Genetic Determinants of Host-Specificity in

*Mesorhizobium loti* NZP2037

James Ting-Chun Chang

MSc

A thesis submitted for the degree of
Doctor of Philosophy
The University of Otago, Dunedin
New Zealand

April 2018
Abstract

The establishment of a successful symbiotic interaction between rhizobia and their respective legume hosts involves complex molecular interactions between the two prospective partners. The determinants behind the specificity between the legume host and its compatible symbiont are yet to be fully understood. Known determinants of specificity that are produced by the bacterial symbiont include modifications to the chitin backbone of Nod factors, signalling through polysaccharides such as exopolysaccharide (EPS), lipopolysaccharide and cyclic beta glucans, and secreted proteins transported via secretion systems such as the Type I, Type III, and Type IV secretion systems.

The *Mesorhizobium loti* strain NZP2037 belongs to a group of *M. loti* strains that exhibit a broad host range that encompasses the legumes *Lotus japonicus* Gifu and the extended host *Lotus pedunculatus*. In contrast, the closely related *M. loti* strain, R7A, belongs to the group of *M. loti* with a narrow host range that encompasses *L. japonicus* Gifu but not *L. pedunculatus*, a host that the strain induces only uninfected nodule primordia on. In addition, truncated EPS in R7A generates an incompatible nodulation phenotype on *L. japonicus* Gifu, while in NZP2037 the same truncated EPS generates only a delayed nodulation phenotype. These observations suggest that NZP2037 harbours unique genetic determinants and is thus of interest to study on a genetic level. Genetic comparisons between the two closely related strains of *M. loti* identified genes unique to NZP2037 hypothesised to be involved in its broad host range.

In this study, it was found that genes required for the nodulation of *L. pedunculatus* and for circumvention of the negative effect of truncated EPS on nodulation of *L. japonicus* Gifu were encoded on the symbiosis island of NZP2037, as a “hybrid” strain with the
chromosome of R7A and the symbiosis island of NZP2037 showed the NZP2037 phenotypes. Genes of interest unique to NZP2037 were targeted for disruption and the effects of these disruptions on nodulation of *L. pedunculatus* and their relation to the circumvention of the negative effect of defective EPS on nodulation of *L. japonicus* were investigated. The genes *nodU, nodFEG*, and *nodA2* were expendable for nodulation and were not involved in the circumvention of defective EPS signalling. Interestingly, *nodU* was found to promote nodule organogenesis, but by itself was not enough to restore nodulation to R7A EPS mutants. Similarly, additional genetic targets unique to NZP2037 relating to effectors of the Type IV secretion system and a Nod factor gene *mln399* were also disrupted. These genes were also expendable for nodulation and were not involved in the circumvention of defective EPS signalling. Furthermore, a strain with mutations in all of the above genes showed the same phenotype as NZP2037, eliminating the possibility that the genes were functionally redundant for nodulation.

The major finding in this study was the involvement of the Type I secretion system PrsDE and its respective effector Mln031 in the nodulation of the extended host *L. pedunculatus*. Mutants defective in *prsDE* had the unusual phenotype of inducing effective nodules on approximately half of the *L. pedunculatus* plants inoculated, and only uninfected nodule primordia on the other half. In addition, the initial gene of interest *nodO* was found to not be involved in the nodulation of the hosts tested. However, the presence of a construct encoding *prsDE* and *mln031* was unable to extend the host range of R7A to include that of *L. pedunculatus*, suggesting that other molecular signals must be required for the establishment of a successful symbiotic relationship with this host.

The results from this project provides insights into the unique Nod factor modifications
of NZP2037 and how they and the Type IV secretion system are expendable for nodulation of the extended host *L. pedunculatus* and circumvention of defective EPS signalling. The importance of the Type I secretion system PrsDE and Mln031 are highlighted as important determinants of symbiosis bringing the importance of these systems into the forefront of rhizobium-legume interactions.
Acknowledgments

I would like to thank my primary supervisor Professor Clive Ronson for taking me under his wing and allowing me to be part of lab607 and taking time out of his extremely busy schedule to provide guidance when required. My other supervisor Dr. John Sullivan was also a huge help during my study and his wealth of knowledge was extremely useful during my study.

I would also like to thank Professor Jens Stougaard and his lab at Aarhus University, in Denmark, and the CARB program that helped support my visits to two international conferences and to visit his Laboratory to help expand my horizons. I would also like to thank Dr. Simon Kelly for accommodating me during my stay over at Aarhus and showing me around the Stougaard lab during my stay.

I would like to thank the University of Otago for providing the PhD scholarship that supported me during my time of study and the Department of Microbiology and Immunology department for also providing funding for my PhD studies and travel expenses to help support myself during my period of study.

I would like to thank my friends and family for supporting me during my PhD study. The journey took longer than I had expected and their emotional and financial support helped me to overcome the pressures that I had taken upon myself.

Lastly, I would like to thank the 6th floor in the department for all the years of support that they have provided and made me feel like a welcome member of the floor.
# Table of Contents

Abstract ........................................................................................................................................... i  
Acknowledgments .............................................................................................................................. iv  
Table of Contents ........................................................................................................................... v  
List of Tables .................................................................................................................................... xii  
List of Figures ................................................................................................................................... xiii  
List of Abbreviations ....................................................................................................................... xvii

1 Introduction ..................................................................................................................................... 1  
1.1 Background ................................................................................................................................. 2  
1.2 Rhizobia ......................................................................................................................................... 3  
  1.2.1 *Mesorhizobium loti* .............................................................................................................. 4  
1.3 The host genus *Lotus* .................................................................................................................. 6  
1.4 Establishment of a symbiotic relationship .................................................................................... 8  
  1.4.1 The plant signalling molecule - Flavonoids ............................................................................ 8  
  1.4.2 NodD and *nod* gene induction ............................................................................................ 9  
  1.4.3 Entry into the host ................................................................................................................... 11  
  1.4.4 Nod factor ............................................................................................................................. 12  
  1.4.5 Recognition of Nod factors and *Ca^{2+}* signalling ............................................................... 21  
  1.4.6 Surface polysaccharides and their roles in nodulation ......................................................... 23  
  1.4.7 Secretion systems .................................................................................................................. 29  
1.5 Aims of this study ......................................................................................................................... 49  

2 Methods ......................................................................................................................................... 51  
2.1 Strains and plasmids used in this study ...................................................................................... 52
2.2 Media and growth conditions ..............................................................57
  2.2.1 Storage of bacterial strains .........................................................58
2.3 Enzymes and chemicals ..................................................................58
2.4 Restriction enzyme digestion .........................................................59
2.5 Agarose gel electrophoresis ............................................................59
2.6 Polymerase chain reaction (PCR) ....................................................59
  2.6.1 Primers used in this study ............................................................60
  2.6.2 Phusion PCR program: ...............................................................64
2.7 Plasmid DNA extraction ..................................................................64
  2.7.1 Alkaline lysis method .................................................................64
  2.7.2 Commercial plasmid/cosmid isolation kit ....................................65
2.8 Genomic DNA extraction ..................................................................65
  2.8.1 PrepMan™ Ultra genomic DNA preparations .............................65
  2.8.2 Commercial genomic DNA extraction kit .....................................65
2.9 Spectroscopy ....................................................................................65
2.10 Ligations .........................................................................................66
2.11 Electroporation ................................................................................66
  2.11.1 Preparation of E. coli .................................................................66
  2.11.2 Electroporation of E. coli ...........................................................67
2.12 Transfer of vectors by conjugation ...............................................67
2.13 DNA sequencing .............................................................................67
2.14 Computer analysis .........................................................................68
2.15 Mutagenesis ..................................................................................68
2.15.1 Homologous recombination protocol for deletion mutants ..................68
2.15.2 Insertion Duplication Mutants (IDM) ..............................................72
2.15.3 Mobilisation of the NZP2037 symbiosis island ..................................73
2.16 Cosmid library screen and transfer into *M. loti* ..................................73
2.17 Plant experiments ..................................................................................74
2.17.1 Seedling preparation and inoculation ..................................................74
2.17.2 Plant growth conditions ......................................................................75
2.17.3 Plant nodulation observations ..............................................................75
2.17.4 Isolation of rhizobia from nodules ......................................................76
2.17.5 Infection thread assays .........................................................................76
2.17.6 Regeneration of roots from seedling cuttings .......................................77

3 Symbiotic genes unique to NZP2037 and their effects on the nodulation of *L. pedunculatus* and *L. japonicus* Gifu .................................................................79
3.1 Introduction ................................................................................................80
3.2 NZP2037 and circumvention of defective EPS signalling on *L. japonicus* Gifu 82
3.3 The NZP2037 unique genes: *nodFEG, nodA2, nodU* .................................82
3.3.1 Construction and nodulation phenotypes of *nod exoU* double mutants...84
3.3.2 Further characterization of NZP2037 Δ *nodU exoU::pFUS2* ...................86
3.3.3 Complementation of *nodU* and *exoU* mutations ...............................89
3.4 Effect of *nodU* in a R7A genetic background ...........................................94
3.5 The transconjugant R7A<sup>C</sup>:NZP2037<sup>I</sup>.................................................................97

3.5.1 R7A<sup>C</sup>:NZP2037<sup>I</sup> exoU::pFUS2 does not exhibit a delayed nodulation phenotype on *L. japonicus* Gifu ........................................................................................................99

3.5.2 R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ*nodU* and R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ*nodU exoU::pFUS2* ...100

3.6 Deletion of unique genes involved in modification of the non-reducing end of the Nod factor in the R7A<sup>C</sup>:NZP2037<sup>I</sup> background .................................................................103

3.6.1 Generation of the mutant JC01 (R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ*nodUFEGA2*)...........103

3.6.2 Nodulation phenotypes of JC01 and JC02 on *L. japonicus* Gifu and *L. pedunculatus* ........................................................................................................................................105

3.7 A new target - Mln399, a putative glycosyl transferase .............................................105

3.7.1 Generation of mutants R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ*mln399* and R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ*mln399 exoU::pFUS2* ..................................................................................................................107

3.7.2 Nodulation phenotypes of R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ*mln399* and

R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ*mln399 exoU::pFUS2* on *L. japonicus* Gifu and *L. pedunculatus* ........................................................................................................................................108

3.8 The Type IV secretion system ......................................................................................109

3.8.1 Generation of the mutant strains R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔTypeIV and

R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔTypeIV exoU::pFUS2..............................................................................................110
3.8.2 Nodulation phenotypes of R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔTypeIV and R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔTypeIV <i>exoU</i>:pFUS2 on <i>L. japonicus</i> Gifu and <i>L. pedunculatus</i>

3.9 Deletion of <i>nodUFEGA2</i> and <i>min399</i>/Type IV secretion system within a single strain

3.9.1 Generation of mutants JC03 and JC04

3.9.2 Generation of mutants JC05 and JC06

3.9.3 Nodulation phenotypes of JC03, JC04, JC05, and JC06 on <i>L. japonicus</i> Gifu and <i>L. pedunculatus</i>

3.10 Continuation of markerless deletions targets unique to the NZP2037 island in the R7A<sup>C</sup>:NZP2037<sup>I</sup> transconjugant background

3.10.1 Generation of mutants JC07 and JC08

3.10.2 Verification of JC07

3.10.3 Nodulation phenotypes of JC07 and JC08 on <i>L. japonicus</i> Gifu and <i>L. pedunculatus</i>

3.11 Introduction of ‘<i>nodO</i>-9.3kb’ into JC07

3.11.1 Generation of mutants JC09 and JC10

3.11.2 Generation of R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ(<i>nodO</i>-9.3kb) and R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ(<i>nodO</i>-9.3kb) <i>exoU</i>:pFUS2
3.11.3 Nodulation phenotypes of JC09, JC10, R7A\textsuperscript{C}:NZP2037\textsuperscript{I} \(\Delta (nodo-9.3kb)\), and R7A\textsuperscript{C}:NZP2037\textsuperscript{I} \(\Delta (nodo-9.3kb)\ exoU::pFUS2\) on \textit{L. japonicus} Gifu and \textit{L. pedunculatus} .......................................................... 119

3.12 Discussion ........................................................................................................ 120

4 \textbf{The journey from NodO, to the Type I secretion system PrsDE and its secreted effector Mln031} ................................................................. 135

4.1 Introduction ...................................................................................................... 136

4.2 Results ........................................................................................................... 137

4.2.1 Phenotype of NZP2037 \(\Delta (nodo-9.3kb)\) ...................................................... 137

4.2.2 The mutant \(\Delta (nodo-9.3kb)\) is affected in its ability to initiate infection threads ........................................................................................................... 142

4.2.3 Plant genetics have little role in the phenotype of \(\Delta (nodo-9.3kb)\) ...... 143

4.2.4 Recovery of \(\Delta (nodo-9.3kb)\) from plants.............................................. 145

4.2.5 Exopolysaccharide plays no role in the phenotype of \(\Delta (nodo-9.3kb)\) ... 147

4.2.6 Complementation of \(\Delta (nodo-9.3kb)\) ...................................................... 147

4.2.7 Cosmid complementation of nodo ............................................................. 151

4.2.8 Generation and effect of a single markerless nodo deletion in NZP2037 .................................................................................................................. 155

4.2.9 The genes noe\textit{KJ} as the next target for mutagenesis......................... 159

4.2.10 The next target, the Type I secretion system genes prs\textit{DE}............... 162
4.2.11 Complementation of NZP2037 ΔprsDE with a functional mln031 .......172

4.3 Discussion ........................................................................................................177

5 Concluding remarks .............................................................................................191

6 References .............................................................................................................197

7 Appendix ................................................................................................................218
List of Tables

Table 1. Modifications of Nod factor structures produced by *Mesorhizobium* in relation to Figure 1.1 ................................................................. 15
Table 2. Bacterial strains used in this study .................................................................................................................. 52
Table 3. Plasmid and Cosmids used in this study ........................................................................................................... 55
Table 4. Antibiotics used in this study ............................................................................................................................ 58
Table 5. Primers used in this study ................................................................................................................................. 60
Table 6. gBlock used for Gibson assembly to generate the construct pFAJ1700-nodOprsDEmln031 ........................................................................ 71
Table 7. Homologues of PrsE within NZP2037 and their corresponding homologue in R7A ........................................................................................................... 170
Table 8. Predicted putative effectors of T1SS containing a RTX Ca^{2+} binding domain ........................................................................................................... 172
List of Figures

Figure 1.1 General structure of Nod factor produced by mesorhizobia. .......................... 13
Figure 1.2 Distribution of bacterial surface polysaccharides involved in symbiosis ... 24
Figure 1.3 R7A EPS biosynthesis ................................................................. 25
Figure 1.4 The HlyA Type I secretion system of E. coli .............................................. 33
Figure 1.5 Examples of genetic orientations of Type I secretion systems ...................... 36
Figure 1.6 The VirB/D4 Type IV secretion system .................................................... 43
Figure 3.1 Effect of R7A and NZP2037 exoU mutants on nodulation of L. japonicus Gifu. ......................................................................................................................... 83
Figure 3.2 The nodulation kinetics of exoU double mutants on A) L. pedunculatus and B) L. japonicus Gifu ................................................................. 85
Figure 3.3 Growth of NZP2037 ΔnodU exoU::pFUS2 in G/RDM broth culture............ 87
Figure 3.4 The nodulation phenotype of a nodU exoU double mutant that had a ‘normal’ growth phenotype on A) L. pedunculatus and B) L. japonicus Gifu ......................... 88
Figure 3.5 Attempted complementation of select exoU mutants on L. japonicus Gifu showed that exoU complementation restored wildtype phenotype but nodU did not ... 90
Figure 3.6 Average number of primordia induced is increased in exoU mutants that harboured pPROBEKT-nodU on L. japonicus Gifu ....................................................... 92
Figure 3.7 L. japonicus Gifu inoculated with nodU mutants of NZP2037 .................... 93
Figure 3.8 The heterologous expression of nodU from NZP2037 in trans in the mutant R7A ΔexoU was unable to restore effective nodulation of L. japonicus Gifu, but increases the number of primordia formed ................................................................. 95
Figure 3.9 L. japonicus Gifu inoculated with R7A harbouring a plasmid containing nodU from NZP2037 ......................................................................................... 96
Figure 3.10 The nodulation phenotype of the transconjugant strain R7AΔC:NZP2037Δl 98
Figure 3.11. The presence of truncated EPS in the R7A<sup>C</sup>:NZP2037<sup>I</sup> transconjugant background does not interfere with its ability to nodulate <i>L. japonicus</i> Gifu.

Figure 3.12. PCR confirmation of a successful transconjugant.

Figure 3.13. Nodulation kinetics of strains with a <i>nodU</i> and/or <i>exoU</i> mutation in an R7A<sup>C</sup>:NZP2037<sup>I</sup> transconjugant background.

Figure 3.14. PCR screen for a NZP2037 Δ<i>nodU</i> Δ<i>nodF</i>Δ<i>E</i>Δ<i>G</i>Δ<i>A</i>2 mutant.

Figure 3.15. The nodulation phenotypes of JC01 and JC02.

Figure 3.16. Genetic orientation of the putative <i>nodACIJ-nolO-mln399</i> operon.

Figure 3.17. Restriction digest of pJQ200SK-<i>mln399</i> with the enzymes BamHI and Apal.

Figure 3.18. Nodulation phenotypes of R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ<i>mln399</i> and R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ<i>mln399 exoU::pFUS2</i>.

Figure 3.19. Genetic organization of the T4SS and associated proteins in R7A and NZP2037.

Figure 3.20. Nodulation phenotype of R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ<i>TypeIV</i> and R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ<i>TypeIV exoU::pFUS2</i>.

Figure 3.21. Nodulation phenotypes for JC03, JC04, JC05 and JC06.

Figure 3.22. Reads obtained from whole genome sequencing do not align to deleted genes.

Figure 3.23. Nodulation phenotypes of JC07 and JC08.

Figure 3.24. Nodulation phenotypes of JC09 and JC10.

Figure 3.25. Comparison between the predicted Nod factors produced by A) R7A and B) NZP2037.

Figure 4.1. The genetic organisation of the DNA proximal to <i>nodO</i> and the 9345 bp markerless DNA deletion in the mutant Δ(<i>nodO</i>-9.3kb).
Figure 4.2. Nodulation kinetics of NZP2037 Δ(nodO-9.3kb) on different Lotus hosts. .................................................................................................................................................................................. 139

Figure 4.3. The Δ(nodO-9.3kb) mutant has two different phenotypes on L. pedunculatus. ........................................................................................................................................................................................................... 139

Figure 4.4 Comparison between wildtype NZP2037 and plants that formed effective nodules when inoculated with Δ(nodO-9.3kb) on L. pedunculatus and the exclusion of data from uninfected plants. ........................................................................................................................................................................................................... 141

Figure 4.5. The onset of nodule/primordia for Δ(nodO-9.3kb) on L. pedunculatus is similar to each other........................................................................................................................................................................................................... 141

Figure 4.6. Number of infection threads formed by NZP2037 and Δ(nodO-9.3kb) on L. pedunculatus at 11 days post inoculation. ........................................................................................................................................................................................................... 143

Figure 4.7. Nodulation of L. pedunculatus plants after root regeneration at 14 days post inoculation with either the wildtype or the Δ(nodO-9.3kb) mutant. ........................................ 145

Figure 4.8. Bacteria recovered from effective nitrogen fixing nodules of L. pedunculatus plants inoculated with Δ(nodO-9.3kb) and re-inoculated onto L. pedunculatus......... 146

Figure 4.9. Double mutants of Δ(nodO-9.3kb) and exoU and their nodulation kinetics. ........................................................................................................................................................................................................... 148

Figure 4.10. Restriction digest of the vector pFAJ1700-nodOprsDE using HindIII.... 150

Figure 4.11. Attempted complementation of Δ(nodO-9.3kb) on L. pedunculatus...... 151

Figure 4.12. PCR screen for the presence of nodO in the NZP2037 cosmid library.... 152

Figure 4.13. Schematic of NZP2037 DNA present within the cosmids 6B1.5 and 8B4.5. ........................................................................................................................................................................................................... 153

Figure 4.14. Effect of the cosmids 6B1.5 and 8B4.5 on L. pedunculatus ................... 154

Figure 4.15. Effect of cosmids 6B1.5 and 8B4.5 in the NZP2037 and R7A wildtype backgrounds on L. pedunculatus. ........................................................................................................................................................................................................... 155
Figure 4.16 Generation of the deletion construct pJQ200SK-nodO single and PCR confirmation of the nodO single deletion mutant.................................................. 157
Figure 4.17. Plasmid Map for pJQ200SK-nodOs. .................................................. 158
Figure 4.18. The effect of a markerless nodO deletion in NZP2037 on L. pedunculatus. .................................................................................................................. 159
Figure 4.19. PCR confirmation of a markerless deletion mutant of noeKJ. .............. 161
Figure 4.20. The genes noeKJ in the nodulation of L. pedunculatus. ..................... 162
Figure 4.21. PCR results for NZP2037 ΔprsDE......................................................... 164
Figure 4.22. Schematic of the genetic orientation present within NZP2037 ΔprsDE. 166
Figure 4.23. Nodulation kinetics of NZP2037 ΔprsDE on L. pedunculatus .......... 167
Figure 4.24. Complementation of NZP2037 ΔprsDE with the cosmid 6B1.5 and pFAJ1700-nodOprsDE. ...................................................................................... 168
Figure 4.25. Protein alignment of Mln031 to NodO. ............................................ 173
Figure 4.26. Approach to restore full length mlno31 in the construct pFAJ1700- nodOprsDE. ................................................................................................. 174
Figure 4.27. PCR screen for a successful ligation of the plasmid pFAJ1700-nodO prsDEmln031................................................................................................. 175
Figure 4.28. Complementation of NZP2037 ΔprsDE with the vector pFAJ1700- nodOprsDEmln031....................................................................................... 176
Figure 4.29. Predicted conserved domains of Mln031 and alignment to homologues. 181
Figure 4.30. Alignment of NZP2037 NodO with other NodO and Mln031.......... 183
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>°C</td>
<td>Degrees Celsius</td>
</tr>
<tr>
<td>aa</td>
<td>Amino acid</td>
</tr>
<tr>
<td>ABC</td>
<td>ATP binding cassette</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic local alignment search tool</td>
</tr>
<tr>
<td>bp</td>
<td>Base pairs</td>
</tr>
<tr>
<td>BTB</td>
<td>Bromothymol blue</td>
</tr>
<tr>
<td>bv</td>
<td>Biovar</td>
</tr>
<tr>
<td>CBG</td>
<td>Cyclic beta glucans</td>
</tr>
<tr>
<td>cm</td>
<td>Centimetre</td>
</tr>
<tr>
<td>cv</td>
<td>Cultivar</td>
</tr>
<tr>
<td>Da</td>
<td>Daltons</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxyribonucleoside triphosphate</td>
</tr>
<tr>
<td>EPS</td>
<td>Exopolysaccharide</td>
</tr>
<tr>
<td>F</td>
<td>Faraday</td>
</tr>
<tr>
<td>g</td>
<td>Gravitational force</td>
</tr>
<tr>
<td>G/RDM</td>
<td>Glucose rhizobium defined media</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-acetyl-D-glucosamine</td>
</tr>
<tr>
<td>Gm</td>
<td>Gentamicin</td>
</tr>
<tr>
<td>h</td>
<td>Hour</td>
</tr>
<tr>
<td>HMW</td>
<td>High molecular weight</td>
</tr>
<tr>
<td>IDM</td>
<td>Insertion duplication</td>
</tr>
<tr>
<td>IMC</td>
<td>Inner membrane complex</td>
</tr>
<tr>
<td>JGI</td>
<td>Joint genome institute</td>
</tr>
<tr>
<td>JUMPstart</td>
<td>Just upstream of many polysaccharide-associated starts</td>
</tr>
<tr>
<td>kb</td>
<td>Kilobase</td>
</tr>
<tr>
<td>KPS</td>
<td>K-polysaccharides</td>
</tr>
<tr>
<td>L</td>
<td>Litre</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani</td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysaccharides</td>
</tr>
<tr>
<td>LysM</td>
<td>Lysin motif</td>
</tr>
<tr>
<td>M</td>
<td>Molar</td>
</tr>
<tr>
<td>Mb</td>
<td>Megabase</td>
</tr>
<tr>
<td>MCS</td>
<td>Multi-cloning site</td>
</tr>
<tr>
<td>MDa</td>
<td>Megadaltons</td>
</tr>
<tr>
<td>MFP</td>
<td>Membrane fusion protein</td>
</tr>
<tr>
<td>µg</td>
<td>Microgram</td>
</tr>
<tr>
<td>µL</td>
<td>Microlitre</td>
</tr>
<tr>
<td>µM</td>
<td>Micromolar</td>
</tr>
<tr>
<td>µm</td>
<td>Micrometre</td>
</tr>
<tr>
<td>mg</td>
<td>Milligram</td>
</tr>
<tr>
<td>min</td>
<td>Minute</td>
</tr>
<tr>
<td>mL</td>
<td>Millilitre</td>
</tr>
<tr>
<td>mM</td>
<td>Millimolar</td>
</tr>
<tr>
<td>NFR</td>
<td>Nod factor receptor</td>
</tr>
<tr>
<td>nm</td>
<td>Nanometres</td>
</tr>
<tr>
<td>Nod factor</td>
<td>Nodulation factor</td>
</tr>
<tr>
<td>NS</td>
<td>Non-symbiotic</td>
</tr>
<tr>
<td>OMC</td>
<td>Outer membrane complex</td>
</tr>
<tr>
<td>OMF</td>
<td>Outer membrane factor</td>
</tr>
<tr>
<td>ops</td>
<td>Operon polarity suppressor</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>pg</td>
<td>Picogram</td>
</tr>
<tr>
<td>phe-tRNA</td>
<td>Phenylalanine transfer RNA</td>
</tr>
<tr>
<td>pv</td>
<td>Pathovar</td>
</tr>
<tr>
<td>RDM</td>
<td>Rhizobium defined media</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>rpm</td>
<td>Revolutions per minute</td>
</tr>
<tr>
<td>RTX</td>
<td>Repeat in toxin</td>
</tr>
<tr>
<td>s</td>
<td>Second</td>
</tr>
<tr>
<td>S/RDM</td>
<td>Sucrose rhizobium defined media</td>
</tr>
<tr>
<td>Sec system</td>
<td>General secretion system</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>SOB</td>
<td>Super optimal broth</td>
</tr>
<tr>
<td>sp.</td>
<td>Species</td>
</tr>
<tr>
<td>T1SS</td>
<td>Type I secretion system</td>
</tr>
<tr>
<td>T3SS</td>
<td>Type III secretion system</td>
</tr>
<tr>
<td>T4CP</td>
<td>Type IV coupling protein</td>
</tr>
<tr>
<td>T4SS</td>
<td>Type IV secretion system</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-acetate</td>
</tr>
<tr>
<td>TAT</td>
<td>twin arginine translocation</td>
</tr>
</tbody>
</table>
1 Introduction
1.1 Background

Rhizobia are bacteria found in the soil that have the ability to establish a symbiotic relationship with host legumes. This relationship occurs within the plant hosts roots, that develop specialised root organs known as root nodules to accommodate the bacteria. Within these organs, the plant provides nutrients for the growing bacteria and, in turn, the bacteria reduce di-atmospheric nitrogen (N$_2$) into the more biologically available ammonia (NH$_3$) via the rhizobial enzyme nitrogenase.

The establishment of a successful symbiotic relationship between legume and rhizobia involves complex molecular signals from both participants. Major factors involved in successful establishment include flavonoids secreted by plants and Nodulation (Nod) factors produced by the rhizobia. In addition, secondary effectors such as surface polysaccharides and bacterial secretion systems can also determine a successful symbiotic outcome (Perret et al., 2000, Spaink, 2000, Jones et al., 2007, Krehenbrink & Downie, 2008, Downie, 2010, Oldroyd, 2013, Janczarek et al., 2015, Kawaharada et al., 2015, Nelson & Sadowsky, 2015). The complex interactions involving these multiple factors culminate in host-range variations between different rhizobial strains, with some strains exhibiting broad host ranges, such as Sinorhizobium fredii NGR234 which has a host range of at least 112 different genera (Pueppke & Broughton, 1999). In contrast, some strains exhibit a small range of compatible hosts such as Sinorhizobium meliloti that forms nodules on legumes of the genera Medicago, Melilotus and Trigonella. (Denarie et al., 1992). Taken together, multiple variables are involved in the establishment of a symbiotic relationship and it is the culmination of all these variables that ultimately determines a strain’s host range.
In this study, we investigate the differences between two closely related strains of *Mesorhizobium loti*, strain R7A representing the group of *M. loti* with a narrow host range (includes strains MAFF303099 and NZP2213) and strain NZP2037, which represents the group of *M. loti* that has a broad host range (includes strains NZP2014, NZP2042 and SU343). The differences in host range are apparent on *Lotus pedunculatus*, where the NZP2037 group form effective nodules whereas the R7A group only induce uninfected primordia (Pankhurst et al., 1979). The genetic differences between the symbiosis islands of these two strains of *M. loti* have been identified (Kasai-Maita et al., 2013) and in this study experiments were designed to determine what genetic factors determine this difference in host-range.

### 1.2 Rhizobia

Rhizobia are gram-negative bacteria first isolated in the 19\textsuperscript{th} century from plant nodules and named *Bacillus radicicola* (Beijerinck, 1888). This strain has subsequently been renamed *Rhizobium leguminosarum* (Laranjo et al., 2014) and since then, strains that have been isolated from within root nodules were classified under the genus *Rhizobium* (Young & Johnston, 1989). As more knowledge was garnered from these strains and the introduction of more modern techniques, three genera were eventually defined based on their growth phenotypes. The three genera were *Rhizobium* (fast growing), *Bradyrhizobium* (slow growing) and *Azorhizobium* (can fix nitrogen when free living) (Jarvis et al., 1986, Dreyfus et al., 1988, Young & Johnston, 1989). The *Mesorhizobium* was separated from the *Rhizobium* genus with ‘meso’ denoting that species within this genus grow at a slower rate when compared to the fast-growing strains within the *Rhizobium* genus, but not to the extent of the *Bradyrhizobium* (Jarvis et al., 1997).
1.2.1 *Mesorhizobium loti*

There are at least 30 characterised strains of *Mesorhizobium* based on their 16S ribosomal RNA and *nodC* genes (Laranjo *et al.*, 2014). The genus is of interest for scientific research as they have the ability to establish symbiotic relationships with a wide range of host legumes ranging from the large tree *Robina psudoacacia* to the small pasture legume *Biserrula pelecinus*. Mimosoid temperate legumes such as *Cicerareitinum*, *Astragalus* spp., *Amorpha fruticose*, *Caragana* spp. and *Lotus* spp. are also included in the range of potential hosts (Laranjo *et al.*, 2014). In contrast to other rhizobia such as *R. leguminosarum* and *S. fredii* NGR234 which utilise large plasmids that encode the genes required for symbiosis (Freiberg *et al.*, 1997), *M. loti* utilises genes encoded within the chromosome derived from integrative mobile genetic elements known as symbiosis islands (Sullivan & Ronson, 1998), although there are rare cases where symbiotic genes may also be located within plasmids (Zhang *et al.*, 2000). Members of the *Mesorhizobium* genus are of particular interest to study due to their relatively conserved core genes, alluding that the less conserved accessory genes have roles in host specificity (Laranjo *et al.*, 2012).

1.2.1.1 Strain R7A

The *M. loti* strain R7A is a re-isolate of the strain ICMP3513 from a root nodule of *Lotus corniculatus* in Lammermoor, Central Otago, New Zealand. ICMP3513 was the recommended inoculant strain for *Lotus corniculatus* in New Zealand and is also known under the strain names NZP2238 and Lc265Da (Sullivan & Ronson, 1998). R7A forms effective nitrogen fixing nodules with the plants *Lotus tenuis*, *L. corniculatus*, *L. japonicus*, *Lotus filicaulis* and *Lotus burttii*, but is only able to form uninfected nodule primordia on the hosts *Leucaena leucocephala* and *L. pedunculatus* (Hubber *et al.*, 2004,
Rodpothong et al., 2009). In 2014, the genome of R7A was sequenced. It has a size of around 6.5 Mb and does not contain any plasmids. It is predicted to encode 6,398 genes, of which 6,323 encode for proteins while 75 encode RNA only (Kelly et al., 2014b). R7A is able to form symbiotic relationships with host legumes due to the presence of a symbiosis island of 502 kb that is integrated into the phe-tRNA (phenylalanine transfer RNA) gene within the chromosome. Upon integration of the island into the chromosome, the phe-tRNA is reconstituted at the left end of the symbiosis island (Sullivan & Ronson, 1998, Ramsay et al., 2006).

