Genome-wide identification of colonisation determinants in *Rhizobium leguminosarum* using BarSeq



Hayley Knights

Wadham College

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Trinity Term 2021

Abstract

Genome-wide identification of colonisation determinants in *Rhizobium leguminosarum* using BarSeq

Hayley Knights, Wadham College
Submitted for the degree of DPhil., Trinity Term 2021

Many plant-beneficial bacteria must persist in the rhizosphere and colonise the roots of their host plant to exert their beneficial effects. Plants primarily influence bacterial colonisation of the rhizosphere and roots through modulation of their root exudate composition, whilst bacteria-bacteria interactions affect bacterial persistence and root colonisation through metabolic exchanges, secretion of antimicrobial compounds and other processes. *Rhizobium*-legume symbioses are one of the best characterised beneficial interactions, due to their potential to alleviate our reliance on nitrogen fertilisers. Despite this, most studies on *Rhizobium* have focused on its symbiotic lifestyle as an endosymbiont in root nodules and therefore the initial stages of rhizosphere growth and root colonisation remain relatively under characterised.

The multifactorial nature of bacterial growth in the rhizosphere and root colonisation means that transposon insertion sequencing techniques such as random-barcode transposon-site sequencing (RB-TnSeq) provide a unique opportunity to study gene function at the whole genome level. Through incorporation of a random DNA barcode into transposons RB-TnSeq allows barcode sequencing (BarSeq), which can be used to assay mutant fitness in a high throughput manner. The aim of the work described in this thesis was to use RB-TnSeq to characterise how plant species and bacterial

competition affect rhizobial growth in the rhizosphere and root colonisation using the model rhizobial strain *R. leguminosarum* bv. *viciae* 3841 (Rlv3841).

RB-TnSeq was established in Rlv3841 and initial validation *in planta* demonstrated the suitability of BarSeq for identification of genes contributing to rhizosphere growth and root colonisation. Using BarSeq, phenotypes for 80% of genes encoded by the Rlv3841 genome were assessed during growth in the rhizosphere and colonisation of host legumes (pea, lentil, *Lathyrus*), a non-host legume (alfalfa) and a non-legume (barley). A core set of 73 genes were identified as commonly required for growth in all five plant rhizospheres and a further 58 genes contributed to root colonisation. Analysis of these genes revealed that growth in the rhizosphere and colonisation of roots required the synthesis of compounds (amino acids, ribonucleotides and cofactors), alteration of metabolic function, adaptation to various stresses (such as changes in osmolarity) and sensing of external stimuli coupled with modification of gene expression. Additionally, chemotaxis and flagella-mediated motility were common prerequisites for root colonisation. Notably, many genes showed plant-specific dependencies. In contrast to common determinants, genes implicated in transport accounted for a large proportion of these likely reflecting altered resource availability due to differences in root exudate composition between plant species.

Traditionally, bacterial colonisation has been examined under sterile laboratory conditions that mitigate the influences of bacterial interactions. To identify genes required for competitive rhizosphere growth and root colonisation, an RB-TnSeq mutant library was generated in the beneficial plant growth promoting rhizobacterium *Pseudomonas fluorescens* SBW25 and co-inoculated with the Rlv3841 RB-TnSeq

library. This allowed, for the first time, simultaneous assessment of gene fitness contributions for two different strains by BarSeq and highlighted the need for altered transport, metabolic and regulatory requirements during conditions of competition.

Overall, this work highlights the significant adaption required to different plant rhizospheres and the presence of a competitor, thereby providing a greater understanding of factors promoting rhizosphere fitness and root colonisation in two plant-beneficial bacteria.

Acknowledgments

I would like to thank my supervisor Professor Philip Poole, not only for this opportunity, but for his continual support, belief and guidance over the past years. It has been great working with you and a lot of fun to be a part of your lab. I would like to thank Dr Beatriz Jorrín, whose initial mentorship role evolved into the greatest of friendships, thank you Bea for always being there and taking the time to listen to me. I would also like to thank Dr Vinoy Ramachandran, for his advice, assistance, and patience in answering my many bioinformatics related questions; Dr Alison East, for her continued support and feedback; and Barney Geddes, for introducing me to microbiology. I am so grateful to have had such a great teacher during those initial months.

Thank you to all the Poole Lab members, past and present, for making daily life in the Lab so enjoyable. I feel incredibly lucky to work with a group of people I have come to consider an extended family. A special thank you to Andrzej, Tim and Carmen for their friendship and advice as well as Helen and Lida, for all you do for us to keep things running smoothly.

I would like to thank my friends outside of the Lab, whom without, the past five years in Oxford would not have been the same. Katie and Janina, I met you in my first week here, thank you for all the great memories; my DTP cohort for those many pub nights; Lorna and Danielle and my ‘East Avenue’ family, without whom the last year would not have been the same; Laura, for the many beers and Spanish omelettes; and my housemates, for those late-night chats and keeping me fed during the final stages of

writing. Finally, I thank my family for their continued support over the years particularly my parents, my sister Kelley and my grandparents.

Abbreviations

ABC	ATP binding cassette
AHL	N-Acyl homoserine lactone
Amp	Ampicillin
Amp^R	Ampicillin resistance
ATP	Adenosine triphosphate
BarSeq	Barcoded sequencing
bp	Base pair
BNF	Biological nitrogen fixation
c-di-GMP	Cyclic dimeric guanosine monophosphate
cfu	Colony forming units
CoA	Coenzyme A
COGs	Clusters of Orthologous Groups
CPS	Capsular polysaccharides
CUT	Carbohydrate uptake transporter
DNA	Deoxyribonucleic acid
dpi	Days post inoculation
ECF	Extracytoplasmic function
EPS	Exopolysaccharides
GABA	γ-aminobutyrate
gDNA	Genomic DNA
HAAT	Hydrophobic amino acid transporter
HITS	High-throughput insertion tracking sequencing
HMM	Hidden Markov model

HMP	Hydroxymethylpyrimidine
hpi	hours post inoculation
INSeq	Insertion sequencing
Kan	Kanamycin
Kan^R	Kanamycin resistance
LB	Luria Bertani
LPS	Lipopolysaccharides
MCP	Methyl-accepting chemotaxis protein
MFS	Major facilitator superfamily
MOMPs	Major outer membrane proteins
N_2	Dinitrogen
Neo	Neomycin
Neo^R	Neomycin resistance
NF	Nod factor
NFR	Nod factor receptor
OD	Optical density
PAAT	Polar amino acid transporter
PCR	Polymerase chain reaction
PPP	Pentose phosphate pathway
Raps	*Rhizobium*-adhering proteins
RB-TnSeq	Random barcode transposon-site sequencing
RNA	Ribonucleic acid
ROS	Reactive oxygen species
SBP	Solute binding protein
SEM	Standard error of the mean

SOC	Super optimal broth with catabolic repression
SR	Stringent response
Str	Streptomycin
Str^R	Streptomycin resistance
SynComs	Synthetic communities
T4SS	Type IV protein secretion system
TA	Thymine-adenine
TCA	Tricarboxylic acid cycle
TIRs	Terminal inverted repeats
TIS	Transposon insertion sequencing
TnSeq	Transposon sequencing
TraDIS	Transposon-directed insertion site sequencing
TY	Tryptone yeast
QS	Quorum sensing
UMS	Universal minimal salts
UPP	Unipolar polysaccharides
UV	Ultraviolet
VLCFA	Very-long-chain fatty acids
v/v	Volume of solute/volume of solution
w/v	Mass of solute/volume of solution

Table of Contents

Abstract ..... **iii**

Acknowledgments ..... **vii**

Abbreviations ..... **ix**

Table of Contents ..... **xii**

Chapter 1 ..... **1**

Introduction **1**

1.1. Background 2

1.2. Bacterial root colonisation 3

1.2.1. Chemotaxis towards the root 4

1.2.2. Root attachment 6

1.2.2.1. Primary attachment 6

1.2.2.2. Secondary attachment 8

1.2.2.3. Environmental factors influence root attachment 11

1.2.3. Bacterial biofilms 12

1.3. Plant-bacteria interactions influence colonisation 19

1.4. Bacteria-bacteria interactions influence colonisation 21

1.5. Tool to investigate bacterial colonisation of root-niches 23

1.6. *Rhizobium*-legume symbiosis 27

1.6.1. The importance of *Rhizobium*-legume symbioses 27

1.6.2. Establishment of *Rhizobium*-legume symbioses 28

1.7. Research aims 31

Chapter 2 ..... **34**

Materials and methods **34**

2.1. Media and growth conditions 35

2.1.1. Media 35

2.1.2. Antibiotics 35

2.2. Bacterial strains, plasmids and primers 36

2.3. Molecular techniques 40

2.3.1. Isolation of genomic and plasmid DNA 40

2.3.2. DNA amplification by PCR 40

2.3.3. Agarose gel electrophoresis 40

2.3.4. Restriction endonuclease digestions, ligations and High-Fidelity assembly 41

2.3.5. Transformation of competent cells 41

2.3.6. Conjugal transfer of plasmids 42

2.4. Bacterial root colonisation assays 43

2.4.1. Preparation of growth vessels 43

2.4.2. Seed sterilisation 43

2.4.3. Inoculation of plants 44

2.4.4. Retrieval of rhizosphere and root-associated bacteria 45

2.5. RB-TnSeq 45

2.5.1. Generation of *E.coli* barcoded *mariner* transposon donor pool 45

2.5.2. Generation of RB-TnSeq *mariner* transposon mutant libraries in Rlv3841 and SBW25 47

2.5.3. Mapping of RB-TnSeq libraries using INSeq 48

2.5.4. RB-TnSeq library inoculation *in planta* for BarSeq analysis 49

2.5.5. BarSeq library preparation and sequencing 50

2.5.6. Calculation of gene fitness values from BarSeq experiments 51

Chapter 3 54

Development of RB-TnSeq in Rvl3841 and validation of BarSeq *in planta* 54

3.1. Introduction 55

3.2. Results and discussion 62

3.2.1. Adaption of pSAM_RI for RB-TnSeq 62

3.2.2. Generation and mapping of Rlv3841 barcoded transposon mutant library 64

3.2.3. Rlv3841 genes lacking transposon insertion mutants 68

3.2.4. Rlv3841 colonisation of pea 72

3.2.5. BarSeq experimental design to assess gene essentiality during rhizosphere growth and root colonisation 75

3.2.6. BarSeq to identify Rlv3841 genes required for growth in the rhizosphere and colonisation of pea 76

3.2.7. Maintenance of genetic diversity *in planta* 78

3.2.8. Identification of genes required for Rlv3841 growth in the rhizosphere and root colonisation of pea 80

3.2.9. Genes required for Rlv3841 growth in the pea rhizosphere 85

3.2.9.1. Biosynthesis, transport and metabolism 86

3.2.9.2. Adaptation, regulation and cellular factors 96

3.2.10. Rlv3841 genes required for colonisation of pea roots 100

3.2.10.1. Motility and chemotaxis 101

3.2.10.2. Cell surface components 107
3.2.10.3. Adaptation, regulation and metabolism 110
3.2.11. Genes that upon mutation provide an advantage for rhizosphere growth and root colonisation 112
3.3. Conclusion 119
Chapter 4 123
Identifying common and plant-specific colonisation determinants in *R. leguminosarum* among host legumes, a non-host legume and a non-legume 123
4.1. Introduction 124
4.2. Results and discussion 127
4.2.1. Rlv3841 colonisation of lentil, *Lathyrus*, fava bean, alfalfa and barley 127
4.2.2. BarSeq experimental design to identify common and plant-specific determinants of Rlv3841 rhizosphere growth and root colonisation 133
4.2.3. Maintenance of Rlv3841 mutant library diversity *in planta* 135
4.2.4. Identification of genes required for growth in the rhizosphere and colonisation of lentil, *Lathyrus*, alfalfa and barley 138
4.2.5. Common determinants of rhizosphere growth and root colonisation 141
4.2.5.1. Common requirements for growth in the rhizosphere of five plant species 150
4.2.5.2. Common requirements for root colonisation of five plant species 159

4.2.5.3. Analysis of central carbon metabolism in the rhizosphere and during root colonisation 167
4.2.6. Plant-specific factors required for rhizosphere growth and root colonisation 171
4.2.6.1. Plant-specific factors contributing to rhizosphere fitness 176
4.2.6.2. Plant-specific factors contributing to root colonisation 186
4.3. Conclusion 195

Chapter 5 205

Identifying competitive colonisation determinants in *R. leguminosarum* and *P. fluorescens* using BarSeq 205

5.1. Introduction 206
5.2. Results and discussion 209
5.2.1. Generation and mapping of *P. fluorescens* SBW25 barcoded transposon mutant library 209
5.2.2. Colonisation dynamics of Rlv3841 and SBW25 during competition on pea and barley 212
5.2.3. BarSeq experimental design to identify Rlv3841 and SBW25 genes important for competitive colonisation of pea and barley 221
5.2.4. Maintenance of genetic diversity of Rlv3841 and SBW25 transposon libraries *in planta* 222
5.2.5. Identification of Rlv3841 genes with significant fitness contributions during competitive colonisation of barley with SBW25 226

5.2.6. Identification of SBW25 genes contributing to growth in the rhizosphere and colonisation of pea and barley during single and co-inoculation with Rlv3841 235

5.2.7. Common determinants of SBW25 rhizosphere growth and root colonisation 238

5.2.8. Genes promoting SBW25 rhizosphere growth and root colonisation during competition with Rlv3841 246

5.2.8.1. Common determinants of pea and barley rhizosphere growth and root colonisation during competition with Rlv3841 247

5.2.8.2. Genes contributing to growth in the rhizosphere of pea or barley during competition with Rlv3841 248

5.2.8.3. SBW25 genes contributing to the colonisation of pea or barley during competition with Rlv3841 256

5.3. Conclusion 264

Chapter 6 278

General Discussion 278

6.1. Overview 279

6.2. Development of RB-TnSeq in *Rhizobium* 280

6.3. Identification of Rlv3841 genes contributing to growth in the rhizosphere and colonisation of pea, lentil, *Lathyrus*, alfalfa and barley 282

6.3.1. Validation of RB-TnSeq *in planta* 283

6.3.2. Common determinants of rhizosphere growth and root colonisation 290

6.3.3. Plant-specific genes contributing to rhizosphere growth and root colonisation 298

6.4. Identification of Rlv3841 genes contributing to competitive growth in the rhizosphere and colonisation of pea and barley 299
6.5. Identification of common determinants of SBW25 rhizosphere growth and colonisation and those specifically required during competition with Rlv3841 303
6.6. Limitations 304
6.7. Concluding remarks and future perspectives 306

Bibliography 308

Appendix A. Supplementary material for Chapter 3 337

Appendix B. Supplementary material for Chapter 4 339

Appendix C. Supplementary material for Chapter 5 340

Chapter 1

Introduction

This introduction is adapted from the review article “Deciphering bacterial mechanisms of root colonization” written by myself and edited in collaboration with Jorrin, B., Haskett, T.L. and Poole, P.S. Environmental Microbiolology Reports. 2021.

1.1. Background

Soil provides a diverse habitat for billions of individual microorganisms, many of which form complex interactions with plants spanning the continuum of ecological outcomes from beneficial to pathogenic (Bardgett and Van Der Putten, 2014). To attract beneficial microbes from nutrient-poor bulk-soil, plants exude up to 20% of their photosynthate into the rhizosphere (soil-root interface), providing carbon for microbial growth and proliferation (Estabrook and Yoder, 1998). Some individuals form more intimate associations with plants, colonising the rhizoplane (root-surface) as epiphytes or endosphere (space between root cells) as endophytes (Fig. 1.1A) (Bulgarelli et al., 2013; Reinhold-Hurek et al., 2015; Tkacz et al., 2015). Epiphytic and endophytic lifestyles allow microorganisms to remain anchored in a nutrient rich environment and facilitate the development of beneficial plant-bacteria interactions, thus providing a key advantage over a free-living lifestyle.

Root-associated microbiota positively influence plant health and productivity through various mechanisms including enhancing nutrient acquisition, priming of plant defences and control of plant pathogens (Berendsen et al., 2012; Philippot et al., 2013; Wei et al., 2015; Trivedi et al., 2020). In recent years, metagenome studies have identified a wealth of microorganisms inhabiting the various root-niches and revealed bacteria to be the most prevalent form of root-associated microbiota. Despite the vast bacterial diversity present in soil, bacteria from four phyla; Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria, account for the major fraction of the root-microbiome (Uroz et al., 2010; Bulgarelli et al., 2012; Lundberg et al., 2012). However, taxonomic composition varies widely at the genus and species levels due to

unique selective pressures imposed by host genotype, sub-localisation and abiotic environmental factors (Turner et al., 2013a; Turner et al., 2013b; Schlaeppi et al., 2014; Lebeis et al., 2015; Tkacz et al., 2020).

Exploitation of the root microbiome is often touted for its enormous potential to substitute environmentally deleterious agrochemicals such as fertilisers and pesticides that are crucial for current agricultural productivity (Busby et al., 2017). However, attempts to modify root microbiomes through the introduction of beneficial microbes is often hindered by the fact that introduced species are outcompeted either directly or indirectly by members of the native microbiome (Albright et al., 2021). As such, there is escalating interest regarding the mechanistic characterisation of factors underlying the processes of rhizosphere colonisation, subsequent persistence and root colonisation as well as how plant-bacteria and bacteria-bacteria interactions influence them.

1.2. Bacterial root colonisation

Bacterial colonisation of plants roots is a sequential, multi-step process that begins in the rhizosphere with chemotaxis towards the root, followed by attachment and subsequent biofilm formation (Fig. 1.1). Its importance in the establishment of both beneficial and detrimental plant-bacteria interactions has led to the identification of multiple genes across diverse bacterial species required for successful root colonisation (Table 1.1).

1.2.1. Chemotaxis towards the root

Plant photosynthates secreted into the rhizosphere form gradients that are perceived by bacteria occupying the surrounding soil, resulting in activation of chemosensory pathways and movement of motile bacteria towards the root (Fig. 1.1A). Bacterial motility is primarily mediated by proteinaceous appendages termed flagella that protrude from the cell surface, or through type IV pili (Alexandre, 2015). Genes homologous to known components of the chemosensory and flagella pathways have been identified in most sequenced bacterial genomes suggesting that chemotaxis provides a selective advantage, particularly in nutrient scarce environments such as soil (Armitage, 1999; Szurmant and Ordal, 2004; Wadhams and Armitage, 2004).

Not surprisingly, inactivation of bacterial chemotaxis or motility renders bacteria deficient for colonisation of the rhizoplane (Ames and Bergman, 1981; Bauer and Caetano-Anollés, 1990; de Weert et al., 2002; Allard-Massicotte et al., 2016). This is illustrated in experiments where all the *Bacillus subtilis* chemoreceptors were genetically inactivated leading to a significant reduction in colonisation of *Arabidopsis thaliana* roots four hours post-inoculation (Allard-Massicotte et al., 2016). Chemotaxis and motility also play fundamental roles in the formation of *Rhizobium*-legume symbioses. Rhizobia are motile alpha-proteobacteria that infect legume root-nodules and fix atmospheric di-nitrogen (N_2) into ammonia for plant utilisation in return for carbon (Poole et al., 2018). Successful establishment of nodule symbiosis requires a highly specific molecular dialogue between the two partners. Beginning in the rhizosphere, exudation of chemoattractants by legumes draw rhizobia towards the root hairs which act as the entry point for nodule infection (Armitage et al., 1988; Barbour et al., 1991; Dharmatilake and Bauer, 1992). Mutation of chemosensory components

impairs competitiveness for root colonisation and nodule occupancy in *Rhizobium leguminosarum* (Miller et al., 2007) and *Azorhizobium caulinodans* (Liu et al., 2020). Although these mutant bacteria remain motile, they are unable to sense chemical gradients towards the root or possibly root hairs where most nodule infections are initiated.

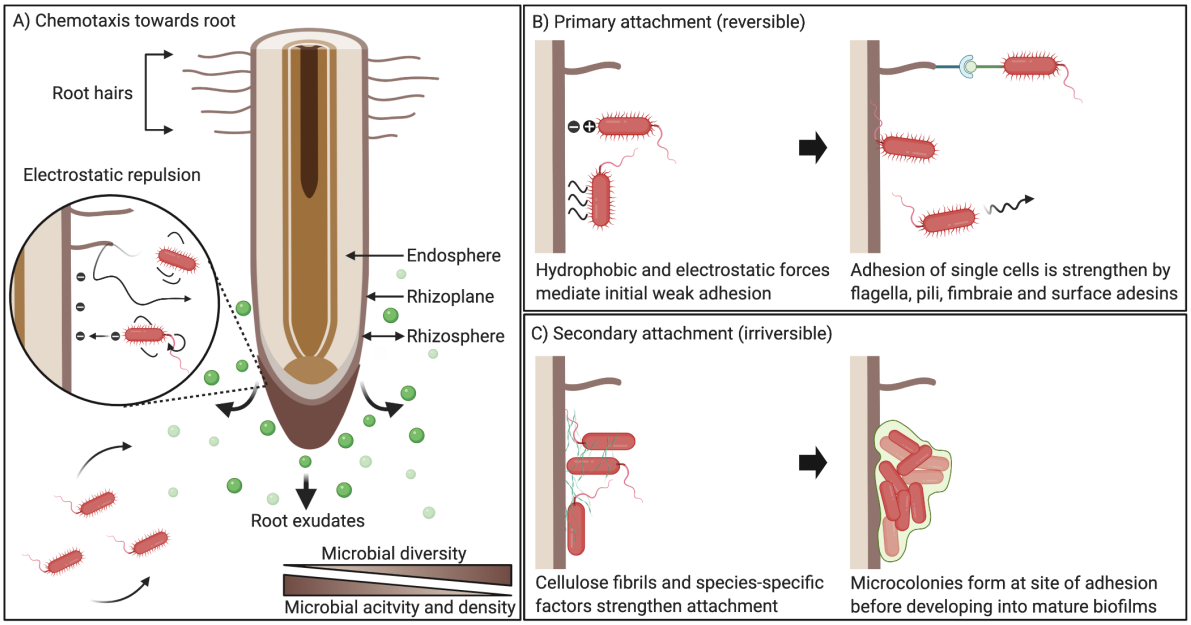


Figure 1.1. Bacterial root colonisation. Bacterial colonisation of plant roots is a multistep process. **A.** Plants secrete photosynthetically fixed carbon into the rhizosphere forming chemical gradients, which chemotactically attract motile bacteria from the soil towards the root surface. Flagella and pili propel bacteria, allowing them to overcome any electrostatic repulsion at the root surface. **B.** Primary attachment results in weak reversible binding of single cells to the root surface. This is initially mediated by hydrophobic and electrostatic interactions and subsequently strengthened by proteinaceous appendages and species-specific surface adhesins. **C.** Secondary attachment leads to strong irreversible binding of bacteria to the root surface, promoting microcolony formation at the initial site of attachment. This process

is mediated by the production of cellulose fibrils and other species-specific factors including polysaccharides and extracellular proteins. Figure from reference (Knights et al., 2021).

1.2.2. Root attachment

Attachment of bacteria to the rhizoplane marks the first physical step in many plant-bacteria interactions, anchoring bacteria in the nutrient-rich environment of the rhizosphere and securing a prime location for the subsequent development of more intimate associations. The molecular mechanisms underlying root attachment have been best defined in the agriculturally important bacterial genera: *Rhizobium*, *Pseudomonas*, *Azospirillum*, *Agrobacterium* and *Salmonella* (Wheatley and Poole, 2018). These proteobacteria share a common biphasic mechanism consisting of two phases: primary attachment, characterised by reversible binding of bacteria to the root surface, followed by secondary attachment which results in their irreversible adhesion.

1.2.2.1. Primary attachment

Primary attachment involves weak, non-specific and reversible binding mediated by hydrophobic and electrostatic interactions between cells and adjacent surface molecules on the root (van Loosdrecht et al., 1987; Kendall and Roberts, 2015). Despite the benefits of root attachment only a small proportion of inoculated isogenic bacteria, typically representing 0.4-3.5% of the population, actually attach to roots in controlled conditions (Rodríguez-Navarro et al., 2007). This is primarily due to electrostatic repulsion, which occurs between the negatively charged bacterial cell envelope and root surface (Berne et al., 2015). To overcome these repulsive forces,

bacteria use flagella and pili to propel themselves towards the root surface (Fig. 1.1A). Following these initial interactions adhesins present on the cell surface mediate a tighter, but still reversible association with the root (Wheatley and Poole, 2018) (Fig. 1.1B). Bacterial adhesins involved in primary attachment include proteinaceous appendages (flagella, pili, fimbriae), surface proteins and polysaccharides (exo- and capsular polysaccharides).

Numerous studies have demonstrated the role of flagella and pili as adhesins, enabling bacteria to not only move to the root, but also to attach and migrate across the root surface. Flagella-defective mutants of *A. brasilense* fail to attach to wheat or maize roots, whereas purified polar flagella bind to wheat roots (Croes et al., 1993; Rossi et al., 2016). Likewise, *P. fluorescens* and *S. enterica* flagella-deficient mutants are unable to competitively colonise alfalfa roots or invade *Arabidopsis* lateral root junctions, respectively (Cooley et al., 2003; Capdevila et al., 2004). In the plant pathogen *P. aeruginosa* (Hahn, 1997; O'Toole and Kolter, 1998) and N_2-fixing endophyte *Azoarcus* sp. BH72 type IV pili act as adhesins (Dörr et al., 1998). However, it is often difficult to elucidate whether adhesion or motility facilitated by the flagella and pili is necessary for primary attachment. Fimbriae also take part in primary attachment, but unlike flagella and pili they do not have an active role in motility. Fimbriae appear to be common primary attachment factors among rhizobacteria (Vesper and Bauer, 1986; Vesper, 1987; Tan et al., 2016), which contain a high proportion of hydrophobic amino acid residues thereby contributing to cell surface hydrophobicity and influencing attachment (Rosenberg and Kjelleberg, 1986; Donlan, 2002).

Considering that bacterial mutants lacking flagella, pili and fimbriae are still able to attach to the root surface (Tan et al., 2016), it is likely that other species-specific factors with adhesive properties such as polysaccharides and surface proteins play a key role in primary attachment. For example, in *A. brasilense* and *P. fluorescens* various major outer membrane proteins (MOMPs) have been implicated in root adhesion and cellular aggregation (De Mot and Vanderleyden, 1991; Burdman et al., 2001; Alvarez Crespo and Valverde, 2009). These MOMPs are exposed on the outer side of the bacterial cell and function by interacting with the surface domains of proteins and polysaccharides located on the root exterior.

1.2.2.2. Secondary attachment

The second phase of attachment involves strong irreversible binding of the bacteria to the root surface, mediated by the synthesis of extracellular cellulose fibrils and species-specific secondary attachment factors. Biosynthesis, secretion or exposure of these cellulose fibrils and secondary attachment factors is typically induced after successful primary attachment (Matthysse, 1983; Ausmees et al., 1999; Martínez-Gil et al., 2010; Monteiro et al., 2012). In *Agrobacterium* for example physical contact with the root surface is thought to result in the secretion of unipolar polysaccharides (UPP) (Tomlinson and Fuqua, 2009). Secondary attachment culminates in the formation of a bacterial microcolony on the root (Fig. 1.1C) and ensures that bacteria remain on the rhizoplane. For many bacteria, this is essential for subsequent endophytic colonisation (Kandel et al., 2017).

Cellulose fibrils, which often extrude from multiple points over the bacterial cell surface, appear to be universal secondary attachment factors among proteobacteria

(Thompson et al., 2018). These fibrils bind tightly to one another, thereby promoting the formation of bacterial aggregates on the rhizoplane. In *Rhizobium,* cellulose fibrils assist bacterial accumulation at the site of infection by tightly adhering rhizobial cells on root-hair tips (Smit et al., 1987; Williams et al., 2008). Similarly, in *Agrobacterium* cellulose fibrils anchor bacteria at the site of primary attachment promoting tumour formation (Matthysse, 1983). Although attachment is a critical early step in *Rhizobium* infection and *Agrobacterium* pathogenesis, the role of cellulose fibrils in attachment is not essential for the establishment of these plant-bacteria interactions. *Rhizobium* and *Agrobacterium* mutants deficient in cellulose fibrils are still able to induce nodulation and tumour formation respectively (Matthysse, 1983; Smit et al., 1987). Nonetheless, it cannot be excluded that cellulose fibrils are important for attachment and development of these plant-bacteria interactions under field conditions since *Agrobacterium* mutants lacking cellulose fibrils are easily removed from the rhizoplane by washing and require inoculation of higher cell densities to induce tumour formation (Minnemeyer et al., 1991).

In addition to the conserved factors described above, species-specific factors play a key role in secondary attachment. These factors include extracellular proteins and polysaccharides that permit accumulation of bacteria at the site of primary attachment (Rodríguez-Navarro et al., 2007). In *P. fluorescens and P. putida* the large adhesin protein LapA defines the transition from reversible polar attachment of single cells to their irreversible adhesion (Hinsa et al., 2003). LapA is a Ca^{2+} binding protein secreted from the bacterial cell through ATP binding cassette (ABC) transporters, that loosely associates with the bacterial cell surface ready to mediate surface interactions. *P. fluorescens lapA* mutants attach to abiotic surfaces at levels comparable to the wild

type 1-hour post inoculation (hpi) but after 5 hpi show a significant reduction in attachment and are defective for biofilm formation, suggesting that LapA is not involved in the initial primary attachment of *P. fluorescens* to roots. In contrast, *P. putida lapA* mutants show reduced attachment to abiotic surfaces and corn seeds and are defective for biofilm formation at 1 hpi implying that LapA plays a role in the initial adhesion of *P. putida* (Espinosa-Urgel et al., 2000; Yousef-Coronado et al., 2008; Duque et al., 2013). Moreover, *P. putida lapA* mutants are at a competitive disadvantage for colonisation of corn roots when in competition with the wild-type strain. In *R. leguminosarum* and *R. etli*, *Rhizobium*-adhering proteins (RAPs) are important species-specific primary attachment factors also thought to bind Ca^{2+} (Ramey et al., 2004). RapA1 is a secreted Ca^{2+}-binding protein that localises on the extracellular surface at the cell poles and is predicted to promote aggregation through binding of exopolysaccharides (EPS) or capsular polysaccharide (CPS) (Russo et al., 2006). Overexpression of *rapA1* in *R. leguminosarum* enhances the number of bacteria attached to host legume roots by up to five-fold (Mongiardini et al., 2008).

Some species-specific adhesins such as EPS and lipopolysaccharides (LPS) play a role in both the primary and secondary attachment of diverse bacterial species. EPS is a major cell surface component composed of carbohydrate polymers, which promote cellular aggregation and irreversible binding to the root surface by forming bridges between bacterial cells (Burdman et al., 2000). The structure of these polymers varies considerably between bacterial species, altering the electrostatic, hydrophobic and steric properties of the cell surface and in turn affecting attachment. In *R. leguminosarum,* mutation of the EPS biosynthesis regulator *pssA*, results in reduced attachment to root hairs and impaired aggregation at root hair tips (Williams

et al., 2008). However, attachment to the root epidermis was still observed. LPSs constitute a major component of the gram-negative outer membrane and are composed of large tripartite glycolipids with a hydrophobic portion called lipid A, a hydrophilic core oligosaccharide, and the hydrophilic O-antigen side chain (Bertani and Ruiz, 2018). LPS plays a critical role in the establishment of effective associations between several plant growth promoting bacteria and their hosts. *Rhizobium* mutants deficient in the production of the monosaccharide rhamnose, an integral component of the LPS O-antigen, displayed reduced colonisation of rice and *Sesbania rostrata* roots, nullifying plant growth promotion and impairing nodulation, respectively (Mitra et al., 2016). Disruption of rhamnose biosynthesis also alters LPS composition in *A. brasilense*, *Herbaspirillum seropedicae* and *A. caulinodans* resulting in reduced colonisation of the host plants (Jofré et al., 2004; Balsanelli et al., 2010) and for *A. caulinodans,* ineffective symbiosis with *S. rostrata* (Gao et al., 2001). Interestingly, several of these LPS mutants were not impaired in attachment to plastic and glass surfaces, indicating that LPS or a component of it, is not required for general attachment (Balsanelli et al., 2010; Mitra et al., 2016).

1.2.2.3. Environmental factors influence root attachment

Attachment can be influenced by environmental factors such as soil pH, divalent cations (Ca^{2+} and Mg^{2+}) and water availability (Caetano-Anollés et al., 1989; Howieson et al., 1993). The effect of pH on attachment has been well characterised in *R. leguminosarum*. Under acidic conditions the UPP glucomannan mediates localised polar attachment of *R. leguminosarum* to pea and vetch root hairs through binding to plant lectins (Laus et al., 2006). The gene locus encoding glucomannan is conserved among *Rhizobium* and *Agrobacterium* where a UPP similar to glucomannan has been

shown to mediate irreversible polar attachment to plant tissue (Tomlinson and Fuqua, 2009; Xu et al., 2012). Under alkaline conditions root lectins are solubilised preventing glucomannan from mediating attachment. It has been proposed than an extracellular Ca^{2+} binding protein termed “rhicadhesin” may mediate attachment under alkaline conditions. Rhicadhesin is predicted to bind the bacterial cell wall via a Ca^{2+} that is thought to dissociate under acidic pH conditions, however, there is currently little evidence pertaining to the identity of this hypothetical protein (Matthysse, 2014; Thompson et al., 2018). In fact, evidence for the existence of rhicadhesin is based entirely on a single set of experiments where crude preparations of rhizobial membrane-proteins inhibited attachment of various rhizobia to pea roots (Smit et al., 1989; Smit et al., 1991). Thus, it remains unclear as to whether a pure rhicadhesin protein facilitates attachment, or whether the concerted action of several proteins is required for this process.

1.2.3. Bacterial biofilms

Following attachment, microcolonies develop into mature biofilms on the root surface. Biofilm formation is a key determinant of successful root colonisation and is a common strategy employed by many soil bacteria. Biofilms provide a physical barrier against detrimental external stimulus such as the diffusion of antimicrobial compounds from the host plant or other microbiome members. They also protect bacteria from environmental stresses including changes in pH, osmotic stress and UV radiation (Davey and O'Toole, 2000). Fundamentally, biofilms consist of dynamic heterogenous communities of bacterial cells embedded in a matrix of EPS which aids in adherence to the root surface and ensures cells remain proximal to one another (Branda et al., 2005; Flemming and Wingender, 2010). Within the biofilm, individual microcolonies

are separated by water channels that facilitate diffusion of nutrients, oxygen, antimicrobial compounds and even DNA via horizontal gene transfer (Donlan, 2002; Flemming and Wingender, 2010); hence biofilms also play a significant role in the functioning of bacteria-bacteria interactions. Large adhesins play a fundamental role in biofilm formation by mediating cell-cell interactions in both gram-negative and gram-positive bacteria. In *P. putida, lapF* mutants are unable to form microcolonies at the initial site of attachment and display reduced colonisation of corn and alfalfa roots when inoculated individually and co-inoculated with the wild type (Martínez-Gil et al., 2010). This suggests that LapF plays a role in both root colonisation and the development of mature biofilms in *P. putida*. LapA has also been implicated in mediating cell-cell as well as cell-surface interactions during biofilm development in *P. putida* (Gjermansen et al., 2010). No orthologues of *lapA* or *lapF* have been identified in *P. syringae*, *P. mendocina*, *P. stutzeri* or *P. aeruginosa* strains suggesting that although both pathogenic and non-pathogenic strains of *Pseudomonas* attach to, colonise and form biofilms on surfaces, the mechanisms by which they do this differ (Duque et al., 2013). Other large adhesins such as Bap and Esp have been shown to mediate surface colonisation and biofilm formation in the gram-positive bacteria *Staphylococcus* and *Enterococcus* respectively (Cucarella et al., 2001; Toledo-Arana et al., 2001). These proteins along with other adhesins have a similar structural organisation to LapA and LapF and are widespread throughout prokaryotes suggesting that a similar mechanism for biofilm formation exists (Lasa and Penadés, 2006; Yousef and Espinosa-Urgel, 2007).

Table 1.1. Genes shown to affect colonisation of plant roots.

Colonisation stage	Category	Gene(s)	Strain(s)	Comments	Reference(s)
Chemotaxis towards the root	Chemotaxis	*cheA*	*P. fluorescens*	*cheA* mutants demonstrate reduced competitive tomato-root-tip colonisation when inoculated 1:1 with wild-type.	(de Weert et al., 2002)
		che1 gene clusters	*R. leguminosarum*	Che1, which encodes the conserved set of *cheAWYRB* homologues, facilitates root colonisation and promotes competitive nodulation of peas.	(Miller et al., 2007)
		mcpA, *mcpB*, *tlpA*, *tlpB*, *mcpC*, *tlpC*, *yvaQ*, *yoaH, yfmS*, *heamAT*	*B. subtilis*	Deletion of all 10 chemoreceptors results in reduced colonisation of *Arabidopsis* roots relative to wild type.	(Allard-Massicotte et al., 2016)
		cheY1, *cheY2*	*A. caulinodans*	Regulate chemotaxis, competitive root colonisation and competitive nodulation of *Sesbania rostrata*.	(Liu et al., 2020)
Movement towards/ over the	Flagella	*fliC*, *fleQ*, *fleS*	*P. fluorescens*	Single mutants are non-motile. Colonise alfalfa roots when inoculated alone, but impaired in colonisation when co-inoculated with wild type. These genes have been found in every pseudomonad analysed.	(Capdevila et al., 2004)

root and potential role in primary attachment as adhesins		*flhD*, *flhC*	*S. enterica*	Double mutant is nonmotile as defective in flagella synthesis and fails to effectively colonise *Arabidopsis* roots or invade lateral root junctions.	(Cooley et al., 2003)
		flmA, *flmB*	*A. brasilense*	Mutation in either gene results in non-motile cells due to altered flagellum assembly. Mutations also affect cells competitive ability to attach to maize roots.	(Rossi et al., 2016)
	Type IV Pili	*pilA*, *pilB*	*Azoarcus* sp. BH72	Single and double mutants failed to colonise rice roots or infect the root epidermal cells.	(Dörr et al., 1998)
Primary attachment	Major outer membrane proteins (MOMPs)	*oprF*	Specific to *Pseudomonas* genus	Purified OprF strongly and selectively binds wheat, barley, maize and sunflower roots, but not leaves. *P. fluorescens oprF* mutants show reduced attachment to cucumber and tomato roots 1-h post inoculation.	(De Mot and Vanderleyden, 1991; Alvarez Crespo and Valverde, 2009)
		omaA	*A. brasilense*	Purified OmaA has a stronger binding affinity for cereal roots relative to legume and tomato roots.	(Burdman et al., 2001)

	Unipolar polysaccharide (UPP)	*gmsA*	*R. leguminosarum*	Determines glucomannan synthesis which under acidic conditions mediates reversible polar attachment of single cells to pea and vetch roots by binding plant lectins.	(Laus et al., 2006)
Secondary attachment and/or microcolony and biofilm formation		*uppABCDEF*	*A. tumefaciens*	Encodes UPP similar to glucomannan but mediates irreversible polar attachment to plant tissue and abiotic surfaces.	(Tomlinson and Fuqua, 2009; Xu et al., 2012)
	Cellulose fibrils	*celA*	*R. leguminosarum*	Encodes cellulose synthase. Mutation does not alter attachment to root hairs but prevents cap formation under acidic and alkaline conditions. Mutants were also able to form biofilms in vitro (on glass) but not on root hairs.	(Williams et al., 2008)
			A. tumefaciens	Mutants attach to carrot tissue culture cells, but unable to form aggregates and easily removed by washing plant tissue.	(Matthysse, 1983)
	Extracellular proteins	*lapA*	Specific to *P. fluorescens* and *P. putida*	Surface adhesin and biofilm matrix component. Drives transition from reversible to irreversible attachment. *P. putida lapA* mutants are impaired in competitive colonisation of corn roots.	(Espinosa-Urgel et al., 2000; Hinsa et al., 2003; Yousef-Coronado et al., 2008; Gjermansen et al., 2010; Duque et al., 2013)

		lapF	Specific to *P. putida*	Mediates cell-cell interactions during biofilm development. Mutants impaired in micrology formation and biofilm development. Also impaired in individual and competitive colonisation for corn and alfalfa roots.	(Martínez-Gil et al., 2010)
		rapA1	Confined to *R. leguminosarum* and *R. etli*	Overexpression of *rapA1* in *R. leguminosarum bv. trifolii* increases attachment to red clover roots 5-fold.	(Mongiardini et al., 2008)
Important for attachment to plant roots, but not to abiotic surfaces	LPS	*rffB*	*Rhizobium* sp. IRBG74 >99% identity to *A. tumefaciens rffB*	Mutants impaired in dTDP-rhamnose synthesis resulting in altered LPS. Show reduced ability to colonise legume and rice roots, which negatively affects nodulation and endophytic colonisation respectively. Mutants not affected in attachment to polypropylene plates so plant attachment specific trait.	(Mitra et al., 2016)
		rmlD	*A. brasilense*	Disruption of dTDP-rhamnose biosynthesis modifies LPS core and increased EPS production resulting in impaired attachment to maize roots and reduced colonisation.	(Jofré et al., 2004)

		rfbB, *rfbC*	*H. seropedicae*	Mutants lack rhamnose-containing LPS and show 100-fold reduction in attachment to maize roots relative to wild type. Also impaired in endophytic colonisation. No difference in attachment to glass fibre suggesting recognition of rhamnose-containing LPS is important for colonisation of hosts.	(Balsanelli et al., 2010)
Root-hair attachment	EPS	*pssA*	*R. leguminosarum*	*pssA* regulates EPS biosynthesis, with mutants showing reduced attachment to root hairs and impaired cap formation. Attachment to the epidermis was still observed.	(Williams et al., 2008; Janczarek and Rachwał, 2013)

1.3. Plant-bacteria interactions influence colonisation

Plant roots grow amongst diverse bacterial communities with up to 10^4 bacterial species and 10^9 bacterial cells per gram of soil (Daniel, 2005). Not surprisingly, the reservoir of bacteria in the surrounding soil is a crucial factor influencing root-associated microbiome structure, as is illustrated by microbiome analyses of 27 inbred field-grown maize lines planted at five geographically distinct locations across the United States (Peiffer et al., 2013). Albeit, plants do have some control over the composition of their root-associated microbiome through modulation of their root exudate composition and mounting of an innate immune response (Lebeis et al., 2015; Reinhold-Hurek et al., 2015; Sasse et al., 2018). As such, there are clear taxonomic differences between microbiota associated with the root and surrounding soil (Peiffer et al., 2013; Bulgarelli et al., 2015; Edwards et al., 2015; de Souza et al., 2016; Cheng et al., 2020). In *Medicago truncatula*, the rhizosphere and rhizoplane community structures become distinct after only one week of growth in soil. The community structure of both environments becomes more robustly established by the second week, and remains stable for a minimum of three weeks thereafter (Tkacz et al., 2020). Crucially, microbiome structure varies across different plants species and even among genotypes within a single species (Kuske et al., 2002; Aira et al., 2010).

Plant root exudates contain a wide range of primary metabolites including carbohydrates, organic acids and amino acids, which preferentially stimulate bacterial growth thereby shaping microbiome assembly (Sasse et al., 2018). In addition, secondary metabolites present in root exudates can have antimicrobial

activity against pathogens (Olanrewaju et al., 2019), act as signals for the establishment of root symbioses (Abdel-Lateif et al., 2012), and have profound effects on the bacterial transcriptome (Ramachandran et al., 2011; Carvalhais et al., 2013). Root exudate composition varies between plant species and is dynamically influenced by developmental stage, environmental conditions and the structure of the root-microbiome (Sasse et al., 2018). Root colonisation by specific microbial communities affects the chemical composition of root exudates through a systemic root-to-root signalling mechanism termed systemically induced root exudation of metabolites (SIREM) (Korenblum et al., 2020). In tomato split root assays, inoculation of *B. subtilis* onto one side of the root induced a systemic signal that results in increased secretion of acyl-sugars from the uninoculated side of the root. Transportation of these signals through shoots to uncolonised areas of the root can modulate colonisation and assembly of SIREM-specific microbial communities.

The evidence supporting plant immune signalling as a regulator of root-associated microbiome structure is less compelling, but nevertheless important (Yu et al., 2019). Mutants of *A. thaliana* impaired in salicylic acid (SA) mediated defence displayed distinct microbiome composition relative to wild-type plants (Lebeis et al., 2015). However, SA-dependent signalling was shown to have the largest impact on endophytic community structure with rhizosphere microbiome structure less affected. In wheat, activation of the jasmonic acid (JA) signalling pathway through application of exogenous JA was found to reduce the diversity of the endophytic microbiome but not rhizosphere microbiome (Liu et al., 2017a). In contrast two *Arabidopsis* mutants disrupted in JA-dependent signalling showed distinct rhizosphere microbiome structures relative to the wild type, though this could be

associated with the fact that their root exudation profiles were also affected (Carvalhais et al., 2015). Additionally, aboveground activation of the immune system by plant pathogens and insects has been shown to alter the rhizosphere microbiome structure of several plant species (Dudenhöffer et al., 2016; Kong et al., 2016; Berendsen et al., 2018; Yuan et al., 2018). Again, alteration of rhizosphere microbiome structure may be linked to alteration of root exudate composition in response to activation of the immune system (Yuan et al., 2018). These results indicate that the immune system can influence epiphytic colonisation, but colonisation of the rhizosphere is indirectly influenced through the modulation of root exudates in response to the mounting of an innate immune defence.

1.4. Bacteria-bacteria interactions influence colonisation

Interactions among bacterial communities can be complex, involving both cooperation and competition (Deines and Bosch, 2016). Cooperative interactions include processes such as metabolite exchanges (Zelezniak et al., 2015), whilst production of antimicrobial toxins and deployment of mechanical weapons are examples of competitive interactions (Granato et al., 2019). Notably, most experiments characterising factors required for bacterial colonisation have done so with single strains under sterile laboratory conditions. A major challenge in understanding bacterial assembly in root-niches is to move from single species studies to those that encompass entire communities to comprehend how individuals interact with each other and their host plant. Due to the enormous complexity of root-associated microbiomes it is extremely challenging to experimentally characterise the molecular mechanisms underlying plant-bacteria and bacteria-

bacteria interactions, and their effects on plant health in natural systems. To facilitate this, synthetic communities (SynComs) have been used to investigate the role of individual species and plant-derived metabolites during colonisation of root-niches (Niu et al., 2017; Voges et al., 2019). Importantly, SynCom assembly is affected by bacterial interactions (Mee et al., 2014), host genotype (Bodenhausen et al., 2014; Lebeis et al., 2015) and niche specificities (Bai et al., 2015). The same factors known to affect assembly of natural microbial communities, validating the use of SynComs as simplified microbiomes.

The power of using SynComs to study plant-bacteria and bacteria-bacteria interactions was recently demonstrated with a simplified community of seven bacterial strains isolated from maize roots (Niu et al., 2017). When inoculated together, all members colonised to form a stable community on maize roots that reduced the prevalence of seedling blight by delaying colonisation of the fungal pathogen *Fusarium verticillioides*. Removal of one specific SynCom member led to failure of the remaining members to form a stable community, resulting in competitive dominance by a single strain which in isolation, was less capable of warding off the pathogen. This study highlights the significance of bacteria-bacteria interactions during colonisation of root-niches and demonstrates that fundamental ecological principles, such as the role of key stone species, can be preserved within simplified SynComs.

1.5. Tool to investigate bacterial colonisation of root-niches

Traditional single-isolate experimental approaches cannot unravel the population level dynamics of bacterial root colonisation or the molecular mechanisms driving them. However, new tools have been developed for marking and monitoring bacterial strains. These include technologies such as engineered transposons for genomic integration of reporter genes (Schlechter et al., 2018) and microfluidic imaging platforms for dynamic mapping of bacterial-root interactions (Massalha et al., 2017; Aufrecht et al., 2018; Noirot-Gros et al., 2020). Transcriptomics approaches such as in vivo expression technology (IVET), microarrays and RNA sequencing have provided insights into to how bacteria modify their gene expression in response to root exudates present in different plant rhizospheres (Rainey, 1999; Barr et al., 2008; Ramachandran et al., 2011; Carvalhais et al., 2013). Moreover, the development of strategies for genome-wide transposon mutagenesis screens are providing unprecedented opportunities to identify population level genetic determinants of colonisation (Cain et al., 2020).

Transposon insertion sequencing (TIS) is one such approach that allows the identification of genetic determinants involved in bacterial colonisation of roots. TIS is a powerful technique whereby libraries of single-insertion transposon mutants, which collectively saturate an organism's genome, are exposed to a specific condition and then analysed with next-generation sequencing to simultaneously estimate the essentiality and/or fitness contribution of each gene in a bacterial genome. Comparison of gene mutation frequency in an 'input pool' relative to an 'output pool' following a challenge such as root colonisation, reveals whether a gene

is essential, non-essential, advantageous, or disadvantageous for growth and colonisation (Fig 1.2). There are several variations of this technique including insertion sequencing (INSeq) and transposon sequencing (TnSeq), each following these same basic principles, allowing determination of gene fitness at the genome scale across a variety of conditions (Cain et al., 2020).

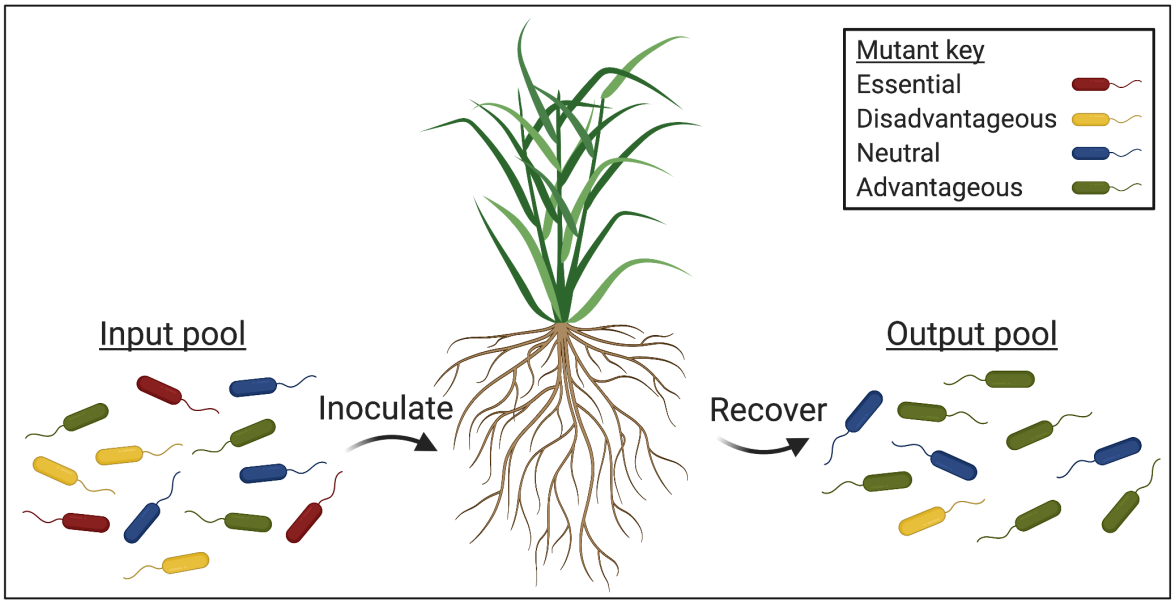


Figure 1.2. Transposon mutagenesis screening to identify genes involved in root colonisation. Libraries containing single-insertion transposon mutants that collectively cover all genes in the bacterial genome are inoculated onto a root system and recovered ‘X' days post-inoculation. Comparison of input and output pools reveals whether a gene is essential (red), non-essential (blue), advantageous (green) or disadvantages (yellow) for root colonisation. Figure adapted from reference (Knights et al., 2021).

INSeq was recently used to identify bacterial genes important in the *Rhizobium*-legume symbiosis at multiple stages of its development (Wheatley et al., 2020). To form a successful symbiosis, rhizobia must undergo several lifestyle changes. To

investigate these, *R. leguminosarum* transposon libraries were inoculated onto its host legume *Pisum sativum* and recovered from four stages of the root-nodule symbiosis: 1) free-living growth in the rhizosphere, 2) colonisation of the root, 3) nodule infection before differentiation into N_2-fixing bacteroids, and 4) terminal differentiation into N_2-fixing bacteroids (Fig. 1.3). While only 27 genes are assigned roles in the organisation and regulation of N_2 fixation, 593 gene were found to be required for the competitive ability to form a successful N_2 fixing symbiosis. Of these, 146 were important for growth in the rhizosphere through to N_2-fixing bacteriods, highlighting that competition in the rhizosphere is critical for establishing plant-bacteria interactions even in near isogenic populations.

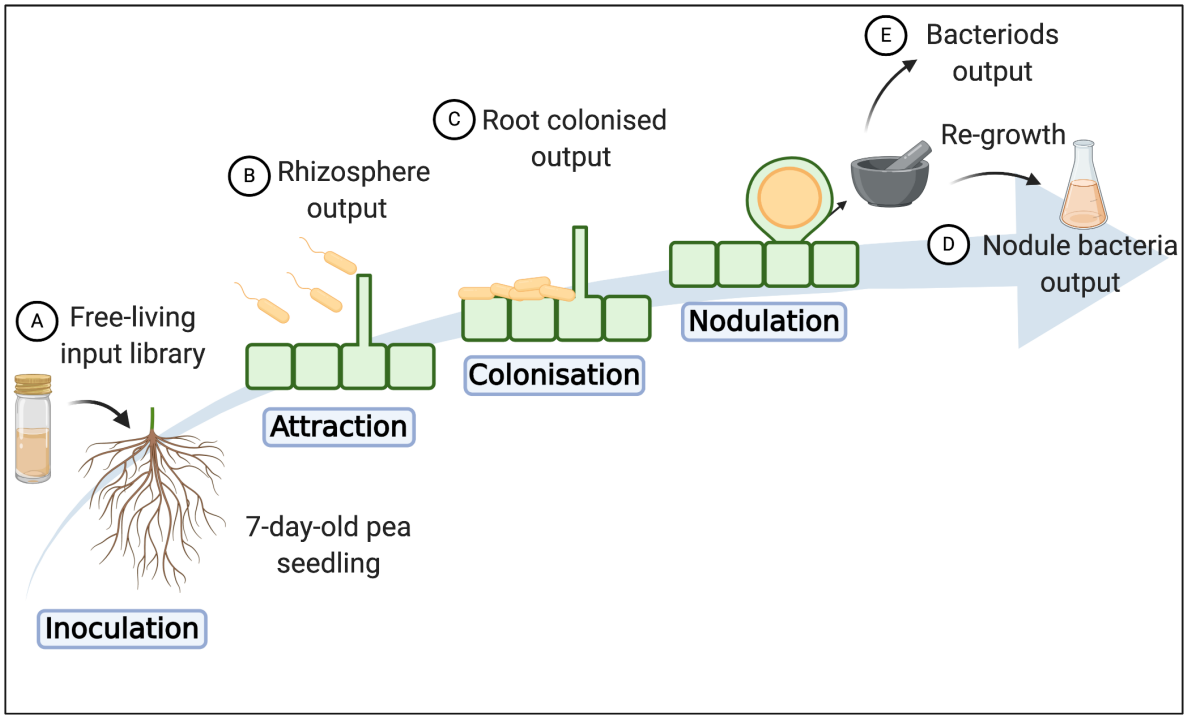


Figure 1.3. Lifestyle adaptations of *Rhizobium* from rhizosphere to symbiosis. Insertion sequencing was used to establish the role of *Rhizobium leguminosarum* genes at multiple stages of symbiosis with *Pisum sativum*. (A) Rlv3841 transposon

library was inoculated onto a 7-day-old pea seedling. Following inoculation bacteria were collected from four stages of symbiosis for analysis: (B) the rhizosphere (5 dpi), (C) the root (5 dpi), (D) nodule bacteria (28 dpi) and (E) N_2-fixing bacteroids (28 dpi). TIS of the input library and four output libraries enabled genome-wide classification of gene fitness contributions at each stage. Figure from reference (Knights et al., 2021).

TIS has also been used to identify genes that contribute to bacterial growth and colonisation in plant niches including *P. simiae* colonisation of *A. thaliana* roots (Cole et al., 2017), *P. syringae* colonisation of the *Phaseolus vulgaris* leaf surface and apoplast (Helmann et al., 2019), *Pseudomonas* sp. WCS365 growth in the rhizosphere of *Arabidopsis* (Liu et al., 2018), *A. tumefaciens* growth during tumour infection of *Solanum lycopersicum* (Gonzalez-Mula et al., 2019) and *Pantoea stewartia* growth in the xylem of corn (Duong et al., 2018).

TIS approaches rely on assessing the effects that loss-of-function mutations have on bacterial fitness. An alternative approach is to assess how a gain-of-function mutation, that results in overexpression of a gene, affects bacterial fitness. Unlike TIS, this allows for the assessment of essential genes under various conditions and can identify roles for genes that are redundant within the genome. However, until recently the nature of these screens has made it unfeasible to achieve them at the genome-wide level. The recent development of a technique termed Dual Barcoded Shotgun Expression Library Sequencing (Dub-seq) allows for high-throughput genome-wide assessment of the effects of gene overexpression in bacteria (Mutalik et al., 2019).

1.6. *Rhizobium*-legume symbiosis

1.6.1. The importance of *Rhizobium*-legume symbioses

Rhizobium-legume symbioses are one of the best characterised plant-bacteria interactions due to their potential to alleviate our reliance on synthetic nitrogen fertilisers. Nitrogen is an essential constituent of amino acids and comprises approximately 16% of total plant protein however, plants are unable to assimilate the abundantly available atmospheric N_2 (Frink et al., 1999). Consequently, nitrogen is a primary limiting factor for plant growth and subsequently crop yields (Kant et al., 2011). To circumvent this limitation, the Haber-Bosch process was developed to reduce N_2 to ammonia, which can be supplied to agricultural systems in the form of fertiliser (Van Breemen, 2002). However, only 30% of applied nitrogen fertiliser is assimilated by crops resulting in significant leaching to the surrounding environment (Raun and Johnson, 1999; Rockström et al., 2009). This leaching is not the only source of contamination, as well, the process itself contributes to the emission of potent greenhouse gases including carbon dioxide and nitrous oxide. In addition, the use of nitrogen fertiliser is the single highest input cost for many crops, making it largely inaccessible to farmers in developing countries. Together, these factors highlight the need to find sustainable and economically viable methods to alleviate the effects of nitrogen limitation of plant growth in agricultural systems. Sustainable incorporation of nitrogen into agricultural systems has been achieved for centuries through the use of legumes in crop rotations. Legumes form symbiotic interactions with soil-dwelling rhizobia, which infect roots where they fix N_2 into biologically accessible ammonia using the enzyme nitrogenase in a process termed biological nitrogen fixation (BNF) (Oldroyd et al., 2011).

1.6.2. Establishment of *Rhizobium*-legume symbioses

To establish symbiosis, rhizobia must colonise legume roots and infect root nodules. Although several modes of rhizobial root infection have been described, infection by "root hair curling" is the best characterised (Jones et al., 2007; Murray, 2011) (Fig. 1.4). Initiation of symbiosis begins in the rhizosphere with the release of flavonoids in legume root exudates, which passively diffuse across the bacterial membrane where they are bound by the bacterial derived transcription factor NodD (Recourt et al., 1989) (Fig. 1.4A). This activates a downstream signalling cascade that results in the biosynthesis and secretion of lipo-chitooligosaccharides called Nod factors (NFs) (D'Haeze and Holsters, 2002; Geurts and Bisseling, 2002) (Fig 1.4B). Attachment of rhizobia to root hairs allows rhizobial NFs to be recognised by plant NF receptors (NFRs) present on the plasma membrane of the root hair. This recognition induces both morphological and biochemical changes that allow infection by root hair curling to proceed (Fig. 1.4C). These changes include oscillations in cytosolic Ca^{2+}, root hair curling and initiation of infection thread development which ultimately results in the entrapment of a single rhizobial cell (Wais et al., 2002; Jones et al., 2007; Oldroyd and Downie, 2008; Murray, 2011). As the root-hair curls, it forms a "Shepard's-crook"-like structure which directs the entrapped rhizobia into an invagination pocket in the root hair cell wall where the infection thread continues to develop from (Jones et al., 2007; Murray, 2011). The continued production of NFs stimulates the development of an intracellular infection thread, through which a single rhizobium divides along, towards the developing nodule (Gage, 2002) (Fig. 1.4D). Infection threads extend between root cells branching and ramifying until they reach the growing meristem, where the bacteria are delivered into the nodule cells via an endocytic process (Gage, 2004; Fournier

et al., 2008) (Fig. 1.4E). Bacteria are surround by the plant membrane forming an organelle called the symbiosome. Here, rhizobia can differentiate into bacteriods at which point the genes required for N_2 fixation are expressed (Alunni and Gourion, 2016).

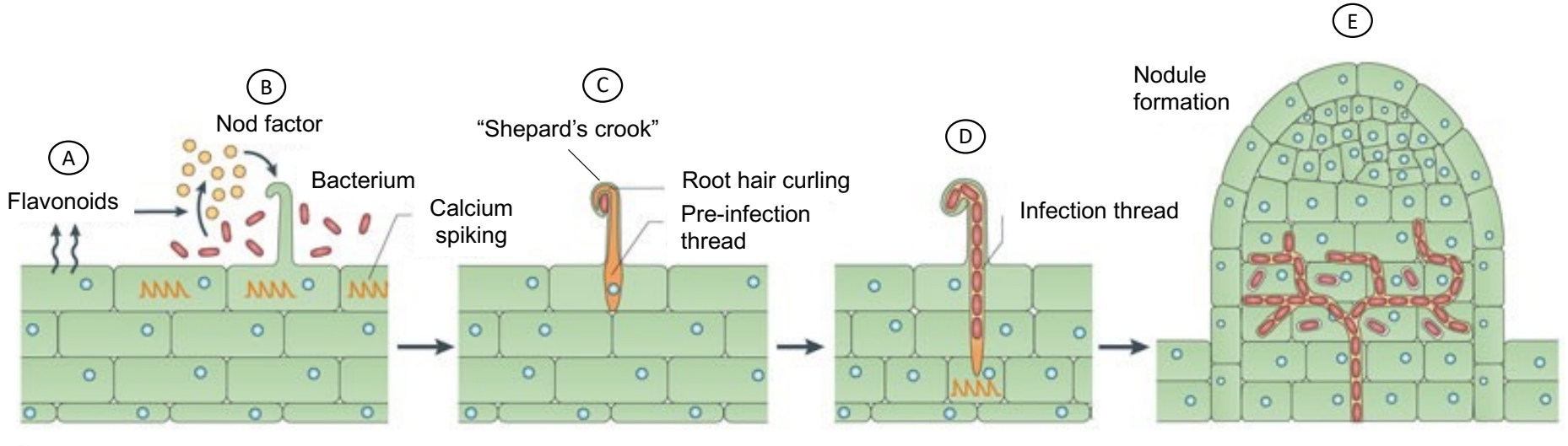


Figure 1.4. Establishment of *Rhizobium*-legume symbioses via the root hair curling mode of infection. A. Flavonoids released by the legume root signal to rhizobia in the rhizosphere, inducing the production of NFs. **B.** Rhizobia attach to plant root hairs in a polar fashion. NF perception by NFRs activates the symbiosis signalling pathway, leading to Ca^{2+} oscillations in the epidermal cells. **C.** This leads to root hair curling, the development of a Shepard’s crook-like structure and infection thread development. **D.** Rhizobia move along the infection thread and are delivered to the nodule. **E.** Here they are enclosed in plant membrane forming the symbiosome and differentiate into N_2-fixing bacteroids. This figure was reproduced from (Oldroyd, 2013).

Most *Rhizobium*-legumes symbioses display stringent specificity meaning that no single rhizobial strain can form symbiosis with all legumes, and vice versa (Wang et al., 2018). Throughout the development of symbiosis there are several checkpoints to ensure infection only proceeds with compatible symbionts. The first checkpoint is provided by flavonoid-NodD interaction. The structure of flavonoids and the structure of NodD vary between legume and rhizobial species respectively, with NF production only induced upon the binding of complementary flavonoid-NodD transcription factor pairs. The second checkpoint is provided by recognition of rhizobial derived NFs by plant NFRs. All NFs share a common backbone encoded by the *nodABC* genes but differ in additional modification along the main structure, whilst NFRs vary between legume species (Long, 1996; Oldroyd and Downie, 2008). During progression through infection threads rhizobial surface polysaccharides provide a further checkpoint for host specificity. Alterations of cell surface polysaccharides results in symbiosis failure, as demonstrated by *S. meliloti* mutants that are defective in EPS production (Jones et al., 2007). Lastly, in the case of indeterminate nodules bacteriods undergo a strictly controlled determination program involving arrest in cell division, genomic endoreduplication, membrane permeabilization and an increase in size. This is mediated in part by the legume encoded nodule-specific cysteine-rich (NCR) peptides providing a final checkpoint in host specificity (Mikuláss et al., 2016).

1.7. Research aims

To date the factors driving rhizosphere fitness and root colonisation in the context of the *Rhizobium*-legume symbiosis remain relatively under characterised, with most studies focusing on the latter stages of nodule infection and N_2 fixation itself. In this

study, we set out to characterise how plant species and bacterial competition affects rhizobial rhizosphere growth and root colonisation using the model rhizobial strain *R. leguminosarum* bv. *viciae* 3841 (Rlv3841). To achieve this, RB-TnSeq (a TIS approach) was used in a genome-wide approach to assess how loss-of-function mutations affect fitness under multiple conditions. Rlv3841 was chosen due to its symbiotic engagement with agriculturally important legumes, the availability of its fully sequenced genome (Young et al., 2006) and access to a wealth of experimental data during both free-living and symbiotic lifestyles (Karunakaran et al., 2006; Williams et al., 2008; Karunakaran et al., 2009; Karunakaran et al., 2010; Ramachandran et al., 2011; Garcia-Fraile et al., 2015; Perry et al., 2016; Wheatley et al., 2017; Parsons, 2019; Wheatley et al., 2020). The growth conditions and genetic tractability of Rlv3841 have also been well defined within the Poole group.

In Chapter 3 the development of RB-TnSeq in Rlv3841 is described and validated *in planta* for genome-wide assignment of gene fitness contributions during growth in the rhizosphere and colonisation of the host legume pea (*Pisum sativum* L.). In Chapter 4 the Rlv3841 RB-TnSeq transposon mutant library was used in barcoded sequencing (BarSeq) screens to assess gene essentiality during growth in the rhizosphere and colonisation of the host legumes pea, lentil (*Lens culinaris* Medik), *Lathyrus* (*Lathyrus latifolius* L.) and fava bean (*Vicia faba* L.), the non-host legume alfalfa (*Medicago sativa* L.), and the non-legume barley (*Hordeum vulgare* L.). This reveals a core set of genes commonly required for growth in the rhizosphere and colonisation of all five plant species, in addition to many genes with plant-specific dependencies. In Chapter 5, the plant growth promoting bacterium *P. fluorescens* SBW25 (SBW25) is co-inoculated with Rlv3841 to facilitate the identification of genes contributing to

colonisation under a competitive setting. Through use of both Rlv3841 and SBW25 RB-TnSeq mutant libraries, gene fitness contributions were simultaneously assigned for both Rlv3841 and SBW25 using BarSeq. These data have provided a greater understanding of the factors contributing to rhizosphere fitness and the colonisation abilities of two beneficial plant growth promoting rhizobacteria, as well as how these genetic dependencies vary in response to the host plant and the presence of a competitor.

Chapter 2

Materials and methods

2.1. Media and growth conditions

Unless otherwise stated all enzymes were purchased from New England Biolabs (NEB) and all chemicals were purchased from Sigma Aldrich.

2.1.1. Media

Escherichia coli strains were cultured in Luria-Bertani (LB) broth at 37°C (Miller, 1972). For growth on solid media, agar was added to 1.4 % w/v prior to autoclaving. *Rhizobium* strains was cultured at 28°C on TY (Beringer, 1974) apart from during the generation of the Rlv3841 RB-TnSeq library for which Rlv3841 was grown on a combination of TY and Universal minimal salts (UMS) combined in equal ratios and supplement with 10mM glucose, 20mM succinate, 10mM ammonium and 10mM glutamate (TY:UMS). UMS medium is derived from AMS (Poole et al., 1994) with the following alterations: $CaCl_2 \cdot 2H_2O$ (0.51 mM), $CoCl_2 \cdot 6H_2O$ (4.2 µM), EDTA-Na_2 (1 µM), and $FeSO_4 \cdot 7H_2O$ (0.04 mM). For growth on solid media, agar was added to 1.75% w/v prior to autoclaving. *Pseudomonas* strains were cultured in either LB or on Kings medium B broth or agar (King et al., 1954).

2.1.2. Antibiotics

Antibiotics were added at the following concentrations where appropriate (µg mL^{-1}): ampicillin (Amp) 100; neomycin (Neo) 50; kanamycin (Kan) 50 for *P. fluorescens* and 20 for *E. coli*; nitrofurantoin (Nitro) 100 for *P. fluorescens* and 20 for *R. leguminosarum* and streptomycin (Str) 500. The antifungal nystatin (Ny) was always used at a concentration of 50 µg mL^{-1}.

2.2. Bacterial strains, plasmids and primers

Bacterial strains used in this work are listed in Table 2.1. All strains were stocked in 15% glycerol and stored at -80°C. Plasmids and primers used are listed in Table 2.2 and Table 2.3 respectively.

Table 2.1. Bacterial strains used in this work.

Strain	Description	Reference
Rlv3841	*Rhizobium leguminosarum* biovar *viciae* 3841; Str^R derivative of strain 300.	(Johnston and Beringer, 1975)
SBW25	*Pseudomonas fluorescens* SBW25; wild-type strain isolated from phyllosphere of sugar beet plant	(Rainey and Bailey, 1996)
DH5α	Competent *E. coli* strain carrying the following mutations; F- *deoR endA1 recA1 relA1 gyrA96 hsdR17(rk-mk+) supE44 thi-1 - phoA Δ(lacZYA-argF)*U169 Φ80*lacZ*ΔM15 λ	Bioline
TransforMax™ EC100D™ *pir*-116	Electrocompetent *E. coli* strain carrying the following mutation; F- *mcrA* Δ(*mrr*-hsdRMS-*mcr*BC) ϕ80d*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara*, *leu*)7697 *gal*U *gal*K λ– *rps*L *nup*G *pir*-116(DHFR).	Lucigen
pSAM_RI	SM10λ*pir* carrying pSAM_RI *Mariner* transposon vector; Neo^r, Amp^r, Kan^r	(Perry and Yost, 2014)
pSAM_RI_RBP1	*E. coli* pool 1 collectively carrying millions of the modified pSAM_RI vectors containing a randomly barcoded transposon.	This work
pSAM_RI_RBP2	*E. coli* pool 2 collectively carrying millions of the modified pSAM_RI vectors containing a randomly barcoded transposon.	This work

Table 2.2. Plasmids used in this work.

Plasmid	Description	Reference
pSAM_RI	*Mariner* transposon vector. Amp^r, Kan/Neo^r	(Perry and Yost, 2014)
pRK2013	Helper plasmid for triparental conjugation. Kan/Neo^r	(Ditta et al., 1980)

Table 2.3. Primers used in this work.

Primers	Sequence 5' → 3'	Description
BCP_1	GATGTCCACGAGGTCTCTNNNNNNNNNN NNNNNNNNNNCGTACGCTGCAGGTCGAC	Oligonucleotide pool of random barcodes flanked by universal primer binding sites
oxp3407	GTCCCCGGTCTTCGTATGCCGTCTTCT GCTTGGCGCGCCCTCGAGGATGTCC ACGAGGTCTCT	BCP_1 amplification primer (forward)
oxp3408	ATTCCGGTTCGCTTGCTGTCCATAAAACC GCCCAGTCTACTCGAGGGGTCGACCTG CAGCGTACG	BCP_1 amplification primer (reverse)
BioSam	/BiotinTEG/CGGTTCGCTTGCTGTCCAT AAAACC	Ion Torrent BioSAM with 5' Biotin TEG (Perry and Yost, 2014)
M12 Top	CTGTCCGTTCCGACTACCCTCCCGAC	M12 adaptor (Goodman et al., 2009)
M12 Bottom	GTCGGGAGGGTAGTCGGAACGGACAG	M12 adaptor (Goodman et al., 2009)
oxp0437	AGATCGGAAGAGCGTCGTGTAGGGAA	INSeq_Adpt_Top (Goodman et al., 2009)

oxp0438	TTCCCTACACGACGCTCTTCCGATCTNN	INSeq_Adpt_Bottom (Goodman et al., 2009)
IT_trP1_FP	CCTCTCTATGGGCAGTCGGTGATTTCCCTACACGACGCTCTTCCGATCT	Ion Proton universal 3'sequencing adaptor (Perry and Yost, 2014)
IT_A_FP_1	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAAGGTAACGATATAAAACCGCCCAGTCTACTCGAGGG	Ion Xpress 5' barcoded sequencing adaptor (Perry and Yost, 2014)
oxp4021	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTAAGGTAACGATGATGTCCACGAGGTCTCT	IT_BarsSeq_1 (Barcode is underlined)
oxp4022	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAAGGAGAACGATGATGTCCACGAGGTCTCT	IT_BarsSeq_2
oxp4023	CCATCTCATCCCTGCGTGTCTCCGACTCAGAAGAGGATTCGATGATGTCCACGAGGTCTCT	IT_BarsSeq_3
oxp4024	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACCAAGATCGATGATGTCCACGAGGTCTCT	IT_BarsSeq_4
oxp4025	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGAAGGAACGATGATGTCCACGAGGTCTCT	IT_BarsSeq_5
oxp4026	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGCAAGTTCGATGATGTCCACGAGGTCTCT	IT_BarsSeq_6
oxp4027	CCTCTCTATGGGCAGTCGGTGATGTCGACCTGCAGCGTACG	IT_BarSeq_3' specific extension
oxp4156	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCGTGATTCGATGATGTCCACGAGGTCTCT	IT_BarsSeq_7

oxp4157	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTCCGATAACGATGATGTCCACGAGGTCTCT	IT_BarsSeq_8
oxp4158	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGAGCGGAACGATGATGTCCACGAGGTCTCT	IT_BarsSeq_9
oxp4159	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACCGAACGATGATGTCCACGAGGTCTCT	IT_BarsSeq_10
oxp4160	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCCTCGAATCGATGATGTCCACGAGGTCTCT	IT_BarsSeq_11
oxp4161	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGGTGGTTCGATGATGTCCACGAGGTCTCT	IT_BarsSeq_12
oxp4162	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAACGGACGATGATGTCCACGAGGTCTCT	IT_BarsSeq_13
oxp4163	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGGAGTGTCGATGATGTCCACGAGGTCTCT	IT_BarsSeq_14
oxp4164	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAGAGGTCGATGATGTCCACGAGGTCTCT	IT_BarsSeq_15
oxp4165	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTGGATGACGATGATGTCCACGAGGTCTCT	IT_BarsSeq_16

2.3. Molecular techniques

2.3.1. Isolation of genomic and plasmid DNA

Genomic DNA was isolated for PCR amplification using a DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's instructions with an additional step of mechanical lysis in a FastPrep-24™ instrument (MP Biomedicals) at speed 6.5 for 90 s after incubation at 56°C. Plasmid DNA was extracted from *E. coli* strains using a GeneJET Plasmid Miniprep Kit (TermoFisher Scientific) as per the manufacturer's recommendations. DNA concentration and purity was analysed using a NanoDrop 1000 (ThermoFisher Scientific) or Qubit 2.0 fluorometer (Invitrogen) where required.

2.3.2. DNA amplification by PCR

Primers used for PCR amplification were designed using Geneious R10 (v10.2.6) and synthesised by Eurofins MWG Operon. Q5® High fidelity DNA polymerase (NEB) was used for the amplification of DNA products required for cloning and sequencing. Cycling conditions were generally 98°C 30 sec (x 1); 98°C 10 sec, 50-72°C 30 sec, 72°C 20 s per kb (x 30); 72°C 2 min (x 1). Analytical PCRs were with OneTaq® Master Mix (NEB). Cycling conditions were as follows unless stated otherwise 94°C 30 sec (x 1); 94°C 30 s, 45-68°C 30 s, 68°C 60 s per kb (x 30); 68°C 5 min (x 1).

2.3.3. Agarose gel electrophoresis

PCR products and plasmid restriction digests were electrophoresed at 100 V for 45 min in 1-2% (w/v) agarose (Sigma Aldrich) dissolved in 1 x TAE buffer (400 mM TRIS, 1mM EDTA, pH 8.3), containing 2 µL 100 mL^{-1} SYBR® Safe DNA Gel Stain (Invitrogen). 1 kb plus or 50 bp DNA ladders (ThermoFisher Scientific) were used as

molecular weight markers and gels were visualized on a Gel Doc EZ System (BioRad). When required, Sanger sequencing of PCR amplicons and plasmids was performed by Eurofins Scientific.

2.3.4. Restriction endonuclease digestions, ligations and High-Fidelity assembly

Amplified PCR products used for cloning were purified prior to digestion and ligations using a Monarch® PCR&DNA Cleanup Kit (NEB) as per the manufacturer's recommendations. Restriction digestions and ligations of DNA were performed using the buffers, temperatures and incubation times indicated by NEB. Restriction enzymes were inactivated by incubation at 65°C for 20 min or removed by purification with a Monarch® PCR&DNA Cleanup Kit (NEB) where products were digested with heat-tolerant enzymes. Ligation of cohesive end fragments was performed using T4 DNA ligase (ThermoFisher Scientific), whilst ligation of blunt end fragments was performed with a CloneJET PCR Cloning kit (ThermoFisher Scientific) both as per the manufacturers' recommendations. NEBuilder HiFi DNA assemblies were performed using the NEBuilder HiFi DNA assembly Master Mix as per the manufacturer's instructions.

2.3.5. Transformation of competent cells

Plasmids were transformed into *E. coli* either by chemical transformation or electroporation. Chemically competent DH5α cells were purchased from Bioline. For chemical transformations, 50 µL of competent cells were combined with 5 µL of a ligation mixture or 5-10ng of DNA (max 5 µL) and incubated on ice for 30 min, heat

shocked at 42°C for 60 s and incubated on ice for a further 2 min. For electroporation, 100 ng (max 5 µl) of DNA was added to 50 µL of TransforMax™ EC100D™ pir-116 electrocompetent *E. coli* cells (Lucigen Simply Genomics) and transferred into a 0.2 mm gap cuvette (Scientific Laboratory Supplies). A single 5 ms 2.5 kV pulse was delivered using a MicroPulser electroporator (BioRad). For both methods, 950 µL of SOC recovery medium (2% tryptone, 0.5% yeast extract, 10mM NaCl, 2.5 mM KCL, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose; Invitrogen) was immediately added directly to the transformation tube or cuvette and the entire contents transferred to a 15 mL falcon tube which was then incubated at 37°C on a rotary shaker (220 rpm) for 1-h. Transformation reactions were serially diluted, spread onto LB agar supplemented with the appropriate selective antibiotics and incubated overnight at 37°C.

2.3.6. Conjugal transfer of plasmids

Triparental matings were used for the transfer of plasmids from *E. coli* donors to *Rhizobium* and *Pseudomonas* recipient cells. *E. coli* donors and *E. coli* containing the helper plasmid pRK2013 were grown overnight in 10 mL LB broth supplemented with appropriate antibiotics. For matings with *Rhizobium*, *E. coli* strains were sub-cultured (1:10 v/v) into 10 mL fresh LB media supplemented with appropriate antibiotics and grown at 37°C with 160 rpm shaking until mid-log phase (OD_{600nm} 0.4-0.6). Cells were harvested by centrifugation (3,800 x g, 5 min), washed twice with 1 mL TY and resuspended in 1 mL TY. *Rhizobium* were grown for 3 days prior to conjugation on a TY slope (10 mL TY set at a 30° angle in a 25 mL universal tube) and resuspended in 5 mL TY on day of conjugation (OD_{600nm} ~3.0). Recipient rhizobia, donor *E. coli* and

helper *E. coli* were mixed in a 2:2:1 ratio, pelleted via centrifugation, resuspended in 30 µl TY, spotted onto a sterile nitrocellulose filter placed on a TY agar plate and incubated ~24 h at 28°C. For matings with *Pseudomonas, Pseudomonas* cultures were grown overnight in 10 mL KB media (OD_{600nm} ~2.0). 3 mL of overnight *E. coli* helper, *E. coli* donor and *Pseudomonas* cultures were harvested, washed twice in KB, resuspended in 1.5 mL KB and heat shocked for 15 min at 37°C and 45°C respectively. Recipient *Pseudomonas*, donor *E. coli* and helper *E. coli* were combined in a 7:3:1 ratio, pelleted via centrifugation, resuspended in 50 µl LB prior to spotting onto a KB agar plate and incubated overnight at 28°C. Mating spots were resuspended in TY or KB as appropriate and serial dilutions were spread onto selective media. Single colonies were passaged to remove potential contamination prior to use.

2.4. Bacterial root colonisation assays

2.4.1. Preparation of growth vessels

Boiling tubes (100 mL, DURAN®) were three-quarters filled with fine vermiculite, supplemented with 25 mL N-free rooting solution ($CaCl_2$•H_2O 1 mM, KCl 100 µM, $MgSO_4$•$7H_2O$ 800 µM, Fe EDTA 100 µM, H_3BO_3 35 µM, $MnCl_2$•$4H_2O$ 9 µM, $ZnCl_2$ 0.8 µM, Na_2MoO_4•$2H_2O$, 0.5 µM $CuSO_4$•$5H_2O$ 0.3 µM, KH_2PO_4 0.5 g L^{-1} and Na_2HPO_4 0.568 g L^{-1}), sealed with a foam bung and autoclaved.

2.4.2. Seed sterilisation

Pea seeds were surface sterilised by immersion in 70% ethanol for 30 s, followed by 2.8% sodium hypochlorite for 5 min, and rinsed 6 times with sterile dH_2O. Lentil seeds were surface sterilized by immersion in 0.56% sodium hypochlorite for 5 min and

washed 6 times with sterile dH_2O. *Lathyrus* seeds were submerged for 8 min in sulfuric acid and washed 6 times with sterile dH_2O. Broad bean seeds were immersed in 2.8% sodium hypochlorite for 5 min, washed 6 times with sterile dH_2O and soaked overnight in sterile water. Alfalfa seeds were immersed in 95% ethanol for 45 min, followed by 5.25% bleach for 15 min, and rinsed 6 times with sterile dH_2O. *H. vulgare* seeds were surface sterilised in 70% ethanol for 2 min, 7% sodium hypochlorite + 0.1% Tween20 (Sigma Aldrich) for 5 min and rinsed 6 times with sterile dH_2O. Sterile seeds were placed directly into sterilised boiling tubes to germinate in a controlled growth chamber set at 23°C with a 16 h light 8 h dark photoperiod.

2.4.3. Inoculation of plants

Seven days post-germination seedlings were inoculated with 1 $x10^5$ cfu of bacterial cells resuspended in 1 mL phosphate buffered saline (PBS). For co-inoculation assays 1 $x10^5$ cfu of each bacterial strain was inoculated. Rlv3841 inoculum was generated by streaking single colonies of Rlv3841 onto TY agar slopes which were grown at 28°C for 3 days. Slopes were washed in 5 mL PBS and the bacterial suspension was transferred to 15 mL falcon tubes and centrifuged (3,800 x g, 5 min). Cells underwent two further cycles of washing in PBS under the same conditions, before normalisation to an OD_{600} 1 (10^9 cfu mL^{-1}) and serial dilution to the desired cell density. SBW25 inoculum was generated by inoculating a single colony of SBW25 into 10 mL LB media and grown overnight at 28°C shaking at 160 rpm. Cultures were transferred to 15 mL falcon tubes and cells were retrieved by centrifugation (3,800 x g, 5 min), then washed 3 x in PBS before normalising to OD_{600} 1 ($5x10^8$ cfu mL^{-1}) and serially diluting to the desired cell density.

