döbereiner, Baldani and Reis 1995). The importance of attachment in PGPR interactions is well studied in *Azospirillum* (Rodríguez-Navarro, Dardanelli and Ruíz-Saínz 2007). Within this genus, the majority of research is focused on *Azospirillum brasilense* and its ability to promote the growth of numerous agriculturally important crop plants, including cereals such as wheat. Whilst azospirilla are able to promote plant growth as free-living soil bacteria by secreting plant phytohormones into the rhizosphere, their attachment to roots maximises their mutual benefit (Burdman *et al.* 2000a,b).

The primary attachment phase of *Azospirillum* is commonly referred to as the 'adsorption phase'. The polar flagellum of *Az. brasilense* is required for initial adsorption to roots, alongside its role in motility (Fig. 1) (Croes *et al.* 1993; Rodríguez-Navarro, Dardanelli and Ruíz-Saínz 2007; Mora *et al.* 2008). Flagella-deficient mutants are impaired in attachment to wheat roots, and the purified polar flagellum binds directly to the wheat root surface (Croes *et al.* 1993). Flagellin is a major component of the polar flagellum. The glycosylation of flagellin is suggested to be carried out by the same genes involved in the biosynthesis of the sugars in lipopolysaccharide (LPS), another cell surface moiety (Rossi *et al.* 2016). Mutation of these genes affects the structure of polar flagellin, the production of LPS and the primary attachment abilities of *Az. brasilense* (Rossi *et al.* 2016). Outer membrane proteins (OMPs) are involved in both the primary and secondary attachment phases in *Azospirillum* (Fig. 1; Fig. 2). OMPs are constitutively present as surface-exposed proteins and involved in initial adsorption onto the root, then their surface-exposed domain enables them to function in cell aggregation through interaction with the surface domains of neighbouring bacteria (Burdman *et al.* 2001). This mode of action is most well characterised for the 47.7 kDa major OMP of *Az. brasilense* (Burdman *et al.* 1999).

Primary attachment in *Pseudomonas*

Models of primary attachment in pseudomonads are relatively poorly characterised compared to later stage secondary attachment. Early research implicates pili have a role in primary attachment in *P. fluorescens* (Vesper 1987) (Fig. 1). However, the nature of the attachment assay means this may be at least partly be due to the importance of pili in motility, opposed to a direct, physical role in attachment itself (Vesper 1987). Both *P. fluorescens*

rom https://academic.oup.com/femsre/article-abstract/42/4/448/4975273
iiversity user
ber 2018

and *P. putida* flagella mutants have colonisation defects on plant surfaces, but again, it is possible this is due to the role of flagella in motility opposed to a direct physical role in attachment. Direct flagella binding studies need to be carried out to clarify this involvement. Initial attachment to plant roots produces a global change in gene expression in *P. putdia*, and isolation and characterisation of these genes shows similarities to those involved in the colonisation of abiotic surfaces (Sauer and Camper 2001). Structural components of flagella, type IV pili and genes involved in polysaccharide biosynthesis are all upregulated in primary attachment (Sauer and Camper 2001).

Outer membrane porin F (OprF) is a major outer membrane porin of pseudomonads and is believed to function as a root adhesion protein in *P. fluorescens* (De Mot and Vanderleyden 1991). OprF demonstrates adhesive properties to wheat, barley, maize and sunflower in short time scale assays of 1 h that suggest it functions in primary attachment (Fig. 1), as this is likely too brief a period for tighter secondary attachments to develop (De Mot and Vanderleyden 1991). OprF also demonstrates adhesiveness to cucumber and tomato roots (Crespo and Valverde 2009). The number of *oprF* mutants able to loosely adhere to the roots is significantly lower than the wild type (Crespo and Valverde 2009), with this loose adherence again implicates a role in primary attachment. Decreased tight adherence to the roots is also observed (Crespo and Valverde 2009), perhaps because reduced primary attachments limits secondary attachments, or because OprF functions in both phases of the attachment mechanism. It is unclear exactly how OprF binds to the root surface and the root surface components it associates with.

Primary attachment in *Salmonella*

Salmonella is widely known for infecting animal species, and *Salmonella* food poisoning in humans is generally believed to be associated with consuming infected animal produce such as meat or eggs. However, *Salmonella* is also able to attach to numerous agriculturally important crop species, the consumption of which can directly infect the grazing animal species and humans (Hernández-Reyes and Schikora 2013; Wiedemann *et al.* 2015). The root attachment mechanism of *Salmonella* is poorly characterised, particularly in comparison to characterisation of the direct infection of animal species (Hernández-Reyes and Schikora 2013; Wiedemann *et al.* 2015). It is believed *Salmonella* is transferred to soil through treatment with a natural, contaminated fertiliser source (such as animal manure) or through contaminated water irrigation (Krtinić, Đurić and Ilić 2010). The percentage of *Salmonella* outbreaks associated with the consumption of contaminated plant produce is on the rise (Barak, Liang and Narm 2008; Nygård *et al.* 2008; Gould *et al.* 2013; Zheng *et al.* 2013), and it is only just becoming apparent the extent to which better understanding this plant–microbe interaction is required.

