The genetic control of root attachment and colonisation in *Rhizobium leguminosarum*.



Clare Cocker

Magdalen College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2023

Abstract

The genetic control of root attachment and colonisation in *Rhizobium leguminosarum*

Clare Cocker, Magdalen College

Submitted for the degree of DPhil., Michaelmas Term 2023

The legume-*Rhizobium* symbiosis is a key source of environmentally clean nitrogen in agriculture. Of up to 70 million tonnes of nitrogen that are fixed into agricultural systems by biological nitrogen fixation each year, 40 million are fixed by rhizobia in symbiosis with legumes. Attachment to and colonisation of host plant roots are essential precursors to formation of a successful nitrogen-fixing symbiosis between *Rhizobium* strains and their host legumes but are less well-characterized than other stages of symbiosis formation. This work investigates the genetic basis for root surface attachment and colonisation in *Rhizobium leguminosarum* on both a whole genome scale and an individual pathway scale.

Development of a differential spin method of separating attached and colonising bacteria from plant root material allows investigation of gene expression that is unique to bacteria that are interacting with the root surface, where previous studies have combined bacterial samples in bulk soil with those attached to the root surface. This method is used to investigate transcriptomic changes in *Rhizobium leguminosarum* bv *viciae* 3841 during colonisation of the barley rhizoplane, and attachment to and colonisation of the pea rhizoplane. Comparison of transcriptomic changes during colonisation of the pea and barley rhizoplanes reveals that Rlv3841 is substantially better adapted to growth on the pea rhizoplane. Downregulation of central carbon metabolism, synthesis of metabolic intermediates, cell division, DNA repair and cell wall synthesis during colonisation of barley roots suggest that Rlv3841 grows much more slowly on the barley rhizoplane than in laboratory culture, or on the pea rhizoplane, possibly due to a lack of adaptation to the carbon sources found in barley root exudates.

The comparison of the Rlv3841 transcriptome during attachment to and colonisation of the pea rhizoplane carried out in this work represents, to my knowledge, the first such study on a rhizobial diazotroph and its host legume. A substantial amount of overlap was shown between the significantly differentially regulated genes during attachment and colonisation transcriptomes in Rlv3841. Functions upregulated during both processes included ABC transport, DNA repair and genes involved in determining cell surface characteristics. Expression of genes involved in motility and chemotaxis was significantly reduced during both attachment and colonisation. However, around 40% of differentially regulated genes differed between the two processes, showing that distinct sets of genes are involved in each stage of interaction with the rhizoplane. In particular, genes involved in breakdown of malonate, GABA and fatty acids were significantly upregulated in attachment, while inositol catabolism was significantly upregulated during colonisation, suggesting that different carbon sources may be important for each process to take place.

Finally, this work characterises the *ecfE* operon, which is upregulated during attachment to and colonisation of the pea rhizoplane, as well as during colonisation of the barley rhizoplane. The promoter (P*lppE*) is only active in cells adhering to the root elongation zone, where rapid cell elongation happens, allowing *R. leguminosarum* to detect a region of the root with new surface area that is free of pre-existing microbial communities. This specific spatial activation happens in response to a root-associated glycoprotein, likely a germin-like protein, that is expressed to optimal levels in the root elongation zone. EcfE is revealed to play a role in mediating expression of an iron-related attachment pathway that facilitates the lifestyle switch between free-living and colonising bacteria.

Acknowledgements

I am grateful for my funding from the BBSRC through the Interdisciplinary Bioscience DTC, without which I would not have been able to carry out this work.

I feel very lucky to have met so many fantastic people over the course of the last four years at Oxford. Firstly, my supervisor, Phil Poole. Thank you so much for all your help and guidance over the last few years, and for your continual enthusiasm for rhizobial metabolism that is so inspiring. Thank you to all the members of the Poole lab who made this experience so enjoyable. When I was stressed out and trying to design a project during lockdown in 2020, Phil wrote that you were the 'best and friendliest lab' in Oxford, and he has been proved right countless times. To Isidro, for always taking the time to help me think through data analysis and next steps, and for joining me in uprooting 800 pea and barley plants, thank you. Thank you to Vinoy and Denis for your help and hard work on the *ecfE* project; Vinoy for introducing me to the project and for your contributions to the chapter and Denis for introducing me to the world of protein purifications. Thank you to Carmen for first introducing me to *Rhizobium* during my undergrad project and inspiring me to want to come back for more. Thank you to Bea, Hayley and Alison for the chats in the office and for always answering my random questions about protocols and databases. Thank you to Raphael for sharing your PCR wisdom and to Tim for your help with RNA library prep. I would like to thank Javi for your moral support during the first RNA extractions and Jessie for your positive energy during the attachment assay optimisation phase. Thank you to both of you and to all the 'kids' – Patrick, Bergthor, Ann-Kathrin, Tom, Olivia and Matt – for all the deep, funny and random chats and for making lab work so enjoyable. Thank you to Helen and Lida for all your support and taking such great care of the lab.

I would also like to thank my friends outside the lab for making the last four years so great. Everyone at MCBC, thank you for making early morning rowing so fun and unforgettable. Grace and Kathy, thank you for all the FaceTime DMC sessions and for always being keen to watch the latest Netflix romcom. My DTC cohort, for the many outdoor lockdown drinks in the park. Lisa, for all the scenic cycles. Ben

and Andrew, thank you for being such fantastic friends from the moment we were assigned seats together at the DTC. I will miss our weekly hot chocolate and debrief sessions so much, and I'll never forget to *not* include print() in a while loop in MATLAB.

I would like to thank Alice for always being there for me, letting me vent about peas and learning the names of all the lab equipment because I talk about it so much. Thank you for making my favourite meal at least twice a week every week while I've been writing my thesis and believing in me the whole time. You mean so much to me and I'm so glad my DPhil gave me the chance to meet you. And lastly, I would like to thank my family, especially my parents, for your continual support and encouragement to pursue my dreams. I couldn't have done this without you.

Statement of Authorship

This thesis is entirely my own work, with the exception of:

Chapter 3	RNA extracted from Rlv3841 during colonisation of pea roots and prepared and sequenced without Nycodenz (Fig. 3.2) by Rachel Wheatley
Chapter 5	RNA sequencing of Rlv3841 wild type and Ω*asfE* mutant grown on UMS by Vinoy Ramachandran
	Images of Rlv3841 on pea roots and agar plates in Figures 5.2 B, 5.2 C and 5.4 C by Vinoy Ramachandran
	Mass spectrometry analysis of P*lppE*-inducing pea root samples by Farnusch Kaschani at the Analytics Core Facility Essen
	Nodulation competition assay on Rlv3841 wild type and Ω*lppE* mutant by Vinoy Ramachandran

Abbreviations

ABC	ATP-binding cassette
AHL	N-acyl homoserine lactone
ATP	Adenosine triphosphate
BarSeq	Barcoded sequencing
BLAST	Basic local alignment search tool
BNF	Biological nitrogen fixation
bp	Base pair
BNF	Biological nitrogen fixation
CCaMK	Calcium- and calmodulin-dependent kinase
cDNA	Complementary DNA
CDS	Coding sequence
cfu	Colony-forming units
CoA	Coenzyme A
cps	Counts per second
CREK	Chitin receptor elicitor kinase
CUT	Carbohydrate uptake transporter
DIMBOA	2,4-dihydroxy-7-methoxy-2*H*-1,4-benzoxazin-3(4*H*)-one
DNA	Deoxyribonucleic acid
dpi	Days post inoculation

DUF	Domain of unknown function
DWA	Distilled water agar
ECF	Extracytoplasmic function
EDTA	Ethylenediaminetetraacetic acid
EPS	Exopolysaccharide
FeCT	Iron chelate uptake transporter
FeT	Ferric ion uptake transporter
FMS	Iron-supplemented universal minimal salts
GABA	γ-aminobutyrate
gDNA	Genomic DNA
GFP	Green fluorescent protein
Gm	Gentamicin
Gm^R	Gentamicin resistance
HAAT	Hydrophobic amino acid transporter
HiFi	High fidelity
hpi	Hours post inoculation
INSeq	Insertion sequencing
Kan	Kanamycin
Kan^R	Kanamycin resistance
LB	Lysogeny broth
LCO	Lipochitooligosaccharide

LFQ	Label-free quantification
LPS	Lipopolysaccharide
LRR	Leucine-rich repeat
MCP	Methyl-accepting chemotaxis protein
MZT	Manganese/zinc/iron chelate uptake transporter
N_2	Dinitrogen
NAD	Nicotinamide adenine dinucleotide
Neo	Neomycin
Neo^R	Neomycin resistance
nt	Nucleotide
PAAT	Polar amino acid transporter
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PGPR	Plant growth-promoting rhizobacteria
PHB	Polyhydroxybutarate
POPT	Polyamine/opine/phosphonate uptake transporter
QAT	Quaternary amine uptake transporter
Rap	*Rhizobium* adhering protein
REZ	Root elongation zone
RLU	Relative luminescence units
RNA	Ribonucleic acid

RNAseq	RNA sequencing
RT-qPCR	Reverse transcription quantitative PCR
SBP	Solute-binding protein
SEM	Standard error of mean
Spec	Spectinomycin
$Spec^R$	Spectinomycin resistance
Str	Streptomycin
Str^R	Streptomycin resistance
SulT	Sulphate/tungstate uptake transporter
SynCom	Synthetic community
TCA	Tricarboxylic acid
Tet	Tetracycline
Tet^R	Tetracycline resistance
TY	Tryptone yeast
UMS	Universal minimal salts
w/v	Mass of solute/volume of solution

Table of Contents

Abstract ..... ***i***

Acknowledgements ..... ***iii***

Statement of Authorship ..... ***v***

Abbreviations ..... ***vi***

1 Introduction ***1***

1.1 Nitrogen in agriculture **2**

1.1.1 The importance of nitrogen in crop growth 2

1.1.2 The nitrogen crisis 2

1.2 Biological nitrogen fixation **4**

1.2.1 The legume-*Rhizobium* mutualism 4

1.2 Colonisation of plant roots by *Rhizobium* strains **11**

1.3.1 How plants affect the rhizosphere 12

1.3.2 Chemotaxis 15

1.3.3 Attachment 16

1.3.4 Growth and biofilm formation 21

1.4 RNA sequencing **24**

1.4.1 Contribution of RNA sequencing to understanding of the legume-*Rhizobium* symbiosis 24

1.4.2 Potential pitfalls 27

1.5 Research objectives **27**

2 Materials and methods ***30***

2.1 Bacterial strains, plasmids and primers **31**

2.2 Media and antibiotics **44**

2.2.1 Media 44

2.2.2 Antibiotics 45

2.3 Molecular techniques 45

2.3.1 Isolation of genomic and plasmid DNA 45

2.3.2 DNA amplification by PCR 45

2.3.3 Agarose gel electrophoresis 47

2.3.4 Restriction endonuclease digestions, ligations and High Fidelity assembly 47

2.3.5 Transformation of competent cells 47

2.3.6 Conjugal transfer of plasmids 48

2.4 Plant growth for attachment and colonisation assays 48

2.4.1 Sterilisation of seeds 48

2.4.2 Preparation of boiling tubes 49

2.5 RNA sequencing 50

2.5.1 Preparation of 'attached' rhizoplane bacteria 50

2.5.2 Preparation of 'colonising' rhizoplane bacteria 50

2.5.3 Preparation of Nycodenz gradients 50

2.5.4 Harvesting of fluorescent-marked bacterial cells from plant root surfaces 51

2.5.5 Total RNA extraction and preparation for sequencing 51

2.5.6 RNA quantification by RT-qPCR 52

2.5.7 Bioinformatic analysis of RNAseq data 52

2.6 Assays using *lux* promoter fusions 53

2.6.1 Endpoint induction assays in liquid culture 53

2.6.2 Continuous induction assays in liquid culture 53

2.6.3 Induction assays on solid media 53

2.6.4 *In planta* induction assays 54

2.6.5 *In planta* induction timelapses 55

2.6.6 Root colonisation assays 55

2.6.7 Root tip attachment assays 55

2.7 Preparation and characterisation of plant protein extracts 56

2.7.1 Crude extraction of pea root protein 56

2.7.2 Purification of glycoproteins 56

2.7.3 Mass spectrometry 57

2.8 Colourimetric enzyme assays 57

2.8.1 Beta-glucuronidase activity assay 57

2.8.2 Alkaline phosphatase activity assay 58

2.8.3 Calculation of Miller units 58

2.9 Fahraeus slides 59

2.10 Bioinformatics, data handling and statistical analysis 59

3 Development and validation of RNA sequencing of plant-colonising bacteria 61

3.1 Introduction 62

3.1.1 Colonisation of the rhizoplane 62

3.1.2 RNA sequencing 64

3.2 Results and Discussion 67

3.2.1 Plant root material can be separated from bacterial cells using a Nycodenz gradient 67

3.2.2 Rlv3841 exhibits distinct transcriptomic signatures during colonisation of pea and barley 69

3.2.3 RNA sequencing results were confirmed by RT-qPCR 70

3.2.4 Transcriptomic changes in Rlv3841 common to colonisation of pea and barley 73

3.2.5 Transcriptomic changes unique to colonisation of the barley rhizoplane 87

3.2.6 Barley-specific substantially differentially regulated genes 96

3.2.7 Changes unique to pea 99

3.3 Summary 99

4 The Rlv3841 transcriptome during attachment to and colonisation of the pea rhizoplane 102

4.1 Introduction 103

4.2 Results and Discussion 104

4.2.1 Rlv3841 transcriptome changes during attachment to and colonisation of the pea rhizoplane 104

4.2.2 Rlv3841 genes upregulated in both processes 106

4.2.3 Rlv3841 genes downregulated in both processes 117

4.2.4 Rlv3841 genes upregulated only in attachment 121

4.2.5 Rlv3841 genes downregulated only in attachment 130

4.2.6 Rlv3841 genes upregulated only in colonisation 132

4.2.7 Rlv3841 genes downregulated only in colonisation 133

4.3 Summary 134

5 The role of the ecfE operon in pea rhizoplane attachment and colonisation 137

5.1 Introduction 138

5.2 Results 141

5.2.1 *lppE*, *ecfE* and *asfE* are expressed as an operon from a promoter upstream of the *lppE* coding sequence 141

5.2.2 The *lppE* promoter region is expressed in *Rhizobium* cells found at plant root elongation zones 143

5.2.3 Transcription from the primary promoter P*lppE* is controlled by EcfE and AsfE 144

5.2.4 EcfE has a broad range of transcriptional targets 148

5.2.5 LppE is transported to the periplasm and indirectly affects EcfE activity 154

5.2.6 P*lppE* expression is activated by a cell surface-associated signal 157

5.2.7 The pea root tip 'induction factor' is a glycoprotein 160

5.2.8 The *ecfE* operon is implicated in attachment to and colonisation of plant roots by *Rhizobium* 170

5.2.9 The *ecfE* operon is not implicated in Rlv3841 nodulation of pea 174

5.3 Discussion 175

6 General discussion and future directions 183

6.1 Overview 184

6.2 A reliable method for obtaining transcriptomic data from root-associated bacteria 184

6.3 Comparison of the Rlv3841 transcriptome during colonisation of pea and barley 186

6.4 The Rlv3841 transcriptome during attachment to and colonisation of host pea roots 187

6.5 Characterisation of the *ecfE* operon and its role in root attachment and colonisation 189

6.6 Future directions 193

6.6.1 Characterisation of the transcriptional targets of EcfE 193

6.6.2 The role of LppE 194

6.6.3 Characterisation of the P*lppE*-inducing glycoprotein 195

6.6.4 Link between iron and rhizoplane attachment and colonisation 196

6.6.5 The roles of ECF sigma factors in Rlv3841 attachment and colonisation of plant rhizoplanes 197

6.6.6 Engineering plant rhizoplane-induced expression of *nifA* 198

Bibliography 199

Appendix: Supplementary information 247

1 Introduction

1.1 Nitrogen in agriculture

1.1.1 The importance of nitrogen in crop growth

Nitrogen is the backbone of protein formation, accounting for about 16% of protein content by mass. As such, it is an essential element for growth of all organisms, including both humans and the plants we consume. For the last hundred years, it has been possible to industrially fix nitrogen from the atmosphere into ammonia using the Haber-Bosch process, which today accounts for 96% of synthetic ammonia [1], and which has allowed the nitrogen content of synthetic fertilisers to substantially increase. Adding nitrogen-rich fertilisers to crops allows for greater yields that support a larger population. Between 1908 and 2008, the number of people supported by one hectare of arable land increased from 1.9 to 4.3 [2], showing the power of boosting yields to facilitate population growth. It is estimated that between 30 and 50% of this yield increase is accounted for solely by the application of nitrogen fertiliser [3].

1.1.2 The nitrogen crisis

The growth in population supported by the significant increase in artificial nitrogen fixation and fertiliser application is ongoing. The UN estimates that the global population will reach 11 billion by 2100 [4], with demand for crops predicted to double by 2050 [5] but with existing agricultural systems unable to meet that need [6]. Nitrogen availability is the most significant limiting factor to growing agricultural production to match a growing population [7], and over the last forty years use of nitrogen fertiliser in agriculture has increased seven-fold [8] to try and match demand. However, this increase in use of synthetic fertiliser to drive increased yields is not sustainable. If current trends continue, global synthetic nitrogen use will increase to around 250 Mt yr^{-1} by 2050 [5], and the food supply chain is already responsible for 26% of anthropogenic greenhouse gas emissions [9], with 11% of this being directly caused by use of nitrogen fertiliser, amounting to emission of 1.13 $GtCO_2e$ in 2018 [10]. Overall, this means that the Haber-Bosch process itself is responsible for 1.2% of global anthropogenic greenhouse gas emissions [1], as it is dependent on fossil fuel consumption. Moreover, application of

synthetic fertiliser is also responsible for further, downstream emissions. Nitrous oxide is a greenhouse gas with around 300 times as much global warming potential of carbon dioxide [11] and is the largest stratospheric depleting substance [12]. Half of anthropogenic N_2O emissions come from microbial nitrification and denitrification of ammonia in agricultural soils [11], the majority of which has been added to soils with the goal of boosting yields. In regions with more intensive agricultural production this figure is greater, with 75% of N_2O emissions in North America coming from agriculture [13]. As synthetic nitrogen application to land increases, release of N_2O also increases exponentially [14]. All of this means that as agriculture is scaled to match population growth, primarily via the application of increased volumes of synthetically fixed nitrogen, its impact on global warming and climate change due to the emission of greenhouse gases will increase enormously.

In addition to its impact on global climate change through emission of greenhouse gases, application of synthetic ammonia to land can cause damage to local environments through the leaching of excess nitrogen away from agricultural land, particularly into aquatic ecosystems. The artificially increased nutrient availability in the water causes increased growth of algae and vegetation which over time results in an increase in dead organic matter sinking to the bottom of lakes and oceans. This dead matter promotes saprotrophic microbial activity, which in turn decreases availability of oxygen towards the bottom of water columns in a process called eutrophication. Food production is responsible for 78% of eutrophication globally [9]. In extreme cases, severe hypoxia causes the formation of 'dead zones' where marine life cannot survive. The number of these 'dead zones' found globally has approximately doubled each decade since the 1960s [15]. Synthetic fertilisers are also damaging to terrestrial environments, with agriculture being responsible for 32% of terrestrial acidification [9].

Finally, the prices of synthetic fertilisers are prohibitive for many farmers, particularly subsistence farmers, having risen by over 300% between 2005 and 2015 [16], due to being tied to the cost of fuel. Given the huge environmental and economic costs of synthetic fertilisers, it is vital that an alternative means is found of increasing staple crop yields to support a growing population.

1.2 Biological nitrogen fixation

Biological nitrogen fixation (BNF) is the process of fixing atmospheric dinitrogen into ammonia, making use of the bacterial enzyme nitrogenase [17]. Nitrogenase is a multi-subunit enzyme synthesised by *nif* gene products under control of the regulator NifA [18] and contains iron-sulphur and iron-molybdenum clusters which are the site of nitrogen fixation [17]. BNF is already a significant contributor of agricultural nitrogen, being responsible for injecting 50-70 million tonnes of nitrogen into agricultural systems every year [19]. However, the availability of BNF is limited, with only certain groups of prokaryotes and no eukaryotes able to synthesise nitrogenase. This means that to benefit from biologically fixed nitrogen plants must act in symbiosis with diazotrophic, or nitrogen-fixing, bacteria. Most plants that do this do so by hosting diazotrophic bacteria in specialised root nodules, each of which can house up to a billion bacterial cells [17]. However, only members of ten out of 380 families of angiosperms can do this, and all of these families are members of the Eurosid clade [20], meaning that the ability to benefit from BNF is very restricted. The one exception is plants of the family Gunneraceae, which can host nitrogen-fixing cyanobacteria in leaf glands, rather than root nodules, and are not part of this clade [20]. The most well-studied symbiotic relationship is that between legumes and rhizobia. Of up to 70 million tonnes of nitrogen that are fixed into agricultural systems by BNF each year, 40 million are fixed by rhizobia in symbiosis with legumes [21], making this relationship the most agriculturally significant, followed by the relationship between diazotrophs of the genus *Frankia* and their host actinorhizal plants [22].

1.2.1 The legume-*Rhizobium* mutualism

The most widely researched and agriculturally significant provider of biologically fixed nitrogen is the symbiotic relationship between legume plants of the family Fabaceae and diazotrophic rhizobia. It is a highly specific relationship, with each legume species only able to be nodulated by a small number of rhizobial strains [23]. Bacteria enter the plant cells and replicate inside them, while cell differentiation in the plant leads to the formation of nodule structures. Bacteria can either enter at places where the

epidermis is broken, such as at sites of lateral root formation, or can attach to root hairs, which they induce to curl around them and then grow into the root through infection threads [24]. Intercellular entry at lateral root bases often happens in legumes growing under flooding conditions, such as the semi-aquatic legume *Sesbania rostrata*, when root hair growth is inhibited [25], [26]. In land plants like pea, clover, vetch, alfalfa and *Medicago*, entry by root hair curling is the most common mode of entry [27]. Plants secrete signalling molecules called flavonoids from their roots, which diffuse through the soil and can passively cross bacterial cell membranes [28]. The flavonoids are perceived by the rhizobial transcriptional activator NodD, which activates expression of *nod* genes whose products synthesise Nod factors, which are lipochitooligosaccharides (LCOs) [29]. Different copies of NodD in different *Rhizobium* strains recognise different flavonoids, helping to create specificity between hosts and symbionts [30], [31]. Nod factors are then perceived by LysM receptor kinases on elongating root hair cells [32], [33], [34]. LysM receptor kinases can distinguish between Nod factors coming from different strains of rhizobia, as they are decorated with different combinations of acetyl, methyl, sulphuryl and glycosyl chains, adding another layer of specificity to the interaction [35]. Two motifs in regions II and IV, in the LysM1 domains of LysM receptors, determine ligand binding and confer specificity. These regions are conserved in legume chitin elicitor receptor kinases (CERKs) which bind chitin, which is structurally invariable, but vary between legume Nod factor receptors, which must recognise highly specific Nod factors that vary between bacterial strains [36]. Indeed, Nod factors are the primary determinants of host-strain specificity in legume-*Rhizobium* interactions [24]. Once bacteria come into contact with root hairs, the tip of the root hair begins to curl around the bacterial cells, creating a microcolony [27]. The concentration of Nod factors builds up within the curled tip, promoting further signal transduction to the plant and triggering plant responses [37]. Perception of Nod factors by plant LysM receptors, such as NFR1/NFP5 in *Lotus japonicus*, leads to signalling via a plant leucine-rich repeat (LRR) receptor, such as SYMRK in *L. japonicus*, which leads to calcium spiking within plant cells. This calcium spiking activates a calcium- and calmodulin-dependent kinase (CCaMK), which in turn activates nodule development via phosphorylation of a transcription factor called CYCLOPS that is

required for the expression of NIN, a transcriptional regulator required for the development of infection threads and nodules [38], [39], [40]. CCaMK allows integration of aboveground and root signalling; for example, in soybean it has been shown to phosphorylate the DNA-binding protein GmSTF3 which is produced in the shoot in response to light, allowing it to form a complex with FLOWERING LOCUS T (GmFT2a) that directly activates expression of NIN [41]. NIN is required at all stages of nodule formation, from initiation of nodule organogenesis to regulation of nodule number and nitrogen fixation. During early nodule development, NIN controls expression of the cytokinin receptor CRE1 and nodule pectate lyase NPL. Later, NIN is cleaved by DNF1 expressed in the infection zone of the nodule, producing a C-terminal fragment containing DNA-binding motifs that activate genes associated with symbiosome development and nitrogen fixation, including leghaemoglobin-encoding genes [42]. However, CCaMK alone is not sufficient for the formation of infection threads, meaning that there must be another form of signalling following Nod factor perception by LysM receptors that is also required [43]. Root cortex cells also dedifferentiate and then divide to form the primordium. Infection threads grow into the cortex where the developing nodule is forming. Bacteria are then taken up into an infection droplet by the developing nodule cells where they differentiate into nitrogen-fixing bacteroids [27]. Further growth and differentiation of plant cells forms a nodule, which can be either determinate or indeterminate. For example, *Medicago* and pea species form indeterminate nodules that have a permanent meristem and in which bacterial cells differentiate into terminally differentiated bacteroids that are unable to replicate or regrow outside the nodule. Meanwhile, soybean and *Lotus* species form determinate nodules that have only a short-lived meristem, in which the bacteroids are not terminally differentiated so that they can still divide [44]. The pathways involved in initiating nodule development overlap with those for lateral root development. Both organs begin to develop as a result of local auxin accumulation, and the gene expression changes during nodule and lateral root development have a 75% overlap. In particular, LBD16 expression is required for development of both and lbd16 mutants show significantly impaired formation of both lateral roots and nodules [45]. However, lateral roots form from defined founder

cells in response to environmental factors including nutrient availability, while nodule formation requires cytokinin recognition by CRE1 leading to CRE1- and NIN-dependent auxin accumulation. This suggests that the initial stages of nodule evolution involved the recruitment of existing signalling and development pathways involved in lateral root development, with modifications to create a unique nodulation pathway [45]. In addition to expression of genes involved in nitrogenase synthesis and nitrogen fixation, bacteroids also switch off expression of glutamine synthetase so that fixed nitrogen is provided to the plant rather than being assimilated by the bacteria. In return, plants provide nitrogen in the form of branched-chain amino acids to the bacteroids so that they do not starve [46]. In order to function, nitrogenase requires a low oxygen environment and plants create this inside the nodule by synthesising leghaemoglobin, which buffers oxygen [47]. They also provide carbon in the form of malate and essential minerals SO_4^{2-}, Fe^{2+}, Mg^{2+} and MoO_4^{2-} [17], [48]. Plants are able to sanction non-fixing nodules, possibly by restricted oxygen availability, resulting in reduced bacterial growth in 'cheating' nodules that are not fixing any nitrogen [49]. Plants not only distinguish between fixers and non-fixers but also to monitor nitrogen provision from each nodule relative to the rest, allowing conditional sanctioning. Thus nodules containing rhizobia that fix N_2 at an intermediate rate are only sanctioned when a superior strain is present [50]. This creates another layer of host-symbiont specificity as legume plants are adapted to prevent 'cheating' strains with reduced effectiveness from forming nodules.

A variety of legumes are farmed, but none of these contribute as significantly to the world's calorie intake as crops from the Poaceae family. Three of these, rice, wheat and maize, account for two thirds of global energy intake [51] but are unable to directly benefit from BNF. Legume plants have been useful in crop rotations over time, being grown for the nitrogen their symbionts fix to provide a nitrogen-rich environment for following crops. This use has been documented as far back and Ancient Rome [52]. There is also ongoing work to modify existing crop plants and bacteria to allow BNF to directly benefit staple cereal crops, rather than indirectly through use in crop rotations. This would allow an increase in yield of crops like rice, which is grown in 100 countries and provides more than

20% of calories consumed worldwide [53], independent of application of synthetic ammonia. Several approaches are being explored to move BNF into cereals and other staple crops. Early attempts to transfer BNF to cereals focussed on removing host-symbiont specificity by enzymatically degrading plant cell walls at root apices. Rhizobia created nodule-like structures on oilseed rape and rice plants treated this way, but only dead plant cells were invaded by the bacteria and very little nitrogenase activity was seen [54]. The presence of rhizobia alone is therefore not sufficient to form symbiotic nodules. A more recently explored alternative to removing all specificity is the creation of a new communication pathway between plants and growth-promoting bacteria. *Medicago truncatula*, a non-host legume, and barley, a model cereal, have been shown to elicit recognition from *R. leguminosarum* bv *viciae* 3841 when engineered to produce the rhizopine scyllo-inosamine [55]. A more complex alternative that is currently being explored is modifying plants to fix their own nitrogen by expressing nitrogenase themselves, cutting out the need for a symbiotic partner. Several independent projects in this area are currently being funded by bodies including the Bill and Melinda Gates Foundation (BMGF), the UK's Biotechnology and Biological Sciences Research Council (BBSRC), and the Indian Council of Agricultural Research (ICAR) [56]. Because nitrogen fixation is such an energetically costly process, plant mitochondrial and chloroplast matrices are good targets for engineered *nif* gene expression [56]. As such, *nif* genes involved in nitrogenase synthesis have been expressed and targeted to the mitochondrial matrix in both the model eukaryote *Saccharomyces cerevisiae* and the model plant *Nicotiana benthamiana*. The gene encoding nitrogenase cofactor biosynthesis protein (NifB) from *Azotobacter* has been expressed and purified from *S. cerevisiae* and shown to be active [57], and sixteen Nif proteins from *Klebsiella pneumoniae* have been successfully expressed in *N. benthamiana* [58]. It is very difficult to construct a multi-gene expression system that will be sufficiently stable in a plant to result in expression of functional nitrogenase, and if this is achieved there are further difficulties with expressing nitrogenase *in planta*. Perhaps the biggest challenge that must be overcome when transferring nitrogenase activity to a plant is the oxygen sensitivity of the iron-sulfur clusters and iron-molybdenum cofactors. The oxygen produced by the chloroplasts during

photosynthesis is a key threat to assembly of functional nitrogenase. This could theoretically be overcome by either temporal or spatial separation of *nif* gene expression and photosynthetic activity [56]. A further possibility for the transfer of nitrogen fixation to staple crop plants is engineering of cereal nodules. Most land plants associate with endomycorrhizal fungi that form arbuscular mycorrhizae and acquire phosphate that can be assimilated by plants [56]. The genetic pathways involved in nodule and arbuscular mycorrhizae formation are very similar, in what is known as the 'common symbiosis' or SYM pathway [59]. The Nod factors involved in signalling between rhizobia and their legume hosts are structurally similar to Myc factors produced by mycorrhizal fungi, and calcium oscillations are central to both pathways. Crucially, at least ten genetic components function in both nodulation and mycorrhizal signalling [60]. The rhizobium-legume symbiosis is believed to be a recent evolutionary innovation that recruited arbuscular mycorrhizal developmental mechanisms [60], and the SYM pathway is known to be present in cereals [61]. It is therefore possible that the SYM pathway could be manipulated in cereals to form nodules that could provide a stable environment for colonisation by symbiotic diazotrophs. It is possible to genetically remove the specificity of legumes for their rhizobial symbionts by artificially inducing autoactivation of nodulation signalling pathways. In normal nodule formation, perception of Nod factors leads to calcium oscillations and transduction of this calcium signal requires presence of a Ca^{2+}/calmodulin-dependent protein kinase (CCaMK). The autoinhibitory domain of this protein negatively regulates its kinase activity, and removal of this domain results in continuous expression of nodulation-related genetic pathways in the plant, without the need for a bacterial signal [62]. It is also known that transferring LysM receptors between legumes allows host-symbiont specificity to be altered, so that plants form associations with bacterial strains that are not their usual partners [60]. This suggests that non-specific nodule formation, or nodule formation with altered specificity, could be induced, allowing nodulating bacteria to colonise plants that are not their hosts, but this would require a great deal of further development. The final main approach to transferring BNF to cereals is modifying plant growth-promoting rhizobacteria (PGPRs), including rhizobia and other species, to fix and release nitrogen to cereal plants. Cereals associate with

a diverse range of bacteria and fungi, which promote plant growth by producing plant hormones, inhibiting pathogens and enhancing mineral availability. Some PGPRs are able to fix atmospheric dinitrogen, but nitrogen fixation levels in these cereal-associated diazotrophs are never as high as those achieved in the legume-*Rhizobium* symbiosis, or in synthetic fertilisers, and are never enough to support the plant's needs [56]. A factor that contributes to this problem is that many species of rhizosphere bacteria that fix nitrogen are not able to release it to benefit a plant host [56]. Engineering bacteria to maximise their nitrogen contribution to a plant host could allow non-legume crops to benefit from BNF. This can be done by increasing nitrogenase activity; for example, *Azotobacter chroococcum* that constitutively expresses the positive *nif* gene regulator *nifA* and does not express the negative regulator *nifL* shows enhanced nitrogenase activity and improved wheat yield [63]. However, constitutive expression of *nif* genes, although it improves plant nitrogen uptake, is extremely energetically costly. It is therefore beneficial to establish inducible nitrogenase activity that can be controlled by an agriculturally relevant signal. Inducible nitrogenase expression has been engineered into *Azorhizobium*, *Rhizobium* and *Pseudomonas* strains, placing *nif* genes under control of eleven sensors responding to fertiliser components, such as DAPG; or root exudate components, such as salicylic acid, naringenin or arabinose [64]. This system can also be controlled by signals such as rhizopine produced by transgenic plants [55]. Nitrogen contribution to a host plant can also be improved by maximising excretion of ammonium that can be taken up by roots. For example, *Azospirillum brasilense* mutants that lack an ammonium uptake system excrete ammonium that can be used by associated plants [65]. This approach avoids some of the complexities of expressing nitrogenase directly in plant cells but does come with challenges of its own. The main such challenge is establishing plant-microbe communication so that fixed nitrogen is released rather than assimilated by the diazotroph. In order to fix and release nitrogen specifically for the use of cereal crops, PGPRs must first be modified to recognise and respond to cereal rhizospheres. It is therefore important to understand how plant growth-promoting diazotrophs sense and interact with their plant hosts and beneficiaries and behave differently in the presence of different plant species. Overall, while there

have been promising developments in several different approaches to conferring the benefits of BNF on staple cereal crops, each has its challenges and none has yet resulted in the development of a commercially available solution. It is therefore clear that it is essential to further understanding of existing symbioses between plants and diazotrophs to be able to leverage these more successfully for the provision of environmentally and economically sustainable nutrition to a growing population. One area that it is particularly important to understand is root colonisation by diazotrophic bacteria, as this process is so specific in existing legume-*Rhizobium* symbioses and is an essential precursor to any sharing of fixed nitrogen, but is less well-characterised than the other stages of symbiosis due to the focus of much research on nodule development and nitrogen fixation [48].

1.2 Colonisation of plant roots by *Rhizobium* strains

Before nodulation can happen, *Rhizobium* strains need to attach to and colonise their plant hosts' root surfaces. The area directly surrounding the plant root and influenced by it is known as the rhizosphere [66]. Plants provide nutrient-rich root exudates, which can contain up to 30% of their photosynthetically fixed carbon [67] and this carbon-rich nutrient source makes the rhizosphere and the root surface, or rhizoplane, an extremely popular place for microorganisms to live. A single gram of rhizosphere soil can contain up to a million fungal cells and a billion bacterial cells [68]. This creates an extremely complex environment in which *Rhizobium* strains must be competitive in order to survive and attach to and grow at the root surface. In general, root microbiomes are dominated by three bacterial phyla: Proteobacteria, Actinobacteria and Bacteroidetes [69]. It is unclear whether rhizobial strains are, as a rule, better colonisers of their host legumes than other microbial species. A study of the microbiome of *Trifolium pratense* roots showed that *Rhizobium* strains accounted for 70% of colonising microorganisms, with the other 30% being made up of *Pantoea*, *Sphingomonas*, *Novosphingobium*, and *Pelomonas* strains [70]. Moreover, a study of the *M. truncatula* microbiome has shown that 60% of colonisers are *Sinorhizobium* (now *Ensifer*) strains [71]. However, other studies have reported much lower values, with 38% of the *Pisum sativum* microbiome

consisting of *Rhizobium* strains, and only around 10% of the *L. japonicus* microbiome consisting of *Mesorhizobium* strains and 10% of the *Lens culinaris* microbiome consisting of *Rhizobium* strains [72], [73]. These differences in reported microbiome composition could stem from differences in sampling and other experimental methodologies, or could reflect the influence of factors such as plant nutrition or soil type. SYM pathway mutants that are unable to form symbioses with rhizobia show altered rhizosphere microbiomes compared with wild type plants, suggesting that legume-rhizobia interactions do shape the interactions of legumes with other microorganisms [72], [74]. There is therefore some suggestion that *Rhizobium* strains are effective and important colonisers of their host legume rhizospheres. Bacterial colonisation of plant roots takes place in four stages. The first is chemotaxis towards the root, during which bacteria follow signals originating from the plant to locate somewhere beneficial to attach. The second is primary attachment, characterised by the formation of weak, reversible bonds between the bacterial cell and the root surface. During secondary attachment, these bonds are replaced by stronger, irreversible bonds. Finally, bacteria divide and grow at the root surface and form biofilms in the true 'colonisation' step. It is extremely important to understand how these interactions take place in this model symbiosis, to understand this specificity and how it differs from less specific plant-diazotroph interactions, in order to be able to leverage this interaction in agriculture and build on it to engineer more efficient interactions with a wider range of plant hosts.

1.3.1 How plants affect the rhizosphere

Most soils are carbon starved [75], but plants can secrete up to 30% of their fixed carbon into the soil directly surrounding their roots [67]. This means that plant rhizospheres can support a much more numerous microbial population than bulk soil, as well as supporting higher microbial activity, in what is known as the 'rhizosphere effect' [66], [75], [76]. It could reasonably be expected that in addition to higher population sizes and higher activity plant rhizospheres are able to support more diverse communities of microorganisms, but this is not the case, with rhizosphere communities being much less diverse than those in bulk soil [75]. This suggests that some microorganisms are better adapted to

make use of the resources available in a specific rhizosphere, allowing plants to select from the diversity available in the soil to build their microbiomes based on the compounds they provide. Therefore, the first stage in bacterial colonisation of plant rhizospheres and root surfaces comes about via selection by the plant itself. Plant species are thought to interact with soil type to shape the composition of the rhizosphere community [77]; however, it is known that some plant species assemble similar communities on different soil types, showing that in at least some cases plant genotype has a greater influence than other environmental factors. For example, alfalfa (*Medicago sativa*) and rye (*Secale cereale*) grown on different soil types had a similar rhizosphere composition regardless of where they were grown, with the major difference in community composition being between plant species, rather than between soil types [78]. Microbial communities in the rhizospheres of different plants growing in the same soil are known to be different [79], [80]. Wheat, pea and oat grown for four weeks on the same soil assembled different microbiomes, with pea and oat assembling communities with a relative abundance of eukaryotes five-fold higher than that found in the wheat rhizosphere, and the pea rhizosphere being particularly enriched for fungi [81].

The rhizosphere microbiome is affected by plant root exudates, which are composed of water, ions, enzymes, mucilage and a range of primary and secondary metabolites [76]. Root exudates differ by plant species and can be affected by nutrient availability and abiotic and biotic stress [82]. The composition of root exudates can also be altered depending on plant developmental stage. For example, secretion of sugars and sugar alcohols by *Arabidopsis* is higher during early development and decreases in older plants [83], with seedlings having a distinct microbiome from plants in later developmental stages [84]. Bacteria have been shown to assimilate root exudates using stable isotope probing [85], meaning that they are able to sense and benefit from plant-secreted components in the soil. Research has shown that individually identified components of root exudates can affect microbiome assembly. For example, cereal roots secrete benzoxazinoids that can inhibit microbial growth, including 2,4-dihydroxy-7-methoxy-2*H*-1,4-benzoxazin-3(4*H*)-one (DIMBOA), the main benzoxazinoid compound in the root exudates of maize. The beneficial microorganism *Pseudomonas*

putida is tolerant to DIMBOA and chemotactically attracted to it, and the roots of DIMBOA-deficient maize plants are significantly less colonised by *P. putida* [86], showing the power of individual compounds found in root exudates to shape the microbiome and preferentially attract beneficial microorganisms to grow in the rhizosphere. Similarly, the root exudates of tomato, cucumber and pepper have been shown to be rich in organic acids including citric acid, and the ability of bacterial strains to grow *in vitro* on citric acid is strongly correlated with ability to colonise roots [87]. In addition to selecting bacterial strains with chemoattractants and specific provision of carbon sources, plants are able to disrupt bacterial communication and behaviour by affecting quorum sensing. Acyl homoserine lactone (AHL) molecules are the primary quorum sensing signals used by Gram-negative bacteria to synchronise their responses to external signals and stresses [88]. Many plants produce AHL-mimicking compounds that interfere with bacterial quorum sensing to modulate their behaviour and control bacterial growth and biofilm formation in rhizospheres. This has been shown to be the case in pea [89], *Medicago* [90], *Chlamydomonas* [91], bean [92] and rice [92], [93], [94]. Plant stress responses can also affect microbiome assembly. Responses to herbivory can include restructuring the microbiome to favour beneficial bacteria, as in pepper (*Capsicum anuum*) plants being fed on by whitefly, where the proportion of fungi and Gram-positive bacteria in the rhizosphere increases [95]. In pepper plants being fed on by aphids the proportion of the beneficial bacterium *Bacillus subtilis* in the rhizosphere increases and the proportion of the pathogen *Ralstonia solanacearum* decreases [96]. Plants can also modulate their microbiomes in response to pathogen infection. When *Arabidopsis* leaves are infected with *Pseudomonas syringae* root colonisation by *B. subtilis* increases due to increased secretion of malic acid by plant roots [97]. Plants are known to activate salicylic and jasmonic acid signalling pathways in response to stress [98] and this may affect microbiome assembly in many of these cases. Artificial activation of the jasmonic acid signalling pathway by application of methyl jasmonate to *Arabidopsis* significantly alters microbiome composition [99], showing a direct link between known plant stress responses and rhizosphere community assembly. Studies investigating the effects of single-gene mutations on microbiome composition have explored the effects of plant

genotype on their colonising bacteria and fungi more deeply and discovered that some aspects of community assembly can be explained by individual genes. An *Arabidopsis* ABC transporter mutant has altered root exudates and an altered microbiome [100], while production of a single exogenous glucosinolate by transgenic *Arabidopsis* significantly alters the microbial community that forms on the roots [101]. Rice CCaMK mutants have an altered rhizosphere microbiome, showing reduced numbers of *Rhizobiales* and *Sphingomonadales* compared with wild type, and higher numbers of *Anaerolineae*, *Clostridia* and *Actinobacteria* [102]. Overall, plant genotypes are shown to contribute hugely to the communities of microorganisms that assemble in their rhizospheres, via differences in root exudates, including exudation of AHL mimics, stress response pathways and signalling pathways like CCaMK. These effects combine to create an initial boundary to colonisation of the rhizosphere by beneficial microorganisms like *Rhizobium* strains, whereby the plant exerts an effect on bacterial colonisation that cannot be ignored. This work will focus on the mechanisms by which the bacteria themselves are able to attach to and colonise plant roots, but it should be noted that the communication involved is two-way and that plant rhizospheres and root surfaces are dynamic environments that in many ways shape their own microbiomes.

1.3.2 Chemotaxis

Bacteria are able to move through liquid or across solid surfaces, propelled by flagella or pili. Flagella are long, helical organelles with a motor at the base that use flux of ions across the cell membrane to power rotation and hence motion. They drive swimming motility in liquid or swarming or gliding motility on solid surfaces [103]. Pili drive twitching motility across surfaces by extension, tethering and retraction of pili to pull the bacteria across the surface [104]. Bacterial cells can direct this motility in response to external factors to preferentially move towards concentrations of nutrients or other chemoattractants and away from compounds such as toxins. Stimuli are detected by receptors called methyl-accepting chemotaxis proteins (MCPs) that bind to ligands and promote autophosphorylation of the chemotaxis protein CheA. In turn, this triggers a phosphorylation cascade that results in changes

in flagellar motion [105]. On average, soil-dwelling bacteria possess two or more chemotaxis systems [106], including *Rhizobium leguminosarum*, which has two [107]. The ability to respond to stimulants in the soil therefore confers an important competitive advantage on rhizosphere bacteria. Plants release a significant proportion – up to 30% – of their carbon into the soil in the form of root exudates [67], [108] and this creates a concentration gradient around their roots that bacteria move towards. Some components of root exudates introduce a degree of specificity into bacterial chemotactic responses, meaning that bacteria preferentially move towards particular species of plants, and that plants themselves preferentially attract certain strains and species of bacteria, as discussed above. In particular, *Rhizobium* strains show different chemotaxis responses to the roots of different plant species. One study tested chemotaxis of six different *Rhizobium* strains towards the roots of ten different legume plants and found that each of the six bacterial strains responded to a different subset of the legume plants, with none responding to all ten [108]. Further investigation has shown that *Rhizobium* strains respond to specific components of plant root exudates. *R. leguminosarum* responds to a range of sugars, carboxylic acids and amino acids found in pea root exudates [109], [110], specifically to homoserine, the most abundant nitrogenous compound in legume root exudates [109]. *R. leguminosarum* can also respond chemotactically to flavonoids, including the *nod* gene inducer naringenin [109]. Similarly, *Sinorhizobium meliloti* responds chemotactically to the flavonoid luteolin, which is present in alfalfa root exudates and induces *S. meliloti nod* genes [111]. Chemotaxis towards plant roots is an important precursor to attachment, colonisation and nodule formation and there is clearly some degree of specificity in which compounds attract which bacterial strains. However, the true determinants of host-symbiont specificity are found downstream of chemotaxis, during root attachment and the signalling processes that happen afterwards.

1.3.3 Attachment

Attachment of *Rhizobium* bacteria to legume root hairs is an important precursor to infection thread formation, and hence nodule development, for three reasons [24]. Firstly, the proximity of the bacteria

to root hairs ensures a consistent nutrient supply, allowing bacteria to form biofilms on the root surface. Secondly, bacteria that are attached at the root hair surface will be those that are trapped when root hair curling is induced, meaning that they will be the ones that go on to replicate within the plant and live in the nodule, which normally contains bacteria of only one clone. Finally, root hair deformation can occur in response to Nod factor signalling, but actual curling of the root hair, which is essential for infection of most legumes by most rhizobia, can only happen when induced by physically attached bacteria. In addition to attaching specifically to the root hairs of their legume hosts, rhizobia are able to attach to the root surfaces of both host and non-host plants as an initial stage in rhizoplane colonisation. This is an important step for rhizobia to be able to benefit from the nutrients available in plant root exudates. Root surface attachment by rhizobia takes places in two stages: primary, reversible attachment, and secondary, irreversible attachment. This biphasic model is common to a range of agriculturally significant bacteria, including *Rhizobium*, *Agrobacterium*, *Azospirillum* and *Pseudomonas* [112]. Initial primary attachment involves binding of bacterial cells to root surfaces via fundamental forces that are common to all plant-microbe interactions. Van der Waals and electrostatic forces interact to cause adhesion between bacterial cells and surfaces in a process that is explained by the DVLO theory of colloid adhesion [113]. Van der Waals forces are a form of ubiquitous attraction between molecules that can also be applied to cell surfaces. The interaction arises from an instantaneously formed dipole in one molecule which interacts with adjacent molecules to induce another weak dipole and hence an attractive force, and this process can act between adjacent cells to hold them together [114]. Meanwhile, the electrostatic forces between cells are initially repulsive, as both bacterial and plant cells are usually negatively charged. As the distance between the two cells decreases, and with growing ionic strength, adhesion becomes more favourable, so that the cells are drawn together rather than repelled by their electrostatic charges [115]. Hydrophobicity also contributes to cell-cell adhesion, including that between bacterial and plant cells [115]. Hydrophobic cells are strongly attracted to hydrophobic surfaces, helping to introduce primary adhesion between cells [116]. In addition to these fundamental forces, constitutively expressed plant and bacterial

surface components also play a more specific role in primary attachment. For example, the role of flagella, pili and fimbriae in root attachment has been well-documented, with flagella and pili playing a role in primary attachment as well as in chemotaxis prior to attachment. Various *A. brasilense* mutants that cannot produce flagella or that produce altered flagella are unable to attach to wheat or maize roots [117], [118]. In addition, damaging existing flagella on wild type *A. brasilense* inhibits attachment, and purified polar flagella can bind to wheat roots *in vitro* [117], showing their importance in primary attachment itself, outside of chemotaxis. Meanwhile the *pilAB* genes whose products form Type IV pili in *Azoarcus* sp BH72 are known to be involved in adhesion to rice roots [119]. Finally, fimbriae are also known as 'attachment pili' and mediate attachment to both biotic and abiotic surfaces [120]. They are common attachment factors in proteobacteria [121], including in rhizobia. For example, antibody-mediated inhibition of *Bradyrhizobium japonicum* pili reduces attachment to soybean roots by 90% and subsequent nodulation by 80% compared with uninhibited bacteria [122].

It was originally thought that host-symbiont specificity was determined using primary attachment, by interaction of bacterial surface polysaccharides with plant lectins. This theory is supported by the finding that soybean lectin binds to 22 out of 25 tested soybean-nodulating *B. japonicum* strains and to none of 23 other tested *Rhizobium* strains that do not nodulate soybean [123]. Cell-surface polysaccharides on *R. leguminosarum* bv *trifolii* have been shown to interact directly with their host clover root lectins [124]. Finally, attachment of *R. leguminosarum* bv *trifolii* to clover roots can be inhibited by addition of 2-deoxyglucose, but this does not inhibit attachment of *S. meliloti* to alfalfa roots, suggesting that the *Rhizobium*-clover symbiosis uniquely involves binding of a bacterial surface polysaccharide to a 2-deoxyglucose-sensitive receptor, while attachment of *S. meliloti* to alfalfa requires a different set of interacting factors [125]. It has since been demonstrated that Nod factors are the primary determinant of host-symbiont specificity in the *Rhizobium*-legume symbiosis, not polysaccharide-lectin interactions. However, primary root attachment is an important step in determining specificity, even if not the predominant one. In addition to the work described above, it has been demonstrated that *R. leguminosarum* bv *viciae* can nodulate transgenic clover roots

expressing pea lectin but not wild type clover [126], showing that the introduction of host lectins into a non-host plant can modify specificity. It may be that the accumulation of rhizobia at the root surface driven by polysaccharide-lectin interactions results in increased delivery of Nod factors to plant receptors which can overcome receptor specificity and induce nodulation signalling [24]. The rhizobial surface polysaccharide glucomannan is particularly important in introducing specificity into primary attachment between rhizobia and legumes. Interaction between *R. leguminosarum* bv *viciae* glucomannan and pea and vetch lectins is known to promote polar attachment of rhizobial cells to the root surface independently of the *nod* gene inducer naringenin [127]. A *gmsA* mutant that is unable to produce glucomannan is unable to attach to or form biofilms on root hairs, despite being able to do so *in vitro*, and is outcompeted by wild type in nodulation, showing the importance of glucomannan for competitive symbiosis formation [128]. This shows that primary attachment is a key precursor to development of a symbiosis with beneficial bacteria, and that interactions taking place during primary attachment to contribute to establishing host-symbiont specificity, potentially helping to make rhizobia more competitive and conferring the ability on plants to avoid interactions with 'cheating' or pathogenic bacteria. In addition to bacterial surface polysaccharides, secreted proteins are important for the process of primary attachment. Some of these secreted proteins are bacterial lectins that bind to plant cell surface carbohydrates; for example, in *B. japonicum* a polarly expressed lactose-sensitive lectin mediates specific binding of the bacteria to soybean roots [129], [130]. In another *Bradyrhizobium* strain an L-fucose-sensitive lectin mediates attachment to white lupin [131]. A calcium-binding protein called rhicadhesin also plays a key role in rhizobial primary attachment to legume roots. Rhicadhesin was first recorded as a 14 kDa protein extracted from rhizobial culture that is able to inhibit rhizobial attachment to roots when used as a pre-treatment [132], suggesting that it binds to plant receptors, blocking binding by the bacteria themselves when plants are exposed to bacterial culture. However, attempts to identify the gene encoding rhicadhesin and further characterise it by proteomics, root attachment assays, INSeq screening and confocal imaging have revealed that there are multiple candidates that match what is known about rhicadhesin, suggesting

that early studies may have been identifying multiple proteins, or that the inhibitory effect of purified rhicadhesin on attachment is not indicative of a direct role in attachment itself [133]. Other attempts to identify and characterise rhicadhesin have revealed a family of *Rhizobium*-adhering proteins (Rap) that are involved in root attachment [134], which also suggests that the attachment-related proteome is more complex than one neat attachment factor. Primary attachment is susceptible to external factors, particularly pH, which affects which mechanism plays a greater role. Under alkaline pHs plant lectins are solubilised, meaning that interactions between secreted bacterial proteins such as rhicadhesin and plant receptors become more significant in mediating the interaction [127]. Meanwhile, rhicadhesin is bound to the bacterial cell wall by a calcium ion [135] that dissociates under acidic conditions, meaning that lectin-polysaccharide interactions become more important under these conditions.

Following primary attachment, secondary attachment involves the formation of more specific, irreversible bonds between bacterial and plant cells, and the development of a bacterial microcolony on the root surface [121]. In particular, rhizobia accumulate around root hairs to form a 'root cap' [24]. Pili and fimbriae continue to be involved in secondary attachment, mediating intercellular interactions that result in aggregation on the root surface [120]. Synthesis and secretion of other secondary attachment factors often happens in response to successful primary attachment; most factors important in secondary attachment are not constitutively expressed [112]. Cellulose fibrils are universal secondary attachment factors among proteobacteria [136], and production of cellulose fibrils in induced in *R. leguminosarum* bv *trifolii* upon close contact with plant roots [137] and within two hours of attachment of *Agrobacterium tumefaciens* to plant tissue [138]. Cellulose fibrils are required for formation of the 'root cap' on root hairs by attaching rhizobia [128], [139]. Exopolysaccharide (EPS) is also required for the specific secondary attachment of rhizobia to root hairs, with *R. leguminosarum* mutants in the EPS production gene *pssA* showing reduced attachment to root hairs although they are still able to attach to root epidermal cells [128]. *R. leguminosarum* bv *trifolii* mutants in *rosR* produce three times less EPS than wild type cells and show significantly impaired attachment to clover roots

[140]. EPS composition can vary between species and strains, which affects cell surface structure and binding properties and hence adhesion to plant cells [24]. Finally, proteins that are secreted in response to successful primary attachment also play a role in secondary attachment. Proteins secreted by the PrsDE Type I secretion system are particularly important [141], particularly Rap proteins. For example, RapA1 is a 30 kDa calcium-binding protein secreted by PrsDE that localises to one end of the extracellular surface and is predicted to bind to EPS or capsular polysaccharide and promote adhesion and agglutination of *Rhizobium* on root hair cells [141]. Various other Rap proteins are also secreted by PrsDE and all contain a conserved motif predicted to bind to EPS [141]. Some of these proteins have also been shown to play a role in attachment, with a *rapA2rapC* double mutant in *R. leguminosarum* bv *viciae* showing severely impaired attachment to pea roots [142]. However, *rapA1*, *rapA2*, *rapB* and *rapC* all appear as neutral in an INSeq screen for genes required in root attachment, suggesting that there is some redundancy among Rap proteins in their role in adhesion [133]. Overall, initial attachment of rhizobia to host plant roots is clearly a complex process that involves a variety of different interactions and components, as well as a source of host-symbiont specificity and a key determinant of competitive success for symbionts.

1.3.4 Growth and biofilm formation

Following attachment, the microcolonies formed during secondary attachment mature into biofilms [121]. Most bacteria in both natural and artificial environments live in biofilms, rather than in planktonic culture [143], with over 99% of bacteria in natural environments estimated to exist in biofilms [144]. Biofilms are three-dimensional assemblies of microbial cells that are irreversibly associated with a surface and surrounded by a matrix made up primarily of polysaccharide [145]. In addition to polysaccharides, the matrix consists of proteins, lipids and free DNA separated by water channels that allow exchange of nutrients, enzymes and signals and disposal of toxins [146], [147]. Biofilms protect bacteria from a variety of stresses, including desiccation, pH changes, osmotic stress, ultraviolet radiation, antibiotics, heavy metals, host immune responses and protozoan grazers [143],

[148]. In addition, growth in a biofilm allows bacteria to interact more closely than in a planktonic lifestyle; for example, allowing communication, horizontal gene transfer and cooperative metabolic activity [148].

The main component of the biofilm matrix is polysaccharides, particularly EPS, which is responsible for adhesion of the biofilm to abiotic and biotic surfaces as well as cohesion between cells within the biofilm. Bacteria that are unable to produce EPS are deficient in biofilm formation. A *S. meliloti exoY* mutant deficient in EPS production showed a 60% reduction in formation of biofilms on abiotic surfaces compared with wild type, despite showing no difference in growth rate. Biofilms produced by *exoY* mutants were also much less stable than wild type biofilms, being easily washed off surfaces, and did not mature correctly, forming only small clusters of cells [149]. A *S. meliloti exoS96* mutant, meanwhile, produces eight times as much EPS as wild type and produces larger biofilms which, however, are less structurally sound and robust than wild type [149]. Different species of bacteria produce different types of EPS, which can be homopolysaccharides, made up of many repeats of the same sugar, or, more commonly, heteropolysaccharides, made up of a variety of different sugars [148]. This means that EPS produced by different bacterial species can have different surface charges and molecular weights. Even different strains of the same species can produce different EPS. For example, 25 strains of *Streptococcus thermophilus* produced EPS of different monomer composition of molecular mass. These could be broadly characterised into three groups, but members of the same group still showed differences in their EPS [150]. *S. meliloti* produces two exopolysaccharides, EPS I (succinoglycan) and EPS II (galactoglucan) [151]. *S. meliloti* Rm1021 produces large amounts of EPS I but no EPS II, while *S. meliloti* Rm8530 does produce EPS II alongside smaller amounts of EPS I [152]. These differences in EPS production can affect the formation and structure of bacterial biofilms, which can in turn affect rhizobial interactions with host plants. *S. meliloti* Rm8530 creates larger biofilms on abiotic surfaces than Rm1021, and forms biofilms all over alfalfa roots, including root hairs, while Rm1021 forms clusters of cells only on the main root [152]. In *R. leguminosarum*, the main type of EPS produced is acidic EPS synthesised by *pss* gene products [153]. *pssA* encodes an isoprenylphosphate

glycosyl transferase that is directly responsible for synthesis of acidic EPS and is essential for biofilm formation, which is completely abolished in *pssA* mutants [128], [141]. A mutant in several other *pss* genes, but not *pssA*, was also unable to produce biofilms [141], showing how central acidic EPS production is to biofilm formation in *R. leguminosarum*. Mutants in the *R. leguminosarum* type I secretion system encoded by *prsD* and *prsE* exhibit disrupted biofilm formation [141], [154], suggesting that the components PrsDE secretes are essential in developing biofilms. Indeed, PrsDE is responsible for the secretion of the EPS glycanases PlyA and PlyB which cleave acidic EPS, and a mutant in *plyB* has altered biofilm formation, with both the timing of biofilm development and its structure being different from wild type [141].

In addition to EPS, lipopolysaccharide (LPS) is also required for colonisation of plant roots by bacterial cells. A *Rhizobium* sp IRBG74 mutant which has deficient production of O-antigen, the portion of LPS that varies between bacterial species, was unable to successfully colonise its legume host *Sesbania cannabina* or rice [155]. *R. leguminosarum* O-antigen mutants showed premature microcolony formation with fewer lateral contacts between bacteria *in vitro*, as well as altered biofilm formation on plant roots[156]. An *lpsB* mutant in *S. meliloti* has a defective core portion of LPS and showed reduced biofilm formation and an altered biofilm structure on abiotic surfaces compared with wild type. In addition, the *lpsB* mutant showed increased attachment to alfalfa roots, perhaps due to increased 'stickiness', but this did not translate into improved nodulation compared with wild type, showing the ability of LPS to alter symbiotic interactions through modified biofilm development [157]. Proteins are also an important component of biofilm matrices. These include enzymes that break down biopolymers and modify EPS, and structural proteins such as lectins [148]. Flagella, pili and fimbriae may also play a structural role in biofilms by interacting with EPS and other matrix components [145]. *S. meliloti* mutants unable to synthesise flagella showed reduced biofilm development; however, they also showed impairment in initial attachment to surfaces [149]. It is therefore unclear exactly what the role of flagella and other chemotaxis proteins is in biofilm formation. They could be playing a direct role in developing and maintaining matrix structure, but the impaired biofilm phenotypes seen in

mutants could also be due to deficient chemotaxis and attachment affecting the overall process of colonisation, rather than biofilm formation specifically.

It has been demonstrated that Nod factors affect the formation of rhizobial biofilms. Individual *S. meliloti* mutants in *nodD*, *nodA*, *nodB* and *nodC*, whose products are the components of the core Nod factor, all showed impaired biofilm formation, as did a *nodDABC* mutant which formed 50-70% smaller biofilms than wild type on both abiotic surfaces and plant roots [158]. *nodD* is also required for biofilm formation and colonisation of soybean roots by *Sinorhizobium fredii* [159]. *nodDABC* are found in almost all bacteria that nodulate legumes, suggesting that development of three-dimensional biofilm structures is an essential part of nodulation. Quorum sensing also plays an important role in biofilm assembly [145]. Overexpression of the *A. tumefaciens* quorum sensing regulator TraR in *Mesorhizobium huakii* interferes with *M. huakii* signalling by binding to AHLs; this overexpressor shows reduced biofilm formation compared with wild type, showing the importance of quorum sensing in biofilm development [160]. Meanwhile, the quorum sensing *cinIR* genes control EPS biosynthesis in *R. leguminosarum* via transcription of *cinS*, which is cotranscribed with *cinI* in response to AHL levels [161].

1.4 RNA sequencing

1.4.1 Contribution of RNA sequencing to understanding of the legume-*Rhizobium* symbiosis

For approximately fifteen years, RNA sequencing has been a leading method of transcriptome analysis that has shaped understanding of genomic function in both eukaryotes and prokaryotes [162], [163], [164]. In this process, RNA is extracted, converted to cDNA using reverse transcriptase and sequenced using a high-throughput next-generation sequencing platform [165]. Two other key techniques for obtaining transcriptional data are microarrays and quantitative reverse transcription PCR (RT-qPCR). In microarray analysis, mRNA is extracted, converted to cDNA and fluorescently labelled, before being hybridised to a set of oligomer probes on glass (the microarray). Fluorescence is quantified and ratio of fluorescence used to calculate relative abundance of transcripts [166]. In RT-qPCR, mRNA is

extracted and converted to cDNA which is then amplified using PCR with primers designed to bind to a gene of interest. The amplified cDNA is quantified using a fluorescent dye and compared with cDNA amplified from the same sample using primers binding to a housekeeping gene that is known to have consistent expression. This shows relative expression of the gene of interest compared with background cellular function [167]. Both microarrays and RNA sequencing provide whole-genome transcriptomic data, but RNA sequencing has a greater sensitivity than microarrays [162], [163]. RT-qPCR is extremely sensitive but a specific set of primers is required for every gene tested, meaning that it is better suited to precise analysis of a few genes than whole transcriptome analysis. Moreover, both RT-qPCR and microarrays are hybridisation-based, while RNA sequencing is not, meaning that it can detect transcription independent of annotation, allowing discovery of new coding regions and antisense transcripts [168]. These advantages, combined with falling costs, give RNA sequencing the potential to contribute hugely to understanding of how the legume-*Rhizobium* symbiosis develops and how it may be manipulated to benefit sustainable agriculture. RNA sequencing provides a snapshot of the transcriptome at a particular moment, and this ability to capture samples at different timepoints means that it is theoretically possible to compare how gene expression changes during the different stages of symbiosis initiation and development. In particular, attachment to and colonisation of host plant roots by *Rhizobium* strains are less well-researched than other parts of the symbiosis, as the majority of research has focussed on nodule formation and nitrogen fixation [48]. Recent research has directed greater focus to these aspects of symbiosis, in the form of insertion sequencing (INSeq) screens that identify genes that are important for attachment to [133] and colonisation of [169] plant roots in competition. While these studies have greatly advanced understanding of certain genetic processes underlying attachment and colonisation, they are not a substitute for RNA sequencing. INSeq is based on mutant libraries generated by random mutagenesis. A library contains, in theory, mutants in every gene in the genome. The library is grown in laboratory culture, then subjected to a challenge, such as being exposed to plant roots for attachment or colonisation, before being harvested and the mutants present in the harvested sample quantified [170], [171]. Strains that were unable to

attach or colonise represent mutations in genes that are essential for survival of the challenge, in this case, interaction with plant roots. However, a gene being important for attachment or colonisation does not mean that it is expressed during these processes, as it may have been expressed at an earlier stage, or its expression may not be enhanced during interaction with a plant root despite being necessary for the interaction to take place. To obtain a more complete understanding of the genetic regulation and gene expression changes that take place during these stages of symbiosis formation, it is essential to further investigate transcriptome changes. Looking at transcriptomic data alongside INSeq data can paint a more complete picture of what genes and pathways are most crucial to the specific affinity of *Rhizobium* strains for their host legume species, and the genetic basis of more general colonisation of other plant roots. An additional option for understanding how the processes of attachment and colonisation take place would be proteomic analysis, such as high-throughput mass spectrometry, which would be used to show the relative abundance of all proteins being expressed during these processes. This technique has been used to analyse protein expression in *S. fredii* in laboratory culture and bacteroids [172]. However, knowledge of which genes are being expressed in a particular stage of symbiosis has applications in engineering of biological nitrogen fixation that cannot be matched by other types of data. The ability to identify promoter regions that are active during colonisation of both host and non-host plants could be used to identify candidate promoters for fusion to genes controlling agriculturally useful traits such as nitrogen fixation, auxin biosynthesis, phosphate solubilisation and others to create 'super bugs' that could be inoculated onto fields and express these traits only once attached to the roots of target crop plants, rather than indiscriminately in the soil or during colonisation of the roots of weeds. This application of transcriptomic data could help to bypass the need for a cereal nodule, for example, by allowing nitrogen fixation to take place on the root surface, triggered specifically by exposure to cereal roots.

