Characterising root attachment in *Rhizobium*-legume symbioses

Jack Parsons

New College

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Trinity Term 2019

Abstract

Characterising root attachment in *Rhizobium*-legume symbioses

Jack Parsons, New College
Submitted for the degree of DPhil., Trinity Term 2019

In *Rhizobium*-legume symbioses the earliest stage of physical contact between bacteria and the plant root is primary root attachment. This is crucial for nitrogen-fixing symbiosis development and underpins many plant growth-promoting relationships. Rhizobia use a variety of factors for primary attachment including pH-dependent adhesins (such as glucomannan and the hypothesised rhicadhesin), surface proteins and extracellular polymeric substances. However, primary attachment remains an understudied area of symbiosis development.

In this work I use a range of techniques including luminescence-based attachment assays, *mariner* insertion sequencing (INSeq) and real-time imaging to investigate the factors governing primary attachment to plant roots with the model organism *Rhizobium leguminosarum* biovar *viciae* 3841. These techniques demonstrate that bacterial cell surface and extracellular factors are crucial and show extensive pH-condition and plant host specificity. Exopolysaccharide, lipopolysaccharide and peptidoglycan all show different profiles of modification in attachment to pea roots at different pHs and attachment to barley and soybean roots. The glycolytic enzyme TpiA is likely to be surface localized and is an attachment factor required under all conditions. Further, outer membrane protein and Flp/Tad pilus usage in attachment to soybean and barley roots shows that Rlv3841 can use primary attachment mechanisms demonstrated in other bacterial species. Another novel insight is that a filamentous hemagglutinin adhesin factor is also a previously unknown primary attachment factor. Proteomics, attachment assays, INSeq and confocal imaging were used to investigate rhicadhesin, demonstrating that there are multiple attachment factors matching the criteria for this protein. As glucomannan-independent root hair attachment is shown to be both polar and non-polar, these factors are likely distributed across the cell surface. Results from INSeq showed that control of cyclic-di-GMP levels is another important parameter in root attachment. It seems likely that the regulator RL4145 (required for attachment to all plants tested) functions via repression of a cyclic-di-GMP degrading factor to promote attachment. This work also builds on root-microbe interaction imaging technologies by developing a system suitable for *Rhizobium* and legume plants. Results reinforced the idea that the root elongation zone is a crucial region for early stage interactions, and that bacterial cell motility is important for this. Overall, this work significantly enhances our understanding of primary attachment mechanisms in *Rhizobium*-legume symbioses, demonstrating a previously unknown mechanistic complexity.

Acknowledgments

In the first instance, a debt of gratitude is owed to my supervisor Professor Philip Poole for his remarkable support throughout, and for imparting much wisdom in the process, both science and non-science related. Similarly, my appreciation goes to Professor Gail Preston who, in her role as second supervisor, provided invaluable guidance. Thirdly, to Dr Alison East. Your role was not bestowed with the title of supervisor, but without you I have no doubt this project would be a shadow of its present self. Although I will refrain from listing you all here, my thanks also extend to each and every member of the Poole Lab and Department of Plant Sciences who have steered this project and provided an engaging research environment in which to study over the last four years.

The funding sources for this project were twofold; I am particularly grateful to the Bellingburn Trust for their generous support through the Yeotown Scholarship of New College, and to the BBSRC for their financing through the Interdisciplinary Biosciences Doctoral Training Partnership. Receipt of this funding was crucial to my ability to undertake doctoral studies, and it is important that funders understand how critical their support is, both in terms of shaping scientific research and impacting student's lives.

New College, my 'home base' for the last four years in Oxford, deserves a special mention of its own here. The warmth, friendship and intellect of the College community has strengthened my resolve in times of difficulty, challenged me to broaden my horizons and been a thoroughly formative influence. Manners Makyth Man.

In the same vein, I am indebted to Cumberland Lodge and all the donors who made visiting scholarships at this charitable foundation possible. My thanks to Canon Dr Edmund Newell, Dr Owen Gower and Dr Rachel Smillie for believing I had something to offer, and for upholding the Lodge's mantra by encouraging a 'safe space for unsafe discussion' throughout.

Third on the institutional appreciation list is Campion Hall. The generosity of the community is commendable. In particular, I am eternally grateful to Rev'd Dr James Hanvey SJ and to Dr Philip Kennedy. Your wisdom, humanity and, most importantly, friendship is and always will be a pillar of strength. I hope many more can encounter the peace I was able to find inside the beautiful walls of the Hall.

To my friends: I wish I saw you all more often, as you are a crucial piece of this jigsaw puzzle. For Leah Lazar (one of life's great listeners), Izzy Gordon (you can always make me laugh), Hamish Dustagheer (providing Mediterranean escapes and more), Mark Blandford-Baker (being truly excellent company) and Dr Mark Byford (inspiring me on alpine retreats), a special mention is undoubtedly deserved.

Finally, to those who, with their love and support, have enabled all I have achieved in the last 27 years; I dedicate this thesis to my parents.

Abbreviations

AD	Advantaged
ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
Amp	Ampicillin
Amp^r	Ampicillin resistant
Apra	Apramycin
Apra^r	Apramycin resistant
ATP	Adenosine triphosphate
AU	Arbitrary units
BCA	Bicinchoninic acid
BLAST	Basic local alignment search tool
BNF	Biological nitrogen fixation
CA	Crude adhesin
c-di-GMP	Cyclic di-GMP
cDNA	Complementary DNA
CFU	Colony forming units
CRISPR	Clustered regularly interspaced short palindromic repeat
CPS	Capsular polysaccharide
DE	Defective
DNA	Deoxyribonucleic acid
DUF	Domain of unknown function
DWA	Distilled water agar
EDTA	Ethylenediaminetetraacetic acid
EPS	Exopolysaccharide
ES	Essential
FHA	Filamentous hemagglutinin adhesin
FP	Fahräeus plant
FV	Fitness value
gDNA	Genomic DNA
Gent	Gentamycin
Gent^r	Gentamycin resistant
GFP	Green fluorescent protein
GTP	Guanosine triphosphate
HAD	Dehalogenase-hydrolase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HITS	High throughput insertion tracking by deep sequencing

HMM	Hidden Markov model
HTH	Helix-turn-helix
INSeq	Insertion sequencing
IRLC	Inverted repeat-lacking clade
IT	Infection thread
Kan	Kanamycin
Kan^{r}	Kanamycin resistant
Kdo	3-deoxy-D-manno-octulosonate
LB	Lysogeny broth
LC-MS/MS	Liquid chromatography-mass spectrometry
LPS	Lipopolysaccharide
MOPS	3-(N-morpholino)propanesulfonic acid
NAD	Nicotinamide adenine dinucleotide
NAT	N-acetyl transferase
NCR	Nodule-specific cysteine rich
NDK	Nucleoside diphosphate kinase
NE	Neutral
Neo	Neomycin
Neo^{r}	Neomycin resistant
NFR	Nod factor receptor
NO	Nitric oxide
OML	Outer membrane lectin
OMP	Outer membrane protein
PCR	Polymerase chain reaction
PDMS	Polydimethylsiloxane
PE	Phosphatidylethanolamine
PEG	Polyethylene glycol
PGPR	Plant growth promoting rhizobacteria
PHB	Polyhydroxybutarate
POTRA	Polypeptide transport associated
PTS	Phosphotransferase systems
REZ	Root elongation zone
RNA	Ribonucleic acid
RNR	Ribonucleotide reductase
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEM	Standard error of the mean
SILAC	Stable isotope labelling with amino acids in cell culture
SOC	Super optimal broth with catabolite repression
Spec	Spectinomycin
Spec^{r}	Spectinomycin resistant
Str	Streptomycin

Str^r	Streptomycin resistant
TA	Thymine-adenine
TAE	Tris-acetate EDTA
TCA	Tricarboxylic acid cycle
Tet	Tetracycline
Tet^r	Tetracycline resistant
TPS	Two partner secretion
TraDIS	Transposon directed insertion site sequencing
TraSH	Transposon site hybridization
TRIS	Tracking root interactions system
TY	Tryptone yeast
UDP	Uridine diphosphate
UMA	Universal minimal agar
UMS	Universal minimal salts
UPP	Unipolar polysaccharide
UV	Ultraviolet
VMM	Vincents minimal media
VWA	Von Willebrand factor type A

Table of Contents

Chapter 1

Introduction

1.1 Planetary health: climate change and the nitrogen crisis

1.1.1 Climate change and the Anthropocene

Global health (human population health worldwide) and planetary health (encompassing the state of earth's natural systems) are inextricably linked; humanity is dependent on the health of multiple planetary systems for survival [1]. In the Holocene era, modern human civilisations developed and planetary changes (such as temperature fluctuations) were effectively buffered. The industrial revolution (characterised by fossil fuel use and industrial agriculture) enabled a new age, known as the Anthropocene, where humanity's impact became such that the planet began to move away from the stability of the Holocene [1–3].

Present day global health is considered better than at any time beforehand. Reductions in extreme poverty, higher life expectancy, lower child mortality and rapid healthcare and technology advances evidence this [4, 5]. This rapid development is underpinned by earth systems (oceans, forests, wetlands etc.) which provide both direct (e.g. food, fuel, water) and indirect (e.g. nutrient cycling) goods and services [3]. However, rapid population growth (see [6]) and continued unsustainable resource use risks irreversible earth system alteration, which may undermine human societal development and species survival [1].

In a comprehensive evaluation of planetary health, Rockström *et al* (2009) developed the planetary boundaries (PBs) framework and defined the current status of each boundary based on risk of large negative planetary perturbation [2, 7]. The PBs are shown in Figure 1-1, below. Note that climate change and land system change sit in the zone of uncertainty, whilst biosphere integrity and biogeochemical flows (especially nitrogen) have moved into the high-risk categories.

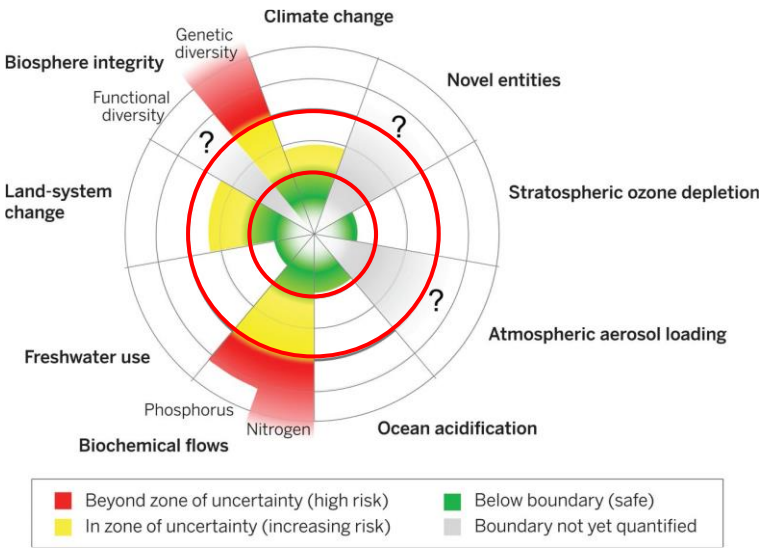


Figure 1-1. The current status of the nine planetary boundaries. The green zone represents the safe operating space. The yellow zone represents the zone of uncertainty and the red zone is the high-risk zone. Processes for which global-level boundaries cannot yet be quantified are represented by question marks. Reproduced from Steffen et al. (2015), [7].

The further these PBs are transgressed beyond the safe operating space the more likely dramatic planetary changes become. This, in turn, impacts on the security of humanity. Climate change itself (encapsulating all PBs) represents the greatest threat to humanity in the 21st century and is largely driven by fossil fuel combustion and expansion of agriculture [2]. Over 30% of non-ice or desert land globally has been converted for agricultural purposes, and this conversion continues, causing soil degradation, forest loss and water pollution [8–10].

1.1.2 Biogeochemical flows and the nitrogen cycle

Disruptions to the nitrogen and phosphorous cycle are major drivers of ecosystem change and decreases in planetary health. The nitrogen cycle is the biogeochemical flows of

different forms of nitrogen through the atmosphere (including terrestrial ecosystems) [11] (Figure 1-2).

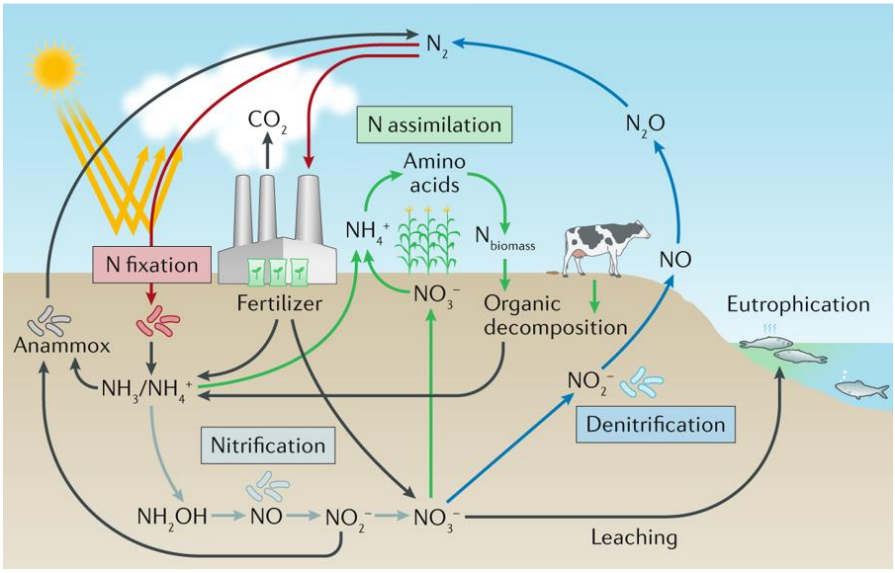


Figure 1-2. The Nitrogen cycle. Red arrows indicate reduction of atmospheric dinitrogen to ammonia (through biological or industrial processes), which provides fertilizer for plants (green arrows). Excess ammonia is processed by soil microorganisms in nitrification (light blue arrows) and denitrification (dark blue arrows), which, along with leaching, converts nitrogen-containing fertilizers into pollutants. Reproduced from Lehnert *et al*. (2018), [12].

Whereas biological processes for the conversion of atmospheric dinitrogen into reactive forms are largely driven by microorganisms and lightning strikes, anthropogenic forms are mostly due to the Haber-Bosch process. This process demands high temperatures and pressures for the conversion of nitrogen to ammonia and is responsible for up to 2 % of global energy use per year [13, 14].

Plants require biologically active forms of nitrogen (including ammonia) for growth [15]; indeed, production of nitrogen fertilizers is a large contributing factor to the increases in global food output over the last 10 years [16]. However, only 30-50 % of reactive nitrogen applied to fields ever reaches the intended crops. The rest leaches into terrestrial and aquatic ecosystems, with myriad negative consequences [17]. These include biodiversity loss, eutrophication of water systems and death of aquatic life [17, 18]. Oxides of nitrogen are particularly damaging given their potent greenhouse gas activity and direct negative impact on human respiratory health [3, 19].

1.2 Evolutionary answers: the nitrogen fixers

1.2.1 Biological nitrogen fixation

Biological nitrogen fixation (BNF) carried out by bacteria is responsible for ~65 % of the available nitrogen in the biosphere [20]. Although BNF can be carried out by various diazotrophic bacteria [21], the largest contribution to BNF comes from symbioses between *Rhizobium* (soil bacteria of the *Rhizobiaceae* family) and legume plants [22–24]. Within this symbiosis, rhizobia reduce atmospheric dinitrogen to ammonia (NH_4^+) via the nitrogenase enzyme complex. This ammonia is provided to the plant (acting as a bio-fertilizer) and bacteria receive carbon sources (mostly as dicarboxylates) in return [25, 26]. Annually, the nitrogen fixing symbiosis provide ~40 million tonnes of bio-available nitrogen into agricultural systems [27]. As fixed nitrogen is delivered directly from the bacteria into plant tissues, BNF reduces the detrimental side effects of fertilizer application, including eutrophication and greenhouse gas release [21, 24]. As these symbioses only occur with legume crops (such as pea, soybean and alfalfa), crop rotation

has traditionally been used (alternating planting of cereal crops with that of legumes) to increase soil nitrogen content, boosting crop yields. However, this technique is rarely employed in intensive agricultural settings [18, 28].

1.3 Trading places: lifestyle switching in the *Rhizobium*-legume symbioses

1.3.1 Signaling and the initiation of symbiosis

Prior to the initiation of symbiosis, rhizobia live a motile, saprophytic lifestyle in the soil in competition with a large microbial community. A single gram of soil contains up to 10^4 bacterial species and 10^9 bacterial cells [25]. The rhizosphere is an interesting zone of soil immediately surrounding plant roots, which is influenced heavily by exudates [29, 30]. It is in the rhizosphere that signaling processes leading to the formation of *Rhizobium*-legume nitrogen fixing symbioses begin.

The process of symbiosis formation is initiated by a two-way molecular dialogue between legume roots and symbiont *Rhizobium*. When soil nitrogen conditions are low (iso)flavonoids are secreted from legume roots which act both as rhizobial chemoattractants and inducers of Nod genes [25] (see Figure 1-3 and 1-4). Flavonoids themselves are functional secondary plant metabolites with different chemical moieties that convey signaling specificity with cognate symbiont rhizobia. Flavonoids are perceived by the NodD transcription factor, which acts as a positive inducer of the *nod* gene cluster in rhizobia [31]. These genes encode Nod factor, a recognition molecule specific to each *Rhizobium*-legume symbiosis which, when perceived by the plant, activates the symbiosis signaling pathway [32, 33]. The Nod factors themselves are lipochitooligosaccharides (LCOs) with an N-acetylglucosamine oligosaccharide backbone

and differential molecular decoration (including sulphonation, glycosylation, methylation, fucosylation and acetylation). They are crucial symbiosis specificity factors and determine recognition between different legumes, which perceive them using Nod factor receptors (NFRs). NFR activation elicits a calcium-based signaling response pathway *in planta* which triggers the development of root nodules and symbiosis formation [26, 33].

1.3.2 Primary root attachment

In the first physical interaction of nitrogen fixing symbiosis, rhizobia attach to root hairs (see Figure 1-3 and 1-4). Attachment also occurs to bulk root epidermis in a further type of non-nitrogen fixing symbiosis [34]. In this symbiosis, plant roots exude up to 20 % of their photosynthate through roots, and these exudates can serve as preferential growth substrates for rhizobia [34, 35]. Host plants also derive benefits of bulk root attachment. For example, the production of indole acetic acid and siderophores by root colonizing *Rhizobium leguminosarum* has been shown to promote the growth of tomato, peppers, maize and lettuce, among others [36–38]. These plant growth promoting rhizobacteria (PGPR) can also protect plant roots from pathogens, likely as a result of plant immune modulation (indirect mechanism) or pathogen exclusion (direct mechanism, see 5.1 for further discussion) [39–43].

Recently, a common biphasic model of root attachment has been proposed which exists across agriculturally relevant microbial species including *Rhizobium, Agrobacterium, Pseudomonas, Azospirillum* and *Salmonella* [34]. Following migration to the root (typically via flagella or pili mediated motility [44, 45]), universal, non-specific binding forces mediate interaction with the root surface. These include Van der Waals forces, electrostatic and hydrophobic interactions [34, 44, 46]. Motility assists in overcoming

repulsive electrostatic forces caused by bacterial cell envelope charge [44]. These initial, reversible interactions are followed by the first (primary) stage of root attachment, which relies on microbe specific factors.

Primary attachment is characterised by root interactions stronger than those mediated by universal forces, but which are still reversible and often involve surface proteins, polysaccharides and flagella [34]. Primary attachment mechanisms used can be affected by a range of different factors, including nutrient availability and soil pH.

A good example of this is the proposed model for pH dependent primary root attachment in *R. leguminosarum*, the cognate nitrogen-fixing symbiont of legumes including pea and vetch. Under acidic to neutral soil conditions, *R. leguminosarum* uses the polarly located surface polysaccharide glucomannan to bind root hair lectin [47]. Lectins themselves are widespread carbohydrate binding proteins that act as recognition molecules in cell-cell interactions and mediate specific yet reversible binding interactions [48, 49]. It is thought that acidic pH lectin-mediated attachment is also used in other *Rhizobium*-legume symbioses, as *Bradyrhizobium japonicum* can attach to soybean roots in this way [50, 51], and legume species often show high lectin content [48].

Under alkaline pH conditions, root lectins are solubilised, and glucomannan no longer mediates attachment [47]. It has been proposed that an extracellular rhizobial protein termed rhicadhesin mediates rhizobial attachment to an unknown plant receptor under these conditions. In this model rhicadhesin is bound to the bacterial cell wall by a calcium (Ca^{2+}) ion and may disassociate under acidic pH conditions [52]. Rhicadhesin is thought to be a 14 kDa protein and could inhibit rhizobial root attachment to pea roots when used as a pre-treatment [53], but neither the protein nor its encoding gene has yet been identified. *Agrobacterium* is also thought to make use of rhicadhesin under alkaline

conditions [53, 54]. However, the definition of rhicadhesin based on its ability to inhibit attachment may not necessarily show that it has a direct role in attachment, and it is also possible that multiple copies of the gene could contribute to difficulties in identifying this factor [55].

Multiple further genes are implicated in primary root attachment in *R. leguminosarum*, though to what extent these are required for root hair vs. bulk root attachment is not always clear. The importance of EPS was highlighted by a *pssA* (acidic EPS biosynthesis gene) mutation in *R. leguminosarum* biovar *viciae* 3841 (Rlv3841), strongly defective in root attachment at acidic or alkaline pH [56]. *Rhizobium* adhering proteins, the PlyB EPS glycanase, RosR cell surface component regulator and predicted cadherin attachment proteins were also implicated [57–59].

In *Agrobacterium*, surface proteins, molecular adhesins, pili and capsular polysaccharides (CPS) are involved in primary attachment [60]. In *Azospirillum*, primary attachment is also called adsorbtion, and is mediated by the polar flagellum [60–62]. Glycosylation of flagellin (the flagellar subunit) is known to be important, and occurs via the same genes involved in lipopolysaccharide (LPS) biosynthesis; mutation of these genes inhibits primary attachment [63]. Various outer membrane proteins (OMPs), which function both in root attachment and cellular aggregation, are also required [64]. It is additionally proposed that the LPS O-antigen of *Azospirillum* directly binds maize root lectin to mediate primary attachment [65].

Pseudomonas fluorescens utilises pili for primary attachment along with flagella and various surface polysaccharides, although for this bacteria the process is not particularly well characterised [66]. Outer membrane porin F (OprF) from *P. fluorescens* shows adhesive properties to wheat, barley, maize and tomato roots, among others [67, 68] and

implicates this as a primary attachment factor. In *Salmonella*, flagella, fimbrae and pili have all been shown to be important in attachment to *Arabidopsis* roots [69, 70]. A summary of these different factors is shown in 1-3, below.

Overall, this highlights the wide variety in primary attachment mechanisms used across different bacterial species. Despite this knowledge, it remains the case that primary attachment mechanisms in *Rhizobium*-legume symbioses (nitrogen fixing or bulk root) are vastly under-characterised in relation to other symbiotic stages [25].

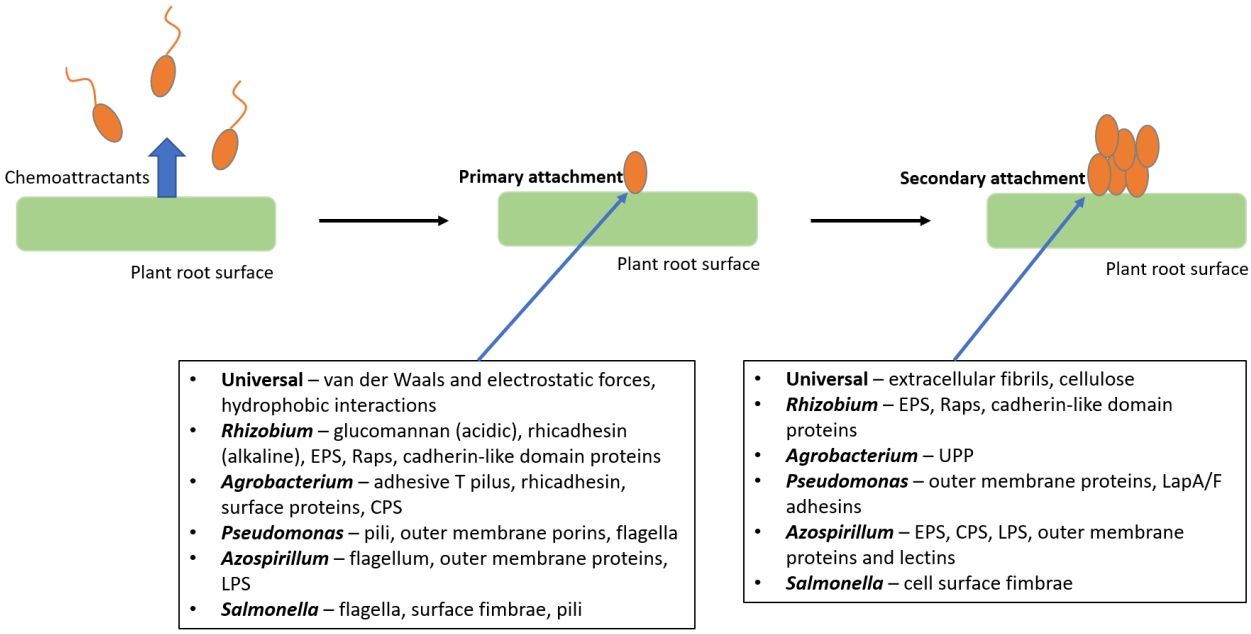


Figure 1-3. Bacterial root attachment mechanisms. In the first stage, chemoattractants exuded from roots cause bacteria to migrate to the root surface. Primary attachment of single cells to the root surface follows this and can be polar, with a variety of different factors involved depending on bacteria and environmental conditions. Secondary attachment follows, with bacterial aggregates becoming tightly bound to the root surface. Secondary attachment factors differ depending on bacteria. EPS – exopolysaccharide, Raps – *Rhizobium* adhering proteins, CPS – capsular polysaccharide, LPS –

lipopolysaccharide, UPP – unipolar polysaccharide. Adapted from Wheatley *et al*. (2018), [34].

1.3.3 Secondary root attachment

Secondary attachment follows primary attachment (often being dependent on it as a precursor stage, [60]) and is characterised by tight binding to roots, which typically involves extracellular fibril synthesis [71]. Secondary attachment secures microbial association with the root and can be a precursor to endophyte colonization [34]. For nitrogen-fixing rhizobia, cellulose fibrils, polysaccharides and secreted proteins permit the accumulation of bacteria at the site of initial primary attachment [34, 72]. Secreted proteins involved often contain Cadherin-like domains, and this has led Rap proteins to be classed as secondary attachment factors [73, 74]. However, the requirement for RapA2 and RapC proteins in a two hour attachment assay gives weight to the idea that they are likely involved in both primary and secondary attachment [59]. RapA1 (the most well characterised of the Raps) is a 30 kDa polar surface protein which functions in cell agglutination through binding EPS or capsular polysaccharide (CPS) [58], and overexpression of Raps enhances both stages of root attachment [59, 75].

Polysaccharides also fall into this dual-role category, particularly EPS. This major cell surface carbohydrate polymer is known as acidic EPS in *R. leguminosarum* and is an octasaccharide repeating unit of glucose, glucuronic acid and galactose in a 5:2:1 molar ratio [76]. On both abiotic and biotic surfaces, EPS deficient mutant strains show defective attachment and colonization; root attachment itself is compromised in both primary and secondary attachment stages [77].

In *Agrobacterium*, unipolar polysaccharide (UPP) is a polarly located polysaccharide analogous to rhizobial glucomannan [47], although the attachment it mediates is irreversible [55]. UPP production is thought to be stimulated by increased cellular cyclic di-GMP (c-di-GMP), which has a key role in attachment and biofilm formation [78, 79]. Cellulose fibrils are also utilised and facilitate bacterial aggregate formation on the plant surface; this mechanism of biofilm formation is also under c-di-GMP regulation [55]. The correlation between high c-di-GMP concentration in cells and biofilm formation (as well as low c-di-GMP levels and motility) has been demonstrated in multiple bacteria, including *E. coli*, *P. aeruginosa* and *S. enterica* [80]. Specific attachment and biofilm formation factors under the regulation of c-di-GMP signalling include type IV pili, EPS production and surface adhesin production, among others [81]. Intracellular c-di-GMP concentrations are modified by altering its rate of synthesis or degradation. Diguanylate cyclases synthesise c-di-GMP from two GTP molecules, and phosphodiesterases degrade it [79]. In terms of protein domains, a GGDEF active site motif typically represents diguanylate cyclase activity, whilst an EAL or HD-GYP active site motif represents phosphodiesterase activity [82, 83]. Whilst proteins can carry both GGDEF and EAL/HD-GYP domains, it is usually the case that only one is functional, or that a third regulatory domain manages protein activity [84, 85]. Recently, Little *et al.* (2019) highlighted the importance of c-di-GMP signalling in competitive wheat rhizosphere colonization of *P. fluorescens* [86]. The relationship between high c-di-GMP levels and positive regulation of attachment and biofilm formation has also been described for *Agrobacterium* and *S. meliloti* [87, 88].

In *Azospirillum* the biosynthesis of polysaccharides and extracellular fibrils is similarly associated with secondary attachment, including EPS and CPS, as well as LPS, OMPs and

outer membrane lectins (OMLs) [34]. Burdman *et al*. (2000) demonstrated that EPS concentration and composition directly affects cell aggregation, with higher arabinose content having a positive correlation with the levels of attachment and colonization [89]. A 67 kDa OML in *A. brasilense* specifically recognises EPS, and is thought to mediate adhesion between attaching cells through the formation of EPS bridges, promoting biofilm formation [62].

In *Pseudomonas*, cellulose fibrils contribute heavily to secondary attachment, alongside two large adhesin proteins, LapA and LapF. LapA is thought to act in both initial polar attachment and the initiation of biofilm formation. Initiation of LapF production in the latter attachment stage mediates cell-cell interactions to secure biofilms [90, 91].

Cellulose fibrils are also crucial to *Salmonella* secondary attachment and surface aggregate fimbrae nucleator proteins also aid in microcolony and biofilm formation [69].

Figure 1-3 (above) provides a summary of these different factors.

1.3.4 Infection thread formation

Following attachment of rhizobia to root hairs, Nod factor signaling induces the influx of calcium (Ca^{2+}) at the root hair tip [27]. This disrupts the apical growth profile of root hairs and leads to curling, which entraps the bacteria in the shepherd's crook structure [26] (see Figure 1-4). From here they derive entry to the plant through an infection thread structure [92]. As the infection thread is formed by an invagination of the plant cell wall, bacteria remain topologically external to the plant, even whilst traversing toward the inner root cortical cells [26, 72].

The infection threads are long passages which allow bacteria to migrate toward the base of the root hair cell. Division of infecting rhizobia at the leading edge of the infection thread

is thought to provide the force for infection thread progression, whilst plant cytoskeletal rearrangement in underlying cells directs the path of infection threads [93, 94].

An infection thread will travel the length of a cell, with bacteria released in the extracellular space. From here, further infection threads form to transport rhizobia toward the developing nodule, through epidermal and cortical cells [26, 95].

Infection thread formation is usually clonal, meaning that one initial *Rhizobium* cell multiplies within the thread and populates the nodule, free from competitor bacteria [72].

Given the intimate nature of the symbiosis and the resources invested by the plant in nodulation, checkpoints aside from Nod factor specificity exist to protect against non-symbiont or pathogen infection. It is thought that specific rhizobial surface polysaccharide recognition may play a role here, with cognate symbiont EPS suppressing plant defense responses. An example of this is the interaction of EPS receptor 3 (EPR3) of *Lotus japonicus* with the EPS of *Mesorhizobium loti*. The extracellular LysM domains of EPR3 selectively recognize cognate symbiont *Rhizobium* EPS, and an intracellular kinase domain transduces a signaling cascade that is required to sustain the infection thread in *L. japonicus* [96]

1.3.5 Nodule development and bacteroid formation

Within the root cortex itself, symbiosis signaling pathways leads to the concomitant cellular differentiation and formation of root nodules (specialized structures in which BNF takes place) during the growth of the bacterial infection thread [25]. Upon reaching the nodule in an infection thread structure, bacteria bud off from the tip of the thread and enter developing nodules as membrane enclosed symbiotic bacterial cells [72].

Two types of cortical nodule exist: indeterminate and determinate [72, 97]. Indeterminate nodules are most commonly formed on legumes found in temperate and tropical regions, are elongated and have different zones of development and persistent meristematic activity [72]. Determinate nodules, however, show a loss of meristem function after development, and are spherical, with no distinct zones [97]. The symbiosis between Rlv3841 and *Pisum sativum* results in the formation of indeterminate nodules, whereas in the *M. loti* and *L. japonicus* symbiosis determinate nodules are formed [72, 97].

In the process of becoming bacteroids, rhizobial cells undergo a differentiation process which results in the production of nitrogenase and accessory components. These permit the fixation of atmospheric dinitrogen into ammonia (see 1.4) [27]. Bacteroids can be viewed as nitrogen-fixing organelles; they are enclosed by a plant-derived symbiosome membrane across which ammonium is provided for the plant in return for carbon sources (mostly C4 dicarboxylates), which provide an energy source for this process [25, 26].

The terminal differentiation that rhizobia undergo to become bacteroids in indeterminate nodules involves genome endoreduplication, loss of cell division and large morphological changes [98]. Plant derived nodule specific cysteine rich peptides (NCRs) are thought to drive these changes in bacterial physiology. NCRs show homology to antimicrobial peptides, and therefore represent an adaptation of plant immunity in controlling endosymbionts [99, 100]. The presence of NCR peptides in inverted repeat lacking clade (IRLC) legumes as well as the older Dalbergioid lineage likely demonstrates convergent evolution; the independent evolution of terminal bacteroid differentiation in two plant lineages has developed to occur by similar mechanisms [101, 102].

In both determinate and indeterminate nodules, bacteria become metabolically dependent on their hosts in the nodule, which provides a protective niche for the bacteria. The

activation of many nitrogen-fixation related genes (including *nif* and *fix* genes) in the bacteria is governed by low oxygen conditions, maintained by leghaemoglobin. This high oxygen affinity protein buffers nodule oxygen concentration (important for nitrogenase function, see 1.4) and causes pink coloration of the nodule due to the presence of the protein heme group [103, 104].

Upon nodule senescence, bacteria from determinate and indeterminate nodules recolonize the soil. In the former this is via de-differentiation back to their free-living state, whilst in the latter this is via the release of undifferentiated bacteria from the infection thread [56, 105, 106].

An overview of the process from attachment to nodule formation, as well as a diagram of an indeterminate nodule, is shown in Figure 1-4, below.

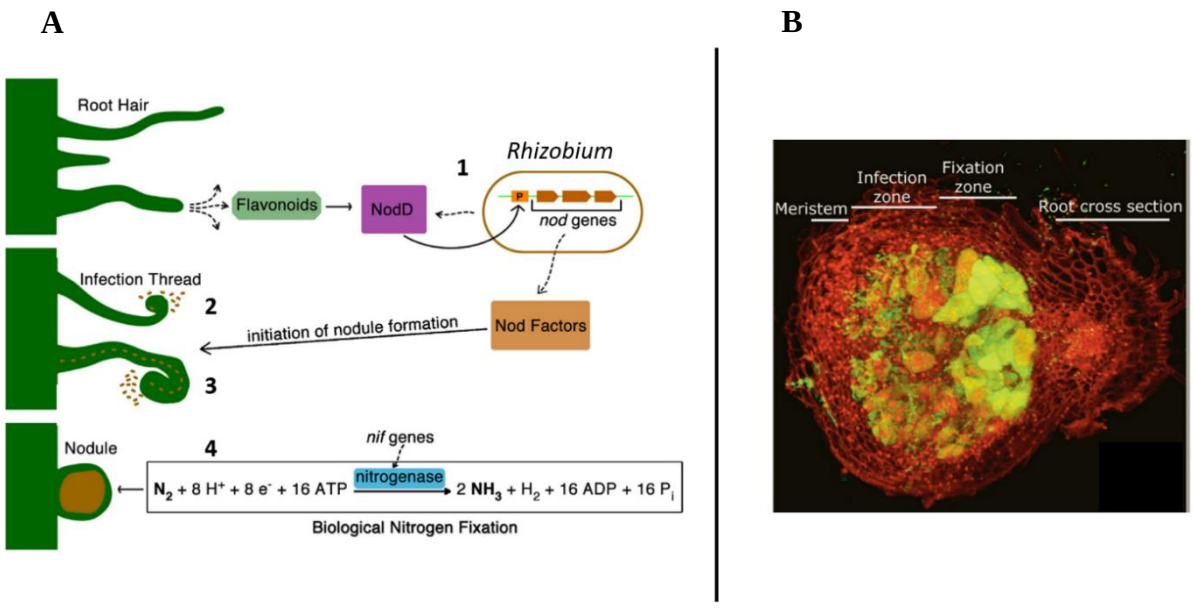


Figure 1-4. The formation of nitrogen-fixing symbiosis and structure of indeterminate nodules. **A** – symbiosis development, from free-living soil bacteria to nitrogen-fixing bacteroid. 1 – *Rhizobium* perceive flavonoid signals from host legume roots via the transcriptional regulator NodD, which induces Nod factor transcription. 2 – rhizobia attach

to roots and plant Nod factor perception induces root hair curling, trapping the bacteria in a shepherd's crook. 3 – infection thread formation begins with the invagination of the plant cell wall, which expands inwards toward the root cortex and allows bacteria to proceed toward the developing nodule structure. 4 – nodulation is completed as bacteria bud off from the infection thread and enter the nodule, differentiating to become nitrogen-fixing bacteroids. Fixation occurs via the nitrogenase enzyme complex in the reaction shown. Adapted from Laranjo *et al*. (2014), [107]. **B** – A longitudinal section from a 10-day old indeterminate alfalfa nodule infected with GFP labelled *S. meliloti* with the meristem, infection and fixation zones indicated. Plant tissue is stained red with propidium iodide. Adapted from Gage *et al*. (2004), [72].

1.4 Nitrogenase biochemistry

Nitrogenase is the metalloenzyme complex which is the catalyst for BNF. It catalyses the conversion of atmospheric dinitrogen to ammonia in the following reaction [108]:

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

The molybdenum-nitrogenase is the best characterised, being composed of one iron-molybdenum protein heterodimer and two iron protein dimers (each containing an iron-sulphur cluster and an ATP binding site) [109] (Figure 1-5). The iron-molybdenum protein is an $\alpha_2\beta_2$ tetramer with two metalloclusters in each $\alpha\beta$ subunit pair. These metalloclusters are the M cluster ($MoFe_7S_9C$-homocitrate) and the P cluster (Fe_8S_7) [110]. The iron protein dimer acts as the electron donor for N_2 substrate reduction in the M cluster [110].

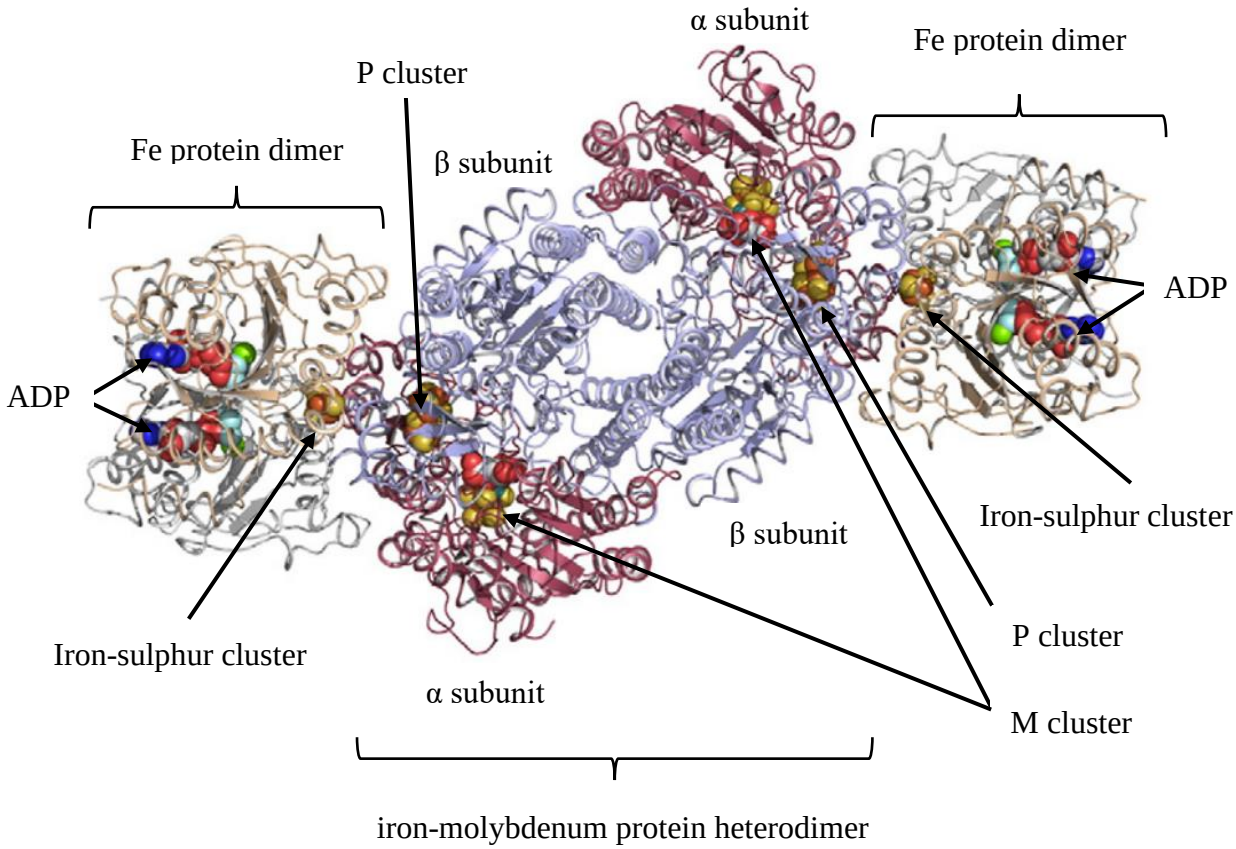


Figure 1-5. Molybdenum nitrogenase. The two iron (Fe) protein dimers are shown at either end of the complex with their iron-sulphur (Fe_4S_4) clusters. The central iron-molybdenum protein heterodimer is composed of two α and two β subunits, with two metalloclusters (one P - Fe_8S_7 – and one M - $MoFe_7S_9C$-homocitrate) in each αβ pair. Electrons for N_2 fixation are donated from the ATP binding site (where ADP is shown) of the Fe protein dimers inwards through the iron-sulphur cluster, to the P cluster and then to the M cluster. Figure adapted from Hu and Ribbe, 2013, [110]

Bacteroids are unable to synthesise their own homocitrate (a key M cluster component) and rely on plant provision for nitrogen fixation. Iron and sulphur are also provided by the plant across the symbiosome membrane [111].

Nodule leghaemoglobin is crucial for nitrogenase function, as oxidation of the enzyme iron-sulphur clusters blocks catalysis [27, 112]. The outer nodule layers form a physical barrier to oxygen diffusion and plant mitochondrial oxygen consumption also lowers nodule oxygen concentration [27]. Note, however, that leghaemoglobin also functions to deliver oxygen to respiring bacteroids which must meet the high ATP demands for nitrogen fixation [113, 114].

Of the dicarboxylic acid carbon sources provided in exchange for fixed nitrogen, malate and succinate are the most abundant, being imported into bacteroids via the high affinity dicarboxylic acid transport system [115–117]. These dicarboxylates enter the tricarboxylic acid (TCA) cycle, from which ATP and reductants are produced [118].

Ammonia is thought to be provided to the plant by diffusion out of the bacteroid across the lipid bilayer or through protein channels non-specific to ammonia, though these have not been identified [27, 109]. Outside the bacteroid, ammonia is protonated in the symbiosome space separating the bacteroid and the plant before transport across the symbiosome membrane through protein channels such as AmtB [109, 111, 119]. Following transport to the plant cell, ammonia is assimilated into organic forms before transport throughout the plant tissue [111].

1.5 Can we fix it? Harnessing nitrogen fixing symbioses

1.5.1 Enhancing existing symbioses

Given the environmental problems caused by overuse of nitrogenous fertilizers, as well as the need to produce more food for a growing population (see 1.1), naturally evolved nitrogen-fixing symbioses present significant opportunities for increasing the

sustainability and output of agricultural systems. For the legumes, this could be achieved via enhancement of existing symbioses. Various factors affect the efficiency of nitrogen fixing symbioses. These include phosphate availability, temperature, moisture, light, soil acidity and soil salinity [120]. The diversity of these abiotic factors highlights the importance of good management of agricultural soils to optimize these parameters. Optimizing soil parameters would also mean that crop planting could be managed more effectively.

Further, identification of the most effective rhizobial nodulating strains for different plant species will be important for optimizing agricultural inoculants. This is because legumes are not nodulated by the strains most effective in nitrogen fixation, but by those strains most competitive in nodulation, which are not always the same [120]. A longer-term goal (due to the complexity of research) is to fully characterise how plant microbiome composition affects symbiotic efficiency. Different microbiome components can prove positive or detrimental for nitrogen fixing symbiosis efficiency, and characterising these effects more fully could allow tailoring of the microbiome to maximize symbiotic benefits [120, 121].

A further possible avenue for enhancing symbioses is genetic engineering of the plant or endosymbiont. Currently, the focus of this research area is to develop a more holistic understanding of both legume and bacterial genetics, and their complex interplay in symbiosis development. Emerging new technologies in genetic screening and manipulation, particularly with CRISPR/Cas genome editing technology, show large promise for agricultural application, including in nitrogen fixing symbiosis enhancement [122, 123].

1.5.2 Synthetic symbiosis approaches

Despite the importance of nitrogen-fixing symbioses for legume crops, cereals (the highest contributors to global calorie intake, providing almost half of all calories consumed, [124, 125]) do not possess this ability and must rely on biologically available nitrogen in the soil or the addition of nitrogenous fertilizers [35]. Thus, the introduction of nitrogen fixation abilities to cereals has long been a goal of agricultural biotechnology research.

There are three avenues which form the main bulk of research into developing synthetic symbioses for non-legume crops. The first of these is to introduce the entire symbiosis pathway from legumes into cereals and engineer rhizobia for compatibility with cereal nodulation. This could be achieved in part by modification of the ubiquitous plant-arbuscular mycorrhizal symbiosis signaling path, which shows extensive overlap with nitrogen fixing symbiosis signaling [126–128]. The second approach is to engineer *in planta* nitrogenase expression. This necessitates not only correct expression and assembly of the multiple enzyme subunits, but also engineering of a suitable low oxygen environment for enzyme function (although plant organelles, notably the chloroplast, could prove suitable for this application, [128]). However, the complexity associated with these two approaches means that it may be another 15 years before genuine solutions can be delivered [128].

The most promising approach to synthetic symbiosis is the re-engineering of pre-existing cereal-endophyte or root colonizer associations, such that bacteria are provided with nitrogen fixation ability. This was reportedly achieved in 2013 with the transfer of nitrogenase biosynthesis genes to *Psuedomonas protegens*, which was able to release ammonium and promote growth of alfalfa and maize [129]. The shortcoming of this strategy is the high metabolic cost of nitrogen fixation to bacteria, meaning engineered

bacteria are likely to be rapidly outcompeted by other soil microbes under field conditions [35, 130]. One approach to overcoming this issue is to engineer specificity signaling between plants and bacteria which could bias the rhizosphere in their favor and act as a signal to express nitrogenase genes. Rhizosphere biasing has been demonstrated in transgenic plants producing opines, with the rhizosphere community remodeled in favor of opine catabolizing bacteria [131]. Placing the expression of nitrogen fixation genes under the control of a specific plant signal which can act as a carbon source for a 'synthetic symbiont' could help overcome limitations to this approach [35, 130, 132].

1.6 The *Rhizobium* genus

1.6.1 Rhizobial taxonomy

The *Rhizobiaceae* family are gram negative soil proteobacteria, of which ~100 species have been described in five main genera: *Rhizobium*, *Bradyrhizobium*, *Azorhizobium*, *Mesorhizobium* and *Sinorhizobium* [133, 134]. The rhizobial genomes are large and consist of a circular chromosome and a series of plasmids [135, 136]. The complex rhizobial genomes reflect the challenging soil environment in which they are found, where microbial resource competition can be high and energy/nutrient sources take many forms [137, 138]. The nodulation (*nod*) and nitrogen fixation (*nif* and *fix*) genes are either encoded on a symbiosis plasmid or colocalised as a 'symbiotic island' in the chromosome [135].

R. leguminosarum is a species with distinct symbiovars and has been extensively researched [138]. The symbiovar *viciae* nodulates *Viciae* legumes including pea (*Pisum sativum*) and vetch (*Vicia cracca* and *hirsuta*, among others) [139, 140]. Rlv3841 was

used throughout this work because of its fully sequenced genome, variety of transcriptomic and physiological data available, well characterised growth conditions and symbiotic engagement with agriculturally important legumes [138, 139, 141–143].

1.6.2 Rlv3841

The genome of Rlv3841 is 7.75 Mb and consists of a circular chromosome (5.08 Mb) and six plasmids: pRL7 (0.15 Mb), pRL8 (0.15 Mb), pRL9 (0.35 Mb), pRL10 (0.49 Mb), pRL11 (0.68 Mb) and pRL12 (0.87 Mb) (Figure 1-6). Sequencing of this genome enabled prediction of 7263 protein coding genes, 65% of which are chromosomally encoded [138].

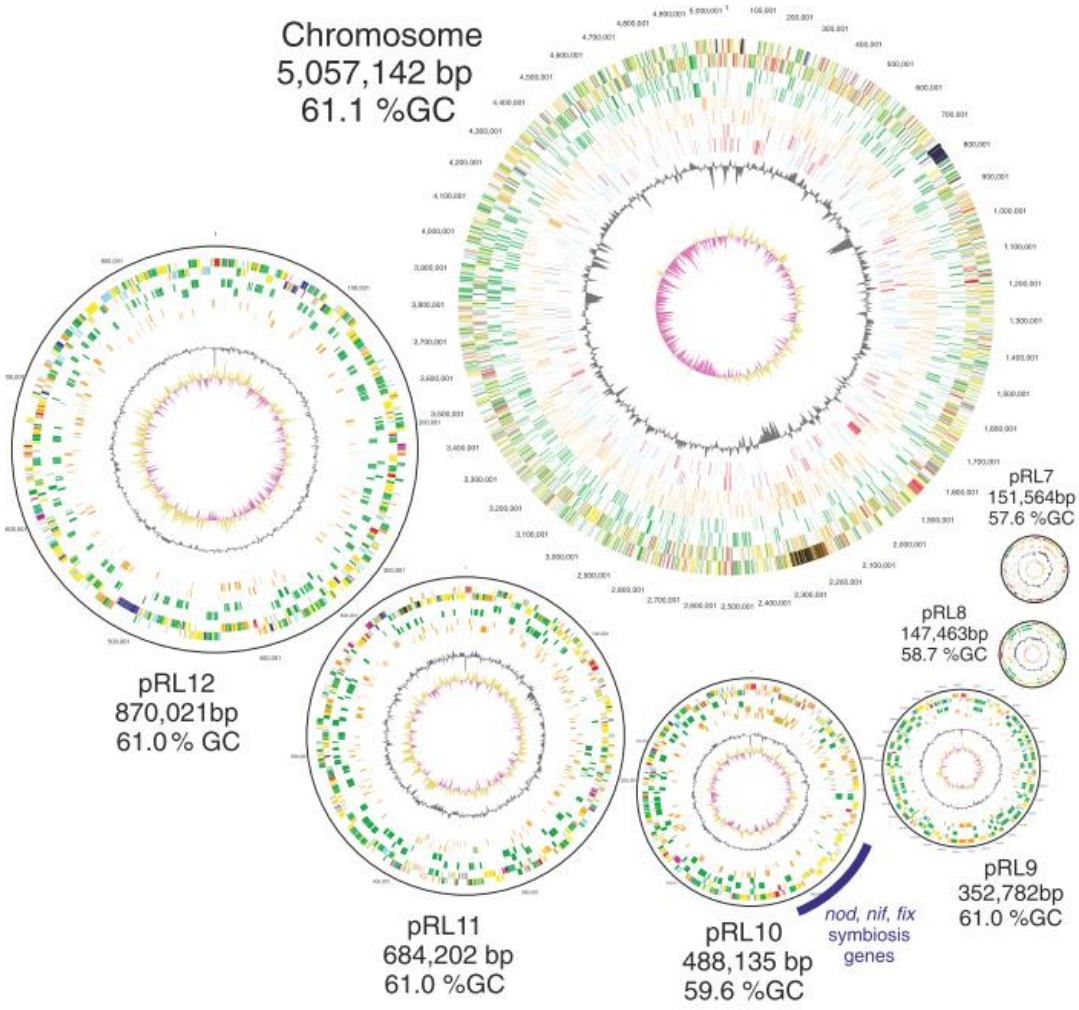


Figure 1-6. The chromosome and plasmids or Rlv3841. Plasmids are shown at relative scale, the chromosome as 25% relative scale. Gene classifications are indicated by color: membrane proteins (bright green), conserved and non-conserved hypotheticals (brown and pale green, respectively), phage and transposons (pink), DNA transcription/restriction/helicases (red, shown for chromosome only) and transcriptional regulators (blue, shown for chromosome only). Inner circles describe deviations in GC content (black) and GC skew (olive/maroon). Figure reproduced from Young *et al.* (2006), [138].

Most genes essential for cellular function (e.g. ribosomal subunits) are chromosomally encoded. Plasmids replicate and partition to daughter cells via the *repABC* system [144]. pRL10 is known as the symbiosis plasmid, and encodes the *nod* genes, as well as nitrogenase enzyme components (the *nifHDKEN* cluster) and electron transfer proteins (*fixABCX*) [138]. pRL8 has been classified as a rhizosphere specific plasmid, with genes selectively expressed in this environment, although many remain of unknown function [141]. pRL11 and 12 are considered the most 'chromosomal' of the plasmids, showing the most chromosome-like gene and nucleotide composition, and have thus been designated as 'chromids' [145]. pRL7 is the least well characterised of all the plasmids, with the majority of genes having no known function [138].

Around ~35% of the Rlv3841 genome is considered 'core', indicating evolutionary conservation. The remaining ~65% is considered 'accessory' (generally encoded on plasmids or chromosomal islands) and is not widely conserved [138, 140]. The high proportion of accessory genes likely reflects the environmental adaptations of Rlv3841. Interestingly, ~35% of the annotated Rlv3841 genes remain uncharacterised, and assigning function to these will help develop a more holistic understanding of organism function [138].

1.7 'Omics approaches to understanding gene function in rhizobia

Improving our understanding of gene function's relation to phenotype in symbioses between *Rhizobium* and legumes is important on multiple fronts. For nitrogen fixing symbiosis development, a more complete characterisation of mechanisms could aid with both enhancing existing symbioses and developing synthetic ones (see 1.5). The symbiosis that occurs when rhizobia attach to roots is also important. Here, carbon-rich exudates are

gained whilst rhizobia can shape a beneficial microbiome, produce plant growth promoting hormones, provide nutrients and protect plants from pathogens [36–38, 41–43]. A better understanding of these characteristics has clear applications for improving plant health and yields. Primary attachment is a particularly important, and under-characterised stage of both these symbioses [25, 34].

Relevant to the improved characterisation of primary attachment mechanisms are genetic, transcriptomic and proteomic approaches which can link gene function to phenotype. The availability of sequenced genomes for Rlv3841 and other rhizobia is critical in allowing such ‘omics based experimental approaches. Functional ‘omics approaches are now particularly important as, whilst gene discovery and sequenced genome availability has increased rapidly, knowledge of gene function has lagged behind [146].

Genetic approaches are often based on mutagenesis strategies, whereby gene function is disrupted, and subsequent phenotype observed. Transcriptomic approaches quantify gene transcripts (as RNA or cDNA) to show how gene expression changes under different conditions, which can also point to gene function and/or regulatory networks. Proteomics generally focusses on large scale protein identification (which shows functional protein, rather than transcript, expression changes) but can also be used for investigating protein localization, interaction partners and structure [147].

1.7.1 Genetic and genomic approaches

As discussed above, genetic approaches often use mutagenesis to assign gene function based on phenotypic effects. Multiple studies demonstrate the usefulness of such an approach in *Rhizobium*. Williams *et al.* (2008) used a Tn5 transposon cassette to disrupt the Rlv3841 glucomannan gene. Using the resulting strain in root attachment and

nodulation competitiveness assays at different pHs allowed them to better define the role of glucomannan in primary attachment and downstream symbiosis formation [56]. Hosie *et al*, (2002) used similar mutagenesis approaches (as well as various transporter activity assays) to define the role of MctP as the first permease of a new transporter subfamily, important for monocarboxylate transport in Rlv3841 [148]. White *et al*. (2009) used mutagenesis to disrupt the *gtsR* gene in Rlv3841, demonstrating (in combination with transcriptomic approaches) that this gene regulates a γ-aminobutyric acid uptake operon [142]. Whilst there are many more examples from single gene mutagenesis studies, such approaches are time and labour intensive, making them more suited for targeted studies than genome screens for entirely new functions.

This limitation can be overcome by isolating and screening large mutant libraries under test conditions, with identification of mutations leading to phenotypes of interest [149]. Multiple methods exist for this, including chemical and UV mutagenesis, as well as transposon insertion. These techniques are often faster and cheaper than targeted mutagenesis approaches, and are particularly useful when genomes are not well characterised [150]. In genome-wide transposon insertion studies, the site of mutagenesis can be identified by high throughput sequencing of the transposon and its neighbouring gDNA, and this principle has proven key to unlocking large amounts of knowledge on gene function in different organisms [149].

Of the available transposon mutagenesis tools, Tn5 is particularly notable for studies in *Rhizobium* [150, 151]. Tn5 transposon mutagenesis is random and can be highly saturating, with almost no insertion bias, allowing the whole genome to be reliably queried under test conditions [152–154]. Prell *et al*. (2012) used a Tn5 mutant library to isolate colonies with a dry morphology on TY media. By cloning and sequencing the

flanking gDNA from these transposon mutants it was shown that the *ptsP* gene had been disrupted. This genotype-phenotype linkage, combined with subsequent transporter assays, identified a key role for the PTS^{Ntr} system in the global regulation of Rlv3841 ATP transporter activity [155].

Whilst Tn5 mutagenesis screens are powerful tools, there are limitations: random saturating mutagenesis does not allow targeting of gene subsets, library composition can be biased to exclude low fitness mutants and compromised strains face a selective pressure to develop second site suppressor mutations, which can mask the effects of transposon insertion [143, 146, 151].

1.7.2 Transcriptomic approaches

Transcriptome level techniques aim to use measures of RNA abundance to quantify the expression of genes in absolute or relative terms [147]. Identifying the conditions under which a gene is transcribed can be useful in inferring its function, but also in determining its regulatory relationships with other genes. The techniques reported most widely are microarrays and RNA sequencing (RNASeq) [156, 157]. Microarray technology relies on the reverse transcription of cellular RNA to cDNA, which is labelled with a fluorescent probe. Genomic probe sequences are printed on a glass microarray slide and used as hybridization targets for labelled cDNAs. Fluorescent signal from each probe spot is used as a proxy for transcript abundance at the probe site [156]. One of the major drawbacks of microarrays is that quantification of gene expression is limited to genomic regions of known sequence. Microarrays have been successfully applied to the studying Rlv3841 gene expression, including in metabolic studies with glucose, pyruvate, succinate or

acetate as carbon sources, and to investigate rhizobial adaptations in the rhizosphere of host and non-host plants [139, 141].

RNASeq has largely supplanted microarrays and enables unbiased, high-throughput and direct quantitation of cellular transcripts via sequencing and bioinformatic analysis [157]. By applying RNASeq to a wild-type and *rosR* mutant strain of *R. leguminosarum* biovar *trifolii*, it was shown that the *rosR* gene functions as a transcriptional repressor with a role in regulating polysaccharide production, motility and aspects of metabolism [158]. More recently, RNASeq was used to compare bacteroid gene expression profiles in determinate and indeterminate nodules using two *R. leguminosarum* strains isogenic apart from their sym plasmid. This shed light on the different conditions bacteria face in each nodule type, including higher levels of metabolite detoxification activity required in determinate (*Phaseolus* bean) nodules and increased expression of DNA replication genes in indeterminate (pea) nodules, consistent with endoreduplication [159].

1.7.3 Proteomic approaches

An advantage of proteomic studies in relation to transcriptomics is their ability to show changes in gene expression at the protein level, which is often the functional unit of the gene. This is an important difference, as protein production can be modified by many factors additional to expression of the encoding gene (see [160] as an example). Further, proteomic studies can shed light on protein localization in the cell, interaction partners and posttranslational modifications. The latter are difficult to predict from transcriptional or genomic data, but can have a large effect on protein function [161].

Often, proteins in biological samples have been studied with 2D gel electrophoresis, which separates and purifies them based on molecular weight and isoelectric focusing

point. Relative protein abundance can then be compared using peptide staining in Western blots. Mass spectrometric peptide sequencing can also be used to characterise proteins and their modifications [147]. Using peptide sequencing and the known codon preferences of an organism can allow encoding genes to be identified from their protein products, and such an approach could be of use in better characterising rhicadhesin (see section 1.3.2). Liquid chromatography protein separation techniques and tandem mass spectrometry (LC-MS/MS) is a more sensitive method of protein separation and analysis and can be applied to protein bands partially purified by gel electrophoresis. Several studies have demonstrated the application of proteomic approaches to characterising nitrogen-fixing symbioses. Comparative proteomics studies of *S. meliloti* in free living state or in symbiosis with alfalfa demonstrated the cellular remodelling that occurs, with nitrogen fixation proteins, amino acid ABC transporters and stress related proteins all upregulated [162–164].

1.8 High throughput whole-genome screening with insertion sequencing

A key aim of this work was to better characterise the range of primary attachment mechanisms displayed by Rlv3841 under different environmental conditions. This is important from the perspective of both basic research and subsequent symbiosis development (be it nitrogen fixing, or bulk root-attached). In the development of this project, it became clear that a high throughput genome wide screen would be appropriate to address this aim. High throughput screening techniques can provide large mutant libraries and allow screening of both characterised and uncharacterised genes for function under test conditions. Four methods using this approach in bacteria were published in

2009: transposon sequencing (Tn-seq, [165]), insertion sequencing (INSeq, [166]), high throughput insertion tracking by deep sequencing (HITS, [167]) and transposon directed insertion site sequencing (TraDIS, [168]).

Both Tn-seq and INSeq use a transposon cassette with a type IIS restriction enzyme site, such as MmeI for the *mariner* transposon used in INSeq [165, 166]. As type IIS restriction enzymes cleave target DNA a defined number of bases downstream of their recognition site (which is, in this case, in the *mariner* transposon) a gDNA sequence of 16 bp adjacent to the transposon insertion is captured. Sequencing this tag and aligning it to the genome reveals the location of *mariner* insertion, and therefore which gene is mutated in different cells [169]. HITS and TraDIS differ in that the transposons used (*himar1* and a derivative of EZ-Tn5, respectively) do not have type IIS restriction sites. Instead, random shearing of DNA by sonication yields gDNA tags for mapping [167, 168]. Whilst the sonication method makes HITS and TraDIS compatible with any transposable element, Tn-seq and INSeq require the type IIS restriction site modification. However, the mapping tag isolated remains at 16 bp, ensuring that all reads generated are mappable [146].

Fundamentally, these techniques harness the power of high-throughput short-read sequencing platforms to allow millions of diagnostic DNA fragments to be sequenced simultaneously. These approaches have now replaced previous methods, some of which used tag arrays. An example of this is transposon site hybridization (TraSH) and derivatives, which used microarrays to detect insertions. By comparing the hybridization signal intensity of insertion locations before and after selection in test conditions, relative gene importance could be assessed. However, different hybridization characteristics for each microarray spot and missing of smaller genes are key disadvantages [170, 171].

1.8.1 *Mariner* insertion sequencing

Forms of the *mariner* transposon insertion sequencing approach have been demonstrated for assessing gene requirements in several bacterial species. These include defining the essential genes for *Porphyromonas gingivalis* survival [172], as well as for *Mycobacterium tuberculosis* growth and cholesterol catabolism [173] and sRNAs required for pathogenesis in *S. pneumoniae* [174]. Perry and Yost (2014) described a modification of the MmeI-adapted *mariner* transposon delivery vector pSAM_Bt [151, 166]. By replacing the erythromycin resistance cassette with a neomycin cassette and the *Bacteroides thetaiotaomicron rpoD* promoter with that of Rlv3841, they produced the *Rhizobium* compatible pSAM_Rl vector. This is a suicide vector, meaning that in a bacterial mutant library each cell will contain only one genomic insertion [151]. A map of this vector is shown in Figure 1-7, below.

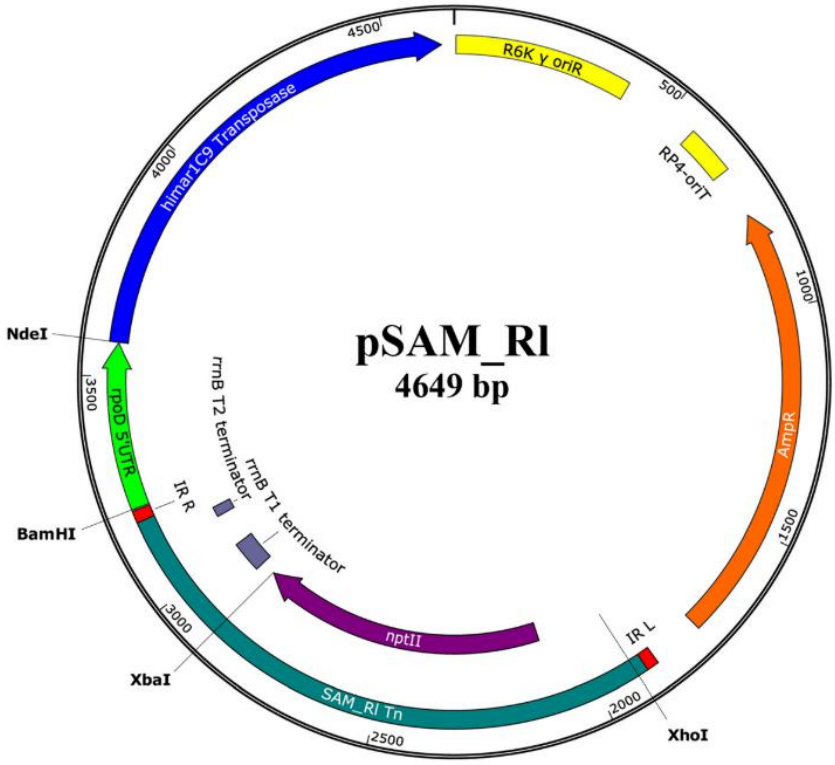


Figure 1-7. Plasmid map of pSAM_Rl. Antibiotic markers (ampicillin, AmpR, neomycin/kanamycin, *nptII*) are shown, as well as origin of replication (R6K y oriR), transposase (himar1C9), transposase promoter (*rpoD* 5'UTR), MmeI-adapted *mariner* inverse repeats (IR_R, IR_L), and the transposon borne Rho-independent terminator (*rrnB* T1 and *rrnB* T2). Reproduced from Perry and Yost (2014), [151].

Since this time, pSAM_Rl has been used to describe gene requirements for Rlv3841 growth on TY media [151], Vincent's minimal media (VMM) [175] and under 21% and 1% oxygen conditions with glucose or succinate as carbon sources (relevant to understanding metabolism under bacteroid conditions) [143].

The key benefit of *mariner* transposons is their specific insertion at thymine-adenine (TA) motifs [176]. Unlike the use of Tn5 (with no insertion preference), the use of *marniner* elements allows transposition events to be comprehensively modelled *in silico* prior to mutant library analysis [151]. This allows for robust statistical analysis of gene essentiality, as well as analysis of specific intra-gene disruption effects at different TA sites and focusing of analysis on defined regions of the genome [151]. The overall advantage of this is that defining gene requirements for specific test conditions does not just rely on input and output mutant pool comparison. Rather, it can use statistical inference based on the known number and location of insertion sites in the genome to infer under or over represented insertions in gene regions when determining gene essentiality [143, 151].

Barquist *et al*. (2013) argued that the specific TA site requirement for *mariner* insertion provides a limit on potential insertion sites that can bias screen results, particularly for GC rich genome regions. This may result in genes with very few or no insertions, and

therefore unreliable classifications in downstream statistical analysis [150]. However, they also note the advantages of defined site insertion for statistical analysis; insertions specific to TA sites allow more accurate predictions of gene essentiality in near-saturated libraries using statistical methods (see 1.8.3), given that every potential integration site is known and the probability of insertion at any given site can usually be assumed to be almost equal.

The genome of Rlv3841 is particularly suitable for *mariner* transposon studies, with 140,056 TA sites in total, an average of 15 per gene, and only 21 genes lacking TA sites [151, 175].

1.8.2 Principles and methodology of insertion sequencing

As reviewed by Barquist *et al*. (2013), all transposon mutagenesis and sequencing methods rely on similar principles. Large, highly saturated libraries of insertion mutants (the input library) are subject to a selection pressure, with transposon tag DNA isolated from both the input and subsequent output library. Sequencing transposon tags and comparing gene insertion frequency before and after selection pressure allows genes to be classified, as deleterious insertions will be reduced in the output [150]. Figure 1-8 provides an overview of this workflow, whilst Figure 1-9 provides a schematic of the process of library DNA preparation, which is carried out before sequencing of input and output samples for INSeq.

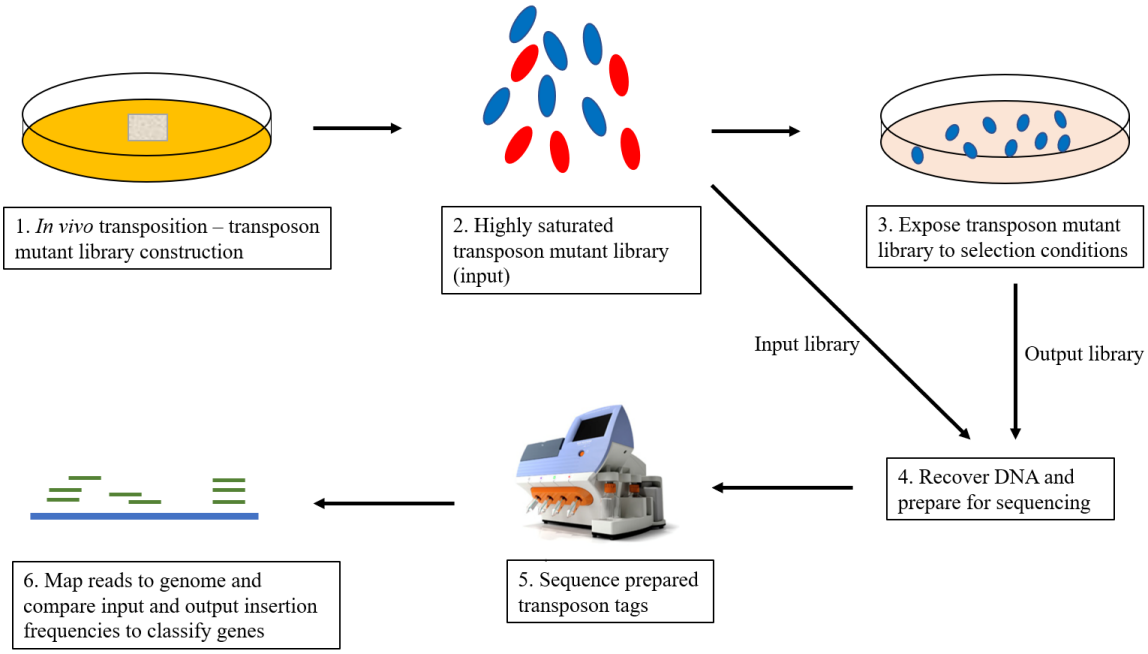


Figure 1-8. A simplified workflow for transposon insertion mutagenesis studies. **1** – transposon cassettes are conjugated into target organism cells for mutagenesis. Here, a filter conjugation is depicted on a rich media plate. **2** – From the conjugation a library of highly saturated transposon mutants is isolated. Shown in blue are cells carrying mutations that produce no phenotype under selection conditions (step 3), whilst red cells are those with deleterious insertions. DNA is isolated directly from a sample of this input library (step 4). **3** – Expose transposon mutants to selection conditions. Here, selection on a media type is shown, and the deleterious mutants (red cells) disappear from the population. **4** – Recover DNA from both input and output library and prepare for sequencing (see also Figure 1-9, below). **5** – Sequence prepared transposon tags (an Ion Proton sequencing system is shown here as an example). **6** – Map reads to genome and compare input and output library insertion frequencies. Loss of insertions in the output library indicates genes deleterious when mutated under selection conditions and allows gene classification.

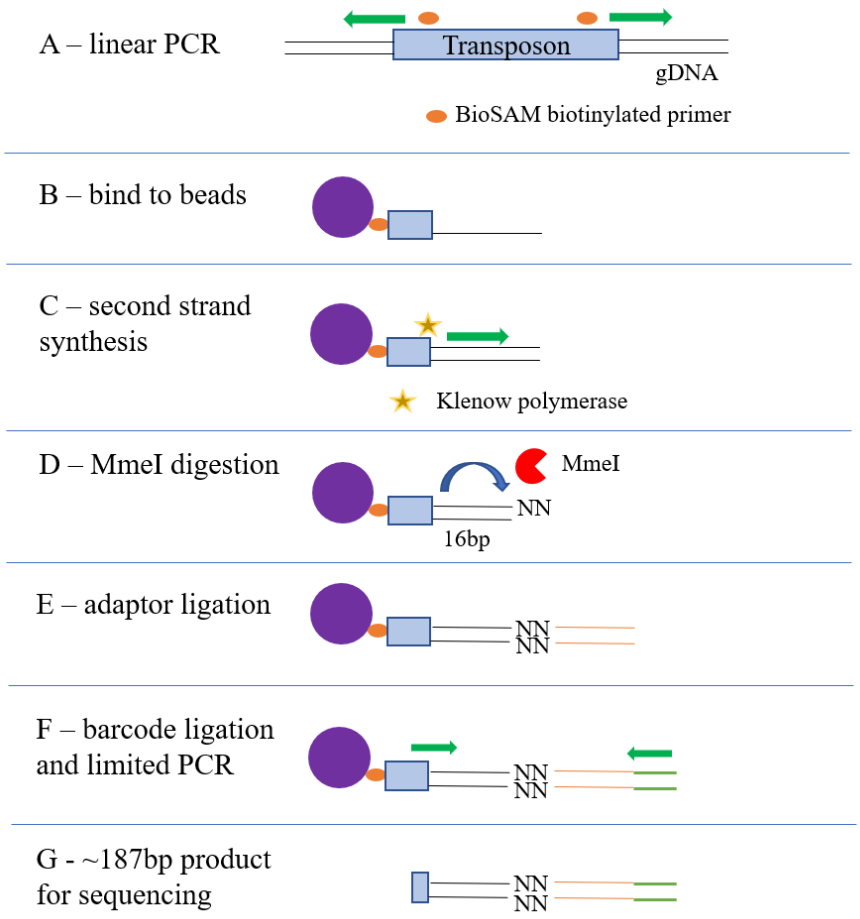


Figure 1-9. A workflow of library DNA (transposon tag) preparation for INSeq. Genomic DNA is extracted from input and output mutant libraries. **A** – linear PCR amplifies out from the *mariner* transposon using a BioSAM 5' biotinylated primer. **B** – Linear PCR products are purified with streptavidin beads, which bind the biotin tag. **C** – Hexanucleotide primers and Klenow polymerase are used for second strand synthesis. **D** – DNA bound to streptavidin beads is digested with the type IIS restriction enzyme MmeI. This cuts 16 bp downstream of the *mariner* insertion, forming the genomic 'transposon tag' that is used for mapping. A two nucleotide 'NN' overhang is left after cleavage. **E** – Adaptor sequences are ligated to the NN overhang (shown in orange). **F** – Sample specific sequencing barcodes (shown in green) are ligated to adaptors, allowing identification of

which library sequencing reads originated from. A final limited PCR removes samples from streptavidin beads and yields ~187 bp products which include the transposon tag for sequencing (**G**). For simplicity, only one side of the transposon is shown in B-G, but both sides are amplified and sequenced using this method. Figure adapted from Goodman *et al*. (2011) [169].

A highly saturated mutant library is desirable in order to allow accurate assessment of gene insertion effects before and after exposure to selection conditions. To achieve this, multiple independent transposon conjugations (isolated following *in vivo* conjugation of the transposon construct to cells of the target organism, see Figure 1-8) can be pooled together to generate a 'master' input library. In addition, isolation of libraries under rich media conditions can minimize the number of insertions deleterious in library construction, again permitting higher input saturation [143].

1.8.3 Statistical approaches to analyzing insertion sequencing data

Using transposon mutagenesis screens, the effect of disrupting a gene under test conditions can be deduced by comparing the frequency of transposon insertions present before and after exposure of mutants to selective conditions. The central rationale is that insertion events deleterious under the selection condition will reduce in frequency compared to the input, allowing a gene to be classified. Using the TraSH approach, probes generated from insertion sequences in the output library (after exposure to selection conditions) were hybridized to microarrays with randomly labelled chromosomal DNA probes, which hybridize to every microarray spot. Spots hybridizing with the genomic, but not output, probe indicated genes required under selection conditions [171].

Newer technologies relying on quantitative, high-throughput sequencing of input and output libraries make it possible to apply far more nuanced and precise statistical analyses to categorize gene disruption effects. It is important to note that all these statistical methods rely on three central assumptions [177]:

- abundance of insertions in the input / output library indicates gene fitness changes under test conditions
- transposon insertion in unbiased (meaning, in INSeq, that all TA sites have equal likelihood of being targeted by *mariner*)
- reads obtained from sequencing are representative of the real insertion profiles of the input and output libraries.

These statistical models vary in complexity. 'ESSENTIALS' was one of the first developed as a simple, automated tool for transposon insertion sequencing data analysis. The software predicts gene essentiality by mapping and comparing the sequenced reads per gene with the expected number of sequenced reads. This is calculated based on library sequencing depth, size and number of insertion sites per gene. The cutoff for gene essentiality is calculated based on the extent of gene underrepresentation in the output library. The model is benchmarked on Tn-seq data from *S. pneumoniae* with 49 known essential and 49 known non-essential genes [178]. ESSENTIALS thereby classifies genes into binary essential / non-essential categories under test conditions.

Bayesian models offer more nuance than ESSENTIALS and rely on the principle that the probability of gene essentiality can be refined continually as higher numbers of insertion sites in each gene are sequentially included in the analysis. This method identifies the

largest consecutive stretch of TA sites in a gene and estimates the likelihood that such an insertion-lacking region could occur by chance. Each individual TA site is treated as independent and classed as essential or non-essential based on defined background insertion frequency in non-essential genes [179]. This approach allows the statistical significance of runs of TA sites lacking insertions in libraries to be assessed, and gene essentiality defined. Using this method allows genes to be classed as essential even if insertions are present at the gene termini. This is important, as previous work has shown that transposon insertions in the extreme N and C termini of essential genes can sometimes be tolerated, and this could confound transposon insertion sequencing results [179–182].

Both ESSENTIALS and Bayesian approaches categorize genes as essential or non-essential. Whilst useful for inferring function, these models do not reflect the more nuanced ways in which gene disruption can affect bacteria under different conditions. A hidden Markov model (HMM) can be applied to transposon insertion sequencing data to classify genes into one of four categories, indicating how their mutation affects bacterium fitness: essential, defective, neutral or advantaged [183].

HMMs are applicable to datasets where a sequence of observed values (such as read counts at TA sites) is explained by an underlying state sequence (the state call of each gene, which is hidden, in that it is not directly observable). A genome is viewed by an HMM as an alternating sequence of different state call regions. The HMM of DeJesus *et al*. (2013) uses read count information from individual TA sites to infer the probability distribution of state calls over a string of TA sites in a gene. A Viterbi algorithm (a dynamic programming algorithm in which the probability of each TA site state call is based on the state probability distribution from the preceding string of TA sites) then

infers the most likely state call for each TA site, and for a gene. The sequential dependence of state calls afforded by the HMM (defined by the Viterbi algorithm and dependent on previous sites) enables a 'smoothing' of read count data. This allows TA sites with no insertions which are in non-essential genes to be classed as neutral because neighboring sites have insertions. This helps prevent erroneous labelling of TA sites as essential just because they are, purely by chance, absent from the sequencing library. It also helps hedge against other sources of variability such as spikes in the data, which could be generated by PCR bias producing an artificial overrepresentation of reads at one site, or insertions in extreme N and C termini which may be tolerated in essential genes. Only if a consecutive sequence of TA sites with no or low insertions is long enough will the Viterbi algorithm label sites as growth essential or defective. [183]

For the HMM to assign state calls accurately, the probability of read counts in non-essential regions as well as the transition probabilities (the point at which the Viterbi algorithm designates a TA as different in classification from previous sites) must be well defined. A key strength of the HMM described by DeJesus *et al*. (2013) is that it is calibrated such that it generates reasonable and robust state calls across datasets with high or low insertion density and high or low read counts. Further, incorporating the four-state model allows regions with low (but not no) or higher than average read counts to be differentially classified, as defective or advantaged in this case. This allows identification of genes where insertion impaired fitness under test conditions, as well as those that have some 'cost' under test conditions, and therefore enhance organism fitness when mutated [183].

To increase model robustness when applied to different datasets, geometric distributions are used with data smoothing to define likelihood functions for non-neutral states. For an

essential state call, site read count is set very near to 1 (meaning sites with 0 reads are likely to be essential, but that sites with 1 or 2 reads can be tolerated as neutral if immediately preceding sites show higher read counts). For defective calls, read counts must be approximately 100 times lower than the mean for a dataset (which is the neutral mark), whilst for advantaged states, read counts must be 5 times the mean read count or higher for a string of sites [183].

Combining this principle with the Viterbi algorithm generates the Markov chain, the system in which transitions from one state call to another are dependent on probabilistic rules which define the state calls in the previous TA sites. The geometric likelihood functions of each state call at a given TA site, as well as the Markov chain for this model and an example of read count data and state calls are given in Figure 1-10, below.

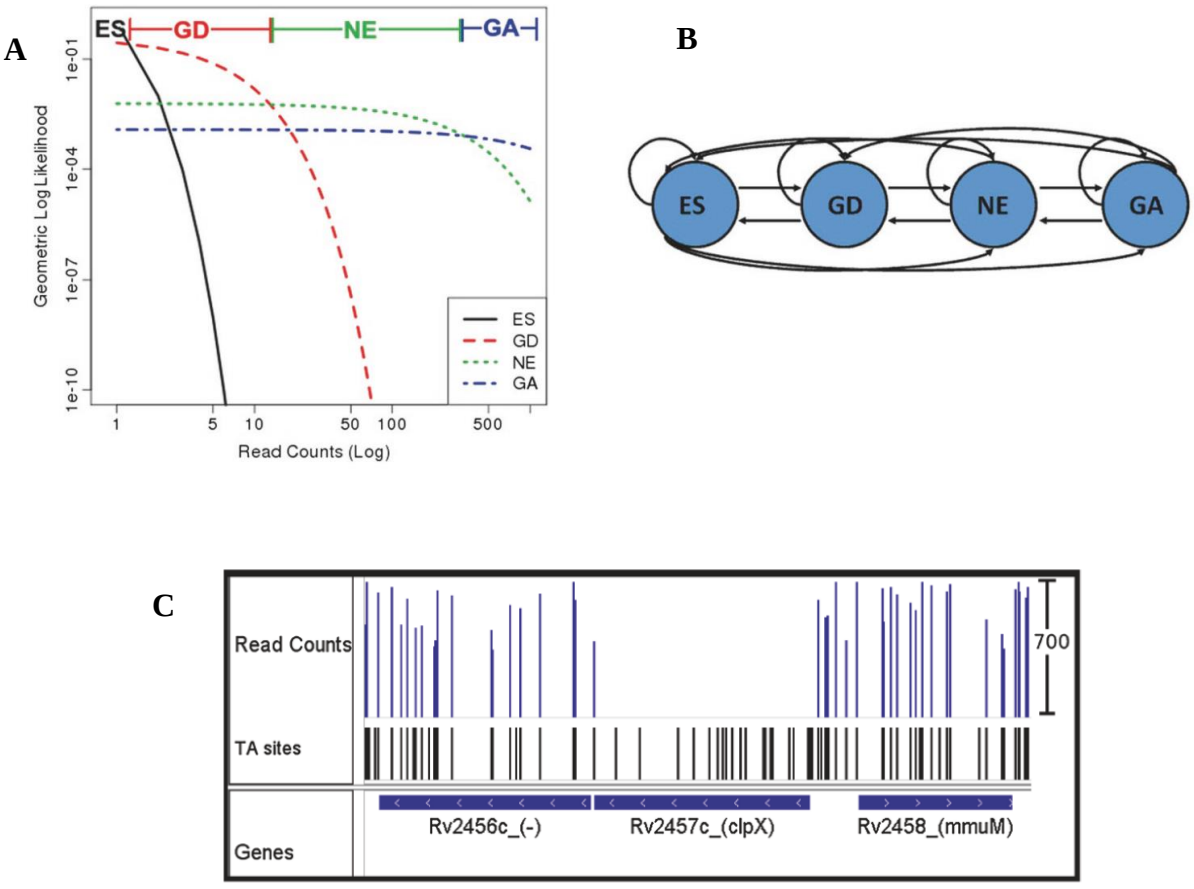


Figure 1-10. Principles and examples of the HMM. **A** – Log plot of the geometric likelihood function for each state call at a given TA site, independent of the Viterbi algorithm. Overlapping densities of the four likelihood functions allow four distinct classifications with one dominating individually at the boundary of each read count parameter. **B** – The Markov chain for the four-state HMM. From left to right, state calls represent read counts of increasing magnitude. The interlinked nature of all nodes means that state transition can occur from one node to any other node in the model. Longer arrows indicate a larger change in read count number that must occur at a TA site for the state call to change from those at preceding sites. **C** – Example read counts for three genes in the *M. tuberculosis* genome obtained from transposon insertion sequencing. Whilst the left and right genes are classified as neutral, the 19 consecutive TA sites in the central gene with zero read counts lead to an essential classification. The 'smoothing' of the data afforded by the Viterbi algorithm of the HMM allows the extreme C terminal insertion mapping read of this gene to be tolerated and maintains the essential classification. Figures reproduced from DeJesus *et al*. (2013), [183].

1.9 Imaging early-stage root-microbe interactions

Prior to attachment, rhizobia must position themselves in proximity to the host plant root [44]. Root exudates can act as chemotactic signals as well as carbon sources for rhizobia [34, 35]. Root exudates are not uniformly exuded along root length. Rather, there seems to be spatial distribution of exudation, which is also dependent on developmental time point [184–187]. Given the importance of root exudates as rhizosphere community influencing chemoattractants, and their differential presence along root length, it is possible that

interesting spatio-temporal interaction dynamics exist between host legume roots and rhizobia, and that characterising these could shed more light on symbiosis stages immediately preceding attachment.

Massalha *et al*. (2017) reported root-microbe interaction imaging of *Arabidopsis thaliana* with *Bacillus subtilis* and/or *E. coli* using a novel microfluidics imaging system [188]. This system used polydimethylsiloxane to cast microfluidics chambers from a master mould featuring channels for plant roots to grow through and inlet/outlet holes for addition of labelled bacteria. By combining the microfluidics device with confocal imaging, the preferential interaction of *B. subtilis* with a specific root zone (termed the elongation zone) of *Arabidopsis* was demonstrated. Further, exclusion of *E. coli* from the root surface by *B. subtilis* was shown, demonstrating that there may be a spatial aspect to pathogen exclusion from roots [188].

Multiple further microfluidics imaging systems for plant roots have been reported (see 6.1), the most enlarged version of which (RMI-chip) was used to image root-microbe interactions between *P. fluorescens* and *Populus tremuloides* (aspen tree) roots [189]. RMI-chip allowed root-microbe interaction imaging over a period of one month and showed that different *B. subtilis* biosensors could demonstrate root exudate composition changes over time [189].

Whilst microfluidics imaging platforms have demonstrated potential in characterising root-microbe interactions, there have been no reports of such imaging in *Rhizobium*-legume symbioses. An appropriate imaging system applied in this context could examine root interaction dynamics as well as strain competition for root colonization and interaction between PGPR and pathogens. The final experimental chapter of this work aimed to develop an imaging system that could be applied to these questions.

1.10 Research objectives

To date the process of primary root attachment, important both for nitrogen fixing symbioses and bulk root interaction, remains one of the least well characterised of all the symbiosis stages [25]. In this study, I aimed to better characterise primary attachment mechanisms of rhizobia under different environmental conditions using the model organism Rlv3841.

Proteomics work (chapter 3) was used to investigate rhicadhesin, a factor hypothesized to be important for root hair attachment to legumes at alkaline pHs [34]. To better characterise root attachment, a whole-root luminescence-based (Lux) one-hour attachment assay was developed with Rlv3841 and pea and validated at different pHs and with mutant rhizobial strains of known attachment phenotype. A bioinformatic screen was used to identify putative novel primary attachment factors, and mutants in these were tested in Lux attachment assays.

Chapter 4 expanded on the findings of chapter 3 by taking a high-throughput and genome-wide approach to characterising primary attachment determinants using an INSeq screen at different pHs. Appropriate experimental parameters were determined and INSeq results compared with published literature and Lux attachment assay results. The screen identified the attachment factors required for Rlv3841 interacting with pea roots at three pHs and was also used to further investigate rhicadhesin and the regulatory requirements for attachment.

Whilst attachment to cognate symbiont partners is important for nitrogen fixation, rhizobia can interact with many different plant roots in the soil [36–38], and evidence from the literature suggests that mechanisms may be, at least in part, plant specific [34, 96]. Therefore, chapter 5 applied the principles of an INSeq attachment assay to Rlv3841 with

a non-host legume (soybean) and a non-legume (barley). This screen comprehensively identified the genes that are universally needed for attachment to different plant roots and shed new light on regulatory mechanisms underpinning Rlv3841 transition from a free-living to root-attached state.

Given the recent advances in imaging technologies and largely uncharacterised spatio-temporal interaction dynamics of rhizobia with legume roots (see 6.1), chapter 6 aimed to develop a real-time imaging system that could examine these processes. Growth conditions for highly motile bacterial cultures were established and spatio-temporal dynamics of reporter gene activation were demonstrated in an existing [188] and newly developed imaging platform. The role of motility in early interaction dynamics and root hair attachment polarity under different conditions were examined with the new system, and all these results were used to inform the design for an updated imaging system. This system will be applicable for future plant-PGPR-pathogen studies and examining root preference between different biovars of rhizobia.

Overall, this data has provided a comprehensive characterisation of primary attachment determinants in Rlv3841 under different environmental conditions and in interaction with different host plants.

Chapter 2

Materials and Methods

2.1 Bacterial strains, plasmids and primers

Bacterial strains used are listed in Table 2-1. All strains were stocked in 15% glycerol and flash frozen in liquid nitrogen before being stored at -80 ^{o}C. Plasmids used are listed in Table 2-2 and all primers used are listed in Table 2-3.

Table 2-1. *Rhizobium* and *E. coli* strains used in this work

Strain	Description	Reference
	Rhizobium leguminosarum	
Rlv3841	*Rhizobium leguminosarum* biovar *viciae* 3841 (Rlv3841); Str^r derivative of strain 300. Nodulates *Viciae* family legumes such as pea (*Pisum sativum*) and Vetch (*Vicia cracca*)	[190]
A1045	Mutant of Rlv3841, *gmsA*::Tn*5*ΩKan/Neo. Str^r, Kan/Neo^r	[56]
A1480	*Rhizobium leguminosarum* biovar *viciae* 300 *rapA2*ΩSpec *rapC::*Tn*5*ΩApra. $Spec^r$, $Apra^r$	[59]
A963	Rlv3841 *praR::*Tn*5*ΩKan/Neo. Str^r, Kan/Neo^r	[191]
D5250	Rlv3841[pIJ11282]	[59]
LMB310	Rlv3841 *pssA*::TnΩSpec. Str^r, $Spec^r$	[155]
LMB349	pLMB211 integration into Rlv3841 disrupting *RL3273*. Str^r, Kan/Neo^r	[192]
LMB487	Rlv3841[pLMB579] (375bp promoter of *lppE* - *RL3234*) cloned into pIJ11268. Str^r, Tet^r	[193]

OPS0111	A963[pIJ11282]. Str^r, Kan/Neo^r, Tet^r	Dr Vinoy Ramachandran, postdoctoral researcher, Poole Group, unpublished
OPS0167	Rlv3841[pOPS0065] (lppE GFP promoter fusion, pTac mCherry). Str^r, Gent^r	Dr Vinoy Ramachandran, postdoctoral researcher, Poole Group, unpublished
OPS0296	A1480[pIJ11282]. Spec^r, Apra^r, Tet^r	Dr Vinoy Ramachandran, postdoctoral researcher, Poole Group, unpublished
OPS0804	LMB310[pIJ11282]. Str^r, Spec^r, Tet^r	Dr Carmen Sanchez-Canizares, postdoctoral researcher, Poole group, unpublished
OPS0914	A1045[pIJ11282]. Str^r, Kan/Neo^r, Tet^r	This work
OPS1131	OPS2051[pIJ11282]. Str^r, Kan/Neo^r, Tet^r	This work
OPS1235	OPS1783[pIJ11282]. Str^r, Kan/Neo^r, Tet^r	This work
OPS1236	OPS1782[pIJ11282]. Str^r, Kan/Neo^r, Tet^r	This work
OPS1237	LMB349[pIJ11282]. Str^r, Kan/Neo^r, Tet^r	This work
OPS1238	OPS2054[pIJ11282]. Str^r, Kan/Neo^r, Tet^r	This work

OPS1239	A1480::A963 phage transduction; *Rhizobium leguminosarum* biovar *viciae* 300::*rapA2*ΩSpec *rapC::*Tn*5*ΩApra *praR::*Tn*5*ΩKan/Neo. Spec^r, Apra^r, Kan/Neo^r	This work
OPS1266	OPS1239[pIJ11282]. Spec^r, Apra^r, Kan/Neo^r, Tet^r	This work
OPS1290	OPS2052[pIJ11282]. Str^r, Kan/Neo^r, Tet^r	This work
OPS1291	OPS2053[pIJ11282]. Str^r, Kan/Neo^r, Tet^r	This work
OPS1709	Rlv3831 *RL0703*::pK19mob (pOPS1095), Str^r, Kan/Neo^r. *RL0703*: chemotaxis motility protein MotA	This work
OPS1710	Rlv3831 *RL0728*::pK19mob (pOPS1096), Str^r, Kan/Neo^r. *RL0728*: flagellar hook protein FlgE	This work
OPS1730	A1045[pLMB449]. Str^r, Kan/Neo^r, Gent^r	This work
OPS1734	Rlv3841[pLMB449]. Str^r, Gent^r	This work
OPS1736	OPS170[pLMB449]. Str^r, Kan/Neo^r, Gent^r	This work
OPS1738	OPS1710[pIJ11282]. Str^r, Kan/Neo^r, Tet^r	This work
OPS1739	OPS1709[pIJ11282]. Str^r, Kan/Neo^r, Tet^r	This work

OPS1782	Rlv3831 *RL4382*::pK19mob (pOPS0489), Str^r, Kan/Neo^r. *RL4382*: putative filamentous hemagglutinin adherence factor	This work
OPS1783	Rlv3831 *RL2969*::pK19mob (pOPS0487), Str^r, Kan/Neo^r. *RL2969*: putative transmembrane protein	This work
OPS1878	RU4062[pIJ11282]. Str^r, Kan/Neo^r, Tet^r	This work
OPS1907	Rlv3831 *RL3453*::pK19mob (pOPS1294), Str^r, Kan/Neo^r. *RL3453*: putative two-component sensor histidine kinase transcriptional regulatory protein	This work
OPS1908	Rlv3831 *RL4145*::pK19mob (pOPS1295), Str^r, Kan/Neo^r. *RL4145*: putative LacI family HTH-type transcriptional repressor	This work
OPS1909	Rlv3831 *pRL100406:*:pK19mob (pOPS1296), Str^r, Kan/Neo^r. *pRLl100406*: transcriptional regulatory protein MctR	This work
OPS1967	OPS1907[pIJ11282]. Str^r, Kan/Neo^r, Tet^r	This work
OPS1968	OPS1908[pIJ11282]. Str^r, Kan/Neo^r, Tet^r	This work
OPS1969	OPS1909[pIJ11282]. Str^r, Kan/Neo^r, Tet^r	This work
OPS2051	Rlv3831 *pRL110543*::pK19mob (pOPS0480), Str^r, Kan/Neo^r. *pRL110543*: conserved hypothetical protein	This work

OPS2052	Rlv3831 *pRL110071*::pK19mob (pOPS0479), Str^r, Kan/Neo^r. *pRL110071*: conserved hypothetical protein	This work
OPS2053	Rlv3831 *pRL100053*::pK19mob (pOPS0478), Str^r, Kan/Neo^r. *pRL100053*: putative transmembrane protein	This work
OPS2054	Rlv3831 *RL0109*::pK19mob (pOPS0483), Str^r, Kan/Neo^r. *RL0109*: conserved hypothetical protein	This work
RU4062	*pRL100162* (*nifH*) mutant of Rlv3841 (pRU2056 integrated). Str^r, Kan/Neo^r	[139]
	E. 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
pSAM_Rl	SM10λ*pir* carrying pSAM_Rl *Mariner* transposon vector; Kan/Neo^r, Amp^r.	[151]

Str^r: Streptomycin resistance, Kan/Neo^r: Kanamycin and Neomycin resistance (conferred by a neomycin phosphotransferase II gene, *nptII*), Tet^r: Tetracycline resistance, Gent^r: Gentamycin resistance, Apra^r: Apramycin resistance, Spec^r: Spectinomycin resistance, Amp^r: Ampicillin resistance. Plasmid numbers are included for pK19mob strains isolated in Rlv3841 (see Table 2-2).