2.4.4. Retrieval of rhizosphere and root-associated bacteria

The population of bacteria in the rhizosphere and associated to roots was assayed at 1-, 3-, 5- and 7-days post inoculation (dpi). To obtain the rhizosphere population the vermiculite surrounding the roots was separated by removing the seedling, 30 mL PBS was added to the boiling tube. The PBS and vermiculite were vortexed thoroughly for 30 s and filtered through 4 layers of muslin cloth into a sterile 50 mL falcon tube. To retrieve bacteria loosely associated to the rhizoplane, roots were detached at the seed, dipped in sterile dH_2O to remove vermiculite, and vortexed vigorously in 25 mL PBS for 1 min. Following this, roots were removed and ground in 10 mL PBS using a pestle and mortar to retrieve bacteria strongly associated to the rhizoplane and those in the endosphere. Both samples were independently filtered through 4 layers of muslin cloth into sterile falcon tubes. To calculate bacterial populations 25 µL drop counts of appropriate serial dilutions were plated onto TY Str Ny for Rlv3841 or LB Nitro N for SBW25. Plates were incubated at 28°C and colonies counted after ~24 h for SBW25 and after ~48 h for Rlv3841.

2.5. RB-TnSeq

2.5.1. Generation of *E.coli* barcoded *mariner* transposon donor pool

To generate a barcoded transposon mutant library, first an *E.coli* pool was generated in which each cell contained a plasmid carrying a barcoded mariner transposon. To generate an *E. coli* barcoded transposon pool, an oligonucleotide pool (BCP_1) in which each oligonucleotide consists of a random 20-nuclotide barcode flanked by universal primer binding sites was ordered from Eurofins genomics (Table 2.3). The *mariner* delivery vector pSAM_RI was digested with XhoI and the barcode pool was

PCR-amplified with primers oxp3407 and oxp3408 to add extensions required for HiFi cloning directly into XhoI digested pSAM_RI (Table 2.3). For each *E. coli* barcoded transposon library five independent PCRs were completed, each with 5 ng of the oligonucleotide library. PCR reactions and XhoI digested pSAM_RI were run on 2% and 0.8% agarose gels respectively, then gel extracted using the Monarch® DNA Gel Extraction Kit (NEB) as per the manufacturer's instructions. Five independent HiFi reactions were set up with a 5:1 insert to vector ratio so that each reaction contained 100ng of XhoI digested PSAM_RI and 15.77ng of the amplified barcode pool. HiFi reactions were pooled, purified by ethanol precipitation (Green and Sambrook, 2016), resuspended in 30 µl TE (10mM TRIS-HCL pH 8.0 and 0.1 mM EDTA) and quantified using the Qubit fluorometer. At least 100ng (5µl maximum) of the purified HiFi reaction was electroporated into 50 µl of TransforMax™ EC100D™ pir-116 electrocompetent cells as previously described. Following 1-h recovery in SOC, 10 µl were taken for serial dilution and spread plated onto LB (Kn Amp) for quantification. The remainder of the cells were diluted in 10 mL LB supplemented with Kn Amp and grown overnight at 37 °C with orbital shaking at 160 rpm. The following morning, glycerol was added to 15% and 1 mL aliquots were flash frozen in liquid nitrogen then stored at -80 °C. Two *E. coli* pools containing the modified pSAM_RI mariner delivery vector were independently generated following this method, namely pSAM_RI_RBP1 and pSAM_RI_RBP2. This was done to ensure minimal overlap in random barcodes between the two pools which will subsequently be used to generate RB-TnSeq transposon mutant libraries in Rlv3841 and SBW25.

2.5.2. Generation of RB-TnSeq *mariner* transposon mutant libraries in Rlv3841 and SBW25

Conjugation of *E. coli* pSAM_RI_RBP1 with Rlv3841 was conducted on TY:UMS agar as described above with one modification, to prepare the donor *E.coli* 1 vial of *E.coli* pSAM_RI_RBP1 was thawed and inoculated into 100 mL of LB supplement with Kan + Amp. After growth overnight, the *E. coli* pSAM_RI_RBP1 culture was sub-cultured (2:10 v/v) into 10 mL fresh LB broth supplemented with Kan + Amp. After 24 h incubation at 28^{o}C the mating spot was scraped, resuspended in 1 mL UMS supplemented with 15% glycerol and stored at -80^{o}C. Total number of transposon insertion mutants in the pool was calculated by thawing a small chip of the resuspended conjugation and plating serial dilutions onto Neo + Nitro TY:UMS. To clean the library, ~3.6 x10^{6} mutants were plated onto 12 separate 245 x 245 mm (Corning) Neo + Nitro TY:UMS agar plates (300,000/ plate) and incubated at 28^{o}C until colonies were pinprick in size. Following incubation, colonies were scrapped off each plate, combined and resuspended in TY:UMS. The resuspension was diluted in Neo + Nitro TY:UMS to an OD_{600} 0.1 and grown at 28 ^{o}C with shaking at 150 rpm until the culture reached an OD_{600} 1.0, at which point glycerol was added to a final concentration of 20% and 1 mL aliquots were made and store at -80^{o}C. The same process was followed to construct a barcoded mariner library in SBW25 with the following modifications: *E. coli* pSAM_RI_RBP2 was conjugated into SBW25, KB was used in place of TY:UMS and libraries were cleaned on Nitro + Kan KB agar.

2.5.3. Mapping of RB-TnSeq libraries using INSeq

To map the *mariner* insertion sites within each transposon library, the protocol for INSeq describe in Wheatly et al. (2017) was used. Briefly, gDNA was isolated from the transposon library and 2000 ng was used in a linear PCR reaction during which linear PCR products were amplified with biotin tags using the BioSAM primer (Table 2.3). Biotinylated PCR products were PCR purified and bound to Pierce Streptavidin Magnetic Beads (Thermo Fisher). Samples were denatured by heating prior to Klenow fragments (3'→5' exo$^-$, NEB) being used for second strand synthesis. Samples were subsequently digested with MmeI, using a M12 fragment (M12 top and M12 bottom primers annealed) *in trans* for efficient DNA cleavage and an Ion Proton sequencing adaptor was ligated (oxp0437 and oxp0438 annealed; Table 2.3). A final round of PCR amplification was conducted to add the 5' sample specific Ion Xpress barcode adaptor (IT_A_FP_1) (Thermo Fisher), to allow automated read separation by samples and the universal Ion Proton 3'sequencing adaptor (IT_trP1_FP) (Table 2.3). The resultant 251 bp fragments were PCR purified then size selected using a E-Gel SizeSelect II agarose gel (Thermo Fisher). Sample concentration was determined using a Agilent High Sensitivity DNA kit (Agilent Technologies), prior to dilution to 75 pM. An Ion Chef and Ion PI chip kit V3 (Thermo Fisher) was used for template preparation of the libraries and sequencing was perform on an Ion Proton system (Thermo Fisher).

Transposon insertion locations and associated barcodes were mapped and identified following the previously described RB-TnSeq pipeline with some modifications (Wetmore et al., 2015); the code is available at https://bitbucket.org/berkeleylab/feba/. Briefly, reads were analysed using a perl script (MapTnSeq.pl), which identifies the random 20-nuclotiede barcode along with the junction between the transposon and

the genome, and maps the remainder of the read i.e., the genomic tag, to the genome. To identify the 20-nucleotide barcode the script looks for the U1 and U2 sequences flanking it. The 6 nucleotides on each side of the barcode must exactly match the corresponding U1 and U2 sequence and each nucleotide of the barcode must have a minimum quality score of 10 (90% base call accuracy). Next, the script identifies the end of the transposon by requiring an exact match for the last 5 nucleotides. If the sequence past the end of the transposon is at least 16 nucleotides long, it is then mapped to the genome using BLAT (Kent, 2002). The result is a table of mapped mutants (reads) with their barcodes and associated genomic insertion locations.

Since the *mariner* transposon only inserts at TA sites, reads that did not strictly map to a TA site were computationally removed from the MapTnSeq output file. The filtered MapTnSeq output file was then passed to a perl script (DesignRandomPool.pl) to identify barcodes that consistently map to a unique genomic location. A barcode is considered suitable for use in the calculation of gene fitness values if it has at least 5 reads supporting its mapping to the primary genomic location, the primary matches account for 75% or more of the reads associated with this barcode and the minimum ratio of reads mapping to the primary genomic location over the second most frequent mapping site is at least 0.8.

2.5.4. RB-TnSeq library inoculation *in planta* for BarSeq analysis

RB-TnSeq libraries were thawed from -80^{o}C stocks and inoculated onto plants as described in 2.4.3 with the exception that 2x10^{5} cfu were inoculated per plant for single inoculation experiments on lentil, *Lathyrus*, alfalfa and barley to maximise

maintenance of diversity at the initial inoculum stage. For pea, $1x10^5$ cfu was inoculated per plant. For co-inoculation experiments $2.5x10^2$ cfu of each RB-TnSeq library was inoculated per plant. The remainder of the thawed libraries were pelleted and stored at -20^{o}C until DNA extraction. The rhizosphere and root-associated libraries were retrieved 7 dpi as described in 2.4.4 however, retrieved cells were harvested by centrifugation (17,000 x g, 15 min) and the resultant pellets stored at -20^{o}C ready for gDNA extraction.

2.5.5. BarSeq library preparation and sequencing

For BarSeq experiments, gDNA was isolated from mutant library samples as described in section 2.3.1 using a DNeasy Blood and Tissue kit (Qiagen). Samples were eluted in 50 µl of water and 1 µl of this was used in subsequent PCR reactions. In some instances, samples were diluted further to prevent inhibition of PCR reactions. BarSeq PCRs were performed in 50 µl reactions with 20 µmol of each primer and 1 µl of template gDNA. All samples were amplified with a common reverse primer (oxp4027) which binds the U1 region and adds the Ion Proton 3' sequencing adaptor and one of 16 forward primers (oxp4021-26 and oxp4156-4165). The forward primers bind the U2 region, adding the 5' Ion Proton adaptor along with 1 of 16 Ion Express barcodes. Cycling conditions were as follows 98°C 4 min (x 1); 98°C 30 sec, 68°C 30 sec, 72°C 30 s (x 30); 72°C 5 min (x 1). For each sample, at least 5 independent PCR reactions were conducted, and equal volumes were pooled prior to purification with a Monarch® PCR&DNA Cleanup Kit (NEB). The final BarSeq library was eluted in 50 µl of water and quantified using the Agilent High Sensitivity DNA kit (Agilent Technologies). BarSeq libraries were diluted to 75 pM and sequenced on an Ion

Proton system following template preparation using the Ion Chef and Ion PI chip kit V3 (Thermo Fisher). Four samples were typically pooled per Ion PI chip.

2.5.6. Calculation of gene fitness values from BarSeq experiments

To determine a fitness value for each gene under a given condition, the relative abundance of barcodes for each mutant that maps to that gene is compared after treatment vs time zero. A perl script (MultiCodes.pl) is used to identify the barcode in each read and make a table containing all unique barcodes and how often each were seen (i.e. the number of reads with that barcode). This script identifies the barcode in a read by searching for the universal primer binding sites (U1 and U2). For a barcode to be identified, the 8 bps immediately upstream and downstream of the 20-nucleotide barcode must match exactly to the expected sequence of the universal primer binding sites. Next, a second perl script (combineBarSeq.pl) takes the table of barcodes and combines it with the table of mapped mutants (the output of the DesignRandomPool.pl script; section 2.5.3) to make a new table of how often each mutant was seen following growth in each treatment or at time zero. Lastly, a R script (BarSeqR.pl) combines this table with a genes table and uses a second custom R script (FEBA.R) to first calculate mutant fitness values and then gene fitness values followed by a measure of their reliability (*t*-score). For each experiment, strain fitness values are calculated as the $\log_2$ ratio of barcode relative abundance following library growth in a given condition divided by the relative barcode abundance in the initial inoculum at time zero. The fitness of the gene is then calculated as the weighted average of the fitness values calculated for the mutants that have insertions within that gene. The data is normalised across the genome and independently for each scaffold so that the typical 'neutral' gene has a fitness of zero meaning that there is no difference in the number of barcode

reads associated with that gene before and after treatment (Wetmore et al., 2015). Genes with a fitness value greater than -2 are considered to have strongly disadvantaged phenotypes, with mutation resulting in at least a 75% reduction in growth relative to the average mutant in the population (Table 2.4). Genes with a fitness value between -1 and -2 are considered to be mildly disadvantaged since mutation of that gene results in 50 to 75% less growth relative to the average mutant. Genes with a fitness value between -1 and 1 are considered to be 'neutral', whilst genes with a fitness value between 1 and 2 are considered to be mildly advantageous and those with a fitness value greater than 2 are considered strongly advantageous meaning these genes, when not mutated, are in some way deleterious for growth in that particular condition.

Table 2.4. Gene fitness values related to the percentage of change in barcodes between time zero and following treatment.

Mutant phenotype	Gene fitness value	Percentage change of barcodes after treatment vs time zero[a]
Strongly advantaged	2	+ 400% (x4)
Mildly advantaged	1	+ 200% (x2)
Neutral	0	0
Mildly disadvantaged	-1	-50% (x0.5)
Strongly disadvantaged	-2	-75% (x0.25)

[a]The percentage change represents the growth of a strain between time zero and after exposure to a given condition, relative to an average mutant in the pool. Fold change relative to time zero is shown in brackets.

Each gene fitness value has an associated *t*-score, representing a *t*-like statistic used to assess the reliability of the assigned gene fitness value. The *t*-score is calculated as the genes fitness divided by the standard error, where the standard error is based on the consistency of the fitness scores for the mutants in that gene (Wetmore et al., 2015; Price et al., 2018). Generally, for genes with fitness values less than -2 there is high confidence in the prediction if the associated *t*-score is less than -3 (Helmann et al., 2019). For genes with fitness values between -1 and -2 there is high confidence in the prediction if the associated *t*-score is less than -2.5 (Helmann et al., 2019). For genes with advantaged phenotypes high confidence in gene fitness values between 1 and 2 or greater than 2 are indicted by *t*-scores greater than 2.5 and 3 respectively.

BarSeq experiments must pass several quality controls for there to be confidence in their gene fitness predictions. The median gene must have at least 50 reads, the correlation between the GC content and gene fitness value must be less than 0.2 and the Spearman correlation of adjacent genes on different strains must be no more than 0.25 (Wetmore et al., 2015).

Chapter 3

Development of RB-TnSeq in Rvl3841 and validation of BarSeq *in planta*

.

3.1. Introduction

The past decade has seen a surge in microbial genome sequencing, made possible by advances in sequencing technologies combined with a reduction in the associated costs. Consequently, there has been a demand for tools capable of relating genome content to function. This resulted in the parallel development of several techniques that allow the study of gene function at the whole genome level, namely: transposon-directed insertion site sequencing (TraDIS), high-throughput insertion tracking sequencing (HITS), TnSeq and INSeq (van Opijnen and Camilli, 2013). They all follow the same basic principles of generating large libraries of single insertion transposon mutants which collectively saturate the organism's genome with transposition events. Afterwards, these libraries are exposed to a particular set of conditions and next-generation sequencing is used to assess how the insertion mutation in a particular gene affects the survival of the bacteria in a given condition. INSeq and TnSeq vary from HITS and TraDIS in the fact that they make use of restriction enzymes during library preparation and therefore do not require the mechanical shearing of DNA. TnSeq and INSeq are very similar although modification of INSeq library preparation has reduced the amount of sample and enzyme needed leading to increased sensitivity relative to TnSeq (van Opijnen and Camilli, 2013; Cain et al., 2020).

To date, most studies have used either the Tn5 or *mariner* transposon for library construction. The major difference between the two is that Tn5 inserts at random within the genome, whilst the *mariner* transposon only inserts into thymine-adenine ('TA') motifs (Muñoz-López and García-Pérez, 2010). This can serve as a disadvantage in guanine-cytosine rich species. However, the defined insertion motif provides the

advantage of allowing transposition events to be modelled *in silico*. Thus, data analysis does not rely on the comparison of an input and output mutant pool (Barquist et al., 2013; Perry and Yost, 2014).

The *mariner* transposon is a Class II transposon that consists of a transposase gene flanked by two terminal inverted repeats (TIRs) (Fig. 3.1A) and moves via a 'cut and paste' mechanism (Fig. 3.1B).

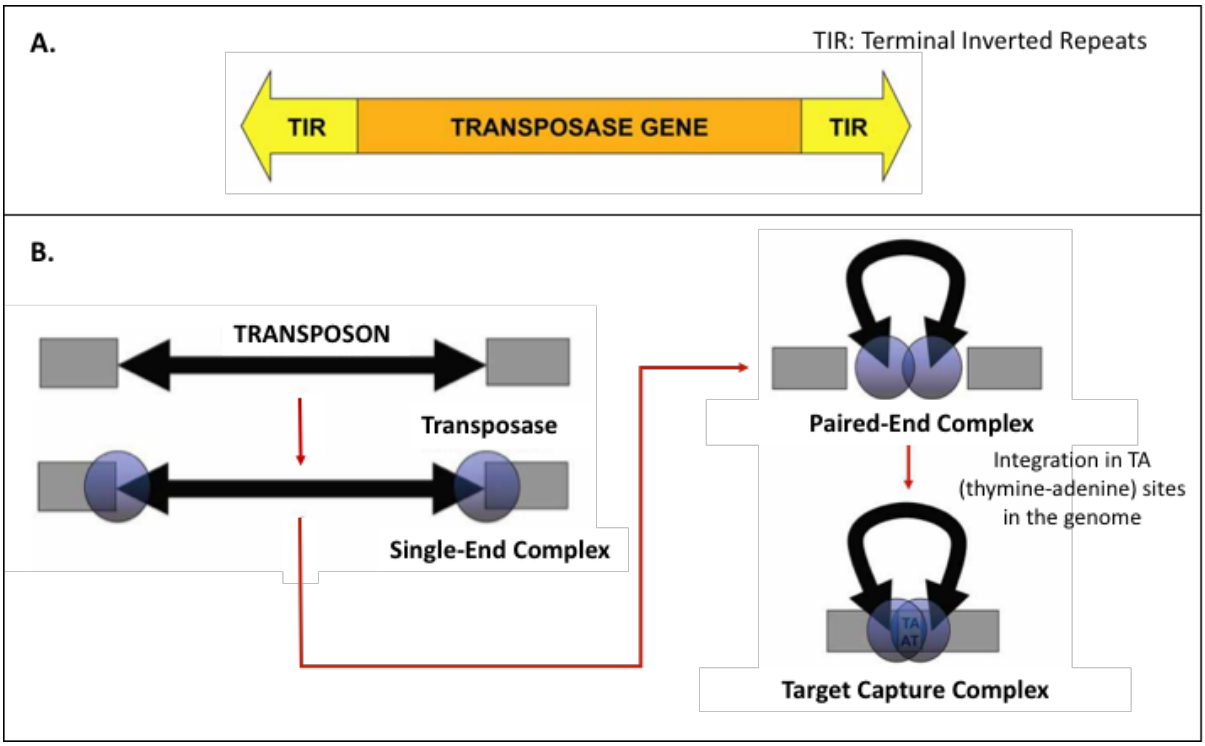


Figure 3.1 Transposition events of the *mariner* transposon. A. The *mariner* transposon consists of the transposase gene flanked by two TIRs. **B.** The mechanism of transposition. Two transposase monomers recognise and bind the TIRs to form a 'single-end' complex. The two transposase monomers then bind to form a dimer, bringing together the two ends of the transposon to form a 'paired-end' complex resulting in excision of the transposon. Lastly, the transposase dimer recognises a TA

site within the genome, which it binds to form the 'target capture complex' resulting in insertion of the transposon. This figure was adapted from reference (Muñoz-López and García-Pérez, 2010).

The realisation that a single guanine to cytosine transversion in the *mariner* transposon TIRs would generate a recognition sequence for the Type IIs restriction enzyme MmeI led to the development of INSeq (Goodman et al., 2009). When directed to this site MmeI cleaves 16 bps outside of the transposon. This allows for the capture of the 16 bps of adjacent genomic DNA and thereby the identification of the genomic location into which the transposon has inserted upon sequencing. These 16 bps of genomic DNA are sufficient to uniquely map each transposition event in bacteria where the genome sequence is known. Construction of a sequencing-adapted *mariner* transposon delivery vector (pSAM) enabled efficient delivery of the *mariner* transposon and the possibility to change the promoter driving expression of the transposase gene (Goodman et al., 2009) (Fig. 3.2A). More recently, pSAM has been adapted (pSAM_RI) and demonstrated to work within the *Rhizobiaceae* family, which includes the agriculturally important species *Rhizobium*, *Sinorhizobium* and *Agrobacterium* (Fig. 3.2B) (Perry and Yost, 2014).

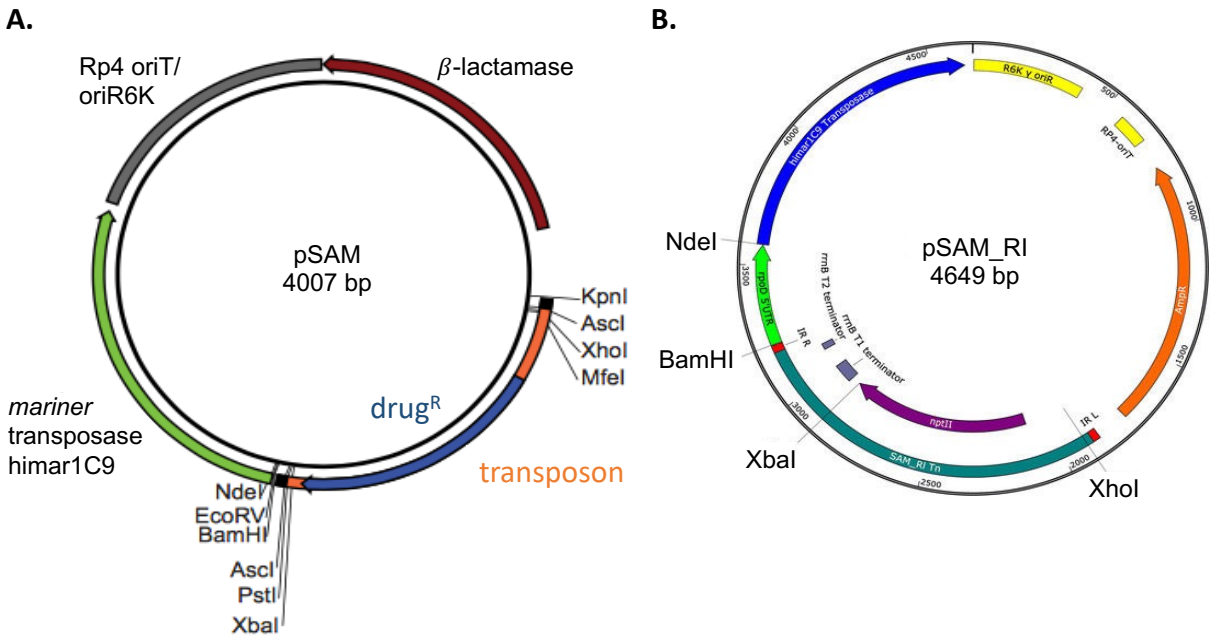


Figure 3.2. Features of pSAM and pSAM_RI. A. pSAM. The three major features are: an antibiotic resistance cassette flanked by MmeI-modified *mariner* TIRs, a multiple cloning site immediately upstream of the *himar1C9 mariner* transposase, and machinery for replication in the donor strain and transfer by conjugation. Marked restriction sites are unique, except for AscI. **B.** pSAM_RI. The original ampicillin resistance gene was replaced with a neomycin/kanamycin resistance cassette and the promoter driving transposase expression was switched for the Rlv3841 *rpoD* promoter region. The restriction sites used are annotated. This figure was reproduced from references (Goodman et al., 2009; Perry and Yost, 2014).

The use of INSeq to screen for bacterial fitness traits across many conditions is largely limited by the requirement of mapping the transposon insertion location following exposure of the mutant pool to each treatment. This process is not only time-consuming, but also technically challenging for the identification of genes involved in

rhizosphere fitness and root colonisation. There are two reasons for this: firstly, recovery of mutant libraries from these environments is relatively low compared to laboratory grown cultures and so retrieved mutants must be cultured for several generations prior to DNA extraction to ensure sufficient quantities are obtained for INSeq library preparation; secondly, plant DNA inhibits certain steps of the INSeq library preparation and therefore recovering and culturing of bacteria retrieved from root niches is essential to reduce the ratio of plant to bacterial DNA.

Random barcode transposon-site sequencing (RB-TnSeq) can be thought of as an 'extension' of INSeq (Fig. 3.3) (Wetmore et al., 2015; Price et al., 2018). It overcomes many of the barriers discussed above by use of a unique, but random, 20-nucleotide DNA 'barcode'. The barcode is positioned within the transposon thereby allowing the identification of each individual insertion mutant within the transposon mutant library (Fig. 3.3A). Universal primer-binding sites (U1 and U2) flank the barcode and therefore the complex library preparation required for mapping of transposon insertion locations i.e., INSeq or Tn-Seq need only be done once to identify the insertion location (TA site) and the associated barcode (Fig. 3.3B). From then onwards, when using the same mutant pool, the universal primer binding sites can simply be used to amplify the barcode in a single PCR reaction and the resulting product is sequenced (Fig. 3.3C). It is then possible to identify the frequency of a barcode in the input and output pools and relate these back to the associated transposon insertion location (TA site). This latter phase of the RB-TnSeq process, in which the barcodes are amplified and sequenced, is commonly referred to as barcoded sequencing (BarSeq). Simplification of the sample processing steps enables increased throughput as demonstrated by the application of BarSeq in the assessment of gene fitness contributions for 32 bacterial

species grown in over 150 different conditions (Price et al., 2018). BarSeq has also been successfully applied to study bacterial root colonisation, with the identification of 115 genes in *P. simiae* required for optimal competitive colonisation of *A. thaliana* roots (Cole et al., 2017).

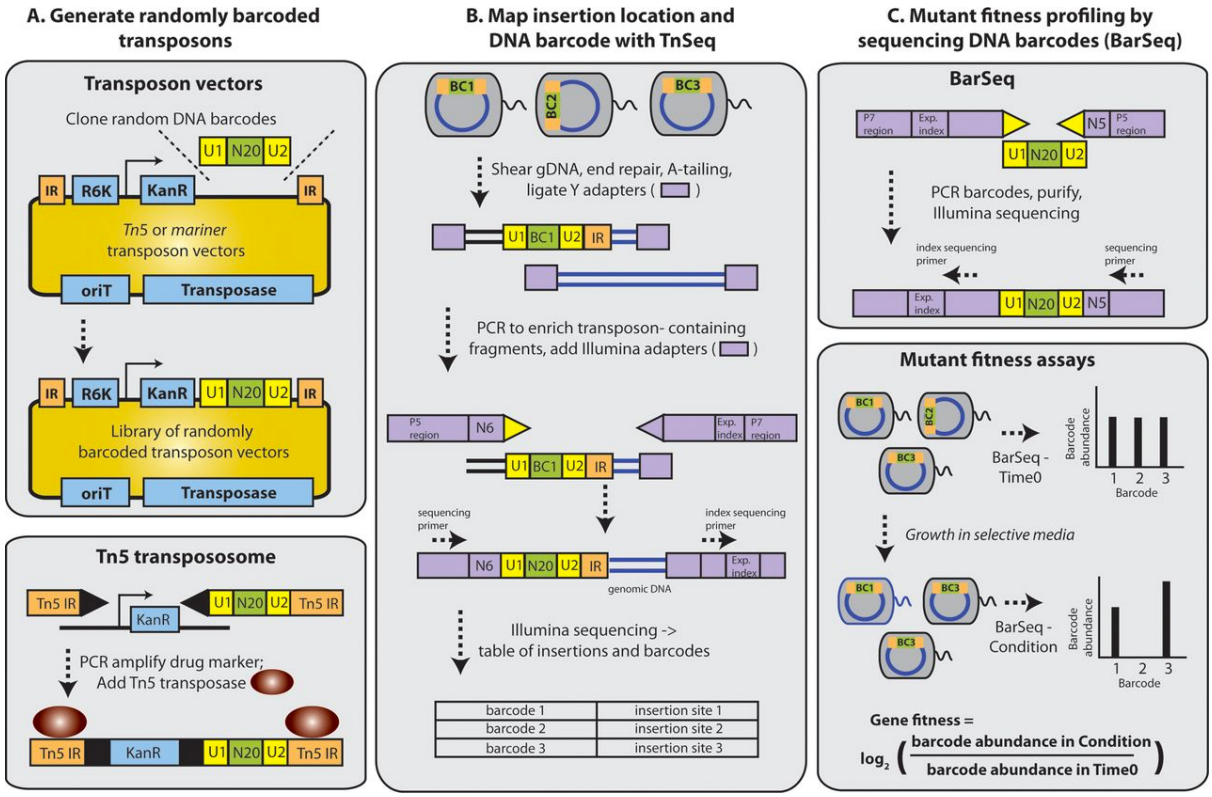


Figure 3.3. Overview of RB-TnSeq. A. Millions of *mariner* and Tn5 transposon delivery vectors were converted to RB-TnSeq vectors by cloning a unique DNA barcode (N20) flanked by common PCR priming sites (U1 and U2) near to the TIR. **B.** Characterisation of the randomly barcoded transposon mutant library is done using TnSeq or similar. **C.** In BarSeq, the DNA barcodes are PCR-amplified using oligonucleotides that bind the common U1 and U2 regions. Both oligonucleotides contain adapter sequences for Illumina sequencing. Competitive fitness assays are performed by comparing the abundance of the DNA barcodes with BarSeq before

(time zero) and after (condition) selective growth. In this example, the gene associated with barcode two has reduced fitness whilst the gene associated with barcode three has enhanced fitness. This figure was reproduced from reference (Wetmore et al., 2015).

Although these experiments are effectively large intraspecies competition assays with thousands of strains competing against one another, the use of random barcodes provides the possibility of conducting interspecies competition assays. The 20-nucleotide random barcode theoretically provides 10^{13} unique barcodes, theoretically allowing the generation of mutant libraries in multiple bacterial species with minimal chance the same barcode will appear among libraries. These libraries can then be co-inoculated in BarSeq experiments to simultaneously assess the gene fitness requirements of multiple bacterial species. Such multiplexing experiments will be crucial for analysing how bacteria interact with the plant and one another during competitive rhizosphere growth and root colonisation.

In this chapter I describe the development of RB-TnSeq in Rlv3841 and the validation of BarSeq *in planta* for the determination of Rlv3841 gene fitness contributions during growth in the rhizosphere and colonisation of pea (host legume). Pea was chosen for initial validation of BarSeq due to a wealth of experimental data available regarding its interactions with pea during the early stages of symbiosis. This includes INSeq screens for pea rhizosphere growth (5 dpi), root colonisation (5 dpi) (Wheatley et al., 2020) and root attachment (1 hpi) (Parsons, 2019); transcriptomics data for Rlv3841 during growth in the pea rhizosphere (7 dpi) (Ramachandran et al., 2011), and individual mutant studies that have identified several genes involved in Rlv3841 pea

rhizosphere fitness and root attachment (Williams et al., 2008; Frederix et al., 2014; Garcia-Fraile et al., 2015).

3.2. Results and discussion

3.2.1. Adaption of pSAM_RI for RB-TnSeq

In RB-TnSeq barcoded transposons are used to generate large mutant pools in which each mutant contains a single transposon insertion, ideally carrying a unique barcode. To convert the *mariner* transposon delivery vector, pSAM_RI, into an RB-TnSeq vector, millions of random 20-nucleotide barcodes were cloned into a unique restriction site located near to the TIR towards the end of the transposon (Fig 3.4). These barcoded vectors were transformed into *E. coli* and subsequently conjugated into Rlv3841 to generate single insertion mutants. It is important that most transposon mutants contain unique barcodes since in BarSeq experiments, only the barcode itself is sequenced and not the genomic location into which the barcoded transposon has inserted. To maximise the chance that each mutant contains a transposon carrying a unique barcode, the number of barcoded transposons in the *E. coli* pool must be sufficiently higher than the number of TA sites present in the Rlv3841 genome. The *E. coli* pool was estimated to contain 2.17 $x10^7$ (SEM: 7.41 $x10^6$) transformants. If each transformant carried the modified pSAM_RI vector with a unique barcode, this would equate to the presence of approximately 155 unique barcodes for each of the 140,056 TA sites in the Rlv3841 genome. In reality, it is likely that there is at least some repetition of barcodes within the *E. coli* pool. Nonetheless, this high ratio of barcodes to TA sites will ensure minimal instances in which the same barcode is found in more than one insertion mutant.

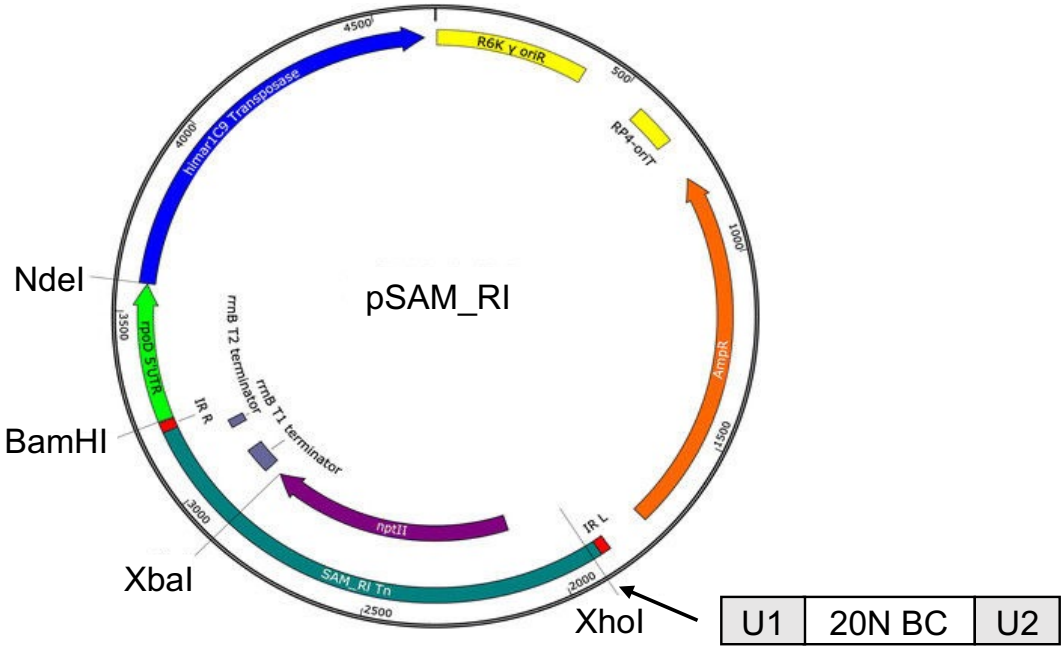


Figure 3.4 Adaptation of pSAM_RI for RB_TnSeq. Random 20-ncleotides barcodes flanked by common universal primer binding sties (U1 and U2) were cloned into the *mariner* transposon at the XhoI restriction site located close to the transposons left TIR. This figure was adapted from reference (Perry and Yost, 2014).

Prior to BarSeq experiments, a mapping run is used to simultaneously identify the transposon insertion location and associated barcodes for each mutant in the pool. This process enables the generation of a database of mutants in which every barcode sequence is present, along with its associated genomic location (TA site). In subsequent BarSeq experiments, the barcode is PCR-amplified, sequenced, and the corresponding genomic location is identified using the mutant database. Therefore, if the same barcode sequence is initially found to be present at multiple genomic locations during the initial mapping run, any data for those barcoded mutants cannot be used for the calculation of gene fitness values in BarSeq experiments. This is

because it is not possible to distinguish which mutant, and therefore which gene, the barcode is associated with.

3.2.2. Generation and mapping of Rlv3841 barcoded transposon mutant library

The barcoded *mariner* transposons were introduced into Rlv3841 through conjugation with the *E. coli* pool carrying the modified pSAM_Rl vector. A mapping run was conducted using INSeq to simultaneously map the genomic location of each insertion mutant and identify the associated barcode. Approximately 2.7 $x10^6$ unique barcodes were identified, of which 26% (711,108) were mapped to the genome. Due to chimeric PCR, whereby extension of an amplicon is aborted and the aborted product then functions as a primer in the next round of PCR, a barcode typically won't map to a single location 100% of the time (Wetmore et al., 2015). Nonetheless, barcodes that legitimately map to multiple locations in the genome must be removed. For a mutant and its associated barcode to be identified, uniquely mapped to the genome and deemed suitable for use in the calculation of gene fitness values it has to meet four requirements. First, a barcode must be seen at least five times. Only 57% (1.5 $x10^6$) of barcodes fell into this category with 43% (1.2 $x10^6$) seen less than five times. Second, a barcode must have at least five 'good' reads supporting its mapping to the preferred location. A read is considered good if it has a unique hit to the genome and the first base pair in the read, after the transposon sequence is removed, matches the genome. Third, the fraction of total reads for a given barcode going to the preferred mapping location, as opposed to any other site in the genome, must be at least 0.75; and fourth, the minimum ratio of reads mapping to the preferred genomic location over the second most frequent mapping site in the genome must be at least 0.8. Approximately 25% (669,399) of barcodes were excluded because they did not meet

the second and third requirements i.e., the barcode had less than five good reads supporting its mapping to a primary location and/or less than 75% of the total reads for that barcode mapped to the preferred location. A further 6% (152,300) were excluded for failing to meet the fourth requirement. Lastly, barcodes which map to the same location and are identical apart from a single nucleotide error are identified. The more common barcode is assumed to be correct and the least common barcode is removed. A further 1,396 barcodes were removed due to this, leaving 26% (711,108) of identified barcodes deemed suitable for use in the calculation of gene fitness values.

The 711,108 transposon mutants, each containing a unique barcode identifier, mapped to 100,662 (72%) of the 140,056 TA sites present in the Rlv3841 genome. The Rlv3841 genomes contains 7,528 protein coding genes distributed across the chromosome (4,927) and six plasmids: pRL12 (806), pRL11 (681), pRL10(479) pRL9 (306), pRL8 (149), pRL7 (180). Insertion densities within the genome ranged from 57-77% across the seven replicons, with an average insertion density of 0.72 (Table 3.1). BarSeq analysis only considers mutants containing insertions in the central 10-90% region of a gene, since even essential genes can often tolerate insertions in the N- and C- terminal regions without disruption to the function of the encoded protein (Miravet-Verde et al., 2020). Of the 711,108 mapped insertions mutants, 463,550 mapped to 63,162 distinct locations in the central 10-90% region of protein coding genes. There are 109,324 potential TA sites within the protein coding genes of Rlv3841, 85,935 of which are located within the central 10%-90% region. Therefore, the Rlv3841 mapped barcode pool contains insertion mutants covering over 73% of these central insertion sites. The high coverage of this mutant library makes it ideal for

use in assessing the fitness of nearly all nonessential protein-coding genes in the Rlv3841 genome by BarSeq.

Table 3.1. Insertion density of the Rlv3841 barcoded transposon mutant library.

Replicon	Size (bp)	Protein coding genes	Observed insertions[a] (mutants generated)	Potential unique insertions[b]	Unique insertions[c]	Insertion density[d]
Chromosome	5,057,142	4,927	479,540	89,660	63,382	0.71
pRL12	870,021	806	69,972	14,845	11,426	0.77
pRL11	686,202	681	63,098	12,456	9,514	0.76
pRL10	488,135	479	40,846	10,003	7,139	0.71
pRL9	352,782	306	30,807	6,359	4,754	0.75
pRL8	147,463	149	14,782	3,118	2,388	0.77
pRL7	151,546	180	12,063	3,615	2,059	0.57
Total	7,753,291	7,528	711,108	140,056	100,622	0.72

[a]Observed insertions is the number of uniquely barcoded mutants mapped to that replication.

[b]Potential unique insertions is the total number of TA sites present in the replicon.

[c]Unique insertions is the number of TA sites for which a barcoded transposon mutant was identified to have inserted.

[d]Insertion density is the fraction of TA insertion sites containing a barcoded transposon insertion.

3.2.3. Rlv3841 genes lacking transposon insertion mutants

Transposon mutants were mapped to 6,703 (89%) of the 7,528 protein coding genes (Table S3.1). The number of mutants for each gene correlated with the number of TA sites (Pearson correlation coefficient r = 0.7) (Fig. 3.5). Only 636 (8%) genes greater than 300 nucleotides long lacked insertions, 10 of which lack TA sites in the central 10-90% region (Table S3.2). The remaining 189 (3%) genes lacking mapped insertions were less than 300 nucleotides (Table S3.3).

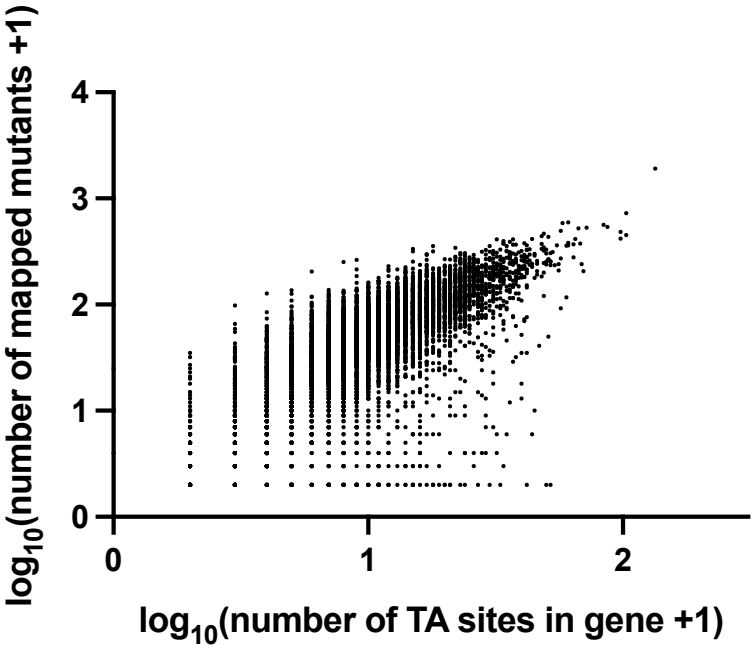


Figure 3.5. The number of TA sites in a gene correlates with the number of mapped barcoded mutants. The Rlv3841 *mariner* transposon library contains 463,550 insertions mapped to the central 10-90% region of protein coding genes. The number of mutants correlates with the number of TA sites present in the gene. The Pearson correlation coefficient was calculated for the 6,703 genes that contain at least one mutant mapped to the central 10-90% region: r = 0.70.

Genes which lack insertions are likely to be essential for the growth of Rlv384. Of the 636 genes (over 300 nucleotides) lacking insertions, mutation of 483 (76%) of these has previously been shown to impair Rlv3841 growth on TY media (Table S3.2) (Wheatley et al., 2020). These genes can be considered part of the core functional genome of Rlv3841, required for growth even in rich media such as TY. Functional analysis of these 636 genes revealed that they primarily have roles in fundamental cellular processes such as nucleic acid replication/repair/synthesis (22.2%), protein/other biosynthesis and modification (11.8%), cell membrane/envelope (9.0%), transport (6.3%), ribosomal proteins (5.2%), amino acid biosynthesis (4.6%), regulation (4.4%) electron transfer (3.5%) and aerobic respiration/TCA cycle (3.1%) (Fig. 3.6). The 153 (24%) of genes not previously found to be growth impaired on TY may lack insertions for one of two reasons: 1) they are required for growth on the media used to generate this insertion library, which was a mixture TY and UMS supplemented with various carbon and nitrogen sources, but not for growth on 100% TY, or 2) no barcodes were reliably mapped to TA sites within these genes. The latter scenario is more probable for genes that possess few TA sites within their central 10-90% region, since this reduces the chance of a transposition event occurring within the gene (Fig. 3.5). For example, *cheW4* (RL3289), *cheW5* (RL4386) and *fliE* (RL0709), which have functional roles in chemotaxis and motility, do not form part of the core functional genome of Rlv3841 but do possess limited TA sites reducing the chance of a transposition event occurring (Table S3.2). Of these 153 genes, 79 contain four or less TA sites.

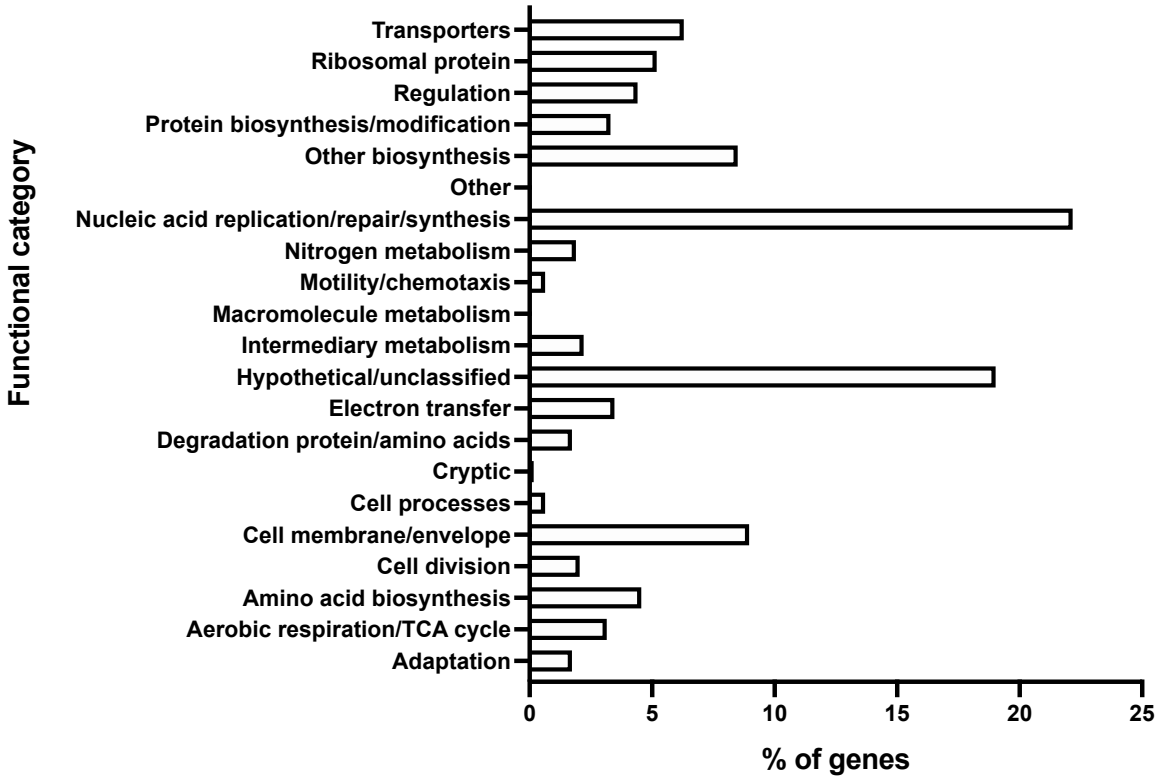


Figure 3.6. Functional classification of 636 genes lacking insertions in the Rlv3841 barcoded transposon mutant pool. Genes lacking insertions less than 300 nucleotides in length are not considered. Functional classification is based on Riley codes (Young et al., 2006). Genes are listed in Table S3.2.

Of the 636 genes lacking insertions, 84% are chromosomally encoded and include those encoding fundamental cellular enzymes essential for growth such as DNA polymerase III (*dnaE3*; RL1723, *holB*; RL2475, *dnaX*; RL0134 and *dn*aN; RL0334), DNA helicase (*ruvAB*; RL3989-90), DNA gyrase (*gyrB*; RL0012, *gyrA*; RL2401 and *gyrB2*; RL2515), adenosine triphosphate (ATP) synthase (*atpIB*; RL0924-25, *aptG2F*; RL0927-28, *atpC*; RL4405, *atpHAGD*; RL4410-07) and tRNA synthetase (*pheST*; RL0269-70) (Table S3.2). Numerous genes encoding factors involved in the transcription and translation of DNA, as well as cell division lacked insertions.

Examples include the ribosomal protein large (*rpl*) and ribosomal protein small (*rps*) genes, which encode the 50s and 30s ribosomal subunits respectively; *rpoD* (RL3402), which encodes the housekeeping σ factor; chromosomal partitioning genes, *parA1B* (RL4736-5); cell division proteins *ftsQAZ1* (RL3300-298), *ftsW* (RL3308), *ftsH* (RL3965), *ftsEX* (RL4330-31) and *ftsY* (RL4543); and the two-component transduction system *chvIG* (RL0036-35). Mutation of either *chvG* or *chvI* in *A. tumefaciens* abolishes growth on complex media (Charles and Nester, 1993) (Table S3.2).

Genes belonging to serval key biosynthesis pathways including LPS (*lpxK*; RL0904, *lpxD*; RL2229, *lpxA*; RL2231, *lpxB*; RL2233, *lpxXL*; RL2812 and *lpxC*; RL3297), EPS (*pssSRMLK*; RL3643-47, *pssI*; RL3649, *pssC*; RL3653, *pssP*; RL3662 and *pssN*; RL3664) and cobalamin (*cobU*; RL2781, cobV; RL2781a, *cobABCD*; RL2824-21, *cobE*; RL2826, *cobO*; RL2829, *cobPW*; RL2832-1 and *cobST*; RL4349-48) lacked insertions, as did multiple genes encoding components of key secretory pathways including the general secretory pathway (*secDF2*; RL0680, *secA*; RL4298 and *secY*; RL1794, *secE*; RL1759), the twin-arginine translocation pathway (*tatAB;* RL2046-47), and the entire Tol system (*tolQ*; RL3973, *tolA*; RL3971 and *tolB*; RL3969) (Krehenbrink and Downie, 2008).

Notable plasmid-encoded genes that lacked insertions include seven of an eight gene cluster, located on pRL11, involved in cobalamin biosynthesis (*cobJKLM*; pRL110628-25 and *cobFGH*; pRL110632-30) (Table S3.2). The eighth gene, *cobI* (pRL110629), only had one insertion which was sequenced nine times during the mapping process implying that this too is essential for Rlv3841 growth (Table S3.1). All six plasmids contain a set of three replication (*rep*) genes apart from pRL7 which contains two

complete sets and a third partial set (Young et al., 2006). All *rep* genes lacked insertions except those on pRL7, *repA* (pRL80001) on pRL8 and *repC* (pRL100003) on pRL10. In the case of pRL80001 and pRL100003 only one barcode was mapped to these genes and these were seen 11 and five times respectively (Table S3.1). The lack of mutants for most *rep* genes provides validation of the BarSeq mutant pool since disruption of *rep* genes will result in loss of the plasmid from the mutant pool.

3.2.4. Rlv3841 colonisation of pea

The more generations of growth a mutant population undergoes during a BarSeq experiment, the more distinguishable their phenotypes become. To determine the optimal timepoint at which to retrieve mutants from the pea rhizosphere and root, colonisation assays were carried out at two-day intervals until bacterial populations reached saturation. A major challenge in using BarSeq to assess the relative fitness of individual mutants is ensuring that all mutants present in the mutant pool are initially well-represented at the time of inoculation, thereby allowing gene fitness values to be assigned based on the behaviour of multiple independent insertion mutants. A loss of diversity at the time of inoculation results in a subsequent loss in statistical power during the analysis of a given gene. The aim was therefore to maximise the total number of inoculated bacterial cells to maintain mutant population diversity, whilst ensuring that the initial inoculum into the plant environment was low enough to allow for substantial growth over which time growth defects can be evaluated. An initial inoculum of 10^5 cfu was chosen, since this allowed for sufficient coverage of genes with a manageable number of plants and avoided an initial 'flooding' of the rhizosphere with bacteria. Inoculation of 75 plants each with 10^5 cfu for example, would provide approximately a 250-fold coverage of the 7,528 genes, even when taking into account

the fact that only 26% of the barcodes present in the Rlv3841 are considered suitable for use in the calculation of gene fitness values (discussed in Chapter 3.2.2).

Following inoculation of seven day-old pea seedling with 10^5 cfu or Rlv3841, the rhizosphere population reached saturation around 5 dpi at 2.88 x10^7 cfu (SEM: 3.84 x10^6), with no significant difference seen between 5 and 7 dpi (*P*=0.9108) (Fig. 3.7). Sampling both the loosely root-associated (vortexed root) and strongly root-associated (ground root after vortexing) populations showed that these too reached saturation by 5 dpi at 8.25 x10^6 (SEM: 1.65 x10^6) and 4.65 x10^6 (SEM: 6.69 x10^5) respectively (Fig. 3.7). To ensure maximal saturation of the mutant populations whilst remaining in the early stages of symbiosis, 7 dpi was chosen as the timepoint at which to harvest barcoded mutant libraries from the rhizosphere and root. During seven days of growth the total number of bacteria present in all fractions increased from 1 x10^5 cfu to 5.27 x10^7 (SEM: 3.52 x10^6), representing approximately nine generations of growth. This selective pressure on the insertion library over nine generations of growth provides a significant period of selection that will allow for the clear resolution of mutant phenotypes.

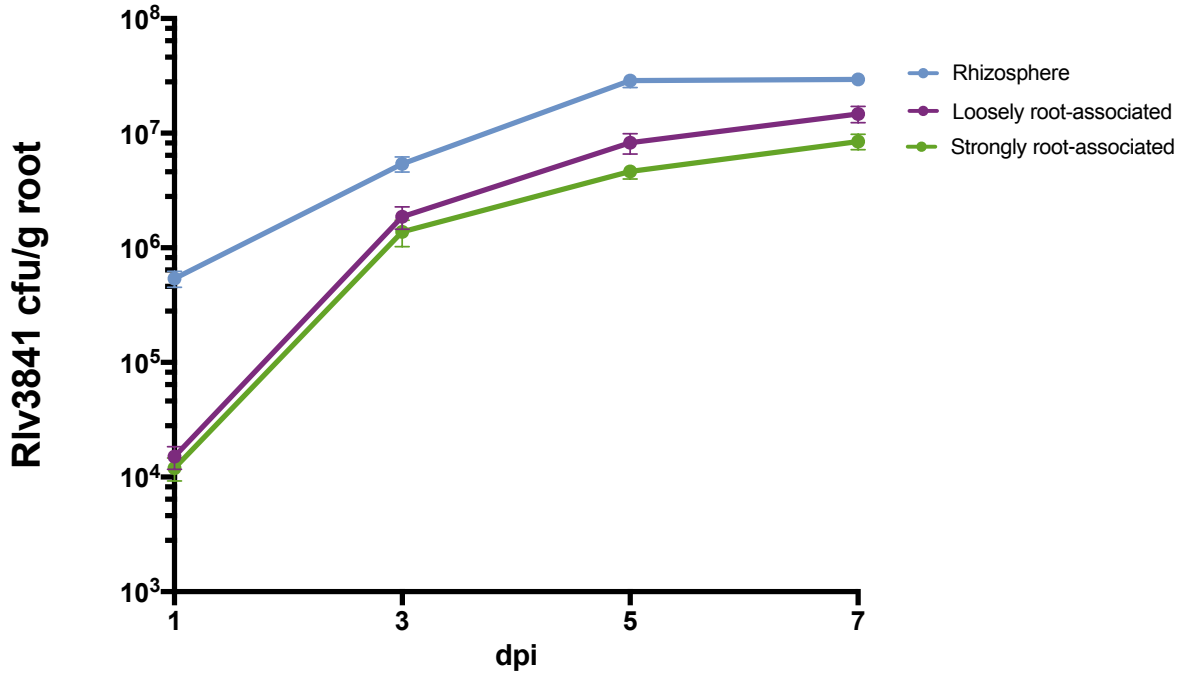


Figure 3.7. Rlv3841 colonisation of the pea rhizosphere and root. 7 day-old pea seedlings were inoculated with 10^5 cfu of Rlv3841. Populations reached saturation in all fractions around 5-dpi, with no significant difference between 5- and 7-dpi. Error bars represent SEM. n=14. One-way ANOVA analysis was performed for each sample followed by Tukey's *post hoc* test. A significant increase in Rlv3841 cfu/g root was seen from 1 to 7 dpi for the rhizosphere ($F = 185.7$, $R^2 = 0.9116$, $P<0.0001$), loosely root-associated ($F = 129.4$, $R^2 = 0.8779$, $P=<0.0001$), and strongly root-associated populations ($F = 123$, $R^2 = 0.8723$, $P=<0.0001$). For all samples a significant increase in cfu/g root was seen on all days apart from 5 to 7 dpi (rhizosphere $P=0.9108$, loosely root-associated $P=0.2461$ and strongly root-associated $P=0.3982$).

3.2.5. BarSeq experimental design to assess gene essentiality during rhizosphere growth and root colonisation

A pool of 10^5 barcoded *mariner* transposon insertion mutants was inoculated onto seven day-old pea seedlings and retrieved from the rhizosphere and root 7 dpi (Fig. 3.8). In total, 96 plants were inoculated providing a total input inoculum of 9.6 x10^6 cfu. Accounting for the fact that data for only 26% of these mutants will be used in the calculation of gene fitness values (due to mapping; see Chapter 3.2.2), this represents approximately an 18-fold coverage of the 140,056 TA sites in Rlv3841 and a 330-fold coverage of the 7,528 genes in Rlv3841. At 7 dpi, transposon mutants were retrieved from the rhizosphere and root (Fig. 3.8). Samples from the 96 plants were pooled prior to gDNA extraction and barcodes were PCR-amplified for sequencing. Input, rhizosphere (vermiculite), loosely root-associated and strongly root-associated libraries were sequenced and analysed following the BarSeq data analysis pipeline (Wetmore et al., 2015). Due to the lower coverage of the loosely and strongly root-associated libraries, data for these two libraries was combined and analysed collectively to identify genes required for root colonisation (7 dpi).

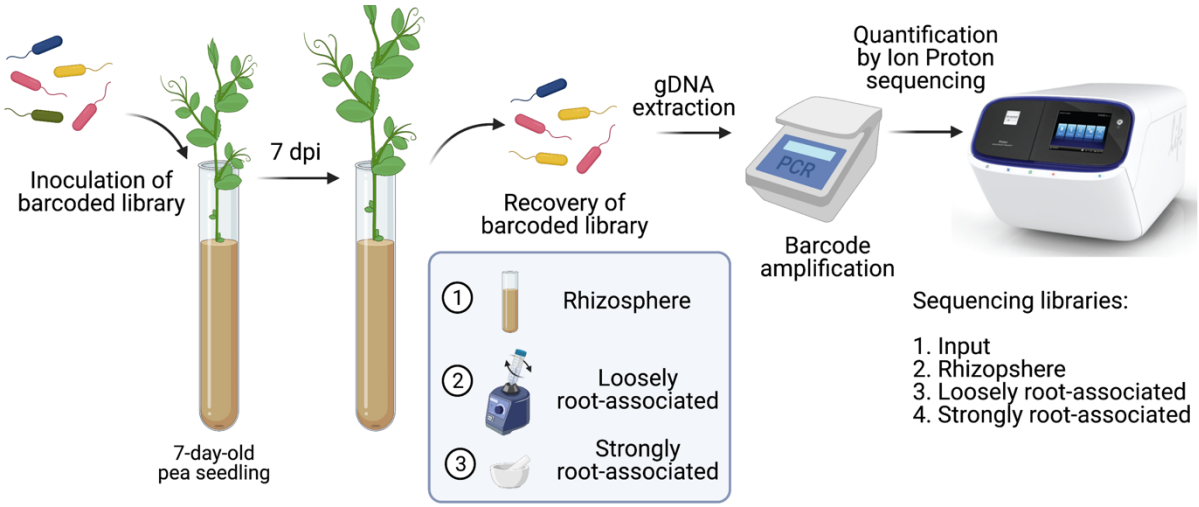


Figure 3.8. BarSeq experimental design. A total of 96 7 day-old pea seedlings were inoculated with 10^5 barcoded *mariner* insertion mutants and retrieved from the rhizosphere and roots at 7 dpi. Samples were pooled across the 96 plants, gDNA extracted, and barcodes amplified by PCR. Input, rhizosphere, loosely root-associated and strongly root-associated samples were sequenced by Ion Proton to quantify barcode frequencies. Data for loosely and strongly root-associated samples was combined and analysed to give an overview of factors important for root colonisation (7 dpi).

3.2.6. BarSeq to identify Rlv3841 genes required for growth in the rhizosphere and colonisation of pea

BarSeq was used to calculate gene fitness values for Rlv3841 during growth in the rhizosphere of its host legume, pea, and colonisation of its roots (described in Chapter 2.5.7). Briefly, gene fitness values are calculated as the $\log_2$ ratio of barcode abundance after growth in each condition divided by the barcode abundance in the initial inoculum at time zero. Fitness values are normalised across the genome so that the typical gene has a neutral fitness value close to zero. Genes with a fitness value less than -2 are considered to have strongly disadvantaged phenotypes, with mutation resulting in at least a 25% reduction in growth relative to the average mutant in the population. Genes with a fitness value between -1 and -2 are considered to be mildly disadvantaged since mutation of that gene results in 50 to 75% less growth relative to the average mutant. Whilst genes with a fitness value between -1 and 1 are considered to be ‘neutral’ and those with a fitness value greater than 1 are considered to be

advantageous meaning these genes, when not mutated, are in some way deleterious for growth in that particular condition.

Fitness values were successfully assigned to 6,019 (80%) of the 7,528 coding genes present in the Rlv3841 genome (Table S3.4), and over 90% of the genes to which barcoded mutants were uniquely mapped in the central 10-90% region. Fitness values were not calculated for 684 of the 6,703 genes containing at least one mapped insertion mutant due to insufficient reads at time zero, suggesting that they were relatively unfit in vitro and therefore account for a small proportion of the mutant library. The percentage of reads containing a barcode varied between samples, ranging from 36% to 66.4% (Table 3.2). Of the reads identified to contain a barcode approximately 24% were successfully linked to a mutant in the mutant database. This is consistent with the fact that 26% of the unique barcodes seen during the initial mapping passed all four criteria required for successful mapping to the genome. Further sequencing of the Rlv3841 insertion library will increase the number of mutants in the mutant database, thereby increasing the number of reads linked to a mutant in subsequent BarSeq experiments. In addition, PCR amplification of the barcode and subsequent purification step can be improved to increase the proportion of reads containing the sample and barcode. Improvement of both factors will increase the number of reads containing a barcode, as well as the fraction of these reads that are subsequently linked to a mutant in the mutant table and used to calculate gene fitness values. Nonetheless, a median of over 220 reads per gene was obtained for each sample (Table 3.2). This is well above the minimum requirement (median 50) considered to be sufficient for calculating gene fitness values by BarSeq (Wetmore et al., 2015).

Table 3.2. BarSeq statistics for Rlv3841 growth in the pea rhizosphere and colonisation of pea roots.

BarSeq statistics	Input	Rhizosphere	Colonisation
% of reads containing a barcode (median)	36.00%	66.39%	59.99%
% of barcodes that map to a mutant in the pool (median)	23.80%	24.71%	24.38%
Median number of reads per gene	375	221	484

3.2.7. Maintenance of genetic diversity *in planta*

Assessment of the mutant populations recovered from each library revealed that there were some bottlenecks with regards to maintaining mutant diversity. This was particularly evident for the rhizosphere population, in which only 22% of the 711,108 mapped mutants were present relative to 97% in the input inoculum (Table 3.3). However, this barcoded transposon library has a particularly high number of uniquely mapped mutants with a median of 60 and mean of 77 unique mutants per gene as opposed to a median of 15 in previous barcoded transposon libraries (Wetmore et al., 2015). Therefore, even when 22% (155,514) of mutants are recovered there is sufficient internal replication to robustly calculate gene fitness values (mean of 21 unique mutants per gene). Further to this, in TnSeq screens mutant libraries with insertion densities above 25% are considered sufficient for accurate determination of gene essentiality (DeJesus and Ioerger, 2013). All recovered libraries were determined to have insertion densities well above this limit, validating the experimental design (Table 3.3).

Table 3.3. Read coverage for genes in input, rhizosphere and root colonised samples.

Library	Unique barcodes sequenced	Unique barcodes matching mutant in pool	Barcodes as % of total mapped mutant pool	Insertion density (% of TA sites with at least 1 read)
Input	2,523,439	691,836 (27.4%)	97%	70%
Rhizosphere	676,113	155,514 (23.0%)	22%	45%
Root colonised	1,198,531	265,510 (22.2%)	37%	56%

3.2.8. Identification of genes required for Rlv3841 growth in the rhizosphere and root colonisation of pea

Of the 6,019 genes for which fitness values were calculated, 84% were assigned neutral fitness values (fitness value between -1 and 1) for pea rhizosphere growth and 87% for root colonisation (Table S3.4). Theoretically, in this system where the Rlv3841 mutant population underwent approximately nine generations of growth, mutants completely impaired in growth such as an auxotroph should have a fitness value close to -9. In reality, very few genes had such an extreme fitness value with only 18 genes assigned a fitness value less than -4 for rhizosphere growth and eight for root colonisation (Fig. 3.9.). This is likely due to the fact that these mutants, previously grown in rich media, initially contain enough of that nutrient in their cell to temporarily support growth in an environment that lacks it.

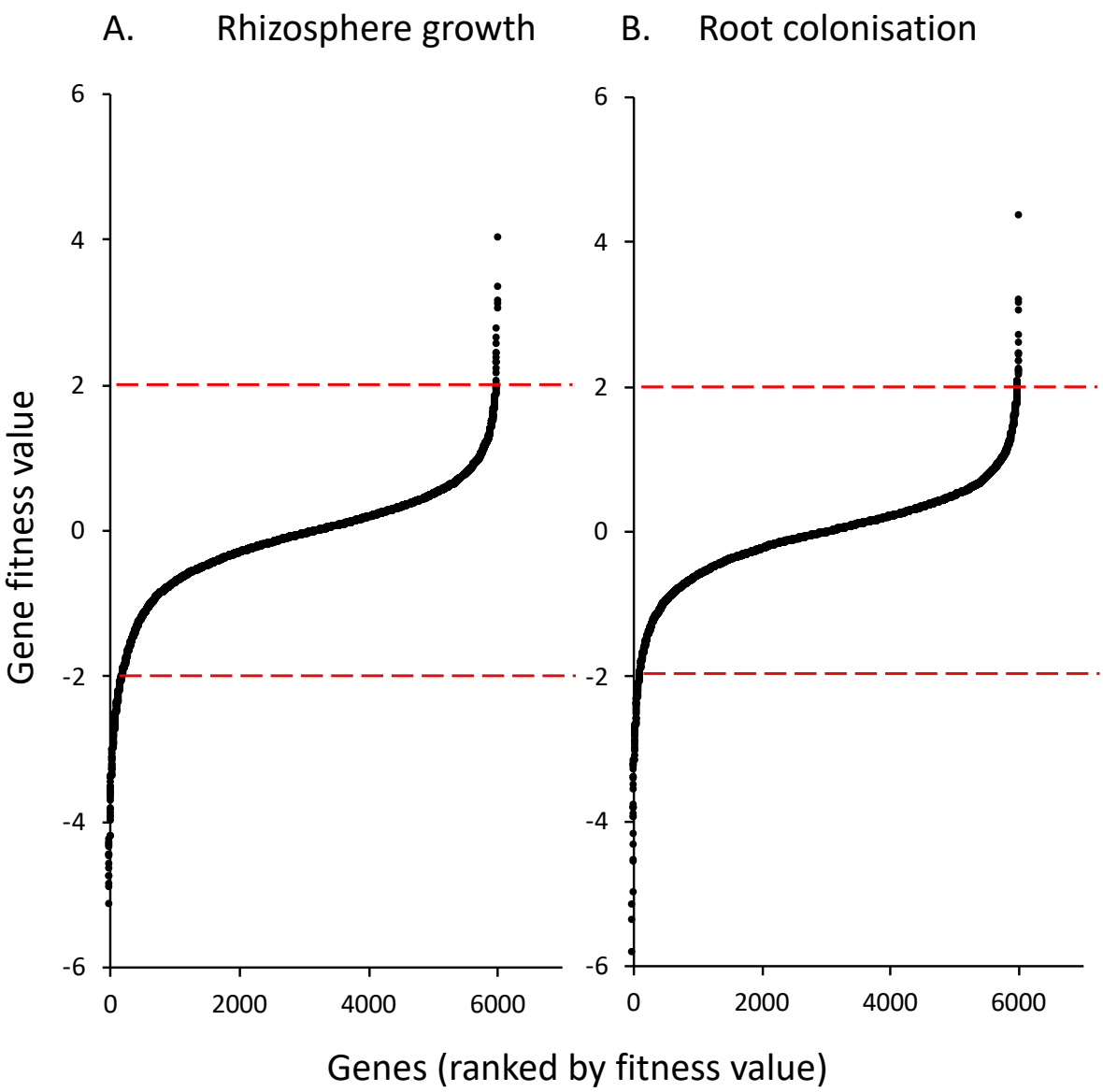


Figure 3.9. Rank ordered gene fitness values for Rlv3841 following pea rhizosphere growth and root colonisation. A. Rhizosphere growth. **B.** Root colonisation. Red dashed lines at -2 and 2 highlight genes with strong fitness defects or strong fitness advantages respectively under that condition.

A total of 667 genes had a fitness value less than -1 for growth in the pea rhizosphere and were therefore considered to be of importance (Table S5.7). The majority of these genes (69%; 461 genes) had mildly disadvantaged phenotypes with fitness values between -1 and -2 (Table 3.4). Mutants in these genes exhibit between 50% and 75% less growth relative to that of a typical mutant in the mutant pool. The other 31% (206 genes) exhibited strongly disadvantaged phenotypes indicating 25% or less growth relative to a typical mutant in transposon pool (fitness value less than -2) (Table 3.4). For colonisation of pea roots, 494 genes were found to contribute (Table S3.6). Again, most of these genes had mildly disadvantaged phenotypes (76%; 376 genes) for root colonisation with only 24% (118 genes) displaying strongly disadvantaged phenotypes (Table 3.4).

Table 3.4. Distribution of Rlv3841 genes fitness values for pea rhizosphere growth and root colonisation.

	Selective condition	
***BarSeq phenotype**	**Pea rhizosphere growth**	**Pea root colonisation**
Strongly disadvantaged	206 (3.4%)	118 (2.0%)
Mildly disadvantaged	461 (7.7%)	376 (6.2%)
Neutral	5,081 (84.4%)	5,265 (87.5%)
Mildly advantaged	253 (4.2%)	235 (3.9%)
Strongly advantaged	18 (0.3%)	25 (0.4%)

*Number of strongly disadvantaged genes (fitness value less than -2), mildly disadvantaged genes (fitness value between -1 and –2), neutral (fitness value between 1 and -1), mildly advantaged genes (fitness value between 1 and 2) and

strongly advantaged genes (fitness value greater than 2) are shown. Percentages shown in brackets are of the total 6,019 genes for which fitness values were assigned.

A previous INSeq screen identified 170 genes as contributing to Rlv3841 growth in the pea rhizosphere (Wheatley et al., 2020). Analysis of gene fitness values calculated for these, revealed that 41 had a fitness value of less than -1 for rhizosphere growth (Table S3.7). Forty had neutral fitness values, though nine of these had a fitness value close to -1 (< -0.7), whilst two genes had a fitness value greater than one. Fitness values were not calculated for the remaining 87 of the 170 genes due to insufficient reads at time zero, suggesting that they were relatively unfit in vitro and therefore account for a small proportion of the mutant library. The same INSeq screen identified 209 genes contributing to Rlv3841 colonisation of pea. Of these, 49 were assigned a fitness value less than -1 for root colonisation (Table S3.8). Sixty-six were assigned neutral fitness values, 13 of which had a fitness value close to -1 (< -0.7), whilst 3 genes were assigned a fitness value greater than one. Fitness values were not calculated for the remaining 91 genes due to insufficient reads at time zero. To distinguish between genes specifically required for rhizosphere growth or root colonisation, from those required for general growth in the media that the mutant library was generated on, BarSeq libraries were grown for several generations prior to storage at -80°C. INSeq libraries however, were not grown for multiple generations and so one reason that fitness values were not calculated for many of the genes identified as growth defective for rhizosphere growth or root colonisation during the initial INSeq screen is likely due to the fact that they were required for general growth in the media that the BarSeq library was generated in.

Comparison of the 667 genes important for growth in the pea rhizosphere (Table S3.5) and 494 genes contributing to pea root colonisation (Table S3.6) identified 231 genes to be common to both categories (Fig. 3.10 and Table S3.9). These 231 genes were classed as rhizosphere-progressive, whereby mutation resulted in reduced rhizosphere growth and subsequent impairment of root colonisation. Genes that fall into the rhizosphere-progressive category may have general roles in rhizosphere growth and root colonisation, or mutation of that gene may impede growth to such an extent that although the gene product has no direct role in root colonisation the mutant remains disadvantaged.

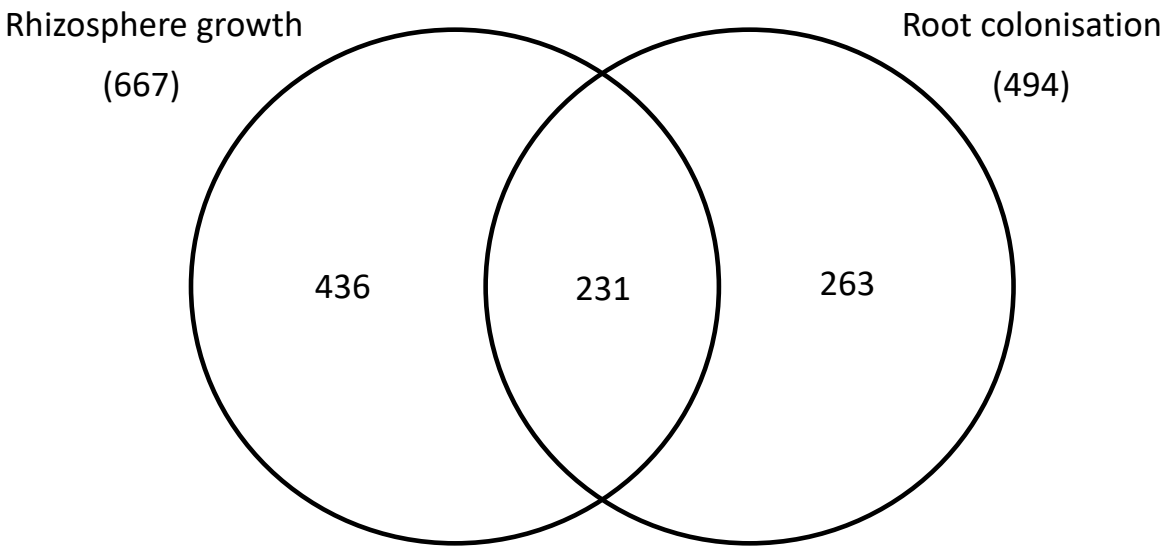


Figure 3.10. Venn diagram of the 667 genes required for pea rhizosphere growth and 494 genes required for pea root colonisation shows 231 genes are required for both (rhizosphere-progressive). Rhizosphere specific, rhizosphere-progressive and colonisation specific genes are listed in Table S3.9.

3.2.9. Genes required for Rlv3841 growth in the pea rhizosphere

The 667 genes identified as rhizosphere-specific and rhizosphere-progressive will be discussed collectively to give an overview of the genetic factors required for optimal growth of Rlv3841 in the pea rhizosphere (Table S3.9). Functional classification of these genes revealed that hypothetical/unclassified proteins accounted for the largest category containing over 35% of genes (Fig. 3.11). Transporters accounted for the second largest group containing 14% of genes, followed by regulation (11%), intermediary metabolism (9%), nucleic acid replication/repair/synthesis (8%) and cell membrane/envelope related functions (7%). The remaining 15 categories individually accounted for less than 3% of the total genes required for growth in the pea rhizosphere.

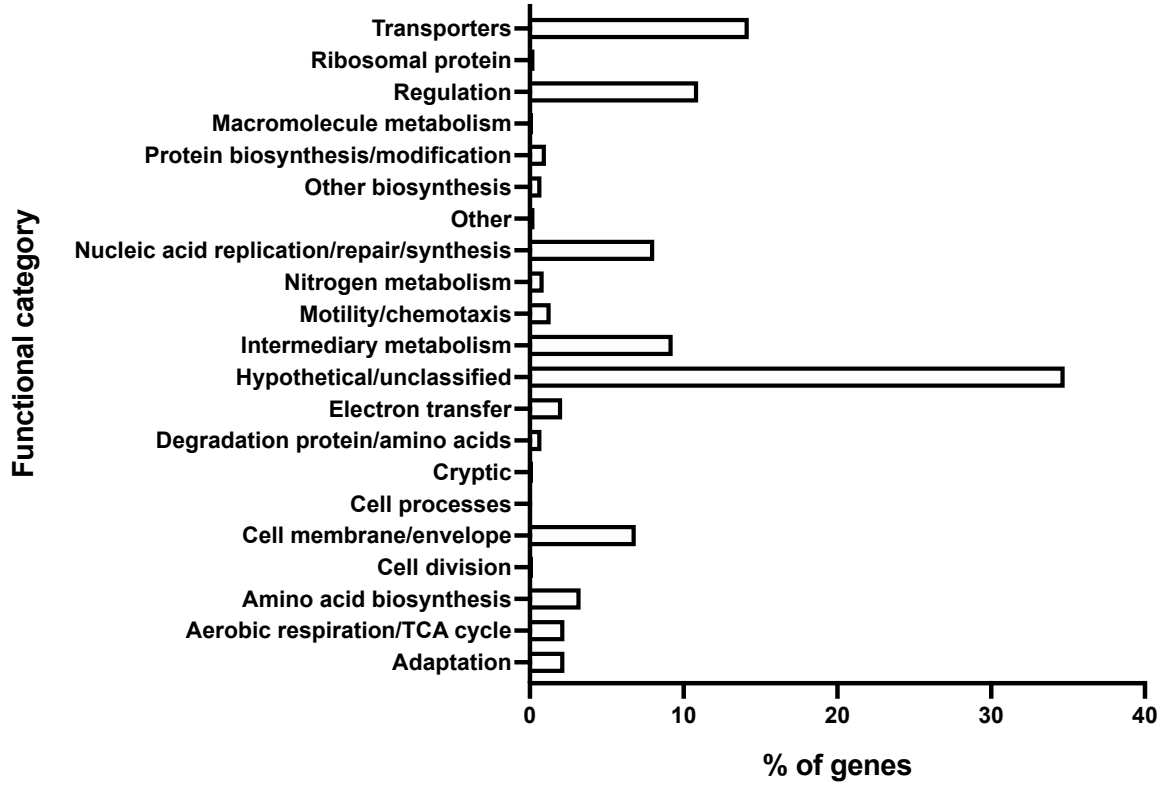


Figure 3.11. Functional classification of 667 genes important for Rlv3841 growth in the pea rhizosphere. All genes with average fitness value less than -1 for growth in the pea rhizosphere. Functional classification based on Riley codes (Young et al., 2006). Genes are listed in Table S3.9.

3.2.9.1. Biosynthesis, transport and metabolism

Amino acid biosynthesis is a clear perquisite for Rlv3841 growth in the pea rhizosphere with mutation of 22 genes involved in the biosynthesis of cysteine (2 genes), glutamate (1), glutamine (1), isoleucine (2), leucine (4), lysine (2), phenylalanine (2), proline (2) and tryptophan (6) resulting in reduced growth (Table 3.5 and Table S3.9). This indicates that whilst these amino acids are present in pea root exudates (Pini et al., 2017), the availability is too low to support the growth of auxotrophic mutants. Biosynthesis of ribonucleotides (purine and pyrimidines) and cofactors (cobalamin/vitamin B_{12}, thiamine/vitamin B_1 and molybdopterin) also contributed to rhizosphere fitness (Table 3.5).

Table 3.5. Genes involved in the biosynthesis of amino acids, ribonucleotides, vitamins, and cofactors that contribute to Rlv3841 growth in the pea rhizosphere. All genes have fitness values less than -1 for rhizosphere growth. These genes are taken from Table S3.9.

Gene ID	Gene Name	Gene Description	Function	Functional Category[a]
pRL80059		putative NifS-like cysteine desulfurase/selenocysteine lyase	Cysteine biosynthesis	Amino acid biosynthesis
RL0900		putative cysQ, an inositol monophosphatase family protein (controls internal levels of PAPS, an intermediate in cysteine biosynth)		
RL4084	*gltD*	putative glutamate synthase [NADPH] small chain	Glutamine biosynthesis	
RL4085	*gltB*	putative glutamate synthase [NADPH] large chain	Glutamate biosynthesis	
RL1803	*ilvD1*	putative dihydroxy-acid dehydratase	Isoleucine biosynthesis	
RL3245	*ilvI*	putative acetolactate synthase isozyme III large subunit		
RL3205	*ilvC*	putative ketol-acid reductoisomerase	Leucine biosynthesis	
RL3513	*leuA1*	putative 2-isopropylmalate synthase		
RL4555		putative 3-isopropylmalate dehydratase		

RL4705	*leuD*	putative 3-isopropylmalate dehydratase small subunit		
pRL100268		putative dihydrodipicolinate synthase	Lysine biosynthesis	
RL1811	*dapA1*	putative dihydrodipicolinate synthase		
RL0139		putative P-protein [includes: chorismate mutase and prephenate dehydrogenase]	Phenylalanine biosynthesis	
RL2847		putative shikimate dehydrogenase		
pRL120517		putative pyrroline-5-carboxylate reductase	Proline biosynthesis	
RL3460	*proC1*	putative pyrroline-5-carboxylate reductase		
RL0003		putative shikimate dehydrogenase	Tryptophan biosynthesis	
RL0020		putative N-(5'-phosphoribosyl)anthranilate		
RL0021	*trpB*	putative tryptophan synthase beta subunit		
RL0022	*trpA*	putative tryptophan synthase alpha subunit		
RL2686	*aroG*	putative phospho-2-dehydro-3-deoxyheptonate aldolase		
RL3521	*trpE*	putative anthralinate synthase		
RL0847	*guaB*	putative inosine-5'-monophosphate dehydrogenase		Nucleic acid replication/repair/

RL1546	*purF*	putative amidophosphoribosyltransferase	Purine ribonucleotide biosynthesis	synthesis
RL1595	*purN*	putative 5'-phosphoribosylglycinamide formyltransferase		
RL1596	*purM*	putative phosphoribosylformylglycinamidine cyclo-ligase		
RL1620	*glyA*	putative serine hydroxymethyltransferase		
RL2606	*purC1*	putative phosphoribosylaminoimidazole-succinocarboxamide synthase (saicar synthetase)		
RL2612	*purL*	putative phosphoribosylformylglycinamidine synthase II		
RL3768	*purA*	putative adenylosuccinate synthetase		
RL4044	*purE*	putative phosphoribosylaminoimidazole carboxylase catalytic subunit		
RL4045	*purK*	putative phosphoribosylaminoimidazole carboxylase ATPase subunit		
RL0572	*pyrD*	putative dihydroorotate dehydrogenase	Pyrimidine ribonucleotide biosynthesis	
RL1738	*pyrC2*	putative dihydroorotase		
RL1739	*pyrB*	putative aspartate carbamoyltransferase		
pRL110441	*thiD*	putative phosphomethylpyrimidine kinase	Thiamine biosynthesis	Nucleic acid replication/repair/ synthesis
pRL110442	*thiE*	putative thiamine-phosphate pyrophosphorylase		

RL2273	*apbE*	putative thiamine biosynthesis protein		
RL4040	*thiE2*	putative thiamine-phosphate pyrophosphorylase		
pRL110127		putative cobalamin synthesis protein (CobW family)	Cobalamin biosynthesis	Other biosynthesis
RL0147		putative molydopterin synthase sulfurylase	Molybdopterin biosynthesis	Protein biosynthesis/ modification
RL0934	*moaB*	putative molybdenum cofactor biosynthesis protein B		
RL1584	*moaE*	putative molybdopterin converting factor subunit E		
RL2711	*moaA2*	putative molybdenum cofactor biosynthesis protein A		

[a]Functional categories are based on riley classification (Young et al., 2006).