1.2.1.2 Strain NZP2037

The M. loti strain NZP2037 was originally isolated in 1982 from root nodules of the plant Lotus divaricatus (Jarvis et al., 1982). In contrast to R7A, NZP2037 has a broad host-range that encompasses the R7A compatible hosts in addition to extended hosts including L. pedunculatus, L. leucocephala, Lotus subbiflorus, Lotus angustissimus and some species of Carmichaelia, Ornithopus, Clianthus, and Vigna (Pankhurst et al., 1987). At the beginning of this study, the symbiosis island of NZP2037 had been sequenced by two different groups (Kasai-Maita et al., 2013, Kelly et al., 2014a) and the two sequences were largely comparable to one another, albeit with some assembly differences present in highly repetitive regions, such as that in the nodO locus (this study). The genome of NZP2037 is around 7.46 Mb in length and is predicted to contain 7,388 genes of which 7,318 encode proteins, while 70 encode RNAs. The symbiosis island within NZP2037 is around 530 kb and has been termed tri-partite as the island integrates into the chromosome at three separate regions (Kasai-Maita et al., 2013, Kelly et al., 2014a, Haskett et al., 2016). The NZP2037 genome has been noted to harbour a plasmid with a size of ~240 MDa (~360 kb) that was named pRlo2037 but is not essential for the nodulation of L.
*pedunculatus* (Pankhurst *et al.*, 1983, Pankhurst *et al.*, 1986, Kelly *et al.*, 2014a). Although pRlo2037 has no observable role involved in host-specificity, strains of NZP2037 cured of pRlo2037 exhibited an increase in the number of nodules formed, suggesting some influence on nodulation. In addition, an increase in the competitiveness of the strain as a symbiont on *L. pedunculatus* was also observed, and this phenotype was abated with the re-introduction of pRlo2037 (Pankhurst *et al.*, 1986).

### 1.3 The host genus *Lotus*

The genus *Lotus* includes over 200 species and is found world-wide, except in the cold extremities and the low-land tropical areas of Southeast Asia, South America and Central America. The probable regional centre for the genus is proposed to be from the Mediterranean basin, wherein the greatest diversity of species is present (Handberg & Stougaard, 1992, Díaz *et al.*, 2005). The genus consists of both annual and perennial species that have adapted to a wide range of ecological habitats. Plants feature erect or decumbent stems with penta-foliate leaves that are generally a green/grey-green colour and have two leaflets at the petiole base. Seedpods are about 2 cm long and change colour from green to brown as they mature and each seedpod can accommodate around 20 seeds (Díaz *et al.*, 2005). The *Lotus* hosts that were primarily used in this study are *L. japonicus* Gifu and *L. pedunculatus*.

The legume *L. japonicus* Gifu is an internationally recognised model legume for symbiotic studies and was chosen due to its ideal characteristics. *L. japonicus* Gifu is a perennial, small bushy plant with a relatively short generation time, taking around 7 weeks to develop from a seed to a flowering plant. It is a diploid with a small chromosome number (*2n* = 12) and a small genome size of around 0.5 pg per haploid genome.
Additionally, it is susceptible to genetic manipulation via transformation with *Agrobacterium tumefaciens*. Mature plants develop large self-fertile flowers, that produce large seed numbers (Handberg & Stougaard, 1992, Udvardi et al., 2005). As an internationally accepted model system for symbiotic studies, experiments are largely focused on root development and the root nodule organ that develops in an effective symbiosis. Effective nitrogen fixing nodules formed on this plant have a determinate and regular spherical shape that are yellow/brown in colouration. The central zone of the nodules primarily contains infected host cells, packed with bacterial endosymbionts that are enclosed within a central vacuole (Szczyglowski et al., 1998).

The legume *L. pedunculatus* (also known under the names *Lotus uliginosis*, and *Lotus major*) is a perennial, five-leaved legume that has been naturalised in New Zealand and originated from European and North African regions (Armstrong, 1974). It was used as livestock forage in pasture within the New Zealand rough land, where the usual foraging plant, white clover, fails to perform (Charlton & Stewart, 1999). *L. pedunculatus* is of particular value as livestock forage as the plant incorporates condensed tannins that are non-bloating when ingested. These tannins also provide additional nutrition in the rumen of livestock which contributes to increased lean meat and wool production (Charlton & Stewart, 1999). The condensed tannins (also known as flavolans or proanthocyanin) have bactericidal effects on some rhizobial species, and may contribute to the host-range of this cultivar (Pankhurst et al., 1982). In this study, the main cultivar used for plant studies was *L. pedunculatus* Grasslands Trojan, which is a tetraploid cultivar of *L. pedunculatus* (Widdup et al., 2004). Effective nitrogen fixing nodules that form on *L. pedunculatus* are determinate with a regular spherical shape and exhibit a red-pinkish hue in colouration (Pankhurst et al., 1979).
1.4 Establishment of a symbiotic relationship

The main signalling molecules involved in symbiosis secreted by the plant are flavonoids (or flavonoid like molecules such as iso-flavonoids). Flavonoids are recognised by compatible rhizobia in the rhizosphere, which in turn, produce signalling molecules in the form of Nod factors that are in turn recognized by the plant (Spaink, 2000, Downie, 2010, Oldroyd et al., 2011, Oldroyd, 2013). The end result of a successful symbiotic relationship is the eventual formation of the nodule on plant roots and differentiation of bacteria into their nitrogen-fixing bacteroid form that then fix atmospheric nitrogen into biologically available forms of nitrogen.

1.4.1 The plant signalling molecule - Flavonoids

One of the first steps in the establishment of a successful symbiotic relationship is the production and release of signalling molecules by plants in the form of flavonoids (or flavonoid-like molecules) into the rhizosphere. Flavonoid production is linked to recognition of environmental stresses such as starvation of nitrogen or phosphorus (Coronado et al., 1995, Juszczuk et al., 2004). Flavonoids are secondary metabolites produced by plants, with over 10,000 different flavonoids documented thus far (Winkel-Shirley, 2001, Hassan & Mathesius, 2012). This large diversity in flavonoid variation is derived from the multitude of basal flavonoid structures available including: flavones, flavonols, flavan-3-ols, flavanones, isoflavonoids, isoflavans and pterocarpanes (Hassan & Mathesius, 2012). In addition, the basal flavonoid structure can undergo further modification including: glycosylation, malonylation, methylation, hydroxylation, acylation, prenylation, and polymerization (Winkel-Shirley, 2001). The main flavonoids known to be involved in symbiosis are mostly derived from 2-phenyl-1,4-benzopyrone
(Perret et al., 2000). The main role for flavonoids are to act as a diffusible signalling molecule that are recognised by the bacterial transcriptional regulator NodD which up-regulates multiple nodulation genes that lead to an effective symbiotic relationship (Spaink, 2000). Interestingly, flavonoids may also have a role in host-range with differing effects on different rhizobia. An example of this can be observed in the flavonoids daidzein and genistein which act as inducers for *S. fredii* NGR234 and *Bradyrhizobium japonicum*. However, the same flavonoids act as anti-inducers on *R. leguminosarum* bv. *trifolii* and *R. leguminosarum* bv. *viciae* (Cooper, 2007). Flavonoids may also act as chemo-attractants for rhizobia during attachment and colonisation of the plant roots (Caetano-Anolles et al., 1988). The natural flavonoid inducer produced by *Lotus* that induces *M. loti* has yet to be identified (Aoki et al., 2000).

### 1.4.2 NodD and nod gene induction

Flavonoids secreted by compatible plants are recognized by NodD. NodD belongs to a family of LysR-like transcriptional regulators that recognise and bind to a highly conserved 47-bp DNA motif termed the *nod* box (Rostas et al., 1986, Fisher & Long, 1993). NodD is usually localised to the membrane of a bacterial cell (e.g. in *R. leguminosarum* bv. *viciae*), although some NodD have been found in soluble cell fractions (e.g. *S. meliloti*) (Schlaman et al., 1992). Regardless of NodD localisation, the general consensus is that NodD requires a flavonoid inducer in order to convert into its active form. The interactions between NodD and its compatible flavonoid occurs within the bacterial membrane, but the specifics regarding the interactions between NodD and flavonoids have yet to be fully understood (Schlaman et al., 1998). Activated NodD binds to the *nod* box, which generates a bend in the DNA assisting RNA polymerase recognition of the promoter region, allowing for the transcription of these genes (Fisher & Long, 1993,
Truncation of the nod box generally leads to inactivation of downstream genes, with examples observed in the promoter region of nodSU in *Sinorhizobium fredii* (Krishnan et al., 1992). Additionally, even if a flavonoid interacts with NodD, transcription does not always occur, and it is suggested that non-inducing flavonoids may compete for binding sites of functional gene-inducing flavonoids resulting in decreased expression (Peck et al., 2006). Some flavonoids only interact with certain NodD homologues, to counter this some bacterial strains harbour multiple copies of nodD within their genomes enabling response to different flavonoids (Spaink et al., 1987, Hartwig et al., 1990, Györgypal et al., 1991, Perret et al., 2000). Active NodD up-regulates a plethora of downstream genes involved in symbiosis, such as the Type III secretion system (T3SS) and a multitude of nod genes involved in Nod factor synthesis (detailed in section 1.4.4), however, many genes that are upregulated have yet to be properly characterised (Cooper, 2007).

Although NodD is the most well understood regulatory protein involved in the induction of symbiotic genes, other regulators have also been identified. In the NodVW two component regulatory system, NodV recognises the flavonoid genistein and phosphorylates NodW, which acts as a transcriptional regulator that up-regulates nod genes (Sanjuan et al., 1994, Loh et al., 1997). Similarly, the protein SymR forms a self-amplifying regulatory circuit with NodD3 to induce transcription of nod genes (Kondorosi et al., 1991a, Swanson et al., 1993). NolR is a repressor of nod gene induction and binds to the promoter regions of nodD (Kondorosi et al., 1991b). NolR regulates the core nod genes involved in Nod factor synthesis and also acts as a regulator of the T3SS (Cren et al., 1995, Lopez-Baena et al., 2008). Lastly, in *Medicago sativa*, the transcriptional regulator EmrR, which is a requirement for effective nodulation,
recognises the flavonoids apigenin, naringenin, galangin, and quercetin and may provide hints towards a different NodD independent signalling pathway (Rossbach et al., 2014).

1.4.3 Entry into the host

In order for effective symbiosis to occur, the bacteria must first find a way to enter the plant cell. One of the first hurdles for potential symbionts is the attachment and colonization of the plant root hair. Attachment is an important step for effective nodulation as the adhesion of rhizobia to the root hairs allows for the amplification of localised rhizobial cell numbers within the rhizosphere, and this in turn, increases the probability of strains being infective (Oldroyd & Downie, 2008). Bacterial adhesion to the root hair can be promoted via the recognition of specific lectins present on the surface of host-plant root hair tips, and binding to these lectins allows for direct cell-cell contact between compatible bacteria and the host. Upon binding to these lectins, the increase in the localized concentration of signals involved in nodulation would then trigger downstream effects that determine the outcome of symbiosis (Hirsch, 1999, van Rhijn et al., 2001, Rodríguez-Navarro et al., 2007).

After attachment of the rhizobial cell to the plant root hair, there are multiple approaches that a potential symbiont may use to enter the plant cell. Of the potential infection routes, the two main ones presented here are root hair crack entry and infection threads. Crack entry is relatively simple and occurs when the epidermis of the plant cell is disrupted (the ‘crack’) which naturally occur during the development and growth of lateral plant root cells. The cracks provide an entry point from which the bacteria can enter the cortical plant cells (Oldroyd & Downie, 2008).
The second method of entry into the plant cell is through the formation of infection threads. During formation of an infection thread, the bacterial cell communicates with the host via the expression of the Nod factors (discussed in more detail below). If both participants are compatible, the plant root hair will curl and form into a shape reminiscent of a ‘shepherds crook’ encapsulating the bacterial cell (Heidstra \textit{et al.}, 1994, Esseling \textit{et al.}, 2003). Enzymes from either bacterial origin (e.g. polygalacturonase and carboxymethyl cellulase) or plant origin (e.g. pectate lyase) proceed to degrade the plant cell wall and invagination of the plant cell wall will occur, forming a tubular structure known as an infection thread, within which the bacterial cells proliferate at the tip of the infection thread structure (Mateos \textit{et al.}, 1992, Xie \textit{et al.}, 2012). The elongation of the infection thread directs the bacteria through epidermal cells, into the cortical cells and are eventually released from the infection thread via un-walled intracellular structures termed infection droplets. Upon release from the infection thread the bacteria are enclosed in plant-derived membranes in a mechanism similar to endocytosis and develop into what are termed symbiosomes. The bacteria within these symbiosomes differentiate into the specialised bacteroid form and nitrogen fixation will then occur via expression of the enzyme nitrogenase (Brewin, 2004, Oldroyd \textit{et al.}, 2011).

\textbf{1.4.4 Nod factor}

Nod factors belong to a family of lipo-chito-oligosaccharides and are the main signalling molecules used in the communication between rhizobia and their potential hosts. Nod factors are essential in the establishment of most functional symbiotic relationships and its involvement is required throughout both the early (root hair deformation) and late phases (infection thread formation and elongation) of the symbiotic infection process (Denarie \textit{et al.}, 1992, Broughton \textit{et al.}, 2000, Perret \textit{et al.}, 2000).
1.4.4.1 Nod factor backbone biosynthesis

The basic Nod factor structure consists of four to five β-1,4-linked \( N \)-acetyl-D-glucosamine (GlcNAc) saccharides that make up the backbone and a fatty acid side chain on the non-reducing residue (basic structure shown in Figure 1.1). Nod factor backbone synthesis is catalysed by enzymes encoded by the universally conserved core \( nod \) genes, \( nodABC \). The first step of Nod factor synthesis is catalysed by NodC, an \( N \)-acetyl-glucosaminyl transferase that elongates the \( N \)-acetyl-D-glucosamine backbone, with the addition of monosaccharide subunits onto the non-reducing end of the developing Nod factor. After elongation, NodB, a de-acetylase removes any \( N \)-acetyl moieties on the terminal non-reducing GlcNAc. Lastly, NodA, an acyltransferase transfers fatty acyl chains onto the C2 of the terminal GlcNAc residue at the non-reducing end of the Nod factor (Broughton et al., 2000).

![Diagram of Nod factor backbone biosynthesis](image)

**Figure 1.1 General structure of Nod factor produced by mesorhizobia.** R1-R6 are potential modifications made to the oligosaccharide backbone. \( n \) = number of repeated \( N \)-acetyl-D-glucosamine subunits in the backbone chain.
1.4.4.2 Nod factor modifications

Whilst the core *nod* genes are involved in the synthesis of the Nod factor backbone, less conserved accessory *nod* genes encode enzymes involved in modification of the basic structure. These modifications generate unique Nod factors that can have crucial roles in the determination of host-range specificity (D'Haeze & Holsters, 2002). Examples of modifications to Nod factors produced by *Mesorhizobium* are listed in Table 1 in relation to Figure 1.1. The genes involved in these Nod factor modifications and their effects on the outcomes of nodulation will be described in more detail below.

1.4.4.3 Sulfation of Nod factor

Sulfation of the Nod factor has been observed in multiple rhizobial strains including; *S. meliloti, Sinorhizobium teranga* bv. *acaciae, S. fredii* NGR234, *Rhizobium tropici, Mesorhizobium* sp. N33, *Mesorhizobium huakuii* and *Rhizobium gallicum* (Brelles-Marino & Ané, 2008). The most well studied of these is the Nod factor of *S. meliloti* which carries a 6-*O*-sulfate group on the terminal GlcNAc at the reducing end of the Nod factor. The genes *nodH* and *nodPQ* are responsible for the transfer and synthesis, respectively, of the sulfate group (Roche *et al.*, 1991, Schwedock & Long, 1992, Schultze *et al.*, 1995). The sulfate modification is important for the host-specificity of *S. meliloti* on *M. sativa* and non-sulfated Nod factor are unable to illicit nodulation of this host (Roche *et al.*, 1991).
Table 1. Modifications of Nod factor structures produced by *Mesorhizobium* in relation to Figure 1.1

<table>
<thead>
<tr>
<th><em>Mesorhizobium</em> strain</th>
<th>n</th>
<th>Nod factor substitutions at R positions</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZP2037</td>
<td>2</td>
<td>C18:0, C18:1</td>
</tr>
<tr>
<td>NZP2213</td>
<td>0, 1, 2</td>
<td>C20:0, C20:1, C22:1, OH-C-C18:1</td>
</tr>
<tr>
<td>R7A</td>
<td>2</td>
<td>C18:0, C18:1, C22:0-OH, C16:0, C16:1, C18:0, C18:1, C20:0, C20:1</td>
</tr>
<tr>
<td>sp. N33</td>
<td>2</td>
<td>Iso-C15:0, Iso-C17:0</td>
</tr>
<tr>
<td>JRL501</td>
<td>2</td>
<td>C18:1</td>
</tr>
<tr>
<td>E1R</td>
<td>2</td>
<td>C18:1, C18:0</td>
</tr>
</tbody>
</table>

In *S. fredii* NGR234 the gene *noeE* encodes a sulfotransferase that catalyses the transfer of a sulfate group from 3’-phosphoadenosine 5’-phosphosulfate onto the 2-O-methylfucosyl group located at the C3 position of the reducing terminal GlcNAc of the Nod factor (Hanin *et al.*, 1997, Quesada-Vincens *et al.*, 1998). Mutants disrupted in *noeE* exhibit a defective nodulation phenotype on *Pachyrhizus tuberosus*, suggesting that the modification mediated by *noeE* is essential for this host (Hanin *et al.*, 1997).

**1.4.4.4 Fatty acid moiety**

The fatty acid moiety attached to the non-reducing terminal GlcNAc can vary in their chain length and in the number of (poly)unsaturated double bonds. The synthesis of (poly)unsaturated fatty acid side chains are controlled by NodFE which shares homology to an acyl-carrier protein and a β-keto-acylsynthase, respectively (Spaink & Sheeley, 1991, Spaink, 1995). After synthesis of the fatty acid moiety is complete, it is transferred to the Nod factor via NodA (Ritsema *et al.*, 1996, Brelles-Marino & Ané, 2008). In *R. leguminosarum* bv. *viciae* and *S. meliloti*, nodE mutants are unable to generate Nod factors containing an unsaturated fatty acid moiety (Demont *et al.*, 1993). The difference in the hydrophobicity of polyunsaturated fatty acyl moieties on the *R. leguminosarum* Nod factor was found to be a determinant for host-specificity on the hosts, *Vicia* and *Trifolium* (Spaink, 1995, Spaink *et al.*, 1995). Interestingly, although there is high similarity (78%) between the nodE of *R. leguminosarum* bv trifolii and *R. leguminosarum* bv *viciae*, analysis of hybrid NodE proteins found that the two do not confer the same host-range, and this was attributed to a region of 44 non-conserved amino acids (aa) that conferred the ability to nodulate *Trifolium* and *Vicia*, suggesting that allele dependent variations of nodE was a factor in host-specificity (Spaink *et al.*, 1989).
Another protein involved in the synthesis of the unsaturated fatty acid moiety is NodG. NodG is a 3-oxoacyl-acyl carrier protein reductase that is hypothesised to be a duplication of the constitutively expressed FabG. FabG is involved in the first reduction step required for the elongation of fatty acids. Based on similarities to FabG, NodG was hypothesised to be involved in the synthesis of an unsaturated fatty acid moiety (López-Lara & Geiger, 2001). Mutants disrupted in nodG do not exhibit major nodulation phenotypes due to functional redundancy between NodG and FabG. An example of this functional redundancy was observed in nodG mutants of S. meliloti that did not exhibit changes in their Nod factor acyl chain compositions, and the mutation had no major effects on the ability to nodulate alfalfa (Swanson et al., 1987, Demont et al., 1993). However, some lesser phenotypes were still observed, as seen in Rhizobium sp. N33 nodG mutants which exhibit a slight defect in nodulation represented by a reduction in the number of nodules on Astragalus cicer. Similarly, nodules formed on Onobrychis viciifolia by nodG mutants were smaller but retained their shape (Cloutier et al., 1997).

1.4.4.5 Carbamoylation

The addition of carbamoyl groups to the terminal GlcNAc at the non-reducing end can occur at positions C3, C4 and C6. NolO is responsible for the addition of a carbamoyl group at either position C3 or C4, but is unable to transfer two carbamoyl groups onto the same Nod factor concurrently (Jabbouri et al., 1998). The inability to transfer two carbamoyl groups onto the same Nod factor was hypothesised to be as an inability of NolO to recognise Nod factor that had already undergone the first carbamoyl modification (Jabbouri et al., 1998). The addition of carbamoyl via NolO is important for host specificity as shown when nolO from S. fredii NGR234 was heterologously expressed in an S. fredii USDA257 background, the host-range was extended to include the previously incompatible C. caeruleum and Lablab purpureus (Jabbouri et al., 1998).
NodU is another carbamoyl transferase responsible for the addition of a carbamoyl group to the C6 position of the non-reducing terminal GlcNAc (Jabbouri et al., 1995, D'Haeze et al., 1999, D'Haeze et al., 2000). Mutants of nodU in S. fredii NGR234 exhibit an altered nodulation phenotype that formed smaller nitrogen fixing nodules on the host L. leucocephala. In addition, the nodU mutants were delayed in nodulation on Vigna unguiculate, but total nodule numbers were unaffected (Lewin et al., 1990). Strains of S. fredii USDA257 that expressed nodU from S. fredii NGR234 gained the ability to nodulate the previously incompatible L. leucocephala, suggesting a role for this modification in host-specificity (Jabbouri et al., 1995). Lastly, in Azorhizobium caulinodans ORS571, carbamoylation at the C6 position of the terminal non-reducing GlcNAc was observed to be aid in recognition of the Nod factor by plant receptors. However, this effect could be replicated by an acetyl group substitution at the same position via NodL and it was suggested that the modification at this position was important, but not the presence of the carbamoyl group itself (D'Haeze et al., 2000).

1.4.4.6 Fucosylation

Fucosylation can occur on the terminal GlcNAc at the reducing end of the Nod factor. This modification is transferred and synthesised by NodZ and NolK/NoeK/NoeJ/NoeL, respectively. (Perret et al., 2000, Sullivan et al., 2002, Rodpothong et al., 2009). When NodZ from B. japonicum was heterologously expressed in R. leguminosarum bv. viciae the host range was expanded to include the previously incompatible L. japonicus (Bras et al., 2000). In M. loti R7A the deletion of nodZ brings about the loss of the acetylfucosyl modification and Nod factors lacking this acetylfucosyl group were unable to form
nodules on *L. filicaulis* and *L. corniculatus*, whilst a delay was observed on *L. japonicus* (Rodpothong *et al.*, 2009).

In the *M. loti* strain NZP2213, addition of a fucosyl group is located on the non-terminal GlcNAc proximal to the non-reducing end of the Nod factor. The presence of this fucosyl group does not appear to play a role in the host-range of NZP2213, and it is hypothesised that the fucosyl group may instead be involved in the protection of the Nod factor from environmental pressures that are involved in the degradation of Nod factors (Olsthoorn *et al.*, 1998).

### 1.4.4.7 Methylation

Methylation of the Nod factor can occur at both the reducing and non-reducing ends of the Nod factor. In *S. fredii* NGR234, Noel was identified as a methyl transferase involved in the transfer of a 2-O-methyl group onto the fucosyl group present on C6 of the terminal GlcNAc at the reducing end. However, it is not yet understood whether the methylation of this fucosyl group is involved in host-specificity as the substitution was not essential for the nodulation of the hosts *C. caeruleum*, *L. leucocephala*, and *Vigna unguiculata* (Jabbouri *et al.*, 1998).

On the non-reducing terminal GlcNAc, the transfer of a methyl group to the nitrogen atom linked to C2 is catalysed by the enzyme NodS. NodS is a *S*-adenosyl-l-methionine dependent methyltransferase that transfers an *N*-methyl group to the C2 position (Geelen *et al.*, 1995, Jabbouri *et al.*, 1995). Introduction of *S. fredii* NGR234 *nodS* into *S. fredii* USDA257 generated pentameric Nod factors with *N*-methylation of the acyl group. The *N*-methylated Nod factor expanded the host-range to include *Leucaena*, suggesting that
\( N \)-methylation was a determinant of host specificity (Lewin et al., 1990, Krishnan et al., 1992, Jabbouri et al., 1995).

### 1.4.4.8 Acetylation

Acetylation of the Nod factor can occur at position C6 of both terminal GlcNAc residues. NodX is an \( O \)-acetyl transferase responsible for the transfer of an acetyl group to the terminal GlcNAc at the reducing end (Firmin et al., 1993, Ovtsyna et al., 1999). Disruption of \( nodX \) in \( R. \) leguminosarum \textit{bv. viciae} TOM leads to an inability to nodulate \textit{Pisum sativum} cv Afghanistan, suggesting that the acetylation was essential for the nodulation of this host (Davis et al., 1988). Interestingly, expression of \( nodX \) is regulated in a temperature dependent manner, and increased expression of \( nodX \) generates an increased ratio of Nod factors containing the 6-\( O \)-acetyl moiety (Olsthoorn et al., 2000). Furthermore, when \( nodZ \) from \( B. \) japonicum was expressed in \( nodX \) mutants of \( R. \) leguminosarum \textit{bv. viciae}, rescue of a defective nodulation phenotype was observed reinforcing the conclusion that modifications at this position is important in the host-specificity (Ovtsyna et al., 1998).

Acetylation substitution at the C6 position of the non-reducing terminal GlcNAc is performed by the \( O \)-acetyl transferase, NodL (Bloemberg et al., 1994). Disruption of \( nodL \) in \( S. \) meliloti generated mutants with a decreased nodulation phenotype on \textit{M. sativa} and this was attributed to a decrease in infection thread formation. However, it should be noted that mutants disrupted in \( nodL \) still retained the ability to form effective nodules, suggesting that \( nodL \) was not essential for the nodulation of this host (Ardourel et al., 1994).
NoLL is responsible for the acetylation of a fucosyl group at position C6 of the reducing terminal GlcNAc (Berck et al., 1999, Corvera et al., 1999, Rodpothong et al., 2009). In R. leguminosarum bv viciae, the expression of noLL was a determinant in nodulation of L. japonicus (Bras et al., 2000). Similarly, in M. loti R7A, noLL disrupted mutants were unable to nodulate L. filicaulis and L. corniculatus but retained the ability to nodulate L. japonicus (Rodpothong et al., 2009). In the M. loti strains NZP2037 and NZP2213, the presence of NoLL was required for the nodulation of L. leucocephala and L. pedunculatus (Scott et al., 1996).

Lastly, in Neorhizobium galegae the gene noeT encodes a putative acetyl transferase that transfers an acetyl group onto position C3 of the GlcNAc proximal to the terminal non-reducing end, paralleling the fucosyl substitution that of M. loti NZP2213 (Olsthoorn et al., 1998, Österman et al., 2014). Mutants disrupted in noeT were unaffected in their host-range specificity and it was hypothesised that the presence of the acetyl modification was to provide protection against degradation (Österman et al., 2014).

1.4.5 Recognition of Nod factors and Ca²⁺ signalling

Nod factors that are produced by symbiotically compatible rhizobia are recognised by plant receptors, and upon recognition, a myriad of downstream effects occur that culminates to the organogenesis and the eventual formation of the plant root nodule organ. The plant receptors involved in the recognition of Nod factors belong to the serine/threonine kinase family and contain conserved extracellular Lysin motifs (LysM), which are involved in the recognition and binding of peptidoglycan and chitin via their characteristic N-acetyl glucosamine subunits (Buist et al., 2008).
In *L. japonicus* the receptors that were identified to bind and recognise Nod factors were designated as NFR1 (Nod factor receptor 1) and NFR5 (Nod factor receptor 5). In *L. japonicus*, plants that had defective NFR1 or NFR5 lost the ability to form symbiotic relationships with *M. loti* (Radutoiu *et al.*, 2003). Furthermore, transfer of *L. japonicus* NFR1/NFR5 into *Medicago truncatula* and *L. filicaulis* increased the list of compatible symbionts to include *M. loti* and *R. leguminosarum* bv. *viciae* DZL which were previously limited to *L. japonicus* only (Radutoiu *et al.*, 2007). Recognition of compatible Nod factor leads to several downstream responses, including among other things; rapid pH changes, de-polarisation of the membrane potential, changes in ion flux (in particular Ca\(^2+\), discussed in more detail below), root hair formation and shape, production of reactive oxygen species, induction of nodulin genes, cytoplasmic bridge formation and formation of nodule primordia (Ehrhardt *et al.*, 1996, D’Haeze & Holsters, 2002, Oldroyd *et al.*, 2011, Miller & Oldroyd, 2012).

One of the responses upon Nod factor recognition by the plant receptors is a rapid change in ion flux that occurs at the plasma membrane, with a rapid uptake of Ca\(^{2+}\) into the host cells (Ehrhardt *et al.*, 1996). This influx of Ca\(^{2+}\) is accompanied by an efflux of Cl\(^-\) and K\(^+\) ions out of the plant cell and a general alkalinisation of the cellular cytoplasm (Felle *et al.*, 1998). The influx of Ca\(^{2+}\) is hypothesised to be involved in the initiation of the infection thread structure (Miwa *et al.*, 2006, Morieri *et al.*, 2013). In the nucleus of plant cells, Ca\(^{2+}\) oscillations at regular intervals (spikes) can also be observed, and a threshold of Ca\(^{2+}\) spikes must be reached in order for downstream gene activation to occur (Miwa *et al.*, 2006). It should be noted that the influx of calcium does not always occur concurrently with the Ca\(^{2+}\) oscillations and this appears to be Nod factor dose-dependent, with higher Nod factor concentrations required to trigger the Ca\(^{2+}\) influx response.
compared to Ca^{2+} spiking (Shaw & Long, 2003, Miwa et al., 2006). Examples of downstream effects of the Ca^{2+} spiking signal include the activation of the calcium/calmodulin-dependent kinase (CCaMK), also known as DMI3 (Mitra et al., 2004). CCaMK associates with multiple downstream targets, one of which is CYCLOPS, which is involved in rhizobial infection during organogenesis of the nodule structure (Yano et al., 2008, Singh et al., 2014).

1.4.6 Surface polysaccharides and their roles in nodulation

While Nod factors and their respective receptors are the major determinants of a successful symbiotic relationship, other elements can also contribute to outcome. One such element are the surface polysaccharides produced by the rhizobial cell (shown in Figure 1.2) that are present in close proximity to the bacterial cell wall.

1.4.6.1 Exopolysaccharides

Exopolysaccharides (EPS) are polysaccharides produced in abundant amounts that are secreted by the bacteria into their surroundings. EPS differs from other polysaccharides in that it is not anchored to the cell and instead accumulates at the cell surface. EPS is readily secreted by rhizobia cultured on media containing simple carbohydrates (e.g. glucose) and the presence of full length EPS bestows upon rhizobia a mucoid colony morphology. Different species as well as differences within strains of rhizobia produce varying types of EPS, with various modifications such as different types of sugars ranging from D-glucose, D-galactose, D-mannose, D-glucuronic and D-galacturonic-acids, linkages to the subunit, the length of the repeating unit and polymerization, non-carbohydrate decorations that are responsible for acidification of EPS e.g. O-acetyl groups, ketal-linked pyruvate and succinyl half ester groups (Lepek & Dantuono, 2005,
EPS can be produced in two forms, characterised by their molecular masses, a high molecular weight (HMW) form that contains polymers of 106-107 Da and a low molecular weight (LMW) form that consist of monomers, dimers and trimers of the repeating unit (Wang et al., 1999, Janczarek, 2011).

Figure 1.2. Distribution of bacterial surface polysaccharides involved in symbiosis. EPS – Exopolysaccharide, KPS – K-polysaccharide, LPS – Lipopolysaccharide, OM – Outer membrane, CBG – cyclic β glucans, IM – Inner membrane. Figure based on data compiled from Lepek & Dantuono (2005) and Rodríguez-Navarro et al. (2007).
1.4.6.1.1 EPS of *M. loti*

The EPS produced by *M. loti* strain R7A is composed of branched octasaccharide subunits made of glucose, galactose, glucuronic acid and riburonic acid (simple diagram depicting structure is shown in Figure 1.3, for a more detailed analysis see Muszyński *et al.* (2018). The structure of the basic subunit contains a main backbone that consists of glucose and galactose while the branch chain contains glucose, glucuronic acid and riburonic acid. The EPS produced by NZP2037 has been partially characterised by Hotter & Scott (1991) and largely correlates with the R7A EPS structure, sharing the same constituents and the lack of succinyl and pyruvyl groups (Kelly, 2012).