Table 2-2. Plasmids used in this work

Plasmid	Description	Reference
pIJ11268	Vector with promoterless reporter luciferase genes. Tet^r	[59]
pIJ11282	pIJ11268 with luciferase genes under control of the *nptII* promoter. Tet^r	[59]
pK19mob	Mobilizable vector used for integration mutagenesis in *Rhizobium*. Kan/Neo^r	[194]
pLMB449	Reporter vector containing a *gfp* gene under control of pTac. $Gent^r$	Karunakaran Ramakrishnan, postdoctoral researcher, Poole group, unpublished
pLMB579	375bp promoter of *lppE* (*RL3234*) cloned into pIJ11268. Tet^r	[193]
pOPS0065	Reporter vector containing a *gfp* gene under control of the 375bp promoter of *lppE* (*RL3234*) with an *mCherry* gene under control of pTac. $Gent^r$	Karunakaran Ramakrishnan / Alison East, postdoctoral researchers, Poole group, unpublished
pOPS0478	opx1751/oxp1752 PCR product of Rlv3841 *pRL100053* BD ligated into *Hin*dIII digested pK19mob. Kan/Neo^r	This work
pOPS0479	oxp1753/oxp1754 PCR product of Rlv3841 *pRL110071* BD ligated into *Hin*dIII digested pK19mob. Kan/Neo^r	This work

pOPS0480	oxp1755/oxp1756 PCR product of Rlv3841 *pRL110543* BD ligated into *Hin*dIII digested pK19mob. Kan/Neo^r	This work
pOPS0483	oxp1785/oxp1786 PCR product of Rlv3841 *Rl0109* BD ligated into *Hin*dIII digested pK19mob. Kan/Neo^r	This work
pOPS0487	oxp1767/oxp1768 PCR product of Rlv3841 *Rl2969* BD ligated into *Hin*dIII digested pK19mob. Kan/Neo^r	This work
pOPS0489	oxp1777/oxp1778 PCR product of Rlv3841 *Rl4382* BD ligated into *Hin*dIII digested pK19mob. Kan/Neo^r	This work
pOPS1095	oxp2817/oxp2818 PCR product of Rlv3841 *Rl0703* BD ligated into *Hin*dIII digested pK19mob. Kan/Neo^r	This work
pOPS1096	oxp2819/oxp2820 PCR product of Rlv3841 *Rl0728* BD ligated into *Hin*dIII digested pK19mob. Kan/Neo^r	This work
pOPS1294	oxp3235/oxp3236 PCR product of Rlv3841 *Rl3453* BD ligated into *Hin*dIII digested pK19mob. Kan/Neo^r	This work
pOPS1295	oxp3231/oxp3232 PCR product of Rlv3841 *Rl4145* BD ligated into *Hin*dIII digested pK19mob. Kan/Neo^r	This work
pOPS1296	oxp3239/oxp3240 PCR product of Rlv3841 *pRL100406* BD ligated into *Hin*dIII digested pK19mob. Kan/Neo^r	This work

pRK2013	Helper plasmid for tripartental conjugation. Kan/Neo^r	[195]
pSAM_Rl	*Mariner* transposon vector. Amp^r, Kan/Neo^r	[151]

Kan/Neo^r: Kanamycin and Neomycin resistance (conferred by a neomycin phosphotransferase II gene, *nptII*), Tet^r: Tetracycline resistance, Gent^r: Gentamycin resistance, Amp^r: Ampicillin resistance. Primer numbers are included for PCR amplified gene products ligated into pK19mob (see Table 2-3).

Table 2-3. Primers used in this work

Primer Name	Sequence 5' → 3'	Description
hah-2	AAACGGGAAAGGTTCCGTCCA	Universal mapping primer for Tn5 insertions
INSeq_Adpt_Top	AGATCGGAAGAGCGTCGTGTAGGGAA	INSeq Adapter (Top) [169]
INSeq_Adpt_Bottom	TTCCCTACACGACGCTCTTCCGATCTNN	INSeq Adapter (Bottom) [169]
Ion Torrent BioSAM	/BiotinTEG/CGGTTCGCTTGCTGTCCATAAAACC	Ion Torrent BioSAM with 5' Biotin TEG [151]
IT_A_FP_1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAAGGTAACGATATAAAACCGCCCAGTCTACTCGAGGG	Ion express forwards barcode primer 1 [151]. Barcode sequence is underlined
IT_A_FP_2	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGGAGAACGATATAAAACCGCCCAGTCTACTCGAGGG	Ion express forwards barcode primer 2 [151]
IT_A_FP_3	CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGAGGATTCGATATAAAACCGCCCAGTCTACTCGAGGG	Ion express forwards barcode primer 3 [151]
IT_A_FP_4	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACCAAGATCGATATAAAACCGCCCAGTCTACTCGAGGG	Ion express forwards barcode primer 4
IT_A_FP_5	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGAAGGAACGATATAAAACCGCCCAGTCTACTCGAGGG	Ion express forwards barcode primer 5

IT_A_FP_6	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGCAAGTTCGATATAAAACCGCCCAGTCTACTCGAGGG	Ion express forwards barcode primer 6
IT_A_FP_7	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGTGATTCGATATAAAACCGCCCAGTCTACTCGAGGG	Ion express forwards barcode primer 7
IT_A_FP_8	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCGATAACGATATAAAACCGCCCAGTCTACTCGAGGG	Ion express forwards barcode primer 8
IT_A_FP_9	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGAGCGGAACGATATAAAACCGCCCAGTCTACTCGAGGG	Ion express forwards barcode primer 9
IT_A_FP_10	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACCGAACGATATAAAACCGCCCAGTCTACTCGAGGG	Ion express forwards barcode primer 10
IT_A_FP_11	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTCGAATCGATATAAAACCGCCCAGTCTACTCGAGGG	Ion express forwards barcode primer 11
IT_A_FP_12	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGGTGGTTCGATATAAAACCGCCCAGTCTACTCGAGGG	Ion express forwards barcode primer 12
IT_trP1_FP	CCTCTCTATGGGCAGTCGGTGATTTCCCTACACGACGCTCTTCCGATCT	Universal reverse barcode primer [151]
M12 Top	CTGTCCGTTCCGACTACCCTCCCGAC	M12 adapter [169]

M12 Bottom	GTCGGGAGGGTAGTCGGAACGGACAG	M12 adapter [169]
M13 uni (-21)	TGTAAAACGACGGCCAGT	Sequencing/mapping primer for pK19mob containing vectors
M13 rev (-29)	CAGGAAACAGCTATGACC	Sequencing/mapping primer for pK19mob containing vectors
oxp1751	TGATTACGCCAAGCTATGGTTGCCATCAAGC	Amplification primer (forward) for *pRL100053*::pK19mob
oxp1752	GCAGGCATGCAAGCTTCTTTGAAGCGATCACGGGC	Amplification primer (reverse) for *pRL100053*::pK19mob
oxp1753	TGATTACGCCAAGCTTGCGGGAAGGGGCGT	Amplification primer (forward) for *pRL110071*::pK19mob
oxp1754	GCAGGCATGCAAGCTCGAAATTCGCTGCGGAAAAC	Amplification primer (reverse) for *pRL110071*::pK19mob
oxp1755	TGATTACGCCAAGCTCTGCCGAAGCCAAGG	Amplification primer (forward) for *pRL110543*::pK19mob
oxp1756	GCAGGCATGCAAGCTCGTTCGGCCTTTCGGG	Amplification primer (reverse) for *pRLl110543*::pK19mob

oxp1767	TGATTACGCCAAGCTGTCCAGTCGGCGCCG	Amplification primer (forward) for *RL2969*::pK19mob
oxp1768	GCAGGCATGCAAGCTAGCGGCTGTAGCAATAGTCG	Amplification primer (reverse) for *RL2969*::pK19mob
oxp1777	TGATTACGCCAAGCTGATCTCCGGAACCATTGCC	Amplification primer (forward) for *RL4382*::pK19mob
oxp1778	GCAGGCATGCAAGCTGGTGCGGCGGGTGTG	Amplification primer (reverse) for *RL4382*::pK19mob
oxp1785	TGATTACGCCAAGCTACCGGCAAGAAGTTTTATGATCT	Amplification primer (forward) for *RL0109*::pK19mob
oxp1786	GCAGGCATGCAAGCTATGATGTCGCTCATGTCGTCATC	Amplification primer (reverse) for *RL0109*::pK19mob
oxp2052	TATCTCCCCCGCCGCGTTAT	*RL2969*::pK19mob mapping primer (forward)
oxp2053	GCGAGGCTTGCTCCGAT	*RL2969*::pK19mob mapping primer (reverse)
oxp2058	TTCGCTCGATTTTACCAAGC	*RL4382*::pK19mob mapping primer (forward)

oxp2059	AGATTGCGGACCGACGT	*RL4382*::pK19mob mapping primer (reverse)
oxp2062	GAAGGAGATGGACACTGCAC	*pRL100053*::pK19mob mapping primer (forward)
oxp2063	CACCAGGGATGAAAGCTTGA	*pRL100053*::pK19mob mapping primer (reverse)
oxp2064	AGTTCGATTGACAGGCTCTC	*pRL110071*::pK19mob mapping primer (forward)
oxp2065	GTGGAATTCTGCTGCTTCG	*pRL110071*::pK19mob mapping primer (reverse)
oxp2066	CGAAGTCAAACAGTCAGGAA	*pRL110543*::pK19mob mapping primer (forward)
oxp2067	CGCTGCCTGTCCTGAC	*pRL110543*::pK19mob mapping primer (reverse)
oxp2068	GGCTATTCACACGAGGCTC	*RL0109*::pK19mob mapping primer (forward)
oxp2069	TTTTTCCTCGCCGGGC	*RL0109*::pK19mob mapping primer (reverse)

oxp2423	CGGATATAGGGCTCGACGAC	*praR* (*RL0390*)::Tn5 mapping primer (forward)
oxp2424	GATCTTCGAGACGCATCTGA	*praR* (*RL0390*)::Tn5 mapping primer (reverse)
oxp2817	TGATTACGCCAAGCTTGCTTGTTGCGCACGAC	Amplification primer (forward) for *RL0703*::pK19mob
oxp2818	GCAGGCATGCAAGCTATCGACAATCCGGCCGAATC	Amplification primer (reverse) for *RL0703*::pK19mob
oxp2819	TGATTACGCCAAGCTCAATCAGGAAGGGCTGACCG	Amplification primer (forward) for *RL0728*::pK19mob
oxp2820	GCAGGCATGCAAGCTTCCATGTCGATCATCTGGCC	Amplification primer (reverse) for *RL0728*::pK19mob
oxp2821	GCCGACCATCAGCTTCGATA	*RL0703*::pK19mob mapping primer (forward)
oxp2822	GCGAAATCGAGCATGTTGCC	*RL0703*::pK19mob mapping primer (reverse)
oxp2823	GCGATCACCAACTACACCGA	*RL0728*::pK19mob mapping primer (forward)

oxp2824	GGCATTGTTCAGCGAGTTGG	*RL0728*::pK19mob mapping primer (reverse)
oxp3231	TGATTACGCCAAGCTATTGACATAGCCCTGGTAGC	Amplification primer (forward) for *RL4145*::pK19mob
oxp3232	GCAGGCATGCAAGCTCGGGAGAAGATCAAGGAACA	Amplification primer (reverse) for *RL4145*::pK19mob
oxp3233	CTCATCGTTCAGCACATCAG	*RL4145*::pK19mob mapping primer (forward)
oxp3234	CCTGAGCAATGTCATATCGC	*RL4145*::pK19mob mapping primer (reverse)
oxp3235	TGATTACGCCAAGCTGAAATCGGATTCCAGCTTCG	Amplification primer (forward) for *RL3453*::pK19mob
oxp3236	GCAGGCATGCAAGCTTGCCTCGAACACACATATCT	Amplification primer (reverse) for *RL3453*::pK19mob
oxp3237	GAAAGAAGGTACCCAGACGA	*RL3453*::pK19mob mapping primer (forward)
oxp3238	CAGATATTTCCGACGATGCG	*RL3453*::pK19mob mapping primer (reverse)

oxp3239	TGATTACGCCAAGCTATGCCGAGCTTCTTTTTGAG	Amplification primer (forward) for *pRL100406*::pK19mob
oxp3240	GCAGGCATGCAAGCTGACCATATCGAAGTTGCCG	Amplification primer (reverse) for *pRL100406*::pK19mob
oxp3241	CCAGACCCCATTCATCGATA	*pRL100406*::pK19mob mapping primer (forward)
oxp3242	CATCAAGGAAATCAGACGGC	*pRL100406*::pK19mob mapping primer (reverse)

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

2.2 Media and Antibiotics

2.2.1 Media

E. coli strains were grown at 37°C in Luria Bertani broth (LB) (10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 5 g L^{-1} NaCl). Cultures were shaken at 200 rpm unless otherwise stated. For growth on solid media, agar was added 1.4 % w/v prior to autoclaving.

R. leguminosarum strains were grown on tryptone yeast (TY) media at 28°C (5 g L^{-1} tryptone, 3 g L^{-1} yeast extract, 6 mM $CaCl_2$). For growth on solid media, agar was added 1.75 % w/v before autoclaving.

Where a minimal growth media was required, *R. leguminosarum* strains were grown in Universal Minimal Salts (UMS) at 28°C. UMS was prepared as follows: 0.5 g L^{-1} $MgSO_4 \bullet 7H_2O$, 0.2 g L^{-1} NaCl, 4.19 g L^{-1} MOPS, 0.5 mM K_2HPO_4, and 1 mL of trace elements (0.375 g L^{-1} EDTA-Na_2, 016 g L^{-1} $ZnSO_4 \bullet 7H_2O$, 0.2 g L^{-1} $NaMoO_4$, 0.25 g L^{-1} H_3BO_3, 0.2 g L^{-1} $MnSO_4 \bullet 4H_2O$, 0.02 g L^{-1} $CuSO_4 \bullet 5H_2O$, 1 g L^{-1} $CoCl_2 \bullet 6H_2O$). pH was adjusted to 7.0 before autoclaving. Media was supplemented after autoclaving with 1 mL of calcium stock solution (75 g L^{-1} $CaCl_2 \bullet 2H_2O$), 1 mL of iron stock solution (12 g L-1 $FeSO_4 \bullet 7H_2O$ dissolved in 1M HCl), and 1 mL of vitamin stock solution (1 g L^{-1} Thiamine hydrochloride, 2 g L^{-1} D-Pantothenic acid calcium salt, 100 mg L^{-1} Biotin). UMS was then supplemented with 30 mM pyruvate (or 3 mM pyruvate where low-carbon availability conditions were required) and 10 mM ammonium chloride. Cultures were shaken at 200

rpm unless otherwise stated. For growth on solid media, agar was added 1.75 % w/v before autoclaving. Where a different media pH was required, pH was adjusted before use.

For experiments examining interactions of Rlv3841 strains with plant roots, liquid FP media was used for strain resuspension. FP media was prepared per litre as follows: 2.5 mL 0.27 M $CaCl_2 \bullet 2H_2O$, 3 mL 0.16 M $MgSO_4 \bullet 7H_2O$, 3.33 mL 0.22 M KH_2PO4, 3.33 mL 0.32 M Na_2HPO_4, 2 mL 0.01 M $FeC_6H_5O_7$ and 1 mL Gibson's trace (2.86 g H_3BO_3, 2.03g $MnSO_4 \bullet 4H_2O$, 220 mg $ZnSO_4 \bullet 7H_2O$, 80 mg $CuSO_4 \bullet 5H_2O$ and 80 mg H_2MoO_4). pH was adjusted to 7.0 before autoclaving. For plant growth on solid media, agar was added 1.0 % w/v before autoclaving.

2.2.2 Antibiotics

For a selection of bacterial strains, antibiotics were added to the media at the concentrations shown in Table 2-4.

Table 2-4. Concentrations (in µg mL^{-1}) of antibiotics used in liquid and solid media.

Antibiotic	**Antibiotic suspended in**	***Rhizobium leguminosarum***	***E. coli***
Ampicillin (Amp)	Water	-	100
Apramycin (Apra)	Water	50	-

Gentamycin (Gent)	HEPES	20	10
Kanamycin (Kan)	Water	50	20
Neomycin (Neo)	HEPES	80/240[a]	20
Spectinomycin (Spec)	HEPES	100	50
Streptomycin (Str)	HEPES	500	25
Tetracycline (Tet)	Water	5	10

a: higher concentration of neomycin used for selection of interposon mutants.

2.3 DNA techniques

2.3.1 Isolation of genomic DNA

Genomic DNA was extracted from Rlv3841 strains using a Qiagen DNeasy Blood and Tissue kit according to the manufacturer's instructions. 1 x 10^9 cells were used per isolation. Genomic DNA was diluted to 1/100 with sterile water for use in PCR amplifications. Plasmid DNA was extracted using a Thermo Fisher GeneJET Plasmid Miniprep kit according to manufacturer's instructions.

2.3.2 PCR amplification

Primers for PCR amplification (Table 2-3) were designed using Geneious R10 [196] and synthesized by Eurofins MWG Operon. Phusion ® High-fidelity PCR Master Mix (Thermo Fisher) was used for amplification of DNA products for downstream cloning reactions. GoTaq ® Green Master Mix (Promega) was used for mapping reactions.

Primer annealing temperatures were determined by Geneious R10. A Verti® thermocycler (Applied Biosystems) was used for all reactions with conditions determined by manufacturer's guidelines for polymerases. PCR products were purified using a Thermo Fisher GeneJET PCR purification kit according to manufacturer's instructions.

Colony PCR was used for mapping reactions with *E. coli* strains; a single colony was transferred into a PCR reaction using a sterile pipette tip. For mapping reactions with Rlv3841 strains, alkaline polyethylene glycol (PEG) was used for DNA extraction [197]. A single colony was resuspended in 15 μL alkaline PEG and heated at 60^{o}C for 10 minutes (min). 3 μL of this extract was used per 10 μL total PCR reaction volume.

2.3.3 Gel electrophoresis

Gel electrophoresis was used to separate PCR products and restriction digest fragments in 0.9 % agarose (Sigma Aldrich) with added Sybr ® Safe DNA dye (Invitrogen, according to manufacturer's instructions), in TAE buffer (400 mM TRIS acetate, 1 mM EDTA). Gels were run at 120 mV for 20-30 min. Gels were visualized using a GelDoc EZ system (BioRad).

2.3.4 Restriction digests and DNA ligation

Restriction digests were carried out using NEB, Roche or Thermo Fisher restriction endonucleases and their respective buffers according to manufacturer's instructions. DNA ligation was performed with T4 DNA ligase and 10X T4 DNA ligase buffer (Thermo Fisher) according to manufacturer's instructions.

2.4 Cloning techniques

2.4.1 Transformation

E. coli DH5α cells were thawed on ice for 15 min and transformation DNA added in a 1:10 v/v ratio. Cells were incubated on ice for 30 min then heat-shocked at 42 ^{o}C for 45 seconds (sec) then incubated on ice for 2 min. SOC medium was added at 9-fold the original transformation volume and cells shaken at 37 ^{o}C for 1 hour (hr) at 200 rpm before plating on LB agar with appropriate antibiotics. Colonies were grown overnight at 37^{o} C before PCR screening and sequencing to confirm the expected construct.

2.4.2 Conjugation to transfer a plasmid from *E. coli* to *R. leguminosarum*

Triparental mating was used for conjugation of chromosomal integration plasmids from *E. coli* into Rlv3841 using helper plasmid pRK2013. Three days (d) before conjugation the recipient Rlv3841 strain was grown on a TY agar slope containing relevant antibiotics. One day before conjugation the conjugative donor and pRK2013 containing *E. coli* strains were inoculated in 10 mL LB with relevant antibiotics and grown overnight at 37^{o}C with shaking at 200 rpm. On the day of conjugation, overnight *E. coli* cultures were sub-

cultured (1:10 v/v) into 5 mL fresh LB with antibiotics and grown with shaking until reaching OD_{600} 0.4-0.6. *E. coli* strains were pelleted at 1,000 x g for 5 min and washed with TY 3x to remove traces of antibiotic. The Rlv3841 slope was resuspended in 3-5 mL TY and 400 µL recipient Rlv3841 mixed with 400 µL donor *E. coli* and 200 µL pRK2013. This mixture was pelleted at 2,500 x g for 5 min and the pellet resuspended in 30 µL TY and pipetted onto a sterile nitrocellulose filter on a TY plate and incubated at 28^{o} C overnight. Bacteria were then streaked from the filter onto a TY plate containing the appropriate transformant selection antibiotics and incubated for 3-4 d at 28^{o}C.

2.4.3 Mutagenesis by pK19mob integration

pK19mob (Table 2-2) was used to generate integration mutants. Purified pK19mob vector was digested with *Hin*dIII. An internal fragment (~500 bp) of the target gene was amplified with PCR primers including 15 bp of homology to the digested pK19mob vector at the 3'- end of each primer. The internal gene fragment was cloned into the digested pK19mob vector using InFusion ® HD cloning according to the manufacturer's instructions. The resulting vector was transformed into chemically competent DH5α cells and plated on LB kanamycin. Restriction digest and sequencing with M13 uni (-21) and M13 rev (-29) primers was used to confirm correct vector ligation. The correct pK19mob construct was conjugated into Rlv3841 and correct conjugants selected on TY str and neo. Genomic DNA was isolated from conjugants and used for mapping and sequencing with relevant mapping primers (Table 2-3) and M13 uni (-21) and M13 rev (-29) to confirm correct insertion.

2.3.5 Transduction of *R. leguminosarum*

Phage transduction was used to generate a triple mutant strain (OPS1239) from two parent mutant strains: A1480 (*R. leguminosarum* biovar *viciae*
300 *rapA2::*Tn*5*ΩSpec *rapC::*Tn*5*ΩApra) and A963 (Rlv3841 *praR::*Tn*5*ΩKan/Neo). The donor strain (A963) was grown for 3 d on a TY slope at 28ºC before being resuspended in 3 mL sterile distilled water. 100 µL A963 resuspension was added to 900 µL serial dilutions (10^{-2} – 10^{-6}) of RL38 phage stock [198]. 4 mL of half-agar TY was added to each dilution and these were plated on TY for 2-3 d incubation at 28ºC. The plate with a bacterial lawn nearing confluence was taken and 10 mL sterile distilled water added. After rocking for 2 hr the supernatant (now the phage lysate) was taken and stored at 4ºC after addition of 3 drops of chloroform.

A1480 (the recipient strain) was grown for 3 d on a TY slope at 28ºC and resuspended in 3 mL TY. 200 µL A1480 resuspension was mixed with a serial dilution (1-0.001 µL) of phage lysate and incubated for 1 hr at 28ºC. 100 µL of each dilution was plated on TY spec, apra and neo plates before incubation at 28ºC for 4 days. Single colony transductants were carefully removed and purified on selective plates before genomic DNA was isolated and used for mapping and sequencing with relevant primers (hah-2, targeting 9 base pair repeats at the Tn5 cassette termini and oxp2363 and 2364, see Table 2-3).

2.5 Proteomics with mass spectrometry

2.5.1 Crude adhesin isolation

A crude adhesin protein fraction was isolated from Rlv3841 cells as described by Smit *et al.* (1989) [53] for use in root section attachment assays. Briefly, Rlv3841 was grown in

TY media and cells harvested by centrifugation at OD_{600} 0.7. The cell pellet was washed with and resuspended in 25 mM phosphate buffer. Cells were sheared for 5 min with a sonic dismembrator 705 (Fisher Scientific) at an amplitude of 5. The suspension was centrifuged at 12,000 x g for 10 min at 4°C and the supernatant then centrifuged at 100,000 x g (Beckmann TL-100 ultracentrifuge) for 2 hr at 4°C. The supernatant obtained from this step was termed the crude adhesin fraction. SDS-PAGE gels were used for protein visualization by molecular weight according to the protocol of Smith (1994) [199]. Gels were run at 180 mV for 45 min and stained with SYPRO® Ruby (Thermo Fisher) according to manufacturer's instructions. Gels were imaged using a GelDoc EZ system (BioRad). Crude adhesin protein concentration was measured using a Pierce® BCA protein assay kit (Thermo Fisher) according to manufacturer's instructions.

2.5.2 LC-MS/MS

A 14 kDa band from the crude adhesin protein fraction was sent for liquid chromatography – mass spectrometry (LC-MS) at the Analytics Core Facility at the University of Duisburg-Essen. Experiments were performed on an Orbitrap Elite instrument (Thermo Fisher) that was coupled to an EASY-nLC 1000 liquid chromatography (LC) system (Thermo Fisher) as described in [200] with the following adjustments: MS/MS spectra data were searched against the UniProt reference database UP000006575_216596.fasta (Rlv3841, 7091 entries). The MS/MS target protein range was set to 12-17 kDa. For data analysis, only protein groups with at least three identified unique peptides over all runs were considered for further analysis. Non-Rlv3841 proteins and hits to the decoy database were removed.

2.6 Root attachment assays

2.6.1 Buffering capacity of vermiculite

The buffering capacity of fine vermiculite (Sinclair Pro, 1-3 mm, used throughout this thesis) was tested using nitrogen-free rooting solution adjusted to the desired pH (6.5, 7.0 or 7.5) using hydrochloric acid or sodium hydroxide solution. The composition of rooting solution used for all plant experiments is shown in Table 2-5.

Table 2-5. Composition of nitrogen-free rooting solution used for plant experiments

Chemical	Final concentration
$CaCl_2 \bullet 2H_20$	1 mM
KCl	100 μM
$MgSO_4 \bullet 7H_2O$	800 μM
Fe EDTA	10 μM
H_3BO_3	35 μM
$MnCl_2 \bullet 4H_2O$	9 μM
$ZnCl_2$	0.8 μM
$Na_2MoO_4 \bullet 2H_2O$	0.5 μM
$CuSO_4 \bullet 5H_2O$	0.3 μM
KH_2PO_4	1 g/L
Na_2HPO_4	1.14 g/L

Salts (KH_2PO_4 and Na_2HPO_4) were dissolved in deionised water before the remaining chemical solutions were added.

25 mL rooting solution at the desired starting pH was added to 10 g fine vermiculite in triplicate in 50 mL Falcon tubes. Tubes were rotated at 20 rpm and the pH of rooting solution measured at defined intervals from 1 min to 72 hr using a pH meter (Hanna).

2.6.2 Growth of Rlv3841 strains for Lux and insertion sequencing attachment assays

Rlv3841 strains were grown on UMA slopes adjusted to the desired pH for Lux root attachment assays. For insertion sequencing assays, the Rlv3841 *mariner* transposon mutant library (see 2.7.1) was grown on UMA slopes at pH 7.0.

2.6.3 Root section attachment assays

Pea seeds (*Pisum sativum* variety Avola, used throughout this thesis) were sterilized by immersion in 95 % ethanol in a sterile flask for 30 sec before rinsing with sterile distilled water and immersing in 2 % sodium hypochlorite solution for 5 min. Seeds were rinsed 10x in sterile distilled water, before transferring seeds onto 1 % distilled water agar (DWA) plates and incubating in the dark at room temperature for 3 d. Root section attachment assays were carried out as described in [201]. Briefly, root sections were incubated for 1 hr in phosphate buffer (pH 7.0) with or without 600 μg total crude adhesin protein before washing in fresh phosphate buffer. D5250 (Rlv3841[pIJ11268]) grown on TY slopes for 3 days at 28^{o}C was resuspended in liquid TY. 20 mL of the resuspension (OD_{600} = 0.1) was added to pea roots of each group and roots incubated on a rocking platform for 1 hr. D5250 luminescence was assayed in triplicate using a GloMax ® Multi+ detection system (Promega). Roots were washed and weighed before luminescence was measured using a GloMax ®-Multi Jr single-tube multimode reader (Promega). Data was

normalized for the weight of roots and luminescence of starting culture, which was used to calculate the relative luminescence of a single bacterial cell. Roots were also imaged using a NightOWL II LB 983 imaging system (Berthold).

2.6.4 Sterilisation and germination for whole root attachment assays

Seeds of pea (*P. sativum*), soybean (*Glycine max*) and barley (*Hordeum vulgare*) were sterilized as described in 2.6.3. Boiling tubes (200 mL size, DURAN ®) were filled ¾ full with vermiculite before 25 mL rooting solution (see Table 2-5) adjusted to the desired pH was added. After sealing with foam bungs, boiling tubes were autoclaved. Sterile seeds were planted 1 per tube and germinated under foil for 4 d in a controlled growth chamber at 23^{o}C. Plants were grown for a further 3d with a 16:8 (light:dark) photoperiod.

2.6.5 Colony count whole root attachment assays

To quantify attachment of Rlv3841 strains by colony count, seeds were germinated and grown as described in 2.6.4. Rlv3841 strains were grown on UMA slopes (see 2.2.1) at pH 7.0. Plants were removed from vermiculite and Rlv3841 strains resuspended to OD_{600} 0.1 in 15 mM MES/HEPES buffer (adjusted to pH 7.0). Strain luminescence was tested as described in 2.6.3. Plant roots were washed by dipping in 15 mM MES/HEPES buffer and placed in fresh boiling tubes. 50 mL Rlv3841 strain resuspension was added, with plants secured such that only roots were immersed. Control plants had 50 mL 15 mM MES/HEPES added. After a 1 hr incubation at room temperature with gentle shaking (20 rpm), plants were removed and washed by dipping in MES/HEPES buffer. Shoots were excised and roots weighed before placing in 50 mL Falcon tubes with 20 mL fresh MES/HEPES buffer and vortexing (Heidolph Multi Reax shaker) at the maximum speed

for 10 min. For attachment assays in which roots were also ground, a pestle and mortar were used after vortexing to grind roots in 10 mL fresh MES/HEPES buffer, which was then pooled with the buffer from vortexing. The suspension was centrifuged at 2,500 x g for 5 min to pellet the Rlv3841 cells released from the roots. Each pellet was resuspended in 1 mL TY, with 100 µL volumes of serial dilutions plated onto TY agar containing strep. Plates were incubated at 28°C for 3-4 d before the colonies were counted.

2.6.6 Lux whole root attachment assays

For whole-root attachment assays of Rlv3841 strains using luminescence (Lux assays), seeds were germinated and grown as described in 2.6.4 and plant roots inoculated as described in 2.6.5. Strain luminescence was tested as described in 2.6.3. After 1 hr incubation at room temperature with gentle shaking (20 rpm), plants were removed from inoculum and washed by dipping in MES/HEPES buffer. Shoots were excised and roots imaged using a NightOWL II LB 983 imaging system (Berthold). Lux assays were conducted with plants grown with rooting solution at pH 6.5, 7.0 or 7.5. For these pH groups, Rlv3841 strains were grown on UMA slopes of matching pH, and 15 mM MES/HEPES buffer was also of matching pH. Data was normalized as described in 2.6.3.

2.7 Insertion sequencing

Insertion sequencing protocols were used from Goodman *et al.* (2011) [169] and Perry and Yost (2014) [151] with adjustments listed below.

2.7.1 *Mariner* library construction

Plasmid pSAM_Rl carrying *E. coli* donor cells (Table 2-1) were grown in LB with kan and amp overnight and Rlv3841 recipient cells were grown on a TY strep agar slope and resuspended in TY. Cultures were pelleted at 2,500 x g for 5 min and washed 3x in TY. Donor and recipient cells were pooled in equal ratios, pelleted, resuspended in 30 μL TY and plated on nitrocellulose filters on TY agar for conjugation. After overnight incubation at 28 ^{o}C, bacteria from conjugations were resuspended in 15 % glycerol UMS and serial dilutions plated on TY str, neo to quantify Rlv3841 transposon insertion mutant concentration. Approximately 300,000 colony forming units (CFU) from each conjugation was plated in triplicate on 245 x 245 x 25 mm TY str, neo plates and grown at 28 ^{o}C for 12-18 hours until pinprick colonies formed. Colonies were scraped off plates and resuspended in 15% glycerol UMS. Three master *mariner* transposon Rlv3841 input libraries were formed from the pooling of two independent pSAM_Rl/Rlv3841 conjugations each.

2.7.2 *Mariner* library inoculation for insertion sequencing

Mariner libraries (isolated as described in 2.7.1) were grown from -80 ^{o}C stocks on UMA slopes at pH 7.0. Plants were grown as described in 2.6.4 and inoculated as described in 2.6.5. The input library DNA was extracted at this point. Output libraries (Rlv3841 *mariner* transposon mutants after 1 hr root attachment assay) were recovered from roots by vortexing as described in 2.6.5. To decrease the contamination with plant DNA and increase bacterial DNA for extraction, output libraries were grown in TY for 12 hr at 28^{o}C with shaking at 200 rpm. DNA was extracted from output libraries using a Qiagen DNeasy Blood and Tissue kit according to the manufacturer's instructions with the

following adjustments: 2 μL RNase A (100 mg/mL) was added alongside proteinase K, and DNA was eluted in 100 μL MilliQ water. DNA was quantified using a NanoDrop spectrophotometer.

2.7.3 Library preparation and sequencing

Library preparation was carried out separately for each of the input and output libraries. Transposon-tags were prepared for DNA sequencing using the method described in Wheatley *et al.* (2017) [143] to make the process compatible with the Ion Proton sequencing platform (Thermo Fisher). Briefly, linear PCR products were amplified using BioSAM primers with 500 ng template DNA. Biotinylated PCR products were purified (GeneJet PCR purification kit) and bound to Pierce streptavidin magnetic beads (Thermo Fisher). Klenow fragment (3'→5' exo$^-$, NEB) was used for second strand synthesis before digestion with MmeI (using an M12 fragment *in trans* for efficient DNA cleavage) and ligation of an Ion Proton sequencing adapter. A final round of PCR amplification was performed with a unique barcoded sequencing primer used for each sample. This barcoding (using twelve different IonXpress barcoding sequences, Thermo Fisher) enabled automated sequencing read separation by experimental sample. An Ion Torrent system (Thermo Fisher) reverse sequencing target 'trP1' was used for all samples in conjunction with the unique forward barcoded primer. All primers are listed in Table 2-3. The end sequencing template was 187 bp and was gel purified from the PCR reactions using E-Gel SizeSelect II agarose gels (Thermo Fisher). DNA for sequencing from each library was analysed on a Bioanalyzer high sensitivity DNA chip (Agilent Technologies), diluted to 100 pM and pooled in equimolar ratios. An Ion Chef and Ion PI chip kit V3

(Thermo Fisher) was used for template preparation of pooled barcoded libraries, and sequencing was performed on an IonProton system (Thermo Fisher).

2.7.4 Transposon insertion analysis using a four-state hidden Markov model

Sequencing reads were analyzed on a Linux server as previously described [151, 169]. Cutadapt [202] was used for quality trimming and adapter sequence removal and resulting transposon tags were checked for a leading TA motif with a custom Perl script. Tags were mapped to the Rlv3841 reference genome (downloaded from Rhizobase, http://genome.annotation.jp/Rhizobase) using Bowtie [203] and split by replicon. Files were converted to .wig format using a custom Perl script and analyzed further with the Tn-HMM (Hidden Markov Model) Python module [183]. This module calculates the HMM state of each thymine-adenine (TA) site (at which the *Mariner* transposon can insert, [151]) and then determines the state of all TA sites in a gene to assign a gene to one of four possible classifications, indicating how mutation affects bacterium fitness: essential, defective, neutral or advantaged. Gene annotations were obtained from UniProt and from a lab-curated annotation file.

2.7.5 Transposon insertion analysis with gene fitness value calculation

Fitness values were calculated for genes (reflecting the impact of mutation on bacterium fitness) as described in [143], allowing quantification of mutants retrieved for any gene in the library as well as numerical comparison of gene fitness across different test conditions (after standardising fitness values per million library reads). The equation for fitness value is given by:

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

Potential insertion sites is the number of TA sites within the annotated length of a gene. Insertion density is the proportion of TA sites in a gene with one or more insertions. Mean read count is the mean number of reads for TA sites with at least one insertion. Fitness values can be standardised per million library reads

2.8 *R. leguminosarum* root interaction imaging

Tracking root interactions system (TRIS) experiments were conducted as described in Massalha *et al* (2017) [188], with adjustments described below.

2.8.1 Preparation of tracking root interactions systems chambers

TRIS chambers were prepared as described in [188]. Briefly, photolithography was used to pattern microchannels onto a layer of SU-8 2100 negative tone resist (Microchem) on a silicon wafer (Agar Scientific). A mask aligner was used to expose the photoresist layer through a custom mask designed by AutoCAD. The crosslinked photoresist formed the master mould. Polydimethylsiloxane (PDMS) and crosslinker (Sylgard 184 silicon elastomer kit, Dow-Corning) were mixed according to manufacturer's instructions and poured into the master mould. PDMS was cured by heating at 60 ^{o}C overnight and slabs (now referred to as TRIS chambers) were cut from the master mould and inlet/outlet/root holes added using a 1 mm biopsy punch (Agar Scientific). TRIS chambers were mounted on a glass slide using plasma bonding. A diagram of a completed TRIS chamber can be seen in Figure 6-1 A, page 297.

2.8.2 Seed sterilization and germination for Lux reporter testing and tracking root interactions system

For Lux reporter testing, seeds of vetch (*Vicia cracca* and *Vicia hirsuta*) and clover (*Trifolium repens* and *Trifolium pratense)* were gently scarified with sandpaper before immersion in 2% sodium hypochlorite solution for five min. Seeds were rinsed 10x in sterile distilled water and immersed in sterile distilled water for 5 hr to imbibe. 200 μL pipette tips with the ~5 mm of the thinnest end removed were filled with FP agar and seeds placed using sterile forceps one per tip at the widest end. 100 mL rooting solution was added to a pipette tip box and tips with agar and seeds placed in the box tip holder. Boxes were sealed with parafilm and plants germinated in the dark for 4 d, and with a 16:8 (light:dark) photoperiod for 2 d.

For TRIS, seeds were sterilized as above and 'planted' in pipette tips. Tips were embedded vertically in a sterile magenta filled 1 cm deep with 1% distilled water agar. Seeds were vernalized in the dark at 4^{o}C for 24 hr and germinated for ~60 hr at room temperature before transplanting to TRIS chambers.

2.8.3 Lux reporter testing on roots

Strain LMB487 was grown for 3 d on UMA slopes at pH 7.0 and resuspended to OD_{600} = 0.1 in 15 mM MES/HEPES buffer. Pipette tips containing germinated plants (2.8.2) were removed from boxes and the roots immersed in the LMB487 suspension. Control roots were immersed in buffer alone. After 2 hr roots were removed, washed by dipping in fresh buffer and imaged using a NightOWL II LB 983 camera (Berthold).

2.8.4 Bacterial growth and preparation for tracking root interactions systems and chamber imaging and interaction profiling systems

Bacterial strains were grown from loop inoculation taken from a TY plate into 100 mL UMS media with 3 mM pyruvate (see 2.2.1) with shaking at 200 rpm for 24 hr. Before use, bacteria were filtered through a single layer of Whatman filter paper and 20 mL culture centrifuged at 100 x g for five min. The upper 16 mL of media was gently removed by pipetting and 16 mL filter sterilized FP media added. This process was repeated with 8 mL fresh FP media added the second time. Bacterial motility was observed using an inverted optical microscope with a 10 x dark-field objective (Olympus) and data evaluated with CellSens (Olympus).

2.8.5 Tracking root interactions systems setup and confocal imaging

Plants in pipette tips were transplanted into TRIS chambers (2.8.2) and grown on a 20 ^{o} slant to promote gravitropic root growth toward the chamber outlet in a 16:8 (light:dark) photoperiod for 36 hr. For imaging, 1.2 mm polypropylene tubes filled with FP media were connected to the inlet and outlet holes. 1 mL plastic syringes filled with FP media were connected to outlet tubes using 18-gauge needles. Chambers were mounted on a motorized microscope stage (inverted laser scanning confocal Ti-eclipse, Nikon) and covered with a custom-made transparent lid. Bacterial cells were introduced into the chambers by immersing the inlet tube in prepared culture (2.8.4) and drawing back gently on the outlet syringe.

Images were acquired at ± 40 μM from the root mid-plane using a 10 x objective lens. Laser excitation at 488 nm with an emission window of 500-550 nm was used for GFP

fluorescence and excitation of 561 nm with an emission window of 595-650 nm was used for mCherry.

2.8.6 Seed sterilization and germination for Chamber Imaging and Interaction Profiling Systems (ChIIPS)

Seeds for chamber imaging and interaction profiling systems (ChIIPS) were sterilized as described in 2.8.2. For alfalfa (*Medicago sativa*) and clover (*T. repens*), seeds were placed on sterile filter paper on a layer of 1% DWA in a square plate, sealed with parafilm. Plates were placed upright to promote gravitropic root growth and plants were germinated at room temperature in the dark for 24 hr and in a 16:8 (light:dark) photoperiod for a further 24 hr (*M. sativa*) or 3 d (*T. repens*). For vetch (*V. villosa*) and lotus (*L. japonicus*), seeds were lightly embedded in 1% DWA in a square plate with a layer of sterile cellophane directly below the seed. This allowed root growth over the layer of cellophane and not into the agar. Seeds were germinated in the dark for 24 hr and in a 16:8 (light:dark) photoperiod for a further 3 d (*L. japonicus*) or 5 d (*V. villosa*) before use.

2.8.7 Chamber imaging and interaction profiling systems setup and confocal imaging (including for polarity experiments)

Sterile ChIIPS chambers (adapted from an SPL Life Sciences single well cell culture slide #31301) were filled halfway with 2% distilled water agar to form an artificial base. Plants were transferred to the chamber so they lay horizontal with the root perpendicular to the chamber sides. Plants were secured by embedding a sterile metal hook into the agar over the shoot, and the hook secured with a small spot of water agar. For imaging, the chamber was filled with bacterial suspension and the lid placed on such that the inner lid surface

was in contact with the liquid, allowing clear imaging. The lid and chamber were sealed with parafilm and transferred to the motorized confocal stage. Imaging took place on a Leica TCS SP5 laser scanning confocal microscope. Images were acquired at ± 20 μM from the root mid-plane using a 10 x objective lens. Laser excitation at 488 nm with an emission window of 500-530 nm was used for GFP fluorescence and excitation of 543 nm with an emission window of 600-630 nm was used for mCherry.

2.9 Bioinformatics, data handling and statistical methods

Geneious R10 was used for primer design and local sequence alignment [196]. Global nucleotide and protein sequence alignments were carried out using BLASTn and BLASTp (NCBI) [204]. Protein-protein interaction networks were predicted and visualized using STRING [205]. Cellular protein localization was predicted using pSORTb v 3.0.2 [206]. In bacterial attachment assays, luminescence was evaluated using IndiGO software (Berthold) and subject to statistical testing in GraphPad Prism 8. Unpaired t-tests were used with the following p values: * = $p < 0.05$, ** = $p < 0.001$, *** = $p < 0.0005$, **** = $p < 0.0001$. Imaging data from tracking root interactions systems and chamber imaging and interaction profiling systems was evaluated with FIJI [207] and LAS AF (Leica). Data handling was largely in MS Excel and all graphs were generated using GraphPad Prism 8. Analysis of sequencing reads from INSeq transposon insertion mutagenesis was carried out using a four-state HMM as described in 2.7.4.

Chapter 3

Investigating novel root attachment factors in *Rhizobium* using a new luminescence-based root-attachment assay

3.1 Introduction

Soil bacteria of the order Rhizobiales are gram negative α-proteobacteria able to form nitrogen-fixing symbioses with their host legumes. Within this symbiosis, rhizobia fix atmospheric dinitrogen (N_2) into ammonia (NH_3), which can be used by the host plant for amino acid synthesis, in return for carbon sources in the form of dicarboxylic acids [26]. One of the early stages of nitrogen fixing symbiosis formation is attachment of rhizobia to the plant root hairs; attachment to bulk epidermal root surface also occurs, but by different mechanisms [34]. Attachment to root hairs is thought to be governed by a pH-dependent two-component system conserved among rhizobia. Under acidic soil conditions, a plant lectin localised at root hair tips is bound by glucomannan, polarly located on rhizobia cells [47, 56]. Under alkaline conditions, the plant lectin disassociates from roots and instead it has been proposed that rhizobia use an extracellular rhicadhesin protein, bound to the bacterial surface by calcium ions, to attach to an unknown plant receptor [53, 71, 208–211]. Following attachment, infection thread formation allows rhizobia to infect nodules, in which they form bacteroids and fix nitrogen [26].

Successful attachment to root hairs is important for subsequent symbiosis formation; a glucomannan mutant in Rlv3841 could still nodulate pea, but when co-inoculated with wild-type was almost completely outcompeted for nodule occupancy [56]. Attachment to the bulk root epidermis of various plants is also likely to be important for persistence of rhizobia in the soil; root exudates can serve as preferential growth substrates for rhizobia [34] and as much as 20 % of photosynthate can be exuded from roots [35].

Rhicadhesin has been reported to be important for attachment in *Rhizobium*-legume symbioses in multiple studies [52–54, 209–211]. Smit *et al*. (1987) first hypothesised the existence of a calcium-dependent adhesin when they showed that low calcium growth

conditions reduced attachment to glass and root hair cap formation for multiple rhizobia species under pH 7.5 attachment conditions [52]. A candidate 14 kDa calcium-binding rhicadhesin protein was later isolated from the cell surface of *Rhizobium leguminosarum* biovar *viciae* 248 (Rlv248) and found to inhibit root attachment at pH 7.5 when roots were pre-treated with adhesin. Surface preparations from other rhizobia, but not non-rhizobia, demonstrated the same properties, and were able to inhibit attachment to both legume and non-legume plants. This suggests conservation of rhicadhesin across rhizobia and a common mechanism of attachment to root hairs under alkaline conditions [53]. Calcium ions were shown to be required for anchoring of rhicadhesin to the rhizobia cell surface [211] and purified rhicadhesin rescued the attachment negative phenotype of a *chvG* *Agrobacterium tumefaciens* mutant, deficient in β-1,2 glucan production [54]. A putative plant receptor for rhicadhesin was also identified based on its ability to prevent rhicadhesin-mediated inhibition of Rlv248 attachment to root hairs at pH 7.5 [209]. Despite the reported importance of rhicadhesin, its identity remains unknown at the gene or protein level [34].

A two-component model has been proposed for the primary attachment to root hairs [34]. Primary attachment is defined as the initial, reversible binding of bacteria, whilst secondary attachment is defined as tight, irreversible binding [34]. Secondary attachment to root hairs occurs in rhizobia through cellulose microfibril, polysaccharide and secreted protein synthesis [34, 71].

Multiple other factors have been implicated in the primary and secondary attachment of rhizobia to the epidermis of plant roots (summarised in Figure 1-3). Motility is important for chemotactic attraction and root exudates (including flavonoids) induce positive chemotaxis in strains of *Rhizobium* [212, 213]. Flagella (the proteinaceous protruding

filaments, driven by a membrane-embedded motor and responsible for motility in many bacteria) provide propulsion and hence allow rhizobia to reach root surfaces for attachment [44]. The flagellum itself may also possess adhesin properties [45, 61, 214, 215], though to what extent flagella contribute as adhesins in Rlv3841 is unclear. EPS is also known to play a role in attachment and a mutant in *pssA* (a key acidic EPS biosynthesis gene) of Rlv3841 is strongly defective in attachment to pea roots after 90 min at pH 6.5 or 7.5, especially to root hairs [56]. Ausmees *et al*. (2001) also identified the *Rhizobium* adhering proteins (Raps) [57] which are negatively regulated by *praR* in Rlv3841 [59]. Mutation of *praR* in Rlv3841 leads to a phenotype of hyper-attachment to pea roots, and only combinations of mutations in Rap genes could abolish the increased attachment of a *praR* mutant [59]. PraR was also shown to negatively regulate *plyB*, a glycanase important for correct EPS processing to facilitate attachment [58], *rosR*, a regulator of cell surface components and polysaccharides (mutation of which inhibits *Rhizobium*-clover symbiosis [216]), and predicted 'cadherin' attachment proteins [59]. There is a comparative lack of characterisation of attachment factors in *Rhizobium*-legume symbioses relative to other symbiotic stages [25]. This is possibly due to the difficulty of assessing root attachment with a method that is both simple and has reasonable throughput, where larger numbers of experiments can be performed to give statistical significance. Previously reported root attachment assays have relied on a handful of methods, each of which have limitations. High magnification microscopy for the counting of root-attached bacteria [52, 53, 210, 217–219] relies on extrapolation of data from a small count area and is very time consuming. The same limitations apply to assaying by confocal microscopy [56]. Whilst vortexing can be used to recover root-attached bacteria for plating and enumeration, this method does not allow observation of the dynamics of

spatial attachment. Frederix *et al*. (2014) used a scintillation-based attachment assay, radiolabelling rhizobia [59]. Whilst quantitative, radiolabelling presents obvious limitations. The same group then developed a Lux luminescence-based attachment assay, labelling rhizobia with bioluminescence genes [59]. This method, combined with appropriate imaging techniques, allows light emission from plant roots to be used as a proxy for enumerating the bacterial attached. However, this study used an attachment assay which examined root sections excised from pea seedlings. This excision creates an unnatural exudation source from the wounding site, which is likely to affect attachment dynamics and rapidly alter root physiology.

In this chapter, the rhicadhesin protein fraction from Rlv3841 is characterised using root section attachment assays and proteomic analysis. A major aim of this work was to establish new tools for the investigation of primary root attachment in *Rhizobium*-legume symbioses. To this end, the development of a new whole-root luminescence-based attachment assay, which is compatible with studying bacterial attachment under conditions of different pH, is presented. This method was validated using a variety of Rlv3841 strains mutated in genes encoding known attachment factors. A further key aim was to generate a more holistic overview of primary attachment determinants and their pH dependency. A bioinformatic approach was therefore taken to identifying potential novel attachment factors in Rlv3841 and mutants in a subset of these tested experimentally using the Lux attachment assay. A better understanding of primary attachment determinants is of relevance both to *Rhizobium*-legume symbiosis development and the mechanisms of persistence of rhizobia in the soil.

3.2 Results and discussion

3.2.1 A crude adhesin fraction isolated from Rlv3841 inhibits bacterial attachment to pea root sections

A crude adhesin fraction was isolated from Rlv3841 cells (2.5.1) using the method described in [53] and visualized on an SDS-PAGE gel (Figure 3-1). A 14 kDa band was seen in both the soluble and crude adhesin fractions and was absent from the membrane fraction, in agreement with the results of Smit *et al*. (1989) [53]. The banding pattern of the crude adhesin fraction SDS-PAGE gel showed a high level of similarity to the banding pattern seen in the crude adhesin fraction SDS-PAGE gel of Smit *et al*. (1989) [53] (data not shown).

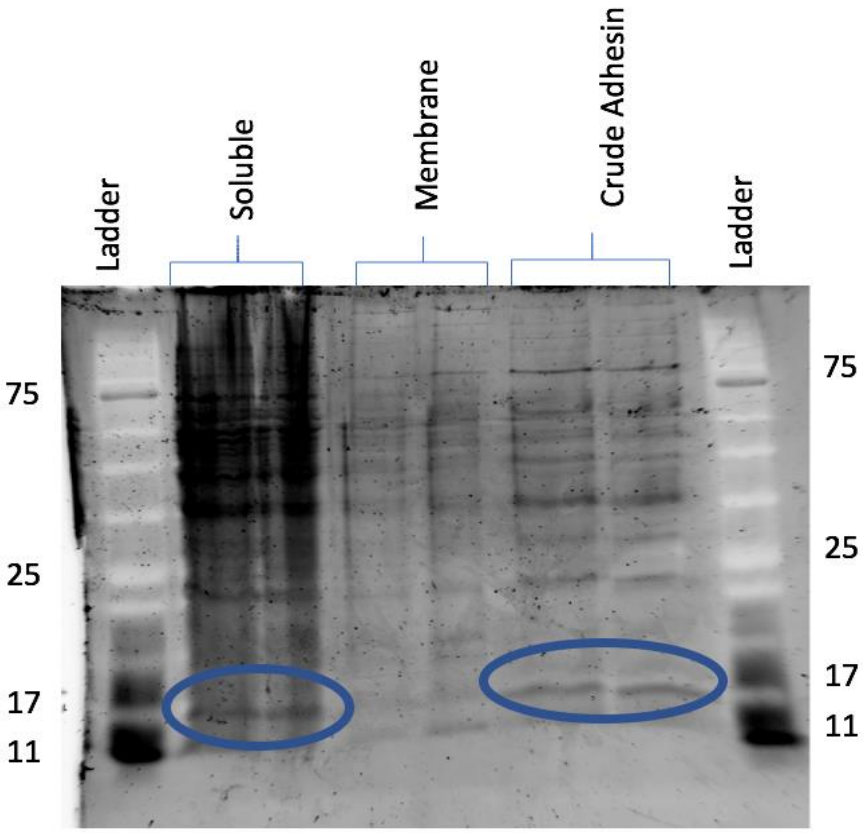


Figure 3-1. Sypro Ruby stained SDS-PAGE gel of the soluble, membrane and crude adhesin fractions isolated from Rlv3841. Smit *et al.* (1989) [53] described a 14 kDa band of interest and bands of this size are circled in blue. Molecular weight markers are shown at either edge of the gel (Ladder). Numbers indicate weight markers (kDa).

This crude adhesin fraction was used in a pea root section attachment assay (2.6.3) at pH 7.0 to determine whether pretreatment of roots with 600 ug crude adhesin protein for 1 hr affected subsequent 1 hr attachment of Rlv3841. Data normalization was carried out as described in Frederix *et al*. (2014) [59]. Preincubation of roots with crude adhesin fraction led to a 72 % reduction in the average number of attached Rlv3841 cells (Figure 3-2).

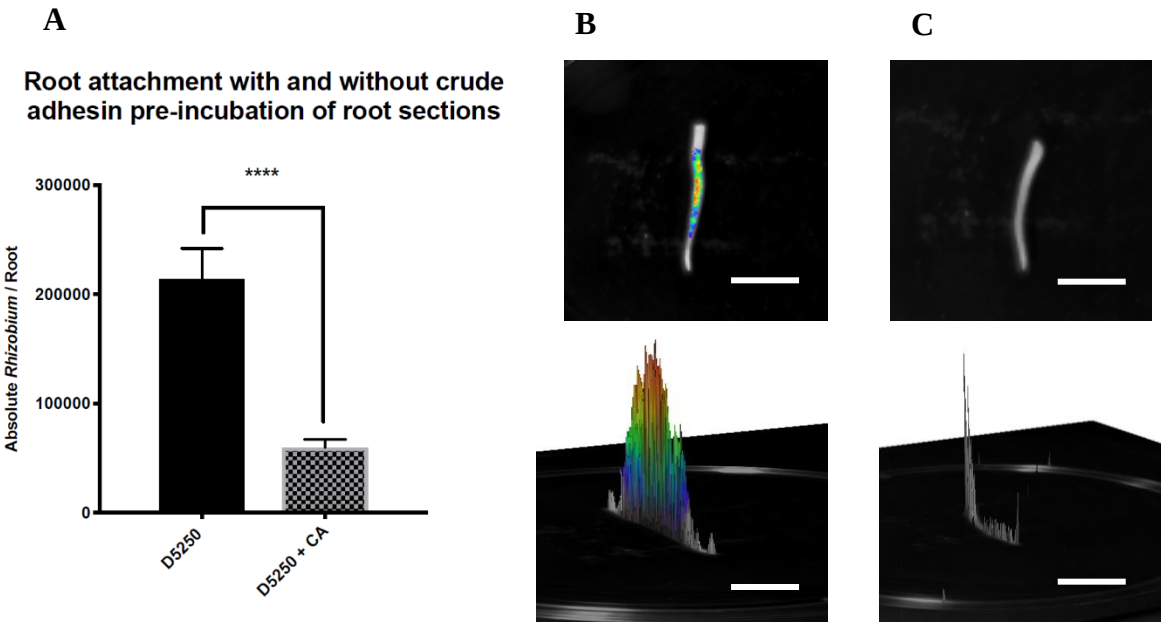


Figure 3-2. Attachment of bacteria to pea root sections, without and with crude adhesin pretreatment. **A** – Number of bacteria - Rlv3841[pIJ11282] (D5250) - attached to pea root sections after 1 hr without (D5250) and with (D5250 + CA) pretreatment with crude adhesin (CA) for 1 hr. **B** – Aerial and lateral view of luminescence signal from a representative root of the D5250 group. **C** - Aerial and lateral view of luminescence signal from a representative root of the D5250 + CA group. Imaging was using a NightOWL II LB 983 camera. White bars are for scale; each represents 1 cm. Data is displayed as mean ± SEM, n=10 roots. An unpaired t-test was used to compare groups. **** = $p < 0.0001$

The ability of a crude adhesin fraction to inhibit the attachment of Rlv3841 to pea roots agrees with the results of Smit *et al*. (1989) [53] although, due to their use of an 'attachment class' metric (where observed levels of attachment were ranked into one of four qualitative categories), it is difficult to make a direct comparison of the data. This result, using a crude adhesin preparation from Rlv3841, also seems to indicate that the adhesin activity is also conserved in this strain of *Rhizobium*. This is important as,

although conservation of rhicadhesin across multiple strains of rhizobia was demonstrated [53], Rlv3841 was not tested. Although the original attachment assays in [53] were carried out at pH 7.5, the ability of the crude adhesin fraction to inhibit Rlv3841 attachment at pH 7.0 is not unexpected, as both the rhicadhesin and glucomannan mechanisms of attachment are thought to operate at neutral pH [34, 71].

3.2.2 The 14 kDa crude adhesin band is made up of at least 15 protein components

The 14 kDa band shown in Figure 3-1 was sent for protein identification by LC-MS/MS at the analytics core facility, University of Duisberg-Essen (see 2.5.2). A total of fifteen protein hits were identified from the isolated SDS-PAGE gel band (Table 3-1).

Table 3-1. Gene designations for the proteins identified by LC-MS/MS from the 14 kDa band of Rlv3841 crude adhesin preparation, ordered by relative abundance.

Gene	Protein description	Protein size (kDa)	Predicted localization	Relative abundance
RL4733	Conserved hypothetical protein	17	Un	100
ndk (*R1580*)	Nucleoside diphosphate kinase	15	Ex	83
pal (*RL3968*)	Putative OmpA family peptidoglycan associated lipoprotein	19	OM	42
RL0770	Putative phasin, phasin-2 superfamily	16	Un	10
rpIQ (*RL1799*)	50 S ribosomal protein L17	15	Cyt	8
RL1635	Putative outer membrane protein	19	Un	7
omp19 (*RL4441*)	Outer membrane lipoprotein Omp19	18	OM	5
atpF1 (*RL0928*)	ATP synthase subunit b 1	19	CytM	5
rpsH (*RL1788*)	30 S ribosomal protein S8	15	Cyt	3
rplT (*RL0268*)	50 S ribosomal protein L20	15	Cyt	3
rpsP (*RL4549*)	30 S ribosomal protein S16	14	Cyt	3

rpsI (*RL1672*)	30 S ribosomal protein S9	17	Cyt	3
rpIU (*RL4676*)	50 S ribosomal protein L21	12	Cyt	2
rpiR (*RL1790*)	50 S ribosomal protein L18	13	Cyt	2
rosR (*RL1379*)	Putative nodulation competitiveness transcriptional regulator	16	Un	2

Predicted localization data is from pSORTb [206] and coded as follows: Cty = cytoplasmic, Un = unknown, CytM = cytoplasmic membrane, Ex = extracellular, OM = outer membrane. Relative abundance reflects intensity of protein hits detected in LC-MS/MS and is indexed to 100 for RL4733, the most abundant of the detected proteins.

Whilst rhicadhesin was reported to be a 14 kDa protein, this was determined by running both crude and purified protein on SDS-PAGE gels against protein standards of known size [53]. However, this method can lead to inaccuracies, especially for small proteins [220]. This is further evidenced by the fact that the protein band running at 14 kDa on an SDS-PAGE gel (Figure 3-1) which was analysed here contained proteins of 12-19 kDa (Table 3-1). Therefore, none of these proteins should be ruled out as a rhicadhesin based on size alone.

This list contains 8 proteins which are predicted by pSORTb to be localized in the cytoplasm or cytoplasmic membrane (RpiQ, AtpF1, RpsH, RpsT, RpsP, RpsI, RpIU and RpiR). These include ribosomal subunit proteins. This would indicate that at least some Rlv3841 cell lysis had occurred during the preparation of crude adhesin fraction. This is in contrast to a previous report [53], where it was claimed that cell viability was not affected, although no data or methodology for testing this was shown. However, given the high sensitivity of MS/MS protein detection methods, high abundance of ribosomal proteins in bacterial cells [221], and the low relative abundance of ribosomal subunit proteins in the sample, this indicates that any cell lysis was likely to be minor. These cytoplasm or cytoplasmic membrane localized proteins are unlikely candidates for rhicadhesin given their described functions and predicted subcellular localization.

Of the remaining candidates, RL0770 (putative phasin) is also an unlikely candidate. Phasins play a role in the accumulation and stabilization of poly-β-hydroxybutyrate granules, an intracellular carbon storage mechanism [222, 223]. Interestingly, phasin-like proteins were highly up-regulated in a *rosR* mutant of *R. leguminosarum* biovar trifolii. However, as a *rosR* mutant is known to be defective in symbiosis [216], this provides further evidence against RL0770 being a rhicadhesin.

RosR itself (RL1379) was also identified as the least abundant protein in the sample in LC-MS/MS but, although known to be involved in transcriptional regulation of cell surface properties [216], there is no evidence for the direct involvement of this protein in attachment.

RL1580 encodes a nucleoside diphosphate kinase (NDK); these are generally cytoplasmic nucleotide metabolism enzymes, although may also be involved in the supply of nucleotide-linked sugars for mycobacteria cell wall polysaccharide synthesis [224]. NDK has been shown to be secreted by the PrsDE type I system in Rlv3841, and proteins secreted through this system (including RapA2 and RapC) have previously been shown to be important for attachment [57, 59, 225]. However, as mutants lacking secreted NDK have been reported to show no symbiosis defect it has not been implicated in attachment [225].

RL3968 is annotated as the Pal component of the characterised Tol-Pal system. This system is critical for outer membrane integrity in Gram-negative bacteria and in cell division [226], but has no characterised role in attachment.

Omp19 (RL4441) is an outer membrane lipoprotein; a lipid anchored protein exported through the localization of lipoproteins (lol) pathway. Although the role of Omp19 in Rlv3841 is unknown, outer membrane lipoproteins have been implicated in iron uptake [227], cell surface modifications [228, 229] and adhesion [230, 231] in different bacteria. Whilst Omp19 could therefore represent a rhicadhesin, its lipid (rather than calcium ion) anchored nature argues against this. However, given the known involvement of Omp proteins in bacterial adhesion [230, 231] this should be investigated further.

RL1635 is a putative outer membrane protein (UniProt annotation, pSORTb was unable to predict localization) of 19kDa, and may represent a rhicadhesin.

Finally, RL4733 is a conserved hypothetical protein of unknown localization and the most likely rhicadhesin candidate, given it was the most abundant protein in the sample and is close to 14 kDa in size.

In order to investigate these candidates (Omp19, RL1635 and RL4733) further an attachment assay capable of profiling attachment factor importance at different pHs (given the pH dependent nature of rhicadhesin mediated attachment, [34, 71]) was developed to investigate this and other possible novel attachment factors further.

3.2.3 Evaluating the suitability of vermiculite for attachment studies at a range of pHs

For the development of an attachment assay capable of determining the pH dependency profile of attachment factors, plants and bacterial strains were buffered during growth at the same pH. Many previous studies featured inconsistencies in pH range attachment assays, such as growing bacteria and plants at neutral pH before resuspending bacteria in a pH 6.5 or 7.5 buffer for the assay [56]. This is unlikely to be representative of more realistic attachment conditions, where plants and bacteria may be exposed to a more consistent pH for time during growth before attachment. It has previously been reported that vermiculite is inappropriate for pH studies due to its high buffering capacity [232], although the type of vermiculite used was unspecified. The buffering capacity of Sinclair Pro fine vermiculite was tested as described in 2.6.1, with results shown in Figure 3-3.

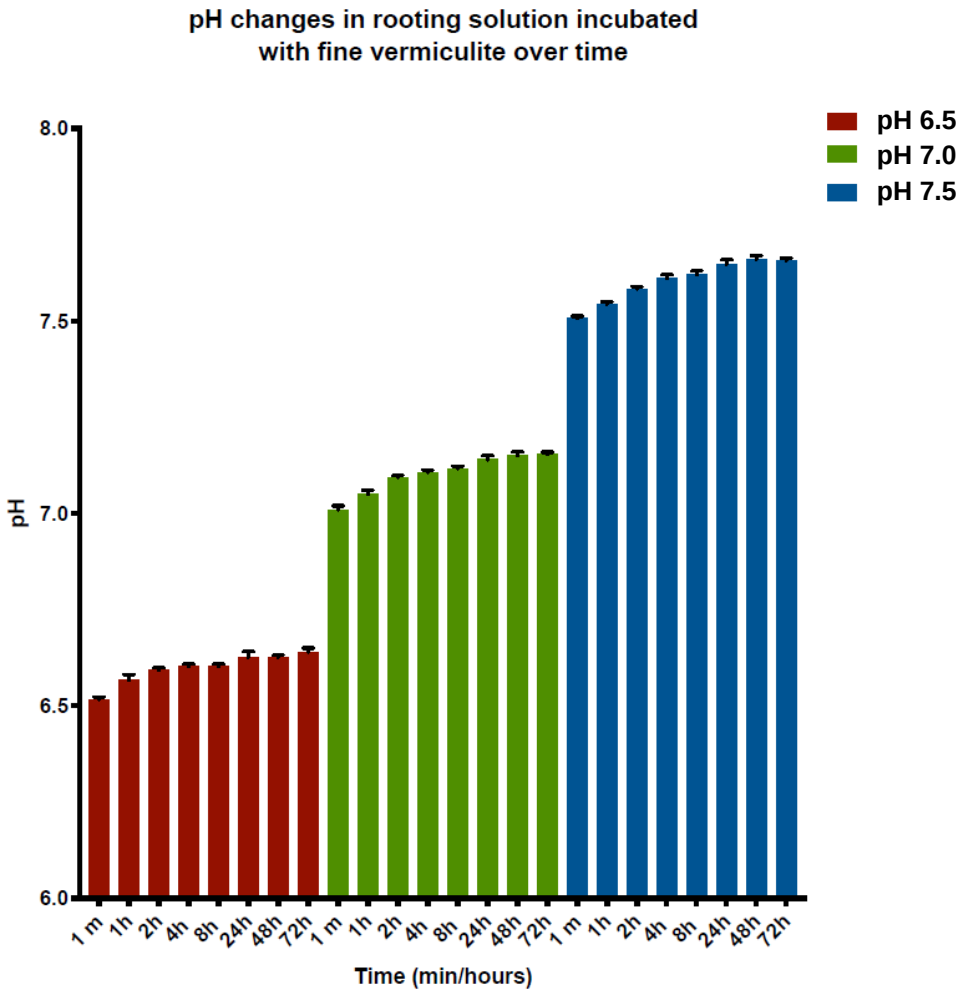


Figure 3-3. pH changes in rooting solution incubated with fine vermiculite over time. 25mL of rooting solution at pH 6.5 (red), 7.0 (green) or 7.5 (blue) was added to 10 g fine vermiculite and the pH measured over time, up to 72 hours. n=3 per group.

In contrast to the results of [232], large changes in the pH of rooting solution were not seen. This could be to do with the brand of vermiculite (not specified in [232]), cleanliness (vermiculite was washed thoroughly and allowed to dry before all experiments in this work) or different concentrations of rooting solution components [232]. Whilst small upward pH changes were observed for all test groups, due to its low buffering capacity the Sinclair Pro vermiculite was deemed appropriate for use in pH attachment studies.

3.2.4 Validating Lux for measuring attachment of bacteria to whole roots

A major aim of this work was to develop a new whole-root luminescence-based attachment assay which could be used for pH attachment studies. Such a tool would be useful given the limitations of previously reported techniques (3.2). Luminescence was previously reported to be a reliable method for measuring Rlv3841 attachment to root sections [59], but this was not validated with whole roots. In order to use luminescence, a plasmid (pIJ11282) which constitutively expresses Lux (using a pNeo promoter) needs to be introduced into each strain assessed. Root attachment measured by vortexing pea roots and counting the colonies released (2.6.5) was compared to assessment of attachment using luminescence data (2.6.6) (Figure 3-4).

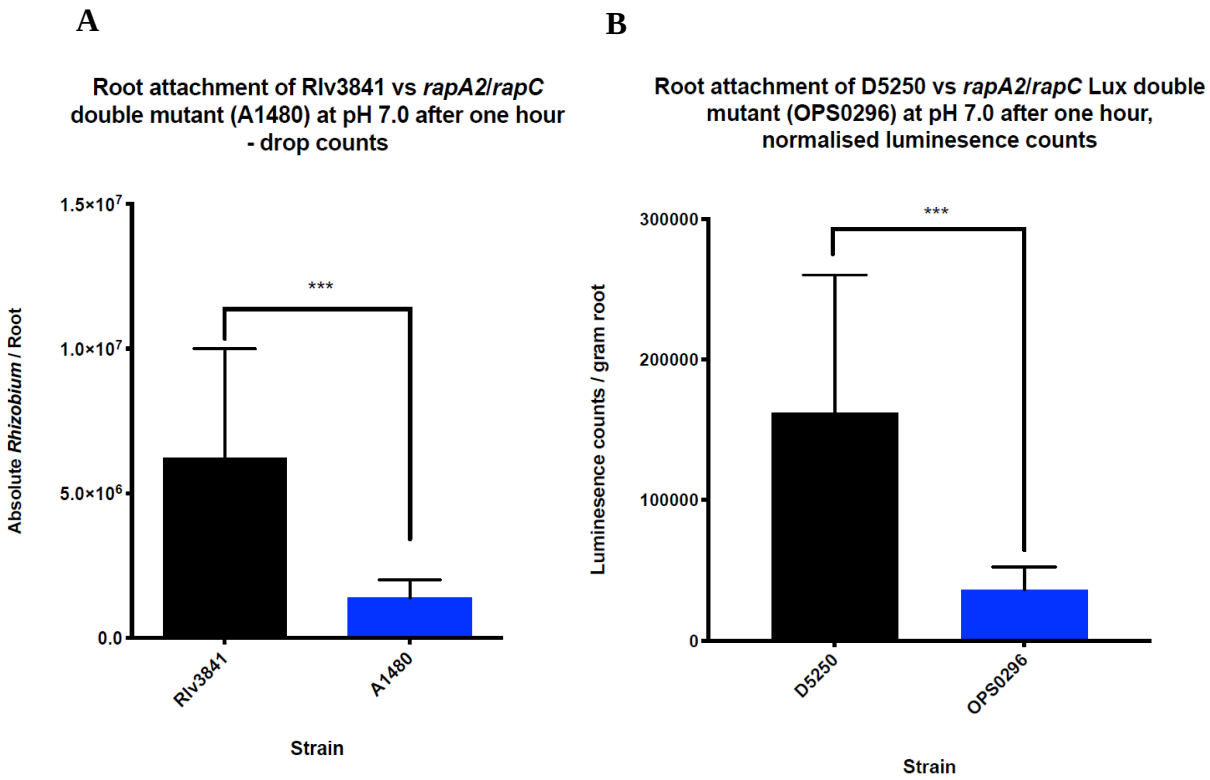


Figure 3-4. Comparison of colony counts and luminescence as a method of assessing attachment of Rlv3841 strains to pea roots. **A** – Evaluation by colony counting of bacterial attachment to whole pea roots after 1 hr, pH 7.0; wild-type Rlv3841 vs A1480 (a *rapA2*/*rapC* double mutant, [59]). A1480 shows, on average, a 77.7% reduction in attachment to roots. **B** – Evaluation by luminescence of bacterial attachment to whole pea roots after 1 hr, pH 7.0; D5250 (Rlv3841[pIJ11282]) vs OPS0296 (A1480[pIJ11282]) following data normalization for weight of roots and luminescence of starting culture. OPS0296 shows, on average, a 74% reduction in attachment to roots. Data is displayed as mean ± SEM, n=10. An unpaired t-test was used to compare groups. *** = $p < 0.0005$

Together with wild-type Rlv3841, strain A1480, a *rapA2*/*rapC* double mutant, known to be defective in root attachment at pH 7.0 [59] was used to assess the two methods. In the colony count method, the mutant strain A1480 showed a 77.7% reduction in attachment compared to Rlv3841 (Figure 3-4, A). In the luminescence method, this reduction was 74% (Figure 3-4, B), mirroring the colony count data.

An example of Lux assay readout from root imaging after inoculation with different Rlv3841 strains is presented for reference in Figure 3-5, below.

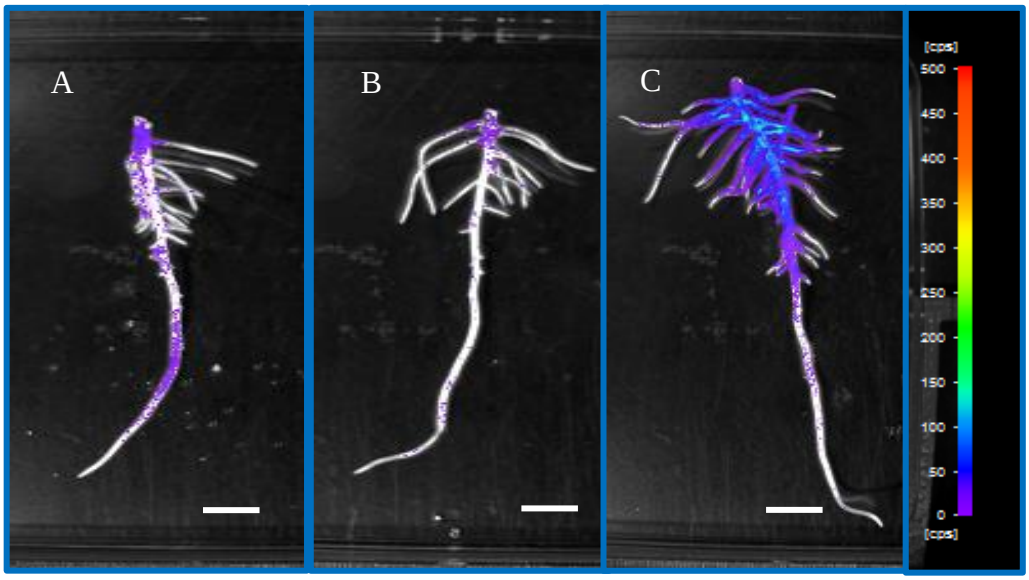


Figure 3-5. A representative example of Lux assay root imaging data following 1 hr attachment assays. **A** - luminescence signal from a pea root inoculated with D5250 (Rlv3841[pIJ11282]) for 1 hr under pH 7.0 test conditions. **B** – with OPS0804 (a *pssA* mutant). **C** – with OPS0111 (a *praR* mutant). Here, luminescence signal indicates reduced attachment of a *pssA* mutant and increased attachment of a *praR* mutant in relation to D5250 even before data normalisation (see 3.2.5 and 3.2.6). All images are scaled to a counts per second cut-off of 500 (scale bar shown right). White bars are for scale; each represents 1 cm. n = 10 per group, representative image provided

3.2.5 Validation of Lux-based attachment assay under different pH conditions using a range of Rlv3841 mutants

As normalised luminescence enabled accurate quantification of bacterial attachment, multiple further strains with characterised attachment phenotypes were tested under different pH conditions (Figure 3-6). Most strains used with this assay are pK19mob integration mutants. To isolate these mutants, an internal fragment of the target gene was

cloned into the pK19mob vector and conjugated into wild-type Rlv3841. Homologous recombination causes single crossover integration mutagenesis and disruption of the target gene. See also 2.4.3.

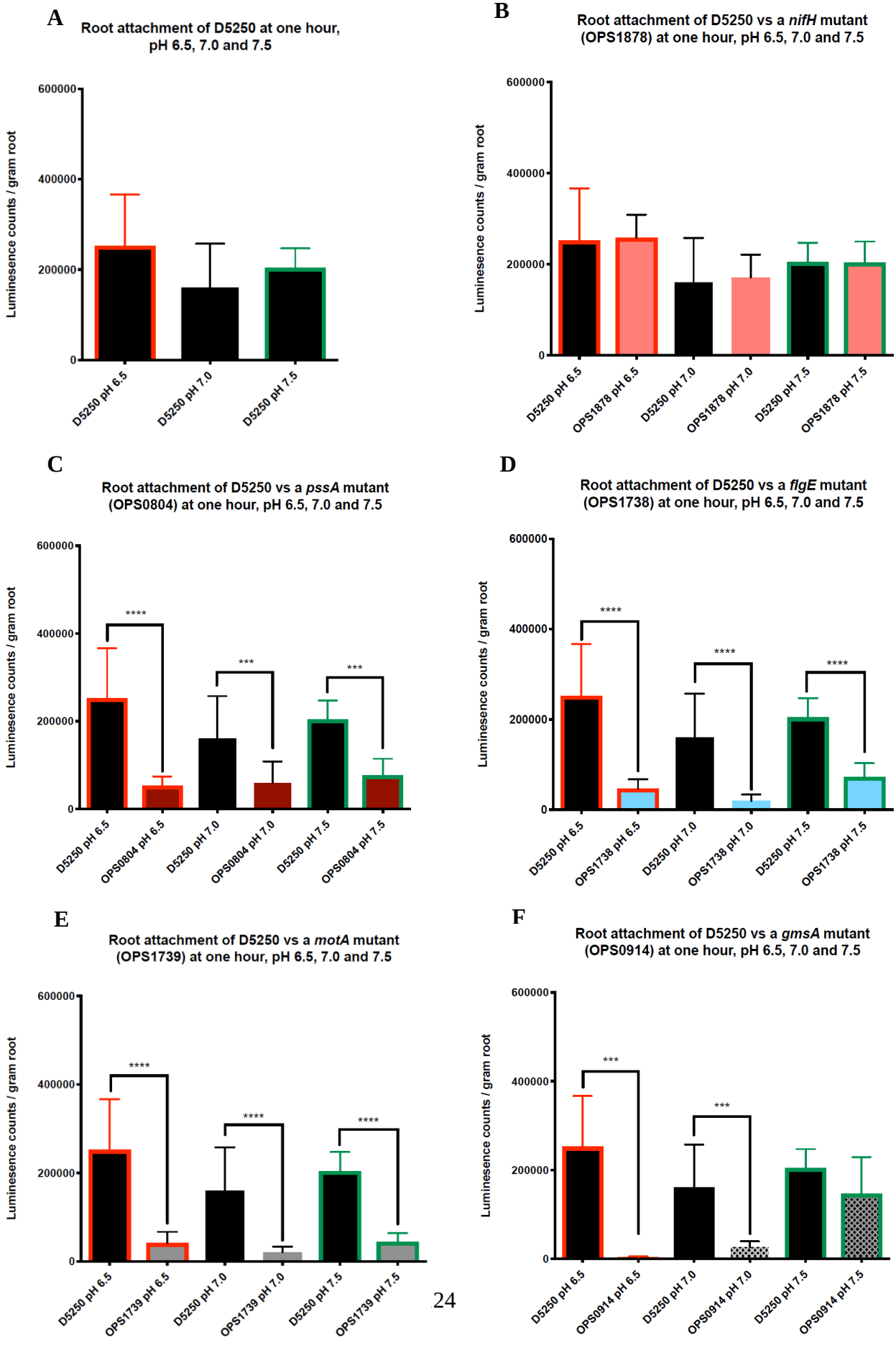

A
Root attachment of D5250 at one hour, pH 6.5, 7.0 and 7.5
Luminescence counts / gram root
600000
400000
200000
0
D5250 pH 6.5
D5250 pH 7.0
D5250 pH 7.5
B
Root attachment of D5250 vs a nifH mutant (OPS1878) at one hour, pH 6.5, 7.0 and 7.5
Luminescence counts / gram root
600000
400000
200000
0
D5250 pH 6.5
OPS1878 pH 6.5
D5250 pH 7.0
OPS1878 pH 7.0
D5250 pH 7.5
OPS1878 pH 7.5
C
Root attachment of D5250 vs a pssA mutant (OPS0804) at one hour, pH 6.5, 7.0 and 7.5
Luminescence counts / gram root
600000
400000
200000
0
****
***
***
D5250 pH 6.5
OPS0804 pH 6.5
D5250 pH 7.0
OPS0804 pH 7.0
D5250 pH 7.5
OPS0804 pH 7.5
D
Root attachment of D5250 vs a flgE mutant (OPS1738) at one hour, pH 6.5, 7.0 and 7.5
Luminescence counts / gram root
600000
400000
200000
0
****
****
****
D5250 pH 6.5
OPS1738 pH 6.5
D5250 pH 7.0
OPS1738 pH 7.0
D5250 pH 7.5
OPS1738 pH 7.5
E
Root attachment of D5250 vs a motA mutant (OPS1739) at one hour, pH 6.5, 7.0 and 7.5
Luminescence counts / gram root
600000
400000
200000
0
****
****
****
D5250 pH 6.5
OPS1739 pH 6.5
D5250 pH 7.0
OPS1739 pH 7.0
D5250 pH 7.5
OPS1739 pH 7.5
F
Root attachment of D5250 vs a gmsA mutant (OPS0914) at one hour, pH 6.5, 7.0 and 7.5
Luminescence counts / gram root
600000
400000
200000
0
***
***
D5250 pH 6.5
OPS0914 pH 6.5
D5250 pH 7.0
OPS0914 pH 7.0
D5250 pH 7.5
OPS0914 pH 7.5

Figure 3-6. Comparison of wild-type Rlv3841 and known attachment deficient strains under different pH conditions in a Lux whole root attachment assay. **A** – Root attachment of D5250 (Rlv3841[pIJ11282]) under pH 6.5, 7.0 and 7.5 conditions. Data from A is presented for D5250 for all further comparisons. **B** – D5250 vs OPS1878 (*nifH* mutant). **C** – D5250 vs OPS0804 (*pssA* mutant). **D** – D5250 vs OPS1738 (*flgE* mutant). **E** – D5250 vs OPS1739 (*motA* mutant). **F** - D5250 vs OPS0914 (*gmsA* mutant). Strains and pH conditions are given on the x axes. Data is displayed as mean ± SEM, n=10. An unpaired t-test was used to compare groups. *** = p <0.0005, **** = p <0.0001

3.2.5.1 Wild-type attachment is the same at pH 6.5, 7.0 and 7.5

No statistically significant difference was seen in the ability of D5250 (Rlv3841[pIJ11282]) to attach to pea roots under any of the pH conditions tested (Figure - 6, A). It is important to note that the pH conditions used here (6.5, 7.0 and 7.5) were chosen because root attachment assays reported in the literature (e.g. [53, 56, 59]) were performed within this pH range.

3.2.5.2 A *nifH* mutant is unchanged in attachment relative to wild-type

A *nifH* mutant was tested in this attachment assay and compared with wild-type Rlv3841 (Figure 3-6, B) at the three pHs. NifH is a nitrogenase iron protein essential for nitrogen fixation. A *nifH* mutant is a non-nitrogen fixing strain but otherwise identical to Rlv3841 [233]. No statistically significant difference was seen in the ability of a *nifH* mutant to attach under any pH conditions in comparison to wild-type. This result is important as it demonstrates that pK19 insertion mutagenesis in itself does not affect the attachment of Rlv3841 strains to pea roots.

3.2.5.3 Mutants in *pssA*, *flgE* and *motA* are impaired in attachment at all pHs relative to wild-type

OPS0804, a *pssA* mutant, and OPS1738, a *flgE* mutant (Figure 3-6, C-D) show significant reductions in attachment under all pH conditions when compared to wild-type, in agreement with the expected phenotypes for these strains given previous literature reports (see section 3.2, [44, 56]). OPS1739 (a mutant in *motA*, a component of the membrane embedded flagellar motor, [234]) is also defective in attachment under all pH conditions (Figure 3-5, E). When comparing Figure 3-6 panels D and E, there is no significant difference between the OPS1738 (*flgE* mutant) and OPS1739 (*motA* mutant) attachment results at each pH condition. This is an interesting result, as it implies that it is lack of motility, rather than loss of flagellar adhesin action, which causes the defective attachment of the OPS1738 (*flgE*) mutant. This assumes that the *motA* mutant strain remains flagellated, as is the case for *motA* mutant *Salmonella* strains [235]. As discussed in [45, 61, 214, 215], the flagellum itself may possess adhesin properties, but it seems that this is either not the case in Rlv3841, or that adhesin properties cannot be attributed to the FlgE flagellar hook subunit, or that flagellar adhesin properties do not contribute to primary root attachment.

3.2.5.4 A *gmsA* mutant is impaired in attachment at pH 6.5 and 7.0 relative to wild-type

A very important result in validating the suitability of the whole-root Lux assay for pH studies was the comparison of D5250 and OPS0914 (a *gmsA* mutant, Figure 3-5, F). *gmsA* encodes glucomannan which is the only fully characterised primary attachment factor of known pH dependency. The results gained here, whereby a glucomannan mutant is impaired in attachment under pH 6.5 and 7.0 conditions, but not at pH 7.5, are in

agreement with the described model of primary root hair attachment in Rlv3841 [25, 34, 71].

The results presented in Figure 3-6 provide strong validation of the whole-root Lux assay as a method for characterising primary attachment factors under different pH conditions.

3.2.6 *praR* regulation of attachment is highly dependent on pH conditions

Two further strains, OPS0296 (*rapA2*/*rapC* double mutant) and OPS0111 (*praR* mutant) were also used to validate the Lux attachment assay under different pH conditions, but the results revealed an interesting pH dependency in *praR* regulation of attachment (Figure 3-7, Figure 3-8).

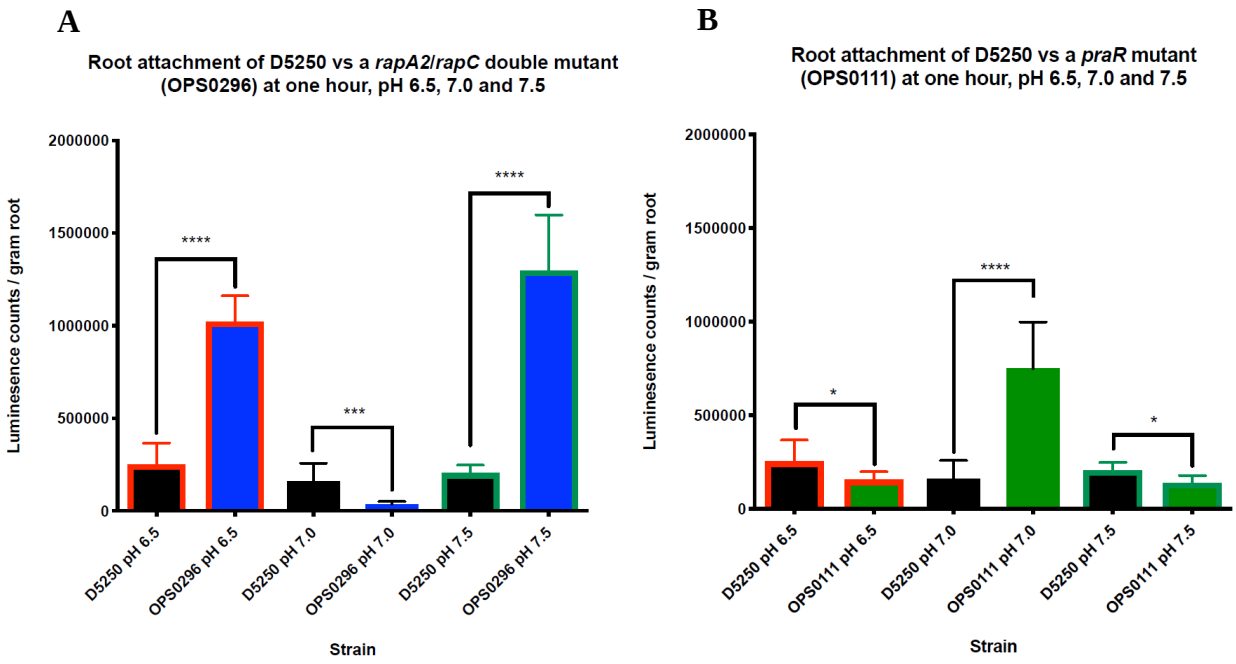


Figure 3-7. Comparison of Rlv3841and mutant strains under different pH conditions in a Lux whole root attachment assay. **A** – Root attachment of D5250 (Rlv3841[pIJ11282]) under pH 6.5, 7.0 and 7.5 conditions vs OPS0296 (*rapA2*/*rapC* mutant). **B** – D5250 vs

OPS0111 (*praR* mutant). Data is displayed as mean ± SEM, n=10. An unpaired t-test was used to compare groups. * = p <0.05, *** = p <0.0005, **** = p <0.0001

A *rapA2*/*rapC* double mutant strain in Rlv3841 has previously been characterised as deficient in attachment to pea roots [59]. Whilst this is seen here at pH 7.0 (Figure 3-7, A), the mutant strain shows a large increase in attachment at pH 6.5 and 7.5. This result was unexpected given the importance of these Raps at neutral pH, though the attachment phenotype of this strain under different pH conditions has not previously been reported. It suggests that RapA2 and RapC proteins interfere with binding promoted by other adhesins at acid and alkaline pH. New hypothetical protein adhesin RapX(s) might result in very strong binding at acid and alkaline pH, while RapA2 and RapC are most important for binding at neutral pH. Given the reported negative regulatory function of *praR* on the *rapA2* and *rapC* genes [59], a *praR* mutant was tested in the Lux attachment assay (Figure 3-7, B) to investigate this further. A *praR* mutant demonstrates the large expected increase in attachment relative to wild-type under pH 7.0 conditions (due to a loss of repression of *rapA2* and *rapC* genes, among others, [59]). However, under pH 6.5 and 7.5 mutation of *praR* results in a significant inhibition of binding, although this inhibitory effect is not as strong as its positive effect at neutral pH. This is intriguing, as it implies that, under pH 6.5 and 7.5 conditions, PraR promotes attachment via RapX(s). A triple mutant strain (a *praR*/*rapA2*/*rapC* mutant) would therefore show defective attachment at pH 7.0 due to the absence of RapA2/C. However, it may also have reduced attachment at pH 6.5 and 7.5 due to the loss of positive regulation of RapX(s). A triple mutant strain (OPS1239) was isolated (see 2.3.5) and conjugated with pIJ11282 to isolate OPS1266, which was tested in a Lux whole-root attachment assay (Figure 3-8).

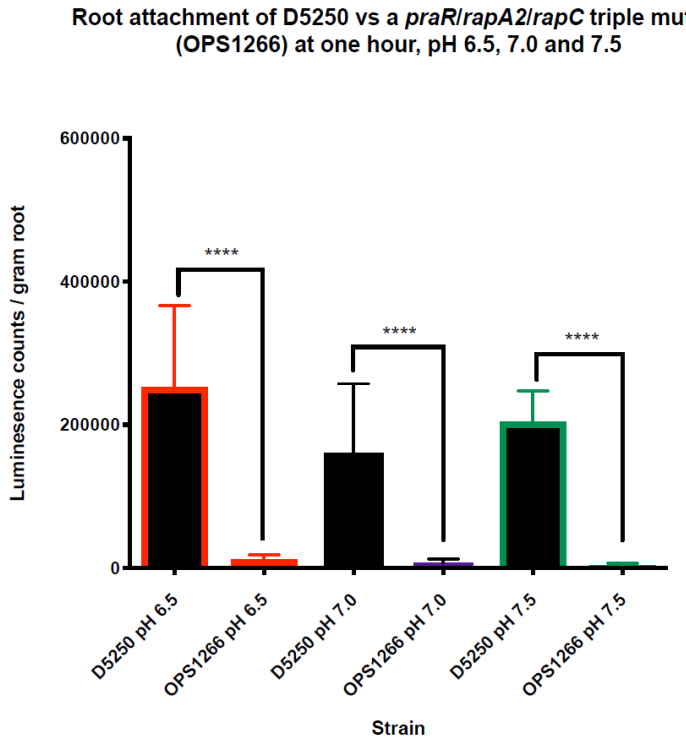


Figure 3-8. Comparison of D5250 (Rlv3841[pIJ11282]) and OPS1266 (*praR*/*rapA2*/*rapC* triple mutant) under different pH conditions in a Lux whole root attachment assay. Data is displayed as mean ± SEM, n=10. An unpaired t-test was used to compare groups. **** = p <0.0001

As shown in Figure 3-8, a *praR*/*rapA2*/*rapC* triple mutant strain is strongly impaired in attachment under all conditions compared to wild-type (D5250). This result is consistent with the multifaceted positive and negative regulatory activity of PraR [59].

praR itself is repressed in a population density dependent fashion by quorum sensing regulators. As population density increases, *cinR*, a regulator, induces the *cinIS* operon. CinS (an antirepressor) binds and inhibits PraR, thereby relieving PraR mediated repression of *rapA2* and *rapC* [59, 191, 236] and allowing attachment and biofilm

formation under neutral conditions. However, studies of PraR regulation have not been reported under different pH conditions. Based on the results presented in Figures 3-7 and 3-8 it may be that, at pH 6.5 and 7.5, PraR activity is not repressed, permitting the expression of unknown RapX factors, which are important for attachment at pH 6.5 and 7.5. Frederix *et al*. (2014) demonstrated that PraR acts as a positive as well as negative regulator [59], and it may be the case that some positively regulated targets of PraR (all of which are uncharacterised, [59]) may act as novel adhesin factors important for attachment at pH 6.5 or 7.5. Based on my results, and those of [59, 191, 236], a simple model of pH dependent PraR regulation of some aspects of primary attachment machinery can be hypothesised (Figure 3-9).