1.4.2 Potential pitfalls

While RNA sequencing has many potential applications and advantages that make it a valuable tool for the understanding of the legume-*Rhizobium* symbiosis, there are issues that need to be addressed to enable it to be leveraged to its full potential. The accepted sequencing read depth to obtain statistically significant information about differential gene expression in bacteria is 5-10 million reads [173]. However, when analysing mixed samples such as bacteria that are colonising the roots of plants it can be difficult to obtain this due to contamination of the bacterial samples with plant RNA. This can be depleted using beads that bind to the polyA tails of eukaryotic mRNA but leave bacterial mRNA unbound, and this method has been used successfully to purify mRNA from rRNA in eukaryotic samples, where sequencing of eukaryotic mRNA is the target [174]. However, unpublished data from the Poole lab shows that even following polyA depletion a significant portion of reads obtained from *R. leguminosarum* biovar *viciae* 3841 isolated from pea roots still map to the pea genome. One solution is simply to sequence to a greater total read depth, so that the total number of reads mapping to the bacterial genome is sufficient to yield a useful dataset. However, this could quickly become extremely expensive. Although the costs of DNA and RNA sequencing have dropped considerably over the last decade, consequent increases in stringency, including the expectation for higher numbers of replicates for publication, have kept overall costs high [175]. It is therefore necessary to develop an additional method of separating contaminating plant material from bacterial samples before sequencing to allow the kind of in-depth, broad-spectrum sequencing project that could be used to significantly further understanding of gene expression during rhizobial attachment to and colonisation of legume roots.

1.5 Research objectives

Attachment to and colonisation of host plant roots are essential precursors to formation of a successful nitrogen-fixing symbiosis between *Rhizobium* strains and their host legumes but are less well-characterized than other stages of symbiosis formation. It is therefore important to further understanding of these processes to allow the legume-*Rhizobium* symbiosis and rhizobial nitrogen

fixation to be utilised as efficiently as possible in the development of more environmentally and economically sustainable agricultural processes. In this study, I therefore aim to further understanding of the genetic basis of attachment to and colonisation of *Pisum sativum* roots by the model *Rhizobium* strain *R. leguminosarum* bv *viciae* 3841 (Rlv3841) [176], on both a whole genome scale and a focussed, individual pathway scale. In Chapter 3 I discuss the development of a new method for separating attached and colonising bacteria from plant roots in order to reduce contamination of RNA sequencing data by plant root material. Use of this method also allows investigation of gene expression that is unique to bacteria that are interacting with the root surface, unlike existing studies that have combined bacterial samples in bulk soil with those attached to the root surface. This will be shown by initial RNA sequencing of Rlv3841 colonising the root surfaces of its host plant (*P. sativum*) and a non-host model cereal, barley (*Hordeum vulgare*). This will allow comparison of the genetic basis of colonisation of a host and a non-host plant by a model nitrogen-fixing *Rhizobium* strain, and further understanding of what pathways are involved in host colonisation as opposed to more general colonisation of plant roots. This data will be important for both understanding of existing BNF systems and application of BNF to cereal growth, as it is impossible to generate a beneficial interaction between a cereal and *Rhizobium* before understanding the genetic basis of its interaction with legumes and what aspects are lacking in its existing interactions with cereals. This dataset will also facilitate engineering of Rlv3841 to carry out agriculturally useful processes when colonising the roots of specific plants, as it will be possible to identify promoters that may be fused to genes such as *nifA*, allowing fixation of nitrogen on the cereal root surface. This method will also make it possible to examine each time point in the colonisation process individually, and a more thorough investigation into the differing gene expression in Rlv3841 during attachment to and colonisation of pea roots will be presented in Chapter 4. This will highlight the very different sets of genes that are required for each stage in colonisation, showing the importance of unpicking this stage of the interaction between *Rhizobium* and its host plant, even though it is only a precursor to nodule formation and nitrogen fixation. It represents an important dataset for furthering understanding of host recognition, attachment and growth at the root

surface in Rlv3841. Chapter 5 presents an example of the kind of insight that can be drawn from the dataset examined in Chapter 4: a closer investigation into a single operon consisting of three genes (RL3234-RL3236) that are somewhat upregulated during attachment to pea roots and more strongly upregulated during colonisation of both pea and barley. The operon consists of RL3234, a lipoprotein, RL3235, an extracytoplasmic sigma factor, and RL3236, an anti-sigma factor. The results presented in this chapter include an in-depth examination of how the operon is regulated, how it perceives the plant root, and its role in attachment and colonisation. Overall, in this work I examine on both a whole genome scale and in finer detail the genetic basis for attachment and colonisation of pea roots by Rlv3841, as well as the overlap and key differences between colonisation of its host and a non-host model cereal, barley. I present a novel method for obtaining transcriptomic data from root-colonising bacteria that, in addition to its applications to Rlv3841, could be applied to understanding other agriculturally significant plant-microbe interactions, both symbiotic and pathogenic.

1.5.1 Hypotheses

- Chapter 3: A differential spin method will allow me to gather transcriptomic data from rhizoplane-colonising bacteria by separating them from plant material
- Chapter 3: There are substantial differences in Rlv3841 gene expression during colonisation of a host legume and a non-host cereal that can contribute to our understanding of how *Rhizobium* strains are adapted to their host plants
- Chapter 4: There are substantial differences in Rlv3841 gene expression during attachment to and colonisation of the pea rhizoplane, showing that these are two distinct processes
- Chapter 5: The *ecfE* operon contributes to attachment to and colonisation of plant rhizoplanes by Rlv3841

2 Materials and methods

2.1 Bacterial strains, plasmids and primers

Bacterial strains used in this study are listed in Table 2.1. Strains were stocked in 15% glycerol and frozen and stored at -80□C. Plasmids are listed in Table 2.2 and primers in Table 2.3.

Table 2.1: Bacterial strains used in this work.

Strain	Description	Reference
Rhizobium leguminosarum		
Rlv3841	*Rhizobium leguminosarum* biovar *viciae* 3841; Str^R derivative of strain 300.	[176]
OPS2453	Rlv3841[pOPS1531] (miniTn7 mCherry). Str^R, Gm^R	Poole lab, unpublished
OPS3439	OPS2968[pIJ11268] (Mutant in RL2602)	This work
OPS3440	OPS3150[pIJ11268] (Mutant in RL0777)	This work
OPS3441	OPS2968[pIJ11282] (Mutant in RL2602)	This work
OPS3442	OPS3150[pIJ11282] (Mutant in RL0777)	This work
LMB653	*lppE* in-frame deletion. Str^R	Poole lab, unpublished
LMB669	*ecfE* in-frame deletion. Str^R	Poole lab, unpublished
LMB527	*asfE* omega Spec mutant. Str^R, $Spec^R$	Poole lab, unpublished
LMB540	*lppE* omega Spec mutant. Str^R, $Spec^R$	Poole lab, unpublished
LMB487	Rlv3841[pLMB579]. (P*lppE* in pIJ11268) Str^R, Tet^R	[177]
LMB743	Rlv3841[pIJ11282]. Str^R, Tet^R	Poole lab, unpublished
LMB542	Rlv3841[pIJ11268]. Str^R, Tet^R	Poole lab, unpublished
LMB747	LMB653[pIJ11282]. Str^R, Tet^R	Poole lab, unpublished

LMB680	LMB653[pLMB579]. Str^R, Tet^R	Poole lab, unpublished
LMB748	LMB669[pIJ11282]. Str^R, Tet^R	Poole lab, unpublished
LMB682	LMB669[pLMB579]. Str^R, Tet^R	Poole lab, unpublished
LMB744	LMB527[pIJ11282]. Str^R, Tet^R	Poole lab, unpublished
OPS0266	LMB527[pLMB579]. Str^R, Tet^R	Poole lab, unpublished
OPS0167	Rlv3841[pOPS0065]. Str^R, Gm^R	Poole lab, unpublished
OPS0149	Rlv3841[pOPS0047]. Str^R, Tet^R	Poole lab, unpublished
OPS0150	Rlv3841[pOPS0048]. Str^R, Tet^R	Poole lab, unpublished
OPS0169	Rlv3841[pOPS0080]. Str^R, Tet^R	Poole lab, unpublished
OPS0190	Rlv3841[pOPS0083]. Str^R, Tet^R	Poole lab, unpublished
OPS0191	Rlv3841[pOPS0084]. Str^R, Tet^R	Poole lab, unpublished
OPS0502	Rlv3841[pOPS0085]. Str^R, Tet^R	Poole lab, unpublished
OPS0503	LMB527[pLMB509][pLMB579]. Str^R, $Spec^R$, Tet^R, Gm^R	Poole lab, unpublished
OPS0504	LMB527[pOPS0050][pLMB579]. Str^R, $Spec^R$, Tet^R, Gm^R	Poole lab, unpublished
OPS3276	LMB540[pOPS2000][pOPS0050]. Str^R, $Spec^R$, Tet^R, Gm^R	This work
OPS3277	LMB540[pOPS2001][pOPS0050]. Str^R, $Spec^R$, Tet^R, Gm^R	This work
OPS3500	LMB540[pOPS2089][pOPS0050]. Str^R, $Spec^R$, Tet^R, Gm^R	This work
OPS3501	LMB540[pOPS2090][pOPS0050], Str^R, $Spec^R$, Tet^R, Gm^R	This work
OPS3502	Rlv3841[pOPS2088]. Str^R, Gm^R.	This work
OPS3586	LMB540[pOPS2126][pOPS0050]. Str^R, $Spec^R$, Tet^R, Gm^R	This work

OPS3587	LMB540[pOPS2127][pOPS0050]. Str^R, $Spec^R$, Tet^R, Gm^R	This work
OPS3588	LMB540[pOPS2128][pOPS0050]. Str^R, $Spec^R$, Tet^R, Gm^R	This work
OPS3589	LMB540[pOPS2129][pOPS0050]. Str^R, $Spec^R$, Tet^R, Gm^R	This work
OPS3590	LMB540[pOPS2126]. Str^R, $Spec^R$, Tet^R	This work
OPS3591	LMB540[pOPS2127]. Str^R, $Spec^R$, Tet^R	This work
OPS3592	LMB540[pOPS2128]. Str^R, $Spec^R$, Tet^R	This work
OPS3593	LMB540[pOPS2129]. Str^R, $Spec^R$, Tet^R	This work
OPS3597	Rlv3841[pOPS2126]. Str^R, Tet^R	This work
OPS3598	Rlv3841[pOPS2127]. Str^R, Tet^R	This work
OPS3599	Rlv3841[pOPS2128]. Str^R, Tet^R	This work
OPS3600	Rlv3841[pOPS2129]. Str^R, Tet^R	This work
OPS3660	LMB540[pOPS2143][pOPS0050]. Str^R, $Spec^R$, Tet^R, Gm^R	This work
OPS3661	LMB540[pOPS2144][pOPS0050]. Str^R, $Spec^R$, Tet^R, Gm^R	This work
Escherichia coli		
DH5α	Competent *E. coli* strain for use in transformations, carrying the following mutations: *F-deoR endA1 recA1 relA1 gyrA96 hsdR17(rk-mk+) supE44 thi-1-phoA Δ(lacZYA-argF) U169 Φ80lacZΔM15 λ.*	Bioline

Table 2.2: Plasmids used in this work.

Plasmid	Description	Reference

pOPS1531	J23104-RBStd-mCherryI-DT16-pOGG276 (pUC18R6K-miniTn7T-Gm). Gm^R	Poole lab, unpublished
pIJ11268	Vector with promoterless reporter luciferase genes. Tet^R	[142]
pIJ11282	pIJ11268 with luciferase genes under control of *nptII* promoter. Tet^R	[142]
pLMB579	375 bp promoter of *lppE* (RL3234) cloned into pIJ11268. Tet^R	[177]
pOPS0065	375 bp *lppE* promoter (oxp0129-oxp0130) cloned as BamHI / KpnI in pLMB426. Gm^R	Poole lab, unpublished
pOPS0047	708 bp *lppE* promoter and coding sequence (oxp0028-oxp0029) cloned as KpnI and BamHI in pIJ11268. Tet^R	Poole lab, unpublished
pOPS0048	108 bp *lppE-ecfE* intergenic region (oxp0143-oxp0144) cloned as KpnI and BamHI in pIJ11268. Tet^R	Poole lab, unpublished
pOPS0080	284 bp *lppE* coding sequence and *lppE-ecfE* intergenic region (oxp0145-oxp0144) cloned as KpnI and BamHI in pIJ11268. Tet^R	Poole lab, unpublished
pOPS0083	285 bp *lppE* promoter region (oxp0147-oxp0149) cloned as KpnI and BamHI in pIJ11268. Tet^R	Poole lab, unpublished
pOPS0084	223 bp *lppE* promoter region (oxp0148-oxp0149) cloned as KpnI and BamHI in pIJ11268. Tet^R	Poole lab, unpublished
pOPS0085	342 bp *lppE* promoter region (oxp0146-oxp0149) cloned as KpnI and BamHI in pIJ11268. Tet^R	Poole lab, unpublished

pLMB509	Vector containing GFPmut3.1 reporter gene downstream of the inducible promoter tauAp and a ribosomal binding site (RBS), followed by a six-His tag and a transcriptional stop codon. Gm^R	[178]
pOPS0050	*lppE* gene (oxp0033-oxp0034) in pLMB509 as BD cloning (NdeI)	Poole lab, unpublished
pOPS2000	*lppE* promoter region (oxp5758-oxp5759) fused to *gusA*. Tet^R	This work
pOPS2001	*lppE* promoter region and truncated coding sequence (oxp5762-oxp5763) fused to *gusA*. Tet^R	This work
pOPS2089	Mutated *lppE* promoter region fused to *gusA*. Tet^R	This work
pOPS2090	Mutated *lppE* promoter region and wild type truncated coding sequence fused to *gusA*. Tet^R	This work
pOPS2088	*lppE* signal peptide fused to *E. coli* K12 *phoA* chain peptide and cloned into pLMB509. Gm^R	This work
pOPS2126	*lppE* promoter region (oxp6265-oxp6266) fused to *lacZ*. Tet^R	This work
pOPS2127	*lppE* promoter region and truncated coding sequence (oxp6269-oxp6270) fused to *lacZ*. Tet^R	This work
pOPS2128	Mutated *lppE* promoter region (oxp6273-oxp6274) fused to *lacZ*. Tet^R	This work
pOPS2129	Mutated *lppE* promoter region and wild type coding sequence (oxp6277-oxp6278) fused to *lacZ*. Tet^R	This work

pOPS2143	*lppE* promoter region (oxp6383-oxp6384) fused to *phoA*. Tet^R	This work
pOPS2144	Mutated *lppE* promoter region (oxp6381-oxp6382) fused to *phoA*. Tet^R	This work
pRK2013	Helper plasmid for triparental conjugation. Kan/Neo^R	[179]

Str^R: Streptomycin resistance, Spec^R: Spectinomycin resistance, Kan/Neo^R: Kanamycin and Neomycin resistance (conferred by a neomycin phosphotransferase II gene, *nptII*), Tet^R: Tetracycline resistance, Gm^R: Gentamicin resistance.

Table 2.3: Primers used in this work.

Primer	Sequence 5' → 3'	Description	Reference
oxp0028	TTTTGGTACCTGGGCTGGCTCTAC AACCGT	Amplification primer for P*lppE* and *lppE* for assembly of pOPS0047 (forward)	This work
oxp0029	TTTTGGATCCCCTCGAAGGTTTCG GGTG	Amplification primer for P*lppE* and *lppE* for assembly of pOPS0047 (reverse)	This work
oxp0033	AGGAGGAAGAACATATGATGAAA TCCGTGAAATTCCTG	Amplification primer for *lppE* to be cloned into pLMB509 (forward)	This work
oxp0034	TGGTGATGATGCATATGTCAAGGC TTGGCGAGGTC	Amplification primer for *lppE* to be cloned into pLMB509 (reverse)	This work
oxp0129	TTTTGGATCCTGGGCTGGCTCTAC AACCGT	Amplification primer for P*lppE* for assembly of pOPS0065 (forward)	This work

oxp0130	TTTTGGTACCTGAAGGTGTAGAG CGTCATGC	Amplification primer for P*lppE* for assembly of pOPS0065 (reverse)	This work
oxp0143	TCGTTTAAACGGATCCAGGATCTG GCCCTCGAAG	Amplification primer for *lppE-ecfE* intergenic region for assembly of pOPS0048 (forward)	This work
oxp0144	GAGTCCGCGGTGGGGGTACCCG GCGGCAACTGGGACCT	Amplification primer for *lppE-ecfE* intergenic region for assembly of pOPS0048 and pOPS0080 (reverse)	This work
oxp0145	GAGTCCGCGGTGGGGGTACCCAA CGACGATTGCGCCAA	Amplification primer for *lppE* and *lppE-ecfE* intergenic region for pOPS0080 (forward)	This work
oxp0146	TTTTGGTACCGGTCTGGGTCGACA TGGCGG	Amplification primer for P*lppE* for assembly of pOPS0085 (forward)	This work
oxp0147	TTTTGGTACCCGGACGAAACGAG CTTCGGC	Amplification primer for P*lppE* for assembly of pOPS0083 (forward)	This work
oxp0148	TTTTGGTACCTTTGGTGCGCCTGG AATAAA	Amplification primer for P*lppE* for assembly of pOPS0084 (forward)	This work
oxp0149	TTTTGGATCCTGAAGGTGTAGAGC GTCATG	Amplification primer for P*lppE* for assembly of pOPS0083, pOPS0084, pOPS0085 (reverse)	This work

oxp5758	CTATGACCATGATTACGCCAGAGA ATAGCGGCAAGGCG	Amplification primer for P*lppE* for HiFi assembly of pOPS2000 (forward)	This work
oxp5759	TAATTCTAGATTAGTCGACTTCCTT GACGGG	Amplification primer for P*lppE* for HiFi assembly of pOPS2000 (reverse)	This work
oxp5760	AGTCGACTAATCTAGAATTAAAGA GGAGAAATTAACC	Amplification primer for *gusA* for HiFi assembly of pOPS2000 (forward)	This work
oxp5761	AGTCGACCTGCAGGCATGCATCAT TGTTTGCCTCCCTG	Amplification primer for *gusA* for HiFi assembly of pOPS2000 (reverse)	This work
oxp5762	CTATGACCATGATTACGCCAGAGA ATAGCGGCAAGGCG	Amplification primer for P*lppE* for HiFi assembly of pOPS2001 (forward)	This work
oxp5763	CAGGACGGACGTCGACTTCCTTG ACGGG	Amplification primer for P*lppE* for HiFi assembly of pOPS2001 (reverse)	This work
oxp5764	GGAAGTCGACGTCCGTCCTGTAG AAACC	Amplification primer for *gusA* for HiFi assembly of pOPS2001 (forward)	This work
oxp5765	AGTCGACCTGCAGGCATGCATCAT TGTTTGCCTCCCTG	Amplification primer for *gusA* for HiFi assembly of pOPS2001 (reverse)	This work

oxp6145	CTATGACCATGATTACGCCACGGC ACGAGAACTTCTGG	Amplification primer for mutated P*lppE* for HiFi assembly of pOPS2090 (forward)	This work
oxp6146	CGGATTTCATGATATCTCTCCTGAG GTTCGG	Amplification primer for mutated P*lppE* for HiFi assembly of pOPS2090 (reverse)	This work
oxp6147	GAGAGATATCATGAAATCCGTGAA ATTCC	Amplification primer for *gusA* for HiFi assembly of pOPS2090 (forward)	This work
oxp6148	AGTCGACCTGCAGGCATGCACTG GATTCTCACCAATAAAAAAC	Amplification primer for *gusA* for HiFi assembly of pOPS2090 (reverse)	This work
oxp6149	CTATGACCATGATTACGCCACGGC ACGAGAACTTCTGG	Amplification primer for mutated P*lppE* for HiFi assembly of pOPS2089 (forward)	This work
oxp6150	CGGATTTCATGATATCTCTCCTGAG GTTCGG	Amplification primer for mutated P*lppE* for HiFi assembly of pOPS2089 (reverse)	This work
oxp6151	GAGAGATATCATGAAATCCGTGAA ATTCC	Amplification primer for *gusA* for HiFi assembly of pOPS2089 (forward)	This work
oxp6152	AGTCGACCTGCAGGCATGCACTG GATTCTCACCAATAAAAAAC	Amplification primer for *gusA* for HiFi assembly of pOPS2089 (reverse)	This work

oxp6153	GTGATGATGGTGATGATGCATTAT TTCAGCCCCAGAGC	Amplification primer for *lppE*::*phoA* fusion for HiFi assembly of pOPS2088 (forward)	This work
oxp6154	TAACGACAGGAGGAAGAACAATG AAATCCGTGAAATTCC	Amplification primer for *lppE*::*phoA* fusion for HiFi assembly of pOPS2088 (reverse)	This work
oxp6265	CTATGACCATGATTACGCCAGAGA ATAGCGGCAAGGCG	Amplification primer for P*lppE* for HiFi assembly of pOPS2126 (forward)	This work
oxp6266	AAACGACGGCCATTGGTTAATTTC TCCTCTTTAATTCTAGATTAG	Amplification primer for P*lppE* for HiFi assembly of pOPS2126 (reverse)	This work
oxp6267	TTAACCAATGGCCGTCGTTTTACA ACGTC	Amplification primer for *lacZ* for HiFi assembly of pOPS2126 (forward)	This work
oxp6268	AGTCGACCTGCAGGCATGCATTAT TTTTGACACCAGACCAAC	Amplification primer for *lacZ* for HiFi assembly of pOPS2126 (reverse)	This work
oxp6269	CTATGACCATGATTACGCCAGAGA ATAGCGGCAAGGCG	Amplification primer for P*lppE* for HiFi assembly of pOPS2127 (forward)	This work
oxp6270	AAACGACGGCGTCGACTTCCTTG ACGGG	Amplification primer for P*lppE* for HiFi assembly of pOPS2127 (reverse)	This work

oxp6271	GGAAGTCGACGCCGTCGTTTTAC AACGTC	Amplification primer for *lacZ* for HiFi assembly of pOPS2127 (forward)	This work
oxp6272	AGTCGACCTGCAGGCATGCATTAT TTTTGACACCAGACCAAC	Amplification primer for *lacZ* for HiFi assembly of pOPS2127 (reverse)	This work
oxp6273	CTATGACCATGATTACGCCACGGC ACGAGAACTTCTGG	Amplification primer for mutated P*lppE* for HiFi assembly of pOPS2128 (forward)	This work
oxp6274	AAACGACGGCCATTGGTTAATTTC TCCTCTTTAATTCTAG	Amplification primer for mutated P*lppE* for HiFi assembly of pOPS2128 (reverse)	This work
oxp6275	TTAACCAATGGCCGTCGTTTTACA ACGTC	Amplification primer for *lacZ* for HiFi assembly of pOPS2128 (forward)	This work
oxp6276	AGTCGACCTGCAGGCATGCATTAT TTTTGACACCAGACCAAC	Amplification primer for *lacZ* for HiFi assembly of pOPS2128 (reverse)	This work
oxp6277	CTATGACCATGATTACGCCACGGC ACGAGAACTTCTGG	Amplification primer for mutated P*lppE* for HiFi assembly of pOPS2129 (forward)	This work
oxp6278	CGACGGCCAGGTCGACTTCCTTG ACGGG	Amplification primer for mutated P*lppE* for HiFi assembly of pOPS2129 (reverse)	This work

oxp6279	GGAAGTCGACCTGGCCGTCGTTTTACAAC	Amplification primer for *lacZ* for HiFi assembly of pOPS2129 (forward)	This work
oxp6280	AGTCGACCTGCAGGCATGCATTATTTTTGACACCAGACCAAC	Amplification primer for *lacZ* for HiFi assembly of pOPS2129 (reverse)	This work
oxp6291	CACTCATTAGGCACCCCAGG	Sequencing primer for *lppE*::*lacZ* fusion plasmids (forward)	This work
oxp6292	GGCTTCTGCTTCAATCAGCG	Sequencing primer for *lppE*::*lacZ* fusion plasmids (reverse)	This work
oxp6383	CTATGACCATGATTACGCCACGCCATGAAAACTTTTGG	Amplification primer for P*lppE* for HiFi assembly of pOPS2143 (forward)	This work
oxp6384	CGGATTTCATGATGTCTCTCCTCAGATTG	Amplification primer for P*lppE* for HiFi assembly of pOPS2143 (reverse)	This work
oxp6381	CTATGACCATGATTACGCCACGGCACGAGAACTTCTGG	Amplification primer for mutated P*lppE* for HiFi assembly of pOPS2144 (forward)	This work
oxp6382	CGGATTTCATGATATCTCTCCTGAGGTTCGG	Amplification primer for mutated P*lppE* for HiFi assembly of pOPS2144 (reverse)	This work

oxp6387	GAGAGACATCATGAAATCCGTGAAATTCC	Amplification primer for *phoA* for HiFi assembly of pOPS2143 (forward)	This work
oxp6388	AGTCGACCTGCAGGCATGCATTATTTCAGCCCCAGAGC	Amplification primer for *phoA* for HiFi assembly of pOPS2143 (reverse)	This work
oxp6385	GAGAGATATCATGAAATCCGTGAAATTCC	Amplification primer for *phoA* for HiFi assembly of pOPS2144 (forward)	This work
oxp6386	AGTCGACCTGCAGGCATGCATTATTTCAGCCCCAGAGC	Amplification primer for *phoA* for HiFi assembly of pOPS2144 (reverse)	This work
oxp3173	GCCGAATTCGGATCCG	Sequencing primer for insertions into pOGG093 (forward)	Poole lab, unpublished
oxp0283	AGCGTTCTGAACAAATCC	Sequencing primer for insertions into pOGG093 (reverse)	Poole lab, unpublished
oxp6295	CGACCGCATCATCTCCATGA	RT-qPCR primer for *gyrA* (forward)	This work
oxp6296	TGACTTCTTCACCGACCAGC	RT-qPCR primer for *gyrA* (reverse)	This work
oxp6297	GCTGCTGTCGAAGTCTTTGC	RT-qPCR primer for *lovR* (forward)	This work
oxp6298	CGGTCATGGTATCGAAGGGG	RT-qPCR primer for *lovR* (reverse)	This work
oxp6299	CGGCCCGTAGATTTCGATGA	RT-qPCR primer for *recA* (forward)	This work
oxp6300	GAAGCGGCACTCTCACAGAT	RT-qPCR primer for *recA* (reverse)	This work

oxp6301	TGAAGGCGATCGACATCCTG	RT-qPCR primer for RL0770 (forward)	This work
oxp6302	GATGAACGAGGTCTGCAGCT	RT-qPCR primer for RL0770 (reverse)	This work
oxp6303	GGTCATCCAGGAACCGATCG	RT-qPCR primer for RL3684 (forward)	This work
oxp6304	CACGCGTTGGTTGTTGATGA	RT-qPCR primer for RL3684 (reverse)	This work
oxp6305	GCGGTTGTAGCTCAGTTGGT	RT-qPCR primer for RLt30 (forward)	This work
oxp6306	GGTACGGTTGGGGGGTCT	RT-qPCR primer for RLt30 (reverse)	This work
oxp6307	GGTGGAGCAGCCCGGTAG	RT-qPCR primer for RLt41/43/45 (forward)	This work
oxp6308	TGGTTGCGGGGGCAGGAT	RT-qPCR primer for RLt41/43/45 (reverse)	This work
oxp6309	CGAAAACGAACACGAGATCA	RT-qPCR primer for *infA* (forward)	[180]
oxp6310	GTAGGGCGTCATTTCCACAA	RT-qPCR primer for *infA* (reverse)	[180]

2.2 Media and antibiotics

2.2.1 Media

E. coli strains were grown in lysogeny broth (LB) at 37□C, shaken at 170 rpm [181]. For growth on solid media agar was added at 1.4% w/v prior to autoclaving. *Rhizobium* strains were gown on tryptone

yeast (TY) at 28□C, shaken at 130 rpm [182]. For growth on solid media agar was added at 1.75% w/v prior to autoclaving.

For assays requiring a minimal growth media, *Rhizobium* strains were grown on universal minimal salts (UMS) media supplemented with 10 mM glucose and 10 mM ammonium. UMS media is derived from AMS [183] with the following alterations: $CaCl_2$□$2H_2O$ (0.51 mM), $CoCl_2$□$6H_2O$ (4.2 □M), EDTA-Na_2 (1 □M), $FeSO_4$□$7H_2O$ (0.04 □M). For growth on solid minimal media agar was added at 1.75% w/v before autoclaving. For iron-supplemented minimal growth media (FMS), $FeSO_4$□$7H_2O$ was replaced by $C_6H_8O_7$ (0.04 mM) and $C_6H_5FeO_7$ (0.04 mM).

2.2.2 Antibiotics

Antibiotics were added at the following concentrations where appropriate (□g mL^{-1}): neomycin (Neo) 50; kanamycin (Kan) 20; tetracycline (Tet) 1 for *Rhizobium* and 5 for *E. coli*; gentamicin (Gm) 20 for *Rhizobium* and 10 for *E. coli*; streptomycin (Str) 500; spectinomycin (Spec) 100.

2.3 Molecular techniques

2.3.1 Isolation of genomic and plasmid DNA

Genomic DNA was isolated for PCR amplification using the DNeasy Blood and Tissue kit (Qiagen) following the manufacturer's instructions. Plasmid DNA was extracted from *E. coli* strains using the GeneJET Plasmid Miniprep kit (ThermoFisher Scientific) following the manufacturer's instructions. DNA concentration and purity was analysed using a NanoDrop One (ThermoFisher Scientific).

2.3.2 DNA amplification by PCR

Primers for PCR amplification (Table 2.3) were designed using Geneious R10 [184] and synthesised by Eurofins MWG Operon.

Q5® High fidelity DNA polymerase (New England Biolabs) was used for the amplification of DNA products for downstream cloning reactions. Cycling conditions were generally as follows:

Denaturation	98□C	1’
15 cycles	98□C	15”
	68□C (-1□C/cycle after first two cycles)	20”
	72□C	30”/kb
20 cycles	98□C	15”
	54□C	15”
	72□C	30”/kb
Extension	72□C	3’

GoTaq® Green Master Mix (Promega) was used for mapping reactions in colony PCRs. For mapping reactions in *E. coli* strains a single colony was transferred into a PCR reaction using a sterile pipette tip. For mapping reactions in *Rhizobium* strains alkaline polyethylene glycol (PEG) was used for DNA extraction [185] and 1 □L of extract used per 10 □L total PCR reaction volume. Cycling conditions were generally as follows:

Denaturation	94□C	1’
15 cycles	94□C	30”
	68□C (-1□C/cycle after first two cycles)	30”
	68□C	1’/kb
20 cycles	94□C	30”
	54□C	30”
	68□C	1’/kb

Extension	68□C	5’

2.3.3 Agarose gel electrophoresis

Gel electrophoresis was used to separate PCR products and restriction digest fragments in 1% w/v agarose (Sigma Aldrich) dissolved in 1x TAE buffer (400 mM TRIS, 1 mM EDTA, pH 8.3), containing 2 □L 100 mL^{-1} SYBR® Safe DNA Gel Stain (Invitrogen). Gels were run at 100 V for 30 minutes. 1 kb plus DNA ladder (ThermoFisher Scientific) was used as a molecular weight marker and gels were visualised on a Gel Doc EZ System (BioRad). When required, Sanger sequencing of PCR amplicons and plasmids was performed by Source Bioscience.

2.3.4 Restriction endonuclease digestions, ligations and High Fidelity assembly

Restriction digests were carried out using New England Biolabs, Roche or ThermoFisher restriction endonucleases and their respective buffers according to the manufacturer’s instructions. Amplified PCR products used for cloning were purified prior to digestion and ligations using a Monarch® PCR and DNA Cleanup Kit (New England Biolabs) following the manufacturer’s instructions. NEBuilder HiFi DNA assemblies were performed using the NEBuilder HiFi DNA assembly Master Mix as per the manufacturer’s instructions.

2.3.5 Transformation of competent cells

Plasmids were transformed into *E. coli* by chemical transformation, using chemically competent DH5α cells purchased from Bioline and thawed on ice for 10 minutes. Competent cells (50 □L) were combined with 5 □L of a ligation mixture and incubated on ice for 30 minutes, heat shocked at 42□C for 60 seconds and incubated on ice for a further 2 minutes. SOC recovery medium (200 □L) (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose; Invitrogen) was added to the transformation tube, which was then shaken at 170 rpm, 37□C for one hour. Transformation reactions were serially diluted and spread onto LB agar supplemented with

appropriate selective antibiotics and incubated overnight at 37□C. Colony PCR and sequencing was used to confirm the presence of the correct construct.

2.3.6 Conjugal transfer of plasmids

Triparental mating was used for conjugation of plasmids from *E. coli* into *R. leguminosarum* strains using the helper plasmid pRK2013. *Rhizobium* strains were grown on TY agar slopes containing the appropriate antibiotics for three days prior to the conjugation, and *E. coli* strains grown in 50 mL LB media with appropriate antibiotics overnight. TY agar slopes were washed in 3 mL TY media and all strains (*Rhizobium* and *E. coli*) transferred to Falcon tubes and spun down at 4500 rcf for 5 minutes. Strains were resuspended in appropriate media using a cut pipette tip and the process repeated twice to remove traces of antibiotics. Donor *E. coli* (400 □L), recipient *Rhizobium* strain (400 □L) and helper *E. coli* (containing pRK2013) (200 □L) were incubated together in a microcentrifuge tube and spun down at 3500 rcf for 5 minutes to pellet. The pellet was resuspended with a cut pipette tip in 30 □L TY and the mix transferred to an antibiotic-free TY agar plate and incubated overnight at 28□C before being streaked on a TY agar plate containing the appropriate selective antibiotics to remove traces of *E. coli*.

2.4 Plant growth for attachment and colonisation assays

2.4.1 Sterilisation of seeds

Pea (*Pisum sativum*) seeds were sterilised as follows:

Ethanol	70%	30”
dH_2O wash		1x
Bleach	2%	5’
dH_2O wash		6x

Barley (*Hordeum vulgare*) seeds were sterilised as follows:

Ethanol	70%	2’
dH_2O wash		2x
Bleach	7% + 0.1% Tween	5’
dH_2O wash		6x

Clover (*Trifolium pratense*) seeds were sterilised as follows:

Ethanol	70%	1’
dH_2O wash		1x
H_2O_2	35%	5’
dH_2O wash		10x

After sterilisation, seeds were transferred directly to prepared boiling tubes (for colonisation RNA sequencing) or germinated in the dark at 28□C on 0.8% distilled water agar (DWA) for 48-72 hours.

2.4.2 Preparation of boiling tubes

100 mL boiling tubes (DURAN®) were three-quarters filled with rinsed fine vermiculite and supplemented with 25 mL nitrogen-free rooting solution ($CaCl_2$□H_2O 1 mM; KCl 100 □M; $MgSO_4$□$7H_2O$ 800 □M; Fe EDTA 100 □M; H_3BO_3 35 □M; $MnCl_2$□$4H_2O$ 9 □M; $ZnCl_2$ 0.8 □M; Na_2MoO_4□$2H_2O$ 0.5 □M; $CuSO_4$□$5H_2O$ 0.3 □M; KH_2PO_4 0.5 g L^{-1}; Na_2HPO_4 0.568 g L^{-1}), sealed with a foam bung and autoclaved. Sterilised seeds or pre-germinated seedlings were transferred to prepared boiling tubes under sterile conditions, foam bungs replaced, and boiling tubes transferred to a controlled growth chamber set at 23□C with a 16 h light 8 h dark photoperiod.

2.5 RNA sequencing

2.5.1 Preparation of 'attached' rhizoplane bacteria

Seven days post sterilisation and transfer to boiling tubes, seedlings were uprooted and roots washed in dH_2O to remove attached vermiculite. *Rhizobium* suspension was generated by growing strains overnight in UMS containing appropriate antibiotics at 28□C. Cultures were centrifuged at 4500 rcf for 5 minutes to pellet cells, washed twice and resuspended to a final OD_{600} of 0.1 in 50 mL 1x PBS per plant. Plant roots were immersed in bacterial suspension with the shoots taped above the surface and left for 6 hours at 100 rpm.

2.5.2 Preparation of 'colonising' rhizoplane bacteria

Seven days post sterilisation and transfer to boiling tubes, seedlings were inoculated with 10^5 cfu bacterial cells resuspended in 1 mL UMS media. *Rhizobium* inoculum was generated by growing strains for three days on UMS agar slopes supplemented with glucose (10 mM) and ammonium chloride (10 mM) and containing appropriate antibiotics at 28□C for three days, washing in 5 mL UMS and centrifuging at 4500 rcf for 5 minutes. The pellet was resuspended and washed two times to ensure that traces of antibiotics had been removed before normalisation to an OD_{600} of 1 and serial dilution to the required cell density.

Seven days post inoculation, plants were uprooted and the root cut just below the seed.

2.5.3 Preparation of Nycodenz gradients

1x PBS was produced by adding 1 tablet 100 mL^{-1} to autoclaved milliQ water. Nycodenz (Serumwerk Bernburg AG) was dissolved in PBS to 14%, 25% and 60% w/v. A Pasteur pipette was used to transfer 5 mL each concentration Nycodenz to a 50 mL Falcon tube, slowly so as to avoid mixing of the layers. Falcon tubes were left undisturbed in a rack overnight to allow formation of a continuous density gradient.

2.5.4 Harvesting of fluorescent-marked bacterial cells from plant root surfaces

Following bacterial attachment or colonisation, the roots of three plants were placed in 25 mL 1x phosphate buffered saline (PBS; ThermoFisher) in a 50 mL Falcon tube. Tubes were vortexed for 1 minute at full speed in a Heidolph Multi Reax vortexer to detach colonising root surface or 'rhizoplane' bacteria. Roots were removed from the Falcon tube and discarded. The roots from three more plants were added to the same Falcon tube and vortexed again. This process was repeated to concentrate the 'rhizoplane' bacteria from ~20 plants into 25 mL PBS.

25 mL PBS (now containing a suspension of root-associated *Rhizobium* as well as some contaminating vermiculite particles and plant matter) was added carefully to the top of a Nycodenz gradient using a Pasteur pipette and centrifuged for 30 minutes at 2600 rcf, at 4□C. A transilluminator was used to visualise the fluorescent band which was then transferred to autoclaved centrifuge tubes. Control samples were prepared by growing *Rhizobium* overnight in UMS supplemented with glucose (10 mM) and ammonium chloride (10 mM), harvesting at an OD_{600} 0.4-0.6 and passing 25 mL liquid culture through a Nycodenz gradient to ensure consistency with the treatment of 'rhizoplane' bacteria.

2.5.5 Total RNA extraction and preparation for sequencing

Rhizobium strains were prepared for RNA sequencing either by growth on pea roots and harvesting as described above, or by culturing in TY or UMS media as described in 2.2.1. Harvested cells were spun down for 10 minutes at 10000 rpm, 4□C (Beckman Coulter Avanti JXN-30, rotor JA25.50) to pellet and resuspended in 250 □L tris hydrochloride. RNA isolation was then carried out using an RNeasy Plus Mini kit (Qiagen), following the manufacturer's instructions. Briefly, cells were lysed using a ribolyser (FastPrep FP120, ThermoFisher), passed through gDNA eliminator spin columns to remove genomic DNA and then through RNeasy spin columns to bind RNA and allow contaminants to be washed off. Following isolation, a further round of gDNA removal was carried out using an Ambion TURBO DNA-Free kit (ThermoFisher) and RNA was cleaned using a Clean and Concentrator kit (Zymo). RNA quality

and concentration was checked using the RNA 6000 Nano assay on an Agilent Bioanalyzer 2100. RNA sequencing was carried out by Novogene.

2.5.6 RNA quantification by RT-qPCR

Primers for PCR amplification (Table 2.3) were designed using Geneious R10 [184] and synthesised by Eurofins MWG Operon. *infA* (RL0616), *gyrA* (RL2401) and *recA* (RL2637) were chosen as reference genes. RNA was extracted as in 2.5.4. Reverse transcription was carried out using Invitrogen SuperScript™ IV Reverse Transcriptase. Briefly, 50 □M Random Hexamers (Invitrogen) were annealed to 500 ng template RNA at 65□C. Reverse transcription was carried out for 10 minutes at 23□C and 10 minutes at 52□C. Enzyme was inactivated at 80□C for 10 minutes. PCR amplification was carried out in 20 □L reactions containing 1 □L cDNA. Cycling conditions were as follows:

Denaturation	95□C	10’
40 cycles	95□C	15”
	60□C	1’
Melt curve	95□C	15”
	60□C	1’
	95□C	15”

2.5.7 Bioinformatic analysis of RNAseq data

FastQC was used for quality control and reads from sequencing by Novogene were uploaded onto a Linux server for analysis. mRNA sequence reads were mapped using protein translation to the Rlv3841 genome and the degree of gene expression quantified using EDGE-pro [186]. Further analysis was carried out using the DESeq2 package in R and Microsoft Excel. Gene set enrichment analysis (GSEA)

was carried out using the FUNAGE-pro tool [187]. Raw data (.fq) files were stored in the department server and will be uploaded to RefSeq.

2.6 Assays using *lux* promoter fusions

2.6.1 Endpoint induction assays in liquid culture

Strains harbouring *lux* promoter fusions were grown on UMS agar slopes supplemented with 10 mM ammonium and 10 mM glucose and containing appropriate antibiotics at 28□C for three days, then resuspended in 4 ml UMS media. Suspension was filtered through a 40 □m cell strainer (Falcon). Cells were inoculated into 10 mL UMS media supplemented with 10 mM glucose and 10 mM ammonia in 36 mL universal bottles at a starting OD_{600} of 0.01. Cultures were incubated at 28□C, 130 rpm for approximately 16 hours, until an OD_{600} of 0.4-0.6 was reached. OD_{600} was measured using a Genesys 150 spectrophotometer (Thermo Scientific) and luminescence read using a Glomax Multi JR reader (Promega). Promoter induction was expressed as luminescence normalised by absorbance (RLU/OD_{600}).

2.6.2 Continuous induction assays in liquid culture

Strains harbouring *lux* promoter fusions were grown on UMS agar slopes, then in liquid culture overnight as in 2.6.1. 50 □L cells were added to each well of a white, clear-bottomed 96-well plate (Greiner Bio-One) along with 150 □L UMS media. Cultures were grown in a FLUOstar Omega plate reader (BMG LABTECH) for 24 hours at 28□C with 700 rpm rotary shaking. Luminescence and OD_{600} were measured every 30 minutes.

2.6.3 Induction assays on solid media

Strains harbouring *lux* promoter fusions were grown on UMS agar slopes supplemented with 10 mM ammonium and 10 mM glucose and containing appropriate antibiotics at 28□C for three days, then resuspended in 4 ml UMS media. Suspension was filtered through a 40 □m cell strainer (Falcon). OD_{600}

was measured using a Genesys 150 spectrophotometer (Thermo Scientific) and luminescence read using a Glomax Multi JR reader (Promega).

2.6.4 *In planta* induction assays

Rhizobium inoculum was generated by growing strains harbouring lux promoter fusions in pIJ11268 on UMS agar slopes supplemented with 10 mM ammonium and 10 mM glucose and containing appropriate antibiotics at 28□C for three days, washing in 5 mL UMS and centrifuging at 4500 rcf for 5 minutes. Pellets were resuspended and washed a following two times to ensure that traces of antibiotics had been removed before normalisation to an OD_{600} of 1 and serial dilution to the required cell density.

12.5 cm square plates were prepared by pouring FP agar ($CaCl_2$□$2H_2O$ 40 g L^{-1}; $MgSO_4$□$7H_2O$ 40 g L^{-1}; KH_2PO_4 30 g L^{-1}; Na_2HPO_4□$12H_2O$ 45 g L^{-1}; $FeC_6H_5O_7$ 2.5 g L^{-1}; H_3BO_3 2.86 g L^{-1}; $MnSO_4$□$4H_2O$ 2.03 g L^{-1}; $ZnSO_4$□$7H_2O$ 220 mg L^{-1}; $CuSO_4$□$5H_2O$ 80 mg L^{-1}; H_2MoO_4 80 mg L^{-1} with 5 g L^{-1} agar) into plates set on a slant so that agar solidified as a slight slope. Once dry, a red-hot scalpel was used to cut a 5 mm^2 section from the top edge of the plate, by the narrowest edge of the FP agar slant. Seedlings were sterilised and germinated on 0.8% DWA as in 2.4.1, then transferred to FP plates with the seed at the top of the plate so the shoot pointed out of the cut-out. Autoclaved Whatman filter paper was placed above and below the seedling to wick moisture from the surface of the agar and ensure the root did not dry out. Each seedling was inoculated with 2x 10^7 cfu *Rhizobium* in 200 □L UMS media before placing the top piece of filter paper. Plates were taped closed with micropore tape, wrapped in aluminium foil to keep the roots in the dark and transferred to a controlled growth chamber set at 23□C with a 16 h light 8 h dark photoperiod.

7 days post inoculation plates were imaged using a NightOWL camera (Berthold). Image analysis was carried out using IndiGO software.

2.6.5 *In planta* induction timelapses

Rhizobium inoculum, pea seedlings and FP square plates were prepared as described above. Following 48-72 h germination on 0.8% DWA, pea seedlings were transferred to FP square plates (one per plate) and grown for a further 3 days. 1 mL *Rhizobium* inoculum normalised to an OD_{600} of 1 was added to 60 mL 0.1% agarose heated to 30□C. A pea seedling was transferred to a new FP plate with no cutout and *Rhizobium*-agarose mix poured slowly over it. Plates were placed in the NightOWL with the lid off and a schedule set to record images every 30 minutes for 16 hours. Images were compiled into a timelapse using IndiGO software.

2.6.6 Root colonisation assays

Colonisation assays were carried out using the same method as that for *in planta* induction assays, above, but plants were inoculated with *Rhizobium* strains harbouring the constitutive *lux* plasmid pIJ11282 rather than a lux promoter fusion strain. Images were taken using the NightOWL (Berthold) 7 days post inoculation and analysed using IndiGO software.

2.6.7 Root tip attachment assays

Sterile pea seeds were germinated on DWA for five days, or until approximately 5 cm in length. *Rhizobium* strains were grown overnight at 28□C, 130 rpm in UMS with 10 mM glucose and 10 mM ammonia, supplemented with appropriate antibiotics. Strains were harvested at 4000 x g, then washed twice in UMS to remove traces of antibiotics and normalised to an OD_{600} of 1, then diluted to OD_{600} 0.1 in 12 mL PBS per plant treated. OD_{600} (Genesys 150 spectrophotometer, Thermo Scientific) and luminescence (Glomax Multi JX Reader, Promega) of each resuspension was measured for normalisation. Roots were cut off pea seedlings and ten roots placed in a 30 mL universal glass vial with 10 mL bacterial cell suspension. This process was repeated for each bacterial strain to be tested. Universal vials were sealed closed with parafilm and placed horizontally in the dark at 100 rpm for two hours. Plants were removed, rinsed twice in PBS and patted dry on Whatman filter paper. Roots treated with each bacterial strain or treatment were placed on a 12.5 cm square plate. Plates were left in the

dark for 5 minutes, then imaged using a 20 second exposure using the NightOWL (Berthold). Images were analysed using IndiGO software.

2.7 Preparation and characterisation of plant protein extracts

2.7.1 Crude extraction of pea root protein

Pea seedlings were sterilised and grown in boiling tubes for two weeks as described above. Seedlings were uprooted and rinsed in dH_2O to remove excess vermiculite. Roots were cut off just below the seedling. Roots were weighed and blended for 45 seconds in a Waring blender with 150 mL 1x PBS and 20 mM 2-mercaptoethanol. Root mix was strained through four layers of muslin cloth, then stirred at 4□C for one hour. The mix was centrifuged at 500 x g, 10 minutes and the supernatant collected and saturated with 90% ammonium sulphate to precipitate proteins. Precipitation was carried out overnight at 4□C, stirring continuously. Precipitated proteins were pelleted by centrifugation at 10,000 rpm, 4□C (Beckman Avanti JXN-26, rotor JA25.50) and resuspended in 1x PBS. Ammonium sulphate was removed by dialysis against 1x PBS using 30 mL 3.5 kDa MWCO Slide-a-Lyzer dialysis cassette (Thermo Scientific). Dialyzed samples were spun down at 20,000 x rcf, 20 minutes (Beckman Avanti JXN-26, rotor JA25.50) and supernatant retained and filtered through a 0.45 □m filter and a 0.22 □m filter. Samples were concentrated using a 50 kDa filter (Amicon) at 4000 x g, 4□C.

2.7.2 Purification of glycoproteins

Glycoproteins were purified from crude protein extract using a HiTrap ConA 4B column (Cytiva), following the manufacturer's instructions. Gycoproteins were then purified further using a HiTrap Q HP anion exchange chromatography column (Cytiva) or Superdex 200 size exchange chromatography column (Cytiva). Output fractions were tested for promoter induction activity following the protocol outlined in 2.6.2. Protein concentration was assayed at each stage of purification using the Qubit 2.0 Fluorometer with the Qubit Protein Assay kit (Invitrogen).

2.7.3 Mass spectrometry

Fractions exhibiting promoter induction activity were prepared for proteomic analysis using mass spectrometry by in solution digestion. Briefly, proteins were pelleted overnight in a DNA SpeedVac 110 (Savant), then resuspended in 1 mL denaturing buffer (6 M urea, 2 M thiourea, 50 mM ammonium bicarbonate) per 5 □g protein. Proteins were reduced by addition of 10 mM TCEP, then incubated at room temperature in the dark for 30 minutes. 50 mM iodoacetamide was then added and samples incubated in the dark for a further 20 minutes. Samples were then diluted with 50 mM ammonium bicarbonate buffer to reduce the urea concentration to 1 M and $CaCl_2$ added to a final concentration of 1 mM. Trypsin/LysC mix (Promega) was diluted to 1 □g □L^{-1} and added to the samples at a 1:100 ratio of protease to protein. Digestion reactions were incubated overnight at 37□C, 1000 rpm. Digests were then transferred to -20 □C to inactivate enzyme. Peptides were then thawed and clean-up performed using Waters Sep-Pak C18 Plus Short cartridges (Thermo Scientific). Following elution from the C18 matrix in 65% acetonitrile, 0.1% TFA, purified peptides were dried in the SpeedVac. Mass spectrometry and analysis was carried out by the Analytics Core Facility Essen.

2.8 Colourimetric enzyme assays

2.8.1 Beta-glucuronidase activity assay

Beta-glucuronidase activity was measured using the reaction with 2-nitrophenyl β-D-glucuronide (PNPG). Bacterial strains were grown for approximately 16 hours from a starting OD_{600} of 0.01 in 10 mL UMS liquid culture supplemented with 10 mM ammonia and 10 mM glucose at 28□C, 130 rpm. Final OD_{600} was read using a Genesys 150 spectrophotometer (Thermo Scientific) to confirm an OD_{600} of 0.4-0.6 had been reached. An appropriate volume (0.1-1 mL) cells was harvested and spun down then resuspended in 1 mL NaCl (0.9%). 85 □L sample cell suspension was added to 915 □L assay buffer (Na_2HPO_4 28.9 mM; NaH_2PO_4□$2H_2O$ 21.1 mM; EDTA 2 mM; sarkosyl 0.1%; Triton X-100 0.1%; β-mercaptoethanol 10 mM; pH 7.0) in 2 mL Eppendorf tubes. A blank tube was also set up with no cells added. 30 □L chloroform was added to each sample and tubes vortexed. 200 □L PNPG (4.2 mg mL^{-1} in

water) was added to each assay tube and the blank tube and the tubes vortexed and placed in a heat block at 37□C. As each reaction turned yellow it was stopped by adding 500 □L stopping reagent (2 M Na_2CO_3) and time taken recorded. After 30 minutes all reactions were terminated. Reactions were centrifuged for 5 minutes at 13,000 rpm to remove cell debris, and absorbance of supernatant read at 405 nm. Beta-galactosidase activity was expressed in Miller units.

2.8.2 Alkaline phosphatase activity assay

Alkaline phosphatase activity was measured using the reaction with *p*-nitrophenylphosphate (*p*-npp). Bacterial strains were grown on UMS agar slopes supplemented with 10 mM ammonia and 10 mM glucose at 28□C for three days, then resuspended in 0.1 M Tris chloride, pH 8.0 and OD_{600} measured. 20 □L chloroform and 10 □L 10% SDS were added to 1 mL resuspended culture and vortexed. 100 □L cells were added to 700 □L Tris chloride and incubated at 30□C for five minutes. A blank tube was set up with no cells added. 100 □L 1 M K_2HPO_4 (stopping reagent) was added to the blank tube, then 100 □L *p*-npp (4 mg mL^{-1} in 0.1 M Tris chloride) was added to each assay tube and to the blank tube. As each reaction turned yellow it was stopped by adding 100 □L stopping reagent and time taken recorded. After 20 minutes all reactions were terminated. Reactions were centrifuged for 10 minutes at 11,000 rpm to remove cell debris, and absorbance of supernatant read at 420 nm. Alkaline phosphatase activity was expressed in Miller units.

2.8.3 Calculation of Miller units

Enzyme activity measurements were expressed using Miller units, calculated as follows:

$$\frac{1000 * OD_x}{vol_{mL} * OD_{600} * time_{minutes}}$$

For the beta-glucuronidase assay, OD_x was OD_{405}, while for the alkaline phosphatase assay OD_x was OD_{420}.

2.9 Fahraeus slides

Fahraeus slides were prepared as described [188]. Briefly, 25 mL nitrogen-free rooting solution (described in 2.4.2) was added to 100 mL glass boiling tubes, sealed with a foam bung and autoclaved. Clover (*T. pratense*) seeds were sterilised and germinated as described in 2.4.1. Fluorescent-labelled *Rhizobium* strains were prepared by growing on TY or UMS agar slopes for three days at 28□C, then resuspended in UMS media and OD_{600} measured. Cells were diluted in molten FP agar to a final concentration of 10^8 cfu mL^{-1} and 2 g L^{-1} FP. 300 □L cell suspension was pipetted onto a microscope slide sterilised with 70% ethanol and a clover seedling placed on top. A coverslip sterilised with 70% ethanol was added and slides left to dry before being transferred to boiling tubes containing 50 mL sterile N-free rooting solution (as in 2.4.2). Boiling tubes were transferred to a controlled growth chamber set at 23□C with a 16 h light 8 h dark photoperiod and left overnight before imaging using a confocal microscope (LEICA TCS SP5).

2.10 Bioinformatics, data handling and statistical analysis

The Rlv3841 genome annotation was acquired from a lab-curated genome annotation file. Global nucleotide and protein alignments were carried out using BLASTn and BLASTp (NCBI) respectively [189]. Geneious R10 was used to carry out local alignments [184]. Protein-protein interaction networks were predicted using STRING [190]. Gene coexpression networks were generated using a lab-developed tool [191]. Multiple sequence alignments were generated using the Pro-Coffee tool [192], [193], [194]. Protein interaction modelling was carried out using Alphafold Multimer via ColabFold with MMSeqs 2 [195]. Protein membrane topology predictions were generated using TOPCONS [196]. Data handling was carried out in Excel and statistical tests were carried out and graphs generated using GraphPad Prism 9. ANOVA and t tests were used with the following p values: ns = $p > 0.05$, * = $p \leq 0.05$, ** = $p \leq 0.01$, *** = $p \leq 0.001$, **** = $p \leq 0.0001$. Analysis of RNA sequencing data was carried out using DESeq2 in R, Microsoft Excel and FUNAGE-pro [187]. Confocal imaging data was analysed using FIJI

[197] and LAS AF (LEICA). NightOWL imaging data was analysed using IndiGO (Berthold) and Microsoft Excel.

3 Development and validation of RNA sequencing of plant-colonising bacteria

3.1 Introduction

3.1.1 Colonisation of the rhizoplane

The area of soil surrounding plant roots, known as the rhizosphere, is occupied by a diverse and complex range of microorganisms, all of which are attempting to benefit from the nutrients in plant root exudates. A subset of these microorganisms are able to attach to the root surface, or rhizoplane, to benefit from a closer association with host roots. Colonisation of the rhizosphere or the rhizoplane is an extremely competitive process. As not all microorganisms are beneficial, the plant also exerts some control over which microorganisms are able to establish, creating a layer of host-microbe specificity in addition to the microbe-microbe competition. The process of rhizoplane colonisation is generally split into four stages: (i) root exudates drive chemotaxis towards the roots, (ii) bacterial primary attachment, (iii) secondary attachment and (iv) formation of a biofilm, a bacterial community surrounded by a matrix composed primarily of polysaccharides [121]. Bacterial colonisation is defined as growth in the rhizosphere or at the root surface over a period of days, while attachment is a faster process that prevents bacterial growth or cell division over a period of hours.

Plant growth-promoting rhizobacteria (PGPRs) are a subgroup of soil bacteria that act in symbiosis with plants, being provided with nutrients by the plant in exchange for providing beneficial services and promoting plant growth. This growth promotion can be direct, through the synthesis and secretion of phytohormones or phytohormone mimics. For example, bacteria of the genus *Azospirillum* secrete auxin which promotes growth in the roots, in particular leading to increased length of lateral roots and an increased number of root hairs, providing a higher surface area for nutrient uptake [198]. It can also be indirect; for example, through biological control of pests and parasites by the production of antifungal and antibacterial compounds by some *Pseudomonas* species. Some *Pseudomonas* species also promote growth by increasing solubility of nutrients such as phosphates, which improves their availability to plant hosts [199]. Because of these properties, PGPRs are widely studied and can be

used as agricultural inoculants to improve plant growth. For example, *P. fluorescens* and *P. putida* are widely used biological inoculants in agriculture [199].

One of the most widely studied and agriculturally significant benefits that PGPRs can convey upon plant hosts is the provision of biologically fixed nitrogen. Nitrogen is an essential element for crop growth, but its availability is often a limiting factor. Biological nitrogen fixation (BNF) has been an important contributor of nitrogen to agricultural systems throughout history. This process is driven by the bacterial enzyme nitrogenase breaking the triple bond in atmospheric dinitrogen and converting it into ammonia that can be assimilated by plants. However, a significant proportion of agricultural nitrogen is supplied as synthetic fertiliser containing reactive nitrogen generated using the Haber-Bosch process. This reactive nitrogen has been extremely important in growing agricultural output to meet the demands of a growing population, with reactive nitrogen synthesis growing by 20% between 1995 and 2005 and cereal production rising by a corresponding 20% during the same period [200]. However, application of synthetic fertiliser is linked with emissions of the greenhouse gas nitrous oxide (N_2O), meaning that this rising use of synthetic nitrogen fertiliser is likely to be extremely damaging to the global climate. Expansion of the exploitation of biological nitrogen fixation in the agricultural industry could reduce reliance on synthetic fertiliser and therefore this segment of anthropogenic greenhouse gas emissions. Studying these BNF systems is therefore essential to allow them to be more efficiently incorporated into agricultural practice. For example, greater knowledge could allow more informed selection of PGPRs to be applied to crops and the design of more effective PGPR communities. Genetic modification of PGPRs could improve their colonisation of plant roots above that of non-beneficial microorganisms, allowing improved transfer of nitrogen to plants.

The most widely studied and agriculturally significant example of BNF is in rhizobia that associate with the roots of legume plants. Around 40 million tonnes of nitrogen is fixed by rhizobia in symbiosis with legumes every year [21]. However, only 10 out of 380 families of angiosperms are able to associate with rhizobia, and all but one, the Gunneraceae, are members of the Eurosid clade [20]. Other plants, including crop plants, can benefit from the nitrogen fixed in the *Rhizobium*-legume symbiosis as

legumes are often grown in rotations to enrich agricultural soils with nitrogen that can be used by subsequent crops. However, staple crops such as rice, wheat and maize cannot directly benefit from biologically fixed nitrogen. Work is ongoing to introduce BNF into these staple crops so that their yields can be increased independently of synthetically fixed nitrogen. Three main approaches are being pursued. The first is engineering plants to express their own nitrogenase, often in chloroplasts, and fix their own nitrogen [57], [58]. Secondly, engineering both plants and bacteria with the aim of generating nitrogen-fixing nodules on staple crop plants, particularly on cereal plants [54], [60], [62]. Finally, engineering diazotrophic rhizosphere PGPRs, including *Rhizobium* species, to fix and release nitrogen either constitutively [63], [65] or in the presence of a specific crop plant [55], [64] while colonising plant rhizoplanes. For any approach that involves engineering Rhizobium to provide nitrogen to non-host plants to work, it will be important to understand how these species interact with their legume hosts, as well as how that differs from their existing ability to colonise non-host plants. In addition to colonisation and nodulation of their host legumes, *Rhizobium* strains have been shown to be able to colonise the roots of staple crops including maize [201], [202] and rice [203]. Since these interactions do not result in nodulation they must differ from the interactions of *Rhizobium* with host legumes. It is vital to understand the differences between these interactions in order to be able to make use of them in an agricultural setting.

3.1.2 RNA sequencing

RNA sequencing can be used to generate transcriptomic data that provides insight into which genes and pathways are involved in attachment to and colonisation of legume plants by rhizobia. RNA sequencing and other expression analysis techniques such as microarrays have been used to characterise the transcriptomes of a variety of different PGPRs, including *Pseudomonas* spp and *Rhizobium* spp, during colonisation of the roots of a range of plant species, including maize, canola, soybean and pea [204], [205], [206], [207], [208]. In general, to isolate root-colonising bacteria, two key methodological approaches have been implemented, as shown in Figure 3.1. The first is based on

growing plants either in hydroponic systems or in inert, solid substrates such as vermiculite, inoculated with bacterial liquid culture. Following an appropriate time interval for bacterial colonisation, the growth media, whether solid or liquid and plant roots are separated, and either the liquid media is centrifuged to produce bacterial pellets that can be pooled before RNA extraction (e.g. *Paenibacillus* on barley [208]) or the solid growth media is strained off to recover bacteria that are pelleted before pooling for RNA extraction (e.g. *P. putida* on maize [204]; *R. leguminosarum* on barley [205]) (Figure 3.1A). Alternatively, the plants can be grown in be grown in a hydroponic system or solid media inoculated with bacteria for an appropriate time period, then uprooted and the roots vortexed in buffer to remove attached or colonising microorganisms (Figure 3.1B). The resulting bacterial suspension can be centrifuged to pellet bacteria and extract RNA. This approach has been used to characterise the transcriptome changes in *Herbaspirillum seropedicae* [209] and *P. protegens* [206] during colonisation of wheat. In general, these approaches are used to characterise transcriptome changes during longer-term colonisation of plant roots, meaning that bacteria are allowed to grow on plant roots for a period of 6-7 days before harvesting. However, as demonstrated by [208], transcriptomic data can also be obtained from bacteria during early attachment if harvesting of cells takes place after a period of six hours.

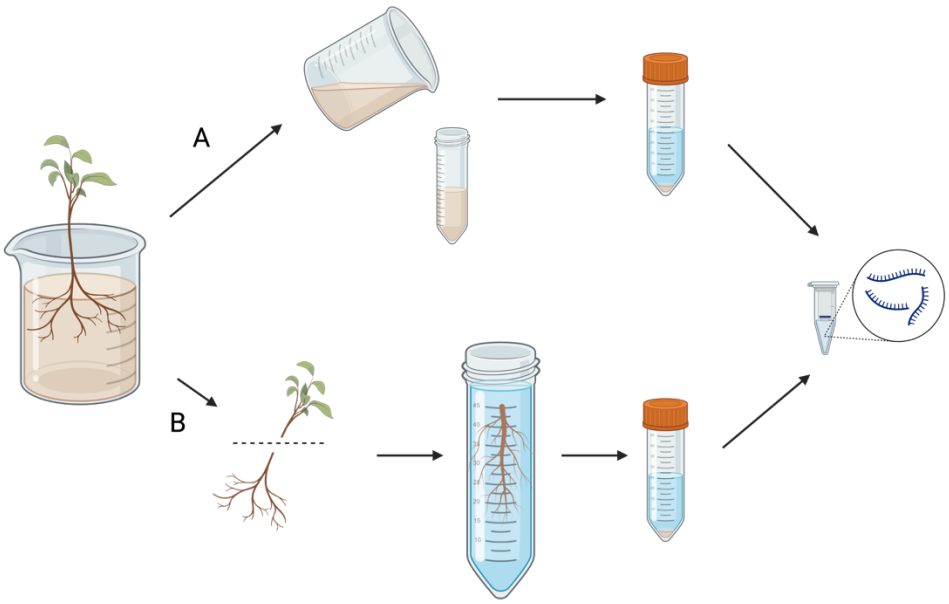


Figure 3.1: Two methodological approaches used to prepare samples for RNA sequencing of root-colonising bacteria. *A, Pelleting of bacterial cells from inoculated plant growth media and extraction from the resultant pellet. B, Vortexing of the plant root to remove firmly attached bacteria, followed by pelleting of these cells and RNA extraction.*

The approach represented in Figure 3.1A, where bacterial RNA is extracted from the inoculated media, has been used to characterise the transcriptome of 'rhizosphere' bacteria that colonise the area of soil or media immediately surrounding the plant root but do not themselves attach to the root surface. However, it cannot provide specific expression data for rhizoplane bacteria that are firmly attached to the root surface. To obtain a view of the transcriptomes of rhizoplane bacteria during colonisation, it is necessary to use the approach represented in Figure 3.1B, where vortexing the root to remove microorganisms from its surface will result in a bacterial pool that is enriched for those cells that have been attached to the root or grown on the root surface. It is important to isolate this fraction of cells to search for genes and pathways that are directly involved in interactions with the root surface specifically, as opposed to those that are related solely to metabolism of root exudates or survival in the growth media used for the experiment. However, unpublished Poole lab data show that using this

method of sampling on *R. leguminosarum* following seven-day colonisation of pea roots results in RNA sequencing datasets that are highly enriched for reads mapping to the pea genome as a result of a high presence of plant cells in the samples. RNA sequencing is not hybridisation-based, meaning that probes are not used to guide which RNA fragments are amplified and sequenced. This gives RNA sequencing an advantage over hybridisation-based methods such as microarrays by allowing sequencing to be carried out independent of annotation, facilitating discovery of new coding regions and antisense transcripts. However, it also means that any RNA in the sample is sequenced regardless of whether it is of plant or bacterial origin, meaning that contaminating plant cells must be removed before sequencing. An excessive number of reads mapping to the plant genome makes the transcriptome sequencing data obtained uninformative due to the lack of reads mapping to the bacterial genome, resulting in a lack of statistical significance for any comparative analysis. It is therefore necessary to develop a new method of separation of colonising bacterial cells from plant roots in order to obtain meaningful transcriptomic data from RNA sequencing in this model system.

