Understanding unipolar growth within the rod-shaped bacterium *Rhizobium leguminosarum*

Hannah Jane Wright

Doctor of Philosophy

University of East Anglia

October 2023

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there-from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract

Unipolar growth within the alphaproteobacterial family *Rhizobiaceae* is believed to be coordinated by a largely unknown suite of proteins distinct from those required for dispersed cell elongation (a growth mechanism widely observed amongst other Gram-negative rods). Recent studies within *Agrobacterium* and *Sinorhizobium* have identified several growth-associated polar proteins, yet the mechanisms underpinning unipolar growth within other *Rhizobiaceae* genera remain poorly understood. In this thesis, *Rhizobium leguminosarum* homologs of the chromosome partitioning ATPase ParA1 and pole-organising protein PopZ, and a novel *Rhizobium* cytoskeletal element (Rcy), were identified as candidate growth-associated polar proteins. Their involvement during unipolar growth was then investigated by studying the impact of their deletion and overexpression on cell growth and morphology, and by observing the localisation of ParA1-, PopZ-, and Rcy-EGFP fusion proteins.

The findings of this work support the classification of PopZ as a polarity protein, proposed here to possibly localise to the *R. leguminosarum* growth pole. Both PopZ and ParA1 help to maintain rod morphology in *R. leguminosarum*, possibly by facilitating spatial coordination of cell growth and division. Overexpression of PopZ was associated with minicell formation, while ParA1 overexpression was correlated with asymmetric cell division and occasional cellular branching. Within this thesis, the candidate cytoskeletal element Rcy was characterised as a unipolar-localised protein, hypothesised to determine the site of cell growth at the growth pole. Overexpression of Rcy within *R. leguminosarum* resulted in the formation of enlarged spherical and irregularly shaped cells containing multiple Rcy foci around the perimeter of the cell. Furthermore, difficulties recovering *rcy* knockout mutants suggest that this protein may be essential for cell viability. Overall, this work highlights the involvement of ParA1, PopZ and Rcy in the coordination of *Rhizobium* growth and division, shedding light on the proteins which may regulate unipolar growth and the maintenance of rod morphology within the *Rhizobiaceae*.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate ‘take down’ action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Table of Contents

Abstract **2**
Table of Contents **3**
List of Figures **7**
List of Tables **10**
Glossary of Abbreviations **11**
Acknowledgements **13**
Chapter One – Introduction **14**
1.1 Rhizobia, *Rhizobium* and the *Rhizobiaceae* 14
1.1.1 The rhizobium-legume symbioses 15
1.1.2 Nodule organogenesis and bacteroid differentiation 18
1.2 Coordinating growth, division, and morphology: The bacterial cytoskeleton 22
1.2.1 Tubulin homologs 23
1.2.2 Actin homologs 27
1.2.3 Intermediate filament-like proteins 31
1.2.4 The ParA/MinD family - Walker A cytoskeletal ATPases 33
1.3 Growth mechanisms in rod-shaped bacteria 37
1.3.1 Lateral (Dispersed) elongation 37
1.3.2 Polar elongation 41
1.4 Polarity markers and pole organisers 47
1.4.1 The *Streptomyces* TIPOC 48
1.4.2 Determining polarity in *Caulobacter* 49
1.4.3 Organising the *Agrobacterium* growth pole 53
1.5 Understanding unipolar growth in *Rhizobium leguminosarum* 58
1.5.1 *R. leguminosarum* PopZ 59
1.5.2 A novel *Rhizobium* cytoskeletal element? 60
1.5.3 *R. leguminosarum* ParA1 62
1.6 Experimental Aims 63
Chapter Two – Materials and Methods **65**
2.1 Bacterial strains, media and growth conditions 65
2.1.1 Bacterial strains 65
2.1.2 Growth medium and conditions 66
2.1.3 *R. leguminosarum* growth curve generation 67
2.2 Plasmids & antibiotics 67
2.2.1 Plasmids 67

2.2.2 Antibiotics, additives, and inducers 69
2.3 DNA preparation and handling 70
2.3.1 Isolation of DNA from bacteria 70
2.3.2 Agarose gel electrophoresis 71
2.3.3 Purification of DNA from agarose gel 72
2.3.4 Restriction enzyme digestion 72
2.3.5 DNA ligation 73
2.3.6 Sanger sequencing of DNA 74
2.3.7 Artificial Gene Synthesis 74
2.4 Bacterial Transformation and Conjugation 76
2.4.1 Transformation of *E. coli* 76
2.4.2 Conjugation of plasmids into *Rhizobium leguminosarum* 78
2.5 Polymerase Chain Reaction 79
2.5.1 Primer design 79
2.5.2 High-fidelity amplification of DNA 81
2.5.3 Diagnostic colony PCR (GoTaq) 83
2.5.4 Purification of PCR products 85
2.6 Synthesis of plasmids using traditional cloning 85
2.6.1 Generation and blue-white screening of pK18*mobsacB*-derived allelic exchange vectors 86
2.6.2 Introduction of a tetracycline resistance gene into pK18*mobsacB*-derived allelic exchange vectors 86
2.6.3 Cloning of pIND6-derived expression plasmids 87
2.7 EGFP-tagging using phage lambda (λ) Red-mediated homologous recombination 87
2.8 Homologous recombination-based gene deletion in *R. leguminosarum* 89
2.8.1 Integration of allelic exchange vectors into the *R. leguminosarum* genome 89
2.8.2 Generation of gene knockout mutants 90
2.9 Polyacrylamide gel electrophoresis 91
2.9.1 Sample preparation 91
2.9.2 Buffer Composition 91
2.9.3 Separation of proteins using SDS-PAGE 92
2.9.4 Anti-His western blot analysis 93
2.10 Microscopic analysis of samples 94
2.10.1 Slide preparation 94
2.10.2 Microscopy 94
2.10.3 Image analysis 95

2.11 Statistical analysis and presentation of cell morphology data 95

Chapter Three – Characterisation of ParA1, PopZ and Rcy by whole-gene deletion..... 96

3.1 Introduction 96

3.2 Attempted generation of unmarked knockout mutants 96

3.2.1 Design and synthesis of unmarked knockout vectors 100

3.2.2 Screening for primary integration events 107

3.2.3 Screening for knockout mutants 114

3.3 Generation of marked knockout mutants 118

3.3.1 Cloning of tetracycline-resistant marked knockout vectors 118

3.3.2 Screening for tetracycline-resistant knockout mutants 122

3.4 Discussion 127

3.4.1 Unsuccessful deletion of *parA1* may be attributed in part to its genomic context 127

3.4.2 Deletion of *rcy* in *R. leguminosarum* may reduce cell viability 129

3.4.3 Potential off-target integration of the pK18-*popZ*-tet^{R}-KO allelic exchange vector 130

3.4.4 Other complications: antibiotic resistance, reduced sucrose-sensitivity, and mixed colonies 131

3.4.5 Alternative gene deletion approaches 133

Chapter Four – Characterisation of ParA1, PopZ and Rcy by protein overexpression..... 136

4.1 Introduction 136

4.2 Generation of *R. leguminosarum* gene overexpression strains 136

4.2.1 The IPTG-inducible expression vector pIND6 136

4.2.2 Generation of pIND6-derived expression vectors 139

4.3 Impact of *parA1*, *popZ* and *rcy* overexpression on *R. leguminosarum* growth 145

4.3.1 Establishment of experimental parameters 145

4.3.2 Impact of gene overexpression 148

4.4 Confirmation of protein expression 156

4.5 Impact of ParA1, PopZ and Rcy overexpression on *R. leguminosarum* cell morphology 161

4.5.1 Morphological phenotypes following protein overexpression 161

4.5.2 Quantitative analysis of overexpression phenotypes 164

4.6 Discussion 173

4.6.1 Severe lag phase growth defects suggest a role of these proteins in preparing for growth 173

4.6.2 The Rcy overexpression growth curve seemingly did not reflect population changes 175

4.6.3 Impacts of ParA1, PopZ and Rcy overexpression support role of these proteins in controlling growth and morphology 176

Chapter Five – Localisation of EGFP fusion proteins 179

5.1 Introduction 179

5.2 EGFP-tagging of ParA1, PopZ and Rcy proteins 180

5.2.1 Bacteriophage lambda (λ) Red recombination-mediated genetic engineering 180

5.2.2 Generation of the ParA1-, PopZ-, and Rcy-EGFP expression plasmids 184

5.3 Localisation of overexpressed EGFP fusion proteins 192

5.4 Protein localisation during low-level overexpression 198

5.4.1 Induction of overexpression with low IPTG concentrations 198

5.4.2 Localisation of EGFP fusion proteins following leaky expression 202

5.5 Identification of an intermediate Rcy overexpression phenotype 206

5.6 Discussion 208

5.6.1 Inconsistent localisation of *R. leguminosarum* ParA1 may reflect its dynamic nature 208

5.6.2 Consistent unipolar localisation of PopZ supports its involvement during unipolar growth of *Rhizobium* 209

5.6.3 Rcy is a unipolar protein which may dictate the site of cell growth 211

Chapter Six - General Discussion and Future Perspectives 213

6.1 ParA1 is required for the coordination of cell division 213

6.2 The polarity determinant PopZ may coordinate spatial regulation of cell division 214

6.3 The polar protein Rcy determines the site of peptidoglycan synthesis 215

6.4 ParA1, PopZ, Rcy, and the *Rhizobium* polar complex 217

6.5 Future perspectives 221

References 224

Appendix 253

A.1 Synthetic gene knockout constructs 253

A.2 Complications with the synthetic *parA1*-KO construct 256

A.3 His-tagged Protein expression constructs 258

A.4 Phase contrast microscopy of protein overexpression phenotypes 263

List of Figures

Figure 1.1. Phylogenetic tree of rhizobial and non-rhizobial Proteobacteria (Pseudomonadota). 15
Figure 1.2. Infection thread formation and progression during the establishment of rhizobium-legume symbiosis. 18
Figure 1.3. Transmission electron microscopy of *R. leguminosarum* bacteroids within pea nodules. 20
Figure 1.4. Structural comparison of eukaryotic tubulin and the prokaryotic tubulin homolog FtsZ. 24
Figure 1.5. Treadmilling of FtsZ filaments at the midcell during cell division. 26
Figure 1.6. Structural comparison of eukaryotic actin and the prokaryotic actin homolog MreB. 28
Figure 1.7. Representation of bacterial cytoskeletal actin filaments. 31
Figure 1.8. Hydrophobic seams of canonical and non-canonical coiled-coil domains. 33
Figure 1.9. Plasmid segregation by the Type I ParABS segregation system. 36
Figure 1.10. The elongasome and divisome protein complexes of *E. coli*. 40
Figure 1.11. Polar growth mechanisms in rod-shaped bacteria. 42
Figure 1.12. Proposed model of the polar growth complex within Hyphomicrobiales. 45
Figure 1.13. Organisation of *Caulobacter* cell poles. 53
Figure 1.14. Localisation of *Agrobacterium* PopZ and PodJ proteins. 55
Figure 1.15. Similarities in the predicted domain architecture of GPR and Rcy. 61
Figure 2.1. Plasmid map of EC41308. 75
Figure 2.2. Plasmid map of pUC57. 76
Figure 3.1. Schematic representation of the generation of whole-gene knockout mutants via homologous recombination using pK18*mobsacB*. 98
Figure 3.2. Plasmid map of the mobilizable sucrose-sensitive suicide vector pK18*mobsacB*. 99
Figure 3.3. Unmarked knockout construct generation, and validation of the knockout vectors pK18-*popZ*-KO and pK18-*rcy*-KO. 102
Figure 3.4. Plasmid map of the allelic exchange vector pK18-*popZ*-KO. 103
Figure 3.5. Plasmid map of the allelic exchange vector pK18-*rcy*-KO. 104
Figure 3.6. Restriction digest validation of resynthesized pK18-*parA1*-KO clones. 105
Figure 3.7. Plasmid map of the allelic exchange vector pK18-*parA1*-KO. 106

Figure 3.8. PCR-based confirmation of the integration of pK18-*popZ*-KO into the genome of *R. leguminosarum*. 109
Figure 3.9. PCR-based confirmation of integration of pK18-*rcy*-KO into the genome of *R. leguminosarum*. 112
Figure 3.10. Colony PCR screening of potential *popZ* and *rcy* knockout mutants..... 117
Figure 3.11. Directional PCR screen of the tetracycline resistance cassette within the knockout vector pK18-*popZ*-tet^{R}-KO..... 121
Figure 3.12. Colony PCR screening of potential tetracycline-resistant *popZ* and *rcy* knockout mutants. 125
Figure 4.1. The IPTG-inducible expression vector pIND6. 138
Figure 4.2. Generation and validation of pIND6-derived *parA1*, *popZ* and *rcy* expression plasmids. 141
Figure 4.3. The IPTG-inducible *parA1* and *popZ* expression vectors pIND6-*parA1* and pIND6-*popZ*. 143
Figure 4.4. The IPTG-inducible *rcy* expression vector pIND6-*rcy*. 144
Figure 4.5. Growth curve of *R. leguminosarum* in TY medium in the presence and absence of IPTG. 147
Figure 4.6. Impact of the overexpression of *parA1* on the growth of *R. leguminosarum*..... 149
Figure 4.7. Impact of *popZ* overexpression on the growth of *R. leguminosarum*. 151
Figure 4.8. Impact of the overexpression of rcy on the growth of *R. leguminosarum*. 152
Figure 4.9. SDS-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of whole-cell *R. leguminosarum* protein samples following overexpression of pIND6-*parA1*, pIND6-*popZ* and pIND6-*rcy*. 157
Figure 4.10. Western blot analysis of whole-cell *R. leguminosarum* protein samples following overexpression of *parA1* and *popZ*. 158
Figure 4.11. SDS-PAGE analysis of *R. leguminosarum* total protein samples following induction of *rcy* overexpression. 160
Figure 4.12. Morphological phenotypes of *R. leguminosarum* overexpressing ParA1, PopZ and Rcy**.** 163
Figure 4.13. Cell area and perimeter phenotypes of ParA1, PopZ and Rcy overexpression *R. leguminosarum* strains..... 167
Figure 4.14. Cell length and diameter phenotypes arising from ParA1, PopZ and Rcy overexpression within *R. leguminosarum*. 170
Figure 5.1. The generation of translational EGFP fusions using phage lambda (λ) Red recombination-based genetic engineering. 183

Figure 5.2. Extension the *egfp*-Apra^R cassette using PCR amplification. 185
Figure 5.3. XbaI validation digests of the ParA1-, PopZ- and Rcy-EGFP fusion protein expression plasmids**.** 188
Figure 5.4. The IPTG-inducible ParA1-EGFP expression plasmid pIND6-*parA1-egfp*. 189
Figure 5.5. The IPTG-inducible PopZ-EGFP expression plasmid pIND6-*popZ-egfp*..... 190
Figure 5.6. The IPTG-inducible Rcy-EGFP expression plasmid pIND6-*rcy-egfp*. 191
Figure 5.7. Confirmation of the absence of detectable green autofluorescence within *R. leguminosarum*. 193
Figure 5.8. Overexpression localisation of ParA1-EGFP within *R. leguminosarum*. 194
Figure 5.9. Overexpression localisation of PopZ-EGFP within *R. leguminosarum*..... 196
Figure 5.10. Overexpression localisation of Rcy-EGFP within *R. leguminosarum*. 197
Figure 5.11. Localisation of ParA1-EGFP following induction with 0.0025 mM IPTG..... 199
Figure 5.12. Localisation of PopZ-EGFP following induction with 0.01 mM IPTG. 200
Figure 5.13. Localisation of Rcy-EGFP following induction with 0.0025 mM IPTG..... 201
Figure 5.14. Leaky expression of ParA1-EGFP from pIND6-*parA1-egfp* in *R. leguminosarum*. 203
Figure 5.15. Leaky expression of PopZ-EGFP from pIND6-*popZ-egfp* within *R. leguminosarum*. 204
Figure 5.16. Leaky expression of Rcy-EGFP from pIND6-*rcy-egfp* within *R. leguminosarum*... 205
Figure 5.17. Overexpression localisation of Rcy-EGFP within *R. leguminosarum* cells displaying an intermediate branched overexpression phenotype. 207
Figure 6.1. Proposed model of the *R. leguminosarum* growth pole complex**.** 219

List of Tables

Table 2.1. Bacterial strains used in this thesis. 65
Table 2.2. Composition of bacterial growth media. 66
Table 2.3. List of plasmids used. 68
Table 2.4. Table of antibiotic, additive and inducer concentrations used throughout this thesis. 70
Table 2.5. Components of a typical restriction enzyme digestion. 73
Table 2.6. Constituents of sticky-ended DNA ligation reactions. 74
Table 2.7. Table of primers. 80
Table 2.8. Constituents of a Phusion polymerase PCR reaction**.** 82
Table 2.9. Thermal cycling parameters for a standard Phusion PCR reaction. 82
Table 2.10. Thermal cycling parameters for a Phusion PCR reaction with extension primers**.** . 83
Table 2.11. Constituents of a diagnostic PCR using GoTaq G2 Hot Start Polymerase. 84
Table 2.12. Cycle parameters of PCR reactions using GoTaq® G2 Hot Start Polymerase. 84
Table 2.13. Composition of buffers used for the analysis of proteins. 91
Table 2.14. Composition of a 10% (v/v) SDS-PAGE polyacrylamide gel. 92
Table 3.1. Genomic context of *parA1*, *popZ* and *rcy* within *R. leguminosarum* 3841**.** 128
Table 4.1. Summary of the impacts of *parA1*, *popZ* and *rcy* overexpression on *R. leguminosarum* cultures. 155
Table 4.2. Summary of morphological overexpression phenotypes. 172

Glossary of Abbreviations

ADP	Adenosine diphosphate
Amp	Ampicillin
Apra	Apramycin
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
Chl	Chloramphenicol
cPCR	Colony polymerase chain reaction
C-terminus	Carboxyl-terminus
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
EGFP	Enhanced green fluorescent protein
EtBr	Ethidium bromide
gDNA	Genomic deoxyribonucleic acid
GDP	Guanosine diphosphate
GFP	Green fluorescent protein
GOI	Gene of interest
GTP	Guanosine triphosphate
IPTG	Isopropyl β-d-1-thiogalactopyranoside
Kan	Kanamycin
KO	Knockout
LB	Luria Broth
MCS	Multiple cloning site
N-terminus	Amino-terminus
NTP	Nucleoside triphosphate
OD_{600}	Optical density (measured at a wavelength of 600 nm)
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PVDF	Polyvinylidene fluoride
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulfate

SOB	Super optimal broth
SOC	Super optimal broth with catabolite repression
Spc	Spectinomycin
Str	Streptomycin
T_a	Annealing temperature
TAE	Tris-Acetate-EDTA
TBS	Tris-buffered saline
TBS-T	TBS-Tween
Tet	Tetracycline
T_m	Melting temperature
Tris	Tris(hydroxymethyl)aminomethane
TY	Tryptone-yeast extract broth
UV	Ultra-violet
WT	Wildtype
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

Acknowledgements

The completion of this thesis could not have been possible without the support and guidance of my supervisors Dr Ben Miller, Dr Andrew Gates and Dr Gabriella Kelemen. My sincere thanks go to all members of their research groups, as well as the many others who have helped me along the way, listened to my problems, and celebrated my successes with me.

I would also like to acknowledge the Processors and Growers Research Organisation, who supported this work. I am particularly grateful for those at PGRO who kindly offered not only their advice and encouragement, but also their friendship.

Last, but by no means least, I owe a special debt of gratitude to my family and my partner for their unconditional support throughout this journey.

Chapter One – Introduction

1.1 Rhizobia, *Rhizobium* and the *Rhizobiaceae*

The *Rhizobiaceae* is an important and diverse family of Gram-negative bacteria within the Alphaproteobacteria order Hyphomicrobiales (formerly Rhizobiales). Members of this family are typically aerobes, rod-shaped in morphology, non-spore forming, and often exhibit flagella-driven motility (*1*, *2*). Historically, the *Rhizobiaceae* was generally considered to be made up of five 'main' genera (*Rhizobium*, *Ensifer* (*Sinorhizobium*), *Agrobacterium*, *Shinella*, and *Carbophilus*), which can be divided into two major superclades grouping *Rhizobium* and *Agrobacterium* into one superclade, and *Shinella* and *Ensifer* within another (*2*, *3*). However, like many bacterial families, the taxonomic classification of species within this family is poor. Up-to-date phylogenetic analysis suggests that the *Rhizobiaceae* is actually formed of 19 distinct and accepted genera, with several additional novel genera and genospecies having been proposed in recent years (*4*–*7*). The notable scientific interest and agricultural importance of this family stems from the fact that many species that lie within it are able to modulate plant growth. Not only does this family possess plant pathogens such as *Agrobacterium tumefaciens*, *Rhizobium rhizogenes* (formerly *Agrobacterium rhizogenes*) and *Agrobacterium rubi* (the causal agents of crown gall, hairy root and cane gall diseases respectively) (*8*), but many other members of the *Rhizobiaceae* instead form mutualistic endosymbiotic interactions with leguminous plants (Fabales) within which fixed nitrogen (predominantly in the form of ammonia) is supplied to the host plant in exchange for carbon sources (primarily malate) (*9*, *10*).

This group of nitrogen-fixing (diazotrophic) facultative endosymbionts are collectively referred to as rhizobia. Within the *Rhizobiaceae*, rhizobia can be found within several genera including *Rhizobium*, *Ensifer* (*Sinorhizobium*), and *Shinella* (*2*). Of particular interest within this thesis is the *Rhizobium* species complex *Rhizobium leguminosarum*. In 1888, Martinus Beijerinck isolated and cultured the first known 'rhizobia' from a root nodule, which he named *Bacillus radicicola* (*11*). This species was later renamed *Rhizobium leguminosarum* (*12*). Since then, a great number of strains have been categorised as *Rhizobium leguminosarum* species*,* and it is now believed that this species complex contains at least 18 distinct genospecies, as well as a number of outlying species (*7*, *13*). Although the *Rhizobiaceae* will remain the focus of this thesis, it is important to mention that rhizobial species are also found within a number of other genera beyond the *Rhizobiaceae*, including *Azorhizobium (Xanthobacteraceae*)*, Bradyrhizobium*

(Nitrobacteraceae) and Mesorhizobium (Phyllobacteriaceae) (Figure 1.1) (*14*, *15*). Furthermore, symbiotic nitrogen fixation (SNF) is not solely restricted to the rhizobia, with members of the Actinobacteria genus *Frankia* capable of forming symbiotic relationships with actinorhizal plants (16).

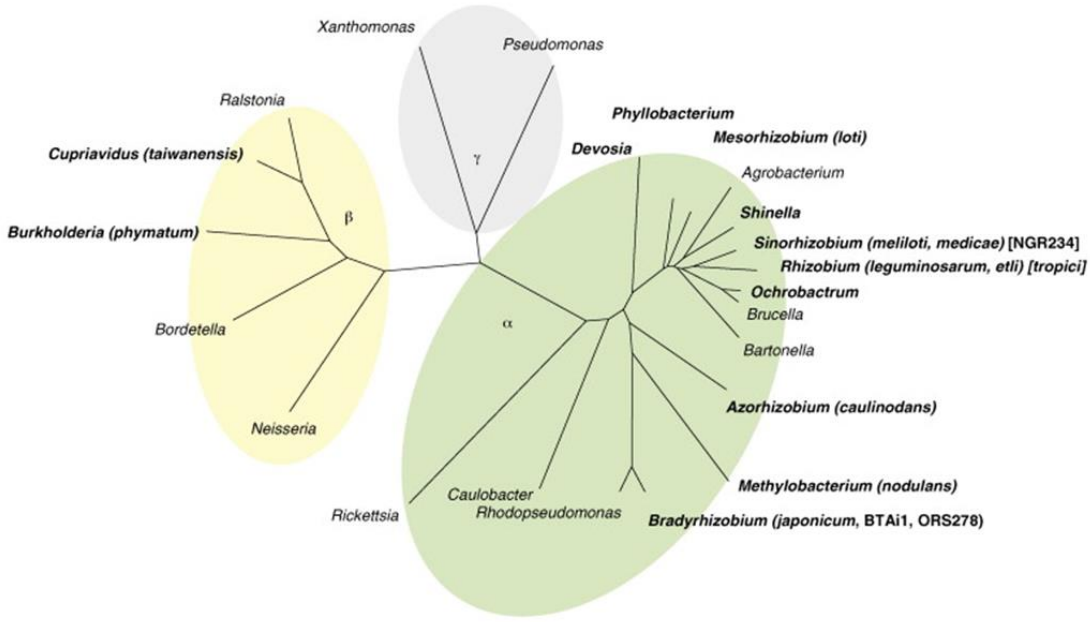


Figure 1.1. Phylogenetic tree of rhizobial and non-rhizobial Proteobacteria (Pseudomonadota). The 16S ribosomal RNA gene sequences of various Alpha(α)-, Beta(β)- and Gamma(γ)-proteobacteria species were used to generated this unrooted phylogenetic tree, published by Masson-Boivin *et al.* (2009) (*15*). In this tree, genera in bold are those which are known to contain rhizobial species (i.e. nitrogen-fixing species which form nodules within legumes).

1.1.1 The rhizobium-legume symbioses

The endosymbiotic relationship between rhizobia and legumes, referred to as the rhizobium-legume symbiosis (not to be confused with the genus *Rhizobium*), is of great economic and environmental importance. It is estimated that rhizobium-legume symbiosis is responsible for the input of somewhere in the region of 20-80 million tonnes of fixed nitrogen into global agricultural systems annually (*17*, *18*). However, global agriculture still remains dependent on the application of synthetic nitrogenous fertilisers, with over 90 million tonnes used each year (*19*, *20*). To fix the vast quantities of nitrogen required for the manufacture of fertilisers, an industrial process called the Haber-Bosch process is used to convert gaseous nitrogen (N_2) into

ammonia via a reaction with hydrogen under high temperature and pressure (*21*). This reaction is incredibly energy-intensive, accounting for more than 1% of the world's total energy consumption (*22*), and relies heavily on natural gas (methane) as a source of both energy and hydrogen, consuming between 3 to 5% of global natural gas supplies (*23*). Not only does the manufacture, transport and application of synthetic nitrogenous fertilisers emit the greenhouse gas (GHG) carbon dioxide (CO_2), but the metabolism of nitrogenous compounds within these fertilisers by soil microorganisms results in the production of nitrous oxide (N_2O), a GHG approximately 300x more potent than CO_2 (*24*–*27*). Alongside GHG emissions, excessive application of nitrogenous fertilisers can pose further environmental risks. The introduction of ammonia- and nitrate-rich agricultural runoff (formed by leaching of these nutrients from the soil following heavy rainfall) into waterways can lead to aquatic pollution, eutrophication, and the eventual loss of aquatic biodiversity (*25*, *26*, *28*, *29*). Enhancing biological nitrogen fixation through rhizobium-legume symbioses can therefore help to improve agricultural sustainability by reducing the requirement for synthetic fertiliser application.

Our current understanding of the mechanisms controlling the establishment and maintenance of rhizobium-legume symbioses has been extensively reviewed (see (*30*–*33*) for notable examples). The initiation of these important endosymbiotic relationships is coordinated via molecular signalling between the host plant and the bacterial symbiont. Under nitrogen-deficient conditions, leguminous plants begin this two-way communication by secreting signalling molecules (predominantly flavonoids) into the soil (*34*, *35*). These molecules, which act as chemoattractants, are detected by rhizobia within the surrounding rhizosphere and trigger the migration and subsequent attachment of these bacteria to the root surface (*36*–*38*). In turn, the detection of flavonoids by rhizobial cells induces expression of the rhizobial *nod* gene cluster, resulting in the production and secretion of nodulation (Nod) factors (lipo-chitooligosaccharide molecules) which are subsequently detected by the host plant (*39*–*43*). Following Nod factor perception, cytoskeletal rearrangements (e.g. actin rearrangement and microtubule array formation) and cell wall remodelling within the tip of root hair cells results in the formation of 180° to 360° root hair curls within which rhizobia (typically a single cell) become entrapped (*44*–*49*). Proliferation of the entrapped rhizobial cell gives rise to a microcolony within the curl, referred to as an infection chamber or pocket, and subsequent invagination of the plant cell wall surrounding the infection chamber generates a tubular plant-derived structure called an infection thread (IT) (Figure 1.2) (*31*, *50*). Ahead of the IT is an unpopulated region referred to as a pre-infection thread. These structures, formed of cytoplasmic 'bridges'

surrounded by longitudinal arrays of microtubules, connect one end of the plant cell to the other and serve as a template for the formation of infection threads (*47*, *51*, *52*). During IT progression, the pre-infection thread is populated by a roughly single-file thread of free-living rhizobia, originating from a restricted division zone behind the growing IT tip (*14*, *52*, *53*). It is worth noting that while growing ITs typically exhibit a single-file rhizobial thread, mature ITs can contain multiple files of rhizobia (*52*, *54*).

In order for the infection thread to pass from one cell to the next, plant cell wall modification at the site of the growing IT tip is required. This process is proposed to have two stages: local cell wall modification by Nod factor-induced plant enzymes; and cell wall degradation by secreted rhizobial enzymes (*50*, *55*–*59*). Transcellular progression of the IT from the root epidermis into the root tissue (and subsequent proliferation of free-living rhizobia within this structure) directs the rhizobial cells from the root hair surface down into the cortex where the root nodule is ultimately formed (Figure 1.2) (*47*, *52*). Although infection threads are plant-derived structures, rhizobia play a vital role during both IT formation and progression, primarily via the ongoing synthesis of Nod factors (*31*, *43*, *60*).

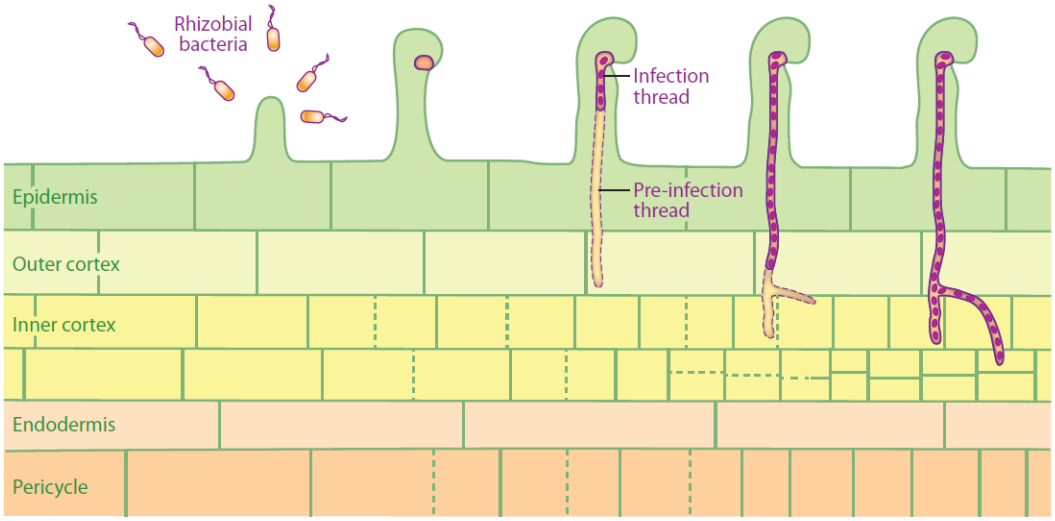


Figure 1.2. Infection thread formation and progression during the establishment of rhizobium-legume symbiosis. The process of rhizobial infection during the establishment of rhizobium-legume symbioses is well-reviewed (*31*–*33*, *52*, *53*). First, the secretion of plant signalling molecules (flavonoids) into the surrounding rhizosphere triggers the migration and attachment of nearby rhizobia to the root surface. In turn, rhizobia produce and secrete nodulation (Nod) factors, which are detected by the host plant. The detection of Nod factors triggers the formation of root hair curls, within which a rhizobial cell becomes entrapped and proliferates. Invagination of the plant cell membrane surrounding this infection pocket gives rise to a plant-derived tubular infection thread, which is subsequently populated by a single-file stream of free-living rhizobia (purple). A pre-infection thread within the root cells guides the direction of the growing infection thread from the root epidermis down into the cortex. Cell divisions (dashed lines) within the inner cortex and pericycle (initiated early in the infection process following Nod factor perception) give rise to the nodule primordium from which the nodule is derived. Figure published by Oldroyd *et al.*, 2010 (*31*).

1.1.2 Nodule organogenesis and bacteroid differentiation

To successfully achieve nodulation, careful coordination of bacterial infection and nodule formation is required. In parallel with triggering morphological changes within root hair cells (i.e. curl formation), the initial perception of Nod factors early on in the rhizobium-legume signalling exchange also initiates the division of cells within the root cortex and pericycle, giving rise to nodule primordium (*39*, *61*, *62*). Once the infection thread reaches these nodule primordium cells, specialised outgrowths of the IT which lack plant cell wall are formed, termed infection

droplets (*46*, *60*, *63*). As observed during root hair curling and infection thread progression, the formation of infection droplets is also associated with dynamic cytoskeletal arrangements within the plant cell (*64*). Rhizobia are then released from infection droplets into the surrounding nodule primordium, but remain separated from the cytoplasm by a plant-derived membrane (*60*, *65*, *66*). Within these membrane-bound structures, referred to as symbiosomes, rhizobia differentiate into highly specialised nitrogen-fixing bacteroids, undergoing substantial physiological and in some cases morphological changes (*66*, *67*).

Leguminous plants form one of two types of root nodules, categorised based on the presence/absence of an active meristem that continually gives rise to new nodule primordia. Determinate nodules, which lack an active meristem, are spherical nodules with a single nitrogen-fixing zone throughout the nodule. In contrast, indeterminate nodules (which possess an active meristem) are elongated nodules with a gradient of zones of activity throughout the nodule, only one of which (zone III) is a nitrogen-fixation zone (*14*, *68*, *69*). Within nitrogen-fixing regions of nodules, careful regulation of free oxygen is vital due to the oxygen-sensitive nature of the nitrogen-fixing enzyme nitrogenase (*70*–*72*). Hence, a physical barrier within the nodule (the oxygen diffusion barrier (ODB)) controls the diffusion of free oxygen into the active zone of the nodule, while the phytoglobin leghaemoglobin facilitates the transport of oxygen (which is required for rhizobial respiration) directly to the bacteroids (*73*, *74*).

Broadly speaking, symbiosomes within determinate nodules contain multiple bacteroids of a similar size and shape to free-living rhizobia, while those within indeterminate nodules usually contain a single, swollen, pleomorphic bacteroid (*14*, *67*). Bacteroid morphotypes can be divided into three main types; E-type: enlarged, elongated bacteroids, which are often also branched; S-type: enlarged, spherical bacteroids; and U-type: bacteroids of a size and shape similar to free-living rhizobia (*75*, *76*). While U-type bacteroids are not terminally-differentiated, the genomes of E and S morphotypes undergo multiple rounds of endoreduplication, resulting in terminally differentiated bacteroids which are incapable of replication (*67*, *75*, *76*). Terminal bacteroid differentiation is typically associated with more efficient symbiotic nitrogen fixation, and is not dependent on the rhizobial species but is instead determined by signalling molecules within the nodule (*77*–*79*). In nodules of legumes belonging to the inverted repeat lacking clade (IRLC) such as *Medicago* and members of the Fabeae tribe (e.g. pea, lentil, broad bean and vetch), the differentiation of bacteroids is controlled predominantly by the action of nodule-specific

cysteine-rich (NCR) peptides within the nodule environment (*77*–*80*). For instance, the nodulation of pea (*Pisum sativum*) by *Rhizobium leguminosarum* species gives rise to indeterminate root nodules filled with E-type bacteroids (Figure 1.3), yet the nodulation of bean (*Phaseolus vulgaris*) by members of this genospecies generates determinate nodules filled with U-type bacteroids (*76*). Similarly, *Bradyrhizobium* sp. strain ORS285 can form both E- and S-type bacteroids during symbioses with *Aeschynomene afraspera* and *Aeschynomene indica* respectively (*75*, *81*). Although rhizobia to not determine the type of nodule formed, rhizobial exopolysaccharides do play a key role in the successful development of mature indeterminate nodules (*14*, *82*, *83*).

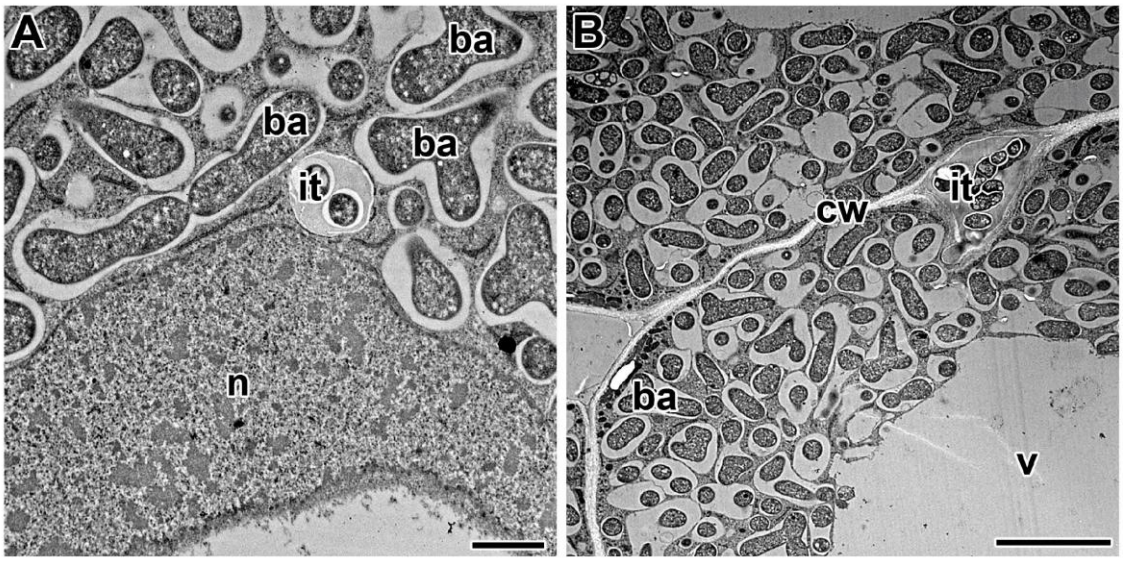


Figure 1.3. Transmission electron microscopy of *R. leguminosarum* bacteroids within pea nodules. This figure shows two panels from a figure by Tsyganova *et al.* (2021) (*60*), showing infection thread (it) and bacteroid (ba) structures within symbiotic nodules of *Pisum sativum*. *R. leguminosarum* bacteroids are separated from the plant cell cytosol by the symbiosome membrane, a specialised structure derived from the plant cell plasma membrane (*84*). Other plant structures shown in these figures include the nucleus (n), a vacuole (v) and the cell wall (cw). Scale bars represent 1 µm (A) and 5 µm (B) (*60*).

Regardless of morphotype, all bacteroids still undergo significant metabolic changes, essentially becoming highly-specialised nitrogen-fixing organelles within the plant cell (*66*). Many key processes required for cell growth and division are shut down, including the synthesis of ribosomal proteins, nucleic acid synthesis and repair, and the synthesis of outer membrane proteins and peptidoglycan (*31*, *85*). Expression of the nitrogen fixation (*nif* and *fix*) gene clusters, which encode nitrogenase (*nif* genes) and a collection of additional proteins required for nitrogen fixation (*fix* genes), is switched on, while the assimilation of ammonia into amino acids is switched off (*86*–*90*). Transcriptomic and proteomic analysis of *Rhizobium* bacteroids within pea nodules has demonstrated increased expression of genes and proteins associated with DNA replication, membrane synthesis and the TCA (tricarboxylic acid) cycle within these bacteroids, while the expression of housekeeping and cell division genes was downregulated (*86*, *91*). Studies of *Bradyrhizobium* E- and S-type bacteroids (within *Aeschynomene* nodules) demonstrated that independent of morphotype, bacteroids undergo metabolic changes involving the upregulation of a suite of genes involved in energy production, nitrogen fixation, and inorganic ion transport and metabolism, coupled with downregulation of genes involved in translation, secondary metabolite biosynthesis and cell motility (*81*). Furthermore, differential expression of a number of cell cycle and peptidoglycan biosynthesis genes was observed between each bacteroid morphotype and free-living *Bradyrhizobium*, highlighting that the regulation of these genes is an important step in bacteroid development (*81*). Although rhizobial maturation does ensure efficient nitrogen fixation, these physiological changes do come at a cost. Once differentiation of rhizobia is underway, the maturation and survival of the ensuing bacteroid relies heavily upon the supply of amino acids (particularly branched-chain amino acids) and reduced carbon (predominantly in the form of dicarboxylic acids) by the plant (*89*, *92*). Furthermore, while bacteroids within determinate nodules (i.e. U-type bacteroids) are capable of resuming free-living growth upon release from the nodule, the differentiation of rhizobia into E- and S-type bacteroids is terminal, preventing the survival of these bacteroids outside of the nodule environment (*67*, *75*, *93*).

1.2 Coordinating growth, division, and morphology: The bacterial cytoskeleton

The ability to not only maintain but also regulate cell size and shape is essential for the survival of bacterial cells (*94*, *95*). In the case of rhizobia, the maintenance of rod-shaped morphology and uniform cell size is crucial for the survival of free-living cells, yet the capacity to modify cell morphology in response to environmental factors or signals (e.g. differentiation into bacteroids within the nodule environment) is equally as important for successful endosymbiotic interactions (*75*, *76*). As reviewed in Section 1.1, the establishment of rhizobium-legume symbiosis depends upon a series of highly coordinated cell proliferations and morphological changes within both the host plant and its endosymbiont. Although the role of the eukaryotic cytoskeleton in bringing about the morphological changes which permit colonisation such as rearrangements of actin filaments and microtubule arrays during root hair curling, infection thread progression and the formation of infection droplets is relatively well documented (*45*, *47–49*, *64*, *96–99*), literature surrounding how rhizobia may coordinate similar and/or accompanying processes such as population of infection threads and differentiation into bacteroids is very limited.

Over the last 30 years, technological advances in the ability to visualize prokaryotic structures has led to the identification of a diverse cytoskeletal protein network within prokaryotes, similar to that of eukaryotic cells, which is essential for an array of fundamental cellular processes including division, growth, motility, morphogenesis and DNA partitioning (*100–102*). Within eukaryotes, the cytoskeletal network is comprised of three groups of protein filaments: tubulin, actin, and intermediate filaments (*103*). The prokaryotic cytoskeleton consists of homologues of these three groups of proteins, as well as a fourth group of proteins with no distinct eukaryotic cytoskeletal homolog. This cytoskeletal network is made up of a broad variety of proteins ranging from those with relatively static structural roles, to proteins with highly dynamic organisational roles (*102*). A member of the tubulin superfamily, the cell division protein FtsZ of *Escherichia coli* (a Gram-negative rod), was the first discovered prokaryotic cytoskeletal protein (*104–106*). The actin homolog MreB was identified a few years later within the Gram-positive rod-shaped bacterium *Bacillus subtilis*, and was determined to be a cell morphology determinant (*107*). Shortly after, the intermediate filament-like protein crescentin (CreS) was identified within the crescent-shaped Gram-negative bacterium *Caulobacter crescentus* (formally *Caulobacter vibrioides*), and was attributed to the control of cell curvature within this species (*108*). As FtsZ, MreB and CreS were the first identified prokaryotic homologs of tubulin,

actin, and intermediate filaments respectively, they remain some of the most well-characterised prokaryotic cytoskeletal components, with *E. coli*, *B. subtilis* and *C. vibrioides* becoming key model species for the study of bacterial cytoskeletal proteins (*100–102*). Within this literature review, the term bacterial cytoskeleton has often been used in favour of prokaryotic cytoskeleton, as the archaeal homologs of this protein network have not been extensively reviewed here. In recent years, the cytoskeletal systems of a much broader range of bacterial species have been better-characterised, including *Streptomyces coelicolor* (Gram-positive filamentous bacteria), *Vibrio cholerae* (Gram-negative curved rod), *Staphylococcus aureus* (Gram-positive coccoid), and the *Rhizobiaceae* member *Agrobacterium tumefaciens* (Gram-negative rod) (*100*, *102*, *109*, *110*), yet our knowledge of this protein network within *Rhizobium* species remains very limited. As will become apparent throughout this literature review, bacterial cytoskeletal systems are very diverse, with different species possessing different suites of cytoskeletal proteins. This literature review will therefore primarily focus on the known cytoskeletal proteins within Gram-negative and rod-shaped model species.

1.2.1 Tubulin homologs

Bacterial tubulin homologs, much like their eukaryotic counterparts, are a widely-distributed group of protein monomers that form dynamic cytomotive filaments which push or pull other molecules within the cell (*102*, *111*). Polymerisation of tubulin monomers generates one-subunit-thick protofilaments which can assemble to form higher-order structures such as rings, bundles, sheets and tubes (*111–113*). Eukaryotic tubulin and bacterial tubulin-like proteins share a common domain structure consisting of two globular domains (an N-terminal GTP-binding domain (GTPase domain) and a C-terminal GTPase activation domain) separated by a central core helix and synergy loop (T7 loop) (Figure 1.4) (*110*, *114*). Polymerization of bacterial tubulin monomers is facilitated by the binding of GTP to the N-terminal GTPase domain (*114*). Upon GTP binding, the interaction of the N-terminal GTPase domain of one tubulin monomer with the C-terminal activation domain of the next monomer forms an active GTPase site around the GTP molecule (*115*). Shortly after binding, hydrolysis of GTP by the GTPase domain (facilitated by the interaction of the T7 loop of one monomer with the GTPase domain of the next) results in the formation of a GDP-tubulin subunit (*116*). This GDP-tubulin subunit is more weakly-associated with the polymer than GTP-tubulin, and typically will quickly dissociate from the end of the

filament unless stabilised by other regulatory proteins (*113*). Cycles of polymerisation and depolymerisation result in the generation of a dynamic protein filament, capable of growth, shrinkage, and molecular treadmilling (*110*, *114*, *117*, *118*). Treadmilling of bacterial tubulin filaments is often essential for their cytomotive function, and is the result of simultaneous growth (polymerisation) and one end of the filament and shrinkage (depolymerisation) at the opposite end, leading to an overall motion of the filament in the direction of filament growth (*119*, *120*).

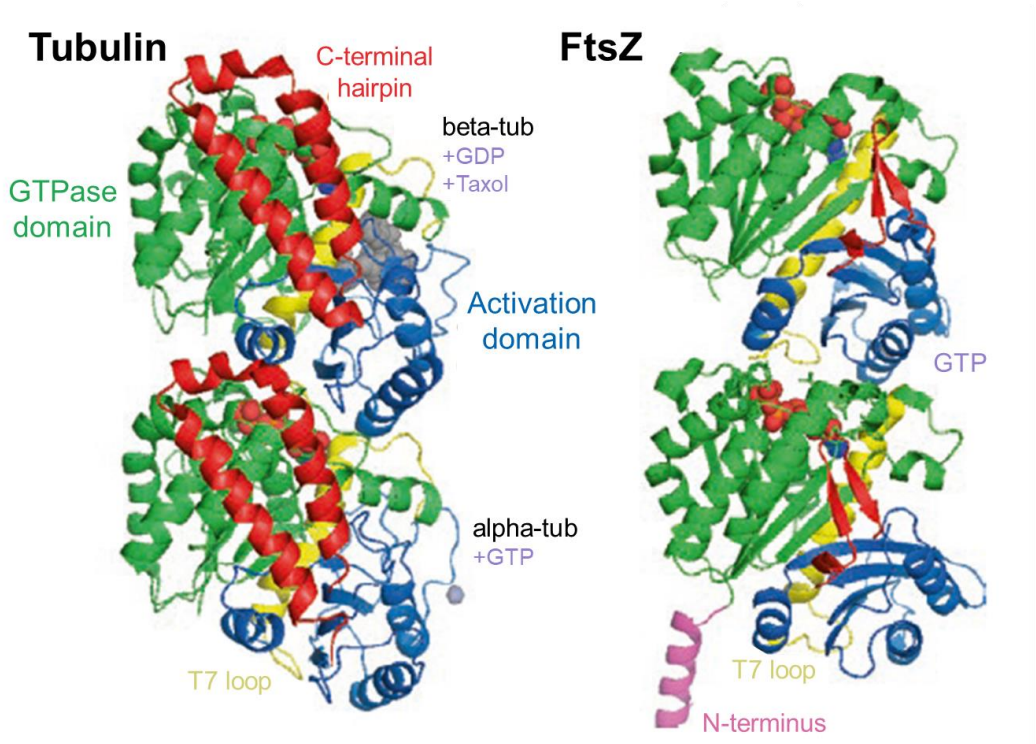


Figure 1.4. Structural comparison of eukaryotic tubulin and the prokaryotic tubulin homolog FtsZ. This figure, modified from Löwe and Amos (2017), shows ribbon diagrams of the alpha- and beta-tubulin (tub) heterodimer (left) and an FtsZ homodimer (right) (*121–123*). Both eukaryotic and prokaryotic tubulin monomers share common a domain structure consisting of a GTPase domain (green) and a GTPase-activation domain (blue) connected by a central helix (yellow) and T7 loop. Both proteins also possess a C-terminal hairpin (red). One molecule of GTP, GDP and taxol (a small molecule which stabilizes tubulin lattices) are shown bound to tubulin, with one molecule of GTP shown bound to FtsZ (represented as ball models). Structures shown are for proteins from *Bos taurus* (tubulin) and *Methanococcus jannaschii* (FtsZ).

The first prokaryotic tubulin homolog identified was the cell division protein FtsZ of *E. coli* (*104*–*106*). FtsZ mediates cell division (cytokinesis) by facilitating the recruitment of cell wall remodelling enzymes to the division site. Within the cytoplasm, FtsZ monomers polymerise into short, polar filaments (*104*, *124*). Prior to cytokinesis, FtsZ filaments assemble (with the assistance of a number of regulatory factors) at the midcell to form a contractile ring structure consisting of short, loosely-bundled FtsZ polymers, referred to as the Z ring (Figure 1.5) (*113*). During cytokinesis, FtsZ filaments treadmill around the circumference of the midcell, carrying cell wall remodelling enzymes such as peptidoglycan biosynthesis enzymes with them (*119*, *125*–*127*). These FtsZ filaments form an integral part of a membrane-associated protein complex called the divisome (see Section 1.3.1), comprised of a large number of proteins which coordinate cell wall remodelling and cell division (*128*, *129*). Circumferential cell wall synthesis at the division plane (septum formation) and subsequent constriction of the Z ring results in eventual separation of the daughter cells (*119*, *125*–*127*).

While most bacterial species possess only a single copy of FtsZ, *Agrobacterium tumefaciens* possesses three homologs, two of which localise to the midcell and form filaments. However, only one of these two filament-forming FtsZ homologs (the canonical FtsZ homolog, termed $FtsZ_{At}$) is essential for cell division (*127*, *130*). Within *Agrobacterium*, $FtsZ_{At}$ exhibits localisation at the growth pole for the majority of the cell cycle, then relocates to the midcell prior to cell division (*127*, *131*). Species within the rhizobial genera *Rhizobium*, *Sinorhizobium*, *Mesorhizobium* and *Bradyrhizobium* also possess a second FtsZ homolog (FtsZ2), the acquisition of which is believed to have been a critical step in the evolution of rhizobium-legume symbiosis (*132*–*135*). Although FtsZ2 is not essential for cell viability, it is proposed that this homolog may play an important role during bacteroid formation. Two independent gene duplications within the order Hyphomicrobiales are proposed to have given rise to these two FtsZ homologs. The first, predating the divergence of *Rhizobiaceae* and *Phyllobacteriaceae*, yielded the FtsZ1 and FtsZ2 paralogues found within *Mesorhizobium* and the *Rhizobiaceae* genera *Sinorhizobium*, *Rhizobium*, and *Agrobacterium*. The second, predicted to have occurred within the *Bradyrhizobiaceae*, yielded the FtsZ1a and FtsZ1b homologs found within *Bradyrhizobium* species (*135*). *Agrobacterium* and *Sinorhizobium* FtsZ1 proteins have been found to interact with one another, and both localise to the Z ring when expressed within *E. coli*, suggesting a highly-conserved function of FtsZ proteins across bacterial species (*136*).

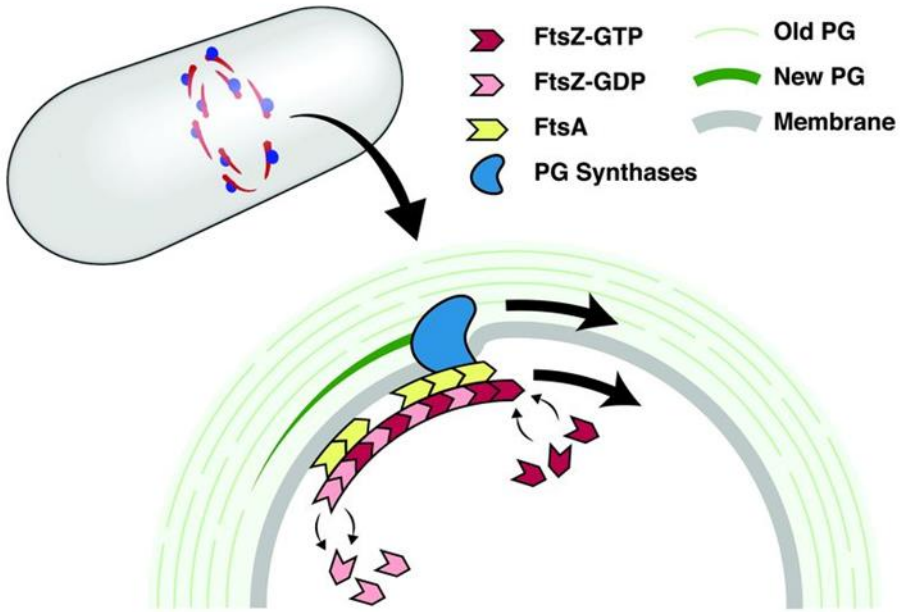


Figure 1.5. Treadmilling of FtsZ filaments at the midcell during cell division. During cytokinesis, multiple FtsZ filaments treadmill around the circumference of the division plane, carrying peptidoglycan (PG) synthases with them (*125*). The association of FtsZ filaments with the cell membrane is facilitated by interactions with the bacterial actin homolog FtsA (*126*). FtsZ treadmilling is the result of simultaneous incorporation of FtsZ-GTP monomers to one end of the filament, and release of FtsZ-GDP monomers (following GTP hydrolysis) from the other, resulting in net movement of the filament in the direction of filament growth (*119*). This treadmilling of FtsZ around the cell results in constricting rings of PG (i.e. cell wall) synthesis at the septum site, eventually dividing the cell into two daughter cells. Figure modified from Bisson-Filho *et al.,* 2017 (*125*).

In addition to FtsZ, several other bacterial tubulins have been identified including TubZ, FtsZm and the BtubAB proteins. However, these tubulin-like proteins are nowhere near as ubiquitous or well-characterised as FtsZ. FtsZm and BtubAB and not widely distributed, having only been characterised within *Magnetospirillum* and *Prosthecobacter* respectively, and their function within these species remains poorly understood. FtsZm, an FtsZ-like protein which may interact with FtsZ, is proposed to be required for magnetosome formation and distribution, and is believed to be important for denitrification, redox control, and iron uptake within magnetotactic bacteria (*137–139*). The biological function of the BtubAB proteins (BtubA and BtubB, homologs of the α- and β-tubulin proteins) is still not fully-understood, but it has been observed that these proteins form dynamic mini-microtubule structures within the cell, similar to microtubule arrays within eukaryotes (*140–142*). TubZ however is more widely distributed and better characterised.

TubZ proteins are typically encoded on rapidly evolving bacterial plasmids such as the virulence plasmids of *Bacillus* and *Clostridium* species. TubZ monomers polymerise into double helical spindle-like filaments which form part of the type III DNA partitioning system responsible for faithful segregation of these plasmids (*120*, *143*, *144*).

1.2.2 Actin homologs

Prokaryotic actin homologs are a diverse group of filament-forming cytoskeletal proteins, all of which share a common ATPase domain also present within hexokinase and hsp70 (heat shock protein 70) protein families (*145*). Polymerisation of actin-like monomers generates protofilaments which are capable of assembling into a wide variety of dynamic higher-order structures (*102*, *111*). Prokaryotic actin monomers share a conserved domain structure, similar to that of eukaryotic actin, composed of two domains (I and II) each of which are divided into two subdomains (IA and IB, and IIA and IIB) which function as pairs (Figure 1.6) (*145*–*147*). ATP binds within the cleft between domains IB and IIB (the interdomain or 'nucleotide-binding' cleft) which closes upon nucleotide binding, bringing the two subdomains together (*146*). Much like tubulin monomers, actin monomers favour polymerisation in their NTP-bound form. The binding of ATP within the interdomain cleft facilitates head-to-tail polymerisation of prokaryotic actin monomers into two-stranded polar helical filaments (*148*). During polymerisation, residues at the bottom of subdomains IA and IIA insert into the cleft between subdomains IB and IIB of the monomer below (*147*). Following polymerisation, hydrolysis of ATP by the ATPase domain forms an ADP-actin subunit. As observed with GDP-tubulin, ADP-actin subunits are more weakly associated with the polymer and quickly dissociate from the filament, resulting in a dynamic prokaryotic actin filament similar to that of eukaryotic F-actin (*148*, *149*).

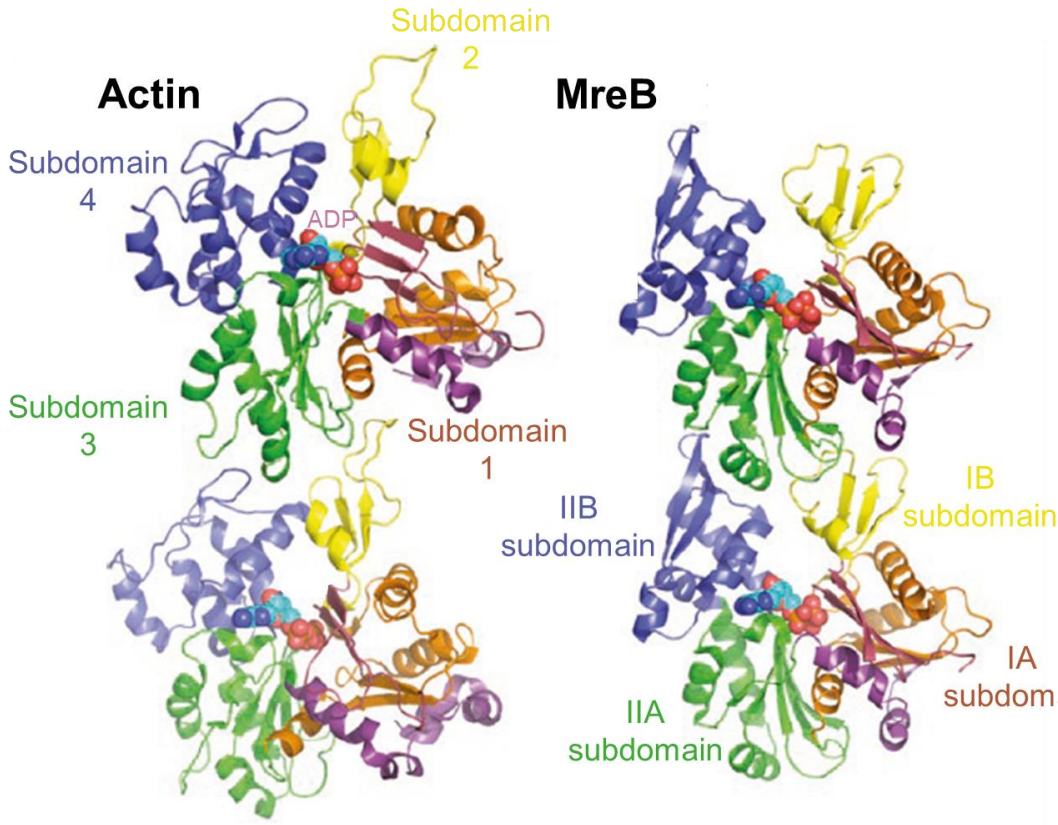


Figure 1.6. Structural comparison of eukaryotic actin and the prokaryotic actin homolog MreB. This figure, modified from Löwe and Amos (2017), shows ribbon diagrams of actin (left) and MreB (right) homodimers (*123*, *147*, *150*). Both eukaryotic and prokaryotic actin monomers share common a domain structure consisting of two domains, each divided into two subdomains; IA (pink & orange) and IB (yellow), and IIA (green) and IIB (purple) (actin subdomains 1, 2, 3 and 4 respectively); which function as pairs. Polymerisation of monomers is catalysed by binding of ATP within the cleft between domains IIB and IB, which is subsequently hydrolysed into ADP. One molecule of ADP is shown bound to each monomer (represented as a ball model). Structures shown are for proteins from *Oryctolagus cuniculus* (actin) and *Thermotoga maritima* (MreB).

Several bacterial actin homologs have been identified including MreB, FtsA, MamK and ParM, which are involved in a wide array of biological functions (Figure 1.7). Of these, MreB was the first prokaryotic actin homolog to be identified (*107*). Originally discovered in *Bacillus subtilis*, it is now known that MreB homologs are present in almost all non-coccoid bacteria, serving an important role in the organisation of cell wall synthesis and maintenance of cylindrical morphology during cell elongation (*148*, *151*). However, members of the Gram-positive phylum Actinobacteria and Gram-negative order Hyphomicrobiales (of which the *Rhizobiaceae* belong) lack MreB (*152*, *153*). It has been proposed that the loss of MreB, followed by the acquisition of

a second copy of the prokaryotic tubulin homolog FtsZ (see Section 1.2.1), were key events during the evolution of the rhizobium-legume symbiosis (*132*). The interaction of exogenous MreB with endogenous FtsZ1 and FtsZ2 within rhizobia (*Mesorhizobium huakuii*) has been shown to be detrimental for bacteroid differentiation, halting the establishment of rhizobium-legume symbioses during the late stages of bacteroid development (*132*). Members of the MreB family (including MreB-like (Mbl) proteins) form antiparallel double filaments around the perimeter of the cylindrical region of cells, binding to the plasma membrane (*154*, *155*). MreB filaments move circumferentially with the cell wall synthesis enzymes of the elongasome (see Section 1.3.1), coordinating local cell wall synthesis around the edge of the cell. However, unlike FtsZ which is driven by molecular treadmilling, the movement of MreB is instead driven directly by cell wall synthesis (*156*–*158*).

One actin homolog which is present within members of the Hyphomicrobiales (but not Actinobacteria) is the cell division protein FtsA. FtsA is a slightly unusual actin homolog, lacking the IB subdomain present in MreB and other canonical actin homologs, and also possessing a large insertion in subdomain IA (*153*, *159*). However, despite these differences, FtsA is still able to polymerise. Filaments of FtsA co-polymerase with FtsZ filaments within the Z ring (Figure 1.5), with FtsA tethering FtsZ filaments to the cell membrane via interaction with a conserved C-terminal region within FtsZ (*126*, *159*, *160*). Interestingly, FtsA filaments are destabilized by unassembled FtsZ monomers, suggesting that the polymerisation of FtsA and FtsZ is co-regulated and may serve as a checkpoint for division progression (*161*). Much like $FtsZ_{At}$, FtsA within *Agrobacterium* exhibits polar localisation for the majority of the cell cycle, then relocates to the Z ring prior to cell division (*127*, *131*). Within *Agrobacterium tumefaciens*, FtsZ, FtsA and the membrane protein FtsW have been found to regulate the transition from cell growth to cell division, providing further evidence that the co-regulation of FtsZ and FtsA polymerisation is a key checkpoint for cell cycle progression within bacteria (*127*, *130*).

A third group of bacterial cytoskeletal actin homologs is the ParM and Alf (actin-like filaments) family. Typically, the segregation of bacterial plasmids is achieved using one of two types of DNA partitioning system, referred to as the Type I and Type II partition systems. Type I partitioning systems are more common, and require members of a group of proteins called the Walker A cytoskeletal ATPases (WACAs) (described within Section 1.2.4), while Type II systems use actin-like ATPases (e.g. ParM and AlfA) (*162*, *163*). These plasmid-encoded actin homologs form actin-

like spindles within the cell which push plasmid DNA to the cell poles (*164*). ParM forms part of an independent tripartite DNA segregation system encoded on some low-copy-number plasmids referred to as the ParMRC system (*146*). This system is composed of three components: the bacterial actin ParM, a DNA-binding adaptor protein (ParR) and a centromere-like DNA region (*parC*). First, ParR adaptor proteins bind to the *parC* region of the plasmid. Plasmid segregation is then driven by ParM spindles, which are stabilised by the binding of a ParR–*parC* plasmid complex at each end of the filament. The now-stabilised ParM spindle then extends via the incorporation of additional ParM filaments to the spindle bundle, pushing the two plasmids in opposite directions towards the poles of the cell (*165*, *166*). The AlfA protein of *Bacillus* is similar in function to ParM in that it forms dynamic filaments required for effective plasmid segregation, however, Alf proteins like AlfA are only distantly related to ParM (*163*).

Finally, although not wholly relevant to this thesis due to its limited distribution, an additional actin homolog, MamK, can be found within magnetotactic bacteria. Much like the other prokaryotic actins, ManK monomers polymerise to form dynamic protein filaments which are used for the alignment and segregation of magnetosomes (organelles filled with iron-rich magnetic particles) (*167*, *168*). The segregation of magnetosomes from the new poles towards midcell following cell division is believed to be achieved by pole-to-midcell treadmilling of MamK filaments (*169*).

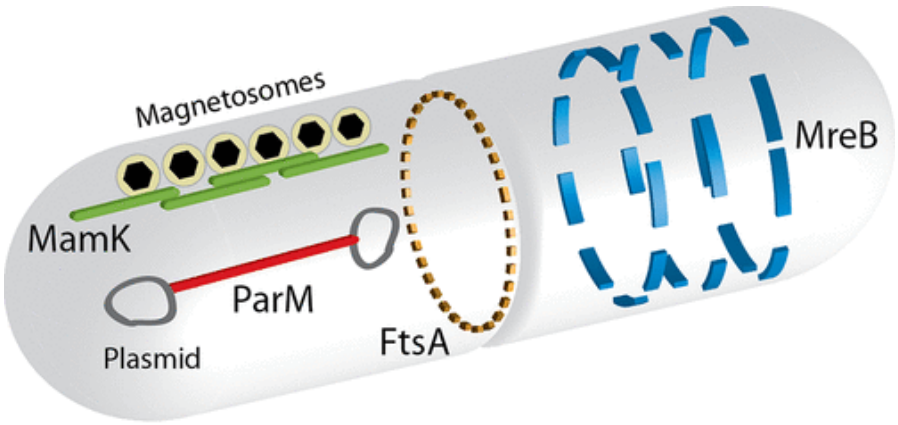


Figure 1.7. Representation of bacterial cytoskeletal actin filaments. Bacterial actin filaments can associate to form a wide variety of dynamic high-order filament structures with diverse biological functions. MreB, present in the majority of non-spherical cells (but not the *Rhizobiaceae*) coordinates local cell wall synthesis around the circumference of the cell, maintaining rod morphology during cell elongation (*107*, *145*, *147*). ParM spindles form part of the ParMRC (Type II) plasmid segregation system. Elongation of ParM spindles following attachment of plasmids to each end results in segregation of the plasmids (*146*, *165*, *166*). FtsA filaments co-polymerise with FtsZ filaments (not shown) at the septum site during cell division, tethering FtsZ filaments to the cell membrane (*126*, *159*). MamK filaments, present in magnetotactic bacteria, are required for the segregation and organisation of magnetosomes within the cell (*167–169*). Figure adapted from Ozyamak *et al.*, 2013 (*148*).

1.2.3 Intermediate filament-like proteins

The intermediate filament-like (IF-like) proteins, characterised by the presence of a coiled-coil structural motif, are the final component of the bacterial cytoskeleton which have a eukaryotic equivalent (*102*, *170*). While prokaryotic tubulins and actins are considered homologs of their eukaryotic counterparts, intermediate filament-like proteins are instead considered relatives of the eukaryotic intermediate filaments, and hence the term IF-like has been adopted to describe these prokaryotic cytoskeletal proteins. Unlike actin and tubulin filaments, IF-like filaments are not highly dynamic in nature, and instead typically perform scaffolding, structural or hub-like organisational roles within the cell (*170–172*). The function of eukaryotic intermediate filament proteins is conserved not through their amino acid sequence, but instead by a common domain structure comprising of N- and C-terminal disordered head and tail domains, and a central α-helical coiled-coil rod domain (*173*, *174*). Although their structure is not as highly conserved, bacterial IF-like proteins share a similar common structure to eukaryotic intermediate filaments,

but often exhibit more variation (in both length and amino acid sequence) within their central coiled-coil rod domains (*170*, *171*, *175*). Two or more α-helical IF-like monomers can form parallel or antiparallel associations, twisting around each other to form a coiled-coil subunit which can then associate into larger filament structures (*174*, *176*). The key to the formation of these coiled-coil subunits is the presence of a hydrophobic 'seam' within the α-helical rod domain which allows the α-helices of individual monomers to associate with one another. To generate this seam, the amino acid sequence of canonical coiled-coil domains, although not highly conserved, contain a heptad-based repeat of hydrophobic residues (a repeating pattern of seven amino acids containing two hydrophobic residues) (Figure 1.8A) (*174*, *176*, *177*).

The first cytoskeletal intermediate filament-like protein to be identified within prokaryotes was the *Caulobacter crescentus* protein crescentin (CreS), which controls the slight curvature of these crescent-shaped cells (*108*). Similar to eukaryotic intermediate filaments, CreS has disordered head (N-terminal) and tail (C-terminal) regions, and coiled-coil central rod domain which allows for the formation of coiled-coil filaments (*178*). Self-assembled CreS filaments localise close to the cell membrane along the concave edge of the cell, and are believed to be anchored to the membrane by the actin homolog MreB (*108*, *172*). It is proposed that crescentin controls cell curvature by exerting a compressive force on the cell wall (*179*). This force is hypothesized to reduce strain on the peptide bonds within the cell wall, making the insertion of new peptidoglycan less favourable. Consequently, cell wall synthesis is favoured along the edge of the cell lacking CreS, resulting in bending (and hence curvature) of the cell as it grows (*172*, *179*, *180*).

In addition to canonical intermediate filament-like proteins, the bacterial cytoskeletal network also contains a broader class of IF-like proteins possessing a non-canonical coiled-coil repeat structure within their rod domain (*171*, *175*). Well-characterised examples of this group of proteins are currently limited to those identified within the Gram-positive phylum Actinobacteria, with Scy and FilP the most well documented (*171*, *181*, *182*). The *Streptomyces* cytoskeletal protein Scy possesses both a short N-terminal canonical, coiled-coil domain and a long C-terminal non-canonical coiled-coil rod domain. Unlike canonical coiled-coil proteins like CreS and the eukaryotic intermediate filaments, which have a heptad-based repeat of hydrophobic residues, Scy's non-canonical coiled-coil rod domain has an unusual repeat of 51 residues containing 14 roughly equally-spaced hydrophobic residues which form a shallower

hydrophobic seam (*175*, *182*) (Figure 1.8B). Scy filaments localise to the tip of *Streptomyces* hyphae during hyphal growth, and are required for the coordination of polar hyphal growth and *de novo* tip formation during hyphal branching (*182*). These filaments form a molecular scaffold that can interact with and organise numerous proteins at the hyphal tip, providing support for an organisational hub called the tip-organising centre (TIPOC), sometimes referred to as a polarisome (a term typically used for yeast and other fungi) (*182*, *183*). Another component of the *Streptomyces* TIPOC is the IF-like protein FilP, which interacts directly with Scy (*182*, *184*) However, less is known about the role of FilP filaments beyond their proposed involvement in the structural regulation of the TIPOC, and providing structural support for the growing hyphal tip (*171*, *181*, *184*).

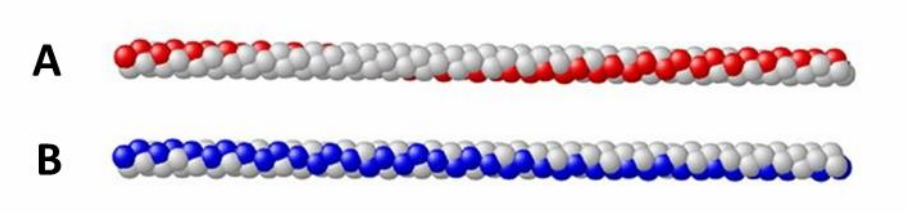


Figure 1.8. Hydrophobic seams of canonical and non-canonical coiled-coil domains. (A) In canonical coiled-coil domains, the α-helix exhibits a left-handed hydrophobic seam (red) resulting from a heptad-based repeat of hydrophobic amino acids. (B) Within Scy, a non-canonical 51-residue repeat results in the formation of a shallow, right-handed hydrophobic seam (blue). Figure modified from Holmes *et al.*, 2013 (*175*, *182*).

1.2.4 The ParA/MinD family - Walker A cytoskeletal ATPases

The final group of prokaryotic cytoskeletal proteins is the ParA/MinD family, also referred to as the Walker A cytoskeletal ATPases (WACAs). Unlike prokaryotic actin, tubulin and the intermediate filament-like proteins, this family of cytoskeletal proteins is unique to the prokaryotes, with no eukaryotic cytoskeletal equivalent (*185*, *186*). The WACAs belong to the AAA+ superfamily of proteins, contain a P-loop NTPase domain, and get their name from the Walker A homology sequences within their AAA+ (ATPases Associated with diverse cellular Activities) domain (*187–189*). The Walker A cytoskeletal ATPases are involved in the spatial organisation of other cellular components, and are widely distributed among prokaryotic species (*185*, *188*, *189*). Several bacterial WACA proteins have been characterised to date, many of which are functional homologs of either ParA or MinD, hence this family of proteins is often referred to simply as the ParA/MinD family. An important feature of the ParA/MinD proteins is

their requirement to function as homodimers. Homodimerization of ParA/MinD ATPases brings together two half-sites (one from each monomer) to form a functional ATPase domain (*185*, *187*). Importantly, members of this family of proteins also function in combination with a specific binding partner (e.g. ParB for ParA, and MinE for MinD), which acts as an ATPase-activator protein upon interaction with its respective WACA (*190*, *191*).

The ATPase MinD forms part of a dynamic protein system referred to as the MinCDE system, which predominantly functions as a spatial regulator of the Z ring, and thus determinant of the site of cell division (*192*, *193*). While this system was first identified within *E. coli*, it is now known that MinCDE systems are present in almost all Gram-negative rods (*185*). The MinCDE system of *E. coli* remains one of the most well-characterised examples of this system. Within *E. coli*, MinD and its activator protein MinE both oscillate between the poles of the cell. ATP-bound MinD dimers bind to the cell membrane (via a C-terminal helix) and are able to recruit MinC, while MinD-ADP monomers (generated following ATP hydrolysis triggered by the binding of MinE to MinD) diffuse freely in the cytoplasm and cannot recruit MinC (*194*, *195*). The MinC protein therefore 'follows' the MinD/MinE oscillation pattern, generating a MinC concentration gradient across the cell which is greatest at the poles, and weakest at the midcell (*192*). The presence of MinC inhibits FtsZ polymerization, hence the established gradient of MinC across the cell restricts Z ring formation to the midcell (*193*, *196*). This oscillatory MinCDE system can be found within the *Rhizobiaceae*, however the precise contribution of this system in this family is not entirely clear and possible redundancy of the MinCDE system within the *Rhizobiaceae* has been proposed (*131*, *197*, *198*). Within *Agrobacterium tumefaciens*, defects in cell morphology and FtsZ-positioning are less severe in cells with deletions of the entire MinCDE system than in those with individual deletions of each component of this system, with MinE knockout strains exhibiting the most pronounced phenotypes including greater cell length variation and FtsZ misplacement (*197*). Interestingly, in *Sinorhizobium meliloti*, deletion of MinC, MinD and the entire MinCDE system does not impact the growth or morphology of free-living cells. However, cells lacking MinE exhibit free-living morphological defects (i.e. swollen and branched cells), and reduced symbiotic nitrogen fixation within alfalfa nodules (*198*).

Within species which lack MinD such as *Caulobacter crescentus*, another member of the WACA family called MipZ facilitates the regulation of the Z ring (*199*). MipZ, much like MinC, inhibits assembly of the Z ring. This protein follows chromosomes as they segregate to the poles (via

interaction with ParA's binding partner ParB), allowing FtsZ to assemble in middle of the cell (*199*, *200*). Although widespread amongst Gram-negative rods, the oscillatory mechanism of the MinCDE system is not conserved in Gram-positive rods. For example, *B. subtilis* uses a static MinCD system which lacks MinE to localise FtsZ to the midcell (*201*). Furthermore, in species which lack ParA-driven Type I segregation systems such as *E. coli*, the MinCDE system has also been proposed to be required for chromosome segregation in addition to Z ring positioning (*202*).

While MinD proteins facilitate spatial regulation of other proteins within the cell, ParA proteins form part of a Type I segregation system referred to as the ParABS system (*parABS* locus), and are required for the segregation of chromosomal and other low copy number plasmids to the cell poles during division (*162*, *185*). The ParABS segregation system is a tripartite system comprised of ParA, the DNA-binding (and ATPase-activator) protein ParB, and centromere-like *parS* sequences which are present on these plasmids (*162*, *203*). Similar to the Type II ParMRC segregation system (Section 1.2.2), ParA is able to associate with plasmid and chromosomal DNA via direct interactions with ParB, which binds to the *parS* sequence to form ParB-*parS* nucleoprotein complexes (*203–205*). However, the exact mechanism of DNA segregation by ParA proteins is not completely understood. Initially, it was proposed that ParA forms helical or linear spindle-like filaments which pull DNA towards the poles of the cell (*206–209*). ParA subunits were first believed to polymerise into spindle-like filaments along the cell length which capture ParB-*parS* nucleoprotein complexes. It was thought that the binding of ParB to ParA triggered ATP hydrolysis and subsequent depolymerisation of the ParA filament, pulling the ParB–*parS* nucleoprotein complexes along the retracting filament towards the cell pole (*206*, *208–210*). However, it is now believed that a diffusion-ratchet (Brownian ratchet) model may best explain the mechanism of plasmid segregation by ParA (211–213). In this model, depicted in Figure 1.9, ParA monomers do not form filaments, but instead generate ATP-bound homodimers which bind non-specifically to DNA throughout the nucleoid (*214*, *215*). ParB-*parS* nucleoprotein complexes bind to ParA-ATP homodimers, thereby stimulating ATPase activity and the dissociation of ParA-ADP monomers (which no longer bind DNA) from the nucleoid (*216*). The ParB-*parS* complex then rebinds to a nearby ParA homodimer further along the nucleoid. Repeated binding of ParB-*parS* complexes to proximal ParA homodimers leaves behind a region of ParA-ADP monomers incapable of binding ParB-*parS* complexes, resulting in directional movement of the plasmid towards the cell pole (*211–213*).

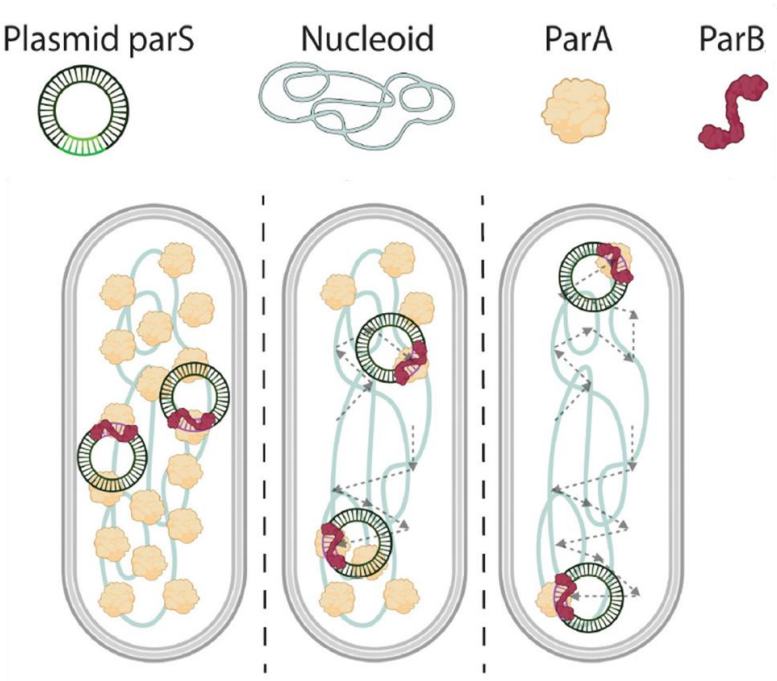


Figure 1.9. Plasmid segregation by the Type I ParABS segregation system. ParA-ATP homodimers (peach) bind non-specifically to DNA throughout the nucleoid (*214*, *215*). These homodimers are capable of binding ParB (red) associated with the *parS* sequences (green) on the plasmid. Binding of ParB to ParA triggers ATP hydrolysis, and subsequent release of ParA-ADP monomers from the nucleoid (not shown) (*216*). ParB-*parS* complexes then bind to nearby ParA nucleoid-bound homodimers. Repeated binding and release of ParB-*parS* to ParA homodimers throughout the nucleoid results in net movement of plasmids towards the pole (arrows) (*211–213*). Figure modified from Gogou *et al*., 2021 (*217*).

In species with multipartite genomes such as those within the genera *Rhizobium*, *Sinorhizobium*, *Mesorhizobium*, *Bradyrhizobium*, and *Agrobacterium* (*218*, *219*), the ParABS (Type I) segregation system is only required for the segregation of the primary chromosomal plasmid. Instead of the ParABS system, secondary chromosomes and additional plasmids within multipartite genomes encode their own replication and partitioning system referred to as the *repABC* locus, similar to the *parABS* locus described above (*220*). In this system, the RepA and RepB proteins are homologous to ParA and ParB respectively (*220–222*). However, *parS* and *repC* are not similar in function, with *repC* required for plasmid replication instead of partitioning. Not only does *repC* encode a replication initiator protein (RepC), but the *repC* gene sequence also contains the origin of replication (oriV) of the plasmid (*220*, *223*, *224*). Furthermore, unlike ParB of the ParABS system which binds to *parS* sequences, the RepB protein of the RepABC system does not bind to *parC* but instead binds to *parS* sites elsewhere on the plasmid (*225*). Although also a member of the rhizobia, forming symbioses with tropical legumes of the genus *Sesbania*, the model

rhizobial species *Azorhizobium caulinodans* ORS571 possesses a genome comprised of only a single chromosomal plasmid (*226*). Within this species, ParA is not essential for cell viability, but the loss of ParA has been found to result in polyploid cells with variations in nucleoid size, likely as a direct result of severe defects in chromosome partitioning (*227*).

1.3 Growth mechanisms in rod-shaped bacteria

The bacterial cell wall, which is primarily composed of peptidoglycan (PG) (a mesh of linear glycan strands crosslinked by peptide bridges), is essential for the maintenance of turgor pressure and cell morphology in bacteria (*228*, *229*). The chemical composition of peptidoglycan is typically conserved across the majority of species, but the thickness of this layer varies, with Gram-negative species having only a single monolayer of peptidoglycan within their cell wall, and Gram-positive species possessing several concentric layers of peptidoglycan (*229*, *230*). In order to undergo successful proliferation and avoid cell lysis, bacteria must be able to modify their cell wall whilst simultaneously maintaining its integrity. To achieve this, careful spatial and temporal coordination of cell wall remodelling is required, a process which is especially important for Gram-negative bacteria which have only a single, thin layer of PG within their cell walls (*231*, *232*). Within coccoid bacteria, cell wall synthesis and division are coupled at the site of septum formation (*233*). However, to maintain cylindrical morphology, rod-shaped bacteria add a distinct elongation phase between each round of cell division. Within these bacteria, one of two elongation mechanisms are adopted: localisation of growth and peptidoglycan synthesis to the cell poles (polar growth), or more commonly, patchy or helical deposition of new cell wall throughout the cylindrical region of the cell (dispersed growth, sometimes referred to as lateral growth) (*94*, *151*, *234*).

1.3.1 Lateral (Dispersed) elongation

As touched upon throughout Section 1.2, bacterial growth, morphology and division have been extensively studied in the model rod species *Escherichia coli* (Gram-negative) and *Bacillus subtilis* (Gram-positive), both of which exhibit dispersed insertion of new cell wall along the cell length (*151*). Hence, our knowledge of the proteins which coordinate this mechanism of bacterial growth is relatively comprehensive (for up-to-date reviews see refs. (*232*, *235–237*)). Dispersed

elongation occurs in two stages. Firstly, localised patches of cell wall synthesis occur along the cylindrical region of cell, leading to cell elongation. Then, once the cell has reached the appropriate length, septum formation (i.e. cell wall synthesis) at the midcell divides the two daughter cells (*234*). This approach to cell elongation is dependent upon a large membrane-associated protein complex called the elongasome, the main structural (cytoskeletal) component of which is MreB (Figure 1.10) (*237*). Cytoplasmic MreB filaments interact with the bitopic membrane protein MreC, which also interacts with the integral membrane protein MreD, forming the membrane-bound MreBCD complex (*238*). Peptidoglycan modification during cell wall elongation and septum formation requires glycosyltransferases (GTases) to generate glycan polymers, and transpeptidases (TPases) to crosslink these new polymers into the existing PG mesh (*235*). At the core of the elongasome is a protein complex composed of the membrane-associated proteins RodA (a Shape, Elongation, Division and Sporulation (SEDS) glycosyltransferase) and PBP2 (a Class B Penicillin Binding Protein (PBP) transpeptidase) (*239*). MreC and MreD are believed to moderate the interaction between RodA and PBP2 (*238*, *240*). Together with the transmembrane protein RodZ, these proteins (RodZ, RodA, PBP2, and MreBCD) form a supramolecular complex named the Rod complex (*241*, *242*). Importantly, PBP2 works in cooperation with the GTase-TPase PBP1A. PBP2 stimulates the glycosyltransferase activity of PBP1A, while PBP1A activates the peptidoglycan transpeptidase activity of PBP2 (*243*). Furthermore, PBP1A is also regulated by the outer membrane lipoprotein activator LpoA (*244*). Also associated with the elongasome are a set of enzymes required for PG precursor synthesis (MurA-F, MraY, and MurG), and PBP5 (a carboxypeptidase Class C PBP) which together facilitate localised PG synthesis (*245*, *246*).

Once the length of a mature cell has been reached, cell division is then initiated by targeting a large septum-specific protein complex termed the divisome to the division site (Figure 1.10) (*128*, *129*, *236*). In contrast to the elongasome where actin-like MreB filaments provide scaffolding, the main structural element of the divisome is the tubulin homolog FtsZ (i.e. the Z ring) (*161*). Within *E. coli*, the core divisome proteins are recruited to the midcell in two stages (*247*, *248*). Some of the first divisome proteins to arrive at the midcell include FtsZ, the actin homolog FtsA (which guides and tethers FtsZ to the cell membrane), the membrane-anchored ZipA and ZapA (which stabilise FtsZ filaments), and the FtsEX membrane complex (which regulates FtsA) (*161*, *249–251*). Stabilisation of the Z ring is then followed by sequential recruitment of the late divisome proteins FtsK, FtsBLQ, FtsW, PBP3 (also known as FtsI) and FtsN to the divisome (*247*, *248*, *252*). Along with the GTase-TPase PBP1B, FtsW (a GTase closely

related to the elongasome protein RodA) and PBP3 (TPase) provide the peptidoglycan synthesis activity during septum synthesis (*253*, *254*). Similar to LpoA and PBP1A of the elongasome, an outer membrane lipoprotein called LpoB is required for the activation of PBP1B (*243*). Also present within the divisome are the septal-specific amidases AmiB and AmiC, which facilitate PG remodelling through the breaking of amide bonds between PG strands (*255*). Although the majority of proteins within the divisome and elongasome differ, a handful of proteins are found within both complexes including MurA-F, MraY, MurG, and PBP5 (*245*, *246*).

Of the late divisome proteins, FtsN is the last to be recruited to the divisome and its arrival triggers the onset of cytokinesis, potentially by simultaneously activating both the early and late divisome proteins (*256*, *257*). Although the precise mechanism is not yet known, it has been proposed that the FtsQLB complex is also important for the regulation and progression of cell division (*258*, *259*). In the majority of species which grow by dispersed growth, including *E. coli*, an intermediate phase between lateral elongation and septum formation termed pre-septal elongation occurs. During this phase, the early divisome proteins coordinate PBP3-independent peptidoglycan synthesis at the future cell division site (*237*, *260*). Within Gram-negative bacteria, the final stage of cell division is the invagination of the outer membrane (*261*). A large set of proteins are required for this process, including the Tol-Pal system which consists of the inner membrane motor complex TolQRA, the periplasmic protein TolB, and the PG-binding outer membrane lipoprotein Pal (*262*, *263*). However, whether or not the divisome coordinates the recruitment of the Tol-Pal system to constriction sites is not yet fully-understood (*264*, *265*).

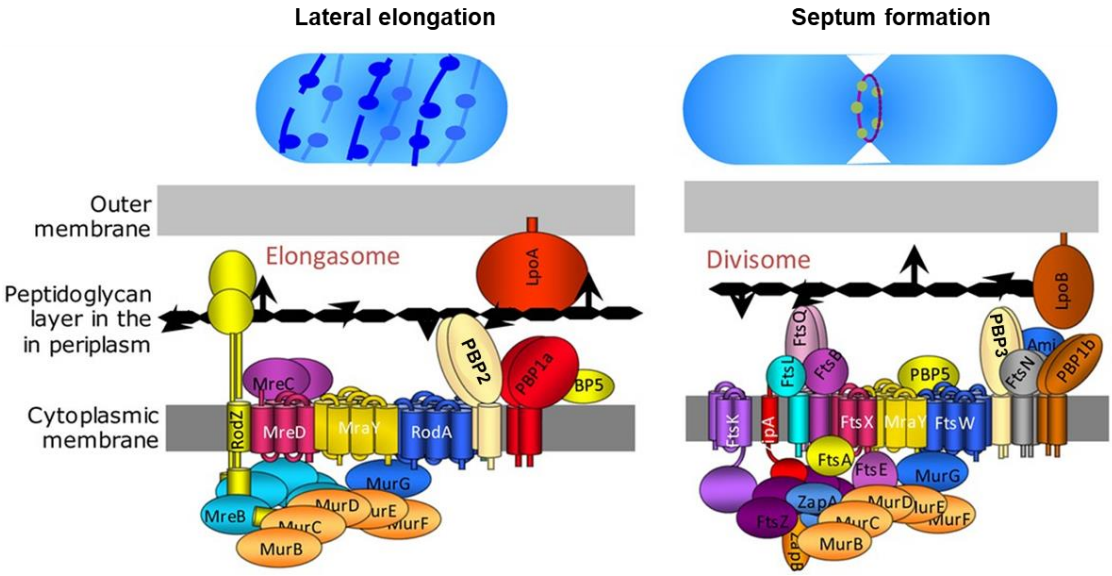


Figure 1.10. The elongasome and divisome protein complexes of *E. coli*. Lateral cell wall elongation (left) within rod-shaped bacteria is coordinated by a large protein complex called the elongasome. The key scaffolding element of this complex is filaments of MreB (light blue), which circumferentially treadmill around the cylindrical region of the cell (*231*, *235*). Other core components of this complex include RodZ (yellow), RodA (dark blue), PBP2 (cream), and the MreCD proteins (purple and pink respectively), which collectively form the Rod complex (*241*). Upon initiation of cell division (right), septal peptidoglycan (PG) synthesis is coordinated by the divisome protein complex. During the formation of the Z ring at the midcell (FtsZ filaments, dark purple), an assortment of core divisome proteins are recruited to the midcell in two stages. First, the early divisome proteins FtsA, ZipA, ZapA, FtsE and FtsX, which collectively stabilise the Z ring, are recruited to the division site. Then, sequential recruitment of the late divisome proteins FtsK, FtsBLQ, FtsW, PBP3 (FtsI) and FtsN occurs (*161*, *236*). Several PG-synthesis proteins are common to both the divisome and elongasome, including MurA-F, MraY, MurG, and PBP5 (*129*). Figure modified from Hugonnet *et al.*, 2016 (*266*).