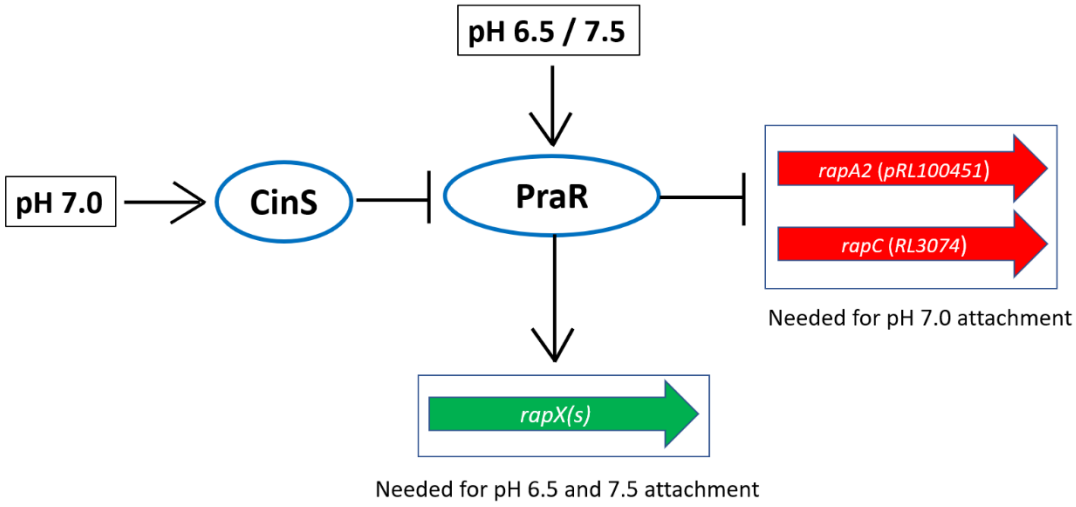


Figure 3-9. A simple model for PraR pH dependent regulation of some aspects of primary attachment machinery in Rlv3841. At pH 7.0, CinS (induced by CinR, not shown) inhibits PraR. This prevents repression of *rapA2* and *rapC*, which are important for attachment under pH 7.0 conditions. However, at pH 6.5 and 7.5, PraR remains active and activates the transcription of unknown and possibly plural '*rapX*' adherence factors which are important for attachment under non-neutral pH conditions.

Whilst this model provides a possible explanation for the pH dependencies of strain attachment (Figures 3-7, 3-8 and 3-9) it is likely that the regulatory mechanisms underlying these results are more complex. PraR is known to be a direct regulator of *rapA2* and *rapC* [59], but it is unknown whether other genes up or downregulated in a *praR* mutant (including possible '*rapX*' genes) are directly or indirectly regulated by PraR. In addition to *rapA2* and *rapC*, Rlv3841 has two further annotated *rap* genes: *rapA1* (*RL3660*) and *rapB* (*RL3911*). A *rapB* mutant strain shows no root attachment defects at neutral pH [59], but has not been tested in acid or alkaline conditions. Given the role of PraR in regulating *rap* genes, *rapB* represents a *rapX* candidate, and testing a *rapB* mutant for attachment at pH 6.5 and 7.5 could shed further light on PraR pH-dependent regulation of primary attachment. Additionally, there may be other uncharacterised regulatory aspects of quorum sensing and PraR regulation which act as relays of external conditions and permit the activation of different attachment mechanisms at different pHs.

3.2.7 Attempted mutation of possible rhicadhesin genes

The three most promising rhicadhesin candidates from the proteins identified from LC-MS/MS of a crude adhesin band were Omp19 (RL4441, 18 kDa), RL1635 (putative outer membrane protein, 19 kDa), and RL4733 (conserved hypothetical protein, 17 kDa). Unfortunately, despite several attempts, it did not prove possible in this work to isolate pK19mob mutants in the genes encoding these proteins. This could be because genes are essential for membrane stability or other important, uncharacterised functions, meaning that mutation is lethal. A second possibility is that small gene size may present a problem for interposon mutagenesis. Gene sizes are as follows: *omp19* (*RL4441*): 513 bp, *RL1635*:

531 bp, *RL4733*: 471 bp. It is important that an internal fragment of the target gene is cloned into the pK19mob vector, rather than the gene in its entirety, to ensure that interposon mutagenesis leads to a functional disruption of the gene. For this reason, it is normal to exclude ~100 bp from each end of the target gene when cloning an internal fragment. For these genes, the resulting internal fragment sizes were therefore in the range of 271-331 bp in length. Efficiency of interposon mutagenesis drops rapidly when the target gene internal fragment used for interposon mutagenesis is below 350 bp (Poole lab, internal communication). Such findings have also been reported elsewhere [237]. Note that these genes were successfully targeted by the *mariner* transposon for mutagenesis in INSeq, and attachment classifications are presented in Chapter 4, Table 4-17.

Having validated a Lux whole-root attachment assay suitable for pH studies and demonstrated the involvement of a broad range of factors in primary root attachment, it was therefore decided to undertake a bioinformatic screen for possible novel Rlv3841 attachment factors in order to continue this work.

3.2.8 Bioinformatic identification of possible novel root attachment factors

Genome scale datasets were analysed to identify likely novel root attachment factors. Data from one RNASeq experiment, six microarrays and two INSeq conditions (Table 3-2) were analysed by the two approaches described in Figure 3-10, resulting in the identification of 136 putative novel root attachment genes.

Table 3-2. RNASeq, microarray and INSeq experiments used in a bioinformatic screen for potential novel root attachment factors in Rlv3841.

Experiment	Source
1) RNASeq 7-day pea rhizosphere, 7-day post-inoculation	Unpublished data, this lab
2) 7-day pea rhizosphere, 1-day post-inoculation microarray [E-MEXP-2845, E-MEXP-2848]	[141]
3) 7-day pea rhizosphere, 3-day post-inoculation microarray [E-MEXP-2848]	[141]
4) 7-day pea rhizosphere, 7-day post-inoculation microarray [E-MEXP-2848, E-MEXP-2852, E-MEXP-2854]	[141]
5) 14-day pea rhizosphere, 1-day post-inoculation microarray [E-MEXP-2845]	[141]
6) 21-day pea rhizosphere, 1-day post-inoculation microarray [E-MEXP-2845]	[141]
7) 7-day pea rhizosphere inoculation, 21-day post-inoculation bacteroid microarray [E-MEXP-1918]	[139]
8) INSeq rhizosphere, 7-day pea rhizosphere inoculation, 5-day post-inoculation	[238]
9) INSeq root colonization, 7-day pea rhizosphere inoculation, 5-day post-inoculation	[238]

For microarrays, Array Express accession numbers for datasets are given in square brackets. Where two or more accession numbers are given, they form part of different time courses

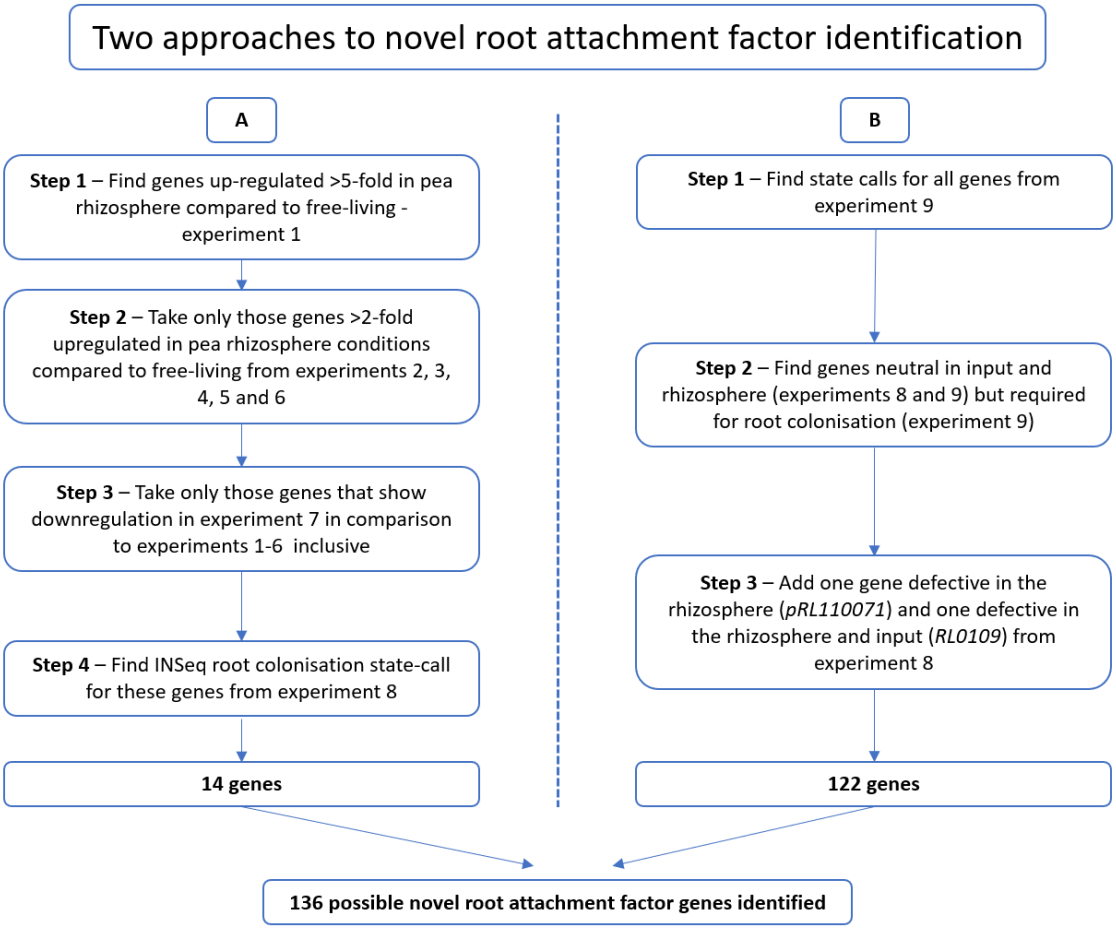


Figure 3-10. Flow charts of the two approaches (A and B) taken to identify bioinformatically novel root attachment factor genes in Rlv3841. Experiment numbers refer to those listed in Table 3-2.

The rationale behind approach A (Figure 3-10) was that genes involved in early stage root attachment should be consistently upregulated in the rhizosphere but subsequently downregulated in the bacteroid stage of symbiosis. The rationale behind approach B was that genes involved in early stage root attachment and/or colonization exclusively should have a neutral phenotype when mutated in INSeq input and rhizosphere libraries. One gene defective in the rhizosphere (*pRL110071*) and one defective in the rhizosphere and

input (*RL0109*) was added to the list of genes generated by approach B (Step 3). Testing mutants in these genes in a Lux assay will determine whether fitness reductions in input and rhizosphere libraries translate into reduced root attachment, as would be expected. All genes identified by approaches A and B (Figure 3-10) can be found in Tables A1 and A2 (Appendix 1) respectively.

3.2.9 Testing possible novel Rlv3841 adhesin factor mutants in Lux whole-root attachment assays

pK19mob interposon mutants were isolated in seven of the genes identified in section 3.3.8 and tested in Lux whole-root attachment assays. Details of these seven genes and the rationale for their choice are summarized in Table 3-3, and attachment assay results are shown in Figure 3-11.

Table 3-3. Summary of bioinformatically identified genes tested in Lux whole-root attachment assay

Strain	Gene	Protein description	Approach	Rationale for choice
OPS1131	*pRL110543*	conserved hypothetical	B	Essential for colonization, neutral in input and rhizosphere (experiments 8 and 9) – likely root attachment factor.
OPS1235	*RL2969*	putative transmembrane	A	Upregulated in the rhizosphere in experiments 1-6, downregulated in experiment 7 – rhizosphere induction indicates possible attachment role.
OPS1236	*RL4382*	putative filamentous hemagglutinin adherence factor precursor	B	Defective for colonization, neutral in input and rhizosphere (experiments 8 and 9) – likely root attachment factor.
OPS1237	*RL3273*	putative von Willebrand factor type A	A	Upregulated in the rhizosphere in experiments 1-6, downregulated in experiment 7 (Table 3-2). Rhizosphere induction indicates possible attachment role and annotation suggestive of role in attachment.
OPS1238	*RL0109*	Conserved hypothetical	B	Defective in input, rhizosphere and colonization, experiments 8 and 9. Testing will determine whether input and rhizosphere defects decrease primary root attachment.
OPS1290	*pRL110071*	Conserved hypothetical	B	Defective in rhizosphere and colonization, experiments 8 and 9. Testing will determine whether rhizosphere and colonization defects decrease primary root attachment.
OPS1291	*pRL100053*	Putative transmembrane	B	Essential for colonization, neutral in input and rhizosphere (experiments 8 and 9) - likely root attachment factor.

Approach refers to the two approaches (A and B) shown in Figure 3-10. Experiment numbers refer to experiments listed in Table 3-2.

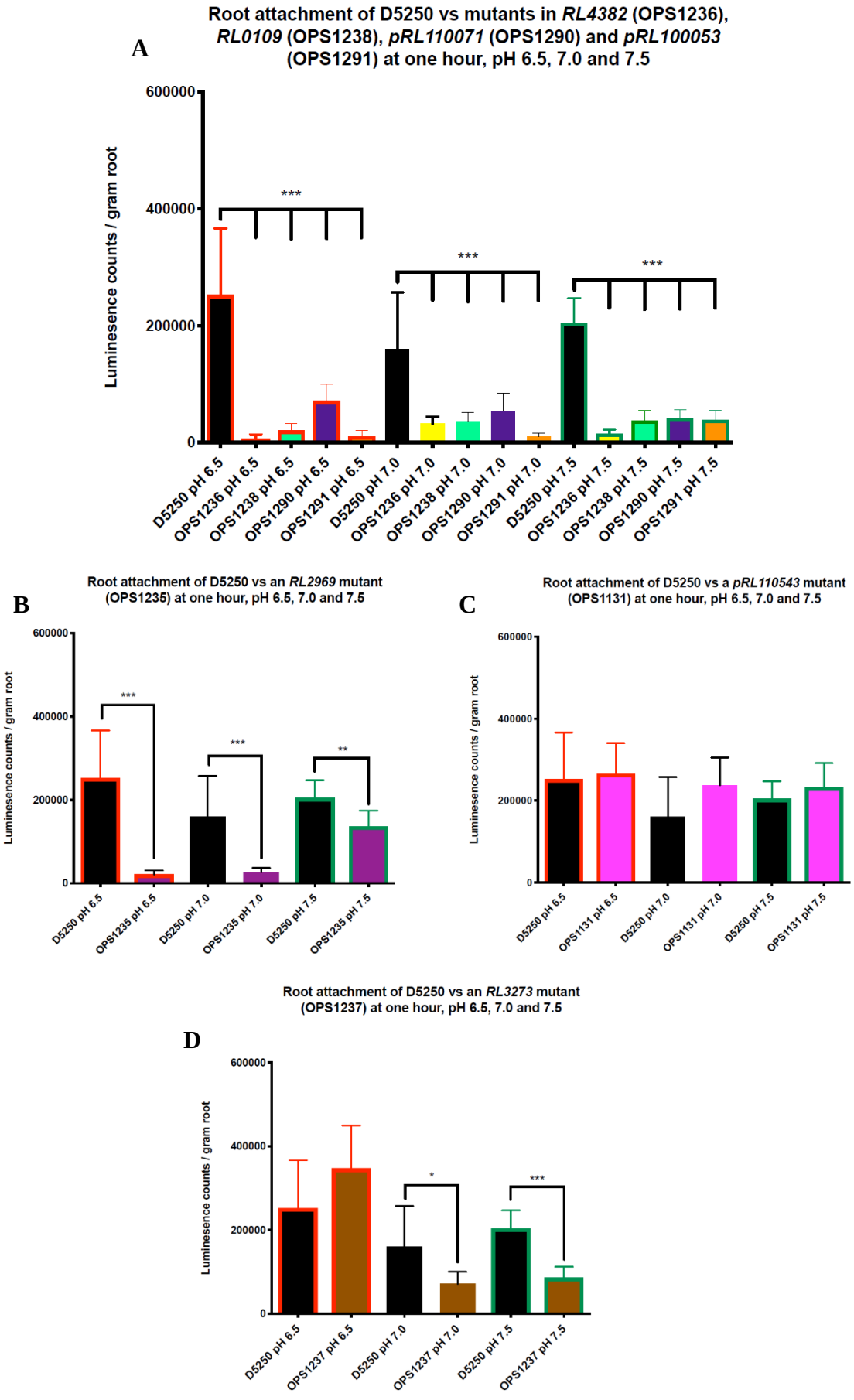

A
Root attachment of D5250 vs mutants in RL4382 (OPS1236), RL0109 (OPS1238), pRL110071 (OPS1290) and pRL100053 (OPS1291) at one hour, pH 6.5, 7.0 and 7.5
Luminescence counts / gram root
600000
400000
200000
0
***
***
***
D5250 pH 6.5
OPS1236 pH 6.5
OPS1238 pH 6.5
OPS1290 pH 6.5
OPS1291 pH 6.5
D5250 pH 7.0
OPS1236 pH 7.0
OPS1238 pH 7.0
OPS1290 pH 7.0
OPS1291 pH 7.0
D5250 pH 7.5
OPS1236 pH 7.5
OPS1238 pH 7.5
OPS1290 pH 7.5
OPS1291 pH 7.5
B
Root attachment of D5250 vs an RL2969 mutant (OPS1235) at one hour, pH 6.5, 7.0 and 7.5
Luminescence counts / gram root
600000
400000
200000
0
***
***
**
D5250 pH 6.5
OPS1235 pH 6.5
D5250 pH 7.0
OPS1235 pH 7.0
D5250 pH 7.5
OPS1235 pH 7.5
C
Root attachment of D5250 vs a pRL110543 mutant (OPS1131) at one hour, pH 6.5, 7.0 and 7.5
Luminescence counts / gram root
600000
400000
200000
0
D5250 pH 6.5
OPS1131 pH 6.5
D5250 pH 7.0
OPS1131 pH 7.0
D5250 pH 7.5
OPS1131 pH 7.5
D
Root attachment of D5250 vs an RL3273 mutant (OPS1237) at one hour, pH 6.5, 7.0 and 7.5
Luminescence counts / gram root
600000
400000
200000
0
*
***
D5250 pH 6.5
OPS1237 pH 6.5
D5250 pH 7.0
OPS1237 pH 7.0
D5250 pH 7.5
OPS1237 pH 7.5

Figure 3-11. Comparison of Rlv3841 and novel interposon mutant strains under different pH conditions in a Lux whole root attachment assay. A – Root attachment of D5250 (Rlv3841[pIJ11282]) under pH 6.5, 7.0 and 7.5 conditions vs OPS1236 (*RL4382* mutant), OPS1238 (*RL0109* mutant), OPS1290 (*pRL110071* mutant) and OPS1291 (*pRL100053* mutant). B – D5250 vs OPS1235 (*RL2969* mutant). C – D5250 vs OPS1131 (*pRL110543* mutant. D – D5250 vs OPS1237 (*RL3273* mutant). Strains and pH conditions are given on the x axes. Data is displayed as mean ± SEM, n=10. An unpaired t-test was used to compare groups. * = $p < 0.05$, ** = $p < 0.001$, *** = $p < 0.0005$

OPS1236 (*RL4382* mutant, Figure 3-11, A) is defective in root attachment under all pH conditions relative to D5250. *RL4382* encodes a putative filamentous hemagglutinin (FHA) adherence factor precursor. FHA is well described in *Bordetella pertussis* where it has been heavily used in acellular component vaccines [239]. In *Bordetella*, FHA can either be cell surface associated, or extracellular. It is translocated across the bacterial outer membrane by FhaC, a polypeptide transport associated (POTRA) domain containing protein and member of the two-partner secretion (TPS) pathway, common in gram-negative bacteria. From there it can be proteolytically processed and released, and is important for adherence to host cells and cell agglutination [239]. Intriguingly, it has been demonstrated that the plant pathogen *Xanthomonas axonopodis* pathovar *citri* requires a similar FHA for host plant tissue colonization [240]. Further, the rhizosphere colonising *P. putida* is known to use related proteins to colonise maize roots [241]. In Rlv3841, a neighbouring gene of FHA is *RL4381*, a putative outer membrane POTRA domain containing protein. These POTRA domains are often found in beta-barrel transporters [242]. The results presented in Figure 3-11, as well as the presence of co-localised gene

encoding a putative POTRA domain transporter, indicate that Rlv3841 also makes use of an FHA system in primary attachment to pea roots, which has not previously been documented.

Rlv3841 strains carrying mutations in *RL0109* (OPS1238), *pRL110071* (OPS1290), *pRL100053* (OPS1291) (Figure 3-11, A) and *RL2969* (OPS1235, Figure 3-11, B) were also defective in attachment under all pH conditions. Little is known about these genes; they encode a putative TIGRO2300 family protein, a beta-lactamase domain containing protein and two putative transmembrane domain containing proteins respectively, all of unknown function. A *pRL110071* mutant was shown to be defective in the rhizosphere (Table 3-2, experiments 8 and 9), whilst an *RL0109* mutant was defective in the rhizosphere and input in the same experiments. The defects in primary attachment shown for OPS1290 (*pRL110071* mutant strain) and OPS1238 (*RL0109* mutant strain) here could therefore result either from a reduced strain fitness in terms of growth, or from loss of a primary attachment factor which also plays a separate role in input and/or rhizosphere conditions. Overall, these results indicate that the bioinformatic approaches taken in 3.3.8 to identifying putative novel Rlv3841 adhesins is very successful.

OPS1131, a mutant in *pRL110543* (encoding a conserved uncharacterised coiled-coil domain containing protein), shows no statistical difference in attachment ability under any conditions compared to D5250 (Figure 3-11, C). Given this gene was essential for colonization but neutral in input and rhizosphere libraries in INSeq [238], it is most likely to represent a secondary attachment determinant.

Finally, OPS1237 (*RL3273* mutant, Figure 3-11, D) was defective in attachment under pH 7 and 7.5 conditions. This gene encodes a putative von Willebrand factor type A (VWA) domain containing protein. VWA domain proteins have been widely implicated in cell

adhesion and intracellular enzyme activity in eukaryotes, and are widely conserved [243]. Within rhizobia, *R. loti* has 12 VWA domain proteins (one magnesium chelatase and 11 uncharacterised) but *R. meililoti* has just three. Both Vanderlinde (2011) and Neudorf (2015) have linked mutations in *RL3501*, a VWA domain protein, to morphology defects and impaired cell envelope function in Rlv3841 [244, 245]. Thus, it may be that mutation in *RL3273* produces its negative effect on root attachment by direct disruption of an adhesin factor, or indirectly through cell envelope disruption, or by a further unknown mechanism, and that these effects are dependent on pH conditions.

Testing of mutants in putative novel adhesin factors has revealed the involvement of several in primary root attachment which had not previously been identified. This illustrates the usefulness of the Lux technique combined with genomic scale datasets for investigating such factors.

3.3 Conclusion

In summary, a crude adhesin fraction isolated from Rlv3841 was able to inhibit attachment when preincubated with root sections. The 14 kDa protein band thought to represent this adhesin was found to be made up of 15 robustly identified protein components, although these were unable to be characterised further at this stage due to difficulties in isolating mutant strains in the encoding genes. Particularly for the three most likely rhicadhesin candidates (*RL4733*, *omp19 – RL4441* and *RL1635*) an in-frame mutagenesis strategy should be applied (such as the double recombination approach described by Link *et al.* (1997)) to investigate these further [246]. This is not limited by gene size, unlike the pK19mob insertion technique used in this study.

However, regarding the rhicadhesin hypothesis, the definition of rhicadhesin based on ability to inhibit attachment [53] may be a flaw in its identification as an adhesin. The ability of a protein mixture to block attachment does not necessarily imply a 'lock and key' type interaction where rhicadhesin occupies plant root binding sites. Rather, interference with the plant root or bacterial surface (surface protein aggregation, for example) or some other indirect effect could be responsible. As an example of this, in one case it was demonstrated that added rhicadhesin could rescue attachment of an *A. tumefaciens chvB* mutant to pea root hairs [54]. However, as discussed by Matthysse (2014), it is likely that the effects of *chvA* or *chvB* mutations are indirect and result from multiple defects caused by the absence of cyclic-β-1,2-D-glucan from the periplasmic space rather than the absence of rhicadhesin [55].

Although proteomics results do not conclusively identify rhicadhesin, the candidate root attachment factors identified bioinformatically (Appendix 1 Tables 1 and 2) are informative. Two approaches (Figure 3-10) led to the identification of 136 possible novel root attachment genes. Testing mutants in seven of these revealed that six showed defects in attachment compared to wild-type Rlv3841. Given how many root attachment factors there are likely to be in Rlv3841, preparation of a membrane fraction, even if it is focussed around 14 kDa proteins, may isolate one or more proteins capable of inhibiting root attachment when pre-incubated with roots. However, this does not indicate that a protein with this property is an important adhesin. Although it may have adhesin properties and be directly responsible for attachment of bacterial cells to root surfaces, it may also have an indirect mode of action. In this case, it could alter the bacterial cell surface in such a way that it alters the presentation, structure or activity of another factor which is a direct

adhesin. This distinction was never thoroughly investigated in the literature reports of rhicadhesin [53, 54, 208, 211].

Whilst a singular rhicadhesin may exist, it is also possible that there are multiple functional genes encoding rhicadhesin, or that it is expressed as a larger precursor and subsequently processed. These would be confounding factors in the identification of a rhicadhesin gene. Combined with the caveats of previous research, the identification of rhicadhesin is extremely difficult. A further complication is that Smit *et al*. (1989) never demonstrated conclusively that their purified rhicadhesin was a homogenous single protein [53]. Although their purified rhicadhesin ran as a ~14 kDa band on an SDS-PAGE gel, it could still have been a mixture of similar-sized proteins, further confounding the definition of a rhicadhesin. Therefore, whilst a rhicadhesin with the properties described by Smit *et al.* (1989) may exist, there are also multiple alternative explanations as to how an isolated protein fraction could inhibit bacterial root attachment, and there are likely to be many different factors that could be rhicadhesin candidates.

To investigate bacterial root attachment further, a luminescence-based whole root attachment assay was developed and validated through comparison with a bacterial colony counting method and by testing with known attachment factor mutants. Crucially, the assay was able to demonstrate the known pH dependent profile of a glucomannan mutant in attachment. This work also highlighted the importance of motility in primary root attachment, and indicated that it is motility of strains, rather than flagellar adhesin properties, that is important for this process.

Whilst mirroring the characterised attachment phenotype of a *rapA2*/*rapC* double mutant and a *praR* mutant, this assay also shed new light on the complexities and pH dependency of PraR regulation. By investigating the attachment phenotype of a *rapA2*/*rapC*/*praR*

triple mutant strain, additions to the known PraR regulatory circuit were able to be proposed which may account for the unexpected attachment phenotypes seen.

Using a bioinformatic approach to investigate genome-scale RNASeq, microarray and INSeq datasets, 136 putative novel root attachment factor genes were identified in Rlv3841. Testing of pK19mob mutants in seven of these genes in a Lux assay revealed that six showed reduced attachment under two or more pH conditions in comparison to wild-type. This indicates that the bioinformatic approaches taken to identifying putative novel Rlv3841 attachment factors was very successful. Further, this testing implicated genes and factors previously unreported in the primary root attachment process, including FHA, a VWA domain protein and multiple uncharacterised proteins.

Given these findings, and the low-throughput nature of individual pK19mob mutant isolation and testing, this chapter provides justification for taking a genome-scale approach to characterising primary Rlv3841 attachment determinants at different pHs. INSeq (Chapters 4 and 5) is a powerful method for such investigation and provides a further experimental avenue by which to investigate the possible identity of rhicadhesin.

Chapter 4

Genome-scale characterisation of the primary attachment determinants in the *R. leguminosarum* symbiosis under acid, neutral and alkaline pH conditions

4.1 Introduction

To attach to the roots of host legume plants rhizobia use multiple mechanisms. The earliest stages of physical interactions between bacteria and plant roots rely on universal, non-specific binding forces [34]. These include Van der Waals forces (caused by momentary changes in molecular electron density [247]), electrostatic interactions (caused by differences in molecular electric charge [44]) and hydrophobic interactions (the strongest of these universal interactions [46, 248, 249]) [34].

Following this, soil bacteria exhibit a range of different specific molecular mechanisms of primary root attachment. These include adhesive pili (*Agrobacterium tumefaciens*, [250]), flagella and outer membrane proteins (*Azospirillum* spp., [60, 89]), pili and outer membrane porins (*Pseudomonas* spp., [67, 214]) and flagella and proteinaceous fimbriae (*S. enterica*, [69]).

For rhizobia, primary root attachment factors include not only the glucomannan system, important for attachment to root hairs at acidic pHs [56, 71], but also the proposed rhicadhesin system [53, 210], thought to be required for root hair attachment at alkaline pHs. Rlv3841 uses additional attachment factors important for primary attachment to root hairs and/or bulk root. Characterised examples include *Rhizobium* adhering proteins (Raps) [57, 59], acidic exopolysaccharides [56] and PraR regulated genes [59]. The work presented in Chapter 3 developed a new method for assaying primary attachment and provided evidence for the involvement of many factors that have yet to be included in primary attachment models (see 3.2.9). Successful primary root attachment is important for both nitrogen-fixing symbiosis formation with legumes [56], but also accessing root exudates, which can serve as growth substrates [34]. Therefore, a detailed understanding of primary root attachment will improve understanding of the early nitrogen-fixing

symbioses stages and persistence of rhizobia in the soil. Increased understanding of these processes is timely. Soil degradation as result of intensive agriculture [1] and climate change induced acidification [251] may alter the balance of primary root attachment factors relied on by rhizobia. This could be important for agricultural inoculant formation. Further, development of synthetic symbioses seeking to engineer root-associative interactions with nitrogen-fixing bacteria to boost crop yields [35, 109] could leverage the genetic circuitry of primary root attachment mechanisms to increase effectiveness.

The presence of diverse primary root attachment mechanisms in rhizobia is not surprising considering the heterogeneity of the soil environment and the diversity of plant roots encountered. A single gram of bulk soil can contain 10^8 bacterial cells [252] and varies widely in soil particle size, nutrient concentration and pH [253–256]. Competition for plant root attachment is intense and can involve the interaction of plant beneficial rhizobia and other soil bacteria, pathogenic bacteria, mycorrhizal fungi and many more [257–260]. Additionally, plant roots may also undergo physical changes in different soil conditions which necessitate adaptation by attaching bacteria. One example of this is the alkaline pH-induced dissociation of plant root hair lectin, thought to prevent glucomannan-mediated attachment of Rlv3841 [25, 47, 60].

The complexity of the soil growth environment is reflected in the genome architecture of Rlv3841. The 7.75 Mb genome consists of a circular chromosome (4800 genes) and six plasmids: pRL7 (186 genes), pRL8 (142 genes), pRL9 (313 genes), pRL10 (471 genes), pRL11 (644 genes) and pRL12 (790 genes) [138, 151]. This is considerably larger than the typical bacterial genome, which is around 5 Mb and encodes ~5,000 proteins [261]. Various ‘omics studies have indicated a plethora of gene functions in Rlv3841 ([139, 141, 143] as examples). However, despite this, ~25% of Rlv3841 genes remain

uncharacterised. Although some gene functions have been inferred from regulatory profiles in transcriptomics [139, 141], this does not identify the importance or role of genes under specific conditions, such as primary root attachment.

Mariner transposon insertion sequencing (INSeq) is a technique that overcomes these limitations. This method allows the effect of mutation on individual genes to be analyzed under conditions of interest using libraries of single insertion bacterial mutants coupled with high-throughput sequencing [169]. By combining sequencing data with a hidden Markov model (HMM), the importance of individual genes in tested conditions can be ascertained.

A benefit of this technique in relation to Tn5 transposon insertion screens is the specific targeting by the *mariner* transposon of thymine-adenine (TA) motifs [176]. This defined insertion site requirement allows *in silico* prediction of insertion site density in any genome sequence and robust statistical analysis based on a defined number of insertion sites [151]. The HMM used here expands on the binary essential/non-essential classifications of other statistical methods by assigning each gene to one of four classification states. These are based on the inferred fitness effect on a bacterium of mutation to a given gene and are as follows: essential (ES, the bacterium is unable to tolerate insertion), defective (DE, insertion impairs fitness of the bacterium), neutral (NE, insertion has no fitness effect) and advantaged (AD, insertion increases fitness of the bacterium). Successful application of an HMM to transposon insertion mutagenesis screens has been reported in a variety of bacterial species [183, 262, 263]. Figure 4-1 summarizes how an INSeq experiment and HMM gene assignments can give information on the bacterial genes involved in primary attachment to roots.

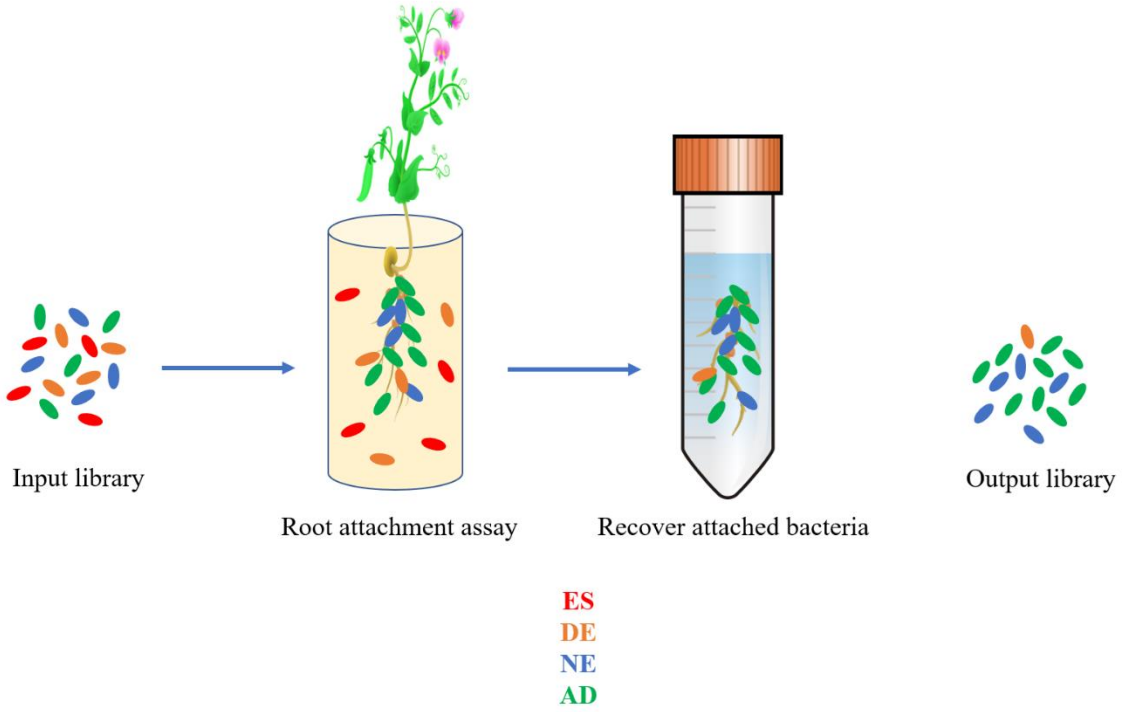


Figure 4-1. Using INSeq and HMM-assigned gene classifications to investigate bacterial root attachment. Mutated bacteria containing a single transposon insertion are represented as ovals where color indicates the HMM classification for root attachment. Red = ES (essential, mutants do not attach to roots and are lost from the total population), orange = DE (defective, mutants reduce in frequency in the population), blue = NE (neutral, no change in the frequency of the mutants), green = AD (advantaged, mutants increase in frequency in the total population).

The first use of INSeq in rhizobia was reported by Perry and Yost (2014), who adapted the INSeq mutagenesis vector from Goodman *et al*. (2011) [169] with a Rhizobiaceae specific promoter driving transposase activity and an *nptII* neomycin resistance cassette. This resulted in the pSAM_Rl *mariner* transposon mutagenesis vector (see Table 2-1 and 2-2), used to investigate gene requirements for Rlv3841 growth on TY media [151]. Wheatley *et al*. (2017) also used pSAM_Rl INSeq in Rlv3841 to characterise the influence of

atmospheric and 1 % oxygen concentrations on growth with glucose or succinate as a carbon source [143].

Here, pSAM_Rl is used in an INSeq experiment to characterise the primary attachment determinants of Rlv3841 to pea roots at three different pHs: 6.5, 7.0 and 7.5. This demonstrates that this process is far more complex than previously reported and uncovers multiple novel mechanisms involved in attachment.

The raw data for this INSeq experiment can be found in Appendix 2, Table 1.

4.2 Results and discussion

4.2.1 Root attachment assays – determining inoculum density and bacterial recovery method for INSeq

Root attachment assays with Rlv3841 (2.6.5) were carried out to determine the optimum starting inoculum density for INSeq analysis after recovery of root-attached bacteria (Figure 4-2).

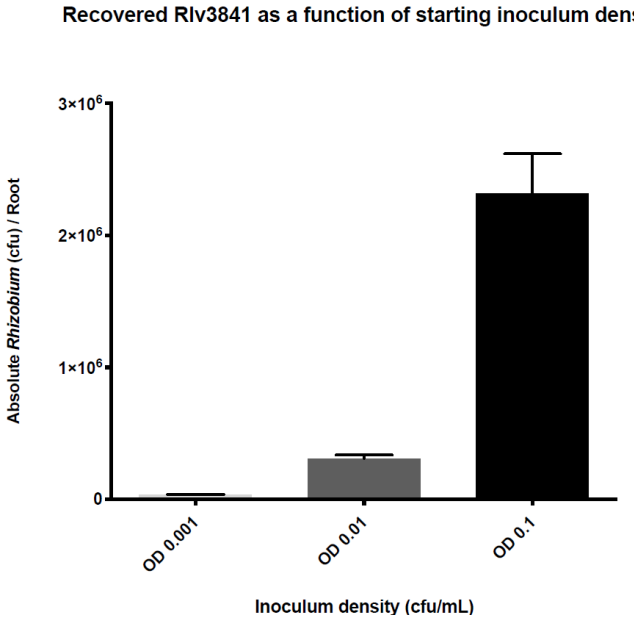


Figure 4-2. Recovered Rlv3841 as a function of starting inoculum density after 1 hr attachment to pea roots at pH 7.0. Inoculum density is given as OD_{600}. n = 5 for all groups.

As demonstrated in Figure 4-2, the higher the starting inoculum density of Rlv3841, the greater the number of Rlv3841 cells recovered from roots after 1 hr attachment. For statistically robust HMM state assignments for genes in an INSeq experiment, it is desirable to have at least 500-750-fold coverage of the genome represented in the output

library. The number of plants required to achieve this for a given starting inoculum density can be calculated using the data shown in Figure 4-2 and the following equation:

$$Pn = Fc \div \left(\frac{\bar{x}}{n}\right)$$

where *Pn* is the number of plants needed for the desired fold coverage, *Fc* is the desired fold coverage, $\bar{x}$ is the mean of the number of root-attached bacteria and *n* is the number of genes in the genome.

For the starting inoculum densities shown in Figure 4-2, Table 4-1 provides the number of plants needed for 500 or 750-fold coverage using the equation above.

Table 4-1. Number of plants needed for 500- or 750-fold coverage of 7,300 genes in the genome of Rlv3841 based on input inoculum density (OD_{600})

Input inoculum (OD_{600})	**Plants needed for 500-fold coverage of 7,300 genes**	**Plants needed for 750-fold coverage of 7,300 genes**
0.001	107	161
0.01	12	18
0.1	2	3

Note that, for Rlv3841, an OD_{600} of 0.1 = 10^8 cfu/mL (data not shown). Given that 500-750-fold coverage of a genome is considered a minimum requirement for robust statistical analysis, and that Lux attachment assays were conducted with an inoculum of OD_{600} 0.1, it was decided to use a starting inoculum of OD_{600} 0.1 with 10 plants per test condition.

Each test condition was performed in triplicate (see 4.2.2). This starting inoculum density and plant number should allow for > 3,000 fold coverage of the ~7,300 Rlv3841 genes (and >165-fold coverage of the 140,056 TA sites).

Primary attachment is defined by relatively weak and reversible interactions between bacteria and plant roots. Secondary attachment is defined by a more 'irreversible' interaction [34, 264, 265], meaning that it is difficult to remove bacteria from roots. Any INSeq assay seeking to define gene roles in root attachment must recover as much of the root-attached bacterial population as possible or risk mis-classifying genes. To this end, the numbers of Rlv3841 that could be recovered from roots after a 1 hr attachment assay by either vortexing alone, or by vortexing and grinding roots with a pestle and mortar, were compared (see 2.6.5, Figure 4-3).

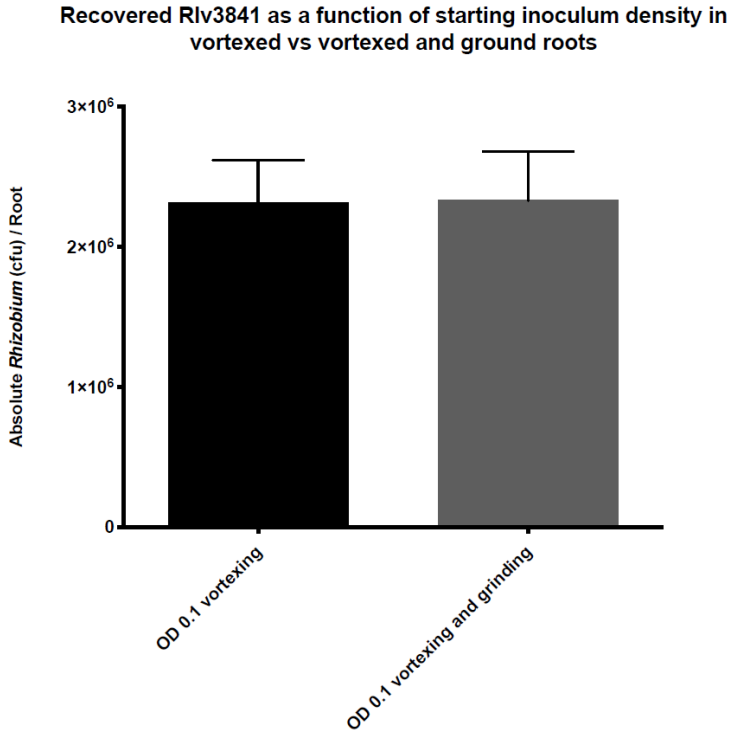


Figure 4-3. Rlv3841 recovered by vortexing alone or vortexing and grinding using a starting inoculum density of OD_{600} 0.1 in a 1 hr attachment assay to pea roots at pH 7.0. n = 5 for all groups

No statistically significant difference was seen between the number of bacteria recovered by vortexing alone compared to vortexing and grinding of roots (Figure 4-3). It can therefore be concluded that vortexing alone is sufficient to recover attached Rlv3841 cells in this INSeq root attachment assay.

4.2.2 INSeq experimental design

To characterise primary attachment determinants at different pH conditions, an INSeq experiment was designed (Figure 4-4). A population of insertion mutants (library 1-3, 500 mL of an OD_{600} 0.1 population) was used to inoculate 10 pea plants for each test condition (pH 6.5, 7.0 and 7.5), resulting in each being performed in triplicate.

Insertion mutants were recovered from roots after 1 hr and grown in liquid TY for 12 hr before the 12 sample libraries (3x3 experimental conditions + 3 input libraries) underwent barcoding, *mariner* library preparation and sequencing (2.7.3, Figure 4-4).

HMM gene classifications were averaged across the three replicates to increase robustness of analysis.

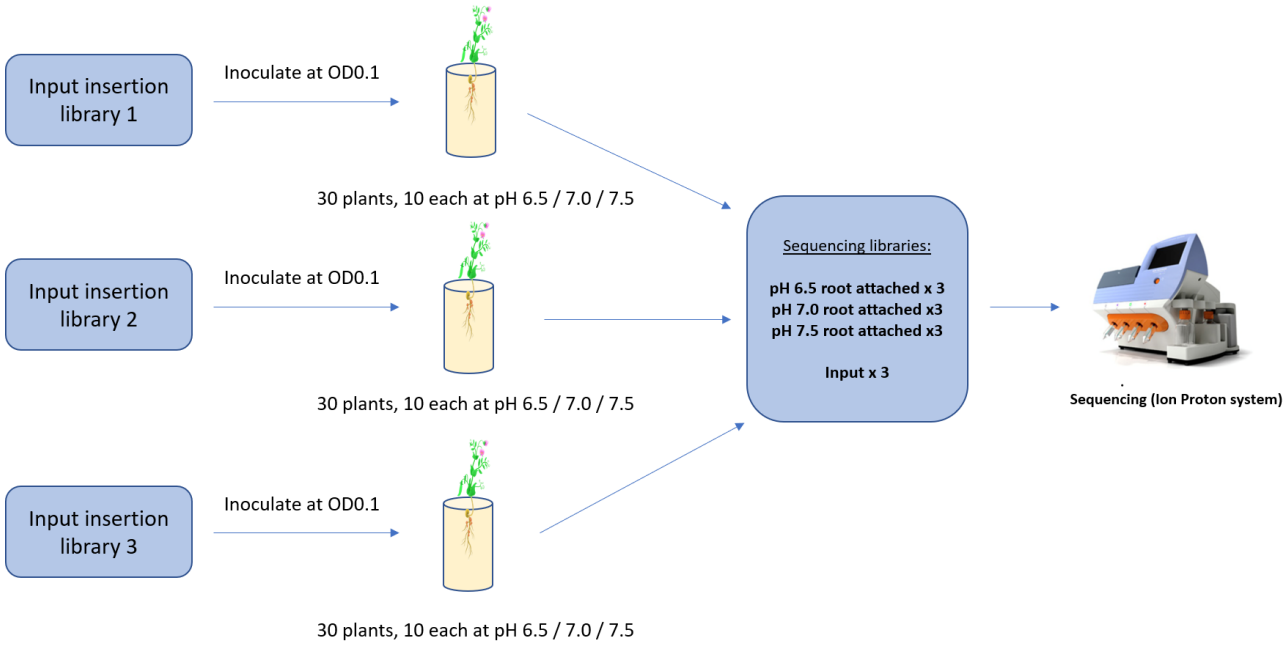


Figure 4-4. INSeq experimental design. 50 mL of insertion mutant inoculum (library 1-3, OD_{600} 0.1) was inoculated onto each pea plant root at pH 6.5, 7.0 or 7.5 and incubated for 1 hr before recovery by vortexing (resulting in three replicates). Recovered bacteria for each attachment group were pelleted and grown on TY agar for 12 hr to increase bacterial gDNA concentration (and effectively decrease plant DNA contamination, see 4.2.4 for further discussion) before DNA preparation (not shown). Input library gDNA samples were extracted directly from input inoculum. Each sample (input and 3 x root-attached) underwent DNA extraction and library preparation before sequencing (see 2.7.2, 2.7.3 and 2.7.4).

4.2.3 HMM analysis of INSeq data

HMM analysis of sequencing reads assigned 7,319 genes (99.7%) in the Rlv3841 genome to one of the following categories, indicating how their mutation affects bacterium fitness: ES, DE, NE or AD (see Figure 4-1). The 0.3% of genes (21) unrepresented did not contain TA sites [151], and could not be analyzed using INSeq. In total, 87 million barcoded

sequencing reads were obtained from a total of twelve samples (nine from plants and three input libraries).

Isolation of *mariner* libraries produced highly saturated mutant pools. The input libraries were shown by sequencing to have an average insertion density (percentage of TA motif insertion sites that carry one or more insertions) of 82% (Table 4-2). The greater the insertion density in the input library, the higher the coverage of TA motif insertion sites. Insertion densities <25% are considered insufficient for robust statistical analysis [146, 150, 169, 179]. Note that for the chromosome or any plasmids individually the insertion density obtained did not differ more than ± 5% from the average input figure given in Table 4-2, indicating unbiased *mariner* insertions across the genome. In this INSeq experiment, it was important that sufficiently large mutant libraries could be recovered from roots so that HMM classifications of genes could be compared between the input and root-attached samples. The insertion densities obtained for the root-attached libraries (Table 4-2) provides early validation of the inoculum density parameters chosen for this experiment.

Table 4-2. Percentage insertion density in the input and root-attached (pH 6.5, 7.0 and 7.5) libraries.

Library	Insertion density
Input	82 %
pH 6.5	48 %
pH 7.0	50 %
pH 7.5	50%

Insertion density is defined as the proportion of total TA motifs that contain at least one insertion in each sequenced library.

4.2.4 INSeq gene classifications

The distribution of HMM gene classifications across the four different states (ES, DE, NE and AD) in input and root-attached libraries is shown in Table 4-3, below

Table 4-3. Distribution of HMM assignment of Rlv3841 genes in input and root-attached (pH 6.5, 7.0 and 7.5) libraries.

HMM assignment	**Library**			
	Input	**pH 6.5**	**pH 7.0**	**pH 7.5**
ES	1 %	6 %	7 %	5 %
DE	10 %	7 %	6 %	7 %
NE	86 %	86 %	86 %	87 %
AD	3 %	1 %	1 %	1 %

Values are given as percentage of the 7,319 genes which contain TA sites.

Additional processing of the input library data was necessary to allow valid comparison between the input and root-attached libraries because of the 12 hr TY growth step for the bacteria recovered from roots. This step increases bacterial gDNA concentration and dilutes contaminating plant DNA before extraction (see 2.7.2). TY was chosen for this as it has been shown, following bacterial centrifugation and resuspension, to promote rapid growth of Rlv3841 (Poole Lab, data not shown). Regrowth in TY, even though it was for

a relatively short time (12 hr, ~ 4 generations for Rlv3841) could lead to an inadvertent selection pressure on mutants and affect downstream HMM classifications.

To prevent this, input library data was curated post-HMM processing to remove any mutations that result in a non-NE gene classification in the HMM when strains are grown on TY. At the outstet, a total of 954 genes were classified as non-NE (meaning ES, DE or AD) in the input library (UMA media) (2.2.1). Perry and Yost (2014) reported 759 genes which were non-NE for growth on TY. Of these, 540 shared a non-NE classification in the input library and 219 were uniquely non-NE on TY. Taking these 219 genes into account (where HMM classification could be skewed due to the TY growth step), 14% of the genome was classified as non-NE in the input library.

When using INSeq to identify primary root-attachment determinants, it is genes that are NE in the input but non-NE in attachment which are of interest.

4.2.5 Validation of INSeq predictions

Initial validation of these INSeq results was provided by comparing the HMM classifications of well-characterised genes with their known or predicted phenotypes from previous relevant literature (Table 4-4)

Table 4-4. S**ummary of HMM classifications of well-characterised genes with known or expected phenotypes.**

Gene(s)	INSeq classification	Description and references
murB, -C, -G, -D, -F, -E (RL3305, -06, -07, -09, -11, -12, respectively)	DE or ES - all conditions	Peptidoglycan biosynthesis genes; key structural role in the cell wall, of critical importance to cell viability [266]
RL1769-99 (50S and 30S ribosomal subunits)	DE or ES - all conditions	Ribosomal subunit proteins are essential for protein synthesis and cellular viability [267]
repABC modules of all plasmids	DE or ES – all conditions	RepABC systems maintain rhizobial plasmids in daughter cells; previous work has identified ES genes on all plasmids; plasmid maintenance should be essential for cell function [151]
kdtA (*RL0902*)	DE – all conditions	KdtA catalyzes the transfer of 3-deoxy-D-manno-octulosonate (Kdo) to the precursor of Lipid A, the anchor of LPS. The Lipid A Kdo domain is required for growth [268]
dnaC (*RL1551*) and *dnaX* (*RL0134*)	DE or ES – all conditions	DnaC is a putative DNA helicase and DnaX a putative polymerase, essential for DNA replication [269]

sufS (*RL2578*), *sufC* (*RL2580*), *sufB* (*RL2582*)	DE or ES – all conditions	Iron sulphur biogenesis and transport related proteins, shown to have a lethal phenotype on mutation [192]
aapJ (*RL2204*)	NE – all conditions	Putative solute binding protein of PAAT L-amino acid transporter, mutation has non-lethal phenotype [270]
mntH (*RL0940*)	NE – all conditions	Putative manganese transporter, mutant has comparable growth rate to Rlv3841 in non-manganese limited conditions [271]

All conditions = input, attachment at pH 6.5, 7.0 and 7.5

Various cellular components are critical for viability, including peptidoglycan (the main shape-maintaining element of the bacterial cell wall) [266] and ribosomal subunits (required for protein synthesis) [267]. In addition, RepABC for each plasmid, KdtA and DNA separation and replication proteins are also required [151, 268, 269]. Fe-S clusters are used as cofactors by various proteins, with Suf proteins involved in oxidative stress responses. Rlv3841 has only one Suf system, so any mutation is deleterious [192]. Given the importance of these cell components, it would be expected that mutants in any of these factors would show reduced fitness in INSeq, which is indeed the case (Table 4-4). Conversely, mutations in the transporters *aapJ* and *mntH* (involved in the transport of polar amino acids and manganese ions, respectively) are not deleterious under free-living conditions, and would be expected to be NE in INSeq, which is what was found in these experiments (Table 4-4).

Further to this literature validation, HMM classifications of known secondary attachment factors were also compared with literature. In this case it was hoped that established secondary attachment factors would not be affected in this INSeq experiment to discover primary attachment factors. This would show that the two stages can be separated experimentally. Cellulose fibril deposition is a characterised secondary attachment process in *R. leguminosarum*, enabling root colonization [71]. The genes encoding cellulose synthesis regulator PleD (*RL1730*, also known as CelR) as well as cellulose synthase CelABC (*RL1646-48*) were classified as NE in INSeq assessing primary attachment at all pHs, as were RhiABC genes (*pRL100169-71*) (Appendix 2, Table 1). The *rhi* genes, named for being strongly rhizosphere-induced, have been implicated in secondary root attachment and are required for successful five-day colonization of pea roots [141, 238]. The NE classification of these known secondary attachment factors provides validation of the sampling (1 hr post-inoculation) as suitable for investigating primary attachment.

4.2.6 Primary attachment gene requirements and functional classifications

After filtering genes which were assigned a non-NE HMM classification in the input library, 6146 genes remained. Some genes were required (ES/DE) for attachment at discrete pH conditions with others required at multiple pHs. In total, 292 genes classified NE in the curated input library were classified as ES/DE in one or more of the root-attached libraries; the distribution of these is shown in Figure 4-5. These genes and their HMM classifications from experiments listed in Table 4-5 and 4-7 are given in Appendix 1 Table A3 (all pHs), Table A4 (pH 6.5), Table A5 (pH 7.0), Table A6 (pH 7.5), Table A7 (pH 6.5 and 7.0), Table A8 (pH 7.0 and 7.5) and Table A9 (pH 6.5 and 7.5).

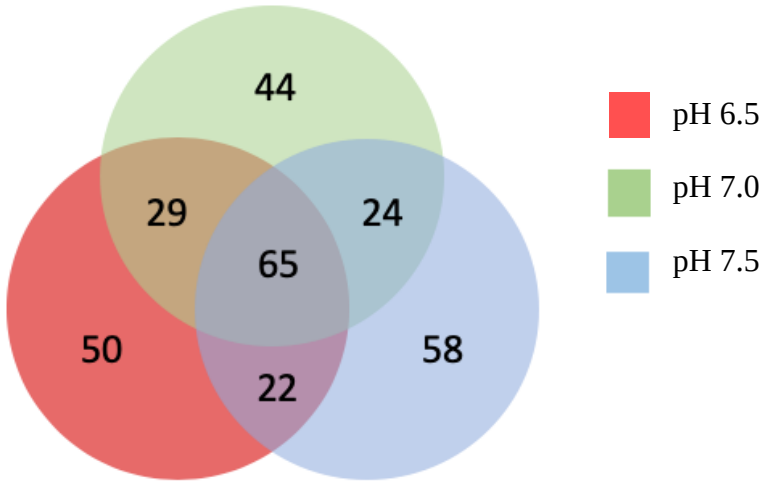


Figure 4-5. Genes classified as NE in the input library and required (ES/DE) in one or more of the root-attached libraries. Total genes = 292. Venn diagram circle color indicates pH; red = pH 6.5, green = pH 7.0, blue = pH 7.5.

Thus far, the only described primary attachment determinants with a pH-dependent profile of activity are glucomannan [56] and the hypothesized rhicadhesin [34, 53, 56, 71, 208, 210, 211]. Given that there are large numbers of genes specifically required at different pHs (50, 44 and 58 at pH 6.5, 7.0 and 7.5, respectively, Figure 4-5), this is likely to be a vast underrepresentation of the complexity of primary root attachment. Differences in plant root surfaces at different pHs, exemplified by the dissociation of root hair lectins at alkaline pH, [25, 47, 60], may at least partially account for this. However, it appears that there are many unexplored factors involved.

The functional classifications of these 292 genes (Figure 4-4) were investigated, and comparison made with those genes required (ES/DE) in the input library. Riley codes (classifying genes into one of seven broad functional categories based on their putative function [138, 272]) are useful for this purpose. Figure 4-6 shows the Riley classification

of the 1,173 genes required (ES/DE) in the input library compared with that of the 65 genes required (ES/DE) for attachment at all pHs.

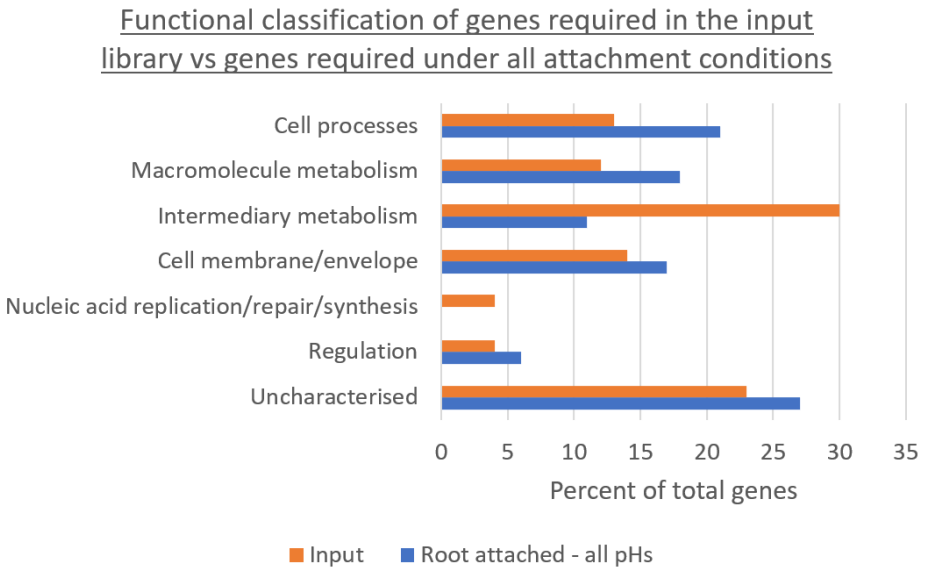


Figure 4-6. Functional classification (using Riley codes) of the 1,173 genes required (ES/DE) in the input library compared with the 65 genes required (ES/DE) for primary root attachment at pH 6.5, 7.0 and 7.5. Data is presented as percentage of total genes in each category.

Functional classifications (Figure 4-6) reveal that the most abundant ES/DE gene class in the input library is intermediary metabolism, implying that metabolic pathway functions are the most important determinants in 'free-living' growth. In contrast, the 65 genes ES/DE under all root attachment conditions show higher representation of cell process, cell membrane/envelope and macromolecule metabolism, as well as uncharacterised

genes, indicating these classes are the most importance in successful primary root attachment.

4.2.7 Genomic localization of genes required for primary root attachment

Further insight can be gained by comparing the genetic loci of genes required (ES/DE) for primary attachment (this work) and secondary attachment (5 d colonization of pea roots, [238]). The experiments of Wheatley (2018) [238] were at neutral pH, so comparison was made to 162 genes required (ES/DE) for primary attachment at pH 7.0 (shown in Figure 4-7).

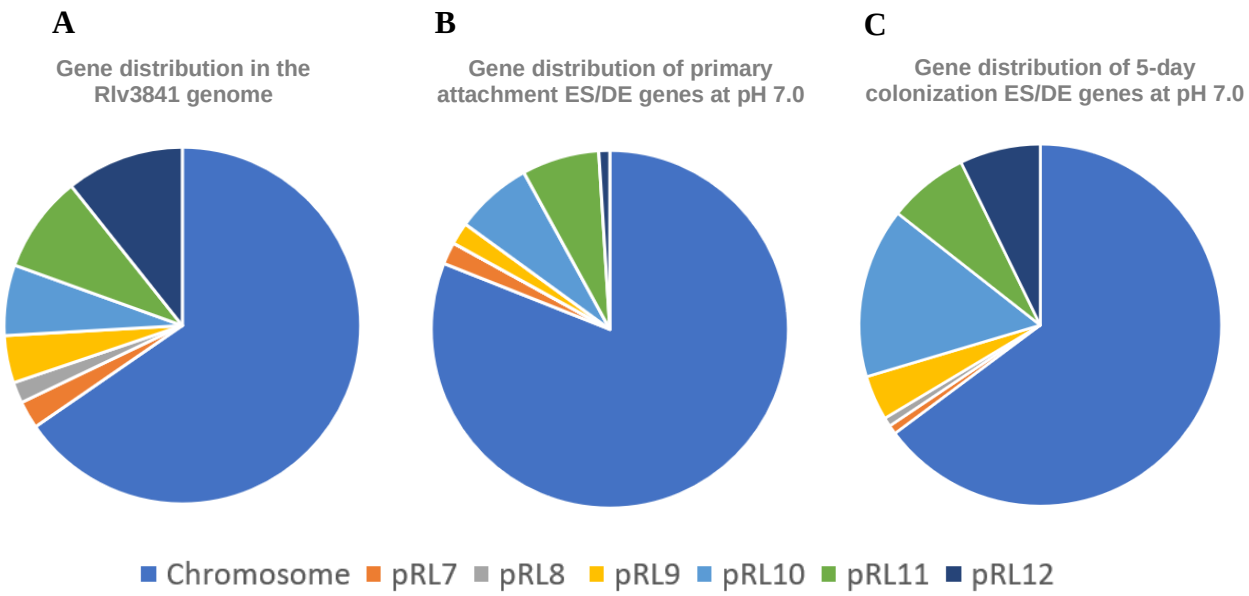


Figure 4-7. Genomic localization of genes (as a proportion of the total). A – 7,340 Rlv3841 genes. B – 162 primary attachment genes required (ES/DE) at pH 7.0. C – 125 secondary attachment genes required (ES/DE) at pH 7.0 determined from 5 d colonization INSeq [238].

In the Rlv3841 genome, 65.4 % of annotated genes are located on the chromosome, with the remainder divided between the six plasmids (Figure 4-7, A). At pH 7.0, the distribution of the genes required for primary attachment (Figure 4-7, B) shows that, compared the genome, the location is skewed towards the chromosome (~81%) and away from the plasmids, with no pRL8 representation. In contrast, genes required for secondary attachment (Figure 4-7, C) show a distribution more like that of the genome, with ~65% encoded on the chromosome and representation of all plasmids, although there is a clear skew to pRL10 (~15% compared to ~6% in the genome).

This demonstrates that different factors are involved in primary and secondary attachment and that genes required for the latter are likely to be induced during the attachment process. The fact that most of the nodulation and nitrogen fixation genes are encoded on pRL10 [141] as well as the *rhi* genes, important root colonization determinants [238] may explain the skew to pRL10 in 5 d colonization. This agrees with a model in which secondary attachment factors (e.g. extracellular fibrils) are synthesized in response to successful primary attachment [60, 71], and provides further validation of the experimental separation of attachment stages.

4.2.8 Mapping gene requirements at different symbiosis stages from INSeq data

INSeq datasets (detailed in Table 4-5) are available for multiple stages of the Rlv3841/pea symbiosis and development of bacteroids. It is possible to map gene requirements for primary attachment with those of other symbiosis stages to see if they are primary attachment-specific or also needed for later symbiosis stages (Figure 4-8). Of those genes required (ES/DE) only in primary attachment, 104 of 292 genes (Figure 4-5) are likely to

represent factors important for attachment to the bulk root epidermis. Those 49 genes which are 'progressive' through infection threads and nodulation (i.e. needed for primary root attachment and all subsequent symbiosis stages, Figure 4-8) are likely to be involved in attachment to root hairs since these are the site of infection thread development [26]. The remaining 139 of 292 genes needed in one or more of the root attached libraries (Figure 4-5) are required (ES/DE) at non-contiguous stages of the symbiosis, and not shown in Figure 4-8. These genes are likely to have discrete functions at different stages of symbiosis development. For example, they may be required in primary attachment and in nodulation. INSeq HMM designation for the genes discussed can be found in Appendix 1, Tables A3-A9.

Table 4-5. Details of the five INSeq libraries used to map the gene requirements to stages of Rlv3841 symbiosis with pea plants.

Library	Condition	Genes required (ES/DE)	Source
Rhizosphere	Pea rhizosphere	904	[238]
Primary root attachment	1 hr root attachment, pH 6.5, 7.0, 7.5	292	This work
Colonization (secondary root attachment)	5 d root colonization	911	[238]
Infection thread	Infection thread	1131	Unpublished data, Poole Lab
Nodulation	Nodule bacteroids	1043	Unpublished data, Poole Lab

Conditions refer to the environment the output libraries were extracted from prior to DNA isolation and library preparation/sequencing.

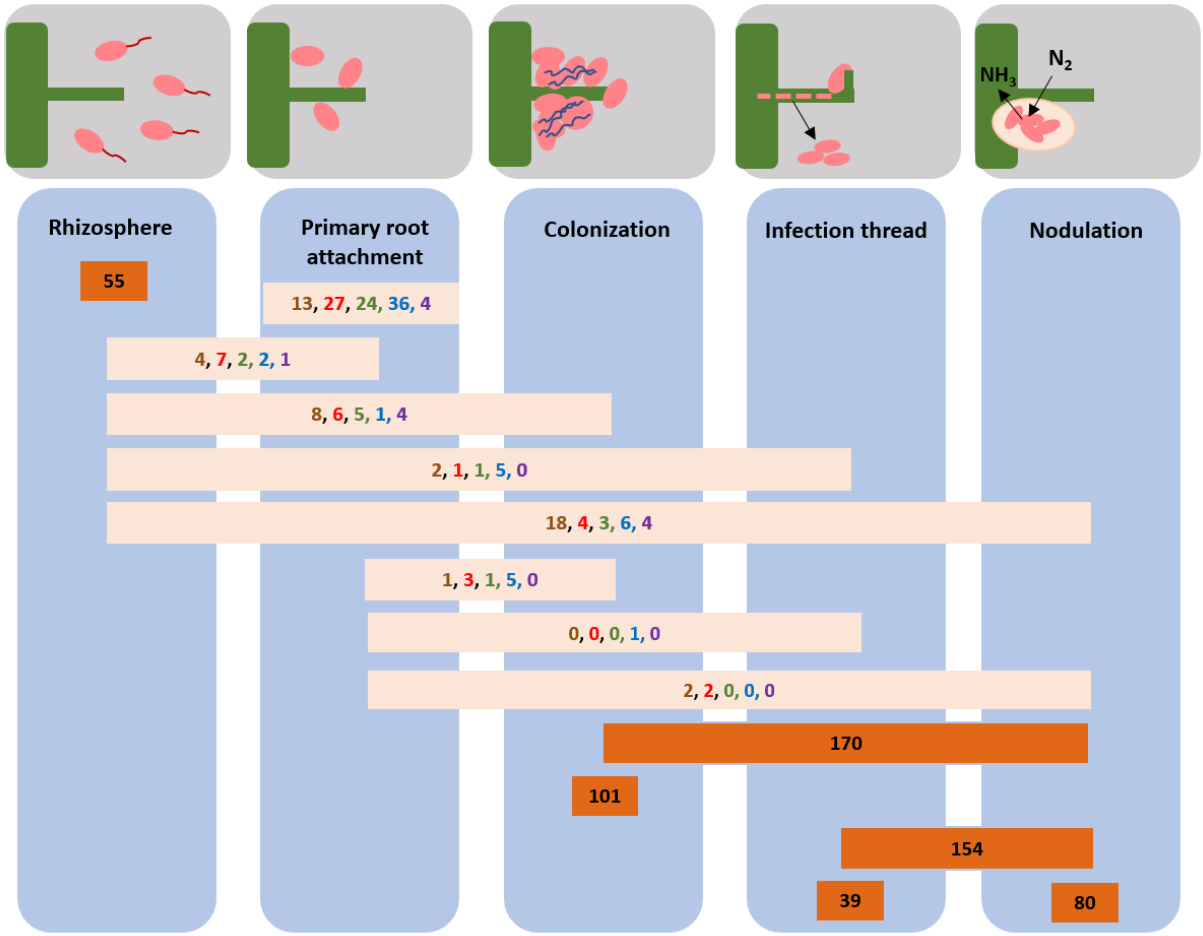


Figure 4-8. Genes required (ES/DE) at each stage of Rlv3841/pea symbiosis. Light boxes show genes required (ES/DE) in primary root attachment (and other stages where applicable) at different combinations of pHs. Color code: brown - pH 6.5, 7.0 and 7.5. Red – pH 6.5 only. Green – pH 7.0 only. Blue – pH 7.5 only. Purple – pH 6.5 and 7.5. Dark orange boxes show genes required in other conditions (Table 4-5). All genes were NE in the input library.

4.2.9 Comparison of INSeq predictions and Lux attachment assays

To compare results from Lux attachment assays and HMM classifications, attachment assay data from Chapter 3 (Figure 3-5, 3-7 and 3-11) was compared to the HMM gene classifications for primary root attachment obtained in this chapter.

INSeq investigates the effect of single-insertion *mariner* transposon mutants, so meaningful comparison with the double and triple mutant strains tested in Lux assays was not possible. It is also difficult to compare the HMM classifications (ES, DE, NE, AD) directly with the simpler statistical testing of results (as a percentage of wild-type attachment) in Lux assays. To enable comparison, the following assumptions were made: genes classified as NE in INSeq should show no statistical difference in attachment in a Lux assay from wild-type (Rlv3841[pIJ11282], D5250), while genes required (ES/DE) should show a statistically significant reduction from wild-type in Lux attachment assays. Significant or non-significant differences in the case of Lux assays are based on unpaired t-test results with the upper threshold for p value significance set at $p < 0.05$. Table 4-6 gives a comparison of the 13 single mutant strains tested in Lux whole-root attachment assays in Chapter 3 with the gene classification from primary attachment INSeq data.

Table 4-6. Comparison of INSeq and Lux attachment assay data.

Gene	Strain	Luminescence from mutant as a percentage of wild-type [a]			INSeq classification		
		pH 6.5	pH 7.0	pH 7.5	pH 6.5	pH 7.0	pH 7.5
praR (*RL0390*)	OPS0111	61%	467%	67%	NE	NE	NE
pssA (*RL3752*)	OPS0804	21%	37%	38%	DE	DE	DE
gmsA (*RL1661*)	OPS0914	1%	16%	70% *	DE	DE	NE
pRL110543	OPS1131	105% *	133% *	114% *	NE	NE	NE
RL2969	OPS1235	8%	16%	67%	NE	NE	NE
RL4382	OPS1236	3%	20%	8%	DE	DE	DE
RL3273	OPS1237	138% *	45%	42%	NE	NE	NE
RL0109	OPS1238	8%	23%	18%	DE	ES	ES
pRL110071	OPS1290	28%	43%	21%	DE	DE	DE
pRL100053	OPS1291	4%	7%	19%	DE	DE	DE
flgE (*RL0728*)	OPS1738	19%	12%	36%	NE	NE	NE

motA (*RL0703*)	OPS1739	16%	13%	12%	NE	NE	NE
nifH (*pRL100162*)	OPS1878	102% *	107% *	99% *	NE	NE	NE

[a] Wild-type in Lux attachment assay is Rlv3841[pIJ11282] (D5250). Results are expressed as average root attachment of test strain in percent, indexed to 100% attachment for D5250 under each condition. Lux attachment results in black show agreement with INSeq data, while those in red do not when using the criteria defined in 4.2.9. * indicates that in Lux attachment assay the strain does not show a statistically significant difference from D5250 using an unpaired t-test with the upper threshold for p value significance set at $p < 0.05$.

Comparing results from the Lux attachment assay to the HMM classifications for primary attachment reveals that, using the criteria defined, there is close agreement in the results from the two techniques for 25 out of 39 comparisons (13 strains tested under three attachment conditions, Table 4-6). Whilst this is a reasonable degree of agreement (~66%), there are various reasons why the two would not completely agree since they are measuring slightly different things. This is due to the design of each experimental technique, each with limitations that need to be considered.

The conflict in results for OPS0111 *praR* (*RL0390*, mutant strain OPS0111), could be due to the low gene TA site number (only four in this case) which affects the degree of certainty in HMM gene classification. Although the HMM incorporates neighboring TA site data to smooth read counts, low TA site number may still reduce the certainty of classification [183]. Using TnSeq, Griffin *et al*. (2011) and Zhang *et al*. (2012) reported uncertainty in assigning state call to bacterial genes with fewer than six TA sites [173, 273]. Therefore, despite the relatively high saturation of the input library (Table 4-2), low TA site number may present an issue for HMM classification, and care should be taken for genes with under six TA sites. 83% of the annotated Rlv3841 genes have >six TA sites. The use of a richer media for *mariner* transposon library construction, or pooling more independent conjugations (as described by DeJesus *et al*. in [274]) would increase input saturation and reduce this problem. For the remaining lack of agreement between these two techniques (genes *RL2969, RL3273, RL0728* and *RL0703*), TA site number is unlikely the issue, each gene having >20 TA sites.

Another possible reason for results disagreement is the different nature of the two assays. In a Lux assay, inoculum is an OD_{600} 0.1 resuspension of a single strain. In INSeq, inoculum is the *mariner* transposon input library, a heterogeneous population of ~115,000

different mutant strains (140,845 genomic TA sites and 82% insertion density). This raises the possibility of *in trans* complementation, where bacteria possessing a functional copy of a gene may act to 'rescue' bacteria with a non-functional copy of the same gene (providing them with a secreted factor, as described for lethal factor in *Bacillus anthracis*, for example [275]). To test if *in trans* complementation could rescue strains OPS1235 (mutated in *RL2969*), OPS1237 (mutated in *RL3273*), OPS1738 (mutated in *flgE*) or OPS1739 (mutated in *motA*) in a Lux attachment assay, further assays were performed as described in 2.6.6. However, inoculation was at 1:1 or 1:100 mixture of pIJ11282-labelled test strain (mutant) and Rlv3841 (unlabeled), thereby providing wild-type Rlv3841 cells with functional copies of all genes which would be able to rescue mutant cells reduced attachment phenotypes if *in trans* complementation is occuring. Complementation was seen for OPS1235 at a 1:100 inoculum, returning attachment to wild-type levels at all pHs tested (Figure 4-9). As *RL2969* encodes a putative transmembrane protein, it seems that either this protein could be released from wild-type cells surfaces and bind cell surfaces of mutant strains to increase attachment ability by acting as an attachment factor. Alternatively, this protein could act to aggregate wild-type cells with mutants on the root surface, and thereby yield a complementation phenotype in attachment. Precisely how this *in trans* complementation is occurring remains unclear; more detailed characterisation of *RL2969* could shed further light on this. However, this result shows that *in trans* complementation may skew INSeq results for a subset of genes and mask their role in primary attachment.

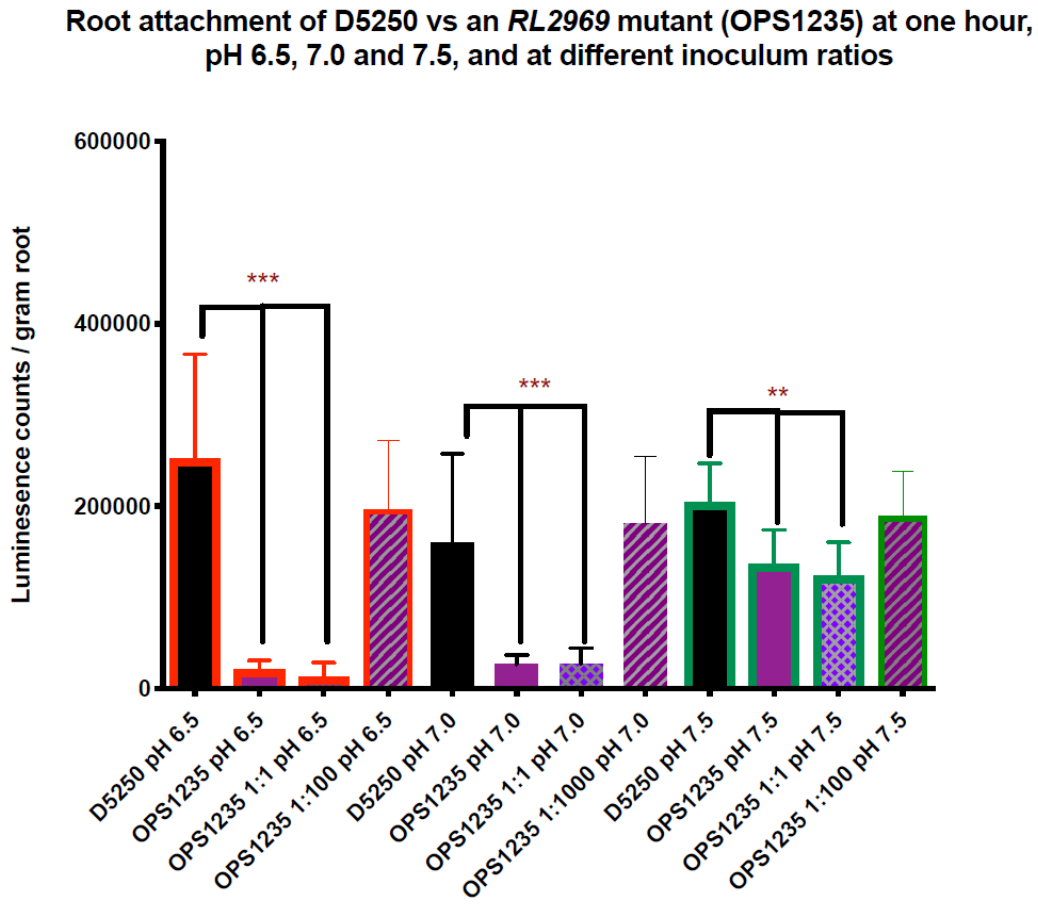


Figure 4-9. Comparison of D5250 (Rlv3841[pIJ11282]) and OPS1235 (*RL2969* mutant) in single inoculum, or in 1:1 or 1:100 mixed inoculum with unlabelled Rlv3841 in a Lux whole root attachment assay. Strains, inoculum ratios and pH conditions are given on the x-axis. Data is displayed as mean ± SEM, n=10. An unpaired t-test was used to compare groups.

For the remaining three strains (*RL3273, flgE and motA* mutants) complementation was not seen under any tested conditions (data not shown). Although the reasons for this are unclear, it may be that a lower inoculum ratio would be needed to see effects. In INSeq, mutants in a given TA site will be present at a frequency of (in this case) ~1:115,000.

However, Lux assay sensitivity precludes testing at such low ratios of labelled cells. Note that a method such bacterial recovery from roots and plating on antibiotics could be used to investigate rescuing of attachment in mutants at extremely low inoculum ratios. However, this approach was not pursued here.

Particularly intriguing are the results for OPS1738 (*flgE* mutant) and OPS1739 (*motA* mutant). These non-motile strains are classified as NE in INSeq but are strongly defective in a Lux attachment assay (Table 4-6, Figure 3-5). Lux results mirror the characterised motility requirement for migration towards roots and positioning for primary attachment [34, 44, 45, 276]. In a mixed inoculum it may be possible that the presence of a majority-motile population could result in a biased Brownian motion or 'aided motility'. This could occur through cell-cell collisions between chemotaxing cells and non-motile cells and lead to positioning of non-motile cells in proximity to the root, permitting primary attachment. In combination with the above, motile bacteria could disrupt a boundary layer around the plant root (caused by replusive hydrodynamic forces at the root surface, [44]). Rotating flagella are known to act as an active propeller to overcome these boundaries [44], and non-motile cells may be able to benefit from this boundary disruption. Therefore, 'aided motility' could be thought of as a combination of biased Brownian motion and root boundary effect disruption in a mixed motile/non-motile bacterial cell population, which can allow non-motile cells to attach to roots by reaching the root surface. If this happens for enough non-motile cells, TA site representation may be high enough in the output library for NE classification. Here, the fitness values (see 2.7.5) for *flgE* and *motA* did decrease (from 248 in the input to an average of 144 in the outputs for *flgE,* and from 756 to 231 for *motA*). Although the HMM uses more sophisticated statistics for gene classifications [183], a decrease in fitness value does indicate a decrease in strain presence

in the output library. This may support the notion of some ‘aided motility’ occurring. A mathematical modelling exercise could shed more light on the extent to which this is possible in a mixed motile/non-motile bacterial population undergoing chemotaxis.

4.2.10 Increasing the specificity for identification of primary root attachment factors from INSeq results – pleiotropy filtering

Many of the factors identified in INSeq show pleiotropic effects on other stages of symbiosis (188 of 292 genes, Figure 4-5). This could indicate that they have discrete roles at other stages of symbiosis, or that loss of primary attachment function leads to loss of mutant strains when assaying at later symbiosis stages using INSeq.

However, if a *mariner* mutant shows pleiotropic effects in growth on different media or under different metabolic conditions, this may be an indicator that a gene may not be a specific primary attachment determinant. Rather, that mutation impairs cellular function in a way that reduces attachment in a general, non-specific way. To increase the specificity of primary attachment factor identification, HMM classifications were cross-referenced against other INSeq datasets (listed in Tables 4-5 and 4-7). Results were used to assign a ‘pleiotropy filter’, where genes showing a non-NE classification in any of the experiments listed in Table 4-7 were considered non-specific primary attachment determinants. This filter can be seen in Tables A3-A9. The distribution of the updated list of116 post-filtering genes is shown in Figure 4-10.

Table 4-7. Details of INSeq libraries used to increase the specificity of primary root attachment factor identification

Library	Condition	Genes required (ES/DE)	Source
VMM	Vincents media	845	[175]
21% oxygen 10mM glucose	UMS/glucose/ NH_4, 21% O_2	1168	[238]
1% oxygen 10mM glucose	UMS/glucose/ NH_4, 1% O_2	1100	[238]
21% oxygen 20mM succinate	UMS/succinate /NH_4, 21% O_2	1166	[238]
1% oxygen 20mM succinate	UMS/succinate /NH_4, 1% O_2	1192	[238]

Selection conditions refer to the environment the output libraries were extracted from prior to DNA isolation and library preparation/sequencing.

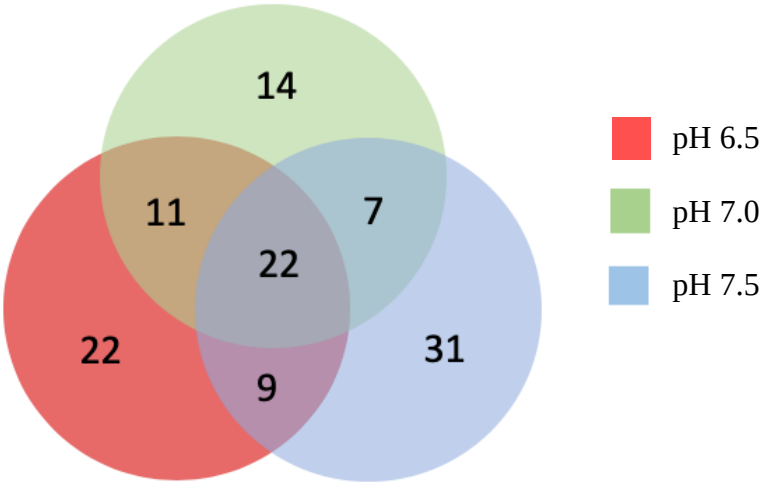


Figure 4-10. Specific primary attachment determinants. These genes have been classified as NE in the input library, required (ES/DE) in one or more of the root-attached libraries (pH 6.5, 7.0 or 7.5) and have been classified as NE in the following INSeq datasets: VMM, 21% oxygen 10mM glucose, 1% oxygen 10mM glucose, 21% oxygen 20mM succinate and 1% oxygen 20mM succinate (see Table 4-7). Total genes = 116. Venn diagram circle color indicates the pH at which the attachment assay was performed; red = pH 6.5, green = pH 7.0, blue = pH 7.5

4.2.11 Primary attachment determinants required under different pH conditions

Filtering increased the specificity of primary attachment determinant identification (4.2.10). Remaining genes (Figure 4-10) were investigated in more detail to determine what roles they play and how pH changes influence Rlv3841 primary attachment mechanisms. Table 4-8 provides a summary of the tables and figures in which this data is shown. Note that only two genes were classified as AD and are discussed in the next chapter (5.2.12, Table 5-13 and Table A17).

Table 4-8. Summary of where results are presented for genes required (ES/DE) for primary attachment at pH 6.5, 7.0 and 7.5 and combinations thereof.

Primary attachment pHs	Genes required (ES/DE)	Reference Tables and Figures
pH 6.5, 7.0 and 7.5	22	Table 4-9, Appendix Table A3, Figure 4-11
pH 6.5 only	22	Table 4-10, Appendix Table A4, Figure 4-12
pH 7.0 only	14	Table 4-11, Appendix Table A5, Figure 4-13
pH 7.5 only	31	Table 4-12, Appendix Table A6, Figure 4-14
pH 6.5 and 7	11	Table 4-13, Appendix Table A7, Figure 4-12 and 4-13
pH 7.0 and 7.5	7	Table 4-14, Appendix Table A8, Figure 4-13 and 4-14
pH 6.5 and 7.5	9	Table 4-15, Appendix Table A9, Figure 4-12 and 4-14

Table 4-9. 22 genes identified as required (ES/DE) for primary attachment at pH 6.5, 7.0 and 7.5

Gene(s)	Symbiosis defects	Description
pRL100053	PA, Col, IT, Nod	Putative transmembrane domain containing protein, helix-turn-helix 37 domain. See Figure 3-11 and section 3.2.9. Unknown protein localization.
pRL100174	RA	Hypothetical protein, no known conserved domains. Unknown protein localization.
hslO (*RL0551*)	Rhi, PA, Col	Putative Hsp33-like chaperonin. Redox regulated molecular chaperone protecting thermally unfolding and oxidised proteins from aggregation – defence against oxidative stress [277, 278]. May assist in coping with a diverse rhizosphere environment (see 4.1)
RL0876	RA	Conserved hypothetical protein, no known conserved domains. Cytoplasmic protein localization.
RL1381	RA	Uncharacterised protein. Unknown protein localization.
amn (*RL1478*)	PA, IT, Nod	AMP nucleosidase; catalyses hydrolysis of AMP to form adenine and ribose 5-phosphate. Changes in AMP levels allow rapid adjustments to changing metabolic conditions [279].
RL2400	RA, Nod	Putative MarC family transmembrane protein, not involved in antibiotic resistance [280], function unknown. Cytoplasmic membrane protein localization.
tpiA (*RL2513*)	Rhi, PA, Col	Putative triosephosphate isomerase. Upregulated in *Staphylococcus aureus* biofilm, possibly due to oxygen limitation [281]. Glycolytic enzymes play additional roles when localised on the cell surface (e.g. α-enolase plasminogen binding in streptococci and GAPDH transferrin binding activity in *S. aureus*) [282, 283]. Surface localised glycolytic enzymes are multifaceted and can be involved in substrate binding. TpiA has been shown to be surface localised and have a direct role in attachment to host cells in *Mycoplasma gallisepticum* [284].

recA (*RL2637*)	Rhi, PA, Col, IT, Nod	RecA is needed for DNA repair and the SOS response. The major activity of RecA in DNA metabolism is the promotion of DNA strand exchange [285]. Requirements for all symbiosis stages suggest that RecA may assist in coping with a diverse rhizosphere environment (see 4.1). However, *recA* disruption has been shown to reduce adherence and colonization of host cells by *Vibrio cholerae*, although the mechanism underlying this remains unknown [286].
pfp (*RL3322*)	PA, IT	Putative pyrophosphate-fructose 6-phosphate 1-phosphotransferase. Catalyses the first committed step in glycolysis, the phosphorylation of D-fructose-6-phosphate [287]. Like *RL2513* (see above), could have multifaceted role
pssA (*RL3752*)	PA	Glycosyl transferase involved in EPS biosynthesis [288, 289]. Mutants are deficient in EPS production and form biofilms slowly compared to Rlv3841. Does not attach to root hairs [56]. Biofilms are flat and unstructured [58]. See also Figure 3-5 and section 3.2.5.
rpoH (*RL3766*)	Rhi, PA, Col	Putative RNA polymerase sigma-32-factor, heat shock. Involved in the regulation of expression of heat shock genes and stress response; may also confer pH change and osmotic stress tolerance [290]. May assist in coping with a diverse rhizosphere environment (see 4.1)
RL3987-90	Rhi, PA, Col, IT, Nod	*RL3987* – uncharacterised, SpoVT-AbrB domain. *RL3988* – uncharacterised, PINc domain. PIN domains function as single stranded RNA nucleases [291]. In prokaryotes they are usually the toxin of toxin-antitoxin operons, helping free-living prokaryotes cope with nutritional stress [292]. *RL3989, RL3990* – Holliday junction ATP dependent DNA helicases RuvA and RuvB; DNA damage repair mechanism. May be required for osmotic shock responses [293, 294]
RL4065	Rhi, Col, RA	Conserved hypothetical protein, no known conserved domains. Cytoplasmic protein localization.
RL4145	RA	Putative conserved LacI type transcriptional regulator (repressor). Regulatory targets unknown. Cytoplasmic protein localization.

RL4362/ dacC (*RL4363*)	PA, IT, Nod	*RL4362* – putative cobalamin (vitamin B12) synthesis protein, CobW domain [295]. Required by *S. meliloti* for symbiosis with *M. sativa* [296]. Only one cobalamin dependent enzyme (*nrdJ*, *S.meliloti* cobalamin dependent ribonucleotide reductase, RNR) affects symbiosis. Removal of *ndrJ* impairs symbiosis; rhizobia are lysed in the plant cytoplasm [297]. Loss of *RL4362* may reduce fitness for competitive primary attachment. Note that cobalamin synthesis genes may be misclassified due to high homology with glutamine amidotransferases, which are involved in peptidoglycan amidation [298]. *RL4363* – *dacC*, putative penicillin binding protein, peptidase S11 domain. In *E. coli dacC* processes sugar-peptide cell wall precursors; involved in peptidoglycan biosynthesis [299].
RL4381/RL4382	PA/PA, Col	Putative POTRA domain transporter (*RL4381*) and filamentous hemagglutinin adhesin (*RL4382*). See Figure 3-11 and section 3.2.9

Genes are listed together where co-localised. Symbiosis defects describe the test conditions under which these genes are classified as ES/DE. Rhi = rhizosphere, RA = (primary) root attachment, Col = 5 d colonization, IT = infection thread, Nod = nodulation. Protein localization was predicted using pSORTb v 3.0.2 [206].

The genes described in Table 4-9 constitute the core primary attachment factors for Rlv3841 in symbiosis with pea, being required at all pHs. The requirements for *RL4381* and *RL4382* were previously unknown, but this result (along with those in 3.2.9, where a mutant in *RL4382* is impaired in attachment in a Lux assay) shows that FHA is crucial to root attachment. The requirement for *pRL100053* (Table 4-6) was demonstrated in 3.2.9, although its exact function remains unknown.

The requirement for *pssA* is unsurprising. This glycosyltransferase catalyses the transfer of glucose-1-phosphate from UDP-glucose to the isoprenylphosphate lipid carrier, the first stage in EPS biosynthesis [288]. This again (in addition to Lux attachment assay results Table 4-6 and 3.2.9) highlights the central role of EPS in primary attachment.

RL4362 is annotated as a putative cobalamin synthesis protein, but may actually represent a peptidoglycan amidation factor [298]. This seems likely given the co-involvement of *RL4363*, a peptidoglycan biosynthesis factor in *E. coli* [299], and points to the importance of peptidoglycan function and modification in primary attachment.

Of interest is *RL2513*, encoding the glycolytic enzyme TpiA. The demonstrated moonlighting function of this enzyme [282, 283], particularly in *Mycoplasma gallisepticum* attachment [284], indicates a novel function in Rlv3841/pea interactions, where it may be surface localised. This may also be the case for *pfp* (*RL3322*), though this is less clear.

Further genes - *RL0051*, *rpoH* - *RL3766*, *amn* - *RL1478*, *RL3987-90* and *recA* - *RL2637* - demonstrate a requirement for stress responses, metabolic changes and DNA repair in primary attachment. Of these, *recA* has been shown to play a role in *Vibrio cholerae* cell adherence, although it is unclear as to whether this is due to the role of RecA in stress responses or its ability to alter the expression of colonization factors [286]. For others,

their requirement may reflect stresses encountered in assay conditions where inoculum is resuspended in buffer before root inoculation, which could induce nutritional and oxidative stress. However, considering the broad functions of these genes, the diversity of stresses in a real rhizosphere environment (including oxidative, nutritional and temperature-related), and that many are required for subsequent symbiosis stages (particularly colonization, where experiments were conducted in vermiculite [238]), it seems possible that they may also be important under field conditions, although this cannot be definitively concluded at this stage. The remaining genes involved (Table 4-9) are largely of unknown function.

In addition to the genes described in Table 4-9, several genes known to be important for primary attachment were also required (ES/DE) at all pHs but were also classified as ES/DE in the input libraries. This is presumably due to their importance for normal cell functioning. These factors include *exoR* (*RL2037*) (involved in succinoglycan synthesis, flagellar gene expression and biofilm formation [300, 301]), *chvA* (*RL4640*) (cyclic-β-1,2-D-glucan synthase; one of the first known attachment mutants, also hypersensitive to osmotic stress [55, 302–304]) and *rosR* (*RL1379*) (nodulation competitiveness and polysaccharide transcriptional regulator, [216]).

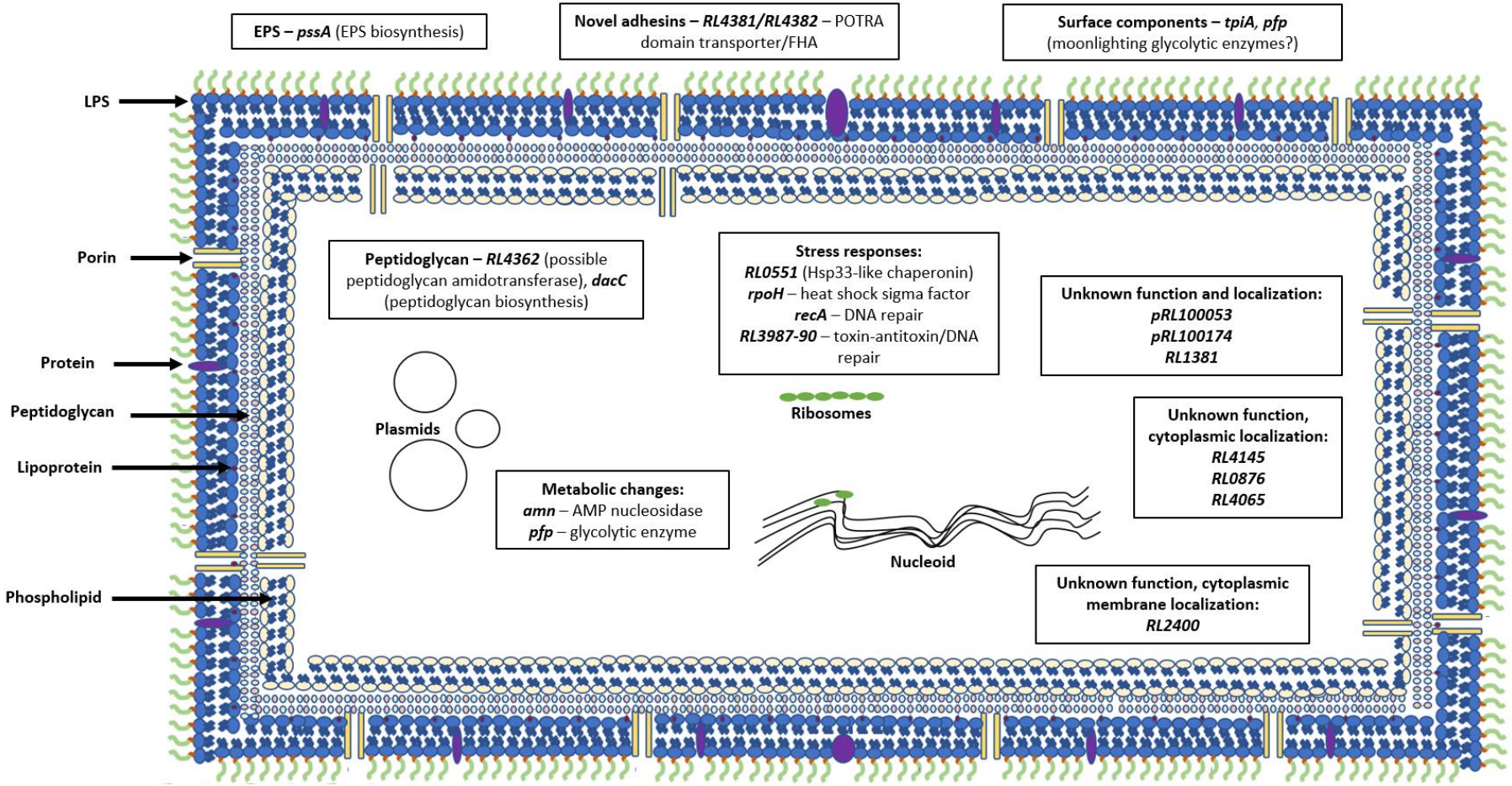


Figure 4-11. Diagram of a stylized gram-negative Rlv3841 cell showing the primary attachment determinants at all pH 6.5, 7.0 and 7.5 (see Table 4-9), grouped by function. Cell outer membrane, peptidoglycan layer, inner membrane, plasmids, nucleoid and other cellular factors are shown, not to scale.