3.2 Results and Discussion

3.2.1 Plant root material can be separated from bacterial cells using a Nycodenz gradient

In order to separate *R. leguminosarum* cells from pea roots, a differential spin method using Nycodenz was developed. Nycodenz is a non-ionic, non-toxic, inert substance that is water-soluble and can be dissolved to different concentrations and used to form density gradients for the separation of different cell types [210]. It has been extensively used for the separation of different blood cell types [211], [212], and a similar substance, Gentodenz, was successfully used for the separation of *Dickeya dadantii* from *Arabidopsis thaliana* leaves [213]. In this work, I used a continuous density gradient of Nycodenz dissolved in phosphate-buffered saline (PBS) to separate pea root fragments from mCherry-labelled Rlv3841 (OPS2453). A fluorescent-tagged strain was used to allow easy visualisation of the rhizobia during separation from plant cells. Comparison of samples obtained from the pea rhizoplane and

processed with Nycodenz with samples prepared by vortexing pea roots in PBS to dislodge bacterial cells, pelleting the cells and proceeding directly to RNA extraction shows a significant decrease in reads aligning to the pea genome following treatment with Nycodenz (Figure 3.2). On average, samples treated with Nycodenz show 23% of reads mapping to the pea genome, as opposed to an average of 70% of reads in samples prepared with direct RNA extraction. This enabled me to reliably obtain statistically significant transcriptome data with approximately 1/3 of the sequencing depth needed for samples prepared by direct extraction, reducing costs and the processing power and time necessary to analyse larger data files.

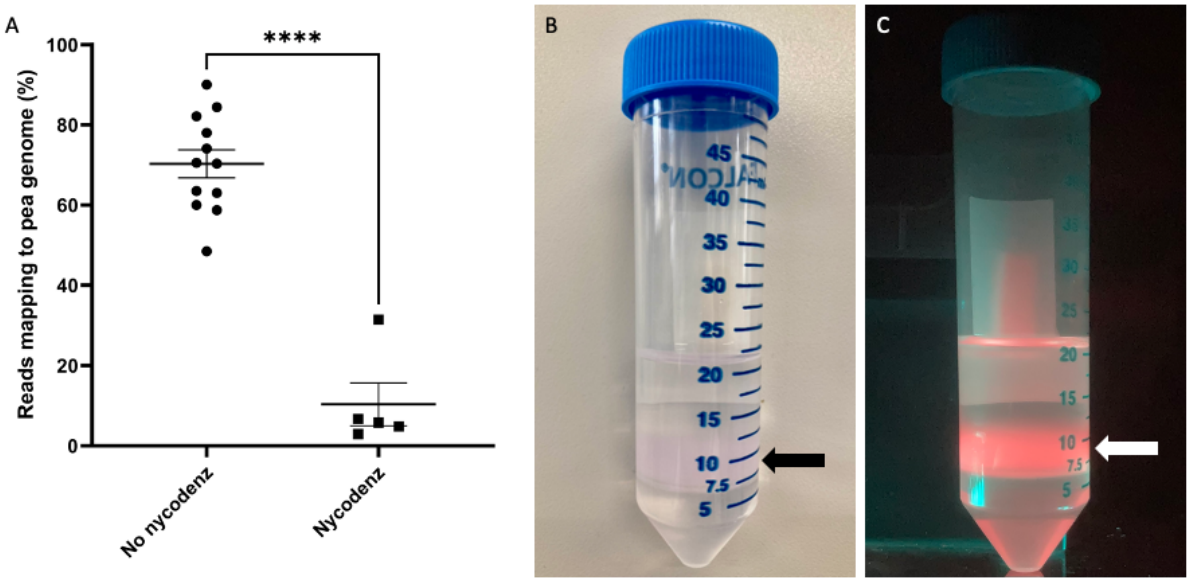


Figure 3.2: Separation of plant material from bacterial cells using a Nycodenz gradient. *A, percentage of reads mapping to the pea genome. The RNA samples were extracted directly from pelleted cells vortexed off pea roots, and from cells vortexed off pea roots then further separated from plant material by differential spin in a Nycodenz gradient. Values are mean ± SEM. Significance calculated using an unpaired t-test. n≥5. ns=p>0.05; *=p≤0.05; **=p≤0.01; ***=p≤0.001; ****=p≤0.0001. B, Nycodenz gradient showing distinct band of mCherry-labelled bacterial cells. C, Nycodenz gradient with distinct bacterial band imaged with UV transillumination.*

3.2.2 Rlv3841 exhibits distinct transcriptomic signatures during colonisation of pea and barley

RNA was isolated from Rlv3841 during colonisation of the rhizoplanes of its host legume pea (*P. sativum*) and the model cereal barley (*H. vulgare*). Cells were inoculated onto 7-day-old seedlings, then harvested by vortexing roots 7 days post inoculation (dpi). Both pea and barley datasets were compared with cells grown in laboratory culture on minimal media. Three biological replicates were obtained for colonisation of pea and one for barley. RNA sequencing was obtained to an average depth of 53.2 x 10^6 reads for Rlv3841 extracted from the pea rhizoplane, and 12.9 x 10^6 for Rlv3841 extracted from the barley rhizoplane. Genes that were over 4-fold differentially regulated, with an adjusted p-value <0.05 (in the case of the pea samples) were further analysed. Data concerning genes discussed in this chapter is summarised in Supplementary Table 1. Gene set enrichment analysis was carried out on differentially regulated gene sets using the FUNAGE-pro tool [187] and genes classified by ontology [214] to establish pathways that were significantly differentially regulated. Principal component analysis (PCA) showed that colonisation of a plant rhizoplane accounts for the majority of difference in gene expression, with samples harvested from pea and barley roots being substantially different from that harvested from laboratory culture (Figure 3.3). This is reflected in the amount of overlap between differentially expressed genes during colonisation of pea and barley. A total of 365 genes were significantly upregulated during colonisation of the barley rhizoplane and 832 during colonisation of pea. Of these, 291 overlapped, accounting for 80% of genes upregulated on barley roots and 35% of genes upregulated on pea roots. 1581 genes were significantly downregulated during colonisation of barley and 346 during colonisation of pea. Of these, 327 overlapped, accounting for 21% of genes downregulated on barley roots and 95% of genes downregulated on pea roots. With only one sample harvested from barley roots, it is difficult to establish whether growth on different plant rhizoplanes causes significantly different gene expression using PCA. However, examination of the data shows that there are substantial differences between growth on the two plants. Overall, it is clear that a much larger proportion of the Rlv3841 genome was downregulated during colonisation of barley, and upregulated during colonisation of pea. However, there were genes that were significantly upregulated

on barley but not pea, and genes that were significantly downregulated on pea but not barley, creating distinct transcriptomic signatures for Rlv3841 during colonisation of each model plant.

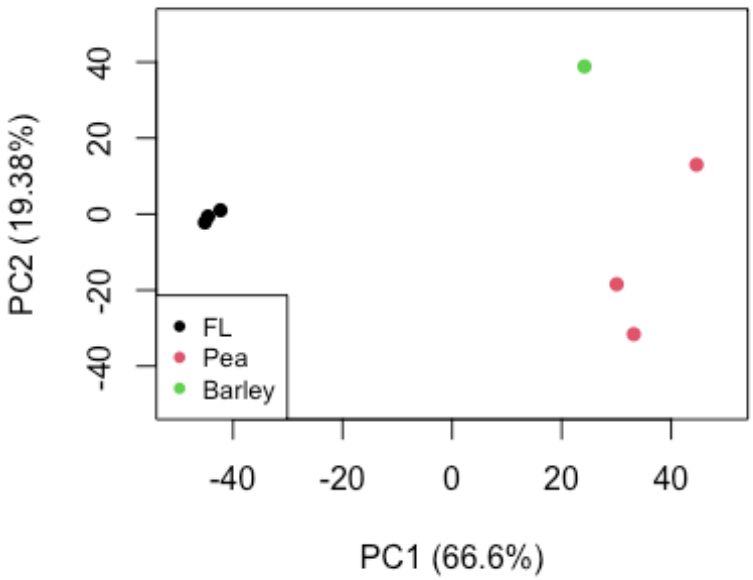


Figure 3.3: Principal Component Analysis of Rlv3841 gene expression during colonisation of pea and barley and in laboratory culture ('FL') (UMS supplemented with glucose (10 mM) and ammonium chloride (10 mM).

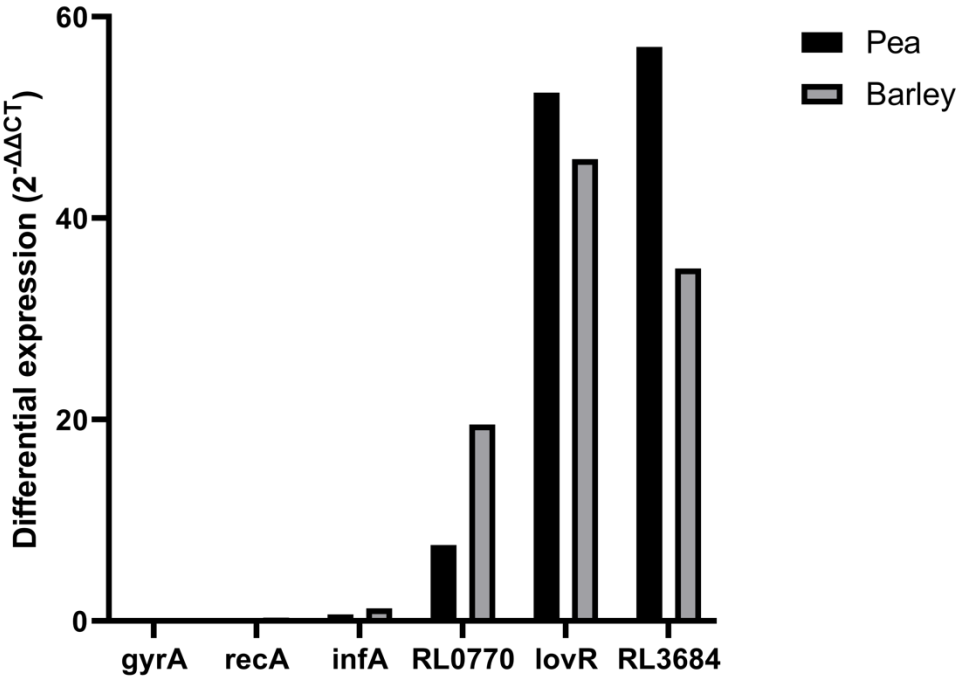
3.2.3 RNA sequencing results were confirmed by RT-qPCR

RT-qPCR was used to confirm upregulation of key genes during colonisation of the pea and barley rhizoplanes (Figure 3.4). RL0770, RL2946 (*lovR*) and RL3684 were all upregulated during colonisation of both pea and barley according to RNAseq data and were selected for RT-qPCR to confirm that this data was reliable. RL0770 was 5.8-fold upregulated during colonisation of pea and 4.4-fold upregulated during colonisation of barley according to RNAseq results. According to RT-qPCR it was 7.5-fold upregulated during colonisation of pea and 19.5-fold upregulated during colonisation of barley. RT-qPCR therefore confirms upregulation of RL0770 during colonisation of both plants, but suggests that there is no significant difference in its expression during colonisation of each plant, as it is not consistently more upregulated on one than the other when compared with laboratory culture. RL0770 encodes a phasin 2 family protein (IPR018968). Phasin 2 family proteins associate with

polyhydroxyalkanoate (PHA) inclusions, which are polymers used by bacteria to store carbon under nutrient-limited growth conditions. Phasin 2 family proteins are specifically associated with short-chain PHAs, including the most common PHA, polyhydroxybutyrate (PHB) [215]. RL0770 shares 72.109% protein sequence identity with the *S. meliloti* phasin-encoding gene SMc00777 (*phaP1*). SMc00777 is also regulated by the quorum sensing receptor ExpR [216]. The ExpR regulatory network plays a role in biofilm formation in *S. meliloti* [152], making it logical for Rlv3841 orthologues of these genes to be upregulated during colonisation of plant roots. The SMc00777 mutant grows more slowly than wild type and shows reduced PHB synthesis but makes more EPS and glycogen than wild type [217]. This suggests that PhaP1 is necessary for storage of carbon as PHB and in its absence this carbon is stored elsewhere. The SMc00777 mutant also forms fewer nodules on alfalfa roots than wild type [217], implicating it in symbiosis formation. Both SMc00777 and RL0770 were significantly upregulated under nitrogen-limited conditions [218], [219]. This suggests that RL0770 is important for survival in the nutrient-limited rhizosphere as well as for colonisation of plant roots and, in the case of legume colonisation, for nodule formation.

RL2946 (*lovR*) encodes the regulator component of a two-component sensor/transcriptional regulator pair. According to RNAseq, it was 115.9-fold upregulated during colonisation of the pea rhizoplane and 29.1-fold upregulated during colonisation of the barley rhizoplane. Upregulation on both plants, as well as more substantial upregulation during growth on pea roots, was confirmed by RT-qPCR (Figure 3.4). RL2946 is also substantially upregulated during growth on acidic pHs, during growth on phenylalanine, a component of pea root exudates, and in pea bacteroids [219]. This suggests that, like RL0770, it is important for a range of life stages, including survival of stress conditions, colonisation of plant roots, and progression of symbiosis with legumes. Finally, RL3684, encoding a transmembrane protein, was 712.0-fold upregulated on pea roots and 82.0-fold upregulated during growth on barley roots. Again, upregulation on both plants and more significant upregulation on pea was confirmed by RT-qPCR (Figure 3.4). RL3684 encodes a DUF1236 domain-containing protein. Proteins containing DUF1236 domains are significantly enriched in the proteomes of PGPRs on soybean and rice

phyllospheres [220], [221]. A membrane-localised DUF1236 domain protein plays an important role in aluminium tolerance in a strain of *R. phaseoli* that is tolerant to heavy metal stress [222]. However, it is unclear what the function of DUF1236 is or how it contributes to the role of RL3684 in plant colonisation. Like RL2946, RL3684 is also significantly upregulated in pea bacteroids, as well as under saline stress and nitrogen limitation [219]. This suggests that as with the other two genes tested it plays a role in survival in the rhizosphere under stress conditions, colonisation of plant roots and progression of symbiosis. It is also somewhat upregulated in the rhizospheres of sugar beet and alfalfa [205]. Together with its upregulation on both pea and barley roots, this suggests that it plays a general role in Rlv3841 colonisation of plant rhizospheres and root surfaces, rather than being specific to interaction with any one plant or group of plants.



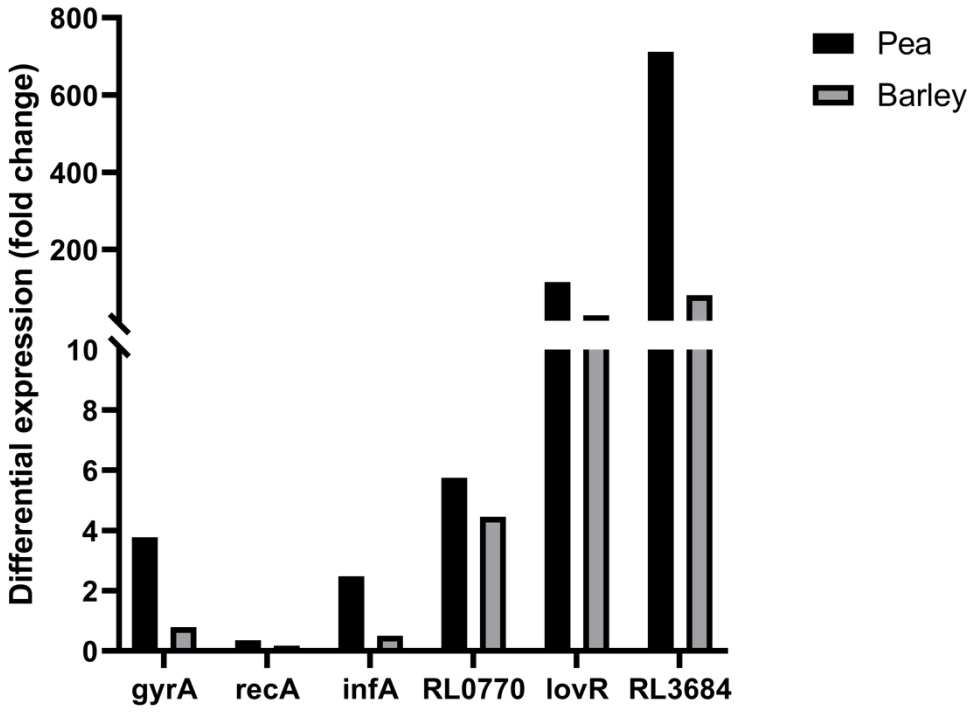


Figure 3.4: Differential expression of example genes during colonisation of pea and barley. *Top, expression measured using RT-qPCR. Bottom, expression measured using RNA sequencing. n=1.*

3.2.4 Transcriptomic changes in Rlv3841 common to colonisation of pea and barley

3.2.4.1 Genes encoding transcriptional regulators are upregulated during colonisation of both pea and barley

A number of genes encoding transcriptional regulators that activate initiation of transcription (GO:0006352) were upregulated during colonisation of both the pea and barley rhizoplanes. The extracytoplasmic sigma factors (ECF) RL3020 (*ecfR*) and RL3703 (*ecfZ*) were significantly upregulated on both. RL3020 was 7.9-fold upregulated on pea and 12.6-fold upregulated on barley, and its adjacent anti-sigma factor RL3021 (*asfR*) was 3.4-fold upregulated on pea and 6.2-fold upregulated on barley. RL3703 was 33.6-fold upregulated during colonisation of pea and 8.5-fold upregulated during colonisation of barley, and its anti-sigma factor RL3704 (*asfZ*) was 32.7-fold upregulated on pea and 10.9-fold upregulated on barley. The anti-anti-sigma factor of *ecfZ*, RL3705 (*phyR*), was 36.0-fold upregulated on pea and 13.4-fold upregulated on barley. *phyR* is part of a two-component sensor-

regulator system and the histidine kinase component, RL3706, was also significantly upregulated during colonisation of both plants. Interestingly, *ecfR* was more substantially upregulated during colonisation of barley, while *ecfZ* was more substantially upregulated during colonisation of pea. Each of these ECF sigma factors has a distinct regulatory network. The *S. meliloti* homologue of *ecfR*, SMb20531 (*rpoE7*) controls transcription of the operon SMb20527-SMb20530, a DoxX domain oxidoreductase, a DUF2063 (DNA-binding domain) protein, a DUF692 (methanobactin synthesis domain) protein and DUF2282 (integral membrane domain) protein [223], [224], [225], [226]. A similar operon in *Legionella pneumophila* encoding DoxX, DUF2063, DUF692 and DUF2282 proteins is involved in response to copper toxicity [224]. This suggests that *rpoE7* may be involved in coordinating metal ion toxicity response. The *Caulobacter crescentus* homologue of *ecfR*, *sigF*, controls response to oxidative stress during stationary phase growth [227]. The *ecfZ* homologue in *S. meliloti*, SMc01506 (*rpoE2*), is induced in response to heat and saline stresses and in response to carbon and nitrogen starvation during stationary phase growth [228]. The homologue in *R. etli*, CH03273 (*rpoE4*), is induced in response to oxidative, saline and osmotic stresses and during stationary phase growth [229]. *ecfR* and *ecfZ* are therefore likely to be responding to overlapping but different sets of conditions and eliciting different responses, and their distinct upregulation patterns during colonisation of the pea and barley rhizoplanes suggests that Rlv3841 comes under unique stresses in different plant microbiomes. The ECF sigma factor RL3235 (*ecfE*) was also 6.3-fold upregulated during colonisation of barley and 3.8-fold upregulated during colonisation of pea. *ecfE* has been shown to be substantially more strongly upregulated in the pea rhizosphere than in the sugar beet and alfalfa rhizospheres [205], leading to the hypothesis that it is a pea-specific regulator. However, my findings would suggest that this may not be the case and that the role of *ecfE* in adaptation to plant rhizospheres and rhizoplanes merits further investigation.

Several two-component sensor/regulators were also upregulated during colonisation of both plant rhizoplanes. In addition to the anti-anti-sigma factor and transcriptional regulator RL3705 and its histidine kinase component RL3706, RL0448 and RL1162-RL1163 were significantly upregulated on

both the pea and barley rhizoplanes. RL0448 encodes the histidine kinase component of a two-component sensor/regulator and was 10.2-fold upregulated during colonisation of pea and 8.2-fold upregulated during colonisation of barley. RL1162 and RL1163 encode the transcriptional regulator and histidine kinase components of another two-component sensor/regulator. RL1162 was 47.8-fold upregulated during colonisation of pea and 13.6-fold upregulated during colonisation of barley, while RL1163 was 34.9-fold upregulated during colonisation of pea and 2.7-fold upregulated during colonisation of barley. While RL1162-RL1163 were upregulated during colonisation of both plants it is possible that they play a greater role in colonisation of the pea rhizoplane, as they were much more strongly upregulated in Rlv3841 on pea roots. Both RL0448 and RL1163 contain PAS domains, a signalling domain that has been implicated in sensing of oxygen, redox potential and light [230], [231]. A range of other rhizobial regulator proteins that play a role in symbiosis with plants have PAS domains, including FixL which is involved in regulation of nitrogen fixation in response to oxygen conditions [232], and pRL110320 which regulates attachment and is expressed in response to blue and UVA light [180].

3.2.4.2 Transketolases are upregulated during colonisation of both pea and barley

Transketolases with transferase activity (GO:0016740) were significantly upregulated during colonisation of both pea and barley rhizoplanes. RL2718-RL2719 encode the alpha and beta subunits of a transketolase predicted to be involved in terpenoid biosynthesis. RL2718 was 35.7-fold upregulated during colonisation of pea roots, and 8.0-fold upregulated during colonisation of barley. RL2719 was 24.6-fold upregulated during growth on pea and 6.3-fold upregulated during growth on barley. Transketolases are involved in the nonoxidative pentose phosphate pathway, which interconverts pentose and hexose sugars and is important in cycling carbon to polysaccharides and in the metabolism of erythritol, ribose and xylulose. Transketolases are also important for the establishment of nitrogen-fixing symbioses with legume plants. A transketolase mutant in *S. meliloti* was unable to establish a wild type symbiosis with its host plant, alfalfa, showing reduced and slower

nodule formation and reduced ability to fix nitrogen [233]. The other key enzyme in the nonoxidative pentose phosphate pathway is transaldolase, encoded by RL4203 (*talB*) in Rlv3841. RL4203 was 2.3-fold upregulated during colonisation of pea roots, but 2.2-fold downregulated during colonisation of barley. The Rlv3841 genome encodes two other transketolases, RL4006 (*cbbT*) and pRL100453. RL4006 was 4.8-fold downregulated during colonisation of barley and no significant result for differential expression was obtained during growth on pea. pRL100453 was 2.7-fold downregulated during growth on barley and 2.3-fold upregulated during growth on pea. This suggests that any role of the transketolase encoded by RL2718-RL2719 in the nonoxidative pentose phosphate pathway is not significant for colonisation at least of barley roots, as all the other components are downregulated. It may be the case that the transketolase encoded by RL2718-RL2719 has a different role that is significant during colonisation of plant roots which explains the upregulation of these genes but not the rest of the nonoxidative pentose phosphate pathway. The adjacent genes, RL2720-RL2722, encode components of an ABC transporter that is induced on arabinogalactan [205], suggesting that it is involved in arabinogalactan import. They were 32.1-fold, 42.9-fold and 48.8-fold upregulated during growth on pea and 7.9-fold, 7.4-fold and 8.0-fold upregulated during growth on barley. RL2718-RL2719 may be involved in arabinogalactan degradation, but are not induced on arabinogalactan themselves, suggesting that they may only respond to components produced by the early part of the degradation pathway [205]. Arabinogalactan proteins are found both inside and on the surfaces of plant cells, and cell-surface arabinogalactan proteins in particular have been shown to play a role in barley root hair development [234]. Upregulation of this one transketolase specifically during Rlv3841 colonisation of pea and barley roots may therefore be a response to root surface components rather than an upregulation of the whole nonoxidative pentose phosphate pathway.

3.2.4.3 Carbohydrate uptake and catabolism genes are upregulated during colonisation of both pea and barley

In addition to the carbohydrate uptake transporter-2 (CUT2) ABC transporter encoded by RL2720-RL2722, which is thought to be an arabinogalactan transporter, expression of other carbohydrate ABC transporters was upregulated during colonisation of both plant rhizoplanes. The CUT2 transporter encoded by RL4172-RL4174 was upregulated on both pea and barley rhizoplanes, with RL4172 being 4.9-fold upregulated on pea and 5.1-fold upregulated on barley, RL4173 being 3.1-fold upregulated on pea and 5.3-fold upregulated on barley and RL4174 being 25.4-fold upregulated on pea and 25.2-fold upregulated on barley. The components of the CUT2 transporter encoded by RL4655-RL4653 (*intABC*), which transports inositol were 13.0-fold, 8.8-fold and 4.0-fold upregulated on pea and 9.8-fold, 5.8-fold and 2.9-fold upregulated on barley. Other components involved in the metabolism of inositol were also upregulated during rhizoplane colonisation. RL0776 (*iolA*) was 6.3-fold upregulated during colonisation of barley and 11.6-fold upregulated during colonisation of pea. RL1496 (*iolR*) and RL1495-RL1492 (*iolCBED*) were not significantly upregulated on barley but were all over 4-fold upregulated during colonisation of pea. *intB*, *iolA* and *iolD* have all been shown to be required for catabolism of *myo*-inositol [235]. The inositol catabolism pathway, comprising *iolA* and *iolCDEB*, is thought to be induced by an end product of catabolism rather than by inositol itself [235], while the transport system is thought to be induced by inositol, with the *S. meliloti* homologue of RL4655, SMb20712, being induced by inositol [236]. Upregulation of the same carbohydrate uptake genes during colonisation of both pea and barley suggests overlap in the composition of their root exudates. Arabinogalactans are present on plant cell surfaces, including root surfaces, and are known to play a role in interactions with microorganisms [237]. Inositol is present in high concentrations in pea root exudates prior to nodulation [238], and is present in barley root exudates [239], meaning that it could be an important carbon source for Rlv3841 during colonisation of both rhizoplanes. Inositol may be provided in smaller quantities by barley than by pea, meaning that the end products of its degradation are not able to build up to the concentrations required for upregulation of the catabolism pathway. However, to my

knowledge there is no quantitative data showing the relative abundance of inositol in pea and barley root exudates, meaning that this hypothesis cannot yet be validated. Catabolic genes are also often less induced than their corresponding transporters, needing a higher concentration of substrate to promote their induction [205].

Other genes involved in catabolism of carbohydrates were induced during colonisation of the pea and barley rhizoplanes. The sugar alcohol dehydrogenase RL0634 was 35.8-fold upregulated during colonisation of pea roots and 11.6-fold upregulated during colonisation of barley roots. RL0634 is also induced during colonisation of the pea and sugar beet rhizospheres [205] and a mutant showed reduced growth in the pea rhizosphere [240]. RL0634 may be involved in metabolism of a compound that is ubiquitous in root exudates. On the other hand, RL0507 (*manB*) encodes a beta-mannosidase and was 11.7-fold upregulated during colonisation of pea but shows no differential expression during colonisation of barley. Beta-mannosidases cleave mannose units from the ends of mannans and other polysaccharides, and the upregulation of *manB* during colonisation of pea but not barley roots suggests a difference in the type of carbon source that is being provided as part of the root exudates. Overall, different changes in expression of transketolases and other carbohydrate-degrading enzymes such as beta-mannosidase alongside changes in expression of carbohydrate ABC transporters suggests that Rlv3841 is exposed to different carbon sources during colonisation of pea and barley, with some overlap between the two, and remodels its transportome and metabolic pathways accordingly. Interestingly, while *manB* and the inositol catabolism pathway are upregulated during colonisation of the pea rhizoplane but not during colonisation of barley, there are no carbohydrate degradation pathways that are significantly upregulated during colonisation of barley but not pea, suggesting that Rlv3841 is well-adapted to growth on the pea rhizoplane but not to the barley rhizoplane.

3.2.4.4 Cell surface characteristics are affected during colonisation of both pea and barley

Factors determining cell surface characteristics are affected during colonisation of both pea and barley rhizoplanes by Rlv3841. In particular, expression of L,D-transpeptidases was downregulated during

colonisation of both plants, shown in Table 3.1. Of the 16 L,D-transpeptidases encoded by Rlv3841, six (38%) were downregulated during colonisation of pea, barley, or both.

Table 3.1: L,D-transpeptidase genes that are downregulated during colonisation of pea and/or barley.

Gene	Differential expression on pea (fold change)	Differential expression on barley (fold change)
RL0055 (*ldt1*)	-1.2 (n.s.)	-4.5
RL0869 (*ldt2*)	-3.8	-6.3
RL0870 (*ldt3*)	-4.1 (n.s.)	-8.6
RL1296 (*ldt4*)	-6.1	-14.0
RL1297 (*ldt5*)	-2.0 (n.s.)	-11.4
RL1458 (*ldt6*)	-6.6	-12.3

L,D-transpeptidases localise to outer membrane vesicles in *R. etli* [241] and are involved in the formation of peptidoglycan cross-links in bacterial cell walls, a function that is important in antibiotic resistance [242]. The L,D-transpeptidase of *Liberibacter asiaticus* has also been shown to have a role in modification of lipopolysaccharide (LPS) lipid A [243]. L,D-transpeptidases are therefore clearly important in determining cell wall and cell surface structure, as well as potentially affecting biofilm formation as LPS is an important component of bacterial biofilms [244]. The BA14K-like immunoreactive protein RL0963 was significantly upregulated during colonisation of both pea and barley roots, being 39.9-fold upregulated during colonisation of pea and 12.0-fold upregulated during colonisation of barley. BA14K is a *Brucella abortus* protein that elicits an immune response in mice [245] and binds carbohydrates, particularly mannose [246]. The BA14K mutant produces altered LPS, suggesting a role for the protein in LPS synthesis or structure [246]. Altered expression of genes

encoding cell surface components suggests that Rlv3841 undergoes some form of cell surface remodelling in response to perception of and interaction with a plant cell surface to aid colonisation.

i3.2.4.5 Expression of genes encoding tRNAs is substantially changed during colonisation of pea and barley

When comparing the transcriptome of Rlv3841 cells during growth on a plant rhizoplane with the transcriptome of an exponential phase laboratory culture, it might be expected that expression of tRNAs would not be substantially changed or, if anything, that tRNA expression would be reduced during growth on the rhizoplane as translation might be expected to be happening most quickly in an exponential culture in a laboratory setting. However, during colonisation of the pea rhizoplane, no tRNA-encoding genes were significantly downregulated compared with laboratory culture and several were upregulated. During colonisation of barley, there was a mix of significant upregulation and downregulation of tRNA-encoding genes. This could be reflecting an increase in growth during growth on the pea rhizoplane compared with in laboratory culture and on the barley rhizoplane enabling increased translation. This would suggest that Rlv3841 is highly specifically adapted to the pea rhizoplane, being able to make better use of the carbon sources and other nutrients that are supplied in pea root exudates than those in minimal laboratory culture or barley root exudates. The Rlv3841 genome encodes 52 tRNAs, nine (17%) of which are upregulated during colonisation of pea. Four tRNAs (8%) are upregulated during colonisation of barley, and eight (15%) are downregulated (Table 3.2).

Table 3.2: Differential expression of tRNA-encoding genes in Rlv3841 during colonisation of pea and barley rhizoplanes. *Values in green were significantly upregulated and values in red significantly downregulated.*

Gene	Anticodon	Amino acid	Fold change on pea	Fold change on barley

RLt10	TGC		+3.1	-60.7
RLt11	GGC		+1.6	-3.3
		Ala		
RLt12	TGC		+3.1	-60.7
RLt13	TGC		+3.1	-60.7
RLt14	ACG		+4.6	-2.6
RLt15	TCT		-2.4	-6.6
		Arg		
RLt16	CCG		+1.2	-2.5
RLt17	CCT		+1.2	-4.4
RLt18	GTT	Asn	+5.7	-1.8
RLt19	GTC		+2.6	-1.2
		Asp		
RLt20	GTC		+2.1	-1.7
RLt21	CGA	Cys	+1.2	-1.2
RLt22	TTG		+1.6	-2.4
		Gln		
RLt23	CTG		+1.5	-3.3
RLt24	TTC		+1.6	-2.0
RLt25	CTC	Glu	+2.2	-1.5
RLt26	TTC		+1.9	-2.1
RLt27	CCC		-1.4	-3.8
RLt28	TCC	Gly	+2.8	-2.7
RLt29	GCC		+1.9	-1.7
RLt30	GTG	His	+8.7	+6.4

RLt31	GAT		+2.9	-63.7
RLt32	GAT	Ile	+2.9	-65.9
RLt33	GAT		+2.9	-65.9
RLt34	TAA		+3.7	-1.0
RLt35	TAG		+1.8	-1.6
RLt36	GAG	Leu	+2.1	-1.8
RLt37	CAA		+10.1	-2.7
RLt38	CAG		+1.3	-2.4
RLt39	TTT		+3.6	-3.1
RLt40	CTT	Lys	+1.6	-3.7
RLt41	CAT		+6.5	+9.7
RLt42	CAT		-1.0	-2.5
RLt43	CAT	Met	+3.9	+4.3
RLt44	CAT		+1.7	-2.2
RLt45	CAT		+6.5	+9.7
RLt46	GAA	Phe	+2.6	-1.9
RLt47	CGG		+1.3	-3.9
RLt48	GGG	Pro	+11.5	+1.1
RLt49	TGG		+5.3	-1.2
RLt50	GCT		+2.5	-1.6
RLt51	TGA	Ser	+2.1	-1.3

RLt52	GGA		-1.4	-3.8
RLt53	CGA		-1.4	-2.5
RLt54	CGA		+1.3	-2.4
RLt55	TGT		+5.0	+1.0
RLt56	CGT	Thr	+3.8	-1.3
RLt57	GGT		+3.3	-1.0
RLt58	CCA	Trp	+1.9	-1.9
RLt59	GTA	Tyr	+3.2	-2.1
RLt60	GAC		-1.2	-2.5
RLt61	TAC	Val	+1.0	-3.5

Expression of genes encoding histidine and methionine tRNAs were upregulated during colonisation of both pea and barley rhizoplanes. RLt30, the only histidine-encoding tRNA, was 8.7-fold upregulated on pea roots and 6.4-fold upregulated on barley roots. RLt41 and RLt45 (methionine) were both 6.5-fold upregulated on pea and 9.7-fold upregulated on barley. RLt43 (methionine) was 4.3-fold upregulated during colonisation of barley and 3.9-fold upregulated on pea. Expression of genes encoding asparagine, leucine, proline and threonine tRNAs were significantly upregulated during growth on pea roots, but not barley. RLt18, the only asparagine tRNA in the Rlv3841 genome, was 5.7-fold upregulated during colonisation of pea roots. RLt36 (leucine) was 10.1-fold upregulated during colonisation of pea, and RLt34 and RLt36 (leucine) were 3.7-fold and 2.1-fold upregulated respectively. RLt48 and RLt49 (proline) were 11.5-fold and 5.3-fold upregulated respectively during growth on pea roots. RLt55 (threonine) was 5.0-fold upregulated during colonisation of pea roots, and RLt56 and RLt57 3.8-fold and 3.3-fold upregulated, respectively. Genes encoding alanine, arginine and isoleucine

tRNAs were significantly downregulated during colonisation of barley roots. Three of the four alanine tRNAs, RLt10, RLt12 and RLt13, were 60.7-fold downregulated. Two of the four arginine tRNA genes were downregulated. RLt15 was 6.6-fold downregulated and RLt17 4.4-fold downregulated. All three isoleucine tRNA genes were significantly downregulated. RLt31 was 63.7-fold downregulated and RLt32 and RLt33 65.9-fold downregulated. GSEA revealed that the pathways for synthesis and ligation of arginine, glycine and glutamine tRNAs were significantly downregulated during colonisation of the barley rhizoplane. The arginyl-tRNA synthetase gene RL2041 (*argS*) was 4.2-fold downregulated and RL1058 (*argS2*) was 3.1-fold downregulated. The glycyl-tRNA synthetase genes RL0943 (*glyQ*) and RL0951 (*glyS*) were 5.6-fold and 8.9-fold downregulated respectively. The glutamyl-tRNA amidotransferase genes RL2076 (*gatA*), RL2082 (*gatB*) and RL2075 (*gatC*) were 4.9-fold, 7.5-fold and 4.9-fold downregulated. These changes in expression of both the tRNA genes themselves and genes involved in tRNA processing could be indicative of changes in translation during colonisation of plant rhizoplanes due to the different nutrient sources provided in different root exudates. Changes in synthesis of specific tRNAs could also be due to an alternative role for the tRNA molecules outside translation. tRNA-derived small RNA fragments (tRFs) are a class of non-coding RNAs (ncRNAs) that have been implicated in inter-kingdom signalling during both pathogenic and symbiotic interactions [247], [248], [249], [250]. Human tRFs are released by keratinocytes in response to infection by the Gram-negative bacterial pathogen *Fusobacterium nucleatum* and have been shown to inhibit its growth. However, these tRFs do not inhibit growth of the Gram-positive pathogenic bacterium *Streptococcus mitis*, suggesting that they are part of a pathogen-specific immune response in humans [247]. Production of tRFs is also induced in wheat during infection by *Fusarium graminearum* [249]. It is not only the host organism that produces tRFs during inter-kingdom interactions. During infection of barley, the fungal pathogen *Blumeria hordei* shows enrichment of tRFs and these tRFs are found in infected host epidermis tissue, showing that they are transmitted from the pathogen to the host and suggesting a role in signalling and virulence [250]. *Bradyrhizobium japonicum* has also been shown to use tRFs to interfere with soybean RNA silencing machinery to affect expression of host genes and

promote nodulation [248]. tRFs have been implicated in a range of plant functions, including regulation of cell processes like transcription and translation, response to abiotic and biotic stresses and protection of the genome from transposable elements [251]. It may therefore be that infecting pathogens or symbionts are capable of modulating these existing plant signalling pathways via the use of tRFs to promote plant responses that best allow them to colonise. This could explain the upregulation of synthesis of some tRNAs during Rlv3841 colonisation of pea and barley roots, as well as why expression of tRNAs differs between colonisation of a host legume and a non-host plant.

3.2.4.6 Motility and chemotaxis genes are downregulated during colonisation of both pea and barley

Genes involved in motility (GO:0071973, GO:0097588) and chemotaxis (GO:0006935) were significantly downregulated during colonisation of both plants. Genes encoding flagellum components RL0695 (*fliF*), RL0700-RL0701 (*fliGN*), RL0704 (*flgF*), RL0707-RL0710 (*flgBC*, *fliE*, *flgG*), RL0712 (*flgI*), RL0714 (*flgH*), RL0716 (*fliP*), RL0718-RL0721 (*flaABCD*), RL0724 (*motC*) and RL0728 (*flgE1*) were all over 4-fold downregulated during colonisation of both plants. In addition, RL3268 (*flaH*) and RL4729 (*flaG*) were 6.5-fold and 4.3-fold downregulated during colonisation of barley. Methyl-accepting chemotaxis proteins RL0972 (*mcpK*), RL2683 (*mcpD*) and RL2931 (*mcpN*) were over 4-fold downregulated during colonisation of both plants. RL1065 (*mcpL*) and RL1318 (*mcpM*) were 4.3-fold and 4.2-fold downregulated during colonisation of barley and 2.2-fold and 2.7-fold downregulated during colonisation of pea. Methyl-accepting chemotaxis proteins detect attractants and repellents and direct bacterial movement by signalling to flagellar components via chemotaxis regulatory proteins [252], [253]. Chemotaxis proteins RL0687-RL0689 (*cheY1A1W1*), RL0691-RL0693 (*cheB1Y2D1*), RL4029-RL4035 (*cheR2W2b*, *mcrABC*, *cheW2aA2*) and RL4386 (*cheW4*) were all also significantly downregulated during colonisation of both plants. All key steps in sensing of attractants and chemotaxis and motility were therefore downregulated. Methyl-accepting chemotaxis protein genes have been shown to be downregulated in pea bacteroids [252], as are other genes involved in motility

and chemotaxis, including chemotaxis protein encoding genes and flagellum component encoding genes [219], showing that downregulation of these functions is a feature of the Rlv3841 transcriptome throughout symbiosis. Moreover, in *S. meliloti*, expression of chemotaxis and motility genes including *flaF*, *flbT*, *cheY1*, *flaC*, *flgB*, *flgF* and *motA* is repressed at high population density by the quorum sensing system ExpR/Sin, which also plays a role in EPS and biofilm formation [152], [254]. This suggests that the downregulation of chemotaxis genes and flagellum synthesis is a necessary part of biofilm formation, which is central to colonisation of plant rhizoplanes, and explains the consistent downregulation seen during growth on the roots of both plants tested.

3.2.4.7 Genes involved in cell division are downregulated during colonisation of both pea and barley

FtsZ-dependent cytokinesis (GO:0043093) was significantly downregulated in Rlv3841 during colonisation of both pea and barley rhizoplanes. RL3298-RL3300 (*ftsZ1AQ*) were 15.8-fold, 8.8-fold and 9.7-fold downregulated during colonisation of barley. No significant result was obtained for expression of *ftsZ1* during colonisation of pea, but *ftsA* was 4.4-fold downregulated and *ftsQ* 5.3-fold downregulated. RL4330 (*ftsE*) was 6.5-fold downregulated during colonisation of pea and 22.2-fold downregulated during colonisation of barley. RL4331 (*ftsX*) was 4.8-fold downregulated during colonisation of pea and 8.5-fold downregulated during colonisation of barley. FtsZ is a tubulin-family protein involved in assembly of the Z-ring, a cytoskeletal structure that forms at the site of bacterial cytokinesis. *In vivo* it is recruited by the actin-related protein FtsA and works in conjunction with other cell division proteins to assemble the Z-ring, but purified FtsZ can also assemble artificial Z-rings in the absence of any other cell division proteins, making it the key cell division protein in bacteria [255]. Expression of the cell cycle regulator RL3758 (*ctrA*) was also 2.6-fold downregulated in Rlv3841 during colonisation of pea and 5.7-fold downregulated during colonisation of barley. CtrA controls DNA methylation, cell wall remodelling and cell division, and directly regulates expression of *ftsZ* in *Caulobacter crescentus* [256]. Interestingly, the Rlv3841 genome encodes two copies of *ftsZ*, RL3298

(*ftsZ*) and RL3071 (*ftsZ2*). Expression of *ftsZ2* was 3.3-fold upregulated during colonisation of pea and showed no change in expression during colonisation of barley compared with growth in laboratory culture. *S. meliloti* also encodes two *ftsZ* genes, which differ from each other primarily in that *ftsZ2* lacks a region of approximately 200 bp that is present in the C-terminal region of *ftsZ1* and a second C-terminal region that is conserved in the *Escherichia coli* and *B. subtilis ftsZ* sequences [257], [258]. Clustal alignment shows that RL3298 and SMc01874 (*ftsZ1*) share 81.13% protein sequence identity and RL3701 and SMc04296 (*ftsZ2*) share 78.47% protein sequence identity. The two Rlv3841 *ftsZ* gene products share 70.54% protein sequence identity and, as with *S. meliloti*, the main difference is that RL3071 (*ftsZ2*) lacks the *ftsZ1* C-terminal region. *S. meliloti* FtsZ2 is not essential, as an *ftsZ2* mutant is still viable. Its function is predicted to differ somewhat from FtsZ1, as overexpression of *ftsZ2* in *E. coli* results in a coiled phenotype not caused by overexpression of *ftsZ1* [258]. This is also supported by my finding that *ftsZ2* is expressed under different conditions than *ftsZ1* and the rest of the FtsZ-associated cell division machinery, suggesting that it is not regulated by CtrA and it plays a different role in cytoskeletal assembly or cell cycle, which may be specific to colonisation of roots.

3.2.5 Transcriptomic changes unique to colonisation of the barley rhizoplane

3.2.5.1 Certain genes involved in carbohydrate metabolism were significantly upregulated during colonisation of barley

Genes involved in D-galactonate catabolism (GO:0034194) were significantly upregulated during colonisation of the barley rhizoplane. RL3864-RL3866 encode muconate lactonising family proteins and were 4.7-fold, 5.6-fold and 4.3-fold upregulated during growth on barley roots. Muconate lactonising family proteins are involved in catabolism of aromatic acids, suggesting that these genes are upregulated in response to compounds that are provided in barley but not pea root exudates. Genes involved in other carbohydrate metabolic processes (GO:0005975) were also upregulated during colonisation of barley. RL0065, encoding a glycosyl transferase, was 4.4-fold upregulated,

RL0066, encoding xylulose-5-phosphate was 5.6-fold upregulated, and RL2717 (*glpK3*), encoding glycerol kinase was 6.1-fold upregulated.

3.2.5.2 Genes involved in response to oxidative stress were upregulated during colonisation of barley

Genes that are implicated in cellular responses to oxidative stress (GO:0006979) were upregulated during Rlv3841 colonisation of the barley rhizoplane. RL2024 (*katE*) encodes a catalase and was 5.1-fold upregulated during colonisation of barley. RL4265 (*msrB2*) encodes a peptide methionine sulfoxide reductase and was 4.8-fold upregulated during colonisation of barley roots. RL4269 encodes a ThiJ/PfpI family protein and was 6.1-fold upregulated. The function of ThiJ/PfpI family proteins in bacteria is not understood, but in eukaryotes these proteins are implicated in stress response, particularly redox stress [259]. This suggests that Rlv3841 is experiencing an increased amount of stress during colonisation of the barley rhizoplane than during colonisation of its host plant, pea.

3.2.5.3 Cell division processes are downregulated during colonisation of the barley rhizoplane

A wide range of genes involved in cell division processes (GO:0051301) were downregulated during Rlv3841 colonisation of the barley rhizoplane. Although *ftsZ1* and associated septum-formation genes were downregulated during colonisation of both plants, the group of cell division-related genes that were significantly downregulated during colonisation of barley roots was substantially broader than those downregulated during colonisation of pea. In particular, genes involved in chromosome segregation (GO:0007059) were significantly downregulated. DNA translocase proteins RL0161 and RL4563 were 21.3-fold and 8.5-fold downregulated. RL1609 (*parC*), a DNA topoisomerase IV subunit, was 5.2-fold downregulated. Chromosome segregation and condensation proteins RL2044 (*scpA*) and RL2045 (*scpB*) were 9.1-fold and 4.9-fold downregulated respectively. Chromosome partitioning proteins RL3343 (*parA2*) and RL4736 (*parA1*) were 4.5-fold and 4.1-fold downregulated. Finally, the tyrosine recombinase RL4354 (*xerD*) was 7.6-fold downregulated.

3.2.5.4 Genes involved in DNA repair are downregulated during colonisation of barley

Genes involved in DNA catabolic processes (GO:0006308) were significantly downregulated during colonisation of the barley rhizoplane. RL0282 (*xseA*), the large subunit of exodeoxyribonuclease VII, was 6.3-fold downregulated and RL0976 (*xseB*), the small subunit, was 6.6-fold downregulated. Exodeoxyribonuclease VII degrades single-stranded DNA into large oligonucleotides [260]. It is also involved in DNA mismatch repair, and in particular in repair of damage caused by exogenous factors including UV radiation, antibiotic compounds and oxidative stress [261], [262], [263], [264]. Downregulation of this DNA repair mechanism may be linked to the downregulation of cell division and DNA replication, particularly as other genes involved in response to oxidative stress were upregulated during colonisation of barley.

3.2.5.5 Genes involved in synthesis of cell wall and cell surface characteristics are downregulated during colonisation of the barley rhizoplane

Expression of genes involved in the synthesis of cell wall components that are needed for cell expansion and division was also downregulated during colonisation of the barley rhizoplane. In addition to the L,D-transpeptidases that were downregulated during colonisation of both pea and barley, genes involved in biosynthesis of the cell wall component peptidoglycan (GO:0009252) and biofilm matrix components extracellular polysaccharide (EPS) (GO:0045226), cellulose (GO:0030244) and lipopolysaccharide (LPS) (GO:0009103) were also significantly downregulated. D-alanine-D-alanine ligases involved in peptidoglycan synthesis, RL1016, RL2541, RL2801 (*ddlA*) and RL3301 (*ddlB*), were all over 4-fold downregulated. EPS synthesis genes RL1623-RL1626 (*rmlA*, *rfbD*, *rmlBC*) were 5.6-fold, 5.3-fold, 3.7-fold and 6.0-fold downregulated, and RL3662 (*pssP*) was 4.9-fold downregulated, with RL3663-RL3665 (*pssONT*) being all being over twofold downregulated also. Cellulose synthase RL1646 (*celA*) was 4.3-fold downregulated, with RL1647-RL1648 (*celBC*) being over twofold downregulated. GGDEF sensory box protein genes RL1577, RL1687 and RL2657 were 4.2-fold, 8.8-fold and 5.5-fold downregulated. LPS biosynthesis genes RL0140 (*kdsB*), RL0684, RL1469-RL1471, RL1570,

RL1664 (*exoP*), RL3646-RL3649 (*pssLKJI*) were all over 4-fold downregulated. Genes involved in dUTP catabolism (GO:0046081) and dUMP synthesis (GO:0006226) were significantly downregulated. RL2627 (*murI*) was 7.9-fold downregulated. RL3303 (*galE*) was 3.8-fold dowregulated and RL3305-RL3312 (RL3305, *murCG*, *ftsW*, *murD*, *mraY*, *murFE*) were all over 4-fold downregulated. Genes involved in wider pyrimidine metabolism, including RL0331 (*pyrF*), orotidine 5'-phosphate decarboxylase, RL0572, dihydroorotate dehydrogenase, and RL1739 (*pyrB*), aspartate carbamoyltransferase, were all significantly downregulated during barley colonisation. Changes in expression of genes involved in pyrimidine metabolism and the UMP biosynthesis pathway have been linked to modulation of biofilm formation in response to cell starvation in *E. coli* [265]. RL2975, the ATP binding component of an Export ABC transporter involved in secretion of polysaccharides to the extracellular matrix, was 15.7-fold downregulated. The other two components of the ABC transporter, RL2976-RL2977, were 3.4-fold and 2.6-fold downregulated, respectively. This ABC transporter is required for biofilm formation and desiccation tolerance [266] and, interestingly, was upregulated during colonisation of the pea rhizoplane. This suggests either that a different set of genes is required for biofilm formation on the barley root or that Rlv3841 struggles to form biofilms on the barley rhizoplane as it is not as well adapted to barley as to its host plants.

Genes involved in diaminopimelate biosynthesis (GO:0019877) were also significantly downregulated during colonisation of the barley rhizoplane. The dihydropicolinate reductase gene RL0180 (*dapB1*) was 4.3-fold downregulated, the 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase gene RL0437 (*dapD*) was 5.4-fold downregulated, the dihydropicolinate synthase gene RL1502 (*mosA*) was 5.6-fold downregulated, the dihydrodipicolinate synthase gene RL3002 (*dapB2*) was 4.0-fold downregulated, the diaminopimelate decarboxylase gene RL4325 (*lysA*) was 4.5-fold downregulated, and the dihydrodipicolinate synthase gene RL4423 was 4.1-fold downregulated. Diaminopimelate is a precursor of lysine and a component of the bacterial cell wall compound peptidoglycan, and is documented as being an important constituent of the ‘rigid layer’ of bacterial cell walls [267], [268].

Cell wall biosynthesis is significantly reduced during colonisation of the barley rhizoplane, which may be the reason for repression of the diaminopimelate synthesis pathway.

3.2.5.6 Central carbon metabolism is downregulated during colonisation of the barley rhizoplane

Genes encoding TCA cycle enzymes were significantly downregulated during colonisation of the barley rhizoplane. As shown in Figure 3.5, the majority of enzymes involved in the central TCA cycle were over 4-fold downregulated, as well as those involved in glycolysis.

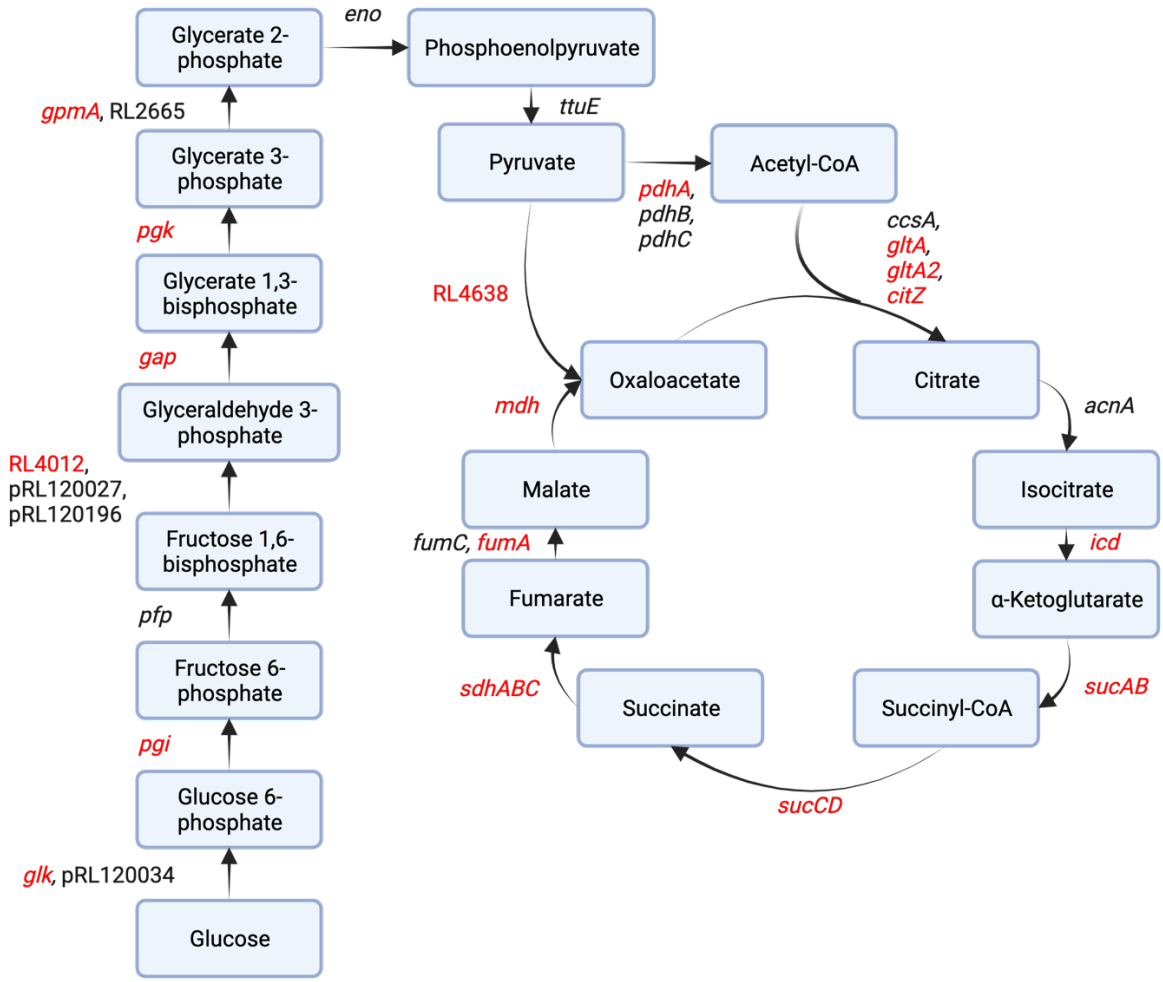


Figure 3.5: TCA cycle and glycolysis genes downregulated in Rlv3841 during colonisation of the barley rhizoplane. *Genes shown in red were downregulated during colonisation of barley; expression of genes shown in black was not significantly altered. Figure made in BioRender.*

Genes involved in ATP synthesis and electron transport were also downregulated, with the ATP synthase subunits RL0925-RL0928 (*atpBEG2F*), RL4405 (*atpC*) and RL4407-RL4409 (*atpDGA*) all being over 4-fold downregulated. NAD^+ synthetase genes RL2688 (*nadE1*) and RL1332 (*nadE2*) were 4.3-fold and 6.5-fold downregulated respectively. Genes with NAD^+ binding function (GO:0070403), including RL1574, RL4135-RL4136 (*pntABAA*) and RL4337 (*tyrC*) were also downregulated compared with laboratory culture. The genes encoding NADH-quinone oxidoreductase, RL1700-RL1714 (*nuoABCDEFGHUIJKLMN*) were all between 4-fold and 15-fold downregulated. Downregulation of these genes correlates with their function in the TCA cycle and central carbon metabolism, and downregulation of all these functions suggests that Rlv3841 is not growing as quickly on the barley rhizoplane as in laboratory culture or on the pea rhizoplane. Glycogen biosynthesis (GO:0005978) was repressed, with RL1662, glycosyltransferase, and RL4115-RL4117 (*glgB1CA*) being significantly downregulated. Polyhydroxybutyrate (PHB) biosynthesis (GO:0042619) was also downregulated, with RL2094 (*phaC*) being 5.5-fold downregulated and RL4621-RL4620 (*phaAB*) being 5.7-fold and 3.8-fold downregulated. The carbon flux regulator RL4622 (*phaR*) was 2.9-fold downregulated. Glycogen and PHB are the two main carbon storage polymers in Rlv3841 [269], so repression of these functions shows that the TCA cycle is not being downregulated in order to maximise flux to a carbon store, but that during colonisation of barley Rlv3841 may lack sufficient carbon to perform any of these processes.

3.2.5.7 Peptide uptake and degradation is downregulated during colonisation of the barley rhizoplane

Expression of genes involved in peptide transport (GO:0015833) was significantly downregulated during Rlv3841 colonisation of the barley rhizoplane. RL0226-RL0228, the components of a peptide (PepT) ABC transporter were 4.6-fold, 3.9-fold and 5.1-fold downregulated. RL0778 (*dppA*) and RL0779 (*dppB*) encode the solute-binding protein (SBP) and permease components of a PepT transporter and were 6.5-fold and 4.5-fold downregulated. The other components of the transporter, RL0780-RL0782

(*dppCDF*) were also over 2-fold downregulated. The *S. meliloti* homologue of *dppA*, SMc00786, is induced by hydroxyproline and casamino acids [236]. RL3543 (*oppA*) encodes the SBP of another PepT transporter and was 4.5-fold downregulated during colonisation of barley. Its *S. meliloti* homologue, SMb21196, is induced by methionine [236]. The other components of the transporter, RL3544-RL3547 (*oppBCDF*), were all 3-fold or more downregulated. RL3557 (*bacA*), a transmembrane transporter protein required for development of pea bacteroids [270], was 4.7-fold downregulated. Finally, RL4647-RL4651, the components of a PepT transporter, were all over 5-fold downregulated.

In addition to genes involved in peptide uptake, those involved in degradation of proteins and peptides were also repressed during Rlv3841 colonisation of the barley rhizoplane. Peptidase encoding genes including RL0184, RL0286, RL1016, RL1446 (*pepN*), RL2069 (*map1*), RL2541, RL2656 and RL4507 (*dcp*) and protease encoding genes including RL0049 (*hslV*), RL0051 (*hslU*), RL1360 (*ptrB*), RL1806 (*degQ*) and RL1877 were over 4-fold downregulated. GSEA showed that functions including carboxypeptidase activity (GO:0004180), metallopeptidase activity (GO:0008237), endopeptidase activity (GO:0004175) and proteolysis (GO:0006508) were significantly downregulated. Downregulation of this variety of peptide transporters and of peptide and protein catabolism suggests either that barley root exudates do not provide many peptide compounds such as hydroxyproline and methionine, or that the reduced growth and respiration demonstrated by the downregulation of the cell cycle and TCA cycle means that Rlv3841 is unable to allocate energy resources to the uptake of these compounds.

3.2.5.8 Iron uptake and assimilation is downregulated during colonisation of the barley rhizoplane

Genes involved in uptake of iron were downregulated during Rlv3841 colonisation of the barley rhizoplane. RL3696-RL3700 (*hmuPSTUV*) were all between 5-fold and 7-fold downregulated and encode an iron chelate (FeCT) ABC transporter responsible for the uptake of hemin. RL4584-RL4583 (*fbpAB*) were 14.4-fold and 4.8-fold downregulated, and RL4582 (*fbpC*) was 3.7-fold downregulated. FbpABC is an ABC transporter responsible for the uptake of iron (III) ions. pRL120314-pRL120322

(*vbsOSADL*, *rpoI*, *vbsC*, *fhuFA2*) are involved in synthesis (*vbs*) and uptake (*fhu*) of the siderophore vicibactin and were all over 7-fold downregulated. The iron response regulator RL0115 (*irrA*) was 3.1-fold downregulated during colonisation of barley, and its known targets, RL0777 (*rirA*), RL4379 (*hemA1*) and RL2583-RL2576 (*sufS2BCDS1XA*) [271] showed no significant change in expression compared with laboratory culture. Since IrrA represses transcription of its targets under iron limitation, this suggests that iron availability on the barley rhizoplane is higher than that in laboratory culture or on the pea rhizoplane. This would also explain the downregulation of iron uptake genes, as in iron-sufficient conditions overexpression of transporters that accumulate iron from the environment would not be necessary.

3.2.5.9 Biosynthesis of metabolic intermediates is downregulated during colonisation of the barley rhizoplane

Genes involved in chorismate biosynthetic processes (GO:0009423), encoding shikimate pathway enzymes, were significantly downregulated in Rlv3841 during colonisation of the barley rhizoplane. RL0003, which encodes shikimate dehydrogenase, was 5.2-fold downregulated. RL0945 (*aroA2*), RL1007 (*aroC*) and RL4352-RL4353 (*aroBK*) encode 3-phosphoshikimate 1-carboxyvinyltransferase, chorismate synthase, 3-dehydroquinate synthase and shikimate kinase I and were all also over 4-fold downregulated. The shikimate pathway converts D-erythrose 4-phosphate and phosphoenolpyruvate to chorismate via shikimate [272]. Downregulation of chorismate synthesis may therefore be due to a lack of available precursors during colonisation of barley, as glycolysis and the TCA cycle were also downregulated. Chorismate is a precursor for a range of products including the amino acids phenylalanine, tyrosine and tryptophan, and iron-chelating siderophores. Since iron uptake is also downregulated during colonisation of the barley rhizoplane, chorismate biosynthesis may also be downregulated due to reduced requirement for its downstream products.

Genes involved in urea catabolic processes (GO:0043419) were also downregulated during colonisation of the barley rhizoplane. The three subunits of urease, RL3731 (*ureC*), RL3734 (*ureB*) and

RL3735 (*ureA*) were 6.4-fold, 7.2-fold and 6.4-fold downregulated. The urease accessory protein RL3727 (*ureG*) was 10.4-fold downregulated and RL3727 (*ureF*) and RL3736 (*ureD*) were also over 3-fold downregulated. Urea is a byproduct of the ornithine cycle, a part of the pathway for synthesis of amino acids including arginine and ornithine. Ornithine is a precursor of the siderophore vicibactin [273]. Downregulation of urea catabolism therefore suggests, like the downregulation of chorismate biosynthesis, that expression of several wider metabolic pathways is reduced during Rlv3841 colonisation of barley.

3.2.5.10 Glutathione biosynthesis is downregulated during colonisation of the barley rhizoplane

Biosynthesis of glutathione (GO:0006750) was significantly downregulated during colonisation of the barley rhizoplane. RL0855 (*gshA*), RL3458, RL4050 (*ggt1*) and RL4340 (*gloB*) were 7.1-fold, 4.4-fold, 4.4-fold and 9.4-fold downregulated. Glutathione is present in almost all Gram-negative bacteria, where it plays in important role in bacterial oxidative stress response [274], [275], [276]. Glutathione has also been reported to play a role in virulence in pathogenic bacteria, and in bacterial biofilm formation [277]. Glutathione is synthesised from cysteine and glutamate, so the repression of its synthesis may be due to a lack of precursor availability. Moreover, rhizobial glutathione synthesis has been shown to be important in interactions with host legume plants [278], [279], [280], [281], but this may not be the case in interactions with non-host plants such as barley.

3.2.5.11 Trends of Rlv3841 gene expression on the barley rhizoplane

Rlv3841 transcriptome changes that were unique to the barley rhizoplane were mainly downregulation of a variety of metabolic pathways, including parts of central carbon metabolism such as the glycolysis pathway and the TCA cycle, and synthesis of metabolic intermediates including chorismate. Cell division processes, DNA repair and cell wall synthesis were also downregulated. This all suggests that Rlv3841 grows much more slowly on the barley rhizoplane than in laboratory culture, or on the pea rhizoplane, possibly due to a lack of available carbon in barley root exudates. When co-inoculated onto

barley roots as part of a seven-member synthetic community (SynCom), Rlv3841 is unable to colonise, while it is able to become established as the second-most abundant strain on pea roots [Jorrin et al., in preparation]. Rlv3841 may therefore simply not be as well adapted to growth on the barley rhizoplane as it is to colonisation of its host plant, pea.