**Bio**

**Figure 1.3. R7A EPS biosynthesis.** Enzymes predicted to be involved in the biosynthesis of EPS produced by R7A. Enzymes involved in synthesis of the backbone, branch chain and the riburonic acid are indicated in blue, red and green, respectively. Gal = galactose, Glc = Glucose, GlcA, glucuronic acid, RibA = riburonic acid. Figure adapted from Wightman (2015).
1.4.6.1.2 EPS as a signalling molecule

The role of EPS in the symbiosis has been proposed to include passive roles ranging from protection of the rhizobial cell from reactive oxygen species, aiding in attachment of bacterial cells to host surfaces, nutrient gathering and cellulose-mediated agglutination (Fraysse et al., 2003, Laus et al., 2005, Janczarek, 2011, Lehman & Long, 2013, López-Baena et al., 2016). In addition, EPS is also proposed to play a role in actively escaping plant defence responses (Niehaus et al., 1993, D'Haeze & Holsters, 2004). Most importantly, EPS can act as a signalling molecule in symbiosis.

In *S. meliloti*, disruption of the genes exoA, exoH, exoB, and exoF generates mutants that are unable to form nitrogen fixing nodules on *M. sativa* and effective nodulation could be restored with the addition of purified succinoglycan from the parent strain (Battisti et al., 1992, Urzainqui & Walker, 1992). A similar observation was seen in *S. fredii* NGR234 exo mutants where effective nodulation was restored with the addition of parent EPS or the oligosaccharide repeating unit (Djordjevic et al., 1987). Interestingly, EPS that was obtained from non-parent strains or structurally altered homologous EPS could not restore nodulation to exo mutants, providing support for EPS as an active signalling molecule, instead of only a passive role (Djordjevic et al., 1987, Battisti et al., 1992, Urzainqui & Walker, 1992).

The most compelling evidence for EPS as a signalling molecule is seen in *M. loti* strain R7A. R7A exoB mutants, which generate no EPS, retains the ability to form nitrogen fixing nodules on *L. japonicus*. However, in R7A mutants that produced a truncated form of EPS, displayed a defective nodulation phenotype that suggested that EPS was important in the signalling of a symbiotic relationship (Kelly et al., 2013). Similarly, in
M. loti strain NZP2037, mutants deficient in acidic EPS production that were generated using a transposon mutagenesis exhibited an inability to form nitrogen fixing nodules on L. leucocephala (Hotter & Scott, 1991). Subsequently, a receptor for EPS was identified in the host L. japonicus Gifu and designated as Epr3. Epr3 was shown to bind and recognise both full length and truncated EPS, and responds as either a promoter or negative regulator of nodulation, respectively (Kawaharada et al., 2015). The recognition of incompatible EPS results in a decreased number of infection threads as well as abortion of the infection thread. Plants that do not express Epr3 exhibited less infection thread formation in the presence of full length EPS, suggesting that activation of Epr3 with full length EPS function as a positive regulator of nodulation (Kawaharada et al., 2015).

1.4.6.2 Other polysaccharides involved in symbiosis

Other surface polysaccharides are known to have roles in symbiosis include K-antigens, or K-polysaccharides (KPS), Lipopolysaccharides (LPS) and Cyclic β-glucans (CBGs). However, these were not investigated in detail in this study, and so are not reviewed in detail here. For more detailed reviews please see Kannenberg et al. (1998), D’Haeze & Holsters (2004), Becker et al. (2005), and Janczarek et al. (2015).

KPS, also commonly referred to as capsular polysaccharides, are closely associated with the rhizobial outer membrane, this is in contrast to EPS which is loosely associated with the bacterial cell (Kannenberg et al., 1998). KPS functions as a signal molecule during for symbiosis. In the symbiosis between S. meliloti and M. sativa, exoB mutants defective in EPS production were rescued with the expression of a KPS production gene, lpsZ, suggesting that KPS could compensate for EPS signalling in this strain (Reuhs et al., 1995). Further support for KPS signalling is observed in S. meliloti KPS mutants that
exhibit an abortive infection thread phenotype on *M. sativa* (Putnoky *et al.*, 1988). In *S. fredii*, mutations in genes involved in synthesis of KPS, *rkpG* and *rkpH* (which encode a 2-aminod-3-ketobutirate-CoA-ligase and a ribitol type dehydrogenase, respectively) lead to a decreased nodulation phenotype, in contrast, an *exoA* mutant (unable to produce EPS) was not significantly affected in nodulation, suggesting that KPS was the dominant signalling polysaccharide (Parada *et al.*, 2006). Lastly, in *S. fredii* NGR234 mutants that lack the *rkpMNO* gene cluster, which encodes enzymes that synthesise the KPS precursor, pseudaminic acid, exhibit a decreased nodulation phenotype (Le Quere *et al.*, 2006).

LPS are polysaccharides that are chemically anchored to the outer membrane of the bacterial cell, this is in contrast to EPS which is only loosely associated to the membrane. LPS can participate in the establishment of symbiosis, as seen between *R. leguminosarum* bv. *trifoli* 0403 and the host white clover. In this symbiotic relationship, small amounts of LPS (5 mg per plant) was sufficient to promote the binding of cells to the root hair tips assisting cell passage across root cell walls, encouraging the formation of infection threads. Mutants that produce LPS lacking in the O-antigen moiety fail to infect host cells and are unable to fix nitrogen, suggesting that it is functionally important (Priefer, 1989, Perotto *et al.*, 1994, Noel & Duelli, 2000, Lerouge & Vanderleyden, 2002). Modulation of plant defences by LPS may also be involved in the later stages of infection. Application of *S. meliloti* LPS can suppress the oxidative burst and alkalinisation of *M. sativa* cell suspension cultures that had been induced to generate hydrogen peroxide via the addition of small amounts of yeast elicitors (Albus *et al.*, 2001).

CBGs are sugar oligomer polymers secreted by rhizobia into the periplasmic space within the bacterial cell wall. CBGs are involved in the establishment of symbiosis, as
demonstrated by ndvA/ndvB mutants of S. meliloti that no longer formed effective nitrogen-fixing nodules on M. sativa. The genes ndvA/ndvB encodes an ABC-type inner membrane transporter for CBGs and a glycosyltransferase required for synthesis of CBGs, respectively (Dylan et al., 1986). However, around 3-5% of M. sativa inoculated with S. meliloti ndv mutants still formed nitrogen fixing nodules. Bacteria isolated from these nodules were termed ‘pseudorevertants’ as they had regained the ability to nodulate M. sativa but had not regained the ability to produce CBGs (Dylan et al., 1990b). In M. loti, CBG mutants form white, empty, nodule primordia on L. tenuis (D'Antuono et al., 2005). Similarly, mutants disrupted in expression of a putative cell envelope protein that resulted in reduced CBG production were unable to form effective nitrogen-fixing nodules on L. japonicus (Kawaharada et al., 2007). Other proposed functions of CBGs include: i) Increased solubility of hydrophobic molecules such as flavonoids, ii) Protection from environmental pressures such as osmotic protection, and iii) Modulation of plant defence responses (Dylan et al., 1990a, Mithöfer et al., 2001, Fraysse et al., 2003).

1.4.7 Secretion systems

Over the countless millennia, bacteria have evolved multiple systems to help them survive and better take advantage of their environments. One such system is the ability to transfer compounds and/or proteins across the bacterial outer and inner membranes and into their extracellular surroundings or directly into the cytosols of their target organisms.

Gram-negative bacteria achieve the transfer of proteins across their membranes via the general secretion pathway or the twin arginine translocation (TAT) secretion pathway, in addition to multiple specialized secretion systems, of which six have been studied in more detail (Tseng et al., 2009). This introduction will primarily focus on the Type I and Type
IV secretion systems. However, other secretion systems such as the Type III secretion system and the general secretion pathway will also be briefly covered.

1.4.7.1 The general secretion pathway and the TAT secretion system

The general secretion system (also known as the Sec system) is a ubiquitous system present amongst all domains of life (Papanikou et al., 2007). The basic molecular machinery consists of a membrane-bound translocase and an ATPase that hydrolyses ATP to provide the energy for translocation of these proteins across the membrane in an unfolded state (Papanikou et al., 2007, Tseng et al., 2009). An example of a symbiotically important protein exported by the Sec pathway is the cellulase, CelC2. CelC2 is required for _R. leguminosarum_ bv. _trifolii_ to form nodules on white clover where it assists entry into the plant via the erosion of the root hair-cell wall (Robledo et al., 2008, Downie, 2010).

The Twin-Arginine Translocation (TAT) pathway is a general secretion system involved in the secretion of proteins in their folded forms. This is an important distinction that separates it from the Sec system, as post-translational modifications may only be available within the secreting cell (Berks et al., 2005). Proteins transported by the TAT pathway usually contain a motif of consecutive twin arginine residues (S-R-R-x-F-L-K) recognized by the cellular machinery for secretion (Tseng et al., 2009). The TAT secretion pathway is found across most bacterial species and the _Rhizobiaceae_ are no exception. Thus far, TAT secretion systems have been identified in _S. meliloti_ (Galibert et al., 2001), _Agrobacterium tumefaciens_ (Ding & Christie, 2003), _Rhizobium etli_ (González et al., 2006), _M. loti_ (Kaneko et al., 2000) and _R. leguminosarum_ bv _viciae_ (Meloni et al., 2003). In _R. leguminosarum_, _tacC_ mutants have defective TAT secretion systems that exhibit an
inability to form nodules on peas, suggesting a role for the TAT secretion systems in symbiosis (Meloni et al., 2003, Krehenbrink & Downie, 2008).

1.4.7.2 The Type I secretion system

The Type I secretion system (T1SS) is involved in the transport of substrates across the bacterial outer and inner membranes into the surrounding extracellular environment and this transport occurs in a one-step process without the formation of stable periplasmic intermediates (Delepelaire, 2004). T1SS are relatively widespread amongst gram-negative bacteria, and effectors secreted by these systems can vary greatly in size, ranging from 78 to 8682 aa residues (Delepelaire, 2004).

Effectors of T1SS commonly contain an RTX (Repeat in ToXin) domain that is made up of repeated nonapeptide glycine rich repeat motif (GGXGXDXUX, U represents a large hydrophobic residue) that binds Ca\(^{2+}\) ions to form β-roll secondary structures (Welch, 2001). The C-terminus for effectors of the T1SS usually contains their secretion signals. However, determinants of the secretion signal are still poorly understood due to lack of conserved primary sequence between known effectors. It is hypothesised that the secretion signal is recognised based on the formation of secondary structures such as amphipathic helices, and helix loop helix motifs (Welch, 2001, Thomas et al., 2014). Due to the lack of an obvious universal conserved secretion signal, the presence of an RTX domain has been used to infer potential effectors that are likely to be secreted by T1SS (Welch, 2001, Delepelaire, 2004, Thomas et al., 2014). In general, T1SS closely resemble the large protein family of ATP-binding cassette (ABC) transporters that are involved in the export of small molecules such as antibiotics (Green & Mecsas, 2016).
The disruptions of the T1SS can result in a defective nodulation phenotype. In *R. leguminosarum* bv. *viciae* mutants disrupted in *prsD* exhibited the inability to form effective nitrogen fixing nodules on *Pisum sativum*, forming greenish-white nodules instead of the healthy pink nodules (Finnie *et al.*, 1997). Similarly, in the related strain *R. leguminosarum* bv. *trifolii* TA1, mutants disrupted in *prsD* exhibit a defective nitrogen fixation phenotype on *T. pratense*. The mutants still formed fully infected nodules, but the nodules lacked nitrogen fixation activity (Mazur *et al.*, 1998).

### 1.4.7.2.1 Structure of the Type I secretion system

The molecular machinery that comprises a T1SS consists of three main components: an ABC transporter, a membrane fusion protein (MFP), and an outer membrane factor (OMF). Bacterial genomes may encode multiple T1SS and each may only transfer a certain subset of effectors secreted by the bacteria (Ma *et al.*, 2003, Delepelaire, 2004). The most well understood T1SS is the HlyA (haemolysin A) secretion system first described in uro-pathogenic *E. coli* (schematic shown in Figure 1.4) and will be used as the model T1SS (Delepelaire, 2004, Tseng *et al.*, 2009, Thomas *et al.*, 2014).
Figure 1.4. The HlyA Type I secretion system of *E. coli*. The HlyA system consists of three protein components, TolC the outer membrane factor is anchored in the outer membrane (OM), HlyD is the membrane fusion protein spanning the periplasmic space. HylB is the ABC transporter and forms a dimer anchored to the inner membrane (IM). HlyA is the effector protein transported across the membranes into the extracellular environment. Figure compiled from data in Delepelaire (2004), Dalbey & Kuhn (2012), and Thomas et al. (2014).

The first component of the T1SS is the ABC transporter which belongs to a family of proteins involved in the binding and/or hydrolysis of ATP, the action of which provides the molecular energy required for transfer of substrates across the bacterial membranes (Kerr, 2002). ABC transporters are minimally composed of four domains, two transmembrane domains and two nucleotide binding domains (Kerr, 2002). In the HlyA secretion system, the ABC transporter is HlyB, and two monomers of HlyB assemble in the inner membrane to form a functional dimer (Thomas *et al.*, 2014). The cytosolic domain of HlyB comes into direct contact with effectors and recognises their secretion signal and this interaction helps determine substrate specificity (Zhang *et al.*, 1993).
The second component of a T1SS is the MFP that acts as the periplasmic adaptor linking the ABC transporter to the OMF. The general structure of an MFP consists of a short cytoplasmic domain at the N-terminus followed by a membrane anchor and a large periplasmic domain (Delepelaire, 2004). In the HlyA system, the MFP is HlyD. The cytoplasmic domain of HlyD has an active role in the recognition of substrates for transfer, and truncation of HlyD at this cytosolic domain results in disruption of an amphipathic helix secondary structure that is required for substrate export (Balakrishnan et al., 2001). HlyD is also involved in recruitment of the OMF and the deletion of a cluster of charged residues located within the N-terminus of HlyD impaired recruitment of the OMF, and as a consequence, the translocational channel spanning the membranes could not be formed, resulting in ineffective substrate transfer (Balakrishnan et al., 2001).

Upon recognition of a substrate candidate that is ready for transfer, the MFP/ABC complex recruits the OMF. Following recruitment of OMF, the protein complex undergoes interactions and assembles into a translocation channel allowing for passage of substrate across the two membranes of the secreting cell (Delepelaire, 2004, Koronakis et al., 2004). In the HlyA system, the OMF is the multi-purpose protein TolC which is a component in multiple outer-membrane processes ranging from drug transport and toxin secretion as observed in the AcrAB drug efflux pump system and the HlyA and Colicin V toxins, respectively (Husain et al., 2004, Koronakis et al., 2004).

TolC forms homotrimers that are arranged to form long water channels which run throughout the outer membrane extending into the periplasm. The protein structure is composed mainly of short β-barrels at the outer membrane and α-helical structures at the
periplasm. Opening of the water channel involves an iris-like movement that rearranges the α-helices, widening the previously non-accessible periplasmic entrance and allowing for the movement of substrates across the channel and into the extracellular space (Koronakis et al., 2000, Thomas et al., 2014).

1.4.7.2.2 Genetic structure and regulation of the Type I secretion system

Genes that encode effectors and secretion machinery of T1SS can be found on both the chromosome or within mobile genetic elements (Thomas et al., 2014). These genes are usually genetically orientated within a single operon such as the hly operon (Figure 1.5A). The hly operon encodes the secreted effector hlyA directly followed by hlyB and hlyD which encode the exporter machinery. The gene hlyC encodes an acyltransferase that is not a required for substrate transfer but is instead required for the haemolysin activity of HlyA (Stanley et al., 1998). It should be noted that TolC expression is not genetically linked to the hly operon and this is likely due to the multiple roles that TolC fulfils in other secretory systems (Thomas et al., 2014).

Genes that encode T1SS substrates are not always genetically orientated in close proximity to their respective secretion systems. An example can be seen in the Serratia marcescens Lip system (Figure 1.5B). In this system lipBCD encodes the exporters and the OMF machinery of the T1SS required for the export of the effector SlaA (an S-layer protein). SlaA is encoded by slaA which is genetically located upstream of lipBCD. However, the Lip system also transports two other substrates, the Lipase, LipA and the protease, PrtA encoded by the genes lipA and prtA, respectively. The genes lipA and prtA are not located in close proximity to the slaA operon (Kawai et al., 1998).
Figure 1.5. Examples of genetic orientations of Type I secretion systems. The *E. coli* hly operon. The JUMPstart domain and the ops domain that are required for the transcriptional regulator RfaH are shown in grey and black, respectively. (B) The *S. marcescens* slaA operon. The operon contains genes that encode the effector (slaA) and the secretion system (lipBCD). The genes lipA and prtA that encode known effectors transported by the Lip secretion system are not genetically linked (dashed line). Figure adapted from Thomas et al. (2014), permission is granted under the open archive Elsevier user license (https://www.elsevier.com/about/our-business/policies/open-access-licenses/elsevier-user-license).

Transcriptional regulation of the Hly operon is controlled by RfaH. RfaH was initially identified as a positive regulator of the *rfa* operon and proposed to be involved in the upregulation of glycosyl transferase genes required for LPS synthesis (Lindberg & Hellerqvist, 1980). Subsequently, RfaH was found to be required for the transcription of the *hly* operon (Bailey et al., 1992). Specifically, RfaH is involved in the transcription of *hlyB* and *hlyD*, located at the distal end of the *hly* operon. RfaH acts as a transcriptional anti-terminator that elongates mRNA transcripts. In the context of the *hly* operon, RfaH extends the mRNA transcript to cover the *hlyBD* region, thus *hlyBD* is uncoupled from
hylAC expression in the absence of RfaH (Leeds & Welch, 1996, Thomas et al., 2014).

RfaH recognizes a 39-bp sequence present in the upstream non-coding region of genes that it regulates termed the JUMPstart (just upstream of many polysaccharide-associated starts) motif. In the hly operon, disruption of the JUMPstart motif abolished RfaH dependent hlyC transcription (Leeds & Welch, 1997). In addition to the JUMPstart motif, RfaH also requires a second smaller 8-bp motif (GGCGGTAG) termed ops (operon polarity suppressor) that is located just downstream of the JUMPstart motif (Bailey et al., 1996, Bailey et al., 1997). It is proposed that RfaH recognizes and binds to the ops motif exposed on the non-coding strand when the RNA polymerase is paused at the JUMPstart sequence. Upon binding, interactions between RfaH and the RNA polymerase stabilises the polymerase complex, and this alleviates stalling of the RNA polymerase complex allowing elongation of the mRNA transcript to proceed (Artsimovitch & Landick, 2002).

1.4.7.2.3 Rhizobial effectors exported by Type I secretion systems

T1SS have been identified and predicted to be encoded within the genomes of multiple rhizobial species and in some cases, multiple T1SS may be present within a single strain. For example, in R. leguminosarum bv. viciae 3841 four T1SS: PrsDE, BltDE, ToaDE, TobDE are predicted to encode the exporter machinery (the ABC-transporter and MFP) while no exclusive OMFs have yet to be identified (Krehenbrink & Downie, 2008). In comparison, only two T1SS, PrsDE and ExpD1/D2, have been predicted to be present in S. meliloti (York & Walker, 1998, Moreira et al., 2000).

The first rhizobial T1SS effector involved in nodulation that was identified was NodO. NodO was first described in R. leguminosarum bv. viciae where it was found to be
secreted into the media of cell cultures after exposure to the flavonoid naringenin, suggesting that NodO expression was NodD dependent (de Maagd et al., 1988). Further characterisation of NodO found that the protein shared many similarities to haemolysins (e.g. HlyA) including the ability to bind Ca\(^{2+}\) ions (Economou et al., 1990). Subsequently NodO was analysed to insert into liposomes, and formed cation selective channels that allowed for movement of the cations K\(^+\) and Ca\(^{2+}\) across lipid membranes (Sutton et al., 1994). The T1SS responsible for the transfer of NodO across the rhizobial membrane was later identified as the PrsDE transport system, the genes for which are genetically unlinked to nodO (Finnie et al., 1997).

NodO is implicated to be involved in the rhizobium-legume symbiosis. In a *R. leguminosarum* bv. *trifolii* cured of its symbiosis plasmid, the presence of a construct that harboured a region of the *R. leguminosarum* bv. *viciae* symbiosis plasmid containing the nod region allowed for nodulation of *Vicia sativa*, and the disruption of nodO severely decreased this nodulation (de Maagd et al., 1989). Similarly, in a nodE mutant of *R. leguminosarum* bv. *trifolii*, the heterologous expression of NodO from *R. leguminosarum* bv. *viciae* allowed for the nodulation of the previously incompatible host *Vicia hirsute*, while expression in the wildtype had no effect (Economou et al., 1994). In *R. leguminosarum* bv. *viciae*, disruption of nodO by itself had no major effects, however a nodE nodO double mutant exhibited formation of abnormal infection threads and expression of nodO alleviated this (Walker & Downie, 2000).

The heterologous expression of NodO from *Rhizobium* sp. BR816 in *R. etli* CE3 broadened the host range to include the previously incompatible *L. leucocephala* (van Rhijn et al., 1996). NodO from BR816 was also able to suppress nodulation defects in
mutants affected in Nod factor production, specifically, when NodO was heterologously expressed in nodU mutants in the S. fredii NGR234 and R. tropici CIAT899 backgrounds, restoration of nodule numbers to wildtype levels was observed on L. leucocephala (Vlassak et al., 1998). NodO also partially supressed the delayed nodulation phenotype of a nodE mutant in the R. leguminosarum bv trifolii background on white clover (Vlassak et al., 1998).

Due to these observations, it was proposed that NodO functions as a complement to Nod factor via transport of the Nod factor directly, or is involved in the promotion of Ca$^{2+}$ movement across the plant cell membrane (Sutton et al., 1994). The second function is more likely, given that NodO associates with Ca$^{2+}$. Movement of Ca$^{2+}$ generates an ion flux that is observed early on during Nod factor signalling (previously covered in section 1.4.5); thus NodO may amplify the ion flux that occurs during establishment of symbiosis, and in the case of NodO, this amplification effect is likely to occur during infection thread development (Sutton et al., 1994, Walker & Downie, 2000, Miwa et al., 2006).

Other rhizobial proteins excreted by T1SS include a group of glycanases involved in the cleavage and processing of EPS. In R leguminosarum bv. viciae the glycanases PlyA and PlyB were found to be secreted by the PrsDE secretion system (Finnie et al., 1997). Mutants of plyA and plyB were severely affected in their ability to form biofilms, likely due to the increased number of long chain EPS molecules (Finnie et al., 1998, Russo et al., 2006). However, the plyA plyB double mutant was not affected in its ability to nodulate Vicia hirsute and so the involvement of these glycanases as determinants of host specificity in symbiosis (if any) has yet to be determined (Finnie et al., 1998).

In S. meliloti a glycanase identified to be exported by the PrsDE system is ExsH, which
is involved in the cleavage of nascent EPS, generating LMW fractions of EPS. However, mutagenesis studies of *prsDE* and *exsH* indicated that the genes were not essential for the nodulation of *M. sativa* and the authors propose that this may be due to parallel pathways that contribute to EPS synthesis (York & Walker, 1997, York & Walker, 1998). Lastly the protein ExpE1 which is secreted by the ExpD1/2 T1SS (a homologue of the PrtD/E system), acts as a glycanase required for the synthesis and/or secretion of galactoglucan which can act as a substitute for succinoglycan signalling on *M. sativa* (Glazebrook & Walker, 1989, Becker et al., 1997, Moreira et al., 2000).

Proteins secreted by the T1SS involved in adhesion to the plant root include the rhizobia adhering proteins (RAPs) produced by *R. leguminosarum* bv. *trifolii* that are secreted by the PrsDE T1SS (Russo et al., 2006). RapA1 is involved in the adhesion and colonisation of root hairs, however, RapA1 was not a requirement for efficient nodulation of *Trifolium pratense* (Mongiardini et al., 2008).

Lastly, some effectors secreted by the T1SS were predicted based on the presence of RTX motifs present in effectors that commonly associate with T1SS. Of the effectors identified using this predictive approach, known secreted substrates such as NodO, PlyA/B were identified, alongside 7 new effectors. However, the effects of these new effectors on symbiosis have yet to be fully investigated and characterised (Krehenbrink & Downie, 2008).

**1.4.7.3 The Type IV secretion system**

Unlike the T1SS, which secretes effector proteins into the extracellular environment, the Type IV secretion system (T4SS) directly transports its effectors into the cytoplasm of its
targets, through the membranes of both the secreting cell and the target cell. T4SS are the most versatile of the gram-negative bacterial secretion systems with substrates ranging from DNA to macromolecular substrates, and target cells including: bacterial, fungal, plant and mammalian cells (Cascales & Christie, 2003, Alvarez-Martinez & Christie, 2009).

The T4SS can be sub-divided into three families based on their functions. However, it should be noted that a T4SS may belong to multiple families.

The first family are involved in the conjugation process and transport DNA via direct cell to cell contact. The DNA transported via these conjugational T4SS include conjugative plasmids and mobile genetic elements responsible for genome plasticity within bacterial communities. (Cascales & Christie, 2003).

The second family are responsible for the export of effector proteins into target cells. Examples of systems that belong to this family can be seen in bacterial pathogens such as Legionella pneumophila, Brucella suis, and Helibacter pylori that utilise these systems to export virulence factors. These virulence factors aid in successful infection usually through the modulation of host cell processes (Backert & Meyer, 2006).

The third family has the ability to import DNA or protein substrates from the extracellular environment without the need to make contact with other cells, an examples of a T4SS that belongs to this family is the ComB system of Helicobacter pylori (Alvarez-Martinez & Christie, 2009).

The most well understood T4SS is the VirB/D4 system present in Agrobacterium
*tumefaciens* and consists of 12 proteins, VirB1-11 and VirD4. Multiple gram-negative bacterial T4SS share homology to the *A. tumefaciens* VirB/D4 system, including rhizobial species such as *M. loti* (Hubber *et al.*, 2004, Bhattty *et al.*, 2013). The known substrates transferred via the VirB/D4 system of *A. tumefaciens* include both DNA and proteins that are usually involved in pathogenicity and the ability to cause crown gall disease in plants (Gelvin, 2012).

### 1.4.7.3.1 Structure of the VirB/D4 Type IV secretion system

The general structure of the Type IV molecular machinery (shown in Figure 1.6) can be separated into 4 subassemblies: i) The Type IV coupling protein (T4CP) which acts as an ATPase and is also involved in the recognition and processing of potential substrates, ii) an inner membrane complex (IMC) involved in transfer of substrates across the inner membrane, iii) the outer membrane complex (OMC) or core complex that transports substrates across the periplasm and the outer membrane, and iv) the pilus structure that adheres to target cells (Christie *et al.*, 2014).

### 1.4.7.3.2 Rhizobial effectors of the Type IV secretion system and their effects on symbiosis

Here I will briefly cover known T4SS effectors involved in the establishment of a symbiotic relationship between the *Rhizobiaceae* and their respective hosts. In *M. loti* strain R7A, the T4SS is involved in the transfer of the effectors Msi059 and Msi061 (Hubber *et al.*, 2004). R7A mutants disrupted in the machinery of the VirB/D4 machinery, or the effector encoding genes (*msi059* and *msi061*) exhibit a delayed nodulation phenotype on *L. corniculatus*. In contrast, the same mutants were able to nodulate *L. leucocephala* a host that is incompatible with the wildtype R7A (Hubber *et al.*, 2004).
Figure 1.6. The VirB/D4 Type IV secretion system. Above: The four sub-assemblies of the VirB/D4 system: The Type IV coupling protein (T4CP), Inner membrane complex (IMC), outer membrane complex (OMC)/Core and the pilus structure together form the translocation channel. The layers of the secreting cell are designated as Outer membrane (OM), Peptidoglycan (P) and Inner membrane (IM). Below: Genetic organization of the VirB/D4 genes in A. tumefaciens and the proteins that they encode. Figure adapted from Christie et al. (2014), permission is granted under the open access Elsevier user license (https://www.elsevier.com/about/our-business/policies/open-access-licenses/elsevier-user-license).
Msi059 is composed of 1798 aa and has a complex repeat structure over the first 1300 aa. Similar proteins that share this repeat structure belong to hypothetical proteins in *B. japonicum*. The residues within 1482-1665 contains highly conserved residues that make up the catalytic domain belonging to the C48 cysteine protease family, which includes XopD of *Xanthomonas campestris*. XopD is predicted to target proteins for degradation via a SUMO (small ubiquitin-like modifier) mechanism (Hubber *et al.*, 2004).

Msi061 is composed of 287 aa and shares sequence similarity to VirF related proteins found in *A. tumefaciens*, *Agrobacterium rhizogenes* and *Allorhizobium vitis*. VirF contains an F-box domain that is involved in the ubiquitination of proteins, the action of which targets these proteins for degradation. The F-box domain requires the presence of highly conserved Leucine and Proline residues which are conserved within Msi061 and its homologues, suggesting that Msi061 acts in a similar fashion (Hubber *et al.*, 2004).

In *S. meliloti* strain KH46c, a mutant disrupted in the genes *virB6-9* exhibited a phenotype where the plants had a decreased nodule number and an increase in nodule masse on the hosts *M. truncatula* cv. A17 and *Medicago tricine* (Sugawara *et al.*, 2013). This same mutant also exhibited a less effective fixation phenotype on *M. sativa* cv. Agate (Sugawara *et al.*, 2013). The effector was recently identified as TfeA (Nelson *et al.*, 2017). TfeA shares 59% aa homology to Msi061 of R7A and was predicted to function in targeted proteolysis via ubiquitination of targets, and subsequent degradation of these targets may alter the host environment to favour a symbiotic relationship (Nelson *et al.*, 2017). In *S. medicae* a mutant disrupted in *virB6-9* also displayed a decreased nodulation phenotype on *M. truncatula* cv A17 and an increase in nodule mass an plant biomass was observed on the closely related *M. truncatula* cv F83005-5 (Sugawara *et al.*, 2013).
1.4.7.4 The Type III secretion system

The secretion systems that have been covered thus far are the T1SS and the T4SS. Here, the focus will shift to the Type III secretion system (T3SS) and some of the effectors that are transported by this system that are known to involved in the establishment of symbiosis will be briefly covered within this section.

The T3SS parallels the T4SS, in that both secretion systems encode molecular machinery involved in the transfer of effectors directly across membranes of the secreting and target cells. Rhizobia that are known to encode T3SS include *S. fredii*, *B. japonicum*, *R. etli*, *B. elkanii*, *Cupriavidus taiwanensis* and *M. loti*. Homologues of effectors secreted by these systems may be found amongst distantly related strains, although some effectors can be unique to certain strains (Deakin & Broughton, 2009, Nelson & Sadowsky, 2015). Here, I will cover some of the rhizobial encoded effectors secreted by T3SS in the strains *S. fredii* NGR234 and *M. loti* MAFF303099 and the effects they have on the outcome of an effective symbiosis.

1.4.7.4.1 Effectors of *S. fredii* NGR234

One of the most extensively studied T3SS that are involved in the establishment of symbiosis is the one present within *S. fredii* NGR234. The known effectors secreted by the T3SS include NopP, NopL, NopJ, NopT, and NopM. Other proteins often associated with these systems such as NopA, NopB and NopX have been shown to be essential for Type III mediated secretion, but NopABX are involved in the formation of the pilus apparatus and/or translocon, and thus are unlikely to function as functional effectors that are secreted into target host cells (Büttner & Bonas, 2002, Saad *et al.*, 2008).
NopP and NopL are Type III effectors that upon entry into the cytoplasm of the host, undergoes phosphorylation by host-cell kinases. It is hypothesised that phosphorylation activates NopP and NopL to either alter the plants host defences or to aid in the entry of rhizobial cell into the plant via alterations to the host cytoskeleton (Skorpil et al., 2005). NopP and NopL are somewhat functionally redundant and the two can act as determinants of host-specificity where they promote the symbiosis of the hosts *Tephrosia vogelii* and *Flemingia congesta*. In contrast, NopP and NopL impede symbiosis on *Pachyrhizus tuberosus* and *Crotalaria juncea* (Marie et al., 2001, Skorpil et al., 2005).

NopJ bears homology to proteins that belong to the YopJ family of proteins that are secreted by the T3SS of *Yersinia pestis*. Based on the similarities between NopJ and YopJ, it was hypothesised that NopJ was involved in the cleavage of proteins or in the inactivation of kinases required for the activation of plant defence responses (Orth, 2002, Mukherjee et al., 2006, Kambara et al., 2009). Mutants of *nopJ* exhibit a phenotype of an increase in the number of average nodules on *Lablab purpureus* suggesting that NopJ has a negative effect on the symbiosis of this host (Kambara et al., 2009).

NopM shares homology to proteins that belong to the IpaH–SspH–YopM family of effectors secreted by the T3SS of known animal pathogens (Deakin & Broughton, 2009). Although the exact mechanism of how NopM functions has yet to be ascertained, it is hypothesised that NopM may be involved in either modification of the host cell gene expression, modification of host kinase activity, or ubiquitin ligase activity based on the presence of a cysteine residue that is required for ligase activity (Deakin & Broughton, 2009, Kambara et al., 2009). Mutants that were disrupted in *nopM* exhibited an increase in the number of nodules on the host *Pachyrhizus tuberosus* suggesting that NopM has a
negative effect on this host. In contrast, \textit{nopM} mutants exhibited a decrease in the number of nodules on \textit{Lablab purpureus}, suggesting that NopM promotes nodulation on this host (Kambara \textit{et al.}, 2009).