Table 4-10. 22 genes identified as required (ES/DE) for primary attachment at pH 6.5 only.

Gene(s)	Symbiosis defects	Description
pRL100176	Rhi, RA	Pseudogene – generally considered non-functional, but can encode partial activity or affect mRNA stability of homologues genes [305]. Unknown protein localization.
pRL100177	RA	Putative homologue of eukaryotic tubulin. Unknown protein localization.
fucA (*pRL100274*)	RA	Putative α-L-fucosidase. Cleaves fucosidic bonds in glycans (particularly in peptidoglycan structures)[306]; likely involved in remodelling of the cell surface. Fucose rich EPS has been documented in *R. sullae* [307].
pRL110283	RA, IT, Nod	Putative DNA binding protein. Cytoplasmic protein localization.
pRL120021	RA, Col	Uncharacterised protein. Unknown protein localization.
impA (*pRL120475*)	RA	Inner membrane protein ImpA. Mutation of *impA* causes outer membrane disruption in *Actinobacillus actinomycetemcomitans* [308]. Expressed from the pathogenic injection type VI secretion system operon in *V. cholerae* [309]. Nicolsamide reduces *Xanthomonas oryzae* leaf blight disease in rice by downregulating xanthan, EPS and impA expression [310]
npr (*RL0032*) / *manX* (*RL0033*)	Rhi, RA, IT, Nod	Components of the phosphenolpyruvate phosphotransferase (PTS) system regulate lifestyle switches [155]. Phosphorelay between Npr and ManX is believed to regulate carbon metabolism [311]. Npr and ManX are co-localised with the ChvI operon [312] and globally regulate ATP-dependent ABC transporter activity in a pos-translational fashion [155, 313]. Mutation of PTS system components causes dry colony morphology due to reduced EPS secretion [313].
RL0398	RA, IT, Nod	Putative N-acetyltransferase (NAT). STRING database indicates functional interaction with RL0399 (putative endopeptidase, interaction score 0.91) and MurE (peptidoglycan biosynthesis factor, interaction score 0.71) – possibly related to alterations in peptidoglycan structure

RL0614	RA	Unknown function. Cytoplasmic protein localization.
RL0726	RA, IT, Nod	Conserved hypothetical exported protein, transglycosylase Slt domain. Degrades peptidoglycan via β 1-4 glycosidic bond cleavage. Linked to biofilm formation in *S. enterica, E. coli* and *Acinetobacter baumannii* [314–316]. Lytic transglycosylases participate extensively in cell wall remodelling, recycling of peptidoglycan and space-making for insertion of cell-envelope spanning structures [317]. A lytic transglycosylase (RL4716) was characterised in Rlv3841 as required for cell envelope function and biofilm formation [318].
ctaE (*RL1026*)	RA	Putative cytochrome c oxidase polypeptide III. Biofilm formation is promoted by some cytochrome c oxidases under anoxic conditions in *P. aeruginosa*, possibly due to nitric oxide (NO) accumulation [319]
RL2211	RA, Nod	Uncharacterised protein. Phasin 2 domain commonly found in phasin proteins which stabilise polyhydroxybutarate storage granules [222]. Unknown protein localization.
hflX (*RL2285*)	RA	GTPase HflX. Heat-shock induced ribosome splitting factor; rescues translationally stalled ribosomes under heat shock [320].
ccdA (*RL2303*)	Rhi, RA, Col, IT, Nod	Putative cytochrome c-type biogenesis protein, DsbD transmembrane domain, disulphide bond formation and redox condition maintenance [321]
RL2316	RA	Putative guanylate cyclase. Catalyses formation of cyclic di-GMP (c-di-GMP), key second messenger in biofilm formation / motile to sessile lifestyle switch [79]. NO can stimulate biofilm formation by regulating c-di-GMP levels in *Shewanella oneidensis* [322]
nnrE (*RL2394*)	Rhi, RA	Multifunctional fusion protein NnrE. Repairs epimers of NAD(P)HX, a damaged form of NAD(P)H
RL2564	RA	Hypothetical exported protein. Unknown protein localization.
RL2595	RA	Putative MutT/nudix family protein. Mutagenesis of nudix proteins in *P. syringae* str DC3000 and *P. aeruginosa* display defects in motility and biofilm formation [323]

dgkA (*RL2780*)	RA	Diacylglycerol kinase DgkA. In *E.coli*, DgkA mutants are defective in biofilm formation [316], and DgkA function has been linked to phospholipid recycling and LPS modifications [324]. In *B. subtillis* it is important for lipoteichoic acid synthesis [325]
RL3179	RA	Putative cobalamin (vitamin B12) synthesis protein, CobW domain, may also be involved in peptidoglycan amidation. See Table 4-9, *RL4362*
gelA (*RL4404*)	RA	Gel forming EPS production protein GelA. Regulated by RosR [158]. Mutation has no effect on 2-9 day biofilm formation or 90 minute pH 6.5 or 7.5 attachment to root hairs [56]. Appears to have role in primary bulk root attachment at pH 6.5 based on this data.

Genes are listed and described together where co-localised. Rhi = rhizosphere, RA = (primary) root attachment, Col = 5 d colonization, IT = infection thread, Nod = nodulation. Protein localization was predicted using pSORTb v 3.0.2 [206].

Genes required for primary attachment at pH 6.5 suggest specific Rlv3841 adaptations under these conditions. Cell wall and surface modification requirements are demonstrated, including EPS, peptidoglycan and LPS modifications (*fucA -pRL100274*, *RL3179* and *dgkA - RL2780*, respectively) and cell wall remodelling (*RL0726*), possibly for the insertion of cell envelope spanning structures, which can include flagella and secretion systems [317]. This could theoretically be linked to *impA* (*pRL120475*) requirement, as this is known to be involved in membrane function and attachment [308, 310]. However, *RL2564* (hypothetical exported protein) or one of the other uncharacterised proteins could also be involved.

The *gelA* (*RL4404*) requirement is of particular interest as, whilst annotation and RosR regulation is suggestive of a role in attachment, no role had been previously defined for this EPS production protein [56, 158].

Biofilm formation factors (*ctaE - RL1026* and *RL2595*) required only at pH 6.5 may reflect adaptation to changing root surface profiles (as hypothesised in the glucomannan/rhicadhesin system [34, 56]), or bacterial pH adaptations.

The ES/DE classifications of *npr* (*RL0032*) and *manX* (*RL0033*) are of interest, as mutation reduces EPS secretion and ATP dependent ABC transporter activity [313], and it is unclear why these PTS system components should be required for attachment only at pH 6.5.

RL2316 was also required for attachment at pH 6.5 and is interesting given its role as a likely catalyst of c-di-GMP formation. This is a key second messenger in biofilm formation and the motile to sessile lifestyle switch in many bacterial species [79] (c-di-GMP in attachment is discussed in detail in 5.2.12 and 7.8). Further genes in Table 4-10 are also important, though their precise role remains unknown.

Table 4-11. 14 genes identified as required (ES/DE) for primary attachment at pH 7.0 only.

Gene(s)	Symbiosis defects	Description
pRL100112	Rhi, RA, Col	Putative dehalogenase-hydrolase (HAD), member of a large superfamily of phosphohydrolases. HADs display activity against various phosphorylated metabolites [326]. A mutant in a *Xanthomonas axonopodis* pv. *citri* HAD is defective in biofilm formation [327]
pRL120518	RA	Putative TetR family transcriptional regulator
RL0052	RA	Uncharacterised protein. Unknown protein localization.
RL1013	Rhi, RA, Col, IT	Uncharacterised protein, 17kDa Anti 2 motif. These motifs are common membrane proteins and include surface antigens in *Rickettsia* [328]. Unknown protein localization.
RL1052	RA	Uncharacterised protein. Unknown protein localization.
pspA (*RL1106*)	Rhi, RA	Putative PspA family regulator, phage shock protein A. Involved in antibiotic resistance and biofilm formation in *E.coli* and pathogenesis in *S. typhimurium* [329, 330]
RL1371	Rhi, RA, Col	Putative transmembrane protein
scpA (*RL2044*)	RA	Segregation and condensation protein A, participates in chromosomal division during cell partition
anmK (*RL2587*)	RA	Anhydro-N-acetylmuramic acid kinase AnmK. Catalyses the phosphorylation of 1,6-anhydro-N-acetylmuramic acid (anhMurNAc), cleaving the 1,6 anhydro ring and generating MurNAc-6-P. Required for cell wall recycling [331, 332]
tyrS (*RL2588*)	Rhi, IT, Nod	Tyrosine-tRNA ligase TyrS. Catalyses attachment of tyrosine to tRNA
gor (*RL2694*) */RL2695*	RA, IT, Nod / RA	*RL2694* – Glutathione reductase (Gor). *RL2695* – uncharacterised protein. Unknown protein localization.
ahpD (*RL3226*)	RA	Alkyl hydroperoxide reductase AhpD

RL4704	RA	Putative glyoxylase family protein, member of the VOC superfamily. Members of this family are known to detoxify methylglyoxal, formed as a by-product of lipid metabolism [333]

Genes are listed and described together where co-localised. Rhi = rhizosphere, RA = (primary) root attachment, Col = 5 d colonization, IT = infection thread, Nod = nodulation. Protein localization was predicted using pSORTb v 3.0.2 [206].

At pH 7.0 primary attachment factors include the cell wall recycling AnmK protein (*RL2587*), which cleaves peptidoglycan structures [331, 332]. This is indicative of further cell wall and surface modifications required for attachment competence at this pH. In line with this, *RL4704* requirements (detoxifying by products of lipid metabolism) may indicate LPS modification and turnover, though this is speculative.

Further factors important in biofilm formation (*pspA* and *pRL100112*) and a possible surface antigen-like adhesin (*RL1013*) are also necessary specifically at pH 7.0, with the function of remaining factors given in Table 4-11 being unclear.

Table 4-12. 31 genes identified as required (ES/DE) for primary attachment at pH 7.5 only.

Gene(s)	Symbiosis defects	Description
fhuA2 (*pRL120322*)	RA, Col	Outer membrane siderophore receptor FhuA2
pRL120795	RA	Uncharacterised protein. Unknown protein localization.
miaB (*RL0395*)	RA	tRNA-2-methylthio-N(6)-dimethylallyladenosine synthase
RL0401	RA, Nod	Putative universal stress protein, UspA family.
RL0561	RA, Col	Putative AraC family transcriptional regulator
rpiI (*RL1552*)	Rhi, RA, Col, IT, Nod	50S ribosomal protein L9 RpiI
RL1164 / RL1165	RA	*RL1164* – uncharacterised protein, unknown protein localization. *RL1165* – uncharacterised protein, 93% identity to gene RLV_3555 from *R. leguminosarum* biovar *viciae*, PepSY domain containing. These domains are likely to have a protease inhibitory function and may be cell wall associated [334]. Biofilm metalloprotease 1 (BmpI) from *Pseudoalteromonas* contains a PepSY domain required for biofilm formation [335]
pmtA (*RL1338*) / *RL1339*	RA	*rl1338* – putative phosphatidylethanolamine (PE) N-methyltransferase PmtA. Significantly upregulated in a *praR* mutant of Rlv3841, suggesting attachment role [59]. Phosphatidylcholine is found in the membranes of Rhizobiaceae and is synthesized from PE by PmtA. *pmtA* mutation in *Bradyrhizobium japonicum* disrupts symbiosis formation [336]. *rl1339*- uncharacterised protein
rpoZ (*RL1505*)	Rhi, RA	DNA directed RNA polymerase subunit omega, RpoZ. The smallest RNA polymerase subunit. However, additional roles have been identified: deletion of RpoZ impairs biofilm formation and sliding motility in *M.*

		smegmatis [337] and biofilm formation in *E. coli* [338] and *S. aureus* [339] through transcriptional changes. In *S. aureus*, *rpoZ* deletion impairs production of two tributyrin lipases. Loss of these lipases impairs biofilm formation; this may partly explain the *rpoZ* mutant phenotype [340].
RL2080 / RL2081 / RL2083	RA / RA, IT, Nod / Rhi, RA, Col, IT	*rl2080* – putative acetyltransferase, *RL2081* – putative transmembrane protein, *RL2083* – putative acetyltransferase
RL2094 /RL2095	RA	*RL2094* – putative polyhydroxybutarate (PHB) synthase. *RL2095* – uncharacterised protein. Unknown protein localization.
RL2489A	RA	Uncharacterised, 100% identity to transglycosylase associated protein Rleg_2013 from *R. leguminosarum* biovar *trifolii* WSM1325. Transglycosylation is a key step in bacterial peptidoglycan synthesis, catalysing glycan chain polymerisation [341, 342]
RL2491	RA	Conserved hypothetical exported protein
sixA (*RL2644*)	RA	Phosphohistidine phosphatase SixA, conserved. SixA is the only known bacterial phosphohistidine phosphatase, and dephosphorylates Npr in *E.coli* [343]. Implicated in biofilm formation in *E. coli* [344]. See also Table 4-10, *RL0032 / RL0033*
RL2777 / RL2778	RA	*RL2777* – uncharacterised protein. Cytoplasmic protein localization. *RL2778* - Putative exopolysaccharide biosynthesis protein
RL2857 / RL2858	RA	*RL2857* – putative LysR family transcriptional regulator. *RL2858* – conserved hypothetical exported protein, coiled-coil domain
hflC (*RL3253*), *hflK* (*RL3254*)	RA / RA, IT	Putative transmembrane serine proteases HflC and HflK. Functional HflC and HflK also modulate HflB activity. HflB, an AAA metalloprotease, is involved in membrane protein regulation, LPS biosynthesis and biofilm formation in *E. coli*, *B. subtillis* and others (where it is often called FtsH) [345, 346]. indicating a role

		in membrane regulation and biofilm formation [347–349]. HflB itself (encoded by *RL3965*) is ES/DE under all conditions in this work, including input (Appendix 2 Table 1).
RL4018	RA, IT, Nod	Putative ATP binding component of ABC transporter. 96% identity to Lipid A ABC exporter from *R. leguminosarum* biovar *trifolii* WSM2304, gene Rleg2_3249. Mutants with reduced lipid A show a delay in nodulation onset and impaired bacteroid shape [350]. Further, defects in lipid A production reduce surface attachment and motility [351]
RL4062	RA, IT, Nod	Putative amidohydrolase
RL4063	RA	Uncharacterised protein. Unknown protein localization.
RL4075	DE	Putative 5'-nucleotidase
RL4383	RA	Putative AsnC family transcriptional regulator
RL4497	Rhi, RA, Nod	Putative transmembrane protein, coiled-coil domain

Genes are listed and described together where co-localised. Rhi = rhizosphere, RA = (primary) root attachment, Col = 5 d colonization, IT = infection thread, Nod = nodulation. Percentage identities describe similarity at the protein level (see 2.9). Protein localization was predicted using pSORTb v 3.0.2 [206].

Cell wall and surface modifying factors also feature heavily in the primary attachment requirements at pH 7.5. *RL1165* (encoding a possible cell wall protease modulator) and likely phosphatidylcholine production *pmtA* (*RL1378*) genes fit this category. The upregulation of *pmtA* in a *praR* mutant (PraR is a known attachment factor regulator, [59]) supports its role in primary attachment. *RL2489A* (encoding a likely transglycosylase), *RL4018* (encoding a likely lipid A exporter) and *hflC* (*RL3253*) and *hflK* (*RL3254*) – transmembrane serine proteases, also involved in FtsH/HflB regulation, see Table 4-12 – also fit into this category.

Interestingly, the smallest DNA polymerase subunit *rpoZ* (*RL1505*) is also required for attachment at pH 7.5. Aside from DNA synthesis, additional roles in biofilm formation (and therefore attachment) have been identified for RpoZ [337–339]. Based on the published literature, this is the first time RpoZ has been implicated in attachment in rhizobia.

The requirement for *sixA* (*RL2644*) is particularly intriguing. SixA is a phosphohistidine phosphatase that dephosphorylates Npr in *E. coli* [343]. Both Npr and ManX were required for primary attachment at pH 6.5, and an activating phospho-relay exists between Npr and ManX [311]. In *S. meliloti* loss of ManX increases succinoglycan accumulation [352] and it is known that succinoglycan production affects EPS I composition and host colonization in this *Rhizobium* [353–356]. Theoretically, dephosphorylating Npr would deactivate phosphorelay signaling, altering EPS I composition [357]. However, EPS structures display high strain-specificity and, whilst succinoglycan EPS I is well characterised in *S. meliloti*, EPS in *R. leguminosarum* is markedly different (an octasaccharide repeating unit of glucose, glucuronic acid and galactose in a 5:2:1 molar ratio, [76]). It has been shown that mutating components of the PTS^{Ntr} system in Rlv3841

globally reduces ATP-dependent ABC transporter activity in a post-translational manner, and that these mutants displayed dry colony morphology [155]. The ‘dry’ colony morphology is most likely as a result of reduced EPS secretion through an unknown ABC exporter [313]. Thus, although the effects of PTS^{Ntr} signaling are complex, it seems that regulatory control of this switch is important in facilitating primary attachment to pea roots under pH 6.5 or 7.5 conditions, and this may be linked to levels of EPS secretion and/or modification in a pH dependent manner.

The remaining genes required for attachment at pH 7.5 are largely uncharacterised and of unknown function in attachment.

Table 4-13. 11 genes identified as required (ES/DE) for primary attachment at pH 6.5 and 7.0 only.

Gene(s)	Symbiosis defects	Description
pRL100220	Rhi, RA, IT	Uncharacterised protein
pRL100242	Rhi, RA, Col	Uncharacterised protein. Cytoplasmic protein localization
pRL110043	RA	Putative transmembrane transporter protein. 99% similarity to arabinose efflux permease family protein in *R. leguminosarum* biovar *trifolii* WSM597. EPS arabinose content regulates cell aggregation in *Azospirillum* [358]
ppx (*RL1600*)	RA, Col, IT, Nod	Putative exopolyphosphatase Ppx; can function in the hydrolysis of EPS [359]
gmsA (*RL1661*)	RA	Glucomannan biosynthesis protein GmsA. Characterised factor important for polar primary root attachment at acidic but not alkaline pH [56]. See also Figure 3-5
ecfE (*RL2227*)	Rhi, RA, Col, IT	Zinc metalloprotease EcfE. STRING database indicates functional interaction with CdsA (*RL2266*, phospholipid metabolism, score 0.92) and BamA (*RL2228*, outer membrane assembly factor, score 0.87) suggesting a role in membrane lipid processing
hfq (*RL2284*)	RA, Col	RNA binding protein Hfq. Global post transcriptional regulator. Loss of Hfq in *S. meliloti* delays nodulation and reduces competitiveness for attachment to alfalfa roots [360]
RL2520	RA	Putative transmembrane protein, ABC transporter permease. Many ABC transporters are involved in lipid transport to the outer membrane [361]
RL3277	RA	Putative transmembrane protein

RL4309	RA, IT, Nod	Putative transmembrane protein. 94% identity to *R. hidalgonense* DedA family protein (CO674_30990). DedA proteins appear to function in membrane homeostasis; mutants show altered membrane lipid composition in multiple bacterial species [362].
RL4335	RA	Uncharacterised protein. Unknown protein localization.

Genes are listed and described together where co-localised. Rhi = rhizosphere, RA = (primary) root attachment, Col = 5 d colonization, IT = infection thread, Nod = nodulation. Percentage identities describe similarity at the protein level (see 2.9). Protein localization was predicted using pSORTb v 3.0.2 [206].

Among the genes required for primary attachment at both pH 6.5 and 7.0 is *pRL110043*, a likely arabinose efflux protein which increases EPS arabinose content. Exactly how this influences attachment is not clear, but the change to EPS composition is clearly important, and similar EPS profiles are important for *Azospirillum* cell aggregation [358].

RL4309, *ppx* (*RL1600*), *RL2520* and *RL2227* are all likely involved in EPS or membrane lipid modifications and stress the importance of cell surface and secreted factor changes for attachment competence at different pHs.

The involvement of glucomannan (*gmsA*, *RL1661*) fits the characterised profile for this attachment factor [56] and indicates that INSeq was able to accurately classify the gene in this case. Well characterised genes giving rise to known phenotypes when mutated were analysed in terms of INSeq classification (Table 4-4) and gave further confidence in INSeq's ability accurately classify genes using the HMM.

The remaining genes in Table 4-13 are uncharacterised.

Table 4-14. Seven genes identified as required (ES/DE) for primary attachment at pH 7.0 and 7.5 only.

Gene(s)	Symbiosis defects	Description
cycM (*RL0141*)	RA, IT, Nod	Membrane-bound cytochrome c CycM
RL0617	Rhi, RA, Col	Putative dTTP/UTP pyrophosphatase
RL1504	RA, Col	Uncharacterised protein, NYN domain. Possibly novel RNAse with regulatory role [363]. Cytoplasmic protein localization.
clpS (*RL2212*)	RA, IT, Nod	ATP dependent Clp protease adaptor protein ClpS. ClpS modifies ClpA substrate specificity [364]. A *clpS* mutant of *P. aeruginosa* showed 70% reduction in biofilm formation and significant impairment in attachment to abiotic surfaces after one hour [365]
RL2642 / dksA (*RL2643*)	RA	*RL2642* – uncharacterised protein. *RL2643* – putative DnaK suppressor protein DksA. Studies from various bacteria indicate that these chaperones can be secreted and bind ligands, contributing to cell adherence and biofilm formation [366–369]. Cytoplasmic protein localization.
RL4083	RA	Uncharacterised protein, SGHN family esterase domain. Unknown protein localization.

Genes are listed and described together where co-localised. Rhi = rhizosphere, RA = (primary) root attachment, Col = 5 d colonization, IT = infection thread, Nod = nodulation. Protein localization was predicted using pSORTb v 3.0.2 [206].

Only 7 genes are needed for primary attachment exclusively at both pH 7.0 and 7.5. Although most are of uncharacterised or unknown function, two are better defined. One of these, *dksA* (*RL2643*), encodes a cellular chaperone factor which can be secreted, potentially contributing to attachment via ligand binding [366–369], and may also play this role in Rlv3841. As this seems to be the first indication of chaperone secretion and ligand binding involved in primary attachment in rhizobia, it would be of interest to fuse this gene to a fluorescent protein and confirm its export with microscopy before further characterisation.

The second gene is *clpS* (*RL2212*), a modifier of the substrate preference of the ClpA protease [364]. ClpA (RL2213) itself is also required for attachment at pH 7.0 and 7.5 (but also DE in 1% oxygen 20mM succinate INSeq and therefore filtered from analysis, see Table A8). The proteolytic targets of ClpS modulated ClpA remain unclear [364, 365], but are of interest in understanding pH 7.0/7.5 primary attachment further.

Table 4-15. Nine genes identified as required (ES/DE) for primary attachment at pH 6.5 and 7.5

Gene(s)	Symbiosis defects	Description
pRL100162A / *pRL100163*	Rhi, RA, Col	*pRL100162A* – hypothetical protein. Unknown protein localization. *pRL100163* – uncharacterised protein, asparagine synthetase family domain. Cytoplasmic protein localization.
pRL100470	RA, Rhi, IT, Col	Uncharacterised protein. Uncharacterised protein localization.
pRL110045 / *pRL110046*	RA, Rhi, Col / RA, Rhi, Col, IT, Nod	*pRL110045* – uncharacterised protein. Unknown protein localization. *pRL110046* – putative transcriptional regulator. Cytoplasmic protein localization.
RL1805 / *degQ* (*RL1806*)	RA	*RL1805* – putative transmembrane protein. *RL1806* – Periplasmic serine endoprotease DegQ. In *B. subtilis* DegQ stimulates phosphotransfer to a transcriptional regulator affecting biofilm formation, promoting transition from a motile to sessile attached state [370]
RL2098	RA	Putative transmembrane protein
dacF (*RL2477*)	RA	Putative penicillin-binding protein DacF. Functions as a D-alanyl-D-alanine carboxypeptidase in *E. coli* and is thought to be involved in cell wall synthesis and modifications [371]

Genes are listed and described together where co-localised. Rhi = rhizosphere, RA = (primary) root attachment, Col = 5 d colonization, IT = infection thread, Nod = nodulation. Protein localization was predicted using pSORTb v 3.0.2 [206].

A small number of genes are important for attachment at pH 6.5 and 7.5, though the reasons for these non-linear pH requirements are unclear. Despite most of the genes in this category being uncharacterised or of unknown function, *degQ* (*RL1805*) seems to promote biofilm formation and *dacF* (*RL2477*) is involved in cell wall modification. It may be the case that DacF promotes different cell wall modifications at pH 6.5 and 7.5 which are required for primary attachment, or that the same modifications promote attachment under both conditions.

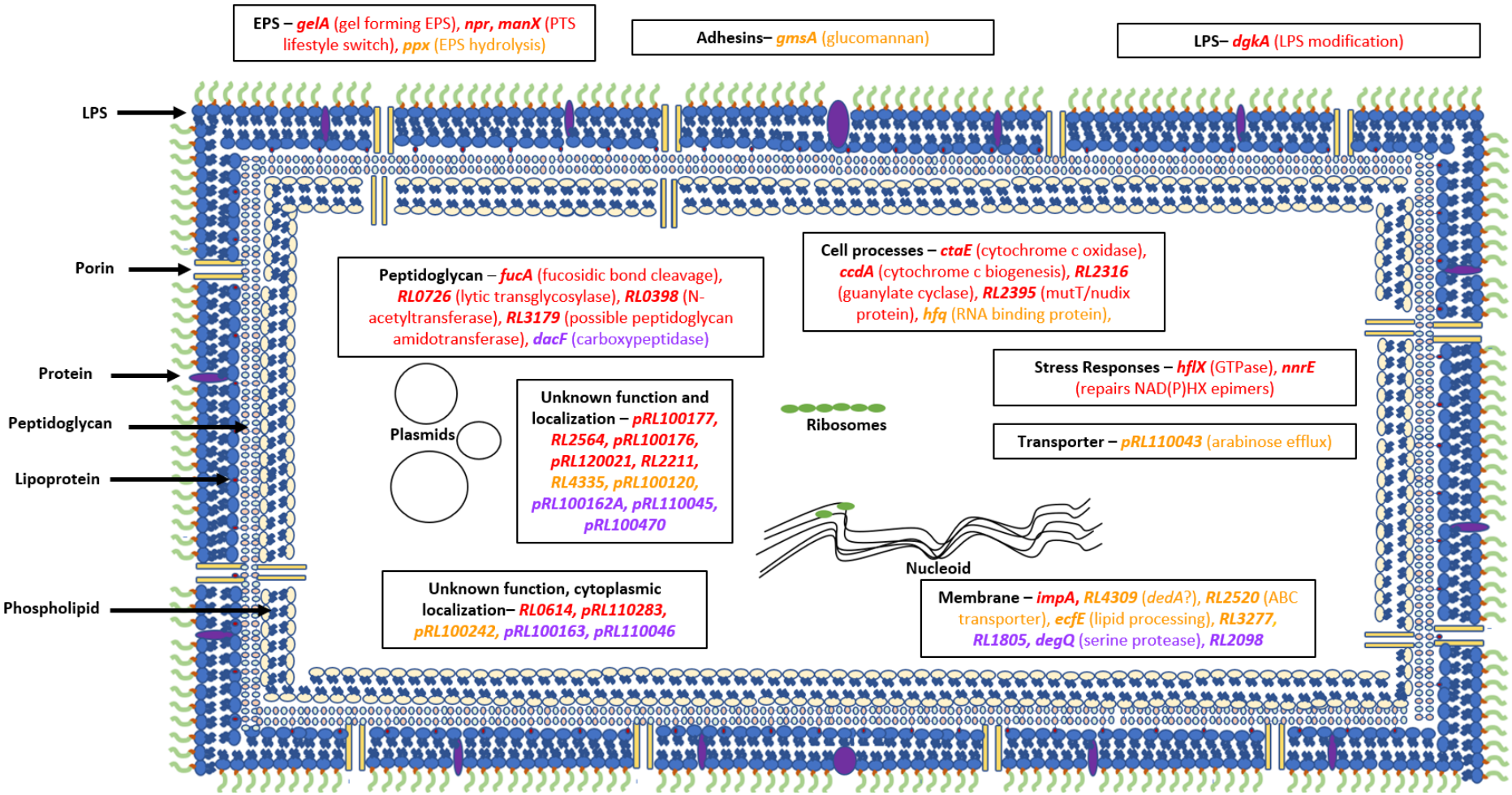


Figure 4-12. Diagram of a stylized gram-negative Rlv3841 cell showing the primary attachment determinants identified as ES/DE at pH 6.5 (red, see Table 4-10), 6.5 and 7 (orange, see Table 4-13) and 6.5 and 7.5 (purple, see Table 4-15), grouped by function. Cell outer membrane, peptidoglycan layer, inner membrane, plasmids, nucleoid and other cellular factors are shown, not to scale.

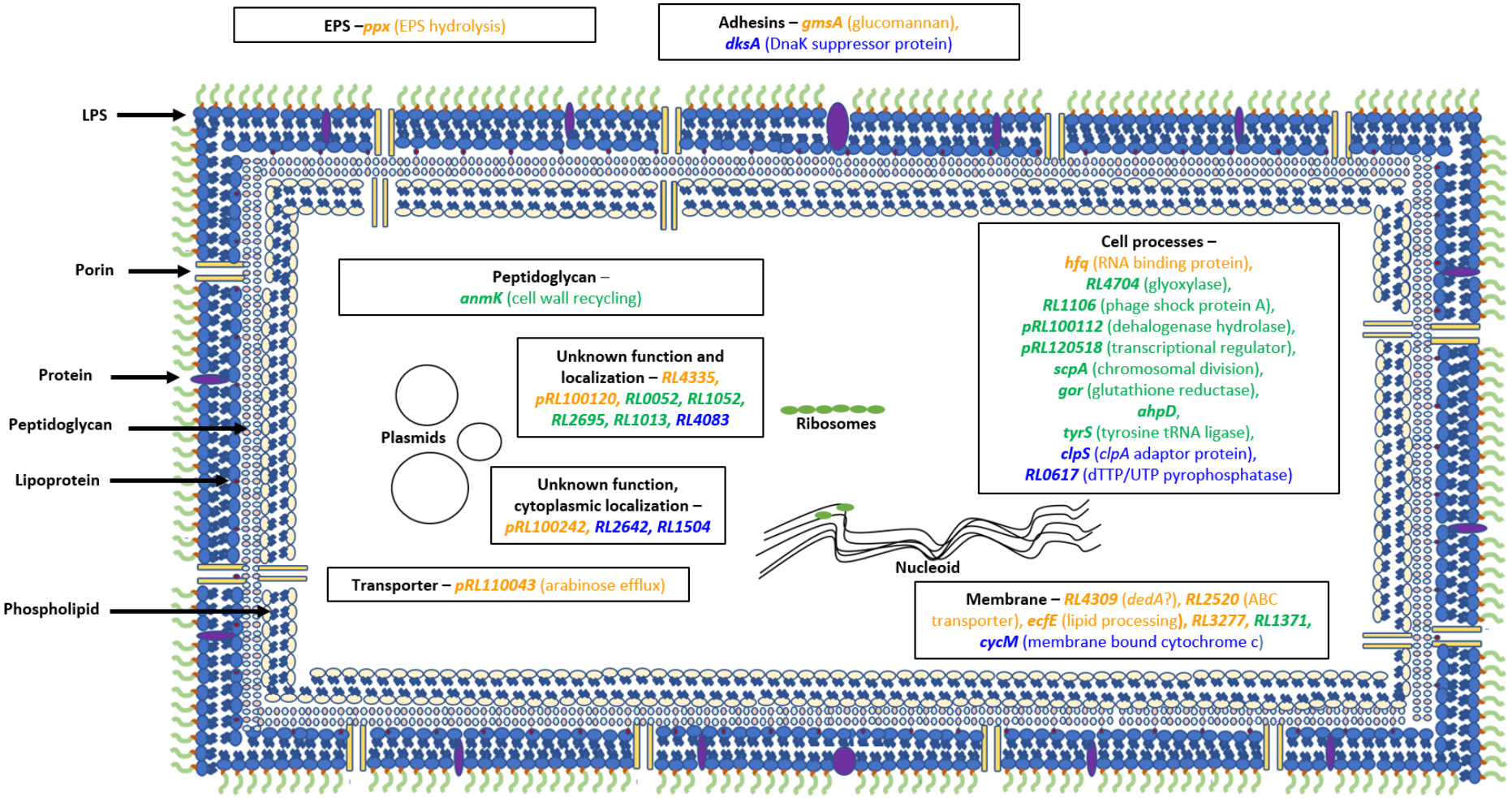


Figure 4-13. Diagram of a stylized gram-negative Rlv3841 cell showing the primary attachment determinants identified as ES/DE at pH 7.0 (green, see Table 4-11), 7 and 7.5 (blue, see Table 4-14) and 6.5 and 7 (orange, see Table 4-13), grouped by function. Cell outer membrane, peptidoglycan layer, inner membrane, plasmids, nucleoid and other cellular factors are shown, not to scale.

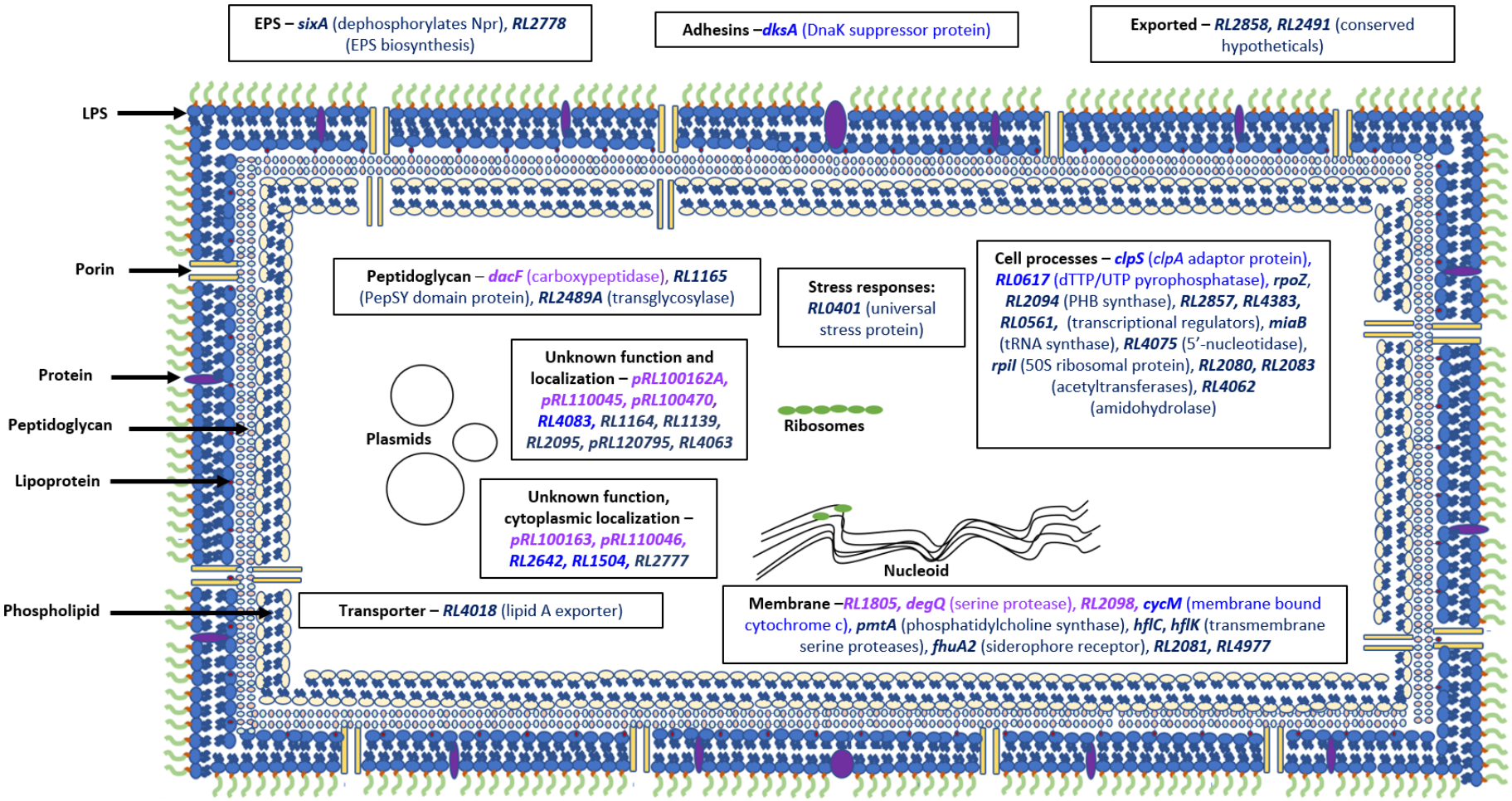


Figure 4-14. Diagram of a stylized gram-negative Rlv3841 cell showing the primary attachment determinants identified as ES/DE at pH 7.5 (dark blue, see Table 4-12), 7 and 7.5 (blue, see Table 4-14) and 6.5 and 7.5 (purple, see Table 4-15), grouped by function. Cell outer membrane, peptidoglycan layer, inner membrane, plasmids, nucleoid and other cellular factors are shown, not to scale.

4.2.12 Regulators required for primary attachment under all pH conditions

Three transcriptional regulators were required (ES/DE) for primary attachment at all pHs in INSeq (see Table 4-9 and Table A3). These were: *RL4145* (putative conserved LacI type transcriptional regulator), *RL3453* (putative two-component sensor-regulator; histidine kinase) and *mctR* (*pRL100406*, putative two component sensor/regulator, transcriptional regulator). Note that *RL3453* was classified DE in a 21% oxygen 20mM succinate INSeq (Poole Lab, unpublished data) and *mctR* (*pRL100406*) was classified AD in a VMM INSeq [175] and so neither are featured in the post-filtering list of genes given in Table 4-9. However, due to the low number of transcriptional regulators involved in root attachment at all pHs, all three were examined here. Only MctR (pRL100406) has any known function, as part of a two-component sensor-regulator system activating *mctP* expression (MctP is a transporter of alanine and other monocarboxlates) [148]. What role this may play in primary attachment is unclear.

Transcriptional regulators are often located adjacent to their regulatory targets. Genes in the vicinity of the two uncharacterised regulators were screened to see if they might represent regulated primary attachment determinants. Genes had to be in close proximity to the regulator, have an ES/DE HMM classification at one or more pHs and have annotation suggestive of possible role in primary attachment (see Table 4-16).

Although the putative regulatory targets of *RL3453* and *RL4145* are mostly conserved hypothetical, hypothetical exported or putative transmembrane (making precise assignment of function difficult), one gene, *RL3463* (a possible regulatory target of *RL3453*), has a role in lipoprotein biosynthesis.

Table 4-16. Putative targets of the transcriptional regulators RL3453 and RL4145 identified as required for primary root attachment.

Transcriptional regulator	Putative regulatory targets		HMM classification from INSeq			
	Locus	Description	Input	pH 6.5	pH 7.0	pH 7.5
RL3453	*RL3461*	Conserved hypothetical protein	DE	DE	DE	DE
	RL3462	Conserved hypothetical protein	NE	DE	DE	DE
	RL3463	Putative prolipoprotein diacylglyceryl transferase	DE	DE	ES	ES
	RL3464	Conserved hypothetical protein	DE	DE	ES	ES
RL4145	*RL4147*	Conserved hypothetical protein	DE	DE	ES	DE
	RL4158	Conserved hypothetical exported protein	DE	ES	DE	DE
	RL4159	Putative transmembrane protein	DE	ES	DE	DE

ES = essential, DE = defective, NE = neutral

To investigate transcriptional regulatory requirements further, pK19 interposon mutagenesis was used to disrupt *RL3453* and *RL4145* (see 2.4.3), and each mutant strain was labelled with the Lux plasmid (by conjugating with pIJ11282), giving OPS1967 and 1968, respectively. OPS1967 and OPS1968 were tested in a Lux root attachment assay (Figure 4-15). Both strains showed a statistically significant reduction in attachment under all conditions compared to wild-type (D5250, Rlv3841[pIJ11282]), a result which agrees with those of INSeq (Figure 4-15).

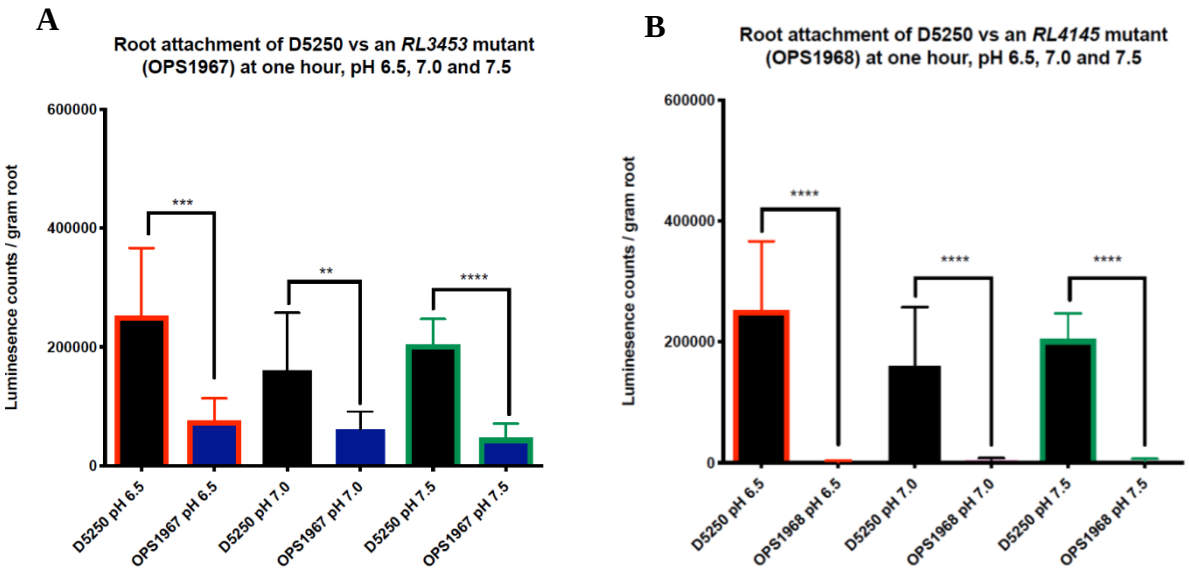


Figure 4-15. Comparison of Rlv3841and mutant strains at pH 6.5, 7.0 and 7.5 in a Lux root attachment assay. A – Root attachment of D5250 (Rlv3841[pIJ11282]) vs OPS1967 (*RL3453* mutant). B – D5250 (Rlv3841[pIJ11282]) vs OPS1968 (*RL4145* mutant). Data is displayed as mean ± SEM, n=10. An unpaired t-test was used to compare groups. ** = p <0.001, *** = p <0.0005, **** = p <0.0001

Whilst a Lux attachment assay confirms the importance of *RL3453* and *RL4145* for primary root attachment, RNASeq of these mutant strains combined with protein-protein and/or protein-DNA interaction studies, among other techniques, would be required to comprehensively elucidate their regulatory targets.

4.2.13 Using INSeq to investigate rhicadhesin

The proposed rhicadhesin primary attachment factor mediates attachment under alkaline conditions and is a protein of ~14 kDa [34, 53, 71, 210]. Following the inconclusive attempts to identify rhicadhesin using proteomic approaches (see 3.2.2), INSeq primary

attachment results were screened to identify rhicadhesin candidates. As glucomannan (*gmsA*) (the proposed 'partner' of rhicadhesin, mediating attachment under acidic conditions [56, 71]) was required for root attachment both at pH 6.5 and 7.0 in INSeq (see Table 4-13), it follows that rhicadhesin should be required at one or both of pH 7.0 and 7.5. A size filter of 12-16 kDa was applied. Based on these criteria the candidates in Table 4-17 were identified (first seven entries, above red line). The INSeq HMM classifications for the eight rhicadhesin candidates identified in LC-MS/MS (in 3.2.2, Table 3-1) that were not predicted to be in the cytoplasm or cytoplasmic membrane are also included (Table 4-17, below red line).

Table 4-17. Rhicadhesin candidates identified from INSeq (above red line) and proteomics (below red line)

Gene	Protein description	Protein size (kDa)	INSeq classification			
			Input	pH 6.5	pH 7.0	pH 7.5
RL1165	Conserved hypothetical protein	12	NE	NE	NE	DE
RL1166	Putative ribonuclease-L-PSP family protein	14	NE	NE	NE	DE
RL1339	Conserved hypothetical protein	15	NE	NE	NE	DE
RL2095	Conserved hypothetical protein	15	NE	NE	NE	DE
dksA (*RL2643*)	Putative DnaK suppressor protein	15	NE	NE	DE	DE
RL2858	Conserved hypothetical exported protein	13	NE	NE	NE	DE
RL4383	Putative AsnC family transcriptional regulator	16	NE	NE	NE	DE
RL0770	Putative phasin, phasin-2 superfamily	16	NE	NE	NE	NE
ndk (*RL1580*)	Nucleoside diphosphate kinase	15	NE	NE	NE	NE
RL1635	Putative outer membrane protein	19	DE	DE	DE	DE
pal (*RL3968)*	Putative OmpA family peptidoglycan associated lipoprotein	19	DE	ES	DE	ES
omp19 (*RL4441*)	Outer membrane lipoprotein Omp19	18	DE	DE	DE	DE
RL4733	Conserved hypothetical protein	17	DE	ES	DE	ES

ES = essential, DE = defective, NE = neutral.

There was no overlap in the rhicadhesin candidates identified from INSeq and proteomic analysis using the criteria described (Table 4-17, 3.3.2), showing that none seem to fit the properties required for a rhicadhesin (required for attachment only under alkaline or neutral-alkaline conditions [34, 53, 71]). However, this result may be misleading, as rhicadhesin is hypothesised to be a root hair specific attachment factor. Given that the INSeq experiment examined attachment to the entire root (and root hairs likely constitute a relatively small part of this in terms of surface area), genes required only for attachment to root hairs, or root hairs in specific root zones, may be misclassified and therefore overlooked as rhicadhesin candidates. It is unclear as to why some rhicadhesin candidate genes identified in proteomics should be classified DE in the input (Table 4-17), as rhicadhesin is not thought to be an integral component of the bacterial cell [53]

Although proteomics results do not identify rhicadhesin, the candidates extracted from the INSeq data (Table 4-17, above red line) are informative. These demonstrate that multiple factors in the 12-16 kDa range display the pattern of defects expected for rhicadhesin and contribute to attachment under alkaline conditions. They may not all be direct adhesins in themselves; effects may be indirect, but this seems less likely for candidates such as *dksA* (*RL2643*– see Table 4-14) and *RL2858* (encoding a conserved hypothetical exported protein). As none of these were identified via proteomic approaches, it is impossible to say whether any of these represent the rhicadhesin(s) first putatively isolated by Smit *et al.* (1989) [53]. As rhicadhesin is thought to be important for attachment to root hairs, more detailed attachment studies (using high magnification microscopy) would be needed to characterise the spatial attachment characteristics of mutants in these genes at alkaline pHs.

INSeq could be applied to identifying primary attachment factors directly involved in root hair attachment at different pHs. Ramos *et al.* (2003) and Qiao *et al.* (2013) describe the isolation of purified root hairs from *M. truncatula* and *G. max* using liquid nitrogen [372, 373]. A primary attachment INSeq followed by isolation of root hairs and recovery of bacteria would identify attachment factors required for root hair attachment. However, there are technical challenges. Ramos *et al.* (2003) reported 40 μg root hair yield from-100 *M. truncatula* root tips, whilst this was 800-1000 mg from 60 whole soybean roots for Qiao *et al*. (2013) [372, 373]. The yield of root hairs, number of bacteria attached, and whether bacteria remain attached during root hair isolation all represent significant challenges to experimental design.

Whilst this approach could be used to identify rhicadhesin, the caveats of previous research and the technical challenges associated with root-hair INSeq meant that the characterisation of rhicadhesin was not pursued further in this work. However, pursuing the characterisation of rhicadhesin as part of this work has enabled far greater insights into the vast array of root attachment factors utilized by Rlv3841 in attachment to pea roots at different pHs.

4.3 Conclusion

In summary, experimental parameters for a primary root attachment INSeq experiment at different pHs were determined. Approximately 87 million barcoded sequencing reads were obtained from twelve sample libraries with high (82%) insertion density. HMM classifications were obtained for 7,319 (99.7%) of Rlv3841 genes. Input library data was curated using a published dataset [151] to account for a TY bacterial recovery step

following isolation from plant roots; this left 86% of the genome classified as NE in the input library.

INSeq results were compared with the phenotypes of mutants characterised in the literature. Lux data for mutant strain attachment from Chapter 3 was also compared to the INSeq gene classifications obtained in this chapter. There was a broad overlap between the INSeq and Lux attachment assay results, although there were some discrepancies. In one case (*praR*), this was likely due to low TA site number reducing HMM classification accuracy. The disagreement of Lux and INSeq results for two motility mutant strains (OPS1738, *flgE* and OPS1739, *motA*) was surprising, but could possibly be explained by the differing nature of the experimental assays. Lux assays use a homogenous single inoculum while INSeq involves inoculation with a mixed population of mutants. This was clearly a factor in the case of an *RL2969* mutant. An *RL2969* mutant (strain OPS1235) was identified as attachment-deficient in a single inoculum Lux assay but was NE in the INSeq experiment. Rescue by *in trans* complementation was suggested to be occurring in INSeq, as the OPS1235 mutant strain was similarly able to be rescued in a Lux attachment assay with 1:100 (mutant:wild-type) inoculation. After accounting for this (and with the regulator mutants tested in Lux assays in 4.2.12), Lux and INSeq showed a ~75% agreement rate using the defined criteria (see 4.2.9).

In total, 292 genes were identified as required (ES/DE) for primary attachment at one or more pHs, with over half of these required specifically at one pH (50 at pH 6.5, 44 at pH 7.0 and 58 at pH 7.5). Examining functional classifications revealed that genes encoding proteins involved in cell processes, cell membrane/envelope and macromolecule metabolism, as well as a high proportion of uncharacterised genes, were involved in primary attachment. Comparison of the genomic localization of primary attachment vs 5-

day colonization genes that seemed to be important included some primary attachment factors that may be constitutively expressed, as opposed to secondary attachment determinants which are largely inducible.

Using data from INSeq experiments across symbiosis development stages, it was possible to determine the genes required only for primary attachment (104) or for primary attachment and subsequent symbiosis stages (49). This is useful in identifying which factors may be involved in attachment to bulk root epidermis (generalist attachment factors) as oppose to root hairs, from which infection threads form. The latter can be considered symbiosis determining primary attachment factors.

INSeq data from experiments with different growth media or metabolic conditions was leveraged to increase the specificity of primary attachment factor identification. This filtering reduced the 292 identified factors to 116, with over half specific to one attachment pH. Examining these in more detail shed light on a range of different factors used by Rlv3841 for primary attachment to pea roots in a pH dependent fashion. A summary of the different gene functions needed at different pHs for the primary attachment of Rlv3841 to pea roots is shown in Figure 4-16.

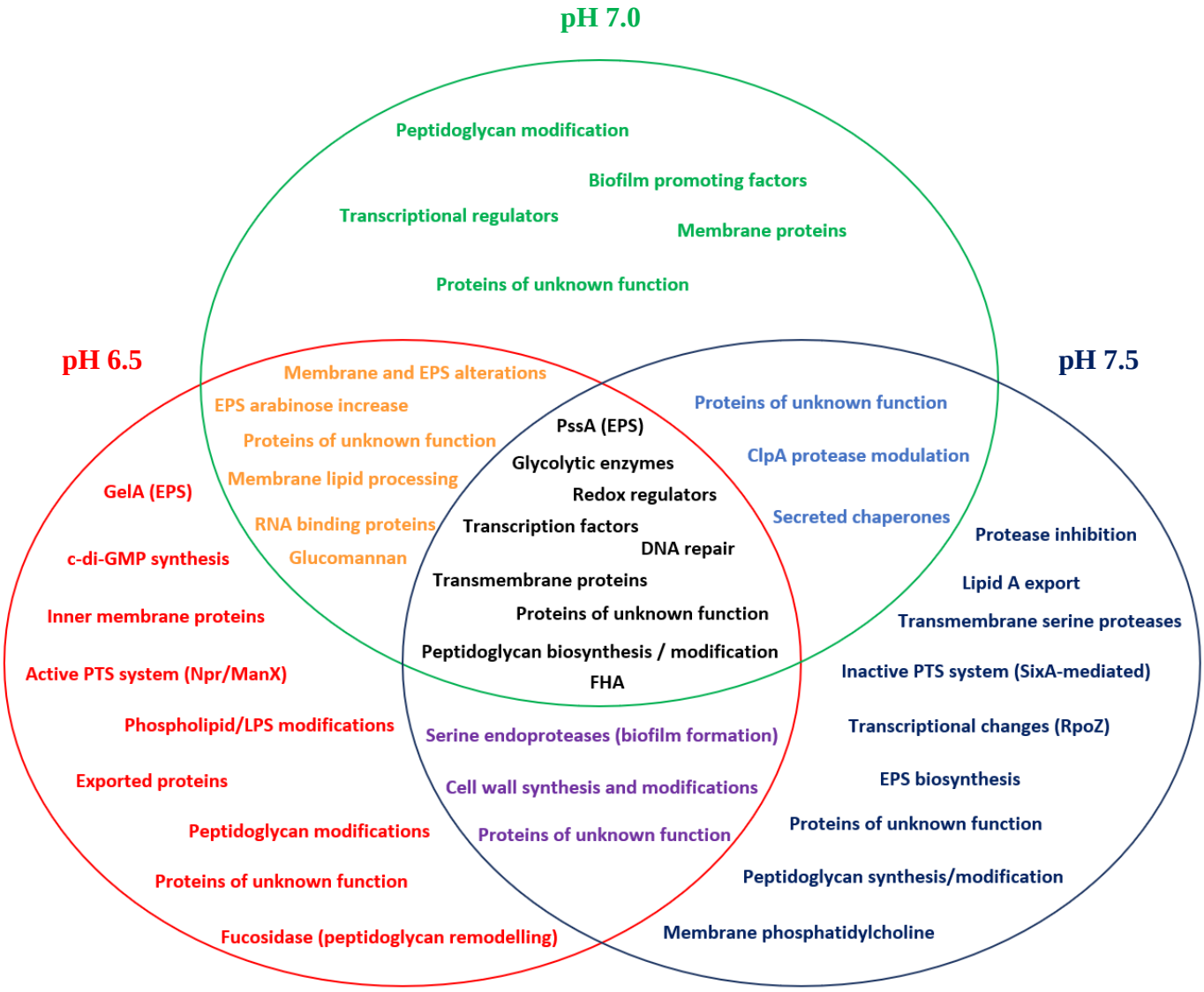


Figure 4-16. Gene functions needed for primary attachment of Rlv3841 to pea roots at pH 6.5, 7.0, 7.5 and combinations thereof. Circle color indicates pH: red = pH 6.5, green = pH 7.0, dark blue = pH 7.5. Text color indicates pH requirement of indicated gene function for primary attachment: red = pH 6.5, orange = pH 6.5 and 7.0, green = pH 7.0, blue = pH 7.0 and 7.5, dark blue = pH 7.5, purple = pH 6.5 and 7.5, black = pH 6.5, 7.0 and 7.5. Gene functions are drawn from Tables 4-9 to 4-15, see also Figures 4-11 to 4-14.

For attachment at all pHs, the requirement for PssA highlighted the core importance of EPS. The demonstrated requirement for FHA (*RL4382*) is novel and should be built into models of primary attachment after further corroborating evidence is obtained. Similarly, moonlighting glycolytic enzymes (particularly TpiA) also seem to be important, as do likely peptidoglycan modification factors (*RL4362*/*RL4363*). Further genes required with functions in cell stress responses (*hslO – RL0551*) and DNA repair (*recA – RL2637*) may reflect stresses encountered in the assay, and similar stresses may also be encountered under field conditions. Transcription factors (notably *RL4145*) are also likely required, though regulatory targets need to be identified in future transcriptomic studies.

At pH 6.5, EPS, LPS and peptidoglycan modifications were required for attachment, notably glycan fucosidic bond cleavage (*fucA, pRL100274*), diacylglycerol kinase activity (*dgkA, RL2780*, linked to phospholipid recycling and LPS modifications [324]) and transglycosylase mediated cell wall alterations (*RL0726*). Of importance was the demonstration of GelA (*RL4404*) function as a primary attachment factor at pH 6.5, likely involved in bulk root epidermis, as oppose to root hair attachment.

At both pH 6.5 and 7.0, pRL110043 (an arabinose efflux protein) indicated a specific EPS structural change needed for attachment, whilst other EPS and membrane lipid modifications were also necessary. These included *ppx* (*RL1600*), a putative exopolyphosphatase, and *RL4309*, a likely DedA family protein which functions in membrane homeostasis [362]. The RNA binding protein Hfq (encoded by *RL2284*) is also required. This is of interest, as loss of this factor has been shown to reduce competitiveness for attachment to alfalfa roots in *S. meliloti*, likely through metabolic pathway perturbation and loss of stress-associated chaperone regulation [360]. INSeq

highlighted the pH dependent profile of glucomannan activity reported previously [56], demonstrating the resolving power of this technique.

AnmK (RL2587, which cleaves peptidoglycan structures, [331, 332]) was required at pH 7.0. Also needed were biofilm promoting factors such as *pspA* (*RL1106*), which functions to promote biofilm formation by stabilizing the cytoplasmic membrane [329], and membrane proteins including *RL1013*, which features a protein motif also present in surface antigens of other bacterial species [328].

The ClpS/ClpA protease system was needed at pH 7.0 and 7.5, although the proteolytic targets remain unclear. As with TpiA, the DnaK suppressor protein DksA seems to have a moonlighting role, being secreted to bind ligands and contribute to cell adherence at these pHs.

At pH 7.5, higher membrane phosphatidylcholine requirements were highlighted (PmtA), along with various protease-driven cell wall changes (Table 4-12). These include those induced by *hflC* (*RL3253*) and *hflK* (*RL3254*), transmembrane serine proteases. As well as their own proteolytic activity, HflC and HflK also modulate HflB, a metalloprotease known to be involved in regulation of membrane proteins and biofilm formation in *E. coli* and other bacteria [345, 346]. HflB itself (encoded by *RL3965*) is classified ES/DE in the input library and at all attachment pHs. As HflC and HflK are only needed pH 7.5 (whilst HflB is needed at all pHs) it seems that HflC and HlfK proteolytic activity and/or modulation of HflB activity is occurring to promote pH 7.5 attachment. Another novel finding was the requirement for RpoZ. A DNA polymerase subunit, RpoZ has been linked to transcriptional regulation of attachment factors [337–339] and is likely to be functioning similarly in Rlv3841.

The requirement for SixA at pH 7.5 and Npr/ManX at pH 6.5 seems to indicate an opposing signalling system. The PTS^{Ntr} system (of which these components are part) controls the complex metabolic phenotypes. At pH 6.5, Npr/ManX requirements indicate EPS secretion through ATP dependent ABC transporters [313]. At pH 6.5, SixA likely deactivates this system, reducing transporter activity [343]. This may be compensated for by other factors required at pH 7.5, such as PmtA, RL2489A and RL4018 (likely lipid A exporter). A summary of the different gene functions needed at different pHs for the primary attachment of Rlv3841 to pea roots is shown in Figure 4-16.

These results largely indicate the disparate cell wall, EPS and LPS requirements for primary attachment to pea roots at different pHs and demonstrate a greatly increased complexity of primary attachment mechanisms compared to the published literature. Two uncharacterised regulators, encoded by *RL3453* and *RL4145*, were needed for attachment at all pHs. Using genomic proximity and HMM classifications to infer regulatory targets indicated that exported, transmembrane and lipoprotein biosynthesis proteins were likely to be regulated by these proteins. Mutants in *RL3453* and *RL4145* (OPS1967 and OPS1968) were confirmed to be deficient in root attachment using Lux whole-root attachment assays with a single inoculum. These mutants will be useful for RNASeq to investigate the transcriptional networks important in primary attachment further.

Using the INSeq dataset to investigate possible rhicadhesin genes identified seven candidates encoding proteins between 12-16 kDa which were required for attachment only at pH 7.0 and/or 7.5. There was no overlap between these genes and those previously identified from proteomic analysis of a protein preparation band thought to contain rhicadhesin. Whilst this does not necessarily mean proteins identified by proteomics

cannot be a rhicadhesin, it does highlight the numerous caveats of previously published [53, 54, 208–211] rhicadhesin research:

I) defining an adhesin based on ability of a protein preparation to inhibit attachment does not conclusively demonstrate it is an adhesin

II) Given the high number of factors involved in Rlv3841 root attachment, preparation of a membrane fraction may isolate root attachment factors, but this does not indicate they are more important for primary attachment than any other factor

III) There may be multiple functional genes encoding rhicadhesin, or multiple rhicadhesins

IV) Smit *et al*. (1989) never conclusively demonstrated they were working with a single purified protein; their 'purified adhesin' could have been a protein mixture [53]

V) There are multiple Rlv3841 factors of 12-16 kDa that contribute to attachment at pH 7.0 and/or pH 7.5, all of which could be considered rhicadhesins

Whilst pursuing the characterisation of rhicadhesin as part of a wider strategy to identify root attachment factors has been very informative, the identity of the protein(s) isolated by Smit *et al*. (1989) [53] remain unclear. Given the number and diversity of Rlv3841 primary attachment factors, they may well have isolated a protein that plays a role in this process, but whether in a direct or indirect capacity is unknown. Pending further, in depth characterisation of small surface localized root attachment factors (particularly for calcium binding activity and involvement in attachment specifically to root hairs, postulated characteristics of rhicadhesin - Smit *et al*. (1989) [53]) it seems unhelpful to describe primary root attachment in Rlv3841 as a two component, pH-dependent glucomannan/rhicadhesin system. More holistically, it should be thought of as a process

employing diverse cellular factors, with a focus on EPS, LPS and peptidoglycan biosynthesis and modification, as well as direct adhesins (such as FHA), membrane proteins and uncharacterised factors, all of which show different pH dependencies. Whilst glucomannan is clearly important for attachment at acidic and neutral pHs [56] (also seen here in a Lux attachment assay – Figure 3-6 – and INSeq), there is likely to be a multitude of other factors that contribute to root hair attachment at alkaline pHs.

This chapter highlights the power of INSeq for investigating early stage symbiosis determinants on a genome-wide scale and has revealed far greater complexity than previously known in primary attachment mechanisms.

Chapter 5

Genome-scale characterisation of the primary attachment determinants of *R. leguminosarum* to roots of a non-host legume and non-legume

5.1 Introduction

The first physical interaction in *Rhizobium*-legume symbioses occurs when bacteria attach to their hosts roots [26]. A comparative lack of research into this interaction [25] means that relatively few factors involved in primary root attachment have been fully characterised. The use of INSeq to investigate primary root attachment determinants at different pHs in Rlv3841/pea symbiosis (described in Chapter 4 of this thesis) revealed far greater complexity than previously thought. By enabling genome-wide characterisation of attachment determinants in a single screen, it provided a powerful tool to unearth novel biological roles in symbiosis formation for >100 genes.

It is not just attachment leading to nitrogen-fixing symbioses (in which rhizobia provide fixed nitrogen in return for carbon sources [25]) which can bring benefits to rhizobia. The ability to attach to bulk root epidermis, or the roots of non-host plants, neither of which results in nodule formation and nitrogen-fixing symbiosis development, is also likely to give nutritional benefits as exudates from roots can serve as preferential growth substrates for rhizobia [34]. This interaction could be considered a symbiosis in itself, as the host plant may also derive benefits. For example, *R. leguminosarum* is able to colonise the roots of tomato and pepper plants and promote plant growth, likely through indole acetic acid and siderophore production [38]. Further studies have reported similar effects of *R. leguminosarum* in maize, lettuce, rice, canola and more [36–38]. Plant growth promoting rhizobacteria (PGPR) may also confer benefits by protecting plants from pathogens. *R. leguminosarum* biovar *viciae* inoculation of pea gives protective effects against *Didymella pinodes* pathogenesis [43]. The protective effects against *D. pinodes*-induced wilt disease are in addition to the plant growth promotion resulting from nitrogen-fixing symbiosis formation [43]. Sistani *et al*. (2017) demonstrated that inoculation of pea seeds with *R.*

leguminosarum biovar *viciae* induced changes in the seed proteome and metabolome. By inducing higher ROS responses, plant cell wall structural changes and production of antimicrobial seed flavonoids and triterpenoids (notably soyasapogenol), significant reductions in *D. pinodes* infection were observed [43]. In addition, different rhizobial isolates have been reported to protect chickpea against *Fusarium* wilt [374]. In this case Arfaoui *et al*. (2006) screened 21 field isolates of *R. leguminosarum* biovar *ciceri*. 17 were able to directly inhibit *Fusarium oxysporum* growth and protect plants from disease via production of volatiles and/or cyanide and some also displayed phosphate solubilising capabilities, meaning plant growth promotion could also occur via nutrient provision [374]. Recent imaging work has even demonstrated that, at least in the short term (up to 12 hours post-inoculation), *B. subtilis* (acting as a PGPR) was able to physically exclude *E. coli* from interacting with *Arabidopsis* roots, suggesting a possible physical protection against plant pathogens is possible in some cases [188]. Such physical exclusion has been documented for *B. subtilis* acting against the plant pathogen *P. syringae* pv tomato DC3000, which is able to infect Arabidopsis roots. Bais *et al*. (2004) demonstrated that root protection was effected by *B. subtilis* forming extensive biofilms on the root surface (thus physically preventing pathogen-root contact) and secreting surfactin, a lipopeptide antimicrobial which is toxic to *P. syringae* [375].

Intriguingly, recent microbiome research has demonstrated that the presence of a nitrogen fixing symbiosis, in this case between *L. japonicus* and *M. loti*, is able to influence the wider microbiome structure of organisms interacting with roots in the rhizosphere. In plants deficient in symbiosis formation the microbiome structure is dramatically altered, implying that a legume host selects a broad taxonomic range of root associated bacteria that could contribute to plant growth [121].

Attachment of different bacteria to roots is a diverse process with different factors thought to be involved, which show considerable variation [34]. For example, whilst *Agrobacterium* rely on adhesive pili and capsular polysaccharide (among other factors) for root attachment, *Pseudomonas* is able to use outer membrane porins as well as pili and flagella, and LPS and outer membrane proteins plays a role in *Azospirillum* [34] (see also Figure 1-3).

As shown in Chapter 4, the factors required for attachment of Rlv3841 to the roots of one plant (pea) also show considerable diversity under different conditions (in this case, pH changes). A good example of this is the variation seen in EPS. EPS is known to be of crucial importance in attachment of Rlv3841 to roots [58, 77]; in this thesis a mutant in *pssA* (*RL3752*), a glycosyltransferase involved in the first step of EPS biosynthesis [288] was seen to be deficient in root attachment both in a Lux whole-root attachment assay (Figure 3-5) and using INSeq (Table 4-9). However, EPS requirements are more nuanced than this. Several genes involved in EPS processing were required for primary root attachment at different pHs. These included *gelA* (*RL4404*, gel forming EPS production protein important in attachment at pH 6.5), *ppx* (*RL1600*, putative exopolyphosphatase functioning in EPS hydrolysis [359], important at pH 6.5 and 7) and *RL2778* (a putative EPS biosynthesis protein, important for attachment at pH 7.5). In addition to processing of EPS, multiple genes were identified as required for attachment at different pHs with likely roles in LPS and peptidoglycan modification, as well as functional roles in the membrane, highlighting the importance of polymeric substances and the bacterial cell membrane and wall structures in primary attachment. There is considerable diversity in EPS structures between different *Rhizobium* (*R. leguminosarum* biovars *triofolii* and *viciae* and *S. meliloti*

all demonstrate distinct EPS structures [376]), and this can influence the specificity of the nitrogen fixing symbiosis [96].

Given that so many different factors are required to attach to the roots of pea at different pHs (Chapter 4), and that different bacteria use different factors for primary attachment to different plant roots [34], it may also be the case that different Rlv3841 factors are required to attach to the roots of different plants.

Given the broad role of PGPR-root interactions in increasing crop yields in diverse plants and eliminating pathogens (see also [39–42]), improving our understanding of root-bacteria interactions is of major interest to both basic and applied plant research.

Very little is known about the mechanistic determinants of interactions of rhizobia with non-host plants in terms of primary attachment factor requirements. Albareda *et al* (2008) examined the effects of salt concentration, culture age, pH and inoculum composition on total numbers of rhizobia colonising the roots of non-host plants [377]. Further, Villacieros *et al* (2002) investigated the colonization of *P. fluorescens* in the alfalfa rhizosphere and how this was influenced by the presence of *S. meliloti* [378]. This demonstrated that, whilst both strains colonize the alfalfa rhizosphere efficiently alone, in co-inoculation *S. meliloti* came to dominate the rhizosphere and reduced the deleterious effects of *P. fluorescens* single inoculation on alfalfa growth. 16S rRNA sequencing has been used to investigate the rhizosphere community influence of rhizobial strain co-inoculation on *Dalbergia odorifera* roots (demonstrating that different rhizobial inoculums cause different shifts in rhizosphere strain abundance [379]). Further, the influence of *Bradyrhizobium* on arbuscular mycorrhizal fungi colonization of soybean roots has also been characterised, showing that bacterial co-inoculation could significantly enhance fungal colonization and increase plant dry weight [380]. However, whilst

providing overviews of changes in community structure and levels of attachment to roots under different conditions, among other parameters, these studies provide no mechanistic information in terms of how bacteria physically attach to the roots of different plants. The large diversity seen in *Rhizobium* EPS structures (given as an example above) hint that interactions with different plant species may involve very different primary attachment determinants. This idea is reinforced by the different primary attachment determinants used by different rhizobacteria when attaching to plant roots [34] (see also 4.1). In this chapter, the use of INSeq to investigate the primary attachment determinants of Rlv3841 to soybean (a non-host legume) and barley (a non-legume) roots at pH 7.0 is described. The major aim of this work was to investigate whether primary attachment determinants of Rlv3841 are unique to different plants, or whether mechanistic overlap is seen, and to improve our understanding of root-bacteria interactions. By combining this data with a subset of the pea primary root attachment data from Chapter 4, the core determinants required for primary attachment to multiple plant species and the variable determinants, which show specificity depending on the plant species, are defined. To the best of my knowledge, this represents the first study to define the determinants of primary attachment to roots of different plants for a given rhizobial species.

The raw data for this INSeq experiment can be found in Appendix 2, Table 2.

5.2 Results and discussion

5.2.1 Attachment assays of Rlv3841 – determining inoculum density and bacterial recovery method for INSeq

Attachment assays with Rlv3841 (2.6.5) were carried out to determine the optimum starting inoculum density for INSeq after recovery of root-attached bacteria for both soybean and barley (Appendix 1, Figures A1 and A2, see also Figure 4-2 for pea). As was the case for pea (where recovered bacteria rose with increasing inoculum density, leading to the recovery of ~ 2 x 10^6 cells/root after 1 hr with 50 mL of an OD_{600} 0.1 inoculum), higher starting inoculum density gave greater numbers of Rlv3841 cells recovered from roots after one hour.

As discussed in 4.2.1, there is a minimum fold-coverage requirement (500-750) of genes in INSeq for robust downstream HMM classification. The numbers of plants required to achieve this 750-fold coverage or higher for soybean and barley is calculated using the equation from 4.2.1 and the data in Appendix 1, Figures A1 and A2. For a given starting inoculum density, Table 5-1 provides the number of plants needed for 750-fold coverage for both soybean and barley.

Table 5-1. Number of plants needed for 750-fold coverage of 7,300 genes in the genome of Rlv3841 based on input inoculum density (OD_{600}) for soybean and barley

Input inoculum (OD_{600})	**Plants needed for 750-fold coverage of 7,300 genes**	
	Soybean	**Barley**
0.001	30	18
0.01	16	10
0.1	1	2

Given the results presented in Table 5-1 and the setup of Lux attachment assays and INSeq work presented in Chapter 4, it was decided to use a starting inoculum of OD_{600} 0.1 with 10 plants per test condition. Each test condition was repeated in triplicate (see 4.2.2). This starting inoculum density and plant number should allow for > 3,000 fold coverage of the ~7,300 Rlv3841 genes for each plant.

As in Chapter 4, no statistically significant difference was seen between the number of Rlv3841 recovered by vortexing compared to vortexing and grinding of roots for either plant (for soybean shown in Appendix 1, Figure A3 and barley in Appendix 1, Figure A4), meaning that vortexing alone should be sufficient to recover Rlv3841 cells in INSeq primary root attachment assays.

5.2.2 INSeq experimental design

To characterise Rlv3841 primary attachment determinants with barley and soybean roots, an INSeq experiment was designed (Figure 5-1). Results were compared with primary attachment determinants identified using INSeq for pea roots at pH 7.0 (Chapter 4). An OD_{600} 0.1 (10^8 cfu/mL) population of insertion mutants (library 1-3, 500 mL) was used to inoculate 10 soybean or barley plants in triplicate. Insertion mutants were recovered from roots after 1 hr and grown in liquid TY for 12 hr before the 9 sample libraries (2x3 experimental conditions + 3 input libraries) underwent barcoding, *mariner* library preparation and sequencing (2.7.3) (Figure 5-1). HMM gene assignments were averaged across three replicates for each plant as described in 4.2.2.

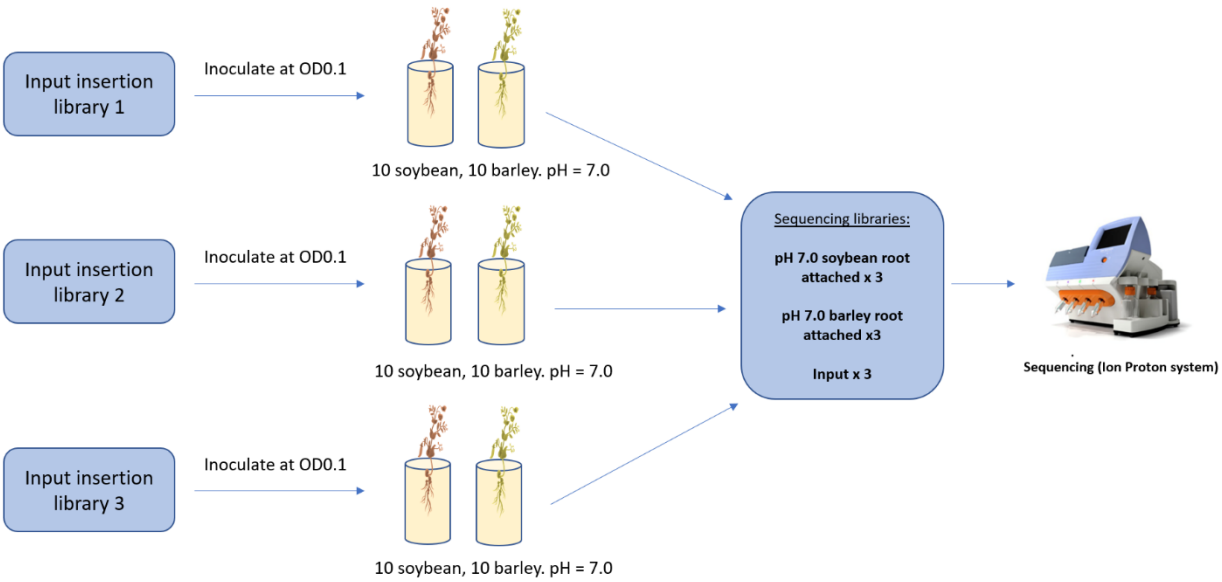


Figure 5-1. INSeq experimental design, 50mL of insertion mutant inoculum (library 1-3, OD_{600} 0.1) was inoculated onto each soybean or barley root at pH 7.0 and incubated for 1 hr before recovery by vortexing (resulting in three replicates). Recovered bacteria for each attachment group were pelleted and grown on TY for 12 hours to increase bacterial gDNA concentration (and effectively decrease plant DNA contamination) before DNA

preparation (not shown). Input library gDNA samples were extracted directly from input inoculum. Each sample (input and 3 x root-attached) underwent DNA extraction and library preparation before sequencing (see 2.7.2, 2.7.3 and 2.7.4).

5.2.3 HMM genome analysis

For primary attachment to roots of soybean or barley, HMM analysis of sequencing reads assigned each of the 7,319 genes in the Rlv3841 genome to one of the following categories, based on how gene mutation affects bacterium fitness: ES, DE, NE or AD (see Figure 4-1), as in the previous chapter for pea plants. In total, 127 million barcoded sequencing reads were obtained from a total of nine samples (six from plants and three input libraries). The insertion densities (proportion of total TA motifs that contain at least one insertion in each sequenced library) for libraries were as follows: input – 87 %, soybean pH 7 – 54%, barley pH 7 – 56%. These insertion densities are more than sufficient, based on multiple reviews of Tn-Seq and INSeq data analyzed using a HMM approach and their respective library insertion densities, for robust statistical analysis and assignment of HMM gene classification [146, 150, 169, 179].

5.2.4 INSeq gene classifications

The distribution of HMM gene classifications across the four different states (ES, DE, NE and AD) in input and root-attached libraries for the two plants is shown in Table 5-2. Note that for the chromosome or any plasmids individually the insertion density obtained did not differ more than ± 3% from the average input figure given in Table 5-2, indicating unbiased *mariner* insertions across the genome.

Table 5-2. Distribution of HMM assignment of Rlv3841 genes in the input and root-attached (pea, soybean and barley, pH 7.0) libraries.

HMM gene assignment	Library condition			
	Input	**pH 7.0 pea**	**pH 7.0 soybean**	**pH 7.0 barley**
ES	1 %	6 %	7 %	8 %
DE	10 %	7 %	8 %	6 %
NE	86 %	86 %	84 %	85 %
AD	3 %	1 %	1 %	1 %

Data for pea plants grown at pH 7.0 is from Chapter 4 (see Table 4-3, section 4.2.4). Values are given as percentage of the 7,319 genes which contain TA sites.

As explained in the previous chapter, due to the nature of the experimental design, the input library data was curated using the published dataset from Perry and Yost (2014) [151] to account for the 12 hour TY growth step for recovered root-attached populations. This 12 hr growth step was used for bacteria recovered from roots to increase bacterial gDNA concentration and dilute contaminating plant DNA before DNA extraction. Curation removed any inadvertent selection on genes from this step which could affect downstream HMM classification (see also 4.2.4). When using INSeq to identify primary root-attachment determinants, it is genes that are NE in the input but non-NE in attachment which are of interest.

5.2.5 Literature validation of INSeq predictions

In order to provide preliminary validation of the HMM gene classifications for soybean and barley root attachment, the INSeq classifications of well-characterised genes (see 4.2.5) were compared with their known or predicted phenotypes according to previous publications. It was not expected that HMM classifications should be plant dependent. These genes are listed in Table 4-4, and include genes encoding peptidoglycan biosynthesis factors, ribosomal subunits, DNA helicases and polymerases, all critical for cell viability. Mutation in these genes leads to an ES or DE HMM classification in attachment to soybean and barley, as is the case for pea. Similarly, genes encoding non-essential factors, such as the putative manganese transporter *mntH* (*RL0940*) are classified as NE in attachment to soybean and barley, as well as pea. Classifications for these genes in attachment to soybean and barley were the same as for attachment to pea (Table 4-4), providing both initial validation of this dataset and an indication of reproducibility in the HMM classifications.

5.2.6 Primary attachment gene requirements and functional classifications

In total, 464 genes classified as NE in the curated input libraries were classified as ES/DE in one or more of the pH 7.0 root-attached libraries (input library data from chapter 4 was used for pea, and input library data from this chapter used for soybean and barley). The distribution of these is shown in Figure 5-2, below. These genes and their HMM classifications from experiments listed in Table 4-5 (pea rhizosphere [238], 5 d root colonization [238], infection thread and nodule bacteroids - unpublished data, Poole Lab) and Table 4-7 (Vincents media [175], 21% or 1% oxygen with 10mM glucose or 20mM succinate, [238]) are listed in the following Appendix 1 tables: Table A10: all plants,

Table A11: pea only, Table A12: soybean only, Table A13: barley only, Table A14: pea and soybean, Table A15: pea and barley, Table A16: barley and soybean.

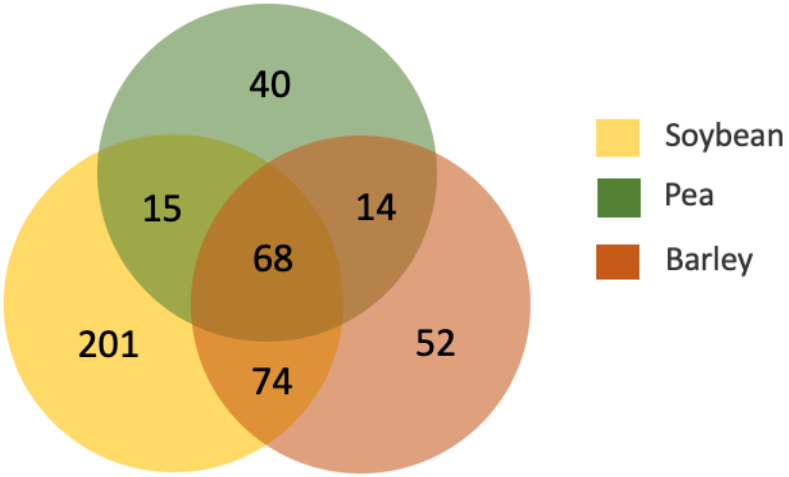


Figure 5-2. Genes required (ES/DE) for primary attachment to roots of pea, soybean and barley at pH 7.0. Total genes = 464. Color indicates plant; green = pea, yellow = soybean, brown = barley.

The distribution of genes required for attachment to pea, soybean and barley roots (Figure 5-2) is interesting because it indicates that, whilst there are a large number of genes (68) required for attachment to all plants, there are also a large number of genes required exclusively for attachment to one or two of the three plants. This demonstrates both the conservation of some primary attachment mechanisms for Rlv3841 across plant species, but also some mechanistic specificity in attachment, which is dependent on plant host. Functional classifications were examined for genes required for attachment to all plants (68 genes) or to pea (40 genes), soybean (201 genes) or barley (52 genes) exclusively using Riley codes [138, 272] as described in 4.2.6 (Figure 5-3).

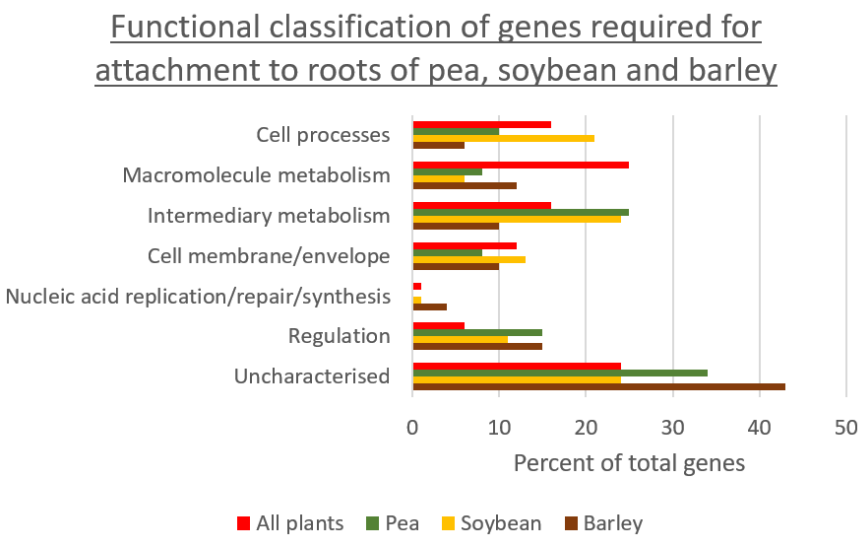


Figure 5-3. Functional classification (using Riley codes) of the 68 genes required (ES/DE) for attachment to pea, soybean and barley roots at pH 7.0 compared with the 40, 201 and 52 genes required for attachment only to pea, soybean or barley roots respectively at pH 7.0. Data is presented as percentage of total genes in each category.

Functional classifications (Figure 5-3) reveal that the most abundant ES/DE gene classes for attachment to all plants are uncharacterised genes and macromolecule metabolism. However, differences in the functional classifications of genes required for attachment to individual plants are visible. Like the finding from pea (Figure 4-6), it seems that uncharacterised genes play an important role in determining successful primary root attachment in all cases, although this is especially the case for barley.

5.2.7 Genomic localization of genes required for primary root attachment to pea, soybean and barley

Further insight can be gained by comparing the loci of genes required (ES/DE) for attachment to the roots of these three plants (Figure 5-4). As for the functional classifications above, comparisons were made for the genes required for attachment to all plants (68 genes) with the genes required for attachment only to one of the three plants (40 genes for pea, 201 genes for soybean and 52 genes for barley, Figure 5-2).

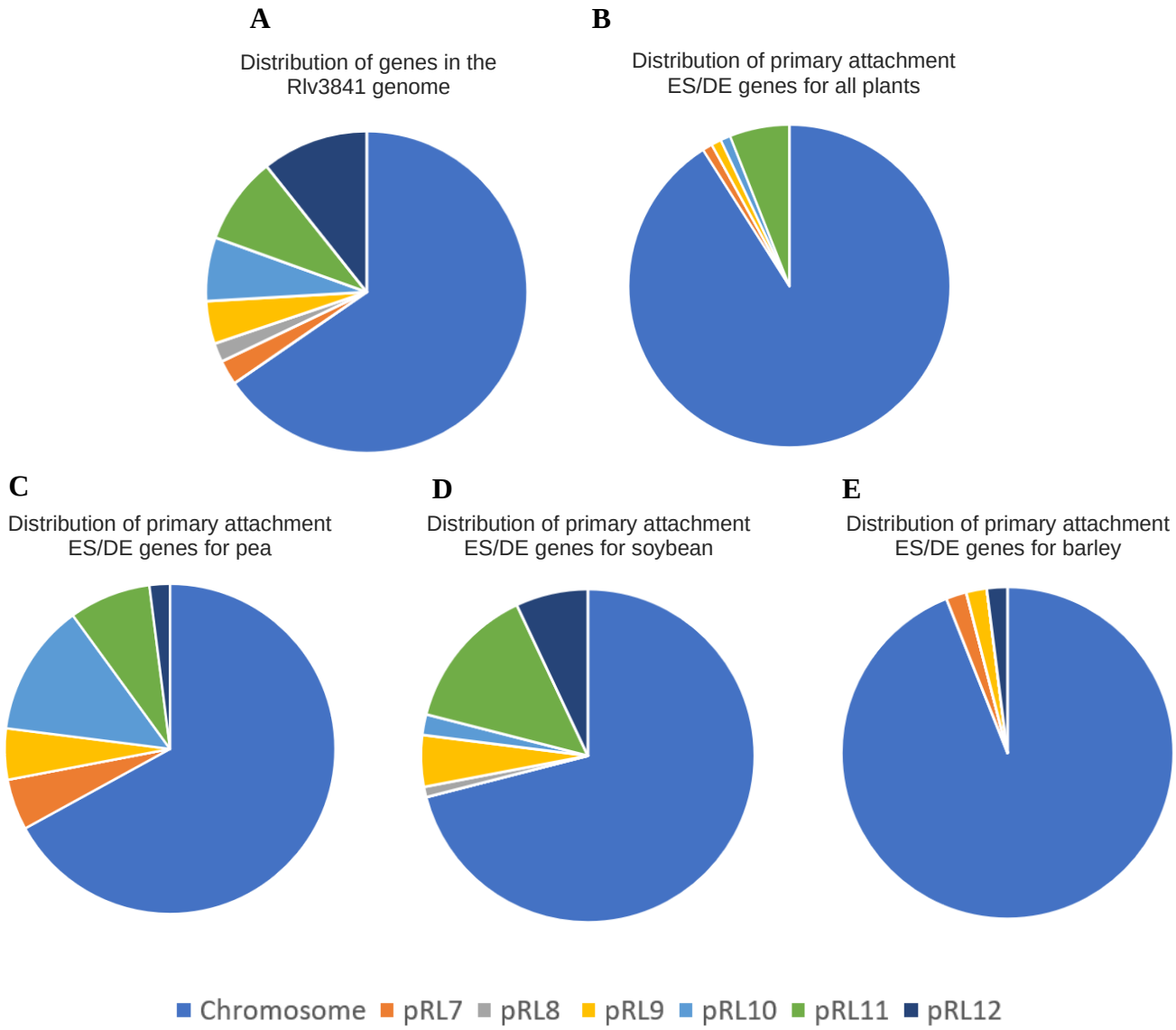


Figure 5-4. Genomic localization of Rlv3841 genes (as a proportion of the total). A – genome of Rlv3841 (7,340 genes). Genes required (ES/DE) for attachment to roots at pH 7.0 to B – all plants (68 genes), C – pea (40 genes), D – soybean (201 genes), E – barley (52 genes).

As discussed previously, 65.4% of the annotated encoded genes in the Rlv3841 genome are located on the chromosome, with the remainder divided between six plasmids (Figure 5-4, A). The distribution of primary attachment genes required for attachment to all plants tested (pea, soybean and barley, Figure 5-4, B) indicates that, compared to the genome the location of these genes is skewed toward the chromosome (91%) and away from the plasmids, with no representation for pRL12 or 8. This most closely mirrors the distribution seen for barley (Figure 5-4, E), where the chromosomally encoded factor requirement is 94%. Figure 5-4 C demonstrates that, for pea, chromosomally encoded factors form the bulk of primary attachment determinants (as they do in all cases), but plasmid representation is also the highest of all plant groups. Plasmid pRL10 is the most well represented; this is of interest as pRL10 is referred to as the symbiosis plasmid, encoding symbiotic functions in the *nod*, *nif* and *fix* genes [381].

Gene localization of soybean only attachment determinants is particularly interesting, as it demonstrates a large overrepresentation of pRL11 genes (28 genes in total) and is the only attachment condition featured in this work to require any genes encoded on pRL8 (three genes) (Table 5-3, Figure 5-4, D). These three genes are presented in Table 5-3, below, along with six genes from pRL11 which are sequentially encoded and may play a complementary role in promoting attachment based on annotation functionally linking them to the pRL8 genes.

The three pRL8 genes required for attachment to soybean roots encode conjugal transfer or pilus stabilising/secretion machinery and the six pRL11 genes cluster functionally around Flp pilus function (Table 5-3). The Flp pilus is part of the Flp/Tad system, which is known to be important for initial adhesion and biofilm formation in many Gram positive and negative bacteria [382]. *S. meliloti* mutants in the Flp pili cluster are impaired in nodulation with *M. sativa*, which is likely to be due to attachment and colonization defects [383]. *A. tumefaciens* with mutations in the *ctp* locus (which encodes the Flp/Tad system in this bacterium) shows reduced adherence and biofilm formation on abiotic surfaces [44]. Deletions in the Flp/Tad genes in many other bacteria, including *Pectobacterium* (a potato pathogen), *Vibrio vulnificus*, *Aggregatibacter actinomycetemcomitans* and *Caulobacter crescentus*, have also been shown to reduce surface adhesion [44, 384, 385]. Studies such as these have led to the general conclusion that Flp/Tad (here the pRL11 genes in Table 5-3) and Flp/Tad related genes (here putatively the pRL8 genes in Table 5-3, which may function in pilus stabilisation and/or secretion) are involved in attachment and biofilm formation [87]. It therefore seems that Rlv3841 requires pilus formation and type IV secretion system components for primary attachment to soybean roots at pH 7.0. Although this has not been documented in the Rlv3841 interaction with pea, *Agrobacterium* and *Pseudomonas* are known to use pili in primary attachment to plant roots [34].

Table 5-3. Genes from pRL8 and pRL11 required (ES/DE) for Rlv3841 attachment to soybean roots at pH 7.0 which may play a complementary role in promoting attachment.

Gene	Description
trbLp8 (*pRL80133)*	P-type conjugative transfer protein TrbL. Found in the *trb* locus of *Agrobacterium* Ti plasmids, involved in the type IV secretion system for conjugative transfer. A homologue of VirB9, a type VI secretion system protein evolved from the conjugative apparatus which exports virulence proteins in pathogenic bacteria. Essential for virulence in *Agrobacterium*, also stabilises the pilin conjugative structure [386–389].
trbHp8 (*pRL80134)*	Putative conjugal transfer protein TrbH. Contains a putative membrane lipoprotein lipid attachment site.
trbGp8 (*pRL80135*)	Putative conjugal transfer protein TrbG. A homologue of VirB6, an essential component of type IV secretion machinery for pilus formation in *Agrobacterium* [388].
pRL110569	Putative transmembrane protein. 95% identity to a pilus assembly protein (gene EHH54_1685) from *R. leguminosarum*.
pRL110570	Hypothetical exported protein.
pRL110571	Uncharacterised protein. 96% identity to Flp pilus assembly protein CpaB from *R. leguminosarum* biovar *viciae*. RcpC family protein expressed from the tight-adherence Tad locus in gram negative bacteria.; likely has a role in modifying and assembling the Flp pilin [390].
pRL110572	Putative Flp pilus assembly protein. T2SP-E domain, containing components of the type II secretion system, type 4 pili and type IV secretion system.
pRL110573	Putative transmembrane pilus component protein. T2SSF domain, found in the type II secretion system and type 4 pili.
pRL110574	Putative transmembrane pilus component protein. 100% identity to Flp pilus assembly protein TadB from *R. leguminosarum* biovar *viciae* WSM1455, predicted to form part of a large molecular structure for Flp pili assembly and secretion [391].

5.2.8 Increasing specificity of primary root attachment factor identification from INSeq – pleiotropy filtering

As discussed in 4.2.10, factors identified as primary attachment determinants which also show pleiotropic effects under different media or metabolic test conditions may be less likely to represent specific primary attachment factors. To increase the specificity of primary attachment factor identification and provide a starting point for in-depth analysis, results from this INSeq experiment were cross-referenced against the INSeq datasets listed in Table 4-7 (Vincents media [175], 21% or 1% oxygen with 10mM glucose or 20mM succinate, [238]). Results were used to assign a 'pleiotropy filter' (see 4.2.10) where genes showing a non-NE classification in any of the experiments listed in Table 4-7 were considered as non-specific primary attachment determinants. Use of this filter can be seen in Appendix 1, Tables A10-A16. Filtering resulted in 312 genes which are required specifically for root attachment. Their distribution over the three plants is shown in Figure 5-5.

