Investigating the basis of competitive nodulation & bacteroid development in *Rhizobium leguminosarum*



Brandon L Ford

Oriel College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2021

"...the highest function of ecology is

the understanding of consequences"

Frank Herbert

Abstract

Some soil-dwelling bacteria in the genus *Rhizobium* can form symbioses with leguminous plants, including important crop species such as peas and beans. During symbiosis, rhizobial bacteroids inhabiting root nodules fix atmospheric dinitrogen into ammonium, which is exchanged with the host plant for fixed carbon from photosynthesis. Inoculation of leguminous crops with highly effective rhizobia can significantly increase crop yields and reduce inorganic fertiliser usage. However, potential yield benefits are reduced or lost if inoculant strains are unable to compete with indigenous rhizobia.

High-throughput transposon insertion sequencing (INSeq) techniques were used to assess genes for importance in different stages of symbiosis, including poorly understood genetic factors influencing competitive nodulation ability and development of nitrogen-fixing nodules. INSeq classifications were used as starting points for experimental investigation of mutations found to affect abundance of rhizobia in nodules, including mutations which have both positive and negative effects on fitness. Previous results have revealed that timing of infective and developmental events is likely crucial to nodulation ability, hence a novel mutant screening strategy, aiming to identify mutants delayed in these events, was developed using Lux- and GFP-based biosensors then tested in a combined growth and non-invasive imaging system.

Rhizobial mutant strains were generated for 47 genes of interest using established and novel mutagenesis systems. Phenotypes were found for 17 of 35 mutants examined, including seven mutants with significantly impaired competitive nodulation ability, two mutants (ΔRL0339 & ΔRL3565) with enhanced competitive nodulation ability and one mutant (ΔRL3624) showing enhanced growth promotion under nitrogen-limited conditions. The results in this thesis provided the basis for further investigation of nodulation kinetics and characterisation of genes which may be utilised for future development of more effective inoculant strains.

Acknowledgements

This work was funded by the BBSRC via the Oxford Interdisciplinary Bioscience DTP (Grant BB/M011224/1). As an iCASE project, funding was supplemented by the industrial partner Legume Technology Ltd.

I must acknowledge the invaluable inspiration, mentorship and guidance from my supervisor, Prof. Philip Poole. His leadership and example have led to the Poole Lab becoming the best lab group in Oxford and possibly the world. My secondary supervisors and assessment committee, Prof. Gail Preston, Dr Bruce Knight, Prof. Renier van der Hoorn and Prof. Steve Kelly, have also provided astute advice on the project.

All Poole Lab members have commensurably contributed to making my time with the group so enjoyable and fulfilling, but I must especially acknowledge the indispensable inputs from Alison, Vinoy and Rachel on my project, the help from Bea with teaching me the prerequisite pea-growing proficiency, Marcela with the colourful competition assays, Nick with the advanced acetylene reductions, Raphael with his marvellous mutagenesis, Javi with his clever coexpression analysis and Andrzej for his contemplative career counselling. To Hayley, Tim, Laura, Sam, Carmen, Jack, Marta, Paul, Anna, Patrick, Clare, Bergthor and Isidro, I hope our paths will cross again. This work would not have been at all possible without the significant support from Helen and Lida. I am additionally grateful to the staff at Plant Sciences, including Gem, John, and Maria, who have provided all-important assistance during my time at the department.

I would like to thank my housemates Josh, Michael, Carolin, Pippa and Joe for their fabulous friendship and keeping me (mostly) sane during the lockdowns. I will miss living with such awesome people, and all the kitchen chats and film/musketeers/pizza nights with them. The DTP (esp. Nic, Alex, & Thomas), Oxford University Pistol Club (esp. Ioana, Mary, Howell, Thomas & Tatiana) and Oriel College (esp. Ryan, Suzie, Rannya & Dan) have also provided foundational friendships during my time with them and beyond. I must acknowledge Buster and Lucy in particular for their close companionship. I am thankful to my parents, my family and my previous teachers for bringing me so far, allowing me to reach this point.

Contents

Abstract v
Acknowledgements vii
List of Figures xii
List of Tables xv
List of Abbreviations xvi
Publications xx

1 General Introduction **1**
1.1 The Nitrogen Cycle 1
1.1.1 The Nitrogen Crisis 1
1.2 *Rhizobium*-Legume Symbiosis 2
1.2.1 Establishment of Symbiosis 6
1.2.2 Bacteroid Development 7
1.3 Use of Inoculants in Agriculture 10
1.4 *Rhizobium* Competitiveness 12
1.4.1 Impact of Competition on Inoculants 12
1.4.2 Measurement of Competitive Nodulation Ability 13
1.4.3 Improvement of Competitive Nodulation Ability 15
1.5 Transposon Insertion Sequencing 17
1.5.1 INSeq as a Genome-Wide Mutagenesis Screen 19
1.5.2 Experimental Findings From INSeq on the Model Strain Rlv3841 . 21
1.6 General Aims of This Research 23

2 Materials and Methods **25**
2.1 Bacterial Growth Conditions 25
2.2 Plant Growth Conditions 26
2.3 Molecular Techniques 28
2.4 Plant and Microbiological Experiments 33
2.5 Computational Methods 37
2.6 Bacterial Strains 38

2.7 Plant Lines 44
2.8 Plasmids 45

3 Exploration of Mutant Phenotypes Through INSeq Data 50

3.1 Introduction 50
3.2 Genes of Interest 53
3.3 Generation of Mutants with the pK19mob system 62
3.4 Examination of Nodulation Ability by Single-Strain Inoculations 64
3.5 Measurement of Mutant Nitrogenase Activity by Acetylene Reduction Assays 70
3.6 Quantification of Mutant Competitive Nodulation Ability by Co-Inoculation Experiments 72
3.7 Investigation of Mutant Viability From Nodules by Nodule Bacteria Recovery Assays 77
3.8 Measurement of Mutant Growth Rate as Free-Living Bacteria 79
3.9 Discussion 81
3.9.1 Summary 81
3.9.2 Future Perspectives 83
3.9.3 Significance 86

4 Characterisation of Advantaged Mutants 87

4.1 Introduction 87
4.2 Selection of Advantaged Genes For Mutagenesis 88
4.3 Advantanged Genes of Interest 92
4.4 Generation of Advantaged Mutants with the pK19mob System 96
4.5 Measurement of Host Biomass Accumulation as a Proxy for Advantaged Mutant Nitrogen Fixation Ability 96
4.6 Quantification of the Competitive Nodulation Abilities of Advantaged Mutants by Co-Inoculation Experiments 100
4.7 Comparison of Nodule Size as an Indicator of Host Sanctioning 102
4.8 Comparison of Nodule Number as a Proxy for Host Nitrogen Availability . 105

4.9 Characterisation of the Growth of an RL3624 Mutant . . . 107
4.10 Coexpression Analysis of RL3565 Microarray Data . . . 110
4.11 Discussion . . . 120
4.11.1 Summary . . . 120
4.11.2 Future Perspectives . . . 123
4.11.3 Significance . . . 125

5 Mutant Screen to Investigate Nodule Developmental Kinetics 126
5.1 Introduction . . . 126
5.2 Biosensors to Monitor Key Stages in Symbiosis . . . 127
5.3 Selection of Candidate Genes for Mutagenesis . . . 133
5.4 An Improved System for Mutagenesis of *Rhizobium* . . . 137
5.5 Nodulation Kinetics Time-Courses . . . 142
5.6 Monitoring Nodule Development with P*nodA* and Ps*nifH* Biosensors in Concert . . . 148
5.7 Monitoring Nodule Development with P*nodA* and Ps*nifH* Biosensors Individually . . . 150
5.8 Kinetics of Nodulation . . . 153
5.9 Monitoring Nodule Development for Two Further Batches . . . 156
5.10 Discussion . . . 159
5.10.1 New Mutagenesis System with pOPS1670/pOPS1713 . . . 159
5.10.2 Nodule Kinetics Time-Courses . . . 160
5.10.3 Future Perspectives . . . 163
5.10.4 Significance . . . 165

6 General Conclusions and Future Perspectives 166
6.1 Impact of the Covid-19 Pandemic . . . 166
6.2 Context of Research . . . 166
6.3 INSeq Data-Informed Experimental Investigation . . . 167
6.4 Dual Mutagenesis Strategies Were Utilised . . . 169
6.5 Defining the Stage of Symbiosis Affected . . . 170

6.5.1 Identified Phenotypes at the Stage of Competitive Colonisation . . 177
6.5.2 Identified Phenotypes at the Stage of Nodulation Initiation 178
6.5.3 Identified Phenotypes at the Stages of IT and Nodule Bacteria . . . 179
6.5.4 Identified Phenotypes at the Stage of Nodule Bacteroids 180
6.5.5 Identified Phenotypes at the Stage of Free-Living Growth 181
6.6 Future Directions 182
6.7 Conclusions 184

7 Bibliography **185**

8 Appendix **216**
8.1 Wheatley et al., 2020 216
8.1.1 Author Contributions 216
8.1.2 Full Text 216

List of Figures

1 Schematic diagram depicting the establishment of *Rhizobium*-legume symbiosis 4

2 Plasmid map of pSAM_Rl, the vector used to generate mariner transposon mutant libraries in *Rhizobium leguminosarum* 20

3 General overview of the stages of the *Rhizobium* symbiotic cycle 51

4 Integration of pK19mob into *Rhizobium* genome by single-crossover homologous recombination 63

5 Nodule morphology from different strains three weeks post inoculation onto pea 65

6 Nodules per plant from single strain mutant inoculations 67

7 Example plates showing the recovery of bacteria from two surface-sterilised nodules following single-strain inoculation with the *nodB* mutant 69

8 Whole plant nitrogenase activity as measured by acetylene reduction assays 71

9 Double-sequentially stained roots of a pea plant 73

10 Ratio between the proportion of each strain in nodule occupancy and the proportion of that strain in the inocula in competitive nodulation assays . 76

11 Growth of wild type and four mutant strains following recovery of bacteria from nodules 78

12 Growth rate in liquid TY media of wild type and four mutant strains . . . 80

13 Screenshot of the integrated genome viewer (IGV) view of RL2036 89

14 Mean dry weight of pea shoots following inoculation with rhizobial strains 97

15 Mean dry weight of pea shoots following inoculation with rhizobial strains 99

16 Ratio between the proportion of each AD strain in nodule occupancy and the proportion of that strain in the inocula in competitive nodulation assays 101

17 Mean areas of the 5, 10, and 15 largest nodules from wild type and the RL3624 mutant 104

18 Total number of nodules per plant of plants inoculated with the strain mutated in RL3624 and wild type 106
19 Comparison of growth in liquid AMS media, supplemented with one of six carbon sources of wild type and ST1 mutant 108
20 Coexpression matrix of the 80 genes most highly coexpressed with RL3565 in the *standard* dataset 112
21 Coexpression matrix of the 72 genes most highly coexpressed with RL3565 in the *complete* dataset 113
22 Pea roots with nodules occupied by rhizobia carrying Ps*nifH*:*sfgfp* reporter on a multicopy plasmid 128
23 Overview of the dual-biosensor system for monitoring timing of nodule development 129
24 Vetch (*Vicia hirsuta*) seedlings inoculated with Rlv3841 carrying pLMB712 131
25 Plasmid maps from the pOPS1670/pOPS1713 system 140
26 Double homologous recombination of the mutagenic plasmid with the flanking regions of the gene of interest 141
27 Pea growth in pouches during nodule development screen 143
28 Fluorescence microscopy images of nodules occupied by OPS2905 on pea roots 145
29 Onset of nodulation and nitrogen fixation in Batch C strains monitored by P*nodA*:*lux* and Ps*nifH*:*sfgfp* biosensors 149
30 Onset of nodulation and nitrogen fixation in Batch C strains monitored by the Ps*nifH*:*sfgfp* biosensor 151
31 Onset of nodulation in Batch C strains monitored by the P*nodA*:*lux* biosensor 152
32 Kinetics of nodulation from following green fluorescence for strains from Batch C 154
33 Mean number of green fluorescent mature nodules per plant at 14 dpi for strains from Batch C 155

34 Onset of nodulation and nitrogen fixation in Batch B strains monitored by the Ps*nifH*:*sfgfp* biosensor 157

35 Onset of nodulation and nitrogen fixation in Batch D strains monitored by the Ps*nifH*:*sfgfp* biosensor 158

36 Timeline of key experimental and biological milestones during screening experiments performed in Chapter 5 162

37 Mutant phenotypes affecting general growth and four stages of *Rhizobium*-legume symbiosis 171

List of Tables

1 Golden Gate parts used for the construction of background strains OPS2575 and OPS2576 . . . 33
2 Strains used in this work . . . 38
3 Plasmids used in this work . . . 45
4 INSeq classifications of genes selected for mutagenesis and characterisation 54
5 Summary of hypothetical affected stages of symbiosis initiation for mutants . . . 55
6 Summary of the mutant phenotypes observed in the experiments described in Chapter 3 . . . 82
7 INSeq classifications of AD genes selected for mutagenesis and characterisation . . . 90
8 Comparison of growth in liquid AMS media, supplemented with one of six carbon sources of seven strains mutated in sugar transporters and wild type . . . 110
9 INSeq classifications of genes highlighted by coexpression analysis of the *standard* dataset that were most strongly and broadly correlated with the RL3565 according to the three distinct analysis methods . . . 115
10 INSeq classifications of genes highlighted by coexpression analysis of the *complete* dataset that were most strongly and broadly correlated with the RL3565 according to the three distinct analysis methods . . . 117
11 Summary of AD mutant phenotypes observed in the experiments described in Section 4 . . . 121
12 Golden Gate parts used for the construction of background strains OPS2575 and OPS2576 . . . 132
13 INSeq classifications of genes selected for mutagenesis as part of the screen for strains with delayed nodule development . . . 134
14 Division of mutants into experimental batches for screening . . . 146
15 Summary of the INSeq classification and mutant phenotypes . . . 172

List of Abbreviations

Excluding *Système International* units and chemical symbols.

Abbreviation	Description
ABC	ATP-Binding Cassette (transporter)
AD	Advantaged (INSeq classification)
Amp	Ampicilin
AMS	Acid Minimal Salts (media)
ANOVA	Analysis of Variance
ARA	Acetylene Reduction Assay
ATP	Adenosine triphosphate
BBSRC	Biotechnology and Biological Sciences Research Council
Bd	Bacteroid (INSeq library)
bp	Base pair(s)
bv.	Biovar
CD	Community Detection
celB	*Pyrococcus furiosus* β-galactosidase gene
CFU	Colony Forming Units
Cm	Chloramphenicol
Cpl	Complete (coexpression analysis dataset)
CUT	Carbohydrate Uptake Transporter
cv.	Cultivar
DAPI	4',6-diamidino-2-phenylindole
DCA	Dicarboxylic Acid
DE	Defective (INSeq classification)
DNA	Deoxyribose Nucleic Acid
DOPS	Department of Plant Sciences (University of Oxford)
dpi	Days post inoculation
dpL	Days post Luminescence
Dsgn	(Gene) Designation

continued on next page

continued from previous page

Abbreviation	Description
DUF	Domain of Unknown Function
DWA	Distilled Water Agar
EAL	EAL domain (name derived from conserved peptide sequence)
EPS	Exopolysaccharide
ES	Essential (INSeq classification)
FIJI	FIJI Is Just ImageJ
Fix^+	Nitrogen-fixing rhizobial phenotype
Fix^-	Non-fixing rhizobial phenotype
GFP	Green Fluorescent Protein
GGDEF	GGDEF domain (name derived from conserved peptide sequence)
Gm	Gentamycin
GMP	Guanosine monophosphate
GOI	Gene of Interest
gusA	*E. coli* β-glucuronidase gene
HD	HD domain (name derived from conserved histidine and aspartate residues)
*Hin*dIII	*Hin*dIII restriction enzyme
HMM	Hidden Markov Model
IGV	Integrated genome viewer
INSeq	(Transposon) Insertion Sequencing
Km	Kanamycin
LB	Lysogeny Broth (media)
LCO	Lipochitooligosaccharide
LHG	Leghaemoglobin
LPS	Lipopolysaccharide
LT	Legume Technology Ltd
lux	*Vibrio fischeri* luciferase operon
Magenta-GlcA	5-bromo-6-chloro-3-indolyl-β-D-glucuronide
MCS	Multiple Cloning Site

continued on next page

continued from previous page

Abbreviation	Description
*Mly*I	*Mly*I restriction enzyme
*Mme*I	*Mme*I restriction enzyme
µt	Mutant
n	Number of replicates
NAD	Nicotinamide adenine dinucleotide
NADH	Reduced nicotinamide adenine dinucleotide
NB	Nodule Bacteria (INSeq library)
NCR	Nodule Cysteine-Rich (peptides)
ND	Not Determined (phenotype)
NE	Neutral (INSeq classification)
Nm	Neomycin
Nod^+	Nodulating rhizobial phenotype
Nod^-	Non-nodulating rhizobial phenotype
OD	Optical Density
OD_{600}	Optical Density at 600 nm
OPS	Oxford Plant Sciences (strain)
PAS	PAS domain (found in Period, Arnt and Sim proteins)
PCR	Polymerase Chain Reaction
pLMB	Laboratory of Molecular Biology (Cambridge) plasmid
P*nodA*	Promoter sequence of *nodA*
pOGG	Oxford Golden Gate part plasmid
pOPS	Oxford Plant Sciences plasmid
PR	PageRank
pRU	Reading University plasmid
PSM	Phosphate-Sucrose-$MgCl_2$ buffer
Ps*nifH*	Synthetic promoter sequence of *nifH*
r	Resisitant (to antibiotic)
RC	Root Colonisation (INSeq library)

continued on next page

continued from previous page

Abbreviation	Description
RLK	Receptor-Like Kinase
Rlv3841	*Rhizobium leguminosarum* bv. *viciae* 3841
RNA	Ribose Nucleic Acid
RS	Rhizosphere (INSeq library)
RU	Reading University (strain)
s	Sensitive (to antibiotic)
*Sal*I	*Sal*I restriction enzyme
SBP	Solute-Binding Protein
SD	Standard Deviation
SEM	Standard Error of the Mean
sfgfp	Super-folder Green Fluorescent Protein
Spec	Spectinomycin
ST	Sugar Transporter, as in ST1
Std	Standard (coexpression analysis dataset)
Str	Streptomycin
T3SS	Type 3 Secretion System
T4SS	Type 4 Secretion System
TA	Thymine-Adenine motif (Mariner transposon insertion site)
TCA	Tricarboxylic Acid (cycle)
Tet	Tetracycline
TY	Tryptone Yeast (media)
UMA	Universal Minimal Agar (media)
UMS	Universal Minimal Salts (media)
Vm	Vincent's Minimal (media)
WE	Weight Edges
WT	Wild Type
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
X-GlcA	5-bromo-4-chloro-3-indolyl-β-D-glucuronide

Publications

Wheatley et al., 2020

R. M. Wheatley*, **B. L. Ford***, Li Li*, S. T. N. Aroney, H. E. Knights, R. Ledermann, A. K. East, V. K. Ramachandran, & P. S. Poole, Lifestyle adaptations of *Rhizobium* from rhizosphere to symbiosis ***PNAS*** 117, 38 (2020) DOI: 10.1073/pnas.2009094117

Abstract

By analyzing successive lifestyle stages of a model *Rhizobium*–legume symbiosis using mariner-based transposon insertion sequencing (INSeq), we have defined the genes required for rhizosphere growth, root colonization, bacterial infection, N_2-fixing bacteroids, and release from legume (pea) nodules. While only 27 genes are annotated as *nif* and *fix* in *Rhizobium leguminosarum*, we show 603 genetic regions (593 genes, 5 transfer RNAs, and 5 RNA features) are required for the competitive ability to nodulate pea and fix N_2. Of these, 146 are common to rhizosphere growth through to bacteroids. This large number of genes, defined as rhizosphere-progressive, highlights how critical successful competition in the rhizosphere is to subsequent infection and nodulation. As expected, there is also a large group (211) specific for nodule bacteria and bacteroid function. Nodule infection and bacteroid formation require genes for motility, cell envelope restructuring, nodulation signaling, N_2 fixation, and metabolic adaptation. Metabolic adaptation includes urea, erythritol and aldehyde metabolism, glycogen synthesis, dicarboxylate metabolism, and glutamine synthesis (GlnII). There are 17 separate lifestyle adaptations specific to rhizosphere growth and 23 to root colonization, distinct from infection and nodule formation. These results dramatically highlight the importance of competition at multiple stages of a *Rhizobium*–legume symbiosis.

1 General Introduction

1.1 The Nitrogen Cycle

Nitrogen is an essential nutrient for all organisms because it is a major component in the biosynthesis of amino acids (proteins) and nucleotides (DNA, RNA, ATP, NAD; Burén and Rubio, 2018). Despite making up almost 80% of air, atmospheric N_2 is biochemically unavailable to most organisms due to the high stability of the di-nitrogen triple bond. This is why nitrogen is often considered the most limiting nutrient for plant growth (Lupwayi et al., 2006). Biologically available nitrogen is usually in the form of ammonia (NH_3) or ammonium (NH_4^+), products of the limited number of bacterial and archaeal species with the ability to 'fix' atmospheric N_2 into these biologically available forms using the nitrogenase enzyme (reviewed in Galloway et al., 2008). It is estimated that this microbe-mediated process, known as biological nitrogen fixation, fixes approximately 200 Teragrams (Tg) of nitrogen per year, however only 50-65 Tg of this total is directly beneficial to agricultural soils (Fowler et al., 2013; Peoples et al., 2021; Rascio & La Rocca, 2013).

1.1.1 The Nitrogen Crisis

The other major input of fixed nitrogen is from the Haber-Bosch process, which adds approximately 121 Tg of nitrogen into global agriculture per year, being the principal method of nitrogen input in the developed world since the early 20th century (reviewed in Galloway et al., 2008). The long-term sustainability of the Haber-Bosch process has come under question in recent decades, as the energy-intensive process contributes an estimated 1.2% of all anthropogenic CO_2 emissions (Smith et al., 2020). This is due to the Haber-Bosch process requiring high temperatures (400-650 °C) and pressures (200-400

atmospheres) to fix nitrogen and hydrogen into ammonia over an iron catalyst, resulting in the emission of 2.6-3.7 tonnes of CO_2 for every tonne of ammonia produced (Jensen et al., 2012). Furthermore, less than half of the artificial fertiliser applied to fields may be utilised by the crop under standard usage practices, with the remainder leaching and being wasted (Hu et al., 2010, *Zea mays*). This wastage causes environmental damage through leaching into aquatic ecosystems, where fertiliser runoff can cause toxic algal blooms (eutrophication), or through denitrification, a microbe-mediated conversion into harmful greenhouse gases such as N_2O (Clúa et al., 2018; Rockström et al., 2009; Schwenke et al., 2015). Excessive inorganic fertiliser usage has also been found to result in detrimental effects on soil microbiota important for soil and crop resilience (Jangid et al., 2008; Naik et al., 2019).

1.2 *Rhizobium*-Legume Symbiosis

Plants in the family Fabaceae, commonly known as legumes, are some of the world's most important crop plants (reviewed in Graham and Vance, 2003). Legumes contribute up to a third of the protein in the human diet and make up around 27% of global grain production (Smýkal et al., 2015). Important crops for human nutrition in the legume family include pea (*Pisum sativum*), chickpea (*Cicer arietinum*), cowpea (*Vigna unguiculata*), pigeon pea (*Cajanus cajan*), common bean (*Phaseolus vulgaris*), broad bean (*Vicia faba*), soybean (*Glycine max*), and lentil (*Lens esculenta*) among others (Graham & Vance, 2003). Many of these crops are also important sources of nutrition for livestock as forage or fodder, especially soybean, clovers (*Trifolium*, *Melilotus*, and *Medicago* sp.) and lucerne (*Medicago sativa*).

The importance of legumes is partly due to their ability to obtain the fixed nitrogen they require for growth from symbioses with nitrogen-fixing bacteria such as those from the order Rhizobiales, in return for supplying fixed carbon to the bacterial symbionts (Masson-Boivin & Sachs, 2018). It is estimated that around 80% of the biological nitrogen fixation in agricultural areas occurs as part of the *Rhizobium*-legume symbiosis (O'Hara, 1998). This means legume crops do not require as much nitrogen fertiliser as non-legume crops, reducing the impacts of environmental damage from eutrophication of aquatic systems and greenhouse gas emissions associated with fertiliser use, in addition to the economic benefits of not buying expensive fertilisers (Clúa et al., 2018; Rockström et al., 2009). Farmers have made use of the *Rhizobium*-legume symbiosis to improve soil fertility by incorporating legumes into crop rotation practices for thousands of years (Bullock, 1992), although it was first reported in Western scientific literature that legume crops can grow independently of soil nitrogen levels in the late 19th century (Hellriegel & Wilfarth, 1888).

Research on *Rhizobium* began more than a century ago with the first isolation and characterisation of nodule bacteria (*Sinorhizobium meliloti*) from clover nodules by the great Dutch microbiologist and botanist Martinus Beijerinck (Beijerinck, 1888). Naturally, the *Rhizobium*-legume symbiosis is considerably older, believed to have evolved around 100 million years ago and is one of the most complex plant-microbe interactions (Van Velzen et al., 2018). This complexity impedes its study, but it is vitally important to research because of the potential to use the knowledge to improve crop yields (reviewed in Poole et al., 2018).

Figure 1

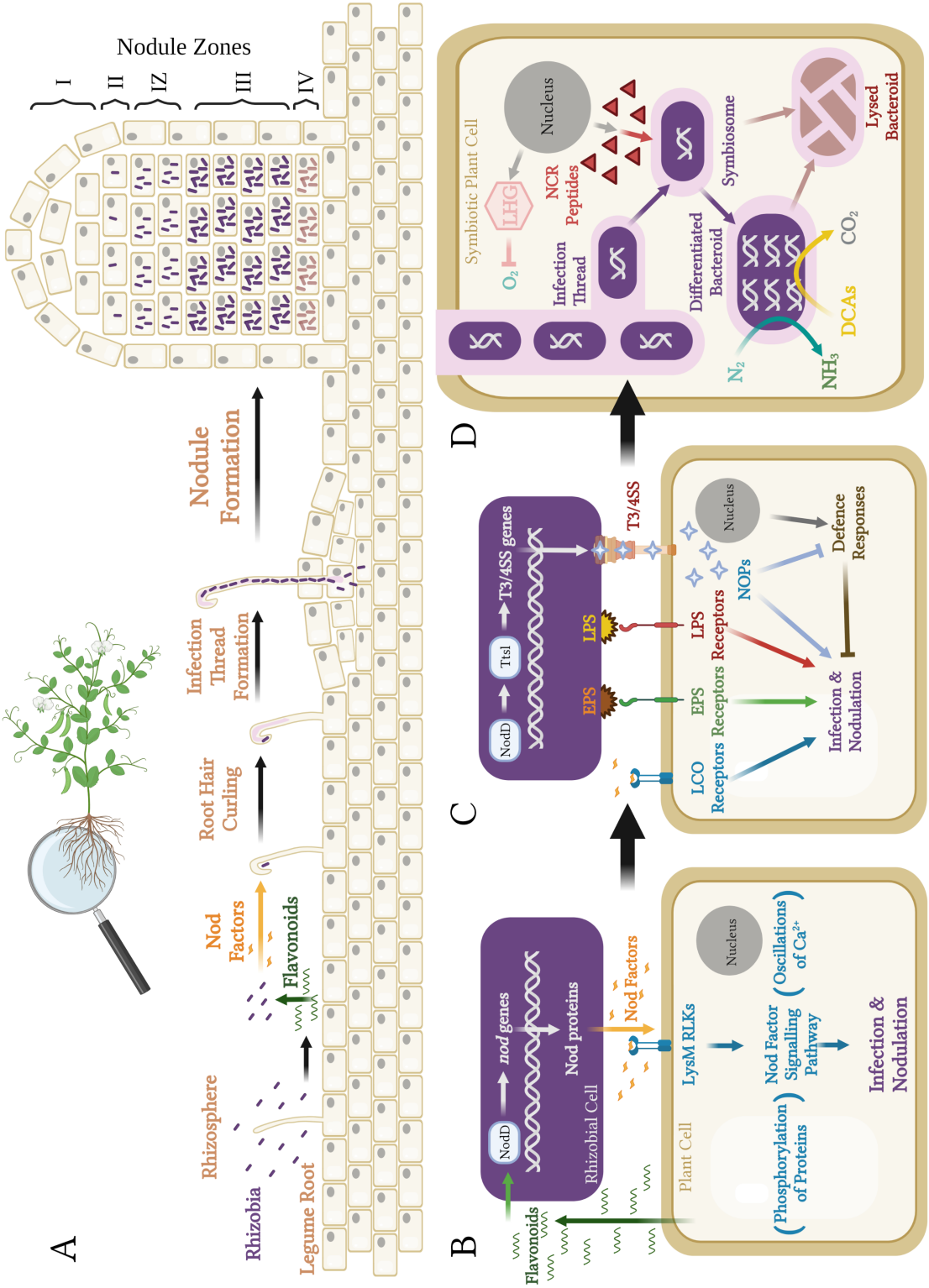

A
Nodule Zones
I
II
IZ
III
IV
Nodule Formation
Infection Thread Formation
Root Hair Curling
Nod Factors
Flavonoids
Rhizosphere
Rhizobia
Legume Root
B
nod genes
NodD
Nod proteins
Rhizobial Cell
Nod Factors
Flavonoids
Plant Cell
LysM RLKs
Nucleus
Nod Factor Signalling Pathway
(Oscillations of Ca^{2+})
(Phosphorylation of Proteins)
Infection & Nodulation
C
T3/4SS genes
NodD
TtsI
T3/4SS
EPS
LPS
LCO Receptors
EPS Receptors
LPS Receptors
NOPs
Nucleus
Defence Responses
Infection & Nodulation
D
Symbiotic Plant Cell
Nucleus
LHG
O_2
NCR Peptides
Infection Thread
Differentiated Bacteroid
Symbiosome
Lysed Bacteroid
N_2
NH_3
DCAs
CO_2

Figure 1. Schematic diagram depicting the establishment of *Rhizobium*-legume symbiosis. (A) The general process by which rhizobia infect legume roots, from flavonoid-induced Nod factor production through Nod-factor-induced root hair curling to infection thread (IT) formation and nodule formation. Nodule zones are as follows: I = meristem zone, II = infection zone, IZ = interzone, III = nitrogen fixation zone, and IV = senescent zone; (B) Initial establishment of *Rhizobium*-legume communication with rhizobial perception of host flavonoids leading to expression of Nod factors, detected by host LysM receptor-like kinases (RLKs) leading to protein phosphorylations and calcium oscillations, root hair curling and nodulation; (C) Host recognition of rhizobial Nod factors (LCOs), exopolysaccharides (EPS), lipopolysaccharides (LPS) in preparation for endocytosis. Rhizobial effector proteins (NOPs) suppress plant immunity and defence responses following injection by type 3 or 4 secretion systems (T3SS/T4SS); and (D) Bacteroid development within symbiosomes after exposure to nodule cysteine-rich (NCR) peptides includes endoreduplication, enlargement and metalobic shifts such as use of dicarboxylic acids (DCAs) as carbon sources and the start of nitrogen fixation, which is dependent on free oxygen in nodules being sequestered by host-provided leghaemoglobin (LGH). Bacteroid/bacterial lysis can occur as a result of NCR peptide toxicity and/or conditions in the senescent zone of the nodule. Adapted from Poole et al., 2018 and Wang et al., 2018. Created with BioRender (biorender.com).

1.2.1 Establishment of Symbiosis

Unlike older endosymbioses such as with chloroplasts and mitochondria, the *Rhizobium*-legume symbioses must be re-established for each new plant generation (Jones et al., 2007). Before symbiosis is established, rhizobia live an oligotrophic lifestyle as saprophytes in soil (Poole et al., 2018). The rhizosphere region around plant roots is comparatively rich in plant root exudates to sustain a plant-influenced microbial ecosystem that includes rhizobia (Philippot et al., 2013; Tkacz et al., 2015; Tkacz & Poole, 2021).

The initiation of symbiosis begins in the rhizosphere, where rhizobia sense plant-produced flavonoids and respond by secreting lipochitooligosaccharides (LCOs) known as Nod (nodulation) factors (Geurts & Bisseling, 2002; Jiménez-Guerrero et al., 2018; Venturi & Keel, 2016). This exchange forms the first part of an inter-domain chemical conversation between legume and rhizobia, which is followed by the perception of Nod factors by LysM receptor-like kinases (Oldroyd & Downie, 2004). The specificity of Nod factor binding to LysM receptor-like kinases contributes to the host-symbiont specificity of the symbiosis, as different rhizobial strains are able to nodulate different legume species (Jorrín, 2016; Wang et al., 2018). The general process of symbiosis initiation is illustrated in Figure 1.

Legumes cannot directly assess the nitrogen fixation ability of potential rhizobial partners before allowing infection and nodule formation (Westhoek et al., 2017). This has implications for the evolutionary stability of the symbiosis, as it is in the host's interest not to reward non-fixing 'cheater' rhizobial strains (Benezech et al., 2020; Chomicki et al., 2020; Sachs et al., 2018). The long-term stability of the symbiosis is driven by the restriction of host resources to under-fixing nodules ('sanctioning') and host specificity to rhizobial

signals associated with mutualist nitrogen-fixing strains (Liu & Murray, 2016; Walker et al., 2020; Westhoek et al., 2021). Similar restrictions on resource provision and further restrictions on nodule development occur when the host legume has access to sufficient mineral nitrogen (Ferguson et al., 2019; Lambert et al., 2020).

Recognition of Nod factors by the host plant activates a complex and incompletely understood signalling pathway that includes protein phosphorylation, secondary messengers and Ca^{2+} oscillations, eventually leading to the plant-side morphological changes that result in the formation of nodule structures (Figure 1; Oldroyd and Downie, 2004). For the infecting rhizobia, chemoattraction to the legume root precedes colonisation, attachment to the plant root surface (rhizoplane) and biofilm formation, where Nod factor-induced root hair curling can trap the bacteria in an infection pocket (Downie, 2010; Knights et al., 2021; Yost et al., 1998). Plant cell membrane invagination forms an infection thread which rhizobia grow down to infect the growing nodule, while further signalling interactions prevent the plant immune system from killing the rhizobia (Figure 1C; Berrabah et al., 2019; Jones et al., 2007; Zipfel and Oldroyd, 2017).

1.2.2 Bacteroid Development

Once inside the nodule, rhizobia are engulfed by cortical cells to form symbiosomes, where one or more bacteria are surrounded by a plant-derived symbiosome membrane. Nodule Cysteine-Rich (NCR) peptides produced by the host plant cell drive the differentiation of rhizobia into nitrogen-fixing bacteroids, resulting in significant changes to rhizobial metabolism (Pan & Wang, 2017; Prell & Poole, 2006). Bacteroids shift from the oligotrophic metabolism of free-living bacteria to the utilisation of carbohydrate-based

photosynthates from their host, while losing the ability to synthesise several amino acids, which are provided by the host plant (Udvardi & Poole, 2013).

Most temperate species of legume, including pea, alfalfa and clover, form 'indeterminate' nodules (Terpolilli et al., 2012; Timmers et al., 2000). Rhizobial differentiation into bacteroids in these legumes is terminal, as bacteria cannot be regrown after differentiation (Poole et al., 2018). NCR peptides alter the cell cycle in these bacteroids, as the whole genome undergoes several cycles of duplication without cell division, resulting in multiple copies of the genome per bacteroid (Alunni & Gourion, 2016; Pan & Wang, 2017). These bacteroids swell to a much larger size than viable bacteria and their membranes become more permeable (Mergaert et al., 2006). Rhizobia in the infection thread, and possibly elsewhere in the nodule, remain undifferentiated and are viable when released from the nodule. Beans and soybeans form 'determinate' nodules, which have a different nodule structure and do not force swelling, endoreduplication or terminal differentiation in their bacteroids (Franssen et al., 1992).

The nitrogenase enzyme complex, responsible for fixing atmospheric nitrogen to ammonium, comprises a homodimer containing an iron-sulphur cluster (Fe protein) and a heterotetramer containing a molybdenum-iron cofactor and two iron-sulphur clusters (FeMo protein; reviewed in Poole et al., 2018). These important metallo-cofactors make both proteins extremely sensitive to oxygen, which can irreversibly poison the nitrogenase complex. Leghaemoglobin produced by the host plant binds oxygen with high affinity and reduces the free oxygen levels inside nodules to the tens of nanomolar range, preventing such poisoning of nitrogenase (Jiang et al., 2021; Ott et al., 2005). However, oxygen is

required for respiration in the bacteroid, which necessitates a fine-tuning of redox balance and oxygen levels to prevent nitrogenase inactivation but allow respiration (Chang et al., 2009; Udvardi & Poole, 2013). It is the change to the micro-aerobic (nanomolar oxygen concentration) conditions of the nodule that is sensed by the oxygen sensitive domains of the regulators FnrN and FixL, which are largely responsible for inducing the expression of the *fixNOPQ* and *fixGHIS* operons required for nitrogen fixation (Cherfils et al., 1989; de Philip et al., 1990; Rutten & Poole, 2019). Other nitrogen fixation gene products are thought to be induced by other signals from host plant regulators such as FixV and downstream signals (Poole et al., 2018). The overall biological nitrogen fixation reaction is described by Equation 1. The process is energy intensive, requiring 16 mols ATP per 2 mols ammonia produced and 8 mols of reductant such as NADH (Seefeldt et al., 2009).

$$N_2 + 16ATP + 16H_2O + 8e^- + 8H^+ \longrightarrow 2NH_3 + H_2 + 16ADP + 16P_i \qquad (1)$$

Equation 1. Biological nitrogen fixation reaction formula.

The shift in metabolism towards growth on host-provided dicarboxylates is key to efficient symbiotic nitrogen fixation (Ledermann et al., 2021). Bacteroids growing on dicarboxylates have decreased flux through the decarboxylating arm of the tricarboxylic acid (TCA) cycle, which decreases the amount of the TCA cycle intermediate 2-oxo-glutarate (Schulte et al., 2021). The intermediate 2-oxo-glutarate is essential for glutamine and glutamate biosynthesis and thus for nitrogen assimilation (Mulley et al., 2011). With bacteroids having reduced ability to assimilate the nitrogen fixed by nitrogenase, ammonia is secreted out of the bacteroid and across the symbiosome membrane into the plant. Additionally, growth on succinate (a dicarboxylate) roughly doubles rhizobial de-

mand for oxygen relative to growth on glucose due to the generation of reductant by the TCA cycle, consequently increasing the NADH/NAD^+ ratio (Schulte et al., 2021). This increased availability of NADH reductant can then be utilised for biological nitrogen fixation.

1.3 Use of Inoculants in Agriculture

The over-arching goal of much investigation into *Rhizobium*-legume symbiosis is to acquire knowledge that may be translated into sustainable improvements for agriculture. One example of this is has been the development and use of rhizobial inoculants for legume crops. Legume crops that form symbioses with rhizobia require compatible rhizobial strains to be present in the soil for optimal growth, a requirement that has been shown to have limited the propagation of symbiotic-legume species outside of their native habitats (Simonsen et al., 2017). Provision of compatible rhizobia has been accomplished with commercial inoculant cultures for over 100 years, with a more primitive soil transfer method for longer before that (Fred et al., 1932; Parnell et al., 2016). Currently, soybean is the crop that is most commercially inoculated worldwide (with *Bradyrhizobium*), but other legume crops (including pea) are also important for inoculant manufacturers (Santos et al., 2019)

Inoculant use aims to provide the legume host with sufficient numbers of effective rhizobia at the germinating seed, in order to nodulate the legume host as rapidly and completely as possible (Lupwayi et al., 2006). Rhizobial inoculants are often used with milled peat to carry and protect the bacteria. This peat is sterilised using gamma irradiation to kill contaminating microbes which would out-grow the *Rhizobium* culture, as heat treatments can produce inhibitory substances that negatively affect rhizobial survival (Deaker et al.,

2004). In addition, the pH, moisture content and salinity of the peat was found to be important for inoculant effectiveness (Roughley & Vincent, 1967).

Inoculants may be applied to soil or directly to legume seeds. Seed treatment may be carried out by farmers, local seed merchants, or inoculant companies as pre-inoculated seed (Deaker et al., 2004). Results are best when seed treatment is carried out close to sowing, as the survival of rhizobia on pre-inoculated seed may be poor over long storage times and this can reduce the efficacy of the inoculant (Brockwell et al., 1975). Rhizobial survival on pre-inoculated seed may be extended by improving rhizobial tolerance to desiccation (Casteriano et al., 2013). Research has been carried out towards this objective by adding sugars (Streeter, 2003) and synthetic polymers (Deaker et al., 2007) to the coating medium, and by investigation of genetic factors that improve desiccation tolerance (Mcintyre et al., 2007).

It is known that inoculating legume crops with compatible rhizobia can significantly increase yields (by 56% in one study; Clayton et al., 2004), with further gains possible if strains with high nitrogen-fixing ability are chosen for the inoculant (Deaker et al., 2004). Many elite, high nitrogen-fixing strains have been developed for this purpose (Ben Romdhane et al., 2008; Maier & Brill, 1978; Pastor-Bueis et al., 2019; Williams & Phillips, 1983). However, in fields with large numbers of indigenous rhizobia, the yield gains from inoculants can be eroded if the indigenous strains are more competitive and have lower nitrogen-fixing ability than the inoculant strain (Lupwayi et al., 2006). This competition problem will need to be overcome before high nitrogen-fixing strains can be most effectively utilised as inoculants (Giller & Wilson, 2001; Mendoza-Suárez et al., 2021).

1.4 *Rhizobium* Competitiveness

The majority (90-98%) of nodules are formed following a single independent infection event involving the entrapment of a single *Rhizobium* and its descendants by root hair curling and infection thread formation (Westhoek et al., 2017). The occurrence of 'mixed' nodules (2-10%), containing rhizobia of heterogeneous origin, can be expected to increase at higher inoculation densities as a result of increased root hair colonisation by rhizobia. These 'mixed' nodules result from the entrapment of multiple rhizobial strains during nodulation. When multiple rhizobial strains are in competition to infect the same plant, the genesis of each nodule is a semi-random event influenced by the relative abundance and relative competitive nodulation abilities (competitiveness, competitiveness for nodulation) of the given strains (Beattie et al., 1989). The relative competitive nodulation abilities of two rhizobial strains can be influenced by myriad environmental and soil conditions, including temperature, moisture, salinity, pH, and the presence of heavy metal ions (Anyango et al., 1998; Frey & Blum, 1994; Lupwayi et al., 2006; Vargas & Graham, 1989). The species of host legume will also modify the apparent competitive nodulation ability of a rhizobial strain through the various mechanisms of host specificity (Boivin et al., 2020; Jorrín, 2016; Walker et al., 2020).

1.4.1 Impact of Competition on Inoculants

It has been known for almost a century that indigenous rhizobia in soil can limit the yield of legume crops, as variation in rhizobial nitrogen-fixing ability means the lowest-performing symbiont strains can be effectively parasitic (Baldwin & Fred, 1929; Dwivedi et al., 2015; Helz et al., 1927; Leonard, 1930). Such strains are the reason why many inoculants fail to improve yield in soils where the indigenous population is significant

(reported by Drew et al., 2012; López-García et al., 2002; May and Bohlool, 1983; Naeem et al., 2004; Thies et al., 1991; Yates et al., 2011 among others). The considerable ability of indigenous strains to out-compete inoculant strains was evident in one study, which found that none of the 543 nodules tested from Wisconsin soybean farms contained the inoculant strains with which the seeds had been treated prior to sowing (Kamicker & Brill, 1986). Another set of field trials concluded that the inoculum may need to exceed the native population by 1000x to achieve 50% nodulation by the inoculant strain (Weaver & Frederick, 1974).

1.4.2 Measurement of Competitive Nodulation Ability

It has been shown that competitiveness is a key factor alongside nitrogen-fixing ability when selecting an inoculant strain, but the quantitative measurement of competitive nodulation ability is less straightforward than for nitrogen-fixing ability (Lupwayi et al., 2006; Vlassak & Vanderleyden, 1997). The experimental measurement of competitiveness is largely through co-inoculation experiments, with the first of these performed almost a century ago (Dunham & Baldwin, 1931; Löhnis, 1930).

Quantitative models have been developed to quantify relative competitiveness between two strains (Amarger & Lobreau, 1982). Later refinement of these models has allowed mixed nodules to be taken into account (Beattie et al., 1989). These models define relative competitiveness as the relationship between the proportional representation of strains in the inoculum and the proportional representation of strains in nodules. When equally competitive strains are placed in competition, the proportional representation of a strain in nodules is expected to be equal to its proportional representation in the inoculum. A

strain with greater competitiveness than its competitor is expected to be over-represented in nodules compared to the inoculum (Mendoza-Suárez et al., 2021).

Nodule occupancy can be assessed by antibiotic resistance of bacteria recovered from nodules or, more recently, by chromogenic staining of marked rhizobia within nodules (Sessitsch et al., 1998; Sessitsch et al., 1996). Chromogenic marker systems relying on *gusA* and *celB* activity allow higher throughput than earlier, more labour-intensive methods based on discrimination of morphological, serological or antibiotic markers (various methods used by Graham, 1969; Holland, 1966; Josey et al., 1979; Wilson et al., 1995; Wilson, 1995). The β-glucuronidase encoded by *gusA* converts colourless Magenta-GlcA into a magenta-coloured product, while the thermostable β-galactosidase encoded by *celB* converts colourless X-Gal into a blue product, following heat treatment to inactivate endogenous β-galactosidases (Jefferson et al., 1986; Voorhorst et al., 1995). This chromogenic marker system has been established in Rlv3841, a model strain used by the Poole lab (Sánchez-Cañizares & Palacios, 2013).

To quantify relative competitiveness, the models involve regressing a function of the nodule occupancies of the strains on a function of the inoculum ratio (Beattie et al., 1989). This is used to generate a competitiveness index as a quantitative measurement of relative competitiveness. One model is described in Equation 2 (Beattie et al., 1989). In this model, the competitive index between strain A and strain B is represented as $C_{A:B}$, which is calculated from the proportional representation of strain A in nodules (P_A), strain B in nodules (P_B) and mixed nodules (P_{both}). The ratio of strains A and B in the inoculum (I_A/I_B) are also taken into account, allowing the integration of data from competetion

assays with different inoculation ratios. This is essential for precise comparisions of strains with vastly different competitive abilities.

$$log\left(\frac{P_A + P_{both}}{P_B + P_{both}}\right) = C_{A:B} + k \cdot log\left(\frac{I_A}{I_B}\right) \quad (2)$$

Equation 2. Model for the quantification of competitiveness from Beattie et al., 1989.

1.4.3 Improvement of Competitive Nodulation Ability

Research has been carried out to select highly competitive and effective strains for inoculants. One strategy is to isolate highly competitive indigenous strains from a particular geographical region, then conduct chemical mutagenesis and screen for mutants with high nitrogen-fixing ability (Paau, 1989). This strategy takes advantage of the natural selection that has taken place to adapt indigenous rhizobial strains for the specific soil, conditions and cultivars of fields in an area and has been recognised as an effective strategy and utilised successfully for other soil-types (Batista et al., 2015). Another way competitiveness has been improved is with the introduction of bacteriocin genes to rhizobia. The *tfx* genes, encoding trifolitoxin anti-rhizobial peptides, were transferred by researchers to the clover symbiont *R. leguminosarum* bv. *trifolii* TA1. The trifolitoxin-producing strain was able to significantly outcompete the wildtype for nodulation, occupying approximately 90% of nodules when inoculated at an approximately one-to-one ratio (Triplett, 1990). The first example of genetic improvement of competitiveness in field conditions came from following work, in which trifolitoxin-producing strains of the common bean symbiont, *Rhizobium etli*, were shown to have increased ability to competitively nodule in real soil and agricultural conditions (Robleto et al., 1997; Robleto et al., 1998).

Exported toxins affecting host legumes can also enhance the competitive nodulation ability of the nodulating rhizobia. The soybean symbiont *Bradyrhizobium elkanii* produces rhizobitoxine, an inhibitor of soybean 1-Aminocyclopropane-1-Carboxylate Deaminase (an enzyme involved in ethylene production). Since ethylene has been reported to be an inhibitor of nodulation in legumes, rhizobitoxine inhibition of host ethylene production allowed *B. elkanii* to form more nodules than mutants lacking rhizobitoxine production during co-inoculation experiments (Lee & LaRue, 1992; Nukui et al., 2000; Okazaki et al., 2004).

Identification of genes positively contributing to competitiveness has led to increases in competitiveness being achieved by the overexpression of these genes. The *rosR* gene, a transcriptional activator for exopolysaccharide (EPS) production was shown to be necessary for successful nodulation of clover by *R. leguminosarum* bv. *trifolii*. Overexpression of *rosR* via plasmid transformation resulted in significantly improved competitiveness for nodulation as well as a significant increase of host plant biomass (Janczarek et al., 2009). Similar results have been found by constitutively overexpressing the proline dehydrogenase *putA* of the alfalfa symbiont *S. meliloti*. Overexpression of *putA* resulted in more nodules being formed earlier than the parental strain. After 34 days the parental strain caught up in nodule occupancy (van Dillewijn et al., 2001). This 'catching' up is likely a consequence of total nodule number being controlled by the host legume in response to nitrogen availability, which would limit the influence of nodulation speed on total nodule number as its nitrogen requirements are met by mature nodules (Ferguson et al., 2019).

Perhaps counter-intuitively, *Rhizobium* genomes include genes which can negatively contribute to competitiveness under certain conditions. These genes have not been removed by natural selection, therefore it is likely that these genes are important for fitness under other conditions. For example, the *praR* gene represses several genes involved in EPS production and cell-surface attachment factors (Frederix et al., 2011). Deletion of *praR* in *R. leguminosarum* releases these genes from repression and leads to a hyper-attachment phenotype from EPS overexpression, increasing biofilm formation and competitiveness for nodulation of pea, much like *rosR* overexpression strains (Frederix et al., 2014; Janczarek et al., 2009). This increase in competitiveness was found when plants were grown in vermiculite, but it is possible that hyper-attachment strains may be competitively disadvantaged in soils with more organic material, as they could hyper-attach to this instead of host roots. In this work, Chapter 4 focuses on identifying and characterising other genes that negatively contribute to competitiveness, using observations from the Poole Lab's transposon insertion sequencing experiments identifying 'Advantaged' mutants (Wheatley et al., 2020).

1.5 Transposon Insertion Sequencing

The process of transitioning from a free-living saprophyte, colonising and infecting plant roots, then undergoing the huge physiological and metabolic changes required to become a nitrogen-fixing bacteroid, overall necessitates a relatively large bacterial genome at 7.75 Mb and 7,263 genes for the model pea-symbiont *Rhizobium leguminosarum* bv. *viciae* 3841 (Rlv3841; Land et al., 2015; Young et al., 2006). These include thousands of interacting parts working together under tight regulation to move rhizobia efficiently from the beginning to the end of this complex process. It is therefore clear that the competitive-

ness of a rhizobial strain, the ability to quickly and successfully infect the host legume over other rhizobial strains, can be affected by myriad genetic and environmental factors (factors reviewed in Dowling and Broughton, 1986; Ji et al., 2017; Onishchuk et al., 2017).

The traditional methods for identifying genes important for competitiveness under specific conditions and experimentally demonstrating their function are low-throughput, which limits their utility for efficiently investigating the many *Rhizobium* genes which are uncharacterised or poorly characterised (Chang et al., 2016; Royet et al., 2019). High-throughput transposon insertion sequencing techniques have been developed for rhizobia to address this (Perry & Yost, 2014).

Transposon insertion sequencing techniques utilise mobile genetic elements to generate mutant libraries comprising bacteria each with one transposon disrupting one of hundreds of thousands of loci where the transposon may potentially insert (Cain et al., 2020; Chao et al., 2016). The Tn*5* and mariner transposons are the most commonly used for this kind of screen (Fabian et al., 2020). This mutant library can be DNA sequenced and analysed to determine the frequency of bacteria with an insertion at each locus (Gawronski et al., 2009). The initial library can be passed through one or more sets of conditions then sequenced and analysed in the same way afterwards (Wheatley et al., 2017). Transposon insertion into a gene essential the growth conditions results in transposon mutants in these genes being lost from the output library (Barquist et al., 2013). By comparing the frequency of transposon mutants for each locus before and after the experimental conditions, genes important for survival in those conditions can be identified in a high-throughput manner (Van Opijnen & Camilli, 2013).

1.5.1 INSeq as a Genome-Wide Mutagenesis Screen

Insertion Sequencing (INSeq) is one of the transposon insertion techniques that uses the Himar I mariner transposon, a type II transposable element that reliably and randomly inserts in thymine-adenine (TA) motifs in prokaryotes, eukaryotes and archaea (Goodman et al., 2009; Plasterk et al., 1999). There are 140,056 TA motifs in the 7.75 Mb genome of the model strain Rlv3841, so there is a potential insertion site every 55 bp on average (Perry & Yost, 2014). Only ten genes in the Rlv3841 genome completely lack TA sites and thus cannot be investigated by this transposon. Genes with few TA sites (<5) or large regions of sequence duplication may lead to ambiguities that prevent gene classifion by the analysis software with high confidence (Perry et al., 2016).

The plasmid pSAM_Rl (Figure 2) was developed to carry a modified mariner transposon into *R. leguminosarum* (Perry & Yost, 2014). The transposase gene is outside of the transposon and will not integrate into the host genome, preventing further transposition events after pSAM_Rl is lost during library preparation (Goodman et al., 2009). The transposon contains *nptII* as a selectable marker, conferring resistance to kanamycin and neomycin.

The mariner transposon was chosen partly because a single base mutation in the inverted repeat regions can introduce an *Mme*I restriction site. The *Mme*I type IIs restriction enzyme does not cut directly at this recognition site, but 16 bp outside of the transposon in this case (Morgan et al., 2009). Combining *Mme*I digestion and linear PCR with biotinylated primers allows the isolation of uniformly amplified sequences, which can be used to map the transposon insertion loci in DNA samples extracted from input and output

mutant libraries (Goodman et al., 2011). The use of *Mme*I to generate uniformly long fragments reduces PCR amplification bias compared to alternative transposon insertion sequencing methods that use random DNA shearing and circularisation.

Figure 2

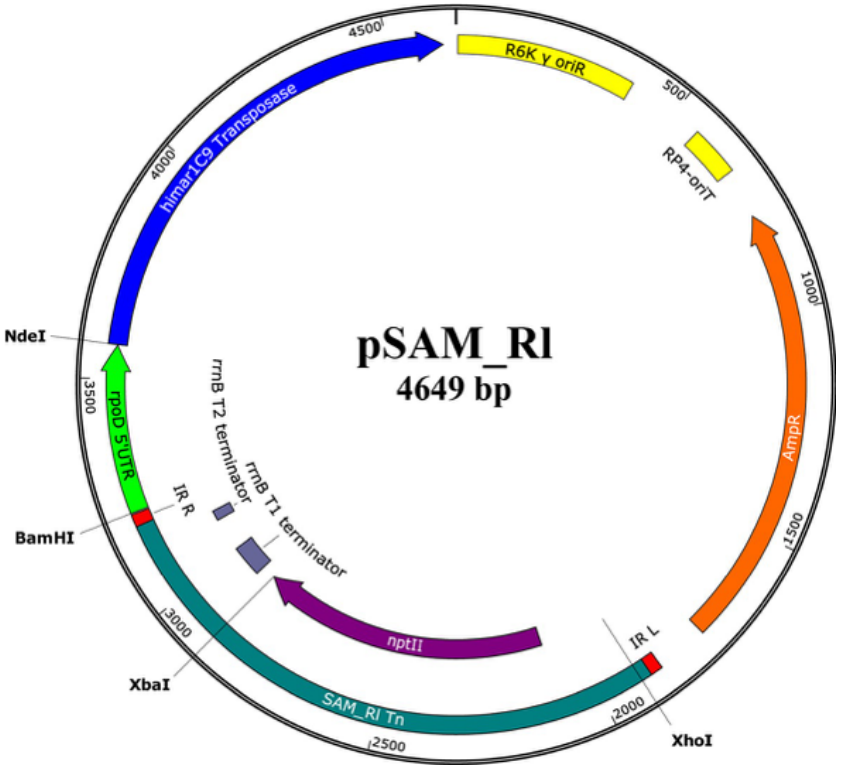


Figure 2. Plasmid map of pSAM_Rl, the vector used to generate mariner transposon mutant libraries in *Rhizobium leguminosarum*. Figure taken from Perry and Yost, 2014.

After this processing, the Ion-Torrent sequencing platform is employed to sequence the mutant libraries and scripts are used to map the reads to TA sites in the genome (Wheatley et al., 2017). The frequencies of reads for each TA site in the input and output library are compared by a Hidden Markov Model, which assigns fitness states (INSeq classifications) to the gene associated with a number of TA sites (Perry & Yost, 2014).

1.5.2 Experimental Findings From INSeq on the Model Strain Rlv3841

The first INSeq experiments on Rlv3841 were performed aiming to find out which genes were essential for free-living growth on a minimal mannitol-containing medium (Vincent's media) in contrast with a peptide rich medium (TY media; Perry et al., 2016). These results revealed that transposon insertion mutation could be tolerated in approximately 90% of Rlv3841 genes in both conditions, however, genes that were required for one medium specifically were also identified. For Vincent's media, 661 genes were required for growth, compared to 563 genes for TY media. The finding that fewer genes are required for growth in TY media is in agreement with TY media being the richer medium. The two lists of genes were contrasted to define an Rlv3841 core functional genome of 491 genes needed for growth irrespective of specific growth conditions (Perry et al., 2016). Almost half (237 genes) of the core functional genome comprised genes involved in macromolecule synthesis and modification, energy and carbon metabolism, ribosome composition or the cell envelope, important core cellular functions (Perry et al., 2016). Interestingly, 50 uncharacterised hypothetical proteins were included, indicating that there is still much that is unknown about many processes essential for basic *Rhizobium* physiology.

In the Poole Lab, INSeq has been used to investigate the genes affecting the growth of Rlv3841 on different substrates and at different oxygen concentrations. Two substrates, succinate and glucose, were investigated at both 21% and 1% O_2. The experiments identified many genes that were specifically important for growth under those conditions (Wheatley et al., 2017).

Some of these genes had previously known or predicted roles in carbohydrate metabolism, strengthening the validity of INSeq as a technique, while others were uncharacterised and highlighted for further investigation. The results were used to make new inferences about the interactions between core metabolic pathways in Rlv3841 that would be difficult to uncover with single-mutational studies (Wheatley et al., 2017). The data on succinate at 1% O_2 are of particular interest because those conditions most closely match those which a nitrogen-fixing bacteroid encounters inside legume nodules (Poole et al., 2018). Identification of ammonia assimilation regulation as essential in these conditions matched microarray data showing that it is switched off in bacteroids, which was suggested to be due to the need to avoid the removal of intermediates from the TCA cycle under low oxygen conditions (Green et al., 2019; Wheatley et al., 2017).

These first INSeq experiments in Rlv3841 laid the groundwork for an ambitious INSeq experiment to carried out with the aim of determining which genes are important for symbiosis with a legume host (Wheatley et al., 2020). Mariner-transposon mutant libraries were passed through and recovered from four stages of symbiosis with pea: rhizosphere growth, root colonisation, bacterial infection (nodule bacteria) and mature nodule bacteroid.

As a result of this study, 603 genetic regions were shown to be required for the ability to competitively nodulate pea and fix nitrogen, far more than the 27 genes previously annotated as nitrogen fixation genes (Wheatley et al., 2020). Mutants in 146 genes were found to impair growth in the rhizosphere to a degree that all subsequent stages of symbiosis were affected, highlighting how important competition in the rhizosphere is to nodulation. In addition, mutations in 211 genes were found to specifically impair the fitness of rhizobia in nodules as bacteria and bacteroids.

This INSeq experiment has classified some genes as expected from previously published literature (e.g. *fix*, *nif*, & *nod* genes), illustrating the validity and reliability of the study, but also identified as yet little characterised genes as candidates for closer examination (Wheatley et al., 2020). The work described in this thesis made use of data generated by the symbiosis INSeq experiment as the basis for such closer investigations of several genes of interest, both aiding in the selection of this genes and providing phenotypic data for comparison with the experimental data generated in this work.

1.6 General Aims of This Research

The first aim of this work was to determine which stages of symbiosis were affected by mutations in a set of genes of interest, identified with the aid of INSeq dataset. This investigation made use of small-scale experiments, such as competition and N-fixation assays, to define the affected stages of symbiosis in greater resolution than allowed by INSeq, while at the same time contributing towards the publication of a selection of data from the INSeq symbiosis experiment (Wheatley et al., 2020).

Another aim was to characterise a set of genes which, if mutated, have the potential to increase the competitiveness and/or effectiveness of a rhizobial strain during symbiosis. These genes are considered particularly interesting because the way inoculants have been used is very different to how rhizobia have existed for millions of years of their evolutionary history, so genes useful to long term survival in soil are not necessary for performance as an inoculant and may be detrimental to this end (Deaker et al., 2004; Frederix et al., 2014). Completion of this aim was of interest to the project's industrial partner, Legume Technology Ltd, who were interested in ways to improve their inoculant strains. By establishing a list of genes detrimental to performance as an inoculant, elite inoculant strains may be better screened from isolate collections, translating into economic and environmental benefits to farmers, as well as increased food security.

One further aim was to identify and characterise mutants with altered timing of nodule development. Identification of mutants with altered timing of nodule development would provide new insights into the regulation of bacteroid development and assist with the global effort to transfer *Rhizobium* symbiosis to cereals (Pankievicz et al., 2019). Investigation towards this aim has previously been impeded by a lack of tools to follow the kinetics of nodulation. In this work, a screening method was developed to monitor multiple milestones of nodule development and nitrogen fixation over days and weeks, and was used to assess mutants for altered timing of nodulation.

2 Materials and Methods

2.1 Bacterial Growth Conditions

R. leguminosarum - TY Media

R. leguminosarum strains were generally grown at 28°C on tryptone-yeast extract (TY) agar or liquid media (Beringer, 1974). TY liquid media comprises 5 g/l tryptone, 3 g/l yeast extract, and 6 mM $CaCl_2$, while TY agar is the same with the addition of 15 g/l bacteriological agar.

R. leguminosarum - UMS and UMA Media

Universal minimal salts (UMS) and universal minimal agar (UMA) were used as minimal liquid and solid media respectively for *R. leguminosarum* strains (Poole et al., 1994). UMS/UMA comprises 0.5 mM K_2HPO_4, 0.5 g/l $MgSO_4 \cdot 7H_2O$, 0.2 g/l NaCl, 4.19 g/l MOPS, and 15 g/l bacteriological agar (for UMA only), pH adjusted to 7.0. After 1 litre of UMS/UMA base has been autoclaved, the following solutions are filter sterilised and added: 1 ml of calcium stock solution (75 g/l $CaCl_2 \cdot 2H_2O$), 1 ml of iron stock solution (12 g/l $FeSO_4 \cdot 7H_2O$), 1 ml of UMS stock solution (1 g/l thiamine hydrochloride, 2 g/l D-pantothenic acid calcium salt, 100 mg/l biotin, 10 mM glucose, 10 mM NH_4Cl), and 1 µl of 1000x vitamin stock solution (0.375 g/l EDTA-Na_2, 0.16 g/l $ZnSO_4 \cdot 7H_2O$, 0.2 g/l $NaMoO_4$, 0.25 g/l H_3BO_3, 0.2 g/l $MnSO_4 \cdot 4H_2O$, 0.02 g/l $CuSO_4 \cdot 5H_2O$, and 1 g/l $CoCl_2.6H_2O$).

E. coli - LB Media

E. coli strains were grown at 37°C on lysogeny broth (LB) agar or liquid media (Bertani, 1951). LB comprises 10 g/l tryptone, 5 g/l NaCl, and 5 g/l yeast extract, with or without 15 g/l bacteriological agar.

Antibiotics

Antibiotics were used at the following concentrations (in µl ml^{-1}) unless otherwise stated: chloramphenicol, 10; gentamycin, 10; kanamycin, 20; neomycin, 80 (240 for conjugant selection); nitrofurantoin, 20; streptomycin, 500; and tetracycline, 10.

Other Chemicals

The dyes X-Gal, Magenta-GlcA and X-GlcA were used at 40 µg ml^{-1}. Hesperitin was used at 2 µM.

2.2 Plant Growth Conditions

Seed Sterilisation

Peas (*Pisum sativum* cv *Avola*) were surface sterilised by immersion in 70% ethanol for 45 seconds, then 2% bleach for 5 minutes, followed by 10 washes with sterile distilled water.