1.3.2 Polar elongation

In contrast to dispersed growth where cell wall elongation occurs evenly throughout the cell length, polar growth is the result of peptidoglycan synthesis specifically localised at the poles of the cell. Polar growth can be observed within a range of bacteria, but it is particularly prevalent in the Gram-positive order Actinomycetales, and the Gram-negative order Hyphomicrobiales (*131*). Interestingly, polar-growing Actinomycetales such as *Mycobacterium* and *Corynebacterium* exhibit growth from both cell poles (bipolar growth), while members of the Hyphomicrobiales (e.g. the *Rhizobiaceae* and *Brucellaceae*) grow from only a single pole (unipolar growth) (Figure 1.11) (*152*, *267*, *268*). Within the unipolar-growing Hyphomicrobiales, the synthesis of new peptidoglycan occurs across the entire curvature of the growth pole (*152*, *269*). Conversely, bipolar cell elongation within *Mycobacterium* is the result of subpolar incorporation of new peptidoglycan (*270*, *271*). However, subpolar modification of the cell wall is not (or possibly is yet to be) observed within *Corynebacterium* (*271*). Further differences between these two mechanisms of polar growth include the nature of the poles formed following cell division. Newly-formed *Agrobacterium* growth poles (formed at the septal site) become active growth poles shortly after separation of the daughter cells, whilst in *Mycobacterium*, septal PG synthesis generates inert cell poles which do not become sites of rapid cell elongation until subsequent cell cycles (Figure 1.11) (*152*, *269*, *272*)**.** Furthermore, *Agrobacterium* growth poles are slightly tapered, and newly-formed growth poles of daughter cells emerge narrower than the parental pole (i.e. the old pole) (*127*, *152*, *273*, *274*).

Polar growth is a form of asymmetric cell growth, yielding two daughter cells which are physiologically (and in some cases morphologically) distinct, wherein one daughter cell receives the majority of newly-synthesised PG, cell wall components and outer membrane proteins (OMPs) (*275*). In the case of *Agrobacterium*, unipolar growth also typically yields a daughter cell which is slightly smaller than the parent cell (*152*). The generation of heterogeneous daughter cells within a population is thought to provide certain advantages such as antibiotic resistance (*Mycobacterium*), improved population health by avoiding 'ageing' of the population, and it has been proposed that within the Hyphomicrobiales differing allocations of OMPs and lipopolysaccharides between daughter cells permits cell survival during host–microbe interactions by evading host defence responses (*152*, *275*, *276*).

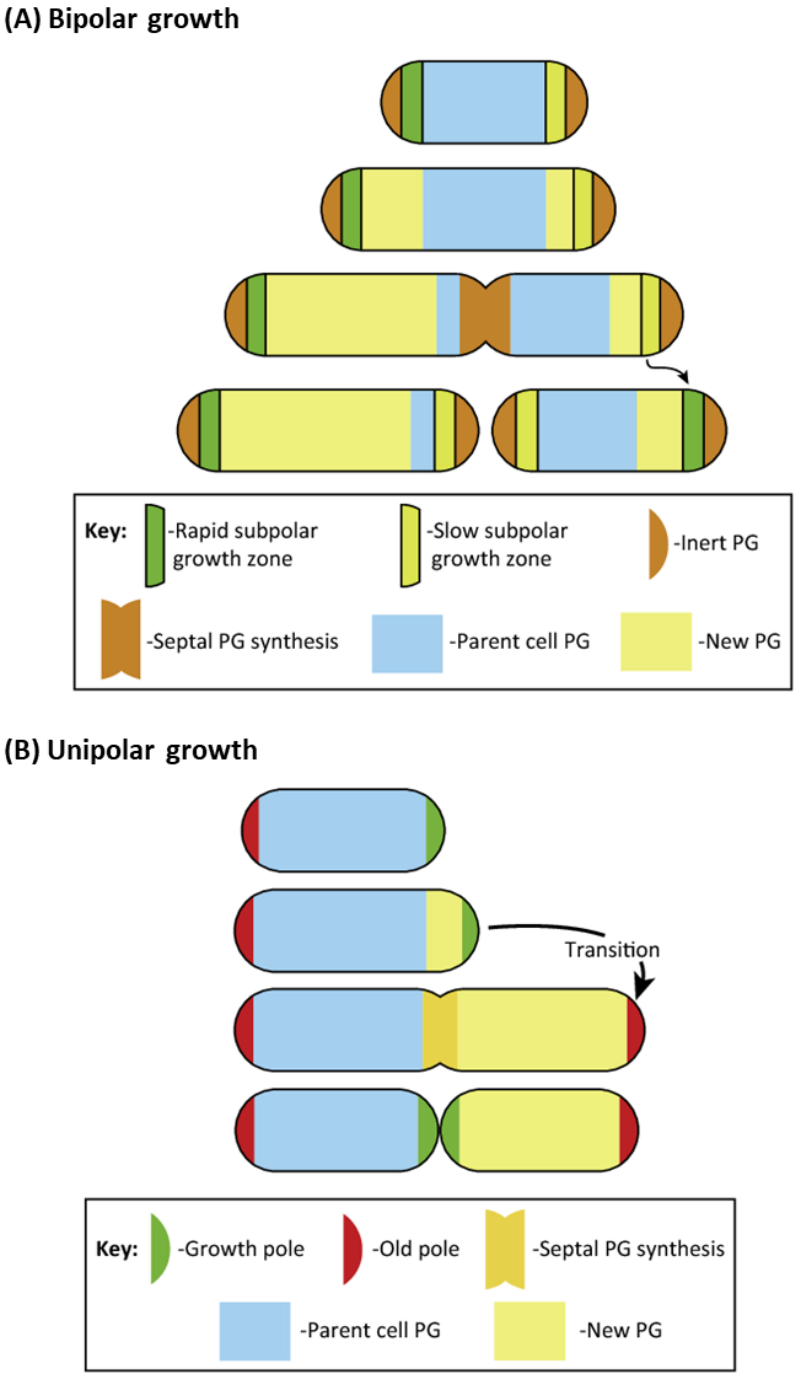


Figure 1.11. Polar growth mechanisms in rod-shaped bacteria. (A) Within species which undergo bipolar growth (e.g. *Mycobacterium*), subpolar synthesis of peptidoglycan (PG) drives cell elongation. Rapid PG synthesis (green band) occurs below the older pole, while slower PG synthesis (yellow band) occurs below the newer pole. During division, septal PG synthesis creates poles comprising of inert PG (orange), which do not become rapidly-growing poles until the next cell cycle (*267*). (B) In species which exhibit unipolar growth (e.g. *Agrobacterium, Rhizobium*), cell elongation occurs from a single growth pole (green). During cell division, the growth pole transitions into a non-growing old pole (red), and septal PG synthesis at the midcell occurs. Following cell division, new growth poles are formed at the division site, and polar PG synthesis begins at the new poles of both daughter cells (*152*). Figure modified from Cameron *et al.*, 2015 (*131*).

Despite its biological significance, relatively little is known about the proteins which coordinate unipolar growth within the Hyphomicrobiales. Genomic analysis has highlighted that members of this order lack homologs of a number of the canonical scaffolding and elongasome proteins required for localised cell wall synthesis during dispersed growth including MreBCD, RodA, RodZ and PBP2 (i.e. the Rod complex) (*131*, *151*, *152*). *Corynebacterium* and *Mycobacterium* also lack several components of the Rod complex, yet these species have retained the RodA-PBP2 PG synthesis complex (*151*, *268*). It has been proposed that instead of MreB, the coiled-coil protein DivIVA generates a cytoskeletal-like scaffold which directs polar cell wall synthesis within *Corynebacterium* and *Mycobacterium* (*267*, *270*, *277*). However, Gram-negative bacteria lack DivIVA, leading to the suggestion that unipolar growth in the Hyphomicrobiales is coordinated by a non-canonical suite of highly conserved proteins (*102*, *131*, *152*). Recent research within the *Rhizobiaceae* and *Brucellaceae* (predominantly *Agrobacterium tumefaciens*, *Sinorhizobium meliloti*, and *Brucella abortus*) has led to the identification of a generally poorly-characterised group of proteins proposed to be required for unipolar growth within this order (Figure 1.12) (*276*). For the remainder of this section, those proteins characterised within members of the *Rhizobiaceae* specifically will be the primary focus.

Structural support for the polar growth complex may be provided by the growth pole ring protein (GPR), which has been found to be required for unipolar growth within *A. tumefaciens* (*278*). GPR is a large membrane-anchored cytoplasmic protein, believed to form a hexameric ring structure at the growth pole which serves as an organisation centre for polar growth. Although it has no predicted enzymatic function, apolipoprotein domains throughout its length may sequester lipids during polar growth (*278*, *279*). Within *S. meliloti*, a less well-characterised homolog of GRP called RgsE has also been identified as a cytoplasmic, membrane anchored protein required for polar growth (*280*, *281*). GPR has been shown to colocalise at the growth pole with the pole-organising protein PopZ, which has itself been linked to the coordination of unipolar growth within *A. tumefaciens* (*273*, *278*, *282*). Both GPR and PopZ have been discussed in further detail within Section 1.4.3. Along with RgsE, protein-protein interaction studies have identified a large, but not yet fully characterised, group of *S. meliloti* polar growth proteins termed the Rgs (rhizobial growth and septation) proteins (RgsA, RgsB, RgsC, RgsD, RgsF, RgsG, RgsH, RgsM, RgsP, RgsS and of course RgsE) (*280*, *281*, *283*). Of these proteins, RgsA, RgsM, RgsP and RgsS are of particular interest. Pull-down assays have indicated that RgsE directly interacts with RgsS and RgsA (*280*)**.** RgsS is an FtsN-like transmembrane protein which contains a conserved SPOR (sporulation-related repeat) domain known to bind amidase-processed

peptidoglycan at the septum site (*281*, *284*). Therefore, it has been proposed that RgsE (which exhibits unipolar localisation at the growth pole) may be recruited to the septum site by RgsS, initiating the assembly of new polar protein complexes at the newly-defined growth poles (*281*). Although the precise function of RgsA is not clear given its lack of homology to other proteins of known function, RgsA has been found to be essential for growth and hence is proposed to be required for polar PG synthesis, possibly forming a complex with RgsP and RgsM (*280*). RgsP, a cyclic di-GMP phosphodiesterase important for the regulation of PG synthesis, interacts with the membrane-anchored peptidoglycan endopeptidase RgsM (*280*, *283*, *285*, *286*). Together, this 'RgsAMP' complex may regulate polar PG synthesis.

During dispersed elongation within *E. coli*, lateral PG synthesis is driven by the RodA-PBP2 complex, with PBP1A having a supporting role in this complex (*239*, *243*). However, within the Hyphomicrobiales (which lack RodA-PBP2), PBP1A is thought to serve as the primary PG-synthesis enzyme during polar growth. Within *Agrobacterium*, PBP1A has been shown to be essential for the addition of new PG to the cell wall during polar elongation (*287*). Additionally, a possible alternative mode of cell wall synthesis by L,D-transpeptidases has also been proposed (*287*). The genomes of Hyphomicrobiales have been found to be enriched in genes encoding L,D-transpeptidase (LDT) enzymes (which crosslink PG), some of which have been shown to localise to the cell poles (*274*, *287*, *288*). The peptidoglycan meshwork within the cell walls of Hyphomicrobiales such as *Agrobacterium* contains a substantially greater proportion of L,D crosslinks than PG within *E. coli* cell walls, suggesting that LDTs may play an important role during cell wall elongation within this order of bacteria (*152*).

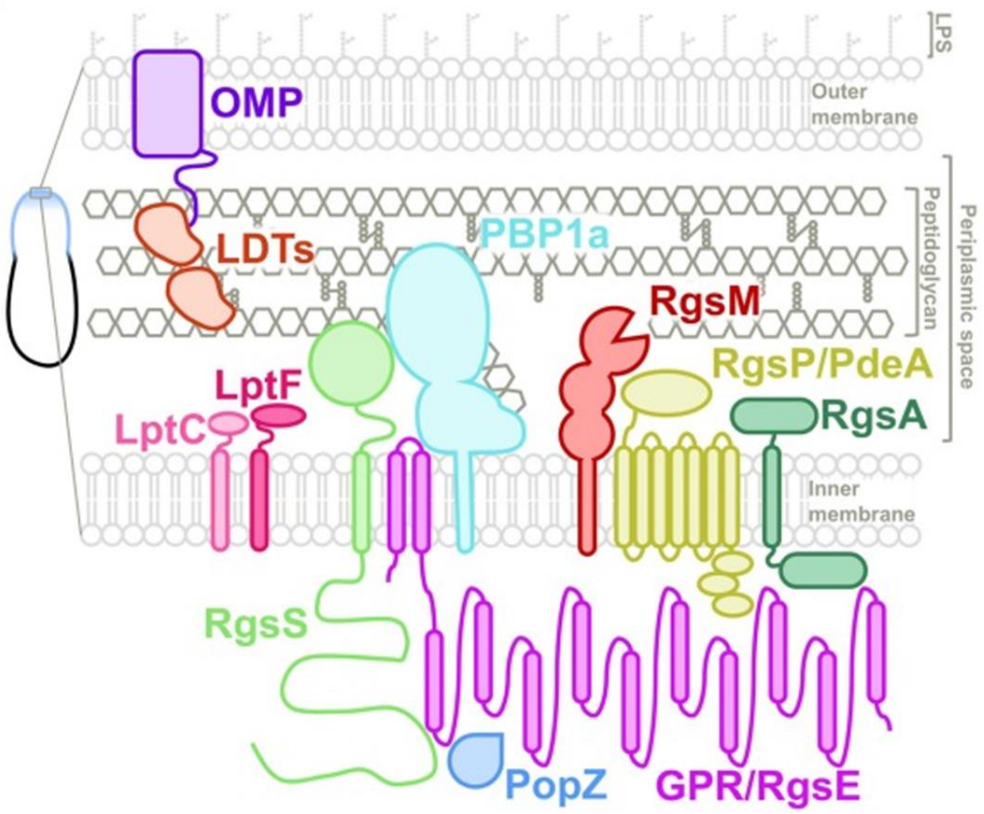


Figure 1.12. Proposed model of the polar growth complex within Hyphomicrobiales. In contrast to the well-characterised protein complex responsible for peptidoglycan (PG) synthesis during dispersed cell elongation (i.e. the elongasome), knowledge of the protein complex which coordinates unipolar growth within Gram-negative rods is limited. In this model, proposed by Amstutz *et al*. (2024) (*276*), proteins previously identified as essential for polar growth within various Hyphomicrobiales species were combined to generate a hypothetical model of an elongasome-like polar growth complex. The proposed scaffolding protein GPR/RgsE (magenta), the pole-organising protein PopZ (blue), the transmembrane receptor protein RgsP/PdeA (yellow), and the SPOR (sporulation-related repeat) domain protein RgsS (light green) provide structural support and temporal regulation of polar growth (*273*, *278*, *281*, *283*). The peptidoglycan (PG) endopeptidase RgsM (red), the GTase-TPase PBP1a (light blue) and L,D-transpeptidase (LDT) enzymes (orange) are responsible for PG synthesis and modification (*274*, *285*, *287*). Here, Amstutz *et al*. also propose that unipolar growth coordinates the synthesis of peptidoglycan with outer membrane proteins (OMPs, purple) and lipopolysaccharide insertion via the activities of LptC and LptF (pink). Also included in this model is the transmembrane protein RgsA (dark green), which despite its unknown function has been found to be essential for unipolar growth (*280*).

Canonical cell division proteins have also been found to regulate polar growth within the Hyphomicrobiales. During unipolar growth, cell division and cell elongation are at times closely linked, with new growth poles rapidly established at the septum site of each new cell following cell division (*152*, *269*). Within *Agrobacterium*, three Fts proteins (FtsZ, FtsA and FtsW) have been found to mediate the transition of peptidoglycan synthesis at the growth pole to the midcell during the initiation of cell division (*127*, *130*, *152*, *274*). Both $FtsZ_{At}$ (the primary FtsZ homolog) and FtsA localise to the growth pole during unipolar growth, but transition to the septum site (i.e. midcell) prior to cell division (*127*, *130*, *152*). In contrast to the non-sequential recruitment of early divisome proteins within *E. coli*, $FtsZ_{At}$ arrives at the *Agrobacterium* midcell substantially earlier than FtsA, indicating that $FtsZ_{AT}$ may be able to initiate Z ring formation in the absence of FtsA and that the correct placement of $FtsZ_{At}$ is required for the recruitment of FtsA to the midcell (*130*, *274*). Characterisation of $FtsZ_{At}$ has revealed that $FtsZ_{At}$ is required not only for the initiation of cell division, but also for the termination of polar peptidoglycan biosynthesis, coordinating growth pole to old pole transition. In the absence of $FtsZ_{At}$, peptidoglycan synthesis is not redirected from the growth pole to the midcell (*130*). Both FtsA and FtsW are involved in cell division and the regulation of polar growth, with FtsW proposed to mediate septal PG synthesis. However, neither of these proteins are essential for the redirection of growth to the septum site (*130*).

The function of the septal amidase AmiC may also be important for the formation of new growth complexes at the site of cell division. Typically, AmiC facilitates cell division by splitting peptidoglycan (*255*). As touched upon above, amidase-processed PG may serve as a landmark for the recruitment of RgsS to the septum (*281*, *285*). Subsequent recruitment of RgsE by RgsS may then permit the formation of polar growth complexes at the septum site (*280*, *281*). Hence, much like the initiation of cell division upon the arrival of FtsN to the *E. coli* divisome, arrival of RgsS to the Hyphomicrobiales divisome also initiates cell division, and is indispensable during the early stages of cell division (*256*, *257*, *280*). Although the Hyphomicrobiales lack the elongasome-specific penicillin-binding protein PBP2, members of this order do possess homologs of the divisome-specific PBP3. Within *Agrobacterium*, which has two copies of PBP3 (PBP3a and PBP3b), PBP3 has been demonstrated to be the driver of septal PG synthesis during cell division (*274*, *287*). Finally, RgsE also interacts directly with TolQ of the Tol-Pal system, likely promoting the coordination of the later stages of cell division within *S. meliloti* (*280*). Interestingly, *S. meliloti* TolQ and Pal localise to the growth pole during cell elongation and to

the septum within predivisional cells, and colocalise with RgsP, suggesting that the Tol-Pal system may also be involved in coordinating polar growth as well as cell division (*280*, *281*).

1.4 Polarity markers and pole organisers

For polar growth to occur, the proteins responsible for mediating PG synthesis and remodelling must first be localised to the pole(s) of the cell. The recruitment of growth complexes to the correct pole is of particular importance during unipolar growth where growth occurs from only a single cell pole (*289*). Furthermore, the identity of cell poles is not fixed, with newly-formed poles generated following cell division later becoming old poles after subsequent rounds of division (*289*, *290*). Therefore, the ability of a cell to determine and mark polar identity must be dynamic. Often, polar proteins are recruited to the pole via associations with other polarity determinants or polar proteins (*290*, *291*). However, this approach first requires the presence of a polar protein already localised to the pole. There are several ways in which these initial polarity proteins (often referred to as polar landmark proteins) can localise to the cell pole including recognition of intrinsic polar features such as negative membrane curvature, detection of pole-specific elements of the cell envelope e.g. unique phospholipids or newly-synthesised PG, or inheritance from the division site (*291*–*293*). The composition of cell poles can also serve as an additional cue for the localisation of polar proteins. Cell poles are typically devoid of nucleosomes and other densely-packed cellular components, allowing for the self-assembly of large protein structures within these areas (nucleoid occlusion) (*294*, *295*). Ultimately, bacteria use a variety of polar landmark proteins to recruit additional members of their polar protein complexes to the poles, relying on a variety of mechanisms to initially localise these early landmark proteins to the cell pole (*289*–*291*).

1.4.1 The *Streptomyces* TIPOC

In the filamentous actinomycete *Streptomyces* (a common model for polar growth studies), organisation of the hyphal tip involves three proteins (Scy, DivIVA and FilP) which form a polar complex referred to as the tip organising centre (TIPOC) (*182*). Within *Streptomyces*, the coiled-coil protein DivIVA functions as an essential polarity determinant, localising to the hyphal tip (*296*, *297*). Here, DivIVA forms apical assemblies which determine the site of new cell wall growth, possibly by recruiting PG synthesis machinery to the hyphal tip (*296*, *298*). DivIVA is also able to dissociate from these apical assemblies and occasionally localises to hyphal sidewalls, marking future sites of hyphal branching (*296*, *298*). The role of DivIVA in determining the site of new PG synthesis is conserved within other polar-growing Actinobacteria (e.g. *Corynebacterium* and *Mycobacterium*), but seemingly not beyond (*267*, *270*, *277*). Within *B. subtilis*, which grows by dispersed growth, DivIVA is instead required for the positioning of its static MinCD system (which determines the site of cell division), localising to the cell poles via the recognition of negative membrane curvature (*299–301*).

Although DivIVA is essential for marking the site of new hyphal PG synthesis, DivIVA alone is not sufficient to maintain normal polarised growth at the *Streptomyces* hyphal tip (*182*). DivIVA has been found to directly interact with the intermediate filament-like proteins Scy and FilP (*181*, *182*). Scy is a large IF-like protein, with a long (1,227 residue) C-terminal coiled-coil domain (*182*). Much like DivIVA, Scy has been shown to localise to hyphal tips and influence branching patterns and cell wall synthesis, with the loss of Scy resulting in irregular branching and uneven hyphal diameter (*182*, *184*). Interestingly, seemingly normal hyphal tips can occasionally be established in the absence of Scy, possibly as a result of the creation of DivIVA-based polarisomes (*184*). It has been proposed that Scy may therefore coordinate the number and organisation of DivIVA complexes within *Streptomyces* hyphae, thereby controlling the number of tips established (but not directly recruiting cell wall synthesis enzymes to these tips) (*182*). Additionally, the long C-terminal coiled-coil domain of Scy may function as a molecular scaffold for other polar proteins (*182*). Indeed, *in vitro* experiments indicate that Scy interacts with FilP (*182*). FilP (which also interacts with DivIVA) forms apical concentration gradients just below DivIVA within actively growing hyphal tips (*171*, *181*, *184*). In contrast to Scy, which serves as a relatively static polar scaffold, apical assemblies of FilP are highly dynamic (*181*, *184*). The loss of FilP results in delayed germination, slower vegetative growth, and curved or curly (but not branched) hyphae, suggesting that unlike DivIVA and Scy, FilP is not required for determining new sites of PG

synthesis (*171*, *184*). Instead, FilP filaments are thought to provide structural support and rigidity to the growing hyphae (*171*, *181*). Although Scy and FilP interact, the localisation of FilP at the hyphal tip is not directly dependent upon Scy, and FilP may instead be recruited to the tip by DivIVA (*181*, *184*). However, the precise nature of interactions between DivIVA and FilP within the hyphal tip are not entirely clear, as FilP also appears to affect the size and position of DivIVA localisations (*184*).

Within *Streptomyces*, the chromosome partitioning protein ParA exhibits localisation to the tip of vegetative and young aerial hyphae (*183*, *302*). ParA is believed to be recruited to the TIPOC by Scy, possibly through interactions with Scy's C-terminal domain (*183*). Here, Scy is proposed to maintain the polar concentration gradient of ParA by regulating ParA polymerisation (*183*). It is important to note that the above studies support the now-outdated spindle-based mechanism of ParA function, therefore the precise role of Scy in maintaining polar ParA gradients may be up for debate. Despite widespread distribution throughout Gram-positive species, DivIVA is not found within Gram-negative bacteria, and no known homologs of FilP and Scy have been identified or characterised within species beyond the filamentous Actinobacteria (*102*). Hence, polarity determination and pole organisation in Gram-negative rods likely depends upon an entirely different suite of proteins altogether.

1.4.2 Determining polarity in *Caulobacter*

Much of our knowledge of polarity determination within Gram-negative bacteria has come from studies within *Caulobacter crescentus* (*289*, *290*). *C. crescentus* has a unique lifestyle wherein asymmetric cell division gives rise to two morphologically distinct daughter cells: a non-replicative flagellated swarmer cell, and a replicating stalked cell (*303*). Although this species exhibits dispersed growth during cell elongation, localised polar elongation is required during stalk formation (*304*, *305*). Progression of the *C. crescentus* lifecycle is highly regulated and coordinated by several polar determinants and regulatory proteins which ensure correct spatial and temporal organisation of polar organelle development, chromosome replication and segregation, and cell division (*304*, *306*). In this section, the polarity determinants which allow for the distinction between the new and old poles have been discussed. *Caulobacter* polarity is primarily determined by two proteins: TipN and PopZ (*294*, *307–309*). TipN, a transmembrane

protein with a predicted coiled-coil region, functions as a polar landmark protein which marks new poles (*307*, *308*). Here, TipN marks the site of flagellum assembly, prevents stalk formation, and coordinates cell division, with the loss of TipN resulting in elongated cells, misplacement of flagellum, and the formation of cells with a stalk at both poles (*307*, *308*). During cell division (late predivisional stage), TipN relocates to the septum site (in an FtsZ and Tol-Pal-dependent manner), not only ensuring its recruitment to the new poles following division but also permitting the transition of the former new pole to an old pole (*307*, *308*, *310*). Ectopic localisation of TipN following its overexpression results in cell branching and the formation of additional cell poles, indicating that TipN also functions as a landmark protein for the assembly of growth-associated complexes (e.g. PG synthesis machinery) (*308*).

In addition to TipN, *Caulobacter* possesses a second polar landmark protein called PopZ. PopZ is an intrinsically disordered proline-rich protein which self-assembles into an irregular network at the cell pole (*294*, *309*). This network serves as a molecular signalling hub, with the intrinsically disordered regions of PopZ believed to be responsible for recruiting several different proteins including the signalling and regulatory proteins CckA, DivJ and SpmX, which are required for cell cycle progression (*294*, *311*, *312*). One of the primary functions of PopZ is the maintenance of chromosome organisation (*294*, *309*). Within swarmer cells, PopZ forms a unipolar matrix at the old pole. A portion of PopZ then migrates to the opposite pole (termed the new pole) during swarmer-to-stalked cell transition, exhibiting bipolar localisation until cell division occurs (Figure 1.13) (*294*, *309*). Within the dividing cell, PopZ tethers chromosomes to the new pole by directly interacting with ParB-*parS* nucleoprotein complexes at the chromosome origin (*309*, *313*). Despite acting as new and old pole markers respectively, the localisation of TipN or PopZ does not dependent upon the other (*294*, *309*, *314*). Although neither PopZ nor TipN are essential for cell viability, mutants lacking both proteins are not viable (*294*, *315*). Important for the regulation of PopZ is the polar scaffolding protein PodJ, which interacts directly with PopZ and triggers its bipolarisation (*316*). PodJ is a transmembrane protein with a cytoplasmic N-terminus composed of a coiled-coil-rich region and an intrinsically disordered region, and a C-terminal periplasmic domain (*317*, *318*). PodJ was first characterised as a polar landmark protein required for the localisation of the pili assembly and regulatory proteins CpaE and PleC (respectively) to the new pole in early predivisional cells (a function attributed to its C-terminus) (*317*, *318*). However, recent work has demonstrated direct interactions between PodJ's coiled-coil region and PopZ, with PodJ regulating PopZ. PodJ was found to localise to the new cell pole slightly before PopZ, and the loss of PodJ prevented PopZ relocation in around 80% of cells, as well as

impairing the ability of PopZ to tether ParB (*316*). Interestingly, this study also suggested that new cell pole accumulations of PopZ could actually be the result of *de novo* PopZ synthesis as opposed to PopZ relocalisation, however the preexisting relocation model of PopZ's bipolar localisation pattern has been adopted in this thesis.

Within *Caulobacter*, the spatial organisation of cell poles is tightly coupled to chromosome replication and segregation (Figure 1.13) (*319*). In newly formed swarmer cells, chromosomes are initially tethered to the old pole by PopZ, which directly binds ParB proteins within ParB-*parS* nucleoprotein complexes (*309*, *313*). Increased concentrations of ParB (bound to PopZ) at the old pole also generates a concentration gradient of the ParB-interacting protein MipZ, a member of the MinD-like ATPase family (*199*). ParB mediates ATP-dependent dimerization of MipZ, allowing MipZ to bind DNA. Hence, MipZ dimers bind predominantly to the polar regions of the nucleoid near the old pole where ParB is concentrated by PopZ (*200*). At the new pole, TipN interacts directly with ParA and promotes the formation of a ParA concentration gradient which is greatest at the new pole (*206*, *315*). The mechanism behind the formation of ParA concentration gradients by TipN is not too well understood. Previous (and possibly now outdated) hypotheses have suggested that TipN may stabilise ParA structures at the new pole, while others suggest that TipN binds and thereby concentrates ParA, or sequesters dissociated ParA (ParA-ADP) away from the nucleoid, generating a concentration of DNA-binding ParA-ATP dimers near the new pole (*206*, *314*).

Upon pole maturation during swarmer-to-stalked transition, ParB is released from the PopZ matrix and DNA replication is initiated (*313*). During this stage, a proportion of PopZ also accumulates at the new pole, altering the localisation of PopZ from unipolar to bipolar (*294*, *309*). TipN-mediated accumulation of ParA at the new pole has been proposed to promote the accumulation and eventual polymerisation of PopZ at the new pole (*319*). Interestingly, the localisation of PopZ at the new pole has also been proposed to further promote the generation of a ParA concentration gradient at the new pole, facilitating ParA-mediated chromosome segregation (*314*, *315*). PopZ has been shown to directly interact with both ParA and ParB via a short, conserved N-terminal region, however while ParA can enter the PopZ matrix, ParB instead only accumulates on the cytoplasmic edge of the matrix (102, 111). 'Inactive' ParA-ADP monomers released from the nucleoid following ParB-mediated ATP hydrolysis are sequestered by the PopZ matrix, which may also function to promote the regeneration of ParA-ATP (*314*).

Although PopZ and TipN display some functional redundancy when it comes to recruiting ParA to the new pole, with the overexpression of one being able to functionally compensate for a lack of the other, TipN is still likely to be the significant factor in new-pole-specific ParA localisation as in the absence of TipN, ParA is more evenly distributed between bipolar localisations of PopZ (*294*, *314*, *315*). The ParA gradient established by TipN and PopZ allows for the migration of ParB-*parS* complexes (and their associated replicated chromosomes) across the cell towards the new pole (*206*, *315*). Details of the mechanism of chromosome segregation by the ParABS system can be found within Section 1.2.4. PopZ then tethers the ParB-*parS* nucleoprotein complexes upon their arrival at the new pole, anchoring chromosomes to this pole (*319*). As mentioned above, the bipolar distribution of ParB formed during chromosome segregation also induces a bipolar concentration gradient of MipZ. As MipZ inhibits the polymerisation of FtsZ filaments (by stimulating FtsZ GTPase activity), bipolar localisation of MipZ restricts Z ring formation to the midcell where the concentration of MipZ is weakest (*199*, *200*). Formation of the Z ring and subsequent recruitment of the divisome proteins not only initiates cell division, but also recruits TipN to midcell ready for its inheritance at the new poles of daughter cells following cell division (*307*, *308*, *310*).

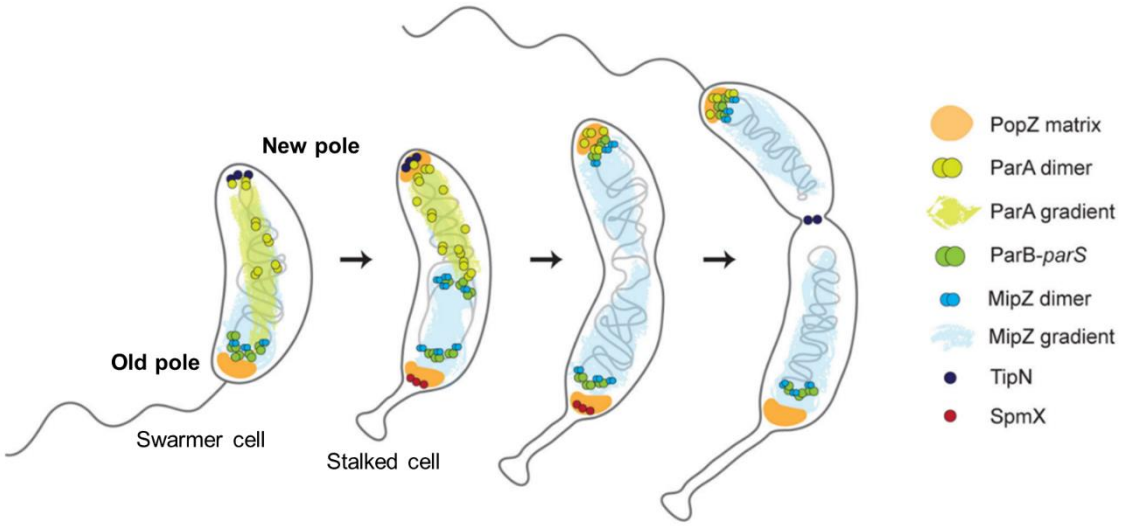


Figure 1.13. Organisation of *Caulobacter* cell poles. In swarmer cells, PopZ (orange) exhibits unipolar localisation, tethering chromosomes (grey lines) to the old pole via direct interactions with the ParB-*parS* nucleoprotein complex (darker green) (*309*). Bound non-specifically to the chromosomes are ParA (light green) and MipZ (blue) dimers. MipZ is concentrated at the old pole by ParB (which is bound to PopZ), while ParA is concentrated at the new pole by the new-pole polarity determinant TipN (black) (*200*, *315*). Following the initiation of chromosome replication, a portion of PopZ relocates to the new pole, while the remaining PopZ recruits SpmX (burgundy) to the old pole, driving stalk formation (*313*). ParB (attached to one of the replicated chromosomes) moves up the ParA concentration gradient towards the new pole, bringing MipZ along with it. The formation of a MipZ concentration gradient which is weakest at the midcell allows for the assembly of FtsZ (the Z ring) and cell division proteins at the division site (not shown) (*200*). TipN is subsequently recruited to the division site by FtsZ, ensuring its inheritance at the new pole following cell division (*307*, *308*). Figure modified from Sundararajan and Goley, 2017 (*320*).

1.4.3 Organising the *Agrobacterium* growth pole

Unlike in *Caulobacter*, the determination of cell polarity and organisation of cell poles within members of the Hyphomicrobiales is generally not well characterised. However, recent studies within the *Rhizobiaceae* member *Agrobacterium tumefaciens* have greatly advanced our understanding of how these important cellular processes are coordinated. Homologs of the *Caulobacter* polarity proteins PopZ and PodJ can also be found within the Hyphomicrobiales, however they often exhibit modified functions within this bacterial order (*273*, *276*, *321*, *322*). Within *A. tumefaciens*, PopZ identifies the new/growth pole (GP), while PodJ identifies the old pole (OP) (*273*). In contrast to *Caulobacter crescentus* PodJ ($PodJ_{Cc}$), the *Agrobacterium* PodJ

protein ($PodJ_{At}$) localises to the old pole early in the cell cycle, but also accumulates at the growth pole later in the cell cycle, exhibiting bipolar localisation (*273*). Colocalisation of PodJ and PopZ at the *Agrobacterium* growth pole is observed during septation when existing GPs transition into non-growing OPs (Figure 1.14) (*273*). Cell division then generates two daughter cells which each have a single focus of PodJ at the old pole and PopZ at the new pole, highlighting that the transition of polar identity requires the loss of PopZ from the previous growth pole (*273*). Deletion of *podJ* results in the formation of ectopic growth poles (leading to cell branching), GPs which fail to transition into OPs and instead continue growing, and elongated cells which do not divide (*322*). In the absence of PodJ, PopZ remains at non-transitioned 'old' poles, indicating that PodJ is required for successful GP to OP transition and the accompanying exclusion/removal of PopZ from the transitioning growth pole (*322*). The observation of active growth at these non-transitioned poles suggests that PodJ ultimately acts as a negative regulator of polar growth in *A. tumefaciens* (*322*). PodJ is also required for the coordination of chromosome segregation and cell division within *Agrobacterium*. During the early stages of polar growth, PodJ plays an essential role in tethering the origins of both the circular and linear *A. tumefaciens* chromosomes to the old pole, ensuring faithful chromosome segregation (*323*, *324*). Loss of PodJ also impacts Z ring formation and stability, resulting in the formation of anucleate minicells, likely as a result of aberrant cell division (*322*).

A homolog of PodJ has also been characterised within *Sinorhizobium*. In *Sinorhizobium meliloti* strain Rm1021, the PodJ gene is split into two adjacent reading frames (*podJ1* and *podJ2*), a gene structure unique to SU47-derived *Sinorhizobium* strains and not commonly observed within other *Sinorhizobium* or *Rhizobiaceae* species (*321*). PodJ2, a protein similar to the C-terminal periplasmic domain of $PodJ_{Cc}$, was found to have no identifiable function. The PodJ1 protein exhibits similarities to the N-terminal cytoplasmic domain of $PodJ_{Cc}$, possessing a transmembrane segment and coiled-coil regions. However, despite PodJ1 and $PodJ_{At}$ both being documented homologs of $PodJ_{Cc}$, notable differences in the localisation of PodJ1 and $PodJ_{At}$ were observed. PodJ1 localises to the newer pole (i.e. growth pole) during cell elongation, and was not found to exhibit the bipolar localisation observed with the *Agrobacterium* PodJ protein, suggesting that PodJ1 may function more like $PodJ_{Cc}$ than $PodJ_{At}$ (*273*, *321*). Within *Sinorhizobium*, PodJ1 is attributed to the control of flagella motility, cell envelope integrity, cell division and morphology, but is seemingly not required for endosymbiotic interactions with its host legume *Medicago sativa* (alfalfa) (*321*).

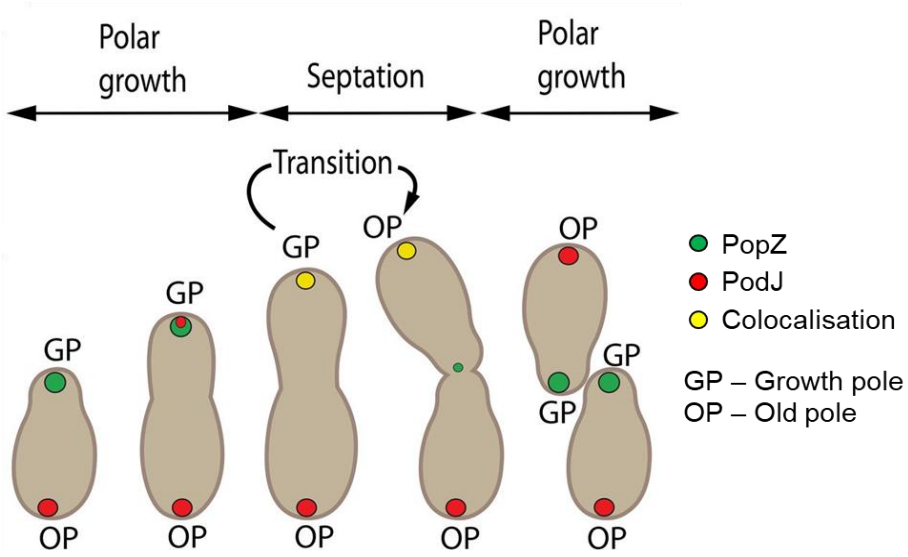


Figure 1.14. Localisation of *Agrobacterium* PopZ and PodJ proteins. Within *Agrobacterium tumefaciens*, PopZ (green) serves as a growth pole (GP) marker, exhibiting strict GP localisation. In newly formed cells, PodJ (red) exhibits unipolar localisation at the old pole (OP), but later in the cell cycle (i.e. during GP to OP transition) PodJ also accumulates at the growth pole with PopZ (colocalisation, yellow). During GP to OP transition, PopZ moves to the midcell, where new GPs will emerge. Cell division gives rise to daughter cells with unipolar localisations of PopZ (GP) and PodJ (OP). Figure modified from Grangeon *et al.*, 2015, who also characterised the localisation of these two proteins (*273*).

In *A. tumefaciens*, PopZ exhibits strict growth pole localisation throughout the cell cycle, and rapidly relocalises to the new growth pole as daughter cells separate (*273*, *325*, *326*). The loss of PopZ from the previous growth pole has been shown to be a result of the movement of PopZ to the new growth pole, as opposed to its removal from the old pole by proteolysis (116). The deletion of *popZ* from *Agrobacterium* results in several morphological and growth defects including varied cell length, minicell formation, ectopic growth poles and in some instances cellular branching (*325*, *326*). Interestingly, despite its consistent growth pole localisation, PopZ is not required for the direction of polar PG synthesis, and in its absence PG synthesis at the growth pole resumes with minimal impact on PG composition (*325*). Therefore, it is thought that PopZ's primary function is the coordination of cell division. Many of the phenotypes arising from the deletion of *popZ* are indicative of cell division defects, including minicell formation, elongated cells and Y-shaped cell branching (*325*, *326*). In *Agrobacterium*, the division proteins $FtsZ_{At}$ and FtsA localise to the growth pole for the majority of the cell cycle, then relocate to the midcell prior to cell division (*127*, *131*). Within *popZ* deletion mutants, $FtsZ_{At}$ and FtsA are often

misplaced, leading to misplacement of the Z ring and hence asymmetric cell constriction (*325*). However, PopZ does not seem to be required for polar localisation of FtsA during cell elongation (*325*). Although $FtsZ_{At}$ can still form rings in the absence of PopZ, these Z rings are often misplaced and less stable, and around 40% of cells exhibit Z ring formation near the pole (leading to minicell formation following cell division) (*325*). In the absence of $FtsZ_{At}$, PopZ accumulates at the growth pole indicating that $FtsZ_{At}$ may be required for the recruitment of PopZ to the midcell (i.e. the site of new growth pole formation) during GP to OP transition (*130*).

In addition to its role in coordinating cell division, PopZ is also required for effective chromosome segregation (*324–326*). It was observed that the minicells formed as a result of the loss of PopZ typically lacked DNA, indicating possible chromosome segregation defects or the occurrence of misplaced (polar) cell division prior to chromosome segregation (*325*). Indeed, it has been shown that PopZ is essential for the effective segregation of both the circular and linear chromosomes of *Agrobacterium tumefaciens*, tethering these genetic elements to the new pole (*323*, *324*, *326*). The tethering of chromosomes by $PopZ_{At}$ to the new (growth) pole is a function shared with $PopZ_{Cc}$, which binds ParB-*parS* nucleoprotein complexes present on the *C. crescentus* chromosome (*319*). As $popZ_{At}$ deletion strains exhibit aberrant ParB localisation patterns, with around 30% of cells lacking a ParB focus, the interaction of $PopZ_{At}$ with ParB may serve as a conserved mechanism of chromosome anchoring by PopZ proteins (*326*). Despite differences in their localisation patterns, $PopZ_{At}$ and $PopZ_{Cc}$ both share a conserved C-terminal domain, which in $PopZ_{Cc}$ is required for the polymerisation of PopZ proteins into a matrix (*273*, *327*). In $PopZ_{At}$, this C-terminal region is believed to be important for proper polar localisation of PopZ (*325*). Heterologous expression of $PopZ_{Cc}$ in *E. coli* does not impair its ability to form matrixes at cell poles, highlighting that the polymerisation of $PopZ_{Cc}$ relies upon intrinsic features of the protein (i.e. the C-terminal domain) (*294*, *311*, *327*). Conversely, when expressed within *E. coli*, $PopZ_{At}$ displays a whole-cell diffuse localisation pattern, suggesting that polar localisation of $PopZ_{At}$ may require species-specific polarity proteins (*273*).

One such protein may be the growth pole ring protein (GPR). As mentioned in Section 1.3.2, GPR is a membrane-anchored cytoplasmic protein proposed to serve as an organising centre for the polar growth protein complex within *A. tumefaciens* (*278*, *279*). Although initially identified following a search for homologs of the *Caulobacter* polarity determinant TipN, GPR was determined to be distinct from TipN based on their limited sequence similarity, large

discrepancies in size (882 aa for TipN vs. 2,115 aa for GPR), and the distinctive localisation of GPR (i.e. polar hexameric ring) (*278*). However, TipN and GPR do share some structural similarities including N-terminal transmembrane domains, N- and C-terminal disordered regions, and central regions of continuous alpha helices (the coiled-coil region of TipN) (*278*, *307*, *308*). GPR exhibits unipolar localisation at the growth pole for most of the cell cycle, but localises at the midcell just before cell division, ensuring its localisation at the new pole following cell division. During its relocalisation, GPR can still be observed at the old pole for a period of time, displaying a period of bipolar localisation (*278*). Cells with a deletion of *gpr* grow slowly, are coccoid in morphology, are often enlarged, and display little evidence of normal unipolar growth. Within these coccoid cells, PG synthesis is abnormally localised around circumference of the cell, supporting the role of GPR as an organiser of polar PG synthesis machinery (*279*). Additionally, even small amounts of GPR overexpression leads to the formation of ectopic growth poles within the cell (*278*). The function of GPR as a polar scaffold for growth machinery is primarily attributed to two structural features of this protein: its coiled-coil-containing apolipoprotein domains, and its disordered C-terminal domain. GPR possesses five domains with similarity to human apolipoprotein A-IV, two of which contain predicted coiled-coil regions (*279*). Deletion of either of these coiled-coil-containing A-IV domains from full-length GPR results in the defects in cell morphology and growth described above, with the deletion of one (A-IV-1) also impacting the ability of GPR to form a hexameric structure at the growth pole. However, the coiled-coil region within A-IV-1 is not required for hexamer formation (*279*). Additionally, deletion of the C-terminal domain also impairs the function of GPR in regulating polar growth and cell morphology, but seemingly does not impact hexameric ring formation (*279*). The abundance of apolipoprotein domains throughout the alpha-helical length of GPR led to the proposal that GPR may play a role in lipid sorting and membrane synthesis, but the mechanism of how GPR is involved in these processes remains unknown (*279*).

Importantly, GPR likely serves as the polar landmark protein required for the localisation of PopZ, with GPR localising to the new pole slightly before PopZ (*278*). At the growth pole, the localisation of PopZ and GPR mostly overlaps, with a small amount of PopZ extending beyond GPR towards the tip of the cell. This localisation pattern represents a central focus of PopZ surrounded by the hexameric GPR structure (*278*). In GPR-depleted cells, PopZ exhibits a diffuse localisation pattern, and PopZ cannot localise to the midcell during septation if GPR is absent (*278*). Finally, much like other polarity landmarks and pole organisers (e.g. PopZ, PodJ), GPR also

plays a role in the organisation of chromosomes, with GPR having been found to be important for the localisation of replication origins to the growth pole (*323*).

1.5 Understanding unipolar growth in *Rhizobium leguminosarum*

As highlighted throughout this literature review, a small number of well-characterised bacterial cytoskeletal (and associated) proteins have also been functionally characterised within rhizobial species both within and beyond the *Rhizobiaceae* including ParA (*Azorhizobium caulinodans*), PodJ and the MinCDE system (*Sinorhizobium meliloti*), FtsZ (*Rhizobium meliloti*) and the RepA and RepB proteins (*R. leguminosarum*) (*136*, *198*, *222*, *227*, *321*, *328*). Yet, the precise requirement of these proteins during unipolar growth within the *Rhizobiaceae*, including possible interactions between these proteins, remains somewhat a mystery. Essential to unipolar growth is the determination of cell polarity and pole identity, allowing for the distinction between the old pole and the growth pole. Hence, a key step in determining the mechanisms which underpin the spatial and temporal coordination of unipolar growth within the genera *Rhizobium* is the identification of proteins which may associate to form polar protein complexes within these species.

In this thesis, *Rhizobium leguminosarum* bv. *viciae* (strain 3841) (NCBI:txid216596) was chosen as the model *R. leguminosarum* species, in part due to the availability of a complete whole-genome sequence of this species (*329*). For many years, this species has served as an important model species for the study of rhizobium-legume symbioses, biological nitrogen fixation, nodule biochemistry and rhizobial genetics, with over 100 publications directly studying various facets of this species' lifestyle (data obtained using Dimensions (*330*)). In July 2023, whole-genome phylogenetic analysis of the *Rhizobium leguminosarum* species complex resulted in a proposed taxonomic reclassification of three previously unnamed *R. leguminosarum* clades (*13*). Consequently, *R. leguminosarum* bv. *viciae* 3841 was renamed *Rhizobium johnstonii* 3841, however its former name (abbreviated to *R. leguminosarum* 3841 hereafter) has been used throughout this thesis. *R. leguminosarum* 3841 has a 7.75 Mb genome, comprised of a circular chromosome (5 Mb) and 6 accessory plasmids (*329*). The organisation of genomes as multiple replicons (i.e. multipartite genomes) is a characteristic feature of rhizobial species (*329*, *331–333*). The genes for host-specificity, nodulation (e.g. *nod*, *nol* and *noe* gene

clusters), and nitrogen fixation (*nif* and *fix* gene clusters) are present on a single plasmid (pRL10), referred to as the symbiotic plasmid, which permits the establishment of endosymbiotic interactions with legumes in the tribe Fabeae (Vicieae), including pea (*Pisum sativum*), lentil (*Lens culinaris*), broad bean (*Vicia faba*) and vetch (*Vicia cracca*) (*329*, *334*).

1.5.1 *R. leguminosarum* PopZ

Given its established role as a new-pole marker and pole organiser within the *Rhizobiaceae* member *Agrobacterium tumefaciens* (*273*, *324–326*), PopZ was hypothesised to be a very likely candidate of a growth-associated polar protein complex within *Rhizobium leguminosarum*. To identify homologs of PopZ within *R. leguminosarum* 3841, the *Agrobacterium tumefaciens* C58 PopZ ($PopZ_{At}$) amino acid sequence (Atu1720) (*273*) was used as the query sequence for a Protein BLAST (BLASTP) against the *R. leguminosarum* 3841 genome. This search identified a 293 amino acid (aa) hypothetical protein encoded by the *R. leguminosarum* 3841 gene *RL2034*. A BLASTP search of the *Caulobacter crescentus* CB15 PopZ ($PopZ_{Cc}$) sequence (Cc1319) (*294*, *309*) against *R. leguminosarum* 3841 confirmed the identity of the *RL2034* gene product as a PopZ homolog. Hereafter, the gene *RL2034* has been designated *popZ*, with its gene product named PopZ. Protein structure prediction using Phyre2 (*335*) suggested that the majority of this protein is disordered, a characteristic feature of PopZ proteins (*311*, *336*). The *Agrobacterium tumefaciens* and *Rhizobium leguminosarum* PopZ sequences (333 and 293 aa respectively) exhibit a pairwise identity of 41%, with more highly conserved N- and C-terminal regions. The conservation of N- and C-terminal regions of PopZ has also been documented between *Agrobacterium* and *Caulobacter* PopZ proteins (*325*), and is likely due to their function. Within $PopZ_{Cc}$, a 75 aa region within the C-terminus is required for its polymerisation and higher order assembly, while a 24 aa N-terminal region confers binding specificity to other proteins (*311*, *336*). In $PopZ_{At}$, the C-terminal region (residues 259-333) is required for polar localisation of this protein (*282*, *325*). Despite an overall sequence homology of 38%, the C-terminal regions of $PopZ_{At}$ and *R. leguminosarum* 3841 PopZ (residues 259-333 and 219-293 respectively) are 73% identical, demonstrating that this region is highly conserved within the *Rhizobiaceae*.

1.5.2 A novel *Rhizobium* cytoskeletal element?

It is widely accepted that members of the *Rhizobiaceae* (and more broadly the rhizobia) lack homologs of certain canonical cytoskeletal proteins including MreB (present in most other Gram-negative rods) and DivIVA (not present in Gram-negative species). However, little is known about whether this family of bacteria possess homologs of the non-canonical coiled-coil IF-like TIPOC proteins Scy and FilP, which to date have only been documented within Actinobacteria (*102*). A key feature of this broad class of relatively-uncharacterised proteins is the periodic occurrence of hydrophobic residues within their amino acid sequence which give rise to non-canonical coiled-coil motifs (*175*). These proteins typically share little sequence homology, and can vary greatly in length, making sequence-similarity searches such as BLAST ineffective for the identification of homologs in more distantly related species. Therefore, Walshaw *et al.* (*175*) used fast Fourier transforms, an algorithm which can be used to calculate periodicities of hydrophobic residues within amino acid sequences (*337*), to identify potential homologs of Scy. Amongst these homologs was a hypothetical protein encoded by the *R. leguminosarum* 3841 gene *RL2654* (Kelemen, personal communication. Data not published). In contrast to the 1,326 aa protein Scy, *RL2654* encodes a protein of 2,376 aa which has been named *Rhizobium* cytoskeletal element (Rcy) in accordance with the nomenclature of Scy (*Streptomyces* cytoskeletal element). Hereafter, the gene *RL2654* has been named *rcy*.

Scy is predicted to possess an N-terminal canonical, coiled-coil domain (residues 1–64), a short, flexible “hinge” (residues 64–72) and a long (1,227 residue) C-terminal non-canonical coiled-coil domain. The majority of Scy is predicted to be α-helical, with a high coiled-coil forming probability throughout (*175*, *182*). Structural predictions using PredictProtein (*338*) and Phyre2 identified that 80% of Rcy is predicted to be α-helical. However, in contrast to Scy, only a handful of short regions of Rcy (residues 469-545, 1447-1477, 1649-1735 and 1973-1993) were predicted to form coiled-coils, potentially suggesting that the coiled-coil repeat architecture of Rcy may differ from that of Scy and other known coiled-coil proteins. Pfam (*339*) and InterPro (*340*) domain predictions predicted that Rcy has an N-terminal disordered region (residues 27-88), two N-terminal transmembrane domains (residues 112-130 and 142-166), and a second C-terminal disordered region (residues 2143-2261). The central length of Rcy is predicted to contain 13 overlapping or adjacent apolipoprotein (lipid-binding) domains.

Surprisingly, both the length and predicted domain architecture of Rcy is strikingly similar to that of the growth pole ring protein GPR (Figure 1.15). GPR is 2,115 aa long (compared to the 2,376 aa-long Rcy), with two N-terminal transmembrane domains, N- and C-terminal proline-rich disordered regions, and a large 1,700 aa central α-helical region containing 12 predicted apolipoprotein domains (*278*). Within this central domain are two, 21 aa coiled-coil domains (residues 413-433 and 1,100-1,120) that lie within the predicted apolipoprotein domains (*279*). However, although attributing GPR to the control of morphology and growth within *Agrobacterium*, and proposing that it functions as an organising centre for membrane and cell wall synthesis during polar growth, Zupan *et al.* did not explicitly comment on whether they categorise GPR as a member of the prokaryotic cytoskeleton or an intermediate filament-like protein (*278*, *279*).

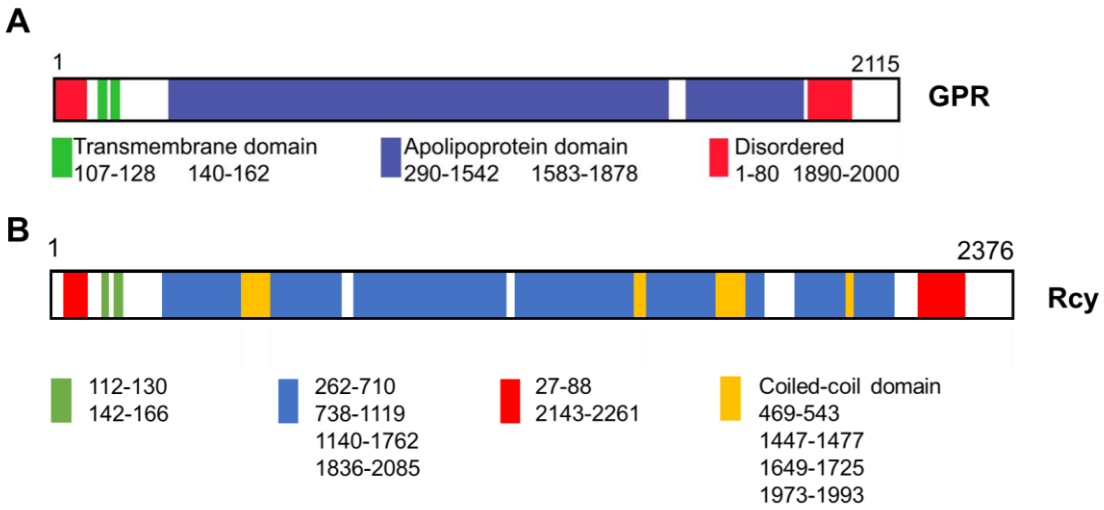


Figure 1.15. Similarities in the predicted domain architecture of GPR and Rcy. (A) Predicted domain architecture of growth pole ring protein (GPR), published by Zupan *et al.* (*278*). GPR is predicted to contain an N- and C-terminal disordered region (red), two N-terminal transmembrane domains (green) and 12 overlapping or adjacent apolipoprotein domains (blue). Not shown within this figure are two predicted 21 amino acid coiled-coil domains (residues 413-433 and 1,100-1,120) which lie within the predicted apolipoprotein domains (*279*). (B) Predicted domain architecture of Rcy, designed to be in-keeping with Part A of this figure (published by Zupan *et al.*) to facilitate comparisons between the two proteins. Rcy is predicted to possess an N- and C-terminal disordered region (red), two N-terminal transmembrane domains (green), and a long central region containing 13 overlapping or adjacent apolipoprotein domains (blue). Unlike Rcy which contains two equal-length coiled-coil domains, Rcy has four predicted coiled-coil domains (yellow) of varying lengths, all of which lie within the predicted apolipoprotein domains.

1.5.3 *R. leguminosarum* ParA1

In this thesis, ParA has been proposed as a third *Rhizobium* polar complex protein. Although not an immediately apparent polar complex protein due to its nucleoid-wide binding and localisation pattern (*215*, *341*, *342*), ParA has been shown to exhibit localisation foci towards the cell poles later in the cell cycle (*208*, *315*, *341*). Additionally, ParA proteins have been found to directly interact with known polarisome/polar complex proteins including Scy (*Streptomyces*), PopZ (*Caulobacter*) and TipN (*Caulobacter*) (*183*, *206*, *314*, *315*). Further support for ParA as a possible unipolar growth-associated protein within *Rhizobium* comes from characterisation studies within *Azorhizobium*, where a null mutation of the *parA* gene within the *Azorhizobium caulinodans* genome was associated with changes in cell morphology and reduced growth rate (*227*). To identify ParA homologs within *R. leguminosarum* 3841, the *Agrobacterium tumefaciens* ParA homolog (CAA28769) was first identified following a search of the UniProt protein database for known ParA proteins within the *Rhizobiaceae*. BLASTP and DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST) searches of the amino acid sequence of this protein against *R. leguminosarum* 3841 highlighted several potential ParA homologs, two of which were encoded by genes *RL4736* and *RL3343* located on the chromosomal plasmid. A number of other plasmid-encoded 'ParA' homologs were also identified, synonymous with the RepA protein of the RepABC plasmid partitioning systems within multipartite genomes (*220*, *221*). To determine which of the two chromosomal '*parA*' genes encodes the primary copy of ParA, their downstream genes were investigated. Typically, ParA proteins are encoded within an operon along with ParB and the *parS* sequence (the *parABS* locus) (*220*, *343*, *344*). The downstream gene of *RL4736* is predicted to encode a ParB protein (RL4735) (*329*), therefore *RL4736* was presumed to encode the primary ParA protein, and was designated *parA1* accordingly. The other *parA* gene (*RL3343*), seemingly not located within a ParABS operon, was named *parA2* but was not studied further within this thesis. *R. leguminosarum* 3841 *parA1* encodes a 264 amino acid protein (ParA1). Pfam and InterPro predictions highlighted that ParA1 contains a P-loop NTPase domain (residues 2-258), within which lies an AAA+ domain (residues 7-183). These domains are characteristic of Walker A cytoskeletal ATPase proteins (*187–189*), confirming the identity of this protein as a member of the ParA family of cytoskeletal proteins.

1.6 Experimental Aims

Recent research within the *Rhizobiaceae* species *Agrobacterium tumefaciens* and *Sinorhizobium meliloti* has highlighted several polar growth and polarity proteins essential for unipolar growth and division within this family (and more broadly the Hyphomicrobiales) including PopZ, PodJ, GPR and the Rgs proteins (*273*, *278*, *280*, *281*, *321*). There is no doubt that these studies have been invaluable in expanding our knowledge of unipolar growth within the *Rhizobiaceae*, but our understanding of this growth mechanism is by no means comprehensive, and the precise role of many of these proteins within other members of the *Rhizobiaceae* including *Rhizobium* species remains incredibly limited. This thesis therefore aims to shed light on unipolar growth within this agriculturally significant bacterial family through the identification and characterisation of proteins which are required for the coordination of growth, division and cell morphology within the model *Rhizobium* species *Rhizobium leguminosarum* (*R. johnstonii*) strain 3841. Within this thesis, three candidate polar proteins have been studied. These are the *R. leguminosarum* homologs of the chromosome partitioning ATPase ParA1 and the polar organising protein PopZ, and a putative homolog of the *Streptomyces* TIPOC protein Scy, named *Rhizobium* cytoskeletal element (Rcy), which also exhibits striking similarity to the *Agrobacterium* polar protein GPR.

The first step taken towards the characterisation of ParA1, PopZ and Rcy was to assess the physiological impact of the loss of these proteins via the attempted generation of whole gene *parA1*, *popZ* and *rcy* knockout strains (Chapter Three). As an alternative approach to elucidate gene function, inducible *parA1*, *popZ* and *rcy* overexpression *R. leguminosarum* 3841 strains were then generated for the purpose of studying the consequence of the misregulation of these genes and their encoded proteins. Novel overexpression phenotypes including impacts on population growth and cell morphology within these strains were analysed to establish the involvement of these proteins in the regulation of unipolar growth within *R. leguminosarum* (Chapter Four). Finally, localisation studies of ParA1-EGFP, PopZ-EGFP and Rcy-EGFP translational fusion proteins were performed to provide valuable insights into not only the normal localisation of these proteins (and hence their likelihood of being polarity complex proteins), but also how their misregulation may alter the growth of *R. leguminosarum* (Chapter Five). Taken together, these research approaches aim to not only address whether or not ParA1, PopZ and Rcy are required for coordinated unipolar growth within *R. leguminosarum* 3841, but may also provide supporting or even new evidence for the precise functions of these proteins

within this species. Overall, by characterising these proteins within *Rhizobium leguminosarum*, the work performed within this thesis sets out to not only advance our understanding of this relatively unique and poorly understood mechanism of growth, but also contribute insights into how members of the bacterial cytoskeletal network (and its associated proteins) bring about the important morphological changes within rhizobia required for the effective establishment of endosymbiotic interactions such as bacteroid differentiation.

Chapter Two – Materials and Methods

2.1 Bacterial strains, media and growth conditions

2.1.1 Bacterial strains

In this thesis, *Rhizobium leguminosarum* bv. *viciae* strain 3841 (referred to hereafter as *R. leguminosarum* 3841) was used as the model *Rhizobium leguminosarum* strain. This strain is a spontaneous streptomycin-resistant mutant of *Rhizobium leguminosarum* bv. *viciae* strain JB300 (*334*, *345*). Multiple *Escherichia coli* (*E. coli*) strains were used to facilitate various molecular biology and genetic manipulation techniques. The key characteristics and genotypes of all strains used in this thesis have been summarised in the table below (Table 2.1).

Table 2.1. Bacterial strains used in this thesis.

Strain	Key features	Ref.
R. leguminosarum 3841	Streptomycin resistant (str^R) 7.75 Mb genome, comprised of 1 circular chromosome (5 Mb), and six accessory plasmids The plasmid pRL10 (symbiosis plasmid, aka pSym), enables symbiotic interactions with members of the plant tribe Vicieae	(*329*, *334*, *345*)
E. coli DH5α	*recA1*, *endA1* *lacZΔM15*, allowing for blue-white selection	(*346*, *347*)
E. coli DH10B	*recA1*, *endA1* *lacZΔM15* Str^R	Thermo Fisher Scientific
E. coli pRK2013	*E. coli* DH5α strain carrying the helper plasmid pRK2013 Kanamycin resistant (kan^R) (conferred by pRK2013)	(*348*)
E. coli BW25113/pIJ790	*E. coli* BW25113 carrying the plasmid pIJ790 $recA^+$ *ΔlacZ4787* $lacI^+$ Δara*BAD* Chloramphenicol resistant (chl^R) (conferred by pIJ790)	(*349–351*)

2.1.2 Growth medium and conditions

Typically, *E. coli* strains were cultured in Luria Broth (Miller) (LB) (*352*) at 37°C for approximately 14-18 hours. *R. leguminosarum* 3841 was cultured in Tryptone-Yeast extract broth (TY) (*353*) at 30°C for a period of approximately 48-60 hours (2-3 days). An important deviation from the standard *E. coli* growth conditions stated above was the culture of strain BW25113/pIJ790. To ensure retention of the temperature-sensitive plasmid pIJ790 (*349*, *351*), this strain was grown at 30°C. Bacterial cultures were routinely grown either on 90 mm Agar plates, or as 10 mL liquid cultures within 30 mL universal glass vials. For volumes > 10 mL, cultures were grown in conical flasks. To ensure aerobic respiration, all liquid cultures were incubated with shaking at 180 rpm within an Innova 44 Incubator Shaker (New Brunswick). During the generation of knockout mutants (Section 2.8), a modified version of TY containing 10% (w/v) sucrose (referred to as TY-sucrose) was used for the selection of double-crossover events. In addition, Super Optimal Broth (SOB) (*346*) was used for the culture of *E. coli* BW25113/pIJ790 during the generation of EGFP fusion proteins (Chapter Five), while Super Optimal broth with Catabolite repression (SOC) medium, a modified version of SOB containing additional glucose (*346*), was used for the recovery of cells following heat-shock transformation (Section 2.4.1). The composition of all bacterial growth media is described in Table 2.2. For solid medium, referred to as Agar medium, 15 g Agar was added per 1 L of media (1.5% (w/v)).

Table 2.2. Composition of bacterial growth media. For solid medium, 15 g Agar was added per 1 L of media. Sucrose solution was filter sterilised with polyethersulfone (PES) Minisart® syringe filters (pore Size 0.22 µm) (Sartorius).

Medium	Composition per 1 litre of de-ionised water	Ref.
Luria Broth (LB)	Tryptone 10 g, Yeast Extract 5.0 g, $NaCl_2$ 10 g	(*352*)
Tryptone-Yeast extract broth (TY)	Tryptone 5.0 g, Yeast Extract 3.0 g, $CaCl_2.6H_2O$ 1.32 g	(*353*)
TY-sucrose (10%)	Tryptone 5.0 g, Yeast Extract 3.0 g, $CaCl_2.6H_2O$ 1.32 g, Sucrose 100 g (200 mL of filter-sterile 50% (w/v) sucrose solution added after autoclaving. Volume of de-ionised water adjusted accordingly)	-
Super Optimal Broth (SOB)	Bacto Tryptone 20 g, Yeast Extract 5.0 g, NaCl 0.58 g, KCl 0.186 g, $MgSO_4$ 1.2 g, $MgCl_2$ 0.95 g	(*346*)
Super Optimal broth with Catabolite repression (SOC)	Bacto Tryptone 20 g, Yeast Extract 5.0 g, NaCl 0.58 g, KCl 0.186 g, $MgSO_4$ 1.2 g, $MgCl_2$ 0.95 g, Glucose 4 g (8 mL filter-sterile 50% glucose solution (w/v) added after autoclaving)	(*346*)

2.1.3 *R. leguminosarum* growth curve generation

The extent of bacterial growth was determined by measuring the optical density of culture samples at a wavelength of 600 nm (OD_{600}) using a Jenway™ 7315 UV/Visible Spectrophotometer. To generate *R. leguminosarum* growth curves, 10 mL starter cultures were grown as described in Section 2.1.2. Starter cultures were then standardised to an OD_{600} of approximately 1.0 by dilution with sterile TY medium. A 1% inoculum (v/v) was taken from the starter cultures and transferred into 250 mL conical flasks containing 50 mL sterile TY medium with 200 µg mL^{-1} kanamycin. Where kanamycin selection was not applicable (i.e. culture of wildtype *R. leguminosarum* 3841), 500 µg mL^{-1} streptomycin was added to the culture. Flasks were transferred to a shaker incubator set to 30°C and 180 rpm, and grown for approximately 90 hours. OD_{600} readings were taken a minimum of every 12 hours throughout the growth period. To ensure accurate OD_{600} measurements were taken, culture samples with an OD_{600} greater than 1.0 were diluted with fresh TY medium until the OD_{600} reading obtained was below 1.0. Actual OD_{600} values were then calculated accordingly. Mean OD_{600} readings of 3-4 biological replicates were calculated and plotted against time (hours) to generate growth curves.

2.2 Plasmids & antibiotics

2.2.1 Plasmids

The majority of plasmids used throughout this study were derived from one of two cloning vectors: pK18*mobsacB* (*354*), a sucrose-sensitive vector used during the generation of knockout mutants (Chapter Three); and pIND6 (Kelemen, unpublished), a derivative of the IPTG-inducible expression vector pIND4 (*355*), used for the overexpression of genes within *R. leguminosarum* (Chapters Four and Five). A comprehensive list of all plasmids used throughout this thesis can be found in Table 2.3. Vectors used solely for the delivery of artificially synthesised gene sequences (refer to Section 2.3.7) have not been included in this table.

Table 2.3. List of plasmids used.

Plasmid	Features and functions	Source
pK18*mobsacB*	Sucrose-sensitive (*sacB*) plasmid used for allelic exchange during conjugation *LacZa*, *mob*+, kan^R	(*354*)
pK18-*popZ*-KO	pK18*mobsacB*-derivative possessing the *popZ* knockout construct, used for the unmarked deletion of *popZ* within *R. leguminosarum*	This study
pK18-*rcy*-KO	pK18*mobsacB*-derivative possessing the *rcy* knockout construct, used for the unmarked deletion of *rcy* within *R. leguminosarum*	This study
pK18-*parA1*-KO	pK18*mobsacB*-derivative possessing the *parA1* knockout construct, used for the unmarked deletion of *parA1* within *R. leguminosarum*	This study
pRK2013	Conjugal helper plasmid, used for the mobilisation of non-transmissible plasmids into *R. leguminosarum* *mob*+, *tra*+, kan^R	(*348*)
pSUP3535	Tet^R transcriptional *lacZ* fusion suicide vector Source of tetracycline resistance (tet^R) cassette Derivative of pSUP202 (*356*)	(*357*)
pK18-*parA1*-tet^R-KO	pK18-*parA1*-KO derivative used for generation of tetracycline-resistant marked *parA1* knockout mutants	This study
pK18-*popZ*-tet^R-KO	pK18-*popZ*-KO derivative used for generation of tetracycline-resistant marked *popZ* knockout mutants	This study
pK18-*rcy*-tet^R-KO	pK18-*rcy*-KO derivative used for generation of tetracycline-resistant marked *rcy* knockout mutants	This study
pIND6	IPTG-inducible-expression plasmid, used for expression of proteins – A derivative of pIND4 (*355*) Kan^R, *lacI^q, lac* promoter	Kelemen, Unpublished
pIND6-*parA1*	pIND6-derivative expressing N-terminal His-tagged *R. leguminosarum* ParA1 protein Kan^R	This study
pIND6-*popZ*	pIND6-derivative expressing N-terminal His-tagged *R. leguminosarum* PopZ protein Kan^R	This study
pIND6-*rcy*	pIND6-derivative expressing N-terminal His-tagged *R. leguminosarum* Rcy protein Kan^R	This study
pIJ790	Temperature-sensitive recombination plasmid expressing the phage λ Red recombinase system Chl^R, λ Red genes (*gam*, *bet*, *exo*), *araC*, *rep101^{ts}* Used to facilitate generation of EGFP fusion proteins	(*349*, *358*)

Table 2.3 (continued)

pIND6-*parA1-egfp*	Derivative of pIND6-*parA1* expressing an N-terminal His-tagged ParA1-EGFP fusion protein Used for localisation studies in *R. leguminosarum* Apra^R, kan^R	This study
pIND6-*popZ-egfp*	Derivative of pIND6-*popZ* expressing an N-terminal His-tagged PopZ-EGFP fusion protein Used for localisation studies in *R. leguminosarum* Apra^R, kan^R	This study
pIND6-*rcy-egfp*	Used for inducible expression of an N-terminal His-tagged Rcy-EGFP fusion protein Used for localisation studies in *R. leguminosarum* Apra^R, kan^R	This study

2.2.2 Antibiotics, additives, and inducers

Unless otherwise stated, all *R. leguminosarum* cultures were grown in the presence of 500 µg mL^{-1} streptomycin (str). In some instances, such as blue-white screening of *E. coli* colonies or the induction of gene expression, additional reagents and inducers were added to the growth media. A full list of all antibiotics, inducers and other media additives, and their respective final concentrations for *E. coli* and *R. leguminosarum* are listed in Table 2.4. All solutions were filter-sterilised with polyethersulfone (PES) Minisart® syringe filters, pore size 0.22 µm (Sartorius), and stored at -20°C.

Table 2.4. Table of antibiotic, additive and inducer concentrations used throughout this thesis. Several antibiotics and other media additives were used during the culture of *E. coli* and *R. leguminosarum*. The concentrations used during the culture of the relevant species are detailed in this table. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was used for both blue-white screening, and as an inducer of gene expression. For IPTG and L-Arabinose, molar concentrations have been provided instead of mg/μg mL^{-1}.

	Solvent	[Stock] (mg mL^{-1})	[Final] (μg mL^{-1}) *E. coli*	[Final] (μg mL^{-1}) *R. leguminosarum*
Antibiotics				
Kanamycin (kan)	dH_2O	100	50	200
Spectinomycin (spc)	dH_2O	100	100	-
Streptomycin (str)	dH_2O	500	-	500
Tetracycline (tet)	Ethanol	10	10	5
Ampicillin (amp)	dH_2O	100	100	-
Chloramphenicol (chl)	Ethanol	25	25	-
Apramycin (apra)	dH_2O	50	50	-
Additives and Inducers				
IPTG (induction)	dH_2O	1 M	-	Typically 1 mM
IPTG (screening)	dH_2O	1 M	1 mM	-
X-gal	DMSO	200	200	-
L-Arabinose	dH_2O	1 M	10 mM	-

2.3 DNA preparation and handling

2.3.1 Isolation of DNA from bacteria

Throughout this thesis, plasmid DNA was routinely isolated from *E. coli* cultures (Section 2.3.1.1), with genomic DNA isolated from *R. leguminosarum* cultures (Section 2.3.1.2). Following the isolation of both plasmid and genomic DNA, the quantity and quality of the purified DNA was determined by measuring absorbance at wavelengths of 260 and 280 nm using a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fisher Scientific).

2.3.1.1 Plasmid DNA

Plasmid DNA was isolated from microbial cultures (3-5 mL) using the QIAGEN QIAprep Spin Miniprep Kit. This kit uses chemical lysis of cells to release plasmid DNA. Following lysis and subsequent neutralisation of the reaction, lysates and solubilised DNA are separated by centrifugation, and the supernatant is applied to a silica membrane spin column. Plasmid DNA is bound to the membrane in the presence of a high salt binding buffer. Following washing of the column with an ethanol-based wash buffer to remove impurities, plasmid DNA was eluted into 30 µL nuclease-free water. All centrifugation steps were performed using an Eppendorf 5424 benchtop microcentrifuge, at the speed and duration specified by kit's protocol (typically 17,900 x g).

2.3.1.2 Genomic DNA

Genomic DNA (gDNA) was extracted from bacterial cells using the QIAGEN DNeasy UltraClean Microbial Kit. This kit uses a combination of mechanical (bead-beating) and chemical lysis to extract DNA from cells. Cell debris is then precipitated by centrifugation. Soluble proteins and any residual cell debris are then precipitated from the solution. The resulting supernatant, containing the solubilised genomic DNA is loaded onto a silica membrane spin column in the presence of a high salt buffer to allow for binding of DNA to the membrane. Following washing of the membrane with an ethanol-based wash solution, purified gDNA was eluted into 50 µL nuclease-free water. All centrifugation steps were performed using an Eppendorf 5424 benchtop microcentrifuge, at the speed and duration specified by kit's protocol (10,000 x g).

2.3.2 Agarose gel electrophoresis

Agarose gel electrophoresis of DNA was performed using 1% (w/v) agarose gels (1g agarose per 100 mL TAE (Tris-Acetate-EDTA) buffer (40mM Tris, 20mM acetic acid, 1mM EDTA)). 1X TAE buffer was prepared from a 50X stock solution (Thermo Fisher Scientific). Prior to pouring of the gels, ethidium bromide (EtBr) was added (final concentration 0.5 µg mL^{-1}) to allow for visualisation of the DNA under ultraviolet (UV) light. Once solidified, gels were transferred to an electrophoresis tank and submerged in TAE buffer. Samples were mixed with 6X Purple Gel Loading Dye (New England Biolabs; NEB) and loaded on the gel alongside up to 5 µL GeneRuler

1 kb Plus DNA Ladder (Thermo Fisher Scientific). Gels were run at 90-110V for between 45 min to 1 hr 30 min depending on gel size. To visualise DNA, gels were imaged under UV light using a Syngene G:Box F3 Gel & Blot imaging system. In instances where EtBr was not added to the gel prior to pouring, gels were soaked in an ethidium bath (0.5 µg mL^{-1} EtBr solution) for approximately 10 minutes before visualisation. Whether or not EtBr was added directly to gel was solely dependent upon the working practices of the laboratory the work was being performed in.

2.3.3 Purification of DNA from agarose gel

DNA fragments (70 bp to 10 kb) excised from agarose gels were extracted using the QIAGEN QIAquick Gel Extraction Kit. In this protocol, gel slices are dissolved in a dual-purpose solubilisation and binding buffer and applied to a silica membrane. To ensure efficient binding of DNA to the membrane, this buffer contains and integrated pH indicator which turns purple when the pH of the solution is too high (> 7.5). If required, pH is adjusted with 3 M sodium acetate (pH 5.0) until the solution is no longer purple. Following loading onto the column, nucleic acids are bound to the membrane under the high-salt conditions provided by this buffer. Following washing of the membrane, nucleic acids were eluted into 30 µL nuclease-free water. Where increased concentrations of DNA were required, an elution volume of 15-20 µL was used. The quantity and quality of eluted DNA was determined by measuring absorbance at wavelengths of 260 and 280 nm using a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fisher Scientific).

2.3.4 Restriction enzyme digestion

All restriction enzymes used in this study were supplied by New England Biolabs (NEB) and were compatible with digestion in either CutSmart™ Buffer or NEBuffer™ r3.1. Digestion reactions were set up as detailed in Table 2.5. For diagnostic digestions, up to 1 µg DNA was digested for at least 1 hour (typically 1-2 hours). For molecular cloning, or previously incomplete digestions, 1 µg DNA was digested overnight (16-18 hours). All digestion reactions were carried out at 37°C. Following digestion, restriction enzymes were inactivated by either heat inactivation (if applicable) or gel purification (typically used during molecular cloning). For digestions with

multiple restriction enzymes, up to 1 µL of each enzyme was added to the reaction, ensuring that the total volume of restriction enzyme within the 50 µL reaction did not exceed 5 µL.

Table 2.5. Components of a typical restriction enzyme digestion.

Component	Quantity (per 50 µL reaction)
DNA	1 µg
10X CutSmart™ Buffer OR 10X NEBuffer™ r3.1	5 µL
Endonuclease	1.0 µL (10 units)
Nuclease-free water	Up to 50 µL

2.3.5 DNA ligation

Ligation of DNA fragments was performed using either T4 DNA ligase (Roche) or Hi-T4™ DNA Ligase (NEB). Although the reaction constituents and conditions for these two enzymes differ slightly (Table 2.6), their suitability for downstream protocols (i.e. heat inactivation, transformation) was identical, and no notable differences in ligation efficiencies were observed between the two enzymes. Therefore, the two enzymes were used interchangeably dependent upon availability. For the ligation of sticky (cohesive) ended DNA fragments (applicable to all cloning in this thesis), a 3:1 insert to vector ratio was typically used. In cases where the insert fragment was much smaller than the vector (approximately ≤ 1/10th of vector length), a 5:1 insert to vector ratio was used. For large inserts, i.e. those similar in size to the vector, a 1:1 insert to vector ratio was used. Reactions using Roche T4 DNA ligase (primarily those within Chapter 3) were incubated at 16°C for 1 hr, then transferred to 4°C overnight to complete the ligation reaction. For ligations using NEB Hi-T4™ DNA ligase (Chapter 4 onwards), reactions were incubated at 25°C for 1-2 hours. Following incubation, reactions were heat inactivated at 65°C for 10 min before transforming 1-5 µL of the reaction volume into *E. coli* for replication and subsequent isolation of the ligated plasmid (Section 2.4.1.2).

Table 2.6. Constituents of sticky-ended DNA ligation reactions. The components of a typical ligation reaction using either Roche T4 DNA ligase (left) or New England Biolabs Hi-T4™ DNA ligase (right) are detailed in this table. The quantity of insert DNA required (*X*) was determined by the following equation:

insert mass (ng)(*X*) = desired insert ratio x $\frac{\text{insert length (bp)}}{\text{vector length (bp)}}$ x vector mass (ng).

Roche T4 DNA ligase		**NEB Hi-T4™ DNA ligase**	
Component	**Quantity (per 30 µL reaction)**	**Component**	**Quantity (per 20 µL reaction)**
Vector DNA	50 ng	Vector DNA	50 ng
Insert DNA	*X* ng	Insert DNA	*X* ng
T4 DNA Ligase Reaction Buffer (10X)	3 µL	T4 DNA Ligase Reaction Buffer (10X)	2 µL
T4 DNA ligase	1 µL	Hi-T4™ DNA ligase	1 µL
Nuclease-free water	Up to 30 µL	Nuclease-free water	Up to 20 µL

2.3.6 Sanger sequencing of DNA

All DNA sequencing was carried out by Eurofins Genomics, using their Mix2Seq sequencing service. Eurofins were provided with 5 µL of either purified PCR product/DNA fragment (1-10 ng μL^{-1}) or plasmid DNA (50-100 ng μL^{-1}), pre-mixed with 5 µL of sequencing primer at 5 pmol μL^{-1}. Analysis and alignment of sequencing reads was performed using Benchling (benchling.com), a cloud-based biotechnology platform, and the molecular biology software SnapGene (Dotmatics). Where applicable, regions of poor-quality sequencing reads (i.e. read ends) were trimmed prior to alignment.

2.3.7 Artificial Gene Synthesis

The allelic exchange constructs used for the generation of unmarked *parA1*, *popZ* and *rcy* knockout mutants (Section 3.2.1) and the gene expression constructs required for the expression of His-tagged ParA1, PopZ and Rcy proteins (Section 4.2.2) were synthesised using one of two artificial gene synthesis services. Invitrogen's GeneArt gene synthesis service was used for the synthesis of the gene knockout allelic exchange constructs. For delivery, the synthesised constructs were cloned into the DraIII sites of the vector EC41308 (*359*) (Figure 2.1).

Following complications with the *parA1* knockout construct synthesised by Invitrogen (Appendix A.2), this construct was resynthesized directly into the XmaI and HindIII sites of pK18*mobsacB* using GenScript's gene synthesis service. A plasmid map of pK18*mobsacB* can be found in Figure 3.2, within Section 3.2. The His-tagged ParA1, PopZ and Rcy expression constructs were also synthesised by GenScript. These constructs were cloned into the EcoRV site of pUC57 (Figure 2.2), in the opposite orientation to the *lac* promoter. Once received, the lyophilized (GenScript) or dried (Invitrogen) plasmids were resuspended in accordance with the provided recommendations and transformed into *E. coli* for plasmid amplification and subsequent plasmid preparation. To select for cells transformed with EC41308-derived plasmids, transformations were plated onto LB Agar containing 100 µg mL^{-1} spectinomycin (spc). Transformations with pUC57-derived plasmids were plated onto LB Agar containing 100 µg mL^{-1} ampicillin (amp).

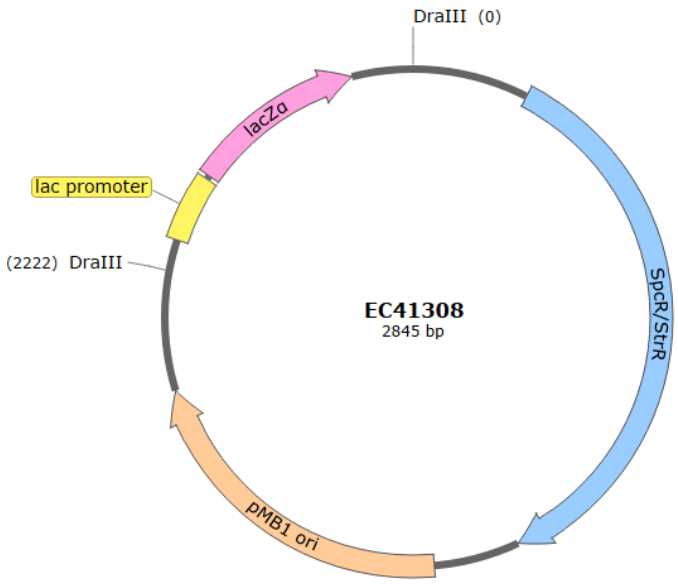


Figure 2.1. Plasmid map of EC41308. Simplified map of the plasmid EC41308 (*359*), used as the delivery vector for the synthetic gene knockout constructs synthesised by Invitrogen. This vector confers resistance to both spectinomycin (SpcR) and streptomycin (StrR) (blue). The pMB1 origin of replication (peach) results in a high copy number of this plasmid within *E. coli*. The *lac* promoter region (yellow) and *lacZα* fragment (pink) were removed following cloning of synthetic constructs into the DraIII sites of this vector.

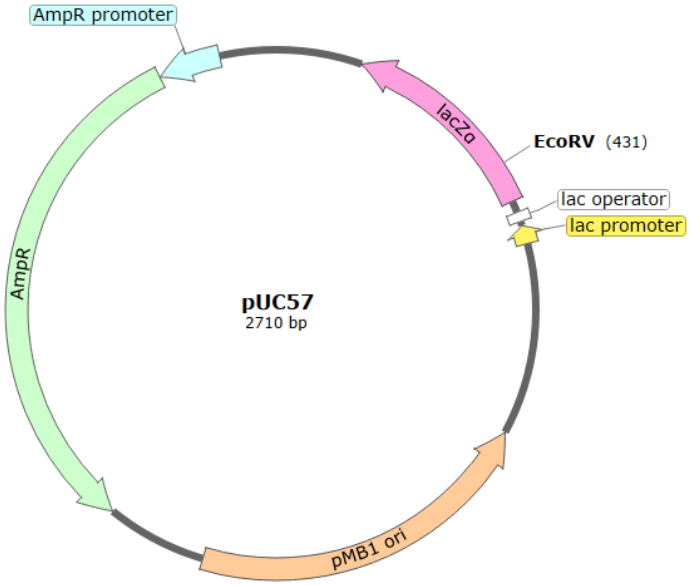


Figure 2.2. Plasmid map of pUC57. Simplified map of the plasmid pUC57, used as the delivery vector for the synthetic His-tagged expression constructs synthesised by GenScript. This vector confers resistance to ampicillin (AmpR, green). The pMB1 origin of replication (peach) allows for a high copy number in *E. coli*. Synthetic constructs were cloned into the EcoRV site, in the reverse orientation to the *lac* promoter (yellow). The *lac* promoter, *lac* operator (white) and *lacZa* fragment (pink) allow for blue-white screening following transformation of this vector into a compatible *E. coli* strain (not used in this thesis).

2.4 Bacterial Transformation and Conjugation

2.4.1 Transformation of *E. coli*

2.4.1.1 Preparation of chemically competent cells

Calcium chloride competent *E. coli* DH5α cells for heat-shock transformation were prepared following a protocol by Chang *et al.* (*360*). In this protocol, 10 mL of sterile LB was inoculated with *E. coli*, and incubated overnight (12-16 hrs) at 37°C with shaking (200 rpm). A 250 mL conical flask containing 99 mL of sterile LB medium was inoculated with 1 mL of the overnight culture, and incubated (as above) until the OD_{600} of the culture reached 0.4 (typically 3-4 hours). Once reached, the culture was split equally across four pre-chilled 50 mL conical centrifuge tubes and cooled on ice for 20 min. Cells were pelleted by centrifugation at 4000 rpm (approximately 3700 x g (max)) for 10 min (4°C), and the supernatant was discarded. Cell pellets were resuspended

in 20 mL ice-cold 0.1 M $CaCl_2$, and incubated on ice for 30 min. Cells were harvested by centrifugation (as above), the supernatant discarded, and the cell pellets resuspended in 5mL ice-cold 0.1 M $CaCl_2$ containing 15% (v/v) glycerol. The resulting competent cells were aliquoted into 50 µL aliquots and flash-frozen in liquid nitrogen before being stored at -80°C. All centrifugation steps were performed using an Avanti J-15R benchtop centrifuge (Beckman Coulter).

2.4.1.2 Heat-shock transformation

Transformation of *E. coli* strains DH5α and DH10B with plasmid DNA was performed using heat-shock transformation. Frozen competent cell aliquots were placed on ice to thaw prior to use. Between 1-5 µL of plasmid DNA was gently mixed with 50 µL of competent cells within a sterile 1.5 mL microcentrifuge tube, and incubated on ice for 45 min. Cells were heat-shocked in a 42°C water bath for exactly 45 sec, then placed back on ice for 2 min. Following heat-shock, 750 µL of pre-warmed (37°C) SOC medium (Table 2.2) was added to the cells. Transformation reactions were then incubated at 37°C with shaking (180 rpm) for 1.5 hours to allow for the recovery of cells following transformation. After recovery, cells were plated onto LB Agar plates containing appropriate antibiotic selection (Section 2.2). For transformation with purified plasmid DNA, 100 µL aliquots of 10^0-10^{-2} serial dilutions were plated. For transformations with subcloned plasmids (i.e. heat-inactivated ligation reactions), 100 and 250 µL aliquots of the transformation mixture were plated. Plates were incubated overnight at 37°C, or until single colonies appeared.