NopT shares homology to a family of cysteine proteases, defined by YopT of \textit{Yersinia} sp. and AvrPphB of \textit{Psuedomonas syringae pv. phaseolicola}. NopT is hypothesised to act in a similar manner to YopT which recognizes and cleaves Rho family GTPases, the action of which releases them from the plasma membrane and also deactivates the GTPases (Kambara \textit{et al.}, 2009). Mutants disrupted in \textit{nopT} exhibit an increased number of nodules on \textit{C. juncea} and a decreased number of nodules on \textit{T. vogelii} suggesting its involvement as a repressor and promoter of nodulation, respectively (Kambara \textit{et al.}, 2009).

\textbf{1.4.7.4.2 Effectors of \textit{M. loti} MAFF303099}

In \textit{Mesorhizobium loti} MAFF303099 the T3SS shares many similarities to the T4SS of \textit{M. loti} R7A and exports the effectors Mlr6316, Mlr6361, Mlr6331 and Mlr6358 (Hubber \textit{et al.}, 2004, Sánchez \textit{et al.}, 2009). Mlr6316 is a member of the C48 cysteine protease family and is predicted to be a functional homologue of \textit{M. loti} R7A Msi059. Mutation of \textit{mlr6316} resulted in the formation of a small number of nodules on a small proportion of \textit{L. leucocephala}, in contrast, the wildtype only formed small white tumour-like structures (Hubber \textit{et al.}, 2004). These results suggest that Mlr6316 had a positive effect on the nodulation of \textit{L. leucocephala}. However, a mutant disrupted in the Type III secretion machinery in MAFF303099; effectively disrupting the transfer of all effectors, exhibited an effective nodulation phenotype on \textit{L. leucocephala} suggesting the involvement of a negatively acting effector yet to be identified that was likely to be transported by the T3SS (Hubber \textit{et al.}, 2004, Sánchez \textit{et al.}, 2009).
The effector Mlr6361 contains a conserved multi-domain (PRK09169) containing an internal domain repeated 15 times and a Shikimate kinase-like domain at its carboxyl terminus (Okazaki et al., 2010). The PRK09169 multi-domain has not yet been properly characterized but is hypothesised to be involved in protein-protein interactions based on the leucine-rich repeats within the domain common among other Type III effectors. The Shikimate-kinase domain is hypothesised to be involved in the synthesis of aromatic aa, and may be involved in the alteration of the environment to favour the growth in an infection environment (Okazaki et al., 2010). Mutants disrupted in mlr6361 exhibited an increased nodulation phenotype on Lotus halophilus suggesting that Mlr6361 has a negative effect on the nodulation of this host (Okazaki et al., 2010). Similarly, Mlr6361 also has a negative effect on the nodulation of L. japonicus Gifu MG20 (Sánchez et al., 2012).

The effector Mlr6331 shares many similarities to Mlr6361, both are predicted to contain a conserved multi-domain PRK09169 and a Shikimate kinase-like domain at its C-terminus, thus is hypothesised to function in a similar fashion to Mlr6361 (Okazaki et al., 2010). Interestingly, despite sharing many similarities to Mlr6361, Mlr6331 has a positive effect on the outcome of nodulation of L. japonicus Gifu MG20, contrasting that of Mlr6361 and suggests that the two effectors have different targets within these hosts (Sánchez et al., 2012).

Mlr6358 is predicted to encode an 812 aa protein that contains about 54% similarity to the 684 aa residues of the N-terminal of Mlr6361. However, no conserved domains could be identified within the ORF of Mlr6358, so the predicted mechanism of action has yet
to be determined (Sánchez et al., 2009). Disruption of mlr6358 indirectly affects strain competitiveness on L. teniuius cv. Esmeralda (Sánchez et al., 2012).

Overall, the T3SS directly transports effectors into target cells that can influence symbiosis in either a positive or negative fashion and it is the cumulative effects of all these effectors that determine the outcome of symbiosis. Taken together, there are multiple complex signalling pathways through which rhizobia use to communicate with their hosts and the basis for this PhD project was to investigate these potential pathways on a genetic level with the end goal of an increased understanding of these complex interactions and in particular their contribution to the nodulation of Lotus pedunculatus.

1.5 Aims of this study

In this PhD project, the focus was on the interactions between two closely related Mesorhizobium loti strains and the genetics that pertain to their ability to form effective nitrogen fixing nodules on species within the genus Lotus. Specifically, two highly similar strains of M. loti, R7A and NZP2037 that exhibit different host-ranges were compared on a genetic basis and genes of interest were investigated for their potential as determinants of host-specificity. The M. loti strain NZP2037 is a representative of the broad host-range group of M. loti strains that also includes the strains NZP2014, NZP2042 and SU343. The representative host for this broad host-range was Lotus pedunculatus. In contrast to NZP2037, the M. loti strain R7A was used as the representative of the narrow host-range group of M. loti, that also includes the strains MAFF303099 and NZP2213. Both strains share a common host in Lotus japonicus Gifu but R7A is incompatible with L. pedunculatus, inducing only uninfected nodule primordia.
From previous studies, it was known that R7A exoU mutants produced truncated EPS and as a consequence exhibited a ‘rough’ colony morphology and were unable to form effective nitrogen fixing nodules on the host *L. japonicus* Gifu (Kelly *et al.*, 2013). However, it was observed that EPS mutants of NZP2037 retained the ability to form nodules on *L. japonicus* Gifu despite a rough colony morphology characteristic of defective EPS (Kelly, 2012). It was hypothesised that NZP2037 encoded genes that allowed for the accommodation of defective EPS and were likely to be unique to NZP2037. Therefore, an ongoing theme in this study was the investigation of genes that may be involved in this circumvention of defective EPS signalling.

The overall aims of this PhD project were to understand the genetic basis for the differences in host-range between the highly similar strains R7A and NZP2037, and thus gain insight into host-specificity determinants that may be involved in symbiosis. This PhD project primarily utilised a targeted markerless genetic deletion approach to disrupt genes of interest unique to NZP2037. The effects these mutations have on the symbiotic relationship of the hosts *L. pedunculatus* and *L. japonicus* Gifu were then determined. Genes of interest were identified through comparisons between the symbiosis islands of R7A and NZP2037. Genes unique to NZP2037 predicted to be under NodD regulation via the presence of putative *nod* boxes were identified (Kasai-Maita *et al.*, 2013). These genes of interest *nodU, nodFEG, nodA2* are involved in the modification of the Nod factor structure, while *nodO* encoded an effector of a T1SS that was genetically located in close proximity to *nodO*. During the course of this study, new additional genetic targets unique to NZP2037 were also identified: *mln452, mln454* and *mln399* that encoded two putative T4SS effectors and a glycosyltransferase, respectively. These additional gene targets were also targeted for mutagenesis and the effects of these mutations were investigated.
2 Methods
2.1 Strains and plasmids used in this study

*Escherichia coli* and *Mesorhizobium loti* strains used in this study are described in Table 2. Plasmids and cosmids used in this study are described in Table 3.

### Table 2. Bacterial strains used in this study

<table>
<thead>
<tr>
<th>Strain name</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>S17-1/λpir</td>
<td>Tp&lt;sup&gt;R&lt;/sup&gt; Sm&lt;sup&gt;R&lt;/sup&gt; recA thi pro hsdR-M&lt;sup&gt;*&lt;/sup&gt; recA::RP4-2-Tc::Mu nptII::Tn7 λpir</td>
<td>(Herrero et al., 1990)</td>
</tr>
<tr>
<td>ST18</td>
<td>S17 λpir ΔhemA</td>
<td>(Thoma &amp; Schobert, 2009)</td>
</tr>
<tr>
<td>Epi300</td>
<td>ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ&lt;sup&gt;-&lt;/sup&gt; rpsL nupG trfA tonA, StrR</td>
<td>Epicentre</td>
</tr>
<tr>
<td>R7A</td>
<td>Field re-isolate of ICMP3153 (NZP2238; Lc265Da) originally isolated in Ireland</td>
<td>(Sullivan et al., 2002)</td>
</tr>
<tr>
<td>NZP2037</td>
<td>Wildtype isolate from <em>L. divaricatus</em> in New Zealand</td>
<td>(Jarvis et al., 1982)</td>
</tr>
<tr>
<td>R7ANS</td>
<td>Non-symbiotic derivative of R7A that has lost the symbiosis island</td>
<td>(Ramsay et al., 2006)</td>
</tr>
<tr>
<td>R7A ΔexoU</td>
<td>R7A markerless <em>exoU</em> mutant</td>
<td>(Kelly et al., 2013)</td>
</tr>
<tr>
<td>NZP2037 Δ(nodO-9.3kb)</td>
<td>NZP2037 markerless deletion that has lost 9.3 kb of DNA including <em>nodO</em></td>
<td>Liam Harold and John Sullivan, personal communication</td>
</tr>
<tr>
<td>NZP2037 ΔnodU</td>
<td>NZP2037 markerless <em>nodU</em> mutation</td>
<td>(Fowler, 2013)</td>
</tr>
<tr>
<td>NZP2037 ΔexoU</td>
<td>NZP2037 markerless <em>exoU</em> mutant</td>
<td>(Kelly, 2012)</td>
</tr>
<tr>
<td>NZP2037 ΔnodA2</td>
<td>NZP2037 markerless <em>nodA2</em> mutant</td>
<td>(Fowler, 2013)</td>
</tr>
<tr>
<td>Strain name</td>
<td>Description</td>
<td>Reference</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>NZP2037</td>
<td>NZP2037 markerless <em>nodFEG</em> mutant</td>
<td>(Fowler, 2013)</td>
</tr>
<tr>
<td>Δ<em>nodFEG</em></td>
<td>NZP2037 markerless deletion of <em>nodO</em></td>
<td>This study</td>
</tr>
<tr>
<td>NZP2037 Δ<em>nodO</em></td>
<td>NZP2037 markerless deletion of <em>nodFEG</em> and <em>nodU</em></td>
<td>Liam Harold and John Sullivan, personal communication</td>
</tr>
<tr>
<td>NZP2037 Δ<em>nodU</em></td>
<td>NZP2037 Δ<em>nodO</em> with an <em>exoU</em> insertion mutation</td>
<td>This study</td>
</tr>
<tr>
<td><em>exoU</em>:pFUS2</td>
<td>NZP2037 Δ<em>nodO</em> with an <em>exoU</em> insertion mutation. Gm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>NZP2037 Δ<em>nodO</em>-9.3kb</td>
<td>NZP2037 Δ<em>nodO</em>-9.3kb with an <em>exoU</em> insertion mutation. Gm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td><em>exoU</em>:pFUS2</td>
<td>NZP2037 Δ<em>nodU</em> with an <em>exoU</em> insertion mutation. Gm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>NZP2037</td>
<td>NZP2037 Δ<em>nodFEG</em> with an <em>exoU</em> insertion mutation. Gm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>NZP2037 Δ<em>nodU</em></td>
<td>NZP2037 with markerless deletions of <em>nodFEG</em>, <em>nodU</em>, and <em>nodA2</em></td>
<td>This study</td>
</tr>
<tr>
<td>Δ<em>nodFEG</em> Δ<em>nodA2</em></td>
<td>Transconjugant with R7A chromosome and R7A&lt;sup&gt;C&lt;/sup&gt;:NZP2037&lt;sup&gt;l&lt;/sup&gt; island.</td>
<td>John Sullivan, personal communication</td>
</tr>
<tr>
<td>NZP2037 Δ<em>prsDE</em></td>
<td>NZP2037 markerless <em>prsDE</em> disruption</td>
<td>This study</td>
</tr>
<tr>
<td>NZP2037 Δ<em>noeKJ</em></td>
<td>NZP2037 markerless <em>noeKJ</em> deletion</td>
<td>This study</td>
</tr>
<tr>
<td>Strain name</td>
<td>Description</td>
<td>Reference</td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>R7A^C::NZP2037^I(\Delta nodU)</td>
<td>R7A^C::NZP2037^I with a markerless deletion in nodU</td>
<td>This study</td>
</tr>
<tr>
<td>R7A^C::NZP2037^I(\Delta nodU)</td>
<td>R7A^C::NZP2037^I (\Delta nodU) with an exoU insertion mutation. Gm(^R)</td>
<td>This study</td>
</tr>
<tr>
<td>JC01 (\Delta)exO::pFUS2</td>
<td>R7A^C::NZP2037^I that has markerless deletions in nod(\text{FEG},) nod(U), and nod(A2)</td>
<td>This study</td>
</tr>
<tr>
<td>JC02</td>
<td>JC01 with an exoU insertion mutation.</td>
<td>This study</td>
</tr>
<tr>
<td>JC03</td>
<td>R7A^C::NZP2037^I that has markerless deletions in (\text{nod(\text{FEG},)}) nod(U), nod(A2) and mln399</td>
<td>This study</td>
</tr>
<tr>
<td>JC04</td>
<td>JC03 with an exoU insertion mutation. Gm(^R)</td>
<td>This study</td>
</tr>
<tr>
<td>R7A^C::NZP2037^I(\Delta mln399)</td>
<td>R7A^C::NZP2037^I with a markerless deletion of mln399</td>
<td>This study</td>
</tr>
<tr>
<td>R7A^C::NZP2037^I(\Delta mln399)</td>
<td>R7A^C::NZP2037^I (\Delta mln399) with an exoU insertion mutation. Gm(^R)</td>
<td>This study</td>
</tr>
<tr>
<td>(\Delta)exO::pFUS2</td>
<td>R7A^C::NZP2037^I with markerless deletion of vir(B1-11) and vir(A)</td>
<td>This study</td>
</tr>
<tr>
<td>R7A^C::NZP2037^I(\Delta TypeIV)</td>
<td>R7A^C::NZP2037^I (\Delta TypeIV) with an exoU insertion mutation. Gm(^R)</td>
<td>This study</td>
</tr>
<tr>
<td>(\Delta)exO::pFUS2</td>
<td>R7A^C::NZP2037^I with markerless deletions in (\text{nod(\text{FEG},)}) nod(U), nod(A2), \text{vir(B1-11) and vir(A)}</td>
<td>This study</td>
</tr>
<tr>
<td>JC05</td>
<td>JC05 with an exoU insertion mutation. Gm(^R)</td>
<td>This study</td>
</tr>
</tbody>
</table>
Strain name | Description | Reference
--- | --- | ---
JC07 | R7A<sup>C</sup>:NZP2037<sup>I</sup> with markerless deletions in nod<sup>FEG</sup>, nod<sup>U</sup>, nodA<sub>2</sub>, mln399, virB1-11 and virA | This study
JC08 | JC07 with an exo<sup>U</sup> insertion mutation. Gm<sup>R</sup> | This study
JC09 | R7A<sup>C</sup>:NZP2037<sup>I</sup> with markerless deletions in nod<sup>FEG</sup>, nod<sup>U</sup>, nodA<sub>2</sub>, mln399, virB1-11 and virA and the 9.3kb region including nodO | This study
JC10 | JC09 with an exo<sup>U</sup> insertion mutation. Gm<sup>R</sup> | This study
R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔprsDE | Transconjugant with R7A chromosome and NZP2037 island that has a markerless prsDE disruption | This study

**Table 3. Plasmid and Cosmids used in this study**

<table>
<thead>
<tr>
<th>Plasmid/Cosmid</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cosmid 6B1.5</td>
<td>Derivative of pIJ3200 containing parDE. Contains nodO Te&lt;sup&gt;R&lt;/sup&gt;.</td>
<td>(Kelly, 2012)</td>
</tr>
<tr>
<td>Cosmid 8B4.5</td>
<td>Derivative of pIJ3200 containing parDE. Contains nodO. Te&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Kelly, 2012)</td>
</tr>
<tr>
<td>pFUS-exoU</td>
<td>Suicide vector derived from pFUS2 for generation of IDM mutants of exo&lt;sup&gt;U&lt;/sup&gt;. Gm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Kelly, 2012)</td>
</tr>
<tr>
<td>pJQ200SK</td>
<td>pACYC derivative, contains sacB gene. Gm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Quandt &amp; Hynes, 1993)</td>
</tr>
<tr>
<td>pJQ200SK-nodA2</td>
<td>pJQ200SK-nodA2 construct for markerless deletion of nodA&lt;sub&gt;2&lt;/sub&gt;, Gm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Fowler, 2013)</td>
</tr>
<tr>
<td>Plasmid/Cosmid</td>
<td>Description</td>
<td>Reference</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>pJQ200SK-nodO-9.3</td>
<td>pJQ200SK-nodO-9.3 construct for markerless deletion of <em>nodO</em>-9.3kb, Gm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Liam Harold and John Sullivan, personal communication</td>
</tr>
<tr>
<td>pJQ200SK-nodOsingle</td>
<td>pJQ200SK-nodO construct for markerless deletion of <em>nodO</em>, Gm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pJQ200SK-noeKJ</td>
<td>pJQ200SK-noeKJ construct for markerless deletion of <em>noeKJ</em>, Gm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>John Sullivan, personal communicatio n</td>
</tr>
<tr>
<td>pJQ200SK-mln399</td>
<td>pJQ200SK-mln399 construct for markerless deletion of <em>mln399</em>, Gm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pJQ200SK-TypeIV</td>
<td>pJQ200SK-TypeIV construct for markerless deletion of <em>virB1-11</em> and <em>virA</em>, Gm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pFAJ1700</td>
<td>Stable RK2-derived cloning vector, Ap&lt;sup&gt;R&lt;/sup&gt;, Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Dombrecht <em>et al.</em>, 2001)</td>
</tr>
<tr>
<td>pSKGFP</td>
<td>pFAJ1700 containing <em>gfp</em> expressed from the <em>nptII</em> promoter, Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Kelly <em>et al.</em>, 2013)</td>
</tr>
<tr>
<td>pFAJ1700-nodOprsDE</td>
<td>pFAJ1700 containing <em>nodO</em>, <em>prsDE</em> and a truncated <em>mln031</em>, Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pFAJ1700-noeKJ</td>
<td>pFAJ1700 containing <em>noeKJ</em>, Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td><strong>Plasmid/Cosmid</strong></td>
<td><strong>Description</strong></td>
<td><strong>Reference</strong></td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------</td>
<td>--------------</td>
</tr>
<tr>
<td>pJQ200SK-prsDE</td>
<td>pJQ200SK-prsDE construct for markerless deletion of <em>prsDE</em>; Gm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
<tr>
<td>pIJ3200-SacB</td>
<td>pIJ3200 with <em>sacB</em> fragment; Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td>John Sullivan, personal communication</td>
</tr>
<tr>
<td>pPROBE-KT</td>
<td>Empty pPROBE-KT vector; Neo&lt;sup&gt;R&lt;/sup&gt;/Km&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Miller <em>et al.</em>, 2000)</td>
</tr>
<tr>
<td>pPROBE-KT-nodU</td>
<td>pPROBE-KT containing <em>nodU</em> from NZP2037; Neo&lt;sup&gt;R&lt;/sup&gt;/Km&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Kelly, 2012)</td>
</tr>
<tr>
<td>pFAJ1700-exoU</td>
<td>pFAJ1700 containing expressing the <em>exoU</em> gene; Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td>(Kelly <em>et al.</em>, 2013)</td>
</tr>
<tr>
<td>pFAJ1700-nodOprsDEmln031</td>
<td>pFAJ1700 containing <em>nodO, prsDE</em> and the full length <em>mln031</em>; Tc&lt;sup&gt;R&lt;/sup&gt;</td>
<td>This study</td>
</tr>
</tbody>
</table>

### 2.2 Media and growth conditions

*E. coli* strains were incubated at 37°C. Media used for the growth of *E. coli* strains were Luria-Bertani (LB) agar/broth (Miller, 1972) and Tryptone yeast (TY) agar/broth (Beringer, 1974). Super Optimal Broth (SOB) was used for generation of electrocompetent cells. For the growth of *E. coli* strain ST18, the growth medium was supplemented with 50 μg/mL of 5-aminolevulinic acid. *Mesorhizobium loti* strains were incubated at 28°C. Growth of *Mesorhizobium* was on rhizobium defined media (RDM) broth and agar plates (Ronson *et al.*, 1987) supplemented with 0.4% glucose (G/RDM). RDM supplemented with 5% sucrose (S/RDM) instead of glucose as the carbon source was used for selection of markerless deletion mutants. Antibiotics (Table 4) were added.
to media as required. Agar was made via addition of 16 g/L of Davis Agar to broth.
Recipes for broths are presented in the Appendix.

2.2.1 Storage of bacterial strains

Bacterial strains were stored at -80°C and prepared via addition of 70 μL of dimethyl sulfoxide (DMSO) to 800 μL of bacterial cultures that had been incubated in appropriate media and temperature until stationary phase (~ 1 day for *E. coli* grown in LB broth and ~2 days for *M. loti* grown in G/RDM broth).

2.3 Enzymes and chemicals

Enzymes were purchased from Roche Diagnostics and New England Biolabs. Antibiotics and chemicals were purchased from Sigma. Antibiotics were dissolved in water and filter sterilized via a 0.45 μm filter at appropriate concentrations, the exception being tetracycline which was dissolved in 100% methanol. Antibiotic concentrations are detailed on Table 4.

**Table 4. Antibiotics used in this study**

<table>
<thead>
<tr>
<th>Antibiotic and abbreviation</th>
<th>Concentration used (μg/mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>M. loti</em></td>
</tr>
<tr>
<td>Gentamicin (Gm)</td>
<td>50</td>
</tr>
<tr>
<td>Tetracycline (Tc)</td>
<td>2</td>
</tr>
<tr>
<td>Kanamycin (Km)</td>
<td>-</td>
</tr>
<tr>
<td>Phosphomycin</td>
<td>25</td>
</tr>
<tr>
<td>Neomycin (Neo)</td>
<td>100</td>
</tr>
</tbody>
</table>
2.4 Restriction enzyme digestion

DNA was digested with restriction enzymes in appropriate buffer for at least 1 h at 37°C.

2.5 Agarose gel electrophoresis

DNA samples were mixed with one-tenth volume bromophenol blue tracking dye (1 mL H₂O, 1 mL glycerol, 1 mL bromophenol blue [10 mg/mL]) prior to loading onto gels made up of 1% agarose dissolved in 1x Tris-acetate (TAE) buffer (40 mM Tris [pH 8], 20 mM acetic acid, 1 mM EDTA [pH 8]) containing 1 μg/mL ethidium bromide. Gels were electrophoresed at 90 V for approximately 1 hour. DNA was visualised and captured on an El Logic 200 Gel Documentation system (Kodak Ltd.) or an Alphalmager® HP (ProteinSimple).

2.6 Polymerase chain reaction (PCR)

PCR was used to amplify DNA for use in mutant construction, complementation, and as a tool for the confirmation of successful mutants. The sequences of PCR primers used in this study are listed in Table 5. Primers were ordered from Invitrogen (Life Technologies) and/or Integrated DNA Technologies (IDT). PCR reactions utilised the Phusion High-Fidelity PCR system (Finnzymes). Standard 100 μL PCR reactions typically contained 10 μL 5x Phusion HF buffer, 1 μL of supplied dNTPs (10 mM), 1 μL of each primer (final concentration 0.5 μM), 0.5 μL of template, 0.25 μL of Phusion DNA polymerase enzyme, filter-sterile Milli-Q H₂O was used to fill out the volume totals. PCR products were purified with the Illustra GFX PCR DNA and Gel Band Purification Kit (GE healthcare product# 28-9034-71) according to the manufacturer’s instructions.
2.6.1 Primers used in this study

Table 5. Primers used in this study

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>nodOcheckL</td>
<td>CAACTGCTTGCTAGGAGCAA</td>
<td>S</td>
</tr>
<tr>
<td>nodOinternalcheckR</td>
<td>CGAATCTTTGGATGTTAGTG</td>
<td>S</td>
</tr>
<tr>
<td>nodOcheckR</td>
<td>GATATGCGGTCTCTCAACA</td>
<td>S</td>
</tr>
<tr>
<td>nodOcompH3L</td>
<td>AATTAAGCTTGCAATGAGATTAGGACCTGTC</td>
<td>C</td>
</tr>
<tr>
<td>hltcAbpH3rev</td>
<td>TTAAGCTTTGCATTACCAGGCTCGAGT</td>
<td>C</td>
</tr>
<tr>
<td>NodOLLapa1</td>
<td>AAATTTGGCGGTAGCATAGCAAGGCTATCA</td>
<td>D</td>
</tr>
<tr>
<td>nodOsinglegeneRL</td>
<td>ATAAGATACCTTGTTCCCTCGGTACCTGCTTGATT</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>GGAA</td>
<td></td>
</tr>
<tr>
<td>nodOseqL1</td>
<td>CTCGGCGCGGCTATGAGTGAC</td>
<td>S</td>
</tr>
<tr>
<td>nodOseqL3</td>
<td>CGGACGCGTTACGATGCTAT</td>
<td>S</td>
</tr>
<tr>
<td>nodOseqR2</td>
<td>CGATTCGTCGCAGGTCGTTTCTCABC</td>
<td>S</td>
</tr>
<tr>
<td>nodOseqR1</td>
<td>CGGTGCCCACGAGGAACA</td>
<td>S</td>
</tr>
<tr>
<td>nodOseqL2</td>
<td>CGAGAACCAGCGCGAGGACTT</td>
<td>S</td>
</tr>
<tr>
<td>nodOseqR3</td>
<td>CGCCGGCTTTCACTCATCATTCT</td>
<td>S</td>
</tr>
<tr>
<td>nodOsinglecheckR</td>
<td>GCACCGCGCTGCTCGCCAA</td>
<td>S</td>
</tr>
<tr>
<td>nodOsinglegeneLR</td>
<td>TTTCCGATCAGACGAGGTACCGGAACAGGT</td>
<td>S</td>
</tr>
<tr>
<td></td>
<td>ATCTAT</td>
<td></td>
</tr>
<tr>
<td>nodOseqR4</td>
<td>TCWGAGATGAGCTCGTGTT</td>
<td>S</td>
</tr>
<tr>
<td>nodOsinglegeneRR</td>
<td>AATTTACTAGGATAGCACCCTTCACGGATCCT</td>
<td>D</td>
</tr>
<tr>
<td>Cosmid check Ca2+</td>
<td>ACAGCACTGCTGCTCAACGT</td>
<td>S</td>
</tr>
<tr>
<td>Cosmid check parDE</td>
<td>AGGTCAAGGGCTGATTTC</td>
<td>S</td>
</tr>
<tr>
<td>noeKJcloneL</td>
<td>AGAAGCATCCAGATCCCAGAATACGGGAA</td>
<td>C</td>
</tr>
<tr>
<td>Primer</td>
<td>Sequence</td>
<td>Use</td>
</tr>
<tr>
<td>----------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-----</td>
</tr>
<tr>
<td>noeKJcloneR</td>
<td>GATGAGCTCAGGGCATGACGTTGGACTAT</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>CA</td>
<td></td>
</tr>
<tr>
<td>noeKJLL</td>
<td>GCGGCGCCTCTAGAACTAGTGATCGCGGCC</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>AATACGGGA</td>
<td></td>
</tr>
<tr>
<td>noeKJLR</td>
<td>AGGGCATGTGACGTTGGACTATCA</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>GTGCCAGCGCAGCGTAGGT</td>
<td></td>
</tr>
<tr>
<td>noeKJRL</td>
<td>ACCTACGCTGCGCTGGCAC</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>TGATAGTCAACGTACATGCCCT</td>
<td></td>
</tr>
<tr>
<td>noeKJRR</td>
<td>GCGAATTGGGTACCCGGCGGCCGGTTCACGC</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>GCCGCTTA</td>
<td></td>
</tr>
<tr>
<td>NoeKJFlankcheckL</td>
<td>CCGCATACTGGAACAGCAGGTGG</td>
<td>S</td>
</tr>
<tr>
<td>NoeKJFlankcheckR</td>
<td>TCCGAGCUCTGGTGTTAAACGN</td>
<td>S</td>
</tr>
<tr>
<td>NoeKJInternalCheckL</td>
<td>ACCAACAGTTCCGGCCCGGA</td>
<td>S</td>
</tr>
<tr>
<td>NoeKJInternalCheckR</td>
<td>TCCGAGCUCTGGTGTTAAACGN</td>
<td>S</td>
</tr>
<tr>
<td>NoeKJFlankcheckR</td>
<td>CGATCACCTTCCGCAACTCGA</td>
<td>S</td>
</tr>
<tr>
<td>M13 forward</td>
<td>CCCAGTCACGACGTTGTAACAG</td>
<td>S</td>
</tr>
<tr>
<td>M13 reverse</td>
<td>AGCGGATAAAACTTTCCACAG</td>
<td>S</td>
</tr>
<tr>
<td>prsDEdeletionLL</td>
<td>GGCAGGCGCCTCTAGAACTAGTGGAATCCGAC</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>TTTATCGACGCGGTCCCA</td>
<td></td>
</tr>
<tr>
<td>prsDEdeletionLR</td>
<td>GGGTGGGAAGGGATCGCACTCTAAGGTCTCC</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>ACCTGAAATGTTCGAG</td>
<td></td>
</tr>
<tr>
<td>prsDEdeletionRL</td>
<td>CTCGAACATTTCAGTGGAGACCCTAGAG</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>GCAGATCCCTTCCACCC</td>
<td></td>
</tr>
<tr>
<td>prsDEdeletionRR</td>
<td>TATAGGGCGAATTGGGTACCGGGGCGACGGCTC</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>AAGCTCAACCTAGGACACC</td>
<td></td>
</tr>
<tr>
<td>Primer</td>
<td>Sequence</td>
<td>Use</td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-----</td>
</tr>
<tr>
<td>prsDE_CheckL</td>
<td>CGCCGCGAATGCCATCTCCG</td>
<td>S</td>
</tr>
<tr>
<td>prsDE_CheckR</td>
<td>AGCATGTCGTGCCGCCGCCTC</td>
<td>S</td>
</tr>
<tr>
<td>prsDE_InternalCheckR</td>
<td>GTATCGCGTATCGGGCGTTGCG</td>
<td>S</td>
</tr>
<tr>
<td>nodA2flankcheckL</td>
<td>TGGTCTAGCAACGCAACTGA</td>
<td>S</td>
</tr>
<tr>
<td>nodA2flankcheckR</td>
<td>ATCGCGCTCGGATCGATTTA</td>
<td>S</td>
</tr>
<tr>
<td>nodA2InternalCheckL</td>
<td>CATTCAATGCCTGTTGTA</td>
<td>S</td>
</tr>
<tr>
<td>6.2islandcheckL</td>
<td>CCTCGCGAATGCAACATCC</td>
<td>S</td>
</tr>
<tr>
<td>6.2islandcheckR</td>
<td>CGTCTTGGCTTCCGCAATCA</td>
<td>S</td>
</tr>
<tr>
<td>27.5islandcheckR</td>
<td>CATGTTTCATGTCGATCA</td>
<td>S</td>
</tr>
<tr>
<td>27.5islandcheckL</td>
<td>CGCCTACAACTTATGCACTT</td>
<td>S</td>
</tr>
<tr>
<td>52.8islandcheckR</td>
<td>CATCAGCAATCGTTACCAA</td>
<td>S</td>
</tr>
<tr>
<td>downstreamisland</td>
<td>TCTGACCTACGGCATGCTCT</td>
<td>S</td>
</tr>
<tr>
<td>mln399_del_LL</td>
<td>GGCGGGCCGCTCTAGAACTTAGTGATCCCCCG</td>
<td>D</td>
</tr>
<tr>
<td>mln399_del_LR</td>
<td>TGATCGGCTGGTGCAA</td>
<td>D</td>
</tr>
<tr>
<td>mln399_del_RL</td>
<td>TGCAGCGACTGTCTATGTGGAAGGCAGCA</td>
<td>D</td>
</tr>
<tr>
<td>mln399_del_RR</td>
<td>AGAACAGTGCGCGCA</td>
<td>D</td>
</tr>
<tr>
<td>mln399_checkL</td>
<td>TATAGGGCGAATGGGTACCGGCCCACAG</td>
<td>D</td>
</tr>
<tr>
<td>mln399_InternalCheckR</td>
<td>ACTCGGTCCGGTCAACCGTGCTCT</td>
<td>S</td>
</tr>
<tr>
<td>mln399_checkR</td>
<td>GGCATAGTCCGGCGGTCGCA</td>
<td>S</td>
</tr>
<tr>
<td>LL_TypeIVSS_del</td>
<td>GGCAGCGGCTCTAGAACTTAGTGATCCGTGTA</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>GAACAGCTGCCCATGATCCA</td>
<td></td>
</tr>
<tr>
<td>Primer</td>
<td>Sequence</td>
<td>Use</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------------------</td>
<td>-----</td>
</tr>
<tr>
<td>LR_TypeIVSS_del</td>
<td>ACGGTGTCATAACCGGTCAGTGTC</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>TCCTATCCCCTACACCTTTCCAAGCGT</td>
<td></td>
</tr>
<tr>
<td>RL_TypeIVSS_del</td>
<td>ACGCTTGAAAAGTAGGGGATAGGA</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>CACAAAGACGGCTTAGACACCGT</td>
<td></td>
</tr>
<tr>
<td>RR_TypeIVSS_del</td>
<td>TATAGGGCGAATTGGGTACCCGGCCGGCG</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>CACATCTTTGCTGCAA</td>
<td></td>
</tr>
<tr>
<td>TypeIV_checkL</td>
<td>GCGAGACCTCCACATGGATCCTG</td>
<td>S</td>
</tr>
<tr>
<td>TypeIV_checkR</td>
<td>TGCTCCCTCGGGACATGGCC</td>
<td>S</td>
</tr>
<tr>
<td>TypeIV_Internal_checkR</td>
<td>TTGCGGCAAACCTCGACCTTGC</td>
<td>S</td>
</tr>
<tr>
<td>faj3</td>
<td>GGACAACTGCTACGAGATGTCTTG</td>
<td>S</td>
</tr>
<tr>
<td>faj5</td>
<td>AGGTCTCCAGACTGGAAAGC</td>
<td>S</td>
</tr>
<tr>
<td>Ca2+_check</td>
<td>GGAAGCGATCTGCTTTG</td>
<td>S</td>
</tr>
</tbody>
</table>

D = Deletion, S = Sequencing/Primers used for confirmation of mutants, C = Cloning for complementation
2.6.2 Phusion PCR program:

Step 1: 98°C for 30 seconds,
Step 2: 98°C for 10 seconds
Step 3: X°C for 30 seconds (X°C Temperature varies depending on primers)
Step 4: 72°C for 15 seconds per 1 kb of product
Step 5: Repeat Steps 2 to 4 for around 30 times
Step 6: 72°C for 5 minutes

2.7 Plasmid DNA extraction

2.7.1 Alkaline lysis method

Plasmid DNA was extracted using a method based on (Feliciello & Chinali, 1993). Overnight cultures were incubated in LB/TY broths at 37°C and 3 mL of culture was harvested via centrifugation at 16,000 g for 1 minute. Pelleted cells were re-suspended in 250 μL of Buffer P1 (QIAGEN) and the cells were lysed via addition of 250 μL lysing solution (0.2 M NaOH, 1% [w/v] SDS) and inverted 6 times. Three hundred and fifty microliters of neutralising solution (3 M Potassium Acetate, 5% [w/v] formic acid) was added to the cell mixture and inverted 6 times. The cell mixture was then centrifuged at 16,000 g for 5 min, following centrifugation, 800 μL of the bacterial supernatant was transferred to a new microfuge tube. DNA was precipitated from the supernatant via addition of 600 μL of isopropanol and the samples were centrifuged again at 16,000 g for 10 min and the supernatant was discarded. The DNA pellets were then washed in 700 μL of 70% ethanol and centrifuged at 16,000 g for 5 minutes, the supernatant was removed and the tube was air-dried at 37°C for at least 10 min. The DNA pellets were then re-suspended in 50 μL of sterile Milli-Q water.
2.7.2 Commercial plasmid/cosmid isolation kit

High-quality plasmid DNA was isolated using commercial kits. For plasmids, the Zyppy™ Plasmid Miniprep Kit (Zymo research) was used and for Cosmids the QIAGEN Plasmid Midi Kit (25) (Cat No./ID: 12143) was used. Extractions were performed according to the manufacturer instructions that were provided with their respective kits.