Nitrogen-free Rooting Solution

N-free rooting solution for plant growth was made up as described in (Poole et al., 1994). N-free rooting solution comprises 1 mM $CaCl_2 \cdot 2H_2O$, 100 µM KCl, 800 µM $MgSO_4 \cdot 7H_2O$, 10 µM Fe·EDTA, 35 µM H_3BO_3, 9 µM $MnCl_2.4H_2O$, 0.8 µM $ZnCl_2$, 0.5 µM $Na_2MoO_4 \cdot 2H_2O$, 0.3 µM $CuSO_4 \cdot 5H_2O$, 0.5 g/l KH_2PO_4, and 0.568 g/l Na_2HPO_4. Rooting

solution was autoclaved before use. For vermiculite-sand pots and paper pouches, rooting solution was used at 2.67x concentration.

Growth in Vermiculite

200 ml rooting solution was added to 500 ml beakers filled with fine vermiculite, or for competition assays 75 ml of 2.67x concentrated rooting solution was added to a 1:1 fine vermiculite to silver sand mixture. Two sterile peas were pressed into the substrate of each pot then pots were sealed with cling film. Pots were grown in a growth chamber at 21°C with a 16 hour photoperiod. After 7 days the cling film was cut open and the smaller pea seedling discarded.

Growth in Plastic Pouches

25 ml distilled water was added to wet each plastic pouch (large CYG^{TM} seed germination pouch, Mega International) before sets of six pouches were wrapped in aluminium foil and autoclaved. Pea seedlings, pre-germinated on 1% distilled water agar plates in the dark at 28°C for seven days, were placed in the paper fold at the top of the pouches (one seedling per pouch) with the root fed down into the main part of the pouch. 10 ml of 2.67x N-free rooting solution was pipetted into each pouch, then a support stick was affixed horizontally to the top of each pouch with tape and pouches were placed in cardboard boxes. These boxes were placed in a growth chamber at 21°C with a 16 hour photoperiod.

Buffers

Phosphate buffer buffer for root staining comprises 7 g/l NaH_2PO_4, 7.2 g/l Na_2HPO_4, 1 mM EDTA (pH 8.0), Sarkosyl 1%, 1 ml/L Triton (Westhoek et al., 2017).

X-Gal, X-GlcA and Magenta-GlcA solutions were made up with 20 mg ml^{-1} of the relevant dye in DMF (Westhoek et al., 2017). PSM comprises sucrose 0.73 mM, K_2HPO_4 8 mM, KH_2PO_4 2 mM, and $MgCl_2$ 5 mM in distilled water, filter sterilised with a 0.22 µm filter and stored at 4°C (Karunakaran et al., 2009).

2.3 Molecular Techniques

DNA Isolation

Genomic DNA was extracted from *R. leguminosarum* using the Qiagen DNeasy® Blood and Tissue kit, according to the manufacturer's protocol. Plasmid DNA was extracted from *E. coli* using the Thermo Scientific GeneJET™ Plasmid Miniprep kit, according to the manufacturer's protocol. PCR products were purified using the Zymo Clean and Concentrator® kit, according to the manufacturer's protocol. DNA concentration measurement was performed with a Qubit assay (Thermo Fisher Scientific), according to the manufacturer's protocol.

Primer Design

PCR primers were designed using Geneious 10.0.9 or the NEBuilder 2.0 online tool (nebuilder.neb.com). Primer oligonucleotides were synthesised by Eurofins MWG.

DNA Amplification by PCR

Amplification of homologous regions for cloning were carried out using Phusion® (NEB) or Q5® (NEB) polymerase in 25 or 50 µl reactions. Mapping PCR was generally carried out as 10 µl colony PCR reactions with GoTaq® (Promega) or OneTaq® (NEB). Thermocycler conditions were set according to the polymerase manufacturers' protocols but

were often modified to include gradient or touchdown PCR steps to amplify fragments which failed to amplify on the first attempt.

DNA Gel Electrophoresis

Electrophoresis was carried out on gels with 1% agarose (Sigma Aldrich) in TAE buffer (400 mM Tris acetate, 1 mM EDTA), with Sybr® Safe 1:5000 (Invitrogen) to stain the DNA. GeneRuler 1 kb (Thermo Scientific) was used as a ladder. Gels were run for 45 minutes at 120 mV. For fragments smaller than 1 kb, electrophoresis was carried out in gels with 1.5% agarose, GeneRuler 100 bp ladder (Thermo Scientific) was used, and gels were run at 100 mV for 60 minutes. Run gels were visualised with the GelDoc EZ System (BioRad).

Restriction Digests and DNA Ligations

Restriction enzyme digests were performed according to manufacturers' instructions (Thermo Scientific, NEB, or Roche). DNA ligations were performed with T4 DNA ligase according to the manufacturer's protocol (Thermo Scientific).

Standard Transformation

50 µl aliquots of chemically competent *E. coli* DH5α were thawed on ice and 2 µl of ice-cold plasmid or ligation mixture added. Cells were incubated for 30 minutes on ice then heat-shocked for 45 seconds at 42°C. After the heat-shock, cells were incubated on ice for 2 minutes before 400 µl SOC media was added (Hanahan, 1983). Cells were shaken at 200 rpm for 1 hour at 37°C. Cells were spread on LB agar selection plates containing relevant antibiotics before being grown at 37°C overnight.

Standard Conjugation (Triparental Mating)

The recipient strains of *R. leguminosarum* to be conjugated were grown on TY agar plates with relevant antibiotics for 3 days prior to conjugation. The *E. coli* plasmid donor strains and *E. coli* pRK2013 helper strain were grown on LB agar plates with relevant antibiotics overnight before the day of conjugation. Single colonies of the *R. leguminosarum* recipient, *E. coli* donor and *E. coli* helper were transferred to a TY agar plate without antibiotics and spread in a $2 \sim 3$ cm diameter circle in the centre of the plate with inoculation loops. Conjugation plates were grown overnight at 28°C, before samples from the conjugation plaque were streaked out on TY agar selection plates containing the appropriate antibiotics to select for rhizobial conjugants. Selection plates were incubated at 28°C for 3 days.

Mutant Construction with pK19mob

20 bp primers were designed using Geneious to amplify a 300-800bp central region of the gene to be mutated. The 5' linkers TGATTACGCCAAGCT and GCAGGCATGCAAGCT were added to the forward and reverse primers respectively. The primers were used to amplify from Rlv3841 genomic DNA with Phusion® polymerase. PCR products were checked by electrophoresis and correctly sized fragments were purified using a Zymo Clean & Concentrator® kit. InFusion was used to clone fragments into *Hin*dIII-digested pK19mob (Schäfer et al., 1994), then DH5α was transformed and selected on LB·Km_{20}·$Xgal_{40}$ plates. Colony PCR with pK19A & B primers was used to verify transformants. Plasmids were sequenced with pK19A & B primers by Eurofins TubeSeq to verify the insert. Triparental mating was carried out with donor *E. coli*, helper pRK2013 *E. coli* and recipient 3841*gusA*. Conjugants were selected on TY·Str_{500}·Nm_{240}

agar plates and mapping colony PCR was used to verify mutagenesis. Correctly-sized PCR products were sequenced by TubeSeq after purification with the Zymo kit.

Mutant Construction with pOPS1670/pOPS1713

20 bp primers were designed using NEBuilder to amplify 700-1000bp flanks upstream and downstream of each target gene. The following 5' adapter sequences were added to the relevant primer: CTCGGTACCCGGGGATCCTCTAGAG to the upstream forward primer; GGTAACCACTCCCACCTGACCCTCC to the upstream reverse primer; TAGTGGAGGTTCTGCGGGTTCAGGG to the downstream forward primer; and CGTTGTAAAACGACGGCCAGTGCCA to the downstream reverse primer. These primers were used in touchdown PCR to amplify from Rlv3841 genomic DNA with Q5 polymerase. PCR products were checked by electrophoresis and correctly sized fragments were purified using a Zymo Clean & Concentrator kit. pOPS1670 and pOPS1713 were digested with *Hin*dIII & *Sal*I and *Mly*I respectively. Gel extractions were carried out on the digested plasmids to select the 5.9 kb backbone fragment from pOPS1670 and the 1.1 kb cassette fragment from pOPS1713. InFusion was used to assemble the upstream flank, cassette, downstream flank, and backbone in a 2:2:2:1 molar ratio with 25 ng backbone in 5 µl reactions. DH5α was transformed with 2 µl of the assembly mix and was selected on LB·Km_{20}·Cam_{10} agar plates. Transformants with kanamycin resistance, chloramphenicol resistance, and red fluorescence from *mCherry* expression were purified and used in triparental mating with helper *E. coli* pRK2013 and recipient OPS2575. Conjugants were selected on TY·Str_{500}·Nm_{240}·$Sucrose_{1\%}$ agar plates and verified with mapping colony PCR.

Construction of P*nifH*:*sfgfp* Reporter with Golden Gate Cloning

Level 1 Golden Gate reactions were carried out in 15 µl reactions containing 1.5 µl T4 DNA ligase buffer, 1.5 µl BSA diluted 10x, 1 µl *Bsa*I, and 1 µl T4 DNA ligase (Thermo Scientific), as well as DNA equal to 40 fmol for each Golden Gate part. Assembly reactions were performed in a thermocycler with the following protocol: 25 cycles of 3 minutes at 37°C and 4 minutes at 16°C, followed by 5 minutes at 50°C then 5 minutes at 80°C. This was based on the 'one-tube one-step' reaction protocol (Engler et al., 2008). 1 µl of the assembly reaction was then used to transform chemically competent *E. coli* cells. To generate OPS2575 (Rlv3841 Ps*nifH*:*sfgfp*), Golden Gate parts from the following plasmids were combined into pOPS1699: pOGG037, pOGG043, pOGG157, and pOGG276. To generate OPS2576 (Rlv3841 Ps*nifH*:*mCherry*), Golden Gate parts of the following plasmids were combined into pOPS1700: pOGG043, pOGG063, pOGG157, and pOGG276. These Golden Gate parts are described in Table 1. The resultant mini-Tn*7* vectors, pOPS1699 and pOPS1700, were then conjugated into Rlv3841 to generate the biosensor strains OPS2575 and OPS2576 respectively.

Table 1

Plasmid	Module	Description	Source
pOGG037	SC	pL0M-SC-*sfgfp*	Weber et al., 2011
pOGG043	PU	pL0M-PU-Ps*nifH* (synthetic P*nifH*)	Mendoza-Suárez et al., 2020
pOGG063	SC	pL0M-SC-*mCherry*	Weber et al., 2011
pOGG157	T	pL0M-T-DT16	Geddes et al., 2019
pOGG276	Level 1 Vector	pL1V-Lv1-pUC18R6K-miniTn*7*T-Gm	Beatriz Jorrín, Poole Lab

Table 1: Golden Gate parts used for the construction of background strains OPS2575 and OPS2576. OPS2575 is Rlv3841 with the Ps*nifH*:*sfgfp* biosensor integrated on the chromosome using the mini-Tn*7* vector (Choi & Schweizer, 2006), while OPS2576 was constructed with the *mCherry*-carrying SC module in place of the *sfgfp* SC module. GFP-labelled OPS2575 is the background strain for all mutagenesis described in Chapter 5. Level 0 modules for Golden Gate Cloning are as follows: PU = promoter with ribosomal binding site; SC = open reading frame not intended for tagging; and T = terminator.

2.4 Plant and Microbiological Experiments

Nodulation Assays

500 ml Azlon® plastic beakers were filled with fine vermiculite and 200 ml 1x N-free rooting solution, which were then autoclaved. Rhizobial strains were grown on TY agar slopes with relevant antibiotics for 3 days before being washed with sterile distilled water. Cell suspensions were diluted to 10^8 cells ml^{-1}. Two sterile peas were pressed into the substrate of each pot then 500 µl of cell suspension was pipetted over each seed. 500 µl sterile distilled water was pipetted over seeds for water controls. Pots were sealed with

cling film and grown for 7 days before the cling film was cut open and the smaller pea seedling discarded. Growth continued for a further 3 weeks before plants were harvested.

Acetylene Reduction Assays

4-week-old plants were harvested and placed in 500 ml Duran bottles with some wet tissue paper. Bottles were closed with rubber septum lid. 15.25 ml of air was drawn from each bottle using a syringe, then 12.2 ml acetylene was injected. Bottles were left for 1 hour before 3 samples of 1 ml were drawn from each bottle and injected into the Perkin Elmer Clarus® 480 Gas Chromatograph as 3 technical replicates. The ratio between the areas under the acetylene and ethylene peaks were recorded and moles of acetylene reduced was used as a proxy for nitrogenase activity (Shah et al., 1975).

Single Staining for *gusA*

50 ml Falcon tubes were filled with 40 ml phosphate buffer and 400 µl X-GlcA solution. Plant roots were cut from the shoot and placed upside-down in these Falcon tubes, which were then covered in foil and incubated overnight at 28°C. Roots were then transferred to Falcon tubes filled with 50% bleach and left for 5 minutes before washing in distilled water.

Nodule Counting

Stained roots were cut below the lowest nodule and spread out on 25 cm square plates before photography of the root systems. The 'cell counter' plugin of FIJI was used to count nodules on root photographs.

Competititive Nodulation Assays

Competitive nodulation assays were carried out as described in (Westhoek et al., 2017) with a total inoculum of 10^4 cells per 500 ml pot. A 1:1 mixture of fine vermiculite and silver sand was used as a substrate, with 75 ml of a 2.67x concentrated N-free rooting solution. Inoculum density and ratio was tested by plating onto TY·Str_{500}·$XGlcA_{40}$ plates and grown before blue and white colonies were counted. Plants were grown for 3 weeks before harvesting.

Plant Dry Weight Assays

2000 ml plastic beakers were filled with medium vermiculite and 800 ml 1x N-free rooting solution, then autoclaved. Two sterile pea seeds were pressed 2 cm deep into the substrate, inoculated with 500 µl of *Rhizobium* (strain of interest) cell suspension at 10^4 cells ml^{-1} each. Plants were grown for 6 weeks before harvesting. Roots were cut from shoots just above the seed. Shoots were placed in paper bags and dried in a 70°C oven for 2 days. The bags of plant material were weighed on a fine balance, then the mean weight of 5 empty paper bags was subtracted from this weight to find the dry weight of the shoot.

Double Sequential Staining (*gusA* & *celB*)

Double sequential *gusA* and *celB* staining was carried out on plants from competition assays as described in (Westhoek et al., 2017). Briefly, 50 ml Falcon tubes were filled with 40 ml phosphate buffer and 400 µl Magenta-GlcA solution. Plant roots were cut from the shoot and placed upside-down in these Falcon tubes, which were then covered in foil and incubated overnight at 28°C. Roots were then transferred to Falcon tubes containing 40 ml phosphate buffer pre-heated to 70°C and incubated at 70°C for 2 hours, before transfer

to Falcon tubes filled with 40 ml phosphate buffer and 400 µl X-Gal solution. Tubes were covered in foil and incubated overnight at 37°C. Roots were then transferred to Falcon tubes filled with 50% bleach and bleached for 5 minutes before washing in distilled water.

Recovery of Bacteria From Nodules

10 of the largest single nodules were picked from each of 4-week-old plants, which had previously been inoculated with 10^4 cells of a single strain. These nodules were surface-sterilised with 2% bleach for 5 minutes, followed by 6 washes in sterile distilled water. In sterile 1.5ml tubes, nodules were suspended in 100 µl PSM and ground with sterile plastic pestles. 900 µl PSM was used to rinse material from each pestle after grinding to bring volume to 1 ml. Tubes were vortexed then allowed to settle for several minutes. 25 µl of liquid from the top of each tube was added to flasks containing 50 ml TY·Str_{500}, taking care to avoid including solid debris. The initial optical densities of the cultures were taken at 600 nm. 30 µl of liquid from each tube was similarly removed for a series of dilutions (100 through 10^{-7}) for 25 µl drop counts on TY·$Strep_{500}$ agar plates. The optical densities at 600 nm were measured for the liquid cultures after growth for 36 hours in a shaker at 28°C. Photographs of the drop count plates were taken at 1x magnification with a Leica M165 FC microscope, using Leica Application Suite V4, after growth for 48 hours at 28°C. The cell counter plugin for FIJI was used to count colonies at the different dilutions on each plate.

Bacterial Growth Assays

Cells were grown on TY agar slopes with appropriate antibiotics for 3 days at 28°C before being washed with 5 ml liquid TY media. These cell suspensions were used to inoculate liquid TY to an OD_{600} of 0.1 and a volume of 400 µl. Cultures were grown in a BMG

Labtech FLUOstar® Omega plate reader at 28°C over 48 hours, with OD_{600} measurements taken every 30 minutes to provide data for plotting growth curves.

2.5 Computational Methods

In silico cloning was performed using Geneious v10.0.9. Data manipulation was carried out using Microsoft Excel and statistical analysis was performed using GraphPad PRISM 9. Figures were made using GraphPad PRISM 9, BioRender (biorender.com), SAVVY (bioinformatics.org/savvy), and paint.net (getpaint.net).

Image Acquisition

Microscope images were taken on a Leica M165 FC, with excitation light for fluorescence supplied by a Leica EL6000 fibre optic fluorescence light source. Leica Application Suite v4.12.0 was used to acquire images.

Luminescence was measured using a Berthold NightOWL camera system with IndiGO v2.0.5.0 image analysis software. The minimum threshold for detected luminescence was set to 150 counts per second.

2.6 Bacterial Strains

Table 2

Strain	Description	Source	Ch.
	***E. coli* strains**		
α-Select Gold	Competent cells. Genotype:F$^-$ φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_K^-, m_K^+) *pho*A *sup*E44 λ$^-$*thi*-1 *gyr*A96 *rel*A1	Bioline	5
DH5α	Competent cells. Genotype: F$^-$ φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(r_K^-, m_K^+) *pho*A *sup*E44 λ$^-$*thi*-1 *gyr*A96 *rel*A1	Invitrogen	All
	***Rhizobium* strains**		
Rlv3841	*Rhizobium leguminosarum* bv. *viciae*, derivative of strain 300, Str^r	Johnston and Beringer, 1975	All
3841gusA	Rlv3841 derivative carrying *gusA* integrated in the chromosome, Str^r	Sánchez-Cañizares and Palacios, 2013	All
3841celB	Rlv3841 derivative carrying *celB* integrated in the chromosome, Str^r	Sánchez-Cañizares and Palacios, 2013	All
LMB109	Rlv3841 with pLMB26 (possible glucose transporter III RL1824) inserted omega tet cassette confirmed by PCR, Str^r Tet^r	Ramakrishnan Karunakaran, Poole Lab	4
LMB110	Rlv3841 glucose transporter triple mutant (RL3624::pK19mob) RL4249::Omega Spec; RL1826::Omega Tet) , Str^r Nm^r Spec^r Tet^r	Ramakrishnan Karunakaran, Poole Lab	4
LMB493	Rlv3841 glucose transporter double mutant combining RU4182 (omega spec) into LMB109 ST2a/ST2b double mutant colony, Str^r Spec^r	Ramakrishnan Karunakaran, Poole Lab	4
LMB504	Rlv3841 glucose transporter double mutant combining RU4184 (pk19) into LMB109 (by transduction) ST1/ST2b double mutant colony, Str^r Nm^r Tet^r	Ramakrishnan Karunakaran, Poole Lab	4
RU4174	Rlv3841 glucose transporter II (RL3624 mutant) (donor pRU2089) was made via pK19mob cloning, Str^r Nm^r	Ramakrishnan Karunakaran, Poole Lab	4
RU4184	Rlv3841 with pRU2087(RL4249 glucose transporter) showed spec resistance and gentamycin sensitivity on sucrose plates, Str^r Spec^r	Ramakrishnan Karunakaran, Poole Lab	4
RU4223	Rlv3841 double mutant of possible glucose transporters (RL3624/pK19 and RL4249 omega spec) confirmed by PCR, Str^r Nm^r Spec^r	Ramakrishnan Karunakaran, Poole Lab	4

continued on next page

Table 2 – continued from previous page

Strain	Description	Source	Ch.
OPS1598	pK19 mutant of RL0688-*cheA1* (pOPS0820 conjugated into 3841gusA), Str^r Nm^r	This work	3
OPS1599	pK19 mutant of pRL100199-*fixB* (pOPS0822 conjugated into 3841gusA), Str^r Nm^r	This work	3
OPS1600	pK19 mutant of pRL100195-*nifB* (pOPS0823 conjugated into 3841gusA), Str^r Nm^r	This work	3
OPS1601	pK19 mutant of pRL100186-*nodB* (pOPS0824 conjugated into 3841gusA), Str^r Nm^r	This work	3
OPS1602	pK19 mutant of pRL90058-*rmrR* (pOPS0825 conjugated into 3841gusA), Str^r Nm^r	This work	3
OPS1603	pK19 mutant of pRL120205-*eryB* (pOPS0831 conjugated into 3841gusA), Str^r Nm^r	This work	3
OPS1604	pK19 mutant of RL0149 (pOPS0833 conjugated into 3841gusA), Str^r Nm^r	This work	3
OPS1606	pK19 mutant of RL1545 (pOPS0836 conjugated into 3841gusA), Str^r Nm^r	This work	3
OPS1607	pK19 mutant of RL2606-*purC1* (pOPS0837 conjugated into 3841gusA), Str^r Nm^r	This work	3
OPS1608	pK19 mutant of RL3549-*glnII* (pOPS0838 conjugated into 3841gusA), Str^r Nm^r	This work	3
OPS1609	pK19 mutant of RL3654-*pssD* (pOPS0839 conjugated into 3841gusA), Str^r Nm^r	This work	3
OPS1610	pK19 mutant of pRL120198 (pOPS0846 conjugated into 3841gusA), Str^r Nm^r	This work	3
OPS1611	pK19 mutant of RL1496-*iolR* (pOPS0849 conjugated into 3841gusA), Str^r Nm^r	This work	3
OPS1612	pK19 mutant of RL3986-*ruvC* (pOPS0851 conjugated into 3841gusA), Str^r Nm^r	This work	3
OPS1613	pK19 mutant of RL1340-*sodB* (pOPS1107 conjugated into 3841gusA), Str^r Nm^r	This work	3
OPS1898	pK19 mutant of RL0748-*aglA1* (pOPS1271 conjugated into 3841gusA), Str^r Nm^r	This work	4
OPS1899	pK19 mutant of RL3565 (pOPS1272 conjugated into 3841gusA), Str^r Nm^r	This work	4
OPS1900	pK19 mutant of RL4139 (pOPS1273 conjugated into 3841gusA), Str^r Nm^r	This work	4

continued on next page

Table 2 – continued from previous page

Strain	Description	Source	Ch.
OPS1901	pK19 mutant of RL0339 (pOPS1274 conjugated into 3841gusA), Str^r Nm^r	This work	4
OPS1903	pK19 mutant of RL2036 (pOPS1276 conjugated into 3841gusA), Str^r Nm^r	This work	4
OPS1904	pK19 mutant of RL3624 (pOPS1278 conjugated into 3841gusA), Str^r Nm^r	This work	4
OPS1905	pK19 mutant of pRL70182 (pOPS1279 conjugated into 3841gusA), Str^r Nm^r	This work	4
OPS2575	Rlv3841 with mini-Tn*7* integration containing Ps*nifH*:*sfgfp* reporter (pOPS1699 conjugation), Str^r Gm^r	This work	5
OPS2576	Rlv3841 with mini-Tn*7* integration containing Ps*nifH*:*mCherry* reporter (pOPS1700 conjugation), Str^r Gm^r	This work	5
OPS2637	Marker replacement mutant of (gene A - pRL100051) made with pOPS1714 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2638	Marker replacement mutant of (gene D - RL3820) made with pOPS1715 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2654	Marker replacement mutant of (gene 1 - pRL100133) made with pOPS1715 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2655	Marker replacement mutant of (gene 4 - pRL110147) made with pOPS1715 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2656	Marker replacement mutant of (gene 7 - pRL120397) made with pOPS1715 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2657	Marker replacement mutant of (gene 8 - pRL70027) made with pOPS1715 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2658	Marker replacement mutant of (gene 9 - pRL70050) made with pOPS1715 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2659	Marker replacement mutant of (gene 10 - pRL70154) made with pOPS1715 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5

continued on next page

Table 2 – continued from previous page

Strain	Description	Source	Ch.
OPS2660	Marker replacement mutant of (gene 11 - pRL70155) made with pOPS1715 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2866	Marker replacement mutant of (gene 3 - pRL100432) made with pOPS1750 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2867	Marker replacement mutant of (gene 13 - pRL80080) made with pOPS1752 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2868	Marker replacement mutant of (gene 15 - pRL90320) made with pOPS1754 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2869	Marker replacement mutant of (gene 18 - RL0381) made with pOPS1757 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2870	Marker replacement mutant of (gene 21 - RL0501) made with pOPS1760 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2871	Marker replacement mutant of (gene 24 - RL0572) made with pOPS1763 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2890	Marker replacement mutant of (gene 14 - pRL90263) made with pOPS1753 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2891	Marker replacement mutant of (gene 17 - RL0143) made with pOPS1756 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2897	Marker replacement mutant of (gene 26 - RL0726) made with pOPS1798 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2898	Marker replacement mutant of (gene 27 - RL0727) made with pOPS1799 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2899	Marker replacement mutant of (gene 28 - RL0776) made with pOPS1800 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2900	Marker replacement mutant of (gene 29 - RL0903) made with pOPS1801 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5

continued on next page

Table 2 – continued from previous page

Strain	Description	Source	Ch.
OPS2901	Marker replacement mutant of (gene 31 - RL0994) made with pOPS1803 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2902	Marker replacement mutant of (gene 32 - RL1060) made with pOPS1804 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2903	Marker replacement mutant of (gene 33 - RL1217) made with pOPS1805 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2904	Marker replacement mutant of (gene 34 - RL1470) made with pOPS1806 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), Str^r Gm^r Nm^r	This work	5
OPS2905	Rlv3841 with mini-Tn*7* integration containing Ps*nifH*:*sfgfp* reporter (pOPS1699), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r	This work	5
OPS2906	Marker replacement mutant of (gene A - pRL100051) made with pOPS1714 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2907	Marker replacement mutant of (gene D - RL3820) made with pOPS1715 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2908	Marker replacement mutant of (gene 1 - pRL100133) made with pOPS1715 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2909	Marker replacement mutant of (gene 4 - pRL110147) made with pOPS1715 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2910	Marker replacement mutant of (gene 7 - pRL120397) made with pOPS1715 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2911	Marker replacement mutant of (gene 8 - pRL70027) made with pOPS1715 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2912	Marker replacement mutant of (gene 9 - pRL70050) made with pOPS1715 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5

continued on next page

Table 2 – continued from previous page

Strain	Description	Source	Ch.
OPS2913	Marker replacement mutant of (gene 10 - pRL70154) made with pOPS1715 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2914	Marker replacement mutant of (gene 11 - pRL70155) made with pOPS1715 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2915	Marker replacement mutant of (gene 3 - pRL100432) made with pOPS1750 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2916	Marker replacement mutant of (gene 13 - pRL80080) made with pOPS1752 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2917	Marker replacement mutant of (gene 15 - pRL90320) made with pOPS1754 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2918	Marker replacement mutant of (gene 18 - RL0381) made with pOPS1757 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2919	Marker replacement mutant of (gene 21 - RL0501) made with pOPS1760 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2920	Marker replacement mutant of (gene 24 - RL0572) made with pOPS1763 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2921	Marker replacement mutant of (gene 14 - pRL90263) made with pOPS1753 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2922	Marker replacement mutant of (gene 17 - RL0143) made with pOPS1756 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2923	Marker replacement mutant of (gene 26 - RL0726) made with pOPS1798 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5

continued on next page

Table 2 – continued from previous page

Strain	Description	Source	Ch.
OPS2924	Marker replacement mutant of (gene 27 - RL0727) made with pOPS1799 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2925	Marker replacement mutant of (gene 28 - RL0776) made with pOPS1800 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2926	Marker replacement mutant of (gene 29 - RL0903) made with pOPS1801 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2927	Marker replacement mutant of (gene 31 - RL0994) made with pOPS1803 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2928	Marker replacement mutant of (gene 32 - RL1060) made with pOPS1804 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2929	Marker replacement mutant of (gene 33 - RL1217) made with pOPS1805 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5
OPS2930	Marker replacement mutant of (gene 34 - RL1470) made with pOPS1806 in OPS2575 background (Ps*nifH*:*sfgfp* reporter), with pLMB712 (P*nodA*:*lux*), Str^r Gm^r Nm^r Tet^r	This work	5

Table 2: Strains used in this work

2.7 Plant Lines

All plant experiments used the *Avola* cultivar of pea (*Pisum sativum*).

2.8 Plasmids

Table 3

Plasmid	Description	Source	Ch.
pK19mob	pK19-based mobilisable plasmid for single recombination in *Rhizobium*, Km^r	Schäfer et al., 1994	All
pLMB712	pIJ11268 with the promoter of pRL100185 (nodABC) in *Xho*I *Kpn*I, Tet^r	Pini et al., 2017	5
pOGG037	Golden gate plasmid pL0M-SC-*sfgfp* superfolder GFP for cloning as ORF SC module without tags in golden gate EC58031, Spec^r	Weber et al., 2011	5
pOGG043	pL0M-PU-Ps*nifH* (synthetic P*nifH* for Golden Gate) EC58034, Spec^r	Mendoza-Suárez et al., 2020	5
pOGG063	pL0M-SC-*mCherry*, Spec^r	Weber et al., 2011	5
pOGG157	pL0M-T-DT16, Spec^r	Geddes et al., 2019	5
pOGG276	pL1V-Lv1-pUC18R6K-miniTn*7*T-Gm, Amp^r Gm^r	Beatriz Jorrín, Poole Lab	5
pOPS1107	PCR oxp1186 & oxp1187 for RL1340-sodB cloned in pK19mob-*Hin*dIII, Km^r	This work	3
pOPS0820	PCR amplified internal fragment of CheA1 (with oxp2200-oxp2201), cloned into pK19mob/*Hin*dIII as BD cloning.	Vinoy Ramachandran, Poole Lab	3
pOPS0822	PCR amplified internal fragment of fixB (with oxp2204-oxp2205), cloned into pK19mob/*Hin*dIII as BD cloning.	Vinoy Ramachandran, Poole Lab	3
pOPS0823	PCR amplified internal fragment of nifB (with oxp2206-oxp2207), cloned into pK19mob/*Hin*dIII as BD cloning.	Vinoy Ramachandran, Poole Lab	3
pOPS0824	PCR amplified internal fragment of nodB (with oxp2208-oxp2209), cloned into pK19mob/*Hin*dIII as BD cloning.	Vinoy Ramachandran, Poole Lab	3
pOPS0825	PCR amplified internal fragment of pRL90058 (with oxp2210-oxp2211), cloned into pK19mob/*Hin*dIII as BD cloning.	Vinoy Ramachandran, Poole Lab	3
pOPS0831	PCR amplified internal fragment of pRL120205 (with oxp2222-oxp2223), cloned into pK19mob/*Hin*dIII as BD cloning.	Vinoy Ramachandran, Poole Lab	3
pOPS0833	PCR amplified internal fragment of RL0149 (with oxp2226-oxp2227), cloned into pK19mob/*Hin*dIII as BD cloning.	Vinoy Ramachandran, Poole Lab	3

continued on next page

Table 3 – continued from previous page

Plasmid	Description	Source	Ch.
pOPS0836	PCR amplified internal fragment of RL1545 (with oxp2232-oxp2233), cloned into pK19mob/*Hin*dIII as BD cloning.	Vinoy Ramachandran, Poole Lab	3
pOPS0837	PCR amplified internal fragment of RL2606 (with oxp2234-oxp2235), cloned into pK19mob/*Hin*dIII as BD cloning.	Vinoy Ramachandran, Poole Lab	3
pOPS0838	PCR amplified internal fragment of RL3549 (with oxp2236-oxp2237), cloned into pK19mob/*Hin*dIII as BD cloning.	Vinoy Ramachandran, Poole Lab	3
pOPS0839	PCR amplified internal fragment of RL3654 (with oxp2238-oxp2239), cloned into pK19mob/*Hin*dIII as BD cloning.	Vinoy Ramachandran, Poole Lab	3
pOPS0846	PCR amplified internal fragment of pRL120198 (with oxp2252-oxp2253), cloned into pK19mob/ *Hin*dIII as BD cloning.	Vinoy Ramachandran, Poole Lab	3
pOPS0849	PCR amplified internal fragment of RL1496 (with oxp2258-oxp2259), cloned into pK19mob/*Hin*dIII as BD cloning.	Vinoy Ramachandran, Poole Lab	3
pOPS0851	PCR amplified internal fragment of RL0397 (with oxp2262-oxp2263), cloned into pK19mob/*Hin*dIII as BD cloning.	Vinoy Ramachandran, Poole Lab	3
pOPS1271	PCR amplified internal fragment of Rlv3841 RL0748-aglA1 (with oxp2829-oxp2830), cloned into pK18mob/*Hin*dIII as BD cloning, Km^r	This work	4
pOPS1272	PCR amplified internal fragment of Rlv3841 RL3565 (with oxp2831-oxp2832), cloned into pK18mob/*Hin*dIII as BD cloning, Km^r	This work	4
pOPS1273	PCR amplified internal fragment of Rlv3841 RL4139 (with oxp2833-oxp2834), cloned into pK18mob/*Hin*dIII as BD cloning, Km^r	This work	4
pOPS1274	PCR amplified internal fragment of Rlv3841 RL0339 (with oxp2835-oxp2836), cloned into pK18mob/*Hin*dIII as BD cloning, Km^r	This work	4
pOPS1275	PCR amplified internal fragment of Rlv3841 RL1130 (with oxp2837-oxp2838), cloned into pK18mob/*Hin*dIII as BD cloning, Km^r	This work	4
pOPS1276	PCR amplified internal fragment of Rlv3841 RL2036 (with oxp2839-oxp2840), cloned into pK18mob/*Hin*dIII as BD cloning, Km^r	This work	4

continued on next page

Table 3 – continued from previous page

Plasmid	Description	Source	Ch.
pOPS1278	PCR amplified internal fragment of Rlv3841 RL3624 (with oxp2841-oxp2842), cloned into pK18mob/*Hin*dIII as BD cloning, Km^r	This work	4
pOPS1279	PCR amplified internal fragment of Rlv3841 pRL70182 (with oxp2843-oxp2844), cloned into pK18mob/*Hin*dIII as BD cloning, Km^r	This work	4
pOPS1670	Backbone plasmid for facilitated marker exchange mutagenesis, Cm^r	Raphael Ledermann, Poole Lab	5
pOPS1713	Donor plasmid for Km cassette (eBFP2 removed), Km^r	Raphael Ledermann, Poole Lab	5
pOPS1714	pOPS1670-based mutagenic plasmid containing flanking regions of pRL100051 around casette from pOPS1713	This work	5
pOPS1715	pOPS1670-based mutagenic plasmid containing flanking regions of RL3820 around casette from pOPS1713	This work	5
pOPS1728	pOPS1670-based mutagenic plasmid containing flanking regions of pRL100133 around casette from pOPS1713	This work	5
pOPS1729	pOPS1670-based mutagenic plasmid containing flanking regions of pRL110047 around casette from pOPS1713	This work	5
pOPS1730	pOPS1670-based mutagenic plasmid containing flanking regions of pRL120397 around casette from pOPS1713	This work	5
pOPS1731	pOPS1670-based mutagenic plasmid containing flanking regions of pRL70027 around casette from pOPS1713	This work	5
pOPS1732	pOPS1670-based mutagenic plasmid containing flanking regions of pRL70050 around casette from pOPS1713	This work	5
pOPS1733	pOPS1670-based mutagenic plasmid containing flanking regions of pRL70154 around casette from pOPS1713	This work	5
pOPS1734	pOPS1670-based mutagenic plasmid containing flanking regions of pRL70155 around casette from pOPS1713	This work	5
pOPS1750	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 3 - pRL100432) around casette from pOPS1713	This work	5

continued on next page

Table 3 – continued from previous page

Plasmid	Description	Source	Ch.
pOPS1751	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 12 - pRL80079) around casette from pOPS1713	This work	5
pOPS1752	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 13 - pRL80080) around casette from pOPS1713	This work	5
pOPS1753	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 14 - pRL90263) around casette from pOPS1713	This work	5
pOPS1754	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 15 - pRL90320) around casette from pOPS1713	This work	5
pOPS1755	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 16 - RL0123) around casette from pOPS1713	This work	5
pOPS1756	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 17 - RL0143) around casette from pOPS1713	This work	5
pOPS1757	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 18 - RL0381) around casette from pOPS1713	This work	5
pOPS1758	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 19 - RL0398) around casette from pOPS1713	This work	5
pOPS1759	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 20 - RL0493) around casette from pOPS1713	This work	5
pOPS1760	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 21 - RL0501) around casette from pOPS1713	This work	5
pOPS1761	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 22 - RL0502) around casette from pOPS1713	This work	5
pOPS1762	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 23 - RL0546) around casette from pOPS1713	This work	5
pOPS1763	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 24 - RL0572) around casette from pOPS1713	This work	5

continued on next page

Table 3 – continued from previous page

Plasmid	Description	Source	Ch.
pOPS1797	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 25 - RL0652) around casette from pOPS1713	This work	5
pOPS1798	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 26 - RL0726) around casette from pOPS1713	This work	5
pOPS1799	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 27 - RL0727) around casette from pOPS1713	This work	5
pOPS1800	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 28 - RL0776) around casette from pOPS1713	This work	5
pOPS1801	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 29 - RL0903) around casette from pOPS1713	This work	5
pOPS1802	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 30 - RL0962) around casette from pOPS1713	This work	5
pOPS1803	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 31 - RL0994) around casette from pOPS1713	This work	5
pOPS1804	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 32 - RL1060) around casette from pOPS1713	This work	5
pOPS1805	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 33 - RL1217) around casette from pOPS1713	This work	5
pOPS1806	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 34 - RL1470) around casette from pOPS1713	This work	5
pOPS1807	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 35 - RL1528) around casette from pOPS1713	This work	5
pOPS1808	pOPS1670-based mutagenic plasmid containing flanking regions of (gene 36 - RL1717) around casette from pOPS1713	This work	5

Table 3: Plasmids used in this work

3 Exploration of Mutant Phenotypes Through INSeq Data

3.1 Introduction

In order to investigate which of approximately seven thousand *R. leguminosarum* genes are involved in symbiosis with legumes, the Poole Lab performed a high-throughput, genome-wide INSeq screen on the model strain Rlv3841. This screen has generated an INSeq dataset covering four successive stages of symbiosis: rhizosphere growth, root colonisation, undifferentiated bacteria within the host and terminally-differentiated bacteroids in nodules.

In this chapter, research was carried out with the objective of defining in which stages of symbiosis were affected by mutation of the selected genes of interest (Table 4; selection rationale in Chapter 3.2) in greater resolution than the symbiosis INSeq screen had allowed. Figure 3 depicts in general how different experiments aimed to identify which stages are affected, as well as the possible biological mechanisms for the INSeq phenotype observed.

The INSeq screen was performed by generating two Mariner-transposon mutant libraries of Rlv3841 (input1 & input2) and inoculating seven-day-old pea seedlings. The first mutant library was recovered five days later from the rhizosphere fraction of substrate (rhizosphere library) and from the washed and ground roots of the plants (root colonised library). The second mutant library was recovered at 28 days-post-inoculation (dpi) from a total of 150,000 picked and surface-sterilised nodules. Rhizobia were extracted from

these nodules and either sequenced directly (nodule bacteroid library) or regrown in TY liquid media for 12 hours before DNA extraction and sequencing (nodule bacteria library). This regrowth step was carried out in order to expand the population of undifferentiated bacteria so that they would predominate over differentiated bacteroids in the nodule bacteria library (Wheatley et al., 2020).

Figure 3

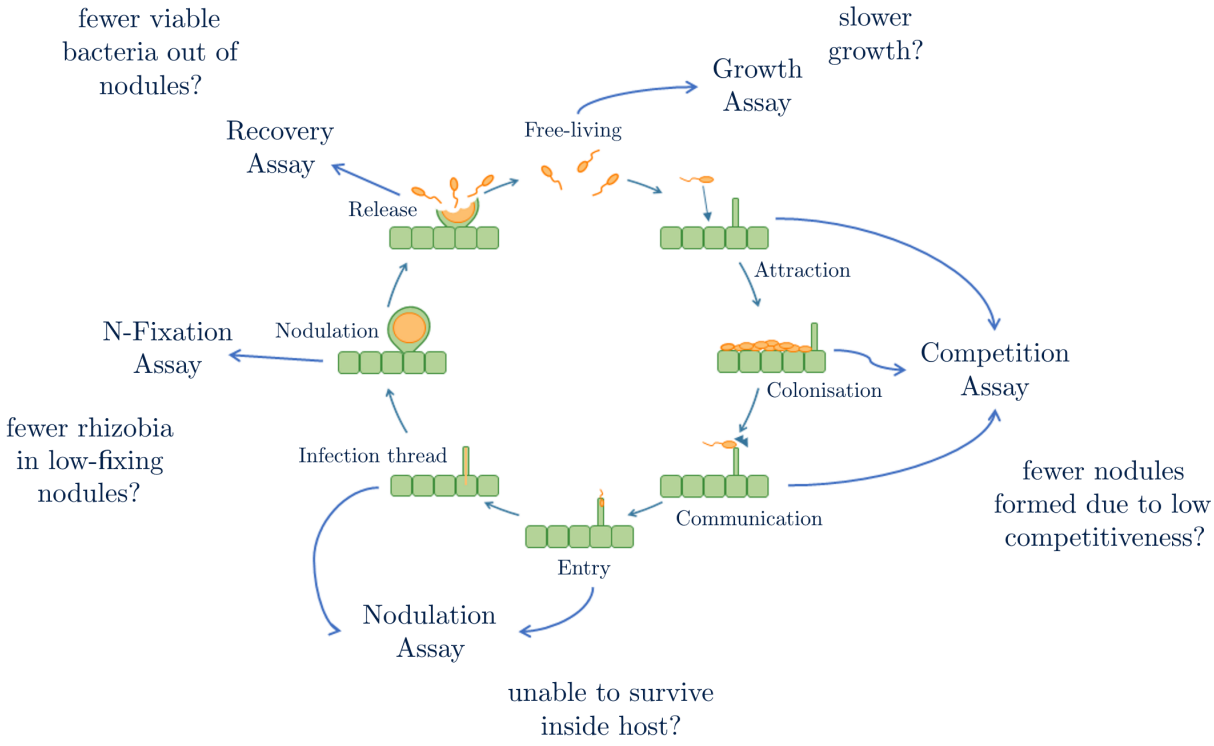


Figure 3. General overview of the stages of the *Rhizobium* symbiotic cycle (core cycle), which experiments were performed to identify effects at these stages (middle layer), and possible biological explanations for a effects at these stages (periphery). Adapted from Aroney, 2021.

The prepared and sequenced transposon mutant libraries were analysed with a Hidden Markov Model (HMM) (DeJesus et al., 2015). This compared the frequencies of reads for TA sites in each gene with the frequencies of reads of the rest of the library. Each gene was then assigned one of four states for that library, dependent on the outcome of the comparison: colors 'Neutral' (**NE**), 'Advantaged' (**AD**), 'Defective' (**DE**), or 'Essential' (**ES**). Genes were labelled as DE when the HMM found fewer reads than expected, indicating that a mutation to this gene conferred a reduction to fitness. For AD genes the opposite was the case, where there were more reads than expected, suggesting the mutant had a fitness advantage relative to wild type. Genes where TA sites had few or no reads were classified as ES, as it was likely that mutations to these genes are lethal, or at least severely deleterious. If the frequencies of reads for the TA sites of a gene were not considerably different from those of the library, then that gene was marked as NE to reflect the neutral impact of a mutation on fitness. An INSeq classification was given to each gene in each library, so the impact of mutation on fitness could be compared across different conditions, i.e. the four stages of symbiosis. It was often the case that a mutation would have a more severe effect in some stages than others, giving clues to at which stages of symbiosis these genes are involved (Wheatley et al., 2020).

3.2 Genes of Interest

Rationale of Selection

The smaller scale fitness and competition assays in this work required specific mutants to be generated. Fifteen genes were selected from the dataset for this purpose. Eleven genes were selected from the 211 genes classified as ES or DE by INSeq for the nodule bacteria and/or bacteroid output libraries (Wheatley et al., 2020). Four genes were selected as representatives of the 'nodule bacteria specific' group of genes (142 genes), meaning they were classified as ES or DE in the nodule bacteroid library specifically (NE in the nodule bacteroid library). The selected genes were chosen as representatives of a variety of functional roles and gene classes (catalysis, transport, regulation, structural, signalling), and lie on a spectrum from poorly characterised (RL1545) to well-studied (*fixB*/*nifB*/*nodB*). Genes without specific annotation ('hypothetical proteins') were not selected due to the additional difficulty of analysis for such genes. The 15 genes and their INSeq classifications are listed in Table 4.

pRL100186-*nodB* - Nodulation

The *nodABCIJ* cluster contains genes for the biosynthesis and export of lipochitooligosaccharide Nod factors, required for the establishment of symbiosis. The *nodB* gene encodes a putative chitooligosaccharide deacetylase, needed to synthesise the chitin backbone of Nod factors (reviewed in Poole et al., 2018; Wang et al., 2018). Disruption of any of the genes in this cluster would be expected to significantly impair a mutant's ability to nodulate host legumes, specifically at the 'communication' stage of symbiosis intiation (Figure 3), and result in a Nod$^-$ phenotype, thus *nodB* was selected as a Nod$^-$ mutant for comparison (Geurts & Bisseling, 2002).

Table 4

Gene ID	Dsgn	Description	TA	Vm	TY	RS	RC	NB	Bd
pRL100186	*nodB*	Putative chitooligosaccharide deacetylase NodB	17	NE	NE	NE	NE	DE	DE
pRL100195	*nifB*	FeMo cofactor biosynthesis protein NifB	43	NE	NE	NE	NE	NE	DE
pRL100199	*fixB*	Electron transfer protein FixB	25	NE	NE	NE	NE	DE	DE
pRL120205	*eryB*	Putative erythritol phosphate dehydrogenase EryB	27	ES	NE	NE	NE	ES	ES
pRL90058	*rmrR*	Putative TetR family regulator of RND family efflux transporter	9	NE	NE	NE	NE	ES	ES
RL0149		Putative XRE family transcriptional regulator	5	NE	NE	NE	NE	DE	DE
RL0688	*cheA1*	Putative chemotaxis protein CheA	14	NE	NE	NE	NE	ES	ES
RL1545		Putative short-chain dehydrogenase/reductase	7	NE	NE	NE	NE	ES	ES
RL2606	*purC1*	Putative phosphoribosylaminoimidazolesuccinocarboxamide synthetase	12	ES	NE	NE	NE	ES	ES
RL3549	*glnII*	Putative glutamine synthetase	25	NE	NE	NE	NE	DE	ES
RL3654	*pssD*	Putative polysacharide biosynthesis protein	10	NE	NE	NE	NE	ES	ES
pRL120198		Putative L-xyulose kinase	23	NE	NE	NE	NE	ES	NE
RL1340	*sodA*	Putative superoxide dismutase SodA	13	ES	NE	NE	DE	ES	NE
RL1496	*iolR*	Putative sugar-isomerase RpiR family regulator	13	NE	NE	NE	NE	ES	NE
RL3986	*ruvC*	Putative Holliday junction endodeoxyribonuclease RuvC	8	NE	NE	NE	NE	DE	NE

Table 4: INSeq classifications of genes selected for mutagenesis and characterisation. Dsgn = Designation; TA = Number of TA sites; VM = Vincent's media (minimal; Perry et al., 2016); TY = Tryptone Yeast media (rich); RS = Rhizosphere; RC = Root Colonisation; NB = Nodule Bacteria; Bd = Bacteroid; NE = Neutral; DE = Defective; ES = Essential. Data from Wheatley et al., 2020.

Table 5

Mutant	**Description**	**FL**	**At**	**Cl**	**Cm**	**En**	**IT**	**Nd**	**Re**
Δ*nodB*	Putative chitooligosaccharide deacetylase NodB				✓				
Δ*nifB*	FeMo cofactor biosynthesis protein NifB							✓	
Δ*fixB*	Electron transfer protein FixB							✓	
Δ*eryB*	Putative erythritol phosphate dehydrogenase EryB			✓					
Δ*rmrR*	Putative TetR family regulator of RND family efflux transporter			✓	✓	✓			
ΔRL0149	Putative XRE family transcriptional regulator			✓					
Δ*cheA1*	Putative chemotaxis protein CheA		✓						
ΔRL1545	Putative short-chain dehydrogenase/reductase								
Δ*purC1*	Putative phosphoribosylaminoimidazolesuccinocarboxamide synthetase	✓		✓				✓	
Δ*glnII*	Putative glutamine synthetase	✓		✓					
Δ*pssD*	Putative polysacharide biosynthesis protein				✓	✓	✓	✓	
ΔpRL120198	Putative L-xyulose kinase								✓
Δ*sodA*	Putative superoxide dismutase SodA								✓
Δ*iolR*	Putative sugar-isomerase RpiR family regulator								✓
Δ*ruvC*	Putative Holliday junction endodeoxyribonuclease RuvC								✓

Table 5: Summary of hypothetical affected stages of symbiosis for mutants. A checkmark (✓) indicates a predicted effect from mutation for a given gene at that stage of symbiosis. An effect at one stage may indirectly affect subsequent stages. Stages of symbiosis (Figure 3) are denoted as: FL = Free Living; At = Attraction; Cl = Colonisation; Cm = Communication; En = Entry; IT = Infection Thread; Nd = Nodulation; Re = Release.

pRL100195-*nifB* - Nitrogen fixation

The *nifB* gene lies on Rlv3841's symbiotic plasmid, pRL10, and encodes a FeMo cofactor biosynthesis protein known to be an essential component of the nitrogenase complex (Hirsch & Smith, 1987; Weidenhaupt et al., 1996). It has been predicted that the sanctioning of non-fixing nodules will lead to fewer rhizobia in these nodules compared to nodules formed by wild type (Westhoek et al., 2017), and this is reflected in the INSeq classification for *nifB* as DE in nodule bacteroids (Table 4).

pRL100199-*fixB* - Nitrogen fixation

As with *nifB*, the electron transfer protein encoded by *fixB* has been known for some time to be essential for symbiotic nitrogen fixation, but its mutation does not affect growth in free-living bacteria (Earl et al., 1987; Weidenhaupt et al., 1996). A *fixB* mutant has been included as another Fix$^-$ control for the phenotypic comparison experiments. A *fixB* mutant has been included as another Fix$^-$ control for the phenotypic comparison experiments. The *fixB* gene has been classifed as DE in both nodule bacteria and bacteroids (Table 4).

pRL120205-*eryB* - Sugar catabolism

The *eryB* gene is part of an operon involved in erythritol catabolism and thought to encode an erythritol phosphate dehydrogenase. It has been previously shown that mutation of the *R. leguminosarum* gene *eryF*, involved in erythritol transport, led to significantly impaired competitive nodulation of pea (Yost et al., 2006). An *eryB* mutant may have a similar phenotype, or suffer from erythritol toxicity that affects some erythritol catabolism mutants (Sperry & Robertson, 1975). Erythritol presence in root exudates has been

detected (Clayton et al., 2008), therefore erythritol toxicity may impair *eryB* mutants during the 'colonisation' stage of symbiosis initiation (Figure 3).

pRL90058-*rmrR* - Transport, regulation

The *rmrR* gene, encoding a TetR-family transcriptional repressor, is thought to regulate the expression of two multidrug efflux pumps, *rmrA* and *rmrB* (Ramos et al., 2005). These lux pumps are believed to allow rhizobia to resist the toxic effects of phytoalexin accumulation and other antimicrobial compounds produced by host legumes as defence against pathogens (Kuc, 1995). In the common bean symbiont *Rhizobium etli* CFN42, the *rmrAB* double mutant has been found to have significantly compromised competitiveness for nodulation (González-Pasayo & Martínez-Romero, 2000). Disruption of *rmrR* could lead to a phenotype similar to the *rmrAB* double mutant.

RL0149 - Structural, regulation

RL0149 encodes an XRE-family transcriptional regulator that is notable for being upregulated (by 11-fold) in mutants where the repressor RL0390-*praR* is knocked out (Frederix, 2010). *praR* mutants show a hyperattachment phenotype thought to be due to overproduction of EPS and cell-surface attachment factors which are no longer repressed by *praR*, with this leading to greater biofilm formation, root attachment and competitiveness against wild type for nodulating pea (Frederix et al., 2014). If RL0149 is involved in the induction of the EPS and cell-surface attachment factors that *praR* represses, disruption of RL0149 might be expected to lead to defective attachment and reduced competitiveness for nodulation.

RL0688-*cheA1* - Chemotaxis

The gene products of the chemotaxis gene clusters enable rhizobia to swim towards chemoattractants, such as root exudates released from potential host legumes (Aroney et al., 2021; Vande Broek & Vanderleyden, 1995). Mutations in Rlv3841 disrupting the major chemotaxis gene cluster, *che1*, were found to significantly reduce the competitiveness of the mutant when competing against wild type to nodulate pea, although disruption of the second chemotaxis gene cluster, *che2*, did not (Miller et al., 2007). A *cheA1* gene

RL1545 - Metabolism

The protein encoded by RL1545 is not well characterised, but a role as a short-chain reductase/dehydrogenase is inferred from homology. BlastP analysis shows 40% identity to the product of the *E. coli* gene *yciK*, placing RL1545 in the NADB-Rossman family of oxidoreductases. Lack of characterisation of *yciK* precludes further speculation about the role of RL1545.

RL2606-*purC1* - Nucleotide biosynthesis

The gene product of *purC1* catalyses the reaction

5-amino-1-(5-phospho-D-ribosyl)imidazole-4-carboxylate + ATP + L-aspartate $\rightleftharpoons$ (2S)-2-[5-amino-1-(5-phospho-β-D-ribosyl)imidazole-4-carboxamido]succinate + ADP + 2 H^+ + phosphate

as part of de novo purine biosynthesis. Adenine auxotrophy from the disruption of purMN has been previously shown in *R. leguminosarum* to result severely reduced competitiveness

for nodulating pea and nodules unable to fix nitrogen (Stevens et al., 2000). Disruption of *purC1* might be expected to result in a similar phenotype in N-fixation and competitive nodulation assays. Rlv3841 possesses a second putative phosphoribosylamidoimidazole-succinocarboxamide synthase gene, RL3872-*purC2*, but PurC2 only has 34.39% peptide sequence identity with PurC1, while *purC2* has milder INSeq classifications than *purC1* (*purC2* is DE where *purC1* is ES), overall suggesting that *purC1* is the primary gene for this function.

RL3549-*glnII* - Amino acid biosynthesis

The *glnII* gene encodes one of the key glutamine synthetases in rhizobia, an enzyme that is unusual among bacterial glutamine synthases by having a structure similar to the eukaryotic enzyme (Carlson et al., 1987). In Rlv3841, *glnII* was upregulated in N-limited conditions, as the initial rhizosphere of pea would be during the vermiculite/N-free rooting solution INSeq experiments (Ramachandran et al., 2011). A reduction in the ability to synthesis glutamine in this N-limited environment would be expected to lead to impaired growth and a reduction in ability to competitively nodulate.

RL3654-*pssD* - Structural biosynthesis

The *pssD* gene is part of the *pssCDE* gene cluster, which is involved in exopolysaccharide (EPS) biosynthesis (Janczarek et al., 2015). Mutation of *pssD* has been previously shown to prevent successful nodulation of clover by the *R. leguminosarum* bv. *trifolii* strain TA1. The non-fixing nodules induced by the mutant strain were found to lack symbiosomes, suggesting EPS was essential for successful endocytosis of rhizobia into plant nodule cells (Król et al., 1998). This phenotype may extend to *R. leguminosarum* bv. *viciae* 3841 for the nodulation of pea.

pRL120198 - Sugar catabolism

The pRL120198 gene is predicted to encode an L-xylulose kinase from sequence homology. Xylulose kinases are required for xylulose catabolism into the pentose-phosphate pathway (Lawlis et al., 1984). The ability to utilise all available plant-supplied photosynthates in the nodule could be expected to influence the vigour of recovered nodule bacteria (Duncan, 1981), which could be one mechanism behind the decreased fitness of pRL120198 mutant nodule bacteria observed in the INSeq screen (Table 4).

RL1340-*sodA* - Oxidative protection

The superoxide dismutase encoded by *sodA* is known in many bacteria to protect against reactive oxygen species (Hassett et al., 1995; McCord & Fridovich, 1969). This is important for the infection of legume hosts by rhizobia, as the infecting rhizobia need to survive an oxidative burst from the host plant immune system early in the infection process (Berrabah et al., 2019; Santos et al., 2001). In *R. leguminosarum*, the SodA enzyme is targeted to the periplasm for this function (Krehenbrink et al., 2011). Disruption of *sodA* could lead to a mutant becoming more vulnerable to plant immune responses and less healthy after recovery from nodules.

RL1496-*iolR* - Sugar catabolism

In the alfalfa symbiont *Sinorhizobium meliloti*, the RpiR-like repressor IolR has been found to regulate a set of genes involved in inositol catabolism (Kohler et al., 2011). High levels of inositol compounds have been found in legume rhizospheres and in nodules (Wood & Stanway, 2001). *S. meliloti* and *R. leguminosarum* mutants defective in inositol catabolism and transport have been found to be unable to compete with wild type to

nodulate host legumes, possibly due to the putative importance of myo-inositol in early nodule development (Fry et al., 2007; Kohler et al., 2010).

RL3986-*ruvC* - DNA repair

The *ruvC* gene has been found in *E. coli* to encode an endonuclease responsible for resolving Holliday junctions, an important process in DNA repair and recombination (Dunderdale et al., 1991). Unresolved Holliday DNA secondary structures in *ruvC* mutants arrest the movement of nucleic acid polymerases and lead to the induction of DNA repair mechanisms (Iwasaki et al., 1991). If the stressful environment of a legume nodule is damaging to rhizobial DNA, then mutants defective in DNA repair would not be expected to grow as quickly after recovery from nodules (Berrabah et al., 2019; Santos et al., 2001). The growth of *ruvC* mutants might also be impaired as a the result of DNA synthesis being slowed or stopped by unresolvable Holliday structures (Asai & Kogoma, 1994).

3.3 Generation of Mutants with the pK19mob system

The pK19mob mutagenesis system (Schäfer et al., 1994) was used to generate 15 mutants with specific disruptions of the chosen genes (Table 4). Using this technique, the *E. coli* plasmid is conjugated from *E. coli* into *Rhizobium*, where it is integrated by homologous recombination into the Rlv3841 genome at specific sites with homology to internal fragments of the gene of interest (GOI) cloned into its multiple cloning site (MCS). This system is well-established for use in multiple *Rhizobium* species (Becker et al., 1993; Karunakaran et al., 2010; Sudtachat et al., 2009; Wang et al., 2019), having advantages in speed and simplicity due to the requirement for a single internal gene fragment for a single homologous recombination (Figure 4). Additionally, insertion mutations generated by pK19mob are structurally similar to mariner transposon insertion mutations generated by pSAM_Rl, making pK19mob mutants more comparable to the INSeq mutant library (Goodman et al., 2009; Perry & Yost, 2014).

Fifteen mutants were made in a *gusA*-marked Rlv3841 background strain using this protocol and confirmed by sequencing (Sánchez-Cañizares & Palacios, 2013). Figure 4 shows the integration of pK19mob into the rhizobial genome and the primers used to verify mutagenesis by PCR. From the observation of the bacteria, it was noted that the *pssD* mutant had visibly flat, dry bacterial colonies when grown on TY agar plates, as could be expected for a mutant defective in EPS production (Janczarek & Skorupska, 2007). All other mutants were indistinguishable as colonies on plates from wild type bacteria by eye. These 15 mutants were then assessed by a series of five experiments and assays, described in the following sections, to determine at which stage of symbiosis the mutation has an effect.

Figure 4

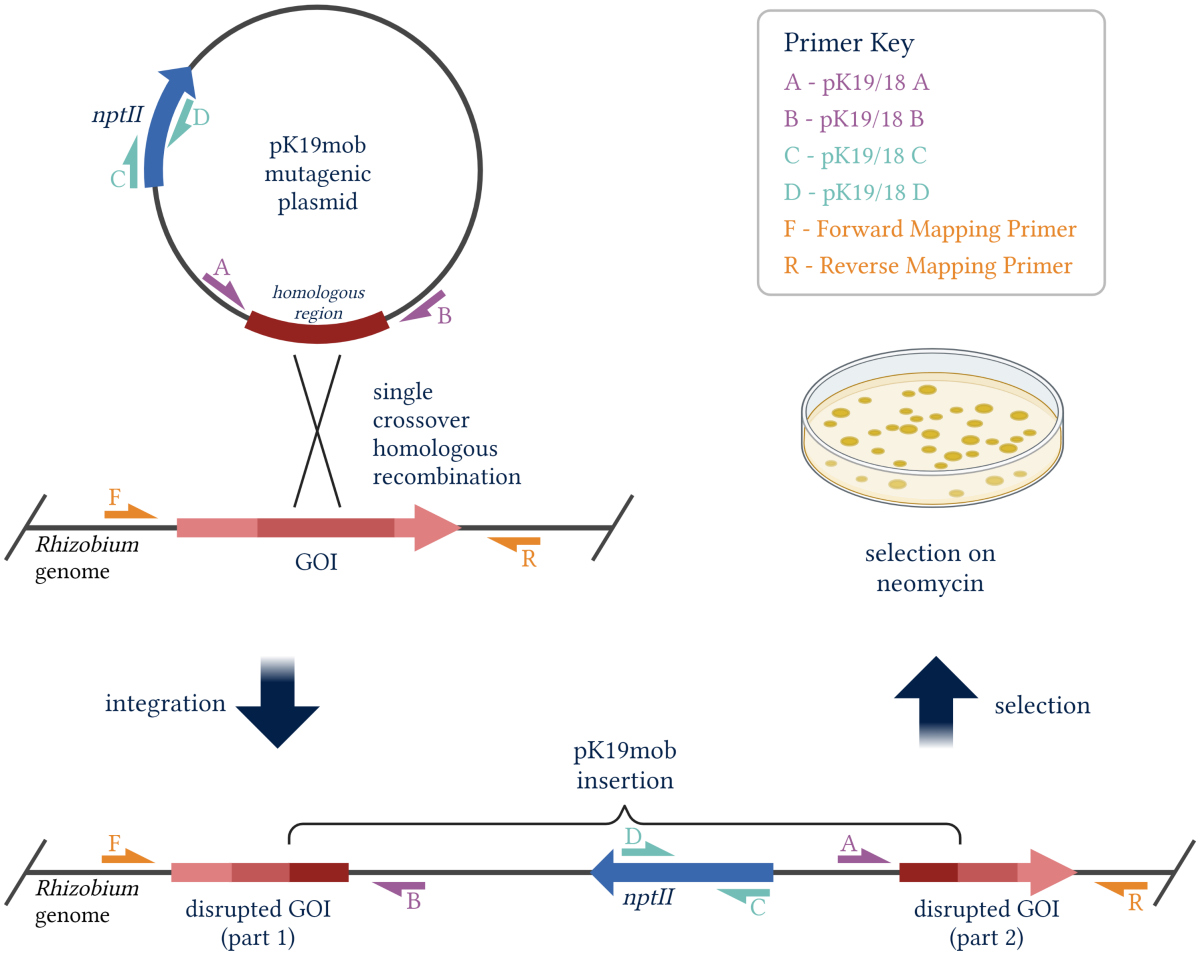


Figure 4. Integration of pK19mob into *Rhizobium* genome by single-crossover homologous recombination. Primers pK19/18 A and pK19/18 B are used to sequence the internal gene fragment cloned into the MCS. Mapping primers (unique to each GOI) are used in pairs or with pK19/18 A/B primers to verify the integration of the plasmid into the target gene. pK19/18 C and pK19/18 D can be used to distinguish spontaneous neomycin resistant mutants from mutants carrying pK19mob. Created with BioRender (biorender.com).

3.4 Examination of Nodulation Ability by Single-Strain Inoculations

Each of the fifteen mutants (Table 4) was initially assessed in a single-strain nodulation assay to determine if the strain was able to form nodules on pea roots. Peas were inoculated with 10^8 CFU of mutant or wild type bacteria (detailed protocol in Chapter 2.4). Plants were harvested after four weeks' growth and the size and shape of nodules were examined (Figure 5).