2.4.1.3 Preparation of electrocompetent cells

Unlike DH5α and DH10B, *E. coli* BW25113/pIJ790 cells were transformed via electroporation. To generate electrocompetent *E. coli* BW25113/pIJ790 cells, a single colony was inoculated into 10 mL sterile LB containing 25 µg mL^{-1} chloramphenicol (chl) and incubated overnight at 30°C with shaking (180 rpm). The overnight culture was then subcultured, inoculating 10 mL sterile LB (+chl) with 100 µL overnight culture. The number of subcultures set up was dependent on how many transformation reactions were to be performed, with each 10 mL subculture culture yielding 100 µL of competent cells (two transformations worth). Subcultures were incubated at 30°C with shaking at 200 rpm for 4-5 hours, until an OD_{600} of approximately 0.4 was reached. When the desired OD_{600} was reached, cultures were transferred to sterile pre-chilled 15 mL conical centrifuge tubes, and cooled on ice for 5 min. Cells were harvested by centrifugation at

4000 rpm (approximately 3700 x g (max)) for 10 min at 4°C in an Avanti J-15R benchtop centrifuge (Beckman Coulter), and the supernatant discarded. Cell pellets were resuspended in 10 mL ice-cold sterile 10% glycerol, and incubated on ice for 5 min. Again, cells were harvested, resuspended in 10 mL ice-cold 10% glycerol, and incubated for a further 5 min before harvesting. Following centrifugation, the supernatant was discarded, and all remaining liquid removed by aspiration. Cells were then resuspended in sterile 10% glycerol to a final volume of 100 µL. Competent BW25113/pIJ790 cells were then kept on ice until their use shortly after generation.

2.4.1.4 Electroporation of E. coli

The transformation of BW25113/pIJ790 was performed using electroporation. Either 2-4 µL of purified PCR product or 1-2 µL of purified plasmid DNA was gently mixed with 45 µL of competent cells in a sterile 1.5 mL microcentrifuge tube (on ice). The cell mixture was then transferred to an ice-cold sterile electroporation cuvette (2 mm gap), ensuring the mixture evenly covered the bottom of the cuvette. Care was taken to avoid the introduction of air bubbles. Prior to electroporation, cuvette walls were thoroughly dried of condensation. Cells were electroporated using a BioRad Gene Pulser, set to the following conditions: 200 Ω, 25 µF and 2.5 kV (expected time constant: 4.5-4.9 ms). Immediately after electroporation, 500 µL of ice-cold sterile LB was added to the cells within the cuvette, and gently mixed. Cells were then transferred to a sterile microcentrifuge tube, and incubated at 30°C with shaking (200 rpm) for 1–1.5 hrs. Following incubation, 100-200 µL aliquots of 10^0 and 10^{-1} dilutions were spread onto LB Agar plates containing 25 µg mL^{-1} chloramphenicol and any other appropriate antibiotic selection (see Section 2.2) and incubated overnight at 30°C.

2.4.2 Conjugation of plasmids into *Rhizobium leguminosarum*

Tri-parental mating was used to conjugate plasmids into *R. leguminosarum* 3841. Plasmids were transferred from *E. coli* DH5α (donor) into *R. leguminosarum* (recipient) with the assistance of an *E. coli* DH5α helper strain possessing the helper plasmid pRK2013 (*348*) (Table 2.3). Tri-parental mating was performed using a filter-based protocol. Bacterial strains were cultured in 10 mL culture volumes, as described in Section 2.1, in the absence of antibiotic selection. *E. coli* cultures were grown overnight (14-18 hours), with *R. leguminosarum* cultures grown for 2-3 days. To set up the conjugation, 500 µL of each *E. coli* strain (donor and helper) were pooled

within a sterile 1.5 mL microcentrifuge tube. Cells were pelleted by centrifugation in an Eppendorf 5424 benchtop microcentrifuge for 3 min at 8000 rpm (approximately 6000 x g). All supernatant was removed by aspiration, and 1 mL of *R. leguminosarum* 3841 culture was added to the pellet. Cells were gently resuspended, then the cell mixture pelleted once more by centrifugation. The supernatant was removed by pouring, and the pellet resuspended in the remaining media (approximately 50 µL). The entire cell mixture was gently pipetted onto the centre of a nitrocellulose membrane filter (Whatman, 0.22 µm pore size) placed on top of a TY Agar plate (no antibiotic selection) and allowed to dry. Once dry, plates were transferred to an incubator (right-side up) and incubated at 30°C for 2 days.

Following incubation, filters were removed from the plates with sterile forceps and transferred into a sterile 50 mL conical centrifuge tube. Cells were recovered by rinsing of the membrane and subsequent resuspension in 1 mL filter-sterile (0.22 µm) 50% glycerol. Serial dilutions (10^{-1}-10^{-4}) of the recovered cells were set up, and 100 µL aliquots were plated onto TY Agar plates containing 500 µg mL^{-1} streptomycin, and any other appropriate antibiotic selection (refer to Section 2.2 for plasmid and antibiotics details). Dilution plates were incubated at 30°C until single colonies appeared (typically 3 days).

2.5 Polymerase Chain Reaction

2.5.1 Primer design

Primers were designed within the molecular biology software SnapGene (Dotmatics) and ordered from Eurofins genomics. Where possible, primers were designed to be approximately 20-22 nucleotides (nt) in length, have a GC-content of less than 60%, and a melting temperature (T_m) of between 60-65°C. To increase binding-specificity, the 3' nucleotide of all primers was either a guanine (G) or cytosine (C). For primers designed to incorporate flanking restriction enzyme recognition sites to the PCR product, 3-6 nucleotides were added upstream of the recognition site to improve digestion efficiency. As all endonucleases used in this thesis were sourced from New England Biolabs (NEB), these additional nucleotides were determined based on NEB-recommended sequences. The sequences of all primers used in this have been included in Table 2.7.

Table 2.7. Table of primers. Sequences in lowercase indicate primer overhangs, with sequences in uppercase annealing to the template sequence. Restriction enzyme recognition sequences within these overhangs have been highlighted in bold. Primers have been listed in the order of their use within this thesis.

Name	Sequence/region annealed	Sequence (5' – 3')
pK18_MCS_Fwd	pK18*mobsacB,* upstream of the multiple cloning site	GAGCGGATAACAATTTCACACAGG
pK18_MCS_Rev	pK18*mobsacB*, downstream of the multiple cloning site	GATTAAGTTGGGTAACGCCAGGG
*popZ*_KO_screen_Fwd	*R. leguminosarum* gDNA, upstream of the 5' *popZ* flank	AACGACGTCGAGATCTCCGC
*popZ*_3'-flank_Rev	Within the 800 bp 3' flanking region of *popZ*	TCGACGGCAGCGGAATCATAGG
*popZ*_5'-flank_Fwd	Within the 800 bp 5' flanking region of *popZ*	CGGGTGGCAAACATCAGGTAAC
*popZ*_KO_screen_Rev	*R. leguminosarum* gDNA, downstream of 3' *popZ* flank	CAGCATGGTCTCCGGTCGTG
*rcy*_KO_screen_Fwd	*R. leguminosarum* gDNA, upstream of the 5' *rcy* flank	TGGGACAACGAGATGCATCG
*rcy*_3'-flank_Rev	Within the 800 bp 3' flanking region of *rcy*	CGCATTCTGTACGTCGATGATC
*rcy*_5'-flank_Fwd	Within the 800 bp 5' flanking region of *rcy*	TCCGTTTGAAAAAGCGGCTATC
*rcy*_KO_screen_Rev	*R. leguminosarum* gDNA, downstream of 3' *rcy* flank	CCATTCGCTTTGATGTGAGCGG
pSUP_tet_PacI_Fwd	Tet^R-determining region of pSUP3535	ggg**ttaattaA**CAGGCCATGTTTGACAGC
pSUP_tet_PacI_Rev	Tet^R-determining region of pSUP3535	cgc**ttaattaa**GTGGCCAGGACCCAACG
*parA1*_5'-flank_Fwd	Within the 800 bp 5' flanking region of *parA1*	AAAGGCCGGGATTACCTGAGGG
egfp-parA1-Fwd	*egfp*-$Apra^R$ cassette	ctccgaggtcatccagcgagaacgtcaaaggctgg ccgcgCCGGTCGCCACCGTGAGCAAGG
egfp-parA1-Rev	*egfp*-$Apra^R$ cassette	tcgtcattcatgatctacactcacctacacccctaaa atgCATATGTGTAGGCTGGAGCTGC
egfp-popZ-Fwd	*egfp*-$Apra^R$ cassette	gcgcgaggaaatcgagcgcgtggcgcgcggcccgc gccgcCCGGTCGCCACCGTGAGCAAGG
egfp-popZ-Rev	*egfp*-$Apra^R$ cassette	cgccccggagaccggagcggcttttcaggaaaaag caatcCATATGTGTAGGCTGGAGCTGC
egfp-rcy-Fwd	*egfp*-$Apra^R$ cassette	ggtctacacgatgctcgctcatgccgcaggcaggct cggcCCGGTCGCCACCGTGAGCAAGG
egfp-rcy-Rev	*egfp*-$Apra^R$ cassette	cacgtggtcaagcggctttccaatcggcgggtacca gtcaCATATGTGTAGGCTGGAGCTGC

2.5.2 High-fidelity amplification of DNA

For the amplification of DNA fragments during molecular cloning, Polymerase Chain Reaction (PCR) was performed using Phusion™ High-Fidelity DNA Polymerase (Thermo Fisher Scientific). This enzyme was chosen due to its low error rate, accredited to the 3'→5' exonuclease activity of this enzyme. PCR reactions were set up on ice in 0.5 mL flat-capped PCR tubes as described in Table 2.8. Phusion™ polymerase is compatible with two different buffers: HF and GC. The Phusion HF buffer was used for routine DNA amplification, while the GC buffer was used for the amplification of sequences with a total GC content of >60%, or those containing multiple GC-rich regions. PCR cycling parameters of a standard PCR reaction are outlined in Table 2.9. For amplifications with primers designed to introduce flanking sequences to the PCR product (i.e. overhang primers), a modified version of this cycle was used, outlined in Table 2.10. Here, two annealing temperatures were used to ensure efficient annealing of these primers to the template DNA. First, PCR reactions were subjected to 10 PCR cycles with an annealing step at the annealing temperature (T_a) required for the region of the primer complimentary to the template DNA. Then, once the flanking nucleotide sequences were incorporated into template, reactions were subjected to a further 15 cycles with the T_a of the full-length primer. All PCR reactions were performed in a Bio-Rad C1000 Touch Thermal Cycler. Following thermal cycling, 2-3 µL of each PCR reaction was analysed by agarose gel electrophoresis to determine whether successful amplification of the template DNA occurred (Section 2.3.2).

Table 2.8. Constituents of a Phusion polymerase PCR reaction. High-fidelity polymerase chain reactions (PCRs) were performed using Phusion DNA polymerase (Thermo Fisher Scientific). Phusion HF Buffer was used for routine amplification, while Phusion GC Buffer was used for the amplification of GC-rich template sequences. Both buffers provide a $MgCl_2$ concentration of 1.5 mM in the final reaction, so no additional $MgCl_2$ was added to the reaction volume.

Component	Amount (per 50 µL reaction)
Genomic DNA or Plasmid DNA	Up to 250 ng Up to 10 ng
5X Phusion HF Buffer or 5X Phusion GC Buffer	10 µL
dNTPs, 25 mM each	0.4 µL
Forward primer, 10 µM	2.5 µL
Reverse Primer, 10 µM	2.5 µL
DMSO	0.75 µL
Phusion DNA Polymerase (2u/µL)	0.5 µL
Nuclease-free water	Up to 50 µL

Table 2.9. Thermal cycling parameters for a standard Phusion PCR reaction. Annealing temperature (T_a) (X) was calculated using the Thermo Fisher Scientific T_m Calculator. Typically for Phusion DNA polymerase, T_a is equivalent to the lower melting temperature (T_m) of the two primers.

Step	Temperature (°C)	Duration	No. of cycles
Initial denaturation	98	30 sec	1
Denaturation	98	10 sec	25
Annealing	X	30 sec	
Extension	72	30 sec per kb	
Final extension	72	5 min	1
Hold	4	∞	-

Table 2.10. Thermal cycling parameters for a Phusion PCR reaction with extension primers. For amplifications with extension primers, two distinct cycle parameters were used within the PCR reaction, differing by annealing temperature (T_a). In this table, X denotes the T_a of the primer regions complimentary to the initial DNA template, and Y denotes the T_a of the full-length primer. Annealing temperatures were calculated using the Thermo Fisher Scientific T_m Calculator.

Step	Temperature (°C)	Duration	No. of cycles
Initial denaturation	98	30 sec	1
Denaturation	98	10 sec	10
Annealing	X	30 sec	
Extension	72	30 sec per kb	
Denaturation	98	10 sec	15
Annealing	Y	30 sec	
Extension	72	30 sec per kb	
Final extension	72	5 min	1
Hold	4	∞	-

2.5.3 Diagnostic colony PCR (GoTaq)

Both standard PCR and colony PCR (cPCR) were used routinely to verify both plasmid and genomic DNA. For diagnostic PCR reactions, GoTaq® G2 Hot Start Taq Polymerase (Promega) was used. Typically, up to 0.25 µg template DNA was amplified in a 25 µL reaction volume (Table 2.11). For cPCR, template DNA was prepared by picking a small volume of a bacterial colony and resuspending in 10 µL of nuclease-free water. Samples were incubated at 95°C for 5 min, then briefly centrifuged to pellet cell debris. Once cooled, 1 µL of the supernatant was used as the template DNA for cPCR. All reactions were performed within Green GoTaq® Flexi Reaction Buffer (Promega). This buffer contains two dyes and increases sample density, allowing for samples to be loaded directly onto an agarose gel without the need to add additional loading dye. The cycle parameters for these reactions have been detailed in Table 2.12. Following cycling, up to 10 µL of each PCR reaction was analysed by agarose gel electrophoresis (Section 2.3.2).

Table 2.11. Constituents of a diagnostic PCR using GoTaq G2 Hot Start Polymerase. Diagnostic PCR reactions were performed on both purified DNA and bacterial colony template using GoTaq® G2 Hot Start Polymerase (Promega). Although multiple reaction buffers are compatible with this polymerase, Green GoTaq® Flexi Buffer (Promega) was used to allow for PCR reactions to be loaded directly onto agarose gels for electrophoresis.

Component	Amount (per 25 µL reaction)
Template DNA	<0.25 µg (or 1 µL diluted colony)
5X Green GoTaq Flexi Buffer	5 µL
$MgCl_2$ Solution, 25mM	2 µL
PCR Nucleotide Mix, 25mM each	0.2 µL
Forward primer, 10 µM	2 µL
Reverse primer, 10 µM	2 µL
DMSO	0.75 µL
GoTaq G2 Hot Start Polymerase (5u/µL)	0.125 µL
Nuclease-free water	Up to 25 µL

Table 2.12. Cycle parameters of PCR reactions using GoTaq® G2 Hot Start Polymerase. The primer annealing temperature (T_a) (X) used within this PCR reaction was approximately 5°C below the melting temperature (T_m) of the primers. Primer melting temperatures were determined using the Promega T_m calculator, which takes into consideration the polymerase and buffer system used within the reaction to calculate an accurate T_m.

Step	Temperature (°C)	Duration	No. of cycles
Initial denaturation	95	2 min	1
Denaturation	95	45 sec	35
Annealing	X	30 sec	
Extension	72	1 min per kb	
Final extension	72	5 min	1
Hold	4	∞	-

2.5.4 Purification of PCR products

For PCR reactions that yielded multiple DNA fragments (i.e. non-specific amplification occurred), the PCR products within the remaining reaction volume were separated by gel electrophoresis, and the desired fragment was excised and purified from the gel (Section 2.3.3). For PCR reactions where no non-specific amplification occurred, PCR products were purified using the QIAquick PCR Purification Kit (QIAGEN). This kit uses silica membrane columns to remove primers, nucleotides, polymerase, and other PCR reaction constituents from the DNA samples. PCR reactions are combined with a high-salt binding buffer and applied onto a spin column. Upon application to the column, the high-salt conditions provided by the buffer allow for the binding of DNA to the silica membrane. Membranes are then washed with an ethanol-based wash buffer to remove impurities. Following washing, the purified DNA was eluted into 30 µL nuclease-free water. Where increased concentrations were required, DNA was eluted into 15-20 µL of nuclease-free water.

2.6 Synthesis of plasmids using traditional cloning

The majority of plasmids used throughout this thesis were generated using traditional molecular cloning techniques. Wherever possible, directional cloning of inserts into vectors/plasmids was performed. The exception to this was the introduction of a tetracycline resistance gene into pK18*mobsacB*-derived allelic exchange vectors (Section 2.6.2). Vector and insert DNA was prepared for cloning by overnight restriction enzyme digestion (Section 2.3.4), with all restriction enzymes used within this thesis generating cohesive ('sticky') ends. Digested DNA fragments were isolated and purified by agarose gel electrophoresis and subsequent gel purification (Sections 2.3.2 and 2.3.3). Following purification, insert DNA fragments were ligated into linearised plasmids within ligation reactions catalysed by T4 DNA ligase (Section 2.3.5). The ligated plasmids were then transformed into *E. coli* (typically DH5α) via heat-shock transformation (Section 2.4.1). Finally, transformation reactions were plated on LB Agar plates containing appropriate antibiotic selection and any other applicable screening reagents to identify cells carrying plasmids possessing the correct insert (Section 2.2). A list of all plasmids generated within this thesis can be found in Table 2.7.

2.6.1 Generation and blue-white screening of pK18*mobsacB*-derived allelic exchange vectors

Synthetic allelic exchange constructs for the generation of gene knockouts (described in detail in Section 3.2.1) were subcloned from their delivery vector (EC41308) into the multiple cloning site (MCS) of pK18*mobsacB* as either XmaI-HindIII (*parA1*, *popZ*) or EcoRI-BamHI (*rcy*) fragments. The *lacZα* fragment of pK18*mobsacB*, which encodes the first 146 amino acids of the enzyme ß-galactosidase, enabled blue-white pre-screening of *E. coli* DH5α and DH10B colonies carrying this vector. Following transformation of *E. coli* with ligated plasmid, cells were plated onto LB Agar plates containing 50 µg mL^{-1} kanamycin (resistance conferred by pK18*mobsacB*), 200 µg mL^{-1} IPTG (Isopropyl β-D-1-thiogalactopyranoside) and 1 mM X-gal (5-bromo-4-chloro-3-indolyl-β- D-galactopyranoside). The *E. coli* strains DH5α and DH10B possess the *lacZΔM15* deletion mutation, meaning they lack the first 146 amino acids of the ß-galactosidase sequence (i.e. the α subunit). When pK18*mobsacB* is taken up by these cells, α-complementation of the β-galactosidase fragment within DH5α/DH10B forms a functional β-galactosidase enzyme. IPTG-induced expression of this enzyme results in the hydrolysis of X-gal, forming 5-bromo-4-chloro-indoxyl. Spontaneous dimerization of this product produces an insoluble blue pigment (5,5'-dibromo-4,4'-dichloro-indigo). As the MCS of pK18*mobsacB* lies within the *lacZα* fragment, introduction of an insert into this site disrupts this fragment, preventing α-complementation of β-galactosidase and subsequent hydrolysis of X-gal. Cells carrying pK18*mobsacB* plasmids possessing an insert therefore remain white. Blue-white screening plates were incubated overnight at 37°C until blue and white colonies appeared. White colonies were then picked for further screening. Successful ligation of the artificially synthesised allelic exchange constructs into pK18*mobsacB* generated the vectors pK18-*parA1*-KO, pK18-*popZ*-KO and pK18-*rcy*-KO (used for the generation of gene knockouts within Chapter Three).

2.6.2 Introduction of a tetracycline resistance gene into pK18*mobsacB*-derived allelic exchange vectors

The tetracycline resistance (tet^R) gene of pSUP3535 (*tetA*) was PCR amplified with the primers pSUP_tet_PacI_Fwd and pSUP_tet_PacI_Rev (Table 2.7). These primers were designed to incorporate 5' and 3' flanking PacI recognition sequences to the PCR product, allowing for the cloning of this gene as a PacI-PacI fragment into the PacI site located between the homologous flanking regions of the pK18*mobsacB*-derived allelic exchange vectors (see Section 3.3.1). The

purified tet^R PCR product (Section 2.5.4) and the vectors pK18-*parA1*-KO, pK18-*popZ*-KO and pK18-*rcy*-KO were digested with PacI to generate complimentary overhangs. Following ligation of the PacI-digested tet^R PCR product into these vectors, ligated plasmids were transformed into *E. coli* and plated onto LB Agar plates containing 50 µg mL^{-1} kanamycin and 10 µg mL^{-1} tetracycline to select for colonies carrying plasmids possessing the tet^R insert. As this cloning was non-directional, further screening to determine the orientation of tet^R gene within these plasmids was performed. Clones possessing a tet^R cassette in the reverse orientation were used within this thesis, referred to as pK18-*parA1*-tet^R-KO, pK18-*popZ*-tet^R-KO and pK18-*rcy*-tet^R-KO.

2.6.3 Cloning of pIND6-derived expression plasmids

To generate the expression plasmids pIND6-*parA1*, pIND6-*popZ* and pIND6-*rcy*, synthetic gene expression constructs (described in detail in Section 4.2.2) were subcloned from their delivery vector (pUC57) into the multiple cloning site of pIND6 as XbaI-BamHI fragments. To generate complimentary overhangs, pIND6 was digested with XbaI and BglII. Although their recognition sequences differ, BamHI and BglII both generate an identical overhang sequence (GATC). Following ligation, plasmids were transformed into *E. coli* DH5α, plated onto LB Agar containing 50 µg mL^{-1} kanamycin and incubated at 37°C until transformant colonies appeared.

2.7 EGFP-tagging using phage lambda (λ) Red-mediated homologous recombination

Translational enhanced green fluorescent protein (EGFP) fusions, used for protein localisation studies within *R. leguminosarum* (Chapter Five), were generated by introducing an *egfp*-$Apra^R$ (apramycin resistance) cassette into the plasmids pIND6-*parA1*, pIND6-*popZ* and pIND6-*rcy*. Integration of this cassette into these plasmids was achieved using phage lambda (λ) Red-mediated homologous recombination. The protocol used in this thesis is an adaptation of a gene disruption protocol initially developed by Datsenko and Wanner (2000) and modified by Gust *et al.* (2002) (*349*, *358*). A gene cassette comprising of the *egfp* gene and an $Apra^R$ gene (*aac(3)-IV*) was amplified with the primer pairs *egfp_parA1*_Fwd + *egfp_parA1*_Rev, *egfp_popZ*_Fwd + *egfp_popZ*_Rev and *egfp_rcy*_Fwd + *egfp_rcy*_Rev. These overhang primers were designed to introduce 40 bp flanking sequences to the *egfp*-$Apra^R$ cassette, homologous to the target regions

of the pIND6-derived plasmids. The sequences of these primers can be found in Table 2.7. PCR reactions were performed as described in Section 2.5.2.

Homologous recombination-mediated plasmid modification was performed within the *E. coli* strain BW25113/pIJ790 (Table 2.1). First, the plasmids pIND6-*parA1*, pIND6-*popZ* and pIND6-*rcy* were electroporated into BW25113/pIJ790. The generation and electroporation of electrocompetent BW25113/pIJ790 cells is described within Section 2.4.1. Following electroporation, cells were plated on LB Agar containing 50 μg mL^{-1} kanamycin (kan) and 25 μg mL^{-1} chloramphenicol (chl) to select for cells carrying pIND6-derived plasmids, and incubated overnight at 30°C. A single colony was transferred into 10 mL LB (50 μg mL^{-1} kan, 25 μg mL^{-1} chl), and incubated overnight at 30°C (180 rpm). A 100 μL aliquot of each overnight culture was then transferred into 10 mL SOB medium (Table 2.2) containing kan and chl (as above) and 10 mM L-arabinose. The presence of L-arabinose induces the expression of the λ Red genes (*beta*, *gam*, *exo*) encoded on pIJ790. Cultures were grown for 3-4 hours at 30°C with shaking (200 rpm) until an OD_{600} of approximately 0.4 was reached. Once cultures reached the required density, electrocompetent cells were prepared (as described within Section 2.4.1).

Extended *egfp*-Apra^R cassettes were electroporated into BW25113/pIJ790 cells carrying the corresponding pIND6-derived expression vector. Recombination between homologous regions of the extended *egfp*-Apra^R cassettes and the pIND6-derived plasmids, mediated by the products of the λ Red genes, resulted in integration of the *egfp*-Apra^R cassette immediately downstream of the target gene generating a translational fusion. Cells were plated onto LB Agar containing kan (as above) and 50 μg mL^{-1} apramycin (apra) and incubated overnight at 37°C. Incubation at 37°C instead of 30°C promotes the loss of the temperature sensitive plasmid pIJ790 from the cells. As the integration of *egfp*-Apra^R will not occur within every copy of the pIND6-derived plasmid within a cell, plasmid DNA was isolated from a single colony and transformed into *E. coli* DH5α (heat-shock, Section 2.4.1). These transformations were plated onto LB Agar containing kan and apra (as above), and incubated overnight at 37°C to select for DH5α cells carrying pIND6-derived plasmids with successful integrations of *egfp*-Apra^R. Further details of this methodology can be found within Chapter Five (Section 5.2).

2.8 Homologous recombination-based gene deletion in *R. leguminosarum*

The generation of *R. leguminosarum* gene knockout mutants was attempted using a homologous recombination-based approach, mediated by the sucrose-sensitive cloning vector pK18*mobsacB* (*354*). Allelic exchange constructs, designed for the generation of whole-gene *parA1*, *popZ* and *rcy* deletions within *R. leguminosarum*, were cloned into pK18*mobsacbB*. The design and generation of these allelic exchange vectors has been covered extensively in Chapter Three. In summary, two types of allelic exchange vectors were generated: unmarked, and those marked with a tetracycline-resistance (tet^R) cassette. All allelic exchange constructs possessed 5' and 3' 800 bp flanking regions, homologous to the 800 bp immediately upstream and downstream of the target gene respectively. Located between these flanking regions was either a short nucleotide sequence containing a PacI recognition site (unmarked), or a tet^R cassette sequence (marked). Homologous recombination between both the 5' and 3' flanking regions of the cassette and the *R. leguminosarum* genome results in the replacement of the target gene with the nucleotide sequence located between the flanking regions of the exchange construct, leading to the deletion of the gene from the genome.

2.8.1 Integration of allelic exchange vectors into the *R. leguminosarum* genome

The methodology for the generation of marked and unmarked gene knockouts differed only by the antibiotic selection used. Firstly, pK18*mobsacB*-derived allelic exchange vectors (see Table 2.3) were conjugated into *Rhizobium leguminosarum* 3841 as described in Section 2.4.2. Following recovery of transconjugants (Section 2.4.2), 100 µL aliquots of 10^0-10^{-4} serial dilutions were plated onto TY Agar plates containing 500 µg mL^{-1} streptomycin (str) and 200 µg mL^{-1} kanamycin (kan). For the generation of tet^R knockout mutants, 5 µg mL^{-1} tetracycline was added to these plates. The antibiotic selection used here selected for *R. leguminosarum* cells possessing a primary integration (single crossover) of the allelic exchange vector into the genome. The vector pK18*mobsacB* confers kanamycin resistance. However, it cannot replicate within *R. leguminosarum*. Therefore, the only way for *R. leguminosarum* cells to survive in the presence of kanamycin is to undergo a homologous recombination event in which pK18*mobsacB* is integrated into the genome (i.e. a primary integration). Serial dilution plates were incubated at 30°C until single primary integration colonies appeared (typically 3-4 days).

2.8.2 Generation of gene knockout mutants

For gene deletion to occur, a second recombination event between the integrated pK18*mobsacB*-derived vector and the *R. leguminosarum* genome is required. To drive this recombination event, between 3 and 5 primary integration colonies were transferred into 10 mL TY medium (without antibiotic selection) and cultured at 30°C for 2 days (180 rpm). Serial dilutions of the non-selective cultures were set up (10^0–10^{-3}), and 100 µL of each plated onto TY-sucrose (10%) plates (Table 2.2) containing 500 µg mL^{-1} streptomycin. For the generation of tet^R-marked mutants, 5 µg mL^{-1} tetracycline was added to these plates. Sucrose counter-selection drives the excision of pK18*mobsacB* from the genome via a second recombination event (i.e. double-crossover). The *sacB* gene of pK18*mobsacB* encodes the enzyme levansucrase. This enzyme breaks down sucrose into a toxic product, therefore retention of pK18*mobsacB* in the genome results in cell death. This second recombination event has one of two possible outcomes: whole-gene deletion, or reversion to the wildtype genotype. For unmarked mutants, selection with streptomycin allows for the growth of both knockout mutants and wildtype revertant colonies. For tet^R mutants, the addition of tetracycline to these plates selects against wildtype revertant colonies.

TY-sucrose plates were incubated at 30°C until single colonies appeared. Between 50 and 100 colonies from these plates were then transferred to a TY master plate containing 500 µg mL^{-1} streptomycin. For tet^R mutant generation, 5 µg mL^{-1} of tetracycline was added to the master plate. To screen for any primary integration colonies which may have survived sucrose counter-selection, master plates were replicated onto TY Agar plates containing streptomycin and kanamycin (unmarked) or streptomycin, kanamycin, and tetracycline (marked). Master and replica plates were incubated at 30°C for 2 days. Kanamycin-sensitive colonies were taken forward for further screening (i.e. colony PCR) to validate their genotype. Any modifications to, or deviations from, this protocol have been stated within the relevant results chapters.

2.9 Polyacrylamide gel electrophoresis

2.9.1 Sample preparation

Confirmation of protein expression was determined by polyacrylamide gel electrophoresis (PAGE) of *R. leguminosarum* culture samples. To prepare samples for PAGE analysis, cells were harvested from 1 mL of culture by centrifugation at 8000 rpm for 3 min (6000 x g). Supernatant was discarded, and the resulting cell pellets were resuspended in 300 µL of 10 mM Tris, 5 mM $MgCl_2$ resuspension buffer. Aliquots of these resuspensions were then standardised based on the OD_{600} of the cultures at the point of sample collection to ensure roughly equal total protein within all samples. For cultures where the OD_{600} at the point of sampling was approximately 1.0, 10 µL of cell resuspension was taken. The volume of resuspension taken for cultures with different OD_{600} values was adjusted accordingly (e.g. 20 µL of resuspension taken from a culture of OD_{600} = 0.5). The final volume of all samples was then equalised with resuspension buffer. Prior to loading, Laemmli (sample) buffer (Table 2.13) was added to each sample (5 µL per 20 µL of sample). Samples were heated to 95°C for 5 min to lyse cells. Once cooled, samples were centrifuged briefly to separate excess cell debris from the soluble protein samples.

2.9.2 Buffer Composition

The composition of all buffers used throughout the analysis of proteins have been detailed in Table 2.13.

Table 2.13. Composition of buffers used for the analysis of proteins.

Buffer	Composition per 1 litre of de-ionised water
5X Laemmli buffer (sample buffer)	For 20 mL: 5 mL 0.5 M Tris (pH 6.8), 0.8 g sodium dodecyl sulfate (SDS), 4 mL glycerol, 4 mL 1M dithiothreitol (DTT), 0.4 mg Bromophenol blue
SDS-PAGE buffer	3.03 g Tris, 14.41 g glycine, 1 g SDS
Transfer buffer	4.66 g Tris, 2.34 g glycine, 1 mL 10% SDS, 200 mL ethanol, pH 9.2
Tris-Buffered Saline (TBS)	1 mL 1M Tris, 15 mL 5M NaCl, pH 7.5
TBS-Tween (TBS-T)	1 mL 1M Tris, 15 mL 5M NaCl, 1 mL Tween 20, pH 7.5

2.9.3 Separation of proteins using SDS-PAGE

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was used for the separation of proteins from *R. leguminosarum*. Discontinuous polyacrylamide gels (8.3 x 7.3 x 1.5 cm) comprising of a resolving and stacking gel were cast using the Bio-Rad Mini-PROTEAN® Tetra Handcast System. The composition of these gels is detailed in Table 2.14. Typically, 10% (v/v) acrylamide resolving gels were used. However, where required, the acrylamide percentage within the resolving gel was increased to 15% to optimise the resolution of smaller proteins (around 10kDa). Once solidified, gels were placed within an electrode assembly and transferred to a Mini-PROTEAN® Tetra cell electrophoresis tank (Bio-Rad). Gel assemblies were filled with freshly-prepared SDS-PAGE buffer (Table 2.13), ensuring gels were fully submerged. The electrophoresis tank was then filled with buffer to the required level (550 mL for 2 gels or 680 mL for 4 gels). Following sample preparation (Section 2.9.1), 20 µL of soluble protein sample was loaded into each well, alongside 5 µL of BLUeye Prestained Protein Ladder (Sigma-Aldrich). Gels were run at room temperature at a constant voltage of 100 V for approximately 1 hr. For polyacrylamide gels where the transfer of protein samples to membranes (i.e. for western blotting) was not required, proteins were visualised by Coomassie staining. SDS-PAGE gels were incubated in InstantBlue® Coomassie Protein Stain (Expedeon) for at least 20 min with gentle agitation. If required, gels were destained by washing in dH_2O (with gentle shaking) until the background staining was reduced.

Table 2.14. Composition of a 10% (v/v) SDS-PAGE polyacrylamide gel.

Component	Resolving gel (10%)	Stacking gel (5%)
1.5 M TRIS pH 8.8	2.5 mL	-
0.5 M TRIS pH 6.8	-	0.75 mL
dH_2O	3.95 mL	1.7 mL
10% SDS	0.1 mL	30 µL
30% Acrylamide/Bisacrylamide solution	3.35 mL	0.5 mL
Tetramethylethylenediamine (TEMED)	5 µL	3 µL
10% ammonium persulphate (APS)	100 µL	30 µL

2.9.4 Anti-His western blot analysis

Following electrophoresis, proteins were transferred from polyacrylamide gels to polyvinylidene difluoride (PVDF) membranes (Amersham™ Hybond® P, GE Healthcare) using a wet electroblotting system. PVDF membranes were activated in ethanol for 3 min, rinsed with dH_2O to remove any excess ethanol, and equilibrated in transfer buffer (Table 2.13) for 5 min prior to assembling in the transfer cassette. Transfer cassettes were stacked as follows: (cathode) sponge, two Whatman blotting papers, acrylamide gel, PVDF membrane, two Whatman blotting papers, sponge (anode). To avoid air bubbles, all components were soaked in transfer buffer before their assembly into the stack. Assembled transfer cassettes were inserted into the transfer tank and submerged in pre-chilled (5°C) transfer buffer (Table 2.13). The transfer of proteins was performed at a constant voltage of 100 V for 1 hr 30 min at 5°C.

Following transfer, membranes were rinsed with TBS-T (Table 2.13) for 5 min to remove any residual transfer buffer. All membrane incubation steps were performed on a rocking platform shaker with gentle rocking. To confirm successful transfer of proteins, membranes were incubated in Ponceau S Solution (Expedeon) for 5 min. Following imaging (Syngene G:Box F3 imager), membranes were subsequently destained in TBS-T (Table 2.13) for 5 min. Membranes were then transferred to a blocking solution of TBS-T containing 5% (w/v) skimmed milk powder and incubated for 1 hr. Following blocking, membranes were washed twice in TBS-T (2 min incubation per wash) and transferred into TBS-T containing a 1:500 dilution of Anti-His_6-Peroxidase Monoclonal Antibody (Sigma-Aldrich). Membranes were incubated in the antibody solution for 1 hr, then washed with three 15 min incubations in TBS-T. The peroxidase-catalysed chemiluminescent signal was generated by the application of Amersham ECL™ Prime Western Blotting Detection Reagent (GE Healthcare Life Sciences) to the membranes. Equal volumes of Solution A (luminol solution) and Solution B (peroxide solution) were combined, and 1 mL of the solution pipetted directly across the membrane. The resulting chemiluminescent signal was detected using a Syngene G:Box F3 Gel & Blot imaging system.

2.10 Microscopic analysis of samples

2.10.1 Slide preparation

To minimise physiological stresses arising from conventional fixing techniques, liquid *R. leguminosarum* culture samples were mounted on agarose pad microscopy slides. Agarose slices were generated following a protocol by Skinner *et al.* (*361*). Here, 1.5% (w/v) agarose gel slices were made by melting 0.3 g high-purity low melting-point agarose in 20 mL filter-sterilised dH_2O. The use of high-purity agarose was of particular importance for epifluorescence microscopy to avoid background fluorescence signal from impurities within the agarose. Gel slices were trimmed to approximately 15 x 15 mm with a scalpel and transferred to the centre of a microscope slide (75 x 25 mm, 1 mm thick). A border of vacuum grease was syringed around the agarose pad, and 3 µL of culture sample pipetted onto the centre of the pad. A cover slip (22 x 22 mm, 0.13-0.16 mm thick) was then placed on top, and gentle pressure applied to secure the cover slip in place.

Where microscopic analysis of bacterial samples from solid media was required, a small volume (approximately 2 µL) of dH_2O was placed in the centre of a microscope slide. A small amount of bacterial sample was taken with a sterile toothpick, and thoroughly mixed with the water on slide. A cover slip was placed on top, and the edges sealed with transparent nail varnish. This mounting technique was not used to generate any of the microscopy images within this thesis, but was used within Section 5.2.2 to check for EGFP signal within *E. coli*.

2.10.2 Microscopy

All microscopy was performed using a Carl Zeiss Axio Scope.AI upright microscope coupled with an Axiocam MRC camera. To visualise bacterial samples, an A-Plan 100x 1.25 Oil Ph3 objective lens was used. Images were captured within the ZEISS microscopy software ZEN. To observe cell morphology, cell samples were visualised using phase contrast microscopy. The localisation of EGFP-tagged proteins within *R. leguminosarum* was visualised by epifluorescence microscopy, using a GFP fluorescence filter set (450-490 nm). EGFP has a single excitation peak of 488 nm, and an emission peak of 507 nm.

2.10.3 Image analysis

Microscopy images were processed using the image processing software ImageJ. This software was also used to obtain cell measurement data. To generate accurate cell size data, any out-of-focus regions were cropped from the microscopy images to be analysed. Images were adjusted to enhance the contrast between the cells and the background. Cell measurements were then obtained using the following analysis workflow: Image type, 8-bit > Adjust threshold, B&W > Convert to mask > Watershed > Analyse Particles (all sizes). On occasion, this workflow interpreted single cells as two separate cells, or two closely neighbouring cells as one cell. Therefore, all analysed images were compared to their original source image, and data from incorrectly identified 'cells' was omitted from the dataset.

2.11 Statistical analysis and presentation of cell morphology data

The optimal number and width of histogram bins was determined using Freedman–Diaconis rule, calculated using data from the control sample dataset. Using Freedman–Diaconis rule, optimal bin width was determined to be 0.0638 (μm), with 29 bins. To improve the visual presentation of this data, bin width was rounded to 0.08 μm (26 bins). Bin sizes and numbers were kept consistent across all samples to facilitate direct visual comparisons of the datasets. For the generation of box-and-whisker blots, exclusive interquartile ranges were calculated. Kolmogorov-Smirnov tests for normality were performed on all datasets to determine appropriate statistical comparisons of the datasets (i.e. parametric t-tests vs non-parametric Mann-Whitney U tests). Median values were compared using independent samples median tests, with Yate's continuity correction. All statistical analysis was conducted using IBM SPSS Statistics software.

Chapter Three – Characterisation of ParA1, PopZ and Rcy by whole-gene deletion.

3.1 Introduction

To begin to determine the role of ParA1, PopZ and Rcy within *Rhizobium leguminosarum* and elucidate their potential involvement in unipolar growth, the generation of whole-gene deletion mutants was attempted. Should these proteins be required for the coordination and control of *Rhizobium* growth, loss of the protein coding sequences from the *R. leguminosarum* 3841 genome could be expected to impair cell growth, division, and possibly morphology. This chapter describes how a homologous recombination-based approach was used to attempt to generate whole-gene *parA1*, *popZ* and *rcy* knockout mutants. Initially, the generation of 'unmarked' knockout strains was attempted, in which entire protein coding sequence (7128 bp for *rcy*, 882 bp for *popZ*, 795 bp for *parA1*) was replaced with a short (15 bp) nucleotide sequence containing a novel PacI restriction site. Following difficulties with the generation and isolation of unmarked knockout mutants, this experimental approach was redesigned to enable the generation of mutants marked with a tetracycline resistance cassette, intended to facilitate mutant screening.

3.2 Attempted generation of unmarked knockout mutants

Initially, the generation of unmarked, whole-gene *parA1*, *popZ*, and *rcy R. leguminosarum* 3841 knockout mutants was used to facilitate the characterisation of the proteins encoded by these genes. By generating unmarked gene deletions, possible polar effects on the expression of neighbouring genes arising from the introduction of an antibiotic resistance cassette (or other marker) into the genome can be avoided (*362*). In addition, whole-gene deletion was chosen over partial gene deletion or the generation of truncation mutations. By deleting the entire gene length, altered or retained functionality of any residual gene product is avoided. This increases the certainty that any phenotypes observed in the resulting knockout strains were the direct consequence of gene loss.

To generate *parA1*, *popZ* and *rcy* knockout strains, a plasmid-mediated homologous recombination-based method of gene knockout was adopted. The first step in this approach is

the generation of an allelic exchange construct, also referred to as a knockout construct. These target-specific constructs consist of a nucleotide sequence homologous to the upstream (5') flanking region of the gene of interest, followed by a sequence (of equal length) homologous to the downstream (3') flanking region (Figure 3.1A). This construct is then cloned into a suitable vector and introduced into the host organism.

For successful integration of the vector into the host genome, and subsequent excision of the target sequence from the genome (i.e. gene deletion), the vector of choice must contain both a selectable and counter-selectable marker (*356*). Importantly, the vector must not be able to replicate within the host organism to ensure genomic integration of the vector and not independent plasmid retention. In this study, the sucrose-sensitive vector pK18*mobsacB* was chosen (*354*) (Figure 3.1A, Figure 3.2). The use of pK18*mobsacB*-mediated gene deletion has been successfully employed to delete a number of genes within *R. leguminosarum* including those encoding riboflavin and nickel transporters, calcium-binding proteins, a cation diffusion facilitator protein, and a diaminobutyrate aminotransferase (*363–367*). The *nptII* gene of pK18*mobsacB*, encoding neomycin phosphotransferase II, confers kanamycin resistance (*368*). However, as pK18*mobsacB* cannot replicate within *Rhizobium*, the only way for cells to survive in the presence of kanamycin is by the integration of the vector into the host genome via a single homologous recombination event (i.e. a single crossover) (Figure 3.1B). Following the integration of the vector into the genome, sucrose selection can then be used to drive a second recombination event. The sucrose-sensitivity of pK18*mobsacB* is the result of the expression of *Bacillus subtilis* levansucrase, encoded by *sacB*. This enzyme catalyses the conversion of sucrose to levan, which when accumulated becomes toxic to the cell (*369*, *370*). Therefore, the presence of sucrose drives a second recombination event which excises the vector backbone and hence the *sacB* gene from the genome (i.e. a double-crossover event), allowing for the survival of the cell (Figure 3.1C). If this second recombination occurs in the same flank as the primary integration occurred, only the vector is excised, leaving the native gene intact and restoring the parental genotype. However, if the second recombination even occurs within the opposite flank, excision of the vector from the genome also results in the loss of the native gene, resulting in a stable genomic knockout mutant.

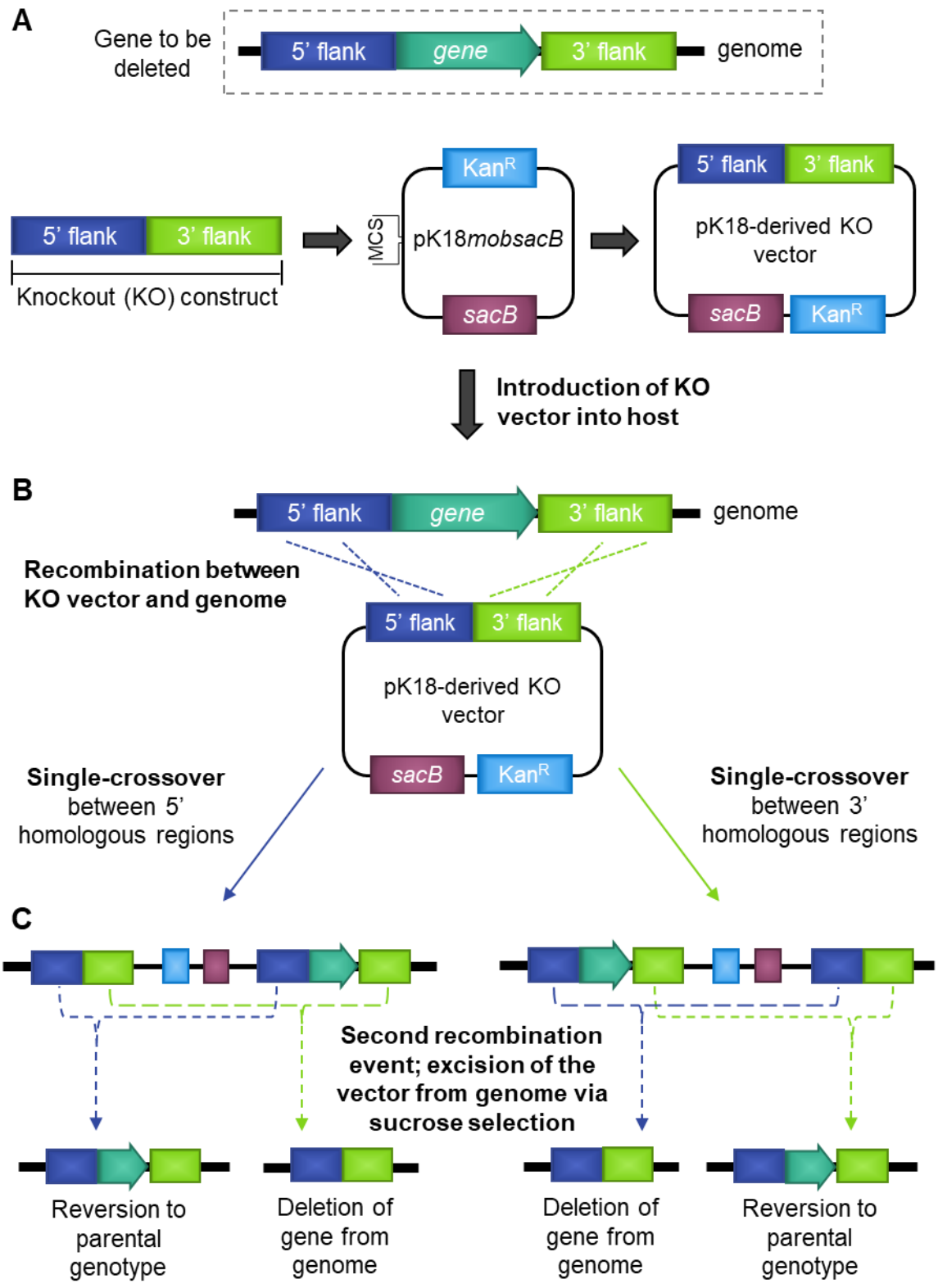


Figure 3.1. Schematic representation of the generation of whole-gene knockout mutants via homologous recombination using pK18mobsacB. (A) Generation of the knockout (KO) vector. A KO construct, comprising of sequences homologous to the target gene flanking regions, is cloned into the multiple cloning site (MCS) of pK18*mobsacB*. (B) Integration of the KO vector into the host's genome as the result of a homologous recombination event. (C) Excision of the vector from the genome following a second recombination event, resulting in either KO mutant generation or reversion to the parental genotype.

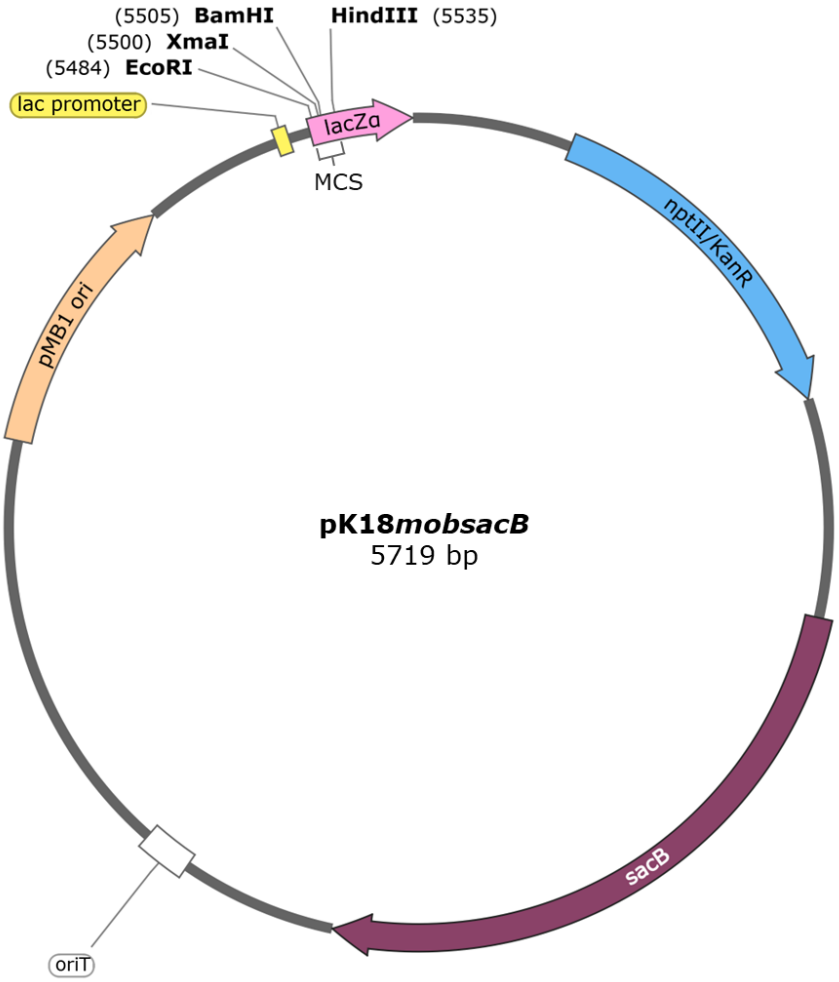


Figure 3.2. Plasmid map of the mobilizable sucrose-sensitive suicide vector pK18*mobsacB*. Allelic exchange constructs for the generation of knockout mutants via homologous recombination were cloned into the multiple cloning site (MCS) of this vector. Relevant restriction enzyme recognition sites have been shown on this map. The MCS lies within *lacZa* (pink), the alpha(α) fragment of the *lacZ* gene (complete *lacZ* gene encodes ß-galactosidase). Kanamycin resistance (KanR) is conferred by *nptII* (blue) (encodes neomycin phosphotransferase II). The *sacB* gene (burgundy) which encodes the *Bacillus subtilis* levansucrase enzyme confers sucrose-sensitivity in cells carrying the vector. Other features are as follows; pMB1 ori (orange): *E. coli* origin of replication, oriT (white): origin of transfer, *lac* promoter (yellow).

3.2.1 Design and synthesis of unmarked knockout vectors

In this study, the allelic exchange constructs used for the generation of gene knockouts were designed *in-silico* and generated using commercial DNA synthesis (Section 2.3.7). The homologous flanking regions of knockout constructs comprised of nucleotide sequences identical to the 800 bp regions immediately flanking the gene of interest (*parA1*, *popZ*, *rcy*) (Figure 3.3A). A novel PacI recognition site (TTAATTAA) was introduced between the two homologous regions, flanked by a start (ATG) and stop (TGA) codon (5' and 3' of the PacI site respectively). An additional nucleotide (guanine, G) was introduced between the end of the start codon and the PacI recognition site to ensure that the construct remained in-frame. These features totalled to a 15 bp sequence located between the two flanks. Finally, restriction endonuclease sites were introduced to the 5' and 3' ends of the knockout construct to allow for the cloning of these constructs into the multiple cloning site (MCS) of pK18*mobsacB*. Here, XmaI (5') and HindIII (3') were used to generate the *parA1* and *popZ* constructs, while EcoRI (5') and BamHI (3') were used for the *rcy* construct (Figure 3.3B). The nucleotide sequences of these knockout (KO) constructs, referred to hereafter as *parA1*-KO, *popZ*-KO and *rcy*-KO, can be found in Appendix A.1.

Throughout this thesis, the term 'unmarked' has been used to define any knockout constructs or mutants which do not possess a selectable marker (e.g. antibiotic resistance, reporter gene etc.) to easily distinguish these constructs from the tetracycline-resistant constructs generated later in this chapter (Section 3.3). However, it is important to note that the constructs described here are not truly 'unmarked' due to the introduction of the short (15 bp) non-native nucleotide sequence in place of the target gene.

The synthetic *popZ*-KO and *rcy*-KO constructs were liberated from their vector backbones by restriction digestion with the corresponding flanking enzymes and subcloned as XmaI-HindIII (*popZ*) or EcoRI-BamHI (*rcy*) fragments into pK18*mobsacB* to generate the KO vectors pK18-*popZ*-KO and pK18-*rcy*-KO respectively. The methodology followed for this subcloning step can be found in Sections 2.6 and 2.6.1. Whereas the subcloning of *popZ*-KO and *rcy*-KO into pK18*mobsacB* was readily achieved, subcloning of *parA1*-KO highlighted unresolvable issues with this construct (see Appendix A.2 for details). Therefore, an alternative method of construct generation was adopted to generate pK18-*parA1*-KO. This alternative approach is described

later in this section. The presence of the *lacZa* gene on pK18*mobsacB* enabled the pre-screening of successful pK18-*popZ*-KO and pK18-*rcy*-KO clones using blue-white screening of colonies following transformation of ligated plasmids into a compatible *E. coli* strain (e.g. DH5α). This screening process is described in Section 2.6.1. Plasmid DNA isolated from three independent white colonies (indicate of positive clones), was further validated by three independent restriction digestions designed to; 1) linearise the plasmid, 2) cleave the 5' homologous flank from the plasmid, and 3) cleave the full-length knockout construct from the pK18*mobsacB* backbone (Figure 3.3C).

Digest products were analysed by agarose gel electrophoresis to confirm whether correct pK18-*popZ*-KO and pK18-*rcy*-KO clones were obtained (Figure 3.3D). For pK18-*popZ*-KO, the digest of plasmid isolated from a single colony has been shown. Incomplete digestion of this plasmid was observed, indicated by the presence of undigested supercoiled plasmid at 5000 bp and a 7000 bp linear plasmid in the PacI-digested (linearised) and double-digest samples respectively. Despite evidence of incomplete digestion, bands corresponding to the 5' homologous flank (814 bp; XmaI-PacI digested sample) and entire knockout construct (1621 bp; XmaI-HindIII digested sample) were clearly identified, leading to the conclusion that this was a correct pK18-*popZ*-KO clone. For pK18-*rcy*-KO, the digestion of plasmid DNA from two independent clones has been shown in Figure 3.3D. The first plasmid (pK18-*rcy*-KO(1)) although linearised by PacI and possessing an insert of the same size as the knockout construct (1.6 kb), did not produce a band indicative of the 5' flank following digestion with EcoRI and PacI. Therefore, this plasmid was deemed incorrect. However, all expected fragments were observed following digestion of pK18-*rcy*-KO(2) and hence this plasmid was assumed to be a correct clone of pK18-*rcy*-KO.

The correct plasmid clones identified in Figure 3.3 were subsequently sequenced using the pK18*mobsacB*-specific primers pK18_MCS_Fwd and pK18_MCS_Rev, designed to amplify inserts within the MCS of pK18*mobsacB* (Table 2.7). Alignment of these sequencing reads to their corresponding template sequences confirmed that both pK18-*popZ*-KO and pK18-*rcy*-KO possessed the correct knockout construct insert, with no mismatches identified across the entire construct read. Plasmids maps of pK18-*popZ*-KO and pK18-*rcy*-KO can be seen in Figures 3.4 and 3.5 respectively.

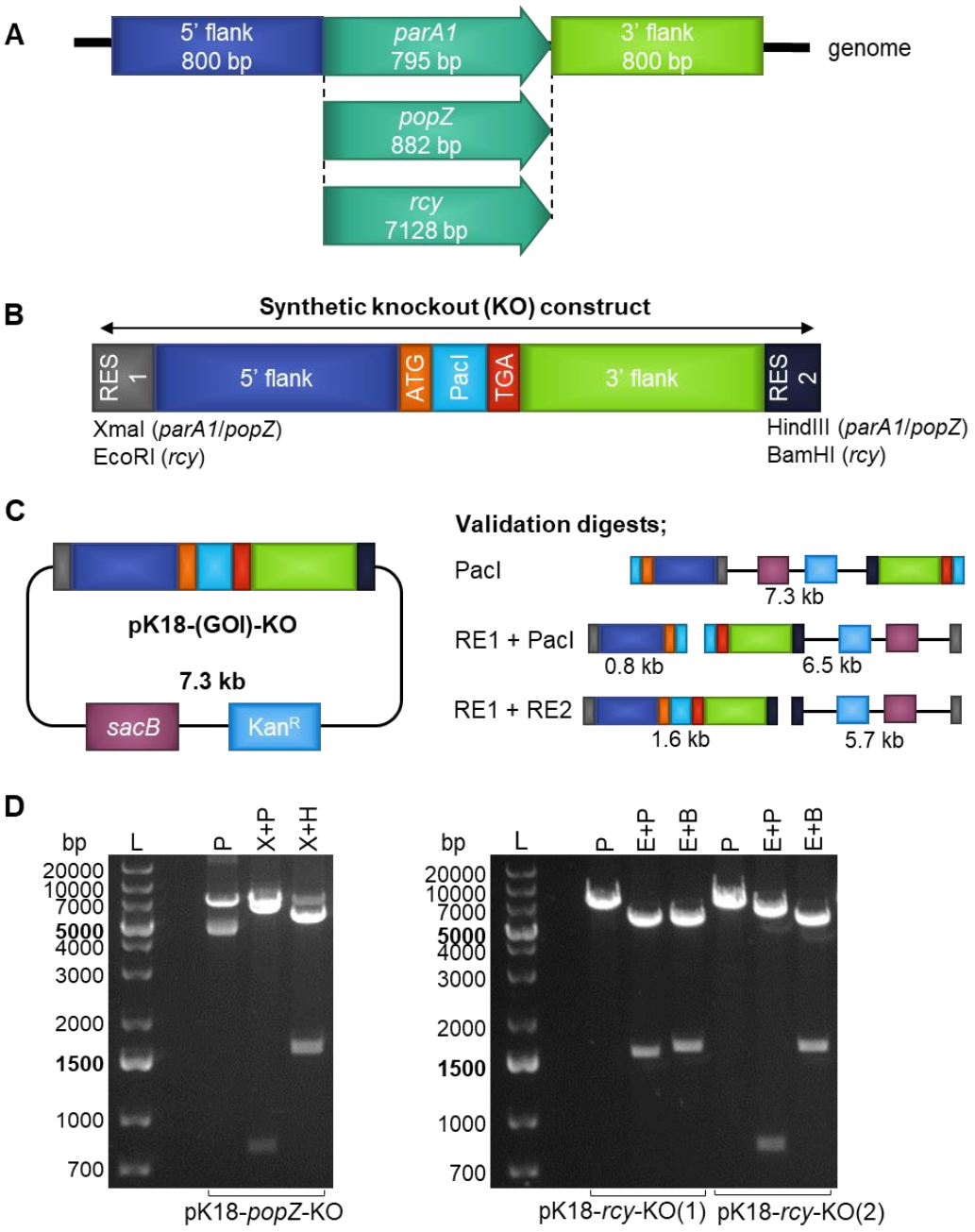


Figure 3.3. Unmarked knockout construct generation, and validation of the knockout vectors pK18-*popZ*-KO and pK18-*rcy*-KO. (A) 800 bp flanking regions immediately upstream and downstream of the genes of interest (GOI) were selected as the homologous flanking sequences to be used in the knockout (KO) construct. (B) KO constructs comprising of an 800 bp 5' flank (purple), start codon (orange), PacI restriction site (blue), stop codon (red) and a 800 bp downstream flanking region (green) were designed for each GOI. Flanking restriction endonuclease sites (RES) were added to allow for cloning into pK18*mobsacB*, and the constructs generated by gene synthesis. (C) Following subcloning of KO constructs into pK18*mobsacB*, the resulting KO vectors were validated by three restriction enzyme digestions. A schematic representation of the digest products has been shown here. (D) Agarose gel electrophoresis (1% w/v) of restriction digest products. Lane L: GeneRuler 1kb Plus DNA ladder. Letters above sample lanes indicate the enzymes used for digestion (P: PacI, X: XmaI, H: HindIII, E: EcoRI, B: BamHI). Exact expected digest product sizes (bp) were as follows: (pK18-*popZ*-KO) P: 7305; X+P: 6491, 814; X+H: 5684, 1621; (pK18-*rcy*-KO) P: 7319; E+P: 6505, 814; E+B: 5698, 1621.

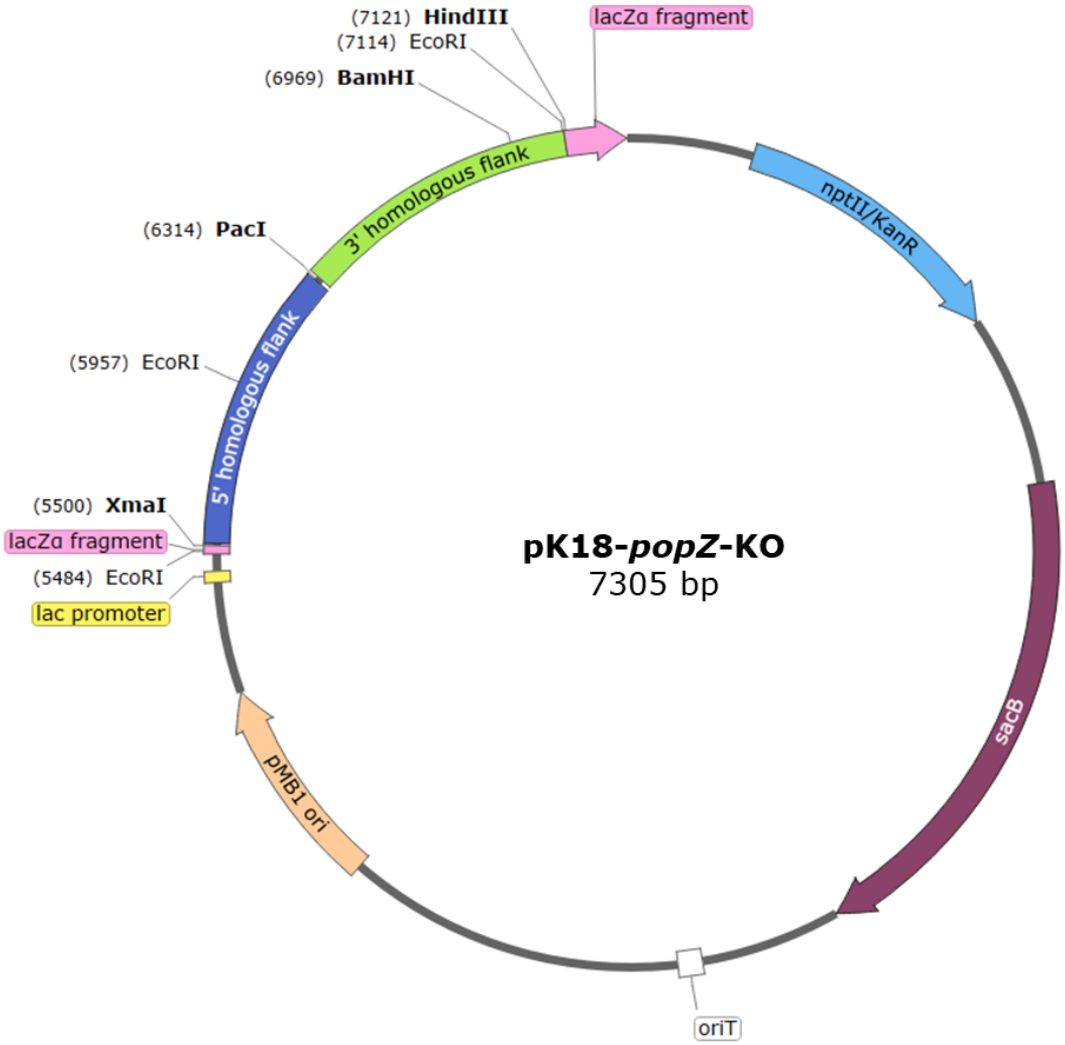


Figure 3.4. Plasmid map of the allelic exchange vector pK18-*popZ*-KO. This vector is a derivative of pK18*mobsacB* (*354*) which possesses the knockout construct required for the generation of unmarked *popZ* deletion mutants within *Rhizobium leguminosarum* 3841. This construct, comprising of a PacI recognition site located between a 5' (indigo) and 3' (green) homologous flank (identical in sequence to the 800 bp regions immediately upstream (5') and downstream (3') of the *R. leguminosarum* 3841 *popZ* gene), was cloned between the XmaI and HindIII sites within the MCS of pK18*mobsacB*, disrupting *lacZα* (pink). Other features of this vector are as follows; *nptII* (blue): confers kanamycin resistance (Kan^R), *sacB* (burgundy): confers sucrose sensitivity, pMB1 ori (orange): origin of replication, oriT (white): origin of transfer, *lac* promoter (yellow). Unique restriction sites are shown in bold.

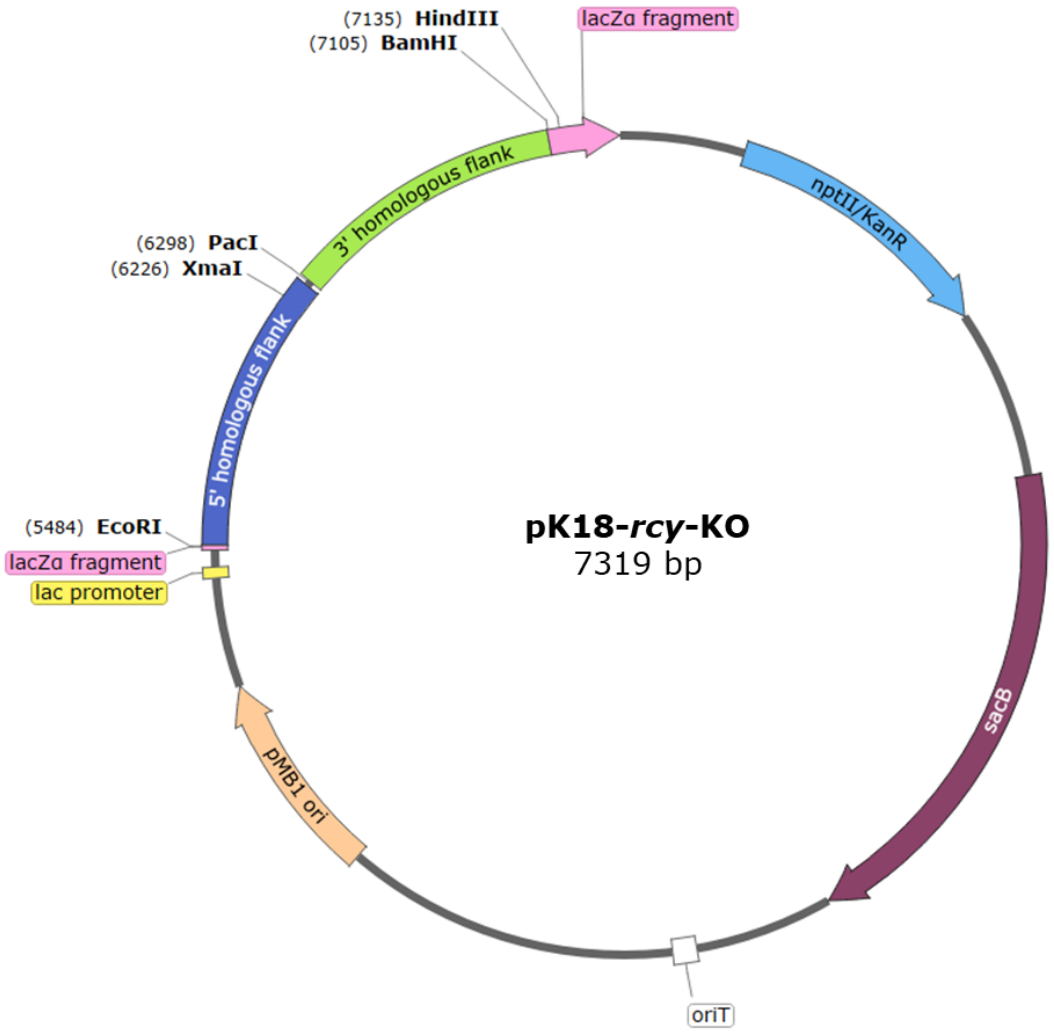


Figure 3.5. Plasmid map of the allelic exchange vector pK18-*rcy*-KO. This vector is a derivative of pK18*mobsacB* (*354*) which possesses the knockout construct required for the generation of an unmarked *rcy* knockout mutant within *Rhizobium leguminosarum* 3841. This construct contains a PacI recognition site located between a 5' (indigo) and 3' (green) homologous flank (identical in sequence to the 800 bp regions immediately upstream (5') and downstream (3') of the *R. leguminosarum* 3841 *rcy* gene). The construct was cloned between the EcoRI and BamHI sites within the MCS of pK18*mobsacB*, disrupting *lacZα* (pink). Other features of this vector are as follows; *nptII* (blue): confers kanamycin resistance (Kan^R), *sacB* (burgundy): confers sucrose sensitivity, pMB1 ori (orange): origin of replication, oriT (white): origin of transfer, *lac* promoter (yellow). Unique restriction sites are shown in bold.

As mentioned previously, initial attempts to subclone the synthetic *parA1*-KO construct into pK18*mobsacB* were unsuccessful. Details of these complications can be found in Appendix A.1. Despite exhaustive attempts to troubleshoot this cloning step, issues with this vector could not be resolved. Therefore, the *parA1* knockout construct was resynthesized and directly cloned commercially into pK18*mobsac*B to generate the knockout vector pK18-*parA1*-KO in a single step (Section 2.3.7). This newly-synthesised vector was validated by sequencing by GenScript following DNA synthesis and a confirmatory restriction digest of four independent clones following transformation of this vector into DH5α (Figure 3.6). This confirmatory digest demonstrated that all four clones were linearised by PacI, and an XmaI-PacI-HindIII triple digest successfully liberated both 800 bp homologous flanks from the pK18*mobsacB* backbone. Consequently, it was determined that all clones were correct copies of pK18-*parA1*-KO (see Figure 3.7 for plasmid map). In parallel to the troubleshooting and eventual resynthesis of *parA1*-KO, the generation of unmarked *popZ* and *rcy* gene knockout mutants was attempted using pK18-*popZ*-KO and pK18-*rcy*-KO (described later in Sections 3.2.2 and 3.2.3). Ultimately, the generation of unmarked *parA1* knockouts was not attempted, however the synthesis of pK18-*parA1*-KO was still an essential step in the generation of marked knockout mutants, described later within this chapter (Section 3.3).

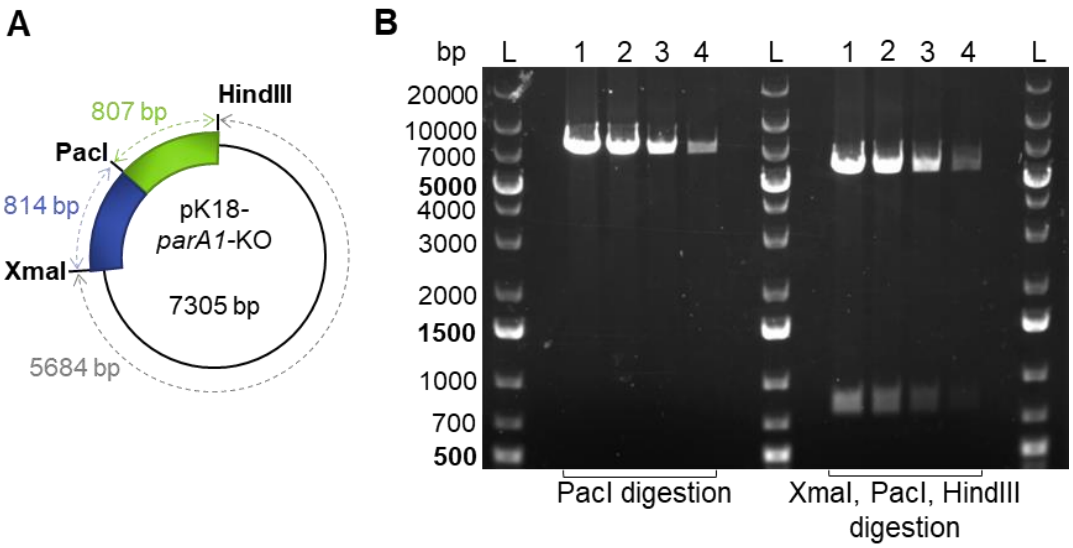


Figure 3.6. Restriction digest validation of resynthesized pK18-*parA1*-KO clones. (A) Schematical representation of pK18-*parA1*-KO, highlighting the XmaI, PacI and HindIII restriction sites used for validation, and the resulting DNA fragment lengths. (B) Agarose gel electrophoresis (1% w/v) of restriction digestion products. Lane L; GeneRuler 1kb Plus DNA ladder. Lanes 1-4; digestions of four independent clones of the pK18-ParA1-KO vector, obtained following transformation of the resynthesized pK18-*parA1*-KO vector into DH5α.

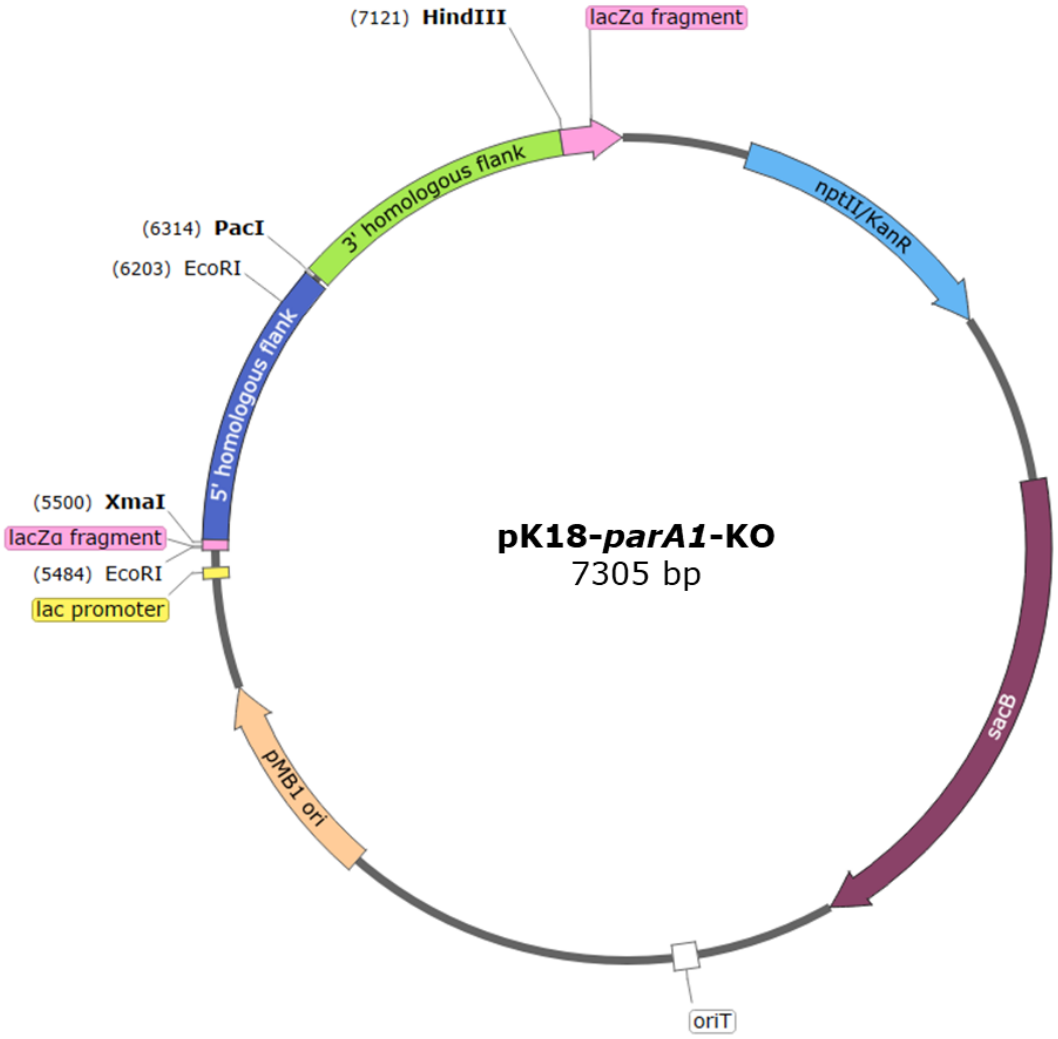


Figure 3.7. Plasmid map of the allelic exchange vector pK18-*parA1*-KO. This vector is a derivative of pK18*mobsacB* (*354*) possessing the knockout construct required for the generation of unmarked *parA1* knockout mutants within *Rhizobium leguminosarum* 3841. This construct contains a PacI recognition site located between 5' (indigo) and 3' (green) homologous flanks. These flanks are identical in sequence to the 800 bp regions immediately upstream (5') and downstream (3') of the *R. leguminosarum* 3841 *parA1* gene. The construct was commercially synthesised directly into pK18*mobsacB* via the XmaI and HindIII sites within the MCS, disrupting *lacZα* (pink). Other features of this vector are as follows; *nptII* (blue): confers kanamycin resistance (Kan^R), *sacB* (burgundy): confers sucrose sensitivity, pMB1 ori (orange): origin of replication, oriT (white): origin of transfer, *lac* promoter (yellow). Unique restriction sites are shown in bold.

3.2.2 Screening for primary integration events

To generate unmarked *popZ* and *rcy* knockout mutants, the gene-specific pK18*mobsacB*-derived knockout vectors described above (Section 3.2.1) were introduced independently into *R. leguminosarum* 3841 by bacterial conjugation (Section 2.4.2). This process involves a tri-parental mating between the recipient strain (*R. leguminosarum* 3841), an *E. coli* (DH5α) 'donor' strain possessing the knockout vector, and the 'helper' strain *E. coli* pRK2013 (Section 2.1.1). The helper plasmid pRK2013 (*348*) facilitates the mobilisation of non-self-transmissible plasmids (i.e. pK18*mobsacB*) into the recipient strain. Transfer (*tra*) genes on this plasmid encode a variety of conjugation-related proteins including those required for pili assembly and conjugal transfer of genetic material, while mobility (*mob*) genes encode proteins which nick DNA at the origin of transfer (oriT), and initiate pili-mediated transfer of single-stranded DNA into the recipient cell via pili (*348*, *356*, *371*). Following mating, transconjugants were plated onto TY Agar containing both streptomycin (str) and kanamycin (kan) to select for *R. leguminosarum* cells that had undergone primary integration events of the knockout vector into the genome (referred to as primary integrants) (Section 2.8). *R. leguminosarum* 3841 is intrinsically resistant to high concentrations of streptomycin (up to 500 ug mL^{-1}), allowing for highly effective str-based selection against the *E. coli* donor and helper strains which are not resistant to this antibiotic. Given that pK18*mobsacB* cannot replicate within *Rhizobium*, any colonies which grow following selection with kanamycin were expected to be primary integrants.

Four kanamycin-resistant primary integrants per knockout vector (pK18-*popZ*-KO and pK18-*rcy*-KO) were inoculated into 10 mL TY and grown for 2 days under non-selective conditions to allow for recombination between the vector and the genome. Following this non-selective growth, 5 mL of each culture was used to prepare purified genomic DNA (gDNA) (Section 2.3.1), with the remaining culture used for the next step of mutant generation (see Sections 3.2.3 and 2.8). Isolated gDNA was subjected to a PCR-based screen to confirm integration of the vector into the genome. The premise of this screen was to amplify regions of gDNA from each culture with two independent pairs of PCR primers. Each primer was specifically designed to screen for integrations within either the 5' or 3' homologous flanking regions. A schematical representation for the use of this PCR screen to confirm the integration of pK18-*popZ*-KO has been shown in Figure 3.8A. Here, the primers *popZ*_KO_screen_Fwd and *popZ*_3'-flank_Rev were used to screen for integrations within the 5' gene-flanking regions, with primers *popZ*_5'-flank_Fwd and *popZ*_KO_screen_Rev used to screen for 3' flank integrations. Sequences of these primers can be found in Table 2.7.

Within each of these pairs, a single primer anneals to the genomic DNA upstream/downstream of the 5'/3' flank (respectively), with the other primer annealing within the opposite flanking region. This primer configuration allows for the amplification of a DNA fragment comprising of the entire 800 bp target flank, a short (approx. 200 bp) region of the opposite flank, and any nucleotides located between the two flanks. As demonstrated in Figure 3.8A, integration of pK18-*popZ*-KO within the corresponding homologous flank would yield a PCR product of approximately 1 kb, referred to here as an 'integration' PCR product. In instances where integration of the vector did not occur within the targeted flank, a larger 1.8 kb 'wildtype' PCR product would be obtained following amplification of both the targeted flanking region(s) and the *popZ* gene (882 bp).

In addition to the four primary integrant gDNA samples, gDNA isolated from wildtype (WT) *R. leguminosarum* 3841 was also subjected to this screen. In the case of WT gDNA, amplification with either primer pair should yield a PCR product of approx. 1.8 kb. Following amplification, the products generated by this PCR screen were then analysed by agarose gel electrophoresis (Figure 3.8B). This screen clearly identified integration of pK18-*popZ*-KO within the 3' flanking region in all four cultures screened, demonstrated by the presence of a 1 kb fragment in all four integrant gDNA samples amplified with the 3' integration screening primers (*popZ*_5'-flank_Fwd and *popZ*_KO_screen_Rev). The absence of vector integration within the 5' flanking region in these samples, inducated by the presence of a 1.8 kb 'WT' PCR product following amplification with the 5' integration screening primers (*popZ*_KO_screen_Fwd and *popZ*_3'-flank_Rev), was consistent with the presence of a band of identical size within the WT gDNA sample.

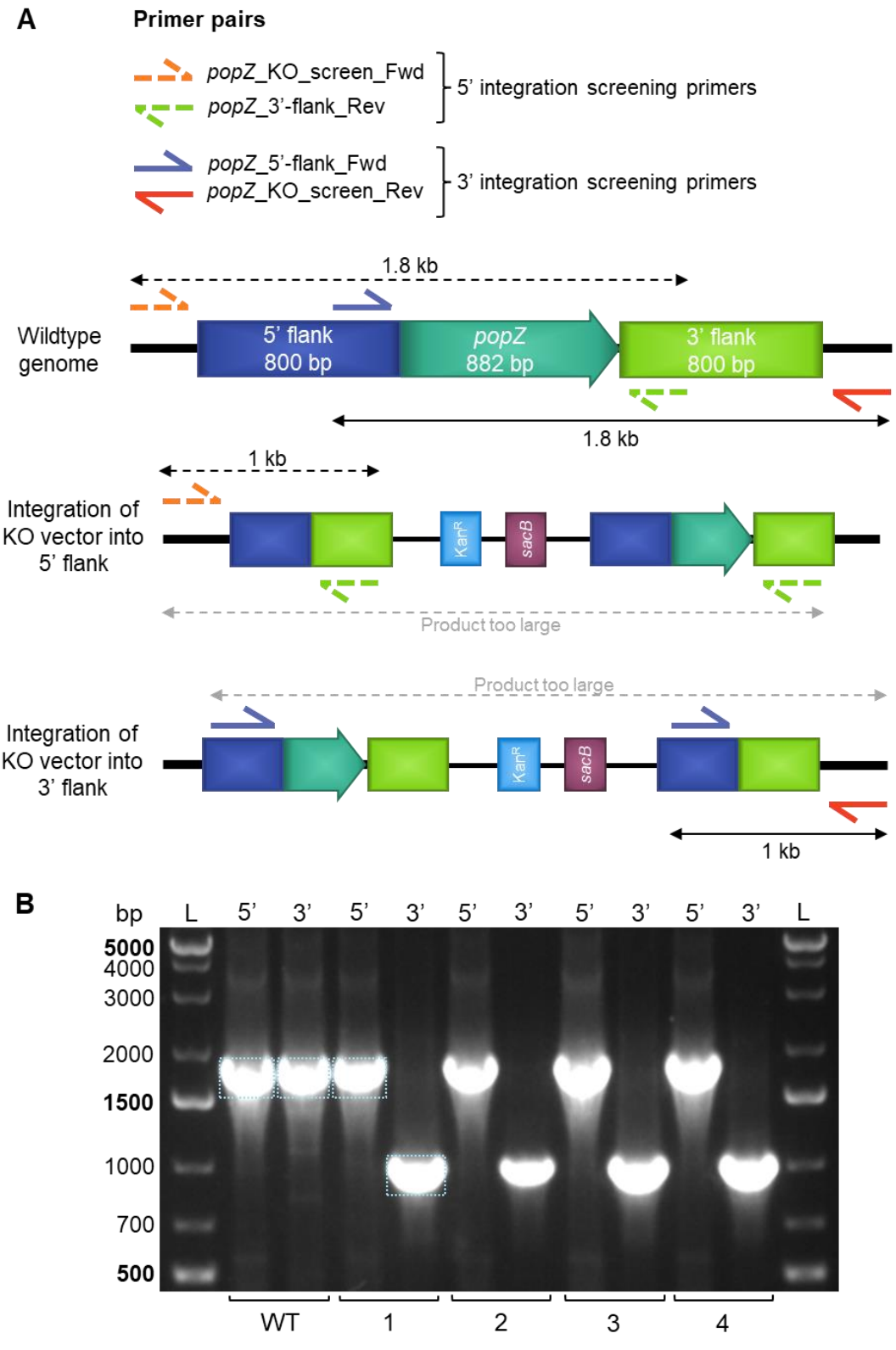


Figure 3.8. PCR-based confirmation of the integration of pK18-*popZ*-KO into the genome of *R. leguminosarum*. (A) Integration of the knockout (KO) vector pK18-*popZ*-KO within each flank was screened for using primer pairs (half arrows) comprising of one genome-specific primer (orange dashed/red solid), and one flank-specific primer (indigo solid/green dashed). Although flank-specific primers have two binding sites, binding to the distant site would result in amplification of the entire vector, including the *sacB* gene (burgundy) and kan resistance

cassette (blue), generating a product too large to be obtained in this PCR reaction. Amplification of the wildtype (WT) *popZ* region using either primer pair would generate a 1.8 kb PCR product. If no integration occurred, this same product would be generated. However, if integration of pK18-*popZ*-KO occurred within the targeted flank, a 1 kb PCR product would be obtained. (B) Genomic DNA (gDNA) isolated from four primary integrants (following a 2-day non-selective culture) and WT *R. leguminosarum* 3841 was amplified using the primers depicted in Part A, and the PCR products analysed by agarose gel electrophoresis (1% w/v). Lane L; GeneRuler 1kb Plus DNA ladder. 5' and 3' lane annotations depict the screening primers used. Square brackets (bottom) indicate the gDNA sample amplified, with 1-4 depicting the four primary integrants. DNA fragments in dashed boxes were purified and sent for Sanger sequencing.

Evidence of integration of pK18-*rcy*-KO was screened for using the following primer pairs: *rcy*_KO_screen_Fwd and *rcy*_3'-flank_Rev, to screen for integrations within the 5' flank; and *rcy*_5'-flank_Fwd and *rcy*_KO_screen_Rev, to screen for 3' flank integrations. Details of these primers can be found in Table 2.7. These primers anneal to the same respective locations within *rcy* gene region as their *popZ* equivalents, generating a screening process identical in principle to the *popZ* screen depicted in Figure 3.8. The integration of pK18-*rcy*-KO within the targeted flank (i.e. flank corresponding to the primer pair used) would yield a 1 kb PCR product, as observed with *popZ* (Figure 3.8A). Within wildtype *R. leguminosarum* 3841, or in instances where no vector integration occurred within the targeted flank, a PCR product of approximately 8 kb would be obtained following amplification of the full-length *rcy* gene (7128 bp) located between the flanking regions amplified. However, in contrast to the 1.8 kb *popZ* WT product, amplification of this 8 kb *rcy* WT product would not be possible within the cycle parameters used in this PCR screen. Therefore, an absence of a PCR product was considered indicative of a lack of integration/WT genotype within the screened flanking region.

Agarose gel electrophoresis analysis of the PCR products obtained in this screen (Figure 3.9) identified evidence of integration of pK18-*rcy*-KO within both the 5' and 3' gene-flanking regions in all four primary integrant cultures, indicated by the presence of a 1 kb PCR product in all culture gDNA samples amplified (Figure 3.9A). Non-specific amplification with the 5' flank screening primers (*rcy*_KO_screen_Fwd and *rcy*_3'-flank_Rev) was observed within this first PCR screen (Figure 3.9A), yielding a product of approximately 1.4 kb in all gDNA samples amplified.

However, this non-specific amplification was resolved following optimisation of the PCR cycle parameters for this primer pair (Figure 3.9B). This second PCR screen generated clear, single 1 kb DNA fragments indicative of plasmid integration within the 5' flanking regions of all primary integrant cultures with clear absence of product in the WT control sample (as expected).

Evidence of the integration of pK18-*rcy*-KO within both the 5' and 3' gene-flanking regions of the gDNA isolated from these primary integrant cultures was suggestive of one of two possible scenarios. Firstly, and preferably, further recombination of pK18-*rcy*-KO may have occurred. Although the initial primary integrant colonies picked should have possessed either a 5' or a 3' integration event (i.e. a homogenous colony), further recombination between the vector and the genome during non-selective growth would give rise to a heterogenous culture, comprising of primary integrants, wildtype revertants, and even the possibility of successful knockout mutants. Genomic DNA isolated from these cultures would therefore contain a mixture of these genotypes. If knockout gDNA was present, in which recombination between the knockout vector and genome has occurred within both flanks, PCR amplification with either of the primer pairs described above would yield a 1 kb PCR product. Alternatively, evidence of vector integration within both flanking regions of a given gDNA sample could have arisen if the initial primary integrant colonies picked for non-selective growth were not derived from a single colony forming unit (i.e. an isogenic colony), but were in fact a mixed colony derived from a mixture of cells with 5' and 3' primary integrations. Screening of gDNA isolated from a culture derived from such a population would therefore produce evidence of integration events of pK18-*rcy*-KO within both the 5' and 3' flanking regions, even if no further recombination occurred between the knockout vector and the genome.