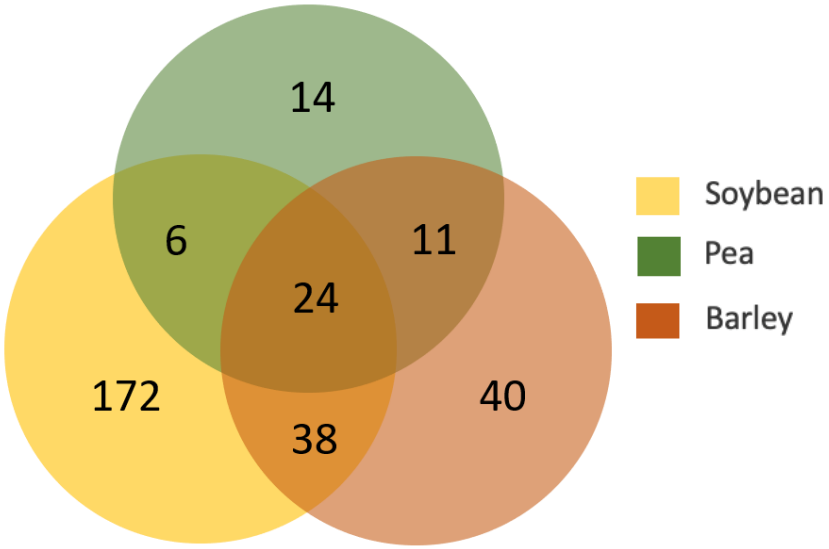


Figure 5-5. Genes specific for primary attachment to roots of pea, soybean and barley. These 312 genes have been classified as NE in the respective input library, required (ES/DE) in one or more of the root attached libraries (pea, soybean or barley, all at pH 7.0) and have been classified as NE in the following INSeq datasets: VMM [175], 21% oxygen 10mM glucose, 1% oxygen 10mM glucose, 21% oxygen 20mM succinate and 1% oxygen 20mM succinate, [238] (see Table 4-7). Color indicates plant; green = pea, yellow = soybean, brown = barley.

5.2.9 Specificity in Rlv3841 primary attachment factor requirements is highly plant-dependent

While there are clear differences in the genes specific for primary attachment to different plants, there is also substantial overlap between the plant species (Figure 5-5).

In order to visualize the specificity of the interaction for each plant species, the number of genes involved only in attachment to one plant's roots can be expressed as a percentage of the total number of genes involved in attachment to those roots. For example, after filtering a total of 55 genes are required (ES/DE) for attachment to pea roots (14 + 6 + 11 + 24, see Figure 5-5). Of these, 14 are specific for attachment to pea roots alone (31%). 69% are therefore also involved in attachment to barley and/or soybean and can be considered non-specific for pea. This calculation illustrating the specificity for attachment to pea, soybean and barley is summarized in Figure 5-6.

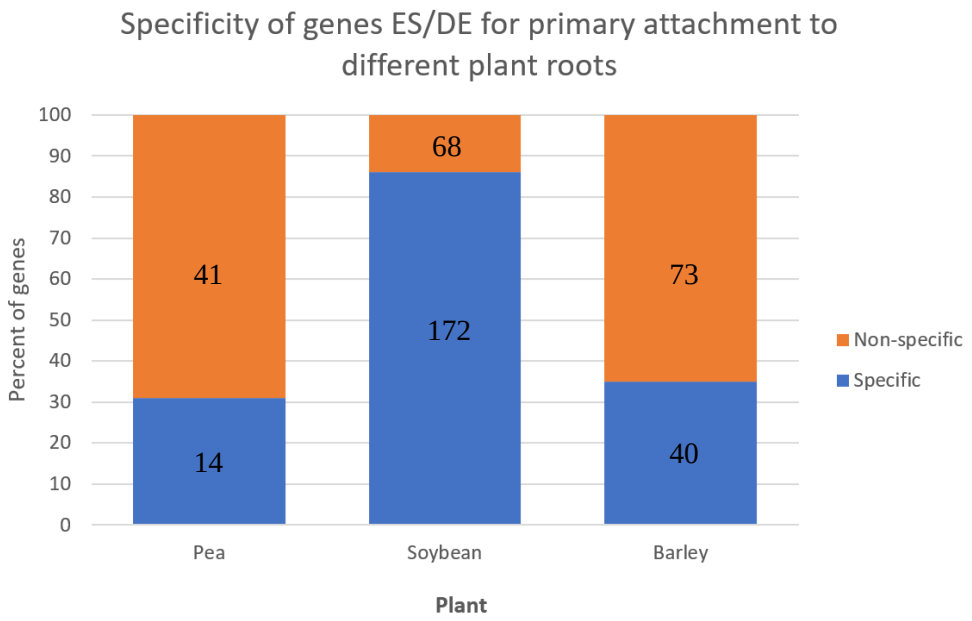


Figure 5-6. Specificity of primary attachment factors for Rlv3841 attaching to pea, soybean and barley roots. Data are expressed as percentage of genes required (ES/DE) for primary attachment to plant roots post-filtering (5.2.8, Figure 5-5) which are specific (i.e. required for attachment to one plant type only, blue) or non-specific (i.e. required for attachment to more than one plant type, orange). Gene numbers are given within the bars for each plant type. Total n = 55 genes for pea, 240 genes for soybean and 113 genes for barley.

Dividing the data in this way (Figure 5-6) demonstrates that Rlv3841 displays the highest unique primary attachment factor requirements (~86%) in interaction with the roots of soybean (non-host legume). The roots of pea plants (host legume) and barley (non-legume) show a similar proportion (~35% in each case), of specific primary attachment factors. This is surprising, given that primary attachment factor requirements for legume roots might be expected to be more similar than for non-legume roots. However, given the far higher unique primary attachment factor requirements, and the demonstrated

requirement for pilus formation and type IV secretion system components for primary attachment to soybean roots at pH 7.0 (see 5.2.7), it seems that this is not necessarily the case. Primary attachment factor requirement may therefore be governed by features of the root surface which are not conserved along legume/non-legume lines. This could also be influenced by differential gene induction in response to different plant exudation profiles, although to what extent this is occurring remains unclear.

5.2.10 Regulatory requirements for Rlv3841 primary attachment to pea, soybean and barley roots

Given the large differences in gene specificity requirements for primary attachment (described in 5.2.9 and Figure 5-6), it seems logical that genes with differing regulatory functions play a role in attachment to different plant roots. Given that soybean shows the highest requirement for specific genes (~86%, Figure 5-6), it may be the case that soybean has the highest number of specific regulators involved. Using Riley codes [138, 272] to identify regulators showed that, post-filtering (5.2.8), there is one regulator required for primary attachment to all plants (*RL4145*, Table 5-4), three required only for pea (*pRL120518*, *RL1105* and *pspA* – *RL1106*, Table 5-4), 22 required only for soybean (Table 5-4), and six required only for barley (Table 5-4). These figures may support the idea that higher specificity in primary attachment factor requirements is coupled to increased regulatory requirements as the numbers for soybean are higher than those of other plants. Only six have functional annotation. These are *pspA* (*RL1106*), putatively required for attachment to pea, *ecfM* (*pRL100385*), *mur* (*RL0397*), *cheY1* (*RL0687*) and *aldR* (*RL1965*), required for attachment to soybean and *msiR* (*RL2857*), required for attachment to barley.

Table 5-4. Genes with regulatory function (based on Riley codes) that are required (ES/DE) in primary attachment to all plants, or to pea, soybean or barley roots exclusively, at pH 7.0.

Gene	Plant	Description
RL4145	All	putative LacI family transcriptional regulator (repressor) (see also Figure 4-15)
pRL120518	Pea	putative TetR family transcriptional regulator
RL1105	Pea	putative TetR family transcriptional regulator
pspA (*RL1106*)	Pea	putative (phage shock protein A) PspA family regulator by protein-protein interactions (see Table 4-11)
ecfM (*pRL100385*)	Soybean	putative RNA polymerase ECF sigma factor, family ECF20/ECF01. As a member of the clade VI sigma factors, EcfM may play a role in stress response or host interactions [392]
pRL110046	Soybean	putative FNR/CRP family transcriptional regulator
pRL110283	Soybean	putative ArsR family transcriptional regulator
RL0229	Soybean	putative LysR family transcriptional regulator
mur (*RL0397*)	Soybean	putative FUR-like transcriptional regulator, manganese uptake regulator [393]
cheY1 (*RL0687*)	Soybean	putative two-component sensor/regulator; chemotaxis transcriptional regulator CheY
RL0762	Soybean	putative XRE family (HipB) transcriptional regulator
RL1040	Soybean	putative LysR family transcriptional regulator

RL1162	Soybean	putative two-component sensor/regulator; transcriptional regulator
RL1163	Soybean	putative two-component sensor/regulator; histidine kinase
aldR (*RL1965*)	Soybean	putative AsnC family positive transcriptional regulator of alanine dehydrogenase [394]
RL2486	Soybean	putative LysR family transcriptional regulator
RL2766	Soybean	putative ArsR family transcriptional regulator
RL2779	Soybean	putative two-component sensor/regulator; histidine kinase
RL3146	Soybean	putative LysR family transcriptional regulator
RL3148	Soybean	putative ArsR family transcriptional regulator
RL3196	Soybean	putative MarR family transcriptional regulator
RL3263	Soybean	putative LysR family transcriptional regulator
RL3265	Soybean	putative AraC family transcriptional regulator (activator)
RL4189	Soybean	putative LysR family transcriptional regulator
RL4219	Soybean	putative DeoR family transcriptional regulator (repressor) of sorbitol/mannitol operon
RL4604	Soybean	putative GntR family transcriptional regulator
RL0159	Barley	putative MarR family transcriptional regulator
RL0561	Barley	putative AraC family transcriptional regulator (activator)
RL2657	Barley	putative GGDEF/GAF sensory box protein

msiR (*RL2857*)	Barley	putative ArsR family transcriptional regulator MsiR, regulates MsiA canavanine (found in seed exudate) exporter
RL3455	Barley	putative MarR family transcriptional regulator
RL3595	Barley	putative LacI family transcriptional regulator (repressor)

Where available, gene annotation is given in brackets next to gene number.

RL4145 (the regulator important for Rlv3841 primary root attachment to all plants, Table 5-4), has been investigated using a Lux attachment assay (Figure 4-15 and Table 4-9) and was required for attachment to pea roots under all pH conditions. Although the regulatory targets of *RL4145* remain unknown (the only predicted functional interaction partner from the STRING database with a confidence score >0.6 is *RL4144*, a putative oxidoreductase/dehydrogenase) this gene seems to be a very important regulator of attachment to pea, soybean and barley roots, including at different pHs.

One regulator involved in attachment only to pea is annotated: *pspA (RL1106*). This phage shock protein A gene functions to stabilise the cytoplasmic membrane in *E. coli* and promote biofilm formation [329]. It is also upregulated under membrane stress conditions in *R. leguminosarum* biovar *trifolii*, where it may improve desiccation tolerance [395]. Although the link with root attachment in this case remains unclear, desiccation tolerance has previously been linked in Rlv3841 with higher EPS accumulation [396]. Given the importance of EPS in root attachment [34, 71, 77], if loss of *pspA* affects EPS production, this could corroborate the importance of this factor for primary attachment to pea roots.

Of the 21 regulators required for primary attachment to soybean only four are annotated: *ecfM* (*pRL100385*), *mur* (*RL0397*), *cheY1* (*RL0687*) and *aldR* (*RL1965*) (Table 5-4). Of interest is the requirement for *cheY1* (*RL0687*), a response regulator controlling flagellar rotation in chemotaxis. In the previous chapter, motility and chemotaxis components were not required for Rlv3841 attachment to pea roots in INSeq, even though they were required in the Lux root attachment assay (see Figure 3-5). Possible reasons for this, mainly concerning the mixed inoculum setup of INSeq compared with the homogenous single inoculum for Lux assays, were discussed in 4.2.9. CheY1 is a minor chemotaxis regulator, whilst CheY2 is of greater importance; deletion of *cheY1* reduces swimming

ability by 30%, whereas this figure is 70% in the case of *cheY2* deletion [397]. If chemotaxis is necessary for primary attachment to soybean roots, *cheY2* would also be expected to be required. However, *cheY2* has only a single TA site (*cheY1* has three). As discussed in 4.2.9, low TA site number (< six) can confound HMM classification.

To investigate this further, several other chemotaxis genes shown to be required by this INSeq experiment in the soybean library and with more TA sites were compared to the same genes in the input, pea and barley libraries using fitness values (FV). FVs allow numerical comparison of the fitness of bacteria carrying mutations in particular genes under different test conditions, providing a simplified visualization of the data underlying HMM classifications [143] (see 2.7.5). Comparisons of gene TA site number and HMM classifications are shown in Table 5-5 and comparison of FVs for the same genes are shown in Figure 5-7.

Table 5-5. Comparison of TA site number and HMM classifications for five chemotaxis-related genes under different INSeq conditions

		FV (HMM classification)				
Gene	**#TA**	**I 1**	**Pea**	**I 2**	**Bar**	**Soy**
mcpR (*pRL120056*) – putative methylation accepting chemotaxis protein	29	NE	NE	NE	NE	DE
icpA (*RL0685*) – putative chemoreceptor protein	15	NE	NE	NE	NE	DE
cheX1 (*RL0686*) – putative chemotaxis related CheX protein	3	NE	NE	NE	NE	DE
cheY1 (*RL0687*) – putative two component sensor/regulator; chemotaxis regulator CheY	3	NE	NE	NE	NE	DE
cheW3 (*RL4030*) – putative chemotaxis protein	9	NE	NE	NE	NE	DE

HMM state call is given in brackets next to each FV. #TA = number of gene TA sites. I 1 = input library 1 (for pea). I 2 = input library 2 (for barley and soybean). Bar = barley. Soy = soybean.

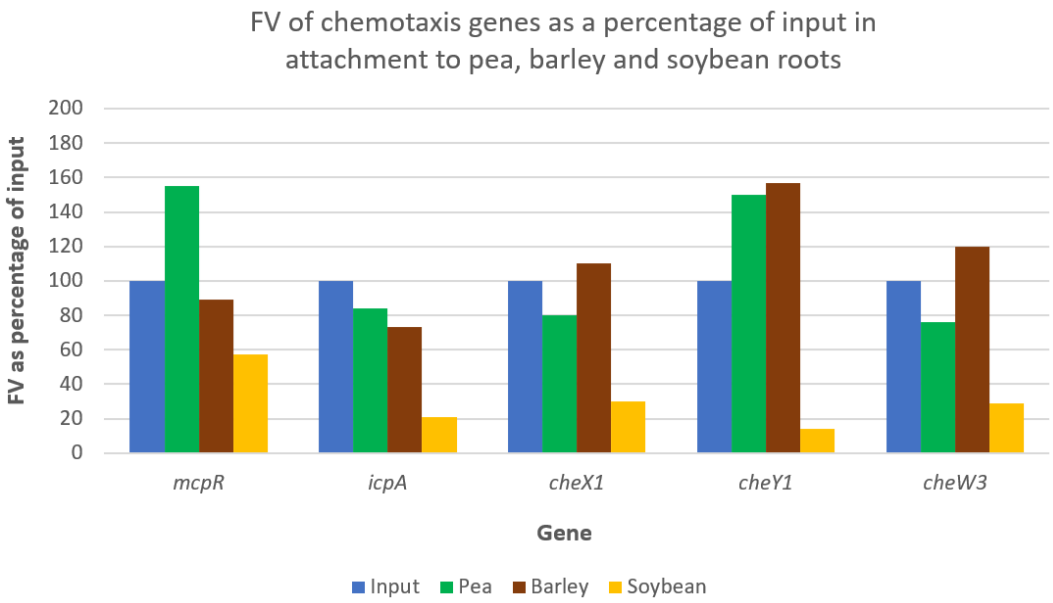


Figure 5-7. Fitness value (FV) of five Rlv3841 chemotaxis-related genes (reflecting how mutation affects organism fitness under test conditions) from INSeq attachment to pea, barley and soybean roots. FVs are expressed as a percentage of input, where input is indexed at 100%. For pea, indexing is to input library 1 and for barley and soybean indexing is to input library 2. HMM classifications can be found in Table 5-5.

Using fitness values to analyze INSeq results (Figure 5-7) demonstrates that the requirement for chemotaxis in Rlv3841 attachment to soybean does not seem to be an artefact of low TA site number in genes. However, genes such as *RL0695 – RL0725* (encoding flagellar and flagellar motor structural proteins) have a NE classification for attachment to soybean, as well as pea and barley (Appendix 2 Table 2). This indicates that specific aspects of chemotaxis are needed for attachment to soybean roots, independent of

non-chemotactic motility. For motile bacteria with mutations in chemotaxis related genes, a negative attachment phenotype may arise due to aberrant swimming, whereby bacteria swim randomly and are therefore not correctly positioned in proximity to the root surface for attachment. Non-motile bacteria may still be able to reach the root via the 'aided motility' mechanism (biased Brownian motion and overcoming of repulsive root boundary forces) hypothesized in 4.2.9, accounting for the NE classification of flagellar genes such as *RL0695 – RL0725*. However, the question of why chemotaxis requirements should be seen only for attachment to soybean, and not pea or barley roots, remains. This could be explained by differential exudation between plants. Plants including pea, soybean and barley exude carbon-rich molecules from their roots which act as chemoattractants, chemorepellents and carbon sources for soil bacteria, and likely play a role in preferential recruitment of certain bacteria to form the root microbiome [34, 212, 213, 398]. In this study, soybean roots are by far the largest of the three root types and may therefore exude higher levels of all compounds, including any chemoattractants. This in turn could provide a stronger advantage to those bacteria able to sense these chemoattractants in this INSeq experiment, making them more likely to reach roots than for plants with lower levels of root exudation. A stronger selection for root attachment on chemo-sensing bacteria combined with the aberrant swimming of motile but non-chemotactic bacteria could potentially result in under-representation of chemotaxis mutants in the soybean root-attached library.

Of the six regulators specific for barley primary root attachment only one is annotated: *msiR* (*RL2857*, Table 5-4). MsiR regulates the MsiA-mediated export of canavanine, a toxin found in seeds and exudates of some plants, including many legumes [141]. MsiA (*RL2856*) is also required (DE) specifically for primary attachment to barley roots.

Canavanine export is important for root attachment and rhizosphere persistence for *Rhizobium* interacting with canavanine-producing legumes [141, 399]. Previous work has shown that soil extracts from barley-cropped areas contain unidentified phytotoxins which can inhibit plant root growth. It may be that barley roots produce a phytotoxic canavanine analogue, accounting for the requirement for the canavanine export system in Rlv3841.

5.2.11 Primary attachment determinants required for interaction with different plants

Filtering increased the specificity of primary attachment determinant identification (5.2.8). There remains a distribution of gene requirements for attachment to different plant roots. These were investigated in more detail to determine what roles they may play and how different host roots influence Rlv3841's mechanisms of primary attachment. For brevity, only factors with roles discernible from their annotation are discussed below. The full lists of factors (including uncharacterised and hypothetical proteins) can be found in Appendix 1, Tables A10-A16.

Table 5-6 provides a summary of the tables in which the results are stored.

Table 5-6. Summary of tables of genes required (ES/DE) in primary attachment to pea, soybean and barley roots (Tables 5-7 to 5-13) and Appendix 1, Tables (A10 to A16)

Plant(s)	Number of genes required	Reference tables
Pea, soybean, barley	24	Table 5-7, Figure 5-8, Appendix Table A10
Pea only	19	Table 5-8, Appendix Table A11
Soybean only	172	Table 5-9, Appendix Table A12
Barley only	40	Table 5-10, Appendix Table A13
Pea and soybean	6	Table 5-11, Appendix Table A14
Pea and barley	11	Table 5-12, Appendix Table A15
Barley and soybean	38	Table 5-13, Appendix Table A16

Number of genes are from the Venn diagram shown in Figure 5-5.

Table 5-7. Twenty-four genes[a] identified as required (ES/DE) for primary attachment to pea, soybean and barley roots.

Gene(s)	Description
pRL110043	Putative transmembrane transporter protein. 99% similarity to arabinose efflux permease family protein in *R. leguminosarum* biovar *trifolii* WSM597. EPS arabinose content regulates cell aggregation in *Azospirillum* [358]. See also Table 4-13 and Figures 4-12 and 4-13
pRL100112	Putative dehalogenase-hydrolase (HAD), member of a large superfamily of phosphohydrolases. HADs display activity against various phosphorylated metabolites [326]. A mutant in a *Xanthomonas axonopodis* pv. *citri* HAD is defective in biofilm formation [327]. See also Table 4-11 and Figure 4-13
hslO (*RL0551*)	Putative Hsp33-like chaperonin. Redox regulated molecular chaperone protecting thermally unfolding and oxidised proteins from aggregation – defence against oxidative stress [277, 278]. May assist in coping with a diverse rhizosphere environment (see 4.1). See also Table 4-9 and Figure 4-11
RL0617	Putative dTTP/UTP pyrophosphatase. See also Table 4-14 and Figures 4-13 and 4-14
RL1013	Uncharacterised protein, 17kDa Anti 2 motif. These motifs are common membrane proteins and include surface antigens in *Rickettsia* [328]. Unknown protein localization. See also Table 4-11 and Figure 4-13
amn (*RL1478*)	AMP nucleosidase; catalyses hydrolysis of AMP to form adenine and ribose 5-phosphate. Changes in AMP levels allow rapid adjustments to changing metabolic conditions [279]. See also Table 4-9 and Figure 4-11
ppx (*RL1600*)	Putative exopolyphosphatase Ppx; can function in the hydrolysis of EPS [359]. See also Table 4-13 and Figures 4-12 and 4-13
clpS (*RL2212*)	ATP dependent Clp protease adaptor protein ClpS. ClpS modifies ClpA substrate specificity [364]. A *clpS* mutant of *P. aeruginosa* showed 70% reduction in biofilm formation and significant impairment in attachment to abiotic surfaces after one hour [365]. See also Table 4-14 and Figures 4-13 and 4-14.
hfq (*RL2284*)	RNA binding protein Hfq. Global post transcriptional regulator. Loss of Hfq in *S. meliloti* delays nodulation and reduces competitiveness for attachment to alfalfa roots [360]. See also Table 4-13 and Figures 4-12 and 4-13

tpiA (*RL2513*)	Putative triosephosphate isomerase. Upregulated in *Staphylococcus aureus* biofilm, possibly due to oxygen limitation [281]. Glycolytic enzymes play additional roles when localised on the cell surface (e.g. α-enolase plasminogen binding in streptococci and GAPDH transferrin binding activity in *S. aureus*) [282, 283]. Surface localised glycolytic enzymes are multifaceted and can be involved in substrate binding. TpiA has been shown to be surface localised and have a direct role in attachment to host cells in *Mycoplasma gallisepticum* [284]. See also Table 4-9 and Figure 4-11.
recA (*RL2637*)	RecA is needed for DNA repair and the SOS response. The major activity of RecA in DNA metabolism is the promotion of DNA strand exchange [285]. Requirements for all symbiosis stages suggest that RecA may assist in coping with a diverse rhizosphere environment (see 4.1). However, *recA* disruption has been shown to reduce adherence and colonization of host cells by *Vibrio cholerae*, although the mechanism underlying this remains unknown [286]. See also Table 4-9 and Figure 4-11.
dksA (*RL2643*)	*RL2643* – putative DnaK suppressor protein DksA. Studies from various bacteria indicate that these chaperones can be secreted and bind ligands, contributing to cell adherence and biofilm formation [366–369]. Cytoplasmic protein localization. See also Table 4-14 and Figures 4-13 and 4-14
gor (*RL2694*)	*RL2694* – Glutathione reductase (Gor). See also Table 4-11 and Figure 4-13
pfp (*RL3322*)	Putative pyrophosphate-fructose 6-phosphate 1-phosphotransferase. Catalyses the first committed step in glycolysis, the phosphorylation of D-fructose-6-phosphate [287]. Like *tpiA* (*RL2513*) (see Table 4-9), could have multifaceted role. See also Table 4-9 and Figure 4-11
RL3987-90	*RL3987* – uncharacterised, SpoVT-AbrB domain. *RL3988* – uncharacterised, PINc domain. PIN domains function as single stranded RNA nucleases [291]. In prokaryotes they are usually the toxin of toxin-antitoxin operons, helping free-living prokaryotes cope with nutritional stress [292]. *RL3989, RL3990* – Holliday junction ATP dependent DNA helicases RuvA and RuvB; DNA damage repair mechanism. May be required for osmotic shock responses [293, 294]. See also Table 4-9 and Figure 4-11

RL4065	Conserved hypothetical protein, no known conserved domains. Cytoplasmic protein localization. See also Table 4-9 and Figure 4-11
RL4145	Putative conserved LacI type transcriptional regulator (repressor). Regulatory targets unknown. Cytoplasmic protein localization. See also Table 4-9 and Figure 4-11
RL4362/ dacC (*RL4363*)	*RL4362* – putative cobalamin (vitamin B12) synthesis protein, CobW domain [295]. Required by *S. meliloti* for symbiosis with *M. sativa* [296]. Only one cobalamin dependent enzyme (*nrdJ*, *S.meliloti* cobalamin dependent ribonucleotide reductase, RNR) affects symbiosis. Removal of *ndrJ* impairs symbiosis; rhizobia are lysed in the plant cytoplasm [297]. Loss of *RL4362* may reduce fitness for competitive primary attachment. Note that cobalamin synthesis genes may be misclassified due to high homology with glutamine amidotransferases, which are involved in peptidoglycan amidation [298]. *RL4363* – *dacC*, putative penicillin binding protein, peptidase S11 domain. In *E. coli dacC* processes sugar-peptide cell wall precursors; involved in peptidoglycan biosynthesis [299]. See also Table 4-9 and Figure 4-11.
RL4381	Putative POTRA domain transporter. See also Table 4-9, Figures 3-11 and 4-11 and section 3.2.9

[a]Number of genes are from the Venn diagram shown in Figure 5-5. Red = Genes also required (ES/DE) in attachment to pea roots at all pHs tested in Chapter 4.

These genes, being ES/DE in attachment to pea, are discussed in the relevant tables and figures in Chapter 4, which are referenced in Table 5-7. The genes listed in Table 5-7 are 'core' determinants of primary attachment in Rlv3841, as they are required not only for attachment to pea, but also to soybean and barley roots. Of these 25 genes, 14 (highlighted in red) are required for attachment to pea roots under all pH conditions and are therefore some of the most important primary attachment determinants in the Rlv3841 genome. These are shown, along with brief functional descriptions, in Figure 5-8, below.

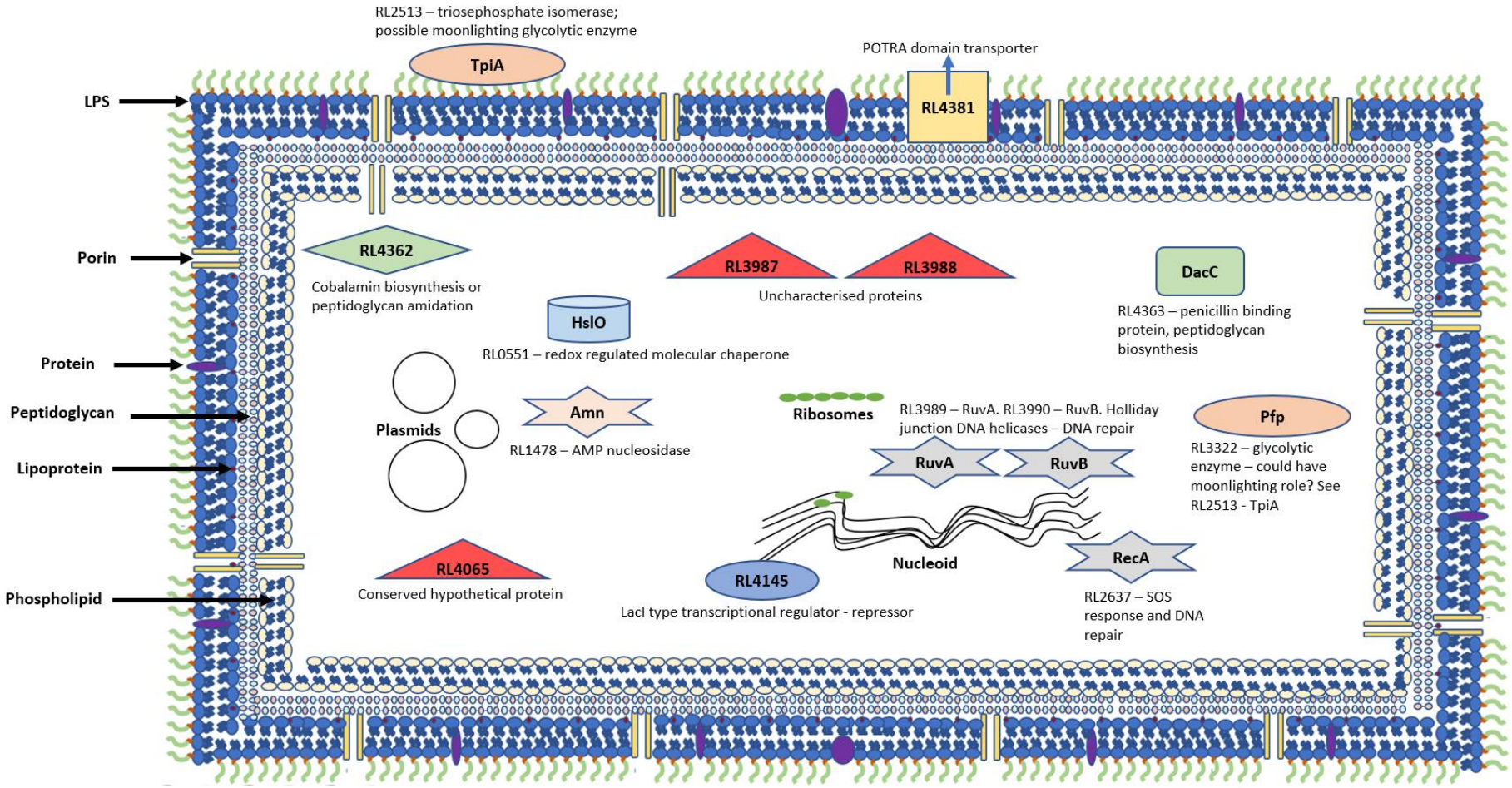


Figure 5-8. Diagram of a stylized gram-negative Rlv3841cell showing the primary attachment determinants whose genes were identified as required (ES/DE) in attachment to pea (pH 6.5, 7.0 and 7.5), soybean and barley (see Table 4-9). Cell outer membrane, peptidoglycan layer, inner membrane, plasmids, nucleoid and other cellular factors are shown, not to scale.

Table 5-8. Fourteen genes[a] identified as required (ES/DE) for primary attachment to pea roots only (pH 7.0).

Gene	Description
pRL100053	Putative transmembrane domain containing protein, helix-turn-helix 37 domain. Unknown protein localization. See Table 4-9 and Figures 3-11 and 4-11 and section 3.2.9.
pRL100174	Hypothetical protein, no known conserved domains. Unknown protein localization. See also Table 4-9 and Figure 4-11
pRL100242	Uncharacterised protein. Cytoplasmic protein localization. See also Table 4-13 and Figures 4-12 and 4-13
pRL120518	Putative TetR family transcriptional regulator. See also Table 4-11 and Figure 4-13.
cycM (*RL0141*)	Membrane-bound cytochrome c CycM. See also Table 4-14 and Figures 4-13 and 4-14
RL1105	putative TetR family transcriptional regulator. See also Table 4-11 and Figure 4-13
pspA (*RL1106*)	Putative PspA family regulator, phage shock protein A. Involved in antibiotic resistance and biofilm formation in *E.coli* and pathogenesis in *S. typhimurium* [329, 330]. See also Table 4-11 and Figure 4-13
RL1504	Uncharacterised protein, NYN domain. Possibly novel RNAse with regulatory role [363]. Cytoplasmic protein localization. See also Table 4-14 and Figures 4-13 and 4-14
gmsA (*RL1661*)	Glucomannan biosynthesis protein GmsA. Characterised factor important for polar primary root attachment at acidic but not alkaline pH [56]. See also Tables 4-13 and Figures 3-5, 4-12 and 4-13
scpA (*RL2044*)	Segregation and condensation protein A, participates in chromosomal division during cell partition. See also Table 4-11 and Figure 4-13
anmK (*RL2587*)	Anhydro-N-acetylmuramic acid kinase AnmK. Catalyses the phosphorylation of 1,6-anhydro-N-acetylmuramic acid (anhMurNAc), cleaving the 1,6 anhydro ring and generating MurNAc-6-P. Required for cell wall recycling [331, 332]. See also Table 4-11 and Figure 4-13
ahpD (*RL3226*)	Alkyl hydroperoxide reductase AhpD. See also Table 4-11 and Figure 4-13

pssA (*RL3752*)	Glycosyl transferase involved in EPS biosynthesis [288, 289]. Mutants are deficient in EPS production and form biofilms slowly compared to Rlv3841. Does not attach to root hairs [56]. Biofilms are flat and unstructured [58]. See also Table 4-9 and Figures 3-5 and 4-11 and section 3.2.5
RL4704	Putative glyoxylase family protein, member of the VOC superfamily. Members of this family are known to detoxify methylglyoxal, formed as a by-product of lipid metabolism [333]. See also Table 4-11 and Figure 4-13

[a]Number of genes are from the Venn diagram shown in Figure 5-5.

These genes, being defective in attachment to pea when mutated, are discussed in the relevant reference tables and figures in Chapter 4, which are referenced in Table 5-8.

Table 5-9. Twenty-seven genes (out of 172[a]) with discernible roles identified as required (ES/DE) for primary attachment to soybean roots only (pH 7.0).

Gene	Description
redAh (*pRL110048*)	*R. etli* has two *redAh* genes [400] which encode proteins similar to MurG, a transferase involved in the membrane steps of peptidoglycan biosynthesis [401]
pRL110395	Putative succinoglycan biosynthesis protein. Glycosyltransferase family 2 domain, can transfer sugars to diverse substrate including cellulose and teichoic acids
pRL110439	Putative glycosyltransferase. 92% identity to beta-N-acetylhexosaminidase from Rlv *trifolii* WSM2304. Role in cell wall recycling; hydrolyses β-1,4 glycosidic bonds in peptidoglycan [402]
pRL110453	Uncharacterised protein, 91% identity to regulator of enolase from Rlv *trifolii* WSM2304. Enolase is involved in glycolysis and gluconeogenesis but also has a moonlighting role. In surface localised form it mediates surface attachment and biofilm formation in *S. aureus*, *Actinomyces naeslundii* and even the fungi *Candida albicans*, though the mechanism remains unclear [282, 283, 403–405]. Enolase (*rl2239*) is ES/DE in input and all plant libraries (it is a central metabolic enzyme) but, as its regulator is only required in attachment to soybean roots, this may suggest that enolase plays a role.
mcpR (*pRL120056*)	Methyl accepting chemotaxis protein, can mediate chemotaxis to malate and fumarate [406], both common root exudates [407]. Malate release from Arabidopsis and tomato roots induces root colonization by soil microbes [408, 409].
lnt (*RL0391*)	Apolipoprotein N-acyltransferase. Catalyses N-acylation of apolipoprotein, the last step in lipoprotein maturation. Most lipoproteins are anchored in the outer membrane [410]
RL0664	Putative transmembrane acetyltransferase, shows homology to EPS modifying gene *pssR* and may acetylate EPS, CPS, N-acetylglucosamine or other targets [411]
cheX1 (*RL0686*)	Chemotaxis related CheX protein. CheY-P phosphatase, role in flagellar directional change [412]

cheY1 (*RL0687*)	Chemotaxis response regulator, minor role in flagellar reorientation [397]
icpA (*RL0865*)	Chemotaxis protein with a broad attractant spectrum [413, 414]
RL0963	Putative transmembrane/surface protein – BA14K family protein, likely with a role in LPS biosynthesis [415]
RL1015	Putative polysaccharide deacetylase protein. 95% identity to xylanase/chitin deacetylase from Rlv WSM1455.
RL1016	Putative D-alanyl-D-alanine carboxypeptidase. Carboxypeptidases are extensively involved in peptidoglycan biosynthesis and modification [416]
RL1090	Putative transmembrane protein, DUF1499 family. Domain of unknown function (DUF) family proteins are often conserved (indicating an important role) and function often only becomes apparent under particular conditions [417]
RL1108	Putative transmembrane AMP-binding acyltransferase, PlsC (phosphate acetyltransferase) domain. Functions in membrane biogenesis and modification, using fatty acid chains to form membrane phospholipids [418]
RL1155	Putative glycosyltransferase. 96% identity to glycosyltransferase 25 (LPS biosynthesis protein) from *R. leguminosarum*. LPS strongly influences adhesive properties in this bacteria [419]
RL1549	Putative transmembrane efflux protein, 92% identity to LysE family translocator from *R. hidalgonenese*, which excretes excess l-Lysine as a result of natural flux imbalance or peptide hydrolysis [420]
dgkA (*RL2780*)	Diacylglycerol kinase DgkA. In *E.coli*, DgkA mutants are defective in biofilm formation [316], and DgkA function has been linked to phospholipid recycling and LPS modifications [324]. In *B. subtillis* it is important for lipoteichoic acid synthesis [325]
RL3149	Putative adenylate cyclase/guanylate cyclase
flaH (*RL3268*)	Flagellin, flagella subunit. Rlv3841 *flaH* mutants show shorter flagella and reduced motility [421]; loss of motility and/or flagella adhesin function may lead to ES/DE classification
RL4030 (*cheW3*)	Mcp type chemoreceptor

RL4075	Uncharacterised LysM domain containing protein. A highly conserved carbohydrate binding module, LysM domains often bind peptidoglycan in prokaryotes [422]
nodN2 (*RL4120*)	NodN2, an additional copy of NodN (*pRL100179*), which participates in Nod factor biogenesis. The role of NodN remains unclear [423]. However, core Nod biosynthesis genes (producing the chitin oligomer backbone of Nod factor) are important for biofilm formation in *S. meliloti* [424]. The protein encoded by this gene may also possess enoyl-CoA hydratase activity.
cysZ (*RL4210*)	Putative cysteine biosynthesis protein. Mutation of *cysZ* in *V. fischeri* were defective in biofilm formation which could be complemented by the *cysK* gene but not by cysteine, suggesting unknown additional roles for CysK [425]
gelA (*RL4404*)	Gel forming EPS production protein GelA. Regulated by RosR [158]. Also impaired in attachment to pea roots under pH 6.5 conditions. (Table 4-10)
exoD (*RL4420*)	Putative EPS biosynthesis protein ExoD. This protein has been shown to be involved in EPS biosynthesis, but its exact function remains unknown [426]
nagA (*RL4602*)	Putative N-acetylglucosamine-6-phosphate deacetylase NagA. Required for cell wall peptidoglycan and teichoic acid biosynthesis in *L. monocytogenes* [427].

[a]Number of genes are from the Venn diagram shown in Figure 5-5.

The genes with a role in chemotaxis/motility (*mcpR – pRL120056, cheX1 – RL0686, cheY1 – RL0687, icpA – RL0856, flaH – RL3268* and *cheW3 – RL4030*) required for attachment to soybean roots are discussed in 5.2.10.

Various genes required indicate modification of EPS for attachment to soybean roots. *pRL110395* is a putative succinoglycan biosynthesis protein, and succinoglycan production is known to affect EPS composition in *S. meliloti* [353–356]. *RL0064* has high homology to the EPS modifying *pssR* gene, loss of which reduces EPS acetylation in *R. leguminosarum* biovar *viciae* VF39 and decreases the efficiency of nitrogen-fixing symbiosis formation with alfalfa, possibly through a reduction in root attachment [411]. Conversely, *RL1015* is a polysaccharide deacetylase, indicating that, for attachment to soybean roots, a balance of EPS acetylation level may be required that is different from the EPS acetylation levels required for attachment to pea or barley roots, likely reflecting differences in root surfaces which require different or modified bacterial factors for successful attachment. *RL0064* and *RL1015* may interact to fulfil this role, although this has not been demonstrated. *exoD* (*RL4420*) is an EPS biosynthesis protein of unknown function [426]. Also required is the gel-forming EPS production protein GelA (RL4404). Previously, the role of this gene in Rlv3841 attachment to vetch root hairs was investigated. A mutant displayed normal attachment, meaning a role for this gene was not assigned, meaning its function remained unknown [158]. The requirement for GelA in Rlv3841 attachment to pea at pH 6.5 seen in Chapter 4 (Table 4-10), combined with the results of Williams *et al*. (2008) [158], suggested a role for this gene in attachment to bulk epidermal pea root surface (as opposed to root hairs) at pH 6.5. The requirement for this gene in attachment to soybean roots at pH 7.0 is interesting, although the similarities

between pea root surfaces at pH 6.5 and soybean root surfaces at pH 7.0 that would necessitate GelA for primary attachment are unclear.

Genes likely involved in peptidoglycan modification include *redAh* (*pRL110048*), a homologue of *murG* which transfers N-acetylglucosamine to a lipid intermediate in the biosynthesis of peptidoglycan [401]. *pRL110439* may play a role in cell wall recycling by hydrolysing β-1,4 glycosidic bonds in peptidoglycan [402]. *RL1016* encodes a carboxypeptidase; these proteases have been heavily implicated in peptidoglycan biosynthesis and modification in many bacterial species [416]. *nagA* (*RL4062*) has been shown to influence both peptidoglycan biosynthesis in *L. monocytogenes* [427], and may also be playing this role in Rlv3841. RL4075 likely encodes a peptidoglycan binding factor, though its exact role remains unclear [422].

A further subset of genes is involved in LPS biosynthesis and modification. *RL0963* is a BA14K family protein, named after homology to a 14 kDa protein from *Brucella abortus*. In *B. abortus*, this protein shows lectin-like carbohydrate binding activity and is essential for virulence, most likely due to its direct or indirect effects on LPS biosynthesis [415]. Whilst it is unclear what role *RL0963* plays in Rlv3841, BA14K homology is suggestive of a function in LPS production. *RL1155* encodes a glycosyltransferase likely involved in LPS biosynthesis, whilst *dgkA* (*RL2780*) has been linked to LPS modifications in *E. coli* [324], with mutants defective in biofilm formation in a 24 hr assay [316].

Two genes are likely involved in membrane function: *lnt* (*RL0391*), involved in the production of outer membrane associated lipoproteins [410] and *RL1108*, functioning in membrane phospholipid production [418].

Similar to TpiA (a glycolytic enzyme which can be cell surface localized, function in substrate binding [284] and is required for attachment to pea roots at all pHs – Table 4-9)

enolase may also be involved in a surface localised capacity in attachment to soybean roots. It has been shown to play this role in various other organisms [282, 283, 403–405] and, unless *pRL110453* (putative regulator of enolase) has other regulatory targets, it is likely to also be playing this role here.

As a possible guanylate cyclase, *RL3149* may be linked to cellular c-di-GMP levels, a well-known signalling and regulatory mechanism governing bacterial motile-sessile lifestyle changes [79].

The requirement for *nodN2* (*RL4120*) could indicate that a core Nod biosynthetic gene product is involved in attachment. Although NodN function is unclear, such a finding is not unprecedented as it has been shown for *S. meliloti* that the core Nod chitin backbone has a second role of stabilising biofilms [424]. Alternatively, this gene could represent a fatty acid metabolising enoyl-CoA hydratase, as also suggested by UniProt annotation.

The role of *RL1549* (transmembrane efflux protein), *RL1090* (DUF family protein) and the cysteine biosynthesis protein CysZ (RL4210) in root attachment is unknown.

Table 5-10. Ten genes (out of 40[a]) with discernible roles identified as required (ES/DE) for attachment to barley roots only (pH 7.0).

Gene(s)	Description
RL0726	Conserved hypothetical exported protein, transglycosylase Slt domain. Degrades peptidoglycan via β 1-4 glycosidic bond cleavage. Linked to biofilm formation in *S. enterica*, *E. coli* and *Acinetobacter baumannii* [314–316]. Lytic transglycosylases participate extensively in cell wall remodelling, recycling of peptidoglycan and space-making for insertion of cell-envelope spanning structures [317]. A lytic transglycosylase (RL4716) was characterised in Rlv3841 as required for cell envelope function and biofilm formation [318]. Also impaired in attachment to pea roots under pH 6.5 conditions
celB (*RL1647*)	Cyclic di-GMP binding protein, cellulose synthesis protein. Binds the cellulose synthase activator and is involved in cellulose biosynthesis. Also has a role in the synthesis of lipid-linked intermediates, although the exact function in this role is unknown [428]
degQ (*RL1806*)	Periplasmic serine endoprotease DegQ. In *B. subtilis* DegQ stimulates phosphotransfer to a transcriptional regulator affecting biofilm formation, promoting transition from a motile to sessile attached state [370]. Also impaired in attachment to pea roots under pH 6.5 and 7.5 conditions
RL2039	Putative HesB family protein, Fe/S biosynthesis. *E. coli* show a link between Fe/S biogenesis mutants and cell surface properties, which may be impaired in biofilm formation and/or motility, although the mechanisms underpinning this remain unknown [429]
sixA (*RL2644*)	Phosphohistidine phosphatase SixA, conserved. SixA is the only known bacterial phosphohistidine phosphatase, and dephosphorylates Npr in *E.coli* [343]. Implicated in biofilm formation in *E. coli* [344]. See also Table 4-10, *rl0032 / rl0033*. Also impaired in attachment to pea under pH 7.5 conditions
RL2656	Putative D-alanyl-D-alanine carboxypeptidase. Carboxypeptidases are extensively involved in peptidoglycan biosynthesis and modification [416]

RL2657	Putative GGDEF/GAF domain sensory box protein, downstream of *RL2656* (above). Likely catalyses synthesis and/or hydrolysis of cyclic di-GMP (c-di-GMP), key second messenger in biofilm formation / motile to sessile lifestyle switch [79]
RL3252	Periplasmic serine endoprotease, DegP-like. In *V. cholerae* DegP associates with the outer membrane and is an important determinant of the biofilm matrix structure [430], facilitating secretion of proteases which may help biofilms adhere to surfaces by processing extracellular components [431]
hflC (*RL3253*), *hflK* (*RL3254*)	Putative transmembrane serine proteases HflC and HflK. Functional HflC and HflK also modulate HflB activity. HflB, an AAA metalloprotease, is involved in membrane protein regulation, LPS biosynthesis and biofilm formation in *E. coli*, *B. subtillis* and others (where it is often called FtsH) [345, 346], indicating a role in membrane regulation and biofilm formation [347–349]. HflB itself (*RL3965*) is ES/DE under all conditions in this work, including input (Appendix 2 Table 2). Also impaired in attachment to pea under pH 7.5 conditions

[a]Number of genes are from the Venn diagram shown in Figure 5-5.

Amongst the list of genes required for attachment to barley roots (Table 5-10), requirement of *celB* (*RL1647,* encoding a cellulose synthesis protein) was interesting as cellulosic fibrils are normally considered a secondary attachment factor [34]. It may be that cellulosic fibril deposition is important in primary root attachment in barley, or that this gene is playing an unidentified role in synthesis of an attachment-required lipoprotein [428] in attachment to these roots.

RL2656 (carboxypeptidase) is likely to be involved in altering peptidoglycan structures. Two serine endoproteases (encoded by *RL3252* and *degQ - RL1806*) are also needed. The former seems to process extracellular components for attachment competence, while the latter promotes activity of transcriptional regulators involved in attachment [370, 431]. The requirement for *RL0726* (transglycosylase) as well as *hflC* (*RL3253*) and *hflK* (*RL3254*) indicate membrane and cell wall alterations are involved in primary attachment to barley roots. The phosphohistidine phosphatase SixA (RL2644), whose function is to dephosphorylate Npr [343], is required for attachment to barley and also for attachment to pea at pH 7.5 (Table 4-12). This is likely to indicate the involvement of down-regulation of ATP-dependent ABC transporters and export of certain EPS structures and is discussed more fully in 4.2.11.

The role of *RL2039* (an Fe/S biosynthesis protein) remains unclear, though it has been linked previously to impaired attachment in *E. coli* [429].

Table 5-11. Six[a] genes identified as required (ES/DE) for attachment to pea and soybean roots at pH 7.0.

Gene	Description
pRL100220	Uncharacterised protein
RL1052	Uncharacterised protein
RL1371	Putative transmembrane protein
RL2400	Putative MarC family transmembrane protein, not involved in antibiotic resistance [280], function unknown
RL2520	Putative transmembrane protein, ABC transporter permease. Many ABC transporters are involved in lipid transport to the outer membrane [361]
RL4083	Uncharacterised protein, SGHN family esterase domain

[a]Number of genes are from the Venn diagram shown in Figure 5-5.

After pleiotropy filtering (section 5.2.8) only six genes are required (ES/DE) for attachment to both legumes (pea and soybean) used in this study. Their functions are largely unknown, although *RL2520* may be involved in outer membrane lipid transport [361]. Given the importance of outer membrane characteristics, including lipid profiles, for surface attachment [44], mutation in this gene may reduce attachment via alterations to the outer membrane lipid profile.

Table 5-11. Two (out of 11[a]) genes with discernible roles required (ES/DE) for attachment to pea and barley roots (pH 7.0).

Gene(s)	Description
RL4309	Putative transmembrane protein. 94% identity to *R. hidalgonense* DedA family protein (CO674_30990). DedA proteins appear to function in membrane homeostasis; mutants show altered membrane lipid composition in multiple bacterial species [362].
RL4382	Putative filamentous hemagglutinin adhesin (*rl4382*). See Figure 3-11 and section 3.2.9

[a]Number of genes are from the Venn diagram shown in Figure 5-5.

For attachment to pea and barley roots, only two genes had discernible function. *RL4382* encodes FHA, which is a newly defined important factor for attachment to pea roots at pH 6.5, 7.0 and 7.5 (discussed in detail in 3.2.9). The second gene, *RL4309*, is likely to encode a DedA protein which is linked to membrane lipid composition [362].

Table 5-12. Ten genes (out of 38[a]) with discernible roles required (ES/DE) for attachment to barley and soybean roots (pH 7.0).

Gene	Description
RL1165	*rl1165* – uncharacterised protein, 93% identity to gene RLV_3555 from *R. leguminosarum* biovar *viciae*, PepSY domain containing. These domains are likely to have a protease inhibitory function and may be cell wall associated [334]. Biofilm metalloprotease 1 (BmpI) from *Pseudoalteromonas* contains a PepSY domain required for biofilm formation [335]. Also impaired in attachment to pea under pH 7.5 conditions
RL1440	Serine endoprotease, DegP-like. In *V. cholerae* DegP associates with the outer membrane and is an important determinant of the biofilm matrix structure [430], facilitating secretion of proteases which may help biofilms adhere to surfaces by processing extracellular components [431]
dacF (*RL2477*)	Putative penicillin binding protein DacF. 92% identity to *Rlv trifolii* WSM2297 D-alanyl-D-alanine carboxypeptidase. Carboxypeptidases are extensively involved in peptidoglycan biosynthesis and modification [416]
RL2595	Putative MutT/nudix family protein. Mutagenesis of nudix proteins in *Pseudomonas syringae* str DC3000 and *P. aeruginosa* display defects in motility and biofilm formation [323]. Also impaired in attachment to pea under pH 6.5 conditions
RL2778	Putative exopolysaccharide biosynthesis protein. Also impaired in attachment to pea under pH 7.5 conditions
RL3267	Putative OmpA family outer membrane protein. OmpA family proteins have diverse roles, including in signal transduction, primary surface adhesion and biofilm formation [265, 432, 433]
RL3320	Putative signalling and peptidoglycan binding protein. N-acetylmuramoyl-L-alanine amidase domain; cleaves the amide bond between N-acetylmuramoyl and L-amine acids in bacterial cell walls [434]
RL4018	Putative ATP binding component of ABC transporter. 96% identity to Lipid A ABC exporter from *R. leguminosarum* biovar *trifolii* WSM2304, gene Rleg2_3249. Mutants with reduced lipid A show a delay in nodulation onset and impaired

	bacteroid shape [350]. Further, defects in lipid A production reduce surface attachment and motility [351]. Also impaired in attachment to pea under pH 7.5 conditions
RL4356	Uncharacterised protein, YkuD domain. This domain can act as an L,D-transpeptidase, giving rise to an alternative pathway for peptidoglycan cross-linkage. In *S. typhi*, toxin secretion requires the localised editing of peptidoglycan by a specific YkuD family L,D-transpeptidase, and in *E. coli* L,D-transpeptidases anchor the Braun lipoprotein to peptidoglycan.
sirA (*RL4357*)	Putative transcriptional regulator SirA. Conserved (with different names) throughout γ-proteobacteria where it positively regulates virulence gene expression, exoenzyme and antibiotic production, motility and biofilm formation [435]

[a]Number of genes are from the Venn diagram shown in Figure 5-5.

From this list of requirements for attachment to soybean and barley roots (Table 5-12), membrane, cell wall and EPS modifications are clearly important. Genes required include *sirA* (*RL4357*), *RL1165* (encoding PepSY-domain protein), *RL1440* (encoding a DegP-like serine endoprotease), *RL2477* (*dacF*, encoding a penicillin binding protein), *RL2778* (encoding a putative EPS biosynthesis protein), *RL3320*, *RL4018* (likely encoding a lipid A exporter) and *RL4356* (encoding a transpeptidase). Also involved are *RL3267*, encoding an OmpA family outer membrane protein of which many have adhesin roles [265, 432, 433] and *RL2595*, encoding a MutX/Nudix family protein.

5.2.12 Mutation of some Rlv3841 genes leads to an increase in primary attachment to different plants

In INSeq experiments, genes are classified as AD when insertion into a gene increases bacterium fitness in, in this case, attachment to roots. This indicates that gene disruption is beneficial to bacterial root attachment and that an AD gene, when functional, somehow inhibits/reduces attachment. After classification and filtering of genes for pleiotropic effecs (see 5.2.6 and 5.2.8), 17 genes remained with an AD classification under one or more conditions, represented in Figure 5-9 (full list of genes in Appendix 1, Table A17). For brevity, only factors with putative functions are listed (Table 5-13) and discussed below.

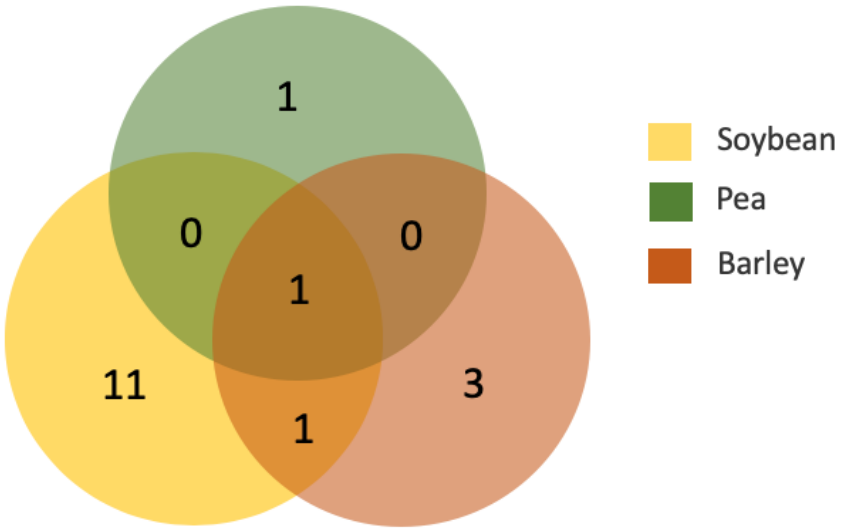


Figure 5-9. Genes classified as NE in the input library and AD in one or more of the root-attached libraries which are also classified as NE in the following INSeq datasets: VMM, 21% oxygen 10mM glucose, 1% oxygen 10mM glucose, 21% oxygen 20mM succinate and 1% oxygen 20mM succinate (see Table 4-7).Venn diagram circle color indicates attachment condition; yellow = soybean roots (pH 7.0), green = pea roots (pH 7.0), brown = barley roots (pH 7.0).

Table 5-13. Three (out of 17[a]) genes with discernible roles identified where mutation is advantageous (AD) for primary attachment to pea, soybean and barley roots (pH 7.0).

Gene	Plants	Description
pRL70156	Pea	Putative conjugative DNA transfer protein TrbGp7. A homologue of VirB9, a type VI secretion system protein in *Agrobacterium tumefaciens* [225]
RL3927	Soybean	Soluble lytic murein transglycosylase protein, Slt domain. Degrades peptidoglycan via β 1-4 glycosidic bond cleavage. Lytic transglycosylases participate extensively in cell wall remodelling, recycling of peptidoglycan and space-making for insertion of cell-envelope spanning structures [317].
RL4139	Pea, Soybean, Barley	Putative transmembrane sensory box GGDEF/EAL protein. The GGDEF domain is likely to catalyze synthesis and/or hydrolysis of c-di-GMP, whereas EAL may function as a diguanylate phosphodiesterase [79]. c-di-GMP is a known positive regulator of biofilm formation [436–438]. However, an EAL domain protein in *Pseudomonas putida* strongly inhibited biofilm formation and could also lead to rapid biofilm dispersal, indicating these proteins also have a role in transitioning cells from a sessile to motile state [439]

[a] Number of genes is from the Venn diagram shown in Figure 5-9.

The three genes in which mutation leads to increased attachment with discernable roles (Table 5-13) include, specifically for pea roots, *pRL70156* (putative *TrbGp7*), which encodes a putative DNA transfer protein and is homologous to the type IV secretion system component VirB9 in *A. tumefaciens* [225]. Other Trb proteins (involved in pilus formation and type IV secretion system function) are required for attachment to soybean roots (5.2.7 and Table 5-3), but this gene impairs Rlv3841 attachment to pea.

Showing an increased attachment only in soybean is mutation of *RL3927*. RL3927 seems to have a role in cell wall remodelling. Although another similar Slt domain transglycosylase protein (RL0726) was required (ES/DE) for attachment to roots of pea (pH 6.5, Table 4-10) and barley (pH 7.0, Table 5-10), the cell wall remodelling promoted by this gene is acting in the opposite way, to impair/reduce Rlv3841 attachment to soybean (and is thus classified AD when mutated), though the reason remains unclear. What is clear, is that factors affecting cell surface, such as cell walls and pili, influence primary attachment and that disruption of these structures perturbs, either to decrease or increase, the ability of Rlv3841 to attach to roots.

Of interest is *RL4139* (putative transmembrane sensory box GGDEF/EAL protein), AD when mutated in attachment to all plant roots at pH 7.0. Levels of the second messenger c-di-GMP have a role in biofilm formation [436–438] and lifestyle changes from a motile to sessile state [87, 88, 439]. In the cell, c-di-GMP is formed by GGDEF domain proteins and broken down by EAL domain proteins [82, 83]. Some proteins (including RL4139) carry both GGDEF and EAL domains. For these proteins, it is usually the case that only one of the two domains is functional, or that a third regulatory domain manages protein activity [84, 85]. An EAL domain protein was shown to inhibit biofilm formation and promote single celled, motile lifestyle in *P. putida*, likely through decreases in

intracellular c-di-GMP [439]. EAL protein mediated decreases in c-di-GMP have also been demonstrated in *P. aeruginosa*, *Yersinia pestis* and *S. enterica* [80, 440, 441].

The regulator *RL4145* (UniProt annotation: putative LacI type transcriptional regulator, repressor) is the only regulator required for attachment to all plant roots (Table 5-7, figure 5-8). Putative targets of this regulator were discussed in 4.2.12. However, if this regulator is a repressor then HMM classifications and genomic proximity link it to *RL4139*. Given the data presented, the following model is possible: under neutral pH conditions functional *RL4145* represses *RL4139* expression, enabling intracellular c-di-GMP levels to increase, facilitating successful primary root attachment. When *RL4145* is mutated an ES/DE classification results, as *RL4139* expression is not repressed and intracellular c-di-GMP decreases, impairing attachment. When *RL4139* is mutated an AD classification results, as c-di-GMP levels increase without the action of the protein's EAL domain. This model would account for the opposing HMM classifications of *RL4145* and *RL4139*. If correct, this interplay is extremely important in governing pH 7.0 primary root attachment (the HMM classification of *RL4139* at pH 6.5 and 7.5 in pea root attachment is NE), and an RNASeq study with Rlv3841 and an *RL4145* mutant as well as promoter binding assays could investigate this further. Alternative signalling pathways are likely to be operating under non-neutral conditions, especially considering the differential regulatory requirements and involvement of other putative c-di-GMP signalling factors (see 4.2.11) Overall, this reinforces the notion that high levels of intracellular c-di-GMP are very important for root attachment (as seen with *RL2316* – pH 6.5 attachment to pea – and *RL2657* – pH 7.0 attachment to barley) and that reduction of c-di-GMP signalling inhibits this process.

5.3 Conclusion

In summary, appropriate experimental parameters were established to screen in an INSeq study for bacterial primary root attachment to soybean and barley roots at pH 7.0. 127 million barcoded sequencing reads were obtained from nine sample libraries, achieving HMM classifications for 99.7% of the Rlv3841 genes. Validation of the results from INSeq by comparison with the literature indicated that the HMM method returned the expected classification for genes of known function (such as ribosomal subunits and plasmid replication systems) under the experimental conditions used here.

Prior to filtering based on pleiotropy effects under media/metabolic conditions, 464 genes classified as NE in the input libraries were classified as required (ES/DE) for attachment to the roots of one or more plants. Their distribution (Figure 5-2) indicated that there were multiple different attachment requirements unique to each plant. Examining the functional classification of these genes indicated that, whilst there were different functional classifications based on plant root attachment conditions, there were also significant requirements for uncharacterised genes in all cases.

The genomic localization of these genes revealed that, whilst chromosomal factors were the largest constituent of genes required for attachment to different plant's roots, the plasmid with the highest number of required genes in pea was pRL10. Plasmid pRL10 is referred to as the symbiosis plasmid, containing *nod*, *nif* and *fix* genes [381]. After pleiotropy filtering to remove any genes affected in growth under different conditions or substrates, the only pRL10 genes remaining were transmembrane proteins (and uncharacterised proteins). This work suggests a role for the symbiosis plasmid in primary root attachment to peas, although the precise function of these genes (*pRL100053, pRL100174* and *pRL100242*) remains unknown.

Soybean attachment factors showed the largest plasmid requirement for pRL11, along with a requirement for pRL8, which was not seen under any other condition tested. Analysis of the likely function of pRL8 and pRL11 genes indicates a requirement for Flp/Tad pili and type IV secretion system-mediated attachment to soybean roots. This mechanism is not obviously used by Rlv3841 in attachment to other plant roots, but is known to be used for attachment by different rhizobia (including *S. meliloti* and *A. tumefaciens*) [44, 383].

After filtering of genes based on pleiotropy phenotype under media/metabolic conditions (see 5.2.8), 312 genes remained ES/DE in attachment to plant roots, showing a mixture of plant specific and non-specific requirements. Visualizing the 'specificity' of genes involved in Rlv3841 attachment to different plant roots revealed that a high percentage of Rlv3841 genes required (ES/DE) for soybean primary root attachment are not required for attachment to pea or barley. 86% of the genes required for attachment to soybean were involved only in attachment to soybean whereas, for pea and barley, this figure was ~30%. This suggests that legume and non-legume plants do not necessarily cluster together in terms of similarity of primary attachment determinants.

The high specificity of genes required for soybean root attachment was reflected in the high number of regulators involved (22 in total). This figure was three for pea and six for barley. Only one regulator (*RL4145*) was required for attachment to all plant roots. Analyzing the possible regulatory targets of this gene based on genomic proximity and HMM classification (see section 4.2.12) revealed that this regulator may control expression of nearby hypothetical proteins and a putative transmembrane protein. Of these three possible regulatory targets, only *RL4147* (a conserved hypothetical protein) was required for attachment to all plant roots. *RL4145* can be thought of as an important

regulator involved in primary root attachment and represents a good target for future RNASeq-based investigation of transcriptional networks underpinning attachment. Interestingly, one of the regulators required for soybean root attachment was CheY1, a chemotaxis response regulator. Analysis of further chemotaxis genes with higher TA site number (for which HMM classifications are likely to be more accurate) by both HMM classification and fitness values indicated a requirement for chemotactic motility in Rlv3841 attachment to soybean, not seen with other plant roots under these INSeq conditions. One possible explanation for this is a higher level of chemoattractant exudation from soybean roots (due to large root size) providing a stronger advantage, or selection, for those bacteria able to sense them, which may not be seen in experiments with smaller rooting systems. The requirement for the MsiR regulator (*RL2857*) in attachment to barley roots indicates that phytotoxin export is likely necessary for attachment and colonization of this root system. A summary of the different gene functions needed for the primary attachment of Rlv3841 to the roots of pea, soybean and barley is shown in Figure 5-10.

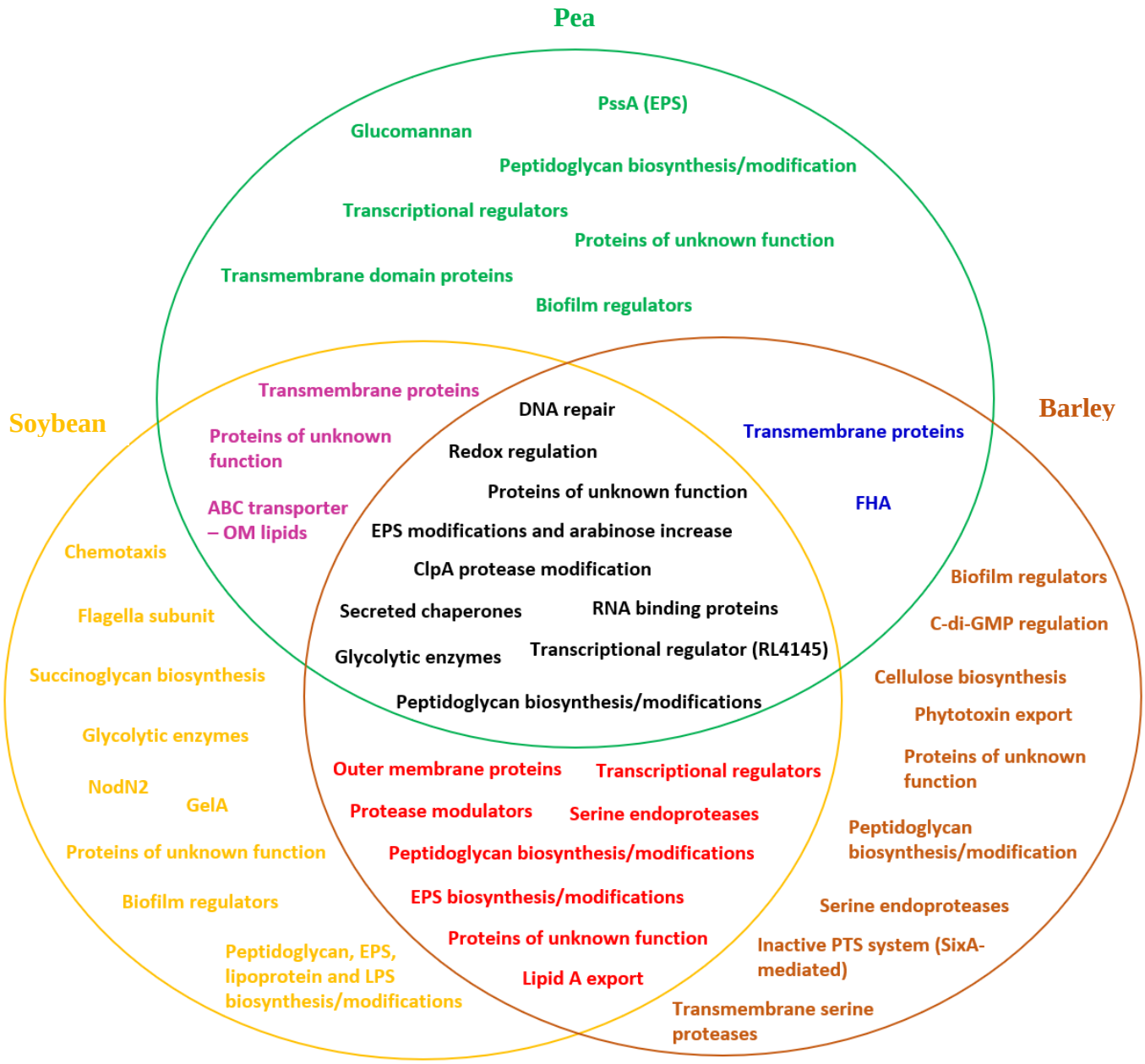


Figure 5-10. Gene functions needed for primary attachment of Rlv3841 to pea, soybean or barley roots and combinations thereof at pH 7.0. Circle color indicates plant: yellow = soybean, green = pea, brown = barley. Text color indicates plant requirement of indicated gene function for primary attachment: yellow = soybean, mauve = soybean and pea, green = pea, blue = pea and barley, brown = barley, red = barley and soybean, black = pea, soybean and barley. OM = outer membrane. Gene functions are drawn from Tables 5-7 to 5-13. See also Figure 5-8.

Twenty-four genes were classed as ES/DE in attachment to all plant roots, and of these 14 were required for attachment to pea under all pH conditions. These can be thought of as core attachment determinants. Among the functions indicated by these genes is peptidoglycan modification. Specifically, RL4362, a putative cobalamin synthesis protein which may actually represent a peptidoglycan amidation factor, and *dacC* (*RL4363*), a penicillin binding protein likely involved in peptidoglycan biosynthesis via sugar-peptide cell wall precursor processing [299]. EPS modification, particularly an increased arabinose content due to *pRL110043* (an arabinose efflux permease), is also required. This is of interest as arabinose content in known to be an important regulator of cell aggregation in *Azospirillum* [358] and is also important in Rlv3841 primary root attachment. Modulation of the ClpA protease (RL2213) by ClpS (RL2212) is necessary for attachment to pea, soybean and barley, though the targets of ClpA (with or without ClpS modulation) remain unclear [364, 365]. Glycolytic enzymes (such as TpiA, which may have a surface localized 'moonlighting' role in attachment [282–284]) and uncharacterised proteins were also implicated. Additionally, DNA repair factors which, although they may be required for coping with cellular stresses in the assay setup, may also play an uncharacterised role in attachment [286], are also required. Interestingly, RL4381 (a POTRA domain transporter) but not RL4382 (FHA, filamentous hemagglutinin adhesin) which it is thought to export, was required for attachment to all plants. This indicates that, whilst the role of FHA is important (being required for attachment to pea and barley roots), RL4381 may additionally export other primary attachment factor determinant(s).

As well as chemotaxis and Tad/Flp pili, attachment to soybean roots also required various sugar transferases and EPS/LPS/peptidoglycan biosynthesis and modification genes. *RL0064* (*pssR*) promotes EPS acetylation whilst *RL1015* (a polysaccharide deacetylase) is also required, indicating a possible balancing of EPS acetylation levels required to facilitate soybean root attachment. *RL0963* (a BA14K family protein) and *RL1155* (a glycosyltransferase) are involved in LPS biosynthesis whilst *dgkA* (*RL2780*) is an LPS modifier, with *E. coli dgkA* mutants defective in biofilm formation [316]. *redAh* (*pRL110048*) is a transferase involved in the membrane steps of peptidoglycan biosynthesis [401], whilst *pRL110439* and *RL1016* are peptidoglycan modifying factors. The involvement of GelA defines another role for this previously uncharacterised EPS biosynthesis protein [158] which is needed both for attachment to pea roots at pH 6.5 and soybean roots at pH 7.0. The requirement for *pRL110453* indicated that exported enolase may play a role in attachment and NodN2 (*RL4120*) indicated that a Nod biosynthetic gene product is likely also to be involved.

In attachment to barley roots, various genes involved in c-di-GMP synthesis as well as proteases were required. A Deg-P like serine endoprotease (*RL3252*) may process extracellular components to facilitate adherence, a Slt domain transglycosylase likely participates in cell wall remodelling and the serine proteases DegQ, HflC and HflK (all with roles in membrane regulation/biofilm formation) are also required. *RL2656* (a carboxypeptidase) seems to be involved in peptidoglycan biosynthesis and modification [416]. The involvement of CelB (*RL1647*), a cellulose synthesis protein, is of interest as, whilst cellulosic fibrils are more often considered as a secondary attachment factor [34], they may also to play a role in barley primary root attachment, although *celB* could be playing a role in unknown lipoprotein synthesis [428]. The phosphohistidine phosphatase

SixA (discussed in 4.3), with a putative role in Npr dephosphorylation and levels of EPS secretion and/or modification also seems to be involved.

A putative transmembrane DedA protein (*RL4309*) with roles in regulatory membrane lipid composition was required for attachment to pea and barley (but not soybean) roots, indicating differential membrane requirements for interactions between these plants. Additionally, *RL4382* (encoding FHA) was also required. This result means that, whilst FHA is a key factor at all pHs for pea root attachment (4.2.11), it also plays a role in Rlv3841/barley interactions.

Genes required for attachment to soybean and barley roots were also mainly of the cell wall/extracellular protein/peptidoglycan synthesis and modification classes. These included *RL1165*, a cell wall associated PepSY domain protein likely to modulate proteases and promote biofilm formation [334, 335] and *RL1440*, a DegP-like serine endoprotease. In *V. cholerae* DegP facilitates processing of extracellular components to encourage biofilm formation [431]. Further, *dacF* (*RL2477*, a likely carboxypeptidase) and *RL3320* (a peptidoglycan binding and cleavage factor [434]) are also necessary. The SirA virulence and biofilm regulator (*RL4357*), OmpA adhesin factor (*RL3267*), and a lipid A exporter (*RL4018*) were also involved.

As well as the Rlv3841 factors required for attachment to different plant roots, it is also informative to examine the factors required for attachment to pea roots (the host legume) which are not required for attachment to soybean (a non-host legume) and barley (a non-legume). One of these is glucomannan (*gmsA*, *RL1661*), the cell surface factor important for primary root attachment at acidic but not alkaline conditions [56]. This factor binds root hair lectins, which are thought to disassociate from roots under alkaline conditions [34]. A glucomannan mutant was demonstrated to be deficient in root attachment at pH

6.5 and 7.0, but not 7.5, in a Lux attachment assay (Figure 5-3). *gmsA* was classified as NE in attachment to soybean and barley roots at pH 7.0 (Appendix 2 Table 2). As *gmsA* is not required for attachment to these plants, it can be inferred that root lectins present are not recognized by Rlv3841 glucomannan. As lectins act as recognition molecules in cell-cell interactions [48, 49], this suggests that there is specificity in the Rlv3841 glucomannan/lectin interaction for attachment to the host legume, pea. *B. japonicum* attaches to the root hairs of its host legume (soybean) with a lectin-mediated system at acidic to neutral pHs [50, 51], but must use a factor different from Rlv3841 glucomannan to allow soybean lectin recognition and binding. This indicates that there is some specificity in attachment to host legume root hairs.

PssA (encoded by *RL3752*), a glycosyltransferase involved in the first steps of EPS biosynthesis, is also required for attachment to pea roots only. PssA is responsible for the addition of glucose-1-phosphate to a polyprenyl phosphate carrier, one of the first steps in EPS synthesis [442]. An *R. leguminosarum* biovar *trifolii* Rt270 strain carrying a *pssA* mutation was deficient in EPS production, and led to host clover plants forming fewer nodules [442]. Various EPS production and modifying factors are important for Rlv3841 attachment to soybean and barley. This, combined with the NE classification of *pssA* in attachment of Rlv3841 to these plants, indicates that enough EPS must be produced by alternative pathways to facilitate root attachment. In *R. leguminosarum* biovar *trifolii* 5599, *pssA* mutants still produce capsular polysaccharide (structurally very similar to EPS) at wild-type levels [443], indicating that production of closely related polysaccharides remains possible.

A further Rlv3841 factor involved only in attachment to pea is *pRL100053*, a putative transmembrane domain containing protein. Testing in a Lux assay confirmed a mutant in

this gene to be defective in attachment to pea roots at pH 6.5, 7.0 and 7.5 (Figure 3-11). Given that the INSeq classification for this gene is NE in the rhizosphere but ES/DE in pea root attachment, colonization, infection thread and nodule experiments (see Table 4-5), the product of this gene likely represents a novel root hair attachment factor. As *pRL100053* is not required for attachment to soybean or barley, this reinforces the idea of specificity in mechanisms of attachment to host legume root hairs.

After pleiotropy filtering, 17 genes were classified as AD, indicating that mutation was beneficial to primary root attachment. Of these, only three had discernible roles (see 5.2.12). The most important of these was *RL4139*, classified as AD in attachment to all plants when mutated. This is likely because of a c-di-GMP depleting role of this EAL domain protein when functional. Considering the genomic proximity and opposing HMM state calls of *RL4139* and the regulator *RL4145*, it seems possible that, when both are functional, the latter represses expression of the former. If correct, this is an important transcriptional regulatory insight into the process of primary root attachment and warrants further investigation.

This chapter has demonstrated that primary attachment determinants show a high degree of plant root specificity and has identified the ‘core’ factors that are important for attachment to all plant roots tested. These findings indicate that different host plants likely present very different root surfaces for attachment by Rlv3841 and, therefore, have differential attachment requirements. Intriguingly, it does not seem to hold that primary attachment mechanisms to legumes are more similar than to non-legumes, as soybean had the most differential requirements in comparison to pea and barley.

Especially considering the relatively small number of characterised primary attachment determinants in the literature [25], this work highlights the astonishing diversity of

primary attachment mechanisms that exist under different conditions and in interaction with different plants.

Chapter 6

Using real-time imaging to track early-stage interaction dynamics of *R. leguminosarum* with plant roots

6.1 Introduction

Plant roots are one of the most productive ecosystem environments in the topsoil and also have a dominant role in shaping the rhizosphere [188]. Multiple inter-species interactions between bacteria, fungi and plants are found in the rhizosphere and these are often shaped by differential plant root exudation, which has a strong selective pressure on microorganisms [444, 445]. Especially in agricultural crops, the constituents of the rhizosphere are important determinants of overall crop productivity [36–38].

The rhizosphere is a heterogeneous environment (even between plants of the same species) and this is often due to root exudate diversity and differential availability in spatial root zones. As an example, *A. thaliana* exudate profiling demonstrated the presence of many different chemical classes (amino acids, secondary metabolites and organic acids among them [446]). Further, Moussaieff *et al.* (2013) demonstrated that *Arabidopsis* root exudation varies widely by tissue [447]. This provides weight to the notion (also proposed by others) that root exudation varies spatially along roots as well as according to developmental timepoint [185–187]. Spatial variation in root exudation has been demonstrated in the annual grass *Avena barbata*, with significantly higher sugar exudation nearing the root tip and tryptophan exudation increasing toward lateral regions [184]. More recently, Pini *et al.* (2017) used 14 different luminescent biosensor-carrying strains of Rlv3841 to demonstrate spatial and temporal characteristics of exudation for compounds including sugars, amino acids and flavonoids from pea roots [201].

Given the importance of root exudates in shaping the rhizosphere community and acting as chemoattractants for soil microorganisms [34, 212, 213, 398], it would be expected that these exudates might play a key role in establishing root-microbe spatio-temporal interaction dynamics. Not only this, but differential exudation gradients may also govern

root site preference for microbial attachment and colonization. Imaging techniques have the potential to enable examination not only of physical root microbe interaction dynamics, but also patterns of bacterial gene expression if appropriate reporters are used. This would be of interest for further dissecting the primary/secondary root attachment processes (secondary root attachment typically involves upregulation of multiple inducible factors, [34]) and examining any spatio-temporal dynamics in bacterial gene expression which may occur during bacterial root attachment or colonization.

Recently, microfluidic approaches have emerged as powerful imaging tools to investigate these questions. Such systems usually rely on miniaturized and transparent chambers into which plant roots grow before imaging, which can be with or without interacting microorganisms. The principal advantages of microfluidics systems for root imaging include the tightly controlled microenvironment of the imaging chamber (allowing precise control of nutrient parameters and establishment of laminar fluid flows, for example, [448]), and the miniaturization of the experimental setup, permitting relatively easy microscopic imaging of the plant roots.

An early example of this is RootChip [449]. This microfluidic device featured an array of small channels through which *Arabidopsis* roots could grow linked to control pipes which allowed rapid delivery of different medias to different roots. Using transgenic plants with fluorescent glucose and galactose biosensors, plant cytosolic sugar levels could be monitored at the subcellular level as environmental conditions changed [449].

A development of this technology came from Busch *et al*. (2012), who developed RootArray. In the microfluidics setup, 64 *Arabidopsis* seedlings were grown simultaneously, and roots imaged with confocal microscopy. By using transgenic reporter lines and imaging regions of interest at high magnification, data from hundreds of roots

could be leveraged to investigate expression patterns of genes during plant root development [450].

In addition to this, Jiang *et al*. (2014) demonstrated TipChip. This device, also using *Arabidopsis*, is a vertical microfluidics device in which gravitropic root growth could be characterised in wild-type and *immutans* mutant lines. However, this study went further in that interactions of *Arabidopsis* with a fungal plant pathogen (*Phytophthora sojae*) were observed, with zoospores visible on roots at 31 hours [451].

Massalha *et al*. (2017) became the first to report the imaging of bacterial interaction with plant roots in a microfluidics setup termed tracking root interactions system (TRIS). This used a transparent polydimethylsiloxane (PDMS) device cast from a silicon master mould featuring 160 μM channels for roots to grow through, as well as inlets and outlets for addition of fluorescently labelled bacteria. Each TRIS device contained nine independent channels, allowing imaging of multiple roots in parallel with confocal microscopy. By inoculating roots with labelled *Bacillus subtilis* (a well-known PGPR, [452–454]), directed attraction of bacteria toward the root elongation zone (REZ) was observed, likely driven by chemotaxis toward a high exudate concentration in this region. It was further observed that *B. subtilis* excluded *E. coli* from the root surface (a possible physical protection against plant pathogens) and that a dual-channel TRIS setup could be used to investigate root-microbe interactions in the presence of different root genotypes [188]. A schematic of the TRIS setup as well as an example image of the root-microbe interactions observed, are shown in Figure 6-1.

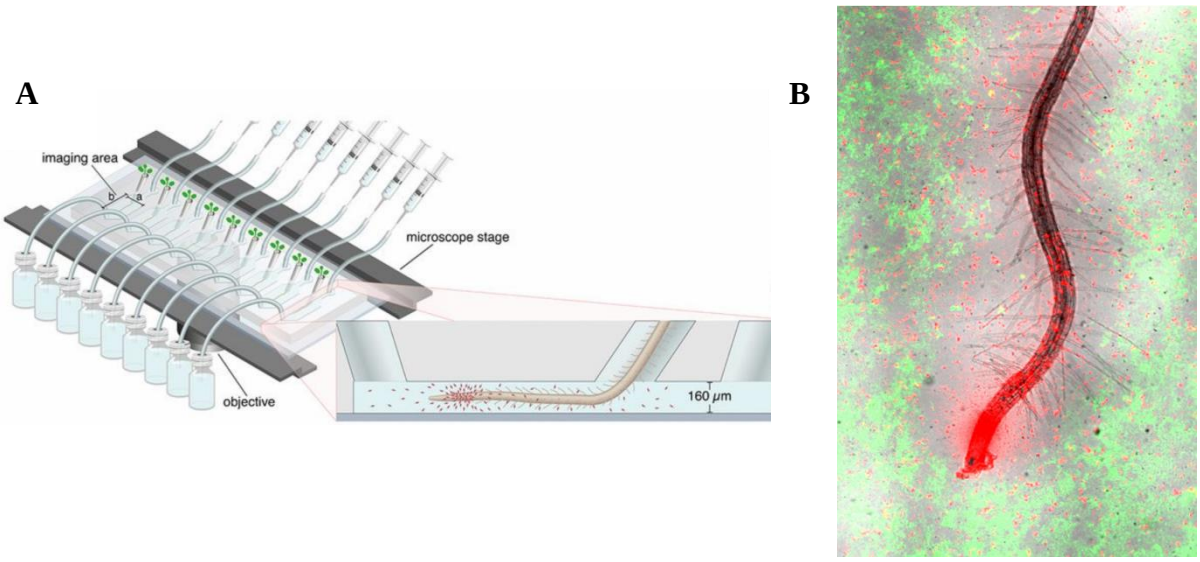


Figure 6-1. TRIS and example root-microbe interactions. A – illustration of the TRIS device on a microscope stage with imaging area and objective also indicated. Inset - section of a TRIS channel with an *Arabidopsis* root and *Bacillus* cells (red, not to scale) with the inlet and outlet channels also visible at either end of the channel. B – Interaction of an *Arabidopsis* root with *B. subtilis* (mKate labelled; red) and *E. coli* (GFP labelled; green) cells 12 hours post-inoculation demonstrating *B. subtilis* interaction with the REZ, and an exclusion of *E. coli* from the root surface. Images adapted from [188].

Since this time, further studies have implemented similar microfluidics setups to investigate *Arabidopsis*-PGPR interactions [455], *Arabidopsis* root development [456], bacterial communication between separated microenvironments [457] and nematode behaviors in soil microenvironments [458]. An enlarged version of the TRIS system has even been used to study the interaction of *P. fluorescens* with *Populus tremuloides* (aspen tree) roots [189]. Despite the extraordinary insights provided by these imaging technologies, there remains no reported application of these techniques to the study of root-microbe interactions in *Rhizobium*-legume symbioses. Instead, techniques such as

confocal or scanning electron microscope imaging of *Rhizobium* on roots have been used, sometimes in conjunction with fluorescence *in situ* hybridization of probes to bacteria to visualize colonization patterns [459]. These techniques usually result in partial imaging of plant roots and do not provide 'live' continuous imaging. Therefore, only a snapshot in time of root-microbe interactions is gained [459].

Such interactions are of great importance to symbiosis formation, as positioning of bacteria in physical proximity to the root is required for attachment [34]. Little is known about which region of the root might be important for this, how interaction dynamics change in the presence of different bacterial species or strains, and what spatio-temporal patterns of gene expression may guide attachment to different root zones.

In this chapter the applicability of microfluidics devices and real time imaging for the visualization of spatio-temporal reporter gene expression in Rlv3841 interactions with legume roots is demonstrated. The development of a new imaging platform and investigation of the role of bacterial motility in establishing early root interaction dynamics is also undertaken. This work constitutes an exploratory investigation of applying imaging systems to *Rhizobium*-legume interactions and provides a basis for further development and testing of such systems.

The root-microbe interaction video data referenced in this Chapter can be found in the Oxford Research Archive at the following link (also given in Appendix 2):

https://doi.org/10.5287/bodleian:w4AxBzN4k

6.2 Results and discussion

6.2.1 Establishing growth conditions for motile Rlv3841 cultures

Massalha *et al* (2017) reported that bacterial culture motility was important for observing root-microbe interaction dynamics between *Bacillus* and *Arabidopsis* roots [188]. Prior to investigating root-microbe interaction dynamics with Rlv3841, it was also important to ensure bacterial culture conditions were appropriate for TRIS inoculum preparation. Bacterial motility is strongly linked to growth phase and carbon source/nutrient availability. Later growth phases and lower nutrient availability can act as triggers for cell motility, enabling them to seek out more favorable environments [421, 460]. Therefore, low-carbon liquid minimal media culture conditions were used (see 2.2.1 and 2.8.4). In TRIS, it is also important that bacterial inoculum is washed such that remaining nutrients from growth are as dilute as possible, minimizing interference with chemotactic signals from roots [188]. Here, bacterial cultures were washed as described in 2.8.4. A similar washing protocol was used by Massalha *et al.* (2017), and remaining trace media components were not seen to interfere with root interaction dynamics [188]. Bacterial motility was visualized using a dark field microscope (2.8.4), and an example processed image demonstrating high culture motility is shown in Figure 6-2, below.

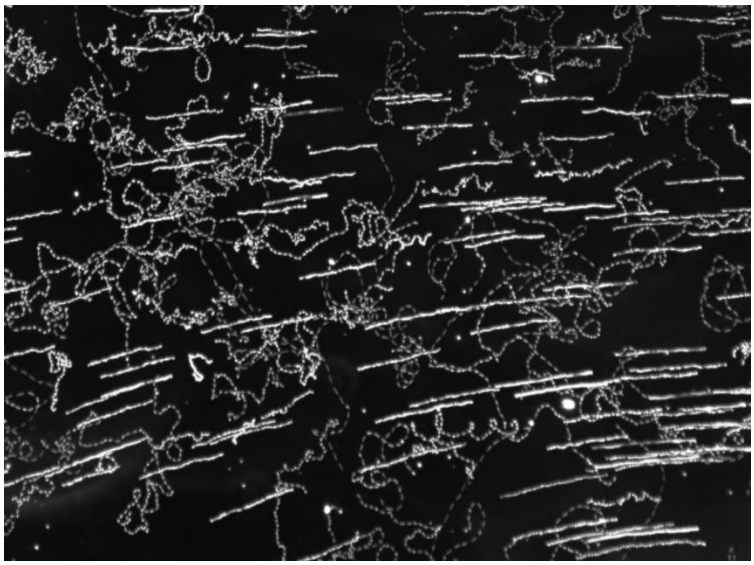


Figure 6-2. Tracking motility of Rlv3841 culture grown in low-carbon conditions. Bacteria were washed and prepared for TRIS (2.8.4) and observed using 10 x magnification with dark field imaging. CellSens software (Olympus) processed images for cell tracking. This image shows a representative culture sample where individual cell movements have been tracked for 5 seconds. Dashed trails (in white) overlaid on the image by CellSens show bacterial motility.

6.2.2 Evaluating root diameter for TRIS compatibility

The channels for root growth in TRIS are 160 μM in diameter. This is large enough to accommodate small *Arabidopsis* roots (which had a reported diameter of ~100 μM when germinated through pipette tips for the TRIS setup [188]) but legume plant roots (particularly vetch, also a cognate symbiont for Rlv3841) are typically larger in diameter. To evaluate root suitability for use in TRIS, seeds of four plants; two vetches (*Vicia cracca* and *Vicia hirsute*) and two clovers (*Trifolium repens* and *Trifolium pratense*) were germinated through pipette tips as described in [188] (see also 2.8.2). This method allows tips with growing roots to be connected to TRIS chambers without disturbing roots. Once

roots had emerged to ~10 mm below the tip, they were imaged using a Leica M165 FC microscope with LAS AF software for measuring root diameter in images. The results of these measurements are given in Table 6-1.

Table 6-1. Average root diameter of legume plants germinated through pipette tips.

Plant	Average root diameter
V. cracca	166 ± 40 μM
V. hirsuta	250 ± 60 μM
T. repens	104 ± 15 μM
T. pratense	125 ± 30 μM

Data is provided as mean ± SEM, n=5.

Based on these measurements, the roots of both vetch species were too large for TRIS as originally developed meaning that, of plants tested here, only the clovers were suitable.

6.2.3 Preliminary reporter gene testing using a luminescence promoter fusion

Microfluidics imaging systems examining root-microbe interactions could be used to examine spatio-temporal profiles of reporter gene expression, which would be useful in characterising gene expression patterns underpinning attachment and colonization in different root zones. *lppE* (*RL3234*, encoding a putative lipoprotein) has been reported to be significantly upregulated in the pea rhizosphere [141]. Further, the *lppE* promoter has been shown to be active specifically near the REZ, likely in response to specific exudate compounds (Poole lab, unpublished data). The REZ is where newly derived cells from the root meristem elongate before becoming part of mature root tissue [461]. Although the role of *lppE* in attachment and colonization remains unclear (*lppE* was classified as NE

under all INSeq conditions tested in this work), its expression does seem to play a role in early interactions with the REZ. To validate this expression pattern, LMB487 (Rlv3841[pLMB579] - a 375bp promoter of *lppE* cloned upstream of the *luxCDABE* operon in pIJ11268, a luminescence reporter plasmid) was prepared and inoculated onto roots of *T. repens* and *T. pratense* as described in 2.8.3. The results of root imaging using a NightOWL II LB 983 camera are shown in Figure 6-3.

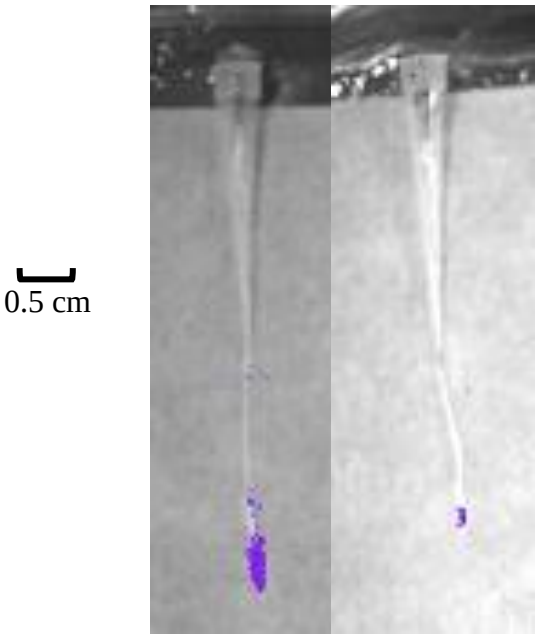


Figure 6-3. Imaging LMB487 on clover roots. Activation of the *lppE* reporter construct can be seen in the REZ of both *T. repens* (left) and *T.pratense* (right) at 2 hr post-inoculation. Representative images from 5 replicate experiments with each plant.

lppE reporter activation was seen localized to the REZ in both clover plant species tested. A larger signal was detected from *T. repens* (Figure 6-3, left), though this may be due to a larger root having higher exudation of promoter activating signals as opposed to a genuine stronger activation of *lppE* in *T. repens*. The activation of *lppE* in response to non-cognate

host roots is interesting, as it suggests a role for *lppE* in interaction with legume roots more broadly. Although the REZ boundaries can be defined by *in planta* reporter genes (as in [188]), such reporter lines were not available for use here.

Therefore, the more general definition of the REZ as given above was used to designate it in this work.

6.2.4 Reporter gene testing using TRIS

Having validated the spatio-temporal expression pattern of *lppE* in interaction with *T. repens* roots using a luminescence method, a GFP reporter strain (OPS0167) was used to investigate *lppE* activation using TRIS. *T. repens* were germinated and grown into TRIS channels, and roots inoculated with motile OPS0167 culture for imaging (see section 2.8). Although *lppE* expression was seen in the REZ after approximately 1 hr (data not shown), signal from the GFP reporter was at its strongest after 39 hrs, as shown in Figure 6-4 (see also Appendix 2 Video 1).

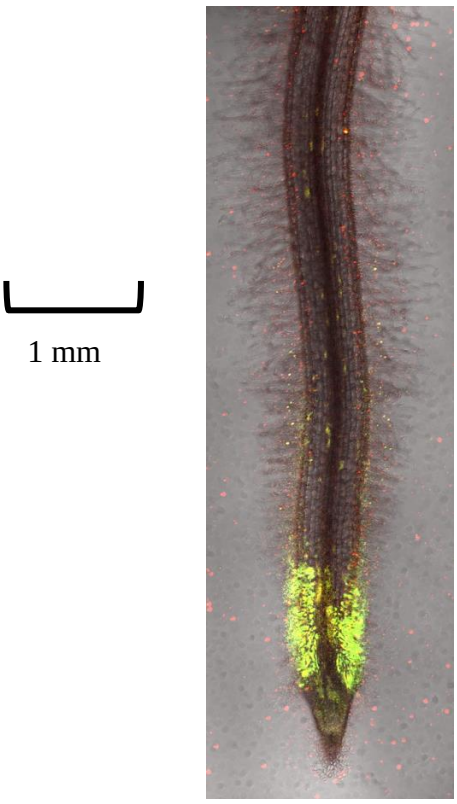


Figure 6-4. Imaging of OPS0167 with a *T. repens* root in the TRIS system. The Rlv3841 strain (OPS0167) is marked with a constitutively expressed mCherry gene, with GFP expression driven by the *lppE* promoter. The activation of the *lppE* reporter construct can be seen in the REZ, here at 39 hrs post-inoculation. See also Appendix 2 Video 1.