3.2.6 Barley-specific substantially differentially regulated genes

One difficulty with engineering rhizosphere PGPRs to fix and release nitrogen for the benefit of crop plants is that constitutive nitrogen fixation is energetically extremely costly, which might impair the ability of these beneficial engineered strains to competitively colonise the rhizosphere of their target plants. However, approaches to engineer specificity so far have included engineering the plant to produce a signal that can be taken up and recognised by bacteria, which are then engineered to begin fixing nitrogen in response to it. An example of this is the family of bacterial signalling molecules, rhizopines, that have been engineered into barley [55]. This has the advantage of providing a signal that is produced in measurable quantities by the crop plant, and because it is a signal of bacterial origin it will not be produced by other, non-GM, plants, thus conferring true specificity on the interaction. However, a disadvantage to this approach is that GM plants require significant testing before they can be implemented in agriculture and in some regions, such as the EU, cannot be licensed for agricultural use [282]. Moreover, many staple crop plants, such as wheat, have polyploid genomes, making them particularly difficult to modify [283]. An alternative to this approach, therefore, could be to identify bacterial genes and their promoter regions that are substantially upregulated during colonisation of a specific plant by PGPRs and generate fusions of these promoters to agriculturally beneficial genes such as *nifA* for nitrogen fixation. This approach could still provide a level of specificity and inducible control of beneficial behaviours but without the need to engineer crop plants. With this in mind, I searched the Rlv3841 transcriptome during colonisation of barley for genes that are substantially (defined as over 16-fold) upregulated, but that are not over substantially upregulated during colonisation of pea.

In total, 62 genes were over 16-fold upregulated during Rlv3841 colonisation of barley, and 244 were over 16-fold upregulated with an adjusted p-value < 0.05 during colonisation of pea. Of these genes, 50 overlapped between the two plants, leaving only 12 genes that were defined as more substantially upregulated during colonisation of barley than during colonisation of pea.

Table 3.3: Candidate 'barley-specific' genes in Rlv3841. *Genes that are over 16-fold upregulated during colonisation of barley and below 16-fold upregulated during colonisation of pea.*

Gene	Differential expression on barley (fold change)	Differential expression on pea (fold change)
RL1407	37.2	11.2
RL1408	25.8	7.7
RL2197	21.7	10.8
RL3018	18.4	15.1
RL3234	25.8	12.0
RL3593	20.8	1.4 (n.s.)
RL3594 (*dapA*)	53.4	1.9
RL4223	17.5	14.8
pRL90129	20.5	4.9 (n.s.)
pRL100243	23.0	12.4
pRL110077	18.3	2.5
pRL120099	24.7	2.2

Several of these genes are promising candidates that may have 'barley-specific' promoter regions. In particular, those genes that are substantially upregulated during colonisation of barley and show no significant differential regulation during colonisation of pea. Of the 12 candidates presented in Table 3.3, RL3594 (*dapA*) was the most upregulated during colonisation of barley, being 53.4-fold upregulated, but only 1.9-fold upregulated during colonisation of pea. RL3594 encodes a dihydrodipicolinate synthase, which is involved in lysine and diaminopimelate biosynthesis. Interestingly, the wider diaminopimelate biosynthesis pathway (GO:0019877) is downregulated during colonisation of barley roots, suggesting that *dapA* is involved in synthesis of an intermediate that is either accumulated or redirected to another branch of metabolism during colonisation of barley. In *R. phaseoli*, all three copies of *dapA* are upregulated in response to maize root exudates but not common bean root exudates [284], suggesting that *dapA* expression in *Rhizobium* strains is important for growth on Poaceae but not Fabaceae. This makes its promoter region a promising one for fusion to genes encoding agriculturally beneficial characteristics to generate *Rhizobium* strains that specifically benefit cereal crop plants. RL3593, which encodes a transmembrane transporter protein, was 20.8-fold upregulated and 1.4-fold upregulated on pea (padj > 0.05). This expression pattern suggests that it may be expressed from the same promoter as RL3594. Several other genes show substantial upregulation during colonisation of barley and very little upregulation during colonisation of pea, meaning that they may also warrant further investigation as targets for engineering. In particular, pRL110077 encodes the permease component of a polar amino acid uptake (PAAT) ABC transporter and was 18.3-fold upregulated during colonisation of barley and only 2.5-fold upregulated during colonisation of pea. The other components of the ABC transporter, pRL110078 and pRL110079, were 10.1-fold and 8.8-fold upregulated respectively during colonisation of barley and both were 2.8-fold upregulated during colonisation of pea. It is likely that all three components are expressed together from one promoter and upregulation of all three significantly more on barley than on pea suggests that this promoter region would be a good target for fusion to agriculturally beneficial genes. Finally, pRL120099 encodes the solute-binding protein of a hydrophobic amino acid (HAAT) ABC transporter

and was 24.7-fold upregulated during colonisation of barley and 2.2-fold upregulated during colonisation of pea. There are several other genes in Table 3.3 that were much more upregulated during colonisation of barley than during colonisation of pea, but the higher the upregulation during growth on pea roots the more likely any fusion would be to generate a strain that expresses agriculturally significant and possibly energetically costly genes on the rhizoplanes of a wide range of plants. This would reduce specificity and possibly also competitiveness by increasing expression of genes that are normally only expressed only a narrow range of conditions due to the cost to doing so. However, as discussed above, wild type Rlv3841 does not seem able to thrive on the barley rhizoplane, particularly when in competition with other strains, expression of agriculturally beneficial genes from a cereal-specific promoter will have to be coupled with a means of better adapting rhizobia to growth on cereal roots to create a strain that is useful in the field.

3.2.7 Changes unique to pea

Transcriptomic changes unique to Rlv3841 colonisation of pea will be discussed in more detail in Chapter 4, which will examine the transcriptome during multiple stages of colonisation of pea roots. However, it is important to note here that substantially more genes and pathways were significantly upregulated during colonisation of pea roots than colonisation of barley roots. Pathways that were upregulated during colonisation of pea but not colonisation of barley include catabolism of different carbon sources, including inositol (GO:0019310) and aromatic compounds (GO:0019439), protein and peptide uptake (GO:0015031, GO:0071916) and sulphur metabolic processes (GO:0006790).

3.3 Summary

There is overlap in differential regulation of gene expression in Rlv3841 during colonisation of the pea and barley rhizoplanes. In particular, the majority of genes (95%) that were downregulated during colonisation of pea were also downregulated during colonisation of barley, and the majority of genes (80%) upregulated during colonisation of barley were also upregulated during colonisation of pea.

However, substantially more genes were upregulated during colonisation of pea and substantially more were downregulated during colonisation of barley, creating a distinct transcriptomic signature for Rlv3841 during colonisation of each plant. This suggests that there is a set of genes whose expression is required for the colonisation of plant roots in general, and that colonisation of the host plant, pea, requires increased expression of an additional set of genes that may be involved in adaptation to the host root surface and root exudates, or in symbiosis-specific pathways. Expression of transcriptional regulators was upregulated during colonisation of both plants. In particular, the ECFs RL3020 (*ecfR*) and RL3703 (*ecfZ*) were significantly upregulated on both rhizoplanes, with *ecfR* being more strongly upregulated on barley and *ecfZ* being more strongly upregulated on pea, suggesting that both transcriptional regulators play a role in colonisation of both plants, but that the regulatory network of each is more important during growth on one rhizoplane than the other. [227][228]Other transcriptional regulators, containing PAS domains that have been implicated in the Rhizobium-legume symbiosis, were also upregulated during colonisation of both plants. [180], [232]

Uptake of specific carbohydrates was upregulated during colonisation of both plants, including the inositol transporter IntABC, and RL2720-RL2722, which encode an ABC transporter predicted to mediate uptake of arabinogalactan [205]. [238][239][237]Expression of genes involved in determining cell surface characteristics was also altered during Rlv3841 colonisation of both pea and barley, with 38% of L,D-transpeptidase genes, which are involved in peptidoglycan cross-linking, being downregulated during colonisation of at least one plant, and the BA14K-like immunoreactive protein encoded by RL0963 being upregulated during colonisation of both. [242], [243][246]Finally, and particularly interestingly, expression of a number of tRNA encoding genes was increased during colonisation of pea and barley. [248], [250]

Both central and secondary metabolism, and functions required for growth such as cell wall biosynthesis and cell division, were downregulated during colonisation of barley, suggesting that Rlv3841 in not well-adapted to growth on a non-host cereal rhizoplane. I also identified several promising genes whose promoters could be candidates for fusion to *nifA*, in particular the PAAT

transporter pRL110077-pRL110079, and RL3594 (*dapA*), both of which are significantly upregulated during colonisation of barley but not pea. However, the downregulation of so many genes involved in metabolism and growth shows that work is needed to ensure that any PGPR containing a construct for cereal-specific *nifA* expression would be able to survive and competitively colonise cereal rhizoplanes.[284]

Overall, this chapter shows the importance of investigating transcriptomic changes during colonisation of both host and non-host plants, and the power of a method that allows cells that are attached to the root surface to be separated from both contaminating plant matter and cells in the surrounding growth media. This method and the datasets it can generate have the potential to be an important resource for the understanding of the genetic basis of root colonisation by PGPRs, and can be used to inform their engineering for agricultural use.

4 The Rlv3841 transcriptome during attachment to and colonisation of the pea rhizoplane

4.1 Introduction

R. leguminosarum is a widely studied and agriculturally significant PGPR due to its ability to form nodules on legume roots and fix and release nitrogen for the benefit of its host. However, *R. leguminosarum* and the majority of its diazotrophic relatives can only form nitrogen-fixing nodules on a narrow range of specific legume hosts. Attachment to and colonisation of legume roots by rhizobial species happens in the same four stages explained above. Initially, rhizobia chemotactically follow concentration gradients of root exudates towards the root; in the legume-*Rhizobium* symbiosis, plant-secreted flavonoids are a particularly important component of root exudates that helps to establish specificity early in the interaction [112]. Primary attachment is characterised by the formation of weak, reversible bonds, including non-specific electrostatic, hydrophobic and Van der Waals forces and more specific bonds, such as the binding of legume lectins to the rhizobial surface polysaccharide glucomannan [127]. During secondary attachment, these interactions are supplemented by stronger, irreversible bonds between components that are synthesised in response to the detection of primary attachment. In particular, the synthesis of extracellular fibrils leads to the accumulation and aggregation of bacterial cells around plant root hairs. Extracellular polysaccharide (EPS) also plays a significant role in secondary attachment, being a major component of bacterial cell surfaces and highly specific to strains and species [24]. Rhizobial proteins with specific domains such as *Rhizobium*-adhering (Rap) domains have also been shown to mediate secondary attachment [285]. Once rhizobia have attached to the root surface, they can begin to grow and form biofilms, the final step in successful colonisation. A deeper understanding of the genetic pathways involved in each stage of colonisation of host plants by *R. leguminosarum* could help explain how this host specificity comes about and what factors are significant to formation of a successful symbiosis. Transcriptomic profiling of different stages in colonisation has been carried out on *Delftia acidovorans* during colonisation of soybean and canola roots [207] but never, to my knowledge, on a diazotrophic *Rhizobium* strain. As described in Chapter 3, I optimised a method for obtaining transcriptomic data from *R. leguminosarum* during long-term colonisation of the rhizoplane, seven days post inoculation. I then applied this method to obtain

transcriptomic data from *R. leguminosarum* during early attachment to the pea rhizoplane, in order to compare expression during different stages of colonisation of the host rhizoplane.

4.2 Results and Discussion

4.2.1 Rlv3841 transcriptome changes during attachment to and colonisation of the pea rhizoplane

RNA was isolated and sequenced from Rlv3841 during both attachment to and colonisation of *P. sativum* roots following the two methods described above. Sequencing was obtained to an average depth of 20.5 x 10^6 reads for Rlv3841 harvested during attachment and 53.2 x 10^6 reads for Rlv3841 harvested during colonisation. Both attachment and colonisation datasets were compared with the transcriptome of free-living Rlv3841 grown in minimal media. Gene set enrichment analysis was carried out using the FUNAGE-pro tool [187], and genes classified by ontology [214] to identify pathways and functions that are significantly differentially expressed during root attachment and colonisation. Genes that were over four-fold up- or downregulated in root-associated compared with free-living samples and had an adjusted p-value < 0.05 were considered significantly differentially expressed. Data concerning genes discussed in this chapter is summarised in Supplementary Table 2. Principal component analysis (PCA) showed that association with the pea rhizoplane accounts for the majority of difference in gene expression, with samples harvested during both attachment and colonisation being substantially different from that harvested from laboratory culture. However, there was also substantial difference between samples harvested during rhizoplane attachment and those harvested during colonisation, showing these to be two distinct processes involving expression of distinct sets of genes (Figure 4.1)

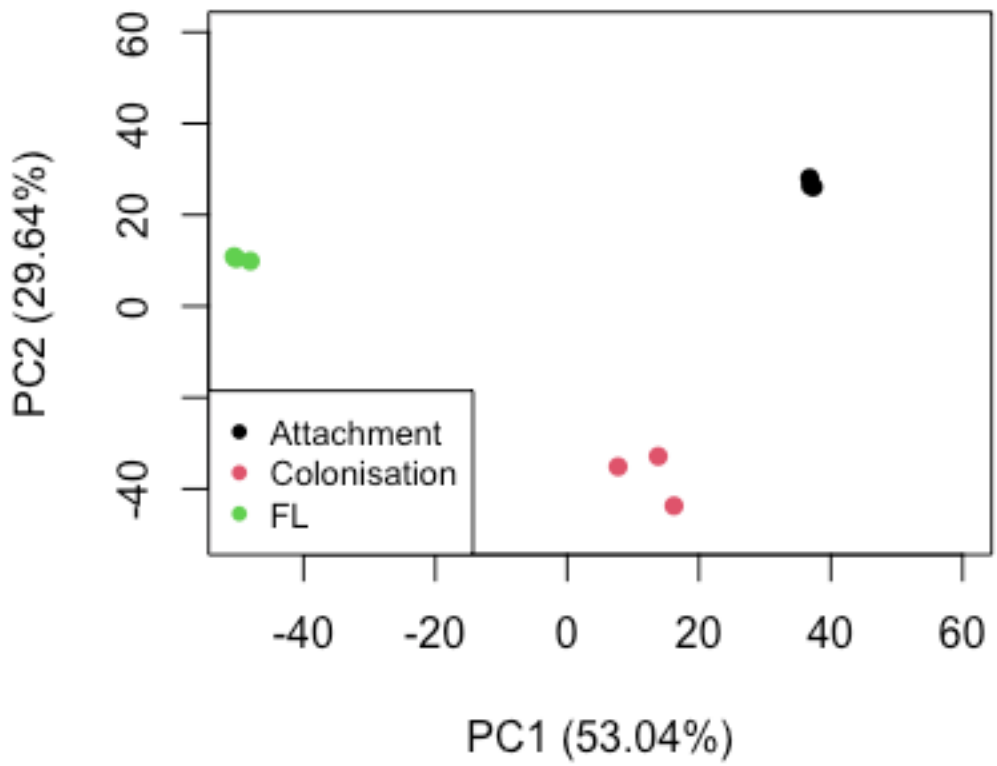


Figure 4.1: Principal Component Analysis of Rlv3841 gene expression during attachment to and colonisation of the pea rhizoplane, and in laboratory culture ('FL') (UMS supplemented with glucose (10 mM) and ammonium chloride (10 mM).

My results show that a total of 897 genes were significantly upregulated during attachment and 832 during colonisation. Of these, 494 (55% of genes upregulated during attachment; 59% of genes upregulated during colonisation) were upregulated in both samples. A total of 577 genes were downregulated during attachment and 346 during colonisation. Of these, 249 were downregulated during both stages of the interaction. Interestingly, only two genes, RL3889 (encoding a tricarboxylate transporter) and RL3891 (hypothetical exported protein), were significantly upregulated during attachment (4.5-fold, 5.4-fold) but downregulated during colonisation (7.9-fold, 8.7-fold), while none of the genes that showed increased expression during colonisation were repressed during attachment. RL3890, encoding another tricarboxylate transporter, was also significantly downregulated during colonisation (12.6-fold) and was upregulated during attachment, although below the threshold for

significance in this study (3.0-fold). Overall, 20% of the Rlv3841 genome was differentially regulated during attachment and 16% during colonisation. This result is in agreement with a previous study on Rlv3841 colonisation of the pea rhizosphere conducted in the Poole group. Of the genes upregulated during rhizoplane colonisation, 190 (23%) were shown in previous work to be upregulated in the rhizosphere, a sample which included cells vortexed off plant roots as well as those filtered out of the growth media [205]. Of the 346 genes downregulated during rhizoplane colonisation in my work, 85 (25%) were also significantly downregulated in the study on rhizosphere colonisation. Although this overlap is substantial, a significant majority of genes that were seen to be differentially expressed in this work did not appear in the previous study, showing the importance of generating a specific sample of rhizoplane-colonising cells for transcriptomic analysis that is enriched for those cells strongly attached to the root surface, rather than those that are growing in the surrounding media. In particular, ribosomal proteins encoded by RL1731, RL1762, RL1786, RL1789, RL1790 and RL1793 were downregulated in the pea rhizosphere but upregulated during rhizoplane colonisation, showing the difference between cells that are surviving around the root and those that are actively growing at the root surface. 27 genes involved in motility and chemotaxis were upregulated in the pea rhizosphere and, as previously mentioned, downregulated during colonisation of the rhizoplane. Interestingly, RL2977, encoding the permease component of an ABC transporter required for desiccation tolerance and biofilm formation [266], was significantly upregulated during rhizoplane colonisation and downregulated in the rhizosphere. This gene was not differentially expressed during initial attachment to the rhizoplane, suggesting that its involvement only begins during growth and replication on the root. Survival in the rhizosphere or at the rhizoplane clearly requires expression of a largely different set of genes, as colonising the root surface requires different processes than persistence in the soil.

4.2.2 Rlv3841 genes upregulated in both processes

Genes that were differentially expressed during both attachment and colonisation are shown in Figure 4.2.

4.2.2.1 Several key regulators of gene expression are upregulated in both attachment and colonisation

Genes involved in sigma factor activity (GO:0016987) are significantly upregulated during both attachment to and colonisation of pea roots by Rlv3841. In total, Rlv3841 has 20 genes encoding sigma factors, 4 of which are housekeeping sigma factors and 16 of which are extracytoplasmic (ECF) sigma factors [286]. Both Rlv3841 copies of the housekeeping sigma factor-encoding the heat shock response gene *rpoH* are significantly upregulated: RL3766 (*rpoH1*) is 4.8-fold upregulated during attachment and 8.6-fold upregulated during colonisation, while RL4614 (*rpoH2*) is 34.2-fold upregulated during attachment and 46.4-fold upregulated during colonisation. In *E. coli*, the main role of RpoH is to maintain protein structure and folding under high temperature and other protein-denaturing conditions, as well as under normal conditions [287]. A number of *Rhizobium* species have two or more RpoH-encoding genes, and in Rlv3841 RpoH1 and RpoH2 share 47% amino acid identity. Both *rpoH1* and *rpoH2* from *Rhizobium etli* [288] and from *S. meliloti* [289], [290] can complement the temperature-sensitive phenotype of an *E. coli rpoH* mutant, although in *S. meliloti* complementation is only viable at 30□C, not at 37□C [289], [290]. The *R. etli rpoH1* mutant is sensitive to heat shock, sodium hypochlorite and hydrogen peroxide, while the *rpoH2* mutant is sensitive to salt stress and sucrose, suggesting that in *R. etli rpoH1* is the main heat shock and oxidative stress response factor, with *rpoH2* being involved in osmotic stress response [288]. However, the *rpoH1H2* double mutant has increased sensitivity to both heat and oxidative stress compared with the *rpoH1* single mutant, suggesting an epistatic interaction with *rpoH2* also. An interaction between the two copies of *rpoH* is also implicated in nodule function, with *rpoH1* and *rpoH1H2* mutants in *R. etli* also showing reduced nitrogenase activity and premature senescence compared with *rpoH2* mutants and wild type [288]. Similarly, in *S. meliloti* mutants in *rpoH1* show impaired colonisation of plant roots and poor survival in nodules, with *rpoH2* mutants showing similar symbiosis phenotypes to wild type [290], but *rpoH1H2* double mutants show a more severe symbiosis phenotype than the *rpoH1* single mutant, with nodules rarely forming and those that do form contain suppressor mutants [291]. This suggests that the two

copies of *rpoH* interact to regulate the establishment of the symbiosis, which is borne out by our finding that transcription of both copies in Rlv3841 is significantly increased during attachment and colonisation of pea roots. An INSeq experiment in Rlv3841 showed that *rpoH1* is essential for attachment to pea roots at pHs 6.5-7.5 [133]. Moreover, five genes that are also required for root attachment in Rlv3841 have RpoH1 recognition sequences upstream of their coding regions: pRL100053, RL2520, RL2694, RL3766 (*rpoH1*) and RL4062. pRL120322 and RL4382 have an RpoH1/H2 binding motif in their upstream regions [133]. pRL100053, which encodes a transmembrane protein, was 2.3-fold upregulated in colonisation and not upregulated during attachment. However, it is required for attachment to pea roots at a range of pHs [133], as well as for colonisation of the pea rhizoplane and survival in nodules [169]. RL2520 encodes another transmembrane protein that was not upregulated during colonisation but was 2.4-fold upregulated during attachment. It is important for attachment to pea roots at acidic and neutral pHs and to soybean roots at pH 7 [133]. RL4382 (filamentous hemagglutinin adherence factor precursor) was 4.6-fold upregulated during attachment but was not upregulated during colonisation. In keeping with this result, it was not required for colonisation of pea roots [169], but was important for pea root attachment at a range of pHs as well as for attachment to barley roots [133]. RL2694 (glutathione reductase) and pRL120322 (*fhuA2*; siderophore receptor precursor) were 2.3-fold and 8.4-fold downregulated in attachment respectively, suggesting that they may be regulated by something else as well as RpoH1 that allows them to be regulated separately from the rest of RpoH1's targets. Notably, RL2694 was only important for attachment to pea roots at pH 7 and pRL120322 at pH 7.5 [133]. These results combined with our findings suggest that they may be involved in attachment only in a very narrow range of conditions. It is likely that *rpoH1* is essential for root attachment due to the involvement of these downstream targets and is therefore upregulated during attachment in order to activate this range of required genes. In a different INSeq experiment, *rpoH1* was not required for growth in the rhizosphere or colonisation of the root surface [169], despite being upregulated during colonisation. However, it was essential for undifferentiated bacteria living in nodules, which may explain its continued upregulation

beyond primary attachment. *rpoH2* alone is not required for attachment [133] or colonisation [169]; however, its significant upregulation during attachment to and colonisation of pea roots suggests that in Rlv3841, RpoH2 may regulate some or all of the same targets as RpoH1 to ensure the formation of a successful symbiosis, in the same way as it appears to in *R. etli* and *S. meliloti*, based on the symbiosis phenotypes of *rpoH1H2* double mutants. A degree of functional redundancy between RpoH1 and RpoH2 would explain the neutral INSeq phenotype particularly of *rpoH2* as a mutation in just one of the *rpoH* genes might be rescued by the presence of the other, functioning, copy.

A range of extracytoplasmic (ECF) sigma factors are also upregulated during both attachment and colonisation. Of the 16 total ECF sigma factors, two are significantly upregulated during both attachment and colonisation (i.e., above 4-fold upregulated, with an adjusted p-value < 0.05). A further four were over two-fold upregulated during both attachment and colonisation. Interestingly, the only ECF sigma factor that was significantly downregulated during either attachment to or colonisation of pea roots was pRL120319 (*rpoI*), which regulates expression of iron uptake genes and was 4-fold downregulated during attachment. The most strongly upregulated ECF sigma factor during attachment and colonisation was RL3703 (*ecfZ*), which was 16.1-fold upregulated during attachment and 33.6-fold during colonisation. *ecfZ* is part of an operon with an anti-sigma factor (RL3704), a transcriptional regulator that acts as an anti-anti-sigma factor and is part of a two-component sensing system (RL3705; *phyR*), and the histidine kinase component of the two-component system (RL3706). All four members of the operon are substantially upregulated during both attachment and colonisation. EcfZ is the General Stress Response sigma factor (also known as EcfG or RpoE). Activation of a general stress response transcription factor, such as *ecfZ*, requires phosphorylation of the anti-anti-sigma factor PhyR in a cascade initiated at the top level by sensory histidine kinases. These sensory histidine kinases respond to a variety of signals, including osmotic, oxidative, nutrient, pH, or heat stress [293] [294]. The general stress response is involved in adaptation to and colonisation of animals and plants. In *Methylobacterium extorquens*, it is needed for successful leaf colonisation [295], whereas in *B. japonicum*, it is needed during infection of soybean [296][297]. The activation of transcription of *ecfZ*

in *R. leguminosarum* during attachment and root colonisation can be predicted to be due to one or more signals detected by the top level sensory histidine kinases, of which *R. leguminosarum* possesses seven. These are likely to transphosphorylate PhyT and ultimately control phosphorylation of PhyR. PhyT was essential during all tested growth conditions [298], probably by controlling phosphorylation of PhyR and preventing lethal over-activation of *ecfZ* as shown in characterised systems [299] [300]. In *C. crescentus* the top-level sensory histidine kinases also activate LovK, which is involved in holdfast production and attachment [300]. An orthologue of LovK, encoded by pRL110320, has already been shown to be involved in attachment of *R. leguminosarum* Rlv3841 to pea roots [180]. Thus the sensory histidine kinases in *R. leguminosarum* may promote root colonisation both by increasing stress tolerance and LovK activation. pRL110007 (*ecfN*) was also significantly upregulated during both attachment and colonisation, being 17.9-fold upregulated during attachment and 4.2-fold upregulated during colonisation. Its homologue in *Mesorhizobium loti*, mlr0407, has been shown to be induced by heat stress [301]. It is predicted to be expressed as an operon with pRL110006, which encodes the anti-sigma factor AsfN, and pRL110005, encoding a transmembrane YccA-like protein with a role in membrane integrity. Both pRL110005 and pRL110006 are also significantly upregulated during attachment and colonisation. All three genes are much more strongly differentially expressed during attachment than colonisation, being between 4- and 7-fold upregulated during colonisation and between 17- and 95-fold upregulated during attachment. This raises the possibility that the role of pRL110006 and its downstream targets is more focussed in early attachment, with decreasing relevance as cells begin to divide at the root surface and biofilm formation begins.

A further four sigma factors were over two-fold upregulated during both attachment and colonisation. pRL110418 (*ecfO*) was 2.3-fold upregulated during attachment and 2.2-fold upregulated during colonisation. The *ecfO* mutant has been shown to be growth advantaged in bacteroids [169], but its increased expression during the earlier stages of symbiosis suggests that it may play a role before bacteroid formation that means it is still important to a successful symbiosis. pRL120580 (*ecfJ*) was 2.6-fold upregulated during attachment and 6.2-fold upregulated during colonisation. Its homologue

in *P. syringae*, PSPTO1043, is induced by oxidative stress and controls expression of genes involved in the oxidative stress response [302]. A double mutant in PSPTO1042/PSPTO1043 showed slight impairment of the early stages of colonisation of *Arabidopsis* seedlings [302]. RL0446A (*ecfA*) was 2.8-fold upregulated during attachment and 2.9-fold upregulated during colonisation. Its orthologue in *Pseudomonas putida*, RpoT, responds to solvent stress, while its orthologue in *B. japonicum*, blr2203, is clustered with the flagella synthesis genes *fliI-J* [286], which are known to be important in attachment [133]. RL3235 (*ecfE*) was 2.2-fold upregulated during attachment and 3.8-fold upregulated during colonisation. The *ecfE* mutant is growth advantaged in nodule bacteria and its anti-sigma factor, *asfE*, is essential for bacterial growth in the nodule [169], suggesting a complex role in the *Rhizobium*-legume symbiosis. Its homologue in *S. meliloti*, SMc01419 (*rpoE1*), regulates its own expression, as well as that of the anti-sigma factor SMc01420 and two adjacent hypothetical proteins SMc01418 and SMc01421. *rpoE1* is expressed during stationary phase in response to sulphite and controls expression of the sulphite dehydrogenase-encoding gene *sorT* [303]. The ECF sigma factor RL0788 (*ecfB*) was only differentially expressed during attachment, being 8.1-fold upregulated, but showing no significant change in expression during colonisation compared with growth in laboratory culture. RL3020 (*ecfR*), meanwhile, was 7.9-fold upregulated during colonisation but showed no significant change in expression in attachment. The homologue of RL3020 in *C. crescentus*, CC3253 (*sigF*), mediates oxidative stress response during stationary phase by regulation of genes including *msrA*, which encodes methionine sulfoxide reductase, and *sodA*, which encodes superoxide dismutase [227]. Overall, half of the 16 ECF sigma factor-encoding genes in Rlv3841 showed some form of upregulation during attachment, colonisation, or both, with *ecfZ* and *ecfN* being substantially upregulated. As each of these sigma factors will activate transcription of a range of downstream targets, this shows the complexity of the transcriptional response of Rlv3841 during its initial interaction with host plants and highlights the importance of continuing to unpick the genetic basis of these interactions to further understanding.

4.2.2.2 DNA binding and initiation of transcription functions are upregulated during both attachment and colonisation

Genes involved in double-strand break repair via nonhomologous end joining (NHEJ) (GO:0006303) are significantly enriched in the set of genes upregulated during attachment and colonisation, with both *ku* genes encoded in the Rlv3841 genome (RL2141 and RL2142) as well as a putative *ligD* gene (pRL11306) being upregulated during both phases of interaction with pea roots. RL2141 was 106.8-fold upregulated during attachment and 49.8-fold during colonisation, RL2142 62.2-fold upregulated during attachment and 59.4-fold during colonisation, and pRL110306 34.3-fold upregulated during attachment and 20.5-fold during colonisation. NHEJ is one of two types of bacterial DNA repair, during which Ku proteins bind the broken ends of DNA and recruit DNA ligase (LigD) which reattaches them [304]. NHEJ has been shown to be more important for DNA repair during stationary phase than during exponential growth, with *ku* gene mutants in *Mycobacterium smegmatis* showing increased sensitivity to ionizing radiation than wild type during stationary phase, but not during exponential phase [305]. This suggests that the upregulation of the Rlv3841 *ku* and *ligD* genes during attachment and colonisation experiments carried out over a period of hours or days, when compared with an exponential laboratory culture, may simply be because these gene products are required for bacterial persistence during stationary phase. Ku proteins are also implicated in desiccation resistance in *S. meliloti*, with a *ku* gene mutant showing 1000-fold reduced persistence under desiccation stress compared with wild type [306]. However, *ku* genes are not required for development of healthy symbiotic nodules of *S. meliloti* on *M. truncatula* [306]. In Rlv3841, RL2142 is 29.5-fold upregulated during growth in the pea rhizosphere, while RL2142 is 26.4-fold upregulated and pRL110306 8.0-fold upregulated [205], showing that these genes are not only upregulated during direct interaction with host plant roots but also in the surrounding soil. This, combined with the finding that NHEJ is important for desiccation resistance in *S. meliloti* but not essential for symbiosis, suggests that NHEJ may be most important for persistence of Rlv3841 in the soil and the initial stages of interaction with plant roots, rather than for survival in the nodule.

4.2.2.3 Aldo/keto reductases involved in toxin degradation are upregulated during both attachment to and colonisation of pea roots

Genes associated with thiamine metabolic processes (GO:0006772) are significantly upregulated during both pea root attachment and colonisation in Rlv3841 ($p = 2.7x10^{-3}$ in attachment; $5.1x10^{-3}$ in colonisation). The two genes in this set that are most significantly upregulated during both processes are RL0272 and RL2983, both of which encode aldo/keto reductases. RL0272 is 5.8-fold upregulated during attachment and 6.0-fold upregulated during colonisation, while RL2983 is 5.1-fold upregulated during attachment and 5.8-fold upregulated during colonisation. Aldo/keto reductases are able to degrade both endogenous and exogenous toxins, including the byproducts of central metabolism such as methylglyoxal [307] and environmental toxins such as the mycotoxin deoxynivalenol [308], [309]. Indeed, RL2983 is predicted to contain a methyglyoxal reductase-related domain (PTHR43364), while RL0272 is predicted to be a member of the aflatoxin B1 aldehyde reductase family (PTHR43625) [310]. Bacterial aldo/keto reductases are also known to play a role in the degradation of aldehydic plant phytoalexins [311], which may explain the enrichment of these genes in the attachment and colonisation transcriptomes in Rlv3841. RL0272 has also been shown to be 4.7-fold upregulated in response to growth on salicylic acid, a component of plant root exudates [312] and to be significantly upregulated both in Rlv3841 bacteroids extracted from pea nodules and in *R. leguminosarum* bv *phaseoli* bacteroids extracted from bean nodules [219], [270], suggesting a role that begins during primary attachment and extends throughout the development of the nodule.

4.2.2.4 ABC transporters involved in import of a range of compounds are upregulated during both attachment and colonisation

Genes encoding ATP-binding cassette (ABC) transporters were significantly upregulated during both attachment and colonisation. ABC transporters are ATP-dependent transporters responsible for the translocation of a variety of substances across membranes. Rhizobial genomes encode significantly more ABC transporters than those of other bacteria, mainly due to nutrient limitation in soils [236]. In

particular, ABC transporters involved in protein transport (GO:0015031) were significantly upregulated during attachment to and colonisation of the pea rhizoplane. RL2338-RL2342, which encode the solute-binding protein (SBP), permease, and ATP-binding components of a PepT family ABC transporter, were between 53 and 103-fold upregulated during attachment and between 28 and 57-fold upregulated during colonisation of pea roots. These ABC transporter components RL2338-RL2342 form a predicted operon with the upstream gene RL2337, which is also upregulated during both attachment and colonisation, encoding a GntR family transcriptional regulator that may be responsible for activation of transcription of these genes. Moreover, the adjacent genes, RL2335-RL2336, were also upregulated during both processes and they are predicted to be expressed as an operon, potentially also controlled by RL2337. A further two peptide ABC transporters were upregulated solely during attachment, showing no differential expression during colonisation. These were RL3543-RL3547, which were 2-6-fold upregulated during attachment, and RL4575-RL4580, which were 6-16-fold upregulated. Genes encoding ABC transporters involved in import of D-ribose (GO:0015611) were upregulated during both attachment to and colonisation of pea roots. In particular, RL2720-RL2722, which encode the permease, SBP and ATP-binding components respectively of a CUT2 family ABC transporter, were between 3- and 10-fold upregulated during attachment and between 32- and 49-fold upregulated during colonisation. RL4172-RL4174, which encode a second CUT2 family ABC transporter, were between 3- and 14-fold upregulated during attachment and between 3- and 26-fold upregulated during colonisation. Expression of RL2720-RL2722 is known to be induced in the presence of arabinogalactan [205]. Cell wall-localised arabinogalactan proteins are common on plant root surfaces and have been documented to be important in cell expansion, root hair development, environmental response and interactions with other organisms, including microorganisms [234], [237], [313]. The quaternary amine ABC transporter *qat6* genes (pRL100079-pRL100083) were between 14- and 61-fold upregulated during attachment and between 19- and 35-fold upregulated during colonisation. Qat transporters are induced by the quaternary amines glycine betaine and choline, and Qat6 is a high affinity transporter of both compounds [314]. Qat6 is also the most strongly upregulated Qat ABC

transporter under osmotic stress [314] and is significantly upregulated under iron-limiting conditions [219], suggesting that it may contribute to nutrient uptake under stress conditions. Mutants in *qat6* genes are growth advantaged in the rhizosphere [169] and during root attachment [133], despite these genes being upregulated during attachment and colonisation. This suggests that the Qat6 transporter may be useful under stress conditions in the environment, leading it to be upregulated during root colonisation, but not actually be necessary under controlled laboratory conditions. pRL110373 and pRL110374, which encode *sulB* and *sulA*, the permease and SBP components of an ABC transporter, respectively, were significantly upregulated during both attachment and colonisation. The other two components, pRL110371 and pRL110372, were 2-fold upregulated or above during both attachment and colonisation also. The *S. meliloti* homologue of the permease component, SMb21133, is induced by sulphate limitation [236]. This once again suggests that ABC transporters that are important under stress conditions play a role during host root attachment and colonisation by Rlv3841. Other ABC transporters that were over two-fold upregulated during attachment and colonisation include the peptide transporters pRL110434-pRL110438 and pRL110454-pRL110456, the polyamine/opine/phosphonate uptake (POPT) transporter pRL120145-pRL120147, the polar amino acid uptake (PAAT) transporters pRL80063-pRL80064, whose closest homologue is a mimosine transporter in *Rhizobium* sp. TAL1145, and RL1042-RL1045. This range of differentially expressed ABC transporters during initial interaction with the plant root shows the significant adjustment in nutrient uptake associated with the lifestyle shift from free-living soil bacteria to root-associated symbionts.

4.2.2.5 Genes encoding carbohydrate-binding products (GO:0030246) are upregulated during interactions with pea roots

ABC transporters are not the only transporters upregulated during attachment to and colonisation of host roots. The TRAP transporter *dctPQM* (pRL120400-pRL120402) was also significantly upregulated during both, as was the GntR family transcriptional regulator pRL120399, which may activate expression of the transporter-encoding genes. pRL120400 is induced by addition of acetoacetate to

growth media [219], while pRL120400-pRL120402 are all induced by pea root exudates [205]. The *dctP* homologue in *S. meliloti*, SMb20442, is induced by fucose and mannose [236]. RL0507 (*manB*), which encodes the mannose-degrading enzyme mannosidase, was 9.7-fold upregulated during attachment and 11.7-fold upregulated during colonisation. Overall, this suggests that expression of these genes during interactions with plant roots represents a switch to metabolising different sugar types found in root exudates and at the root surface. The glucose 1-dehydrogenase-encoding gene RL0634 was also 21.4-fold upregulated during attachment and 35.8-fold upregulated during colonisation. Moreover, the succinate dehydrogenase operon RL4443-RL4446 (*sdhBACD*) was downregulated during both attachment and colonisation, suggesting a change in sugar uptake and metabolism.

A second set of carbohydrate-binding proteins that are upregulated during attachment and colonisation is a range of BA14K family proteins. BA14K is a 14 kDa *Brucella abortus* protein that has lectin-like properties and a particular affinity for mannose and is predicted to play a role in lipopolysaccharide (LPS) biosynthesis as the mutant shows altered LPS formation leading to decreased virulence [246]. BA14K family protein genes RL0963, RL1842 and RL2969 were all upregulated during Rlv3841 attachment to and colonisation of pea roots. A mutant in RL0963 was deficient in attachment to soybean roots [133]. Although it is unclear what role these proteins play in Rlv3841, links to BA14K suggest a role in LPS biosynthesis which may be involved in ability to attach to and grow on plant root surfaces.

4.2.2.6 Photoreceptor-related genes are upregulated during attachment to and colonisation of pea roots

Histidine kinase-encoding genes predicted to have a role in photoreceptor activity (GO:0009881) were significantly upregulated during both attachment and colonisation. RL1162 was 22.2-fold upregulated during attachment and 47.8-fold upregulated during colonisation, RL3706 8.6-fold upregulated during attachment and 30.8-fold upregulated during colonisation, and pRL110320 46.1-fold upregulated during attachment and 51.0-fold upregulated during colonisation. The pRL110320 gene product

absorbs blue and UVA light and regulates production of flagellar proteins and EPS and attachment to abiotic surfaces [180]. RL3706 expression is responsive to phenylalanine and acidic pHs [219], suggesting that although it is predicted to be classified in the gene ontology function of photoreceptor activity, it may be responding to plant root exudates in this instance rather than light. However, the deoxyribopyrimidine photo-lyase-encoding gene *phr* (RL1985) was also 26.3-fold upregulated during attachment and 45.4-fold upregulated during colonisation. Photolyases utilise blue and UVA light to repair DNA damage [315] and upregulation of *phr* together with pRL110320 suggests that response to these wavelengths of light plays a role during attachment to and colonisation of host plant roots.

4.2.3 Rlv3841 genes downregulated in both processes

4.2.3.1 Genes involved in motility and chemotaxis are significantly downregulated during both attachment and colonisation.

The sets of genes downregulated in attachment and colonisation were significantly enriched for genes encoding products involved in flagellum-dependent cell motility (GO:0071973, GO:0097588) and chemotaxis (GO:0006935). Flagellum assembly genes RL0687, RL0695 (*fliF*), RL0700-RL0701 (*fliGN*), RL0704 (*fliF*), RL0705 (*fliI*), RL0707-RL0716 (*flgBC*, *fliE*, *flgG*, *flgA*, flagellar P-ring protein, flagellar protein, *flgH*, *fliLP*), RL0718-RL0721 (*flaABCD*), RL0728 (*flgE*), RL0732 (*flbT*), RL0735 (*flhA*) were downregulated during both attachment and colonisation. Pilus assembly genes RL0214-RL0218 (*rcpA*, *cpaDEF*) were downregulated during both attachment and colonisation. Chemotaxis genes RL0429 (*hemAT*), RL0605 (*icpA*), RL0687-RL0689 (*cheY1A1W1*), RL0690 (*cheR1*) RL0691 (*cheB1*), RL0692 (*cheY2*), RL0700-RL0703 (*fliG*, *fliN*, *fliM*, *motA*), RL0723-RL0725 (*motBC*, *fliK*), RL0972 (*mcpK*), RL2683 (*mcpD*), RL2391 (*mcpN*), RL4029-RL4035 (*cheR2*, *cheW2b*, *mcrABC*, *cheW2A*, *cheA2*), RL4386 (*cheW4*) were downregulated during both attachment and colonisation. The master regulators of flagella and chemotaxis, RL0696 (*visN*) and RL0697 (*visR*), were 9.7-fold and 8.0-fold downregulated during colonisation and 3.0- and 3.7-fold downregulated during attachment. Many flagella and chemotaxis genes, including the *che1* cluster RL0686-RL0693 and *visNR*, are needed for function in bacteria and

differentiated bacteroids in nodules, but not earlier in colonisation, suggesting that they play a role in chemotaxis towards the root or early entry into the infection thread but not during actual colonisation of the root surface [169]. In *S. meliloti*, chemotaxis and motility genes including *flaF*, *flbT*, *cheY1*, *flaC*, *flgB*, *flgF* and *motA* are controlled by the quorum sensing system ExpR/Sin, which downregulates their expression at high population density [254]. The ExpR/Sin system controls biofilm formation in response to population density in *S. meliloti* by regulating expression of genes involved in exopolysaccharide (EPS) synthesis [152]. The findings in *S. meliloti* suggest that the chemotaxis genes are not only not essential for rhizoplane colonisation, as noted by Wheatley et al [169], but that their expression is actively repressed during this process, which supports our finding that these genes are significantly downregulated in Rlv3841 isolated from the rhizoplane.

4.2.3.2 Components involved in polar cell wall synthesis are downregulated during attachment and colonisation.

Genes encoding products with peptidoglycan LD-transpeptidase activity (GO:0071972) are downregulated during both attachment and colonisation. LD-transpeptidases are a class of cell wall synthesising enzymes, which create crosslinks between peptidoglycans, and Rhizobiales genomes encode a larger number of LD-transpeptidases than most bacterial species [316]. Of the 16 LD-transpeptidases encoded in the Rlv3841 genome, 7 are more than two-fold downregulated during attachment and 6 in colonisation, with RL1296 (*ldt4*) and RL1458 (*ldt6*) both being 8.0-fold downregulated during attachment and 6.1-fold and 6.6-fold downregulated during colonisation respectively. Only RL1845 (*ldt9*) and RL3342 (*ldt11*) were upregulated during either attachment or colonisation, with *ldt9* being 37.6-fold upregulated during attachment and 36.7-fold upregulated during colonisation, and *ldt11* being 3.3-fold upregulated during attachment and 3.2-fold during colonisation. These changes in expression of LD-transpeptidase genes may suggest alterations in cell wall composition during the lifestyle switch from free-living to root-colonising bacteria. Further suggesting changes in cell wall composition, various genes involved in LPS synthesis were

downregulated. In particular, RL1662, encoding a glycosyltransferase, was 5.4-fold downregulated during attachment and 5.5-fold downregulated during colonisation, and the sugar transferase RL1661 (*gmsA*) 7.0-fold downregulated during attachment and 3.5-fold downregulated during colonisation. GmsA controls synthesis of the surface polysaccharide glucomannan [128]. Further, the predicted operon RL0242-RL0246 involved in O antigen biosynthesis (GO:0009243), the portion of the LPS that varies between bacterial strains and species [317], was downregulated during both attachment and colonisation. RL0240, encoding a perosamine synthetase, was 8.8-fold downregulated during attachment and 4.6-fold downregulated during colonisation. RL0245, encoding a CDP-glucose 4,6 dehydratase, was 5.6-fold downregulated during attachment and 9.2-fold downregulated during colonisation, and RL0246, encoding the glucose-1-phosphate cytidylyltransferase DdhA, was 12.0-fold downregulated during attachment and 12.2-fold downregulated during colonisation. The O antigen biosynthesis protein RL0242 was 7.1-fold downregulated during attachment. Components involved in the synthesis of dTDP-rhamnose (GO:0019305), a precursor for O antigen biosynthesis [318] were also downregulated. RL0795 was 7.8-fold downregulated during attachment and 5.0-fold downregulated during colonisation, and RL1623-RL1627 were over fourfold downregulated during attachment and over twofold downregulated during colonisation. This downregulation of a range of genes involved in cell surface characteristics suggests that during initial interaction with plant roots Rlv3841 undergoes cell surface remodelling. However, the majority of genes involved in LPS synthesis were substantially more downregulated during attachment than colonisation, when both samples were compared with free-living laboratory culture. This means that the majority of these genes were expressed more strongly during colonisation than during attachment to the rhizoplane. For example, *gmsA* was 2-fold more highly expressed during colonisation than during attachment. This suggests that these genes are repressed during initial attachment but that their expression begins to increase again during biofilm development, consistent with a well-documented role for LPS in biofilm formation [156], [319].

4.2.3.3 Sulphur uptake changes during both attachment to and colonisation of pea roots.

Genes involved in arylsulfatase activity (GO:0004065) were significantly downregulated during both attachment and colonisation, with the exported arylsulfatase gene RL1149 being 11.3-fold downregulated during attachment and 9.9-fold downregulated during colonisation and the arylsulfatase genes RL1236 being 21.1-fold downregulated during attachment and 4.1-fold downregulated during colonisation, and RL1237 being 7.9-fold downregulated during attachment and 5.0-fold downregulated during colonisation. The Rlv3841 genome encodes a total of 7 arylsulfatase genes. Of the remaining 4, RL2267 was 2.1-fold downregulated during attachment and 3.0-fold downregulated during colonisation, and the remaining three showed no significant differential expression. Arylsulfatases are used by bacteria to satisfy their sulphate requirements by releasing sulphate from aromatic sulphate esters. The putative sulphate transporter encoded by *sulABCD* was upregulated during both attachment and colonisation, and the ABC transporter encoded by members of the *ssu* operon (RL2767-RL2770, *ssuBCDA*) was significantly upregulated during colonisation of pea roots by Rlv3841, with the four genes being 7.2-fold, 9.8-fold, 9.1-fold and 13.4-fold upregulated. All four genes were not significantly upregulated during attachment, but *ssuB* was 4.7-fold upregulated and *ssuA* 6.7-fold upregulated, with *ssuC* and *ssuD* being over twofold upregulated also. The *ssuB* homologue is upregulated in *R. leguminosarum* A34 during colonisation of the pea rhizosphere [320], and during colonisation of the maize rhizosphere by *P. putida* [321], while Rlv3841 mutants in *ssuB* and *ssuC* show impaired attachment to soybean roots at pH 7.0 [133]. This suggests that this sulphate transporter is significant for plant-microbe interactions, and that a change in sulphur uptake may be common during microbial colonisation of plant roots. *ssuB*, *ssuC* and *ssuA* encode the ATP-binding, permease and SBP components of the ABC transporter which moves alkanesulfonates into the cell, while *ssuD* encodes the enzyme alkanesulfonate monooxygenase, which bacteria use to obtain sulphur from alkanesulfonates during sulphur limited conditions [322]. The genes encoding both subunits of sulphate adenylyltransferase, RL1260 and RL1261 (*cysND*), are 9.6-fold and 11.6-fold upregulated during colonisation. RL1262, which is 11.3-fold upregulated, is predicted to be expressed as an operon

with RL1260-RL1261 and encodes phosphoadenosine phosphosulphate reductase. Sulphate adenylyltransferase catalyses the first step of intracellular sulphate assimilation, with phosphoadenosine phosphosulphate reductase acting downstream. This suggests that Rlv3841 may adjust its sulphur source during interaction with plant roots, as the genes involved in its uptake and metabolism are differentially expressed.

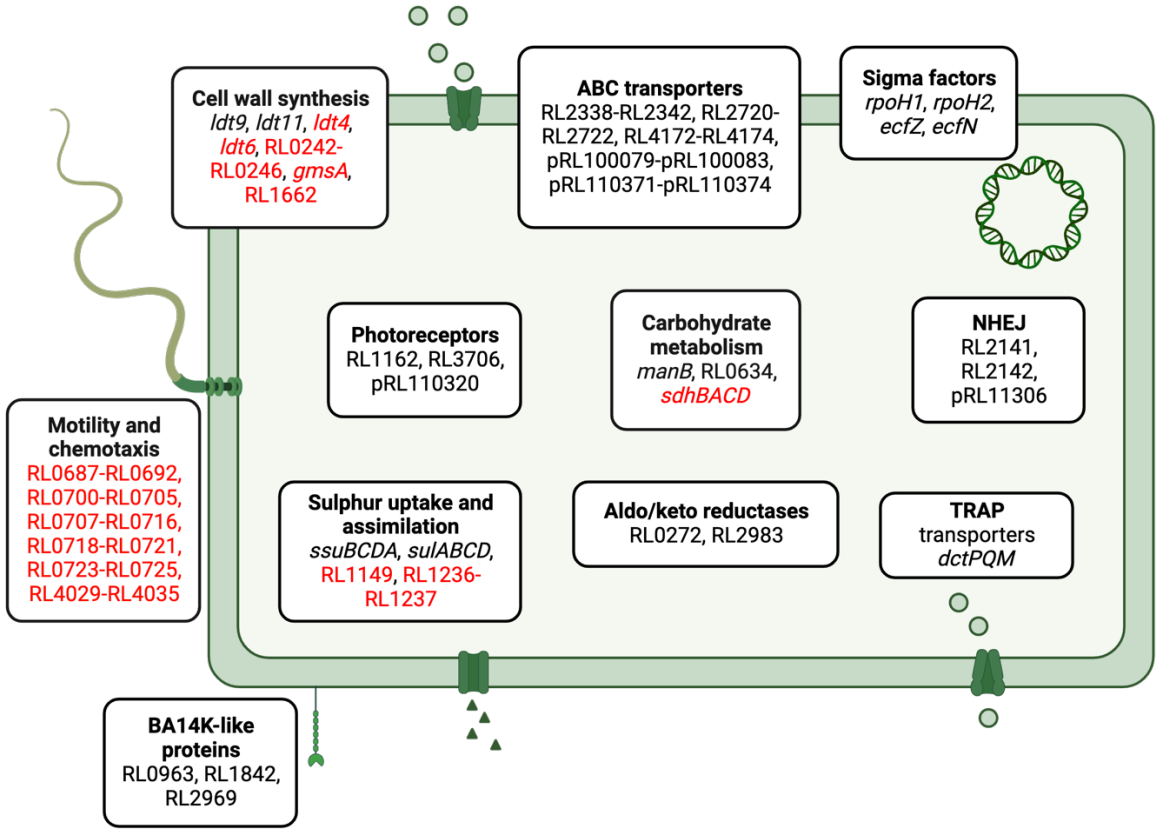


Figure 4.2: Genes differentially expressed during both attachment to and colonisation of the pea rhizoplane. *Genes shown in black were significantly upregulated during both processes and those shown in red were significantly downregulated.*

4.2.4 Rlv3841 genes upregulated only in attachment

Genes that were differentially expressed during attachment only are shown in Figure 4.3.

4.2.4.1 Transcriptional regulators are significantly upregulated during attachment.

32 transcriptional regulators were significantly upregulated during attachment but not colonisation of pea roots (Table 4.1). This shows the significant difference in transcriptome modification between primary attachment, over a period of hours, and biofilm formation and colonisation, over a period of days, and the importance of distinguishing between the two as separate stages of symbiosis development.

Table 4.1: Transcriptional regulators significantly (> 4x) upregulated during pea root attachment by Rlv3841.

Gene	Family	Fold change	Adjusted p-value
RL2879	AraC	13.7	1.8×10^{-34}
RL1133		39.0	3.9×10^{-30}
RL1168	ArsR	5.8	1.9×10^{-9}
RL1214		6.4	7.7×10^{-25}
RL0989 (*matR*)	GntR	4.9	2.6×10^{-6}
RL2337		5.8	2.9×10^{-23}
RL4574		14.2	1.9×10^{-27}
RL0083	LacI	15.2	3.9×10^{-28}
RL1820		48.1	1.2×10^{-38}
RL1995		6.4	1.1×10^{-22}
RL2343		11.3	2.7×10^{-18}
RL0273	LysR	19.4	2.6×10^{-81}

RL0851		4.2	$3.7x\ 10^{-16}$
RL1040		6.1	$9.7x\ 10^{-7}$
RL1202		4.5	$9.8x\ 10^{-15}$
RL1209		5.9	$1.6x\ 10^{-29}$
RL1404		4.7	$6.9x\ 10^{-21}$
RL1674		5.3	$5.9x\ 10^{-23}$
RL1960		6.7	$7.7x\ 10^{-43}$
RL2443		8.8	$1.7x\ 10^{-19}$
RL2486		5.4	$2.5x\ 10^{-27}$
RL2501		6.5	$4.7x\ 10^{-13}$
RL2839		4.3	$1.3x\ 10^{-6}$
RL3161		4.3	$6.8x\ 10^{-7}$
RL3712		6.5	$1.7x\ 10^{-23}$
RL4055		8.8	$7.0x\ 10^{-22}$
RL4189		4.6	$1.6x\ 10^{-17}$
RL2550	MerR	10.8	$1.3x\ 10^{-49}$
RL0500		5.5	$8.1x\ 10^{-13}$
RL1233	TetR	5.8	$2.7x\ 10^{-14}$
RL2667A		4.2	$2.3x\ 10^{-18}$
RL1441	Two component sensor/regulator; transcriptional regulator	18.4	$6.1x\ 10^{-26}$

4.2.4.2 Genes involved in glutathione metabolism are upregulated during attachment.

Genes involved in glutathione synthesis and degradation are upregulated during Rlv3841 attachment to pea roots. Both gamma-glutamyl transpeptidase genes RL4050 (*ggt1*) and RL4578 (*ggt2*) were significantly upregulated. RL4050 was 8.3-fold upregulated and RL4578 14.7-fold. Glutamyl transpeptidases are responsible for catalysis of the breakdown of the tripeptide glutathione and can also be virulence factors in pathogenic species such as *Helicobacter pylori*, which use their host's glutathione as a nitrogen and nutrient source [323]. RL4578 is predicted to be expressed with the operon RL4576-RL4580, all members of which were significantly upregulated during attachment. RL4576, RL4577 and RL4580 encode the permease and ATP-binding components of a peptide uptake ABC transporter and RL4575, which encodes the SBP of the same transporter, was also significantly upregulated, as was the GntR-family regulator gene RL4574. This suggests the transcriptional regulator may be responsible for activating expression of this operon, and that the peptide transporter may be involved in glutathione uptake. One glutathione synthetase gene RL0855 (*gshA*) was 4.3-fold upregulated during attachment and the other, RL0338 (*gshB*), was 2.7-fold upregulated. Glutathione synthetase has been shown to be essential for nodule formation and health in a range of rhizobial strains including *R. leguminosarum* [278], [279], [280], [281], as glutathione is important in protecting cells from oxidative stress. The *gshA* mutant has impaired attachment to pea roots at a range of pHs, as well as to soybean and barley roots [133], as well as being unable to colonise root surfaces or survive in nodules [169], while the *gshB* mutant has impaired attachment to soybean roots [133] and is deficient in colonisation of root surfaces and unable to survive in nodules [169]. Glutathione synthesis is therefore clearly important from attachment onwards, despite these genes not being upregulated during colonisation. Glutathione degradation, however, seems uniquely important during root attachment.

4.2.4.3 Transport of a range of compounds is upregulated during attachment.

Genes involved in carbohydrate transport are upregulated during attachment to pea roots. The genes encoding the components of the carbohydrate uptake transporter-1 (CUT1) family ABC transporter RL0092-RL0094 were all upregulated. The genes encoding the CUT1 family transporter responsible for the import of alpha-glucoside, RL0745-RL0748 were upregulated, with RL0745 (*aglE*) being significantly (4.2-fold) upregulated and the other components being over twofold upregulated, with the exception of RL0748, which was not significantly changed. The *S. meliloti* homologue of RL0745, SMc03061, is induced by maltotriose and turanose [236]. The components of a further CUT1 transporter, RL1824-RL1827, were 40.1-fold, 23.9-fold, 11.4-fold and 3.3-fold upregulated. The *S. meliloti* homologue of the SBP encoded by RL1824, SMc04396, is induced by dextrin [236]. The components of the CUT1 transporter RL2345-RL2348 were between 4-fold and 12-fold upregulated, and the CUT1 transporter components RL2793-RL2796 were between 3-fold and 9-fold upregulated. The *S. meliloti* homologue of the SBP RL2796, SMc04295, is induced by gentiobiose and salicin and binds cellobiose [236]. A number of carbohydrate uptake transporter-2 (CUT2) family ABC transporters were also upregulated. RL1996-RL1999, encoding the components of a CUT2 transporter, were all upregulated during attachment, as were RL2720-RL2722, which are known to be induced by arabinogalactan [205], and RL4172-RL4174. The MFS family sugar transport protein encoded by RL2445 was also 4.5-fold upregulated.

Transporters involved in import of a range of other compounds were also upregulated during attachment but not colonisation. Two transporters involved in transport of glycine betaine, which has roles both as an energy source and in osmotic stress response [324], were significantly upregulated. The QAT1 operon, RL3533-RL3535 (*qatX1W1V1*), was significantly upregulated, with *qatX1* and *qatW1* being 4.2-fold upregulated and *qatV1* 3.9-fold upregulated. The QAT2 operon genes, RL0509-RL0512 (*qatY2V2W2X2*), were all between 5-fold and 6-fold upregulated. A mutant in the QAT1 operon is known to be unable to transport choline or glycine betaine [314]. RL1522, which encodes a

transmembrane MscS class ion channel, was also upregulated. MscS channels are responsible for the release of internal ions under hypo-osmotic conditions [325] and upregulation of RL1522 in combination with upregulation of the QAT1 and QAT2 transporters suggests that an osmotic stress response is important during root attachment. The polyamine/opine/phosphonate uptake (POPT) transporter encoded by RL4190-RL4193 was upregulated. The two permease genes, RL4192-RL4193, were 5.0-fold and 5.5-fold upregulated. The SBP and ATP-binding protein were not significantly upregulated but were above 3-fold up. The *S. meliloti* homologue of the SBP component, SMc01652, is induced by putrescine and agmatine [236]. The uncharacterised transmembrane transporter genes RL0353 and RL0529 were also both 6.6-fold upregulated. The peptide transporters encoded by RL4575-RL4580 and RL3543-RL3547 were upregulated, as previously mentioned. The RL3543 homologue in *S. meliloti*, *oppA* (SMb21196), is induced by methionine [236] and co-purifies with the sulphite dehydrogenase SorT [326]. It also has high sequence similarity to the SBP of the *B. subtilis* transport system Opp [236], which is implicated in cell wall turnover and the initiation of *B. subtilis* sporulation [327]. The Opp system in *S. meliloti* is required for the uptake of tetrapeptides and some tripeptides, including 3Ala, but not for formation of healthy symbiotic nodules on alfalfa [328]. The zinc ion transport system encoded by RL3176-RL3178 (*znuBCA*) was upregulated, as was the transcriptional regulator gene RL3175 (*zurR*) which regulates response to zinc. The manganese/zinc/iron chelate uptake (MZT) transporter encoded by RL1047-RL1049 was also significantly upregulated, with RL1047 being 109.3-fold upregulated, RL1048 17.5-fold and RL1049 47.4-fold. This suggests that there is a significant response to zinc during initial attachment of rhizobia to roots, but not later in colonisation.

A number of genes involved in the export of compounds were also upregulated in cells attached to pea roots. The MFS family permease RL0996 was 10.8-fold upregulated in attached cells. RL0996 has 82% to an *Agrobacterium vitis* tartrate transporter. RL1329-RL1330 (*salBA*) were 2.0-fold and 4.7-fold upregulated and the LysR family transcriptional regulator RL1328 (*salR*) slightly downregulated. The RND family efflux transporters encoded by RL1453-RL1454, RL2364-RL2365, RL3269-RL3270 and RL4223-RL4224 were significantly upregulated. RL2364-RL2365 were 6.3-fold and 5.0-fold upregulated

and are predicted to be involved in xenobiotic export (GO:1990961). The permease components of export ABC transporters encoded by RL0524 and RL3488 were 87.5-fold and 5.0-fold upregulated.

4.2.4.4 Malonate degradation is upregulated during attachment.

Genes involved in degradation of the dicarboxylic acid malonate were upregulated during attachment. The GntR family transcriptional regulator RL0989 (*matR*) was 4.9-fold upregulated, and the members of the operon it controls, RL0990-RL0992 (*matABC*), were 46.7-fold, 22.9-fold and 7.1-fold upregulated, respectively. RL0990 encodes malonyl-CoA decarboxylase, RL0991 malonyl-CoA synthetase and RL0992 a transmembrane malonate carrier protein. RL0990-RL0992 play an essential role in degradation of malonate, which is an inhibitor of succinate dehydrogenase, meaning that if they are mutated Rlv3841 is unable to grow on media containing both succinate and malonate [329]. Malonate is the predominant acid in soybean roots, and its concentration in the root increases as soil nitrogen availability decreases [330], meaning that expression of the *matABC* genes may be needed during attachment to remove this competitive inhibitor and allow cells to continue to metabolise succinate. Malonate is secreted by pea roots early on in colonisation but is not present in nodules [238], and Rlv3841 mutants in *matA*, *matB* or *matC* are able to form healthy nodules that fix nitrogen to the same levels as wild type [329]. This, combined with significant upregulation of RL0990-RL0992 during rhizoplane attachment but not colonisation, suggests that malonate is an important carbon source during early attachment but does not play a role in the later stages of symbiosis. Genes involved in acetolactate synthesis (GO:0003984) were also significantly upregulated during attachment. RL2442, encoding acetolactate synthase, was 5.5-fold upregulated, while RL3244-RL3245 (*ilvHI*), which are expressed as an operon and encode two subunits of acetolactate synthase, were 6.2-fold and 3.8-fold upregulated. RL0915-RL0916 (*dgoA*, *dgoK*), which play a role in catabolism of D-galactonate (GO:0034194), were 4.1-fold and 5.8-fold. No significant result was obtained for RL3850 (*dgoD*).

4.2.4.5 GABA degradation is upregulated during attachment.

Genes involved in the breakdown of gamma-aminobutyric acid (GABA) (GO:0009450) were upregulated during root attachment. Succinate semialdehyde dehydrogenase is a key enzyme in the GABA degradation pathway and the Rlv3841 genome encodes six succinate semialdehyde dehydrogenases [107], four of which were significantly (above 4-fold) upregulated during attachment. RL0101 (*gabD2*) was 5.9-fold upregulated and is predicted to be expressed as an operon with RL0102 (*gabT*), which was 2.1-fold upregulated. No significant result was obtained for the regulator RL0103 (*gabR*). RL0101 encodes a succinate semialdehyde dehydrogenase and RL0102 encodes a GABA aminotransferase. pRL100134-pRL100136 (*aatKLM*) were 10.5-fold, 17.9-fold and 20.5-fold upregulated, and their regulator pRL100133 was 4.6-fold upregulated. pRL100134 encodes another succinate semialdehyde dehydrogenase, pRL100135 a 4-hydroxybutarate dehydrogenase and pRL100136 a lactonase. The *A. tumefaciens* orthologues of pRL100134-pRL100136, *attKLM* (*blcABC*), are expressed in response to GABA and *attM* mediates inactivation of an AHL signal involved in quorum sensing to modify bacterial behaviour in response to GABA signalling by plants [331], suggesting that Rlv3841 may also be recognising and responding to GABA as an interkingdom signal. pRL120044 (*gabD6*) encodes a succinate semialdehyde dehydrogenase and was 3.9-fold upregulated, and pRL120603 (*gabD3*) was 4.2-fold upregulated. Of the final two succinate semialdehyde dehydrogenase genes, pRL120628 (*gabD5*) was 2.4-fold upregulated, and pRL100252 (*gabD4*) was 2.0-fold upregulated. GABA is a key component of plant root exudates [332], [333] and can affect the microbiome assembly [333]. It is secreted by early pea roots, and present in high concentrations in nodules, but is not present in the root exudates of older pea roots [238]. GABA metabolism is clearly significant during primary attachment to plant roots, as every succinate semialdehyde dehydrogenase was to some extent upregulated, along with genes encoding other enzymes in the pathway for GABA catabolism, suggesting that this is part of Rlv3841 recognition of pea root exudates and establishment in the microbiome.