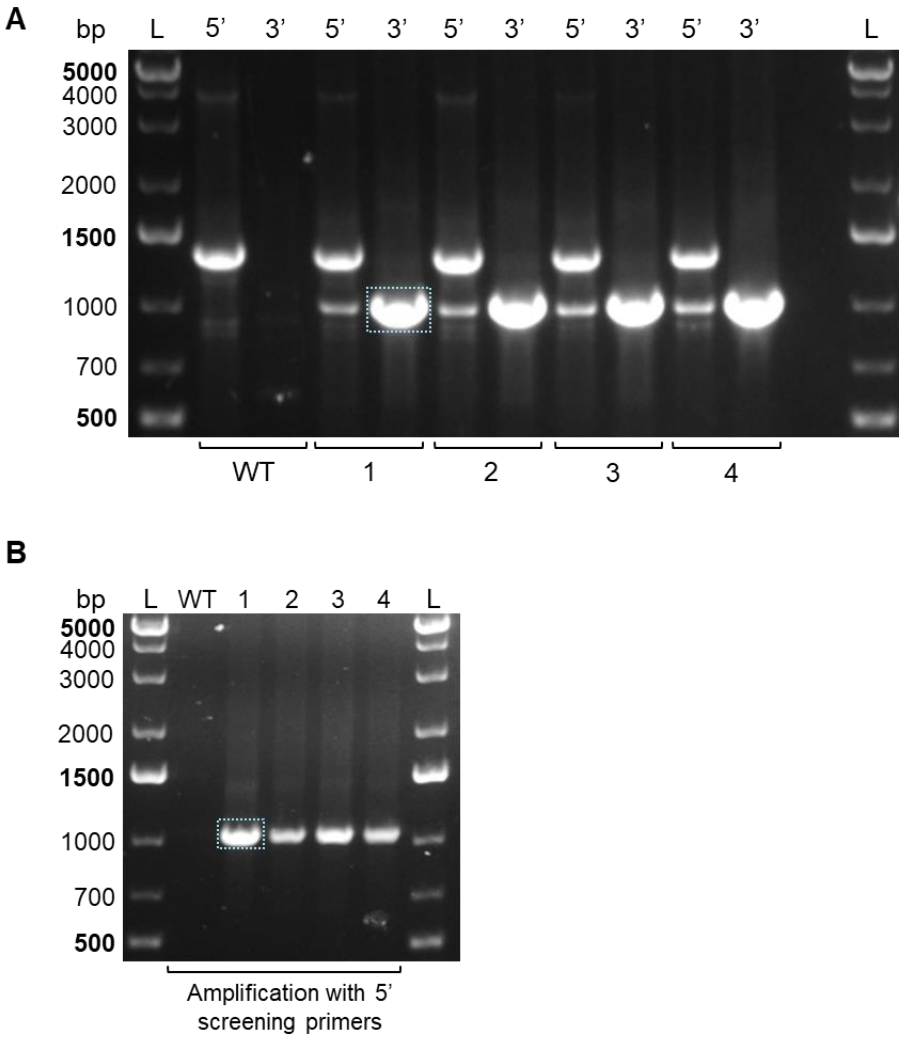


Figure 3.9. PCR-based confirmation of integration of pK18-*rcy*-KO into the genome of *R. leguminosarum*. (A) Genomic DNA (gDNA) isolated from four primary integrants (following a 2-day non-selective culture) and wildtype (WT) *R. leguminosarum* 3841 was amplified using the following primers: *rcy*_KO_screen_Fwd / *rcy*_3'-flank_Rev, to screen for integrations within the 5' flank; and *rcy*_5'-flank_Fwd / *rcy*_KO_screen_Rev, to screen for 3' flank integrations. In WT samples, and those where no integration occurred in the targeted flank, no PCR product should be obtained. If integration of the vector occurred within the targeted flank, a 1 kb PCR product should be generated. PCR reactions were then analysed by agarose gel electrophoresis (1% w/v). Lane L; GeneRuler 1kb Plus DNA ladder. 5' and 3' lane annotations depict the screening primers used. Square brackets (bottom) indicate the gDNA sample amplified, with 1-4 depicting the four primary integrant cultures. (B) All gDNA samples were amplified with the 5' integration screening primers (*rcy*_KO_screen_Fwd / *rcy*_3'-flank_Rev), using slightly modified PCR cycle parameters optimised for this primer pair. Lane L; GeneRuler 1kb Plus DNA ladder. WT; WT gDNA template. Lanes 1-4; primary integrant culture gDNA. DNA fragments in dashed boxes were purified and sent for Sanger sequencing.

A DNA fragment representative of each of the unique PCR products obtained in these PCR screens (Figures 3.7B, 3.8A & 3.8B, dashed boxes) was excised from the agarose gel and purified (Section 2.3.3). These six fragments - three 1.8 kb bands, two obtained from amplification of WT gDNA with both *popZ*-specific primer pairs and a third from the amplification of pK18-*popZ*-KO integrant gDNA with the respective 5' flank screening primers; one 1 kb fragment from the amplification of a pK18-*popZ*-KO integrant with the 3' flank screening primers; and two 1 kb fragments from the screening of a pK18-*rcy*-KO integrant with both *rcy*-specific primer pairs – were then sent for Sanger sequencing with their respective PCR primers, generating a forward and reverse sequencing read for each (Section 2.3.6). For all 1.8 kb fragments, indicative of an absence of vector integration (i.e. WT genotype), the full-length *popZ* gene sequence was clearly identified between the two flanking regions, confirming that no disruption of genome occurred within these regions. Furthermore, sequencing reads of all three 1 kb fragments, indicative of vector integration, clearly identified disruption to these regions had occurred. The PacI recognition sequence located within the short nucleotide sequence between the 5' and 3' homologous flanks of the knockout constructs (Figure 3.3) was present in all three fragments, confirming that the knockout vectors had successfully integrated into their target genomic regions within the conjugants screened.

3.2.3 Screening for knockout mutants

Following the conjugation of the knockout vectors pK18-*popZ*-KO and pK18-*rcy*-KO into *R. leguminosarum* 3841, cells which had undergone primary integration events of the vector into the genome (described above in Section 3.2.2) were subsequently cultured for two days under non-selective conditions to allow for further recombination between the homologous regions of the integrated knockout vector and the genome. These subsequent recombination events, also referred to as double-crossover events, result in resolution of the vector backbone generating either wildtype (WT) revertants, or successful knockout mutants (Figure 3.1). To isolate cells which had undergone a double-crossover event, sucrose counter-selection was used. As described at the beginning of this Chapter (Section 3.2), expression of the *sacB* gene on pK18*mobsacB* following integration of this vector into the host genome results in lethal sucrose sensitivity. Therefore, the presence of sucrose inhibits the growth of any cells in which the vector has been retained (i.e. primary integrations). By contrast, cells in which the vector has been excised from the genome following a double-crossover event will no longer be sensitive to sucrose.

Serial dilutions of the non-selective primary integrant cultures were plated onto TY-sucrose (10%) Agar plates containing streptomycin (str) (Section 2.8) and incubated for 2-3 days until single colonies appeared. Typically, the colonies which arose on these sucrose plates varied in diameter, but exhibited uniform morphology (circular, convex colonies). In comparison to colonies on TY Agar plates, colonies on TY-sucrose Agar plates were more fluid in texture. To avoid bias, colonies of all sizes were picked onto TY Agar (+ str) master plates, taking note of the size of each initial colony picked. Master plates were then incubated for 1-2 days until visible colonies were obtained. Colony growth on the master plates did not appear to correlate with the original colony size, suggesting that small colonies picked from the sucrose plates were not necessarily slow-growing colonies. One drawback of *sacB*-mediated sucrose counter-selection is the sensitivity of *sacB* to suppressor mutations, giving rise to false positives i.e. sucrose-resistant primary integration colonies (*372*). Therefore, all colonies from these master plates were replica plated onto TY Agar plates containing kanamycin (kan), where resistance to kan indicated that the knockout vector was still integrated into the genome. Following replica plating, a few kan-resistant colonies were identified but the vast majority were kan-sensitive as expected.

To distinguish between wildtype revertants and knockout mutants, kan-sensitive colonies from replica plate screening were subjected to a colony PCR screen (Section 2.5.3). In this screen, illustrated in Figure 3.10A, colony genomic DNA was amplified using the following primer pairs: *popZ*_KO_screen_Fwd / *popZ*_KO_screen_Rev, and *rcy*_KO_screen_Rev / *rcy*_KO_Screen_Fwd. These primer pairs, each consisting of a forward primer which anneals upstream of the 5' flank and a reverse primer annealing downstream of the 3' flank, amplify regions of the *R. leguminosarum* genome encompassing the target gene and both 800 bp flanking regions. In the event of a *popZ* or *rcy* knockout mutant, amplification of the colony template with the respective primer pair would result in the generation of a 1.8 kb PCR fragment. In the event of a wildtype revertant, amplification with of the *popZ* gene-region (primers *popZ*_KO_screen_Fwd and *popZ*_KO_screen_Rev) would yield a fragment of 2.5 kb, while amplification of the *rcy* gene region (primers *rcy*_KO_screen_Rev and *rcy*_KO_Screen_Fwd) would generate a fragment of 8.9 kb. Similar to the primary integration PCR screen used in Section 3.2.2 (Figure 3.8), this 8.9 kb WT *rcy* fragment would not be successfully amplified within the cycle parameters used in this screen (extension time too short). Therefore, an absence of PCR product in colony samples amplified with the *rcy*-specific primers was indicative of a WT genotype.

Figure 3.10 shows DNA electrophoresis analysis of products obtained from colony PCR of six double-crossover colonies derived from pK18-*popZ*-KO (Part B) and pK18-*rcy*-KO (Part C) integration events. A significant drawback of this PCR screen is the lack of positive control for PCR reactions using primers *rcy*_KO_screen_Rev and *rcy*_KO_Screen_Fwd. Although primer specificity and functionality were not a concern (successful amplification with these primers previously demonstrated in Section 3.2.2), a lack of positive control would make difficult to accurately determine whether an absence of product is a true knockout mutant, or the result failed PCR reactions. Therefore, a small control PCR was set up in which three pK18-*rcy*-KO-derived colonies were amplified using the *popZ* screening primers (Figure 3.10D). Here, a 2.5 kb WT *popZ* product should be obtained, as no disruption to this genomic region should have occurred within these samples. Of the 12 double-crossover colonies analysed in Figure 3.10 (6 per gene), only WT revertants were identified. For all six pK18-*popZ*-KO-derived colonies, a 2.5 kb WT *popZ* DNA fragment was obtained (Figure 3.10B). This same product was also obtained following amplification of gDNA isolated from wildtype *R. leguminosarum* 3841, confirming that these colonies were in fact WT revertant colonies. In the case of pK18-*rcy*-KO-derived colonies, amplification with *rcy* screening primers primers *rcy*_KO_screen_Rev and *rcy*_KO_Screen_Fwd did not yield a PCR product (Figure 3.10C). Successful amplification of the 2.5 kb control PCR

product (Figure 3.10D) confirmed that the absence of PCR product in the *rcy* screening reactions was the result of these colonies also being WT revertants.

Figures 3.10B and 3.10C show only a representative sample of the colonies screened. In total, approximately 200 colonies per conjugation were screened using this colony PCR screen. Of the ~400 colonies screened, all pK18-*rcy*-KO double-crossover events yielded clear wildtype revertants (indicated by absence of PCR product), while all but three pK18-*popZ*-KO double-crossover colonies produced clear 2.5 kb bands demonstrating these also were WT revertants. The three (of ~200) colonies which did not yield a clear 2.5 kb band instead exhibited an unexpected triple banding pattern, with all three bands between 1.5 - 3 kb in size (data not shown). This triple banding was consistent across all three colonies, and more pronounced following amplification of gDNA isolated from these colonies. Although no biological/genetic explanation of the middle-sized band (approx. 2 kb) could be determined (other than non-specific primer binding), the presence of a larger (approx. 2.5-2.8 kb) and smaller (between 1.5-2 kb) PCR product may have been indicative of heterogenous populations, comprising of both WT revertants and knockout mutants. However, multiple attempts to separate these (potentially) mixed populations into single homogenous colonies and isolate knockout mutants was unsuccessful. Therefore, it was ultimately determined that the generation of unmarked *popZ* and *rcy* knockout mutants using the knockout vectors pK18-*popZ*-KO and pK18-*rcy*-KO was unsuccessful.

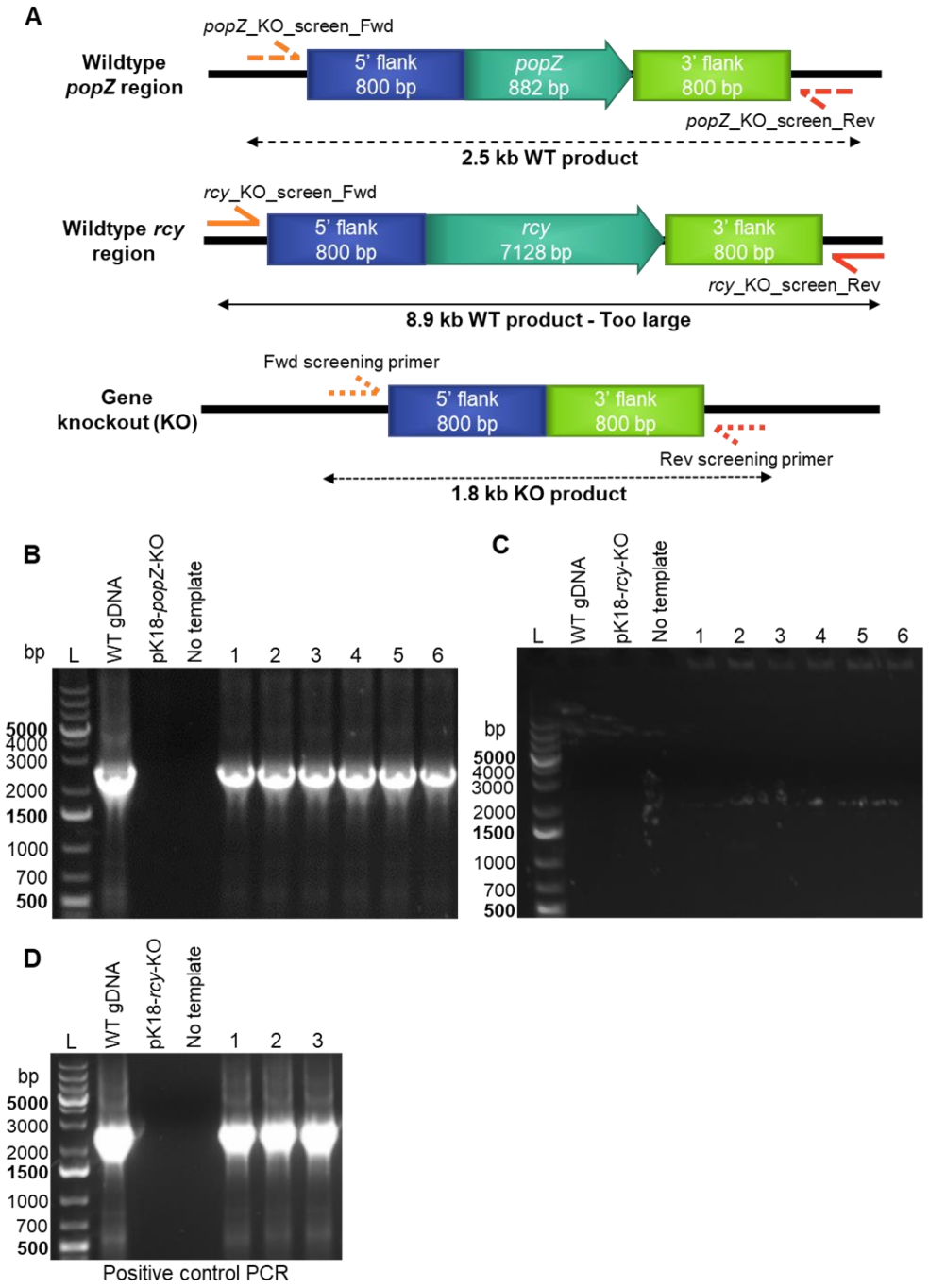


Figure 3.10. Colony PCR screening of potential *popZ* and *rcy* knockout mutants. (A) Double-crossover colony template was amplified in colony PCR reactions using genome-specific primer pairs, comprising of a forward primer (orange half arrows) annealing upstream of the 5' flank (indigo), and a reverse primer (red half arrows) annealing downstream of the 3' flank (green). Primer names are noted next to each primer. Amplification of the wildtype (WT) *popZ* region (i.e. WT revertant) with corresponding primers would yield a PCR product of 2.5 kb. Amplification of WT *rcy* colonies would generate a theoretical product of 8.9 kb. However, this product is too large to be successfully amplified within the extension time used. Hence, an absence of PCR product is indicative of WT *rcy* colonies. (B-D) Agarose gel (1% w/v) electrophoresis analysis of the colony PCR products. Lane L; GeneRuler 1 kb Plus DNA Ladder. For amplification with each primer pair, three control reactions were performed using the following templates: WT gDNA (positive for popZ, negative for rcy), pK18-*popZ*-KO/pK18-*rcy*-

KO (negative) and dH_2O (i.e. no template) (negative). (B) Six pK18-popZ-KO double-crossover colonies were screened with the primers *popZ*_KO_screen_Fwd and *popZ*_KO_screen_Rev. (C) Six pK18-*rcy*-KO double-crossover colonies were screened with the primers *rcy*_KO_screen_Fwd and *rcy*_KO_screen_Rev. (D) Positive control PCR of three pK18-*rcy*-KO double-crossover colonies amplified with the *popZ* screening primers.

3.3 Generation of marked knockout mutants

Given the difficulty of generating and isolating unmarked *popZ* and *rcy* mutants, a different strategy of mutant generation was undertaken to enable generation of marked knockout mutants. In these mutants, the entire gene length is replaced with a tetracycline resistance (tet^R) cassette, intended to facilitate the isolation and identification of a successful gene knockout. To generate these mutants, the same pK18*mobsacB*-mediated homologous recombination-based system (as described in Section 3.2) is used, but this time using a modified tet^R knockout vector. This section describes the generation of these tet^R knockout vectors, and their subsequent use in the attempted generation of marked *parA1*, *popZ* and *rcy* knockout mutants.

3.3.1 Cloning of tetracycline-resistant marked knockout vectors

To generate the knockout constructs required for the generation of tet^R marked mutants, the tet^R gene (*tetA*) of pSUP3535, a pSUP202 derivative (*356*, *357*), was subcloned into the previously-generated unmarked knockout vectors (pK18-*parA1*-KO, pK18-*popZ*-KO and pK18-*rcy*-KO, Section 3.2.1). This vector was chosen as the source of the *tetA* gene as pSUP3535/pSUP202 has been demonstrated to confer effective tetracycline resistance in members of the Hyphomicrobiales, such as various *Bradyrhizobium* and *Rhizobium* species (including *R. leguminosarum* 3841) (*224*, *343*, *357*, *373*, *374*). The tet^R cassette was amplified from pSUP3535 using the primers pSUP_tet_PacI_Fwd and pSUP_tet_PacI_Rev (Table 2.7), designed to introduce flanking PacI recognition sites to each end of the tet^R cassette fragment. The purified PCR product was then subcloned as a PacI-PacI fragment directly into the PacI site of the unmarked knockout vectors, introducing the tet^R cassette between the 5' and 3' homologous flanks of the knockout construct (see Figures 3.4, 3.5 and 3.7 for plasmid maps of

pK18-*popZ*-KO, pK18-*rcy*-KO, and pK18-*parA1*-KO respectively). Successful clones, referred to from here on as pK18-*parA1*-tet^R-KO, pK18-*popZ*-tet^R-KO and pK18-*rcy*-tet^R-KO, were then identified based on their resistance to tetracycline. An overview of the cloning of these vectors can be found in Section 2.6.2.

Due to the non-directional nature of this cloning, two possible orientations of the tet^R cassette within the knockout vectors could be obtained; a forward and reverse compliment (Figure 3.11A). Although either orientation would confer tetracycline resistance within a successful knockout mutant, the reverse orientation of this cassette within the knockout construct was desired for the generation of knockout mutants. By introducing the tet^R cassette into the genome in the opposite direction to neighbouring genes, polar effects on downstream gene expression as a result of the introduction of the tet^R promoter are avoided. Therefore, to confirm the orientation of the tet^R cassettes within the knockout constructs, another PCR screen was employed. In Figure 3.11B, a schematic of the use of this PCR screen on pK18-popZ-tet^R-KO clones has been shown. In this screen, the pK18-*popZ*-tet^R-KO vectors were each amplified with two pairs of primers. The same 5' homologous flank-specific forward primer (*popZ*_5'-flank_Fwd) was used within each primer pair. To screen for clones possessing a tet^R cassette in the forward orientation, the second primer within this pair was a reverse primer annealing to the 3' end of the tet^R cassette (pSUP_tet_PacI_Rev). To screen for the reverse-orientation tet^R cassette, the second primer within this pair was a 'reverse' primer which anneals to the 5' end of the tet^R cassette (pSUP_tet_PacI_Fwd). Although this primer is technically a 'forward' primer when described in the context of a forward-oriented tet^R cassette, it is important to note that if the tet^R cassette is in the reverse orientation, the tet^R-specific primers take on the opposite function i.e. a forward primer functions as a reverse primer. Following amplification, if the screening primer pair and orientation of tet^R cassette are complimentary, a 1.6 kb PCR product would be generated. If the primers and tet^R cassette orientation are not compatible, no PCR product would be obtained. Therefore, each vector should produce a 1.6 kb band with one of each primer pair, but not both.

Figure 3.11C shows the results of this PCR screen performed on seven pK18-*popZ*-tet^R-KO clones. In this screen, amplification of clones 1, 3, 4, 5 and 7 with the reverse-orientation screening primers produced a clear 1.6 kb band, with clones 2 and 6 generating this band following amplification with the forward-orientation screening primers. An identical screen was

performed on several pK18-*parA1*-tet^R-KO and pK18-*rcy*-tet^R-KO clones. The same tet^R cassette-specific primers (pSUP_tet_PacI_Fwd and pSUP_tet_PacI_Rev) were used for each of these screens, along with the *parA1* 5' flank-specific primer *parA1*_5'-flank_Fwd for pK18-*parA1*-tet^R-KO screening, and the *rcy* 5' flank-specific primer *rcy*_5'-flank_Fwd for pK18-*rcy*-tet^R-KO screening (data not shown). Following screening, two clones for each pK18-tet^R-KO vector (one for each tet^R orientation) were sent for Sanger sequencing to confirm their identity. Each vector was sequenced with four sequencing primers (pK18_MCS_Fwd, pSUP_tet_PacI_Fwd, pSUP_tet_PacI_Rev and pK18_MCS_Rev) to ensure sequencing reads spanning the entire knockout construct were obtained. Analysis of sequencing reads confirmed no mutations within the knockout constructs within all six vectors sequenced.

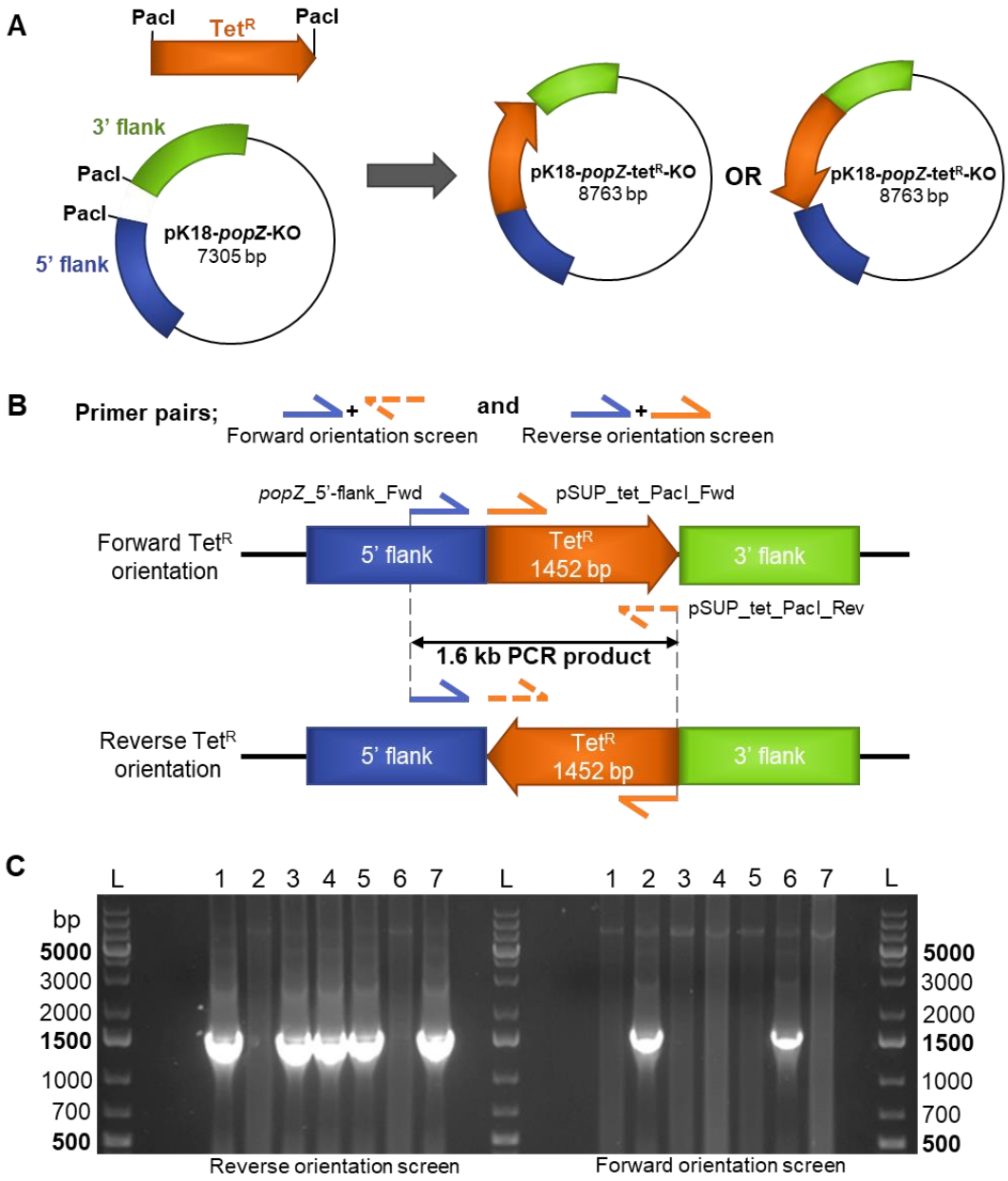


Figure 3.11. Directional PCR screen of the tetracycline resistance cassette within the knockout vector pK18-*popZ*-tet^R-KO. (A) Following non-directional cloning of the tetracycline resistance cassette (Tet^R) (orange) into the PacI site of the pK18-*popZ*-KO vector two possible orientations of the cassette within the resulting vector (pK18-popZ-tet^R-KO) may be obtained. (B) To determine the orientation of tet^R cassettes, pK18-*popZ*-tet^R-KO vectors were screened with two primer pairs (primers indicated with half arrows), where each primer pair screened for a single orientation. If the orientation of the cassette and primer pair were complementary (i.e. forward-oriented cassette screened with forward orientation screening primers), a 1.6 kb PCR product would be obtained. If cassette orientation and primer pair not complementary, no PCR product would be obtained. (C) Agarose gel electrophoresis (1% w/v) of the PCR products following the directional screening of seven pK18-*popZ*-tet^R-KO vectors. Lane L; GeneRuler 1 kb Plus DNA Ladder. Lane annotations 1-7 denote each of the seven pK18-*popZ*-tet^R-KO vectors screened. Amplifications with the reverse-orientation screening primers *popZ*_5'-flank_Fwd and pSUP_tet_PacI_Fwd are shown on the left-hand of this gel. Amplifications with the forward-orientation screening primers *popZ*_5'-flank_Fwd and pSUP_tet_PacI_Rev are shown on the right-hand side.

3.3.2 Screening for tetracycline-resistant knockout mutants

As described during the attempted generation of unmarked mutants, the marked knockout vectors pK18-*parA1*-tet^R-KO, pK18-*popZ*-tet^R-KO and pK18-*rcy*-tet^R-KO (collectively referred to as the pK18-tet^R-KO vectors) were introduced into *R. leguminosarum* 3841 via tri-parental conjugation. *Rhizobium* transconjugants possessing a primary integration event of the knockout vector into the genome were selected by plating the conjugation mixture onto TY Agar containing streptomycin (str), tetracycline (tet) and kanamycin (kan) (Section 2.8). Primary integrants were subsequently grown under non-selective conditions for 2 days to allow for further recombination between the knockout vectors and the genome. Following this non-selective growth, serial dilutions of the pK18-tet^R-KO primary integrant cultures were then plated onto TY Agar medium containing 10% (w/v) sucrose, str and tet. Excision of the pK18*mobsacB* vector backbone from the genome (i.e. double crossover event) would result in either allelic exchange of the target gene with the tet^R gene, or a reversion to the WT genotype. Therefore, only double-crossover events which yielded successful knockout mutants should grow on these plates as sucrose counter-selection selects against cells expressing *sacB* (i.e. primary integrants), and the presence of tet selects against any WT revertants.

Following incubation of the TY-sucrose plates (2-3 days), a surprisingly high number of single colonies were observed. Many of these serial dilution plates had similar, and in some cases greater, numbers of colonies than their unmarked knockout vector counterparts (comparison only applicable for *popZ* and *rcy* as unmarked *parA1* knockouts were not attempted). As the antibiotic selection in these plates selected for knockout mutants only, these high colony counts were very unexpected given the previous difficulty in generating unmarked knockout mutants. This raised concerns about this reliability of this screening step. A control plating of each tri-parental mating strain (WT *R. leguminosarum* 3841, *E. coli* pRK2013 and each *E. coli* pK18-tet^R-KO donor strain) onto TY-sucrose containing str only, tet only, and both str and tet combined (at concentrations used for *R. leguminosarum* selection) confirmed that this antibiotic selection was sufficient to select against each of these strains independently. To determine the reliability of the *sacB*-mediated counter-selection, approximately 100 colonies from each conjugation (300 total) were picked onto TY Agar master plates containing both streptomycin and tetracycline. Colonies were then replica plated onto TY Agar plates containing streptomycin and kanamycin, alongside a WT *R. leguminosarum* 3841 control streak. In contrast to observations made during the generation of unmarked knockout mutants where replica plating of double-crossovers

identified only a few kan-resistant colonies, the vast majority of the tet-resistant 'knockout mutants' obtained here were found to be resistant to kanamycin, suggesting that these colonies were not true knockout mutants but were in fact still primary integrants. This observation highlighted potential complications with the sucrose sensitivity of all three tet^R knockout vectors following their integration into the genome.

In an attempt to address this reduced sucrose-sensitivity, the selection for double-crossovers was reattempted using TY-sucrose containing 15% (w/v) sucrose. Although the number of emerging colonies was only slightly (but not significantly) reduced when compared to 10% sucrose, the increased sucrose concentration had a markedly detrimental impact on colony morphology. Instead of the round, off-white, convex colonies observed on 10% sucrose, colonies on 15% sucrose were more irregular in morphology, with a watery consistency. Within these 'colonies' were off-white granule-like structures. These observations were consistent with the formation *Rhizobial* micro-colonies surrounded by secreted exopolysaccharides (EPS), often observed in the presence of high sucrose and/or during biofilm formation (*375–377*). The secretion of EPS protects these cells from osmotic stresses, such as high sugar concentrations. This altered colony morphology made the isolation of single colonies difficult. Therefore, this concentration of sucrose was simply not viable for this protocol.

As increasing sucrose concentration was not viable, the generation of tet^R mutants was attempted once more using an extended selection process. In this modified protocol, non-selective primary integrant cultures were plated onto TY-sucrose (10%) Agar plates containing streptomycin only. Three colonies (per knockout attempt) were picked from the TY-sucrose plates and streaked onto TY Agar plates containing streptomycin (str) and tetracycline (tet). These plates were incubated until single colonies appeared (2-3 days). One colony from each plate was then streaked back onto TY-sucrose (10%) plates containing streptomycin, and incubated for a further 2 days before streaking a single colony back onto TY (+str and tet) plates. This re-streaking of a single colony between TY-sucrose (+str) and TY (+str and tet) plates was repeated until the colony had been exposed to each media type a total of 4 times. As discussed previously, *sacB* is sensitive to suppressor mutations which can render this counter-selection method ineffective (*372*), therefore this repeated exposure to sucrose runs the risk of increasing the likelihood of false positives. As a result, and sucrose-sensitivity/resistance not used as a marker for whether a colony possessed a primary integration or double-crossover event. Instead, antibiotic selection was used to identify mutants. As the driving of homologous recombination using *sacB*-mediated sucrose counter-selection was more likely to be ineffective,

repeated re-streaking increased the likelihood of generating a knockout mutant by allowing more time for resolution of the pK18*mobsacB* vector from the genome to occur.

Between 100-150 colonies for each gene knockout (taken across all three final streak plates) were picked onto TY Agar master plates containing streptomycin and tetracycline. Streak plates were then incubated for a further 1-3 days, or until the plate became over-populated. Each day, the original streak plates were checked, and any later-emerging colonies were also picked onto master plates. Colonies on the TY (+str and tet) master plates were replica plated onto TY Agar plates containing streptomycin and kanamycin identify whether these colonies were kanamycin-sensitive, demonstrating successful knockout generation. Unfortunately, no kanamycin-sensitive colonies were identified following the integration of pK18-*parA1*-tet^{R}-KO into the *R. leguminosarum* genome. Therefore, all colonies were determined to be primary integrations still possessing this vector. As a result, no further screening of these colonies was conducted, and the generation of marked *parA1* knockout mutants was considered unsuccessful.

When screening for *popZ* and *rcy* knockout mutants, a small number of kan-sensitive, tet-resistant colonies were identified. To confirm whether these colonies were in fact true knockout mutants, colonies were screened using the same colony PCR screen as described in Section 3.2.3, Figure 3.10. Although the amplification of WT revertant gDNA would still produce a PCR product of the same size as in the previous PCR screen (i.e. 2.5 kb for *popZ,* and an unobtainable 8.9 kb fragment for *rcy*), the substitution of *popZ*/*rcy* with the tet^{R} gene cassette following successful gene knockout would result in a PCR fragment of 3.3 kb, as opposed to the 1.8 kb fragment generated following amplification of gDNA from an unmarked knockout mutant. Analysis of the products obtained from this PCR screen demonstrated that despite being resistant to tetracycline, all nine kan-sensitive *popZ* 'knockout' colonies identified were wildtype in genotype at the targeted region. Four of the nine colonies screened gave a distinct 2.5 kb WT band, with the remaining five producing a faint band of the same size (Figure 3.12A). A fragment of this size was also produced following amplification of WT template gDNA, supporting conclusion these colonies were wildtype in genotype at this region. Conversely, all eight potential *rcy* knockout colonies screened produced a faint 3.3 kb PCR product, indicating that successful allelic exchange and subsequent deletion of the *rcy* gene had occurred within these samples. A small degree of non-specific amplification was also observed, indicated by a very faint 1.5 kb band in all samples (including the WT gDNA sample), but this was not of concern given intensity of band.

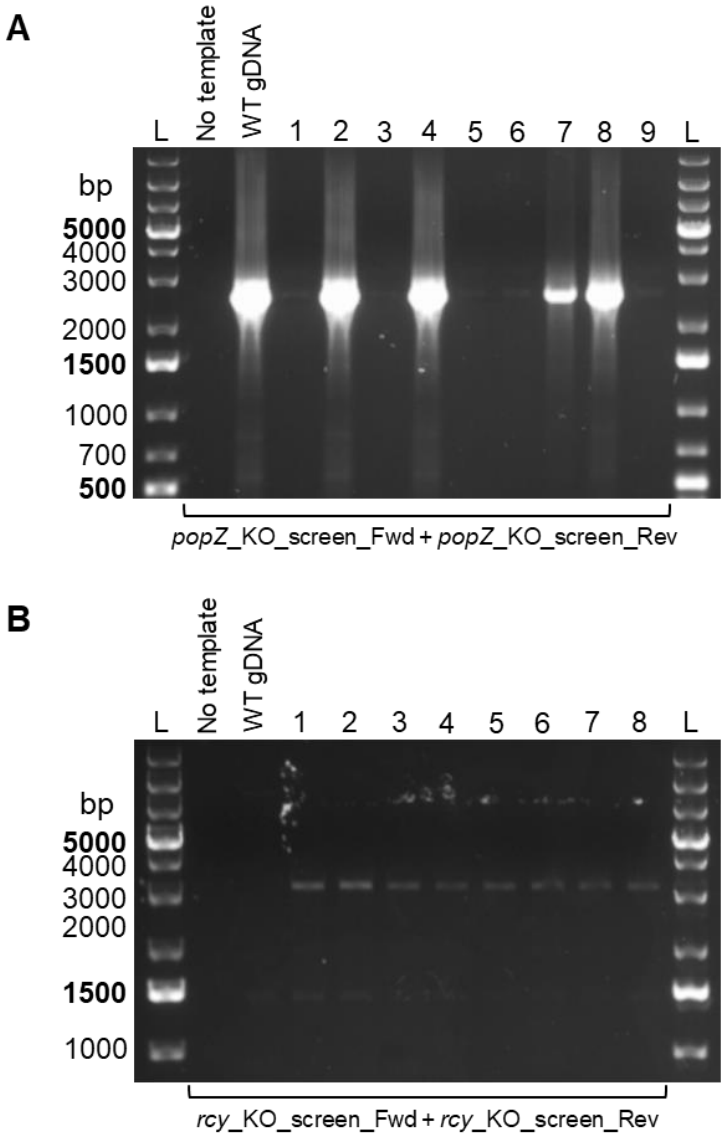


Figure 3.12. Colony PCR screening of potential tetracycline-resistant *popZ* and *rcy* knockout mutants. Potential *popZ* and *rcy* marked knockout mutants were subjected to a colony PCR screen designed to identify successful gene deletion. In this PCR, genome-specific primer pairs comprising of a forward primer annealing upstream of the 5' flank, and a reverse primer annealing downstream of the 3' flank were used. PCR reactions were then analysed by electrophoresis on a 1% (w/v) agarose gel. (A) Nine potential *popZ* mutants (Lanes 1-9) were screened with the primers *popZ*_KO_screen_Fwd and *popZ*_KO_screen_Rev. If template colony was a knockout, a 3.3 kb band would be obtained. If a WT revertant, a band of 2.5 kb would be obtained. (B) Eight potential *rcy* mutants (Lanes 1-8) were amplified with primers *rcy*_KO_screen_Fwd and *rcy*_KO_screen_Rev. Again, if the template colony was a knockout mutant, a 3.3 kb band would be obtained. However, if a WT revertant, theoretical band of 8.9 kb would be generated (too large to obtain in the PCR cycle). Common annotations: Lane L; GeneRuler 1 kb Plus DNA Ladder. For each PCR screen, two control reactions were performed using dH_2O (i.e. no template) (negative control) and WT *R. leguminosarum* 3841 gDNA (positive for *popZ*, negative for *rcy*).

Unfortunately, the results of this PCR screen were not as promising or straightforward as first hoped. All tetracycline-resistant, kanamycin-sensitive colonies screened in this PCR were identified following a 2-day incubation of the replica plates. Following an additional overnight incubation, all 17 colonies (9 '*popZ*' and 8 '*rcy*') exhibited growth in the presence of kanamycin. This raised one of two questions: how were tetracycline and kanamycin-resistant WT colonies obtained following attempted deletion of *popZ*; and how did *rcy* 'knockout mutants' grow on kanamycin?

Firstly, the growth of *R. leguminosarum* colonies resistant to both tetracycline and kanamycin following attempted generation of *popZ* knockout mutants demonstrated primary integrations of the pK18-*popZ*-tet^{R}-KO vector within the genome. However, the presence of a WT PCR product (Figure 3.12A) following amplification of the *popZ* gene region suggests that integration may not have occurred within this region of the genome (i.e. off-target integration). Secondly, the presence of kanamycin-resistant *rcy* 'knockout' colonies in conjunction with the faint knockout PCR product bands observed in Figure 3.12B indicated that these colonies may once again have been mixed populations containing a small number of knockout mutants amongst mostly primary integrant colonies. Similar to attempts to isolate potential unmarked *popZ* knockout colonies in Section 3.2.3, multiple attempts to isolate these marked *rcy* knockout colonies (including re-streaking and serial dilutions) did not yield distinct kan-sensitive, tet-resistant colonies which produced a clear PCR product indicative a successful knockout (data not shown). The inability to isolate knockout mutants despite their apparent presence within these colonies may suggest that these mutants were sick or even non-viable, but that trace amounts of their gDNA present within the colonies was still amplified by the colony PCR screen shown in Figure 3.12B.

3.4 Discussion

3.4.1 Unsuccessful deletion of *parA1* may be attributed in part to its genomic context

During the attempted generation of marked *parA1* gene deletions (Section 3.3.2), the resolution of pK18-*parA1*-tet^R-KO from the *R. leguminosarum* genome following primary integration was not observed. All colonies screened following sucrose counter-selection were resistant to both kanamycin and tetracycline, demonstrating that *sacB*-mediated excision of the vector from the genome had not occurred. In contrast to unmarked gene deletion attempts (Section 3.2.2, not performed for *parA1*), tet^R-marked primary integrants were not screened to identify which homologous flank vector integration occurred within. Nevertheless, the genomic context of *parA1* sheds a potential light on why the deletion of this gene was not successful.

The *R. leguminosarum parA1* gene is located only 19 bp downstream and 32 bp upstream of its neighbouring genes, with all three genes encoded on the same DNA strand (Table 3.1). While the expression of the upstream gene (*rsmG*) was unlikely to be affected, the deletion of *parA1* from the *R. leguminosarum* 3841 genome was likely to have impacted the expression of the downstream gene. Almost immediately downstream of *parA1* gene is the gene *parB*, which encodes the DNA binding protein ParB. As reviewed in Chapter One (Section 1.2.4), the ParA and ParB proteins work together (in conjunction with the *parS* sequence) to segregate chromosomal plasmids during cell replication (*206*, *344*). The ParA and ParB proteins are typically encoded within the same operon, with the downstream *parS* sequence either within or nearby this operon (*220*, *343*, *344*). Given their proximity, it appears that the ParABS system of *R. leguminosarum* 3841 is also encoded within an operon, possibly under the control of a single promoter.

The disruption of genomic regions within operons, even if in-frame, can impact the expression of downstream genes often resulting in either up- or downregulation of the downstream genes (*378*, *379*). The integration of the pK18-*parA1*-tet^R-KO allelic exchange vector into the 5' flanking region (i.e. upstream) of *parA1* is likely to have allowed the *R. leguminosarum* 3841 *parABS* operon structure to remain intact. However, an integration event of this vector within the 3' (downstream) flank would disrupt the operon downstream of *parA1*. Any double crossover event which resulted in the deletion of *parA1* may also have disrupted this operon and

subsequently the expression of *parB* downstream. Within several Gram-negative species including *Caulobacter crescentus*, the *parABS* locus has been demonstrated to be essential for cell viability (*380–383*). In species where the ParABS system is not essential, disruption or loss of the *parABS* locus impacts cell growth and DNA segregation, often resulting in the formation of cells lacking nucleoid DNA (*384–389*). Hence, it seems likely that the deletion of *parA1* within *R. leguminosarum 3841* and subsequent disruption of the *parABS* locus would be detrimental or potentially even lethal, making the generation of a whole-gene *parA1* knockout mutant unfavourable. That being said, a *parA* null mutant (in-frame partial gene deletion) within the rhizobial species *Azorhizobium caulinodans* (belonging to the *Xanthobacteraceae* family) has been successfully generated (*227*). Although *ΔparA A. caulinodans* cells displayed pleiomorphic morphologies and dramatic defects in chromosome segregation, the deletion of *parA* was seemingly not lethal within this species. In support of this, ParA deletion (truncation mutants) in *Pseudomonas aeruginosa* (Gram-negative rod) impacts growth rate and chromosome segregation, leading to the formation of anucleate cells. ParA inactivation also made ParB unstable, leading to increased degradation of ParB (*390*). However, it is very important to note that the gene deletions mentioned above were not whole-gene deletions like those attempted in this study, but instead partial gene deletions. Hence, disruption to the *parABS* locus downstream of *parA* may have been minimised within these other studies.

Table 3.1. Genomic context of *parA1*, *popZ* and *rcy* within *R. leguminosarum* 3841. The distances of *parA1*, *popZ* and *rcy* from their neighbouring genes have been detailed in this table. Predicted/hypothetical gene products have been provided instead of gene names as many of these genes remain uncharacterised. The *Rhizobium leguminosarum* bv. viciae 3841 genome sequence and gene predictions were sourced from ref. (*329*).

Target gene	Distance from 5' gene (bp)	5' gene orientation	5' gene product	Distance from 3' gene (bp)	3' gene orientation	3' gene product
parA1	19	Forward	16s rRNA methyltransferase G	32	Forward	ParB
popZ	334	Forward	Protein-L-isoaspartate O-methyltransferase	115	Forward	Valine—tRNA ligase
rcy	462	Reverse	Hpt domain-containing protein	111	Forward	Cysteine hydrolase

During the attempted generation of unmarked *popZ* knockout mutants, the integration of pK18-*popZ*-KO only occurred within the 3' flanking region of this gene and attempts to delete this gene were unsuccessful. Unfortunately, unlike *parA1*, the genomic context of *popZ* provides little insight as to why this may have been the case. The *popZ* gene sits over 100 bp from each of its neighbouring genes (Table 3.1) and does not appear to be encoded within an operon. Within *Agrobacterium*, the loss of *popZ* and results aberrant cell division, defects in chromosome segregation, and even cell mortality (*326*). Potential detrimental deletion phenotypes of *popZ* within *R. leguminosarum* may be the reason double crossover events yielding successful *popZ* deletion were not easily isolated, but the reason behind the lack of plasmid integration within the 5' flank remains unknown.

3.4.2 Deletion of *rcy* in *R. leguminosarum* may reduce cell viability

Of all three gene deletion attempts, *rcy* was the only gene which demonstrated the potential for successful knockout mutant generation using the pK18*mobsacB*-mediated knockout protocol adopted within this thesis. Although the generation of unmarked mutants was unsuccessful yielding only wildtype (WT) revertants, evidence of integration of the unmarked allelic exchange vector within both flanking regions was observed. Attempts to generate tetracycline-resistant marked knockout mutants resulted in the identification of *R. leguminosarum* colonies which appeared to produce a colony PCR product indicative of a successful knockout genotype. Yet, unexpected resistance to kanamycin led to the conclusion that these colonies comprised of a mixed population of knockout mutants and primary integrants. Attempts to isolate knockout mutants from primary integrants were unsuccessful, ultimately leading to the suggestion that *rcy* knockout mutants were likely to be very sick or potentially non-viable (Section 3.3.2).

The possibility of a severely detrimental phenotype following *rcy* deletion is consistent with phenotypes observed following the loss or depletion of other 'Rcy-like' proteins. Within *Agrobacterium*, the loss of GPR (a protein with very similar domain architecture to Rcy - see Section 1.5.2) was often lethal (*278*). In cells which did survive, loss of GPR resulted in slow-growing cells with doubling time 2-3 times longer than wildtype cells. Not only were these *Δgpr* cells slow-growing, but they also exhibited significant morphological and growth defects (*278*). Additionally, the loss of Scy within *Streptomyces* causes developmental defects including smaller

colony size and reduced sporulation (*182*). Ultimately, a reduced viability of *rcy* knockout mutants may explain why even though potential mutants were identified following colony PCR, attempts to isolate these mutants were unsuccessful. It is probable that *rcy* knockout mutants, if obtained, were repeatedly outcompeted by colonies still expressing Rcy.

3.4.3 Potential off-target integration of the pK18-*popZ*-tet^R-KO allelic exchange vector

Following the attempted generation of marked *popZ* deletions, a number of *R. leguminosarum* colonies were identified which were resistant to both kanamycin and tetracycline (indicative of a primary integration of the allelic exchange vector) yet appeared to be undisrupted at the genomic region surrounding *popZ* (Section 3.3.2, Figure 3.12A). These observations were suggestive of primary integrations of the pK18-*popZ*-tet^R-KO vector at a location other than the targeted genomic region (i.e. the *popZ* flanking regions). In contradiction, the integration of the unmarked *popZ* allelic exchange vector (pK18-*popZ*-KO) appeared to occur within targeted regions only. Evidence of pK18-*popZ*-KO integration within the targeted flanking regions was observed following colony PCR screening (Section 3.2.2, Figure 3.8) (albeit only within the 3' flank), and the resolution of the vector from the genome following sucrose counter-selection restored kanamycin sensitivity (Section 3.2.3).

A nucleotide BLAST (BLASTN) of the tet^R cassette introduced into these vectors against the *R. leguminosarum* 3841 genome highlighted two short regions with some degree of homology; a 66 bp region with 77% homology, and 83 bp region with 72% homology. However, given their short length and relatively low homology, these regions are not enough to be a convincing site of vector integration. Interestingly, both of these regions lie within a gene (*RL3414*) predicted to encode a tetracycline MFS (major transporters facilitator superfamily) efflux pump, similar in function to that encoded by the *tetA* gene introduced into these vectors (*329*, *356*, *391*). Despite its predicted function, this putative tetracycline-resistance gene did not confer resistance to tetracycline within wildtype *R. leguminosarum* 3841 at the concentrations used within this study (5 µg mL^{-1} in TY media), and therefore does not in itself explain the tetracycline resistance observed within WT revertants.

3.4.4 Other complications: antibiotic resistance, reduced sucrose-sensitivity, and mixed colonies

The *sacB* counter-selection system used in this thesis (i.e. pK18*mobsacB*, with selection on TY containing 10% sucrose) has been successfully used to delete the *R. leguminosarum* 3841 genes *ribN* (*RL1692*), encoding the riboflavin transporter RibN (*363*) and *rapD* (*RL2702*) which encodes a biofilm matrix-associated calcium-binding protein RapD (*364*). Although the deletion of *parA1*, *popZ* and *rcy* from *R. leguminosarum* 3841 was ultimately unsuccessful, the use of this system during the attempted generation of unmarked knockout mutants appeared to work as expected. Colonies carrying primary integrations of the unmarked knockout vector were kanamycin resistant, those which had undergone a double crossover event (i.e. WT revertants) were kanamycin sensitive, and *sacB*-mediated resolution of vectors from the genome was reasonably efficient. A few primary integrants remained following sucrose counter-selection, but this is not unusual given the high suppressor mutation rate of *sacB* (*372*).

Despite observing effective antibiotic selection and sucrose counter-selection during the generation of unmarked mutants, issues with these screening steps were encountered following the introduction of the tetracycline resistance (tet^R) cassette to the unmarked pK18-derived allelic exchange vectors. Following sucrose counter-selection of primary integrants carrying these tet^R-marked vectors, all isolated colonies appeared to be resistant to both kanamycin and tetracycline. As resistance to both antibiotics indicated primary integration of the vector into the genome, this demonstrated a significant reduction in sucrose sensitivity within cells carrying these vectors. The exact reason for this sudden loss of sucrose sensitivity following introduction of *tetA* remains unknown, but it is likely to have been the result of poor optimisation of the selection media. The use of *tetA-sacB* selection/counter-selection systems has been demonstrated to be highly sensitive to selection conditions, often requiring species-specific optimisation of the selection media (*392–394*). TetA has also been associated with increased osmotic sensitivity of cells (*395*), which may have further influenced the effectiveness of sucrose counter-selection within this study.

While the sensitivity of *tetA-sacB* selection systems and the use of an inappropriate or poorly-optimised selection media may explain the issues with sucrose-sensitivity and/or tetracycline resistance observed within this thesis, the question still remains as to why widespread

kanamycin resistance was observed. Although the introduction of additional antibiotic resistance genes can also increase a cell's resistance to other antibiotics, the introduction of *tetA* has been linked to an increase in sensitivity to kanamycin, not resistance (*394*, *396*). Therefore, it seems most likely that the widespread resistance to kanamycin observed during the attempted generation of tet^R-marked knockout mutants is not a direct result of *tetA* introduction, but is instead the consequence of reduced sucrose sensitivity, allowing for the vector to remain integrated within the genome following sucrose counter-selection.

One final issue encountered during the generation of gene knockouts was genetically heterogeneous colonies. The observation of such colonies was not unique to either marked or unmarked gene deletion attempts. Additionally, the use of higher concentrations of sucrose within the selection media gave rise to unusual colony morphologies comprising of 'micro-colonies' surrounded by exopolysaccharides (EPS) (Section 3.3.2). These colony morphologies are common in the formation of biofilms, and in the presence of high sucrose (or similar osmotic stresses) (*375*–*377*). If these knockout attempts were to be repeated, the use of a minimal or low-salt media may help to resolve the difficulties with isolating single colonies in presence of sucrose. The use of low salt selection media can enhance sensitivity to sucrose during *sacB* counter selection (*397*), allowing for the use of lower sucrose concentrations, as demonstrated by the successful deletion of regions within an uncharacterised *R. leguminosarum* 3841 operon (*RL3499-RL3502*) associated with cell membrane integrity (*398*). Here, Vanderlinde *et al.* selected for double crossover mutants on Vincent's minimal medium (VMM) with only 5% sucrose. In this study, sucrose sensitivity was conferred by the *sacB* gene of the vector pJQ200SK (*398*, *399*). A reduction in sucrose concentration may therefore help isolation of single colonies by minimising the formation of heterogeneous biofilm-like colonies and the secretion of EPS which may have interfered with various screening steps and mutant colony isolation throughout this chapter.

3.4.5 Alternative gene deletion approaches

There are of course a great number of other gene deletion approaches which could have been used to generate *parA1*, *popZ* and *rcy* deletions, ranging from the generation of different types of mutations (e.g. partial deletions, truncations and frameshifts), to the use of different gene editing techniques altogether. Within this section, a handful of accessible alternative approaches which align most closely to those used within this thesis have been discussed. Despite unsuccessful attempts to generate *parA1*, *popZ* and *rcy* gene knockouts within *R. leguminosarum* (this study), viable deletions of *parA*, *popZ* and the putative *rcy* homolog *gpr* have been achieved within other members of the Hyphomicrobiales. Within *Agrobacterium tumefaciens*, both whole-gene and various in-frame deletions of *popZ* have been generated, as well as a partial deletion of *gpr* (2378 bp deletion consisting of the first 2140 bp of *gpr* (6345 bp) and 238 bp upstream) (*278*, *282*, *326*). In the rhizobial species *Azorhizobium caulinodans*, deletion of ParA was achieved by generating a partial deletion (codons 7-150) of the *parA* gene (*227*). The generation of each of these mutants was achieved using various *sacB*-based suicide plasmid-mediated genome editing approaches, and in the case of *parA* deletion, a pK18*mobsacB*-mediated gene deletion approach was used. However, one key difference between the knockouts attempted in this thesis and those described above is the type of deletion generated. With the exception of a *popZ* whole-gene deletion, all other deletions were instead partial gene deletions. Attempts to generate similar partial gene deletions within the *R. leguminosarum* genome may prove a successful alternative approach to gene knockout, particularly for *parA1* where whole-gene deletion was predicted to have downstream polar effects on the *parABS* locus (see Section 3.4.1).

Despite widespread use of *sacB* counter-selection systems (including pK18*mobsacB*) within the Hyphomicrobiales, potential issues with this selection system were encountered within this thesis. Many of the suicide vectors available for use in Gram-negative bacteria utilise the *sacB* gene (*399*, *400*). However, alternative compatible counter-selectable markers such as *pheS* (confers sensitivity to the phenylalanine analog *p*-chloro-phenylalanine) and *thyA* (confers sensitivity to trimethoprim) are available (*401–405*). Substitution of the *sacB* gene encoded on the pK18*mobsacB*-derived knockout plasmids generated in this chapter with an alternative counter-selectable marker may therefore provide a relatively straightforward solution to the *sacB*/sucrose-related complications encountered within this study. Such counter-selectable markers could also be used to generate a valuable additional resource intended to facilitate the

characterisation of these potentially essential *R. leguminosarum* genes. To delete genes which encode proteins essential for cell viability, a complementing copy of the gene must first be introduced to the cell (typically encoded on a plasmid). However, in order to study the effects of gene deletion, the complementing copy must subsequently be inactivated or lost (*406*, *407*). Typically, the loss of this copy can be achieved using either a counter-selectable marker also encoded on the plasmid, or by expressing the gene under the control of a tightly regulated inducible promoter. In Chapter Four, IPTG-inducible *parA1*, *popZ* and *rcy* expression plasmids were generated (see Section 4.2 for further details). Although the IPTG-inducible promoter of these plasmids was not tightly regulated within *Rhizobium* (see Section 5.4.2), the incorporation of a counter-selectable marker would generate a gene expression plasmid suitable for complementation during genomic gene deletion attempts. Transformation of these counter-selectable complementation plasmids into *R. leguminosarum* 3841 cells may permit pK18*mobsacB*-mediated deletion of the genomic *parA1*, *popZ* and *rcy* genes should they indeed by essential. Adoption of this approach may be of particular value for the characterisation of *rcy*, which has been proposed here to encode a protein essential for cell viability.

In addition to suicide plasmid-mediated techniques, other commonly-used genome editing techniques include Red homologous recombination and CRISPR/Cas9-based systems (*362*, *408*, *409*). Of these two methodologies, Red homologous recombination (also referred to as Red recombineering) is particularly relevant to this thesis. Red recombineering exploits the functions of three lambda (λ) Red bacteriophage proteins (Gam, Exo and Beta) to facilitate recombination between genomic (or plasmid) DNA and a linear allelic exchange construct *in vivo* (*410*, *411*). Within Chapter Five, a λ Red recombineering system developed by Datsenko and Wanner (2000) (and further modified by Gust *et al.* (2002)) was successfully employed to incorporate EGFP cassettes into plasmids within *E. coli* (*349*, *351*). However, this exact system (which has been described in detail in Section 5.2.1) cannot readily be applied to the generation of gene knockouts within *R. leguminosarum* without substantial modification. Firstly, expression of the λ Red genes (*bet*, *gam*, *exo*) (encoded on a plasmid referred to as the λ Red recombination plasmid) was driven by an arabinose-inducible promoter (pBAD). This λ Red gene expression system has previously been demonstrated to be ineffective within *Agrobacterium,* and therefore is likely to be ineffective within *Rhizobium* also given the close phylogenetic relationship of these genera (*412*). Secondly, a temperature sensitive replicon present on the λ Red recombination plasmid was used to inhibit expression of the λ Red genes following plasmid modification. In this step, a temperature shift from 30°C to 37°C was required, which would not be compatible with

R. leguminosarum which has an optimal incubation range of 25°C to 31°C (*413*). Finally, host compatibility issues resulting from differences in the function of the λ Red proteins within different species makes Red recombineering in species more distantly related to *E. coli* substantially less effective (*362*, *414*). It is worth mentioning that the modification and optimisation of phage recombineering for use within 'non-conventional' species is not impossible, and both λ Red and ET (the RecE/RecT recombinase system from the Rac prophage) phage recombinase gene editing systems have been successfully employed within *Agrobacterium* (*362*, *412*, *414*). However, these systems are still a long way off becoming commonplace, and therefore have not been considered an accessible alternative approach to gene knockout within the scope of this thesis.

Chapter Four – Characterisation of ParA1, PopZ and Rcy by protein overexpression

4.1 Introduction

Following unsuccessful attempts to generate *parA1*, *popZ* and *rcy* gene knockout mutants (Chapter Three), the overexpression of these genes was employed as an alternative approach to study gene function. In this study, protein overexpression was achieved *in vivo* by expression of the full-length *parA1*, *popZ* and *rcy* genes from the isopropyl ß-D-1-thiogalactopyranoside (IPTG)-inducible promoter of plasmid pIND6. Following the introduction of these plasmids into *R. leguminosarum* 3841, novel overexpression phenotypes within these strains were analysed to determine the involvement of these genes, and their encoded proteins, in the regulation of growth and cell morphology within *R. leguminosarum*. In this chapter, the growth of these IPTG-inducible gene overexpression strains was monitored through the generation of growth curves. Total protein extracts were then isolated from culture samples and analysed using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) to identify expression of the ParA1, PopZ and Rcy proteins following induction of gene overexpression. Finally, *R. leguminosarum* cells overexpressing ParA1, PopZ and Rcy were viewed using phase contrast microscopy to observe any changes in cell morphology resulting from the overexpression of these proteins.

4.2 Generation of *R. leguminosarum* gene overexpression strains

4.2.1 The IPTG-inducible expression vector pIND6

Overexpression of *parA1*, *popZ* and *rcy* within *R. leguminosarum* 3841 was achieved using the IPTG-inducible expression vector pIND6 (Kelemen, unpublished. Figure 4.1). This vector is a direct derivative of the vector pIND4 (*355*), differing only by a few modifications to restriction sites within the multiple cloning site (MCS). The vector pIND4 was originally developed for use in the Alphaproteobacteria species *Rhodobacter sphaeroides* and *Paracoccus denitrificans*, thereby making its derivative pIND6 a suitable choice of expression vector for use within *R. leguminosarum*. Within pIND4, cloning of a given gene into the NcoI site within the MCS allows for the expression of C-terminally histidine-tagged proteins from the IPTG-inducible *lac*

promoter of this vector. To generate pIND6, additional NdeI and EcoRI sites were introduced into the MCS of pIND4 (not utilised in this thesis, therefore not included in Figure 4.1). However, it important to note that these modifications to pIND4 resulted in the removal of the hexa-histidine tag (6xHis-tag) within this vector, meaning pIND6 lacks this feature.

IPTG-inducible expression from this vector is achieved by the presence of *lacI*^{q} (*lacI* driven by the modified *lacI*^{q} promoter), in conjunction with an IPTG-inducible expression cassette. This cassette (comprising of the $P_{A1/04/03}$ promoter (a modified *lac* promoter), a ribosome binding site (RBS), a multiple cloning site, and a transcriptional terminator) allows for IPTG-inducible expression of genes following cloning of these genes downstream of the $P_{A1/04/03}$ promoter within this cassette. Importantly, expression levels from this plasmid are sensitive to the concentration of IPTG present in the growth medium, allowing for dose-dependent induction of gene expression (*355*). In addition to IPTG-inducible gene expression, pIND6 also confers kanamycin resistance via the presence of a kanamycin resistance gene (*kan*^{R}) encoding aminoglucoside phosphotransferase. The backbone of this vector is derived from the naturally occurring *Rhodobacter* plasmid pMG160 allowing for stable replication of this vector within *R. leguminosarum* 3841 (*355*, *415*), while the presence of the ColE1 origin of replication allows for stable replication within *E. coli* (Figure 4.1).

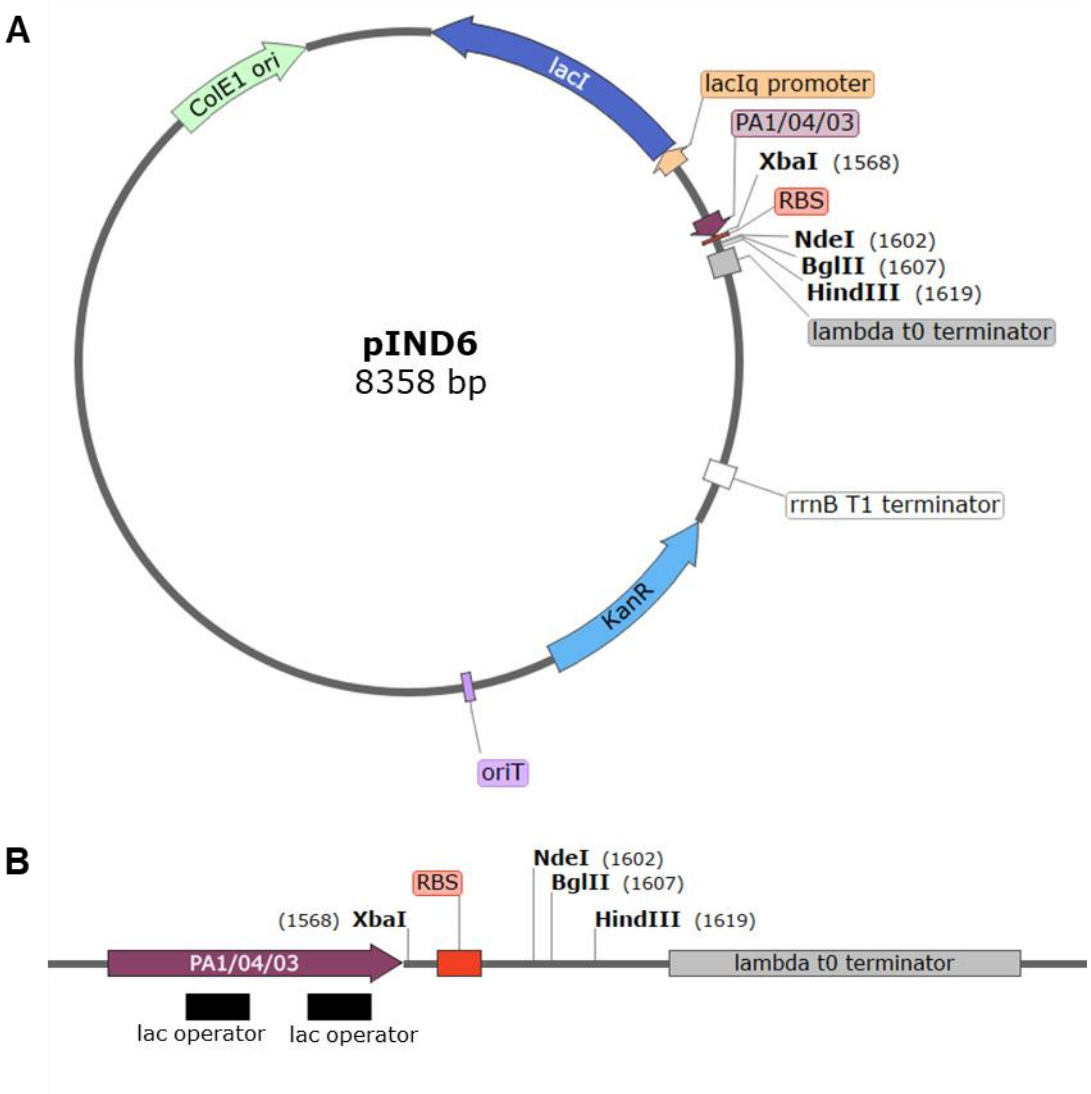


Figure 4.1. The IPTG-inducible expression vector pIND6. Inducible expression of genes within *R. leguminosarum* 3841 was achieved using the vector pIND6, a derivative of pIND4 (*355*). A plasmid map of this vector has been shown in Part A of this figure. Gene sequences were cloned downstream of the $P_{A1/04/03}$ promoter region (burgundy) allowing for IPTG-inducible gene expression. Key features of this plasmid include the *lacI* gene (encodes *lac* repressor protein) (indigo), driven by *lacI*^{q} promoter (orange), a kanamycin resistance (Kan^{R}) gene (blue), a ribosomal binding site (RBS) (red), origin of transfer (oriT, purple), and the lambda t0 and rrnB T1 transcriptional terminators (grey & white). The ColE1 origin of replication (light green) allows for replication of this plasmid in *E. coli*. The backbone of this vector (between oriT and ColE1) is derived from the naturally occurring *Rhodobacter* plasmid pMG160, allowing for stable replication of this vector within *R. leguminosarum* 3841 (*355*, *415*). Restriction sites relevant to this thesis have been shown in this plasmid map. (B) Linear representation of the IPTG-inducible expression cassette of pIND6, showing the *lac* operator regions (black) within the $P_{A1/04/03}$ promoter region, an RBS (red), and the lambda t0 transcriptional terminator (grey).

4.2.2 Generation of pIND6-derived expression vectors

Similar to the generation of the unmarked knockout constructs (Section 3.2.1), gene expression constructs were designed *in silico* and generated by commercial gene synthesis (Section 2.3.7). In this study, the expression of N-terminal His-tagged *R. leguminosarum* proteins (ParA1, PopZ and Rcy) from pIND6 was required to allow for immunodetection of the expressed protein. To achieve this, relatively simple expression constructs were generated, comprising of an RBS, 6xHis-tag, a thrombin site, and the full-length wildtype *R. leguminosarum* 3841 gene sequence (Figure 4.2A). Incorporation of 5' XbaI and 3' BamHI restriction sites to each of these constructs allowed for the subcloning of these synthetic constructs (as XbaI-BamHI fragments) into the XbaI and BglII sites of pIND6 (located downstream of the $P_{A1/04/03}$ promoter) (Figure 4.2B), generating the plasmids pIND6-*parA1*, pIND6-*popZ* and pIND6-*rcy* (Figures 4.3 and 4.4). Even though the recognition sequences of BamHI and BglII differ, these enzymes are isocaudomers and generate complimentary overhangs allowing for the ligation of the BamHI-digested construct fragments into BglII-digested pIND6 (Figure 4.2C). Despite pIND6 possessing an RBS within its IPTG-inducible expression cassette (Figure 4.1), incorporation of an RBS into the synthetic expression constructs was necessary as XbaI-BglII digestion of pIND6 resulted in the excision of the RBS from this vector. Here, the bacteriophage T7 gene 10 ribosome binding site (RBS) sequence was used to enhance translation of these genes (*416*). Although not relevant in this thesis, a thrombin cleavage site was included within the linker region between the hexa-histidine (6xHis) tag and the gene sequence to allow for cleavage of the His tag from the protein if desired in future work. The full sequence of these expression constructs can be found within Appendix A.3.

Following subcloning of the synthetic constructs into pIND6 (Section 2.6.3), the resulting pIND6-*parA1*, pIND6-*popZ* and pIND6-*rcy* clones were validated by restriction enzyme digestion. In this digest, illustrated in Figure 4.2D, XbaI digestion was used to linearise the ligated plasmids. This endonuclease was also used for a double-digestion of these plasmids which was designed to liberate the gene expression constructs from the pIND6 vector backbone. However, as the BamHI and BglII recognition sites were lost during subcloning, a HindIII site downstream of the expression constructs was used as the second enzyme for this double-digest. For the plasmids pIND6-*parA1* and pIND6-*popZ*, XbaI-HindIII digestion cleaves the complete expression construct from pIND6 backbone, generating two DNA fragments of 8.3 kb (backbone) and either 958 bp (*parA1* expression construct) or 1 kb (*popZ* expression construct). However, due to an internal

HindIII site within the *rcy* gene sequence, XbaI-HindIII digestion of pIND6-*rcy* generates three DNA fragments of 8.3, 5.9 and 1.4 kb.

Figure 4.2E shows gel electrophoresis of DNA fragments generated following the digestion of a single pIND6 (control), pIND6-*parA1*, pIND6-*popZ* and pIND6-*rcy* plasmid. Except for a small 51 bp fragment generated following the digestion of pIND6 with XbaI and HindIII which was not resolved, all expected fragments described above were identified. Following successful validation, pIND6-derived expression vectors were then conjugated into *R. leguminosarum* 3841 via triparental mating (Section 2.4.2) to generate the *R. leguminosarum* strains used for gene overexpression studies throughout the remainder of this chapter. Hereafter, these strains are referred to as *R.leg*-pIND6-*parA1*, *R. leg*-pIND6-*popZ* and *R.leg*-pIND6-*rcy*. In addition to the pIND6-derived expression vectors described here, an empty pIND6 vector was also conjugated into *R. leguminosarum* 3841 to generate the empty vector control strain *R.leg*-pIND6.

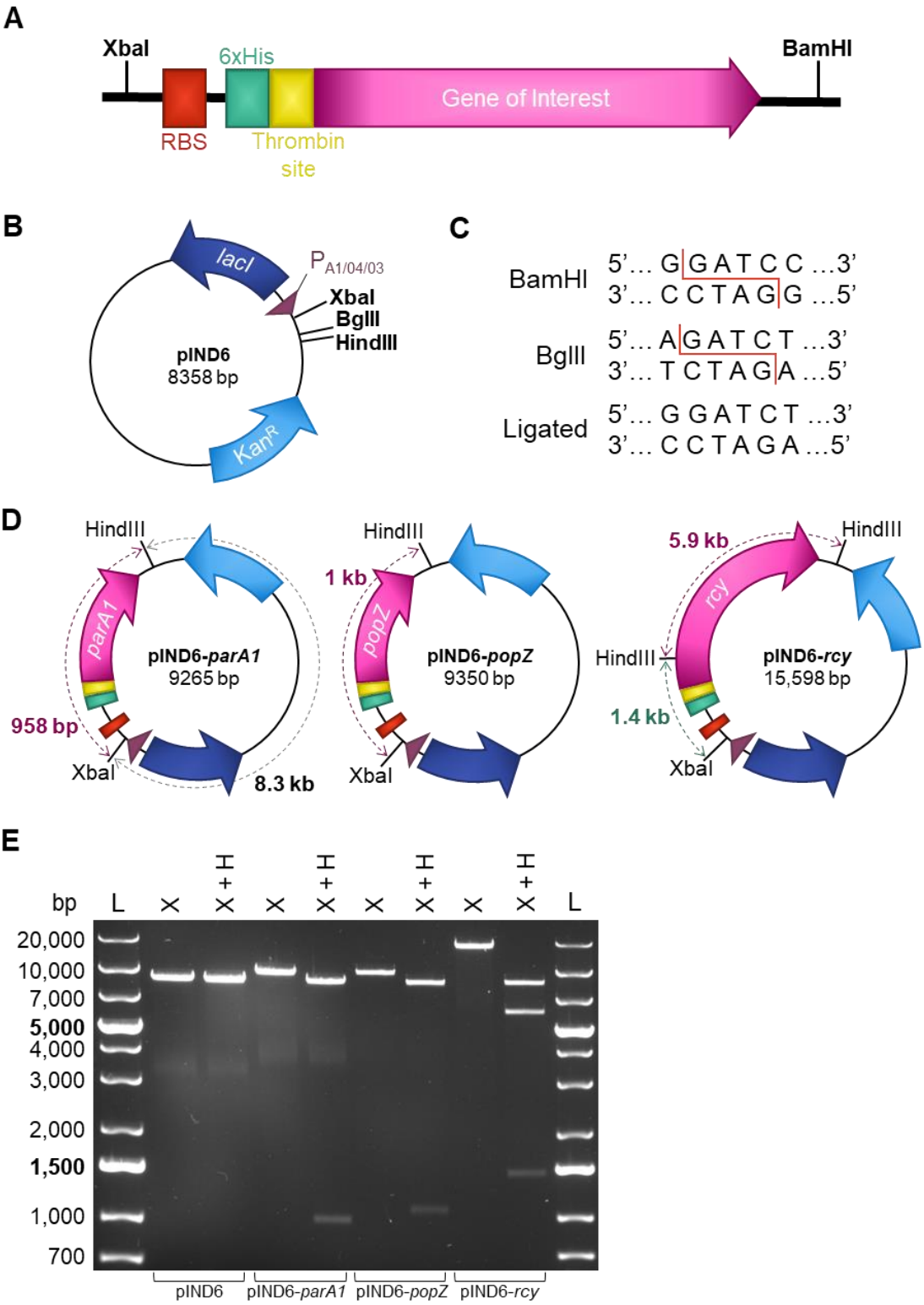


Figure 4.2. Generation and validation of pIND6-derived *parA1*, *popZ* and *rcy* expression plasmids. (A) Expression constructs comprising of a ribosomal binding site (red) (TTTGTTTAACTTTAAGAAGGAGA), a hexa-histidine tag (CATCATCATCATCATCAC) (turquoise), thrombin cleavage site (CTGGTGCCGCGCGGCAGC) (yellow), and the full-length *R. leguminosarum* gene sequence *parA1*, *popZ* or *rcy* (pink) were synthesised by artificial gene synthesis. Flanking XbaI (5') and BamHI (3') restriction sites allowed for subcloning into the XbaI and BglII sites of pIND6 (Part B). (C) BamHI and BglII generate complementary overhangs,

allowing for ligation of fragments produced by these enzymes. However, ligation of these fragments destroys each recognition sequence. (D) Plasmids pIND6-*parA1*, pIND6-*popZ* and pIND6-*rcy* were subjected to two validation digests: an XbaI digest to linearise the vectors, and an XbaI/HindIII double digest designed to cleave the expression constructs from the pIND6 backbone. Approximate fragment sizes are shown in this schematic. (E) Agarose gel electrophoresis (1% w/v) of restriction digest products. Lane L: GeneRuler 1kb Plus DNA ladder. Letters above sample lanes indicate the enzymes used for digestion (X: XbaI, H: HindIII). Exact digest product sizes (bp) were as follows: (pIND6) X: 8358, X+H: 8307, 51; (pIND6-*parA1*) X: 9265, X+H: 8307, 958; (pIND6-*popZ*) X: 9350, X+H: 8307, 1043; (pIND6-*rcy*) X: 15598, X+H: 8307, 5868, 1423.

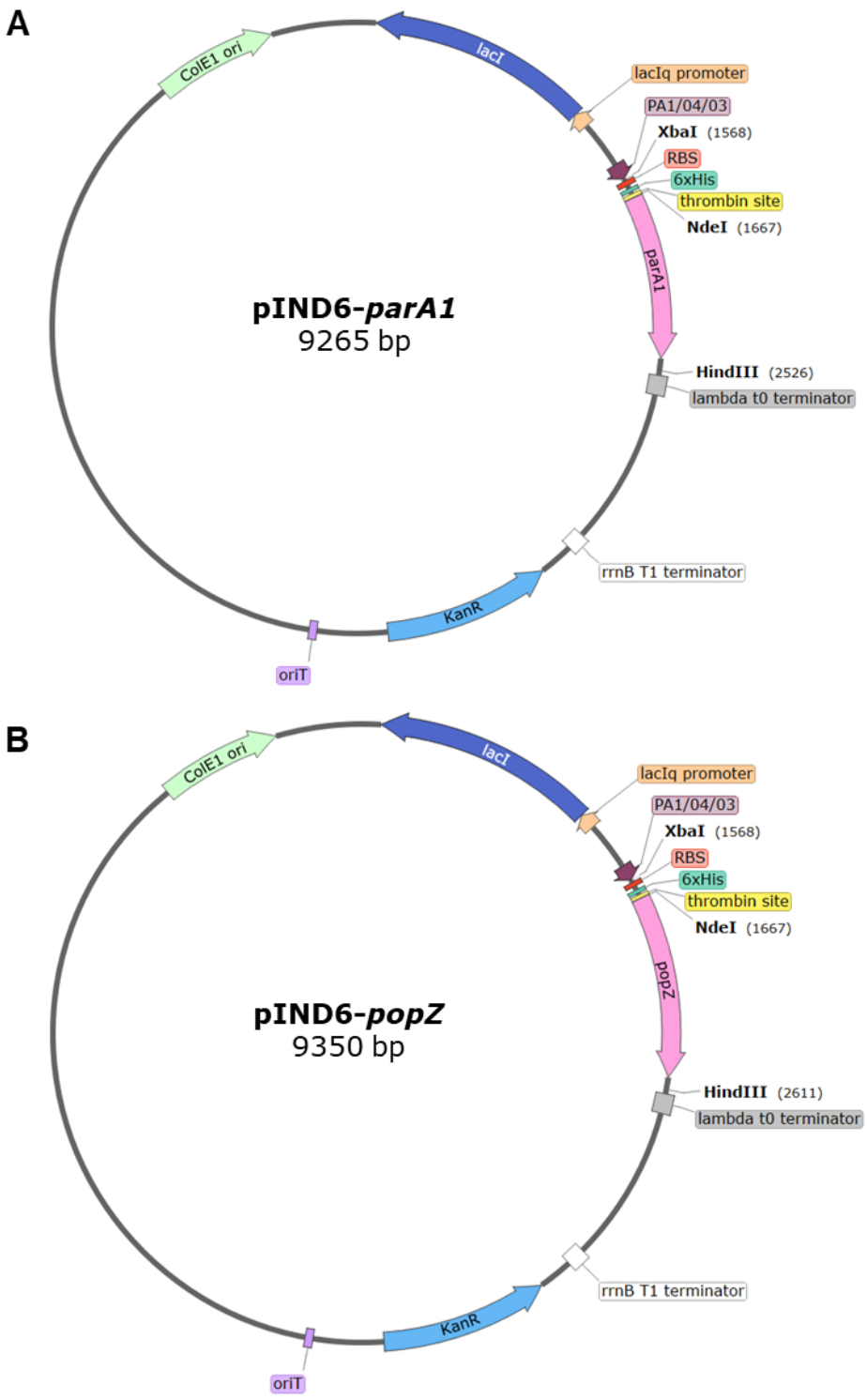


Figure 4.3. The IPTG-inducible *parA1* and *popZ* expression vectors pIND6-*parA1* and pIND6-*popZ*. Synthetic gene expression constructs comprising of a ribosomal binding site (red), 6xHis-tag (turquoise), thrombin cleavage site (yellow) and the *parA1* or *popZ* gene sequence were subcloned downstream of the $P_{A1/04/03}$ promoter region (burgundy) of pIND6 generating the plasmids pIND6-*parA1* (Part A) and pIND6-*popZ* (part B). Key features of these plasmids include the *lacI* gene (indigo) driven by *lacI*^{q} promoter (orange), a kanamycin resistance (Kan^{R}) gene (blue), origin of transfer (oriT, purple), the lambda t0 and rrnB T1 transcriptional terminators (grey & white) and a ColE1 origin of replication (light green).

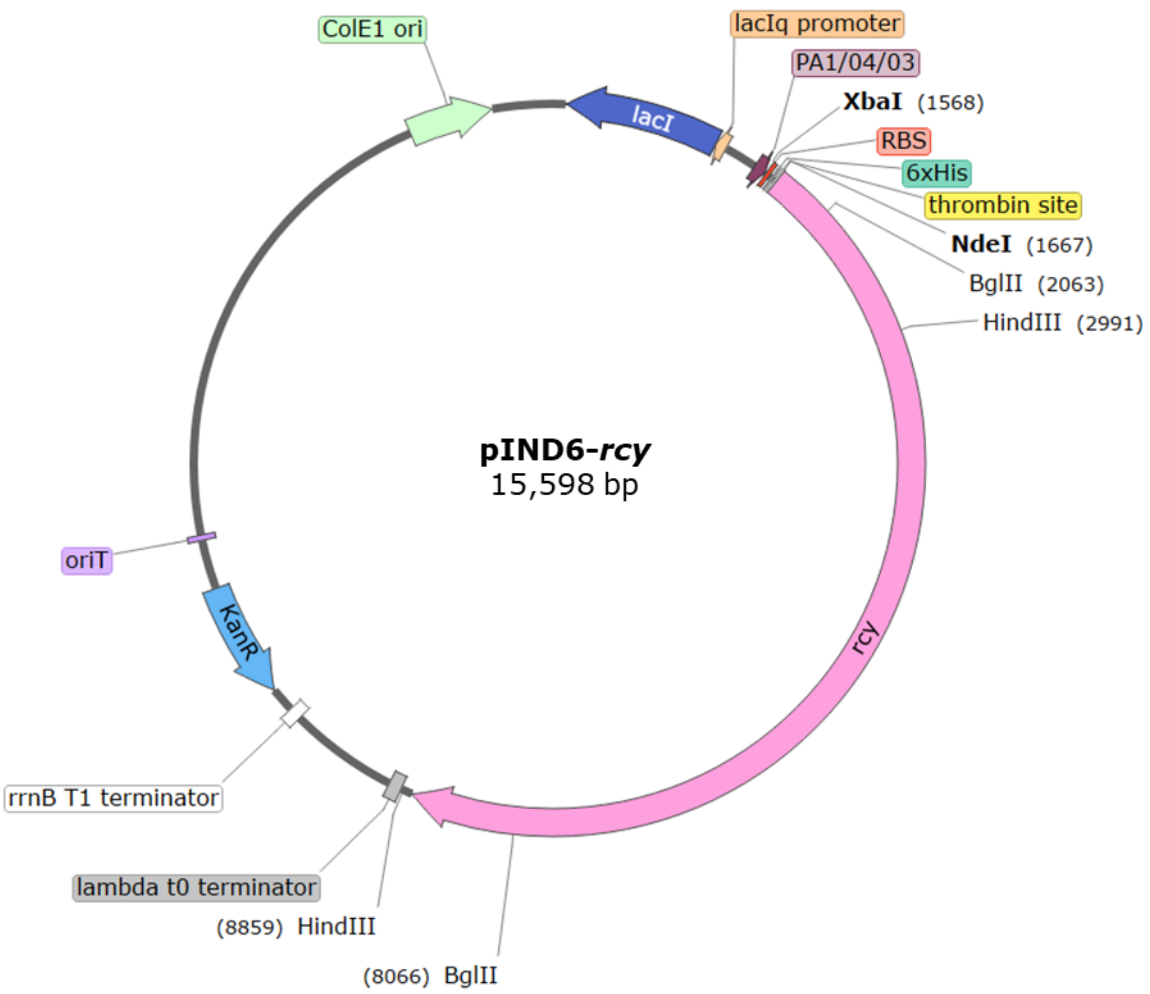


Figure 4.4. The IPTG-inducible *rcy* expression vector pIND6-*rcy*. The synthetic *rcy* expression construct, consisting of a ribosomal binding site (red), 6xHis-tag (turquoise), thrombin cleavage site (yellow) and the *rcy* gene, was subcloned downstream of the $P_{A1/04/03}$ promoter region (burgundy) of pIND6. Key features of this plasmid include the *lacI* gene (indigo) driven by *lacI*^{q} promoter (orange), a kanamycin resistance (Kan^{R}) gene (blue), origin of transfer (oriT, purple), the lambda t0 and rrnB T1 transcriptional terminators (grey & white) and a ColE1 origin of replication (light green).

4.3 Impact of *parA1*, *popZ* and *rcy* overexpression on *R. leguminosarum* growth

The first step in the effort to characterise *parA1*, *popZ* and *rcy* within *Rhizobium* was to study the impact of their overexpression on the growth of *R. leguminosarum.* To achieve this, growth curves of the gene overexpression strains described in Section 4.2 were generated to monitor the effect of overexpression of these genes. If the proteins encoded by these genes contribute to the coordination of polar growth, then it is likely that disruption of their normal regulation (i.e. through overexpression) would subsequently alter the growth of cells. Furthermore, to develop a more thorough understanding of the roles of these genes, gene overexpression was induced at various phases of growth to determine any growth phase-dependent overexpression phenotypes.

4.3.1 Establishment of experimental parameters

For *R. leguminosarum* growth curves, the gene overexpression strains *R.leg*-pIND6-*parA1*, *R.leg*-pIND6-*popZ*, *R.leg*-pIND6-*rcy* and *R.leg*-pIND6 (empty vector control) were each cultured in 50 mL of Tryptone-Yeast extract broth (TY), a rich growth medium. The procedure for the set-up and monitoring of these growth curve cultures is described in Section 2.1.3. To ensure high levels of overexpression, the induction of gene expression from the pIND6-derived plasmids was achieved by the addition of a final concentration of 1 mM IPTG to the growth medium (*355*). The impact of gene overexpression on the growth of these strains was then assayed by monitoring the optical density of the cultures at a wavelength of 600 nm (OD_{600}) as a function of time.

Before studying the overexpression of *parA1*, *popZ* and *rcy*, a preliminary growth curve was performed. In this curve, shown in Figure 4.5, the pIND6 empty vector control strain *R.leg*-pIND6 and wildtype *R. leguminosarum* 3841 were each grown in the presence and absence of 1 mM IPTG. This preliminary assay served multiple purposes, including checking the suitability of experimental conditions, determining the approximate durations of each growth phase, and ultimately deciding whether the *R.leg*-pIND6 control strain was a suitable comparison to the wildtype strain *R. leguminosarum* 3841. To ensure the retention of pIND6 and pIND6-derived expression plasmids, all pIND6-possessing *R. leguminosarum* strains were cultured in the presence of kanamycin (200 µg mL^{-1}). Therefore, it was important to understand how the growth

of these strains in the presence of kanamycin may differ from non-selective growth of wildtype *R. leguminosarum* 3841.

These preliminary growth curves (Figure 4.5) showed that the presence of 200 µg mL^{-1} kanamycin slightly slowed the growth of *R. leguminosarum* when compared to non-selective wildtype *R. leguminosarum* 3841 cultures, delaying exponential growth phase by approximately 6 hours. This extended lag phase is most likely a result of the metabolic burden associated with the replication and expression of exogenous plasmids i.e. pIND6 and the Kan^R resistance gene (aminoglucoside phosphotransferase) encoded on this plasmid (*417*). However, the addition of IPTG (final concentration of 1 mM) to the growth medium did not significantly impact the growth of either strain. Furthermore, the presence of either kanamycin or IPTG in these cultures did not impact the max OD_{600} reached by all cultures. Therefore, it was concluded that these experimental conditions (i.e. 200 µg mL^{-1} kanamycin, 1 mM IPTG) were suitable for this study, but it is important to be mindful that direct time-dependent comparisons between *R.leg*-pIND6 and *R. leguminosarum* 3841 should not be made due to slightly delayed growth in the presence of kanamycin. For *R.leg*-pIND6 (grown in the presence of kanamycin), the transition from lag phase to exponential growth occurred between approximately 18 and 24 hours, with cultures reaching stationary phase by around 60 hours. In these cultures, a maximum OD_{600} of 2.0 was reached. By 84 hours of growth, no clear death phase had been observed, with cultures remaining in stationary phase. These growth phase timepoints were subsequently used as a basis for determining the point of induction of *R.leg*-pIND6-*parA1*, *R.leg*-pIND6-*popZ* and *R.leg*-pIND6-*rcy* in later growth experiments.

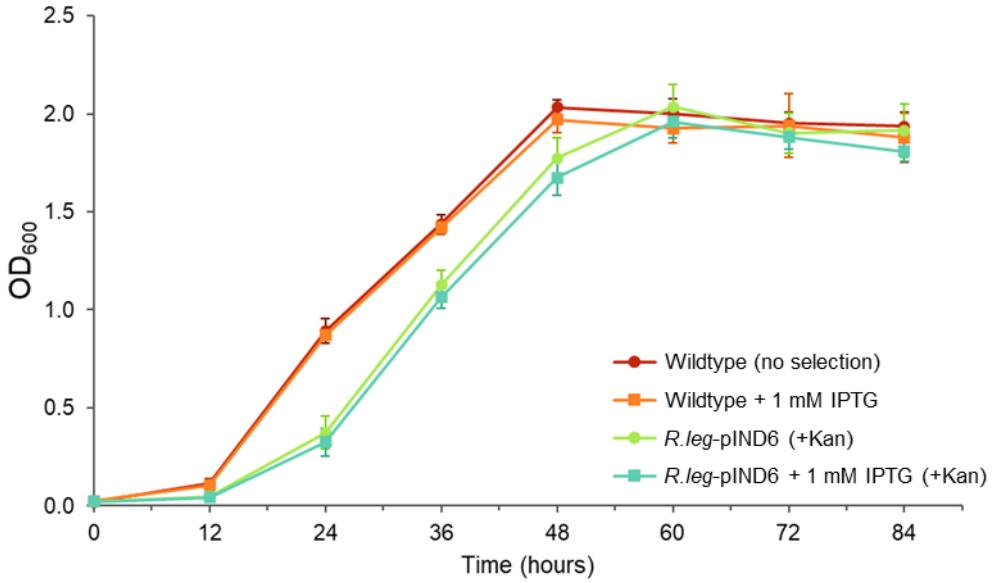


Figure 4.5. Growth curve of *R. leguminosarum* in TY medium in the presence and absence of IPTG. *R. leguminosarum* 3841 (wildtype) and *R.leg*-pIND6 (empty vector control strain) were grown in TY medium in the presence and absence of 1 mM IPTG. Cultures were incubated at 30°C with shaking. To ensure plasmid retention, strains possessing pIND6 were grown in the presence of 200 µg mL^{-1} kanamycin (+Kan). The optical density of samples was measured at a wavelength of 600 nm (OD_{600}). Data points represent means ± SD (n = 4).