Expression of *lppE* in the REZ of *T. repens* could be seen clearly using TRIS. This result highlights the potential of such microfluidics-based imaging systems to characterise the spatio-temporal expression patterns of genes responsive to interactions with roots. If applied to the study of primary and secondary attachment factor expression patterns, this could generate insights into how attachment and colonization occurs in different root

regions. However, investigations of this sort would warrant a larger system than TRIS, where only ~6-8 mm of root length can be imaged due to the design of the chamber.

6.2.5 Developing Chamber Imaging and Interaction Profiling Systems (ChIIPS)

Work with TRIS demonstrated the potential for investigating root-microbe interaction dynamics between *Rhizobium* and legumes using microfluidics imaging platforms. However, the roots of some plants are too large to fit into the narrow TRIS channels (Table 6-1). Additionally, the short length imaging window (~6-8mm) means that, for larger roots, it may be difficult to image across different root zones. Further, if imaging is to take place with an upright confocal microscope (rather than an inverted microscope, as was used for TRIS), then the chamber structure must be more compact so as not to collide with the objective lens.

Given these considerations (as well as time constraints on this project which prevented a comprehensive reengineering of TRIS) a potential new alternative to TRIS was trialed which is lower cost, has reduced setup time, enables imaging of larger roots and is compatible with upright confocal microscopes. Named ChIIPS (chamber imaging and interaction profiling system), this method makes use of commercially available SPL Life Sciences cell culturing chamber slides as containers for plants and labelled bacteria. Plants are germinated in sterile conditions outside of the chamber before being aseptically transferred to a bed of 3% water agar solidified inside the chamber to 1cm depth. This post-germination transfer ensures that only plants with straight roots that can be more easily imaged are used. Bacterial cultures are prepared by filtering, washing and resuspending in FP media (see 2.8.4) before adding directly to fill the chamber, which is then sealed with parafilm for imaging. Such a simple design with commercially available

components ensures rapid and inexpensive experimental setup. The length of the chamber (~ 5 cm) and vastly increased depth compared to TRIS allows for imaging of much larger roots. A diagram of the ChIIPS chamber and a photo of a plant in the chamber are shown in Figure 6-5.

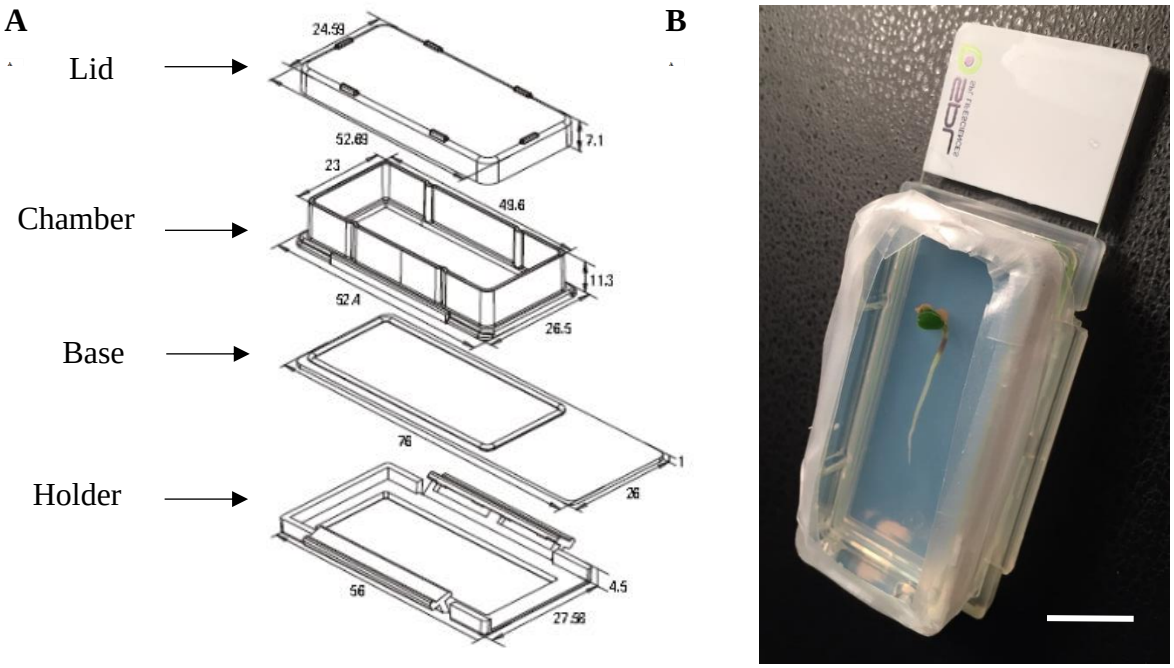


Figure 6-5. A diagram of the ChIIPS chamber and a photographic example. A – Design of the SPL Life Sciences cell culturing slide used as the basis for ChIIPS. The lid, chamber, base and holder are indicated. Measurements are given in millimeters. Image adapted from the SPL Life Sciences product catalogue. B – a photographic example of a clover plant in ChIIPS with bacterial inoculum, ready for imaging. White bar is for scale and represents 1 cm.

Although not strictly speaking a 'microfluidics' device, as there is no fluid flow once imaging begins, this was also the case for TRIS, where fluid flow was only on initial

addition of bacteria [188]. Unlike in TRIS, when using ChIIPS the aerial plant parts sit within the bacterial inoculum. Although this likely alters plant physiology and photosynthetic function, the direct transfer of pre-grown plants to the chambers allowed far quicker experimental setup in this preliminary work. Despite the limitations of plant leaf submersion (which will, if photosynthetic output is reduced, likely also reduce exudation [35, 462]), distinct root-microbe interaction patterns were still observable (see below), and this work provided a useful platform from which to develop *Rhizobium*-legume imaging technology further.

6.2.6 Rlv3841 interaction dynamics with legume roots in ChIIPS

Having developed a preliminary test system to allow the imaging of larger roots over a greater area, the interaction dynamics of OPS1734 (Rlv3841[pLMB449] – a reporter plasmid containing a *gfp* gene under control of the pTac promoter) with various legume roots were examined as described in 2.8.7. Interaction dynamics with *L. japonicus*, *Vicia villosa* and *M. sativa* revealed a strong preference for bacterial accumulation in the putative REZ. In all cases this was visible after 2.5 hrs (and often before) and generally peaked around 7.5 hours. Confocal images as well as plots of bacterial fluorescence intensity along the root axis for lotus, vetch and alfalfa are presented in Figure 6-6. See also Appendix 2 Videos 2, 3 and 4 respectively.

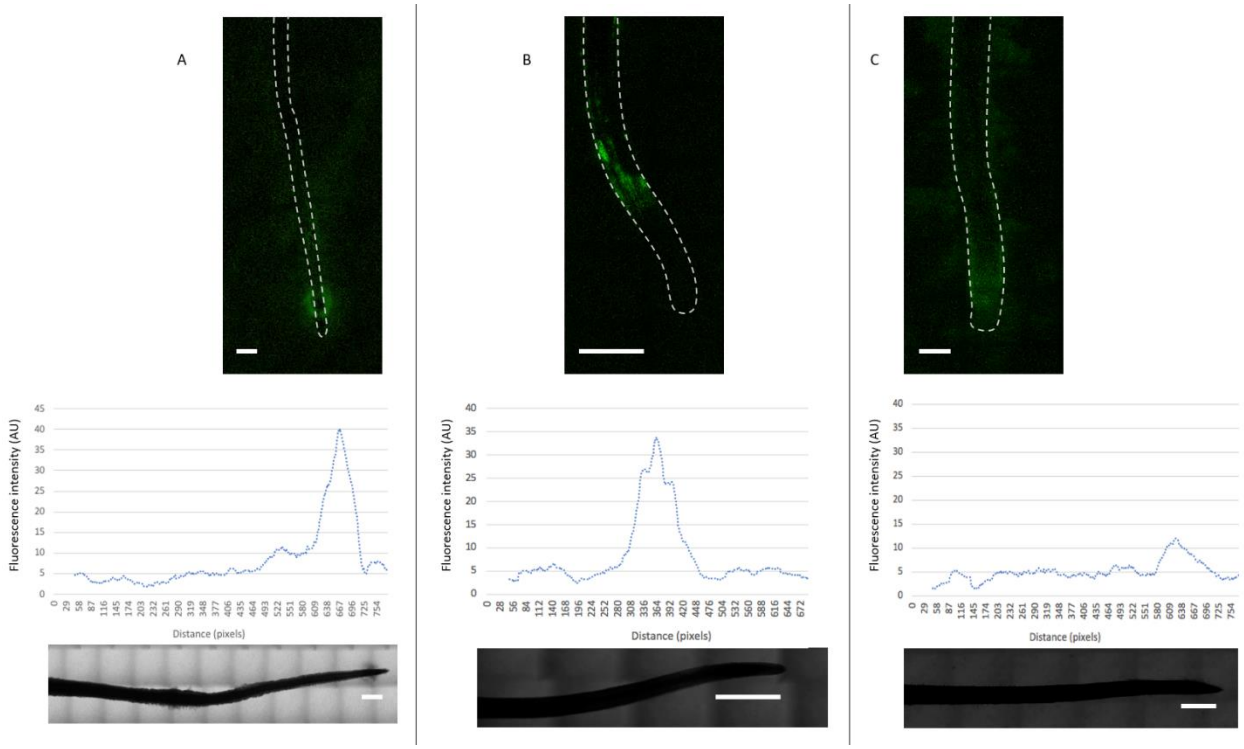


Figure 6-6. Rlv3841 interaction dynamics with legume roots in ChIIPS. A – upper panel: GFP signal from *L. japonicus* root (outlined in dashed white) imaged with OPS1734 inoculum at 7.5 hours post-inoculation. Lower panel: GFP fluorescence intensity (in arbitrary units, AU) along the edge of the root (shown below the x axis for reference) from mature zone to growing tip (Appendix 2 Video 2). B – as for A, but with *V. villosa* (Appendix 2 Video 3), C – as for A, but with *M. sativa* (Appendix 2 Video 4). White bars are for scale; each represents 3 mm.

As shown in Figure 6-6, a peak of bacterial fluorescence intensity (indicating higher density of cells) is seen ~ 3 mm back from the growing tip of the root, indicating a preference for interaction with the REZ. This was not unexpected; the REZ is a hotspot for sugar and other chemoattractant exudation [184, 461], and a similar interaction dynamic

was seen between *B. subtilis* and *Arabidopsis* [188]. However, this result indicates that this interaction pattern is likely to be widely conserved, particularly between soil bacteria and uncolonized roots. This interaction peaked at ~ 7.5 hrs post-inoculation in all cases, and generally decreased slightly thereafter (Appendix 2 Videos 2-4). This may have been due to reduced root exudation as a result of leaf submersion (discussed above) and/or loss of bacterial cell viability due to extended containment in the air-tight chamber, or laser irradiation from extended confocal imaging. Differences in peak signal intensity at 7.5 hrs between the three plants are likely due to differences in levels of exudation or different exudate components having higher or lower chemotactic potential with Rlv3841. Despite the unfavourable physiological conditions that submersion may expose plants to, this experimental setup was still able to demonstrate root-microbe interaction dynamics (Figure 6-6). A higher magnification image of OPS1734 interacting with an *M. sativa* root at 6 hrs is shown below (Figure 6-7) and illustrates this point further. However, for future work (which may wish to examine interactions over longer periods of time) it would be recommended to incorporate a design enabling the aerial plant parts to remain external from the chamber, and a future possible 'ChIIPS2' design is given in 6.2.9.

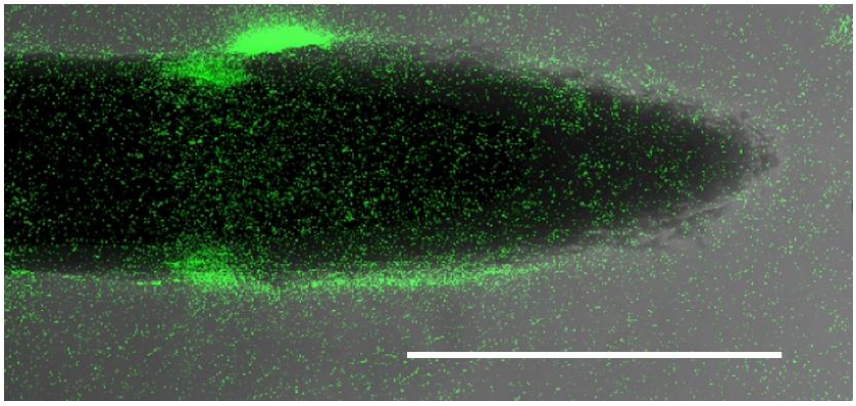


Figure 6-7. Imaging *M. sativa* root with OPS1734 6 hours post inoculation, GFP signal overlaid onto brightfield image. White bar is for scale and represents 3 mm.

6.2.7 The role of motility in early-stage interaction dynamics

Section 6.2.6 demonstrated the *Rhizobium*-legume interaction dynamics between Rlv3841 and various legume roots, with strong attraction to the REZ seen in all cases (Figure 6-6, 6-7). If chemoattraction toward higher root exudation from this zone is driving this interaction dynamic, then a non-motile inoculum should not show this attraction profile when interacting with legume roots. To test this hypothesis, a non-motile GFP marked strain, OPS1736 (OPS1709 (*motA*::pK19)[pLMB449] – *gfp* driven by a pTac promoter) was tested in the ChIIPS system with *V. villosa* and the results compared to interaction dynamics seen with OPS1734 (the motile Rlv3841 GFP strain). The results of confocal imaging and plotting bacterial fluorescence intensity along the root at 7.5 hrs are shown in Figure 6-8.

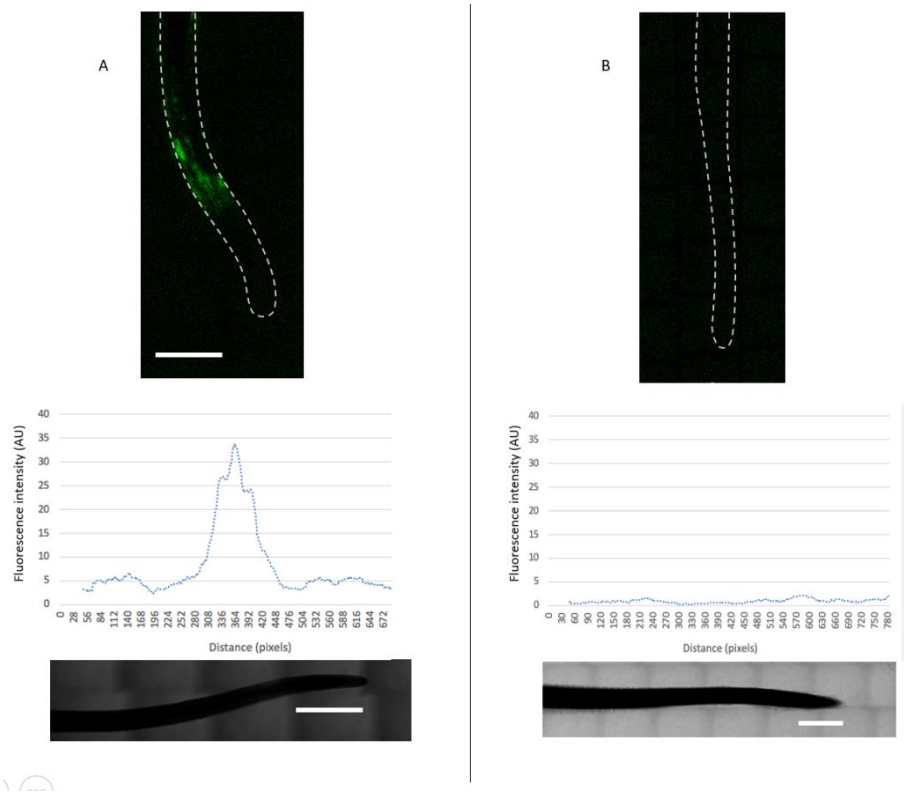


Figure 6-8. Motile and non-motile Rlv3841 interaction dynamics with vetch roots in ChIIPS. A – upper panel: GFP signal from *V. villosa* root (outlined in dashed white) imaging with OPS1734 inoculum at 7.5 hrs post inoculation. Lower panel: GFP fluorescence intensity (in arbitrary units, AU) along the edge of the root (shown below the x axis for reference) from mature zone to growing tip (see also Appendix 2 Video 3). B – as for A, but with *V. villosa* and non-motile OPS1736 (Appendix 2 Video 5).

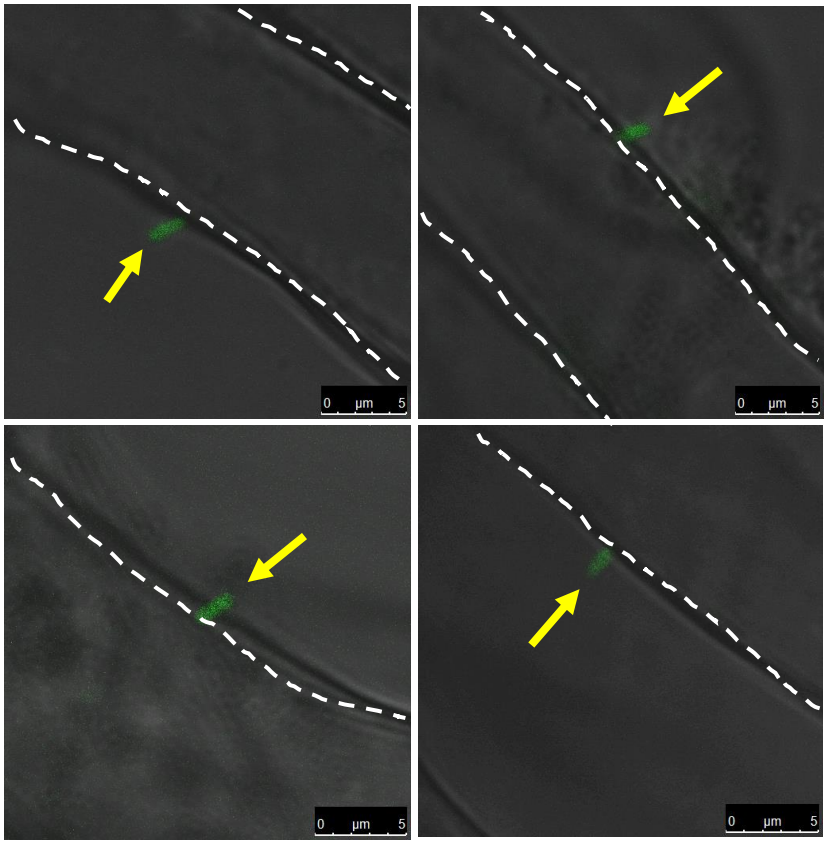
Whereas the motile strain showed a clear REZ interaction profile with a vetch root, the motility mutant strain showed no such profile at any time point. (Figure 6-8). This strongly indicates that bacterial chemotaxis drives the development of the characteristic *Rhizobium*-REZ interaction profile demonstrated in this work.

6.2.8 Using ChIIPS to investigate root hair attachment polarity

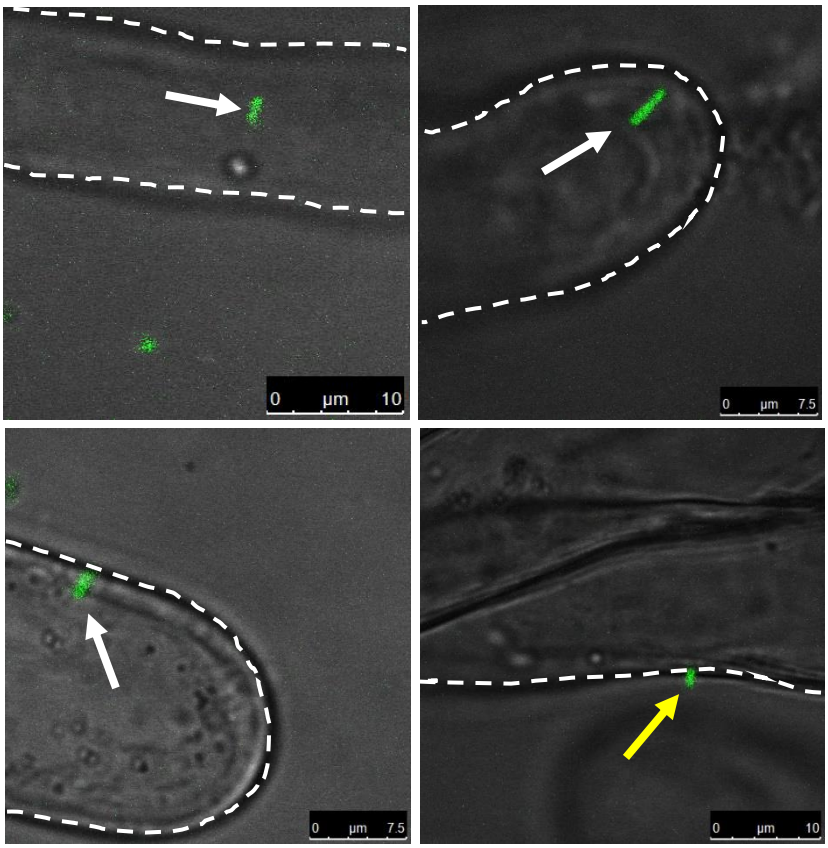
The existing literature model of Rlv3841 attachment to legume root hairs indicates that, under acidic to neutral conditions, polar glucomannan on the bacterial cell surface binds root hair lectin. Under alkaline conditions these lectins disassociate, and Rlv3841 uses rhicadhesin to bind root hairs [34, 71]. This work has investigated rhicadhesin extensively (see section 4.2.13). Despite uncertainty around the identity of factor(s) promoting alkaline pH root hair attachment, basic physiology of this process (such as whether attachment to root hairs is also polar, as with glucomannan) remains uncharacterised.

To investigate this, ChIIPS was used to image the attachment of both Rlv3841 and a *gmsA* (glucomannan) mutant strain to vetch root hairs at high magnification (see 2.8.7; here a 64x objective lens was used). Representative images and graphed percentage attachment classification (polar or non-polar) for Rlv3841 and the glucomannan mutant strain are shown in Figure 6-9.

A



B



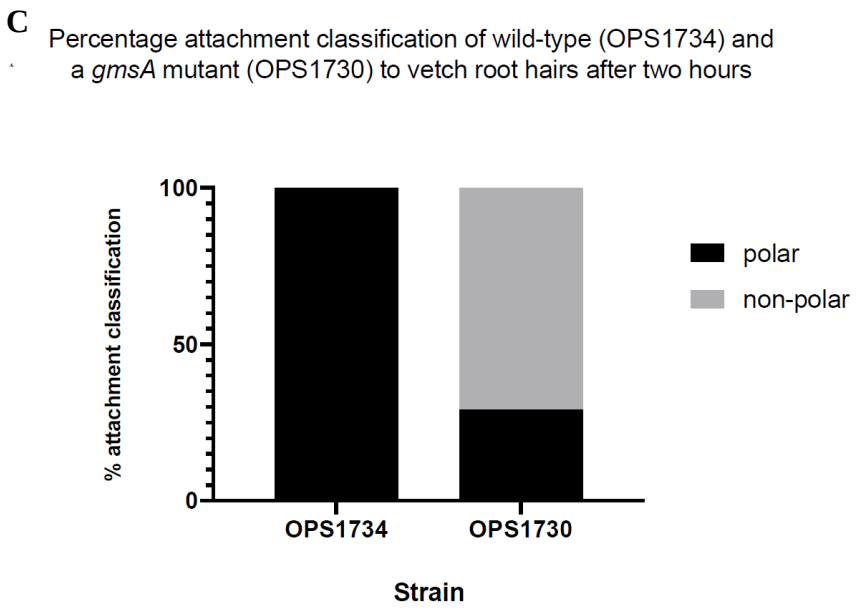


Figure 6-9. Attachment polarity of OPS1734 (Rlv3841[pLMB449] – pTac *gfp*) and OPS1730 (A1045- *gmsA*::Tn*5*ΩKan/Neo[pLMB449] – pTac *gfp*) to *V. villosa* root hairs after two hours. A – Representative images of OPS1734 attached to root hairs. Root hairs are outlined in dashed white and attached bacteria are indicated with arrow. Yellow arrow colour indicates polar attachment; white arrow colour indicates non-polar attachment. B – as for A, but with OPS1730. C – Percentage attachment classification of both strains (lateral vs polar) for all counted bacteria. n = 35 for OPS1734 and 42 for OPS1730.

These results demonstrate that, under neutral conditions, all Rlv3841 cells observed with functioning glucomannan attached to root hairs in an end-on, polar fashion. However, in a glucomannan mutant strain, this figure was ~ 30%, with the remainder of observed cells being attached in a non-polar (i.e. lateral) fashion. This indicates that, whilst glucomannan mediated attachment is polar, when Rlv3841 cells lack glucomannan they can attach to root hairs in a polar or non-polar fashion. There are two possible conclusions that can be drawn from this: rhicadhesin is not exclusively located at cell poles and can therefore mediate root hair attachment in a polar or non-polar fashion, or that there are multiple,

differentially localized cell surface components which govern attachment in the absence of glucomannan. This is a possibility given the diversity of attachment factors required for Rlv3841 attachment to pea roots under alkaline conditions (when glucomannan will not be functioning) seen in this work (see chapter 4 and 4.2.11).

These findings are comparable to those described by Matthysse (2014), reviewing work with *A. tumefaciens*. In this bacteria, unipolar polysaccharide (UPP, a glucomannan analogue) mediates polar binding of bacteria to root hairs under low calcium, low phosphate or acidic pH conditions. When UPP is not produced, attachment is mediated by unknown molecules and is both polar and lateral [55]. It was noted that, in UPP-independent attachment, the total number of bacterial cells attaching was also lower than with UPP-producing bacteria [55]. This fits with the results of a Lux assay and INSeq for a glucomannan mutant, where attachment is reduced in comparison to Rlv3841 under neutral pH conditions (see Figure 5-3, Table 4-6).

Therefore, although factors governing glucomannan or UPP-independent root hair attachment remain to be identified, it seems both Rlv3841 and *A. tumefaciens* can attach in a polar or lateral fashion in the absence of these factors.

6.2.9 ChIIPS2 design for future work

The ChIIPS chamber used in this work proved suitable for imaging early-stage *Rhizobium*-legume root interaction dynamics. However, for imaging over longer timescales and for more realistic test conditions, it would be desirable for the aerial plant parts to remain external to the chamber. A schematic for a new chamber design (ChIIPS2) which would overcome the limitations of ChIIPS whilst maintaining its advantages is shown in Figure 6-10.

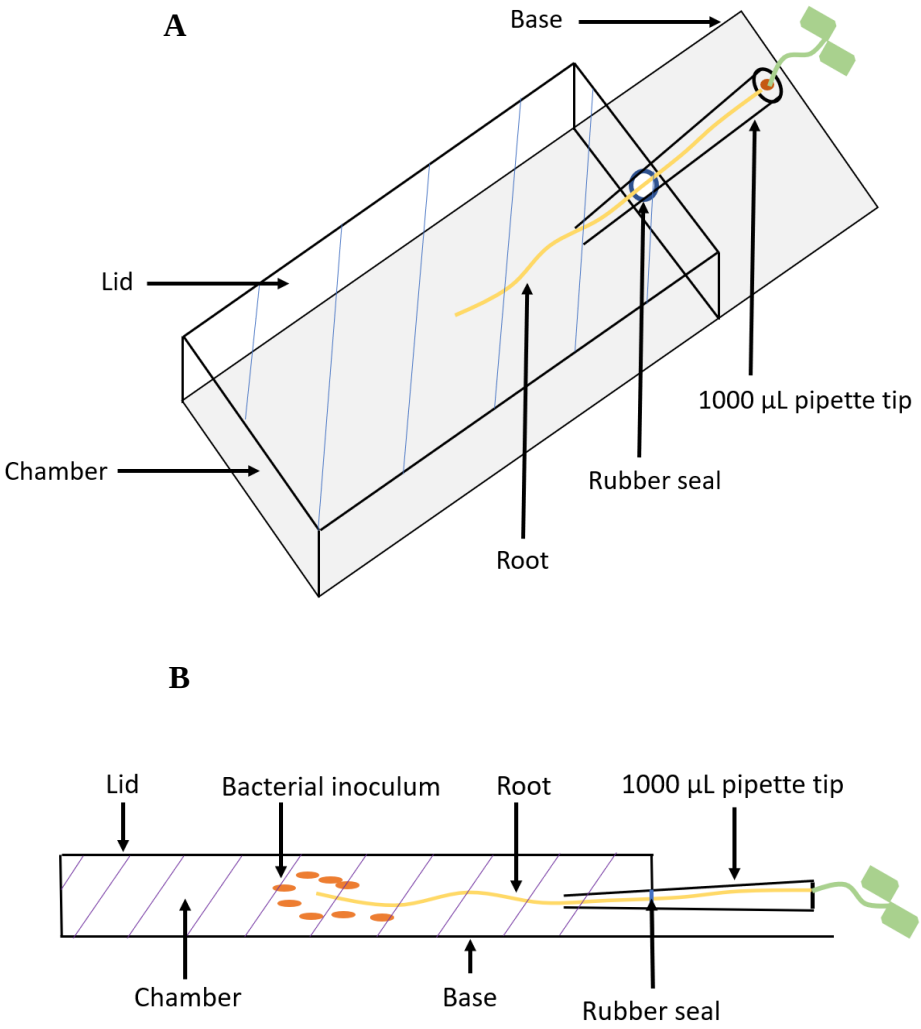


Figure 6-10. ChIIPS2 design. A – Schematic of the proposed ChIIPS2 shown from above. Plants are germinated upright in 1000 µL pipette tips to promote gravitropic root growth. These tips are transferred to the ChIIPS2 chamber (before the root emerges from the tip) through a hole in the chamber wall and held in place by the rubber seal. Roots continue to grow into the chamber and can then be imaged. B – longitudinal section of the proposed ChIIPS2 showing the plant root inside the chamber with bacterial inoculum (brown ovals, not to scale).

6.3 Conclusion

Here, the applicability of microfluidics systems to the study of root interaction dynamics and spatio-temporal mapping of interactions between bacteria and roots (the first step in *Rhizobium*-legume symbioses) is demonstrated.

For mapping the spatio-temporal activation of the *lppE* gene (*RL3234*) on roots, clover plants (*T. repens* and *T. pratense*) were used with Rlv3841 reporter strains. Having validated the specific activation of *lppE* (*RL3234*) gene expression in the REZ using a luminescence method, spatiotemporal gene expression visualization was also tested in the TRIS setup. Strong activation of the *lppE* promoter was seen in the REZ from 1 hr onward, and this peaked at ~39 hrs post inoculation. Although the role of *lppE* in root interaction remains unknown, its expression is clearly related to REZ signaling molecules. Particularly in secondary root attachment, different factors are expressed and upregulated in different rhizobia to facilitate the colonization process. These include extracellular fibrils, cellulose, cadherin-like proteins, fimbrae and outer membrane proteins, among others [34]. However, very little is known about either the precise timings of gene expression in root interactions, or whether genes show specific spatial requirements for attachment to different root zones. As the data for *lppE* shows, a specific spatial expression does exist. By using microfluidics setups with a suite of reporter genes, the process of primary and secondary attachment could be more specifically defined.

The REZ is a zone of interest for root-microbe interactions. This is not just because of its high exudation and identity as a preferential microbe interaction site in several systems, but also because of the role it may play in primary attachment. In a soil system with large microbial load, it seems likely that the REZ, emerging directly behind the meristematic

growing tip, will represent one of the only uncolonized root zones. The high level of exudation from this site may serve to rapidly recruit new root microbiota preferentially and act as a method to protect against pathogen colonization. If this hypothesis is correct, the elongation zone presents an interesting area of focus for characterising root-microbe interaction profiles, and spatio-temporal gene expression patterns of attaching bacteria in this region could provide novel mechanistic insights.

Given the limitations of TRIS (where narrow channels with a short imaging window prevent imaging of roots much larger than Arabidopsis, and chamber design precludes use with an upright confocal microscope), a preliminary new test setup called ChIIPS was trialed. Based on commercially available cell culturing slides, ChIIPS is lower cost than TRIS, compatible with much larger root systems and easier to use with upright confocal microscopes. Despite the limitations of this preliminary system, preferential interaction of Rlv3841 with the REZs of lotus, vetch and alfalfa were clearly visible. This result is important in confirming the REZ as a hotspot for interaction across multiple plant species. A motility mutant of Rlv3841 did not show these preferential interaction dynamics with vetch, confirming that chemotactic motility is required on the part of interacting bacteria. This result is interesting when compared with data from Lux root attachment assays (Chapter 3) and INSeq experiments (Chapters 4 and 5). In a Lux root attachment assay, loss of motility lead to a large reduction in attachment to pea roots at all pHs in a *motA* mutant strain (Figure 3-6). However, INSeq classified *motA* (and other motility genes) as NE (not required for root attachment) in attachment to pea roots at all pHs (see 4.2.9) and to barley roots at pH 7.0, although chemotaxis was needed for attachment to soybean roots (see 5.2.10). The hypotheses explaining these results are as follows: in a homogenous inoculum of non-motile cells (as in a Lux attachment assay with a *motA* mutant), bacteria

cannot move toward the root surface for attachment. However, in a mixed INSeq inoculum (where most of the cells will be motile), 'aided motility' occurs as cell-cell collisions between chemotaxing and non-motile cells (as well as disruption of the repulsive hydrodynamic boundary layer) leads to positioning of non-motile cells in proximity to the root which can permit primary attachment. For soybean, the large root size and higher exudation could result in a stronger selection for root attachment on chemo-sensing bacteria, resulting in under-representation of chemotaxis mutants in the soybean root-attached library and classifying chemotaxis genes as ES/DE. A ChIIPS system could be used to investigate these hypotheses further, especially with regards to 'aided motility'. A simple experiment using single and mixed (differentially labelled) wild-type and non-motile Rlv3841 inoculums with different roots could quantify how many non-motile cells migrate to within close proximity of the root over time. This would enable quantification of any 'aided motility' effect and could provide additional evidence to explain disparities between Lux attachment assay and INSeq results.

As a final test of ChIIPS, high magnification confocal microscopy was used to investigate how Rlv3841 and a glucomannan mutant attach to root hairs at neutral pH. Interestingly, whereas Rlv3841 showed only polar, end-on root hair attachment, a glucomannan mutant showed a mixture of lateral (~65 %) and polar (~ 35 %) attachment. This mirrors the characterised attachment profiles of UPP proficient and deficient strains of *Agrobacterium*, respectively. This suggests that, whilst the identity of any 'rhicadhesin(s)' mediating UPP or glucomannan independent attachment remains unknown, there is likely to be mechanistic similarities between the two species. If there is a single rhicadhesin, it is not only polarly located but also permits lateral cell attachment to root hairs.

To conclude this chapter, an improved ChIIPS2 design was suggested, which would overcome ChIIPS limitations. The main new feature of this design is the provision for aerial plant parts to remain separate from the bacterial inoculum. This is likely to enable longer term imaging of root-microbe interactions within a less physiologically disruptive environment for the plant.

The use of ChIIPS2 would enable many experimental questions to be addressed. Alongside spatiotemporal gene expression patterns, interactions between *Rhizobium*, roots and pathogens could be examined. Also of interest would be to use a dual channel setup (not dissimilar to that seen in TRIS, [188]), whereby two different roots can be imaged side by side. This would enable inoculum preferences to be examined; for example, is root preference seen in a differentially labelled inoculum of *R. leguminosarum* biovar *trifolii* and Rlv3841 interacting with vetch and clover roots? Does one strain outcompete another for root interaction when its cognate symbiont root is present? Such questions remain unexplored.

One limitation of all systems mentioned in this work is the long-term immersion of roots in liquid, which is not particularly representative of soil conditions. An exciting prospect is the combination of ChIIPS2 with transparent soil. This is a form of highly treated fluoropolymer which has very similar water and nutrient retention characteristics to real soil, and is suitable for sustaining plant growth and imaging root-microbe interactions [463]. However, for the moment this transparent soil remains expensive and technically challenging to produce, particularly in large quantities.

This work highlights the huge potential of novel imaging platforms in the study of *Rhizobium*-legume root microbe interactions. Further work in this area is likely to yield novel insights into these inter-kingdom relationships.

Chapter 7

General discussion

7.1 Overview

In this work a range of techniques, including proteomic and bioinformatic analysis, newly developed Lux whole-root attachment assays and INSeq were used to define the primary root attachment determinants needed for Rlv3841 interaction with pea at different pHs and with those of a non-host legume and non-legume at pH 7.0. These approaches revealed a far greater diversity of primary attachment mechanisms than previously recognized for Rlv3841, which demonstrate considerable pH and plant-host dependency. A summary of the most important findings regarding attachment factors from all the experimental chapters of this work is presented in Figure 7-1. These findings are discussed and evaluated in the subsequent sections of this chapter, starting with extracellular/surface-localized factors and moving to intracellular and uncharacterised factors. A critique of the experimental techniques employed is followed by an outline of further research directions.

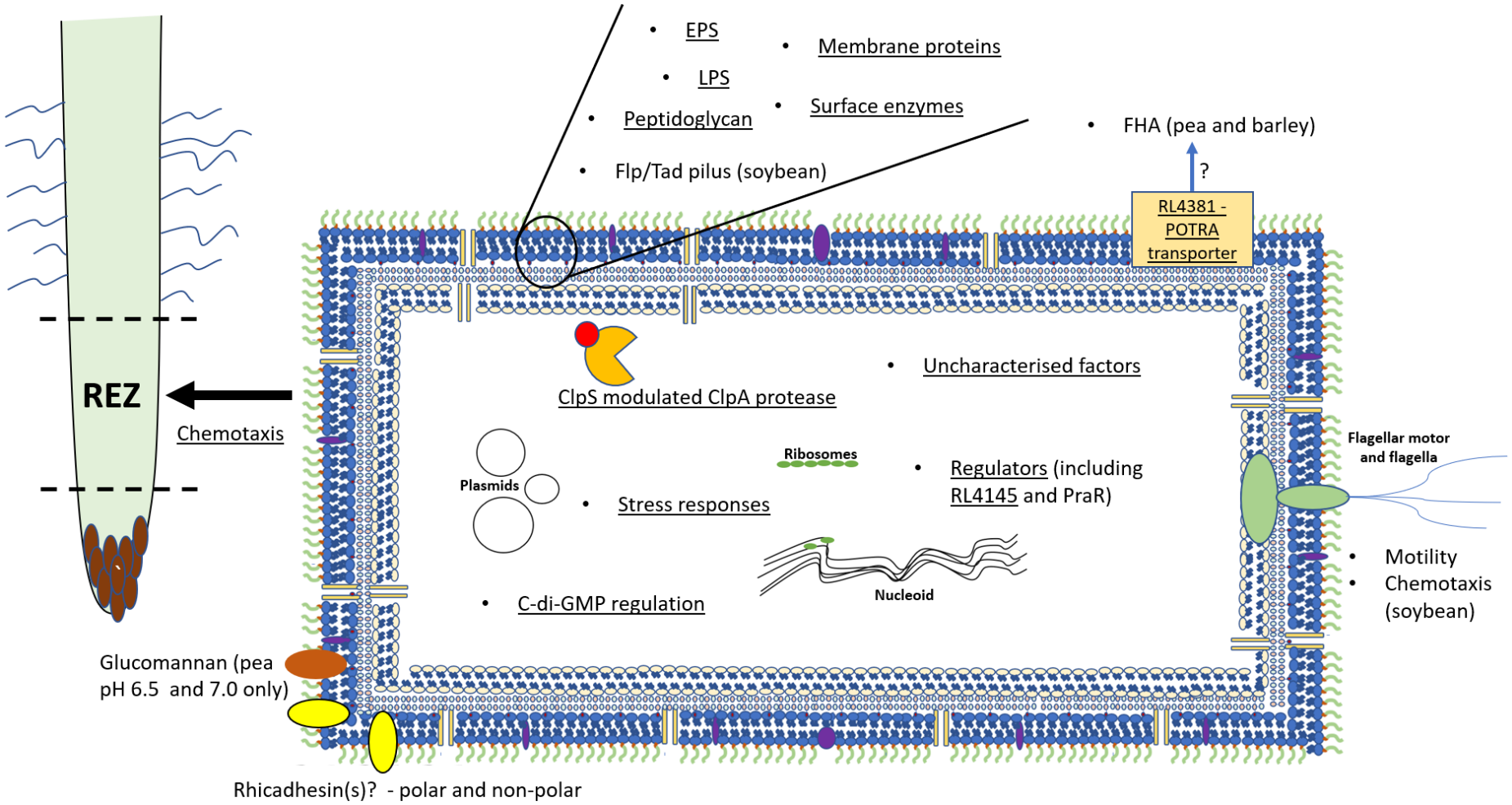


Figure 7-1. Summary of the main findings of this work regarding interaction and primary root attachment of Rlv3841. A stylized gram-negative Rlv3841 cell and a plant root are shown (not to scale) with the major factors required for interaction with all plants underlined. For other factors, attachment requirements are given in parenthesis. These factors are discussed in greater detail throughout in the remaining sections of Chapter 7. REZ = root elongation zone.

7.2 Extracellular/surface localized primary attachment factor requirements

7.2.1 EPS and peptidoglycan

Chapters 4 and 5 demonstrated that Rlv3841 primary root attachment at different pHs and to different plant roots is vastly more complex than previously reported. Whilst there are multiple follow up points to this work, several target genes stand out for further investigation. In terms of cell surface and exported factors, gene classifications from INSeq suggested that activation of the Npr/ManX system is needed for pea root attachment at pH 6.5, but not at pH 7.5, where they are likely deactivated by SixA. The most likely effect of this at pH 7.5 is a reduced EPS secretion through an unknown ABC exporter [313]. However, this could be confirmed by quantification of EPS in Rlv3841 cell cultures at pH 6.5 and 7.5. Methods such as liquid chromatography-organic carbon detection–organic nitrogen detection and colorimetric assays of protein, sugar and phenolic content (among others) [464, 465] can also be applied and demonstrate at a molecular level the EPS changes needed for pea root attachment at different pHs. A large diversity of different factors involved in EPS biosynthesis and modification was shown to be required for attachment to pea roots at different pHs and to different plants roots using both Lux and INSeq approaches. Therefore, the EPS category encompasses crucial cell surface alterations (Figure 7-1) promoting attachment under different conditions. Similarly, INSeq experiments highlighted a ubiquitous requirement for peptidoglycan biosynthesis and modification factors for primary attachment under all plant attachment conditions tested. Of particular note was the *dacC* (*RL4363*) gene, encoding a putative penicillin binding protein, required for attachment to all plant roots and at all pH

conditions. In *E. coli*, DacC processes sugar-peptide cell wall precursors and is involved in peptidoglycan biosynthesis [299]. What is clear is that peptidoglycan composition is another crucial factor promoting attachment under different conditions. Ramstedt *et al*. (2011) have described a method of cryo-x-ray photoelectron spectroscopy analysis combined with a curve resolution analysis of the carbon spectra. This technique is able to closely monitor the chemical composition of bacterial cell walls and predict changes in peptidoglycan structures [466]. Such a technique could be useful in characterising the peptidoglycan changes required for root attachment under different conditions and with different plant hosts in Rlv3841.

Not discussed so far in this work is the role of cell shape, which can play a crucial role in regulating bacterial attachment [467]. For example, *C. crescentus* has a crescent cell shape determined by the production of a protein called crescentin [468]. Being an organism largely residing in freshwater lakes and streams, this crescent shape specifically enhances surface colonization under fluid-flow conditions. This is due to favourable cellular orientation caused by shape relative to the target surface under these conditions [468]. Peptidoglycan is the main stress-bearing structure that dictates bacterial cell shape [467]. It would be interesting (using microscopy) to investigate the cellular morphology in Rlv3841 peptidoglycan mutants shown to be deficient in attachment under certain conditions to determine if any common altered morphologies are observed. These would likely disrupt attachment by interfering with bacterial cell-surface factor interactions with plant roots.

7.2.2 Surface enzymes

Several enzymes (such as TpiA, a glycolytic enzyme with a role in central carbon metabolism) were hypothesised to play an important bacterial surface-associated moonlighting role in primary attachment to all plant roots based on INSeq results. TpiA was very interesting, being required for attachment to all plant roots at all pHs tested (Figure 7-1, included in 'surface enzymes'). If correct, this would represent a further novel primary attachment mechanism for Rlv3841. Surface localised glycolytic enzymes are multifaceted and can be involved in substrate binding. TpiA has been shown to be surface localised and have a direct role in adherence to host cells in *Mycoplasma gallisepticum* [284]. An Rlv3841 mutant in TpiA carrying pIJ11282 (luminescence cassette) could be used to validate this in a Lux attachment assay, whilst fluorescent protein tagging could investigate the hypothesised secretion of TpiA.

7.2.3 Flp/Tad pili, outer membrane proteins and LPS

Further to the novel attachment factor requirements described above, INSeq robustly demonstrated that surface localised primary attachment factors previously identified in other bacterial species can be used by Rlv3841 in a context dependent manner. Primary attachment to soybean roots showed a Flp/Tad pilus requirement (Figure 7-1) which was demonstrated by the need for encoding genes found on plasmids pRL11 and pRL8. This shows an overlap whereby Rlv3841 can make use of attachment mechanisms (namely adhesive Flp/Tad pili) reported in *Agrobacterium*, *Pectobacterium* (a potato pathogen), *V. vulnificus*, *A. actinomycetemcomi* and *C. crescentus*, among others [44, 384, 385].
A second example of this is the requirement of the OmpA protein RL3267 for attachment to barley and soybean roots (included under 'outer membrane proteins, Figure 7-1). Such

outer membrane proteins acting as primary attachment factors are well characterised in *Azospirillum*, where they function both in root attachment and cellular aggregation [64]. The third example is the requirement for various LPS biosynthesis and modification factors for attachment to different plant roots under different conditions (Figure 7-1). As an example, *dgkA* (*RL2780*) was needed for attachment to pea roots at pH 6.5 and soybean roots at pH 7.0. In *E.coli, dgkA* mutants are defective in biofilm formation [316], and DgkA function has been linked to phospholipid recycling and LPS modifications [324]. In *B. subtillis* it is important for lipoteichoic acid synthesis [325]. However, LPS has not previously been implicated in primary attachment in Rlv3841, instead being more associated with *Azospirillum*. Indeed, it has been proposed that the LPS O-antigen of *Azospirillum* directly binds maize root lectin to mediate primary attachment [65].

A powerful conclusion of this work overall is therefore that Rlv3841 can make use of primary attachment mechanisms which, until now, have only been associated with other bacterial species. It does so in an environmental condition and plant host-dependent manner. Given this finding, it seems possible that primary attachment mechanisms are more conserved across soil bacteria than realised but demonstrate greater context dependency than previous experimental work has been able to investigate. Characterising this further would require inverse experiments to be conducted, for example using INSeq to examine the primary attachment mechanisms of *Azospirillum* to different plant roots under different conditions.

7.2.4 FHA

RL4382 (filamentous hemagglutinin adhesin - FHA) was important for pea root attachment at all pHs and to barley roots (Figure 7-1). This was also demonstrated in a

Lux assay in Chapter 3. Previous work in other bacteria suggests that FHA plays a direct attachment role and agglutinates cells, leading to higher numbers of cells attached [239–241]. This has not previously been documented in *Rhizobium*, making this a novel attachment factor. Various scanning-electron microscopy techniques are available for high-resolution imaging of biofilm structures [469], and applying such techniques to wild-type Rlv3841 and an *RL4382* mutant could shed more light on the mechanisms behind FHA mediated attachment. *RL4381* (putative POTRA domain exporter of FHA) was needed for attachment to all plant roots whilst FHA was not (Figure 7-1), suggesting this exporter my interact with other attachment factors. Application of TRANSIT analysis (a method combining multiple different statistical approaches to analysing INSeq data into one platform [470], see also 7.9.2) could reinforce or challenge this idea by enabling comparison of gene classifications with different methods of data analysis. If reinforced, a bioinformatic screen of other Rlv3841 genes encoding TPS domains (required for interaction with POTRA domain exporters, [239]) could indicate the identity of other exported factors. Further, RL4381 should be confirmed as the transporter for FHA. One approach to this would be to express *RL4382* fused to a fluorescent protein in both wild-type and *RL4381* mutant Rlv3841 strains. Changes in fluorophore location (intracellular/extracellular) would show whether the transporter encoded by *RL4381* is necessary and sufficient for FHA export.

7.2.5 Motility

In a Lux attachment assay, *flgE* and *motA* mutant strains were defective in attachment to pea roots at pHs 6.5, 7.0 and 7.5 (see Figure 7-1, motility). These mutant strains showed no significant differences to one another in terms of attachment level. It was therefore

hypothesised that it was motility itself (rather than flagellar adhesin properties) contributing to attachment in Rlv3841/pea symbiosis. This suggests that flagellar adhesin action is not contributing directly to attachment, as is seen in some other rhizobacteria [34]. No flagellar subunits were classified as required (ES/DE) in pea root attachment at any pHs in INSeq (Chapter 4). However, the disagreement (possibly caused by 'aided motility', 4.2.9) between Lux and INSeq means that a role for flagellar subunits in adhesion cannot be completely ruled out (indeed *flaH*, a flagellar subunit, was needed for soybean root attachment, see Table 5-9). The bacterial flagellum is made up of more than 20 different proteins, of which FliC is the major subunit. FliC has been reported to have adhesin properties in *E. coli* and *P. aeruginosa* [235]. To investigate possible flagellar adhesin properties in Rlv3841, further single flagellar subunit mutants could be tested in Lux attachment assays and compared to a *motA* mutant strain. Mutants showing reduced attachment versus the *motA* mutant would indicate both loss of motility and loss of flagellar adhesin function.

Motility was not required for Rlv3841 attachment to pea or barley roots in INSeq, but aspects of chemotaxis were needed for soybean root attachment (Figure 7-1). The reasons for this (and the disagreement with Lux attachment assay data showing motility as required for effective pea root attachment) remain unclear. A form of 'aided motility' in INSeq (biased Brownian motion as well as disruption of the repulsive hydrodynamic boundary layer around roots, see 4.2.9), along with aberrant swimming of chemotaxis mutants and higher levels of exudation from soybean roots (providing a stronger selection for chemotaxis-competent bacteria than for other plant roots), was hypothesized to explain these results. Exudation levels from soybean, pea and barley roots could potentially be examined using Rlv3841 luminescence biosensor strains, such as those developed by Pini

et al. (2017), with levels of light emission acting as a proxy for exudate levels [201]. Further, a mathematical modelling exercise of bacterial motion in mixed population of chemotaxing/non-chemotaxing cells could examine the plausibility of the 'aided motility' hypothesis. Using differentially labelled motile and non-motile Rlv3841 strains in a ChIIPS setup could also help resolve this question, with confocal imaging able to determine how many non-motile cells were able to reach the plant root surface. Inoculating a motile wild-type culture of Rlv3841 labelled with GFP with lotus, vetch or alfalfa roots showed preferential bacterial interaction with the REZ (Figure 7-1), indicating that the REZ exudes high levels of chemoattractants. This result was also demonstrated for *B. subtilis* interacting with Arabidopsis roots in TRIS [188]. This interaction profile was not seen with a *motA* mutant of Rlv3841. Whilst this further suggests that motility is important for establishing early-stage root microbe interaction dynamics, this further highlights the need to clarify INSeq results and determine cell movement dynamics where non-motile strains are a minority cell type in a mixed population interacting with roots.

7.2.6 Rhicadhesin

Rhicadhesin is a proposed 14 kDa protein which facilitates attachment of *Rhizobium* to legume root hairs at alkaline pHs [34], but has never been identified at the gene level. A proteomics approach was taken to identifying rhicadhesin. Using a 14 kDa protein band (thought to contain rhicadhesin) for LC-MS/MS identified 15 proteins, of which three (*RL4733, omp19 – RL4441* and *RL1635*) were the most likely rhicadhesin candidates. Due to gene size limitations, pK19mob mutagenesis was unsuccessful in targeting rhicadhesin candidate genes identified. An in-frame mutagenesis approach could be used to overcome

these issues, as very small genes (such as the *hdeA* gene in *E. coli*, 110 bp) have been mutated by this approach in the past [246].

The preparation of a crude adhesin fraction from Rlv3841 cells relied on isolating cell surface proteins, as described by Smit *et al* (1989) [53]. However, here the 14kDa band from the crude adhesin fraction also contained ribosomal subunit and other predicted intracellular proteins at low levels, suggesting that some minor cell lysis had occurred. Siciliano *et al.* (2019) describe several innovative proteomics approaches to investigating bacterial surface-exposed proteins (the proteosurfaceome) which could be more appropriate to the study of cell surface adhesins. These include cell shaving, in which surface exposed protein fragments are released by proteolytic digestion using trypsin with intact bacterial cells. This approach allows comprehensive analysis of the proteosurfaceome as membrane and cell wall embedded proteins can be released, further digested with trypsin and subjected to LC-MS/MS. This approach has been applied to both Gram-positive and negative bacteria [221]. However, tryptic digest is a limitation of this technique as is it dependent on trypsin binding sites being accessible in target proteins to generate peptides for mass spectrometry. Further, functional proteins will not be recovered. A more advanced approach uses biotin tag-containing reagents to target surface proteins before lysing cells and isolating surface proteins using streptavidin affinity chromatography. This method is similarly suitable for both Gram-positive and negative bacteria [221]. Combining these methods could allow the identification at the gene level of ~14kDa Rlv3841 surface proteins as well as the recovery of intact surface proteins for functional assays.

Regarding rhicadhesin, the original research defining its role (Smit et al. (1989), [53]) was certainly not conclusive. Defining an adhesin based on its ability to inhibit attachment is

not definitive. Further, given the number of root attachment factors identified bioinformatically in Chapter 3, it is likely that membrane preparations may isolate proteins that can inhibit attachment when pre-incubated with roots, but this does not imply these proteins are important adhesins. Multiple genes encoding rhicadhesin or use of an impure protein preparation when defining its activity also confound definitive identification further. When combined with INSeq results from Chapter 4 (where multiple 12-16 kDa factors were important for root attachment at neutral and/or alkaline pHs) it is apparent that there are multiple attachment factors with rhicadhesin-like properties. Further, work in Chapter 6 demonstrated that a glucomannan mutant of Rlv3841 can attach to vetch roots in both a polar and non-polar fashion (whilst glucomannan is required for attachment at acidic to neutral pHs in a polar fashion, demonstrated with Lux assays, INSeq and ChIIPS, see also Figure 7-1). Therefore, whilst the definitive identification of a rhicadhesin as originally proposed by Smit et al. (1989) [53] was not possible, this work demonstrates there are multiple small proteins with rhicadhesin-like profiles of attachment activity. Hence, in Figure 7-1, rhicadhesin(s) are shown on both the polar and lateral surfaces of the Rlv3841 cell.

7.3 Intracellular primary attachment factors

7.3.1 Regulators - PraR

Intracellular factors important for root attachment include a variety of regulators required at different pHs and with different roots (Chapters 3, 4 and 5). A mutant in the important *praR* regulator demonstrated a previously uncharacterised pH dependency in Lux assays. Attachment phenotypes for *praR* and *rapA2*/*rapC* mutants were as previously described at

pH 7.0 [59] in a Lux assay with pea roots. However, these strains showed very different attachment profiles at pH 6.5 and 7.5, with the *praR* mutant defective in attachment and the *rapA2*/*rapC* mutant a hyper-attaching strain (hence the PraR regulator is shown in Figure 7-1). Testing a *praR*/*rapA2*/*rapC* triple mutant in a Lux assay (defective in attachment at all pHs) allowed an updated model of regulation to be proposed. At pH 7.0, CinS inhibits PraR which relieves inhibition on *rapA2* and *rapC*, permitting increased attachment. However, at pH 6.5 and 7.5 PraR is not inhibited, and induces expression of unknown '*rapX*' factor(s), possible given the positive and negative regulatory activity of PraR [59]. Whilst this model has not been validated, mutation of *praR* leads to hyper-attachment at pH 7.0 and reduced attachment at pH 6.5 and 7.5. RNASeq has previously been used to investigate the transcriptome profile of a *praR* mutant [59], and expanding this work to pH 6.5 and 7.5 conditions could refine the model of PraR/Rap interactions and primary root attachment at different pHs.

7.3.2 c-di-GMP regulation and regulators – *RL4145*

Various c-di-GMP synthesis proteins were required for root attachment at different pHs and to different plant roots (Figure 7-1, 'c-di-GMP regulation'). Whist c-di-GMP signalling is well characterised, particularly in biofilm formation [78, 79], how the different c-di-GMP synthesis proteins in Rlv3841 are contributing to attachment under different conditions and to different roots is unknown, and warrants further investigation. c-di-GMP regulated biofilm determinants range from flagella rotation to EPS and surface adhesin production and type IV pili retraction [81]. Reconciling the global effect of cellular c-di-GMP concentration on biofilm formation with the discrete actions of c-di-GMP is difficult, but various hypotheses exist. One is that diguanylate cyclases (DGCs)

and phosphodiesterases (PDEs) show differential expression and enzyme activity and so have discrete impacts on the cellular pool of c-di-GMP. Another is that different effectors show different c-di-GMP binding affinities and are differentially responsive to intracellular c-di-GMP levels. One of the most intriguing theories is that different c-di-GMP pools are sequestered in multi-protein complexes at distinct cellular sites [79]. This has been observed for the LapA adhesin of *P. fluorescens*, where LapA activity is regulated post-translationally by the LapD/LapG c-di-GMP effector system. This system activates LapA in the outer membrane on receipt of a specifically localized c-di-GMP signal. This signal in turn is likely generated by activation of DGCs co-localized with LapD/LapG [471].

What is clear is that understanding c-di-GMP effects is not as simple as measuring cellular c-di-GMP content. Tightly regulated signalling systems, localized protein complexes and distinct subcellular changes may all be involved in producing different outcomes in terms of primary root attachment.

One regulator (*RL4145*) was required for attachment to all plants (Figure 7-1), whilst many were required for attachment to pea at different pHs. In the case of the protein encoded by *RL4145*, a likely regulatory target is *RL4139* (a putative c-di-GMP degrading EAL domain protein). If it is the case that RL4145 represses RL4139 (as was suggested by their inverse HMM classifications), then the requirement for RL4145 in attachment to all plants can be explained by the maintenance of high intracellular c-di-GMP levels by this protein. Investigating *RL4145* regulation further could be approached using RNASeq to infer *RL4145* regulatory targets as well as protein-protein interaction studies or promoter binding assays to see if there is direct activity between *RL4145* and *RL4139*. An RNASeq

approach would also be well suited to investigating transcriptional regulatory networks underpinning attachment in mutant Rlv3841 strains of other important regulators.

7.3.3 ClpS-modulated ClpA protease

INSeq demonstrated a requirement for the ClpS/ClpA protease system for attachment to pea roots at pH 7.0 and 7.5, as well as to soybean and barley roots (Figure 7-1). The targets of ClpA (with or without modulation by ClpS) are unclear and understanding these would aid in defining the cellular changes induced by this system. Multiple experimental approaches could be taken to characterising ClpS/ClpA targets. These include experimental methods such as gel separation of control and ClpS/ClpA proteolyzed bacterial protein preparations, where proteolytic targets are identified by gel band shifts between the two samples. Further, radiolabelling approaches such as SILAC in wild-type and ClpS/ClpA mutant Rlv3841 strain, followed by MS-based peptide quantitation, could also be applied [472]. There are also multiple bioinformatic approaches which could be applied to predicting protease targets [473].

7.4 Uncharacterised primary attachment factors

Amongst the ‘uncharacterised factors’ is *pRL100053* encoding a putative transmembrane protein of unknown function (Figure 7-1), which is defective in attachment in a Lux assay at all pHs and required for primary attachment (all pHs), colonization, infection thread and nodulation in INSeq experiments (Table 4-9). This data strongly suggests that *pRL100053* encodes a novel root hair attachment factor which is critical for subsequent nitrogen fixing symbiosis development.

For the 292 genes required for primary attachment to pea roots (INSeq, Chapter 4), multiple datasets from different stages of Rlv3841/pea symbiosis were used to map gene requirements and infer which primary attachment factors could be involved in attachment to root hairs (the entry-point for nitrogen-fixing symbioses [25]) or bulk root epidermis. This approach uses the rationale that factors required for root hair attachment will also be required in subsequent symbiosis stages. However, INSeq data alone does not directly demonstrate this, and the picture is complicated by factors which are classified as required in non-contiguous symbiotic stages. A potentially high-impact way of resolving this further would be to follow up on INSeq results with in-depth studies of gene function, and *pRL100053* represents a good candidate for this.

Expressing a vector-borne *pRL100053* construct whereby *pRL100053* is fused to a fluorescent protein in Rlv3841 could visualize protein localization. Using biotin-tagged *pRL100053* as a bait protein for pea root protein extracts could identify its plant root binding partners, and purification studies could also define protein structure.

7.5 Experimental techniques and future research directions

7.5.1 Lux whole-root attachment assay

The Lux whole-root attachment assay developed here allowed the primary attachment abilities of Rlv3841 strains to be tested at different pHs. This assay represents a considerable advance on previously described techniques which relied on high magnification microscopy and counting bacteria [52, 53, 210, 217–219], confocal microscopy [56], bacterial recovery and plating or radiolabeling approaches [59]. In comparison, Lux is higher throughput, preserves spatial attachment information, is safer

than radiolabelling and is more appropriately established for studies at different pHs. The activation of the *lppE* promoter selectively in the REZ along with the REZ attraction profile of bacteria (Chapter 6) suggests spatio-temporal interaction dynamics between Rlv3841 and legume roots, which may extend to attachment. Although not pursued here, Lux assays at different time points could evaluate the spatio-temporal attachment profile of different Rlv3841 strains to roots. This would increase our understanding not just of attachment factors, but also root zone importance in the attachment and colonization process. However, it is also worth noting that interpretation of Lux data over longer time periods may become more complex than in a one-hour attachment assay as bacterial replication and metabolic activity may affect Lux signal intensity [201].

7.5.2 INSeq

Chapters 4 and 5 used INSeq with HMM analysis to classify gene requirements in primary root attachment. Although there are various advantages to using an HMM for analysis over other published methods (see 1.8.3), low TA site number in a gene can still be a confounding factor. DeJesus *et al*. (2015) developed a software package for analyzing TnSeq data (which can also be applied to INSeq) called TRANSIT [470]. TRANSIT provides a graphical interface with three different methods for analyzing INSeq data, allowing combinations of methods or direct comparisons and simpler manual curation of results that differ between analysis methods. A Bayesian/Gumbel method for data analysis is complemented by a HMM for assigning gene classifications in individual test conditions, and a resampling (permutation) test allows comparative analysis of gene classifications between conditions [470]. The latter would be particularly applicable in the case of attachment at different pHs, as the model creates a distribution of read count

differences that could occur by chance between conditions before analyzing data with the null hypothesis that gene classifications are not different. Genes with truly different classifications between conditions fall out of the bounds of the resampling distribution and are classified as such. This offers a different and complimentary approach to the HMM for INSeq data analysis. Using this method with the INSeq data collected in Chapters 4 and 5 could allow comparison of results between analysis techniques for low TA site genes, with a consensus result being assigned to increase classification certainty.

7.5.3 ChIIPS and ChIIPS2

A preliminary new imaging system called ChIIPS was developed, better suited to imaging work with larger legume roots and upright confocal microscopes compared to TRIS. The specific REZ activation of the *lppE* promoter illustrated that devices such as ChIIPS and TRIS can be used to investigate spatio-temporal activation of genes on the root surface. To understand root-microbe interactions and root attachment better, spatio-temporal activation dynamics of genes upregulated in the rhizosphere and primary/secondary attachment factors should be investigated further. Plant roots display multiple different zones with very different properties [461]. However, how attachment factor requirements vary by root zone is largely unknown. Using reporter fusion strains could investigate not only this, but also timing of gene expression, and thus help determine the primary-secondary attachment switch more precisely.

Results in Chapter 6 indicate that the REZ exudes high levels of chemoattractants. However, the composition of these chemoattractants and how their exudation varies along the root length is not understood. An interesting further application of the biosensors developed by Pini *et al*. (2014) [201] would be for study of spatial root exudate

composition in ChIIPS. This could be achieved by inoculating roots with fluorescent protein exudate biosensor strains and observing the location and intensity of reporter induction along the plant root and over time. Ideally this experimental approach would use ratiometric imaging, with strains carrying a constitutive gene and a biosensor gene present, allowing a better quantitation of reporter expression relative to a background gene expression level.

An improved ChIIPS2 design was presented to enable future *Rhizobium*-legume interaction studies. ChIIPS2 enables the aerial plant parts to remain external to the inoculum-containing imaging chamber. Based on the work in Chapter 6, ChIIPS2 should be applicable to more complex questions than examining chemotactic responses to root zones. The question of root choice remains unexplored. For example, are a vetch and a clover root in the same chamber preferentially 'chosen' for interaction by *R. leguminosarum* biovars *viciae* and *trifolii*, respectively? Competition between strains could also be examined, for example by inoculating a vetch root with a mixed *R. leguminosarum* biovars *viciae* and *trifolii* population and observing whether one outcompetes the other for root interaction. From the plant perspective, use of transgenic reporter gene-carrying or mutant plant lines (e.g. Massalha *et al*. (2017, [188]) in ChIIPS2 could help define the gene-level responses of plants to different bacterial inoculums and the pathways needed for production of exudates that drive REZ interaction profiles. *Rhizobium*-pathogen-root interactions could also be investigated. Massalha *et al*. (2017) demonstrated the (likely surfactin-mediated) exclusion of *E. coli* from the roots of Arabidopsis by *B. subtilis* [188], but no such interactions have been investigated with *Rhizobium* and plant or human pathogens. In order to conduct such experiments

effectively, motility inducing culture conditions (such as the low carbon supply conditions used in this work) would have to be established for each bacterial strain and species.
A particularly innovative experimental setup could be gained by combining ChIIPS2 with transparent soil [463] when asking the questions posed above. This would vastly increase the physiological relevance of the environment under study by providing a heterogeneous, porous and transparent substrate for plant root growth which is directly suitable for imaging root-microbe interactions and has soil-like properties in terms of nutrient content and root growth morphology.

7.5.4 The plant perspective of primary root attachment

The work in this thesis has examined root attachment from the bacterial perspective. It is likely that plant root surface changes at different pHs (such as lectin disassociation, [71]) are one of the major drivers of bacterial attachment factor requirements, and that roots of different plants show wide variation in surface architecture. A holistic understanding of rhizobial primary root attachment would therefore also include plant root surface structures and their interaction with attachment factors. A lack of well-characterised mutant libraries for many plant species hampers this work and conducting bacterial attachment assays with a plant mutant library would be extremely time intensive. Other approaches that could be taken to investigating primary attachment from the plant perspective include labelling bacterial attachment factor proteins of interest and using them as bait in a plant root protein preparation, followed by analysis of binding partners. Ausmees *et al*. (2001) used a phage display approach to identify rhizobial proteins binding bacterial cell surface receptors [57]. Here, bacterial peptides of interest are displayed on the surface of a phage carrying the peptide gene. As the peptide and its coding DNA are in

the same phage particle, with a library of phages it is possible to enrich for particles interacting with an affinity substrate. For this research, isolated plant root proteins (such as lectins) could be used as affinity substrates to test for specific bacterial protein interacting partners, although the exact experimental setup would have to be carefully designed.

7.6 Concluding remarks

In this work proteomic and bioinformatic approaches have been combined with a newly developed Lux whole-root attachment assay and INSeq to define the primary root attachment determinants needed for Rlv3841 interaction with pea at different pHs and with a non-host legume and non-legume. An important contribution of this work is to suggest a new approach to defining primary root attachment for Rlv3841 that does not revolve around a glucomannan/rhicadhesin hypothesis. More holistically, primary attachment of Rlv3841 to plant roots should be thought of as a process employing diverse cellular factors, with a focus on EPS, LPS and peptidoglycan biosynthesis and modification, as well as direct adhesins, membrane proteins and uncharacterised factors, all of which can show different pH as well as plant host dependencies. Additionally, Rlv3841 is also able to use primary attachment mechanisms previously demonstrated in other bacterial species, such as Flp/Tad pili (used by *Agrobacterium, Pectobacterium, V. vulnificus, A. actinomycetemcomi* and *C. crescentus*, among others [44, 384, 385]), outer membrane proteins and LPS (used by *Azospirillum* [64, 65]). These are used in both an environment and plant host-dependent manner.

Further, the applicability of real-time imaging setups, notably the newly developed ChIIPS system, to investigating spatio-temporal root interaction and gene expression

dynamics has been demonstrated. By showing conserved preference for microbe interaction with the REZ, spatio-temporal reporter gene activation on roots, and providing a design for the ChIIPS2 system, this sets the stage for further high-impact developments in both imaging capabilities and our understanding of root-microbe interactions.

This work provides clear future avenues for further research. In the first instance, using RNASeq, protein-protein interaction and promoter binding assays would define regulatory networks underpinning primary attachment and provide new insights into PraR regulation and the activity of the important attachment regulator RL4145, among others. Secondly, an in-depth study of *pRL100053* function would likely represent a high-impact contribution to our understanding of symbiosis formation, given that it may function as a novel pea root hair attachment factor which also impacts downstream colonisation, infection thread formation and nodulation. Further, luminescence and confocal imaging assays at different time points could be used to shed light on spatio-temporal attachment and colonization dynamics. These and INSeq could be used in conjunction with reporter genes to better define the primary-secondary temporal attachment shift, which remains loosely defined at present. Combining new imaging technologies such as ChIIPS2 with experiments using mixed bacterial inoculums, different plant roots, plant pathogens, bioreporters and transparent soil would provide detailed insights into how and when root-microbe interactions are established and regulated under different conditions. Finally, studying both bacterial and plant root surface proteins could be used to define modes of action for attachment factors and explain differential pH/plant root attachment requirements.

Bibliography

1. Parsons J (2018) Global and Planetary Health. *Good Health and Well-Being. Encyclopedia of the UN Sustainable Development Goals.* In Leal W, Wall T, Azeiteuro U, Azul A, Brandi L, Oezuyar P (eds) Good Health and Well-Being, pp 1–12, Springer, Cham.
2. Rockström J, Steffen W, Noone K, Persson Å, Chapin FS, Lambin E, Lenton TM, Scheffer M, Folke C, Schellnhuber HJ, Nykvist B, de Wit CA, Hughes T, van der Leeuw S, Rodhe H, Sörlin S, Snyder PK, Costanza R, Svedin U, et al. (2009) Planetary boundaries: Exploring the safe operating space for humanity. *Ecology and Society* **14** (2): 474-475.
3. Whitmee S, Haines A, Beyrer C, Boltz F, Capon AG, De Souza Dias BF, Ezeh A, Frumkin H, Gong P, Head P, Horton R, Mace GM, Marten R, Myers SS, Nishtar S, Osofsky SA, Pattanayak SK, Pongsiri MJ, Romanelli C, et al. (2015) Safeguarding human health in the Anthropocene epoch: Report of the Rockefeller Foundation-Lancet Commission on planetary health. *The Lancet* **386** (10007): 1973–2028.
4. You D, Hug L, Chen Y, Wardlaw T, Newby H (2014) Levels and trends in child mortality. *New York: United Nations Inter-agency Group for Child Mortality Report 2014*.
5. Olinto P, Beegle K, Sobrado C, Uematsu H (2013) The State of the Poor: Where Are The Poor, Where Is Extreme Poverty Harder to End, and What Is the Current Profile of the World's Poor? *The World Bank* **125**: 1–8.
6. Gerland P, Raftery AE, Ševčíková H, Li N, Gu D, Spoorenberg T, Alkema L, Fosdick BK, Chunn J, Lalic N, Bay G, Buettner T, Heilig GK, Wilmoth J (2014) World population stabilization unlikely this century. *Science* **346** (6206): 234–237.
7. Steffen W, Richardson K, Rockström J, Cornell SE, Fetzer I, Bennett EM, Biggs R, Carpenter SR, De Vries W, De Wit CA, Folke C, Gerten D, Heinke J, Mace GM, Persson LM, Ramanathan V, Reyers B, Sörlin S (2015) Planetary boundaries: Guiding human development on a changing planet. *Science* **347** (6223): 736-747.
8. Foley JA, Monfreda C, Ramankutty N, Zaks D (2007) Our share of the planetary pie. *Proceedings of the National Academy of Sciences* **104** (31): 12585–12586.
9. Lambin EF, Meyfroidt P (2011) Global land use change, economic globalization, and the looming land scarcity. *Proceedings of the National Academy of Sciences* **108** (9): 3465–3472.
10. Suweis S, Rinaldo A, Maritan A, D'Odorico P (2013) Water-controlled wealth of nations. *Proceedings of the National Academy of Sciences* **110** (11): 4230-4233.
11. Gruber N, Galloway JN (2008) An Earth-system perspective of the global nitrogen cycle. *Nature* **451** (7176): 293–296.
12. Lehnert N, Dong HT, Harland JB, Hunt AP, White CJ (2018) Reversing nitrogen fixation. *Nature Reviews Chemistry* **2** (10): 278–289.
13. Smil V (2011) Global Energy: The Latest Infatuations. *American Scientist* **99** (3): 212–219.
14. Scheibel MG, Schneider S (2012) New insights into the biological and synthetic fixation of nitrogen. *Angewandte Chemie - International Edition* **51** (19): 4529–4531.
15. Hayatsu M, Tago K, Saito M (2008) Various players in the nitrogen cycle: Diversity and functions of the microorganisms involved in nitrification and denitrification. *Soil Science and Plant Nutrition* **54** (1): 33–45.
16. Erisman JW, Sutton MA, Galloway J, Klimont Z, Winiwarter W (2008) How a century of ammonia synthesis changed the world. *Nature Geoscience* **1** (10): 636–639.
17. Tilman D, Cassman KG, Matson PA, Naylor R, Polasky S (2002) Agricultural sustainability and intensive production practices. *Nature* **418** (6898): 671–677.
18. Stagnari F, Maggio A, Galieni A, Pisante M (2017) Multiple benefits of legumes for

agriculture sustainability: an overview. *Chemical and Biological Technologies in Agriculture* **4** (2): 1–13.
19. Gillespie-Bennett J, Pierse N, Wickens K, Crane J, Howden-Chapman P, Shields H, Viggers H, Free S, Phipps R, Fjallstrom P, Boulic M, Cunningham M, Lloyd B, Cunningham C, Chapman R, Bullen C, Woodward A (2011) The respiratory health effects of nitrogen dioxide in children with asthma. *European Respiratory Journal* **38** (2): 303–309.
20. E. M. Lodwig, A. H. F. Hosie, A. Bourde`s, K. Findlay, D. Allaway, R. Karunakaran, J. A. Downie PSP (2003) Amino-acid cycling drives nitrogen fixation in the legume-*Rhizobium* symbiosis. *Nature* **422**: 722-726.
21. Santi C, Bogusz D, Franche C (2013) Biological nitrogen fixation in non-legume plants. *Annals of Botany* **111** (5): 743–767.
22. Scott Fields (2004) Global nitrogen cycle: Cycling out of Control. *Environmental Health Perspectives* **112** (10): 556–563.
23. Galloway JN, Townsend AR, Erisman JW, Bekunda M, Cai Z, Freney JR, Martinelli LA, Seitzinger SP, Sutton MA (2008) Transformation of the nitrogen cycle: recent trends, questions, and potential solutions. *Science* **320**: 889–892.
24. Fowler D, Coyle M, Skiba U, Sutton M, Cape JN, Reis S, Sheppard L, Jenkins A, Grizetti B, Galloway JN, Vitousek P, Leach A, Bouwman L, Butterbach-Bahl K, Dentener F, Stevenson D, Amann M, Voss M (2013) The global nitrogen cycle in the 21th century. *Philisophical Transactions of the Royal Society of London, B Biological Sciences* **368** (1621): 1–13.
25. Poole P, Ramachandran V, Terpolilli J (2018) Rhizobia : from saprophytes to endosymbionts. *Nature Publishing Group* **16** (5): 291–303.
26. Oldroyd GED, Murray JD, Poole PS, Downie JA (2011) The Rules of Engagement in the Legume-Rhizobial Symbiosis. *Annual Review of Genetics* **45** (1): 119–144.
27. Udvardi M, Poole PS (2013) Transport and metabolism in legume-rhizobia symbioses. *Annual review of plant biology* **64**: 781–805.
28. Gan Y, Hamel C, O'Donovan JT, Cutforth H, Zentner RP, Campbell CA, Niu Y, Poppy L (2015) Diversifying crop rotations with pulses enhances system productivity. *Scientific Reports* **5** (14625): 1–14.
29. Walker TS, Bais HP, Grotewold E, Vivanco JM (2003) Update on Root Exudation and Rhizosphere Biology Root Exudation and Rhizosphere Biology 1. *Plant Physiology* **132**: 44–51.
30. Estabrook EM, Yoder JI (1998) Update on Plant-Plant Communication Plant-Plant Communications : Rhizosphere Signaling between Parasitic Angiosperms and Their Hosts. *Plant Physiology* **116**: 1–7.
31. Theunis M, Kobayashi H, Broughton WJ, Prinsen E (2007) Flavonoids, NodD1, NodD2, and Nod -Box NB15 Modulate Expression of the y4wEFG Locus That Is Required for Indole-3-Acetic Acid Synthesis in *Rhizobium* sp. strain NGR234. *Molecular Plant-Microbe Interactions* **17** (10): 1153–1161.
32. Brencic A, Winans SC, Colonization P (2005) Detection of and Response to Signals Involved in Host-Microbe Interactions by Plant-Associated Bacteria. *Microbiology and molecular biology reviews* **69** (1): 155–194.
33. Oldroyd GED (2013) Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nature* **11** (4): 252–263.
34. Wheatley RM, Poole PS (2018) Mechanisms of bacterial attachment to roots. *FEMS microbiology reviews* **42**: 448–461.
35. Geddes BA, Ryu MH, Mus F, Garcia Costas A, Peters JW, Voigt CA, Poole P (2015) Use of plant colonizing bacteria as chassis for transfer of N2-fixation to cereals. *Current Opinion in Biotechnology* **32**: 216–222.

36. Chabot R, Antoun H, Cescas MP (1996) Growth promotion of maize and lettuce by phosphate-solubilizing *Rhizobium leguminosarum* biovar. *phaseoli*. *Plant and Soil* **184** (2): 311–321.
37. Mishra RPN, Singh RK, Jaiswal HK, Kumar V, Maurya S (2006) *Rhizobium*-mediated induction of phenolics and plant growth promotion in rice (Oryza sativa L.). *Current Microbiology* **52** (5): 383–389.
38. García-Fraile P, Carro L, Robledo M, Ramírez-Bahena MH, Flores-Félix JD, Fernández MT, Mateos PF, Rivas R, Igual JM, Martínez-Molina E, Peix Á, Velázquez E (2012) *Rhizobium* promotes non-legumes growth and quality in several production steps: Towards a biofertilization of edible raw vegetables healthy for humans. *PLoS ONE* **7** (5): 1–7.
39. Weyens N, van der Lelie D, Taghavi S, Vangronsveld J (2009) Phytoremediation: plant-endophyte partnerships take the challenge. *Current Opinion in Biotechnology* **20** (2): 248–254.
40. Mei C, Flinn B (2010) The use of beneficial microbial endophytes for plant biomass and stress tolerance improvement. *Recent patents on biotechnology* **4** (1): 81–95.
41. Hardoim PR, van Overbeek LS, Elsas JD van (2008) Properties of bacterial endophytes and their proposed role in plant growth. *Trends in Microbiology* **16** (10): 463–471.
42. Santhanam R, Luu VT, Weinhold A, Goldberg J, Oh Y, Baldwin IT (2015) Native root-associated bacteria rescue a plant from a sudden-wilt disease that emerged during continuous cropping. *Proceedings of the National Academy of Sciences* **112** (36): 5013–5020.
43. Sistani N, Kaul H-P, Desalegn G, Wienkoop S (2017) *Rhizobium* Impacts on Seed Productivity, Quality, and Protection of *Pisum sativum* upon Disease Stress Caused by Didymella pinodes: Phenotypic, Proteomic, and Metabolomic Traits. *Frontiers in Plant Science* **8** (1961): 1–15.
44. Cécile Berne et al., Berne C, Ducret A, Hardy GG, Brun Y V., Hardy GG (2015) Adhesins Involved in Attachment to Abiotic Surfaces by Gram-Negative Bacteria. *Microbiology Spectrum* **3** (4): 1–27.
45. Zheng H, Mao Y, Teng J, Zhu Q, Ling J, Zhong Z (2015) Flagellar-dependent motility in *Mesorhizobium tianshanense* is involved in the early stage of plant host interaction: Study of an flgE mutant. *Current Microbiology* **70** (2): 219–227.
46. Kochkodan V, Tsarenko S, Potapchenko N, Kosinova V, Goncharuk V (2008) Adhesion of microorganisms to polymer membranes: a photobactericidal effect of surface treatment with TiO2. *Desalination* **220**: 380–385.
47. Laus MC, Logman TJ, Lamers GE, Van Brussel AAN, Carlson RW, Kijne JW (2006) A novel polar surface polysaccharide from *Rhizobium leguminosarum* binds host plant lectin. *Molecular Microbiology* **59** (6): 1704–1713.
48. Lagarda-Diaz I, Guzman-Partida AM, Vazquez-Moreno L (2017) Legume lectins: Proteins with diverse applications. *International Journal of Molecular Sciences* **18** (6): 1–18.
49. Sharon N, Lis H (2004) History of lectins: From hemagglutinins to biological recognition molecules. *Glycobiology* **14** (11): 53–62.
50. Lodeiro AR, López-García SL, Vázquez TEE, Favelukes G (2000) Stimulation of adhesiveness, infectivity, and competitiveness for nodulation of *Bradyrhizobium japonicum* by its pretreatment with soybean seed lectin. *FEMS Microbiology Letters* **188** (2): 177–184.
51. Lodeiro AR, Favelukes G (1999) Early interactions of *Bradyrhizobium japonicum* and soybean roots: specificity in the process of adsorption. *Soil Biology and Biochemistry* **31**: 1405–1411.
52. Smit G, Kijne JW, Lugtenberg BJ (1987) Involvement of both cellulose fibrils and a Ca2+-dependent adhesin in the attachment of *Rhizobium leguminosarum* to pea root hair tips. *Journal of Bacteriology* **169** (9): 4294–4301.

53. Smit G, Logman TJ, Boerrigter ME, Kijne JW, Lugtenberg BJ (1989) Purification and partial characterization of the *Rhizobium leguminosarum* biovar *viciae* Ca2+-dependent adhesin, which mediates the first step in attachment of cells of the family Rhizobiaceae to plant root hair tips. *Journal of Bacteriology* **171** (7): 4054–4062.
54. Swart S, Smit G, Lugtenberg BJJ, Kijne JW (1993) Restoration of attachment, virulence and nodulation of Agrobacterium tumefaciens chvB mutants by rhicadhesin. *Molecular Microbiology* **10** (3): 597–605.
55. Matthysse AG (2014) Attachment of Agrobacterium to plant surfaces. *Fronteirs in Plant Science* **5**: 1–8.
56. Williams A, Wilkinson A, Krehenbrink M, Russo DM, Zorreguieta A, Downie JA (2008) Glucomannan-mediated attachment of *Rhizobium leguminosarum* to pea root hairs is required for competitive nodule infection. *Journal of Bacteriology* **190** (13): 4706–4715.
57. Ausmees N, Jacobsson K, Lindberg M (2001) A unipolarly located, cell-surface-associated agglutinin, RapA, belongs to a family of Rhizobium-adhering proteins (Rap) in *Rhizobium leguminosarum* bv. *trifolii*. *Microbiology* **147** (3): 549–559.
58. Russo DM, Williams A, Edwards A, Posadas DM, Finnie C, Dankert M, Downie JA, Zorreguieta A (2006) Proteins exported via the PrsD-PrsE type I secretion system and the acidic exopolysaccharide are involved in biofilm formation by *Rhizobium leguminosarum*. *Journal of Bacteriology* **188** (12): 4474–4486.
59. Frederix M, Edwards A, Swiderska A, Stanger A, Karunakaran R, Williams A, Abbruscato P, Sanchez-Contreras M, Poole PS, Downie JA (2014) Mutation of praR in *Rhizobium leguminosarum* enhances root biofilms, improving nodulation competitiveness by increased expression of attachment proteins. *Molecular Microbiology* **93** (3): 464–478.
60. Rodriguez-Navarro DN, Dardanelli M, Ruiz-Sainz JE (2007) Attachment of bacteria to the roots of higher plants. *FEMS Microbiology Letters* **272**: 127-136
61. Croes C, Moens S, Bastelaere E, Vanderleyden J, Michiels W (1993) The polar flagellum mediates Azospirillum brasilense adsorption to wheat roots. *Journal of general microbiology* **139** (1993): 2261–2269.
62. Mora P, Rosconi F, Franco Fraguas L, Castro-Sowinski S (2008) Azospirillum brasilense Sp7 produces an outer-membrane lectin that specifically binds to surface-exposed extracellular polysaccharide produced by the bacterium. *Archives of Microbiology* **189** (5): 519–524.
63. Rossi FA, Medeot DB, Liaudat JP, Pistorio M, Jofré E (2016) In Azospirillum brasilense, mutations in flmA or flmB genes affect polar flagellum assembly, surface polysaccharides, and attachment to maize roots. *Microbiological Research* **190**: 55–62.
64. Burdman S, Dulguerova G, Okon Y, Jurkevitch E (2001) Purification of the Major Outer Membrane Protein of Azospirillum brasilense, Its Affinity to Plant Roots, and Its Involvement in Cell Aggregation. *Molecular Plant-Microbe Interactions* **14** (4): 555–561.
65. Balsanelli E, Tuleski TR, de Baura VA, Yates MG, Chubatsu LS, de Oliveira Pedrosa F, de Souza EM, Monteiro RA (2013) Maize Root Lectins Mediate the Interaction with Herbaspirillum seropedicae via N-Acetyl Glucosamine Residues of Lipopolysaccharides. *PLoS ONE* **8** (10): 1–9.
66. Sauer K, Camper AK (2001) Characterization of phenotypic changes in *Pseudomonas putida* in response to surface-associated growth. *Journal of Bacteriology* **183** (22): 6579–6589.
67. Alvarez Crespo MC, Valverde C (2009) A single mutation in the oprF mRNA leader confers strict translational control by the Gac/Rsm system in *Pseudomonas fluorescens* CHA0. *Current Microbiology* **58** (2): 182–188.
68. De Mot R, Vanderleydenl J, Schoofs G, Roelandt A, Declerck P, Proost P, Van Damme J, Vanderleyden J (1994) Molecular characterization of the major outer- membrane protein OprF from plant root- colonizing *Pseudornonas fluorescens*. *Microbiology* **140**: 1377–

1387.
69. Barak JD, Gorski L, Naraghi-Arani P, Charkowski AO (2005) Salmonella enterica virulence genes are required for bacterial attachment to plant tissue. *Applied and environmental microbiology* **71** (10): 5685–91.
70. Cooley MB, Miller WG, Mandrell RE (2003) Colonization of *Arabidopsis thaliana* with *Salmonella enterica* and Enterohemorrhagic *Escherichia coli* O157 : H7 and Competition by *Enterobacter asburiae*. *Applied and Environmental Microbiology* **69** (8): 4915–4926.
71. Downie JA (2010) The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. *FEMS Microbiology Reviews* **34** (2): 150–170.
72. Gage DJ (2004) Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiology and molecular biology reviews* **68** (2): 280–300.
73. Poole P (2017) Shining a light on the dark world of plant root–microbe interactions. *Proceedings of the National Academy of Sciences* **114** (17): 4281-4283.
74. Abdian PL, Caramelo JJ, Ausmees N, Zorreguieta A (2013) RapA2 Is a Calcium-binding Lectin Composed of Two Highly Conserved Cadherin-like Domains That Specifically Recognize *Rhizobium leguminosarum* Acidic Exopolysaccharides. *Journal of Biological Chemistry* **288** (4): 2893–2904.
75. Mongiardini EJ, Ausmees N, Pérez-Giménez J, Julia Althabegoiti M, Ignacio Quelas J, López-García SL, Lodeiro AR (2008) The rhizobial adhesion protein RapA1 is involved in adsorption of rhizobia to plant roots but not in nodulation. *FEMS Microbiology Ecology* **65** (2): 279–288.
76. Skorupska A, Janczarek M, Marczak M, Mazur A, Król J (2006) Rhizobial exopolysaccharides: Genetic control and symbiotic functions. *Microbial Cell Factories* **5** (7): 1–19.
77. Janczarek M, Rachwał K, Cieśla J, Ginalska G, Bieganowski A (2015) Production of exopolysaccharide by *Rhizobium leguminosarum* bv. *trifolii* and its role in bacterial attachment and surface properties. *Plant and Soil* **388** (1–2): 211–227.
78. Xu J, Kim J, Koestler BJ, Choi JH, Waters CM, Fuqua C (2013) Genetic analysis of agrobacterium tumefaciens unipolar polysaccharide production reveals complex integrated control of the motile-to-sessile switch. *Molecular Microbiology* **89** (5): 929–948.
79. Valentini M, Filloux A (2016) Biofilms and Cyclic di-GMP (c-di-GMP) signaling: Lessons from *Pseudomonas aeruginosa* and other bacteria. *Journal of Biological Chemistry* **291** (24): 12547–12555.
80. Simm R, Morr M, Kader A, Nimtz M, Römling U (2004) GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessibility to motility. *Molecular Microbiology* **53** (4): 1123–1134.
81. Romling U, Galperin MY, Gomelsky M (2013) Cyclic di-GMP: the First 25 Years of a Universal Bacterial Second Messenger. *Microbiology and Molecular Biology Reviews* **77** (1): 1–52.
82. Ryjenkov DA, Tarutina M, Moskvin O V, Gomelsky M (2005) Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: Insights into biochemistry of the GGDEF protein domain. *Journal of Bacteriology* **187** (5): 1792–1798.
83. Schmidt AJ, Ryjenkov DA, Gomelsky M (2005) The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: Enzymatically active and inactive EAL domains. *Journal of Bacteriology* **187** (14): 4774–4781.
84. Chou S-H, Galperin MY (2016) Diversity of Cyclic Di-GMP-Binding Proteins and Mechanisms. *Journal of Bacteriology* **198** (1): 32–46.
85. Galperin MY (2010) Diversity of structure and function of response regulator output domains. *Current Opinion in Microbiology* **13** (2): 150–159.
86. Little RH, Woodcock SD, Campilongo R, Fung RKY, Heal R, Humphries L, Pacheco-

Moreno A, Paulusch S, Stigliano E, Vikeli E, Ward D, Malone JG (2019) Differential Regulation of Genes for Cyclic-di-GMP Metabolism Orchestrates Adaptive Changes During Rhizosphere Colonization by *Pseudomonas fluorescens*. *Frontiers in Microbiology* **10**: 1–14.

87. Heindl JE, Wang Y, Heckel BC, Mohari B, Feirer N, Fuqua C (2014) Mechanisms and regulation of surface interactions and biofilm formation in Agrobacterium. **5**: 1–21.
88. Schäper S, Krol E, Skotnicka D, Kaever V, Hilker R, Søgaard-Andersen L, Becker A (2016) Cyclic Di-GMP Regulates Multiple Cellular Functions in the Symbiotic Alphaproteobacterium *Sinorhizobium meliloti*. *Journal of bacteriology* **198** (3): 521–35.
89. Burdman S, Okon Y, Jurkevitch E (2000) Surface characteristics of Azospirillum brasilense in relation to cell aggregation and attachment to plant roots. *Critical Reviews in Microbiology* **26** (2): 91–110.
90. Fuqua C (2010) Passing the baton between laps: Adhesion and cohesion in *Pseudomonas putida* biofilms: MicroCommentary. *Molecular Microbiology* **77** (3): 533–536.
91. Martínez-Gil M, Yousef-Coronado F, Espinosa-Urgel M (2010) LapF, the second largest *Pseudomonas putida* protein, contributes to plant root colonization and determines biofilm architecture. *Molecular Microbiology* **77** (3): 549–561.
92. Robledo M, Jiménez-Zurdo JI, Velázquez E, Trujillo ME, Zurdo-Piñeiro JL, Ramírez-Bahena MH, Ramos B, Díaz-Mínguez JM, Dazzo F, Martínez-Molina E, Mateos PF (2008) *Rhizobium* cellulase CelC2 is essential for primary symbiotic infection of legume host roots. *Proceedings of the National Academy of Sciences* **105** (19): 7064 – 7069.
93. Fournier J, Timmers ACJ, Sieberer BJ, Jauneau A, Chabaud M, Barker DG (2008) Mechanism of Infection Thread Elongation in Root Hairs of *Medicago truncatula* and Dynamic Interplay with Associated Rhizobial Colonization. *Plant Physiology* **148** (4): 1985–1995.
94. Brewin NJ (2004) Plant cell wall remodelling in the *Rhizobium*-legume symbiosis. *Critical Reviews in Plant Sciences* **23** (4): 293–316.
95. Gage DJ, Margolin W (2000) Hanging by a thread: Invasion of legume plants by rhizobia. *Current Opinion in Microbiology* **3** (6): 613–617.
96. Kawaharada Y, Kelly S, Nielsen MW, Hjuler CT, Gysel K, Muszyński A, Carlson RW, Thygesen MB, Sandal N, Asmussen MH, Vinther M, Andersen SU, Krusell L, Thirup S, Jensen KJ, Ronson CW, Blaise M, Radutoiu S, Stougaard J (2015) Receptor-mediated exopolysaccharide perception controls bacterial infection. *Nature* **523** (7560): 308–312.
97. Oke V, Long SR (1999) Bacteroid formation in the *Rhizobium*–legume symbiosis. *Current Opinion in Microbiology* **2** (6): 641–646.
98. Kondorosi E, Roudier F, Gendreau E (2000) Plant cell-size control: Growing by ploidy? *Current Opinion in Plant Biology* **3** (6): 488–492.
99. Maunoury N, Redondo-Nieto M, Bourcy M, Van De Velde W, Alunni B, Laporte P, Durand P, Agier N, Marisa L, Vaubert D, Delacroix H, Duc G, Ratet P, Aggerbeck L, Kondorosi E, Mergaert P (2010) Differentiation of symbiotic cells and endosymbionts in *Medicago truncatula* nodulation are coupled to two transcriptome-switches. *PLoS ONE* **5** (3): 1–14.
100. Van De Velde W, Zehirov G, Szatmari A, Debreczeny M, Ishihara H, Kevei Z, Farkas A, Mikulass K, Nagy A, Tiricz H, Satiat-Jeunemaître B, Alunni B, Bourge M, Kucho KI, Abe M, Kereszt A, Maroti G, Uchiumi T, Kondorosi E, et al. (2010) Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science* **327** (5969): 1122–1126.
101. Montiel J, Downie JA, Farkas A, Bihari P, Herczeg R, Bálint B, Mergaert P, Kereszt A, Kondorosi É (2017) Morphotype of bacteroids in different legumes correlates with the number and type of symbiotic NCR peptides. *Proceedings of the National Academy of Sciences* **114** (19): 5041–5046.
102. Czernic P, Gully D, Cartieaux F, Moulin L, Guefrachi I, Patrel D, Pierre O, Fardoux J,

Chaintreuil C, Nguyen P, Gressent F, Da Silva C, Poulain J, Wincker P, Rofidal V, Hem S, Barrière Q, Arrighi J-F, Mergaert P, et al. (2015) Convergent Evolution of Endosymbiont Differentiation in Dalbergioid and Inverted Repeat-Lacking Clade Legumes Mediated by Nodule-Specific Cysteine-Rich Peptides. *Plant Physiology* **169** (2): 1254–1265.
103. Roponen I (1970) The Effect of Darkness on the Leghemoglobin Content and Amino Acid Levels in the Root Nodules of Pea Plants. *Physiologia Plantarum* **23** (3): 452–460.
104. Downie JA (2014) Legume nodulation. *Current Biology* **24** (5): R184–R190.
105. Paau AS, Bloch CB, Brill WJ (1980) Developmental fate of *Rhizobium meliloti* bacteroids in alfalfa nodules. *Journal of Bacteriology* **143** (3): 1480–1490.
106. Müller J, Boller T, Wiemken A (2001) Trehalose becomes the most abundant non-structural carbohydrate during senescence of soybean nodules. *Journal of Experimental Botany* **52** (358): 943–947.
107. Laranjo M, Alexandre A, Oliveira S (2014) Legume growth-promoting rhizobia: An overview on the *Mesorhizobium* genus. *Microbiological Research* **169** (1): 2–17.
108. Hoffman BM, Lukoyanov D, Yang Z-Y, Dean DR, Seefeldt LC (2014) Mechanism of Nitrogen Fixation by Nitrogenase: The Next Stage. *Chemical Reviews* **114** (8): 4041–4062.
109. Mus F, Crook MB, Garcia K, Costas AG, Geddes BA, Kouri ED, Paramasivan P, Ryu MH, Oldroyd GED, Poole PS, Udvardi MK, Voigt CA, Ane JM, Peters JW (2016) Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. *Applied and Environmental Microbiology* **82** (13): 3698–3710.
110. Hu Y, Ribbe MW (2013) Nitrogenase assembly. *Biochimica et Biophysica Acta - Bioenergetics* **1827** (8–9): 1112–1122.
111. Clarke VC, Loughlin PC, Day DA, Smith PMC (2014) Transport processes of the legume symbiosome membrane. *Frontiers in Plant Science* **5** (15): 1–9.
112. Ott T, Van Dongen JT, Günther C, Krusell L, Desbrosses G, Vigeolas H, Bock V, Czechowski T, Geigenberger P, Udvardi MK (2005) Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. *Current Biology* **15** (6): 531–535.
113. Kundu S, Hargrove MS (2003) Distal heme pocket regulation of ligand binding and stability in soybean leghemoglobin. *Proteins: Structure, Function and Genetics* **50** (2): 239–248.
114. Kundu S, Trent JT, Hargrove MS (2003) Plants, humans and hemoglobins. *Trends in Plant Science* **8** (8): 387–393.
115. Glenn AR, Brewin NJ (1981) Succinate-resistant Mutants of *Rhizobium leguminosarum*. *Journal of general microbiology* **126** : 237–241.
116. Finan TM, Wood JM, Jordan DC (1983) Symbiotic properties of C4-dicarboxylic acid transport mutants of *Rhizobium leguminosarum*. *Journal of Bacteriology* **154** (3): 1403–1413.
117. Ronson CW, Astwood PM, Downie JA (1984) Molecular cloning and genetic organization of C4-dicarboxylate transport genes from *Rhizobium leguminosarum*. *Journal of Bacteriology* **160** (3): 903–909.
118. Lodwig E, Poole P (2003) Metabolism of *Rhizobium* Bacteroids. *Critical Reviews in Plant Sciences* **22** (1): 37–78.
119. Niemietz CM, Tyerman SD (2000) Channel-mediated permeation of ammonia gas through the peribacteroid membrane of soybean nodules. *FEBS Letters* **465** (2–3): 110–114.
120. Al-Falih AMK (2002) Factors affecting the efficiency of symbiotic nitrogen fixation by *Rhizobium*. *Pakistan journal of biological sciences* **5** (11): 1277–1293.
121. Zgadzaj R, Garrido-Oter R, Jensen DB, Koprivova A, Schulze-Lefert P, Radutoiu S (2016) Root nodule symbiosis in *Lotus japonicus* drives the establishment of distinctive rhizosphere, root, and nodule bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America* **113** (49): 7996–8005.