4.2.4.6 Fatty acid metabolism is upregulated during attachment.

Genes involved in fatty acid beta-oxidation (GO:0006635) were significantly upregulated during pea root attachment. RL0505 (*fadD*) encodes a long-chain fatty acid-CoA ligase and was 16.2-fold upregulated. The operon RL0592-RL0594 encodes a thiolase, a cupin domain-containing protein, and a 3-hydroxyacyl-CoA dehydrogenase that were 25.0-fold, 29.7-fold and 31.0-fold upregulated. RL0965-RL0966 are predicted to be expressed as an operon and encode an acyl-CoA dehydrogenase and an enoyl-CoA hydratase that are 7.6- and 7.4-fold upregulated respectively. The operon RL1820-RL1823 encodes a LacI family transcriptional regulator, a CoA transferase, an enoyl-CoA hydratase and an aldehyde dehydrogenase that were 48.1-, 36.3-, 48.8- and 38.1-fold upregulated. In many prokaryotes, including *E. coli*, *P. aeruginosa* and *Haloferax mediterranei*, enoyl-CoA hydratases are involved in synthesis of polyhydroxyalkanoate, which can be used as an energy store under nutrient-starved conditions, or in resisting saline stress [334], [335]. However, this is not known to be the case in *Rhizobium*. RL4621 (*phaA*) encodes an acetyl-CoA acetyltransferase and was 4.0-fold upregulated during attachment. The acetyl-CoA reductase RL4620 (*phaB*) was 2.6-fold upregulated. The *phaBA* gene products are known to play a role in PHB synthesis in rhizobia [336], suggesting that carbon storage as PHB is upregulated during attachment to the pea rhizoplane.

4.2.4.7 Glycogen biosynthesis is upregulated during attachment.

Genes involved in glycogen biosynthesis (GO:0005978) were upregulated during root attachment. RL4114-RL4119 (*glgPBCA*, *pgm*, *glgX*) are expressed as an operon and RL4114-RL4117 were all significantly upregulated (6.8-fold, 9.0-fold, 6.6-fold, 7.2-fold) and RL4118 was 3.6-fold upregulated, with no significant result being seen for RL4119. *glgPBCA* and *glgX* are not required for bacteroid development or for bacterial persistence in nodules [169], but glycogen is thought to be the main carbon storage compound in pea nodules [269]. RL4114-RL4119 are expressed in Rlv3841 bacteroids, being between 1.8- and 4.0-fold upregulated in the bacteroid compared with free-living cultures [219]. However, they are not expressed in the pea rhizosphere [205] or during pea root colonisation.

4.2.5 Rlv3841 genes downregulated only in attachment

4.2.5.1 Rlv3841 cells are in stationary phase directly following primary attachment to pea roots.

Many of the changes in gene expression seen during attachment were those to be expected from a stationary culture that has not yet begun to divide at the plant root surface or form biofilms. Both operons encoding subunits of ATP synthase were downregulated. RL4405-RL4410 (*atpCDGAH*) was between 5- and 9-fold downregulated, while RL0924-RL0928 (*atpIBEG2F*) was between 2- and 7-fold downregulated, with the exception of RL0924, for which no significant result was obtained. Components of ribosomal subunits were also significantly downregulated, with RL0267-RL0268 (*rpmI*, *rplT*), RL1554-RL1555 (*rpsR*, *rpsF*), RL1672-RL1673 (*rpsI, rpsM*), RL1765 (*rplL*), RL1769-RL1770 (*rpsL*, *rpsG*), RL1773-RL1778 (*rpsJ*, *rplC*, *rplD*, *rplW*, *rplB*, *rpsS*), RL1783-RL1789 (*rpsQ*, *rplN*, *rplX*, *rplE*, *rpsN*, *rpsH*, *rplF*), RL1797 (*rpsK*), RL2624 (*rpsD*), and RL4552 (*rplS*) being more than 4-fold downregulated. Other genes involved in translation were also downregulated, including elongation factors RL1757 (*tufB1*), RL1771-RL1772 (*fus*, *tufB2*) and RL4600 (*efp*), peptide deformylase RL0430 (*def1*), peptide chain release factor RL0347 (*prfC*), and synthesis of the majority of tRNAs. Of the 52 tRNA-encoding genes, 13 (25%) were more than 4-fold downregulated during root attachment. The operon for biosynthesis (GO:0008616) of queuosine, which is required for tRNA modification, was also downregulated during attachment. RL4254-RL4255 (*exsBC*) were 4.1- and 10.3-fold downregulated respectively, and RL4256 (*exsD*) was 2.6-fold downregulated. Regulation of nitrogen utilisation (GO:0006808) was significantly downregulated, with RL0032 (*npr*) being 5.2-fold downregulated and RL4565 (*glnK*) being 19.7-fold downregulated. This may be an artefact of harvesting conditions, as the cells making up the attachment sample were collected from the rhizoplane after six hours in high cell density (10^7 cfu mL^{-1}) while those making up the colonisation sample were collected following seven days of growth on the rhizoplane. However, it also seems likely that directly following initial attachment

to the rhizoplane, rhizobia would not yet have begun to benefit from the nutrient density found there and therefore might well repress expression of functions involved in cell growth and division.

4.2.5.2 Iron transport and assimilation are downregulated during attachment of Rlv3841 to pea roots.

The ECF sigma factor that controls the response of Rlv3841 to iron limitation, *rpoI* (pRL120319), is the only ECF sigma factor whose expression was downregulated during either attachment to or colonisation of pea roots, being 4.0-fold reduced during attachment compared with free-living culture. Expression of several iron transporters was also repressed during attachment, including expression of the siderophore-dependent Fe (III) importers encoded by RL2713-RL2715 (*irp6A6B6C*), whose components were 42.7-, 28.1- and 9.2-fold downregulated and by RL3050-RL3052, whose components were 2.6-, 5.0- and 5.1-fold downregulated and RL3696-RL3700 (*hmuPSTUV*), whose components were 13.5-, 5.0-, 5.9-, 9.1- and 5.9-fold downregulated. Expression of the Fe (III) ion transporter encoded by RL4582-RL4584 (*fbpABC*) was also downregulated, with the three genes being 8.0-, 5.3- and 10.6-fold downregulated. Interestingly, RL2713-RL2715 and RL3696-RL3700 are over 5-fold upregulated in an Rlv3841 mutant in RL4145 which shows impaired root attachment, suggesting that the attachment phenotype exhibited by this mutant may be in part explained by its lack of sensitivity to iron availability. RL2602, which is annotated as a low affinity Fe^{2+} transporter, was 9.3-fold upregulated during attachment and 19.5-fold upregulated during colonisation, suggesting that the downregulation of Fe (III) transport genes may be partly due to a switch to importing Fe (II), which is also thought to be the main form of iron that is provided to nodules by plants [337]. The *P. syringae* ECF sigma factor *pvdS* controls biosynthesis of the siderophore pyoverdine and, like *rpoI*, is induced in response to iron limitation. Interestingly, *pvdS* is suppressed by plant immune responses during colonisation of *Arabidopsis* leaves, while constitutive expression of *pvdS* results in improved growth on plant hosts [338]. Regulation by the plant immune system could therefore be playing a role in *rpoI* suppression during Rlv3841 root attachment, as well as changing iron availability.

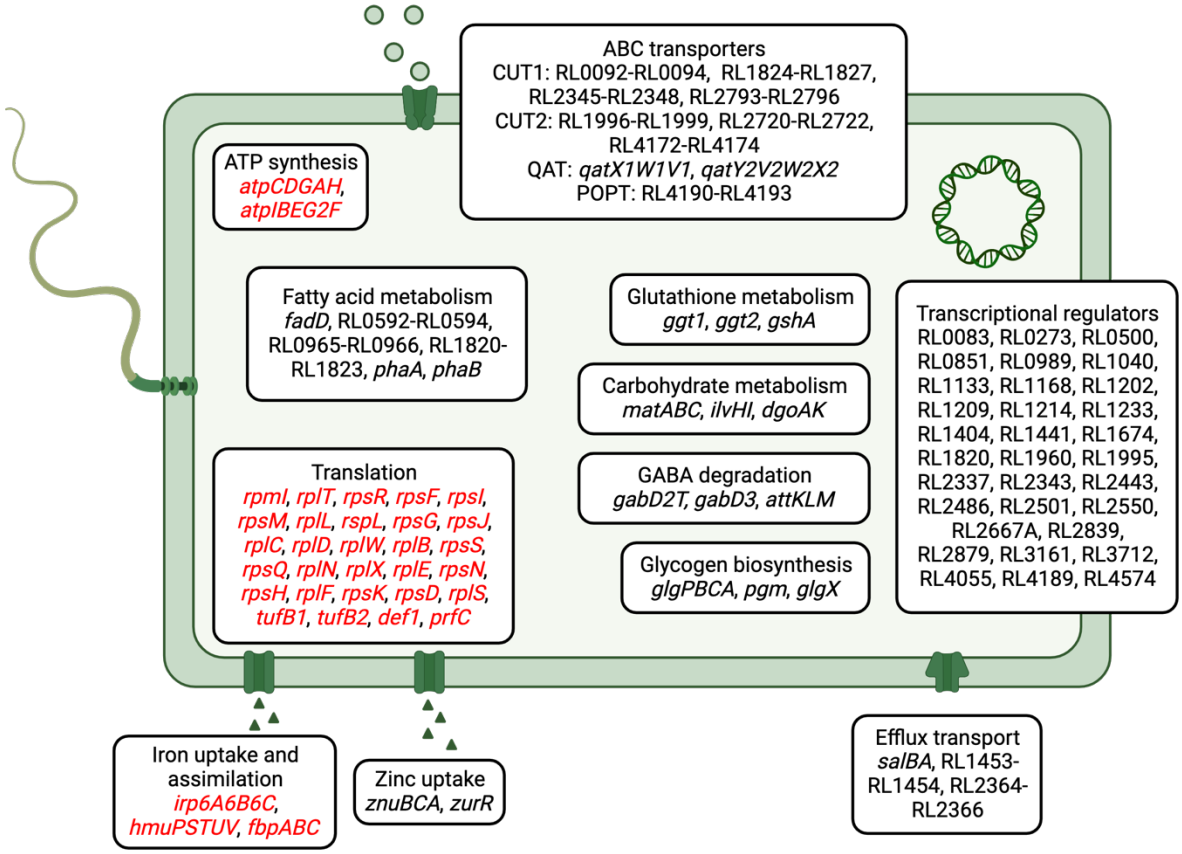


Figure 4.3: Genes uniquely differentially regulated during attachment to the pea rhizoplane. *Genes shown in black were significantly upregulated; those shown in red were significantly downregulated.*

4.2.6 Rlv3841 genes upregulated only in colonisation

Genes that were differentially expressed only during colonisation are shown in Figure 4.4.

4.2.6.1 Fatty acid metabolism is significantly upregulated during colonisation.

Genes involved in fatty acid elongation (GO:0030497) were significantly upregulated during colonisation. The short-chain dehydrogenase/reductase genes RL0482A, RL2723, RL2725 and RL4235 were 8.1-fold, 17.9-fold, 16.2-fold and 12.9-fold upregulated, respectively. The oxidoreductase genes RL1041 and RL4186 were 15.7-fold and 17.3-fold upregulated. RL0482A, RL2723 and RL2725 were all induced by arabinogalactan [219]. Genes involved in fatty acid metabolism, including PHB synthesis, were also upregulated during attachment. However, a distinct set of fatty acid metabolism-related genes was upregulated during each process, suggesting that there are differences in how energy is

stored and these stores utilised during each process. This may be due to improved nutrient availability to cells living in a biofilm following several days of growth on plant roots, meaning that these cells have no need to metabolise storage polymers and are instead able to store carbon.

4.2.6.2 Inositol catabolism is significantly upregulated during pea root colonisation.

Genes associated with inositol catabolism (GO:0019310) were significantly upregulated during colonisation of pea roots by Rlv3841. The inositol transporter components *intA*, *intB* and *intC* (RL4655, RL4654 and RL4653), encoding the SBP, ATP-binding and permease components of the transporter respectively, were all significantly upregulated during colonisation, but not during attachment. *intA* was upregulated 13.0-fold, *intB* 8.8-fold and *intC* 4.0-fold. Moreover, the operon *iolCDEB* (RL1495-RL1492), all the genes of which are involved in the *myo*-inositol catabolism pathway, was also significantly upregulated in colonisation but not attachment, as was the regulator *iolR* (RL1496). Inositol has been shown to be the main carbohydrate present in *R. leguminosarum* nodules [339], as well as being observed in plant root exudates to high levels prior to nodule formation [238]. *iolD* and *iolA* mutants are significantly impaired in nodulation competitiveness when inoculated onto pea plants in competition with wild type, although they can grow just as well in the rhizosphere as wild type [235]. It therefore appears that inositol uptake and metabolism only become important during colonisation and later nodule formation, that are the more irreversible stages of symbiosis after primary attachment.

4.2.7 Rlv3841 genes downregulated only in colonisation

Expression of some ABC transporters is significantly downregulated during colonisation but not attachment. Components of the PepT ABC transporter encoded by RL0185-RL0188 were significantly downregulated during colonisation. RL0187 and RL0188, which encode the permease and ATP-binding components respectively, were over four-fold downregulated, while RL0186 was 3.6-fold downregulated. Expression of the *S. meliloti* homologue of the SBP component is induced by taurine, valine and isoleucine [236], while the Rlv3841 SBP component RL0185 is downregulated in bacteroids

[219]. RL0185-RL0188 are predicted to be involved in microcin transport (GO:0042884). All four components of the CUT1 ABC transporter RL4249-RL4252 were significantly downregulated. The *S. meliloti* homologue of the SBP component, SMc04396, is induced by dextrin [236]. All three components of the CUT2 ABC transporter RL3624-RL3626 were significantly downregulated (37.8-fold, 15.3-fold, 15.3-fold). The *S. meliloti* homologue of the SBP, SMb20902, is induced by mannose and lyxose [236]. As rhizobial colonies begin to form at the root surface, the transcriptome changes to reflect the differing nutrient sources that are being used for growth.

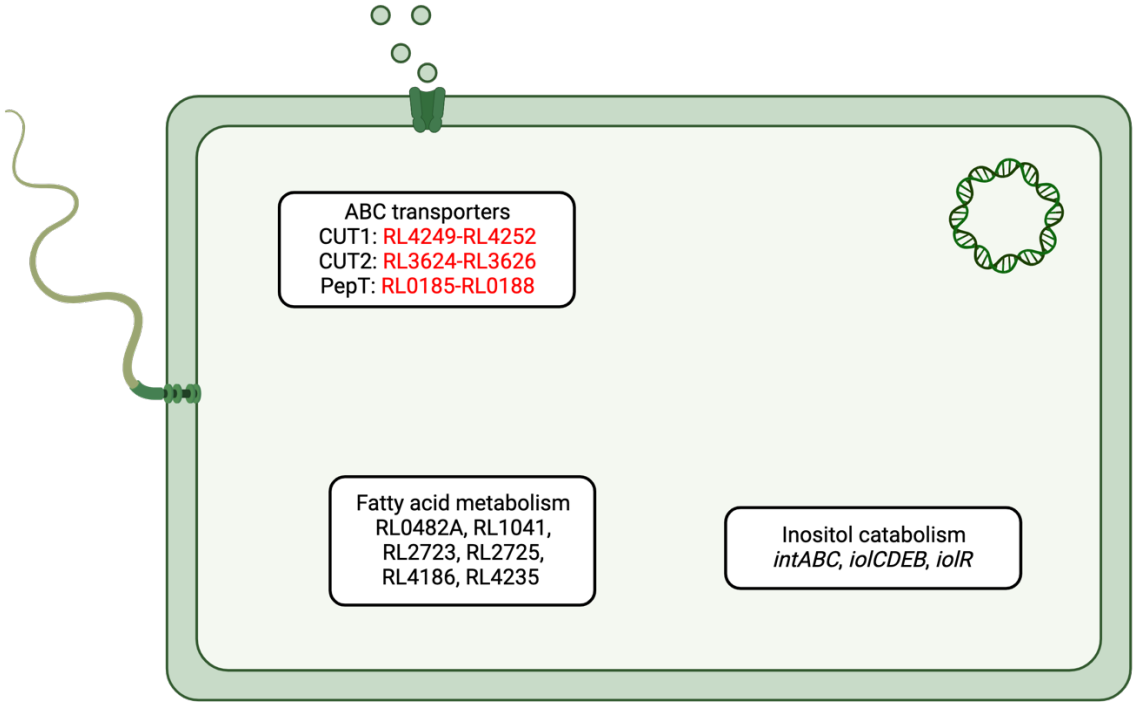


Figure 4.4: Genes uniquely differentially regulated during colonisation of the pea rhizoplane. *Genes shown in black were significantly upregulated; those shown in red were significantly downregulated.*

4.3 Summary

Overall, there is a substantial amount of overlap between the significantly differentially regulated genes in the attachment and colonisation transcriptomes in Rlv3841. Around 50% of differentially regulated genes were affected during both stages of interaction with the plant host. In particular, a

significant number of both housekeeping and ECF sigma factors were upregulated during both attachment and colonisation. This significant change in expression of transcriptional regulators may partially account for the substantial number of other differentially transcribed genes during attachment to and colonisation of the pea rhizoplane. Key functions upregulated during both attachment and colonisation included transmembrane transport, with a wide range of ABC transporters increasing the uptake of a variety of substances in response to the changing environment and lifestyle of rhizobial cells [236], [314]. Genes related to DNA repair functions, aldo/keto reductase activity and cell surface characteristics in the form of BA14K-like proteins were also significantly upregulated during both processes. Expression of genes involved in motility and chemotaxis was significantly reduced during both attachment and colonisation, confirming the switch between motile, free-living cells and irreversible attachment to the rhizoplane [254]. Sulphur uptake was also affected during both stages of interaction, with arylsulfatases being significantly downregulated but sulphur transporters upregulated, in common with other documented plant-microbe interactions [320], [321]. A key difference between the attachment and colonisation transcriptomes is that cells harvested directly following primary attachment have more traits of stationary phase culture than those harvested following a week of colonisation of the root surface, showing downregulation of genes involved in ATP synthesis and translation. Genes involved in iron uptake and assimilation were also significantly downregulated during attachment. Another key difference between attachment and colonisation was in the types of compounds being catabolised. Genes involved in breakdown of malonate, GABA and fatty acids were significantly upregulated in attachment, while inositol catabolism was significantly upregulated during colonisation. This suggests that different compounds are used as a carbon source in different stages of rhizoplane colonisation, perhaps because due to different levels of availability in root exudates versus at the root surface. [331][205][266]

This chapter shows the importance of investigating transcriptomic changes during each individual phase of root infection, and the power of a method that allows cells that are attached to the root surface to be separated from both contaminating plant matter and cells in the surrounding growth

media. These datasets have the potential to be an important resource for the understanding of the genetic basis of root colonisation by PGPRs.

5 The role of the *ecfE* operon in pea rhizoplane attachment and colonisation

5.1 Introduction

The soil surrounding plant roots is colonised by a huge number of diverse microbial species, with a gram of rhizosphere soil containing up to a million fungal cells and a billion bacterial cells [68]. All these organisms compete to inhabit this area in order to benefit from plant root exudates, which are rich in carbon, accounting for up to 30% of a plant's photosynthates [67]. Plant growth-promoting rhizobacteria (PGPRs), including diazotrophic rhizobia living in symbiosis with legumes that can contribute up to 40 million tonnes of fixed nitrogen per year to agricultural systems [21], must therefore compete with a huge range of other microorganisms to colonise the rhizosphere and root surface and confer their benefits on their hosts. This makes understanding how colonisation happens of vital importance in furthering our knowledge of the plant microbiome and how it can be manipulated to move towards more sustainable and higher-yielding agriculture. As described in Chapter 4, colonisation of plant roots takes place in four stages. The first is chemotaxis towards the roots, during which bacteria move along a concentration gradient of root exudates, which contain a range of chemoattractants including organic acids, phenolics, amino acids and phytosiderophores. In the *Rhizobium*-legume symbiosis, plant-secreted flavonoids are a particularly important component of root exudates that is involved in the chemical signalling between both partners to allow the establishment of the symbiosis [112]. Rhizobia travel chemotactically along gradients, allowing them to find and attach to root hair tips, the optimum position for exposure to root exudates and plant surface components that mediate successful further interactions [340], [341]. Cell surface components such as flagella and type IV pili act as propellors that bacteria can use to actively migrate along gradients. Mutants lacking flagella and type IV pili attach to surfaces less efficiently than wild type bacteria [342], [343], [344]. Active propulsion by flagella and type IV pili overcomes repulsive electrostatic forces and other barriers that might otherwise prevent cells coming close enough to roots to attach [112], [341]. Once bacterial cells reach the root surface, primary attachment takes place by the formation of weak, reversible bonds. This includes non-specific bonds such as Van der Waals, electrostatic and hydrophobic forces between the negatively charged bacterial cell envelope and the

root [112], [344]. There are also more specific, but still reversible, interactions that take place during primary attachment, mediated by surface proteins and polysaccharides, and surface structures such as flagella. One such interaction that contributes to specificity in binding of rhizobia to legume roots is the interaction between the rhizobial surface polysaccharide glucomannan and pea and vetch lectins [127]. Following primary attachment, secondary attachment is characterised by the formation of irreversible bonds between the bacterial cell surface and the root. Unlike primary attachment, which is mediated by constitutively present cell surface components, secondary attachment happens following primary attachment-induced biosynthesis and secretion of attachment factors [345]. In particular, the synthesis of extracellular fibrils leads to the accumulation and aggregation of bacterial cells around plant root hairs. Extracellular polysaccharide (EPS) also plays a significant role in secondary attachment, being a major component of bacterial cell surfaces and highly specific to strains and species [24]. Rhizobial proteins with specific domains such as Rhizobium-adhering (Rap) domains have also been shown to mediate secondary attachment [285]. Once rhizobia have attached to the root surface, they can begin to grow and form biofilms, the final step in successful colonisation. The successful colonisation of plant roots by *Rhizobium* species is important both as a first step in formation of nitrogen-fixing nodules on legume roots and as a way for *Rhizobium* to act as a growth-promoting commensal on the roots of non-host plants including agriculturally significant cereal crops.

Following on from the work presented in Chapters 3 and 4, I identified a set of three genes, RL3234-RL3236, that were upregulated during colonisation of both the barley and pea rhizoplanes, and attachment to the pea rhizoplane. They are also upregulated in the rhizospheres of pea, alfalfa and sugar beet [205]. These three genes encode a lipoprotein, *lppE*, an extracytoplasmic sigma factor, *ecfE*, and an anti-sigma factor, *asfE*. The three genes were not induced in any other microarray studies, including in pea bacteroids, apart from being weakly upregulated in cells grown in free-living culture with phenylalanine as a nitrogen source [205], [219]. Extracytoplasmic sigma and anti-sigma factors generally act as pairs that regulate transcription in response to a signal from outside the cell [346]. In Gram-negative bacteria such as *Rhizobium*, anti-sigma factors are anchored in the inner membrane,

with a cytoplasmic and periplasmic domain. The cytoplasmic domain interacts with the sigma factor, holding it at the membrane and preventing it from carrying out its role as a transcriptional activator. In response to changes in external conditions, or to a specific external signal, factors in the periplasm interact with the periplasmic domain of the anti-sigma factor and cause a conformational change so that the sigma factor is released into the cytoplasm. There, it activates RNA polymerase to transcribe its targets [347]. A well-studied example of regulation by a sigma/anti-sigma factor pair is σ^E and RseA in *Escherichia coli*. Under non-stressed conditions, RseA binds to σ^E and anchors it at the cell membrane [348]. In response to protein misfolding in the cell envelope, RseA is cleaved by two signalling proteins, DegS and RseP, releasing σ^E and allowing it to activate transcription of its target genes [349], [350], [351]. Extracytoplasmic sigma factors have been implicated in response to heat stress, saline stress, iron limitation and oxidative stress [228], [286], [301], [302], [352] and are clearly important in governing bacterial responses to changing environments and lifestyles as are found during the switch from free-living soil bacteria to root colonisation. Significant upregulation of an extracytoplasmic sigma factor during colonisation of plant roots, and particularly during attachment to and colonisation of host legume roots, in the model diazotroph Rlv3841, warrants further investigation. The processes of attachment to and colonisation of plant roots by *Rhizobium* species are less well characterised than the rest of the symbiosis, as research efforts have mainly focussed on nodule development and nitrogen fixation [48]. However, attachment and colonisation are key determinants of competitive success and understanding the genetic factors that govern these processes will be important in furthering knowledge of the *Rhizobium*-legume symbiosis and its agricultural applications. In this chapter, I show that the *lppE-ecfE-asfE* are expressed in bacteria attaching to and colonising the root elongation zone, the region of the root where rapid cell growth occurs, making new surface area available for colonising microorganisms. The *ecfE* system is therefore clearly important in bacterial recognition of a key area for colonisation and this study will demonstrate how this process takes place.

5.2 Results

5.2.1 *lppE*, *ecfE* and *asfE* are expressed as an operon from a promoter upstream of the *lppE* coding sequence

RL3234 (*lppE*), RL3235 (*ecfE*) and RL3236 (*asfE*) are all upregulated during colonisation of the pea and barley rhizoplanes. This suggests that they may be expressed as an operon from a single promoter upstream of *lppE*. Computational operon prediction also suggests that *lppE-ecfE-asfE* are expressed as an operon [353]. In order to verify that this is the case, promoter fusions were generated by inserting several predicted promoter regions in front of the *luxCDABE* luminescence cassette of plasmid pIJ11268 [142] (Fig. 5.1A). A secondary transcriptional start site is also predicted inside the coding region of *lppE* [Poole lab, unpublished data] and three fusions were generated that incorporated this possible secondary promoter region. Expression of all three genes is known to be weakly induced during growth on phenylalanine as a nitrogen source [205], [219] and this property was used to assay for promoter activity in each of the *lux* fusion assemblies. Rlv3841 wildtype derivative strains harbouring the different *lux* promoter fusions were grown in free-living liquid culture supplemented with 5 mM phenylalanine and the production of luminescence was measured during the exponential growth phase. Both the upstream and intragenic regions of *lppE* confirmed the presence of promoter regions able to drive expression of *lux* cassette individually (Fig. 5.1B), with the strain harbouring a *lux* fusion to fragment 6, containing both regions, exhibiting higher luminescence than any strain containing a fusion to only one of these two regions. Luminescence was approximately two-fold higher in those strains harbouring *lux* fusions to the promoter region upstream of *lppE*, the first gene in the putative operon, than in the strains harbouring fusions to the promoter region within the *lppE* coding sequence, confirming that these three genes form an operon, with the primary promoter located in the upstream region and the one within the internal region representing a secondary, weaker promoter that may drive expression of *ecfE* and *asfE*. All further expression studies were therefore carried out using the primary promoter region (P*lppE,* fragment 1) only, since *lppE* was shown to be

upregulated during root colonisation in addition to *ecfE* and *asfE*, implying that expression of the operon from the primary promoter region is significant during colonisation of plant roots.

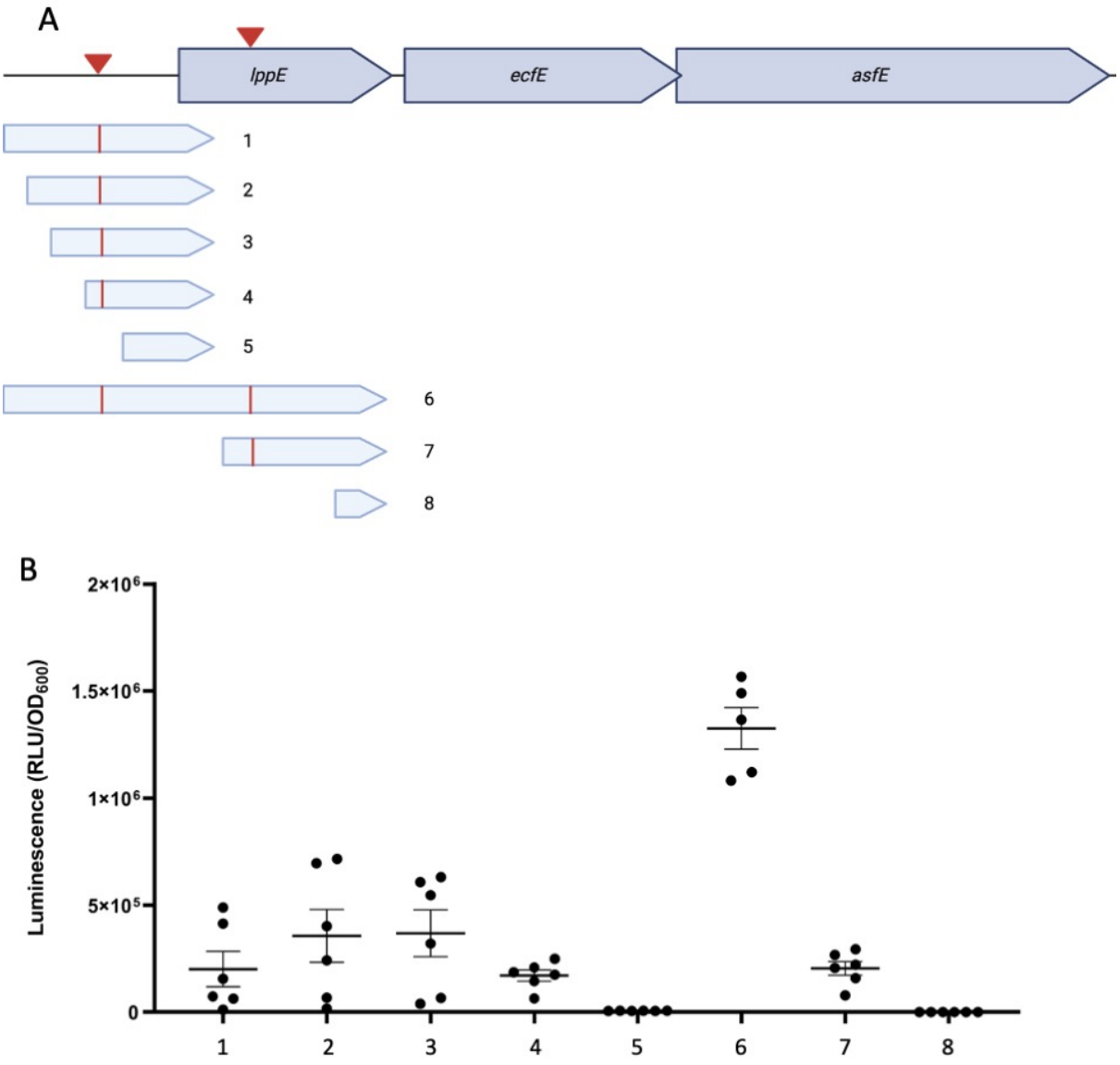


Figure 5.1: Operonic expression and location of the promoter region of the ecfE operon. *A, Genetic map of Rlv3841 RL3234-36 (ecfE operon genes) showing fragments tested for promoter activity. Red arrows ▼ indicate predicted transcription start sites. B, Luminescence driven by upstream fragments fused to luxCDABE genes following induction with 5 mM phenylalanine. All assays performed on free-living cultures. Values are the mean ± SEM expressed as Relative luminescence units/OD_{600}. Biological replicates n=3.*

5.2.2 The *lppE* promoter region is expressed in *Rhizobium* cells found at plant root elongation zones

Microarray data showed expression of all three genes in the operon to be upregulated during colonisation of a range of plant species [205]. The *lux* fusion to the primary promoter region upstream of all three genes (P*lppE,* fragment 1) was used to verify this finding. A suspension of Rlv3841 harbouring this fusion in 0.1% agarose was inoculated onto the roots of 5-day-old pea seedlings grown on square plates and images taken over 16 hours show P*lppE*-driven luminescence. These images (Fig. 5.2A, Supplementary videos 1-2) show that detectable luminescence driven by expression from P*lppE* begins approximately six hours post inoculation. Images of peas taken 7 days post inoculation show that P*lppE* expression remains active over this time period (Fig. 5.2B). Even more striking than this temporal expression pattern, P*lppE* has a distinct spatial expression pattern seen during both short-term and longer-term *in planta* induction assays. Luminescence driven by the P*lppE* promoter is seen solely at the ends of the roots. Indeed, higher magnification shows that this luminescence is actually restricted just above the root tips, around the root elongation zone (Fig. 5.2B). The root elongation zone is the region where rapid cell expansion occurs, with cells growing eight times more quickly than in the root apical meristem [354].

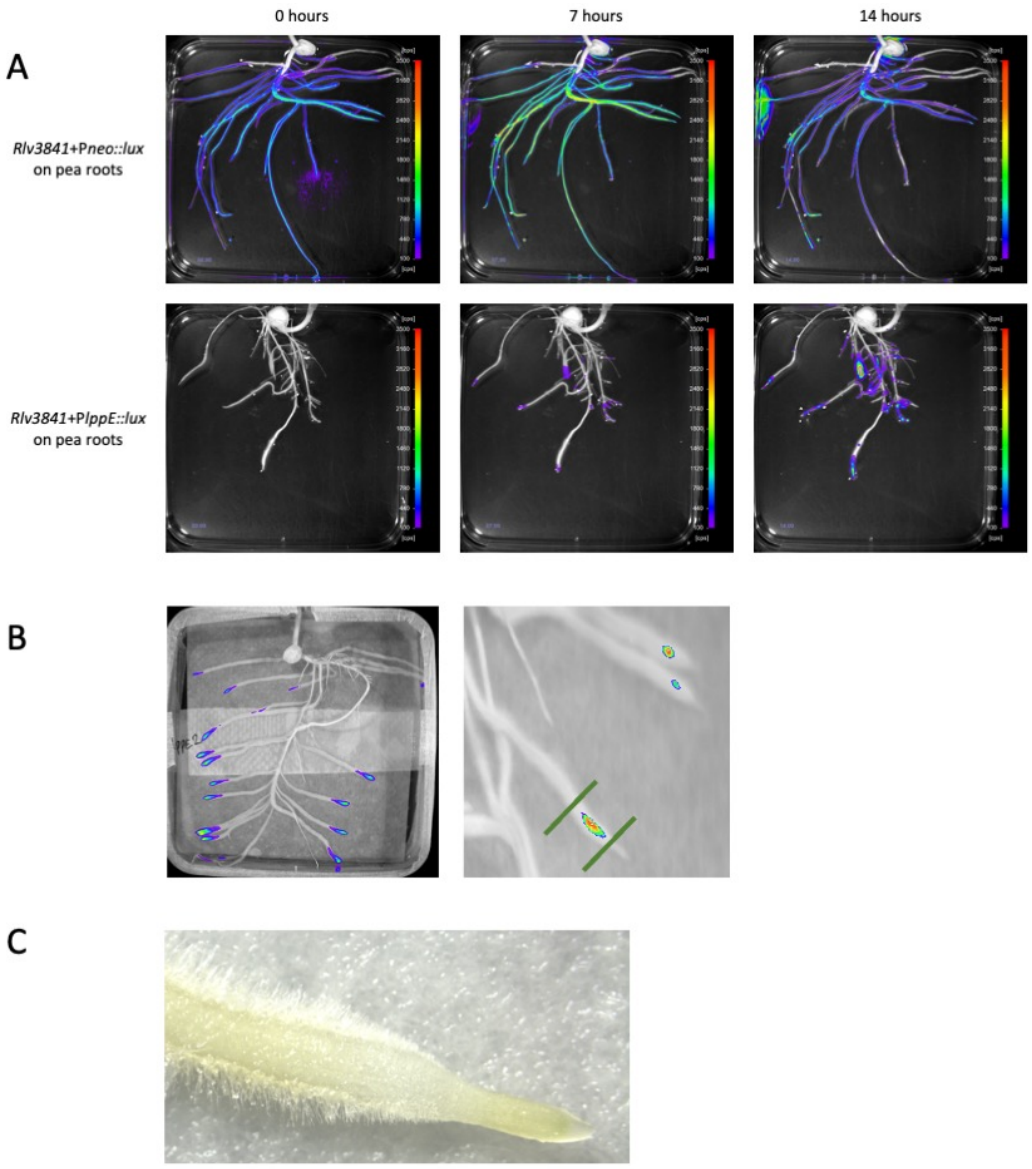


Figure 5.2: Luminescence driven by expression from PlppE in Rlv3841 inoculated onto pea roots. *A, Stills from timelapses taken at 0h, 7h and 14h post inoculation. Rlv3841 expressing constitutive Pneo::lux as control. B, Still images taken 7 dpi. Images taken using NightOWL (Berthold) and processed using IndiGO. C, REZ of a pea root. Image taken using stereo microscope (LEICA M165 FC).*

5.2.3 Transcription from the primary promoter P*lppE* is controlled by EcfE and AsfE

Pairs of sigma and anti-sigma factors often control expression from their own promoters [228]. To investigate whether this was the case for EcfE and AsfE, I used the same P*lppE*::*lux* promoter fusion, expressed in Rlv3841 mutants in each of the three genes in the operon (Δ*lppE*, Δ*ecfE* and Ω*asfE*).

Strains were grown in liquid cultures supplemented with 5 mM phenylalanine to induce P*lppE* activity and luminescence was measured to determine relative transcription. EcfE induced transcription of its own promoter, as the Δ*ecfE* strain had 77-fold reduced luminescence compared with the wild type (Fig. 5.3A). AsfE is a transcriptional repressor of its own promoter, with P*lppE*-driven luminescence increased in Ω*asfE* 270-fold compared with wild type levels. In order to verify that this transcriptional regulation also takes place during colonisation of pea roots, these induction assays were repeated *in planta*, inoculating the mutant strains harbouring P*lppE*::*lux* fusions onto pea plants grown on square plates and imaging luminescence at 7 days post inoculation. The same trends are seen in P*lppE* expression during colonisation of pea roots as in laboratory cultures, with EcfE activating expression of P*lppE* and AsfE preventing it (Fig. 5.3B, C). Computational predictions of the membrane topology of AsfE were carried out using the TOPCONS web server [196]. The cytosolic domain is predicted to be the N terminal region, and the periplasmic domain the C terminal region, with a 21-amino acid transmembrane domain (Fig. 5.3D). Modelling of the hypothesised interaction between EcfE and AsfE using Alphafold 2 [195] suggests that the interface is between EcfE and the AsfE N terminal region, placing the interaction in the cytosol as in other sigma factor/anti-sigma factor interactions (Fig. 5.3E).

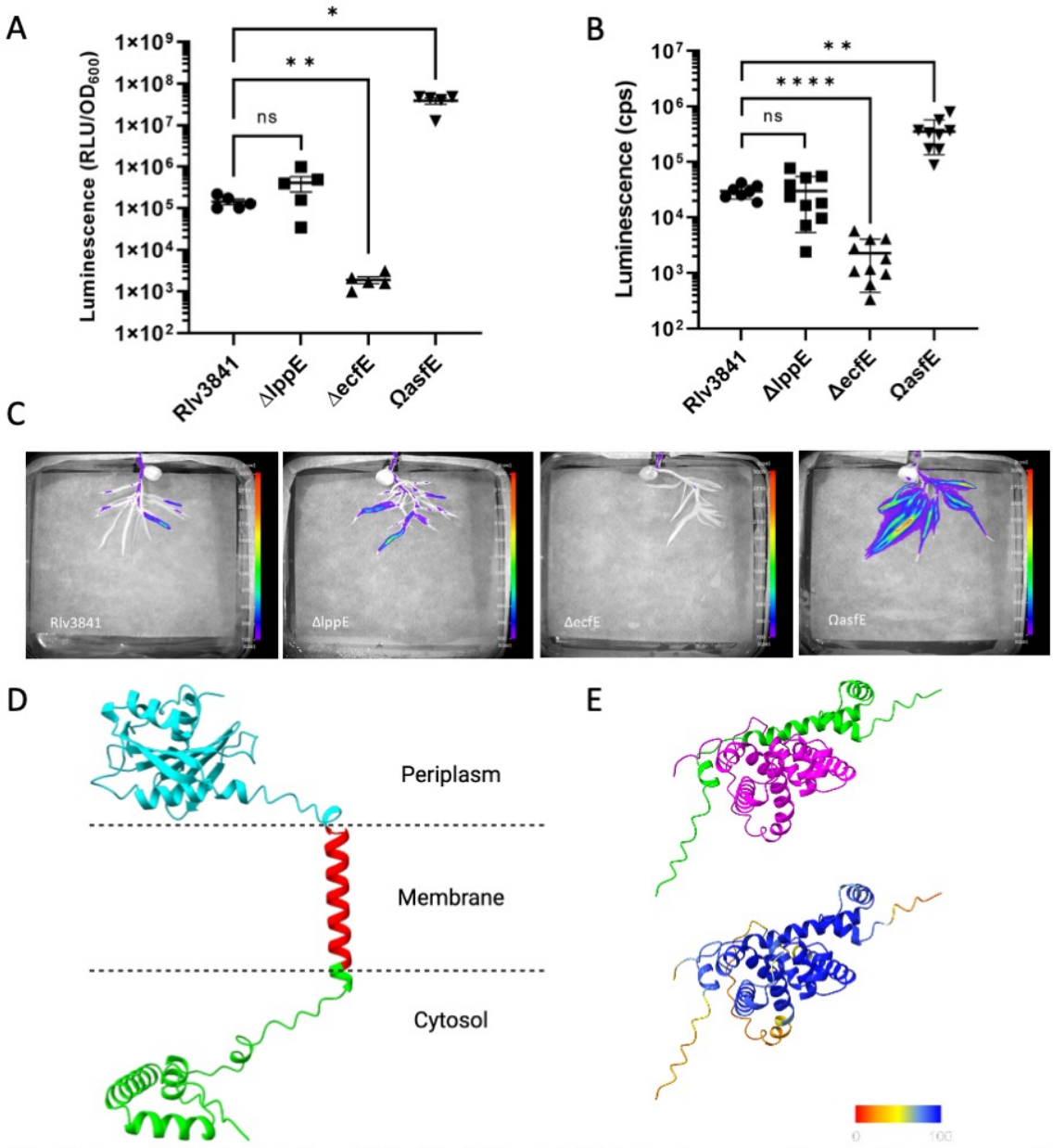


Figure 5.3: Transcriptional regulation of P*lppE *by EcfE and AsfE. *A, Luminescence driven by PlppE expressed in Rlv3841 wild type and mutants in each ecfE operon gene. Strains grown on agar containing 5 mM phenylalanine for three days, then resuspended before measurement. Values are mean ± SEM (Relative luminescence units/OD_{600}). Biological replicates n=5. Significance calculated using Brown-Forsythe and Welch ANOVA tests and Dunnett's multiple comparisons test. B, Luminescence driven by PlppE expressed in Rlv3841 wild type and mutants in each ecfE operon gene. Assays carried out on strains inoculated onto pea roots then grown for 7 days. Values are mean ± SEM (Luminescence counts per second). Biological replicates n=7. Significance calculated using unpaired t*

*tests. ns=p>0.05; *=p≤0.05; **=p≤0.01; ***=p≤0.001; ****=p≤0.0001. C, Representative images of plant roots inoculated with each strain taken using NightOWL 7 dpi. D, Predicted membrane topology of AsfE generated using TOPCONS. E, Predicted interaction between EcfE and the cytoplasmic domain of AsfE. AsfE coloured in green and EcfE in pink (top); interaction coloured by confidence (pLDDT) (bottom). Protein structure and interaction predictions generated using AlphaFold Multimer via ColabFold with MMSeqs2.*

A putative EcfE sigma factor binding site was identified upstream of the *lppE* coding sequence, in the primary promoter region identified in 5.2.1. The same operon structure was identified in 8 other *Rhizobium* and *Sinorhizobium* species. A region including 100 nt of the 5' untranslated region (UTR) upstream of *lppE* and the first 30 nt of the coding sequence was taken from these 8 species and aligned (Fig. 5.4). The sigma factor binding site, GGAATAA-N_{14}-GTGTCT (3384348-3384373), was conserved in all 8 species studied.

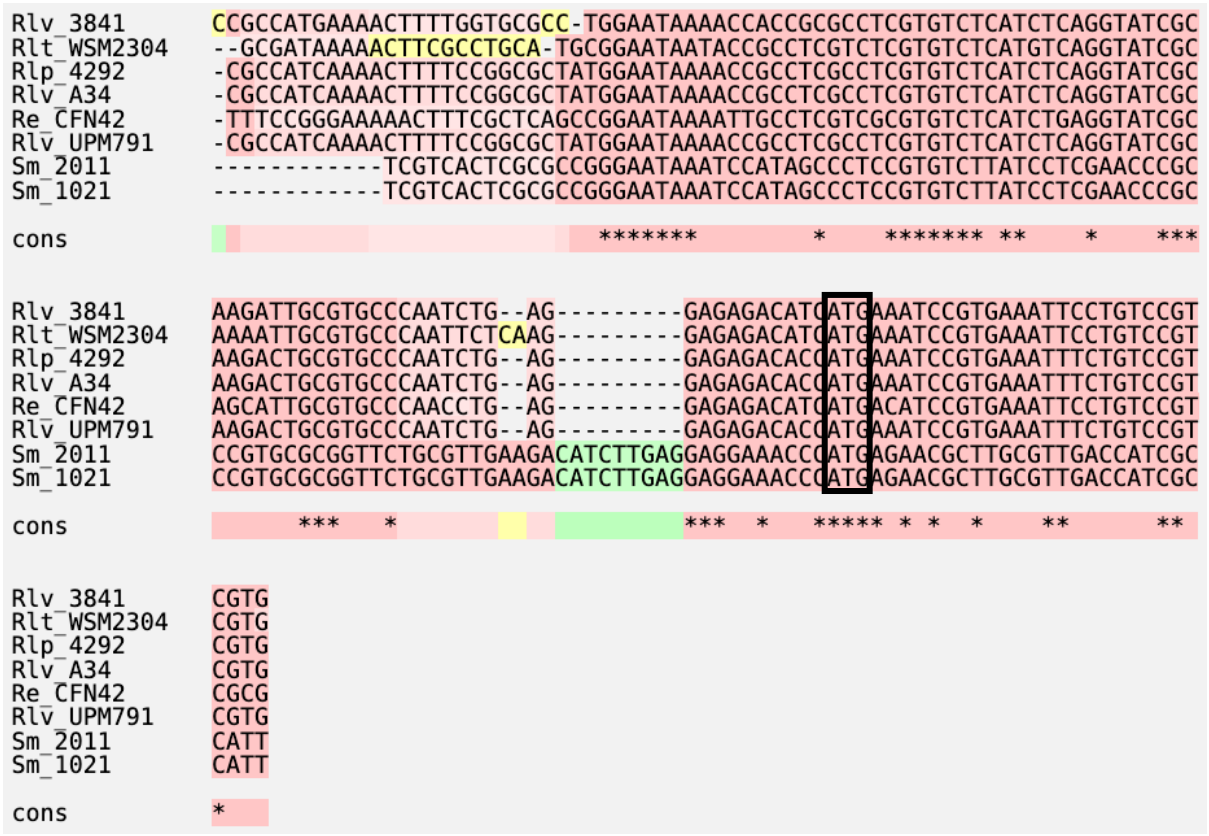


Figure 5.4: Conserved putative EcfE binding region upstream of the* lppE *coding sequence. *Multiple sequence alignment showing the* lppE *CDS and 100 nt of 5' UTR in* Rlv3841 *and 7 other* Rhizobium *and* Sinorhizobium *strains. Alignment generated using Pro-Coffee, a software for alignment of homologous promoter regions* [192], [194]*. Conserved region starred. Start codon outlined in black.*

5.2.4 EcfE has a broad range of transcriptional targets

ECF sigma factors often have a broad range of transcriptional targets beyond their own promoters, and having established that EcfE activates transcription of its own operon, I used RNA sequencing of a Ω*asfE* mutant to search for additional transcriptional targets. Since AsfE represses expression of *ecfE* and prevents the EcfE protein from initiating transcription of its targets, I predicted that in Ω*asfE*, EcfE would be overexpressed and free to activate expression of its target genes. Thus, any genes upregulated (>4-fold increase in expression, padj<0.05) in the *asfE* mutant compared to wild type may be transcriptional targets of EcfE. As expected from the expression data obtained using lux fusions above, *lppE* and *ecfE* were upregulated 194.0-fold and 157.6-fold respectively in the *asfE* mutant

compared with the wild type, suggesting that the major transcriptional target of EcfE is its own promoter. Five further genes were upregulated in Ω*asfE* relative to the wild type: RL1136, RL1137, RL4584 (*fbpA*), RL1802 and RL1951 (*wrbA*). RL1136 and RL1137 were 19.7-fold and 5.7-fold upregulated respectively. RL1138 and RL1139 were also 2.3-fold and 1.8-fold upregulated in Ω*asfE* relative to wild type, suggesting that the four genes may be expressed together as an EcfE-dependent operon. RL1138 and RL1139 encode another sigma factor-anti-sigma factor pair, EcfC and AsfC. RL1136 is annotated as an exported protein and consists of three alpha-helices with a DUF305 domain, a domain of unknown function that is associated with ferritin-like metalloproteins, suggesting it may be involved in iron or copper binding [310]. RL1137 is also annotated as an exported protein containing a YVTN beta-propeller repeat (IPR011964), which in its AlphaFold predicted structure consists of beta sheets forming a pore-like structure with a lid. RL1137 also has 90% identity to YncE family proteins in six closely related *Rhizobium* strains. *fbpA* was 5.7-fold upregulated in Ω*asfE* and encodes the solute-binding protein of the iron (III) importer FbpABC. RL1802 was 4.3-fold upregulated in Ω*asfE*. It encodes the catalytic subunit MsrP, part of the MsrPQ enzymatic complex that catalyses the reduction of methionine sulfoxide residues on periplasmic proteins using quinones, to protect them from oxidative stress. The system localises to the inner membrane of the periplasm and is important in maintaining envelope integrity [310]. RL1801, encoding the haem-binding subunit MsrQ, was 3.5-fold upregulated in Ω*asfE* compared with the wild type, suggesting that both genes may be coregulated by EcfE. *wrbA*, encoding an NADH quinone oxidoreductase flavoprotein, was 4.0-fold upregulated in Ω*asfE* and is also involved in response to oxidative stress.

Table 5.1: Upregulated genes in Ω*asfE *compared with Rlv3841 wild type. *Differential regulation expressed as fold change. Data analysed using DeSeq2 and values with adjusted p-value < 0.05 and fold change > 4.0 taken forward for further analysis.*

Locus tag/ gene name	**Differential regulation in Ω*asfE* (minimal media)/ adjusted p-value**	**Differential regulation in Ω*asfE* (iron-sufficient media)/ adjusted p-value**
RL3234/ *lppE*	194.0/ 3.5E-255	2195.0/ 0
RL3235/ *ecfE*	157.6/ 0	256/ 0
RL1136	19.7/ 9.5E-36	22.6/ 5.2E-150
RL4584/ *fbpA*	5.7/ 4.4E-23	8.0/ 1.4E-189
RL1137	5.7/ 1.0E-28	24.3/ 9.6E-178
RL1802/ *msrP*	4.3/ 1.6E-14	6.1/ 4.2E-91
RL1951/ *wrbA*	4.0/ 1.5E-6	1.0/ 0.9 (n.s.)
RL0880	4.0/ 2.6E-3	0.9/ 0.7 (n.s.)
RL1801/ *msrQ*	3.5/ 1.5E-6	5.3/ 1.6E-64
RL1800	1.6/ 1.0 (n.s.)	5.3/ 5.1E-38
RL1138/ *ecfC*	2.3/ 0.2 (n.s.)	4.9/ 3.3E-27

Searching for the putative EcfE binding site upstream of the genes in Table 5.1 revealed that it is present upstream of the RL1136 coding sequence (1215261-1215286) (Fig. 5.5), but not any of the other possible targets. This suggests that RL1136-RL1139 may be the only direct transcriptional target of EcfE, with the other genes that were upregulated in Ω*asfE* being targets of EcfC (encoded by RL1138).



Figure 5.5: The putative EcfE binding site is also found upstream of the RL1136 coding sequence. *Multiple sequence alignment showing 100 bp of the* lppE *5' UTR in Rlv3841 and seven other* Rhizobium *and* Sinorhizobium *strains, together with 100 bp of the RL1136 5' UTR from Rlv3841. The motif (GGAATAA-14$_n$-GTGTCT) is conserved except for the substitution of the final A for a G in the RL1136 5' UTR. MSA generated using Pro-Coffee, a software for alignment of homologous promoter regions* [192], [194]. *Conserved region starred.*

Analysis of RNA sequencing data also identified 6 genes that are significantly downregulated in Ω*asfE* relative to wild type (>4 fold decrease in expression, padj<0.05). Five of these genes, *vbsADL* and *vbsSO*, encode components of the siderophore vicibactin. The sixth, RL3692, encodes an uncharacterised protein with >90% sequence identity to 42 proteins found in other *Rhizobium* strains [310]. Below the cutoff threshold of a 4-fold change in expression, a number of other genes involved in iron transport and assimilation are differentially expressed, including the regulatory genes *rpoI* (3.6-

fold downregulated) and *tonB* (2.9-fold downregulated) and haem uptake genes *hmuPSTUV* (1.8-2.9-fold downregulated). ECF sigma factors such as EcfE are transcriptional activators, making it unlikely that reduced expression of these genes results from direct regulation by EcfE. However, downregulation of all these genes in Ω*asfE* suggests that regulation of these genes is carried out by a pathway that includes EcfE.

Table 5.2: Downregulated genes in Ω*asfE *compared with Rlv3841 wild type. *Differential regulation expressed as absolute fold change. Data analysed using DeSeq2 and values with adjusted p-value < 0.05 and absolute fold change < -4.0 taken forward for further analysis.*

Locus tag/ gene name	**Differential regulation in Ω*asfE* (minimal media)/ adjusted p-value**	**Differential regulation in Ω*asfE* (iron-sufficient media)/ adjusted p-value**
pRL120316/ *vbsA*	-7.0/ 2.1E-17	-1.2/ 0.6 (n.s.)
pRL120315/ *vbsS*	-6.1/ 1.8E-23	-1.3/ 0.05 (n.s.)
pRL120318/ *vbsL*	-6.1/ 1.3E-16	-1.4/ 0.04
pRL120317/ *vbsD*	-4.9/ 1.3E-10	1.1/ 0.9 (n.s.)
pRL120314/ *vbsO*	-4.3/ 1.3E-13	-1.1/ 0.5 (n.s.)
RL3692	-4.3/ 7.4E-4	-1.1/ 0.9 (n.s.)
pRL110007/ *ecfN*	-1.9/ 0.8 (n.s.)	-5.3/ 4.4E-34
RL0470	-1.0/ 1.0 (n.s.)	-4.9/ 2.0E-66
pRL120245	-0.9/ 1.0 (n.s.)	-4.6/ 3.2E-29
RL3446	-1.1/ 1.0 (n.s.)	-4.6/ 2.6E-33
RL4183	-2.5/ 2.2E-6	-4.3/ 6.5E-78

pRL110321	-1.1/ 1.0 (n.s.)	-4.3/ 8.2E-41
pRL110006/ *asfN*	-1.5/ 1.0 (n.s.)	-4.0/ 3.2E-50
pRL120342/ *hspD*	-1.9/ 6.1E-3	-4.0/ 2.3E-27

The minimal medium (UMS) used for bacterial growth during the RNA sequencing experiment was iron-limited due to the $FeSO_4 \cdot 7H_2O$ becoming rapidly oxidised to iron (III), which is insoluble and therefore not accessible to Rlv3841 growing in the media. Therefore, to further investigate the links between the *ecfE* operon and iron uptake I carried out the same transcriptomic comparison between Ω*asfE* and wild type but this time grew the cultures on media supplemented with 0.04 mM soluble iron (III) citrate and citric acid instead of iron sulfate (FMS), making it iron sufficient. Interestingly, none of the six genes that were significantly downregulated in Ω*asfE* on iron-limited media were differentially regulated on iron-sufficient media. The vicibactin synthesis genes are the most highly upregulated genes in Rlv3841 under conditions of iron limitation [Poole lab, unpublished data]. However, *fbpA*, which is also involved in iron uptake, was 8-fold upregulated in iron-sufficient conditions, showing that its expression is regulated by EcfE regardless of iron availability. Upregulation of *fbpA* in Ω*asfE* likely leads to improved iron uptake by Ω*asfE*, meaning that the *vbs* genes do not need to be upregulated in iron-limited media in the mutant like they are in the wild type. Other changes seen under iron-sufficient conditions were that some genes that were below the threshold of being considered significantly differentially regulated became significant. In particular, pRL110007 (*ecfN*) and its anti-sigma factor pRL110006 (*asfN*), which were 1.9- and 1.5-fold downregulated in Ω*asfE* under iron-limited conditions, were 5.3-fold and 4.0-fold downregulated under iron-sufficient conditions. This links the *ecfE* operon into a wider chain of cross-regulation by sigma and anti-sigma factors in response to environmental conditions including iron limitation. Finally, RL1951 (*wrbA*), which was 4.0-fold upregulated in Ω*asfE* under iron-limited conditions, did not show any differential expression under iron-sufficient conditions. Overexpression of *ecfE* caused by mutation of *asfE* clearly

causes an altered response of the cell to iron availability, probably due to improved ability to take up iron because of the increased expression of *fbpA*.

5.2.5 LppE is transported to the periplasm and indirectly affects EcfE activity

LppE is annotated as a lipoprotein, but protein sequence analysis did not find any obvious lipobox in the N-terminal region. However, it does contain a tandem repeating lipoprotein domain (IPR005297) which itself incorporates a Yx(FWY)xxD motif, a motif of unknown function that is shared among bacterial adhesins [355]. In Gram-negative bacteria, lipoproteins can be anchored in the periplasm at the inner membrane or peripherally at the outer membrane, or they can be secreted and anchored on the extracellular surface [356]. Examination of the *lppE* sequence revealed a signal peptide sequence predicted to target the protein to the periplasm. To confirm that this is the case, a fusion of the LppE signal peptide to *E. coli* alkaline phosphatase (*phoA*) was generated and expressed in Rlv3841 under control of P*tau*. Alkaline phosphatase is only active in the periplasm, meaning that fusion of the *E. coli phoA* gene to a signal peptide of interest can be used to confirm whether the signal peptide targets its protein to the periplasm [357]. Alkaline phosphatase activity in this strain grown in the presence and absence of taurine confirmed that the LppE signal peptide targets the protein to the periplasm (Fig. 5.6A). The lack of a lipobox suggests that rather than being truly anchored in the membrane, LppE may be localised to a membrane by a hydrophobic patch that interacts with the heads of phospholipids.

Mutation of *lppE* did not increase P*lppE* activity (Fig. 5.3), but overexpression of LppE from the taurine-inducible promoter *Ptau* in pLMB509 [178] resulted in repression of P*lppE*-driven luminescence from the lux cassette in pIJ11268 in *R. leguminosarum* (Fig. 5.6B). To confirm that this is not due to a transcriptional regulatory effect of LppE itself, I generated a constitutively expressed transcriptional fusion of *lppE* to *gusA*. In order to remove any confounding transcriptional effects of EcfE and AsfE, this fusion was expressed in a mutant in all three genes in the operon (Ω*lppE*) which expressed *lppE* from P*tau* in pLMB509. Cultures were grown both with and without the addition of taurine resulting in either no expression of *lppE* or overexpression, and no significant difference in beta-glucuronidase

activity was observed (Fig. 5.6C). Therefore, overexpression of LppE does not directly alter expression from its own promoter. In order to investigate whether the effect seen in Fig. 6B might be due to a translational regulatory effect of LppE, a constitutively expressed translational fusion of *phoA* to P*lppE* and the N terminus of the *lppE* coding sequence (22 amino acids), which encodes the signal peptide transporting LppE to the periplasm, was generated. This fusion was expressed in Ω*lppE* harbouring *lppE* under P*tau* in pLMB509. Alkaline phosphatase activity was measured in the absence and presence of taurine. There was no significant difference in alkaline phosphatase activity generated by the translational fusion in the presence or absence of LppE expressed from P*tau*, showing that LppE does not directly regulate translation of its own operon (Fig. 5.6D). Therefore, the effect of overexpressing *lppE* on P*lppE*-driven luminescence seen in Fig. 6B is not due to a direct regulatory role for LppE. To investigate the possibility that LppE interacts with AsfE to stabilise its binding to EcfE, I carried out a computational simulation using Alphafold2 with MMSeqs2. This model shows that LppE may interact with the periplasmic domain of AsfE (Fig. 5.6E). The residues predicted to be involved in the interaction are highly conserved across a range of *Rhizobium* species, supporting the idea that an interaction takes place and suggesting it may be important for the functioning of LppE and AsfE. It is therefore likely that LppE indirectly regulates expression of P*lppE* by binding to AsfE to stabilise anchoring of EcfE at the membrane, preventing its release in the absence of an external signal, meaning that it is not able to activate expression of P*lppE*.

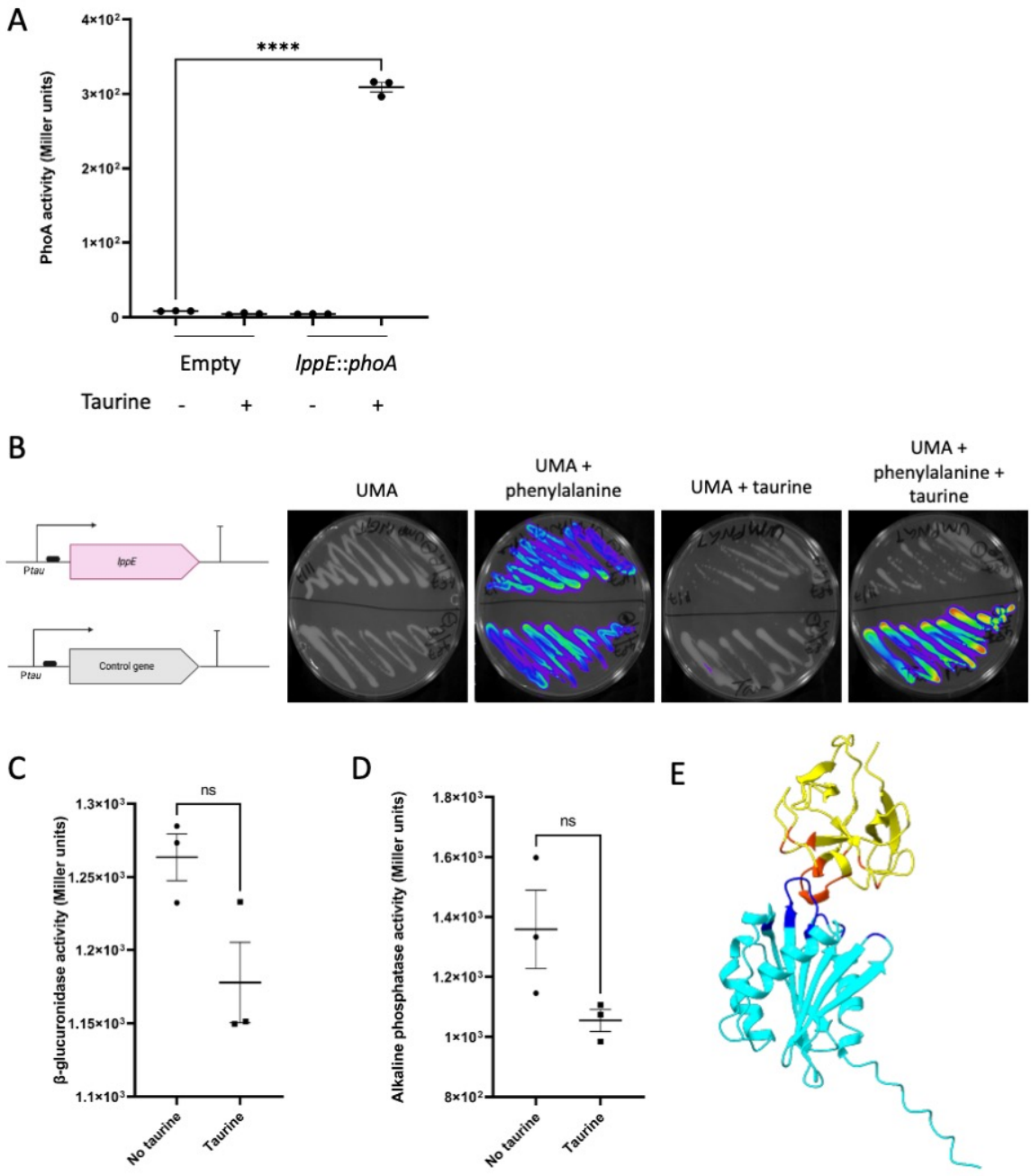


Figure 5.6: The role of LppE in regulation of the ecfE operon. *A, Activity of alkaline phosphatase fused to LppE signal peptide. B, Luminescence in Rlv3841 expressing PlppE::lux and either taurine-inducible lppE (pOPS0050) or a taurine-inducible control gene (pLMB509) streaked on agar and imaged using NightOWL. Taurine added at 10 mM; phenylalanine added at 5 mM. C, Activity of beta-glucuronidase fused transcriptionally to lppE in ΩlppE with taurine-inducible lppE expression. D, Activity of alkaline phosphatase fused translationally to lppE in ΩlppE with taurine-inducible lppE expression. E, Predicted interaction between LppE and the periplasmic domain of AsfE, modelled using AlphaFold Multimer via*

*ColabFold with MMSeqs2. LppE in yellow; AsfE in blue. Conserved residues involved in interactions are coloured dark. All values given as mean ± SEM (Miller units). Biological replicates n=3. Significance calculated using unpaired t tests. ns=p>0.05; *=p≤0.05; **=p≤0.01; ***=p≤0.001; ****=p≤0.0001*

5.2.6 P*lppE* expression is activated by a plant cell surface-associated signal

To further elucidate the mechanism by which P*lppE* expression is activated in rhizobia attaching at the root elongation zone, I characterised the plant signal that induces expression. *In planta* luminescence assays suggest that P*lppE* expression is restricted to cells that are at or near the root surface, as shown in the images in Fig. 5.2. However, the NightOWL camera used to image luminescent bacteria is not sufficiently spatially resolved to establish whether luminescence is strictly restricted to the root surfaces or whether a signal is diffusing out into the media surrounding the plant root, allowing non-attached cells to express *lppE*. To establish whether the plant signal is restricted to the root surface or diffusible, Rlv3841 constitutively marked with mCherry and harbouring a P*lppE*::*gfp* fusion was inoculated onto *Trifolium pratense* roots and imaged in Farhaeus slides 24 hours post inoculation. *T. pratense* was used for confocal imaging as the smaller root systems are more suited to growth on Fahraeus slides. Images showed that while *Rhizobium* cells localise around the root, only those in the root elongation zone expressed *lppE* (Fig. 5.7A). This confirms that the proposed 'induction factor' is anchored at the root surface, rather than being secreted in the root exudate.