4.3.2 Impact of gene overexpression

First, the impact of *parA1*, *popZ* and *rcy* overexpression during early exponential growth was studied. During this growth phase, cultures transition from lag phase, within which cells adjust and prepare for growth, to a period of active cell division (*418*). Therefore, induction of overexpression during this growth phase should help to elucidate the role of these genes during cell division. The preliminary growth curve of *R.leg*-pIND6 (Figure 4.5) indicated that early exponential phase occurred between 18-24 hours. Due to the low resolution of this growth curve the exact time-point of the transition from lag phase to exponential phase could not be determined, therefore 20 hours was chosen as an appropriate timepoint for induction of gene overexpression. At this timepoint, cultures of the overexpression strains were likely to have progressed from lag phase into early exponential phase.

Of all three genes, overexpression of *parA1* during early exponential phase resulted in the most substantial growth defect. In uninduced cultures, *R.leg*-pIND6-*parA1* and *R.leg*-pIND6 exhibited no clear difference in growth, with both strains reaching a maximum OD_{600} of 2.0 (Figure 4.6A). Following induction with 1 mM IPTG, the growth of *R.leg*-pIND6-*parA1* was impacted, with cultures showing a reduction in growth from around 12 hours post-induction (Figure 4.6B). In contrast to *R.leg*-pIND6 and uninduced *R.leg*-pIND6-*parA1* cultures where a max OD_{600} of approximately 2.0 was reached, induced *R.leg*-pIND6-*parA1* cultures reached a max OD_{600} of only 1.1. However, it is important to note that any conclusions drawn from Figure 4.6B are limited by the fact that the *R.leg*-pIND6-*parA1* curve in this figure was generated from only two biological replicates following the omission of the third replicate which displayed abnormal growth. Hence, this curve (and the corresponding uninduced curve) was repeated. Again, uninduced *R.leg*-pIND6 and *R.leg*-pIND6-*parA1* cultures displayed no clear difference in growth (Figure 4.6C), although a slightly lower max OD_{600} of ~1.7 for both strains was observed (compared to an OD_{600} of 2.0 observed in Figure 4.6A). From around 12 hours post-induction, *R.leg*-pIND6-*parA1* cultures once more showed a reduction in growth, while the induction of *R.leg*-pIND6 did not impact the growth of this strain (max OD_{600} of 1.7, reached at 50 hours). By 24 hours post-induction (44 hours) the OD_{600} of these cultures was significantly lower than those of *R.leg*-pIND6 cultures, with *R.leg*-pIND6-*parA1* reaching max OD_{600} of only 0.9 after 68 hours.

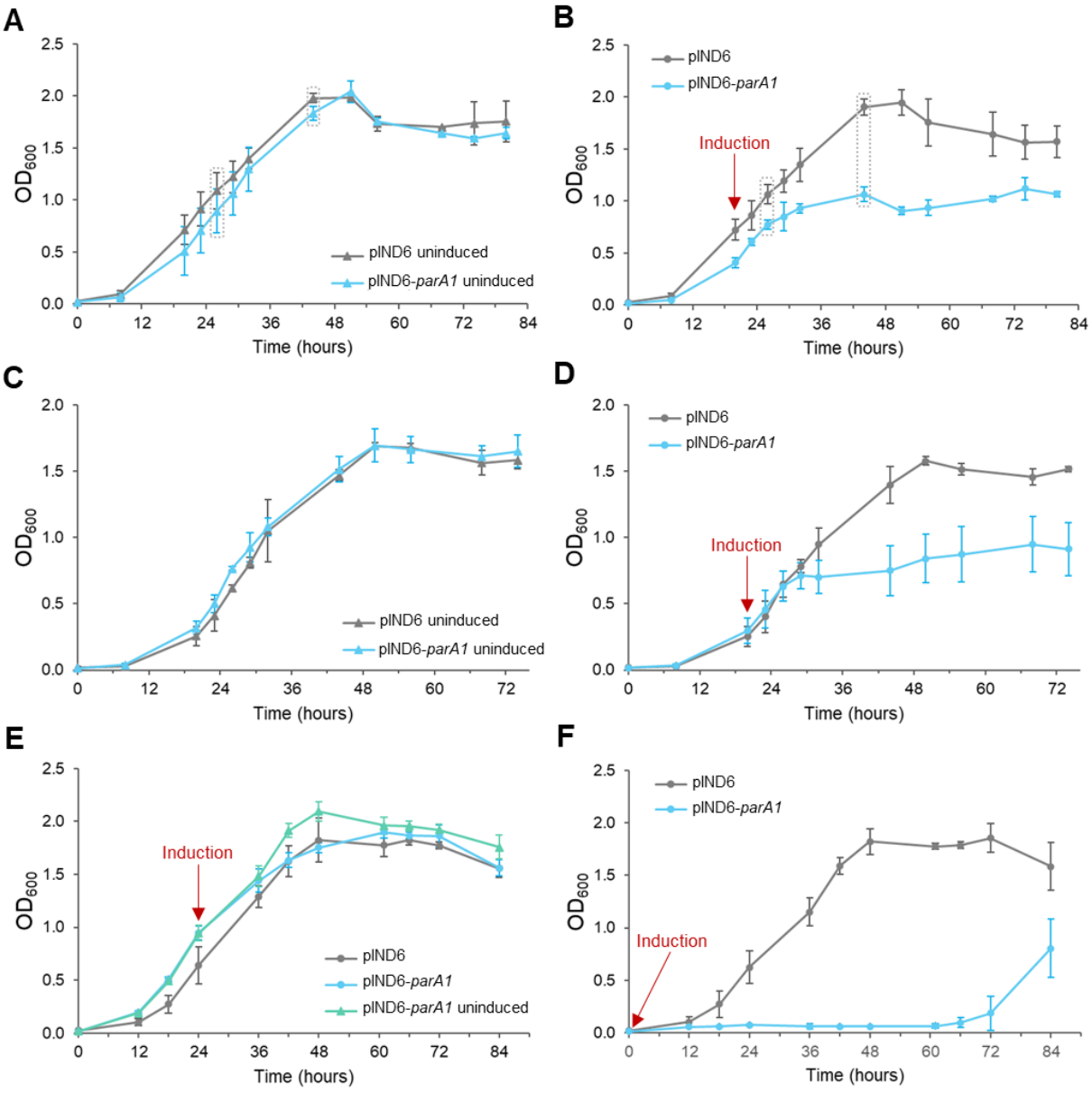


Figure 4.6. Impact of the overexpression of *parA1* on the growth of *R. leguminosarum*. *R. leguminosarum* strains *R.leg*-pIND6 (empty vector control) and *R.leg*-pIND6-*parA1* (*parA1* expression strain) were grown in TY medium (30°C), in the presence of 200 µg mL^{-1} kanamycin. To monitor growth, the optical density of culture samples was measured at a wavelength of 600 nm (OD_{600}). The overexpression of *parA1* was achieved by the induction of cultures with IPTG at a final concentration of 1 mM. Points of induction have been indicated by red arrows. Data points and error bars represent means ± SD ((A) n = 3, (B) n = 2, (C & D) n = 3, (E & F) n = 4). Dashed boxes in Parts A and B indicate timepoints where samples were taken for further analysis (SDS-PAGE and microscopy).

By contrast, the overexpresison of *popZ* and *rcy* at this timepoint (20 hours) did not result in such distinct impacts to the growth of *R. leguminosarum*. Although no clear changes in the perceived rate of growth were observed, the induction of *R.leg*-pIND6-*popZ* may have slightly impacted the OD_{600} of these cultures (Figure 4.7B). Induced *R.leg*.pIND6-*popZ* cultures reached a slightly lower max OD_{600} of 1.7 compared to uninduced *R.leg*-pIND6-*popZ* (1.9, Figure 4.7A) and induced *R.leg*-pIND6 (1.95, Figure 4.7B) cultures. Cultures overexpressing *popZ* appeared to have a lower OD_{600} at both 44 and 51 hours (24 and 31 hours post-induction). However, as uninduced *R.leg*-pIND6-*popZ* cultures also exhibited a lower OD_{600} than uninduced *R.leg*-pIND6 cultures at 44 hours (Figure 4.7A), the impact of *popZ* overexpression on OD_{600} at this timepoint was inconclusive. Similarly, the induction of *R.le*g-pIND6-*rcy* did not result in a noticeable change to the growth of this strain (Figure 4.8A & 4.8B). While the OD_{600} of induced *R.le*g-pIND6-*rcy* cultures at 24 hours post-induction appeared to be slightly lower than that of induced *R.leg*-pIND6 cultures, suggested by non-overlapping standard deviation error bars (Figure 4.8B), the OD_{600} readings at all other timepoints were not distinctly different.

Despite efforts to standardise the generation of these growth curves, the growth of these strains was not always consistent between replicates. As a result, despite a consistent timepoint of induction (20 hours, pressumed to be early exponential phase), the OD_{600} values and respective growth phases at induction actually varied substantially between strains. The mean OD_{600} values at the point of induction of pIND6-*parA1* cultures were 0.40 (Figure 4.6B) and 0.30 (Figure 4.6D). The growth cruves of this strain indicate that as expected, these cultures were within early exponential phase when induction of gene overexpression occurred. Conversely, *R.leg*-pIND6-*popZ* and *R.leg*-pIND6-*rcy* cultures were actually induced at a higher OD_{600}, with a mean OD_{600} at induction of 0.65 (Figure 4.7B) and 0.53 (Figure 4.8B) respectively. Unlike *R.leg*-pIND6-*parA1* cultures which were within early exponential phase, *R.leg*-pIND6-*popZ* and *R.leg*-pIND6-*rcy* cultures were further into exponential phase at the point of induction. Interestingly, the overexpression of *parA1* (which occurred at the lowest OD_{600}) had the clearest negative impact on the growth of *R. leguminosarum*, raising the question whether the severity of overexpression phenotypes is not only dependent on the gene expressed, but also the growth phase of the culture during overexpression.

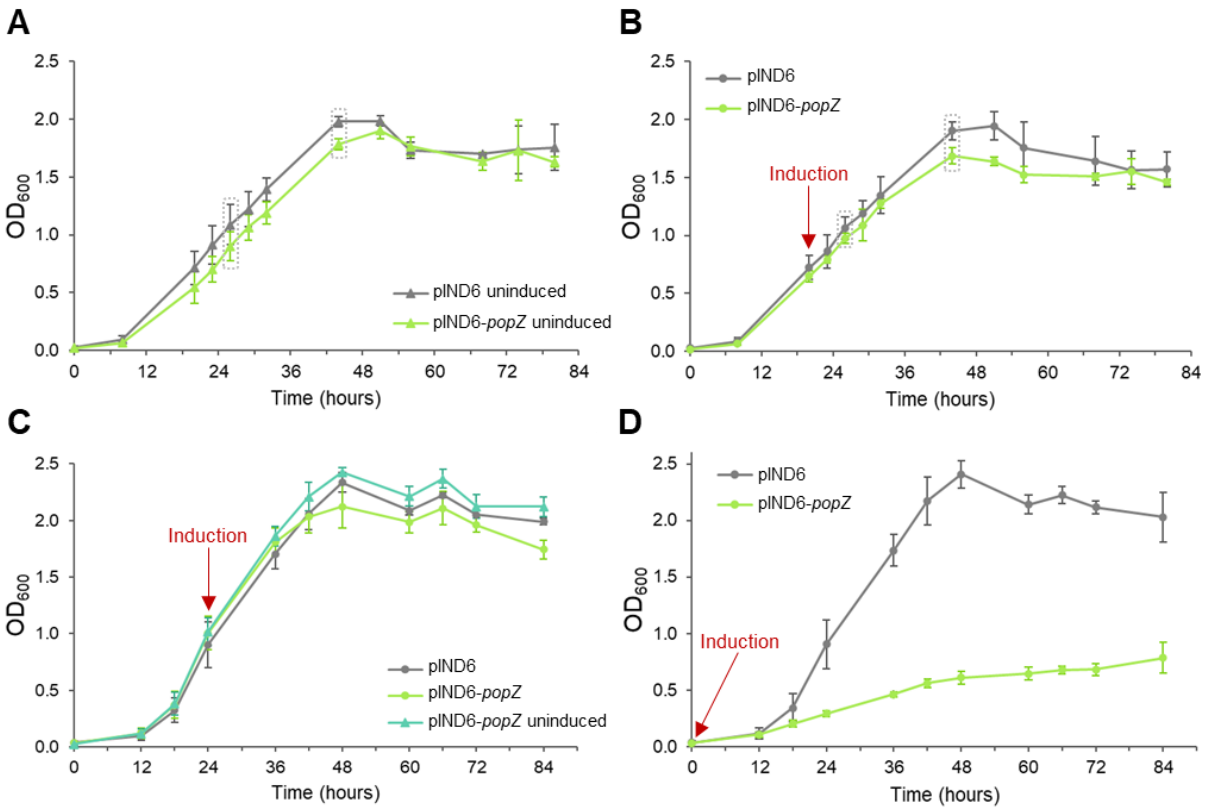


Figure 4.7. Impact of *popZ* overexpression on the growth of *R. leguminosarum*. *R. leguminosarum* strains *R.leg*-pIND6 (empty vector control) and *R.leg*-pIND6-*popZ* (*popZ* expression strain) were grown in TY medium (30°C), in the presence of 200 µg mL^{-1} kanamycin. To monitor growth, the optical density of samples was measured at a wavelength of 600 nm (OD_{600}). Overexpression of *popZ* was achieved by the induction of cultures with 1 mM IPTG. Time points of induction are indicated by red arrows. Data points and error bars represent means ± SD ((A & B) n = 3, (C & D) n = 4). Dashed boxes in Parts A and B indicate timepoints where samples were taken further analysis by SDS-PAGE and microscopy.

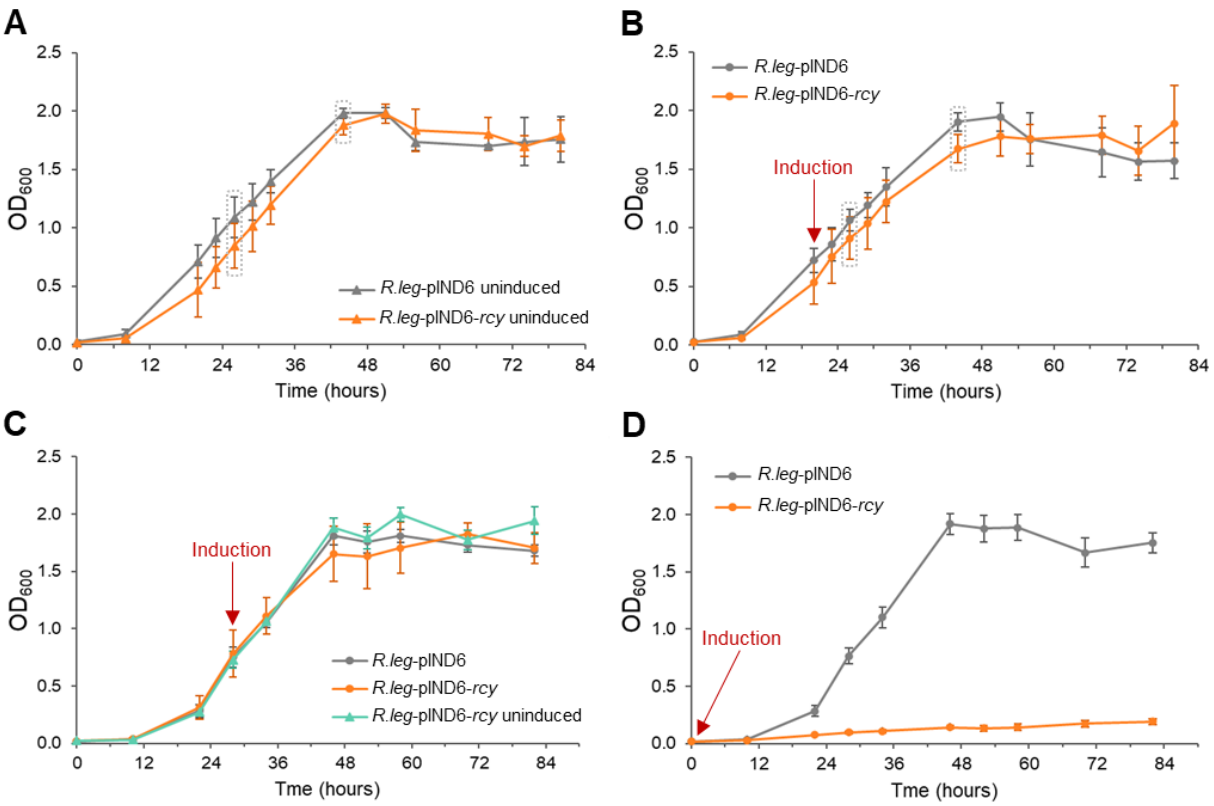


Figure 4.8. Impact of the overexpression of rcy on the growth of *R. leguminosarum*. *R. leguminosarum* strains *R.leg*-pIND6 (empty vector control) and *R.leg*-pIND6-*rcy* (*rcy* expression strain) were grown in TY medium (30°C), in the presence of 200 µg mL^{-1} kanamycin. To monitor growth, the optical density of samples was measured at a wavelength of 600 nm (OD_{600}). Overexpression of *rcy* was achieved by the induction of *R.leg*-pIND6-*rcy* cultures with 1 mM IPTG. Time points of induction are indicated by red arrows. Data points and error bars represent means ± SD ((A & B) n = 3, (C & D) n = 4). Dashed boxes in Parts A and B indicate timepoints where samples were taken for further analysis (SDS-PAGE and microscopy).

To determine whether the growth phase at the point of induction influences the severity of overexpression phenotpyes, gene overexpression during both lag and mid-exponential phase was studied. In contrast to exponential phase, where cells divide at a constant rate, lag phase is a period of limited cell division in which cells adapt to new environmental conditions. Although minimal cell division occurs during this phase, cells are still metabolically active, prioritising the synthesis of cellular components required for growth (*419*, *420*). To study the impact of gene overexpression during lag phase, IPTG was added to the culture media at the point of inoculation (i.e. 0 hrs). To study overexpression during mid exponential phase, cultures were induced between 24-28 hours, aiming for an approximate OD_{600} of between 0.8 and 1.0. As the typical max OD_{600} of (uninduced) *R. leguminosarum* cultures was approximately 2.0, an OD_{600} of 0.8-1.0 (i.e. ½ of max OD_{600}) served as a good approximate for mid-exponential phase.

Overexpression of *parA1* at 24 hours (OD_{600} 0.95) resulted in a noticable reduction in growth between 36 and 60 hours (12-36 hours post-induction) when compared to uninduced *R.leg*-pIND6-*parA1* cultures (Figure 4.6E). Unexpectedly, the *R.leg*-pIND6 control strain cultured appeared to exhibit poorer growth when compared to *R.leg*-pIND6-*parA1* cultures. This was not observed in prior growth curves (Figures 4.6A-D), and was most likely due to an issue with the *R.leg*-pIND6 starter culture. Consequently, comparisons to this strain (*R.leg*-pIND6) cannot justifiably be made for this cruve. In addition to a reduction in growth rate, induction of *R.leg*-pIND6-*parA1* cultures also led to a slight reduction in max OD_{600} (2.1 vs 1.9 of uninduced *R.leg*-pIND6-*parA1* cultures) (Figure 4.6E). The induction of *R.leg*-pIND6-*popZ* cultures at 24 hours (OD_{600} 1.0), shown in Figure 4.7C, also appeared result in a small reduction in max OD_{600} of these cultures. Induced *R.leg*-pIND6-*popZ* cultures reached a max OD_{600} of 2.1, with (induced) *R.leg*-pIND6 and uninduced *R.leg*-pIND6-*popZ* cultures reaching a max OD_{600} of 2.3 and 2.4 respectively. However, overlapping standard deviation bars suggest that this reduction in max OD_{600} is unlikely to actually be significant. Finally, the overexpression of *rcy* during mid-exponential phase did not appear to significantly impact the growth of *R. leguminosarum*. Shown in Figure 4.8C, the induction of *R.leg*-pIND6-*rcy* at 28 hours (OD_{600} 0.72) did not alter the growth of this strain, with no clear differences in the growth of *R.leg*-pIND6, *R.leg*-pIND6-*rcy* and uninduced *R.leg*-pIND6-*rcy* cultures observed.

In contrast, the overexpression of *parA1*, *popZ* and *rcy* during lag phase all greatly inhibited the growth of *R. leguminosarum*. Growth curves generated following the overexpression of these

genes have been shown in Figures 4.6F (*parA1*), 4.7D (*popZ*) and 4.8D (*rcy*). When cultured in the presence of IPTG, *R.leg*-pIND6-*parA1* cultures exhibited little to no growth for the first 60 hours (Figure 4.6F). However, after 60 hours these cultures began to display the beginnings of a 'typical' growth curve. The OD_{600} of these cultures gradually increased between 60 and 72 hours (from 0.07 to 0.19), suggestive of a lag phase, followed by a sharp increase in OD_{600} (from 0.19 to 0.80) between 72 and 84 hours, suggestive of an exponential growth phase. The overexpression of *popZ* during lag phase resulted in reduced but steady growth of *R. leguminosaurm* (Figure 4.7D). *R.leg*-pIND6-*popZ* cultures displayed a gradual increase of OD_{600} over the observed growth period, reaching a final OD_{600} of approximately 0.8 by 84 hours. In comparison, *R.leg*-pIND6 cultures reached a max OD_{600} of 2.41 (Figure 4.7D), while uninduced *R.leg*-pIND6-*popZ* cultures reached a max OD_{600} of 2.43 (Figure 4.7C). As observed with the overxpression of both *parA1* and *popZ*, the induction of *rcy* overexpression during lag phase also significantly impeded growth. *R.leg*-pIND6-*rcy* cultures grown in the presence of 1 mM IPTG displayed very little growth, reaching a final OD_{600} of only 0.2 by 82 hours (Figure 4.8D).

A summary of the effects of gene overexpression on population growth described in this section has been shown in Table 4.1. Overall, the growth of *R. leguminosarum* was most strongly impacted by gene overexpression during lag phase, with induction of gene expression during this growth phase inhibiting growth. Of all three genes, the overexpression of *parA1* was most detrimental to growth (Figure 4.6), independent of the growth phase of the culture when overexpression occurred. A small caveat of this was the overexpression of *parA1* during lag phase. Although next to no growth was observed for the first 60 hours, *R.leg*-pIND6-*parA1* cultures appeared to display 'normal' growth towards the very end of the growth period observed (Figure 4.6F). It is also important to note that due to unexpected differences in the growth of cultures between independent growth curve experiments, only *R.leg*-pIND6-*parA1* cultures were induced during early exponential phase. Differences between the impact of overexpression of *popZ* and *rcy* were harder to distinguish. Generally, the overexpression of *popZ* appeared to be the least detrimental to the growth of *R. leguminosarum* (Figure 4.7). The induction of *R.leg*-pIND6-*popZ* cultures during lag phase inhibited growth the least of all three strains, while induction during exponential phase had only minimal impact on the growth of these cultures. Interestingly, despite having no observable impact on OD_{600} or growth when overexpressed during exponential phase, the overexpression of *rcy* during lag phase almost completely inhibited the growth of *R. leguminosarum* (Figure 4.8). In conclusion, these findings demonstrate that the effect of misregulation of these genes through their overexpression are

not only gene-specific, but the severity of the given impact is also influenced by the OD_{600} of the culture (and ultimately growth phase) at the point of gene overexpression. The observation of growth-related phenotypes, such as delayed growth and reduction in max OD_{600} following overexpression support the involvement of these genes in controlling and coordinating the growth of *R. leguminosarum*.

Table 4.1. Summary of the impacts of *parA1*, *popZ* and *rcy* overexpression on *R. leguminosarum* cultures. The impact of gene overexpression was determined by the generation of growth curves of the gene expression strains *R.leg*-pIND6-*parA1*, *R.leg*-pIND6-*popZ* and *R.leg*-pIND6-*rcy*, and the empty vector strain *R.leg*-pIND6. This table summarises the changes in each growth curve, and the OD_{600} and corresponding growth phase of each culture at the point of induction of gene overexpression. Comparisons to 'max OD_{600}' refer to the max OD_{600} of uninduced cultures and the control strain *R.leg*-pIND6.

Gene	OD_{600}	Growth phase	Impact(s) of gene overexpression
parA1	0.01	Lag	No growth between 0-60 hours 'Normal' growth between 60-84 hours
	0.30-0.40	Early exp.	Reduction in growth rate (~12 hours post-induction) ~50% reduction in max OD_{600}
	0.95	Mid exp.	Slight reduction in growth rate (12-36 hours post-induction) ~10% reduction in max OD_{600}
popZ	0.03	Lag	Minimal growth Final OD_{600} of 0.8 by 84 hours (~30% of max OD_{600} of uninduced cultures
	0.65	Mid exp.	~10% reduction of max OD_{600}
	1.0	Mid exp.	No clear impact
rcy	0.02	Lag	Severely inhibited growth Final OD_{600} of 0.2 by 82 hours (~10% of max OD_{600} of uninduced cultures
	0.53	Mid exp.	No clear impact
	0.72	Mid exp.	No clear impact

4.4 Confirmation of protein expression

To determine whether gene overexpression from pIND6-derived expression vectors corresponded to protein overexpression, *R. leguminosarum* total protein extracts were analysed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Section 2.9). Samples were isolated from *R.leg*-pIND6-*parA1*, *R.leg*-pIND6-*popZ*, *R.leg*-pIND6-*rcy*, and *R.leg*-pIND6 cultures at 24 hours post-induction, from cultures induced at 20 hours (Figures 4.6B, 4.7B and 4.8B), with control samples taken from uninduced cultures at the equivalent timepoint (Figures 4.6A, 4.7A and 4.8A). Samples taken 24 hours post-induction were chosen for the initial protein analysis as growth-related phenotypes were typically most noticeable at this time point. The preparation and standardisation of samples for SDS-PAGE analysis is described in Section 2.9.1. In short, cells from 1 mL of culture were pelleted and resuspended in a 10 mM Tris, 5 mM $MgCl_2$ buffer. Samples were then standardised to ensure equal loading of protein into each well, before boiling in sample buffer to lyse cells. Polyacrylamide gels for SDS-PAGE were prepared as described in Section 2.9.3. To achieve the best resolution of the proteins of interest, *R.leg*-pIND6-*parA1* and *R.leg*-pIND6-*popZ* samples were analysed on 15% polyacrylamide gels (His-ParA1: 31.1 kDa, His-PopZ: 33.7 kDa), while *R.leg*-pIND6-*rcy* samples were analysed on 10% polyacrylamide gels (His-Rcy: 255.6 kDa). As all proteins expressed from the pIND6 expression vectors were His-tagged, all protein masses stated throughout this section refer to the His-tagged versions of these proteins.

Coomassie staining of acrylamide gels following SDS-PAGE of whole-cell protein samples, shown in Figure 4.9, identified a distinct protein band at approximately 255 kDa, consistent with His-Rcy (255.6 kDa) in the induced *R.leg*-pIND6-*rcy* sample (Figure 4.9B). Faint bands of this size in the uninduced *R.leg*-pIND6-*rcy* sample and both (induced and uninduced) *R.leg*-pIND6 samples clearly demonstrate the overexpression of this protein following induction. Although Coomassie staining highlighted clear overexpression of His-Rcy in induced *R.leg*-pIND6-*rcy* cultures, no distinct differences in banding patterns were observed between *R.leg*-pIND6-*parA1, R.leg*-pIND6-*popZ* and *R.leg*-pIND6 samples (Figure 4.9A). Many other strong protein bands similar in size to His-ParA1 (31.1 kDa) and His-PopZ (33.7 kDa) were present in these samples, making the identification of these proteins challenging.

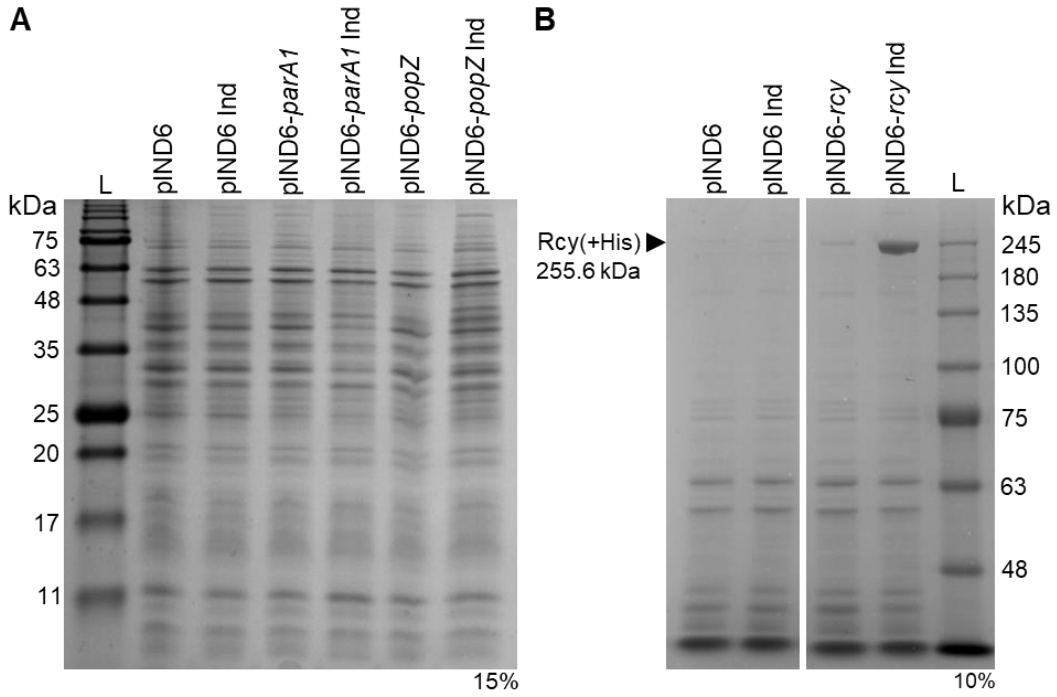


Figure 4.9. SDS-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of whole-cell *R. leguminosarum* protein samples following overexpression of pIND6-*parA1*, pIND6-*popZ* and pIND6-*rcy*. Whole-cell protein samples, isolated from induced (Ind) and uninduced *R. leguminosarum* cultures 24 hours post-induction, were analysed by SDS-PAGE. This figure shows InstantBlue™ Coomassie staining of SDS gels. (A) 15% polyacrylamide gel for the visualisation of His-ParA1 (31.1 kDa) and His-PopZ (33.7 kDa). (B) 10% polyacrylamide gel for the visualisation of His-Rcy (255.6 kDa). Lane legends denote the gene expression vectors carried by each *R. leguminosarum* strain. Lane L; BLUeye Prestained Protein Ladder.

Since Coomassie staining of protein samples following SDS-PAGE was not sufficient to identify the expression of His-ParA1 and His-PopZ in these samples (Figure 4.9A), western blotting was used instead (Section 2.9.4). Following SDS-PAGE, separated proteins were transferred onto a polyvinylidene difluoride (PVDF) membrane. Proteins were then probed with an anti-His_6-Peroxidase monoclonal antibody (Sigma-Aldrich) to detect the His-tagged ParA1 and PopZ proteins. Figure 4.10 shows the western blot of these samples, and an accompanying Ponceau S stain of the membrane. As observed with Coomassie staining, Ponceau S staining of the membrane did not reveal any noticeable differences in protein expression between samples. In addition to *R.leg*-pIND6-*parA1, R.leg*-pIND6-*popZ* and *R.leg*-pIND6 control samples, a sample of wildtype *R. leguminosarum* 3841 (isolated from a stationary-phase culture) was also analysed. Ponceau S staining indicated that less protein was loaded onto the gel for the wildtype sample, most likely as a result of poor standardisation of this sample.

Although Ponceau S staining did not identify differences in the expression of ParA1 and PopZ between induced and uninduced *R.leg*-pIND6-*parA1* and *R.leg*-pIND6-*popZ* samples, western blotting identified a clear signal in the induced culture samples (Figure 4.10). Unexpectedly, the western signal for His-PopZ was much stronger than that of His-ParA1, potentially suggesting a higher abundance of His-PopZ in these samples (compared to His-ParA1 in *R.leg*-pIND6-*parA1*). In this blot, His-PopZ appeared to be slightly larger than expected (appearing approximately between 5 and 15 kDa larger). While this may have been the result of incomplete denaturation or post-translational modifications, it was likely due to the biochemical properties of this protein. His-PopZ has a pI (isoelectric point) of 4.75, and it is well-documented that acidic proteins (i.e. those with a higher abundance of acidic residues) have slower mobility during SDS-PAGE (*421*, *422*)**.** As a result, these proteins can appear larger than their true size following electrophoresis.

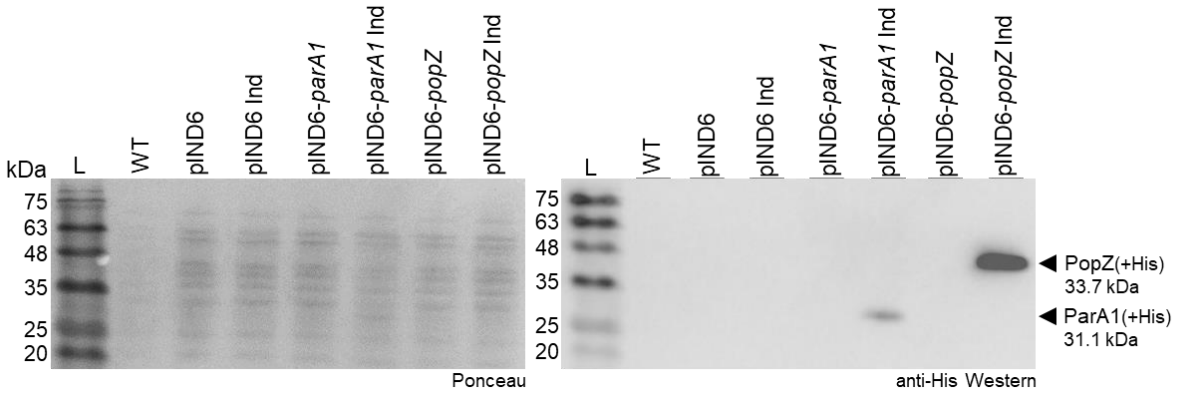


Figure 4.10. Western blot analysis of whole-cell *R. leguminosarum* protein samples following overexpression of *parA1* and *popZ*. Whole-cell protein samples, isolated from induced (Ind) and uninduced *R. leguminosarum* cultures 24 hours post-induction, were analysed by western blot analysis. PVDF membranes were sained with Ponceau S (left), destained, then probed with an anti-His_6-Peroxidase monoclonal antibody (right). Lane labels denote the gene expression vectors possessed by each strain. Lane L; BLUeye Prestained Protein Ladder. WT; wildtype *R. leguminosarum* 3841.

Following the confirmation of ParA1, PopZ and Rcy overexpression at 24 hours post-induction, protein expression at this time point was compared to that observed 6 hours post-induction. *R. leguminosarum* is categorised as a "fast growing" species of rhizobia, broadly defined as those with a mean generation time (MGT) of less than 6 hours, with "slow growing" species having a MGT of greater than 6 hours (*423*, *424*). Therefore, by 6 hours post-induction, most *R. leguminosarum* cells were likely to have undergone one round of division in the presence of the inducer (IPTG). Culture samples were taken from induced *R.leg*-pIND6-*rcy* cultures at 0, 6 and 24 hours post-induction, with control samples taken from uninduced *R.leg*-pIND6-*rcy* cultures at the equivalent timepoints. As before, these samples were isolated from cultures induced 20 hours into their growth i.e. exponential phase (Figure 4.8). ParA1 and PopZ expression was not assessed here as these proteins could only be identified following Western blot analysis (Figure 4.10). Again, whole-cell total protein samples were analysed by SDS-PAGE, then Coomassie stained to visualise protein. This gel image, shown in Figure 4.11, confirmed once more that overexpression of His-Rcy at both 6 and 24 hours post-induction was clearly observed, with little difference in band intensity, suggesting a similar abundance of protein at both timepoints.

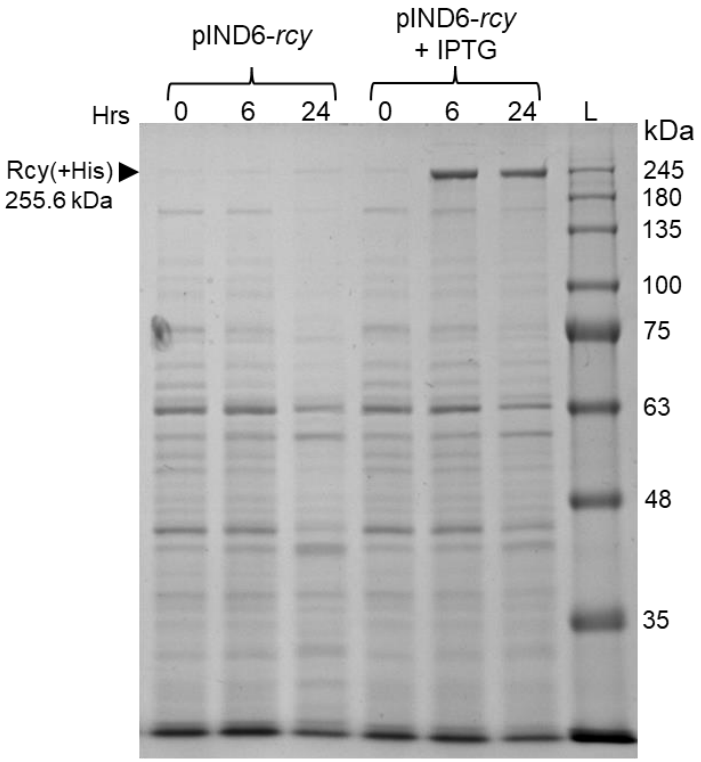


Figure 4.11. SDS-PAGE analysis of *R. leguminosarum* total protein samples following induction of *rcy* overexpression. Culture samples were taken from induced (+ IPTG) *R.leg*-pIND6-*rcy* cultures at 0, 6 and 24 hours post induction. Control samples were taken from uninduced *R.leg*-pIND6-*rcy* cultures at the equivalent timepoints. Whole-cell protein samples were analysed by SDS-PAGE (15% polyacrylamide gel), then gels stained with InstantBlue™ Coomassie stain. Lane L; BLUeye Prestained Protein Ladder. Numbers above lanes indicate hours (hrs) post-induction.

4.5 Impact of ParA1, PopZ and Rcy overexpression on *R. leguminosarum* cell morphology

4.5.1 Morphological phenotypes following protein overexpression

While the growth-related impacts of protein overexpression described in Section 4.3.2 shed light on the importance of controlled regulation of *parA1*, *popZ* and *rcy* (and their encoded proteins) during growth, these observations alone are not sufficient to characterise the exact role of these proteins during cell growth and division. Therefore, morphological phenotypes arising from ParA1, PopZ and Rcy overexpression were studied to characterise the role of these proteins *in vivo* and determine the mechanisms underpinning the growth characteristics described in Section 4.3.2. Samples isolated from cultures at 6 and 24 hours post-induction (Figures 4.6, 4.7 and 4.8, parts A & B, induced at 20 hours) were analysed by phase-contrast microscopy to study any morphological changes following protein overexpression. To minimise physiological stresses to cells such as dehydration, which can arise from traditional fixing methods (e.g. heat or methanol), culture samples were mounted on top of agarose pads (Section 2.10.1). This enabled the 'fixing' of motile cell samples while preserving cell morphology.

Phase-contrast microscopy identified distinct morphological defects following overexpression of ParA1, PopZ and Rcy (Figure 4.12). By 6 hours post-induction (or equivalent timepoint), cells within all uninduced cultures and the induced control *R.leg*-pIND6 culture displayed rod-shaped morphology, with no distinct differences in cell size. Induced *R.leg*-pIND6-*popZ* cells exhibited very little difference in morphology to uninduced *R.leg*-pIND6-*popZ* and *R.leg*-pIND6 cells, maintaining a rod morphology. Cells overexpressing PopZ appeared slightly shorter than the uninduced control samples but remained similar in width. While the overexpression of ParA1 did not alter cell morphology, with cells exhibiting a rod-shaped morphology, an increased prevalence of what appeared to be morphologically asymmetric cell division was observed (Figure 4.12, hollow arrowheads). Rcy overexpression resulted in the most distinct change to *R. leguminosarum* morphology. At 6 hours post-induction, the majority of induced *R.leg*-pIND6-*rcy* cells appeared to be enlarged. Although similar in length, these cells appeared broader than their uninduced counterparts and the *R.leg*-pIND6 control cells. Furthermore, some cells also exhibited an unusual, 'budding-like' morphology (Figure 4.12, asterisks).

At 24 hours post-induction, induced *R.leg*-pIND6-*parA1* samples continued to display evidence of asymmetric cell division, yielding daughter cells of notably different size. Typically, cells overexpressing ParA1 retained their normal rod-shaped morphology. However, ParA1 overexpression occasionally resulted in branched Y-shaped cells (Figure 4.12, solid arrowhead). Although not shown in Figure 4.12, V-shaped branched cells were also observed within these samples. While Y- and V-shaped branched cells were markedly more prominent in induced *R.leg*-pIND6-*parA1* samples, these cells only made up approximately 1 in every 100 cells. Hence, although cell branching was observed, it was not considered a predominant phenotype of ParA1 overexpression within *R. leguminosarum*.

Overexpression of PopZ resulted in an increased prevalence of small, spherical cellular bodies within cultures 24 hours post-induction (Figure 4.12, arrows). Although these cells appeared shorter than uninduced *R.leg*-pIND6-*popZ* cells at 6 hours post-induction, cell length between both induced and uninduced *R.leg*-pIND6-*popZ* cells at 24 hours did not appear to be so distinctly different. This 'shortening' of uninduced cells between 6 and 24 hours was observed (to varying extents) across all strains and was most likely due to the age of these cultures. It has been demonstrated previously that rod-shaped bacterial regulate their cell morphology dependent on their growth rate. During stationary phase, where nutrients are limited and growth rate slows, cells within a culture are typically smaller (*425*). At 24 hours post-induction (44 hours total growth), the cultures observed here were all approaching (or already in) stationary phase (Figures 4.6, 4.7 and 4.8). By contrast, cells observed at 6 hours post-induction (26 hours growth) were taken from cultures within exponential phase, where growth rate is fastest and cells are typically at their largest (*425*). Given this, comparisons of cell morphology (particularly cell size) were made only between cells isolated from cultures within the same growth phase.

Rcy overexpression gave rise to by far the most distinct morphological change. At 24 hours post-induction, most induced *R.leg*-pIND6-*rcy* cells displayed a phenotype which mimicked 'swelling' of these cells. In contrast to the rod-shaped cells observed in within the uninduced *R.leg*-pIND6-*rcy* culture sample, the overexpression of Rcy resulted in the formation of enlarged coccoid cells. In addition, some cells also displayed the same budding-like morphology as observed at 6 hours post-induction (Figure 4.12, asterisks). As opposed to true budding, it is most likely that these cells were in fact dividing cells where each of the two developing daughter cells displayed an irregular or spherical morphology and a variation in size, giving the appearance of budding.

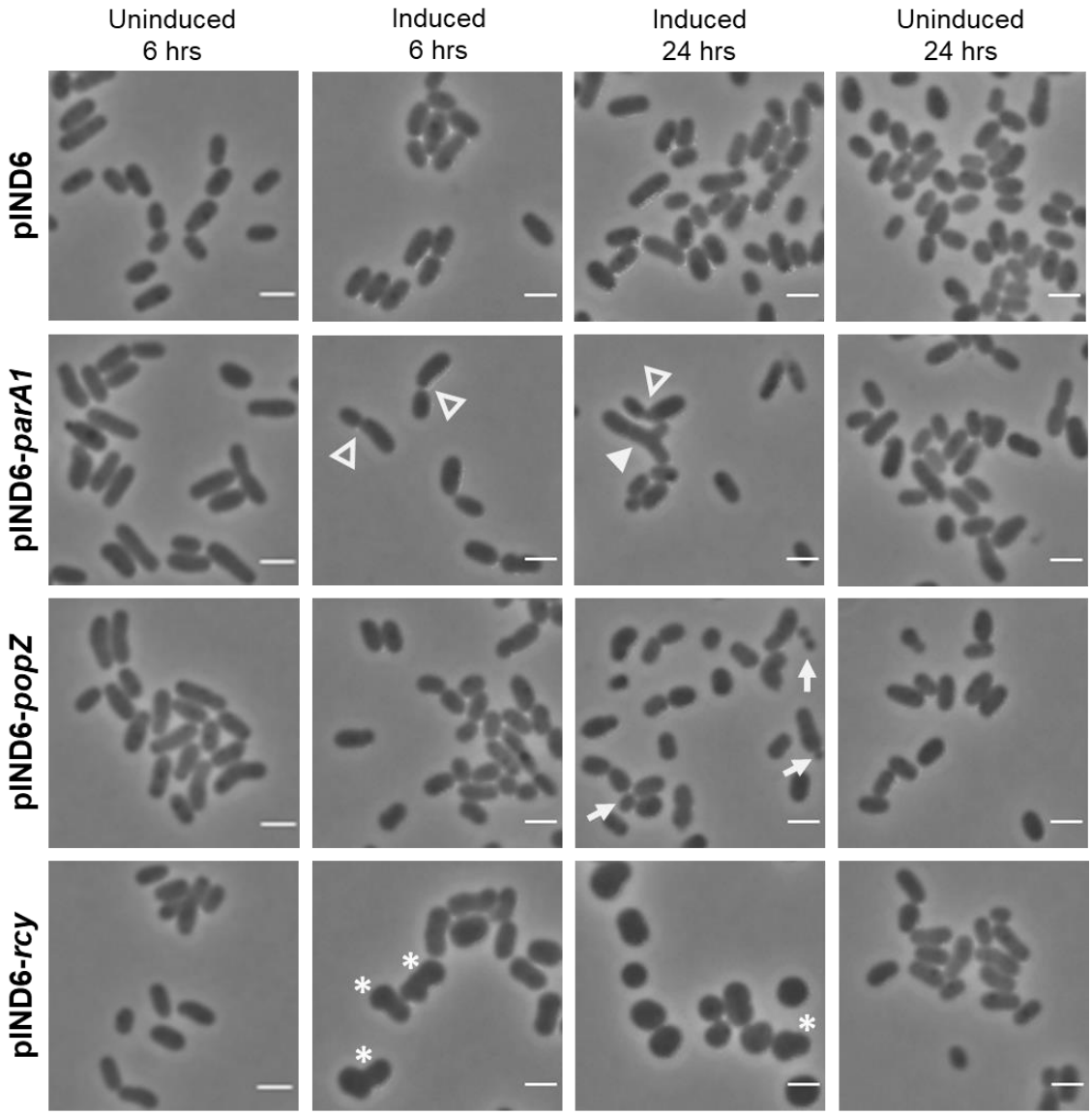


Figure 4.12. Morphological phenotypes of *R. leguminosarum* overexpressing ParA1, PopZ and Rcy. The *R. leguminosarum* strains *R.leg*-pIND6-*parA1*, *R.leg*-pIND6-*popZ*, *R.leg*-pIND6-*rcy* and *R.leg*-pIND6 (control) were cultured in TY medium, with and without induction (1 mM IPTG). Samples were isolated from cultures at 6 and 24 hours post-induction (or equivalent timepoints for uninduced cultures), and viewed using phase-contrast microscopy. Phenotypes observed include asymmetric division (hollow arrowheads), branching (solid arrowhead), small cellular bodies (arrows) and budding-like morphologies (asterisks). Scale bars represent 2 µm.

4.5.2 Quantitative analysis of overexpression phenotypes

The cell morphology images shown in Figure 4.12 display only a representative snapshot of the cells observed within these samples. To obtain a better understanding of the morphologies of all cells within these populations, whole field of view phase-contrast microscopy images were analysed using the image processing software ImageJ. This software was used to measure a range of parameters including cell area, perimeter, length, and width for all cells within these samples. In this study, Feret diameter (also referred to as calliper diameter) was used as a measure of cell length and width (covered later in this section). The image analysis workflow used to obtain these measurements has been outlined in Section 2.10.3, with examples of the images used for analysis displayed in Appendix A.4. As part of this analysis, cells smaller than 0.3 x 0.3 µm (Feret diameter) were omitted from datasets. 'Cells' of this size were identified in all samples, and corresponded to either debris or falsely-identified cells during the image processing workflow.

As described in Section 4.5.1, the overexpression of ParA1, PopZ and Rcy resulted in various morphological changes to *R. leguminosarum*. By 24 hours post-induction, two distinct ParA1 overexpression phenotypes were observed; asymmetric division, and occasional branching of cells. The overexpression of PopZ appeared to increase frequency of small cellular bodies in these samples, while overexpression of Rcy resulted in enlarged, spherical cells. To determine whether the observed differences in cell size were in fact genuine overexpression phenotypes representative of the whole population, measurement data from 580 cells per overexpression culture were analysed. Even though analysis of protein expression did not clearly identify elevated expression in uninduced cultures (Section 4.4), it is still possible that leaky expression of genes from the pIND6-derived plasmids could result in low-level overexpression of these proteins (*355*). Even at a low level, protein overexpression could result in slight morphological changes of cells within uninduced cultures. Therefore, all cell morphology data collected from the induced expression strains (*R.leg*-pIND6-*parA1*, *R.leg*-pIND6-*popZ* and *R.leg*-pIND6-*rcy*) were compared to the *R.leg*-pIND6 control strain, and not the corresponding uninduced cultures. For the remainder of this section, the terms *R.leg*-pIND6, *R.leg*-pIND6-*parA1*, *R.leg*-pIND6-*popZ* and *R.leg*-pIND6-*rcy* refer to samples collected from induced cultures.

The first step in analysis of the overexpression phenotypes observed in Figure 4.12 was the plotting of cell area histograms to visualise the overall size distribution of cells within these populations. Within *R.leg*-pIND6 cultures, approximately half of cells analysed had an area of between 0.64 and 0.88 μm^2, with over 90% of the cells within the sampled population between 0.48 and 1.36 μm^2 in size (Figure 4.13A). Cells observed in cultures overexpressing ParA1 exhibited a more even distribution of cell area (Figure 4.13B). Unlike *R.leg*-pIND6, where around 50% of cells had an area of between 0.64 and 0.88 μm^2, only 20% of *R.leg*-pIND6-*parA1* cells were this size. A higher frequency of smaller cells (≤ 0.64 μm^2) was observed, likely reflecting an increase in smaller cells arising from asymmetric cell division (Figure 4.12), along with a small increase in cells > 1.52 μm^2. Cross-referencing of area data to microscopy images confirmed that these larger cells corresponded to the branched Y- and V-shaped cells observed within these samples. Analysis of cells isolated from *R.leg*-pIND6-*popZ* cultures, shown in Figure 4.13C, highlighted an increased frequency of cells with an area of less than 0.5 μm^2 within this population, corresponding to the small cellular bodies observed in Figure 4.12. In contrast to *R.leg*-pIND6 samples, where only around 3% of cells were smaller than 0.5 μm^2, 16% of cells within the *R.leg*-pIND6-*popZ* culture were < 0.5 μm^2. No distinct difference in the frequency of cells > 1.2 μm^2 was observed, suggesting that the overexpression of PopZ did not result in an increase in cell area. Similar to ParA1, the overexpression of Rcy also resulted in a more evenly-spread distribution of cell area within the *R.leg*-pIND6-*rcy* population (Figure 4.13D). However, in contrast to ParA1 overexpression which gave rise to an increase in smaller cells within this population (Figure 4.13B), Rcy overexpression resulted in a distinct increase in the frequency of much larger cells (Figure 4.13D). While only approximately 3% of *R.leg*-pIND6 cells were larger than 1.36 μm^2, 42% of *R.leg*-pIND6-*rcy* cells had an area of > 1.36 μm^2, suggesting a strong Rcy overexpression phenotype.

The area of cells within the overexpression strain samples were then compared to those of the *R.leg*-pIND6 control strain, visualised with box-and-whisker plots (Figure 4.13E). Additional details of all statistical analyses performed in this thesis can be found in Section 2.11. As Kolmogorov-Smirnov (K-S) tests demonstrated that none of the four datasets were normally distributed ($p < 0.01$), differences in the datasets were determined using Mann–Whitney U (MWU) tests. This statistical analysis suggested that the observed differences in cell areas within *R.leg*-pIND6-*parA1* and *R.leg*-pIND6-*rcy* cultures were statistically distinct from the areas of *R.leg*-pIND6 control cells ($p < 0.01$). However, despite an apparent increase in smaller cellular bodies within *R.leg*-pIND6-*popZ* samples, overall differences in cell area between this population

and *R.leg*-pIND6 cells was not statistically significant ($p = 0.13$). In conjunction with comparisons of the distribution of these datasets with Mann–Whitney U tests, median area values were also compared with independent samples Median tests. As expected, the median cell areas for *R.leg*-pIND6-*popZ* and *R.leg*-pIND6 sample (0.82 and 0.83 μm^2 respectively) were not significantly different ($p = 0.597$). The median *R.leg*-pIND6-*parA1* cell area (0.70 μm^2) was significantly smaller than that of *R.leg*-pIND6 ($p < 0.01$), with the *R.leg*-pIND6-*rcy* median cell area (1.22 μm^2) significantly larger than that of *R.leg*-pIND6 ($p < 0.01$).

In addition to comparisons of cell areas, the perimeters of these same cells were also analysed (Figure 4.13F). K-S tests showed that perimeter data for *R.leg*-pIND6, *R.leg*-pIND6-*parA1* and *R.leg*-pIND6-*popZ* cultures were not normally distributed ($p < 0.01$), whereas *R.leg*-pIND6-*rcy* data was normally distributed ($p = 0.200$). Therefore, Mann-Whitney U and Medians tests were again employed to compare these datasets. Typically, differences in the perimeters of cells overexpressing these proteins were consistent with observations made for the areas of these cells. As observed with the cell area data, the perimeters of cells within *R.leg*-pIND6-*parA1* cultures were significantly more varied than those of *R.leg*-pIND6 cells ($p < 0.01$), with the median *R.leg*-pIND6-*parA1* cell perimeter of 3.40 µm significantly smaller than that of *R.leg*-pIND6 samples (3.57 µm) ($p < 0.05$). The perimeters of *R.leg*-pIND6-*rcy* cells were larger ($p < 0.01$), as clearly visualised by the box-and-whisker plot of this strain (Figure 4.13F). Unsurprisingly, the median perimeter of this strain (4.27 µm) was significantly larger than that of the *R.leg*-pIND6 culture ($p < 0.01$). Mean cell perimeter of this strain was 4.29 µm. This mean perimeter value is likely a slightly more accurate reflection of this population given the normal distribution of this dataset. Based on the very small difference in mean and median values (difference of 0.02 µm), and large difference between these values and the median of the *R.leg*-pIND6 data (*R.leg*-pIND6 median: 3.57 µm), it is possible to conclude that this mean perimeter value is also statistically larger than the median *R.leg*-pIND6 cell perimeter.

The morphological impact of PopZ overexpression within *R. leguminosarum* was less conclusive. In contrast to the area of cells within this population, which did not differ from the *R.leg*-pIND6 population ($p = 0.13$), the perimeters of *R.leg*-pIND6-*popZ* cells were significantly smaller than those of pIND6 ($p < 0.01$). However, comparison of median values suggested that the median *R.leg*-pIND6-*popZ* cell perimeter of 3.50 µm was unlikely to be significantly smaller than that of *R.leg*-pIND6 cells (3.57 µm) ($p = 0.06$).

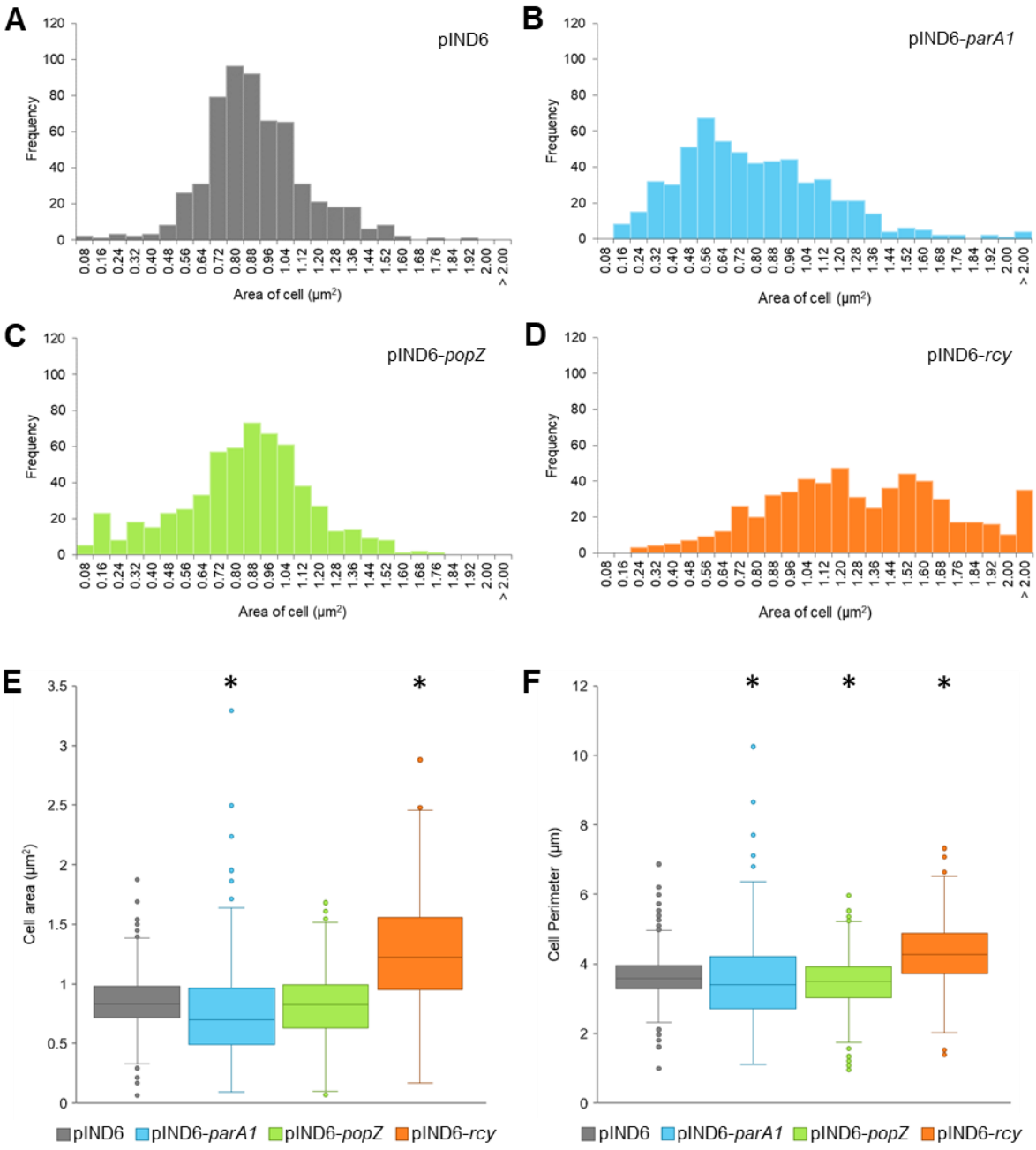


Figure 4.13. Cell area and perimeter phenotypes of ParA1, PopZ and Rcy overexpression *R. leguminosarum* strains. Computational analysis of phase-contrast microscopy images was used to obtain the cross-sectional cell area and perimeter measurements of cells within liquid cultures of the *R. leguminosarum* strains *R.leg*-pIND6 (control), *R.leg*-pIND6-*parA1*, *R.leg*-pIND6-*popZ* and *R.leg*-pIND6-*rcy*. Cell samples were isolated from cultures at 24 hours post-induction of protein overexpression. Data from 580 cells per strain are shown in these figures. The distribution of cell areas within these populations was visualised with histograms (A – D). Optimal bin size of histograms was determined using the *R.leg*-pIND6 dataset. All bin labels represent the upper limit of each bin. Differences in the distribution of the cell area (E) and their corresponding perimeters (F) were then compared. Asterisks indicate significant differences with *R.leg*-pIND6 cells (Mann–Whitney U tests, $p < 0.01$).

Following comparisons of overall cell size (area and perimeter), cell length and width measurements were analysed to obtain a better understanding of how the overexpression of ParA1, PopZ and Rcy alters the shape of *R. leguminosarum* cells. In this study maximum Feret (F_{max}) diameter, defined as the greatest distance between two parallel tangents (Figure 4.14A) was selected as the measure of cell length, with the minimum Feret (F_{min}) diameter representing cell width. To study the proportions of these cells, length (F_{max}) and width (F_{min}) measurements for each cell were plotted against each as scatter plots. Figure 4.14B shows a combined scatter plot of all four *R. leguminosarum* strains, with each dataset then plotted individually in Figures 4.14C-F. On first inspection, the combined plot (Figure 4.14B) suggests that the overexpression of these proteins had a greater impact on the width of *R. leguminosarum* cells than their length. The *R.leg*-pIND6-*parA1* culture sample appeared to contain more cells with a smaller width (F_{min}) than the *R.leg*-pIND6 control sample. A few very large outliers were also observed within this sample, corresponding to the irregularly-shaped branched cells observed previously. The *R.leg*-pIND6-*popZ* sample contained a few cells smaller in both length and width, but generally cells within this sample appeared to be similar in size to *R.leg*-PIND6 cells. Most strikingly, the overexpression of Rcy resulted in a much higher proportion of wider cells within this population. With the exception of the few large *R.leg*-pIND6-*parA1* cells, the maximium cells lengths observed in these samples did not appear to differ much between all four strains.

Most *R.leg*-pIND6 cell parameters clustered at an F_{min} of 0.7-0.9 µm and F_{max} of 1.2-1.5 µm (Figure 4.14C). Although the *R.leg*-pIND6-*popZ* sample appeared to contain a greater proportion of cells < 1 x 0.5 µm, many of the *R.leg*-pIND6-*popZ* cells also appeared to cluster within the same range of F_{min} and F_{max} values as the *R.leg*-pIND6 sample (Figure 4.14E). *R.leg*-pIND6-*parA1* cells exhibited less clustering, suggesting a more even distribution of cell sizes within this population (Figure 4.14D). Two very large cells were observed in this sample (1.85 x 3.86 µm and 1.02 x 3.82 µm) along with several other large cells including cells of 1.36 x 2.96 µm, 1.24 x 2.89 µm and 0.96 x 2.93 µm. Consistent with the *R.leg*-pIND6 and *R.leg*-pIND6-*popZ* samples, no *R.leg*-pIND6-*rcy* cells longer than 3 µm were observed (Figure 4.14F). In the *R.leg*-pIND6 control sample, no cells were found to have a F_{min} value of above 1.1 µm. However, many of the cells overexpressing Rcy were found to have a F_{min} > 1.1 µm (with cells up to 1.65 µm wide) suggesting that the overexpression of Rcy results in cells which are not necessarily much longer, but are markedly broader.

Comparison of the F_{min} and F_{max} values of cells within these populations, visualised as box-and-whisker plots (Figures 4.14G & 4.14H respectively) highlighted the greater variation in cell width (F_{min}) than length (F_{max}) following ParA1, PopZ and Rcy overexpression. The *R.leg*-pIND6 control sample was found to have a median cell width of 0.84 µm, with a median length of 1.37 µm, consistent with the size of wildtype *R. leguminosarum* strains (*426*). Comparisons of cells isolated from the overexpression cultures to those from the *R.leg*-pIND6 control sample using Mann-Whitney U and Medians tests (non-normal distribution of all datasets, $p < 0.01$) demonstrated that cells within the *R.leg*-pIND6-*parA1* culture typically had a smaller minimum Feret diameter ($p < 0.01$). The median cell width within this culture was 0.72 µm, significantly smaller than that of *R.leg*-pIND6 cells ($p < 0.01$). Cultures overexpressing ParA1 exhibited a greater variation in cell length. However, the median length of cells within of these cultures (1.36 µm) was not significantly different to that of *R.leg*-pIND6 control cells ($p = 0.518$). *R.leg*-pIND6-*popZ* cultures exhibited a greater range of cell widths ($p < 0.01$), and a greater median F_{min} of 0.86 µm ($p < 0.01$). Although wider, cells overexpressing PopZ were generally shorter than *R.leg*-pIND6 cells ($p < 0.01$), with a statistically smaller median length of 1.31 µm ($p < 0.01$). Finally, both the width and length of cells within the *R.leg*-pIND6-*rcy* culture sample were statistically greater than those of *R.leg*-pIND6 ($p < 0.01$), with cells from this culture having a significantly greater median width (1.16 µm) and length (1.49 µm) ($p < 0.01$). The box-and-whisker plots of this data highlight that the increase in cell length was not as pronounced as the increase in cell width (Figures 4.14G & 4.14H). This observation, in conjunction with fact that the width of these cells was generally similar to their length, reflects the spherical 'swollen' morphologies exhibited by these cells (Figure 4.12).

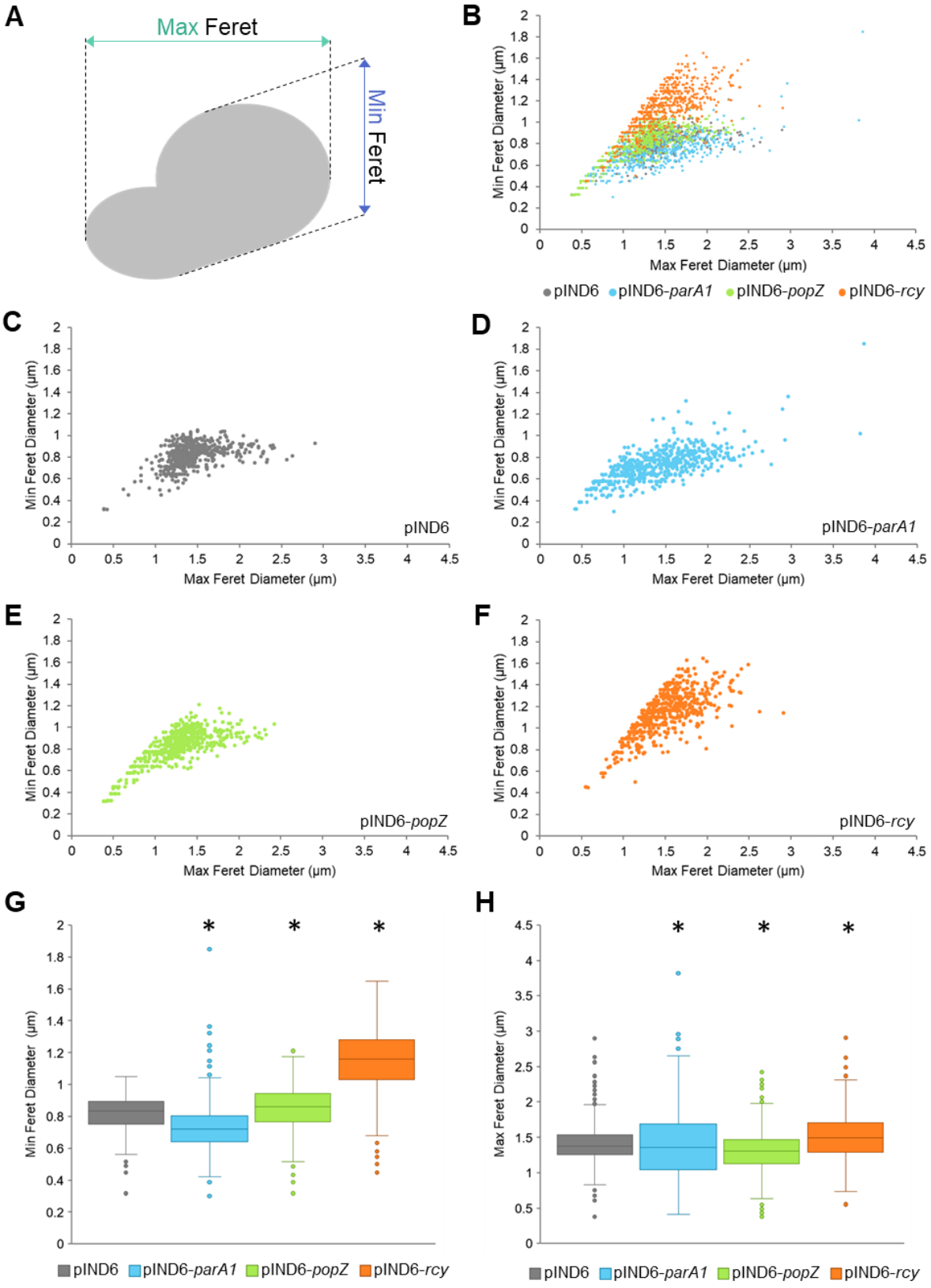


Figure 4.14. Cell length and diameter phenotypes arising from ParA1, PopZ and Rcy overexpression within *R. leguminosarum*. Computational analysis of phase-contrast microscopy images was used to obtain data on the length and width of cells within liquid cultures of the *R. leguminosarum* strains *R.leg*-pIND6 (control), *R.leg*-pIND6-*parA1*, *R.leg*-

pIND6-*popZ* and *R.leg*-pIND6-*rcy*. Cell samples were isolated from cultures 24 hours after induction of protein overexpression. Data from the same 580 cells per strain are shown in this figure as in Figure 4.13. (A) In this study, minimum (min) and maximum (max) Feret diameters were used as measurements of cell width and length respectively. Combined (B) and individual (C-F) scatter plots of min and max Feret diameters of cells from these strains were plotted to visualise and compare the proportions of these cells. Differences in the distribution of min (G) and max (H) Feret diameters were then compared. Asterisks indicate significant differences with *R.leg*-pIND6 cells (Mann–Whitney U test, $p < 0.01$).

A summary of the median values of all cell parameters analysed in this section (area, perimiter, length and width) has been shown in Table 4.2. Indisputably, the overexpression of Rcy within *R. leguminosarum* resulted in the most distinct morphological change of all three proteins. These cells were larger, with phase-contrast microscopy releaving a distinct enlarged coccoid morphology of these cells 24 hours post-induction of protein overexpression (Figure 4.12). The overexpression of ParA1 gave rise to a number of possible cell morphologies. It could be considered that there were three overlapping cell types within these cultures: 'normal' cells, rod-shaped in morphology and similar in size to *R.leg*-pIND6 cells; 'smaller' rod-shaped cells which arose from asymmetric cell division, and the occasional (approximately 1 in 100 cells) large Y- and V-shaped branched cells. These morphologies were not detectable as defined sub-populations, but instead resulted in an even distruption of cell sizes within this population. Overall, cells overexpressing ParA1 were statistically more likely to be shorter than *R.leg*-pIND6 (and by extension wildtype *R. leguminosarum* 3841) cells, with a smaller area and perimiter. Lastly, the overexpression of PopZ had the least significant impact on the morphology of *R. leguminosarum*. Cells within the *R.leg*-pIND6-*popZ* population sample were statisitcally shorter, but wider, than those of the *R.leg*-pIND6 control sample. The median area and perimiter values were not significantly different to those of the *R.leg*-pIND6 sample, suggesting the overexpressopn of PopZ resulted a slight change in cell shape, but not nescessarily overal size. A decrease in cell length coupled with an increase in cell width suggests that cells within this population were more spherical in morphology than the rod-shaped *R.leg*-pIND6 cells, which was consistent with the cell morphologies observed in Figure 4.12. Although an increased prevalence of very small cellular bodies was observed in cultures overexpressing PopZ, the formation of these small cells was not a considered a prominent phenotype.

Table 4.2. Summary of morphological overexpression phenotypes. This table summarises the median area, perimeter, width (minimum Feret diameter) and length (maximum Feret diameter) measurements for cells of the *R. leguminosarum* strains *R.leg*-pIND6 (empty vector control), *R.leg*-pIND6-*parA1*, *R.leg*-pIND6-*popZ*, *R.leg*-pIND6-*rcy*. All samples were taken from cultures 24 hours post induction of protein overexpression. To allow for quick visual comparisons of the data, annotations are included next to values to identify values which are statistically greater than (↑), smaller than (↓) or the same as (-) the *R.leg*-pIND6 empty vector control strain. N = 580 cells (per strain). Asterisks indicate significant differences of medians compared to *R.leg*-pIND6 (empty vector) values following independent sample Medians tests (* denotes $p < 0.05$, ** denotes $p < 0.01$).

Protein	**Area** (μm^2)	**Perimeter** (μm)	**Width** (μm)	**Length** (μm)
Empty vector	0.83	3.57	0.84	1.37
ParA1	↓ 0.70**	↓ 3.40*	↓ 0.72**	- 1.36
PopZ	- 0.82	- 3.50	↑ 0.86**	↓ 1.31**
Rcy	↑ 1.22**	↑ 4.27**	↑ 1.16**	↑ 1.49**

4.6 Discussion

4.6.1 Severe lag phase growth defects suggest a role of these proteins in preparing for growth

Lag phase remains one of the less understood bacterial growth phases. However, it is widely acknowledged that this growth phase serves as a period of adaptation to a new environment and preparation for cellular division. Although minimal cell division occurs during this phase, bacterial cells are still highly metabolically active, prioritising the synthesis of cellular components required for growth and the coordination of the first cellular divisions (*419*, *420*). If ParA1, PopZ and Rcy are solely required during active cell divison, it might be expected that their misregulation during lag phase would have minimal impact on the growth of *R. leguminosaurm* cultures. However, as demonstrated within Section 4.3.2, this is clearly not the case. In this thesis, it was found that the overexpresison of *parA1*, *popZ* and *rcy* during lag phase severly impacted the growth of *R. leguminosarum* cultures, suggesting that the proteins encoded by these genes are not only involved in active cell growth and division, but may also play key roles as cell organisers during preperation for the replication of *R. leguminosarum* cells.

Before cell division can occur, a vast number of cellular components including plasmid and genomic DNA, cell wall synthesis machinery and respiratory complexes need to be localised to the correct regions of the cell, be that the poles or the midcell (*131*, *427*). As reviewed within Chapter One, it is widely documented that a vast array of cytoskeletal (and associated) proteins are involved in the subcellular organisation of bacterial cells, many of which are involved in multiple growth-related cellular processes. ParA proteins are primarily attributed to the segregation of chromosomal plasmids during cell division (reviewed in refs. (*203*, *206*, *428*)), but can also serve roles during the transport of other non-genetic cellular cargo (*429*). Initially described as a polarity determinant which marks the new (growth) pole, the *Agrobacterium tumefaciens* homolog of PopZ has also been linked to coordinating cytoplasmic organisation and chromosome segregation (*273*, *324*, *326*). Although very few 'homologs' of Rcy have been extensively characterised, it is believed that both Scy and GPR function as polar hub proteins within *Streptomyces* and *Agrobacterium* respectively, facilitating the organisation of other polar-localised proteins including membrane and peptidoglycan synthesis proteins (*175*, *182*, *278*, *279*). The observation of such strong lag phase overexpression phenotypes within this

chapter supports the suggestion that the *Rhizobial* homologs of these proteins may serve similar organisational roles within this genus.

On the other hand, it is possible that the severe growth defects observed following the induction of overexpression during lag phase may have been exaggerated by metabolic overloading of the cells. The translation, folding, and modification of proteins is a resource-demanding process, therefore strong overexpression of proteins can drain cellular resources and subsequently impact growth (*430*). Throughout this chapter, 'strong' protein overexpression was induced with a final concentration of 1 mM IPTG regardless of the growth phase of the culture. During lag phase, culture density (i.e. cell number) is low and therefore the ratio of IPTG to cell biomass is high. This relatively high concentration of IPTG can result in an increased uptake of the inducer, thereby causing stronger overexpression and consequently a greater metabolic burden of protein overexpression (*431*). During exponential phase, the ratio of IPTG to cells is lower, overexpression is more likely to occur at a comparatively lower level, and the metabolic burden of overexpression is lessened.

Finally, one of the more surprising observations following the induction of protein overexpression during lag phase was the sudden growth of cultures overexpressing ParA1 after around 60 hours of growth (Figure 4.6F). While this could have been the result of a drastically extended lag phase, alternative hypotheses may include an accumulation of suppressor mutations which rendered the overexpressed ParA1 protein non-functional, a degradation of the antibiotic selection within the media allowing for cells to lose the *parA1* expression plasmid, or the introduction of a contaminant to the media during sampling. Unfortunately, without further experimentation it is impossible to ascertain the cause of this seemingly delayed growth, therefore it should not be assumed that this sudden growth was a direct result of ParA1 overexpression.

4.6.2 The Rcy overexpression growth curve seemingly did not reflect population changes

The impact of ParA1 and PopZ overexpression on bacterial population growth and cell morphology generally appear to have been consistent with one another. The overexpression of PopZ during exponential growth phase was associated with the formation of cells which although similar in area (cross-sectional), were typically slightly shorter yet wider than *R. leguminosarum* control cells (Figure 4.12, Table 4.2). Not only did these cells most closely resemble the rod-shaped morphology of wildtype *R. leguminosarum*, but the overexpression of PopZ (during exponential phase) had minimal impact on the growth of these cultures (Figures 4.7B and 4.7C, Table 4.1). ParA1 overexpression gave rise to alterations in cell morphology which were slightly more pronounced. With the exception of occasional elongated branched cell morphologies, cells overexpressing ParA1 were typically smaller in area than their control strain counterparts, with cells similar in length but narrower (Figure 4.12, Table 4.2). This more pronounced morphological impact of ParA1 overexpression was reflected by the corresponding growth curves, with the overexpression of ParA1 clearly inhibiting population growth (Figures 4.6B, 4.6D and 4.6E, Table 4.1).

Although the overexpression of ParA1 and PopZ resulted in changes in cell morphology which were determined to be statistically significant in at least one parameter e.g. length, width, area, or perimeter, the majority of cells within these overexpression populations typically still resembled a rod or rod-like morphology. By contrast, the overexpression of Rcy (during exponential phase) resulted in the formation of enlarged, coccoid cells. Yet, despite clear morphological defects, these cells seemingly exhibited no population growth-related defects (Figures 4.8B and 4.8C, Table 4.1). Although possible, it seems highly unlikely that such drastic changes in cell morphology would not also impact the growth of *R. leguminosarum* cultures given how closely-linked cell morphology and growth are. This disparity between the observed impacts on cell morphology and culture growth is instead most likely to be a consequence of the method in which the growth of these cultures was determined.

In this thesis, the extent of bacterial growth was determined based on the optical density (measured at a wavelength of 600 nm, OD_{600}) of the bacterial culture. OD_{600} is a widely used measure of bacterial culture growth, and for cultures with uniform cell size it provides a

reasonably good estimate of the density of a culture. However, as OD is influenced by the scatter of light as well as absorbance, OD_{600} readings are sensitive to the size and shape of cells as well as simply the number of cells present. Changes in cell morphology can alter OD_{600} readings, even if the number of cells within a culture does not increase, thereby augmenting the perceived density of the culture (*432*). Instead, alternative measurements of bacterial growth such as those which determine the number of cells within a culture (e.g. flow cytometry or colony-forming unit (CFU) counts) would likely have served as a better approximation for the impact of protein overexpression on bacterial growth within this thesis. It is also worth noting that discrepancies between changes in cell morphology and OD_{600} readings may have been further influenced by changes in cell volume, which was not directly assessed within this thesis.

4.6.3 Impacts of ParA1, PopZ and Rcy overexpression support role of these proteins in controlling growth and morphology

Overall, the data collected within this chapter supports the proposed involvement of ParA1 and PopZ in coordinating growth and division within *R. leguminosarum*. The overexpression of ParA1 not only impacted the growth of *R. leguminosarum* cultures but was also associated with asymmetric cell division and occasional cellular branching. PopZ overexpression typically resulted in formation of slightly smaller cells. Although not a prominent phenotype, clusters of small spherical cells were also observed within these populations (Figure 4.12), potentially as a result of uncoordinated cell division. The morphological impact of PopZ overexpression within the *Rhizobiaceae* is not well-documented. However, the loss or depletion of PopZ within *Agrobacterium tumefaciens* (a member of the *Rhizobiaceae*) has been linked to aberrant cell division and minicell formation (*282*), suggesting role of PopZ may be conserved within *R. leguminosarum* and by extension across the *Rhizobiaceae*. Within *Caulobacter*, although the precise mechanism of PopZ function differs to that of *Agrobacterium*, the overexpression of PopZ was also associated with minicell formation, with multiple cell divisions occurring at the new (growth) pole (*294*). Somewhat conflictingly however, although the loss of *parA* within *Azorhizobium caulinodans* (strain ORS571) has been documented to result in defects in chromosome segregation and variable cell morphology, Chein *et al.* reported that the overexpression of this gene had no impact on growth, cell morphology or even chromosome segregation within free-living cells (*227*, *433*). Naturally, relative overexpression levels between the system used by Chein *et al.* and that used within this thesis will vary, which may account for

differences in the observed morphologies following overexpression. However, both systems involved 'strong' overexpression of this gene, either via constitutive expression from the strong promoter P*nptII* (Chein *et al.* (*433*)), or the induction of expression from the IPTG-inducible $P_{A1/04/03}$ promoter with relatively high IPTG concentrations (this study). Therefore, this may highlight that the precise function or requirement of ParA proteins may differ between different rhizobial families, potentially as a result of their genome structure. Unlike *R. leguminosarum* 3841 which possesses a multipartite genome, the *A. caulinodans* ORS571 genome consists of a single chromosomal plasmid (*226*, *329*).

The role of Rcy plays in controlling cell morphology and growth within *Rhizobium* is a little less clear. Drastic changes in cell morphology coupled with minimal impacts in culture growth suggest that Rcy may be more heavily involved in regulating *R. leguminosarum* morphology than in coordinating its division. However, as discussed in Section 4.6.2, it may be the case that growth-related impacts (measured by OD_{600} of overexpression cultures) were masked by significant changes in cell morphology. Although the reliability of the observed impact of Rcy overexpression on growth is up for debate, there is little doubt that the observed cell morphologies are not a direct result of the overexpression of this protein. Not only were these morphologies unique to the overexpression of Rcy, but both the overexpression and deletion of GPR (considered here to be a homolog of Rcy given its similarity) has been associated with the formation of enlarged, coccoid *Agrobacterium* cells very similar to those observed within this chapter (*278*). Such striking changes in cell morphology following Rcy overexpression may also indicate that this protein is usually present at much lower concentrations within *R. leguminosarum* cells than ParA1 and PopZ. Although the expression of all three proteins was induced with the same concentration of IPTG from the same IPTG-inducible promoter (i.e. 'absolute' overexpression of these proteins was the same), the relative overexpression of these proteins would differ based upon the native abundance of each protein within the cell, with proteins that are usually present at lower molecular concentrations exhibiting a much stronger relative overexpression than those which were already more abundant (*430*). Interestingly, similar enlarged spherical morphologies within *R. leguminosarum* 3841 have been observed following the disruption of genes encoding proteins predicted to be involved in cell envelope development and membrane integrity (*398*, *434*). These proteins include the carboxy terminal protease CtpA (*434*), and an uncharacterised operon (genes *RL3499-RL3502*) encoding a putative moxR-like AAA(+) ATPase, two hypothetical membrane proteins, and an protein of unknown function (*398*). Similarities between Rcy overexpression phenotypes and those

observed following the disruption of these genes may indicate a tentative link between Rcy and the organisation of these cell membrane proteins. This proposed link is further strengthened by the presence of two predicted transmembrane domains within the N-terminus of Rcy which may anchor this protein within the cell membrane (Section 1.5.2).

Ultimately, while the overexpression phenotypes observed within this chapter support the importance of ParA1, PopZ and Rcy in controlling *R. leguminosarum* growth, division and/or cell morphology, these observations alone are not enough to determine their function. Therefore, the next chapter of this thesis will study the localisation of these proteins, with the aim of better understanding their function within *R. leguminosarum* 3841.

Chapter Five – Localisation of EGFP fusion proteins

5.1 Introduction

As demonstrated throughout Chapter Four, the overexpression of ParA1, PopZ and Rcy within *Rhizobium leguminosarum* 3841 resulted in both growth and morphological defects. The overexpression of each of these three proteins during lag phase significantly impeded the growth of *R. leguminosarum* cultures. Although the impact on growth following their overexpression during exponential phase was less pronounced, the overexpression of these proteins during this growth phase did give rise to cell morphologies which significantly differed from the uniform rod-shaped morphology of wildtype *R. leguminosarum* 3841. The most distinct of these morphologies was the formation of enlarged spherical cells following the overexpression of Rcy. Furthermore, the overexpression of ParA1 was associated with an increased occurrence of asymmetric cell division (yielding daughter cells of differing size) and occasional cellular branching.

Having therefore established that these proteins are likely involved in the coordination of growth, division, and cell morphology within *Rhizobium*, it was imperative to determine their intracellular localisation in order to better understand their function and how their misregulation following overexpression resulted in the previously-observed phenotypes. To localise ParA1, PopZ and Rcy within *R. leguminosarum* 3841, proteins were tagged with an enhanced green fluorescent protein (EGFP) tag. Enhanced GFP is a modified variant of the wildtype *Aequorea* GFP protein with improved folding efficiency and a stronger emission signal, making it a more suitable fluorescent marker for intracellular protein localisation than wildtype GFP particularly at low molecular concentrations (*435*, *436*). Imaging of fluorescently tagged proteins in both live and fixed cells has been a core technique for the visualisation and characterisation of numerous members of the bacterial cytoskeleton. The localisation of ParA-YFP (yellow fluorescent protein), CFP (cyan fluorescent protein) and GFP fusions within *Caulobacter crescentus* and *Vibrio cholerae* facilitated the formulation of the now out-of-favour spindle-based mechanism of plasmid segregation by ParA (*208*, *315*, *437*). Later localisation studies of ParA-YFP within *Caulobacter* subsequently provided early support for the now-favoured diffusion-ratchet model of ParA function (*342*). Localisation of PopZ-YFP in *Caulobacter* was vital for its initial characterisation as a pole-organising protein within this species (*294*, *309*), while studies of PopZ-GFP enabled the determination of PopZ as new-pole-specific marker in

Agrobacterium tumefaciens (*273*). Finally, observations of EGFP- and mCherry-Scy, and GFP-GPR fusion proteins enabled the characterisation of these proteins as polar organisers within *Streptomyces* and *Agrobacterium* respectively, as well as revealing the hexameric ring structure of GPR (*182*, *278*).

In this thesis, translational EGFP fusions were generated by the introduction of an EGFP-Apramycin resistance ($Apra^R$) cassette into the existing pIND6-derived expression plasmids used throughout the overexpression studies performed within Chapter Four. The introduction of this cassette into these plasmids allowed for IPTG-inducible expression of N-terminal His-tagged proteins with a C-terminal EGFP tag. Unlike previous plasmid synthesis steps which used traditional cloning methods, these plasmids were generated using bacteriophage lambda (λ) Red recombineering. In this chapter, the intracellular localisation of the ParA1-EGFP, PopZ-EGFP and Rcy-EGFP fusion proteins following their overexpression during exponential phase was studied. The expression of these EGFP fusion proteins was then induced with minimal IPTG concentrations in order to determine the localisation of these proteins at lower molecular concentrations within the cell. By studying changes in cell morphology and protein localisation following both 'high' and 'low' levels of overexpression, the localisation studies performed within this chapter aimed to shed light on the role of these proteins during unipolar growth, and how their misregulation gave rise to the overexpression phenotypes and growth defects previously described. Furthermore, any observation of similar localisation patterns between ParA1-EGFP, PopZ-EGFP and Rcy-EGFP may suggest potential co-localisation or interaction of these proteins within *R. leguminosarum*.

5.2 EGFP-tagging of ParA1, PopZ and Rcy proteins

5.2.1 Bacteriophage lambda (λ) Red recombination-mediated genetic engineering

Incorporation of EGFP tags to the protein coding sequences within the pIND6-derived expression plasmids was achieved using an adaptation of a bacteriophage lambda (λ) Red recombinase-mediated protocol for gene disruption. This genetic engineering protocol was initially developed by Datsenko and Wanner (2000) for use in *E. coli*, and subsequently modified by Gust *et al.* (2002) for gene disruption within *Streptomyces* (*349*, *358*). In short, this protocol exploits the λ Red recombinase genes (*gam*, *bet*, *exo*) to promote recombination between a target gene region

and a linear DNA fragment (i.e. a disruption cassette). The λ Red recombinase genes *bet* and *exo* promote recombination, while *gam* prevents the endogenous nucleases RecBCD and SbcCD degrading the linear DNA fragment introduced into the cell (*351*, *438*, *439*). The λ Red genes are encoded on the λ Red recombination plasmid pIJ790 (*349*, *358*), a modified version of the plasmid pKD20 (*351*) in which the ampicillin resistance gene *bla* was replaced with the chloramphenicol resistance gene *cat*. This plasmid also encodes the L-arabinose operon regulator protein AraC, enabling the expression of the λ Red genes from the L-arabinose-inducible promoter P_{araBAD}. Genetic manipulation using this protocol is therefore performed within the *E. coli* strain BW25113, which is unable to metabolise arabinose (*ΔaraBAD*) (Table 2.1). Finally, the temperature sensitive low-copy origin of replication (oriR101 & repA101-ts) of pIJ790 facilitates the loss of this plasmid from *E. coli* BW25113 following incubation at an elevated temperature (37°C as opposed to 30°C).