Differences in nodule morphology from those of wild type were observed only for the *pssD*, *fixB*, and *nifB* mutant strains (Figure 5). The *pssD* mutant (Figure 5D) did not form nodules, instead forming small, white growths on the roots. This failure to nodulate fits with the observed INSeq phenotype of the *pssD* mutant (Table 4) and shows that it is due to a direct effect of the disrupted gene and not due to the effect of competition from other rhizobial strains. The *fixB* and *nifB* mutants (Figure 5B & 5C) induced nodules that were smaller and paler than wild type nodules (Figure 5A) to form. Such nodules are typical of a Fix^- strain and are likely to be the result of sanctioning, as the plant does not provide resources to grow nodules that do not fix nitrogen (Westhoek et al., 2017). These small, white nodules might be expected to contain fewer bacteroids and bacteria than wild type nodules, which would give rise to the INSeq phenotypes predicted (both DE in bacteroid, Table 4). The nitrogen-fixing abilities of these mutants are measured later in Chapter 3.5.

Figure 5

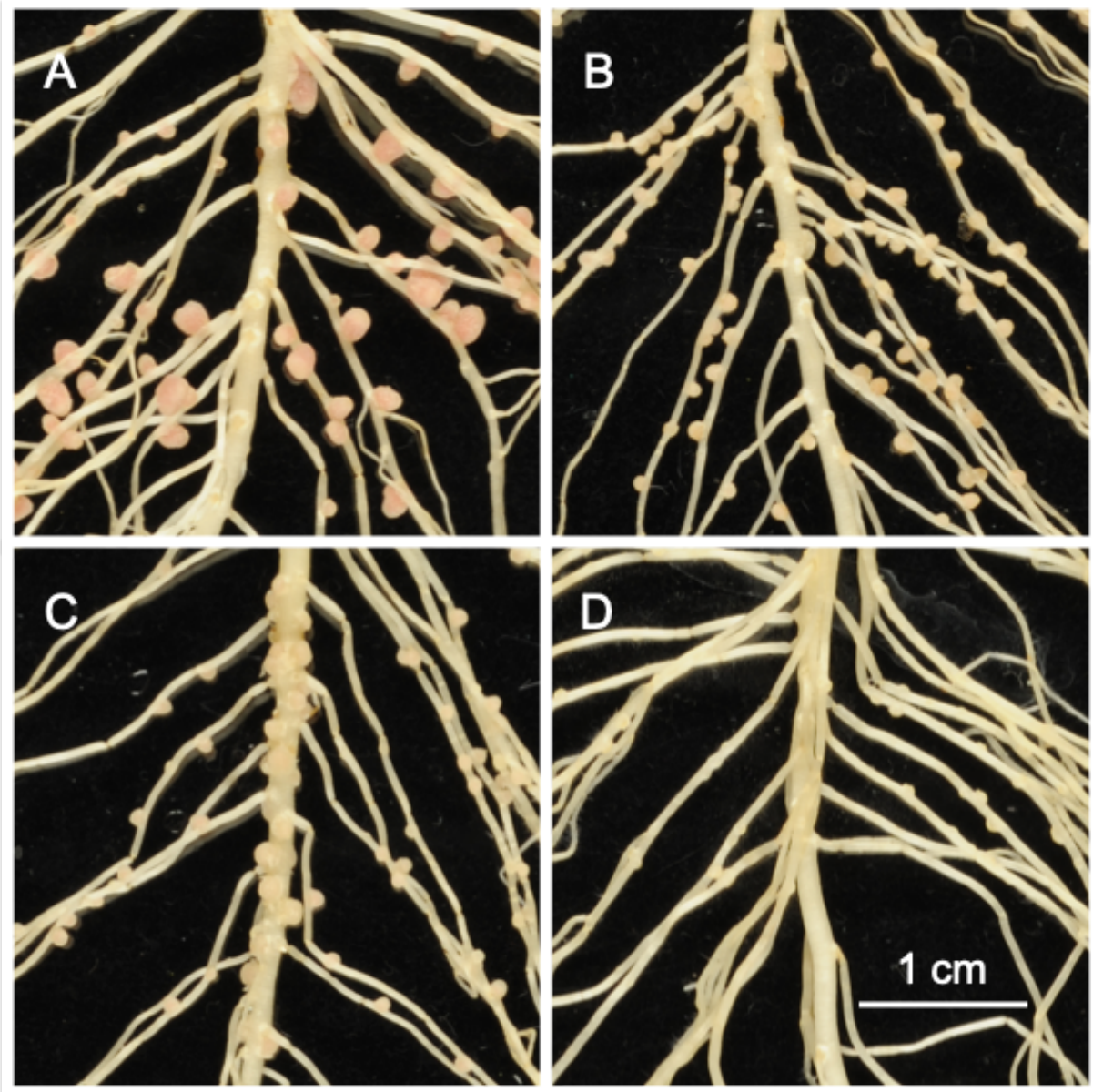


Figure 5. Nodule morphology from different strains three weeks post inoculation onto pea. The control strain 3841*gusA* (A) induced large healthy pink nodules while the Fix$^-$ mutants of *nifB* (B) and *fixB* (C) induced around twice the number of smaller, whiter nodules. The *pssD* mutant (D) induced small white growths on roots which did not show *gusA* activity when stained with X-GlcA. The plants photographed here were independent from the nodulation assay represented in Figure 6. The scale bar represents 1 cm.

Roots were stained with X-GlcA to turn nodules blue, making it easier to count the numbers of nodules per plant. Figure 6 shows that the *fixB* and *nifB* mutants induced around twice as many nodules on the plant as wild type. This can be explained by the plant allowing more nodules to form as it becomes progressively nitrogen-starved. All other mutant strains were statistically the same as wild type (Dunnett's multiple comparisons test).

Figure 6

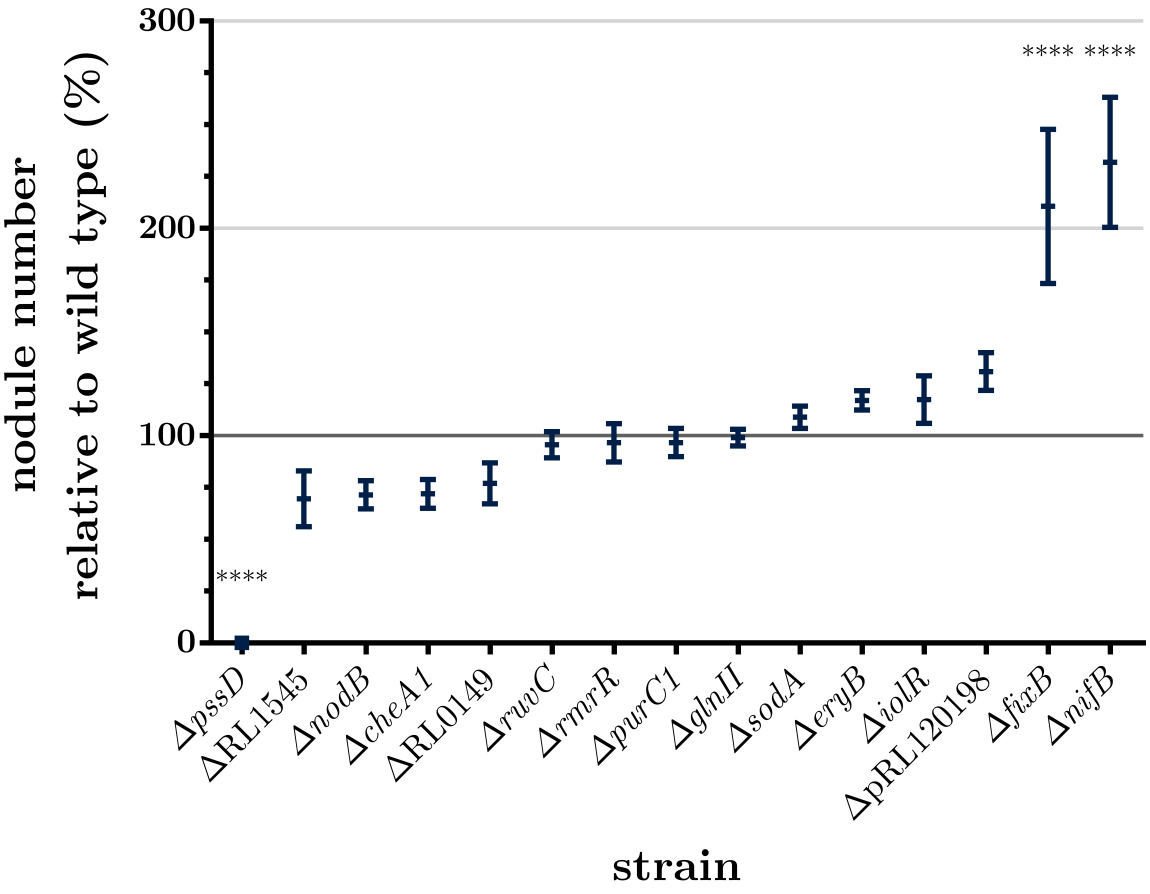


Figure 6. Nodules per plant from single strain mutant inoculations. Nodule numbers are presented relative to wild type (3841*gusA*, 180 nodules per plant ±19.2, 100%). Error bars represent ±SEM. Statistical significance was assessed by carrying out a Dunnett's test comparing each mutant strain to the wild type. An asterisk (*) indicates a statistically significant (p-value < 0.05) difference from 3841*gusA*. n = 3 biological replicates, each consisting of a single plant.

It is surprising that the *nodB* mutant forms approximately 75% of wild type nodule numbers. It should not be able to form any nodules as it is unable to produce the lipochitooligosaccharides (Nod factors) which are essential to initiate signalling for establishing symbiosis (Geurts & Bisseling, 2002). In order to try to resolve this discrepency, bacteria from six surface-sterilised *nodB* mutant nodules were extracted using a nodule extraction protocol and plated on non-selective TY agar plates to generate single colonies.

Approximately 24 colonies from each nodule were transferred to selective TY·Str_{500}·Nm_{80} agar plates, where only rhizobia containing a pK19 insertion should grow. It was found that three nodules contained only rhizobia lacking pK19, while the other three contained a 50-50 mixture of rhizobia with and without pK19 (example plates shown in Figure 7). Thus, nodulation is likely to result from bacterial revertants arising from the instability of pK19 integrated into the genome. This, combined with the potential for Nod^- rhizobia to be complemented by Nod factors produced by nearby Nod^+ rhizobia, is likely to have led to the 'mixed' nodules observed (Geurts & Bisseling, 2002).

Figure 7

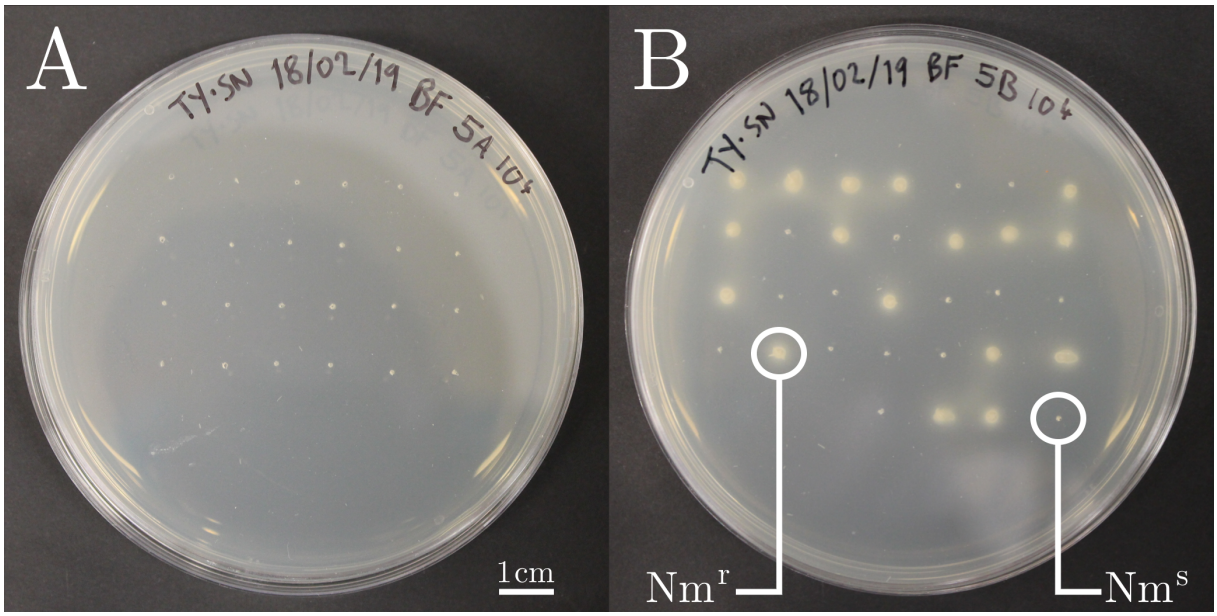


Figure 7. Example plates showing the recovery of bacteria from two surface-sterilised nodules following single-strain inoculation with the *nodB* mutant. Recovered bacteria were first streaked out to single colonies on non-selective media, then transferred to neomycin plates where only pK19 mutants are able to grow. (A) Plate containing neomycin which shows none of the transferred colonies have grown, indicating the bacteria recovered from the nodule are all revertants. (B) Plate containing neomycin showing about half of the transferred colonies have grown, indicating a 'mixed' nodule containing both pK19 mutant (Nm^r) and revertant (Nm^s) bacteria. The scale bar represents 1 cm.

3.5 Measurement of Mutant Nitrogenase Activity by Acetylene Reduction Assays

Having established that all mutant strains apart from the *pssD* and *nodB* mutants are able to form nodules, nitrogen-fixing ability of the mutant strains was investigated. Acetylene reduction assays were carried out to measure whole-plant nitrogenase activity as a proxy for nitrogen fixation (Shah et al., 1975). The *nodB* mutant was not included in this assay as the issue with revertants had been identified (Figure 7). The results of these assays are shown in Figure 8.

As expected, the plants inoculated with *pssD* exhibited no nitrogenase activity as they formed no nodules. In addition, the Fix$^-$ strains carrying mutation in *fixB* and *nifB* showed a lack of nitrogenase activity. Impairment in nitrogen-fixing ability could lead to mutant nodules being sanctioned and starved of resources by the host plant (Westhoek et al., 2017). This would result in fewer mutant bacteroids and bacteria in those nodules and lead to the INSeq phenotypes observed (Table 4). The other 11 mutants showed nitrogenase activity that was not significantly different from that of wild type, according to a Dunnett's multiple comparisons test.

Figure 8

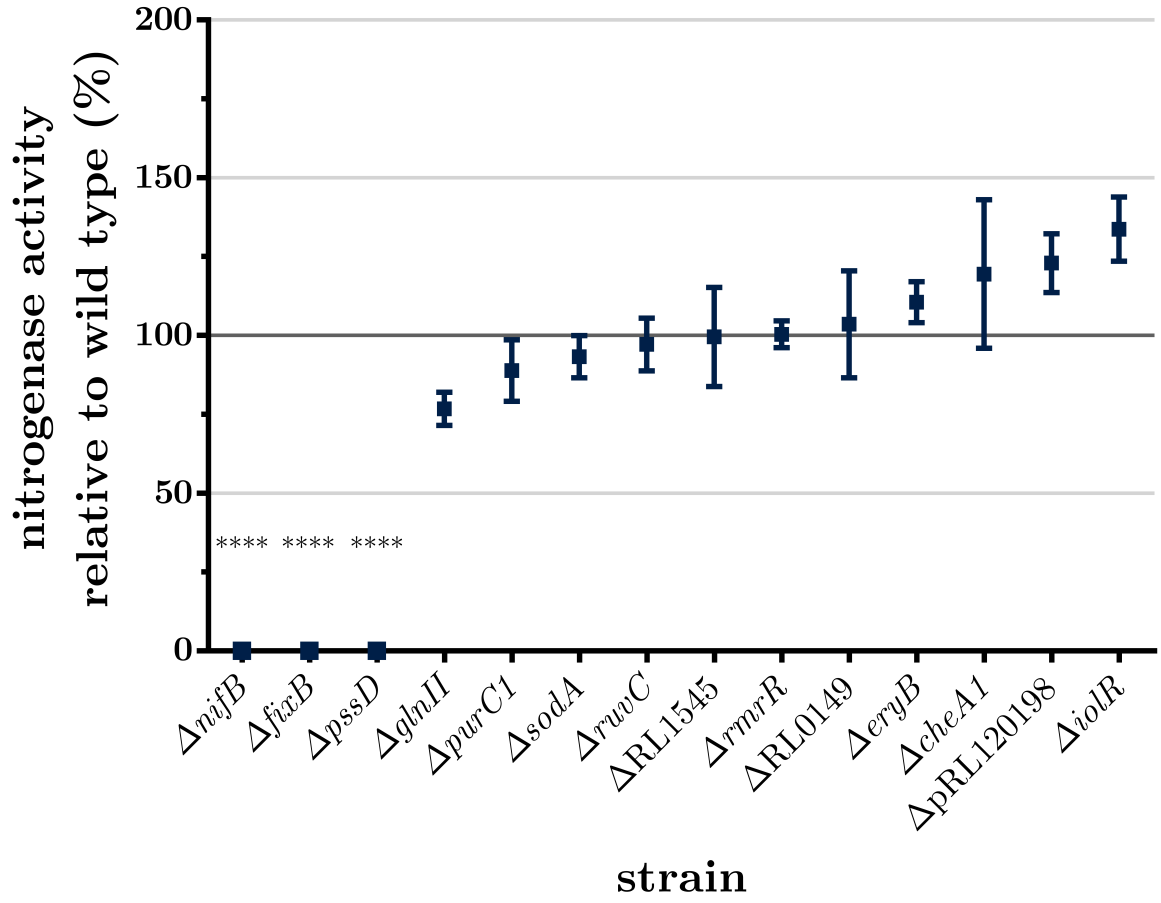


Figure 8. Whole plant nitrogenase activity as measured by acetylene reduction assays. Nitrogenase activity is presented relative to wild type (3841*gusA*, 5.18 µmoles ethylene per plant per hour ±0.380, 100%). Error bars represent ±SEM. Statistical significance was assessed by carrying out a Dunnett's test comparing each mutant strain to the wild type (3841*gusA*). Four asterisks (****) indicate a statistically significant (p-value < 0.0001) difference from 3841*gusA*. n = 3 biological replicates, each consisting of a single plant.

3.6 Quantification of Mutant Competitive Nodulation Ability by Co-Inoculation Experiments

Competitive nodulation assays were carried out to assess the competitiveness of a subset of mutant strains against wild type. This subset excluded the four mutants which were NE for the nodule bacteroid dataset (the *iolR*, *ruvC*, *sodA*, and pRL120198 mutants), and excluded the *fixB*, *nifB* and *nodB* strains as these were previously shown to be unable to form nitrogen-fixing nodules (Chapters 3.4 and 3.5). The *pssD* mutant was included as a negative control.

In competitive nodulation assays, peas were inoculated with a 1:1 mixture of *gusA*-marked mutant and *celB*-marked wild type (3841*celB*) cells. The ratios of strains in the inocula were verified by plating on TY agar plates containing X-GlcA, a colourless compound which is converted to a blue pigment by the GusA activity of the *gusA*-marked strain, staining these colonies blue. Nodule occupancy was determined by double-sequential staining to resolve nodules occupied by a *gusA*-marked strain as magenta and nodules occupied by a *celB*-marked strain as blue. Mixed nodules containing both *gusA*- and *celB*-marked strains were stained a dark blue/purple colour. This double-sequential staining is illustrated by Figure 9, which shows the stained roots of a plant inoculated with wild type Rlv3841 marked with *celB* and the *rmrR* mutant marked with *gusA*.

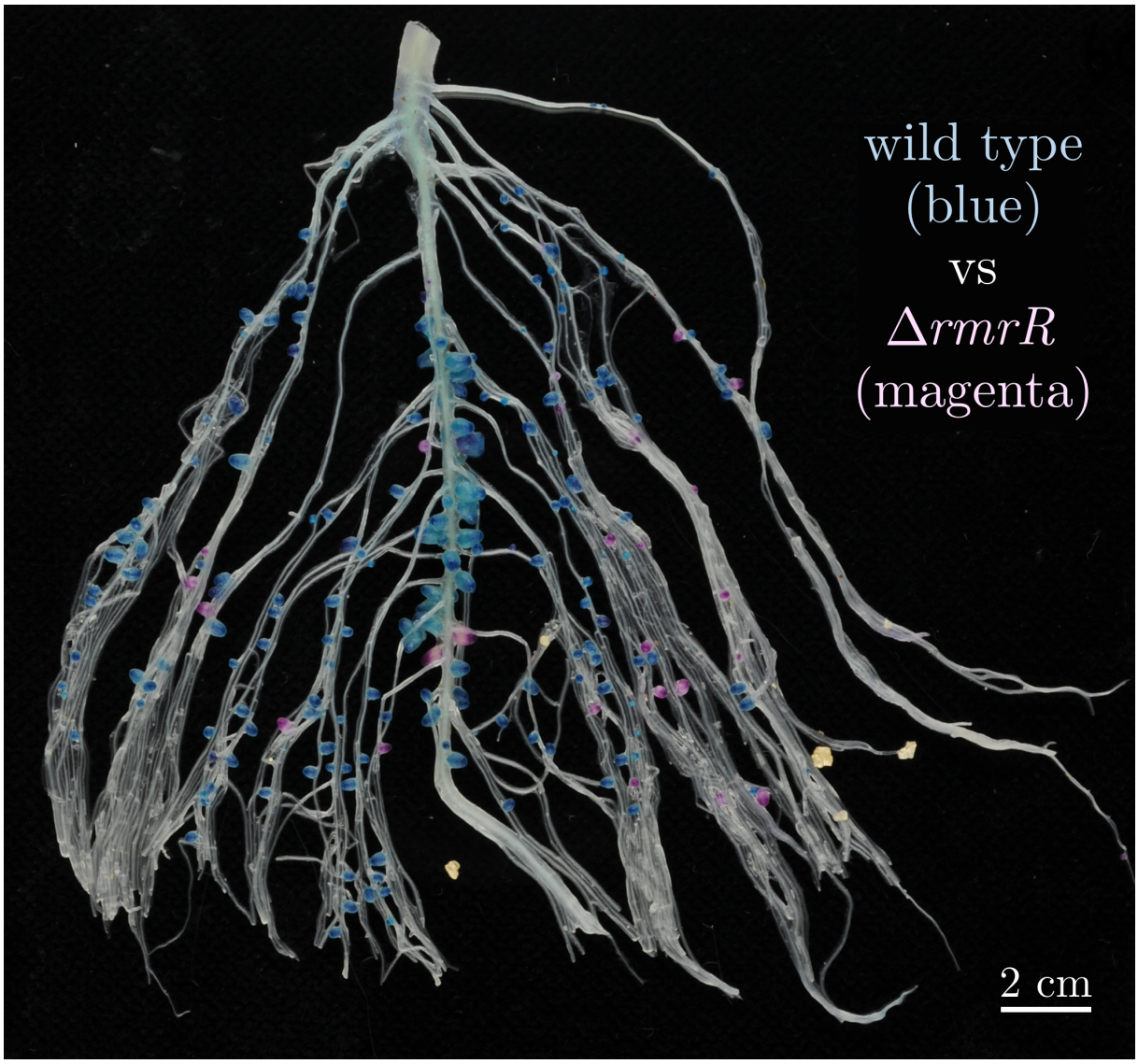


Figure 9. Double-sequentially stained roots of a pea plant. This plant was inoculated with a 1:1 mixture of wild type Rlv3841 marked with *celB* (3841*celB*) and the *rmrR* pK19mob mutant marked with *gusA* (Δ*rmrR*). Nodules occupied by the wild type strain are stained blue by the β-galactosidase activity of CelB on X-Gal, while nodules occupied by Δ*rmrR* are stained magenta by the β-glucuronidase activity of GusA on Magenta-GlcA. Nodules with both strains present are stained purple by the combined activities of both enzymes on their substrates. In this photograph there are visibly fewer magenta nodules than blue nodules, illustrating the competitive defect of Δ*rmrR*. The scale bar represents 2 cm. Photograph by John Baker.

The ratio between the proportion of a strain in nodule occupancy and the proportion of that strain in the inocula was used as a simple measure of competitiveness (a ratio below 1.0 is less competitive than wild type). It was decided to use this ratio instead of the competitive index from Equation 2 for two reasons (Beattie et al., 1989). Firstly, the constant k is derived from the slope of the curve between nodule occupancy and inoculum, which requires multiple inoculum ratios to be tested, e.g. 1:9 and 9:1. These competition assays were carried out only with 1:1 inoculum ratios for speed and resolving power, so it was not possible to accurately calculate k. Secondly, for strains that were unable to nodulate under competitive conditions ($\Delta pssD$ and $\Delta eryB$), the competitive index would be infinitely negative. These competitive indices cannot be easily represented with the rest of the data on a graph.

The competitive ratios for the eight mutants tested are shown in Figure 10. It was found that all the mutant strains tested showed a significant competitive disadvantage against wild type (mixed effects model comparing inoculum proportion to nodule proportion), except for RL1545, which encodes a short chain reductase. Looking more closely at the INSeq results, RL1545 has a low number of TA sites where the mariner transposon could insert (only 7 sites, Table 4), and these TA sites had a very low number of reads in the input library (average 6 and maximum 20, compared to a library non-zero average of 24.3). These factors make this particular INSeq classification less reliable as the low number of mutants present in the input library make it more likely that the mutants could be lost from the output library by chance (Perry & Yost, 2014). However, during the original INSeq experiment the mutant strains were competing against hundreds of thousands of other mutant strains, many of which would have a transposon insertion that

does not affect competitiveness (Wheatley et al., 2020). This competitive environment could lead to mutants with relatively subtle deficiencies in fitness or nodulation ability to be lost from the mutant library, so it is possible that the RL1545 mutant has a competitive defect that is too subtle to be identified in this assay. Alternatively, the RL1545 mutant does not have a competitive defect and another mechanism is responsible for the INSeq classification of the RL1545 gene. The other mutants showed varying degrees of competition deficiency, from the mild deficiency of the RL0149 (encoding a putative XRE family transcriptional regulator) mutant to the complete inability of the *eryB* mutant to nodulate under competitive conditions (Figure 10).

Figure 10

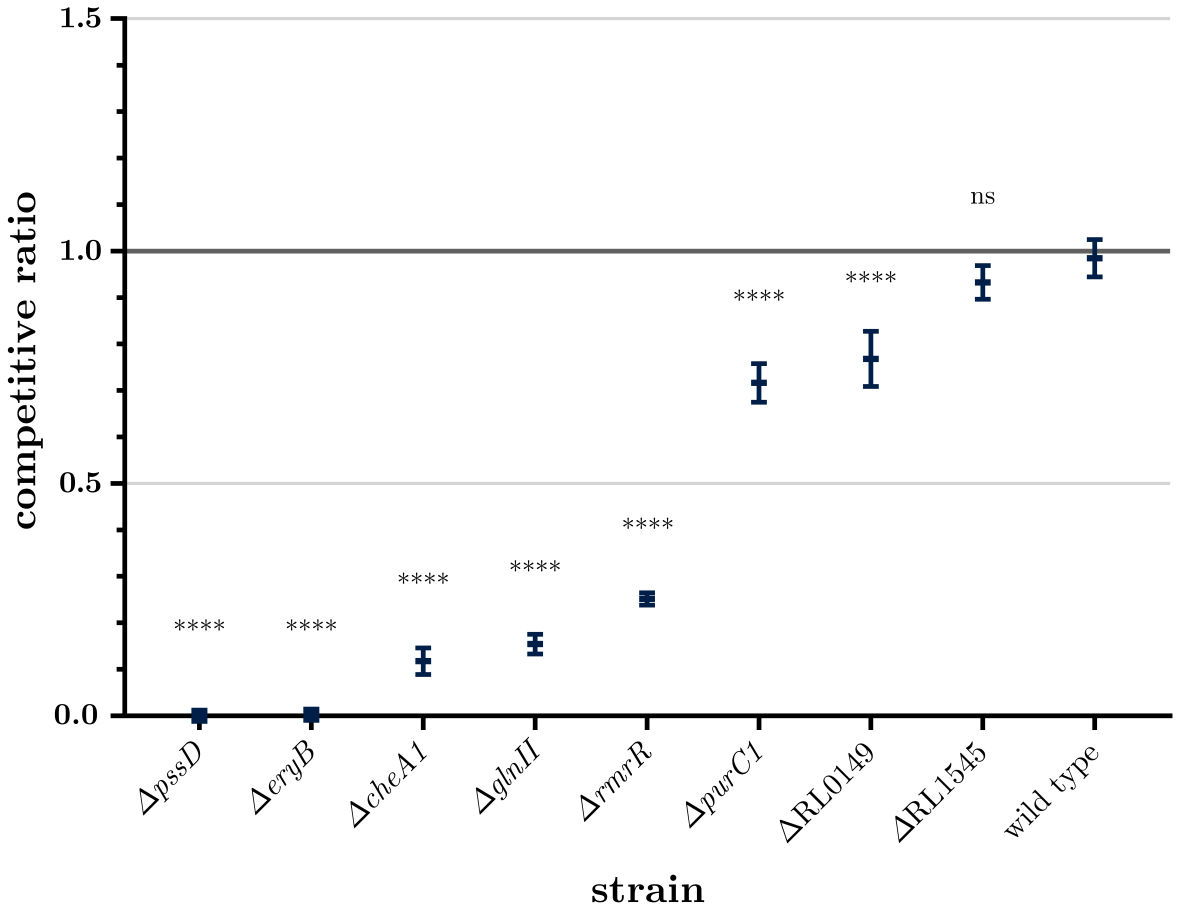


Figure 10. Ratio between the proportion of each strain in nodule occupancy and the proportion of that strain in the inocula in competitive nodulation assays. The competing *gusA*-marked strains were combined with 3841*celB* in a 1:1 ratio. In this measure of competitiveness, a ratio below 1.0 is less competitive than wild type. Error bars represent ±SEM. Statistical significance was assessed by carrying out a mixed model analysis between the proportion of each inoculum and its nodules as the competing strain. Four asterisks (****) indicate a statistically significant (p-value < 0.0001) difference between the proportion of a strain in the inoculum and the proportion of nodules of that strain from a mixed effects model. n = 5 biological replicates, each consisting of a single plant.

3.7 Investigation of Mutant Viability From Nodules by Nodule Bacteria Recovery Assays

Another group of genes identified by INSeq are the recovery-specific genes. These genes have been classified as ES or DE for recovered nodule bacteria, but NE for other conditions including nodule bacteroids. Four genes with this phenotype were included in the 15 pK19 mutants made: *iolR*, *ruvC*, *sodA*, and pRL120198 (Table 4).

In the original INSeq experiment, surface-sterilised nodules were ground to extract undifferentiated nodule bacteria. These extracts were used to inoculate liquid TY media which was then grown at 28°C for 12 hours. DNA was then extracted from this culture and sequenced for INSeq. Recovery assays were carried out using nodules picked from plants inoculated with either of these four mutants (*iolR*, *ruvC*, *sodA*, and pRL120198) or a wild type. The recovery assays were carried out similarly to the original INSeq recovery but were cultured for 36 hours instead of 12 hours. After 36 hours, the number of bacteria in each culture was estimated by measuring the optical density of each culture at 600 nm (Figure 11). It was found that the cultures of all four mutants had significantly fewer cells than the wild type cultures (Dunnett's multiple comparisons test). These results agree with the phenotypes predicted by INSeq. The observed phenotypes may be due to fewer viable undifferentiated mutant bacteria present in each nodule, or that mutant bacteria are slower to recover and start growing after passage through the hostile environment of the nodule.

Figure 11

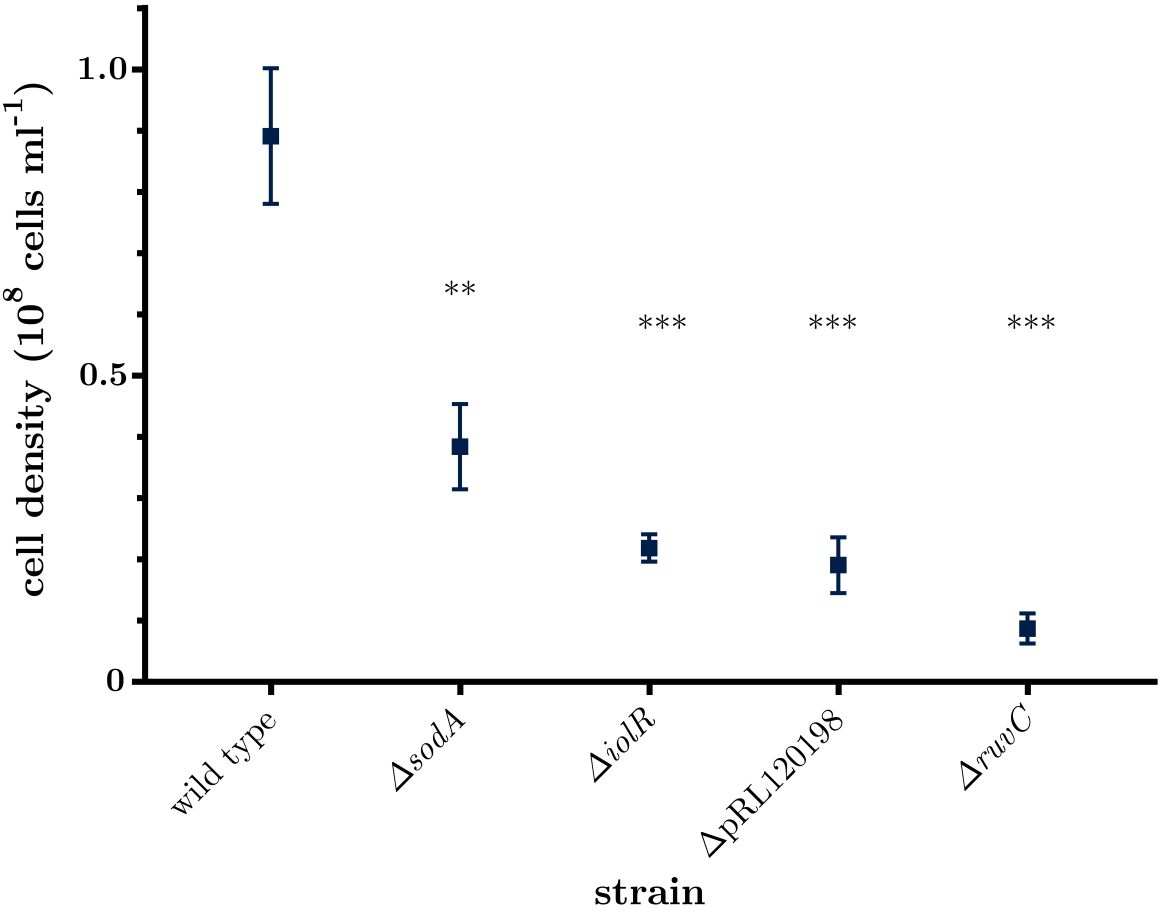


Figure 11. Growth of wild type and four mutant strains following recovery of bacteria from nodules. Cell densities were estimated following measurement of optical densities of 36-hour liquid TY cultures. Error bars represent ±SEM. Statistical significance was assessed by carrying out a Dunnett's test comparing each mutant strain with wild type (3841 *gusA*). Double-asterisks(**) represents significance where p-value < 0.01, triple-asterisks (***) represents significance where $p < 0.001$. n = 3 biological replicates, each consisting of a single plant.

3.8 Measurement of Mutant Growth Rate as Free-Living Bacteria

While none of the four selected recovery-specific genes were classified as ES or DE in TY media alone, the Hidden Markov Model can be conservative, and it is possible that the phenotypes observed in the recovery assay are manifestations of underlying growth defects in liquid TY media. To investigate this possibility, the four recovery-specific mutants were grown in liquid TY media in a plate reader, with OD_{600} measured every 30 minutes for 48 hours. These growth data were used to plot growth curves, which were then used to find the mid-logarithmic phase growth rate for the strains (Figure 12).

Analysis of the growth data showed that the strain mutated in *ruvC* had a significantly longer mean generation time than wild type (3.55 hours compared to 2.98 hours), while the other mutants did not show a significant difference (Dunnett's multiple comparisons test, Figure 12). It is therefore likely that the *ruvC* mutant is not a true recovery-specific mutant, but instead carries a general defect for growth in TY media.

Figure 12

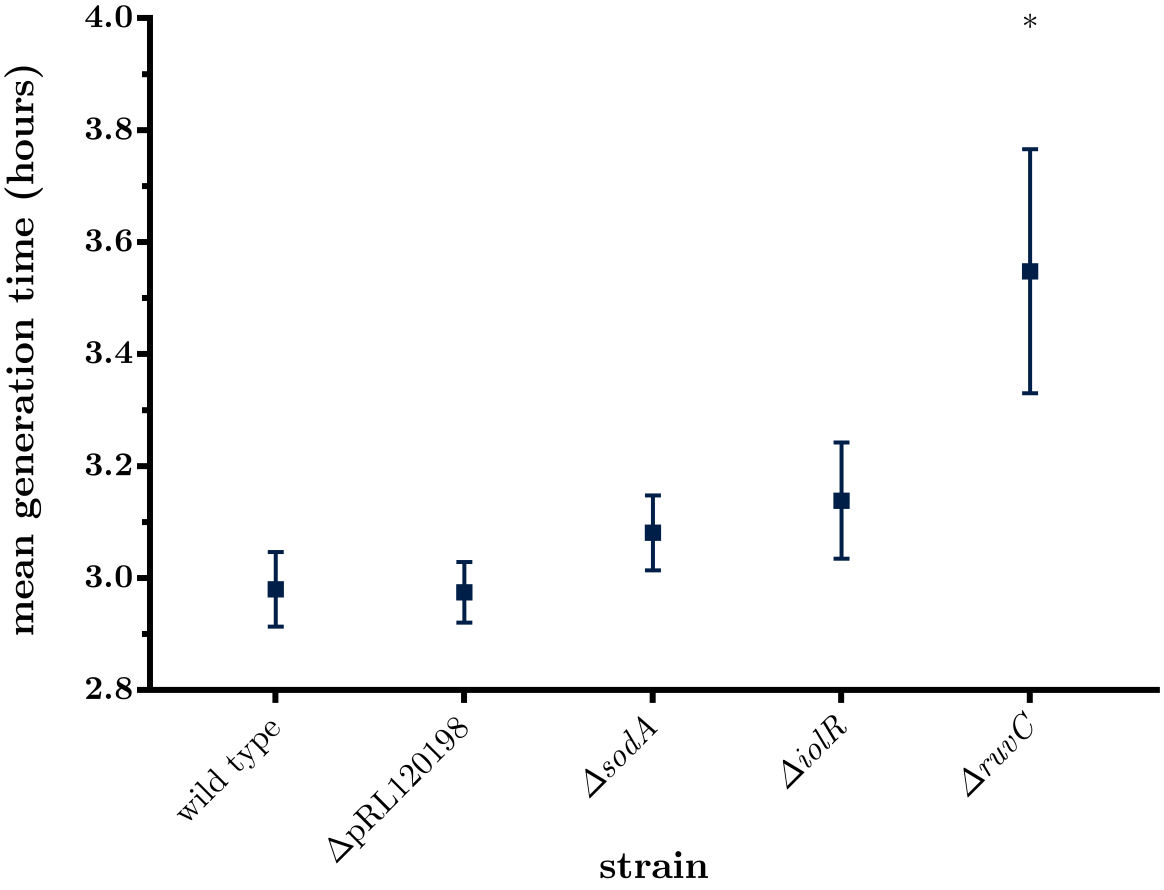


Figure 12. Growth rate in liquid TY media of wild type and four mutant strains. A plate reader was used to follow growth in liquid TY at 28°C. Error bars represent ±SEM. Statistical significance was assessed by carrying out a Dunnett's test comparing each mutant strain with wild type (3841*gusA*). An asterisk (*) indicates statistical significance where $p < 0.05$). n = 4 biological replicates.

3.9 Discussion

3.9.1 Summary

In this chapter, research was carried out with the objective of generating comparable data for selected predictions from the INSeq dataset and using this to determine which stages of symbiosis were affected for each mutant. These results were in agreement with INSeq predictions for 14 out of 15 mutants investigated, in addition to improving the definition of which stages of symbiosis were affected by these mutations (Figure 3). These results are summarised in Table 6.

The *eryB*, *cheA1*, *glnII*, *rmrR*, *purC1* and RL0149 mutants showed significantly impaired competitive nodulation ability, suggesting the importance of those genes for nodule initiation (Figure 10, Table 6). The *pssD* and *nodB* mutants were unable to nodulate at all, indicating a total loss of recognition of the rhizobia as symbiotic partners to the plant (Figure 6, Table 6). In the case of the *pssD* mutant, the mutant may be unable to survive in the infection thread, while the *nodB* mutant is likely incapable of producing a recognisable Nod factor (Geurts & Bisseling, 2002; Król et al., 1998; Venturi & Keel, 2016; Zipfel & Oldroyd, 2017). After nodulation, the *fixB* and *nifB* mutants showed no nitrogenase activity, while their nodules had visible signs of sanctioning by the plant (Figures 5 & 8, Table 6). All of this could be expected to reduce the amount of bacteroid DNA recovered from the nodules of a plant following inoculation with an INSeq library (Wheatley et al., 2020).

Strains mutated in pRL120198, *iolR*, *ruvC*, and *sodA* regrew significantly less well from nodule extracts than wild type, with the caveat that the *ruvC* mutant showed a significant

growth defect in the TY media used for the regrowth assay (Table 6). These mutations may lead to a reduced number of viable bacteria per nodule, or potentially causing metabolic changes in those bacteria to slow their transition from the intra-nodule environment to free-living growth in TY media.

Table 6

Mutant	**Nodulation**	**Nitrogenase Act.**	**Competition**	**Recovery**	**Growth**
Δ*nodB*	− −*	ND	ND	ND	ND
Δ*nifB*	+	− −	ND	ND	ND
Δ*fixB*	+	− −	ND	ND	ND
Δ*eryB*	WT	WT	− −	ND	ND
Δ*rmrR*	WT	WT	−	ND	ND
ΔRL0149	WT	WT	−	ND	ND
Δ*cheA1*	WT	WT	−	ND	ND
ΔRL1545	WT	WT	WT	ND	ND
Δ*purC1*	WT	WT	−	ND	ND
Δ*glnII*	WT	WT	−	ND	ND
Δ*pssD*	− −	− −	− −	ND	ND
Δ*pRL120198*	WT	WT	ND	−	WT
Δ*sodA*	WT	WT	ND	−	WT
Δ*iolR*	WT	WT	ND	−	WT
Δ*ruvC*	WT	WT	ND	−	−

Table 6: Summary of the mutant phenotypes observed in the experiments described in Section Chapter 3: Nodulation assay (Chapter 3.4), nitrogenase activity assay (Chapter 3.5), competitive nodulation assay (Chapter 3.6), recovery assay (Chapter 3.7), and growth assay (Chapter 3.8). 'WT' represents wild type levels; a plus sign (+) significant enhancement; a minus sign (–) significant impairment; two minus signs (– –) no measurable activity; and ND stands for 'Not Determined'. *The *nodB* mutant was unable to form nodules that did not contain revertants.

3.9.2 Future Perspectives

Further investigation of individual mutants would be specific to each mutant. For example, the expression of *glnII* is upregulated in N-limited conditions (Ramachandran et al., 2011). The growth rate of the *glnII* mutant and a wild type control could be measured on N-limiting media, such as UMS with 1 mM NH_4Cl and 20 mM succinate or with 0.5 mM NH_4Cl and 10 mM glucose, while N-excess media with NH_4Cl increased to 100 mM would serve as a control. The growth rates of the *glnII* mutant could be compared to wild type for these two conditions with the aim of finding out if the *glnII* mutant is more impaired for growth than wild type in N-limited media. A follow-on experiment could be performed to test the effect of N-excess rooting solution on the competitive nodulation ability of the *glnII* mutant. This would involve conducting nodulation competition assays with the *glnII* mutant and wild type on plants grown in N-limited and N-excess rooting solution. It is possible that the competitive-defect phenotype could be rescued by N-excess rooting solution and the competitive nodulation ability of the *glnII* mutant therefore raised to equivalency with wild type.

It is interesting that the mutation of *iolR*, a gene thought to repress the operon required for myo-inositol catabolism, has resulted in a similar competitive defect phenotype as that shown by myo-inositol catabolic mutants (Fry et al., 2007; Kohler et al., 2011). The *iolR* mutant could be further investigated by growing it on myo-inositol as a carbon source, in order to find out if this deregulation of myo-inositol catabolism has led to impaired growth on that substrate. The *iolR* mutant's ability to transport myo-inositol could be measured with ^{14}C-radiolabelled substrate as was performed for myo-inositol catabolic mutants in previous work (Fry et al., 2007).

The *sodA* and *ruvC* gene products are both thought to have roles in protection against DNA damage, which could be important for surviving oxidative bursts from the host plant during infection (Hassett et al., 1995; Iwasaki et al., 1991; Santos et al., 2001). Thus, experiments could be performed to confirm if their mutants are more susceptible to DNA damage. The mutants and wild type controls would have their growth rates measured in the presence and absence of an oxidative species such as hydrogen peroxide, and/or subjected to DNA-damaging UV treatments. It would be expected that the *sodA* and *ruvC* mutants would have their growth impaired by the oxidative/DNA-damaging conditions more than wild type. DNA damage could be measured by RT-qPCR measurement of the induction of DNA repair systems, e.g. recA, or more directly by methods such as Terminal Deoxyribonucleotide Transferase-Mediated dUTP Nick End Labelling (TUNEL), which reports DNA damage as fluorescence (Rohwer & Azam, 2000; Tapias et al., 1997).

The *cheA1* mutant has already been investigated in many further experiments, including swimming assays, in which the ability of the bacterial mutant to swim in the direction of a chemoattractant is measured (Aroney et al., 2021). Through root attachment and colonisation assays, it was found that disruption of the *che1* cluster was severely attenuating for early attachment to roots, competitive colonisation, and movement from the root surface to root hair tips during the initiation of symbiosis. However, *che1* mutants were not found to be impaired in infection thread formation nor to lead to altered nodule composition as a result, in contrast to *che2* mutants (Aroney, 2021). Follow-on experiments are suggested to compare the exudates of root hairs and with general root exudates, with the objective of identifying the chemoattractants driving *che1*-mediated chemotaxis to the root hair tips.

The *eryB* mutant was unable to form any nodules under competitive conditions, yet it was able to form similar numbers of nodules comparable to wild type when inoculated alone. Such a severe phenotype in the competition assays raises the possibility that the nodules formed during the single-strain nodulation assays were occupied by revertants. Therefore, one follow-on experiment could be to recover nodule bacteria from plants inoculated with the *eryB* mutant alone and check if these bacteria were mutants or revertants, as was performed for the *nodB* mutants (Figure 7). If revertants are present, the mutant could be re-made as a marker-replacement or in-frame deletion mutant, which would not be able to give rise to revertants in this way.

The *eryB* mutant could be further investigated by testing its growth on media containing different concentrations of erythritol, in order to test its susceptibility to erythritol toxicity. This could be carried out in comparison with other erythritol catabolism mutants such as an *eryF* mutant (Sperry & Robertson, 1975). Similarly, the pRL120198 mutant could be grown on xylulose as a carbon source to help test if the prediction that it is an L-xylulose kinase involved in xylulose catabolism.

The *pssD* mutant showed a phenotype of dry, flat colonies, suggesting a severe reduction in EPS production (Król et al., 1998). This could be investigated by isolating EPS from cultures of the *pssD* mutant and wild type using 96% ethanol (Janczarek & Skorupska, 2007). The precipitate could then be analysed for carbohydrates and the total sugars content measured by the Anthrone method (Loewus, 1952). It is expected that the *pssD* mutant would show a significantly reduced level of EPS production compared to wild type. Further characterisation of the *pssD* mutant could be accomplished through mi-

croscopic imaging experiments. The small, white, non-fixing growths it induced on the host plant (Figure 5D) were unable to be stained for GusA activity and are therefore unlikely to contain living rhizobia. However, it may be interesting to section these failed nodules and use confocal microscopy or scanning electron microscopy to investigate how their structure may differ from healthy nodules at different developmental stages.

3.9.3 Significance

The phenotypic comparison experiments described in this chapter have identified the affected stage of symbiosis for 14 of 15 mutants investigated. This worked towards the project's overall aim of investigating the basis of competitive nodulation and bacteroid development, and also highlighted the importance of competition at multiple stages of a *Rhizobium*-legume symbiosis. The data generated by these experiments have contributed to the publication of a selection of Rlv3841 INSeq data in Wheatley et al., 2020, which characterised the impact of mutations on fitness for 7,357 genes across four stages of symbiosis and identified 603 genetic regions required for competitive nodulation. It was hoped that this dataset would prove a valuable resource to inform the planning of further investigations into the *Rhizobium*-legume symbiosis, especially for identifying genes important for rhizobial competitiveness and use for inoculants. In this regard, INSeq data have already been utilised to select genes of interest as candidates for the mutant studies described in the following two chapters.

4 Characterisation of Advantaged Mutants

4.1 Introduction

One class of genes of particular interest from the INSeq dataset are the 'Advantaged' (AD) genes: genes that can be mutated to confer an apparent advantage to fitness. These genes are defined by a higher-than-expected number of reads across TA sites in an output library. For example, Figure 13 shows an overview of the region surrounding RL2036, a gene classified as AD for the nodule bacteroid dataset. The tracks show the number of reads for each TA site in the input (top), bacteroid (middle) and recovered nodule bacteria (bottom) libraries. The range for each track has been normalised to 16x the non-zero mean number of reads across all TA sites in that library. It can be seen for RL2036 that the normalised numbers of reads for most of the TA sites are higher than expected for the output libraries, the nodule bacteroid library especially.

A total of 761 genes have been classified as AD by INSeq in at least one of the conditions tested by Poole Lab members (Wheatley et al., 2020). No gene has been classified as AD in all conditions tested, and 146 of the AD genes have been classified as 'Essential' (ES) or 'Defective' (DE) for at least one other condition. It is likely that AD genes, being detrimental for one set of conditions, are beneficial for another set of conditions since natural selection would be expected to remove purely deleterious genes (Crook et al., 2012). Additionally, the INSeq experiments were carried out in specific conditions that may not reveal all effects of a mutation, such as phenotypes that would be related to competition against other soil microbe species, utilisation of (or tolerance to) compounds not present, fitness in different pHs or temperatures, or long-term survival.

There are several phenotypes which could lead to the observed increased frequency of reads. Firstly, the mutation could increase competitive nodulation ability under the conditions used for INSeq. This would lead to mutants occupying more nodules than expected, and more mutant DNA being extracted and sequenced from the pool of nodules. It is also possible that mutations increase the number of bacteroids and bacteria within individual nodules. One way this could happen is by the mutation increasing the nitrogen-fixing ability of bacteroids. This would lead to the plant providing more resources to nodules occupied by the mutant strain and growing larger in volume, with more rhizobia (Westhoek et al., 2021). It is also possible that the ratio of bacteria capable of growth outside of the nodule may be increased relative to nitrogen-fixing bacteroids, perhaps by an altered response to NCR peptides (Alunni & Gourion, 2016). A further explanation for AD phenotypes in the bacteroid dataset could be an increase in endoreduplication resulting from the mutation (Mergaert et al., 2006). This would result in more copies of the mutant genome within bacteroids and increase the representation of these TA sites in the output library without necessarily increasing the fitness of the mutant.

4.2 Selection of Advantaged Genes For Mutagenesis

This research has focused on the genes that have been classified as AD for the nodule bacteria and bacteroid datasets, since increasing the abundance of rhizobia in nodules may be useful in development of improved inoculant strains. There are 34 genes classified as AD in the nodule bacteroid set, 206 genes for the recovered nodule bacteria set, including 14 genes for both sets (Wheatley et al., 2020). However, the Hidden Markov Model does not further quantify the degree to which a gene is AD beyond assigning it that classification (DeJesus et al., 2015).

Figure 13

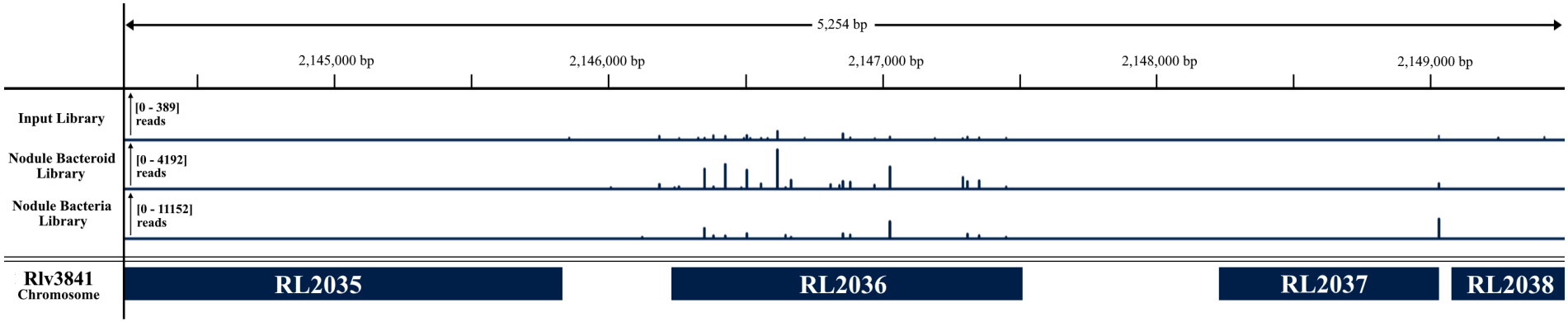


Figure 13. Screenshot of the integrated genome viewer (IGV) view of RL2036, showing the number of reads for TA sites from the input library (top track), bacteroid library (middle track) and nodule bacteria library (bottom track). The y-axis scales for each track have been normalised to 16x the non-zero mean number of reads for TA sites in that library. The peaks represent the number of reads for a TA site at that location on the genome. Many peaks for TA sites within RL2036 are visibly taller in the bacteroid and nodule bacteria libraries than in the input library.

Table 7

Gene ID	**Dsgn**	**Description**	**TA**	**Vm**	**TY**	**RS**	**RC**	**NB**	**Bd**
pRL70182		conserved hypothetical exported protein	22	NE	NE	NE	NE	NE	AD
RL0339		putative Mg^{2+} chelatase family protein	12	NE	NE	NE	NE	NE	AD
RL0748	*algA1*	Putative alpha-glucosidase	33	NE	NE	NE	NE	AD	AD
RL1130		conserved hypothetical protein	5	NE	NE	NE	NE	NE	AD*
RL2036		putative outer membrane transport protein	40	NE	NE	NE	NE	AD	NE
RL3565		Putative sensory box HD domain protein	14	NE	NE	NE	NE	NE	AD
RL3624		putative SBP of ABC transporter CUT2 GLU2 ST1	11	NE	NE	NE	NE	NE	AD*
RL4139		putative transmembrane GGDEF/EAL sensory box protein	32	NE	NE	NE	NE	AD	AD

Table 7: INSeq classifications of AD genes selected for mutagenesis and characterisation. Dsgn = Designation; TA = Number of TA sites; VM = Vincent's media (minimal, Perry et al., 2016); TY = Tryptone Yeast media (rich); RS = Rhizosphere; RC = Root Colonisation; NB = Nodule Bacteria; Bd = Bacteroid; NE = Neutral; AD = Advantaged. *RL1130 and RL3624 were initially classified as NE for the bacteroid library, but manual curation of the INSeq reads resulted in reclassification to AD. Data from Wheatley et al., 2020.

To select a small number of genes to mutate for closer investigation, the sequencing data for all potential transposon insertion sites (TA sites) were examined. The number of reads from each TA site was normalised to the total reads from the library, then TA sites were highlighted as 'Strongly Advantaged' where normalised numbers of reads in the bacteroid or nodule bacteria output libraries were 10x greater than in the input library. Eight genes with the greatest proportion of their TA sites highlighted as 'Strongly Advantaged' were selected for mutation and further investigation. This included six genes classified as AD in the bacteroid and/or nodule bacteria libraries, together with RL1130 and RL3624. RL1130 and RL3624 were initially classified with the Hidden Markov Model as NE, potentially because of a low total number of TA sites or reads. These two genes were reclassified as AD in bacteroids as part of manual curation due to the high proportion of strongly overrepresented TA sites in those genes. Such manual curation with IGV software (Figure 13) has been previously required for INSeq datasets (Wheatley et al., 2020). The INSeq classifications of these genes are described in Table 7.

4.3 Advantanged Genes of Interest

pRL70182 - Unknown

The protein encoded by pRL70182 has not been characterised previously. Peptide sequence analysis reveals an N-terminal signal sequence associated with protein export, as well as two domains of unknown function: DUF1254 and DUF1214. DUF1254 is described by InterPro as being an "immunoglobulin-like beta-sandwich domain" that is found in a few proteins of unknown function, while DUF1214 can be found in the C-terminal region of several bacterial proteins of unknown function. BLAST analysis reveals 60-72% identity to a number of proteins annotated only as 'hypothetical', 'uncharacterised' or 'DUF1254-containing' proteins, mainly from soil dwelling proteobacteria such as Burkholderiales, Pseudomonadales and Enterobacteriales. The lack of characterisation of this family of genes prevents speculation over the mechanism of the AD phenotype predicted by the INSeq analysis.

RL0339 - Mg^{2+} chelation

Sequence analysis indicates that the RL0339-encoded protein has homology to the YifB superfamily of Mg^{2+} chelatases. This family, named for *yifB* in *E. coli*, is not well studied but is predicted to have ATPase activity and similarity to the ChlI domain of other magnesium chelatases. The exact role of ChlI domains is not clear, but as part of a three-component magnesium chelatase they catalyse the insertion of Mg^{2+} into protoporphyrin IX, an essential step in bacteriochlorophyll synthesis (Walker & Willows, 1997). ChlI domains are also present in some competence proteins and proteases in bacteria and archaea. YifB-family Mg^{2+} chelatases have a structural similarity to archaeal Lon proteases that suggests that suggests a proteolytic function (Iyer et al., 2004). The amino acid se-

quence of the protein encoded by RL0339 has 40% homology to *E. coli* ATP-dependent proteases, supporting this as a potential role for the gene. However, it is not clear how knocking out RL0339 could lead to an AD phenotype.

RL0748-*aglA1* - Sugar catabolism

The *aglA1* gene sits in the *aglEFGA1K* operon, which is regulated by the LacI-family repressor *aglR*. This operon contains components of an alpha-glucoside transporter CUT1, while the *aglA1* gene is predicted to encode an alpha-glucosidase due to the presence of an alpha-amylase domain. RL3542 (*aglA2*) sits far away from this operon in the genome. AglA2 has 56% identity to AglA1 but is still predicted to have an alpha-amylase domain. The activity of AglA2 relative to AglA1 is unknown in Rlv3841. Alpha-glucosidases are important in rhizobia for the catabolism of sucrose, the most abundant source of reduced carbon in legume root nodules (Streeter, 1991; Willis & Walker, 1999).

Expression of the CUT1 transporter operon orthologous to *aglEFGA1K* in *S. meliloti* is induced by maltotriose and turanose, while microarray data shows that *aglA1* expression appears to be downregulated in the rhizosphere and nodule compared to free-living culture, suggesting that the substrate sugars of AglA1 are not present in these environments (Willis & Walker, 1999). Knocking out *aglA1* which is not needed in these conditions may lead to increased fitness and competitiveness by the lessening of a burden associated with *aglA1* expression.

RL1130 - Unknown

RL1130 has been annotated as encoding an 'uncharacterised protein' or 'conserved hypothetical protein' because very little is known about it. The protein is small, with

131 residues, and is predicted by this sequence to have a P-loop containing a nucleoside triphosphate hydrolase. The protein has some homology to kinases in other species of bacteria, but the role of RL1130 in *R. leguminosarum* is unknown, precluding speculation about why it is apparently advantageous to disrupt this gene.

RL2036 - Unknown

The gene product of RL2036 contains an N-terminal signal peptide targeting it to the outer membrane. The rest of the gene appears to encode a OMPP1/FadL/TodX-family transporter. Members of this family have been reported to transport fatty acids and other hydrocarbons across membranes (Hearn et al., 2008). It is possible that the cargo of the RL2036 protein was not present in the experimental conditions of the INSeq experiment and that expression of RL2036 was a burden on the cell. More research is needed to understand the role of RL2036 in *R. leguminosarum* and why it has been classified as AD by INSeq.

RL3565 - Unknown

The protein encoded by RL3565 is not well characterised, although peptide sequence analysis reveals that it has a Histidine-Aspartate (HD) domain indicative of a sensory box protein. Orthologues of the protein encoded by RL3565 were found across many proteobacteria species including saprophytic and plant pathogenic bacteria. Proteins in the HD domain superfamily are believed to have phosphohydrolase activity and possible roles in signal transduction or nucleic acid metabolism (Nagata et al., 2008). It is not clear how disruption of RL3565 could lead to an AD classification.

RL3624 - Sugar transport

RL3624 is thought to encode the solute-binding protein (SBP) component of the CUT2 glucose uptake transporter Sugar Transporter 1 (ST1, encoded by RL3624-26), one of three ABC transporters of Rlv3841 for the uptake of glucose and related sugars (RL1824-27 and RL4249-52, encoding CUT1 transporters ST2A and ST2B respectively). Expression of all three of these transporters is downregulated in rhizobia when in the rhizosphere relative to in rhizobia grown in media containing glucose (Ramachandran et al., 2011). The three glucose transporters were all strongly repressed by growth on either succinate or pyruvate, indicating that both these carbon sources are preferred over glucose and similar sugars (Karunakaran et al., 2009). The permease and ATP-binding components of ST1 are RL3625 and RL3626, respectively, both classified as NE across all tested conditions. No phenotype of RL3624 has been reported yet which would account for the AD classification.

RL4139 - Sensor

The gene product of RL4139 is poorly characterised but predicted to have five transmembrane helices at its N-terminus and Per-Arnt-Sim (PAS), GGDEF (named for conserved residues) and EAL (also named for conserved residues) domains towards the C-terminus. PAS domains are indicative of sensory proteins in bacteria, archaea and eukaryotes, which can sense diverse signals from bound ligands to redox potential and light (Taylor & Zhulin, 1999). GGDEF domains have been reported to have di-guanylate cyclase activity for signal transduction by that secondary messenger (Ausmees et al., 2001). EAL domains are often found with GGDEF domains as they are thought to break down cyclic di-GMP in opposition to a GGDEF domain (Simm et al., 2004).

The level of cyclic di-GMP in the cell, determined by the relative activities of GGDEF and EAL domains, has been implicated in the regulation of cell surface structures, including attachment factors (Galperin et al., 2001; Jenal, 2004). It is possible that disruption of RL4139 makes rhizobia insensitive to an unknown signal and/or changes the cell surface in some way. Increased expression of cell surface attachment factors has been previously reported to increase competitiveness for nodulation, so RL4139 mutants may have a similar phenotype (Frederix et al., 2014).

4.4 Generation of Advantaged Mutants with pK19mob

Mutagenesis using pK19mob was used successively to mutate seven out of the eight selected genes (Table 7; RL1130 failed) and these were confirmed by sequencing. RL1130 is a small gene (396 bp), making design of a region of homology long enough for successful homologous recombination difficult. While an alternative strategy could be employed to make this mutant, it was decided to move forward with plant inoculation experiments with the seven confirmed mutants.

4.5 Measurement of Host Biomass Accumulation as a Proxy for Advantaged Mutant Nitrogen Fixation Ability

An increase in nitrogen-fixing ability is one potential reason for an AD classification, as this could lead to the host plant providing more resources to mutant nodules over other nodules, resulting in a higher number of bacteroids and nodule bacteria in each nodule (Westhoek et al., 2021). In order to determine if any of the seven AD mutant strains had this phenotype, plant inoculation experiments were set up using dry weight as a proxy for nitrogen-fixing ability.