A clear requirement for the translocation of substrates across the bacterial membrane, particularly by ABC transporters, suggests that active transport provides a competitive advantage in the rhizosphere (Fig. 3.11). ABC transporters required for rhizosphere growth include carbohydrate uptake transporters (CUT) belonging to the CUT1 and CUT2 subfamilies (Table S3.9). Examples include *aglF* (RL0746), encoding a component of a CUT1 α-glucoside transporter the *S. meliloti* homologue of which is involved in the uptake of sucrose, maltose and trehalose (Willis and Walker, 1999); pRL100440 and pRL100442, encoding a component of a CUT1 transporter and the associated solute binding protein (SBP) respectively; and pRL120285-86, encoding components of a CUT2 transporter possibly involved in the uptake of galactose, arabinose or fucose (Mauchline et al., 2006). Genes encoding components of ABC transporters involved in the uptake of substrates other than carbohydrates also contributed to pea rhizosphere fitness. These include pRL120406-07 (*braEF*), encoding components of a hydrophobic amino acid transporter (HAAT) involved in the transport of branched chain amino acids (Hosie et al., 2002); pRL90302-03, encoding a component of a polar amino acid transporter (PAAT) and the associated SBP respectively; as well as SBPs involved in the uptake of phosphate (RL0542), zinc (RL1049) and peptides (pRL120131). A clear requirement for active transport in the rhizosphere is presumably linked to the need to translocate nutrients for support of bacterial growth. This is consistent with previous transcriptomics analyses of Rlv3841, in which genes encoding components of various active transporters were induced in the pea rhizosphere (Ramachandran et al., 2011).

Other transporters of note contributing to rhizosphere fitness are DctA (RL3424), a C_4-dicarboxylate transporter and BacA (RL3557), a transmembrane transporter protein

involved in the incorporation of very-long-chain fatty acids (VLCFA) into lipid A of LPS. Rlv3841 *dctA* mutants show reduced competitiveness in the pea rhizosphere (Ramachandran et al., 2011) suggesting that despite the requirement for active uptake of carbohydrates, organic acids serve as an important carbon source in this environment. BacA is essential for *R. leguminosarum* bacteroid development during symbiosis with pea, where it acts to protect rhizobia from the antimicrobial activity of plant-derived nodule-specific cysteine-rich peptides (Karunakaran et al., 2010; Haag et al., 2013). Identification of BacA as contributing to rhizosphere fitness suggests it may play a role during earlier stages of symbiosis. This finding is in agreement with a previous INSeq screen, which also resulted in the identification of BacA as important for Rlv3841 growth in the pea rhizosphere (Wheatley et al., 2020).

After hypothetical proteins and transporters, genes involved in intermediary metabolism formed the third largest functional group contributing to rhizosphere growth and consisted of 62 genes (Fig. 3.11 and Table S3.9). This high requirement likely reflects the significant adaptation bacteria must undergo during transition from growth in rich media to the relatively nutrient deplete conditions of the rhizosphere. Pea root exudates contain a variety of carbon sources capable of supporting bacterial growth (Pini et al., 2017). The ability of rhizobia to catabolise several of these, including erythritol, myo-inositol, homoserine and arabinose, is required for competitive nodulation of pea by its symbiotic partners (Fry et al., 2001; Yost et al., 2006; Vanderlinde et al., 2014; Garcia-Fraile et al., 2015). The requirement of erythritol catabolism for competitive nodulation of pea was demonstrated in *R. leguminosarum* VF39SM, which is closely related to Rlv3841. Both genomes encode three operons involved in erythritol uptake and catabolism: (1) *eryEFG*, an erythritol ABC-transporter

operon; (2) *eryABCD,* an erythritol catabolic operon which is induced in the presence of erythritol and negatively regulated by the repressor EryD; and (3) an operon encoding a DeoR family transcriptional regulator (*eryR*), a putative triosephosphate isomerase (*tpiA2*) and a putative ribose-5-phosphate isomerase (*rpiB2*) (Yost et al., 2006). Mutation of *eryA*, *eryC*, *eryR*, *eryE*, *eryF*, *tpiA2* and *rpiB2* inhibits *R. leguminosarum* VF39SM growth on erythritol as a sole carbon source, whilst addition of erythritol to minimal media inhibits the growth of *eryC*, *eryR*, *tpiA2* and *rpiB2* mutants but not *eryA*, *eryD*, *eryE* or *eryF* mutants. This growth inhibition is due to a build-up of toxic phosphorylated intermediates in the erythritol catabolism pathway which is dependent on erythritol kinase (*eryA*) encoded by the first gene in the pathway. Analysis of gene fitness values for the three operons involved in erythritol uptake and catabolism in Rlv3841 revealed that as expected *eryA* has a neutral phenotype for growth in the pea rhizosphere, whilst *eryB* (pRL120205), *eryC* (pRL120206), *eryR* (pRL120208), *tpiA2* (pRL120209) and *rpiB2* (pRL120210) mutants showed a clear impairment in rhizosphere growth and subsequent root colonisation (Table 3.6). This suggests that the ability to catabolise erythritol provides a competitive advantage in the pea rhizosphere. However, the erythritol ABC-transporter operon (*eryEFG*; pRL120200-02) was not important for rhizosphere growth, perhaps indicating that the ability to catabolise erythritol provides a competitive advantage by preventing erythritol toxicity or the presence of an alternative erythritol transport system in Rlv3841.

Table 3.6. Rlv3841 gene fitness values for erythritol uptake and catabolism operons following growth in the rhizosphere and colonisation of pea. These genes are taken from Table S3.4.

Gene ID	Gene Name	Gene Description	Gene Fitness Value	
			Rhizosphere	Colonisation
pRL120200	*eryG*	putative SBP of ABC transporter CUT2 erythritol	+0.59	+0.55
pRL120201	*eryF*	putative permease component of ABC transporter CUT2 erythritol transporter	-0.03	+0.19
pRL120202	*eryE*	putative ATP-binding:ATP-binding (ABC:ABC) component of ABC transporter CUT2, erythritol	+0.81	+0.51
pRL120204	*eryA*	putative erythritol kinase	-0.12	-0.01
pRL120205	*eryB*	putative erythritol phosphate dehydrogenase	-1.42	-1.64
pRL120206	*eryC*	putative erythrulose 4-phosphate dehydrogenase	-1.05	-1.19
pRL120207	*eryD*	putative DeoR family transcriptional regulator (repressor)	-0.11	-0.31
pRL120208	*eryR*	putative DeoR family transcriptional regulator (repressor)	-2.68	-0.73
pRL120209	*tpiA2*	putative triosephosphate isomerase	-0.91	-1.50
pRL120210	*rpiB2*	putative ribose-5-phosphate isomerase B	-2.52	-2.99

An arabinose dehydrogenase encoded by *araE* (RL3613) was identified as rhizosphere-progressive (Table S3.9). This is consistent with previous findings that Rlv3841 *araE* mutants are unable to grow on arabinose, display reduced growth in the pea rhizosphere during competition with wild-type Rlv3841 and demonstrate impaired competitiveness for nodule infection (Garcia-Fraile et al., 2015). Rhamnose catabolism was also of importance as demonstrated by the requirement of *rhaD* (pRL110415), encoding an oxidoreductase essential for growth on rhamnose (Richardson et al., 2004). The ability to catabolise rhamnose is required for competitive nodulation of clover by *R. leguminosarum* bv. *trifolii* and a Rlv3841 INSeq screen led to the identification of *rhaD* as important for nodulation of pea (Oresnik et al., 1998; Wheatley et al., 2020). Collectively, this implies the rhamnose catabolism is also important for Rlv3841 competitive nodulation of pea most likely due to increased rhizosphere fitness.

Mutation of five key enzymes involved in the major metabolic pathways glycolysis/gluconeogenesis led to reduced rhizosphere fitness. These enzymes are a pyrophosphate-fructose-6-phosphate 1-phosphotransferase (*pfp*, RL3322), a putative triosephosphate isomerase (*tpiA*; RL2513), a putative phosphoglycerate kinase (*pgk*, RL4011), a 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase (*gpmA*, RL0179) and an enolase (*eno*, RL2239) (Table S3.9). Transcriptome analyses have demonstrated clear induction of glycolysis/gluconeogenesis in Rlv3841 during growth in plant rhizospheres (Ramachandran et al., 2011).

3.2.9.2. Adaptation, regulation and cellular factors

The plant rhizosphere provides a varied environment for bacteria to navigate, with variation in nutrient availability, pH, osmotic conditions and oxidative stress often at the microscale level (Hinsinger et al., 2003; Santoyo et al., 2017). This stringent response (SR) protein RelA (RL1506) was important for rhizosphere fitness and root colonisation (Table S3.9). RelA has previously been found to be important during all stages of Rlv3841 symbiosis with pea (Wheatley et al., 2020), from rhizosphere growth to N_2-fixing bacteroids, and *R. etli relA* mutants display altered bacteroid morphology combined with reduced N_2-fixing activity (Moris et al., 2005). This is consistent with its fundamental role in cellular physiology, catalysing the synthesis of hyperphosphorylated guanosine derivatives ((p)ppGpp) which mediate the SR. The SR enables bacteria to rapidly reprogram transcription in response to changes in nutrient availability, with (p)ppGpp directly binding to RNA polymerase to downregulate genes involved in cell division and upregulate those involved in stress responses (Boutte and Crosson, 2013).

Maintenance of osmotic balance within the cell is essential for bacterial survival under any condition. Accordingly, mutation of *ndvA* (RL4640) which is responsible for the biosynthesis of cyclic ß-1-2-glucan resulted in impaired rhizosphere growth. Cyclic ß-1-2-glucans are cell surface carbohydrates found almost exclusively in the *Rhizobiaceae* family where they play a major role in osmotic adaptation (Dylan et al., 1990; Breedveld and Miller, 1994). Transcriptional analyses in Rlv3841 have previously shown *ndvA* to be upregulated in various plant rhizospheres including pea (Ramachandran et al., 2011) and *ndvA* mutants demonstrate impaired growth ability

in hypoosmotic media (Breedveld and Miller, 1994). Osmotically-induced *osmC* (RL4226) was also important for rhizosphere fitness.

Adaption to oxidative stress was important in the pea rhizosphere, with the requirement of pRL100335, encoding a glutathione S-transferase involved in the quenching of reactive molecules (Kumar and Trivedi, 2018) and RL2003, encoding a putative peroxidase. Plants are known to release reactive oxygen species (ROS) such as hydrogen peroxide, superoxide and phytoalexins in response to biotic and abiotic stresses (Nanda et al., 2010). Copper tolerance promoted rhizosphere fitness with the requirement of *cusF* (RL2432), encoding a copper binding protein; *copA* (RL2434), encoding a putative copper resistance protein; and *cutC* (RL0514), encoding a putative copper homeostasis protein. Copper resistance mechanisms are widespread in plant-associated bacteria (Rainey, 1999). It is likely that these genes play an important role in copper homeostasis within the cell as whilst copper is a key micronutrient, at high concentrations it becomes toxic. A putative cold-shock protein, RL2594 (*deaD*), also promoted growth in the pea rhizosphere.

During free-living growth of rhizobia electron transfer from the cytochrome bc_1 complex to cytochrome aa_3 typically occurs via the transmembrane cytochrome *c* CycM (Delgado et al., 1998). Accordingly, *petA* (RL3486) encoding a component of the cytochrome bc_1 complex, and *cycM* (RL0141) had rhizosphere-progressive phenotypes (Table 3.7). Genes encoding the remaining components of the cytochrome bc_1 complex did not show a clear requirement for rhizosphere growth, but *petC* (RL3484) was important for root colonisation (Table 3.7). Several systems for cytochrome *c* maturation have been characterised in bacteria (Yurgel et al., 2007).

Those involved in the assembly of *c*-type cytochromes in the periplasm include: (1) the cytochrome *c* biogenesis operon *cycHJKL* (RL1436-39), most likely required for the covalent attachment of heme to cytochrome *c* apoproteins (Delgado et al., 1995); (2) an ABC transporter encoded by *ccmAB* (RL4537-38), required for export of heme-binding proteins to the periplasm (Aguilar and Soberón, 1996); and (3) *cycZ*, (RL4539) a putative heme exporter protein. Deletion of the *cycHJKL* operon in *Bradyrhizobium japonicum* results in an inability to synthesise *c*-type cytochromes, as does mutation of *cycH* and *cycK* in *R. leguminosarum*, and mutation of *ccmA*, *ccmB* and *cycZ* in *S. meliolti* (Delgado et al., 1995; Ritz et al., 1995; Yurgel et al., 2007). With the exception of *cycZ*, all of the aforementioned genes were important for growth in the pea rhizosphere and many were classed as rhizosphere-progressive (Table 3.7). This suggests that novel cytochrome *c* biosynthesis and export to the periplasm is important for rhizosphere fitness and subsequent root colonisation. However, *R. leguminosarum cyc* mutants also show impairment in high-affinity iron acquisition due to an inability to make or export siderophores (Yeoman et al., 1997). Synthesis and export of siderophores is essential to obtain sufficient amounts of iron to support bacterial growth and ability to utilise siderophores is thought to affect inter-strain competition in the rhizosphere (Guerinot, 1991). Therefore, requirement of the *cycHJKL* operon for rhizosphere growth and root colonisation may reflect a growth impairment due to iron limitation, a requirement for novel cytochrome *c* biosynthesis or both.

Table 3.7. Rlv3841 gene fitness values for electron transfer and cytochrome *c* biogenesis operons during growth in the rhizosphere and colonisation of pea. These genes are taken from Table S3.4.

Gene ID	Gene Name	Gene Description	Gene Fitness Value	
			Rhizosphere	Colonisation
RL0141	*cycM*	putative cytochrome c	-1.13	-1.15
RL1436	*cycH*	putative cytochrome c-type biogenesis protein	-1.11	-1.27
RL1437	*cycJ*	putative cytochrome c-type biogenesis protein	-2.19	-1.79
RL1438	*cycK*	putative cytochrome c-type biogenesis protein	-1.22	-0.49
RL1439	*cycL*	putative cytochrome c-type biogenesis protein	-1.54	-2.95
RL3484	*petC*	putative cytochrome c1 precursor	-0.56	-1.21
RL3485	*petB*	putative cytochrome b	-0.71	-0.72
RL3486	*petA*	putative ubiquinol-cytochrome c reductase iron-sulfur subunit	-2.42	-3.10
RL4537	*ccmA*	putative ATP-binding component of ABC transporter Export cytochrome c biogenesis	-2.62	-0.07
RL4538	*ccmB*	putative permease component of ABC transporter Export cytochrome c binding export protein	-3.37	-2.35
RL4539	*cycZ*	putative permease component of ABC transporter Export heme exporter protein c (cytochrome c-type biogenesis protein)	-0.70	-0.85

3.2.10. Rlv3841 genes required for colonisation of pea roots

Colonisation of the rhizoplane is a perquisite for many plant-bacteria interactions and allows bacteria to benefit from root exudates. A total of 494 genes contributed to root colonisation (Fig. 3.10 and Table S3.6), but only 263 genes contributed specifically (Table S3.9). These 263 genes will be discussed to give an overview of factors important for colonisation of pea roots. Functional classification of the 263 genes involved in root colonisation revealed that as with rhizosphere growth hypothetical proteins accounted for the largest functional group containing 37% of genes (Fig. 3 12). Similarly to rhizosphere growth, this was followed by transporters (10%) and intermediary metabolism (9%). However, genes involved in chemotaxis and motility accounted for the fourth largest category (8%) followed jointly by regulation, cell membrane/envelop and nucleic acid replication/repair/synthesis (all 7%).

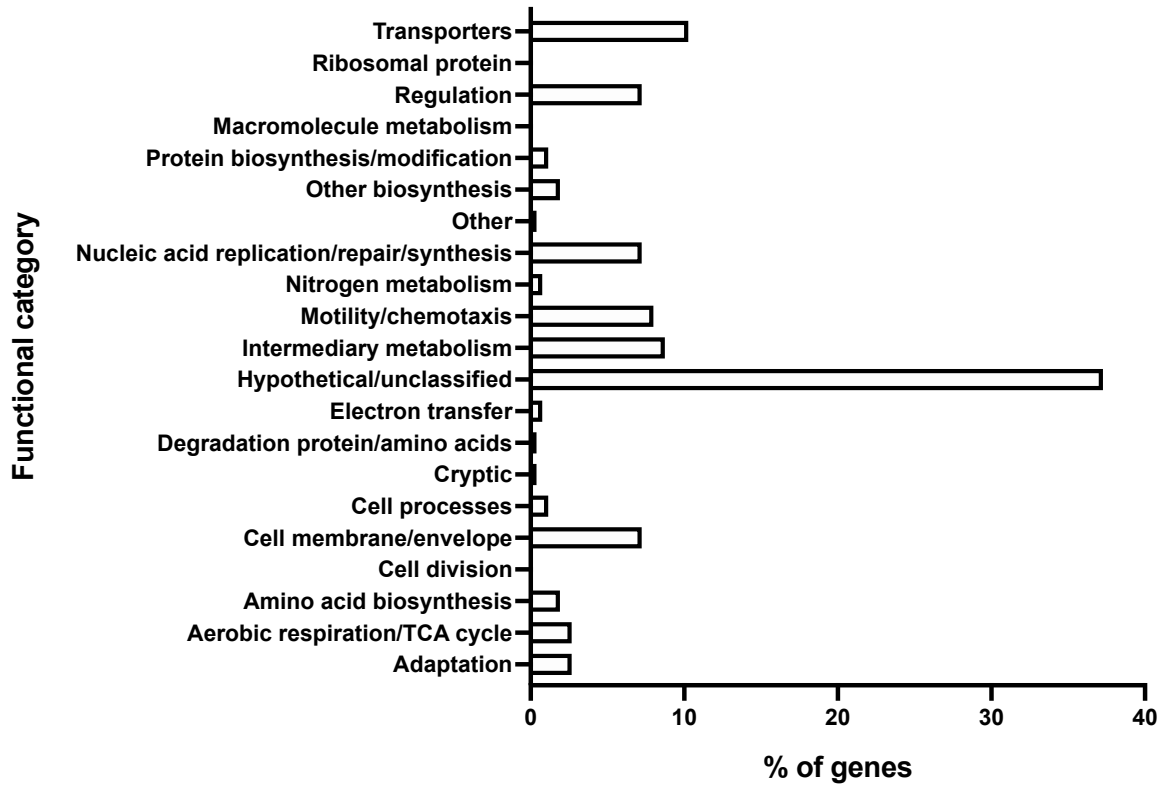


Figure 3.12. Functional classification of 263 genes important for Rlv3841 colonisation of pea roots. All genes with average fitness value less than -1. Functional classification based on Riley codes (Young et al., 2006). Genes are listed in Table S3.9.

3.2.10.1. Motility and chemotaxis

Plant photosynthates secreted into the rhizosphere form chemical gradients that are perceived by bacteria, which respond by moving along these gradients towards the root. The Rlv3841 genome contains 92 genes predicted to have a role in chemotaxis and flagella-mediated motility (Fig 3.13 and Table S3.10). These include genes encoding chemotaxis proteins (*che*, 16 genes), methyl-accepting chemotaxis receptors (*mcp*, (24 genes) and *mcr* (3 genes)), flagella biosynthesis and motor function proteins (*fla* (9 genes), and *flg* (16 genes), *flh* (2 genes), *fli* (12 genes), *mot* (5 genes)). A previous INSeq screen for Rlv3841 colonisation of pea found that all of the aforementioned genes were not important for rhizosphere growth or root colonisation (Wheatley et al., 2020). In contrast, analysis of gene fitness values obtained in these BarSeq screens identified 21 genes involved in chemotaxis and motility as contributing specifically to pea root colonisation (Table 3.8). A further nine with putative roles in chemotaxis and motility contributed to rhizosphere growth, six of which were rhizosphere-progressive (Table 3.8 and Table S3.10). Though it is probable that chemotaxis towards root exudates promotes rhizosphere fitness, chemotaxis towards the rhizoplane is fundamental for root colonisation likely explaining this higher requirement for genes involved in chemotaxis and motility. Requirement for chemotaxis and flagella-mediated motility during bacterial root colonisation is well-established in the literature for various soil bacteria (de Weert et al., 2002; Miller et al.,

2007; Allard-Massicotte et al., 2016; Liu et al., 2020). In Rlv3841 the *che1* cluster is the major pathway controlling chemotaxis with mutants impaired in competitive nodulation of pea, implying that the *che1* operon processes the environmental cues that trigger chemotaxis of Rlv3841 towards pea roots, thereby facilitating root colonisation (Miller et al., 2007). The same study demonstrated that although the *che2* cluster contributes to swimming bias, it only has a minor effect on chemotaxis. In accordance, all genes belonging to the *che2* cluster were found to have gene fitness values greater than -1 for both rhizosphere growth and root colonisation (Fig. 3.13 and Table S3.10).

Table 3.8. Genes with predicted roles in chemotaxis and motility important for pea root colonisation and/or rhizosphere growth. 3 genes rhizosphere specific, 6 genes rhizosphere-progressive and 21 genes colonisation specific. Genes are taken from Table S3.9.

Gene ID	Gene Name	Gene Description	Gene Fitness Value	
			Rhizosphere	Colonisation
Rhizosphere growth				
RL0218		putative pilus assembly protein	-1.39	-0.49
RL0564	*mcpE*	putative methyl-accepting chemotaxis protein	-1.32	-0.65
RL0719	*flaB*	putative flagellin B	-2.02	-0.88
Rhizosphere-progressive				
RL0685	*icpA*	putative chemoreceptor protein, transducer-like protein	-1.26	-2.44
RL0686	*cheX1/cheS*	putative chemotaxis related CheX protein	-3.39	-3.82
RL0693	*cheD1*	putative chemotaxis related protein	-1.89	-2.15
RL0698		putative flagellin related gene	-1.81	-1.03
RL0711	*flgA*	putative flagellar protein precursor FlgA	-2.06	-2.19
RL0731	*flaF*	flagellar biosynthesis regulator FlaF	-1.59	-1.11
Root colonisation				
RL0213	*cpaB*	putative pilus assembly protein	+1.42	-1.42
RL0689	*cheW1*	putative chemotaxis protein CheW family	-0.57	-2.54
RL0690	*cheR1*	putative chemotaxis protein methyltransferase	-0.48	-1.78

RL0695	*fliF*	putative flagellar M-ring protein	+0.16	-1.64
RL0699	*flhB*	putative flagellar biosynthetic protein	-0.05	-1.15
RL0700	*fliG*	putative flagellar motor switch protein	+0.39	-1.10
RL0702	*fliM*	putative flagellar motor switch protein	+0.63	-1.26
RL0704	*flgF*	putative flagella-associated protein FlgF	-0.34	-3.57
RL0705	*fliI*	putative flagellum-specific ATP synthase	+0.18	-1.21
RL0707	*flgB*	putative flagellar basal-body rod protein FlgB	+0.90	-1.53
RL0708	*flgC*	putative flagellar basal-body (proximal) rod protein FlgC	+0.05	-1.74
RL0710	*flgG*	putative flagellar protein FlgG	+0.03	-1.00
RL0714	*flgH*	putative flagellar basal body L-ring protein precursor FlgH	-0.51	-2.17
RL0720	*flaC*	putative flagellin A	-0.91	-1.16
RL0724	*motC*	putative motility protein precursor MotC	-0.62	-2.68
RL0728	*flgE1*	putative flagellar hook protein FlgE	-0.02	-1.53
RL0729	*flgK1*	putative flagellar hook-associated protein FlgK	+0.36	-1.18
RL0730	*flgL1*	putative flagellar hook-associated protein FlgK	0.00	-1.25
RL0735	*flhA*	putative flagellar biosynthesis protein FlhA	0.92	-1.51
RL4292		putative pilus component protein	-0.61	-1.19
RL4579		putative pilin protein	-0.46	-1.88

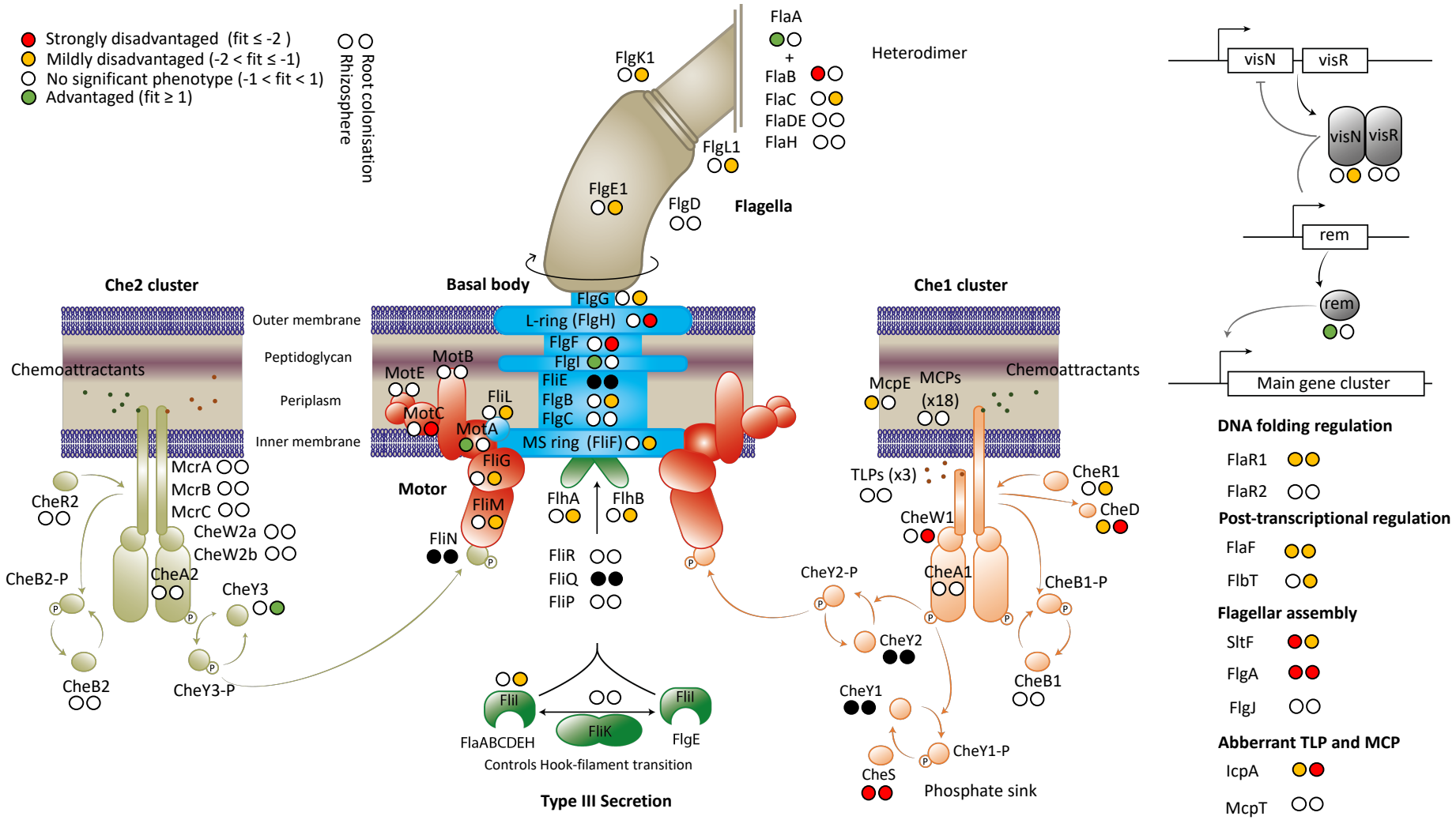

Strongly disadvantaged (fit ≤ -2)
Mildly disadvantaged (-2 < fit ≤ -1)
No significant phenotype (-1 < fit < 1)
Advantaged (fit ≥ 1)
Root colonisation
Rhizosphere
FlaA
+
FlaB
FlaC
FlaDE
FlaH
Heterodimer
FlgK1
FlgL1
FlgE1
FlgD
Flagella
Che2 cluster
Basal body
Che1 cluster
Outer membrane
Peptidoglycan
Periplasm
Inner membrane
Chemoattractants
FlgG
L-ring (FlgH)
FlgF
FlgI
FliE
FlgB
FlgC
MS ring (FliF)
MotB
MotE
MotC
MotA
FliL
FliG
FliM
FliN
Motor
FlhA
FlhB
FliR
FliQ
FliP
McpE
MCPs
(x18)
TLPs (x3)
McrA
McrB
McrC
CheR2
CheW2a
CheW2b
CheA2
CheB2-P
CheY3
CheB2
CheY3-P
CheR1
CheD
CheW1
CheA1
CheB1-P
CheY2-P
CheY2
CheB1
CheY1
CheY1-P
CheS
Phosphate sink
FliI
FliK
FlgE
FlaABCDEH
Controls Hook-filament transition
Type III Secretion
visN
visR
rem
Main gene cluster
DNA folding regulation
FlaR1
FlaR2
Post-transcriptional regulation
FlaF
FlbT
Flagellar assembly
SltF
FlgA
FlgJ
Abberrant TLP and MCP
IcpA
McpT

Figure 3.13. Classification of genes encoding chemotaxis and flagella mediated motility in rhizosphere and root colonised libraries. Genes with a fitness value less than -2, where mutation has a strong disadvantage are shown in red; genes with a fitness value between -1 and -2, where mutation provides a mild disadvantage are shown in yellow; genes with a fitness value greater than -1 but less than 1, where mutation results in a neutral phenotype are shown in white; genes with a fitness value greater than 1, where mutation provides an advantage are shown in green; and genes for which fitness values were not calculated are shown in black. Genes are listed in Table S3.10. This figure was adapted from reference (Wheatley et al., 2020).

3.2.10.2. Cell surface components

Nineteen genes with predicted functions in cell membrane/envelope structure were identified as important for root colonisation (Table S3.9). LPSs constitutes a major component of the gram-negative outer membrane and several genes involved in LPS biosynthesis contributed to pea root colonisation including RL1626 (*rmlC*), encoding a dTDP-rhamnose-3,5-epimerase; RL0825 (*gmd*), encoding a GDP-d-mannose dehydratase; and RL3304 (*galE*), encoding a putative UDP-glucose 4-epimerase. RmlC is required for dTDP-rhanmose biosynthesis, which has been shown to promote root colonisation in several bacterial genera including *Rhizobium*, *Azospirillum* and *Herbaspirillum* (Jofré et al., 2004; Balsanelli et al., 2010; Mitra et al., 2016). *R. tropici gmd* mutants have a truncated LPS core and are impaired in maize root colonisation (Ormeño-Orrillo et al., 2008), whilst *B. japonicum galE* mutants showed altered LPS profiles and were impaired in nodulation (Chang et al., 2008).

The outer membrane porins RopA2 (RL1499) and RopB (RL1589) contributed to pea root colonisation. Rlv3841 RopA2 and RopB share 77.48% and 98.6% identity to *R. leguminosarum* bv. *viciae* RCAM1026 RopA and RopB respectively, which form amyloid fibrils (Kosolapova et al., 2019). Although the precise function of these proteins remain elusive, they are predicted to play a role in host plant colonisation (Kosolapova et al., 2019). This is consistent with the finding that bacterial amyloid fibrils in various bacteria, including CsgA in *E.coli*, form extracellular fimbriae called curli with direct roles as adhesins and biofilm constituents (Chapman et al., 2002; Blanco et al., 2012). A previous INSeq screen also led to the identification of *ropA2* as important for 1-h attachment to pea roots at pH 6.5-7.5 (Parsons, 2019), providing support for RopA2 functioning as an adhesin in Rlv3841.

Other genes affecting cell surface structure important for root colonisation include pRL90136, encoding a putative glucosyl transferase; pRL120461, encoding a capsule polysaccharide export protein; pRL90144, encoding an exopolysaccharide biosynthesis-related protein; and RL4381, encoding a putative autotransporter secretion outer membrane protein TamA. Whilst this gene has not been characterised in *R. leguminosarum*, in *E.coli* TamA has a role in the secretion of autotransporters which are universally used by gram-negative bacteria to export virulence and colonisation factors (van Ulsen et al., 2018).

Several genes encoding factors affecting the rhizobial cell surface have previously been identified to play an important role in attachment to pea root hairs and the subsequent development of a bacterial aggregates termed root hair caps (Smit et al., 1987). These include *pssA* (RL3752), which catalyses the first step in acidic EPS biosynthesis; *celA* (RL1646), required for the synthesis of cellulose fibrils; and *gmsA*, encoding the UPP glucomannan (RL1661). Analysis of fitness values for these genes revealed that none were important for root colonisation 7 dpi (Table 3.9). This is not unexpected since although *pssA* is important for 1-h attachment to pea roots (Parsons, 2019), mutants are still able to colonise the root epidermis (Williams et al., 2008). Similarly, whilst *celA* mutants are impaired in root cap formation they can individually attach to root hairs, show no impairment in root-attachment (1 hpi) and form biofilms indistinguishable from those formed by the wild-type Rlv3841 in vitro (Williams et al., 2008; Parsons, 2019). Rlv3841 *gmsA* mutants also display reduced attachment to pea roots (1 hpi) but only at pH 6.5 and 7.0, not at pH 7.5 (Parsons, 2019). The pH of this experimental system was measured to be neutral (pH 7.0), however, *gmsA* mutants form normal biofilms in vitro and a previous INSeq screen found no role for *gmsA*

during colonisation of pea roots (5 dpi) (Wheatley et al., 2020). Collectively, these results suggest the existence of alternative mechanisms for rhizobial attachment to root hairs (1-2 hpi) and long-term colonisation over the entire root surface (7 dpi). It is possible however that these factors play a more prominent role in soil environments where interspecies competition is present.

Raps are secreted Ca^{2+}-binding lectins, containing at least one *Rhizobium*-adhering or cadherin-like domain, that promote root attachment and biofilm development (Ausmees et al., 2001; Frederix et al., 2014). The Rlv3841 genome encodes five predicted Raps: RapA1 (RL3660), RapA2 (pRL100451), RapB (RL3911), RapC (RL3074) and a putative RapD (RL2702) (Abdian et al., 2013). Three of these: *rapA2*, *rapB* and *rapC,* are under the control of the transcriptional repressor PraR (RL0390). Mutation of *praR* results in increased expression of these *rap* genes, which in turn leads to enhanced root attachment and competitiveness for nodulation relative to the wild-type strain (Frederix et al., 2014). Gene fitness values were calculated for RapA2, RapB and RapD, with all showing neutral phenotypes for root colonisation (Table 3.9). These results are consistent with previous findings that showed mutation of *rapA2*, *rapB* or *rapC* in the Rlv3841 *praR* mutant background had no effect on the enhanced root attachment phenotype and that abolishment of this phenotype required the mutation of at least two *rap* genes suggesting a level of 'redundancy' among the *rap* proteins (Frederix et al., 2014).

Table 3.9. Root colonisation fitness values for genes previously implicated in Rlv3841 attachment to pea roots. Genes are taken from Table S3.9.

Gene ID	Gene Name	Gene Description	Gene Fitness Value
RL3752	*pssA*	putative transferase involved in exopolysaccharide biosynthesis	-0.47
RL1646	*celA*	putative cellulose synthase catalytic subunit [UDP-forming]	+0.21
RL1661	*gmsA*	putative transmembrane sugar transferase protein, glucomannan	-0.52
pRL100451	*rapA2*	putative autoaggregation protein	+0.15
RL3911	*rapB*	putative rhizobium adhering protein	+0.37
RL2702	*rapD*	putative rhizobium adhering-like protein	-0.15

3.2.10.3. Adaptation, regulation and metabolism

As with rhizosphere growth, genes involved in adaptation and detoxification of compounds were important during root colonisation (Table S3.9). These included a putative superoxide dismutase SodC (RL1521). SodC is involved in protection against ROS and has been best characterised in *S. meliloti* where it directly detoxifies ROS of plant origin (Ampe et al., 2003). Plants are known to release ROS in their root exudates (Nanda et al., 2010) and so bacteria anchored to the root surface are likely to be exposed to particularly high concentrations, making detoxification systems such as SodC essential for persistence.

Nineteen transcriptional regulators were found to contribute specifically to pea root colonisation (Table S3.9). Those with known function include VisN (RL0696), a global transcriptional activator of flagella, motility and chemotaxis genes (Sourjik et al., 2000; Tambalo et al., 2010); CasR (pRL120066) a TetR family transcriptional regulator; and

FlbT (RL0732) a repressor of flagellar biosynthesis. VisN is thought to form a heterodimer with VisR, which promotes the expression of *rem* that, in turn, positively regulates expression of *fla*, *mot* and *che1* genes (Sourjik et al., 2000; Tambalo et al., 2010). Rlv3841 *visN* mutants lack flagella and are consequently non-motile explaining their deficiency for colonisation of the pea rhizoplane (Tambalo et al., 2010). In *R. elti,* CasR represses the expression of *casA* under free-living conditions (Xi et al., 2000). CasA is a Ca^{2+}-binding protein called calsymin important for symbiosis, with *R. elti casA* mutants impaired in bacteroid development and nitrogen fixation (Xi et al., 2000). During *R. elti* colonisation of *P. vulgaris*, CasR repression of *casA* is lifted due to the presence of an inducing compound most likely of plant origin (Xi et al., 2000). This derepression is specific to colonisation of *P. vulgaris*, with no *casA* expression observed during colonisation of *M. sativa* or wheat roots. The role of CasA (pRL120065) in Rlv3841 is undetermined. A INSeq screen found no deleterious phenotype for Rlv3841 *casA* mutants at any stage of symbiosis with pea (Wheatley et al., 2020) and transcriptomic analysis of Rlv3841 pea bacteroids found no induction of *casA* (Green et al., 2019). FlbT has been characterised in *Caulobacter crescentus* where it binds to flagellin mRNA thereby repressing its translation (Anderson and Gober, 2000). In doing so, it couples flagellum assembly to gene expression, suggesting that proper regulation of flagella biosynthesis is essential for root colonisation (Mangan et al., 1999). *C. crescentus flbT* mutants overproduce flagellins and whilst motile show impairment in chemotaxis (Schoenlein and Ely, 1989).

Catabolism of *myo*-inositol provides a competitive advantage during nodulation of pea by Rlv3841, though the stage of symbiosis at which this advantage arises remains unclear (Fry et al., 2001; Pini et al., 2017). Genes involved in *myo*-inositol catabolism

are not upregulated in bacteroids, indicating that *myo*-inositol catabolism is important during rhizosphere growth or early in infection (Karunakaran et al., 2009). This is consistent with the fact that *myo*-inositol is present in pea root exudates (Pini et al., 2017). However, plate counts indicate that *myo*-inositol catabolism mutants have similar growth rates to wild type in the pea rhizosphere and investigation into whether *myo*-inositol or catabolic intermediates may be toxic in these mutants revealed no differences in mean generation times when grown on *myo*-inositol suggesting this not to be the case (Fry et al., 2001). Analysis of gene fitness values led to the identification of inositol dehydrogenase (*idhA*, RL3622) and 2-keto-*myo*-inositol dehydratase (*iolE*; RL1493) as important for Rlv3841 colonisation of pea roots (Table S3.9). Rlv3841 *iolE* mutants are unable to grow on *myo*-inositol as a sole carbon source and *S. melitoli* IdhA has been shown to catalyse the first step of *myo*-inositol degradation (Galbraith et al., 1998; Fry et al., 2001). Collectively this suggests that myo-inositol catabolism by Rlv3841 provides a competitive advantaged during root colonisation. Notably, *idhA* and *iolE* had fitness values of -0.90 and -0.93 respectively for rhizosphere growth. These are close to the cut-off limit of -1 used to identify genes of importance. This coupled with the fact these mutants were subsequently impaired for root colonisation suggests that the ability to catabolise myo-inositol does in fact provide a mild competitive advantage for growth in the pea rhizosphere and that this promotes competitive colonisation of pea roots as well as nodulation by Rlv3841.

3.2.11. Genes that upon mutation provide an advantage for rhizosphere growth and root colonisation

In addition to identification of genes important under a given condition, BarSeq also allows the identification of genes whereby mutation results in improved bacterial

fitness meaning their expression typically has a negative effect under the test condition. In total, 271 genes were found to have advantaged phenotypes for growth in the pea rhizosphere (Table S3.11) and 260 genes led to increased root colonisation upon mutation (Table S3.12). Most of these genes had mildly advantaged phenotypes with just 7% (18 genes) of those involved in rhizosphere growth and 10% (25 genes) of those in root colonisation assigned a strongly growth advantaged phenotype (Table 3.4, Table S3.11 and S3.12). Comparison of genes with advantaged phenotypes for pea rhizosphere growth and root colonisation revealed that 69 genes were common to both categories (Fig. 3.14).

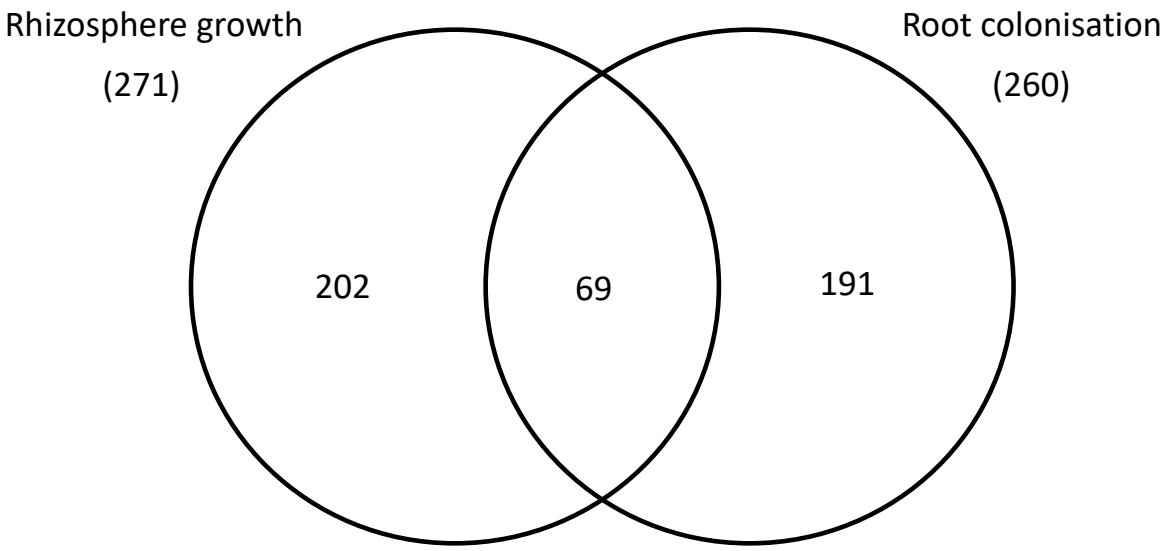


Figure 3.14. Venn diagram of the 271 genes that upon mutation provide an advantage for pea rhizosphere growth and 260 genes that when mutated result in increased root colonisation shows that 69 genes are common to both (rhizosphere-progressive). These genes are listed in Table S3.13.

Functional analysis revealed that for both rhizosphere growth and root colonisation over 40% of genes with advantaged phenotypes encode hypothetical proteins with no known function (Fig. 3.15). Genes encoding proteins involved in transport (11% rhizosphere, 10% colonisation), intermediary metabolism (9% rhizosphere, 10% colonisation), regulation (13% rhizosphere, 8% colonisation), cell membrane/envelope (8% rhizosphere, 9% colonisation) and nucleic acid replication/repair/synthesis (6% rhizosphere, 9% colonisation) accounted for the five largest functional categories of genes showing growth advantaged phenotypes for pea rhizosphere growth or root colonisation (Fig. 3.15 and Table S3.13).

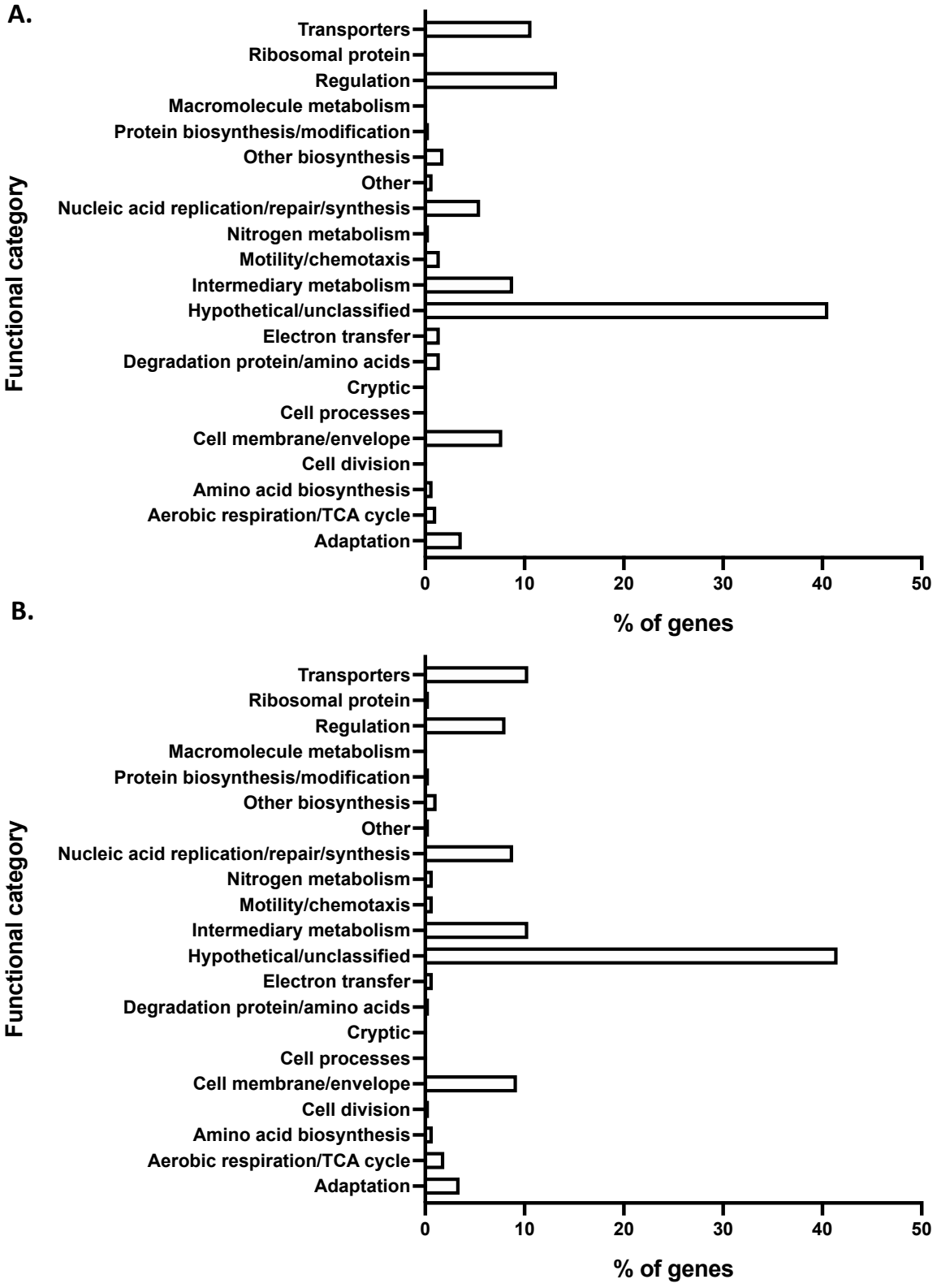

A.
Functional category
Transporters
Ribosomal protein
Regulation
Macromolecule metabolism
Protein biosynthesis/modification
Other biosynthesis
Other
Nucleic acid replication/repair/synthesis
Nitrogen metabolism
Motility/chemotaxis
Intermediary metabolism
Hypothetical/unclassified
Electron transfer
Degradation protein/amino acids
Cryptic
Cell processes
Cell membrane/envelope
Cell division
Amino acid biosynthesis
Aerobic respiration/TCA cycle
Adaptation
0
10
20
30
40
50
% of genes
B.
Functional category
Transporters
Ribosomal protein
Regulation
Macromolecule metabolism
Protein biosynthesis/modification
Other biosynthesis
Other
Nucleic acid replication/repair/synthesis
Nitrogen metabolism
Motility/chemotaxis
Intermediary metabolism
Hypothetical/unclassified
Electron transfer
Degradation protein/amino acids
Cryptic
Cell processes
Cell membrane/envelope
Cell division
Amino acid biosynthesis
Aerobic respiration/TCA cycle
Adaptation
0
10
20
30
40
50
% of genes

Figure 3.15. Functional classification of Rlv3841 genes where mutation leads to an advantage for growth in the rhizosphere or colonisation of pea. A. Rhizosphere growth. **B.** Root colonisation. All genes with average fitness values greater than 1. Functional classification based on Riley codes (Young et al., 2006). Genes are listed in Table S3.13.

The reason for a growth-advantaged phenotype is often less clear relative to the identification of genes required under a given condition. Mutation of transporters may result in a growth-advantaged phenotype because the functional transporter imports a toxic substrate. Alternatively, the substrate may not be important for growth in the rhizosphere, if present at all, and so alleviating the metabolic cost of synthesising the transporter may result in a competitive advantage. Mutation of genes involved in sensing or metabolism of specific compounds can also provide an advantage, particularly in nutrient poor conditions such as the rhizosphere, because energy is not unnecessarily wasted on the expression of genes involved in these pathways (Wheatley et al., 2020).

Examples of genes that provide a growth advantage upon mutation for rhizosphere growth and/or root colonisation include ABC transporters involved in the uptake of carbohydrates, RL3804-05 (CUT1) and pRL120193-94 (CUT2); peptides, RL4466, pRL110434, pRL110479 and pRL110516; and amino acids, pRL80060 and RL3908-10 (Table S3.13). pRL80060 encodes a PAAT SBP with homology to a mimosine transporter of *Rhizobium* sp. TAL 1145 (Borthakur et al., 2003), whilst RL3908-10 encodes a HAAT involved in the transport of protocatechuate (Garcia-Fraile et al., 2015). Under these experimental conditions, Rlv3841 can competitively grow without

the ability to import the substrates of these transporters. The SBP encoded by pRL80060 was previously found to confer a growth advantaged upon mutation for Rlv3841 growth in the pea rhizosphere (5 dpi) (Wheatley et al., 2020).

Other genes that upon mutation provided a growth advantage in the pea rhizosphere include CpaB (RL0213), a pilus assembly protein; MsiA, a putative canavanine exporter; and PhoB (RL0547), a regulator of the Pho regulon (Table S3.13). Canavanine is the toxic analogue of the amino acid arginine and has been found to be present in the root exudates of some plant species such as alfalfa, where it is capable of restricting bacterial growth (Emmert et al., 1998). In *Rhizobium msiA* is required for growth in the rhizosphere of canavanine producing plants and transcriptomics analyses have shown that Rlv3841 *msiA* is induced in the alfalfa rhizosphere but not pea or sugar beet (Cai et al., 2009; Ramachandran et al., 2011). Whilst canavanine is present in pea root exudates (Pini et al., 2017), it suggests that the concentration is too low to cause toxicity and therefore mutation of *msiA* does not have a deleterious effect on Rlv3841 growth in this environment. In *E. coli* the Pho regulon regulates the expression of a wide range of genes in response to phosphate limitation. These genes include those involved in phosphate metabolism, the biosynthesis of extracellular enzymes capable of obtaining inorganic phosphate from organic phosphates and phosphate-specific transporters including Ugp (glycerolphosphodiesters) and Phn (phosphonates uptake) (Santos-Beneit, 2015). RL0173 (*phnK*) encoding a component of a phosphate transporter also displayed increased root colonisation upon mutation (Table S3.13). The growth-advantaged phenotype for *phoB* and *phnK* suggests that inorganic phosphate is not limiting under

these conditions, which is consistent with the fact that the rooting solution supplied to these plants contained inorganic phosphate in the form of KH_2PO_4 and Na_2HPO_4.

Mutation of *casA* (pRL120065) also resulted in enhanced root colonisation. This is particularly interesting since the negative regulator of *casA*, CasR, was found to be disadvantaged for root colonisation upon mutation (Chapter 3.2.10.2 and Table S3.9). Perhaps surprisingly, mutation of the transcriptional regulator *praR* (RL0390) did not result in enhanced colonisation (Table S3.4). In Rlv3841 PraR represses expression of Raps, that promote attachment and aggregation of rhizobia. Mutation of *praR* therefore results in derepression of Rap proteins, resulting in increased biofilm formation on pea roots (Frederix et al., 2014). This enhanced attachment phenotype was not observed during the analysis of data obtained from the BarSeq screen or previous INSeq screens for Rlv3841 colonisation (5 dpi) and attachment to pea roots (1 hpi) (Parsons, 2019; Wheatley et al., 2020). Luminescence-based attachment assays have however demonstrated a clear increase in attachment for *praR* mutants 1 hpi relative to wild type at a neutral pH 7.0, though, under slightly acidic (pH 6.5) or alkaline (pH 7.5) conditions *praR* mutants showed reduced attachment to pea roots (Parsons, 2019). PraR only contains four TA sites and several TnSeq screens have demonstrated uncertainty in determining gene essentiality for those with less than six TA sites. Therefore, the lack of observed phenotype for *praR* may be the result of insufficient TA sites and thereby insufficient mutants to accurately calculate gene essentiality (Griffin et al., 2011; Zhang et al., 2012). This example highlights a limitation of the *mariner* transposon, however the majority of Rlv3841 genes (over 70%) contain at least six TA sites in the central 10-90% region.

3.3. Conclusion

In summary, a highly saturated barcoded transposon mutant library was generated in Rlv3841 and successfully validated *in planta* (Table 3.1 and Fig. 3.16). The Rlv3841 barcoded transposon pool contains 711,108 mapped transposon mutants, each containing a unique barcode identifier, mapped to 100,662 unique locations within the genome. The average insertion density within the central 10-90% of coding genes was over 73% making this library ideal for the determination of gene fitness values by BarSeq. Nonetheless, only approximately 26% of the unique barcodes present in the library were successfully mapped with the remainder failing to meet at least one of the four mapping criteria (Chapter 3.2.2). The primary reason for barcodes failing to map to the genome was the fact that they were seen less than five times. It is expected that further sequencing of the original mutant pool will result in the successful mapping of significantly more barcodes and therefore the retention of more sequencing data from subsequent BarSeq experiments, as well as the re-analysis of the data discussed here.

BarSeq was successfully used to identify 667 genes in Rlv3841 required for optimal growth in the pea rhizosphere (Fig. 3.10 and Table S3.5) and 494 genes contributing to root colonisation (Fig. 3.10 and Table S3.6). Most of these genes, 69% for rhizosphere growth and 76% for root colonisation, had mild effects, with mutants displaying a 50-75% reduction in rhizosphere growth or root colonisation (fitness value -1 to -2) relative to the average mutant in the population (Table 3.4). Only 31% and 24% of genes implicated in rhizosphere growth and root colonisation, respectively, showed strong phenotypes in which mutants displayed at least a 75% reduction in

growth or colonisation (fitness value less than -2) (Table 3.4). Forty-one genes previously linked to Rlv3841 pea rhizosphere fitness in INSeq screens and 49 genes shown to contribute to pea root colonisation were identified as important in these experiments, confirming the success of this approach using BarSeq (Tables S3.7 and S3.8) (Wheatley et al., 2020). These included genes involved in the biosynthesis of amino acids; RelA, the stringent response protein; *ropB*, encoding amyloid fibrils; NdvA, involved in the biosynthesis of cyclic ß-1-2-glcans and BacA, the VLCFA transporter.

In contrast to a previous INSeq screen (Wheatley et al., 2020), analysis of BarSeq data identified a clear role for chemotaxis and flagella mediated motility (Fig 3.13 and Table S3.10). The importance of the flagella and the *che1* chemotaxis cluster during Rlv3841 root colonisation is well established in the literature (Miller et al., 2007). One reason as to why the results from the previous INSeq screen and results described here differ, may be due to the way in which the data is analysed. To calculates gene fitness values from data obtained from BarSeq screens, the number of reads for a mutant after growth in each condition are compared to those at time zero. In contrast, the data obtained from the previous INSeq screen was analysed using the Hidden Markov Model (HMM) to assess gene essentiality. This method does not compare the frequency of reads for a given mutant after growth in a condition to those at time zero. Instead, it relies on the fact that the *mariner* transposon can only insert at TA sites. This allows *in silico* modelling of expected insertion frequencies within the genome whereby the read counts for individual TA sites are compared to the mean read count for that insertion library. In simplified terms, genes classed as growth-essential will have zero or very few reads. Growth-defective genes will have less than 1/100th of the

mean read count, genes with no phenotype when mutated will have a read count around the mean and genes that when mutated provide a growth-advantage will have over five times the mean read count (DeJesus and Ioerger, 2013). For screens examining bacterial colonisation of roots this HMM approach appears to be more conservative, classifying genes with documented roles in root colonisation, that have been experimentally validated, as neutral (e.g., *dctA*, *che1* and *flg* gene clusters (Miller et al., 2007; Ramachandran et al., 2011)). In contrast, BarSeq facilitates the identification of genes with milder, but still biologically relevant, fitness contributions. This is made possible by the high levels of internal replication provided by the ability to identify each individual mutant, calculate mutant fitness values and then use multiple mutant fitness values in the calculation of a gene fitness value. Furthermore, the fact that mutant libraries do not need to be cultured for several generations after retrieval from root niches and that fitness profiling by BarSeq requires a single PCR amplification of the barcode, collectively make BarSeq an ideal tool for the subsequent identification of plant-specific and competitive colonisation determinants in Rlv3841.

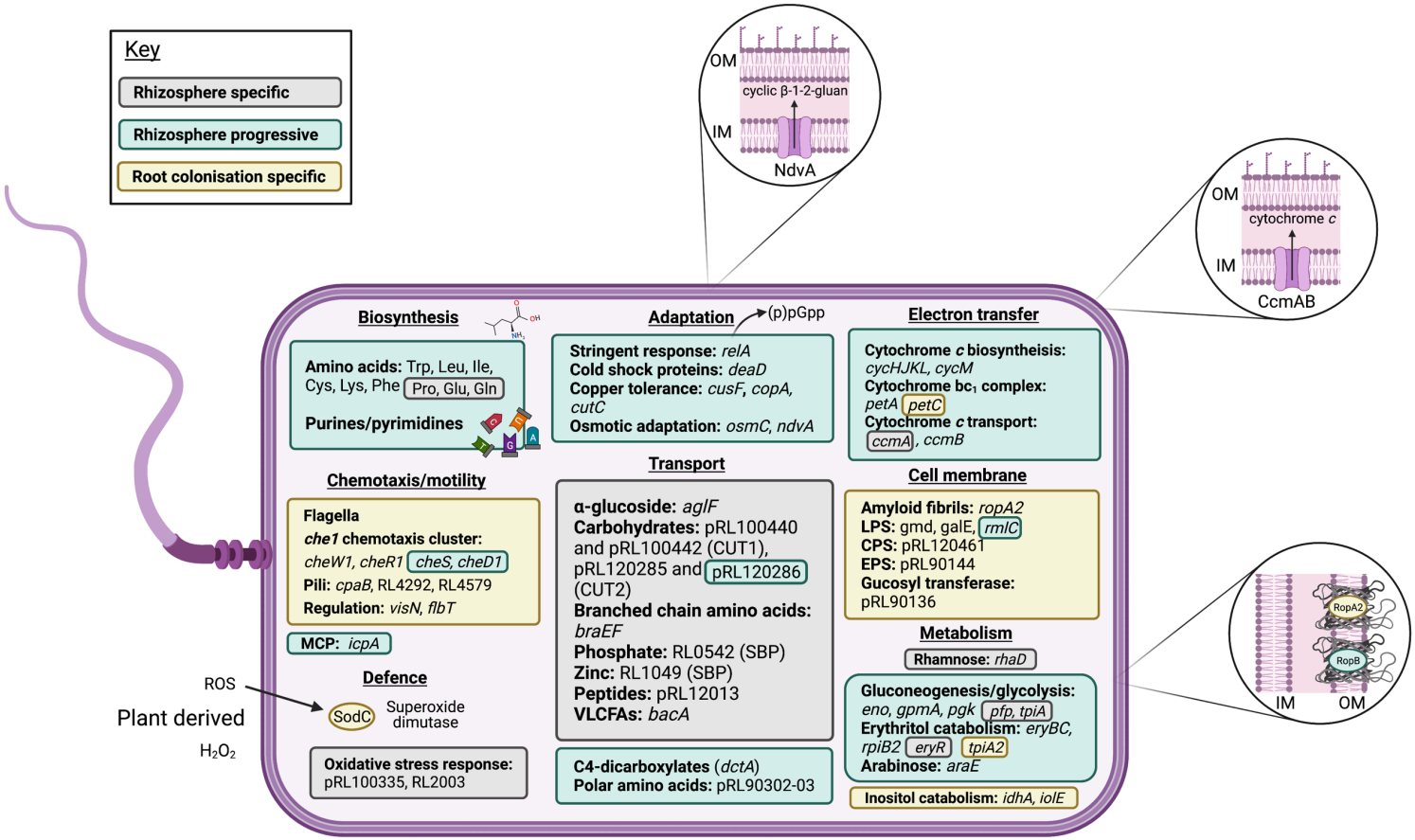


Figure 3.16. Summary of genes required for Rlv3841 pea rhizospheres growth and root colonisation. Genes shown in grey are specifically required for rhizosphere growth, genes shown in green are rhizosphere-progressive and genes shown in yellow are specific to root colonisation.

Chapter 4

Identifying common and plant-specific colonisation determinants in *R. leguminosarum* among host legumes, a non-host legume and a non-legume

.

4.1. Introduction

Rhizobium leguminosarum bv. *viciae* (*Rlv*) can form N_2-fixing symbioses with members of the Fabeae legume tribe which consists of *Pisum*, *Lens*, *Lathyrus*, and *Vicia* species (Andrews and Andrews, 2017). Although *Rlv* strains can form effective symbioses with all Fabeae, certain *Rlv* genotypes are differentially selected by specific plant genera suggesting host-specific variation in competitiveness to form nodules among *Rlv* genotypes (Laguerre et al., 2003; Mutch and Young, 2004; Jorrin and Imperial, 2015). For example, isolation of *Rlv* populations from bulk soils, rhizospheres, and nodules of fava bean and pea revealed that different Fabeae members demonstrate a preference for specific *Rlv* genotypes present in the soil population (Laguerre et al., 2003). For symbiosis with fava bean the symbiotic (*sym*) plasmid was identified as the primary determinant of successful nodule occupancy by *Rlv*, independent of its background genotype (chromosome and resident plasmids), the plant genotype and the soil. More specifically, the ability to nodulate fava bean correlates with the *nodDF* genotype (Laguerre et al., 2003; Mutch and Young, 2004). In contrast, analysis of *Rlv* populations isolated from pea nodules revealed that both the *sym* plasmid and genomic background contribute to competitiveness for nodulation of pea. This is consistent with a previous study that introduced the same *sym* plasmid into different *R. leguminosarum* genetic backgrounds and found that both the *sym* plasmid and background genotype contributed to competitiveness for nodulation of pea (Brewin et al., 1983). Studies aimed at characterising the stage in symbiosis at which certain *Rlv* genotypes are selected have shown that for pea selection may occur during competitive growth in the rhizosphere and root colonisation, since pea-selected *Rlv* genotypes are higher in the rhizosphere relative

to in the bulk soil. In contrast, fava bean *Rlv* populations are thought to be selected at a later stage of symbiosis i.e., in the infection thread, since they are no more competitive than other *Rlv* genotypes available in the bulk soil for rhizosphere colonisation (Laguerre et al., 2003).

Several studies provide support for the hypothesis that rhizosphere fitness contributes to *Rlv* competitiveness for nodulation of pea. First, inoculation of 104 *Rlv* genotypes onto various *Pisum* spp. demonstrated that compatible symbionts have various levels of competitiveness to form root nodules and that this is independent of their BNF efficiency (Bourion et al., 2018). Second, single nucleotide polymorphism analysis of *Rlv* populations isolated from pea, fava bean, lentil and vetch revealed that *Rlv* strains selected by pea had similar genotypes among non *sym*-encoded genes implicated in rhizosphere fitness, and that these were distinct from those found in other *Rlv* populations isolated from fava bean, lentil and vetch (Jorrin, 2016). Third, *Rlv* mutants impaired in the catabolism of arabinose and protocatechuate display reduced rhizosphere growth and decreased competitiveness for nodulation of pea (Garcia-Fraile et al., 2015). Additionally, catabolism of homoserine, an abundant amino acid in pea root exudates, corelates with competitiveness to nodulate pea (Hynes and O'Connell, 1990). Fourth, microarray analysis of *Rlv* during growth in pea, alfalfa and sugar beet rhizospheres identified a set of genes uniquely induced in the pea rhizosphere (Ramachandran et al., 2011). These genes had roles in transport, metabolism and adaptation. Subsequent mutation of two of these genes resulted in reduced competitiveness for pea rhizosphere colonisation, but not alfalfa, demonstrating their host-specific role in rhizosphere fitness. A recent study however did show that competitiveness for nodulation of pea is, at least in part, associated with

variation in specific regions of the *nod* locus, though additional genetic determinants were also implicated (Boivin et al., 2020) and *Rlv* growth on the root surface is attributed to the strains genetic background, not the *sym* plasmid (Brewin et al., 1983).

Deciphering the molecular mechanisms underlying competitiveness for nodulation is of great interest since compatible symbionts are not always the most effective at BNF, often resulting in suboptimal crop yields in domesticated legume species (Bourion et al., 2018). Traits that generally influence bacterial competitiveness during rhizosphere persistence and root colonisation include chemotaxis and motility, use of ABC transporters, EPS production and autoaggregation (Turnbull et al., 2001). However, a comparative lack of research into these initial stages of *Rhizobium*-legume interactions means that there is little knowledge regarding the host-specific factors required for rhizosphere persistence or root colonisation. Identification of factors that drive persistence in the rhizosphere of host legumes will also be of interest for the development of biological fertilisers that are commonly applied to agricultural systems to improve yields of leguminous crops. When applied to agricultural systems the introduced rhizobia, chosen for their superior N_2 fixation rates, are often outcompeted by native strains (Meade et al., 1985). A better understanding of the genetic determinants required for rhizosphere persistence and root colonisation would facilitate the targeted engineering of rhizobial strains and host legumes, to both enhance the existing symbiosis and transfer it to non-legumes.

In this chapter, the use of BarSeq to identify genes required for Rlv3841 rhizosphere growth and root colonisation among four host legumes from the Fabace tribe: pea, lentil, *Lathyrus* and fava bean, as well as the non-host legume alfalfa and non-legume

barley will be described. This work has identified plant-specific genes implicated in rhizosphere growth and root colonisation, in addition to a core set of genes required for rhizosphere persistence and root colonisation, irrespective of the plant species.

4.2. Results and discussion

4.2.1. Rlv3841 colonisation of lentil, *Lathyrus*, fava bean, alfalfa and barley

To assess if 7 dpi was a suitable timepoint at which to retrieve mutant libraries from the rhizosphere and roots for BarSeq analysis, the colonisation dynamics of Rlv3841 on lentil*, Lathyrus*, fava bean (host legumes), alfalfa (non-host legume) and barley (non-legume) were investigated (Fig. 4.1). Seven-day-old seedlings were inoculated with 10^5 cfu of Rlv3841 and populations were quantified at 1, 3, 5, and 7 dpi for the rhizosphere, loosely root-associated and strongly root-associated fractions. To account for the differences in the size of roots between plant species, bacterial numbers were normalised by root weight. As seen with pea (Chapter 3.2.4), a significant increase was observed from 1 to 7 dpi for each population (Fig. 4.1). Rlv3841 reached maximum colonisation of the rhizosphere and root by 5 dpi for all plant species, with no significant differences between 5 and 7 dpi (Table 4.1). For BarSeq experiments it is essential that mutant libraries undergo several generations of growth in the test condition to allow clear resolution of mutant phenotypes. After seven days of growth the total number of bacteria present had increased by a minimum of six generations, which is sufficient for resolution of mutant phenotypes (Table 4.1).

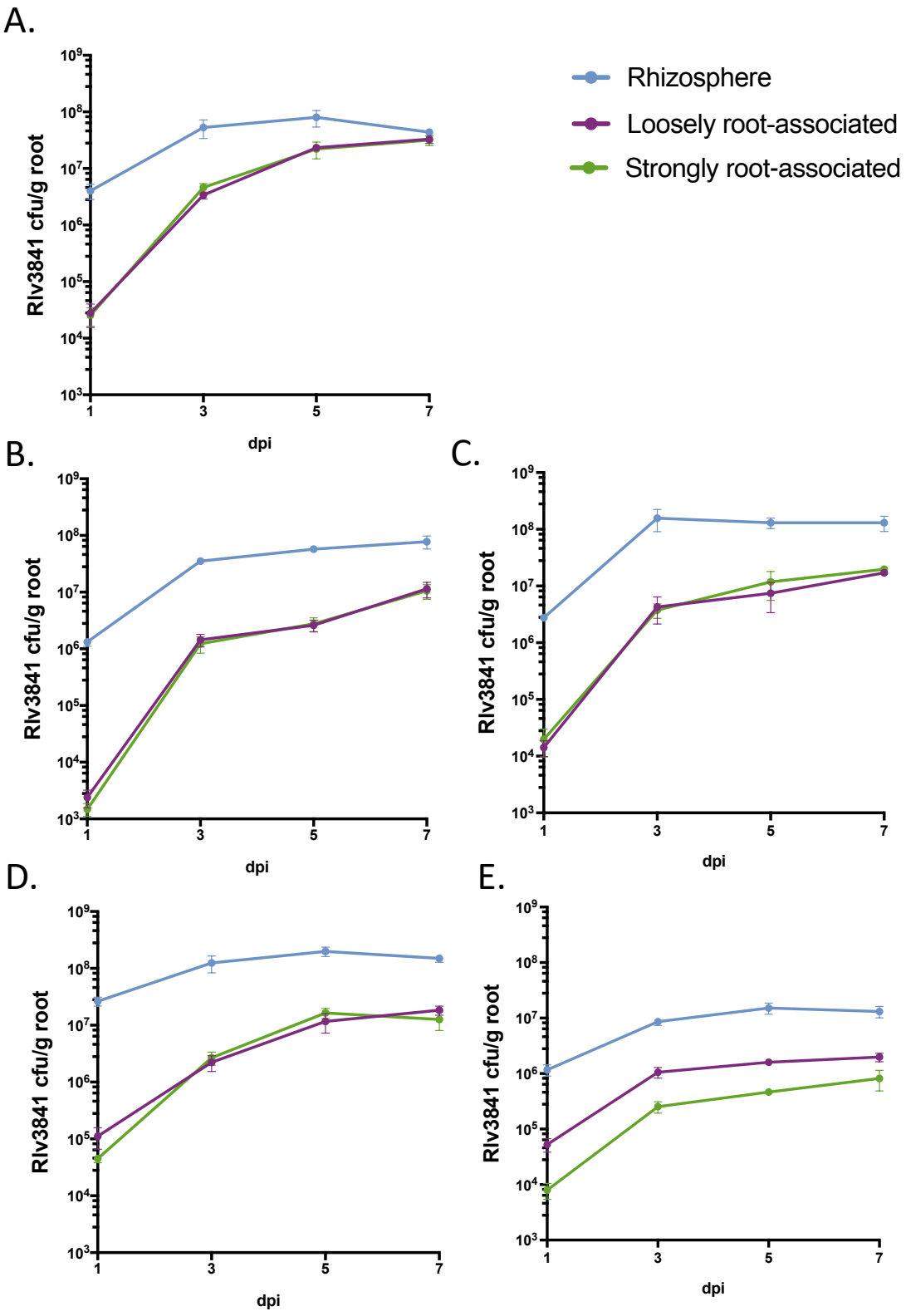

A.
B.
C.
D.
E.
Rhizosphere
Loosely root-associated
Strongly root-associated
Rlv3841 cfu/g root
dpi
10^9
10^8
10^7
10^6
10^5
10^4
10^3
1
3
5
7

Figure 4.1. Rlv3841 colonisation of lentil, *Lathyrus*, fava bean, alfalfa and barley. 7-day-old seedlings were inoculated with 10^5 cfu of Rlv3841. Mean populations of Rlv3841 in the rhizosphere and associated with the root are shown for **A.** lentil, **B.** *Lathyrus*, **C.** fava bean, **D.** alfalfa, and **E.** barley at 1, 3, 5 and 7 dpi. Error bars represent SEM. N ≥ 4. One-way ANOVA analysis was performed for each sample followed by Tukey's *post hoc* test. A significant increase in Rlv3841 cfu/g root was seen from 1 to 7 dpi for the rhizosphere ([lentil: F = 3.483, R^2 = 0.4106, p = 0.0425], [*Lathyrus*: F = 96.43, R^2 = 0.9538, p <0.0001], [Fava bean: F = 30.03, R^2 = 0.8573, p <0.0001] [alfalfa: F = 18.37, R^2 = 0.7750, p <0.0001] and [barley: F = 31.72, R^2 = 0.7425, p <0.0001]), loosely root-associated ([lentil: F = 39.41, R^2 = 0.8874, p <0.0001], [*Lathyrus*: F = 95.97, R^2 = 0.9568, p <0.0001], [Fava bean: F = 64.83, R^2 = 0.9284, p = 0.0008] [alfalfa: F = 65.12, R^2 = 0.9243, p <0.0001] and [barley: F = 51.08, R^2 = 0.8228, p <0.0001]), and strongly root-associated ([lentil: F = 9.671, R^2 = 0.6592, p <0.0001], [*Lathyrus*: F = 124, R^2 = 0.9637, p <0.0001], [Fava bean: F = 65.95, R^2 = 0.9295, p <0.0001], [alfalfa: F = 116.1, R^2 = 0.9561, p <0.0001] and [barley: F = 59.63, R^2 = 0.8443, p <0.0001]).

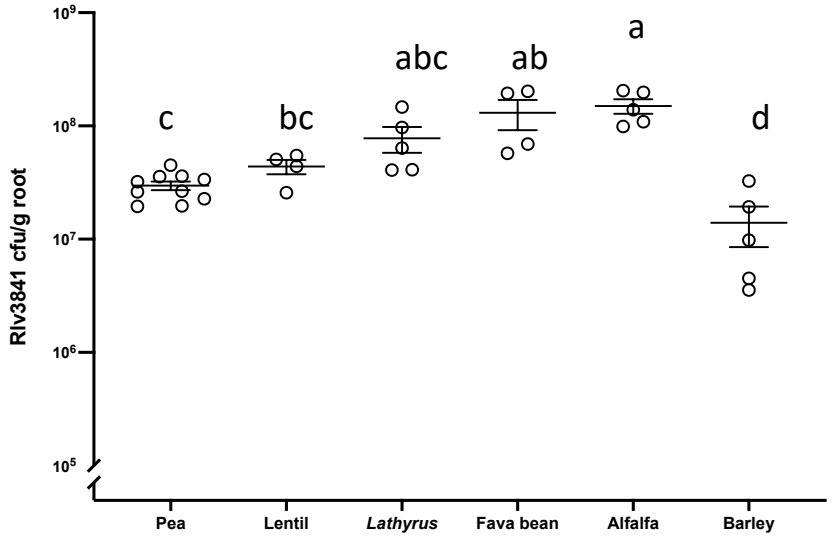
Table 4.1. RIv3841 population at 5 and 7 dpi for pea, lentil *Lathyrus,* fava bean, alfalfa and barley.

	Rhizosphere (cfu/g root)			**Loosely root associated** (cfu/ g root)			**Strongly root associated** (cfu/ g root)				
Plant	5 dpi	7 dpi	*p*-value	5 dpi	7 dpi	*p*-value	5 dpi	7 dpi	*p*-value	**Total** (cfu/plant)	**Number of generations**
Pea	$2.3x10^7$ ± $3.3x10^6$	$3.0x10^7$ ± $2.6x10^6$	0.4513	$8.9x10^6$ ± $1.7x10^6$	$1.4x10^7$ ± $3.0x10^6$	0.4841	$5.3x10^6$ ± $8.0x10^5$	$6.5x10^6$ ± $1.3x10^6$	0.9700	$4.45x10^7$ ± $5.7x10^6$	8
Lentil	$8.0x10^7$ ± $2.6x10^7$	$4.4x10^7$ ± $6.4x10^6$	0.4984	$2.3x10^7$ ± $2.0x10^6$	$3.3x10^7$ ± $5.4x10^6$	0.0675	$2.2x10^7$ ± $2.3x10^6$	$3.2x10^7$ ± $6.3x10^6$	0.5172	$2.32x10^7$ ± $2.3x10^6$	7
Lathyrus	$5.7x10^7$ ± $7.6x10^6$	$7.8x10^7$ ± $2.0x10^7$	0.8500	$2.6x10^6$ ± $6.1x10^5$	$1.1x10^7$ ± $3.5x10^6$	0.0548	$2.8x10^6$ ± $4.5x10^5$	$1.1x10^7$ ± $3.1x10^6$	0.1010	$8.94x10^6$ ± $1.7x10^6$	6
Fava bean	$1.3x10^8$ ± $2.8x10^7$	$1.3x10^8$ ± $3.9x10^7$	0.9997	$7.5x10^6$ ± $4.1x10^6$	$1.7x10^7$ ± $1.8x10^6$	0.1244	$1.2x10^7$ ± $6.2x10^6$	$1.9x10^7$ ± $1.7x10^6$	0.3606	$6.79x10^7$ ± $9.1x10^6$	9
Alfalfa	$1.9x10^8$ ± $3.8x10^7$	$1.5x10^8$ ± $2.2x10^7$	0.8339	$1.2x10^7$ ± $4.4x10^6$	$1.8x10^7$ ± $3.3x10^6$	0.4727	$1.6x10^7$ ± $3.3x10^6$	$1.3x10^7$ ± $4.6x10^6$	0.6342	$1.15x10^7$ ± $1.6x10^6$	7
Barley	$1.8x10^7$ ± $5.9x10^6$	$1.4x10^7$ ± $5.4x10^6$	0.9259	$1.8x10^6$ ± $2.8x10^5$	$1.3x10^6$ ± $3.0x10^5$	0.9994	$4.3x10^5$ ± $6.0x10^4$	$4.3x10^5$ ± $9.6x10^4$	0.9995	$9.72x10^6$ ± $3.0x10^6$	6

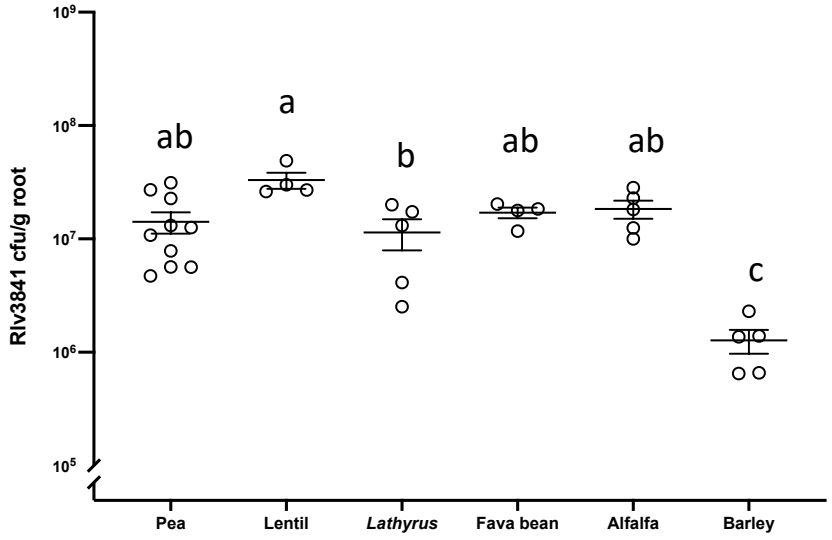
All pea data is from Chapter 3. SEM shown. One-way ANOVA analysis was performed for each sample followed by Tukey's *post hoc* test.

Comparison of Rlv3841 rhizosphere and root-associated populations of the six plant species revealed that at 7 dpi Rlv3841 was less effective at colonising barley, relative to host legumes and alfalfa (Fig. 4.2 and Table 4.1). The reason for this is unknown, but possibly Rlv3841 is better adapted for growth in the rhizosphere of these plants and this, in turn, promotes colonisation.

A.


Rlv3841 cfu/g root
10^9
10^8
10^7
10^6
10^5
c
bc
abc
ab
a
d
Pea
Lentil
Lathyrus
Fava bean
Alfalfa
Barley

B.


Rlv3841 cfu/g root
10^9
10^8
10^7
10^6
10^5
ab
a
b
ab
ab
c
Pea
Lentil
Lathyrus
Fava bean
Alfalfa
Barley

C.

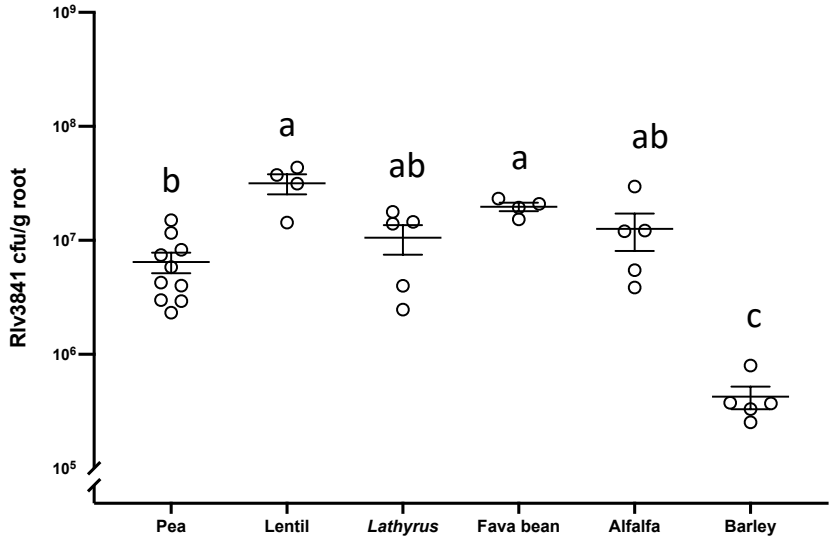

Rlv3841 cfu/g root
10^9
10^8
10^7
10^6
10^5
b
a
ab
a
ab
c
Pea
Lentil
Lathyrus
Fava bean
Alfalfa
Barley

Figure 4.2. Comparison of Rlv3841 populations among pea, lentil, *Lathyrus*, fava bean, alfalfa and barley after seven days root colonisation. A. rhizosphere, **B.** loosely root-associated, **C.** strongly root-associated. Error bars represent SEM. N ≥ 4. One-way ANOVA analysis was performed for each sample followed by Tukey's *post hoc* test. A significant difference in Rlv3841 populations among plants was seen for rhizosphere growth and root colonisation ([rhizosphere: $F = 17.25$, $R^2 = 0.7616$, $p < 0.0001$], [loosely root-associated: $F = 17.29$, $R^2 = 0.7620$, $p < 0.0001$] and [strongly root-associated: $F = 27.40$, $R^2 = 0.8354$, $p < 0.0001$]).