To establish where on the root the 'induction factor' localises, 5-day-old pea seedlings were cut into 1 cm sections, representing regions at 1 cm intervals from the root tip. Each of these sections was homogenised in phosphate-buffered saline to create a fraction of root material. Fractions were normalised for protein content and added to liquid cultures of Rlv3841 expressing the P*lppE*::*lux* fusion and luminescence measured (Fig. 5.7B). In the fraction taken from nearest the root tip, only the lowest masses of homogenised root (2 □g and 5 □g) were able to induce P*lppE* expression. In the fraction taken from 1-2 cm away from the root tip, all concentrations of homogenised root were able to induce expression but the highest expression was induced by an intermediate mass of 7.5 □g. Fractions taken

from >2 cm away from the root tip did not induce P*lppE* expression unless added at very high mass (10-20 □g). This suggests that the 'induction factor' is present on root surfaces in a concentration gradient with the concentration being highest at the root tip and lowest near the seed, with P*lppE* expression only being induced at intermediate concentrations, as found 1-2 cm away from the root tip. Furthermore, it suggests that the *lppE* operon is exquisitely sensitive to the absolute concentration of inducing factor, such that too little or too much root factor reduces induction.

Comparison between P*lppE*-driven luminescence in bacterial liquid culture and in bacteria grown on solid agar slopes shows that P*lppE* activity is higher in cells grown on solid media (Fig. 5.7C). This suggests that there is a mechanical element to P*lppE* induction, where expression from the promoter is affected by changes to the bacterial cell surface that come about by interaction with a solid surface. To further investigate this possible cell surface interaction effect, the P*lppE*::*lux* promoter fusion plasmid was conjugated into Rlv3841 mutants in *pssA* (LMB310, [358]) and *lpcAB* (RU301, [359]). The *pssA* mutant is deficient in EPS production, and the *lpcAB* mutant is deficient in LPS production. Both mutants therefore have extremely altered cell surface characteristics compared with Rlv3841. Wild type and mutant strains harbouring P*lppE*::*lux* were grown in the presence of 10 □g of pea protein obtained from root tips ground in phosphate-buffered saline and luminescence measured (Fig. 5.7D). Both mutants exhibited higher and much more variable P*lppE* expression in response to the addition of pea root protein than the wild type strain, suggesting that their altered cell surfaces in some way interfere with the ability of the cells to respond to the pea root signal.

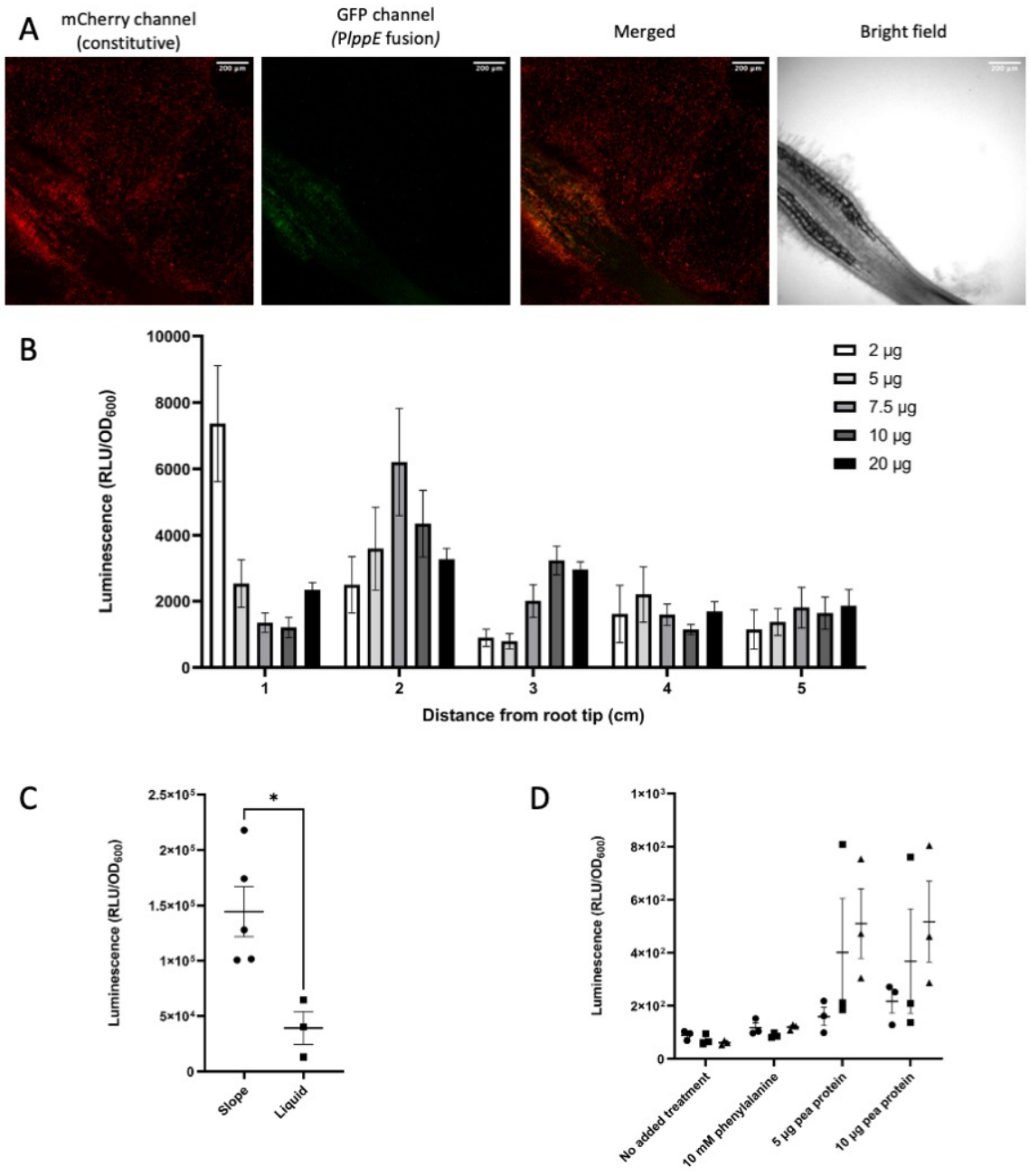


Figure 5.7: PlppE is induced by a component anchored to the root surface. *A, GFP fluorescence driven by expression from PlppE in Rlv3841 expressing constitutive mCherry inoculated onto T. pratense roots. Seedlings transferred to Fahraeus slides and inoculated 36h before imaging. Images taken using LEICA TCS SP5 and processed with FIJI. B, Luminescence driven by PlppE expression in Rlv3841 induced with ground pea root material, taken in sections 1, 2, 3, 4 or 5 cm away from the root tip. Values are mean ±SEM (Relative luminescence units/OD_{600}). Biological replicates n=3. C, Luminescence driven by expression from PlppE in Rlv3841 grown in liquid culture and on solid media. Biological replicates ≥ 3.*

*Significance calculated using an unpaired t test. ns=p>0.05; *=p≤0.05; **=p≤0.01; ***=p≤0.001; ****=p≤0.0001. D, Luminescence driven by expression from PlppE in Rlv3841 (circles), LPS (squares) and EPS (triangles) mutants, grown in liquid media supplemented with 10 mM phenylalanine, 5 □g pea root protein or 10 □g pea root protein. Values are mean ±SEM (Relative luminescence units/OD_{600}). Biological replicates n≥3.*

5.2.7 The pea root tip 'induction factor' is a glycoprotein

To further characterise the nature of this 'induction factor', P*lppE*-driven luminescence was measured in *Rhizobium* grown in the presence of pea roots homogenised in phosphate-buffered saline and treated in various ways (Fig. 5.8). Treating the ground root with viscozyme, to remove carbohydrates, or proteinase, to remove proteins, reduces P*lppE*-driven luminescence to significantly below the levels seen in cells grown with untreated ground root, showing that the 'induction factor' must contain both carbohydrate and protein components. Proteins smaller than 50 kDa in mass also elicit significantly reduced P*lppe*-driven luminescence, showing that the 'induction factor' must be greater than 50 kDa. Heat treating the root does not significantly impair its ability to induce P*lppE* expression. I therefore conclude that P*lppE* expression is driven by interaction of the bacterial cell surface with a heat-stable pea root surface glycoprotein, greater than 50 kDa. Cell wall glycoproteins have previously been shown to play a role in a range of plant-microbe interactions [360], [361], [362], [363], [364].

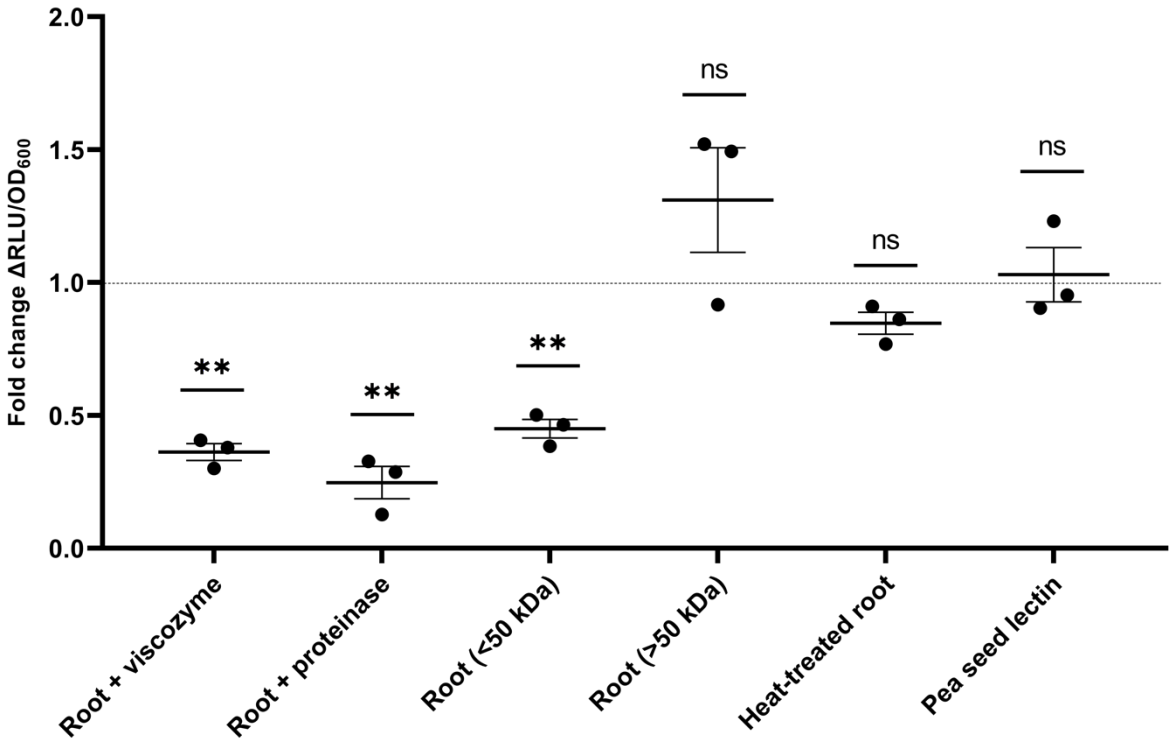


Figure 5.8: PlppE is induced by a glycoprotein ≥ 50 kDa. *Roots of pea seedlings were ground in PBS and treated in various ways before being added to liquid cultures of Rlv3841 harbouring the PlppE::lux fusion. Luminescence was measured relative to that induced by untreated ground pea root. Values are mean ± SEM (fold change in relative luminescence units/OD_{600}). Biological replicates n=3. Significance calculated using a one-sample t test. ns=p>0.05; *=p≤0.05; **=p≤0.01; ***=p≤0.001; ****=p≤0.0001.*

To further purify the glycoprotein, total pea root proteins were extracted from 7-day old roots homogenised in phosphate-buffered saline, then glycoproteins separated by binding with concanavalin. Total glycoproteins were further separated by QAE anion exchange chromatography or on the basis of mass by gel filtration (Superdex 200 column) into fractions that could be tested for P*lppE*-inducing activity using the P*lppE*::*lux* fusion. Fractions both able and unable to induce P*lppE* expression were analysed using mass spectrometry to identify proteins more abundant in fractions that induced P*lppE* expression than in those that did not. This generated a list of 16 candidate groups for the 'induction factor', shown in Table 5.3. Filtering these candidates by those predicted to assemble into glycoproteins, and by cellular localisation to the membrane, reduced the initial 16 candidate

groups to five of particular interest (shown in red in Table 5.3), containing a total of 10 candidate proteins. Predicted mass was calculated for each protein based on amino acid sequence, as the target protein is known to be greater than 50 kDa. However, since these predictions do not take into account the extra mass added by glycosylation or multimerisation, proteins with a smaller predicted mass cannot be ruled out. Two germin-like proteins, A0A9D5B2D0 and A0A9D4XZ80, had an average differential label-free quantification (ΔLFQ) of 0.0256. Their protein sequences are identical excepting one amino acid at position 192, which is a leucine in A0A9D5B2D0 and a proline in A0A9D4XZ80. They share 21 relatives with over 90% identity listed in UniProt, all of which are found in legume species. Germins are a family of plant glycoproteins with oxalate oxidase activity found in cereals; germin-like proteins (GLPs) have a germin motif but do not exhibit oxalate oxdiase activity [365]. Many GLPs exhibit superoxide dismutase activity and as such germins and GLPs are thought to be involved in plant defense responses [366]. GLPs are targeted to the cell wall [366], and both A0A9D5B2D0 and A0A9D4XZ80 have predicted signal peptide domains, suggesting they are expressed on the cell surface, a necessary feature to interact with Rlv3841 during attachment and colonisation. Alignment using the Clustal Omega program showed that A0A9D5B2D0 and A0A9D4XZ80 share 47.0% identity with PsGER1, a nodule-associated germin-like protein in pea which is a predicted rhicadhesin receptor [367], [368]. PsGER1 has been purified from the cell walls of pea roots [367] and also localises to the epidermis of the nodule, where it is most strongly expressed in the zone of cell expansion directly adjacent to the nodule apical meristem [368], which is interesting as the P*lppE*-inducing activity of the glycoprotein characterised in this study takes place in the root elongation zone, another area of rapid cell expansion. This makes A0A9D5B2D0 and A0A9D4XZ80, relatives of PsGER1, good candidates for P*lppE* induction activity in the root elongation zone. Three proteins with alpha-glucosidase domains, A0A9D5GZC8, A0A9D5GZF3, and A0A9D5GZP8, were also identified as candidate P*lppE*-inducing proteins. A0A9D5GZC8 and A0A9D5GZF3 have 97.2% identity to each other and both have 95.0% identity to A0A9D5GZP8, which is shorter but otherwise identical. There is no documentation of alpha-glucosidases playing a role in *Rhizobium*-legume symbioses or other plant-microbe interactions in the

literature; indeed, alpha-glucosidases are known to be inhibited by flavonoids, which also play a key role in host-symbiont recognition in the *Rhizobium*-legume symbiosis, suggesting that alpha-glucosidase activity is not important in this process. The laccase domain-containing protein A0A9D4VZY6 was a further candidate for the role of P*lppE* inducer. Laccases are multicopper oxidases that catalyse oxidation of a range of substances. They are found across all kingdoms of life, and plants usually express multiple laccase genes [369], [370]. In plants, laccases have been implicated in lignin synthesis, pigmentation, jasmonic acid biosynthesis, flavonoid biosynthesis and secondary cell wall biosynthesis [370], [371]. There is no documented role for laccases in the *Rhizobium*-legume symbiosis, but a study of all possible soybean laccases showed that the majority are secretory proteins [372], and A0A9D4VZY6 has a signal peptide that is predicted to localise it to the apoplast. A0A9D4VZY6 also has 68.2% identity to Glyma.01G108200 and 57.7% identity to Glyma.07G142600, two soybean laccases that are specifically expressed in root tissues, and only 24.1% identity to Glyma.07G225300, which is expressed in flowers [372]. Three LysM domain-containing proteins, A0A9D4ZUU2, A0A9D4VK66, and A0A9D4VKE7, were also candidate glycoproteins that may exhibit P*lppE* induction activity. A0A9D4VK66 and A0A9D4VKE7 have 96.7% sequence identity to each other and over 65% identity to A0A9D4ZUU2. Legume Nod factor receptors contain LysM domains [32], [33], [373] that perceive rhizobial exopolysaccharides [34], meaning that they are already known to have a significant role in the *Rhizobium*-legume symbiosis. The final candidate was the FAS1 domain-containing protein A0A9D4X3K2. FAS1 domain proteins are involved in cell-cell adhesion in a variety of contexts [374]. Bacterial FAS1 domain proteins have been shown to be important in *Rhizobium*-legume interactions [375], [376], in other plant-microbe interactions [220], and in *Nostoc*-lichen symbioses [377]. There is no evidence in the literature that plant FAS1 domain proteins play a role in legume interactions with rhizobia, but given that the bacteria involved in these interactions have increased expression of these proteins it seems plausible that a plant FAS1 domain protein could also be important in bacterial recognition. Overall, it is difficult to narrow down which of these candidate protein groups is the one responsible for P*lppE* induction in the root elongation zone, let alone which

individual protein within the group is the target glycoprotein, if it is one and not multiple, which is also a possibility given the level of sequence similarity between some of the candidates. However, based on the information available in the literature, one of the germin-like proteins or LysM domain-containing proteins seems the most likely candidate, as these types of proteins are already known to be involved in bacterial recognition during *Rhizobium*-legume interactions. As the germin-like proteins are around 40 times more concentrated in the inducing fractions than the non-inducing fractions, while the LysM domain-containing proteins are only around twice as concentrated in the inducing fractions than the non-inducing fractions (Table 5.3), one of the Germin-like proteins is the most promising candidate for the P*lppE* 'induction factor'.

Table 5.3: Pea protein candidates for the PlppE 'induction factor'. *Candidates obtained by purification and mass spectrometry. Candidates highlighted in red are of particular interest based on membrane localisation and predicted glycan modification. Values are average differential label-free quantification expressed as (LFQ in inducing fraction)/(LFQ in paired non-inducing fraction).*

Protein IDs	Annotation	Average differential LFQ	Glycoprotein	Localisation	Function	Mass (kDa)
A0A9D5B2D0; A0A9D4XZ80	Germin-like protein	39.0413035	Yes	Apoplast	Structural proteins; auxin receptors; protein receptors; oxalate oxidases; superoxide dismutases	24; 24
A0A9D4XEF7; A0A9D4XE88	Alpha-mannosidase	27.8598898	Unknown	Unknown	Mannose degradation	114; 65
A0A9D5GZC8; A0A9D5GZF3; A0A9D5GZP8	Alpha-glucosidase	13.588988	Yes	Membrane	Hydrolysis of glycosidic bonds	102; 100; 72
A0A9D4Y6X5; A0A9D4Y877	Polyvinylalcohol dehydrogenase	7.81717415	Unknown	Membrane	Degradation of alcohols	61; 59

A0A9D4VMF1	Thioredoxin-like fold domain-containing protein	6.41515407	Unknown	Unknown	Change redox state of thiol groups to regulate enzyme activity	25
A0A9D4XBS9	Amine oxidase	4.75021415	Unknown	Unknown	Catalyse cleavage of alkylamines into aldehydes and ammonia	76
A0A9D5B2B4; A0A9D4XWN3; A0A9D4VVJ8; A0A9D4VVQ7; A0A9D4WYZ3; A0A9D4XUI6; A0A9D4XSR3; A0A9D4WYP7; A0A9D5ALQ7; A0A9D4VYM9; A0A9D4XKS5;	Aspartate-semialdehyde dehydrogenase-like protein	3.80816692	Unknown	Unknown	Bind carbohydrates and hydrolyse glycosidic bonds	51; 58; 56; 28; 41; 41; 39; 20; 17; 21; 42; 32; 21; 31; 31; 31; 50

A0A9D4XKZ8; A0A9D5AXH2; A0A9D4WYT3; A0A9D4XLP7; A0A9D4XTH4; A0A9D4W7U3						
A0A9D4WCY7	Peroxidase	2.96965911	Yes	Secreted	Removal of hydrogen peroxide from chloroplasts and cytosol; oxidation of toxic compounds; biosynthesis of the cell wall; defence responses towards wounding; indole-3-acetic acid (IAA) catabolism; ethylene biosynthesis	35
A0A9D5AMW1	Epidermis-specific secreted glycoprotein EP1-like	2.62826736	Yes	Unknown	Binds mannose	47

A0A9D4VZY6	Laccase	2.10102244	Yes	Apoplast	Lignin degradation and detoxification of lignin-derived products	64
A0A9D4Y5R0	Glucan endo-1,3-beta-D-glucosidase	2.077695	Unknown	Unknown	Hydrolysis of glucan chains	74
A0A9D4Y6B7; A0A9D4VUR0; A0A9D5BF84; A0A9D5BF82; A0A9D5BF74	Neutral ceramidase	2.02012089	Unknown	Membrane	Hydrolysis of the sphingolipid ceramide into sphingosine and free fatty acid	86; 12; 15; 17; 50
A0A9D4ZUU2; A0A9D4VK66; A0A9D4VKE7	LysM domain-containing protein	1.84097207	Yes	Membrane	Peptidoglycan binding; bacterial cell wall degradation; plant-bacterial signalling	40; 16; 17
A0A9D4WHT8	Uncharacterized protein	1.43314996	Unknown	Membrane	Unknown	56
A0A9D4X921; A0A9D4X772	Amine oxidase	1.3847514	Unknown	Unknown	Catalyse oxidation of biogenic amines	76; 32

A0A9D4X3K2	FAS1 domain-containing protein	1.35442706	Yes	Membrane	Cell surface adhesion	26

5.2.8 The *ecfE* operon is implicated in attachment to and colonisation of plant roots by *Rhizobium*

To establish whether the *ecfE* operon genes have a quantifiable effect on attachment to pea roots, attachment assays were carried out on mutants in each of the three genes harbouring the lux cassette *luxCDABE* under control of the constitutive promoter P*neo* [142]. Strains were suspended in phosphate-buffered saline and root tips (5 cm) immersed in this bacterial suspension for two hours before imaging to visualise bacterial attachment. Mutants in both *ecfE* and *asfE* exhibited reduced attachment to pea roots compared with the wild type strain (Fig. 5.9A, B). A mutant in the operon *fbpABC* also exhibited reduced attachment to pea roots compared with the wild type. *fbpA* is 5.6-fold upregulated in Ω*asfE* compared with wild type, suggesting that it may be part of the EcfE regulatory network. Δ*lppE*, however, showed significantly increased attachment.

Both the differentially regulated genes in Ω*asfE* and the significant reduction in root attachment by Δ*fbpABC* suggest a link between initial root attachment and iron availability and uptake. To investigate this further, the attachment of Rlv3841 (grown in UMS minimal medium) to pea roots was measured in iron-limited and iron-sufficient resuspension buffer. All attachment assays described in this work were normally carried out using bacteria suspended in PBS alone, but for this experiment this was supplemented with 0.04 mM iron (III) citrate to create an iron-sufficient resuspension buffer. In iron-sufficient conditions, Rlv3841 attached approximately twice as effectively to pea roots, confirming the link between iron availability and root attachment (Fig. 5.9C). To further investigate this, attachment assays were carried out on mutants in the iron uptake regulator RL0777 (*rirA*) and iron (II) transporter RL2602 (Fig. 5.9D). Neither of these mutants showed altered attachment compared with wild type.

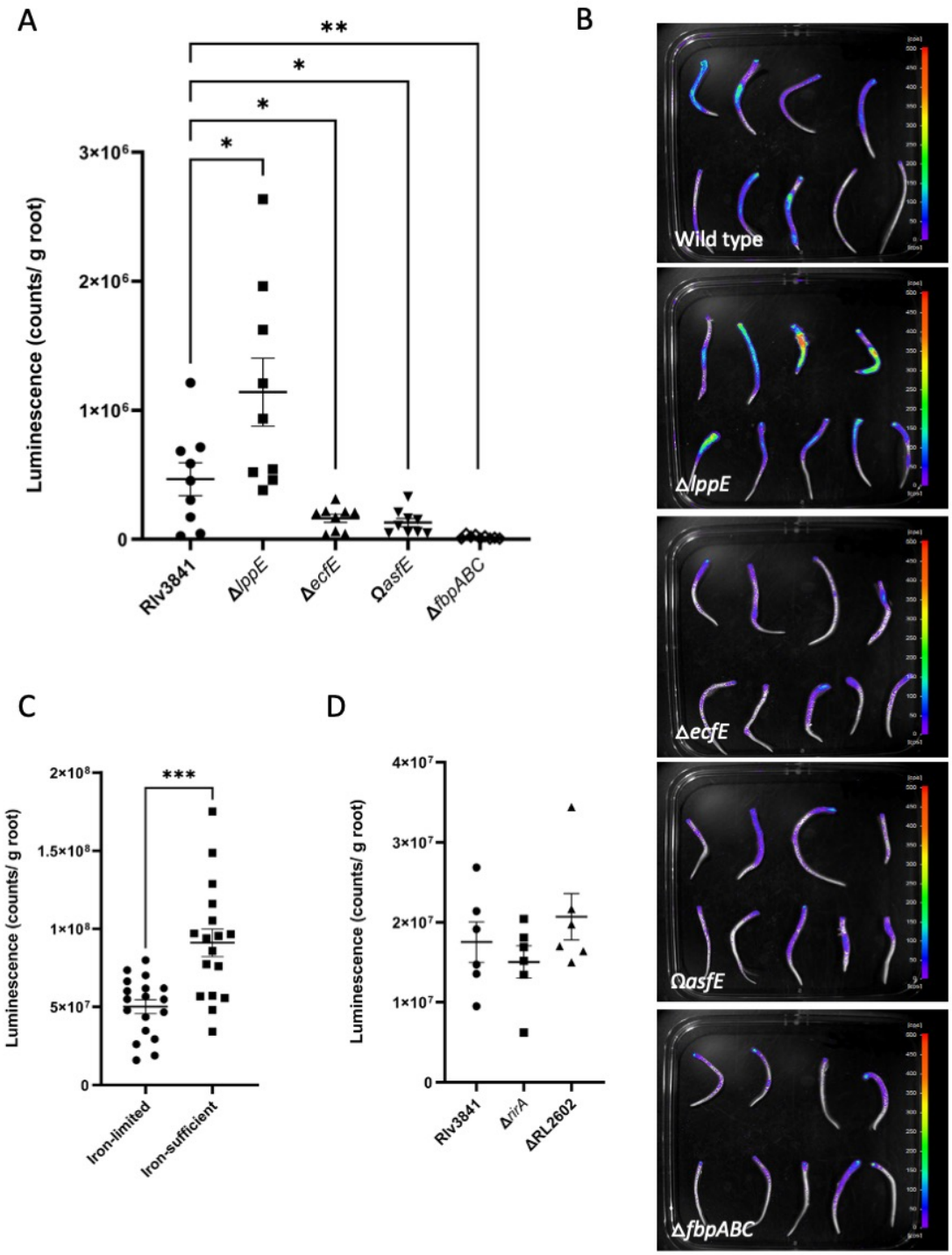


Figure 5.9: Attachment to pea roots is affected by mutation in the ecfE system and iron availability. *A, Root attachment by ecfE operon mutants and fbpABC mutant harbouring Pneo::lux fusions. Biological replicates n=9. B, Representative images taken using NightOWL camera to show luminescence driven by constitutive Pneo in attaching bacteria. C, Effect of iron on pea root attachment. Biological replicates n=18. D, Effect of mutation in iron uptake-related genes on pea root attachment.*

*Biological replicates n=6. Values are luminescence counts/g root ± SEM. Significance calculated using Brown-Forsythe and Welch ANOVA tests and Dunnett's multiple comparisons test. ns=p>0.05; *=p≤0.05; **=p≤0.01; ***=p≤0.001; ****=p≤0.0001.*

To investigate the effects of the operon on attachment in longer term colonisation, Δ*lppE*, Δ*ecfE* and Ω*asfE* harbouring P*neo*::*lux* fusions were inoculated onto the roots of pea plants grown on square plates and imaged 7 days post inoculation (Fig. 5.10) . While the *lppE* mutant showed no change in colonisation compared with the wild type, the *ecfE* and *asfE* mutants showed significantly decreased colonisation of pea roots. As the operon was also shown to be upregulated during colonisation of the barley rhizoplane, Ω*asfE* was also assayed for colonisation of barley, showing significantly reduced colonising ability compared with wild type.

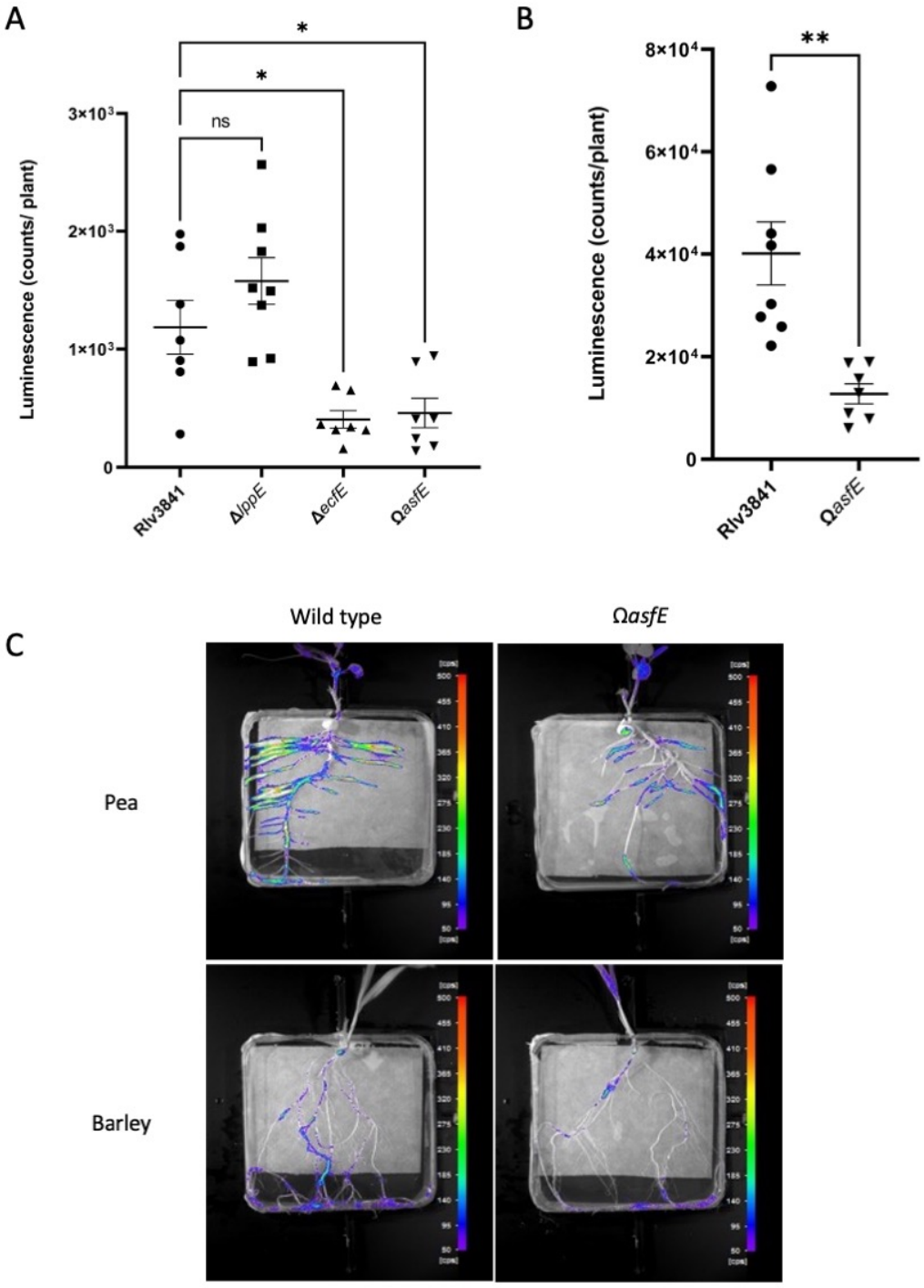


Figure 5.10: Colonisation of pea and roots is affected by mutations in the ecfE system. *A, Colonisation of pea by wild type and mutants in all three genes after 7 days. B, Colonisation of barley by wild type and ΩasfE after 7 days. Values are mean ± SEM (Luminescence counts per second per g root). Biological replicates n=10. Significance calculated using unpaired t tests. ns=p>0.05; *=p≤0.05; **=p≤0.01;*

****=p≤0.001; ****=p≤0.0001. C, Representative images taken using NightOWL camera 7 dpi to show luminescence driven by constitutive Pneo in colonising bacteria.*

5.2.9 The *ecfE* operon is not implicated in Rlv3841 nodulation of pea

To establish whether the effects of the *ecfE* operon on attachment and colonisation affect competition for nodule occupancy, Rlv3841 and Ω*lppE* strains marked with *gusA* and *celB* were co-inoculated on peas. The Ω*lppE* strain was not impaired in nodulation compared with wild type (Table 5.4). The effects of the *ecfE* operon are therefore restricted to rhizoplane attachment and colonisation.

Table 5.4: Effect of mutation of* ecfE *operon on nodulation competitiveness. *Nodulation competition assay carried out as described in* [378]*. Briefly,* gusA-*labelled Ω*lppE *(OPS0750, Poole lab) and* celB-*labelled Rlv3841 (3841*celB*,* [378]*) were inoculated onto pea roots in the ratios below. Plants were harvested 21 dpi and stained with Magenta-glc and X-gal and nodules counted. Values are percentages with SEM shown. Biological replicates n ≥ 2.* X^2 *tests were used to test for significant deviation of nodule percentages from expected values based on the ratio of strains in the inoculum. ns=p>0.05; *=p≤0.05; **=p≤0.01; ***=p≤0.001; ****=p≤0.0001.*

Rlv3841: Ω*lppE* ratio in inoculant	Rlv3841 nodules (%)	Ω*lppE* nodules (%)	SEM	X^2 p-value
10:1	89.85	10.15	0.006	0.88 (n.s.)
1:1	57.94	42.06	0.025	0.27 (n.s.)
1:10	21.92	78.08	0.039	0.13 (n.s.)

5.3 Discussion

In this chapter, I characterised a three-gene operon that RNA sequencing revealed to be upregulated during attachment to and colonisation of the pea and barley rhizoplanes (Chapters 3 and 4). It is also shown by microarray data to be upregulated during colonisation of the rhizospheres of pea, alfalfa and sugar beet [205]. Validation of this finding using a *lux* promoter fusion showed that the primary promoter of the operon begins to be expressed within 6 hours of Rlv3841 cells being inoculated onto pea roots and continues to be expressed for at least 7 days after inoculation. This suggests that the three genes in the operon are involved in the transition from a free-living lifestyle in the soil to longer-term colonisation on pea roots, as the promoter is induced early in attachment and continues to drive gene expression over a period of days, during which bacteria are known to switch from forming nonspecific, reversible bonds to forming irreversible bonds with plant root surfaces and beginning to divide and form biofilms [112], [345], [379]. Even more interestingly, the promoter is only switched on in bacterial cells adhering to the root elongation zone. This is where rapid cell elongation happens, with cells growing eight times more quickly than in the root apical meristem [354], resulting in the creation of new surface area and meaning that in field conditions it may be one of very few locations on the plant root that has not already been colonised by a range of soil microorganisms. I posit that a set of genes capable of detecting the narrow section of root surface which will be relatively free of other, competing, microbes will confer a competitive advantage on *R. leguminosarum* by allowing it to preferentially invest in forming permanent bonds with and colonising a niche where space and resources will be more freely available. The root elongation zone has also been linked to other plant-microbe interactions; for example, *Bacillus subtilis* has been shown to aggregate at the elongation zone of *Arabidopsis* roots in less than 20 minutes post inoculation, before spreading to other parts of the root, showing a preferential targeting of this region [380]. Meanwhile, chitin and flg22 induce transient calcium signals that initiate from the root elongation zone [381]. This suggests that the root elongation zone is an important site of two-way communication between plants and bacteria during colonisation.

The operon consists of genes encoding a lipoprotein, *lppE*, an extracytoplasmic sigma factor, *ecfE*, and an anti-sigma factor, *asfE*. Assaying for promoter activity in mutants of each of the three genes shows that the sigma factor EcfE activates expression from the primary promoter region of the operon, while the anti-sigma factor AsfE prevents expression. This is in line with other ECF sigma/anti-sigma factor pairs that often regulate their own expression [228]. This suggests that EcfE and AsfE may play a role in the spatiotemporal response of P*lppE* to plant root surfaces, in the manner of other ECF sigma and anti-sigma factor pairs, which control bacterial transcriptional responses to external stimuli such as envelope stress (*E. coli* σ^E, *B. subtilis* sigma W), oxidative stress (*Streptomyces coelicolor* sigma R, *Rhodobacter sphaeroides* σ^E) [346] and heat shock (*E. coli* σ^E) [382]. RNA sequencing of the *asfE* mutant also identified several candidate genes whose expression appears to be activated by EcfE. Four contiguous genes, RL1136-RL1139, were all upregulated in Ω*asfE*. RL1138 (*ecfC*) and RL1139 (*asfC*) encode the sigma factor and anti-sigma factor pair EcfC and AsfC, respectively. Thus, EcfE apparently activates a second sigma factor and anti-sigma factor pair. RL1136 has a predicted signal peptide that targets it for export, as well as a DUF305 domain (IPR005183), a domain of unknown function that is shared by bacterial secreted proteins. The predicted structure of RL1137 generated with Alphafold 2 is pore-like, meaning that the protein encoded by RL1136 might be exported through that encoded by RL1137. RL1137 has over 90% identity to YncE proteins in related *Rhizobium* species. *yncE* was first discovered in *E. coli* in studies on iron-dependent gene expression, as its expression is regulated by TonB, itself encoded by an iron-regulated gene [383]. RL1137 also contains a predicted haem domain. The DUF305 domain in RL1136 is also associated with ferritin-like metalloproteins, including iron-binding proteins. RL4584 (*fbpA*), which encodes the solute-binding protein of an Fe (III) transporter, was also 6-fold upregulated in Ω*asfE*. *fbpA* is expressed as an operon with *fbpBC*, which encode the permease and ATP-binding components of the transporter respectively [384]. Neither *fbpB* nor *fbpC* were differentially expressed in Ω*asfE*; however, it is common for solute-binding proteins to be differentially expressed without the other components of their transporters. A range of genes involved in uptake and assimilation of iron were downregulated in Ω*asfE* compared with wild type, suggesting

their expression may be indirectly affected by EcfE. These included the vicibactin synthesis genes *vbsADL* and *vbsSO*, regulatory genes *rpoI* and *tonB* and haem uptake genes *hmuPSTUV*. In *Pseudomonas aeruginosa*, the siderophore pyoverdine is known to be involved in signalling and regulation of virulence factors that are required to cause disease in animals, including exotoxin A and an endoprotease as well as regulation of its own expression, via the ECF PvdS [385], [386]. The ability of bacteria to interact with their host organisms can therefore be mediated by ECF-regulated iron sensing. When RNA sequencing was carried out on Ω*asfE* in iron-sufficient media, these iron uptake-related genes no longer appeared as downregulated. However, *fbpA* remained upregulated in Ω*asfE* in iron-sufficient conditions. This suggests that expression of *fbpA* is either directly or indirectly activated by EcfE. During iron-limited growth on UMS, increased expression of *fbpA* in Ω*asfE* results in increased ability to take up iron, relieving the iron limitation imposed by the media, and resulting in reduced need for *vbs* gene expression. The role of the *ecfE* operon in mediating attachment and colonisation may be in changing the mechanism of iron uptake and assimilation in response to binding to a plant root, facilitating the lifestyle switch between free-living and colonising bacteria, whether via a change in mineral uptake or via regulation of more direct host-microbe interactions.

The other putative transcriptional targets of EcfE suggest that it may play a role in oxidative stress response. RL0880 (encoding a quinol oxidase subunit), RL1800-RL1802 (encoding a transmembrane protein and the protein methionine sulfoxide reductase MsrPQ) and RL1951 (encoding the quinone/NAD(P)H dehydrogenase WrbA) were all upregulated in Ω*asfE* on iron-limited or iron-sufficient media, or both. This is similar to the *S. meliloti* homologue of *ecfE*, SMc01419 (*rpoE1*), which has putative transcriptional targets including SMc04047 (encoding a copper-binding pseudoazurin), SMc04049 (encoding a sulfite oxidase) and SMc04291 (encoding an oxidoreductase) [225]. This suggests that a role for *ecfE* in oxidative stress response may be conserved among rhizobia.

The LppE protein has a signal peptide that localises it to the periplasm, as shown via alkaline phosphatase fusions. Localisation of LppE to the periplasm indicates that it may be involved in signal

transduction from the root elongation zone to AsfE which is anchored at the inner membrane. In other bacteria, anti-sigma factors must be bound by signal proteins to mediate their interactions with sigma factors and their responses to external signalling. In *E. coli*, RseA is bound by a periplasmic protein, RseB, which prevents its cleavage by RseP and stabilises its interaction with σ^E, increasing its affinity by 2-3-fold [348], [387]. This is thought to increase control of the RseA degradation pathway, ensuring that σ^E is only released under stress conditions [388]. It is possible that LppE could be similarly involved in the AsfE-EcfE interaction, stabilising AsfE by binding its periplasmic domain and allowing it to bind EcfE more tightly in the cytosol. I showed that an interaction between LppE and the periplasmic domain of AsfE is possible using modelling with AlphaFold2, and that the residues predicted to be involved in this interaction are highly conserved among *Rhizobium* species. This suggests that LppE does interact with AsfE and may contribute to its ability to bind and release EcfE in the absence and presence of the plant glycoprotein signal.

The primary promoter of the operon is active solely in bacteria attaching to root elongation zones following recognition of a plant root glycoprotein that is localised at the root surface. This glycoprotein is expressed in a concentration gradient, with the highest concentration being found at the root tip and the lowest concentration nearest the seed. Expression from P*lppE* is activated by a specific concentration of the glycoprotein that is found in the root elongation zone; too much or too little glycoprotein will not activate expression of P*lppE*. This response to a signal that is localised so specifically and is anchored at root surfaces rather than being secreted allows expression from P*lppE* to be specific to bacteria attaching to that region of the root, rather than being activated in surrounding soil or on surrounding root tissues. I suggest that this allows the operon to play a role in allowing bacterial recognition of an advantageous location to begin root colonisation. Cell wall glycoproteins have previously been shown to play a role in a range of plant-microbe interactions. Carbohydrate-binding lectins have been documented for many years as being involved in recognition of *Rhizobium* strains by host plants [360], [361], [362], as well as being involved in recognition of pathogens [363]. Extensins, part of the hydroxyproline-rich glycoprotein (HGRP) family, are found across the plant

kingdom and throughout all plant tissues and have been shown to be involved in plant defence against microbial pathogens, in particular via changes in gene expression following exposure to pathogens [364]. I observed that the background expression of P*lppE* in the absence of either phenylalanine or a plant signal is stronger in bacteria grown on solid media than in liquid culture, suggesting that the mechanism of recognition of the plant signal may be partly mechanical, via a physical interaction that conformationally changes the bacterial cell wall. Physical interaction with the bacterial cell surface could therefore be the mechanism by which the glycoprotein is able to drive P*lppE* activity in cells bound to the root elongation zone. This was supported by the finding that the response of P*lppE* to pea root protein seems to be higher and is more variable in EPS and LPS mutants, which have altered cell surface characteristics, than in wild type Rlv3841. Other sigma factors are known to respond to altered cell membrane structures, such as σ^E in both *E. coli* and *Streptomyces coelicolor* [349], [350], [351], [389]. Physical interaction with the bacterial cell surface could therefore be the mechanism by which the glycoprotein is able to drive P*lppE* activity in cells bound to the root elongation zone. Further purification of the P*lppE*-inducing fraction of pea roots and sequencing by mass spectrometry identified 10 possible candidates for the inducer, and further investigation of these suggested that the target is likely a germin-like protein, which are known to play a significant role in bacterial recognition by plant roots in the *Rhizobium*-legume symbiosis [367], [368].

I showed that both the *ecfE* and *asfE* mutants are deficient in attachment to and colonisation of pea roots. EcfE and AsfE regulate transcription in opposing ways, with EcfE being a transcriptional activator of P*lppE* and AsfE being an anti-sigma factor that prevents EcfE from acting except under certain conditions. I therefore suggest that P*lppE* induction and transcription of *ecfE* and its downstream targets in response to the root glycoprotein are essential to successful attachment and colonisation. A perturbation of the system prior to bacterial exposure to roots, as caused by mutating either a transcriptional activator or anti-sigma factor, leads to an altered or diminished response to the glycoprotein, preventing *Rhizobium* from attaching and colonising successfully. As on pea roots, colonisation of barley roots by the *asfE* mutant was significantly impaired compared with the wild type.

AsfE has been shown to be important for colonisation of the host and non-host plants pea, lentil, lathyrus, alfalfa and barley in BarSeq experiments [Knights et al., in preparation]. The *ecfE* operon genes may therefore be important for colonisation of a range of plants, both host-legumes and non-host crops including agriculturally significant cereals. I also showed that Δ*fbpABC* is deficient in attachment to pea roots compared with Rlv3841 wild type. *fbpA* was 6-fold upregulated in the *asfE* mutant compared with wild type, suggesting that its transcription is positively regulated by EcfE. This suggests that the attachment phenotypes of the *asfE* and *ecfE* mutants are due to changes in expression of genes such as *fbpA* in the mutants, rather than being caused by a direct effect of the *ecfE* operon products themselves. In an InSeq screen, the *fbpA* mutant was found to be deficient in attachment to soybean and barley roots at pH 7.0 and in attachment to pea roots at pH 6.5, 7.0 and 7.5, in addition to being growth defective on minimal media, although not on rich media or in the pea rhizosphere [133]. This suggested a link between iron availability and uptake and attachment to pea roots in Rlv3841. I confirmed this link by showing that the addition of iron (III) citrate to the suspension buffer used for attachment assays substantially improved root attachment in Rlv3841 wild type. However, attempts to further investigate this effect by carrying out attachment assays on mutants in RL0777 (*rirA*) and RL2602 showed that neither of these mutants has an effect on attachment. RirA negatively regulates expression of iron uptake-related genes including *rpoI*, *fhuCDB* and *hmuPSTUV* [390]. It also negatively regulates expression of *fbpA* [391]. In a *rirA* mutant, therefore, expression of these genes is upregulated. RL2602, meanwhile, is annotated as an iron (II) uptake transporter. The fact that mutation in neither of these genes affects Rlv3841 root attachment suggests that EcfE-mediated regulation of *fbpA* expression in response to interaction with the root glycoprotein is part of an iron-related attachment pathway that is separate from the RirA-mediated iron limitation pathway. Moreover, *fbpA* is 2.7-fold upregulated in iron-limited media relative to iron-sufficient media (Poole lab, unpublished data), confirming its role in the RirA-mediated iron limitation response. However, this is not as strong as its upregulation in Ω*asfE*, supporting the hypothesis of a separate and important EcfE-related regulatory pathway. One purpose for the iron taken up by the FbpABC transporter

following EcfE-related upregulation of expression may be binding with the exported proteins encoded by RL1136 and RL1137 which contain a ferritin-like domain and a haem domain respectively. As RL1137 is predicted to be a beta barrel protein it may form a pore in the outer membrane that might allow passage of RL1136 to the outside of the cell, where it might promote binding to roots. The increase in rhizoplane attachment caused by additional iron (III) citrate in the suspension buffer is likely due to effects on this iron-related attachment pathway that is activated by association with plant roots, rather than the RirA regulated pathway that is involved in survival in iron-poor soil and media. Further characterisation of the iron-related attachment pathway is clearly required to determine its exact mechanism.

Preliminary data show that Ω*lppE*, in which *lppE-ecfE-asfE* are all absent, is not impaired in nodule development. The role of the *ecfE* system is therefore limited to rhizoplane attachment and colonisation. This confirms that rhizoplane attachment is a distinct process from root hair attachment, which is an essential precursor to nodule formation [17], as expression of distinct sets of genes is clearly required for each process. As Ω*asfE* is defective in colonisation of both pea and barley, the *ecfE* system seems to be required for colonisation of the rhizoplanes of a range of plants, not just those capable of forming nodules with Rlv3841, meaning that it may be important to agricultural applications of rhizobia as PGPRs.

Overall, the *ecfE* operon clearly plays a central role in attachment to and colonisation of the plant rhizoplane by *R. leguminosarum*, as is summarised in Figure 5.11. Through sensing of precise concentrations of a plant root-associated glycoprotein, it enables rhizobia to identify the root elongation zone and begin localised expression of the *ecfE* operon. EcfE then activates transcription of an iron-related attachment pathway. EcfE-mediated activation of *fbpA* expression allows increased iron uptake during attachment to the root elongation zone. Mutation of *fbpABC* has been shown to drastically reduce attachment to pea roots, while application of exogenous iron results in a twofold increase in attachment, placing iron uptake at the centre of successful attachment. This may be due to

a role for iron in the functioning of the exported proteins encoded by RL1136 and RL1137, as expression of these genes is directly activated by EcfE. The *ecfE* operon genes affect rhizoplane attachment, rather than root hair attachment which precedes nodulation. This places EcfE as a master regulator of attachment to and colonisation of a range of plants beyond pea, including agriculturally significant cereals.

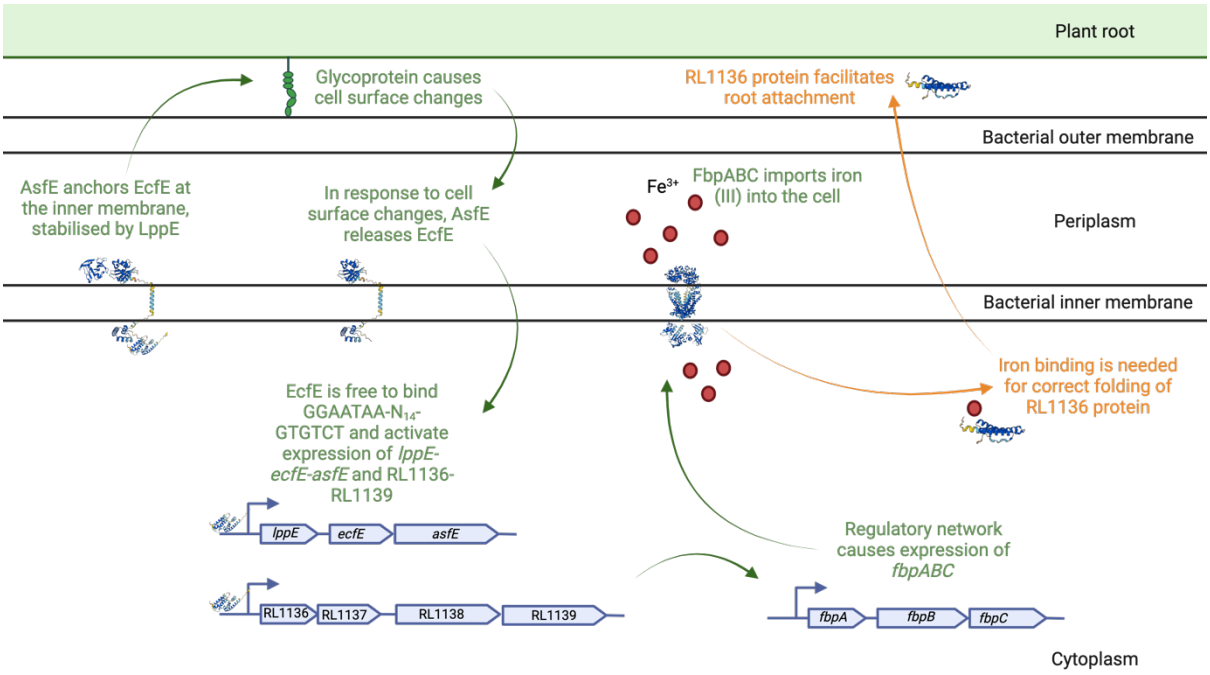


***Figure 5.11: Summary of the hypothetical iron-related attachment pathway regulated by* EcfE.** *Stages characterised in this chapter are shown in green, and those that remain for investigation are shown in orange. Figure generated using BioRender.*

6 General discussion and future directions

6.1 Overview

In this project I made use of whole transcriptome characterisation and more detailed investigation of a single operon to further understanding of attachment to and colonisation of plant rhizoplanes by the model *Rhizobium* strain Rlv3841. *Rhizobium* spp are agriculturally significant PGPRs, primarily due to their ability to form nitrogen-fixing nodules on the roots of legume plants, including the crop plants soybean, pea, lentil and chickpea. In addition to being crop plants in their own right, legumes have been used in crop rotations for thousands of years, to enrich the soil with nitrogen that subsequent, non-legume crops can benefit from [52]. More recently, research efforts have focussed on allowing non-legume staple crops including wheat, maize and rice to benefit directly from biological nitrogen fixation, utilising strategies including engineering of the plants themselves to express the nitrogen fixing enzyme nitrogenase, and engineering PGPRs to fix and release nitrogen for uptake by a wider range of hosts [55], [57], [58], [63], [65], reviewed by [54], [56]. Any solution relying on fixation and release of nitrogen by PGPRs is dependent on the ability of the microorganism in question to competitively colonise its target host. However, the mechanisms by which rhizobia attach to and colonise the rhizoplanes of both host and non-host plants are much less well-studied than the processes by which they form nodules and fix nitrogen [48]. It is therefore vital to further our understanding of how these processes take place so that staple crop plants can benefit from nitrogen fixation by diazotrophic bacteria.

6.2 A reliable method for obtaining transcriptomic data from root-associated bacteria

Pinpointing which bacterial genes and pathways are differentially regulated during attachment to and colonisation of plant rhizoplanes will help further our understanding of how these important processes take place and are regulated, and was a key goal of this project. INSeq screens have been able to uncover more about which genes are important for these processes [133], [169], but a gene being required for attachment or colonisation to take place does not mean it is necessarily expressed during

these stages of infection. It is therefore important to have expression data as well to gain a more holistic view of the genetic basis for attachment and colonisation of rhizoplanes. Moreover, transcriptomic data can be applied directly to engineering of *Rhizobium* to fix nitrogen under specific circumstances, by supplying information about promoter regions that are active in particular environments, including the rhizoplanes of agriculturally significant plants. Whole transcriptome data does exist for Rlv3841 during colonisation of the rhizosphere of its host plant, pea [205] as well as for Rlv3841 bacteroids isolated from pea root nodules [219], but not data is available for colonisation of the rhizoplane specifically. Since distinct microbial communities form in the rhizosphere and at the rhizoplane, it is clear that these are two separate environments [392]. This suggests that colonisation of each of these fractions may require expression of a separate set of genes and that they should therefore be studied separately. However, unpublished data from the Poole lab shows that removing the rhizoplane fraction of bacteria by vortexing roots and then extracting total RNA yields sequencing data where the majority (70%) of reads map to the plant genome, making it difficult to obtain data with a sufficient number of reads mapping to the bacterial genome to generate statistically significant results and enable informative conclusions to be drawn. The accepted sequencing read depth to obtain significant information about differential gene expression in bacteria is between five and 10 million reads [173]. It could therefore be possible to simply sequence to a greater depth to obtain more reads mapping to the bacterial genome. However, despite drops in costs of sequencing, this would still end up being extremely expensive when all necessary biological replicates are taken into account. It was therefore important to develop a method of separating plant and bacterial material before RNA extraction, reducing contamination in the final RNA sample and hence increasing the proportion of reads mapping to the target bacterial genome. I developed a simple method using a differential spin gradient to separate plant and bacterial material before RNA extraction and was able to obtain rhizoplane-derived samples with an average of only 23% of reads mapping to the plant genome. This method made it possible to obtain transcriptome datasets for attachment to and colonisation of the rhizoplane specifically. This data was used to further understanding of how rhizoplane colonisation

differs between plants, to separate the genetic basis of initial attachment to and established colonisation of the rhizoplane, and to examine how colonisation of the host rhizoplane differs from colonisation of the rhizosphere.

6.3 Comparison of the Rlv3841 transcriptome during colonisation of pea and barley

The method for separating plant and bacterial material prior to total RNA extraction was initially validated by carrying out RNA sequencing of Rlv3841 during colonisation of the pea and barley rhizoplanes. I showed that there are significant differences in rhizobial gene expression on the two rhizoplanes. Most strikingly, although Rlv3841 is able to colonise the barley rhizoplane when applied as a single inoculant, the transcriptomic data makes it clear that it struggles to make use to the resources provided in barley root exudates and at the root surface. Central carbon metabolism, including the TCA cycle and glycolysis, was significantly repressed during colonisation of the barley rhizoplane, as were carbon storage pathways including glycogen and PHB synthesis. This suggests that during colonisation of the barley rhizoplane Rlv3841 is starved for carbon, which in turn results in downregulation of cell processes including cell division, DNA repair and cell wall synthesis, as well as a range of secondary metabolic pathways. This was not the case during colonisation of the pea rhizoplane, showing that Rlv3841 is much better adapted to survival on the rhizoplane of its host plant. This result was borne out by experimental work that describes the dynamics of pea and barley colonisation by a synthetic community including Rlv3841. During colonisation of the pea rhizoplane, Rlv3841 is able to become established among the synthetic community, whereas during colonisation of barley it dies out almost immediately after inoculation, becoming undetectable in bacterial samples recovered from the rhizoplane one dpi [Jorrin et al., in preparation]. The differential regulation indicative of carbon starvation seen in Rlv3841 during colonisation of barley clearly means that it is unable to survive a competitive scenario on resources it is poorly adapted to make use of even as a single inoculant. This is particularly interesting as it shows that Rlv3841 would be a poor choice of

target for engineering to generate a diazotroph that fixes nitrogen for the benefit of cereal crops, even if it were able to fix nitrogen either constitutively or in response to a cereal-specific signal. In non-sterile soil already colonised by competing microbial strains, Rlv3841 would be unlikely to successfully colonise a cereal rhizoplane. This shows the importance of knowing which genes are being expressed and hence which metabolic and cellular processes are taking place during colonisation.

Based on the expression data for Rlv3841 during colonisation of barley, I also identified several genes that were much more strongly upregulated during colonisation of the barley rhizoplane than during colonisation of the pea rhizoplane. Some of these are promising targets for use in promoter fusions to genes involved in nitrogen fixation, such as the positive regulator *nifA*, to generate a strain that would fix nitrogen exclusively in response to colonisation of cereal rhizoplanes. In particular, pRL110077-pRL110079, encoding components of a PAAT transporter, were between 8-fold and 19-fold upregulated during colonisation of barley and between 2-fold and 3-fold upregulated during colonisation of pea, suggesting that their expression is barley-specific and their promoter may be a useful target of promoter fusions. pRL120099 encodes the solute-binding protein of a HAAT transporter and was 24.7-fold upregulated during colonisation of barley and 2.2-fold upregulated during colonisation of pea, making it another promising target. Most interestingly, RL3594 (*dapA*), encoding a dihydrodipicolinate synthase, was 53.4-fold upregulated during colonisation of the barley rhizoplane and only 1.9-fold upregulated during colonisation of the pea rhizoplane. *dapA* is also known to be upregulated in *R. phaseoli* in response to maize root exudates, but not in response to root exudates from bean, its host legume [284].

6.4 The Rlv3841 transcriptome during attachment to and colonisation of host pea roots

This work was, to my knowledge, the first to carry out transcriptomic profiling of a diazotrophic *Rhizobium* strain during two different stages of rhizoplane colonisation. I sampled Rlv3841 during early attachment to and established colonisation of the rhizoplane of its host plant, pea, and investigated

the overlap and differences in gene expression during these two processes. Although there is a substantial amount of overlap between the two processes, with around 50% of differentially regulated genes being shared between them, each stage of colonisation did have its own distinct transcriptome, demonstrating the importance of separating these stages to further understanding of how the whole process takes place. In particular, a number of housekeeping and ECF sigma factors were upregulated during both stages, with *rpoH1*, *rpoH2*, *ecfZ* and *ecfN* were over 4-fold upregulated during both stages, and *ecfO*, *ecfJ*, *ecfA* and *ecfE* over 2-fold upregulated in both. Other functions that were upregulated in both attachment and colonisation, possibly as a result of transcriptional activation of the relevant genes by one of the above sigma factors, were transmembrane transport, DNA repair, aldo/keto reductase activity and cell surface characteristics defined by BA14K-like proteins. Motility and chemotaxis related genes were significantly downregulated during both attachment and colonisation, as is the case in *S. meliloti*, where these genes are downregulated during biofilm formation by the quorum sensing ExpR/Sin system [254]. However, the carbon source being utilised by Rlv3841 seems to be different during attachment and colonisation, with malonate, GABA and fatty acid degradation genes being upregulated during attachment and inositol catabolism being upregulated during colonisation. Interestingly, comparison of the Rlv3841 transcriptome during colonisation of the pea rhizoplane with its transcriptome during colonisation of the pea rhizosphere [205] showed that only around 25% of differentially regulated genes overlapped between these two processes, half the number that overlap between rhizoplane attachment and colonisation. This suggests that a substantially different set of genes is required for colonisation of the rhizoplane compared with persistence in the soil, and that a significant lifestyle switch takes place during initial rhizoplane attachment that is then carried forward into rhizoplane colonisation.

6.5 Characterisation of the *ecfE* operon and its role in root attachment and colonisation

The *ecfE* operon was selected as a target for further study, as expression of the three genes was upregulated during colonisation of both pea and barley rhizoplanes. *lppE*, encoding a lipoprotein, was 12.0-fold upregulated during colonisation of pea and 25.8-fold upregulated during colonisation of barley. *ecfE*, encoding an ECF sigma factor, was 3.8-fold and 6.3-fold upregulated during colonisation of pea and barley respectively. *asfE*, encoding an anti-sigma factor, was 1.8-fold and 2.6-fold upregulated during colonisation of pea and barley respectively. Both *lppE* and *ecfE* were also over 2-fold upregulated during attachment to the pea rhizoplane. The three genes have also been shown to be upregulated during colonisation of the pea, sugar beet and alfalfa rhizospheres [205]. This suggests that their products play a part in colonisation of the roots of a range of plants, making them relevant to the role of *Rhizobium* as a PGPR. The primary promoter for the operon was identified, and a fusion to the *luxCDABE* cassette in plasmid pIJ11268 [142] used to investigate expression of the three genes *in planta*. Luminescence imaging revealed that, interestingly, the operon is only expressed in Rlv3841 during colonisation of the root elongation zone, where rapid cell elongation occurs [354]. This is where new surface area is being generated, and the ability to recognise this region and respond by activating transcription of downstream target genes as ECF sigma factors do may be an important property for Rlv3841, enabling it to competitively colonise plant roots. P*lppE* expression begins within seven hours of inoculation on to roots, when initial attachment is taking place, and continues for at least 7 days. This suggests that the system plays a role in recognising the root elongation zone and mediating both initial attachment and longer-term colonisation of the rhizoplane.

EcfE is an ECF sigma factor, which generally play a role in sensing and responding to environmental changes by activating transcription of target genes [228], [286], [301], [302], [346], [352]. I used RNA sequencing of the Ω*asfE* mutant on minimal media to identify the transcriptional targets of EcfE. As is the case with most ECF sigma-anti-sigma factor pairs, EcfE activates expression from its own promoter, as well as several other targets. In general, these targets were related either to oxidative stress

response (RL0880, RL1801-RL1802, RL1951) or to iron binding and uptake (RL1136-RL1137, RL4584). The role of EcfE in oxidative stress response may be conserved, as putative targets of the *ecfE* homologue in *S. meliloti*, SMc01419 (*rpoE1*) include a range of genes with related functions. RpoE1 binding sites are found upstream of its own operon, SMc04047 (encoding a copper-binding pseudoazurin), SMc04049 (encoding a sulfite oxidase) and SMc04291 (encoding an oxidoreductase) [225]. The presence of another ECF sigma factor, *ecfC* (RL1138), among the putative transcriptional targets of EcfE places EcfE within a regulatory network that may be integrating a number of environmental signals and bacterial responses during the lifestyle switch from free-living soil bacteria to rhizoplane colonisation. In *S. meliloti*, there is also an RpoE1 binding site upstream of SMc04051, encoding the ECF sigma factor RpoE4, suggesting that like EcfE, RpoE1 is part of a regulatory cascade [225]. Some of the genes upregulated in Ω*asfE* may be transcriptional targets of EcfC rather than EcfE. Indeed, a putative EcfE binding motif, GGAATAA-N_{14}-GTGTCT, was found in the 5' UTR of *lppE-ecfE-asfE* and of RL1136-RL1139 but none of the other upregulated genes, suggesting that they may be targets of EcfC. The presence of RL1136-RL1137 and RL4584 among the putative transcriptional targets of EcfE suggests that the *ecfE* operon is involved in a switch in iron uptake during rhizoplane attachment and colonisation. RL1136 has a DUF305 domain, which is associated with ferritin-like metalloproteins, suggesting it may be involved in iron or copper binding [310]. RL1137 has a haem domain, and over 90% protein sequence identity to several YncE proteins in closely related *Rhizobium* strains. Expression of *yncE* in *E. coli* is regulated by TonB, itself encoded by an iron-regulated gene [383]. Most compellingly, *fbpA* encodes the solute-binding protein of the ABC transporter responsible for uptake of Fe (III) ions. A link between the *ecfE* system and iron uptake was further suggested by examination of the set of genes downregulated in Ω*asfE*, which included siderophore-encoding genes *vbsADL* and *vbsSO*, regulatory genes *rpoI* and *tonB* and haem uptake genes *hmuPSTUV*. These genes were not downregulated in Ω*asfE* grown on iron-sufficient media. The *vbs* genes are usually upregulated during conditions of iron limitation and this suggests that EcfE-induced expression of *fbpA*, independent of iron availability, frees Ω*asfE* from the iron limitation that is being exhibited by the wild type. It is

therefore probable that EcfE is involved in activation of an iron-related attachment pathway that includes uptake of iron by FbpABC, potentially for use in binding to and enabling folding of the products of RL1136 and RL1137.