Salmonella enterica is found in the rhizosphere of numerous crop plants (Berg *et al.* 2005). The bacteria exhibits positive chemotaxis towards the root exudates of lettuce and microarray analysis indicates pathogenicity-related genes are induced (Klerks *et al.* 2007). Flagella are important for the root attachment of *S. typhimurium*; flagella mutants are unable to invade the root junctions of *Arabidopsis thaliana* plants (Cooley, Miller and Mandrell 2003). It is possible this decreased attachment is a result of reduced motility, as opposed to a direct physical role of the flagella in surface adherence. It is unclear which phase of root attachment the flagella may function in, although from the understanding of other rhizobacteria it seems most likely flagella will be involved in initial primary attachment, and then potentially play a role in secondary attachment too.

A transposon mutant screen to identify *S. enterica* genes with a role in attachment to alfalfa roots sheds more light on the root attachment capabilities of *Salmonella* (Fig. 1) (Barak *et al.* 2005). Twenty mutants with reduced adherence were identified, 13 of which are uncharacterised genes (Barak *et al.* 2005). Identified genes with a probable role in primary attachment are within an aggregative fimbral (*agf*) operon, encoding proteinaceous surface fimbriae with the ability to function as bacterial adhesions, and a stationary-phase sigma factor (RNA polymerase, sigma S; *rpoS*) with a number of regulatory roles, including regulation of the *agf* fimbral operon and other adhesive factors such as *Salmonella* pili (Barak *et al.* 2005). Similarly to flagella and pili, fimbriae are filamentous proteinaceous structures that extend from the bacterial cell surface. Fimbriae are shorter and more numerous per cell than pili, and unlike flagella and pili, no do function in active motility (Van Gerven, Waksman and Remaut 2011; Berne *et al.* 2015). Fimbriae and pili are also important factors in the animal pathogenicity of *Salmonella*, and appear to be relatively common primary attachment factors in numerous rhizobacteria. At least 13 fimbral operons are present in the genome of *S. typhimurium*, and it is probable some of these also function in plant root attachment but are yet to be characterised.

MOLECULAR MECHANISMS GOVERNING SECONDARY ATTACHMENT

Secondary attachment is characterised by bacteria binding in a tight and irreversible manner to the roots. Typical of this switch to a stronger and more specific binding mode is the synthesis of extracellular fibrils aiding accumulation of bacterial aggregates and the formation of microcolonies on the root (Fig. 2) (Downie 2010). For many root-attaching bacteria, the synthesis of secondary attachment factors is induced in response to successful primary attachment (Rodríguez-Navarro, Dardanelli and Ruíz-Saínz 2007). Many microbes use secondary attachment simply as a means of secure adhesion to the root surface, while for others it may be a pre-requisite of endophytic root colonisation, pathogenic infection or in the case of *A. tumefaciens*, genetic transfer.

Secondary attachment in *Rhizobium*

Nitrogen-fixing rhizobia exhibit accumulation of additional bacteria at the site of initial adhesion and the synthesis of extracellular fibrils largely constructed from cellulose (Fig. 2). Rhizobia attach specifically to the tips of developing root hairs, a region distinct from the epidermal surface of the root. Root hair binding is important as the first step in legume infection, leading to nodule formation and biological nitrogen fixation (Smit, Kijne and Lugtenberg 1986; Gage 2004). Two types of nodules exist: determinant and indeterminant. Determinant nodules are found on a number of tropical legumes and lose meristematic activity shortly after formation to produce spherical mature nodules. Indeterminant nodules are found on the majority of other legumes and maintain an active meristem that continues to generate new cells; this produces more cylindrical mature nodules separated into distinct developmental zones.

Tight secondary attachment allows infection thread formation when rhizobia become trapped between two root hair cells walls after nodulation signalling induces curling of the root hair (Gage 2004). Extracellular fibrils aid formation of bacterial

rom https://academic.oup.com/femsre/article-abstract/42/4/448/4975273
iversity user
ber 2018

aggregates as a 'cap' on the root hair, a structure that can be likened to a biofilm, but has a very distinct attachment process compared to biofilm formation on an inert surface (Smit, Kijne and Lugtenberg 1987; Laus *et al.* 2006; Williams *et al.* 2008). Secondary attachment factors in *Rhizobium* include cellulose fibrils, polysaccharides and secreted proteins. These secreted proteins commonly contain *Rhizobium*-adhering or Cadherin-like domains such as the *Rhizobium*-adhering protein (Rap) family (Poole 2017). Cellulose fibrils help facilitate tight adherence between rhizobial cells as they form a cap on the root hair (Poole 2017), but are not essential for symbiosis. In *R. leguminosarum*, mutants deficient in cellulose fibril production are still able to form nodules (Smit, Kijne and Lugtenberg 1987). It is likely cellulose fibril formation helps facilitate the colonisation of fast-growing root hairs but is not strictly essential for the symbiosis (Laus *et al.* 2006).

A polysaccharide that plays a significant role in secondary attachment is EPS. EPS is a major component of the cell surface and composed of carbohydrate polymers that demonstrate highly variable composition between strains and species (Downie 2010). Polymers are formed from linear or branched molecules of either a homopolysaccharide sugar or different sugars as a heteropolysaccharide (Balsanelli *et al.* 2014). EPS helps facilitate attachment and cellular aggregation on both abiotic and biotic surfaces. EPS-deficient mutants of indeterminate-nodule forming *R. leguminosarum* have defective root attachment and strongly reduced numbers of infection foci (Williams *et al.* 2008). The EPS of *R. leguminosarum* is known as acidic EPS (Bogino *et al.* 2013), and is composed of octasaccharide repeating units of D-glucose, D-galactose and D-glucuronic acid residues in a 5:1:2 molar ratio (Robertsen *et al.* 1981; Philip-Hollingsworth, Hollingsworth and Dazzo 1989; O'Neill, Darvill and Albersheim 1991; Breedveld *et al.* 1993). A *R. leguminosarum* acidic EPS mutant has an interesting phenotype of near-zero attachment and cap formation at root hairs, but is still able to attach along the boundaries of adjacent root epidermal cells (Williams *et al.* 2008). This suggests that whilst acidic EPS is required for attaching to root hairs, it is not essential for attaching at the boundaries of root epidermal cells (Williams *et al.* 2008). The differing structural physiologies of the root hair versus root epidermis leave them receptive to different attachment mechanisms.