4.2.2. BarSeq experimental design to identify common and plant-specific determinants of Rlv3841 rhizosphere growth and root colonisation

Having established 7 dpi as a suitable time point at which to retrieve mutant libraries from the rhizosphere and root, a pool of 2×10^5 barcoded *mariner* insertion mutants was inoculated onto seven-day-old lentil, *Lathyrus*, fava bean, alfalfa and barley seedlings, and retrieved from the rhizosphere and root 7 dpi (Fig. 4.3). Samples for each plant species were pooled prior to gDNA extraction and barcodes were PCR-amplified for sequencing. Input, rhizosphere, loosely and strongly root-associated libraries were sequenced and analysed following the BarSeq data analysis pipeline (described in Chapter 2.5.7; (Wetmore et al., 2015)). To account for the differences in root sizes among plant species, three seeds were planted per boiling tube for lentil and *Lathyrus*, 10 for alfalfa and one for fava bean and barley. Increasing the number of seedlings sown per boiling tube for lentil, *Lathyrus* and alfalfa ensured that rooting systems grew fully throughout the vermiculite. In total, 96 boiling tubes were inoculated per plant species, providing a total input inoculum of 2.4×10^7 cfu. Considering that

only approximately 26% of the total data will be used in calculation of gene fitness values (due to constraints of mapping mutations discussed in Chapter 3.2.2), this represents approximately a 45-fold coverage of the 140,056 TA sites in Rlv3841, and an 830-fold coverage of the 7,528 genes. As with pea, due to low coverage of the strongly root-associated library, data for loosely and strongly root-associated libraries was combined and analysed collectively to identify genes required for root colonisation.

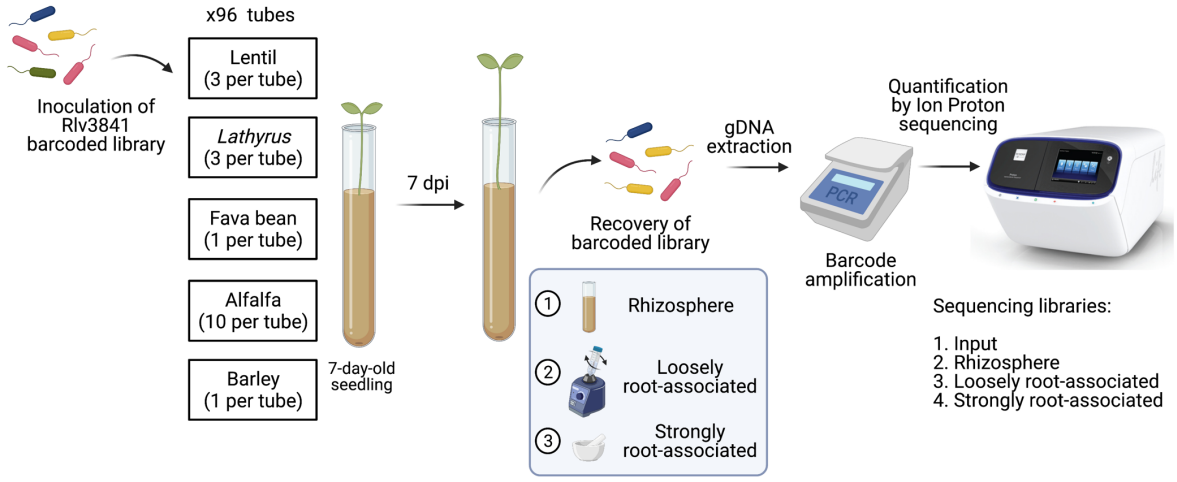


Figure 4.3. BarSeq experimental design. A total of 96 boiling tubes containing 7-day-old lentil, *Lathyrus*, fava bean, alfalfa or barley seedlings were inoculated with 10^5 barcoded *mariner* insertion mutants and retrieved from the rhizosphere and roots at 7 dpi. Samples were pooled, gDNA extracted, and barcodes amplified by PCR. Input, rhizosphere, loosely and strongly root-associated samples were sequenced by Ion Proton to quantify barcode frequencies.

4.2.3. Maintenance of Rlv3841 mutant library diversity *in planta*

Fitness values were assigned to 6,019 (80%) of 7,528 Rlv3841 genes (Table S4.1). Approximately 11% of genes were not assigned fitness values because they lacked insertions due to their essential role in Rlv3841 growth (Chapter 3.2.3). A further 9% of genes contained at least one mapped insertion, but lacked sufficient reads at time zero (input) for calculation of fitness values, indicating that they were relatively unfit in vitro compared to the other mutants in the pool. All experiments passed quality thresholds (Chapter 2.5.6), except for Rlv3841 rhizosphere growth and root colonisation of fava bean, which failed due to an insufficient number of reads per gene (median less than 50) for calculation of gene fitness. Subsequent analysis therefore excludes these two datasets. As seen with Rlv3841 BarSeq analysis of pea rhizosphere growth and root colonisation (Chapter 3.2.7), there were some bottlenecks in maintaining mutant diversity (Table 4.2). The mutant population retrieved from lentil roots contained the least diversity, with 20% (139,509) of the 711,108 mapped mutants present in the recovered population. However, as previously discussed (Chapter 3.2.7), the number of uniquely mapped mutants is much higher than those used in previous BarSeq screens and so even when only 20% of mapped mutants are recovered there is sufficient data to calculate gene fitness values. Additionally, the insertion density for this pool was 41% which is considered sufficient for calculation of gene essentiality (DeJesus and Ioerger, 2013).

In the subsequent sections, data obtained for Rlv3841 pea rhizosphere growth and root colonisation (Chapter 3) will be analysed in combination with that obtained for lentil, *Lathyrus*, alfalfa and barley to firstly identify genes commonly required for

rhizosphere growth and root colonisation, irrespective of the plant species, and secondly, to identify those that demonstrate plant-specific dependencies.

Table 4.2. Read coverage for genes in input, rhizosphere and root-colonised samples.

Library	No. unique barcodes sequenced	No. barcodes matching mutant in pool	No. barcodes as % of total mutant pool	Insertion density (No. TA sites with at least 1 read)	Median no. reads per gene
Lentil					
Input	2,942,159	595,161	84%	65%	267
Rhizosphere	1,176,410	255,303 (21.7%)	36%	52%	143
Root-colonised	691,698	139,509	20%	41%	259
Lathyrus					
Input	2,051,061	457,276	64%	62%	164
Rhizosphere	1,197,816	295,573	42%	54%	154
Root-colonised	740, 203	165,155	23%	44%	237
Alfalfa					
Input	3,426,918	631,153	89%	66%	841
Rhizosphere	763,588	176,680	25%	45%	99
Root-colonised	775,974	181,443	26%	46%	228
Barley					
Input	2,346,710	501,887	71%	63%	255
Rhizosphere	761,199	180,637	25%	46%	239
Root-colonised	1,028,639	232,476	33%	50%	233

4.2.4. Identification of genes required for growth in the rhizosphere and colonisation of lentil, *Lathyrus*, alfalfa and barley

Genes which when mutated lead to a strong disadvantage (fitness value less than -2) or mild disadvantage (fitness value between -1 and -2) for rhizosphere growth and root colonisation were identified for lentil, *Lathyrus*, alfalfa and barley (Fig. 4.4). Growth in the pea and alfalfa rhizosphere required the largest number of genes with 667 contributing to fitness, followed by barley (635), lentil (316) and *Lathyrus* (299) (Tables S3.5 and S4.2-S4.5). For root colonisation, 689 genes contribute to colonisation of alfalfa, 659 to *Lathyrus*, 628 to lentil, 524 to barley and 494 to pea (Tables S3.6 and S4.2-S4.5). Comparison of genes required for growth in the rhizosphere and colonisation of each plant species allowed identification of those that contribute to both. These genes were classed as rhizosphere-progressive (Fig. 4.4).

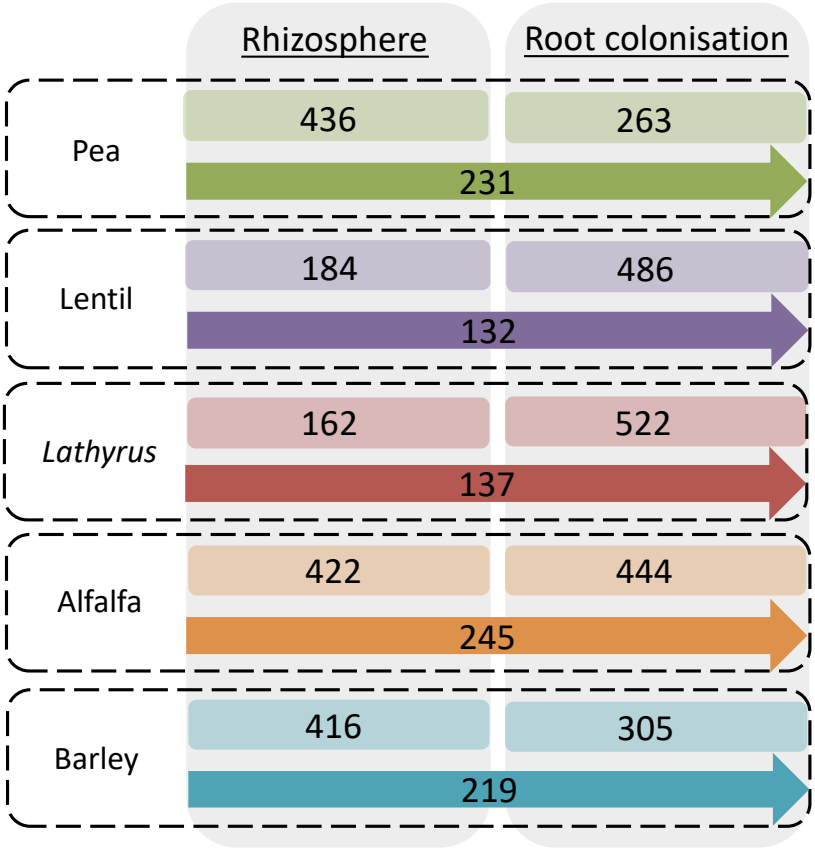


Figure 4.4. Number of genes required for growth in the rhizosphere and root colonisation of pea, lentil, *Lathyrus*, alfalfa and barley. All genes have fitness values less than -1. The number of rhizosphere-progressive genes are shown on the arrow. Genes are listed in Tables S3.9 (pea), S4.2 (lentil), S4.3 (*Lathyrus*), S4.4 (alfalfa) and S4.5 (barley).

Most genes implicated in rhizosphere fitness and root colonisation for each plant species had mild fitness contributions with a fitness value between -1 and -2, representing a 50-75% reduction in growth relative to the average mutant in the pool (Table 4.3). Only 16% of genes required for growth in the *Lathyrus* rhizosphere showed a strong fitness contribution (fitness value less than -2), 17% for lentil, 23% for both alfalfa and barley, whilst pea had the highest fraction of genes with strong fitness contributions at 31% (Tables S3.9 and S4.2-S4.5). The same was true for root colonisation, with only 22% of genes showing strong contributions to colonisation of *Lathyrus* and barley, 24% for pea, 26% for alfalfa and 28% for lentil. Most genes, over 82%, were assigned neutral fitness values, whilst 2-6% were assigned advantaged fitness values depending on the sample (Table 4.3).

Table 4.3. Distribution of Rlv3841 gene fitness values for pea, lentil *Lathyrus*, alfalfa and barley rhizosphere growth and root colonisation.

	Rhizosphere					Root colonised				
	Pea	Lentil	*Lathyrus*	Alfalfa	Barley	Pea	Lentil	*Lathyrus*	Alfalfa	Barley
Strong disadvantage	206 (3.4%)	53 (0.9%)	49 (0.8%)	154 (2.6%)	143 (2.4%)	118 (2%)	173 (2.9%)	144 (2.4%)	181 (3%)	113 (1.9%)
Mild disadvantage	461 (7.7%)	263 (4.4%)	250 (4.2%)	513 (8.5%)	492 (8.2%)	376 (6.2%)	445 (7.4%)	515 (8.5%)	508 (8.4%)	411 (6.8%)
Neutral	5,081 (84.4%)	5,594 (92.9%)	5,615 (93.3%)	5,011 (83.2%)	5,018 (83.4%)	5,265 (87.5%)	5,147 (85.5%)	4,987 (82.9%)	4,987 (82.9%)	5,223 (86.8%)
Advantaged	271 (4.5%)	109 (1.8%)	105 (1.7%)	341 (5.7%)	366 (6.0%)	260 (4.3%)	254 (4.2%)	373 (6.2%)	343 (5.7%)	272 (4.5%)

Number of strongly disadvantaged genes (fitness value < -2), mildly disadvantaged genes (fitness value between -1 and –2), neutral (fitness value between 1 and -1) and advantaged genes (fitness value > 1) are shown. Percentages shown in brackets are of the total 6,019 genes for which fitness values were assigned. Genes are listed in Table S4.1.

4.2.5. Common determinants of rhizosphere growth and root colonisation

Comparison of genes with fitness values less than -1 for a) rhizosphere growth and b) root colonisation among all five plant species led to the identification of 11 genes commonly required for optimal rhizosphere growth and 16 genes contributing to root colonisation of all five plant species (Fig.4.5 and Table 4.4). Comparison of these genes showed that four were required for both growth in the rhizosphere and colonisation of all five plant species (rhizosphere-progressive). The seven genes commonly required for rhizosphere growth, four genes which are rhizosphere-progressive, and 12 genes required for the colonisation of all five plant species will be discussed to give an overview of common factors required for Rlv3841 rhizosphere growth and root colonisation (Table 4.4).

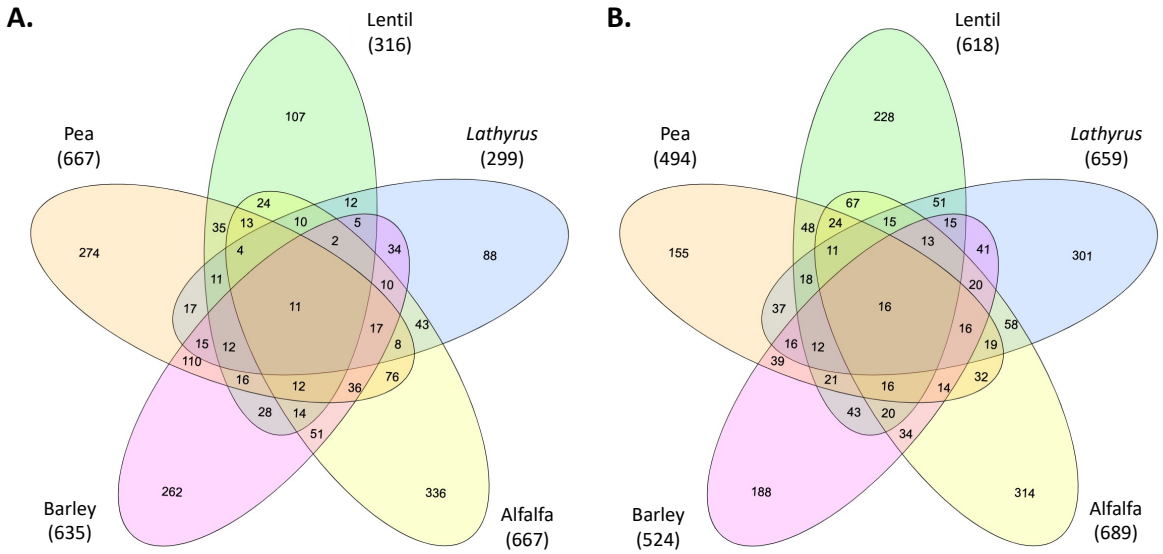


Figure 4.5. Venn diagram showing the number of genes contributing to growth in the rhizosphere and root colonisation of pea, lentil, *Lathyrus*, alfalfa and barley. A. Rhizosphere growth **B.** Root colonisation. All genes have a fitness value <

-1. Genes are listed in Tables S3.9 (pea), S4.2 (lentil), S4.3 (*Lathyrus*), S4.4 (alfalfa) and S4.5 (barley).

In the rhizosphere, biosynthesis of the amino acids leucine (RL1803; *ilvD1* and RL3245; *ilvI*) and isoleucine contributed to fitness, as did ribonucleotide biosynthesis (RL1596; *purM* and RL4045; *purK*) and riboflavin biosynthesis (RL1622; *ribC*) (Table 4.4). Riboflavin promotes rhizobial survival in the rhizosphere and root colonisation, with both the addition of exogenous riboflavin or overproduction by *S. meliloti* strains resulting in enhanced root colonisation (Streit et al., 1996; Yang et al., 2002; García Angulo et al., 2013). Also, *R. leguminosarum* strains impaired in riboflavin transport show reduced nodule occupancy of pea (García Angulo et al., 2013). The exact mechanism(s) by which riboflavin promotes rhizobial rhizosphere growth, root colonisation and nodule occupancy remain unclear. Only one gene with a putative role in transport, RL4402 (*thiY*), was commonly required. RL4402 encodes an SBP of a nitrate/nitrite/cyanate ABC transporter predicted to be involved in the transport of hydroxymethylpyrimidine (HMP) (Bale et al., 2010; Novichkov et al., 2013). HMP is a precursor of thiamine (vitamin B_1), an essential cofactor required for the metabolism of branched-chain amino acids and carbohydrates (Karunakaran et al., 2006). The lack of commonality in transporters required for rhizosphere growth likely reflects diversity in root exudates among the five plant species.

The remaining four genes commonly required for rhizosphere growth are RL4006 (*cbbT*), a putative transketolase involved in the non-oxidative branch of the pentose phosphate pathway; pRL110545 (*minD*), a putative septum site-determining protein; RL0380 (*rph*), a putative ribonuclease; and pRL80128, encoding a conserved

hypothetical protein which was rhizosphere-progressive (Table 4.4). pRL80128 is located within a cluster of *tra* genes on pRL8 and has previously been identified in an INSeq screen as important for growth of nodule bacteria and bacteroid development during the Rlv3841-pea symbiosis (Wheatley et al., 2020). The *tra* genes encode a type IV pili, which in gram negative bacteria partake in numerous processes including biofilm formation, adhesion, motility and secretion of factors involved in colonisation (Hahn, 1997; Dörr et al., 1998; O'Toole and Kolter, 1998; Craig et al., 2004), though the function of pRL80128 remains unknown.

A further 12 genes were required for root colonisation, but not growth in the rhizosphere of all plants (Table 4.4). These include RL0707 (*flgB*) and RL0686 (*cheX1*/*cheS*) with roles in chemotaxis and motility, essential bacterial traits for efficient root colonisation (de Weert et al., 2002; Miller et al., 2007; Allard-Massicotte et al., 2016; Liu et al., 2020). FlgB is required for the biosynthesis of the membrane-embedded flagella basal body, forming part of the rod structure that connects the motor to the hook (Liu and Ochman, 2007), whilst CheS modulates chemosensory responses (Dogra et al., 2012). *S. meliloti cheS* mutants swim 20% slower than the wild type (Dogra et al., 2012) and so mutation of *cheS* in Rlv3841 may impede movement towards the root surface.

The cyclic β-(1,2) glucan transporter NdvA and two genes with putative roles in intermediary metabolism, namely RL0257 a putative acetyltransferase and pRL90210 a putative short-chain dehydrogenase, were also common determinants of root colonisation (Table 4.4). As previously discussed, cyclic β-(1,2) glucans are important for adaptation to changes in osmolarity, however, *S. meliloti ndvA* mutants show

reduced attachment to alfalfa roots after 1-h incubation, suggesting that they also function as adhesins, possibly binding plant root lectins (Planqué and Kijne, 1977; Dylan et al., 1990; Breedveld and Miller, 1994). Whilst RL0257 and pRL90210 are involved in metabolism and therefore do not have direct roles in colonisation, they may be important for bacterial fitness on the rhizoplane. Several genes involved in the biosynthesis of leucine (RL3205; *ilvC* and RL4555), tryptophan (RL0022; *trpA*), and ribonucleotides (RL0572, RL1738; *pyrC2* and pRL2612; *purLI*) were also important for root colonisation.

Table 4.4. 23 genes commonly required for growth in the rhizosphere and/or root colonisation of pea, lentil, *Lathyrus* alfalfa and barley. All genes have fitness values less than < -1 under the corresponding conditions.

Gene	Description	Function
Rhizosphere growth		
RL0380 (*rph*)	putative ribonuclease	Degradation of RNA
RL1596 (*purM*)	putative phosphoribosylformylglycinamidine cyclo-ligase	Purine ribonucleotide biosynthesis
RL1803 (*ilvD1*)	putative dihydroxy-acid dehydratase	Amino acid biosynthesis
RL3513 (*leuA1*)	putative 2-isopropylmalate synthase	Amino acid biosynthesis
RL4006 (*cbbT*)	putative transketolase	Aerobic respiration
RL4045 (*purK*)	putative phosphoribosylaminoimidazole carboxylase ATPase subunit	Purine ribonucleotide biosynthesis
pRL110545 (*minD*)	putative septum site-determining protein MinD	Cell division
Rhizosphere-progressive		
RL1622 (*ribC*)	putative riboflavin biosynthesis protein	Electron transfer
RL3245 (*ilvI*)	putative acetolactate synthase isozyme III large subunit	Amino acid biosynthesis
RL4402 (*thiY*)	putative SBP of ABC transporter NitT (Nitrate/Nitrite/Cyanate)	Transport/binding proteins

pRL80128	conserved hypothetical protein	Conserved in organism other than *E. coli*
Root colonisation		
RL0022 (*trpA*)	putative tryptophan synthase alpha subunit	Amino acid biosynthesis
RL0257	putative acetyltransferase	Intermediary metabolism
RL0572	putative dihydroorotate dehydrogenase	Ribonucleotide biosynthesis
RL0686 (*cheS*)	putative chemotaxis related CheX protein	Motility/chemotaxis
RL0707 (*flgB*)	putative flagellar basal-body rod protein FlgB	Motility/chemotaxis
RL1738 (*pyrC2*)	putative dihydroorotase	Ribonucleotide biosynthesis
RL_RS12565	hypothetical protein	Unknown function
RL2612 (*purL*)	putative phosphoribosylformylglycinamidine synthase II	Ribonucleotide biosynthesis
RL3205 (*ilvC*)	putative ketol-acid reductoisomerase	Amino acid biosynthesis
RL4555	putative 3-isopropylmalate dehydratase	Amino acid biosynthesis
RL4640 (*ndvA*)	putative ATP-binding:permease (ABC:IMP) component of ABC transporter Export beta-(1→2) glucan export	Transporter
pRL90210	putative short-chain dehydrogenase	Intermediary metabolism

From this screen we have identified a total of 23 genes that are commonly required for growth in the rhizosphere and/or root colonisation of these five plant species (Table 4.4). This seems to be too few for such similar processes. It is likely that many genes involved are simply excluded through the strict criteria used to determine the cut-off. The initial cut-off fitness value of -1 represents at least a 50% reduction in growth relative to the average mutant in the pool. To identify further genes commonly required for rhizosphere growth and/or root colonisation this was relaxed to include genes with a fitness value of -1 or less for at least two plant species, and a fitness value between -0.42 and -1 for the remaining. A fitness value of -0.42 represents a 25% reduction in growth or root colonisation relative to a typical mutant in the pool. By relaxing the criteria a further 62 genes with putative roles in rhizosphere growth and 74 genes contributing to root colonisation were identified. Comparison of the total 73 genes with a role in rhizosphere growth (62 genes, together with the 11 genes previously identified), and 90 genes with a role in root colonisation (74 genes, plus 16 previously identified), revealed 32 were rhizosphere-progressive for colonisation of all five plant species (Fig.4.6 and Table 4.5).

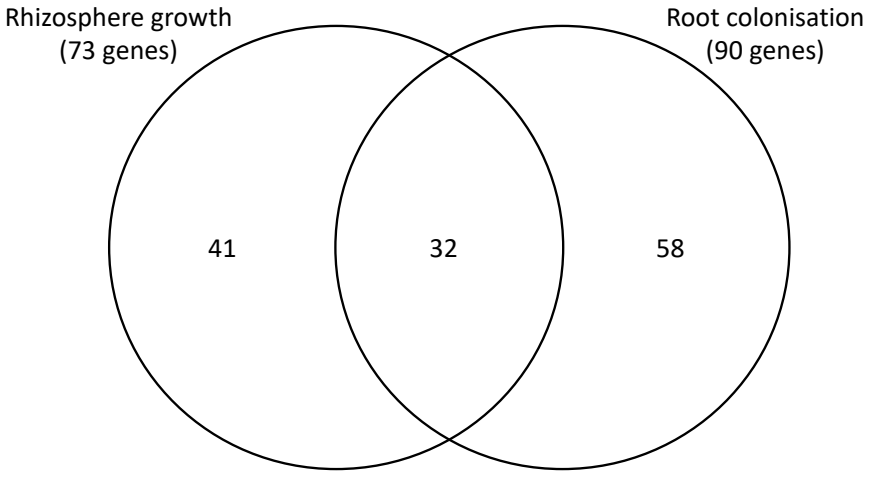
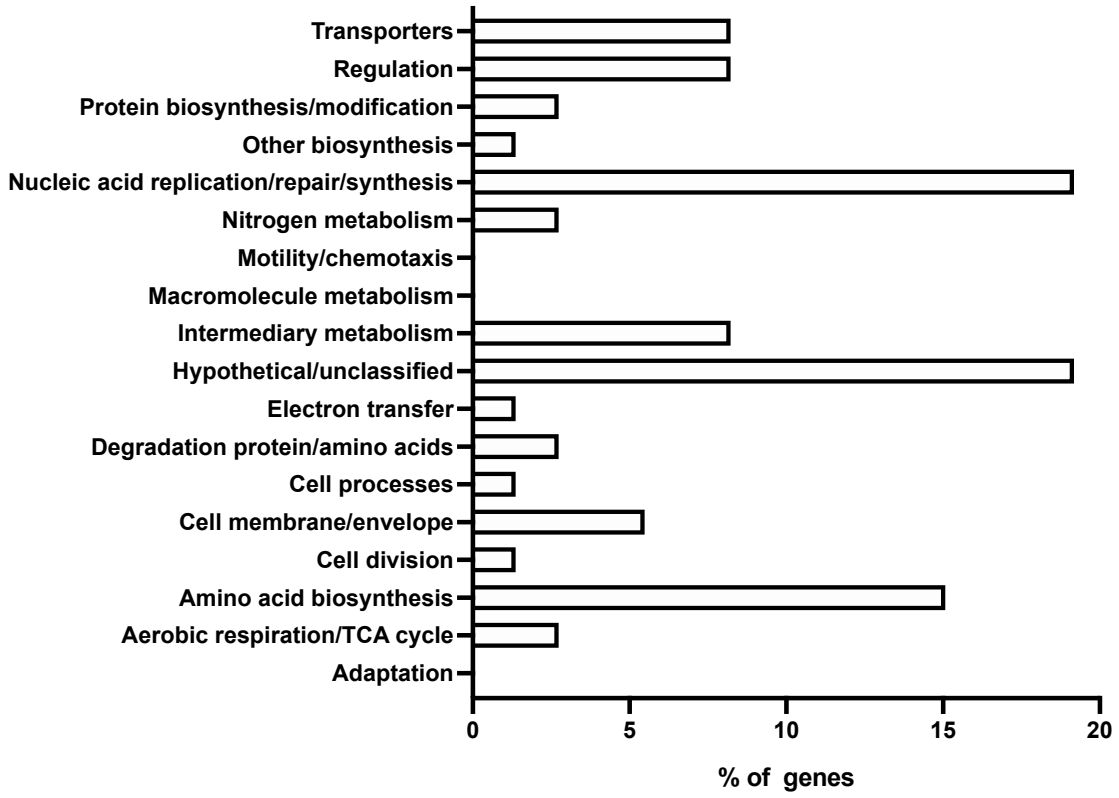


Figure 4.6. Venn diagram of the 73 genes commonly required for rhizosphere growth and 90 genes commonly required for root colonisation shows 32 genes are required for both (rhizosphere-progressive). These genes are listed in Tables 4.5 and 4.6 (see Table S4.6 for genes listed with associated fitness values).

Functional classification of the 73 genes contributing to rhizosphere growth (rhizosphere and rhizosphere-progressive) revealed that the majority fell into one of seven major categories: hypothetical/unclassified (19%), nucleic acid replication/repair/synthesis (19%), amino acid biosynthesis (15%), regulation (8%), intermediary metabolism (8%), transporters (8%) and cell membrane/envelope (5%) (Fig. 4.7A). The remaining categories contained less than 3% of the total 73 genes, with no common requirement of genes involved in adaptation or chemotaxis and motility. Functional classification of the 58 genes involved only in root colonisation revealed that the top categories were hypothetical/unclassified proteins (31%), regulation (12%), motility/chemotaxis (10%) and intermediary metabolism (10%) (Fig. 4.7B). The distribution of genes required for growth in plant rhizospheres contrasts with those needed for plant colonisation. In the rhizosphere, no genes involved in chemotaxis and motility were commonly important (compared to 10% in plant colonisation), hypothetical proteins only accounted for 19% of genes (compared to 31%), and genes involved in nucleic acid replication/repair/synthesis accounted for 19% (compared to 5%).

A.


Transporters
Regulation
Protein biosynthesis/modification
Other biosynthesis
Nucleic acid replication/repair/synthesis
Nitrogen metabolism
Motility/chemotaxis
Macromolecule metabolism
Intermediary metabolism
Hypothetical/unclassified
Electron transfer
Degradation protein/amino acids
Cell processes
Cell membrane/envelope
Cell division
Amino acid biosynthesis
Aerobic respiration/TCA cycle
Adaptation
0
5
10
15
20
% of genes

B.

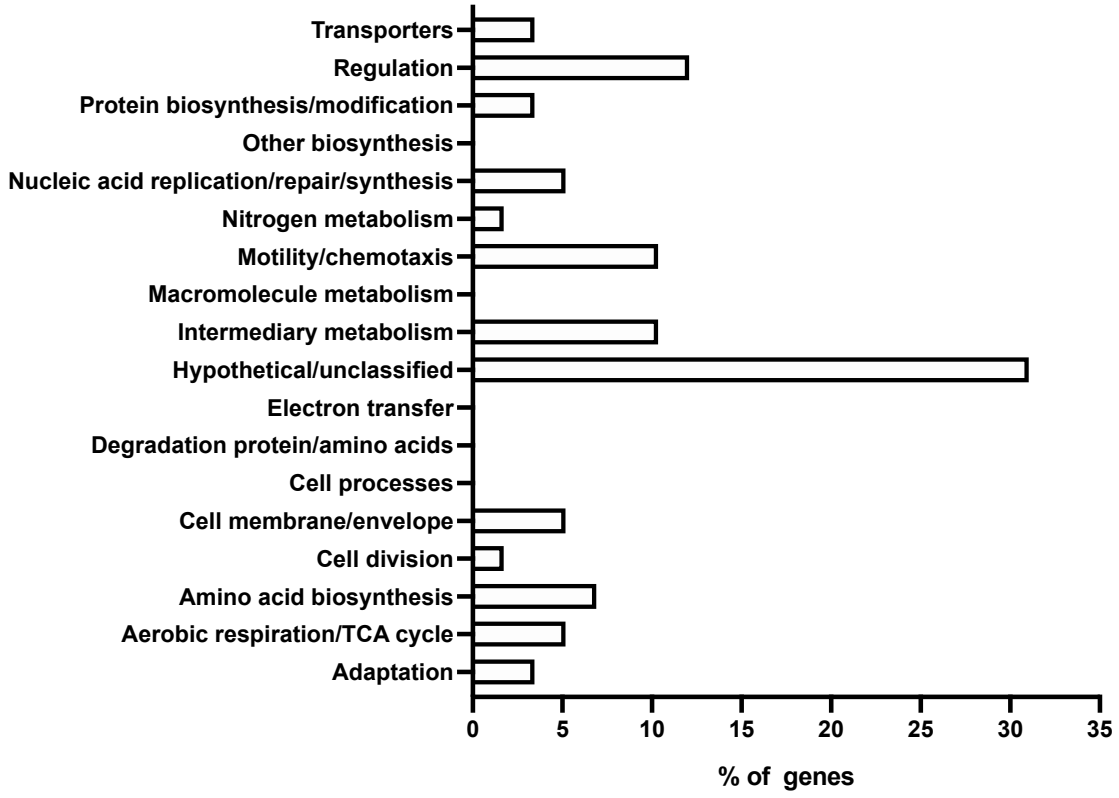

Transporters
Regulation
Protein biosynthesis/modification
Other biosynthesis
Nucleic acid replication/repair/synthesis
Nitrogen metabolism
Motility/chemotaxis
Macromolecule metabolism
Intermediary metabolism
Hypothetical/unclassified
Electron transfer
Degradation protein/amino acids
Cell processes
Cell membrane/envelope
Cell division
Amino acid biosynthesis
Aerobic respiration/TCA cycle
Adaptation
0
5
10
15
20
25
30
35
% of genes

Figure 4.7. Functional classification of genes required for growth in the rhizosphere and colonisation of pea, lentil, *Lathyrus*, alfalfa and barley. A. 73 genes required for rhizosphere growth (rhizosphere and rhizosphere-progressive; genes listed in Table 4.5). **B.** 58 genes specific for root colonisation (genes listed in Table 4.6). Functional classification is based on Riley codes (Young et al., 2006).

4.2.5.1. Common requirements for growth in the rhizosphere of five plant species

The expanded list of 41 genes required for the growth in all five plant rhizospheres and 32 genes identified as rhizosphere-progressive for all five plant species will be discussed collectively to give an overview of genes commonly required for rhizosphere fitness (Table 4.5). Amino acid biosynthesis is a clear prerequisite for rhizosphere growth. Further to the three genes previously identified (Table 4.4), genes involved in the biosynthesis of leucine (RL3205, RL4555 and RL4705), tryptophan (RL0003 and RL0022; *trpA*), proline (RL3460; *proC1*), glutamine (RL4082; *gltD*) and glutamate (RL4085; *gltB*) contributed to rhizosphere fitness. Similarly, further genes involved in ribonucleotide (RL0847; *guaB*, RL1546; *purF*, RL1595; *purN*, RL2606; *purC1*, RL2612; *purI*, RL3768; *purA*, RL0572, RL1738; *purC2* and RL1739; *pyrB*) and cofactor biosynthesis (RL1584; *moaE*) were important (Table 4.5).

In addition to RL4402 involved in the transport of HMP, a further five genes were identified with putative roles in transport for all plant rhizospheres (Table 4.5). These include pRL80047, a major facilitator superfamily (MFS) transporter involved in the efflux of unknown substrate(s); RL2435 (*copB*), a TolC efflux protein which confers

copper tolerance (Rensing and Grass, 2003); and pRL90110 (*ugpC2*), involved in the import of glycerol-3-phosphate. Glycerol-3-phosphatre is an intermediate in phospholipid biosynthesis and a precursor of glyceraldehyde-3-phosphate, an intermediate in the glycolysis pathway (Lemieux et al., 2004). A hypothetical protein RL4399 belonging to the YKOF superfamily and showing 30% similarity to *B. subtilis* YkoF, a SBP involved in the transport of HMP/thiamine, was also identified (Devedjiev et al., 2004).

A putative erythritol phosphate dehydrogenase (*eryB*; pRL120205) was commonly required for optimal rhizosphere growth. As discussed in Chapter 3, erythritol is present in pea root exudates and the ability to catabolise it is important for *R. leguminosarum* competitive nodulation of pea (Yost et al., 2006). The common requirement for *eryB* in all plant rhizosphere suggests that lentil, *Lathyrus*, alfalfa and barley root exudates also contain erythritol and that the ability to catabolise this is important for optimal rhizosphere fitness.

LPS plays an important role as a protective barrier against threats present in the external environment (Bertani and Ruiz, 2018). This can include antimicrobial compounds released by plants in response to the triggering of their defences, a process which is not only induced through interaction with pathogenic bacteria but also symbiotic partners (Azpilicueta et al., 2004). RL0826 (*fcl*) a putative GDP-L-fucose synthetase involved in LPS O-antigen biosynthesis was commonly required for Rlv3841 rhizosphere growth (Muszynski et al., 2011) (Table 4.5). The importance of GDP-L-fucose biosynthesis in rhizosphere fitness has been demonstrated in *R. tropici*, where mutants unable to synthesis GDP-L-fucose have altered O-antigen structure

and demonstrated reduced growth in the maize rhizosphere (Ormeño-Orrillo et al., 2008). This phenotype was at least in part due to increased susceptibility to maize antimicrobials. RL1155, a putative glycosyltransferase, and RL1383, a peptidoglycan binding transmembrane protein, were also important (Table 4.5). Glycosyltransferases constitute a large family of proteins that transfer sugar molecules to acceptor molecules. They are required for the biosynthesis of polysaccharide (EPS/LPS/CPS), glycolipids and glycoproteins, thereby playing an essential role in cell wall biogenesis among other processes (Zhu et al., 2013).

The rhizosphere-induced protein RhiB (pRL100170) was important for growth in all plant rhizospheres. The gene encoding RhiB belongs to the *rhi* cluster consisting of *rhiABCR*, which is only found in *Rlv* strains and is located on the symbiosis plasmid pRL10 between the *nod* and *nif* genes. RhiA and RhiB are cytoplasmic proteins whilst RhiC is periplasmic. Their expression is regulated in a cell density dependent manner via RhiR, which positively regulates their transcription in response to the *N*-acyl-homoserine lactones; a class of signalling molecules involved in quorum sensing that can inhibit rhizobial growth (Cubo et al., 1992; Rodelas et al., 1999). The functions of *rhiABC* remain unknown, but due to their location on the symbiotic plasmid it has been suggested that they mediate a currently uncharacterised host-specific interaction between *Rlv* and its symbiotic partners (pea, vetch, lentil and *Lathyrus* spp.) (Wisniewski-Dyé and Downie, 2002). However, this requirement for RhiB in all plant rhizospheres suggests more general importance for this protein during rhizosphere growth.

Bacteria make use of two-component sensor kinase systems to sense environmental stimuli and adapt accordingly. One two-component system FeuPQ (RL1433-34) was important for growth in the rhizosphere of all five plant species (Table 4.5). Whilst *R. leguminosarum feuQ* mutants have been shown to have reduced rates of iron uptake (Yeoman et al., 1997), the system is best characterised in *S. meliloti* where it senses extracellular changes in osmolarity and, in response, alters the transcription of 16 genes including *ndvA* (Griffitts et al., 2008). NdvA is a cyclic β-(1,2) glucan transporter and *S. meliloti feuP* mutants are unable to transport cyclic β-(1,2) glucans to the periplasm, resulting in impaired growth under hypoosmotic conditions and formation of Fix^- nodules (Griffitts et al., 2008). Rlv3841 FeuP shares 86% identity to *S. meliloti* FeuP and so a similar mechanism may exist in Rlv3841, in which FeuPQ senses changes in osmolarity in addition to affecting iron uptake, though this is yet to be confirmed. NtrB (RL2256), which phosphorylates NtrC (RL2257) in response to nitrogen limitation, was required in all plant rhizospheres (Fischer, 1994). Phosphorylation of NtrC results in altered transcription of nitrogen-regulated genes (Patriarca et al., 1992; Brown et al., 2014). Plants were grown in N-free rooting solution and so it makes sense that this general nitrogen regulatory system was important. Other transcriptional regulators of common importance for rhizosphere fitness include pRL120618, a LacI family transcriptional repressor that may be involved in the regulation of sugar utilisation (Novichkov et al., 2013) and RL3453, part of a two-component sensor/regulator (histidine kinase) with low homology (29% identity) to the well-characterised EnvZ protein of *E. coli* that coordinates responses to osmotic stress (Cai and Inouye, 2002) (Table 4.5).

Table 4.5. 73 genes required for growth in the rhizosphere of pea, lentil, *Lathyrus*, alfalfa and barley.

Gene	Description	Function
Rhizosphere growth		
RL0003	putative shikimate dehydrogenase	Amino acid biosynthesis
RL0346 (*dut*)	putative deoxyuridine' 5'triphosphate nucleotidohydrolase	Nucleic acid replication/repair/synthesis
RL0486	hypothetical protein	Hypothetical
RL0826 (*fcl*)	putative GDP-L-fucose synthetase	Cell membrane/envelope
RL0860	putative adenosylhomocysteinase (S-adenosyl-L-homocysteine hydrolase)	Intermediary metabolism
RL1067	hypothetical protein	Hypothetical
RL1085	hypothetical protein	Hypothetical
RL1155	putative glycosyl transferase	Cell membrane/envelope
RL1383	putative peptidoglycan binding transmembrane protein	Cell membrane/envelope
RL1433 (*feuP*)	putative two-component sensor/regulator; transcriptional regulator	Regulation
RL1595 (*purN*)	putative' 5'-phosphoribosylglycinamide formyltransferase	Nucleic acid replication/repair/synthesis
RL1901	hypothetical protein	Hypothetical
RL2007	epimerase	Hypothetical
RL2084	putative acetyltransferase	Intermediary metabolism

RL2212 (*clpS1*)	putative ATP-dependent Clp protease adaptor protein	Degradation protein/amino acids
RL2227	putative transmembrane protease	Degradation protein/amino acids
RL2256 (*ntrB*)	putative two-component sensor/regulator; histidine kinase NtrB	Regulation
RL2290 (*cycI*)	putative sulfite reductase	Intermediary metabolism
RL2567 (*rLuC*)	putative ribosomal large subunit pseudouridine synthase C	Nucleic acid replication/repair/synthesis
RL2568 (*crcB*)	putative CrcB family transmembrane protein (fluoride ion transporter)	Cell membrane/envelope
RL2787	putative ATP-binding component of ABC transporter Export	Transporters
RL3321	putative DnaJ family chaperone	Cell processes
RL3453	putative two-component sensor/regulator; histidine kinase	Regulation
RL3460 (*proC1*)	putative pyrroline-5-carboxylate reductase	Amino acid biosynthesis
RL3479	putative GTP-dependent nucleic acid-binding protein	Transporters
RL3980	putative phosphatase	Other biosynthesis
RL4084 (*gltD*)	putative glutamate synthase [NADPH] small chain	Amino acid biosynthesis
RL4085 (*gltB*)	putative glutamate synthase [NADPH] large chain	Amino acid biosynthesis
RL4399	conserved hypothetical protein YKOF super family	Hypothetical
RL4526	pterin-4-alpha-carbinolamine dehydratase	Hypothetical
RL4603	putative aminotransferase	Intermediary metabolism
pRL80047	putative major facilitator superfamily (MFS) transporter transmembrane component	Transporters

pRL90110	putative ATP-binding component of ABC transporter CUT1	Transporters
pRL100005	conserved hypothetical protein with PIN domain	Hypothetical
pRL100170 (*rhiB*)	rhizosphere induced protein RhiB	Nitrogen metabolism
pRL110031	putative LacI family transcriptional regulator (repressor)	Regulation
pRL110338	hypothetical protein	Hypothetical
pRL110545 (*minD*)	putative septum site-determining protein MinD	Cell division
pRL110556 (*glxC*)	putative glutamate synthase subunit	Nitrogen metabolism
pRL120618	putative LacI family transcriptional regulator (repressor)	Regulation
pRL120641	conserved hypothetical protein	Hypothetical
Rhizosphere-progressive		
RL0022 (*trpA*)	putative tryptophan synthase alpha subunit	Amino acid biosynthesis
RL0380 (*rph*)	putative ribonuclease	Nucleic acid replication/repair/synthesis
RL0571	conserved hypothetical protein	Hypothetical
RL0572 (*pyrD*)	putative dihydroorotate dehydrogenase§	Nucleic acid replication/repair/synthesis
RL0816	conserved hypothetical protein	Hypothetical
RL0822	putative DegT family aminotransferase	Intermediary metabolism
RL0847 (*guaB*)	putative inosine'-5'-monophosphate dehydrogenase	Nucleic acid replication/repair/synthesis

RL1434 (*feuQ*)	putative two-component sensor/regulator; histidine kinase	Regulation
RL1546 (*purF*)	putative amidophosphoribosyltransferase	Nucleic acid replication/repair/synthesis
RL1584 (*moaE*)	putative molybdopterin converting factor subunit E	Protein biosynthesis/modification
RL1596 (*purM*)	putative phosphoribosylformylglycinamidine cyclo-ligase	Nucleic acid replication/repair/synthesis
RL1622 (*ribC*)	putative riboflavin biosynthesis protein	Electron transfer
RL1738 (*pyrC2*)	putative dihydroorotase	Nucleic acid replication/repair/synthesis
RL1739 (*pyrB*)	putative aspartate carbamoyltransferase	Nucleic acid replication/repair/synthesis
RL1803 (*ilvD1*)	putative dihydroxy-acid dehydratase	Amino acid biosynthesis
RL2435 (*copB*)	putative TolC efflux protein, copper tolerence protein, CopB	Transporters
RL_RS12565	hypothetical protein	Hypothetical
RL2606 (*purC1*)	putative phosphoribosylaminoimidazole-succinocarboxamide synthase (saicar synthetase)	Nucleic acid replication/repair/synthesis
RL2612 (*purL*)	putative phosphoribosylformylglycinamidine synthase II	Nucleic acid replication/repair/synthesis
RL3205 (*ilvC*)	putative ketol-acid reductoisomerase	Amino acid biosynthesis
RL3245 (*ilvI*)	putative acetolactate synthase isozyme III large subunit	Amino acid biosynthesis
RL3513 (*leuA1*)	putative 2-isopropylmalate synthase	Amino acid biosynthesis

RL3768 (*purA*)	putative adenylosuccinate synthetase	Nucleic acid replication/repair/synthesis
RL4006 (*cbbT*)	putative transketolase	Aerobic respiration/TCA cycle
RL4011 (*pgk*)	putative phosphoglycerate kinase	Aerobic respiration/TCA cycle
RL4045 (*purK*)	putative phosphoribosylaminoimidazole carboxylase ATPase subunit	Nucleic acid replication/repair/synthesis
RL4281 (*hemK*)	putative protein methyltransferase	Protein biosynthesis/modification
RL4402	putative SBP of ABC transporter NitT	Transporters
RL4555	putative 3-isopropylmalate dehydratase	Amino acid biosynthesis
RL4705 (*leuD*)	putative 3-isopropylmalate dehydratase small subunit	Amino acid biosynthesis
pRL80128	conserved hypothetical protein	Hypothetical
pRL120205 (*eryB*)	putative erythritol phosphate dehydrogenase	Intermediary metabolism

Forty-one genes commonly required for rhizosphere growth of all five plant species and 32 genes are commonly rhizosphere-progressive. Genes are listed in Table S4.6 with their associated fitness values for each condition.

4.2.5.2. Common requirements for root colonisation of five plant species

The expanded list of 58 genes involved in root colonisation (Table 4.6) reveals that chemotaxis and motility are essential bacterial traits for efficient root colonisation. In addition to *flgB* and *cheS*, RL0712 (*flgI*), RL0704 (*flgF*), RL0699 (*flhB)* and RL0716 (*fliP*) were required for effective root colonisation. FlgI is involved in the formation of the membrane-embedded flagella basal body, assembling around the rod to form the P-ring which protects the motor/basal body from shearing forces during rotation (Liu and Ochman, 2007). FlgF forms the M-ring of the flagella basal body, whilst FlhB and FliP form part of the flagella export apparatus (Ward et al., 2018).

Properties of the bacterial cell surface influence adhesion to roots (Wheatley and Poole, 2018). A putative transmembrane protein encoded by *bacA* (RL3557), previously shown to be required for bacteroid development, was important for root colonisation of all plants (Karunakaran et al., 2010) (Table 4.6). Rlv3841 *bacA* mutants display impaired membrane function, with reduced incorporation of VLCFA into the lipid A component of LPS and altered peptide uptake (Karunakaran et al., 2010). The exact role of *bacA* is not clear, but lack of VLCFA incorporation into *R. leguminosarum* lipid A results in impaired membrane integrity and increased sensitivity to environmental stressors such as changes in osmolarity and the presence of antimicrobial peptides (Bourassa et al., 2017). Ability to adapt and tolerate such stresses at the rhizoplane surface is likely critical for successful colonisation and subsequent persistence. LPS also plays an important role in adhesion and therefore altered LPS structure may impede this function (Gao et al., 2001; Jofré et al., 2004; Balsanelli et al., 2010; Mitra et al., 2016). Other cell membrane components commonly required for root colonisation include RL2778, a putative EPS biosynthesis protein;

RL1499 (*ropA2*), an outer membrane porin that likely forms amyloid fibrils and may have a direct role in attachment to roots (Kosolapova et al., 2019); and RL0811A, a LPS glycosyltransferase (Table 4.6). Mutation of RL0811A results in reduced infection of pea nodules relative to wild type, though mutants have slightly reduced growth rates (Garcia-Fraile et al., 2015).

Once on the rhizoplane bacteria continue to grow forming microcolonies that subsequently develop into biofilms, in which bacterial cells are embedded in a matrix of EPS, aiding adherence to the rhizoplane (Branda et al., 2005; Flemming and Wingender, 2010). Factors involved in biosynthesis and metabolism therefore play a critical role during colonisation of the rhizoplane, in addition to those directly involved in adhesion to the root surface. A further six genes were implicated in the biosynthesis of amino acids or ribonucleotides: RL0020 (tryptophan), RL0139 (phenylalanine), RL4210 (cysteine), RL4707 (leucine), RL0747 (*purD*) and RL1620 (*glyA*) (Table 4.6). Cytochrome *c* biogenesis was also important, with the requirement of RL4540 (*cycY*), which has previously been found to contribute to Rlv3841 colonisation of pea and subsequent development of symbiosis (Wheatley et al., 2020).

Genes with putative roles in metabolism include RL3614 (*araD*), encoding a putative hydratase elevated by growth on L-arabinose and galactose (Ramachandran et al., 2011) and pRL110415 (*rhaD*), encoding a putative short-chain dehydrogenase/oxidoreductase (Table 4.6). *R. leguminosarum rhaD* mutants are unable to grow on rhamnose as a sole carbon source, show rhamnose toxicity when grown on glycerol-rhamnose due to a build-up of toxic phosphorylated intermediates in the rhamnose catabolism pathway, and are impaired in competitive nodulation

(Richardson et al., 2004). A common requirement for *rhaD* implies that rhamnose is present in the root exudates of all five plant species and that the ability to catabolise it provides a competitive advantage during root colonisation.

RL4400 was the only transport-related gene commonly required for root colonisation. It is part of the same transport system as RL4402, identified to have a rhizosphere-progressive phenotype for all plant species (Table 4.5), involved in the transport of the vitamin B_1 precursor HMP. A previous INSeq screen identified this transporter (RL4400-02) as important during all stages of Rlv3841 symbiosis with pea (Wheatley et al., 2020).

Genes involved in adaption during colonisation of the rhizoplane include a cold shock protein (*cspA4*; RL3449), a copper tolerance protein (*cusF*; RL2432) and a putative dioxygenase (RL0962), the latter of which was previously identified in an INSeq screen as essential during Rlv3841-pea symbiosis for growth of nodule bacteria and bacteroid development (Wheatley et al., 2020) (Table 4.6). These results indicate that RL0962 may have a more general role in defence, namely the degradation of toxic aromatic compounds, earlier in symbiosis during root colonisation. Cold shock proteins play a wide role in stress tolerance and can contribute to osmotic, oxidative, starvation and pH stress tolerance (Keto-Timonen et al., 2016).

Seven transcriptional regulators were commonly required for root colonisation across all five plant species (Table 4.6). These include PhoU (RL0546), a putative negative regulator of phosphate transport; AsfR (RL3021), an anti-sigma factor; RL0149, an XRE family transcriptional regulator; and RL3206, a TetR family transcriptional

regulator predicted to regulate transcription of the adjacent gene RL3205 (*ilvC*). RL3205 is involved in leucine biosynthesis and accordingly was found to be rhizosphere-progressive for all plant species. PhoU has been best characterised in *E. coli* where it has roles in phosphate homeostasis, stress responses and bacterial virulence (Santos-Beneit, 2015), though its function in Rlv3841 is yet to be characterised. Mutation of *asfR* in Rlv3841 results in the upregulation of 16 genes including the extracytoplasmic function (ECF) sigma factor *ecfR,* thought to regulate the expression of genes involved in oxidative stress (Stanger, 2008). However, none of the 16 genes showed a clear phenotype for the colonisation of all plant species (Table S4.7). RL0149 is induced 11-fold in *praR* mutants, which show a hyper attachment phenotype (Frederix et al., 2014). PraR interacts with the promoter of RL0149, likely repressing its expression, however, the function of RL0149 remains elusive since RL0149 mutants showed no major changes in gene expression or biofilm formation (Frederix et al., 2014). RL0149 mutants are however impaired in nodule infection (Wheatley et al., 2020). Identification of a clear dependency for RL0149 during the colonisation of all five plant species indicates that the symbiosis defective phenotype of RL0149 mutants may in part be due to impaired root colonisation, though the mechanism through which this occurs is unknown.

Table 4.6. 58 genes required for colonisation of pea, lentil, *Lathyrus*, alfalfa and barley.

Gene	Description	Function
RL0020	putative N-(5'-phosphoribosyl)anthranilate	Amino acid biosynthesis
RL0130	putative transcriptional regulator/acetyltransferase	Regulation
RL0139	putative P-protein [includes: chorismate mutase and prephenate dehydrogenase]	Amino acid biosynthesis
RL0149	putative XRE family transcriptional regulator, up reg in RL0390 (PraR) mutant MA	Regulation
RL0257	putative acetyltransferase	Intermediary metabolism
RL0378	putative HAM1 family protein	Not classified
RL0504 (*pgi*)	putative glucose-6-phosphate isomerase	Aerobic respiration/TCA cycle
RL0546 (*phoU*)	putative phosphate uptake regulator PhoU, unknown mechanism to regulate expression of high-affinity ABC systems	Regulation
RL0686 (*cheX1/cheS*)	putative chemotaxis related CheX protein	Motility/chemotaxis
RL0699 (*flhB*)	putative flagellar biosynthetic protein	Motility/chemotaxis
RL0704 (*flgF*)	putative flagella-associated protein FlgF	Motility/chemotaxis
RL0707 (*flgB*)	putative flagellar basal-body rod protein FlgB	Motility/chemotaxis
RL0712 (*flgI*)	putative flagellar basal body P-ring protein precursor FlgI	Motility/chemotaxis
RL0716 (*fliP*)	putative flagellar biosynthetic protein FliP	Motility/chemotaxis
RL0722	conserved hypothetical protein	Hypothetical

RL0811A	putative glycosyltransferase	Cell membrane/envelope
RL0939	methyltransferase	Hypothetical
RL0947 (*purD*)	putative phosphoribosylamine--glycine ligase	Nucleic acid replication/repair/synthesis
RL0957	conserved hypothetical protein	Hypothetical
RL0962	putative ring hydroxylating dioxygenase subunit	Adaptation
RL1499 (*ropA2*)	putative outer membrane porin protein RopA	Cell membrane/envelope
RL1620 (*glyA*)	putative serine hydroxymethyltransferase	Nucleic acid replication/repair/synthesis
RL2219	ATP-binding protein	Hypothetical
RL2239 (*eno*)	putative enolase	Aerobic respiration/TCA cycle
RL2289	conserved hypothetical protein	Hypothetical
RL2432 (*cusF*)	putative periplasmic copper binding protein, CusF in E. coli involved in Cu and Ag resistance	Hypothetical
RL2492 (*ppiD*)	putative peptidyl-prolyl cis-trans isomerase D	Protein biosynthesis/modification
RL_RS13445	phosphoribosylformylglycinamidine synthase	Not classified
RL2609	conserved hypothetical exported protein	Hypothetical
RL2778	putative exopolysaccharide biosynthesis protein	Cell membrane/envelope
RL3021 (*asfR*)	putative transmembrane anti-sigma factor DUF1109 (6 membrane-spanning domains, typical of ECF16)	Regulation
RL3113	cupin	Hypothetical

RL3206	putative TetR family transcriptional regulator	Regulation
RL3260	conserved hypothetical protein	Hypothetical
RL3339	putative ribonuclease-L-PSP family protein	Not classified
RL3449 (*cspA4*)	putative cold shock protein	Adaptation
RL3557 (*bacA*)	putative transmembrane transporter protein required for bacteroid development	Nitrogen metabolism
RL3614 (*araD*)	putative hydratase, has hydrolase (aligned to COG3802) (elevated on MA L-arabinose and galactose Rlv3841)	Intermediary metabolism
RL3763	putative rRNA methyltransferase	Nucleic acid replication/repair/synthesis
RL4066	conserved hypothetical protein	Hypothetical
RL4147	creatininase	Hypothetical
RL_RS21450	hypothetical protein	Hypothetical
RL4210 (*cycZ*)	putative cysteine biosynthesis protein	Amino acid biosynthesis
RL4400	putative permease component of ABC transporter NitT	Transporters
RL4540 (*cycY*)	putative thiol:disulfide interchange protein CycY precursor (cytochrome c biogenesis protein)	Protein biosynthesis/modification
RL4542 (*ispZ*)	putative intracellular septation protein	Cell division
RL4640 (*ndvA*)	putative ATP-binding:permease (ABC:IMP) component of ABC transporter Export beta-(1-->2) glucan export	Transporters
RL_RS23920	hypothetical protein	Hypothetical
RL4707 (*leuB*)	putative 3-isopropylmalate dehydrogenase	Amino acid biosynthesis

RL_RS25920	hypothetical protein	Hypothetical
pRL90210	putative short-chain dehydrogenase	Intermediary metabolism
pRL100447	putative LysR family transcriptional regulator	Regulation
pRL110149	putative isomerase	Intermediary metabolism
pRL110224	putative two-component sensor/regulator; transcriptional regulator	Regulation
pRL110415 (*rhaD*)	putative short-chain dehydrogenase/oxidoreductase involved in competition for nodulation	Intermediary metabolism
pRL110433	spermidine synthase	Hypothetical
pRL120210 (*rpiB2*)	putative ribose-5-phosphate isomerase B	Aerobic respiration/TCA cycle
pRL120297	putative epoxide hydrolase	Intermediary metabolism

Genes are listed in Table S4.6 with their associated fitness values for each condition.

4.2.5.3. Analysis of central carbon metabolism in the rhizosphere and during root colonisation

The requirement of genes involved in the central carbon pathways of Rlv3841 were analysed for rhizosphere growth and root colonisation across all five-plant species (Fig. 4.8.). As discussed in Chapter 3, transcriptome analyses have demonstrated clear induction of Rlv3841 genes involved in the glycolysis/gluconeogenesis pathways during growth in plant rhizospheres (Ramachandran et al., 2011). Using BarSeq we were able to identify numerous enzymes belonging to these pathways as important for Rlv3841 growth in the pea rhizosphere, including *eno* (RL2239), *pfp* (RL3322), *tpiA* (RL2513), *gpmA* (RL0179) and Pgk (RL4401). Pgk was found to be essential for rhizosphere growth and root colonisation across all five-plant species indicating that glycolysis/gluconeogenesis is a general requirement for rhizosphere fitness and subsequent root colonisation (Fig. 4.8 and Table 4.5). Two other enzymes belonging to this pathway, Pgi and Eno, were also commonly required for root colonisation (Fig. 4.8 and Table 4.6).

As previously mentioned, RL4006 (*cbbT*), encoding a putative transketolase involved in the non-oxidative branch of the pentose phosphate pathway (PPP), was commonly required for rhizosphere growth and root colonisation (Fig. 4.8 and Table 4.5). The PPP plays a central role in maintenance of carbon homeostasis and is divided into two enzymatic branches, namely the oxidative branch and the nonoxidative branch. Transketolase is a major enzyme involved in the nonoxidative branch and is required for the production of the RNA backbone precursor ribose 5-phosphate and the aromatic amino acid precursor erythrose 4-phosphate (Stincone et al., 2015). In *E. coli*, two transketolases encoded by *tktA* and *tktB* are involved in the PPP. TktA

accounts for 70-90% of the transketolase activity with *tktA* mutants unable to grow on pentose sugars, whilst TktB plays a minor role, with mutants showing no visible phenotype when grown on pentose sugars (Iida et al., 1993). The Rlv3841 genome contains four genes encoding putative transketolases: RL2718, RL2719, RL4006 (*cbbT*) and pRL100453. Only *cbbT* was required for rhizosphere growth and root colonisation of all five-plant species. This gene has also previously been shown to be essential for Rlv3841-pea nodule infection (Wheatley et al., 2020). Comparison of protein sequences revealed CbbT to share the greatest homology with TktA (51.61%) and pRL100453 to share the greatest homology with TktB (50.46%). RL2718 and RL2719 showed less than 35% homology to TktA or TktB. This, together with the growth impairment phenotype for *cbbT* mutants in all plant rhizospheres, suggests that *cbbT* is a homologue of *tktA* and the primary transketolase involved in the nonoxidative branch of the PPP in Rlv3841. In addition, pRL100453 is likely a homologue of *tktB* possibly playing a minor role in the PPP, with its transketolase activity not sufficient to compensate for lack of CbbT. This hypothesis is further supported by the fact that CbbT shows 76.86% homology to Tkt2*,* the primary transketolase of the PPP in *S. meliloti*, identified as a homologue of *E. coli* TktA (Hawkins et al., 2017). *S. meliloti tkt2* mutants are growth-impaired on erythritol and ribose, auxotrophic for aromatic amino acids and unable to form effective symbioses with their host legumes; the latter of which was strongly correlated with their inability to produce erythrose-4-phosphate (Hawkins et al., 2017). Erythrose-4-phosphate is required for chorismate production, the common first step in the biosynthesis of the three aromatic amino acids: tyrosine, phenylalanine and tryptophan (Tzin and Galili, 2010). These results suggest that *cbbT* plays a fundamental role in Rlv3841 rhizosphere fitness, potentially by allowing growth

on a wider variety of carbon sources/pentose sugars, such as erythritol and ribose and/or through production of aromatic amino acids.

Another gene encoding an enzyme of the PPP (non-oxidative branch), pRL120210 (*rpiB2*), was required for the colonisation of all five-plant species and growth in the rhizosphere of all but alfalfa (Fig. 4.8 and Table 4.6). RpiB2 has a central role in carbon metabolism and is required for erythritol catabolism with *rpiB2* mutants unable to utilise erythritol as a sole carbon source (Yost et al., 2006).

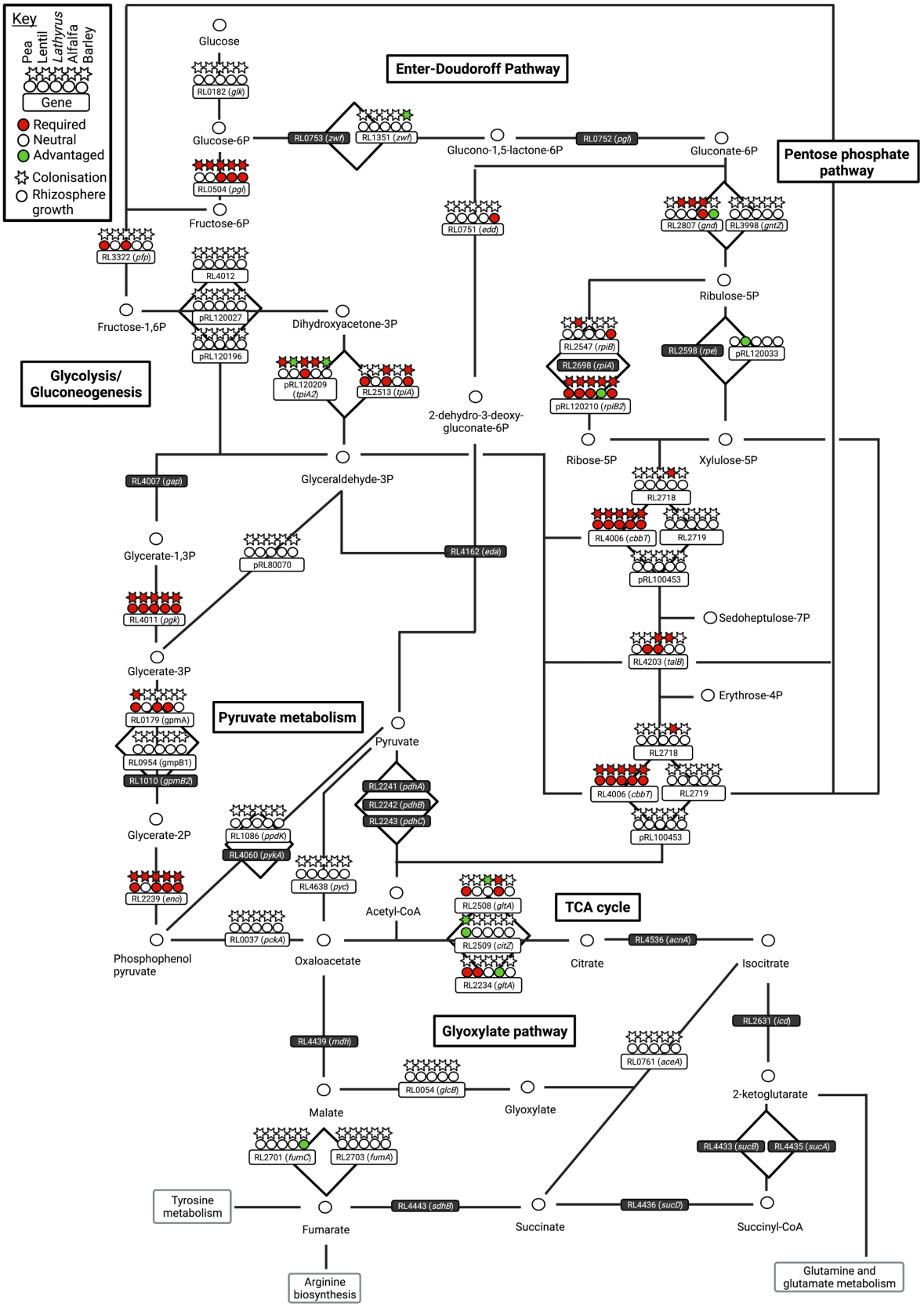

Key
Pea
Lentil
Lathyrus
Alfalfa
Barley
Gene
Required
Neutral
Advantaged
Colonisation
Rhizosphere growth
Glucose
RL0182 (glk)
Glucose-6P
RL0753 (zwf)
RL1351 (zwf)
Glucono-1,5-lactone-6P
RL0752 (pgl)
Gluconate-6P
Enter-Doudoroff Pathway
Pentose phosphate pathway
RL0504 (pgi)
Fructose-6P
RL0751 (edd)
RL2807 (gnd)
RL3998 (gntZ)
RL3322 (pfp)
RL4012
pRL120027
pRL120196
Fructose-1,6P
Dihydroxyacetone-3P
Ribulose-5P
RL2547 (rpiB)
RL2698 (rpiA)
pRL120210 (rpiBZ)
RL2598 (rpe)
pRL120033
Glycolysis/ Gluconeogenesis
pRL120209 (tpiAZ)
RL2513 (tpiA)
2-dehydro-3-deoxy-gluconate-6P
Ribose-5P
Xylulose-5P
RL4007 (gap)
Glyceraldehyde-3P
RL2718
RL4006 (cbbT)
RL2719
pRL100453
Glycerate-1,3P
pRL80070
RL4162 (eda)
Sedoheptulose-7P
RL4011 (pgk)
RL4203 (talB)
Glycerate-3P
Erythrose-4P
RL0179 (gpmA)
RL0954 (gmpB1)
RL1010 (gpmBZ)
Pyruvate metabolism
Pyruvate
RL2718
RL4006 (cbbT)
RL2719
pRL100453
Glycerate-2P
RL2241 (pdhA)
RL2242 (pdhB)
RL2243 (pdhC)
RL1086 (ppdK)
RL4060 (pykA)
RL2239 (eno)
RL4638 (pyc)
Acetyl-CoA
RL2508 (gltA)
TCA cycle
RL2509 (citZ)
RL2234 (gltA)
RL0037 (pckA)
RL4536 (acnA)
Phosphophenol pyruvate
Oxaloacetate
Citrate
Isocitrate
RL4439 (mdh)
Glyoxylate pathway
RL2631 (icd)
RL0761 (aceA)
RL0054 (glcB)
2-ketoglutarate
Malate
Glyoxylate
RL2701 (fumC)
RL2703 (fumA)
RL4433 (sucB)
RL4435 (sucA)
Tyrosine metabolism
RL4443 (sdhB)
RL4436 (sucD)
Fumarate
Succinate
Succinyl-CoA
Arginine biosynthesis
Glutamine and glutamate metabolism

Figure 4.8. Central carbon metabolism pathways for Rlv3841 during rhizosphere growth and root colonisation of pea, lentil *Lathyrus*, alfalfa and barley. Candidate genes for enzymes performing the catalytic steps are shown in boxes. Genes lacking fitness values due to essentiality for growth even in rich media are shown in black boxes. Mutants with fitness value less than -1 are shown in red (or < -0.42 if identified as commonly required among all five plant species), fitness values between -1 and 1 are shown in white and fitness values greater than 1 are shown in green.

4.2.6. Plant-specific factors required for rhizosphere growth and root colonisation

To identify genes with plant-specific roles during rhizosphere growth and/or root colonisation, genes with fitness values less than -1 for each plant species during a) rhizosphere growth and b) root colonisation (Fig. 4.4) were independently filtered to remove any with fitness values less than -0.42 for either rhizosphere growth or root colonisation of the other four-plant species. Whilst genes with fitness values between 1 and -1 are generally considered to have relatively neutral phenotypes, a filter of -0.42 was applied to ensure that genes selected are truly plant-specific since a fitness value of -0.42 represents a 25% reduction in growth relative to the average mutant in the library. Therefore, any gene identified as plant-specific for rhizosphere growth or root colonisation will have a fitness value less than -1 for that plant and a fitness value greater than -0.42 for all other plant species. These genes were then further filtered based on their associated *t*-score, a measure of consistency in the strain fitness values used to calculate the gene fitness value, to retain those which have high confidence in the associated fitness value. Genes with fitness values less than -2 were filtered to retain those with an associated *t*-score < -3 and those with fitness values between -1

and –2 were filtered to retain those with a *t*-score < -2.5 (described in Chapter 2.5.7). For the rhizosphere, alfalfa had the highest requirement for plant-specific genes (76 genes; Table S4.11), followed by barley (53; Table S4.12), pea (54; Table S4.8), lentil (20; Table S4.9) and *Lathyrus* (18; Table S4.10). For root colonisation, *Lathyrus* had the highest demand for plant-specific requirements (70 genes; Table S4.10), followed by alfalfa (58; Table S4.11), lentil (46; Table S4.9), barley (25; Table S4.12) and pea (22; Table S4.8) (Fig. 4.9).

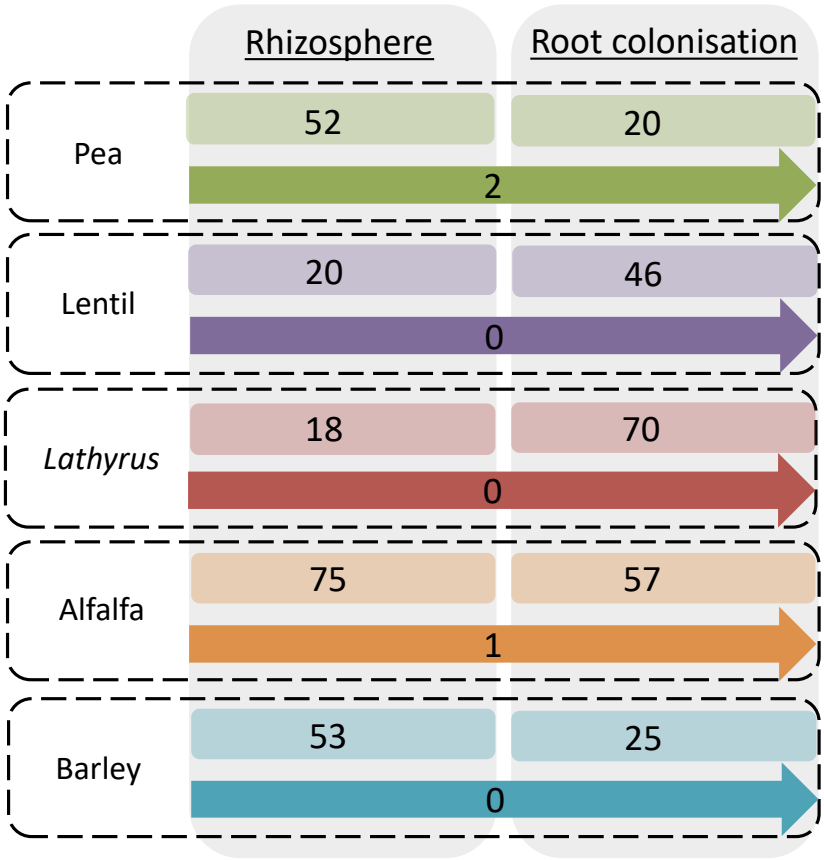


Figure 4.9. Number of plant-specific genes required for rhizosphere growth and root colonisation for each plant species by RIv3841. Genes are classed as plant-specific if they have a fitness value less than -2 (*t*-score < -3) or fitness value between -1 and -2 (*t*-score < -2.5) for one plant species, but greater than -0.42 for all other plant

species. The number of plant-specific rhizosphere-progressive genes are shown on the arrow. These genes are listed in Tables S4.8-S4.12.

Functional classification of plant-specific genes required for rhizosphere growth (including rhizosphere-progressive genes) (Fig. 4.10A) and root colonisation (Fig. 4.10B) revealed that hypothetical proteins account for the largest number of plant-specific genes for all species except pea, where it forms the second largest category after genes involved in transport. After hypothetical proteins, plant-specific genes were predominantly involved in either transport, regulation or intermediary metabolism. There are, however, clear differences in functional requirements among the plant species, for example, growth in the *Lathyrus* rhizosphere required many genes involved in cell envelope functions accounting for 17% of *Lathyrus* rhizosphere-specific genes. In the pea rhizosphere, there was a high demand for transporters accounting for 28% of pea-specific genes, compared to just 10% for lentil. Barley had the highest demand for genes involved in adaption at 8%, whereas no genes belonging to this functional category were identified specifically for growth in the rhizosphere of the legumes pea, lentil, and alfalfa. A similar trend was seen for root colonisation, with hypothetical proteins, those involved in transport, regulation, and intermediary metabolism accounting for the largest functional groups for each plant species.

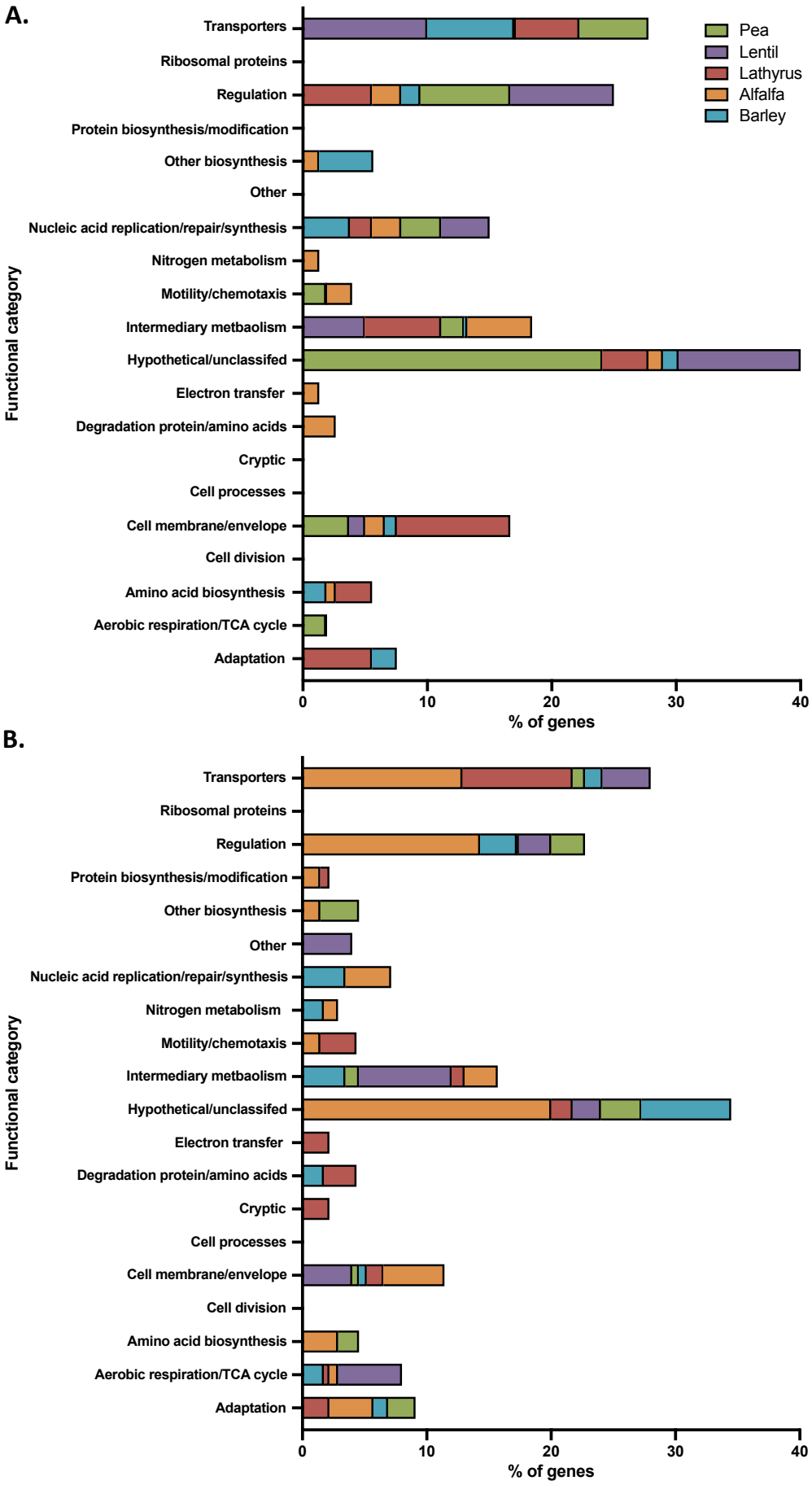

A.
Transporters
Ribosomal proteins
Regulation
Protein biosynthesis/modification
Other biosynthesis
Other
Nucleic acid replication/repair/synthesis
Nitrogen metabolism
Motility/chemotaxis
Intermediary metbaolism
Hypothetical/unclassifed
Electron transfer
Degradation protein/amino acids
Cryptic
Cell processes
Cell membrane/envelope
Cell division
Amino acid biosynthesis
Aerobic respiration/TCA cycle
Adaptation
Functional category
Pea
Lentil
Lathyrus
Alfalfa
Barley
0
10
20
30
40
% of genes
B.
Transporters
Ribosomal proteins
Regulation
Protein biosynthesis/modification
Other biosynthesis
Other
Nucleic acid replication/repair/synthesis
Nitrogen metabolism
Motility/chemotaxis
Intermediary metbaolism
Hypothetical/unclassifed
Electron transfer
Degradation protein/amino acids
Cryptic
Cell processes
Cell membrane/envelope
Cell division
Amino acid biosynthesis
Aerobic respiration/TCA cycle
Adaptation
Functional category
0
10
20
30
40
% of genes

Figure 4.10. Functional classification of plant-specific genes required for rhizosphere growth. A. Rhizosphere growth. **B.** Root colonisation. Functional classification is based on Riley codes (Young et al., 2006). These genes are listed in Tables S4.8-S4.12.

4.2.6.1. Plant-specific factors contributing to rhizosphere fitness

Root exudate composition varies among plant species, and it is therefore not surprising that numerous plant-specific genes required for rhizosphere fitness are implicated in nutrient uptake from the rhizosphere (Fig. 4.11, Tables S4.13). The demand for specific transporters varies among plant species with 28% of pea-specific genes implicated in transport compared to just 10% for lentil-specific genes (Fig. 4.10A). Plant-specific transporters with known substrates include RL4583 (*fbpB*), a ferric cation transporter and RL3546 (*oppD*), involved in the transport of oligopeptides required for pea rhizosphere fitness. A SBP (pRL120752) involved in phosphonate uptake and RL4229 (*araH*), which shares 98% homology *R. etli araH* involved in arabinose uptake, were required specifically in the lentil rhizosphere. In the alfalfa rhizosphere, Rlv3841 required the transport of phosphate and potassium via pRL100343 (*pit2*) and pRL110379 (*kdpC*) (Pedersen et al., 2019), respectively, in addition to pRL90234, a carbohydrate transporter (CUT1 family) previously shown not to be important for growth in the pea rhizosphere (Garcia-Fraile et al., 2015). VsbD (pRL120317), a multidrug and toxic compound efflux transporter putatively involved in vicibactin siderophore export, was required for growth in the barley rhizosphere, along with pRL120515 (*qatW3*), involved in quaternary amine uptake (Fox et al., 2008) and RL2000, a tripartite ATP-independent periplasmic (TRAP) C4-dicarboxylate transporter. Under iron limitation, bacteria secrete siderophores to scavenge iron from the environment. *In R. leguminosarum* expression of *vbsD* is increased under low iron conditions and VbsD is predicted to have a role in vicibactin siderophore export suggesting iron limitation in the barley rhizosphere (Carter et al., 2002). Whilst *vbsOSA* (pRL120314-16), required for siderophore biosynthesis in Rlv3841, had neutral phenotypes for growth in the barley rhizosphere, other mutants in the population are

capable of siderophore production and therefore likely compensate for this. Identification of plant-specific transporters reveals the molecular basis of the adaptive ability of Rlv3841, with its genome encoding numerous transporters, for survival and growth in the multiple and varied plant rhizospheres it may encounter during its free-living lifestyle in soil.