2.8 Genomic DNA extraction

2.8.1 PrepMan™ Ultra genomic DNA preparations

Crude genomic DNA for use in PCR was extracted from bacterial pellets harvested from 200 μL of overnight broth cultures that were spun down at 16,000 g for 3 minutes. The pellets were then re-suspended in 100 μL of PrepMan™ Ultra reagent (Applied Biosystems) and vortexed for ~30 seconds or until the pellets have been re-suspended. The re-suspended cells were then boiled at 100°C for 10 min and harvested at 16,000 g for 3 min and 75 μL of the supernatant was transferred into a clean tube for use in downstream applications. DNA was also extracted from colonies using a similar protocol. However, instead of bacterial overnight broth cultures, single colonies were re-suspended in 50 μl of PrepMan™ Ultra reagent and boiled at 100°C for 10 min. Cells were then harvested at 4700 g for 4 minutes and 30 μl of supernatant was transferred to a clean tube ready for downstream applications.

2.8.2 Commercial genomic DNA extraction kit

High-quality genomic DNA was isolated using a commercial kit from Mo Bio laboratories, Inc., UltraClean® Microbial DNA Isolation Kit (Catalog # 12224-50).

2.9 Spectroscopy

Absorbance readings and concentrations of DNA were measured using a Nanodrop ND-
100 Spectrophotometer (Nanodrop Technologies, USA) and/or a NanoDrop One/One Spectrophotometer (Thermo Scientific).

2.10 Ligations

Appropriate volumes of prepared plasmid and insert DNA were mixed with 2.5 μL of 10x ligation buffer, 1.5 μL of T4 DNA ligase (Roche cat# 481220) in a total volume of 20 μL and incubated overnight at 12°C. Alternatively, 1.5 μL T4 DNA ligase (New England Biolabs cat# M0202) was used and this was incubated at room temperature for 20 minutes. Ligated DNA was ethanol-precipitated with the addition of 50 μL 100% ethanol, 2 μL of 3 M sodium acetate and 1 μL of Pellet paint® co-precipitant (Novagen), pelleted at 16,000 g and then washed with 100 μL 70% ethanol and the liquid decanted before resuspension in 5 μL of filter-sterile Milli-Q water for use in downstream applications.

2.11 Electroporation

2.11.1 Preparation of E. coli

Electrocompetent cell stocks of E. coli EPI300 and ST18 were prepared using a protocol adapted from (Sheng et al., 1995). Strains were grown to stationary phase in 5 mL LB broths and 0.5 mL of this culture was used to seed 500 mL broths of SOB media and incubated at 37°C with shaking at 200 rpm. Cultures were incubated until they had reached an OD_{600} of ~0.6-0.8 and were harvested by centrifugation at 4°C for 10 min at 7000 g. Following two washes in 500 mL chilled 10% [w/v] glycerol, cells were washed in 30 mL of chilled 10% glycerol, pelleted at 4°C for 10 min at 7000 g, then resuspended in 1 mL of 10% glycerol. Aliquots (50 μL) of the now electrocompetent bacterial suspensions were then stored at -80°C for further applications.
2.11.2 Electroporation of *E. coli*

Electrocompetent cells (50 µL) were thawed on ice. DNA (1-5 µL) was added to the thawed cells and incubated on ice for ~5-10 minutes and then transferred to a pre-chilled 1 mm gapped electroporation cuvette (Biorad). Cells were then transformed using a Gene Pulser Xcell™ electroporation system (Biorad) using pre-set parameters (1800 V, 200 Ω, 25 µF). Immediately following the electrical shock, *E. coli* cells were resuspended in 1 mL of LB broth and incubated at 37°C with shaking for 60 minutes before dilutions of the cell suspension were plated out onto LB plates containing appropriate antibiotics.

2.12 Transfer of vectors by conjugation

Plasmids containing an RK6 origin of replication were transferred from *E. coli* strains into *M. loti* strains through biparental spot-matings. For spot-matings *E. coli* and *M. loti* strains were grown to stationary phase in TY broth. Aliquots of each culture (25 µL) were dispensed together as a spot onto a TY agar plate and incubated at 28°C overnight. The resultant bacterial growth was then streaked onto selective media containing appropriate antibiotics.

2.13 DNA sequencing

Plasmid and PCR products were sequence verified using the primers described in Table 5. Primers were mixed with template DNA and filter-sterile Milli-Q water, and the mixture sent to the Allan Wilson Centre Genome Service (Massey University, Palmerston North, NZ) for sequencing. For whole genome sequencing the DNA template was prepared using the UltraClean® Microbial DNA Isolation Kit (Catalog # 12224-50) from Mo Bio laboratories, Inc. The DNA template was then sent to Mr. DNA (Molecular Research LP), 503 Clovis Rd, Shallowater, TX 79363, USA for whole genome sequencing using the Illumina MiSeq platform.
2.14 Computer analysis

DNA sequence was viewed and edited using 4peaks software (Mekentosj.com) and/or SnapGene software (from GSL Biotech; available at snapgene.com). Further manipulation of DNA sequences was performed using the Lasergene suite of software (DNASStar). The National Center for Biotechnology Information (NCBI) databases were utilised and searches for similar nucleotide or amino-acid sequences were performed using the BLAST N, X and P algorithms (Altschul et al., 1997). Protein domain predictions utilised the Interpro database (Finn et al., 2016). The software suite Geneious (Kearse et al., 2012) was used for the mapping and alignment of reads that were generated from whole genome sequencing and used in conjunction with FLASH software (Magoč & Salzberg, 2011). Genome sequences and genomic comparisons for relevant strains were obtained from NCBI and the Joint Genome Institute (JGI) database (Grigoriev et al., 2011).

2.15 Mutagenesis

2.15.1 Homologous recombination protocol for deletion mutants

Markerless deletion mutants were constructed through allelic replacement of the wild-type gene with a mutated form through homologous recombination as detailed below.

2.15.1.1 Overlap extension PCR and Gibson assembly

Two approaches for generation and integration of DNA fragments that were ligated into suicide vectors were used in this study, overlap extension PCR and Gibson cloning.
2.15.1.1.1  **Overlap extension PCR**

In this approach, generation of markerless deletion mutant fragments was done with primer pairs xLL/xLR and xRL/xRR to PCR amplify left and right arms of ~1 kb which encoded respective flanking regions either side of the target gene. Primer design incorporated restriction enzyme sites at the outermost ends of the left and right arm PCR products and ~20 bp of overlapping sequence between the internal regions of the two arms.

The first PCR reactions generate two PCR products that are then used as template DNA for a further PCR reaction using the primer pair containing the outermost left-arm (xLL) and right-arm (xRR) primers. The end result of the two step PCR is an amplified ~2-kb product made up of the two arms joined by an overlapping sequence. The ~2 kb overlap extension PCR product was digested with the appropriate restriction enzymes and ligated into the multi cloning site of the suicide vector pJQ200SK (Quandt & Hynes, 1993). Electrocompetent *E. coli* cells were then transformed with the mutant construct and selection on LB containing gentamicin was performed. Plasmid DNA extracted from gentamicin resistant clones were confirmed as containing the overlap extension PCR product via restriction digestion and sequencing was used to verify the integrity of the sequence that had inserted within the vector.

2.15.1.1.2  **Gibson assembly**

The second approach for generation of DNA fragments for ligation into a vector was through the commercially available Gibson Assembly® Cloning kit (Cat# E5510S) or NEBuilder® HiFi DNA Assembly Tool (Cat# E2621S). Similar to the overlap extension PCR method, primer pairs xLL/xLR and xRL/xRR were designed that amplify left and right arms of ~1 kb which encoded respective flanking regions either side of the target.
The difference between Gibson and overlap extension is that the outermost primers (xLL/xRR) also contained sequence that overlaps that of the vector construct. The primer pair was then used to generate the first PCR products of ~ 1kb in length. The two ~1kb PCR products that were generated from the previous PCR reactions were then both ligated into the MCS pJQ200SK (Quandt & Hynes, 1993) using the commercial kit as per their manufacturer’s instructions. Electrocompetent ST18 *E. coli* cells were then transformed with the mutant construct and selection on LB containing gentamicin was performed.

Gibson assembly was also used to generate the plasmid construct pFAJ1700-nodOprsDEmln031. Construction of the vector took advantage of the activity of the exonuclease within the Gibson assembly mix to trim 5’ ends of DNA fragments to remove extra unwanted bases. The gBlock® Gene Fragment (synthetic double-stranded DNA fragments ordered from IDT (sequence shown in Table 6) was designed to restore full length *mln031* and once verified was ligated onto the end of the plasmid construct pFAJ1700-nodOprsDE treated with *BamHI*. The gBlock and plasmid were ligated using the NEBuilder® HiFi DNA Assembly Tool (Cat# E2621S) as per the manufacturer’s instructions.
Table 6. gBlock used for Gibson assembly to generate the construct pFAJ1700-nodOprsDEmln031

<table>
<thead>
<tr>
<th>gBlock Name</th>
<th>Sequence</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ca2+ bp</td>
<td>GTTGAACCTGAGGGTGTAATACGATCACGCTCTCCA</td>
<td>Restoration of full length mln031 for complementation</td>
</tr>
<tr>
<td></td>
<td>ACGTGAGTGCTGCAGACGTTTCAACGCGCATCGAAACC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TTTATTCTTCTCGTCTAGGTGTTCAGTCCCGAGGTCAG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TGGACCCGAATCGATCCCGAAGCGATCTGGCTGTTTG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CCGATCCGCTGCAATTAGCTTCGTAAGGTGTAAGGCCA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TGAAGGCTTTGAGCCGCTGGCCGAAGTTGAGGATCCAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CGAGCAGAGGCGGCAAGGTGATTTCAGTTGGGCCGCTG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GGTCGGAAAACCGGCAGGATCGGGTAATACGCGACCGACGA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ACGTGCCGTGCAATTGACCGCCGATCAACCCGGGATTG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GGATCCGCGCGCGCCGCCGCGTTAA</td>
<td></td>
</tr>
</tbody>
</table>

Plasmid DNA extracted from tetracycline resistant clones were isolated and sequencing was used to confirm the sequence integrity of the DNA insert within pFAJ1700-nodOprsDEmln031.

2.15.1.2 Isolation of markerless deletion mutants

Appropriate *M. loti* strains were spot-mated with ST18 *E. coli* strains that harboured the suicide vector to allow for transfer of the suicide vector via conjugation. Twenty-five µL volumes of both recipient (*M. loti*) and donor strains (*E. coli*) were spotted together and grown on TY agar supplemented with 5-aminolevulinic acid for ~ 24 hours before bacterial spots were streaked out for antibiotic selection on G/RDM containing
gentamicin and incubated at 25°C for ~4 days. Single colonies were then passaged another 2 times on G/RDM containing gentamicin to ensure that the colonies were pure.

To complete the markerless deletion process, a second homologous recombination event was required to remove the integrated pJQ200SK vector from the genome and form double-crossover clones. Selection for the second-crossover event was obtained by incubating gentamicin resistant strains obtained from the first mating in TY broth at 28°C with shaking for ~2 days. Broth cultures were then diluted and plated onto sucrose/RDM (S/RDM) agar plates and incubated at 28°C for at least 96 hours. The sucrose allowed for the selection of strains that had lost the suicide vector through a second-crossover event and therefore lost the sacB gene. Single colonies from the S/RDM plates were then passaged onto S/RDM for 2 more times and also checked for gentamicin sensitivity. Genomic DNA was then recovered from potential mutants and this DNA was used as template for a PCR with a primer pair that binds to the flanking regions of the wildtype gene. This primer pair would generate smaller PCR products than those in the wildtype control. Sequencing of the larger PCR product was then performed to confirm a successful deletion. In most cases, another primer pair was also used as a negative control. One primer binds to the flanking region of the gene, and the other primer binding within the gene of interest. This primer pair would only produce a PCR product in the wildtype background and no PCR product in successful mutants.

2.15.2 Insertion Duplication Mutants (IDM)

IDMs, generate insertion mutants that inactivate a gene of interest through the integration of the plasmid construct. Most IDMs made in this study were constructed using the suicide plasmid pFUS-exoU. E. coli S17-1/λpir containing the suicide vector (donor) and
M. loti (recipient) were spot-mated and incubated on TY agar plates at 28°C for at least 24 hours. Single-colony purifications of the mating spots was performed to isolate desired mutant strains via passaging on G/RDM containing gentamicin and phosphomycin for at least 2 passages. Selection bias for colonies that exhibited a rough colony phenotype was performed to select for mutants that had successfully integrated the vector into the exoU gene of the recipient.

2.15.3 Mobilisation of the NZP2037 symbiosis island

Stationary-phase TY broth cultures of the donor (M. loti strain that contains the island for transfer) and recipient (R7ANS containing the vector pIJ3200-SacB) were dispensed (500 μL each culture) onto the surface of a 0.45 μm filter (Type HA, 47 mm, Millipore Corporation, USA) via a syringe. Filters were then transferred to a TY agar plate and incubated at 28°C for at least 24 h. Filter growth was resuspended in 3 mL of sterile MilliQ H₂O using a flamed spreader and dilutions up to 10⁻² of the suspension were plated onto G/RDM medium containing tetracycline and incubated at 28°C. Colonies that appeared at an early onset of ~4 days incubation were passaged an additional 2 times on G/RDM containing tetracycline. Single colonies were then inoculated into TY broths which were incubated at 28 °C with shaking for 2 days. Dilutions of the cell culture were plated out onto S/RDM to select for colonies that had lost the vector pIJ3200-SacB. Single colonies were then passaged onto S/RDM 2 more times and checked for tetracycline sensitivity. PCR of the three regions where the tri-partite island integrates was then performed using the islandcheck primers listed in Table 5 to screen for successful integration of the island.

2.16 Cosmid library screen and transfer into M. loti

The NZP2037 genomic cosmid library was contained within 20 x 96 well plates that were
stored at -80 °C. Half of the 96 well plates (48 wells) were screened via first incubation of bacteria from the 48 wells into 20 mL of LB broth containing Tc using a flame sterilized stamper. Ten mL of the LB broth was then transferred into a sterile glass universal and incubated at 37°C overnight. Cosmids were then purified from the overnight broths using the QIAGEN Plasmid Midi Kit (25) (Cat No./ID: 12143). A screen with primers that amplified the region of interest was used to identify half plates that contained cosmids containing the region of interest. The 48 wells from the half plates were then separated into 6 columns each containing 8 wells and these were once again screened for the region of interest, narrowing down the list of potential cosmids to respective columns. Lastly, 1 well was isolated from the 8 wells that contained the *E. coli* strain that harboured the cosmid of interest. The cosmid of interest was then purified and transformed into the *E. coli* strains Epi300 and ST18 via electroporation. Epi300 was used to generate large amounts of plasmid for DNA sequence analysis while ST18 was used as the donor strain for conjugation of the cosmid into *M. loti* strains of interest.

### 2.17 Plant experiments

#### 2.17.1 Seedling preparation and inoculation

Plant seeds were surface-sterilised to prevent contamination that may influence the results of plant studies. Surface sterilisation was performed by washing seeds in 70% ethanol for 1 minute. The ethanol was then decanted and another wash in a 1:1 mix of 95% ethanol and 30 % hydrogen peroxide was performed for 4 minutes. The liquid was decanted and the seeds were washed with sterile H₂O by inverting the seed container, a step that was repeated 3 times. After the final wash, the liquid was removed and the seeds were transferred into an empty sterile petri dish. The seeds were then submerged in sterile H₂O, and the water replaced with fresh sterile H₂O every 30 min, for 3 times. The seeds were then transferred onto 0.8% watery agar (H₂O with agar only) and left to germinate in the
dark at \( \sim 25^\circ C \) for 24-48 h. Following germination, seeds were placed into sterile 18 mm test tubes containing 8 mL slant of Jensen’s seedling agar (Vincent, 1970) and incubated on the slants for 24 h in the plant room. Seedling roots were then inoculated with bacterial growth from fresh G/RDM plate cultures that was resuspended in \( \sim 5\)-10 mL sterile water using sterile loops. Each seedling root was inoculated with 100 \( \mu \)L of the appropriate bacterial suspension. Plant seeds used in this study were *Lotus japonicus* ecotype Gifu provided by Dr. S. Kelly University of Aarhus, Denmark., and *L. pedunculatus* Grasslands cv. Trojan supplied by PGG Wrightson Seeds Ltd.

### 2.17.2 Plant growth conditions

Plants were grown in a room with a controlled environment. Relative humidity was kept at 70\% with the temperature maintained at 22-25\(^\circ\)C during the day and 14\(^\circ\)C at night on a 16 h day / 8 h night cycle. During the period of time covered by this study, the plant room conditions were altered to reduce the temperature during the day to 21\(^\circ\)C allow for better growth conditions for the model host plant *L. japonicus* Gifu. The lights were also slightly raised to lower light intensity.

### 2.17.3 Plant nodulation observations

Various host plants inoculated with *Mesorhizobium* strains were observed periodically for nodule formation over a period of 4-7 weeks starting at day 7 post inoculation. The symbiotic effectiveness of strains was determined via comparison to plants inoculated with wild-type positive controls and the uninoculated negative control. Plants that formed effective nitrogen-fixing nodules that were considered healthy exhibited green foliage and seedlings had larger wet weights. In contrast, plants that formed no nodules or small nodule primordia exhibited stunted growth and yellowing foliage.
2.17.4 Isolation of rhizobia from nodules

Nodules that had formed on plants inoculated with strains of interest at the conclusion of plant assays (usually 6 weeks, post inoculation) were randomly selected. Nodules were surface-sterilised for 1 min in 70% ethanol, followed by washing in a 1:1 mixture of 95% ethanol and 30% hydrogen peroxide for 1 min. Liquid was decanted and the nodules washed four times with sterile H₂O. Individual nodules were then crushed in 50 μL sterile water using sterile wooden sticks. The resultant milky exudates (50 μL) were then streaked onto G/RDM plates containing antibiotics (if applicable) and incubated at 28°C. All strains that harboured IDM of exoU were re-isolated from nodules and confirmed to exhibit a ‘rough’ colony morphology indicative of disruption of exoU via the pFUS2 construct. No strains were found to have reverted to a ‘smooth’ colony morphology, indicating that the pFUS2 construct was relatively stable within the nodule environment.

2.17.5 Infection thread assays

2.17.5.1 Pillow system

Seedlings that were used for infection thread assays via the pillow system were grown between two sterile nylon pillows (5 x 5 x 19 cm) filled with a 6:1 vermiculite:perlite mix that were placed in sterile glass trays (Szczyglowski et al., 1998). The pillows were soaked in Hoagland’s solution [recipe taken from (Gibson, 1980), shown in Appendix] for 30 min prior to planting of seedlings, and excess liquid was decanted before placing the seedlings between the two pillows. Ten mL of bacterial suspension that was taken from fresh G/RDM plates was used to inoculate the seedlings, and the pillows were then sandwiched to provide the roots with shelter. Pillows were irrigated every two days, alternating with 200 mL H₂O and 200 mL Hoagland’s solution.
2.17.5.2 Plate system

*L pedunculatus* seedlings for infection thread assays were grown in a plate system. Around 50 ml of Jensen’s agar was aliquoted into square plates measuring 10 x 10 cm and allowed to set at an angle to form an agar slant. Sterilised Kimwipes™ Delicate Task Wipers were then placed on the top of the agar slants and flattened using a flamed spreader to remove air bubbles. Newly germinated seedlings were then placed onto the slants and incubated for 24 h. Plants were then inoculated with 100 µL of bacterial suspension from fresh G/RDM plates per seedling. Plates were incubated facing upright so as to allow the root to grow downwards and harvested when ready. It should be noted that the roots of the seedlings were exposed to the overhanging lights during incubation.

2.17.5.3 Microscopy

When needed, plants were carefully removed from plates and the roots were submerged and stained with 10 µg/mL of propidium iodide for 20 min. The roots were then placed on a microscope slide and covered with a glass cover slip. The slides were examined using an Olympus microscope (model BX51TRF) with fluorescence illuminator (model BX-RFA). GFP-expressing cells were visualised using a fluorescence mirror unit (model U-MWIB3) consisting of a 460-495 nm bandpass exciter, a 505 nm longpass dichroic mirror and a 510 nm longpass emitter. An Olympus digital camera (model DP70) was used to view and capture images when required.

2.17.6 Regeneration of roots from seedling cuttings

Glass Agee pint jars were filled to 3/4 of their total volume with vermiculite that was moistened with sterile H₂O and covered with glass lids prior to autoclaving. Following sterilisation, the vermiculite was allowed to dry at room temperature for around 2 days before enrichment with Hornum’s solution (Márquez *et al.*, 2005) (components of
Hornum’s solution are presented in the Appendix). The volume of Hornum’s solution was variable due to small variability in jar sizes, but generally Hornum’s solution was added until the meniscus of the liquid barely covered the top most vermiculite layer. Seedlings chosen for regeneration were removed from tubes using flame sterilised tweezers and their roots removed just below the cotyledon with a flame sterilised scalpel blade before the seedling was placed firmly into the vermiculite of the jars and incubated in the plant room conditions. After around 12 days incubation, seedlings that had produced new roots were carefully transferred onto a slant of Jensen’s Agar in an 18 mm test tube and inoculated with 100 µL of relevant bacterial suspension from fresh G/RDM plates. The newly inoculated seedlings were then incubated as per normal in plant room conditions and monitored after 2 weeks incubation.
3 Symbiotic genes unique to NZP2037 and their effects on the nodulation of *L. pedunculatus* and *L. japonicus* Gifu
3.1 Introduction

The DNA sequence of the symbiosis island of NZP2037 (representative of broad host range \textit{M. loti}) was compared to that of R7A (representative of narrow host range) and potential symbiotic genes that were unique to NZP2037 and therefore likely to play a role in host-specificity were identified (Kasai-Mita \textit{et al.}, 2013, Kelly \textit{et al.}, 2014a). A list of these genes is located in the Appendix. The genes identified via this approach with probable \textit{nod} boxes were \textit{nodFEG}, \textit{nodA2}, \textit{nodU} and \textit{nodO}. \textit{NodFEG} and \textit{NodA2} are involved in the synthesis and transfer of a putative unsaturated fatty acid side chain to the non-reducing end of the Nod factor (Spaink & Sheeley, 1991, Demont \textit{et al.}, 1993, Ritsema \textit{et al.}, 1996) and the presence of this unsaturated fatty acid moiety on the Nod factor is likely to be involved in host-specificity (Spaink \textit{et al.}, 1989, Spaink, 1995). \textit{NodU} is responsible for the addition of a 6-\textit{O}-carbomyl group to the non-reducing end of the Nod factor structure (Jabbouri \textit{et al.}, 1995). \textit{NodO} is a cation-binding protein that is involved in the transport of cations across lipid bilayers and is a determinant of nodulation in specific mutant backgrounds of \textit{R. leguminosarum} bv. \textit{viciae} (Economou \textit{et al.}, 1994, Sutton \textit{et al.}, 1994). Experiments pertaining to \textit{nodO} are covered in Chapter 4.

Previous experiments in the Ronson laboratory had shown that when \textit{nodU}, \textit{nodFEG} and \textit{nodA2} were individually disrupted, no significant effects on the nodulation ability of NZP2037 on \textit{L. japonicus} Gifu or \textit{L. pedunculatus} were observed (Fowler, 2013). This result suggested that these genes, by themselves, were not essential for the extended host-range of NZP2037. A hypothesis was proposed that these genes may function via the same signalling pathways, given that these genes were all involved in Nod factor modification (specifically, they modify the terminal GlcNAc of the non-reducing end). Therefore, the disruption of one gene by itself may not have observable effects as other unique
modifications may compensate for the disruption.

In R7A, production of truncated EPS via the disruption of the gene exoU leads to the inability of the mutant to form effective nitrogen fixing nodules on L. japonicus Gifu. It was also shown that defective EPS was recognised via the plant receptor EPR3 (Kawaharada et al., 2015). In contrast, early on in this PhD project it was observed that when truncated EPS was produced by NZP2037 due to the same exoU mutation, the mutant exhibited a delay in nodulation but retained the ability to form nodules on L. japonicus Gifu, suggesting that NZP2037 was likely to encode gene(s) that allowed for the circumvention of the negative signalling effect of defective EPS. These observations lead to an additional aim to identify the gene(s) unique to NZP2037 that are involved in the strain’s ability to circumvent defective EPS signalling.

In this chapter, the hypothesis that modifications to the Nod factor may have functional redundancy was pursued further. In addition, further genes that were identified during this project are detailed, and their effects on the host-range of NZP2037 on L. pedunculatus and L. japonicus Gifu are documented. The host range conferred by the symbiosis island of NZP2037 when in an R7A genomic background that lacked a symbiosis island was also explored. The genes unique to the NZP2037 island were also investigated for their ability to circumvent the negative effects of defective EPS on L. japonicus Gifu. Lastly, the potential role of EPS signalling in the nodulation of L. pedunculatus was also investigated. The results gathered from all these investigations resulted in the conclusion that the unique genes of NZP2037 documented in this section were not essential for establishment of a symbiotic relationship with the extended host L. pedunculatus, either alone or in combination, nor were they essential for the circumvention of defective EPS
signalling. However, the results did provide insights into the symbiotic involvement of
\textit{nodU}, as well as confirmation of a lack of signalling by defective EPS in \textit{L. pedunculatus}.

3.2 NZP2037 and circumvention of defective EPS signalling on \textit{L. japonicus} Gifu

In R7A, mutants disrupted in \textit{exoU} produce a truncated form of EPS and these mutants exhibit an inability to form effective (nitrogen-fixing) nodules, but instead induce uninfected nodule primordia on \textit{L. japonicus} Gifu due to signalling through EPR3 (Kelly \textit{et al.}, 2013, Kawaharada \textit{et al.}, 2015). In contrast, mutants of \textit{M. loti} NZP2037 disrupted in \textit{exoU} retained the ability to form nodules on \textit{L. japonicus} Gifu after a delay (Figure 3.1A), and induced primordia that eventually developed into nodules, represented by a plateau and decrease in primordia over time (Figure 3.1B). These results suggest that NZP2037 encoded unique gene(s) that allowed NZP2037 to circumvent defective EPS signalling.

3.3 The NZP2037 unique genes: \textit{nodFEG}, \textit{nodA2}, \textit{nodU}

As noted above, single markerless deletion mutants of the genes unique to the NZP2037 island (\textit{nodU}, \textit{nodFEG} and \textit{nodA2}) had no significant effects on the nodulation of \textit{L. pedunculatus} and \textit{L. japonicus} Gifu ((Fowler (2013), John Sullivan, personal communication). It was hypothesised that the genes disrupted may be involved in the circumvention of defective EPS signalling in \textit{L. japonicus} Gifu, compensating for or masking potential EPS signalling that may occur through EPR3. Therefore, double \textit{nod exoU} mutants were constructed and assayed for their symbiotic properties on \textit{L. japonicus} Gifu and \textit{L. pedunculatus}.
Figure 3.1. Effect of R7A and NZP2037 *exoU* mutants on nodulation of *L. japonicus* Gifu. A) Percentage of plants that formed nodules over time. *p* values for R7A vs R7A *ΔexoU*, and NZP2037 vs NZP2037*ΔexoU* are both <0.05. B) Average number of primordia formed over time. *p* values for R7A vs R7A *ΔexoU* and NZP2037 vs NZP2037*ΔexoU* are both <0.05. Strains used as inocula are indicated in the key. Results shown here were calculated from an average across three separate experiments, each consisting of a sample size of 15 plants. Error bars represent the standard error of the mean. Note that a decrease in primordia numbers over time is indicative of primordia developing into effective nodules.
3.3.1 Construction and nodulation phenotypes of *nod exoU* double mutants

To test the hypothesis that *nodU*, *nodFEG* and *nodA2* were involved in the ability of NZP2037 to circumvent the negative symbiotic effects of truncated EPS, IDMs of *exoU* were made in the strains NZP2037 Δ*nodU*, NZP2037 Δ*nodFEG*, and NZP2037 Δ*nodA2* using the suicide vector pFUS-exoU (methods as detailed in section 2.15.2). The introduction of the vector into the gene generates an insertion mutation that inactivates the gene of interest (*exoU*). The resultant strains NZP2037 Δ*nodU exoU::pFUS2*, NZP2037 Δ*nodFEG exoU::pFUS2*, and NZP2037 Δ*nodA2 exoU::pFUS2* (collectively referred to as ‘*exoU* double mutants’) were used as inocula for plant studies on *L. pedunculatus* and *L. japonicus* Gifu (Figure 3.2).