P*lppE* expression is activated in the presence of a pea root surface glycoprotein, which is expressed in a concentration gradient along the root, with the highest concentration found at the root tip and the lowest at the base of the root. P*lppE* expression is only induced by a specific concentration of glycoprotein, with too much or too little resulting in the system being switched off. This creates an elegant system whereby Rlv3841 expresses EcfE and its transcriptional targets only when the 'perfect' concentration of glycoprotein is available, at the root elongation zone. Plant cell wall glycoproteins have been implicated in a range of plant-microbe interactions, including interactions with rhizobia [360], [361], [362] and with pathogens [363]. I identified 10 possible candidates for the pea glycoprotein that induces expression of P*lppE*, including two germin-like proteins (A0A9D5B2D0, A0A9D4XZ80), which are around 40-fold more concentrated in P*lppE*-inducing fractions than in non-inducing fractions and, as such, particularly promising candidates. The two proteins have 47.0% protein sequence identity with PsGER1, a germin-like protein that is most strongly expressed in the zone of cell expansion in the nodule and is a predicted rhicadhesin receptor [367], [368]. Therefore, there is strong evidence that these proteins, which are identical except for one amino acid, could be the P*lppE* inducer, warranting further investigation.

The *ecfE* system was shown to have a role in both attachment to and colonisation of pea roots, with mutants in all three genes showing altered attachment and colonisation phenotypes. Mutants in *ecfE* and *asfE* exhibited reduced attachment and colonisation compared with wild type, while the *lppE* mutant showed increased attachment. An *fbpABC* mutant also showed significantly reduced root attachment compared with wild type, showing that at least one of the putative transcriptional targets of EcfE has a role in root attachment and supporting the hypothesis of an EcfE-regulated iron-related attachment pathway where *fbpABC* plays a central role. Moreover, wild type Rlv3841 showed significantly increased root attachment when supplemented with 0.04 mM iron (III) citrate during

attachment compared with attachment in unsupplemented PBS. Iron is known to affect colonisation of plant rhizoplanes by a range of agriculturally relevant bacteria, with the majority of agricultural soils being iron-limited [393], meaning that bacteria must compete to obtain sufficient iron. Plant iron nutritional status can influence bacterial community assembly, with iron-starved barley plants showing a significantly different microbial community composition than iron-sufficient plants [394]. Both plant and bacterial iron response mechanisms have been implicated in cross-kingdom signalling and community assembly. Expression of the *P. syringae* iron limitation response gene *pvdS* is suppressed during growth on *Arabidopsis* leaves [338]. The *Arabidopsis* genes MYB72 and BGLU42 are required for synthesis and excretion of iron-mobilising coumarins under iron limitation, but *myb72* and *bglu42* mutants also show altered root microbiome assembly, suggesting a link between plant defence and iron limitation responses [395]. Moreover, the ability of bacteria to take up iron from their plant hosts has been linked to improved colonisation ability, with *Bacillus velezensis* SQR9 secreting a protein that inserts into the plant plasma membrane, promoting iron leakage and improving bacterial colonisation [396]. Interestingly, in Chapter 4 I observed that genes encoding iron uptake systems were significantly downregulated in Rlv3841 during attachment to the pea rhizoplane, although not during colonisation. *irp6A6B6C*, RL3050-RL3052 and *fbpABC* all encode Fe (III) importers and were all downregulated, as were the haem transporter genes *hmuPSTUV*, and the ECF sigma factor controlling Rlv3841 response to iron limitation, *rpoI*. This is particularly interesting as *fbpA* is one of the putative targets of EcfE, expression of which is upregulated during colonisation and weakly upregulated during attachment. The RNA sequencing described in Chapter 4 was carried out on whole roots, while *ecfE* expression is unique to root elongation zones, and this could mean that local expression patterns caused by EcfE-mediated transcriptional activation were not picked up in the RNA sequencing experiment. EcfE may facilitate an iron uptake response that is unique to the root elongation zone and not seen in bacteria localised to other parts of the root surface. *P. fluorescens* colonising lupin roots is more iron stressed at the root tip than in the lateral root zone [397]. It may be that iron availability differs across the root

surface and that a system that responds to this is important for rhizobial attachment to growing cells with available surface area in the REZ.

Preliminary data shows that the *ecfE* system does not play a role in Rlv3841 nodulation of peas, despite its clear involvement in rhizoplane attachment and colonisation. This is in agreement with the finding that an *S. meliloti* mutant in all sigma and anti-sigma factors encoded in the genome nodulates and fixes nitrogen as wild type [225]. Attachment to root hairs is an essential precursor to root hair curling, infection thread formation and nodule development [17], and is clearly not affected in an *ecfE* operon mutant, which is able to nodulate as effectively as wild type. This suggests that rhizoplane and root hair attachment are two distinct processes, requiring expression of separate sets of genes. The fact that the *ecfE* operon plays a role in rhizoplane but not root hair attachment means that it is important for the role of *Rhizobium* as a plant-general PGPR, but not as a specific symbiont of pea. The specific spatial activation of P*lppE* at the REZ, as well as the EcfE-mediated regulation of a suite of iron-related genes places EcfE as a master regulator of an iron-related attachment pathway that may enable specific and targeted attachment to elongating cells in a range of agriculturally important PGPRs.

6.6 Future directions

6.6.1 Characterisation of the transcriptional targets of EcfE

As outlined above, I identified several possible transcriptional targets for EcfE by carrying out RNA sequencing of the Ω*asfE* mutant. However, the putative EcfE binding motif was only found upstream of RL3234-RL3236 (*lppE-ecfE-asfE*) and RL1136-RL1139, suggesting that the other upregulated genes may in fact be targets of EcfC. In order to verify this, it would be advisable to carry out RNA sequencing on a mutant in *asfC* (RL1139), as any upregulated genes in this mutant would be likely to be transcriptional targets of EcfC. It would also be beneficial to use an electrophoretic mobility shift assay (EMSA) to confirm binding of EcfE to the GGAATAA-N_{14}-GTGTCT putative binding motif upstream of its own operon and RL1136-RL1139. My work has also shown that mutating *fbpABC* causes a significant

reduction in attachment of Rlv3841 to pea roots, and it will be important to investigate whether the other targets of EcfE (whether direct or via EcfC) have an impact on root attachment or colonisation, by generating mutants and carrying out attachment and colonisation assays. Moreover, since EcfE expression is upregulated on barley as well as pea, and Ω*asfE* showed impaired colonisation of barley as well as pea, it would be informative to investigate attachment and colonisation of mutants in EcfE transcriptional target genes on both plants, to find out whether there is any difference between plants downstream of EcfE regulation.

6.6.2 The role of LppE

The roles of EcfE and AsfE in regulating their own expression are well defined by this project: they are canonical sigma and anti-sigma factors, with EcfE activating expression from its own promoter and AsfE being localised to the inner membrane and preventing this activation except in the presence of an inducer such as the pea glycoprotein or phenylalanine. However, the role of LppE is less well-characterised. I showed that LppE is transported to the periplasm and, though it does not have a transmembrane domain, has a hydrophobic patch domain (Yx(FWY)xxD) that may anchor it to either the inner or outer membrane. Mutation of *lppE* does not affect P*lppE* expression, but overexpression of *lppE* does result in inhibition of P*lppE* activity, suggesting that LppE does have a regulatory role in its own operon. Using transcriptional and translational fusions, I showed that this is not a direct transcriptional or translational regulatory effect. However, there are a number of ways in which LppE could have an indirect regulatory effect on EcfE activity and hence P*lppE* expression. In *Streptomyces coelicolor*, the sigma factor σ^E is encoded in an operon consisting of four genes, including *cseA*, which encodes a lipoprotein that localises to the extracellular surface of the membrane. CseA negatively regulates expression of *sigE*, the gene encoding σ^E, possibly by interaction with another regulatory factor, CseC, or because loss of CseA under membrane stress conditions results in membrane destabilisation and induction of CseC [389]. Moreover, the well-studied *E. coli* anti-sigma factor RseA must be bound by the periplasmic protein RseB to stably bind σ^E [348], [387]. LppE may have a similar

role to CseA or RseB, interacting with other signalling factors or stabilising AsfE to prevent release of EcfE and expression from P*lppE* and its other target promoters. I showed that an interaction between LppE and the periplasmic domain of AsfE is possible, using modelling with AlphaFold2, and that the residues predicted to be involved in this interaction are conserved among *Rhizobium* strains. However, in order to confirm this interaction it would be necessary to observe the interaction *in vivo*, using a co-immunoprecipitation or bacterial two-hybrid experiment. An experiment of this type could shed light on the role of LppE and further understanding of how P*lppE* expression is regulated in response to the pea root glycoprotein, as well as mechanisms by which ECF sigma factors respond to environmental changes.

6.6.3 Characterisation of the P*lppE*-inducing glycoprotein

One of the most interesting aspects of the *ecfE* system is how expression of the operon is restricted to bacteria colonising the root elongation zone, because it is dependent on a specific concentration of a pea root surface glycoprotein. In this work, I identified several candidate glycoproteins that could be responsible for P*lppE* induction using mass spectrometry and narrowed down the possibilities to five candidates using available literature. However, it would provide an important insight into the mechanism by which P*lppE* expression is governed if the glycoprotein could be definitively identified. Throughout this work, I used a P*lppE*::*lux* fusion to examine the conditions under which P*lppE* is active. Preliminary data from the Poole lab shows that ground *Nicotiana benthamiana* leaves do not elicit P*lppE*-driven luminescence. Transient expression of glycoprotein candidates in *N. benthamiana* by agroinfiltration, followed by grinding of the leaves and use in a P*lppE*::*lux* fusion induction assay, could therefore be a promising method of testing the candidates individually and identifying which protein or proteins are responsible for P*lppE* activity in the root elongation zone. Identifying this and clarifying the role of LppE might also enable further work to elucidate how the extracellular signal from the plant glycoprotein is transmitted to AsfE, which is anchored in the inner membrane. This is currently the

largest gap remaining in our understanding of how the *ecfE* operon is regulated and filling it would provide a complete model for this branch of interkingdom signalling.

6.6.4 Link between iron and rhizoplane attachment and colonisation

A further interesting point raised by this work was the link between iron and attachment to and colonisation of plant rhizoplanes by Rlv3841. Iron availability clearly affects pea root attachment, as supplementing bacterial cell suspensions with iron (III) citrate during attachment assays results in significantly increased attachment. This is despite the downregulation of a number of Fe (III) transporters encoded by *irp6A6B6C*, RL3050-RL3052 and *fbpABC* during rhizoplane attachment. I showed that mutation of RL0777 (*rirA*) or RL2602 (encoding an Fe (II) permease) does not affect attachment to the pea rhizoplane by Rlv3841, despite RirA being a major regulator of iron uptake [390] and RL2602 being upregulated during attachment to the rhizoplane. This suggests that the iron limitation response pathway controlled by RirA does not play a role in rhizoplane attachment or colonisation. However, the *ecfE* system regulates *fbpA* and activates Rlv3841 iron uptake as part of an iron-related attachment pathway induced by contact with root elongation zones. Upregulation of *fbpA* in response to increased expression of *ecfE* on root elongation zones would not be seen in the transcriptomic data from whole root attachment presented in Chapter 4, as any localised increase in expression of *fbpA* would be diluted out by the reduced expression seen in cells attaching to other parts of the root. Generating a reporter fusion to the *fbpABC* promoter in Rlv3841 and inoculating the strain on plants grown on square plates, as shown in Chapter 5, would enable more precise examination of possible spatially specific expression of *fbpABC*. The iron-related attachment pathway is clearly separate from the RirA-regulated iron limitation response, although *fbpA* is involved in both. Further characterisation of this pathway, including the RL1136 and RL1137 gene products, which may be a target for the iron taken up by FbpABC, is clearly necessary. No effect has been documented of mutation of RL1136 or RL1137 on attachment based on INSeq screening [133]. However, RL1136 encodes an exported protein that could be shared between members of a bacterial community,

meaning that an effect of RL1136 mutation on attachment would not be seen in a competitive INSeq screen. Mutation of RL1136 and RL1137, as well as RL1138-RL1139 (*ecfC*, *asfC*) which are part of the operon, would allow root tip attachment assays to be carried out with the mutants in a single inoculum to confirm whether these genes are involved in iron-related attachment to the root elongation zone.

6.6.5 The roles of ECF sigma factors in Rlv3841 attachment and colonisation of plant rhizoplanes

In this work, I investigated the *ecfE* operon and its role in attachment to and colonisation of the rhizoplane by Rlv3841. However, the RNA sequencing experiments described in Chapters 3 and 4 show that *ecfE* is not the only sigma factor encoding gene that is differentially regulated during interactions with plant rhizoplanes. In fact, there are several other ECF sigma factors that warrant further investigation. In particular, *ecfZ* was the most strongly upregulated ECF sigma factor-encoding gene during both attachment to and colonisation of the pea rhizoplane. It is known to be upregulated in the *praR* mutant, which itself exhibits improved attachment and nodulation competitiveness compared with wild type [142], possibly due to relief of repression of *ecfZ* and hence its downstream targets. The most strongly upregulated ECF sigma factor-encoding gene during colonisation of the barley rhizoplane was *ecfR*. The *ecfR* homologue in *C. crescentus*, *sigF*, is involved in response to oxidative stress [227], while *ecfZ* and its *S. meliloti* homologue are expressed in response to heat and saline stress [228], [286]. The different levels of expression of these two sigma factors during colonisation of different plants therefore suggests that Rlv3841 is subject to different stresses on different plant rhizoplanes. Further investigation of when these genes are expressed and what their downstream targets are could aid with understanding why Rlv3841 is better adapted to the pea rhizoplane than to the barley rhizoplane, and whether it would be possible to engineer a 'cereal-adapted' *Rhizobium* strain, or what to look for in a naturally occurring strain that might be better adapted to cereal rhizoplanes.

6.6.6 Engineering plant rhizoplane-induced expression of *nifA*

Finally, one of the most applied follow up directions for this work would be to generate *nifA* fusions to the promoter regions of some of the genes identified in Chapter 3 that were substantially more strongly upregulated during colonisation of the barley rhizoplane than the pea rhizoplane. In particular, RL3594 (*dapA*) was strongly upregulated on barley and not differentially regulated on pea. *dapA* expression seems likely to be conserved as cereal-specific and therefore may be a useful target of promoter fusions in other rhizobial strains. As I have shown that Rlv3841 is poorly adapted to utilisation of carbon provided by barley roots, future work involving fusions of *nifA* to these target promoter regions would have to include identification of another strain that is better adapted to the barley rhizoplane. A *Rhizobium* strain that is a better coloniser of cereal rhizoplanes than Rlv3841 and that harbours a *dapA* gene that is expressed solely in response to cereal roots could be an excellent target of engineering.

Bibliography

[1] C. Smith, A. K. Hill, and L. Torrente-Murciano, “Current and future role of Haber–Bosch ammonia in a carbon-free energy landscape,” *Energy Environ Sci*, vol. 13, no. 2, pp. 331–344, 2020, doi: 10.1039/c9ee02873k.

[2] J. W. Erisman, M. A. Sutton, J. Galloway, Z. Klimont, and W. Winiwarter, “How a century of ammonia synthesis changed the world,” *Nat Geosci*, vol. 1, no. 10, pp. 636–639, 2008, doi: 10.1038/ngeo325.

[3] W. M. Stewart, D. W. Dibb, A. E. Johnston, and T. J. Smyth, “The contribution of commercial fertilizer nutrients to food production,” *Agron J*, vol. 97, no. 1, pp. 1–6, 2005, doi: 10.2134/agronj2005.0001.

[4] UN, “World Population Prospects.” Accessed: Nov. 15, 2023. [Online]. Available: https://population.un.org/wpp

[5] D. Tilman, C. Balzer, J. Hill, and B. L. Befort, “Global food demand and the sustainable intensification of agriculture,” *Proc Natl Acad Sci U S A*, vol. 108, no. 50, pp. 20260–20264, 2011, doi: 10.1073/pnas.1116437108.

[6] D. K. Ray, N. D. Mueller, P. C. West, and J. A. Foley, “Yield trends are insufficient to double global crop production by 2050,” *PLoS One*, vol. 8, no. 6, p. e66428, 2013, doi: 10.1371/journal.pone.0066428.

[7] M. Maheswari, A. N. G. Murthy, and A. K. Shanker, “Nitrogen nutrition in crops and its importance in crop quality,” in *The Indian Nitrogen Assessment: Sources of Reactive Nitrogen, Environmental and Climate Effects, Management Options, and Policies*, Elsevier, 2017, pp. 175–186. doi: 10.1016/B978-0-12-811836-8.00012-4.

[8] F. Gastal, G. Lemaire, J. L. Durand, and G. Louarn, “Quantifying crop responses to nitrogen and avenues to improve nitrogen-use efficiency,” in *Crop Physiology: Applications for Genetic*

Improvement and Agronomy: Second Edition, Elsevier Inc., 2015, pp. 161–206. doi: 10.1016/B978-0-12-417104-6.00008-X.

[9] J. Poore and T. Nemecek, "Reducing food's environmental impacts through producers and consumers," *Science (1979)*, vol. 360, no. 6392, pp. 987–992, 2018, doi: science.aaq0216.

[10] S. Menegat, A. Ledo, and R. Tirado, "Greenhouse gas emissions from global production and use of nitrogen synthetic fertilisers in agriculture," *Scientific Reports 2022 12:1*, vol. 12, no. 1, pp. 1–13, 2022, doi: 10.1038/s41598-022-18773-w.

[11] IPCC, "AR4 Climate Change 2007: The Physical Science Basis." Accessed: Mar. 16, 2023. [Online]. Available: https://www.ipcc.ch/report/ar4/wg1/

[12] A. R. Ravishankara, J. S. Daniel, and R. W. Portmann, "Nitrous oxide (N2O): The dominant ozone-depleting substance emitted in the 21st century," *Science (1979)*, vol. 326, no. 5949, pp. 123–125, Oct. 2009, doi: 10.1126/SCIENCE.1176985/SUPPL_FILE/RAVISHANKARA.SOM.PDF.

[13] E. A. Davidson *et al.*, "Excess nitrogen in the U.S. environment: trends, risks, and solutions," *Issues in Ecology*, vol. 15, pp. 1–16, 2012.

[14] I. Shcherbak, N. Millar, and G. P. Robertson, "Global metaanalysis of the nonlinear response of soil nitrous oxide (N2O) emissions to fertilizer nitrogen," *Proc Natl Acad Sci U S A*, vol. 111, no. 25, pp. 9199–9204, 2014, doi: 10.1073/pnas.1322434111.

[15] R. J. Diaz and R. Rosenberg, "Spreading dead zones and consequences for marine ecosystems," *Science (1979)*, vol. 321, no. 5891, pp. 926–929, 2008, doi: science.1156401.

[16] A. R. Fox *et al.*, "Major cereal crops benefit from biological nitrogen fixation when inoculated with the nitrogen-fixing bacterium *Pseudomonas protegens*," *Environ Microbiol*, vol. 18, no. 10, pp. 3522–3534, 2016, doi: 10.1111/1462-2920.13376.

[17] J. A. Downie, "Legume nodulation," *Current Biology*, vol. 24, no. 5, pp. R184–R190, 2014, doi: 10.1016/j.cub.2014.01.028.

[18] G. N. Gussin, C. W. Ronson, and F. M. Ausubel, “Regulation of nitrogen fixation genes,” *Annu Rev Genet*, vol. 20, no. 1, pp. 567–591, 1986, doi: 10.1146/annurev.ge.20.120186.003031.

[19] D. F. Herridge, M. B. Peoples, and R. M. Boddey, “Global inputs of biological nitrogen fixation in agricultural systems,” *Plant Soil*, vol. 311, no. 1, pp. 1–18, 2008, doi: 10.1007/s11104-008-9668-3.

[20] D. E. Soltis *et al.*, “Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms,” *Proc Natl Acad Sci U S A*, vol. 92, pp. 2647–2651, 1995.

[21] M. Udvardi and P. S. Poole, “Transport and metabolism in legume-rhizobia symbioses,” *Annu Rev Plant Biol*, vol. 64, no. 1, pp. 781–805, 2013, doi: 10.1146/annurev-arplant-050312-120235.

[22] M. L. Clawson, A. Bourret, and D. R. Benson, “Assessing the phylogeny of *Frankia*-actinorhizal plant nitrogen-fixing root nodule symbioses with *Frankia* 16S rRNA and glutamine synthetase gene sequences,” *Mol Phylogenet Evol*, vol. 31, no. 1, pp. 131–138, 2004, doi: 10.1016/j.ympev.2003.08.001.

[23] D. Wang, S. Yang, F. Tang, and H. Zhu, “Symbiosis specificity in the legume – rhizobial mutualism,” *Cell Microbiol*, vol. 14, no. 3, pp. 334–342, 2012, doi: 10.1111/J.1462-5822.2011.01736.X.

[24] J. A. Downie, “The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots,” *FEMS Microbiol Rev*, vol. 34, no. 2, pp. 150–170, 2010, doi: 10.1111/j.1574-6976.2009.00205.x.

[25] S. Goormachtig, W. Capoen, and M. Holsters, “*Rhizobium* infection: Lessons from the versatile nodulation behaviour of water-tolerant legumes,” *Trends Plant Sci*, vol. 9, no. 11, pp. 518–522, 2004, doi: 10.1016/j.tplants.2004.09.005.

[26] S. Goormachtig, W. Capoen, E. K. James, and M. Holsters, “Switch from intracellular to intercellular invasion during water stress-tolerant legume nodulation,” *Proc Natl Acad Sci U S A*, vol. 101, no. 16, pp. 6303–6308, 2004, doi: 10.1073/pnas.0401540101.

[27] D. J. Gage, “Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes,” *Microbiology and Molecular Biology Reviews*, vol. 68, no. 2, pp. 280–300, 2004, doi: 10.1128/mmbr.68.2.280-300.2004.

[28] K. Recourt, A. A. van Brussel, A. J. Driessen, and B. J. Lugtenberg, “Accumulation of a nod gene inducer, the flavonoid naringenin, in the cytoplasmic membrane of *Rhizobium leguminosarum* biovar viciae is caused by the pH-dependent hydrophobicity of naringenin.,” *J Bacteriol*, vol. 171, no. 8, pp. 4370–4377, 1989, doi: 10.1128/jb.171.8.4370-4377.1989.

[29] H. R. M. Schlaman, R. J. H. Okker, and B. J. J. Lugtenberg, “Regulation of nodulation gene expression by nodD in rhizobia,” *J Bacteriol*, vol. 174, no. 16, pp. 5177–5182, 1992, doi: 10.1128/JB.174.16.5177-5182.1992.

[30] H. P. Spaink, C. A. Wijffelman, E. Pees, R. J. H. Okker, and B. J. J. Lugtenberg, “*Rhizobium* nodulation gene nodD as a determinant of host specificity,” *Nature 1987 328:6128*, vol. 328, no. 6128, pp. 337–340, 1987, doi: 10.1038/328337a0.

[31] X. Perret, C. Staehelin, and W. J. Broughton, “Molecular basis of symbiotic promiscuity,” *Microbiology and Molecular Biology Reviews*, vol. 64, no. 1, pp. 180–201, 2000, doi: 10.1128/mmbr.64.1.180-201.2000.

[32] E. Limpens, C. Franken, P. Smit, J. Willemse, T. Bisseling, and R. Geurts, “LysM domain receptor kinases regulating rhizobial Nod factor-induced infection,” *Science (1979)*, vol. 302, no. 5645, pp. 630–633, 2003, doi: 10.1126/science.1090074.

[33] E. B. Madsen *et al.*, “A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals,” *Nature*, vol. 425, no. 6958, pp. 637–640, 2003, doi: 10.1038/nature02045.

[34] S. Kelly, S. Radutoiu, and J. Stougaard, “Legume LysM receptors mediate symbiotic and pathogenic signalling,” *Curr Opin Plant Biol*, vol. 39, pp. 152–158, 2017, doi: 10.1016/j.pbi.2017.06.013.

[35] G. E. D. Oldroyd and J. A. Downie, “Calcium, kinases and nodulation signalling in legumes,” *Nature Reviews Molecular Cell Biology 2004 5:7*, vol. 5, no. 7, pp. 566–576, 2004, doi: 10.1038/nrm1424.

[36] Z. Bozsoki *et al.*, “Ligand-recognizing motifs in plant LysM receptors are major determinants of specificity,” *Science (1979)*, vol. 369, no. 6504, pp. 663–670, Aug. 2020, doi: 10.1126/science.abb3377.

[37] H. Miwa, J. Sun, G. E. D. Oldroyd, and J. A. Downie, “Analysis of Nod-factor-induced calcium signalling in root hairs of symbiotically defective mutants of *Lotus japonicus*,” *Molecular Plant-Microbe Interactions*, vol. 19, no. 8, pp. 914–923, 2006, doi: 10.1094/MPMI-19-0914.

[38] L. Schauser, A. Roussis, J. Stiller, and J. Stougaard, “A plant regulator controlling development of symbiotic root nodules,” *Nature*, vol. 402, no. 6758, pp. 191–195, 1999, doi: 10.1038/46058.

[39] J. F. Marsh *et al.*, “*Medicago truncatula* NIN is essential for rhizobial-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase,” *Plant Physiol*, vol. 144, no. 1, pp. 324–335, 2007, doi: 10.1104/pp.106.093021.

[40] C. Laffont, A. Ivanovici, P. Gautrat, M. Brault, M. A. Djordjevic, and F. Frugier, “The NIN transcription factor coordinates CEP and CLE signaling peptides that regulate nodulation antagonistically,” *Nature Communications 2020 11:1*, vol. 11, no. 1, pp. 1–13, 2020, doi: 10.1038/s41467-020-16968-1.

[41] T. Wang *et al.*, “Light-induced mobile factors from shoots regulate *Rhizobium*-triggered soybean root nodulation,” *Science (1979)*, vol. 374, no. 6563, pp. 65–71, 2021, doi: 10.1126/science.abh2890.

[42] J. Feng, T. Lee, K. Schiessl, and G. E. D. Oldroyd, “Processing of NODULE INCEPTION controls the transition to nitrogen fixation in root nodules,” *Science (1979)*, vol. 374, no. 6567, pp. 629–632, 2021, doi: 10.1126/science.abg2804.

[43] T. Hayashi, Y. Shimoda, S. Sato, S. Tabata, H. Imaizumi-Anraku, and M. Hayashi, “Rhizobial infection does not require cortical expression of upstream common symbiosis genes responsible for the induction of Ca(2+) spiking,” *The Plant Journal*, vol. 77, no. 1, pp. 146–159, 2014, doi: 10.1111/tpj.12374.

[44] B. J. Ferguson *et al.*, “Molecular analysis of legume nodule development and autoregulation,” *J Integr Plant Biol*, vol. 52, no. 1, pp. 61–76, 2010, doi: 10.1111/j.1744-7909.2010.00899.x.

[45] K. Schiessl *et al.*, “NODULE INCEPTION recruits the lateral root developmental program for symbiotic nodule organogenesis in *Medicago truncatula*,” *Current Biology*, vol. 29, no. 21, pp. 3657–3668, 2019, doi: 10.1016/j.cub.2019.09.005.

[46] J. Prell, J. P. White, A. Bourdes, S. Bunnewell, R. J. Bongaerts, and P. S. Poole, “Legumes regulate *Rhizobium* bacteroid development and persistence by the supply of branched-chain amino acids,” *Proc Natl Acad Sci U S A*, vol. 106, no. 30, pp. 12477–12482, 2009, doi: 10.1073/pnas.0903653106.

[47] C. A. Appleby, “Leghemoglobin and *Rhizobium* respiration,” *Ann. Rev. Plant Physiol*, vol. 35, pp. 443–478, 1984, doi: 10.1146/annurev.pp.35.060184.002303.

[48] P. Poole, V. Ramachandran, and J. Terpolilli, “Rhizobia: from saprophytes to endosymbionts,” *Nat Rev Microbiol*, vol. 16, no. 5, pp. 291–303, 2018, doi: 10.1038/nrmicro.2017.171.

[49] E. T. Kiers, R. A. Rousseau, S. A. West, and R. F. Denlson, “Host sanctions and the legume–*Rhizobium* mutualism,” *Nature*, vol. 425, no. 6953, pp. 78–81, 2003, doi: 10.1038/nature01931.

[50] A. Westhoek *et al.*, “Conditional sanctioning in a legume-*Rhizobium* mutualism,” *Proc Natl Acad Sci U S A*, vol. 118, no. 19, p. e2025760118, 2021, doi: 10.1073/pnas.2025760118.

[51] K. G. Cassman, “Ecological intensification of cereal production systems: Yield potential, soil quality, and precision agriculture,” *Proc Natl Acad Sci U S A*, vol. 96, no. 11, pp. 5952–5959, 1999, doi: 10.1073/pnas.96.11.5952.

[52] White KD, “Fallowing, crop rotation and crop yields in Roman times,” *Agric Hist*, vol. 44, pp. 281–290, 1970.

[53] N. K. Fukagawa and L. H. Ziska, “Rice: importance for global nutrition,” *J Nutr Sci Vitaminol (Tokyo)*, vol. 65, pp. 2–3, 2019, doi: 10.3177/jnsv.65.S2.

[54] D. Dent and E. Cocking, “Establishing symbiotic nitrogen fixation in cereals and other non-legume crops: the greener Nitrogen Revolution,” *Agric Food Secur*, vol. 6, no. 7, pp. 1–9, 2017, doi: 10.1186/s40066-016-0084-2.

[55] B. A. Geddes *et al.*, “Engineering transkingdom signalling in plants to control gene expression in rhizosphere bacteria,” *Nat Commun*, vol. 10, no. 1, 2019, doi: 10.1038/s41467-019-10882-x.

[56] M. Rosenblueth *et al.*, “Nitrogen fixation in cereals,” *Front Microbiol*, vol. 9, no. 1794, pp. 1–19, 2018, doi: 10.3389/fmicb.2018.01794.

[57] S. Burén, X. Jiang, G. López-Torrejón, C. Echavarri-Erasun, and L. M. Rubio, “Purification and in vitro activity of mitochondria targeted nitrogenase cofactor maturase NifB,” *Front Plant Sci*, vol. 8, no. 1567, pp. 1–16, 2017, doi: 10.3389/fpls.2017.01567.

[58] R. S. Allen *et al.*, “Expression of 16 nitrogenase proteins within the plant mitochondrial matrix,” *Front Plant Sci*, vol. 8, no. 287, pp. 1–14, 2017, doi: 10.3389/fpls.2017.00287.

[59] K. Markmann and M. Parniske, “Evolution of root endosymbiosis with bacteria: how novel are nodules?,” *Trends Plant Sci*, vol. 14, no. 2, pp. 77–86, 2009, doi: 10.1016/j.tplants.2008.11.009.

[60] C. Rogers and G. E. D. Oldroyd, “Synthetic biology approaches to engineering the nitrogen symbiosis in cereals,” *J Exp Bot*, vol. 65, no. 8, pp. 1939–1946, 2014, doi: 10.1093/jxb/eru098.

[61] M. Banba *et al.*, “Divergence of evolutionary ways among common sym genes: CASTOR and CCaMK show functional conservation between two symbiosis systems and constitute the root of a common signaling pathway,” *Plant Cell Physiol*, vol. 49, no. 11, pp. 1659–1671, 2008, doi: 10.1093/pcp/pcn153.

[62] C. Gleason, S. Chaudhuri, T. Yang, A. Muñoz, B. W. Poovaiah, and G. E. D. Oldroyd, “Nodulation independent of rhizobia induced by a calcium-activated kinase lacking autoinhibition,” *Nature*, vol. 441, no. 7097, pp. 1149–1152, 2006, doi: 10.1038/nature04812.

[63] U. K. Bageshwar *et al.*, “An environmentally friendly engineered *Azotobacter* strain that replaces a substantial amount of urea fertilizer while sustaining the same wheat yield,” *Appl Environ Microbiol*, vol. 83, no. 15, p. e00590, 2017, doi: 10.1128/AEM.00590-17.

[64] M.-H. Ryu *et al.*, “Control of nitrogen fixation in bacteria that associate with cereals,” *Nat Microbiol*, vol. 5, no. 2, pp. 314–330, 2019, doi: 10.1038/s41564-019-0631-2.

[65] C. Christiansen-Weniger and J. A. Van Veen, “NH4+-excreting *Azospirillum brazilense* mutants enhance the nitrogen supply of a wheat host,” *Appl Environ Microbiol*, vol. 57, no. 10, pp. 3006–3012, 1991, doi: 10.1128/aem.57.10.3006-3012.1991.

[66] L. Hiltner, “Über neuere Erfahrungen und Probleme auf dem Gebiete der Bodenbakteriologie unter besonderer Berücksichtigung der Gründüngung und Brache,” *Arbeiten der Deutschen Landwirtschaftlichen Gesellschaft*, vol. 98, pp. 59–78, 1904.

[67] J. Whipps, “Carbon Economy,” in *The Rhizosphere*, J. Lynch, Ed., Essex: John Wiley & Sons Ltd, 1990, pp. 59–97.

[68] R. Mendes, P. Garbeva, and J. M. Raaijmakers, “The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms,” *FEMS Microbiol Rev*, vol. 37, no. 5, pp. 634–663, 2013, doi: 10.1111/1574-6976.12028.

[69] S. Hacquard *et al.*, “Microbiota and host nutrition across plant and animal kingdoms,” *Cell Host Microbe*, vol. 17, no. 5, pp. 603–616, May 2015, doi: 10.1016/j.chom.2015.04.009.

[70] K. Hartman, M. G. A. van der Heijden, V. Roussely-Provent, J. C. Walser, and K. Schlaeppi, "Deciphering composition and function of the root microbiome of a legume plant," *Microbiome*, vol. 5, no. 2, pp. 1–13, Jan. 2017, doi: 10.1186/s40168-016-0220-z.

[71] S. P. Brown, M. A. Grillo, J. C. Podowski, and K. D. Heath, "Soil origin and plant genotype structure distinct microbiome compartments in the model legume *Medicago truncatula*," *Microbiome 2020 8:1*, vol. 8, no. 1, pp. 1–17, 2020, doi: 10.1186/S40168-020-00915-9.

[72] R. Zgadzaj, R. Garrido-Oter, D. B. Jensen, A. Koprivova, P. Schulze-Lefert, and S. Radutoiu, "Root nodule symbiosis in *Lotus japonicus* drives the establishment of distinctive rhizosphere, root, and nodule bacterial communities," *Proc Natl Acad Sci U S A*, vol. 113, no. 49, pp. E7996–E8005, 2016, doi: 10.1073/pnas.1616564113.

[73] J. Cordero, J. R. de Freitas, and J. J. Germida, "Bacterial microbiome associated with the rhizosphere and root interior of crops in Saskatchewan, Canada," *Can J Microbiol*, vol. 66, no. 1, pp. 71–85, 2020, doi: 10.1139/cjm-2019-0330.

[74] A. Tkacz, R. Ledermann, A. Martyn, S. Schornack, G. E. D. Oldroyd, and P. S. Poole, "Nodulation and nitrogen fixation in *Medicago truncatula* strongly alters the abundance of its root microbiota and subtly affects its structure," *Environ Microbiol*, vol. 24, no. 11, pp. 5524–5533, 2022, doi: 10.1111/1462-2920.16164.

[75] R. L. Berendsen, C. M. J. Pieterse, and P. A. H. M. Bakker, "The rhizosphere microbiome and plant health," *Trends Plant Sci*, vol. 17, no. 8, pp. 1360–1385, 2012, doi: 10.1016/j.tplants.2012.04.001.

[76] H. P. Bais, T. L. Weir, L. G. Perry, S. Gilroy, and J. M. Vivanco, "The role of root exudates in rhizosphere interactions with plants and other organisms," *Annu Rev Plant Biol*, vol. 57, pp. 233–66, 2006, doi: 10.1146/annurev.arplant.57.032905.105159.

[77] G. Berg and K. Smalla, "Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere," *FEMS Microbiol Ecol*, vol. 68, pp. 1–13, 2009, doi: 10.1111/j.1574-6941.2009.00654.x.

[78] R. Miethling, G. Wieland, H. Backhaus, and C. C. Tebbe, “Variation of microbial rhizosphere communities in response to crop species, soil origin, and inoculation with *Sinorhizobium meliloti* L33,” *Microb Ecol*, vol. 40, no. 1, pp. 43–56, 2000, doi: 10.1007/s002480000021.

[79] M. Viebahn, C. Veenman, K. Wernars, L. C. Van Loon, E. Smit, and P. A. H. M. Bakker, “Assessment of differences in ascomycete communities in the rhizosphere of field-grown wheat and potato,” *FEMS Microbiol Ecol*, vol. 53, no. 2, pp. 245–253, 2005, doi: 10.1016/j.femsec.2004.12.014.

[80] P. Garbeva, J. D. Van Elsas, and J. A. Van Veen, “Rhizosphere microbial community and its response to plant species and soil history,” *Plant Soil*, vol. 302, no. 1, pp. 19–32, 2008, doi: 10.1007/s11104-007-9432-0.

[81] T. R. Turner *et al.*, “Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants,” *ISME J*, vol. 7, no. 12, pp. 2248–2258, 2013, doi: 10.1038/ismej.2013.119.

[82] V. Lakshmanan, G. Selvaraj, and H. P. Bais, “Functional soil microbiome: belowground solutions to an aboveground problem,” *Plant Physiol*, vol. 166, no. 2, pp. 689–700, 2014, doi: 10.1104/pp.114.245811.

[83] J. M. Chaparro, D. V. Badri, M. G. Bakker, A. Sugiyama, D. K. Manter, and J. M. Vivanco, “Root exudation of phytochemicals in *Arabidopsis* follows specific patterns that are developmentally programmed and correlate with soil microbial functions,” *PLoS One*, vol. 8, no. 2, p. e55731, 2013, doi: 10.1371/journal.pone.0055731.

[84] J. M. Chaparro, D. V. Badri, and J. M. Vivanco, “Rhizosphere microbiome assemblage is affected by plant development,” *ISME J*, vol. 8, no. 4, pp. 790–803, 2014, doi: 10.1038/ismej.2013.196.

[85] F. E. Z. Haichar *et al.*, “Plant host habitat and root exudates shape soil bacterial community structure,” *ISME J*, vol. 2, no. 12, pp. 1221–1230, 2008, doi: 10.1038/ismej.2008.80.

[86] A. L. Neal, S. Ahmad, R. Gordon-Weeks, and J. Ton, “Benzoxazinoids in root exudates of maize attract *Pseudomonas putida* to the rhizosphere,” *PLoS One*, vol. 7, no. 4, p. e35498, 2012, doi: 10.1371/journal.pone.0035498.

[87] F. Kamilova, L. V. Kravchenko, A. I. Shaposhnikov, T. Azarova, N. Makarova, and B. Lugtenberg, “Organic acids, sugars, and L-tryptophane in exudates of vegetables growing on stonewool and their effects on activities of rhizosphere bacteria,” *Molecular plant-microbe interactions*, vol. 19, no. 3, pp. 250–256, 2006, doi: 10.1094/MPMI-19-0250.

[88] Z. Li and S. K. Nair, “Quorum sensing: How bacteria can coordinate activity and synchronize their response to external signals?,” *Protein Science*, vol. 21, no. 10, pp. 1403–1417, 2012, doi: 10.1002/pro.2132.

[89] M. Teplitski, J. B. Robinson, and W. D. Bauer, “Plants secrete substances that mimic bacterial N-acyl homoserine lactone signal activities and affect population density-dependent behaviors in associated bacteria,” *Molecular Plant-Microbe Interactions*, vol. 13, no. 6, pp. 637–648, 2000, doi: 10.1094/MPMI.2000.13.6.637.

[90] M. Gao, M. Teplitski, J. B. Robinson, and W. D. Bauer, “Production of substances by *Medicago truncatula* that affect bacterial quorum sensing,” *Molecular Plant-Microbe Interactions*, vol. 16, no. 9, pp. 827–834, 2003, doi: 10.1094/MPMI.2003.16.9.827.

[91] M. Teplitski *et al.*, “*Chlamydomonas reinhardtii* secretes compounds that mimic bacterial signals and interfere with quorum sensing regulation in bacteria,” *Plant Physiol*, vol. 134, no. 1, pp. 137–146, 2004, doi: 10.1104/pp.103.029918.

[92] F. Pérez-Montaño *et al.*, “Rice and bean AHL-mimic quorum-sensing signals specifically interfere with the capacity to form biofilms by plant-associated bacteria,” *Res Microbiol*, vol. 164, no. 7, pp. 749–760, Sep. 2013, doi: 10.1016/j.resmic.2013.04.001.

[93] G. Degrassi, G. Devescovi, R. Solis, L. Steindler, and V. Venturi, “*Oryza sativa* rice plants contain molecules that activate different quorum-sensing N-acyl homoserine lactone biosensors and

are sensitive to the specific AiiA lactonase," *FEMS Microbiol Lett*, vol. 269, no. 2, pp. 213–220, Apr. 2007, doi: 10.1111/j.1574-6968.2006.00624.x.

[94] S. Ferluga and V. Venturi, "OryR is a LuxR-family protein involved in interkingdom signalling between pathogenic *Xanthomonas oryzae* pv. oryzae and rice," *J Bacteriol*, vol. 191, no. 3, p. 890, 2009, doi: 10.1128/JB.01507-08.

[95] J. W. Yang *et al.*, "Whitefly infestation of pepper plants elicits defence responses against bacterial pathogens in leaves and roots and changes the below-ground microflora," *Journal of Ecology*, vol. 99, no. 1, pp. 46–56, 2011, doi: 10.1111/j.1365-2745.2010.01756.x.

[96] B. Lee, S. Lee, and C. M. Ryu, "Foliar aphid feeding recruits rhizosphere bacteria and primes plant immunity against pathogenic and non-pathogenic bacteria in pepper," *Ann Bot*, vol. 110, no. 2, pp. 281–290, Jul. 2012, doi: 10.1093/aob/mcs055.

[97] T. Rudrappa, K. J. Czymmek, P. W. Paré, and H. P. Bais, "Root-secreted malic acid recruits beneficial soil bacteria," *Plant Physiol*, vol. 148, no. 3, pp. 1547–1556, Nov. 2008, doi: 10.1104/pp.108.127613.

[98] J. L. Dangl and J. D. G. Jones, "Plant pathogens and integrated defence responses to infection," *Nature*, vol. 411, no. 6839, pp. 826–833, 2001, doi: 10.1038/35081161.

[99] L. C. Carvalhais, P. G. Dennis, D. V. Badri, G. W. Tyson, J. M. Vivanco, and P. M. Schenk, "Activation of the jasmonic acid plant defence pathway alters the composition of rhizosphere bacterial communities," *PLoS One*, vol. 8, no. 2, p. e56457, Feb. 2013, doi: 10.1371/journal.pone.0056457.

[100] D. V. Badri *et al.*, "An ABC transporter mutation alters root exudation of phytochemicals that provoke an overhaul of natural soil microbiota," *Plant Physiol*, vol. 151, no. 4, pp. 2006–2017, 2009, doi: 10.1104/pp.109.147462.

[101] M. Bressan *et al.*, "Exogenous glucosinolate produced by *Arabidopsis thaliana* has an impact on microbes in the rhizosphere and plant roots," *ISME J*, vol. 3, no. 11, pp. 1243–1257, Jun. 2009, doi: 10.1038/ismej.2009.68.

[102] S. Ikeda *et al.*, "The genotype of the calcium/calmodulin-dependent protein kinase gene (CCaMK) determines bacterial community diversity in rice roots under paddy and upland field conditions," *Appl Environ Microbiol*, vol. 77, no. 13, pp. 4399–4405, Jul. 2011, doi: 10.1128/aem.00315-11.

[103] S. Nakamura and T. Minamino, "Flagella-driven motility of bacteria," *Biomolecules*, vol. 9, no. 7, pp. 279–302, 2019, doi: 10.3390/biom9070279.

[104] J. S. Mattick, "Type IV pili and twitching motility," *Annu Rev Microbiol*, vol. 56, pp. 289–314, 2002, doi: 10.1146/annurev.micro.56.012302.160938.

[105] K. K. Compton and B. E. Scharf, "Rhizobial chemoattractants, the taste and preferences of legume symbionts," *Front Plant Sci*, vol. 12, no. 686465, 2021, doi: 10.3389/FPLS.2021.686465.

[106] A. Buchan, B. Crombie, and G. M. Alexandre, "Temporal dynamics and genetic diversity of chemotactic-competent microbial populations in the rhizosphere," *Environ Microbiol*, vol. 12, no. 12, pp. 3171–3184, 2010, doi: 10.1111/J.1462-2920.2010.02290.X.

[107] J. P. W. Young *et al.*, "The genome of *Rhizobium leguminosarum* has recognizable core and accessory components," *Genome Biol*, vol. 7, no. 4, pp. 1–20, 2006, doi: 10.1186/GB-2006-7-4-R34.

[108] W. W. Currier and G. A. Strobel, "Chemotaxis of *Rhizobium* spp. to plant root exudates," *Plant Physiol*, vol. 57, pp. 820–823, 1976, doi: 10.1104/pp.57.5.820.

[109] J. P. Armitage, A. Gallagher, and A. W. B. Johnston, "Comparison of the chemotactic behaviour of *Rhizobium leguminosarum* with and without the nodulation plasmid," *Mol Microbiol*, vol. 2, no. 6, pp. 743–748, 1988, doi: 10.1111/j.1365-2958.1988.tb00085.x.

[110] E. T. Gaworzewska and M. J. Carlile, "Positive chemotaxis of *Rhizobium leguminosarum* and other bacteria towards root exudates from legumes and other plants," *J Gen Microbiol*, vol. 128, no. 6, pp. 1179–1188, 1982, doi: 10.1099/00221287-128-6-1179.

[111] A. J. Dharmatilake and W. D. Bauer, “Chemotaxis of *Rhizobium meliloti* towards nodulation gene-inducing compounds from alfalfa roots,” *Appl Environ Microbiol*, vol. 58, no. 4, pp. 1153–1158, 1992, doi: 10.1128/aem.58.4.1153-1158.1992.

[112] R. M. Wheatley and P. S. Poole, “Mechanisms of bacterial attachment to roots,” *FEMS Microbiol Rev*, vol. 42, no. 4, pp. 448–461, 2018, doi: 10.1093/femsre/fuy014.

[113] M. A.-S. Vigeant, R. M. Ford, M. Wagner, and L. K. Tamm, “Reversible and irreversible adhesion of motile *Escherichia coli* cells analyzed by total internal reflection aqueous fluorescence microscopy,” *Appl Environ Microbiol*, vol. 68, no. 6, pp. 2794–2801, 2002, doi: 10.1128/aem.68.6.2794-2801.2002.

[114] K. Kendall and A. D. Roberts, “van der Waals forces influencing adhesion of cells,” *Philosophical Transactions B*, vol. 370, p. 20140078, 2015, doi: 10.1098/rstb.2014.0078.

[115] M. C. M. Van Loosdrecht, W. Norde, and A. J. B. Zehnder, “Physical chemical description of bacterial adhesion,” *J Biomater Appl*, vol. 5, no. 2, pp. 91–106, 1990, doi: 10.1177/088532829000500202.

[116] E. Giaouris, M.-P. Chapot-Chartier, and R. Briandet, “Surface physicochemical analysis of natural *Lactococcus lactis* strains reveals the existence of hydrophobic and low charged strains with altered adhesive properties,” *Int J Food Microbiol*, vol. 131, pp. 2–9, 2008, doi: 10.1016/j.ijfoodmicro.2008.09.006.

[117] C. L. Croes, S. Moens, E. Van Bastelaere, J. Vanderleyden, and K. W. Michiels, “The polar flagellum mediates *Azospirillum brasilense* adsorption to wheat roots,” *J Gen Microbiol*, vol. 139, no. 9, pp. 2261–2269, 1993, doi: 10.1099/00221287-139-9-2261.

[118] F. A. Rossi, D. B. Medeot, J. P. Liaudat, M. Pistorio, and E. Jofré, “In *Azospirillum brasilense*, mutations in flmA or flmB genes affect polar flagellum assembly, surface polysaccharides, and attachment to maize roots,” *Microbiol Res*, vol. 190, pp. 55–62, 2016, doi: 10.1016/j.micres.2016.05.006.

[119] J. Dörr, T. Hurek, and B. Reinhold-Hurek, “Type IV pili are involved in plant–microbe and fungus–microbe interactions,” *Mol Microbiol*, vol. 30, no. 1, pp. 7–17, 1998, doi: 10.1046/j.1365-2958.1998.01010.x.

[120] C. Berne, A. Ducret, G. G. Hardy, and Y. V Brun, “Adhesins involved in attachment to abiotic surfaces by Gram-negative bacteria,” *Microbiol Spectr*, vol. 3, no. 4, 2015, doi: 10.1128/microbiolspec.MB-0018-2015.

[121] H. E. Knights, B. Jorrin, T. L. Haskett, and P. S. Poole, “Deciphering bacterial mechanisms of root colonization,” *Environ Microbiol Rep*, vol. 13, no. 4, pp. 428–444, 2021, doi: 10.1111/1758-2229.12934.

[122] S. J. Vesper and W. D. Bauer, “Role of pili (fimbriae) in attachment of *Bradyrhizobium japonicum* to soybean roots,” *Appl Environ Microbiol*, vol. 52, no. 1, pp. 134–141, 1986, doi: 10.1128/aem.52.1.134-141.1986.

[123] B. B. Bohlool and E. L. Schmidt, “Lectins: a possible basis for specificity in the *Rhizobium*—legume root nodule symbiosis,” *Science (1979)*, vol. 185, no. 4147, pp. 269–271, 1974, doi: 10.1126/science.185.4147.269.

[124] F. B. Dazzo and D. H. Hubbell, “Cross-reactive antigens and lectin as determinants of symbiotic specificity in the *Rhizobium*-clover association,” *Appl Microbiol*, vol. 30, no. 6, pp. 1017–1033, 1975, doi: 10.1128/am.30.6.1017-1033.1975.

[125] F. B. Dazzo, C. A. Napoli, and D. H. Hubbell, “Adsorption of bacteria to roots as related to host specificity in the *Rhizobium*-clover symbiosis,” *Appl Environ Microbiol*, vol. 32, no. 1, pp. 166–171, 1976, doi: 10.1128/aem.32.1.166-171.1976.

[126] C. L. Díaz, L. S. Melchers, P. J. J. Hooykaas, B. J. J. Lugtenberg, and J. W. Kijne, “Root lectin as a determinant of host–plant specificity in the *Rhizobium*–legume symbiosis,” *Nature 1989 338:6216*, vol. 338, no. 6216, pp. 579–581, 1989, doi: 10.1038/338579a0.

[127] M. C. Laus, T. J. Logman, G. E. Lamers, A. A. N. Van Brussel, R. W. Carlson, and J. W. Kijne, "A novel polar surface polysaccharide from *Rhizobium leguminosarum* binds host plant lectin," *Mol Microbiol*, vol. 59, no. 6, pp. 1704–1713, 2006, doi: 10.1111/j.1365-2958.2006.05057.x.

[128] A. Williams, A. Wilkinson, M. Krehenbrink, D. M. Russo, A. Zorreguieta, and J. A. Downie, "Glucomannan-mediated attachment of *Rhizobium leguminosarum* to pea root hairs is required for competitive nodule infection," *J Bacteriol*, vol. 190, no. 13, pp. 4706–4715, 2008, doi: 10.1128/jb.01694-07.

[129] J. T. Loh, S. C. Ho, A. W. De Feijter, J. L. Wang, and M. Schindler, "Carbohydrate binding activities of *Bradyrhizobium japonicum*: unipolar localization of the lectin BJ38 on the bacterial cell surface.," *Proc Natl Acad Sci U S A*, vol. 90, no. 7, pp. 3033–3037, 1993, doi: 10.1073/pnas.90.7.3033.

[130] S. -C Ho, J. L. Wang, M. Schindler, and J. T. Loh, "Carbohydrate binding activities of *Bradyrhizobium japonicum*. III. Lectin expression, bacterial binding, and nodulation efficiency," *The Plant Journal*, vol. 5, no. 6, pp. 873–884, 1994, doi: 10.1046/j.1365-313x.1994.5060873.x.

[131] J. P. Wisniewski and F. M. Delmotte, "Modulation of carbohydrate-binding capacities and attachment ability of *Bradyrhizobium* sp. (lupinus) to white lupin roots," *Can J Microbiol*, vol. 42, no. 3, pp. 234–242, 1996, doi: 10.1139/m96-035.

[132] G. Smit, T. J. Logman, M. E. Boerrigter, J. W. Kijne, and B. J. Lugtenberg, "Purification and partial characterization of the *Rhizobium leguminosarum* biovar viciae Ca2+-dependent adhesin, which mediates the first step in attachment of cells of the family Rhizobiaceae to plant root hair tips," *J Bacteriol*, vol. 171, no. 7, pp. 4054–4062, 1989, doi: 10.1128/jb.171.7.4054-4062.1989.

[133] J. Parsons *et al.*, "Factors governing attachment of *Rhizobium leguminosarum* to legume roots," *BioRxiv*, 2022, doi: 10.1101/2022.11.21.517457.

[134] N. Ausmees, K. Jacobsson, and M. Lindberg, "A unipolarly located, cell-surface-associated agglutinin, RapA, belongs to a family of *Rhizobium*-adhering proteins (Rap) in *Rhizobium*

leguminosarum bv. trifolii," *Microbiology (Reading)*, vol. 147, pp. 549–559, 2001, doi: 10.1099/00221287-147-3-549.

[135] G. Smit, D. M. J. Tubbing, J. W. Kijne, and B. J. J. Lugtenberg, "Role of Ca2+ in the activity of rhicadhesin from *Rhizobium leguminosarum* biovar viciae, which mediates the first step in attachment of Rhizobiaceae cells to plant root hair tips," *Arch Microbiol*, vol. 155, no. 3, pp. 278–283, 1991, doi: 10.1007/bf00252212.

[136] M. A. Thompson, M. C. Onyeziri, and C. Fuqua, "Function and regulation of *Agrobacterium tumefaciens* cell surface structures that promote attachment," *Curr Top Microbiol Immunol*, vol. 418, pp. 143–184, 2018, doi: 10.1007/82_2018_96.

[137] N. Ausmees, H. Jonsson, S. Höglund, H. Ljunggren, and M. Lindberg, "Structural and putative regulatory genes involved in cellulose synthesis in *Rhizobium leguminosarum* bv. trifolii," *Microbiology (Reading)*, vol. 145 (Pt 5), no. 5, pp. 1253–1262, 1999, doi: 10.1099/13500872-145-5-1253.

[138] A. G. Matthysse, "Role of bacterial cellulose fibrils in *Agrobacterium tumefaciens* infection," *J Bacteriol*, vol. 154, no. 2, pp. 906–915, 1983, doi: 10.1128/jb.154.2.906-915.1983.

[139] G. Smit, J. W. Kijne, and B. J. Lugtenberg, "Involvement of both cellulose fibrils and a Ca2+-dependent adhesin in the attachment of *Rhizobium leguminosarum* to pea root hair tips," *J Bacteriol*, vol. 169, no. 9, pp. 4294–4301, 1987, doi: 10.1128/jb.169.9.4294-4301.1987.

[140] M. Janczarek, J. Kutkowska, T. Piersiak, and A. Skorupska, "*Rhizobium leguminosarum* bv. trifolii rosR is required for interaction with clover, biofilm formation and adaptation to the environment," *BMC Microbiol*, vol. 10, no. 284, pp. 1–23, 2010, doi: 10.1186/1471-2180-10-284.

[141] D. M. Russo *et al.*, "Proteins exported via the PrsD-PrsE type I secretion system and the acidic exopolysaccharide are involved in biofilm formation by *Rhizobium leguminosarum*," *J Bacteriol*, vol. 188, no. 12, pp. 4474–4486, 2006, doi: 10.1128/jb.00246-06.

[142] M. Frederix *et al.*, "Mutation of praR in *Rhizobium leguminosarum* enhances root biofilms, improving nodulation competitiveness by increased expression of attachment proteins," *Mol Microbiol*, vol. 93, no. 3, pp. 464–478, 2014, doi: 10.1111/mmi.12670.

[143] M. E. Davey and G. A. O'Toole, "Microbial biofilms: from ecology to molecular genetics," *Microbiology and Molecular Biology Reviews*, vol. 64, no. 4, p. 847, 2000, doi: 10.1128/mmbr.64.4.847-867.2000.

[144] J. W. Costerton *et al.*, "Bacterial biofilms in nature and disease," *Annu Rev Microbiol*, vol. 41, pp. 435–464, 1987, doi: 10.1146/annurev.mi.41.100187.002251.

[145] L. V Rinaudi, W. Giordano, L. V Rinaudi, W. Giordano, and M. S. Khan, "Bacterial biofilms: role in *Rhizobium*–legume symbiosis," in *Microbes for Legume Improvement*, Springer, Vienna, 2010, pp. 325–335. doi: 10.1007/978-3-211-99753-6_13.

[146] J. W. Costerton, Z. Lewandowski, D. E. Caldwell, D. R. Korber, and H. M. Lappin-Scott, "Microbial biofilms," *Annu Rev Microbiol*, vol. 49, pp. 711–745, 1995, doi: 10.1146/annurev.mi.49.100195.003431.

[147] N. R. Stanley and B. A. Lazazzera, "Environmental signals and regulatory pathways that influence biofilm formation," *Mol Microbiol*, vol. 52, no. 4, pp. 917–924, 2004, doi: 10.1111/j.1365-2958.2004.04036.x.

[148] H. C. Flemming and J. Wingender, "The biofilm matrix," *Nat Rev Microbiol*, vol. 8, no. 9, pp. 623–633, 2010, doi: 10.1038/nrmicro2415.

[149] N. A. Fujishige, N. N. Kapadia, P. L. De Hoff, and A. M. Hirsch, "Investigations of *Rhizobium* biofilm formation," *FEMS Microbiol Ecol*, vol. 56, no. 2, pp. 195–206, 2006, doi: 10.1111/j.1574-6941.2005.00044.x.

[150] F. Vaningelgem *et al.*, "Biodiversity of exopolysaccharides produced by *Streptococcus thermophilus* strains is reflected in their production and their molecular and functional characteristics," *Appl Environ Microbiol*, vol. 70, no. 2, pp. 900–912, 2004, doi: 10.1128/AEM.70.2.900-912.2004.

[151] L. V Rinaudi and W. Giordano, “An integrated view of biofilm formation in rhizobia,” *FEMS Microbiol Lett*, vol. 304, pp. 1–11, 2010, doi: 10.1111/j.1574-6968.2009.01840.x.

[152] L. V Rinaudi and J. E. González, “The low-molecular-weight fraction of exopolysaccharide II from *Sinorhizobium meliloti* is a crucial determinant of biofilm formation,” *J Bacteriol*, vol. 191, no. 23, pp. 7216–7224, 2009, doi: 10.1128/jb.01063-09.

[153] D. Borthakur, R. F. Barker, J. W. Latchford, L. Rossen, and A. W. B. Johnston, “Analysis of pss genes of *Rhizobium leguminosarum* required for exopolysaccharide synthesis and nodulation of peas: Their primary structure and their interaction with psi and other nodulation genes,” *Molecular & General Genetics*, vol. 213, no. 1, pp. 155–162, Jul. 1988, doi: 10.1007/BF00333413.

[154] M. Krehenbrink and J. A. Downie, “Identification of protein secretion systems and novel secreted proteins in *Rhizobium leguminosarum* bv. viciae,” *BMC Genomics*, vol. 9, no. 1, pp. 1–19, Jan. 2008, doi: 10.1186/1471-2164-9-55/TABLES/4.

[155] S. Mitra *et al.*, “A rhamnose-deficient lipopolysaccharide mutant of *Rhizobium* sp. IRBG74 is defective in root colonization and beneficial interactions with its flooding-tolerant hosts *Sesbania cannabina* and wetland rice,” *J Exp Bot*, vol. 67, no. 19, pp. 5869–5884, 2016, doi: 10.1093/JXB/ERW354.

[156] D. M. Russo *et al.*, “Lipopolysaccharide O-chain core region required for cellular cohesion and compaction of *in vitro* and root biofilms developed by *Rhizobium leguminosarum*,” *Appl Environ Microbiol*, vol. 81, no. 3, pp. 1013–1023, 2015, doi: 10.1128/aem.03175-14.

[157] F. Sorroche *et al.*, “Cell autoaggregation, biofilm formation, and plant attachment in a *Sinorhizobium meliloti* lpsB mutant,” *Molecular Plant-Microbe Interactions*, vol. 31, no. 10, pp. 1075–1082, 2018, doi: 10.1094/mpmi-01-18-0004-r.

[158] N. A. Fujishige, M. R. Lum, P. L. De Hoff, J. P. Whitelegge, K. F. Faull, and A. M. Hirsch, “*Rhizobium* common nod genes are required for biofilm formation,” *Mol Microbiol*, vol. 67, no. 3, pp. 504–515, Feb. 2008, doi: 10.1111/j.1365-2958.2007.06064.x.

[159] F. Pérez-Montaño *et al.*, "The symbiotic biofilm of *Sinorhizobium fredii*SMH12, necessary for successful colonization and symbiosis of *Glycine max* cv Osumi, is regulated by quorum sensing systems and inducing flavonoids via NodD1," *PLoS One*, vol. 9, no. 8, p. e105901, Aug. 2014, doi: 10.1371/journal.pone.0105901.

[160] H. Wang, Z. Zhong, T. Cai, S. Li, and J. Zhu, "Heterologous overexpression of quorum-sensing regulators to study cell-density-dependent phenotypes in a symbiotic plant bacterium *Mesorhizobium huakuii*," *Arch Microbiol*, vol. 182, no. 6, pp. 520–525, Dec. 2004, doi: 10.1007/S00203-004-0735-8.

[161] A. Edwards, M. Frederix, F. Wisniewski-Dyé, J. Jones, A. Zorreguieta, and J. Allan Downie, "The cin and rai quorum-sensing regulatory systems in *Rhizobium leguminosarum* are coordinated by ExpR and CinS, a small regulatory protein coexpressed with CinI," *J Bacteriol*, vol. 191, no. 9, pp. 3059–3067, 2009, doi: 10.1128/JB.01650-08.

[162] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics," *Nat Rev Genet*, vol. 10, no. 1, p. 57, 2009, doi: 10.1038/nrg2484.

[163] N. J. Croucher and N. R. Thomson, "Studying bacterial transcriptomes using RNA-seq," *Curr Opin Microbiol*, vol. 13, no. 5, pp. 619–624, 2010, doi: 10.1016/j.mib.2010.09.009.

[164] R. Stark, M. Grzelak, and J. Hadfield, "RNA sequencing: the teenage years," *Nat Rev Genet*, vol. 20, no. 11, pp. 631–656, 2019, doi: 10.1038/s41576-019-0150-2.

[165] K. R. Kukurba and S. B. Montgomery, "RNA sequencing and analysis," *Cold Spring Harb Protoc*, vol. 11, pp. 951–969, 2015, doi: 10.1101/pdb.top084970.

[166] R. Lowe, N. Shirley, M. Bleackley, S. Dolan, and T. Shafee, "Transcriptomics technologies," *PLoS Comput Biol*, vol. 13, no. 5, May 2017, doi: 10.1371/JOURNAL.PCBI.1005457.

[167] Y. Mo, R. Wan, and Q. Zhang, "Application of Reverse Transcription-PCR and Real-Time PCR in Nanotoxicity Research," *Methods Mol Biol*, vol. 926, p. 99, 2012, doi: 10.1007/978-1-62703-002-1_7.

[168] M. F. Z. Wang *et al.*, “Uncovering transcriptional dark matter via gene annotation independent single-cell RNA sequencing analysis,” *Nat Commun*, vol. 12, no. 1, pp. 1–10, 2021, doi: 10.1038/s41467-021-22496-3.

[169] R. M. Wheatley *et al.*, “Lifestyle adaptations of *Rhizobium* from rhizosphere to symbiosis,” *Proc Natl Acad Sci U S A*, vol. 117, no. 38, pp. 23823–23834, Sep. 2020, doi: 10.1073/pnas.2009094117.

[170] T. Van Opijnen and A. Camilli, “Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms,” *Nat Rev Microbiol*, vol. 11, no. 7, pp. 435–442, 2013, doi: 10.1038/nrmicro3033.

[171] T. Van Opijnen and H. L. Levin, “Transposon insertion sequencing, a global measure of gene function,” *Annu Rev Genet*, vol. 54, pp. 337–365, Nov. 2020, doi: 10.1146/annurev-genet-112618-043838.

[172] B. Gourion *et al.*, “High-throughput mass spectrometric analysis of the whole proteome and secretome from *Sinorhizobium fredii* strains CCBAU25509 and CCBAU45436,” *Front Microbiol*, vol. 10, p. 2569, 2019, doi: 10.3389/fmicb.2019.02569.

[173] B. J. Haas, M. Chin, C. Nusbaum, B. W. Birren, and J. Livny, “How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes?,” *BMC Genomics*, vol. 13, no. 1, pp. 1–11, 2012, doi: 10.1186/1471-2164-13-734.

[174] S. Zhao, Y. Zhang, R. Gamini, B. Zhang, and D. Von Schack, “Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA+ selection versus rRNA depletion,” *Sci Rep*, vol. 8, no. 1, pp. 1–12, 2018, doi: 10.1038/s41598-018-23226-4.

[175] M. Teufel and P. Sobetzko, “Reducing costs for DNA and RNA sequencing by sample pooling using a metagenomic approach,” *BMC Genomics*, vol. 23, no. 1, pp. 613–623, 2022, doi: 10.1186/s12864-022-08831-y.

[176] A. W. B. Johnston and J. E. Beringer, “Identification of the *Rhizobium* strains in pea nodules using genetic markers,” *J Gen Microbiol*, vol. 87, pp. 343–350, 1975, doi: 10.1099/00221287-87-2-343.

[177] A. McNally, “Rhizosphere colonization : assigning genes induced in the rhizosphere to regulatory networks.,” Masters, John Innes Centre (University of East Anglia), Norwich, 2012.