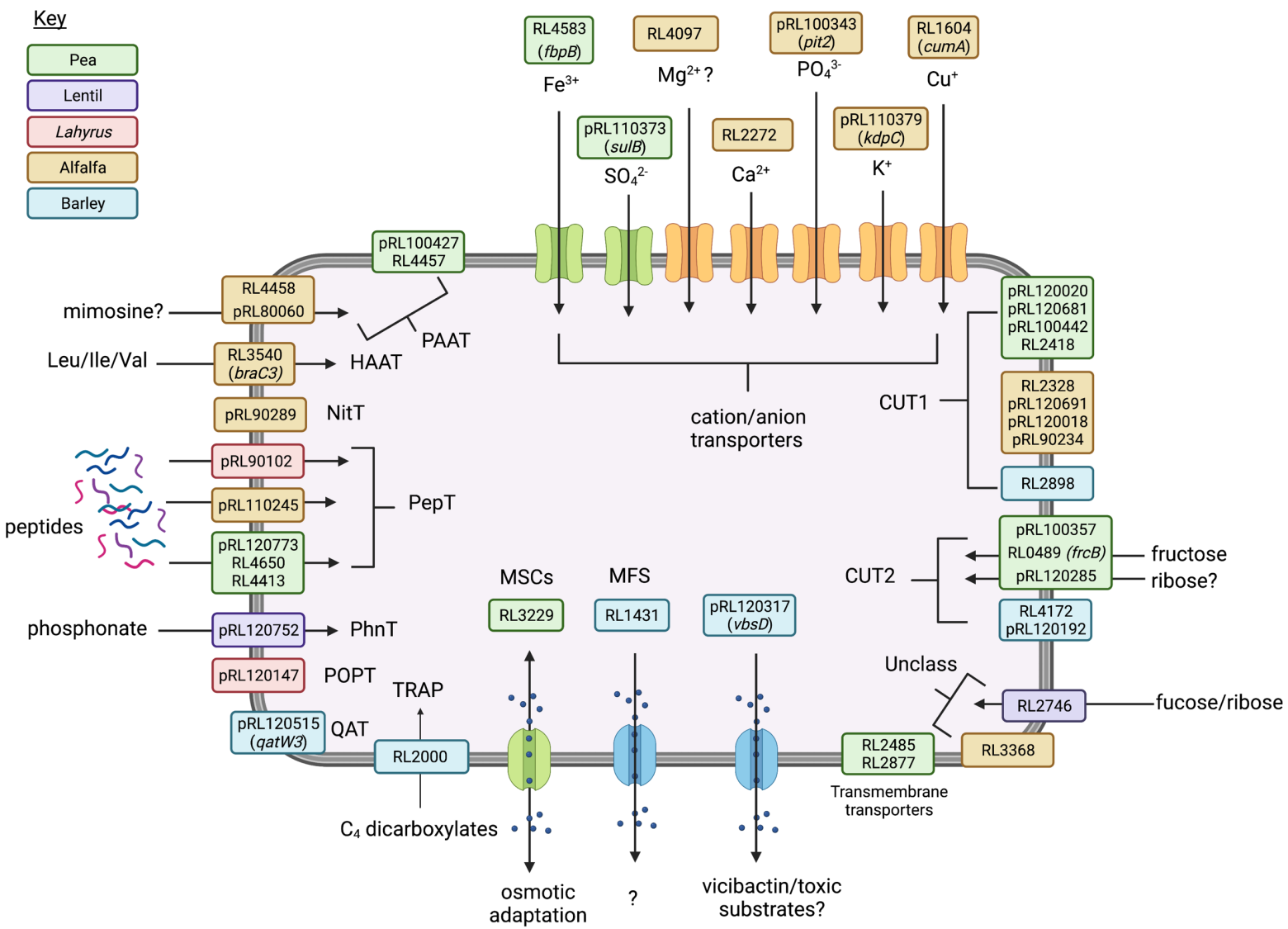

Key
Pea
Lentil
Lahyrus
Alfalfa
Barley
RL4583 (fbpB)
Fe3+
RL4097
Mg2+ ?
pRL100343 (pit2)
PO4 3-
RL1604 (cumA)
Cu+
pRL110373 (sulB)
SO4 2-
RL2272
Ca2+
pRL110379 (kdpC)
K+
pRL100427 RL4457
PAAT
RL4458 pRL80060
mimosine?
HAAT
RL3540 (braC3)
Leu/Ile/Val
pRL90289
NitT
pRL90102
pRL110245
pRL120773 RL4650 RL4413
PepT
peptides
pRL120752
PhnT
phosphonate
pRL120147
POPT
pRL120515 (qatW3)
QAT
TRAP
RL2000
C4 dicarboxylates
cation/anion transporters
pRL120020 pRL120681 pRL100442 RL2418
CUT1
RL2328 pRL120691 pRL120018 pRL90234
RL2898
pRL100357 RL0489 (frcB) pRL120285
CUT2
fructose
ribose?
RL4172 pRL120192
Unclass
RL2746
fucose/ribose
RL2485 RL2877
RL3368
Transmembrane transporters
MSCs
RL3229
MFS
RL1431
pRL120317 (vbsD)
osmotic adaptation
?
vicibactin/toxic substrates?

Figure 4.11. Rlv3841 transporters with plant-specific roles in rhizosphere fitness. All genes have a fitness value less than -2 (*t*-score < -3) or fitness value between -1 and -2 (*t*-score < -2.5) for either pea, lentil, *Lathyrus*, alfalfa of barley rhizosphere growth and a fitness value greater than -0.42 for growth in all other plant rhizospheres. For uptake ABC transporter family: CUT1, carbohydrate uptake transporter 1; CUT2, carbohydrate uptake transporter 2; HAAT, hydrophobic amino acid transporter; NitT, nitrate/nitrite/cyanate transporter; PAAT, polar amino acid transporter; PepT, peptide/opine/nickel transporter; POPT, polyamine/opine/phosphonate transporter; QAT, quaternary amine transporter; PhnT, phosphonate Uptake Transporter; TRAP, tripartite ATP-independent periplasmic; MSCs, mechanosensitive ion channel and MFS, major facilitator superfamily. All genes shown here are listed in Table S4.13.

Between 5% (lentil) and 18% (alfalfa) of plant-specific genes required for rhizosphere growth had functional roles in intermediary metabolism, with these genes likely reflecting variation in available resources among plant rhizospheres (Fig. 4.10A). Notable examples include pRL100434, encoding an allophanate hydrolase subunit that catalyses the hydrolysis of allophanate to ammonium and CO_2, contributing to pea rhizosphere fitness. Allophanate is an intermediate in urea and cyanuric acid metabolism but it is also predicted to be an intermediate in the catabolism of other nitrogen heterocyclic rings which are common in natural compounds inducing those found in plant root exudates (Cheng et al., 2005; da Silva Lima et al., 2014; Sasse et al., 2020). From INSeq results, pRL100434 was found to be essential for nodule symbiosis with pea, but not required for rhizosphere growth and root colonisation (Wheatley et al., 2020). However, the results from this study suggest that pRL100434 is initially important for growth in the pea rhizosphere. This discrepancy may be accounted for by the fact that the HMM was used for INSeq analysis. In this instance the HMM classes genes as important for growth if they have less than 1% of the mean read count. In contrast, BarSeq analysis classes a gene as important for fitness if it has 50% less reads relative to time zero. RL4384 (*arcA*), a putative arginase that splits arginine into ornithine and urea as part of the urea cycle, and more generally, nitrogen metabolism, was important for *Lathyrus* rhizosphere fitness. RL4123 (*soxA1*) a sarcosine oxidase subunit was important in the barley rhizosphere. Sarcosine oxidases catalyse the conversion of sarcosine to glycine. The ability to metabolise sarcosine, most likely originating from barley root exudates, could be important for enabling utilisation of sarcosine as a carbon and/or nitrogen source in the barley rhizosphere.

Growth in the barley rhizosphere required the largest proportion of plant-specific genes with functional roles in adaption (8%) (Fig. 4.10A). These include RL4013, a putative phenazine biosynthesis protein; RL1686, a putative cytochrome P450; and RL3911 encoding the *Rhizobium*-adhering protein RapB (Table S4.12). The exact function of these genes in rhizosphere fitness is unclear; however, phenazines have roles in various processes from biofilm formation to modification of cellular redox status, thereby promoting bacterial survival (Pierson and Pierson, 2010), cytochrome P450's function in various metabolic pathways (Kelly and Kelly, 2013), and Raps promote bacterial aggregation through interaction with EPS, thereby promoting biofilm development, a known survival mechanism (Abdian et al., 2013). No genes with a functional classification of adaption were exclusively required for growth in pea, lentil, or alfalfa rhizospheres.

Growth in the lentil rhizosphere had the highest requirement for plant-specific transcriptional regulators accounting for 25% of lentil-specific genes, followed by pea (17%), barley (9%), alfalfa (8%) and *Lathyrus* (6%) (Fig. 4.10A). Of the five transcriptional regulators required for lentil rhizosphere fitness, only two have predicted functions. These are RL1995, a LacI family transcription repressor, and RL3235 (*ecfE*), an RNA polymerase ECF sigma factor (Table 4.7). RL1995 is predicted to regulate the adjacent carbohydrate transporter encoded by RL1996-99 (CUT2 family) (Novichkov et al., 2013). EcfE has previously been identified as a master regulator of Rlv3841 pea root attachment in response to a signal derived from pea root hairs (Vinoy Ramachandran, personal communication) though, previous INSeq screens have failed to identify a phenotype for EcfE during symbiosis with pea (Parsons, 2019; Wheatley et al., 2020) (Table 4.7). RL1473 (*ecfD*) a putative RNA

polymerase ECF sigma factor, and RL0789 (*asfB*), the anti-sigma factor of EcfB (RL0788), were required for optimal fitness in the pea rhizosphere (Table 4.7). ECF sigma factors regulate numerous cellular processes in response to external stimuli by coordinating gene transcription, particularly during various stress responses. Anti-sigma factors are antagonistic to their often co-transcribed sigma factor, binding the sigma factor and preventing it from interacting with RNA polymerase (Helmann, 2002). Whilst the targets of EcfB and EcfD are unknown, the genomic organisation of *ecfD* is like that seen in *S. meliloti* RpoE10 and *P. putida* ECF10, the latter of which responds to antibiotic resistance and biofilm formation (Tettmann et al., 2014; Lang et al., 2018). In the barley rhizosphere, both GlpR (pRL90073), a repressor of the *glp* operon required for the active uptake and metabolism of glycerol (Ding et al., 2012), and GabR (RL0103), a negative regulator of *gabT* (RL0102), a γ-aminobutyrate (GABA) aminotransferase involved in GABA metabolism (Prell et al., 2002), were important (Table 4.7). Whilst *R. leguminosarum glpR* mutants can actively transport glycerol and utilise it as a sole carbon source, the *glp* operon is constitutively expressed at high levels suggesting that proper regulation of glycerol transport and/or metabolism is key for Rlv3841 fitness in the barley rhizosphere (Ding et al., 2012). NolR (RL3537) and VisN (RL0696) were important for growth in the alfalfa rhizosphere (Table 4.7). In *R. leguminosarum* bv. *viciae* strain TOM, NolR acts as a transcriptional repressor of *nod*-gene expression following induction by NodD, suggesting that repression of Nod factor biosynthesis is important for growth in the rhizosphere of alfalfa and maybe non-host legumes in general (Kiss et al., 1998). Interestingly, *Rlv* has been shown to respond to flavonoids exuded by alfalfa, with alfalfa-derived flavonoids inducing the expression of a subset of *nod* genes in *Rlv* (Long, 1996). NolR has also been shown to function as a global regulator in *Sinorhizobium* where it alters the expression of genes involved

in the TCA cycle, oxidative stress and cell growth, among other processes, in response to environmental signals (Chen et al., 2000). VisN functions as a global transcriptional activator of flagella, motility and chemotaxis genes and Rlv4841 *visN* mutants are non-motile (Sourjik et al., 2000; Tambalo et al., 2010). The requirement of VisN only in the alfalfa rhizosphere is unclear, but may be related to the fact that alfalfa has the smallest rooting system of the five plant species investigated and so whilst 10 seedlings were planted per tube, perhaps motility was more important for movement towards the roots and for access to root exudates.

Since the targets of these transcriptional regulators are predominantly unknown, their phenotypes from previous INSeq screens with Rlv3841 were considered. These experiments were: 1) during growth on glucose or succinate at 1% and 21% oxygen (Wheatley et al., 2017); 2) during symbiosis with pea (Wheatley et al., 2020); and 3) during attachment to pea roots under alkaline, neutral, and acidic pH conditions (Parsons, 2019) (Table 4.7). Whilst mutation of most of these genes had no effect under these conditions (Table 4.7), the following genes encoding transcriptional regulators were important for growth on minimal media supplemented with glucose or succinate: pea-specific RL0905 (GntR family) and lentil-specific RL1995 (LacI family) (Wheatley et al., 2017). The lentil-specific pRL120695 (TetR family), alfalfa-specific regulator VisN (RL0696) and barley-specific RL2716 (DeoR family) were impaired for nodule symbiosis with (Wheatley et al., 2020). So, whilst these transcriptional regulators are not important for Rlv3841 fitness in the pea rhizosphere, some are required at later stages of symbiosis or during free-living growth.

Table 4.7. Phenotypes for the 26 transcriptional regulators demonstrating plant-specific dependencies for growth in the rhizosphere of pea, lentil, Lathyrus, alfalfa or barley were examined from INSeq screens assessing gene contributions during free-living growth, symbiosis with pea and attachment to pea roots at different pH's.

			INSeq HMM classifications											
			Free-living				Pea symbiosis					Pea attachment		
Plant	**Gene ID**	**Gene Description**	**G21**	**S21**	**G1**	**S1**	**In**	**Rh**	**Ro**	**Bc**	**No**	**6.5**	**7.0**	**7.5**
Pea	RL0485	transmembrane GGDEF/EAL sensory box protein	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL0759	LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL0789 (*asfB*)	transmembrane anti-sigma factor	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL0905	GntR family transcriptional regulator	GD	GD	GD	GD	NE	NE	NE	NE	NE	NE	NE	NE
	RL1209	LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL1473 (*ecfD*)	RNA polymerase ECF sigma factor	NE	NE	NE	NE	NE	NE	NE	NE	AD	-	-	-
	RL3112	ROK family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL3472	transmembrane GGDEF/EAL sensory box protein	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
Lentil	RL1204	LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL1995	LacI family transcriptional regulator (repressor)	GD	GD	ES	ES	NE	AD	NE	NE	NE	NE	NE	NE
	RL3235 (*ecfE*)	RNA polymerase ECF sigma factor	NE	NE	NE	NE	NE	NE	NE	AD	NE	NE	NE	NE
	pRL120077	GntR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	pRL120695	TetR family transcriptional regulator of RND efflux transporter	NE	NE	NE	NE	NE	NE	NE	GD	GD	NE	NE	NE
Lathyrus	pRL120600	GntR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
Alfalfa	RL0480	TetR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE

	RL0696 (*visN*)	LuxR/GerE family transcriptional regulator	NE	NE	NE	NE	AD	NE	NE	ES	ES	NE	NE	NE
	RL0774	LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL2913	GntR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL3537 (*nolR*)	Nodulation repressor	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	pRL110184	LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	pRL110528	GntR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
Barley	RL0103 (*gabR*)	MerR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	AD	NE	NE	NE
	RL0194	AsnC family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL0674	two-component sensor/regulator; transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL2716	DeoR family transcriptional regulator (repressor)	NE	NE	NE	NE	NE	NE	NE	NE	GD	NE	NE	NE
	pRL90073 (*glpR*)	DeoR family transcriptional regulator (repressor) regulator of glycerol uptake and metabolic genes	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE

INSeq experiments are for Rlv3841(i) Free-living, grown on glucose at 21% oxygen (G21), succinate 21% (S21), glucose 1% (G1) and succinate 1% (S1) (Wheatley et al., 2017); (ii) Pea symbiosis: input library (In), 5 dpi rhizosphere growth (Rh), 5 dpi root colonisation (Ro), bacteroid (Bc) and nodule bacteria (No) (Wheatley et al., 2020); and (iii) Pea attachment: 1-hour attachment to pea roots at pH 6.5, 7.0 and 7.5 (Parsons, 2019). Gene classification: essential (ES), growth defective (GD), neutral (NE) and growth-advantaged (AD). These genes are listed in Tables S4.8-S4.12.

The bacterial cell membrane provides a protective barrier against environmental stresses. The rhizospheres of different plant species are likely to present unique challenges and so unsurprisingly, several plant-specific genes had roles in modifying cell membrane function. For *Lathyrus* these include two genes with predicted roles in peptidoglycan biosynthesis, namely RL2627 (*murI*) and pRL110048 (*redAh*) (Table S4.10). MurI synthesises D-glutamate from L-glutamate, a component of peptidoglycan. Whilst the essentiality of *murI* has been demonstrated for several bacterial species (Li et al., 2014), this is not the case for Rlv3841, with mutation causing unaffected growth under various conditions, including the pea rhizosphere (Wheatley et al., 2017; Wheatley et al., 2020). Alfalfa rhizosphere fitness required RL2428 (*plsY2*) which encodes a glycerol-3-phosphate acyltransferase that catalyses the initial step in the assembly of membrane glycerolipids (Yang et al., 2010) (Table S4.11). This enzyme is essential for membrane biogenesis in numerous bacteria (Lu et al., 2006) however, the Rlv3841 genome contains at least one other putative glycerol-3-phosphate acyltransferase encoded by RL1737 (*plsY1*) and so the reason for the requirement for RL2428 solely in the alfalfa rhizosphere is unclear. In the barley rhizosphere, RL3508 (*lppF*) encoding a conserved exported hypothetical protein (possible lipoprotein) of unknown function was important (Table S4.12).

4.2.6.2. Plant-specific factors contributing to root colonisation

The genes with plant-specific roles in root colonisation are listed in Tables S4.8-S4.12. Motility and chemotaxis are prerequisites for root colonisation and therefore, not unexpectedly, six genes involved in these processes were identified as common determinants of root colonisation for all five plant species (Table 4.6). Only three genes with putative roles in these processes showed plant-specific phenotypes and these

were pRL120056 (*mcpR*), a methyl-accepting chemotaxis protein required for *Lathyrus* root colonisation; RL0210 (*pilQ*), encoding a putative pilus subunit transmembrane protein; and RL2683 (*mcpD*), a methyl-accepting chemotaxis protein both specifically required for lentil colonisation. The identification of McpD and McpR as uniquely required for lentil and *Lathyrus* colonisation, respectively, suggest that they recognise a chemoattractant present only in these root exudates.

Bacterial cell surface components such as EPS, CPS and LPS influence bacterial adhesion, root colonisation and host specificity during the establishment of *Rhizobium*-legume symbioses (Wang et al., 2012; Wheatley and Poole, 2018). A putative capsule polysaccharide export protein encoded by pRL120461 was specifically required for colonisation of pea. RL1624 (*rfbD*), encoding a dTDP-4-dehydrorhamnose reductase involved in the biosynthesis of dTDP-L-rhamnose, contributed to lentil colonisation. Rhamnose is a key constituent of the O-antigen of LPS and dTDP-L-rhamnose is its immediate precursor. Rhamnose-containing LPS is important for the establishment of several plant-microbe interactions, with mutants impaired in attachment, endophytic colonisation and nodulation (Jofré et al., 2004; Balsanelli et al., 2010; Mitra et al., 2016). However, this is not always the case since *rfbD* was found to be unimportant during Rlv3841 symbiosis with pea (Wheatley et al., 2020). RL1470, encoding a glycosyl transferase involved in lipid A GalA biosynthesis of LPS (Brown et al., 2013); RL3210, encoding a putative glucosyl transferase; and pRL90140, encoding a putative Ca^{2+}-binding hemolysin-like protein that is secreted by the Type I protein secretion system PrsDE (Krehenbrink and Downie, 2008) were important for *Lathyrus* root colonisation. Ca^{2+}-binding proteins have previously been documented to play a role in bacterial root adhesion (Smit et al., 1991; Downie, 2010). Rlv3841 colonisation of

alfalfa was influenced by RL1664 (*exoP*), encoding a polysaccharide transport protein (Williams et al., 2008); pRL110050 (*redA*), encoding a putative membrane-anchored protein with a likely role in peptidoglycan biosynthesis; and pRL90213, encoding a putative outer membrane protein of unknown function. Colonisation of barley did not specifically require any genes thought to have a direct influence on the bacterial cell surface.

While anchored to the rhizoplane, bacteria will need to tolerate various environmental stresses, such as the release of ROS by plant roots. Several plant-specific genes for root colonisation were identified as having putative roles in adaption. For pea these include RL1521 (*sodC*), encoding a putative superoxide dismutase required for the destruction of ROS possibly from plant origin (Ampe et al., 2003) and RL1333 (*gstA2*), encoding a putative glutathione S-transferase that quenches reactive molecules thereby protecting the cell from oxidative damage (Kumar and Trivedi, 2018) (Table S4.8). Rlv3841 also required specific glutathione S-transferases for the colonisation of *Lathyrus* (RL2759) and alfalfa (RL1348; *gstA4*) (Tables S4.10 and S4.11). A putative choline dehydrogenase, encoded by pRL110491, that plays a role in osmotic adaption through the oxidation of choline to betaine aldehyde was important for colonisation of *Lathyrus* roots (Table S4.10). Osmotic adaption was also important for colonisation of alfalfa which specifically required RL0096, encoding a putative choline/ethanolamine kinase. Additionally, genes involved in detoxification of aromatic compounds, namely pRL110087 (*pcaH1*), encoding a protocatechuate 3,4-dioxygenase; and pRL110085, encoding a putative cycloisomerase, were important for alfalfa colonisation (Table S4.11).

A total of 38 genes encoding transcriptional regulators were identified as playing plant-specific roles during root colonisation, though the regulatory targets are largely unknown (Table 4.8). AglR (RL0744), encoding a negative regulator of the adjacent alpha-glucoside (*agl*) utilisation operon (RL0745-49) was important for pea colonisation (Willis and Walker, 1999) (Table 4.8). PhyR (RL3705) and the zinc uptake regulator Zur (RL3175) were required for lentil colonisation (Table 4.8). Zur is a repressor of zinc uptake genes (RL3176-78), but has also been shown to regulate genes involved in amino acid transport in *B. subtilis* and virulence in *Xanthomonas campestris* (Kandari et al., 2021). The phyllosphere-induced regulator PhyR has been shown to be important for plant colonisation in gram negative bacteria. In *B. japonicum, phyR* mutants are impaired in nodulation, likely due to its role in regulating genes in response to certain stress and growth signal such as nutrient limitation, oxidative and hyperosmotic stress (Gourion et al., 2006; Liu et al., 2017b). For *Lathyrus* colonisation MatR (RL0989), encoding a repressor of *matABC* (RL0990-92), involved malonate uptake and conversion to acetyl-CoA (Lee et al., 2000); RL1980 (*asfL*), encoding an anti-sigma factor predicted to interact with the sigma factor *ecfL* (RL4524); and RL1730 (*celR2*), encoding a putative two-component sensor/regulator, were important (Table 4.8). CelR2 has been best characterised in *R. leguminosarum* bv. *trifolii,* where, upon phosphorylation, it catalyses the production of the second messenger cyclic dimeric guanosine monophosphate (c-di-GMP) which subsequently activates target genes including those involved in cellulose production (Ausmees et al., 1999). Cellulose fibrils play an important role in biofilm formation and appear to be induced upon close contact with the root surface, possibly with a plant-derived compound initiating cellulose synthesis via *celR2*. As with the plant-specific regulators involved in rhizosphere growth (Table 4.7), the INSeq phenotypes for those with plant-

specific roles in root colonisation were investigated. Mutation of most transcriptional regulators resulted in no effect during free-living growth, establishment of symbiosis with pea or attachment to pea roots (Table 4.8). RL0905, a pea-specific GntR family transcriptional regulator; RL3828, a lentil-specific FNR/CRP family transcriptional regulator; and lentil-specific PhyR (RL3705), all showed growth defects during free-living growth on minimal media supplemented with glucose or succinate at 21 and 1% oxygen. Mutation of these genes had no effect under any other condition, except for PhyR which was growth-defective for nodule bacteria (Wheatley et al., 2020) (Table 4.8). Interestingly, RL4145, a *Lathyrus*-specific LacI family transcriptional regulator was important for attachment to pea at pH 6.5-7.5, though showed no importance during symbiosis with pea (Parsons, 2019) (Table 4.8).

As with rhizosphere growth, plant-specific genes identified as important for root colonisation include those encoding components of various ABC transporters. For pea colonisation these include pRL90082, encoding a CUT2 family protein with a putative role in ribose transport and pRL120533 (*qatW4*), encoding a protein involved in the uptake of quaternary amines. RL3696 (*hmuP*), encoding a hemin uptake protein, promoted lentil colonisation. Transport proteins encoded by pRL12755 (*qatV5*), involved in the uptake of quaternary amines; pRL100248 (*gtsA*), involved in GABA uptake; and RL3750 (*braD*), involved in the uptake of branched chain amino acids were specifically required for *Lathyrus* root colonisation. Colonisation of alfalfa was influenced by the import of mannitol (RL4216; *mtlG*) and export of polysaccharides via the transporter encoded by RL2976. Import of phosphate by PhnK (RL0173) promoted barley root colonisation. Genes encoding two components of a type IV secretion system (T4SS) encoded on pRL7 were required for *Lathyrus* colonisation, namely

pRL70087 (*traF*) and pRL70155 (*traJ*), which are homologues of *A. tumefaciens virB10* and *virB8* respectively. VirB8 and VirB10 are periplasmic subunits required for translocation of substrates across the inner membrane, with *A. tumefaciens virb10* mutants impaired in secretion via the T4SS and T pilus biogenesis (Krehenbrink and Downie, 2008) (Jakubowski et al., 2009). Rhizobia inject effector proteins into host plant cells to modify plant signalling. In *M. loti* the T4SS is involved in nodulation in a host-specific manner. Whilst Rlv3841 T4SS mutants are not impaired in the nodulation of *P. sativum*, *V. faba* or *V. hirsta* (Krehenbrink and Downie, 2008), it is possible it may play a role during symbiosis with *Lathyrus spp*. Pili themselves are also known to play a direct role in bacterial attachment and colonisation of surfaces (Warner et al., 1981).

Two genes with putative roles in nitrogen metabolism, namely pRL100161 (*nifD*), encoding the nitrogenase molybdenum-iron protein alpha chain, and RL2270, encoding a putative nodulin-related transmembrane protein with a possible role in iron transport, were important for *Lathyrus* colonisation. The requirement of *nifD* for *Lathyrus* root colonisation is unexpected since it forms part of the nitrogenase complex that catalyses nitrogen fixation in root nodules. Although Rlv3841 can form N_2-fixing symbioses with this host, the stage at which mutant populations were retrieved from the root (7dpi) is too premature for the establishment of N_2-fixing bacteroids. This is not the first time a phenotype has been observed for Rlv3841 *nifD* mutants outside an N_2-fixing symbiosis. A previous INSeq screen with Rlv3841 found *nifD* to have a growth-defective phenotype under free-living conditions (Wheatley et al., 2017).

Table 4.8. Phenotypes for 38 transcriptional regulators demonstrating plant-specific dependencies for colonisation of pea, lentil, *Lathyrus*, alfalfa or barley were examined from INSeq screens assessing gene contributions during free-living growth, symbiosis with pea and attachment to pea roots at different pH's.

			INSeq HMM classifications											
			Free-living				Pea symbiosis					Pea attachment		
Plant	**Gene ID**	**Gene Description**	**G21**	**S21**	**G1**	**S1**	**In**	**Rh**	**Ro**	**Bc**	**No**	**6.5**	**7.0**	**7.5**
Pea	RL0744 (*algR*)	LacI family transcriptional regulator (repressor)	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL0905	GntR family transcriptional regulator	GD	GD	GD	GD	NE	NE	NE	NE	NE	NE	NE	NE
	RL1384	ArcA family transcriptional regulator (activator)	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL2743	DeoR family transcriptional regulator (repressor)	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL3361	GntR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
Lentil	RL3175 (*zur*)	FUR-like zinc response transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL3528	LuxR/GerE transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL3705 (*phyR*)	Two-component sensor/regulator; transcriptional regulator	ES	ES	ES	ES	NE	NE	NE	NE	GD	NE	NE	NE
	RL3771	GGDEF/EAL sensory box protein	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL3780	LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL3828	FNR/CRP family transcriptional regulator	GD	GD	GD	GD	NE	NE	NE	NE	NE			
	RL4253	LacI family transcriptional regulator (repressor)	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	pRL90128	LuxR/GerE transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE

Lathyrus	RL0911	Two-component sensor/regulator; histidine kinase	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL0989 (*matR*)	GntR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL1404	LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL1730 (*celR2*)	Two-component sensor/regulator; transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL1980 (*asfL*)	ChrR-family anti-sigma factor	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL3819	FNR/CRP family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	AD	NE	NE	NE	NE
	RL4145	LacI family transcriptional regulator (repressor)	NE	NE	NE	NE	NE	NE	NE	NE	NE	GD	GD	GD
	pRL100151 (*exsI*)	XRE family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	pRL110577	LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	pRL120325	GntR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
Alfalfa	RL1820	LacI family transcriptional regulator (repressor)	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL2501	LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL3712	LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL3863	GntR family transcriptional regulator	NE	NE	NE	NE	NE	AD	AD	NE	NE	NE	NE	NE
	RL3777	LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL4153	Transmembrane GGDEF sensory box protein	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	pRL90071	LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	pRL100289	Two-component sensor/regulator; histidine kinase	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	pRL100368	Transmembrane GGDEF/EAL sensory box protein	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	pRL110368	Phosphoesterase/regulator	NE	NE	NE	NE	NE	NE	NE	NE	AD	NE	NE	NE

	RL0265	XRE family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL1163	Two-component sensor/regulator; histidine kinase	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
Barley	RL3102	PadR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE
	RL4378	LacI family transcriptional regulator (repressor	NE	NE	NE	NE	NE	NE	NE	NE	AD	NE	NE	NE
	pRL110105	LysR family transcriptional regulator	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE

INSeq experiments are for Rlv3841(i) Free-living, grown on glucose at 21% oxygen (G21), succinate 21% (S21), glucose 1% (G1) and succinate 1% (S1) (Wheatley et al., 2017); (ii) Pea symbiosis: input library (In), 5 dpi rhizosphere growth (Rh), 5 dpi root colonisation (Ro), bacteroid (Bc) and nodule bacteria (No) (Wheatley et al., 2020); and (iii) Pea attachment: 1-hour attachment to pea roots at pH 6.5, 7.0 and 7.5 (Parsons, 2019). Gene classification: essential (ES), growth defective (GD), neutral (NE) and growth-advantaged (AD).

4.3. Conclusion

Whilst most studies on *Rhizobium leguminosarum* have focused on its symbiotic lifestyle as an endosymbiont in root nodules, this is not its only lifestyle. To establish an effective symbiosis, rhizobia must first survive in soil and compete with other microbiome members prior to infecting host legumes and differentiating into N_2-fixing bacteroids. It is predicted that a single rhizobial cell in soil has just a one in a million chance of finding its symbiotic host (Denison and Kiers, 2004) and yet few studies have focused on identifying genes involved in rhizosphere growth and root colonisation, the processes that precede nodule infection itself. BarSeq was successfully used to examine gene requirements for Rlv3841 rhizosphere growth and root colonisation of host legumes (pea, lentil and *Lathyrus*), a non-host legume (alfalfa) and a non-legume (barley). Comparison of the genes identified as contributing to successful rhizosphere growth and root colonisation for each plant species revealed core sets of genes commonly involved in these processes among all five plant species. Key observations of common factors required for rhizosphere fitness (Fig. 4.12) and root colonisation are summarised (Fig 4.12). These show a clear need to synthesise compounds (such as amino acids, ribonucleotides and cofactors), alter metabolic function, sense external stimuli and modify gene expression accordingly, and adapt to various stresses (such as changes in osmolarity). Notably, few genes commonly required for rhizosphere growth and root colonisation had predicted roles in transport of solutes and other molecules. This lack of common transporters suggests variation in the available resources among the five plant rhizospheres, likely because of differences in root exudate composition.

BarSeq also allowed identification of plant-specific genes implicated in rhizosphere growth and/or root colonisation. The plant-specific genes required for rhizosphere growth (Fig. 4.14) and root colonisation (Fig.4.15) emphasise that, in contrast to common determinants involved in these processes, those encoding transporters account for a large proportion. Variation in metabolic, adaptive, cell surface and regulatory requirements are also clear, likely reflecting not only variation in the rhizospheres of different plant species, but also their root surfaces.

The power of BarSeq is that it enables identification of genes required when in competition with other bacteria i.e., other mutants present in the mutant pool. This is essential for investigating determinants of rhizosphere fitness and root colonisation since mutation of many genes may not negatively impact these processes in artificial environments where a single mutant strain is inoculated in isolation. Identification of the rhizosphere-induced protein RhiB as a common determinant of rhizosphere fitness is a classic example of this with previous studies aimed at characterising the role of RhiB failing to observe any phenotype on mutation of the gene (Cubo et al., 1992; Rodelas et al., 1999; Wisniewski-Dyé and Downie, 2002). With regards to engineering rhizobial strains for increased competitiveness in host rhizospheres, it is perhaps daunting, but not surprising, to see so many plant-specific genes contribute to these processes. These BarSeq datasets provide a useful resource for better understanding the genetic factors that contribute to rhizosphere fitness and root colonisation, both prerequisites for establishing effective N_2-fixing rhizobia-legume symbioses.

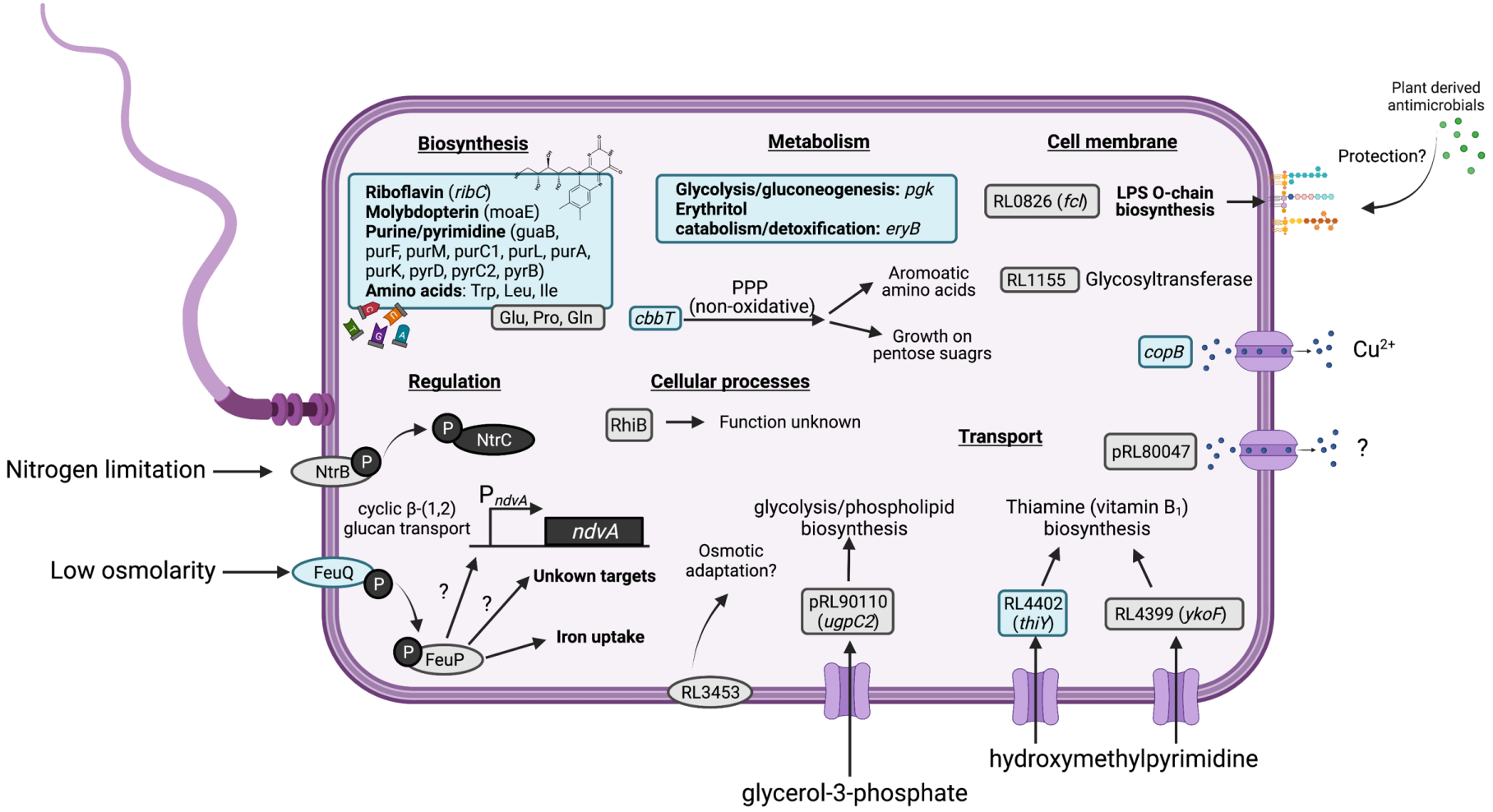

Biosynthesis
Riboflavin (ribC)
Molybdopterin (moaE)
Purine/pyrimidine (guaB, purF, purM, purC1, purL, purA, purK, pyrD, pyrC2, pyrB)
Amino acids: Trp, Leu, Ile
Glu, Pro, Gln
Metabolism
Glycolysis/gluconeogenesis: pgk
Erythritol catabolism/detoxification: eryB
cbbT
PPP (non-oxidative)
Aromoatic amino acids
Growth on pentose suagrs
Cell membrane
RL0826 (fcl)
LPS O-chain biosynthesis
Plant derived antimicrobials
Protection?
RL1155
Glycosyltransferase
copB
Cu2+
Regulation
Cellular processes
RhiB
Function unknown
Transport
pRL80047
?
Nitrogen limitation
NtrB
P
NtrC
cyclic β-(1,2) glucan transport
PndvA
ndvA
Low osmolarity
FeuQ
FeuP
?
Unkown targets
Iron uptake
Osmotic adaptation?
glycolysis/phospholipid biosynthesis
Thiamine (vitamin B1) biosynthesis
pRL90110 (ugpC2)
RL4402 (thiY)
RL4399 (ykoF)
RL3453
glycerol-3-phosphate
hydroxymethylpyrimidine

Figure 4.12. Summary of the common determinants required by Rlv3841 for fitness in the rhizospheres of pea, lentil, *Lathyrus*, alfalfa and barley. Genes shown in grey are commonly required for optimal growth in the rhizosphere of all five plant species. Those that are rhizosphere-progressive for all five plant species are shown in blue. Gene in black were not identified to be common determinants of rhizosphere fitness but based on the literature they interact with those are. Genes are listed in Table 4.6.

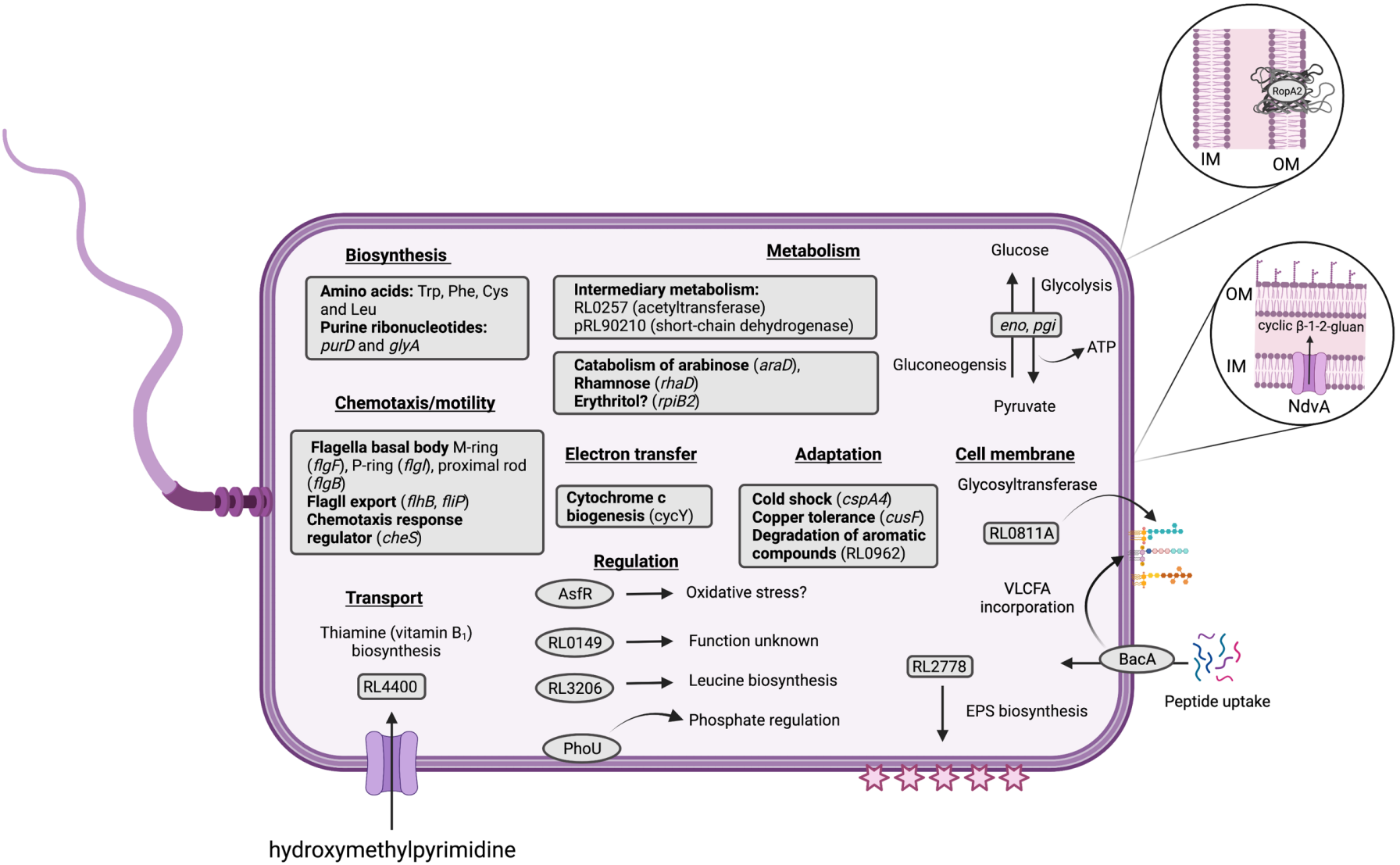

RopA2
IM
OM
OM
cyclic β-1-2-gluan
IM
NdvA
Biosynthesis
Amino acids: Trp, Phe, Cys and Leu
Purine ribonucleotides: purD and glyA
Metabolism
Intermediary metabolism: RL0257 (acetyltransferase) pRL90210 (short-chain dehydrogenase)
Catabolism of arabinose (araD), Rhamnose (rhaD) Erythritol? (rpiB2)
Glucose
Glycolysis
eno, pgi
Gluconeogenesis
ATP
Pyruvate
Chemotaxis/motility
Flagella basal body M-ring (flgF), P-ring (flgI), proximal rod (flgB)
FlagII export (flhB, fliP)
Chemotaxis response regulator (cheS)
Electron transfer
Cytochrome c biogenesis (cycY)
Adaptation
Cold shock (cspA4)
Copper tolerance (cusF)
Degradation of aromatic compounds (RL0962)
Cell membrane
Glycosyltransferase
RL0811A
VLCFA incorporation
Regulation
AsfR
Oxidative stress?
RL0149
Function unknown
RL3206
Leucine biosynthesis
Phosphate regulation
PhoU
Transport
Thiamine (vitamin B1) biosynthesis
RL4400
RL2778
EPS biosynthesis
BacA
Peptide uptake
hydroxymethylpyrimidine

Figure 4.13. Summary of the common determinants of Rlv3841 required for colonisation of pea, lentil, *Lathyrus*, alfalfa and barley roots. Genes are listed in Table 4.6.

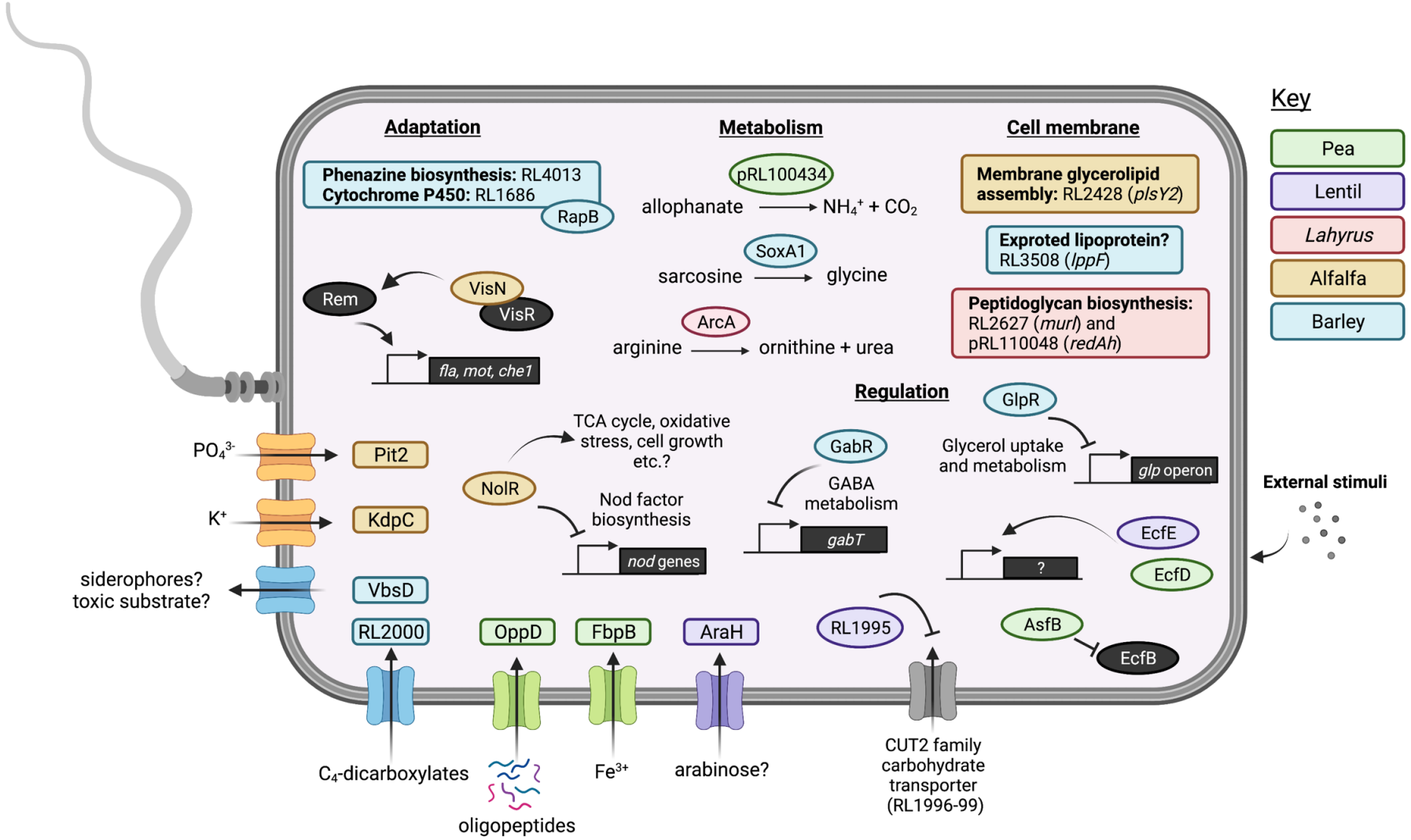

Adaptation
Phenazine biosynthesis: RL4013
Cytochrome P450: RL1686
RapB
Rem
VisN
VisR
fla, mot, che1
Metabolism
pRL100434
allophanate
NH4+ + CO2
SoxA1
sarcosine
glycine
ArcA
arginine
ornithine + urea
Cell membrane
Membrane glycerolipid assembly: RL2428 (plsY2)
Exproted lipoprotein? RL3508 (lppF)
Peptidoglycan biosynthesis: RL2627 (murI) and pRL110048 (redAh)
Regulation
GlpR
Glycerol uptake and metabolism
glp operon
GabR
GABA metabolism
gabT
TCA cycle, oxidative stress, cell growth etc.?
NolR
Nod factor biosynthesis
nod genes
EcfE
EcfD
?
AsfB
EcfB
RL1995
PO4 3-
Pit2
K+
KdpC
siderophores? toxic substrate?
VbsD
RL2000
OppD
FbpB
AraH
C4-dicarboxylates
oligopeptides
Fe3+
arabinose?
CUT2 family carbohydrate transporter (RL1996-99)
External stimuli
Key
Pea
Lentil
Lahyrus
Alfalfa
Barley

Figure 4.14. Summary of the plant-specific factors important in the fitness of Rlv3841 in the rhizospheres of pea, lentil, *Lathyrus*, alfalfa or barley. Genes shown in black were not identified to have plant-specific phenotypes but based on the literature interact with those that do. Genes are listed in Tables S4.8-S4.12.

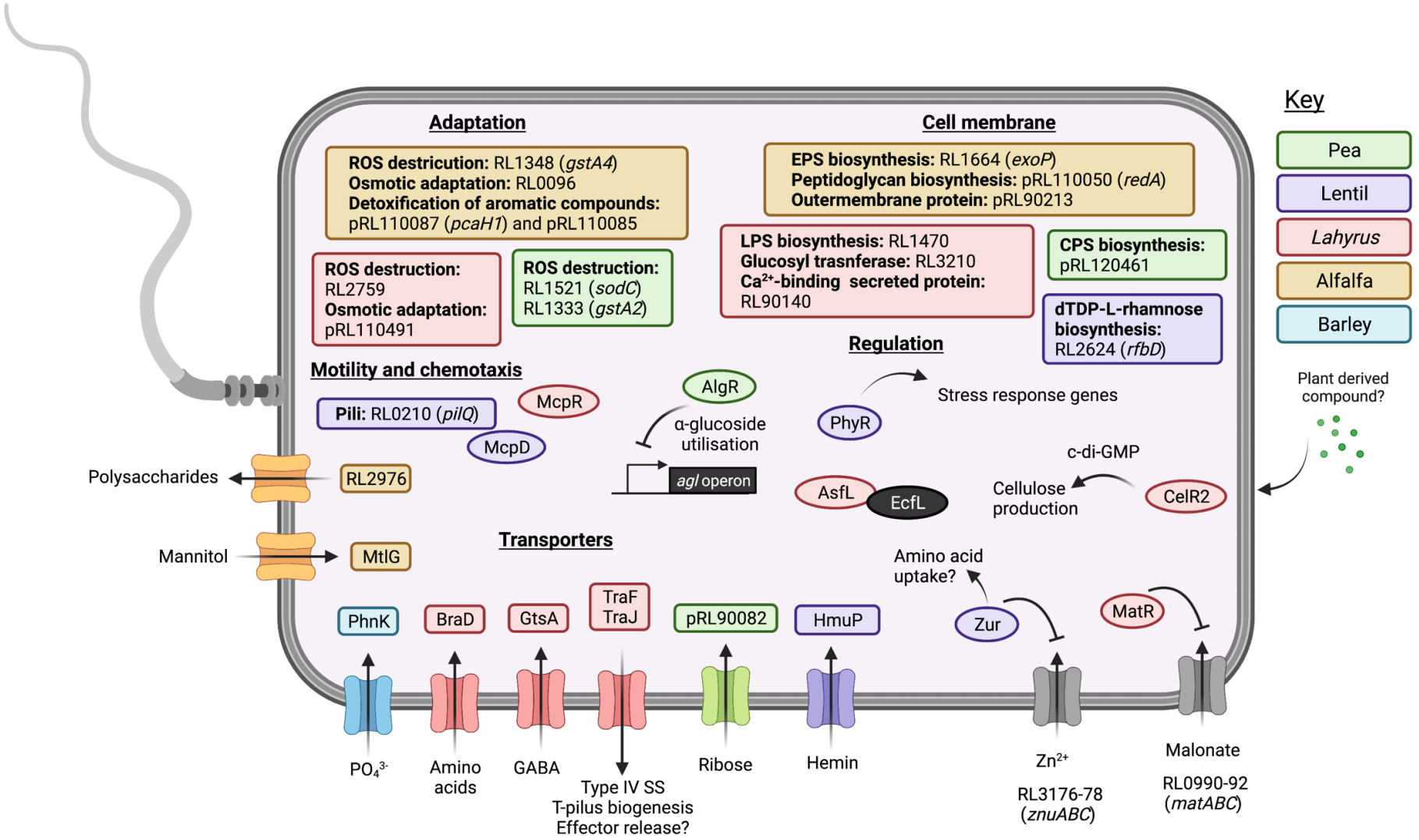

Adaptation
ROS destrication: RL1348 (gstA4)
Osmotic adaptation: RL0096
Detoxification of aromatic compounds: pRL110087 (pcaH1) and pRL110085
ROS destruction: RL2759
Osmotic adaptation: pRL110491
ROS destruction: RL1521 (sodC) RL1333 (gstA2)
Cell membrane
EPS biosynthesis: RL1664 (exoP)
Peptidoglycan biosynthesis: pRL110050 (redA)
Outermembrane protein: pRL90213
LPS biosynthesis: RL1470
Glucosyl trasnferase: RL3210
Ca2+-binding secreted protein: RL90140
CPS biosynthesis: pRL120461
dTDP-L-rhamnose biosynthesis: RL2624 (rfbD)
Regulation
Motility and chemotaxis
Pili: RL0210 (pilQ)
McpR
McpD
AlgR
α-glucoside utilisation
agl operon
PhyR
Stress response genes
c-di-GMP
Cellulose production
CelR2
AsfL
EcfL
Polysaccharides
RL2976
Mannitol
MtlG
Transporters
Amino acid uptake?
Zur
MatR
PhnK
BraD
GtsA
TraF TraJ
pRL90082
HmuP
PO4 3-
Amino acids
GABA
Type IV SS T-pilus biogenesis Effector release?
Ribose
Hemin
Zn2+
RL3176-78 (znuABC)
Malonate
RL0990-92 (matABC)
Key
Pea
Lentil
Lahyrus
Alfalfa
Barley
Plant derived compound?

Figure 4.15. Summary of the plant-specific factors required for Rlv3841 colonisation of roots of pea, lentil, Lathyrus, alfalfa and barley. Genes shown in black were not identified to have plant-specific phenotypes but based on the literature interact with those that do. Genes are listed in Tables S4.8-S4.12.

Chapter 5

Identifying competitive colonisation determinants in *R. leguminosarum* and *P. fluorescens* using BarSeq

.

5.1. Introduction

Competitive rhizosphere growth and root colonisation are key requirements of plant growth promoting rhizobacteria during the establishment of beneficial plant-bacteria interactions such as *Rhizobium*-legume symbioses. The success or failure to establish beneficial plant-bacteria interactions can be influenced by the plant itself (Lebeis et al., 2015; Reinhold-Hurek et al., 2015; Sasse et al., 2018), and both directly or indirectly by other microbiome members (Deines and Bosch, 2016). For example, two individuals may directly compete for the same resources or deploy mechanical weapons, whilst others may modify the environment through alteration of pH or production of a metabolite that can be utilised by another individual (Zelezniak et al., 2015; Granato et al., 2019). Microbiome engineering is an emerging field aimed at improving agricultural activity either through the modification of native microbes or the addition of non-native species (Song et al., 2021). However, these goals are often hindered by the fact that microbial inoculants regularly fail to establish, partly due to antagonistic interactions with native competing bacteria (Albright et al., 2021). For this reason, an emerging research priority in microbiome engineering is to identify such barriers and develop solutions to overcome them. In the case of competition for a specific resource, this could be engineering an introduced strain with transporters to increase uptake efficiency or adding the resources during the initial phase of inoculum establishment (Albright et al., 2021). Regardless of this, most studies characterising genetic determinants of bacterial rhizosphere growth and root colonisation have predominantly done so with single strains under sterile laboratory conditions.

A fundamental challenge in understanding how plant-bacteria and bacteria-bacteria interactions influence the rhizosphere persistence and colonisation ability of an individual species is to move from single to multispecies studies. Whilst the vast complexity present in natural root-associated microbiomes would make this extremely challenging, simplified SynComs have successfully been used to investigate the role of individual species and plant-derived metabolites during colonisation of root niches (Niu et al., 2017; Voges et al., 2019). These studies have demonstrated the importance of both bacteria-bacteria and plant-bacteria interactions during root colonisation and their subsequent effect on the host plants. However, they have not characterised the molecular mechanisms underlying these interactions. For example, Niu *et al.* (2017) demonstrated that removal of one specific SynCom member led to failure of the remaining members to form a stable community on maize roots, resulting in competitive dominance by a single strain which, in isolation, was less capable of warding off a pathogen. They did not however, identify how that particular bacterial species contributed to the stability of the community during colonisation of maize roots. BarSeq provides an exciting opportunity to identify candidate genes contributing to rhizosphere fitness and root colonisation in a bacterial species when co-inoculated with one or more other bacterial species. To date, BarSeq has only been used to identify genes contributing to fitness of a single bacterial species under a given condition. However, the use of a 20N random barcode in the transposon, theoretically provides 10^{13} unique 'identifiers' thus opening the possibility of conducting multispecies analyses, since the diversity in barcodes allows the generation of mutant libraries in multiple strains with minimal chance that the same barcode will repeat. By co-inoculating two or more RB-TnSeq libraries, it should be possible to simultaneously quantify gene fitness contributions for both in BarSeq screens.

The aim of this chapter is to investigate the mechanisms underpinning Rlv3841 rhizosphere fitness and root colonisation in a competitive setting, more analogous to field conditions. To achieve this, Rlv3841 was co-inoculated with *P. fluorescens* SBW25 (SBW25). *P. fluorescens* strains can promote plant growth via various mechanisms, such as eliciting induction of systemic resistance or active exclusion of plant pathogens from the rhizosphere through the production of antimicrobial compounds. SBW25 is a common pseudomonad originally isolated from the phyllosphere of sugar beet and subsequently shown to colonise and persist in the rhizosphere of numerous plant species (De Leij et al., 1995; Thompson et al., 1995; Naseby et al., 2001; Unge and Jansson, 2001; Gal et al., 2003; Humphris et al., 2005; Jackson et al., 2005; Noirot-Gros et al., 2018; Wilton et al., 2018; Carroll et al., 2020). Additionally, its genome has been mapped, it has been used as a model strain for studying bacterial attachment, colonisation dynamics (Carroll et al., 2020) and promoter trapping assays have identified genes induced during growth in the sugar beet rhizosphere (Rainey, 1999). These include genes with predicted functions in high-affinity branched chain amino acid transport, xylose metabolism, copper homeostasis, C_4-dicarboxyilic acquisition as well as the *rsp* gene cluster which shows similarity to type III secretion systems (T3SS).

This chapter will describe the use of BarSeq to identify genetic determinants contributing to Rlv3841 and SBW25 rhizosphere fitness and root colonisation under competitive conditions. This was achieved through co-inoculation of Rlv3841- and SBW25-barcoded mutant libraries onto pea, as a host legume of Rlv3841, and barley, as a model crop species. Two plant species were chosen to give an insight into host influence. The SBW25-barcoded mutant library was also inoculated alone onto pea

and barley. Comparison of these data sets with those obtained in Chapters 3 and 4 regarding Rlv3841 colonisation of pea and barley during single inoculation, will enable the identification of genes specifically required for competitive pea and/or barley rhizosphere growth and root colonisation.

5.2. Results and discussion

5.2.1. Generation and mapping of *P. fluorescens* SBW25 barcoded transposon mutant library

To screen for Rlv3841 and SBW25 genes contributing to rhizosphere fitness and root colonisation under competitive conditions, first a randomly barcoded SBW25 mariner transposon library was generated through conjugation with an *E. coli* donor library generated as previously described (Chapter 3.2.1). Sequencing of the SBW25 transposon library identified 684,486 mutant strains that reliably mapped to the SBW25 genome, each containing a unique barcode identifier. As seen with mapping of the Rlv3841 transposon library, $2.2x10^6$ unique barcodes were sequenced but only approximately 30% (684,486) were reliably mapped. This was primarily due to an insufficient number of reads supporting the mapping of the remaining barcodes, which can be rectified in the future by deep sequencing of the initial mutant library. Since the SBW25 and Rlv3841 mutant libraries were co-inoculated during competition experiments it was not possible to use strains containing an identical barcode present in both mutant libraries for calculation of gene fitness since there is no way to determine if the reads containing that barcode belongs to a Rlv3841 or SBW25 mutant. Comparison of barcodes present in each mutant library revealed less than a 3% overlap in mapped barcodes between the SBW25- and Rlv3841-barcoded mutants.

These were computationally removed, leaving 666,797 mutants that mapped to 88,907 distinct locations in the SBW25 genome. The SBW25 genome contains 144,722 TA sites and so barcodes were mapped to 61% of the potential insertion sites. To calculate gene fitness values, only insertions located within the central 10-90% region of the gene were used. Of the 666,797 barcoded stains identified in the SBW25 library, 392,021 were mapped to 52,812 distinct locations within the central 10-90% region of genes. The number of mutants for each gene correlated with the number of TA sites (Pearson correlation coefficient r = 0.98). Over 91% (5,368 genes) of the 5,921 protein coding genes contained central insertions (Table S5.1), with a median of 51 insertion mutants per gene. A total of 557 genes lacked insertions entirely, one of which does not contain mariner insertion (TA) sites within the central 10-90% region (Table S5.2). These genes are likely to be essential for growth, even in rich media (KB), encoding proteins involved in fundamental cellular processes such as ribosomal biogenesis, translation, coenzyme biosynthesis, energy production, replication, and cell wall biogenesis (Fig. 5.1). Examples include RNA polymerase (*rpoABC*; PFLU5502, PFLU5534-33), sigma factors (*rpoN*; PFLU0882 and *rpoD*; PFLU5592), *rpl* and *rps* ribosomal proteins, ATP synthase required for ATP synthesis (*atpB*; PFLU6124 and *atpFHAGD*; PFLU6122-18), chromosol partitioning genes (*parE*; PFLU0499 and *parC*; PFLU0501) and an operon involved in cell wall biosynthesis and division (PFLU0942-53) (Azzolina et al., 2001).

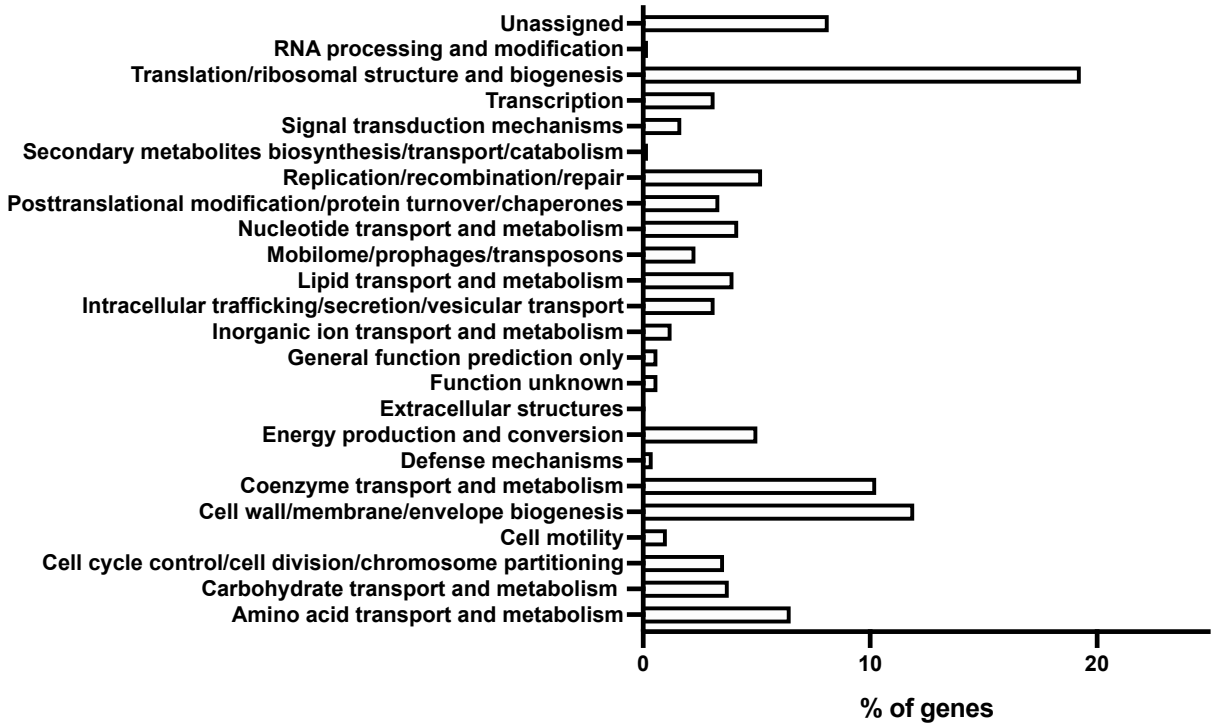


Figure 5.1. Functional classification of genes lacking insertions in the SBW25 barcoded transposon mutant pool. Functional classification based on COG annotation (Tatusov et al., 2000).

5.2.2. Colonisation dynamics of Rlv3841 and SBW25 during competition on pea and barley

To assess whether 7 dpi is a suitable timepoint at which to harvest transposon libraries during competition experiments, the colonisation dynamics of Rlv3841 and SBW25 during competition with each other on pea and barley were assayed. Seven-day-old seedlings were inoculated with 10^5 cfu of both Rlv3841 and SBW25. The rhizosphere, loosely root-associated and strongly root-associated populations were quantified at 1, 3, 5 and 7 dpi. Colonisation assays were also conducted for SBW25 inoculated alone onto pea and barley. These results, along with those from Chapters 3 and 4 for Rlv3841 colonisation of pea and barley, respectively, allowed for comparison of Rlv3841 and SBW25 colonisation dynamics under competitive and non-competitive conditions.

For Rlv3841, a significant increase in cfu/g root was observed from 1 to 7 dpi for each population under all conditions tested (Fig. 5.2). When inoculated alone, Rlv3841 populations reached saturation by 5 dpi (Fig. 5.2A and Fig 5.2C). The same was true during competition with SBW25 on barley (Fig 5.2D). During co-inoculation onto pea however, there was a significant increase in cfu/g root retrieved from all three fractions between 5 and 7 dpi (rhizosphere p = < 0.0001, loosely root-associated p = 0.007, strongly root-associated p = 0.0032) (Fig. 5.2B). Comparison of Rlv3841 total cfu 7 dpi onto pea during competition with SBW25 versus inoculation alone revealed that, whilst there was no significant difference in rhizosphere growth (p = 0.6088), Rlv3841 was significantly impaired in root colonisation (Fig. 5.3A). In contrast, on barley Rlv3841 reached significantly higher cfu in the rhizosphere and demonstrated improved colonisation of roots when in competition with SBW25 (Fig.5.3B).

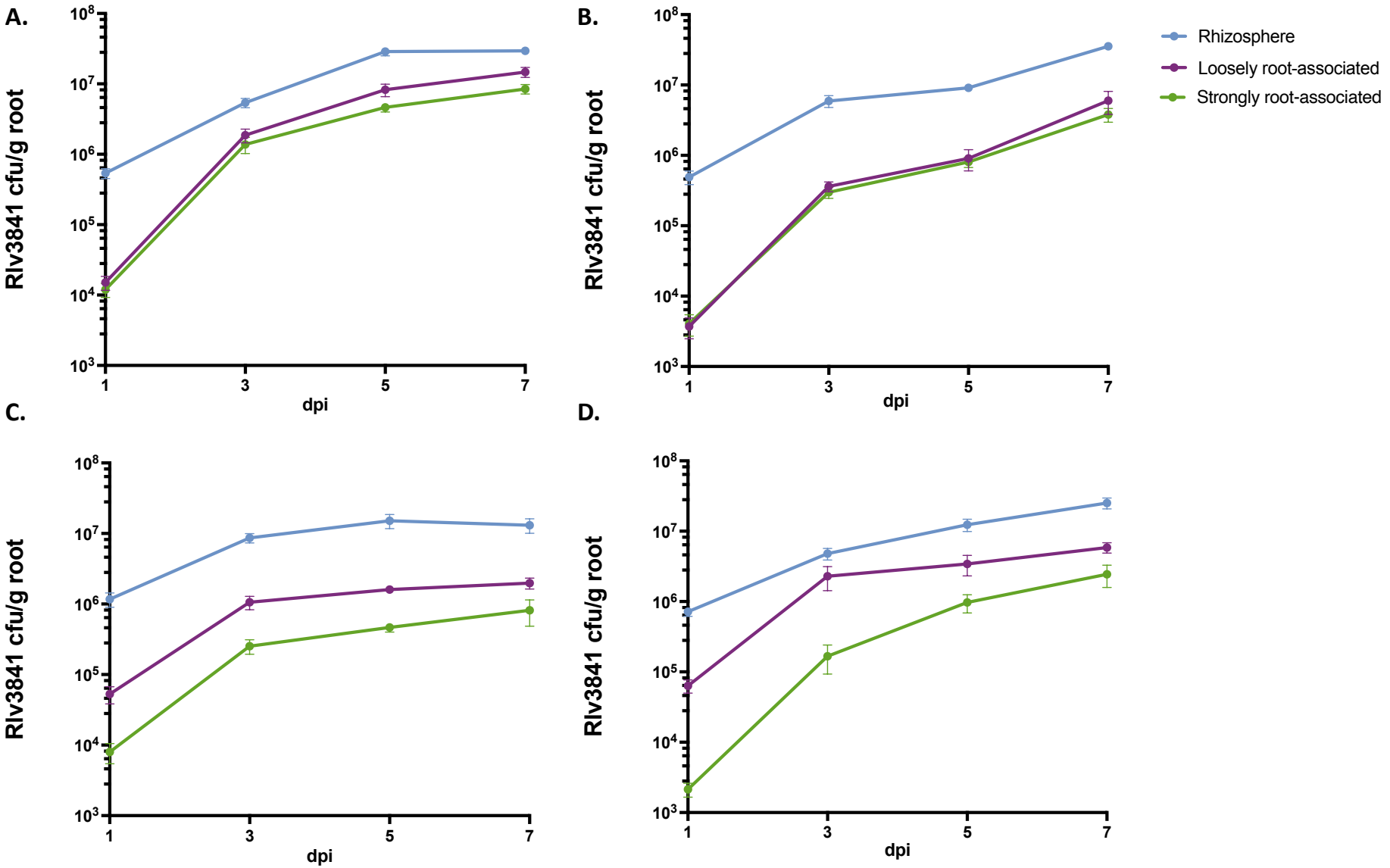

A.
B.
C.
D.
Rlv3841 cfu/g root
dpi
Rhizosphere
Loosely root-associated
Strongly root-associated

Figure 5.2. Rlv3841 colonisation dynamics during single and co-inoculation with SBW25 on pea and barley. Data for single inoculation assays is from Chapters 3 and 4. For co-inoculation assays 7-day-old seedlings were inoculated with 10^5 cfu of Rlv3841 and 10^5 cfu of SBW25. Mean populations of Rlv3841 in the rhizosphere and associated with the root are shown at 1, 3, 5 and 7 dpi for **A.** colonisation of pea during single inoculation; **B.** colonisation of pea during competition with SBW25; **C**. colonisation of barley during single inoculation; and **D.** colonisation of barley during competition with SBW25. Error bars represent ± SEM. n ≥ 7. To assess the difference in SBW25 cfu/g root over 1 to 7 dpi a one-way ANOVA analysis was performed for each sample followed by Tukey's post hoc test. A significant increase in Rlv3841 cfu/g root was seen from 1 to 7 dpi for the rhizosphere ([pea SI: F = 185.7, R^2 = 0.9116, *P*<0.0001], [pea CI: F = 128.7, R^2 = 0.9273, *P*<0.0001], [barley SI: F = 33.13, R^2 = 0.7451, *P*<0.0001] and [barley CI: F = 66.32, R^2 = 0.8541, *P*<0.0001]), loosely root-associated ([pea SI: F = 129.4, R^2 = 0.8779, *P*<0.0001], pea CI: F = 74.53, R^2 = 0.8887, *P*<0.0001], [barley SI: F = 53.46, R^2 = 0.8251, *P*<0.0001] and [barley CI: F = 44.20, R^2 = 0.7959, *P*<0.0001]), and strongly root-associated ([pea SI: F = 123, R^2 = 0.8723, *P*<0.0001], pea CI: F = 146.5, R^2 = 0.9361, *P*<0.0001], [barley SI: F = 115.6, R^2 = 0.9107, *P*<0.0001] and [barley CI: F = 113.9, R^2 = 0.9095, *P* <0.0001]).

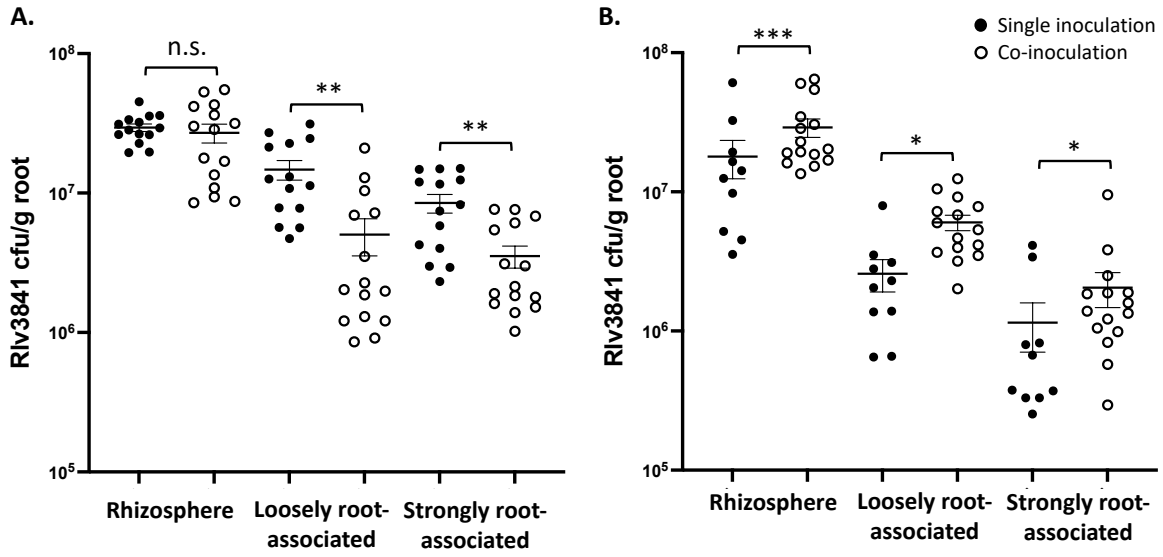


Figure 5.3. Comparison of Rlv3841 populations 7 dpi for colonisation of pea and barley after single and co-inoculation with SBW25. A. Rlv3841 colonisation of pea **B.** Rlv3841 colonisation of barley. To assess statistical significance an unpaired *t*-test was used to compare total cfu at 7 dpi for each fraction after single and co-inoculation with SBW25. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001. Error bars show ± SEM.

For SBW25, not all populations showed a significant increase between 1 and 7 dpi; some reached saturation prior to 1 dpi (Fig. 5.4). Comparison of SBW25 total cfu 7 dpi on pea revealed that when co-inoculated with Rlv3841, SBW25 reached a significantly higher population in the rhizosphere but showed no difference in ability to colonise pea roots relative to inoculation alone (Fig. 5.5A). In contrast, when SBW25 competed with Rlv3841 on barley, SBW25 reached lower populations both in the rhizosphere and on the root relative to inoculation alone (Fig. 5.5B). Since most populations were at saturation by 7 dpi, except for Rlv3841 during competition on pea, this timepoint was deemed suitable for retrieval of mutant libraries. This not only allows for comparison with results obtained in Chapters 3 and 4 for Rlv3841 colonisation of pea and barley, but also ensures that strongly-root associated populations are retrieved prior to nodule development.

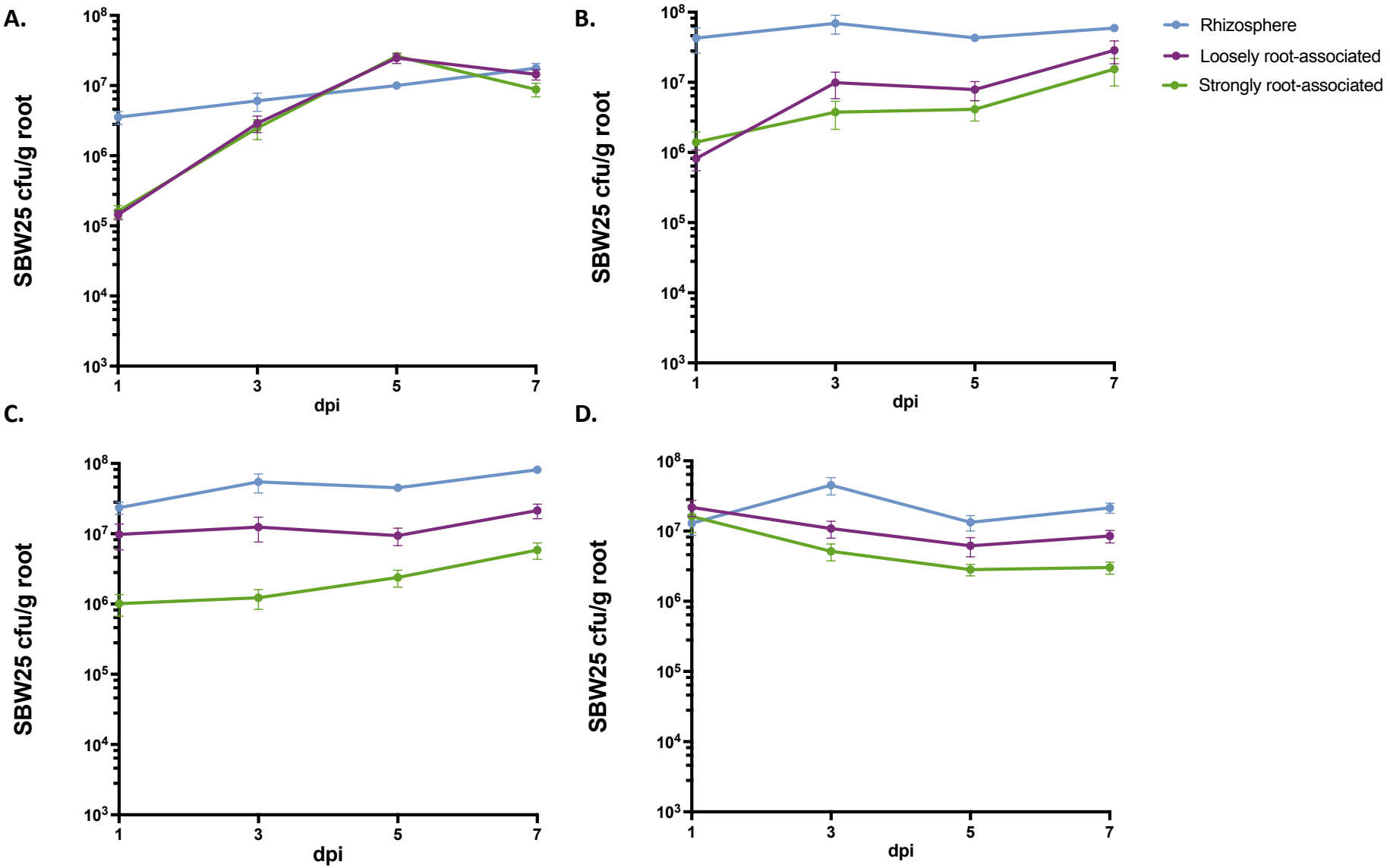

A.
B.
C.
D.
SBW25 cfu/g root
dpi
Rhizosphere
Loosely root-associated
Strongly root-associated

Figure 5.4. SBW25 colonisation dynamics during single and co-inoculation with Rlv3841 on pea and barley. For single inoculation assays 7-day-old seedlings were inoculated with 10^5 cfu of SBW25, whilst for co-inoculation assays 10^5 cfu of SBW25 and 10^5 cfu of Rlv3841 were inoculated. Mean populations of SBW25 in the rhizosphere and associated with the root are shown at 1, 3, 5 and 7 dpi for **A.** colonisation of pea during single inoculation; **B.** colonisation of pea during competition with Rlv3841; **C**. colonisation of barley during single inoculation; and **D.** colonisation of barley during competition with Rlv3841. Error bars represent SEM. n ≥ 5. To assess the difference in SBW25 cfu/g root over 1 to 7 dpi a one-way ANOVA analysis was performed for each sample followed by Tukey's post hoc test. Rhizosphere ([pea SI: F = 8.932, R^2 = 0.4408, *P*<0.0002], [pea CI: F = 2.724, R^2 = 0.2141, *P*=0.0617], [barley SI: F = 5.116, R^2 = 0.3242, *P*<0.0053] and [barley CI: F = 2.467, R^2 = 0.2033, *P*=0.0821]), loosely root-associated ([pea SI: F = 42.26, R^2 = 0.7885, *P*<0.0001], [pea CI: F = 8.373, R^2 = 0.4557, *P*=0.0003], [barley SI: F = 2.410, R^2 = 0.1797, *P*=0.0845] and [barley CI: F = 3.489, R^2 = 0.2354, *P*=0.0261]), and strongly root-associated ([pea SI: F = 29.20, R^2 = 0.7204, *P*<0.0001], pea CI: F = 2.506, R^2 = 0.2004, *P*=0.0780], [barley SI: F = 6.376, R^2 = 0.3670, *P*=0.0016] and [barley CI: F = 0.4061, R^2 = 0.03459, *P*=0.7496]).

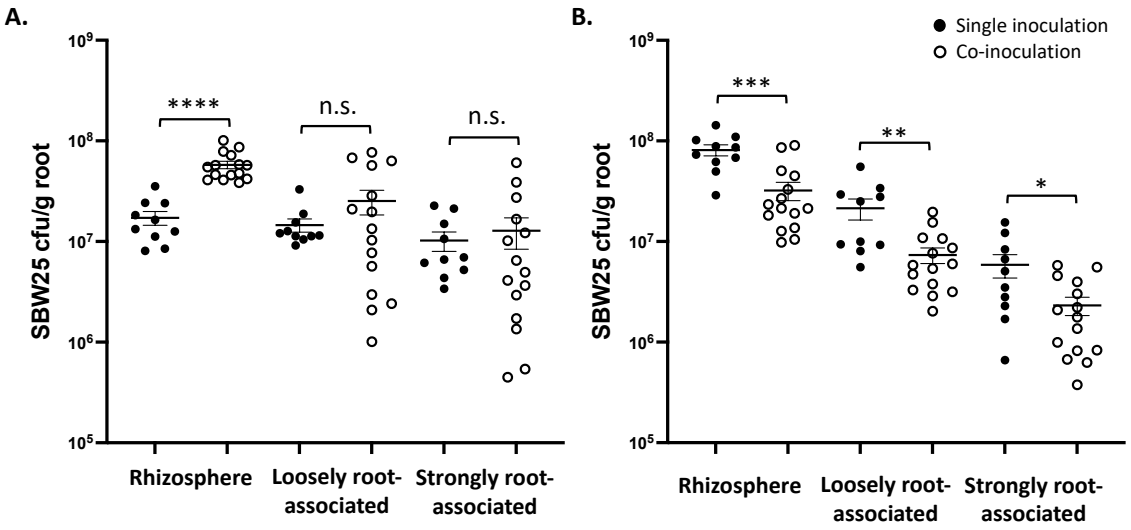


Figure 5.5. Comparison of SBW25 populations 7 dpi for colonisation of pea and barley after single and co-inoculation with Rlv3841. A. SBW25 colonisation of pea. **B.** SBW25 colonisation of barley. To assess statistical significance an unpaired *t*-test was used to compare total cfu at 7 dpi for each fraction after single and co-inoculation Rlv3841. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001. Error bars show ± SEM.

For BarSeq experiments, mutant libraries must undergo several generations of growth under the test condition to ensure clear resolution among mutant phenotypes. After seven days of growth, the total number of bacteria had increased by a minimum of eight generations, which is sufficient for resolution of mutant phenotypes (Table 5.1).

Table 5.1. Rlv3841 and SBW25 populations 7 dpi

Condition*	Total cfu/plant	Generations of growth
SBW25		
Pea SI	$4.2x10^7 \pm 3.5x10^6$	8
Pea CI	$9.6x10^7 \pm 1.2x10^7$	9
Barley SI	$1.08x10^8 \pm 1.3x10^7$	10
Barley CI	$4.18x10^7 \pm 6.5x10^6$	8
Rlv3841		
Pea CI	$3.6x10^7 \pm 4.8x10^6$	8
Barley CI	$3.7x10^7 \pm 5.1x10^6$	8

*SI - single inoculation, CI – co-inoculation

5.2.3. BarSeq experimental design to identify Rlv3841 and SBW25 genes important for competitive colonisation of pea and barley

Having established 7 dpi as a suitable timepoint for retrieval of mutant libraries during competition experiments, an aliquot of Rlv3841 and SBW25 mutant libraries were thawed, combined in a 1:1 ratio and inoculated onto seven day-old pea and barley seedlings (Fig 5.6). A total of 96 plants were inoculated per experiment with 2.5 x 10^5 cfu of both the Rlv3841 and SBW25 mutant library, providing a total inoculum of 2.4 x 10^7 cfu for each. Considering that only data for 26% of the Rlv3841 mutants will be retained during analysis of BarSeq data (due to mapping), the number of mapped mutants in the inoculum represents approximately a 45-fold coverage of the 140,056 TA sites in Rlv3841 and an 830-fold coverage of the 7,528 genes. Whilst for SBW25, data for 30% of the mutants will be retained, providing a 50-fold coverage of the 144,722 TA sites and 1,200-fold coverage of the 5,921 protein coding genes. The 96 samples for each condition were pooled prior to gDNA extraction and barcodes were PCR-amplified for sequencing. Input, rhizosphere, loosely and strongly root-associated libraries were sequenced and analysed following the BarSeq data analysis pipeline (Chapter 2.5.6) (Wetmore et al., 2015). For data analysis, loosely and strongly root-associated libraries were combined and analysed collectively to identify genes required for root colonisation.