The *exoU* double mutants nodulated *L. pedunculatus* with similar kinetics to the wildtype and the single mutant NZP2037 Δ*exoU* controls (Figure 3.2A), suggesting that individually, the genes *nodU*, *nodFEG* and *nodA2* were not essential for the nodulation of *L. pedunculatus* in the presence of defective EPS.
Figure 3.2. The nodulation kinetics of \textit{exoU} double mutants on A) \textit{L. pedunculatus} and B) \textit{L. japonicus Gifu}. Strains used as inocula are indicated in the key. Strains denoted as Δ\textit{exoU} are markerless deletion mutants while \textit{exoU}:pFUS2 strains are IDM mutants that had inactivated \textit{exoU} via insertion of the construct. Percentage of plants was calculated from an experiment consisting of 15 plants per strain and is representative of the two experiments carried out.
When the *exoU* double mutants were used as inocula for *L. japonicus* Gifu, a delay in nodulation comparable to the single mutant NZP2037 Δ*exoU* control was observed for all the mutants tested, with the exception of the *nodU exoU* double mutant which did not form nodules in the time period tested (Figure 3.2B). Disruption of the *nodA2* and *nodFEG* appeared to slightly benefit nodulation efficiency, but restoration of the nodulation phenotype to wildtype levels was not observed. This suggested that *nodA2* and *nodFEG* were not major determinants in the circumvention of defective EPS signalling. However, the results suggested that *nodU* was likely to be a determinant in the circumvention of the effect of defective EPS on this host. Note that the double mutant still induced nodule primordia (see Section 3.3.3 below).

### 3.3.2 Further characterization of NZP2037 Δ*nodU exoU::pFUS2*

Following the observation that the NZP2037 Δ*nodU exoU::pFUS2* double mutant exhibited a defective nodulation phenotype on *L. japonicus* Gifu, the mutant was cultured in G/RDM broth in an attempt to freeze a cell culture for preservation. During incubation, it was observed that the mutant grew poorly compared to mutants that only had a single *exoU* mutation (Figure 3.3). The decreased growth of the mutant was only observed in broth culture and not observed when the mutant was incubated on solid G/RDM medium. The decreased growth of the mutant was unexpected, as genes involved in Nod factor synthesis are unlikely to be expressed in broth culture. One possibility was that the mutant may have undergone a secondary mutation that had affected the growth rate of this strain, and that this may have compromised the nodulation experiments. Therefore, the strain NZP2037 Δ*nodU exoU::pFUS2* was re-made following the same procedure as described in Section 3.3.1.
The ‘new’ NZP2037 ΔnodU exoU::pFUS2 mutant exhibited a growth phenotype in broth culture that was comparable to the single mutant control NZP2037 ΔexoU (data not shown). This ‘normal’ growth phenotype NZP2037 ΔnodU exoU::pFUS2 double mutant was then used as an inoculum on L. pedunculatus (Figure 3.4A) and L. japonicus Gifu (Figure 3.4B). The new NZP2037 ΔnodU exoU::pFUS2 mutant exhibited a defective nodulation phenotype comparable to the previously described ‘slow’ nodU exoU double mutant. This confirmed that the nodU gene unique to NZP2037 is likely involved in the circumvention of effect of defective EPS signalling that would otherwise hinder an effective, but delayed, nodulation phenotype on L. japonicus Gifu.

![Figure 3.3. Growth of NZP2037 ΔnodU exoU::pFUS2 in G/RDM broth culture.](image)

The strain NZP2037 ΔnodU exoU::pFUS2 (left) grows to a lower density in broth culture than the control NZP2037 ΔexoU strain (right). Broth cultures shown here were inoculated with a colony from fresh plates using a toothpick and incubated at 28°C for 3 days with shaking. The experiment was repeated 3 times and the NZP2037 ΔnodU exoU::pFUS2 broths never became dense in the 3 days of incubation.
Figure 3.4 The nodulation phenotype of a nodU exoU double mutant that had a ‘normal’ growth phenotype on A) L. pedunculatus and B) L. japonicus Gifu. Strains used as inocula are indicated in the key. The percentage of plants was calculated from an experiment consisting of 15 plants per strain that was representative of the two experiments carried out.
3.3.3 Complementation of nodU and exoU mutations

To confirm that the defective nodulation phenotype of the NZP2037 nodU exoU double mutant was due to the combined effect of the two mutations, complementation experiments were carried out. The nodU mutation was complemented using the plasmid pPROBEKT-nodU (Table 3), while the exoU mutation was complemented with the plasmid pFAJ1700-exoU (Table 3).

Each of the two plasmids were introduced into NZP2037 ΔnodU exoU::pFUS2 by conjugation and the resultant strains were used as inocula on L. japonicus Gifu. Unexpectedly, the complementation of NZP2037 ΔnodU exoU::pFUS2 with the vector pPROBEKT-nodU was unsuccessful (Figure 3.5A). In contrast, the defective nodulation phenotype of NZP2037 ΔnodU exoU::pFUS2 was complemented with pFAJ1700-exoU (Figure 3.5A, brown line), suggesting that the defective nodulation phenotype was only observed when an exoU mutation was present.

As complementation with pPROBEKT-nodU was unsuccessful, it was hypothesised that this was due to a negative effect of the expression of NodU within the exoU mutant background. To test the hypothesis, the empty vector pPROBEKT and the vector pPROBEKT-nodU were introduced into NZP2037 ΔexoU to generate the strains NZP2037 ΔexoU pPROBEKT and NZP2037 pPROBEKT-nodU, respectively. These strains were then used as inocula on L. japonicus Gifu and the nodulation of these plants were observed (Figure 3.5B).
Figure 3.5. Attempted complementation of select exoU mutants on *L. japonicus* Gifu showed that exoU complementation restored wildtype phenotype but nodU did not. A) Co-complementation of the nodU exoU double mutant with pPROBEKT-nodU and pFAJ1700-exoU. B) Effect of pPROBEKT-nodU in the exoU single mutant. Strains used as inocula are indicated by the key. Percentage of plants nodulated was calculated from an experiment that consisted of 7 plants per strain that was representative of the two experiments carried out.
The results showed that strains that harboured the empty vector retained a nodulation phenotype that was comparable to the single NZP2037 ΔexoU mutant control. In contrast, when the vector pPROBEKT-nodU was present, the strain exhibited a severely delayed nodulation phenotype of around 7 days but still retained the ability to induce nodules in the presence of defective EPS. These results suggest that the inability to restore the nodulation phenotype to the nodU exoU double mutant, was likely due to the burden of nodU expression from pPROBEKT-nodU that resulted in a delay in nodulation.

Although the presence of the pPROBEKT-nodU did not restore the nodulation phenotype of the nodU exoU double mutant, an interesting phenotype was still observed. In strains that expressed nodU within a disrupted exoU background, double the average number of primordia were induced. At 31 days post inoculation, the average primordia numbers of strains that harboured pPROBEKT-nodU was $10 \pm 1$, which contrasted with that of $5 \pm 1$ in strains that did not harbour pPROBEKT-nodU (Figure 3.6).

The nitrogen fixation phenotypes of the strains reflected their kinetics of nodulation. Plants that were inoculated with NZP2037 strains disrupted in exoU exhibited yellow leaves and a dwarfed appearance akin to the phenotype of an uninoculated control. In contrast, plants that were inoculated with strains that produced functional EPS (wildtype NZP2037, NZP2037 ΔnodU, NZP2037 ΔnodU pPROBEKT-nodU, and NZP2037 ΔnodU exoU::pFUS2 pFAJ1700-exoU) exhibited green leaves indicative of nitrogen fixation (Figure 3.7A).
Figure 3.6. Average number of primordia induced is increased in \textit{exoU} mutants that harboured pPROBEKT-nodU on \textit{L. japonicus} Gifu. Average number of primordia was calculated from an experiment consisting of 7 plants per strain that was representative of the two experiments carried out. Error bars represent the standard error of the mean. \textit{p} values = <0.05. Note that a decrease in primordia is observed when primordia develop into effective nodules.

As noted previously, the NZP2037 \textit{ΔnodU exoU::pFUS2} double mutant was only able to form non-infected primordia (Figure 3.7B). When the vector pPROBEKT-nodU was present in NZP2037 \textit{ΔnodU exoU::pFUS2}, an increase in the number of primordia was observed, but rescue of a nodulation phenotype was not achieved (Figure 3.7C). In contrast, a single \textit{exoU} mutant retained the ability to form infected nodules (Figure 3.7D). Taken together, the results suggest that \textit{nodU} is involved in the promotion of symbiosis of \textit{L. japonicus} Gifu, presumably in organogenesis of the root-nodule structure but is not essential for the circumvention of defective EPS signalling observed in wildtype NZP2037.
Figure 3.7. *L. japonicus* Gifu inoculated with *nodU* mutants of NZP2037. A) Nitrogen fixation phenotypes of plants inoculated with: U = uninoculated, WT = wildtype NZP2037, *nodU* = NZP2037 Δ*nodU*, *nodU exoU* = NZP2037 Δ*nodU exoU*, *nodU* complement = NZP2037 Δ*nodU* harbouring the vector pPROBEKT-*nodU*, *nodU exoU* + *nodU* complement = NZP2037 Δ*nodU exoU*::pFUS2 harbouring the vector pPROBEKT-*nodU*, *nodU exoU* + *exoU* complement = NZP2037 Δ*nodU exoU*::pFUS2 harbouring the vector pFAJ1700-*exoU*, *exoU* = NZP2037 Δ*exoU*. B) Close up of *nodU exoU* roots. C) Close up of *nodU exoU* + *nodU* complement roots D) Close up of *exoU*. Red arrows indicate primordia and/or nodules. Photos were taken at 6 weeks post inoculation.
3.4 Effect of *nodU* in a R7A genetic background

Strains that harboured pPROBEKT-nodU induced an increased average number of primordia, suggesting that NodU was likely to have an effect in the establishment of symbiosis. It seemed possible that the heterologous expression of *nodU* may rescue the defective nodulation phenotype that is characteristic of the R7A Δ*exoU* mutant. To test this hypothesis the vector pPROBEKT-nodU was transferred into the wildtype R7A and the R7A Δ*exoU* mutant and the strain tested on *L. japonicus* Gifu. No effect of the plasmid expressing *nodU* was observed (Figure 3.8A and Figure 3.9A)

However, the heterologous expression of *nodU* in R7A Δ*exoU* promoted the formation of nodule primordia in comparison to the R7A Δ*exoU* control (Figure 3.8B, Figure 3.9B, C), a similar phenotype to that observed in the NZP2037 Δ*exoU* background. Taken together, the results suggest that *nodU* was involved in organogenesis, but by itself was unable to overcome the negative signalling that occurs when defective EPS was present in the R7A genomic background.
Figure 3.8. The heterologous expression of nodU from NZP2037 in trans in the mutant R7A ΔexoU was unable to restore effective nodulation of L. japonicus Gifu, but increases the number of primordia formed. A) Percentage of plants nodulated over time B) Average number of primordia per plant formed over time. Strains used as inocula are indicated in the key. Error bars represent the standard error of the mean. $p$ value = <0.05. Results shown were calculated from an experiment consisting of 7 plants per strain and are representative of the two experiments carried out.
Figure 3.9 *L. japonicus* Gifu inoculated with R7A harbouring a plasmid containing *nodU* from NZP2037. A) Nitrogen fixation phenotypes of plants inoculated with: WT = wildtype R7A, $exoU = \text{R7A}\Delta exoU$, $exoU + nodU$ complement = R7A $\Delta exoU$ harbouring the vector pPROBEKT-nodU, WT + nodU complement = R7A harbouring the vector pPROBEKT-nodU. B) Close up of roots on the plant inoculated with the strain R7A $\Delta exoU$. C) Close up of roots on the plant inoculated with the strain R7A $\Delta exoU$ pPROBEKT-nodU. Red arrows highlight uninfected primordia. Photos taken at 6 weeks post inoculation.
3.5 The transconjugant R7A<sup>C</sup>:NZP2037<sup>I</sup>

At the start of this PhD project it was known that the symbiosis island of NZP2037 was fragmented within the genome (Kasai-Maita et al., 2013), leading to the assumption it was non-mobile. However during the PhD it was discovered that the symbiosis island was a tripartite island that was transferable as a single entity to non-symbiotic strains of <i>M. loti</i> (Haskett et al., 2016). Therefore, the NZP2037 symbiosis island was mobilized into a strain of <i>M. loti</i> R7A that had lost its symbiosis island (designated as R7A NS (R7A non-sym)) This generated the transconjugant strain R7A<sup>C</sup>:NZP2037<sup>I</sup> (R7A<sup>Chromosome</sup>:NZP2037<sup>island</sup>; mating performed by Dr. John Sullivan). The transconjugant was sequence verified via genome sequencing performed by Dr. J. P. Ramsay and his group at Curtin University (Accession number LZTH00000000; (Haskett et al., 2016)). A small-scale plant experiment using R7A<sup>C</sup>:NZP2037<sup>I</sup> found that the transconjugant was symbiotically proficient on both <i>L. japonicus</i> Gifu and <i>L. pedunculatus</i> (John Sullivan, personal communication).

The transconjugant strain was then handed to me for detailed experiments. The nodulation phenotypes were replicated on <i>L. japonicus</i> Gifu (Figure 3.10A) and <i>L. pedunculatus</i> (Figure 3.10B). Interestingly, R7A<sup>C</sup>:NZP2037<sup>I</sup> was a more proficient symbiont on <i>L. japonicus</i> Gifu than the wildtype NZP2037, with a faster onset of nodulation and a greater average number of nodules formed on this host (Figure 3.10C). No differences in nodule number or nodulation kinetics were observed for <i>L. pedunculatus</i> (Figure 3.10D).

Overall, the transfer of the NZP2037 symbiosis island into the R7A genomic background expanded the host-range of the strain to include the previously incompatible host <i>L. pedunculatus</i>. This result indicates that the chromosomal genes of NZP2037 are not
Figure 3.10 The nodulation phenotype of the transconjugant strain R7A<sup>C</sup>:NZP2037<sup>I</sup>. A) and B) nodulation kinetics on \emph{L. japonicus} Gifu and \emph{L. pedunculatus}, respectively. C) and D) the average number of nodules and primordia formed on \emph{L. japonicus} Gifu and \emph{L. pedunculatus}, respectively. Strains used are as indicated by the key. The average number of nodules and primordia formed by NZP2037 are shown in filled black bars and unfilled black bars, respectively. Similarly, the average number of nodules and primordia formed by R7A<sup>C</sup>:NZP2037<sup>I</sup> are shown in red filled bars and unfilled red bars, respectively. Stacked bars are used to show the number of primordia and nodules concurrently. The percentage of plants and average number of nodules/primordia shown here was calculated from 3 separate experiments, each comprised of 15 plants. Error bars represent the standard error of the mean. \(p\) values are as follows for: A) R7A vs R7A<sup>C</sup>:NZP2037<sup>I</sup> = 0.0154, NZP2037 vs R7A<sup>C</sup>:NZP2037<sup>I</sup> = 0.0372, NZP2037 vs R7A = 0.007. B) R7A vs R7A<sup>C</sup>:NZP2037<sup>I</sup> = <0.001, NZP2037 vs R7A<sup>C</sup>:NZP2037<sup>I</sup> = 0.4880, NZP2037 vs R7A <0.001. C) <0.001. D) 0.6036.
required for its expanded host range, but rather genes located within the symbiosis island determine the host-range. Furthermore, as R7A\textsuperscript{C}:NZP2037\textsuperscript{I} was a more efficient symbiont than the wildtype NZP2037 strain on \textit{L. japonicus} Gifu, it was decided to use R7A\textsuperscript{C}:NZP2037\textsuperscript{I} as the main strain for future host range studies.

### 3.5.1 R7A\textsuperscript{C}:NZP2037\textsuperscript{I} exo\textit{U}::pFUS2 does not exhibit a delayed nodulation phenotype on \textit{L. japonicus} Gifu

Transfer of the NZP2037 symbiosis island into a R7A genomic background generated a transconjugant strain with a host-range similar to that of NZP2037. It was therefore of interest to determine whether an \textit{exo\textit{U}} mutant of the transconjugant strain displayed the R7A \textit{exo\textit{U}} (uninfected primordia) or NZP2037 \textit{exo\textit{U}} (delayed but effective nodulation) phenotype. To investigate this, strain R7A\textsuperscript{C}:2037\textsuperscript{I} \textit{exo\textit{U}}::pFUS2 was constructed (method detailed in section 3.3.1). The strain retained the ability to nodulate \textit{L. japonicus} Gifu (Figure 3.11). Interestingly, the delay in nodulation that was observed for NZP2037 \textit{Δexo\textit{U}} was not replicated in the R7A\textsuperscript{C}:2037\textsuperscript{I} background and this supported the previous observation that R7A\textsuperscript{C}:NZP2037\textsuperscript{I} was a more robust symbiont than the wildtype NZP2037 for use in nodulation assays. Overall, these results show that genes on the NZP2037 symbiosis island are able to circumvent the negative symbiotic effects of truncated EPS caused by the \textit{exo\textit{U}} mutation.
Figure 3.11. The presence of truncated EPS in the R7A\textsuperscript{C}:NZP2037\textsuperscript{l} transconjugant background does not interfere with its ability to nodulate \textit{L. japonicus} Gifu. Strains used as inocula are indicated by the key. Percentage of plants was calculated from 15 plants per strain and are representative of the two experiments carried out.

3.5.2 R7A\textsuperscript{C}:NZP2037\textsuperscript{l} ΔnodU and R7A\textsuperscript{C}:NZP2037\textsuperscript{l} ΔnodU exoU::pFUS2

It was of interest to see if transfer of the \textit{nodU} disrupted island into R7A NS would replicate the nodulation phenotypes observed for NZP2037 and NZP2037 ΔnodU exoU::pFUS2 (as described in section 3.3). To test this, the NZP2037 ΔnodU symbiosis island was transferred into R7ANS (method detailed in section 2.15.3), and in addition, an IDM \textit{exoU} mutation was introduced into the new transconjugant.

Successful transconjugants were confirmed by PCR of each of the three island regions (Figure 3.12). The successful transconjugant was designated as R7A\textsuperscript{C}:NZP2037\textsuperscript{l} ΔnodU. The suicide vector pFUS-exoU was transferred into R7A\textsuperscript{C}:NZP2037\textsuperscript{l} ΔnodU and the resultant mutant was designated as R7A\textsuperscript{C}:NZP2037\textsuperscript{l} ΔnodU exoU::pFUS2.
Figure 3.12. PCR confirmation of a successful transconjugant. The ladder presented here is HindIII digested λ phage and HaeIII digested φX174, relevant band sizes (in bp) are indicated to the left of the ladder. PCR products from different backgrounds are shown here, R7A and NZP2037 are negative controls that do not form PCR products with the primer pairs used. The numbers 1, 2 and 3 represent the primer pairs 6.2islandcheckL and 6.2islandcheckR, 27.5islandcheckL and 27.5islandcheckR, and 52.8islandcheckR and downstreamisland (Primers listed in Table 5), respectively. PCR products for a successful transconjugant using the primer pairs indicated above for 1, 2, 3 are 700 bp, 454 bp, and 832 bp, respectively.
3.5.2.1 Nodulation phenotypes of $R7A^C$:NZP2037I $\Delta$nodU and $R7A^C$:2037I $\Delta$nodU exoU::pFUS2 on *L. pedunculatus* and *L. japonicus* Gifu

The strains $R7A^C$:NZP2037I $\Delta$nodU and $R7A^C$:NZP2037I $\Delta$nodU exoU::pFUS2 exhibited nodulation phenotypes on *L. japonicus* Gifu and *L. pedunculatus* that were comparable to the non-mutagenized $R7A^C$:NZP2037I on both hosts (Figure 3.13A and Figure 3.13B, respectively).

These results were unexpected as previously, in the NZP2037 background, a $nodU$ exoU double mutant was severely disrupted in its ability to form nodules on *L. japonicus* Gifu (section 3.3.2). Taken together, the results suggested that disruption of $nodU$ within the $R7A^C$:NZP2037I transconjugant background had no effect on the nodulation of *L. japonicus* Gifu and *L. pedunculatus* in the presence or absence of defective EPS.

![Figure 3.13](3.13.png)

Figure 3.13. Nodulation kinetics of strains with a $nodU$ and/or exoU mutation in an $R7A^C$:NZP2037I transconjugant background. A) *L. pedunculatus* and B) *L. japonicus* Gifu. Strains used as inocula are as indicated by the key. Percentage of plants nodulated was calculated from an experiment consisting of 15 plants per strain and is representative of the two experiments carried out.
3.6 Deletion of unique genes involved in modification of the non-reducing end of the Nod factor in the R7A<sup>C</sup>:NZP2037<sup>I</sup> background

Thus far, the experiments in this chapter have shown that a nodU exoU double mutant of NZP2037 was unable to form nodules on L. japonicus Gifu, but no effect was observed for L. pedunculatus. However, when the same nodU exoU gene disruptions were introduced into the R7A<sup>C</sup>:NZP2037<sup>I</sup> transconjugant background, the mutant retained the ability to nodulate L. japonicus Gifu. A hypothesis for this observation was that other modifications to the Nod factor unique to the NZP2037 island may compensate the nodU mutation within the R7A<sup>C</sup>:NZP2037<sup>I</sup> background and this may explain the retention of the ability to nodulate L. japonicus Gifu in the presence of defective EPS or perhaps be involved in the nodulation of the extended host L. pedunculatus.

The most attractive gene targets were the previously described nodFEG and nodA2 genes that are involved in the modification of the non-reducing end of the Nod-factor, a trait that is shared by nodU. Therefore, to test this hypothesis, disruptions in the genes nodFEG, nodU and nodA2 were introduced into a single strain in the R7A<sup>C</sup>:NZP2037<sup>I</sup> background. An exoU mutation was also introduced to these strains.

3.6.1 Generation of the mutant JC01 (R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔnodUFEGA2)

A double mutant of the genes nodU and nodFEG in the NZP2037 genomic background (NZP2037 ΔnodUFEG) was available. This strain was used as the starting point and a nodA2 markerless deletion was introduced via a homologous recombination approach (detailed in section 2.15.1) using the vector pJQ200SK-nodA2 (detailed in Table 3). The loss of nodA2 was screened for using a PCR approach (Figure 3.14). Primers used are
listed in Table 5. The primer nodA2flankcheckL binds to DNA that flanks the 5’ region of nodA2 and when used with primer nodA2flankcheckR produces a PCR product of 1,162 bp in the wildtype and a PCR product of 572 bp in a ΔnodA2 mutant. The primer pair nodA2internalcheckL and nodA2flankcheckR produces a PCR product of 443 bp only in the wildtype background. The 572 bp PCR product was sent for Sanger sequencing to confirm the integrity of the mutant. The symbiosis island that was disrupted in the genes nodUFEGA2 was then transferred into the R7A genomic background (detailed in section 3.5.2) and the successful transconjugant strain was designated as JC01. Additionally, pFUS-exoU was transferred into JC01 to generate an IDM of exoU and this mutant was designated as JC02.

Figure 3.14. PCR screen for a NZP2037 ΔnodU ΔnodFEGA2 mutant. The wildtype control and mutant are shown on the left and right, respectively. Lanes marked 1 and 2 represent PCR reactions using the primer pairs nodA2flankcheckL and nodA2flankcheckR and nodA2internalcheckL and nodA2flankcheckR, respectively.
3.6.2 Nodulation phenotypes of JC01 and JC02 on *L. japonicus* Gifu and *L. pedunculatus*

The strains JC01 and JC02 exhibited phenotypes that were comparable to non-mutagenized transconjugant on *L. japonicus* Gifu and *L. pedunculatus* (Figure 3.15A and Figure 3.15B). These results suggest that in the R7A^C::NZP2037^I transconjugant background, the *nodU*, *nodFEG* and *nodA2* genes are not essential for the nodulation of *L. pedunculatus* and *L. japonicus* Gifu. In addition, these genes were not essential for the circumvention of defective EPS signalling on *L. japonicus* Gifu.

![Graph](image)

Figure 3.15. The nodulation phenotypes of JC01 and JC02. A) *L. japonicus* Gifu and B) *L. pedunculatus*. Strains used as inocula are as indicated by the key. JC01 and JC02 represent the strains R7A^C::NZP2037^I Δ*nodU*, Δ*nodFEG*, Δ*nodA2* and its *exoU::pFUS2* mutant equivalent, respectively. Percentage of plants nodulated was calculated from an experiment consisting of 7 plants per strain and is representative of the 2 experiments carried out.

3.7 A new target - Mln399, a putative glycosyl transferase

During the course of this study, a more thorough comparison of genes shared amongst strains of *M. loti* that belonged to the group with an extended host range was performed as part of a project that sequenced the genomes of several *M. loti* strains that were compatible with *L. pedunculatus* (NZP2037, NZP2014, NZP2042, SU343). Genes
predicted to be on the symbiosis islands of _L. pedunculatus_ compatible strains were compared to non-compatible strains (R7A and MAFF303099) resulting in a list of unique candidate genes that are present or absent from either group. Multiple unique genes belonging to the broad host range _M. loti_ were identified and non-hypothetical genes with _p_ value < 0.5 were investigated as potential genetic targets for host-specificity. From this comparative analysis, a potential unique gene designated as _mln399_, that was predicted to encode a putative glycosyl transferase was identified. _mln399_ is genetically located directly downstream of and in the same orientation as the _nodACIJ-nolO_ operon (Figure 3.16), suggesting that it could be expressed as part of the _nodACIJ-nolO_ operon and therefore under NodD regulation. Furthermore, its location suggests it may be involved in Nod factor synthesis. Therefore, the hypothesis was that _mln399_ encoded a protein involved in the Nod factor modifications, that may act as a determinant of host specificity and/or is involved in EPS circumvention in NZP2037 and the transconjugant.

To test this hypothesis, a markerless deletion of _mln399_ was constructed in the R7A<sup>C</sup>:NZP2037<sup>I</sup> background. Additionally, the introduction of an IDM of _exoU_ within the _mln399_ mutant was also performed.

**Figure 3.16. Genetic orientation of the putative _nodACIJ-nolO-mln399_ operon.** The operon is preceded by a _nod_ box (yellow) followed by _nodACIJ-nolO_ (blue). _mln399_ (red) is located 168 bp downstream of _nolO_.

106
3.7.1 Generation of mutants R7A<sup>C</sup>:NZP2037<sup>I</sup> Δmln399 and R7A<sup>C</sup>:NZP2037<sup>I</sup> Δmln399 exoU::pFUS2

PCR with the primers mln399_del_LL, mln399_del_LR, mln399_del_LL and mln399_del_LR (listed in Table 5) was used to amplify the flanking regions of DNA proximal to mln399. The amplified DNA from the PCR reactions were then ligated via Gibson cloning into the multi cloning site (MCS) of the suicide vector pJQ200SK. Successful ligation into pJQ200SK were screened via a restriction digest using the enzymes BamHI and ApaI with an expected product size of 2022 bp in successful ligations that had integrated into the MCS of the vector (Figure 3.17). Potential vectors were verified via Sanger sequencing and the verified vector was designated as pJQ200SK-mln399.

The plasmid pJQ200SK-mln399 was transferred into R7A<sup>C</sup>:NZP2037<sup>I</sup> via conjugation, and deletion of mln399 was performed (as detailed in section 3.6.1). Potential mutants were screened via PCR with the primer pairs mln399_checkL and mln399_checkR and mln399_checkL and mln399_internalCheckR (listed in Table 5). The successful mln399 markerless deletion mutant was confirmed using Sanger sequencing and designated as R7A<sup>C</sup>:NZP2037<sup>I</sup> Δmln399. The pFUS-exoU derivative was named R7A<sup>C</sup>:NZP2037<sup>I</sup> Δmln399 exoU::pFUS2.
Figure 3.17. Restriction digest of pJQ200SK-mln399 with the enzymes *Bam*HI and *Apa*I. L = ladder which is λ phage digested with *Hind*III, V = Vector that contains the expected insert size of 2022 bp after digestion with the enzymes *Bam*HI and *Apa*I.

### 3.7.2 Nodulation phenotypes of R7A\textsuperscript{C}:NZP2037\textsuperscript{I} Δmln399 and R7A\textsuperscript{C}:NZP2037\textsuperscript{I} Δmln399 exoU::pFUS2 on *L. japonicus* Gifu and *L. pedunculatus*

Both strains R7A\textsuperscript{C}:NZP2037\textsuperscript{I} Δmln399 and R7A\textsuperscript{C}:NZP2037\textsuperscript{I} Δmln399 exoU::pFUS2 exhibited a nodulation phenotype comparable to the non-mutagenized R7A\textsuperscript{C}:NZP2037\textsuperscript{I} on *L. japonicus* Gifu and *L. pedunculatus* (Figure 3.18A and Figure 3.18B). These results suggest that *mln*399 was not essential for nodulation of the hosts tested, nor was it essential in the circumvention of defective EPS signalling.
Figure 3.18. Nodulation phenotypes of R7A<sup>C</sup>:NZP2037<sup>I</sup> Δmln399 and R7A<sup>C</sup>:NZP2037<sup>I</sup> Δmln399 exoU::pFUS2. A) L. japonicus Gifu and B) L. pedunculatus. Strains used as inocula are as indicated in the key. Percentage of plants nodulated was calculated from an experiment consisting of 7 plants per strain and is representative of the 2 experiments carried out.

3.8 The Type IV secretion system

Another potential unique target within the NZP2037 symbiosis island was the T4SS. A T4SS is also present within R7A and this system is known to secrete the effector proteins Msi059 and Msi061 that are involved in the promotion and inhibition of nodulation on *L. corniculatus* and *L. filicaulis*, respectively (Hubber et al., 2004). The NZP2037 genome encodes two unique genes; *mln452* and *mln454* that are likely to encode effectors transported by the T4SS due to the presence of a predicted Type IV secretion signal motif at their 3’ ends (Kasai-Maita et al., 2013). Mln452 has an N-terminal domain that has a 73% amino acid identity to a T3SS effector protein in *Xanthomonas fuscans* subsp. *aurantifolii* while the C-terminal end shares homology to the secretion signal of T4SS (Kasai-Maita et al., 2013). Mln454 shares a high degree of similarity to the C-terminal end of Msi059 and this region of similarity is also present in the effector Mln6316 of the *M. loti* strain MAFF303099 (Kasai-Maita et al., 2013). A comparison of the genetic organization of the region encoding the T4SS and associated proteins in NZP2037 and R7A is shown in Figure 3.19.
The T4SS had previously been investigated by Kasai-Maita et al. (2013) and it was found that disruption of the T4SS in NZP2037 had no effect on the nodulation of a majority of Lotus hosts, including L. pedunculatus and L. japonicus. However, a hypothesis was that proteins secreted by the T4SS may only exhibit a phenotype when in the presence of other mutations, such as in a background that produced truncated EPS. Furthermore, a mutant that was disrupted in the T4SS may perform differently under our experimental conditions. To test this hypothesis, the disruption of the T4SS located within the more robust R7A<sup>C</sup>:NZP2037<sup>I</sup> background was performed and its nodulation phenotypes explored.

### 3.8.1 Generation of the mutant strains R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔTypeIV and R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔTypeIV exoU::pFUS2

The approach chosen to generate a T4SS mutant was to remove the whole secretion system instead of targeting single putative effector genes. The targets chosen for disruption were the structural genes virB1-11 and the regulatory gene virA that is located...
directly downstream of the *virB1-11* operon (Figure 3.19). Note that this was similar to what Kasai-Maita *et al.* (2013) had performed previously, except here *virA* was also targeted for deletion. The primers LL_TypeIVSS_del, LR_TypeIVSS_del, RL_TypeIVSS_del, and RR_TypeIVSS_del (Table 5) were used to amplify DNA that flanked the ends of *virB1* and *virA*, respectively. The PCR products were ligated into the vector pJQ200SK via Gibson cloning and successful ligation was verified via restriction digest (method described in section 3.7.1). Potential clones were confirmed using Sanger sequencing and correct plasmid was designated as pJQ200SK-TypeIV.

Plasmid pJQ200SK-TypeIV was transferred into R7A<sup>C</sup>:NZP2037<sup>I</sup> and deletion of the genes *virA* and *virB1-11* was performed as described in section 3.6.1. Potential markerless deletion mutants were verified via PCR with the primers TypeIV_checkL, TypeIV_checkR, and TypeIV_Internal checkR (Table 5). A markerless deletion mutant of the T4SS was confirmed using Sanger sequencing and designated as R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔTypeIV. It’s pFUS-*exoU* derivative was termed R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔTypeIV *exoU*::pFUS2.

### 3.8.2 Nodulation phenotypes of R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔTypeIV and R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔTypeIV *exoU*::pFUS2 on *L. japonicus* Gifu and *L. pedunculatus*

Both mutants R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔTypeIV and R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔTypeIV *exoU*::pFUS2 exhibited a nodulation phenotype that was comparable to the non-mutagenized R7A<sup>C</sup>:NZP2037<sup>I</sup> transconjugant on *L. japonicus* Gifu and *L. pedunculatus* (Figure 3.20A and Figure 3.20B, respectively). The results suggest that the T4SS was not essential for nodulation of both hosts that were tested, nor was it involved in the circumvention of
defective EPS.