Figure 14

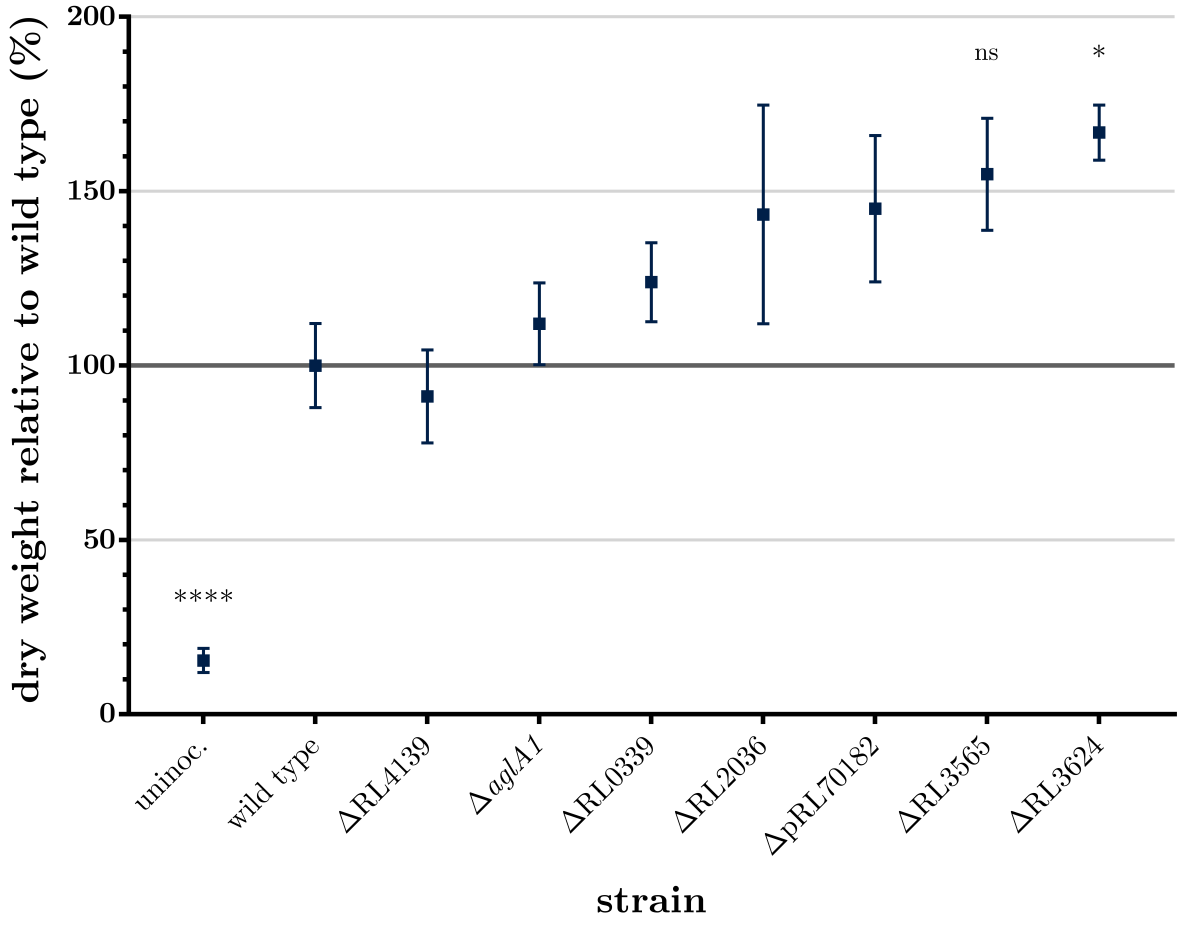


Figure 14. Mean dry weight of pea shoots following inoculation with rhizobial strains. Dry weights were converted to values relative to the mean dry weights of the wild type to allow comparison between two batches of plants (3841*gusA* dry weight was 1920 mg ±244 mg, 100%). Error bars represent ±SEM. Statistical significance was assessed by carrying out a Dunnett's test comparing each mutant strain with wild type. An asterisk (*) indicates statistical significance where $p < 0.05$, four asterisks (****) for $p < 0.0001$. n = 5 biological replicates, each consisting of a single plant.

The seven mutant strains were inoculated onto pea plants and grown for six weeks under controlled conditions. After harvesting, healthy nodules were observed on the roots of each inoculated plant. The shoots were cut from the roots, dried and weighed to determine the dry shoot biomass of each plant (Figure 14). The plants inoculated with the RL3624 mutant were the only plants that had significantly higher dry shoot biomass than plants inoculated with the wild type strain (Dunnett's multiple comparisons test).

The high degree of variation between individual plants in this type of assay means that the data from only five biological replicates were not as conclusive as was hoped. Therefore, further dry weight assays were carried out in order to obtain new data to help in the investigation of the potential enhanced plant growth promoting phenotype of the RL3624 mutant. The RL3565 mutant, which also showed possible enhanced plant growth (although not statistically significant, Figure 14) was also included in this follow-up experiment. Eleven biological replicates were set up and plants were grown for eight weeks rather than the six weeks in the previous experiment. Otherwise, the experiment was carried out as previously described.

When inoculated on pea, the RL3624 mutant showed significantly enhanced plant growth promotion compared to inoculation with wild type bacteria (143% of wild type, Figure 15), possibly due to improved nitrogen-fixing ability of the mutant. The significance of this result was greater than the previous experiment ($p < 0.01$ compared to $p < 0.05$), indicating a higher confidence. The mean shoot dry weight for plants inoculated with the RL3624 mutant was 42.7% greater than that of the wild type, while even the smallest plant was comparable to the wild type mean (99.6%, Figure 15). Plants inoculated with the RL3565 mutant were almost identical to the wild type controls (Figure 15).

Figure 15

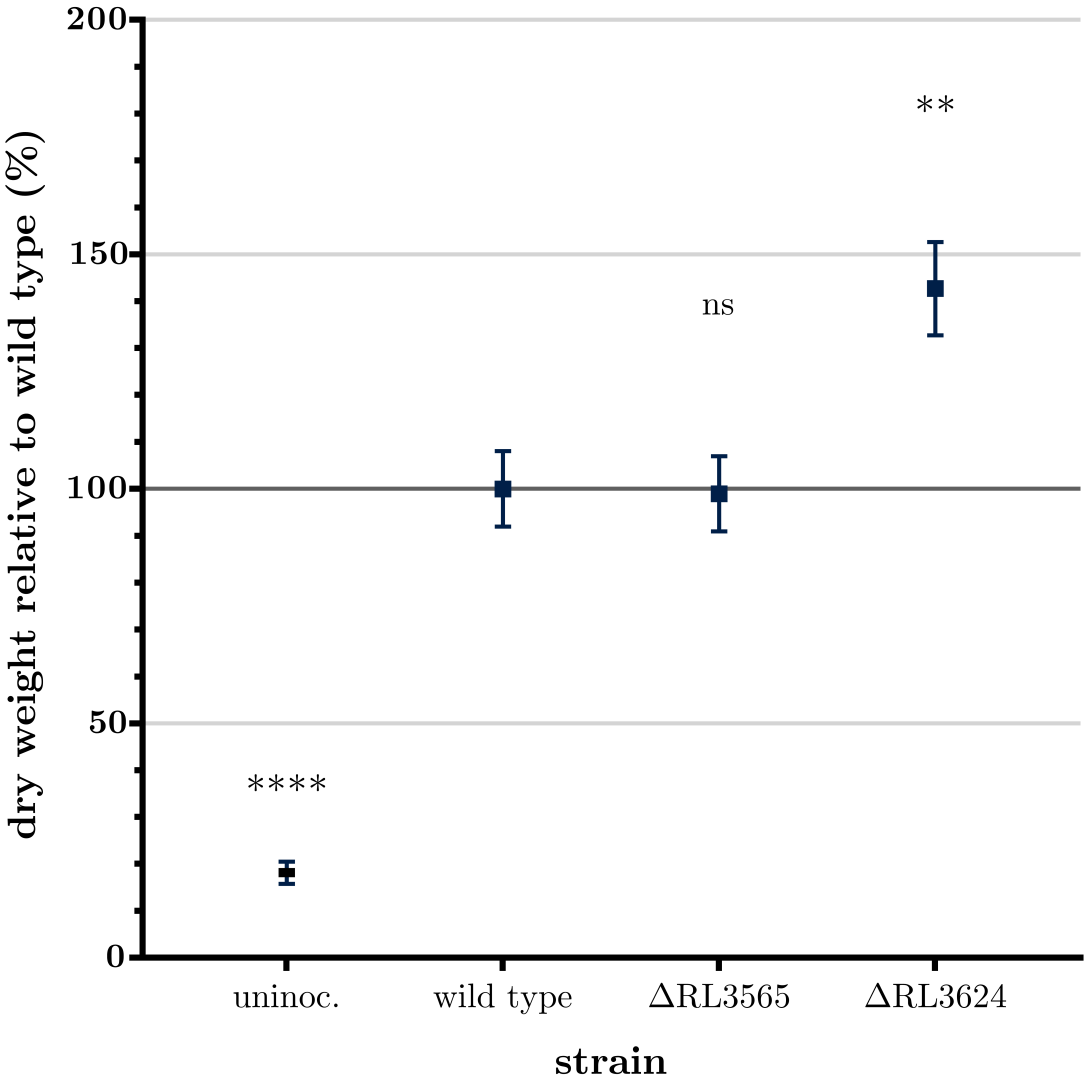


Figure 15. Mean dry weight of pea shoots following inoculation with rhizobial strains. Dry weights were converted to values relative to the mean dry weights of the wild type control set to assist comparison between experiments (3841*gusA* dry weight was 2692 mg ±216 mg, 100%). Error bars represent ±SEM. Statistical significance was assessed by carrying out a Dunnett's test comparing each mutant strain with wild type. Two asterisks (**) indicate statistical significance where $p < 0.01$, four asterisks (****) for $p < 0.0001$. n = 11 biological replicates.

4.6 Quantification of the Competitive Nodulation Abilities of Advantaged Mutants by Co-Inoculation Experiments

Competitive nodulation assays were designed with the aim to investigate another potential mechanism which could lead to AD INSeq classification: an increase in competitiveness. In the original INSeq experiment, mutants with an enhanced ability to nodule under competitive conditions could be expected to occupy a greater number of nodules than other strains, and therefore make up a larger share of reads in the output library.

These assays were carried out in the same manner as described in Chapter 3.6 to test whether the mutants were forming more nodules than wild type under competitive conditions. The results of these competition assays are shown in Figure 16, in the form of the ratio between a strain's proportion of nodule occupancy and the strain's proportion in the inoculum; this means that a strain with greater competitiveness than wild type would have a statistically significant competitive ratio above 1.0. The RL0339 and RL3565 mutants showed a small, but statistically significant increase in their share of nodules compared to their share of the inoculum (ratios of 1.18 and 1.16, respectively, p-values <0.05, mixed effects model, GraphPad PRISM). These were the only mutant strains that showed any difference in ability to compete for nodule occupancy from that of wild type (Figure 16).

The results from these competitive nodulation assays indicate that increased competitiveness for nodulation is the likely cause of the AD INSeq classifications of RL0339 and RL3565.

Figure 16

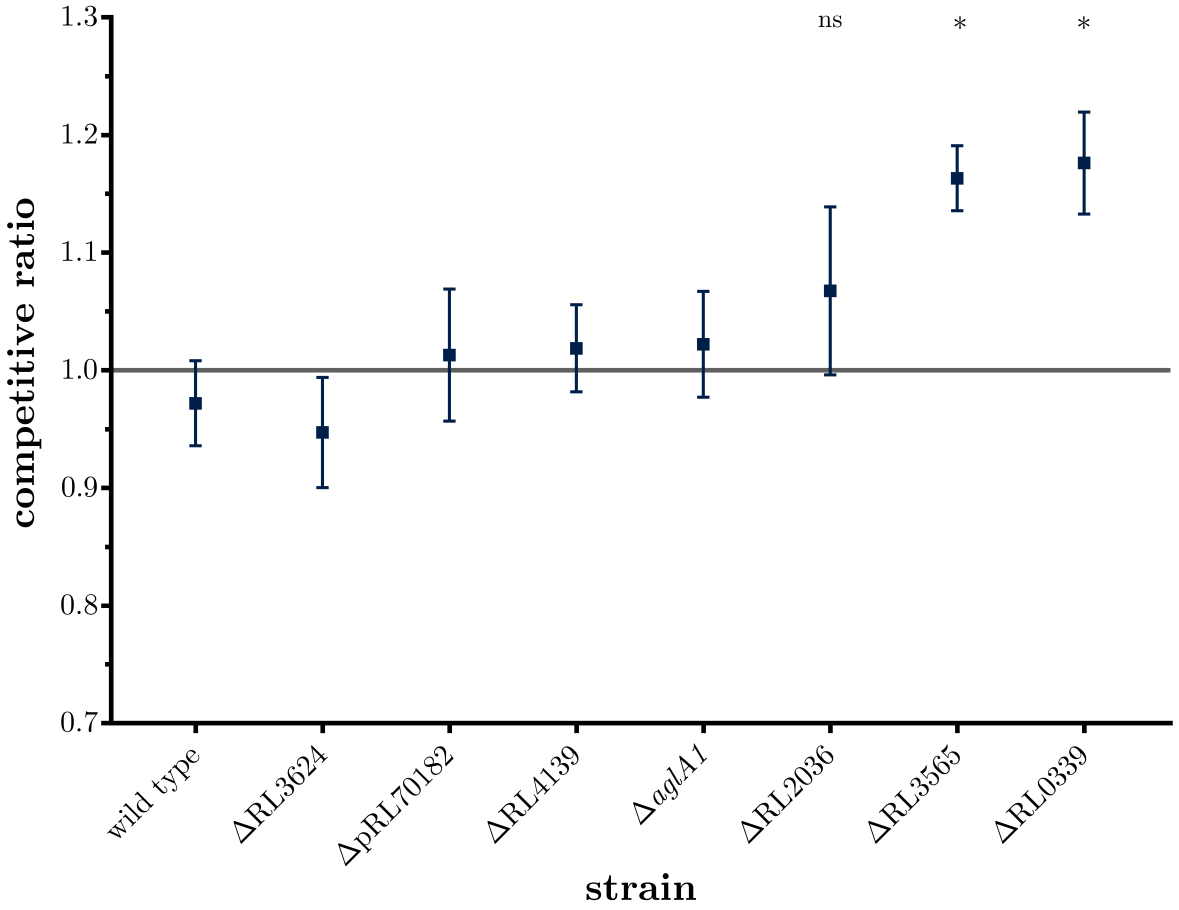


Figure 16. Ratio between the proportion of each AD strain in nodule occupancy and the proportion of that strain in the inocula in competitive nodulation assays. The competing *gusA*-marked strains were combined with 3841*celB* in a 1:1 ratio. In this measure of competitiveness, a ratio above 1.0 is more competitive than wild type. Error bars represent ±SEM. Statistical significance was assessed by carrying out a mixed model analysis between the proportion of each inoculum and its nodules as the competing strain. An asterisk (*) indicates a statistically significant (p-value < 0.05) difference between the proportion of a strain in the inoculum and the proportion of nodules of that strain from a mixed effects model. n = 6 biological replicates.

4.7 Comparison of Nodule Size as an Indicator of Host Sanctioning

The competitive nodulation assay described in Chapter 4.6 used co-inoculation of wild type and mutant strains followed by double sequential staining of the roots to allow the occupancy of each nodule to be determined. For the experiment using co-inoculation with wild type and the strain mutated in RL3624, the images captured during this competitive nodulation assay provided an opportunity to measure the area of nodules in to see if there was a difference in nodule size. A significant difference in nodule size would suggest a difference in host resource allocation to the two bacterial strains. Nodule areas were measured for each mature nodule on these plants (approximately 20-45 nodules per strain per plant). Nodule areas were sorted by size and for each strain the mean areas of the largest 5, 10 and 15 nodules are shown in Figure 17. In a previous similar study showing differences in resource allocation to strains with different nitrogen-fixing abilities, the mean area of the ten largest nodules was compared (Westhoek et al., 2021). Between wild type and the RL3624 mutant in 1:1 inoculation, there was no significant difference in the mean area of the largest 5, 10 or 15 nodules of the two strains (two-way ANOVA, p=0.8684, Figure 17).

Since inoculation of peas with the RL3624 mutant was shown to result in larger plant biomass in nitrogen-limited conditions than wild type (Chapter 4.5), a proposed explanation for its AD classification is that pea nodules occupied by the RL3624 mutant are provided with more resources from the host plant because they fix more nitrogen than other nodules. These nodules would be expected to grow visibly larger and contain more rhizobia relative to other nodules occupied by the INSeq mutant library, either due to a form of 'reverse-sanctioning' or as a result of the plant sanctioning other nodules. This hypothesis

is supported by previous observations that the sanctioning of a strain is conditional on the relative fixation ability of the strain it is co-inoculated with (Westhoek et al., 2021). However, the nodule size data (Figure 17) do not support this hypothesis because there was no observable difference in the sizes of nodules between nodules occupied by wild type and the RL3624 mutant. It may be that the difference in nitrogen-fixation ability between wild type and the RL3624 mutant is too small to result in visible sanctioning of wild type nodules. However, this does not preclude the possibility of increased bacteroid densities within RL3624 mutant-occupied nodules.

Figure 17

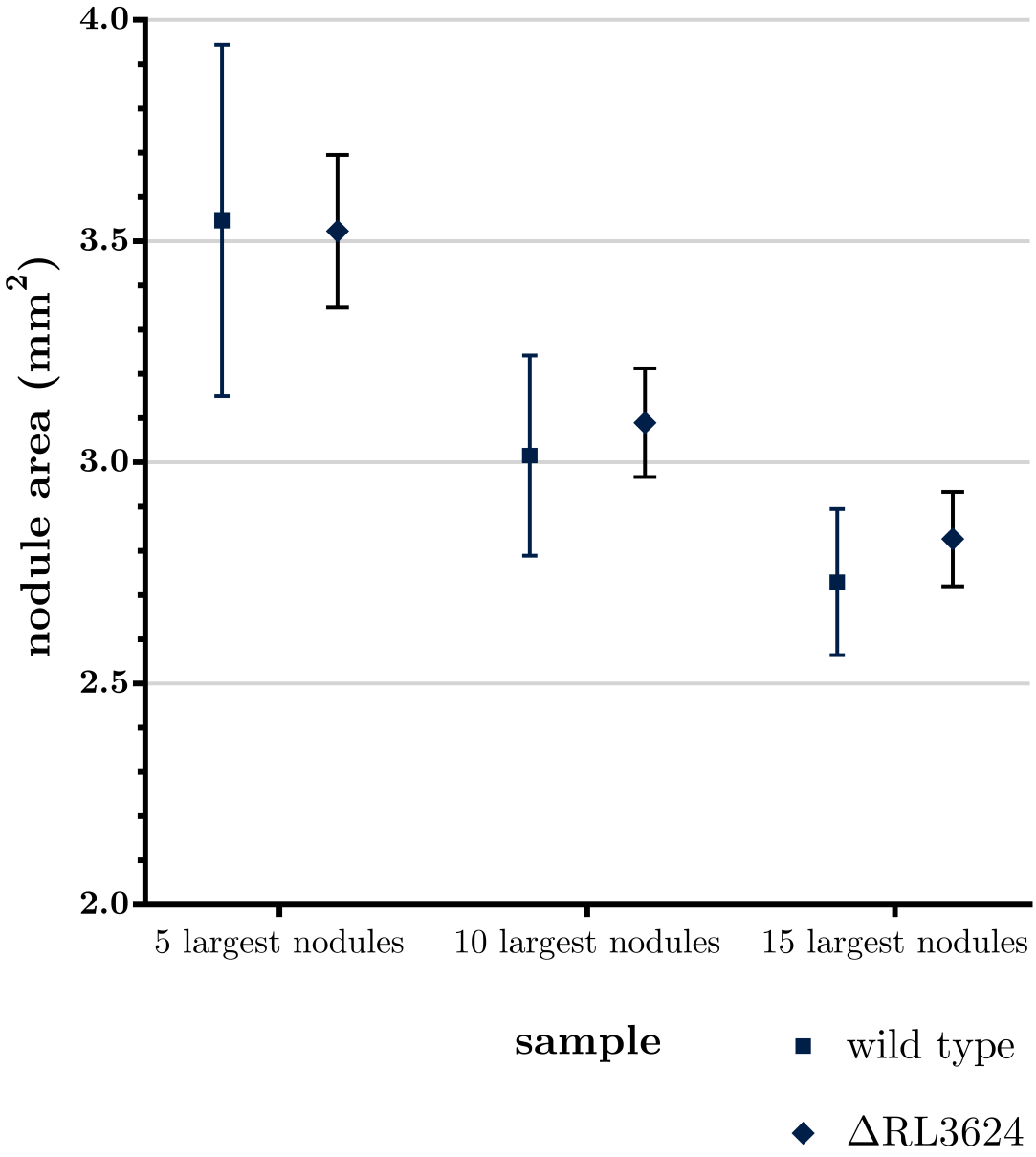


Figure 17. Mean areas of the 5, 10, and 15 largest nodules from wild type and the RL3624 mutant. Average areas were 3.55 ±0.397 mm^2, 3.02 ±0.226 mm^2, and 2.73 ±0.165 mm^2 for the 5, 10, and 15 largest wild-type-occupied nodules, respectively. A two-way ANOVA did not find the variation between the two strains to be significant ($p = 0.8684$). $n = 6$ biological replicates, each consisting of a single plant.

4.8 Comparison of Nodule Number as a Proxy for Host Nitrogen Availability

One further, unexpected result was found in the competitive nodulation assays. It was noticed that plants inoculated with a 1:1 mixture of wild type and the RL3624 mutant had visibly fewer nodules in total than plants inoculated with the wild type strain alone. The mean numbers of total nodules per plant were 191 ±10.2 for wild type alone and 146 ±12.0 for ΔRL3624 + wild type (Figure 18). A Student's t test with Welch's correction was performed, which confirmed the significance of this difference in nodule number ($p = 0.0161$).

It is proposed that this difference in the total number of nodules formed is the result of increased nitrogen-fixing ability of the RL3624 mutant providing the host plant with more nitrogen sooner, leading to the host plant suppressing the formation of new nodules because of high nitrogen availability. It would be expected that the restriction of nodule formation would be even greater for plants inoculated with a pure inoculum of the RL3624 mutant, however, this proposed experiment was not performed due to timing issues relating to Covid-19 pandemic disruption.

Figure 18

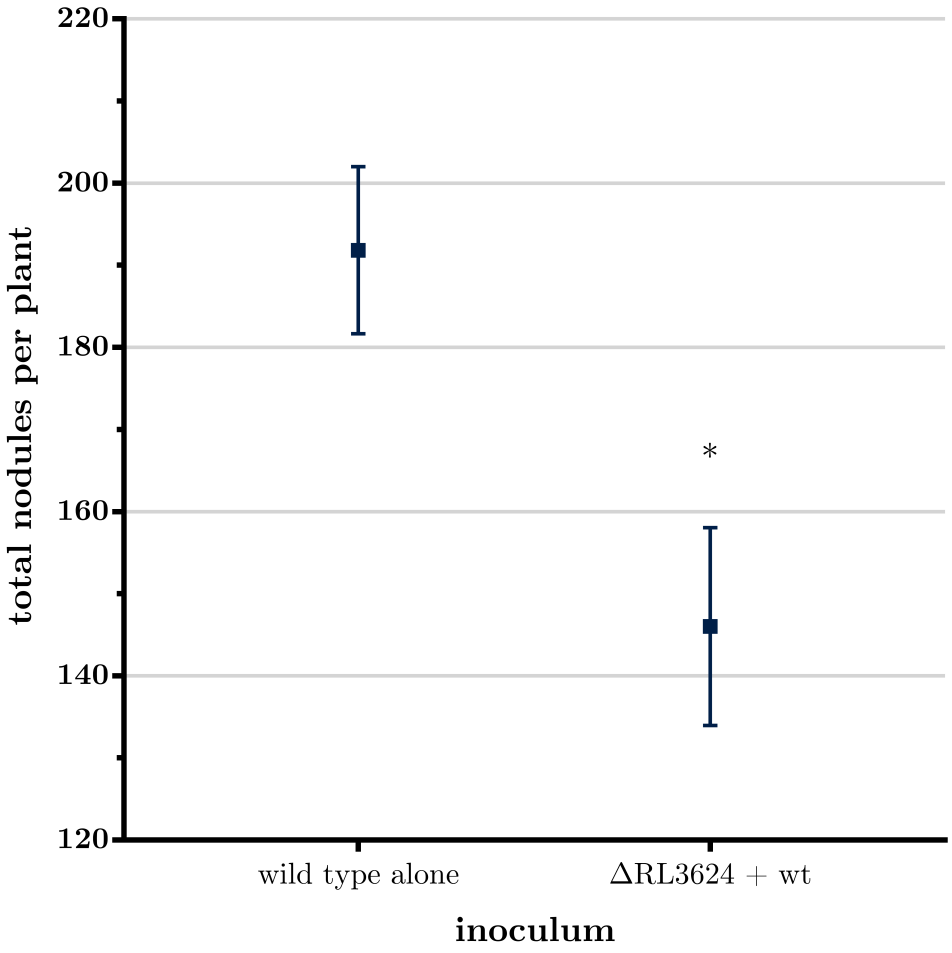


Figure 18. Total number of nodules per plant of plants inoculated with the strain mutated in RL3624 and wild type. Plants were inoculated with either a 1:1 mixture of the two wild type strains, 3841*gusA* and 3841*celB*, or a 1:1 mixture of the RL3624 mutant and the 3841*celB* wild type. Error bars represent ±SEM. Significance was assessed by carrying out a two-tailed student's t test with Welch's correction between the two groups of values, which returned $p = 0.0161$. An asterisk (*) indicates a statistically significant (p-value < 0.05). $n = 6$ biological replicates, each consisting of a single plant.

4.9 Characterisation of the Growth of an RL3624 Mutant

The RL3624 gene is thought to encode the solute-binding component of ST1, a predicted sugar transporter (Davidson et al., 2008; EnsemblBacteria). However, there exists no published data regarding which sugars ST1 may be transporting. Analysis of unpublished growth assays carried out several years ago by former Poole Lab member Ramakrishnan Karunakaran may help to resolve this.

Wild type Rlv3841 and an RL3624 mutant were grown on liquid AMS (defined) media with 10 mM ammonium chloride plus one of five sugars at 10 mM or pyruvate 30 mM as sole carbon sources. The OD_{600} of the growing cultures were followed in a plate reader, with measurements taken every 30 minutes. The RL3624 mutant used in these experiments (RU4174) is distinct from the RL3624 pK19mob mutant discussed previously as ΔRL3624 (OPS1903), and is therefore referred to as ΔST1 in this chapter. It was found that the ST1/RL3624 mutant grew considerably more slowly than wild type on glucose, mannose and xylose, with 212%, 170% and 222% higher mean generation times on those carbon sources respectively (Figure 19). However, the growth rates of this mutant on arabinose, galactose and pyruvate were roughly comparable to the growth rate of wild type. This indicates that ST1 is the primary transporter for the uptake of glucose, mannose and xylose in Rlv3841.

Figure 19

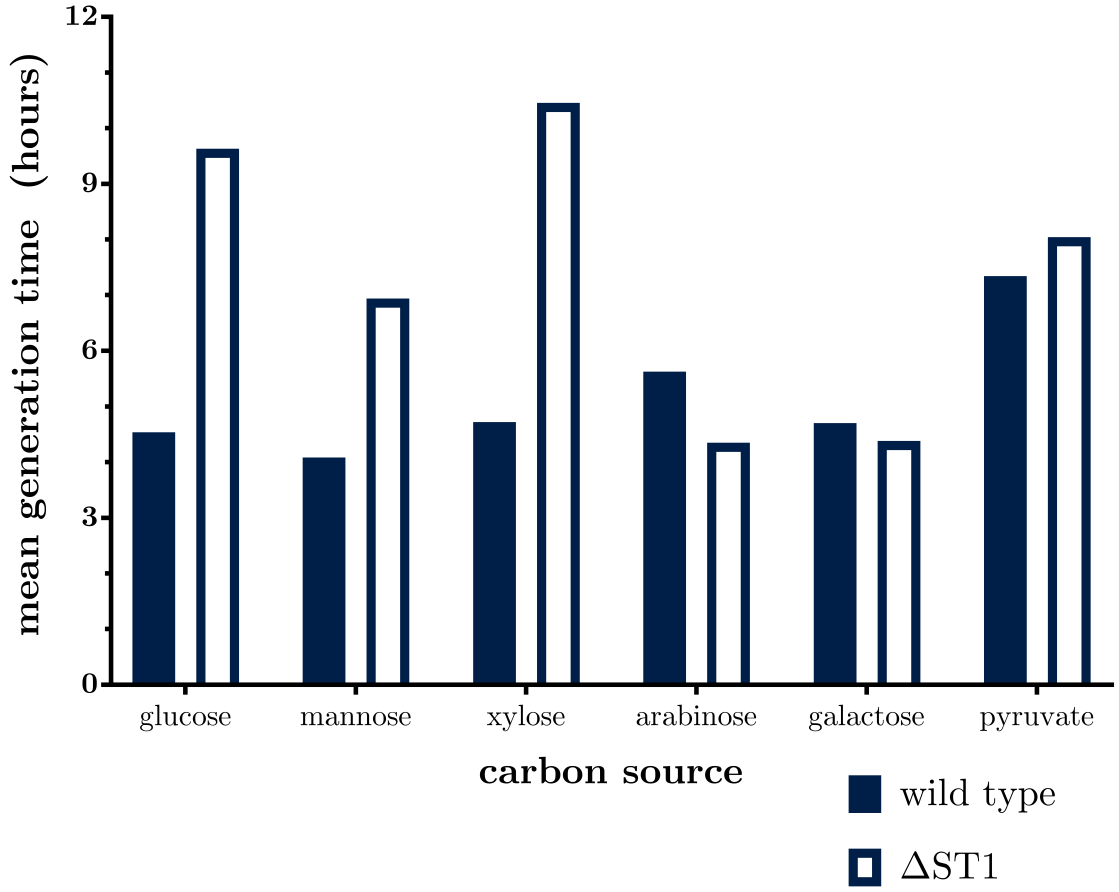


Figure 19. Comparison of growth in liquid AMS media, supplemented with one of six carbon sources of wild type and ST1 mutant (Karunakaran, unpublished). Mean generation time (hours). ΔST1 in this experiment was RU4174, not OPS1903. n = 1 biological replicate. Data from R. Karunakaran, personal communication.

Table 8 shows additional growth data giving the mean generation times for other single, double and triple sugar transporter mutants. Several key findings presented in this table are relevant to the characterisation of ST1/RL3624. Single and double mutants in the two other sugar transporter systems of Rlv3841, ST2A and ST2B, do not exhibit considerably slowed growth on glucose, mannose, or xylose compared to wild type. This further supports the hypothesis that ST1 is the primary transporter for these sugars. However, the triple mutant in ST1/ST2A/ST2B grew considerably slower than the ST1 single mutant on those three sugars, completely losing the ability to grow on xylose. Since a phenotype of this severity was not observed for either ST1/ST2A or ST1/ST2B double mutants, this suggests that both ST2A and ST2B have some ability to transport these three sugars, although this ability is not as strong as ST1 (Table 8). Furthermore, the ST1/ST2A double mutant was apparently unable to grow on arabinose as a sole carbon source, suggesting a role in arabinose uptake for both these systems.

However, the strain carrying mutations in all three transporters was able to grow on arabinose comparably to wild type, which raised the possibility that the result for the ST1/ST2A double mutant on arabinose was anomalous. The significance of these data cannot be assessed by statistical tests because the experiments involved only one biological replicate. Nevertheless, these data suggest that ST1 is the primary transporter for glucose uptake in Rlv3841 and also has a role in the uptake of xylose and mannose.

Table 8

Mutant	**Strain**	**Glc.**	**Man.**	**Xyl.**	**Ara.**	**Gal.**	**Pyr.**
wild type	Rlv3841	4.5	4.1	4.7	5.6	4.7	7.3
ΔST1	RU4174	**9.6**	**6.9**	**10.5**	4.4	4.4	8.0
ΔST2A	RU4184	4.0	4.0	4.7	5.3	4.2	7.4
ΔST2B	LMB109	4.2	4.5	4.6	4.8	4.1	7.7
ΔST1/ΔST2A	RU4223	**11.3**	**5.9**	**11.4**	**NG**	4.2	**10.5**
ΔST1/ΔST2B	LMB504	**9.0**	**7.0**	**10**	5	4.9	**7.9**
ΔST2A/ΔST2B	LMB493	4.0	4.0	5.0	5.5	5.1	**7.7**
ΔST1/ΔST2A/ΔST2B	LMB110	**16**	**11.2**	**NG**	6.4	4.6	7.0

Table 8: Comparison of growth in liquid AMS media, supplemented with one of six carbon sources of seven strains mutated in sugar transporters and wild type (Karunakaran, unpublished). Mean generation times (hours) are presented, with values more than 20% larger than wild type values in **bold**. ΔST1 in this experiment was RU4174, not OPS1903. NG stands for 'No Growth'. Carbon sources were as follows: Glc = glucose; Man = mannose; Xyl = xylose; Ara = arabinose; Gal = galactose; and Pyr = pyruvate. n = 1 biological replicate. Data from Karunakaran, personal communication.

4.10 Coexpression Analysis of RL3565 Microarray Data

The RL3565 mutant was shown to have significantly increased competitive nodulation ability compared to wild type in Chapter 4.6. RL3565 encodes an uncharacterised protein with an HD domain indicative of a sensory box protein, suggesting a role in signal transduction in an unknown regulatory pathway. If this regulatory pathway could be identified, this might aid in the generation of hypotheses for why the RL3565 mutant showed enhanced competitive nodulation ability in Chapter 4.6. Computational analysis of gene expression data could be used to search for genes that are highly coexpressed

with RL3565 across multiple conditions, which would indicate possible regulation by, or of, RL3565.

Coexpression networks were generated by Poole Lab member Javier Pardo-Díaz using signed distance correlation. Two collections of microarray data were used to construct the networks: a 'standard' collection, made up of 54 microarrays across 18 different conditions with three replicates for each condition; and a 'complete' collection, which comprises 87 microarrays across 35 different conditions with fewer replicates (Pardo-Díaz, 2021; Pardo-Díaz et al., 2021). Three different statistical methods were used to analyse the networks: Community Detection (CD), PageRank (PR), and Weight Edges (WE). Coexpression matrices were generated from the top 50 most highly coexpressed genes with RL3565 identified by at least one of the three methods of analysis (Pardo-Díaz, personal communication). The concert of these methods enhances the reliability of the analysis, as genes highlighted by all three methods are more likely to be related to RL3565.

In the standard and complete matrices, 80 and 72 genes were in the top 50 most highly coexpressed genes with RL3565 in at least one of the three methods of analysis of the standard and complete datasets, respectively (Figures 20 & 21; Pardo-Díaz, personal communication). These genes cover diverse functions and putative roles, including transporters, motility genes, conjugal transfer genes, sugar catabolism genes, and transcriptional regulators. A total of 51 genes were shared between the two matrices, while 29 and 20 were unique to the standard and complete matrix respectively. It is worth noting that plasmid-encoded genes appear to be over-represented in both the standard and complete matrices compared to the Rlv3841 genome (51.3% and 54.2% respectively, despite comprising 34.8% of the Rlv3841 genome).

Figure 20

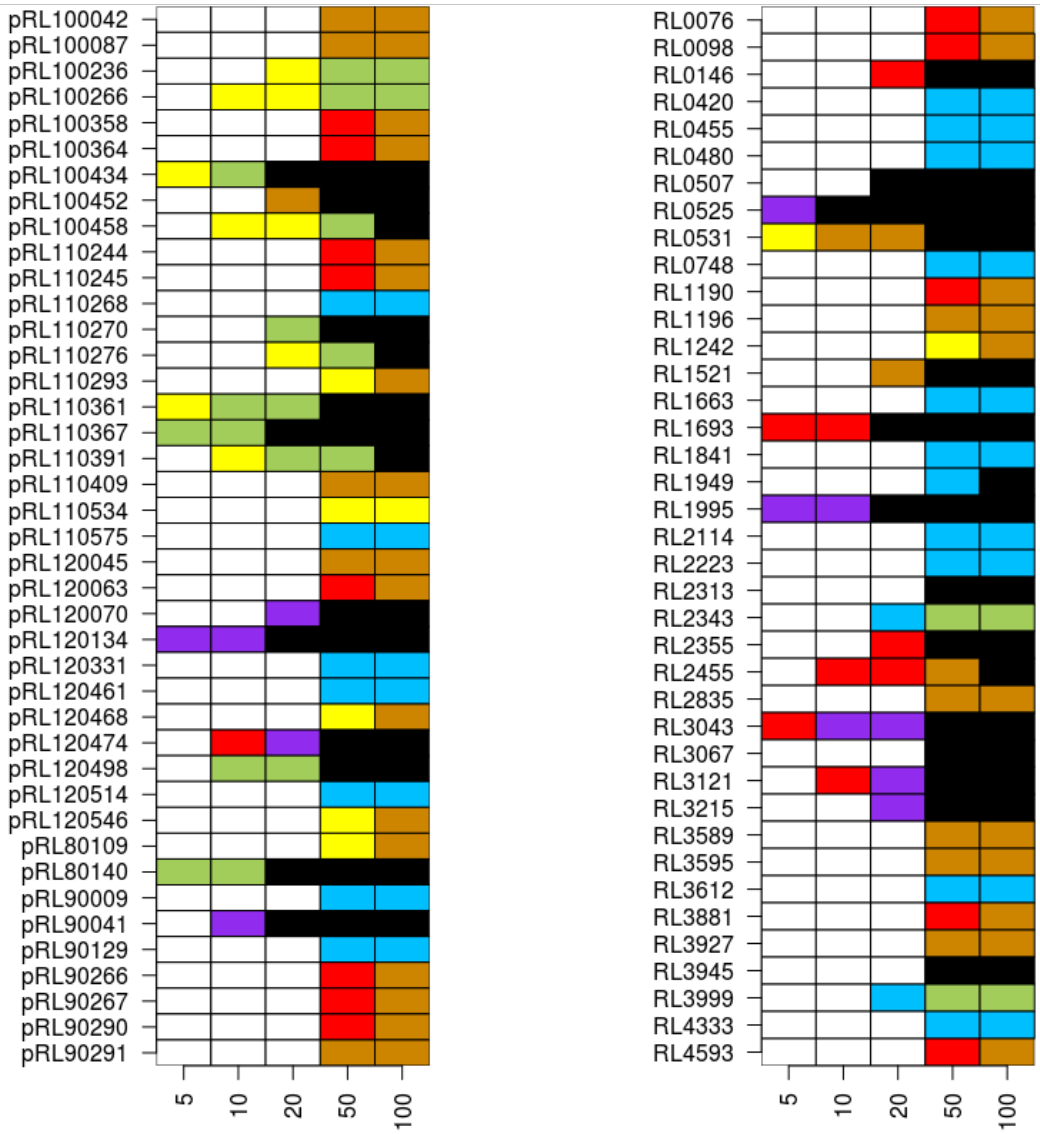


Figure 20. Coexpression matrix of the 80 genes most highly coexpressed with RL3565 in the *standard* dataset. The x axis represents whether a specific gene was in the top x of a particular analysis method (Community Detection, PageRank, and Weight Edges). The three analysis methods are denoted by colours: red = CD only; yellow = PR only; blue = WE only; orange = CD & PR; green = PR & WE; purple = CD & WE; **black = all three methods**; and white = no method. For example, in this dataset pRL100434 was detected in the top 5 genes from the PageRank analysis, in the top 10 from the Weight Edges analysis, and the top 20 from the Community Detection analysis. Figure generated by Javier Pardo-Díaz.

Figure 21

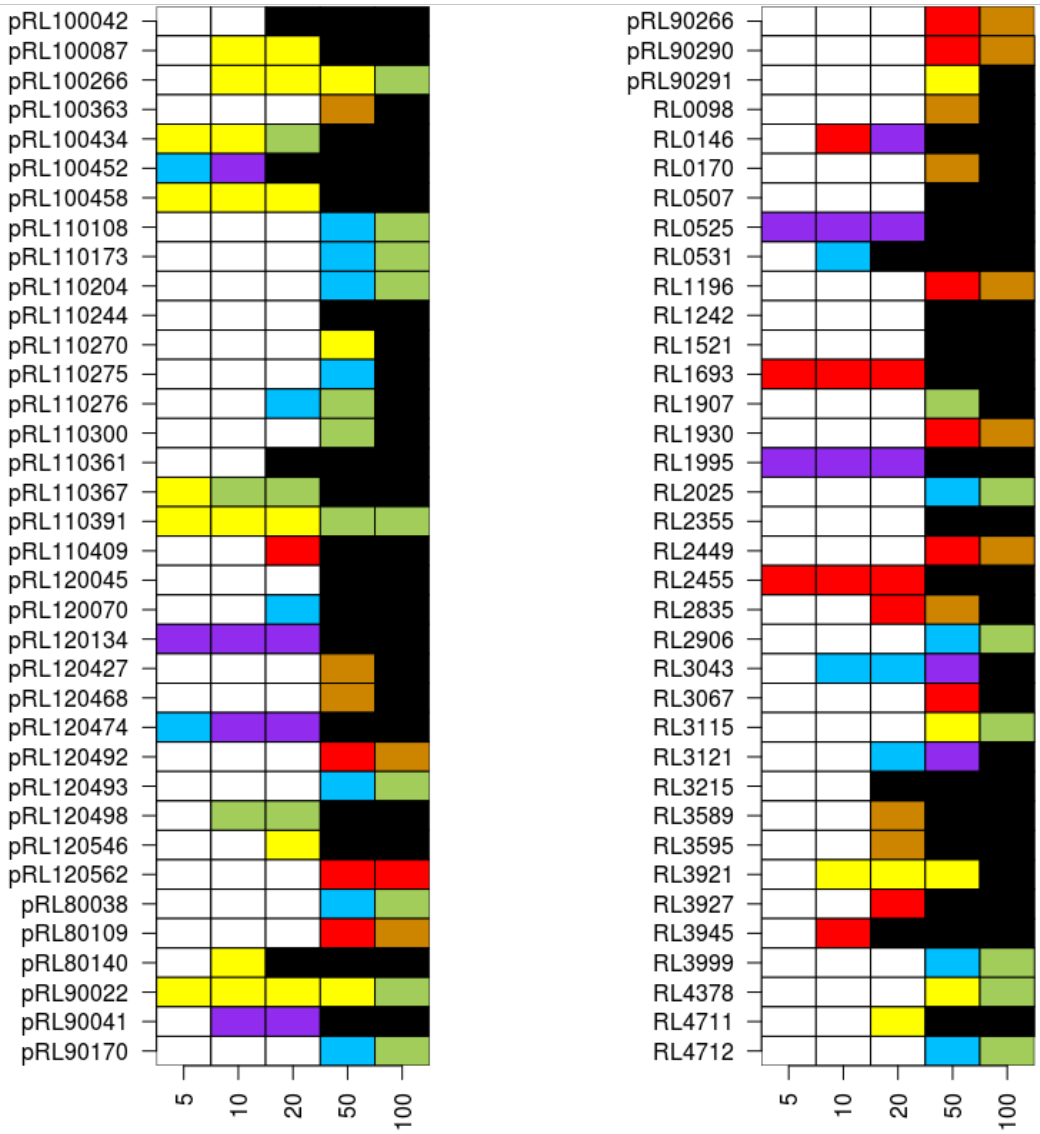


Figure 21. Coexpression matrix of the 72 genes most highly coexpressed with RL3565 in the *complete* dataset. The x axis represents whether a specific gene was in the top x of a particular analysis method (Community Detection, PageRank, and Weight Edges). The three analysis methods are denoted by colours: red = CD only; yellow = PR only; blue = WE only; orange = CD & PR; green = PR & WE; purple = CD & WE; **black = all three methods**; and white = no method. For example, in this dataset pRL100452 was detected in the top 5 genes from the Weight Edges analysis, in the top 10 from the Community Detection analysis, and the top 20 from the PageRank analysis. Figure generated by Javier Pardo-Díaz.

The most closely correlated genes were selected from the data on Figures 20 and 21 by selecting only genes that were in the top 10, 20, or 50 most correlated with RL3565 for all three statistical methods. These genes were more closely examined and their INSeq classifications were compared in Tables 9 and 10. Most genes were classified as NE in all conditions, but a few exceptions stand out.

In the complete dataset shown in Table 10, the hypothetical proteins pRL100458 and RL1242 are classified as AD in both the rhizosphere and root colonisation, while the putative LacI-family repressor RL3595 is AD for nodule bacteria. Those genes having similar AD classifications to RL3565 is evidence in support of a hypothesis that signal from RL3565 regulates the expression of those genes. However, it must be kept in mind that RL3565 was classified as AD for the nodule bacteroid INSeq library, and no gene on these two tables had that classification.

The only gene to be highlighted by this coexpression analysis and classified as AD for nodule bacteroid was *aglA1*, which was found to be in the top 50 genes most correlated to RL3565 by the Weight Edges analysis of the standard microarray dataset. Despite the correlation being unsupported by the other two analysis methods, the similarity of INSeq classifications adds to the possibility of a regulatory link between RL3565 and *aglA1*. However, the *aglA1* mutant did not show significantly improved ability for competitive nodulation in Chapter 4.6, unlike the RL3565 mutant.

Table 9

Gene ID	Dsgn	Description	Riley	TA	Vm	TY	RS	RC	NB	Bd
top 10 in all three analysis methods										
RL0525		putative MarR family transcriptional regulator	6.3.7	3	NE	NE	NE	NE	NE	NE
top 20 in all three analysis methods										
pRL100434		putative allophanate hydrolase subunit 1	3.4.2	9	NE	NE	NE	NE	ES	ES
pRL110367		putative ATP-binding component of ABC transporter CUT1	1.5.3	12	NE	NE	NE	NE	NE	NE
pRL120134	*stbC1*	putative plasmid stability protein StbC	2.2.3	4	NE	NE	NE	NE	NE	NE
pRL80140	*trbDp8*	putative conjugal transfer protein TrbD	5.1.3	7	NE	NE	NE	NE	NE	NE
RL0507	*manB*	putative mannosidase (beta mannosidase?)	2.1.3	48	NE	NE	NE	NE	NE	NE
RL1693		conserved hypothetical exported protein	0.0.2	31	NE	NE	NE	NE	NE	NE
RL1995		putative LacI family transcriptional regulator (repressor)	6.3.5	11	NE	NE	NE	NE	NE	NE
top 50 in all three analysis methods										
pRL100452		conserved hypothetical protein	0.0.2	6	NE	NE	NE	NE	NE	NE
pRL110270		putative ArdC antirestriction protein	2.2.3	13	NE	NE	NE	NE	NE	NE
pRL110361		conserved hypothetical protein	0.0.1	0	-	-	-	-	-	-
pRL120070		putative TetR family transcriptional regulator	6.3.8	10	NE	NE	NE	NE	DE	NE
pRL120474	*impB*	conserved hypothetical protein involved in nitrogen fixation, component of T6SS	3.3.22	3	NE	NE	NE	NE	NE	NE

Continued on next page

Table 9 – continued from previous page

Gene ID	Dsgn	Description	Riley	TA	Vm	TY	RS	RC	NB	Bd
pRL120498		putative large permease (DctM-like) component of TRAP transporter, protocatechuate transporter	1.5.0	31	NE	NE	NE	NE	NE	NE
RL0146		putative acetyltransferase	3.3.15	1	AD	NE	NE	NE	NE	NE
RL0531		putative transmembrane protein	4.1.1	9	NE	NE	NE	NE	NE	NE
RL1521	*sodC*	putative superoxide dismutase	1.4.2	7	NE	NE	NE	NE	NE	NE
RL2313		conserved hypothetical protein	0.0.1	9	NE	NE	NE	NE	NE	NE
RL2355		putative carboxymuconolactone decarboxylase	1.4.2	11	NE	NE	NE	NE	NE	NE
RL3043	*ctaE2*	putative cytochrome c oxidase subunit III	3.5.1	9	NE	NE	NE	NE	NE	NE
RL3067		putative two-component sensor/regulator; transcriptional regulator	6.1.2	5	NE	NE	NE	NE	NE	NE
RL3121		putative acetyltransferase	3.3.15	9	NE	NE	NE	NE	NE	NE
RL3215		putative AsnC family transcriptional regulator	6.3.1	9	NE	NE	NE	NE	NE	NE
RL3945		putative phage DNA primase/helicase	5.1.2	27	NE	NE	NE	NE	NE	NE

Table 9: INSeq classifications of genes highlighted by coexpression analysis of the *standard* dataset that were most strongly and broadly correlated with the RL3565 according to the distinct analysis methods. Dsgn = Designation; Riley = Riley code; TA = Number of TA sites; VM = Vincent's media (minimal); TY = Tryptone Yeast media (rich); RS = Rhizosphere; RC = Root Colonisation; NB = Nodule Bacteria; Bd = Bacteroid; NE = Neutral; ES = Essential; DE = Defective; AD = Advantaged (Wheatley et al., 2020). Constructed using coexpression data from J. Pardo-Díaz, personal communication.

Table 10

Gene ID	Dsgn	Description	Riley	TA	Vm	TY	RS	RC	NB	Bd
top 20 in all three analysis methods										
pRL100042		conserved hypothetical protein	0.0.1	4	NE	NE	NE	NE	NE	NE
pRL100452		conserved hypothetical protein	0.0.2	6	NE	NE	NE	NE	NE	NE
pRL110361		conserved hypothetical protein	0.0.1	0	-	-	-	-	-	-
pRL80140	*trbDp8*	putative conjugal transfer protein TrbD	5.1.3	7	NE	NE	NE	NE	NE	NE
RL0531		putative transmembrane protein	4.1.1	9	NE	NE	NE	NE	NE	NE
RL3215		putative AsnC family transcriptional regulator	6.3.1	9	NE	NE	NE	NE	NE	NE
RL3945		putative phage DNA primase/helicase	5.1.2	27	NE	NE	NE	NE	NE	NE
top 50 in all three analysis methods										
pRL100087	*acdS*	putative 1-aminocyclopropane-1-carboxylate deaminase	3.4.5	28	NE	NE	NE	NE	NE	NE
pRL100434		putative allophanate hydrolase subunit 1	3.4.2	9	NE	NE	NE	NE	ES	ES
pRL100458		hypothetical protein	0.0.0	13	NE	NE	AD	AD	NE	NE
pRL110244		putative permease component of ABC transporter PepT	1.5.0	22	NE	NE	NE	NE	NE	NE
pRL110367		putative ATP-binding component of ABC transporter CUT1	1.5.3	12	NE	NE	NE	NE	NE	NE
pRL110409	*rhaU*	putative L-rhamnose 1-epimerase involved in competition for nodulation, mutarotase	3.3.15	8	NE	NE	NE	NE	NE	NE
pRL120045		putative AsnC family transcriptional regulator	6.3.1	8	NE	NE	NE	NE	NE	NE

Continued on next page

Table 10 – continued from previous page

Gene ID	Dsgn	Description	Riley	TA	Vm	TY	RS	RC	NB	Bd
pRL120070		putative TetR family transcriptional regulator	6.3.8	10	NE	NE	NE	NE	DE	NE
pRL120134	*stbC1*	putative plasmid stability protein StbC	2.2.3	4	NE	NE	NE	NE	NE	NE
pRL120474	*impB*	conserved hypothetical protein involved in nitrogen fixation, component of T6SS	3.3.22	3	NE	NE	NE	NE	NE	NE
pRL120498		putative large permease (DctM-like) component of TRAP transporter protocatechuate transporter	1.5.0	31	NE	NE	NE	NE	NE	NE
pRL120546		putative short-chain dehydrogenase/oxidoreductase	3.3.15	9	NE	NE	NE	NE	NE	NE
pRL90041	*groLp9*	putative GroEL 60 kDa chaperonin	1.3.1	17	NE	NE	NE	NE	NE	NE
RL0146		putative acetyltransferase	3.3.15	1	AD	NE	NE	NE	NE	NE
RL0507	*manB*	putative mannosidase	2.1.3	48	NE	NE	NE	NE	NE	NE
RL0525		putative MarR family transcriptional regulator	6.3.7	3	NE	NE	NE	NE	NE	NE
RL1242		conserved hypothetical protein	0.0.2	4	NE	NE	AD	AD	NE	NE
RL1521	*sodC*	putative superoxide dismutase	1.4.2	7	NE	NE	NE	NE	NE	NE
RL1693		conserved hypothetical exported protein	0.0.2	31	NE	NE	NE	NE	NE	NE
RL1995		putative LacI family transcriptional regulator (repressor)	6.3.5	11	NE	NE	NE	NE	NE	NE
RL2355		putative carboxymuconolactone decarboxylase	1.4.2	11	NE	NE	NE	NE	NE	NE
RL2455		putative epoxide hydrolase	1.4.2	21	NE	NE	NE	NE	NE	NE

Continued on next page

Table 10 – continued from previous page

Gene ID	Dsgn	Description	Riley	TA	Vm	TY	RS	RC	NB	Bd
RL3589	*goxB*	putative glycine oxidase	3.4.2	16	NE	NE	NE	NE	NE	NE
RL3595		putative LacI family transcriptional regulator (repressor)	6.3.5	21	NE	NE	NE	NE	AD	NE
RL3927		conserved hypothetical protein	0.0.2	24	NE	NE	NE	NE	NE	NE
RL4711		putative ATP-binding component of ABC transporter Unclass	1.5.0	6	NE	NE	NE	NE	NE	NE

Table 10: INSeq classifications of genes highlighted by coexpression analysis of the *complete* dataset that were most strongly and broadly correlated with the RL3565 according to three distinct analysis methods. Dsgn = Designation; Riley = Riley code; TA = Number of TA sites; VM = Vincent's media (minimal); TY = Tryptone Yeast media (rich); RS = Rhizosphere; RC = Root Colonisation; NB = Nodule Bacteria; Bd = Bacteroid; NE = Neutral; ES = Essential; DE = Defective; AD = Advantaged (Wheatley et al., 2020). Constructed using coexpression data from J. Pardo-Díaz, personal communication.

4.11 Discussion

4.11.1 Summary

The AD mutants are some of the most interesting results to come out of INSeq dataset with regard to understanding competitive nodulation and nitrogen-fixation ability. The existence of AD mutants may be counterintuitive at first, but it must be remembered that the conditions of the INSeq experiments were very specific and not exactly the natural environment for rhizobial growth and survival. Three major factors stand out as differences between the natural symbiotic cycle of *Rhizobium* and experimental INSeq conditions: 1) The extremely short window of time required for survival in the substrate/rhizosphere; 2) The differences in chemistry between vermiculite with N-free rooting solution and real soil; and 3) The absence of other soil microbiota from the sterile environments of the INSeq experiments. Therefore, one possible explanation for why an AD gene has not been removed by natural selection is that it may be important for fitness in conditions more closely resembling the natural environment of *Rhizobium*.

With the limitations of lab conditions in mind, it is still significant that three AD mutants showed phenotypes in small scale experiments, in agreement with INSeq predictions. The RL0339 and RL3565 mutants showed small, but significant increases in competitive nodulation ability against the wild type strain. This would explain why these genes were classified as AD, since these mutants would be expected to occupy more nodules than other strains in competitive INSeq experiments, leading to higher recovery of their DNA from nodules. As described earlier in Chapter 4.3, it is not obvious why the putative HD domain sensory box protein encoded by RL3565, nor the putative Mg^{2+} chelatase protein encoded by RL0339, would have a detrimental effect on nodulation ability in these

conditions. Coexpression analysis of RL3565 has yielded several candidate genes where transcription may be regulated by signal from RL3565, including *aglA1*, but this was not conclusive enough to identify a clear signal transduction pathway. These results are summarised in Table 11.

Table 11

Mutant	**Competition**	**Dry Weight**	**Nod. Size**	**Nod. Number**
Δ*aglA1*	WT	WT	ND	ND
ΔpRL70182	WT	WT	ND	ND
ΔRL0339	+	WT	ND	ND
ΔRL1130	ND	ND	ND	ND
ΔRL2036	WT	WT	ND	ND
ΔRL3565	+	WT	ND	ND
ΔRL3624	WT	+	WT	−
ΔRL4139	WT	WT	ND	ND

Table 11: Summary of AD mutant phenotypes observed in the experiments described in Chapter 4: Competitive nodulation assay (Chapter 4.6), dry weight assay (Chapter 4.5), nodule size comparison (Chapter 4.7), and nodule number comparison (Chapter 4.8). 'WT' represents wild type levels; a plus sign (+) significantly increased levels; a minus sign (–) significantly reduced levels; and ND stands for 'Not Determined'.

As another key finding, inoculation of the RL3624 mutant resulted in significantly enhanced plant growth promotion in nitrogen-limited conditions, which could be interpreted as increased nitrogen-fixing ability relative to wild type. This provides another explanation for the increased number of reads during INSeq, as over-fixing nodules could be expected to receive more resources from the host plant. This would lead to more rhizobia

and *Rhizobium* DNA in these nodules to be recovered during the INSeq experiments, independent of an advantage in nodulation ability. The RL3624 mutant did not show an advantage in nodulation ability under competitive conditions, so the explanation for the phenotype is likely to be found within the nodule. The over-fixing hypothesis is further supported by the restriction of nodule numbers by plants inoculated with the RL3624 mutant relative to wild type, which could be expected from plants with more access to fixed nitrogen. However, comparison of the size of nodules occupied by the RL3624 mutant and wild type did not reveal significantly smaller wild type nodules, which, if observed, could have been interpreted as plant sanctioning due to the presence of a higher-fixing strain in other nodules (Westhoek et al., 2021). Nevertheless, it is possible that the individual nodules containing RL3624 mutant bacteroids are provided with more resources for longer than other nodules and this leads to more bacteroids being harvested from nodules containing the RL3624 mutant.

The RL3624 gene encodes the solute binding protein of the transporter ST1, which is responsible for glucose uptake in Rlv3841 together with ST2A and ST2B. Unpublished data show that mutation of ST1 alone slows the growth of Rlv3841 to less than half the rate of wild type on glucose as a sole carbon source, indicating that ST1 is the primary transporter for glucose uptake in Rlv3841 (Karunakaran, personal communication). The same data also suggested that ST1 is important for the uptake of xylose and mannose. While glucose is present in pea root exudate, especially during the first ten days of growth, bacteroid metabolism is based on receiving dicarboxylates from the plant (Ledermann et al., 2021; Rovira, 1956).

It is possible that, by knocking out ST1 and restricting the uptake of glucose and related sugars, the metabolism of the RL3624 mutant strain shifts away from the utilisation of these sugars and towards other solutes (e.g. dicarboxylates) at an earlier stage of nodule development. This could accelerate the rate of bacteroid formation and provide the plant with more nitrogen-fixing bacteroids earlier in nodule development, or it may alter the ratio of nitrogen-fixing bacteroids to non-fixing nodule bacteria to result in a more efficient nodule. Alternatively, an earlier and/or greater shift to dicarboxylate metabolism may be increasing bacteroid ammonia secretion through heightened demand for oxygen and more severe restriction of bacteroid glutamate biosynthesis. This would fit with recent metabolic models, which predict that bacteroid ammonia secretion is promoted by the provision of dicarboxylates and tight control of oxygen supply by the host legume to this effect (Schulte et al., 2021). These explanations could be investigated by metabolic modelling and future experiments to assay the abundance of nodule bacteria and bacteroids over time during nodule development.

4.11.2 Future Perspectives

The investigation of the absolute and relative quantities of bacteria and bacteroids in nodules occupied by the RL3624 mutant would be one potential next step for the further characterisation of this mutant. This could be carried out by performing nodule extractions as described in Chapter 2.4, followed by use of Percoll or Nycodenz density gradients to separate the nodule bacteria, bacteroids, and plant debris from the nodule extract mixture (Higashi et al., 1984). If the rhizobia were labelled with fluorescent markers, perhaps driven by bacteria or bacteroid-specific promoters (e.g. P*rhiA* or P*nifH*, respectively), this would aid in the quantification of nodule bacteria and bacteroids by flow cytometry (Economou et al., 1989; Mendoza-Suárez et al., 2020).

One of the major changes involved in the transition from bacterium to bacteroid is the endoreduplication of the entire *Rhizobium* genome, resulting in bacteroids possessing up to 64 or 128 copies of the haploid genome (Mergaert et al., 2006). Since INSeq classifications are derived from the number of sequencing reads from a given sample, any mutation leading to increased endoreduplication in bacteroids could significantly increase the number of reads and result in an AD classification from INSeq, without necessarily increasing rhizobial fitness. Measuring the number of genome copies per bacteroid would be a pertinent way of investigating the four AD mutants which did not show increased competitive nodulation or nitrogen-fixing ability, in addition to any further AD mutants generated. This could be achieved by isolating bacteroids through methods described previously, then treating bacteroid samples with a fluorescent, DNA-binding stain such as DAPI (Kapuscinski, 1995). Stained samples could be analysed by flow cytometry to compare the amount of DNA in wild type and mutant bacteroids, similar to methods that have previously been employed to characterise plant mutants deficient in endoreduplication (Cebolla et al., 1999).

Coexpression analysis of RL3565 has yielded several candidate genes that may be regulated by signal from RL3565, including *aglA1*, but this was not conclusive enough to identify a clear signal transduction pathway. Follow-up experiments could involve measuring how knock-out and/or overexpression mutations of RL3565 affect the level of transcription of *aglA1* and other genes of interest from the coexpression analyses. This gene expression analysis could be achieved through RT-qPCR, comparing cDNA transcript levels between wild type and the mutant(s) after growth in different conditions, e.g. rhizosphere growth or root colonisation. Further investigations could involve perform-

ing genome-wide transcriptomics experiments, such as microarrays and RNAseq, aiming to investigate possible interactions between RL3565 and genes outside of the coexpression analyses. These experiments would augment the coexpression analyses by also being able to identify genes that are negatively correlated with RL3565, i.e. as a possible result of repressor activity in the RL3565 signal pathway.

4.11.3 Significance

It is worth noting that the way that commercial inoculants are used is also different from the natural symbiotic cycle of *Rhizobium* (Lupwayi et al., 2006; Poole et al., 2018). While inoculant strains must still contend with the microbiota and chemistry of real soil, factors such as long-term survival in soil are less important for a commercial inoculant strain (Deaker et al., 2004). On the other hand, qualities such as the ability to reach high cell densities in liquid culture and resistance to desiccation become critical to the effectiveness of inoculants applied to legume seeds (Casteriano et al., 2013; Deaker et al., 2007). It is possible that inoculant manufacturers may be able screen candidate strains in the future for the absence of genes detrimental to these conditions, competitive nodulation ability, or nitrogen-fixing ability, thereby improving the efficiency at which the most effective inoculant strains can be identified from natural isolates. Eventually, it may be possible to improve elite inoculant strains further by knocking out such detrimental genes directly. It is hoped that this investigation into AD genes could assist with both possibilities.

5 Mutant Screen to Investigate Nodule Developmental Kinetics

5.1 Introduction

The aim of the work in this chapter was to identify mutants delayed in bacteroid development. Identification of a mutant with a delayed bacteroid development phenotype would be very useful for elucidating which genes control the timing of nodule bacterial differentiation into bacteroids and advancing fundamental knowledge of bacteroid development. However, are no appropriate molecular tools to follow bacteroid development over time, other than the labour-intensive harvesting of nodules at different timepoints and extracting bacteroids for analysis. The development of promoter fusion biosensors has enabled the possibility of using such tools to track the development of nodules through key checkpoints *in vivo* (Mendoza-Suárez et al., 2020; Pini et al., 2017). In order to assess mutants for delays to nodule development, it was necessary to design and construct two biosensors to monitor key stages of symbiosis.

It is plausible that some of the ES and DE phenotypes in the nodule bacteroid INSeq dataset may be due to mutations delaying bacteroid development and, by extension, nodule development. Nodules in which the start of nitrogen fixation is delayed could be affected through the sanctioning of carbon supplied by the plant and hence contain a lower number of bacteroids, reducing the amount of nodule bacteroid DNA recovered in the INSeq screen (Kiers & Denison, 2008; Westhoek et al., 2017; Wheatley et al., 2020). Therefore the INSeq dataset was used as a starting point for the selection of candidate genes to mutate and screen for delays to nodule development.

5.2 Biosensors to Monitor Key Stages in Symbiosis

To determine when an individual nodule has started to fix nitrogen, a super-folder GFP (*sfgfp*; Pédelacq et al., 2006) reporter driven by the synthetic promoter sequence of *nifH* (Ps*nifH*) was chosen (Mendoza-Suárez et al., 2020).. The synthetic *nifH* promoter was developed from the native rhizobial *nifH* promoter region, a type -24/-12 promoter recognised by RNA polymerase sigma factor RpoN (σ54; Thöny and Hennecke, 1989), which has been successfully used for decades to study *Rhizobium* nitrogen fixation and competition (Denton et al., 2003; Sessitsch et al., 1996; Wilson, 1995). A synthetic consensus sequence for P*nifH* was identified and cloned as Ps*nifH* to allow its use as a reporter across multiple rhizobial strains (Mendoza-Suárez et al., 2020). This synthetic promoter sequence was used to develop a reporter system (Ps*nifH*:*sfgfp*) to assess the nitrogen-fixing ability of a large number of strains in high-throughput experiments (Figure 22), as expression driven by Ps*nifH* was shown to be a good proxy for nitrogen fixation (Mendoza-Suárez et al., 2020). In that study, the Ps*nifH*:*sfgfp* reporter was on a multicopy plasmid, but for the purposes of this experiment it was judged preferable to have this reporter system as a single copy integrated into the Rlv3841 chromosome (Figure 23). This would eliminate the effect of copy number variation on the signal of the reporter, whilst also reducing the likelihood of interference with the mutagenesis system or other fitness costs.