The PrsDE type I protein secretion system is responsible for the export of several secondary attachment proteins in *R. leguminosarum*, including Raps (RapA1, RapA2, RapB, RapC and RapD) and EPS-glycanases (PlyA and PlyB) (Fig. 2) (Russo *et al.* 2006). EPS glycanases cleave cell surface EPS, and *plyB/plyA* mutants produce an EPS of considerably greater length than the wild type (Russo *et al.* 2006). Correct EPS processing is important for attachment in *Rhizobium*. Secretion mutants of *prsD* and *prsE* are unable to develop secure secondary attachment cell aggregations (Russo *et al.* 2006). The *rapA1* gene is the most well-characterised Rap group protein, and encodes a 30-kDa Ca^{2+}-binding protein that localises to the cell pole and promotes cellular aggregation (Russo *et al.* 2006). Rap adhesins are believed to function in agglutination through binding either EPS or capsular polysaccharide (Russo *et al.* 2006). The overexpression of RapA1 in *R. leguminosarum* bv. *trifolii* enhances attachment to its host legume by up to 5-fold (Mongiardini *et al.* 2008), and also enhances attachment to non-cognate host plants, soybean and alfalfa (Mongiardini *et al.* 2008). Increased expression of *rap* genes also enhances the root attachment of *R. leguminosarum* bv. *viciae* to pea (Frederix *et al.* 2014).

Secondary attachment in *Agrobacterium*

A polysaccharide-containing adhesive structure known as UPP mediates secondary attachment in *A. tumefaciens* (Fig. 2) (Heindl *et al.* 2007). UPP localises to the cell poles of *Agrobacterium*, and facilitates an irreversible and polar attachment that has low specificity and is utilised in binding to both biotic and abiotic surfaces (Tomlinson and Fuqua 2009; Matthyssee 2014). UPP is only observed upon surface contact (Berne *et al.* 2015), and it has recently been proposed that UPP secretion is stimulated by elevated intracellular c-di-GMP (Xu *et al.* 2013). UPP is a very closely related polysaccharide to glucomannan of *R. leguminosarum* (Laus *et al.* 2006). N-Acetyl glucosamine has been identified as one of the components of UPP; polarly attached *A. tumefaciens* cells stain with fluorescent wheat germ agglutinin that specifically binds N-acetyl glucosamine residues in polysaccharides (Tomlinson and Fuqua 2009). N-Acetyl glucosamine is a monosaccharide derivative of glucose, and this agrees with carbohydrate analysis of glucomannan showing it to consist of almost exclusively glucose and mannose (Laus *et al.* 2006). There are likely additional sugars besides N-acetyl glucosamine, and these remain to be defined.

The key difference between the UPP and glucomannan is the irreversible and reversible attachment they respectively facilitate. However, both associations do provide sites for the attachment and aggregation of increasing numbers of bacteria, and this is facilitated by cellulose fibrillation (Heindl *et al.* 2007). The cellulose synthase gene (*celA*) of *A. tumefaciens* shares a high degree of homology with the cellulose synthase of rhizobia and other Alphaproteobacteria (Matthyssee 2014). *Agrobacterium tumefaciens* and *Rhizobium* share a common mechanism of cellulose synthesis that produces microfibrils scattering from multiple points over the cell surface (Matthyssee 2014). Cellulose fibrils binding tightly to each other, and result of this is the formation of bacterial aggregates on the plant surface. The extent of similarities between the attachment of *Rhizobium* and *Agrobacterium* is unsurprising considering their close phylogenetic relationship. It is characteristic of cellulose production in *A. tumefaciens* that bacteria in the cellulose-bound aggregates are 'tangled' in random orientations (Matthyssee 2014). Fibril formation helps strengthen attachment but is not essential; cellulose synthesis-deficient mutants demonstrate weakened but still functional binding to plant surfaces (Matthysse 1983; Matthysse and McMahan 1998). The synthesis of cellulose and UPP are believed to be integrated through c-di-GMP regulation (Matthyssee 2014), and both play prominent roles in the secondary attachment phase of *A. tumefaciens*.