![Graph](image)

**Figure 3.20.** Nodulation phenotype of \( \text{R7A}^c: \text{NZP2037}^l \Delta \text{TypeIV} \) and \( \text{R7A}^c: \text{NZP2037}^l \Delta \text{TypeIV} \ exoU::pFUS2 \). A) \( L. \ japonicus \) Gifu and B) \( L. \ pedunculatus \). Strains used as inocula are as indicated by the key. Percentage of plants nodulated was calculated from an experiment consisting of 7 plants per strain and is representative of 2 the experiments carried out.

### 3.9 Deletion of \( nodUFEGA2 \) and \( mln399/\text{Type IV secretion system} \) within a single strain

Thus far, the mutant strains \( \text{R7A}^c: \text{NZP2037}^l \Delta \text{nodUFEGA2} \), \( \text{R7A}^c: \text{NZP2037}^l \Delta \text{mln399} \) and \( \text{R7A}^c: \text{NZP2037}^l \Delta \text{TypeIV} \) in the presence or absence of an \( exoU \) mutation exhibited a phenotype that was comparable to the non-mutagenized \( \text{R7A}^c: \text{NZP2037}^l \) transconjugant. The hypothesis for these observations was that the deletion of these genes was tolerated in the \( \text{R7A}^c: \text{NZP2037}^l \) transconjugant background as other genes unique to the island may compensate for these deletions. It was decided that the unique gene(s) that had been identified thus far would be targeted for mutagenesis in a cumulative fashion, potentially concluding with a mutant that had lost all the genes of interest presented thus far.
3.9.1 Generation of mutants JC03 and JC04

The suicide vector pJQ200SK-mln399 (described in section 3.7) was used to delete  
mln399 from the strain JC01 using the same methods previously described in section 3.7.1. 
The successful markerless deletion of mln399 was confirmed and the mutant with the 
genotype R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔnodUFEGA2, Δmln399 was designated as JC03. Disruption 
of exoU was achieved via the vector pFUS-exoU to generate the mutant JC04.

3.9.2 Generation of mutants JC05 and JC06

The suicide plasmid pJQ200SK-TypeIV (described in section 3.8) was transformed into 
the mutant JC01 to generate strain JC05 (R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔnodUFEGA2, ΔvirB1-11, 
ΔvirA). Disruption of exoU in JC05 via pFUS-exoU was used to generate the mutant JC06.

3.9.3 Nodulation phenotypes of JC03, JC04, JC05, and JC06 on L. japonicus Gifu and L. pedunculatus

Mutants JC03, JC04, JC05 and JC06 exhibited nodulation phenotypes that were 
comparable to the non-mutagenized R7A<sup>C</sup>:NZP2037<sup>I</sup> transconjugant on L. japonicus Gifu 
and L. pedunculatus (Figure 3.21A and Figure 3.21B, respectively), indicating that the 
combination of these various mutations did not impact nodulation of these hosts.

3.10 Continuation of markerless deletions targets unique to the 
NZP2037 island in the R7A<sup>C</sup>:NZP2037<sup>I</sup> transconjugant 
background

All mutants tested thus far within the R7A<sup>C</sup>:NZP2037<sup>I</sup> transconjugant background had 
phenotypes comparable to a non-mutagenized transconjugant on both L. japonicus Gifu 
and L. pedunculatus. As the gene(s) responsible for the host range of NZP2037 and the
ability to circumvent defective EPS were encoded on the symbiosis island, it was decided that the approach of mutagenesis of genes unique to NZP2037 that have been identified thus far was worth continuing, as relatively few genes were left to test.

Figure 3.21. Nodulation phenotypes for JC03, JC04, JC05 and JC06. A) *L. japonicus* Gifu and B) *L. pedunculatus*. Strains used as inocula are indicated by the key. JC03 and JC04 represents the strains R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔnodUFEGA2, Δmln399 and its respective exoU::pFUS2 mutant equivalent. JC05 and JC06 represents the strains R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔnodUFEGA2 ΔTypeIV and its respective exoU::pFUS2 mutant equivalent. Percentage of plants nodulated was calculated from an experiment consisting of 7 plants per strain and is representative of the two experiments carried out.

3.10.1 Generation of mutants JC07 and JC08

The strain JC03 (described in section 3.9.1) was used as a base the introduction of markerless deletion of the T4SS into JC03 to give strain JC07. Disruption of *exoU* was achieved in JC07 using pFUS-*exoU* and this gave the mutant JC08.
3.10.2 Verification of JC07

The time that had been invested into generating the mutant JC07 was extensive, therefore, it was imperative to confirm the genotype of the mutant. Genomic DNA from JC07 sent for whole genome Illumina sequencing by Mr. DNA (Molecular Research LP, Shallowater, Texas USA). Reads were aligned to a complete NZP2037 island DNA sequence with a coverage of 352 x that of a presumed transconjugant using Geneious as described in section 2.14. The alignment confirmed that JC07 had lost all genes that had been targeted for deletion (Figure 3.22), supporting the validity of the mutants presented thus far.

3.10.3 Nodulation phenotypes of JC07 and JC08 on *L. japonicus* Gifu and *L. pedunculatus*

The mutants JC07 and JC08 exhibited a nodulation phenotype that was comparable to the non-mutagenized R7A<sup>C</sup>:NZP2037<sup>l</sup> transconjugant on *L. japonicus* Gifu and *L. pedunculatus* (Figure 3.23A and Figure 3.23B, respectively). These results suggest that disruption of the genes *nodU, nodFEG, nodA2, mln399, virB1-11* and *virA* together in combination within a single mutant strain, in the presence or absence of an IDM of *exoU* were not essential for the nodulation of *L. japonicus* Gifu or *L. pedunculatus* nor were the genes essential for the circumvention of defective EPS within the R7A<sup>C</sup>:NZP2037<sup>l</sup> transconjugant background.
**Figure 3.22. Reads obtained from whole genome sequencing do not align to deleted genes.** Genes and their coding DNA sequences are shown in green and yellow, respectively. From top to bottom: Genes that were deleted: *nodA2, nodFEG, mln399*, and The T4SS (*virA* and *virB1-11*). The first 4 reads that were aligned are shown here as black bars below their respective genes. When no reads could be aligned, this results in a blank space within the alignment. Note that some genes were annotated with a different gene name (e.g. *nodU* is annotated as *tobZ* which shares amino acid homology, due to predicted carbamoyl transferase activity). Although the genes *virB1-11* and *virA* were successfully deleted, the whole operon is not displayed here and was trimmed (*virB5* to the end of *virB11* are not shown) for better visualisation.
117

Figure 3.23. Nodulation phenotypes of JC07 and JC08. A) *L. japonicus* Gifu and B) *L. pedunculatus*. Strains used as inocula are indicated by the key. JC07 and JC08 represents the strains R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔnodUFEGA2, Δmln399, ΔTypeIV and its respective exoU::pFUS2 mutant equivalent. Percentage of plants nodulated was calculated from an experiment consisting of 7 plants per strain and is representative of the two experiments carried out.

3.11 Introduction of ‘nodO-9.3kb’ into JC07

Of all the potential targets that were identified during this PhD project, only one gene target remained, *nodO*. The experiments detailing the *nodO* deletion are described in Chapter 4. Briefly, the deletion of a 9.3-kb region of DNA proximal and including *nodO* within NZP2037 generated a mutant that had no effect on the nodulation of *L. japonicus* Gifu and this was also true for an exoU double mutant. However, a defective nodulation phenotype on *L. pedunculatus* was observed whereby only about 50% of plants formed effective nodules after a delay and the other 50% of plants only formed uninfected nodule primordia. It was of interest to see if the introduction of this deletion into JC07 could replicate the phenotypes observed for the 9.3-kb deletion alone, or perhaps the genes disrupted within JC07 may have provided a synergistic effect and their absence may alter the phenotype of the 9.3-kb mutant, resulting in a strain unable to form effective nodules.
3.11.1 Generation of mutants JC09 and JC10

In Chapter 4, it was found that the defective nodulation phenotype exhibited by the Δ(nodO-9.3kb) mutant was attributed to the disruption of prsDE/mln031. Although the targeted mutagenesis of prsDE would have been ideal, the isolation of a ΔprsDE mutant had been troublesome and the time invested into generating the NZP2037 ΔprsDE mutant had been substantial (around a year). Thus, in the interest of time, deletion of the whole 9.3-kb region was chosen as a more attractive alternative. Furthermore, the vector used to generate a 9.3 kb deletion mutant, pJQ200SK-nodO (Table 3) had posed no problems for mutagenesis in the past (Liam Harold, personal communication).

The mutant JC07 was used for the markerless deletion of the 9.3-kb region using the suicide plasmid pJQ200SK-nodO. PCR using the primers nodOcheckL, nodOcheckR, and nodOinternalcheckR (Table 5) were used to screen for successful mutants. The mutant with the genotype of R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔnodUFEGA2 Δmln399 ΔvirB1-11 ΔvirA Δ(nodO-9.3kb) was designated as JC09. Disruption of exoU in JC09 using pFUS-exoU to generated the mutant strain JC10.

3.11.2 Generation of R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ(nodO-9.3kb) and R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ(nodO-9.3kb) exoU::pFUS2

It was also of interest to see if the NZP2037 symbiosis island disrupted in the nodO-9.3kb region in the R7A genomic background gave the same defective nodulation phenotype on L. pedunculatus as seen in the NZP2037 background. This was important, as the phenotypes may differ between genomic backgrounds (detailed in section 3.5.2). The transfer of the symbiosis island from the mutant NZP2037 Δ(nodO-9.3kb) into the R7A NS background was performed as detailed in section 3.5 to give transconjugant strain
R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ(nodO-9.3kb). It’s pFUS-exoU derivative was named R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ(nodO-9.3kb) exoU::pFUS2.

### 3.11.3 Nodulation phenotypes of JC09, JC10, R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ(nodO-9.3kb), and R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ(nodO-9.3kb) exoU::pFUS2 on *L. japonicus* Gifu and *L. pedunculatus*

The nodulation phenotypes of JC09 and JC10 on *L. japonicus* Gifu was comparable to the non-mutagenized R7A<sup>C</sup>:NZP2037<sup>I</sup> transconjugant and R7A<sup>C</sup>:NZP2037<sup>I</sup> Δ(nodO-9.3kb), in the presence or absence of an exoU mutation (Figure 3.24A). Similarly, the nodulation phenotypes of JC09 and JC10 on *L. pedunculatus* were comparable to mutants disrupted in the 9.3-kb region (Figure 3.24B). Taken together, the nodulation phenotypes caused by the Δ(nodO-9.3kb) mutation was unaffected by disruption of nodU, nodFEG, nodA2, mln399, virB1-11, and virA.

![Figure 3.24. Nodulation phenotypes of JC09 and JC10. A) *L. japonicus* Gifu and B) *L. pedunculatus*.](image)

Strains used as inocula are indicated by the key. JC09 and JC10 represents the strains R7A<sup>C</sup>:NZP2037<sup>I</sup> ΔnodUFEGA2, Δmln399, ΔTypeIV, Δ(nodO-9.3kb) and its respective exoU::pFUS2 mutant equivalent. Percentage of plants nodulated was calculated from an experiment consisting of 15 plants per strain and is representative of the two experiments carried out.

119
3.12 Discussion

In this chapter, the main goals were to identify unique genetic elements that are responsible for the broader host-range (represented by the host *L. pedunculatus*) of *M. loti* strain NZP2037 and the retention of nodulation compatibility on the host *L. japonicus* Gifu in the presence of defective EPS. These phenotypes are in contrast to the genetically similar *M. loti* strain R7A which belongs to the narrow host-range group, forming only uninfected primordia on *L. pedunculatus* and was not compatible with *L. japonicus* Gifu in the presence of defective EPS. Comparisons between the genomes of NZP2037 and R7A identified genes of interest that were unique to NZP2037 as attractive targets for mutagenesis studies.

Several models to explain the differences between R7A and NZP2037 and their *exoU* mutants in the nodulation of *L. pedunculatus* and *L. japonicus* Gifu were considered at the beginning of this study. Previous work using a complementation approach with a cosmid library of NZP2037 genomic DNA in R7A failed to uncover a clone that enabled R7A to nodulate *L. pedunculatus*, while only *exoU* clones allowed R7A *exoU* mutants to nodulate *L. japonicus* Gifu (Kelly, 2012). This suggested that more than one unlinked genetic determinant from NZP2037 may be required for both effects. The suggestion of multiple determinants missing from R7A that are required for nodulation of *L. pedunculatus* was further supported by the finding that NZP2037 contained a suite of genes under NodD control that were lacking in the narrow-host-range strains R7A and MAFF303099 (Kasai-Maita et al., 2013), and the potential importance of these genes was reinforced early in this study when it was found the suite of genes were present in four other broad-host-range strains and missing from the narrow-host-range strains. Another possibility considered was that R7A produced a negative factor that prevented nodulation.
of *L. pedunculatus*. However, against this idea is the fact that no “escape nodules” ever formed on *L. pedunculatus* inoculated with R7A, whereas escape nodules do form on *L. japonicus* Gifu inoculated with R7A *exoU* mutants. Finally, the different reactions of *L. japonicus* Gifu to the R7A versus NZP2037 *exoU* mutants may be informative, as the current model is that *L. japonicus* Gifu perceives the truncated pentasaccharide produced by *exoU* mutants through its EPR3 receptor, leading to a negative plant response that halts infection (Kawaharada et al., 2015). As NZP2037 *exoU* mutants are thought to produce the same truncated pentasaccharide as R7A *exoU* mutants (Kelly, 2012), then NZP2037 must produce a positive factor that overcomes this effect. It seems possible that this positive factor also allows NZP2037 to nodulate *L. pedunculatus*. Its absence from R7A may explain why the wildtype strain only induces uninfected primordia on *L. pedunculatus*, similar to the uninfected primordia induced by R7A *exoU* mutants on *L. japonicus* Gifu. It was on the basis of these considerations that it was decided to sequentially inactivate all potential symbiotic genes present on the NZP2037 symbiosis island (and in the other broad-host-range strains), with the expectation that mutations that eliminated the ability of NZP2037 to nodulate *L. pedunculatus* would be found.

The finding about half-way through this study that the NZP2037 symbiosis island was tripartite and not disrupted as had been thought, allowed us to transfer the island to a derivative of R7A cured of its own island. The findings that the resultant strain R7A<sup>C</sup>:NZP2037<sup>I</sup> and its *exoU* derivative both formed effective nodules on *L. pedunculatus* and *L. japonicus* Gifu strongly supported our chosen approach, as they showed that the genes required to nodulate *L. pedunculatus* and to overcome the negative effect of the *exoU* mutation were encoded on the island. Interestingly, the inefficient nodulation of *L. japonicus* Gifu by NZP2037 and NZP2037 *ΔexoU* was eliminated in
R7A\textsuperscript{C}:NZP2037\textsuperscript{I}, while the strain nodulated \textit{L. pedunculatus} as well as NZP2037. This led to the decision to use R7A\textsuperscript{C}:NZP2037\textsuperscript{I} as the host strain for further work to develop a strain that lacked all the NZP2037 symbiotic genes that were absent in R7A.

The genes \textit{nodU}, \textit{nodFEG}, and \textit{nodA2} by themselves were not required for nodulation of \textit{L. japonicus} Gifu or \textit{L. pedunculatus} (Fowler, 2013). In this study, the effect of an \textit{exoU} mutation introduced in tandem with the single gene deletion mutants was investigated. The results showed that the \textit{nodU} \textit{exoU} double mutant induced uninfected primordia on the host \textit{L. japonicus} Gifu. In contrast, the symbiotic phenotypes of double mutants of \textit{exoU} and \textit{nodFEG/nodA2} were comparable to that of a single \textit{exoU} mutant. All strains remained effective as symbionts on \textit{L. pedunculatus}. This fully effective nodulation phenotype for \textit{L. pedunculatus} was not unexpected, as similar results were observed by Hotter & Scott (1991) for EPS disrupted mutants of NZP2037 generated with a transposon mutagenesis approach. However, it should be noted that the mutants generated by Hotter & Scott (1991) do not produce truncated EPS, a contrast to \textit{exoU} mutants which produce truncated EPS and so may not have been recognised by Epr3. The results suggest that the genes involved in the synthesis (\textit{nodFEG}) and transfer (\textit{nodA2}) of the acyl group located on the non-reducing end of the NZP2037 Nod factor, were not essential for circumvention of defective EPS signalling.

The inability of the NZP2037 double \textit{nodU exoU} mutant to form an effective symbiotic relationship on \textit{L. japonicus} Gifu was likely due to an inability to enter the developing root nodule structure. The inability to enter the root nodule is likely to be manifested during the elongation of the infection thread past the epidermal layer due to the recognition of defective EPS by the plant receptor EPR3 (Kawaharada \textit{et al.}, 2015). Based
on the inability of the *nodU exoU* mutant to nodulate *L. japonicus* Gifu, it was hypothesised that NodU (predicted to as a carbamoyl transferase, involved in the transfer of a carbamoyl group to C6 on the GlcNAc of the non-reducing end of the Nod factor) was responsible for the retention of nodulation on *L. japonicus* Gifu when defective EPS signalling was present. However, complementation experiments that utilised a plasmid-based approach failed to restore an effective nodulation phenotype. In contrast, complementation of truncated EPS via a plasmid encoding *exoU* did restore an effective nodulation phenotype. This suggested that the defective nodulation phenotype exhibited by a *nodU exoU* double mutant was largely due to the disruption of *exoU*. The inability to restore a nodulation phenotype in mutants that harboured the plasmid pPROBEKT-nodU was hypothesised to be due to a compounded phenotype, involving the burden of NodU expression in conjunction with defective EPS signalling, resulting in a severely delayed nodulation phenotype.

An alternative hypothesis was that the presence of an unidentified secondary mutation within the *nodU exoU* double mutant may have affected nodulation of *L. japonicus* Gifu. An argument against this hypothesis was that two independently generated *nodU exoU* double mutants retained the ineffective nodulation phenotype. The second mutant was generated because the first exhibited an unusual growth phenotype in broth culture (detailed in section 3.3.2). The likelihood that the same secondary mutation would occur is small unless there is an unknown selection for it, supporting the suggestion that disruption of *exoU* was the cause of the ineffective nodulation phenotype. However, it is possible that NZP2037 Δ*nodU*, the mutant that was used as the foundation for mutagenesis, had a prior secondary mutation before the introduction of the *exoU* mutation. Thus, subsequent mutations may not be representative of a loss-of-function mutation and
the secondary mutation hypothesis cannot be ruled out without making a new NZP2037 ΔnodU mutant.

Although complementation with pPROBEKT-nodU was not achieved, it was observed that exoU mutants harbouring pPROBEKT-nodU in the exoU mutant background induced an increased average number of nodule primordia on *L. japonicus* Gifu. The observation suggested that the addition of the carbamoyl group to the Nod factor is involved in the promotion of nodule organogenesis on this host. The presence of the carbamoyl group may aid in the recognition of the Nod factor by NFRs. In support of this, in the strain *Azorhizobium caulinodans* ORS571, which generates pentameric carbamoylated Nod factors with the presence of a 6-Ο-carbamoyl group was shown to have greater biological activity than Nod factors that lacked the additional carbamoyl group (D'Haeze *et al.*, 2000). The increased activity provided by the carbamoyl group could be replicated when an acetyl group was replaced at the same position, and it was speculated that the keto function or the lack of a C-6-hydroxyl group may have contributed to the enhanced recognition that was observed (D'Haeze *et al.*, 2000).

If NodU is involved in organogenesis, it is tempting to speculate that the increase in primordia numbers could lead to an increase in the frequency of crack entry. However, EPR3 has been shown to be involved in crack entry and is able to recognise defective EPS during entry into the plant via this route, resulting in negative signalling and subsequent inhibition of infection (Kawaharada *et al.*, 2015). Therefore, alternative entry into host cells via crack entry as a consequence of increased primordia formation would be unlikely to rescue the defective nodulation phenotype of EPS disrupted mutants. This is reflected in the finding that NodU was unable to restore the wildtype nodulation
phenotype in the \textit{exoU nodU} double mutant background. Given that the LysM2 domain within NFR5 of \textit{L. japonicus} is predicted to recognise the non-reducing end of the Nod factor (Radutoiu \textit{et al.}, 2007, Bek \textit{et al.}, 2010), it is likely that the presence of the carbamoyl group promotes the recognition of Nod factors by their respective NFRs. Therefore, the hypothesis is that NodU promotes primordia formation via an increased affinity to NFRs that are involved in organogenesis. However, the expression of carbamoylated Nod factor alone is unable restore a nodulation phenotype to defective EPS mutants nor could it expand the host range of R7A, suggesting that other factors are required for both.

Another possibility is that the introduction of the \textit{nodU} mutation resulted in lowered Nod factor concentrations being produced by the mutant. Decreased Nod factor production due to \textit{nodU} gene disruption was documented in a \textit{nodSU} mutant of \textit{S. fredii} NGR234 and in a \textit{nodU} mutant of \textit{A. caulinodans} ORS571 (Jabbouri \textit{et al.}, 1995, D'Haeze \textit{et al.}, 2000). Interestingly, despite the decrease in Nod factor concentrations, no effect on the nodulation efficiency of \textit{A. caulinodans} ORS571 was observed (D'Haeze \textit{et al.}, 2000). This lack of a phenotype was also observed in the current study for the single \textit{nodU} mutant of NZP2037. Nevertheless, in the double \textit{exoU nodU} mutant, it is possible that the lowered Nod factor concentrations in conjunction with defective EPS resulted in a severely disrupted nodulation phenotype represented by the inability of the mutant to nodulate \textit{L. japonicus} Gifu.

Taken together, the results indicate that \textit{nodU} is not essential in the establishment of a symbiotic relationship with \textit{L. pedunculatus} or \textit{L. japonicus} Gifu. However, \textit{nodU} promoted organogenesis priming the formation of the nodule structure. It is known that
organogenesis is not influenced by EPR3 (Kawaharada et al., 2015). Therefore, the signalling pathway pertaining to carbamoylated Nod factor is likely to be independent of EPR3 signalling. However, the two pathways eventually converge with one another, resulting in a successful symbiosis, and a defect in either pathway prevents effective nodulation (Guinel & Geil, 2002). Potential hypotheses could be that NodU promotes Nod factor affinity to NFRs responsible for organogenesis or that changes in Nod factor concentrations due to nodU mutagenesis were responsible. Future experiments could focus on the purification of the Nod factor from a nodU mutant. This approach would allow for the quantification of Nod factor concentrations generated by nodU mutants or used in exogenous applications to test their activity. Regardless, nodU, is not essential for nodulation of the hosts tested and the presence of nodU was unable to rescue a defective nodulation phenotype in the R7A background.

The further delayed nodulation phenotype for L. japonicus Gifu observed in NZP2037 exoU mutants compared to NZP2037 was not apparent in the R7A^C::NZP2037^l background, as the exoU mutants nodulated with comparable kinetics to the non-mutagenized transconjugant. It may be possible that detrimental factors relating to L. japonicus Gifu nodulation are present within the chromosome of NZP2037. The transfer of these chromosomal genes would not accompany transfer of the symbiosis island, thus the transconjugant strains lack these detrimental genes and as a consequence exhibit a more robust nodulation phenotype. Taken together, this reinforces that gene(s) responsible for the circumvention of defective EPS signalling and extended host-range are located within the NZP2037 symbiosis island.

In the NZP2037 genomic background, the nodU exoU double mutant was no longer able
to nodulate *L. japonicus* Gifu, and there was no effect on the nodulation of *L. pedunculatus*. However, the mutant R7A<sup>C</sup>:NZP2037<sup>Δ</sup> ΔnodU exoU::pFUS2 did not exhibit the defective nodulation phenotype. The lack of a phenotype in R7A<sup>C</sup>:NZP2037<sup>Δ</sup> ΔnodU exoU::pFUS2 is likely due to the proficiency of the transconjugant that lacks the delayed nodulation phenotype of an *exoU* mutant in the NZP2037 background. This supported the decision to use R7A<sup>C</sup>:NZP2037<sup>Δ</sup> as the host strain for further work due to its more robust nodulation capabilities.

During this study, additional genes unique to the NZP2037 island were identified and targeted for mutagenesis. One of the new targets identified were the genes *virB1-11* and *virA* which form the T4SS. The T4SS was of interest to study due to observations that in the R7A genomic background, the effectors of the T4SS inhibited and promoted the nodulation of *L. filicaulis* and *L. corniculatus*, respectively (Hubber et al., 2004). The lack of an observable phenotype for the NZP2037 T4SS mutant on *L. pedunculatus* and *L. japonicus* Gifu had previously been shown by Kasai-Maita et al. (2013) where NZP2037 mutants that were disrupted in the T4SS were found to have no major differences on the nodulation of multiple *Lotus* hosts. In this study it was hypothesized that these effector proteins may instead be involved in the circumvention of defective EPS that was observed in NZP2037 *exoU* mutants or that under our experimental conditions, mutants disrupted in the T4SS may behave differently to what was observed by Kasai-Maita et al. (2013). The results from the NZP2037 ΔTypeIV mutant reflected those observed by Kasai-Maita et al. (2013). The T4SS was not required for nodulation of *L. pedunculatus* nor was it not essential in the circumvention of defective EPS signalling on *L. japonicus* Gifu.
An explanation for these results could be that the effectors may have opposing roles on the nodulation process, similar to the T3SS of *M. loti* strain MAFF303099, which secretes effectors that are highly similar to those secreted by the T4SS of R7A (Hubber et al., 2004). Effectors with conflicting effects on nodulation have been documented in the T3SS effectors NopM NopJ and NopT in *S. fredii* NGR234, where depending on the host, the effectors may elicit a defence response or promote symbiosis and it is the net effect of these effectors that determines the outcome of symbiosis (Kambara et al., 2009). A similar effect was observed for the T3SS of MAFF303099 for the effectors Mlr6361 and Mlr6331 on *L. japonicus* Gifu MG20 (Sánchez et al., 2012). If the effectors of the T4SS in NZP2037 had opposing effects on nodulation then the removal of a single effector may be of interest for future experiments. On the flipside, the simplest explanation would be that the effectors secreted by the T4SS of NZP2037 simply do not have targets within the hosts tested.

The second target that was investigated in this study was the gene *mln399*. Homologues of *mln399* were found to be present only within strains of *M. loti* that belong to the broad host range group (John Sullivan, personal communication) and thus the gene was of interest to study. Unfortunately, the results showed that *mln399* was not essential for the nodulation of *L. pedunculatus*, nor was it involved in circumvention of defective EPS signalling on *L. japonicus* Gifu. Mln399 is predicted to function as a glycosyl transferase and is hypothesized to transfer a sugar residue to the Nod factor. In support of this hypothesis was the genetic context of the gene, which was located directly downstream of the *nodACIJ-nolO* operon involved in Nod factor synthesis, suggesting that it may be co-transcribed as part of the *nodACIJ-nolO* operon. It is possible that *mln399* is simply not co-transcribed with *nodACIJ-nolO*, and not expressed during the symbiotic process.
Alternatively, *mln399* may not encode a functional protein, or there may be functional homologues elsewhere within the genome. The reported NZP2037 Nod factor structure was described to not contain unique glycosyl modifications, nor any unsaturated fatty acid moieties on the non-reducing end (López-Lara *et al.*, 1995) but it would be worthwhile to determine the Nod factor structures from the *mln399* mutant using the more powerful technologies now available (see Bek *et al.* (2010)). For example, the expression of *mln399* may be low as *mln399* is the final gene within the proposed *nodSACIJ-nolO-mln399* operon. Low expression of *mln399* may result in low concentrations of Nod factors that contain the unique glycosyl modification and thus may have escaped identification by López-Lara *et al.* (1995).

If we assume that Mln399 was responsible for the addition of a glycosyl to the Nod factor structure, the question remains of where this glycosyl modification would be located? A clue comes from the *M. loti* strain NZP2213, which produces Nod factors that contain a fucosyl modification located on the second GlcNAc proximal to the non-reducing end, although the gene responsible for this modification has yet to be identified (Olsthoorn *et al.*, 1998). Perhaps *mln399* is a functional homologue to this gene and modifications to the Nod factor at this position may be present within the NZP2037 secreted Nod factors. It was hypothesised that this fucosyl group was unrelated to strain host-specificity as NZP2213 belongs to the group of *M. loti* strain that have a narrow host range, and the fucosyl group was proposed to aid in the stability of the Nod factor (Olsthoorn *et al.*, 1998). The results in this study would support that Mln399 was not involved in host-specificity and its function may instead be involved in the stabilization of the Nod factor.

Multiple gene deletions were made sequentially within a single strain in the
transconjugant background resulting in a mutant that had lost all unique genes of interest 
\(\text{nodU, nodFEG, nodA2, virA, virB1-11, mln399}\) detailed in this study, except for \(\text{nodO-9.3kb}\). However, the mutant (and its \(\text{exoU}\) mutant) exhibited a nodulation phenotype that was comparable to that of the un-mutagenized \(\text{R7A}^{C}:\text{NZP2037}\). The results suggested that the mutated genes collectively were not essential for nodulation of \(L.\ pedunculatus\) nor were they involved in circumvention of defective EPS signalling on \(L.\ japonicus\) Gifu.

When a \(\text{nodO-9.3kb}\) deletion was introduced, the resulting mutant exhibited a nodulation phenotype that was comparable to that of a single \(\text{nodO-9.3kb/prsDE}\) mutant suggesting that the defective nodulation phenotype was due to the 9.3-kb deletion alone.

Interestingly, the Nod factors predicted to be produced by these mutants should be comparable in structure to Nod factor produced by wildtype R7A (predicted Nod factor structures produced by R7A and NZP2037 are shown in Figure 3.25), though the structures should be determined in future work. Based on this prediction, it could be hypothesised that the R7A Nod factor is compatible with \(L.\ pedunculatus\). In support of this hypothesis, the extracellular region of the NFRs that recognise Nod factors for \(L.\ pedunculatus\) and \(L.\ japonicus\) Gifu were interchangeable and both could recognise R7A Nod factor (Bek et al., 2010). However, R7A was still incompatible for the nodulation of \(L.\ pedunculatus\), even when \(L.\ japonicus\) NFR1 and NFR5 were present. These observations lead Bek et al. (2010) to hypothesise that additional symbiotic components were required for infection of \(L.\ pedunculatus\). If the Nod factor structures were identical to R7A Nod factor, an interesting experiment could be to see what Nod factor modifications are required in order for nodulation of both hosts, perhaps only a basic Nod factor structure is required and other factors are required as host-determinants.
The predicted modifications on the NZP2037 Nod factor are presented in purple. The gene *nodU* is responsible for the transfer of a 6-O-carbamoyl group to the 6\(^{th}\) Carbon on the GlcNAc at the non-reducing end. The genes *nodFEG/nodA2* are responsible for the synthesis and transfer of a putative unsaturated fatty acid moiety on to the non-reducing GlcNAc residue of the Nod factor. Note: only two acyl groups are shown here as examples of potential acyl groups that could be transferred.

The incompatibility between R7A and *L. pedunculatus* may be due to the presence of prodelphinidin-rich flavolans (condensed tannins) within the roots of *L. pedunculatus* that are known to have bactericidal effects at high concentrations (Pankhurst et al., 1982). It has been shown that R7A can gain entry into *L. pedunculatus* root hairs, but development of the infection thread is arrested at the epidermal layer (Bek et al., 2010). Perhaps R7A is unable to persist within *L. pedunculatus* due to high concentrations of flavolans present, preventing colonisation of the nodule primordia, similar to that seen in NZP2213 which belongs to narrow host-range *M. loti* (Jones et al., 1987).
In contrast to R7A, NZP2037 is resistant to the flavolans produced by *L. pedunculatus* and this was attributed to a flavolan-binding polysaccharide covalently bound to the peptidoglycan layer of NZP2037 (Pankhurst *et al.*, 1982, Jones *et al.*, 1987). The flavolan-binding polysaccharide was found to not associate with EPS (Jones *et al.*, 1987) and this is supported in this study, where all mutants of NZP2037 that produced truncated EPS retained the ability to nodulate *L. pedunculatus*. The flavolan-binding polysaccharide is tightly associated with the membrane, therefore, genes that encode flavolan-binding polysaccharide may be associated with constitutively expressed structural genes that are not under NodD regulation. The genes targeted for mutagenesis in this study all contained a predicted *nod* box in their promoters, and therefore the gene that encodes this flavolan-binding polysaccharide would not have been targeted. A small caveat to the flavolan hypothesis is that toxic flavolan synthesis was suggested to not occur during the early stages of infection, suggesting that flavolans were not major determinants of nodulation as the symbionts do not encounter them during infection (Cooper & Rao, 1992). However, it should be noted that Cooper & Rao (1992) did not differentiate between cortical cells and epidermal cells within the nodule structure, and their findings do support that flavolan production is increased in ineffective nodules but not in effective nodules. Flavolan was regularly found directly below multiplying bacteria that were incompatible with *L. pedunculatus*. In contrast, no flavolans were observed in nitrogen-fixing nodules that contained bacteroids (Pankhurst *et al.*, 1979). Interestingly, the lack of flavolans within nitrogen fixing nodules suggests that rhizobial involvement modulates flavolan synthesis and that additional signalling would likely occur only within an effective mature nodule containing nitrogen fixing bacteroids (Pankhurst *et al.*, 1979, Cooper & Rao, 1992). Although the question remains as to whether flavolan production is a consequence of ineffective signalling or that it is produced as a general defence response.
Taken together, R7A may produce Nod factor that is compatible with *L. pedunculatus* and is able to undergo early infection, however, R7A is unable to persist within the host due to the presence of the bactericidal flavolans within this host that leads to eventual cell death. In contrast, NZP2037 encodes a flavolan binding protein that protects against the effects of these flavolans allowing for the formation of an effective nitrogen fixing nodule, alongside an as of yet unidentified signalling molecule that is involved in the production of these toxic flavolans. Future experiments to test this hypothesis could be co-inoculation of R7A with NZP2037 (or NZP2037 infection-defective mutant) as the presence of NZP2037 may protect R7A from the flavolans present within *L. pedunculatus*, allowing R7A infection to proceed past the flavolans rich cells.