122. Karkute SG, Singh AK, Gupta OP, Singh PM, Singh B (2017) CRISPR/Cas9 Mediated Genome Engineering for Improvement of Horticultural Crops. *Frontiers in Plant Science* **8** : 1–6.
123. Parsons J (2018) The theory and applications of CRISPR in plant and tree improvement. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources* **13** (7): 1–12.
124. FAO (2009) How to Feed the World in 2050. *Insights from an expert meeting at FAO*: 1–35.
125. Webb J, Misselbrook TH, Tscharntke T, Clough Y, Wanger TC, Jackson L, Motzke I, Perfecto I, Vandermeer J, Whitbread A, Tilman D, Cassman KG, Matson P a, Naylor R, Polasky S, Balzer C, Hill J, Befort BL, Thévenot A, et al. (2011) The Future of Food and Farming: Challenges and choices for global sustainability. *The Government Office for Science, London* **149**: 193–208.
126. Gutjahr C, Banba M, Croset V, An K, Miyao A, An G, Hirochika H, Imaizumi-Anraku H, Paszkowski U (2008) Arbuscular Mycorrhiza-Specific Signaling in Rice Transcends the Common Symbiosis Signaling Pathway. *The Plant Cell* **20** (11): 2989–3005.
127. Maillet F, Poinsot V, André O, Puech-Pagés V, Haouy A, Gueunier M, Cromer L, Giraudet D, Formey D, Niebel A, Martinez EA, Driguez H, Bécard G, Dénarié J (2011) Fungal lipochitooligosaccharide symbiotic signals in arbuscular mycorrhiza. *Nature* **469** (7328): 58–64.
128. Oldroyd GED, Dixon R (2014) Biotechnological solutions to the nitrogen problem. *Current Opinion in Biotechnology* **26**: 19–24.
129. Setten L, Soto G, Mozzicafreddo M, Fox AR, Lisi C, Cuccioloni M, Angeletti M, Pagano E, Díaz-Paleo A, Ayub ND (2013) Engineering *Pseudomonas protegens* Pf-5 for Nitrogen Fixation and its Application to Improve Plant Growth under Nitrogen-Deficient Conditions. *PLoS ONE* **8** (5): 1–14.
130. Parsons J (2017) Can we fix it? Biotechnology and nitrogen fixing symbioses. *The Biologist* **64** (4): 12–16.
131. Mondy S, Lenglet A, Beury-Cirou A, Libanga C, Ratet P, Faure D, Dessaux Y (2014) An increasing opine carbon bias in artificial exudation systems and genetically modified plant rhizospheres leads to an increasing reshaping of bacterial populations. *Molecular Ecology* **23** (19): 4846–4861.
132. Geddes B (2014) Engineering synthetic symbiosis : characterizing rhizopines and developing tools. *Proceedings of Synthetic Symbioses* **2**: 1-17
133. Ormeno-Orrillo E, Servin-Garciduenas LE, Rogal MA, Gonzalez V, Peralta H, Mora J, Martinez-Romero J, Martinez-Romero E (2015) Taxonomy of rhizobia and agrobacteria from the Rhizobiaceae family in light of genomics. *Systematic and Applied Microbiology* **197**: 51–57.
134. Berrada H (2014) Taxonomy of the Rhizobia: Current Perspectives. *British Microbiology Research Journal* **4** (6): 616–639.
135. Mazur A, Stasiak G, Wielbo J, Kubik-Komar A, Marek-Kozaczuk M, Skorupska A (2011) Intragenomic diversity of *Rhizobium leguminosarum* bv. *trifolii* clover nodule isolates. *BMC Microbiology* **11** (123): 1–15.
136. Pérez Carrascal OM, VanInsberghe D, Juárez S, Polz MF, Vinuesa P, González V (2016) Population genomics of the symbiotic plasmids of sympatric nitrogen-fixing *Rhizobium* species associated with Phaseolus vulgaris. *Environmental Microbiology* **18** (8): 2660–2676.
137. González V, Acosta JL, Santamaría RI, Fernández PBJL, Hernández González IL, Díaz R, Flores M, Palacios R, Mora J, Dávila G (2010) Conserved symbiotic plasmid DNA sequences in the multireplicon pangenomic structure of *Rhizobium etli*. *Applied and Environmental Microbiology* **76** (5): 1604–1614.

138. Young JPW, Crossman LC, Johnston AWB, Thomson NR, Ghazoui ZF, Hull KH, Wexler M, Curson ARJ, Todd JD, Poole PS, Mauchline TH, East AK, Quail MA, Churcher C, Arrowsmith C, Cherevach I, Chillingworth T, Clarke K, Cronin A, et al. (2006) The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biology* **7** (4): 1–20.
139. Karunakaran R, Ramachandran VK, Seaman JC, East AK, Mouhsine B, Mauchline TH, Prell J, Skeffington A, Poole PS (2009) Transcriptomic Analysis of *Rhizobium leguminosarum* biovar *viciae* in Symbiosis with Host Plants *Pisum sativum* and *Vicia cracca*. *Journal of Bacteriology* **191** (12): 4002–4014.
140. Kumar N, Lad G, Giuntini E, Kaye ME, Udomwong P, Shamsani NJ, Young PW, Bailly X (2015) Bacterial genospecies that are not ecologically coherent: population genomics of *Rhizobium leguminosarum*. *Open Biology* **5** (1): 140133–140133.
141. Ramachandran VK, East AK, Karunakaran R, Downie JA, Poole PS (2011) Adaptation of *Rhizobium leguminosarum* to pea, alfalfa and sugar beet rhizospheres investigated by comparative transcriptomics. *Genome Biology* **12** (10): 1:12.
142. White JP, Prell J, Ramachandran VK, Poole PS (2009) Characterization of a γ-aminobutyric acid transport system of *Rhizobium leguminosarum* bv. *viciae* 3841. *Journal of Bacteriology* **191** (5): 1547–1555.
143. Wheatley R, Ramachandran VK, Geddes BA, Perry BJ, Yost CK, Poole PS (2017) Role of O2 in the Growth of *Rhizobium leguminosarum* bv. *viciae* 3841 on Glucose and Succinate. *Journal of Bacteriology* **199** (1): 1–17.
144. Ramírez-Romero MA, Soberón N, Pérez-Oseguera A, Téllez-Sosa J, Cevallos MA (2000) Structural elements required for replication and incompatibility of the *Rhizobium etli* symbiotic plasmid. *Journal of Bacteriology* **182** (11): 3117–3124.
145. Harrison PW, Lower RPJ, Kim NKD, Young JPW (2010) Introducing the bacterial "chromid": Not a chromosome, not a plasmid. *Trends in Microbiology* **18** (4): 141–148.
146. van Opijnen T, Camilli A (2013) Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nature Reviews Microbiology* **11** (7): 435–442.
147. Lardi M, Pessi G (2018) Functional Genomics Approaches to Studying Symbioses between Legumes and Nitrogen-Fixing Rhizobia. *High-Throughput* **7** (2): 1–20.
148. Hosie AHF, Allaway D, Poole PS (2002) A Monocarboxylate Permease of *Rhizobium leguminosarum* Is the First Member of a New Subfamily of Transporters. *Journal of Bacteriology* **184** (19): 5436–5448.
149. Gray AN, Koo B, Shiver AL, Peters JM, Osadnik H, Gross CA (2015) High throughput bacterial functional genomics in the sequencing era. *Current Opinion in Microbiology* **27** : 86–95.
150. Barquist L, Boinett CJ, Cain AK (2013) Approaches to querying bacterial genomes with transposon-insertion sequencing. *RNA biology* **10** (7): 1161–1169.
151. Perry BJ, Yost CK (2014) Construction of a mariner-based transposon vector for use in insertion sequence mutagenesis in selected members of the Rhizobiaceae. *BMC Microbiology* **14** (1): 1–11.
152. Johnston AWB, Beynon JL, Buchanan-Wollaston A V., Setchell SM, Hirsch PR, Beringer JE (1978) High frequency transfer of nodulating ability between strains and species of *Rhizobium* [29]. *Nature* **276** (5688): 634–636.
153. Reeve WG, Tiwari RP, Worsley PS, Dilworth MJ, Glenn AR, Howieson JG (1999) Constructs for insertional mutagenesis, transcriptional signal localization and gene regulation studies in root nodule and other bacteria. *Microbiology* **145** (6): 1307–1316.
154. Wilson KJ (1995) Molecular techniques for the study of rhizobial ecology in the field. *Soil Biology and Biochemistry* **27** (4–5): 501–514.
155. Prell J, Poole P, Untiet V, Karunakaran R (2012) The PTSNtr System Globally Regulates ATP-Dependent Transporters in *Rhizobium leguminosarum*. *Biological Nitrogen Fixation*

84 (1): 117-129.
156. Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270** (5235): 467-470.
157. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics Zhong. *Nature Reviews Genetics* **10** (1): 57–63.
158. Rachwal K, Matczyńska E, Janczarek M (2015) Transcriptome profiling of a *Rhizobium leguminosarum* bv. *trifolii* rosR mutant reveals the role of the transcriptional regulator RosR in motility, synthesis of cell-surface components, and other cellular processes. *BMC Genomics* **16** (1111): 1–18.
159. Green RT, East AK, Karunakaran R, Downie JA, Poole PS (2019) Transcriptomic analysis of *Rhizobium leguminosarum* bacteroids in determinate and indeterminate nodules. *Microbial Genomics* **5** (2): 1–16.
160. Kozak M (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361** (1–2): 13–37.
161. Graves PR, Haystead TAJ (2002) Molecular Biologist's Guide to Proteomics. *Microbiology and Molecular Biology Reviews* **66** (1): 39–63.
162. Natera SHA, Guerreiro N, Djordjevic MA (2000) Proteome Analysis of Differentially Displayed Proteins As a Tool for the Investigation of Symbiosis. *Molecular Plant-Microbe Interactions* **13** (9): 995–1009.
163. Djordjevic MA (2004) *Sinorhizobium meliloti* metabolism in the root nodule: A proteomic perspective. *Proteomics* **4** (7): 1859–1872.
164. Djordjevic MA, Chen HC, Natera S, Van Noorden G, Menzel C, Taylor S, Renard C, Geiger O, Weiller GF (2003) A Global Analysis of Protein Expression Profiles in *Sinorhizobium meliloti* : Discovery of New Genes for Nodule Occupancy and Stress Adaptation. *Molecular Plant-Microbe Interactions* **16** (6): 508–524.
165. van Opijnen T, Bodi KL, Camilli A (2009) Tn-seq: High-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods* **6** (10): 767–772.
166. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R, Gordon JI (2009) Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat. *Cell Host and Microbe* **6** (3): 279–289.
167. Gawronski JD, Wong SMS, Giannoukos G, Ward D V, Akerley BJ (2009) Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proceedings of the National Academy of Sciences* **106** (38): 16422–16427.
168. Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK (2009) Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Research* **19** (12): 2308–2316.
169. Goodman AL, Wu M, Gordon JI (2011) Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries. *Nature Protocols* **6** (12): 1969–1980.
170. Chaudhuri RR, Allen AG, Owen PJ, Shalom G, Stone K, Harrison M, Burgis TA, Lockyer M, Garcia-Lara J, Foster SJ, Pleasance SJ, Peters SE, Maskell DJ, Charles IG (2009) Comprehensive identification of essential Staphylococcus aureus genes using Transposon-Mediated Differential Hybridisation (TMDH). *BMC Genomics* **10** (291): 1–18.
171. Sassetti CM, Boyd DH, Rubin EJ (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular Microbiology* **48** (1): 77–84.
172. Klein BA, Tenorio EL, Lazinski DW, Camilli A, Duncan MJ, Hu LT (2012) Identification of essential genes of the periodontal pathogen Porphyromonas gingivalis. *BMC Genomics* **13** (578): 1–17.

173. Griffin JE, Gawronski JD, DeJesus MA, Ioerger TR, Akerley BJ, Sassetti CM (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathogens* **7** (9): 1–9.
174. Mann B, van Opijnen T, Wang J, Obert C, Wang YD, Carter R, McGoldrick DJ, Ridout G, Camilli A, Tuomanen EI, Rosch JW (2012) Control of virulence by small RNAs in Streptococcus pneumoniae. *PLoS Pathogens* **8** (7): 34.
175. Perry BJ, Akter MS, Yost CK (2016) The use of transposon insertion sequencing to interrogate the core functional genome of the legume symbiont *Rhizobium leguminosarum*. *Frontiers in Microbiology* **7**: 1–21.
176. Plasterk RHA, Izsvak Z, Ivics Z (1999) Resident Aliens: The Tc1/marniner superfamily of transposable elements. *Trends in Genetics* **15** (8): 326–332.
177. DeJesus MA, Nambi S, Smith CM, Baker RE, Sassetti CM, Ioerger TR (2017) Statistical analysis of genetic interactions in Tn-Seq data. *Nucleic Acids Research* **45** (11): 1–11.
178. Zomer A, Burghout P, Bootsma HJ, Hermans PWM, van Hijum SAFT (2012) Essentials: Software for rapid analysis of high throughput transposon insertion sequencing data. *PLoS ONE* **7** (8): 1–9.
179. DeJesus MA, Zhang YJ, Sassetti CM, Rubin EJ, Sacchettini JC, Ioerger TR (2013) Bayesian analysis of gene essentiality based on sequencing of transposon insertion libraries. *Bioinformatics* **29** (6): 695–703.
180. Christen B, Abeliuk E, Collier JM, Kalogeraki VS, Passarelli B, Coller JA, Fero MJ, McAdams HH, Shapiro L (2011) The essential genome of a bacterium. *Molecular Systems Biology* **7** (528): 1–7.
181. Akerley BJ, Rubin EJ, Camilli A, Lampe DJ, Robertson HM, Mekalanos JJ (1998) Systematic identification of essential genes by in vitro mariner mutagenesis. *Proceedings of the National Academy of Sciences* **95** (15): 8927–8932.
182. Smith V, Chou KN, Lashkari D, Botstein D, Brown PO (1996) Functional analysis of the genes of yeast chromosome V by genetic footprinting. *Science* **274** (5295): 2069–2074.
183. DeJesus MA, Ioerger TR (2013) A Hidden Markov Model for identifying essential and growth-defect regions in bacterial genomes from transposon insertion sequencing data. *BMC Bioinformatics* **14** (1): 1-12.
184. Jaeger III CH, Lindow SE, Miller W, Clark E, Firestone MK (1999) Mapping sugar and aminon acid exudation from roots in soil using bacterial sensors of sucrose and tryptophan. *Applied an Environmental Microbiology* **65** (6): 2685–2690.
185. Ofek M, Hadar Y, Minz D (2011) Colonization of cucumber seeds by bacteria during germination. *Environmental Microbiology* **13** (10): 2794–2807.
186. Aulakh MS, Wassmann R, Bueno C, Kreuzwieser J, Rennenberg H (2001) Characterization of root exudates at different growth stages of ten rice (Oryza sativa L.) cultivars. *Plant Biology* **3** (2): 139–148.
187. Chaparro JM, Badri D V., Vivanco JM (2014) Rhizosphere microbiome assemblage is affected by plant development. *ISME Journal* **8** (4): 790–803.
188. Massalha H, Korenblum E, Malitsky S, Shapiro OH, Aharoni A (2017) Live imaging of root–bacteria interactions in a microfluidics setup. *Proceedings of the National Academy of Sciences* **114** (17) 4549-4554 .
189. Gros M-FS, Shinde S, Akins C, Johnson JL, Zerbs S, Wilton R, Kemner KM, Noirot PH, Babnigg G (2018) RMI-chip: a microfluidics setup for functional imaging of microbial interactions with tree roots. *BioRxiv*: (pre-print).
190. Johnston AW, Beringer JE (1975) Identification of the *Rhizobium* strains in pea root nodules using genetic markers. *Journal of General Microbiology* **87** (2): 343–50.
191. Frederix M, Edwards A, Mcanulla C, Downie JA (2011) Co-ordination of quorum-sensing regulation in *Rhizobium leguminosarum* by induction of an anti-repressor. *Molecular Microbiology* **81** (4): 994–1007.

192. Hood GA (2013) Physiological response of *Rhizobium leguminosarum* during bacteroid development. Doctoral Thesis submitted to the Department of Molecular Microbiology, John Innes Centre (University of East Anglia), Norwich.
193. Mcnally A (2012) Rhizosphere Colonization : Assigning genes induced in the rhizosphere to regulatory networks. Masters Thesis submitted to the Department of Molecular Microbiology, John Innes Centre (University of East Anglia), Norwich.
194. Schafer A, Tauch A, Jager W, Kalinowski J, Thierbachb G, Puhler A (1994) Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum. *Gene* **145**: 69–73.
195. Ditta G, Stanfield S, Corbin D, Helinski DR, Donald Helinskl CR (1980) Broad host range DNA cloning system for Gram-negative bacteria: Construction of a gene bank of *Rhizobium meliloti* (plasmid RK2/plasmid vehicle/conjugal transfer/nif genes). *Genetics* **77** (12): 7347–735.
196. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28** (12): 1647–1649.
197. Chomczynski P, Rymaszewski M (2006) Alkaline polyethylene glycol-based method for direct PCR from bacteria, eukaryotic tissue samples, and whole blood. *BioTechniques* **40** (4): 454–458.
198. Buchanan-Wollaston V (1979) Generalized Transduction in *Rhizobium leguminosarum*. *Journal of General Microbiology* **112**: 135–142.
199. Keen J (1994) SDS Polyacrylamide Gele Electrophoresis of Proteins. In Walker JM (eds) *Basic Protein and Peptide Protocols, Methods in Molecular Biology*, Humana Press.
200. Buscaill P, Chandrasekar B, Sanguankiattichai N, Kourelis J, Kaschani F, Thomas EL, Morimoto K, Kaiser M, Preston GM, Ichinose Y, van der Hoorn RAL (2019) Glycosidase and glycan polymorphism control hydrolytic release of immunogenic flagellin peptides. *Science* **364** (6436): 1-11
201. Pini F, East AK, Appia-Ayme C, Tomek J, Karunakaran R, Mendoza-Suárez M, Edwards A, Terpolilli JJ, Roworth J, Downie JA, Poole PS (2017) Bacterial Biosensors for in Vivo Spatiotemporal Mapping of Root Secretion. *Plant Physiology* **174** (3): 1289–1306.
202. Martin. M (2011) Cutadapt removes adaptor sequences from high-throughput sequencing reads. *EMBnet* **17** (1): 10–12.
203. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10** (3): 1-10
204. Gish W, States DJ (1993) Identification of protein coding regions by database similarity search. *Natural Genetics* **3** : 266–272.
205. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, Von Mering C (2015) STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research* **43** (1): 447-D452.
206. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Cenk Sahinalp S, Ester M, Foster LJ, Brinkman FSL (2010) PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26** (13): 1608–1615.
207. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A (2012) Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9** (7): 676–82.
208. Smit G, Swart S, Lugtenberg BJJ, Kijne JW (1992) Molecular mechanisms of attachment

of *Rhizobium* bacteria to plant roots. *Molecular Microbiology* **6** (20): 2897–2903.
209. Swart S, Logman TJJ, Smit G, Lugtenberg BJJ, Kijne JW (1994) Purification and partial characterization of a glycoprotein from pea (*Pisum sativum*) with receptor activity for rhicadhesin, an attachment protein of Rhizobiaceae. *Plant Molecular Biology* **24** (1): 171–183.
210. Dardanelli M, Angelini J, Fabra A (2003) A calcium-dependent bacterial surface protein is involved in the attachment of rhizobia to peanut roots. *Canadian Journal of Microbiology* **49** (6): 399–405.
211. Smit G, Tubbing DMJ, Kijne JW, Lugtenbert JJ (1991) Role of Ca 2 + 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. *Archives of Microbiology* **155**: 278–283.
212. Aguilar MM, Ashby AM, Richards GJL, Watson MD, Shaw CH (1988) Chemotaxis of *Rhizobium leguminosarum* biovar *phaseoli* towards Flavonoid Inducers of the Symbiotic Nodulation Genes. *Journal of General Microbiology* **134**: 2741–2746.
213. Dharmatilake AJ, Bauer WD (1992) Chemotaxis of *Rhizobium meliloti* towards Alfalfa Roots. *Applied and Environmental Microbiology* **58** (4): 1153–118.
214. Vesper SJ (1987) Production of Pili (Fimbriae) by *Pseudomonas fluorescens* and Correlation with Attachment to Corn Roots. *Applied and environmental microbiology* **53** (7): 1397–405.
215. Pratt LA, Kolter R (1998) Genetic analysis of *Escherichia coli* biofilm formation: Roles of flagella, motility, chemotaxis and type I pili. *Molecular Microbiology* **30** (2): 285–293.
216. Rachwal K, Boguszewska A, Kopcinska J, Karas M, Tchórzewski M, Janczarek M (2016) The regulatory protein RosR affects *Rhizobium leguminosarum* bv. *trifolii* protein profiles, cell surface properties, and symbiosis with clover. *Frontiers in Microbiology* **7**: 1–21.
217. Dazzo FB, Napoli CA, Hubbell DH (1976) Adsorption of bacteria to roots as related to host specificity in the *Rhizobium*-clover symbiosis. *Applied and Environmental Microbiology* **32** (1): 166–171.
218. Dazzo FB, Truchet GL, Sherwood JE, Hrabak EM, Abe M, Pankratz SH (1984) Specific phases of root hair attachment in the *Rhizobium trifolii*-clover symbiosis. *Applied and Environmental Microbiology* **48** (6): 1140–1150.
219. Smit G, Kijne JW, Lugtenberg BJJ (1986) Correlation between extracellular fibrils and attachment of *Rhizobium leguminosarum* to pea root hair tips. *Journal of Bacteriology* **168** (2): 821–827.
220. Chait BT, Kent SBH (1992) Weighing Naked Proteins : Practical , High-Accuracy Mass Measurement of Peptides and Proteins. *Science* **257** (5078): 1885–1894.
221. Siciliano RA, Lippolis R, Mazzeo MF (2019) Proteomics for the Investigation of Surface-Exposed Proteins in Probiotics. *Frontiers in Nutrition* **6** (52): 1–8.
222. Yoshida KI, Takemoto Y, Sotsuka T, Tanaka K, Takenaka S (2013) PhaP phasins play a principal role in poly-β-hydroxybutyrate accumulation in free-living *Bradyrhizobium japonicum*. *BMC Microbiology* **13** (290): 1–10.
223. Quelas JI, Mesa S, Mongiardini EJ, Jendrossek D, Lodeiro AR (2016) Regulation of Polyhydroxybutyrate Synthesis in the Soil Bacterium *Bradyrhizobium diazoefficiens*. *Applied and Environmental Microbiology* **82** (14): 4299–4308.
224. Chakrabarty AM (1998) Nucleoside diphosphate kinase: role in bacterial growth, virulence, cell signalling and polysaccharide synthesis. *Molecular Microbiology* **28** (5): 875–882.
225. Krehenbrink M, Allan JA (2008) Identification of protein secretion systems and novel secreted proteins in *Rhizobium leguminosarum* bv. *viciae*. *BMC Genomics* **9** (55): 1–19.
226. Dubuisson JF, Vianney A, Hugouvieux-Cotte-Pattat N, Lazzaroni JC (2005) Tol-Pal proteins are critical cell envelope components of Erwinia chrysanthemi affecting cell morphology and virulence. *Microbiology* **151** (10): 3337–3347.

227. Dashper SG, Hendtlass A, Slakeski N, Jackson C, Cross KJ, Brownfield L, Hamilton R, Barr I, Reynolds EC (2000) Characterization of a Novel Outer Membrane Hemin-Binding Protein of. *Journal of Bacteriology* **182** (22): 6456–6462.
228. Pride AC, Guan Z, Trent MS (2014) Characterization of the Vibrio cholerae volA surface-exposed lipoprotein lysophospholipase. *Journal of Bacteriology* **196** (8): 1619–1626.
229. d'Enfert C, Ryter A, Pugsley AP (1987) Cloning and expression in *Escherichia coli* of the Klebsiella pneumoniae genes for production, surface localization and secretion of the lipoprotein pullulanase. *The EMBO Journal* **6** (11): 3531–3538.
230. Jin S, Joe A, Lynett J, Hani EK, Sherman P, Chan VL (2001) JlpA, a novel surface-exposed lipoprotein specific to Campylobacter jejuni, mediates adherence to host epithelial cells. *Molecular Microbiology* **39** (5): 1225–1236.
231. Fuchs H, Wallicht R, Simont MM, Kramer MD (1994) The outer surface protein A of the spirochete Borrelia burgdorferi. *Microbiology* **91**: 12594–12598.
232. Indrasumunar A, Gresshoff PM (2013) Vermiculite's strong buffer capacity renders it unsuitable for studies of acidity on soybean (Glycine max L.) nodulation and growth. *BMC Research Notes* **6** (1): 1-8.
233. Westhoek A, Field E, Rehling F, Mulley G, Webb I, Poole PS, Turnbull LA (2017) Policing the legume-*Rhizobium* symbiosis: a critical test of partner choice. *Scientific Reports* **7** (1): 1419.
234. Blair DF, Berg HC (1990) The MotA Protein of E. coli Is a Proton-Conducting Component of the Flagellar Motor. **60**: 439–449.
235. Haiko J, Westerlund-Wikström B (2013) The Role of the Bacterial Flagellum in Adhesion and Virulence. *Biology* **2** (4): 1242–1267.
236. Edwards A, Frederix M, Wisniewski-Dyé F, Jones J, Zorreguieta A, Allan Downie J (2009) The cin and rai quorum-sensing regulatory systems in *Rhizobium leguminosarum* are coordinated by ExpR and CinS, a small regulatory protein coexpressed with CinI. *Journal of Bacteriology* **191** (9): 3059–3067.
237. Nelson RT, Pryor BA, Lodge JK (2003) Sequence length required for homologous recombination in Cryptococcus neoformans. *Fungal Genetics and Biology* **38** (1): 1–9.
238. Wheatley R (2018) Genome-scale characterisation of symbiotic fitness determinants of *Rhizobium leguminosarum* using INSeq. Doctoral Thesis submitted to the University of Oxford.
239. Scheller E V., Cotter PA (2015) Bordetella filamentous hemagglutinin and fimbriae: critical adhesins with unrealized vaccine potential. *Pathogens and disease* **73** (8): 1–9.
240. Gottig N, Garavaglia BS, Garofalo CG, Orellano EG, Ottado J (2009) A filamentous hemagglutinin-like protein of Xanthomonas axonopodis pv. citri, the phytopathogen responsible for citrus canker, is involved in bacterial virulence. *PLoS ONE* **4** (2): 1-13.
241. Molina MA, Ramos JL, Espinosa-Urgel M (2006) A two-partner secretion system is involved in seed and root colonization and iron uptake by *Pseudomonas putida* KT2440. *Environmental Microbiology* **8** (4): 639–647.
242. Pavlova O, Peterson JH, Ieva R, Bernstein HD (2013) Mechanistic link between barrel assembly and the initiation of autotransporter secretion. *Proceedings of the National Academy of Sciences* **110** (10): 938–947.
243. Whittaker CA, Hynes R (2002) Distribution and Evolution of von Willebrand/Integrin A Domains: Widely Dispersed Domains with Roles in Cell Adhesion and Elsewhere. *Molecular Biology of the Cell* **13**: 3369–3387.
244. Vanderlinde EM, Magnus SA, Tambalo DD, Koval SF, Yost CK (2011) Mutation of a broadly conserved operon (RL3499-RL3502) from *Rhizobium leguminosarum* biovar *viciae* causes defects in cell morphology and envelope integrity. *Journal of Bacteriology* **193** (11): 2684–2694.
245. Neudorf KD, Vanderlinde EM, Tambalo DD, Yost CK (2015) A previously

uncharacterized tetratricopeptide-repeat-containing protein is involved in cell envelope function in *Rhizobium leguminosarum*. *Microbiology* **161** (1): 148–157.
246. Link AJ, Church GM (1997) Methods for Generating Precise Deletions and Insertions in the Genome of Wild-Type. *Microbiology* **179** (20): 6228–6237.
247. Kendall K, Roberts AD (2015) Van der Waals forces influencing adhesion of cells. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**: 1–5.
248. Giaouris E, Chapot-Chartier MP, Briandet R (2009) Surface physicochemical analysis of natural Lactococcus lactis strains reveals the existence of hydrophobic and low charged strains with altered adhesive properties. *International Journal of Food Microbiology* **131** (1): 2–9.
249. Krasowska A, Sigler K (2014) How microorganisms use hydrophobicity and what does this mean for human needs? *Frontiers in Cellular and Infection Microbiology* **4** (112): 1–7.
250. Backert S, Fronzes R, Waksman G (2008) VirB2 and VirB5 proteins: specialized adhesins in bacterial type-IV secretion systems? *Trends in Microbiology* **16** (9): 409–413.
251. Karmakar R, Das I, Dutta D, Rakshit A (2016) Potential Effects of Climate Change on Soil Properties: A Review. *Science International* **4** (2): 51–73.
252. Raynaud X, Nunan N (2014) Spatial ecology of bacteria at the microscale in soil. *PLoS ONE* **9** (1): 1-9.
253. García-Palacios P, Maestre FT, Gallardo A (2011) Soil nutrient heterogeneity modulates ecosystem responses to changes in the identity and richness of plant functional groups. *Journal of Ecology* **99** (2): 551–562.
254. García-palacios P, Maestre FT, Bardgett RD, de Kroon H, De H, de Kroon H (2012) Europe PMC Funders Group Plant responses to soil heterogeneity and global environmental change. *The Journal of Ecology* **100** (6): 1303–1314.
255. Xue W, Huang L, Yu FH (2016) Spatial heterogeneity in soil particle size: Does it affect the yield of plant communities with different species richness? *Journal of Plant Ecology* **9** (5): 608–615.
256. Zhang Z, Hu B, Hu G (2014) Spatial heterogeneity of soil chemical properties in a subtropical karst forest, Southwest China. *The Scientific World Journal* **2014**: 1-9.
257. Liu L, Yu S, Xie ZP, Staehelin C (2016) Distance-dependent effects of pathogenic fungi on seedlings of a legume tree: impaired nodule formation and identification of antagonistic rhizosphere bacteria. *Journal of Ecology* **104** (4): 1009–1019.
258. De Boer W, Folman LB, Summerbell RC, Boddy L (2005) Living in a fungal world: Impact of fungi on soil bacterial niche development. *FEMS Microbiology Reviews* **29** (4): 795–811.
259. Whipps JM (2001) Microbial interactions and biocontrol in the rhizosphere. *Journal of Experimental Botany* **52**: 487–511.
260. Frey-Klett P, Burlinson P, Deveau A, Barret M, Tarkka M, Sarniguet A (2011) Bacterial-Fungal Interactions: Hyphens between Agricultural, Clinical, Environmental, and Food Microbiologists. *Microbiology and Molecular Biology Reviews* **75** (4): 583–609.
261. Land M, Hauser L, Jun SR, Nookaew I, Leuze MR, Ahn TH, Karpinets T, Lund O, Kora G, Wassenaar T, Poudel S, Ussery DW (2015) Insights from 20 years of bacterial genome sequencing. *Functional and Integrative Genomics* **15** (2): 141–161.
262. Long JE, DeJesus M, Ward D, Baker RE, Loerger T, Sassetti CM (2015) Identifying Essential Genes in Mycobacterium tuberculosis by Global Phenotypic Profiling. *Methods in Molecular Biology* **1279**: 79–95.
263. Chao MC, Pritchard JR, Zhang YJ, Rubin EJ, Livny J, Davis BM, Waldor MK (2013) High-resolution definition of the Vibrio cholerae essential gene set with hidden Markov model-based analyses of transposon-insertion sequencing data. *Nucleic Acids Research* **41** (19): 9033–9048.
264. Ellison CK (2018) Bacterial adhesion at the single-cell level. *Nature Reviews Microbiology*

16: 614-627.
265. Petrova OE, Sauer K (2012) Sticky situations: Key components that control bacterial surface attachment. *Journal of Bacteriology* **194** (10): 2413–2425.
266. Scheffers DJ, Pinho MG (2005) Bacterial Cell Wall Synthesis: New Insights from Localization Studies Dirk-Jan. *Microbiology and Molecular Biology Reviews* **69** (4): 585–607.
267. Frazier AD, Champney WS (2013) Inhibition of Ribosomal Subunit Synthesis in *Escherichia co*li by the Vanadyl Ribonucleoside Complex Ashley. *Current Microbiology* **67** (2): 226–233.
268. Raetz CRH, Reynolds CM, Trent SM, Bishop RE (2007) Lipid A modification systems in gram-negative bacteria. *Annual Reviews of Biochemistry* **76**: 295–329.
269. Noirot-Gros M-F, Dervyn E, Wu LJ, Mervelet P, Errington J, Ehrlich SD, Noirot P (2002) An expanded view of bacterial DNA replication. *Proceedings of the National Academy of Sciences* **99** (12): 8342–7.
270. Walshaw DL, Poole PS (1996) The general L-amino acid permease of *Rhizobium leguminosarum* is an ABC uptake system that also influences efflux of solutes. *Molecular Microbiology* **21** (6): 1239–1252.
271. Hood G, Ramachandran V, East AK, Downie JA, Poole PS (2017) Manganese transport is essential for N2-fixation by *Rhizobium leguminosarum* in bacteroids from galegoid but not phaseoloid nodules. *Environmental Microbiology* **19** (7): 2715–2726.
272. Riley M (1993) Function of the gene products in *Escherichia coli*. *Microbiology Reviews* **57** (4): 862–952.
273. Zhang YJ, Ioerger TR, Huttenhower C, Long JE, Sassetti CM, Sacchettini JC, Rubin EJ (2012) Global Assessment of Genomic Regions Required for Growth in Mycobacterium tuberculosis. *PLoS Pathogens* **8** (9): 1–9.
274. DeJesus MA, Gerrick ER, Xu W, Park SW, Long JE, Boutte CC, Rubin EJ, Schnappinger D, Ehrt S, Fortune SM, Sassetti CM, Ioerger TR (2017) Comprehensive Essentiality Analysis of the Mycobacterium tuberculosis Genome via Saturating Transposon Mutagenesis . *mBio* **8** (1): 1–17.
275. Lowe DE, Ya J, Glomski IJ (2014) In trans complementation of lethal factor reveal roles in colonization and dissemination in a murine mouse model. *PLoS ONE* **9** (4): 1–12.
276. Caetano-Anolles G, Wall LG, De Micheli AT, Macchi EM, Bauer WD, Favelukes G (1988) Role of Motility and Chemotaxis in Efficiency of Nodulation by *Rhizobium meliloti*. *Plant Physiology* **86** (4): 1228–1235.
277. Hoffmann JH, Linke K, Graf PCF, Lilie H, Jakob U (2004) Identification of a redox-regulated chaperone network. *EMBO Journal* **23** (1): 160–168.
278. Kumar CMS, Mande SC, Mahajan G (2015) Multiple chaperonins in bacteria—novel functions and non-canonical behaviors. *Cell Stress and Chaperones* **20** (4): 555–574.
279. Zhang Y, Cottet SE, Ealick SE (2004) Structure of *Escherichia coli* AMP nucleosidase reveals similarity to nucleoside phosphorylases. *Structure* **12** (8): 1383–1394.
280. McDermott PF, McMurry LM, Podglajen I, Dzink-Fox JL, Schneiders T, Draper MP, Levy SB (2008) The marC gene of *Escherichia coli* is not involved in multiple antibiotic resistance. *Antimicrobial Agents and Chemotherapy* **52** (1): 382–383.
281. Becker P, Hufnagle W, Peters G, Herrmann M, Rda DN a, Atpase C (2001) Detection of Differential Gene Expression in Biofilm-Forming versus Planktonic Populations of Staphylococcus aureus Using Micro-Representational-Difference Analysis. *Applied and Environmental Microbiology* **67** (7): 2958–2965.
282. Pancholi V, Fischetti V a (1998) α -Enolase, a Novel Strong Plasmin(ogen) Binding Protein on the Surface of Pathogenic Streptococci. *The Journal of Biological Chemistry* **273** (23): 14503–14515.
283. Modun B, Williams P (1999) The staphylococcal transferrin-binding protein is a cell wall

glyceraldehyde-3-phosphate dehydrogenase. *Infection and Immunity* **67** (3): 1086–1092.
284. Bao S, Chen D, Yu S, Chen H, Tan L, Hu M, Qiu X, Song C, Ding C (2015) Characterization of triosephosphate isomerase from Mycoplasma gallisepticum. *FEMS Microbiology Letters* **362** (17): 1–9.
285. Cox MM (2003) The Bacterial RecA Protein as a Motor Protein. *Annual Review of Microbiology* **57** (1): 551–577.
286. Kumar KK, Srivastava R, Sinha VB, Michalski J, Kaper JB, Srivastava BS (1994) recA mutations reduce adherence and colonization by classical and El Tor strains of Vibrio cholerae. *Microbiology* **140** (5): 1217–1222.
287. Duan E, Wang Y, Liu L, Zhu J, Zhong M, Zhang H, Li S, Ding B, Zhang X, Guo X, Jiang L, Wan J (2016) Pyrophosphate: fructose-6-phosphate 1-phosphotransferase (PFP) regulates carbon metabolism during grain filling in rice. *Plant Cell Reports* **35** (6): 1321–1331.
288. Janczarek M, Skorupska A (2004) Regulation of pssA and pssB gene expression in *Rhizobium leguminosarum* bv. *trifolii* in response to environmental factors. *International Journal of General and Molecular Microbiology* **85** (3): 217–227.
289. Ivashina T V., Khmelnitsky MI, Shiyapnikov MG, Kanapin AA, Ksenzenko VN (1994) The pss4 gene from *Rhizobium leguminosarum* bv *viciae* VF39: cloning, sequence and the possible role in polysaccharide production and nodule formation. *Gene* **150** (1): 111–116.
290. Oke V, Rushing BG, Fisher EJ, Moghadam-Tabrizi M, Long SR (2001) Identification of the heat-shock sigma factor RpoH and a second RpoH-like protein in *Sinorhizobium meliloti*. *Microbiology* **147** (9): 2399–2408.
291. Makarova KS, Aravind L, Galperin MY, Grishin N V, Tatusov RL, Wolf YI, Koonin E V (1999) Comparative Genomics of the Archaea (Euryarchaeota): Evolution of Conserved Protein Families, the Stable Core, and the Variable Shell. *Genome research* **9** : 608–628.
292. Clissold PM, Ponting CP (2000) PIN domains in nonsense-mediated mRNA decay and RNAi. *Current Biology* **10** (24): 888–890.
293. Domínguez-Ferreras A, Pérez-Arnedo R, Becker A, Olivares J, Soto MJ, Sanjuán J (2006) Transcriptome profiling reveals the importance of plasmid pSymB for osmoadaptation of *Sinorhizobium meliloti*. *Journal of Bacteriology* **188** (21): 7617–7625.
294. Pittman JR, Buntyn JO, Posadas G, Nanduri B, Pendarvis K, Donaldson D (2014) Proteomic Analysis of Cross Protection Provided between Cold and Osmotic Stress in Listeria monocytogenes. *Journal of Proteome Research* **13** (4): 1896–1904.
295. Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS (2003) Comparative Genomics of the Vitamin B12 Metabolism and Regulation in Prokaryotes. *Journal of Biological Chemistry* **278** (42): 41148–41159.
296. Campbell GRO, Taga ME, Mistry K, Lloret J, Anderson PJ, Roth JR, Walker GC (2006) *Sinorhizobium meliloti* bluB is necessary for production of 5,6-dimethylbenzimidazole, the lower ligand of B12. *Proceedings of the National Academy of Sciences* **103** (12): 4634–4639.
297. Taga ME, Walker GC (2010) *Sinorhizobium meliloti* Requires a Cobalamin-Dependent Ribonucleotide Reductase for Symbiosis With Its Plant Host. *Molecular Plant-Microbe Interactions* **23** (12): 1643–1654.
298. Leisico F, Vieira D V., Figueiredo TA, Silva M, Cabrita EJ, Sobral RG, Ludovice AM, Trincão J, Romão MJ, De Lencastre H, Santos-Silva T (2018) First insights of peptidoglycan amidation in Gram-positive bacteria-The high-resolution crystal structure of Staphylococcus aureus glutamine amidotransferase GatD. *Scientific Reports* **8** (1): 1–13.
299. Pedersen LB, Murray T, Popham DL, Setlow P (1998) Characterization of dacC, which encodes a new low-molecular- weight penicillin-binding protein in *Bacillus subtilis*. *Journal of Bacteriology* **180** (18): 4967–4973.
300. Tomlinson AD, Ramey-Hartung B, Day TW, Merritt PM, Fuqua C (2010) Agrobacterium

tumefaciens ExoR represses succinoglycan biosynthesis and is required for biofilm formation and motility. *Microbiology* **156** (9): 2670–2681.
301. Ghosh PK, Maiti TK (2016) Structure of Extracellular Polysaccharides (EPS) Produced by Rhizobia and their Functions in Legume–Bacteria Symbiosis: — A Review. *Achievements in the Life Sciences* **10** (2): 136–143.
302. Douglas CJ, Halperin W, Nester EW (1982) Agrobacterium tumefaciens mutants affected in attachment to plant cells. *Journal of Bacteriology* **152** (3): 1265–1275.
303. Douglas CJ, Staneloni RJ, Rubin RA, Nester EW (1985) Identification and genetic analysis of an Agrobacterium tumefaciens chromosomal virulence region. *Journal of Bacteriology* **161** (3): 850–860.
304. Puvanesarajah V, Schell FM, Stacey G, Douglas CJ, Nester EW (1985) Role for 2-linked-β-D-glucan in the virulence of Agrobacterium tumefaciens. *Journal of Bacteriology* **164** (1): 102–106.
305. Pei B, Sisu C, Frankish A, Howald C, Habegger L, Mu XJ, Harte R, Balasubramanian S, Tanzer A, Diekhans M, Reymond A, Hubbard TJ, Harrow J, Gerstein MB (2012) The GENCODE pseudogene resource. *Genome Biology* **13** (R51): 1–26.
306. Tsai TI, Li ST, Liu CP, Chen KY, Shivatare SS, Lin CW, Liao SF, Lin CW, Hsu TL, Wu YT, Tsai MH, Lai MY, Lin NH, Wu CY, Wong CH (2017) An effective bacterial fucosidase for glycoprotein remodeling. *ACS Chemical Biology* **12** (1): 63–72.
307. Gharzouli R, Carpéné MA, Couderc F, Benguedouar A, Poinsot V (2013) Relevance of fucose-rich extracellular polysaccharides produced by *Rhizobium sullae* strains nodulating Hedysarum coronarium L. legumes. *Applied and Environmental Microbiology* **79** (6): 1764–1776.
308. Mintz KP, Fives-Taylor PM (2000) impA, a gene coding for an inner membrane protein, influences colonial morphology of *Actinobacillus actinomycetemcomitans*. *Infection and Immunity* **68** (12): 6580–6586.
309. Pukatzki S, Ma AT, Sturtevant D, Krastins B, Sarracino D, Nelson WC, Heidelberg JF, Mekalanos JJ (2005) Identification of a conserved bacterial protein secretion system in Vibrio cholerae using the Dictyostelium host model system. *Proceedings of the National Academy of Sciences* **103** (5): 1528–1533.
310. Sahu SK, Zheng P, Yao N (2018) Niclosamide Blocks Rice Leaf Blight by Inhibiting Biofilm Formation of Xanthomonas oryzae. *Frontiers in Plant Science* **9** : 1–16.
311. Dozot M, Poncet S, Nicolas C, Copin R, Bouraoui H, Mazé A, Deutscher J, de Bolle X, Letesson JJ (2010) Functional characterization of the incomplete phosphotransferase system (PTS) of the intracellular pathogen brucella melitensis. *PLoS ONE* **5** (9): 1–16.
312. Pflüger-Grau K, Görke B (2010) Regulatory roles of the bacterial nitrogen-related phosphotransferase system. *Trends in Microbiology* **18** (5): 205–214.
313. Untiet V, Karunakaran R, Krämer M, Poole P, Priefer U, Prell J (2013) ABC Transport Is Inactivated by the PTSNtr under Potassium Limitation in *Rhizobium leguminosarum* 3841. *PLoS ONE* **8** (5): 2–10.
314. Crépin S, Ottosen EN, Peters K, Smith SN, Himpsl SD, Vollmer W, Mobley HLT (2018) The lytic transglycosylase MltB connects membrane homeostasis and in vivo fitness of Acinetobacter baumannii. *Molecular Microbiology* **109** (6): 745–762.
315. Monteiro C, Fang X, Ahmad I, Gomelsky M, Römling U (2011) Regulation of biofilm components in Salmonella enterica serovar typhimurium by lytic transglycosylases involved in cell wall turnover. *Journal of Bacteriology* **193** (23): 6443–6451.
316. Niba ETE, Naka Y, Nagase M, Mori H, Kitakawa M (2007) A Genome-wide Approach to Identify the Genes Involved in Biofilm Formation in E. coli. *DNA Research* **14** (6): 237–246.
317. Scheurwater E, Reid CW, Clarke AJ (2008) Lytic transglycosylases: Bacterial space-making autolysins. *International Journal of Biochemistry and Cell Biology* **40** (4): 586–

591.
318. Neudorf KD, Yost CK (2017) An uncharacterized gene coding a conserved lytic transglycosylase domain (RL4716) is required for proper cell envelope function in *Rhizobium leguminosarum*. *FEMS Microbiology Letters* **364** (5): 1–7.
319. Hamada M, Toyofuku M, Miyano T, Nomura N (2014) cbb3-type cytochrome c oxidases, aerobic respiratory enzymes, impact the anaerobic life of *Pseudomonas aeruginosa* PAO1. *Journal of Bacteriology* **196** (22): 3881–3889.
320. Dey S, Biswas C, Sengupta J (2018) The universally conserved GTPase HflX is an RNA helicase that restores heat-damaged *Escherichia coli* ribosomes. *Journal of Cell Biology* **217** (7): 2519–2529.
321. Missiakas D, Schwager F, Raina S (1995) Identification and characterization of a new disulfide isomerase-like protein (DsbD) in *Escherichia coli*. *The EMBO Journal* **14** (14): 3415–3424.
322. Plate L, Marletta MA (2012) Nitric Oxide Modulates Bacterial Biofilm Formation through a Multicomponent Cyclic-di-GMP Signaling Network. *Molecular Cell* **46** (4): 449–460.
323. Modzelan M, Kujawa M, Głabski K, Jagura-Burdzy G, Kraszewska E (2014) NudC Nudix hydrolase from *Pseudomonas syringae*, but not its counterpart from *Pseudomonas aeruginosa*, is a novel regulator of intracellular redox balance required for growth, motility and biofilm formation. *Molecular Microbiology* **93** (5): 867–882.
324. Wahl A, My L, Dumoulin R, Sturgis JN, Bouveret E (2011) Antagonistic regulation of dgkA and plsB genes of phospholipid synthesis by multiple stress responses in *Escherichia coli*. *Molecular Microbiology* **80** (5): 1260–1275.
325. Jerga A, Lu Y-J, Schujman GE, De Mendoza D, Rock CO (2007) Identification of a Soluble Diacylglycerol Kinase Required for Lipoteichoic Acid Production in *Bacillus subtilis*. *The Journal of Biological Chemistry* **282** (30): 21738–21745.
326. Kuznetsova E, Nocek B, Brown G, Makarova KS, Flick R, Wolf YI, Khusnutdinova A, Evdokimova E, Jin K, Tan K, Hanson AD, Hasnain G, Zallot R, De Crécy-Lagard V, Babu M, Savchenko A, Joachimiak A, Edwards AM, Koonin E V., et al. (2015) Functional diversity of haloacid dehalogenase superfamily phosphatases from Saccharomyces cerevisiae: Biochemical, structural, and evolutionary insights. *Journal of Biological Chemistry* **290** (30): 18678–18698.
327. Li J, Wang N (2011) Genome-wide mutagenesis of xanthomonas axonopodis pv. citri reveals novel genetic determinants and regulation mechanisms of biofilm formation. *PLoS ONE* **6** (7): 1-16.
328. Anderson BE, Baumstark BR, Bellini WJ (1988) Expression of the gene encoding the 17-kilodalton antigen from Rickettsia rickettsii: Transcription and posttranslational modification. *Journal of Bacteriology* **170** (10): 4493–4500.
329. Darwin AJ (2013) Stress Relief during Host Infection: The Phage Shock Protein Response Supports Bacterial Virulence in Various Ways. *PLoS Pathogens* **9** (7): 1-4.
330. Wang Q, Frye JG, McClelland M, Harshey RM (2004) Gene expression patterns during swarming in Salmonella typhimurium : genes specific to surface growth and putative new motility and pathogenicity genes. *Molecular Microbiology* **52** (1): 169–187.
331. Uehara T, Suefuji K, Valbuena N, Donegan M, Park JT (2005) Recycling of the Anhydro-N-Acetylmuramic Acid Derived from Cell Wall Murein Involves a Two-Step Conversion to N-Acetylglucosamine-Phosphate. *Journal of Bacteriology* **187** (11): 3643–3649.
332. Uehara T, Suefuji K, Jaeger T, Mayer C, Park JT (2006) MurQ Etherase Is Required by *Escherichia coli* in Order To Metabolize Anhydro-N-Acetylmuramic Acid Obtained either from the Environment or from Its Own Cell Wall. *Journal of Bacteriology* **188** (4): 1660–1662.
333. Singla-Pareek SL, Reddy MK, Sopory SK (2003) Genetic engineering of the glyoxalase pathway in tobacco leads to enhanced salinity tolerance. *Proceedings of the National*

Academy of Sciences **100** (25): 14672–14677.
334. Yeats C, Rawlings ND, Bateman A (2004) The PepSY domain: A regulator of peptidase activity in the microbial environment? *Trends in Biochemical Sciences* **29** (4): 169–172.
335. Iijima S, Washio K, Okahara R, Morikawa M (2009) Biofilm formation and proteolytic activities of *Pseudoalteromonas* bacteria that were isolated from fish farm sediments. *Microbial Biotechnology* **2** (3): 361–369.
336. Minder AC, Rudder KEE De, Narberhaus F, Fischer H, Hennecke H, Geiger O (2001) Phosphatidylcholine levels in *Bradyrhizobium japonicum* membranes are critical for an efficient symbiosis with the soybean host plant. *Molecular Microbiology* **39** (5): 1186–1198.
337. Mathew R, Mukherjee R, Balachandar R, Chatterji D (2006) Deletion of the rpoZ gene , encoding the v subunit of RNA polymerase, results in pleiotropic surface-related phenotypes in Mycobacterium smegmatis. *Microbiology* **152**: 1741–1750.
338. Bhardwaj N, Syal K, Chatterji D (2018) The role of ω- subunit of *Escherichia coli* RNA polymerase in stress response. *Genes to Cells* **23** : 357–369.
339. Weiss A, Moore BD, Tremblay MHJ, Chaput D, Kremer A, Shaw LN (2017) The Omega Subunit Governs RNA Polymerase Stability and Transcriptional Specificity in Staphylococcus aureus. *Journal of Bacteriology* **199** (2): 1–16.
340. Hu C, Xiong N, Zhang Y, Rayner S, Chen S (2012) Functional characterization of lipase in the pathogenesis of Staphylococcus aureus. *Biochemical and Biophysical Research Communications* **419** (4): 617–620.
341. Heijenoort J Van (2001) Formation of the glycan chains in the synthesis of bacterial peptidoglycan 1. *Glycobiology* **11** (3): 25–36.
342. Takahashi Y, Takechi K, Takio S, Takano H (2016) Both the transglycosylase and transpeptidase functions in plastid penicillin-binding protein are essential for plastid division in Physcomitrella patens. *Proceedings of the Japanese Academy* **92** (10): 499–508.
343. Schulte JE, Goulian M (2018) The Phosphohistidine Phosphatase SixA Targets a Phosphotransferase System. *mBio* **9** (6): 1–13.
344. Beloin C, Valle J, Latour-lambert P, Faure P, Kzreminski M, Balestrino D, Haagensen JAJ, Molin S, Prensier G, Arbeille B, Ghigo J (2004) Global impact of mature biofilm lifestyle on *Escherichia coli* K-12 gene expression. *Molecular Microbiology* **51** (3): 659–674.
345. Yepes A, Schneider J, Mielich B, Koch G, Ramamurthi KS, Vlamakis H, López D (2012) The biofilm formation defect of a *Bacillus subtilis* flotillin- defective mutant involves the protease FtsH. *Molecular Microbiology* **86** (2): 457–471.
346. Skagia A, Zografou C, Vezyri E, Venieraki A, Katinakis P, Dimou M (2016) Cyclophilin PpiB is involved in motility and biofilm formation via its functional association with certain proteins. *Genes to Cells* **21** : 833–851.
347. Schumann W (1999) FtsH - a single-chain charonin? *FEMS Microbiology Reviews* **23**: 1–11.
348. Kihara A, Akiyama Y, Ito K (1996) A protease complex in the *Escherichia coli* plasma membrane : HflKC (HflA) forms a complex with FtsH (HflB), regulating its proteolytic activity against SecY. *The EMBO journal* **15** (22): 6122–6131.
349. Hinderhofer M, Walker CA, Friemel A, Stuermer CAO, Möller HM, Reuter A (2009) Evolution of prokaryotic SPFH proteins. *BMC Evolutionary Biology* **9** (10): 1–18.
350. Vedam V, Haynes JG, Kannenberg EL, Carlson RW, Sherrier DJ (2004) A *Rhizobium leguminosarum* Lipopolysaccharide Lipid-A Mutant Induces Nitrogen-Fixing Nodules with Delayed and Defective Bacteroid Formation. *Molecular Plant-Microbe Interactions* **17** (3): 283–291.
351. Harrison JJ, Vanderlinde EM, Muszyn A, Koval SF, Foreman DL, Ceri H, Kannenberg EL, Carlson RW, Yost CK (2009) *Rhizobium leguminosarum* biovar *viciae* 3841, deficient in 27-hydroxyoctacosanoate-modified lipopolysaccharide, is impaired in desiccation tolerance

, biofilm formation and motility. *Microbiology* **155** : 3055–3069.
352. Pinedo CA, Bringhurst RM, Gage DJ (2008) *Sinorhizobium meliloti* mutants lacking phosphotransferase system enzyme HPr or EIIA are altered in diverse processes, including carbon metabolism, cobalt requirements, and succinoglycan production. *Journal of Bacteriology* **190** (8): 2947–2956.
353. Cheng HP, Walker GC (1998) Succinoglycan Is Required for Initiation and Elongation of Infection Threads during Nodulation of Alfalfa by *Rhizobium meliloti*. *Journal of Bacteriology* **180** (19): 5183–5191.
354. González JE, Reuhs BL, Walker GC (1996) Low molecular weight EPS II of *Rhizobium meliloti* allows nodule invasion in *Medicago sativa*. *Proceedings of the National Academy of Sciences* **93** (16): 8636–41.
355. Leigh JA, Signer ER, Walker GC (1985) Exopolysaccharide-deficient mutants of *Rhizobium meliloti* that form ineffective nodules. *Proceedings of the National Academy of Sciences* **82** (18): 6231–6235.
356. Yao SY, Luo L, Har KJ, Becker A, Rüberg S, Yu GQ, Zhu JB, Cheng HP (2004) *Sinorhizobium meliloti* ExoR and ExoS proteins regulate both succinoglycan and flagellum production. *Journal of Bacteriology* **186** (18): 6042–6049.
357. Niemeyer D, Becker A (2001) The Molecular Weight Distribution of Succinoglycan Produced by *Sinorhizobium meliloti* is Influenced by Specific tyrosine phosphorylation and ATPase activity of the cytoplasmic domain of the ExoP protein. *Journal of Bacteriology* **183** (17): 5163–5170.
358. Bahat-Samet E, Castro-sowinski S, Okon Y (2004) Arabinose content of extracellular polysaccharide plays a role in cell aggregation of Azospirillum brasilense. *FEMS Microbiology Letters* **237** (237): 195–203.
359. Zeng F, Jin W, Zhao Q (2019) Temperature effect on extracellular polymeric substances (EPS) and phosphorus accumulating organisms (PAOs) for phosphorus release of anaerobic sludge. *RSC Advances* **9** : 2162–2171.
360. Torres-Quesada O, Oruezabal RI, Peregrina A, Jofré E, Lloret J, Rivilla R, Toro N, Jiménez-zurdo JI (2010) The *Sinorhizobium meliloti* RNA chaperone Hfq influences central carbon metabolism and the symbiotic interaction with alfalfa. *BMC Microbiology* **10** (71): 1–20.
361. Narita S (2011) ABC Transporters Involved in the Biogenesis of the Outer Membrane in Gram-Negative Bacteria. *Bioscience, Biotechnology and Biochemistry* **75** (6): 1044–1054.
362. Doerrler WT, Sikdar R, Kumar S, Boughner LA (2013) New functions for the ancient DedA membrane protein family. *Journal of Bacteriology* **195** (1): 3–11.
363. Anantharaman V, Aravind L (2006) The NYN domains: Novel predicted RNAses with a PIN domain-like fold. *RNA Biology* **3** (1): 18–27.
364. Dougan DA, Reid BG, Horwich AL, Bukau B (2002) ClpS, a substrate modulator of the ClpAP machine. *Molecular Cell* **9** (3): 673–683.
365. Fernández L, Breidenstein EBM, Song D, Hancock REW (2012) Role of intracellular proteases in the antibiotic resistance, motility, and biofilm formation of *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy* **56** (2): 1128–1132.
366. Henderson B, Fares MA, Lund PA (2013) Chaperonin 60: A paradoxical, evolutionarily conserved protein family with multiple moonlighting functions. *Biological Reviews* **88** (4): 955–987.
367. Ojha A, Anand M, Bhatt A, Kremer L, Jacobs WR, Hatfull GF (2005) GroEL1: A dedicated chaperone involved in mycolic acid biosynthesis during biofilm formation in mycobacteria. *Cell* **123** (5): 861–873.
368. Castanié-Cornet MP, Bruel N, Genevaux P (2014) Chaperone networking facilitates protein targeting to the bacterial cytoplasmic membrane. *Biochimica et Biophysica Acta - Molecular Cell Research* **1843** (8): 1442–1456.

369. Yang CK, Ewis HE, Zhang XZ, Lu CD, Hu HJ, Pan Y, Abdelal AT, Tai PC (2011) Nonclassical protein secretion by *Bacillus subtilis* in the stationary phase is not due to cell lysis. *Journal of Bacteriology* **193** (20): 5607–5615.
370. Kobayashi K (2007) Gradual activation of the response regulator DegU controls serial expression of genes for flagellum formation and biofilm formation in *Bacillus subtilis*. *Molecular Microbiology* **66** (2): 395–409.
371. Davies C, White SW, Nicholas RA (2001) Crystal structure of a deacylation-defective mutant of penicillin-binding protein 5 at 2.3-Å resolution. *Journal of Biological Chemistry* **276** (1): 616–623.
372. Ramos J, Bisseling T (2003) A method for the isolation of root hairs from the model legume *Medicago truncatula*. *Journal of Experimental Botany* **54** (391): 2245–2250.
373. Qiao Z, Libault M (2013) Unleashing the potential of the root hair cell as a single plant cell type model in root systems biology. *Frontiers in Plant Science* **4**: 1–8.
374. Arfaoui A, Sifi B, Boudabous A, Hadrami I, Cherif M (2006) Identification of *Rhizobium* isolates possessing antagonistic activity against *Fusarium oxysporum* F.S.P *ciceris*, the causal agent of Fusarium wilt in chickpea. *Journal of Plant Pathology* **88** (1): 67–75.
375. Bais HP, Fall R, Vivanco JM (2004) Biocontrol of *Bacillus subtilis* against Infection of Arabidopsis Roots by *Pseudomonas syringae* Is Facilitated by Biofilm Formation and Surfactin Production. *Plant Physiology* **134**: 307–319.
376. Marczak M, Mazur A, Koper P, Żebracki K, Skorupska A (2017) Synthesis of rhizobial exopolysaccharides and their importance for symbiosis with legume plants. *Genes* **8** (360): 1–24.
377. Albareda M, Dardanelli MS, Sousa C, Megías M, Temprano F, Rodríguez-Navarro DN (2006) Factors affecting the attachment of rhizospheric bacteria to bean and soybean roots. *FEMS Microbiology Letters* **259** (1): 67–73.
378. Villacieros M, Power B, Sánchez-Contreras M, Lloret J, Oruezabal RI, Martín M, Fernández-Piñas F, Bonilla I, Whelan C, Dowling DN, Rivilla R (2003) Colonization behaviour of *Pseudomonas fluorescens* and *Sinorhizobium meliloti* in the alfalfa (*Medicago sativa*) rhizosphere. *Plant and Soil* **251** (1): 47–54.
379. Lu J, Yang F, Wang S, Ma H, Liang J, Chen Y (2017) Co-existence of rhizobia and diverse non-rhizobial bacteria in the rhizosphere and nodules of *Dalbergia odorifera* seedlings inoculated with *Bradyrhizobium elkanii*, *Rhizobium multihospitium*-Like and *Burkholderia pyrrocinia*-like strains. *Frontiers in Microbiology* **8** (2255): 1–11.
380. Hemmat Jou MH, Besalatpour AA (2018) Interactive effects of co-inoculation of *Bradyrhizobium japonicum* strains and mycorrhiza species on soybean growth and nutrient contents in plant. *Journal of Plant Nutrition* **41** (1): 10–18.
381. Crossman LC, Castillo-Ramirez S, McAnnula C, Lozano L, Vernikos GS, Acosta JL, Ghazoui ZF, Hernández-González I, Meakin G, Walker AW, Hynes MF, Young JPW, Downie JA, Romero D, Johnston AWB, Dávila G, Parkhill J, González V (2008) A common genomic framework for a diverse assembly of plasmids in the symbiotic nitrogen fixing bacteria. *PLoS ONE* **3** (7): 1–9.
382. Tomich M, Planet PJ, Figurski DH (2007) The tad locus: Postcards from the widespread colonization island. *Nature Reviews Microbiology* **5** (5): 363–375.
383. Zatakia HM, Nelson CE, Syed UJ, Scharf BE (2014) ExpR Coordinates the Expression of Symbiotically Important, Bundle-Forming Flp Pili with Quorum Sensing in *Sinorhizobium meliloti*. *Applied and Environmental Microbiology* **80** (8): 2429–2439.
384. Nykyri J, Mattinen L, Niemi O, Adhikari S, Kõiv V, Somervuo P, Fang X, Auvinen P, Mäe A, Palva ET, Pirhonen M (2013) Role and Regulation of the Flp/Tad Pilus in the Virulence of *Pectobacterium atrosepticum* SCRI1043 and *Pectobacterium wasabiae* SCC3193. *PLoS ONE* **8** (9): 1–11.
385. Pu M, Rowe-Magnus DA (2018) A Tad pilus promotes the establishment and resistance of

Vibrio vulnificus biofilms to mechanical clearance. *npj Biofilms and Microbiomes* **4** (1): 1–8.
386. Christie PJ, Whitaker N, González-Rivera C (2014) Mechanism and structure of the bacterial type IV secretion systems. *Biochimica et Biophysica Acta* **1843** (8): 1578–1591.
387. Haase J, Lurz R, Grahn AM, Bamford, DH, Lanka E (1995) Bacterial conjugation mediated by plasmid RP4: RSF1010 mobilization, donor-specific phage propagation, and pilus production require the same Tra2 core components of a proposed DNA transport complex. *Journal of bacteriology* **177** (16): 4779–4791.
388. Aguilar J, Cameron TA, Zupan J, Zambryski P (2011) Membrane and core periplasmic Agrobacterium tumefaciens virulence type IV secretion system components localize to multiple sites around the bacterial perimeter during lateral attachment to plant cells. *mBio* **2** (6): 1–8.
389. Jakubowski SJ, Cascales E, Krishnamoorthy V, Christie PJ (2005) Component of a Type IV Secretion System , Regulates Substrate Selection and T-Pilus Biogenesis. *Journal of bacteriology* **187** (10): 3486–3495.
390. Bernard CS, Bordi C, Termine E, Filloux A, De Bentzmann S (2009) Organization and PprB-dependent control of the *Pseudomonas aeruginosa* tad locus, involved in Flp Pilus Biology. *Journal of Bacteriology* **191** (6): 1961–1973.
391. Clock SA, Planet PJ, Perez BA, Figurski DH (2008) Outer membrane components of the tad (tight adherence) secreton of *Aggregatibacter actinomycetemcomitans*. *Journal of Bacteriology* **190** (3): 980–990.
392. Stanger A (2008) Analysis of the ECF sigma factors of *Rhizobium leguminosarum* biovar *viciae* 3941. Doctoral Thesis submitted to the Department of Molecular Microbiology, John Innes Centre (University of East Anglia), Norwich.
393. Díaz-Mireles E, Wexler M, Sawers G, Bellini D, Todd JD, Johnston AWB (2004) The Fur-like protein Mur of *Rhizobium leguminosarum* is a Mn2+ - responsive transcriptional regulator. *Microbiology* **150** (5): 1447–1456.
394. Lodwig E, Kumar S, Allaway D, Bourdes A, Prell J, Priefer U, Poole P (2004) Regulation of L-Alanine Dehydrogenase in *Rhizobium leguminosarum* bv. *viciae* and Its Role in Pea Nodules. *Journal of Bacteriology* **186** (3): 842–849.
395. Casteriano A, Wilkes MA, Deaker R (2013) Physiological Changes in Rhizobia after Growth in Peat Extract May Be Related to Improved Desiccation Tolerance. *Applied and Environmental Microbiology* **79** (13): 3998–4007.
396. Vanderlinde EM, Harrison JJ, Muszyński A, Carlson RW, Turner RJ, Yost CK (2010) Identification of a novel ABC transporter required for desiccation tolerance, and biofilm formation in *Rhizobium leguminosarum* bv. *viciae* 3841. *FEMS Microbiology Ecology* **71** (3): 327–340.
397. Sourjik V, Schmitt R (1996) Different roles of CheY1 and CheY2 in the chemotaxis of *Rhizobium meliloti*. *Molecular Microbiology* **22** (3): 427–436.
398. Steinauer K, Chatzinotas A, Eisenhauer N (2016) Root exudate cocktails: the link between plant diversity and soil microorganisms? *Ecology and Evolution* **6** (20): 7387–7396.
399. Cai T, Cai W, Zhang J, Zheng H, Tsou AM, Xiao L, Zhong Z, Zhu J (2009) Host legume-exuded antimetabolites optimize the symbiotic rhizosphere. *Molecular Microbiology* **73** (3): 507–517.
400. Vos K, Braeken K, Fauvart M, Ndayizeye M, Verhaert J, Zachurzok S, Lambrichts I, Michiels J (2007) The *Rhizobium etli* opt operon is required for symbiosis and stress resistance. *Environmental Microbiology* **9** (7): 1665–1674.
401. Mengin-Lecreulx D, Texier L, Rousseau M, van Heijenoort J (1991) The murG gene of *Escherichia coli* codes for the UDP-N-acetylglucosamine: N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase involved in the membrane steps of peptidoglycan synthesis. *Journal of bacteriology* **173** (15): 4625–36.

402. Slámová K, Bojarová P, Petrásková L, Křen V (2010) B-N-Acetylhexosaminidase: What's in a name...? *Biotechnology Advances* **28** (6): 682–693.
403. Paddick JS, Brailsford SR, Rao S, Soares RF, Kidd EAM, Beighton D, Homer KA (2006) Effect of biofilm growth on expression of surface proteins of *Actinomyces naeslundii* genospecies 2. *Applied and Environmental Microbiology* **72** (5): 3774–3779.
404. Arrecubieta C, Matsunaga I, Asai T, Naka Y, Deng MC, Lowy FD (2008) Vaccination with Clumping Factor A and Fibronectin Binding Protein A to Prevent Staphylococcus aureus Infection of an Aortic Patch in Mice . *The Journal of Infectious Diseases* **198** (4): 571–575.
405. Silva RC, Padovan ACB, Pimenta DC, Ferreira RC, da Silva C V., Briones MRS (2014) Extracellular enolase of Candida albicans is involved in colonization of mammalian intestinal epithelium. *Frontiers in Cellular and Infection Microbiology* **4**: 1–9.
406. Parales RE, Luu RA, Chen GY, Liu X, Wu V, Lin P, Hughes JG, Nesteryuk V, Parales J V., Ditty JL (2013) *Pseudomonas putida* F1 has multiple chemoreceptors with overlapping specificity for organic acids. *Microbiology* **159**: 1086–1096.
407. Rengel Z, Marschner P (2005) Nutrient availability and management in the rhizosphere: Exploiting genotypic differences. *New Phytologist* **168** (2): 305–312.
408. Rudrappa T, Czymmek KJ, Pare PW, Bais HP (2008) Root-Secreted Malic Acid Recruits Beneficial Soil Bacteria. *Plant Physiology* **148** (3): 1547–1556.
409. Chen Y, Shugeng C, Yunrong C, Jon C, Roberto K, Jian-hua G, Losick R (2012) A *Bacillus subtilis* Sensor Kinase Involved in Triggering Biofilm Formation on the Roots of Tomato Plants. *Molecular Microbiology* **85** (3): 418–430.
410. Robichon C, Vidal-Ingigliardi D, Pugsley AP (2005) Depletion of apolipoprotein N-acyltransferase causes mislocalization of outer membrane lipoproteins in *Escherichia coli*. *Journal of Biological Chemistry* **280** (2): 974–983.
411. Ivashina T V., Ksenzenko VN (2012) Exopolysaccharide Biosynthesis in *Rhizobium leguminosarum*: From Genes to Functions. In *The Complex World of Polysaccharides*, Karunaratne, DN (eds), Intech, pp 99–126.
412. Muff TJ, Foster RM, Liu PJY, Ordal GW (2007) CheX in the three-phosphatase system of bacterial chemotaxis. *Journal of Bacteriology* **189** (19): 7007–7013.
413. Miller LD (2007) Characterization of the Chemotaxis System of the Endosymbiotic Bacterium *Rhizobium leguminosarum* bv . *viciae* Strain 3841. Doctoral thesis submitted to the Georgia Institute of Technology.
414. Meier VM, Muschler P, Scharf BE (2007) Functional analysis of nine putative chemoreceptor proteins in *Sinorhizobium meliloti*. *Journal of Bacteriology* **189** (5): 1816–1826.
415. Vemulapalli TH, Vemulapalli R, Schurig GG, Boyle SM, Sriranganathan N (2006) Role in virulence of a *Brucella abortus* protein exhibiting lectin-like activity. *Infection and Immunity* **74** (1): 183–191.
416. Sauvage E, Kerff F, Terrak M, Ayala JA, Charlier P (2008) The penicillin-binding proteins: Structure and role in peptidoglycan biosynthesis. *FEMS Microbiology Reviews* **32** (2): 234–258.
417. Rorbach J, Gammage PA, Minczuk M (2012) C7orf30 is necessary for biogenesis of the large subunit of the mitochondrial ribosome. *Nucleic Acids Research* **40** (9): 4097–4109.
418. Zhang Y-M, Rock CO (2008) Thematic Review Series: Glycerolipids. Acyltransferases in bacterial glycerophospholipid synthesis. *Journal of Lipid Research* **49** (9): 1867–1874.
419. Russo DM, Abdian PL, Posadas DM, Williams A, Vozza N, Giordano W, Kannenberg E, Allan Downie J, Zorreguietaa A (2015) Lipopolysaccharide O-chain core region required for cellular cohesion and compaction of in vitro and root biofilms developed by *Rhizobium leguminosarum*. *Applied and Environmental Microbiology* **81** (3): 1013–1023.
420. Vrljic M, Sahm H, Eggeling L (1996) A new type of transporter with a new type of cellular

function: L-lysine export from Corynebacterium glutamicum. *Molecular Microbiology* **22** (5): 815–826.
421. Tambalo DD, Bustard DE, Del Bel KL, Koval SF, Khan MF, Hynes MF (2010) Characterization and functional analysis of seven flagellin genes in *Rhizobium leguminosarum* bv. *viciae*. Characterization of *R. leguminosarum* flagellins. *BMC Microbiology* **10** (219): 1–16.
422. Mesnage S, Dellarole M, Baxter NJ, Rouget JB, Dimitrov JD, Wang N, Fujimoto Y, Hounslow AM, Lacroix-Desmazes S, Fukase K, Foster SJ, Williamson MP (2014) Molecular basis for bacterial peptidoglycan recognition by LysM domains. *Nature Communications* **5** (4269): 1–11.
423. Baev N, Schultze M, Barlier I, Dang Cam Ha, Virelizier H, Kondorosi E, Kondorosi A (1992) *Rhizobium* nodM and nodN genes are common nod genes: nodM encodes functions for efficiency of nod signal production and bacteroid maturation. *Journal of Bacteriology* **174** (23): 7555–7565.
424. Fujishige NA, Lum MR, De Hoff PL, Whitelegge JP, Faull KF, Hirsch AM (2008) *Rhizobium* common nod genes are required for biofilm formation. *Molecular Microbiology* **67** (3): 504–515.
425. Singh P, Brooks JF, Ray VA, Mandel MJ, Visick KL (2015) CysK Plays a Role in Biofilm Formation and Colonization by Vibrio fischeri. *Applied and Environmental Microbiology* **81** (15): 5223–5234.
426. Pereira SB, Mota R, Vieira CP, Vieira J, Tamagnini P (2015) Phylum-wide analysis of genes/proteins related to the last steps of assembly and export of extracellular polymeric substances (EPS) in cyanobacteria. *Scientific Reports* **5**: 1–16.
427. Popowska M, Osińska M, Rzeczkowska M (2012) N-acetylglucosamine-6-phosphate deacetylase (NagA) of *Listeria monocytogene*s EGD, an essential enzyme for the metabolism and recycling of amino sugars. *Archives of Microbiology* **194** (4): 255–268.
428. Ausmees N, Jonsson H, Höglund S, Ljunggren H, Lindberg M (1999) Structural and putative regulatory genes involved in cellulose synthesis in *Rhizobium leguminosarum* bv. *trifolii*. *Microbiology* **145** (5): 1253–1262.
429. Roche B, Aussel L, Ezraty B, Mandin P, Py B, Barras F (2013) Iron/sulfur proteins biogenesis in prokaryotes: Formation, regulation and diversity. *Biochimica et Biophysica Acta - Bioenergetics* **1827** (8–9): 923–937.
430. Altindis E, Fu Y, Mekalanos JJ (2014) Proteomic analysis of *Vibrio cholerae* outer membrane vesicles. *Proceedings of the National Academy of Sciences* **111** (15): 1548–1556.
431. Berk V, Fong JCN, Dempsey GT, Develioglu ON, Zhuang X, Liphardt J, Yildiz FH, Chu S (2012) Molecular Architecture and Assembly Principles of *Vibrio cholerae* Biofilms. *Science* **337** (6091): 236–239.
432. Zimaro T, Thomas L, Marondedze C, Garavaglia BS, Gehring C, Ottado J, Gottig N (2013) Insights into *Xanthomonas axonopodis* pv. *citri* biofilm through proteomics. *BMC Microbiology* **13** (186): 1–14.
433. Smith SGJ, Mahon V, Lambert MA, Fagan RP (2007) A molecular Swiss army knife: OmpA structure, function and expression. *FEMS Microbiology Letters* **273** (1): 1–11.
434. Loessner MJ, Wendlinger G, Scherer S (1995) Heterogeneous endolysins in Listeria monocytogenes bacteriophages: a new class of enzymes and evidence for conserved holin genes within the siphoviral lysis cassettes. *Molecular Microbiology* **16** (6): 1231–1241.
435. Teplitski M, Goodier RI, Ahmer BMM (2003) Pathways Leading from BarA/SirA to Motility and Virulence Gene Expression in *Salmonella*. *Journal of Bacteriology* **185** (24): 7257–7265.
436. Whiteley CG, Lee DJ (2015) Bacterial diguanylate cyclases: Structure, function and mechanism in exopolysaccharide biofilm development. *Biotechnology Advances* **33** (1):

124–141.
437. Barnhart DM, Su S, Farrand SK (2014) A signaling pathway involving the diguanylate cyclase CelR and the response regulator DivK controls cellulose synthesis in *Agrobacterium tumefaciens*. *Journal of Bacteriology* **196** (6): 1257–1274.
438. Pérez-Mendoza D, Aragón IM, Prada-Ramírez HA, Romero-Jiménez L, Ramos C, Gallegos MT, Sanjuán J (2014) Responses to elevated c-di-GMP levels in mutualistic and pathogenic plant-interacting bacteria. *PLoS ONE* **9** (3): 1–13.
439. Gjermansen M, Ragas P, Tolker-Nielsen T (2006) Proteins with GGDEF and EAL domains regulate *Pseudomonas putida* biofilm formation and dispersal. *FEMS Microbiology Letters* **265** (2): 215–224.
440. Bobrov AG, Kirillina O, Perry RD (2005) The phosphodiesterase activity of the HmsP EAL domain is required for negative regulation of biofilm formation in *Yersinia pestis*. *FEMS Microbiology Letters* **247** (2): 123–130.
441. Hickman JW, Tifrea DF, Harwood CS (2005) A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. *Proceedings of the National Academy of Sciences* **102** (40): 14422–14427.
442. Janczarek M, Rachwał K (2013) Mutation in the pssA gene involved in exopolysaccharide synthesis leads to several physiological and symbiotic defects in *Rhizobium leguminosarum* bv. *trifolii*. *International Journal of Molecular Sciences* **14** (12): 23711–23735.
443. van Workum WAT, Cremers HCJC, Wijfjes AHM, van der Kolk C, Wijffelman CA, Kijne JW (1997) Cloning and Characterization of Four Genes of *Rhizobium leguminosarum* bv. *trifolii* Involved in Exopolysaccharide Production and Nodulation. *Molecular Plant-Microbe Interactions* **10** (2): 290–301.
444. Bulgarelli D, Rott M, Schlaeppi K, Ver Loren van Themaat E, Ahmadinejad N, Assenza F, Rauf P, Huettel B, Reinhardt R, Schmelzer E, Peplies J, Gloeckner FO, Amann R, Eickhorst T, Schulze-Lefert P (2012) Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature* **488** (7409): 91–5.
445. Lebeis SL, Parades SH, Lundberg DS, Breakfield N, Gehring J, McDonald M, Malfatti S, Glavina del Rio T, Jones CD, Tringe SG, Dangl JL (2015) Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* **349** (6250): 860–864.
446. Strehmel N, Böttcher C, Schmidt S, Scheel D (2014) Profiling of secondary metabolites in root exudates of *Arabidopsis thaliana*. *Phytochemistry* **108**: 35–46.
447. Moussaieff A, Rogachev I, Brodsky L, Malitsky S, Toal TW, Belcher H, Yativ M, Brady SM, Benfey PN, Aharoni A (2013) High-resolution metabolic mapping of cell types in plant roots. *Proceedings of the National Academy of Sciences* **110** (13): 1232–1241.
448. Meier M, Lucchetta EM, Ismagilov RF (2010) Chemical stimulation of the *Arabidopsis thaliana* root using multi-laminar flow on a microfluidic chip. *Lab on a Chip* **10** (16): 2147–2153.
449. Grossmann G, Guo W-J, Ehrhardt DW, Frommer WB, Sit R V., Quake SR, Meier M (2011) The RootChip: An Integrated Microfluidic Chip for Plant Science. *The Plant Cell* **23** (12): 4234–4240.
450. Busch W, Moore BT, Martsberger B, MacE DL, Twigg RW, Jung J, Pruteanu-Malinici I, Kennedy SJ, Fricke GK, Clark RL, Ohler U, Benfey PN (2012) A microfluidic device and computational platform for high-throughput live imaging of gene expression. *Nature Methods* **9** (11): 1101–1106.
451. Jiang H, Xu Z, Aluru MR, Dong L (2014) Plant-Chip for High-Throughput Phenotyping of Arabidopsis. *Lab Chip* **14** (7): 1281–1293.
452. Vullo DL, Coto CE, Sineriz F (1991) Characteristics of an Inulinase Produced by *Bacillus subtilis* 430A, a Strain Isolated from the Rhizosphere of *Vernonia herbacea* (Vell Rusby). *Applied an Environmental Microbiology* **57** (8): 2392–2394.

453. Pandey A, Palni LMS (1997) *Bacillus* species: the dominant bacteria of the rhizosphere of established tea bushes. *Microbiological Research* **152** : 359–365.
454. Cazorla FM, Romero D, Pérez-García A, Lugtenberg BJJ, Vicente A De, Bloemberg G (2007) Isolation and characterization of antagonistic *Bacillus subtilis* strains from the avocado rhizoplane displaying biocontrol activity. *Journal of Applied Microbiology* **103** (5): 1950–1959.
455. Aufrecht JA, Timm CM, Bible A, Morrell-Falvey JL, Pelletier DA, Doktycz MJ, Retterer ST (2018) Microfluidics: Quantifying the Spatiotemporal Dynamics of Plant Root Colonization by Beneficial Bacteria in a Microfluidic Habitat (Adv. Biosys. 6/2018). *Advanced Biosystems* **2** (6): 1–12.
456. Stanley C, Shrivastava J, Brugman R, Heinzelmann E, Frajs V, Bühler A, van Swaay D, Grossmann G (2018) Fabrication and Use of the Dual-Flow-RootChip for the Imaging of Arabidopsis Roots in Asymmetric Microenvironments. *Bio-Protocol* **8** (18): 1–19.
457. Osmekhina E, Jonkergouw C, Schmidt G, Jahangiri F, Jokinen V, Franssila S, Linder MB (2018) Controlled communication between physically separated bacterial populations in a microfluidic device. *Communications Biology* **1** (97): 1–7.
458. Aleklett K, Kiers ET, Ohlsson P, Shimizu TS, Caldas VE, Hammer EC (2018) Build your own soil: Exploring microfluidics to create microbial habitat structures. *ISME Journal* **12** (2): 312–319.
459. Cardinale M (2014) Scanning a microhabitat: Plant-microbe interactions revealed by confocal laser microscopy. *Frontiers in Microbiology* **5** (94): 1–10.
460. Wei X, Bauer WD (1998) Starvation-induced changes in motility, chemotaxis, and flagellation of *Rhizobium meliloti*. *Applied and Environmental Microbiology* **64** (5): 1708–1714.
461. Kong X, Liu G, Liu J, Ding Z (2018) The Root Transition Zone: A Hot Spot for Signal Crosstalk. *Trends in Plant Science* **23** (5): 403–409.
462. Martin BC, Gleeson D, Statton J, Siebers AR, Grierson P, Ryan MH, Kendrick GA (2018) Low light availability alters root exudation and reduces putative beneficial microorganisms in seagrass roots. *Frontiers in Microbiology* **8** : 1–16.
463. Downie H, Holden N, Otten W, Spiers AJ, Valentine TA, Dupuy LX (2012) Transparent Soil for Imaging the Rhizosphere. *PLoS ONE* **7** (9): 1–6.
464. Felz S, Vermeulen P, van Loosdrecht MCM, Lin YM (2019) Chemical characterization methods for the analysis of structural extracellular polymeric substances (EPS). *Water Research* **157**: 201–208.
465. Stewart TJ, Traber J, Kroll A, Behra R, Sigg L (2013) Characterization of extracellular polymeric substances (EPS) from periphyton using liquid chromatography-organic carbon detection-organic nitrogen detection (LC-OCD-OND). *Environmental Science and Pollution Research* **20** (5): 3214–3223.
466. Ramstedt M, Nakao R, Wai SN, Uhlin BE, Boily J-F (2011) Monitoring Surface Chemical Changes in the Bacterial Cell Wall. *Journal of Biological Chemistry* **286** (14): 12389–12396.
467. Huang KC, Mukhopadhyay R, Wen B, Gitai Z, Wingreen NS (2008) Cell shape and cell-wall organization in Gram-negative bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **105** (49): 19282–19287.
468. Yang DC, Blair KM, Salama NR (2016) Staying in Shape: the Impact of Cell Shape on Bacterial Survival in Diverse Environments. *Microbiology and molecular biology reviews* **80** (1): 187–203.
469. Alhede M, Qvortrup K, Liebrechts R, Høiby N, Givskov M, Bjarnsholt T (2012) Combination of microscopic techniques reveals a comprehensive visual impression of biofilm structure and composition. *FEMS Immunology and Medical Microbiology* **65** (2): 335–342.

470. DeJesus MA, Ambadipudi C, Baker R, Sassetti C, Ioerger TR (2015) TRANSIT - A Software Tool for Himar1 TnSeq Analysis. *PLoS Computational Biology* **11** (10): 1–17.
471. Newell PD, Yoshioka S, Hvorecny KL, Monds RD, O'Toole GA (2011) Systematic analysis of diguanylate cyclases that promote biofilm formation by *Pseudomonas fluorescens* Pf0-1. *Journal of Bacteriology* **193** (18): 4685–4698.
472. Agard NJ, Wells JA (2009) Methods for proteomic identification of protease substrates. *Current Opinion in Chemical Biology* **13** (5–6): 503–509.
473. Marini S, Vitali F, Rampazzi S, Demartini A, Akutsu T (2019) Protease target prediction via matrix factorization. *Bioinformatics* **35** (6): 923–929.

Appendix 1. Supplementary material for Chapters 3, 4 and 5

Overview

Chapter 3 – Tables A1 – A2

Chapter 4 – Tables A3 – A9

Chapter 5 – Figures A1 – A4 and Tables A10 – A17

Table A1 – Genes identified by approach 1 (Figure 3-10) in Rlv3841 putative novel adhesin identification

			Fold expression changes vs free living controls						
			RNA sequencing	Microarrays					
Gene	**Protein description**	**INSeq root colonisation**	**RNA seq 7d rhizo 7dpi**	**7d rhizo 1dpi**	**7d rhizo 3dpi**	**7d rhizo 7dpi**	**14d rhizo 1dpi**	**21d rhizo 1dpi**	**21d bacteroid**
pRL80020	Conserved hypothetical protein	DE	242.38	23.04	65.74	34.52	12.3	31.94	1.69
RL2307	Conserved hypothetical protein	DE	154.64	31.29	27.39	22.56	14.71	21.9	1.55
RL0506	Conserved hypothetical protein	DE	138.66	56.13	45.78	42.89	28.5	40.9	1.54
RL1165	Conserved hypothetical protein	DE	97.87	68.52	41.33	24.31	39.44	55.42	1.91
RL1172	Putative transmembrane protein	NE	67.68	35.66	28.52	31.92	17.25	20.85	2.14
RL3186	Putative transmembrane protein	DE	64	42.93	84.33	56.5	37.88	45.19	2.05
RL3982	Conserved hypothetical protein	DE	52.31	32.04	21.59	27.33	23.72	31.45	2.57
RL2969	putative transmembrane protein	NE	50.78	8.78	5.37	10.77	9.6	12.03	0.93
RL3702	putative transmembrane protein	NE	21.37	2.94	1.84	3.02	2.42	3.09	1.13
pRL120724	Putative transmembrane protein	DE	18.18	18.08	17.61	16.88	10.94	19.56	1.77
pRL110268	Conserved hypothetical protein	DE	11.8	149.22	93.33	66.34	39.65	42.52	2.04
RL2554	Hypothetical exported protein	DE	11.56	19.18	37.37	87.24	37.26	123.98	1.59
RL3273	putative von Willebrand factor type A	DE	7.25	8.75	4.42	3.05	4.32	9.38	2.39
RL3384	Conserved hypothetical exported protein	NE	6.29	22.64	49.48	32.6	54.98	75.06	0.44

Rhizo = rhizosphere, dpi = days post inoculation. For INSeq, DE = defective, NE = neutral. Array Express accession numbers for microarrays can be found in Table 3-2.

Table A2– Genes identified by approach 2 (Figure 3-10) in Rlv3841 putative novel adhesin identification

Gene	Description	Input library	Rhizosphere library	Root colonisation library
pRL70001 (*repA*)	putative replication protein RepA	NE	NE	DE
pRL80032	putative LysR family transcriptional regulator	NE	NE	DE
pRL90009	conserved hypothetical protein with CBS domain	NE	NE	ES
pRL90149	putative LuxR/GerE family transcriptional regulator	NE	NE	DE
pRL90204	putative amidase	NE	NE	DE
pRL90243	putative acyl-CoA dehydrogenase	NE	NE	DE
pRL90298	conserved hypothetical protein	NE	NE	DE
pRL100011	hypothetical protein	NE	NE	DE
pRL100012	conserved hypothetical protein	NE	NE	DE
pRL100026	conserved hypothetical protein	NE	NE	DE
pRL100027	putative restriction modification methylase	NE	NE	DE
pRL100053	putative transmembrane protein	NE	NE	ES
pRL100056	putative aromatic amino acid aminotransferase	NE	NE	DE
pRL100057	putative AsnC family transcriptional regulator	NE	NE	DE
pRL100127	putative ATP-binding component of ABC transporter Unclass	NE	NE	DE
pRL100128	putative ATP-binding component of ABC transporter Unclass	NE	NE	DE
pRL100129	putative permease component of ABC transporter Unclass	NE	NE	DE
pRL100164	putative enzyme to synthesise (autoinducer) AHL, RhiI	NE	NE	DE
pRL100170 (*rhiB*)	rhizosphere-induced protein RhiB	NE	NE	DE
pRL100171 (*rhiC*)	rhizosphere-induced protein RhiC	NE	NE	DE
pRL100172 (*rhiR*)	putative transcriptional regulatory protein controlling rhi gene expression RhiR	NE	NE	DE
pRL100179 (*nodN*)	nodulation protein NodN	NE	NE	DE
pRL100309	cadherin domain-containing calcium-binding glycoprotein	NE	NE	DE
pRL100387	putative gluconolactonase precursor	NE	NE	DE
pRL100428	putative permease component of ABC transporter Unclass	NE	NE	DE
pRL100429	putative short-chain dehydrogenase	NE	NE	DE
pRL110142	hypothetical protein	NE	NE	DE
pRL110169	putative SBP of ABC transporter CUT2	NE	NE	DE

pRL110353	conserved hypothetical protein	NE	NE	DE
pRL110543	conserved hypothetical protein	NE	NE	ES
pRL110544 (*minE*)	putative cell division topological specificity factor	NE	NE	ES
pRL110545 (*minD*)	putative septum site-determining protein MinD	NE	NE	ES
pRL110560 (*soxB*)	putative sarcosine oxidase beta subunit	NE	NE	DE
pRL110561 (*soxD*)	putative sarcosine oxidase delta subunit	NE	NE	DE
pRL110564	putative tight adherence protein	NE	NE	ES
pRL120021	hypothetical protein	NE	NE	ES
pRL120058	hypothetical protein	NE	NE	DE
pRL120160	putative DeoR family transcriptional regulator (repressor)	NE	NE	ES
pRL120306	putative ATP-binding component of ABC transporter NitT	NE	NE	DE
pRL100325 (*fhuA*)	outer membrane siderophore receptor precursor	NE	NE	DE
pRL120579	putative transmembrane protein	NE	NE	ES
pRL120584	putative epimerase	NE	NE	DE
pRL120656	hypothetical protein	NE	NE	ES
pRL120692	putative ROK family transcriptional regulator	NE	NE	DE
RL0133	conserved hypothetical protein	NE	NE	DE
RL0256	putative XRE family (HipB) transcriptional regulator with cupin2 domian	NE	NE	DE
RL0561	putative AraC family transcriptional regulator (activator)	NE	NE	DE
RL0682	putative transmembrane ion efflux system protein	NE	NE	DE
RL0799	putative hexapeptide repeat transferase	NE	NE	DE
RL0802	putative deoxygenase	NE	NE	DE
RL0817	putative transmembrane protein	NE	NE	DE
RL0929	hypothetical protein	NE	NE	DE
RL0930 (*rnhB*)	putative ribonuclease HII	NE	NE	DE
RL0934 (*moaB*)	putative molybdenum cofactor biosynthesis protein B	NE	NE	DE
RL0963	putative transmembrane/surface protein BA14K-like immuno reactive protein	NE	NE	DE
RL1001	conserved hypothetical protein	NE	NE	DE
RL1032 (*rnhA*)	putative ribonuclease HI	NE	NE	DE
RL1040	putative LysR family transcriptional regulator	NE	NE	DE
RL1041	putative oxidoreductase/monooxygenase	NE	NE	DE
RL1047	putative ATP-binding component of ABC transporter MZT	NE	NE	DE
RL1048	putative permease component of ABC transporter MZT	NE	NE	DE

RL1084 (*smc*)	putative structural maintenance of chromosomes protein	NE	NE	DE
RL1085	hypothetical protein	NE	NE	DE
RL1091	conserved hypothetical protein	NE	NE	DE
RL1092	conserved hypothetical protein	NE	NE	DE
RL1093	putative beta-lactamase family protein	NE	NE	DE
RL1101	hypothetical protein	NE	NE	DE
RL1340 (*sodB*)	putative superoxide dismutase	NE	NE	DE
RL1504	conserved hypothetical protein	NE	NE	DE
RL1600 (*ppx*)	putative exopolyphosphatase	NE	NE	DE
RL1628	hypothetical protein	NE	NE	ES
RL1631	putative cytochrome c type nitrate reductase NapC exported to periplasm	NE	NE	DE
RL1632 (*ribH*)	putative 6,7-dimethyl-8-ribityllumazine synthase	NE	NE	DE
RL1633 (*nusB*)	putative N utilization substance protein B	NE	NE	DE
RL1937	hypothetical protein	NE	NE	DE
RL1938	putative phosphatase	NE	NE	DE
RL1939	hypothetical protein	NE	NE	DE
RL2109	conserved hypothetical protein	NE	NE	DE
RL2153	conserved hypothetical protein	NE	NE	DE
RL2284 (*hfq*)	putative host factor protein	NE	NE	ES
RL2291	conserved hypothetical protein	NE	NE	DE
RL2304	conserved hypothetical protein	NE	NE	DE
RL2305	hypothetical protein	NE	NE	DE
RL2307	conserved hypothetical protein	NE	NE	DE
RL2368	putative GntR family transcriptional regulator	NE	NE	DE
RL2507	conserved hypothetical exported protein	NE	NE	DE
RL2508 (*gltA*)	putative citrate synthase II	NE	NE	DE
RL2662	putative racemase/isomerase	NE	NE	DE
RL2663	putative permease component of ABC transporter Unclass	NE	NE	DE
RL2664 (*lpxH*)	putative UDP-2,3-diacylglucosamine hydrolase	NE	NE	DE
RL2692	putative DAHP synthetase protein	NE	NE	DE
RL2702 (*rapD*)	putative rhizobium adhering-like protein	NE	NE	DE
RL2776	putative AsnC family transcriptional regulator	NE	NE	DE
RL2828	putative XRE family (HipB) family transcriptional regulator	NE	NE	DE
RL2837	putative SBP of ABC transporter QAT? orphan proline/glycine betaine	NE	NE	DE
RL2930	conserved hypothetical protein	NE	NE	DE

RL2931	putative methyl-accepting chemotaxis protein McpN	NE	NE	DE
RL2932	putative acyl-CoA hydrolase	NE	NE	DE
RL2933	putative transmembrane protein	NE	NE	DE
RL2934	conserved hypothetical exported protein	NE	NE	DE
RL3198	conserved hypothetical protein	NE	NE	DE
RL3199	putative aminotransferase	NE	NE	DE
RL3213	putative tetracycline resistance protein	NE	NE	DE
RL3320	putative signalling and peptidoglycan binding protein	NE	NE	DE
RL3321	putative DnaJ family chaperone	NE	NE	DE
RL3460 (*proC*)	putative pyrroline-5-carboxylate reductase	NE	NE	DE
RL3560 (*map*)	putative methionine aminopeptidase	NE	NE	DE
RL3561	putative bacterial luciferase family protein	NE	NE	DE
RL3591	putative two-component sensor/regulator; histidine kinase	NE	NE	DE
RL3677 (*IspL*)	putative UDP-glucuronate 5'-epimerase	NE	NE	DE
RL3828	putative FNR/CRP family transcriptional regulator	NE	NE	DE
RL4087	putative LysR family transcriptional regulator	NE	NE	DE
RL4183	putative transmembrane protein	NE	NE	DE
RL4382	putative filamentous hemagglutinin adherence factor precursor	NE	NE	DE
RL4632	putative DeoR family transcriptional regulator (repressor)	NE	NE	DE
RL4669	hypothetical protein	NE	NE	DE
RL4670	conserved hypothetical protein	NE	NE	DE
RL4671 (*rci*)	putative shufflon-specific DNA recombinase	NE	NE	DE
RL4689	conserved hypothetical protein	NE	NE	DE
RL4690	conserved hypothetical protein	NE	NE	DE
pRL110071	conserved hypothetical protein	NE	DE	ES
RL0109	conserved hypothetical protein	DE	DE	DE

NE = neutral, DE = defective, ES = essential.