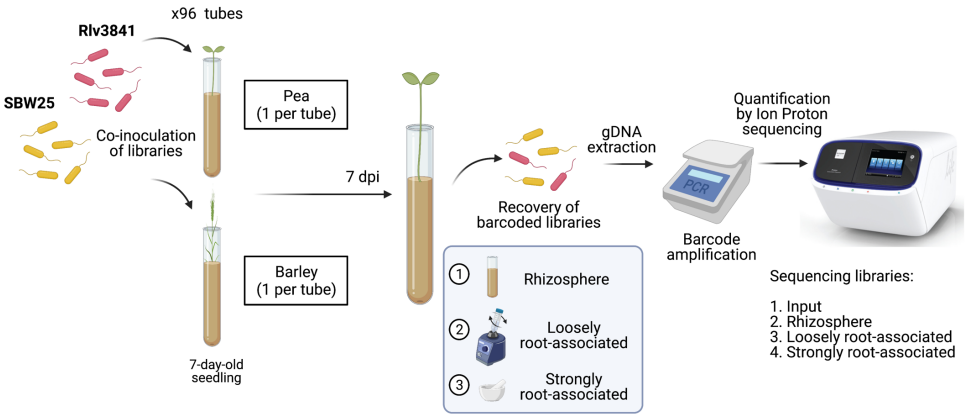


Figure 5.6. Experimental design for BarSeq competition experiments. A total of 96 7-day-old pea and barley seedlings were inoculated with $2.5x10^5$ cfu of Rlv3841- and SBW25-barcoded mariner insertion mutants and retrieved from the rhizosphere and roots at 7 dpi. Samples were pooled across the 96 plants, gDNA extracted, and barcodes amplified by PCR. Input, rhizosphere, loosely and strongly root-associated samples were sequenced by Ion Proton to quantify barcode frequencies. Loosely and strongly root-associated samples were analysed collectively as root colonisation (7 dpi).

5.2.4. Maintenance of genetic diversity of Rlv3841 and SBW25 transposon libraries *in planta*

For SBW25, all experiments passed quality thresholds with fitness values assigned to 80% (4,720 genes) of the total protein coding genes (Table S5.3) (Chapter 2.5.6). The remaining genes were not assigned fitness values either because they lacked insertion mutants in the library due to their essentiality or lacked sufficient reads at time zero for calculation of fitness value. Analysis of SBW25 mutants retrieved from each condition revealed that the SBW25 mutant population isolated from barley roots after competition with Rlv3841 contained the least mutant diversity (Table 5.2). This

sample contained 14% of the total barcoded mutants mapped in the SBW25 mutant pool. However, the insertion density for this sample was 30.8% which is above the limit of 25% generally considered sufficient for calculation of gene fitness and analysis passed BarSeq quality thresholds (Chapter 2.5.6) (DeJesus and Ioerger, 2013).

For Rlv3841, sufficient reads were only obtained for calculation of gene fitness during competitive root colonisation of barley. Fitness values were assigned to 77% (5,770 genes) of the total protein coding genes (Table S5.4). The remaining experiments failed BaRSeq quality controls as they had less than 50 median reads per gene which is insufficient for accurate calculation of gene fitness (Table 5.3). This is because although SBW25 and Rlv3841 transposon libraries were initially inoculated in a 1:1 ratio, after seven days of growth in the pea rhizosphere and colonisation of pea roots, the SBW25 mutant population was significantly higher and therefore accounted for the majority of barcode reads for these samples (Fig 5.7A). In contrast, there was no significant difference in Rlv3841 and SBW25 population retrieved from the barley rhizosphere or root after seven days (Fig 5.7B). Whilst fitness values were not calculated for Rlv3841 growth in the barley rhizosphere during competition with SBW25, a median of 39 reads per gene were obtained and so this condition was close to passing quality controls.

Table 5.2. Read coverage for SBW25 genes in input, rhizosphere and root colonised samples.

Library	No. unique barcodes sequenced	No. barcodes matching mutant pool	No. barcodes as % of total mutant pool	Insertion density (No. TA sites with at least 1 read)[a]	Median no. reads per gene
Pea SI					
Input	1,993,928	456,778	69%	81,812 (56.5%)	267
Rhizosphere	1,159,998	276,074	41%	71,026 (49.1%)	200
Root colonised	929,255	180,721	27%	60,764 (42.0%)	253
Barley SI					
Input	1,174,021	299,488	45%	72,706 (50.2%)	92
Rhizosphere	956,201	217,321	33%	67,947 (47%)	133
Root colonised	890,614	244,010	37%	65,271 (45.1%)	73
Pea competition					
Input SBW25	*2,719,095	553,186	83%	85,435 (59.0%)	248
Rhizosphere	682,800	152,554	23%	56,982 (39.4%)	173
Root colonised	953,279	173,308	26%	59,968 (41.4%)	177
Barley competition					
Input SBW25	1,757,103	421,578	63%	80,153 (55.4%)	91
Rhizosphere	947,650	196,835	30%	63,302 (43.7%)	84
Root colonised	1,267,279	92,429	14%	44,582 (30.8%)	104

[a]Percentage of the total 144,722 TA sites with at least one read. Single inoculation (SI).

Table 5.3. Read coverage for Rlv3841 genes in input, rhizosphere and root colonised samples for competition experiments.

Library	No. unique barcodes sequenced	No. barcodes matching mutant in Rlv3841 pool	No. barcodes as % of total mutant pool	Insertion density (No. TA sites with at least 1 read)[a]	Median no. reads per gene
Pea competition					
Input Rlv3841	2,773,483	594,647	86%	58,405 (65.1%)	333
Rhizosphere	682,800	24,250	3.5%	12,480 (13.9%)	15
Root colonised	953,279	33,270	4.8%	15,931 (17.8%)	6
Barley competition					
Input Rlv3841	1,865,622	435,457	63%	54,380 (60.7%)	97
Rhizosphere	947,650	42,554	6.1%	18,919 (21.1%)	39
Root colonised	1,267,279	95,776	14%	30,884 (34.4%)	59

[a]Percentage of the total 140,056 TA sites with at least one read.

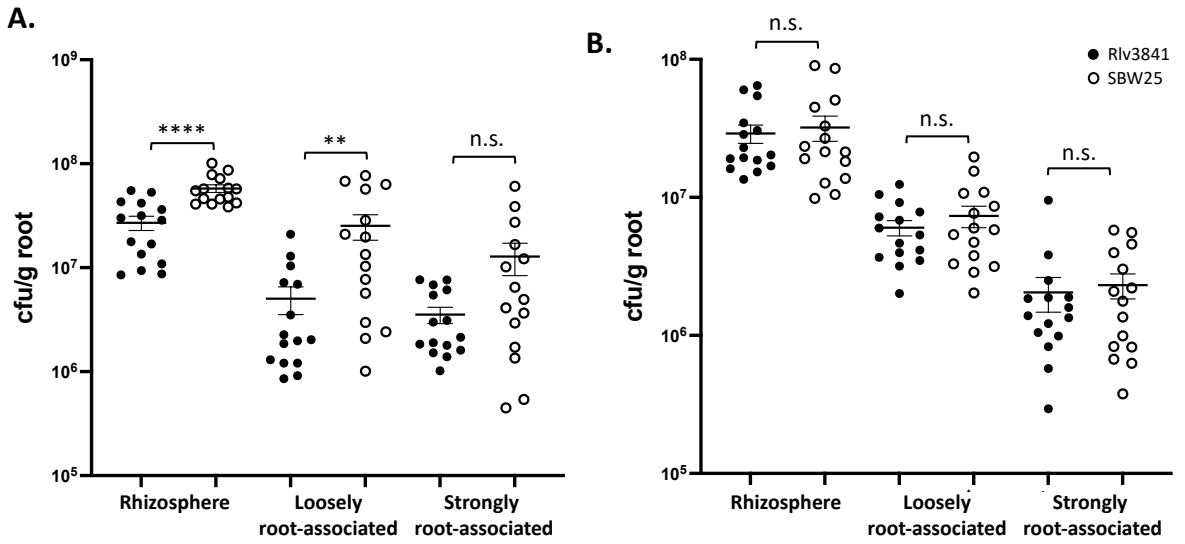


Figure 5.7. Comparison of Rlv3841 and SBW25 populations 7 dpi on pea and barley during co-inoculation with each other. A. Pea. **B.** Barley. To assess statistical significance, an unpaired *t*-test was used to compare Rlv3841 and SBW25 populations for each fraction. **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001. Error bars show ± SEM.

5.2.5. Identification of Rlv3841 genes with significant fitness contributions during competitive colonisation of barley with SBW25

To identify Rlv3841 genes contributing to competitive root colonisation of barley, genes were filtered to select those which had a fitness value less than -1 during competition with SBW25. This led to the identification of 462 genes, most of which had mild fitness contributions (fitness value between -1 and -2), with only 42 genes assigned strong fitness contributions (fitness value < -2) (Table S5.5). To identify those specifically required during competition with SBW25, these genes were compared to the 435 genes found to contribute to Rlv3841 colonisation of barley when inoculated

alone (Table S5.5). This revealed that 92 genes (20%) were required under both conditions (Fig. 5.8 and Table S5.5).

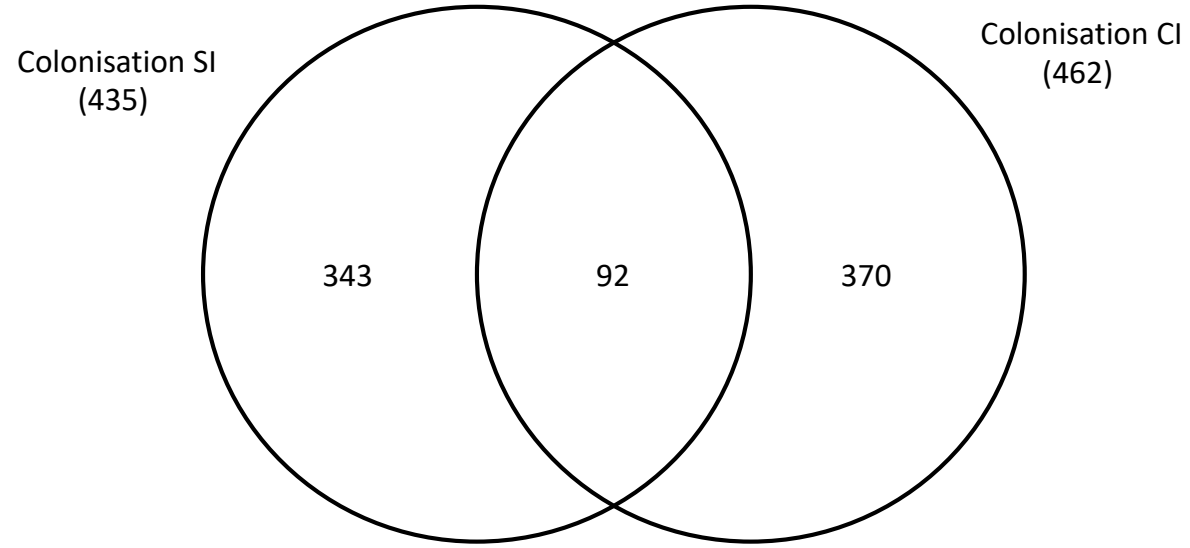


Figure 5.8. Venn diagram of 805 genes contributing to Rlv3841 colonisation of barley under single and co-inoculation with SBW25. Genes are listed in Table S5.5. Single inoculation (SI) and co-inoculation (CI).

As discussed in Chapter 4, genes with fitness values between -1 and 1 are generally classed as neutral. However, a fitness values of -0.42 represents a 25% reduction in growth relative to the average mutant in the pool and therefore the 370 genes specifically identified as contributing to barley root colonisation during competition with SBW25 were filtered to remove those with fitness values less than -0.42 for colonisation of barley when inoculated alone. The remaining genes were then further filtered based on their associated *t*-score, a measure of consistency in the mutant fitness values used to calculate the gene fitness value, to retain those which have high confidence in the predicted fitness value. Genes with fitness values less than -2 were filtered to retain those with an associated *t*-score < -3 and those with fitness values

between -1 and -2 were filtered to retain those with a *t*-score < -2.5 (Chapter 2.5.6). This left 50 genes with high confidence fitness predictions for contribution to barley root colonisation specifically during competition with SBW25 (Table S5.6). Functional classification of the genes revealed that 34% encode hypothetical proteins (Fig.5.9). One of the primary drivers of competition among bacteria is a battle for resources (Hibbing et al., 2010), this is particularly true in relatively nutrient-poor conditions such as the rhizosphere. It is therefore not surprising that after hypothetical proteins, the largest functional group of genes had predicted roles in transport, accounting for 16% (8 genes) of the total genes. This was followed by regulation, which accounted for 14% (7 genes) of genes and intermediary metabolism accounting for 12% (6 genes) of genes.

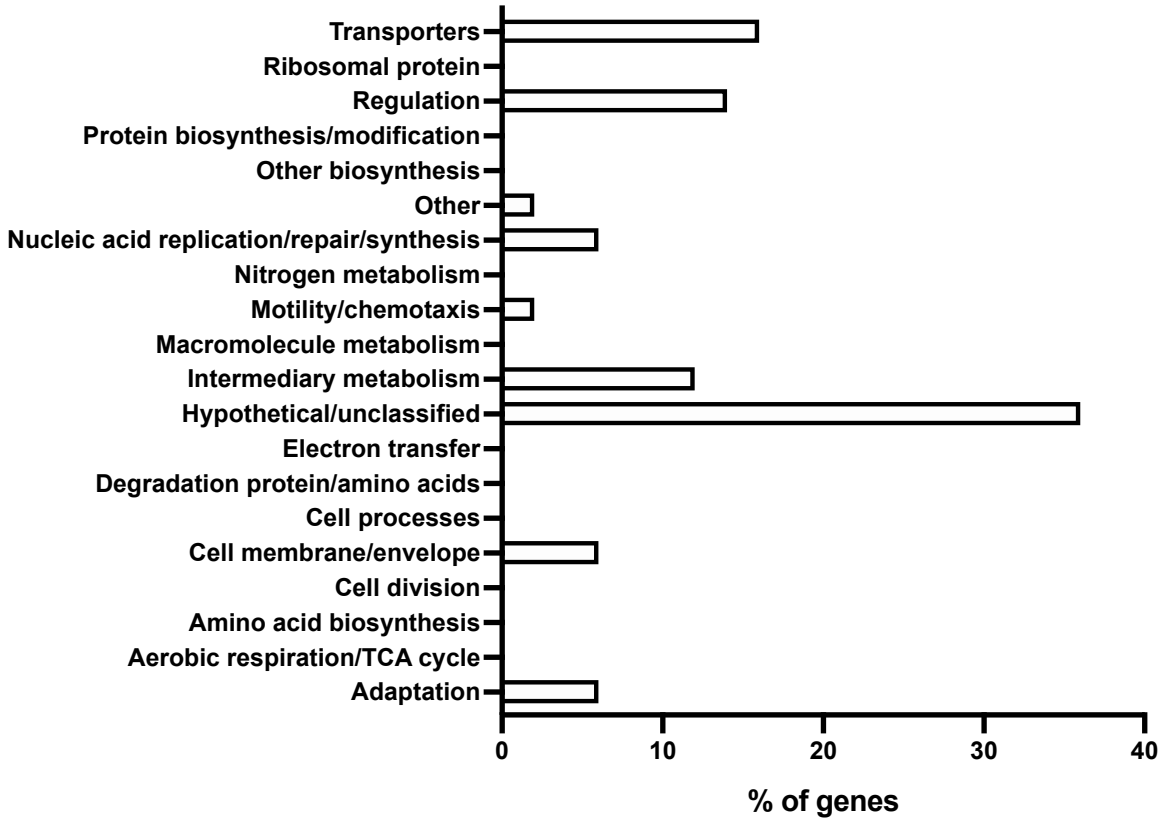


Figure 5.9. Functional classification of 50 genes contributing to Rlv3841 colonisation of barley roots specifically during competition with SBW25.

Functional classification is based on Riley codes (Young et al., 2006). Genes are listed in Table S5.6.

Competition with SBW25 for resources in the rhizosphere is likely facilitated by the fact that the Rlv3841 genome encodes 816 ABC genes (183 complete operons) in contrast to the 335 ABC genes encoded by the SBW25 genome (Young et al., 2006). This huge diversity promotes the acquisition of a broad range of nutrients from the rhizosphere and allows adaption to variation in resource availability caused by the presence of a competitor strain. Rlv3841 genes with roles in transport important during competition with SBW25 include RL0228, a SBP of a peptide transporter; pRL120172, a SBP of a polyamine/opine/phosphonate transporter; RL0168 (*phoD*), the *S. meliloti* homologue of which is induced by phosphate limitation; and RL1993, the *S. meliloti*, homologue of which is induced by nitrogen limitation (Mauchline et al., 2006). Components of several ABC transport systems themselves were also identified, these include pRL100251 (*gtsD*) part of a GABA transporter; RL3545 (*oppC*), a component of an oligopeptide transporter; pRL120515 (*qatW3*), part of a quaternary amine transporter; and pRL100354, a component of an CUT2 family carbohydrate transporter. These altered transport requirements, in combination with the fact that 12% of genes have predicted functional roles in intermediary metabolism, suggest altered resource availability and therefore metabolic demands for Rlv3841 during competition with SBW25.

Seven genes encode proteins with regulatory roles, though their targets are predominantly unknown. These include the anti-sigma factors AsfB (RL0789), and

DgpfD1 (RL1474); three LysR transcriptional regulators, encoded by RL0229, pRL100331 and pRL110388; and two MarR transcriptional regulators, encoded by RL2924 and pRL90275 (Table S5.6). AsfB is the anti-sigma factor of EcfB, though the targets of EcfB are unknown. Analysis of phenotypes for these genes, from various INSeq screens and the BarSeq screens conducted in Chapters 3 and 4, revealed that most had neutral phenotypes under all previously tested conditions. Exceptions were AsfB, which was important for pea rhizosphere growth and root colonisation 7 dpi under single inoculation, but not barley, and RL2924, which was important for pea nodule infection and bacteroid development.

Table 5.4. Phenotypes for seven Rlv3841 transcriptional regulators required for colonisation of barley during competition with SBW25 were examined from INSeq screens assessing gene contributions during free-living growth, symbiosis with pea and attachment to pea roots at different pH's, as well as BarSeq screens assessing growth in the rhizosphere and colonisation of various plant species.

	INSeq HMM classifications												**BarSeq fitness values**									
	Free-living				**Pea symbiosis**					**Pea attachment**			**Pea**		**Lentil**		***Lathyrus***		**Alfalfa**		**Barley**	
Gene ID	G21	S21	G1	S1	In	Rh	Ro	Bc	No	6.5	7.0	7.5	Rhi	Col	Rhi	Col	Rhi	Col	Rhi	Col	Rhi	Col
RL0789 (*asfB*)	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	-2.0	-1.2	-0.3	-0.9	-0.3	-0.3	0.1	-0.1	-0.2	-0.2
RL1474 (*dgpfD1*)	NE	NE	NE	NE	NE	NE	NE	NE	AD	NE	NE	NE	-0.5	0.0	0.1	1.1	0.0	-0.5	0.6	0.1	0.4	0.1
RL0229	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	0.6	-0.1	0.7	-0.1	0.4	-0.3	-0.8	-0.3	0.4	1.3
pRL100331	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	-0.2	-0.3	0.1	0.1	0.1	-0.8	0.1	0.2	0.0	-0.4
pRL110388	NE	NE	NE	NE	NE	NE	AD	NE	NE	NE	NE	NE	0.7	0.4	0.0	-0.5	0.1	0.3	-0.1	-0.5	-0.4	-0.2
RL2924	NE	NE	NE	NE	NE	NE	AD	ES	ES	NE	NE	NE	0.0	0.2	0.1	-0.3	-0.6	-0.1	2.0	1.4	0.0	-0.3
pRL90275	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	NE	0.4	0.6	0.2	0.2	0.2	-0.5	0.9	0.3	-0.3	-0.2

INSeq experiments are for Rlv3841(i) Free-living, grown on glucose at 21% oxygen (G21), succinate 21% (S21), glucose 1% (G1) and succinate 1% (S1) (Wheatley et al., 2017); (ii) Pea symbiosis: input library (In), 5 dpi rhizosphere growth (Rh), 5 dpi root colonisation (Ro), bacteroid (Bc) and nodule bacteria (No) (Wheatley et al., 2020); and (iii) Pea attachment: 1-hour attachment to pea roots at pH 6.5, 7.0 and 7.5 (Parsons, 2019). Gene classification: essential (ES), growth defective (GD), neutral (NE) and

growth-advantaged (AD). BarSeq experiments are for Rlv3841 7 dpi rhizosphere growth (Rhi) and root colonisation (Col) of pea, lentil, *Lathryus*, alfalfa and barley (Chapter 4). Conditions for which mutation of the gene provides a disadvantage are highlighted in orange and those that provide an advantage are shown in green. Genes are from Table S5.6.

Three genes with functional roles in adaption were important for competitive colonisation of barley by Rlv3841. These were pRL110288 (*pcaJ*), involved in the detoxification or catabolism of aromatic compounds such as protocatechuate (Göbel et al., 2002; Garcia-Fraile et al., 2015); RL2050 (*surE*), a putative acid phosphatase; and RL3994 (*atvA*), a putative acid tolerance protein (Table S5.6). The exact physiological role of SurE remains elusive, however the *E. coli* homologue is predicted to function in response to protein damage promoting survival under conditions of stress (Visick et al., 1998), whilst the homologue in *Thermotoga maritima* is proposed to have a role in intracellular remobilisation of phosphate during phosphorus starvation (Zhang et al., 2001). Rlv3841 *atvA* (acid tolerance and virulence) is homologous to *A. tumefaciens* virulence genes *acvB* (agrobacterium chromosomal virulence protein B) and *virJ,* which facilitate the translocation of T-DNA and effector molecules into host plant cells. (Groenewold et al., 2019). AcvB was subsequently shown to act as a lysyl-phosphatidylglycerol (LPG) hydrolase required for the maintenance of low levels of LPG in the bacterial membrane, a requisite for T-DNA transfer (Groenewold et al., 2019). Further to this, *R. tropici* AtvA has been shown to hydrolyse LPG conferring acid tolerance, though the reason for this is unclear (Vinuesa et al., 2003; Sohlenkamp et al., 2007). Plants are known to secrete organic acids and hydrogen ions into the rhizosphere, altering the pH, and so acid tolerance may play a key role in fitness during colonisation of the rhizoplane (Hinsinger et al., 2003). Bacteria can also alter the pH through the release of organic acids for example (Zifcakova, 2020)

Other genes contributing to Rlv3841 root colonisation during competition include pRL80031 (*mcpS*), encoding an MCP; RL3841, encoding a putative UDP-glucose 4-epimerase; and pRL70091 (*traI*), encoding a putative autoinducer synthesis protein.

UDP-glucose 4-epimerase catalyses the conversion of UDP-galactose to UDP-glucose and UDP-glucose 4-epimerase mutants have been shown to be impaired in LPS, EPS and CPS biosynthesis (Canter Cremers et al., 1990). The Rlv3841 genome encodes six other putative UDP-glucose 4-epimerase and so the requirement of RL3841 during competition with SBW25 is unclear. TraI is predicted to catalyse the production of a N-Acyl homoserine lactone (AHL). This gene was neutral for growth in the rhizosphere and colonisation of pea, lentil, *Lathyrus*, alfalfa and barley under single inoculation conditions (Chapter 4; Table S4.1). AHLs are a class of signalling molecules, the synthesis and perception of which enable bacteria to regulate gene expression in a cell-density dependent manner in a process termed quorum sensing (QS) (Fuqua et al., 1994; Cha et al., 1998). QS has been shown to regulate various processes in rhizobia, such as plasmid transfer, motility, polysaccharide production, biofilm formation and symbiotic interactions themselves (Swift et al., 2001). In the closely related strain *R. leguminosarum* A34, *traI* is located on the symbiotic plasmid where it immediately precedes the *trb* operon and has been shown to play a role in QS-regulated plasmid transfer (Danino et al., 2003; McAnulla et al., 2007). In *R. leguminosarum* A34, TriR induces the *traI-trb* operon in response to the *traI* produced AHL 3-oxo-octanoyl-L-homoserinelactone (3-O-C_8-HSL), resulting in plasmid transfer. In Rlv3841 however, *traR* is located on pRL8 and *traI* on prL7, not the symbiosis plasmid pRL10. No role in plasmid transfer has been reported to date for pRL70091 (*traI*), but sequence analysis suggests that it is functional and so it may mediate another, as of yet undefined role in cell-cell signalling particularly important during competition (Sánchez-Cañizares et al., 2018).

5.2.6. Identification of SBW25 genes contributing to growth in the rhizosphere and colonisation of pea and barley during single and co-inoculation with Rlv3841

For SBW25, 471 genes contributed to pea rhizosphere fitness when inoculated alone and 579 genes when in competition with Rlv3841, whilst 593 genes contributed to barley rhizospheres fitness under single inoculation and 404 under co-inoculation (Fig.5.10). During colonisation of pea, 627 genes contributed when inoculated alone and 700 genes contributed during competition, whilst 461 genes promoted colonisation of barley under single inoculation and 491 genes under co-inoculation. Comparison of genes contributing to rhizosphere growth and root colonisation under each condition allowed identification of those that are rhizosphere-progressive (Fig. 5.10). As expected, most genes were assigned neutral fitness values, with no more than 3.4% of genes displaying strong fitness contributions under any of the conditions (Table 5.5).

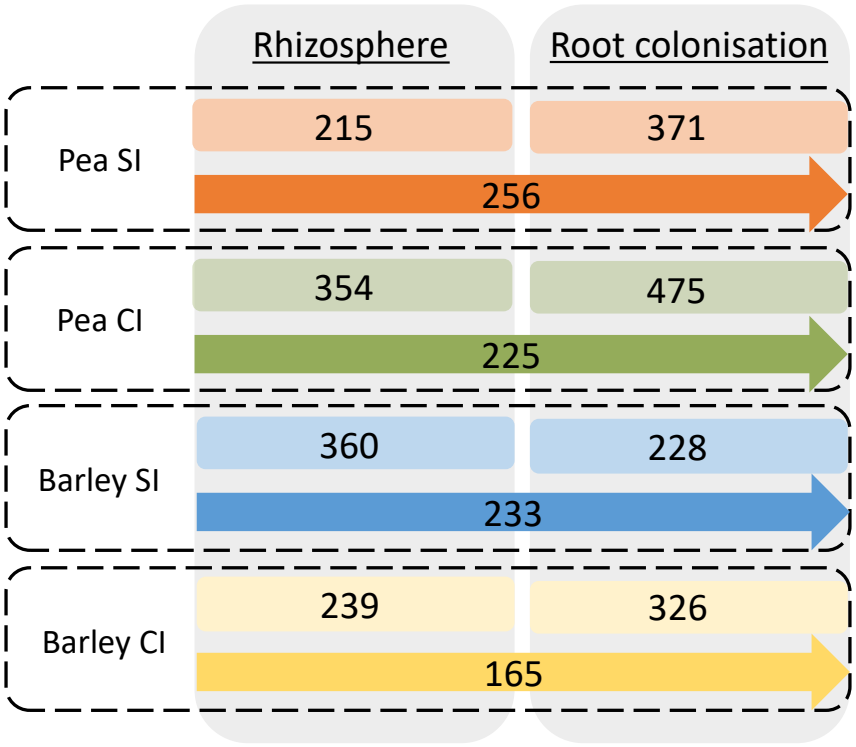


Figure 5.10. Number of SBW25 genes contributing to rhizosphere fitness and root colonisation of pea and barley under single and co-inoculation with RIv3841. All genes have fitness values less than -1. Single inoculation (SI) and co-inoculation (CI). The number of rhizosphere-progressive genes are shown in the arrow. Genes are listed in Tables S5.7 (pea SI), S5.8 (pea CI), S5.9 (barley SI) and S5.10 (barley CI).

Table 5.5. Distribution of SBW25 genes fitness values for pea and barley rhizosphere growth and root colonisation during single and co-inoculation with Rlv3841.

	Rhizosphere				Root colonised			
	Pea SI	**Pea CI**	**Barley SI**	**Barley CI**	**Pea SI**	**Pea CI**	**Barley SI**	**Barley CI**
Strongly disadvantage	100 (2.1%)	162 (3.4%)	91 (1.9%)	38 (0.8%)	159 (3.4%)	150 (3.2%)	47 (1.0%)	47 (1.0%)
Mildly disadvantage	371 (7.9%)	417 (8.8%)	502 (10.6%)	366 (7.8%)	468 (9.9%)	550 (11.6%)	414 (8.8%)	444 (9.4%)
Neutral	4,113 (87.1%)	3,953 (83.8%)	3,998 (84.7%)	4,058 (86.0%)	3,672 (77.8%)	3,652 (77.4%)	4,006 (84.8%)	3,782 (80.1%)
Advantaged	136 (2.9%)	188 (4.0%)	129 (2.8%)	258 (5.4%)	421 (8.9%)	368 (7.8%)	253 (5.4%)	447 (9.5%)

Number of strongly disadvantaged genes (fitness value < -2), mildly disadvantaged genes (fitness value between -1 and –2), neutral (fitness value between 1 and -1) and advantaged genes (fitness value > 1) are shown. Percentage of the total 4,720 genes for which fitness values were assigned is shown in brackets. Single inoculation (SI) and co-inoculation (CI). Genes are listed in Table S5.3.

5.2.7. Common determinants of SBW25 rhizosphere growth and root colonisation

Comparison of genes required for a) growth in the rhizosphere and b) colonisation of pea and barley during single inoculation or competition with Rlv3841 revealed that 58 and 25 genes respectively were commonly required (Fig. 5.11).

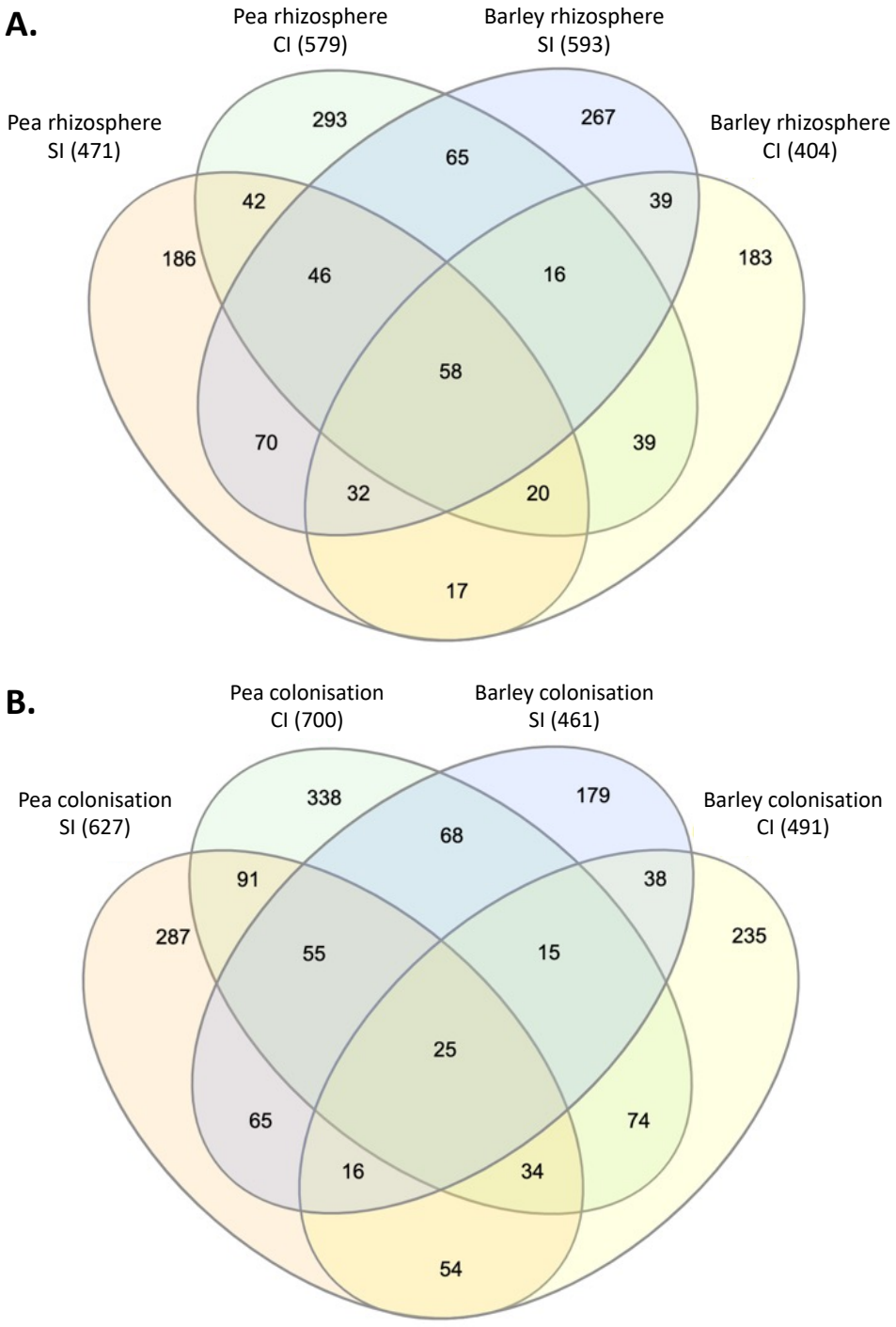


Figure 5.11. Venn diagram showing the number of SBW25 genes contributing to rhizosphere fitness and root colonisation under single and co-inoculation with Rlv3841. A. Rhizosphere growth. **B.** Root colonisation. All genes have a fitness value < -1. Single inoculation (SI) and co-inoculation (CI).

Comparison of the 58 genes commonly required for rhizosphere growth and 25 genes contributing to colonisation under all conditions revealed that 12 genes were rhizosphere-progressive under all conditions (Fig. 5.12 and Table S5.11).

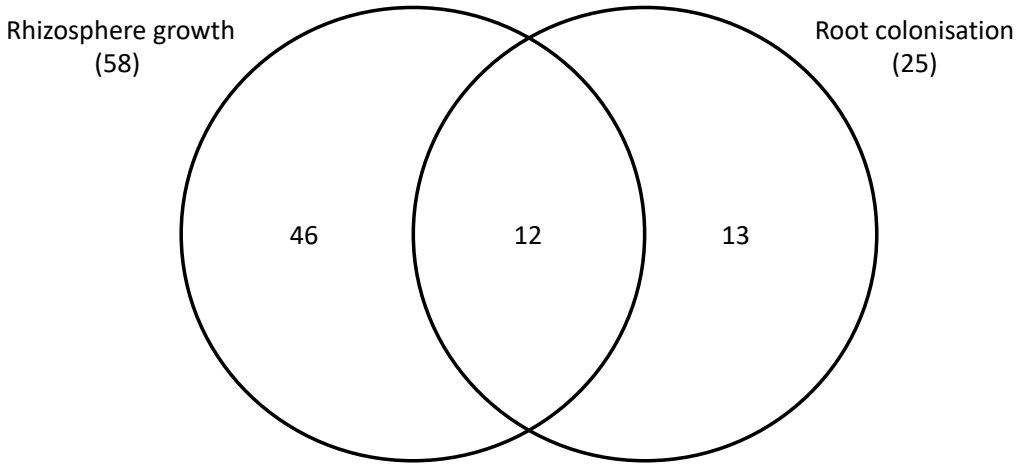


Figure 5.12. Common determinants of SBW25 pea and barley rhizosphere growth and root colonisation required under single inoculation and co-inoculation with Rlv3841. Comparison of 58 genes commonly required for SBW25 (pea and barley) rhizosphere growth and 25 genes commonly required for (pea and barley) root colonisation revealed that 12 SBW25 genes were rhizosphere-progressive under all tested conditions. See full gene list in Table S5.11.

The 46 genes contributing to rhizosphere growth and 12 genes identified as rhizosphere-progressive will be discussed collectively to give an overview of genes important for SBW25 rhizosphere fitness (Table S5.11). Amino acid biosynthesis was a clear prerequisite for rhizosphere fitness with genes involved in the biosynthesis of leucine (PFLU4193; *leub1*, PFLU4196; *leuC and* PFLU5050; *leuA*), isoleucine/valine (PFLU5797; *ilvD*), methionine (PFLU4182; *metZ*, PFLU5721; *metF* and PFLU5766; *metX*), serine (PFLU0504; *serB*), proline (PFLU5168; *proB*), histidine (PFLU0328; *hisD*) and glutamine (PFLU0415; *gltD*) required. *P. fluorescens* amino acid auxotroph's are impaired in colonisation of tomato rhizospheres demonstrating that whilst present in root exudates, amino acids are often not available at sufficient levels to support growth (Simons et al., 1997). Purine ribonucleotide biosynthesis also contributed to rhizosphere growth with the requirement of PFLU4183 (*purF*), PFLU0612 (*purH*) and PFLU5034 (*purL*), as did the biosynthesis of the cofactor's thiamine (vitamin B_1) (PFLU5773; *thiG* and PFLU2925; *thiF*) and biotin (PFLU5614; *bioF*).

As seen with Rlv3841, glycolysis/gluconeogenesis appears to be a determinant of rhizosphere fitness with the requirement of PFLU5243 (*pgi*) encoding glucose-6-phosphate isomerase, a key enzyme of these pathways. Energy production was also of importance in the rhizosphere with the requirement of PFLU0842 (*petB*) encoding cytochrome B of the cytochrome b-c1 complex involved in ATP synthesis/energy production and a cbb3-type cytochrome c oxidase subunit encoded by PFLU4559 (*ccoO1*).

Genes involved in cell wall biogenesis were critical and included PFLU5421, a putative peptidoglycan like protein; PFLU0478, a putative glycosyl transferase; and three putative lipoproteins encoded by PFLU0908, PFLU4495 and PFLU5420, the latter of which shares 64% identity to RlpA *of P. aeruginosa*. RlpA is an endolytic peptidoglycan transglycosylase required for efficient cell division and maintenance of rod shape (Jorgenson et al., 2014). LPS is a key constituent of the outer-membrane, and it is anchored into the membrane via its lipid A terminal. Two lauroyl acyltransferases HtrB (PFLU0013) and LpxL (PFLU4368) involved in lipid A biosynthesis were of common importance. These enzymes play important roles in maintenance of membrane integrity with *P. aeruginosa htrB* mutants displaying increased membrane permeability and therefore susceptibility to antimicrobial peptides (Hittle et al., 2015). Three genes belonging to the *mla* pathway (PFLU0888; *mlaF* and PFLU0890-91; *mlaDC*) involved in the transport of phospholipids were also required (Hughes et al., 2019), as were cell division proteins FtsX (PFLU5778) and MinC (PFLU4369).

Genes with regulatory roles include PFLU4418 (*fleN*), encoding a ParA family ATPase shown to negatively regulate flagella number in *P. aeruginosa* (Dasgupta et al., 2000) and PFLU1467 (*algU*), encoding a RNA polymerase sigma factor shown to positively regulate genes involved in alginate biosynthesis. AlgU controls EPS alginate production in *P. fluorescens* and plays a key role in adaptation to both desiccation and osmotic stress, conditions likely encountered in the harsh environment of the rhizosphere (Schnider-Keel et al., 2001). FleN is predicted to antagonise FleQ, a transcriptional activator of numerous genes involved in flagella biosynthesis and assembly, thereby regulating flagella number (Dasgupta et al., 2000). Deletion of *fleN* in *P. aeruginosa* results in non-motile multiflagellate cells. DksA (PFLU5233), a TraR-

family transcriptional regulator was also of importance and is implicated in the stringent response (SR). SR is a phenomenon in which nutrient or environmental stress results in the increased synthesis of (p)ppGpp, typically by RelA or SpoT, which in turn acts on RNA polymerase (RNAP) to regulate the transcription of various genes. Often the transcription of genes involved in amino acid biosynthesis and transport are induced whilst those involved in DNA replication are repressed. The result is a complete reprogramming of cellular transcription leading to changes in bacterial metabolism and behaviour that promote survival under challenging conditions. DksA has been shown to also bind RNAP, increasing the sensitivity of RNAP to (p)ppGpp thereby resulting in the enhanced expression or repression of genes implicated in the SR (Gourse et al., 2018). A putative two-component sensor kinase encoded by PFLU0958 was also of importance. Immediately upstream of PFLU0958 is the two-component response regulator CusR (PFLU0959) known to play a role in copper homeostasis and so PFLU0958 probably encodes a homologue of CusS. However, it only shares 38% identity to *E. coli* CusS (Munson et al., 2000).

Thirteen genes were commonly implicated in colonisation under all conditions tested (Table S5.11). These included PFLU4188 (*trpF*), involved in tryptophan biosynthesis; PFLU5546 (*anmK*), encoding an anhydro-N-acetylmuramic acid kinase involved in cell wall recycling (Bacik et al., 2011); PFLU0932 and PFLU5702, encoding putative lipoproteins; and PFLU4746 (*rsmA*), a regulator of secondary metabolism. RsmA is an RNA-binding post transcriptional regulator that has been shown to regulate a wide variety of bacterial processes in *Pseudomonas* including metabolism, adherence, chemotaxis and motility, virulence via T3SS, polysaccharide biosynthesis, biofilm formation, stress responses and iron acquisition (Song et al., 2015; Sobrero and

Valverde, 2020; Liu et al., 2021). Further to this SBW25 *rsmA* shares 100% identity to *rsmA* of *P. fluorescens* SS101, which has been shown to regulate lipopeptide biosynthesis in addition to controlling genes involved in motility, competition and survival in plant rhizospheres (Song et al., 2015). A GntR family transcriptional regulator encoded by PFLU2347 was also of importance, though its target(s) are unknown.

Surprisingly no genes involved in chemotaxis or motility were commonly required for root colonisation. The SBW25 genome contains two flagella-mediated chemotaxis clusters namely *che1* and *che2* in addition to a pili-mediated chemotaxis pathway, identified through high synteny with those in the closely related species *P. aeruginosa* (Fig.5.13) (Güvener et al., 2006; Garvis et al., 2009; Sampedro et al., 2014).

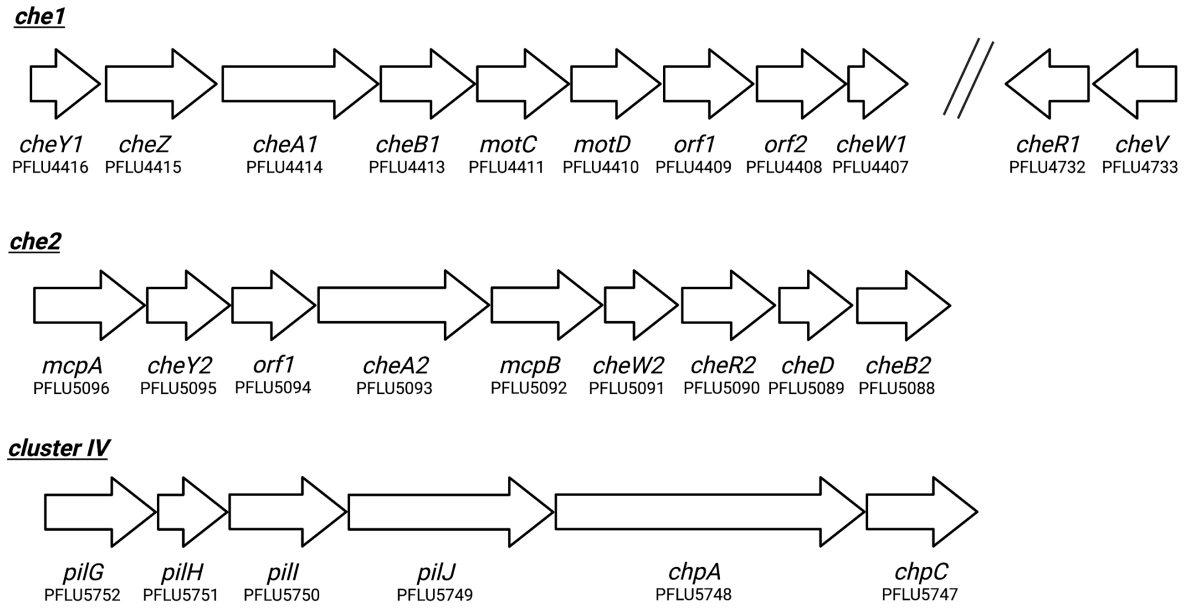


Figure 5.13. Chemotaxis *gene* clusters in *P. fluorescens* SBW25. The SBW25 chemotaxis clusters were inferred based on homology and high synteny to that of

other *P. fluorescens* strains and the closely related species *P. aeruginosa* (Güvener et al., 2006; Garvis et al., 2009; Sampedro et al., 2014).

Analysis of gene fitness values for the chemotaxis clusters revealed that the *che1* cluster is of greater importance in promoting colonisation during competition with Rlv3841, with an average fitness value of -0.71 for colonisation of pea and -0.64 for colonisation of barley (Fig.5.14A). In comparison, when inoculated alone the average fitness value of the *che1* cluster was -0.04 for pea and 0.39 for barley. Accordingly, the associated flagella genes required for flagella biosynthesis and assembly showed the same trend as the *che1* cluster (Fig. 5.14B). Numerous genes of the *che2* and type IV clusters also contributed to colonisation, though there was no clear increased demand for these clusters during competition with Rlv3841 (Fig. 5.14C and 5.14D).

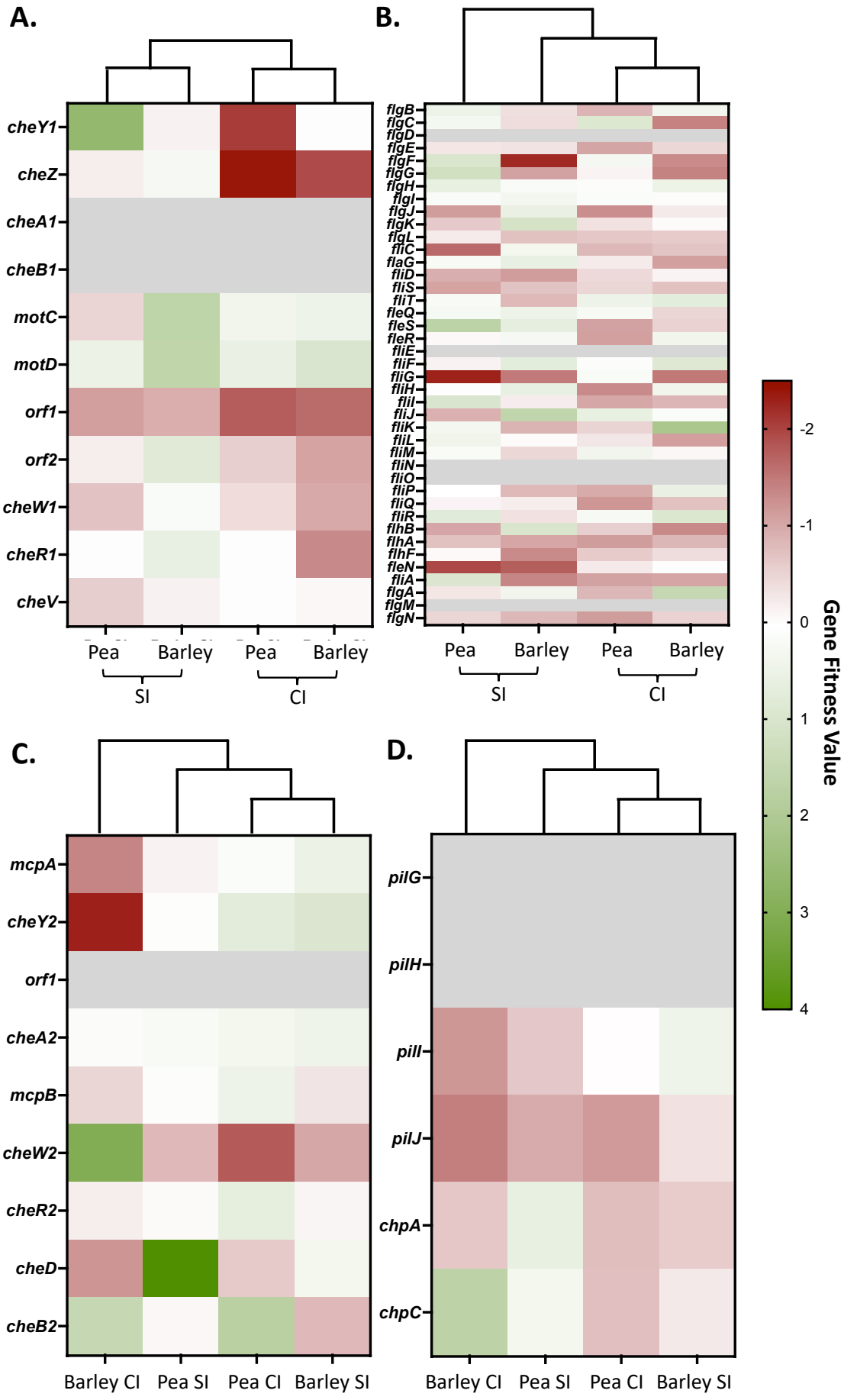

A.
B.
C.
D.
cheY1
cheZ
cheA1
cheB1
motC
motD
orf1
orf2
cheW1
cheR1
cheV
flgB
flgC
flgD
flgE
flgF
flgG
flgH
flgI
flgJ
flgK
flgL
fliC
flaG
fliD
fliS
fliT
fleQ
fleS
fleR
fliE
fliF
fliG
fliH
fliI
fliJ
fliK
fliL
fliM
fliN
fliO
fliP
fliQ
fliR
flhB
flhA
flhF
fleN
fliA
flgA
flgM
flgN
Pea
Barley
Pea
Barley
SI
CI
Pea
Barley
Pea
Barley
SI
CI
mcpA
cheY2
orf1
cheA2
mcpB
cheW2
cheR2
cheD
cheB2
Barley CI
Pea SI
Pea CI
Barley SI
pilG
pilH
pilI
pilJ
chpA
chpC
Barley CI
Pea SI
Pea CI
Barley SI
Gene Fitness Value
-2
-1
0
1
2
3
4

Figure 5.14. Clustered heatmaps of SBW25 chemotaxis and flagella clusters. A. *che1* chemotaxis cluster. **B.** Genes implicated in flagella synthesis and assembly. **C.** *che2* chemotaxis cluster. **D**. Type IV cluster. Single inoculation (SI) and co-inoculation (CI). Cluster analysis was performed using pheatmap R package (Kolde, 2012). Genes from Table S5.3.

5.2.8. Genes promoting SBW25 rhizosphere growth and root colonisation during competition with Rlv3841

To identifying genes contributing to SBW25 competitiveness specifically during competition, genes previously identified with fitness values less than -1 for a) rhizosphere growth and b) root colonisation of pea and barley (Fig. 5.10 and Tables S5.7 to S5.10) were filtered to remove those with fitness values less than -0.42 during single inoculation for the respective condition and retain those with high confidence fitness predictions indicated by their associated *t*-score (fitness value < -2 and *t*-score < -3 or fitness value between -1 and -2 and *t*-score < -2.5). To ensure genes were uniquely required during competition, those with a fitness value less than -1 during single inoculation under the same condition (rhizosphere growth or root colonisation) but for the alternative plant species (pea or barley), were also removed. For rhizosphere fitness, pea had the highest requirement of competition specific genes (83 genes), followed by barley (63 genes) (Fig. 5.15). Comparison of these genes revealed that five were common among pea and barley. For root colonisation 100 genes contributed to the colonisation of pea specifically during competition with Rv3841 and 59 genes for barley, five of which were common to both (Fig. 5.15.).

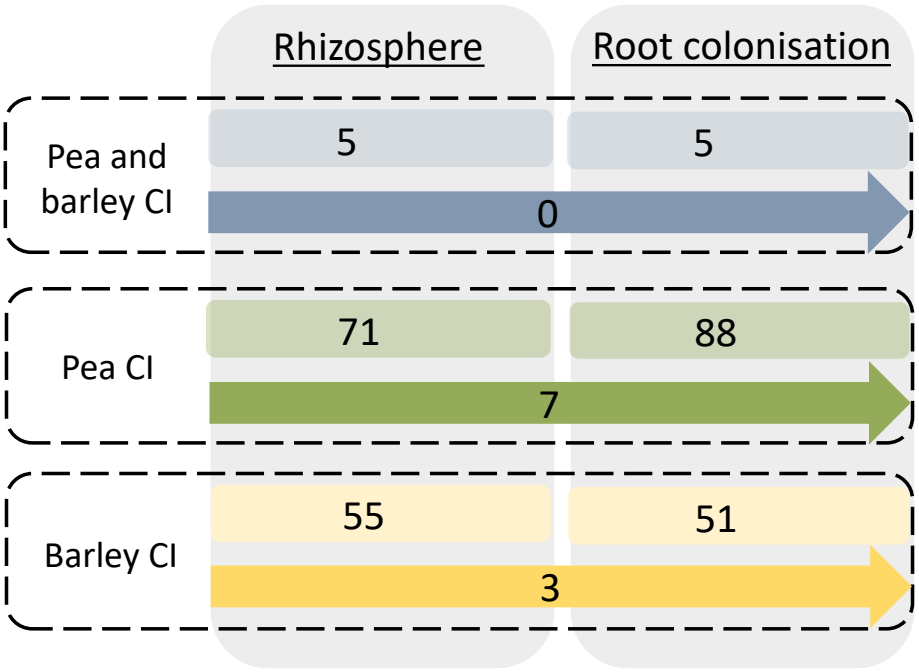


Figure 5.15. Number of SBW25 genes contributing to rhizosphere fitness and root colonisation of pea and barley during competition with Rlv3841 following manual curation. All genes have fitness values less than -2 and *t*-score < -3 or fitness value less than -1 but greater than -2 and *t*-score < -2.5 during competition with Rlv3841, a fitness value greater than -0.42 for the corresponding condition when inoculated alone, and a fitness vale greater than -1 when inoculated alone under the same condition but on the alternative plant species. The number of rhizosphere-progressive genes are shown on the arrow. Single inoculation (SI) and co-inoculation (CI). Genes are listed in Tables S5.12 (pea and barley CI), S5.13 (pea CI) and S5.14 (barley CI).

5.2.8.1. Common determinants of pea and barley rhizosphere growth and root colonisation during competition with Rlv3841

Five genes were identified as commonly required for growth in the rhizosphere of pea and barley when competing with Rlv3841 (Table S5.12). Four of these genes encode

hypothetical proteins of unknown function and other, PFLU0189 (*cysW*), a component of a sulfate ABC transporter. Colonisation of and persistence on the rhizoplane of pea and barley required five genes specifically during competition with Rlv3841. These were PFLU0249 (*betT*), encoding a transporter involved in the uptake of the osmotic protectants choline/betaine; PFLU1342 (*livF1*), a component of a branched chain amino acid transporter; PFLU5048 (*xseA*), a component of a exodeoxyribonuclease complex involved in the degradation of single-stranded DNA; PFLU1152, a putative alkyl hydroperoxide reductase that plays a role in protection against oxidative stress through the detoxification of peroxides; and PFLU1712, encoding a hypothetical protein of unknown function.

5.2.8.2. Genes contributing to growth in the rhizosphere of pea or barley during competition with Rlv3841

During competition with Rlv3841, SBW25 growth in the rhizosphere of pea and barley was promoted by 78 and 58 genes, respectively (Fig 5.15, Tables S5.13 and S5.14). Of the genes implicated in pea and barley rhizosphere fitness, 32% and 34%, respectively, encoded hypothetical proteins. Functional analysis of these 78 genes (Table S5.13) revealed that genes with unknown functions accounted for the largest functional category, followed by amino acid biosynthesis and transport, general function predictions and transcription (Fig 5.16A). Functional analysis of the 59 genes required for growth in the barley rhizosphere during competition with Rlv3841 (Table S5.14) revealed a similar trend, with genes with unknown functions accounting for the largest proportion, followed by those involved in transcription, amino acid biosynthesis and transport, and carbohydrate transport and metabolism (Fig. 5.16B).

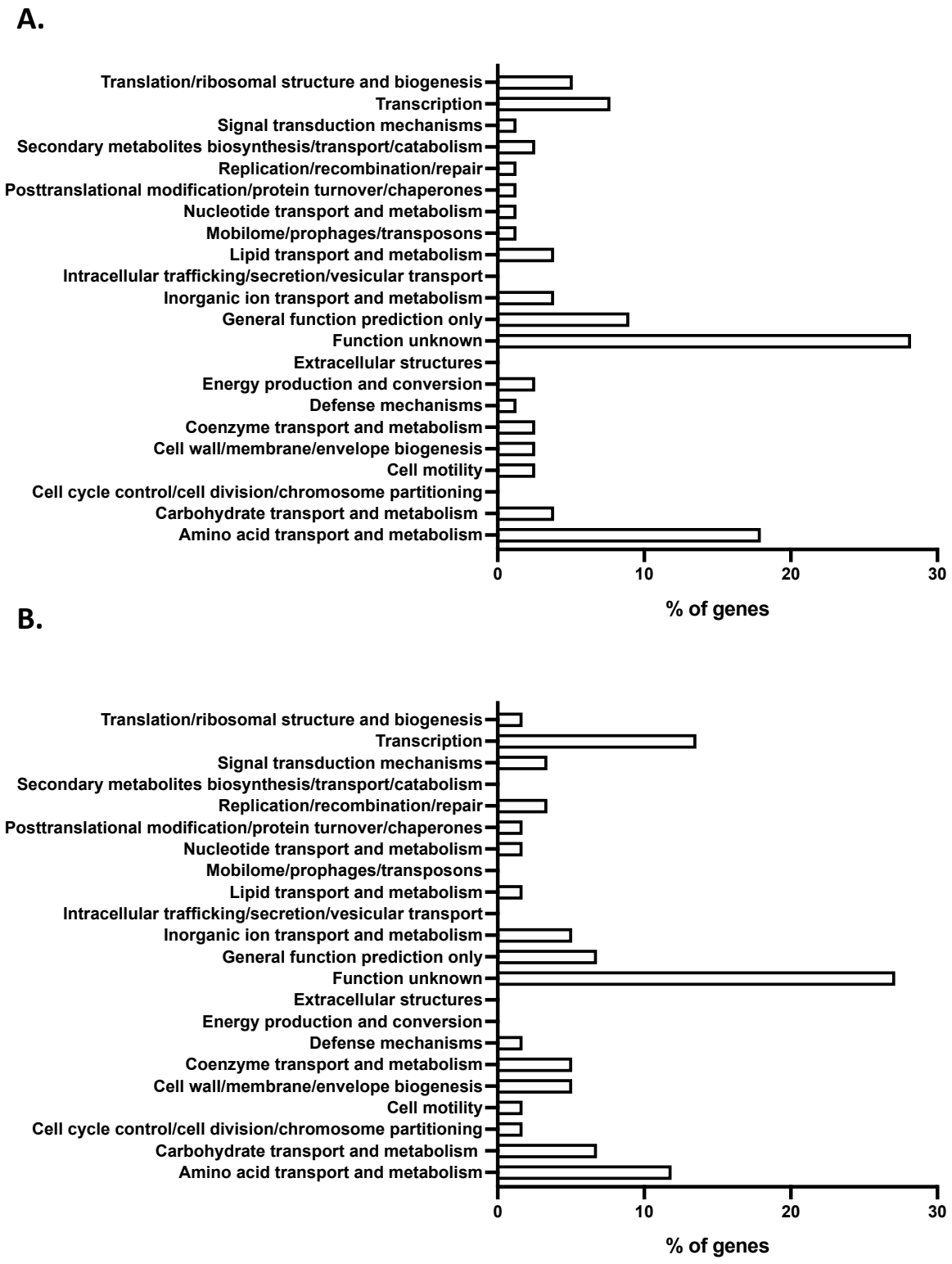


Figure 5.16. Functional classification of SBW25 genes contributing to growth in the rhizosphere of pea or barley during competition with Rlv3841. A. 78 genes contributing to competitive pea **r**hizosphere growth (Table S5.13). **B.** 59 genes

contributing to competitive barley rhizosphere fitness (Table S5.14). Functional classification based on COG annotation (Tatusov et al., 2000). Genes are listed in Tables S5.13 and S5.15.

Genes with functional roles in amino acid transport and metabolism required for growth in the pea rhizosphere during competition with Rlv3841 include *putA* (PFLU0451), which functions in proline catabolism by oxidising proline to glutamate for use as a carbon and nitrogen source; *sstT* (PFLU2139), encoding a transporter implicated in the import of serine and threonine; PFLU1632, encoding a component of a peptide ABC transporter; and PotF1 (PFLU5845), a putrescine binding SBP. Biosynthesis of aromatic amino acids also contributed to pea rhizosphere fitness during competition with Rlv3841, as demonstrated by the requirement of two deoxy-D-arabino-heptulonate-7-phosphate synthase (DHAPS) isoforms encoded by PFLU1704 (*aroG*) and PFLU2184 (*aroF*). DHAPs catalyse the first step in the biosynthesis of aromatic amino acids tryptophan, tyrosine and phenylalanine. The SBW25 genome encodes a third DHAP PFLU1723 (*aroH*), with *aroG*, *aroF* and *aroH* inhibited by the presence of phenylalanine, tyrosine and tryptophan, respectively. Both tyrosine and phenylalanine are present in pea rood exudates and so the requirement of *aroG* and *aroF* during competition with Rlv3841 suggests that uptake by Rlv3841 may reduce availability for SBW25 (Pini et al., 2017). Growth in the barley rhizosphere required PFLU4031, a cysteine binding SBP; PFLU5636, encoding a component of PAAT system; an amino acid permease encoded by PFLU3148; *asnB* (PFLU4320), involved in asparagine biosynthesis; and *srpH* (PFLU0248), involved in the biosynthesis of cysteine. PFLU1987 and PFUL3211 putatively involved in the biosynthesis of cobalamin were

important in the barley rhizosphere. Specific requirements for these genes during competition with Rlv3841 likely reflects differences in amino acid and vitamin availability in the rhizosphere due to the presence of Rlv3841.

Transcription factors accounted for 8% and 14% of genes required for growth in the pea and barley rhizosphere, respectively, during competition with Rlv3841. The targets of these TFs are predominantly unknown, with the exception of PFLU4157, a repressor of the adjacent operon involved in ribose transport and utilisation, which was required for growth in the pea rhizosphere (Liu et al., 2015).

Genes involved in carbohydrate transport and metabolism accounted for 4% of genes contributing to pea rhizosphere fitness and 7% of genes contributing to barley rhizosphere fitness during competition with Rlv3841. Those with known functions contributing to pea rhizosphere growth include PFLU2745, a mannitol binding SBP, and PFLU1953, encoding a MFS transporter (Table S5.13). For barley rhizosphere growth PFLU3734, encoding a sugar efflux transporter; PFLU2540, encoding an MFS transporter; and PFLU2743 (*mtlG*), encoding a component of a mannitol/sorbitol transporter were of importance (Table S5.14).

Two genes had functional roles in cell motility specifically for growth in the pea rhizosphere during competition with Rlv3841. These were PFLU1096, a putative exported fimbrial protein and PFLU1219 (*wspA*), a putative MCP. WspA, which has been best characterised in *P. aeruginosa*, forms part of the *Wsp* (wrinkly spreader) complex which is a chemotaxis-like system that regulates c-di-GMP biosynthesis in response to surface sensing via the chemoreceptor WspA (O'Connor et al., 2012).

Increased levels of the intracellular signalling molecule c-di-GMP promotes biofilm formation and reduced motility. Biofilm formation is a lifestyle strategy employed by bacteria in hostile environments, such as the rhizosphere, to promote survival (Angus and Hirsch, 2013). However, further analysis of gene fitness values for the entire *wsp* operon revealed no clear requirement for growth in the rhizosphere of pea as opposed to barley, or growth during competition with Rlv3841 as opposed to growth during single inoculation (Fig. 5.17). Most genes showed neutral phenotypes, with fitness values between 1 and -1. PFLU4409 was the only gene with a putative role in cell motility important in the barley rhizosphere during competition with Rlv3841. Whilst it is annotated as a putative chromosome partitioning protein, it forms part of the *che1* operon, though its function is yet to be demonstrated (Fig 5.13).

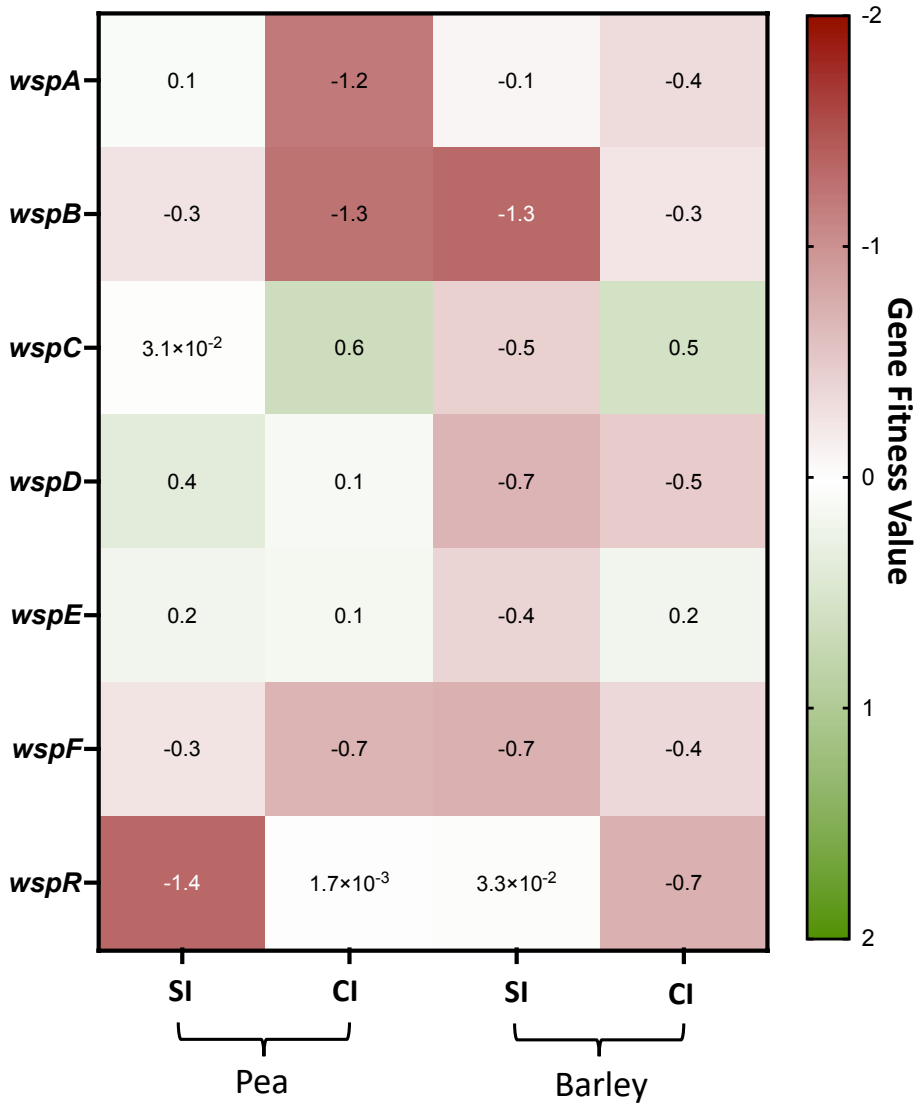


Figure 5.17. Gene fitness values for the *wsp* cluster of SBW25 during growth in the rhizosphere of pea and barley under single and co-inoculation with Rlv3841. Single inoculation (SI) and co-inoculation (CI)**.** Genes from Table S5.3.

Other genes with known function identified as contributing to fitness in the pea rhizosphere during competition with Rlv3841 include PFLU0261 (*rimk*) and PFLU0725 (*rspD*). Rimk is a ribosomal modification protein thought to couple environmental stresses to ribosomal changes thereby resulting in an adaptive remodelling of the

proteome. SBW25 *rimK* mutants are compromised in competitive wheat rhizosphere growth and motility, with *rimK* expression found to be induced during the initial stages of wheat rhizosphere colonisation suggesting that expression is induced by signal(s) present in early wheat rhizospheres (Little et al., 2016). These results suggest that this signal(s) is also present in the pea rhizosphere and that subsequent remodelling of the proteome mediated by *rimk* is important for growth during competition with Rlv3841. RspD belongs to a rhizosphere-expressed gene cluster that resembles the T3SS of pathogenic bacteria. Pathogenic bacteria utilise T3SSs to export proteins capable of supressing host cell immune responses and modulate their metabolism, resulting in release of nutrients for bacterial uptake. The *rsp* gene cluster is conserved in many plant growth promoting rhizobacteria and in SBW25 it has been shown to be specifically induced in plant rhizospheres (Preston et al., 2001; Jackson et al., 2005). However, it lacks genes encoding integral components of the T3SS that are essential for its functionality in protein secretion and so its function remains elusive. Further analysis of the *rsp* cluster under all tested conditions revealed that, apart from growth in the pea rhizosphere during single inoculation, the *rsp* cluster seems to be of greater importance during competition (Fig. 5.18). This is consistent with previous findings that mutation of the *rsp* regulator RspL led to a significant reduction in competitive colonisation of sugar beet root tips, collectively indicating a role during competition in the rhizosphere (Jackson et al., 2005).

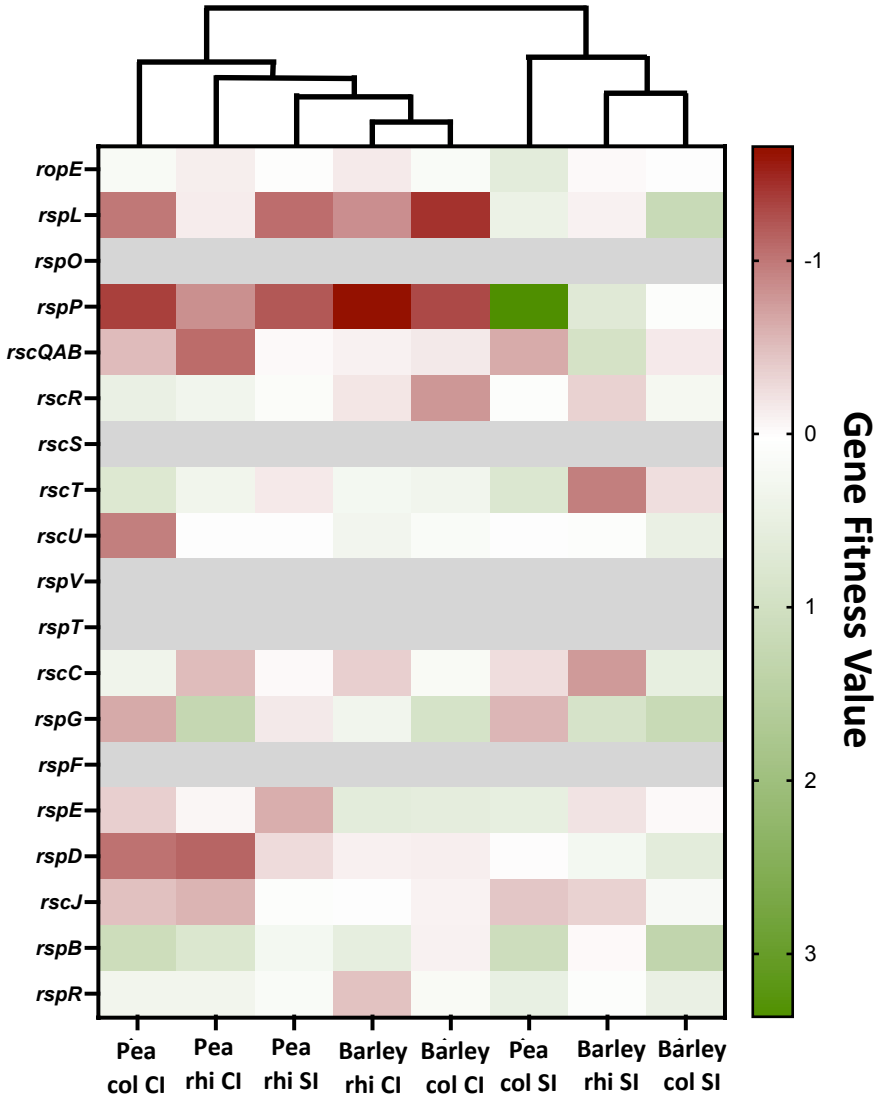


Figure 5.18 Gene fitness values for SBW25 *rsp* operon during growth in the rhizosphere and colonisation of pea and barley under single and co-inoculation with Rlv3841. Single inoculation (S) and co-inoculation (CI). Cluster analysis was performed using pheatmap R package (Kolde, 2012). Genes from Table S5.3.

5.2.8.3. SBW25 genes contributing to the colonisation of pea or barley during competition with Rlv3841

During competition with Rlv3841, SBW25 colonisation of pea and barley was promoted by 88 and 51 genes, respectively (Fig 5.15, Table S5.13 and S5.14). For pea and barley 27% and 29% of genes, respectively, encode hypothetical proteins. Functional analysis of these genes revealed that for both pea (Fig. 5.19A) and barley (Fig. 5.19B), genes with unknown functions accounted for the largest functional category, followed by amino acid biosynthesis and transport and general function predictions and transcription. This is the same trend seen for genes important in rhizosphere growth (Fig. 5 .16) and highlights the lack of knowledge regarding genetic determinants involved in rhizosphere growth and root colonisation.

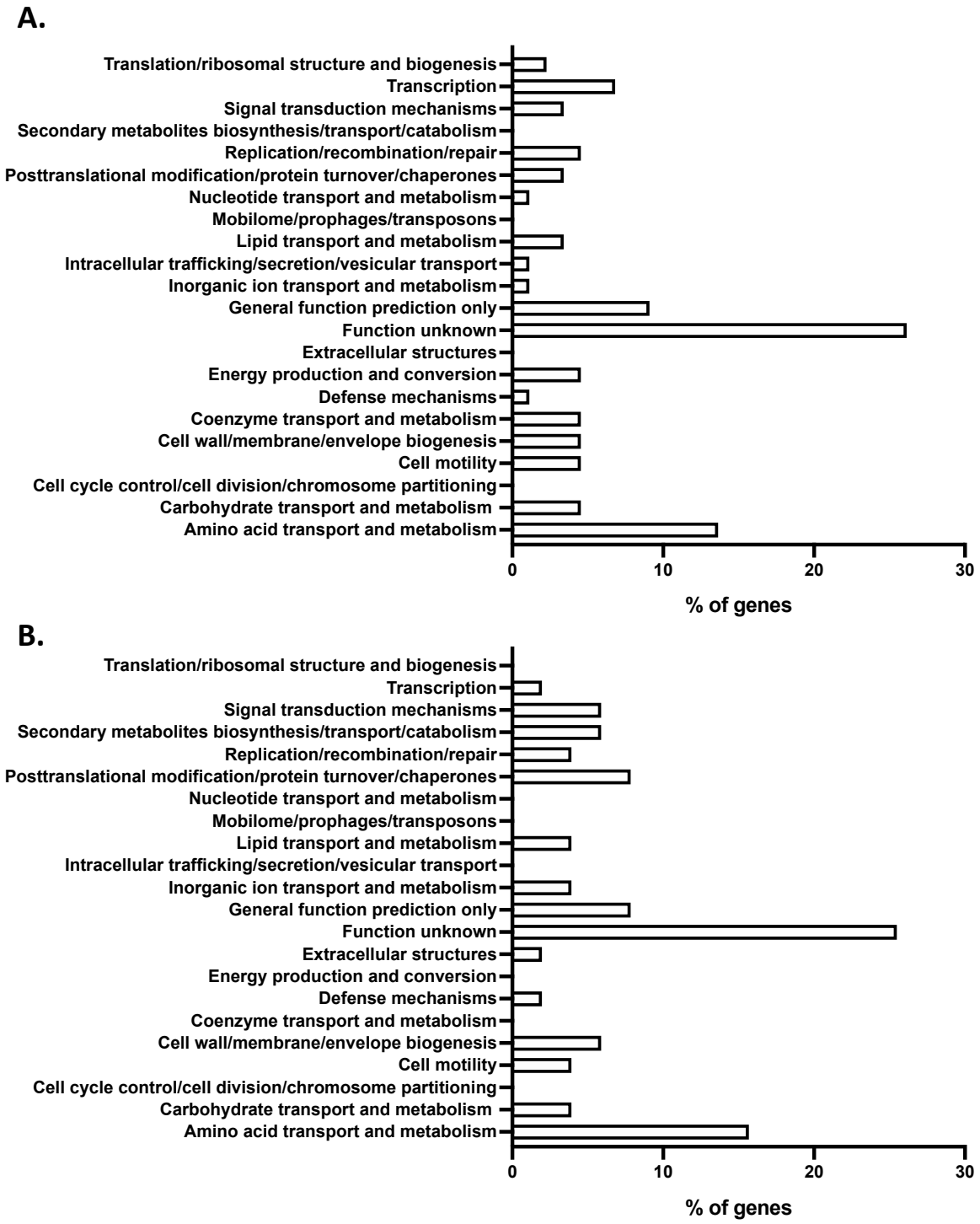


Figure 5.19. Functional classification of SBW25 genes contributing to pea or barley root colonisation during competition with Rlv3841. A. 88 genes (colonisation specific) contributing to pea root colonisation (listed in Table S5.13). **B.** 51 genes (colonisation specific) contributing to barley root colonisation (listed in

Table S5.14). Functional classification based on COG annotation (Tatusov et al., 2000).

Most genes within the functional category of amino acid biosynthesis and transport were involved in import of amino acids rather than their biosynthesis. For pea, a GABA transporter GabP (PFLU0315), a serine transporter encoded by PFLU3490 and three PAAT SBPs (encoded by PFLU0230, PFLU2617 and PFLU5633) were of importance. For barley, a branched-chain amino acid binding SBP encoded by PFLU1346 (*livJ1*) contributed to fitness, along with PFLU2943, encoding a SBP of HAAT class transporter, and two genes, PFLU0123 and PFLU0203, both encoding taurine dioxygenases.

Carbohydrate transport also contributed to root colonisation with the requirement of PFLU4844, encoding a putative glucose transporter, and PFLU2096, encoding a putative MFS transporter important for pea colonisation. A putative MFS transporter encoded by PFLU4106 contributed to barley colonisation. In terms of carbohydrate metabolism, PFLU3732 (*tal*), encoding a putative transaldolase, promoted pea root colonisation during competition with Rlv3841. Transaldolases are one of the major enzymes of the non-oxidative branch of the PPP and whilst their mutation often results in relatively mild phenotypes due to the fact that they are not required to produce any major metabolites, *S. meliloti tal* mutants are found to be growth-impaired when cultured on carbon sources that feed directly into the PPP (Stincone et al., 2015; Hawkins et al., 2017). For barley colonisation, PFLU0030 encoding a 2-keto-3-deoxygalactonate kinase involved in galactose metabolism was important.

Genes with known regulatory roles important for colonisation during competition with Rlv3841 include PFLU4441, which promoted pea colonisation, and PFLU4895 (*tyrR*), which contributed to barley colonisation during competition with Rlv3841. PFLU4441 shares 80% identity to *P. aeruginosa fleR*. FleR, along with FleS, forms a two-component regulatory system that positively regulates expression of genes involved in flagellar synthesis. These include genes encoding the basal body-rod, L ring, hook, hook-cap scaffold and hook-filament junction proteins, thereby affecting traits such as adhesion, motility and swarming (Dasgupta et al., 2003). PFLU4442 encodes a transcriptional regulator with 73% identity to *P. aeruginosa fleS* and this gene also had a fitness value less than -1 specifically for pea colonisation under competitive conditions but the associated *t*-score was -2.08 and therefore this gene was excluded during filtering. Collectively this suggests that flagella-mediated chemotaxis and motility is of particular importance for pea colonisation during competition with Rlv3841. The transcriptional regulator TyrR has been best characterised in *E. coli* where, in response to the presence of aromatic amino acids, it has been shown to represses the expression of genes involved in aromatic amino acid biosynthesis and transport (Pittard and Davidson, 1991). However, in *A. brasilense*, TyrR has also been shown to regulate genes involved in EPS production and biofilm development (Jijón-Moreno et al., 2019).

Transcriptional regulators often regulate the expression of adjacent genes and so those with unknown targets were assessed to see if adjacent genes also showed a phenotype. PFLU3945, identified as important for colonisation of pea during competition with Rlv3841, was found to be immediately upstream of PFLU3946, encoding a copper tolerance protein. PFLU3946 was assigned a fitness value less

than -1 for colonisation of pea during competition with Rlv3841 only, but was excluded due to the low associated *t*-score (-1.4) (Table S5.3). Collectively, this suggest that PFLU3945 may regulate transcription of PFLU3946, and that copper tolerance is therefore important for competitive colonisation of pea.

Genes with roles in chemotaxis and motility important for pea colonisation during competition with Rlv3841 include two putative MCPs encoded by PFLU2262 and PFLU2358, in addition to a putative flagella hook protein encoded by PFLU4728 (*flgE*) and a twitching mobility protein encoded by PFLU5762 (*pilT*). FlgE connects the flagella basal body to the flagella filament, whilst PilT is an ATPase essential for type IV pilus retraction required for twitching motility. *P. aeruginosa pilT* mutants overexpress pili but lose twitching motility (Comolli et al., 1999). For barley colonisation, a MCP methyltransferase belonging to the *che1* cluster encoded by PFLU4732 (*cheR*) was important, as was a putative fimbriae biogenesis protein encoded by PFLU0409 and a putative chemotaxis-like protein encoded by PFLU5096. As previously discussed, flagella-mediated chemotaxis and motility appears to be of increased importance during competitive colonisation.

Other SBW25 genes specifically required for colonisation of pea during competition with Rlv3841 include PFLU1510 (*pobA*), which catalyses to formation of protocatechuate from 4-hydroxybenzoate; PFLU0876 (*sodA*), encoding a superoxide dismutase involved in oxidative stress responses; PFLU1133 (*glpM*), a membrane protein with a predicted role in alginate biosynthesis (Schweizer et al., 1995); PFLU4201, a putative filamentous adhesin; PFLU5109, a putative lipoprotein; and PFLU3550, a type VI secretion system (T6SS) secreted protein Hcp (secreted

cytotoxin). The T6SS of gram-negative bacteria, an effector translocation system that participates in a wide range of processes from antimicrobial activity to the uptake of metal ions, confers a competitive advantage against other bacteria (Gallique et al., 2017). Hemolysin-coregulated proteins (Hcp) are extracellular components of the T6SS that structurally form the inner tube of the contractile sheath and serve as chaperones or receptors for various T6SS effectors (Silverman et al., 2013). In *P. fluorescens* Hcp proteins have been shown to have a role in bacterial competition, with *hcp* mutants displaying reduced EPS and flagellin accumulation, loss of motility and impaired competitor inhibition (Decoin et al., 2015).