Agrobacterium strains defective in UPP synthesis are no longer able to bind in the prominently polar wild-type fashion (Matthyssee 2014). Although a UPP-deficient mutant can still infect a wide range of plants (Tomlinson and Fuqua 2009), the ability to attach is much reduced (Matthyssee 2014). Further investigation into the UPP-deficient mutants shows they are removed from the plant surface by water washing at time points where the wild type is bound irreversibly (Lippincott and Lippincott 1967; Sykes and Matthysse 1986). Nevertheless, an increased incubation time with the root surface allows secondary attachment and extracellular fibril production to later proceed (Lippincott and Lippincott 1967; Matthysse 1983; Neff and Binns 1985), and this points to the presence of an alternative secondary attachment mechanism that allows attachment in a non-polar and UPP-independent manner. Factors involved in UPP-independent attachment remain unclear, but it is evident that the binding of UPP and synthesis of extracellular

from https://academic.oup.com/femsre/article-abstract/42/4/448/4975273
niversity user
mber 2018

fibrils, such as cellulose, constitutes the major mechanism of secondary attachment in *Agrobacterium* species. Separate attachment factors proposed include other exo- and capsular polysaccharides (Reuhs, Kim and Matthysse 1997), other elements on the Ti plasmid, '*att*' gene proteins (Matthysse and McMahan 1998) and proteinaceous fimbrae (Penalver *et al.* 1996; Matthyssee 2014).

Secondary attachment in *Azospirillum*

Secondary attachment in *Azospirillum* is similarly characterised by bacterial aggregates forming on the root surface and associated with the biosynthesis of a polysacchardidic fibrillar material (Fig. 2) (Jofré, Lagares and Mori 2004). The outer surface of *Azospirillum* is composed of two major types of polysaccharide: CPS, a dense capsular polysaccharide that binds tightly to the cells, and EPS, a lighter form that extends out from the CPS (Mora *et al.* 2008). Secondary attachment factors in *Azospirillum* include a combination of EPS and CPS, particularly arabinose-rich variants, as well as LPS, outer membrane proteins (OMPs) and outer membrane lectins (OMLs) (Michiels, Croes and Vanderleyden 1991; Del Gallo and Fendrik 1994; De Troch and Vanderleyden 1996).

The EPS of *Az. brasilense* is proposed to function in cell-to-cell adhesion by acting as either a non-specific flocculant or through forming specific interactions with components of neighbouring bacterial cell envelopes (Burdman *et al.* 1998). EPS synthesis is induced as part of secondary attachment, and analysis of cell aggregation ability and EPS production in different *Az. brasilense* strains shows EPS concentration and composition directly influences the extent of cell aggregation (Burdman *et al.* 2000). Of the monosaccharide components of EPS, L-arabinose plays an especially prominent role (Burdman *et al.* 2000). Arabinose is a component of both EPS and CPS, and the concentration of arabinose in both these cell surface components has a positive correlation with the extent of cell attachment and aggregation (Burdman *et al.* 2000).

LPS is a significant structural component of the outer membrane in Gram-negative bacteria, and is exposed on the cell surface in *Azospirillum* (Schloter *et al.* 1994). LPS is a tripartite molecule consisting of a lipid joined to a polysaccharide 'O-antigen' component by core oligosaccharides. The O-antigen is generally a large and very repetitive polysaccharide domain. Correct structure of these is important for secondary attachment in *Azospirillum* (Balsanelli *et al.* 2013). An LPS mutant in both *Az. brasilense* and *Herbaspirillum seropedicae* with an altered core structure has impaired attachment to maize roots and diminished root colonisation (Jofré, Lagares and Mori 2004; Balsanelli *et al.* 2013). The addition of N-acetyl glucosamine inhibits *Azospirillum* attachment to maize roots, and the N-acetyl glucosamine residues of the LPS O-antigen are proposed to directly bind maize root lectins (Balsanelli *et al.* 2013).

OMPs and OMLs also contribute to the secondary attachment phase in *Azospirillum* (Burdman *et al.* 1999). As previously discussed, constitutively present OMPs are believed to facilitate both primary attachment and cell aggregation (Burdman *et al.* 2001). An OML has recently been identified in *Az. brasilense* that specifically recognises and binds to bacterial EPS (Mora *et al.* 2008). This 67-kDa OML is believed to mediate adhesion between attaching azospirilla cells through EPS bridges (Mora *et al.* 2008), and may provide an early step in the formation of larger aggregates on the root surface. Immobilised EPS used in affinity column experiments demonstrates both high affinity and specificity in the binding between EPS and this OML, and also suggests that the lectin targets a complex oligosaccharide structure as opposed to specific single monosaccharides (Mora *et al.* 2008).

Secondary attachment in *Pseudomonas*

There is considerable variety in the transition between primary and secondary attachment across pseudomonads. These differences may stem from differences in the nature of the plant–microbe interaction and differences in the surface chemistry and saturation levels of plant species rooting systems. First, there are differences in the root colonisation phenotype. *Pseudomonas putida* establishes stable biofilm structures on the root surface, and these reach a maximal size relative to the root after which any further growth is subsequently coupled with growth of the root (Sauer and Camper 2001). In comparison, *P. fluorescens* discontinuously colonises the root surface and develops as a smaller biofilms along epidermal fissures (Bloemberg *et al.* 2000; Bloemberg and Lugtenberg 2004), and finally, *P. syringae* is observed to form dense, merging colonies attached to the surface (Bais, Fall and Vivanco 2004).