Figure 22

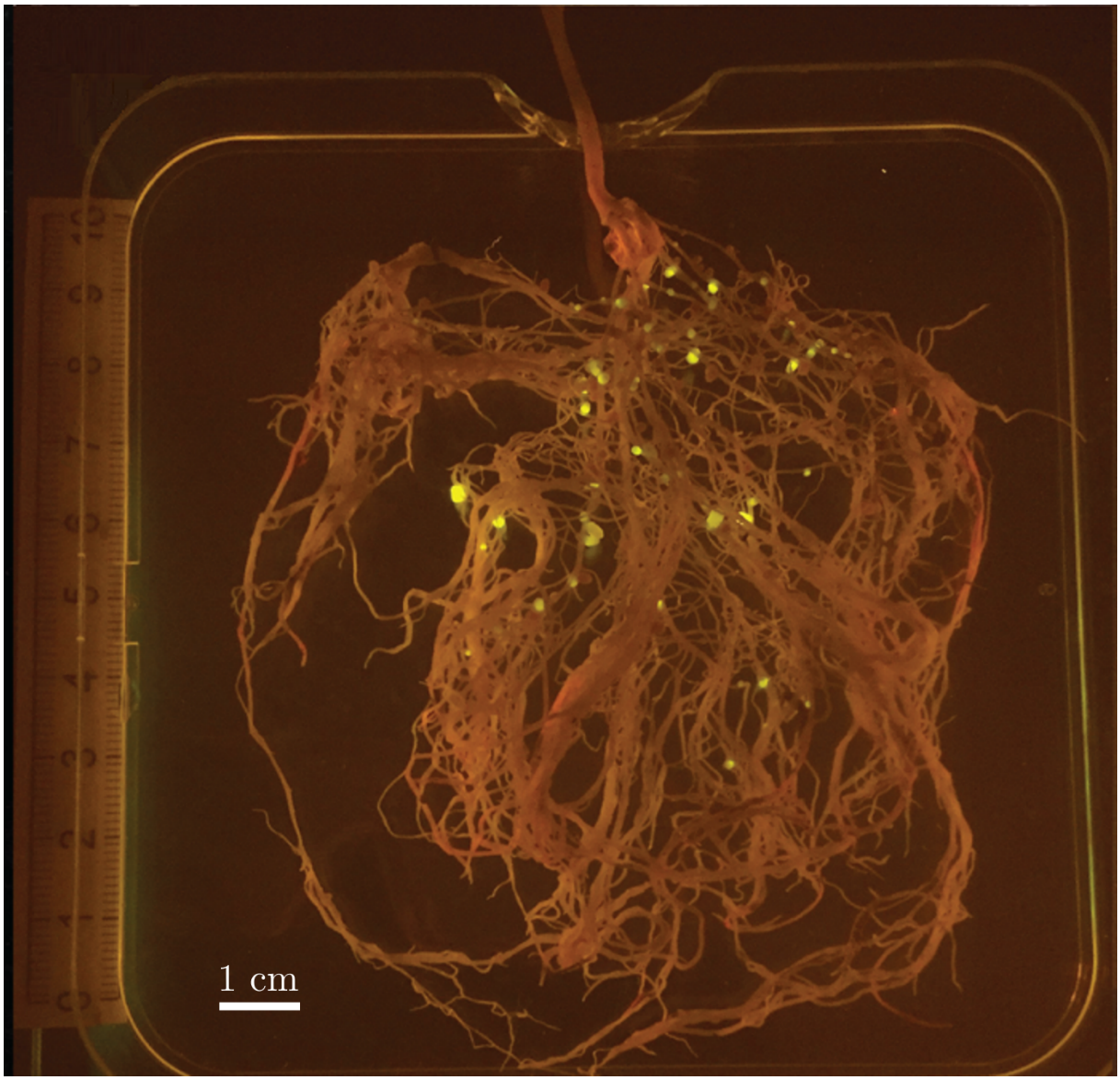


Figure 22. Pea roots with nodules occupied by rhizobia carrying Ps*nifH*:*sfgfp* reporter on a multicopy plasmid. Photo taken through an orange filter with roots spread on a square plate above a transilluminator. The scale bar represents 1 cm. Figure taken from Mendoza-Suárez, 2018.

Figure 23

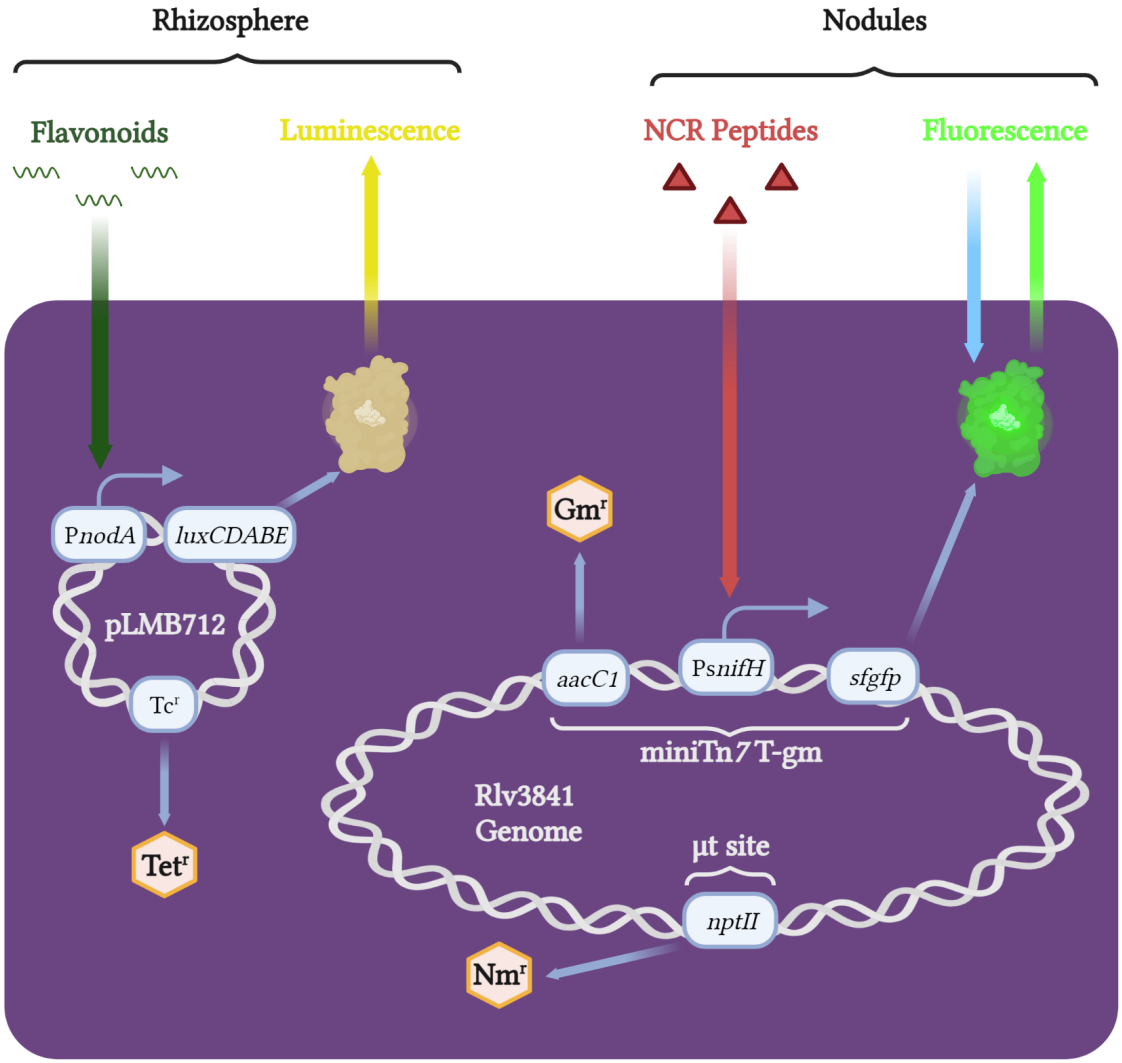


Figure 23. Overview of the dual-biosensor system for monitoring timing of nodule development. On the pLMB712 plasmid, the promoter sequence of *nodA* drives expression of *luxCDABE* upon detection of flavonoids at the start of symbiosis initation. At the mini-Tn*7* insertion site in the Rlv3841 chromosome, the synthetic *nifH* promoter sequence drives *sfgfp* expression during bacteroid development, which is driven by NCR peptides. The marker replacement at the site of mutation (µt site) confers resistance to neomycin, while resistance to tetracycline and gentamycin are conferred by genes on the pLMB712 plasmid and the mini-Tn*7* insertion respectively. Created with BioRender (biorender.com).

A vector for the chromosomal integration of the Ps*nifH*:*sfgfp* reporter was designed to use the mini-Tn*7* (Choi & Schweizer, 2006) system for insertion into the single Tn*7* attachment site in the Rlv3841 chromosome (Figure 23). A second construct, carrying *mCherry* instead of *sfgp*, was also designed to allow for the possibility of future experiments using two strains in competition with each other on the same plant. Both of these reporter systems were constructed using level one Golden Gate Cloning in one-tube one-step reactions with T4 DNA ligase and the restriction enzyme *Bsa*I (Engler et al., 2008; Singer et al., 2010). A pUC18-derived plasmid carrying the miniTn*7*T system with a gentamycin resistance marker was used as the level one vector. This was assembled with level zero parts carrying Ps*nifH* with a ribosomal binding site as the PU module (Mendoza-Suárez et al., 2020), either *sfgfp* or *mCherry* as an SC module (Weber et al., 2011), and the DT-16 terminator as the T module (Table 12; Geddes et al., 2019). The constructed mini-Tn*7* vectors were confirmed by sequencing then conjugated into Rlv3841, where the mini-Tn*7* vectors integrated into the chromosome at the single Tn*7* attachment site. This resulted in the Rlv3841-derived strains OPS2575 (Ps*nifH*:*sfgfp*) and OPS2576 (Ps*nifH*:*mCherry*). OPS2575 was used as the background strain for all mutagenesis in this chapter.

The second reporter system was set up with the intent to track the initiation of nodulation. The pLMB712 plasmid has the *luxCDABE* (Fan et al., 2008) operon driven by the promoter sequence of *nodA* (P*nodA*)::*lux*; Pini et al., 2017). The *nodA* promoter switches on upon perception of host plant flavonoids, driving expression of the genes necessary for the production of Nod factors (Geurts & Bisseling, 2002). This is the start of the chemical conversation between *Rhizobium* and host legume that precedes nodulation, a key checkpoint that was used in this experiment to mark the beginning of a nodule's de-

velopment (Figure 23). The pLMB712 (P*nodA*:*lux*) biosensor plasmid potentially allows the determination of this timepoint for each nodule via a spike in luminescence localised to the site of that future nodule, as depicted for the vetch root in Figure 24 (Pini et al., 2017). The pLMB712 plasmid was conjugated into the OPS2575 wild type control strain and all mutants generated in this chapter using a standard protocol (Chapter 2.3).

Figure 24

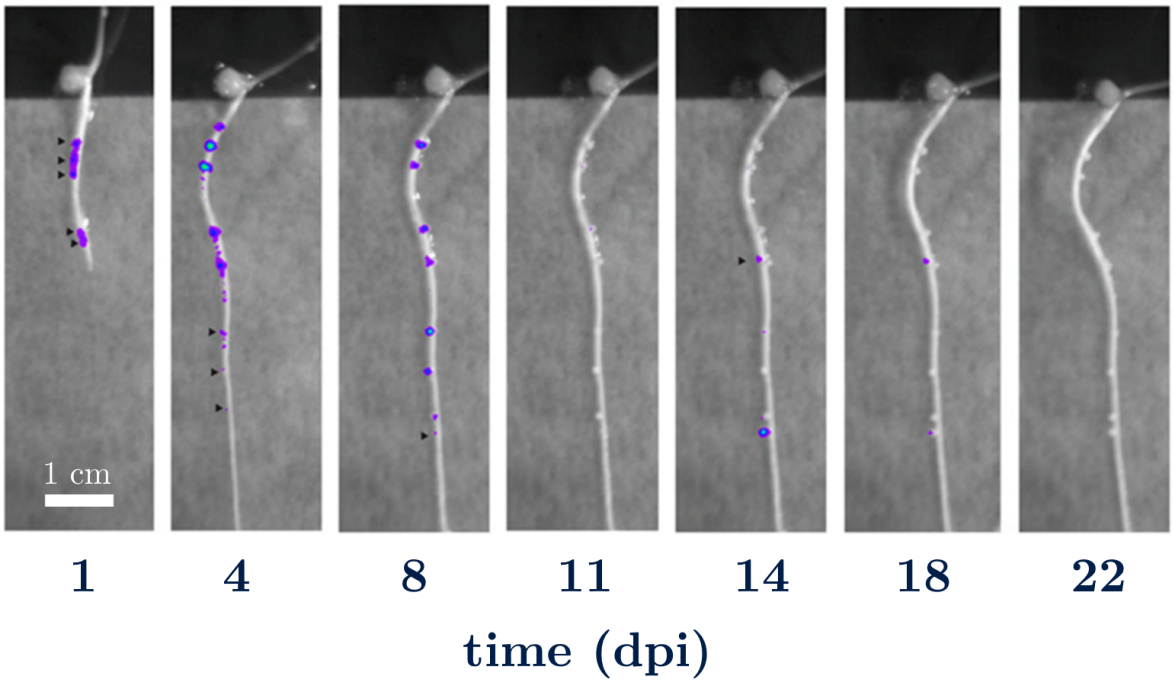


Figure 24. Vetch (*Vicia hirsuta*) seedlings inoculated with Rlv3841 carrying pLMB712. In this biosensor plasmid, the promoter sequence of *nodA* drives expression of *luxCDABE*, resulting in luminescence detected by the NightOWL camera. Detection of plant-exuded flavonoids by *Rhizobium* induces *nodA* promoter activity, leading to luminescence peaks on the vetch root (indicated by black arrowheads) in the days prior to the formation of a visible nodule. Figure adapted from Pini et al., 2017.

Table 12

Plasmid	**Module**	**Description**	**Source**
pOGG037	SC	pL0M-SC-*sfgfp*	Weber et al., 2011
pOGG043	PU	pL0M-PU-Ps*nifH* (synthetic P*nifH*)	Mendoza-Suárez et al., 2020
pOGG063	SC	pL0M-SC-*mCherry*	Weber et al., 2011
pOGG157	T	pL0M-T-DT16	Geddes et al., 2019
pOGG276	Level 1 Vector	pL1V-Lv1-pUC18R6K-miniTn*7*T-Gm	Beatriz Jorrín, Poole Lab

Table 12: Golden Gate parts used for the construction of background strains OPS2575 and OPS2576. Golden Gate parts used for the construction of background strains OPS2575 and OPS2576. OPS2575 is Rlv3841 with the Ps*nifH*:*sfgfp* biosensor integrated on the chromosome using the mini-Tn*7* vector (Choi & Schweizer, 2006), while OPS2576 was constructed with the *mCherry*-carrying SC module in place of the *sfgfp* SC module. GFP-labelled OPS2575 is the background strain for all mutagenesis described in Chapter 5. Level 0 modules for Golden Gate Cloning were as follows: PU = promoter with ribosomal binding site; SC = open reading frame not intended for tagging; and T = terminator.

5.3 Selection of Candidate Genes for Mutagenesis

Candidate genes were selected for mutation and screening with the objective of identifying a mutant delayed in nodule development. These genes were chosen in accordance with the assumption that mutants delayed in the start of nitrogen fixation would be sanctioned and therefore be classified as ES or DE by INSeq. A total of 368 genes were classified this way in nodule bacteroid dataset. This set were reduced to 165 by excluding the genes which were affected for growth on other media and stages of symbiosis (did not show an NE phenotype across TY liquid media, Vincent's liquid media, pea rhizosphere, pea root attachment, and the input libraries of the experimental sets). The set were further reduced to 60 (Table 13) by excluding the genes with previously reported phenotypes (e.g. genes involved in nitrogen fixation, chemotaxis and motility), mutants already investigated by the Poole Lab and the subsequent genes in the same operon of a gene already on this list. These 60 genes, along with their INSeq classifications, are listed in Table 13.

Table 13

Gene ID	Dsgn	Description	NB	Bd	µA	µC
RL0119		putative ribosomal RNA small subunit methyltransferase	NE	ES	✓	
RL0123	*truB*	putative tRNA pseudouridine synthase B	DE	ES	✓	
RL0143		putative PfkB family sugar kinase	ES	DE	✓	✓
RL0381	*hrcA*	putative heat-inducible transcription repressor	DE	DE	✓	✓
RL0398		putative acetyltransferase	DE	DE	✓	
RL0493	*pyrC*	putative dihydroorotase	ES	DE	✓	
RL0501		putative orotate phosphoribosyltransferase	NE	DE	✓	✓
RL0502	*frk*	putative fructokinase	ES	DE	✓	
RL0546	*phoU*	putative phosphate uptake regulator PhoU	ES	ES	✓	
RL0572		putative dihydroorotate dehydrogenase	ES	ES	✓	✓
RL0652		putative LacI family transcriptional regulator (repressor)	DE	DE	✓	
RL0726		conserved hypothetical exported protein	DE	DE	✓	✓
RL0727	*rem*	putative two-component sensor/regulator; transcriptional regulator	DE	DE	✓	✓
RL0776		malonic semialdehyde oxidative decarboxylase	ES	DE	✓	✓
RL0903	*gph*	putative phosphoglycolate phosphatase	ES	ES	✓	✓
RL0962		putative ring hydroxylating dioxygenase subunit	ES	ES	✓	
RL0994		putative hydroxypyruvate reductase	DE	DE	✓	✓
RL1060		putative ribosomal-protein-alanine acetyltransferase	ES	ES	✓	✓
RL1217		putative TetR family transcriptional regulator	NE	ES	✓	✓
RL1470		putative glycosyl transferase	DE	ES	✓	✓
RL1528		conserved hypothetical protein	DE	DE	✓	
RL1717		putative lactoylglutathione lyase	NE	DE	✓	
RL1732		putative transmembrane protein	DE	DE		
RL2212	*clpS1*	putative ATP-dependent Clp protease adaptor protein	DE	DE	✓	
RL2288	*cysG2*	putative siroheme synthase	ES	ES		
RL2290	*cysI*	putative sulfite reductase	ES	ES		

Continued on next page

Table 13 – continued from previous page

Gene ID	Dsgn	Description	NB	Bd	µA	µC
RL2404		putative peptidyl-prolyl cis-trans isomerase (cyclophilin)	DE	DE		
RL2588	*tyrS1*	putative tyrosyl-tRNA synthetase	ES	ES		
RL2611		hypothetical protein	ES	ES		
RL2924		putative MarR family transcriptional regulator	ES	ES		
RL3453		putative two-component sensor/regulator; histidine kinase	DE	ES		
RL3465		conserved hypothetical protein	DE	DE		
RL3479		putative GTP-dependent nucleic acid-binding protein	NE	ES		
RL3539		putative metal-dependent phosphohydrolases with conserved 'HD' motif	ES	ES		
RL3666		conserved hypothetical protein	NE	DE		
RL3820		putative exopolysaccharide production protein	ES	ES	✓	✓
RL4062		putative amidohydrolase	ES	ES		
RL4117	*glgA*	putative glycogen synthase	ES	ES		
RL4309		putative transmembrane protein	ES	ES		
RL4356		conserved hypothetical protein	DE	DE		
RL4363	*dacC*	putative penicillin-binding protein precursor	DE	DE		
RL4493	*gpsA*	putative glycerol-3-phosphate dehydrogenase [NAD(P)+]	ES	DE		
RL4515	*argG*	putative argininosuccinate synthase	DE	ES		
RL4603		putative aminotransferase	ES	ES		
pRL80079		putative DeoR family transcriptional regulator (repressor)	DE	ES	✓	
pRL80080		putative fructokinase	DE	ES	✓	✓
pRL90263		conserved hypothetical protein	ES	ES	✓	✓
pRL90320		hypothetical protein	ES	ES	✓	✓
pRL100051		hypothetical protein	DE	DE	✓	✓
pRL100133		putative IclR family transcriptional regulator	DE	DE	✓	✓
pRL100134	*aatK*	putative succinate semialdehyde dehydrogenase	DE	DE	✓	

Continued on next page

Table 13 – continued from previous page

Gene ID	Dsgn	Description	NB	Bd	µA	µC
pRL100432		putative lactam utilization protein	ES	ES	✓	✓
pRL110047		putative permease component of ABC transporter Export	ES	DE	✓	✓
pRL110415	*rhaD*	putative short-chain dehydrogenase/oxidoreductase	DE	DE	✓	
pRL120291		putative MarR family transcriptional regulator	DE	DE	✓	
pRL120397		putative GntR family transcriptional regulator	ES	ES	✓	✓
pRL70027		hypothetical protein	DE	DE	✓	✓
pRL70050		hypothetical protein	DE	DE	✓	✓
pRL70154		putative transmembrane conjugal transfer protein	NE	DE	✓	✓
pRL70155		putative conjugal transfer protein TraJ	NE	DE	✓	✓

Table 13: INSeq classifications of genes selected for mutagenesis as part of the screen for strains with delayed nodule development. Dsgn = Designation; NB = Nodule Bacteria; Bd = Bacteroid; µA = Mutant Attempted; µC = Mutant Confirmed by mapping PCR; NE = Neutral; DE = Defective; ES = Essential. INSeq data from Wheatley et al., 2020.

5.4 An Improved System for Mutagenesis of *Rhizobium*

Following the issues with revertants when using pK19mob mutants in Chapter 3.4, it was decided to use a new mutagenesis strategy to eliminate the possibility of losing the mutation through the pK19mob plasmid recombining out of the genome and leaving the gene of interest (GOI) intact. The newly developed pOPS1670/pOPS1713 system (Figure 25A & B, plasmids constructed by R. Ledermann) is capable of making marker-replacement mutants where such revertants cannot arise, since the GOI is completely lost from the genome during mutagenesis. The marker-replacement mutants generated by this system can be further processed into in-frame deletions by using the *loxP* sites in the cassette for Cre-Lox recombination (Albert et al., 1995). In-frame deletion mutants do not suffer from polarity effects on downstream genes, which is advantageous for the study of operons.

The pOPS1670/pOPS1713 system uses two plasmids, pOPS1670 as a backbone and pOPS1713 as a cassette donor. The backbone is largely derived from pJQ200SK and encodes chloramphenicol resistance (*cat*), levansucrase (*sacB*) and *mCherry* as selection markers, while the cassette carries *nptII* for both kanamycin and neomycin resistance. At an earlier stage of development, this cassette encoded *mTag* as a blue fluorescent marker, but this was removed after it was found to slow the growth of *E. coli* considerably during testing. The backbone plasmid (pOPS1670) has an *oriC* origin of replication so it can replicate in *E. coli* but not in *Rhizobium*, becoming a suicide vector in Rlv3841 (Kaniga et al., 1991).

As part of Gibson assembly, it is necessary to linearise the pOPS1670 backbone (using *Hin*dIII and *Sal*I restriction digestion), while the pOPS1713 cassette donor is digested

with *Mly*I to release the 1100 bp cassette fragment encoding *nptII*. The fragments required for Gibson assembly (5874 bp and 1100 bp from pOPS1670 and pOPS1713, respectively) were purified following extraction from agarose gels. Construction of mutagenic vectors involves the Gibson assembly of two flanking regions, each approximately one kilobase in length (with appropriate adapters, see Chapter 2.3) with the linear *nptII* cassette into the linear plasmid backbone, generating a circular mutagenic plasmid comprised of backbone-upstream_flank-cassette-downstream_flank, as shown in Figure 25C (Gibson et al., 2009).

Successful *E. coli* transformants are selected on plates containing chloramphenicol and kanamycin. After conjugation into *Rhizobium*, homologous single recombination events may insert the mutagenic plasmid into the upstream or downstream flank of the target gene. Double-crossover events, where the cassette has replaced the GOI (Figure 26), are identified by direct selection on plates containing both 1% sucrose and high levels of neomycin (Nm_{240}) in a single selection step. Selection plates also contained streptomycin to select against the *E. coli* donor and helper strains present in the conjugation mixture. Colonies on the selection plate resulting from single-crossover events, where the target gene is still present, are prevented due to sucrose sensitivity imparted by *sacB* (Ried & Collmer, 1987). Should the sucrose selection fail (e.g. due to random mutation of *sacB*), colonies from single-crossovers are visible due to *mCherry* expression from the backbone of the inserted mutagenic plasmid, allowing these colonies to be avoided. Rhizobial colonies, resistant to neomycin and lacking *mCherry* expression, were selected for mapping by PCR direct from colonies and further purification. Mutants were not confirmed by sequencing in the interest of time.

Due to time constraints brought about by Covid-19 pandemic-related disruptions, attempts were made to generate mutants in only 40 genes from the list of 60 candidates (Table 13), using the OPS2575 strain as a background (Rlv3841 Ps*nifH*:*sfgfp*). Of these 40 attempts, only 25 marker-replacement mutants were successfully generated and confirmed by PCR-mapping of the insertion (Table 13). Of the 15 unsuccessful attempts, 10 were abandoned due to difficulties with the PCR amplification of one of the flanking regions. The flavonoid *lux* biosensor plasmid pLMB712 (P*nodA*:*lux*; Pini et al., 2017) was then successfully conjugated into each of the 25 mutants (Table 13), and the background strain OPS2575 as a wild type control.

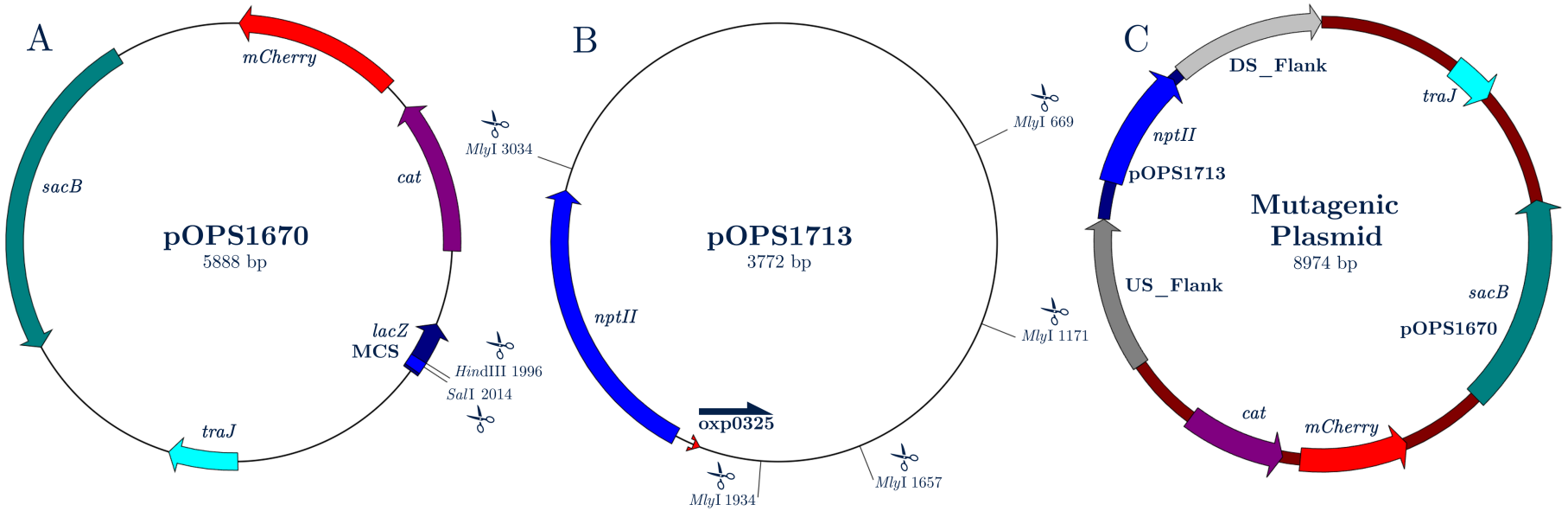


Figure 25. Plasmid maps from the pOPS1670/pOPS1713 system. (A) Backbone plasmid pOPS1670 with *mCherry* and *cat* (encoding chloramphenicol resistance) as selection markers as well as a multiple cloning site within *lacZ*. Restriction enzymes *Sal*I and *Hin*dIII cut at this site during the assembly of the mutagenic plasmid. (B) Plasmid pOPS1713 provides the cassette, which has *nptII* (encoding kanamycin and neomycin resistance) as a selectable marker and is excised by *Mly*I digestion as a 1100 bp fragment. Primer oxp0325 binds upstream of *nptII* to amplify outwards from the cassette in PCR-mapping of insertions in mutants. (C) An example of a mutagenic plasmid showing key marker genes: *nptII*, *cat*, *mCherry*, and *sacB* after Gibson assembly of linear backbone (from pOPS1670), cassette (from pOPS1713) and the PCR-amplified upstream (US_Flank) and downstream (DS_Flank) flanking regions (approximately 1000 bp each) of the GOI to create a circular plasmid. Created with SAVVY (bioinformatics.org/savvy).

Figure 26

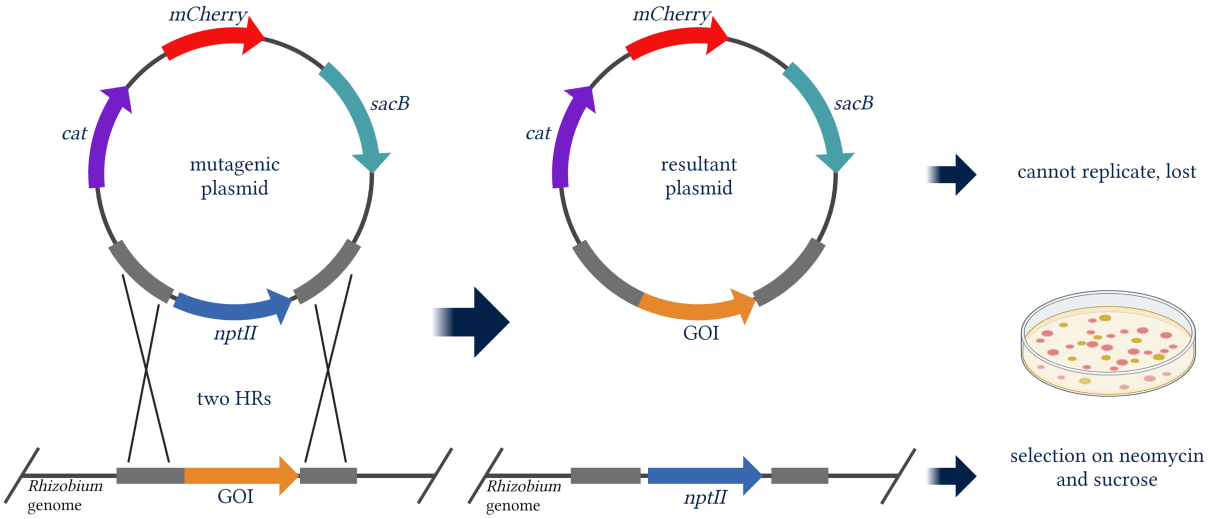


Figure 26. Double homologous recombination of the mutagenic plasmid with the flanking regions of the GOI. These recombinations result in the replacement of the GOI with the kanamycin/neomycin resistance marker *nptII*. The mutagenic plasmid lacks an appropriate origin of replication to persist in *Rhizobium*, thus is lost. Selection media contains both neomycin and sucrose to screen out non-recombinants and recombinants from single crossovers respectively. Created with BioRender (biorender.com).

5.5 Nodulation Kinetics Time-Courses

The mutant strains generated (Table 13) were assessed for defective nodule development of inoculated plants in a time-course experiment. In order to allow the detection of signal from the *gfp* and *lux*-based biosensors present in all strains, the host pea plants were grown in transparent plastic pouches (Figure 27A). These plastic pouches each contained a paper insert to aid moisture distribution and retention, while the whole pouch could be sterilised by autoclaving to avoid contamination.

Surface sterilised pea seeds were pre-germinated on agar for seven days in the dark, then transferred to plastic pouches containing nitrogen-free rooting solution, where they grew for another seven days in a controlled environment growth chamber. After this period, the pea seedlings usually had established healthy roots growing down the wet paper insert, others were discarded. The seedlings were then inoculated with a rhizobial strain. Immediately after inoculation, the pouches were imaged in the NightOWL imaging system and luminescence was measured as the timepoint for day zero. The plants were returned to the growth chamber (Figure 27B) and imaged about once every two days. The roots were also examined under a fluorescence stereomicroscope to check for the appearance of nodules with green fluorescence. Nodules with bright green fluorescence were judged to have *nifH* promoter activity and therefore could be assumed to be fixing nitrogen. Figure 28 shows examples of nodules which were judged to have bright green fluorescence and also some which were not. These nodules were labelled on the plastic pouch and recorded along with the date at which they started to show this fluorescence. Monitoring in this way was continued until no new fluorescent nodules were detected for four consecutive days. Figure 27A shows an example pouch following the labelling of fluorescent nodules.

Figure 27

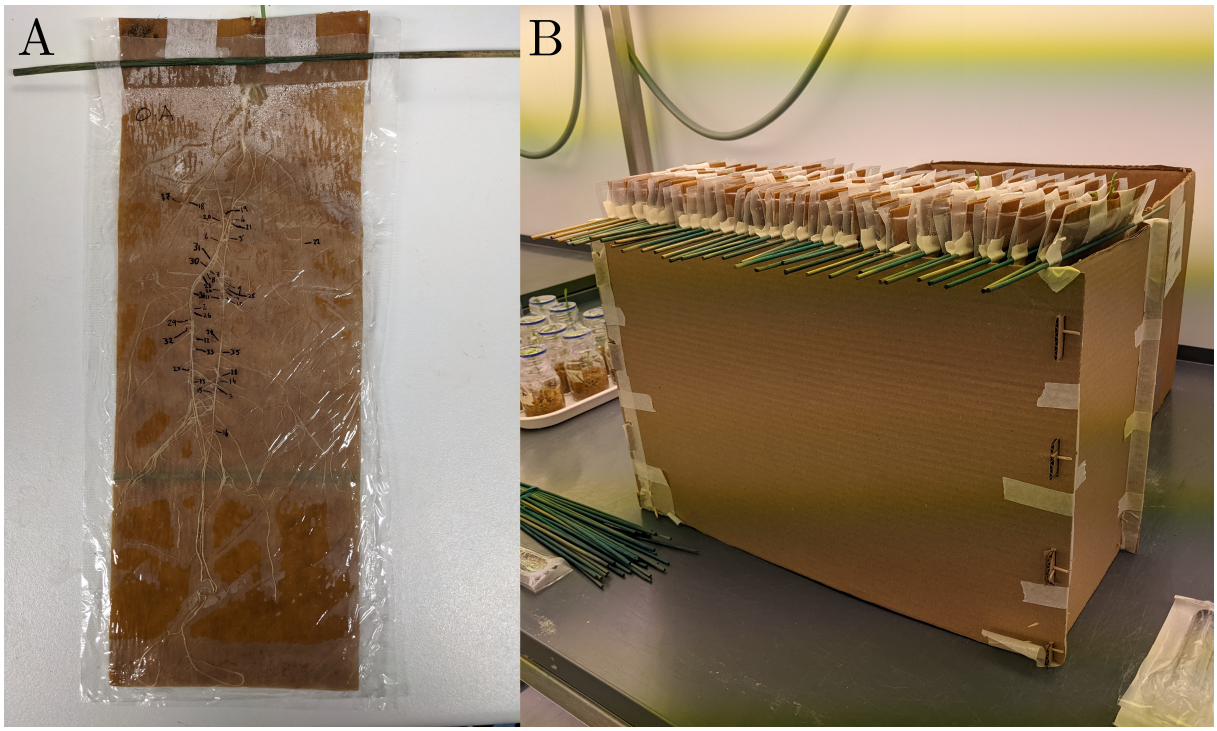


Figure 27. Pea growth in pouches during nodule development screen. A) Roots of a pea, 38-days-old, growing in a plastic pouch. The brown paper has taken up 10 ml of N-free rooting solution to provide a substrate for the roots. This pouch was inoculated with the wild type strain (OPS2905) and 36 nodules were recorded to have formed on this plant in the 24 days post inoculation (indicated by black labels in marker pen, each number corresponding to a date recorded in a key). The horizontal spar is affixed to the top of the pouch to keep the pouch positioned in the growth box. B) A growth box containing 35 plastic pouches, each containing a seven-day-old pea seedling. These seedlings were grown in the controlled environment room for seven days before inoculation with test strains, then for a further 24 days. By this stage, no new nodules had formed on any plant for four consecutive days.

The 25 mutant strains were divided into four batches of six or seven, with each mutant strain inoculated onto four or five plants. This setup is detailed in Table 14. During the course of this first batch of experiments (Batch A), it was discovered that the NightOWL camera suite was not sensitive enough to detect the green fluorescence from fixing nodules as planned. The methodology was altered to utilise a fluorescence stereomicroscope to identify fluorescent nodules (Figure 28). However, by the time of this change to the protocol, most of the nodules on the Batch A plants had already developed and were showing green fluorescence. Batches B, C and D were at early enough stages to make better use of this change in methodology. However, Batch D suffered from unusually high levels of fungal contamination, killing many plants and leaving only the wild type and one mutant strain (ΔRL0994) with more than one usable biological replicate. Due to the impact of departmental closure and restrictions on facilities use, neither Batch A nor Batch D were able to be repeated.

Figure 28

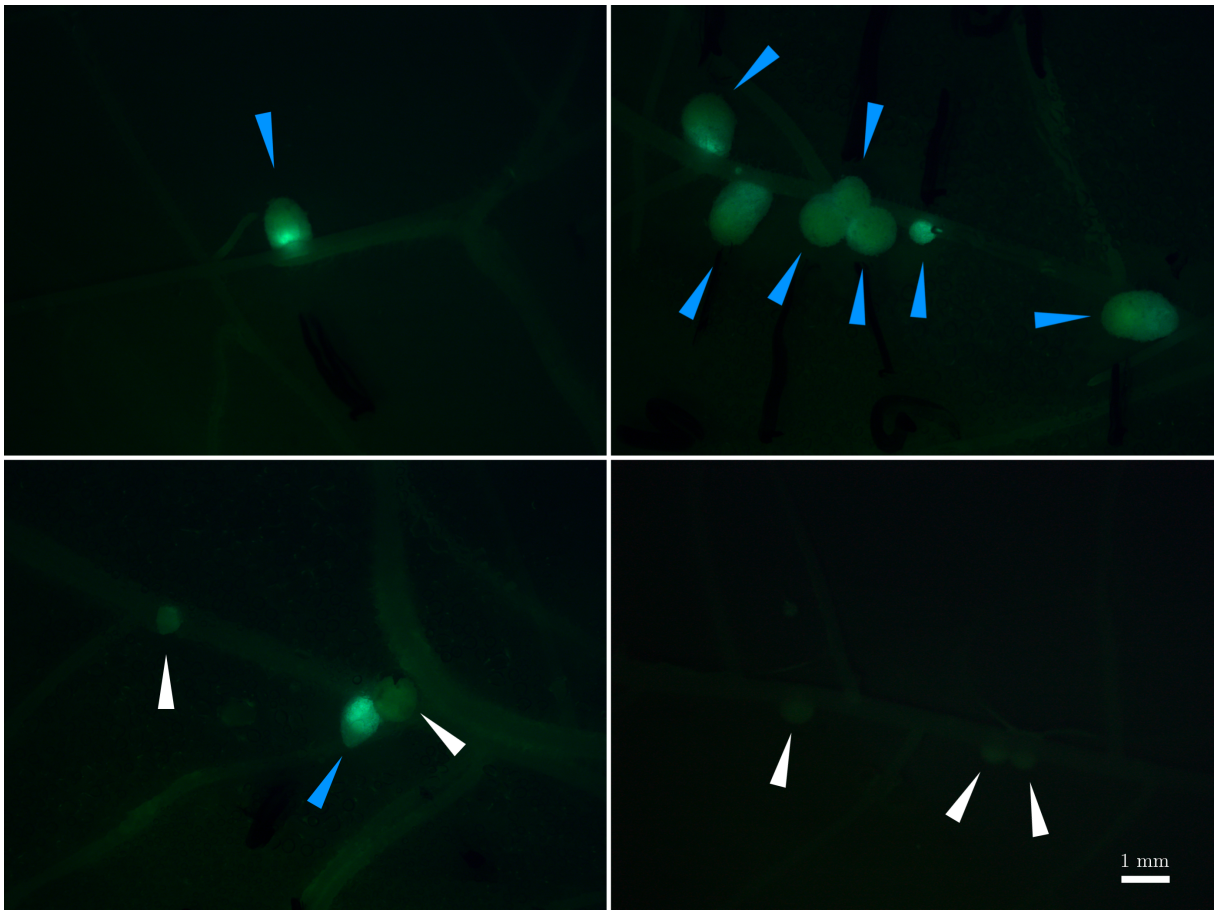


Figure 28. Fluorescence microscopy images of nodules occupied by OPS2905 on pea roots. OPS2905 is a wild type strain marked with sfGFP (under the control of Ps*nifH*:*sfgfp*)and was used as a control to establish the onset of nitrogen fixation for time-course experiments. This biosensor was judged to be *on* in nodules marked with blue arrowheads, and to be *off* in nodules marked with white arrowheads. The scale bar represents 1 mm. Taken with the Leica Application Suite on a Leica M165 FC.

Table 14

Strain	Mutated in	Description	Plant Death
Batch A	**Fluorescent nodules missed, therefore data was not usable**		
OPS2905	None	wild type	0 of 5
OPS2915	pRL100432	putative lactam utilization protein	0 of 4
OPS2916	pRL80080	putative fructokinase	0 of 4
OPS2922	RL0143	putative PfkB family sugar kinase	0 of 4
OPS2918	*hrcA* (RL0381)	putative heat-inducible transcription repressor	0 of 4
OPS2924	*rem* (RL0727)	putative two-component sensor/regulator; transcriptional regulator	0 of 4
OPS2906	pRL100051	hypothetical protein	0 of 4
OPS2911	pRL70027	hypothetical protein	1 of 4
Batch B	**Data quality reduced as a consequence of plants dying**		
OPS2905	None	wild type	0 of 5
OPS2907	RL3820	putative exopolysaccharide production protein	1 of 4
OPS2920	RL0572	putative dihydroorotate dehydrogenase	1 of 4
OPS2925	RL0776	malonic semialdehyde oxidative decarboxylase	2 of 4
OPS2926	*gph* (RL0903)	putative phosphoglycolate phosphatase	1 of 4
OPS2928	RL1060	putative ribosomal-protein-alanine acetyltransferase	1 of 4
OPS2929	RL1217	putative TetR family transcriptional regulator	1 of 4
Batch C	**Best quality data**		
OPS2905	None	wild type	0 of 5
OPS2908	pRL100133	putative IclR family transcriptional regulator	0 of 5
OPS2909	pRL110047	putative permease component of ABC transporter Export	0 of 5
OPS2910	pRL120397	putative GntR family transcriptional regulator	0 of 5
OPS2912	pRL70050	hypothetical protein	0 of 5
OPS2913	pRL70154	putative transmembrane conjugal transfer protein	0 of 4
OPS2914	pRL70155	putative conjugal transfer protein TraJ	0 of 4

Continued on next page

Table 14 – continued from previous page

Strain	Mutated in	Description	Plant Death
Batch D	**Many plants contaminated, only wildtype and ΔRL0994 data usable**		
OPS2905	None	wild type	3 of 5
OPS2921	pRL90263	conserved hypothetical protein	4 of 4
OPS2917	pRL90320	hypothetical protein	3 of 4
OPS2919	RL0501	putative orotate phosphoribosyltransferase	4 of 4
OPS2930	RL1470	putative glycosyl transferase	4 of 4
OPS2923	RL0726	conserved hypothetical exported protein	3 of 4
OPS2927	RL0994	putative hydroxypyruvate reductase	2 of 4

Table 14: Division of mutants into experimental batches for screening. Each batch also included OPS2905 as a wild type control and two pouches which were inoculated with sterile water as uninoculated controls.

5.6 Monitoring Nodule Development with P*nodA* and Ps*nifH* Biosensors in Concert

Following the completion of the time-courses, the NightOWL luminescence data were analysed to find the date at which luminescence above a threshold was first detected over the location of each future nodule. This threshold was set at 150 counts per second, based on the luminescence pattern of one of the wild type control plants over the time-course. This analysis was completed manually for the 442 images that comprise the data of the highest quality batch, Batch C. These date-of-first-luminescence data were combined with the date-of-first-fluorescence data to find the time interval between luminescence (expression driven by P*nodA*, indicating the initiation of nodulation) and nodule fluorescence (expression driven by Ps*nifH* expression, indicating the onset of nitrogen fixation) for each nodule. The individual nodule data were averaged for each plant and used to make Figure 29, showing the experimental results from Batch C.

The mean time interval between detected luminescence and detected nodule fluorescence was 7.05 days post luminescence (dpL) for the wild type control strain (OPS2905, Figure 29). All mutant strains had similar mean time intervals, except for the strain mutated in pRL120397 (OPS2910), whose nodules appeared to develop slightly faster (fluorescence at 5.58 dpL, Figure 29). However, this was not found to be significantly different from wild type following a Dunnett's multiple comparisons test (Figure 29).

Figure 29

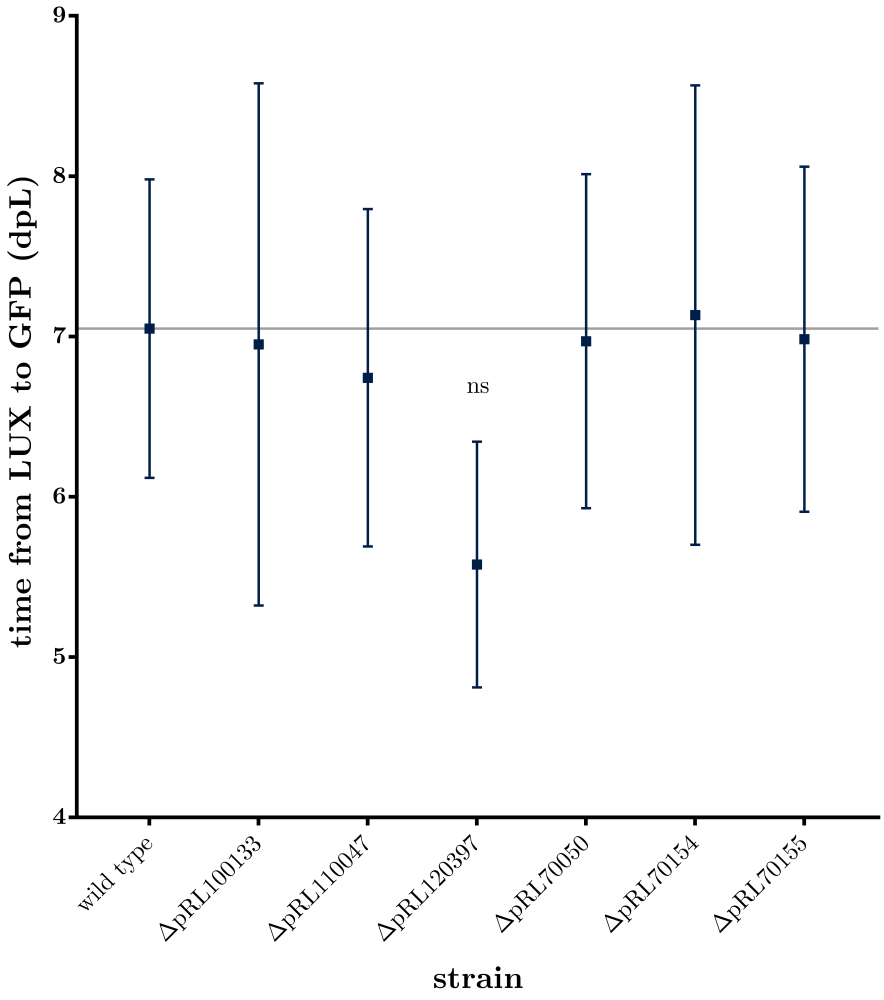


Figure 29. Onset of nodulation and nitrogen fixation in Batch C strains monitored by P*nodA*:*lux* and Ps*nifH*:*sfgfp* biosensors. The mean time intervals (days post luminescence, dpL) between detection of luminescence above a threshold at a nodule formation site and detection of green fluorescence from a mature nodule on that site were measured. Wild type (OPS2905) had a mean time interval of 7.05 ±0.931 dpL. Error bars represent ±SEM. Statistical significance was assessed by carrying out a Dunnett's test comparing each mutant strain with wild type. n = 5 for all strains except ΔpRL120397, ΔpRL70154 and ΔpRL70155, where n = 4. Each biological replicate consisted of a single plant.

5.7 Monitoring Nodule Development with P*nodA* and Ps*nifH* Biosensors Individually

While it was the original intention of the screen to measure the interval between the activity of the two biosensors, this meant there were twice as many opportunities for error and noise to affect the results when compared to a single biosensor time-course. To mitigate this, the activity of each biosensor was individually examined with reference to the inoculation date of the plants. Results from these methods of analysis are shown in Figure 30 for the Ps*nifH*:*sfgfp* biosensor and Figure 31 for the P*nodA*:*lux* biosensor.

For the wild type (OPS2905) the mean time between inoculation and nodule fluorescence was 12.8 ±1.05 days post inoculation (dpi, Figure 30). All mutant strains had similar times of approximately 13 days, with no significant differences in time to nodulate when assessed with a Dunnett's multiple comparisons test (Figure 30). Figure 31 shows the data from the P*nodA*:*lux* biosensor to represent the timing of the earliest stages of nodulation. It took on average 5.88 dpi to detect the luminescence signal for wild type nodule formation sites (Figure 31). Nodule formation sites formed by mutant strains showed broadly similar timing (5.5-6.5 dpi), although the strains mutated in pRL120397 and pRL70050 were slower, detected at 6.67 and 7.01 dpi, respectively. As luminescence was measured every two days, it may be necessary for future experiments to have a higher frequency of measurement, such as daily or every 12 hours, to improve resolution of potentially small differences.

Figure 30

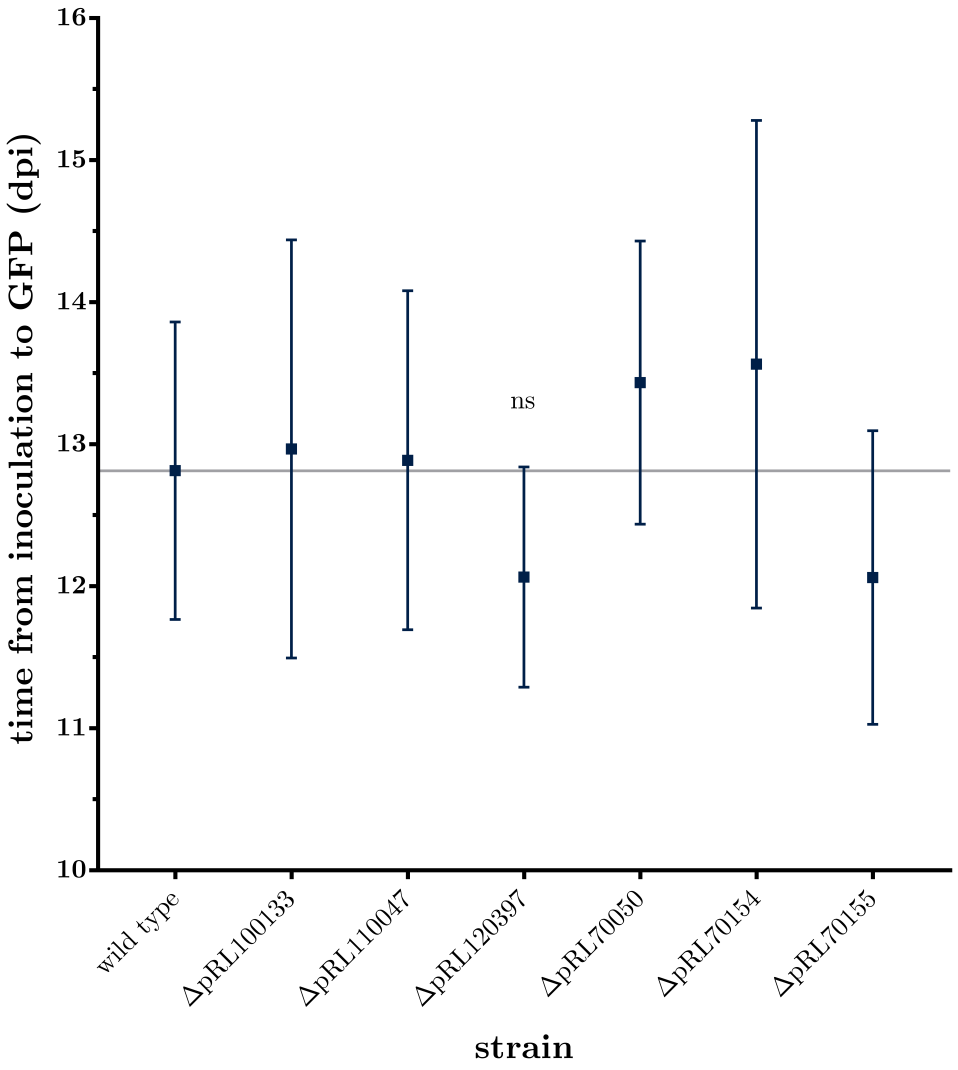


Figure 30. Onset of nodulation and nitrogen fixation in Batch C strains monitored by the Ps*nifH*:*sfgfp* biosensor. The mean times (dpi) are shown for a mature nodule to develop and show green fluorescence (indicating the onset of nitrogen fixation). Wild type (OPS2905) had a mean time interval of 12.8 ±1.05 dpi before nodule fluorescence. Error bars represent ±SEM. Statistical significance was assessed by carrying out a Dunnett's test comparing each mutant strain with wild type. n = 5 biological replicates, each consisting of a single plant, for all strains except ΔpRL70154 and ΔpRL70155, where n = 4.

Figure 31

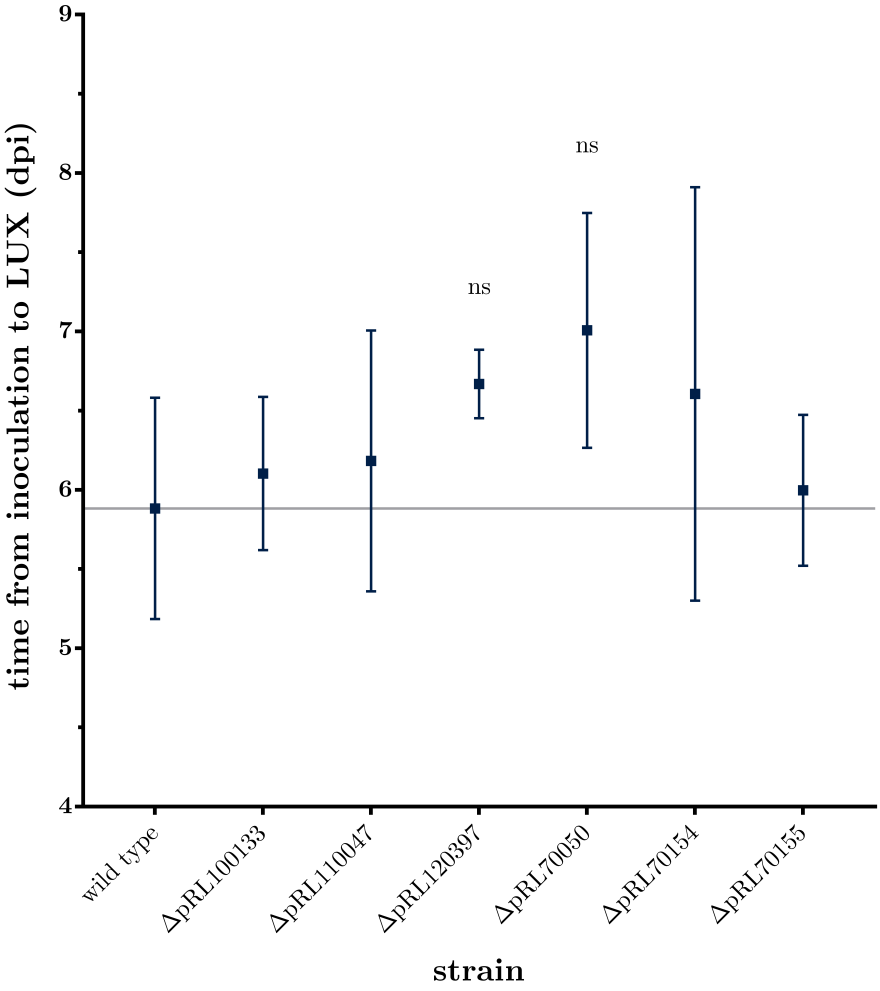


Figure 31. Onset of nodulation in Batch C strains monitored by the P*nodA*:*lux* biosensor. Mean time (dpi) is shown for luminescence above a threshold (150 counts per second, indicating the onset of initiation of nodulation). Wild type (OPS2905) had a mean time interval of 5.88 ±0.699 dpi before nodule formation site luminescence. Error bars represent ±SEM. Statistical significance was assessed by carrying out a Dunnett's test comparing each mutant strain with wild type. n = 5 biological replicates, each consisting of a single plant, for all strains except ΔpRL120397, ΔpRL70154 and ΔpRL70155, where n = 4.

5.8 Kinetics of Nodulation

There was little biological variation in the data from the P*nodA*:*lux* biosensor alone between strains (SD of means = 0.41 days), which reduces the likelihood of resolving differences between strains when considering measurements were taken at intervals of up to two days. However, the data from the Ps*nifH*:*sfgfp* biosensor alone showed a slightly greater range of biological variation between strains (SD of means = 0.59 days). Therefore the Ps*nifH*:*sfgfp* biosensor data was judged to be more informative for following the kinetics of nodulation in detail and potentially resolving differences between strains.

The earliest fluorescent nodules were detected at 8 dpi, while the rate of new nodule formation had slowed considerably by day 16 (Figure 32). Logistic growth curves were plotted to fit the datapoints, but there were no significant differences between these curves for the mutant strains tested (Dunnett's multiple comparisons tests as part of a two-way ANOVA, Figure 32).

It is apparent from the variation in these data that future experiments may be improved by increasing the number of plant replicates and reducing the time interval between measurements, although this would reduce the throughput of this experiment and make it less useful for screening large numbers of mutant strains at a time. On the other hand, if it were necessary to substantially increase the throughput of the experiment, this could be achieved by counting the bright-green-fluorescent nodules at a single timepoint.

It is clear from Figure 32 that the greatest differences in nodule number between faster- and slower-nodulating strains are at 12 and 14 dpi. In Figure 33, the nodules detected by

14 dpi from each strain in Batch C are shown. From this data, some variation between the mutant strains can be observed (Figure 33), though not significant differences from a Dunnett's multiple comparisons test.

Figure 32

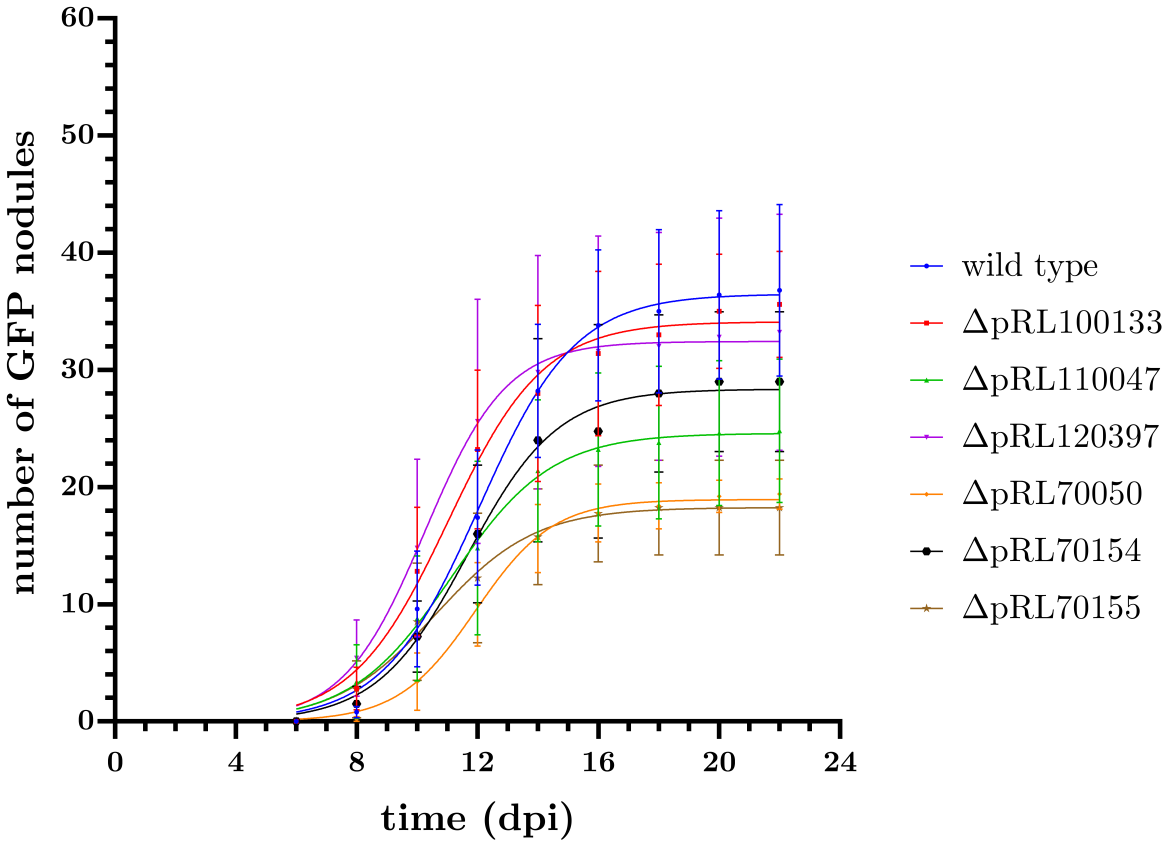


Figure 32. Kinetics of nodulation from following green fluorescence (expression driven by P*snifH*:*sfgfp*) for strains from Batch C. Green fluorescent, mature nodules were counted for each plant over time and the mean number plotted against time (dpi). Logistic growth curves were plotted using the non-linear fit function of GraphPad Prism. Error bars represent ±SEM. Dunnett's multiple comparisons tests failed to find any statistically significant differences between wild type (OPS1905) and any mutant strain, at any time-point, as part of a two-way ANOVA. n = 5 biological replicates, each consisting of a single plant, for all strains except ΔpRL70154 and ΔpRL70155, where n = 4.

Figure 33

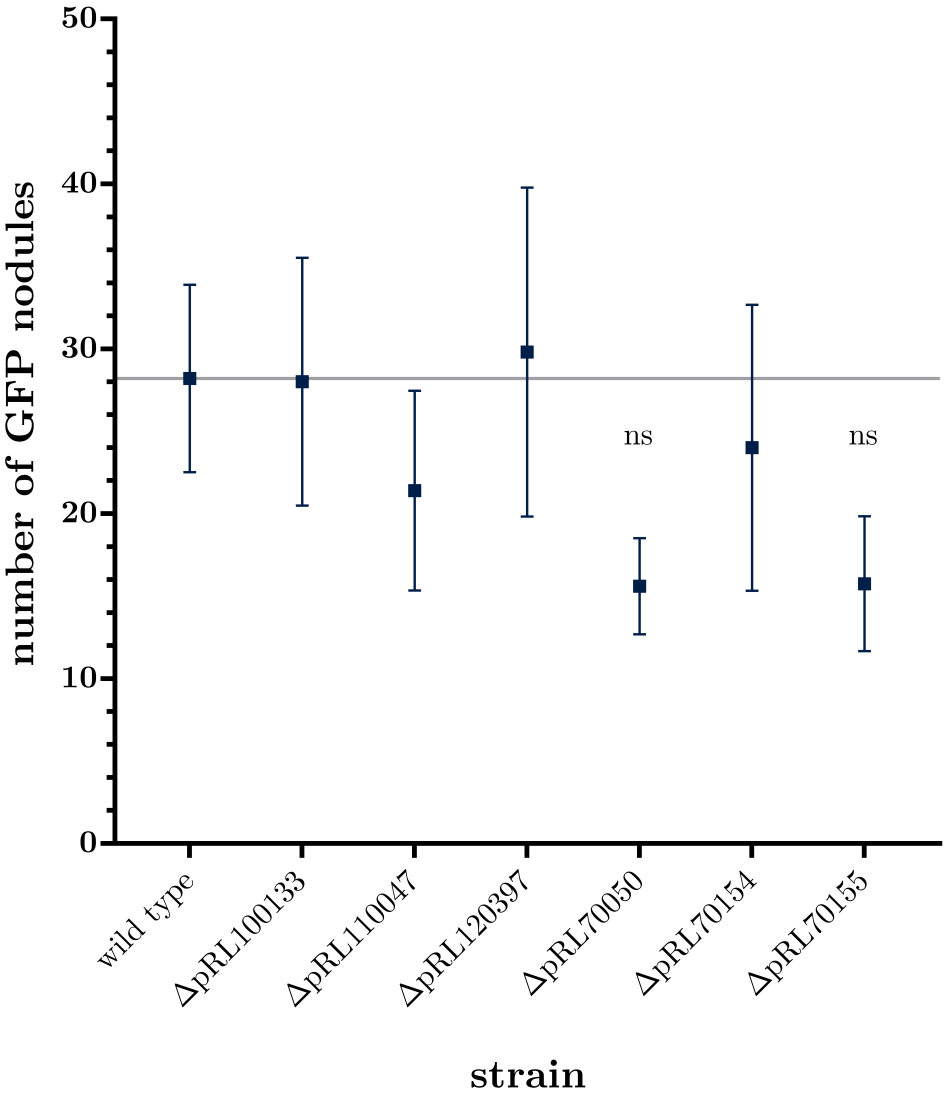


Figure 33. Mean number of green fluorescent (Ps*nifH*:*sfgfp*)nodules per plant at 14 dpi for strains from Batch C. Wild type (OPS2905) had a mean of 28.2 ±5.68 nodules per plant. Error bars represent ±SEM. Statistical significance was assessed by carrying out a Dunnett's test comparing each mutant strain with wild type. The strains mutated in pRL70050 and pRL70155 gave non-significant p values of 0.6004 and 0.6643, respectively. n = 5 for all strains except ΔpRL70154 and ΔpRL70155, where n = 4. Each biological replicate consisted of a single plant.