Other genes contributing to the colonisation of barley during competition with Rlv3841 include several affecting cell wall structure such as PFLU2082, a putative glycosyltransferase; PFLU1233, a putative lipoprotein; and PFLU3668 (*gmd*), encoding a putative GDP-mannose 4,6-dehydratase which catalyses the first step in the pathway that converts GDP-D-mannose to GDP-L-fucose, a key component of glycoproteins and glycolipids (Sturla et al., 1997). PFLU1377 (*hscC),* encoding a Hsp70 chaperone shown to confer resistance to cadmium ions and UV in *E.coli* (Kluck et al., 2002); PFLU2791, encoding a nitrate/sulfonate/bicarbonate binding SBP; and PFLU0648, encoding a pilus assembly protein CpaB, also contributed to barley colonisation during competition with Rlv3841. PFLU0648 forms part of the *tad* (tight-adherence) locus, which encodes a type II secretion system required for the assembly of adhesive Flp (fibril low-molecular-weight) pili. Genes encoded in the *tad* locus have been shown to be important in several processes including biofilm formation, colonisation, and pathogenesis of certain bacteria (Tomich et al., 2007). Comparison of this region to that of *P. aeruginosa* allowed for putative annotation of the SBW25

tad locus (Fig. 5.20) (Bentzmann et al., 2006). Further analyses of these genes revealed that *pilA* (PFLU0649), *cpaF* (PFLU0645) also had a fitness value of less than -1 for barley root colonisation during competition, and *tadC* (PFLU0643) had a fitness value of -0.97 (Fig. 5.21). The average fitness value of this operon was -0.6 for barley during competition with Rlv3841 but greater than -0.12 for colonisation under all other conditions. However, *flpA*, encoding the major structural protein of the Flp pili, had a fitness value less than -1 for colonisation of pea and barley when SBW25 was inoculated alone, suggesting a general importance of the Flp pili during root colonisation.

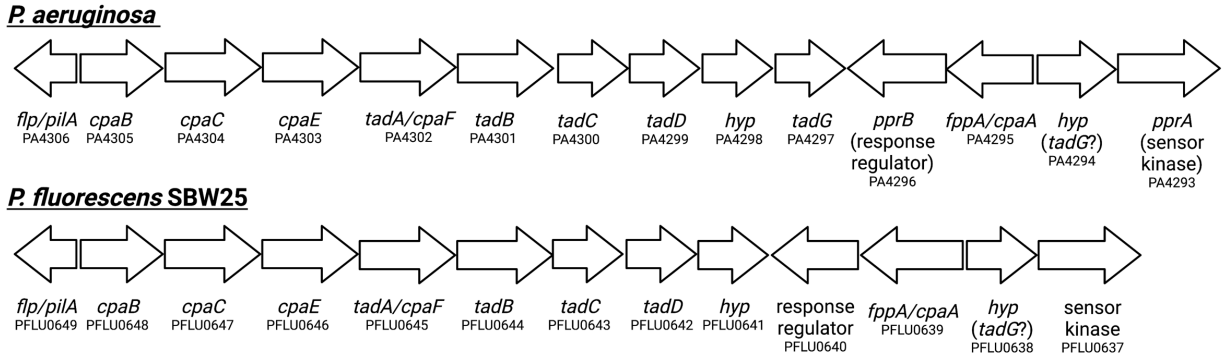


Figure 5.20. Pseudomonas *tad* gene clusters. The SBW25 *tad* gene cluster was inferred based on homology and high synteny to that of closely related *P. aeruginosa* (Bentzmann et al., 2006).

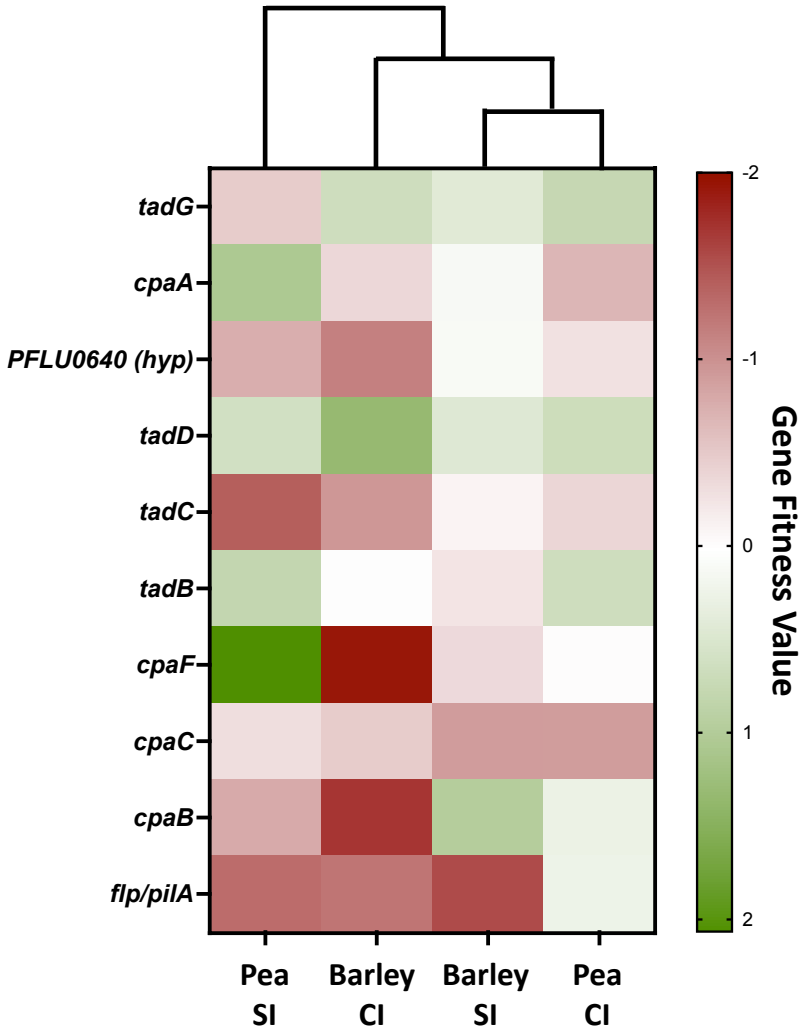


Figure 5.21. Gene fitness values for the SBW25 *tad* cluster for the colonisation of pea and barley during single and co-inoculation with Rlv3841. Single inoculation (SI) and co-inoculation (CI). Cluster analysis was performed using pheatmap R package (Kolde, 2012). Genes from Table S5.3.

5.3. Conclusion

Rhizosphere growth and root colonisation are complex processes to which many bacterial traits contribute. To date, factors identified to be involved in these processes include amino acid biosynthesis, vitamin production, chemotaxis, motility, flagella, Type IV pili, major outer membrane proteins, unipolar polysaccharides, cellulose fibrils, LPS and EPS, among others (Simons et al., 1996; Streit et al., 1996; Simons et al., 1997; Knights et al., 2021). However, these studies have predominantly been conducted under sterile laboratory conditions. The aim of this chapter was to identify genes involved in rhizosphere growth and root colonisation during competition with another bacterial species. To facilitate this, a highly saturated barcoded transposon mutant library was generated in SBW25. The SBW25 barcoded transposon pool consisted of 684,486 mapped transposon mutants, each containing a unique barcode identifier. Comparison of these barcodes to those mapped in the Rlv3841 barcoded mutant pool revealed that less than 3% were common to both and so these were computationally removed from Rlv3841 and SBW25 mutant pools. The insertion density of the SBW25 transposon library was 61%, making it suitable for the determination of gene fitness values by BarSeq. As seen during the mapping of barcodes for the Rlv3841 in Chapter 3, a large proportion (70%) of unique barcodes present in the library failed to meet the primary mapping criteria of having at least five reads supporting its mapping to the genome. These barcodes and therefore the associated mutants were consequently excluded from the pool of mapped barcoded mutants. Further sequencing of the original mutant pool will result in the mapping of significantly more barcodes and therefore the retention of more sequencing data from BarSeq experiments.

For the first time, BarSeq was successfully used to simultaneously assign gene fitness values to two different species whilst in competition with each other. This was achieved by co-inoculating Rlv3841- and SBW25-barcoded mutant pools onto barley and retrieving them from the roots 7 dpi. Rlv3841 genes required for barley root colonisation during competition with SBW25 are summarised (Fig. 5.22 and Table S.4), highlighting the importance of transport, metabolism, and regulation. Variation in these requirements is likely to reflect the need of Rlv3841 to adapt to differences in resource availability during competition with SBW25.

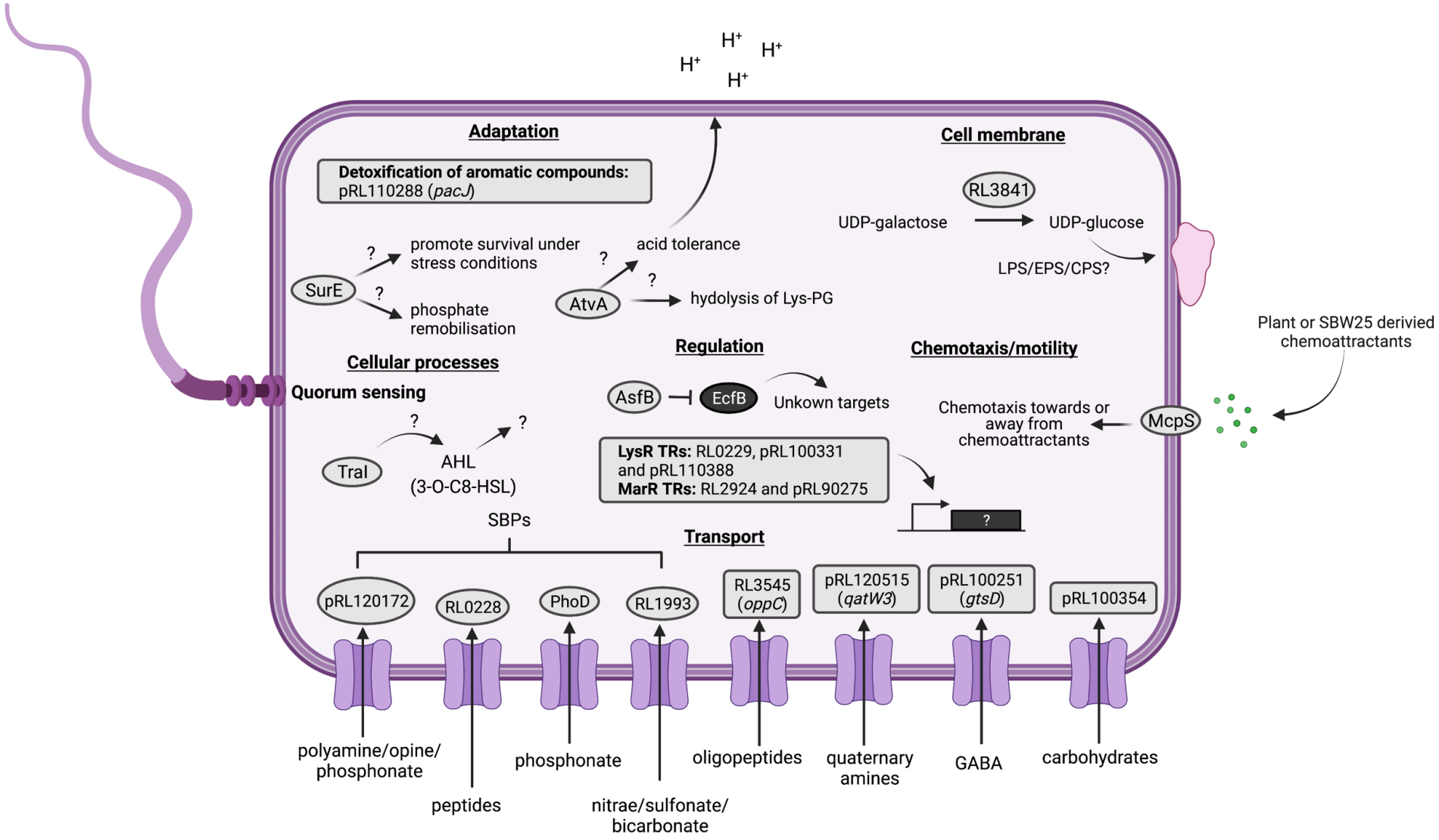

H+
H+
H+
H+
Adaptation
Detoxification of aromatic compounds:
pRL110288 (pacJ)
SurE
promote survival under stress conditions
phosphate remobilisation
AtvA
acid tolerance
hydolysis of Lys-PG
Cell membrane
RL3841
UDP-galactose
UDP-glucose
LPS/EPS/CPS?
Plant or SBW25 derivied chemoattractants
Cellular processes
Quorum sensing
TraI
AHL
(3-O-C8-HSL)
Regulation
AsfB
EcfB
Unkown targets
LysR TRs: RL0229, pRL100331 and pRL110388
MarR TRs: RL2924 and pRL90275
Chemotaxis/motility
Chemotaxis towards or away from chemoattractants
McpS
SBPs
Transport
pRL120172
RL0228
PhoD
RL1993
RL3545 (oppC)
pRL120515 (qatW3)
pRL100251 (gtsD)
pRL100354
polyamine/opine/ phosphonate
peptides
phosphonate
nitrae/sulfonate/ bicarbonate
oligopeptides
quaternary amines
GABA
carbohydrates

Figure 5.22. Summary of Rlv3841 genes contributing to barley root colonisation during competition with SBW25. Genes shown in black were not identified to contribute specifically to Rlv3841 colonisation of barley during competition with SBW25 but based on the literature interact with those that do. Genes are listed in Table S5.6.

Whilst attempts were also made to assess fitness contributions for both strains during competitive growth in the rhizosphere of pea and barley, as well as competitive colonisation of pea, this was only achieved for SBW25 and not Rlv3841. Fitness values were not calculated for Rlv3841 under these conditions because SBW25 accounted for a significantly larger proportion of the retrieved mutants in the samples and therefore the majority of sequenced barcode reads belonged to SBW25 mutants. Consequently, there were insufficient reads to allow for accurate calculation of Rlv3841 gene fitness values under these conditions.

The fact that less than 3% of barcodes were common to both the Rlv3841 and SBW25 mutant pools opens the possibility of expanding the number of strains simultaneously assayed under a given condition, by BarSeq, to larger SynComs. However, to achieve this in situations whereby each strain is not present in equal ratios after undergoing selection in the chosen test condition would require either a significantly greater sequencing depth or adaptation of the current system to enable species specific amplification of barcodes from the retrieved mutant pool prior to sequencing.

For SBW25, BarSeq was successfully used to identify genes contributing to SBW25 rhizosphere growth and root colonisation of pea and barley, both in single inoculation and in competition with Rlv3841. Comparison of genes important under each condition led to the identification of a core set of genes commonly required for rhizosphere growth and root colonisation, in addition to distinguishing genes specifically required during competition with Rlv3841. Common factors required for rhizosphere fitness and root colonisation are summarised (Fig. 5.23 and Table S5.11). As seen with Rlv3841 in Chapter 4, these show a clear need to synthesise compounds (such as amino acids,

ribonucleotides, cofactors and cytochromes), utilise the glycolysis/gluconeogenesis pathway, modify gene expression in response to external stimuli, adapt to changes in osmolarity and alter cell surface structure.

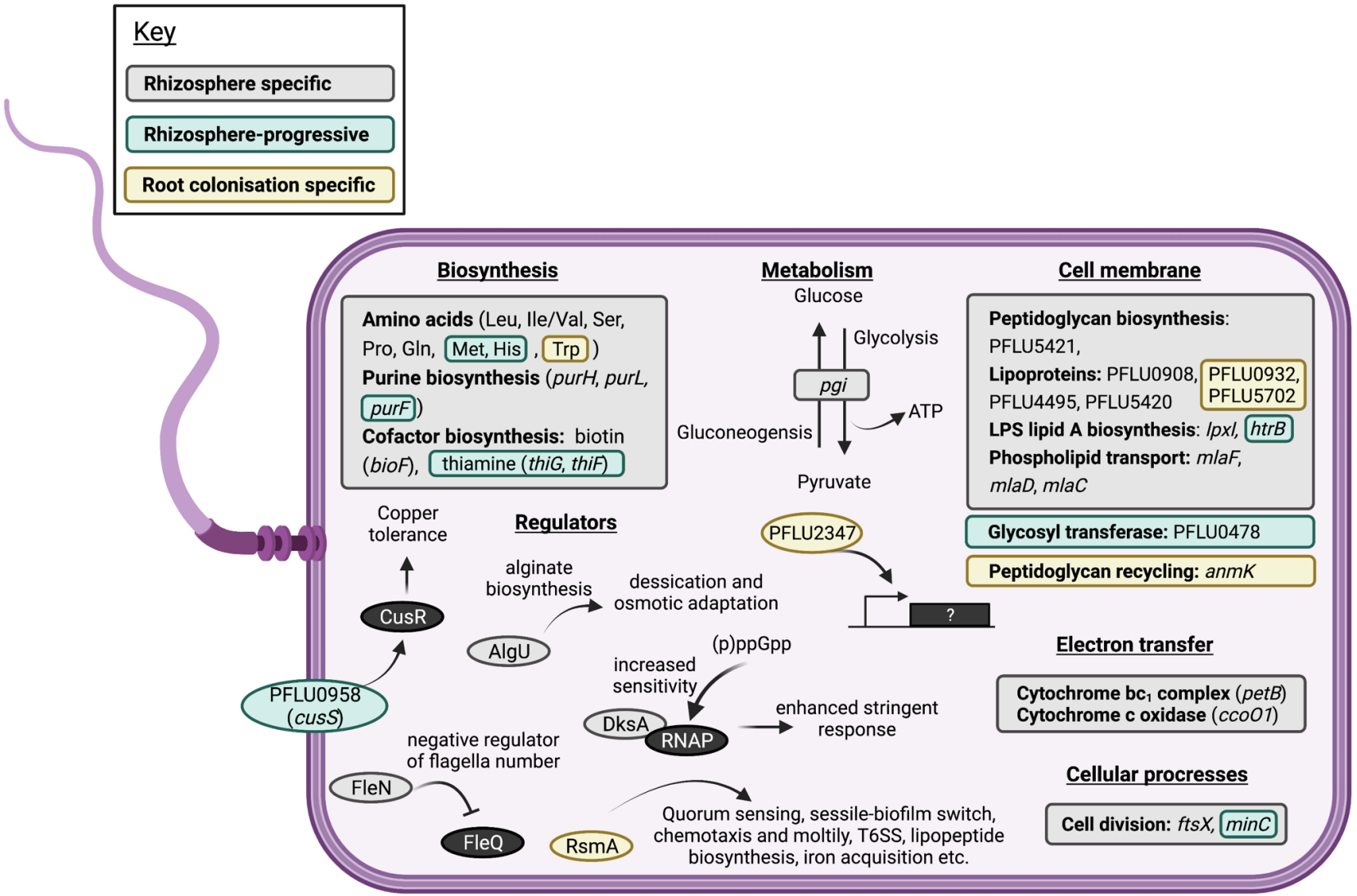

Key
Rhizosphere specific
Rhizosphere-progressive
Root colonisation specific
Biosynthesis
Amino acids (Leu, Ile/Val, Ser, Pro, Gln, Met, His, Trp)
Purine biosynthesis (purH, purL, purF)
Cofactor biosynthesis: biotin (bioF), thiamine (thiG, thiF)
Metabolism
Glucose
Glycolysis
pgi
ATP
Gluconeogenesis
Pyruvate
Cell membrane
Peptidoglycan biosynthesis: PFLU5421,
Lipoproteins: PFLU0908, PFLU0932, PFLU4495, PFLU5420, PFLU5702
LPS lipid A biosynthesis: lpxI, htrB
Phospholipid transport: mlaF, mlaD, mlaC
Glycosyl transferase: PFLU0478
Peptidoglycan recycling: anmK
Copper tolerance
Regulators
PFLU2347
?
CusR
alginate biosynthesis
dessication and osmotic adaptation
AlgU
(p)ppGpp
increased sensitivity
PFLU0958 (cusS)
DksA
RNAP
enhanced stringent response
Electron transfer
Cytochrome bc1 complex (petB)
Cytochrome c oxidase (ccoO1)
negative regulator of flagella number
FleN
FleQ
RsmA
Quorum sensing, sessile-biofilm switch, chemotaxis and moltily, T6SS, lipopeptide biosynthesis, iron acquisition etc.
Cellular processes
Cell division: ftsX, minC

Figure 5.23. Summary of SBW25 common factors required for growth in the rhizosphere and colonisation of both pea and barley during single and co-inoculation with Rlv3841. Under all tested conditions genes shown in grey are required for rhizosphere growth, genes shown in blue were rhizosphere-progressive, and genes shown in yellow were required for root colonisation. Genes shown in black were not identified as common determinants of SBW25 rhizosphere growth or root colonisation but based on the literature interact with those that were. Genes are listed in Table S5.11.

Genes required for either growth in the rhizosphere (Fig. 5.24) or colonisation (Fig. 5.25) of pea or barley during competition with Rlv3841 were also identified. Many of these genes had functional roles in the biosynthesis and transport of amino acids or carbohydrates. This altered requirement in transport reflects the variation in available resources due to difference in root exudate composition between plant species and uptake by Rlv3841. For example, during competition with Rlv3841, two enzymes that catalyse the first step in aromatic amino acid biosynthesis (AroG and AroF) were important for growth only in the rhizosphere of pea (and not barley) and one of these was also important for root colonisation. In contrast, a repressor of aromatic amino acid biosynthesis (TyrR) promoted barley root colonisation during competition with Rlv3841 but was unimportant in pea.

It is worth noting that some traits involved in bacterial competition, such as the production of antimicrobial compounds, would not be identified in a BarSeq screen due to the fact that mutants impaired in the production of such compounds account for a small proportion of the mutant pool and so the majority of mutants would still be capable of producing these.

Notably, very few genes contributing to SBW25 growth in the rhizosphere of pea or barley during competition with Rlv3841 were common to both plants. The same was true for genes implicated in competitive root colonisation. This again highlights the variation in the rhizospheres of different plant species, a fact also demonstrated in Chapter 4 through the identification of many plant-specific genes important for Rlv3841 rhizosphere growth and root-colonisation. Analysis of gene fitness values for SBW25 chemotaxis clusters did, however, highlight that chemotaxis mediated by the

che1 cluster is of increased importance for root colonisation during competition with Rlv3841 (Fig. 5.14). The same was true for genes involved in flagella biosynthesis, export and assembly (Fig. 5.14). Analysis of the *rsp* gene cluster, known to be induced in plant rhizospheres, also suggested increased importance during competition with Rlv3841, for both rhizosphere growth and root colonisation, though its function remains elusive.

The identification of Rlv384 and SBW25 genes specifically required during competition with each other demonstrate the need to adapt and alter various function in response to a competitor strain. Analysis of genes contributing to rhizosphere growth and root colonisation under these conditions revealed that for both Rlv3841 and SBW25, alteration in resource availability appears to be a primary driver of competition, as opposed to direct interactions with each other. This is reflected by the large number of transport related genes.

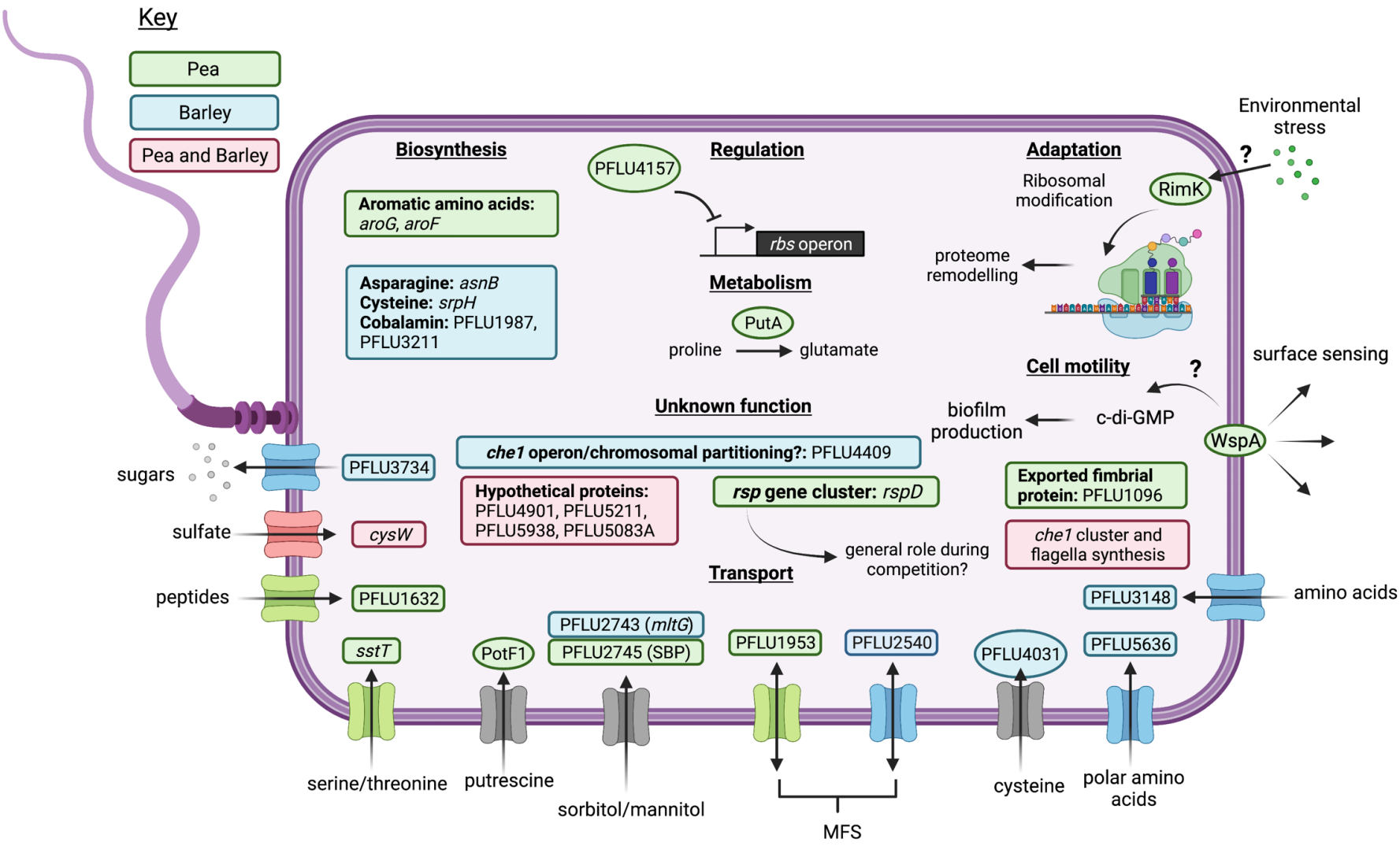

Key
Pea
Barley
Pea and Barley
Biosynthesis
Aromatic amino acids: aroG, aroF
Asparagine: asnB
Cysteine: srpH
Cobalamin: PFLU1987, PFLU3211
Regulation
PFLU4157
rbs operon
Metabolism
PutA
proline
glutamate
Adaptation
Ribosomal modification
RimK
Environmental stress
?
proteome remodelling
Cell motility
?
surface sensing
biofilm production
c-di-GMP
WspA
Unknown function
che1 operon/chromosomal partitioning?: PFLU4409
Hypothetical proteins: PFLU4901, PFLU5211, PFLU5938, PFLU5083A
rsp gene cluster: rspD
general role during competition?
Exported fimbrial protein: PFLU1096
che1 cluster and flagella synthesis
Transport
sugars
PFLU3734
sulfate
cysW
peptides
PFLU1632
sstT
PotF1
PFLU2743 (mltG)
PFLU2745 (SBP)
PFLU1953
PFLU2540
PFLU4031
PFLU3148
PFLU5636
amino acids
serine/threonine
putrescine
sorbitol/mannitol
MFS
cysteine
polar amino acids

Figure 5.24. Summary of SBW25 factors contributing to growth in the rhizosphere of pea and/or barley during competition with Rlv3841. Genes shown in green are important in the pea rhizosphere, genes in blue are important in the barley rhizosphere and those shown in pink are required in both. Genes shown in black were not identified to contribute to SBW25 rhizosphere growth during competition with Rlv3841 but based on the literature interact with those that do. Genes are from Tables S5.12, S5.13 and S5.14.

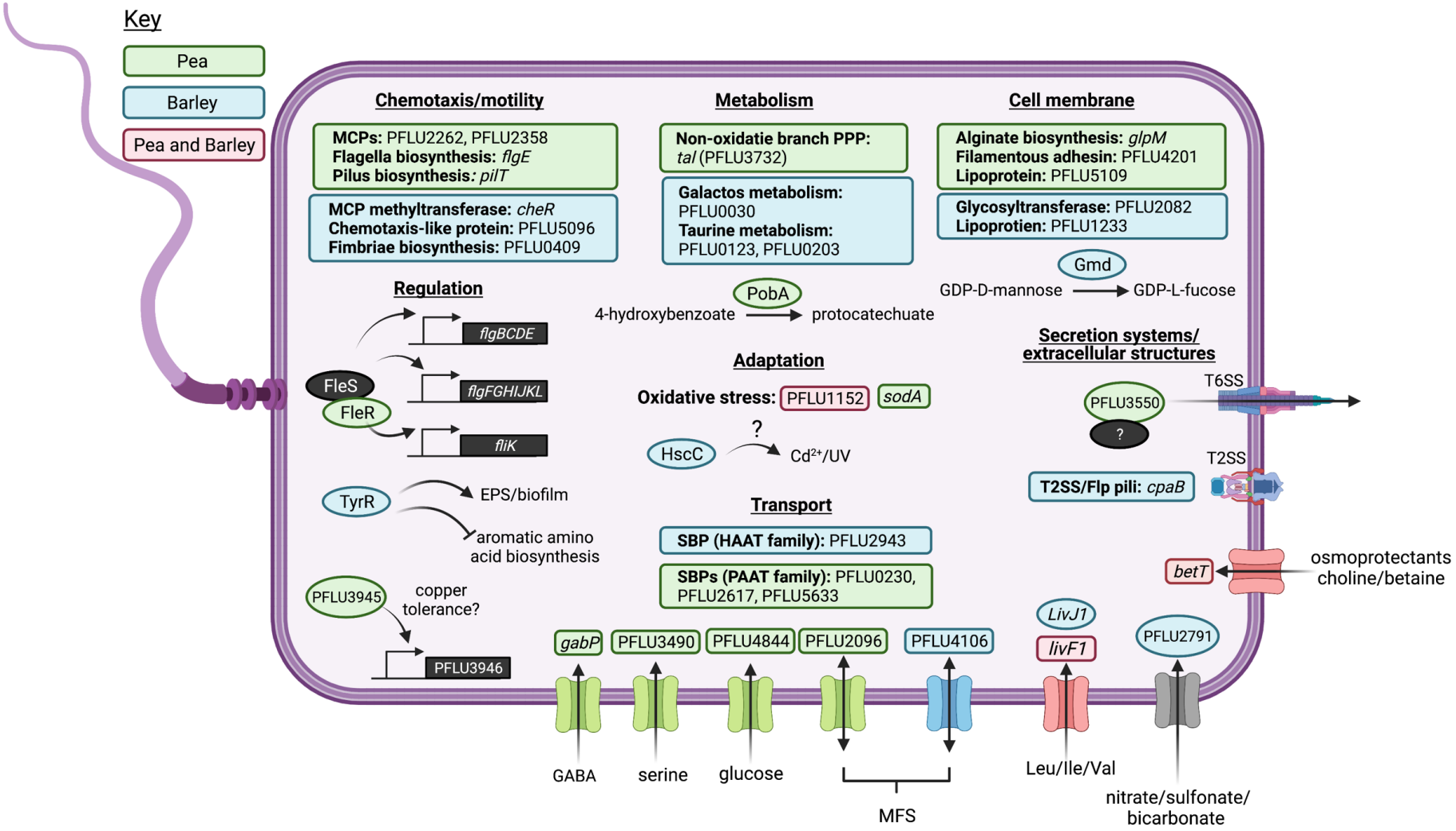

Key
Pea
Barley
Pea and Barley
Chemotaxis/motility
MCPs: PFLU2262, PFLU2358
Flagella biosynthesis: flgE
Pilus biosynthesis: pilT
MCP methyltransferase: cheR
Chemotaxis-like protein: PFLU5096
Fimbriae biosynthesis: PFLU0409
Metabolism
Non-oxidatie branch PPP: tal (PFLU3732)
Galactos metabolism: PFLU0030
Taurine metabolism: PFLU0123, PFLU0203
Cell membrane
Alginate biosynthesis: glpM
Filamentous adhesin: PFLU4201
Lipoprotein: PFLU5109
Glycosyltransferase: PFLU2082
Lipoprotien: PFLU1233
Gmd
GDP-D-mannose
GDP-L-fucose
Regulation
flgBCDE
FleS
FleR
flgFGHIJKL
fliK
TyrR
EPS/biofilm
aromatic amino acid biosynthesis
PFLU3945
copper tolerance?
PFLU3946
PobA
4-hydroxybenzoate
protocatechuate
Adaptation
Oxidative stress: PFLU1152
sodA
HscC
?
Cd2+/UV
Transport
SBP (HAAT family): PFLU2943
SBPs (PAAT family): PFLU0230, PFLU2617, PFLU5633
Secretion systems/ extracellular structures
PFLU3550
?
T6SS
T2SS
T2SS/Flp pili: cpaB
betT
osmoprotectants choline/betaine
gabP
PFLU3490
PFLU4844
PFLU2096
PFLU4106
LivJ1
livF1
PFLU2791
GABA
serine
glucose
MFS
Leu/Ile/Val
nitrate/sulfonate/ bicarbonate

Figure 5.25. Summary of SBW25 factors contributing to colonisation of pea and/or barley during competition with Rlv3841. Genes shown in green are important in the pea colonisation, genes in blue are important in the barley colonisation and those shown in pink are required for the colonisation of both. Genes shown in black were not identified to contribute to SBW25 colonisation during competition with Rlv3841 but based on the literature interact with those that do. Genes are from Tables S5.12, S5.13 and S5.14.

Chapter 6

General Discussion

6.1. Overview

Bacterial persistence in the rhizosphere and root colonisation are critical for the establishment of many beneficial plant-bacteria interactions including those between *R. leguminosarum* and its host legumes. Despite this, most studies on *R. leguminosarum* have focused on its symbiotic lifestyle as endosymbiont in root nodules. Increasingly, it has become apparent that although *R. leguminosarum* can form N_2-fixing symbioses with members of the Fabeae legume tribe (*Pisum*, *Lens*, *Lathyrus*, and *Vicia* species) different *Rlv* strains display altered competitiveness for nodulation of host legumes (Laguerre et al., 2003; Mutch and Young, 2004; Jorrin and Imperial, 2015; Andrews and Andrews, 2017). This has been proposed to at least in part be due to altered competitiveness for rhizosphere growth and/or root colonisation (Hynes and O'Connell, 1990; Garcia-Fraile et al., 2015; Jorrin, 2016; Bourion et al., 2018), which is not surprising since both plant-bacteria and bacteria-bacteria interactions have been shown to influence these processes (Knights et al., 2021). In agricultural environments this can result in suboptimal yields since the most competitive strains are not always the most effective at BNF meaning introduced rhizobia, chosen for their superior rates of N_2 fixation, are often outcompeted by native strains (Meade et al., 1985; Bourion et al., 2018). The two primary aims of the work described in this thesis were to: 1) assess the effect of plant species on *R. leguminosarum* rhizosphere growth and root colonisation by identifying genetic determinants contributing to these processes in the presence of host legumes (pea, lentil and *Lathyrus*), a non-host legume (alfalfa) and a non-legume (barley) and 2) identifying genes contributing to rhizosphere growth and root colonisation in a competitive setting more analogous to field conditions. To achieve this RB-TnSeq was

used in a genome-wide approach to assess how loss-of-function mutations affect fitness under multiple conditions. Rlv3841 was chosen as a model *R. leguminosarum* strain for use in this work due to the availability of a complete genome sequence, a requirement of RB-TnSeq, and its regular use as a model for the investigation of *Rhizobium*-legume symbioses. This has resulted in a wealth of experimental data regarding its interactions with host legumes at various stages of symbiosis (Williams et al., 2008; Ramachandran et al., 2011; Garcia-Fraile et al., 2015; Parsons, 2019; Wheatley et al., 2020).

6.2. Development of RB-TnSeq in *Rhizobium*

To facilitate the identification of genes contributing to Rlv3841 rhizosphere persistence and root colonisation, the *Rhizobiaceae*-compatible *mariner* transposon delivery vector, pSAM_RI (Perry and Yost, 2014), was converted into an RB-TnSeq vector through inclusion of unique 20-nucleotide barcodes. These modified vectors were used to generate a highly saturated Rlv3841 transposon library in which barcoded transposons mapped to 72% of 140,056 TA sites present in the genome (Table 3.1). To the best of my knowledge, this represents the first application of RB-TnSeq in *Rhizobium*. RB-TnSeq is similar to other transposon mutagenesis tools, such as TnSeq and INSeq, in that it allows the identification of conditionally important genes at the whole genome level. This is achieved by generating single insertion transposon mutant libraries with high insertion densities, which are subsequently exposed to a condition of interest and then sequenced to assess how the insertion mutation in a particular gene effects the survival of the bacteria in that condition. The benefit of RB-TnSeq however, is that each individual transposon mutant within the library is

identifiable by a unique, but random, 20-nuceltoide barcode. This enables distinction and quantification of individual mutants within the mutant pool by BarSeq, meaning that the relative fitness of each individual mutant can be assessed (Wetmore et al., 2015). In this way, RB-TnSeq provides a high level of internal replication since multiple independent insertion mutants for each TA site within a gene can be used in the calculation of genes fitness, increasing the confidence in fitness predictions. Furthermore, RB-TnSeq does not require the mapping of transposon locations following every experiment. Instead, transposon insertion locations need only be mapped once to the genome. In future BarSeq experiments, using the same mutant pool, the random 20-nucleotide barcode is PCR-amplified, sequenced and linked to the previously determined location. RB-TnSeq therefore provides a rapid high-throughput method of assaying the importance of genes under many different conditions. In this thesis alone for example, BarSeq was used to quantify gene fitness contributions for Rlv3841 during growth under 11 *in planta* conditions and SBW25 under four. BarSeq also provides an advantage for *in planta* experiments in that retrieved mutant populations need not be regrown in laboratory culture prior to preparation for sequencing. When using techniques such as INSeq to assay gene fitness *in planta*, recovered mutant libraries require culturing in rich media for several generations of growth to ensure sufficient bacterial DNA for library preparation as well as dilution of plant DNA, which interferes with this process (Wheatley et al., 2020). Instead, the work described here demonstrates that the single PCR amplification of barcodes required by BarSeq can be done directly on the mutant population recovered from *in planta* environments following DNA extraction.

6.3. Identification of Rlv3841 genes contributing to growth in the rhizosphere and colonisation of pea, lentil, *Lathyrus*, alfalfa and barley

Quantification of Rlv3841 growth in the root niches of these plant species over a seven-day period demonstrated that growth was sufficient for determination of gene fitness by BarSeq and so BarSeq screens were conducted for Rlv3841 rhizosphere growth and root colonisation of pea, lentil *Lathyrus*, alfalfa and barley (Table 4.1). Assessment of gene contributions was achieved for 6,019 (80%) of the 7,528 protein coding genes (Chapter 4). The ability to resolve differences in mutant fitness by BarSeq increases with the number of generations of growth the mutant population undergoes. *In planta* mutant populations underwent between six and nine generations of growth. Since fitness contributions were identified for many genes, the analysis described in this worked focused on those with the largest contributions i.e., a fitness values less than -1 representing at least a 50% reduction in growth relative to a typical mutant within the population. It is entirely plausible that genes with milder fitness contributions, for example those with a fitness value of -0.42 representing a 25% reduction in growth, have biological significance. However, previous BarSeq screens have shown that there is higher statistical confidence in fitness estimates for genes with stronger fitness contributions (Helmann et al., 2019). Conducting multiple replicates under a given conditions to see if genes with weaker contributions are consistently identified would provide greater support in the identification of these genes with weak phenotypes.

6.3.1. Validation of RB-TnSeq *in planta*

To validate the use of BarSeq in identifying genes contributing to Rlv3841 rhizosphere fitness and root colonisation, BarSeq assays were initially conducted with the host legume pea, for which, numerous genes involved in these processes have previously been identified (Williams et al., 2008; Ramachandran et al., 2011; Frederix et al., 2014; Garcia-Fraile et al., 2015; Parsons, 2019; Wheatley et al., 2020). Comparison to results obtained from previous INSeq screens found that 41 genes implicated in pea rhizosphere growth and 49 genes contributing to root colonisation were assigned fitness values less than -1 under their respective condition (Wheatley et al., 2020). Assessment of fitness values for some of these key genes, as well as others previously identified in Rlv3841 single mutant studies, during growth in the rhizosphere and colonisation of lentil, *Lathyrus*, alfalfa and barley revealed that many were determinants of rhizosphere growth or root colonisation for multiple plant species (Table 6.1). Several genes however, previously implicated in Rlv3841 attachment to pea root hairs showed neutral phenotypes under most conditions. These include *gmsA, celA* and *pssA*, suggesting the existence of alternative mechanisms for rhizobial attachment to root hairs (1-2 hpi) and long-term colonisation over the entire root surface (7 dpi) (Smit et al., 1987).

Notably, in comparison to the previous INSeq screens, BarSeq screens led to the identification of far more genes involved in Rlv3841-pea rhizosphere growth and root colonisation. A further 626 genes contributing to Rlv3841-pea rhizosphere fitness and 445 genes contributing to pea root colonisation were identified. This is most likely due to differences in how the data generated by INSeq and BarSeq is analysed. For example, the HMM was used to analyse the INSeq results. In this instance the HMM

classes genes as important for growth if they have less than 1% of the mean read count. In contrast, during the analysis of BarSeq data a gene is classed as important for fitness if it has 50% less reads relative to time zero and therefore genes with milder contributions to Rlv3841 rhizosphere fitness or root colonisation are identified.

Table 6.1. Summary of BarSeq phenotypes for key genes previously implicated in *R. leguminosarum* rhizosphere growth, root attachment or root colonisation. RS, rhizosphere specific; RP, rhizosphere progressive; CS, colonisation specific; NE, neutral phenotype under all conditions tested and ND, fitness value not determined due to essentially in initial growth media or insufficient reads at time zero.

Category	Gene(s)	BarSeq phenotype					Comments
		Pea	Lentil	*Lathyrus*	Alfalfa	Barley	
Adaptation	*rapB* (RL3911)	NE	NE	NE	NE	RS	Rhizobial adherence proteins that promote root attachment and biofilm development by Rlv3841. Appears to be redundancy among *rap* proteins encoded by the Rlv3841 genome, with no observable phenotype upon mutation of a single *rap* gene (Frederix et al., 2014).
	rapA2 (pRL100451)	NE	NE	NE	NE	NE	
	relA (RL1506)	RP	RS	RS	NE	RS	RelA catalyses the synthesis of (p)ppGpp, which mediates the SR. Rlv3841 *relA* mutants are impaired in pea rhizosphere growth, root colonisation and nodule symbiosis (Wheatley et al., 2020).
Amino acid biosynthesis	*trpA* (RL0022)	RP	RP	RP	RP	RP	Required for tryptophan biosynthesis. Essential for growth in the pea rhizosphere and establishment of symbiosis (Wheatley et al., 2020).
	leuA1 (RL3513)	RP	RP	RP	RP	RP	Required for leucine biosynthesis. Essential for growth in the pea rhizosphere and establishment of symbiosis (Wheatley et al., 2020).
	ilvC (RL3205)	RP	RP	RP	RP	RP	Required for isoleucine biosynthesis. Essential for growth in the pea rhizosphere and establishment of symbiosis (Wheatley et al., 2020).

Chemotaxis	*cheX1/cheS* (RL0686)	RP	CS	RP	RP	RP	Part of the *che1* operon, which encodes the conserved set of *cheAWYRB* homologues, facilitates root colonisation and promotes competitive nodulation of peas (Miller et al., 2007).
Cellulose fibrils	*celA* (RL1646)	NE	NE	NE	NE	NE	CelA is required for the synthesis of cellulose fibrils. Rlv3841 *celA* mutants are impaired in root cap formation, but can attach to root hairs individually, show no impairment in root-attachment (1 hpi) and form biofilms indistinguishable from those formed by wild-type RLv3841(Williams et al., 2008; Parsons, 2019).
Cell membrane/envelope	RL0811A	CS	RP	RP	CS	CS	An LPS glycosyltransfe involved in LPS biosynthesis. Mutation of RL0811A results in reduced pea root colonisation and nodule infection relative to wild type (Garcia-Fraile et al., 2015; Wheatley et al., 2020).
	gmd (RL0825)	CS	CS	NE	RP	CS	Gmd has a role in LPS biosynthesis. Rlv3841 *gmd* mutants are impaired in pea rhizosphere growth, root colonisation and nodule symbiosis (Wheatley et al., 2020).
	ropB (RL1589)	RP	CS	NE	RS	NE	RopB is an outer membrane porin thought to form amyloid fibrils (Kosolapova et al., 2019). Important for Rlv3841 pea rhizosphere growth, root colonisation and nodule infection (Wheatley et al., 2020).
	gmsA (RL1661)	NE	NE	NE	NE	NE	Encodes UPP glucomannan, which under neutral and acidic conditions mediates reversible polar attachment of single cells to pea and vetch root hairs by binding plant lectins (Laus et al., 2006).
	pssA (RL3752)	NE	NE	NE	RS	NE	PssA is an EPS biosynthesis regulator. Rlv3841 *pssA* mutants are impaired in attachment to pea

							roots 1-h post inoculation however, mutants can colonise the root epidermis (Williams et al., 2008; Parsons, 2019).
Energy production	*cycY* (RL4210)	CS	CS	CS	CS	RP	Involved in cytochrome *c* production. Previously shown to be important for Rlv3841 pea root colonisation and nodule symbiosis (Wheatley et al., 2020).
Flagella	*flgB* (RL0707)	CS	CS	CS	CS	CS	Encodes part of the flagella apparatus. Flagella mediated motility facilities Rlv3841-pea root colonisation (Miller et al., 2007).
Hypothetical proteins	RL4399	RP	RS	RP	CA	RS	Encodes a hypothetical protein belonging to the YKOF superfamily and showing 30% similarity to *B. subtilis* YkoF, a SBP involved in the transport of HMP/thiamine (Devedjiev et al., 2004). Mutants are impaired in pea rhizosphere growth, root colonisation and nodule symbiosis (Wheatley et al., 2020).
Riboflavin biosynthesis	*ribC* (RL1622)	RP	RP	RP	RP	RP	Putative riboflavin biosynthesis protein. Riboflavin promotes *S. meliloti* rhizosphere survival and root colonisation (Streit et al., 1996; Yang et al., 2002; García Angulo et al., 2013). Rlv3841 *ribC* mutants are impaired in pea rhizosphere growth, root colonisation and nodule symbiosis (García Angulo et al., 2013; Wheatley et al., 2020).
Transporters	*dctA* (RL3424)	RP	CS	NE	RS	NE	A C_4-dicarboxylate transporter. Rlv3841 *dctA* mutants display reduced competitiveness in the pea rhizosphere (Ramachandran et al., 2011).
	bacA (RL3557)	RP	CS	CS	RP	RP	A transmembrane transporter involved in the incorporation of VLCFA into lipid A of LPS. Rlv3841 *bacA* mutants are impaired in pea rhizosphere

							growth, root colonisation and nodule symbiosis (Karunakaran et al., 2010; Wheatley et al., 2020).
	RL4400	CS	RP	CS	CS	RP	Component of an ABC transporter involved in the import of the thiamine precursor HMP. Mutation results in impaired pea rhizosphere growth, root colonisation and nodule symbiosis (Wheatley et al., 2020).
	thiY (RL4402)	RP	RP	RP	RP	RP	SBP involved in transport of the thiamine precursor HMP. Rlv3841 *thiY* mutants are impaired in pea rhizosphere growth, root colonisation and nodule symbiosis (Wheatley et al., 2020).
	ndvA (RL4640)	RP	CS	CS	RP	CS	Responsible for the synthesis of cyclic ß-1-2-glucans, cell surface carbohydrates involved in osmotic adaptation (Dylan et al., 1990; Breedveld and Miller, 1994). In Rlv3841 *ndvA* is upregulated in plant rhizospheres and mutants are impaired in pea rhizosphere growth, root colonisation and nodule symbiosis (Ramachandran et al., 2011; Wheatley et al., 2020).
Metabolism	*araE* (RL3613)	RP	NE	RS	RP	RS	Encodes an arabinose dehydrogenase involved in arabinose catabolism. Rlv3841 *araE* mutants display reduced growth in the pea rhizosphere during competition with wild-type Rlv3841 and demonstrate impaired competitiveness for nodule infection (Garcia-Fraile et al., 2015).
	cbbT (RL4006)	RP	RP	RP	RP	RP	Encodes a putative transketolase involved in the non-oxidative branch of the pentose phosphate pathway. Mutation results in impaired pea rhizosphere growth, root colonisation and nodule symbiosis (Wheatley et al., 2020).

Protein modification/ biosynthesis	*cycY* (RL4540)	CS	CS	CS	CS	RP	Involved in cytochrome *c* biosynthesis. Rlv3841 *cycY* mutants are impaired in pea root colonisation and nodule symbiosis (Wheatley et al., 2020).
Regulation	*feuP* (RL1433)	RS	RS	RS	RS	RS	Two-component sensor kinase shown to regulate iron uptake in Rlv3841, though likely has additional functions as seen in *S. meliloti* (Yeoman et al., 1997; Griffitts et al., 2008). Mutants are impaired in pea rhizosphere growth, root colonisation and nodule symbiosis (Wheatley et al., 2020).
	feuQ (RL1434)	RP	RP	RP	RP	RP	
	RL3206	RP	CS	RP	RP	RP	TetR family transcriptional regulator predicted to regulate the adjacent gene, RL3205 (*ilvC*), involved in leucine biosynthesis. Rlv3741 mutants are impaired in pea rhizosphere growth, root colonisation and nodulation (Wheatley et al., 2020).
	pRL120618	RP	RP	RS	RS	RP	LacI family transcriptional repressor that may be involved in the regulation of sugar utilisation (Novichkov et al., 2013). Mutation results in impaired pea rhizosphere growth, root colonisation and nodule symbiosis (Wheatley et al., 2020).
Ribonucleotide biosynthesis	*pyrB* (RL1739)	RP	RP	RP	RP	RP	Required for pyrimidine biosynthesis. Mutation results in impaired pea rhizosphere growth, root colonisation and nodule symbiosis (Wheatley et al., 2020).

6.3.2. Common determinants of rhizosphere growth and root colonisation

Analysis of genes contributing to Rlv3841 rhizosphere growth and root colonisation of all five plant species allowed the identification of a core set of 73 genes commonly contributing to rhizosphere growth and 90 genes contributing to root colonisation (Fig. 4.6 and summary Fig. 4.12 and Fig. 4.13). Comparison of these genes revealed that 32 were rhizosphere-progressive whereby mutation impaired both growth in the rhizosphere and root colonisation. For growth in the rhizosphere genes primarily had roles in the biosynthesis of compounds such as amino acids, ribonucleotides and cofactors. The identification of these auxotrophic mutants as having severe fitness impairments provides an indication as to the nutrient availability in the rhizosphere of these plant species. The observation of many amino acid auxotroph's for example suggests that whilst most amino acids are present in root exudates (Canarini et al., 2019), they are not at concentrations sufficient to support bacterial growth. Variation in the severity of the growth impairment displayed by various amino acid auxotroph's also provides an indication of the most limiting amino acids. For example, leucine, isoleucine and tryptophan auxotroph's showed the strongest phenotypes across all five plant rhizospheres with multiple genes involved in these pathways identified as part of the core set of genes commonly required for rhizosphere fitness. Several of these were also identified as rhizosphere-progressive, not only affecting growth in the rhizosphere, but also colonisation of these plant species. This is consistent with a previous INSeq screen which found Rlv3841 leucine, isoleucine and tryptophan auxotroph's to be impaired in pea rhizosphere growth and root colonisation (Wheatley et al., 2020), but also suggests that these three amino acids are severely limited in diverse plant rhizospheres. Proline, glutamine and glutamate auxotroph's also showed

impaired growth among all plant rhizospheres whilst biosynthesis of phenylalanine and cysteine were commonly important for root colonisation.

Genes involved in regulation also provide insights into the rhizosphere environment. Requirement of the sensor kinase NtrB for example, which activates transcription of nitrogen-regulated genes via NtrC under conditions of nitrogen limitation, indicates that nitrogen is limited in all five plant rhizospheres (Patriarca et al., 1992; Fischer, 1994; Brown et al., 2014). Plants were grown in nitrogen-free rooting solution to promote the initial phases of *Rhizobium*-legume signalling and so this is not surprising. Similarly, requirement of the two-component sensor-kinase system FeuPQ indicates either iron limitation and/or a need to adapt to hypoosmotic conditions. The copper tolerance protein CopB was also important for growth in all plant rhizospheres and subsequent root colonisation (Rensing and Grass, 2003). Copper resistance mechanisms are widespread in plant-associated microbes since whilst essential at low levels, copper is toxic at higher concentrations and likely therefore plays an important role in cellular copper homeostasis (Rainey, 1999).

Genes involved in polysaccharide biosynthesis were important for both rhizosphere growth and root colonisation. Polysaccharides (LPS/EPS/CPS) play and important role in adaption to environmental stresses such as oxidative stress, in addition to influencing bacterial adhesion, root colonisation and host specificity during the establishment of *Rhizobium*-legume symbioses (Davey and O'Toole, 2000; Wang et al., 2012; Wheatley and Poole, 2018). Several of these genes were involved in the synthesis of the LPS O-antigen which is well documented to play a role in rhizosphere growth and root colonisation of diverse bacterial species (Lerouge and Vanderleyden,

2002; Mitra et al., 2016). Similarly, incorporation of VLCFA into the lipid A component of LPS was important for root colonisation. This is consistent with the fact that Rlv3841 mutants impaired in VLCFA incorporation into LPS display impaired membrane integrity coupled with increased sensitivity to environmental stresses (Karunakaran et al., 2010). Requirement of these polysaccharide modifications demonstrate the stressful environment of rhizosphere and the importance of membrane integrity in protection against these stressors.

Chemotaxis and flagella mediated motility are well documented in the literature to be important traits required for root colonisation among diverse bacteria including Rlv3841 (Ames and Bergman, 1981; Bauer and Caetano-Anollés, 1990; de Weert et al., 2002; Miller et al., 2007; Allard-Massicotte et al., 2016). Whilst no genes with roles in chemotaxis and motility were commonly required for rhizosphere growth, they account for 10% of genes commonly required for the colonisation of all five plant species (Fig. 4.7). Most of these genes had roles in flagella biosynthesis, export and assembly or belonged to the *che1* clusters which is the major pathway controlling chemotaxis in Rlv3841 (Miller et al., 2007). Whilst chemotaxis and motility are likely essential for rhizosphere colonisation in the soil environment, where bacteria must move from bulk soil into the rhizosphere, they were not found to be important for rhizosphere growth under these experimental conditions. This is probably due to the fact that plants were grown in narrow boiling tubes, allowing their rooting systems to grow throughout the vermiculite, and so the rhizosphere was defined as the population retrieved from the vermiculite fraction.

In comparison to genes involved in the biosynthesis of essential compounds such as amino acids, those involved in the catabolism of compounds often have milder phenotypes. This is most likely due to the presence of diverse and interchangeable carbon sources present in plant root exudates. Nonetheless genes involved in the catabolism of arabinose, rhamnose and erythritol were identified as important for the colonisation of all five plant species. Catabolism of these carbon sources has previously been shown to be important for competitive nodulation of pea by Rlv3841 or in the case of erythritol closely related *R. leguminosarum* VF39SM (Richardson et al., 2004; Yost et al., 2006; Garcia-Fraile et al., 2015). Not only does this suggests that arabinose, rhamnose and erythritol are present in the root exudates of lentil, *Lathyrus*, alfalfa and barley, but also that catabolism of them by Rlv3841 provides a competitive advantage for colonisation of their roots and in the case of pea at least this is essential for competitive nodulation.

Notably, very few genes with predicted role in transport were identified as common determinants of rhizosphere growth or root colonisation indicating that altered root exudate composition among the five plant species results in altered transport requirements in these rhizospheres. One notable exception was the common requirement for the import of the thiamine (vitamin B_1) precursor HMP. Unlike many other bacteria, Rlv3841 does not have a pathway that allows de novo synthesis of thiamine and instead relies of a salvage pathway for which HMP is an intermediate (Karunakaran et al., 2006).

Analysis of Rlv3841's central carbon metabolism pathways revealed that the glycolysis/gluconeogenesis pathway was of particular importance in all plant

rhizospheres, as was the metabolism of pentose sugars via the pentose phosphate pathway (Fig. 4.8). Additionally, *cbbT* (RL4006), encoding a putative transketolase of the non-oxidative branch of the pentose phosphate pathway was commonly required for rhizosphere growth and root colonisation. This suggests that CbbT is the primary transketolase in Rlv3841, and that the activity of the other three transketolases encoded by RL2718, RL2719 and pRL100453 are not sufficient to compensate for the lack of CbbT.

Whilst several of these genes have previously been implicated in Rlv3841-pea rhizosphere fitness or root colonisation (Table 6.1), many novel genes were identified as contributing to these processes for the first time. Of these, several have been shown to be essential for the establishment of N_2-fixing nodule symbioses with pea (Table 6.2). These results indicate that these genes play an important role during rhizosphere growth or root colonisation that subsequently impairs these mutants throughout the development of symbiosis. Two genes of particular interest for further characterisation are RL0149, encoding a transcriptional regulator previously shown to be 11-fold upregulated in the Rlv3841 *praR* hyper attachment mutant, and pRL80128, encoding a hypothetical protein that was required for rhizosphere growth and root colonisation of all five plant species tested (Frederix et al., 2014).

Table 6.2. Primary candidate genes predicted to be involved in Rlv3841 rhizosphere growth and/or root colonisation of pea, lentil, *Lathyrus*, alfalfa and barley based on BarSeq screens. RS, rhizosphere specific; RP, rhizosphere progressive; CS, colonisation specific and NE, neutral phenotype.

Category	Gene(s)	BarSeq phenotype					Comments
		Pea	Lentil	*Lathyrus*	Alfalfa	Barley	
Adaptation	RL0962	CS	CS	CS	CS	CS	Encodes a putative dioxygenase, mutation of which results in impaired nodule symbiosis with pea (Wheatley et al., 2020). This is the first evidence for a role during root colonisation.
	rhiB (pRL100170)	RS	RS	RP	RS	RS	Rlv3841 *rhiB* is rhizosphere induced though its function remains unknown. Expression is regulated in a cell density dependent manner via RhiR (Cubo et al., 1992; Rodelas et al., 1999; Wisniewski-Dyé and Downie, 2002). This is the first evidence that *rhiB* contributes to rhizosphere fitness, likely observed due to the competitive nature of BarSeq screens.
Cell membrane/ envelope	*fcl* (RL0826)	RS	RS	RS	RS	RP	Encodes a putative GDP-L-fucose synthase with a role in LPS O-antigen biosynthesis. *R. tropici* mutants impaired in GDP-L-fucose synthesis demonstrate reduce growth in the maize rhizosphere, which is at least in part due to increased susceptibility to maize antimicrobials (Ormeño-Orrillo et al., 2008). Rlv3841 *fcl* mutants are impaired in the establishment of nodule symbiosis with pea (Wheatley et al., 2020).
	ropA2 (RL1499)	CS	CS	CS	CS	RP	Predicted to encode an outer membrane porin thought to form amyloid fibrils (Kosolapova et al., 2019). Rlv3841 *ropA2* is essential for nodule symbiosis with pea and 1-h attachment to pea roots

							(Parsons, 2019; Wheatley et al., 2020). This indicates that it also plays a key role in root colonisation that may be progressive during the establishment of symbiosis.
	rmlC (RL1626)	CS	NE	CS	NE	RS	Required for dTDP-rhamnose biosynthesis, an LPS constituent, shown to promote root colonisation in *Rhizobium*, *Azospirillum* and *Herbaspirillum* (Jofré et al., 2004; Balsanelli et al., 2010; Mitra et al., 2016). First evidence for a role in *R. leguminosarum* root colonisation.
Hypothetical protein	pRL80128	RP	RP	RP	RP	RP	Located within a cluster of *tra* genes on pRL8, though its function is unknown. Mutation results in impaired nodule symbiosis with pea (Wheatley et al., 2020).
Metabolism	*araD* (RL3614)	CS	CS	CS	RP	CS	Encodes a putative hydratase. Rlv3841 *araD* mutants are impaired in nodule symbiosis with pea (Wheatley et al., 2020). These results also suggest an important role for *araD* during rhizosphere growth/root colonisation.
	rhaD (pRL110415)	RP	RP	RP	CS	RP	Encodes an oxidoreductase required for the metabolism of rhamnose. *Rhizobium leguminosarum bv. trifolii rhaD* mutants are impaired in competitive nodulation of clover and Rlv3841 *rhaD* mutants are impaired in competitive nodulation of pea (Oresnik et al., 1998; Wheatley et al., 2020). This is the first evidence suggesting that rhamnose catabolism plays an important role in Rlv3841 rhizosphere fitness and root colonisation.
	eryB (pRL120205)	RP	RP	RP	RP	RP	Encodes a putative erythritol phosphate dehydrogenase involved in erythritol metabolism. *R. leguminosarum* VF39SM *eryB* mutants are impaired in erythritol catabolism and show reduced

							competitiveness for nodulation of pea (Yost et al., 2006). This is the first evidence that erythritol catabolism initially plays an important role in rhizosphere fitness, which may be progressive through to nodulation under competitive conditions.
Regulator	RL0149	CS	RP	CS	RP	CS	A transcriptional regulator previously shown to be 11-fold upregulated in the *praR* mutant which shows a hyper attachment phenotype (Frederix et al., 2014). Mutation of RL0149 also results in impaired nodule symbiosis with pea (Wheatley et al., 2020). This is the first evidence for a role during rhizosphere growth/root colonisation.
	phoU (RL0546)	RP	CS	RP	CS	RP	Encodes a putative negative regulator of phosphate transport. Rlv3841 *phoU* mutants are impaired in nodule symbiosis with pea (Wheatley et al., 2020). This is the first evidence for a role during rhizosphere growth/root colonisation.

6.3.3. Plant-specific genes contributing to rhizosphere growth and root colonisation

Many Rlv3841 genes were found to have plant-specific roles in rhizosphere growth or root colonisation (Fig. 4.14 and Fig. 4.15). In contrast to common determinants of rhizosphere growth and root colonisation where only 8% and 3% of genes respectively had predicted roles in transport, between 10% (lentil) and 28% (pea) of plant-specific genes required for rhizosphere growth and 13% (alfalfa) to 28% (lentil) of genes required for root colonisation were involved in transport (Fig. 4.10). Similarly, plant-specific genes with roles in intermediary metabolism were identified. Altered requirements in transport and metabolism among plant rhizospheres likely reflects adaption of Rlv3841 to variation in resource availability due to differences in root exudate composition between plant species. The ability of Rlv3841 to adapt in such a way is likely facilitated by the 269 genes encoded in its genome that are involved in the active uptake off solutes from the environment. Other plant-specific genes had roles in adaption and regulation indicating altered external stimuli among plant rhizospheres, that demand different responses through reprogramming of cellular functions. Some of these regulatory genes had roles in altering Rlv3841 metabolism, for example optimal growth in the barley rhizosphere required repression of glycerol uptake and metabolism by GlpR whilst colonisation of pea was promoted by repression of α-glucoside utilisation via AglR (Willis and Walker, 1999; Ding et al., 2012). Interestingly, growth in the alfalfa rhizosphere required the transcriptional repressor NolR, which in *R. leguminosarum* bv. *viciae* strain TOM has been shown to repress the synthesis of nod factors (Kiss et al., 1998). Nod factors play an important role in ensuring symbioses only proceeds between compatible symbiotic partners and

so it is interesting that NolR was only required in the rhizosphere of alfalfa as a non-host legume of Rlv3841. Whilst the function of NolR has not been confirmed for Rlv3841, it is known that *Rlv* responds to alfalfa derived flavonoids resulting in the triggering of nod factor production and so it is interesting that repression of nod factor biosynthesis is important in the alfalfa rhizosphere (Long, 1996). For root colonisation, plant-specific genes involved in the modification of the cell surface were identified. These could alter the adhesive properties of the cell with different requirements in polysaccharide structure or cell surface proteins possibly reflecting variation in root surface properties between plant species. As expected, very few genes with roles in chemotaxis and motility displayed plant-specific phenotypes for root colonisation. Those that did were two putative MCPs McpD and McpR which promoted colonisation of lentil and *Lathyrus* respectively. These MCPs are likely required for the detection of a chemoattractant unique to the root exudates of these plant species.

6.4. Identification of Rlv3841 genes contributing to competitive growth in the rhizosphere and colonisation of pea and barley

To establish a successful symbiosis, rhizobia must first compete with other microbiome members to colonise the rhizosphere and root surface of its host legume. The success or failure of an individual to establish a successful symbiosis can be influenced both directly or indirectly by other microbiome members through competition for resources or, production of antimicrobial compounds or production of a metabolite that can be utilised by a competitor for example (Zelezniak et al., 2015; Deines and Bosch, 2016; Granato et al., 2019). Nonetheless most studies characterising bacterial colonisation have done so under sterile conditions. To assess

factors required for Rlv3841 colonisation under a competitive setting more analogous to field conditions Rlv3841 was co-inoculated with the robust root coloniser SBW25 onto pea (host legume) and barley (non-legume). SBW25 is a plant growth promoting rhizobacterium shown to colonise and persist in the rhizosphere of numerous plant species (De Leij et al., 1995; Thompson et al., 1995; Naseby et al., 2001; Unge and Jansson, 2001; Gal et al., 2003; Humphris et al., 2005; Jackson et al., 2005; Noirot-Gros et al., 2018; Wilton et al., 2018; Carroll et al., 2020). Its genome has been mapped and it has been used as a model strain for studying bacterial attachment and colonisation dynamics (Silby et al., 2009; Carroll et al., 2020). Importantly, the use of a random 20-nucleotide barcode by RB-TnSeq opens the possibility of conducting multispecies screens in which gene fitness contributions for more than one species are simultaneously assayed at the genome-wide scale. This is possible because the random 20-nucelotide barcode provides 10^{13} potentially unique barcode combinations and therefore it is theoretically possible to generate multiple RB-TnSeq mutant libraries in different bacterial species with minimal repetition of barcodes between them. A random barcoded mutant library was therefore made in SBW25 and mapped barcodes compared to those identified in the Rlv3841-barcoed mutant library (Chapter 5). This revealed that less than 3% of barcodes were common to both libraries and therefore both strains could be used in co-inoculation experiments to assess both Rlv3841 and SBW25 fitness determinants for growth in the rhizosphere and colonisation of pea (host legume) and barley (non-legume).

Gene fitness contributions were successfully determined for SBW25 during competition with Rlv3841 for rhizosphere growth and colonisation of pea and barley. For Rlv3841 however, fitness values were only calculated for colonisation of barley

during competition with SBW25. Rlv3841 genes identified to have a role in barley root colonisation specifically during competition with SBW25 highlighted the need for altered transport, metabolic and regulatory requirements (Fig. 5.22). Interestingly TraI which is predicted to catalyse the production of N-acyl homoserine lactones (AHL), key quorum sensing molecules, was specifically required for barley root colonisation during competition with SBW25. Whilst *traI* mediates quorum-sensing regulated plasmid transfer in the closely related strain *R. leguminosarum* A34 this does not appear to be the case in Rlv3841 (Danino et al., 2003; McAnulla et al., 2007). It is possible that it may be involved in the mediation of a yet unidentified process in a cell density dependent manner. Whilst it is difficult to identify genes involved in quorum sensing using screen such as BarSeq, since other mutants within the population produce the quorum signal, a recent study using BarSeq enable identification of a T3SS as a requirement for infection of the apoplast by the plant pathogen *P.* syringe. This suggested that cell densities within the apoplast were too low for enough effectors to be produced to trans-complement these mutants impaired in effector release and so a similar phenomenon may be true for AHL production via *traI* during competitive colonisation of barley.

In the remaining three experiments an insufficient number of barcode reads for Rlv3841 mutants were obtained to support calculation of gene fitness values. The reason for this was that SBW25 made up a far greater proportion of the mutant population retrieved under these experimental conditions. There are several ways to overcome this limitation. Firstly, increased sequencing depth would theoretically provide sufficient data for calculation of Rlv3841 gene fitness contributions however, the number of reads required for certain conditions could become unfeasible. For

example, whilst the number of reads required for calculation of gene fitness values will vary among bacteria according to the size of their genome, a sequencing depth of 10 million reads is generally sufficient. Therefore, if a given strain accounts for 10% of the population a sequencing depth of 100 million reads would be required. Secondly, the Rlv3841 barcoded library could be co-inoculated with wild-type SBW25. The downside to this solution is that it would not be possible to simultaneously assay gene fitness values for both strains and therefore the experiment would have to be repeated with reversed barcoding in order to determine fitness contributions for both strains i.e., co-inoculation of a SBW25 barcoded transposon library with wild-type Rlv3841. Given that a long-term goal would be to use BarSeq to simultaneously assay gene fitness contribution for small SynComs containing approximately seven strains, such a solution would require seven independent experiments to ascertain fitness contributions for all seven strains in just one conditions. The third and most practical option would be to redesign the barcode itself. Currently the 20-nucleotide random barcode is flanked by universal primer binding sites (U1 and U2) and these are the same for both the Rlv3841- and SBW25-barcoded libraries. Modification of the universal primer binding sites, so that each RB-TnSeq library is constructed with a unique pair, will allow selective PCR amplification in a species-specific manner. In this way you can simultaneously assay gene fitness contributions for multiple strains without the need to significantly increase sequencing depth or experimental number. It would also remove the requirement to filter out mutants containing barcodes present in other RB-TnSeq libraries during BarSeq data analysis, since based on the universal primer binding sequence it would be possible to identify which strain that barcode, and therefore mutant, belonged to.

6.5. Identification of common determinants of SBW25 rhizosphere growth and colonisation and those specifically required during competition with Rlv3841

SBW25 is of agricultural interest due to is plant growth prompting capabilities. The BarSeq screens conducted here represent the first genome-wides analyses of factors contributing to SBW25 fitness in plant rhizospheres and during root colonisation. Analysis of SBW25 genes important for growth in the rhizosphere and colonisation of pea and barley during single inoculation and competition with Rlv3841, allowed the identification of a core set of genes required for rhizosphere growth and/or root colonisation under all conditions (Fig. 5.23). As seen with Rlv3841, these genes included those involved in the biosynthesis of amino acids, ribonucleotides, cofactors and peptidoglycan. Metabolism of glucose via the glycolysis/gluconeogenesis pathway was also important for rhizosphere growth, as was phospholipid transport via the *mla* pathway (Hughes et al., 2019). Genes with rhizosphere-progressive phenotypes included CusS, a sensor kinase involved in coper homeostasis (Munson et al., 2000), and several genes involved in LPS biosynthesis or modification. Attachment to the root surface required several lipoproteins which may have a direct role in adhesion and RsmA a RNA-binding post transcriptional regulator shown to regulate many processes in *Pseudomonas* including metabolism, adherence, chemotaxis and motility, virulence via T3SS, polysaccharide biosynthesis, biofilm formation, stress responses and iron acquisition (Song et al., 2015; Sobrero and Valverde, 2020; Liu et al., 2021).

During competition with Rlv3841, SBW25 showed a clear increased reliance on chemotaxis and flagella-mediated motility during colonisation of pea and barley (Fig. 5.14). This was demonstrated by stronger fitness impairments for genes belonging to both the *che1* cluster and those involved in flagella biosynthesis, export and assembly. Whilst its function is still elusive the *rsp* gene cluster, previously shown to be induced in plant rhizospheres (Preston et al., 2001; Jackson et al., 2005), appeared to contribute more strongly to rhizosphere growth during competition with Rlv3841 (Fig. 5.17). SBW25 also showed altered requirements in the biosynthesis of aromatic amino acids. During competitive colonisation of the pea rhizosphere biosynthesis of aromatic amino acids contributed to fitness whilst during competitive colonisation of the barley rhizosphere TyrR dependent repression of aromatic amino acid biosynthesis contributed to fitness (Fig.5.24 and Fig 5.25).

6.6. Limitations

Several limitations were highlighted throughout the work described in this thesis. Firstly, only approximately 26% and 30% of unique barcodes seen during the initial mapping of the Rlv3841- and SBW25-barcoded mutant libraries respectively were successfully mapped to the genome. This was primarily because unmapped barcodes had less than five reads mapping to the genome. Increased sequencing of these initial mutant libraries should result in the mapping of a higher proportion of barcodes to the genome and therefore allow the identification of more strains which can subsequently be used for the calculation of gene fitness values. This will allow the reanalysis of the data already obtained and provide even higher confidence in gene fitness predictions

due to increased internal replication. It may also facilitate the calculation of gene fitness values for those which currently have insufficient reads at time zero.

Another limitation observed during these BarSeq screens was the maintenance of mutant diversity *in plant*. Whilst the initial time zero inoculum for each experiment contained between an average of 73% of the total mapped barcodes present in the mutant pool, libraries retrieved form the rhizosphere and root that passed quality controls contained an average of 31% and 27% respectively. This is a common problem with transposon mutagenesis screens conducted *in planta* (Cole et al., 2017; Helmann et al., 2019). To overcome this a larger initial inoculum would be required either through inoculation of higher densities at time zero or through inoculation of a larger number of plants. Nonetheless these libraries had particularly high numbers of mapped barcoded strains and so even with a loss in mutant diversity, populations retrieved from these root environments were sufficient for robust calculation of gene fitness and retained relatively high insertion densities.

Lastly, BarSeq is unlikely to identify factors that can be complemented *in trans*. For example a strain impaired in siderophore production, widely known to be important during competition in the rhizosphere (Guerinot, 1991), would not be identified in BarSeq screens since near isogenic strains present in the mutant population synthesise and release siderophores. This would also apply to genes involved in the production quorum sensing signal, antimicrobial compounds or metabolic intermediates that may be taken up and utilised by a nonproducing strain.

6.7. Concluding remarks and future perspectives

The work described in this thesis represents the first use of RB-TnSeq in *Rhizobium* and provides a better understanding of genes at the genome-wide scale that contribute to growth in the rhizosphere and colonisation of its host legumes (pea, lentil and *Lathyrus*), the non-host legume (alfalfa) and the non-legume (barley) during single inoculations and in the case of barley root colonisation, during competition with SBW25. This work also represents the first time that BarSeq has been used to simultaneously assay gene fitness contributions for two different species during competition. Whilst there were some limitations to this, a slight modification to the design of the barcode region should allow efficient assay of gene fitness contributions in mixed cultures where each species is present in unequal ratios.

The results discussed here highlight that multiple factors contribute to Rlv3841 and SBW25 rhizosphere growth and root colonisation, far more than previously described in the literature, and that these are influenced by both plant species and the presence of competitors. Notably though, most of these genes have mild fitness contributions (fitness value between -1 and -2). They also reinforce that during root colonisation the processes of transport, metabolism and adaption to environmental stresses play an equally important role as factors involved in movement towards the root, adhesion itself and subsequent biofilm formation. Whilst the analysis described here focused on genes with strong (fitness value less than -2) and mild (fitness value between -1 and -2) fitness contributions, these datasets may also be used to identify genes with weaker but likely biologically significant contributions (i.e., fitness values between -0.42 and -1). Furthermore, these datasets provide the opportunity to identify genes

that when mutated provide a growth advantage in the rhizosphere or result in increased root colonisation. This is not surprising as despite the general conception that bacteria fine-tune their gene expression as they adapt and respond to environmental conditions, this is generally not the case (Price et al., 2013). In fact, a recent study assessing *P. syringae* colonisation of plant niches found a lack of correlation between gene expression, measured by microarray analysis or RNA sequencing, and gene essentiality determined from BarSeq experiments (Helmann et al., 2019).

In the wider context, the work described here provides a greater understanding of the factors contributing to rhizosphere persistence and root colonisation for two plant-beneficial bacteria. Since both factors are prerequisites for the establishment of beneficial plant-microbe interactions, this work provides a valuable resource to further our understanding of how these imitate associations develop and to possibly exploit them for agricultural benefit through the engineering of strains with enhanced rhizosphere competitiveness and root colonising abilities.

Bibliography

Abdel-Lateif, K., Bogusz, D., and Hocher, V. (2012) The role of flavonoids in the establishment of plant roots endosymbioses with arbuscular mycorrhiza fungi, rhizobia and *Frankia* bacteria. *Plant Signal Behav* **7**: 636-641.

Abdian, P.L., Caramelo, J.J., Ausmees, N., and Zorreguieta, A. (2013) RapA2 is a calcium-binding lectin composed of two highly conserved cadherin-like domains that specifically recognize *Rhizobium leguminosarum* acidic exopolysaccharides. *J Biol Chem* **288**: 2893-2904.

Aguilar, G.R., and Soberón, M. (1996) Cloning and sequence analysis of the *Rhizobium etli ccmA* and *ccmB* genes involved in c-type cytochrome biogenesis. *Gene* **182**: 129-135.

Aira, M., Gómez-Brandón, M., Lazcano, C., Bååth, E., and Domínguez, J. (2010) Plant genotype strongly modifies the structure and growth of maize rhizosphere microbial communities. *Soil Biol Biochem* **42**: 2276-2281.

Albright, M.B.N., Louca, S., Winkler, D.E., Feeser, K.L., Haig, S.J., Whiteson, K.L. et al. (2021) Solutions in microbiome engineering: prioritizing barriers to organism establishment. *ISME J*.

Alexandre, G. (2015) Chemotaxis control of transient cell aggregation. *J Bacteriol* **197**: 3230-3237.

Allard-Massicotte, R., Tessier, L., Lécuyer, F., Lakshmanan, V., Lucier, J.-F., Garneau, D. et al. (2016) *Bacillus subtilis* early colonization of *Arabidopsis thaliana* roots involves multiple chemotaxis receptors. *mBio* **7**: e01664-01616.

Alunni, B., and Gourion, B. (2016) Terminal bacteroid differentiation in the legume-*rhizobium* symbiosis: nodule-specific cysteine-rich peptides and beyond. *New Phytol* **211**: 411-417.

Alvarez Crespo, M.C., and Valverde, C. (2009) A single mutation in the *oprF* mRNA leader confers strict translational control by the Gac/Rsm system in *Pseudomonas fluorescens* CHA0. *Curr Microbiol* **58**: 182-188.

Ames, P., and Bergman, K. (1981) Competitive advantage provided by bacterial motility in the formation of nodules by *Rhizobium meliloti*. *J Bacteriol* **148**: 728-729.

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

Anderson, P.E., and Gober, J.W. (2000) FlbT, the post-transcriptional regulator of flagellin synthesis in *Caulobacter crescentus*, interacts with the 5′ untranslated region of flagellin mRNA. *Mol Microbiol* **38**: 41-52.

Andrews, M., and Andrews, M.E. (2017) Specificity in legume-rhizobia symbioses. *Int J Mol Sci* **18**: 705.

Angus, A.A., and Hirsch, A.M. (2013) Biofilm formation in the rhizosphere: multispecies interactions and implications for plant growth. In *Molecular Microbial Ecology of the Rhizosphere*, pp. 701-712.

Armitage, J.P. (1999) Bacterial tactic responses. *Adv Microb Physiol* **41**: 229-289.

Armitage, J.P., Gallagher, A., and Johnston, A.W.B. (1988) Comparison of the chemotactic behaviour of *Rhizobium leguminosarum* with and without the nodulation plasmid. *Mol Microbiol* **2**: 743-748.

Aufrecht, J.A., Timm, C.M., Bible, A., Morrell-Falvey, J.L., Pelletier, D.A., Doktycz, M.J., and Retterer, S.T. (2018) Quantifying the spatiotemporal dynamics of plant root colonization by beneficial bacteria in a microfluidic habitat. *Adv Biosyst* **2**: 1800048.

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

Ausmees, N., Jonsson, H., Höglund, S., Ljunggren, H., and Lindberg, M. (1999) Structural and putative regulatory genes involved in cellulose synthesis in *Rhizobium leguminosarum* bv. *trifolii*. *Microbiology* **145**: 1253-1262.

Azpilicueta, C.E., Zawoznik, M.S., and Tomaro, M.a.L. (2004) Phytoalexins synthesis is enhanced in groundnut plants inoculated with *Bradyrhizobium* sp. (*Arachis*). *Crop Protection* **23**: 1069-1074.

Azzolina, B.A., Yuan, X., Anderson, M.S., and El-Sherbeini, M. (2001) The cell wall and cell division gene cluster in the *Mra* operon of *Pseudomonas aeruginosa*: cloning, production, and purification of sctive enzymes. *Protein Expr Purif* **21**: 393-400.

Bacik, J.P., Whitworth, G.E., Stubbs, K.A., Yadav, A.K., Martin, D.R., Bailey-Elkin, B.A. et al. (2011) Molecular basis of 1,6-anhydro bond cleavage and phosphoryl transfer by *Pseudomonas aeruginosa* 1,6-anhydro-N-acetylmuramic acid kinase. *J Biol Chem* **286**: 12283-12291.

Bai, Y., Muller, D.B., Srinivas, G., Garrido-Oter, R., Potthoff, E., Rott, M. et al. (2015) Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature* **528**: 364-369.

Bale, S., Rajashankar, K.R., Perry, K., Begley, T.P., and Ealick, S.E. (2010) HMP binding protein ThiY and HMP-P synthase THI5 are structural homologues. *Biochemistry* **49**: 8929-8936.

Balsanelli, E., Serrato, R.V., De Baura, V.A., Sassaki, G., Yates, M.G., Rigo, L.U. et al. (2010) *Herbaspirillum seropedicae rfbB* and *rfbC* genes are required for maize colonization. *Environ Microbiol* **12**: 2233-2244.

Barbour, W.M., Hattermann, D.R., and Stacey, G. (1991) Chemotaxis of *Bradyrhizobium japonicum* to soybean exudates. *Appl Environ Microbiol* **57**: 2635–2639.

Bardgett, R.D., and Van Der Putten, W.H. (2014) Belowground biodiversity and ecosystem functioning. *Nature* **515**: 505-511.

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

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

Bauer, W.D., and Caetano-Anollés, G. (1990) Chemotaxis, induced gene expression and competitiveness in the rhizosphere. *Plant Soil* **129**: 45-52.

Bentzmann, S.d., Aurouze, M., Ball, G., and Filloux, A. (2006) FppA, a novel *Pseudomonas aeruginosa* prepilin peptidase involved in assembly of Type IVb Pili. *J Bacteriol* **188**: 4851-4860.

Berendsen, R.L., Pieterse, C.M., and Bakker, P.A. (2012) The rhizosphere microbiome and plant health. *Trends Plant Sci* **17**: 478-486.

Berendsen, R.L., Vismans, G., Yu, K., Song, Y., de Jonge, R., Burgman, W.P. et al. (2018) Disease-induced assemblage of a plant-beneficial bacterial consortium. *ISME J* **12**: 1496-1507.

Beringer, J.E. (1974) R Factor transfer in *Rhizobium leguminosarum*. *J Gen Microbiol* **84**: 188-198.

Berne, C., Ducret, A., Hardy, G.G., and Brun, Y.V. (2015) Adhesins involved in attachment to abiotic surfaces by gram-negative bacteria. *Microbiol Spectr* **3**.

Bertani, B., and Ruiz, N. (2018) Function and biogenesis of lipopolysaccharides. *EcoSal Plus* **8**: 10.1128/ecosalplus.ESP-0001-2018.

Blanco, L.P., Evans, M.L., Smith, D.R., Badtke, M.P., and Chapman, M.R. (2012) Diversity, biogenesis and function of microbial amyloids. *Trends Microbiol* **20**: 66-73.

Bodenhausen, N., Bortfeld-Miller, M., Ackermann, M., and Vorholt, J.A. (2014) A synthetic community approach reveals plant genotypes affecting the phyllosphere microbiota. *PLoS Genet* **10**: e1004283.