Table A3 – Genes ES/DE under all pHs in primary root attachment which were also NE in the input library (not shown).

Gene	Description	G21	S21	G1	S1	VMM	TY	Rhi	RA 6.5	RA 7.0	RA 7.5	Col	IT	Nod
RL0876	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE	NE
pRL100174	hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE	NE
RL1381	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	NE	NE
RL3752 (pssA)	putative transferase involved in exopolysaccharide biosynthesis	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE	NE
RL4145	putative LacI family transcriptional regulator (repressor)	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE	NE
RL4381	putative cell surface protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE	NE
RL4382	putative filamentous hemagglutinin adherence factor precursor	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	NE	NE
RL0551 (hslO)	putative Hsp33-like chaperonin	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	NE	NE
RL2513 (tpiA)	putative triosephosphate isomerase	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	NE	NE
RL4065	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	NE	NE
RL3766 (rpoH1)	putative RNA polymerase sigma-32 factor (heat shock)	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	NE	NE
pRL100053	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	ES	DE	DE
RL1478 (amn)	putative AMP nucleosidase	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	ES	ES
RL2400	putative MarC (multiple antibiotic resistance) family transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	DE	AD
RL3322 (pfp)	putative pyrophosphate--fructose 6-phosphate 1-phosphotransferase	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE	DE
RL3987	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	DE	DE
RL3988	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	DE	DE
RL3989 (ruvA)	putative Holliday junction DNA helicase RuvA	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	DE	DE
RL3990 (ruvB)	putative Holliday junction DNA helicase RuvB	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	DE	DE
RL4362	putative cobalamin synthesis protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	ES	NE	DE	DE
RL4363 (dacC)	putative penicillin-binding protein precursor	NE	NE	NE	NE	NE	NE	NE	DE	DE	ES	NE	DE	DE

RL2637 (recA)	putative recombinase	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	ES	ES
pRL100405	putative transmembrane protein	NE	NE	NE	NE	AD	NE	NE	DE	DE	DE	NE	NE	NE
pRL100406 (mctR)	putative two-component sensor/regulator; transcriptional regulator	NE	NE	NE	NE	AD	NE	NE	DE	DE	DE	NE	NE	NE
pRL110044	conserved hypothetical protein	NE	NE	NE	NE	DE	NE	DE	DE	DE	DE	DE	NE	NE
pRL100404 (mctP)	putative transmembrane sodium-solute symporter	NE	NE	NE	NE	AD	NE	AD	DE	DE	DE	NE	NE	NE
pRL110071	conserved hypothetical protein	DE	DE	DE	DE	ES	NE	DE	DE	DE	DE	ES	NE	DE
pRL70055	hypothetical protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	DE	ES
pRL70100	hypothetical protein	ES	DE	ES	DE	NE	NE	DE	DE	DE	DE	NE	NE	NE
RL0160 (polA)	putative DNA polymerase I	DE	DE	NE	DE	NE	NE	NE	DE	DE	DE	NE	DE	NE
RL0187	putative permease component of ABC transporter PepT	DE	DE	DE	DE	NE	NE	NE	DE	DE	DE	NE	ES	ES
RL0430 (def1)	putative peptide deformylase (polypeptide deformylase)	DE	DE	DE	DE	DE	NE	DE	DE	DE	DE	DE	DE	ES
RL0431	putative plasmid stability protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	DE	ES
RL0432	putative plasmid stability protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	DE	ES
RL0855 (gshA)	putative ADmma-glutamylcysteine synthetase precursor	DE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	ES	ES
RL0921	putative cationic transport protein, CorA family	DE	DE	NE	NE	NE	NE	DE	DE	DE	DE	DE	DE	DE
RL1389	conserved hypothetical protein	DE	DE	NE	DE	NE	NE	NE	DE	DE	DE	NE	NE	ES
RL1390	hypothetical protein	NE	DE	NE	DE	NE	NE	NE	DE	DE	DE	NE	NE	NE
RL1391	putative transmembrane protein	NE	DE	NE	DE	NE	NE	NE	DE	DE	DE	NE	NE	NE
RL1392	putative transmembrane protein	NE	DE	NE	DE	NE	NE	NE	DE	DE	DE	NE	NE	NE
RL1506 (relA)	putative stringent response protein	NE	NE	DE	DE	NE	NE	DE	DE	DE	DE	DE	ES	ES
RL1507	conserved hypothetical protein	NE	NE	NE	DE	NE	NE	DE	DE	DE	DE	DE	DE	NE
RL1508	putative transmembrane protein	NE	NE	NE	DE	NE	NE	DE	DE	DE	DE	NE	ES	ES
RL1572	putative DNA polymerase III subunit	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	ES	ES
RL2062 (tig)	putative chaperone trigger factor	DE	DE	NE	DE	NE	NE	DE	DE	DE	DE	NE	NE	NE
RL2406 (queA)	putative S-adenosylmethionine:tRNA ribosyltransferase-isomerase	DE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	DE	ES
RL2407 (tgt)	putative queuine tRNA-ribosyltransferase	DE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	DE	ES

RL2473 (metG)	putative methionyl-tRNA synthetase	DE	DE	DE	NE	NE	NE	NE	DE	DE	DE	NE	NE	ES
RL2474	putative transmembrane protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	NE	NE	NE
RL2514	conserved hypothetical protein	DE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	NE	NE
RL2638 (rbsK3)	putative ribokinase	ES	ES	NE	DE	NE	NE	DE	DE	DE	DE	DE	NE	NE
RL3257	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	ES	DE
RL3453	putative two-component sensor/regulator; histidine kinase	NE	DE	NE	NE	NE	NE	NE	DE	DE	DE	NE	ES	DE
RL3462	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	DE	NE
RL3596	conserved hypothetical protein	DE	DE	NE	NE	NE	NE	DE	DE	DE	DE	NE	NE	DE
RL3597	putative DEAD-box ATP-dependent RNA helicase protein	DE	DE	NE	NE	NE	NE	DE	DE	DE	DE	DE	NE	DE
RL3677 (IspL)	putative UDP-glucuronate 5'-epimerase	NE	NE	NE	DE	NE	NE	NE	DE	DE	DE	DE	ES	ES
RL3763	putative rRNA methyltransferase	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	NE	NE
RL3872 (purC2)	putative phosphoribosylamidoimidazole-succinocarboxamide synthase	DE	DE	DE	DE	DE	NE	NE	DE	DE	DE	NE	NE	NE
RL4040 (thiE2)	putative thiamine-phosphate pyrophosphorylase	ES	DE	DE	DE	NE	NE	DE	DE	DE	ES	DE	ES	ES
RL4066	conserved hypothetical protein	DE	DE	DE	DE	DE	NE	DE	DE	DE	DE	DE	ES	ES
RL4073 (hss)	putative homospermidine synthase	DE	DE	NE	DE	NE	NE	DE	DE	DE	DE	NE	ES	ES
RL4582 (fbpC)	putative ATP-binding component of ABC transporter Unclass ferric cations transporter	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	DE	DE
RL4583 (fbpB)	putative permease component of ABC transporter Unclass ferric cations transporter	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	DE	DE
RL4584 (fbpA)	putative SBP of ABC transporter Unclass ferric cations transporter	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	NE	NE	NE

INSeq experiments are coded as follows: G21 = 21% oxygen, 10 mM glucose. S21 = 21 % oxygen, 20 mM succinate. G1 = 1% oxygen, 10 mM glucose. S1 = 1% oxygen, 20 mM succinate. VMM = Vincents media. TY = Tryptone yeast media. Rhizo = rhizosphere. RA 6.5 = root attachment, pH 6.5. RA 7.0 = root attachment, pH 7.0. RA 7.5 = root attachment, pH 7.5 Colon = 5 day root colonisation. IT = infection thread. Nod = nodulation. Thick red line = pleiotropy filter cutoff, as described in 4.2.8

Table A4 – Genes ES/DE under pH 6.5 conditions only in primary root attachment (data for primary root attachment pH 7.0 and 7.5 conditions not shown) which were also NE in the input library (not shown).

Gene	Description	G21	S21	G1	S1	VMM	TY	Rhi	RA 6.5	Col	IT	Nod
pRL100177	putative homologue of eukaryotic tubulin	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL100274 (fucA)	putative alpha-L-fucosidase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL0614	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1026 (ctaE1)	putative cytochrome c oxidase polypeptide III (cytochrome aa3 subunit 3)	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2285 (hflX)	putative GTP-binding protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2316	putative cyclase/kinase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2564	conserved hypothetical exported protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2595	putative MutT/Nudix family protein (phosphohydrolases)	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2780 (dgkA)	putative transmembrane diacylglycerol kinase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3179	putative cobalamin synthesis protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL120475 (impA)	inner membrane protein ImpA, component of T6SS	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL4404 (pssA3)	putative exopolysaccharide production protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL120021	hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	ES	NE	NE
pRL100176	pseudogene, incomplete ATP-binding protein	NE	NE	NE	NE	NE	NE	DE	DE	NE	0	0
RL2394	putative carbohydrate kinase protein	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
pRL110283	putative ArsR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	NE	ES	DE
RL0032	putative phosphocarrier protein HPr for mannose	NE	NE	NE	NE	NE	NE	DE	ES	NE	ES	ES
RL0033	putative phosphotransferase system component, mannose PTS component IIA	NE	NE	NE	NE	NE	NE	DE	ES	NE	ES	ES
RL0398	putative acetyltransferase	NE	NE	NE	NE	NE	NE	NE	DE	NE	DE	DE
RL0726	conserved hypothetical exported protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	DE	DE
RL2211	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	DE
RL2303 (ccdA)	putative cytochrome c-type biogenesis protein	NE	NE	NE	NE	NE	NE	ES	DE	DE	ES	ES
pRL100158 (nifN)	putative nitrogenase iron-molybdenum cofactor biosynthesis protein NifN	DE	DE	DE	DE	NE	NE	DE	DE	NE	DE	DE
RL0019	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	NE	DE	NE	DE	ES
RL0226	putative permease component of ABC transporter PepT	DE	DE	DE	DE	NE	NE	DE	DE	NE	ES	ES
RL0227	putative permease component of ABC transporter PepT	DE	DE	DE	DE	NE	NE	DE	DE	NE	ES	ES
RL0228	putative SBP of ABC transporter PepT	DE	DE	DE	DE	NE	NE	DE	DE	NE	ES	ES
RL0501	putative orotate phosphoribosyltransferase	DE	NE	NE	NE	DE	NE	NE	ES	NE	NE	DE

RL0570	conserved hypothetical exported protein	DE	NE	NE	DE	NE	NE	NE	DE	NE	NE	NE
RL0856	conserved hypothetical protein	DE	NE	NE	NE	NE	NE	AD	DE	NE	AD	NE
RL1020	putative invasion associated protein	DE	DE	NE	DE	NE	NE	DE	DE	NE	NE	NE
RL1025 (ctaG)	putative cytochrome c oxidase assembly protein	DE	DE	DE	DE	NE	NE	NE	DE	NE	NE	NE
RL1548 (radA)	putative DNA repair protein RadA homologue	DE	DE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2032	hypothetical protein	DE	DE	NE	DE	NE	NE	NE	DE	NE	NE	NE
RL2045 (scpB)	putative chromosome segreADtion and condensation protein B	DE	DE	DE	DE	NE	NE	DE	ES	NE	NE	NE
RL2117A	hypothetical protein	DE	DE	NE	DE	ES	NE	NE	DE	NE	ES	DE
RL2152	hypothetical protein	DE	DE	ES	DE	NE	NE	NE	DE	NE	ES	ES
RL2324	putative ROK family transcriptional regulator	DE	NE	NE	NE	NE	NE	DE	DE	NE	DE	NE
RL2404	putative peptidyl-prolyl cis-trans isomerase (cyclophilin)	DE	NE	NE	NE	NE	NE	NE	DE	NE	DE	DE
RL2526	putative oxidoreductase	DE	NE	DE	DE	NE	NE	NE	DE	NE	DE	NE
RL3180	hypothetical protein	DE	DE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3181	conserved hypothetical protein	DE	DE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL3295 (recN)	putative DNA repair protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	NE	NE
RL3334 (rnsA)	putative ribonuclease I	DE	DE	NE	DE	NE	NE	DE	DE	NE	NE	NE
RL3335	putative lysophospholipase	DE	DE	NE	DE	NE	NE	NE	DE	NE	NE	NE
RL3465	conserved hypothetical protein	NE	DE	NE	DE	NE	NE	NE	DE	NE	DE	DE
RL3667	putative UDP-glucose 6-dehydrogenase	DE	DE	DE	DE	NE	NE	DE	DE	DE	ES	DE
RL3669	conserved hypothetical protein	NE	DE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL3678	hypothetical protein	NE	NE	NE	DE	NE	NE	NE	DE	NE	NE	NE
RL4011 (pgk)	putative phosphoglycerate kinase	NE	NE	NE	NE	DE	NE	NE	ES	NE	NE	DE

INSeq experiments are coded as follows: G21 = 21% oxygen, 10 mM glucose. S21 = 21 % oxygen, 20 mM succinate. G1 = 1% oxygen, 10 mM glucose. S1 = 1% oxygen, 20 mM succinate. VMM = Vincents media. TY = Tryptone yeast media. Rhizo = rhizosphere. RA 6.5 = root attachment, pH 6.5. Colon = 5 day root colonisation. IT = infection thread. Nod = nodulation. 0 = data not available. Thick red line = pleiotropy filter cutoff, as described in 4.2.8

Table A5 – Genes ES/DE under pH 7.0 conditions only in primary root attachment (data for primary root attachment pH 6.5 and 7.5 conditions not shown) which were also NE in the input library (not shown).

Gene	Description	G21	S21	G1	S1	VMM	TY	Rhi	RA 7.0	Col	IT	Nod
pRL120518	putative TetR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL0552	hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1052	hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2044 (scpA)	putative chromosome segreADtion and condensation protein A	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2587	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2695	hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3226	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL4704	putative glyoxalase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1106 (pspA)	putative (phage shock protein A) PspA family regulator by protein-protein interactions	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
pRL100112	putative dehalogenase/hydrolase	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE
RL1371	putative transmembrane protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE
RL1013	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	ES	NE
RL2588 (tyrS1)	putative tyrosyl-tRNA synthetase	NE	NE	NE	NE	NE	NE	NE	DE	NE	ES	ES
RL2694 (gor)	putative glutathione reductase	NE	NE	NE	NE	NE	NE	NE	DE	NE	ES	ES
pRL100149	conserved hypothetical protein	NE	NE	NE	NE	ES	NE	NE	ES	NE	NE	NE
pRL100388	putative LacI family transcriptional regulator (repressor)	NE	DE	NE	DE	NE	NE	NE	DE	NE	NE	NE
pRL110073	putative GntR family transcriptional regulator	DE	DE	DE	DE	NE	NE	NE	DE	NE	NE	NE
pRL110389	putative exopolysaccharide production protein	DE	DE	DE	DE	NE	NE	NE	ES	NE	ES	ES
pRL90040	pseudogene, conserved hypothetical protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	0	0
pRL90280	conserved hypothetical protein	ES	ES	DE	DE	ES	NE	NE	ES	NE	NE	NE
RL0417 (ihfB)	putative integration host factor beta subunit	NE	NE	NE	NE	ES	NE	ES	DE	NE	NE	NE
RL0563	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	NE	DE	NE	NE	NE
RL0564 (mcpE)	putative methyl-accepting chemotaxis protein	DE	DE	DE	DE	NE	NE	NE	DE	NE	NE	NE
RL1001	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	NE	DE	DE	NE	NE
RL1002 (bioY)	putative transmembrane biotin biosynthesis protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	ES	ES

RL1003	putative permease component of ABC transporter Unclass	DE	DE	DE	DE	NE	NE	DE	DE	DE	ES	ES
RL1004	putative ATP-binding component of ABC transporter Unclass	DE	DE	DE	DE	NE	NE	DE	DE	DE	ES	ES
RL1060	putative ribosomal-protein-alanine acetyltransferase	DE	DE	NE	DE	ES	NE	NE	ES	NE	ES	ES
RL1105	putative TetR family transcriptional regulator	NE	NE	NE	NE	ES	NE	DE	DE	NE	NE	ES
RL1450	hypothetical protein	DE	DE	DE	ES	NE	NE	DE	ES	DE	NE	NE
RL1451	putative uracil-DNA glycosylase	DE	DE	DE	ES	NE	NE	DE	ES	DE	NE	NE
RL1585 (moaD)	putative molybdopterin converting factor subunit D	NE	NE	NE	NE	ES	NE	NE	DE	NE	NE	NE
RL1618	putative MarR family transcriptional regulator	DE	DE	DE	DE	NE	NE	DE	DE	NE	DE	DE
RL2071	putative transmembrane protein	DE	NE	NE	NE	NE	NE	NE	ES	NE	NE	NE
RL2072	putative transmembrane protein	DE	NE	NE	NE	NE	NE	NE	ES	NE	NE	NE
RL2073 (ruvX)	putative Holliday junction resolvase RuvX	DE	NE	NE	NE	NE	NE	NE	ES	NE	NE	NE
RL2478	putative outer membrane protein	NE	DE	NE	DE	NE	NE	NE	DE	NE	NE	NE
RL2652 (fdxB2)	putative ferredoxin, 2Fe-2S	DE	DE	DE	DE	NE	NE	NE	DE	NE	NE	NE
RL2653	putative HPt (Histidine Phosphotransfer) domain protein of two-component system (multistage phosphorelay can result)	DE	DE	DE	DE	NE	NE	NE	DE	NE	NE	NE
RL2696	putative acylphosphatase	ES	DE	ES	ES	NE	NE	NE	DE	NE	NE	NE
RL2953	putative adenylate/guanylate cyclase	DE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL4304	putative transmembrane transporter protein	NE	ES	DE	NE	NE	NE	NE	DE	NE	NE	NE
RL4546	conserved hypothetical protein	ES	DE	DE	NE	NE	NE	NE	DE	NE	NE	NE

INSeq experiments are coded as follows: G21 = 21% oxygen, 10 mM glucose. S21 = 21 % oxygen, 20 mM succinate. G1 = 1% oxygen, 10 mM glucose. S1 = 1% oxygen, 20 mM succinate. VMM = Vincents media. TY = Tryptone yeast media. Rhizo = rhizosphere. RA 7.0 = root attachment, pH 7.0. Colon = 5 day root colonisation. IT = infection thread. Nod = nodulation. 0 = data not available. Thick red line = pleiotropy filter cutoff, as described in 4.2.8

Table A6 – Genes ES/DE under pH 7.5 conditions only in primary root attachment (data for primary root attachment pH 6.5 and 7.0 conditions not shown) which were also NE in the input library (not shown).

Gene	Description	G21	S21	G1	S1	VMM	TY	Rhi	RA 7.5	Col	IT	Nod
pRL120795	hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL0395 (miaB)	putative MiaB protein (methylthiolation of isopentenylated A37 derivatives in rRNA)	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1164	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1165	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1338 (pmtA)	putative phosphatidylethanolamine N-methyltransferase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1339	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2094 (phaC)	putative poly(3-hydroxyalkanoate) polymerase (PHA synthase)	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2095	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2489A	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2491	conserved hypothetical exported protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2644	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2777	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2778	putative exopolysaccharide biosynthesis protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2857 (msiR)	putative ArsR family transcriptional regulator MsiR, regulates MsiA canavanine (found in seed exudate) exporter	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2858	conserved hypothetical exported protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3253 (hflC)	putative transmembrane serine protease	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL4063	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL4075	putative 5'-nucleotidase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL4383	putative AsnC family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL120322 (fhuA2)	outer membrane siderophore receptor precursor	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
RL0561	putative AraC family transcriptional regulator (activator)	NE	NE	NE	NE	NE	NE	NE	ES	DE	NE	NE
RL1505	putative polymerase	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL0401	putative universal stress protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	DE
RL1552 (rplI)	putative 50S ribosomal protein L9	NE	NE	NE	NE	NE	NE	DE	ES	DE	DE	DE
RL2080	putative acetyltransferase	NE	NE	NE	NE	NE	NE	AD	DE	AD	NE	NE

RL2081	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	AD	ES	ES
RL2083	putative acetyltransferase	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	NE
RL3254 (hflK)	putative transmembrane serine protease	NE	NE	NE	NE	NE	NE	NE	DE	NE	DE	NE
RL4018	putative ATP-binding:permease (ABC:IMP) component of ABC transporter Export	NE	NE	NE	NE	NE	NE	NE	DE	NE	DE	DE
RL4062	putative amidohydrolase	NE	NE	NE	NE	NE	NE	NE	DE	NE	ES	ES
RL4497	putative transmembrane protein	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	DE
pRL100037	hypothetical protein	DE	DE	DE	DE	NE	NE	NE	DE	NE	NE	NE
pRL100199 (fixB)	electron transfer protein FixB	DE	DE	DE	DE	NE	NE	NE	DE	NE	DE	DE
pRL100200 (fixA)	electron transfer protein FixA	DE	DE	DE	DE	NE	NE	NE	DE	NE	DE	DE
pRL120319 (rpol)	putative RNA polymerase ECF sigma factor involved in iron uptake, ECF09 gene arrangment (no asf) induced in MA of iron limitation in Rlv3841	DE	DE	DE	DE	DE	NE	NE	DE	DE	NE	NE
pRL120796	hypothetical exported protein	DE	DE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL90144	putative exopolysaccharide biosynthesis-related protein	DE	DE	DE	DE	NE	NE	NE	DE	NE	NE	NE
RL0344	putative AsnC family transcriptional regulator	NE	DE	DE	ES	NE	NE	DE	DE	NE	ES	ES
RL0345	conserved hypothetical protein	NE	DE	DE	ES	NE	NE	NE	DE	NE	ES	NE
RL0346 (dut)	putative deoxyuridine 5'triphosphate nucleotidohydrolase	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE
RL0546 (phoU)	putative phosphate uptake regulator PhoU, unknown mechanism to regulate expression of high-affinity ABC systems	NE	NE	DE	DE	NE	NE	NE	DE	NE	ES	ES
RL0547 (phoB)	putative two-component sensor/regulator; phosphate regulon transcriptional regulator PhoB	DE	NE	DE	NE	NE	NE	NE	DE	NE	NE	NE
RL0903 (gph)	putative phosphoglycolate phosphatase	DE	DE	DE	DE	NE	NE	NE	DE	NE	ES	ES
RL0922 (kup1)	putative potassium uptake transport system protein	DE	DE	NE	NE	NE	NE	DE	DE	DE	DE	NE
RL1166	putative ribonuclease-L-PSP family protein	DE	DE	DE	DE	NE	NE	NE	DE	NE	DE	NE
RL1167	putative TetR family transcriptional regulator	DE	DE	DE	DE	NE	NE	NE	DE	NE	DE	NE
RL1393 (pbpF)	putative peptidoglycan biosynthesis/penicillin binding protein	NE	DE	NE	DE	NE	NE	NE	DE	NE	NE	NE
RL1564 (ksAD)	putative dimethyladenosine transferase	DE	DE	DE	DE	NE	NE	NE	DE	NE	NE	NE
RL2236 (lexA)	putative LexA repressor (SOS regulatory protein)	NE	DE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2237	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	NE	DE	NE	NE	NE
RL2496 (moeA)	putative molybdopterin synthesis protein MoeA	ES	ES	DE	ES	ES	NE	NE	DE	NE	NE	NE
RL2818 (fnrN)	putative FNR/CRP family transcriptional regulator, 100% id to VF39 FnrN	DE	DE	DE	DE	NE	NE	NE	DE	NE	AD	NE

RL3549 (glnII)	putative glutamine synthetase II	DE	DE	DE	DE	NE	NE	NE	DE	NE	DE	ES
RL4297	putative foldase/peptidyl-prolyl cis-trans isomerase	DE	DE	DE	DE	NE	NE	DE	DE	DE	ES	ES
RL4430	putative 3-oxoacyl-[acyl-carrier-protein] reductase	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	NE
RL4431	conserved hypothetical exported protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	NE
RL4432	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	DE	DE	NE	ES	NE
RL4493 (gpsA)	putative glycerol-3-phosphate dehydrogenase [NAD(P)+]	NE	DE	NE	NE	ES	NE	NE	DE	NE	ES	DE

INSeq experiments are coded as follows: G21 = 21% oxygen, 10 mM glucose. S21 = 21 % oxygen, 20 mM succinate. G1 = 1% oxygen, 10 mM glucose. S1 = 1% oxygen, 20 mM succinate. VMM = Vincents media. TY = Tryptone yeast media. Rhizo = rhizosphere. RA 7.5 = root attachment, pH 7.5. Colon = 5 day root colonisation. IT = infection thread. Nod = nodulation. Thick red line = pleiotropy filter cutoff, as described in 4.2.8

Table A7 – Genes ES/DE under pH 6.5 and 7.0 conditions only in primary root attachment (data for primary root attachment pH 7.5 conditions not shown) which were also NE in the input library (no shown).

Gene	Description	G21	S21	G1	S1	VMM	TY	Rhi	RA 6.5	RA 7.0	Col	IT	Nod
pRL110043	putative transmembrane transporter protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL1661 (pssA2)	putative transmembrane suADr transferase protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2520	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	ES	DE	NE	NE	NE
RL3277	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL4335	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2284 (hfq)	putative host factor protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	ES	NE	NE
pRL100242	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	NE	NE
pRL100220	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	ES	DE	NE	DE	NE
RL2227	putative transmembrane protease	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	NE
RL4309	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	ES	ES
RL1600 (ppx)	putative exopolyphosphatase	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	ES	ES
pRL100161 (nifD)	nitrogenase molybdenum-iron protein alpha chain NifD	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	ES	DE
pRL110072	putative GntR family transcriptional regulator	NE	NE	DE	DE	NE	NE	NE	DE	DE	NE	NE	NE
pRL110565	putative transmembrane protein	ES	NE	ES	NE	NE	NE	NE	DE	DE	NE	NE	NE
pRL70037	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	NE	NE
pRL70166	conserved hypothetical protein	DE	DE	DE	DE	ES	NE	NE	DE	DE	NE	DE	DE
RL0123 (truB)	putative tRNA pseudouridine synthase B (tRNA pseudouridine 55 synthase) (Psi55 synthase) (pseudouridylate synthase) (uracil hydrolyase)	DE	DE	NE	DE	NE	NE	NE	DE	DE	NE	DE	ES
RL0186	putative permease component of ABC transporter PepT (S. mel SBP homologue SMc02832 induced by taurine, valine, isoleucine)	DE	DE	DE	DE	NE	NE	NE	DE	DE	NE	ES	ES
RL0188	putative ATP-binding:ATP-binding (ABC:ABC) componentof ABC transporter PepT (S. mel SBP homologue SMc02832 induced by taurine, valine, isoleucine)	DE	DE	DE	DE	NE	NE	NE	DE	DE	NE	NE	NE
RL0885	putative hydrolase	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	ES	DE
RL1628	hypothetical protein	NE	NE	NE	NE	ES	NE	NE	DE	DE	ES	ES	NE
RL2405	putative peptidyl-prolyl cis-trans isomerase B (cyclophilin-related protein)	DE	NE	NE	NE	NE	NE	NE	DE	DE	NE	ES	DE

RL2514	conserved hypothetical protein	DE	NE	NE	NE	NE	NE	DE	DE	DE	DE	NE	NE
RL2638 (rbsK3)	putative ribokinase	ES	ES	NE	DE	NE	NE	DE	DE	DE	DE	NE	NE
RL3294	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	NE	NE
RL3462	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	NE	DE
RL3677 (IspL)	putative UDP-glucuronate 5'-epimerase	NE	NE	NE	DE	NE	NE	NE	DE	DE	DE	ES	ES
RL4322 (tlpA)	putative thiol:disulfide interchange protein	NE	NE	NE	NE	ES	NE	NE	DE	ES	ES	NE	NE
RL4675	hypothetical protein	ES	ES	ES	ES	NE	NE	NE	DE	ES	NE	ES	ES

INSeq experiments are coded as follows: G21 = 21% oxygen, 10 mM glucose. S21 = 21 % oxygen, 20 mM succinate. G1 = 1% oxygen, 10 mM glucose. S1 = 1% oxygen, 20 mM succinate. VMM = Vincents media. TY = Tryptone yeast media. Rhizo = rhizosphere. RA 6.5 = root attachment, pH 6.5. RA 7.0 = root attachment, pH 7.0. Colon = 5 day root colonisation. IT = infection thread. Nod = nodulation. Thick red line = pleiotropy filter cutoff, as described in 4.2.8

Table A8 – Genes ES/DE under pH 7.0 and 7.5 conditions only in primary root attachment (data for primary root attachment pH 6.5 conditions not shown) which were also NE in the input library (not shown).

Gene	Description	G21	S21	G1	S1	VMM	TY	Rhi	RA 7.0	RA 7.5	Col	IT	Nod
RL2642	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2643 (dksA2)	putative DnaK suppressor protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL4083	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL1504	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE
RL0617	putative Maf septum inhibitor protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	NE	NE
RL2212 (clpS1)	putative ATP-dependent Clp protease adaptor protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	DE	DE
RL0141 (cycM)	putative cytochrome c	NE	NE	NE	NE	NE	NE	NE	DE	ES	NE	ES	ES
pRL110615 (nadC)	putative nicotinate-nucleotide pyrophosphorylase [carboxylating]	DE	DE	NE	DE	NE	NE	NE	DE	DE	NE	NE	NE
pRL110616 (nadB)	putative L-aspartate oxidase	DE	DE	NE	DE	NE	NE	NE	DE	DE	NE	NE	NE
pRL110617 (nadA)	quinolinate synthetase A	DE	DE	NE	DE	NE	NE	NE	DE	DE	NE	NE	NE
pRL110618	conserved hypothetical protein	DE	DE	NE	DE	NE	NE	NE	DE	DE	NE	NE	NE
pRL90143	putative transposase	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	DE
RL0945 (aroA2)	putative 3-phosphoshikimate 1-carboxyvinyltransferase	ES	DE	DE	NE	ES	NE	NE	ES	DE	NE	ES	NE
RL0956	putative para-hydroxybenzoate--polyprenyltransferase	NE	DE	NE	DE	NE	NE	AD	DE	DE	AD	NE	NE
RL0957	conserved hypothetical protein	NE	DE	NE	DE	NE	NE	DE	DE	DE	ES	ES	ES
RL1340 (sodB)	putative superoxide dismutase	NE	NE	NE	NE	ES	NE	NE	ES	DE	DE	ES	NE
RL1617	putative transmembrane protein	ES	ES	ES	ES	NE	NE	NE	DE	DE	NE	DE	DE
RL2213 (clpA)	putative ATP-dependent Clp protease ATP-binding subunit	NE	NE	NE	DE	NE	NE	NE	DE	DE	NE	ES	NE
RL2087	putative leucyl/phenylalanyl-tRNA--protein transferase	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	NE
RL2495 (moaC)	putative molybdenum cofactor biosynthesis protein C	ES	ES	DE	ES	ES	NE	NE	DE	DE	NE	NE	NE
RL2776	putative AsnC family transcriptional regulator	NE	NE	NE	NE	ES	NE	NE	DE	DE	DE	NE	NE
RL4017 (rpmE)	putative 50S ribosomal protein L31	DE	DE	DE	DE	NE	NE	NE	DE	DE	NE	DE	DE
RL4074	conserved hypothetical protein	DE	DE	NE	DE	NE	NE	DE	DE	DE	NE	NE	NE
RL4295	putative acetyltransferase	DE	DE	DE	DE	NE	NE	NE	DE	DE	NE	ES	ES
RL4296 (argJ)	putative arginine biosynthesis bifunctional protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	ES	ES

INSeq experiments are coded as follows: G21 = 21% oxygen, 10 mM glucose. S21 = 21 % oxygen, 20 mM succinate. G1 = 1% oxygen, 10 mM glucose. S1 = 1% oxygen, 20 mM succinate. VMM = Vincents media. TY = Tryptone yeast media. Rhizo = rhizosphere. RA 7.0 = root attachment, pH 7.0. RA 7.5 = root attachment, pH 7.5. Colon = 5 day root colonisation. IT = infection thread. Nod = nodulation. Thick red line = pleiotropy filter cutoff, as described in 4.2.8

Table A9 – Genes ES/DE under pH 6.5 and 7.5 conditions only in primary root attachment (data for primary root attachment pH 7.0 conditions not shown) which were also NE in the input library (not shown).

Gene	Description	G21	S21	G1	S1	VMM	TY	Rhi	RA 6.5	RA 7.5	Col	IT	Nod
RL1805	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL1806 (degQ)	putative protease DegQ precursor	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2477 (dacF)	putative penicillin-binding protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
pRL100162A	hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	0	0
pRL100163	hypothetical protein with homology to asparagine synthase (glutamine hyrdrolysing) at C-term	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	NE	NE
pRL110045	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	NE	NE
pRL100470	hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	ES	ES
pRL110046	putative FNR/CRP family transcriptional regulator	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	ES	ES
RL2098	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	AD	NE
pRL100275	putative LacI family transcriptional regulator (repressor)	DE	DE	DE	DE	DE	NE	DE	DE	DE	DE	NE	NE
pRL100404 (mctP)	putative transmembrane sodium-solute symporter	NE	NE	NE	NE	AD	NE	AD	DE	DE	NE	NE	NE
pRL70099	putative phage-derived invertase/resolvase	ES	DE	ES	DE	NE	NE	NE	DE	DE	NE	NE	NE
RL0618	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	NE	DE
RL0921	putative cationic transport protein, CorA family	DE	DE	NE	NE	NE	NE	DE	DE	DE	DE	DE	DE
RL1731 (rpmG)	putative 50S ribosomal protein L33	DE	DE	DE	DE	NE	NE	NE	DE	DE	NE	DE	DE
RL2208	putative hydrolase	NE	DE	NE	NE	NE	NE	DE	DE	DE	NE	AD	AD
RL2209 (cysE1)	putative serine acetyltransferase	DE	DE	DE	ES	DE	NE	DE	DE	DE	DE	ES	DE
RL2210	conserved hypothetical protein	DE	DE	DE	ES	DE	NE	DE	DE	DE	DE	NE	DE
RL2389	conserved hypothetical exported protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	NE	NE	NE
RL4040 (thiE2)	putative thiamine-phosphate pyrophosphorylase	ES	DE	DE	DE	NE	NE	DE	DE	ES	DE	ES	ES
RL4117 (glAD)	putative glycogen synthase	DE	NE	DE	NE	DE	NE	NE	DE	DE	NE	ES	ES

INSeq experiments are coded as follows: G21 = 21% oxygen, 10 mM glucose. S21 = 21 % oxygen, 20 mM succinate. G1 = 1% oxygen, 10 mM glucose. S1 = 1% oxygen, 20 mM succinate. VMM = Vincents media. TY = Tryptone yeast media. Rhizo = rhizosphere. RA 6.5 = root attachment, pH 6.5. RA 7.5 = root attachment, pH 7.5 Colon = 5 day root colonisation. IT = infection thread. Nod = nodulation. 0 = data not available. Thick red line = pleiotropy filter cutoff, as described in 4.2.8

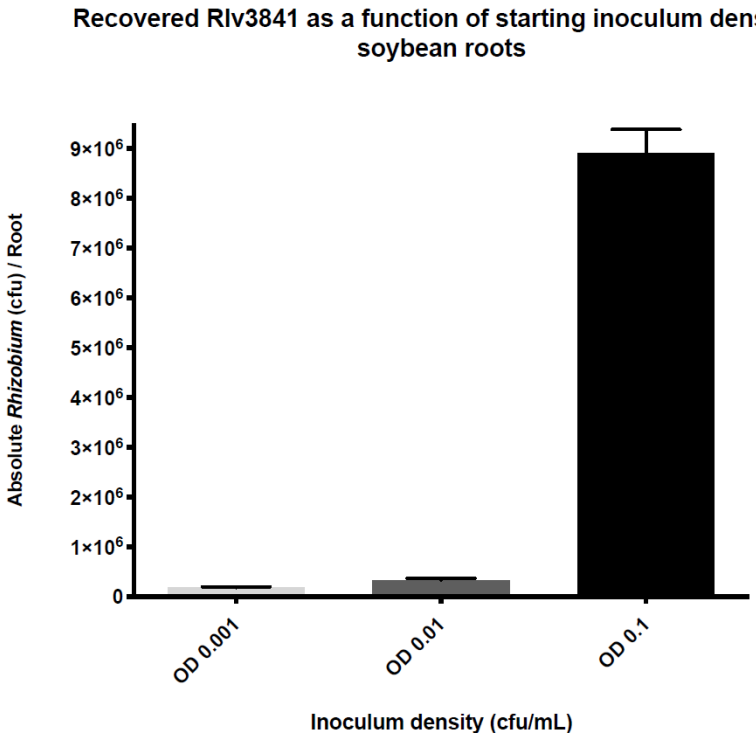


Figure A1. Recovered Rlv3841 as a function of starting inoculum density after one hour attachment to soybean roots under pH 7 conditions. Inoculum density is given as OD (600). N = 5 for all groups

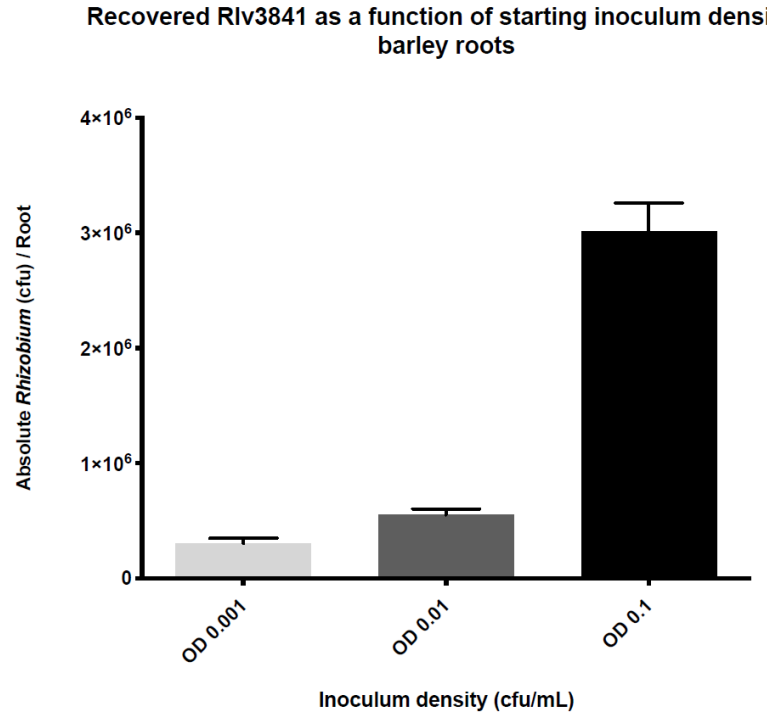


Figure A2. Recovered Rlv3841 as a function of starting inoculum density after one hour attachment to barley roots under pH 7 conditions. Inoculum density is given as OD (600). N = 5 for all groups

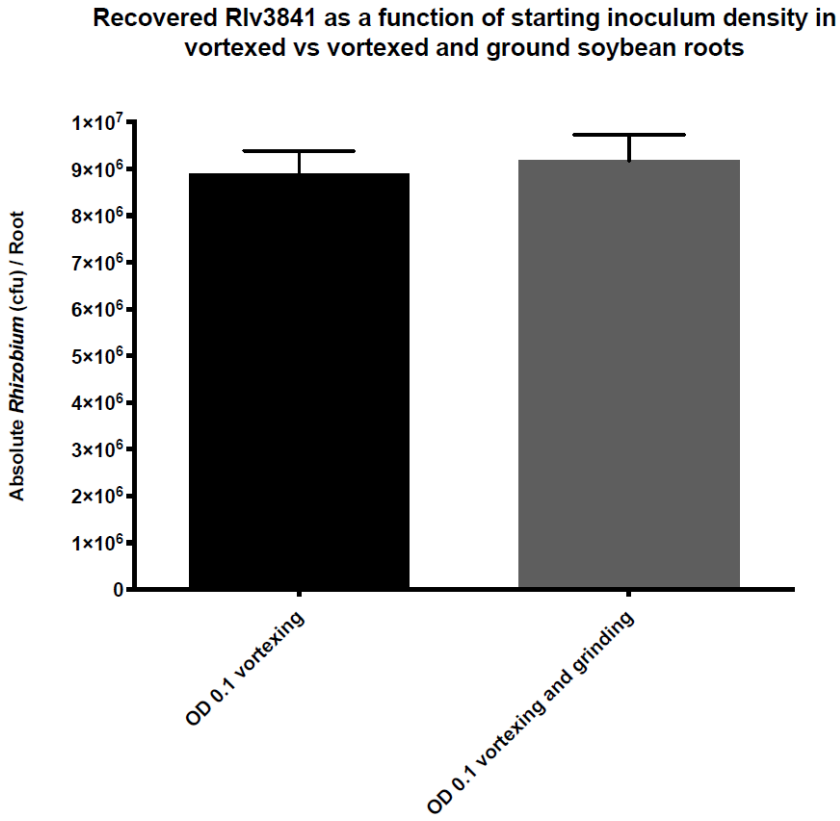


Figure A3. Recovered Rlv3841 by vortexing or vortexing and grinding using a starting inoculum density of OD (600) 0.1 and a one hour attachment assay to soybean roots at pH 7. Inoculum density is given as OD (600). N = 5 for all groups

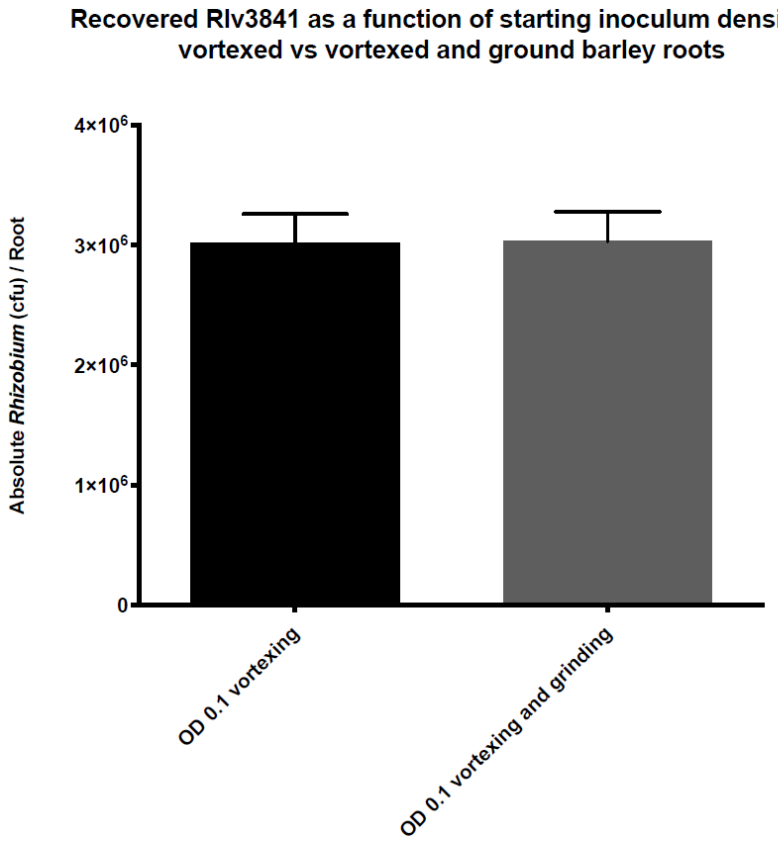


Figure A4. Recovered Rlv3841 by vortexing or vortexing and grinding using a starting inoculum density of OD (600) 0.1 and a one hour attachment assay to barley roots at pH 7. Inoculum density is given as OD (600). N = 5 for all groups

Table A10 – Genes ES/DE in primary root attachment to pea, soybean and barley roots which were also NE in the input library (not shown).

Gene	Description	G21	S21	G1	S1	VMM	TY	Rhi	Pea 7.0	Soy 7.0	Bar 7.0	Col	IT	Nod
pRL110043	putative transmembrane transporter protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE	NE
RL2643 (dksA2)	putative DnaK suppressor protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE	NE
RL4145	putative LacI family transcriptional regulator (repressor)	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE	NE
RL4381	putative cell surface protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE	NE
RL0551 (hslO)	putative Hsp33-like chaperonin	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	NE	NE
pRL100112	putative dehalogenase/hydrolase	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	NE	NE
RL0617	putative Maf septum inhibitor protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	NE	NE
RL4362	putative cobalamin synthesis protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	DE	DE
RL4363 (dacC)	putative penicillin-binding protein precursor	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	DE	DE
RL2694 (gor)	putative glutathione reductase	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	ES	ES
RL2212 (clpS1)	putative ATP-dependent Clp protease adaptor protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	DE	DE
RL3322 (pfp)	putative pyrophosphate--fructose 6-phosphate 1-phosphotransferase	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	DE	NE
RL1013	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	ES	NE
RL1478 (amn)	putative AMP nucleosidase	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	ES	ES
RL1600 (ppx)	putative exopolyphosphatase	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	ES	ES
RL2284 (hfq)	putative host factor protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	ES	NE	NE
RL2513 (tpiA)	putative triosephosphate isomerase	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	NE	NE
RL2637 (recA)	putative recombinase	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	ES	ES
RL3987	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	DE	DE
RL3988	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	DE	DE
RL3989 (ruvA)	putative Holliday junction DNA helicase RuvA	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	DE	DE
RL3990 (ruvB)	putative Holliday junction DNA helicase RuvB	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	DE	DE

RL4065	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	NE	NE
RL2776	putative AsnC family transcriptional regulator	NE	NE	NE	NE	ES	NE	NE	DE	DE	DE	DE	NE	NE
pRL110044	conserved hypothetical protein	NE	NE	NE	NE	DE	NE	DE	DE	DE	DE	DE	NE	NE
pRL110617 (nadA)	quinolinate synthetase A	DE	DE	NE	DE	NE	NE	NE	DE	DE	DE	NE	NE	NE
pRL110618	conserved hypothetical protein	DE	DE	NE	DE	NE	NE	NE	DE	DE	DE	NE	NE	NE
pRL70055	hypothetical protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	ES	DE
pRL90143	putative transposase	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	DE	DE
RL0160 (polA)	putative DNA polymerase I	DE	DE	NE	DE	NE	NE	NE	DE	DE	DE	NE	NE	DE
RL0187	putative permease component of ABC transporter PepT	DE	DE	DE	DE	NE	NE	NE	DE	DE	DE	NE	ES	ES
RL0430 (def1)	putative peptide deformylase (polypeptide deformylase)	DE	DE	DE	DE	DE	NE	DE	DE	DE	DE	DE	ES	DE
RL0431	putative plasmid stability protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	ES	DE
RL0432	putative plasmid stability protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	ES	DE
RL4546	conserved hypothetical protein	ES	DE	DE	NE	NE	NE	NE	DE	DE	DE	NE	NE	NE
RL4582 (fbpC)	putative ATP-binding component of ABC transporter Unclass ferric cations transporter	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	DE	DE
RL4583 (fbpB)	putative permease component of ABC transporter Unclass ferric cations transporter	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	DE	DE
RL4584 (fbpA)	putative SBP of ABC transporter Unclass ferric cations transporter	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	NE	NE	NE
RL4675	hypothetical protein	ES	ES	ES	ES	NE	NE	NE	ES	ES	ES	NE	ES	ES
RL0563	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	NE	DE	DE	DE	NE	NE	NE
RL0564 (mcpE)	putative methyl-accepting chemotaxis protein	DE	DE	DE	DE	NE	NE	NE	DE	DE	DE	NE	NE	NE
RL0855 (gshA)	putative ADmma-glutamylcysteine synthetase precursor	DE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	ES	ES
RL0956	putative para-hydroxybenzoate--polyprenyltransferase	NE	DE	NE	DE	NE	NE	AD	DE	DE	DE	AD	NE	NE
RL0957	conserved hypothetical protein	NE	DE	NE	DE	NE	NE	DE	DE	DE	DE	ES	ES	ES
RL1507	conserved hypothetical protein	NE	NE	NE	DE	NE	NE	DE	DE	DE	DE	DE	NE	DE
RL1508	putative transmembrane protein	NE	NE	NE	DE	NE	NE	DE	DE	DE	DE	NE	ES	ES
RL1572	putative DNA polymerase III subunit	DE	DE	DE	DE	NE	NE	DE	DE	DE	ES	DE	ES	ES

RL1617	putative transmembrane protein	ES	ES	ES	ES	NE	NE	NE	DE	DE	DE	NE	DE	DE
RL1618	putative MarR family transcriptional regulator	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	NE	DE	DE
RL2087	putative leucyl/phenylalanyl-tRNA--protein transferase	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	DE	NE
RL2213 (clpA)	putative ATP-dependent Clp protease ATP-binding subunit	NE	NE	NE	DE	NE	NE	NE	DE	DE	DE	NE	ES	NE
RL2405	putative peptidyl-prolyl cis-trans isomerase B (cyclophilin-related protein)	DE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	ES	DE
RL2406 (queA)	putative S-adenosylmethionine:tRNA ribosyltransferase-isomerase	DE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	ES	DE
RL2407 (tgt)	putative queuine tRNA-ribosyltransferase	DE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	ES	DE
RL2473 (metG)	putative methionyl-tRNA synthetase	DE	DE	DE	NE	NE	NE	NE	DE	DE	DE	NE	ES	NE
RL2474	putative transmembrane protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	NE	NE	NE
RL2478	putative outer membrane protein	NE	DE	NE	DE	NE	NE	NE	DE	DE	DE	NE	NE	NE
RL2514	conserved hypothetical protein	DE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	NE	NE
RL2638 (rbsK3)	putative ribokinase	ES	ES	NE	DE	NE	NE	DE	DE	DE	DE	DE	NE	NE
RL2652 (fdxB2)	putative ferredoxin, 2Fe-2S	DE	DE	DE	DE	NE	NE	NE	DE	DE	DE	NE	NE	NE
RL2653	putative HPt (Histidine Phosphotransfer) domain protein of two-component system (multistage phosphorelay can result)	DE	DE	DE	DE	NE	NE	NE	DE	DE	DE	NE	NE	NE
RL3294	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	NE	NE
RL3596	conserved hypothetical protein	DE	DE	NE	NE	NE	NE	DE	DE	DE	DE	NE	DE	NE
RL3597	putative DEAD-box ATP-dependent RNA helicase protein	DE	DE	NE	NE	NE	NE	DE	DE	DE	DE	DE	DE	NE
RL4017 (rpmE)	putative 50S ribosomal protein L31	DE	DE	DE	DE	NE	NE	NE	DE	DE	DE	NE	DE	DE
RL4073 (hss)	putative homospermidine synthase	DE	DE	NE	DE	NE	NE	DE	DE	DE	ES	NE	ES	ES
RL4296 (argJ)	putative arginine biosynthesis bifunctional protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	ES	ES
RL3677 (lspL)	putative UDP-glucuronate 5'-epimerase	NE	NE	NE	DE	NE	NE	NE	DE	DE	DE	DE	ES	ES

INSeq experiments are coded as follows: G21 = 21% oxygen, 10 mM glucose. S21 = 21 % oxygen, 20 mM succinate. G1 = 1% oxygen, 10 mM glucose. S1 = 1% oxygen, 20 mM succinate. VMM = Vincents media. TY = Tryptone yeast media. Rhizo = rhizosphere. Pea 7.0 = pea root attachment, pH 7.0. Soy 7.0 = soybean root attachment, pH 7.0. Bar 7.0 = barley root attachment, pH 7.0. Colon = 5 day pea root colonisation. IT = pea infection thread. Nod = pea nodulation. Thick red line = pleiotropy filter cutoff, as described in 5.2.8

Table A11 – Genes ES/DE in primary root attachment to pea only (data for primary root attachment to soybean and barley not shown) which were also NE in the input library (not shown).

Gene	Description	G21	S21	G1	S1	VMM	TY	Rhi	Pea 7.0	Col	IT	Nod
RL4704	putative glyoxalase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3752 (pssA)	putative transferase involved in exopolysaccharide biosynthesis	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3226	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2587	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1661 (pssA2)	putative transmembrane suADr transferase protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2044 (scpA)	putative chromosome segreADtion and condensation protein A	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL120518	putative TetR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL100174	hypothetical protein, no known homology	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1504	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
RL1106 (pspA)	putative (phage shock protein A) PspA family regulator by protein-protein interactions	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
pRL100242	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE
pRL100053	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	ES	DE	DE
RL0141 (cycM)	putative cytochrome c	NE	NE	NE	NE	NE	NE	NE	DE	NE	ES	ES
RL2588 (tyrS1)	putative tyrosyl-tRNA synthetase	NE	NE	NE	NE	NE	NE	NE	DE	NE	ES	ES
RL0417 (ihfB)	putative integration host factor beta subunit	NE	NE	NE	NE	ES	NE	ES	DE	NE	NE	NE
RL1105	putative TetR family transcriptional regulator	NE	NE	NE	NE	ES	NE	DE	DE	NE	NE	ES
RL1585 (moaD)	putative molybdopterin converting factor subunit D	NE	NE	NE	NE	ES	NE	NE	DE	NE	NE	NE
RL1628	hypothetical protein	NE	NE	NE	NE	ES	NE	NE	DE	ES	ES	NE
pRL100161 (nifD)	nitrogenase molybdenum-iron protein alpha chain NifD	DE	DE	DE	DE	NE	NE	DE	DE	DE	ES	DE
pRL100388	putative LacI family transcriptional regulator (repressor)	NE	DE	NE	DE	NE	NE	NE	DE	NE	NE	NE
pRL110072	putative GntR family transcriptional regulator	NE	NE	DE	DE	NE	NE	NE	DE	NE	NE	NE
pRL110073	putative GntR family transcriptional regulator	DE	DE	DE	DE	NE	NE	NE	DE	NE	NE	NE

pRL110389	putative exopolysaccharide production protein	DE	DE	DE	DE	NE	NE	NE	ES	NE	ES	ES
pRL70037	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	NE	NE
pRL70166	conserved hypothetical protein	DE	DE	DE	DE	ES	NE	NE	DE	NE	DE	DE
pRL90040	pseudogene, conserved hypothetical protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	0	0
pRL90280	conserved hypothetical protein	ES	ES	DE	DE	ES	NE	NE	ES	NE	NE	NE
RL1001	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	NE	DE	DE	NE	NE
RL1002 (bioY)	putative transmembrane biotin biosynthesis protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	ES	ES
RL1003	putative permease component of ABC transporter Unclass	DE	DE	DE	DE	NE	NE	DE	DE	DE	ES	ES
RL1004	putative ATP-binding component of ABC transporter Unclass	DE	DE	DE	DE	NE	NE	DE	DE	DE	ES	ES
RL1389	conserved hypothetical protein	DE	DE	NE	DE	NE	NE	NE	DE	NE	ES	NE
RL1390	hypothetical protein	NE	DE	NE	DE	NE	NE	NE	DE	NE	NE	NE
RL2495 (moaC)	putative molybdenum cofactor biosynthesis protein C	ES	ES	DE	ES	ES	NE	NE	DE	NE	NE	NE
RL2696	putative acylphosphatase	ES	DE	ES	ES	NE	NE	NE	DE	NE	NE	NE
RL2953	putative adenylate/guanylate cyclase	DE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3872 (purC2)	putative phosphoribosylamidoimidazole-succinocarboxamide synthase	DE	DE	DE	DE	DE	NE	NE	DE	NE	NE	NE
RL4295	putative acetyltransferase	DE	DE	DE	DE	NE	NE	NE	DE	NE	ES	ES
RL4304	putative transmembrane transporter protein	NE	ES	DE	NE	NE	NE	NE	DE	NE	NE	NE

INSeq experiments are coded as follows: G21 = 21% oxygen, 10 mM glucose. S21 = 21 % oxygen, 20 mM succinate. G1 = 1% oxygen, 10 mM glucose. S1 = 1% oxygen, 20 mM succinate. VMM = Vincents media. TY = Tryptone yeast media. Rhizo = rhizosphere. Pea 7.0 = pea root attachment, pH 7.0. Colon = 5 day pea root colonisation. IT = pea infection thread. Nod = pea nodulation. 0 = data not available. Thick red line = pleiotropy filter cutoff, as described in 5.2.8

Table A12 – Genes ES/DE in primary root attachment to soybean only (data for primary root attachment to pea and barley not shown) which were also NE in the input library (not shown).

Gene	Description	G21	S21	G1	S1	VMM	TY	Rhi	Soy 7.0	Col	IT	Nod
pRL110197	pseudogene	NE	NE	NE	NE	NE	NE	NE	DE	NE	0	0
pRL110205 (hutH1)	putative histidine ammonia-lyase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL110206 (hutI)	putative imidazolonepropionase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL110207	putative histidine degradation related amidohydrolase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL110213	putative permease component of ABC transporter PAAT	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL110214	putative ATP-binding component of ABC transporter PAAT	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL110393	hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL110394	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL110395	putative succinoglycan biosynthesis protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL110439	putative glycosyl transferase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL110440	putative acetyltransferase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL110453	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL110541	putative glycine degradation aminomethyltransferase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL110569	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL110572	putative flp pilus assembly protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL110573	putative transmembrane pilus component protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL110574	putative transmembrane pilus component protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL110575	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL120056 (mcpR)	putative methylation accepting chemotaxis protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL120074	putative ATP-binding component of ABC transporter PAAT (S. mel SBP homologue SMb21135 induced by ADlactosamine, glucosamine)	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL120075 (stbC)	putative plasmid stability protein StbC	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL120487	putative salicylaldehyde dehydrogenase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE

pRL120488	putative 3-alpha-hydroxysteroid dehydrogenase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL120489	putative permease component of ABC transporter HAAT	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL120490	putative permease component of ABC transporter HAAT	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL120491	putative ATP-binding component of ABC transporter HAAT	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL90242	putative nitrilotriacetate monooxygenase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL0229	putative LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL0230	putative peptidase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL0376	putative L-lysine dehydrogenase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL0391 (int)	putative transmembrane acyltransferase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL0589	conserved hypothetical protien	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL0590	putative short-chain dehydrogenase/reductase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL0664	putative transmembrane acyltransferase (lipid metabolism)	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL0762	putative XRE family (HipB) transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL0909	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL0910 (mutL)	putative DNA mismatch repair protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL0964	putative Mg2+ transporter protein (uptake and/or export, other ions can be transported), CorA family	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1015	putative polysaccharide deacetylase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1016	putative D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein)	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1033	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1034	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1038	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1039	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1108	putative transmembrane AMP-binding acyltransferase family protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1154	conserved hypothetical exported protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1155	putative glycosyl transferase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE

RL1162	putative two-component sensor/regulator; transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1163	putative two-component sensor/regulator; histidine kinase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1549	putative transmembrane efflux protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1644	conserved hypothetical exported protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1918	putative exported aryl-sulfatase protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1919	putative protein with alpha/beta hydrolase fold (DUF900)	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2485	putative transmembrane transporter protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2486	putative LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2505 (deaD)	putative cold-shock DEAD-box protein A (ATP-dependent RNA helicase DEAD)	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2506 (cyaA12)	putative adenylate cyclase 1 (ATP pyrophosphate-lyase 1) (adenylyl cyclase 1)	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2521	putative IAA acetyltransferase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2562	putative DoxD family transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2563 (pccB)	putative propionyl-CoA carboxylase beta subunit	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2570	conserved hypothetical exported protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2571	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2572 (gcvT)	putative aminomethyltransferase (glycine cleavage system t protein)	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2764	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2765	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2766	putative ArsR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2767 (ssuB)	putative ATP-binding component of ABC transporter NitT aliphatic sulfonates	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2768 (ssuC)	putative permease component of ABC transporter NitT aliphatic sulfonates	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2779	putative two-component sensor/regulator; histidine kinase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2875	putative ion efflux protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2876	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE

RL2877	putative transmembrane transporter protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2900 (dhaK1)	putative dihydroxyacetone kinase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2901 (dhaL1)	putative PTS-dependent dihydroxyacetone kinase, ADP-binding subunit	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2902	putative PTS-dependent dihydroxyacetone kinase, phosphotransferase subunit	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2903 (ptsH)	putative multiphosphoryl transfer/phosphocarrier protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2904 (ptsI)	putative phosphoenolpyruvate-protein phosphotransferase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2999	putative glyoxylase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3000	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3006 (rhbB)	putative L-2,4-diaminobutyrate decarboxylase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3146	putative LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3147	putative transmembrane transporter protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3149	putative adenylate cyclase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3192	putative ATP-binding component of ABC transporter Unclass	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3193	putative permease component of ABC transporter Unclass	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3262 (mocA)	putative MocA family oxidoreductase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3263	putative LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3264	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3265	putative AraC family transcriptional regulator (activator)	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3268 (flaH)	putative flagellin protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL4120 (nodN2)	putative nodulation protein NodN2	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL4189	putative LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL4219	putative DeoR family transcriptional regulator (repressor) of sorbitol/mannitol operon	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL4220 (exoD)	putative exopolysaccharide biosynthesis protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL4604	putative GntR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE

pRL100385 (ecfM)	putative RNA polymerase ECF sigma factor, family ECF20/ECF01	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE
pRL100183 (nodF)	nodulation protein NodF acyl carrier protein (ACP) used in Nod factor synthesis	NE	NE	NE	NE	NE	NE	NE	DE	NE	DE	DE
pRL110045	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE
pRL110046	putative FNR/CRP family transcriptional regulator	NE	NE	NE	NE	NE	NE	DE	DE	DE	ES	ES
pRL110048 (redAh)	UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	DE
pRL110049 (redB)	putative glycosyltransferase	NE	NE	NE	NE	NE	NE	DE	DE	NE	ES	DE
pRL110283	putative ArsR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	NE	ES	DE
pRL110441 (thiD)	putative phosphomethylpyrimidine kinase	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
pRL110560 (soxB2)	putative sarcosine oxidase beta subunit	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
pRL110561 (soxD2)	putative sarcosine oxidase delta subunit	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
pRL110570	hypothetical exported protein	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
pRL110571	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
pRL120057	putative plasmid stability protein	NE	NE	NE	NE	NE	NE	ES	DE	DE	NE	NE
pRL120058	hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
pRL120306	putative ATP-binding component of ABC transporter NitT	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
pRL80133 (trbLp8)	putative conjuADl transfer protein TrbL	NE	NE	NE	NE	NE	NE	ES	DE	ES	NE	NE
pRL80134 (trbHp8)	putative conjuADl transfer protein TrbH	NE	NE	NE	NE	NE	NE	ES	DE	ES	NE	NE
pRL80135 (trbGp8)	putative conjuADl transfer protein TrbG	NE	NE	NE	NE	NE	NE	ES	DE	ES	NE	NE
pRL90243	putative acyl-CoA dehydrogenase	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
pRL90298	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	AD	NE
RL0338 (gshB)	putative glutathione synthetase	NE	NE	NE	NE	NE	NE	ES	DE	NE	ES	ES
RL0397 (mur)	putative FUR-like transcriptional regulator, iron response regulator	NE	NE	NE	NE	NE	NE	NE	DE	NE	DE	DE
RL0482A	putative short-chain dehydrogenase/oxidoreductase	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL0685 (hemAT)	putative chemoreceptor protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	ES	ES
RL0686 (cheX1)	putative chemotaxis related CheX protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	ES	ES

RL0687 (cheY1)	putative two-component sensor/regulator; chemotaxis transcriptional regulator CheY	NE	NE	NE	NE	NE	NE	NE	DE	NE	ES	ES
RL0962	putative ring hydroxylating dioxygenase subunit	NE	NE	NE	NE	NE	NE	NE	DE	NE	ES	ES
RL0963	putative transmembrane/surface protein BA14K-like immuno reactive protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
RL1040	putative LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
RL1089	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	AD	NE
RL1090	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	AD	NE
RL1215	putative transmembrane protein	NE	NE	NE	NE	AD	NE	NE	DE	NE	NE	NE
RL1438 (cycK)	putative cytochrome c-type biogenesis protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	ES	ES
RL1965 (aldR)	putative AsnC family transcriptional regulator	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL1966 (aldA)	putative alanine dehydrogenase	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2051 (pcm)	putative protein-L-isoaspartate O-methyltransferase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	AD
RL2102 (cspA5)	putative cold shock protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	DE	NE
RL2303 (ccdA)	putative cytochrome c-type biogenesis protein	NE	NE	NE	NE	NE	NE	ES	DE	DE	ES	ES
RL2305	hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
RL2394	putative carbohydrate kinase protein	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2507	conserved hypothetical exported protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
RL2508 (gltA2)	putative citrate synthase II	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
RL2564	conserved hypothetical exported protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2569	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	DE	NE
RL2608 (purQ)	putative phosphoribosylformylglycinamidine synthase I	NE	NE	NE	NE	ES	NE	NE	DE	NE	DE	NE
RL2609	conserved hypothetical exported protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	DE	NE
RL2662	putative racemase/isomerase	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
RL2663	putative permease component of ABC transporter Unclass	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
RL2664 (lpxH)	putative UDP-2,3-diacylglucosamine hydrolase	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
RL2780 (dgkA)	putative transmembrane diacylglycerol kinase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2819	hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	NE	AD	NE
RL2820	hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE

RL2827	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	AD	NE
RL2933	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
RL2934	conserved hypothetical exported protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
RL3148	putative ArsR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	AD
RL3196	putative MarR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3198	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
RL3560 (map2)	putative methionine aminopeptidase	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
RL3561	putative bacterial luciferase family protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
RL3592	conserved hypothetical exported protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE
RL4025 (suhB)	putative inositol-1-monophosphatase	NE	NE	NE	NE	ES	NE	NE	DE	NE	ES	NE
RL4030 (cheW3)	putative chemotaxis protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	AD	NE
RL4062	putative amidohydrolase	NE	NE	NE	NE	NE	NE	NE	DE	NE	ES	ES
RL4063	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL4075	putative 5'-nucleotidase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL4162 (eda)	putative 2-dehydro-3-deoxyphosphogluconate aldolase	NE	NE	NE	NE	ES	NE	NE	ES	NE	NE	NE
RL4210 (cysZ)	putative cysteine biosynthesis protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	DE	NE
RL4333	putative phospholipid/glycerol acyltransferase	NE	NE	NE	NE	NE	NE	DE	DE	NE	ES	ES
RL4404 (pssA3)	putative exopolysaccharide production protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL4497	putative transmembrane protein	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	DE
RL4538 (ccmB)	putative permease component of ABC transporter	NE	NE	NE	NE	NE	NE	DE	DE	DE	ES	ES
RL4539 (cycZ)	Export cytochrome c binding export protein putative permease component of ABC transporter	NE	NE	NE	NE	NE	NE	DE	DE	DE	ES	ES
RL4597	Export heme exporter protein c (cytochrome c-type biogenesis protein) conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL4602 (naAD)	putative N-acetylglucosamine-6-phosphate deacetylase	NE	NE	NE	NE	NE	NE	NE	DE	NE	ES	ES
RL4603	putative aminotransferase	NE	NE	NE	NE	NE	NE	NE	DE	NE	ES	ES
RL4618	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE
RL4690	conserved hypothetical protein	DE	ES	DE	ES	NE	NE	NE	DE	DE	NE	NE
RL4411	putative transmembrane protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	AD	NE

RL4413	putative ATP-binding component of ABC transporter PepT	DE	DE	DE	DE	NE	NE	AD	DE	AD	NE	NE
RL4493 (gpsA)	putative glycerol-3-phosphate dehydrogenase [NAD(P)+]	NE	DE	NE	NE	ES	NE	NE	DE	NE	ES	DE
pRL100005	conserved hypothetical protein with PIN domain	ES	NE	ES	DE	NE	NE	NE	DE	NE	NE	NE
pRL100006	conserved hypothetical protein	ES	NE	ES	DE	NE	NE	NE	DE	NE	NE	NE
pRL110012	putative penicillin-binding family protein	DE	NE	NE	DE	NE	NE	NE	DE	NE	NE	NE
pRL120584	putative epimerase	DE	DE	DE	DE	NE	NE	NE	DE	DE	NE	NE
pRL90042	pseudogene, conserved hypothetical protein	DE	DE	NE	NE	NE	NE	NE	DE	NE	0	0
pRL90043	putative permease component of ABC transporter Export	DE	DE	DE	DE	NE	NE	NE	DE	NE	NE	NE
pRL90066	putative transmembrane protein	DE	DE	DE	DE	NE	NE	ES	DE	NE	NE	NE
pRL90067	conserved hypothetical exported protein	DE	DE	DE	DE	NE	NE	AD	DE	NE	NE	NE
pRL90068	putative transmembrane protein	DE	DE	DE	DE	NE	NE	DE	DE	NE	NE	NE
pRL90274	putative transmembrane hydrogenase-related protein	NE	NE	NE	DE	NE	NE	NE	DE	NE	NE	NE
pRL90276	putative ATP-binding :permease (ABC:IMP) component of ABC transporter Export cytochrome bd-related	NE	ES	ES	ES	NE	NE	NE	DE	NE	NE	NE
pRL90277	putative ATP-binding :permease (ABC:IMP) component of ABC transporter Export cytochrome bd-related	ES	ES	ES	ES	NE	NE	ES	DE	NE	NE	NE
RL0501	putative orotate phosphoribosyltransferase	DE	NE	NE	NE	DE	NE	NE	ES	NE	NE	DE
RL0887	putative plasmid stability protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	NE	NE
RL0888	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	NE	NE
RL2208	putative hydrolase	NE	DE	NE	NE	NE	NE	DE	DE	NE	AD	AD
RL2509 (citZ)	putative citrate synthase I	NE	DE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2538 (pip2)	putative proline iminopeptidase	DE	DE	DE	DE	NE	NE	NE	DE	NE	NE	NE
RL2614	conserved hypothetical exported protein	DE	DE	DE	DE	NE	NE	NE	DE	NE	NE	NE
RL2639	putative indigoidine A related protein	ES	NE	NE	DE	NE	NE	NE	DE	AD	NE	NE
RL3005	hypothetical protein	ES	NE	ES	NE	NE	NE	NE	DE	NE	NE	NE
RL3295 (recN)	putative DNA repair protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	NE	NE

RL4031 (mcrA)	putative sensory transducer methyl-accepting chemotaxis protein	DE	NE	NE	DE	NE	NE	NE	DE	NE	AD	NE
RL4032 (mcrB)	putative sensory transducer methyl-acccepting chemotaxis protein	DE	NE	NE	DE	NE	NE	NE	DE	NE	AD	NE
RL4117 (glAD)	putative glycogen synthase	DE	NE	DE	NE	DE	NE	NE	DE	NE	ES	ES

INSeq experiments are coded as follows: G21 = 21% oxygen, 10 mM glucose. S21 = 21 % oxygen, 20 mM succinate. G1 = 1% oxygen, 10 mM glucose. S1 = 1% oxygen, 20 mM succinate. VMM = Vincents media. TY = Tryptone yeast media. Rhizo = rhizosphere. Soy 7.0 = soybean root attachment, pH 7.0. Colon = 5 day pea root colonisation. IT = pea infection thread. Nod = pea nodulation. 0 = data not available. Thick red line = pleiotropy filter cutoff, as described in 5.2.8

Table A13 – Genes ES/DE in primary root attachment to barley only (data for primary root attachment to pea and soybean not shown) which were also NE in the input library (not shown).

Gene	Description	G21	S21	G1	S1	VMM	TY	Rhi	Bar 7.0	Col	IT	Nod
RL0159	putative MarR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL0562	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL0592	putative thiolase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL0593	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL0597	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1647 (celB)	putative cellulose synthase protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1859	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1941	hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2038 (xthA)	putative exodeoxyribonuclease III	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2039	putative HesB family protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2525	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2655	hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2656	putative D-alanyl-D-alanine carboxypeptidase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2657	putative DEEEF/ADF sensory box protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2856 (msiA)	putative transmembrane lysine/arginine export protein family 97% id to MsiA of Mesorhizobium tianshanense induced by canavanine from seed exudates. Exports canavanine (an anti-metabolite) out of cell	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2863	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2864	hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2865	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3252	putative serine protease	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3274 (prkA)	putative PrkA family kinase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3278	hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3279 (pncA)	putative pyrazinamidase/nicotinamidase	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL0401	putative universal stress protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	DE

RL0561	putative AraC family transcriptional regulator (activator)	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE
RL0726	conserved hypothetical exported protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	DE	DE
RL0816	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	ES	ES
RL1805	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL1806 (degQ)	putative protease DegQ precursor	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2644	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2645	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2646	putative transmembrane protein	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2647	conserved hypothetical exported protein	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2857 (msiR)	putative ArsR family transcriptional regulator MsiR, regulates MsiA canavanine (found in seed exudate) exporter	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL2858	conserved hypothetical exported protein	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3253 (hflC)	putative transmembrane serine protease	NE	NE	NE	NE	NE	NE	NE	DE	NE	NE	NE
RL3254 (hflK)	putative transmembrane serine protease	NE	NE	NE	NE	NE	NE	NE	DE	NE	DE	NE
RL3455	putative MarR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	NE	DE	NE
RL3595	putative LacI family transcriptional regulator (repressor)	NE	NE	NE	NE	NE	NE	NE	DE	NE	AD	NE
RL3950	putative phage-related protein	NE	NE	NE	NE	NE	NE	AD	DE	NE	NE	NE
RL4599	putative lysyl-tRNA synthetase homolog	NE	NE	NE	DE	NE	NE	DE	DE	NE	ES	ES
pRL120796	hypothetical exported protein	DE	DE	NE	NE	NE	NE	NE	DE	NE	NE	NE
pRL70099	putative phage-derived invertase/resolvase	ES	DE	ES	DE	NE	NE	NE	DE	NE	NE	NE
pRL90149	putative LuxR/GerE family transcriptional regulator, part of two component response regulator?	ES	DE	ES	ES	NE	NE	NE	DE	DE	NE	NE
RL0566	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	NE	DE	NE	NE	NE
RL0821	putative O-antigen transporter	NE	DE	NE	NE	NE	NE	DE	DE	DE	ES	ES
RL1167	putative TetR family transcriptional regulator	DE	DE	DE	DE	NE	NE	NE	DE	NE	DE	NE
RL2036	putative outer membrane transport protein	DE	DE	DE	DE	NE	NE	NE	DE	NE	NE	AD
RL2152	hypothetical protein	DE	DE	ES	DE	NE	NE	NE	DE	NE	ES	ES
RL2526	putative oxidoreductase	DE	NE	DE	DE	NE	NE	NE	DE	NE	DE	NE
RL2553	conserved hypothetical protein	ES	ES	ES	ES	NE	NE	NE	DE	NE	NE	NE

RL2837	putative SBP of ABC transporter QAT? orphan proline/glycine betaine	DE	DE	DE	DE	NE	NE	NE	DE	DE	NE	NE
RL3498	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	NE	DE	NE	AD	NE

INSeq experiments are coded as follows: G21 = 21% oxygen, 10 mM glucose. S21 = 21 % oxygen, 20 mM succinate. G1 = 1% oxygen, 10 mM glucose. S1 = 1% oxygen, 20 mM succinate. VMM = Vincents media. TY = Tryptone yeast media. Rhizo = rhizosphere. Bar 7.0 = barley root attachment, pH 7.0. Colon = 5 day pea root colonisation. IT = pea infection thread. Nod = pea nodulation. Thick red line = pleiotropy filter cutoff, as described in 5.2.8

Table A14 – Genes ES/DE in primary root attachment to pea and soybean roots (data for primary root attachment to barley not shown) which were also NE in the input library (not shown).

Gene	Description	G21	S21	G1	S1	VMM	TY	Rhi	Pea 7.0	Soy 7.0	Col	IT	Nod
RL4083	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
pRL100220	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	DE	NE
RL1052	hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL1371	putative transmembrane protein	NE	NE	NE	NE	NE	NE	DE	DE	ES	DE	NE	NE
RL2400	putative MarC (multiple antibiotic resistance) family transmembrane protein, they may be transporters	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	AD	DE
RL2520	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
pRL100149	conserved hypothetical protein	NE	NE	NE	NE	ES	NE	NE	ES	ES	NE	NE	NE
pRL110615 (nadC)	putative nicotinate-nucleotide pyrophosphorylase [carboxylating]	DE	DE	NE	DE	NE	NE	NE	DE	DE	NE	NE	NE
pRL110616 (nadB)	putative L-aspartate oxidase	DE	DE	NE	DE	NE	NE	NE	DE	DE	NE	NE	NE
RL0188	putative ATP-binding:ATP-binding (ABC:ABC) componentof ABC transporter PepT (S. mel SBP homologue SMc02832 induced by taurine, valine, isoleucine)	DE	DE	DE	DE	NE	NE	NE	DE	DE	NE	NE	NE
RL1060	putative ribosomal-protein-alanine acetyltransferase	DE	DE	NE	DE	ES	NE	NE	ES	ES	NE	ES	ES
RL1392	putative transmembrane protein	NE	DE	NE	DE	NE	NE	NE	DE	DE	NE	NE	NE
RL3462	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	NE	DE
RL3763	putative rRNA methyltransferase	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	NE	NE
RL4074	conserved hypothetical protein	DE	DE	NE	DE	NE	NE	DE	DE	DE	NE	NE	NE

INSeq experiments are coded as follows: G21 = 21% oxygen, 10 mM glucose. S21 = 21 % oxygen, 20 mM succinate. G1 = 1% oxygen, 10 mM glucose. S1 = 1% oxygen, 20 mM succinate. VMM = Vincents media. TY = Tryptone yeast media. Rhizo = rhizosphere. Pea 7.0 = pea root attachment, pH 7.0. Soy 7.0 = soybean root attachment, pH 7.0. Colon = 5 day pea root colonisation. IT = pea infection thread. Nod = pea nodulation. Thick red line = pleiotropy filter cutoff, as described in 5.2.8

Table A15 – Genes ES/DE in primary root attachment to pea and barley roots (data for primary root attachment to soybean not shown) which were also NE in the input library (not shown).

Gene	Description	G21	S21	G1	S1	VMM	TY	Rhi	Pea 7.0	Bar 7.0	Col	IT	Nod
RL2642	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2695	hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL3277	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
pRL100405	putative transmembrane protein	NE	NE	NE	NE	AD	NE	NE	DE	DE	NE	NE	NE
pRL100406 (mctR)	putative two-component sensor/regulator; transcriptional regulator	NE	NE	NE	NE	AD	NE	NE	DE	DE	NE	NE	NE
RL1340 (sodB)	putative superoxide dismutase	NE	NE	NE	NE	ES	NE	NE	ES	ES	DE	ES	NE
RL1381	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	NE	NE
RL2227	putative transmembrane protease	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	NE
RL3453	putative two-component sensor/regulator; histidine kinase	NE	DE	NE	NE	NE	NE	NE	DE	DE	NE	DE	ES
RL4309	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	ES	ES
RL4382	putative filamentous hemagglutinin adherence factor precursor	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE
pRL70100	hypothetical protein	ES	DE	ES	DE	NE	NE	DE	DE	DE	NE	NE	NE
RL0186	putative permease component of ABC transporter PepT (S. mel SBP homologue SMc02832 induced by taurine, valine, isoleucine)	DE	DE	DE	DE	NE	NE	NE	DE	DE	NE	ES	ES
RL4379 (hemA1)	putative 5-aminolevulinate synthase	NE	NE	NE	NE	ES	ES	DE	DE	DE	DE	ES	ES

INSeq experiments are coded as follows: G21 = 21% oxygen, 10 mM glucose. S21 = 21 % oxygen, 20 mM succinate. G1 = 1% oxygen, 10 mM glucose. S1 = 1% oxygen, 20 mM succinate. VMM = Vincents media. TY = Tryptone yeast media. Rhizo = rhizosphere. Pea 7.0 = pea root attachment, pH 7.0. Soy 7.0 = soybean root attachment, pH 7.0. Colon = 5 day pea root colonisation. IT = pea infection thread. Nod = pea nodulation. Thick red line = pleiotropy filter cutoff, as described in 5.2.8

Table A16 – Genes ES/DE in primary root attachment to soybean and barley roots (data for primary root attachment to pea not shown) which were also NE in the input library (not shown).

Gene	Description	G21	S21	G1	S1	VMM	TY	Rhi	Soy 7.0	Bar 7.0	Col	IT	Nod
RL0162	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL0163 (phnN)	putative phosphonate metabolism protein N, ATP-binding	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL0594	putative fatty oxidation complex subunit	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL0595 (cspA1)	putative cold shock protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL0596	putative MarR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL1111	conserevd hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL1645	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2118	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2287	putative guanine/cytosine deaminase	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2774	putative LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2862	conserved hypothetical exported protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL3261	putative MocA family oxidoreductase	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL3266	conserevd hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL3267	putative OmpA family outer membrane protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL4090	putative lysophospholipase	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL4357 (sirA)	putative two-component sensor/regulator; transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL4380	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL0032	putative phosphocarrier protein HPr for mannose	NE	NE	NE	NE	NE	NE	DE	ES	DE	NE	ES	ES
RL0398	putative acetyltransferase	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	DE	DE
RL0892	putative ribosomal large subunit pseudouridine synthase B	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	DE	DE
RL1012 (cbpA)	putative curved DNA-binding protein	NE	NE	NE	NE	NE	NE	DE	DE	DE	DE	NE	NE
RL1026 (ctaE1)	putative cytochrome c oxidase polypeptide III (cytochrome aa3 subunit 3)	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL1164	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE

RL1165	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL1439 (cycL)	putative cytochrome c-type biogenesis protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	ES	ES
RL1440 (degP1)	putative serine protease	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	ES	ES
RL2285 (hflX)	putative GTP-binding protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2477 (dacF)	putative penicillin-binding protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2595	putative MutT/Nudix family protein (phosphohydrolases)	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2777	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2778	putative exopolysaccharide biosynthesis protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2828	putative XRE family (HipB) family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	ES	ES	DE	ES	ES
RL3320	putative signalling and peptidoglycan binding protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE
RL3321	putative DnaJ family chaperone	NE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE
RL3986 (ruvC)	putative Holliday junction endodeoxyribonuclease RuvC	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	DE	NE
RL4018	putative ATP-binding:permease (ABC:IMP) component of ABC transporter Export	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	DE	DE
RL4354 (xerD)	putative tyrosine recombinase	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	DE	ES
RL4356	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	DE	DE	NE	DE	DE
RL1091	conserved hypothetical protein	DE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE
RL1092	conserved hypothetical protein	DE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE
RL1093	putative beta-lactamase family protein	DE	NE	NE	NE	NE	NE	NE	DE	DE	DE	NE	NE
pRL100010	conserved hypothetical protein	DE	DE	DE	DE	DE	NE	NE	DE	DE	NE	DE	NE
pRL100011	hypothetical protein	DE	DE	DE	DE	NE	NE	NE	DE	DE	DE	NE	NE
pRL100275	putative LacI family transcriptional regulator (repressor)	DE	DE	DE	DE	DE	NE	DE	DE	DE	DE	NE	NE
pRL90144	putative exopolysaccharide biosynthesis-related protein	DE	DE	DE	DE	NE	NE	NE	DE	DE	NE	NE	NE
RL0153	putative penicillin-binding transpeptidase/transglycosylase protein	DE	NE	NE	NE	NE	NE	DE	DE	DE	DE	NE	NE
RL0161	putative cell division DNA translocase protein	DE	DE	DE	DE	NE	NE	NE	DE	DE	NE	NE	NE

RL0226	putative permease component of ABC transporter PepT	DE	DE	DE	DE	NE	NE	DE	DE	DE	NE	ES	ES
RL0227	putative permease component of ABC transporter PepT	DE	DE	DE	DE	NE	NE	DE	DE	DE	NE	ES	ES
RL0228	putative SBP of ABC transporter PepT	DE	DE	DE	DE	NE	NE	DE	DE	DE	NE	ES	ES
RL0423	putative transmembrane protein	NE	DE	NE	DE	NE	NE	DE	DE	DE	DE	NE	AD
RL0546 (phoU)	putative phosphate uptake regulator PhoU, unknown mechanism to regulate expression of high-affinity ABC systems	NE	NE	DE	DE	NE	NE	NE	DE	DE	NE	ES	ES
RL0547 (phoB)	putative two-component sensor/regulator; phosphate regulon transcriptional regulator PhoB	DE	NE	DE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL0565	putative SBP of ABC transporter PAAT (S. mel SBP homologue SMc02219 induced by valine, homoserine, isoleucine)	DE	DE	DE	DE	NE	NE	NE	DE	DE	NE	NE	NE
RL0618	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	NE	DE
RL0818	putative lipopolysaccharide biosynthesis protein	NE	DE	NE	NE	NE	NE	DE	DE	DE	DE	ES	ES
RL0819	putative imidazole glycerol phosphate synthase subunit	NE	DE	NE	NE	NE	NE	DE	DE	DE	DE	ES	ES
RL0820 (hisH2)	putative imidazole glycerol phosphate synthase subunit (igp synthase glutamine amidotransferase subunit)	NE	DE	NE	NE	NE	NE	DE	DE	DE	DE	ES	ES
RL1024 (coxF)	putative cox locus protein	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	ES	ES
RL1025 (ctaG)	putative cytochrome c oxidase assembly protein	DE	DE	DE	DE	NE	NE	NE	DE	DE	NE	NE	NE
RL1166	putative ribonuclease-L-PSP family protein	DE	DE	DE	DE	NE	NE	NE	DE	DE	NE	DE	NE
RL1388	conserved hypothetical protein	NE	DE	NE	DE	NE	NE	NE	DE	ES	NE	NE	NE
RL1548 (radA)	putative DNA repair protein RadA homologue	DE	DE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL1642	putative two-component sensor/regulator; transcriptional regulator	NE	DE	NE	DE	NE	NE	NE	DE	ES	NE	NE	NE
RL1731 (rpmG)	putative 50S ribosomal protein L33	DE	DE	DE	DE	NE	NE	NE	DE	DE	NE	DE	DE
RL4430	putative 3-oxoacyl-[acyl-carrier-protein] reductase	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	DE	NE
RL2209 (cysE1)	putative serine acetyltransferase	DE	DE	DE	ES	DE	NE	DE	DE	DE	DE	ES	DE
RL2210	conserved hypothetical protein	DE	DE	DE	ES	DE	NE	DE	DE	DE	DE	NE	DE
RL2236 (lexA)	putative LexA repressor (SOS regulatory protein)	NE	DE	NE	NE	NE	NE	NE	DE	DE	NE	NE	NE
RL2237	conserved hypothetical protein	DE	DE	DE	DE	NE	NE	NE	DE	DE	NE	NE	NE

RL2510	conserved hypothetical protein	ES	DE	DE	DE	NE	NE	DE	DE	ES	DE	NE	NE
RL2818 (fnrN)	putative FNR/CRP family transcriptional regulator, 100% id to VF39 FnrN	DE	DE	DE	DE	NE	NE	NE	DE	DE	NE	AD	NE
RL3260	putative FAD-dependent dehydrogenase/oxidoreductase	DE	DE	DE	NE	NE	NE	DE	DE	DE	DE	ES	ES
RL3667	putative UDP-glucose 6-dehydrogenase	DE	DE	DE	DE	NE	NE	DE	DE	DE	DE	ES	DE

INSeq experiments are coded as follows: G21 = 21% oxygen, 10 mM glucose. S21 = 21 % oxygen, 20 mM succinate. G1 = 1% oxygen, 10 mM glucose. S1 = 1% oxygen, 20 mM succinate. VMM = Vincents media. TY = Tryptone yeast media. Rhizo = rhizosphere. Soy 7.0 = soybean root attachment, pH 7.0. Bar 7.0 = barley root attachment, pH 7.0. Colon = 5 day pea root colonisation. IT = pea infection thread. Nod = pea nodulation. Thick red line = pleiotropy filter cutoff, as described in 5.2.8

Table A17 – Genes AD in primary root attachment to pea, soybean and barley roots which were also NE in the input library (not shown).

Gene	AD classification conditions	Description	G21	G1	S21	S1	VMM	TY	Rhi	Col	IT	Nod
RL4139	Pea, Bar, Soy	putative transmembrane GGDEF/EAL sensory box protein	NE	NE	NE	NE	NE	NE	NE	NE	GA	GA
pRL70156 (trbGp7)	Pea	putative conjugal transfer protein TrbG	NE	NE	NE	NE	NE	NE	NE	NE	NE	GD
RL3927	Soy	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL3928	Soy	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL3929	Soy	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL3930	Soy	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL4464	Soy	putative lyase/mutase	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
pRL70057	Soy	putative transposase	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL0645	Soy	putative short-chain dehydrogenase/reductase	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL3092	Soy	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL3093	Soy	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL3861	Soy	putative permease component of ABC transporter CUT1	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL3885 (sitB)	Soy	putative ATP-binding component of ABC transporter MZT (S. mel SBP homologue SMc02509 induced by manganese limitation)	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL4003	Bar	conserved hypothetical exported protein	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL3091	Soy, Bar	putative transmembrane protein	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL4724 (rsmB)	Soy, Bar	putative ribosomal RNA small subunit methyltransferase B	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL3090	Soy, Bar	putative methyltransferase	NE	NE	NE	NE	NE	NE	GA	GA	NE	NE
pRL100223	Soy	conserved hypothetical protein	NE	NE	NE	NE	GA	NE	NE	NE	NE	NE
pRL100362	Soy	putative LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
pRL120012	Soy	putative permease component of ABC transporter CUT1	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE

pRL70157 (trbLp7)	Soy	putative conjugal transfer protein TrbL	GD	NE	NE	GD	GA	GA	NE	NE	NE	NE
RL0070	Soy	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL0071 (toaE)	Soy	putative type I export system	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL0076	Soy	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	GA	GA	GA	NE
RL0105	Soy	conserved hypothetical protein	NE	NE	NE	NE	GA	NE	GA	GA	GA	NE
RL0241	Soy	conserved hypothetical protein	NE	NE	NE	NE	NE	GA	NE	NE	NE	NE
RL0639	Soy	putative SBP of ABC transporter CUT1	NE	NE	NE	NE	NE	NE	GA	GA	NE	NE
RL0644 (rbtD)	Soy	putative ribitol 2-dehydrogenase	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL0770	Soy	putative phasin, phasin-2 superfamily	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL0980	Soy	putative metalloprotease M24 family	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL3733	Soy	conserved hypothetical protein	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL3886 (sitC)	Soy	putative permease component of ABC transporter MZT (S. mel SBP homologue SMc02509 induced by manganese limitation)	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL4244	Soy	putative permease component of ABC transporter CUT1	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL4649	Soy	putative permease component of ABC transporter PepT	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL4706 (ilvA2)	Soy	putative threonine dehydratase biosynthetic	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
RL4725	Soy	putative heat-shock peptidase	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE

INSeq experiments are coded as follows: G21 = 21% oxygen, 10 mM glucose. S21 = 21 % oxygen, 20 mM succinate. G1 = 1% oxygen, 10 mM glucose. S1 = 1% oxygen, 20 mM succinate. VMM = Vincents media. TY = Tryptone yeast media. Rhizo = pea rhizosphere. Colon = 5 day pea root colonisation. IT = pea infection thread. Nod = pea nodulation. Thick red line = pleiotropy filter cutoff, as described in 5.2.8. Pea= pea root attachment, pH 7.0. Soy = soybean root attachment, pH 7.0. Bar = barley root attachment, pH 7.0.

Appendix 2. Supplementary material for Chapters 4, 5 and 6

Overview

Chapter 4 – Raw INSeq data: Appendix 2, Table 1

Chapter 5 – Raw INSeq data: Appendix 2, Table 2

These datasets have been deposited electronically in the Oxford University Research Archive (ORA) and may be accessed in the repository using the following link:

https://doi.org/10.5287/bodleian:kZYnK2vQ2

Chapter 6 – Videos 1-5

These videos have been deposited electronically in ORA and may be accessed in the repository using the following link:

https://doi.org/10.5287/bodleian:w4AxBzN4k