5.9 Monitoring Nodule Development for Two Further Batches

Batches B and D gave less complete data than Batch C as a consequence of plants becoming infected by pathogens and/or dying during experiments. These problems are a key limitation of the plastic-paper pouch system, as the pouches are open at the top to allow watering and the shoot to grow out, which leaves the nutrient-soaked paper vulnerable to colonisation by fungal species. The problem may have been exacerbated by the tight packing of pouches into the growth boxes (Figure 27B).

The data from the P*nifH*:*sfgfp* biosensor were focused upon for Batches B and D due to observed interference in the P*nodA*:*lux* biosensor by fungal contamination of the growth pouches. The intervals between inoculation and bright green nodule fluorescence for Batches B (Figure 34) and D (Figure 35) were calculated. In Batch B, the wild type strain was slightly later to form mature fluorescent nodules (Figure 34), an average of 18.2 ±1.01 dpi, compared to 12.8 ±1.05 dpi in Batch C. Several mutant strains tested in Batch B showed nodule formation at approximately 15 days (ΔRL0572, Δ*gph* & ΔRL1060), but this was not significantly earlier than the wild type strain when assessed with Dunnett's multiple comparisons tests (Figure 34). Comparable results are seen for Batch D (Figure 35), where the strain mutated in RL0994 (13.7 ±2.61 dpi) was not significantly different from wild type (14.1 ±1.90 dpi). However, with only two usable biological replicates for each strain in this dataset, the confidence of this result is not as hgh as was hoped.

Figure 34

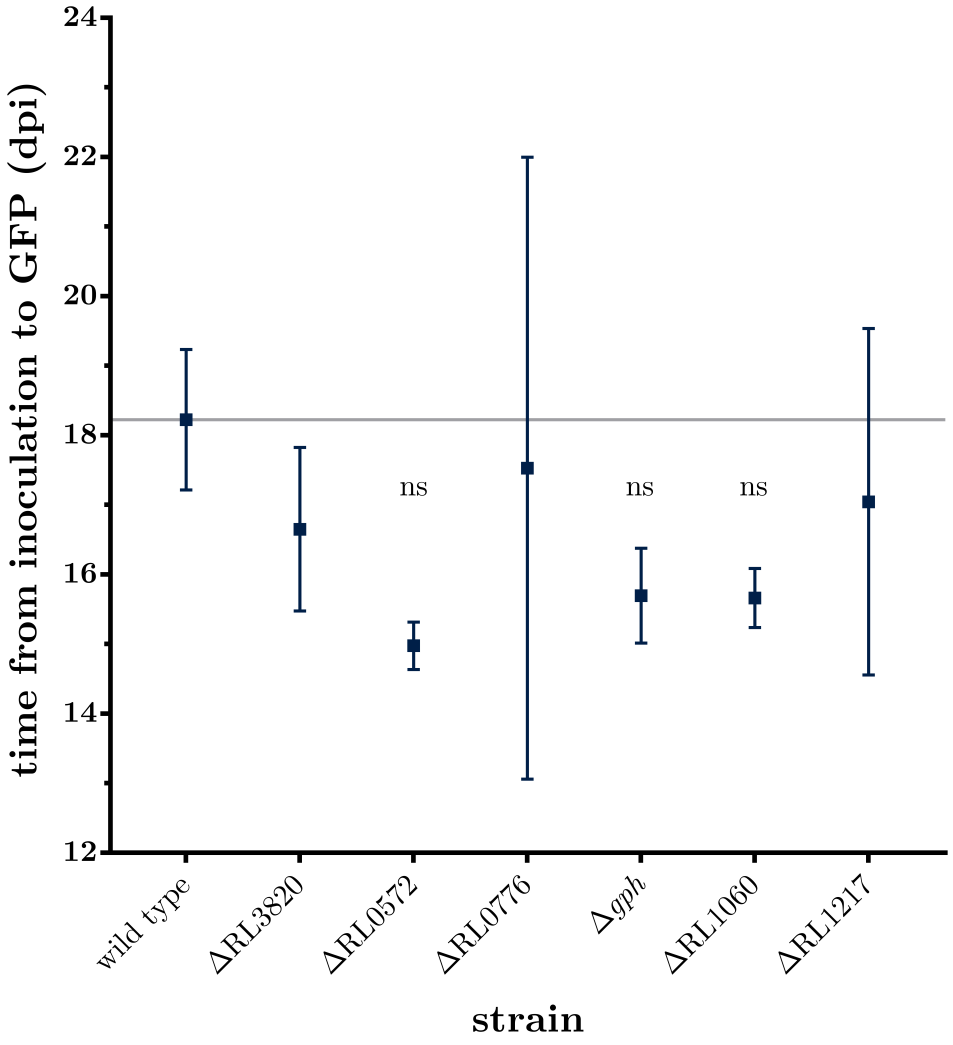


Figure 34. Onset of nodulation and nitrogen fixation in Batch B strains monitored by the Ps*nifH*:*sfgfp* biosensor. The mean times (dpi) are shown for a mature nodule to develop and show green fluorescence (indicating the onset of nitrogen fixation). Wild type (OPS2905) had a mean time interval of 18.2 ±1.01 dpi before nodule fluorescence. Error bars represent ±SEM. Statistical significance was assessed by carrying out a Dunnett's test comparing each mutant strain with wild type. n = 2 for the RL0776 mutant, n = 5 for the wild type control, and n = 3 for all other strains. Each biological replicate consisted of a single plant.

Figure 35

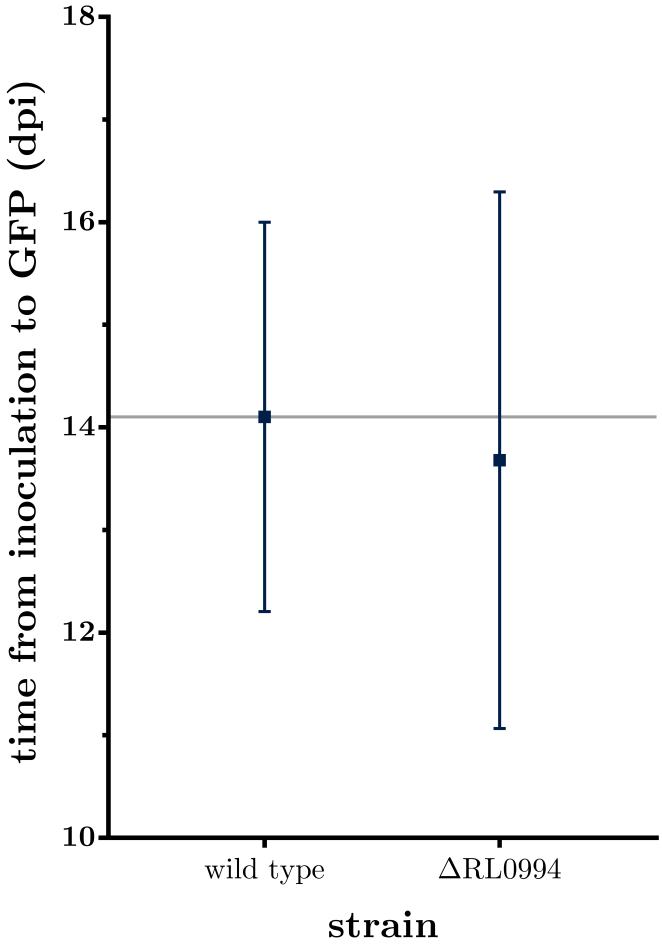


Figure 35. Onset of nodulation and nitrogen fixation in Batch D strains monitored by the Ps*nifH*:*sfgfp* biosensor. The mean times (dpi) are shown for a mature nodule to develop and show green fluorescence (indicating the onset of nitrogen fixation). OPS2905 was used as a wild type control and had a mean time interval of 14.1 ±1.90 dpi before nodule fluorescence. Error bars represent ±SEM. Statistical significance was assessed by carrying out a two-tailed t test with Welch's correction, which did not find a significant difference between the two strains. n = 2 biological replicates for both strains, each consisting of a single plant.

5.10 Discussion

5.10.1 New Mutagenesis System with pOPS1670/pOPS1713

In this chapter, a new mutagenesis strategy was designed for *Rhizobium* and was successfully implemented. The pOPS1670/pOPS1713 system was used to generate mutants in 25 of the 40 target genes attempted (Table 13). Most of the unsuccessful mutagenesis attempts (15 in total) were abandoned due to difficulties amplifying at least one of the GOI flanks by PCR. These problems would likely be overcome by designing new primers to amplify a slightly different length of flanking region, but due to the impact of Covid-19 pandemic-related restrictions, it was decided to press on and screen the phenotypes of as many genes on the list of candidates as possible.

Compared to the pK19mob insertion mutagenesis system, the pOPS1670/pOPS1713 system requires two homologous regions to be successfully amplified by PCR instead of just one (Schäfer et al., 1994). This means there is twice the chance that an ineffective pair of primers will delay the successful mutagenesis of a given GOI. However, redesigning the primers is likely to resolve this, as the marker replacement mutagenesis does not need to remove the entire gene in order to disrupt it.

The pOPS1670/pOPS1713 system has considerable advantages over the pK19mob system. The marker replacement mutants generated with pOPS1670/pOPS1713 by double-crossover homologous recombination are stable deletions and thus cannot give rise to revertants in the same way that pK19mob insertions can be shed from the target gene. Selecting for double crossovers with pOPS1670/pOPS1713 is straightforward due to the red mCherry visible marker and *sacB* (encoding levansucrase for sucrose sensitivity) in

the pOPS1670 backbone, so conjugant colonies without red fluorescence can be selected following a one-step selection and mapped. Cre-Lox recombination can then be optionally used to excise the cassette from genome and create an in-frame deletion without resistance markers and without polarity effects on downstream genes (Albert et al., 1995). This would also allow further mutagenesis with the pOPS1670/pOPS1713 to generate strains with multiple mutations. A future objective is to generate derivatives of the cassette carried by pOPS1713 in which *nptII* (kanamycin/neomycin resistance) is replaced by other resistance markers, e.g. *aacC1* (gentamycin resistance) or *aaD9* (spectinomycin resistance), to increase the versatility of the system by allowing the mutation of strains with pre-existing kanamycin/neomycin resistance.

5.10.2 Nodule Kinetics Time-Courses

Mutants generated from the new system were used to develop a new screening method for identifying mutants with phenotypes in bacteroid and nodule development. The P*nodA*:*lux* and Ps*nifH*:*sfgfp* biosensors were brought together to allow measurement of two key milestones of nodule development: the start of *Rhizobium*-legume communication and the start of nitrogen fixation (Geurts & Bisseling, 2002; Mendoza-Suárez et al., 2020; Pini et al., 2017). This new method was used to screen 13 mutants from the list of candidate genes informed by INSeq (Figures 29 through 35, summarised by Figure 36). Using this method none of these 13 mutants showed a significantly altered rate of nodule development. It is therefore likely that these 13 mutants have impaired competitive nodulation ability, impaired nitrogen-fixing ability, growth defects, or any one of the myriad other phenotypes that may lead to ES or DE INSeq classification. Their phenotypes could be confirmed by experiments such as those described in Chapter 3. Further to

the lack of delayed nodule development mutants identified, evaluation of the effectiveness of this screen is further limited by the lack of a positive control strain (one which shows delayed onset of nitrogen fixation compared to wild type which is statistically significant), although this cannot be helped as the screen is for a novel phenotype.

The Ps*nifH*:*sfgfp* biosensor may be considered more reliable than the P*nodA*:*lux* biosensor, as the latter was often unable to generate a signal before nodules appeared in specific regions of root. This rate of 'non-detection' was 34.1% on average for plants in Batch C, though this was as high as 100% for some plants, leading to the discrepancy in number of biological replicates between the two biosensors. This problem was especially pronounced for plants inoculated with the pRL120397 mutant, where over 60% of nodules were not preceded by detectable luminescence on three plants.

The P*nodA*:*lux* biosensor was also affected by the fungal contamination afflicting some of the plastic pouches. This has been observed in other experiments (Roworth, 2017). It is possible that the biosensor was induced by flavonoids derived from fungus instead of the plant host, since the pouches with the most fungal contamination appeared to have stronger, earlier and more widespread luminescence compared to uncontaminated pouches. Reducing fungal contamination is therefore a key consideration for the screening strategy moving forward, and might be achieved by better sealing the tops of the pouches or the inclusion of fungicide to the rooting solution.

Figure 36

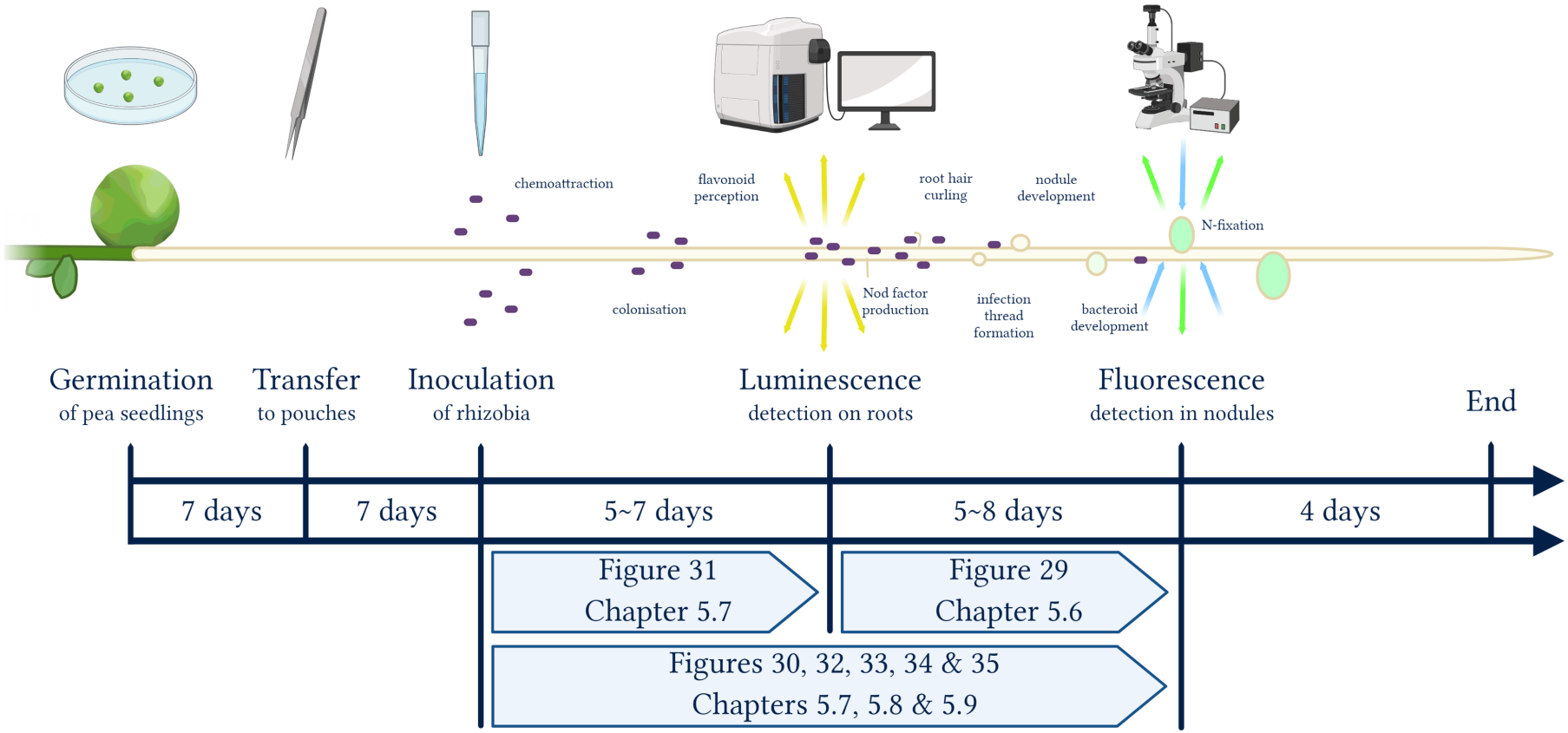


Figure 36. Timeline of key experimental and biological milestones during screening experiments performed in Chapter 5. The P*nodA*:*lux* biosensor (luminescence upon rhizobial perception of plant flavonoids) and the Ps*nifH*:*sfgfp* biosensor (fluorescence as nitrogen fixation begins in bacteroids) were examined both in concert and individually in different analyses. Not to scale. Created with BioRender (biorender.com).

5.10.3 Future Perspectives

Further work must be carried out in order to identify a mutant with delayed nodule development. The 12 mutants which were not screened in this work (being in Batches A and D) provide a starting point towards this, as do the 35 candidate genes for mutagenesis for which mutants were not confirmed (Table 13). This was unable to be accomplished prior to writing as a result of Covid-19 pandemic-related disruption.

The P*nodA*:*lux*/Ps*nifH*:*sfgfp* time-courses proved too complex and low-throughput in practice. This, combined with the observed unreliability of the P*nodA*:*lux* biosensor to measure exactly what it was hoped to, leads to the conclusion that future experiments should use the Ps*nifH*:*sfgfp* biosensor alone. Time-courses with the Ps*nifH*:*sfgfp* biosensor could be adapted to follow the kinetics of competitive nodulation through co-inoculation with wild type marked with the Ps*nifH*:*mCherry* biosensor (OPS2576). In the future, transgenic plant lines with host-side reporter systems, e.g. reporters based on ENOD promoters, could be used to add an additional dataset to the time-course, though this would be a considerable undertaking to develop such lines (Miklashevichs et al., 2001).

The labour-intensive and low-throughput nature of these time-courses could be partly mitigated by carrying out preliminary experiments to pre-emptively identify mutants unlikely to display a delayed development phenotype. Such experiments would aim to confirm or disconfirm other potential phenotypes that could lead to the gene being classified as ES or DE in the nodule bacteroid INSeq dataset: the major selection criteria for the candidate mutants (Wheatley et al., 2020). Examples of these preliminary experiments could include nodulation competition assays (to screen out mutants with impaired

nodulation ability) and acetylene reduction assays at 21 dpi (to screen out mutants with impaired nitrogen-fixing ability); since both impairments could be expected to lead to the INSeq classification alone and their presence reduces the likelihood that a delayed nodule development phenotype is responsible for the INSeq classification.

One of the biggest difficulties with the plastic pouch growth system was keeping the pouches free from fungal contamination, partly due to the pouches being open at the top and the tight packing of pouches in the growth boxes. It is therefore important to choose only the strongest germinated seedlings (longest roots) for transplantation into the pouches, and to set up more pouches than necessary for the experiment. An alternative could be to return to the more sterile square agar plate system with vetch, as used in Pini et al., 2017. However, this square plate system has often had problems with water retention, which necessitates the transfer of the growing plants to new plates when the current plate dries out. Plates can dry out in as quickly as a week, and each transfer is labour-intensive and risky for the plant. After the conclusion of the imaging time-course, plants grown in the square plate system could be measured for whole-plant nitrogenase activity by acetylene reduction assays as a further assessment of nitrogen fixation.

One further alternative could be to grow peas in vermiculite-filled boiling tubes. This system leads to healthy plants and good growth but does not allow non-invasive imaging of roots. Therefore, an experimental setup with this system would use many more replicates with destructive sampling, potentially with the addition of acetylene reduction assays to assess whole-plant nitrogenase activity instead of the Ps*nifH*:*sfgfp* biosensor.

5.10.4 Significance

A novel strategy for screening rhizobial mutants for a phenotype in nodule development and, by extension, bacteroid development has been conceived, developed, and tested. With further refinement and continued use of this strategy, it is hoped that previously unknown mutants with these phenotypes may be found and characterised (Figure 36). Identification of such mutants would provide new insights into the regulation of bacteroid development and assist in the global effort to transfer *Rhizobium* symbiosis to cereals (Haskett et al., 2021; Pankievicz et al., 2019). In addition, the newly-development pOPS1670/OPS1713 mutagenesis system will prove a useful tool for future mutant studies of *Rhizobium* (Figures 25 & 26).

6 General Conclusions and Future Perspectives

6.1 Impact of the Covid-19 Pandemic

Large parts of this work were carried out during and after Covid-19 pandemic-related lockdowns and working restrictions. Laboratory spaces were closed from March 2020 until July 2020, followed by laboratory shifts of approximately one-third hours until October 2020 and approximately half hours until April 2021. These disruptions were equivalent to losing approximately nine months of laboratory time and prevented several planned experiments in Chapters 4 and 5 from being carried out.

6.2 Context of Research

The global nitrogen cycle has been in crisis due to excessive usage of artificial nitrogen fertiliser causing hundreds of millions of tonnes of greenhouse gas emissions and devastating damage to aquatic ecosystems (Clúa et al., 2018; Rockström et al., 2009). Meanwhile, millions of the world's farmers live in poverty and populations suffer from food insecurity whilst their crops remain nitrogen-limited (Mueller et al., 2012; Oldroyd & Dixon, 2014). Improving the utilisation of biological nitrogen fixation could address both problems.

Farmers can use rhizobial inoculants on their legume crops to boost yields and add nitrogen to their soil by increasing biological nitrogen fixation in the field (Clayton et al., 2004; Deaker et al., 2004). In order to extract the most benefit from inoculation, inoculants must comprise rhizobial strains with both the high nitrogen-fixing ability and competitive nodulation ability, lest the inoculant strain be out-competed by indigenous rhizobial strains (Triplett & Sadowsky, 1992).

6.3 INSeq Data-Informed Experimental Investigation

Genetic factors underlying competitive nodulation ability are poorly understood, but establishment of high throughput transposon insertion sequencing techniques (e.g. INSeq), provides an opportunity to investigate on a genomic scale the contribution of genes to competitive nodulation ability (Perry & Yost, 2014; Wheatley et al., 2020). The INSeq data generated by the Poole Lab (published in Wheatley et al., 2020) characterised the impacts of mutations on fitness for 7,357 genes across four stages of symbiosis (rhizosphere, root colonisation, nodule bacteria and bacteroids), identifying 603 genetic regions required for competitive nodulation (Wheatley et al., 2020). These data were used as starting points for investigation into competitive nodulation ability and nodule development described in this work.

INSeq experiments are intrinsically competitive because they involve libaries of thousands of strains competing with one another for survival and, in the specific case of the symbiosis INSeq experiment, nodulation of a legume host. One of the major reasons for non-neutral (ES, DE or AD) INSeq classifications in the nodule bacteroid library is an altered number of nodules occupied by these mutant strains relative to strains with wild type phenotypes and NE classifications (Wheatley et al., 2020). This is because the number of nodules occupied by a strain is directly correlated with the amount of recoverable DNA from that strain and therefore to the subsequent number of sequencing reads. Therefore, competitive nodulation assays, which aimed to measure the number of nodules that a strain could form in a competitive environment, were some of the closest small-scale experiments in this work to the intrinsically competitive INSeq experiment. As a consequence of this similarity, it is not surprising that more mutant phenotypes (eight)

were identified by competitive nodulation assays than any other type of experiment in this work (Table 15).

A second reason behind non-neutral INSeq classifications is an altered number of bacteroids in nodules occupied by a mutant strain. This may be through sanctioning by the host plant in response to reduced levels of nitrogen fixation from bacteroids (Kiers et al., 2003; Westhoek et al., 2017). It is likely in most cases to result in smaller nodules containing the strain with lower nitrogen-fixing ability. The nitrogen-fixing ability of mutant strains was investigated by nitrogenase activity assays and dry weight assays as proxies (Chapters 3.5 & 4.5).

A diverse set of 15 genes (Table 4), which were classified as ES or DE for nodule bacteria and/or bacteroids, were selected for investigation of the phenotype of mutants and are described in Chapter 3.4. In initial experiments, INSeq classifications were compared to the results with single mutant strains to allow the determination of where the mutants were affected in symbiosis in greater resolution than allowed by INSeq.

These methods of investigation were extended to genes with AD INSeq classifications, for which the mutation conferred an apparent fitness advantage during the intrinsically competitive INSeq experiments. Improving understanding of genes with AD classifications may help with the development of more competitive inoculant strains. In Chapter 4, nodule bacteria and bacteroid INSeq data were examined to select eight of the most 'strongly advantaged' genes (Table 7) as candidates for mutagenesis and further characterisation.

INSeq data were also used as the basis for the larger mutant screen described in Chapter 5, where the objective was to identify a mutant with altered nodulation kinetics. It was hypothesised that mutants delayed in bacteroid development would have INSeq classifications of ES or DE for the bacteroid dataset, as a consequence of sanctions on nodules with delayed fixation leading to a reduction in the amount of rhizobial DNA recoverable from such nodules (Liu & Murray, 2016; Walker et al., 2020; Westhoek et al., 2021). A list of 60 relatively uncharacterised genes with this INSeq classification (Table 13) were selected as good candidates and in the time available 25 of them were mutated (Table 15).

6.4 Dual Mutagenesis Strategies Were Utilised

Initial investigations in this work made use of the pK19mob system (Schäfer et al., 1994) to successfully generate mutant strains for 22 genes in Chapters 3 and 4 (Table 15). However, the ability of the inserted pK19mob to recombine out from the rhizobial genome (and potentially restore wild type gene structure and function) was a problem for the investigation of the *nodB* mutant in Chapter 3.4 in particular. This instability of pK19mob insertion mutants under high selection pressures therefore poses problems for investigations of mutants during experiments free from wild type competition. The pK19mob mutants are also highly polar, meaning the expression of genes downstream of the target gene may be disrupted.

In light of pK19mob insertion instability and polarity, it was decided to develop an alternative mutagenesis and screening strategy for nodule development (Chapter 5). This was especially necessary as the screen initially involved single-strain inoculations of pea,

without competition from wild type rhizobia, where revertants would considerably affect results. To this end, the newly developed pOPS1670/pOPS1713 was used to generate 25 marker replacement mutations (Figure 26, Table 13).

The requirements of double-crossover homologous recombination contributed to the comparatively low total success rate for generating pOPS1670/pOPS1713 mutants, with 25 confirmed mutants from 40 attempts (Table 13) compared to 22 confirmed mutants from 23 attempts with pK19mob (Tables 4 & 7). However, time limitations imposed by Covid-19 pandemic-related restrictions prevented more thoroughly troubleshooting cloning issues, so it is likely that the success rate with this mutagenesis system could have been increased if more time had been available. Gibson assembly and conjugation resulted in successful generation of mutants for the vast majority of genes for which flanking regions were successfully amplified (29 of 34). Therefore, the success rate for this new mutagenesis strategy was over 85% after the most difficult step of PCR amplification was completed.

6.5 Defining the Stage of Symbiosis Affected

In this work, a total of 47 mutant strains were generated and experimentally investigated (Table 15). Results from small-scale experiments were compared with INSeq classifications and used to determine the stage of symbiosis affected by mutations in 18 genes (Figure 37 & Table 15).

Figure 37

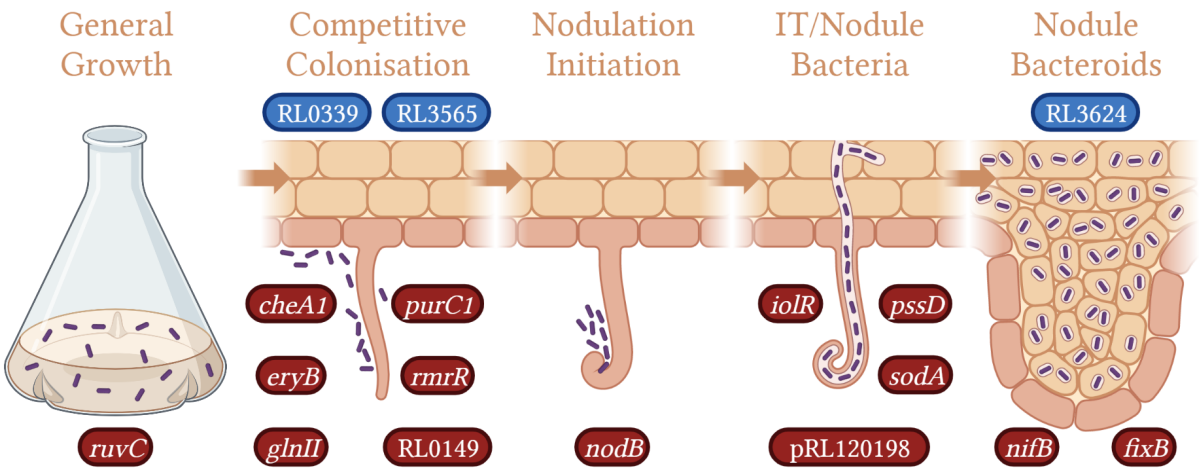


Figure 37. Identified mutant phenotypes affecting general growth and four stages of *Rhizobium*-legume symbiosis. Mutations highlighted above in blue were found to confer fitness advantages for the relevant stage, while mutations highlighted below in red impaired rhizobial fitness. Mutants without observed phenotypes are not shown in this figure, but are included in Table 15 overleaf. Created with BioRender (biorender.com).

Table 15

Gene Information			INSeq		Experimental Results							
Gene ID	**Mutant**	**Description**	**NB**	**Bd**	**Nod.**	**N-ase Act.**	**Comp.**	**Recov.**	**Grow.**	**Dry Wt**	**Nod. Size**	**Nod. Dev.**
Affected Stage: General Growth												
RL3986	Δ*ruvC*	Putative Holliday junction endodeoxyribonuclease RuvC	DE	NE	WT	WT	ND	–	–	ND	ND	ND
Affected Stage: Competitive Colonisation												
pRL120205	Δ*eryB*	Putative erythritol phosphate dehydrogenase EryB	ES	ES	WT	WT	– –	ND	ND	ND	ND	ND
pRL90058	Δ*rmrR*	Putative TetR family regulator of RND family efflux transporter	ES	ES	WT	WT	–	ND	ND	ND	ND	ND
RL0149	ΔRL0149	Putative XRE family transcriptional regulator	DE	DE	WT	WT	–	ND	ND	ND	ND	ND
RL0339	ΔRL0339	Putative Mg2+ chelatase family protein	NE	AD	ND	ND	+	ND	ND	WT	ND	ND
RL0688	Δ*cheA1*	Putative chemotaxis protein CheA	ES	ES	WT	WT	–	ND	ND	ND	ND	ND
RL2606	Δ*purC1*	Putative phosphoribosylaminoimidazolesuccinocarboxamide synthetase	ES	ES	WT	WT	–	ND	ND	ND	ND	ND
RL3549	Δ*glnII*	Putative glutamine synthetase	ES	DE	WT	WT	–	ND	ND	ND	ND	ND
RL3565	ΔRL3565	Putative sensory box HD domain protein	NE	AD	ND	ND	+	ND	ND	WT	ND	ND

Continued on next page

Table 15 – continued from previous page

Gene Information			INSeq		Experimental Results							
Gene ID	**Mutant**	**Description**	**NB**	**Bd**	**Nod.**	**N-ase Act.**	**Comp.**	**Recov.**	**Grow.**	**Dry Wt**	**Nod. Size**	**Nod. Dev.**
Affected Stage: Initiation of Nodulation												
pRL100186	Δ*nodB*	Putative chitooligosaccharide deacetylase NodB	DE	DE	– –*	ND	ND	ND	ND	ND	ND	ND
Affected Stage: Infection Thread & Nodule Bacteria												
pRL120198	ΔpRL120198	Putative L-xyulose kinase	ES	NE	WT	WT	ND	–	WT	ND	ND	ND
RL1340	Δ*sodA*	Putative superoxide dismutase SodA	ES	NE	WT	WT	ND	–	WT	ND	ND	ND
RL1496	Δ*iolR*	Putative sugar-isomerase RpiR family regulator	ES	NE	WT	WT	ND	–	WT	ND	ND	ND
RL3654	Δ*pssD*	Putative polysacharide biosynthesis protein	ES	ES	– –	– –	– –	ND	ND	ND	ND	ND
Affected Stage: Nodule Bacteroids												
pRL100195	Δ*nifB*	FeMo cofactor biosynthesis protein NifB	NE	DE	+	– –	ND	ND	ND	ND	ND	ND
pRL100199	Δ*fixB*	Electron transfer protein FixB	DE	DE	+	– –	ND	ND	ND	ND	ND	ND
RL3624	ΔRL3624	Putative SBP of ABC transporter CUT2 GLU2 ST1	NE	AD†	–	ND	WT	ND	ND	+	–	ND
Affected Stage: Not Determined												
pRL100133	ΔpRL100133	Putative IclR family transcriptional regulator	DE	DE	ND	ND	ND	ND	ND	ND	ND	WT

Continued on next page

Table 15 – continued from previous page

Gene Information			INSeq		Experimental Results							
Gene ID	**Mutant**	**Description**	**NB**	**Bd**	**Nod.**	**N-ase Act.**	**Comp.**	**Recov.**	**Grow.**	**Dry Wt**	**Nod. Size**	**Nod. Dev.**
pRL110047	ΔpRL110047	Putative permease component of ABC transporter Export	ES	DE	ND	ND	ND	ND	ND	ND	ND	WT
pRL120397	ΔpRL120397	Putative GntR family transcriptional regulator	ES	ES	ND	ND	ND	ND	ND	ND	ND	WT
pRL70050	ΔpRL70050	Hypothetical protein	DE	DE	ND	ND	ND	ND	ND	ND	ND	WT
pRL70154	ΔpRL70154	Putative transmembrane conjugal transfer protein	NE	DE	ND	ND	ND	ND	ND	ND	ND	WT
pRL70155	ΔpRL70155	Putative conjugal transfer protein TraJ	NE	DE	ND	ND	ND	ND	ND	ND	ND	WT
pRL70182	ΔpRL70182	Conserved hypothetical exported protein	NE	AD	ND	ND	WT	ND	ND	WT	ND	ND
RL0572	ΔRL0572	Putative dihydroorotate dehydrogenase	ES	ES	ND	ND	ND	ND	ND	ND	ND	WT
RL0748	Δ*aglA1*	Putative alpha-glucosidase	AD	AD	ND	ND	WT	ND	ND	WT	ND	ND
RL0776	ΔRL0776	Malonic semialdehyde oxidative decarboxylase	ES	DE	ND	ND	ND	ND	ND	ND	ND	WT
RL0903	Δ*gph*	Putative phosphoglycolate phosphatase	ES	ES	ND	ND	ND	ND	ND	ND	ND	WT
RL0994	ΔRL0994	Putative hydroxypyruvate reductase	DE	DE	ND	ND	ND	ND	ND	ND	ND	WT

Continued on next page

Table 15 – continued from previous page

Gene Information			INSeq		Experimental Results							
Gene ID	**Mutant**	**Description**	**NB**	**Bd**	**Nod.**	**N-ase Act.**	**Comp.**	**Recov.**	**Grow.**	**Dry Wt**	**Nod. Size**	**Nod. Dev.**
RL1060	ΔRL1060	Putative ribosomal-protein-alanineacetyltransferase	ES	ES	ND	ND	ND	ND	ND	ND	ND	WT
RL1217	ΔRL1217	Putative TetR family transcriptional regulator	NE	ES	ND	ND	ND	ND	ND	ND	ND	WT
RL1545	ΔRL1545	Putative short-chain dehydrogenase/reductase	ES	ES	WT	WT	WT	ND	ND	ND	ND	ND
RL2036	ΔRL2036	Putative outer membrane transport protein	AD	NE	ND	ND	WT	ND	ND	WT	ND	ND
RL3820	ΔRL3820	Putative exopolysaccharide production protein	ES	ES	ND	ND	ND	ND	ND	ND	ND	WT
RL4139	ΔRL4139	Putative transmembrane GG-DEF/EAL sensory box protein	AD	AD	ND	ND	WT	ND	ND	WT	ND	ND

Table 15: Summary of the INSeq classification and mutant phenotypes. INSeq classifications for Nodule Bacteria (**NB**) and Bacteroids (**Bd**) were: Advantaged (AD), Neutral (NE), Defective (DE) and Essential (ES). The following experiments were performed to characterise mutant phenotypes: Nodulation assays (Nod., Chapters 3.4 & 4.8), Nitrogenase activity assays (N-ase Act., Chapter 3.5), Competitive nodulation assays (Comp., Chapters 3.6 & 4.6), Recovery assays (Recov., Chapter 3.7), Growth assays (Grow., Chapter 3.8), Dry weight assay (Dry Wt, Chapters 4.5), Nodule size comparisons (Nod. Size, Chapter 4.7) and Nodule development time-courses (Nod. Dev.,

Chapters 5.6 through 5.9). 'WT' represents wild type levels; a plus sign (+) significant enhancement; a minus sign (−) significant impairment; two minus signs (− −) no measurable activity; and 'ND' stands for 'Not Determined'. *The *nodB* mutant was unable to form nodules that did not contain revertants. †The INSeq reads for RL3624 were manually curated which resulted in reclassification as AD for bacteroids.

6.5.1 Identified Phenotypes at the Stage of Competitive Colonisation

The stages of symbiosis between rhizobial inoculation and the start of *Rhizobium*-legume communication were investigated using competitive nodulation assays. This includes colonisation of the rhizosphere, colonisation of roots and attachment to roots in a competitive environment. Eight mutants were established to affect competitive colonisation, two strains (ΔRL0339, ΔRL3565) showed increased, and six strains (Δ*eryB*, Δ*cheA1*, Δ*glnII*, Δ*rmrR*, Δ*purC1* & ΔRL0149) a decreased ability to form nodules in competition with wild-type Rl3841 (Figures 14 & 15, Table 15).

Double-sequential staining with *gusA*- and *celB*-chromogenically marked strains (Sánchez-Cañizares & Palacios, 2013) were used in the competitive nodulation assays. Experiments carried out with 1:1 inoculations were sensitive enough to allow the detection relatively small increases (ΔRL0339, ΔRL3565) and decreases (Δ*purC1*, ΔRL0149) in competitive nodulation ability (Figure 37, Table 15). This is despite the levels of competition in 1:1 inoculations being far below those in an INSeq experiment, where thousands of mutant stains are competing to nodulate the host legume instead of just two strains. Genetic factors found to influence competitive colonisation included genes encoding chemotaxis proteins (*cheA1*), regulators (*rmrR*, RL3565), catabolic enzymes (*eryB*, RL0149), biosynthetic enzymes (*glnII*, *purC1*) and an Mg^{2+} chelator (RL0339, Figure 37, Table 15).

Signal transduction by the *che1* chemotaxis gene cluster can drive rhizobial movement across a number of chemical gradients (including host legume root exudates) and has been shown to be important for initial attachment to pea roots and chemotaxis towards root hair tips, both critical events in competitive nodulation (Aroney, 2021). The importance

of the *che1* cluster for competitive nodulation of pea was show in previous mutant studies (Miller et al., 2007) and the same has now been specifically shown for RL0688 (*cheA1*, Table 15).

Another protein thought to be involved in signal transduction is the gene product of RL3565, a HD domain sensory box protein (Nagata et al., 2008). The strain mutated in RL3565 was found to have a small but significant increase in competitive nodulation ability (Table 15). The strain mutated in pRL90058 (*rmrR*), encoding another regulator, showed the opposite phenotype: a severe reduction in competitive nodulation ability. This fits with previously reported results from *R. etli* strains mutated in *rmrA* and *rmrB*, the two genes regulated by *rmrR* (Ramos et al., 2005).

Mutants in two biosynthetic genes, RL2606 (*purC1*) and RL3549 (*glnII*), also showed reduced fitness at this stage of symbiosis (Table 15). Therefore, ability to synthesise purines (PurC1; Stevens et al., 2000) and glutamine (GlnII; Carlson et al., 1987) are likely to be important for rhizobial fitness in the nutrient-limited rhizosphere prior to nodulation. Interestingly, *glnII* and *purC2* (homologue of *purC1*) were highlighted as being important for root attachment in alkaline conditions (*glnII*) and across multiple pHs (*purC2*) by another INSeq experiment (Parsons, 2019).

6.5.2 Identified Phenotypes at the Stage of Nodulation Initiation

Nodulation assays generated data which showed that the strain mutated in pRL100186 (*nodB*) was unable to properly nodulate peas as a single-strain inoculant (Table 15, Chapter 3.4), likely due to an inability to produce a Nod factor recognisable by the

host legume (Venturi & Keel, 2016). The investigation of this mutant was hampered by revertant mutants arising from pK19mob insertion instability in the rhizobial genome. While *nodB* mutants could be found inside nodules during nodulation assays, revertant strains were present in all tested nodules (Figure 7). This was an example of Nod^- mutants being complemented by nearby Nod^+ rhizobia (Geurts & Bisseling, 2002).

6.5.3 Identified Phenotypes at the Stages of Infection Thread and Nodule Bacteria

Four mutants (Δ*pssD*, ΔpRL120198, Δ*sodA* and Δ*iolR*) were found to affect rhizobia as bacteria within nodules (Table 15). Analysis of the growth rate of pRL120198, RL1340 (*sodA*) and RL1496 (*iolR*) mutant bacteria recovered from nodule extracts, an analogous assay to the original INSeq experiment, revealed the impaired regrowth ability of these strains compared to wild type. It may be that nodule bacteria mutated in *sodA* are less able to tolerate immune responses (including oxidative bursts) from host legumes (Berrabah et al., 2019; Santos et al., 2001), as the superoxide dismutase SodA is likely to have a role in tolerance to reactive oxygen species (Hassett et al., 1995; McCord & Fridovich, 1969). This provides one potential mechanism which could have led to the recovery of less-healthy nodule bacteria. Potential explanations for the phenotypes of the pRL120198 and RL1496 (*iolR*) mutants could involve inability of nodule bacteria to utilise xylulose (pRL120198-encoded xylulose kinase; Duncan, 1981; Lawlis et al., 1984) or inositol compounds (IolR; Kohler et al., 2011; Wood and Stanway, 2001), although further experiments would be necessary to test these hypotheses.

The strain mutated in RL3654 (*pssD*) was unable to form nitrogen-fixing nodules with wild type morphology (Table 15, Figure 5). Microscopy performed on *R. leguminosarum* bv. *trifolii pssD* mutants in clover nodules revealed that rhizobia mutated in *pssD* appeared unable to move from the infection thread into symbiosomes, possibly as a result of a breakdown of recognition between host and rhizobia (Król et al., 1998). A similar mechanism may be behind the phenotype of *R. leguminosarum* bv. *viciae pssD* mutants.

6.5.4 Identified Phenotypes at the Stage of Nodule Bacteroids

The final stage of symbiosis examined by INSeq was the abundance of bacteroids in nodules (Wheatley et al., 2020). The sanctioning of under-fixing nodules, such as the non-fixing nodules occupied by pRL100195 (*nifB*) and pRL100199 (*fixB*) mutants (Table 15), resulted in visibly smaller, white nodules (Figure 5) that could be expected to contain fewer bacteroids and therefore less rhizobial DNA to be sequenced in INSeq experiments (Wheatley et al., 2020). Nitrogenase activity assays confirmed that plants inoculated with *nifB* or *fixB* mutants had no detectable nitrogenase activity (Table 37), as was expected for these two Fix$^-$ control strains.

Inoculation of the RL3624 mutant led to significantly enhanced plant growth promotion in nitrogen-limited conditions compared to wild type in dry weight assays (Table 15). These experiments measured dry wieght as a proxy for nitrogen fixation (Maier & Brill, 1978; Williams & Phillips, 1983), but these results alone cannot be used to conclude that increased nitrogen fixation is responsible for either the plant growth promotion phenotype or the INSeq classification of RL3624 as AD. This hypothesis may be further tested by measuring a second proxy for nitrogen fixation;, e.g. measuring nitrogenase activity

through acetylene reduction assays (Shah et al., 1975). further evidence that increased nitrogen fixation is responsible for the plant growth promotion phenotype might be obtained by measuring another proxy for nitrogen fixation in the future, e.g. measuring nitrogenase activity by acetylene reduction assays (Shah et al., 1975).

Comparison of the size of nodules occupied by the RL3624 mutant and wild type did not reveal significantly smaller wild type nodules. This can be interpreted as evidence against the hypothesis that increased nitrogen fixation ability is responsible for the RL3624 mutant phenotype, since the host plant could be expected to sanction nodules occupied by a relatively lower-fixing strain (Westhoek et al., 2021).

Thirteen mutant strains were examined by the screening strategy with the aim to find mutants with delayed nodule development (from Chapter 5.6). However, none of these 13 mutant strains showed significantly altered timing of nodulation (Table 15). Identification of such mutants would provide new insights into the regulation of bacteroid development and assist in the global effort to transfer *Rhizobium* symbiosis to cereals (Haskett et al., 2021; Pankievicz et al., 2019). It is hoped that further use and refinement of this mutant screening strategy, particularly with regard to the contamination-prone growth system, could result in the identification of a mutant with this novel phenotype.

6.5.5 Identified Phenotypes at the Stage of Free-Living Growth

The RL3986 (*ruvC*) gene, which encodes a putative Holliday junction endodeoxyribonuclease, was classified by INSeq as DE in nodule bacteria and NE for all other conditions (Table 4; Wheatley et al., 2020). While the strain mutated in *ruvC* did show significantly

impaired regrowth from nodule extracts (Chapter 3.7, Table 15), analysis of plate-reader growth data revealed that free-living growth was also impaired (Chapter 3.8, Table 15). Thus, a mutant in *ruvC* is not specifically affected at a stage of *Rhizobium*-legume symbiosis. In *E. coli*, RuvC is important for resolution of Holliday structures, a crucial process in DNA repair and recombination, disruption of which could be expected to impair growth (Asai & Kogoma, 1994). Analysis of the growth data for the *ruvC* mutant shows that this is likely to be similar for Rlv3841.

6.6 Future Directions

The results described in this work have set up numerous avenues for further investigation of *Rhizobium*-legume symbioses. It was originally hoped that progress could be made towards many of the more obvious lines of further research within this work (e.g. screening more mutants for delayed nodule development), but increased time constraints due to Covid-19 pandemic-related lockdowns and working restrictions thwarted this possibility. The proposed experiments are described in more detail in the discussion sections of the relevant chapter.

The aim of the experiments described in Chapter 3 was to define in which stages of symbiosis each mutant was affected. Whilst the five assays (Table 15) performed have helped towards this aim as previously described, there is the potential for further experiments to enhance the resolution of these definitions. Further investigations would involve follow-up experiments specific to each of the 14 mutants (Table 15), examples of which are given in Chapter 3.9.3.

Future experiments to follow on from the investigations of AD mutants in Chapter 4 can be divided into two categories: experiments to find phenotypes for the four mutants (Δ*aglA1*, ΔpRL70182, ΔRL2036 & ΔRL4139) which did not show a phenotype, and experiments to further investigate the three mutants (ΔRL0339, ΔRL3565 & ΔRL3624) for which phenotypes were discovered. The former category would include investigation of bacteroid endoreduplication using DAPI staining and flow cytometry, as well as quantification of nodule bacteria and bacteroid numbers using fluorescent labelling, density separation gradients and flow cytometry. In the latter category, the links between the expression of RL3565 (competitive nodulation advantaged mutant), *aglA1* and other genes highlighted by the coexpression matrices could be investigated by RT-qPCR or RNAseq on relevant mutants. Acetylene reduction assays could be performed on the growth promoting RL3624 mutant to check if nitrogenase activity is higher, which, if found, would agree with the hypothesis that the RL3624 mutant has higher nitrogen-fixing ability. These proposed experiments are outlined in more detail in Chapter 4.11.2.

The nodule development mutant screen described in Chapter 5 was the part of this work most affected by Covid-19 pandemic-related disruption. As a result, the future avenues of research are clearer. The 12 mutants which were unable to be successfully screened in time could be tested in time-courses with the P*nodA*:*lux* and Ps*nifH*:*sfgfp* biosensors. Second, the 35 genes remaining as candidates for mutagenesis could be mutated then phenotypically screened, with new primers designed for these genes as necessary. The new pOPS1670/pOP1713 mutagenesis system worked well and would continue to be used for further mutagenesis. Preliminary screening experiments would consist of low-replicate competitive nodulation assays to determine if there is an impairment in competitive nod-

ulation ability behind the ES/DE INSeq classification for each mutant. Plasmids with *gusA* and *celB* chromogenic markers could be conjugated into the remaining mutants for this purpose.

For future screening experiments with this method, it is crucial that the impact of fungal contamination be reduced. This could be achieved by better sealing the tops of the growth pounches, spacing growth pouches further apart, adding fungicide to the rooting solution, or setting up more replicates. Moving to a more sterile agar plate or single-time-point boiling tube growth system could also avoid the problem of fungal contamination. The screen could be further extended with competitive co-inoculation experiments, in which mutant strains marked with the Ps*nifH*:*sfgfp* biosensor are co-inoculated with wild type marked with the Ps*nifH*:*mCherry* biosensor (OPS2576) in order to investigate the kinetics of competitive nodulation.

6.7 Conclusions

In conclusion, this work has explored the genetic basis for competitive nodulation and developed strategies for the investigation of nodule development in *Rhizobium leguminosarum*. This work provides the foundations for future studies which could eventually lead to the improvement of rhizobial inoculants, for example through new screening methods to identify competitive *Rhizobium* isolates for use as inoculant strains. Such improved rhizobial inoculants would result in reduced environmental harms, more robust food security and improved economic conditions for farmers.

7 Bibliography

Albert, H., Dale, E.C., Lee, E. & Ow, D.W. (1995). Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. The Plant Journal, *7*(4), 649–659.

Alunni, B. & Gourion, B. (2016). Terminal bacteroid differentiation in the legume-*Rhizobium* symbiosis: nodule-specific cysteine-rich peptides and beyond. The New Phytologist, *211*(2), 411–417.

Amarger, N. & Lobreau, J.P. (1982). Quantitative study of nodulation competitiveness in *Rhizobium* strains. Applied and Environmental Microbiology, *44*(3), 583–588.

Anyango, B., Wilson, K. & Giller, K. (1998). Competition in Kenyan soils between *Rhizobium leguminosarum* biovar *phaseoli* strain Kim5 and *R. tropici* strain CIAT899 using the *gusA* marker gene. Plant and Soil, *204*(1), 69–78.

Aroney, S. (2021). Role and regulation of chemotaxis and motility in the *Rhizobium*-legume symbiosis. (Doctoral dissertation).

Aroney, S.T.N., Poole, P.S. & Sánchez-Cañizares, C. (2021). Rhizobial chemotaxis and motility systems at work in the soil. Frontiers in Plant Science, *12*, 1–15.

Asai, T. & Kogoma, T. (1994). Roles of *ruvA*, *ruvC* and *recG* gene functions in normal and DNA damage-inducible replication of the *Escherichia coli* chromosome. Genetics, *137*(4), 895–902.

Ausmees, N., Mayer, R., Weinhouse, H., Volman, G., Amikam, D., Benziman, M. & Lindberg, M. (2001). Genetic data indicate that proteins containing the GGDEF domain possess diguanylate cyclase activity. FEMS Microbiology Letters, *204*(1), 163–167.

Baldwin, I.L. & Fred, E.B. (1929). Strain variation in the root nodule bacteria of clover, *Rhizobium trifolii*. Journal of Bacteriology, *17*(1), 17–18.

Barquist, L., Boinett, C.J. & Cain, A.K. (2013). Approaches to querying bacterial genomes with transposon-insertion sequencing. RNA Biology, *10*(7), 1161–1169.

Batista, L., Irisarri, P., Rebuffo, M., Cuitiño, M.J., Sanjuán, J. & Monza, J. (2015). Nodulation competitiveness as a requisite for improved rhizobial inoculants of Trifolium pratense. Biology and Fertility of Soils, *51*(1), 11–20.

Beattie, G.A., Clayton, M.K. & Handelsman, J. (1989). Quantitative comparison of the laboratory and field competitiveness of *Rhizobium leguminosarum* biovar *phaseoli*. Applied and Environmental Microbiology, *55*(11), 2755–2761.

Becker, A., Kleickmann, A., Keller, M., Arnold, W. & Pühler, A. (1993). Identification and analysis of the *Rhizobium meliloti exoAMONP* genes involved in exopolysaccharide biosynthesis and mapping of promoters located on the *exoHKLAMONP* fragment. Molecular & General Genetics, *241*(3-4), 367–379.

Beijerinck, M.W. (1888). Cultur des *Bacillus radicicola* aus den Knöllchen. Botanische Zeitung, *46*, 740–750.

Ben Romdhane, S., Aouani, M.E., Trabelsi, M., de Lajudie, P. & Mhamdi, R. (2008). Selection of high nitrogen-fixing rhizobia nodulating chickpea (*Cicer arietinum*) for semi-arid Tunisia. Journal of Agronomy and Crop Science, *194*, 413–420.

Benezech, C., Doudement, M. & Gourion, B. (2020). Legumes tolerance to rhizobia is not always observed and not always deserved. Cellular Microbiology, *22*(1), 1–9.

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

Berrabah, F., Ratet, P. & Gourion, B. (2019). Legume nodules: Massive infection in the absence of defense induction. Molecular Plant-Microbe Interactions, *32*(1), 35–44.

Bertani, G. (1951). Studies on lysogenesis: 1. The mode of phage liberation by lysogenic *Escherichia coli*. Journal of Bacteriology, *62*(3), 293–300.

Boivin, S., Ait Lahmidi, N., Sherlock, D., Bonhomme, M., Dijon, D., Heulin-Gotty, K., Le-Queré, A., Pervent, M., Tauzin, M., Carlsson, G., Jensen, E., Journet, E.P., Lopez-Bellido, R., Seidenglanz, M., Marinkovic, J., Colella, S., Brunel, B., Young, P. & Lepetit, M. (2020). Host-specific competitiveness to form nodules in *Rhizobium leguminosarum* symbiovar *viciae*. New Phytologist, *226*(2), 555–568.

Brockwell, J., Herridge, D.F., Roughley, R.J., Thompson, J.A. & Gault, R.R. (1975). Studies on seed pelleting as an aid to legume seed inoculation. 4. Examination of preinoculated seed. Australian Journal of Experimental Agriculture, *15*(77), 780–787.

Bullock, D.G. (1992). Crop rotation. Critical Reviews in Plant Sciences, *11*(4), 309–326.

Burén, S. & Rubio, L.M. (2018). State of the art in eukaryotic nitrogenase engineering. FEMS Microbiology Letters, *365*(2), 1–9.

Cain, A.K., Barquist, L., Goodman, A.L., Paulsen, I.T., Parkhill, J. & Van Opijnen, T. (2020). A decade of advances in transposon-insertion sequencing. Nature Reviews Genetics.

Carlson, T.A., Martin, G.B. & Chelm, B.K. (1987). Differential transcription of the two glutamine synthetase genes of *Bradyrhizobium japonicum*. Journal of Bacteriology, *169*(12), 5861–5866.

Casteriano, A., Wilkes, M.A. & Deaker, R. (2013). Physiological changes in rhizobia after growth in peat extract may be related to improved desiccation tolerance. Applied and Environmental Microbiology, *79*(13), 3998–4007.

Cebolla, A., Vinardell, J.M., Kiss, E., Oláh, B., Roudier, F., Kondorosi, A. & Kondorosi, E. (1999). The mitotic inhibitor *ccs52* is required for endoreduplication and ploidy-dependent cell enlargement in plants. EMBO Journal, *18*(16), 4476–4484.

Chang, C., Damiani, I., Puppo, A. & Frendo, P. (2009). Redox changes during the legume-*Rhizobium* symbiosis. Molecular Plant, *2*(3), 370–377.

Chang, Y.C., Hu, Z., Rachlin, J., Anton, B.P., Kasif, S., Roberts, R.J. & Steffen, M. (2016). COMBREX-DB: An experiment centered database of protein function: Knowledge, predictions and knowledge gaps. Nucleic Acids Research, *44*(D1), D330–D335.

Chao, M.C., Abel, S., Davis, B.M. & Waldor, M.K. (2016). The design and analysis of transposon insertion sequencing experiments. Nature Reviews Microbiology, *14*(2), 119–128.

Cherfils, J., Gibrat, J.-F..-., Levin, J., Batut, J. & Kahn, D. (1989). Model-building of Fnr and FixK DNA-binding domains suggests a basis for specific DNA recognition. Journal of Molecular Recognition, *2*(3), 114–121.

Choi, K.-h. & Schweizer, H.P. (2006). Mini-Tn*7* insertion in bacteria with single *att* Tn*7* sites: example *Pseudomonas aeruginosa*. Nature Protocols, *1*(1), 153–161.

Chomicki, G., Kiers, E.T. & Renner, S.S. (2020). The evolution of mutualistic dependence. Annual Review of Ecology, Evolution, and Systematics, (51), 409–432.

Clayton, G.W., Rice, W.A., Lupwayi, N.Z., Johnston, A.M., Lafond, G.P., Grant, C.A. & Walley, F.L. (2004). Inoculant formulation and fertilizer nitrogen effects on field pea : Crop yield and seed quality. Canadian Journal of Plant Science, *84*(1), 79–88.

Clayton, S.J., Read, D.B., Murray, P.J. & Gregory, P.J. (2008). Exudation of alcohol and aldehyde sugars from roots of defoliated lolium perenne l. grown under sterile conditions. Journal of Chemical Ecology, *34*(11), 1411–1421.

Clúa, J., Roda, C., Zanetti, M.E. & Blanco, F.A. (2018). Compatibility between legumes and rhizobia for the establishment of a successful nitrogen-fixing symbiosis. Genes, *9*(3).

Crook, M.B., Lindsay, D.P., Biggs, M.B., Bentley, J.S., Price, J.C., Clement, S.C., Clement, M.J., Long, S.R. & Griffitts, J.S. (2012). Rhizobial plasmids that cause impaired symbiotic nitrogen fixation and enhanced host invasion. Molecular Plant-Microbe Interactions, *25*(8), 1026–1033.

Davidson, A.L., Dassa, E., Orelle, C. & Chen, J. (2008). Structure, function, and evolution of bacterial ATP-binding cassette systems. Microbiology and Molecular Biology Reviews, *72*(2), 317–364.

Deaker, R., Roughley, R.J. & Kennedy, I.R. (2004). Legume seed inoculation technology: a review. Soil Biology and Biochemistry, *36*(8), 1275–1288.

Deaker, R., Roughley, R.J. & Kennedy, I.R. (2007). Desiccation tolerance of rhizobia when protected by synthetic polymers. Soil Biology and Biochemistry, *39*(2), 573–580.

DeJesus, M.A., Ambadipudi, C., Baker, R., Sassetti, C. & Ioerger, T.R. (2015). TRANSIT - a software tool for Himar1 TnSeq analysis. PLoS Computational Biology, *11*(10), 1–17.

Denton, M.D., Reeve, W.G., Howieson, J.G. & Coventry, D.R. (2003). Competitive abilities of common field isolates and a commercial strain of *Rhizobium leguminosarum* bv. *trifolii* for clover nodule occupancy. Soil Biology and Biochemistry, *35*(8), 1039–1048.

de Philip, P., Batut, J. & Boistard, P. (1990). *Rhizobium meliloti* FixL is an oxygen sensor and regulates *R. meliloti nifA* and *fixK* genes differently in *Escherichia coli*. Journal of Bacteriology, *172*(8), 4255–4262.

Dowling, D.N. & Broughton, W.J. (1986). Competition for nodulation of legumes. Annual Review of Microbiology, *40*, 131–157.

Downie, J.A. (2010). The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. FEMS Microbiology Reviews, *34*(2), 150–170.

Drew, E., Herridge, D.F., Ballard, R., O'Hara, G.W., Deaker, R., Denton, M.D., Yates, R.J., Gemell, G., Hartley, E. & Phillips, L.A. (2012). *Inoculating legumes: a practical guide*. Grains Research; Development Corporation.

Duncan, M.J. (1981). Properties of Tn*5*-induced carbohydrate mutants in *Rhizobium meliloti*. Journal of General Microbiology, *122*(1), 61–67.

Dunderdale, H.J., Benson, F.E., Parsons, C.A., Sharpies, G.J., Lloyd, R.G. & West, S.C. (1991). Formation and resolution of recombination intermediates by *E. coli* RecA and RuvC proteins. Nature, *354*(6354), 506–510.

Dunham, D.H. & Baldwin, I.L. (1931). Double infection of leguminous plants with good and poor strains of rhizobia. Soil Science, *32*(3), 235.

Dwivedi, S.L., Sahrawat, K.L., Upadhyaya, H.D., Mengoni, A., Galardini, M., Bazzicalupo, M., Biondi, E.G., Hungria, M., Kaschuk, G., Blair, M.W. & Ortiz, R. (2015).

Advances in host plant and *Rhizobium* genomics to enhance symbiotic nitrogen fixation in grain legumes. Advances in Agronomy, *129*, 1–116.

Earl, C.D., Ronson, C.W. & Ausubel, F.M. (1987). Genetic and structural analysis of the *Rhizobium meliloti fixA*, *fixB*, *fixC*, and *fixX* genes. Journal of Bacteriology, *169*(3), 1127–1136.

Economou, A., Hawkins, F.K., Downie, J.A. & Johnston, A.W.B. (1989). Transcription of *rhiA*, a gene on a *Rhizobium leguminosarum* bv. *viciae* Sym plasmid, requires *rhiR* and is repressed by flavanoids that induce nod genes. Molecular Microbiology, *3*(1), 87–93.

Engler, C., Kandzia, R. & Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS ONE, *3*(11).

Fabian, B.K., Tetu, S.G. & Paulsen, I.T. (2020). Application of transposon insertion sequencing to agricultural science. Frontiers in Plant Science, *11*.

Fan, J., Crooks, C. & Lamb, C. (2008). High-throughput quantitative luminescence assay of the growth in planta of *Pseudomonas syringae* chromosomally tagged with *Photorhabdus luminescens luxCDABE*. Plant Journal, *53*(2), 393–399.

Ferguson, B.J., Mens, C., Hastwell, A.H., Zhang, M., Su, H., Jones, C.H., Chu, X. & Gresshoff, P.M. (2019). Legume nodulation: The host controls the party. Plant Cell and Environment, *42*(1), 41–51.

Fowler, D., Coyle, M., Skiba, U., Sutton, M.A., Cape, J.N., Reis, S., Sheppard, L.J., Jenkins, A., Grizzetti, B. & Galloway, J.N. (2013). The global nitrogen cycle in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences, *368*(1621), 20130164.

Franssen, H.J., Vijn, I., Yang, W.C. & Bisseling, T. (1992). Developmental aspects of the *Rhizobium-legume* symbiosis. Plant Molecular Biology, *19*(1), 89–107.

Fred, E.B., Baldwin, I.L. & McCoy, E. (1932). *Root nodule bacteria and leguminous plants*. UW-Madison Libraries Parallel Press.

Frederix, M. (2010). A novel mechanism of coupling quorum sensing systems in *Rhizobium leguminosarum* bv. *viciae* 3841. (Doctoral dissertation).

Frederix, M., Edwards, A., Mcanulla, C. & Downie, J.A. (2011). Co-ordination of quorum-sensing regulation in *Rhizobium leguminosarum* by induction of an anti-repressor. Molecular Microbiology, *81*(4), 994–1007.

Frederix, M., Edwards, A., Swiderska, A., Stanger, A., Karunakaran, R., Williams, A., Abbruscato, P., Sanchez-Contreras, M., Poole, P.S. & Downie, J.A. (2014). Mutation of *praR* in *Rhizobium leguminosarum* enhances root biofilms, improving nodulation competitiveness by increased expression of attachment proteins. Molecular Microbiology, *93*(3).

Frey, S.D. & Blum, L.K. (1994). Effect of pH on competition for nodule occupancy by type I and type II strains of *Rhizobium leguminosarum* bv. *phaseoli*. Plant and Soil, *163*(2), 157–164.