An overview of the use of this protocol to generate the EGFP fusion proteins used for localisation studies within this thesis has been shown in Figure 5.1. A detailed methodology of this protocol can be found in Section 2.7, with the application of this method covered in further detail within Section 5.2.2. EGFP translational fusion proteins were generated by introducing an *egfp*-$Apra^R$ cassette immediately downstream of the *parA1*, *popZ* and *rcy* gene sequences within the pIND6-derived plasmids generated in Section 4.2.2. This cassette consisted of the *egfp* gene, and the apramycin resistance ($Apra^R$) gene *aac(3)-IV*. Immediately upstream of the EGFP coding region was a 12 bp linker sequence (CCGGTCGCCACC), introduced to facilitate proper folding and maturation of both EGFP and the tagged protein (i.e. ParA1, PopZ or Rcy). Also included within this cassette but not used within this thesis (and therefore not illustrated) were FRT sequences (FLP recognition targets) flanking the $Apra^R$ region of the cassette to allow for FLP recombinase-mediated removal of this resistance cassette if desired (*440*). To enable λ Red-mediated homologous recombination between this cassette and the pIND6-derived expression plasmids, the *egfp*-$Apra^R$ cassette was extended by PCR to introduce 40 bp target-specific homologous flanking sequences. The primers used for this PCR amplification comprised of a 5' 40 nucleotide (nt) sequence homologous to the target region of the plasmid, followed by a 22 nt region specific to the cassette (Figure 5.1A). For forward primers, the 5' 40 nt sequence was homologous to the coding strand sequence immediately upstream of the stop codon of the target gene, with the cassette-specific 22 nt sequence annealing to the 12 bp linker sequence and the first 10 bp of *egfp*. For the reverse primers, the 5' 40 nt sequence was homologous to the complimentary strand sequence immediately downstream of the target gene, with the cassette-specific 22 nt

binding downstream of Apra^R. Importantly, the regions of homology incorporated into the extended cassette did not encompass the stop codon at end of the targeted gene, thereby allowing for the generation of a translational fusion following allelic exchange between the plasmid and the cassette.

In order to introduce the *egfp*-Apra^R cassette into pIND6-*parA1*, pIND6-*popZ* and pIND6-*rcy*, these plasmids were first electroporated into *E. coli* BW25113 cells carrying the λ Red recombination plasmid pIJ790 (BW25113/pIJ790) (Figure 5.1B). Expression of the λ Red genes within the resulting BW25113/pIJ790/pIND6 strains was then induced by culturing these strains in the presence of L-arabinose. Following induction, the extended *egfp*-Apra^R cassettes were electroporated into their corresponding BW25113/pIJ790/pIND6 strains (Figure 5.1C). λ Red-mediated recombination between homologous regions of the cassette and pIND6-derived expression plasmid resulted in the integration of the *egfp*-Apra^R cassette immediately downstream of the target gene. During allelic exchange, the stop codon of the targeted gene is lost, generating a translational fusion. Incubation of these cells at 37°C drives the loss of the temperature sensitive plasmid pIJ790 (Figure 5.1D). Plasmid preparations containing only pIND6-derived plasmids can then be isolated from the BW25113 cells for further screening.

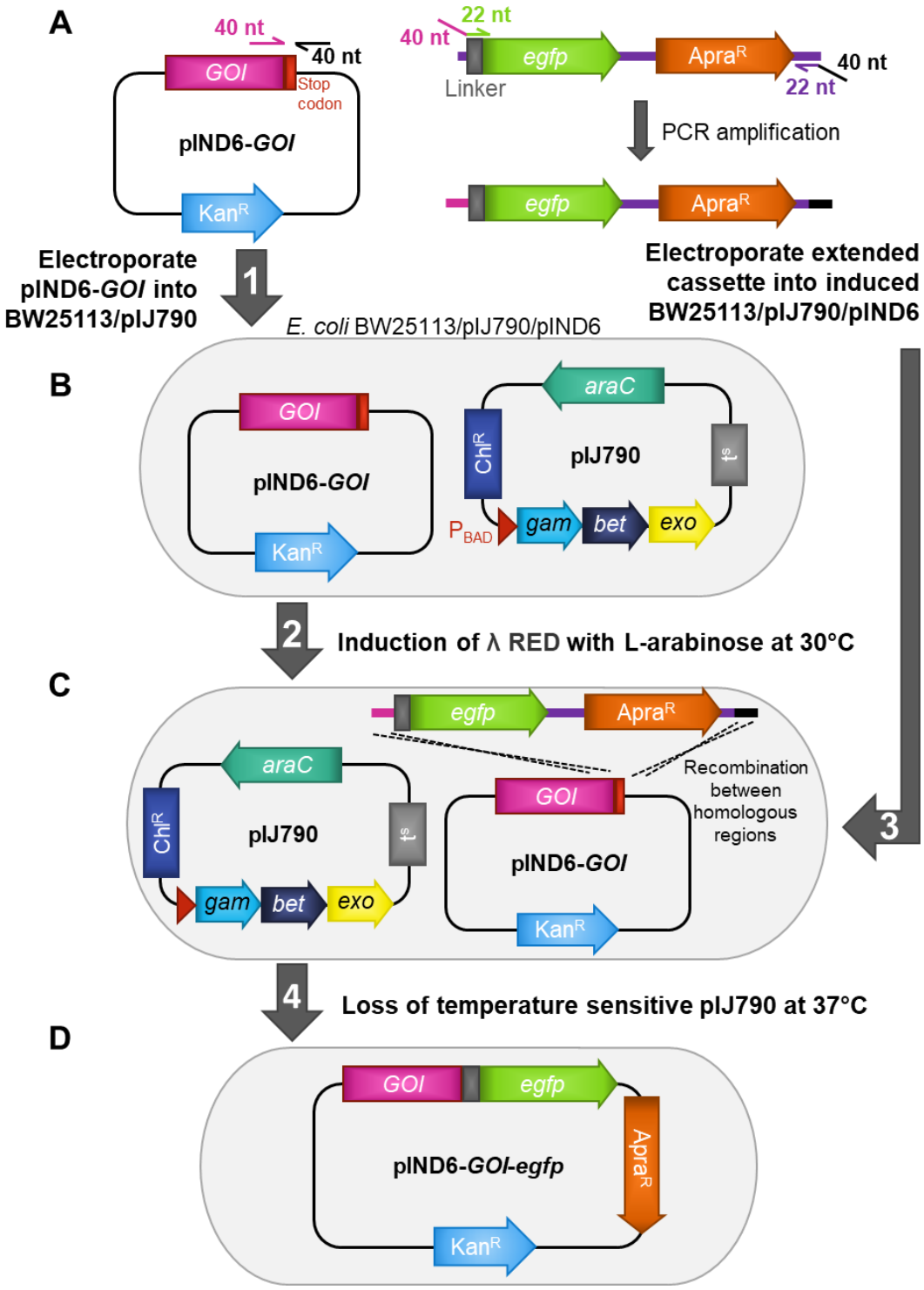


Figure 5.1. The generation of translational EGFP fusions using phage lambda (λ) Red recombination-based genetic engineering. C-terminal EGFP tagged proteins were generated by the introduction of an *egfp*-Apra^R cassette into the pIND6-derived *parA1*, *popZ* and *rcy* expression plasmids using phage lambda (λ) Red-mediated homologous recombination. This protocol was initially developed by Datsenko and Wanner (2000), and subsequently modified by Gust *et al.* (2002) (*349*, *351*, *358*). (A) Homologous flanking regions were introduced to the *egfp*-Apra^R cassette via PCR amplification. These flanking regions were homologous to the 40 bp immediately upstream and downstream of (but not including) the stop codon (red) of the gene of interest (*GOI*, pink). (B) pIND6-derived expression plasmids were electroporated into the *E. coli* strain BW25113, carrying the λ Red recombination plasmid pIJ790

(BW25113/pIJ790). This plasmid carries the λ Red genes *gam*, *bet* and *exo*, driven by the L-arabinose-inducible P_{araBAD} promoter (abbreviated to P_{BAD}). This plasmid also encodes the arabinose activator AraC (*araC*), and a chloramphenicol resistance (Chl^R) gene. Presence of a temperature sensitive replicon (t^s) on this plasmid means cells carrying pIJ790 must be cultured at 30°C to retain this plasmid. Kanamycin resistance (Kan^R) conferred by the pIND6-derived plasmids allows for kanamycin-based selection of transformants. (C) The expression of the λ Red genes is induced in the presence of L-arabinose. Following induction, the extended *egfp*-$Apra^R$ cassette is electroporated into BW25113/pIJ790/pIND6 cells. Red-mediated homologous recombination results in the introduction of the *egfp*-$Apra^R$ cassette into pIND6-*GOI* immediately downstream of the target gene, generating a translational fusion. (D) Incubation at 37°C results in the loss of pIJ790, leaving the newly-generated plasmid pIND6-*GOI-egfp* within *E. coli* BW25113.

5.2.2 Generation of the ParA1-, PopZ-, and Rcy-EGFP expression plasmids

As depicted in Figure 5.1, one of the first steps in the generation of EGFP-tagged ParA1, PopZ and Rcy proteins was the electroporation of the pIND6-derived expression plasmids pIND6-*parA1*, pIND6-*popZ*, and pIND6-*rcy* (Table 2.3) into the *E. coli* strain BW25113/pIJ790 (Table 2.1). Electrocompetent BW25113/pIJ790 cells were prepared as described in Section 2.4.1. Following electroporation of pIND6-*parA1*, pIND6-*popZ* and pIND6-*rcy* into BW25113/pIJ790 (Section 2.4.1), electroporation reactions were plated onto LB Agar containing 50 μg mL^{-1} kanamycin (kan) and 25 μg mL^{-1} chloramphenicol (chl) to select for successful transformants. Due to the temperature sensitive replicon of pIJ790, plates were incubated at 30°C as opposed to 37°C (typically used for *E. coli*) to ensure retention of this plasmid. For simplicity, these transformants have been collectively referred to as BW25113/pIJ790/pIND6 strains within this chapter. To generate the extended *egfp*-$Apra^R$ cassette required for successful integration into pIND6-*parA1*, pIND6-*popZ* and pIND6-*rcy*, the *egfp*-$Apra^R$ cassette was amplified with three distinct PCR primer pairs designed to incorporate target-specific homologous flanks onto this cassette. These primer pairs were as follows: *egfp_parA1*_Fwd & *egfp_parA1*_Rev, *egfp_popZ*_Fwd & *egfp_popZ*_Rev and *egfp_rcy*_Fwd & *egfp_rcy*_Rev. The sequences of these primers can be found in Table 2.7. Successful amplification of this cassette with each of these three primer pairs generated a 2.2 kb PCR product, comprising of the 2142 bp *egfp*-$Apra^R$ cassette with 40 bp 5' and 3' homologous flanking regions (Figure 5.2A).

Following PCR amplification (Section 2.5.2), 3 µL samples of each PCR reaction were analysed by agarose gel electrophoresis to confirm successful amplification of this cassette and determine the appropriate method of purification of the extended cassette. As shown in Figure 5.2B, successful amplification of the cassette with all three gene-specific primer pairs was achieved. However, as non-specific amplification of the cassette also occurred, the extended *egfp*-Apra^R cassettes were purified from the remaining PCR reaction volume by gel purification (Section 2.3.3).

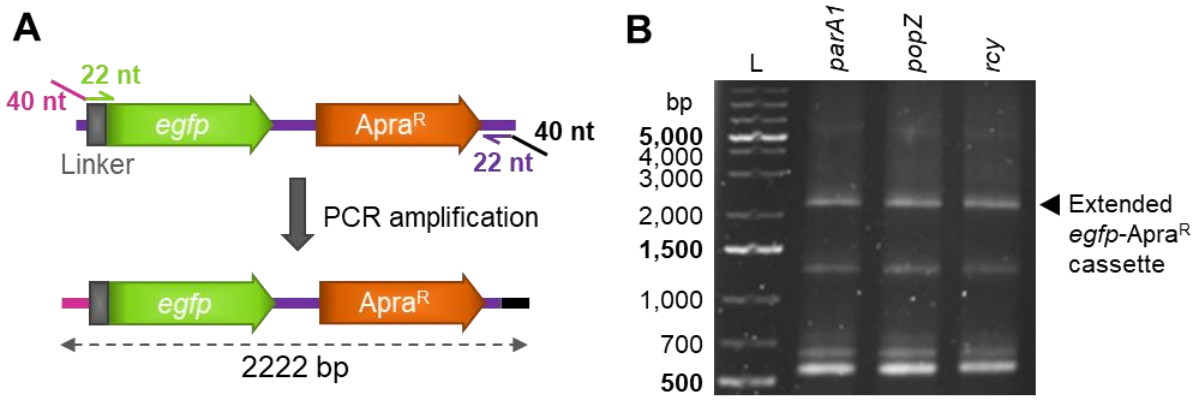


Figure 5.2. Extension the *egfp*-Apra^R cassette using PCR amplification. (A) An extended *egfp*-Apra^R cassette was generated by PCR amplification with primer pairs designed to introduce 40 bp flanking regions to the cassette. The forward PCR primer (pink and green half arrow) was designed to incorporate a sequence homologous to the final 40 bp of the target gene (excluding the stop codon), with the reverse primer (purple and black half arrow) designed to incorporate a 40 bp sequence homologous to the region immediately downstream of the gene. (B) Agarose gel electrophoresis (1% w/v) of the DNA fragments generated following PCR amplification of the *egfp*-Apra^R cassette with the extension primers depicted in Part A. Lane L: GeneRuler 1kb Plus DNA ladder. Remaining lane labels represent the gene-specific primer pairs used for amplification: *parA1*: *egfp_parA1*_Fwd & *egfp_parA1*_Rev; *popZ*: *egfp_popZ*_Fwd & *egfp_popZ*_Rev; and *rcy*: *egfp_rcy*_Fwd & *egfp_rcy*_Rev.

Once the extended *egfp*-Apra^R cassettes were purified, 10 mL overnight cultures of the BW25113/pIJ790/pIND6 strains were set up as described in Section 2.7. These overnight cultures were then subcultured in 10 mL SOB media (Table 2.2) containing 50 µg mL^{-1} kanamycin, 25 µg mL^{-1} chloramphenicol and 10 mM L-arabinose. The presence of L-arabinose within these culture volumes induced the expression of the λ Red genes (*bet*, *gam*, *exo*) encoded on pIJ790. Following a 3-4 hour subculture, cells were prepared for electroporation and the purified extended *egfp*-Apra^R cassettes were subsequently electroporated into the corresponding BW25113/pIJ790/pIND6 strains. Electroporation reactions were plated onto LB Agar containing 50 µg mL^{-1} of both kanamycin and apramycin to select for cells carrying plasmids which had undergone successful recombination events, and incubated overnight at 37°C to promote the loss of pIJ790 from the cells. It is important to note that multiple copies of the pIND6-derived expression plasmids will be present within a single BW25113 cell. However, not all copies will have undergone a successful recombination event. Therefore, BW25113 colonies which grow in the presence of kanamycin (kan) and apramycin (apra) will likely contain a mixture of 'tagged' and 'non-tagged' plasmids. Given this, plasmid was isolated from these colonies, and subsequently transformed (via heat-shock) into *E. coli* DH5α to obtain isogenic *E. coli* colonies. Once more, transformation reactions were plated onto LB Agar plates containing kan and apra (as above) to select for cells carrying successfully 'tagged' plasmids. These tagged plasmids are referred to as pIND6-*parA1-egfp*, pIND6-*popZ-egfp* and pIND6-*rcy-egfp* hereafter.

Plasmid isolated from three DH5α colonies (per gene) was validated with an XbaI restriction digest, illustrated in Figure 5.3A. XbaI-digestion of pIND6-*parA1-egfp*, pIND6-*popZ-egfp* and pIND6-*rcy-egfp* produced three distinct DNA fragments: a fragment of variable size containing the *egfp*-tagged gene sequence (*parA1*: 1688, *popZ*: 1775, *rcy*: 8021 bp), a fragment of 1301 bp containing the Apra^R gene, and a pIND6 backbone fragment of 8415 bp. Agarose gel electrophoresis of the products of these digestions (Figure 5.3B) highlighted incomplete digestion of these plasmids, demonstrated by the presence of a > 10 kb band within pIND6-*rcy-egfp* digests, and an approximately 3 kb band within pIND6-*parA1-egfp* and pIND6-*popZ-egfp* digests. However, although the 8.0 kb (gene fusion) and 8.4 kb (backbone) fragments of pIND6-*rcy-egfp* did not fully resolve, clear *parA1-egfp* and *popZ-egfp* fragment bands were observed (1.7 and 1.8 kb respectively). Furthermore, 1.3 kb Apra^R fragments were clearly identified in all plasmid digests. The observation of these digest fragments was consistent with the isolation and digestion of correct pIND6-*parA1-egfp*, pIND6-*popZ-egfp* and pIND6-*rcy-egfp* clones.

As an additional confirmatory step, the *E. coli* DH5α colonies from which the digested plasmids (Figure 5.3) were isolated were streaked onto LB Agar plates containing 0.5 mM IPTG to induce expression of the EGFP fusion proteins. A reduced IPTG concentration was used in these plates as it was previously found that 1 mM IPTG was detrimental to the growth of *E. coli* cells carrying the pIND6-derived expression plasmids, presumably as a result of overexpression of heterologous ParA1, PopZ and Rcy proteins (data not shown). Following an overnight incubation of these plates, the *E. coli* strains were checked for EGFP expression by epifluorescence microscopy (Section 2.10). EGFP signal was detected in all *E. coli* strains (data not shown), confirming that the plasmids screened within Figure 5.3 had been successfully tagged with functional EGFP. The plasmids pIND6-*parA1-egfp*, pIND6-*popZ-egfp* and pIND6-*rcy-egfp* were then conjugated into *R. leguminosarum* 3841 (Section 2.4.2), to generate the ParA1-, PopZ-, and Rcy-EGFP expression strains (referred to as *R.leg-parA1-egfp*, *R.leg-popZ-egfp* and *R.leg-rcy-egfp* respectively) used for localisation studies throughout the remainder of this chapter. It is worth noting that as EGFP tags were introduced to the His-tagged protein-coding sequences within pIND6-*parA1*, pIND6-*popZ* and pIND6-*rcy*, the resulting fusion proteins possessed both an N-terminal His tag, and a C-terminal EGFP tag (Figures 5.4, 5.5 and 5.6).

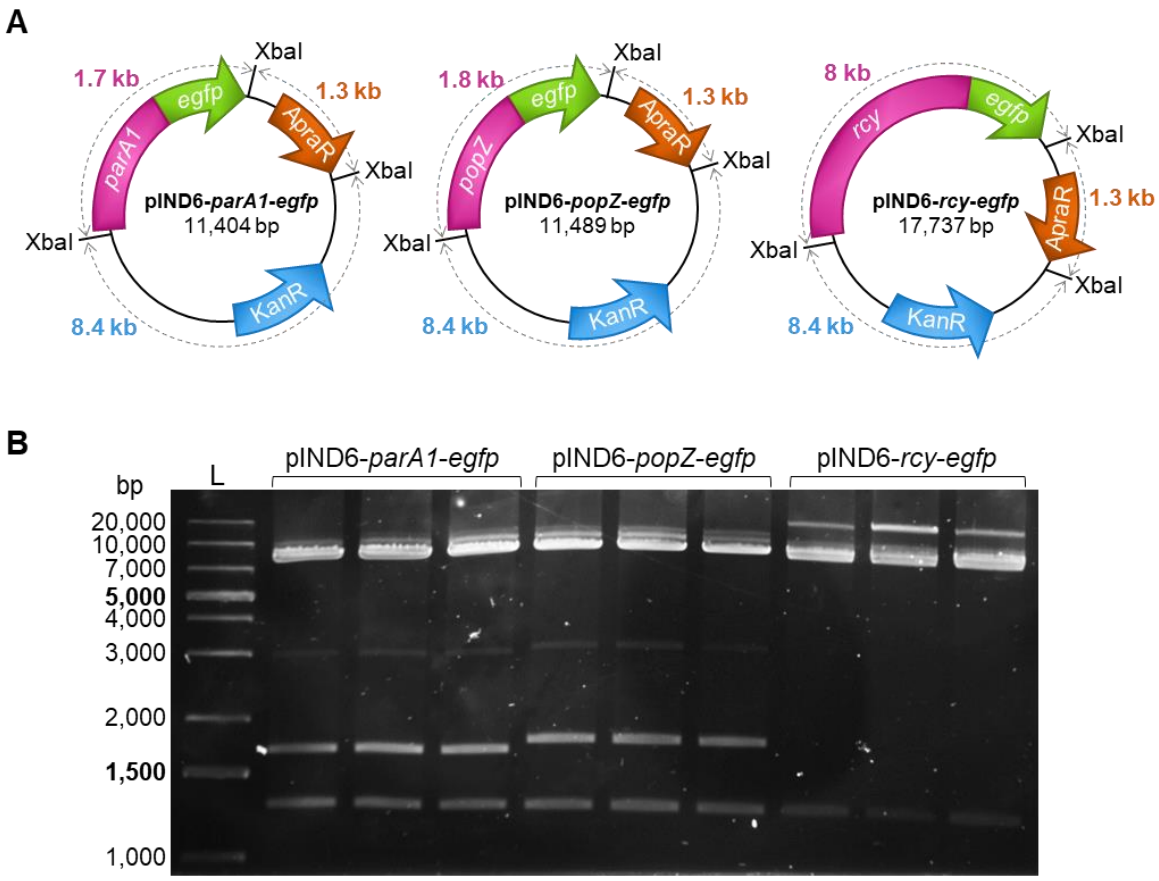


Figure 5.3. XbaI validation digests of the ParA1-, PopZ- and Rcy-EGFP fusion protein expression plasmids. (A) XbaI restriction digestion was used to screen for successful integration of the *egfp*-Apra^{R} cassette into the pIND6-derived *parA1*, *popZ* and *rcy* expression vectors. XbaI digestion of pIND6-*parA1-egfp*, pIND6-*popZ-egfp*, and pIND6-*rcy-egfp* should generate an 8.4 kb pIND6 backbone fragment containing the kanamycin resistance (KanR) gene (blue), a 1.3 kb Apra^{R}-containing fragment (ApraR, orange), and third fragment of variable length containing the *egfp* fusion protein sequence (*parA1-egfp*: 1.7 kb, *popZ-egfp*: 1.8 kb, *rcy-egfp*: 8 kb). (B) Agarose gel electrophoresis (1% w/v) of the XbaI restriction digest products obtained following digestion of three independent putative pIND6-*parA1-egfp*, pIND6-*popZ-egfp*, and pIND6-*rcy-egfp* plasmid preparations. Lane L: GeneRuler 1kb Plus DNA ladder. Labels above wells denote the plasmid samples within each well.

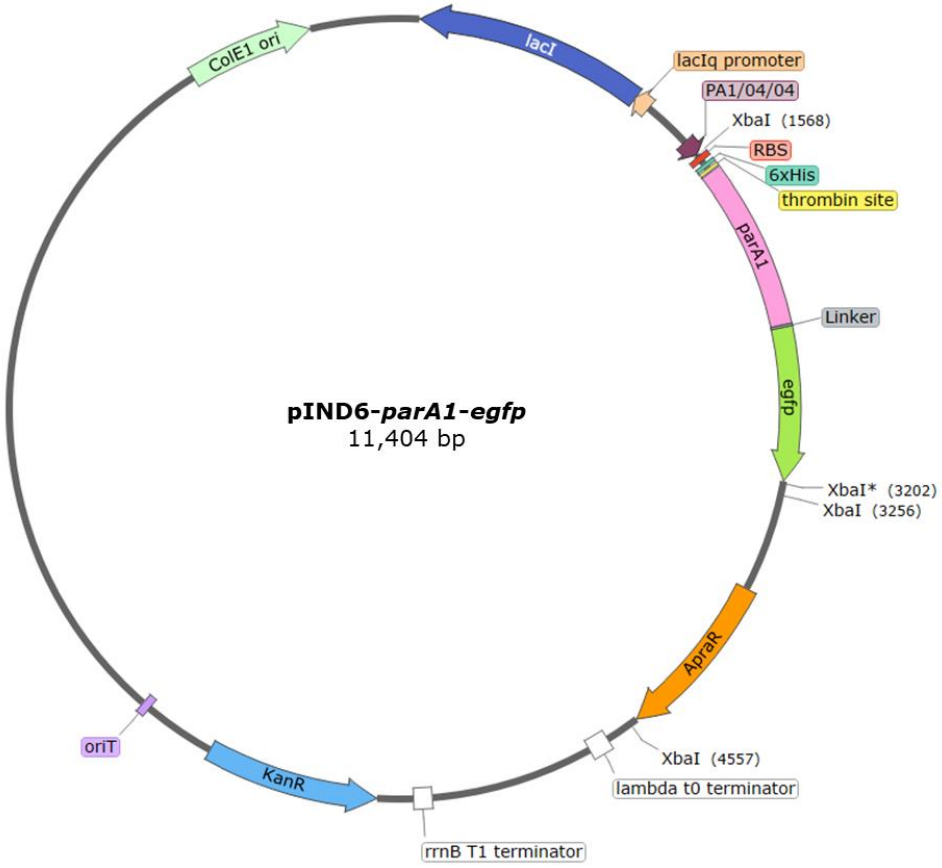


Figure 5.4. The IPTG-inducible ParA1-EGFP expression plasmid pIND6-*parA1-egfp*. An *egfp*-Apra^{R} (apramycin resistance) cassette was introduced into pIND6-*parA1*, generating pIND6-*parA1-egfp*. This plasmid encodes a ParA1-EGFP translational fusion protein with an N-terminal His-tag (turquoise) and thrombin cleavage site (yellow). A 12 bp linker sequence (grey) located between *egfp* (green) and *parA1* (pink) facilitated proper folding of both EGFP and ParA1 in the fusion protein. The Apra^{R} gene (orange) encoded the aminoglycoside acetyltransferase AAC(3)-IV. IPTG-inducible expression of this fusion protein was driven by the $P_{A1/04/03}$ promoter (burgundy). Other key features of this plasmids include the *lacI* gene (indigo) driven by the *lacI*^{q} promoter (light orange), a kanamycin resistance (Kan^{R}) gene (blue), ribosomal binding site (red), origin of transfer (oriT, purple), the lambda t0 and rrnB T1 transcriptional terminators (grey & white) and a ColE1 origin of replication (light green). XbaI recognition sites, used for validation digests have been highlighted. XbaI* denotes an XbaI recognition site which is blocked by dam methylation.

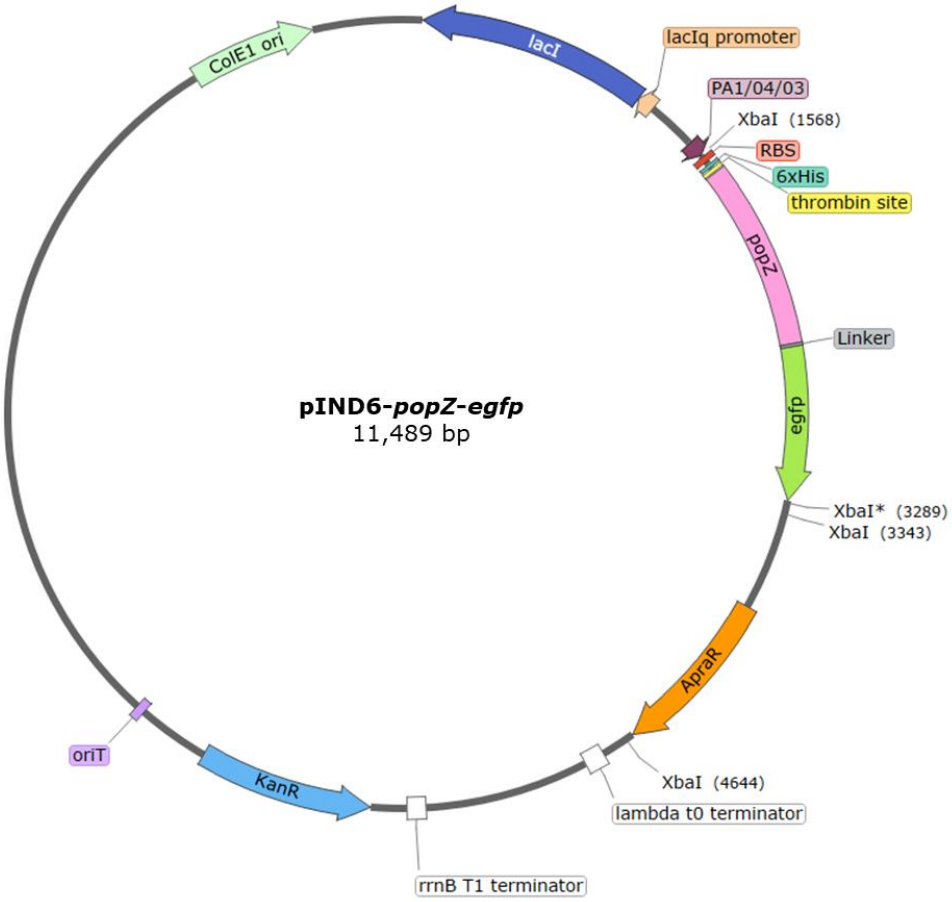


Figure 5.5. The IPTG-inducible PopZ-EGFP expression plasmid pIND6-*popZ-egfp*. An *egfp*-Apra^R (apramycin resistance) cassette was introduced into pIND6-*popZ*, generating pIND6-*popZ-egfp*. This plasmid encodes a PopZ-EGFP translational fusion protein with an N-terminal His-tag (turquoise) and thrombin cleavage site (yellow). A 12 bp linker sequence (grey) located between *egfp* (green) and *popZ* (pink) facilitated proper folding of both EGFP and PopZ in the fusion protein. The Apra^R gene (orange) encoded the aminoglycoside acetyltransferase AAC(3)-IV. IPTG-inducible expression of this fusion protein was driven by the $P_{A1/04/03}$ promoter (burgundy). Other key features of this plasmids include the *lacI* gene (indigo) driven by the *lacI*^{q} promoter (light orange), a kanamycin resistance (Kan^R) gene (blue), ribosomal binding site (red), origin of transfer (oriT, purple), the lambda t0 and rrnB T1 transcriptional terminators (grey & white) and a ColE1 origin of replication (light green). XbaI recognition sites, used for validation digests have been highlighted. XbaI* denotes an XbaI recognition site which is blocked by dam methylation.

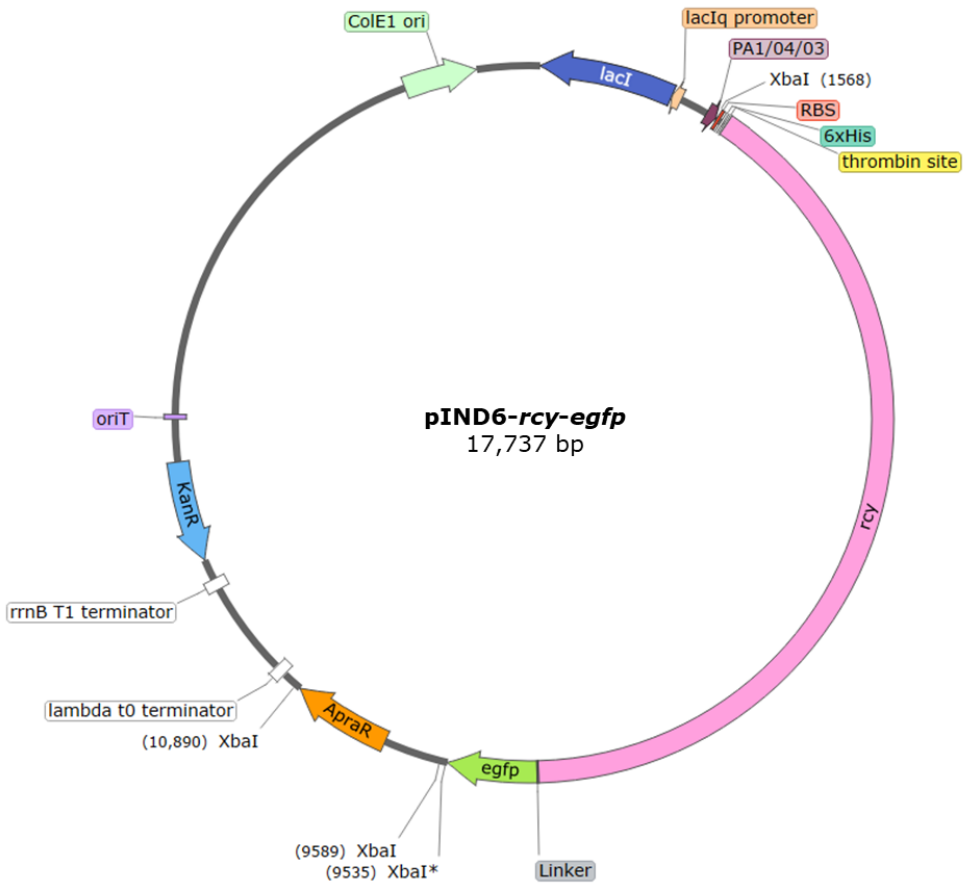


Figure 5.6. The IPTG-inducible Rcy-EGFP expression plasmid pIND6-*rcy-egfp*. An *egfp*-Apra^R (apramycin resistance) cassette was introduced into pIND6-*rcy*, generating pIND6-*rcy-egfp*. This plasmid encodes an Rcy-EGFP translational fusion protein with an N-terminal His-tag (turquoise) and thrombin cleavage site (yellow). A 12 bp linker sequence (grey) located between *rcy* (pink) and *egfp* (green) facilitated correct folding of both EGFP and Rcy in the fusion protein. IPTG-inducible expression of this fusion protein was driven by the $P_{A1/04/03}$ promoter (burgundy). The Apra^R gene (orange) encoded the aminoglycoside acetyltransferase AAC(3)-IV. Other key features of this plasmids include the *lacI* gene (indigo) driven by the *lacI*^{q} promoter (light orange), a kanamycin resistance (Kan^R) gene (blue), ribosomal binding site (red), origin of transfer (oriT, purple), the lambda t0 and rrnB T1 transcriptional terminators (grey & white) and a ColE1 origin of replication (light green). XbaI recognition sites, used for validation digests have been highlighted. XbaI* denotes an XbaI recognition site which is blocked by dam methylation.

5.3 Localisation of overexpressed EGFP fusion proteins

To study the localisation of ParA1, PopZ and Rcy during their overexpression, the *R. leguminosarum* EGFP fusion expression strains *R.leg-parA1-egfp*, *R.leg-popZ-egfp* and *R.leg-rcy-egfp* (Section 5.2.2) were cultured in the same conditions as those used for the overexpression studies performed throughout Chapter Four. Details of these conditions can be found in Sections 2.1.2 and 2.1.3. In short, strains were grown in 50 mL TY cultures in the presence of 200 µg mL^{-1} kanamycin. Although also resistant to apramycin, the EGFP fusion strains were grown in the presence of kanamycin only to ensure that more reliable comparisons between the localisation patterns observed within this chapter and the overexpression phenotypes analysed within Chapter Four could be made. The overexpression of ParA1-EGFP, PopZ-EGFP and Rcy-EGFP was induced with the addition of a final concentration of 1 mM IPTG to the cultures at 20 hours (i.e. during exponential phase). The localisation of these proteins at both 6 and 24 hours post-induction was then determined by epifluorescence microscopy. Details of slide preparation and epifluorescence microscopy can be found within Sections 2.10.1 and 2.10.2 respectively. Before performing any localisation studies, the pIND6 empty vector control strain *R.leg*-pIND6 was checked for any intrinsic cellular fluorescence (autofluorescence) which may interfere with the localisation of EGFP-tagged proteins. To check for autofluorescence, this strain was cultured as described above and analysed by epifluorescence microscopy with a GFP fluorescence filter set. As shown in Figure 5.7, little-to-no detectable green autofluorescence signal was observed within these samples. Therefore, any fluorescence emission signal detected within *R.leg-parA1-egfp*, *R.leg-popZ-egfp* and *R.leg-rcy-egfp* was the direct result of excitation of EGFP and not autofluorescence.

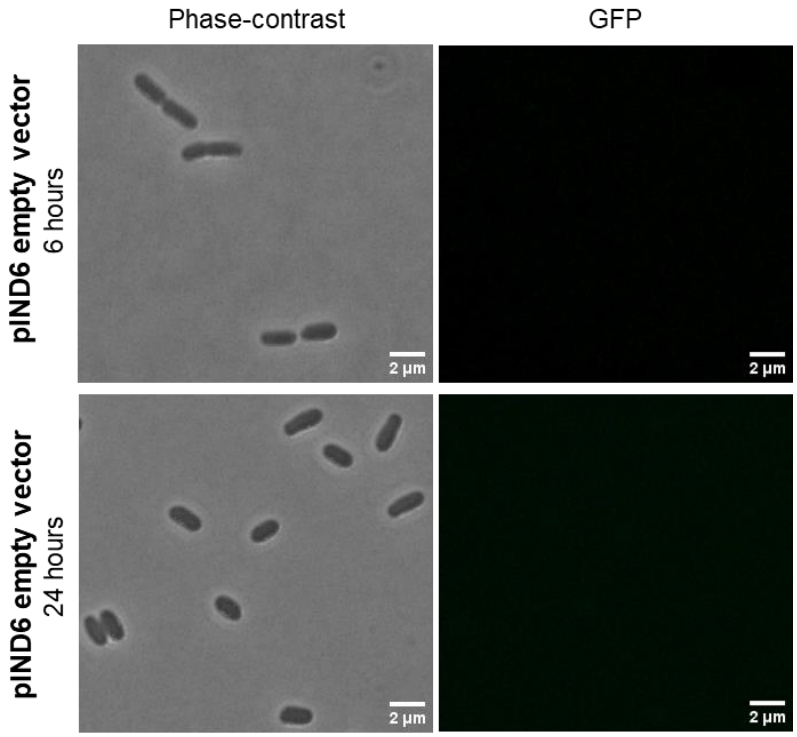


Figure 5.7. Confirmation of the absence of detectable green autofluorescence within *R. leguminosarum*. Phase contrast (left) and epifluorescence (right) microscopy was used to check for autofluorescence within the *R. leguminosarum* pIND6 empty vector control strain (*R.leg*-pIND6). Epifluorescence microscopy was performed using a GFP fluorescence filter set. *R.leg*-pIND6 was cultured in TY media in the presence of kanamycin. At 20 hours, 1 mM IPTG was added to the cultures. Cultures were then sampled at 6 and 24 hours after the addition of IPTG.

When overexpressed, ParA1-EGFP did not exhibit a distinct localisation pattern within *R. leguminosarum* cells (Figure 5.8). At both 6 and 24 hours post-induction, overexpression of ParA1-EGFP resulted in homogeneous fluorescence signal throughout the cytoplasm of the majority of cells sampled. A stronger EGFP signal was observed at 24 hours post induction, suggesting a higher abundance of protein at this timepoint. Within some cells, a single concentration of fluorescence signal (referred to as a focus) was also observed amongst the whole-cell fluorescence signal. However, the localisation of these ParA1-EGFP foci was not consistent between cells. Often foci were localised at a single cell pole, but concentrations of ParA1-EGFP were also observed at the edge of some cells, and within the centre of others. Somewhat surprisingly, no evidence of cellular branching was observed within samples isolated at 24 hours post-induction. Figure 5.8 shows only a representative sample of the *R.leg-parA1-*

egfp cells observed. In total, several hundred cells were imaged, none of which displayed the distinct Y- and V-shaped branched morphologies previously observed following the overexpression of ParA1 within Chapter Four (Section 4.5.1, Figure 4.12). Figure 5.8 does however show an irregularly shaped cell at 6 hours post induction, which may be displaying early signs of cellular branching. While this could suggest that the functionality of ParA1-EGFP differs to that of ParA1, cell branching was an infrequent phenotype of ParA1 overexpression, arising in approximately only 1 in 100 cells, therefore is possible this phenotype was simply not observed within the *R.leg-parA1-egfp* samples taken here. Furthermore, evidence of what appeared to be asymmetric division was frequently observed at both 6 and 24 hours post-induction of ParA1-EGFP expression, indicating that the phenotypic impact of ParA1-EGFP overexpression did not differ substantially from that of ParA1.

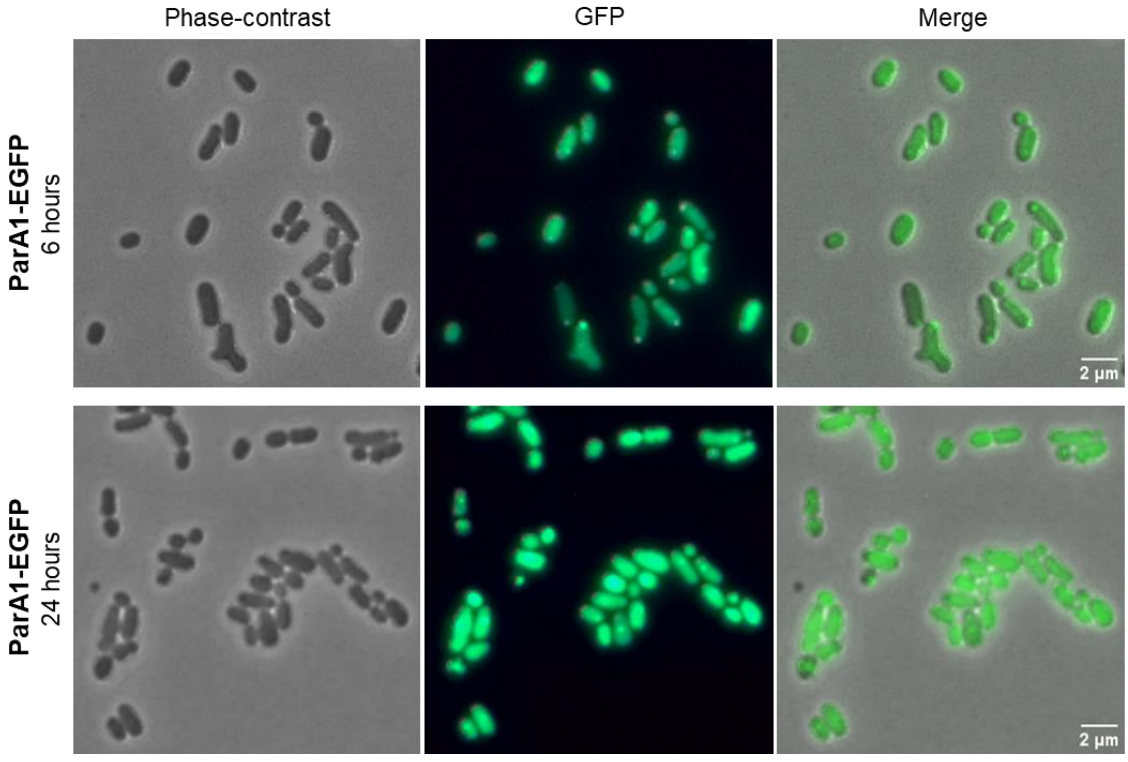


Figure 5.8. Overexpression localisation of ParA1-EGFP within *R. leguminosarum*. The *R. leguminosarum* strain *R.leg-parA1-egfp* carrying the ParA1-EGFP expression plasmid pIND6-*parA1-egfp* was grown in TY media in the presence of kanamycin to maintain selection of this plasmid. At 20 hours, overexpression of ParA1-EGFP was induced with the addition of 1 mM IPTG to the growth media. Culture samples were isolated at 6 and 24 hours post induction and viewed by epifluorescence microscopy (GFP filter set) to assess ParA1-EGFP localisation. Phase contrast, epifluorescence (GFP) and overlayed (merge) images have been shown.

In contrast to the localisation of ParA1-EGFP, PopZ-EGFP exhibited a distinct unipolar localisation pattern in which a single, polar fluorescence focus was observed within the vast majority of cells (Figure 5.9). The localisation pattern of these foci remained consistent between 6 and 24 hours post induction. Small cellular bodies similar to those described in Chapter Four following PopZ overexpression (Figure 4.12) were also observed following the overexpression of PopZ-EGFP. These small spherical cells typically lacked EGFP signal suggesting an absence of PopZ-EGFP within these cells (Figure 5.9, arrows). Analysis of cellular morphology may suggest that PopZ-EGFP foci were located at the growth pole. Although research is more limited within *Rhizobium* species, it is well established that *Agrobacterium* (a closely-related member of the *Rhizobiaceae*) displays a tapered growth pole (*127*, *152*, *273*). As seen in Figure 5.9, many foci were located within what appeared to be the narrower (i.e. tapered) end of the cell, which may be suggestive of growth poles *within R. leguminosarum*. However, as previously demonstrated in Chapter Four, the overexpression of PopZ resulted in defects in cell morphology. Hence, definitive conclusions regarding the localisation of these foci based solely on cellular morphology cannot be made. It was also observed that the signal from cells expressing PopZ-EGFP was generally weaker than that of ParA1-EGFP when same intensity of excitation was used. It is important to note that to optimise image acquisition, the excitation intensity was adjusted between samples. Therefore, any comparisons of fluorescence intensity made throughout this chapter are anecdotal, and not direct comparisons of the epifluorescence microscopy images shown within the figures throughout this chapter.

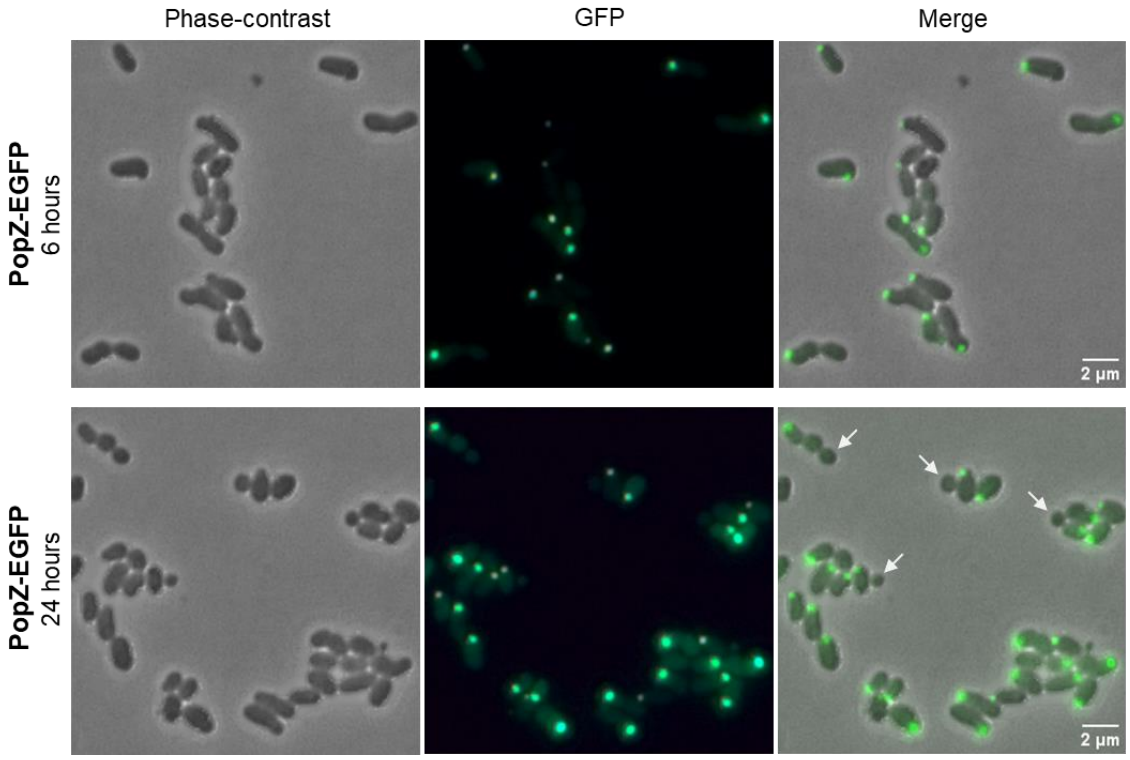


Figure 5.9. Overexpression localisation of PopZ-EGFP within *R. leguminosarum*. An *R. leguminosarum* 3841 strain carrying pIND6-*popZ-egfp* (*R.leg-popZ-egfp*) was grown in TY media in the presence of kanamycin to maintain selection of this plasmid. At 20 hours, overexpression of PopZ-EGFP was induced with the addition of 1 mM IPTG to the growth media. Culture samples were isolated at 6 and 24 hours post induction and viewed by epifluorescence microscopy (GFP filter set) to assess PopZ-EGFP localisation. Phase contrast, epifluorescence (GFP) and overlayed (merge) images have been shown. Arrows highlight small cellular bodies which lack EGFP fluorescence signal.

Overexpressed Rcy-EGFP exhibited the most complex localisation pattern of all three fusion proteins. Within *R.leg-rcy-egfp* samples, cells with morphologies which more closely resembled a typical rod-shaped *R. leguminosarum* cell displayed clusters of Rcy-EGFP foci at each end of the cell (i.e. bipolar localisation) (Figure 5.10). By 24 hours post induction, these foci often appeared to cover the entire region of curvature at the ends of the cell. In rounded or 'swollen' cells, multiple foci were observed around the perimeter of the cell (Figure 5.10, hollow arrowheads). Generally, the more spherical the cell, the more evenly distributed these foci were around the cell perimeter. Within the samples isolated at 6 hours post-induction, a small number of rod-shaped bacteria with only a single Rcy-EGFP foci at each end of the cell were observed (Figure 5.10, asterisk), suggesting a potential association between altered cell morphology and

the number of Rcy-EGFP foci within the cell. This association between the number of foci and cell morphology was supported by the observation of more complex Rcy-EGFP localisation patterns within cells which deviated from the normal rod-shaped morphology of *R. leguminosarum* (as described above). As observed with PopZ-EGFP overexpression, a few cells did display weak fluorescence signal or appeared to lack Rcy-EGFP altogether. However, unlike *R.leg-popZ-egfp* cells where only small, spherical cells seemingly lacked Pop-EGFP signal, there was no clear correlation between the size of *R.leg-rcy-egfp* cells and the intensity of their Rcy-EGFP signal.

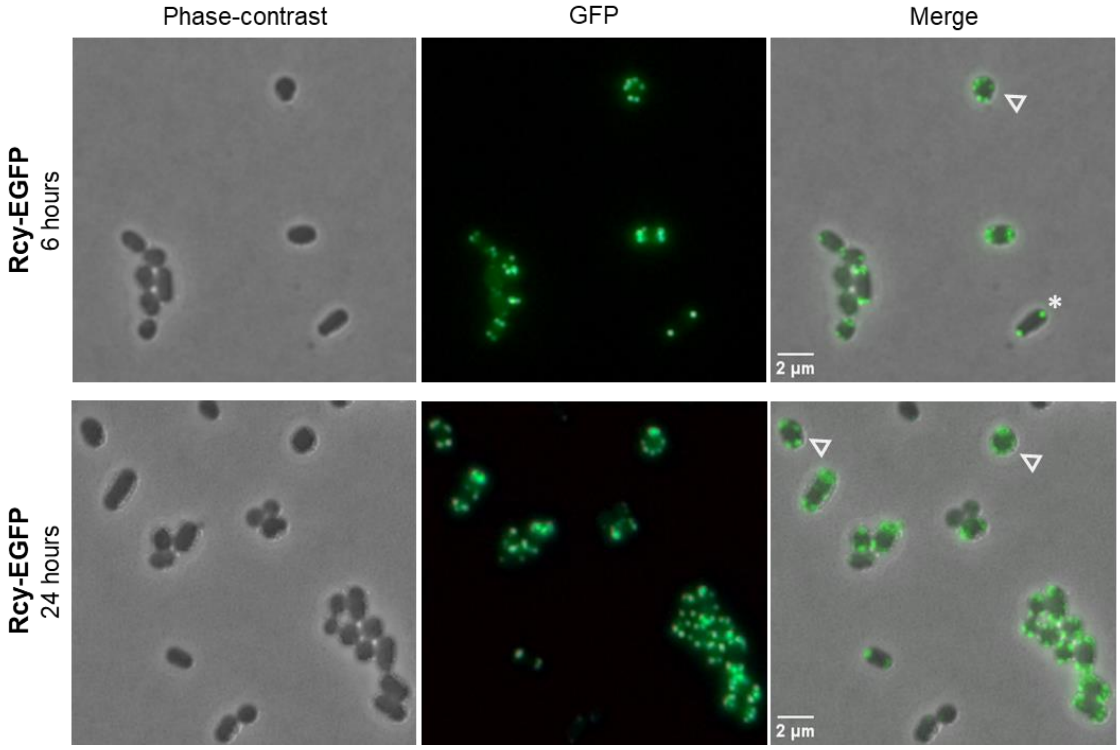


Figure 5.10. Overexpression localisation of Rcy-EGFP within *R. leguminosarum*. An *R. leguminosarum* strain carrying pIND6-*rcy-egfp* (*R.leg-rcy-egfp*) was grown in TY media in the presence of kanamycin. At 20 hours, overexpression of Rcy-EGFP was induced with the addition of 1 mM IPTG to the growth media. Culture samples were isolated at 6 and 24 hours post induction and viewed by epifluorescence microscopy (GFP filter set) to assess Rcy-EGFP localisation. Phase contrast, epifluorescence (GFP) and overlayed (merge) images have been shown. Hollow arrowheads highlight spherical cells with clusters of Rcy-EGFP foci around the cell perimeter. Asterisk marks a rod-shaped cell with one Rcy-EGFP focus at each end of the cell.

5.4 Protein localisation during low-level overexpression

To understand how the overexpression of ParA1, PopZ and Rcy gave rise to the impacts on cell morphology and population growth previously observed within Chapter Four, it was important to identify how the localisation of these proteins during overexpression differed from their native localisation. As the EGFP fusion proteins were encoded by plasmid-borne genes (driven by the IPTG-inducible $P_{A1/04/03}$ promoter), true native levels of protein expression could not be achieved using this system. Nevertheless, lower levels of overexpression of the EGFP fusion proteins from these plasmids may still facilitate the localisation of these proteins under 'normal' conditions.

5.4.1 Induction of overexpression with low IPTG concentrations

To attempt to mimic 'native' localisation, the expression of EGFP fusion proteins was induced with a range of low IPTG concentrations (0.0025, 0.005, 0.01 and 0.02 mM) to identify an expression level where EGFP signal could be detected, but no clear overexpression phenotypes were observed (i.e. cells retained a rod-shaped morphology). Induction of ParA1-EGFP expression with 0.0025 mM IPTG resulted heterogeneous expression of ParA1-EGFP within the *R.leg-parA1-egfp* population, ranging from strong whole-cell signal to cells where the expression of this protein was not detected (Figure 5.11). Interestingly, unlike the overexpression of PopZ-EGFP where smaller cellular bodies typically lacked fluorescence signal (Figure 5.9), ParA1-EGFP was detected even within very small cellular bodies. Despite induction of ParA1-EGFP expression with very little IPTG (0.0025 mM), *R.leg-parA1-egfp* cells still appeared to deviate from their expected morphology (most noticeable at 24 hours post induction), indicating overexpression of this protein. As seen following strong overexpression of this protein with 1 mM IPTG (Figure 5.8), ParA1-EGFP foci were observed within several cells but again did not appear to display a uniform localisation pattern. One possible explanation of the inconsistent presence and localisation of these foci could be the formation of protein aggregates or inclusion bodies in response to the overexpression of this fusion protein (*441*, *442*). Due to their increased density, inclusion bodies within bacterial cells can often be identified by phase contrast microscopy. Within these samples, no convincing evidence of inclusion bodies was observed (Figure 5.11). Ultimately, the localisation of ParA1-EGFP at this expression level was deemed inconclusive, particularly at 24 hours post induction, and unlikely to be an accurate reflection of native localisation of ParA1.

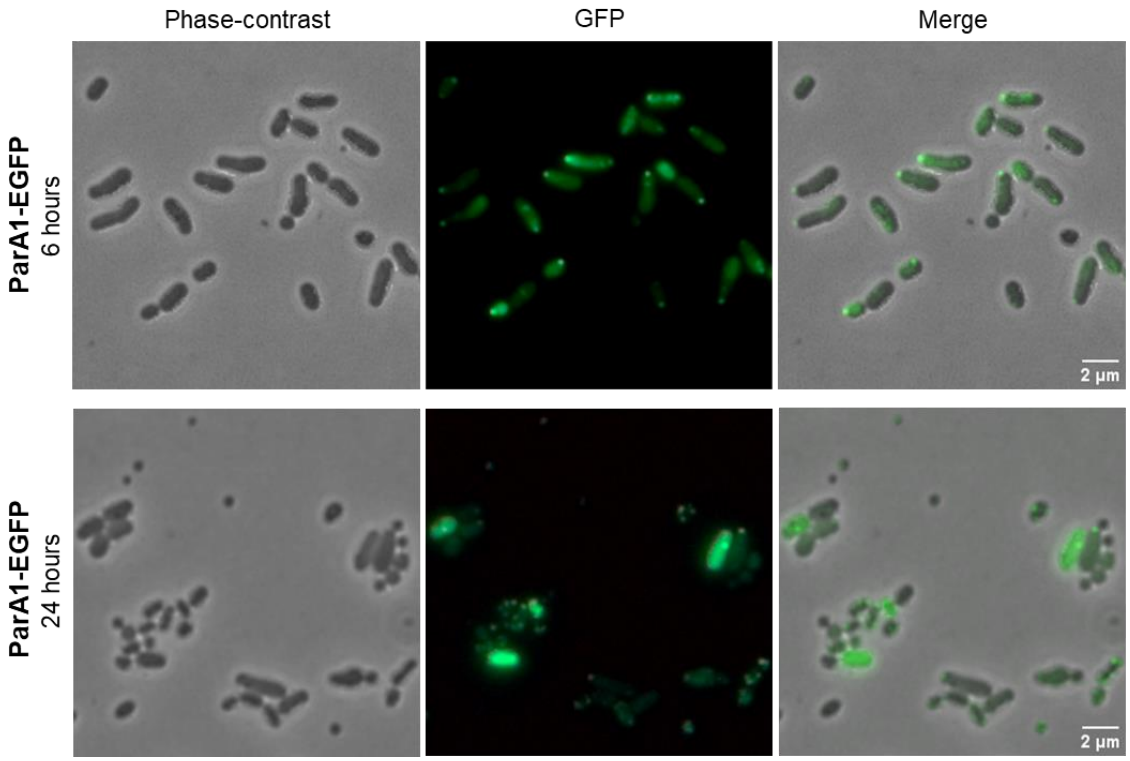


Figure 5.11. Localisation of ParA1-EGFP following induction with 0.0025 mM IPTG. The ParA1-EGFP expression strain *R.leg-parA1-egfp* was cultured in TY media in the presence of kanamycin. At 20 hours, expression of ParA1-EGFP was induced with the addition of a final concentration of 0.0025 mM IPTG to the growth media. Bacterial samples were isolated at 6 and 24 hours post induction and viewed using epifluorescence microscopy to assess ParA1-EGFP localisation. Phase contrast, epifluorescence (GFP) and overlayed (merge) images have been shown.

The induction of PopZ-EGFP expression with 0.0025 mM and 0.005 mM IPTG resulted in very weak fluorescence signal within *R.leg-popZ-egfp* cells, therefore induction with a final IPTG concentration of 0.01 mM IPTG was used for these localisation studies. Consistent with observations following higher levels of overexpression (Figure 5.9), PopZ-EGFP displayed almost exclusively unipolar localisation following induction of overexpression with 1 mM IPTG (Figure 5.12). Again, the localisation of these foci appeared to be correlated with tapered cell poles, which may be suggestive of PopZ-EGFP localisation at the growth pole. A small number of *R.leg-popZ-egfp* cells showed PopZ-EGFP localisation at both poles (bipolar localisation) (Figure 5.12, hollow arrows), as well as a few smaller cells with no signal (also observed in Figure 5.9). Potentially one of the most interesting observations within these samples was the presence of pairs of potential daughter cells which appeared to be undergoing (or had recently undergone)

cellular division which displayed only one PopZ-EGFP foci within a single cell (Figure 5.12, solid arrows). In these cells, PopZ-EGFP was localised at the tapered pole opposite the site of division. During cell division in unipolar-growing bacteria, the existing growth pole (GP) transitions into a non-growing old pole (OP) once the mature cell is of the correct length for cell division. Following cell division, new growth poles are then formed at the division sites of the two daughter cells (*152*). Therefore, these PopZ-EGFP foci were likely to have been located within either transitioning (GP to OP) or newly-defined old poles during and shortly after cell division.

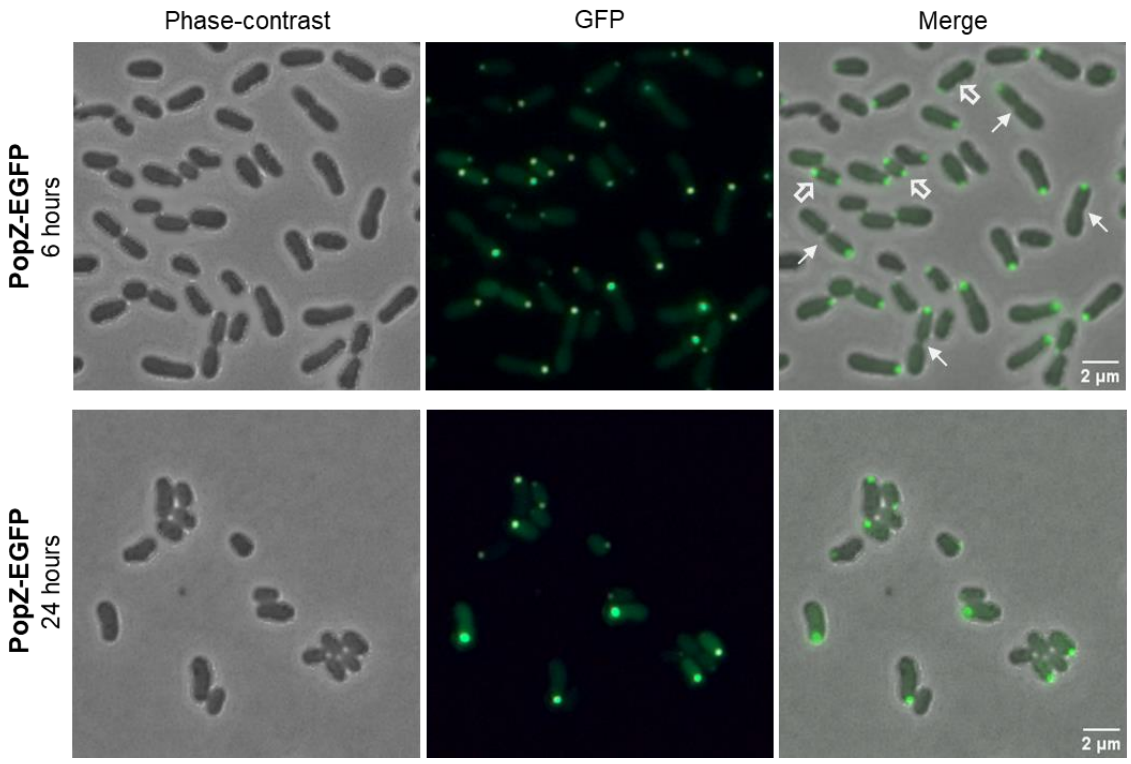


Figure 5.12. Localisation of PopZ-EGFP following induction with 0.01 mM IPTG. The PopZ-EGFP expression strain *R.leg-popZ-egfp* was cultured in TY media in the presence of kanamycin. At 20 hours, expression of PopZ-EGFP was induced with the addition of a final concentration of 0.01 mM IPTG to the growth media. Bacterial samples were isolated at 6 and 24 hours post induction and viewed using epifluorescence microscopy to assess PopZ-EGFP localisation. Phase contrast, epifluorescence (GFP) and overlayed (merge) images have been shown. Hollow arrowheads highlight cells displaying bipolar localisation of PopZ-EGFP. Solid arrows highlight cells which appear to be undergoing, or have recently undergone, cellular division.

As with ParA1-EGFP, the induction of Rcy-EGFP expression with 0.0025 mM IPTG was sufficient to give a clear fluorescence signal (Figure 5.13). However, Rcy-EGFP signal was not detected in every cell within the *R.leg-rcy-egfp* population. Although notably less severe than those

observed with strong overexpression (1 mM IPTG), morphological defects were still observed following induction with only 0.0025 mM IPTG. At 6 hours post induction, unipolar localisation of Rcy-EGFP foci was observed within several cells. However, due to alterations in cell morphology as a result of Rcy-EGFP overexpression, it is difficult to predict which pole these foci may have been located at. Unipolar localisation of Rcy-EGFP was not observed following strong overexpression (Section 5.3, Figure 5.10), suggesting that this localisation pattern could be a more accurate reflection of the native localisation of Rcy. By 24 hours post induction, an array of Rcy-EGFP localisation patterns was observed. These localisation patterns included both unipolar and bipolar localisation, as well as the presence of irregular-shaped cells with foci located at the end of each cellular protrusion. In some instances, these irregular shaped cells contained more than one Rcy-EGFP focus within a single protrusion (Figure 5.13, hollow arrowheads).

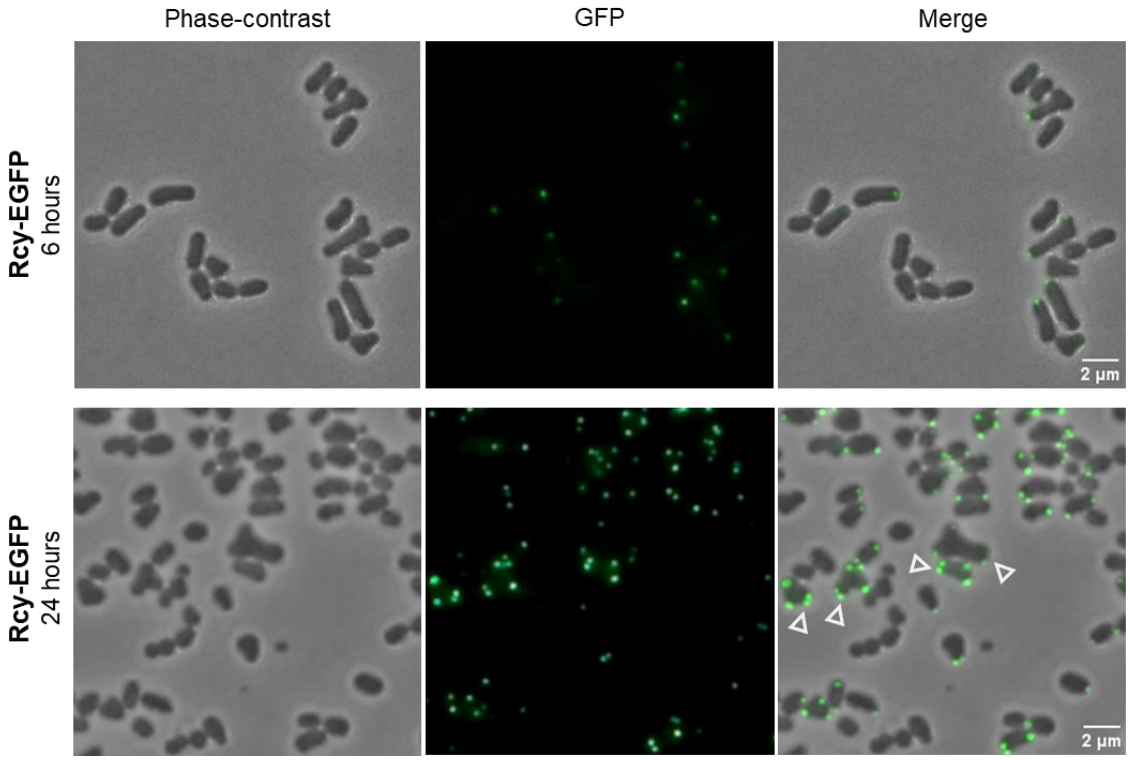


Figure 5.13. Localisation of Rcy-EGFP following induction with 0.0025 mM IPTG. The Rcy-EGFP expression strain *R.leg-rcy-egfp* was cultured in TY media in the presence of kanamycin. At 20 hours, expression of Rcy-EGFP was induced with the addition of a final concentration of 0.0025 mM IPTG to the growth media. Bacterial samples were isolated at 6 and 24 hours post induction and viewed using epifluorescence microscopy to assess Rcy-EGFP localisation. Phase contrast, epifluorescence (GFP) and overlayed (merge) images have been shown. Hollow arrowheads highlight cells with an irregular morphology where all protrusions of the cell possess at least one Rcy-EGFP focus.

5.4.2 Localisation of EGFP fusion proteins following leaky expression

Within Section 5.4.1, it was observed that the induction of ParA1-, PopZ- and Rcy-EGFP expression even with very low IPTG concentrations gave rise to cells which deviated from the expected rod morphology of *R. leguminosarum*, presumably as a direct result of the overexpression of these proteins. This was particularly noticeable for ParA1-EGFP an Rcy-EGFP expression, where induction with as little as 0.0025 mM IPTG resulted in distinct morphological defects (Figures 5.11 and 5.13 respectively) consistent with those observed within Chapter Four. Instead of inducing expression with even lower IPTG concentrations, the possibility of leaky expression of pIND6-*parA1-egfp*, pIND6-*popZ-egfp* and pIND6-*rcy-egfp* within *R. leguminosarum* was investigated. Detectable leaky expression from pIND4 (the 'parental' vector used to generate pIND6) was identified within *Paracoccus denitrificans*, but not in *Rhodobacter sphaeroides* (*355*). To check for leaky expression of the pIND6-derived expression plasmids within *R. leguminosarum*, *R.leg-parA1-egfp*, *R.leg-popZ-egfp* and *R.leg-rcy-egfp* were cultured as described previously, but this time without induction. Culture samples were then taken at timepoints equivalent to 6 and 24 hours post induction (i.e. 26 and 44 hours of growth) and checked for EGFP signal. At 26 hours, unipolar localisation of ParA1-EGFP was observed within many cells (Figure 5.14). In a small number of cells, bipolar localisation of ParA1-EGFP foci was observed (Figure 5.14, arrowheads). In these cells, weak ParA1-EGFP signal was also detected throughout the cytoplasm. The presence of single ParA1-EGFP foci amongst whole-cell cytoplasmic signal was concurrent with the previously observed localisation of this protein following induction with IPTG. However, bipolar localisation of ParA1-EGFP foci was not typically observed following induction with IPTG (Figures 5.8 and 5.11). By 44 hours, ParA1-EGFP signal appeared to have dissipated. Fluorescence foci were less distinct, and the localisation of ParA1-EGFP could not clearly be determined.

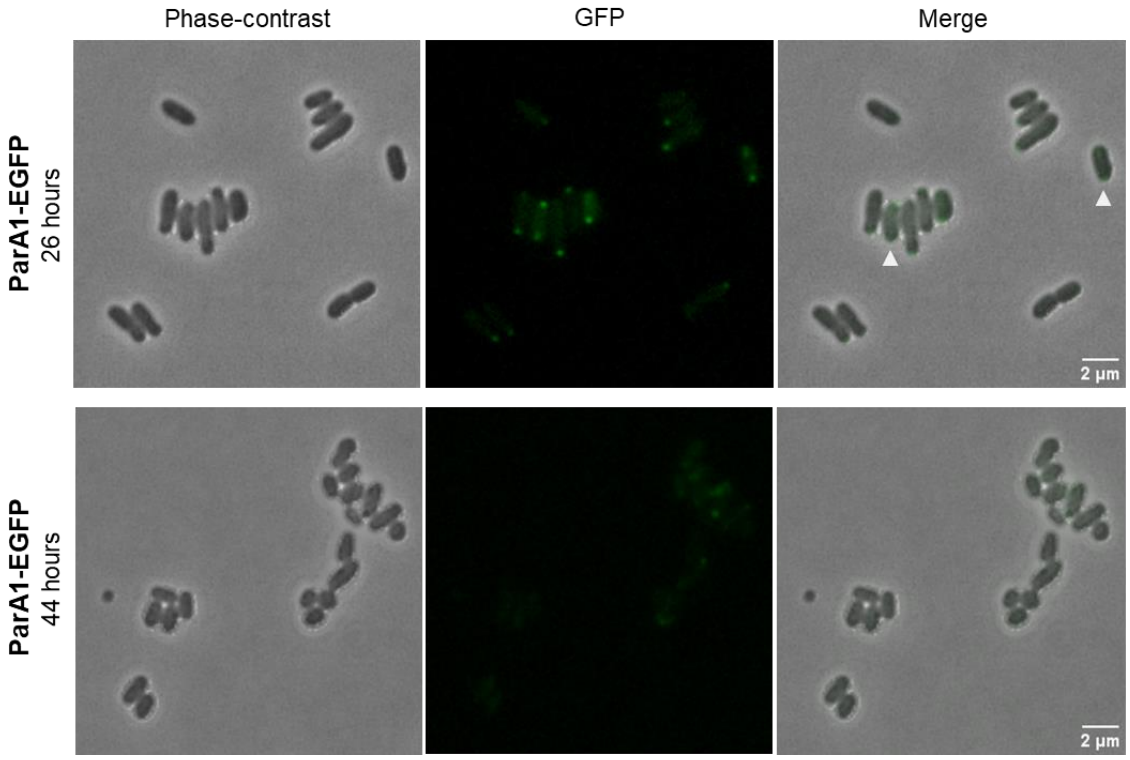


Figure 5.14. Leaky expression of ParA1-EGFP from pIND6-*parA1-egfp* in *R. leguminosarum*. *R.leg-parA1-egfp* was cultured in TY media in the presence of kanamycin to maintain selection of pIND6-*parA1-egfp*. To assess whether leaky expression of ParA1-EGFP occurred from this plasmid, cultures were not induced with IPTG. Bacterial samples were isolated at 26 and 44 hours (equivalent timepoints to 6 and 24 hours post induction). Epifluorescence microscopy was used to assess the presence and localisation of ParA1-EGFP. Phase contrast, epifluorescence (GFP) and overlayed (merge) images have been shown. Where detected, most cells exhibited unipolar ParA1-EGFP localisation. Solid arrowheads highlight cells with bipolar localisation of ParA1-EGFP foci.

Minimal evidence of leaky expression from pIND6-*popZ-egfp* within *R. leguminosarum* was observed, with almost no detectable PopZ-EGFP signal identified at either timepoint. Figure 5.15 shows potential polar localisation of PopZ-EGFP within a single cell sampled at 44 hours. However, weak emission signal made the localisation of this fluorescence focus almost entirely indistinguishable within the merged image. Given that induction of PopZ-EGFP expression with 0.0025 and 0.005 mM IPTG also resulted in weak EGFP signal within cells (data not shown), it is not surprising that leaky expression of this protein did not result in detectable EGFP signal.

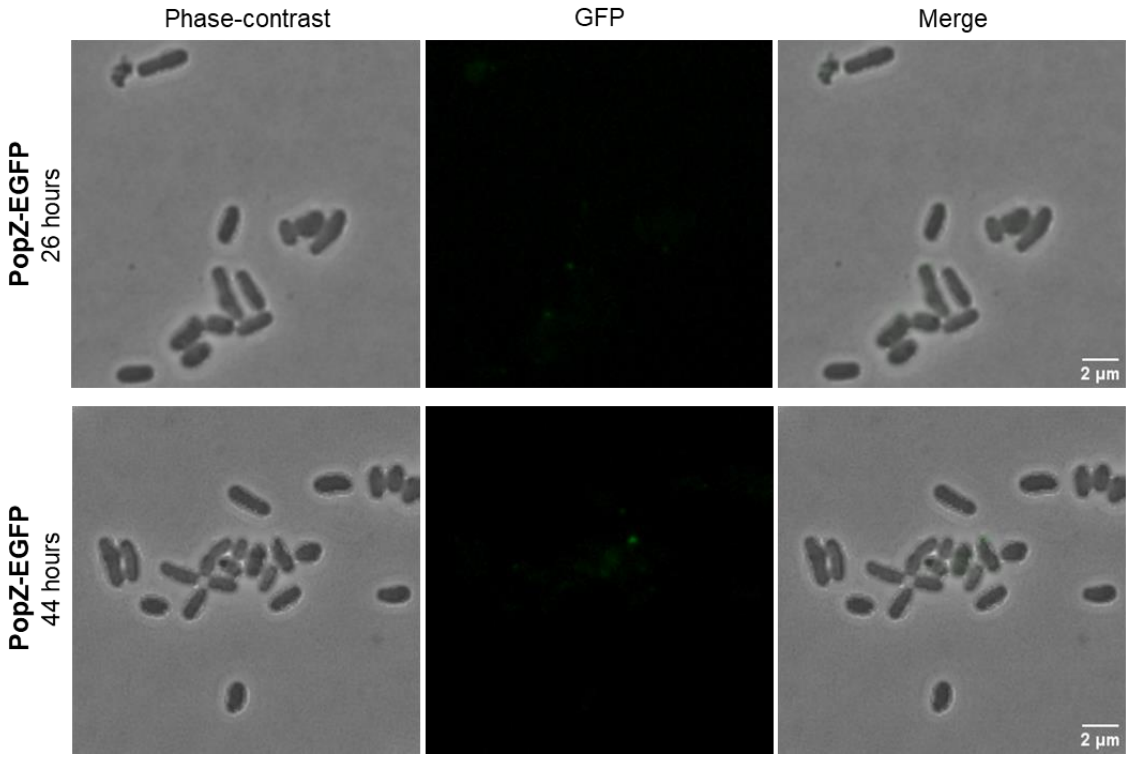


Figure 5.15. Leaky expression of PopZ-EGFP from pIND6-*popZ-egfp* within *R. leguminosarum*. *R.leg-popZ-egfp* was cultured in TY media in the presence of kanamycin to maintain selection of pIND6-*popZ-egfp*. To assess whether leaky expression of PopZ-EGFP occurred, cultures were not induced with IPTG. Bacterial samples were isolated at 26 and 44 hours (equivalent timepoints to 6 and 24 hours post induction). Epifluorescence microscopy was used to assess the presence and localisation of PopZ-EGFP. Phase contrast, epifluorescence (GFP) and overlayed (merge) images have been shown.

Leaky expression of Rcy-EGFP from pIND6-*rcy-egfp* gave the strongest fluorescence signal of all three fusion proteins, most noticeable at 44 hours (Figure 5.16). Fluorescence signal was not detected within every cell, but where detected, Rcy-EGFP typically displayed unipolar localisation. While most cells within this *R.leg-rcy-egfp* culture were rod-shaped, some irregular cell morphologies were observed indicating that overexpression of Rcy-EGFP may still have been an issue within these cultures. Figure 5.16 shows one of these irregular shaped cells, highlighted with a hollow arrowhead. Within this cell, Rcy-EGFP foci were observed within each protrusion of the cell. This localisation pattern was consistent with previous localisation patterns of overexpressed Rcy-EGFP (Figures 5.10 and 5.13). A slight right shift in the EGFP signal image as a result of a small shift in the focal plane resulted in a few foci falsely appearing to be located on the outer edges of cells within the merged image. Widespread unipolar localisation of Rcy-EGFP was only observed within uninduced cultures (Figure 5.16), and samples isolated at 6 hours post

induction with 0.0025 mM IPTG (Figure 5.13). Cells within these samples most closely resembled wildtype rod-shaped *R. leguminosarum* cells, suggesting that unipolar localisation is most likely to be the native localisation pattern of Rcy.

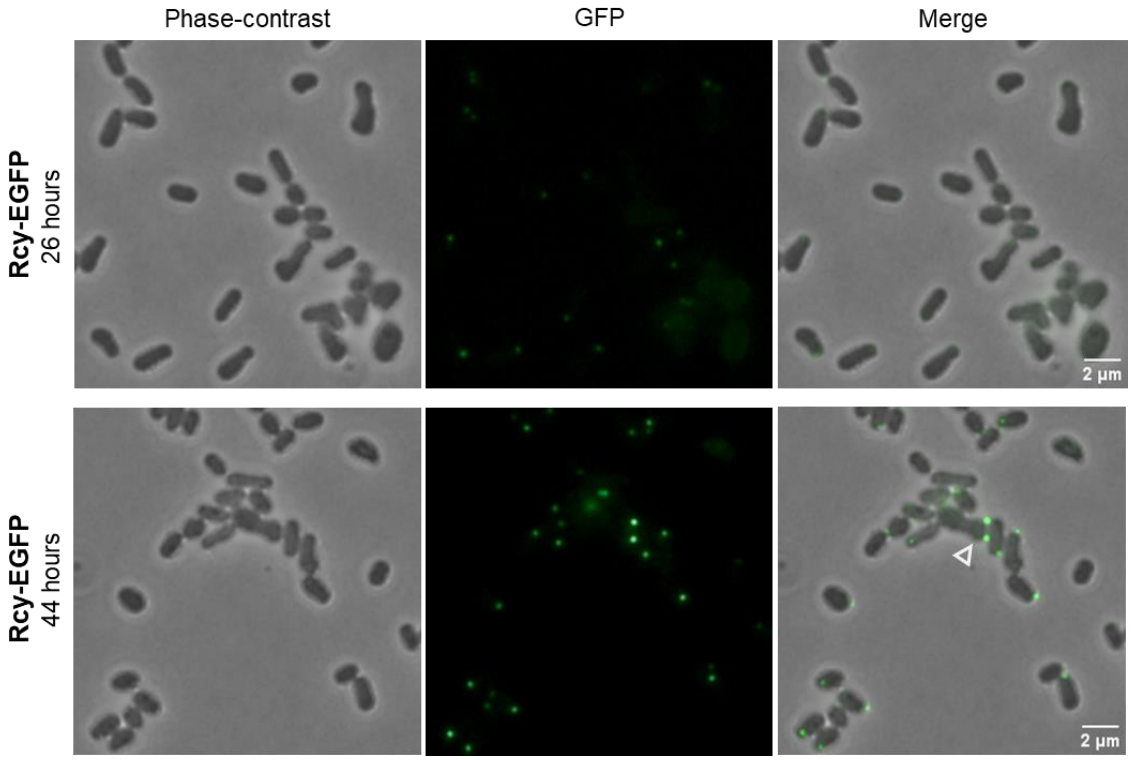


Figure 5.16. Leaky expression of Rcy-EGFP from pIND6-*rcy-egfp* within *R. leguminosarum*. *R.leg-rcy-egfp* was cultured in TY media in the presence of kanamycin to maintain selection of pIND6-*rcy-egfp*. To assess whether leaky expression of Rcy-EGFP occurred, cultures were not induced with IPTG. Bacterial samples were isolated at 26 and 44 hours (equivalent timepoints to 6 and 24 hours post induction). Epifluorescence microscopy was used to assess the presence and localisation of Rcy-EGFP. Phase contrast, epifluorescence (GFP) and overlayed (merge) images have been shown. The hollow arrowhead highlights an irregular shaped *R. leguminosarum* cell in which Rcy-EGFP localised to each end of the cell.

5.5 Identification of an intermediate Rcy overexpression phenotype

Induction of Rcy-EGFP expression with a range of low IPTG concentrations resulted in the observation of an 'intermediate' overexpression phenotype. As described in Chapter Four, strong Rcy overexpression (induced with 1 mM IPTG) typically resulted in swollen cells with irregular morphologies at 6 hours post induction and enlarged spherical cells by 24 hours post induction. However, analysis of *R.leg-rcy-egfp* cells at 6 hours post induction with 0.01 and 0.02 mM IPTG led to the identification of an overexpression phenotype not previously described within this thesis. These cells, shown in Figure 5.17, displayed a three- or occasionally four-armed branched morphology. However, these branched cells were less like the elongated Y- and V-shaped branched cells observed following ParA1 overexpression (Chapter Four, Figure 4.12), and instead more closely resembled three- or four-pointed stars with roughly equally-spaced protrusions. Rcy-EGFP foci were located at the ends of each of these protrusions. Following induction with 0.01 mM IPTG, a single focus was typically observed within each end of the cell. However, following induction with 0.02 mM IPTG, the resulting branched cells often had multiple foci at the end of each branch of the cell. These localisation patterns, in combination with the observation of numerous foci at the ends of the poles and protrusions of cells induced with 1 mM IPTG (Figure 5.10) gave rise to the conclusion that instead of forming larger foci following an increased abundance of protein (as observed with PopZ-EGFP expression), Rcy-EGFP instead forms secondary (and eventually multiple) foci within the cell pole. By 24 hours post induction, cells within cultures induced with 0.01 or 0.02 mM IPTG resembled the enlarged, spherical and irregular swollen morphologies observed previously in this chapter, with multiple foci located along regions of cell curvature (i.e. ends of irregular cells or the perimeter of spherical cells) (data not shown).

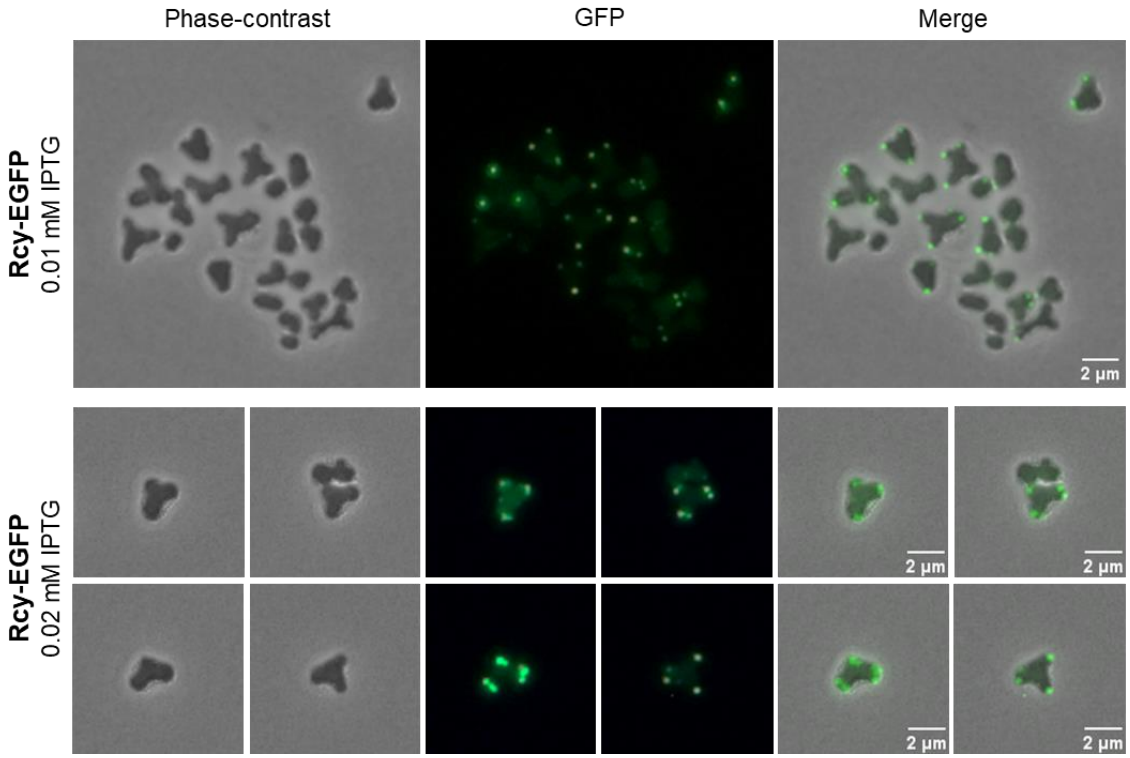


Figure 5.17. Overexpression localisation of Rcy-EGFP within *R. leguminosarum* cells displaying an intermediate branched overexpression phenotype. *R.leg-rcy-egfp* was grown in TY media in the presence of kanamycin to maintain selection of the Rcy-EGFP expression plasmid pIND6-*rcy-egfp*. At 20 hours, cultures were induced with either 0.01 or 0.02 mM IPTG. Bacterial samples were then isolated at 6 hours post induction and viewed by epifluorescence microscopy (GFP filter) to determine Rcy-EGFP localisation. Phase contrast, epifluorescence (GFP) and overlayed (merge) images have been shown.

5.6 Discussion

5.6.1 Inconsistent localisation of *R. leguminosarum* ParA1 may reflect its dynamic nature

ParA1-EGFP within *R. leguminosarum* exhibited an inconsistent localisation pattern, which appeared to differ depending on the extent of induction of protein overexpression. Following strong overexpression (1 mM IPTG), homogeneous cytoplasmic ParA1-EGFP signal was observed within almost all cells, with some cells also displaying a single, concentrated ParA1-EGFP focus within the cell (often but not always polar-localised) (Figure 5.8). Low-level overexpression of this protein (0.0025 mM IPTG) resulted in an inconsistent expression and localisation pattern within the *R. leguminosarum* population, with some cells displaying strong whole-cell ParA1-EGFP signal, others exhibiting patchy cytoplasmic localisation with polar foci, and many lacking fluorescence signal all-together (Figure 5.11). Finally, although fluorescence signal was weak and not widely detected throughout the population, leaky expression of ParA1-EGFP from the pIND6-derived expression plasmid within *R. leguminosarum* resulted in predominantly unipolar (but occasionally bipolar) localisation of ParA1-EGFP foci, with patches of weaker ParA1-EGFP signal throughout the cytoplasm (Figure 5.14).

Early localisation studies within *Caulobacter crescentus* and *Vibrio cholerae* demonstrated that ParA formed localisation foci at the new pole within *C. crescentus* or at both poles within *V. cholerae*, often surrounded by patchy or 'cloud-like' cytoplasmic localisations from the pole towards the midcell (*208*, *315*, *437*). In *V. cholerae*, localisation of ParA was demonstrated to be highly dynamic, and dependent upon the cell cycle. Younger cells were found to have foci at a single pole, and a dense patch of ParA extending from the other pole towards the midcell. Predivisonal cells displayed almost exclusively bipolar localisation, while older cells undergoing cytokinesis had polar foci with smaller patches of ParA localised near the midcell (*208*). Improved understanding of the mechanism of ParA and its localisation have yielded the hypothesis that ParA proteins exhibit nucleoid-wide localisation throughout the cell (*215*, *341*, *342*). Increased concentrations of ParA are typically associated with high-density chromosomal regions, as well as the cell poles during later stages of the cell cycle (*341*), with polar concentration gradients of ParA proposed to be maintained by direct interactions with other polar proteins such as TipN, PopZ and Scy (*183*, *206*, *314*, *315*).

The dynamic, cell cycle-dependent localisation of ParA proteins could therefore explain (at least in-part) why an inconsistent localisation pattern of ParA1-EGFP within *R. leguminosarum* was observed within this thesis. With the exception of strongly overexpressed ParA-EGFP (Figure 5.8), the localisation patterns observed within this chapter may be a reasonably-accurate representation of the localisation of ParA within *R. leguminosarum*. Various ParA1-EGFP localisation patterns observed following induction of ParA1-EGFP expression with 0.0025 mM IPTG or its leaky expression from pIND6-*parA1-egfp* (Figures 5.11 and 5.14) such as polar ParA-EGFP foci amongst patchy cytoplasmic localisations show a strong resemblance to previously-observed ParA localisation patterns within other Gram-negative model species (*208*, *315*, *437*).

The observation of polar ParA1-EGFP foci within *R. leguminosarum* and documented interactions of ParA with polarity proteins within other polar-growing species may support the hypothesis that ParA1 is associated (even if only in a cell cycle-dependent manner) with *R. leguminosarum* polar protein complexes. However, although such bodies were not clearly observed by phase contrast microscopy, it is not possible to entirely rule out the possibility that the foci observed were not insoluble protein inclusion bodies. Unfortunately, the ParA-EGFP localisation studies performed within this chapter ultimately offer little explanation as to precisely how the overexpression of ParA produced the growth and morphological defects observed throughout Chapter Four. Although localisation and expression patterns did vary with inducer concentration, differing localisation of ParA did not appear to correlate with cell morphology or size.