Boivin, S., Ait Lahmidi, N., Sherlock, D., Bonhomme, M., Dijon, D., Heulin-Gotty, K. et al. (2020) Host-specific competitiveness to form nodules in *Rhizobium leguminosarum* symbiovar *viciae*. *New Phytol* **226**: 555-568.

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

Bourassa, D.V., Kannenberg, E.L., Sherrier, D.J., Buhr, R.J., and Carlson, R.W. (2017) The lipopolysaccharide lipid A long-chain fatty acid is important for *Rhizobium leguminosarum* growth and stress adaptation in free-living and nodule environments. *Mol Plant Microbe Interact* **30**: 161-175.

Bourion, V., Heulin-Gotty, K., Aubert, V., Tisseyre, P., Chabert-Martinello, M., Pervent, M. et al. (2018) Co-inoculation of a pea core-collection with diverse rhizobial strains shows competitiveness for nodulation and efficiency of nitrogen fixation are distinct traits in the interaction. *Front Plant Sci* **8**: 2249.

Boutte, C.C., and Crosson, S. (2013) Bacterial lifestyle shapes stringent response activation. *Trends Microbiol* **21**: 174-180.

Branda, S.S., Vik, Å., Friedman, L., and Kolter, R. (2005) Biofilms: the matrix revisited. *Trends Microbiol* **13**: 20-26.

Breedveld, M.W., and Miller, K.J. (1994) Cyclic beta-glucans of members of the family *Rhizobiaceae*. *Microbiol Rev* **58**: 145-161.

Brewin, N.J., Wood, E.A., and Young, J.P.W. (1983) Contribution of the symbiotic plasmid to the competitiveness of *Rhizobium leguminosarum*. *Microbiology* **129**: 2973-2977.

Brown, D.B., Muszynski, A., and Carlson, R.W. (2013) Elucidation of a novel lipid A α-(1,1)-GalA transferase gene (*rgtF*) from *Mesorhizobium lot*i: heterologous expression of *rgtF* causes *Rhizobium etli* to synthesize lipid A with α-(1,1)-GalA. *Glycobiology* **23**: 546-558.

Brown, D.R., Barton, G., Pan, Z., Buck, M., and Wigneshweraraj, S. (2014) Nitrogen stress response and stringent response are coupled in *Escherichia coli*. *Nat Commun* **5**: 4115.

Bulgarelli, D., Schlaeppi, K., Spaepen, S., Van Themaat, E.V.L., and Schulze-Lefert, P. (2013) Structure and functions of the bacterial microbiota of plants. *Annu Rev Plant Biol* **64**: 807-838.

Bulgarelli, D., Garrido-Oter, R., Philipp, Weiman, A., Dröge, J., Pan, Y. et al. (2015) Structure and function of the bacterial root microbiota in wild and domesticated barley. *Cell Host Microbe* **17**: 392-403.

Bulgarelli, D., Rott, M., Schlaeppi, K., Ver Loren van Themaat, E., Ahmadinejad, N., Assenza, F. et al. (2012) Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. *Nature* **488**: 91-95.

Burdman, S., Okon, Y., and Jurkevitch, E. (2000) Surface characteristics of *Azospirillum brasilense* in relation to cell aggregation and attachment to plant roots. *Crit Rev Microbiol* **26**: 91-110.

Burdman, S., Dulguerova, G., Okon, Y., and Jurkevitch, E. (2001) Purification of the major outer membrane protein of *Azospirillum brasilense*, its affinity to plant roots, and its involvement in cell aggregation. *Mol Plant Microbe Interact* **14**: 555-561.

Busby, P.E., Soman, C., Wagner, M.R., Friesen, M.L., Kremer, J., Bennett, A. et al. (2017) Research priorities for harnessing plant microbiomes in sustainable agriculture. *PLOS Biol* **15**: e2001793.

Caetano-Anollés, G., Lagares, A., and Favelukes, G. (1989) Adsorption of *Rhizobium meliloti* to alfalfa roots: dependence on divalent cations and pH. *Plant Soil* **117**: 67-74.

Cai, S.J., and Inouye, M. (2002) EnvZ-OmpR interaction and osmoregulation in *Escherichia coli*. *J Biol Chem* **277**: 24155-24161.

Cai, T., Cai, W., Zhang, J., Zheng, H., Tsou, A.M., Xiao, L. et al. (2009) Host legume-exuded antimetabolites optimize the symbiotic rhizosphere. *Mol Microbiol* **73**: 507-517.

Cain, A.K., Barquist, L., Goodman, A.L., Paulsen, I.T., Parkhill, J., and Van Opijnen, T. (2020) A decade of advances in transposon-insertion sequencing. *Nat Rev Genet* **21**: 526-540.

Canarini, A., Kaiser, C., Merchant, A., Richter, A., and Wanek, W. (2019) Root exudation of primary metabolites: mechanisms and their roles in plant responses to environmental stimuli. *Front Plant Sci* **10**: 157.

Canter Cremers, H.C., Batley, M., Redmond, J.W., Eydems, L., Breedveld, M.W., Zevehuizen, L.P. et al. (1990) *Rhizobium leguminosarum exoB* mutants are deficient in the synthesis of UDP-glucose 4'-epimerase. *J Biol Chem* **265**: 21122-21127.

Capdevila, S., Martínez-Granero, F.M., Sánchez-Contreras, M., Rivilla, R., and Martín, M. (2004) Analysis of *Pseudomonas fluorescens* F113 genes implicated in flagellar filament synthesis and their role in competitive root colonization. *Microbiology* **150**: 3889-3897.

Carroll, D., Holden, N., Gifford, M.L., and Dupuy, L.X. (2020) Framework for quantification of the dynamics of root colonization by *Pseudomonas fluorescens* isolate SBW25. *Front Microbiol* **11**: 2403.

Carter, R.A., Worsley, P.S., Sawers, G., Challis, G.L., Dilworth, M.J., Carson, K.C. et al. (2002) The *vbs* genes that direct synthesis of the siderophore vicibactin in *Rhizobium leguminosarum*: their expression in other genera requires ECF σ factor RpoI. *Mol Microbiol* **44**: 1153-1166.

Carvalhais, L.C., Dennis, P.G., Badri, D.V., Kidd, B.N., Vivanco, J.M., and Schenk, P.M. (2015) Linking jasmonic acid signaling, root exudates, and rhizosphere microbiomes. *Mol Plant Microbe Interact* **28**: 1049-1058.

Carvalhais, L.C., Dennis, P.G., Fan, B., Fedoseyenko, D., Kierul, K., Becker, A. et al. (2013) Linking plant nutritional status to plant-microbe interactions. *PLOS One* **8**: e68555.

Cha, C., Gao, P., Chen, Y.C., Shaw, P.D., and Farrand, S.K. (1998) Production of acyl-homoserine lactone quorum-sensing signals by gram-negative plant-associated bacteria. *Mol Plant Microbe Interact* **11**: 1119-1129.

Chang, W.S., Park, K.M., Koh, S.C., and So, J.S. (2008) Characterization of the *Bradyrhizobium japonicum galE* gene: its impact on lipopolysaccharide profile and nodulation of soybean. *FEMS Microbiol Lett* **280**: 242-249.

Chapman, M.R., Robinson, L.S., Pinkner, J.S., Roth, R., Heuser, J., Hammar, M. et al. (2002) Role of *Escherichia coli* curli operons in directing amyloid fiber formation. *Science* **295**: 851-855.

Charles, T.C., and Nester, E.W. (1993) A chromosomally encoded two-component sensory transduction system is required for virulence of *Agrobacterium tumefaciens*. *J Bacteriol* **175**: 6614-6625.

Chen, H., Higgins, J., Kondorosi, E., Kondorosi, A., Djordjevic, M.A., Weinman, J.J., and Rolfe, B.G. (2000) Identification of *nolR*-regulated proteins in *Sinorhizobium meliloti* using proteome analysis. *Electrophoresis* **21**: 3823-3832.

Cheng, G., Shapir, N., Sadowsky, M.J., and Wackett, L.P. (2005) Allophanate hydrolase, not urease, functions in bacterial cyanuric acid metabolism. *Appl Environ Microbiol* **71**: 4437-4445.

Cheng, Z., Lei, S., Li, Y., Huang, W., Ma, R., Xiong, J. et al. (2020) Revealing the variation and stability of bacterial communities in tomato rhizosphere microbiota. *Microorganisms* **8**: 170.

Cole, B.J., Feltcher, M.E., Waters, R.J., Wetmore, K.M., Mucyn, T.S., Ryan, E.M. et al. (2017) Genome-wide identification of bacterial plant colonization genes. *PLOS Biol* **15**: e2002860-e2002860.

Comolli, J.C., Hauser, A.R., Waite, L., Whitchurch, C.B., Mattick, J.S., and Engel, J.N. (1999) *Pseudomonas aeruginosa* gene products PilT and PilU are required for cytotoxicity in vitro and virulence in a mouse model of acute pneumonia. *Infect Immun* **67**: 3625-3630.

Cooley, M.B., Miller, W.G., and Mandrell, R.E. (2003) Colonization of *Arabidopsis thaliana* with *Salmonella enterica* and enterohemorrhagic *Escherichia coli* O157:H7 and competition by *Enterobacter asburiae*. *Appl Environ Microbiol* **69**: 4915-4926.

Craig, L., Pique, M.E., and Tainer, J.A. (2004) Type IV pilus structure and bacterial pathogenicity. *Nat Rev Microbiol* **2**: 363-378.

Croes, C.L., Moens, S., van Bastelaere, E., Vanderleyden, J., and Michiels, K.W. (1993) The polar flagellum mediates *Azospirillum brasilense* adsorption to wheat roots. *Microbiology* **139**: 2261-2269.

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

Cucarella, C., Solano, C., Valle, J., Amorena, B., Lasa, I., and Penadés, J.R. (2001) Bap, a *Staphylococcus aureus* surface protein involved in biofilm formation. *J Bacteriol* **183**: 2888-2896.

D'Haeze, W., and Holsters, M. (2002) Nod factor structures, responses, and perception during initiation of nodule development. *Glycobiology* **12**: 79r-105r.

da Silva Lima, L., Olivares, F.L., Rodrigues de Oliveira, R., Vega, M.R.G., Aguiar, N.O., and Canellas, L.P. (2014) Root exudate profiling of maize seedlings inoculated with *Herbaspirillum seropedicae* and humic acids. *Chem biol technol agric* **1**: 23.

Daniel, R. (2005) The metagenomics of soil. *Nat Rev Microbiol* **3**: 470-478.

Danino, V.E., Wilkinson, A., Edwards, A., and Downie, J.A. (2003) Recipient-induced transfer of the symbiotic plasmid pRL1JI in *Rhizobium leguminosarum* bv. *viciae* is regulated by a quorum-sensing relay. *Mol Microbiol* **50**: 511-525.

Dasgupta, N., Arora, S.K., and Ramphal, R. (2000) *fleN*, a gene that regulates flagellar number in *Pseudomonas aeruginosa*. *J Bacteriol* **182**: 357-364.

Dasgupta, N., Wolfgang, M.C., Goodman, A.L., Arora, S.K., Jyot, J., Lory, S., and Ramphal, R. (2003) A four-tiered transcriptional regulatory circuit controls flagellar biogenesis in *Pseudomonas aeruginosa*. *Mol Microbiol* **50**: 809-824.

Davey, M.E., and O'Toole, G.A. (2000) Microbial biofilms: from ecology to molecular genetics. *Microbiol Mol Biol* **64**: 847-867.

De Leij, F.A.A.M., Sutton, E.J., Whipps, J.M., Fenlon, J.S., and Lynch, J.M. (1995) Field release of a genetically modified *Pseudomonas fluorescens* on wheat: establishment, survival and dissemination. *Bio/Technology* **13**: 1488-1492.

De Mot, R., and Vanderleyden, J. (1991) Purification of a root-adhesive outer membrane protein of root-colonizing *Pseudomonas fluorescens*. *FEMS Microbiol Lett* **81**: 323-327.

de Souza, R.S., Okura, V.K., Armanhi, J.S., Jorrin, B., Lozano, N., da Silva, M.J. et al. (2016) Unlocking the bacterial and fungal communities assemblages of sugarcane microbiome. *Sci Rep* **6**: 28774.

de Weert, S., Vermeiren, H., Mulders, I.H.M., Kuiper, I., Hendrickx, N., Bloemberg, G.V. et al. (2002) Flagella-driven chemotaxis towards exudate components is an

important trait for tomato root colonization by *Pseudomonas fluorescens*. *Mol Plant Microbe Interact* **15**: 1173-1180.

Decoin, V., Gallique, M., Barbey, C., Le Mauff, F., Poc, C.D., Feuilloley, M.G.J. et al. (2015) A *Pseudomonas fluorescens* type 6 secretion system is related to mucoidy, motility and bacterial competition. *BMC Microbiol* **15**: 72-72.

Deines, P., and Bosch, T.C.G. (2016) Transitioning from microbiome composition to microbial community interactions: the potential of the metaorganism hydra as an experimental model. *Front Microbiol* **7**: 1610.

DeJesus, M.A., and Ioerger, T.R. (2013) A Hidden Markov Model for identifying essential and growth-defect regions in bacterial genomes from transposon insertion sequencing data. *BMC Bioinformatics* **14**: 303.

Delgado, M.J., Bedmar, E.J., and Downie, J.A. (1998) Genes involved in the formation and assembly of rhizobial cytochromes and their role in symbiotic nitrogen fixation. *Adv Microb Physiol* **40**: 191-231.

Delgado, M.J., Yeoman, K.H., Wu, G., Vargas, C., Davies, A.E., Poole, R.K. et al. (1995) Characterization of the *cycHJKL* genes involved in cytochrome c biogenesis and symbiotic nitrogen fixation in *Rhizobium leguminosarum*. *J Bacteriol* **177**: 4927-4934.

Denison, R.F., and Kiers, E.T. (2004) Lifestyle alternatives for rhizobia: mutualism, parasitism, and forgoing symbiosis. *FEMS Microbiol Lett* **237**: 187-193.

Devedjiev, Y., Surendranath, Y., Derewenda, U., Gabrys, A., Cooper, D.R., Zhang, R.G. et al. (2004) The structure and ligand binding properties of the *B. subtilis* YkoF gene product, a member of a novel family of thiamin/HMP-binding proteins. *J Mol Biol* **343**: 395-406.

Dharmatilake, A.J., and Bauer, W.D. (1992) Chemotaxis of *Rhizobium meliloti* towards nodulation gene-inducing compounds from alfalfa roots. *Appl Environ Microbiol* **58**: 1153-1158.

Ding, H., Yip, C.B., Geddes, B.A., Oresnik, I.J., and Hynes, M.F. (2012) Glycerol utilization by *Rhizobium leguminosarum* requires an ABC transporter and affects competition for nodulation. *Microbiology* **158**: 1369-1378.

Ditta, G., Stanfield, S., Corbin, D., and Helinski, D.R. (1980) Broad host range DNA cloning system for gram-negative bacteria: construction of a gene bank of *Rhizobium meliloti*. *Proc Natl Acad Sci U S A* **77**: 7347-7351.

Dogra, G., Purschke, F.G., Wagner, V., Haslbeck, M., Kriehuber, T., Hughes, J.G. et al. (2012) *Sinorhizobium meliloti* CheA complexed with CheS exhibits enhanced binding to CheY1, resulting in accelerated CheY1 dephosphorylation. *J Bacteriol* **194**: 1075-1087.

Donlan, R.M. (2002) Biofilms: microbial life on surfaces. *Emerg Infect Dis* **8**: 881-890.

Dörr, J., Hurek, T., and Reinhold-Hurek, B. (1998) Type IV pili are involved in plant-microbe and fungus-microbe interactions. *Mol Microbiol* **30**: 7-17.

Downie, J.A. (2010) The roles of extracellular proteins, polysaccharides and signals in the interactions of rhizobia with legume roots. *FEMS Microbiol Rev* **34**: 150-170.

Dudenhöffer, J.-H., Scheu, S., and Jousset, A. (2016) Systemic enrichment of antifungal traits in the rhizosphere microbiome after pathogen attack. *J Ecol* **104**: 1566-1575.

Duong, D.A., Jensen, R.V., and Stevens, A.M. (2018) Discovery of *Pantoea stewartii* ssp. *stewartii* genes important for survival in corn xylem through a Tn-Seq analysis. *Molecular Plant Pathology* **19**: 1929-1941.

Duque, E., de la Torre, J., Bernal, P., Molina-Henares, M.A., Alaminos, M., Espinosa-Urgel, M. et al. (2013) Identification of reciprocal adhesion genes in pathogenic and non-pathogenic *Pseudomonas*. *Environ Microbiol* **15**: 36-48.

Dylan, T., Helinski, D.R., and Ditta, G.S. (1990) Hypoosmotic adaptation in *Rhizobium meliloti* requires beta-(1----2)-glucan. *J Bacteriol* **172**: 1400-1408.

Edwards, J., Johnson, C., Santos-Medellín, C., Lurie, E., Podishetty, N.K., Bhatnagar, S. et al. (2015) Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc Natl Acad Sci U S A* **112**: E911-E920.

Emmert, E.A.B., Milner, J.L., Lee, J.C., Pulvermacher, K.L., Olivares, H.A., Clardy, J., and Handelsman, J. (1998) Effect of canavanine from alfalfa seeds on the population biology of *Bacillus cereus*. *Appl Environ Microbiol* **64**: 4683-4688.

Espinosa-Urgel, M., Salido, A., and Ramos, J.-L. (2000) Genetic analysis of functions involved in adhesion of *Pseudomonas putida* to seeds. *J Bacteriol* **182**: 2363-2369.

Estabrook, E.M., and Yoder, J.I. (1998) Plant-plant communications: rhizosphere signaling between parasitic angiosperms and their hosts. *Plant Physiol* **116**: 1-7.

Fischer, H.M. (1994) Genetic regulation of nitrogen fixation in rhizobia. *Microbiol Rev* **58**: 352-386.

Flemming, H.-C., and Wingender, J. (2010) The biofilm matrix. *Nat Rev Microbiol* **8**: 623-633.

Fournier, J., Timmers, A.C.J., Sieberer, B.J., Jauneau, A., Chabaud, M., and Barker, D.G. (2008) Mechanism of infection thread elongation in root hairs of *Medicago truncatula* and dynamic interplay with associated rhizobial colonization. *Plant Physiol* **148**: 1985-1995.

Fox, M.A., Karunakaran, R., Leonard, M.E., Mouhsine, B., Williams, A., East, A.K. et al. (2008) Characterization of the quaternary amine transporters of *Rhizobium leguminosarum* bv. *viciae* 3841. *FEMS Microbiol Lett* **287**: 212-220.

Frederix, M., Edwards, A., Swiderska, A., Stanger, A., Karunakaran, R., Williams, A. et al. (2014) Mutation of *praR* in *Rhizobium leguminosarum* enhances root biofilms, improving nodulation competitiveness by increased expression of attachment proteins. *Mol Microbiol* **93**: 464-478.

Frink, C.R., Waggoner, P.E., and Ausubel, J.H. (1999) Nitrogen fertilizer: retrospect and prospect. *Proc Natl Acad Sci U S A* **96**: 1175-1180.

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

Fuqua, W.C., Winans, S.C., and Greenberg, E.P. (1994) Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *J Bacteriol* **176**: 269-275.

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

Gage, D.J. (2004) Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiol Mol Biol Rev* **68**: 280-300.

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

Galbraith, M.P., Feng, S.F., Borneman, J., Triplett, E.W., de Bruijn, F.J., and Rossbachl, S. (1998) A functional *myo*-inositol catabolism pathway is essential for rhizopine utilization by *Sinorhizobium meliloti*. *Microbiology* **144**: 2915-2924.

Gallique, M., Bouteiller, M., and Merieau, A. (2017) The type VI secretion system: a dynamic system for bacterial communication? *Front Microbiol* **8**: 1454.

Gao, M., D'Haeze, W., De Rycke, R., Wolucka, B., and Holsters, M. (2001) Knockout of an azorhizobial dTDP-L-rhamnose synthase affects lipopolysaccharide and extracellular polysaccharide production and disables symbiosis with *Sesbania rostrata*. *Mol Plant Microbe Interact* **14**: 857-866.

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

Garcia-Fraile, P., Seaman, J.C., Karunakaran, R., Edwards, A., Poole, P.S., and Downie, J.A. (2015) Arabinose and protocatechuate catabolism genes are important for growth of *Rhizobium leguminosarum* biovar *viciae* in the pea rhizosphere. *Plant Soil* **390**: 251-264.

Garvis, S., Munder, A., Ball, G., de Bentzmann, S., Wiehlmann, L., Ewbank, J.J. et al. (2009) *Caenorhabditis elegans* semi-automated liquid screen reveals a specialized

role for the chemotaxis gene *cheB2* in *Pseudomonas aeruginosa* virulence. *PLOS Pathogens* **5**: e1000540.

Geurts, R., and Bisseling, T. (2002) *Rhizobium* nod factor perception and signalling. *The Plant cell* **14 Suppl**: S239-S249.

Gjermansen, M., Nilsson, M., Yang, L., and Tolker-Nielsen, T. (2010) Characterization of starvation-induced dispersion in *Pseudomonas putida* biofilms: genetic elements and molecular mechanisms. *Mol Microbiol* **75**: 815-826.

Göbel, M., Kassel-Cati, K., Schmidt, E., and Reineke, W. (2002) Degradation of aromatics and chloroaromatics by *Pseudomonas* sp. strain B13: cloning, characterization, and analysis of sequences encoding 3-oxoadipate:succinyl-coenzyme A (CoA) transferase and 3-oxoadipyl-CoA thiolase. *J Bacteriol* **184**: 216-223.

Gonzalez-Mula, A., Lachat, J., Mathias, L., Naquin, D., Lamouche, F., Mergaert, P., and Faure, D. (2019) The biotroph *Agrobacterium tumefaciens* thrives in tumors by exploiting a wide spectrum of plant host metabolites. *New Phytol* **222**: 455-467.

Goodman, A.L., McNulty, N.P., Zhao, Y., Leip, D., Mitra, R.D., Lozupone, C.A. et al. (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6**: 279-289.

Gourion, B., Rossignol, M., and Vorholt, J.A. (2006) A proteomic study of *Methylobacterium extorquens* reveals a response regulator essential for epiphytic growth. *Proc Natl Acad Sci U S A* **103**: 13186-13191.

Gourse, R.L., Chen, A.Y., Gopalkrishnan, S., Sanchez-Vazquez, P., Myers, A., and Ross, W. (2018) Transcriptional Responses to ppGpp and DksA. *Annu Rev Microbiol* **72**: 163-184.

Granato, E.T., Meiller-Legrand, T.A., and Foster, K.R. (2019) The evolution and ecology of bacterial warfare. *Curr Biol* **29**: R521-R537.

Green, M.R., and Sambrook, J. (2016) Precipitation of DNA with ethanol. *Cold Spring Harb Protoc* **2016**.

Green, R.T., East, A.K., Karunakaran, R., Downie, J.A., and Poole, P.S. (2019) Transcriptomic analysis of *Rhizobium leguminosarum* bacteroids in determinate and indeterminate nodules. *Microb Genom* **5**: e000254.

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

Griffitts, J.S., Carlyon, R.E., Erickson, J.H., Moulton, J.L., Barnett, M.J., Toman, C.J., and Long, S.R. (2008) A *Sinorhizobium meliloti* osmosensory two-component system required for cyclic glucan export and symbiosis. *Mol Microbiol* **69**: 479-490.

Groenewold, M.K., Hebecker, S., Fritz, C., Czolkoss, S., Wiesselmann, M., Heinz, D.W. et al. (2019) Virulence of *Agrobacterium tumefaciens* requires lipid homeostasis mediated by the lysyl-phosphatidylglycerol hydrolase AcvB. *Mol Microbiol* **111**: 269-286.

Guerinot, M.L. (1991) Iron uptake and metabolism in the rhizobia/legume symbioses. *Plant Soil* **130**: 199-209.

Güvener, Z.T., Tifrea, D.F., and Harwood, C.S. (2006) Two different *Pseudomonas aeruginosa* chemosensory signal transduction complexes localize to cell poles and form and remould in stationary phase. *Mol Microbiol* **61**: 106-118.

Haag, A.F., Arnold, M.F.F., Myka, K.K., Kerscher, B., Dall'Angelo, S., Zanda, M. et al. (2013) Molecular insights into bacteroid development during *Rhizobium*–legume symbiosis. *FEMS Microbiol Rev* **37**: 364-383.

Hahn, H.P. (1997) The type-4 pilus is the major virulence-associated adhesin of *Pseudomonas aeruginosa* – a review. *Gene* **192**: 99-108.

Hawkins, J.P., Ordonez, P.A., and Oresnik, I.J. (2017) Characterization of mutations that affect the nonoxidative pentose phosphate pathway in *Sinorhizobium meliloti*. *J Bacteriol* **200**: e00436-00417.

Helmann, J.D. (2002) The extracytoplasmic function (ECF) sigma factors. In *Adv Microb Physiol*: Academic Press, pp. 47-110.

Helmann, T.C., Deutschbauer, A.M., and Lindow, S.E. (2019) Genome-wide identification of *Pseudomonas syringae* genes required for fitness during colonization of the leaf surface and apoplast. *Proc Natl Acad Sci U S A*: 201908858.

Hibbing, M.E., Fuqua, C., Parsek, M.R., and Peterson, S.B. (2010) Bacterial competition: surviving and thriving in the microbial jungle. *Nat Rev Microbiol* **8**: 15-25.

Hinsa, S.M., Espinosa-Urgel, M., Ramos, J.L., and O'Toole, G.A. (2003) Transition from reversible to irreversible attachment during biofilm formation by *Pseudomonas fluorescens* WCS365 requires an ABC transporter and a large secreted protein. *Mol Microbiol* **49**: 905-918.

Hinsinger, P., Plassard, C., Tang, C., and Jaillard, B. (2003) Origins of root-mediated pH changes in the rhizosphere and their responses to environmental constraints: a review. *Plant Soil* **248**: 43-59.

Hittle, L.E., Powell, D.A., Jones, J.W., Tofigh, M., Goodlett, D.R., Moskowitz, S.M., and Ernst, R.K. (2015) Site-specific activity of the acyltransferases HtrB1 and HtrB2 in *Pseudomonas aeruginosa* lipid A biosynthesis. *Pathog Dis* **73**: ftv053.

Hosie, A.H.F., Allaway, D., Galloway, C.S., Dunsby, H.A., and Poole, P.S. (2002) *Rhizobium leguminosarum* has a second general amino acid permease with unusually broad substrate specificity and high similarity to branched-shain amino acid transporters (Bra/LIV) of the ABC Family. *J Bacteriol* **184**: 4071-4080.

Howieson, J.G., Robson, A.D., and Ewing, M.A. (1993) External phosphate and calcium concentrations, and pH, but not the products of rhizobial nodulation genes, affect the attachment of *Rhizobium meliloti* to roots of annual medics. *Soil Biol Biochem* **25**: 567-573.

Hughes, G.W., Hall, S.C.L., Laxton, C.S., Sridhar, P., Mahadi, A.H., Hatton, C. et al. (2019) Evidence for phospholipid export from the bacterial inner membrane by the Mla ABC transport system. *Nat Microbiol* **4**: 1692-1705.

Humphris, S.N., Bengough, A.G., Griffiths, B.S., Kilham, K., Rodger, S., Stubbs, V. et al. (2005) Root cap influences root colonisation by *Pseudomonas fluorescens* SBW25 on maize. *FEMS Microbiol Ecol* **54**: 123-130.

Hynes, M.F., and O'Connell, M.P. (1990) Host plant effect on competition among strains of *Rhizobium leguminosarum*. *Can J Microbiol* **36**: 864-869.

Iida, A., Teshiba, S., and Mizobuchi, K. (1993) Identification and characterization of the *tktB* gene encoding a second transketolase in *Escherichia coli* K-12. *J Bacteriol* **175**: 5375-5383.

Jackson, R.W., Preston, G.M., and Rainey, P.B. (2005) Genetic characterization of *Pseudomonas fluorescens* SBW25 *rsp* gene expression in the phytosphere and in vitro. *J Bacteriol* **187**: 8477-8488.

Jakubowski, S.J., Kerr, J.E., Garza, I., Krishnamoorthy, V., Bayliss, R., Waksman, G., and Christie, P.J. (2009) *Agrobacterium* VirB10 domain requirements for type IV secretion and T pilus biogenesis. *Mol Microbiol* **71**: 779-794.

Janczarek, M., and Rachwał, K. (2013) Mutation in the *pssA* gene involved in exopolysaccharide synthesis leads to several physiological and symbiotic defects in *Rhizobium leguminosarum* bv. *trifolii*. *Int J Mol Sci* **14**: 23711-23735.

Jijón-Moreno, S., Baca, B.E., Castro-Fernández, D.C., and Ramírez-Mata, A. (2019) TyrR is involved in the transcriptional regulation of biofilm formation and D-alanine catabolism in *Azospirillum brasilense* Sp7. *PLOS One* **14**: e0211904.

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

Johnston, A.W.B., and Beringer, J.E. (1975) Identification of the *Rhizobium* strains in pea root nodules using genetic markers. *Microbiology* **87**: 343-350.

Jones, K.M., Kobayashi, H., Davies, B.W., Taga, M.E., and Walker, G.C. (2007) How rhizobial symbionts invade plants: the *Sinorhizobium-Medicago* model. *Nat Rev Microbiol* **5**: 619-633.

Jorgenson, M.A., Chen, Y., Yahashiri, A., Popham, D.L., and Weiss, D.S. (2014) The bacterial septal ring protein RlpA is a lytic transglycosylase that contributes to rod

shape and daughter cell separation in *Pseudomonas aeruginosa*. *Mol Microbiol* **93**: 113-128.

Jorrin, B. (2016) Genomics of specificity in the symbiotic interaction between *Rhizobium leguminosarum* and legumes. PhD thesis, Licenciatura en Biologia.

Jorrin, B., and Imperial, J. (2015) Population genomics analysis of legume host preference for specific rhizobial genotypes in the *Rhizobium leguminosarum* bv. *viciae* symbioses. *Mol Plant Microbe Interact* **28**: 310-318.

Kandari, D., Joshi, H., and Bhatnagar, R. (2021) Zur: zinc-sensing transcriptional regulator in a diverse set of bacterial species. *Pathogens* **10**: 344.

Kandel, S., Joubert, P., and Doty, S. (2017) Bacterial endophyte colonization and distribution within plants. *Microorganisms* **5**: 77.

Kant, S., Bi, Y.M., and Rothstein, S.J. (2011) Understanding plant response to nitrogen limitation for the improvement of crop nitrogen use efficiency. *J Exp Bot* **62**: 1499-1509.

Karunakaran, R., Ebert, K., Harvey, S., Leonard, M.E., Ramachandran, V., and Poole, P.S. (2006) Thiamine is synthesized by a salvage pathway in *Rhizobium leguminosarum* bv. *viciae* strain 3841. *J Bacteriol* **188**: 6661-6668.

Karunakaran, R., Ramachandran, V.K., Seaman, J.C., East, A.K., Mouhsine, B., Mauchline, T.H. et al. (2009) Transcriptomic analysis of *Rhizobium leguminosarum* biovar *viciae* in symbiosis with host plants *Pisum sativum* and *Vicia cracca*. *J Bacteriol* **191**: 4002-4014.

Karunakaran, R., Haag, A.F., East, A.K., Ramachandran, V.K., Prell, J., James, E.K. et al. (2010) BacA is essential for bacteroid development in nodules of galegoid, but not phaseoloid, legumes. *J Bacteriol* **192**: 2920-2928.

Kelly, S.L., and Kelly, D.E. (2013) Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us? *Philos Trans R Soc Lond B Biol Sci* **368**: 20120476.

Kendall, K., and Roberts, A.D. (2015) Van der Waals forces influencing adhesion of cells. *Philos Trans R Soc Lond B Biol Sci* **370**: 20140078.

Kent, W.J. (2002) BLAT---The BLAST-Like Alignment Tool. *Genome Research* **12**: 656-664.

Keto-Timonen, R., Hietala, N., Palonen, E., Hakakorpi, A., Lindström, M., and Korkeala, H. (2016) Cold shock proteins: a minireview with special emphasis on Csp-family of *Enteropathogenic Yersinia*. *Front Microbiol* **7**: 1151.

King, E.O., Ward, M.K., and Raney, D.E. (1954) Two simple media for the demonstration of pyocyanin and fluorescin. *J Lab Clin Med* **44**: 301-307.

Kiss, E., Mergaert, P., Olàh, B., Kereszt, A., Staehelin, C., Davies, A.E. et al. (1998) Conservation of *noIR* in the *Sinorhizobium* and *Rhizobium* genera of the *Rhizobiaceae* family. *Mol Plant-Microbe Interact* **11**: 1186-1195.

Kluck, C.J., Patzelt, H., Genevaux, P., Brehmer, D., Rist, W., Schneider-Mergener, J. et al. (2002) Structure-function analysis of HscC, the *Escherichia coli* member of a novel subfamily of specialized Hsp70 chaperones. *J Biol Chem* **277**: 41060-41069.

Knights, H.E., Jorrin, B., Haskett, T.L., and Poole, P.S. (2021) Deciphering bacterial mechanisms of root colonization. *Environ Microbiol Rep* **13**: 428-444.

Kolde, R. (2012) Pheatmap: pretty heatmaps. *R package*: R Package Version 1.0.10.

Kong, H.G., Kim, B.K., Song, G.C., Lee, S., and Ryu, C.-M. (2016) Aboveground whitefly infestation-mediated reshaping of the root microbiota. *Front Microbiol* **7**: 1314.

Korenblum, E., Dong, Y., Szymanski, J., Panda, S., Jozwiak, A., Massalha, H. et al. (2020) Rhizosphere microbiome mediates systemic root metabolite exudation by root-to-root signaling. *Proc Natl Acad Sci U S A* **117**: 3874-3883.

Kosolapova, A.O., Belousov, M.V., Sulatskaya, A.I., Belousova, M.E., Sulatsky, M.I., Antonets, K.S. et al. (2019) Two novel amyloid proteins, RopA and RopB, from the root nodule bacterium *Rhizobium leguminosarum*. *Biomolecules* **9**: 694.

Krehenbrink, M., and Downie, J.A. (2008) Identification of protein secretion systems and novel secreted proteins in *Rhizobium leguminosarum* bv. *viciae*. *BMC Genomics* **9**: 55.

Kumar, S., and Trivedi, P.K. (2018) Glutathione s-transferases: role in combating abiotic stresses including arsenic detoxification in plants. *Front Plant Sci* **9**: 751.

Kuske, C.R., Ticknor, L.O., Miller, M.E., Dunbar, J.M., Davis, J.A., Barns, S.M., and Belnap, J. (2002) Comparison of soil bacterial communities in rhizospheres of three plant species and the interspaces in an arid grassland. *Appl Environ Microbiol* **68**: 1854-1863.

Laguerre, G., Louvrier, P., Allard, M.-R., and Amarger, N. (2003) Compatibility of rhizobial genotypes within natural populations of *Rhizobium leguminosarum* biovar *viciae* for nodulation of host legumes. *Appl Environ Microbiol* **69**: 2276-2283.

Lang, C., Barnett, M.J., Fisher, R.F., Smith, L.S., Diodati, M.E., and Long, S.R. (2018) Most *Sinorhizobium meliloti* extracytoplasmic function sigma factors control accessory functions. *mSphere* **3**: e00454-00418.

Lasa, I., and Penadés, J.R. (2006) Bap: A family of surface proteins involved in biofilm formation. *Res Microbiol* **157**: 99-107.

Laus, M.C., Logman, T.J., Lamers, G.E., Van Brussel, A.A.N., Carlson, R.W., and Kijne, J.W. (2006) A novel polar surface polysaccharide from *Rhizobium leguminosarum* binds host plant lectin. *Mol Microbiol* **59**: 1704-1713.

Lebeis, S.L., Paredes, S.H., Lundberg, D.S., Breakfield, N., Gehring, J., McDonald, M. et al. (2015) Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* **349**: 860-864.

Lee, H.Y., An, J.H., and Kim, Y.S. (2000) Identification and characterization of a novel transcriptional regulator, MatR, for malonate metabolism in *Rhizobium leguminosarum* bv. *trifolii*. *Europ J Biochem* **267**: 7224-7230.

Lemieux, M.J., Huang, Y., and Wang, D.-N. (2004) Glycerol-3-phosphate transporter of *Escherichia coli*: structure, function and regulation. *Res Microbiol* **155**: 623-629.

Lerouge, I., and Vanderleyden, J. (2002) O-antigen structural variation: mechanisms and possible roles in animal/plant–microbe interactions. *FEMS Microbiol Rev* **26**: 17-47.

Li, Y., Mortuza, R., Milligan, D.L., Tran, S.L., Strych, U., Cook, G.M., and Krause, K.L. (2014) Investigation of the essentiality of glutamate racemase in *Mycobacterium smegmatis*. *J Bacteriol* **196**: 4239-4244.

Little, R.H., Grenga, L., Saalbach, G., Howat, A.M., Pfeilmeier, S., Trampari, E., and Malone, J.G. (2016) Adaptive remodeling of the bacterial proteome by specific ribosomal modification regulates *Pseudomonas* infection and niche colonisation. *PLoS Genet* **12**: e1005837-e1005837.

Liu, H., Carvalhais, L.C., Schenk, P.M., and Dennis, P.G. (2017a) Effects of jasmonic acid signalling on the wheat microbiome differ between body sites. *Sci Rep* **7**: 41766.

Liu, J., Yu, M., Ge, Y., Tian, Y., Hu, B., and Zhao, Y. (2021) The RsmA RNA-binding proteins in *Pseudomonas* syringae exhibit distinct and overlapping roles in modulating virulence and survival under different nutritional conditions. *Front Plant Sci* **12**: 637595.

Liu, R., and Ochman, H. (2007) Stepwise formation of the bacterial flagellar system. *Proc Natl Acad Sci U S A* **104**: 7116-7121.

Liu, W., Bai, X., Li, Y., Min, J., Kong, Y., and Hu, X. (2020) CheY1 and CheY2 of *Azorhizobium caulinodans* ORS571 regulate chemotaxis and competitive colonization with the host plant. *Appl Environ Microbiol* **86**: e00599-00520.

Liu, Y., Rainey, P.B., and Zhang, X.-X. (2015) Molecular mechanisms of xylose utilization by *Pseudomonas fluorescen*s: overlapping genetic responses to xylose, xylulose, ribose and mannitol. *Mol Microbiol* **98**: 553-570.

Liu, Y., Jiang, X., Guan, D., Zhou, W., Ma, M., Zhao, B. et al. (2017b) Transcriptional analysis of genes involved in competitive nodulation in *Bradyrhizobium diazoefficiens* at the presence of soybean root exudates. *Sci Rep* **7**: 10946.

Liu, Z., Beskrovnaya, P., Melnyk, R.A., Hossain, S.S., Khorasani, S., O'Sullivan, L.R. et al. (2018) A genome-wide screen identifies genes in rhizosphere-associated *Pseudomonas* required to evade plant defenses. *mBio* **9**: e00433-00418.

Long, S.R. (1996) *Rhizobium* symbiosis: nod factors in perspective. *The Plant Cell* **8**: 1885-1898.

Lu, Y.-J., Zhang, Y.-M., Grimes, K.D., Qi, J., Lee, R.E., and Rock, C.O. (2006) Acyl-phosphates initiate membrane phospholipid synthesis in gram-positive pathogens. *Molecular Cell* **23**: 765-772.

Lundberg, D.S., Lebeis, S.L., Paredes, S.H., Yourstone, S., Gehring, J., Malfatti, S. et al. (2012) Defining the core *Arabidopsis thaliana* root microbiome. *Nature* **488**: 86-90.

Mangan, E.K., Malakooti, J., Caballero, A., Anderson, P., Ely, B., and Gober, J.W. (1999) FlbT couples flagellum assembly to gene expression in *Caulobacter crescentus*. *J Bacteriol* **181**: 6160-6170.

Martínez-Gil, M., Yousef-Coronado, F., and Espinosa-Urgel, M. (2010) LapF, the second largest *Pseudomonas putida* protein, contributes to plant root colonization and determines biofilm architecture. *Mol Microbiol* **77**: 549-561.

Massalha, H., Korenblum, E., Malitsky, S., Shapiro, O.H., and Aharoni, A. (2017) Live imaging of root-bacteria interactions in a microfluidics setup. *Proc Natl Acad Sci U S A* **114**: 4549-4554.

Matthysse, A.G. (1983) Role of bacterial cellulose fibrils in *Agrobacterium tumefaciens* infection. *J Bacteriol* **154**: 906-915.

Matthysse, A.G. (2014) Attachment of *Agrobacterium* to plant surfaces. *Front Plant Sci* **5**: 252-252.

Mauchline, T.H., Fowler, J.E., East, A.K., Sartor, A.L., Zaheer, R., Hosie, A.H.F. et al. (2006) Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome. *Proc Natl Acad Sci U S A* **103**: 17933-17938.

McAnulla, C., Edwards, A., Sanchez-Contreras, M., Sawers, R.G., and Downie, J.A. (2007) Quorum-sensing-regulated transcriptional initiation of plasmid transfer and replication genes in *Rhizobium leguminosarum* biovar *viciae*. *Microbiology* **153**: 2074-2082.

Meade, J., Higgins, P., and O'Gara, F. (1985) Studies on the inoculation and competitiveness of a *Rhizobium leguminosarum* strain in soils containing indigenous rhizobia. *Appl Environ Microbiol* **49**: 899-903.

Mee, M.T., Collins, J.J., Church, G.M., and Wang, H.H. (2014) Syntrophic exchange in synthetic microbial communities. *Proc Natl Acad Sci U S A* **111**: E2149-E2156.

Mikuláss, K.R., Nagy, K., Bogos, B., Szegletes, Z., Kovács, E., Farkas, A. et al. (2016) Antimicrobial nodule-specific cysteine-rich peptides disturb the integrity of bacterial outer and inner membranes and cause loss of membrane potential. *Ann Clin Microbiol Antimicrob* **15**: 43-43.

Miller, J.H. (1972) *Experiments in molecular genetics*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

Miller, L.D., Yost, C.K., Hynes, M.F., and Alexandre, G. (2007) The major chemotaxis gene cluster of *Rhizobium leguminosarum* bv. *viciae* is essential for competitive nodulation. *Mol Microbiol* **63**: 348-362.

Minnemeyer, S.L., Lightfoot, R., and Matthysse, A.G. (1991) A semiquantitative bioassay for relative virulence of *Agrobacterium tumefaciens* strains on *Bryophyllum daigremontiana*. *J Bacteriol* **173**: 7723-7724.

Miravet-Verde, S., Burgos, R., Delgado, J., Lluch-Senar, M., and Serrano, L. (2020) FASTQINS and ANUBIS: two bioinformatic tools to explore facts and artifacts in transposon sequencing and essentiality studies. *Nucl Acids Res* **48**: e102-e102.

Mitra, S., Mukherjee, A., Wiley-Kalil, A., Das, S., Owen, H., Reddy, P.M. et al. (2016) A rhamnose-deficient lipopolysaccharide mutant of *Rhizobium* sp. IRBG74 is defective in root colonization and beneficial interactions with its flooding-tolerant hosts *Sesbania cannabina* and wetland rice. *J Exp Bot* **67**: 5869-5884.

Mongiardini, E.J., Ausmees, N., Pérez-Giménez, J., Julia Althabegoiti, M., Ignacio Quelas, J., López-García, S.L., and Lodeiro, A.R. (2008) The rhizobial adhesion protein RapA1 is involved in adsorption of rhizobia to plant roots but not in nodulation. *FEMS Microbiol Ecol* **65**: 279-288.

Monteiro, R.A., Balsanelli, E., Tuleski, T., Faoro, H., Cruz, L.M., Wassem, R. et al. (2012) Genomic comparison of the endophyte *Herbaspirillum seropedicae* SmR1 and the phytopathogen *Herbaspirillum rubrisubalbicans*M1 by suppressive subtractive hybridization and partial genome sequencing. *FEMS Microbiol Ecol* **80**: 441-451.

Moris, M., Braeken, K., Schoeters, E., Verreth, C., Beullens, S., Vanderleyden, J., and Michiels, J. (2005) Effective symbiosis between *Rhizobium etli* and *Phaseolus vulgaris* requires the alarmone ppGpp. *J Bacteriol* **187**: 5460-5469.

Muñoz-López, M., and García-Pérez, J.L. (2010) DNA transposons: nature and applications in genomics. *Curr Genomics* **11**: 115-128.

Munson, G.P., Lam, D.L., Outten, F.W., and O'Halloran, T.V. (2000) Identification of a copper-responsive two-component system on the chromosome of *Escherichia coli* K-12. *J Bacteriol* **182**: 5864-5871.

Murray, J.D. (2011) Invasion by invitation: rhizobial infection in legumes. *Mol Plant-Microbe Interact* **24**: 631-639.

Muszynski, A., Laus, M., Kijne, J.W., and Carlson, R.W. (2011) Structures of the lipopolysaccharides from *Rhizobium leguminosarum* RBL5523 and its UDP-glucose dehydrogenase mutant (*exo5*). *Glycobiology* **21**: 55-68.

Mutalik, V.K., Novichkov, P.S., Price, M.N., Owens, T.K., Callaghan, M., Carim, S. et al. (2019) Dual-barcoded shotgun expression library sequencing for high-throughput characterization of functional traits in bacteria. *Nat Commun* **10**: 308.

Mutch, L.A., and Young, J.P.W. (2004) Diversity and specificity of *Rhizobium leguminosarum* biovar *viciae* on wild and cultivated legumes. *Mol Ecol* **13**: 2435-2444.

Nanda, A.K., Andrio, E., Marino, D., Pauly, N., and Dunand, C. (2010) Reactive oxygen species during plant-microorganism early interactions. *J Integr Plant Biol* **52**: 195-204.

Naseby, D.C., Way, J.A., Bainton, N.J., and Lynch, J.M. (2001) Biocontrol of *Pythium* in the pea rhizosphere by antifungal metabolite producing and non-producing *Pseudomonas* strains. *J Appl Microbiol* **90**: 421-429.

Niu, B., Paulson, J.N., Zheng, X., and Kolter, R. (2017) Simplified and representative bacterial community of maize roots. *Proc Natl Acad Sci U S A* **114**: E2450-E2459.

Noirot-Gros, M.-F., Shinde, S., Larsen, P.E., Zerbs, S., Korajczyk, P.J., Kemner, K.M., and Noirot, P.H. (2018) Dynamics of aspen roots colonization by *Pseudomonads* reveals strain-specific and mycorrhizal-specific patterns of biofilm formation. *Front Microbiol* **9**: 853.

Noirot-Gros, M.-F., Shinde, S.V., Akins, C., Johnson, J.L., Zerbs, S., Wilton, R. et al. (2020) Functional imaging of microbial interactions with tree roots using a microfluidics setup. *Front Plant Sci* **11**: 408.

Novichkov, P.S., Kazakov, A.E., Ravcheev, D.A., Leyn, S.A., Kovaleva, G.Y., Sutormin, R.A. et al. (2013) RegPrecise 3.0--a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics* **14**: 745.

O'Connor, J.R., Kuwada, N.J., Huangyutitham, V., Wiggins, P.A., and Harwood, C.S. (2012) Surface sensing and lateral subcellular localization of WspA, the receptor in a chemosensory-like system leading to c-di-GMP production. *Mol Microbiol* **86**: 720-729.

O'Toole, G.A., and Kolter, R. (1998) Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. *Mol Microbiol* **30**: 295-304.

Olanrewaju, O.S., Ayangbenro, A.S., Glick, B.R., and Babalola, O.O. (2019) Plant health: feedback effect of root exudates-rhizobiome interactions. *Appl Microbiol Biotechnol* **103**: 1155-1166.

Oldroyd, G.E., and Downie, J.A. (2008) Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu Rev Plant Biol* **59**: 519-546.

Oldroyd, G.E., Murray, J.D., Poole, P.S., and Downie, J.A. (2011) The rules of engagement in the legume-rhizobial symbiosis. *Annu Rev Genet* **45**: 119-144.

Oldroyd, G.E.D. (2013) Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat Rev Microbiol* **11**: 252-263.

Oresnik, I.J., Pacarynuk, L.A., O'Brien, S.A.P., Yost, C.K., and Hynes, M.F. (1998) Plasmid-encoded catabolic genes in *Rhizobium leguminosarum* bv. *trifolii*: evidence for a plant-inducible rhamnose locus involved in competition for nodulation. *Mol Plant-Microbe Interact* **11**: 1175-1185.

Ormeño-Orrillo, E., Rosenblueth, M., Luyten, E., Vanderleyden, J., and Martínez-Romero, E. (2008) Mutations in lipopolysaccharide biosynthetic genes impair maize rhizosphere and root colonization of *Rhizobium tropici* CIAT899. *Environ Microbiol* **10**: 1271-1284.

Parsons, J. (2019) Characterising root attachment in *Rhizobium*-legume symbioses. DPhil Thesis, University of Oxford.

Patriarca, E.J., Chiurazzi, M., Manco, G., Riccio, A., Lamberti, A., De Paolis, A. et al. (1992) Activation of the *Rhizobium leguminosarum glnII* gene by NtrC is dependent on upstream DNA sequences. *Mol Gen Genet* **234**: 337-345.

Pedersen, B.P., Stokes, D.L., and Apell, H.-J. (2019) The KdpFABC complex – K+ transport against all odds. *Mol Membr Biol* **35**: 21-38.

Peiffer, J.A., Spor, A., Koren, O., Jin, Z., Tringe, S.G., Dangl, J.L. et al. (2013) Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proc Natl Acad Sci U S A* **110**: 6548-6553.

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

Perry, B.J., Akter, M.S., and Yost, C.K. (2016) The Use of Transposon Insertion Sequencing to Interrogate the Core Functional Genome of the Legume Symbiont Rhizobium leguminosarum. *Front Microbiol* **7**: 1873-1873.

Philippot, L., Raaijmakers, J.M., Lemanceau, P., and van der Putten, W.H. (2013) Going back to the roots: the microbial ecology of the rhizosphere. *Nat Rev Microbiol* **11**: 789-799.

Pierson, L.S., 3rd, and Pierson, E.A. (2010) Metabolism and function of phenazines in bacteria: impacts on the behavior of bacteria in the environment and biotechnological processes. *Appl Microbiol Biotechnol* **86**: 1659-1670.

Pini, F., East, A.K., Appia-Ayme, C., Tomek, J., Karunakaran, R., Mendoza-Suarez, M. et al. (2017) Bacterial biosensors for in vivo spatiotemporal mapping of root secretion. *Plant Physiol* **174**: 1289-1306.

Pittard, A.J., and Davidson, B.E. (1991) TyrR protein of *Escherichia coli* and its role as repressor and activator. *Mol Microbiol* **5**: 1585-1592.

Planqué, K., and Kijne, J.W. (1977) Binding of pea lectins to a glycan type polysaccharide in the cell walls of *Rhizobium leguminosarum*. *FEBS Lett* **73**: 64-66.

Poole, P., Ramachandran, V., and Terpolilli, J. (2018) Rhizobia: from saprophytes to endosymbionts. *Nat Rev Microbiol* **16**: 291-303.

Poole, P.S., Schofiel, N.A., Reid, C.J., Drew, E.M., and Walshaw, D.L. (1994) Identification of chromosomal genes located downstream of *dctD* that affect the requirement for calcium and the lipopolysaccharide layer of *Rhizobium leguminosarum*. *Microbiology* **140**: 2797–2809.

Prell, J., Boesten, B., Poole, P., and Priefer, U.B. (2002) The *Rhizobium leguminosarum* bv. *viciae* VF39 γ-aminobutyrate (GABA) aminotransferase gene (*gabT*) is induced by GABA and highly expressed in bacteroids. *Microbiology* **148**: 615-623.

Preston, G.M., Bertrand, N., and Rainey, P.B. (2001) Type III secretion in plant growth-promoting *Pseudomonas fluorescens* SBW25. *Mol Microbiol* **41**: 999-1014.

Price, M.N., Deutschbauer, A.M., Skerker, J.M., Wetmore, K.M., Ruths, T., Mar, J.S. et al. (2013) Indirect and suboptimal control of gene expression is widespread in bacteria. *Molecular Systems Biology* **9**: 660.

Price, M.N., Wetmore, K.M., Waters, R.J., Callaghan, M., Ray, J., Liu, H. et al. (2018) Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature* **557**: 503-509.

Rainey, P.B. (1999) Adaptation of *Pseudomonas fluorescens* to the plant rhizosphere. *Environ Microbiol* **1**: 243-257.

Rainey, P.B., and Bailey, M.J. (1996) Physical and genetic map of the *Pseudomonas fluorescens* SBW25 chromosome. *Mol Microbiol* **19**: 521-533.

Ramachandran, V.K., East, A.K., Karunakaran, R., Downie, J.A., and Poole, P.S. (2011) Adaptation of *Rhizobium leguminosarum* to pea, alfalfa and sugar beet rhizospheres investigated by comparative transcriptomics. *Genome Biol* **12**: 1-12.

Ramey, B.E., Koutsoudis, M., Bodman, S.B.v., and Fuqua, C. (2004) Biofilm formation in plant–microbe associations. *Curr Opin Microbiol* **7**: 602-609.

Raun, W.R., and Johnson, G.V. (1999) Improving nitrogen use efficiency for cereal production. *Agronomy Journal* **91**: 357-363.

Recourt, K., van Brussel, A.A., Driessen, A.J., and Lugtenberg, B.J. (1989) Accumulation of a nod gene inducer, the flavonoid naringenin, in the cytoplasmic membrane of *Rhizobium leguminosarum* biovar *viciae* is caused by the pH-dependent hydrophobicity of naringenin. *J Bacteriol* **171**: 4370-4377.

Reinhold-Hurek, B., Bünger, W., Burbano, C.S., Sabale, M., and Hurek, T. (2015) Roots shaping their microbiome: global hotspots for microbial activity. *Annu Rev Phytopathol* **53**: 403-424.

Rensing, C., and Grass, G. (2003) *Escherichia coli* mechanisms of copper homeostasis in a changing environment. *FEMS Microbiol Rev* **27**: 197-213.

Richardson, J.S., Hynes, M.F., and Oresnik, I.J. (2004) A genetic locus necessary for rhamnose uptake and catabolism in *Rhizobium leguminosarum* bv. *trifolii*. *J Bacteriol* **186**: 8433-8442.

Ritz, D., Thöny-Meyer, L., and Hennecke, H. (1995) The *cycHJKL* gene cluster plays an essential role in the biogenesis of *c*-type cytochromes in *Bradyrhizobium japonicum*. *Mol Gen Genet* **247**: 27-38.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F.S., Lambin, E.F. et al. (2009) A safe operating space for humanity. *Nature* **461**: 472-475.

Rodelas, B., Lithgow, J.K., Wisniewski-Dye, F., Hardman, A., Wilkinson, A., Economou, A. et al. (1999) Analysis of quorum-sensing-dependent control of rhizosphere-expressed (*rhi*) genes in *Rhizobium leguminosarum* bv. *viciae*. *J Bacteriol* **181**: 3816-3823.

Rodríguez-Navarro, D.N., Dardanelli, M.S., and Ruíz-Saínz, J.E. (2007) Attachment of bacteria to the roots of higher plants. *FEMS Microbiol Lett* **272**: 127-136.

Rosenberg, M., and Kjelleberg, S. (1986) Hydrophobic interactions: role in bacterial adhesion. In *Advances in Microbial Ecology*. Marshall, K.C. (ed). Boston, MA: Springer US, pp. 353-393.

Rossi, F.A., Medeot, D.B., Liaudat, J.P., Pistorio, M., and Jofré, E. (2016) In *Azospirillum brasilense*, mutations in *flmA* or f*lmB* genes affect polar flagellum assembly, surface polysaccharides, and attachment to maize roots. *Microbiol Res* **190**: 55-62.

Russo, D.M., Williams, A., Edwards, A., Posadas, D.M., Finnie, C., Dankert, M. et al. (2006) Proteins exported via the PrsD-PrsE type I secretion system and the acidic exopolysaccharide are involved in biofilm formation by *Rhizobium leguminosarum*. *J Bacteriol* **188**: 4474-4486.

Sampedro, I., Parales, R.E., Krell, T., and Hill, J.E. (2014) *Pseudomonas* chemotaxis. *FEMS Microbiol Rev* **39**: 17-46.

Sánchez-Cañizares, C., Jorrín, B., Durán, D., Nadendla, S., Albareda, M., Rubio-Sanz, L. et al. (2018) Genomic diversity in the endosymbiotic bacterium *Rhizobium leguminosarum*. *Genes* **9**: 60.

Santos-Beneit, F. (2015) The Pho regulon: a huge regulatory network in bacteria. *Front Microbiol* **6**: 402.

Santoyo, G., Hernández-Pacheco, C., Hernandez-Salmeron, J., and Hernández-León, R. (2017) The role of abiotic factors modulating the plant-microbe-soil interactions: Toward sustainable agriculture. A review. *Span J Agric Res* **15**: e03R01.

Sasse, J., Martinoia, E., and Northen, T. (2018) Feed your friends: do plant exudates shape the root microbiome? *Trends Plant Sci* **23**: 25-41.

Sasse, J., Kosina, S.M., de Raad, M., Jordan, J.S., Whiting, K., Zhalnina, K., and Northen, T.R. (2020) Root morphology and exudate availability are shaped by particle size and chemistry in *Brachypodium distachyon*. *Plant Direct* **4**: e00207.

Schlaeppi, K., Dombrowski, N., Oter, R.G., Ver Loren Van Themaat, E., and Schulze-Lefert, P. (2014) Quantitative divergence of the bacterial root microbiota in *Arabidopsis thaliana* relatives. *Proc Natl Acad Sci U S A* **111**: 585-592.

Schlechter, R.O., Jun, H., Bernach, M., Oso, S., Boyd, E., Muñoz-Lintz, D.A. et al. (2018) Chromatic bacteria – a broad host-range plasmid and chromosomal insertion toolbox for fluorescent protein expression in bacteria. *Front Microbiol* **9**: 3052.

Schnider-Keel, U., Lejbølle, K.B., Baehler, E., Haas, D., and Keel, C. (2001) The sigma factor AlgU (AlgT) controls exopolysaccharide production and tolerance towards desiccation and osmotic stress in the biocontrol agent *Pseudomonas fluorescens* CHA0. *Appl Environ Microbiol* **67**: 5683-5693.

Schoenlein, P.V., and Ely, B. (1989) Characterization of strains containing mutations in the contiguous *flaF*, *flbT,* or *flbA-flaG* transcription unit and identification of a novel fla phenotype in *Caulobacter crescentus*. *J Bacteriol* **171**: 1554-1561.

Schweizer, H.P., Po, C., and Bacic, M.K. (1995) Identification of *Pseudomonas aeruginosa glpM*, whose gene product is required for efficient alginate biosynthesis from various carbon sources. *J Bacteriol* **177**: 4801-4804.

Silby, M.W., Cerdeño-Tárraga, A.M., Vernikos, G.S., Giddens, S.R., Jackson, R.W., Preston, G.M. et al. (2009) Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*. *Genome Biol* **10**: R51.

Silverman, Julie M., Agnello, Danielle M., Zheng, H., Andrews, Benjamin T., Li, M., Catalano, Carlos E. et al. (2013) Haemolysin coregulated protein is an exported receptor and chaperone of Type VI secretion substrates. *Molecular Cell* **51**: 584-593.

Simons, M., Permentier, H.P., de Weger, L.A., Wijffelman, C.A., and Lugtenberg, B.J.J. (1997) Amino acid synthesis Is necessary for tomato root colonization by *Pseudomonas fluorescens* strain WCS365. *Mol Plant-Microbe Interact* **10**: 102-106.

Simons, M., van der Bij, A.J., Brand, I., de Weger, L.A., Wijffelman, C.A., and Lugtenberg, B.J. (1996) Gnotobiotic system for studying rhizosphere colonization by plant growth-promoting *Pseudomonas* bacteria. *Mol Plant Microbe Interact* **9**: 600-607.

Smit, G., Kijne, J.W., and Lugtenberg, B.J. (1987) Involvement of both cellulose fibrils and a Ca2+-dependent adhesin in the attachment of *Rhizobium leguminosarum* to pea root hair tips. *J Bacteriol* **169**: 4294-4301.

Smit, G., Tubbing, D.M.J., Kijne, J.W., and Lugtenberg, B.J.J. (1991) Role of Ca2+ in the activity of rhicadhesin from *Rhizobium leguminosarum* biovar *viciae*, which mediates the first step in attachment of *Rhizobiaceae* cells to plant root hair tips. *Arch Microbiol* **155**: 278-283.

Smit, G., Logman, T.J., Boerrigter, M.E., Kijne, J.W., and Lugtenberg, B.J. (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. *J Bacteriol* **171**: 4054.

Sobrero, P.M., and Valverde, C. (2020) Comparative genomics and evolutionary analysis of RNA-binding proteins of the CsrA family in the genus *Pseudomonas*. *Front Mol Biosci* **7**: 127-127.

Sohlenkamp, C., Galindo-Lagunas, K.A., Guan, Z., Vinuesa, P., Robinson, S., Thomas-Oates, J. et al. (2007) The lipid lysyl-phosphatidylglycerol is present in membranes of *Rhizobium tropici* CIAT899 and confers increased resistance to polymyxin B under acidic growth conditions. *Mol Plant Microbe Interact* **20**: 1421-1430.

Song, C., Jin, K., and Raaijmakers, J.M. (2021) Designing a home for beneficial plant microbiomes. *Curr Opin Plant Biol* **62**: 102025.

Song, C., van der Voort, M., van de Mortel, J., Hassan, K.A., Elbourne, L.D.H., Paulsen, I.T. et al. (2015) The Rsm regulon of plant growth-promoting *Pseudomonas fluorescens* SS101: role of small RNAs in regulation of lipopeptide biosynthesis. *Microbial Biotechnology* **8**: 296-310.

Sourjik, V., Muschler, P., Scharf, B., and Schmitt, R. (2000) VisN and VisR are global regulators of chemotaxis, flagellar, and motility genes in *Sinorhizobium* (*Rhizobium*) *meliloti*. *J Bacteriol* **182**: 782-788.

Stanger, A. (2008) Analysis of the ECF sigma factors of *Rhizobium leguminosarum* biovar *viciae* 3841. PhD thesis, University of East Anglia.

Stincone, A., Prigione, A., Cramer, T., Wamelink, M.M.C., Campbell, K., Cheung, E. et al. (2015) The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. *Biol Rev Camb Philos Soc* **90**: 927-963.

Streit, W.R., Joseph, C.M., and Phillips, D.A. (1996) Biotin and other water-soluble vitamins are key growth factors for alfalfa root colonization by *Rhizobium meliloti* 1021. *Mol Plant Microbe Interact* **9**: 330-338.

Sturla, L., Bisso, A., Zanardi, D., Benatti, U., De Flora, A., and Tonetti, M. (1997) Expression, purification and characterization of GDP-d-mannose 4,6-dehydratase from *Escherichia coli*. *FEBS Letters* **412**: 126-130.

Swift, S., Downie, J.A., Whitehead, N.A., Barnard, A.M., Salmond, G.P., and Williams, P. (2001) Quorum sensing as a population-density-dependent determinant of bacterial physiology. *Adv Microb Physiol* **45**: 199-270.

Szurmant, H., and Ordal, G.W. (2004) Diversity in chemotaxis mechanisms among the bacteria and archaea. *Microbiol Mol Biol Rev* **68**: 301-319.

Tambalo, D.D., Del Bel, K.L., Bustard, D.E., Greenwood, P.R., Steedman, A.E., and Hynes, M.F. (2010) Regulation of flagellar, motility and chemotaxis genes in *Rhizobium leguminosarum* by the VisN/R-Rem cascade. *Microbiology* **156**: 1673-1685.

Tan, M.S.F., White, A.P., Rahman, S., and Dykes, G.A. (2016) Role of fimbriae, flagella and cellulose on the attachment of *Salmonella Typhimurium* ATCC 14028 to plant cell wall models. *PLOS One* **11**: e0158311-e0158311.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucl Acids Res* **28**: 33-36.

Tettmann, B., Dötsch, A., Armant, O., Fjell, C.D., and Overhage, J. (2014) Knockout of extracytoplasmic function sigma factor ECF-10 affects stress resistance and biofilm formation in *Pseudomonas putida* KT2440. *Appl Environ Microbiol* **80**: 4911-4919.

Thompson, I.P., Lilley, A.K., Ellis, R.J., Bramwell, P.A., and Bailey, M.J. (1995) Survival, colonization and dispersal of genetically modified *Pseudomonas fluorescens* SBW25 in the phytosphere of field grown sugar beet. *Bio/Technology* **13**: 1493-1497.

Thompson, M.A., Onyeziri, M.C., and Fuqua, C. (2018) Function and regulation of *Agrobacterium tumefaciens* cell surface structures that promote attachment. *Curr Top Microbiol Immunol* **418**: 143-184.

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

Tkacz, A., Bestion, E., Bo, Z., Hortala, M., and Poole, P.S. (2020) Influence of plant fraction, soil, and plant species on microbiota: a multikingdom comparison. *mBio* **11**.

Toledo-Arana, A., Valle, J., Solano, C., Arrizubieta, M.J., Cucarella, C., Lamata, M. et al. (2001) The enterococcal surface protein, Esp, is involved in *Enterococcus faecalis* biofilm formation. *Appl Environ Microbiol* **67**: 4538-4545.

Tomich, M., Planet, P.J., and Figurski, D.H. (2007) The *tad* locus: postcards from the widespread colonization island. *Nat Rev Microbiol* **5**: 363-375.

Tomlinson, A.D., and Fuqua, C. (2009) Mechanisms and regulation of polar surface attachment in *Agrobacterium tumefaciens*. *Curr Opin Microbiol* **12**: 708-714.

Trivedi, P., Leach, J.E., Tringe, S.G., Sa, T., and Singh, B.K. (2020) Plant–microbiome interactions: from community assembly to plant health. *Nat Rev Microbiol* **18**: 607-621.

Turnbull, G.A., Morgan, J.A.W., Whipps, J.M., and Saunders, J.R. (2001) The role of bacterial motility in the survival and spread of *Pseudomonas fluorescens* in soil and in the attachment and colonisation of wheat roots. *FEMS Microbiol Ecol* **36**: 21-31.

Turner, T.R., James, E.K., and Poole, P.S. (2013a) The plant microbiome. *Genome Biol* **14**: 1-10.

Turner, T.R., Ramakrishnan, K., Walshaw, J., Heavens, D., Alston, M., Swarbreck, D. et al. (2013b) Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants. *ISME J* **7**: 2248-2258.

Tzin, V., and Galili, G. (2010) The biosynthetic pathways for shikimate and aromatic amino acids in *Arabidopsis thaliana*. *Arabidopsis Book* **8**: e0132-e0132.

Unge, A., and Jansson, J. (2001) Monitoring population size, activity, and distribution of *gfp*-l*uxAB*-tagged *Pseudomonas fluorescens* SBW25 during colonization of wheat. *Microb Ecol* **41**: 290-300.

Uroz, S., Buee, M., Murat, C., Frey-Klett, P., and Martin, F. (2010) Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. *Environ Microbiol Rep* **2**: 281-288.

Van Breemen, N. (2002) Nitrogen cycle: natural organic tendency. *Nature* **415**: 381-382.

van Loosdrecht, M.C., Lyklema, J., Norde, W., Schraa, G., and Zehnder, A.J. (1987) The role of bacterial cell wall hydrophobicity in adhesion. *Appl Environ Microbiol* **53**: 1893-1897.

van Opijnen, T., and Camilli, A. (2013) Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat Rev Microbiol* **11**: 435-442.

van Ulsen, P., Zinner, K.M., Jong, W.S.P., and Luirink, J. (2018) On display: autotransporter secretion and application. *FEMS Microbiol Lett* **365**: 10.1093/femsle/fny1165.

Vanderlinde, E.M., Hynes, M.F., and Yost, C.K. (2014) Homoserine catabolism by *Rhizobium leguminosarum* bv. *viciae* 3841 requires a plasmid-borne gene cluster that also affects competitiveness for nodulation. *Environ Microbiol* **16**: 205-217.

Vesper, S.J. (1987) Production of pili (fimbriae) by *Pseudomonas fluorescens* and correlation with attachment to corn roots. *Appl Environ Microbiol* **53**: 1397-1405.

Vesper, S.J., and Bauer, W.D. (1986) Role of pili (fimbriae) in attachment of *Bradyrhizobium japonicum* to soybean roots. *Appl Environ Microbiol* **52**: 134-141.

Vinuesa, P., Neumann-Silkow, F., Pacios-Bras, C., Spaink, H.P., Martínez-Romero, E., and Werner, D. (2003) Genetic analysis of a pH-regulated operon from *Rhizobium tropici* CIAT899 involved in acid tolerance and nodulation competitiveness. *Mol Plant-Microbe Interact* **16**: 159-168.

Visick, J.E., Ichikawa, J.K., and Clarke, S. (1998) Mutations in the *Escherichia coli surE* gene increase isoaspartyl accumulation in a strain lacking the *pcm* repair methyltransferase but suppress stress-survival phenotypes. *FEMS Microbiol Lett* **167**: 19-25.

Voges, M.J.E.E.E., Bai, Y., Schulze-Lefert, P., and Sattely, E.S. (2019) Plant-derived coumarins shape the composition of an *Arabidopsis* synthetic root microbiome. *Proc Natl Acad Sci U S A* **116**: 12558-12565.

Wadhams, G.H., and Armitage, J.P. (2004) Making sense of it all: bacterial chemotaxis. *Nat Rev Mol Cell Biol* **5**: 1024-1037.

Wais, R.J., Keating, D.H., and Long, S.R. (2002) Structure-function analysis of nod factor-induced root hair calcium spiking in *Rhizobium*-legume symbiosis. *Plant Physiol* **129**: 211-224.

Wang, D., Yang, S., Tang, F., and Zhu, H. (2012) Symbiosis specificity in the legume: rhizobial mutualism. *Cell Microbiol* **14**: 334-342.

Wang, Q., Liu, J., and Zhu, H. (2018) Genetic and molecular mechanisms underlying symbiotic specificity in Legume-*Rhizobium* Interactions. *Front Plant Sci* **9**: 313.

Ward, E., Renault, T.T., Kim, E.A., Erhardt, M., Hughes, K.T., and Blair, D.F. (2018) Type-III secretion pore formed by flagellar protein FliP. *Mol Microbiol* **107**: 94-103.

Warner, H.R., Duncan, B.K., Garrett, C., and Neuhard, J. (1981) Synthesis and metabolism of uracil-containing deoxyribonucleic acid in *Escherichia coli*. *J Bacteriol* **145**: 687-695.

Wei, Z., Yang, T., Friman, V.-P., Xu, Y., Shen, Q., and Jousset, A. (2015) Trophic network architecture of root-associated bacterial communities determines pathogen invasion and plant health. *Nat Commun* **6**: 8413.

Wetmore, K.M., Price, M.N., Waters, R.J., Lamson, J.S., He, J., Hoover, C.A. et al. (2015) Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *mBio* **6**: e00306-00315.

Wheatley, R.M., and Poole, P.S. (2018) Mechanisms of bacterial attachment to roots. *FEMS Microbiol Rev* **42**: 448-461.

Wheatley, R.M., Ramachandran, V.K., Geddes, B.A., Perry, B.J., Yost, C.K., and Poole, P.S. (2017) Role of O_2 in the growth of *Rhizobium leguminosarum* bv. *viciae* 3841 on glucose and succinate. *J Bacteriol* **199**: e00572-00516.

Wheatley, R.M., Ford, B.L., Li, L., Aroney, S.T.N., Knights, H.E., Ledermann, R. et al. (2020) Lifestyle adaptations of *Rhizobium* from rhizosphere to symbiosis. *Proc Natl Acad Sci U S A* **117**: 23823-23834.

Williams, A., Wilkinson, A., Krehenbrink, M., Russo, D.M., Zorreguieta, A., and Downie, J.A. (2008) Glucomannan-mediated attachment of *Rhizobium*

leguminosarum to pea root hairs is required for competitive nodule infection. *J Bacteriol* **190**: 4706-4715.

Willis, L.B., and Walker, G.C. (1999) A novel *Sinorhizobium meliloti* operon encodes an α-glucosidase and a periplasmic-binding-protein-dependent transport system for α-glucosides. *J Bacteriol* **181**: 4176-4184.

Wilton, R., Ahrendt, A.J., Shinde, S., Sholto-Douglas, D.J., Johnson, J.L., Brennan, M.B., and Kemner, K.M. (2018) A new suite of plasmid vectors for fluorescence-based imaging of root colonizing *Pseudomonads*. *Front Plant Sci* **8**: 2242.

Wisniewski-Dyé, F., and Downie, J.A. (2002) Quorum-sensing in *Rhizobium*. *Antonie van Leeuwenhoek* **81**: 397-407.

Xi, C., Schoeters, E., Vanderleyden, J., and Michiels, J. (2000) Symbiosis-specific expression of *Rhizobium etli casA* encoding a secreted calmodulin-related protein. *Proc Natl Acad Sci U S A* **97**: 11114-11119.

Xu, J., Kim, J., Danhorn, T., Merritt, P.M., and Fuqua, C. (2012) Phosphorus limitation increases attachment in *Agrobacterium tumefaciens* and reveals a conditional functional redundancy in adhesin biosynthesis. *Res Microbiol* **163**: 674-684.

Yang, G., Bhuvaneswari, T.V., Joseph, C.M., King, M.D., and Phillips, D.A. (2002) Roles for riboflavin in the *Sinorhizobium*-alfalfa association. *Mol Plant-Microbe Interact* **15**: 456-462.

Yang, W., Pollard, M., Li-Beisson, Y., Beisson, F., Feig, M., and Ohlrogge, J. (2010) A distinct type of glycerol-3-phosphate acyltransferase with *sn*-2 preference and phosphatase activity producing 2-monoacylglycerol. *Proc Natl Acad Sci U S A* **107**: 12040-12045.

Yeoman, K.H., Delgado, M.-J., Wexler, M., Downie, J.A., and Johnston, A.W.B. (1997) High affinity iron acquisition in *Rhizobium Leguminosarum* requires the *cycHJKL* operon and the *feuPQ* gene products, which belong to the family of two-component transcriptional regulators. *Microbiology* **143**: 127-134.

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

Young, J.P.W., Crossman, L.C., Johnston, A.W.B., Thomson, N.R., Ghazoui, Z.F., Hull, K.H. et al. (2006) The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol* **7**: R34.

Yousef, F., and Espinosa-Urgel, M. (2007) *In silico* analysis of large microbial surface proteins. *Res Microbiol* **158**: 545-550.

Yousef-Coronado, F., Travieso, M.L., and Espinosa-Urgel, M. (2008) Different, overlapping mechanisms for colonization of abiotic and plant surfaces by *Pseudomonas putida*. *FEMS Microbiol Lett* **288**: 118-124.

Yu, K., Pieterse, C.M.J., Bakker, P.A.H.M., and Berendsen, R.L. (2019) Beneficial microbes going underground of root immunity. *Plant Cell Environ* **42**: 2860-2870.

Yuan, J., Zhao, J., Wen, T., Zhao, M., Li, R., Goossens, P. et al. (2018) Root exudates drive the soil-borne legacy of aboveground pathogen infection. *Microbiome* **6**: 156.

Yurgel, S.N., Berrocal, J., Wilson, C., and Kahn, M.L. (2007) Pleiotropic effects of mutations that alter the *Sinorhizobium meliloti* cytochrome *c* respiratory system. *Microbiology* **153**: 399-410.

Zelezniak, A., Andrejev, S., Ponomarova, O., Mende, D.R., Bork, P., and Patil, K.R. (2015) Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proc Natl Acad Sci U S A* **112**: 6449-6454.

Zhang, R.G., Skarina, T., Katz, J.E., Beasley, S., Khachatryan, A., Vyas, S. et al. (2001) Structure of *Thermotoga maritima* stationary phase survival protein SurE: a novel acid phosphatase. *Structure* **9**: 1095-1106.

Zhang, Y.J., Ioerger, T.R., Huttenhower, C., Long, J.E., Sassetti, C.M., Sacchettini, J.C., and Rubin, E.J. (2012) Global assessment of genomic regions required for growth in *Mycobacterium tuberculosi*s. *PLOS Pathogens* **8**: e1002946.

Zhu, F., Wu, R., Zhang, H., and Wu, H. (2013) Structural and biochemical analysis of a bacterial glycosyltransferase. *Methods Mol Biol* **1022**: 29-39.

Zifcakova, L. (2020) Factors affecting soil microbial processes. In *Carbon and Nitrogen Cycling in Soil*. Datta, R., Meena, R.S., Pathan, S.I., and Ceccherini, M.T. (eds). Singapore: Springer Singapore, pp. 439-461.

Appendix A. Supplementary material for Chapter 3

The following supplementary tables for chapter can be found in the excel file entitled 'Chapter 3 Supplementary Tables':

Table number	Description
Table S3.1	6,703 protein coding genes containing insertions within the central 10-90% region.
Table S3.2	636 unhit protein coding genes lacking insertion mutants in the Rlv3841 barcoded transposon pool.
Table S3.3	189 Rlv3841 protein coding genes less than 300 nucleotides long lacking insertions.
Table S3.4	Gene fitness values for 6,019 Rlv3841 genes during growth in the rhizosphere and colonisation of pea.
Table S3.5	667 Rlv3841 genes with fitness values less than -1 for growth in the rhizosphere of pea.
Table S3.6	494 Rlv3841 genes with fitness values less than -1 for pea root colonisation.
Table S3.7	Table S3.7. Gene fitness values for 170 genes previously implicated in an INSeq screen as contributing to Rlv3841 growth in the pea rhizosphere (Wheatley et al., 2020).
Table S3.8	Table S3.8. Gene fitness values for 170 genes previously implicated in an INSeq screen as contributing to Rlv3841 colonisation of pea roots (Wheatley et al., 2020).
Table S3.9	Venn diagram results for the 667 genes with fitness values less than -1 for growth in the pea rhizosphere and 494 genes with fitness values less than -1 for pea root colonisation. 436 genes are rhizosphere specific, 231 genes are rhizosphere-progressive and 263 genes are colonisation specific.
Table S3.10	Rlv3841 gene fitness values for genes involved in chemotaxis and flagella mediated motility during growth in the pea rhizosphere and colonisation of pea roots.
Table S3.11	271 genes that provided a growth advantaged in the pea rhizosphere when mutated.
Table S3.12	260 genes that result in increased colonisation of pea roots when mutated.
Table S3.13	Venn diagram comparison of 271 genes with growth advantaged phenotypes for growth in the pea rhizosphere and 260 genes advantaged for pea root colonisation. 202 genes were growth advantaged specifically for rhizosphere growth, 69 genes were rhizosphere-progressive and 191 genes were colonisation specific.

Appendix B. Supplementary material for Chapter 4

The following supplementary tables for chapter can be found in the excel file entitled 'Chapter 4 Supplementary Tables':

Table number	Description
Table S4.1	Gene fitness values for 6,019 Rlv3841 genes during growth in the rhizosphere and colonisation of pea, lentil, *Lathyrus*, alfalfa and barley.
Table S4.2	802 Rlv3841 genes required for lentil rhizosphere growth and root colonisation. 184 genes are rhizosphere specific, 132 genes rhizosphere progressive and 486 genes colonisation specific.
Table S4.3	Rlv3841 genes required for *Lathyrus* rhizosphere growth and root colonisation. 162 genes rhizosphere specific, 137 genes rhizosphere progressive and 522 genes colonisation specific.
Table S4.4	Rlv3841 genes required for alfalfa rhizosphere growth and root colonisation. 422 genes rhizosphere specific, 245 genes rhizosphere progressive and 444 genes colonisation specific.
Table S4.5	Rlv3841 genes required for rhizosphere growth and colonisation of barley. 416 genes rhizosphere specific, 219 genes rhizosphere progressive and 305 genes colonisation specific.
Table S4.6	Venn diagram comparison of 73 genes required growth in the rhizosphere and 90 genes required for colonisation of pea, lentil, *Lathyrus*, alfalfa and barley.
Table S4.7	16 genes upregulated in Rlv3841 *asfR* mutant background.
Table S4.8	74 Rlv3841 genes with pea-specific phenotypes for rhizosphere growth and/or root colonisation
Table S4.9	66 Rlv3841 genes with lentil-specific phenotypes for rhizosphere growth and/or root colonisation
Table S4.10	88 Rlv3841 genes with *Lathyrus*-specific phenotypes for rhizosphere growth and/or root colonisation
Table S4.11	133 Rlv3841 genes with alfalfa-specific phenotypes for rhizosphere growth and/or root colonisation
Table S4.12	8 Rlv3841 genes with barley-specific phenotypes for rhizosphere growth and/or root colonisation
Table S4.13	Genes with transport related functions displaying a plant-specific phenotype for rhizosphere growth.

Appendix C. Supplementary material for Chapter 5

The following supplementary tables for chapter can be found in the excel file entitled 'Chapter 5 Supplementary Tables':

Table number	Description
Table S5.1	5,368 genes containing at least one mapped barcoded transposon mutant within the central 10-90% region.
Table S5.2	477 of the 557 genes lacking barcoded insertion mutants within their central 10-90% region. Genes less than 300 nucleotides long are not shown.
Table S5.3	Gene fitness values for 4,720 SBW25 genes during growth in the rhizosphere and colonisation of pea and barley, during single and co-inoculation with Rlv3841.
Table S5.4	Gene fitness values for 5,770 Rlv3841 genes for colonisation of barley during single and co-inoculation with SBW25.
Table S5.5	Venn diagram results for comparison of 435 genes contributing to Rlv3841 colonisation of barley under single inoculation to 462 genes contributing to barley colonisation during competition with SBW25. 92 genes were commonly required. All genes have fitness values less than -1.
Table S5.6	50 genes contributing to Rlv3841 colonisation of barley specifically during competition with SBW25 after filtering.
Table S5.7	SBW25 genes required for pea rhizosphere growth and root colonisation when inoculated alone. 215 genes required for rhizosphere growth, 256 genes rhizosphere progressive and 371 genes required for root colonisation.
Table S5.8	SBW25 genes required for pea rhizosphere growth and root colonisation when co-inoculated with SBW25. 354 genes required for rhizosphere growth, 225 genes rhizosphere progressive and 475 genes required for root colonisation.
Table S5.9	SBW25 genes required for barley rhizosphere growth and root colonisation when inoculated alone. 252 genes required for rhizosphere growth, 341 genes rhizosphere progressive and 228 genes required for root colonisation.
Table S5.10	SBW25 genes required for barley rhizosphere growth and root colonisation when co-inoculated with SBW25. 239 genes required for rhizosphere growth, 165 genes rhizosphere progressive and 326 genes required for root colonisation.
Table S5.11	Venn diagram results from comparison of 58 genes required growth of SBW25 in the pea and barley rhizosphere during SI and CI with Rlv3841 and the 25 genes required for root colonisation.
Table S5.12	SBW25 genes required for growth in the rhizosphere or colonisation of pea and barley during competition with Rlv3841. 5

	genes required for rhizosphere growth and 5 genes contributing to root colonisation.
Table S5.13	SBW25 genes required for growth in the rhizosphere or colonisation of pea during competition with Rlv3841. 71 genes required for rhizosphere 7 genes rhizosphere progressive and 88 genes contributing to root colonisation.
Table S5.15	SBW25 genes required for growth in the rhizosphere or colonisation of barley during competition with Rlv3841. 56 genes required for rhizosphere growth, 3 genes rhizosphere progressive and 51 genes contributing to root colonisation.