Fry, J., Wood, M. & Poole, P.S. (2007). Investigation of *myo*-Inositol Catabolism in *Rhizobium leguminosarum* bv. *viciae* and Its Effect on Nodulation Competitiveness. Molecular Plant-Microbe Interactions, *14*(8), 1016–1025.

Galloway, J.N., Townsend, A.R., Erisman, J.W., Bekunda, M., Cai, Z., Freney, J.R., Martinelli, L.A., Seitzinger, S.P. & Sutton, M.A. (2008). Transformation of the Nitrogen Cycle. Science, *320*(May), 889–892.

Galperin, M.Y., Nikolskaya, A.N. & Koonin, E.V. (2001). Novel domains of the prokaryotic two-component signal transduction systems. FEMS Microbiology Letters, *203*(1), 11–21.

Gawronski, J.D., Wong, S.M.S., Giannoukos, G., Ward, D.V. & Akerley, B.J. (2009). 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, *106*(38), 16422–16427.

Geddes, B.A., Mendoza-Suárez, M. & Poole, P.S. (2019). A Bacterial Expression Vector Archive (BEVA) for flexible modular assembly of Golden Gate-compatible vectors. Frontiers in Microbiology, *9*(January), 392423.

Geurts, R. & Bisseling, T. (2002). *Rhizobium* nod factor perception and signalling. The Plant Cell Online, 239–250.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A. & Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods, *6*(5), 343–345.

Giller, K.E. & Wilson, K. (2001). *Nitrogen fixation in tropical cropping systems*. CAB International.

González-Pasayo, R. & Martínez-Romero, E. (2000). Multiresistance genes of *Rhizobium etli* CFN42. Molecular Plant-Microbe Interactions, *13*(5), 572–577.

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

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

Graham, P.H. (1969). Selective medium for growth of *Rhizobium*. Applied Microbiology, *17*(5), 769–770.

Graham, P.H. & Vance, C.P. (2003). Legumes: Importance and constraints to greater use. American Society of Plant Biologists, *131*(3), 872–877.

Green, R.T., East, A.K., Karunakaran, R., Downie, J.A. & Poole, P.S. (2019). Transcriptomic analysis of *Rhizobium leguminosarum* bacteroids in determinate and indeterminate nodules. Microbial Genomics, *5*(2), 1–16.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. Journal of Molecular Biology, *166*(4), 557–580.

Haskett, T.L., Tkacz, A. & Poole, P.S. (2021). Engineering rhizobacteria for sustainable agriculture. ISME Journal, *15*(4), 949–964.

Hassett, D.J., Schweizer, H.P. & Ohman, D.E. (1995). *Pseudomonas aeruginosa soda* and *sodb* mutants defective in manganese- and iron-cofactored superoxide dismutase activity demonstrate the importance of the iron-cofactored form in aerobic metabolism. Journal of Bacteriology, *177*(22), 6330–6337.

Hearn, E.M., Patel, D.R. & van den Berg, B. (2008). Outer-membrane transport of aromatic hydrocarbons as a first step in biodegradation. Proceedings of the National Academy of Sciences, *105*(25), 8601–8606.

Hellriegel, H. & Wilfarth, H. (1888). Untersuchungen uber die Stickstoffnahrung der Gramineen und Leguminosen. Buchdruckerei der Post Kayssler & Co.

Helz, G.E., Baldwin, I.L. & Fred, E.B. (1927). Strain variations and host specificity of the root-nodule bacteria of the pea group. Journal of Agricultural Research, *35*(11), 1039–1055.

Higashi, S., Katahira, S. & Abe, M. (1984). Isolation of pea nodule bacteroids utilizing silica sol (Percoll) density gradient centrifugation. Plant and Soil, *81*(1), 93–99.

Hirsch, A.M. & Smith, C.A. (1987). Effects of *Rhizobium meliloti nif* and *fix* mutants on alfalfa root nodule development. Journal of Bacteriology, *169*(3), 1137–1146.

Holland, A.A. (1966). Serologic characteristics of certain root-nodule bacteria of legumes. Antonie van Leeuwenhoek, *32*(1), 410–418.

Hu, K., Li, Y., Chen, W., Chen, D., Wei, Y., Edis, R., Li, B., Huang, Y. & Zhang, Y. (2010). Modeling nitrate leaching and optimizing water and nitrogen management under irrigated maize in desert oases in Northwestern China. Journal of Environmental Quality, *39*(2), 667–677.

Iwasaki, H., Takahagi, M., Shiba, T., Nakata, A. & Shinagawa, H. (1991). *Escherichia coli* ruvc protein is an endonuclease that resolves the holliday structure. The EMBO Journal, *10*(13), 4381–4389.

Iyer, L.M., Leipe, D.D., Koonin, E.V. & Aravind, L. (2004). Evolutionary history and higher order classification of AAA+ ATPases. Journal of Structural Biology, *146*(1-2), 11–31.

Janczarek, M., Jaroszuk-Ściseł, J. & Skorupska, A. (2009). Multiple copies of *rosR* and *pssA* genes enhance exopolysaccharide production, symbiotic competitiveness and clover nodulation in *Rhizobium leguminosarum* bv. *trifolii*. Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, *96*(4), 471–486.

Janczarek, M., Rachwał, K., Marzec, A., Grzadziel, J. & Palusińska-Szysz, M. (2015). Signal molecules and cell-surface components involved in early stages of the legume-*Rhizobium* interactions. Applied Soil Ecology, *85*, 94–113.

Janczarek, M. & Skorupska, A. (2007). The *Rhizobium leguminosarum* bv. *trifolii* RosR: Transcriptional regulator involved in exopolysaccharide production. Molecular Plant-Microbe Interactions, *20*(7), 867–881.

Jangid, K., Williams, M.A., Franzluebbers, A.J., Sanderlin, J.S., Reeves, J.H., Jenkins, M.B., Endale, D.M., Coleman, D.C. & Whitman, W.B. (2008). Relative impacts of land-use, management intensity and fertilization upon soil microbial community structure in agricultural systems. Soil Biology and Biochemistry, *40*(11), 2843–2853.

Jefferson, R.A., Burgess, S.M. & Hirsh, D. (1986). β-Glucuronidase from *Escherichia coli* as a gene-fusion marker. Proceedings of the National Academy of Sciences, *83*(22), 8447–8451.

Jenal, U. (2004). Cyclic di-guanosine-monophosphate comes of age: A novel secondary messenger involved in modulating cell surface structures in bacteria? Current Opinion in Microbiology, *7*(2), 185–191.

Jensen, E.S., Peoples, M.B., Boddey, R.M., Gresshoff, P.M., Henrik, H.N., Alves, B.J.R. & Morrison, M.J. (2012). Legumes for mitigation of climate change and the provision of feedstock for biofuels and biorefineries. A review. Agronomy for Sustainable Development, *32*(2), 329–364.

Ji, Z.J., Yan, H., Cui, Q.G., Wang, E.T., Chen, W.F. & Chen, W.X. (2017). Competition between rhizobia under different environmental conditions affects the nodulation of a legume. Systematic and Applied Microbiology, *40*(2), 114–119.

Jiang, S., Jardinaud, M.F., Gao, J., Pecrix, Y., Wen, J., Mysore, K., Xu, P., Sanchez-Canizares, C., Ruan, Y., Li, Q., Zhu, M., Li, F., Wang, E., Poole, P.S., Gamas, P. & Murray, J.D. (2021). NIN-like protein transcription factors regulate leghemoglobin genes in legume nodules. Science, *374*(6567), 625–628.

Jiménez-Guerrero, I., Acosta-Jurado, S., del Cerro, P., Navarro-Gómez, P., López-Baena, F.J., Ollero, F.J., Vinardell, J.M. & Pérez-Montaño, F. (2018). Transcriptomic studies of the effect of nod gene-inducing molecules in rhizobia: Different weapons, one purpose. Genes, *9*(1), 1–27.

Johnston, A.W.B. & Beringer, J.E. (1975). Identification of the *Rhizobium* strains in pea root nodules using genetic markers. Journal of General Microbiology, *87*(2), 343–350.

Jones, K.M., Kobayashi, H., Davies, B.W., Taga, M.E. & Walker, G.C. (2007). How rhizobial symbionts invade plants: the *Sinorhizobium-Medicago* model. Nature Reviews Microbiology, *5*(8), 619–633.

Jorrín, B. (2016). Genomics of specificity in the symbiotic interaction between *Rhizobium leguminosarum* and legumes. (Doctoral dissertation).

Josey, D.P., Beynon, J.L., Johnston, A.W.B. & Beringer, J.E. (1979). Strain identification in *Rhizobium* using intrinsic antibiotic resistance. Journal of Applied Bacteriology, *46*(2), 343–350.

Kamicker, B.J. & Brill, W.J. (1986). Identification of *Bradyrhizobium japonicum* nodule isolates from wisconsin soybean farms. Applied and Environmental Microbiology, *51*(3), 487–492.

Kaniga, K., Delor, I. & Cornelis, G.R. (1991). A wide-host-range suicide vector for improving reverse genetics in Gram-negative bacteria: inactivation of the *blaA* gene of *Yersinia enterocolitica*. Gene, *109*(1), 137–141.

Kapuscinski, J. (1995). DAPI: A DNA-Specific fluorescent probe. Biotechnic and Histochemistry, *70*(5), 220–233.

Karunakaran, R., Haag, A.F., East, A.K., Ramachandran, V.K., Prell, J., James, E.K., Scocchi, M., Ferguson, G.P. & Poole, P.S. (2010). BacA is essential for bacteroid development in nodules of galegoid, but not phaseoloid, legumes. Journal of Bacteriology, *192*(11), 2920–2928.

Karunakaran, R., Ramachandran, V.K., Seaman, J.C., East, A.K., Mouhsine, B., Mauchline, T.H., Prell, J., Skeffington, A. & Poole, P.S. (2009). Transcriptomic analysis of *Rhizobium leguminosarum* biovar *viciae* in symbiosis with host plants *Pisum sativum* and *Vicia cracca*. Journal of Bacteriology, *191*(12), 4002–4014.

Kiers, E.T., Rousseau, R.A., West, S.A. & Denison, R.F. (2003). Host sanctions and the legume–*Rhizobium* mutualism. Nature, *425*, 78.

Kiers, E.T. & Denison, R.F. (2008). Sanctions, cooperation, and the stability of plant-rhizosphere mutualisms. Annual Review of Ecology, Evolution, and Systematics, *39*, 215–236.

Knights, H.E., Jorrin, B., Haskett, T.L. & Poole, P.S. (2021). Deciphering bacterial mechanisms of root colonization. Environmental Microbiology Reports, *13*(4), 428–444.

Kohler, P.R.A., Choong, E.-l. & Rossbach, S. (2011). The RpiR-like repressor IolR regulates inositol catabolism in *Sinorhizobium meliloti*. Journal of Bacteriology, *193*(19), 5155–5163.

Kohler, P.R.A., Zheng, J.Y., Schoffers, E. & Rossbach, S. (2010). Inositol catabolism, a key pathway in *Sinorhizobium meliloti* for competitive host nodulation. Applied and Environmental Microbiology, *76*(24), 7972–7980.

Krehenbrink, M., Edwards, A. & Downie, J.A. (2011). The superoxide dismutase SodA is targeted to the periplasm in a SecA-dependent manner by a novel mechanism. Molecular Microbiology, *82*(1), 164–179.

Król, J., Wielbo, J., Mazur, A., Kopcińska, J., Lotocka, B., Golinowski, W. & Skorupska, A. (1998). Molecular characterization of *pssCDE* genes of *Rhizobium leguminosarum* bv. *trifolii* strain TA1: *pssD* mutant is affected in exopolysaccharide synthesis and endocytosis of bacteria. Molecular Plant-Microbe Interactions, *11*(11), 1142–1148.

Kuc, J. (1995). Phytoalexins, stress metabolism, and disease resistance in plants. Annual Review of Phytopathology, *33*(1), 275–297.

Lambert, I., Pervent, M., Le Queré, A., Clément, G., Tauzin, M., Severac, D., Benezech, C., Tillard, P., Martin-Magniette, M.L., Colella, S. & Lepetit, M. (2020). Responses of mature symbiotic nodules to the whole-plant systemic nitrogen signaling. Journal of Experimental Botany, *71*(16), 5039–5052.

Land, M., Hauser, L., Jun, S.-R., Nookaew, I., Leuze, M.R., Ahn, T.-H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S. & Ussery, D.W. (2015). Insights from 20 years of bacterial genome sequencing. Functional & Integrative Genomics, *15*(2), 141–161.

Lawlis, V.B., Dennis, M.S., Chen, E.Y., Smith, D.H. & Henner, D.J. (1984). Cloning and sequencing of the xylose isomerase and xylulose kinase genes of *Escherichia coli*. Applied and Environmental Microbiology, *47*(1), 15–21.

Ledermann, R., Schulte, C.C.M. & Poole, P.S. (2021). How rhizobia adapt to the nodule environment. Journal of Bacteriology, *203*(12).

Lee, K.H. & LaRue, T.A. (1992). Exogenous ethylene inhibits nodulation of *Pisum sativum* L. cv *Sparkle*. Plant Physiology, *100*(4), 1759–1763.

Leonard, L.T. (1930). A failure of Austrian winter peas apparently due to nodule bacteria. Journal of the American Society of Agronomy, *22*, 277–280.

Liu, C.W. & Murray, J.D. (2016). The role of flavonoids in nodulation host-range specificity: An update. Plants, *5*(3), 4045–4053.

Loewus, F.A. (1952). Improvement in anthrone method for determination of carbohydrates. Analytical Chemistry, *24*(1), 219.

Löhnis, M.P. (1930). Can *Bacterium radicicola* assimilate nitrogen in the absence of the host plant? Soil Science, *29*(1), 37–57.

López-García, S.L., Vázquez, T.E.E., Favelukes, G. & Lodeiro, A.R. (2002). Rhizobial position as a main determinant in the problem of competition for nodulation in soybean. Environmental Microbiology, *4*(4), 216–224.

Lupwayi, N.Z., Clayton, G.W. & Rice, W.A. (2006). Rhizobial inoculants for legume crops. Journal of Crop Improvement, *15*(2), 97–125.

Maier, R.J. & Brill, W.J. (1978). Mutant strains of *Rhizobium japonicum* with increased ability to fix nitrogen for soybean. Science, *201*(4354), 448–450.

Masson-Boivin, C. & Sachs, J.L. (2018). Symbiotic nitrogen fixation by rhizobia — the roots of a success story. Current Opinion in Plant Biology, *44*, 7–15.

May, S.N. & Bohlool, B.B. (1983). Competition among *Rhizobium leguminosarum* strains for nodulation of lentils (*Lens esculenta*). Applied and Environmental Microbiology, *45*(3), 960–965.

McCord, J.M. & Fridovich, I. (1969). Superoxide dismutase - an enzymic function for erythrocuprein (hemocuprein). The Journal of Biological Chemistry, *244*(25), 1969–6049.

Mcintyre, H.J., Davies, H., Hore, T.A., Miller, S.H., Dufour, J.-p. & Ronson, C.W. (2007). Trehalose biosynthesis in *Rhizobium leguminosarum* bv . *trifolii* and its role in desiccation tolerance. Applied and Environmental Microbiology, *73*(12), 3984–3992.

Mendoza-Suárez, M.A., Andersen, S.U., Poole, P.S. & Sánchez-Cañizares, C. (2021). Competition, nodule occupancy, and persistence of inoculant strains: key factors in the *Rhizobium*-legume symbioses. Frontiers in Plant Science, *12*(August).

Mendoza-Suárez, M.A., Geddes, B.A., Sánchez-Cañizares, C., Ramírez-González, R.H., Kirchhelle, C., Jorrín, B. & Poole, P.S. (2020). Optimizing *Rhizobium*-legume symbioses by simultaneous measurement of rhizobial competitiveness and N_2 fixation in nodules. Proceedings of the National Academy of Sciences, *117*(18), 9822–9831.

Mendoza-Suárez, M.A. (2018). A molecular toolkit for high-throughput identification of competitive and effective rhizobial strains. (Doctoral dissertation).

Mergaert, P., Uchiumi, T., Alunni, B., Evanno, G., Cheron, A., Catrice, O., Mausset, A.-E., Barloy-Hubler, F., Galibert, F., Kondorosi, A. & Kondorosi, E. (2006). Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. Proceedings of the National Academy of Sciences, *103*(13), 5230–5235.

Miklashevichs, E., Röhrig, H., Schell, J. & Schmidt, J. (2001). Perception and signal transduction of rhizobial NOD factors. Critical Reviews in Plant Sciences, *20*(4), 373–394.

Miller, L.D., Yost, C.K., Hynes, M.F. & Alexandre, G. (2007). The major chemotaxis gene cluster of *Rhizobium leguminosarum* bv. *viciae* is essential for competitive nodulation. Molecular Microbiology.

Morgan, R.D., Dwinell, E.A., Bhatia, T.K., Lang, E.M. & Luyten, Y.A. (2009). The *Mme*I family: Type II restriction-modification enzymes that employ single-strand modification for host protection. Nucleic Acids Research, *37*(15), 5208–5221.

Mueller, N.D., Gerber, J.S., Johnston, M., Ray, D.K., Ramankutty, N. & Foley, J.A. (2012). Closing yield gaps through nutrient and water management. Nature, *490*(7419), 254–257.

Mulley, G., White, J.P., Karunakaran, R., Prell, J., Bourdes, A., Bunnewell, S., Hill, L. & Poole, P.S. (2011). Mutation of GOGAT prevents pea bacteroid formation and N_2-fixation by globally downregulating transport of organic nitrogen sources. Molecular Microbiology, *80*(1), 149–167.

Naeem, F.I., Ashraf, M.M., Malik, K.A. & Hafeez, F.Y. (2004). Competitiveness of introduced *Rhizobium* strains for nodulation in fodder legumes. Pakistan Journal of Botany, *1*(36), 159–166.

Nagata, M., Kaito, C. & Sekimizu, K. (2008). Phosphodiesterase activity of CvfA is required for virulence in *Staphylococcus aureus*. Journal of Biological Chemistry, *283*(4), 2176–2184.

Naik, K., Mishra, S., Srichandan, H., Singh, P.K. & Sarangi, P.K. (2019). Plant growth promoting microbes: Potential link to sustainable agriculture and environment. Biocatalysis and Agricultural Biotechnology, *21*(July), 101326.

Nukui, N., Ezura, H., Yuhashi, K.I., Yasuta, T. & Minamisawa, K. (2000). Effects of ethylene precursor and inhibitors for ethylene biosynthesis and perception on nodulation

in *Lotus japonicus* and *Macroptilium atropurpureum*. Plant and Cell Physiology, *41*(7), 893–897.

O'Hara, G.W. (1998). The role of nitrogen fixation in crop production. Nutrient Use in Crop Production, 115–138.

Okazaki, S., Nukui, N., Sugawara, M. & Minamisawa, K. (2004). Rhizobial strategies to enhance symbiotic interactions: Rhizobitoxine and 1-aminocyclopropane-1-carboxylate deaminase. Microbes and Environments, *19*(2), 99–111.

Oldroyd, G.E.D. & Dixon, R. (2014). Biotechnological solutions to the nitrogen problem. Current Opinion in Biotechnology, *26*, 19–24.

Oldroyd, G.E.D. & Downie, J.A. (2004). Calcium, kinases and nodulation signalling in legumes. Nature Reviews Molecular Cell Biology, *5*(7), 566–576.

Onishchuk, O.P., Vorobyov, N.I. & Provorov, N.A. (2017). Nodulation competitiveness of nodule bacteria: Genetic control and adaptive significance: Review. Applied Biochemistry and Microbiology, *53*(2), 131–139.

Ott, T., Van Dongen, J.T., Günther, C., Krusell, L., Desbrosses, G., Vigeolas, H., Bock, V., Czechowski, T., Geigenberger, P. & Udvardi, M.K. (2005). Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. Current Biology, *15*(6), 531–535.

Paau, A.S. (1989). Improvement of *Rhizobium* inoculants by mutation, genetic engineering and formulation. Biotechnology Advances, *9*(2), 173–184.

Pan, H. & Wang, D. (2017). Nodule cysteine-rich peptides maintain a working balance during nitrogen-fixing symbiosis. Nature Plants, *3*(May).

Pankievicz, V.C.S., Irving, T.B., Maia, L.G.S. & Ané, J.M. (2019). Are we there yet? The long walk towards the development of efficient symbiotic associations between nitrogen-fixing bacteria and non-leguminous crops. BMC Biology, *17*(1).

Pardo-Díaz, J. (2021). Bioinformatics tools for the study of the changes in the gene expression of *Rhizobium leguminosarum*. (Doctoral dissertation).

Pardo-Díaz, J., Bozhilova, L.V., Beguerisse-Díaz, M., Poole, P.S., Deane, C.M. & Reinert, G. (2021). Robust gene coexpression networks using signed distance correlation. Bioinformatics, *37*(14), 1982–1989.

Parnell, J.J., Berka, R., Young, H.A., Sturino, J.M., Kang, Y., Barnhart, D.M. & Dileo, M.V. (2016). From the lab to the farm: An industrial perspective of plant beneficial microorganisms. Frontiers in Plant Science, *7*(AUG2016), 1–12.

Parsons, J. (2019). Characterising root attachment in *Rhizobium*-legume symbioses. (Doctoral dissertation).

Pastor-Bueis, R., Sánchez-Cañizares, C., James, E.K. & González-Andrés, F. (2019). Formulation of a highly effective inoculant for common bean based on an autochthonous elite strain of *Rhizobium leguminosarum* bv. *phaseoli*, and genomic-based insights into its agronomic performance. Frontiers in Microbiology, *10*.

Pédelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C. & Waldo, G.S. (2006). Engineering and characterization of a superfolder green fluorescent protein. Nature Biotechnology, *24*(1), 79–88.

Peoples, M.B., Giller, K.E., Jensen, E.S. & Herridge, D.F. (2021). Quantifying country to global scale nitrogen fixation for grain legumes: I. Reliance on nitrogen fixation of soybean, groundnut and pulses. Plant and Soil, (0123456789).

Perry, B.J., Akter, M.S. & Yost, C.K. (2016). The use of transposon insertion sequencing to interrogate the core functional genome of the legume symbiont *Rhizobium leguminosarum*. Frontiers in Microbiology.

Perry, B.J. & Yost, C.K. (2014). Construction of a mariner-based transposon vector for use in insertion sequence mutagenesis in selected members of the Rhizobiaceae. BMC Microbiology, *14*(1), 1–11.

Philippot, L., Raaijmakers, J.M., Lemanceau, P. & van der Putten, W.H. (2013). Going back to the roots: the microbial ecology of the rhizosphere. Nature reviews. Microbiology, *11*, 789–799.

Pini, F., East, A.K., Appia-Ayme, C., Tomek, J., Karunakaran, R., Mendoza-Suárez, M., Edwards, A., Terpolilli, J.J., Roworth, J., Downie, J.A. & Poole, P.S. (2017). Bacterial biosensors for in vivo spatiotemporal mapping of root secretion. Plant Physiology, *174*(3), 1289–1306.

Plasterk, R.H.A., Izsvák, Z. & Ivics, Z. (1999). Resident aliens: The Tc1/mariner superfamily of transposable elements. Trends in Genetics, *15*(8), 326–332.

Poole, P.S., Blyth, A., Reid, C.J. & Walters, K. (1994). *Myo*-inositol catabolism and catabolite regulation in *rhizobium leguminosarum* bv. *viciae*. Microbiology, *140*(10), 2787–2795.

Poole, P.S., Ramachandran, V.K. & Terpolilli, J.J. (2018). Rhizobia: From saprophytes to endosymbionts. Nature Reviews Microbiology, *16*(5), 291–303.

Prell, J. & Poole, P.S. (2006). Metabolic changes of rhizobia in legume nodules. Trends in Microbiology, *14*(4), 161–168.

Ramachandran, V.K., East, A.K., Karunakaran, R., Downie, J.A. & Poole, P.S. (2011). Adaptation of *Rhizobium leguminosarum* to pea, alfalfa and sugar beet rhizo-

spheres investigated by comparative transcriptomics. Genome Biology, *12*(10), R106.

Ramos, J.L., Martinez-Bueno, M., Molina-Henares, A.J., Teran, W., Watanabe, K., Zhang, X., Trinidad Gallegos, M., Brennan, R. & Tobes, R. (2005). The TetR family of transcriptional repressors. Microbiology and Molecular Biology Reviews, *69*(2), 326–356.

Rascio, N. & La Rocca, N. (2013). Biological nitrogen fixation. *Encyclopedia of ecology* (2e, pp. 264–279).

Ried, J.L. & Collmer, A. (1987). An *nptI-sacB-sacR* cartridge for constructing directed, unmarked mutations in Gram-negative bacteria by marker exchange-eviction mutagenesis. Gene, *57*(2-3), 239–246.

Robleto, E.A., Kmiecik, K., Oplinger, E.S., Nienhuis, J. & Triplett, E.W. (1998). Trifolitoxin production increases nodulation competitiveness of *Rhizobium etli* CE3 under agricultural conditions. Applied and Environmental Microbiology, *64*(7), 2630–2633.

Robleto, E.A., Scupham, A.J. & Triplett, E.W. (1997). Trifolitoxin production in *Rhizobium etli* Strain CE3 increases competitiveness for rhizosphere colonization and root nodulation of *Phaseolus vulgaris* in Soil. Molecular Plant-Microbe Interactions, *10*(2), 228–233.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F.S., Lambin, E., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H.J., Nykvist, B., de Wit, C.A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P.K., Costanza, R., Svedin, U., ... Foley, J. (2009). Planetary boundaries: Exploring the safe operating space for humanity. Ecology and Society, *14*(2).

Rohwer, F. & Azam, F. (2000). Detection of DNA damage in prokaryotes by terminal deoxyribonucleotide transferase-mediated dUTP nick end labeling. Applied and Environmental Microbiology, *66*(3), 1001–1006.

Roughley, R.J. & Vincent, J.M. (1967). Growth and survival of *Rhizobium* spp. in peat culture. Journal of Applied Bacteriology, *30*, 362–376.

Rovira, A.D. (1956). Plant root excretions in relation to the rhizosphere effect - 1. The nature of root exudate from oats and peas. Plant and Soil, *7*(2), 178–194.

Roworth, J.D. (2017). The characterisation of root exudation and colonisation in the rhizosphere of land plants. (Doctoral dissertation).

Royet, K., Parisot, N., Rodrigue, A., Gueguen, E. & Condemine, G. (2019). Identification by Tn-seq of *Dickeya dadantii* genes required for survival in chicory plants. Molecular Plant Pathology, *20*(2), 287–306.

Rutten, P. & Poole, P.S. (2019). Oxygen regulatory mechanisms of nitrogen fixation in rhizobia. Advances in microbial physiology, *75*, 325–389.

Sachs, J.L., Quides, K.W. & Wendlandt, C.E. (2018). Legumes versus rhizobia: a model for ongoing conflict in symbiosis. New Phytologist, *219*(4), 1199–1206.

Sánchez-Cañizares, C. & Palacios, J. (2013). Construction of a marker system for the evaluation of competitiveness for legume nodulation in *Rhizobium* strains. Journal of Microbiological Methods, *92*(3), 246–249.

Santos, M.S., Nogueira, M.A. & Hungria, M. (2019). Microbial inoculants: reviewing the past, discussing the present and previewing an outstanding future for the use of beneficial bacteria in agriculture. AMB Express, *9*(1).

Santos, R., Hérouart, D., Sigaud, S., Touati, D. & Puppo, A. (2001). Oxidative burst in alfalfa-*Sinorhizobium meliloti* symbiotic interaction. Molecular Plant-Microbe Interactions, *14*(1), 86–89.

Schäfer, A., Tauch, A., Jäger, W., Kalinowski, J., Thierbach, G. & Pühler, A. (1994). Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. Gene, *145*, 69–73.

Schulte, C.C.M., Borah, K., Wheatley, R.M., Terpolilli, J.J., Saalbach, G., Crang, N., de Groot, D.H., Ratcliffe, R.G., Kruger, N.J., Papachristodoulou, A. & Poole, P.S. (2021). Metabolic control of nitrogen fixation in *Rhizobium*-legume symbioses. Science Advances, *7*(31), 1–12.

Schwenke, G.D., Herridge, D.F., Scheer, C., Rowlings, D.W., Haigh, B.M. & McMullen, K.G. (2015). Soil N_2O emissions under N_2-fixing legumes and N-fertilised canola: A reappraisal of emissions factor calculations. Agriculture, Ecosystems and Environment, *202*, 232–242.

Seefeldt, L.C., Hoffman, B.M. & Dean, D.R. (2009). Mechanism of Mo-dependent nitrogenase. Annual Review of Biochemistry, *78*, 701–722.

Sessitsch, A., Hardarson, G., De Vos, W.M. & Wilson, K.J. (1998). Use of marker genes in competition studies of *Rhizobium*. Plant and Soil, *204*(1), 35–45.

Sessitsch, A., Wilson, K.J., Akkermans, A.D.L. & De Vos, W.M. (1996). Simultaneous detection of different *Rhizobium* strains marked with either the *Escherichia coli gusA* gene or the *Pyrococcus furiosus celB* gene. Applied and Environmental Microbiology, *62*(11), 4191–4194.

Shah, V.K., Davis, L.C. & Brill, W.J. (1975). Nitrogenase. VI. Acetylene reduction assay: Dependence of nitrogen fixation estimates on component ratio and acetylene concentration. BBA - Enzymology, *384*(2), 353–359.

Simm, R., Morr, M., Kader, A., Nimtz, M. & Römling, U. (2004). GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessibility to motility. Molecular Microbiology, *53*(4), 1123–1134.

Simonsen, A.K., Dinnage, R., Barrett, L.G., Prober, S.M. & Thrall, P.H. (2017). Symbiosis limits establishment of legumes outside their native range at a global scale. Nature Communications, *8*, 1–9.

Singer, J.T., Phennicie, R.T., Sullivan, M.J., Porter, L.A., Shaffer, V.J. & Kim, C.H. (2010). Broad-host-range plasmids for red fluorescent protein labeling of gram-negative bacteria for use in the zebrafish model system. Applied and Environmental Microbiology, *76*(11), 3467–3474.

Smith, C., Hill, A.K. & Torrente-Murciano, L. (2020). Current and future role of Haber-Bosch ammonia in a carbon-free energy landscape. Energy and Environmental Science, *13*(2), 331–344.

Smýkal, P., Coyne, C.J., Ambrose, M.J., Maxted, N., Schaefer, H., Blair, M.W., Berger, J., Greene, S.L., Nelson, M.N., Besharat, N., Vymyslický, T., Toker, C., Saxena, R.K., Roorkiwal, M., Pandey, M.K., Hu, J., Li, Y.H., Wang, L.X., Guo, Y., ... Varshney, R.K. (2015). Legume crops phylogeny and genetic diversity for science and breeding. Critical Reviews in Plant Sciences, *34*, 43–104.

Sperry, J.F. & Robertson, D.C. (1975). Inhibition of growth by erythritol catabolism in *Brucella abortus*. Journal of Bacteriology, *124*(1), 391–397.

Stevens, J.B., De Luca, N.G., Beringer, J.E., Ringer, J.P., Yeoman, K.H. & Johnston, A.W.B. (2000). The *purMN* genes of *Rhizobium leguminosarum* and a superficial link with siderophore production. Molecular Plant-Microbe Interactions, *228*(2), 228–231.

Streeter, J.G. (1991). Transport and metabolism of carbon and nitrogen in legume nodules. Advances in botanical research, *18*, 129–187.

Streeter, J.G. (2003). Effect of trehalose on survival of *Bradyrhizobium japonicum* during desiccation. Journal of Applied Microbiology, *95*, 484–491.

Sudtachat, N., Ito, N., Itakura, M., Masuda, S., Eda, S., Mitsui, H., Kawaharada, Y. & Minamisawa, K. (2009). Aerobic vanillate degradation and C1 compound metabolism in *Bradyrhizobium japonicum*. Applied and Environmental Microbiology, *75*(15), 5012–5017.

Tapias, A., De Henestrosa, A.R. & Barbé, J. (1997). Characterization of the promoter of the *Rhizobium etli recA* gene. Journal of Bacteriology, *179*(5), 1573–1579.

Taylor, B.L. & Zhulin, I.B. (1999). PAS domains: internal sensors of oxygen, redox potential, and light. Microbiology and Molecular Biology Reviews, *63*(2), 479–506.

Terpolilli, J.J., Hood, G.A. & Poole, P.S. (2012). What determines the efficiency of N_2-fixing *Rhizobium*-legume symbioses? Advances in Microbial Physiology, *60*, 325–389.

Thies, J.E., Singleton, P.W. & Bohlool, B.B. (1991). Modeling symbiotic performance of introduced rhizobia in the field by use of indices of indigenous population size and nitrogen status of the soil. Applied and Environmental Microbiology, *57*(1), 29–37.

Thöny, B. & Hennecke, H. (1989). The -24/-12 promoter comes of age. FEMS Microbiology Reviews, *5*(4), 341–357.

Timmers, A.C.J., Soupene, E., Auriac, M.C., De Billy, F., Vasse, J., Boistard, P. & Truchet, G. (2000). Saprophytic intracellular rhizobia in alfalfa nodules. Molecular Plant-Microbe Interactions, *13*(11), 1204–1213.

Tkacz, A., Cheema, J., Chandra, G., Grant, A. & Poole, P.S. (2015). Stability and succession of the rhizosphere microbiota depends upon plant type and soil composition. ISME Journal, *9*(11), 2349–2359.

Tkacz, A. & Poole, P. (2021). The plant microbiome: The dark and dirty secrets of plant growth. Plants People Planet, *3*(2), 124–129.

Triplett, E.W. (1990). Construction of a symbiotically effective strain of *Rhizobium leguminosarum* bv. *trifolii* with increased nodulation competitiveness. Applied and Environmental Microbiology, *56*(1), 98–103.

Triplett, E.W. & Sadowsky, M.J. (1992). Genetics of competition for nodulation of legumes. Annual Review of Microbiology, *46*(1), 399–428.

Udvardi, M.K. & Poole, P.S. (2013). Transport and metabolism in legume-rhizobia Symbioses. Annual Review of Plant Biology, *64*(1), 781–805.

van Dillewijn, P., Soto, M.J., Villadas, P.J. & Toro, N. (2001). Construction and environmental release of a *Sinorhizobium meliloti* strain genetically modified to be more competitive for alfalfa nodulation. Applied and Environmental Microbiology, *67*(9), 3860–3865.

Van Opijnen, T. & Camilli, A. (2013). Transposon insertion sequencing: A new tool for systems-level analysis of microorganisms. Nature Reviews Microbiology, *11*(7), 435–442.

Van Velzen, R., Holmer, R., Bu, F., Rutten, L., Van Zeijl, A., Liu, W., Santuari, L., Cao, Q., Sharma, T., Shen, D., Roswanjaya, Y., Wardhani, T.A.K., Kalhor, M.S.,

Jansen, J., Van Den Hoogen, J., Güngör, B., Hartog, M., Hontelez, J., Verver, J., ... Soltis, D.E. (2018). Comparative genomics of the non-legume *Parasponia* reveals insights into evolution of nitrogen-fixing *Rhizobium* symbioses. Proceedings of the National Academy of Sciences, *115*(20), E4700–E4709.

Vande Broek, A. & Vanderleyden, J. (1995). The role of bacterial motility, chemotaxis, and attachment in bacteria-plant interactions. Molecular Plant-Microbe Interactions, *8*, 800–810.

Vargas, A.A.T. & Graham, P.H. (1989). Cultivar and pH effects on competition for nodule sites between isolates of *Rhizobium* in beans. Plant and Soil, *117*(2), 195–200.

Venturi, V. & Keel, C. (2016). Signaling in the rhizosphere. Trends in Plant Science, *21*(3), 187–198.

Vlassak, K.M. & Vanderleyden, J. (1997). Factors influencing nodule occupancy by inoculant rhizobia. Critical Reviews in Plant Sciences, *16*(2), 163–229.

Voorhorst, W.G.B., Eggen, R.I.L., Luesink, E.J. & De Vos, W.M. (1995). Characterization of the *celB* gene coding for β-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* and its expression and site-directed mutation in *Escherichia coli*. Journal of Bacteriology, *177*(24), 7105–7111.

Walker, C.J. & Willows, R.D. (1997). Mechanism and regulation of Mg-chelatase. The Biochemical Journal, *327*, 321–333.

Walker, L., Lagunas, B. & Gifford, M.L. (2020). Determinants of host range specificity in legume-rhizobia symbiosis. Frontiers in Microbiology, *11*.

Wang, Q., Liu, J. & Zhu, H. (2018). Genetic and molecular mechanisms underlying symbiotic specificity in legume-*Rhizobium* interactions. Frontiers in Plant Science, *9*, 1–8.

Wang, S., Lu, T., Xue, Q., Xu, K. & Cheng, G. (2019). Antioxidation and symbiotic nitrogen fixation function of *prxA* gene in *Mesorhizobium huakuii*. MicrobiologyOpen, *8*(10), 1–9.

Weaver, R.W. & Frederick, L.R. (1974). Effect of inoculum rate on competitive nodulation of *Glycine max* L. Merrill. II. Field studies. Agronomy Journal, *66*, 233–236.

Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. PLoS ONE, *6*(2).

Weidenhaupt, M., Rossi, P., Beck, C., Fischer, H.M. & Hennecke, H. (1996). *Bradyrhizobium japonicum* possesses two discrete sets of electron transfer flavoprotein genes: *fixA*, *fixB* and *etfS*, *etfL*. Archives of Microbiology, *165*(3), 169–178.

Westhoek, A., Clark, L.J., Culbert, M., Dalchau, N., Griffiths, M., Jorrín, B., Karunakaran, R., Ledermann, R., Tkacz, A., Webb, I., James, E.K., Poole, P.S. & Turnbull, L.A. (2021). Conditional sanctioning in a legume-*Rhizobium* mutualism. Proceedings of the National Academy of Sciences, *118*(19), 1–8.

Westhoek, A., Field, E., Rehling, F., Mulley, G., Webb, I., Poole, P.S. & Turnbull, L.A. (2017). Policing the legume-*Rhizobium* symbiosis: A critical test of partner choice. Scientific Reports, *7*(1419), 1–10.

Wheatley, R.M., Ford, B.L., Li, L., Aroney, S.T.N., Knights, H.E., Ledermann, R., East, A.K., Ramachandran, V.K. & Poole, P.S. (2020). Lifestyle adaptations of *Rhizobium* from rhizosphere to symbiosis. Proceedings of the National Academy of Sciences, *117*(38), 23823–23834.

Wheatley, R.M., Ramachandran, V.K., Geddes, B.A., Perry, B.J., Yost, C.K. & Poole, P.S. (2017). The role of O_2 in the Growth of *Rhizobium leguminosarum* bv. *viciae* 3841 on glucose and succinate. ASM Bacteriology, *199*(1), 1–17.

Williams, L.E. & Phillips, D.A. (1983). Increased soybean productivity with a *Rhizobium japonicum* mutant. Crop Science, *23*, 246–250.

Willis, L.B. & Walker, G.C. (1999). A novel *Sinorhizobium meliloti* operon encodes an α-glucosidase and a periplasmic-binding-protein-dependent transport system for α-glucosides. Journal of Bacteriology, *181*(14), 4176–4184.

Wilson, K.J., Sessitsch, A., Corbo, J.C., Giller, K.E., Akkermans, A.D.L. & Jefferson, R.A. (1995). β-glucuronidase (GUS) transposons for ecological and genetic studies of rhizobia and other gram-negative bacteria. Microbiology, *141*(7), 1691–1705.

Wilson, K.J. (1995). Molecular techniques for the study of rhizobial ecology in the field. Soil Biology and Biochemistry, *27*(4-5), 501–514.

Wood, M. & Stanway, A.P. (2001). *Myo*-inositol catabolism by *rhizobium* in soil: hplc and enzymatic studies. Soil Biology and Biochemistry, *33*(3), 375–379.

Yates, R.J., Howieson, J.G., Reeve, W.G. & O'Hara, G.W. (2011). A re-appraisal of the biology and terminology describing rhizobial strain success in nodule occupancy of legumes in agriculture. Plant and Soil, *348*(1-2), 255–267.

Yost, C.K., Rath, A.M., Noel, T.C. & Hynes, M.F. (2006). Characterization of genes involved in erythritol catabolism in *Rhizobium leguminosarum* bv. *viciae*. Microbiology, *152*(7), 2061–2074.

Yost, C.K., Rochepeau, P. & Hynes, M.F. (1998). *Rhizobium leguminosarum* contains a group of genes that appear to code for methyl-accepting chemotaxis proteins. Microbiology, *144*(7), 1945–1956.

Young, J.P.W., Crossman, L.C., Johnston, A.W.B., Thomson, N.R., Ghazoui, Z.F., Hull, K.H., Wexler, M., Curson, A.R.J., Todd, J.D., Poole, P.S., Mauchline, T.H., East, A.K., Quail, M.A., Churcher, C., Arrowsmith, C., Cherevach, I., Chillingworth, T., Clarke, K., Cronin, A., ... Parkhill, J. (2006). The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. Genome Biology, *7*(4).

Zipfel, C. & Oldroyd, G.E. (2017). Plant signalling in symbiosis and immunity. Nature, *543*(7645), 328–336.

8 Appendix

8.1 Wheatley et al., 2020

8.1.1 Author Contributions

R. M. Wheatley, B. L. Ford, S. T. N. Aroney, H. E. Knights, V. K. Ramachandran, and P. S. Poole designed research; R. M. Wheatley, B. L. Ford, Li Li, S. T. N. Aroney, H. E. Knights, and V. K. Ramachandran performed research; R. M. Wheatley and B. L. Ford contributed new reagents/analytic tools; R. M. Wheatley, B. L. Ford, S. T. N. Aroney, H. E. Knights, R. Ledermann, A. K. East, V. K. Ramachandran, and P.S.P. analyzed data; and R. M. Wheatley, B. L. Ford, S. T. N. Aroney, H. E. Knights, R. Ledermann, A. K. East, V. K. Ramachandran, and P. S. Poole wrote the manuscript.

8.1.2 Full Text

Lifestyle adaptations of *Rhizobium* from rhizosphere to symbiosis

Rachel M. Wheatley[a,1], Brandon L. Ford[a,1], Li Li[b,1], Samuel T. N. Aroney[a], Hayley E. Knights[a], Raphael Ledermann[a], Alison K. East[a], Vinoy K. Ramachandran[a,2], and Philip S. Poole[a,2]

[a]Department of Plant Sciences, University of Oxford, OX1 3RB Oxford, United Kingdom; and [b]Chinese Academy of Sciences Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, 430074 Wuhan, People's Republic of China

Edited by Éva Kondorosi, Hungarian Academy of Sciences, Biological Research Centre, Szeged, Hungary, and approved August 4, 2020 (received for review May 7, 2020)

By analyzing successive lifestyle stages of a model *Rhizobium*–legume symbiosis using mariner-based transposon insertion sequencing (INSeq), we have defined the genes required for rhizosphere growth, root colonization, bacterial infection, N_2-fixing bacteroids, and release from legume (pea) nodules. While only 27 genes are annotated as *nif* and *fix* in *Rhizobium leguminosarum*, we show 603 genetic regions (593 genes, 5 transfer RNAs, and 5 RNA features) are required for the competitive ability to nodulate pea and fix N_2. Of these, 146 are common to rhizosphere growth through to bacteroids. This large number of genes, defined as rhizosphere-progressive, highlights how critical successful competition in the rhizosphere is to subsequent infection and nodulation. As expected, there is also a large group (211) specific for nodule bacteria and bacteroid function. Nodule infection and bacteroid formation require genes for motility, cell envelope restructuring, nodulation signaling, N_2 fixation, and metabolic adaptation. Metabolic adaptation includes urea, erythritol and aldehyde metabolism, glycogen synthesis, dicarboxylate metabolism, and glutamine synthesis (GlnII). There are 17 separate lifestyle adaptations specific to rhizosphere growth and 23 to root colonization, distinct from infection and nodule formation. These results dramatically highlight the importance of competition at multiple stages of a *Rhizobium*–legume symbiosis.

Rhizobium | legume | N_2 fixation | nodulation | rhizosphere

Biological N_2 fixation carried out by bacteria provides ~65% of the biosphere's available nitrogen (1), with the single greatest contribution coming from the symbioses between soil-dwelling bacteria of the Rhizobiaceae family and plants of the legume family (1–3). Rhizobia provide the plant with ammonium (NH_4^+) and receive carbon sources as an energy supply in return (4). This effectively functions as a biofertilizer for the legume host, increasing plant growth without the need for exogenously applied nitrogen (3, 5). During symbiosis, rhizobia must adapt to several different lifestyles. These range from free-living growth in the rhizosphere, through root attachment and colonization, to passage along infection threads, differentiation into bacteroids that fix N_2, and, finally, bacterial release from nodules at senescence.

Several decades of intense research have gone into understanding the mechanistic details of signaling between rhizobia and legumes leading to nodulation (6–8). Rhizobia respond to plant flavonoids by synthesizing lipochitooligosaccharides (LCOs) that are detected by plant LysM receptors to initiate nodule morphogenesis and concomitant rhizobial infection (8–11). Rhizobia typically, but not always, attach to the tips of developing root hairs, triggering root hair curling and entrapment of bacteria in an infection pocket (7). The plant cell plasma membrane invaginates to form an infection thread, down which bacteria grow toward the base of the root hair cell through continuous division at the leading edge. Infection is usually a clonal event (12), selecting one *Rhizobium* and allowing it to multiply to high numbers within the ramifying threads. Infection threads deliver rhizobia to the root cortex of the developing nodule. In symbiosis between *Rhizobium leguminosarum* bv. *viciae* and pea, bacteria are internalized by nodule cells and undergo terminal differentiation into N_2-fixing bacteroids (10). Nodules provide a protective microaerobic environment to maintain oxygen-labile nitrogenase (6). In exchange for NH_4^+ and alanine, the legume host provides carbon sources to fuel this process, primarily as dicarboxylic acids (13, 14).

However, nodule infection is only one stage of the lifestyle of rhizobia, and they spend much of their time surviving in the rhizosphere, the zone of soil immediately surrounding roots (15). Plants can release up to a fifth of their photosynthate via their roots (16), making the rhizosphere a nutritionally rich zone, but one where there is strong selective pressure and competition for resources. Attachment and subsequent colonization of roots by rhizobia is almost certainly important for their long-term survival, but it is unclear how important these steps are to infection of legumes (17). Research over several decades has enabled identification of around 8 *nif* and 27 *fix* genes required directly for N_2 fixation, as well as genes essential for recognition by the plant, differentiation, and energization of N_2 fixation. However, it has long been speculated these probably represent a fraction of genes needed by rhizobia for successful survival and growth in competition with other bacteria.

Rhizobiaceae have large genomes that probably reflect their adaptation to the heterogenous soil environment and dramatic lifestyle and developmental changes undergone in symbiosis (18, 19). The 7.75-megabase genome of *R. leguminosarum* bv. *viciae* Rlv3841 consists of a circular chromosome (4,788 genes) and 6 plasmids: pRL12 (790 genes), pRL11 (635 genes), the symbiosis

Significance

Rhizobia are soil-dwelling bacteria that form symbioses with legumes and provide biologically useable nitrogen as ammonium for the host plant. High-throughput DNA sequencing has led to a rapid expansion in publication of complete genomes for numerous rhizobia, but analysis of gene function increasingly lags gene discovery. Mariner-based transposon insertion sequencing has allowed us to characterize the fitness contribution of bacterial genes and determine those functionally important in a *Rhizobium*–legume symbiosis at multiple stages of development.

Author contributions: R.M.W., B.L.F., S.T.N.A., H.E.K., V.K.R., and P.S.P. designed research; R.M.W., B.L.F., L.L., S.T.N.A., H.E.K., and V.K.R. performed research; R.M.W. and B.L.F. contributed new reagents/analytic tools; R.M.W., B.L.F., S.T.N.A., H.E.K., R.L., A.K.E., V.K.R., and P.S.P. analyzed data; and R.M.W., B.L.F., S.T.N.A., H.E.K., R.L., A.K.E., V.K.R., and P.S.P. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

[1]R.M.W, B.L.F., and L.L. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: vinoy.ramachandran@plants.ox.ac.uk or philip.poole@plants.ox.ac.uk.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2009094117/-/DCSupplemental.

Downloaded at University of Oxford on September 9, 2020

plasmid pRL10 (461 genes), pRL9 (313 genes), pRL8 (141 genes), and pRL7 (189 genes). Here, we perform a multistage mariner transposon-insertion sequencing (INSeq) genetic screen of Rlv3841 engaging in symbiosis with pea (*Pisum sativum*) at four different symbiotic lifestyle stages: 1) growth in the rhizosphere, 2) root colonization, 3) undifferentiated nodule bacteria, and 4) terminally differentiated bacteroids. INSeq is a powerful and robust technique that can be used to functionally annotate bacterial genomes (20–24). Libraries of mariner transposon-insertion mutants are analyzed by high-throughput sequencing to assess the effect of mutation of a gene on growth and survival of the bacterium. This has allowed us to identify the genetic requirements for competitiveness at multiple developmental stages during symbiosis.

Results and Discussion

Mariner Library Construction, Sequencing, and Analysis. Rlv3841 mariner-insertion mutants (input1) were inoculated onto 7-d-old pea seedlings and subsequently recovered at 5 d postinoculation (dpi) to construct two libraries for INSeq analysis: 1) rhizosphere and 2) roots (Fig. 1). A second mutant library (input2) was inoculated onto 7-d-old pea seedlings and subsequently recovered at 28 dpi from nodules. A total of 150,000 nodules were picked from ~1,500 plants, and a further two libraries for INSeq analysis were constructed: 3) undifferentiated nodule bacteria, generated by growing bacteria from crushed nodules (which will represent nodule bacteria that are not terminally differentiated along with those recently released from infection threads) for 12 h in rich medium before extracting DNA; and 4) bacteroids, generated from DNA extracted directly from nodules (Fig. 1). A total of six INSeq libraries (input 1, input 2, rhizosphere, roots, nodule bacteria, and bacteroids) were sequenced and analyzed by a Hidden Markov Model (HMM) (25). 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 (25). Each of the 7,357 Rlv3841 genes that contain TA insertion sites (99.9% of the genome) was classified as essential (ES) (tolerates no or very few insertions), defective (DE) (insertion impairs fitness), neutral (NE) (insertion has a neutral impact on fitness), or advantaged (AD) (insertion enhances fitness) (Dataset S1) (25). Genes classified as NE in the respective input library (6,429 for input1 and 6,656 for input2) were considered for each test condition, with genes classified as either ES or DE (ES/DE) regarded as required at that stage of symbiosis and described thus throughout this paper.

Bacteria with a mutation that results in out-competition at any stage of early nodule formation will occupy fewer nodules on peas and lead to an ES/DE classification. However, many mutants unable to fix N_2, such as those defective in *nifH* (encoding Fe protein of nitrogenase), do not alter the competitiveness for nodule infection. Nonetheless, plant sanctioning results in the formation of smaller nodules (26). We therefore predicted at the outset of this work, and subsequently confirm, that with mutants impaired specifically in N_2 fixation but not nodule formation, the result is an ES/DE classification for the gene in bacteroids.

Classification of Symbiotic Fitness Determinants. Across all stages of symbiosis, INSeq reveals a total of 603 genetic regions (593 genes, 5 transfer RNAs [tRNAs], and 5 RNA features) are required to competitively nodulate peas (Fig. 2). A total of 17 are required only in the rhizosphere (*SI Appendix*, Table S1), while 146 are classified as rhizosphere-progressive (*SI Appendix*, Table S2), required not only in the rhizosphere but also at every subsequent stage. Seven genes are required for both rhizosphere and root colonization (rhizosphere and root) (*SI Appendix*, Table S3), with 23 genes specifically required for root colonization (root) (*SI Appendix*, Table S4). Genes classified as rhizosphere- or root-required are important for fitness and persistence in those environments but not linked to nodule infection, per se. There are 33 genes required from root colonization to nodule (root-progressive) (*SI Appendix*, Table S5), while 211 genes are required by both nodule bacteria and bacteroids (nodule-general) (*SI Appendix*, Table S6). There are an additional 142 specific for nodule bacteria, i.e., required only for regrowth of bacteria from crushed nodules (*SI Appendix*, Table S7), with 24 specific for bacteroids (direct isolation from nodules without regrowth) (*SI Appendix*, Table S8 and Fig. 2).

Rhizosphere Genes. The rhizosphere, the zone immediately surrounding plant roots, is an oxidative environment formed through the release of reactive oxygen species (27, 28). Mutation of the 17 genetic regions required specifically in the rhizosphere (*SI Appendix*, Table S1) did not impair subsequent stages of root colonization or nodulation (Fig. 2). Rhizosphere-specific genes involved in the oxidative-stress response, important for survival in this environment, encode a putative oxidoreductase (pRL110539), a superoxide dismutase (RL1340 [*sodB*]), and an RNA-binding protein Hfq (RL2284) involved in control of posttranscriptional gene regulation. Persistence in the rhizosphere also required RL2964 (*cspA3*), encoding a cold shock protein. This rhizosphere-specific group also includes a single tRNA ($tRNA^{Gln}$) and five genes predicted to encode metabolic or biosynthetic functions for pRL80090, a fucose operon protein; pRL90136, a glycosyl transferase; pRL90204, an amidase; pRL100145, an acyl-CoA dehydrogenase; and RL3335, a lysophospholipase.

Rhizosphere-Progressive Genes. The 146 rhizosphere-progressive genes (Fig. 2 and *SI Appendix*, Table S2), required in the rhizosphere but also at subsequent stages, include genes for the biosynthesis of tryptophan (RL0021 [*trpB*], RL0022 [*trpA*], RL2493 [*trpD*], RL2494 [*trpC*], and RL3521 [*trpE*]), riboflavin (RL1621 [*ribG*] and RL1622 [*ribC*]), leucine (RL3513 [*leuA1*], RL4707 [*leuB*], and RL4732 [*leuS*]), isoleucine (RL1803 [*ilvD1*], RL3205 [*ilvC*], and RL3245 [*ilvI*]), and cobalamin (pRL110632 [*cobF*], RL2781 [*cobU*], RL2781A [*cobV*], RL2829 [*cobO*], RL4348 [*cobT*], and RL4349 [*cobS*]). There may be a specific requirement for these biosynthetic genes in the rhizosphere, with mutants subsequently uncompetitive at later stages of symbiosis, or, alternatively, they

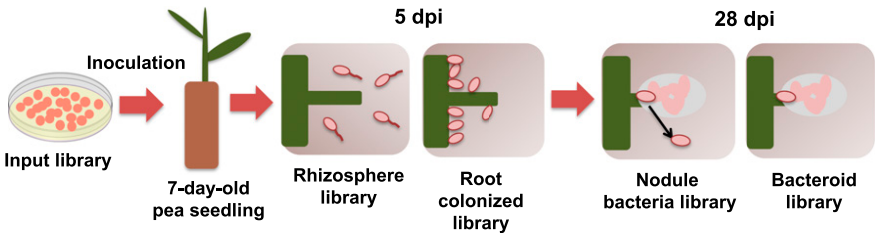


Fig. 1. Collection steps following inoculation of pea plants with the INSeq input libraries. Bacterial DNA was purified from 1) the rhizosphere (5 dpi) and 2) colonized roots (5 dpi) following inoculation with input1; and from 3) nodule bacteria (28 dpi) and 4) bacteroids (28 dpi) following inoculation with input2. These four samples, together with DNA purified from input1 and input2, were used to make libraries for INSeq analysis.

Downloaded at University of Oxford on September 9, 2020

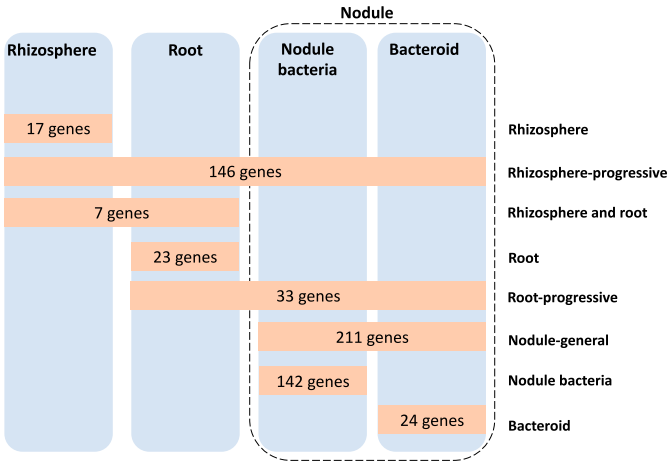


Fig. 2. Rlv3841 genes required during the stages of symbiosis with pea. Orange boxes show the number at each symbiotic stage: rhizosphere (17 genes) (*SI Appendix*, Table S1), rhizosphere-progressive (146 genes) (*SI Appendix*, Table S2), rhizosphere and root (7 genes) (*SI Appendix*, Table S3), root (23 genes) (*SI Appendix*, Table S4), root-progressive (33 genes) (*SI Appendix*, Table S5), nodule-general (211 genes) (*SI Appendix*, Table S6), nodule bacteria (142 genes) (*SI Appendix*, Table S7), and bacteroid (24 genes) (*SI Appendix*, Table S8). All genes shown were classified as NE in the respective input library.

may be required throughout all symbiotic stages. Riboflavin availability influences rhizobial survival in the rhizosphere and root colonization. An *R. leguminosarum* mutant in *ribN* (encoding putative riboflavin transporter protein RibN) showed lower nodule occupancy in pea plants (29).

The outer surface of rhizobia, composed of a variety of polysaccharides, including lipopolysaccharide (LPS) and exopolysaccharide (EPS) (17), is critical for both attachment to root hairs and host-specific recognition. Rhizosphere-progressive genes include those encoding putative proteins: O-antigen ligase (pRL90053); EPS production protein (pRL110389), with high homology (97% identity [id]) to PssF of *R. leguminosarum* TA1 involved in polysaccharide export (30); and O-antigen transporter (RL0821). In addition, both RL3651 (*pssG*) and RL3664 (*pssN*) have well-defined roles in EPS biosynthesis (30), and their requirement, alongside pRL110389 (*pssF*) and RL1600 (*ppx*), highlights the importance of correct EPS processing not only in nodule infection but in all stages of interaction with the plant. The transcriptional regulator RL1379 (*rosR*), classified as rhizosphere-progressive, has a role in nodulation competitiveness (31). RosR regulates *pss* genes encoding EPS, and *R. leguminosarum rosR* mutants produce only one-third of the EPS of wild type (32).

Other requirements include RL1589 (*ropB*) and RL2775 (*ropA*), which are suggested to encode outer membrane amyloid fibrils in *R. leguminosarum* (33). These are conserved across numerous rhizobia (34, 35) and are possibly required for either root attachment or transport. RL2256 (*ntrB*) and RL2257 (*ntrC*), encoding the NtrB/NtrC two component regulatory system, control expression of genes in response to nitrogen limitation. While in this work, pea seedlings were grown in nitrogen-free rooting solution to enable nodulation, this pathway is likely to be important in a natural soil environment where nitrogen availability is low. RL0422 (*rpoN*), the gene encoding RNA polymerase sigma factor 54, is classified as rhizosphere-progressive. Interestingly, this indicates that while essential for *nif* gene expression and dicarboxylate transport (36) (processes that occur in nodules), it is required for processes in the rhizosphere. Bacterial uptake of solutes is important, shown by the requirement for genes encoding ATP-binding cassette (ABC) transporters of nitrogen-containing ions (RL4400 to RL4402) and ferric cations (RL4583 [*fbpB*]). Another ABC system is required (RL1004 to RL1003), but the solute transported is unknown. RL3557 (*bacA*) encoding the BacA transporter protein is well established to be required for bacteroid development (37). BacA protects rhizobia against the antimicrobial activity of nodule-specific cysteine-rich (NCR) peptides produced by the host pea (38, 39), but clearly it is also important at earlier stages of interaction with the plant. The gene encoding the stringent response protein RelA (RL1506 [*relA*]), which synthesizes ppGpp, was similarly required across all symbiotic stages, indicating its central importance in response to stress.

Root Colonization Genes. The root library, formed by collecting bacteria from the entire pea root, consists of colonizing bacteria attached over the whole epidermal surface. The genes specific for rhizosphere and root (7 genes; *SI Appendix*, Table S3) are required for both bacterial rhizosphere growth and root colonization, while 23 genes are required only for root colonization (*SI Appendix*, Table S4; designated as "Root" in Fig. 2). This latter group is important for attachment to roots (8, 17, 40) but is not involved in nodule formation, indicating a pathway of root colonization, distinct from that of nodulation.

Root-specific genes include those encoding putative cell envelope-related functions for pRL90052, polysaccharide biosynthesis O-antigen–related protein; and RL3663 (*pssO*), a polysaccharide transport outer membrane protein. These cell envelope-related functions may actively facilitate adherence to the root surface or may be important for general cell fitness during colonization. RL2512, encoding a putative transmembrane protein, was specifically required for root colonization. RL2512 encodes a putative SecG, which forms part of the general protein secretion (Sec) system in bacteria. The protein encoded by RL2512 shows homology to SecG in *Escherichia coli* (39% id over the amino-terminal 71 amino acids) and *Agrobacterium tumefaciens* (73% id over 175 amino acids). Potential orthologs of the Sec system have been identified in Rlv3841 by homology with *E. coli* proteins (41). However, these chromosomal genes were not located in operons or even close to one another. SecG forms a protein-conducting channel in the SecYEG complex (41). Both the putative Rlv3841 genes encoding SecE (RL1759) and SecY (RL1794) were essential in the input libraries as well as across all stages of symbiosis. The different phenotype observed for RL2512 suggests there may be a second gene encoding a SecG-like function in Rlv3841 (although no gene was found with >26% id).

Root-Progressive Genes. Thirty-three root-progressive genes affect all stages from root colonization onward (Fig. 2 and *SI Appendix*, Table S5). They include RL2382 (*nodM*/*glmS*), encoding a fructose-6-phosphate aminotransferase that forms glucosamine-6-phosphate. There are two homologous *nodM* genes in *R. leguminosarum*, RL2382 and pRL100180, encoded on the chromosome and symbiosis plasmid, respectively (42). Biochemical assays demonstrated either protein provides sufficient glucosamine-6-phosphate (G6P) for LCO Nod factor synthesis (42). A double *nodM* mutant in *R. leguminosarum* no longer nodulated pea or vetch (42). The two genes fall into different categories in this study: pRL100180 (*nodM*) is required only by nodule bacteria, while RL2382 (*nodM*/*glmS*) is root-progressive. Their different INSeq phenotypes suggest differential expression during symbiosis, consistent with plant-flavonoid induction of *nodM*.

This root-progressive group also includes two genes encoding putative cell surface constituents for RL0818, a LPS biosynthesis protein; and RL3677 (*ispL*), a UDP-glucuronate 5′-epimerase required for glycoconjugation in EPS biosynthesis (43). This root-progressive designation suggests these two cell surface–related genes are required to produce the correct bacterial cell

Downloaded at University of Oxford on September 9, 2020

surface needed for root colonization, as well as in subsequent steps, potentially root hair attachment, infection thread ramification, or bacteroid formation in nodules. An ABC transporter (RL4538 to RL4539 [*ccmB*, *cycZ*]) showing homology with a system exporting heme-binding proteins to the periplasm in *Rhizobium etli* (44) is required, together with RL4540 (*cycY*), a cytochrome *c* biogenesis protein responsible for interchange of thiol:disulfide bonds. This suggests novel cytochrome biosynthesis and export is important during initial root colonization.

Nodule-General Genes. A total of 211 genes are classified nodule-general (Fig. 2 and *SI Appendix*, Table S6). These are genes whose mutation impaired recovery from both nodule bacteria (i.e., regrown bacteria) and bacteroids (i.e., isolated from total nodule rhizobacterial DNA without regrowth). This group includes nine *nod* genes, encoding components for the biosynthesis and signaling pathways of LCO nodulation (Nod) factors (45). LCOs are diffusible molecules, and *nod* mutants are still able to initiate symbiosis if coinoculated 1:1 with wild type. However, in mass competition, such as this INSeq experiment, *nod* mutants may either have delayed establishment or reduced cell numbers within nodules, one or both of which could result in a nodule-general classification. The genes are organized in three operons on pRL10: pRL100180 to pRL100178 (*nodMNT*), pRL100181 to pRL100183 (*nodLEF*), and pRL100184 to pRL100189 (*nodD-nodABCIJ*), and in the latter two clusters, every gene (nine in total) is classified as nodule-general (*SI Appendix*, Table S6). These genes include biosynthesis of Nod factor as well as export by the ABC transporter encoded by pRL100188 (*nodI*) and pRL100189 (*nodJ*).