5.6.2 Consistent unipolar localisation of PopZ supports its involvement during unipolar growth of *Rhizobium*

Regardless of the extent of its overexpression, PopZ displayed consistent unipolar localisation within *R. leguminosarum*, robustly supporting its documented identity as a pole organising protein within the *Rhizobiaceae* (*273*, *325*, *326*). Unlike ParA1-EGFP and Rcy-EGFP, where varying the extent of overexpression altered the localisation pattern of these proteins, PopZ-EGFP localisation was seemingly unaffected by its abundance within the cell. This consistent localisation may therefore explain why the overexpression of this protein did not have such a drastic impact on cell morphology or the growth of *R. leguminosarum* cultures (as discussed in Chapter Four). Interestingly, the concentration of IPTG required to induce sufficient PopZ-EGFP

expression to give a clear fluorescence signal was slightly higher than that of ParA1-EGFP and Rcy-EGFP, suggesting that *Rhizobium* may be able to better-regulate the concentration of PopZ within the cell, possibly through targeted protein degradation or differences in protein turnover (*443*, *444*).

By far the biggest limitation of the localisation studies performed within this thesis was the inability to accurately distinguish the growth pole from the old pole. As mentioned within Section 5.3, PopZ localisation may have been associated with the growth pole of *R. leguminosarum* (determined based on cell morphology), but as the overexpression of this protein was previously demonstrated to influence cell morphology, this hypothesis can by no means be taken as fact. Without supporting evidence such as live cell imaging or appropriate cell stains to enable the determination of pole identity (e.g. FM 4-64 staining, which has been demonstrated to stain older halves of *Agrobacterium* cells (*127*)), it is impossible to confirm which pole PopZ localises to within *R. leguminosarum*. Consequently, the involvement of PopZ specifically within the growth pole polar complex of *R. leguminosarum* remains unconfirmed.

Observed following lower-level overexpression of PopZ was the presence of pairs of daughter cells in which PopZ-EGFP foci appeared to have been located within the transitioning or newly-defined old pole of a single daughter cell (Figure 5.12). During cell division in unipolar-growing bacteria, the existing growth pole (GP) transitions into an old pole (OP), and new growth poles are formed at the division sites of the daughter cells (*131*, *152*). During this transitionary process within *Agrobacterium*, PopZ rapidly relocates from the old growth pole to the new growth poles of both daughter cells, and for a brief window of time exhibits bipolar localisation (*273*). This relocation of PopZ was not clearly observed within the highlighted pairs of *R. leguminosarum* daughter cells observed within Figure 5.12, and instead PopZ-EGFP may have temporarily remained at the old pole following cell division. That being said, a small number of isolated cells within this same population did display bipolar localisation of PopZ-EGFP. It would therefore be of interest to know whether these cells had recently undergone division and were consequently undergoing a transitionary phase, or whether bipolar localisation of PopZ was an infrequent phenotype of its overexpression.

5.6.3 Rcy is a unipolar protein which may dictate the site of cell growth

Initially, the overexpression phenotypes observed within Chapter Four indicated that Rcy may be more strongly linked to controlling *R. leguminosarum* morphology than growth. However, the localisation studies performed within this chapter suggest that the observed changes in cell morphology following Rcy overexpression were likely the direct result of uncoordinated polar growth. Localisation studies following low-level overexpression (i.e. leaky expression) of Rcy-EGFP indicated that under normal conditions, Rcy is likely to exhibit unipolar localisation within *R. leguminosarum* (Figure 5.16). However, when overexpressed, Rcy forms secondary and eventually multiple foci within the cell. Lower-level overexpression resulted in the formation of irregular-shaped cells, with Rcy-EGFP foci located at each end of the cell (Figures 5.13), while strong overexpression resulted in the formation of coccoid cells with multiple foci around the perimeter of the cell (Figure 5.10). It is not too clear whether secondary foci are first formed at the same pole as the initial Rcy focus, or instead form at the opposite pole, but the presence of rod-shaped and branched cells with a single Rcy-EGFP focus within each protrusion (Figures 5.13 & 5.17) suggest that secondary Rcy foci are first formed at the opposite pole, with subsequent foci formed adjacent to the existing foci.

Within *R. leguminosarum* cells with an irregular non-coccoid morphology, Rcy-EGFP foci were located at the end of each cellular protrusion, suggesting that Rcy foci may determine the site of cell growth. Within *Streptomyces* and *Agrobacterium* respectively, Scy and GPR are proposed to function as polar hub proteins which organise polar growth. The overexpression of these proteins result in the formation of multiple polarity centres (Scy) and multiple ectopic growth poles (GPR) (*182*, *278*). Interestingly, the deletion of *gpr* resulted in defective cell division. *Agrobacterium* cells lacking GPR grew in multiple directions, generating quartets of daughter cells which remained attached (*278*). This morphology is strikingly similar to the branched cells observed following Rcy-EGFP overexpression (Figure 5.17), indicating that Rcy and GPR may exhibit similar biological functions. Within *Rhizobium leguminosarum* the formation of additional Rcy foci could redirect growth within the cell. Localisation of a second foci within a single pole may split growth machinery across both foci, ultimately resulting in branched polar growth. This hypothesis of cellular branching following secondary Rcy foci formation aligns with the observation of *de novo* tip formation and subsequent hyphal branching following the overproduction of Scy within *Streptomyces*, highlighting that Rcy likely functions as an organiser of polar growth much like Scy (*182*). Following strong overexpression, the localisation of multiple

Rcy foci around the cell perimeter appeared to result in attempted radial growth of *R. leguminosarum* as opposed to unipolar growth, generating enlarged spherical cells. As the expression of all other growth-related proteins such as cell wall synthesis machinery remained unchanged within these cells, it is probable that new or existing growth poles could not be maintained, which may explain the wide diversity of irregular Rcy overexpression phenotypes observed throughout this chapter.

Finally, during localisation studies of GPR, Zupan *et al.* proposed that the generation of C-terminal GFP fusions altered the function of GPR (*278*, *279*). However, as far as the data within this thesis is concerned, this did not appear to be the case for C-terminal EGFP tagging of Rcy. The overexpression of Rcy-EGFP resulted in cell morphologies which were consistent with non-EGFP-tagged Rcy overexpression i.e. enlarged, spherical cells (Figures 4.12 and 5.10), strongly suggesting that the EGFP tag did not impair Rcy function. It may of course be possible that these observations instead indicate differences in function between the C-terminal domains of GPR and Rcy, and that native copies of Rcy could have compensated for the impaired function of plasmid-borne C-terminal EGFP-tagged Rcy proteins. Further investigation into differences in the localisation and function of C- and N-terminal tagged Rcy would therefore be recommended to rule out possible impacts of the introduction of C-terminal fluorescent protein tags on the function of this protein.

Chapter Six - General Discussion and Future Perspectives

Taken together, the data presented throughout this thesis supports the initial hypothesis that ParA1, PopZ and Rcy are required for the coordination of unipolar growth and division within *Rhizobium leguminosarum*. Within *R. leguminosarum*, ParA1 and PopZ are likely primarily involved in the spatial and/or temporal coordination of cell division, while Rcy is instead proposed to determine the site of PG synthesis during polar growth. Consistent polar localisation of both PopZ and Rcy demonstrated that these proteins are indeed polarity proteins, with both proposed to localise to the *Rhizobium* growth pole. Distinct and non-overlapping morphological phenotypes following the overexpression of each of these proteins highlights that ParA1, PopZ and Rcy each serve a unique role during the regulation of *R. leguminosarum* growth. Furthermore, the observation of detrimental growth phenotypes following the overexpression of these proteins during lag-phase suggests that all three proteins are likely required for polar and/or whole-cell organisation in preparation for active cell growth and division.

6.1 ParA1 is required for the coordination of cell division

The overexpression of ParA1 was associated with one of two possible phenotypes: pronounced morphologically asymmetric division, yielding daughter cells of distinctly different sizes; and occasional Y- and V-shaped cell branching (see Section 4.5). Similar branched morphologies have been observed within *Sinorhizobium* when cell division is blocked, suggesting that the overexpression of ParA1 may disrupt cell division (*198*, *445*, *446*). The inhibition of DNA replication, which prevents cell division early on in the cell cycle, also results in the formation of Y-shaped branched morphologies within *Sinorhizobium* (*447*). Possible impacts of ParA1 misregulation (overexpression) on DNA replication and/or segregation within *R. leguminosarum*, leading to the formation of Y- and V-branched cells, is consistent with the widespread role of ParA proteins as chromosome partitioning proteins (*162*, *203*). Additionally, overexpression of ParA1 during exponential growth phase (i.e. when cells are rapidly dividing) significantly impaired the growth of *R. leguminosarum* cultures, slowing growth and reducing the maximum OD_{600} reached (Figure 4.6), further supporting the conclusion that ParA1 is primarily involved (either directly or indirectly) in the regulation of cell division.

Within *R. leguminosarum*, ParA1 typically exhibited diffuse whole-cell localisation, with cells often, but not always, containing a polar ParA1 focus. As already discussed within Section 5.6.1, the seemingly inconsistent localisation pattern of ParA1 within *R. leguminosarum* may simply reflect its dynamic nature as a chromosome partitioning ATPase. Although ParA dimers can bind non-specifically throughout the nucleoid, ParA proteins often exhibit polar concentration gradients in a cell-cycle dependent manner (i.e. during chromosome segregation) (*214*, *215*, *341*). Additionally, polar concentration gradients of ParA proteins are proposed to be maintained by interactions with various polarity proteins including Scy, PopZ and TipN (*183*, *206*, *314*, *315*). Hence, the polar localisations of ParA1 observed within this thesis may indicate the establishment of similar polarity protein-mediated polar concentration gradients.

6.2 The polarity determinant PopZ may coordinate spatial regulation of cell division

One of the most striking features of PopZ was its consistent unipolar localisation. Unipolar localisations of PopZ were typically associated with tapered cell poles, which in the case of *Agrobacterium* are indicative of growth poles, suggesting that PopZ is likely a new pole marker within both *Agrobacterium* and *Rhizobium* (*127*, *152*, *273*, *274*). Consistent unipolar localisation of PopZ-EGFP foci, even when strongly overexpressed, could also be indicative of polar matrix formation by *Rhizobium leguminosarum* PopZ ($PopZ_{Rl}$). Comparison of PopZ-EGFP localisation following induction with 0.1 mM vs 1 mM IPTG (Figures 5.12 and 5.9 respectively) highlight that the localisation pattern of $PopZ_{Rl}$ does not change with an increase in protein abundance, and instead fluorescence foci simply increase in size and intensity. It is widely accepted that *Caulobacter* PopZ ($PopZ_{Cc}$) self-assembles into a disordered matrix at the cell poles, and increasing $PopZ_{Cc}$ concentration has been shown to result in an expansion of the existing PopZ matrix instead of the formation of new assemblies (*311*, *319*). Therefore, the observation of larger/more intense PopZ-EGFP foci following overexpression indicates that matrix formation or PopZ polymerisation may be a conserved function of both $PopZ_{Cc}$ and $PopZ_{Rl}$.

Although the overexpression of PopZ had the least drastic impact on *Rhizobium* morphology and growth, one noteworthy phenotype observed following PopZ overexpression was the formation of minicells. Minicell formation is typically associated with misplacement of the Z ring towards the poles of the cell instead of the midcell, leading to constriction of the cell at the cell pole (*322*,

325, *326*). The formation of such cells, which often lack DNA, has also been observed within *Agrobacterium* following the deletion of *popZ* (*325*, *326*). Similar to the misregulation of *Agrobacterium* PopZ through its deletion, overexpression of $PopZ_{Rl}$ may also result in cell division defects as a result of misplacement or instability of the Z ring. Further similarities between $PopZ_{At}$ and $PopZ_{Rl}$ include the frequency of minicell formation itself. Within this thesis, minicell formation was a common, but not necessarily dominant, phenotype of PopZ overexpression. Similarly, although minicell formation was observed *in Agrobacterium popZ* deletion mutants, most cells lacking $PopZ_{At}$ still divided around the midcell (*325*). In light of the characterisation of $PopZ_{Rl}$ within this thesis, which highlights possible functional conservation of $PopZ_{At}$ and $PopZ_{Rl}$, the involvement of $PopZ_{Rl}$ in coordinating cell division as opposed to polar PG synthesis is not wholly surprising. Despite its role as a marker of the *Agrobacterium* growth pole, it has been shown that $PopZ_{At}$ is not required for the localisation of polar peptidoglycan biosynthesis, and in its absence unipolar PG synthesis continues with minimal impact (*325*).

6.3 The polar protein Rcy determines the site of peptidoglycan synthesis

Of all three proteins characterised within this thesis, Rcy displayed by far the most complex localisation pattern and associated morphological phenotypes following its overexpression. These phenotypes have been described thoroughly within Section 4.5 and throughout Chapter Five. In short, Rcy overexpression was associated with the formation of swollen, irregularly shaped cells, which eventually displayed drastically enlarged, spherical morphologies. The severity of these morphological phenotypes was tightly correlated with the number of Rcy foci observed within the cell, with rod-like or slightly branched cells possessing only one or two foci at each pole/protrusion of the cell, and irregular or spherical cells displaying many foci around the perimeter of the cell (Figure 5.10). Importantly, leaky expression of Rcy-EGFP from the pIND6-derived expression plasmid resulted in the observation of rod-shaped cells with only a single unipolar focus of Rcy-EGFP (Figure 5.16), suggesting that unipolar localisation may in fact be the normal localisation pattern of Rcy within *R. leguminosarum*. However, subtle morphological phenotypes arising from the slight overexpression of Rcy even under leaky expression meant that pole identity was difficult to determine, as very few cell poles appeared tapered. Yet, given the clear correlation between changes in cell morphology and the number of Rcy foci, it was concluded that Rcy most likely determines the site of PG synthesis within *R. leguminosarum* and is therefore likely to localise to the 'growing' pole.

In contrast to the Y- and V-shaped branched morphologies observed following overexpression of ParA1, branched cells resulting from the overexpression of Rcy instead displayed roughly equally-spaced branches which arose from the centre of the cell, more closely resembling X- or T-shaped morphologies. Such morphologies are indicative of a redirection of polar PG synthesis to the midcell, which can also be observed when cell division is blocked at a later stage, and have been observed within *Agrobacterium* following these scenarios (*130*). The observation of unique branched morphologies following the overexpression of ParA1 and Rcy supports the conclusion that these proteins are required for the coordination of (early) cell division and the direction of polar PG synthesis respectively. The observation of these X- and T-shaped branched morphologies may also indicate that Rcy has additional or indirect role in coordinating later stages of cell division, with the irregular branched morphologies representing cells which were unable to divide.

Throughout this thesis, parallels have been drawn between Rcy and GPR. GPR exhibits growth-pole localisation, and is proposed to provide structural support for the *Agrobacterium* polar growth complex (i.e. PG synthesis machinery) (*278*, *279*). Morphological phenotypes arising from the misregulation of GPR display similarities to those observed within this thesis following the overexpression of Rcy. Overproduction of GPR results in branched cells and the formation of ectopic growth poles, while the deletion of *gpr* was associated with cell mortality, and viable cells with a deletion of *gpr* grew slowly, were coccoid in morphology, are were often enlarged (*278*). The loss of GPR also resulted in defective cell division, generating quartets of daughter cells which remained attached – a morphology similar to the branched cells observed following Rcy-EGFP overexpression (Figure 5.17) (*278*). Within coccoid *gpr*-deletion cells, PG synthesis was abnormally localised around circumference of the cell (*279*). Circumferential PG synthesis within cells lacking GPR is reminiscent of the proposed radial growth of spherical Rcy-overexpressing cells as a result of circumferential Rcy localisation (see Section 5.6.3). Similar length, near-identical predicted domain architecture, and functional similarities (i.e. direction of polar PG synthesis) strongly suggest that Rcy and GPR are indeed homologous proteins, thereby making Rcy a putative homolog of the *Sinorhizobium meliloti* protein RgsE. Within *Sinorhizobium*, RgsE overexpression results in irregular, branched cell morphologies, while its depletion results in slow-growing cells which exhibit aberrant cell division (*280*, *281*).

Of course, Rcy was initially identified as a potential 'homolog' of the *Streptomyces* intermediate filament-like protein Scy (Section 1.5.2). Scy, another notably long and predominantly α-helical protein (with high coiled-coil forming probability) forms *de novo* hyphal tips within *Streptomyces* following its overexpression, and is proposed to serve as a scaffolding protein for the TIPOC, which coordinates polar growth at the hyphal tip (*175*, *182*). The role of Scy as a molecular scaffold for the TIPOC resembles the proposed roles of GPR, RgsE and Rcy as polar growth organisers within the Hyphomicrobiales, suggesting that these proteins may form part of a widespread but poorly conserved group of large cytoskeletal scaffolding proteins which organise PG synthesis within the poles/tips of polar-growing species.

In comparison to ParA1 and PopZ, Rcy was the most sensitive to overexpression, suggesting that it is tightly regulated within wildtype cells or at least present at lower concentrations under native conditions (*430*). The observation of Rcy foci at both poles of rod-like cells (bipolar localisation) following even slight overexpression may support this theory. As membrane-anchored proteins are not readily able to migrate across the cytoplasm between cell poles, relocalisation of membrane-bound polarity determinants in transitioning cell poles often relies on their degradation at one pole, and *de novo* synthesis at the other (*278*, *448*). Rcy is predicted to be a membrane-anchored protein, therefore it too may remain at the old pole until its degradation. The presence of bipolar and ectopic Rcy foci following its overexpression could therefore reflect an overabundance of residual membrane-anchored protein, which has not been degraded fast enough as the proteins governing the proteolysis of Rcy are not present at sufficient levels within Rcy-overexpressing cells.

6.4 ParA1, PopZ, Rcy, and the *Rhizobium* polar complex

Based on the data presented and discussed throughout this thesis, it is likely that PopZ and Rcy both colocalise to the *Rhizobium leguminosarum* growth pole, possibly forming part of a cytoplasmic pole-organising complex (Figure 6.1). Within this thesis, Rcy has been proposed to play a key role in the regulation of unipolar growth by determining the site of polar PG synthesis. This role may be facilitated by the coiled-coil domains present within Rcy's predicted apolipoprotein domains. As described within Section 1.5.2 (Figure 1.15), the central length of Rcy is predicted to contain 13 overlapping or adjacent apolipoprotein domains, along with four

short coiled-coil domains which lie within these predicted apolipoprotein domains. Coiled-coil domains are often found within interacting proteins and serve as a highly conserved mechanism of protein-protein interaction (*449*). Interactions between Rcy's coiled-coil domains and coiled-coil domains within Hyphomicrobiales-specific polar peptidoglycan-modifying enzymes such as the glycosyltransferase-transpeptidase (GTase-TPase) PBP1A, L,D-transpeptidases (LDTs) and possibly other penicillin-binding proteins (PBPs) may permit the organisation and polar localisation of PG synthesis machinery, driving unipolar growth (*274*, *287*, *288*). The coiled-coil domains of Rcy may also recruit PopZ monomers to the growth pole, resulting in the formation of a PopZ matrix at this pole. Alternatively, PopZ (or other polar proteins) may be recruited by Rcy's cytoplasmic N- and C-terminal disordered domains. Of the two, the C-terminal domain is arguably a better candidate for these interactions, as the C-terminal domain of GPR has been found to be required for GPR's function (*279*). Of course, further experimentation is required to confirm whether direct interactions between Rcy and PopZ actually occur, as this was not studied within the scope of this thesis.

As discussed above (Section 6.1), polar localisations of ParA1 were occasionally observed amongst a background of whole-cell localisation. Hence, ParA1 is not necessarily a 'polar' protein, but may exhibit polar localisations in a cell cycle-dependent manner. Based on the observation of polar ParA concentration gradients within *Streptomyces* and *Caulobacter*, which are maintained either by Scy (*Streptomyces*), or the polarity determinants PopZ and TipN (*Caulobacter*) (see refs above), it could be the case that growth pole ParA1 concentrations within *R. leguminosarum* are established or maintained by either Rcy (a proposed homolog of Scy) or PopZ. A model of how Rcy, PopZ and ParA1 may interact at the growth pole of *R. leguminosarum* has been proposed in Figure 6.1.

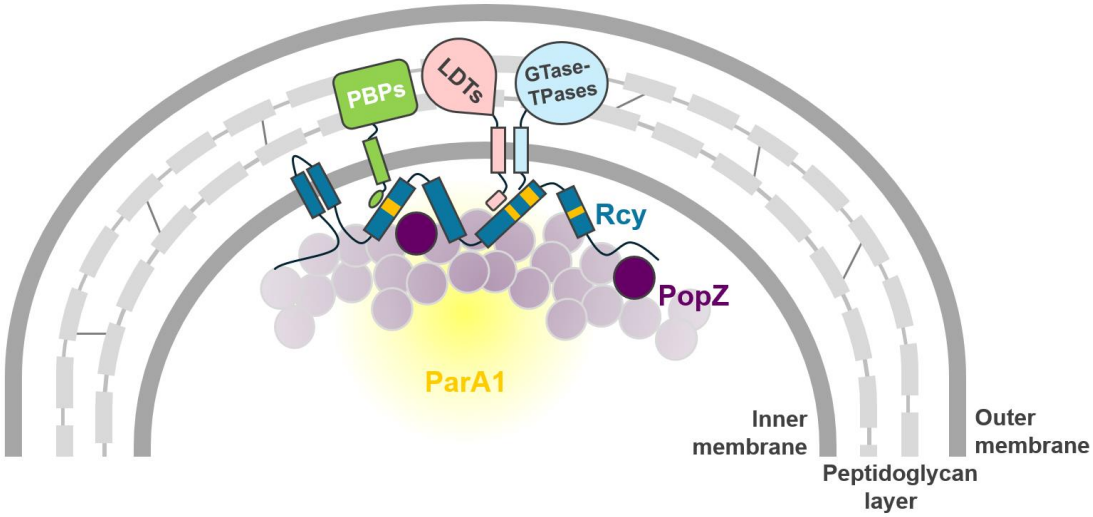


Figure 6.1. Proposed model of the *R. leguminosarum* growth pole complex. Rcy (dark blue) and PopZ (purple) may interact at the *Rhizobium* growth pole to form a polar organisational complex. Two transmembrane domains anchor Rcy to the inner membrane, with the bulk of the protein residing in the cytoplasm. The 13 predicted apolipoprotein domains of Rcy have been represented as four apolipoprotein domain-containing regions, connected by alpha-helical regions. Coiled-coil regions within these domains have been represented as yellow bars. The coiled-coil domains within Rcy may facilitate protein-protein interactions with other polar growth complex proteins, with Rcy serving as an organisational centre for PG remodelling machinery such as penicillin binding proteins (PBPs, green), L,D-transpeptidases (LDTs, pink) and glycosyltransferase-transpeptidase enzymes (GTase-TPases, pale blue), recruiting these proteins to the growth pole. PopZ, which serves as a growth pole marker within *R. leguminosarum*, may be organised by Rcy. PopZ monomers (purple circles) may interact with Rcy either via its coiled-coil regions or disordered N- or C-terminal regions (further research required). Within *R. leguminosarum*, PopZ monomers likely form a large matrix at the growth pole (pale purple). Finally, polar concentrations of ParA1 (yellow) may be established or maintained by either Rcy or PopZ in a cell cycle-dependent manner, similar to the establishment of polar ParA concentration gradients by Scy (*Streptomyces*), and PopZ and Tip (*Caulobacter*) at the growing/new tips/poles of these species (*183*, *206*, *314*, *315*).

Additional members of this polar complex which have not been detailed in the model depicted above may include *Rhizobium* homologs of the *Sinorhizobium* membrane proteins RgsA (unknown function), RgsS (an FtsN-like transmembrane protein) and TolQ (part of the Tol-Pal complex), which have been found to directly interact with the GPR/Rcy homolog RgsE (*280*, *281*). Additionally, this proposed *R. leguminosarum* polar protein complex may play an important role during chromosome segregation. In this thesis, ParA1 and PopZ were both attributed primarily to the control of cell division, which is closely linked to chromosome segregation (*450*). It is widely documented that ParA concentration gradients are required for the effective segregation of chromosomes towards the cell poles by the ParABS segregation system (*203*, *451*). Additionally, $PopZ_{At}$ (which seemingly shares many functional similarities with $PopZ_{Rl}$) has been found to be important for the segregation of *Agrobacterium* chromosomes, possibly tethering ParB-*parS* complexes in a manner similar to the tethering of ParB-*parS* by $PopZ_{Cc}$ within *Caulobacter* (*309*, *323*, *324*, *326*). Hence, an additional interactor of the *R. leguminosarum* polar complex may be ParB, interacting specifically with $PopZ_{Rl}$.

Although not discussed in detailed throughout this thesis, the predicted apolipoprotein domains present throughout the length of Rcy could be required for important addition functions of this protein. As mentioned in Section 4.6.3, enlarged, spherical morphologies within *R. leguminosarum* 3841 have also been observed following the disruption of genes encoding a suite proteins predicted to be required for cell envelope development and membrane integrity (*398*, *434*). The observation of near-identical morphologies following Rcy overexpression, and Rcy's high abundance of apolipoprotein domains (which facilitate lipid sorting) could link Rcy with the maintenance of membrane synthesis or integrity (*452*, *453*). Apolipoprotein domains likely play an important role during unipolar growth both within and beyond the *Rhizobiaceae*, with around 80% of bacterial apolipoproteins residing within the Hyphomicrobiales (*279*). However, further research is required to establish their role in these species.

6.5 Future perspectives

The work performed within this thesis has ultimately provided a valuable insight into the proteins which govern unipolar growth and the coordination of cell division within *R. leguminosarum*. This work has also provided support for the identity of PopZ as a growth pole marker across the *Rhizobiaceae*, and has highlighted the presence of a conserved putative scaffolding protein which organises polar PG synthesis machinery at the growth pole within this family (i.e. Rcy/GPR/RgsE). As the majority of literature regarding the regulation of polar growth within the Hyphomicrobiales has originated from *Agrobacterium* and *Sinorhizobium*, this research within *Rhizobium* has provided essential information into the functional conservation of these proteins. However, there are of course a few limitations of this study which should be addressed in future work. The first of these should be the confirmation of growth-pole localisation of both PopZ and Rcy. The localisation studies performed within this thesis did not formally identify pole identity, and instead growth poles were identified based on their morphology (i.e. tapered poles). Therefore, immediate future work should involve further localisation studies in parallel with the confirmation of pole identity, which could easily be achieved using several experimental approaches including (but not limited to) the use of appropriate cell staining to distinguish poles (e.g. FM 4-64, which stains the old poles of *Agrobacterium*) or live imaging and cell-cycle tracking (*127*). Arguably the other substantial limitation of these studies is the characterisation of these proteins based solely on overexpression studies. For a more robust characterisation of these proteins, the generation and subsequent characterisation of gene deletion or depletion mutants should ideally be attempted once more (see Section 3.4.5 for discussion regarding alternative gene deletion approaches).

Once these limitations have been addressed, several further research avenues could be pursued. Ultimately, these can be boiled down into one of two categories: identification of the proposed *R. leguminosarum* polar growth complex, or further characterisation of the role of ParA1, PopZ and Rcy throughout *R. leguminosarum*'s lifestyle. In the polar growth complex model proposed in Section 6.4 (Figure 6.1), a number of possible interactions between ParA1, PopZ and Rcy have been proposed, along with additional interactions between Rcy and polar PG-synthesis machinery, and PopZ and ParB of the ParABS system. These interactions were proposed based on supporting literature from other species (see references above), but direct interactions between these proteins within *R. leguminosarum* have yet to be proven. The employment of

protein-protein interaction studies (e.g. yeast two hybrid or pull-down assays) to confirm whether ParA1, PopZ, Rcy and other proposed members of this complex do indeed directly interact would provide important information as to how this proposed *R. leguminosarum* polar complex may function. Within this thesis, it has also been proposed that Rcy may recruit PopZ to the growth pole. This hypothesis was primarily based on studies within *Agrobacterium*, where GPR was found to localise to the growth pole prior to PopZ, and the loss of GPR impaired PopZ localisation (*278*). Live or time-lapse colocalisation studies of differentially-tagged PopZ and Rcy proteins (e.g. mCherry and GFP) could be a useful approach to determining the order that these two proteins localise to the growth pole, shedding light on the possible recruitment of PopZ to the growth pole by Rcy.

One of the most important features of the *R. leguminosarum* lifestyle is the establishment of endosymbiotic relationships with legumes belonging to the Fabeae tribe, including pea (*Pisum sativum*), lentil (*Lens culinaris*), broad bean (*Vicia faba*) and vetch (*Vicia cracca*) (*329*, *334*). As reviewed within Section 1.1, the establishment of these interactions requires several coordinated rhizobial processes including population of the growing infection thread by 'free-living' rhizobia, and substantial metabolic and morphological changes during terminal bacteroid differentiation. Symbioses between *Rhizobium leguminosarum 3841* and *Pisum sativum* give rise to indeterminate root nodules filled with E-type bacteroids, which are enlarged, elongated bacteroids that often exhibit branched morphologies (*76*). Notably, the overexpression of ParA1 yielded free-living *Rhizobium* cells with branched morphologies, while Rcy overexpression resulted in enlarged, spherical cells. These phenotypes, especially the swelling of cells following Rcy overexpression, are somewhat reminiscent of E-type and S-type (enlarged, spherical) bacteroids, suggesting that changes in the expression of ParA1 and Rcy may facilitate morphological changes during bacteroid differentiation.

Although studies into the characterisation of rhizobial cytoskeletal proteins in the context of endosymbiotic relationships are limited, a handful of key studies have been highlighted throughout this thesis. Arguably the most relevant of these is the characterisation of ParA within *Azorhizobium caulinodans* during symbioses with *Sesbania rostrata* (*227*, *433*). The deletion of *parA* from *A. caulinodans* resulted in the formation of immature stem nodules on *S. rostrata*, while ParA overexpression significantly altered *A. caulinodans* bacteroid formation. *S. rostrata* nodules typically contain multiple non-terminally differentiated U-type bacteroids, yet

overexpression of ParA resulted in the formation of symbiosomes containing a single, swollen bacteroid reminiscent of terminally differentiated E- and S-type bacteroids (*227*, *433*). As *Azorhizobium* is not a member of the *Rhizobiaceae*, studies of the impact of ParA1 overexpression within *R. leguminosarum* would yield essential data into additional roles of ParA proteins, and whether these supplementary functions are conserved across rhizobial species belonging to different families.

To facilitate similar studies of ParA1, PopZ and Rcy within *R. leguminosarum*, genomic overexpression and deletion strains should ideally be generated. The use of stable overexpression mutants as opposed to inducible, plasmid-based overexpression systems (like those used within this thesis) can help to overcome issues surrounding the retention of possibly detrimental plasmids in the absence of selection pressure (i.e. within the plant). Following inoculation of these *R. leguminosarum* strains onto a common model legume such as *Pisum sativum*, various aspects of the rhizobium-legume symbiosis could be studied. Infection thread tracking and observation of the rhizobia within these structures can highlight differences in the function of these proteins in free-living bacteria within a rapidly changing environment (i.e. the infection thread) vs. carefully controlled laboratory cultures, as well as demonstrating whether possible morphological defects arising from their misregulation may inhibit the progression of rhizobia through the infection thread. The observation of bacteroid morphology within the nodule may provide insight as to whether these proteins do indeed regulate bacteroid differentiation, and nodulation assays can be used to determine the efficiency of nodulation and biological nitrogen fixation within the resulting bacteroids. It is known that members of the eukaryotic cytoskeleton serve vital roles during various stages of the establishment of rhizobium-legume symbioses, from root hair curling and infection thread formation all the way to the release of rhizobia into the nodule, but the involvement of the bacterial cytoskeleton has been almost completely overlooked (*45*, *47–49*, *64*, *96–99*). Overall, such studies would provide a valuable insight into how ParA1, PopZ and Rcy not only regulate growth and division within free-living *Rhizobium* species, but also how they may be required for the establishment and/or maintenance of effective endosymbiotic interactions.

References

1. S. T. N. Aroney, P. S. Poole, C. Sánchez-Cañizares, Rhizobial Chemotaxis and Motility Systems at Work in the Soil. *Front. Plant Sci.* **12** (2021).

2. L. M. Carareto Alves, J. A. M. de Souza, A. D. M. Varani, E. G. D. M. Lemos, "The Family Rhizobiaceae" in *The Prokaryotes* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2014)vol. 9783642301, pp. 419–437.

3. E. Ormeño-Orrillo, L. E. Servín-Garcidueñas, M. A. Rogel, V. González, H. Peralta, J. Mora, J. Martínez-Romero, E. Martínez-Romero, Taxonomy of rhizobia and agrobacteria from the Rhizobiaceae family in light of genomics. *Syst. Appl. Microbiol.* **38**, 287–291 (2015).

4. T. Ma, H. Xue, C. Piao, N. Jiang, Y. Li, Phylogenomic reappraisal of the family Rhizobiaceae at the genus and species levels, including the description of Ectorhizobium quercum gen. nov., sp. nov. *Front. Microbiol.* **14** (2023).

5. N. Kuzmanović, C. Fagorzi, A. Mengoni, F. Lassalle, G. C. Dicenzo, Taxonomy of Rhizobiaceae revisited: proposal of a new framework for genus delimitation. *Int. J. Syst. Evol. Microbiol.* **72** (2022).

6. A. C. Parte, J. Sardà Carbasse, J. P. Meier-Kolthoff, L. C. Reimer, M. Göker, List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int. J. Syst. Evol. Microbiol.* **70**, 5607–5612 (2020).

7. J. P. W. Young, S. Moeskjær, A. Afonin, P. Rahi, M. Maluk, E. K. James, M. I. A. Cavassim, M. H. Rashid, A. A. Aserse, B. J. Perry, Defining the Rhizobium leguminosarum species complex. *Genes (Basel).* **12**, 111 (2021).

8. M. A. Escobar, A. M. Dandekar, Agrobacterium tumefaciens as an agent of disease. *Trends Plant Sci.* **8**, 380–386 (2003).

9. J. I. Sprent, J. Platzmann, *Nodulation in Legumes* (Royal Botanic Gardens Kew, 2001).

10. M. Udvardi, P. S. Poole, Transport and Metabolism in Legume-Rhizobia Symbioses. *Annu. Rev. Plant Biol.* **64**, 781–805 (2013).

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

12. B. Frank, *Ueber Die Pilzymbiose Der Leguminosen* (P. Parey, 1890).

13. J. P. W. Young, B. Jorrin, S. Moeskjær, E. K. James, Rhizobium brockwellii sp. nov., Rhizobium johnstonii sp. nov. and Rhizobium beringeri sp. nov., three genospecies within the Rhizobium leguminosarum species complex. *Int. J. Syst. Evol. Microbiol.* **73**, 5979 (2023).

14. D. J. Gage, Infection and Invasion of Roots by Symbiotic, Nitrogen-Fixing Rhizobia during Nodulation of Temperate Legumes. *Microbiol. Mol. Biol. Rev.* **68**, 280–300 (2004).

15. C. Masson-Boivin, E. Giraud, X. Perret, J. Batut, Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? *Trends Microbiol.* **17**, 458–466 (2009).

16. E. E. Chaia, L. G. Wall, K. Huss-Danell, Life in soil by the actinorhizal root nodule endophyte Frankia. A review. *Symbiosis* **51**, 201–226 (2010).

17. D. F. Herridge, M. B. Peoples, R. M. Boddey, Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil* **311**, 1–18 (2008).

18. A. Shah, M. Nazari, M. Antar, L. A. Msimbira, J. Naamala, D. Lyu, M. Rabileh, J. Zajonc, D. L. Smith, PGPR in Agriculture: A Sustainable Approach to Increasing Climate Change Resilience. *Front. Sustain. Food Syst.* **5** (2021).

19. M. Udvardi, E. L. Brodie, W. Riley, S. Kaeppler, J. Lynch, Impacts of Agricultural Nitrogen on the Environment and Strategies to Reduce these Impacts. *Procedia Environ. Sci.* **29**, 303 (2015).

20. FAO, World Food and Agriculture – Statistical Yearbook. (2023).

21. J. Humphreys, R. Lan, S. Tao, Development and recent progress on ammonia synthesis catalysts for Haber–Bosch process. *Adv. Energy Sustain. Res.* **2**, 2000043 (2021).

22. J. W. Erisman, M. A. Sutton, J. Galloway, Z. Klimont, W. Winiwarter, How a century of ammonia synthesis changed the world. *Nat. Geosci.* **1**, 636–639 (2008).

23. Y. Song, D. Johnson, R. Peng, D. K. Hensley, P. V Bonnesen, L. Liang, J. Huang, F. Yang, F. Zhang, R. Qiao, A physical catalyst for the electrolysis of nitrogen to ammonia. *Sci. Adv.* **4**, e1700336 (2018).

24. E. A. Davidson, The contribution of manure and fertilizer nitrogen to atmospheric nitrous oxide since 1860. *Nat. Geosci.* **2**, 659–662 (2009).

25. J. W. Erisman, J. N. Galloway, S. Seitzinger, A. Bleeker, N. B. Dise, A. M. R. Petrescu, A. M. Leach, W. de Vries, Consequences of human modification of the global nitrogen cycle. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, 20130116 (2013).

26. W. H. Schlesinger, On the fate of anthropogenic nitrogen. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 203–208 (2009).

27. Y. Gao, A. Cabrera Serrenho, Greenhouse gas emissions from nitrogen fertilizers could be reduced by up to one-fifth of current levels by 2050 with combined interventions. *Nat. Food* **4**, 170–178 (2023).

28. M. N. Khan, F. Mohammad, "Eutrophication: Challenges and Solutions" in *Ansari, A., Gill, S. (Eds) Eutrophication: Causes, Consequences and Control: Volume 2*, A. A. Ansari, S. S. Gill, Eds. (Springer Netherlands, Dordrecht, 2014), pp. 1–15.

29. K. C. Cameron, H. J. Di, J. L. Moir, Nitrogen losses from the soil/plant system: a review. *Ann. Appl. Biol.* **162**, 145–173 (2013).

30. G. E. D. Oldroyd, Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol.* **11**, 252–263 (2013).

31. G. E. D. Oldroyd, J. D. Murray, P. S. Poole, J. A. Downie, The Rules of Engagement in the Legume-Rhizobial Symbiosis. *Annu. Rev. Genet.* **45**, 119–144 (2011).

32. Q. Wang, J. Liu, H. Zhu, Genetic and Molecular Mechanisms Underlying Symbiotic Specificity in Legume-Rhizobium Interactions. *Front. Plant Sci.* **9**, 313 (2018).

33. J. Yang, L. Lan, Y. Jin, N. Yu, D. Wang, E. Wang, Mechanisms underlying legume-rhizobium symbioses. *J. Integr. Plant Biol.* **64**, 244–267 (2022).

34. W. Dong, Y. Song, The significance of flavonoids in the process of biological nitrogen fixation. *Int. J. Mol. Sci.* **21**, 5926 (2020).

35. C.-W. Liu, J. D. Murray, The role of flavonoids in nodulation host-range specificity: an update. *Plants* **5**, 33 (2016).

36. S. R. Long, Genes and Signals in the Rhizobium-Legume Symbiosis. *Plant Physiol.* **125**, 69–72 (2001).

37. B. G. Rolfe, Flavones and isoflavones as inducing substances of legume nodulation. *Biofactors* **1**, 3–10 (1988).

38. J. Stougaard, Regulators and regulation of legume root nodule development. *Plant Physiol.* **124**, 531–540 (2000).

39. M. D. Mbengue, C. Hervé, F. Debellé, Nod factor signaling in symbiotic nodulation. *Adv. Bot. Res.* **94**, 1–39 (2020).

40. P. M. C., F. R. F., L. S. R., Diverse Flavonoids Stimulate NodD1 Binding to nod Gene Promoters in Sinorhizobium meliloti. *J. Bacteriol.* **188**, 5417–5427 (2006).

41. H. Kobayashi, Y. N. Graven, W. J. Broughton, X. Perret, Flavonoids induce temporal shifts in gene-expression of nod-box controlled loci in Rhizobium sp. NGR234. *Mol. Microbiol.* **51**, 335–347 (2004).

42. A. A. Begum, S. Leibovitch, P. Migner, F. Zhang, Specific flavonoids induced nod gene expression and pre-activated nod genes of Rhizobium leguminosarum increased pea (Pisum sativum L.) and lentil (Lens culinaris L.) nodulation in controlled growth chamber environments. *J. Exp. Bot.* **52**, 1537–1543 (2001).

43. S. R. Long, Rhizobium symbiosis: nod factors in perspective. *Plant Cell* **8**, 1885 (1996).

44. J. J. Esseling, F. G. P. Lhuissier, A. M. C. Emons, Nod factor-induced root hair curling: continuous polar growth towards the point of nod factor application. *Plant Physiol.* **132**, 1982–1988 (2003).

45. S. M. Schmidt, R. Panstruga, Cytoskeleton functions in plant–microbe interactions. *Physiol. Mol. Plant Pathol.* **71**, 135–148 (2007).

46. N. J. Brewin, Plant cell wall remodelling in the rhizobium-legume symbiosis. *CRC. Crit. Rev. Plant Sci.* **23**, 293–316 (2004).

47. A. C. J. Timmers, M. C. Auriac, G. Truchet, Refined analysis of early symbiotic steps of the Rhizobium-Medicago interaction in relationship with microtubular cytoskeleton rearrangements. *Development* **126**, 3617–3628 (1999).

48. K. Yokota, E. Fukai, L. H. Madsen, A. Jurkiewicz, P. Rueda, S. Radutoiu, M. Held, M. S. Hossain, K. Szczyglowski, G. Morieri, G. E. D. Oldroyd, J. A. Downie, M. W. Nielsen, A. M. Rusek, S. Sato, S. Tabata, E. K. James, H. Oyaizu, N. Sandal, J. Stougaard, Rearrangement of Actin Cytoskeleton Mediates Invasion of Lotus japonicus Roots by Mesorhizobium loti. *Plant Cell* **21**, 267–284 (2009).

49. F. M. Perrine-Walker, M. Lartaud, H. Kouchi, R. W. Ridge, Microtubule array formation during root hair infection thread initiation and elongation in the Mesorhizobium-Lotus symbiosis. *Protoplasma* **251**, 1099–1111 (2014).

50. J. Fournier, A. Teillet, M. Chabaud, S. Ivanov, A. Genre, E. Limpens, F. de Carvalho-Niebel, D. G. Barker, Remodeling of the infection chamber before infection thread formation reveals a two-step mechanism for rhizobial entry into the host legume root hair. *Plant Physiol.* **167**, 1233–1242 (2015).

51. A. A. N. van Brussel, R. Bakhuizen, P. C. van Spronsen, H. P. Spaink, T. Tak, B. J. J. Lugtenberg, J. W. Kijne, Induction of pre-infection thread structures in the leguminous host plant by mitogenic lipo-oligosaccharides of Rhizobium. *Science (80-.).* **257**, 70–72 (1992).

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

53. B. G. Turgeon, W. D. Bauer, Ultrastructure of infection-thread development during the infection of soybean by Rhizobium japonicum. *Planta* **163**, 328–349 (1985).

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

55. P. C. van Spronsen, R. Bakhuizen, A. A. van Brussel, J. W. Kijne, Cell wall degradation during infection thread formation by the root nodule bacterium Rhizobium leguminosarum is a two-step process. *Eur. J. Cell Biol.* **64**, 88—94 (1994).

56. F. Xie, J. Murray, J. Kim, A. Heckmann, A. Edwards, G. Oldroyd, J. Downie, Legume pectate lyase required for root infection by rhizobia. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 633–638 (2011).

57. M. Robledo, J. I. Jimenez-Zurdo, E. Velazquez, M. E. Trujillo, J. L. Zurdo-Pineiro, M. H. Ramirez-Bahena, B. Ramos, J. M. Diaz-Minguez, F. Dazzo, E. Martinez-Molina, P. F. Mateos, Rhizobium cellulase CelC2 is essential for primary symbiotic infection of legume host roots. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 7064–7069 (2008).

58. R. W. Ridge, B. G. Rolfe, Rhizobium sp. Degradation of Legume Root Hair Cell Wall at the Site of Infection Thread Origin. *Appl. Environ. Microbiol.* **50**, 717–720 (1985).

59. C. Su, G. Zhang, M. Rodriguez-Franco, R. Hinnenberg, J. Wietschorke, P. Liang, W. Yang, L. Uhler, X. Li, T. Ott, Transcellular progression of infection threads in Medicago truncatula roots is associated with locally confined cell wall modifications. *Curr. Biol.* **33**, 533–542 (2023).

60. A. V. Tsyganova, N. J. Brewin, V. E. Tsyganov, Structure and development of the legume-rhizobial symbiotic interface in infection threads. *Cells* **10**, 1050 (2021).

61. M. Schultze, A. Kondorosi, Regulation of symbiotic root nodule development. *Annu. Rev. Genet.* **32**, 33–57 (1998).

62. A. V. Tsyganova, V. E. Tsyganov, “Plant Genetic Control over Infection Thread Development during Legume-Rhizobium Symbiosis” in *Symbiosis* (InTechOpen, 2018; http://www.intechopen.com/books/symbiosis/plant-genetic-control-over-infection-thread-development-during-legume-rhizobium-symbiosis), pp. 23–52.

63. B. Bassett, R. N. Goodman, A. Novacky, Ultrastructure of soybean nodules. I: release of rhizobia from the infection thread. *Can. J. Microbiol.* **23**, 573–582 (1977).

64. K. Hlaváčková, J. Šamaj, M. Ovečka, Cytoskeleton as a roadmap navigating rhizobia to establish symbiotic root nodulation in legumes. *Biotechnol. Adv.*, 108263 (2023).

65. A. V Tsyganova, A. B. Kitaeva, V. E. Tsyganov, Cell differentiation in nitrogen-fixing nodules hosting symbiosomes. *Funct. Plant Biol.* **45**, 47–57 (2017).

66. T. Coba de la Pena, E. Fedorova, J. J. Pueyo, M. M. Lucas, The symbiosome: legume and

rhizobia co-evolution toward a nitrogen-fixing organelle? *Front. Plant Sci.* **8**, 2229 (2018).

67. R. Oono, I. Schmitt, J. I. Sprent, R. F. Denison, Multiple evolutionary origins of legume traits leading to extreme rhizobial differentiation. *New Phytol.* **187**, 508–520 (2010).

68. A. M. Hirsch, Developmental biology of legume nodulation. *New Phytol.* **122**, 211–237 (1992).

69. W. Kohlen, J. L. P. Ng, E. E. Deinum, U. Mathesius, Auxin transport, metabolism, and signalling during nodule initiation: indeterminate and determinate nodules. *J. Exp. Bot.* **69**, 229–244 (2018).

70. J. A. Imlay, Iron-sulphur clusters and the problem with oxygen. *Mol. Microbiol.* **59**, 1073–1082 (2006).

71. C. Franche, K. Lindström, C. Elmerich, Nitrogen-fixing bacteria associated with leguminous and non-leguminous plants. *Plant Soil* **321**, 35–59 (2009).

72. K. Marchal, J. Vanderleyden, The" oxygen paradox" of dinitrogen-fixing bacteria. *Biol. Fertil. soils* **30**, 363–373 (2000).

73. A. R. Schwember, J. Schulze, A. del Pozo, R. A. Cabeza, Regulation of Symbiotic Nitrogen Fixation in Legume Root Nodules. *Plants* **8**, 333 (2019).

74. T. Ott, J. T. van Dongen, C. Gu, L. Krusell, G. Desbrosses, H. Vigeolas, V. Bock, T. Czechowski, P. Geigenberger, M. K. Udvardi, others, Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. *Curr. Biol.* **15**, 531–535 (2005).

75. J. Montiel, A. Szűcs, I. Z. Boboescu, V. D. Gherman, É. Kondorosi, A. Kereszt, Terminal Bacteroid Differentiation Is Associated With Variable Morphological Changes in Legume Species Belonging to the Inverted Repeat-Lacking Clade. *Mol. Plant-Microbe Interact.* **29**, 210–219 (2015).

76. P. Mergaert, T. Uchiumi, B. Alunni, G. Evanno, A. Cheron, O. Catrice, A.-E. Mausset, F. Barloy-Hubler, F. Galibert, A. Kondorosi, Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium–legume symbiosis. *Proc. Natl. Acad. Sci.* **103**, 5230–5235 (2006).

77. F. Lamouche, N. Bonadé-Bottino, P. Mergaert, B. Alunni, Symbiotic Efficiency of Spherical and Elongated Bacteroids in the Aeschynomene-Bradyrhizobium Symbiosis. *Front. Plant Sci.* **10**, 377 (2019).

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

79. R. Oono, R. F. Denison, Comparing symbiotic efficiency between swollen versus nonswollen rhizobial bacteroids. *Plant Physiol.* **154**, 1541–1548 (2010).

80. A. Kereszt, P. Mergaert, J. Montiel, G. Endre, É. Kondorosi, Impact of plant peptides on symbiotic nodule development and functioning. *Front. Plant Sci.* **9**, 1026 (2018).

81. F. Lamouche, D. Gully, A. Chaumeret, N. Nouwen, C. Verly, O. Pierre, C. Sciallano, J. Fardoux, C. Jeudy, A. Szücs, Transcriptomic dissection of Bradyrhizobium sp. strain ORS285 in symbiosis with Aeschynomene spp. inducing different bacteroid morphotypes with contrasted symbiotic efficiency. *Environ. Microbiol.* **21**, 3244–3258

(2019).

82. M. C. Laus, A. A. N. Van Brussel, J. W. Kijne, Role of cellulose fibrils and exopolysaccharides of Rhizobium leguminosarum in attachment to and infection of Vicia sativa root hairs. *Mol. Plant-Microbe Interact.* **18**, 533–538 (2005).

83. S. Acosta-Jurado, F. Fuentes-Romero, J.-E. Ruiz-Sainz, M. Janczarek, J.-M. Vinardell, Rhizobial exopolysaccharides: Genetic regulation of their synthesis and relevance in symbiosis with legumes. *Int. J. Mol. Sci.* **22**, 6233 (2021).

84. V. C. Clarke, P. C. Loughlin, D. A. Day, P. M. C. Smith, Transport processes of the legume symbiosome membrane. *Front. Plant Sci.* **5**, 699 (2014).

85. E. Lodwig, P. Poole, Metabolism of Rhizobium bacteroids. *CRC. Crit. Rev. Plant Sci.* **22**, 37–78 (2003).

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

87. R. Dixon, D. Kahn, Genetic regulation of biological nitrogen fixation. *Nat. Rev. Microbiol.* **2**, 621–631 (2004).

88. J. P. Zehr, P. J. Turner, Nitrogen fixation: nitrogenase genes and gene expression. *Methods Microbiol.* **30**, 271–286 (2001).

89. J. Prell, A. Bourdès, S. Kumar, E. Lodwig, A. Hosie, S. Kinghorn, J. White, P. Poole, Role of symbiotic auxotrophy in the Rhizobium-legume symbioses. *PLoS One* **5**, e13933–e13933 (2010).

90. E. M. Lodwig, A. H. F. Hosie, A. Bourdès, K. Findlay, D. Allaway, R. Karunakaran, J. A. Downie, P. S. Poole, Amino-acid cycling drives nitrogen fixation in the legume-Rhizobium symbiosis. *Nature* **422**, 722–726 (2003).

91. D. Duran, M. Albareda, C. García, A.-I. Marina, T. Ruiz-Argüeso, J.-M. Palacios, Proteome analysis reveals a significant host-specific response in Rhizobium leguminosarum bv. viciae endosymbiotic cells. *Mol. Cell. Proteomics* **20** (2021).

92. J. Prell, J. P. White, A. Bourdes, S. Bunnewell, R. J. Bongaerts, P. S. Poole, Legumes regulate Rhizobium bacteroid development and persistence by the supply of branched-chain amino acids. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 12477–12482 (2009).

93. E. Kondorosi, P. Mergaert, A. Kereszt, A paradigm for endosymbiotic life: cell differentiation of Rhizobium bacteria provoked by host plant factors. *Annu. Rev. Microbiol.* **67**, 611–628 (2013).

94. M. T. Cabeen, C. Jacobs-Wagner, Skin and bones: the bacterial cytoskeleton, cell wall, and cell morphogenesis. *J. Cell Biol.* **179**, 381–387 (2007).

95. K. D. Young, The selective value of bacterial shape. *Microbiol. Mol. Biol. Rev.* **70**, 660–703 (2006).

96. V. E. Tsyganov, A. B. Kitaeva, K. N. Demchenko, Comparative analysis of tubulin cytoskeleton rearrangements in nodules of Medicago truncatula and Pisum sativum. *Model Legum. Medicago truncatula*, 543–547 (2020).

97. X. Zhang, L. Han, Q. Wang, C. Zhang, Y. Yu, J. Tian, Z. Kong, The host actin cytoskeleton channels rhizobia release and facilitates symbiosome accommodation during

nodulation in Medicago truncatula. *New Phytol.* **221**, 1049–1059 (2019).

98. A. C. J. TIMMERS, The role of the plant cytoskeleton in the interaction between legumes and rhizobia. *J. Microsc.* **231**, 247–256 (2008).

99. A. B. Kitaeva, A. P. Gorshkov, P. G. Kusakin, A. R. Sadovskaya, A. V Tsyganova, V. E. Tsyganov, Tubulin cytoskeleton organization in cells of determinate nodules. *Front. Plant Sci.* **13**, 823183 (2022).

100. K. Celler, R. I. Koning, A. J. Koster, G. P. van Wezel, Multidimensional view of the bacterial cytoskeleton. *J. Bacteriol.* **195**, 1627–1636 (2013).

101. M. Ingerson-Mahar, Z. Gitai, A growing family: the expanding universe of the bacterial cytoskeleton. *FEMS Microbiol. Rev.* **36**, 256–266 (2012).

102. J. Wagstaff, J. Löwe, Prokaryotic cytoskeletons: protein filaments organizing small cells. *Nat. Rev. Microbiol.* **16**, 187–201 (2018).

103. D. A. Fletcher, R. D. Mullins, Cell mechanics and the cytoskeleton. *Nature* **463**, 485–492 (2010).

104. E. Bi, J. Lutkenhaus, FtsZ ring structure associated with division in Escherichia coli. *Nature* **354**, 161–164 (1991).

105. P. De Boer, R. Crossley, L. Rothfield, The essential bacterial cell-division protein FtsZ is a GTPase. *Nature* **359**, 254–256 (1992).

106. A. Mukherjee, K. Dai, J. Lutkenhaus, Escherichia coli cell division protein FtsZ is a guanine nucleotide binding protein. *Proc. Natl. Acad. Sci.* **90**, 1053–1057 (1993).

107. L. J. F. Jones, R. Carballido-López, J. Errington, Control of cell shape in bacteria: helical, actin-like filaments in Bacillus subtilis. *Cell* **104**, 913–922 (2001).

108. N. Ausmees, J. R. Kuhn, C. Jacobs-Wagner, The bacterial cytoskeleton: an intermediate filament-like function in cell shape. *Cell* **115**, 705–713 (2003).

109. Y.-L. Shih, L. Rothfield, The Bacterial Cytoskeleton. *Microbiol. Mol. Biol. Rev.* **70**, 729–754 (2006).

110. J. Löwe, L. A. Amos, Evolution of cytomotive filaments: The cytoskeleton from prokaryotes to eukaryotes. *Int. J. Biochem. Cell Biol.* **41**, 323–329 (2009).

111. K. K. Busiek, W. Margolin, Bacterial actin and tubulin homologs in cell growth and division. *Curr. Biol.* **25**, R243–R254 (2015).

112. F. van den Ent, L. Amos, J. Löwe, Bacterial ancestry of actin and tubulin. *Curr. Opin. Microbiol.* **4**, 634–638 (2001).

113. K.-H. Huang, J. Durand-Heredia, A. Janakiraman, FtsZ ring stability: of bundles, tubules, crosslinks, and curves. *J. Bacteriol.* **195**, 1859–1868 (2013).

114. E. Nogales, K. H. Downing, L. A. Amos, J. Löwe, Tubulin and FtsZ form a distinct family of GTPases. *Nat. Struct. Biol.* **5**, 451–458 (1998).

115. J. M. Andreu, F. M. Ruiz, C. Fernández-Tornero, Conserved GTPase mechanism in bacterial FtsZ and archaeal tubulin filaments. *FEBS J.* **290**, 3527–3532 (2023).

116. J. Löwe, L. A. Amos, Tubulin-like protofilaments in Ca2+-induced FtsZ sheets. *EMBO J.* **18**, 2364–2371 (1999).

117. A. Mukherjee, J. Lutkenhaus, Guanine nucleotide-dependent assembly of FtsZ into filaments. *J. Bacteriol.* **176**, 2754–2758 (1994).

118. A. Mukherjee, J. Lutkenhaus, Dynamic assembly of FtsZ regulated by GTP hydrolysis. *EMBO J.* **17**, 462–469 (1998).

119. X. Yang, Z. Lyu, A. Miguel, R. McQuillen, K. C. Huang, J. Xiao, GTPase activity–coupled treadmilling of the bacterial tubulin FtsZ organizes septal cell wall synthesis. *Science (80-.).* **355**, 744–747 (2017).

120. R. A. Larsen, C. Cusumano, A. Fujioka, G. Lim-Fong, P. Patterson, J. Pogliano, Treadmilling of a prokaryotic tubulin-like protein, TubZ, required for plasmid stability in Bacillus thuringiensis. *Genes Dev.* **21**, 1340–1352 (2007).

121. M. A. Oliva, S. C. Cordell, J. Löwe, Structural insights into FtsZ protofilament formation. *Nat. Struct. Mol. Biol.* **11**, 1243–1250 (2004).

122. J. Löwe, H. Li, K. H. Downing, E. Nogales, Refined structure of αβ-tubulin at 3.5 Å resolution. *J. Mol. Biol.* **313**, 1045–1057 (2001).

123. J. Löwe, L. A. Amos, *Prokaryotic Cytoskeletons: Filamentous Protein Polymers Active in the Cytoplasm of Bacterial and Archaeal Cells* (Springer, 2017)vol. 84.

124. B. Jindal, D. Panda, Understanding FtsZ assembly: cues from the behavior of its N- and C-terminal domains. *Biochemistry* **52**, 7071–7081 (2013).

125. A. W. Bisson-Filho, Y.-P. Hsu, G. R. Squyres, E. Kuru, F. Wu, C. Jukes, Y. Sun, C. Dekker, S. Holden, M. S. VanNieuwenhze, Y. V Brun, E. C. Garner, Treadmilling by FtsZ filaments drives peptidoglycan synthesis and bacterial cell division. *Science (80-.).* **355**, 739–743 (2017).

126. M. Loose, T. J. Mitchison, The bacterial cell division proteins FtsA and FtsZ self-organize into dynamic cytoskeletal patterns. *Nat. Cell Biol.* **16**, 38–46 (2014).

127. J. R. Zupan, T. A. Cameron, J. Anderson-Furgeson, P. C. Zambryski, Dynamic FtsA and FtsZ localization and outer membrane alterations during polar growth and cell division in Agrobacterium tumefaciens. *Proc. Natl. Acad. Sci.* **110**, 9060–9065 (2013).

128. V. W. Rowlett, W. Margolin, The bacterial divisome: ready for its close-up. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20150028 (2015).

129. P. Szwedziak, J. Löwe, Do the divisome and elongasome share a common evolutionary past? *Curr. Opin. Microbiol.* **16**, 745–751 (2013).

130. M. Howell, A. Aliashkevich, K. Sundararajan, J. J. Daniel, P. J. Lariviere, E. D. Goley, F. Cava, P. J. B. Brown, Agrobacterium tumefaciens divisome proteins regulate the transition from polar growth to cell division. *Mol. Microbiol.* **111**, 1074–1092 (2019).

131. T. A. Cameron, J. R. Zupan, P. C. Zambryski, The essential features and modes of bacterial polar growth. *Trends Microbiol.* **23**, 347–353 (2015).

132. W. Zhao, H. Zhu, F. Wei, D. Zhou, Y. Li, X.-X. Zhang, Investigating the involvement of cytoskeletal proteins MreB and FtsZ in the origin of legume-rhizobial symbiosis. *Mol. Plant-Microbe Interact.* **34**, 547–559 (2021).

133. W. Margolin, S. R. Long, Rhizobium meliloti contains a novel second homolog of the cell division gene ftsZ. *J. Bacteriol.* **176**, 2033–2043 (1994).

134. M. Gao, M. Teplitski, RIVET—a tool for in vivo analysis of symbiotically relevant gene

expression in Sinorhizobium meliloti. *Mol. plant-microbe Interact.* **21**, 162–170 (2008).

135. S. Vaughan, B. Wickstead, K. Gull, S. G. Addinall, Molecular evolution of FtsZ protein sequences encoded within the genomes of archaea, bacteria, and eukaryota. *J. Mol. Evol.* **58**, 19–29 (2004).

136. X. Ma, Q. Sun, R. Wang, G. Singh, E. L. Jonietz, W. Margolin, Interactions between heterologous FtsA and FtsZ proteins at the FtsZ ring. *J. Bacteriol.* **179**, 6788–6797 (1997).

137. F. D. Müller, O. Raschdorf, H. Nudelman, M. Messerer, E. Katzmann, J. M. Plitzko, R. Zarivach, D. Schüler, The FtsZ-Like Protein FtsZm of Magnetospirillum gryphiswaldense Likely Interacts with Its Generic Homolog and Is Required for Biomineralization under Nitrate Deprivation. *J. Bacteriol.* **196**, 650–659 (2014).

138. S. Barber-Zucker, N. Keren-Khadmy, R. Zarivach, From invagination to navigation: The story of magnetosome-associated proteins in magnetotactic bacteria. *Protein Sci.* **25**, 338–351 (2016).

139. Y. Li, Redox control of magnetosome biomineralization. *J. Oceanol. Limnol.* **39**, 2070–2081 (2021).

140. X. Deng, G. Fink, T. A. M. Bharat, S. He, D. Kureisaite-Ciziene, J. Löwe, Four-stranded mini microtubules formed by Prosthecobacter BtubAB show dynamic instability. *Proc. Natl. Acad. Sci.* **114**, E5950–E5958 (2017).

141. D. Schlieper, M. A. Oliva, J. M. Andreu, J. Löwe, Structure of bacterial tubulin BtubA/B: evidence for horizontal gene transfer. *Proc. Natl. Acad. Sci.* **102**, 9170–9175 (2005).

142. C. Jenkins, R. Samudrala, I. Anderson, B. P. Hedlund, G. Petroni, N. Michailova, N. Pinel, R. Overbeek, G. Rosati, J. T. Staley, Genes for the cytoskeletal protein tubulin in the bacterial genus Prosthecobacter. *Proc. Natl. Acad. Sci.* **99**, 17049–17054 (2002).

143. C. H. S. Aylett, Q. Wang, K. A. Michie, L. A. Amos, J. Löwe, Filament structure of bacterial tubulin homologue TubZ. *Proc. Natl. Acad. Sci.* **107**, 19766–19771 (2010).

144. M. A. Oliva, A. J. Martin-Galiano, Y. Sakaguchi, J. M. Andreu, Tubulin homolog TubZ in a phage-encoded partition system. *Proc. Natl. Acad. Sci.* **109**, 7711–7716 (2012).

145. P. Bork, C. Sander, A. Valencia, An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. *Proc. Natl. Acad. Sci.* **89**, 7290–7294 (1992).

146. F. van den Ent, J. Møller-Jensen, L. A. Amos, K. Gerdes, J. Löwe, F-actin-like filaments formed by plasmid segregation protein ParM. *EMBO J.* **21**, 6935–6943 (2002).

147. F. van den Ent, L. A. Amos, J. Löwe, Prokaryotic origin of the actin cytoskeleton. *Nature* **413**, 39–44 (2001).

148. E. Ozyamak, J. M. Kollman, A. Komeili, Bacterial actins and their diversity. *Biochemistry* **52**, 6928–6939 (2013).

149. R. Dominguez, K. C. Holmes, Actin structure and function. *Annu. Rev. Biophys.* **40**, 169–186 (2011).

150. T. Oda, M. Iwasa, T. Aihara, Y. Maéda, A. Narita, The nature of the globular-to fibrous-actin transition. *Nature* **457**, 441–445 (2009).

151. R. A. Daniel, J. Errington, Control of cell morphogenesis in bacteria: Two distinct ways to

make a rod-shaped cell. *Cell* **113**, 767–776 (2003).

152. P. J. B. Brown, M. A. de Pedro, D. T. Kysela, C. Van der Henst, J. Kim, X. De Bolle, C. Fuqua, Y. V Brun, Polar growth in the Alphaproteobacterial order Rhizobiales. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1697–1701 (2012).

153. H. Shi, B. P. Bratton, Z. Gitai, K. C. Huang, How to build a bacterial cell: MreB as the foreman of E. coli construction. *Cell* **172**, 1294–1305 (2018).

154. J. Salje, F. van den Ent, P. de Boer, J. Löwe, Direct Membrane Binding by Bacterial Actin MreB. *Mol. Cell* **43**, 478–487 (2011).

155. F. Van den Ent, T. Izoré, T. A. M. Bharat, C. M. Johnson, J. Löwe, Bacterial actin MreB forms antiparallel double filaments. *Elife* **3**, e02634 (2014).

156. J. Domínguez-Escobar, A. Chastanet, A. H. Crevenna, V. Fromion, R. Wedlich-Söldner, R. Carballido-López, Processive movement of MreB-associated cell wall biosynthetic complexes in bacteria. *Science (80-.).* **333**, 225–228 (2011).

157. E. C. Garner, R. Bernard, W. Wang, X. Zhuang, D. Z. Rudner, T. Mitchison, Coupled, circumferential motions of the cell wall synthesis machinery and MreB filaments in B. subtilis. *Science (80-.).* **333**, 222–225 (2011).

158. S. Van Teeffelen, S. Wang, L. Furchtgott, K. C. Huang, N. S. Wingreen, J. W. Shaevitz, Z. Gitai, The bacterial actin MreB rotates, and rotation depends on cell-wall assembly. *Proc. Natl. Acad. Sci.* **108**, 15822–15827 (2011).

159. P. Szwedziak, Q. Wang, S. M. V Freund, J. Löwe, FtsA forms actin-like protofilaments. *EMBO J.* **31**, 2249–2260 (2012).

160. X. Ma, W. Margolin, Genetic and functional analyses of the conserved C-terminal core domain of Escherichia coli FtsZ. *J. Bacteriol.* **181**, 7531–7544 (1999).

161. J. J. Morrison, J. Conti, J. L. Camberg, Assembly and architecture of Escherichia coli divisome proteins FtsA and FtsZ. *J. Biol. Chem.* **298** (2022).

162. G. Ebersbach, K. Gerdes, Plasmid segregation mechanisms. *Annu. Rev. Genet.* **39**, 453–479 (2005).

163. E. Becker, N. C. Herrera, F. Q. Gunderson, A. I. Derman, A. L. Dance, J. Sims, R. A. Larsen, J. Pogliano, DNA segregation by the bacterial actin AlfA during Bacillus subtilis growth and development. *EMBO J.* **25**, 5919–5931 (2006).

164. A. I. Derman, E. C. Becker, B. D. Truong, A. Fujioka, T. M. Tucey, M. L. Erb, P. C. Patterson, J. Pogliano, Phylogenetic analysis identifies many uncharacterized actin-like proteins (Alps) in bacteria: regulated polymerization, dynamic instability and treadmilling in Alp7A. *Mol. Microbiol.* **73**, 534–552 (2009).

165. J. Salje, P. Gayathri, J. Löwe, The ParMRC system: molecular mechanisms of plasmid segregation by actin-like filaments. *Nat. Rev. Microbiol.* **8**, 683–692 (2010).

166. P. Gayathri, T. Fujii, J. Møller-Jensen, F. Van Den Ent, K. Namba, J. Löwe, A bipolar spindle of antiparallel ParM filaments drives bacterial plasmid segregation. *Science (80-.).* **338**, 1334–1337 (2012).

167. A. Scheffel, M. Gruska, D. Faivre, A. Linaroudis, J. M. Plitzko, D. Schüler, An acidic protein aligns magnetosomes along a filamentous structure in magnetotactic bacteria. *Nature* **440**, 110–114 (2006).

168. A. Komeili, Z. Li, D. K. Newman, G. J. Jensen, Magnetosomes are cell membrane imaginations organized by the actin-like protein MamK. *Science (80-.).* **311**, 242–245 (2006).

169. M. Toro-Nahuelpan, F. D. Müller, S. Klumpp, J. M. Plitzko, M. Bramkamp, D. Schüler, Segregation of prokaryotic magnetosomes organelles is driven by treadmilling of a dynamic actin-like MamK filament. *BMC Biol.* **14**, 1–24 (2016).

170. G. H. Kelemen, "Intermediate filaments supporting cell shape and growth in bacteria" in *Prokaryotic Cytoskeletons* (Springer, 2017), pp. 161–211.

171. S. Bagchi, H. Tomenius, L. M. Belova, N. Ausmees, Intermediate filament-like proteins in bacteria and a cytoskeletal function in Streptomyces. *Mol. Microbiol.* **70**, 1037–1050 (2008).

172. G. Charbon, M. T. Cabeen, C. Jacobs-Wagner, Bacterial intermediate filaments: in vivo assembly, organization, and dynamics of crescentin. *Genes Dev.* **23**, 1131–1144 (2009).

173. K. T. Sapra, O. Medalia, Bend, push, stretch: Remarkable structure and mechanics of single intermediate filaments and meshworks. *Cells* **10**, 1960 (2021).

174. H. Herrmann, U. Aebi, Intermediate filaments: structure and assembly. *Cold Spring Harb. Perspect. Biol.* **8**, a018242 (2016).

175. J. Walshaw, M. D. Gillespie, G. H. Kelemen, A novel coiled-coil repeat variant in a class of bacterial cytoskeletal proteins. *J. Struct. Biol.* **170**, 202–215 (2010).

176. L. Truebestein, T. A. Leonard, Coiled-coils: The long and short of it. *Bioessays* **38**, 903–916 (2016).

177. A. A. Chernyatina, D. Guzenko, S. V Strelkov, Intermediate filament structure: the bottom-up approach. *Curr. Opin. Cell Biol.* **32**, 65–72 (2015).

178. M. T. Cabeen, H. Herrmann, C. Jacobs-Wagner, The domain organization of the bacterial intermediate filament-like protein crescentin is important for assembly and function. *Cytoskeleton* **68**, 205–219 (2011).

179. M. T. Cabeen, G. Charbon, W. Vollmer, P. Born, N. Ausmees, D. B. Weibel, C. Jacobs-Wagner, Bacterial cell curvature through mechanical control of cell growth. *EMBO J.* **28**, 1208–1219 (2009).

180. M. T. Cabeen, M. A. Murolo, A. Briegel, N. K. Bui, W. Vollmer, N. Ausmees, G. J. Jensen, C. Jacobs-Wagner, Mutations in the lipopolysaccharide biosynthesis pathway interfere with crescentin-mediated cell curvature in Caulobacter crescentus. *J. Bacteriol.* **192**, 3368–3378 (2010).

181. K. Fuchino, S. Bagchi, S. Cantlay, L. Sandblad, D. Wu, J. Bergman, M. Kamali-Moghaddam, K. Flärdh, N. Ausmees, Dynamic gradients of an intermediate filament-like cytoskeleton are recruited by a polarity landmark during apical growth. *Proc. Natl. Acad. Sci.* **110**, E1889–E1897 (2013).

182. N. A. Holmes, J. Walshaw, R. M. Leggett, A. Thibessard, K. A. Dalton, M. D. Gillespie, A. M. Hemmings, B. Gust, G. H. Kelemen, Coiled-coil protein Scy is a key component of a multiprotein assembly controlling polarized growth in Streptomyces. *Proc. Natl. Acad. Sci.* **110**, E397 LP-E406 (2013).

183. B. Ditkowski, N. Holmes, J. Rydzak, M. Donczew, M. Bezulska, K. Ginda, P. Kedzierski, J. Zakrzewska-Czerwińska, G. H. Kelemen, D. Jakimowicz, Dynamic interplay of ParA with

the polarity protein, Scy, coordinates the growth with chromosome segregation in Streptomyces coelicolor. *Open Biol.* **3**, 130006 (2013).

184. M. J. Fröjd, K. Flärdh, Apical assemblies of intermediate filament-like protein FilP are highly dynamic and affect polar growth determinant DivIVA in Streptomyces venezuelae. *Mol. Microbiol.* **112**, 47–61 (2019).

185. J. Lutkenhaus, The ParA/MinD family puts things in their place. *Trends Microbiol.* **20**, 411–418 (2012).

186. D. Mishra, R. Srinivasan, Catching a Walker in the act—DNA partitioning by ParA family of proteins. *Front. Microbiol.* **13**, 856547 (2022).

187. A. O. Maggiolo, S. Mahajan, D. C. Rees, W. M. Clemons, Intradimeric walker A ATPases: conserved features of a functionally diverse family. *J. Mol. Biol.* **435**, 167965 (2023).

188. L. M. Iyer, D. D. Leipe, E. V Koonin, L. Aravind, Evolutionary history and higher order classification of AAA+ ATPases. *J. Struct. Biol.* **146**, 11–31 (2004).

189. M. Latterich, S. Patel, The AAA team: related ATPases with diverse functions. *Trends Cell Biol.* **8**, 65–71 (1998).

190. Z. Hu, J. Lutkenhaus, Topological regulation of cell division in E. coli: spatiotemporal oscillation of MinD requires stimulation of its ATPase by MinE and phospholipid. *Mol. Cell* **7**, 1337–1343 (2001).

191. J.-Y. Bouet, B. E. Funnell, P1 ParA interacts with the P1 partition complex at parS and an ATP–ADP switch controls ParA activities. *EMBO J.* **18**, 1415–1424 (1999).

192. S. Arumugam, Z. Petrášek, P. Schwille, MinCDE exploits the dynamic nature of FtsZ filaments for its spatial regulation. *Proc. Natl. Acad. Sci.* **111**, E1192 LP-E1200 (2014).

193. B. Ramm, P. Glock, J. Mücksch, P. Blumhardt, D. A. García-Soriano, M. Heymann, P. Schwille, The MinDE system is a generic spatial cue for membrane protein distribution in vitro. *Nat. Commun.* **9**, 3942 (2018).

194. T. H. Szeto, S. L. Rowland, L. I. Rothfield, G. F. King, Membrane localization of MinD is mediated by a C-terminal motif that is conserved across eubacteria, archaea, and chloroplasts. *Proc. Natl. Acad. Sci.* **99**, 15693–15698 (2002).

195. W. Wu, K. Park, T. Holyoak, J. Lutkenhaus, Determination of the structure of the MinD–ATP complex reveals the orientation of MinD on the membrane and the relative location of the binding sites for MinE and MinC. *Mol. Microbiol.* **79**, 1515–1528 (2011).

196. A. Dajkovic, G. Lan, S. X. Sun, D. Wirtz, J. Lutkenhaus, MinC spatially controls bacterial cytokinesis by antagonizing the scaffolding function of FtsZ. *Curr. Biol.* **18**, 235–244 (2008).

197. S. A. Flores, M. Howell, J. J. Daniel, R. Piccolo, P. J. B. Brown, Absence of the Min system does not cause major cell division defects in Agrobacterium tumefaciens. *Front. Microbiol.* **9**, 681 (2018).

198. J. Cheng, C. D. Sibley, R. Zaheer, T. M. Finan, A Sinorhizobium meliloti minE mutant has an altered morphology and exhibits defects in legume symbiosis. *Microbiology* **153**, 375–387 (2007).

199. M. Thanbichler, L. Shapiro, MipZ, a Spatial Regulator Coordinating Chromosome Segregation with Cell Division in Caulobacter. *Cell* **126**, 147–162 (2006).

200. D. Kiekebusch, K. A. Michie, L.-O. Essen, J. Löwe, M. Thanbichler, Localized dimerization and nucleoid binding drive gradient formation by the bacterial cell division inhibitor MipZ. *Mol. Cell* **46**, 245–259 (2012).

201. A. L. Marston, H. B. Thomaides, D. H. Edwards, M. E. Sharpe, J. Errington, Polar localization of the MinD protein of Bacillus subtilis and its role in selection of the mid-cell division site. *Genes Dev.* **12**, 3419–3430 (1998).

202. B. Di Ventura, B. Knecht, H. Andreas, W. J. Godinez, M. Fritsche, K. Rohr, W. Nickel, D. W. Heermann, V. Sourjik, Chromosome segregation by the Escherichia coli Min system. *Mol. Syst. Biol.* **9**, 686 (2013).

203. A. S. B. Jalal, T. B. K. Le, Bacterial chromosome segregation by the ParABS system. *Open Biol.* **10** (2020).

204. B. E. Funnell, ParB partition proteins: complex formation and spreading at bacterial and plasmid centromeres. *Front. Mol. Biosci.* **3**, 44 (2016).

205. A. Volante, J. C. Alonso, Molecular anatomy of ParA-ParA and ParA-ParB interactions during plasmid partitioning. *J. Biol. Chem.* **290**, 18782–18795 (2015).

206. J. L. Ptacin, S. F. Lee, E. C. Garner, E. Toro, M. Eckart, L. R. Comolli, W. E. Moerner, L. Shapiro, A spindle-like apparatus guides bacterial chromosome segregation. *Nat. Cell Biol.* **12**, 791–798 (2010).

207. G. Ebersbach, K. Gerdes, Bacterial mitosis: partitioning protein ParA oscillates in spiral-shaped structures and positions plasmids at mid-cell. *Mol. Microbiol.* **52**, 385–398 (2004).

208. M. A. Fogel, M. K. Waldor, A dynamic, mitotic-like mechanism for bacterial chromosome segregation. *Genes Dev.* **20**, 3269–3282 (2006).

209. S. Ringgaard, J. van Zon, M. Howard, K. Gerdes, Movement and equipositioning of plasmids by ParA filament disassembly. *Proc. Natl. Acad. Sci.* **106**, 19369–19374 (2009).

210. K. Gerdes, M. Howard, F. Szardenings, Pushing and pulling in prokaryotic DNA segregation. *Cell* **141**, 927–942 (2010).

211. L. Hu, A. G. Vecchiarelli, K. Mizuuchi, K. C. Neuman, J. Liu, Brownian ratchet mechanisms of ParA-mediated partitioning. *Plasmid* **92**, 12 (2017).

212. L. Hu, A. G. Vecchiarelli, K. Mizuuchi, K. C. Neuman, J. Liu, Directed and persistent movement arises from mechanochemistry of the ParA/ParB system. *Proc. Natl. Acad. Sci.* **112**, E7055–E7064 (2015).

213. L. Hu, A. G. Vecchiarelli, K. Mizuuchi, K. C. Neuman, J. Liu, Brownian ratchet mechanism for faithful segregation of low-copy-number plasmids. *Biophys. J.* **112**, 1489–1502 (2017).

214. A. G. Vecchiarelli, Y. Han, X. Tan, M. Mizuuchi, R. Ghirlando, C. Biertümpfel, B. E. Funnell, K. Mizuuchi, ATP control of dynamic P1 ParA–DNA interactions: a key role for the nucleoid in plasmid partition. *Mol. Microbiol.* **78**, 78–91 (2010).

215. A. G. Vecchiarelli, K. Mizuuchi, B. E. Funnell, Surfing biological surfaces: exploiting the nucleoid for partition and transport in bacteria. *Mol. Microbiol.* **86**, 513–523 (2012).

216. L. C. Hwang, A. G. Vecchiarelli, Y. Han, M. Mizuuchi, Y. Harada, B. E. Funnell, K. Mizuuchi, ParA-mediated plasmid partition driven by protein pattern self-organization.

EMBO J. **32**, 1238–1249 (2013).

217. C. Gogou, A. Japaridze, C. Dekker, Mechanisms for chromosome segregation in bacteria. *Front. Microbiol.*, 1533 (2021).

218. A. M. MacLean, T. M. Finan, M. J. Sadowsky, Genomes of the symbiotic nitrogen-fixing bacteria of legumes. *Plant Physiol.* **144**, 615–622 (2007).

219. S. C. Slater, B. S. Goldman, B. Goodner, J. C. Setubal, S. K. Farrand, E. W. Nester, T. J. Burr, L. Banta, A. W. Dickerman, I. Paulsen, Genome sequences of three Agrobacterium biovars help elucidate the evolution of multichromosome genomes in bacteria. *J. Bacteriol.* **191**, 2501–2511 (2009).

220. U. M. Pinto, K. M. Pappas, S. C. Winans, The ABCs of plasmid replication and segregation. *Nat. Rev. Microbiol.* **10**, 755–765 (2012).

221. S. Tabata, P. J. Hooykaas, A. Oka, Sequence determination and characterization of the replicator region in the tumor-inducing plasmid pTiB6S3. *J. Bacteriol.* **171**, 1665–1672 (1989).

222. K. Żebracki, P. Koper, M. Marczak, A. Skorupska, A. Mazur, Plasmid-encoded RepA proteins specifically autorepress individual repABC operons in the multipartite Rhizobium leguminosarum bv. trifolii genome. *PLoS One* **10**, e0131907 (2015).

223. M. A. Ramírez-Romero, J. Téllez-Sosa, H. Barrios, A. Pérez-Oseguera, V. Rosas, M. A. Cevallos, RepA negatively autoregulates the transcription of the repABC operon of the Rhizobium etli symbiotic plasmid basic replicon. *Mol. Microbiol.* **42**, 195–204 (2001).

224. M. A. Ramírez-Romero, N. Soberón, A. Pérez-Oseguera, J. Téllez-Sosa, M. A. Cevallos, Structural elements required for replication and incompatibility of the Rhizobium etli symbiotic plasmid. *J. Bacteriol.* **182**, 3117–3124 (2000).

225. P. Koper, K. Żebracki, M. Marczak, A. Skorupska, A. Mazur, RepB proteins of the multipartite Rhizobium leguminosarum bv. trifolii genome discriminate between centromere-like parS sequences for plasmid segregational stability. *Mol. Microbiol.* **102**, 446–466 (2016).

226. K.-B. Lee, P. De Backer, T. Aono, C.-T. Liu, S. Suzuki, T. Suzuki, T. Kaneko, M. Yamada, S. Tabata, D. M. Kupfer, The genome of the versatile nitrogen fixer Azorhizobium caulinodans ORS571. *BMC Genomics* **9**, 1–14 (2008).

227. C.-T. Liu, K.-B. Lee, Y.-S. Wang, M.-H. Peng, K.-T. Lee, S. Suzuki, T. Suzuki, H. Oyaizu, Involvement of the azorhizobial chromosome partition gene (parA) in the onset of bacteroid differentiation during Sesbania rostrata stem nodule development. *Appl. Environ. Microbiol.* **77**, 4371–4382 (2011).

228. J.-V. Höltje, Growth of the stress-bearing and shape-maintaining murein sacculus of Escherichia coli. *Microbiol. Mol. Biol. Rev.* **62**, 181–203 (1998).

229. W. Vollmer, D. Blanot, M. A. De Pedro, Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.* **32**, 149–167 (2008).

230. T. J. Silhavy, D. Kahne, S. Walker, The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* **2**, a000414 (2010).

231. A. J. F. Egan, R. M. Cleverley, K. Peters, R. J. Lewis, W. Vollmer, Regulation of bacterial cell wall growth. *FEBS J.* **284**, 851–867 (2017).

232. S. Booth, R. J. Lewis, Structural basis for the coordination of cell division with the synthesis of the bacterial cell envelope. *Protein Sci.* **28**, 2042–2054 (2019).

233. M. G. Pinho, M. Kjos, J.-W. Veening, How to get (a) round: mechanisms controlling growth and division of coccoid bacteria. *Nat. Rev. Microbiol.* **11**, 601–614 (2013).

234. F. Cava, E. Kuru, Y. V Brun, M. A. de Pedro, Modes of cell wall growth differentiation in rod-shaped bacteria. *Curr. Opin. Microbiol.* **16**, 731–737 (2013).

235. S. Garde, P. K. Chodisetti, M. Reddy, Peptidoglycan: structure, synthesis, and regulation. *EcoSal Plus* **9** (2021).

236. T. A. Cameron, W. Margolin, Insights into the assembly and regulation of the bacterial divisome. *Nat. Rev. Microbiol.* **22**, 33–45 (2024).

237. A. J. F. Egan, J. Errington, W. Vollmer, Regulation of peptidoglycan synthesis and remodelling. *Nat. Rev. Microbiol.* **18**, 446–460 (2020).

238. T. Kruse, J. Bork-Jensen, K. Gerdes, The morphogenetic MreBCD proteins of Escherichia coli form an essential membrane-bound complex. *Mol. Microbiol.* **55**, 78–89 (2005).

239. R. Nygaard, C. L. B. Graham, M. Belcher Dufrisne, J. D. Colburn, J. Pepe, M. A. Hydorn, S. Corradi, C. M. Brown, K. U. Ashraf, O. N. Vickery, Structural basis of peptidoglycan synthesis by E. coli RodA-PBP2 complex. *Nat. Commun.* **14**, 5151 (2023).

240. X. Liu, J. Biboy, E. Consoli, W. Vollmer, T. den Blaauwen, MreC and MreD balance the interaction between the elongasome proteins PBP2 and RodA. *PLoS Genet.* **16**, e1009276 (2020).

241. R. Ago, D. Shiomi, RodZ: a key-player in cell elongation and cell division in Escherichia coli. *AIMS Microbiol.* **5**, 358 (2019).

242. F. O. Bendezú, C. A. Hale, T. G. Bernhardt, P. A. J. De Boer, RodZ (YfgA) is required for proper assembly of the MreB actin cytoskeleton and cell shape in E. coli. *EMBO J.* **28**, 193–204 (2009).

243. A. Typas, M. Banzhaf, B. van den B. van Saparoea, J. Verheul, J. Biboy, R. J. Nichols, M. Zietek, K. Beilharz, K. Kannenberg, M. von Rechenberg, Regulation of peptidoglycan synthesis by outer-membrane proteins. *Cell* **143**, 1097–1109 (2010).

244. M. Banzhaf, B. van den Berg van Saparoea, M. Terrak, C. Fraipont, A. Egan, J. Philippe, A. Zapun, E. Breukink, M. Nguyen-Distèche, T. den Blaauwen, Cooperativity of peptidoglycan synthases active in bacterial cell elongation. *Mol. Microbiol.* **85**, 179–194 (2012).

245. L. Potluri, A. Karczmarek, J. Verheul, A. Piette, J. Wilkin, N. Werth, M. Banzhaf, W. Vollmer, K. D. Young, M. Nguyen-Distèche, Septal and lateral wall localization of PBP5, the major D, D-carboxypeptidase of Escherichia coli, requires substrate recognition and membrane attachment. *Mol. Microbiol.* **77**, 300–323 (2010).

246. A. Bouhss, A. E. Trunkfield, T. D. H. Bugg, D. Mengin-Lecreulx, The biosynthesis of peptidoglycan lipid-linked intermediates. *FEMS Microbiol. Rev.* **32**, 208–233 (2007).

247. M. E. G. Aarsman, A. Piette, C. Fraipont, T. M. F. Vinkenvleugel, M. Nguyen-Distèche, T. den Blaauwen, Maturation of the Escherichia coli divisome occurs in two steps. *Mol. Microbiol.* **55**, 1631–1645 (2005).

248. S. Du, J. Lutkenhaus, Assembly and activation of the Escherichia coli divisome. *Mol.*

Microbiol. **105**, 177–187 (2017).

249. S. Pichoff, J. Lutkenhaus, Tethering the Z ring to the membrane through a conserved membrane targeting sequence in FtsA. *Mol. Microbiol.* **55**, 1722–1734 (2005).

250. F. J. Gueiros-Filho, R. Losick, A widely conserved bacterial cell division protein that promotes assembly of the tubulin-like protein FtsZ. *Genes Dev.* **16**, 2544–2556 (2002).

251. S. Du, S. Pichoff, J. Lutkenhaus, FtsEX acts on FtsA to regulate divisome assembly and activity. *Proc. Natl. Acad. Sci.* **113**, E5052–E5061 (2016).

252. N. W. Goehring, J. Beckwith, Diverse paths to midcell: assembly of the bacterial cell division machinery. *Curr. Biol.* **15**, R514–R526 (2005).

253. D. K. Ranjit, M. A. Jorgenson, K. D. Young, PBP1B glycosyltransferase and transpeptidase activities play different essential roles during the de novo regeneration of rod morphology in Escherichia coli. *J. Bacteriol.* **199**, 10–1128 (2017).

254. A. Taguchi, M. A. Welsh, L. S. Marmont, W. Lee, M. Sjodt, A. C. Kruse, D. Kahne, T. G. Bernhardt, S. Walker, FtsW is a peptidoglycan polymerase that is functional only in complex with its cognate penicillin-binding protein. *Nat. Microbiol.* **4**, 587–594 (2019).

255. N. T. Peters, T. Dinh, T. G. Bernhardt, A fail-safe mechanism in the septal ring assembly pathway generated by the sequential recruitment of cell separation amidases and their activators. *J. Bacteriol.* **193**, 4973–4983 (2011).

256. P. Muller, C. Ewers, U. Bertsche, M. Anstett, T. Kallis, E. Breukink, C. Fraipont, M. Terrak, M. Nguyen-Disteche, W. Vollmer, The essential cell division protein FtsN interacts with the murein (peptidoglycan) synthase PBP1B in Escherichia coli. *J. Biol. Chem.* **282**, 36394–36402 (2007).

257. B. Liu, L. Persons, L. Lee, P. A. J. de Boer, Roles for both FtsA and the FtsBLQ subcomplex in FtsN-stimulated cell constriction in E scherichia coli. *Mol. Microbiol.* **95**, 945–970 (2015).

258. S. J. Craven, S. G. F. Condon, A. Senes, A model of the interactions between the FtsQLB and the FtsWI peptidoglycan synthase complex in bacterial cell division. *bioRxiv*, 2010–2022 (2022).

259. L. S. Marmont, T. G. Bernhardt, A conserved subcomplex within the bacterial cytokinetic ring activates cell wall synthesis by the FtsW-FtsI synthase. *Proc. Natl. Acad. Sci.* **117**, 23879–23885 (2020).

260. A. Typas, M. Banzhaf, C. A. Gross, W. Vollmer, From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nat. Rev. Microbiol.* **10**, 123–136 (2012).

261. A. J. F. Egan, Bacterial outer membrane constriction. *Mol. Microbiol.* **107**, 676–687 (2018).

262. A. A. Yakhnina, T. G. Bernhardt, The Tol-Pal system is required for peptidoglycan-cleaving enzymes to complete bacterial cell division. *Proc. Natl. Acad. Sci.* **117**, 6777–6783 (2020).

263. M. A. Gerding, Y. Ogata, N. D. Pecora, H. Niki, P. A. J. De Boer, The trans-envelope Tol–Pal complex is part of the cell division machinery and required for proper outer-membrane invagination during cell constriction in E. coli. *Mol. Microbiol.* **63**, 1008–1025 (2007).