An attachment model is proposed in *P. putida* in which two large adhesion proteins (LapF and LapA) act in an orchestrated sequence to facilitate the initiation of secondary attachment and then microcolony formation (Martínez-Gil, Yousef-Coronado and Espinosa-Urgel 2010). LapA is believed to drive a transition from reversible, polar attachment to irreversible attachment on the root surface. LapF production is then initiated during these early surface interactions, and the secreted protein mediates cell–cell interactions and is required to drive the transition from singularly attached cells to microcolony assembly. Indeed, LapF-deficient mutants have competition defects in the rhizosphere and cannot advance beyond single-cell attachments to form aggregated colonies (Fuqua 2010; Martínez-Gil, Yousef-Coronado and Espinosa-Urgel 2010). LapF and LapA share commonalities in adhesive properties, size, secretion, cell-surface localisation and both contain expansive repeat domains, but, however, do not possess any substantial sequence identity (Fuqua 2010). LapA is conserved in both *P. putida* and *P. fluorescens* (Yamamoto *et al.* 2000), although differing in size between the two species (Fuqua 2010), and it is absent in pathogenic *P. syringae* (Hinsa *et al.* 2003). The adhesion domain of LapA is similar to the rhizobium-adhering domain of Rap and cadherin proteins in *Rhizobium* species (Ramey *et al.* 2004).

Alongside this model of cell surface protein driven attachment, extracellular fibril production, in particular cellulose, is identified as contributing to this secondary attachment phase in *Pseudomonas* species (Fig. 2) (Cannon and Anderson 1991; Spiers *et al.* 2003; Matthyssee 2014). In contrast to *Rhizobium* and *Agrobacterium*, cellulose fibrils in *P. fluorescens* are synthesised as a linear array of sites localised to a single side of the cell to form a sheet (Cannon and Anderson 1991; Spiers *et al.* 2003; Matthyssee 2014). The geometry of cellulose production is an important factor in influencing the phenotype of bacterial aggregation, and the differences in production geometry are observed to correlate with coding differences in *celB* (Matthyssee 2014).

Secondary attachment in *Salmonella*

Secondary attachment is comparatively poorly characterised in *Salmonella*. Whilst attachment to plant leaves is reasonably characterised, it is unknown the extent to which root attachment is similar. Cellulose is involved in facilitating secondary attachment to roots (Charkowski *et al.* 2002; Barak *et al.* 2007),

rom https://academic.oup.com/femsre/article-abstract/42/4/448/4975273
iversity user
ber 2018

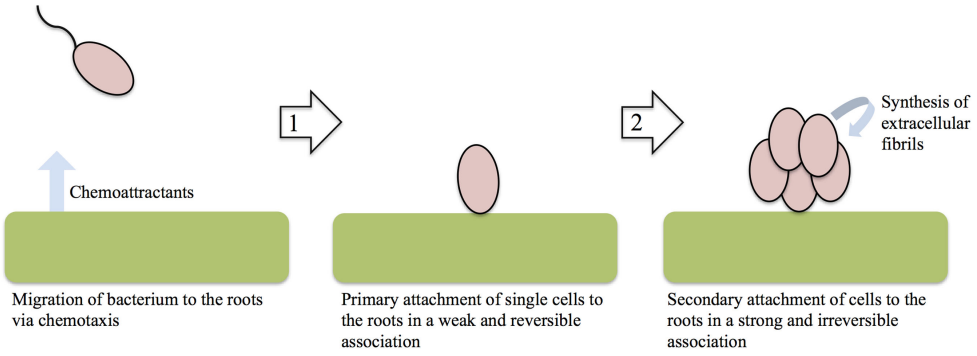


Figure 3. The biphasic model of root attachment. Bacteria migrate towards the roots via chemotaxis in response to root exuded chemoattractants. Step 1 of the biphasic model is the primary attachment phase, where a bacterium attaches to the root in a weak and reversible manner. Root association is initially dictated by physiochemical and electrostatic attractions between the surface molecules of the root and bacterial cell envelope. Stronger microbe-specific primary attachment factors (Fig. 1) may then bind to the root surface and will dictate the overall mechanics of primary attachment. Step 2 of the biphasic model is the secondary attachment phase, where strong and irreversible binding between the bacterium and the root surface occurs. There is the synthesis of extracellular fibrils, cellulose and microbe-specific secondary attachment factors (Fig. 2) and the formation of bacterial aggregates.

similarly to the secondary attachment mechanisms observed in *Rhizobium*, *Agrobacterium*, *Azospirillum* and *Pseudomonas*. The role of cellulose in *Salmonella* root colonisation is probably likewise in cell–cell interactions and aggregation (Barak *et al.* 2007). A recent transposon mutant screen carried out to identify attachment factors in *S. enterica* suggests a number of potential secondary attachment candidates (Fig. 2) (Barak *et al.* 2005). This includes a cell surface aggregate fimbria nucleator (encoded by *agfB*) that may be important in cell aggregation on the root surface and the RpoS, which has a number of regulatory roles, the most relevant of these being in cellulose production. Considering what is known for other rhizobacteria, particularly the more closely related pseudomonads, it is likely we are yet to identify a number of other cell surface proteins and polysaccharides that are involved in *Salmonella* secondary attachment.

DRAWING TOGETHER A COMMON BIPHASIC MODEL OF ROOT ATTACHMENT

In terms of drawing together a common biphasic model of root attachment (Fig. 3), there is the weak and reversible primary attachment phase, followed by the stronger and more specific secondary attachment phase, that is generally accompanied by extracellular fibril synthesis and the formation of bacterial aggregates on the root. Microbes typically use surface appendages such as flagella or pili to migrate towards the root (Fig. 3). Initial primary attachment is dictated by weak electrostatic and physiochemical binding forces developing between surface molecules of the root and bacterial cell envelope (Berne *et al.* 2015). However, these charge-based interactions are only really important on a very early timescale, and significantly stronger associations develop through microbe-specific primary attachment factors, such as surface polysaccharides, proteins, OMPs, flagella and pili (Fig. 1). EPS is an example of a cell surface constituent that can influence both the primary and secondary attachment phases across different bacterial species. The type of EPS present will modify cell surface electrostatic, hydrophobic and steric properties, all of which influence the initial primary interactions, and the polysaccharide is also able to help facilitate cell aggregation and irreversible binding to the surface.