[178] A. J. Tett, S. J. Rudder, A. Bourdès, R. Karunakaran, and P. S. Poole, “Regulatable vectors for environmental gene expression in Alphaproteobacteria,” *Appl Environ Microbiol*, vol. 78, no. 19, pp. 7137–7140, 2012, doi: 10.1128/aem.01188-12.

[179] G. Ditta, S. Stanfield, D. Corbin, D. R. Helinski, and C. R. Donald Helinskl, “Broad host range DNA cloning system for Gram-negative bacteria: Construction of a gene bank of *Rhizobium meliloti*,” *Proc Natl Acad Sci U S A*, vol. 77, no. 12, pp. 7347–735, 1980, doi: 10.1073/pnas.77.12.7347.

[180] H. R. Bonomi *et al.*, “Light regulates attachment, exopolysaccharide production, and nodulation in *Rhizobium leguminosarum* through a LOV-histidine kinase photoreceptor,” *Proc Natl Acad Sci U S A*, vol. 109, no. 30, pp. 12135–12140, 2012, doi: 10.1073/pnas.1121292109.

[181] J. Miller, *Experiments in Molecular Genetics*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory, 1972.

[182] J. E. Beringer, “R factor transfer in *Rhizobium leguminosarum*,” *Microbiology (Reading)*, vol. 84, no. 1, pp. 188–198, 1974, doi: 10.1099/00221287-84-1-188.

[183] P. S. Poole, A. Blyth, C. J. Reid, and K. Walters, “Myo-inositol catabolism and catabolite regulation in *Rhizobium leguminosarum* bv. viciae,” *Microbiology (Reading)*, vol. 140, no. 10, pp. 2787–2795, 1994, doi: 10.1099/00221287-140-10-2787.

[184] M. Kearse *et al.*, “Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data,” *Bioinformatics Applications Note*, vol. 28, no. 12, pp. 1647–1649, 2012, doi: 10.1093/bioinformatics/bts199.

[185] P. Chomczynski and M. Rymaszewski, “Alkaline polyethylene glycol-based method for direct PCR from bacteria, eukaryotic tissue samples, and whole blood,” *Biotechniques*, vol. 40, no. 4, pp. 454–458, 2006, doi: 10.2144/000112149.

[186] T. Magoc, D. Wood, and S. L. Salzberg’, “EDGE-pro: Estimated Degree of Gene Expression in Prokaryotic Genomes,” *Evolutionary Bioinformatics*, vol. 9, p. EBO.S11250, 2013, doi: 10.4137/ebo.s11250.

[187] A. De Jong, O. P. Kuipers, and J. Kok, “FUNAGE-Pro: comprehensive web server for gene set enrichment analysis of prokaryotes,” *Nucleic Acids Res*, vol. 50, 2022, doi: 10.1093/nar/gkac441.

[188] G. Fahraeus, “The infection of clover root hairs by nodule bacteria studied by a simple glass slide technique.,” *J Gen Microbiol*, vol. 16, no. 2, pp. 374–381, 1957, doi: 10.1099/00221287-16-2-374.

[189] W. Gish and D. J. States, “Identification of protein coding regions by database similarity search,” *Nat Genet*, vol. 3, no. 3, pp. 266–272, 1993, doi: 10.1038/ng0393-266.

[190] D. Szklarczyk *et al.*, “STRING v10: protein-protein interaction networks, integrated over the tree of life,” *Nucleic Acids Res*, vol. 43, pp. 447–452, 2014, doi: 10.1093/nar/gku1003.

[191] J. Pardo-Diaz, P. S. Poole, M. Beguerisse-Díaz, C. M. Deane, and G. Reinert, “Generating weighted and thresholded gene coexpression networks using signed distance correlation,” *Network Science*, vol. 10, no. 2, pp. 131–145, 2022, doi: 10.1017/nws.2022.13.

[192] J. F. Taly *et al.*, “Using the T-Coffee package to build multiple sequence alignments of protein, RNA, DNA sequences and 3D structures,” *Nature Protocols 2011 6:11*, vol. 6, no. 11, pp. 1669–1682, Oct. 2011, doi: 10.1038/nprot.2011.393.

[193] C. Notredame, D. G. Higgins, and J. Heringa, “T-Coffee: A novel method for fast and accurate multiple sequence alignment,” *J Mol Biol*, vol. 302, pp. 205–217, 2000, doi: 10.1006/jmbi.2000.4042.

[194] I. Erb, J. R. González-Vallinas, G. Bussotti, E. Blanco, E. Eyras, and C. Notredame, “Use of ChIP-Seq data for the design of a multiple promoter-alignment method,” *Nucleic Acids Res*, vol. 40, no. 7, Apr. 2012, doi: 10.1093/NAR/GKR1292.

[195] M. Mirdita, S. Ovchinnikov, M. Steinegger, K. Schütze, Y. Moriwaki, and L. Heo, “ColabFold: making protein folding accessible to all,” *Nat Methods*, vol. 19, pp. 679–682, 2022, doi: 10.1038/s41592-022-01488-1.

[196] K. D. Tsirigos, C. Peters, N. Shu, L. Käll, and A. Elofsson, “The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides,” *Nucleic Acids Res*, vol. 43, no. W1, pp. W401–W407, 2015, doi: 10.1093/NAR/GKV485.

[197] J. Schindelin *et al.*, “Fiji: an open-source platform for biological-image analysis,” *Nat Methods*, vol. 9, no. 7, pp. 676–682, 2012, doi: 10.1038/nmeth.2019.

[198] A. Van Dommelen and J. Vanderleyden, “Associative nitrogen fixation,” in *Biology of the Nitrogen Cycle*, Elsevier, 2007, pp. 179–192. doi: 10.1016/B978-044452857-5.50013-8.

[199] H.-B. Li *et al.*, “Genetic diversity of nitrogen-fixing and plant growth promoting *Pseudomonas* species isolated from sugarcane rhizosphere,” *Front Microbiol*, vol. 8, no. 1268, pp. 1–20, 2017, doi: 10.3389/fmicb.2017.01268.

[200] J. N. Galloway *et al.*, “Transformation of the nitrogen cycle: Recent trends, questions, and potential solutions,” *Science (1979)*, vol. 320, no. 5878, pp. 889–892, 2008, doi: 10.1126/science.1136674.

[201] M. L. Gutiérrez-Zamora and E. Martínez-Romero, “Natural endophytic association between *Rhizobium etli* and maize (*Zea mays* L.),” *J Biotechnol*, vol. 91, no. 2, pp. 117–126, 2001, doi: 10.1016/S0168-1656(01)00332-7.

[202] M. Rosenblueth and E. Martínez-Romero, “*Rhizobium etli* maize populations and their competitivenes for roof colonization,” *Arch Microbiol*, vol. 181, no. 5, pp. 337–344, 2004, doi: 10.1007/s00203-004-0661-9.

[203] Y. G. Yanni *et al.*, "Natural endophytic association between *Rhizobium leguminosarum* bv. trifolii and rice roots and assessment of its potential to promote rice growth," *Plant Soil*, vol. 194, no. 1, pp. 99–114, 1997, doi: 10.1023/A:1004269902246.

[204] M. A. Matilla, M. Espinosa-Urgel, J. J. Rodríguez-Herva, J. L. Ramos, and M. I. Ramos-González, "Genomic analysis reveals the major driving forces of bacterial life in the rhizosphere," *Genome Biol*, vol. 8, no. 9, pp. 1–13, 2007, doi: 10.1186/gb-2007-8-9-r179.

[205] V. K. Ramachandran, A. K. East, R. Karunakaran, J. A. Downie, and P. S. Poole, "Adaptation of *Rhizobium leguminosarum* to pea, alfalfa and sugar beet rhizospheres investigated by comparative transcriptomics," *Genome Biol*, vol. 12, no. 10, p. R106, 2011, doi: 10.1186/gb-2011-12-10-r106.

[206] P. Vesga, P. Flury, J. Vacheron, C. Keel, D. Croll, and M. Maurhofer, "Transcriptome plasticity underlying plant root colonization and insect invasion by *Pseudomonas protegens*," *ISME J*, vol. 14, no. 11, pp. 2766–2782, 2020, doi: 10.1038/s41396-020-0729-9.

[207] D. M. Suchan *et al.*, "Transcriptomics reveal core activities of the plant growthpromoting bacterium *Delftia acidovorans* RAY209 during interaction with canola and soybean roots," *Microb Genom*, vol. 6, no. 11, pp. 1–13, Nov. 2020, doi: 10.1099/mgen.0.000462.

[208] T. Li, R. Mann, J. Kaur, G. Spangenberg, and T. Sawbridge, "Transcriptome analyses of barley roots inoculated with novel *Paenibacillus* sp. and *Erwinia gerundensis* strains reveal beneficial early-stage plant–bacteria interactions," *Plants*, vol. 10, no. 9, p. 1802, 2021, doi: 10.3390/plants10091802.

[209] V. C. S. Pankievicz *et al.*, "RNA-seq transcriptional profiling of *Herbaspirillum seropedicae* colonizing wheat (*Triticum aestivum*) roots," *Plant Mol Biol*, vol. 90, no. 6, pp. 589–603, 2016, doi: 10.1007/s11103-016-0430-6.

[210] A. Boyum, "Isolation of human blood monocytes with Nycodenz, a new non-ionic iodinated gradient medium," *Scand J Immunol*, vol. 17, no. 5, pp. 429–436, 1983, doi: 10.1111/j.1365-3083.1983.tb00809.x.

[211] E. H. Carter, A. Suhrbier, R. E. Sinden, and P. J. A. Beckers, “The in vitro cultivation of *P. falciparum* ookinetes, and their enrichment on Nycodenz density gradients,” *Parasitology*, vol. 95, no. 1, pp. 25–30, 1987, doi: 10.1017/S0031182000057516.

[212] C. J. Janse and A. P. Waters, “*Plasmodium berghei*: the application of cultivation and purification techniques to molecular studies of malaria parasites,” *Parasitol Today*, vol. 11, no. 4, pp. 138–143, 1995, doi: 10.1016/0169-4758(95)80133-2.

[213] E. Chapelle *et al.*, “A straightforward and reliable method for bacterial in planta transcriptomics: application to the *Dickeya dadantii*/*Arabidopsis thaliana* pathosystem,” *The Plant Journal*, vol. 82, no. 2, pp. 352–362, 2015, doi: 10.1111/TPJ.12812.

[214] The Gene Ontology Consortium *et al.*, “The Gene Ontology knowledgebase in 2023,” *Genetics*, vol. 224, no. 1, 2023, doi: 10.1093/genetics/iyad031.

[215] M. P. Mezzina and M. J. Pettinari, “Phasins: multifaceted polyhydroxyalkanoate granule-associated proteins,” *Appl Environ Microbiol*, vol. 82, no. 17, p. 5060, 2016, doi: 10.1128/aem.01161-16.

[216] M. Gao *et al.*, “*sinI*- and *expR*-dependent quorum sensing in *Sinorhizobium meliloti*,” *J Bacteriol*, vol. 187, no. 23, p. 7931, 2005, doi: 10.1128/JB.187.23.7931-7944.2005.

[217] C. Wang, X. Sheng, R. C. Equi, M. A. Trainer, T. C. Charles, and B. W. S. Sobral, “Influence of the poly-3-hydroxybutyrate (PHB) granule-associated proteins (PhaP1 and PhaP2) on PHB accumulation and symbiotic nitrogen fixation in *Sinorhizobium meliloti* Rm1021,” *J Bacteriol*, vol. 189, no. 24, p. 9050, 2007, doi: 10.1128/JB.01190-07.

[218] M. D’Alessio, R. Nordeste, A. C. Doxey, and T. C. Charles, “Transcriptome analysis of polyhydroxybutyrate cycle mutants reveals discrete loci connecting nitrogen utilization and carbon storage in *Sinorhizobium meliloti*,” *mSystems*, vol. 2, no. 5, 2017, doi: 10.1128/msystems.00035-17.

[219] R. Karunakaran *et al.*, "Transcriptomic analysis of *Rhizobium leguminosarum* biovar viciae in symbiosis with host plants *Pisum sativum* and *Vicia cracca*," *J Bacteriol*, vol. 191, no. 12, pp. 4002–4014, 2009, doi: 10.1128/jb.00165-09.

[220] N. Delmotte *et al.*, "Community proteogenomics reveals insights into the physiology of phyllosphere bacteria," *Proc Natl Acad Sci U S A*, vol. 106, no. 38, pp. 16428–16433, 2009, doi: 10.1073/pnas.0905240106.

[221] C. Knief *et al.*, "Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice," *ISME J*, vol. 6, no. 7, p. 1378, 2012, doi: 10.1038/ISMEJ.2011.192.

[222] C. Wekesa, J. O. Muoma, M. Reichelt, G. O. Asudi, A. C. U. Furch, and R. Oelmüller, "The cell membrane of a novel *Rhizobium phaseoli* strain is the crucial target for aluminium toxicity and tolerance," *Cells*, vol. 11, no. 5, pp. 873–895, 2022, doi: 10.3390/cells11050873.

[223] D. Das *et al.*, "The structure of the first representative of Pfam family PF09836 reveals a two-domain organization and suggests involvement in transcriptional regulation," *Acta Crystallogr Sect F Struct Biol Cryst Commun*, vol. 66, no. Pt 10, p. 1174, 2010, doi: 10.1107/S1744309109022672.

[224] K. Jwanoswki, C. Wells, T. Bruce, J. Rutt, T. Banks, and T. L. McNealy, "The *Legionella pneumophila* GIG operon responds to gold and copper in planktonic and biofilm cultures," *PLoS One*, vol. 12, no. 5, p. e0174245, 2017, doi: 10.1371/journal.pone.0174245.

[225] C. Lang, M. J. Barnett, R. F. Fisher, L. S. Smith, M. E. Diodati, and S. R. Long, "Most *Sinorhizobium meliloti* extracytoplasmic function sigma factors control accessory functions," *mSphere*, vol. 3, no. 5, 2018, doi: 10.1128/mspheredirect.00454-18.

[226] V. Waschulin *et al.*, "Biosynthetic potential of uncultured Antarctic soil bacteria revealed through long-read metagenomic sequencing," *The ISME Journal 2021 16:1*, vol. 16, no. 1, pp. 101–111, 2021, doi: 10.1038/s41396-021-01052-3.

[227] C. E. Alvarez-Martinez, R. L. Baldini, and S. L. Gomes, "A *Caulobacter crescentus* extracytoplasmic function sigma factor mediating the response to oxidative stress in

stationary phase," *J Bacteriol*, vol. 188, no. 5, pp. 1835–1846, 2006, doi: 10.1128/JB.188.5.1835-1846.2006.

[228] L. Sauviac, H. Philippe, K. Phok, and C. Bruand, "An extracytoplasmic function sigma factor acts as a general stress response regulator in *Sinorhizobium meliloti*," *J Bacteriol*, vol. 189, no. 11, p. 4204, 2007, doi: 10.1128/JB.00175-07.

[229] J. M. Martínez-Salazar, E. Salazar, S. Encarnación, M. A. Ramírez-Romero, and J. Rivera, "Role of the extracytoplasmic function sigma factor RpoE4 in oxidative and osmotic stress responses in *Rhizobium etli*," *J Bacteriol*, vol. 191, no. 13, p. 4122, 2009, doi: 10.1128/JB.01626-08.

[230] I. B. Zhulin, B. L. Taylor, and R. Dixon, "PAS domain S-boxes in archaea, bacteria and sensors for oxygen and redox," *Trends Biochem Sci*, vol. 22, no. 9, pp. 331–333, 1997, doi: 10.1016/S0968-0004(97)01110-9.

[231] B. L. Taylor and I. B. Zhulin, "PAS domains: internal sensors of oxygen, redox potential, and light," *Microbiology and Molecular Biology Reviews*, vol. 63, no. 2, pp. 479–506, 1999, doi: 10.1128/mmbr.63.2.479-506.1999.

[232] R. A. Ayers and K. Moffat, "Changes in quaternary structure in the signaling mechanisms of PAS domains," *Biochemistry*, vol. 47, no. 46, p. 12078, 2008, doi: 10.1021/BI801254C.

[233] J. P. Hawkins, P. A. Ordonez, and I. J. Oresnik, "Characterization of mutations that affect the nonoxidative pentose phosphate pathway in *Sinorhizobium meliloti*," *J Bacteriol*, vol. 200, no. 2, 2018, doi: 10.1128/JB.00436-17.

[234] M. Marzec, I. Szarejko, and M. Melzer, "Arabinogalactan proteins are involved in root hair development in barley," *J Exp Bot*, vol. 66, no. 5, pp. 1245–1257, 2015, doi: 10.1093/jxb/eru475.

[235] J. Fry, M. Wood, and P. S. Poole, "Investigation of myo-inositol catabolism in *Rhizobium leguminosarum* bv. viciae and its effect on nodulation competitiveness," *Mol Plant Microbe Interact*, vol. 14, no. 8, pp. 1016–1025, 2001, doi: 10.1094/MPMI.2001.14.8.1016.

[236] T. H. Mauchline *et al.*, “Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome,” *Proc Natl Acad Sci U S A*, vol. 103, no. 47, pp. 17933–17938, 2006, doi: 10.1073/pnas.0606673103.

[237] D. Hromadová, A. Soukup, and E. Tylová, “Arabinogalactan proteins in plant roots – an update on possible functions,” *Front Plant Sci*, vol. 12, p. 674010, 2021, doi: 10.3389/fpls.2021.674010.

[238] F. Pini *et al.*, “Bacterial biosensors for *in vivo* spatiotemporal mapping of root secretion,” *Plant Physiol*, vol. 174, no. 3, pp. 1289–1306, 2017, doi: 10.1104/pp.16.01302.

[239] M. Naveed *et al.*, “Plant exudates may stabilize or weaken soil depending on species, origin and time,” *Eur J Soil Sci*, vol. 68, no. 6, pp. 806–816, 2017, doi: 10.1111/EJSS.12487.

[240] P. Garcia-Fraile, J. C. Seaman, R. Karunakaran, A. Edwards, P. S. Poole, and J. A. Downie, “Arabinose and protocatechuate catabolism genes are important for growth of *Rhizobium leguminosarum* biovar viciae in the pea rhizosphere,” *Plant Soil*, vol. 390, no. 1, p. 251, 2015, doi: 10.1007/S11104-015-2389-5.

[241] H. Taboada *et al.*, “Proteins in the periplasmic space and outer membrane vesicles of *Rhizobium etli* CE3 grown in minimal medium are largely distinct and change with growth phase,” *Microbiology (United Kingdom)*, vol. 165, no. 6, pp. 638–650, 2019, doi: 10.1099/mic.0.000720.

[242] S. Magnet, L. Dubost, A. Marie, M. Arthur, and L. Gutmann, “Identification of the L,D-transpeptidases for peptidoglycan cross-linking in *Escherichia coli*,” *J Bacteriol*, vol. 190, no. 13, pp. 4782–4785, 2008, doi: 10.1128/JB.00025-08.

[243] J. F. Coyle, F. A. Pagliai, D. Zhang, G. L. Lorca, and C. F. Gonzalez, “Purification and partial characterization of LdtP, a cell envelope modifying enzyme in *Liberibacter asiaticus*,” *BMC Microbiol*, vol. 18, no. 1, pp. 1–15, 2018, doi: 10.1186/S12866-018-1348-8.

[244] R. Nakao, M. Ramstedt, S. N. Wai, and B. E. Uhlin, “Enhanced biofilm formation by *Escherichia coli* LPS mutants defective in hep biosynthesis,” *PLoS One*, vol. 7, no. 12, p. 51241, 2012, doi: 10.1371/journal.pone.0051241.

[245] R. L. Chirhart-Gilleland, M. E. Kovach, P. H. Elzer, S. R. Jennings, and R. M. Roop, “Identification and characterization of a 14-kilodalton *Brucella abortus* protein reactive with antibodies from naturally and experimentally infected hosts and T lymphocytes from experimentally infected BALB/c mice,” *Infect Immun*, vol. 66, no. 8, pp. 4000–4003, 1998, doi: 10.1128/IAI.66.8.4000-4003.1998.

[246] T. H. Vemulapalli, R. Vemulapalli, G. G. Schurig, S. M. Boyle, and N. Sriranganathan, “Role in virulence of a *Brucella abortus* protein exhibiting lectin-like activity,” *Infect Immun*, vol. 74, no. 1, p. 183, 2006, doi: 10.1128/iai.74.1.183-191.2006.

[247] X. He *et al.*, “Human tRNA-derived small RNAs modulate host–oral microbial interactions,” *J Dent Res*, vol. 97, no. 11, pp. 1236–1243, 2018, doi: 10.1177/0022034518770605.

[248] B. Ren, X. Wang, J. Duan, and J. Ma, “Rhizobial tRNA-derived small RNAs are signal molecules regulating plant nodulation,” *Science (1979)*, vol. 365, no. 6456, pp. 919–922, 2019, doi: 10.1126/SCIENCE.AAV8907/SUPPL_FILE/AAV8907_REN_TABLE-S4.XLSX.

[249] Z. Sun *et al.*, “tRNA-derived fragments from wheat are potentially involved in susceptibility to *Fusarium* head blight,” *BMC Plant Biol*, vol. 22, no. 1, pp. 1–17, 2022, doi: 10.1186/S12870-021-03393-9.

[250] S. Kusch, M. Singh, H. Thieron, P. D. Spanu, and R. Panstruga, “Site-specific analysis reveals candidate cross-kingdom small RNAs, tRNA and rRNA fragments, and signs of fungal RNA phasing in the barley–powdery mildew interaction,” *Mol Plant Pathol*, vol. 24, no. 6, pp. 570–587, 2023, doi: 10.1111/MPP.13324.

[251] C. S. Alves and F. T. S. Nogueira, “Plant small RNA world growing bigger: tRNA-derived fragments, longstanding players in regulatory processes,” *Front Mol Biosci*, vol. 8, p. 638911, 2021, doi: 10.3389/fmolb.2021.638911.

[252] C. K. Yost, K. L. Del Bel, J. Quandt, and M. F. Hynes, "*Rhizobium leguminosarum* methyl-accepting chemotaxis protein genes are down-regulated in the pea nodule," *Arch Microbiol*, vol. 182, no. 6, pp. 505–513, 2004, doi: 10.1007/S00203-004-0736-7.

[253] S. T. N. Aroney, P. S. Poole, and C. Sánchez-Cañizares, "Rhizobial chemotaxis and motility systems at work in the soil," *Front Plant Sci*, vol. 12, p. 725338, 2021, doi: 10.3389/fpls.2021.725338.

[254] H. H. Hoang, N. Gurich, and J. E. González, "Regulation of motility by the ExpR/Sin quorum-sensing system in *Sinorhizobium meliloti*," *J Bacteriol*, vol. 190, no. 3, pp. 861–871, 2008, doi: 10.1128/JB.01310-07.

[255] H. P. Erickson, D. E. Anderson, and M. Osawa, "FtsZ in bacterial cytokinesis: cytoskeleton and force generator all in one," *Microbiology and Molecular Biology Reviews*, vol. 74, no. 4, pp. 504–528, 2010, doi: 10.1128/mmbr.00021-10.

[256] M. T. Laub, S. L. Chen, L. Shapiro, and H. H. McAdams, "Genes directly controlled by CtrA, a master regulator of the *Caulobacter* cell cycle," *Proc Natl Acad Sci U S A*, vol. 99, no. 7, pp. 4632–4637, 2002, doi: 10.1073/pnas.062065699.

[257] W. Margolin, J. C. Corbo, and S. R. Long, "Cloning and characterization of a *Rhizobium meliloti* homolog of the *Escherichia coli* cell division gene *ftsZ*," *J Bacteriol*, vol. 173, no. 18, p. 5822, 1991, doi: 10.1128/JB.173.18.5822-5830.1991.

[258] W. Margolin and S. R. Long, "*Rhizobium meliloti* contains a novel second homolog of the cell division gene *ftsZ*," *J Bacteriol*, vol. 176, no. 7, pp. 2033–2043, 1994, doi: 10.1128/JB.176.7.2033-2043.1994.

[259] N. Smith and M. A. Wilson, "Structural biology of the DJ-1 superfamily," *Adv Exp Med Biol*, vol. 1037, pp. 5–24, 2017, doi: 10.1007/978-981-10-6583-5_2.

[260] J. W. Chase and C. C. Richardson, "Exonuclease VII of *Escherichia coli*," *Journal of Biological Chemistry*, vol. 249, no. 14, pp. 4553–4561, 1974, doi: 10.1016/S0021-9258(19)42454-X.

[261] M. Viswanathan, V. Burdett, C. Baitinger, P. Modrich, and S. T. Lovett, “Redundant exonuclease involvement in *Escherichia coli* methyl-directed mismatch repair,” *Journal of Biological Chemistry*, vol. 276, no. 33, pp. 31053–31058, 2001, doi: 10.1074/jbc.M105481200.

[262] G. M. Li, “Mechanisms and functions of DNA mismatch repair,” *Cell Research 2008 18:1*, vol. 18, no. 1, pp. 85–98, 2007, doi: 10.1038/cr.2007.115.

[263] S. Y. N. Huang *et al.*, “Exonuclease VII repairs quinolone-induced damage by resolving DNA gyrase cleavage complexes,” *Sci Adv*, vol. 7, no. 10, 2021, doi: 10.1126/sciadv.abe0384.

[264] P. Briaud, T. Gautier, V. Rong, L. Mereghetti, P. Lanotte, and A. Hiron, “The *Streptococcus agalactiae* exonuclease ExoVII is required for resistance to exogenous DNA-damaging agents,” *J Bacteriol*, vol. 205, no. 6, 2023, doi: 10.1128/jb.00024-23.

[265] M. Garavaglia, E. Rossi, and P. Landini, “The pyrimidine nucleotide biosynthetic pathway modulates production of biofilm determinants in *Escherichia coli*,” *PLoS One*, vol. 7, no. 2, p. e31252, 2012, doi: 10.1371/journal.pone.0031252.

[266] E. M. Vanderlinde *et al.*, “Identification of a novel ABC transporter required for desiccation tolerance, and biofilm formation in *Rhizobium leguminosarum* bv. viciae 3841,” *FEMS Microbiol Ecol*, vol. 71, no. 3, pp. 327–340, 2009, doi: 10.1111/j.1574-6941.2009.00824.x.

[267] W. Weidel, H. Frank, and H. H. Martin, “The rigid layer of the cell wall of *Escherichia coli* strain B.,” *J Gen Microbiol*, vol. 22, no. 1, pp. 158–166, 1960, doi: 10.1099/00221287-22-1-158.

[268] M. S. Pavelka and W. R. Jacobs, “Biosynthesis of diaminopimelate, the precursor of lysine and a component of peptidoglycan, is an essential function of *Mycobacterium smegmatis*,” *J Bacteriol*, vol. 178, no. 22, pp. 6496–6507, 1996, doi: 10.1128/JB.178.22.6496-6507.1996.

[269] E. M. Lodwig *et al.*, “Role of polyhydroxybutyrate and glycogen as carbon storage compounds in pea and bean bacteroids,” *Mol Plant Microbe Interact*, vol. 18, no. 1, pp. 67–74, 2005, doi: 10.1094/MPMI-18-0067.

[270] R. Karunakaran *et al.*, "BacA is essential for bacteroid development in nodules of Galegoid, but not Phaseoloid, legumes," *J Bacteriol*, vol. 192, no. 11, p. 2920, 2010, doi: 10.1128/JB.00020-10.

[271] J. D. Todd, G. Sawers, D. A. Rodionov, and A. W. B. Johnston, "The *Rhizobium leguminosarum* regulator IrrA affects the transcription of a wide range of genes in response to Fe availability," *Molecular Genetics and Genomics*, vol. 275, no. 6, pp. 564–577, 2006, doi: 10.1007/S00438-006-0115-Y/FIGURES/5.

[272] F. Dosselaere and J. Vanderleyden, "A metabolic node in action: chorismate-utilizing enzymes in microorganisms," *Crit Rev Microbiol*, vol. 27, no. 2, pp. 75–131, 2001, doi: 10.1080/20014091096710.

[273] M. J. Dilworth, K. C. Carson, R. G. F. Giles, L. T. Byrne, and A. R. Glenn, "*Rhizobium leguminosarum* bv. viciae produces a novel cyclic trihydroxamate siderophore, vicibactin," *Microbiology (Reading)*, vol. 144, no. 3, pp. 781–791, 1998, doi: 10.1099/00221287-144-3-781.

[274] R. C. Fahey, W. C. Brown, W. B. Adams, and M. B. Worsham, "Occurrence of glutathione in bacteria," *J Bacteriol*, vol. 133, no. 3, pp. 1126–1129, 1978, doi: 10.1128/JB.133.3.1126-1129.1978.

[275] J. A. Chesney, J. W. Eaton, and J. R. Mahoney, "Bacterial glutathione: a sacrificial defense against chlorine compounds," *J Bacteriol*, vol. 178, no. 7, pp. 2131–2135, 1996, doi: 10.1128/JB.178.7.2131-2135.1996.

[276] M. Montoya, "Bacterial glutathione import," *Nat Struct Mol Biol*, vol. 20, no. 7, pp. 775–775, Jul. 2013, doi: 10.1038/nsmb.2632.

[277] J. W. K. Ku and Y. H. Gan, "New roles for glutathione: Modulators of bacterial virulence and pathogenesis," *Redox Biol*, vol. 44, p. 102012, 2021, doi: 10.1016/J.redox.2021.102012.

[278] J. Harrison *et al.*, “Glutathione plays a fundamental role in growth and symbiotic capacity of *Sinorhizobium meliloti*,” *J Bacteriol*, vol. 187, no. 1, pp. 168–174, 2005, doi: 10.1128/JB.187.1.168-174.2005.

[279] C. Muglia, G. Comai, E. Spegazzini, P. M. Riccillo, and O. M. Aguilar, “Glutathione produced by *Rhizobium tropici* is important to prevent early senescence in common bean nodules,” *FEMS Microbiol Lett*, vol. 286, no. 2, pp. 191–198, 2008, doi: 10.1111/J.1574-6968.2008.01285.X.

[280] G. Cheng, R. Karunakaran, A. K. East, O. Munoz-Azcarate, and P. S. Poole, “Glutathione affects the transport activity of *Rhizobium leguminosarum* 3841 and is essential for efficient nodulation,” *FEMS Microbiol Lett*, vol. 364, no. 8, p. 45, 2017, doi: 10.1093/femsle/fnx045.

[281] A. Hu *et al.*, “*Rhizobium leguminosarum* glutathione peroxidase is essential for oxidative stress resistance and efficient nodulation,” *Front Microbiol*, vol. 12, p. 627562, 2021, doi: 10.3389/fmicb.2021.627562.

[282] M. Garcia-Alonso, C. Novillo, P. Kostolaniova, M. M. Parrilla, E. Alcalde, and N. Podevin, “The EU’s GM crop conundrum: Did the EU policy strategy to convert EFSA GMO guidance into legislation deliver on its promises?,” *EMBO Rep*, vol. 23, no. 5, 2022, doi: 10.15252/embr.202154529.

[283] E. S. Lee *et al.*, “Engineering homoeologs provide a fine scale for quantitative traits in polyploid,” *Plant Biotechnol J*, 2023, doi: 10.1111/PBI.14141.

[284] J. L. Aguirre-Noyola, M. Rosenblueth, M. G. Santiago-Martínez, and E. Martínez-Romero, “Transcriptomic responses of *Rhizobium phaseoli* to root exudates reflect its capacity to colonize maize and common bean in an intercropping system,” *Front Microbiol*, vol. 12, Oct. 2021, doi: 10.3389/fmicb.2021.740818.

[285] E. J. Mongiardini *et al.*, “The rhizobial adhesion protein RapA1 is involved in adsorption of rhizobia to plant roots but not in nodulation,” *FEMS Microbiol Ecol*, vol. 65, no. 2, pp. 279–288, 2008, doi: 10.1111/J.1574-6941.2008.00467.X.

[286] A. Stanger, “Analysis of ECF sigma factors of *Rhizobium leguminosarum* biovar viciae 3841,” 2008.

[287] E. Guisbert, T. Yura, V. A. Rhodius, and C. A. Gross, “Convergence of molecular, modeling, and systems approaches for an understanding of the *Escherichia coli* heat shock response,” *Microbiology and Molecular Biology Reviews*, vol. 72, no. 3, pp. 545–554, 2008, doi: 10.1128/mmbr.00007-08.

[288] J. M. Martínez-Salazar *et al.*, “The *Rhizobium etli* RpoH1 and RpoH2 sigma factors are involved in different stress responses,” *Microbiology (Reading)*, vol. 155, pp. 386–397, 2009, doi: 10.1099/mic.0.021428-0.

[289] Y. Ono, H. Mitsui, T. Sato, and K. Minamisawa, “Two RpoH homologs responsible for the expression of heat shock protein genes in *Sinorhizobium meliloti*,” *Molecular Genetics and Genomics*, vol. 264, pp. 902–912, 2001, doi: 10.1007/s004380000380.

[290] V. Oke, B. G. Rushing, E. J. Fisher, M. Moghadam-Tabrizi, and S. R. Long, “Identification of the heat-shock sigma factor RpoH and a second RpoH-like protein in *Sinorhizobium meliloti*,” *Microbiology (Reading)*, vol. 147, pp. 2399–2408, 2001, doi: 10.1099/00221287-147-9-2399.

[291] A. N. Bittner and V. Oke, “Multiple groESL operons are not key targets of RpoH1 and RpoH2 in *Sinorhizobium meliloti*,” *J Bacteriol*, vol. 188, no. 10, pp. 3507–3515, 2006, doi: 10.1128/JB.188.10.3507-3515.2006.

[292] A. Francez-Charlot, A. Kaczmarczyk, H. M. Fischer, and J. A. Vorholt, “The general stress response in Alphaproteobacteria,” *Trends Microbiol*, vol. 23, no. 3, pp. 164–171, Mar. 2015, doi: 10.1016/J.TIM.2014.12.006.

[293] J. Wülser *et al.*, “Salt- And Osmo-Responsive Sensor Histidine Kinases Activate the Bradyrhizobium diazoefficiens General Stress Response to Initiate Functional Symbiosis,” *Molecular Plant-Microbe Interactions*, vol. 35, no. 7, pp. 604–615, Jul. 2022, doi: 10.1094/MPMI-02-22-0051-FI/ASSET/IMAGES/LARGE/MPMI-02-22-0051-FIF4.JPEG.

[294] A. Kaczmarczyk, R. Hochstrasser, J. A. Vorholt, and A. Francez-Charlot, "Two-tiered histidine kinase pathway involved in heat shock and salt sensing in the general stress response of Sphingomonas melonis Fr1," *J Bacteriol*, vol. 197, no. 8, pp. 1466–1477, 2015, doi: 10.1128/JB.00019-15.

[295] B. Gourion, M. Rossignol, and J. A. Vorholt, "A proteomic study of Methylobacterium extorquens reveals a response regulator essential for epiphytic growth," *Proc Natl Acad Sci U S A*, vol. 103, no. 35, pp. 13186–13191, Aug. 2006, doi: 10.1073/PNAS.0603530103/SUPPL_FILE/03530TABLE3.PDF.

[296] B. Gourion *et al.*, "The PhyR-σEcfG signalling cascade is involved in stress response and symbiotic efficiency in Bradyrhizobium japonicum," *Mol Microbiol*, vol. 73, no. 2, pp. 291–305, Jul. 2009, doi: 10.1111/J.1365-2958.2009.06769.X.

[297] R. Ledermann, I. Bartsch, B. Müller, J. Wülser, and H. M. Fischer, "A functional general stress response of Bradyrhizobium diazoefficiens is required for early stages of host plant infection," *Molecular Plant-Microbe Interactions*, vol. 31, no. 5, pp. 537–547, May 2018, doi: 10.1094/MPMI-11-17-0284-R/ASSET/IMAGES/LARGE/MPMI-11-17-0284-R_T1.JPEG.

[298] R. M. Wheatley, V. K. Ramachandran, B. A. Geddes, B. J. Perry, C. K. Yost, and P. S. Poole, "Role of O2 in the growth of *Rhizobium leguminosarum* bv. viciae 3841 on glucose and succinate," *J Bacteriol*, vol. 199, no. 1, 2017, doi: 10.1128/jb.00572-16.

[299] L. Gottschlich, P. Geiser, M. Bortfeld-Miller, C. M. Field, and J. A. Vorholt, "Complex general stress response regulation in Sphingomonas melonis Fr1 revealed by transcriptional analyses," *Sci Rep*, vol. 9, no. 1, Dec. 2019, doi: 10.1038/S41598-019-45788-7.

[300] C. Lori, A. Kaczmarczyk, I. de Jong, and U. Jenal, "A Single-Domain Response Regulator Functions as an Integrating Hub To Coordinate General Stress Response and Development in Alphaproteobacteria," *mBio*, vol. 9, no. 3, May 2018, doi: 10.1128/MBIO.00809-18.

[301] A. Alexandre, M. Laranjo, and S. Oliveira, “Global transcriptional response to heat shock of the legume symbiont *Mesorhizobium loti* MAFF303099 comprises extensive gene downregulation,” *DNA Res*, vol. 21, no. 2, p. 195, 2014, doi: 10.1093/dnares/dst050.

[302] B. G. Butcher *et al.*, “The ECF sigma factor, PSPTO_1043, in *Pseudomonas syringae* pv. tomato DC3000 is induced by oxidative stress and regulates genes involved in oxidative stress response,” *PLoS One*, vol. 12, no. 7, p. e0180340, 2017, doi: 10.1371/journal.pone.0180340.

[303] B. Bastiat, L. Sauviac, C. Picheraux, M. Rossignol, and C. Bruand, “*Sinorhizobium meliloti* sigma factors RpoE1 and RpoE4 are activated in stationary phase in response to sulfite,” *PLoS One*, vol. 7, no. 11, 2012, doi: 10.1371/journal.pone.0050768.

[304] S. Shuman and M. S. Glickman, “Bacterial DNA repair by non-homologous end joining,” *Nature Reviews Microbiology 2007 5:11*, vol. 5, no. 11, pp. 852–861, 2007, doi: 10.1038/nrmicro1768.

[305] R. S. Pitcher, A. J. Green, A. Brzostek, M. Korycka-Machala, J. Dziadek, and A. J. Doherty, “NHEJ protects mycobacteria in stationary phase against the harmful effects of desiccation,” *DNA Repair (Amst)*, vol. 6, pp. 1271–1276, 2007, doi: 10.1016/j.dnarep.2007.02.009.

[306] P. Dupuy, B. Gourion, L. Sauviac, and C. Bruand, “DNA double-strand break repair is involved in desiccation resistance of *Sinorhizobium meliloti*, but is not essential for its symbiotic interaction with *Medicago truncatula*,” *Microbiology (United Kingdom)*, vol. 163, no. 3, pp. 333–342, 2017, doi: 10.1099/mic.0.000400.

[307] A. W. Grant, G. Steel, H. Waugh, and E. M. Ellis, “A novel aldo-keto reductase from *Escherichia coli* can increase resistance to methylglyoxal toxicity,” *FEMS Microbiol Lett*, vol. 218, no. 1, pp. 93–99, 2003, doi: 10.1111/J.1574-6968.2003.TB11503.X.

[308] W. J. He *et al.*, “An aldo-keto reductase is responsible for *Fusarium* toxin-degrading activity in a soil *Sphingomonas* strain,” *Sci Rep*, vol. 7, no. 1, pp. 1–13, 2017, doi: 10.1038/s41598-017-08799-w.

[309] N. Abraham *et al.*, “Structure–function characterization of an aldo–keto reductase involved in detoxification of the mycotoxin, deoxynivalenol,” *Sci Rep*, vol. 12, no. 1, pp. 1–13, 2022, doi: 10.1038/s41598-022-19040-8.

[310] A. Bateman *et al.*, “UniProt: the universal protein knowledgebase in 2023,” *Nucleic Acids Res*, vol. 51, no. D1, pp. D523–D531, 2023, doi: 10.1093/nar/gkac1052.

[311] E. M. Ellis, “Microbial aldo-keto reductases,” *FEMS Microbiol Lett*, vol. 216, no. 2, pp. 123–131, 2002, doi: 10.1111/J.1574-6968.2002.TB11425.X.

[312] A. J. Tett, R. Karunakaran, and P. S. Poole, “Characterisation of SalRAB a salicylic acid inducible positively regulated efflux system of *Rhizobium leguminosarum* bv viciae 3841,” *PLoS One*, vol. 9, no. 8, 2014, doi: 10.1371/journal.pone.0103647.

[313] E. Nguema-Ona, M. Vicré-Gibouin, M.-A. Cannesan, and A. Driouich, “Arabinogalactan proteins in root-microbe interactions,” *Trends Plant Sci*, vol. 18, no. 8, pp. 440–449, 2013, doi: 10.1016/j.tplants.2013.03.

[314] M. A. Fox *et al.*, “Characterization of the quaternary amine transporters of *Rhizobium leguminosarum* bv. viciae 3841,” *FEMS Microbiol Lett*, vol. 287, pp. 212–220, 2008, doi: 10.1111/j.1574-6968.2008.01307.x.

[315] S. Weber, “Light-driven enzymatic catalysis of DNA repair: a review of recent biophysical studies on photolyase,” *Biochim Biophys Acta*, vol. 1707, pp. 1–23, 2005, doi: 10.1016/j.bbabio.2004.02.010.

[316] M. A. Williams *et al.*, “Unipolar peptidoglycan synthesis in the Rhizobiales requires an essential class A penicillin-binding protein,” *mBio*, vol. 12, no. 5, p. e0234621, Oct. 2021, doi: 10.1128/mbio.02346-21.

[317] C. R. H. Raetz and C. Whitfield, “Lipopolysaccharide endotoxins,” *Annu Rev Biochem*, vol. 71, pp. 635–700, 2002, doi: 10.1146/annurev.biochem.71.110601.135414.

[318] P. Reeves, “Evolution of *Salmonella* O antigen variation by interspecific gene transfer on a large scale,” *Trends in Genetics*, vol. 9, no. 1, pp. 17–22, 1993, doi: 10.1016/0168-9525(93)90067-R.

[319] E. M. Vanderlinde *et al.*, “*Rhizobium leguminosarum* biovar viciae 3841, deficient in 27-hydroxyoctacosanoate-modified lipopolysaccharide, is impaired in desiccation tolerance, biofilm formation and motility,” *Microbiology (Reading)*, vol. 155, no. Pt 9, p. 3055, 2009, doi: 10.1099/MIC.0.025031-0.

[320] M. Barr, A. K. East, M. Leonard, T. H. Mauchline, and P. S. Poole, “In vivo expression technology (IVET) selection of genes of *Rhizobium leguminosarum* biovar viciae A34 expressed in the rhizosphere,” *FEMS Microbiol Lett*, vol. 282, no. 2, pp. 219–227, 2008, doi: 10.1111/J.1574-6968.2008.01131.X.

[321] M. I. Ramos-González, M. J. Campos, and J. L. Ramos, “Analysis of *Pseudomonas putida* KT2440 gene expression in the maize rhizosphere: In vitro expression technology capture and identification of root-activated promoters,” *J Bacteriol*, vol. 187, no. 12, pp. 4033–4041, 2005, doi: 10.1128/jb.187.12.4033-4041.2005.

[322] C. Park, B. Shin, and W. Park, “Protective role of bacterial alkanesulfonate monooxygenase under oxidative stress,” *Appl Environ Microbiol*, vol. 86, no. 15, 2020, doi: 10.1128/aem.00692-20.

[323] M. Saini, A. Kashyap, S. Bindal, K. Saini, and R. Gupta, “Bacterial gamma-glutamyl transpeptidase, an emerging biocatalyst: insights into structure–function relationship and its biotechnological applications,” *Front Microbiol*, vol. 12, 2021, doi: 10.3389/fmicb.2021.641251.

[324] H. H. Zahran, “*Rhizobium*-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate,” *Microbiology and Molecular Biology Reviews*, vol. 63, no. 4, p. 968, 1999, doi: 10.1128/mmbr.63.4.968-989.1999.

[325] B. Martinac and A. Kloda, “Evolutionary origins of mechanosensitive ion channels,” *Prog Biophys Mol Biol*, vol. 82, no. 1–3, pp. 11–24, 2003, doi: 10.1016/S0079-6107(03)00002-6.

[326] J. J. Wilson and U. Kappler, “Sulfite oxidation in *Sinorhizobium meliloti*,” *Biochim Biophys Acta*, vol. 1787, pp. 1516–1525, 2009, doi: 10.1016/j.bbabio.2009.07.005.

[327] M. Perego, C. F. Higgins, S. R. Pearce, M. P. Gallagher, and J. A. Hoch, “The oligopeptide transport system of *Bacillus subtilis* plays a role in the initiation of sporulation,” *Mol Microbiol*, vol. 5, no. 1, pp. 173–185, 1991, doi: 10.1111/J.1365-2958.1991.TB01838.X.

[328] J. Nogales *et al.*, “Genetic characterization of oligopeptide uptake systems in *Sinorhizobium meliloti*,” *FEMS Microbiol Lett*, vol. 293, no. 2, pp. 177–187, 2009, doi: 10.1111/J.1574-6968.2009.01527.X.

[329] R. Karunakaran, A. K. East, and P. S. Poole, “Malonate catabolism does not drive N2 fixation in legume nodules,” *Appl Environ Microbiol*, vol. 79, no. 14, p. 4496, 2013, doi: 10.1128/aem.00919-13.

[330] D. K. Stumpf and R. H. Burris, “Organic acid contents of soybean: age and source of nitrogen,” *Plant Physiol*, vol. 68, no. 5, p. 989, 1981, doi: 10.1104/PP.68.5.989.

[331] R. Chevrot *et al.*, “GABA controls the level of quorum-sensing signal in *Agrobacterium tumefaciens*,” *Proc Natl Acad Sci U S A*, vol. 103, no. 19, pp. 7460–7464, 2006, doi: 10.1073/pnas.0600313103.

[332] A. Canarini, C. Kaiser, A. Merchant, A. Richter, and W. Wanek, “Root exudation of primary metabolites: mechanisms and their roles in plant responses to environmental stimuli,” *Front Plant Sci*, vol. 10, no. 157, pp. 1–19, 2019, doi: 10.3389/fpls.2019.00157.

[333] P. Wang *et al.*, “Natural variation in root exudation of GABA and DIMBOA impacts the maize root endosphere and rhizosphere microbiomes,” *J Exp Bot*, vol. 73, no. 14, pp. 5052–5066, 2022, doi: 10.1093/jxb/erac202.

[334] T. Tsuge *et al.*, "Molecular cloning of two (R)-specific enoyl-CoA hydratase genes from *Pseudomonas aeruginosa* and their use for polyhydroxyalkanoate synthesis," *FEMS Microbiol Lett*, vol. 184, no. 2, pp. 193–198, 2000, doi: 10.1111/J.1574-6968.2000.TB09013.X.

[335] G. Liu *et al.*, "Enoyl-CoA hydratase mediates polyhydroxyalkanoate mobilization in *Haloferax mediterranei*," *Scientific Reports 2016 6:1*, vol. 6, no. 1, pp. 1–11, 2016, doi: 10.1038/srep24015.

[336] R. Tombolini, S. Povolo, A. Buson, A. Squartini, and M. P. Nuti, "Poly-β-hydroxybutyrate (PHB) biosynthetic genes in *Rhizobium meliloti* 41," *Microbiology (Reading)*, vol. 141, no. 10, pp. 2553–2559, 1995, doi: 10.1099/13500872-141-10-2553.

[337] S. Moreau, D. A. Day, and A. Puppo, "Ferrous iron is transported across the peribacteroid membrane of soybean nodules," *Planta*, vol. 207, no. 1, pp. 83–87, 1998, doi: 10.1007/S004250050458.

[338] T. Nobori *et al.*, "Transcriptome landscape of a bacterial pathogen under plant immunity," *Proc Natl Acad Sci U S A*, vol. 115, no. 13, pp. E3055–E3064, 2018, doi: 10.1073/pnas.1800529115.

[339] L. Skøt and H. Egsgaard, "Identification of ononitol and O-methyl-scyllo-inositol in pea root nodules," *Planta*, vol. 161, pp. 32–36, 1984, doi: 10.1007/BF00951457.

[340] G. Smit, S. Swart, B. J. J. Lugtenberg, and J. W. Kijne, "Molecular mechanisms of attachment of *Rhizobium* bacteria to plant roots," *Mol Microbiol*, vol. 6, no. 20, pp. 2897–2903, 1992, doi: 10.1111/j.1365-2958.1992.tb01748.x.

[341] B. E. Scharf, M. F. Hynes, and G. M. Alexandre, "Chemotaxis signaling systems in model beneficial plant–bacteria associations," *Plant Mol Biol*, vol. 90, no. 6, pp. 549–559, 2016, doi: 10.1007/S11103-016-0432-4/FIGURES/2.

[342] V. B. Tran, S. M. J. Fleiszig, D. J. Evans, and C. J. Radke, "Dynamics of flagellum-and pilus-mediated association of *Pseudomonas aeruginosa* with contact lens surfaces," *Appl Environ Microbiol*, vol. 77, no. 11, pp. 3644–3652, 2011, doi: 10.1128/aem.02656-10.

[343] E. C. Sonnenschein, D. A. Syit, H. P. Grossart, and M. S. Ullrich, “Chemotaxis of *Marinobacter adhaerens* and its impact on attachment to the diatom *Thalassiosira weissflogii*,” *Appl Environ Microbiol*, vol. 78, no. 19, p. 6900, 2012, doi: 10.1128/AEM.01790-12.

[344] J. C. Conrad, “Physics of bacterial near-surface motility using flagella and type IV pili: implications for biofilm formation,” *Res Microbiol*, vol. 163, no. 9–10, pp. 619–629, Nov. 2012, doi: 10.1016/J.RESMIC.2012.10.016.

[345] D. N. Rodríguez-Navarro, M. S. Dardanelli, J. Jos´, and J. E. Ruíz-Saínz, “Attachment of bacteria to the roots of higher plants,” *FEMS Microbiol Lett*, vol. 272, pp. 127–136, 2007, doi: 10.1111/j.1574-6968.2007.00761.x.

[346] M. Paget, “Bacterial sigma factors and anti-sigma factors: structure, function and distribution,” *Biomolecules*, vol. 5, no. 3, pp. 1245–1265, 2015, doi: 10.3390/biom5031245.

[347] K. L. Brown and K. T. Hughes, “The role of anti-sigma factors in gene regulation,” *Mol Microbiol*, vol. 16, no. 3, pp. 397–404, 1995, doi: 10.1111/J.1365-2958.1995.TB02405.X.

[348] A. De Las Peñas, L. Connolly, and C. A. Gross, “The σE-mediated response to extracytoplasmic stress in *Escherichia coli* is transduced by RseA and RseB, two negative regulators of σE,” *Mol Microbiol*, vol. 24, no. 2, pp. 373–385, 1997, doi: 10.1046/J.1365-2958.1997.3611718.X.

[349] S. E. Ades, L. E. Connolly, B. M. Alba, and C. A. Gross, “The *Escherichia coli* sigma E-dependent extracytoplasmic stress response is controlled by the regulated proteolysis of an anti-sigma factor,” *Genes Dev*, vol. 13, pp. 2449–2461, 1999, doi: 10.1101/gad.13.18.2449.

[350] I. L. Grigorova *et al.*, “Fine-tuning of the *Escherichia coli* sigma E envelope stress response relies on multiple mechanisms to inhibit signal-independent proteolysis of the transmembrane anti-sigma factor, RseA,” *Genes Dev*, vol. 18, pp. 2686–2697, 2004, doi: 10.1101/gad.1238604.

[351] Y. Akiyama, K. Kanehara, and K. Ito, “RseP (YaeL), an *Escherichia coli* RIP protease, cleaves transmembrane sequences,” *EMBO Journal*, vol. 23, no. 22, pp. 4434–4442, 2004, doi: 10.1038/sj.emboj.7600449.

[352] K. H. Yeoman, S. Mitelheiser, G. Sawers, and A. W. B. Johnston, “The ECF σ factor RpoI of *R. leguminosarum* initiates transcription of the *vbsGSO* and *vbsADL* siderophore biosynthetic genes in vitro,” *FEMS Microbiol Lett*, vol. 223, no. 2, pp. 239–244, 2003, doi: 10.1016/S0378-1097(03)00386-0.

[353] M. N. Price, K. H. Huang, E. J. Alm, and A. P. Arkin, “A novel method for accurate operon predictions in all sequenced prokaryotes,” *Nucleic Acids Res*, vol. 33, no. 3, pp. 880–892, 2005, doi: 10.1093/nar/gki232.

[354] R. T. Brumfield, “Cell growth and division in living root meristems,” *Am J Bot*, vol. 29, no. 7, p. 533, 1942, doi: 10.2307/2437102.

[355] T. S. Gunasekera *et al.*, “Transcriptomic analyses elucidate adaptive differences of closely related strains of Pseudomonas aeruginosa in fuel,” *Appl Environ Microbiol*, vol. 83, no. 10, May 2017, doi: 10.1128/AEM.03249-16.

[356] L. Mark Wetzler *et al.*, “Analyzing the molecular mechanism of lipoprotein localization in *Brucella*,” *Frontiers in Microbiology | www.frontiersin.org*, vol. 6, p. 1189, 2015, doi: 10.3389/fmicb.2015.01189.

[357] C. S. Hoffman and A. Wright, “Fusions of secreted proteins to alkaline phosphatase: an approach for studying protein secretion,” *Proc Natl Acad Sci U S A*, vol. 82, no. 15, pp. 5107–5111, 1985, doi: 10.1073/PNAS.82.15.5107.

[358] J. Prell *et al.*, “The PTSNtr system globally regulates ATP-dependent transporters in *Rhizobium leguminosarum*,” *Mol Microbiol*, vol. 84, no. 1, pp. 117–129, 2012, doi: 10.1111/J.1365-2958.2012.08014.X.

[359] D. Allaway, B. Jeyaretnam, R. W. Carlson, and P. S. Poole, “Genetic and chemical characterization of a mutant that disrupts synthesis of the lipopolysaccharide core tetrasaccharide in Rhizobium leguminosarum.,” *J Bacteriol*, vol. 178, no. 21, p. 6403, 1996, doi: 10.1128/JB.178.21.6403-6406.1996.

[360] T. V Bhuvaneswari, S. G. Pueppke, and W. D. Bauer, “Role of lectins in plant-microorganism interactions I. Binding of soybean lectin to rhizobia,” *Plant Physiol*, vol. 60, pp. 486–491, 1977, doi: 10.1104/pp.60.4.486.

[361] T. V Bhuvaneswari, W. D. Bauer, and C. F. Kettering, “Role of lectins in plant-microorganism interactions III. Influence of rhizosphere/rhizoplane culture conditions on the soybean lectin-binding properties of rhizobia,” *Plant Physiol*, vol. 62, pp. 71–74, 1978, doi: 10.1104/pp.62.1.71.

[362] A. Ayouba *et al.*, “Interactions of plant lectins with the components of the bacterial cell wall peptidoglycan,” *Biochem Syst Ecol*, vol. 22, no. 2, pp. 153–159, 1994, doi: 10.1016/0305-1978(94)90005-1.

[363] N. Lannoo, E. J. M. Van Damme, R. Tenhaken, and E. J. M. Van, “Lectin domains at the frontiers of plant defense,” *Front Plant Sci*, vol. 5, no. 397, pp. 1–16, 2014, doi: 10.3389/fpls.2014.00397.

[364] R. Castilleux *et al.*, “Cell wall extensins in root-microbe interactions and root secretions,” *J Exp Bot*, vol. 69, no. 18, pp. 4235–4247, 2018, doi: 10.1093/jxb/ery238.

[365] J. M. Dunwell, J. George Gibbings, T. Mahmood, and S. M. Saqlan Naqvi, “Critical reviews in plant sciences germin and germin-like proteins: evolution, structure, and function germin and germin-like proteins: evolution, structure, and function,” *CRC Crit Rev Plant Sci*, vol. 27, no. 5, pp. 342–375, 2008, doi: 10.1080/07352680802333938.

[366] R. M. Davidson, P. A. Reeves, P. M. Manosalva, and J. E. Leach, “Germins: A diverse protein family important for crop improvement,” *Plant Science*, vol. 177, pp. 499–510, 2009, doi: 10.1016/j.plantsci.2009.08.012.

[367] S. Swart, T. J. J. Logman, G. Smit, B. J. J. Lugtenberg, and J. W. Kijne, “Purification and partial characterization of a glycoprotein from pea (*Pisum sativum*) with receptor activity for rhicadhesin, an attachment protein of Rhizobiaceae,” *Plant Mol Biol*, vol. 24, no. 1, pp. 171–183, 1994, doi: 10.1007/BF00040583.

[368] S. Gucciardo, J. P. Wisniewski, N. J. Brewin, and S. Bornemann, “A germin-like protein with superoxide dismutase activity in pea nodules with high protein sequence identity to a putative rhicadhesin receptor,” *J Exp Bot*, vol. 58, no. 5, pp. 1161–1171, 2007, doi: 10.1093/JXB/ERL282.

[369] M. Liu, H. Dong, M. Wang, and Q. Liu, “Evolutionary divergence of function and expression of laccase genes in plants,” *J Genet*, vol. 99, p. 23, 2020, doi: 10.1007/s12041-020-1184-0.

[370] G. Janusz *et al.*, “Laccase properties, physiological functions, and evolution,” *Int J Mol Sci*, vol. 21, p. 966, 2020, doi: 10.3390/ijms21030966.

[371] Y. Bai, S. Ali, S. Liu, J. Zhou, and Y. Tang, “Characterization of plant laccase genes and their functions,” *Gene*, vol. 852, p. 147060, 2023, doi: 10.1016/j.gene.2022.147060.

[372] Q. Wang *et al.*, “The soybean laccase gene family: evolution and possible roles in plant defense and stem strength selection,” *Genes (Basel)*, vol. 10, no. 9, p. 701, 2019, doi: 10.3390/GENES10090701.

[373] C. Gough, “*Rhizobium* symbiosis: Insight into Nod factor receptors,” *Current Biology*, vol. 13, pp. 973–975, 2003, doi: 10.1016/j.cub.2003.11.047.

[374] G. J. Seifert, “Fascinating fasciclins: A surprisingly widespread family of proteins that mediate interactions between the cell exterior and the cell surface,” *Int J Mol Sci*, vol. 19, no. 6, 2018, doi: 10.3390/IJMS19061628.

[375] V. Oke and S. R. Long, “Bacterial genes induced within the nodule during the *Rhizobium*-legume symbiosis,” *Mol Microbiol*, vol. 32, no. 4, pp. 837–849, 1999, doi: 10.1046/J.1365-2958.1999.01402.X.

[376] V. Oke and S. R. Long, “Bacteroid formation in the *Rhizobium*-legume symbiosis,” *Curr Opin Microbiol*, vol. 2, no. 6, pp. 641–646, 1999, doi: 10.1016/S1369-5274(99)00035-1.

[377] P. Paulsrud and P. Lindblad, “Fasciclin domain proteins are present in *Nostoc* symbionts of lichens,” *Appl Environ Microbiol*, vol. 68, no. 4, pp. 2036–2039, 2002, doi: 10.1128/aem.68.4.2036-2039.2002.

[378] C. Sánchez-Cañizares and J. Palacios, “Construction of a marker system for the evaluation of competitiveness for legume nodulation in *Rhizobium* strains,” *J Microbiol Methods*, vol. 92, no. 3, pp. 246–249, 2013, doi: 10.1016/J.MIMET.2012.12.022.

[379] K. W. Michiels, C. L. Croes, and J. Vanderleyden, “Two different modes of attachment of *Azospirillum brasilense* Sp7 to wheat roots,” *J Gen Microbiol*, vol. 137, pp. 2241–2246, 1991, doi: 10.1099/00221287-137-9-2241.

[380] H. Massalha, E. Korenblum, S. Malitsky, O. H. Shapiro, and A. Aharoni, “Live imaging of root-bacteria interactions in a microfluidics setup,” *Proc Natl Acad Sci U S A*, vol. 114, no. 17, pp. 4549–4554, 2017, doi: 10.1073/pnas.1618584114.

[381] N. F. Keinath *et al.*, “Live cell imaging with R-GECO1 sheds light on flg22- and chitin-induced transient [Ca2+]cyt patterns in *Arabidopsis*,” *Mol Plant*, vol. 8, no. 8, p. 1188, 2015, doi: 10.1016/J.MOLP.2015.05.006.

[382] T. L. Raivio and T. J. Silhavy, “Periplasmic stress and ECF sigma factors,” *Annu Rev Microbiol*, vol. 55, pp. 591–624, 2001, doi: 10.1146/annurev.micro.55.1.591.

[383] J. P. McHugh *et al.*, “Global iron-dependent gene regulation in *Escherichia coli*,” *Journal of Biological Chemistry*, vol. 278, no. 32, pp. 29478–29486, 2003, doi: 10.1074/jbc.m303381200.

[384] S. C. Andrews, A. K. Robinson, and F. Rodríguez-Quiñones, “Bacterial iron homeostasis,” *FEMS Microbiol Rev*, vol. 27, no. 2–3, pp. 215–237, 2003, doi: 10.1016/S0168-6445(03)00055-X.

[385] I. L. Lamont, P. A. Beare, U. Ochsner, A. I. Vasil, and M. L. Vasil, “Siderophore-mediated signaling regulates virulence factor production in Pseudomonas aeruginosa,” *Proc Natl Acad Sci U S A*, vol. 99, no. 10, p. 7072, May 2002, doi: 10.1073/PNAS.092016999.

[386] B. E. Brooks and S. K. Buchanan, “Signaling mechanisms for activation of extracytoplasmic function (ECF) sigma factors,” *Biochim Biophys Acta*, vol. 1778, no. 9, p. 1930, 2008, doi: 10.1016/J.BBAMEM.2007.06.005.

[387] B. Collinet, H. Yuzawa, T. Chen, C. Herrera, and D. Missiakas, “RseB binding to the periplasmic domain of RseA modulates the RseA:sigmaE interaction in the cytoplasm and the availability

of sigmaE·RNA polymerase," *Journal of Biological Chemistry*, vol. 275, no. 43, pp. 33898–33904, 2000, doi: 10.1074/JBC.M006214200.

[388] P. Wollmann and K. Zeth, "The structure of RseB: A sensor in periplasmic stress response of *E. coli*," *J Mol Biol*, vol. 372, no. 4, pp. 927–941, 2007, doi: 10.1016/J.JMB.2007.06.039.

[389] M. I. Hutchings, H.-J. Hong, E. Leibovitz, I. C. Sutcliffe, and M. J. Buttner, "The sigma(E) cell envelope stress response of *Streptomyces coelicolor* is influenced by a novel lipoprotein, CseA.," *J Bacteriol*, vol. 188, no. 20, pp. 7222–9, 2006, doi: 10.1128/JB.00818-06.

[390] J. D. Todd, M. Wexler, G. Sawers, K. H. Yeoman, P. S. Poole, and A. W. B. Johnston, "RirA, an iron-responsive regulator in the symbiotic bacterium *Rhizobium leguminosarum*," *Microbiology (Reading)*, vol. 148, no. 12, pp. 4059–4071, 2002, doi: 10.1099/00221287-148-12-4059.

[391] J. C. Crespo-Rivas *et al.*, "*Sinorhizobium fredii* HH103 RirA is required for oxidative stress resistance and efficient symbiosis with soybean," *Int J Mol Sci*, vol. 20, no. 3, pp. 787–803, 2019, doi: 10.3390/IJMS20030787.

[392] A. Tkacz, E. Bestion, Z. Bo, M. Hortala, and P. S. Poole, "Influence of plant fraction, soil, and plant species on microbiota: A multikingdom comparison," *mBio*, vol. 11, no. 1, 2020, doi: 10.1128/MBIO.02785-19.

[393] T. Lurthy, B. Pivato, P. Lemanceau, and S. Mazurier, "Importance of the rhizosphere microbiota in iron biofortification of plants," *Front Plant Sci*, vol. 12, p. 744445, 2021, doi: 10.3389/fpls.2021.744445.

[394] C. H. Yang and D. E. Crowley, "Rhizosphere microbial community structure in relation to root location and plant iron nutritional status," *Appl Environ Microbiol*, vol. 66, no. 1, pp. 345–351, 2000, doi: 10.1128/aem.66.1.345-351.2000.

[395] I. A. Stringlis *et al.*, "MYB72-dependent coumarin exudation shapes root microbiome assembly to promote plant health," *Proc Natl Acad Sci U S A*, vol. 115, no. 22, pp. E5213–E5222, May 2018, doi: 10.1073/pnas.1722335115.

[396] Y. Liu *et al.*, “Plant commensal type VII secretion system causes iron leakage from roots to promote colonization,” *Nature Microbiology 2023 8:8*, vol. 8, no. 8, pp. 1434–1449, 2023, doi: 10.1038/s41564-023-01402-1.

[397] P. Marschner, D. E. Crowley, and B. Sattelmacher, “Root colonization and iron nutritional status of a *Pseudomonas fluorescens* in different plant species,” *Plant Soil*, vol. 196, no. 2, pp. 311–316, 1997, doi: 10.1023/A:1004236410302.

Appendix: Supplementary information

Chapters 3 & 4: Tables 1 & 2, respectively.

Table 1: Reads in transcripts per million (TPM), log2(fold change) and adjusted p-value calculated using DeSeq2 (where applicable) for genes referred to in Chapter 3.

Table 2: Reads in transcripts per million (TPM), log2(fold change) and adjusted p-value calculated using DeSeq2 for genes referred to in Chapter 4.

Chapter 5: Videos 1-2.

Video 1: P*lppE*-driven luminescence in Rlv3841 inoculated on pea roots. Timelapse taken over 14.5 hours. Images captured using NightOWL (Berthold) and timelapse generated using IndiGO.

Video 2: P*neo*-driven luminescence in Rlv3841 inoculated onto pea roots. Timelapse taken over 14.5 hours. Images captured using NightOWL (Berthold) and timelapse generated using IndiGO.