The nitrogenase complex is encoded by genes pRL100162 to pRL100158 (*nifHDKEN*) and pRL100196 to pRL100195 (*nifAB*). These genes are classified as nodule-general (*nifDK* and *nifN*) (*SI Appendix*, Table S6) and bacteroid (*nifH*, *nifE*, and *nifAB*) (*SI Appendix*, Table S8). Although the classification of these latter four as bacteroid specific suggests that their mutation reduces N_2-fixing bacteroids but not the level of undifferentiated bacteria within nodules, as these classifications are closely related, the division of functionally related clusters into separate groups should not be overinterpreted.

Electron-transfer flavoproteins (ETFs) transfer electrons to nitrogenase by bifurcation (46). Genes *fixABCX* (pRL100200 to pRL100197) encoding ETF FixABCX were required in the nodule, although, as above, they fall into two categories: nodule-general (*fixAB*) (*SI Appendix*, Table S6) and bacteroid (*fixCX*) (*SI Appendix*, Table S8). Again, we might tentatively conclude that while all four genes are needed for symbiotic N_2 fixation, mutation of *fixC* and *fixX* does not alter the number of viable nodule bacteria. Mutation of ETF proteins abolishes N_2 fixation in several rhizobia–legume symbioses (47).

The microaerobic environment required for the efficient functioning of the nitrogenase complex is achieved by the combined activities of plant leghemoglobin and the bacteroid's high-affinity CBB_3 cytochrome complex. Rlv3841 has two clusters of *fix* genes that encode the components of CBB_3, one on the symbiosis plasmid *fixN1O1Q1P1G1H1I1S1* (pRL100205 to pRL100211) and the other on pRL9 (pRL90018 to pRL90012 [*fixN2O2Q2P2-G2H2I2S2*]). Only *fixN1* and *fixN2* were required (*SI Appendix*, Tables S2 and S6), suggesting the other duplicated components of the CBB_3 complex can compensate for each other and, possibly, that the level of FixN limits the activity of the complex.

Differentiation of rhizobia into bacteroids requires significant reprogramming of the LPS and outer membrane (48). Genes with a predicted role in cell envelope functions formed an important group of nodule-general genes. These include RL3649 (*pssI*) and RL3650 (*pssH*), encoding putative glycosyltransferases, and RL3820, encoding a putative EPS-production protein. These proteins may restructure the cell surface during infection and bacteroid differentiation. O-antigen biosynthesis also appears to play a role in these structural changes, with pRL110056, encoding a putative UDP-*N*-acetyl-D-glucosamine 6-dehydrogenase involved in O-antigen biosynthesis classified as nodule-general. As with growth in the rhizosphere and root colonization, stress- and defense-response genes are also required in nodules. These include RL0824, encoding a putative oxidoreductase; RL0962, encoding a putative dioxygenase; and RL0994 and RL1545, encoding putative reductases.

In terms of metabolism, nodule-general genes include pRL100434 to pRL100433, which are predicted to encode allophanate hydrolase, catalyzing the hydrolysis of allophanate yielding ammonium and CO_2. Allophanate can be either derived from cyanuric acid degradation or be an alternative to urease in degradation of urea (49), in which urea is first carboxylated to allophanate. Indeed, pRL100436 and pRL100437 (nodule bacteria-specific; *SI Appendix*, Table S6) are annotated to encode a biotin-dependent carboxylase, which may be a urea carboxylase. Urea carboxylase belongs to the biotin-dependent carboxylase family of enzymes, which has 40% id to urea carboxylase of *Saccharomyces cerevisiae* and *Oleomonas sagaranensis* (50). Overall, this suggests urea catabolism is important in nodule bacteria and bacteroids.

Genes pRL100246 and pRL100247, encoding a putative aldehyde dehydrogenase and adjacent conserved hypothetical protein, were nodule-general. This suggests aldehyde metabolism is important for optimal fitness in both nodule bacteria and bacteroids. Soybean nodules contain aldehydes (31 to 53 nM), which are converted to acetate by aldehyde dehydrogenase (51, 52). Erythritol utilization is also required in nodules: pRL120205 (*eryB*), encoding erythritol phosphate dehydrogenase; and pRL120206 (*eryC*), encoding erythrulose 4-phosphate dehydrogenase are nodule-general. This agrees with the known reduction in nodule competitiveness of *ery* mutants (53). RL4117 (*glgA*) encoding glycogen synthase is nodule-general. Glycogen is an important carbon store in rhizobia, and its disruption impairs growth on glucose-containing molecules (54).

Various transport systems are required in nodules. Genes required include those encoding pRL110047 (permease of ABC exporter), pRL90262 (ABC of a hydrophobic amino acid transporter [HAAT] ABC uptake system), RL0186 to RL0187 (permease components of peptide transporter [PepT] ABC uptake system), RL0226 and RL0228 (permease and solute binding protein [SBP] of a second PepT family), and RL3333 (ABC of an unclassified transporter). The importance of NCR peptides in development of mature bacteroids within N_2-fixing nodules is well known (55). While we speculate that the two PepT uptake systems are involved in delivery of plant-derived peptides crucial for plant-controlled bacteroid development, uptake by systems transporting unknown solutes and export of unwanted compounds is clearly also necessary.

Nodule Bacteria Genes. Although most genes with known or obvious roles in nodule infection are classified as nodule-general or bacteroid, there are an additional 142 genes specific for nodule bacteria (Fig. 2 and *SI Appendix*, Table S7). This INSeq library is made up of undifferentiated bacteria, likely to come from inside, or recently released from, infection threads, and capable of growth in tryptone–yeast (TY) (fully differentiated bacteroids are unable to regrow in this way). Classification of genes as specific for nodule bacteria may be for one of two reasons: firstly, that they are needed for optimum bacterial growth within infection threads; or, secondly, that they are required for regrowth in TY medium, with implications for importance in bacterial release as nodules senesce.

This list of genes includes several flagella-related genes that are considered below. A few defense- and stress-related genes, including RL0392 (encoding a putative hemolysin-like protein), RL2102 (*cspA5*) (encoding a putative cold shock protein), and pRL110509 and RL2526 (encoding putative oxidoreductases),

Downloaded at University of Oxford on September 9, 2020

are also specific for nodule bacteria. The gene encoding a resistance–nodule–division (RND) family efflux transporter (pRL120696) is required, and its role is likely to be in defense, exporting toxic/unwanted chemicals from bacterial cells. The gene encoding an ABC export system RL4018 (ABC-permease fusion) may also be involved in a similar task. In fact, genes encoding components of several ABC uptake transporters are classified as nodule bacteria-specific, including pRL100062 (permease of polyamine/opine/phosphonate transporter [POPT] family), pRL100072 (SBP of polar amino acid transporter [PAAT] family), pRL120071 (SBP of PAAT, which may transport galactosamine, glucosamine given its high homology to SMb21135) (56), pRL120560 (ABC carbohydrate uptake transporter 1 [CUT1] family), RL2659 (SBP, class unclassified), and RL3353 (ABC, ferric iron transporter [FeT]). The large number of transporters involved in uptake of amino acids, sucrose, iron, and other unknown solutes may be required for optimum recovery from crushed nodules and subsequent growth in TY and their mutation consequently results in loss from the nodule bacteria library.

Bacteroid Genes. There are 24 genes specifically required by bacteroids (Fig. 2 and *SI Appendix*, Table S8). Mutants in these genes have reduced fitness specifically in formation or maintenance of bacteroids but are not affected in bacterial progression through infection threads or regrowth from nodules. They include several *nif* (pRL100196 and pRL100195 [*nifAB*], pRL100159 [*nifE*], and pRL100162 [*nifH*]) and *fix* (pRL100198 and pRL100197 [*fixCX*]) genes that are needed for nitrogenase assembly and microaerobic respiration (57, 58). This confirms our prediction that mutants impaired in N_2 fixation would result in an ES/DE classification in the libraries recovered from nodules.

The dicarboxylate transport system (Dct) genes (RL3425 [*dctB*] and RL3426 [*dctD*]), enabling dicarboxylic uptake by bacteroids to fuel N_2 fixation (6, 59, 60), are bacteroid-specific. Prior to bacteroid differentiation, rhizobia utilize a wide range of carbon sources and *dct* mutants form well-developed nodules and bacteroids (61). We assume that wild-type and *dct* mutant strains maintain similar populations of viable bacteria in nodules (reflected in the nodule bacteria results), while *dct* mutants have severely reduced N_2 fixation and numbers of bacteroids. RL3424 (*dctA*), encoding the Dct system transporter, gave an NE classification in HMM analysis; however, inspection of the insertion sequencing data showed that out of 17 TA sites in *dctA*, 11 sites had no or very few reads (less than 10). This suggests that 65% of TA sites in *dctA* do not tolerate insertion, emphasizing the significant role of this transporter. When TA insertions are near the gene termini, it may be that mariner insertion does not fully disrupt gene function. Our previous INSeq screen in Rlv3841 similarly reported the termini of RL3424 (*dctA*) to tolerate a large number of insertions (62). RL0037 (*pckA*), encoding phosphoenolpyruvate carboxykinase, which is essential for growth on dicarboxylates (63), was also specifically required in nodule bacteroids (*SI Appendix*, Table S8). This fits with evidence that dicarboxylates drive bacteroid metabolism but not bacterial growth during nodule infection, where multiple carbon sources are probably available.

The cytochrome *bc1* complex is encoded by RL3486 to RL3484 (*petABC*), with RL3486 (*petA*) encoding the ubiquinol–cytochrome *c* reductase iron-sulfur subunit required across all symbiosis growth conditions (rhizosphere-progressive), whereas *petB* is classified as nodule-general. In contrast, RL3484 (*petC*), encoding the cytochrome *c*1 precursor, is specifically required in bacteroids. A mutant of RL3484 (*petC*) has been demonstrated to be Fix^- in *R. leguminosarum* (64). In addition, RL3577 (*guaD*), encoding guanine deaminase, an enzyme that catalyzes the hydrolytic deamination of guanine producing xanthine and ammonia, is also bacteroid-specific (*SI Appendix*, Table S8).

Mutants with Improved Bacterial Colonization or Increased Numbers in Nodules. When mutation in a gene leads to a bacterial strain with improved fitness for a given condition or environment, its INSeq classification is AD. (AD genes at each stage from rhizosphere to nodule [NE or DE in input library] are summarized in *SI Appendix*, Fig. S1 and Table S9.) No genes are classified as AD across all four conditions (rhizosphere-progressive) and none are root-progressive. However, there are considerable groups where mutation leads to better recovery from the rhizosphere (84 genes) and/or increased colonization of roots (172 and 33 genes, respectively), as well as leading to an increased population recovered from crushed nodules: nodule-general (11 genes), nodule bacteria-specific (following regrowth; 207 genes), or bacteroid-specific (22 genes) (*SI Appendix*, Fig. S1 and Table S9). A large proportion (approximately 66%) of AD genes are contiguous and clustered, with many in operons. In contrast to the genes required at a given stage, the reason for an AD classification is far more nuanced. Mutation of genes involved in uptake, sensing, or metabolism of specific compounds can be advantageous to bacteria in a given environment, increasing their fitness, as energy is not expended on pathways or gene expression that is not needed. For example, in contrast to rapid growth in rich broth in vitro, when bacteria are growing slowly, their metabolism will be more restricted, and there will be a greater penalty for synthesizing proteins not absolutely required. AD genes are likely to reflect this. In the rhizosphere, AD-classified genes (84) include ABC homoserine transporter pRL80088, Major Facilitator Superfamily (MSF) nitrate transporter NarK (RL1992), ABC nitrate transporter (NitT) family SBP RL1993, and nitrate reductase transcriptional regulator NasT (RL1994), as well as a large region of 26 genes (RL1870 to RL1895), including genes encoding sensor regulators, heat shock proteins, among others.

Rhizosphere and root AD genes (172; *SI Appendix*, Fig. S1 and Table S9) include ABC transporters involved in uptake of quaternary amines (QAT) pRL100078 to pRL100083 (*qatV6W6X6*); peptides, PepT-family pRL110054 to pRL110053 (*optDA*), pRL110281, and pRL120330; amino acids, pRL120404 (*braC2*) involved in uptake of branched chain amino acids (65) (HAAT), general L-amino uptake system RL2204 to RL2203 (*aapJQ*, polar amino acid [PAAT] family) and pRL80060 (PAAT SBP with homology to a mimosine transporter of *Rhizobium* sp. TAL 1145) (66); carbohydrates RL2895 to RL2896 (CUT1); and phosphate uptake transporter (PhoT) RL1683 and polyamine/opine/phosphonate uptake (POPT) family RL2009 to RL2014. Under these specific experimental conditions, in the rhizosphere and on roots, bacteria can clearly survive without the wide range of compounds that these transporters import. Export systems, RND-family efflux proteins pRL100287 to pRL100286, and multidrug efflux transporter RL3414 (*norM*) are also classified AD under these conditions.

There is a small group of nodule-general AD transport genes, including CUT1 uptake system RL0749 (*aglK*) transporting alpha-glucosides and exporter RL3632. This latter is part of a cluster of five genes (RL3628 to RL3632), which includes a sugar decarboxylase (RL3628) and three glycosyl transferases (RL3629 to RL3631).

When we consider nodule bacteria, even slight changes in the terminal differentiation to bacteroids will affect the proportion of bacteria recovered from nodules. Mutation of a gene that leads to delay of differentiation might lead to an AD gene classification. However, as preparation of the nodule bacteria sample includes recovery in TY broth, any conclusions should be tentative as the influence may be to increase fitness for regrowth in a nutrient-rich medium. Nodule bacteria AD genes (207) include numerous ABC and other transporter systems; CUT1 family systems pRL100440 to pRL100443, contiguous with an inositol-degradation protein (pRL100439); pRL120006 to pRL120007; pRL120557 to pRL120558, contiguous with a glycosyl hydrolase (pRL120559); and RL4378 to RL4375; CUT2

Downloaded at University of Oxford on September 9, 2020

family systems *eryEFG* (pRL120202 to pRL120200), together with *eryH* (pRL120203), a periplasmic lipoprotein involved in erythritol uptake and/or metabolism; pRL90246 to pRL90247; HAAT family components pRL120095 and RL3745 (*braC*) (involved in uptake of branched chain amino acids; ref. 65); and PepT complete systems pRL110517 to pRL110513 and pRL90249 to pRL90253, which could be involved in uptake of NCR peptides as without their uptake, differentiation into bacteroids might be reduced leading to increased bacterial recovery. Tripartite ATP-independent transporter (TRAP) import system RL4359 to RL4361 is likely to be involved in the uptake of methyl pyruvate or pyruvate by homology with *Sinorhizobium meliloti* (56). For uptake of metals, the following genes are classified AD in nodule bacteria; Iron chelate uptake transporter (FeCT) components RL2713 to RL2715; FeT family RL3350 to RL3352; Manganese/Zinc/Iron chelate transporter (MZT) family components RL3885 to RL3886 (*sitAB*), likely to import manganese (from homology to *S. meliloti* transporter; ref. 56), and MgtE-like family cationic transporter RL2550. In addition, AD genes include clusters encoding three extracytoplasmic function (ECF) sigma factors and their associated proteins involved in signal transduction from the external environment: RL1471 to RL1477, including *ecfDdgpfD1D2D3*; RL3233 to RL3235, including *ecfElppE*; RL3508 to RL3512, including *lppFecfFasfF*. These sigma factors control, often large, networks of genes, with their mutation meaning energy is not consumed by unrequired gene expression, thus providing an advantage in the test environment. Genes involved in motility and chemotaxis classified as AD are discussed in detail in *Importance of Motility and Chemotaxis during Symbiosis*.

Bacteroid AD genes (22) include another ECF sigma factor, EcfO (pRL110418), together with nearby gene pRL110416 (*rhaI*) encoding a rhamnose isomerase involved in competition for nodulation in *R. leguminosarum* bv. *trifolii* (67). Contiguous gene pRL110415 (*rhaD*) encodes a dehydrogenase required for nodulation (*SI Appendix*, Table S6), although the contiguous rhamnose transporter (pRL110413 to pRL110410), which might have been expected to be required for nodulation based on the results with *R. leguminosarum* bv. *trifolii*, is not required at any stage examined (Dataset S1). This is presumably because there is redundancy in Rlv3841 where another uptake system (most likely a CUT2) imports rhamnose, but the metabolism carried out by pRL110415 is needed for successful nodule formation.

Importance of Motility and Chemotaxis during Symbiosis. The genome of Rlv3841 contains up to 93 genes with putative roles in chemotaxis and motility, arranged in 7 clusters of 2 genes or more, and 28 single genes (*SI Appendix*, Table S10). Genes encoding putative proteins for chemotaxis (*che* [16 genes]), methyl-accepting chemotaxis receptors (*mcp* [24 genes] and *mcr* [3 genes]), and flagellar biosynthesis and motor function (*fla* [10 genes], *flg* [16 genes], *flh* [2 genes], *fli* [12 genes] and *mot* [5 genes]) have a variety of classifications throughout symbiosis, although the majority seem to fall into 3 clear categories: either not required for symbiosis (39 genes), required for nodule formation and function (nodule-general [39 genes]; *SI Appendix*, Table S6), or their mutation is advantageous in some way for growth in the rhizosphere and root colonization (5 genes) (*SI Appendix*, Table S10).While 10 motility-related genes are encoded on plasmids in Rlv3841 (7 on pRL12, and one each on pRL8, RL10, and pRL11), only chromosomally encoded motility-related genes are required for symbiosis.

Each of the individual genes in *che1* cluster (RL0686 to RL0693 [*icpAcheX1Y1A1W1R1B1Y1D1*]) is classified as nodule-general (*SI Appendix*, Tables S6 and S10), which means that they are all needed for nodule function and also, perhaps counter intuitively, that the cluster is not required for optimal colonization of the rhizosphere or roots (Fig. 3). This confirms the results of Miller et al., who found the *che1* cluster promotes competitive nodulation of peas (68). However, the INSeq results also suggest that Che1 is either required for chemotaxis specifically to root hairs or early progression into infection threads but is not essential for colonization of the root surface. In contrast, all genes in the *che2* cluster RL4036 to RL4028 (*cheY3A2W2mcrCBA-cheW3R2B2*) are classified AD in nodule bacteria (Fig. 3 and *SI Appendix*, Tables S9 and S10). While the reason is not known, we speculate that the *che2* gene cluster is involved in the later stages of bacterial movement through infection threads. A mutation that delays progression from infection threads to differentiated bacteroids would result in their overrepresentation in the nodule bacteria library, leading to the genes' classification as AD.

Most flagellar assembly and basal body apparatus proteins are important in nodules, both in bacteroids and nodule bacteria (Fig. 3 and *SI Appendix*, Table S10). The gene encoding critical motor protein FliG (RL0700 [*fliG*]) is neutral across all stages of symbiosis; however, in contrast, the gene encoding an interacting integral motor protein MotA (RL0703) is required in nodule bacteria and bacteroids (Fig. 3). In addition, the genes encoding master flagella and chemotaxis regulator pair VisNR (RL0696 and RL0697) are required in nodule bacteria and bacteroids (Fig. 3), although the carboxy-terminal GerE-encoding domain of *visN* displays a dramatic increase in transposon insertions, suggesting its deletion does not impair VisN function and may even enhance it. Several flagellar biosynthesis-encoding genes required in nodule bacteria are not needed, or their mutation is even advantageous, in the rhizosphere and for root colonization. Nonmotile bacteria may become stranded in the rhizosphere or on the root surface. Overall, it suggests that under the experimental conditions, motility and chemotaxis are not essential for colonization of the root surface but have a particularly important role in the nodule infection pathway, for movement toward root hairs or, possibly, progression down infection threads (Fig. 3). However, bacterial chemotaxis has previously been shown to provide a competitive advantage in colonization of plant root surfaces (69, 70). To minimize preselection the INSeq library was inoculated from a frozen culture at stationary phase (likely nonmotile) onto sterile vermiculite-grown plants. However, growth in soil, in competition with a motile, complex microbiome, is likely to impose different motility and chemotaxis selection upon the bacteria, with loss of chemotaxis and motility likely to be deleterious for long-term survival in soil and root colonization (71).

Plant Rescue of Auxotrophic Rhizobia. While it is generally accepted that terminally differentiated bacteroids are mainly supplied with dicarboxylic acids, the metabolic status of rhizobia during infection remains largely enigmatic. Root exudates contain a large variety of amino acids, sugars, organic acids, purines, and pyrimidines (72, 73); however, it remains unknown what compounds are available in substantial amounts and used by rhizobia. Therefore, we compared INSeq results from Rlv3841 grown in minimal medium (62) with symbiosis to identify auxotrophic mutants rescued by plant-derived compounds (*SI Appendix*, Table S11). We identified the main biosynthetic pathways for 18 of 20 amino acids and key components for purine, pyrimidine, *S*-adenosyl-methionine, heme, vitamin B_{12}, CoA, riboflavin and flavin mononucleotide/flavin adenine dinucleotide, pyridoxal, and tetrahydrofolate biosynthesis (*SI Appendix*, Table S11). These pathways are required throughout all stages tested (i.e., rhizosphere-progressive), with the exception of purine biosynthesis, where initial steps to 5-aminoimidazole-4-carboxamide-1-β-D-ribofuranosyl 5′-monophosphate (AICAR) are only required in nodules and not in the rhizosphere or on roots (*SI Appendix*, Table S11). As the genes encoding the enzymes of these pathways are rhizosphere-progressive, root exudates presumably contain insufficient amounts of these compounds to sustain

Downloaded at University of Oxford on September 9, 2020

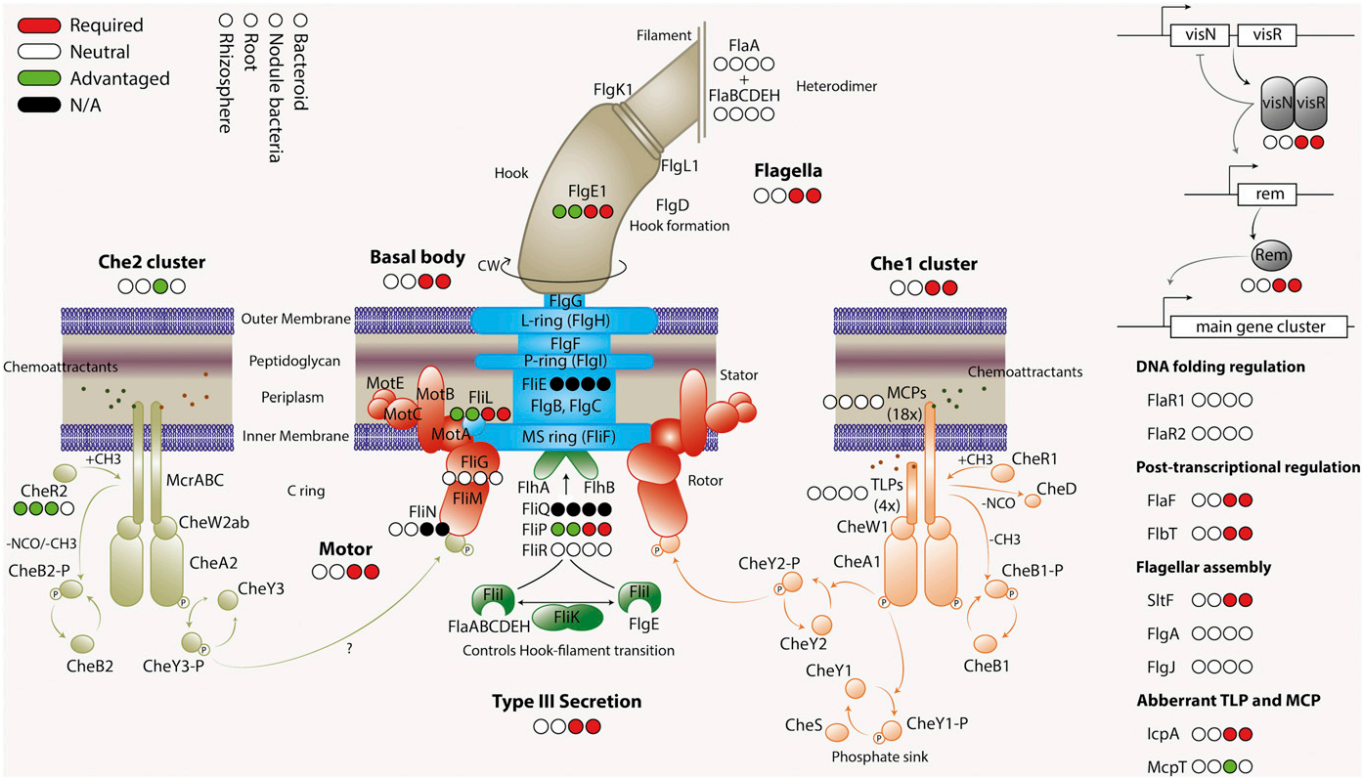


Fig. 3. Classification of genes encoding motility and chemotaxis proteins in rhizosphere, root, nodule bacteria, and bacteroid INSeq libraries. Genes required at a given stage are shown in red; genes where mutation has no effect (NE) are shown in white; genes that, when mutated, are advantageous are in green; and those unclassified due to too few TA sites are in black. (INSeq results are shown in *SI Appendix*, Table S10.) The proteins are colored by functional group, but the color is arbitrary with respect to INSeq phenotype.

competitive growth of these mutants. Only AICAR (or an equivalent metabolite) is released at sufficient levels into the rhizosphere and onto root surfaces to enable bacterial growth, however not in sufficient amounts for growth in infection threads or during bacteroid formation. Interestingly, several genes, NE in minimal or rich (TY) medium are required during stages of symbiosis. Rlv3841 has at least two glutamine synthases, and, of these, RL3549 encoding GlnII, is NE in minimal medium and in colonization of roots or rhizosphere but is ES/DE in nodules and nodule-recovered bacteria (i.e., nodule-general) (*SI Appendix*, Table S6). This is intriguing and suggests that increased glutamine synthesis may be required during infection. Similarly, there are two redundant copies of *argG* (RL2987 and RL4515). ArgG catalyzes the condensation of citrulline and aspartate to argininosuccinate in the urea cycle as part of arginine biosynthesis. As RL4515 is nodule-general (*SI Appendix*, Table S6) but is NE in minimal medium or during colonization, this suggests an increased role for arginine synthesis or degradation during bacteroid formation or function. Intriguingly, urea is also derived from arginine (see above) and urea catabolism is classified nodule-general. This highlights a potential role for arginine and urea catabolism via allophanate (urea-1-carboxylate) in nodule bacteria and/or bacteroids.

Histidine biosynthesis requires formation of imidazole-glycerole-3-phosphate from phosphoribulosyl-formimino-AICAR-phosphate, catalyzed by heterodimeric HisFH. Two copies are encoded in the genome, one of which (RL0042/RL0046) is required for growth in the input conditions. Of the second copy, RL0819 is nodule-general (*SI Appendix*, Table S6), and RL0820 is rhizosphere-progressive. This indicates an increased need for these reaction steps or possibly an incompatibility with other enzymes under N_2-fixing conditions. Glutathione-biosynthesis genes (RL0855 [*gshA*], RL0338 [*gshB*], and RL1571 [*pepA*]; *SI Appendix*, Table S11) are rhizosphere-progressive (*SI Appendix*, Table S2) but are dispensable in minimal medium, indicating that when interacting with plants, rhizobia must cope with increased oxidative stress (74, 75).

Validation of INSeq Predictions. Fifteen genes identified by INSeq as required in different stages of colonization and symbiosis were mutated by pK19 insertion, and their mutant phenotypes were determined experimentally (Table 1). Mutants in genes classified as root-progressive (pRL100053 and pRL80032) showed reduced colonization of roots compared with wild type (Fig. 4 *A* and *B*) but no significant reduction in survival in the rhizosphere (wild type, 2.88×10^7 cfu/mL [SEM: 3.84×10^6; $n = 15$; cfu, colony-forming units], and mutant in pRL100053, 1.81×10^7 cfu/mL [SEM: 2.57×10^6; $n = 8$]; wild type, 2.13×10^8 cfu/mL [SEM: 3.20×10^7; $n = 3$], and mutant in pRL80032, 2.22×10^8 cfu/mL [SEM: 3.10×10^7; $n = 3$]). Mutants in genes classified as nodule-general, together with those specific for nodule bacteria or bacteroids (Table 1), were tested experimentally for the ability to nodulate and to fix nitrogen in single-strain inoculations. Of these 13 mutants, 10 nodulated peas at levels similar to wild type (Fig. 4*C*). The exceptions were mutants in *pssD*, a nodule-general mutant that formed no nodules, and *nifB* (nodule bacteroid-specific) and *fixB* (nodule-general), which induced approximately twice as many, but small white, nodules as wild type (Fig. 4*C* and *SI Appendix*, Fig. S2). Assessing nitrogenase activity, the *nifB*, *fixB*, and *pssD* mutants had no detectable acetylene reduction, while all other mutants showed activity similar (approximately 80 to 130%) to that of wild type (Table 1 and Fig. 4*D*). These three mutant phenotypes show clearly that *pssD*, *nifB*, and *fixB* are required for nodulation and N_2 fixation. It is also apparent why mutants in these genes have been lost

Downloaded at University of Oxford on September 9, 2020

Table 1. Phenotype of mutants affected at different stages of symbiosis

Gene mutated	Description	Experiment					
		Rhizosphere*	Root†	Nodules‡	C_2H_2 reduction§	Nodule occupancy¶	Regrowth from nodules#
Root-progressive							
pRL100053	Transmembrane protein	++	+	++	ND	ND	ND
pRL80032	LysR family transcriptional regulator	++	+	—	ND	ND	ND
Nodule-general							
pRL100199 (*fixB*)	Electron-transfer protein FixB	ND	ND	+++	—	ND	ND
pRL120205 (*eryB*)	Erythritol phosphate dehydrogenase	ND	ND	++	++	—	ND
pRL90058 (*rmrR*)	TetR family transcriptional regulator of RND family efflux transporter	ND	ND	++	++	+	ND
RL0688 (*cheA1*)	Chemotaxis protein CheA	ND	ND	++	++	+	ND
RL0149	XRE family transcriptional regulator	ND	ND	++	++	+	ND
RL1545	Short-chain dehydrogenase/reductase	ND	ND	++	++	++	ND
RL2606 (*purC1*)	Saicar synthetase	ND	ND	++	++	+	ND
RL3549 (*glnII*)	Glutamine synthetase II	ND	ND	++	++	+	ND
RL3654 (*pssD*)	Polysaccharide biosynthesis protein	ND	ND	—	—	—	ND
Nodule bacteria							
pRL120198	L-xylulose kinase	ND	ND	++	++	ND	+
RL1496 (*iolR*)	SIS (sugar isomerase)-RpiR family transcriptional regulator	ND	ND	++	++	ND	+
RL3986 (*ruvC*)	Holliday junction endoDNase RuvC	ND	ND	++	++	ND	+
Bacteroid							
pRL100195 (*nifB*)	FeMo cofactor biosynthesis protein NifB	ND	ND	+++	—	ND	ND

*Rhizosphere colonization by a single strain.
†Root colonization by a single strain.
‡Nodule formation by a single strain.
§C_2H_2 (acetylene) reduction when peas inoculated with a single strain.
¶Nodule occupancy when in competition with wild type bacteria.
#Regrowth of bacteria from crushed nodules. ++, wild-type levels; +, reduced levels; +++, increased levels; —, no measurable activity; ND, not determined.

from nodule bacteria and bacteroid libraries (Fig. 4*C*), as less DNA would be present in the nodules collected from strains that form smaller nodules (*nifB*, *fixB*) or are unable to form nodules at all (*pssD*). The reason for the classification of the other seven genes as nodule-general (Table 1) was unclear until experiments were performed to test their mutant's ability to compete against wild type to form nodules. Mutant strains (marked with *gusA*) were competed against wild-type *celB*-marked bacteria by inoculation in a 1:1 ratio (*SI Appendix*, Fig. S3). Six of the seven nodule-general mutants were significantly impaired in competitive nodulation compared with wild type (Fig. 4*E*). A competitive disadvantage was evident despite the fact that a 1:1 inoculum provides significantly less competition than the INSeq environment, where up to 10^5 strains, with mutations in any 1 of over 6,000 genes (6,429/6,656 classified as NE in input1/input2), compete to form nodules. For these six strains, we have shown that the gene mutated in each case is required for effective symbiosis, and these INSeq experiments have identified the stage at which it is needed. Only a single predicted nodule-general strain (RL1545 mutant) showed no reduction in nodule occupancy in inoculation at this competition ratio, and we can speculate that its deficiency (it encodes a dehydrogenase/reductase) may only become apparent under stiffer competition conditions.

For genes classified as nodule bacteria-specific, we tested regrowth in TY medium since the initial preparation of this library involved a recovery step of growth in TY medium after nodule crushing. Wild type and mutants in three nodule bacteria-specific genes (pRL120198, RL1496 [*iolR*], and RL3986 [*ruvC*]) were grown in TY after extraction from nodules and compared. All mutants had a significantly lower rate of regrowth from crushed nodules than wild type (Fig. 4*F*). While this explains the low recovery of each of these mutants in this library, we did not identify any genes that increased the ratio of bacteria to bacteroids in nodules in such a small sample set. Future studies using flow cytometry to measure this ratio could be a powerful approach in such analyses.

Conclusion

Here, we present an in-depth investigation of N_2-fixing bacterial interactions with a host legume during symbiosis by examining bacterial gene requirements at different stages of rhizosphere growth, root colonization, and nodulation. Key observations are summarized (Fig. 5), illustrating the progressive nature of many mutations impairing effective symbiosis. It emphasizes that rhizobia experience many different soil and plant environments and is consistent with the Rhizobiales being a large family with many members being symbiotic or commensal bacteria (76). The

Downloaded at University of Oxford on September 9, 2020

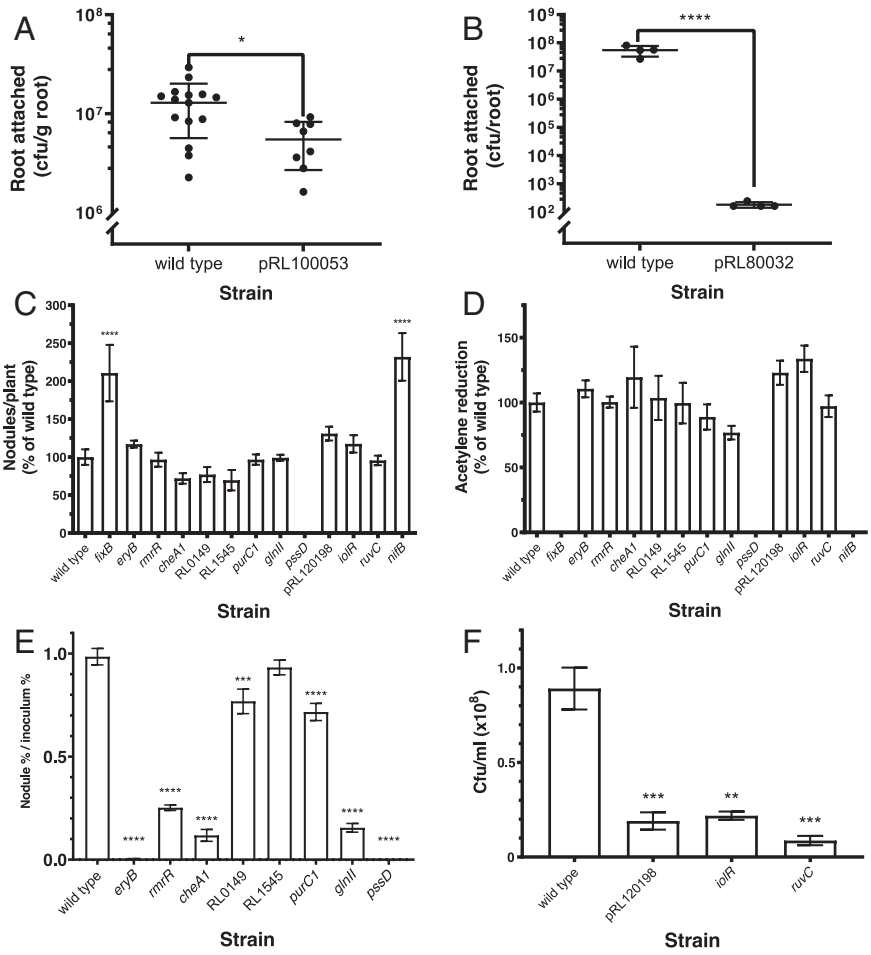


Fig. 4. Experimental characterization of mutant phenotypes predicted from INSeq data. (*A*) Root colonization by bacteria (cfu/g root) from a single-strain inoculation (10^5 cfu) of a 7-d-old pea plant comparing wild type and a mutant in pRL100053 (classified as root-progressive) at 5 dpi ($n \geq 8$). An unpaired *t* test was used to assess statistical significance. (*B*) Root colonization by bacteria (cfu/root) from a single-strain inoculation (10^5 cfu) of a 7-d-old pea plant comparing wild type and a mutant in pRL80032 (root-progressive) at 5 dpi ($n \geq 4$). An unpaired *t* test was used to assess statistical significance. (*C*) Nodule formation on peas (nodules/plant as a percentage of wild type [Rlv3841GusA, 180 ±19.2 nodules per plant]) from a single-strain inoculation (10^4 cfu) for mutants in nodule-general genes (9), nodule bacteria genes (3), and a bacteroid gene (Table 1) at 28 dpi ($n \geq 3$). A Dunnett's test comparing each mutant to wild type was used to assess statistical significance. (*D*) Acetylene reduction as a percentage of wild type (Rlv3841GusA, 5.18 ± 0.380 μmol of ethylene per plant per hour) of mutants in nodule-general genes (9), nodule bacteria genes (3), and a bacteroid gene (Table 1) at 28 dpi ($n \geq 3$). A Dunnett's test comparing each mutant to wild type was used to assess statistical significance. (*E*) Competition for nodule occupancy of mutants with wild type (nodule percentage/inoculum percentage) from 1:1 (mutant:wild type) coinoculation (total, 10^4 cfu) of pea plants with eight nodule-general mutants harvested at 21 dpi ($n \geq 5$). Statistical significance was assessed by a mixed-effects model comparing the proportion of each strain in inoculum and nodules. (*F*) Bacterial recovery (cfu/mL [$\times 10^8$]) from crushed nodules (10 per plant) and 36 h growth in TY broth at 28 °C for wild type and three nodule bacteria mutants ($n \geq 3$). To assess statistical significance, a Dunnett's test comparing each mutant strain to wild type was used. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$. Error bars show ±SEM.

power of INSeq is that it enables identification of those genes required when in competition with other bacteria. This is important when investigating environmental competition, since a mutation may not be deleterious for subsequent nodule formation or effective symbiosis in the artificial environment of inoculation with a single strain. It is sobering to appreciate that each of the 603 genetic regions required for competitive nodulation is individually just as important for a successful symbiosis as the structural components of nitrogenase. This INSeq dataset provides a powerful resource for better understanding the genetic basis of the crucially important N_2-fixing rhizobia–legume symbioses.

Materials and Methods

Bacterial and Plant Growth. Bacterial strains and plasmids shown in *SI Appendix*, Table S12 were grown as described previously (64). Mutant strains in *R. leguminosarum* were constructed as described previously (64), except ligation was performed using a BD In-Fusion cloning kit (Clontech), according to the manufacturer's instructions (primers listed in *SI Appendix*, Table S12). Pea plants (*Pisum sativum* cv. Avola) were grown as described (77).

INSeq Library Construction and Sequencing. Construction of the mariner library for INSeq experiments was as described previously (62), producing input libraries from two independent conjugation pools, input1 and input2. Seven-day-old peas grown in boiling tubes were inoculated with 10^5-cfu mariner-insertion mutants (input1). Rhizosphere and root-colonization samples were recovered from these at 5 dpi. Pea seedlings were grown in boiling tubes whose diameter allowed the rooting system to fully expand throughout the space; the vermiculite surrounding the roots has been defined as the rhizosphere in this study. To recover the rhizosphere population, the vermiculite surrounding the roots was separated, and 10 mL of phosphate-buffered saline (PBS) was added, vortexed (1 min), and filtered through four layers of muslin. To recover the root-colonized population, the root, snipped from the cotyledons, was vortexed (1 min) with 20 mL of PBS, ground by pestle and mortar, and filtered as above. Both were grown on TY (12 h) and spun to pellet cells before DNA extraction and library preparation. Seven-day-old peas grown in 2-L pots were inoculated with 10^6 cfu mariner-insertion mutants (input2). Nodule bacteria and bacteroid samples were

Downloaded at University of Oxford on September 9, 2020

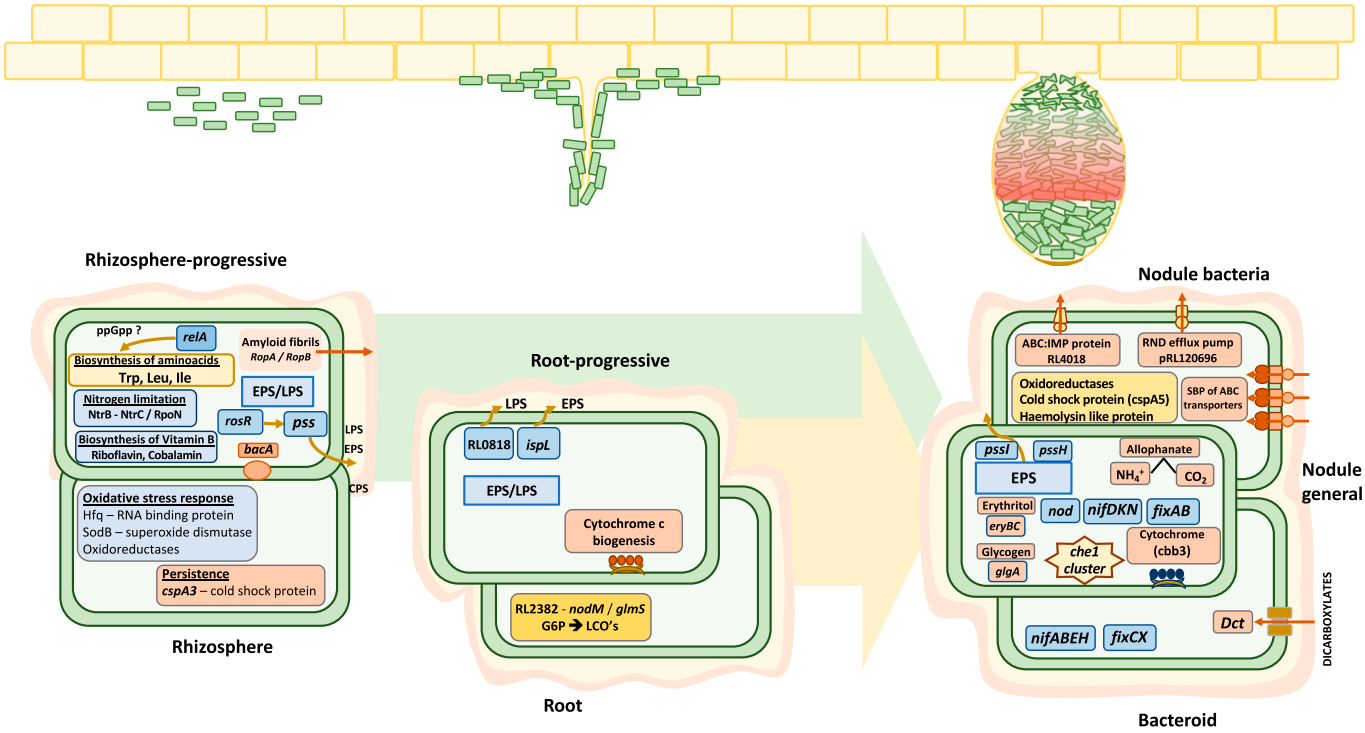


Fig. 5. Rlv3841 genes required at different stages of symbiosis with pea from INSeq analysis. Three environments (rhizosphere, root, and nodule) are shown. Progressive genes alter competition at subsequent stages. Gene requirements at each stage are discussed in *Results*, and full gene lists can be found in *SI Appendix*, Tables S1–S8.

recovered from these at 28 dpi. A total of ∼150,000 nodules were harvested from 1,500 plants and libraries prepared from 1) extract regrown in TY (12 h) before DNA preparation (nodule bacteria) and 2) DNA extracted from nodules (bacteroids). A total of six INSeq libraries (input1, input 2, the rhizosphere, roots, nodule bacteria, and bacteroids) were prepared and sequenced as described previously on an Ion Torrent Proton (62).

Transposon Insertion Analysis with a Four-State HMM. For sequencing data analysis, sequencing reads were uploaded onto a Linux server, and read-quality filtering and mapping (78, 79) and conversion of mapped data to wig files (for each replicon) were as described previously (62). Wig files were analyzed by TRANSIT software (80) using an HMM analysis method as described in ref. 25. In short, the HMM algorithm uses read-mapping statistics for all of the genes in a replicon and assigns one of four state classifications for each gene (ES, DE, NE, AD) (25, 80). In simplified form, ES genes should tolerate zero or very few insertions, and NE genes will sit within a mean parameter of insertions that do not excel past the boundaries set for DE or AD. For the DE state, the boundary is set at less than 1/100th of the read counts for a statistically significant consecutive stretch of insertion sites (25, 80). For the AD state, the boundary is set at over five times the mean read count for a statistically significant consecutive stretch of insertion sites (25, 80). There were a small number of TA sites where insertions gave very high AD read counts that misled the HMM to predict all subsequent TA sites as ES, as these very high read counts skewed the local mean read count. The values for these TA sites were reduced to next local-highest AD TA read count, and it was these wig files that were used for analysis. In TRANSIT, two analysis modes were used: 1) whole gene and 2) core region (considering 10 to 90% of gene, ignoring 10% at 5′ and 3′ ends), with the following parameters: method, HMM (to assign gene status) (25); and normalization, trimmed total reads and corrected for genome positional bias (Dataset S1). Gene status assignment differed by <0.85% between the two analyses. The 25 gene features that lacked TA sites in their core region were excluded from core-region analysis, and whole-gene analysis was used instead. For disagreements between the two analyses, if a gene was designated ES/DE/AD by whole-gene analysis but NE by core region analysis, then the final designation was based on the whole-gene analysis (i.e., ES/DE/AD). However, when assignment by core-region analysis was not NE, then read counts mapped to genes were checked in Integrated Genome Viewer (81), and designation was manually assigned. A fitness value (FV) for each gene was then calculated from the HMM analysis mapping statistics (62). FV was calculated as follows: FV = (number of insertion sites [TA] in a gene) × (mean insertion density of a gene) × (mean read count of a gene). The FV per gene was then standardized per million library reads (Dataset S1). FV gives a representation of the total gene mutants in a population and a numerical quantification for how well a gene may tolerate insertion (62); however, it should be appreciated that FV does not possess the same statistical rigor for analyzing transposon-insertion sequencing data as the HMM method. The HMM outputs for each gene status classification are provided in Dataset S2.

Characterization of Bacterial Mutants. Root colonization was assessed at 5 dpi from inoculation of 7-d-old peas with 10^5 cfu of bacteria. The root, snipped from the seed, was vortexed (1 min) with PBS (25 mL) before being ground (as above) and resuspended in PBS (10 mL). The rhizosphere population was assessed by adding PBS (30 mL) and vortexing thoroughly. Serial dilution and plating on appropriate antibiotics were used to calculate cfu. Acetylene-reduction assays (14) of seed-inoculated peas were performed at 28 dpi, and the same roots, subsequently stained for GusA activity using 5-bromo-4-chloro-3-indolyl-β-D-glucuronide cyclohexylammonium salt (82), were photographed to count blue-stained nodules. Competition between two strains to form nodules was assessed at 21 dpi on seed-inoculated peas (total inoculum, 10^4 cfu) by double-sequential staining for GusA activity using Magenta-GlcA and CelB activity with 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (82), to count magenta- and blue-stained nodules, respectively. Recovery of bacteria from nodules was assessed by picking the 10 largest nodules from a root, which were surface-sterilized (5 min in 2% bleach, followed by 6× washes in sterile distilled water) and crushed in 100 μL of PSM buffer (64) in a sterile 1.5-mL Eppendorf with a plastic pestle, before increasing the total volume to 1 mL of PSM and vortexing thoroughly; 25 μL was used to inoculate 50 mL of TY broth and an optical density at 600 nm (OD_{600}) followed for 36 h of growth, with shaking (140 rpm) at 28 °C (OD_{600} of 1 = 8×10^8 cfu).

Data Availability. The data supporting the findings of the study are available in this article and its *SI Appendix*.

ACKNOWLEDGMENTS. This work was supported by Natural Environment Research Council Grant NE/L501530/1, Biotechnology and Biological Sciences Research Council Grants BB/M011224/1 and BB/N013387/1, and Swiss National Science Foundation Postdoc.Mobility Grant 183901 (to R.L.).

Downloaded at University of Oxford on September 9, 2020

1. D. Fowler *et al*., The global nitrogen cycle in the twenty-first century. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20130164 (2013).
2. S. Fields, Global nitrogen: Cycling out of control. *Environ. Health Perspect.* **112**, A556–A563 (2004).
3. J. N. Galloway *et al*., Transformation of the nitrogen cycle: Recent trends, questions, and potential solutions. *Science* **320**, 889–892 (2008).
4. P. Poole, D. Allaway, Carbon and nitrogen metabolism in *Rhizobium*. *Adv. Microb. Physiol.* **43**, 117–163 (2000).
5. R. W. F. Hardy, U. D. Havelka, Nitrogen fixation research: A key to world food? *Science* **188**, 633–643 (1975).
6. M. Udvardi, P. S. Poole, Transport and metabolism in legume-rhizobia symbioses. *Annu. Rev. Plant Biol.* **64**, 781–805 (2013).
7. G. E. Oldroyd, J. D. Murray, P. S. Poole, J. A. Downie, The rules of engagement in the legume-rhizobial symbiosis. *Annu. Rev. Genet.* **45**, 119–144 (2011).
8. J. A. Downie, The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. *FEMS Microbiol. Rev.* **34**, 150–170 (2010).
9. P. Poole, V. Ramachandran, J. Terpolilli, Rhizobia: From saprophytes to endosymbionts. *Nat. Rev. Microbiol.* **16**, 291–303 (2018).
10. S. Radutoiu *et al*., LysM domains mediate lipochitin-oligosaccharide recognition and Nfr genes extend the symbiotic host range. *EMBO J.* **26**, 3923–3935 (2007).
11. V. D. Via, M. E. Zanetti, F. Blanco, How legumes recognize rhizobia. *Plant Signal. Behav.* **11**, e1120396 (2016).
12. D. J. Gage, Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiol. Mol. Biol. Rev.* **68**, 280–300 (2004).
13. J. White, J. Prell, E. K. James, P. Poole, Nutrient sharing between symbionts. *Plant Physiol.* **144**, 604–614 (2007).
14. D. Allaway *et al*., Identification of alanine dehydrogenase and its role in mixed secretion of ammonium and alanine by pea bacteroids. *Mol. Microbiol.* **36**, 508–515 (2000).
15. T. S. Walker, H. P. Bais, E. Grotewold, J. M. Vivanco, Root exudation and rhizosphere biology. *Plant Physiol.* **132**, 44–51 (2003).
16. E. M. Estabrook, J. I. Yoder, Plant-plant communications: Rhizosphere signaling between parasitic angiosperms and their hosts. *Plant Physiol.* **116**, 1–7 (1998).
17. R. M. Wheatley, P. S. Poole, Mechanisms of bacterial attachment to roots. *FEMS Microbiol. Rev.* **42**, 448–461 (2018).
18. L. M. C. Alves, J. A. M. de Souza, A. de Mello Varani, E. G. de Macedo Lemos, *The Family Rhizobiaceae*, (Springer, 2014), pp. 419–437.
19. J. P. W. Young *et al*., The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol.* **7**, R34 (2006).
20. L. Barquist, C. J. Boinett, A. K. Cain, Approaches to querying bacterial genomes with transposon-insertion sequencing. *RNA Biol.* **10**, 1161–1169 (2013).
21. A. K. Cain *et al*., A decade of advances in transposon-insertion sequencing. *Nat. Rev. Genet.* **21**, 526–540 (2020).
22. M. C. Chao, S. Abel, B. M. Davis, M. K. Waldor, The design and analysis of transposon insertion sequencing experiments. *Nat. Rev. Microbiol.* **14**, 119–128 (2016).
23. A. L. Goodman, M. Wu, J. I. Gordon, Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries. *Nat. Protoc.* **6**, 1969–1980 (2011).
24. B. J. Perry, C. K. Yost, Construction of a mariner-based transposon vector for use in insertion sequence mutagenesis in selected members of the Rhizobiaceae. *BMC Microbiol.* **14**, 298 (2014).
25. M. A. DeJesus, T. R. Ioerger, A Hidden Markov Model for identifying essential and growth-defect regions in bacterial genomes from transposon insertion sequencing data. *BMC Bioinf.* **14**, 303 (2013).
26. A. Westhoek *et al*., Policing the legume-*Rhizobium* symbiosis: A critical test of partner choice. *Sci. Rep.* **7**, 1419 (2017).
27. C. Bhattacharyya *et al*., Genome-guided insights into the plant growth promotion capabilities of the physiologically versatile *Bacillus aryabhattai* strain AB211. *Front. Microbiol.* **8**, 411 (2017).
28. R. F. Doornbos, L. C. van Loon, P. A. Bakker, Impact of root exudates and plant defense signaling on bacterial communities in the rhizosphere. A review. *Agron. Sustain. Dev.* **32**, 227–243 (2012).
29. V. A. García Angulo *et al*., Identification and characterization of RibN, a novel family of riboflavin transporters from *Rhizobium leguminosarum* and other proteobacteria. *J. Bacteriol.* **195**, 4611–4619 (2013).
30. M. Janczarek, K. Rachwał, J. Kopcińska, Genetic characterization of the Pss region and the role of PssS in exopolysaccharide production and symbiosis of *Rhizobium leguminosarum* bv. *trifolii* with clover. *Plant Soil* **396**, 257–275 (2015).
31. M. Janczarek, J. Jaroszuk-Sciseł, A. Skorupska, Multiple copies of rosR and pssA genes enhance exopolysaccharide production, symbiotic competitiveness and clover nodulation in Rhizobium leguminosarum bv. trifolii. *Antonie van Leeuwenhoek* **96**, 471–486 (2009).
32. M. Janczarek, J. Kutkowska, T. Piersiak, A. Skorupska, *Rhizobium leguminosarum* bv. *trifolii rosR* is required for interaction with clover, biofilm formation and adaptation to the environment. *BMC Microbiol.* **10**, 284 (2010).
33. A. O. Kosolapova *et al*., Two novel amyloid proteins, RopA and RopB, from the root nodule bacterium *Rhizobium leguminosarum*. *Biomolecules* **9**, 694 (2019).
34. H. P. Roest, C. J. P. Bloemendaal, C. A. Wijffelman, B. J. J. Lugtenberg, Isolation and characterization of *ropA* homologous genes from *Rhizobium leguminosarum* biovars *viciae* and *trifolii*. *J. Bacteriol.* **177**, 4985–4991 (1995).
35. S. Suzuki *et al*., Rhizobial factors required for stem nodule maturation and maintenance in *Sesbania rostrata-Azorhizobium caulinodans* ORS571 symbiosis. *Appl. Environ. Microbiol.* **73**, 6650–6659 (2007).
36. C. W. Ronson, B. T. Nixon, L. M. Albright, F. M. Ausubel, Rhizobium meliloti ntrA (rpoN) gene is required for diverse metabolic functions. *J. Bacteriol.* **169**, 2424–2431 (1987).
37. R. Karunakaran *et al*., BacA is essential for bacteroid development in nodules of galegoid, but not phaseoloid, legumes. *J. Bacteriol.* **192**, 2920–2928 (2010).
38. M. F. F. Arnold *et al*., Genome-wide sensitivity analysis of the microsymbiont *Sinorhizobium meliloti* to symbiotically important, defensin-like host peptides. *MBio* **8**, e01060-17 (2017).
39. A. F. Haag *et al*., Molecular insights into bacteroid development during *Rhizobium*-legume symbiosis. *FEMS Microbiol. Rev.* **37**, 364–383 (2013).
40. P. De Troch, J. Vanderleyden, Surface properties and motility of *Rhizobium* and *Azospirillum* in relation to plant root attachment. *Microb. Ecol.* **32**, 149–169 (1996).
41. M. Krehenbrink, J. A. Downie, Identification of protein secretion systems and novel secreted proteins in *Rhizobium leguminosarum* bv. *viciae*. *BMC Genom.* **9**, 55 (2008).
42. C. Marie, M. A. Barny, J. A. Downie, Rhizobium leguminosarum has two glucosamine synthases, GlmS and NodM, required for nodulation and development of nitrogen-fixing nodules. *Mol. Microbiol.* **6**, 843–851 (1992).
43. J. I. Quelas *et al*., Lack of galactose or galacturonic acid in *Bradyrhizobium japonicum* USDA 110 exopolysaccharide leads to different symbiotic responses in soybean. *Mol. Plant Microbe Interact.* **23**, 1592–1604 (2010).
44. G. R. Aguilar, M. Soberón, Cloning and sequence analysis of the *Rhizobium etli ccmA* and *ccmB* genes involved in c-type cytochrome biogenesis. *Gene* **182**, 129–135 (1996).
45. R. Geurts, T. Bisseling, *Rhizobium* nod factor perception and signalling. *Plant Cell* **14**, S239–S249 (2002).
46. L. C. Seefeldt, B. M. Hoffman, D. R. Dean, Mechanism of Mo-dependent nitrogenase. *Annu. Rev. Biochem.* **78**, 701–722 (2009).
47. M. Weidenhaupt, P. Rossi, C. Beck, H. M. Fischer, H. Hennecke, *Bradyrhizobium japonicum* possesses two discrete sets of electron transfer flavoprotein genes: *fixA, fixB* and *etfS, etfL*. *Arch. Microbiol.* **165**, 169–178 (1996).
48. R. W. Carlson, L. S. Forsberg, E. L. Kannenberg, *Lipopolysaccharides in Rhizobium-Legume Symbioses*, (Springer, 2010), pp. 339–386.
49. G. Cheng, N. Shapir, M. J. Sadowsky, L. P. Wackett, Allophanate hydrolase, not urease, functions in bacterial cyanuric acid metabolism. *Appl. Environ. Microbiol.* **71**, 4437–4445 (2005).
50. T. Kanamori, N. Kanou, H. Atomi, T. Imanaka, Enzymatic characterization of a prokaryotic urea carboxylase. *J. Bacteriol.* **186**, 2532–2539 (2004).
51. S. Tajima, T. A. Larue, Enzymes for acetaldehyde and ethanol formation in legume nodules. *Plant Physiol.* **70**, 388–392 (1982).
52. J. B. Peterson, T. A. LaRue, Soluble aldehyde dehydrogenase and metabolism of aldehydes by soybean bacteroids. *J. Bacteriol.* **151**, 1473–1484 (1982).
53. C. K. Yost, A. M. Rath, T. C. Noel, M. F. Hynes, Characterization of genes involved in erythritol catabolism in *Rhizobium leguminosarum* bv. *viciae*. *Microbiology* **152**, 2061–2074 (2006).
54. E. M. Lodwig *et al*., Amino-acid cycling drives nitrogen fixation in the legume-*Rhizobium* symbiosis. *Nature* **422**, 722–726 (2003).
55. P. Mergaert *et al*., Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5230–5235 (2006).
56. T. H. Mauchline *et al*., Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17933–17938 (2006).
57. H.-M. Fischer, Genetic regulation of nitrogen fixation in rhizobia. *Microbiol. Rev.* **58**, 352–386 (1994).
58. F. J. De Bruijn, "Biological nitrogen fixation" in *Principles of Plant-Microbe Interactions*, (Springer, 2015), pp. 215–224.
59. C. J. Reid, P. S. Poole, Roles of DctA and DctB in signal detection by the dicarboxylic acid transport system of *Rhizobium leguminosarum*. *J. Bacteriol.* **180**, 2660–2669 (1998).
60. S. N. Yurgel, M. L. Kahn, Dicarboxylate transport by rhizobia. *FEMS Microbiol. Rev.* **28**, 489–501 (2004).
61. C. W. Ronson, P. M. Astwood, J. A. Downie, Molecular cloning and genetic organization of C4-dicarboxylate transport genes from *Rhizobium leguminosarum*. *J. Bacteriol.* **160**, 903–909 (1984).
62. R. M. Wheatley *et al*., Role of O2 in the growth of Rhizobium leguminosarum bv. viciae 3841 on glucose and succinate. *J. Bacteriol.* **199**, e00572-16 (2016).
63. G. Mulley *et al*., Pyruvate is synthesized by two pathways in pea bacteroids with different efficiencies for nitrogen fixation. *J. Bacteriol.* **192**, 4944–4953 (2010).
64. R. Karunakaran *et al*., Transcriptomic analysis of *Rhizobium leguminosarum* biovar *viciae* in symbiosis with host plants *Pisum sativum* and *Vicia cracca*. *J. Bacteriol.* **191**, 4002–4014 (2009).
65. A. H. F. Hosie, D. Allaway, C. S. Galloway, H. A. Dunsby, P. S. Poole, *Rhizobium leguminosarum* has a second general amino acid permease with unusually broad substrate specificity and high similarity to branched-chain amino acid transporters (Bra/LIV) of the ABC family. *J. Bacteriol.* **184**, 4071–4080 (2002).
66. D. Borthakur, M. Soedarjo, P. M. Fox, D. T. Webb, The mid genes of Rhizobium sp strain TAL1145 are required for degradation of mimosine into 3-hydroxy-4-pyridone and are inducible by mimosine. *Microbiology* **149**, 537–546 (2003).
67. J. S. Richardson, M. F. Hynes, I. J. Oresnik, A genetic locus necessary for rhamnose uptake and catabolism in *Rhizobium leguminosarum* bv. *trifolii*. *J. Bacteriol.* **186**, 8433–8442 (2004).
68. L. D. Miller, C. K. Yost, M. F. Hynes, G. Alexandre, The major chemotaxis gene cluster of Rhizobium leguminosarum bv. viciae is essential for competitive nodulation. *Mol. Microbiol.* **63**, 348–362 (2007).
69. A. Kawasaki *et al*., Microbiome and exudates of the root and rhizosphere of brachypodium distachyon, a model for wheat. *PLoS One* **11**, e0164533 (2016).

Downloaded at University of Oxford on September 9, 2020

70. B. E. Scharf, M. F. Hynes, G. M. Alexandre, Chemotaxis signaling systems in model beneficial plant-bacteria associations. *Plant Mol. Biol.* **90**, 549–559 (2016).
71. G. A. Turnbull, J. A. W. Morgan, J. M. Whipps, J. R. Saunders, The role of motility in the in vitro attachment of *Pseudomonas putida* PaW8 to wheat roots. *FEMS Microbiol. Ecol.* **35**, 57–65 (2001).
72. E. T. Gaworzewska, M. J. Carlile, Positive chemotaxis of *Rhizobium leguminosarum* and other bacteria towards root exudates from legumes and other plants. *J. Gen. Microbiol.* **128**, 1179–1188 (1982).
73. F. Z. Haichar, C. Santaella, T. Heulin, W. Achouak, Root exudates mediated interactions belowground. *Soil Biol. Biochem.* **77**, 69–80 (2014).
74. R. Santos, D. Hérouart, S. Sigaud, D. Touati, A. Puppo, Oxidative burst in alfalfa-*Sinorhizobium meliloti* symbiotic interaction. *Mol. Plant Microbe Interact.* **14**, 86–89 (2001).
75. C. Chang, I. Damiani, A. Puppo, P. Frendo, Redox changes during the legume-*rhizobium* symbiosis. *Mol. Plant* **2**, 370–377 (2009).
76. R. Garrido-Oter *et al*.; AgBiome Team, Modular traits of the Rhizobiales Root microbiota and their evolutionary relationship with symbiotic Rhizobia. *Cell Host Microbe* **24**, 155.e5-167.e5 (2018).
77. R. Karunakaran *et al*., Thiamine is synthesized by a salvage pathway in *Rhizobium leguminosarum* bv. *viciae* strain 3841. *J. Bacteriol.* **188**, 6661–6668 (2006).
78. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
79. B. Langmead, C. Wilks, V. Antonescu, R. Charles, Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35**, 421–432 (2019).
80. M. A. DeJesus, C. Ambadipudi, R. Baker, C. Sassetti, T. R. Ioerger, TRANSIT–A software tool for Himar1 TnSeq analysis. *PLoS Comput. Biol.* **11**, e1004401 (2015).
81. J. T. Robinson *et al*., Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
82. C. Sánchez-Cañizares, J. Palacios, Construction of a marker system for the evaluation of competitiveness for legume nodulation in *Rhizobium* strains. *J. Microbiol. Methods* **92**, 246–249 (2013).

Downloaded at University of Oxford on September 9, 2020