Primary attachment is mediated by constitutively present cell surface components, whose protrusive domains form bonds with components of the root cell surface (Fig. 1). For example, glucomannan of *R. leguminosarum* binds to lectin receptors on pea and vetch roots (Laus *et al.* 2006), and the 47.7-kDa major OMP of *Az. brasilense* facilitates initial adsorption onto wheat roots (Burdman *et al.* 1999). Secondary attachment is commonly facilitated by the induced biosynthesis, secretion or exposure of specific secondary attachment factors and fibrillar material (Fig. 2). For example, this is seen with the stimulation of UPP secretion upon surface contact in *Agrobacterium* (Berne *et al.* 2015), and EPS synthesis is induced as part of the secondary attachment mechanism in *Az. brasilense* (Burdman *et al.* 2000). However, this induction is not always the case, and constitutively present cell surface components, such as OMPs, OMLs and EPS, also function in mediating strong secondary attachment associations and cell aggregation. Cellulose is a universal secondary attachment factor for all microbes discussed within this review, and clearly provides important agglutinating activity in plant root colonisation.

In some species, regulation of the transition from primary to secondary attachment is very well defined, such as the orchestrated sequence in which LapA and LapF function in *P. putida* (Martínez-Gil, Yousef-Coronado and Espinosa-Urgel 2010). This understanding of LapA and LapF is one of the most significant recent advances in plant–microbe attachment research. In other species, it is less evident how secondary attachment factors are induced, although in the case of *Agrobacterium*, surface contact and c-di-GMP signalling certainly appear to play a role (Matthyssee 2014). It seems logical in all cases that either secondary attachment associations will form through constitutively present cell surface factors or surface contact as either 'sensed' by the binding of primary attachment factors or through another cell surface component will trigger signalling to induce biosynthesis, secretion or exposure of specific secondary attachment factors.

Alongside a role in motility, flagella and pili can facilitate adhesion, and this has been identified independently in

rom https://academic.oup.com/femsre/article-abstract/42/4/448/4975273
iversity user
ber 2018

numerous species of *Agrobacterium*, *Azospirillum*, *Pseudomonas* and *Salmonella*. Flagella and pili mutants are commonly used to investigate a role in attachment, but in this research it is not always entirely clear whether the structures have a direct physical role in attachment themselves, or if they are required for moving towards the roots. However, many studies are able to demonstrate a direct physical role in attachment, such as with the purified polar flagella of *Az. brasilense* binding directly to the wheat root surface (Croes *et al.* 1993). This role could perhaps be better distinguished by examining motor mutants (*mot*) that have normal flagella that cannot rotate and so are unable to swim, but should still possess any adhesive domain functions. Although it has not yet been shown for all rhizobacteria that flagella and pili possess this duel role and function directly in root attachment, the fact that they function in this adhesive manner in a wide range of other strains suggests it is possible it is a universal trait.

Flagella and pili are known to play a direct adhesive role in attachment to both abiotic surfaces and in animal colonisation across wide range of Gram-negative and Gram-positive bacteria (Berne *et al.* 2015). It has been well characterised for attachment to abiotic surfaces that pili are commonly involved in the lifestyle switch between motile and attached, and this has been studied in depth with the chaperone-usher pathway pili of *Escherichia coli* and the tight adhesion (Tad) pili of *Caulobacter crescentus* (Pratt and Kolter 1998; Entcheva-Dimitrov and Spormann 2004). The motility role of flagella and pili in this review is focused on the 'swimming' of cells towards the roots, but these appendages can also mediate a coordinated 'swarming' of bacterial populations on solid surfaces (Jarrell and McBride 2008). More recently, the specific role of surface swarming in root colonisation has began to be investigated (Verstraeten *et al.* 2008; Dietel *et al.* 2013), and it appears both the swimming and swarming functions of these appendages play important roles in facilitating attachment and colonisation. Swarming is likely more important in plant–microbe interactions involving biofilm formation on the roots as opposed to interactions involving an endophytic or infection-based colonisation.

EXPANSION ACROSS OTHER SPECIES

The five bacterial genera discussed in this review spread across a range of agriculturally important interactions for species whose root attachment mechanisms have received some of the most significant research attention to date. As previously mentioned, the vast majority of root attachment research has thus far been carried out in plant–microbe interactions involving Gram-negative bacteria (Rodríguez-Navarro, Dardanelli and Ruíz-Saínz 2007; Bogino *et al.* 2013; de Jesus Sousa and Olivares 2016), and the root attachment mechanisms of Gram-positive species has only more recently become a bigger focus of attention (Viaene *et al.* 2016; Habib *et al.* 2017). Abiotic biofilm formation has generally been better characterised in species of plant-interacting Gram-positive bacteria, and is shown to follow the same biphasic mechanism of the initial reversible attachment of single cells, followed by irreversible attachment of multiple cells at the site of initial adhesion (Beauregard *et al.* 2013; Viaene *et al.* 2016; Habib *et al.* 2017). However, it is unknown how similar this is to an attachment mechanism that occurs in planta. Many processes in attachment and colonisation seem to be conserved among bacteria (Poole 2017). For instance, numerous species of both Gram-negative and Gram-positive bacteria have been identified with requirements for EPS and cellulose in attaching to plants (Bogino *et al.* 2013; Poole 2017), and pili in Gram-positive bacteria have also been suggested to possess dual roles in both motility and attachment (Mandlik *et al.* 2008).