264. C. A. Hale, L. Persons, P. A. J. de Boer, Recruitment of the TolA protein to cell constriction sites in Escherichia coli via three separate mechanisms, and a critical role for FtsWI activity in recruitment of both TolA and TolQ. *J. Bacteriol.* **204**, e00464-21 (2022).

265. P. Baccelli, R. Rachedi, B. Serrano, M. Petiti, C. S. Bernard, L. Houot, D. Duche, Timing of TolA and TolQ recruitment at the septum depends on the functionality of the Tol-Pal system. *J. Mol. Biol.* **434**, 167519 (2022).

266. J.-E. Hugonnet, D. Mengin-Lecreulx, A. Monton, T. den Blaauwen, E. Carbonnelle, C. Veckerle, Y. Brun V, M. van Nieuwenhze, C. Bouchier, K. Tu, Factors essential for L, D-transpeptidase-mediated peptidoglycan cross-linking and β-lactam resistance in Escherichia coli. *Elife* **5**, e19469 (2016).

267. K. J. Kieser, E. J. Rubin, How sisters grow apart: mycobacterial growth and division. *Nat. Rev. Microbiol.* **12**, 550–562 (2014).

268. J. W. Sher, H. C. Lim, T. G. Bernhardt, Polar growth in Corynebacterium glutamicum has a flexible Cell Wall synthase requirement. *MBio* **12**, 10–1128 (2021).

269. E. Kuru, H. V. Hughes, P. J. Brown, E. Hall, S. Tekkam, F. Cava, M. A. de Pedro, Y. V Brun, M. S. VanNieuwenhze, In situ probing of newly synthesized peptidoglycan in live bacteria with fluorescent D-amino acids. *Angew. Chemie* **124**, 12687–12691 (2012).

270. X. Meniche, R. Otten, M. S. Siegrist, C. E. Baer, K. C. Murphy, C. R. Bertozzi, C. M. Sassetti, Subpolar addition of new cell wall is directed by DivIVA in mycobacteria. *Proc. Natl. Acad. Sci.* **111**, E3243–E3251 (2014).

271. F. M. Meyer, M. Bramkamp, Cell wall synthesizing complexes in Mycobacteriales. *Curr. Opin. Microbiol.* **79**, 102478 (2024).

272. H. Botella, G. Yang, O. Ouerfelli, S. Ehrt, C. F. Nathan, J. Vaubourgeix, Distinct spatiotemporal dynamics of peptidoglycan synthesis between Mycobacterium smegmatis and Mycobacterium tuberculosis. *MBio* **8**, 10–1128 (2017).

273. R. Grangeon, J. R. Zupan, J. Anderson-Furgeson, P. C. Zambryski, PopZ identifies the new pole, and PodJ identifies the old pole during polar growth in Agrobacterium tumefaciens. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 11666–11671 (2015).

274. T. A. Cameron, J. Anderson-Furgeson, J. R. Zupan, J. J. Zik, P. C. Zambryski, Peptidoglycan synthesis machinery in Agrobacterium tumefaciens during unipolar growth and cell division. *MBio* **5**, e01219--14 (2014).

275. D. T. Kysela, P. J. B. Brown, K. C. Huang, Y. V Brun, Biological consequences and advantages of asymmetric bacterial growth. *Annu. Rev. Microbiol.* **67**, 417–435 (2013).

276. J. Amstutz, E. Krol, A. Verhaeghe, X. De Bolle, A. Becker, P. J. Brown, Getting to the point: unipolar growth of Hyphomicrobiales. *Curr. Opin. Microbiol.* **79**, 102470 (2024).

277. M. Letek, E. Ordóñez, J. Vaquera, W. Margolin, K. Flärdh, L. M. Mateos, J. A. Gil, DivIVA is required for polar growth in the MreB-lacking rod-shaped actinomycete Corynebacterium glutamicum. *J. Bacteriol.* **190**, 3283–3292 (2008).

278. J. R. Zupan, R. Grangeon, J. S. Robalino-Espinosa, N. Garnica, P. Zambryski, GROWTH POLE RING protein forms a 200-nm-diameter ring structure essential for polar growth and rod shape in Agrobacterium tumefaciens. *Proc. Natl. Acad. Sci. U. S. A.* **166**, 10962–10967 (2019).

279. J. Zupan, Z. Guo, T. Biddle, P. Zambryski, Agrobacterium tumefaciens Growth Pole Ring Protein: C Terminus and Internal Apolipoprotein Homologous Domains Are Essential for Function and Subcellular Localization. *MBio* **12**, 10.1128/mbio.00764-21 (2021).

280. E. Krol, H. C. L. Yau, M. Lechner, S. Schäper, G. Bange, W. Vollmer, A. Becker, Tol-Pal system and Rgs proteins interact to promote unipolar growth and cell division in Sinorhizobium meliloti. *MBio* **11**, e00306-20 (2020).

281. E. Krol, L. Stuckenschneider, J. M. Kästle Silva, P. L. Graumann, A. Becker, Stable inheritance of Sinorhizobium meliloti cell growth polarity requires an FtsN-like protein and an amidase. *Nat. Commun.* **12**, 545 (2021).

282. R. Grangeon, J. Zupan, Y. Jeon, P. C. Zambryski, Loss of PopZAt activity in agrobacterium tumefaciens by deletion or depletion leads to multiple growth poles, minicells, and growth defects. *MBio* **8** (2017).

283. S. Schäper, H. C. L. Yau, E. Krol, D. Skotnicka, T. Heimerl, J. Gray, V. Kaever, L. Søgaard-Andersen, W. Vollmer, A. Becker, Seven-transmembrane receptor protein RgsP and cell wall-binding protein RgsM promote unipolar growth in Rhizobiales. *PLoS Genet.* **14**, e1007594 (2018).

284. A. Yahashiri, M. A. Jorgenson, D. S. Weiss, The SPOR domain, a widely conserved peptidoglycan binding domain that targets proteins to the site of cell division. *J. Bacteriol.* **199**, 10–1128 (2017).

285. W. M. Figueroa-Cuilan, A. M. Randich, C. M. Dunn, G. Santiago-Collazo, A. Yowell, P. J. B. Brown, Diversification of LytM protein functions in polar elongation and cell division of Agrobacterium tumefaciens. *Front. Microbiol.* **12**, 729307 (2021).

286. S. Schäper, E. Krol, D. Skotnicka, V. Kaever, R. Hilker, L. Søgaard-Andersen, A. Becker, Cyclic di-GMP regulates multiple cellular functions in the symbiotic alphaproteobacterium Sinorhizobium meliloti. *J. Bacteriol.* **198**, 521–535 (2016).

287. M. A. Williams, A. Aliashkevich, E. Krol, E. Kuru, J. M. Bouchier, J. Rittichier, Y. V Brun, M. S. VanNieuwenhze, A. Becker, F. Cava, Unipolar peptidoglycan synthesis in the Rhizobiales requires an essential class A penicillin-binding protein. *MBio* **12**, 10–1128 (2021).

288. A. Aliashkevich, F. Cava, LD-transpeptidases: the great unknown among the peptidoglycan cross-linkers. *FEBS J.* **289**, 4718–4730 (2022).

289. P. J. B. Brown, D. T. Kysela, Y. V Brun, Polarity and the diversity of growth mechanisms in bacteria. *Semin. Cell Dev. Biol.* **22**, 790–798 (2011).

290. B. M. Davis, M. K. Waldor, Establishing polar identity in gram-negative rods. *Curr. Opin. Microbiol.* **16**, 752–759 (2013).

291. G. Laloux, C. Jacobs-Wagner, How do bacteria localize proteins to the cell pole? *J. Cell Sci.* **127**, 11–19 (2014).

292. K. S. Ramamurthi, R. Losick, Negative membrane curvature as a cue for subcellular localization of a bacterial protein. *Proc. Natl. Acad. Sci.* **106**, 13541–13545 (2009).

293. S. Govindarajan, K. Nevo-Dinur, O. Amster-Choder, Compartmentalization and spatiotemporal organization of macromolecules in bacteria. *FEMS Microbiol. Rev.* **36**, 1005–1022 (2012).

294. G. Ebersbach, A. Briegel, G. J. Jensen, C. Jacobs-Wagner, A Self-Associating Protein

Critical for Chromosome Attachment, Division, and Polar Organization in Caulobacter. *Cell* **134**, 956–968 (2008).

295. S. Saberi, E. Emberly, Non-equilibrium polar localization of proteins in bacterial cells. *PLoS One* **8**, e64075 (2013).

296. K. Flärdh, Essential role of DivIVA in polar growth and morphogenesis in Streptomyces coelicolor A3 (2). *Mol. Microbiol.* **49**, 1523–1536 (2003).

297. L. R. Hammond, M. L. White, P. J. Eswara, ¡ vIVA la DivIVA! *J. Bacteriol.* **201**, 10–1128 (2019).

298. A. M. Hempel, S. Wang, M. Letek, J. A. Gil, K. Flärdh, Assemblies of DivIVA mark sites for hyphal branching and can establish new zones of cell wall growth in Streptomyces coelicolor. *J. Bacteriol.* **190**, 7579–7583 (2008).

299. M. A. Oliva, S. Halbedel, S. M. Freund, P. Dutow, T. A. Leonard, D. B. Veprintsev, L. W. Hamoen, J. Löwe, Features critical for membrane binding revealed by DivIVA crystal structure. *EMBO J.* **29**, 1988–2001 (2010).

300. H. Stahlberg, E. Kutejova, K. Muchova, M. Gregorini, A. Lustig, S. A. Müller, V. Olivieri, A. Engel, A. J. Wilkinson, I. Barak, Oligomeric structure of the Bacillus subtilis cell division protein DivIVA determined by transmission electron microscopy. *Mol. Microbiol.* **52**, 1281–1290 (2004).

301. R. Lenarcic, S. Halbedel, L. Visser, M. Shaw, L. J. Wu, J. Errington, D. Marenduzzo, L. W. Hamoen, Localisation of DivIVA by targeting to negatively curved membranes. *EMBO J.* **28**, 2272–2282 (2009).

302. D. Jakimowicz, P. Żydek, A. Kois, J. Zakrzewska-Czerwińska, K. F. Chater, Alignment of multiple chromosomes along helical ParA scaffolding in sporulating Streptomyces hyphae. *Mol. Microbiol.* **65**, 625–641 (2007).

303. Y. V Burn, R. Janakiraman, The dimorphic life cycle of Caulobacter and stalked bacteria. *Prokaryotic Dev.*, 295–317 (1999).

304. S. K. Govers, C. Jacobs-Wagner, Caulobacter crescentus: model system extraordinaire. *Curr. Biol.* **30**, R1151–R1158 (2020).

305. E. A. Klein, S. Schlimpert, V. Hughes, Y. V Brun, M. Thanbichler, Z. Gitai, Physiological role of stalk lengthening in Caulobacter crescentus. *Commun. Integr. Biol.* **6**, e24561 (2013).

306. P. D. Curtis, Y. V. Brun, Getting in the Loop: Regulation of Development in Caulobacter crescentus. *Microbiol. Mol. Biol. Rev.* **74**, 13–41 (2010).

307. E. Huitema, S. Pritchard, D. Matteson, S. K. Radhakrishnan, P. H. Viollier, Bacterial birth scar proteins mark future flagellum assembly site. *Cell* **124**, 1025–1037 (2006).

308. H. Lam, W. B. Schofield, C. Jacobs-Wagner, A landmark protein essential for establishing and perpetuating the polarity of a bacterial cell. *Cell* **124**, 1011–1023 (2006).

309. G. R. Bowman, L. R. Comolli, J. Zhu, M. Eckart, M. Koenig, K. H. Downing, W. E. Moerner, T. Earnest, L. Shapiro, A Polymeric Protein Anchors the Chromosomal Origin/ParB Complex at a Bacterial Cell Pole. *Cell* **134**, 945–955 (2008).

310. Y. C. Yeh, L. R. Comolli, K. H. Downing, L. Shapiro, H. H. McAdams, The Caulobacter Tol-Pal complex is essential for outer membrane integrity and the positioning of a polar

localization factor. *J. Bacteriol.* **192**, 4847–4858 (2010).

311. J. A. Holmes, S. E. Follett, H. Wang, C. P. Meadows, K. Varga, G. R. Bowman, Caulobacter PopZ forms an intrinsically disordered hub in organizing bacterial cell poles. *Proc. Natl. Acad. Sci.* **113**, 12490–12495 (2016).

312. A. M. Perez, T. H. Mann, K. Lasker, D. G. Ahrens, M. R. Eckart, L. Shapiro, A localized complex of two protein oligomers controls the orientation of cell polarity. *MBio* **8**, 10–1128 (2017).

313. G. R. Bowman, L. R. Comolli, G. M. Gaietta, M. Fero, S. Hong, Y. Jones, J. H. Lee, K. H. Downing, M. H. Ellisman, H. H. McAdams, Caulobacter PopZ forms a polar subdomain dictating sequential changes in pole composition and function. *Mol. Microbiol.* **76**, 173–189 (2010).

314. J. L. Ptacin, A. Gahlmann, G. R. Bowman, A. M. Perez, L. von Diezmann, M. R. Eckart, W. E. Moerner, L. Shapiro, Bacterial scaffold directs pole-specific centromere segregation. *Proc. Natl. Acad. Sci.* **111**, E2046–E2055 (2014).

315. W. B. Schofield, H. C. Lim, C. Jacobs-Wagner, Cell cycle coordination and regulation of bacterial chromosome segregation dynamics by polarly localized proteins. *EMBO J.* **29**, 3068–3081 (2010).

316. N. Lu, S. W. Duvall, G. Zhao, K. A. Kowallis, C. Zhang, W. Tan, J. Sun, H. N. Petitjean, D. T. Tomares, G.-P. Zhao, Scaffold-Scaffold Interaction Facilitates Cell Polarity Development in Caulobacter crescentus. *MBio* **14**, e03218-22 (2023).

317. P. H. Viollier, N. Sternheim, L. Shapiro, Identification of a localization factor for the polar positioning of bacterial structural and regulatory proteins. *Proc. Natl. Acad. Sci.* **99**, 13831–13836 (2002).

318. A. J. Hinz, D. E. Larson, C. S. Smith, Y. V Brun, The Caulobacter crescentus polar organelle development protein PodJ is differentially localized and is required for polar targeting of the PleC development regulator. *Mol. Microbiol.* **47**, 929–941 (2003).

319. G. Laloux, C. Jacobs-Wagner, Spatiotemporal control of PopZ localization through cell cycle-coupled multimerization. *J. Cell Biol.* **201**, 827–841 (2013).

320. K. Sundararajan, E. D. Goley, Cytoskeletal proteins in Caulobacter crescentus: spatial orchestrators of cell cycle progression, development, and cell shape. *Prokaryotic Cytoskelet. filamentous protein Polym. Act. cytoplasm Bact. archaeal cells*, 103–137 (2017).

321. A. T. Fields, C. S. Navarrete, A. Z. Zare, Z. Huang, M. Mostafavi, J. C. Lewis, Y. Rezaeihaghighi, B. J. Brezler, S. Ray, A. L. Rizzacasa, The conserved polarity factor PodJ1 impacts multiple cell envelope-associated functions in Sinorhizobium meliloti. *Mol. Microbiol.* **84**, 892–920 (2012).

322. J. C. Anderson-Furgeson, J. R. Zupan, R. Grangeon, P. C. Zambryski, Loss of PodJ in Agrobacterium tumefaciens leads to ectopic polar growth, branching, and reduced cell division. *J. Bacteriol.* **198**, 1883–1891 (2016).

323. Z. Ren, Q. Liao, I. S. Barton, E. E. Wiesler, C. Fuqua, X. Wang, Centromere interactions promote the maintenance of the multipartite genome in Agrobacterium tumefaciens. mBio 13: e00508-22. (2022).

324. J. S. Robalino-Espinosa, J. R. Zupan, A. Chavez-Arroyo, P. Zambryski, Segregation of four

Agrobacterium tumefaciens replicons during polar growth: PopZ and PodJ control segregation of essential replicons. *Proc. Natl. Acad. Sci.* **117**, 26366–26373 (2020).

325. M. Howell, A. Aliashkevich, A. K. Salisbury, F. Cava, G. R. Bowman, P. J. B. Brown, Absence of the Polar Organizing Protein PopZ Results in Reduced and Asymmetric Cell Division in Agrobacterium tumefaciens. *J. Bacteriol.* **199**, 101–118 (2017).

326. H. M. Ehrle, J. T. Guidry, R. Iacovetto, A. K. Salisbury, D. J. Sandidge, G. R. Bowman, Polar Organizing Protein PopZ Is Required for Chromosome Segregation in Agrobacterium tumefaciens. *J. Bacteriol.* **199** (2017).

327. G. R. Bowman, A. M. Perez, J. L. Ptacin, E. Ighodaro, E. Folta-Stogniew, L. R. Comolli, L. Shapiro, Oligomerization and higher-order assembly contribute to sub-cellular localization of a bacterial scaffold. *Mol. Microbiol.* **90**, 776–795 (2013).

328. W. Margolin, J. C. Corbo, S. R. Long, Cloning and characterization of a Rhizobium meliloti homolog of the Escherichia coli cell division gene ftsZ. *J. Bacteriol.* **173**, 5822–5830 (1991).

329. J. P. W. Young, L. C. Crossman, A. W. B. Johnston, N. R. Thomson, Z. F. Ghazoui, K. H. Hull, M. Wexler, A. R. J. Curson, J. D. Todd, P. S. Poole, T. H. Mauchline, A. K. East, M. A. Quail, C. Churcher, C. Arrowsmith, I. Cherevach, T. Chillingworth, K. Clarke, A. Cronin, P. Davis, A. Fraser, Z. Hance, H. Hauser, K. Jagels, S. Moule, K. Mungall, H. Norbertczak, E. Rabbinowitsch, M. Sanders, M. Simmonds, S. Whitehead, J. Parkhill, The genome of Rhizobium leguminosarum has recognizable core and accessory components. *Genome Biol.* **7** (2006).

330. D. W. Hook, S. J. Porter, C. Herzog, Dimensions: Building Context for Search and Evaluation. *Front. Res. Metrics Anal.* **3**, 23 (2018).

331. A. Mazur, P. Koper, Rhizobial plasmids - replication, structure and biological role. *Cent. Eur. J. Biol.* **7**, 571–586 (2012).

332. M. G. López-Guerrero, E. Ormeño-Orrillo, J. L. Acosta, A. Mendoza-Vargas, M. A. Rogel, M. A. Ramírez, M. Rosenblueth, J. Martínez-Romero, E. Martínez-Romero, Rhizobial extrachromosomal replicon variability, stability and expression in natural niches. *Plasmid* **68**, 149–158 (2012).

333. V. Gonzales, R. I. Santamarı, A. Medrano-soto, P. Bustos, I. Herna, G. Moreno-hagelsieb, S. C. Janga, M. A. Ramı, J. Collado-vides, G. Da, The partitioned Rhizobium etli genome : Genetic and metabolic redundancy in seven interacting replicons. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 3834–3839 (2006).

334. A. W. B. Johnston, J. E. Beringer, Identification of the Rhizobium Strains in Pea Root Nodules Using Genetic Markers. *Artic. J. Gen. Microbiol.* **87**, 343–350 (1975).

335. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

336. C. T. Nordyke, Y. M. Ahmed, R. Z. Puterbaugh, G. R. Bowman, K. Varga, Intrinsically disordered bacterial polar organizing protein Z, PopZ, interacts with protein binding partners through an N-terminal Molecular Recognition Feature. *J. Mol. Biol.* **432**, 6092–6107 (2020).

337. A. D. McLachlan, M. Stewart, The 14-fold periodicity in α-tropomyosin and the interaction with actin. *J. Mol. Biol.* **103**, 271–298 (1976).

338. M. Bernhofer, C. Dallago, T. Karl, V. Satagopam, M. Heinzinger, M. Littmann, T. Olenyi, J. Qiu, K. Schütze, G. Yachdav, PredictProtein-predicting protein structure and function for 29 years. *Nucleic Acids Res.* **49**, W535–W540 (2021).

339. J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G. A. Salazar, E. L. L. Sonnhammer, S. C. E. Tosatto, L. Paladin, S. Raj, L. J. Richardson, Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).

340. T. Paysan-Lafosse, M. Blum, S. Chuguransky, T. Grego, B. L. Pinto, G. A. Salazar, M. L. Bileschi, P. Bork, A. Bridge, L. Colwell, InterPro in 2022. *Nucleic Acids Res.* **51**, D418–D427 (2023).

341. A. Le Gall, D. I. Cattoni, B. Guilhas, C. Mathieu-Demazière, L. Oudjedi, J.-B. Fiche, J. Rech, S. Abrahamsson, H. Murray, J.-Y. Bouet, Bacterial partition complexes segregate within the volume of the nucleoid. *Nat. Commun.* **7**, 12107 (2016).

342. H. C. Lim, I. V. Surovtsev, B. G. Beltran, F. Huang, J. Bewersdorf, C. Jacobs-Wagner, Evidence for a DNA-relay mechanism in ParABS-mediated chromosome segregation. *Elife* **2014** (2014).

343. S. L. Turner, J. P. W. Young, The replicator region of the Rhizobium leguminosarum cryptic plasmid pRL8JI. *FEMS Microbiol. Lett.* **133**, 53–58 (1995).

344. C. Bignell, C. M. Thomas, The bacterial ParA-ParB partitioning proteins. *J. Biotechnol.* **91**, 1–34 (2001).

345. A. R. Glenn, P. S. Poole, J. F. Hudman, Succinate uptake by free-living and bacteroid forms of Rhizobium leguminosarum. *J. Gen. Microbiol.* **119**, 267–271 (1980).

346. D. Hanahan, Studies on transformation of Escherichia coli with plasmids. *J. Mol. Biol.* **166**, 557–580 (1983).

347. R. G. Taylor, D. C. Walker, R. R. McInnes, E. coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Res.* **21**, 1677–1678 (1993).

348. D. H. Figurski, D. R. Helinski, Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 1648–1652 (1979).

349. K. Gust, B., Kieser, T., Charter, PCR targeting system in Streptomyces coelicolor A3. *John Innes Cent.* **3**, 1–38 (2002).

350. T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, H. Mori, Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).

351. K. A. Datsenko, B. L. Wanner, One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci.* **97**, 6640–6645 (2000).

352. J. H. Miller, *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1972).

353. J. E. BERINGER, R Factor Transfer in Rhizobium leguminosarum. *Microbiology* **84**, 188–198 (1974).

354. A. Schäfer, A. Tauch, W. Jäger, J. Kalinowski, G. Thierbach, A. Pühler, Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and

pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum. *Gene* **145**, 69–73 (1994).

355. A. C. Ind, S. L. Porter, M. T. Brown, E. D. Byles, J. A. de Beyer, S. A. Godfrey, J. P. Armitage, Inducible-expression plasmid for Rhodobacter sphaeroides and Paracoccus denitrificans. *Appl. Environ. Microbiol.* **75**, 6613–6615 (2009).

356. R. Simon, U. Priefer, A. Pühler, A Broad Host Range Mobilization System for In Vivo Genetic Engineering: Transposon Mutagenesis in Gram Negative Bacteria. *Bio/Technology* **1**, 784–791 (1983).

357. S. Mesa, E. J. Bedmar, A. Chanfon, H. Hennecke, H. M. Fischer, Bradyrhizobium japonicum NnrR, a denitrification regulator, expands the FixLJ-FixK2 regulatory cascade. *J. Bacteriol.* **185**, 3978–3982 (2003).

358. B. Gust, G. L. Challis, K. Fowler, T. Kieser, K. F. Chater, PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci.* **100**, 1541–1546 (2003).

359. E. Weber, C. Engler, R. Gruetzner, S. Werner, S. Marillonnet, A modular cloning system for standardized assembly of multigene constructs. *PLoS One* **6**, e16765 (2011).

360. A. Y. Chang, V. W. Chau, J. A. Landas, Yvonne, Preparation of calcium competent Escherichia coli and heat-shock transformation. *J. Exp. Microbiol. Immunol.* **1**, 22–25 (2017).

361. S. Skinner, L. Sepúlveda, H. Xu, I. Golding, Corrigendum: Measuring mRNA copy number in individual Escherichia coli cells using single-molecule fluorescent in situ hybridization. *Nat. Protoc.* **8**, 1100–1113 (2013).

362. U. Fels, K. Gevaert, P. Van Damme, Bacterial Genetic Engineering by Means of Recombineering for Reverse Genetics. *Front. Microbiol.* **11**, 548410 (2020).

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

364. J. Tarsitano, L. Y. Ramis, L. G. Alonso, D. M. Russo, A. Zorreguieta, RapD Is a Multimeric Calcium-Binding Protein That Interacts With the Rhizobium leguminosarum Biofilm Exopolysaccharide, Influencing the Polymer Lengths. *Front. Microbiol.* **13** (2022).

365. L. Rubio-Sanz, R. I. Prieto, J. Imperial, J. M. Palacios, B. Brito, Functional and Expression Analysis of the Metal-Inducible dmeRF System from Rhizobium leguminosarum bv. viciae. *Appl. Environ. Microbiol.* **79**, 6414–6422 (2013).

366. B. Belén, P. Rosa-Isabel, C. Ezequiel, M.-B. Marie-Andrée, I. Juan, R.-A. Tomás, P. José-Manuel, Rhizobium leguminosarum hupE Encodes a Nickel Transporter Required for Hydrogenase Activity. *J. Bacteriol.* **192**, 925–935 (2010).

367. M. Ballesteros-Gutiérrez, M. Albareda, C. Barbas, Á. López-Gonzálvez, M. F. Dunn, J. M. Palacios, A host-specific diaminobutyrate aminotransferase contributes to symbiotic performance, homoserine metabolism, and competitiveness in the Rhizobium leguminosarum/Pisum sativum system. *Front. Microbiol.* **14** (2023).

368. R. L. Yenofsky, M. Fine, J. W. Pellow, A mutant neomycin phosphotransferase II gene reduces the resistance of transformants to antibiotic selection pressure. *Proc. Natl.*

Acad. Sci. **87**, 3435–3439 (1990).

369. P. Gay, L. C. D, M. Steinmetz, E. Ferrari, J. A. Hoch, Cloning structural gene sacB, which codes for exoenzyme levansucrase of Bacillus subtilis: expression of the gene in Escherichia coli. *J. Bacteriol.* **153**, 1424–1431 (1983).

370. M. Steinmetz, D. Le Coq, H. Ben Djemia, P. Gay, Genetic analysis of sacB, the structural gene of a secreted enzyme, levansucrase of Bacillus subtilis Marburg. *Mol. Gen. Genet. MGG* **191**, 138–144 (1983).

371. Y. H. Fu, M. M. Tsai, Y. N. Luo, R. C. Deonier, Deletion analysis of the F plasmid oriT locus. *J. Bacteriol.* **173**, 1012–1020 (1991).

372. Y. Zhang, F. Buchholz, J. P. P. Muyrers, A. F. Stewart, A new logic for DNA engineering using recombination in Escherichia coli. *Nat. Genet.* **20**, 123–128 (1998).

373. M. J. Torres, E. Bueno, A. Jiménez-Leiva, J. J. Cabrera, E. J. Bedmar, S. Mesa, M. J. Delgado, FixK(2) Is the Main Transcriptional Activator of Bradyrhizobium diazoefficiens nosRZDYFLX Genes in Response to Low Oxygen. *Front. Microbiol.* **8**, 1621 (2017).

374. M. Janczarek, J. Król, A. Skorupska, The pssB gene product of Rhizobium leguminosarum bv. trifolii is homologous to a family of inositol monophosphatases. *FEMS Microbiol. Lett.* **173**, 319–325 (1999).

375. R. Z. Sayyed, D. D. Jamadar, P. R. Patel, Production of Exo-polysaccharide by Rhizobium sp. *Indian J. Microbiol.* **51**, 294–300 (2011).

376. P. Lipa, J.-M. Vinardell, J. Kopcińska, A. Zdybicka-Barabas, M. Janczarek, Mutation in the pssZ Gene Negatively Impacts Exopolysaccharide Synthesis, Surface Properties, and Symbiosis of Rhizobium leguminosarum bv. trifolii with Clover. *Genes (Basel).* **9**, 369 (2018).

377. N. A. Fujishige, N. N. Kapadia, P. L. De Hoff, A. M. Hirsch, Investigations of Rhizobium biofilm formation. *FEMS Microbiol. Ecol.* **56**, 195–206 (2006).

378. J. Lawrence, Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. *Curr. Opin. Genet. Dev.* **9**, 642–648 (1999).

379. A. Mateus, M. Shah, J. Hevler, N. Kurzawa, J. Bobonis, A. Typas, M. M. Savitski, Transcriptional and Post-Transcriptional Polar Effects in Bacterial Gene Deletion Libraries. *mSystems* **6**, e0081321 (2021).

380. D. A. Mohl, J. W. Gober, Cell cycle–dependent polar localization of chromosome partitioning proteins in Caulobacter crescentus. *Cell* **88**, 675–684 (1997).

381. A. Jung, A. Raßbach, R. L. Pulpetta, M. C. F. van Teeseling, K. Heinrich, P. Sobetzko, J. Serrania, A. Becker, M. Thanbichler, Two-step chromosome segregation in the stalked budding bacterium Hyphomonas neptunium. *Nat. Commun.* **10**, 3290 (2019).

382. A. A. Iniesta, ParABS system in chromosome partitioning in the bacterium Myxococcus xanthus. *PLoS One* **9**, e86897 (2014).

383. D. A. Mohl, J. Easter, J. W. Gober, The chromosome partitioning protein, ParB, is required for cytokinesis in Caulobacter crescentus. *Mol. Microbiol.* **42**, 741–755 (2001).

384. P. Jecz, A. A. Bartosik, K. Glabski, G. Jagura-Burdzy, A single parS sequence from the cluster of four sites closest to oriC is necessary and sufficient for proper chromosome segregation in Pseudomonas aeruginosa. *PLoS One* **10**, e0120867 (2015).

385. P. S. Lee, A. D. Grossman, The chromosome partitioning proteins Soj (ParA) and Spo0J (ParB) contribute to accurate chromosome partitioning, separation of replicated sister origins, and regulation of replication initiation in Bacillus subtilis. *Mol. Microbiol.* **60**, 853–869 (2006).

386. A. Minnen, L. Attaiech, M. Thon, S. Gruber, J. Veening, SMC is recruited to oriC by ParB and promotes chromosome segregation in Streptococcus pneumoniae. *Mol. Microbiol.* **81**, 676–688 (2011).

387. C. Donovan, A. Schwaiger, R. Krämer, M. Bramkamp, Subcellular localization and characterization of the ParAB system from Corynebacterium glutamicum. *J. Bacteriol.* **192**, 3441–3451 (2010).

388. M. Donczew, P. Mackiewicz, A. Wróbel, K. Flärdh, J. Zakrzewska-Czerwińska, D. Jakimowicz, ParA and ParB coordinate chromosome segregation with cell elongation and division during Streptomyces sporulation. *Open Biol.* **6**, 150263 (2016).

389. R. Kadoya, J. H. Baek, A. Sarker, D. K. Chattoraj, Participation of chromosome segregation protein ParAI of Vibrio cholerae in chromosome replication. *J. Bacteriol.* **193**, 1504–1514 (2011).

390. K. Lasocki, A. A. Bartosik, J. Mierzejewska, C. M. Thomas, G. Jagura-Burdzy, Deletion of the parA (soj) homologue in Pseudomonas aeruginosa causes ParB instability and affects growth rate, chromosome segregation, and motility. *J. Bacteriol.* **189**, 5762–5772 (2007).

391. P. McNicholas, I. Chopra, D. M. Rothstein, Genetic analysis of the tetA(C) gene on plasmid pBR322. *J. Bacteriol.* **174**, 7926–7933 (1992).

392. V. Khetrapal, K. Mehershahi, S. Rafee, S. Chen, C. L. Lim, S. L. Chen, A set of powerful negative selection systems for unmodified Enterobacteriaceae. *Nucleic Acids Res.* **43**, e83–e83 (2015).

393. T. Podolsky, S.-T. Fong, B. T. O. Lee, Direct Selection of Tetracycline-SensitiveEscherichia coliCells Using Nickel Salts. *Plasmid* **36**, 112–115 (1996).

394. X.-T. Li, L. C. Thomason, J. A. Sawitzke, N. Costantino, D. L. Court, Positive and negative selection using the tetA-sacB cassette: recombineering and P1 transduction in Escherichia coli. *Nucleic Acids Res.* **41**, e204 (2013).

395. T. A. Stavropoulos, C. A. Strathdee, Expression of the tetA(C) tetracycline efflux pump in Escherichia coli confers osmotic sensitivity. *FEMS Microbiol. Lett.* **190**, 147–150 (2000).

396. T. A. Stavropoulos, C. A. Strathdee, Synergy between tetA and rpsL Provides High-Stringency Positive and Negative Selection in Bacterial Artificial Chromosome Vectors. *Genomics* **72**, 99–104 (2001).

397. I. C. Blomfield, V. Vaughn, R. F. Rest, B. I. Eisenstein, Allelic exchange in Escherichia coli using the Bacillus subtilis sacB gene and a temperature-sensitive pSC101 replicon. *Mol. Microbiol.* **5**, 1447–1457 (1991).

398. E. M. Vanderlinde, S. A. Magnus, D. D. Tambalo, S. F. Koval, C. K. Yost, Mutation of a broadly conserved operon (RL3499-RL3502) from Rhizobium leguminosarum biovar viciae causes defects in cell morphology and envelope integrity. *J. Bacteriol.* **193**, 2684–2694 (2011).

399. J. Quandt, M. F. Hynes, Versatile suicide vectors which allow direct selection for gene

replacement in Gram-negative bacteria. *Gene* **127**, 15–21 (1993).

400. A. Pokhrel, L. Li, F. L. Short, I. T. Paulsen, A suite of modular, all-synthetic suicide vectors for allelic exchange mutagenesis in multidrug resistant Acinetobacter strains. *BMC Microbiol.* **23**, 137 (2023).

401. J.-M. Reyrat, V. Pelicic, B. Gicquel, R. Rappuoli, Counterselectable markers: untapped tools for bacterial genetics and pathogenesis. *Infect. Immun.* **66**, 4011–4017 (1998).

402. P. Kast, pKSS—a second-generation general purpose cloning vector for efficient positive selection of recombinant clones. *Gene* **138**, 109–114 (1994).

403. K. Miyazaki, Molecular engineering of a PheS counterselection marker for improved operating efficiency in Escherichia coli. *Biotechniques* **58**, 86–88 (2015).

404. K. A. Stacey, E. Simson, Improved method for the isolation of thymine-requiring mutants of Escherichia coli. *J. Bacteriol.* **90**, 554–555 (1965).

405. Q. N. Y. Wong, V. C. W. Ng, M. C. M. Lin, H. Kung, D. Chan, J.-D. Huang, Efficient and seamless DNA recombineering using a thymidylate synthase A selection system in Escherichia coli. *Nucleic Acids Res.* **33**, e59–e59 (2005).

406. C. K. Murphy, E. J. Stewart, J. Beckwith, A double counter-selection system for the study of null alleles of essential genes in Escherichia coli. *Gene* **155**, 1–7 (1995).

407. F. Arigoni, F. Talabot, M. Peitsch, M. D. Edgerton, E. Meldrum, E. Allet, R. Fish, T. Jamotte, M.-L. Curchod, H. Loferer, A genome-based approach for the identification of essential bacterial genes. *Nat. Biotechnol.* **16**, 851–856 (1998).

408. R. D. Arroyo-Olarte, R. Bravo Rodriguez, E. Morales-Ríos, Genome editing in bacteria: CRISPR-Cas and beyond. *Microorganisms* **9**, 844 (2021).

409. J. P. P. Muyrers, Y. Zhang, A. F. Stewart, Techniques: recombinogenic engineering–new options for cloning and manipulating DNA. *Trends Biochem. Sci.* **26**, 325–331 (2001).

410. R. Li, A. Li, Y. Zhang, J. Fu, The emerging role of recombineering in microbiology. *Eng. Microbiol.* **3**, 100097 (2023).

411. B. J. Caldwell, C. E. Bell, Structure and mechanism of the Red recombination system of bacteriophage λ. *Prog. Biophys. Mol. Biol.* **147**, 33–46 (2019).

412. S. Hu, J. Fu, F. Huang, X. Ding, A. F. Stewart, L. Xia, Y. Zhang, Genome engineering of Agrobacterium tumefaciens using the lambda Red recombination system. *Appl. Microbiol. Biotechnol.* **98**, 2165–2172 (2014).

413. P. Somasegaran, H. J. Hoben, "Quantifying the Growth of Rhizobia" in *Handbook for Rhizobia*, P. Somasegaran, H. J. Hoben, Eds. (Springer New York, New York, NY, 1994; https://doi.org/10.1007/978-1-4613-8375-8_5), pp. 47–57.

414. Z. Bian, S. Li, R. Yang, J. Yin, Y. Zhang, Q. Tu, J. Fu, R. Li, Development of a new recombineering system for Agrobacterium species. *Appl. Environ. Microbiol.* **88**, e02499-21 (2022).

415. M. Inui, K. Nakata, J. H. Roh, A. A. Vertès, H. Yukawa, Isolation and Molecular Characterization of pMG160, a Mobilizable Cryptic Plasmid from Rhodobacter blasticus. *Appl. Environ. Microbiol.* **69**, 725–733 (2003).

416. P. O. Olins, C. S. Devine, S. H. Rangwala, K. S. Kavka, The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in

Escherichia coli. *Gene* **73**, 227–235 (1988).

417. A. San Millan, R. C. MacLean, Fitness costs of plasmids: a limit to plasmid transmission. *Microbiol. Spectr.* **5**, 10–1128 (2017).

418. J. Monod, The growth of bacterial cultures. *Annu. Rev. Microbiol.* **3**, 371–394 (1949).

419. R. L. Bertrand, Lag phase is a dynamic, organized, adaptive, and evolvable period that prepares bacteria for cell division. *J. Bacteriol.* **201**, 10–1128 (2019).

420. M. D. Rolfe, C. J. Rice, S. Lucchini, C. Pin, A. Thompson, A. D. S. Cameron, M. Alston, M. F. Stringer, R. P. Betts, J. Baranyi, M. W. Peck, J. C. D. Hinton, Lag Phase Is a Distinct Growth Phase That Prepares Bacteria for Exponential Growth and Involves Transient Metal Accumulation. *J. Bacteriol.* **194**, 686–701 (2012).

421. P. Tiwari, P. Kaila, P. Guptasarma, Understanding anomalous mobility of proteins on SDS-PAGE with special reference to the highly acidic extracellular domains of human E- and N-cadherins. *Electrophoresis* **40**, 1273–1281 (2019).

422. A. Shirai, A. Matsuyama, Y. Yashiroda, A. Hashimoto, Y. Kawamura, R. Arai, Y. Komatsu, S. Horinouchi, M. Yoshida, Global analysis of gel mobility of proteins and its use in target identification. *J. Biol. Chem.* **283**, 10745–10752 (2008).

423. D. C. Jordan, Transfer of Rhizobium japonicum Buchanan 1980 to Bradyrhizobium gen . nov ., a Genus of Slow-Growing, root nodule bacteria from leguminous plants. *Int. J. Syst. Bacteriol.* **32**, 136–139 (1982).

424. G. H. Elkan, Taxonomy of the rhizobia. *Can. J. Microbiol.* **38**, 446–450 (1992).

425. H. Shi, Y. Hu, P. D. Odermatt, C. G. Gonzalez, L. Zhang, J. E. Elias, F. Chang, K. C. Huang, Precise regulation of the relative rates of surface area and volume synthesis in bacterial cells growing in dynamic environments. *Nat. Commun.* **12**, 1975 (2021).

426. M. Laranjo, A. Alexandre, S. Oliveira, Legume growth-promoting rhizobia: An overview on the Mesorhizobium genus. *Microbiol. Res.* **169** (2013).

427. F. Alberge, L. Espinosa, F. Seduk, L. Sylvi, R. Toci, A. Walburger, A. Magalon, Dynamic subcellular localization of a respiratory complex controls bacterial respiration. *Elife* **4** (2015).

428. A. Badrinarayanan, T. B. K. Le, M. T. Laub, Bacterial Chromosome Organization and Segregation. *Annu. Rev. Cell Dev. Biol.* **31**, 171–199 (2015).

429. L. T. Pulianmackal, J. M. I. Limcaoco, K. Ravi, S. Yang, J. Zhang, M. K. Tran, M. Ghalmi, M. J. O'Meara, A. G. Vecchiarelli, Multiple ParA/MinD ATPases coordinate the positioning of disparate cargos in a bacterial cell. *Nat. Commun.* **14**, 3255 (2023).

430. H. Moriya, Quantitative nature of overexpression experiments. *Mol. Biol. Cell* **26**, 3932–3939 (2015).

431. M. Mühlmann, E. Forsten, S. Noack, J. Büchs, Optimizing recombinant protein expression via automated induction profiling in microtiter plates at different temperatures. *Microb. Cell Fact.* **16**, 220 (2017).

432. J. A. Myers, B. S. Curtis, W. R. Curtis, Improving accuracy of cell and chromophore concentration measurements using optical density. *BMC Biophys.* **6** (2013).

433. H.-L. Chien, W.-Z. Huang, M.-Y. Tsai, C.-H. Cheng, C.-T. Liu, Overexpression of the chromosome partitioning gene parA in Azorhizobium caulinodans ORS571 alters the

bacteroid morphotype in Sesbania rostrata stem nodules. *Front. Microbiol.* **10**, 2422 (2019).

434. K. B. Gilbert, E. M. Vanderlinde, C. K. Yost, Mutagenesis of the carboxy terminal protease CtpA decreases desiccation tolerance in Rhizobium leguminosarum. *FEMS Microbiol. Lett.* **272**, 65–74 (2007).

435. B. P. Cormack, R. H. Valdivia, S. Falkow, FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996).

436. R. A. Cinelli, A. Ferrari, V. Pellegrini, M. Tyagi, M. Giacca, F. Beltram, The enhanced green fluorescent protein as a tool for the analysis of protein dynamics and localization: local fluorescence study at the single-molecule level. *Photochem. Photobiol.* **71**, 771–776 (2000).

437. C. W. Shebelut, J. M. Guberman, S. Van Teeffelen, A. A. Yakhnina, Z. Gitai, Caulobacter chromosome segregation is an ordered multistep process. *Proc. Natl. Acad. Sci.* **107**, 14194–14198 (2010).

438. K. C. Murphy, Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. *J. Bacteriol.* **180**, 2063–2071 (1998).

439. S. K. Kulkarni, F. W. Stahl, Interaction between the sbcC gene of Escherichia coli and the gam gene of phage lambda. *Genetics* **123**, 249–253 (1989).

440. T. Schlake, J. Bode, Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. *Biochemistry* **33**, 12746–12751 (1994).

441. M. Carrió, A. Villaverde, Protein aggregation in recombinant bacteria: biological role of inclusion bodies. *Biotechnol. Lett.*, 1385–1395 (2003).

442. V. Yeong, E. G. Werth, L. M. Brown, A. C. Obermeyer, Formation of biomolecular condensates in bacteria by tuning protein electrostatics. *ACS Cent. Sci.* **6**, 2301–2310 (2020).

443. I. V Hinkson, J. E. Elias, The dynamic state of protein turnover: It's about time. *Trends Cell Biol.* **21**, 293–303 (2011).

444. S. Dong, H. Chen, Q. Zhou, N. Liao, Protein degradation control and regulation of bacterial survival and pathogenicity: the role of protein degradation systems in bacteria. *Mol. Biol. Rep.*, 1–11 (2021).

445. F. Pini, N. J. De Nisco, L. Ferri, J. Penterman, A. Fioravanti, M. Brilli, A. Mengoni, M. Bazzicalupo, P. H. Viollier, G. C. Walker, Cell cycle control by the master regulator CtrA in Sinorhizobium meliloti. *PLoS Genet.* **11**, e1005232 (2015).

446. F. Pini, B. Frage, L. Ferri, N. J. De Nisco, S. S. Mohapatra, L. Taddei, A. Fioravanti, F. Dewitte, M. Galardini, M. Brilli, The DivJ, CbrA and PleC system controls DivK phosphorylation and symbiosis in Sinorhizobium meliloti. *Mol. Microbiol.* **90**, 54–71 (2013).

447. J. N. Latch, W. Margolin, Generation of buds, swellings, and branches instead of filaments after blocking the cell cycle of Rhizobium meliloti. *J. Bacteriol.* **179**, 2373–2381 (1997).

448. J. C. Chen, P. H. Viollier, L. Shapiro, A membrane metalloprotease participates in the sequential degradation of a Caulobacter polarity determinant. *Mol. Microbiol.* **55**,

1085–1103 (2005).

449. P. Mier, G. Alanis-Lobato, M. A. Andrade-Navarro, Protein-protein interactions can be predicted using coiled coil co-evolution patterns. *J. Theor. Biol.* **412**, 198–203 (2017).

450. R. Reyes-Lamothe, E. Nicolas, D. J. Sherratt, Chromosome replication and segregation in bacteria. *Annu. Rev. Genet.* **46**, 121–143 (2012).

451. J. Mierzejewska, G. Jagura-Burdzy, Prokaryotic ParA-ParB-parS system links bacterial chromosome segregation with the cell cycle. *Plasmid* **67**, 1–14 (2012).

452. S. Okuda, H. Tokuda, Lipoprotein sorting in bacteria. *Annu. Rev. Microbiol.* **65**, 239–259 (2011).

453. S. Narita, H. Tokuda, Bacterial lipoproteins; biogenesis, sorting and quality control. *Biochim. Biophys. Acta (BBA)-Molecular Cell Biol. Lipids* **1862**, 1414–1423 (2017).

Appendix

A.1 Synthetic gene knockout constructs

*par*A1-KO - 1627 bp;

```
CCCGGGGCTGAAGGTCAAGCTCGGCACTGCTCGCCCTTTCAATATTGCCCAGGCTGCAAT      60
CGTTGAAGGGATGACACCTGCGGCAATCGCATTGCTGCTCGTGCATTTGCGTCGACTATC     120
CTCTGCCGAGCGGAATAGCGCTTGATACAAGATTTGAGAGTCTTTCAGCATGGAATTAAA     180
CGGTTTGCGTGTTTCACGTGAAACACAGGAACGGTTGCAACACTTTGCAGCGCTGTTTCA     240
GAAGTGGGCAAAAGCAATCAATCTGGTGGCGCCGTCGACACTCGACGATCTATGGCACCG     300
GCACATCGCTGATAGCAGCCAGGTCTTTCAGATTCATCCTCAACCGATTACTTGGGTTGA     360
CCTCGGCAGTGGGGGCGGCTTTCCCGGCGTTATTACGGCAATTTTTCTGGCCGAGCTGCA     420
AGGTGGATGGGTCCATCTGGTCGAAAGCAACCATAAAAAGGCAGCGTTTCTACGGACAGC     480
GCTGCGGGAGACAAATGCACGCGGCAGCGTCCATAGCATACGAATCGAAGATGCTTATAC     540
GGAGGTTCGCGAATGTAGTGCGATTTCAGCGCGCGCCCTTACGGATCTTGATGGGCTCAT     600
CGGATATTCATCGCCCTGGATGCTTGGCAAGGAGAATTGTCGCGGCTTCTTTCACAAAGG     660
CCGGGATTACCTGAGGGAAATTGACAAAGCACGTGGTCGGTGGGAATTCGATCTGTTAGA     720
ACATAAGAGCGCCGTCGAGCAGGAATCTGTTATTCTGGAAATATCAAATCTCCGGCGCTT     780
GGTTTAACAGATCGGATATTGCGGCAATGGTTAATTAATGATATCATTTTAGGGGTGTAG     840
GTGAGTGTAGATCATGAATGACGATGTATCAAAGAGGCGATTGGGTCGCGGCCTTGCGGC     900
GCTGATTGGTGAAATGGATCAGCCCGTCCCGGTGGAAGCCGAGAGGACTATCAGCGCCGA     960
CAGAATGATTCCGATCGAGTTCGTCAGCCGTAACCCCCGCAATCCCCGTCGGTTCTTCGA    1020
TGACTCGGAGCTGCACGATCTCGCAAGCTCCATTCGTCAACACGGTATCGTTCAGCCGAT    1080
TGTTGTGCGTACGATGTCGCGGGATCGATATGAGATCATCGCCGGCGAACGACGTTGGCG    1140
CGCAGCTCAGCTTGCGGGGCTGATCGAGATTCCCGTCATTATCCGTGATGTGGATGATAA    1200
GACGGCGCTGGAGATTGCGATCGTCGAAAACGTGCAGCGCGCGGATCTCAATGCCCTCGA    1260
AGAGGCCTTGGGTTATGAGCAGCTTATCGCGGAATACGGCTACACCCAGAATGACCTCGG    1320
CGAAATCATCGGGAAGAGCCGCAGTCATGTGGCCAACTCGTTGCGCCTGTTGAAATTGCC    1380
TGATCCGGTGCGCGATCTGCTTGCTGCCGGCAGCCTATCGGCTGGGCACGCGCGCGCTTT    1440
GGTTTCAACATCCGACCCGGCAGGTCTTGCCCGTACGATCGTTGCCAAGGGCATGTCGGT    1500
GCGTGATGCGGAAAAGCTGGCGCAAAACAATATCAAGGCCCAATCGGAGCCGCAGCAGGC    1560
GGCCTTGCGGCGAGATCAGAAAGATTCCGACACGCTGGCGCTGGAACGGACACTTTCCGA    1620
TAAGCTT                                                         1627
```

XmaI recognition site, Upstream gene-flanking region, Start codon, PacI recognition site, Stop codon, Downstream gene-flanking region, HindIII recognition site

popZ-KO - 1627 bp;

```
CCCGGGGATGGAAGCATCGCCGCTTGCCAAGCTGCTGCAGCTGGCAGCGATCACCAAAGA      60
CGACTTCGTGCTCGAGGTTGGTTGCGGTACGGGTTACACTTCGGCGCTGCTGTCGATCAT     120
TGCCGGCTCCGTCATCGCGCTCGAATGCGACGAGACGCTGGCCACTGAAGCGAAGGCGCA     180
GCTTGCCGGCTACGCCAAGGTCGAAGTCGTGACCGGCCCGCTCGAAAAGGGCTACGCTGC     240
CGGCGCTCCCTATGATCTGATCTTCATCAACGGCGCGGTCGAAGAGGTTCCGGCCGCCCT     300
TCTCGGTCAATTGCGCGATGGTGGCCGTCTGATCACCGTCGAGGGTTACGGCAATGCCGC     360
CCGCGCCAAGGTCTTCGTCGCCGAGCGCGGCGCCATTTCGGAAAACGTCTTCTTCAACGC     420
CTCGGTCAAGCCGCTGCCGGGTTTCGCCAAGGCGCGCGAATTCGTTTTCTGACATCGTTC     480
TTATCGATGACATTCAGCAACGGGCGCCATTGGCGCCCGTTTTTTGTCAACGCCGCTCGA     540
AAGCCTTCACAAACATCATCTAGGCTGACGCCGCCACGGCGGATCGGACGGATGTTATGC     600
GCTTCAACTTCGGTTTCATGGGATGTTTCCGCCGCAAAATCGCGACAGGCTTTGCGGGTG     660
GCAAACATCAGGTAACAAAACTGTGCATGGCCACATTCATTTGATTCGCGATGATTTTTT     720
TCACGTCTGGGGTATTCTAAGGTCGTGGTGATTCGGATGAATGCTCCACACCATCCGGTC     780
GTTGTTTTGGGATAACGGGGATAGAAATGGTTAATTAATGATCGGATTGCTTTTTCCTGA     840
AAAGCCGCTCCGGTCTCCGGGGCGGCTTTTTTATTGACTTGGCCGTGACCATCCTATTTA     900
CAAACGCCACCCTCGAACCTCGAAATTGGTCAAAAAATGCTCGACAAGACCTATGATTCC     960
GCTGCCGTCGAACCGAAAATCGCCGCGATATGGGATGAGGCGGACGCTTTCCGCGCCGGC    1020
GCGAACGCCAGGCCCGGAGCCGAGACCTTCACCATCGTGATCCCGCCGCCGAATGTCACG    1080
GGTTCGCTGCACATGGGCCACGCGCTGAACAATACGCTGCAGGACATCCTGGTCCGCTTC    1140
GAGCGCATGCGCGGCAAGGACGTGCTCTGGCAGCCGGGCATGGACCACGCCGGCATAGCT    1200
ACGCAGATGGTCGTCGAGCGCAAGCTGATGGAACAGCAGCTGCCGGGCCGCCGCGACATG    1260
GGCCGCGAAGCCTTTATCGACAAGGTCTGGGAATGGAAGGCTGAATCAGGCGGCCTGATC    1320
TTCAACCAGCTGAAGCGCCTGGGGGCGTCATGCGACTGGTCGCGTGAACGTTTCACCATG    1380
GACGAAGGCCTCTCAAAGGCCGTGCTCGAGGTTTTCGTCACGCTCTACAAGGAAGGGCTG    1440
ATCTACAAGGACAAGCGTCTGGTCAATTGGGATCCGAAGTTGCTGACGGCGATTTCCGAT    1500
ATCGAAGTCGAGCAGCACGAGGTCAAGGGCAATCTCTGGCACCTGCGCTATCCGCTGGAA    1560
AAGGGCGTAACCTACCAATACCCGATTGCATTCGATGAGGAAGGCAAGCCGACCGAATTC    1620
GAAGCTT                                                         1627
```

XmaI recognition site, Upstream gene-flanking region, Start codon, PacI recognition site, Stop codon, Downstream gene-flanking region, HindIII recognition site

Rcy-KO - 1627 bp;

```
GAATTCGCGGCACCGAGCGCTGCCGTCGCGCCGGCATCGCCACCGTTTTCCTCGACGGCT      60
TCCGCTGTCTGAGACACGCGGAACGCGCCGACCGAGCTTGCCGCGCCCTTCAGCCGGTGC     120
GCGGCTGCGCCGATCCGGATGGTTTCGCCGCTGGCAATATCCTGAAGGCATGCGCGGGCT     180
TGGCGCGCAAACATCTGAAGGACTTCGAGTTCCAGCGTCTTGTCGCCCATCGTCTGCTTT     240
GCGAGATGGACCAGATCGATCGGCCGGACCTTCGAGGGACACGGTCCCTTTGCATTATCA     300
GGCGCCTCGAATACGATGTTCAATGCTGCCATCTGGGCTGCCATTGCCAATTCTCCCGTC     360
TCTGTCGTCCGCAGTTGCAACAGTTCTGCGACAGGAGGGTGGCCATCACGTTAAATTCCG     420
GCACAGTCAGGGCGGAAATCTCGCCATATGGTTAACATCCGTTAAATATCCCTAAACTAT     480
TGGCTTTTAAGCCGCGCCTAGGATTCAAAGGTGTTAACAAAGGGTTAACGCGCCATGAAT     540
CTCAAGCCTTCATTAATTGTGTGAAGAGCCCAGATGTGTCACTACGATACTGGTGGAGCG     600
GGTTTATGGTACGCGCCTTTGCCGAGCGAGTGGATAATGGTAGTGTTTTGATACCTTCCG     660
TTTGAAAAAGCGGCTATCTACTCTGAAATGACATGCTTATCCGTCCTTCGTTGGGATGGA     720
AGGCTGGAACGGCAAAGTTGTTCCCGGGTGAGGCGGAAGCCCAGGAAGCGCTGTCGGACT     780
GGCTGACAGTAACGAGGCATAACCGAATGGTTAATTAATGAGCCTGACTGGTACCCGCCG     840
ATTGGAAAGCCGCTTGACCACGTGTCGGGCGGCTTTCTCTCTATGAAGCGAGTGTTATCC     900
ATTCGCCATCCAGCCGCGGAAGCCACTCCCGAATGACCAAGGCCCTTCTGATCATCGACG     960
TACAGAATGCGATCCTCCGCGGAAAAGCCTCGCCAGAGCGGCAGCCTCAAGTCGATGCCG    1020
CACTCGACGAGACCGTCGCACGGCTGGCCGCGCTACAGGAAAAAGCTCGGCAGGCGGGCG    1080
CGCCGATCGTGCTCGTTCAGCATGACGGCGACAGCGGCCATCGATTGGCTGTCGGCACGC    1140
CGGGTTGGGCGCTGCGCGACGAGCTCGCGCCGAGGCAGGCCGAGGTCGTTGTGCGCAAGA    1200
AAAGCGCCGACGCCTTCTTCGAGACCGATCTTGCCGAACGCCTGGACGAACGATCGGTCA    1260
CCCACCTCGTCATCGGTGGATGCATGTCGCAATTCTGTGTCGACACGACAGTCAGGCGAG    1320
CCGTGTCGCTCGGATATGACGTGACGTTGATTGCCGATGGTCACACGACCGGCGACACGG    1380
CGACCTTGACCTTCTCCGAGATCATTGCCCATCACAACGAAACACTCAATGGTTTCGATG    1440
CGGGCAAGGCGACGGTAGAAATCCGCCCCGCCGCCGAGATCGACTTTAGATAATTTTTCG    1500
AGCTGCAGGTGACCGGAATTGCCAGGCGGCCCCGGTCTCCTATTTCTTACCGATCGTCCG    1560
CTATGTTCCAGTTGTGGACGCAACAATCGAGGCAGGGTGAGACAAAAGGAATCAGCTGTG    1620
GGGATCC                                                         1627
```

EcoRI recognition site, Upstream gene-flanking region, Start codon, PacI recognition site, Stop codon, Downstream gene-flanking region, BamHI recognition site

A.2 Complications with the synthetic *parA1*-KO construct

Multiple complications with the synthetic *parA1*-KO allelic exchange construct (Section 2.3.7) were encountered during attempts to clone the pK*18-parA1*-KO vector (Table 2.3). Firstly, the transformation of the *parA1*-KO delivery plasmid into *E. coli* DH5α was unsuccessful. Furthermore, transformation of this plasmid into Thermo Fisher Scientific DH10B Competent Cells yielded only two colonies across three independent transformations. Restriction digestion of plasmid isolated from each of these colonies highlighted several issues with this plasmid (Figure A.1, Parts A & B). Undigested plasmids were much larger than expected, producing a band of between 8 and 9 kb. Linearisation of the plasmid with PacI produced a band of approximately 4.5 kb. Not only was the linearised plasmid also larger than expected, but generation of a 'linear' product significantly smaller than the undigested plasmid suggested potential concatermerisation of the plasmid. XmaI and HindIII digestion yielded a product of approximately 1.6 kb, indicative of the correct insert (i.e. the synthetic *parA1*-KO construct). However, instead of the expected 2.2 kb vector backbone fragment, a smaller 1 kb band was observed. Liberation of the 5' (XmaI + PacI digest) and 3' (PacI + HindIII digest) flanks produced the expected 800 bp fragments, supporting the conclusion that the synthetic *parA1*-KO construct was present, but that the vector backbone was incorrect.

The 1.6 kb XmaI/HindIII fragment (presumed to be the *parA1*-KO construct) was isolated from each of the two DH10B-derived plasmid preparations, and subcloned into pK18*mobsacB* (Section 2.6.1). Plasmid was isolated from three transformants per subcloning attempt and subjected to a validation digest alongside a pK18*mobsac*B-derived control vector. A schematic of this vector has been shown in Figure A.1C. Analysis of the digestion (Figure A.1D) revealed that despite successful digestion of the control vector, all six potential pK18-*parA1*-KO clones were not linearised by PacI. For all potential pK18-*parA1*-KO clones, XmaI, HindIII, PacI triple-digest generated a band representing the pK18*mobsacB* backbone (also seen in control digest). Incomplete digestion was also observed, indicated by a band of approximately 7000 bp. The third product of this digest was a 1600 bp fragment, likely corresponding to the insert cloned into pK18*mobsacB*. Failure of this insert to be digested by PacI into two 800 bp fragments confirmed that the 1.6 kb inserts cloned into pK18*mobsacB* were not the correct *parA1*-KO construct. Further screening of transformants did not identify of any correct pK18-*parA1*-KO clones, with subsequent subcloning attempts also unsuccessful.

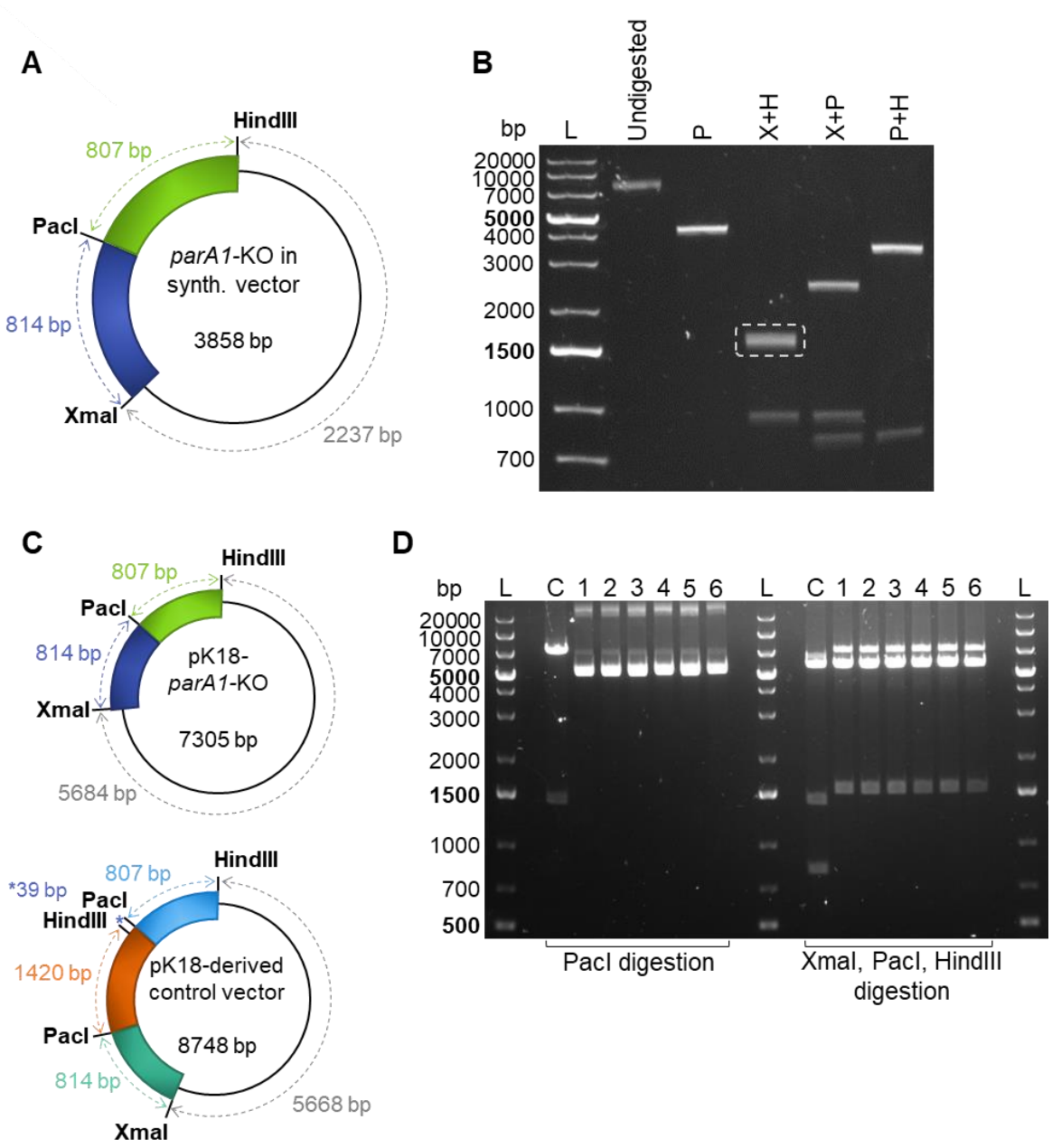


Figure A.1. Diagnostic digests of the synthesised *parA1-KO* construct. (A) Schematic representation of the expected *parA1*-KO synthesis vector received following artificial gene synthesis, highlighting the XmaI, PacI and HindIII restriction sites used for diagnostic digest, and the expected digest products. (B) Agarose gel electrophoresis (1% w/v) of the restriction digestions of a single *parA1*-KO synthesis vector. Lane L; GeneRuler 1kb Plus DNA ladder. Letters above sample lanes indicate endonucleases used for digestion (P: PacI, X: XmaI, H: HindIII). An undigested plasmid sample was also analysed. (C) Schematic representation of the pK18-*parA1*-KO plasmid and a pK18*mobsacB*-derived control vector, highlighting the XmaI, PacI and HindIII restriction sites and the expected products following digestion with these enzymes. (D) Agarose gel electrophoresis (1% w/v) of restriction digestions. Lane L: GeneRuler 1kb Plus DNA ladder. Lane C: pK18-derived control vector. Lanes 1-6: potential pK18-*parA1*-KO clones, derived from two independent subcloning events.

A.3 His-tagged Protein expression constructs

His-tagged ParA1 expression construct - 990 bp;

```
TTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCC        60
ATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCTGGCGAAC       120
GACATCGGATTATCACCATCGCAAATCAGAAGGGTGGCGTTGGCAAGACGACCACAGCAA       180
TTAACCTGGCAACGGCGCTTGCTGCTATCGGTGAGCGGGTCCTAATCGTCGATCTCGACC       240
CGCAGGGCAATGCCAGCACAGGCCTCGGTATCGATCGCCGTGATCGCAAGCTGTCCTCCT       300
ACGATCTGATGGTCGGTGATCGTGGGATTCCCGAAGTGACTCTCGAAACAGCGGTTCCTA       360
ATCTGTTTATCGTCCCATCAACGATGGATCTTCTTGGTATCGAGATGGAAATCTCGCAGC       420
AGAGCGATCGGGTGTTCAAATTACGGAAAGCGCTATCTTCGCCCGAAGCCATGGCGTTTT       480
CTTATATTCTTCTGGATTGCCCGCCTTCGTTCAATCTGCTGACGATGAATGCCATGGCGG       540
CAGCGCATTCGGTCTTGGTGCCGCTGCAATGTGAATTCTTTGCGCTGGAGGGGTTGAGCC       600
AGCTGCTCGAGACGGTCAGCCAGGTTCGGCGGACAGTTAACCCGCGTCTGGATATTCAGG       660
GTATCGTCTTGACGATGTTTGATGCGCGAAACAATCTTGCCCAGCAGGTTGTCAATGACG       720
TGCGTACGCATCTCGGGGACAAGGTTTACCACACGTTAATTCCGCGCAACGTCCGGGTTT       780
CCGAGGCTCCGTCCTACGGTAAGCCGGCTATTCTCTATGACTTGAAGTGCGCGGGCAGTC       840
AGGCCTATCTGCAGCTGGCCTCCGAGGTCATCCAGCGAGAACGTCAAAGGCTGGCCGCGT       900
AATATCATTTTAGGGGTGTAGGTGAGTGTAGATCATGAATGACGATGTATGCGGATCCGA       960
ATTCGAGCTCCGTCGACAAGCTTGCGGCCG
```

XbaI recognition site, Ribosomal binding site, His-tag, Thrombin cleavage site, *parA1* gene sequence, BamHI recognition site

His-tagged PopZ expression construct – 1075 bp;

```
TTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCC        60
ATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCTCAGCCAA       120
GCGTAGCGCGTGAACCGTCCATGGAAGAAATTCTGGCGTCCATCCGCCAGATCATCGAAA       180
GCAATGAGCCCGGCGCCGGCAAGGCGATTTCTGCATCCCTGCCGCCGGTCTACGGCGCCG       240
ATGAGGATGACAATGGCTCCGAAATTCATCTGACGGTCGACGATACCTATGCCGGCGTGG       300
AATTCCCCGAACCGGTCATGCGCTCCTCCGATCCGCGTTTCGTCGCCGCCAATTCGGCCG       360
GCACCGCCCCCGAAGCAGACGTTCCGGCCCGCGCTCTGTCGCTTGCCGATGTCGCCGCCC       420
GCGTGCGCGCCGCCTCCGAGCGCAGCGCCGTGCAGGCCGGTCAGGCCCTGCGCGAAATTC       480
CGAGCGGCTTCCGTCAGCCCGAGCCGCAGCCGGCCGTAATGCCGGAACCGCCGCGCGCCG       540
CGGCACCGCAACCCCAGCCGCAGCCGCAGCCGGCACAGCCTGTTTTCCCGGATGCCGCAG       600
CGTCCATCCAGCATGTTGCGATCCAGCAGCCGGCTGAGCCGGTGTTCACGGAAATGCAAA       660
GGGTTGCCGTCGCCGAGCCGGCGCCTGCTGTCGAAACAGCGCTGCCGGCTATGGAAGCCA       720
CTCAGTCGTCGACCGAGCGTTTCCTGCCGAGCGTTATGGACGAGGTCCAGCCGACGCTGC       780
```

```
TTTCGGAGGATGCCGGCCTGCAGATCAGCCGCTCCTTCGAGGAGCTCGCAGCCGCGATCG        840
ACGGTGCCGAGCGCCGCTCGCTGGACGAGATCGCAGAGGACATGCTTCGCCCGATGCTGC        900
GCGAATGGCTTGATGATAATCTGCCGACACTGGTCGAGCGGTTGGTGCGCGAGGAAATCG        960
AGCGCGTGGCGCGCGGCCCGCGCCGCTGATCGGATTGCTTTTTCCTGAAAAGCCGCTCCG       1020
GTCTCCGGGGCGGCTGCGGATCCGAATTCGAGCTCCGTCGACAAGCTTGCGGCCG            1075
```

XbaI recognition site, Ribosomal binding site, His-tag, Thrombin cleavage site, *popZ* gene sequence, BamHI recognition site

His-tagged Rcy expression construct – 7308 bp;

```
TTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCC         60
ATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCGAACAACA        120
AATATAACGAGTCAATTGAGGACAAGGCGTTCAAGGCATTGGACGAGGCGCTCCAGATCG        180
ACTTTAGCAAGGATAACGTACCGTCTCGCCGCGGCGACGCGCCCCAGGCCCCGGAGGCTA        240
ATGTGTCTGATCCAGTAAATGCGCGTAACCAGCAGGCCCAGGAAGAAGCGCAGCGCCGCA        300
GCCGCCGCGGTGCCGCCGGCGAGCAGGCTTCCAGGGGTCCGGCCTTCGCGCCCGCCAACG        360
ATGCCAGCCGCAATACGCCCGCCTCGATCCTGAAATCCTTTGACGGCGCCTCCAGCCGCT        420
CGGCGGTGCGCACCGCCACCCTGTTTTCCGTGCTTTGGGTTATGGCCGGCCTTGGCCTGA        480
TGTCGCTGCTTTATGCTCCGCAGATCTGGCAGATCCGTTCGCTCGCCGATCTCGTCGCCC        540
TGCCCGGCGTGATCGCCGGCCTCGTCGGCATCATCATTCCCGTGATGATGTTCTATGCCT        600
TCGCCATCATGATCTCCCGCGCCCAGGACATGCGCAACGCCGCCCGCTCCATGGCAGAGG        660
TGGCGTTGCGCCTGGCAGAGCCGGAAACCATCGCCTCCGAGCGCATCATGACCGTCGGCC        720
AGGCCGTGCGCCGTGAGGTCTCGGCCATGAATGAAGGCATCGAGCGCACCATCGCGCGTG        780
CCTCCGAGCTGGAAACGCTCGTCCATTCCGAAGTCAATGCGCTTGAACGTTCTTATGCGG        840
ACAACGAGTTGCGCGTTCGCGGCCTCGTCCACGAACTCGGCTCCGAACGCGAGGCGATCG        900
TCAACCATGCCGACCGTATCCGCACCTCGATCGGCAGCGTGCACGACCAGTTGAAGGAAG        960
AGCTGTCGCTTGCGACCGAAGAGATCGCGGTGCGGCTGGCCACATCGGGCGAAGCCTTCG       1020
CCTCGCTGATCGATACACGCGCCGCAACGATCCTGGAAAAATCGGACAGCGCCTTGCAGT       1080
CGATGGGCAGCCTGCTCGCCGCCAAGACCGATACCCTGCTGCAGACCCTAAACGCATCGG       1140
GTTTTGCGCTGGCCACCGAATTCGACAACCGCCTCGAAGCCCTCTCCGTCAATCTCAACG       1200
ATCATGGCGAAAGGCTGCTCAGCCAGTTCGAGACGCGCGCCTCGACCATGGACAGCAGCA       1260
CCGAAAAGCTGAATGCGGCGCTGAACGAGCGTACGCACCAGCTCAACGAGATCCTGATCG       1320
CCCGCACCCGCGAGATCAATGAGAGCCTGACGTCAGGCGAACGCACCATCGGCGGCACGC       1380
TCGACGACGTGCTATCGAAGCTGAACAACGCGCTCGACGAAAAGGGCGCAAGCTTCCGCC       1440
AGAGCCTGCAGACCACCGCCGACGACGCCGTCATGGATCTCGACGTGCGCTCCGGCTTCT       1500
TCGAGGAGCGGCTGCAGACGACCGTCGCCCAGCTGTCGACCGCCTTCGACGAACGCGTCT       1560
CCGAATTCACCAGCGCCTTCGACAAACGCACCGGCTCGCTCGACACCAAGCTGATGGAGA       1620
GCCTCGCCCGTATCAACGAGACGCTGACCGGCGGATCGGATTCGATCGACGGCATCCTGA       1680
ATTCCGGCATCGACCGCCTTGGTTCGTCGATGACCGACCAGTCGCTGGCGCTCGCCACCG       1740
```

CGCTCGCCACCGGCCACGAAGTGCTGGAAAGCACGCTCGGTAACCGGGCGGACGAAATCA 1800
CCACCGCACTCACCAGCCGCACTGGCGAGCTGACCTCGGCGCTGCAGAGCGCGACCTCGG 1860
AAATCGCGCTGACGCTTGCGACCGGCACCTCCGAGCTGCAGAACAACCTTCAGACACGCT 1920
CGGCCGAATTCCGCGACACGCTGCGCACCACCACCACCGATCTGACGACTGCCGTTGCCG 1980
CTGGCACCGAGCAGGTGACCACGGCCTTCGGCGGTCGCGCCGAGGAACTGAGCGGCGTGC 2040
TTGCCGCCCGTGCGGCCGAAATCAGCGAGGCGCTCGGCACGGCCCATGGCCGTATCGACA 2100
GCGTCATGGCAGAGCGCGGCGGCGCACTGGTCGAGGCGCTGACCACCCATCACGGCCGCT 2160
TCGAGGAAGCGCTGACGACCCGCTCCGACGCCATCATCAATGCCGTCTCCGACACCCATG 2220
ACCGTCTTGCCGAGACGCTCGACGAAAAGTCGATGGCGCTCGCCATCTCTCTGAACGAGA 2280
GCCAGGCCCGCATCGAGGACACGCTCGAGACCCGCTCCGAGGCCTTCCTCAACGCCGTGT 2340
CTGGCACCCATCATCGCCTCTCCGAGACGCTCGACGAGAAGGCGATGGCGCTCGCCATTT 2400
CGCTCAGCGAAAACCAGGCCCGCATCGAGGATACGCTCGAGACCCGCTCCGAGGCCTTCC 2460
TCAACGCCGTGTCTGGCACGCATGATCGCCTGTCGGAAACGCTGGACAGCAAGGCGGCGG 2520
CCCTGACGACCTCGCTCGACGAGCGCCATGCCCGCATCGAGGATGCGCTTGGAACGCGCG 2580
CCGAGGCACTGCTGAACACTGTATCCGGCACGCGCGATCGTCTTGCCGAGACGCTCGACG 2640
AGAAGGCGATGGCGCTCGCCATGTCCTTGAACGAGAGCCATGCCCGTATCGAGGAGACGC 2700
TGGAAACCCGCTCCGCCGCCCTGCTGAACGCTGTCTCCGGCACCCATGATCGTCTCTCGG 2760
AGACTCTGGACGGCAAGGCGGCAGCTTTTGCCACCTCGCTGAGCGAAGGCCAGGCCCGTA 2820
TCGAGAATACGCTCGAAAGCCGCTCCGCAGCGCTGCTGAACGCCGCCTCCGGCATCCATG 2880
ATCGCCTCTCCGAGACGCTGGACGAAAAGGCGATGGCACTCGCCATTTCGCTCAACGAAA 2940
CCCAGACGCGTATCGAGGACACGCTGGAAACCCGTTCGGCGGCCCTGCTGAATGCGGTCT 3000
CCGGCACGCATGATCGCCTCTCCGAGACGCTGGACGAGAAGGCGATGACGCTCGCCATCT 3060
CGCTCAATGAGAGCCAGTCGCGGATCGAAGACACGCTGGAAGCACGCTCCGAAGCCTTCC 3120
TCAAGGCCGTGTCCGGTACGCACGACCGTCTCTCGGAAACGCTCGACGAAAAGGCGACGG 3180
CGATCGCCGCCTCCCTGAACGAGGGCCAGAGCCGCCTGCAGGACACGCTGGAGAGCCGCT 3240
CGGAATCCTTCCTGAATGCGGTCTCTGCCACCCACGACCGCCTGTCCGAGACACTCGACG 3300
ATCGGGCGATGGCGCTTGCCATCTCGCTGAACGAGAGCCAGAGCCGTCTCGAGGAGACGC 3360
TGACAACAGGCGCCGATGCGATCACCAATGCGGTCTCCGGCACGCATGTCGGCCTGAACG 3420
GTGTACTCGACCAGAAGGCCGCCGCCCTCGCCACCTCGCTGAGCGAAGGCCAGGCCCGCC 3480
TCGAGGAAGCCTTGGGCCACCGCACCGCCGCGATCATCGGCGCCGTAACGGCAACCCACG 3540
ATCGGCTCACCGATACGCTCGACGAAAAGACCATGGCGCTCGCCATTTCGCTCGATGATA 3600
ACCAGAGCCGCTTCGACAGTGTGCTCGAAGCCCGCAGCAACGCCATCATGGAGGCCGTCT 3660
CCGGTGCCGAGGCACGAGTCGCCGGCGCCTTCTCCGACAAGACCGATGCGATCCGCACTG 3720
CCTATACCGATAATCAACAGCGGCTCGAAAATGCGCTTTCTGAACATAGCGCCGCCCTCT 3780
CGGGCGTTCTCGATGCTGGCGGCAACCGCTTTGAGGATCTCGTCGGCGGCTTGACCGGCC 3840
GCATCGAAGGCCGTCTGACCGATGCCCATACACGGCTCGGAGGCCTCGCCGACGAAGCCG 3900
CGGCTCGTATCGAGGGCGGGCTCGCCTCTGCCCATGAGCGCATCCGTACGACGCTCGAAG 3960
ACCGCGCCAACGCCATCGACCTATCGCTGAACCAGGCCCATGCGCTGATCAGCGATACGC 4020
TGGCCGAACAGGCGACCAGCATCGGCACCTCGGTGGCGACGAGCGTCGGCATGCTCGAAT 4080
TGTCGCTGGAGCACCGTGAAGCCGCGATCCGCCAGGCGATCGATGCCGGCGCCCAGACAT 4140
TGGAAGATCGCATGCATGCCGGCGCGGGCCAAATCGCCGGCCGCTTCCAGGAGGCGGCAA 4200

ATGCGATTTCGAGCTCCACCCAGGATTTCTCCGCCCACCTCGATCGGTCGGTCGAAAGCC 4260
TGACGGGCCGTTTTGAAGAAACCGGGGCGCGTGTGGAAGCCGGCCTTGCGGCGATCGAGA 4320
CCCGCATCCGCGACGGCGTCGGCGGCGTTGCCGAAAGGGTCGAAGCGGCCAGCGGCCAGC 4380
TGTCGGGCGTGCTTGCCGACGGCGTCTCTCGTATCGGCGCGCTCAGCGACGACGCCACGC 4440
AGCGCATCTCGGCGACGCTCGAAGGCAGTGCCGCAAACCTCACCCAGGCGATCGACAACC 4500
GCACTGCCAATCTGGCCGAGACGCTGGACAGCCGCACGACCAGCCTGACCGACGCTATCG 4560
ACGGCCGCACCGCGAGCCTCGGCGAAACTCTCGACCGCGGCAATGAACGCATCGAGGAAC 4620
GCCTCTCCACCATGGACCGTGCGTTGACCGTCGGCCTCGAGGCCGTCAACCGCACCATCG 4680
AAGGCAAGGCCGCCGGTCTCGCCTCGACGCTGCGCAGCGCGGTTGCCGACGCAGCCCAGG 4740
GAATGGAAGGCGAAGCCACGAGAACCACCGAACTGCTTGGCAAGACCGGCCAGCAATTCG 4800
CCGAGGATCTCAATACGAAAAGCGACGAATTCACCCGCACAATGGATGAGCGCTCGTCGC 4860
AGATCGTAACCCGCGTCGCCGAAGCGCAGAACCGGCTGGCAAGCCAGGCCGCTGCCGTCG 4920
CCCAGACATTCTCGGAAGCCGGCAACGCCATCGTCAACAAGGTCGCCGAAGCCGAGACCA 4980
TCGTCGGCACGCAGGTCAATGCCATCTCCAAGGTGCTGTCGGATGCCGGCCAGTCTCTGG 5040
AGACACGCGGCAACGCCATCCGCTCGACACTGTCGGGCGCCGGCAGCGAGATTTCCGCCA 5100
CCATGGCGGATGTCGACCGGGCGCTCGAAGCACGCAGCAGCGCCATTCGCACCAACCTCG 5160
AAGAGCGTGCCCGCGAGATCGATACGACGTTCTCCGATATCGACCGGGCGCTCGAAGCAC 5220
GCGGCAACTCGATCCGCTCGACGCTCGAGGAGCGCACGCGCGACCTCAACTCGATGCTAT 5280
CAGGCCGTTCGGTCGAATTGACGCGCATCCTCGACGAAACGGCCCGCCCGATCATCGACC 5340
GCTACACCGATGCCGGCGAGCAGGCCGCCGCCCGCATTACCGCCGCGGCCAACCTCAGCG 5400
CCGACCGCCTGCGCGCCGAAAACGAAGCGCTCGCCAGCGCCGTTGCCGCCCGCACCGAGA 5460
ATGTCGCCAATGCCGTCAGCGCCATAGAGAACAGCCTGGTCGGCAACGTCAACGGCCTCG 5520
TCGAGCGCATGTCGGAAAGCAGTGCTGCAATGGCGATGATGATGAACCGCGCCGCCGAGC 5580
AACTGGCCAGCGTCGACGGCCGTCTCGGCGATACCACCACACGCTTTGCCGACTCCGCCA 5640
CCAAGGCCGCTGAAATGGTCTCCGCCTCGACCAGGCTTCTCGAAGGCAAGGTCGACCGCC 5700
TCTCCGACATCTCGGGCCAGACGCTGGCACAGGTCGGCGGCATCATCGGTCGCTTCGACG 5760
AACACTCCAAGGTTCTGACCCAGGCTTCGCAGCTTCTCGGCGCCGCCCAGTCCAACCTCG 5820
CCTCGACGCTCGAAGAGCGCGAGAGCGCTCTGCAGACGCTCTCGGTCGGCCTCGTCCAGC 5880
GCTCCGCCGAGATCGAAAAGACCATGCAAGGCCTCGGCGGCATGATCGAGACCGTGTTCG 5940
AGCGGGCAGAGCAGCGCTCAGCCCAAGTCACCGGCAATCTGCGCCAAGGCGTGCAGTCGT 6000
CCTTCGCCGATATCGGCCGCATTCTCACCGAAACCGAGAAGCGCGCCCAGGAAGCGGCAG 6060
AGACGATGCGCGAGGCGATCACCCGCGCCGGCGACGAGGCCAATACGACGATCGACGGCA 6120
CCTTCGCCAATGTCGAGCGCCGCTCCGGCGATCTCTCCAATCGCATTCGCGGCGGTCTCA 6180
CCGCCTCGCTGTCGGAAGTCGAGCGCATGCTCGGCGAAGCCGGCCGCGCTTCCGACGGCG 6240
CCGCCCAGAATATGCGCGAAGCGCTGCGCGAGGCGATCGATGATGCCGTCGGCCGCTTCT 6300
CCGGCGCCACCGAAGATATCCGCCGCTCCGCCGCCGACATCCGCAGCGAGCTCGACGCCA 6360
CCCGCGCCGAACTGAAGCGCGGCGCCTTCGATCTTCCCGAGGAAGCCAAGGAAAGCGCCT 6420
CGGCCATGCGCCGGGCCGTCGCCGAGCAGATTAAGGCGCTGCAGGATATCTCCCAACTCG 6480
TCGGCCGCTCCAGCCAACAGCTCGAGATCTCCGAACCCGTTGCCCGCACGCTCGCCCAGG 6540
TGCAGCCGGCCCCGCGTCCCTCCAAGCCGGTCGCAGCCCAGCCGCCGCAGGCGGCAGCAC 6600
CCCCGCCGATCGAAGCGACAGGTCTGCGCGGAACGATCGCACCATCTGCGCCGGCTGCCG 6660

CCCCCCAGCGCGCCGCGCCGCAGCCTGCCCCCGCTCGCCAGGAAACAGGCCGCCCGGAAA 6720

CAGGTCGCCAGGAGCCGGCTCGTCCGGAAAGCGGCGGCTGGATCAGCGATCTCCTGCGTG 6780

GCGCATCGCGTGAGGAGGCCGCCGACGAGGCCGCGGCCGCCCGTGTTCCGGCCCGCCCGG 6840

CGGAAACTGCCGCGCCGGCCGCGCGAAGCAATGACAGCCGCAATCCGCGCCACGTCGTCG 6900

AATCGCTGAACTCGCTCTCGGTCGATATCGCCCGCGCCATCGACCACGACGCTTCGGTCG 6960

ATCTCTGGCGCCGTTACCAGCGCGGCGAGCGCGATGTCTTCACCCGTCGCCTCTACACGC 7020

TGAAGGGCCAGCAGACCTTCGACGAGATCAAGCGCAAATATGACCGCGAACCGGAGTTCC 7080

GCACCGCCGTCGACCGCTACATCGGTGATTTCGAAAAGCTGCTCGCCGACGTCGCCCGCA 7140

CCGACCCGAACCGGACGGTGACGCAGTCCTACCTCACCTCGGATACGGGCAAGGTCTACA 7200

CGATGCTCGCTCATGCCGCAGGCAGGCTCGGCTGAGCCTGACTGGTACCCGCCGATTGGA 7260

AAGCCGCTTGACCACGTGTCGGGGCGGATCCGAATTCAAGCTTCAGTC 7308

XbaI recognition site, Ribosomal binding site, His-tag, Thrombin cleavage site, *rcy* gene sequence, BamHI recognition site

A.4 Phase contrast microscopy of protein overexpression phenotypes

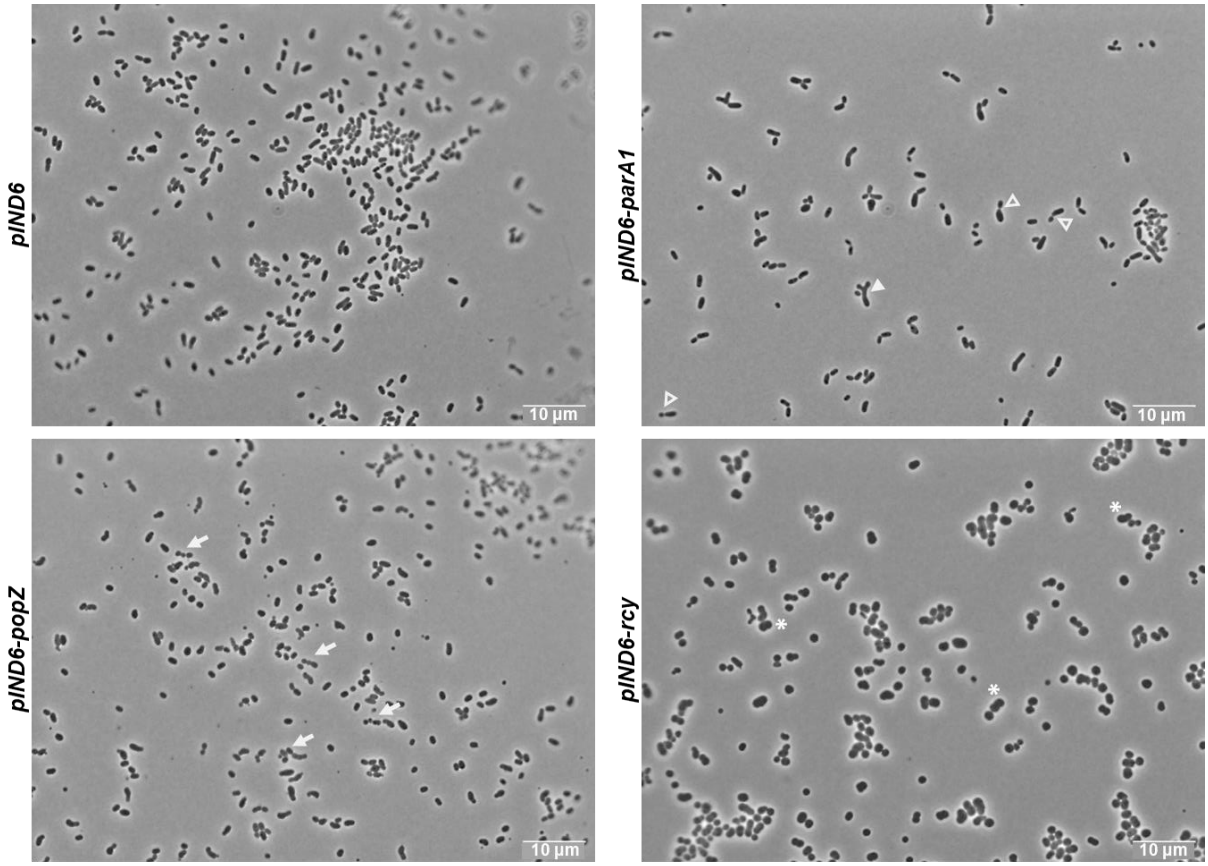


Figure A.2. Whole field-of-view phase contrast images of ParA1, PopZ and Rcy overexpression phenotypes within *R. leguminosarum*. *R. leguminosarum* 3841 strains carrying the plasmids pIND6 (empty vector control), pIND6-*parA1*, pIND6-*popZ* and pIND6-*rcy* were cultured in TY media for 20 hours. Protein overexpression was induced at 20 hours with 1 mM IPTG. Culture samples were isolated at 24 hours post-induction and viewed using phase contrast microscopy. A range of overexpression phenotypes were observed, including asymmetric cell division (hollow arrowheads, pIND6-*parA1*), branched cells (solid arrowhead, pIND6-*parA1*), clusters of small, spherical cellular bodies (arrows, pIND6-*popZ*), and enlarged cells with a 'budding-like' morphology (asterisks, pIND6-*rcy*). These images were used for further analysis to measure cell area, perimeter, width, and length. Scale bars represent 10 µm.