Bacillus subtilis has long been known to provide plant growth-promoting effects through association with plants in the rhizosphere, and only recently evidence has shown that there is also the formation of root-associated biofilms in response to plant polysaccharide signals, which also serve as a source of sugars for matrix EPS synthesis (Beauregard *et al.* 2013; Habib *et al.* 2017). The biofilm matrix of *Ba. subtilis* consists of EPS and protein fibres (TasA and TapA), in which TasA is believed to play a similar structural role as to the secondary attachment lectins in *Az. brasilense* that form EPS bridges between cells. *Streptomyces* are similarly renowned for their plant growth effects; they are commonly found inhabiting the soil and rhizosphere and are able to produce a variety of bioactive secondary metabolites (Viaene *et al.* 2016). Both classical and new techniques convincingly show that *Streptomyces* attaches to roots, but these attachment strategies are poorly understood (Viaene *et al.* 2016). Attachment to abiotic hydrophobic surfaces can be mediated by extracellular fimbriae constructed of chaplin protein monomers assembling along cellulose fibrils (De Jong *et al.* 2009). Cellulose acts as a scaffold for the bundling of these protein monomers and mutational studies indicate the presence of a cellulose synthase gene to be involved in attachment (De Jong *et al.* 2009). Cellulose has been identified as an important attachment component across the various other rhizobacteria discussed in this review, and it can therefore be hypothesised that cellulose may similarly be required for *Streptomyces* attachment in the biotic context of plant roots. The root attachment mechanism of *Streptomyces* needs to be better understood before fit to the common biphasic mechanism outlined in this review can be confidently assessed.

For many species, research into root attachment does not yet extend past *in silico* genome analysis. For instance, a recent analysis of *Ba. aryabhattai* AB211 was able to suggest that the genome does indeed encode the necessary arsenal required for adhesion to root surfaces (Bhattacharyya *et al.* 2017). The genome possesses a repertoire of EPS biosynthesis and cell surface export genes, which may aid multicellular irreversible adhesion as part of a secondary attachment mechanism, but this remains to be validated *in vivo*. *Escherichia coli* is another microbe whose root attachment mechanism has received research attention, and similarly to *Salmonella*, it has been implicated in numerous food-poisoning outbreaks associated with the consumption of contaminated food (Martínez-Vaz *et al.* 2014; Wiedemann *et al.* 2015). A microarray analysis study looking at the interaction between *E. coli* and lettuce roots found that genes involved in curli formation and biofilm modulation were important for attachment (Hou *et al.* 2012). Curli are extracellular proteinaceous extrusions involved in mediating cell adhesion in both *E. coli* and *Salmonella* (Barnhart and Chapman 2006), and biofilm modulators, extracellular CPS, curli and flagella have all been suggested to be involved in the molecular mechanism by which *E. coli* colonises plant roots (Seo 2013; Martínez-Vaz *et al.* 2014). Parallels have been drawn between the attachment mechanisms utilised by *Salmonella* and *E. coli*, and similarly to *Salmonella*, a primary attachment phase driven by extracellular structures, specifically curli/fimbriae and flagella, has also been suggested in *E. coli* (Martínez-Vaz *et al.* 2014). Current research strongly suggests that genes involved in biofilm formation are important for the root attachment of *E. coli* (Hou, Martínez-Vaz *et al.* 2014); however, the formation of mature biofilms by *E. coli* in natural plant environments has not yet been documented (Martínez-Vaz *et al.* 2014). The utilisation of these biofilm modulator genes

rom https://academic.oup.com/femsre/article-abstract/42/4/448/4975273
iiversity user
ber 2018

would most likely occur as part of a secondary attachment mechanism, although it remains possible that different mechanisms using similar genetic responses are being utilised in plant colonisation.

CONCLUDING REMARKS

Attachment to plant roots is a universally important step in plant colonisation across symbiotic, pathogenic and plant growth-promoting interactions. A common biphasic mechanism of root attachment consisting of a primary and a secondary attachment phase is observed in agriculturally important rhizobacteria species. Despite the fact that attachment capabilities are a central factor in determining microbial competitiveness in root colonisation (Rodríguez-Navarro, Dardanelli and Ruíz-Saínz 2007), much remains unclear about the molecular mechanisms governing root attachment, particularly in Gram-positive bacterial species (Bogino *et al.* 2013; Viaene *et al.* 2016). It is evident that our knowledge of attachment factors is incomplete, for instance, what mediates UPP-independent attachment in *Agrobacterium*? Which factors are involved in primary and secondary attachment in *Salmonella*? Is there a rhicadhesin encoded in *Rhizobium* and *Agrobacterium* genomes? The plant–microbe interactions of *Rhizobium* and *Agrobacterium* have been studied for decades, but for *Salmonella*, root attachment is only recently a focus of attention. Within these plant–microbe interactions, it has proven difficult to identify attachment factors, quantify their role within an *in vivo* system and piece together multifaceted molecular mechanisms. The complex networks that underpin key biological processes are often difficult to tease apart, and for studying root attachment, there are also the technical difficulties in protein purification, problems with the accuracy of targeted mutational studies and limitations in the power of attachment assays.

Looking to the future, a number of recent technologies provide promising avenues to improve our understanding of this important biological mechanism. These include transposon insertion mutation studies to carry out high-throughput whole-genome screening approaches (Perry and Yost 2014; Wheatley *et al.* 2017), Tracking root interaction systems to observe real-time attachment dynamics with high-resolution microscopy (Massalha *et al.* 2017), lux-based reporter systems to measure attachment in a medium-throughput fashion (Pini *et al.* 2017) and *omics* technologies to aid the structural determination of bacterial surface molecules (Rodríguez-Navarro, Dardanelli and Ruíz-Saínz 2007). One limitation is that most commonly used measurements of attachment involve microscopy and colony counting, both of which are labour intensive, time intensive and reasonably susceptible to human error. A quantitative real-time PCR assay to measure *A. tumefaciens* attaching has recently been developed (Petrovicheva, Joyner and Muth 2017). This measure lends itself to various applications, including the more high-throughput ability to test effects of mutations on bacterial adhesion molecules and to compare attachment across numerous plant species. Lux-based reporter systems provide significant advantages in terms of ease and cost-efficiency. They allow spatiotemporal mapping of bacteria attaching to the root, and can be used to give a striking visualisation of the phases of attachment (Pini *et al.* 2017). From drawing together a common biphasic model of root attachment, it is clear that the knowledge acquired in a particular species can provide insight into a vast array of other plant–microbe interactions. Hopefully utilising these new investigative technologies will enable a more complete picture of the molecular mechanisms governing root attachment to be acquired, and from this, a better understanding of how to manipulate these plant–microbe interactions for agricultural benefit.

FUNDING

This work was supported by the Natural Environment Research Council [grant number NE/L501530/1].

Conflicts of interest. None declared.

REFERENCES

Aguilar JMM, Ashby AM, Richards AJ *et al.* Chemotaxis of *Rhizobium leguminosarum* biovar phaseoli towards flavonoid inducers of the symbiotic nodulation genes. *Microbiology* 1988;**134**:2741–6.

Armitage J, Gallagher A, Johnston A. Comparison of the chemotactic behaviour of *Rhizobium leguminosarum* with and without the nodulation plasmid. *Mol Microbiol* 1988;**2**:743–8.

Atmakuri K, Cascales E, Burton OT *et al.* Agrobacterium ParA/MinD-like VirC1 spatially coordinates early conjugative DNA transfer reactions. *EMBO J* 2007;**26**:2540–51.

Backert S, Fronzes R, Waksman G. VirB2 and VirB5 proteins: specialized adhesins in bacterial type-IV secretion systems? *Trends Microbiol* 2008;**16**:409–13.

Bais HP, Fall R, Vivanco JM. Biocontrol of Bacillus subtilis against infection of Arabidopsis roots by *Pseudomonas syringae* is facilitated by biofilm formation and surfactin production. *Plant Physiol* 2004;**134**:307–19.

Balsanelli E, de Baura VA, de Oliveira Pedrosa F *et al.* Exopolysaccharide biosynthesis enables mature biofilm formation on abiotic surfaces by *Herbaspirillum seropedicae*. *PLoS One* 2014;**9**:e110392.

Balsanelli E, Tuleski TR, de Baura VA *et al.* Maize root lectins mediate the interaction with *Herbaspirillum seropedicae* via N-acetyl glucosamine residues of lipopolysaccharides. *PLoS One* 2013;**8**:e77001.

Barnhart MM, Chapman MR. Curli biogenesis and function. *Annu Rev Microbiol* 2006;**60**:131–47.

Barak JD, Gorski L, Naraghi-Arani P *et al. Salmonella enterica* virulence genes are required for bacterial attachment to plant tissue. *Appl Environ Microb* 2005;**71**:5685–91.

Barak JD, Jahn CE, Gibson DL *et al.* The role of cellulose and O-antigen capsule in the colonization of plants by *Salmonella enterica*. *Mol Plant Microbe In* 2007;**20**:1083–91.

Barak JD, Liang A, Narm K-E. Differential attachment to and subsequent contamination of agricultural crops by *Salmonella enterica*. *Appl Environ Microb* 2008;**74**:5568–70.

Beauregard PB, Chai Y, Vlamakis H *et al. Bacillus subtilis* biofilm induction by plant polysaccharides. *Proc Natl Acad Sci USA* 2013;**110**:E1621–30.

Begonia M, Kremer R. Chemotaxis of deleterious rhizobacteria to velvetleaf (Abutilon theophrasti Medik.) seeds and seedlings. *FEMS Microbiol Ecol* 1994;**15**:227–36.

Berg G, Eberl L, Hartmann A. The rhizosphere as a reservoir for opportunistic human pathogenic bacteria. *Environ Microbiol* 2005;**7**:1673–85.

Berne C, Ducret A, Hardy GG *et al.* Adhesins involved in attachment to abiotic surfaces by Gram-negative bacteria. *Microbiol Spectrum* 2015;**4**, DOI: 10.1128/microbiolspec.MB-0018-2015.

Bhattacharyya C, Bakshi U, Mallick I *et al.* Genome-guided insights into the plant growth promotion capabilities of the

rom https://academic.oup.com/femsre/article-abstract/42/4/448/4975273
iiversity user
ber 2018