Thus far, this discussion has focused on the assumption that gene(s) unique to NZP2037 were responsible for its extended host-range. However, it is possible that the R7A symbiosis island encodes effectors detrimental to the nodulation of *L. pedunculatus*. If we were to assume that the Nod factor of R7A was compatible with *L. pedunculatus*, perhaps the inability to establish a symbiotic relationship may be related to unique genes within the R7A island that illicit a plant defence response. To test the hypothesis of negative signalling within the R7A symbiosis island, a cosmid library of R7A could be transferred into the wildtype NZP2037 and cosmids that endow interesting phenotypes on *L. pedunculatus* could be investigated in more detail.

In summary, disruptions of the genes *nodU, nodFEG, nodA2, virB1-11, virA, mln399* alone or collectively in the R7A_C:NZP2037_1 transconjugant genomic background, had no effect on the nodulation of *L. japonicus* Gifu and *L. pedunculatus* in the presence or
absence of defective EPS. EPS signalling does not appear to be important (or even recognised) in *L. pedunculatus*. In contrast, EPS signalling has an important role in the nodulation of *L. japonicus* Gifu, but the gene(s) that allow for the circumvention of defective EPS signalling could not be identified within this study. A defective nodulation phenotype was observed for a *nodU exoU* double mutant on *L. japonicus* Gifu, and this suggested that *nodU* was involved in the formation of nodule primordia on this host, but by itself was unable to fully overcome the negative signalling of truncated EPS. The more robust nodulation phenotypes observed in the transconjugant background was hypothesised to be due to the absence of factors encoded within the NZP2037 chromosome. Alternatively, the R7A symbiosis island may encode gene(s) detrimental to nodulation and the removal of these results in a more effective symbiont. Lastly, the Nod factor produced by R7A was hypothesised to be compatible with *L. pedunculatus* and the inability to nodulate *L. pedunculatus* was attributed to an inability to survive the harsh environment that is present within this host as was previously proposed by Pankhurst *et al.* (1982), and Jones *et al.* (1987).
4 The journey from NodO, to the Type I secretion system PrsDE and its secreted effector MIn031
4.1 Introduction

Of the unique genes in the NZP2037 island that were identified to likely play a role in nodulation due to the presence of a nod box upstream of the gene, nodO was the only one not involved in the synthesis and modification of Nod factor. Instead, the nodO gene encodes a secreted protein that contains a calcium binding domain and has been shown to transport cations across lipid bilayers (Economou et al., 1990, Sutton et al., 1994). The ability to transport cations was hypothesised to amplify Ca\(^{2+}\) influx, which is a prerequisite for nodule formation and supports the elongation of the infection thread (Miwa et al., 2006). In R. leguminosarum bv viciae, a defective nodulation phenotype was only observed for a nodO mutant when accompanied by a second nodE mutation (Economou et al., 1994). The observation that nodO mutants only had a phenotype in a nodE mutant lead to the hypothesis that it was responsible for amplification of the signals expressed by plants when they recognise potentially compatible Nod factors (Economou et al., 1994, Walker & Downie, 2000, Miwa et al., 2006). Furthermore, the expression of Rhizobium sp. BR816 NodO increased the host range of R. etli CE3 to include L. leucocephala, a previously incompatible host (Vlassak et al., 1998). Therefore, nodO was of interest to study due to its involvement in the establishment of a successful symbiotic relationship. This chapter details the phenotypes observed in a deletion mutant that had lost a 9,345-bp region of DNA that included nodO. The mutant exhibited a previously-undescribed symbiotic phenotype where only about 50% of inoculated plants formed nodules, whereas the other 50% formed only uninfected primordia. Efforts made to identify the genes within the deleted region responsible for this phenotype resulted in the finding that the Type I secretion system, PrsDE, and an effector it most likely transported, Mln031, as key components involved in the nodulation of the host L. pedunculatus by M. loti strain NZP2037. Surprisingly, the nodO gene was not involved in the phenotype.
4.2 Results

4.2.1 Phenotype of NZP2037 Δ(nodO-9.3kb)

At the start of this study, the DNA sequence of the region spanning nodO within the NZP2037 genome had been assembled by two separate groups, a whole genome sequence that was available on the JGI database (Grigoriev et al., 2011), and another sequence assembly restricted to the symbiosis island that was assembled by a Japanese group (Kasai-Maita et al., 2013). The DNA sequences from the two groups were largely identical; however, the DNA in the region spanning nodO had been assembled differently, likely due to the presence of highly repetitive DNA within the region. The discrepancy in sequence assembly between the two groups made it difficult to design a targeted mutagenesis approach for disruption of nodO. During this study, the correct sequence assembly was identified as the Japanese sequence (pictured in Figure 4.1); the steps that lead to this conclusion are detailed in section 4.2.6 of this chapter.

![Figure 4.1. The genetic organisation of the DNA proximal to nodO and the 9345 bp markerless DNA deletion in the mutant Δ(nodO-9.3kb).](image)

The presence of a nod box (yellow) in the promoter region of nodO suggests NodD regulation. Downstream of nodO (black) are genes encoding a Type I secretion system prsDE (red) and a putative pseudogene mln031 that encodes a small putative Ca²⁺ binding protein (blue), a region of repetitive DNA (green) followed by the nod box at the promoter region of noeKJ (orange). The 9345 bp deletion in Δ(nodO-9.3kb) initiated 61 bp downstream of the 5’ end of nodO and terminated 3 bp upstream of the nod box in the promoter region of noeK (shown as red line).
Previously in the lab, a mutant was made in the _nodO_ region by Liam Harold, a summer student at the time. The mutant (referred to as Δ(_nodO-9.3kb_)) contained a 9,345 bp deletion that initiated 61 bp within the 5’ end of _nodO_ and spanned the rest of _nodO_, an adjacent Type I secretion system _prsDE_, and a putative pseudogene (_mln031_) encoding a small protein containing calcium binding motifs. This region comprised 4.8 kb and the remainder of the deleted region was a repetitive region of DNA that contained fragments of transposons and integrases, but no other clearly discernible ORFs. The deletion terminated 341 bp upstream of _noeK_ and left the _nod_ box in the promoter region intact (Figure 4.1). Note that in the incorrect genome sequence both _nodO_ and _mln031_ were not located proximal to the _prsDE_ genes due to the repetitive DNA that encompassed _prsDEmln031_. The 9,345 bp deletion was designed to avoid possible issues that could have arisen during mutant construction from complications caused by recombination of the highly repetitive DNA that was likely to have caused the discrepancy in assemblies between the two sequences that were available. The successful mutant was confirmed with primers that bound to flanking regions of DNA proximal to the deleted region (John Sullivan, personal communication).

NZP2037 Δ(_nodO-9.3kb_) exhibited a large decrease in the percentage of _L. pedunculatus_ nodulated when compared to the wildtype (Figure 4.2A). In contrast, no effects were observed for nodulation of _L. japonicus_ Gifu (Figure 4.2B). A proportion of _L. pedunculatus_ (~50%) inoculated with Δ(_nodO-9.3kb_) formed effective nitrogen fixing nodules (spherical nodules > 1 mm; plants were green), while the remaining plants formed ineffective nodule primordia (tumour-like irregularly shaped structures, white/pale pink colouration; plants were yellow). Examples of plants are shown in Figure 4.3.
Figure 4.2. Nodulation kinetics of NZP2037 Δ(nodO-9.3kb) on different Lotus hosts. A) Nodulation kinetics of *L. pedunculatus*. Percentage of plants was calculated from a total of 60 plants across 4 separate experiments. *p*-value for NZP2037 vs Δ(nodO-9.3kb) = <0.001. In no experiment did more than 60% of plants form effective nodules. B) Nodulation kinetics of *L. japonicus* Gifu. Percentage of plants was calculated from 45 plants across 3 experiments. *p*-value for NZP2037 vs Δ(nodO-9.3kb) = 0.7296. Error bars represent the standard error of the mean.

Figure 4.3. The Δ(nodO-9.3kb) mutant has two different phenotypes on *L. pedunculatus*. Uninoculated is a water-only negative control, NZP2037 is the wildtype strain that forms effective nitrogen fixing nodules giving rise to healthy green leaves. Δ(nodO-9.3kb) refers to plants inoculated with NZP2037 Δ(nodO-9.3kb) that only formed small nodule primordia. ‘Δ(nodO-9.3kb) ’escape’” refers to plants inoculated with Δ(nodO-9.3kb) that formed effective nitrogen-fixing nodules after a delay.
For comparison between the wildtype and plants that formed effective nodules, plants that did not form nodules were excluded from the data set. From this it could be observed that of the $\Delta(nodO-9.3\text{kb})$ inoculated plants that formed effective nitrogen fixing nodules, $\sim 7$ days delay in nodulation was observed (Figure 4.4A) leading to a decrease in the average number of nodules formed (Figure 4.4B). Interestingly, the onset of primordia formation was similar between plants that developed effective nodules and plants where the primordia did not develop further, suggesting that organogenesis was not likely to be affected (Figure 4.5).

In conclusion, $\Delta(nodO-9.3\text{kb})$ exhibited a defective nodulation phenotype on *L. pedunculatus* that was manifested by a delayed nodulation in 50% of plants and in the formation of only ineffective nodule primordia in the other 50% of plants. In contrast, no effects on nodulation could be observed on *L. japonicus* Gifu. Taken together, these observations suggest that a gene(s) within the 9.3-kb region of DNA that was disrupted in $\Delta(nodO-9.3\text{kb})$ was involved in the establishment of a symbiotic relationship with *L. pedunculatus* but not *L. japonicus* Gifu. In the following sections, the symbiotic phenotypes of $\Delta(nodO-9.3\text{kb})$ on the host *L. pedunculatus* were further characterised.
Figure 4.4 Comparison between wildtype NZP2037 and plants that formed effective nodules when inoculated with Δ(nodO-9.3kb) on *L. pedunculatus* and the exclusion of data from uninfected plants. Wildtype is shown in red and the Δ(nodO-9.3kb) mutant is shown in blue. A) Percentage of plants nodulated. *p*-value = 0.0376. B) Average number of nodules. Percentage of plants was calculated from a total of 60 different plants across 4 experiments. Error bars represent standard error of the mean. *p*-value = 0.0042. Plants that only formed primordia are not included in either figure part.

Figure 4.5. The onset of nodule/primordia for Δ(nodO-9.3kb) on *L. pedunculatus* is similar to each other. Δ(nodO-9.3kb) refers to plants that developed effective nitrogen fixing nodules. Δ(nodO-9.3kb) primordia refers to plants that did not develop into effective nodules. Percentage was calculated from a total of 60 plants across 4 experiments. Error bars represent the standard error of the mean.
4.2.2 The mutant Δ(nodO-9.3kb) is affected in its ability to initiate infection threads

Given that Δ(nodO-9.3kb) had a defective nodulation phenotype on L. pedunculatus, it was of interest to determine the step of the infection process that the mutant had a negative impact on. To test this, the stable vector pSKGFP was introduced into the wildtype strain NZP2037 and the mutant strain Δ(nodO-9.3kb) via conjugation. The presence of pSKGFP generated strains that constitutively expressed GFP, allowing for visualization of the bacteria under a fluorescent microscope. Initially, L. pedunculatus plants were grown in a pillow system. However, it was found that the seedlings formed multiple lateral roots and nodule formation no longer occurred at the expected timeframe of ~14 days post inoculation observed for plants that had been grown in a test tube system (data not shown). Therefore, an agar plate system was used instead. Briefly, germinated seedlings were placed onto slants of Jensen’s agar inside square plates and incubated for at least 24 hours in the plant room conditions.

Following incubation, the seedlings were then inoculated with 100 µl of Δ(nodO-9.3kb) pSKGFP suspended in sterile H2O and incubated for another 11 days before the roots were removed from the plate and placed onto a glass slide for observation under a fluorescent microscope. The results (Figure 4.6) showed that only 2 out of 5 plants tested (i.e. 40%) formed infection threads when inoculated with Δ(nodO-9.3kb). In contrast, 100% of the wildtype control formed infection threads. The percentage of plants observed to have formed infection threads when inoculated with Δ(nodO-9.3kb) was similar to the percentage of plants nodulated when grown in tube assays. No truncated infection threads were observed, suggesting that Δ(nodO-9.3kb) was affected at the initiation stage of infection thread formation on L. pedunculatus.
Figure 4.6. Number of infection threads formed by NZP2037 and $\Delta(nodO-9.3\text{kb})$ on *L. pedunculatus* at 11 days post inoculation. Bacteria containing the construct pSKGFP were used as inoculum on *L. pedunculatus* and the number of infection threads formed by fluorescent bacteria were recorded at 11 days post inoculation under a fluorescent microscope. Data obtained from 5 plants for each inoculum. Error bars represent the standard error of the mean.

4.2.3 Plant genetics have little role in the phenotype of $\Delta(nodO-9.3\text{kb})$

Although the disruption of the 9.3-kb region within $\Delta(nodO-9.3\text{kb})$ generated a defective nodulation phenotype on *L. pedunculatus*, there were still plants that formed effective nodules, albeit with a delay. It was hypothesised that these ‘escape’ plants were able to nodulate due to natural genetic variation within the plant population.
To test this hypothesis, roots were regenerated from plants that had previously shown compatibility with \( \Delta(nodO-9.3\text{kb}) \), generating what was effectively a clone of the original plant root, and this new root system was then inoculated with \( \Delta(nodO-9.3\text{kb}) \). If plant genetics were responsible, we would expect that 100% of plants with regenerated roots derived from previously compatible plants would retain the ability to form effective nitrogen-fixing nodules.

Plant seedlings that had formed nitrogen-fixing nodules with an inoculum of \( \Delta(nodO-9.3\text{kb}) \) that had been incubated for ~6 weeks were removed from their agar slants and their roots were removed with a flamed scalpel blade. These rootless seedlings were then moved to a glass jar containing only vermiculite and watered with Hornum’s solution (Márquez et al., 2005) to stimulate formation of new roots. After new roots had re-grown from the previously rootless seedlings (~14 days incubation), the plants were removed carefully from the vermiculite jars and transferred onto agar slants and re-inoculated with either wildtype NZP2037 or \( \Delta(nodO-9.3\text{kb}) \) and incubated for a further 14 days. The results (Figure 4.7) showed that of the plants with roots that were regenerated from previously compatible seedlings, around 50% of plants formed effective nodules when inoculated with \( \Delta(nodO-9.3\text{kb}) \), whereas all plants inoculated with NZP2037 formed effective nodules. The results indicate that the nodulation phenotypes of \( \Delta(nodO-9.3\text{kb}) \) detailed thus far were not due to plant genetic variation but was instead due to the disruption generated by the 9.3-kb mutation, thus disproving that the plant genotype was responsible for the phenotypes observed.
Figure 4.7. Nodulation of *L. pedunculatus* plants after root regeneration at 14 days post inoculation with either the wildtype or the Δ(*nodO*-9.3kb) mutant. Plants that had formed nitrogen fixing nodules with Δ(*nodO*-9.3kb) had their roots removed and regenerated prior to re-inoculation. Plants were inoculated with either the wildtype, or Δ(*nodO*-9.3kb) and observed 14 days post inoculation. Each bar represents a sample size of 6 plants.

### 4.2.4 Recovery of Δ(*nodO*-9.3kb) from plants

Thus far, it has been shown that Δ(*nodO*-9.3kb) has a defective nodulation phenotype on *L. pedunculatus* and that this phenotype was not directly related to the natural genetic variation within the plant population. A hypothesis was put forward suggesting that strains of Δ(*nodO*-9.3kb) that had formed effective nitrogen fixing nodules on *L pedunculatus* had undergone a secondary mutation that allowed the mutant strains to regain an effective nodulation phenotype.

To test this hypothesis, one nodule from each of six *L. pedunculatus* plants that had formed nodules with Δ(*nodO*-9.3kb) were crushed and the bacteria recovered from within nodules were incubated on G/RDM for 5 days. The bacteria that were recovered from
these nodule crushes were then used as inocula onto new *L. pedunculatus* seedlings and their nodulation proficiency was observed (Figure 4.8). All six isolates from the nitrogen fixing nodules (named recover 1 through 6) exhibited the phenotype of 40-60% nodulation similar to the original Δ(*nodO*-9.3kb) mutant. It was therefore concluded that the defective nodulation phenotype of Δ(*nodO*-9.3kb) was not due to secondary suppressor mutations.

![Graph showing plant nodule percentage over days post inoculation](image)

**Figure 4.8.** Bacteria recovered from effective nitrogen fixing nodules of *L. pedunculatus* plants inoculated with Δ(*nodO*-9.3kb) and re-inoculated onto *L. pedunculatus*. Bacteria within effective nodules from six different *L. pedunculatus* seedlings inoculated with Δ(*nodO*-9.3kb) were re-isolated and named ‘recover’ 1-6. The recovered strains were used as inocula for fresh *L. pedunculatus* seedlings and displayed the same phenotype as the original Δ(*nodO*-9.3kb) mutant strain. Percentage of plants was calculated from 15 plants per strain.
4.2.5 Exopolysaccharide plays no role in the phenotype of Δ(nodO-9.3kb)

Full length EPS is known to have a supportive role in the establishment of the *Mesorhizobium-Lotus* symbiosis (Kelly et al., 2013) and it may be that EPS promotes nodulation via the same pathway as that of the nodO-9.3kb deletion. Therefore, the hypothesis was that the presence of full length EPS may complement the defective nodulation phenotype of the Δ(nodO-9.3kb) mutant, resulting in the ~50% of plants that still formed nitrogen fixing nodules. To test this hypothesis, IDM mutants of exoU were made in the Δ(nodO-9.3kb) mutant background using the construct pFUS-exoU, and colonies that displayed a rough colony phenotype indicative of production of truncated EPS were purified. The results showed that strains that produced truncated EPS, in addition to the Δ(nodO-9.3kb) deletion, displayed a phenotype almost identical to that of the Δ(nodO-9.3kb) mutant alone (Figure 4.9A). These results suggest that the Δ(nodO-9.3kb) deletion phenotype either acts upstream of EPS recognition by the host or that the phenotype of the 9.3-kb deletion mutant is not related to EPS signalling. The double mutant strain was also tested on *L. japonicus* Gifu, to observe if the genes lost in Δ(nodO-9.3kb) were involved in the ability of NZP2037 ΔexoU to nodulate this host (this phenotype is described in more detail in Chapter 3.2). The results showed that the double mutant nodulated with similar kinetics to the exoU mutant (Figure 4.9B).

4.2.6 Complementation of Δ(nodO-9.3kb)

From within the deleted 9.3-kb region of DNA, the nodO and prsDE genes seemed likely candidates to have caused the defective nodulation phenotype of Δ(nodO-9.3kb) on *L. pedunculatus*. Therefore, a complementation experiment was designed to determine if re-introduction of these genes in trans to the Δ(nodO-9.3kb) mutant could restore 100% nodulation on *L. pedunculatus*. 
Figure 4.9. Double mutants of Δ(nodO-9.3kb) and exoU and their nodulation kinetics. A) Nodulation of *L. pedunculatus*. B) Nodulation of *L. japonicus* Gifu. Strains used as inocula are as indicated in the key. Percentage of plants was calculated from 15 plants per strain and is representative of the two experiments carried out.
4.2.6.1 Generation of complementation plasmid pFAJ1700-nodOprsDE

To re-introduce nodOprsDE back into Δ(nodO-9.3kb), a region of DNA containing the genes nodO, prsDE and the majority of mln031 was amplified via PCR using the primer pair nodOcompH3L and hltcabpH3rev (Table 5). The PCR product generated was then purified and ligated into the vector pFAJ1700 forming the complementation vector that was designated as pFAJ1700-nodOprsDE. The plasmid pFAJ1700-nodOprsDE was verified via a restriction digest with HindIII, which produces a product size of 5212 bp (Figure 4.10) in a successful clone. The plasmid was then sent for Sanger sequencing to confirm the insert sequence using the primers faj3, faj5, nodOseqL1, nodOseqL2, nodOseqL3, nodOseqR1, nodOseqR2, nodOseqR3 and nodOseqR4 (Table 5). The sequence data that were obtained from pFAJ1700-nodOprsDE confirmed both the integrity of the insert and the genetic organisation within the area immediately downstream of nodO (as shown in Figure 4.1). The genetic organisation was found to be identical to the sequence data of Kasai-Maita et al. (2013), thus supporting the validity of their assembly in the loci of nodO. It should be noted that the primers used to generate pFAJ1700-nodOprsDE were originally designed according to the JGI genome and as a consequence, mln031 was truncated, lacking 65 bp of DNA from the 3-prime end. Following confirmation of pFAJ1700-nodOprsDE, the plasmid was transferred into Δ(nodO-9.3kb) via conjugation and the nodulation phenotypes of this strain are detailed below.
4.2.6.2 Complementation of Δ(nodO-9.3kb) with pFAJ1700-nodOprsDE

The strain Δ(nodO-9.3kb) pFAJ1700-nodOprsDE was used as inoculum for *L. pedunculatus* and it was found that the plasmid did not restore the phenotype of Δ(nodO-9.3kb) (Figure 4.11). This suggested that pFAJ1700-nodOprsDE either does not contain the gene(s) responsible for the defective nodulation phenotype of Δ(nodO-9.3kb) or that problems arising from expression of the gene(s) *in trans* may have prevented successful complementation.
4.2.7 Cosmid complementation of \textit{nodO}

4.2.7.1 Screen of the NZP2037 library for cosmids that contain \textit{nodO}

As the construct pFAJ1700-nodOprsDE failed to complement $\Delta$(\textit{nodO}-9.3kb) (section 4.2.6.2), another approach for complementation was devised. A cosmid library that was representative of the NZP2037 genome was available within the lab (Kelly, 2012) and cosmids that contained \textit{nodO} were isolated from it for use in complementation experiments. The cosmid library was screened using the primers nodOseqL2 and nodOsinglecheckR (Table 5) which bind to the 3’ end of \textit{nodO} and generate a 337 bp product. Two cosmids of interest were isolated and named 6B1.5 and 8B4.5, reflecting their positions within the library (Figure 4.12).
Figure 4.12. PCR screen for the presence of nodO in the NZP2037 cosmid library. For both gels; M = Marker made from HindIII digested λ phage and HaeIII digested φX174. wt is the wildtype control. Numbers above the lanes denotes the well number within in a column of a 96 well plate. A) Screen for nodO in the cosmid 6B1.5, shown here is the 6th plate, side B (second half of the 96 well plate), column number 1 within the library. As the cosmid was isolated from the 6th plate, side B, column number 1, well #5, it was named 6B1.5. Similarly, B) Screen for nodO in the cosmid 8B4.5 which was from the 8th plate, side B, column number 4, well #5. Bands that represent the presence of nodO are highlighted by red boxes.

The regions of the NZP2037 genomic DNA that had inserted within the cosmids were estimated via Sanger sequencing into the insert within the cosmids using primers: M13 forward, M13 reverse, Cosmid check Ca2+ and Cosmid check parDE (Table 5). Cosmid 6B1.5 was found to contain around ~16,445 bp of DNA, including ~15,000 bp of DNA downstream of nodO (schematic shown in Figure 4.13). The insert within 8B4.5 mapped at one end to DNA located ~23,000 bp downstream of nodO; however, sequence data from the opposite flank of the cosmid mapped to a region that was located within the chromosome instead of the symbiosis island, suggesting that the insert within cosmid 8B4.5 was chimeric. The maximum insert size that can be accommodated is 30 kb, which would suggest that 8B4.5 may contain up to 7 kb of DNA that is from elsewhere in the genome.
Figure 4.13. Schematic of NZP2037 DNA present within the cosmids 6B1.5 and 8B4.5. The base position at the start of the \textit{nodO nod} box (30068) and the end of \textit{noeJ} (42750) are shown at the top. The cosmid 6B1.5 contains around 16,445 bp of DNA covering the whole deletion within Δ\textit{(nodO-9.3kb)} (base positions 28960-45405). The insert within 8B4.5 is known to contain \textit{nodO}, however, sequencing reads obtained from the vector mapped to loci not located in proximity to \textit{nodO} (position 3015645, in purple), suggesting that 8B4.5 was a chimeric insert. Note that the 8B4.5 has dashed lines to represent the insert size as it is too large to fit to the scale.

4.2.7.2 Effects of the cosmids 6B1.5 and 8B4.5 in Δ\textit{(nodO-9.3kb)} on the nodulation of \textit{L. pedunculatus}

The two cosmids 6B1.5 and 8B4.5 were transferred into Δ\textit{(nodO-9.3kb)} via conjugation and resultant strains tested for nodulation of \textit{L. pedunculatus}. The results showed that cosmid 6B1.5 complemented the defective nodulation phenotype of Δ\textit{(nodO-9.3kb)}, but in contrast cosmid 8B4.5 further suppressed nodulation (Figure 4.14). Nevertheless, the successful complementation with 6B1.5 suggested that a gene(s) lost in Δ\textit{(nodO-9.3kb)} was responsible for its defective nodulation phenotype on \textit{L. pedunculatus}. 
Figure 4.14. Effect of the cosmids 6B1.5 and 8B4.5 on *L. pedunculatus*. The presence of the cosmid 6B1.5 in $\Delta(nodO\text{-}9.3\text{kb})$ increased nodulation (green). In contrast, the presence of the cosmid 8B4.5 (black) generated a decrease in nodulation that was below that of the $\Delta(nodO\text{-}9.3\text{kb})$ mutant alone (red). NZP2037 is the wildtype control (blue). Percentage of plants nodulated was calculated from 15 plants per strain and is representative of the two experiments carried out.

4.2.7.3 Presence of the cosmids 6B1.5 and 8B4.5 in wildtype strains of NZP2037 and R7A have no major effects on the nodulation of *L. pedunculatus*

The cosmids 6B1.5 and 8B4.5 had opposite effects on the nodulation of *L. pedunculatus* when they were present within the $\Delta(nodO\text{-}9.3\text{kb})$ mutant background. It was hypothesized that these effects on nodulation may also be expressed in a dominant fashion in the wildtype background, and, conceivably, the genes encoded within these cosmids may be responsible for the extended host range of NZP2037. Therefore, the effects of the cosmids 6B1.5 and 8B4.5 in wildtype NZP2037 and wildtype R7A on nodulation of *L. pedunculatus* were investigated. No major effects were observed, with a slight delay in nodulation of *L. pedunculatus* by NZP2037 the only effect observed (Figure 4.15A, B).
Figure 4.15. Effect of cosmids 6B1.5 and 8B4.5 in the NZP2037 and R7A wildtype backgrounds on *L. pedunculatus*. (A) NZP2037 wildtype control (red), NZP2037 harbouring the cosmid 6B1.5 (green) and NZP2037 harbouring the cosmid 8B4.5 (blue). (B) R7A wildtype (orange) is unable to form effective nitrogen fixing nodules on *L. pedunculatus*, even when harbouring the cosmid 6B1.5 (green) or cosmid 8B4.5 (blue), NZP2037 (red) is the positive control. Percentage of plants nodulated was calculated from 15 plants per strain and is representative of the two experiments carried out.

4.2.8 Generation and effect of a single markerless *nodO* deletion in NZP2037

The mutant Δ(*nodO*-9.3kb) was originally developed due to the presence of the highly repetitive DNA within the *nodO* loci that gave rise to conflicting genomic assemblies within this region. With the knowledge gathered from the DNA sequencing performed on the plasmid pFAJ1700-*nodO*prsDE, it was concluded that the Japanese group had the correct assembly and the possibility to disrupt only *nodO* became available. Therefore, the working hypothesis was that the disruption of only *nodO* should generate a mutant that would exhibit the characteristic defective nodulation phenotype of Δ(*nodO*-9.3kb).
4.2.8.1 Generation of a *nodO* markerless deletion mutant

The primers pairs NodOLLapa1 and nodOsinglegeneLR, and nodOsinglegeneRL and nodOsinglegeneRR (Table 5), were used to amplify DNA regions of around 1 kb that flanked the 5’ and 3’ ends of *nodO*, respectively. The 1 kb PCR products were then used as templates for an overlap extension PCR with the primer pair NodOLLapa1 and nodOsinglegeneRR (Table 5) to generate a construct of ~2 kb that was then ligated into the suicide vector pJQ200SK to generate the construct pJQ200SK-nodOsingle. Successful constructs were verified using a restriction digest assay with the restriction enzyme *Asp*718. From the restriction digest assay, 3 expected bands with sizes of 4734 bp, 1799 bp, and 1128 bp are shown (Figure 4.16A). Plasmid map that for expected band sizes in a successful vector is shown in Figure 4.17. The construct pJQ200SK-nodOsingle was purified and sent for DNA sequencing and the integrity of the insert was verified.

The suicide plasmid pJQ200SK-nodOsingle was transferred into the wildtype NZP2037 background via conjugation, and strains that had undergone homologous recombination were selected for on media containing gentamicin. Mutants that had lost *nodO* via a second round of homologous recombination were then selected for by plating on media containing sucrose and screening was performed using the primer pair nodOCheckL and nodOsinglecheckR (Table 5), with a PCR product of around 470 bp expected in a successful deletion mutant (Figure 4.16B). The 470 bp PCR product was then purified and sent out for DNA sequencing to confirm the sequence integrity of the mutant.
Figure 4.16 Generation of the deletion construct pJQ200SK-nodOsingle and PCR confirmation of the nodO single deletion mutant. A) Restriction digest of pJQ200SK-nodOsingle with the restriction enzyme Asp718. L = ladder with sizes noted on the left. P = plasmid digested with Asp718, arrows indicates the bands of sizes: 4734 bp, 1799 bp, and 1128 bp (indicated by red arrows in descending order). B) PCR confirmation of NZP2037 ΔnodO was screened with the primer pair nodOcheckL and nodOsinglecheckR. Wt = wildtype which gives the expected size of 1343 bp (indicated by red arrow in lane). M = mutant that gives the expected product size of 470 bp (indicated by red arrow in lane) in a successful mutant.
4.2.8.2 Deletion of \( \text{nodO} \) has no effect on the nodulation of \( \text{L. pedunculatus} \)

The mutant strain that had lost \( \text{nodO} \) in a markerless deletion was designated NZP2037 \( \Delta \text{nodO} \) and was used as the inoculum for plant nodulation experiments. Unexpectedly, plants inoculated with NZP2037 \( \Delta \text{nodO} \) exhibited a phenotype similar to that of the wildtype (Figure 4.18). These results suggested that \( \text{nodO} \), by itself, was not essential for the nodulation of \( \text{L. pedunculatus} \) and that the defective nodulation phenotype exhibited by \( \Delta(\text{nodO}-9.3\text{kb}) \) was likely due to other gene(s) that were deleted in the 9.3-kb deletion.
Figure 4.18. The effect of a markerless nodO deletion in NZP2037 on *L. pedunculatus*. The wildtype control (blue) and the single nodO mutant (green) showed similar nodulation kinetics, while Δ(nodO-9.3kb) (red) had the defective nodulation phenotype. Percentage of plants nodulated was calculated from 15 plants per strain and is representative of the two experiments carried out.

### 4.2.9 The genes noeKJ as the next target for mutagenesis

Given the lack of phenotype of the nodO deletion strain and the lack of complementation of the defective Δ(nodO-9.3kb) phenotype by pFAJ1700-nodOprsDE, it was hypothesised that the removal of the 9.3-kb region of DNA may have instead disrupted the expression of gene(s) downstream of the deletion. The hypothesis of a polar mutation in the downstream genes noeKJ was plausible given that the mutant Δ(nodO-9.3kb) had been made with a construct that recombined within the nod box (17 bp overlap directly with the nod box) of the promoter region of noeKJ, thus potentially affecting the expression of these genes. Furthermore, the hypothesis that polar mutations of the genes noeKJ was attractive because the noeKJ genes are involved in the synthesis of precursors involved in the acetyl-fucosyl group that is transferred onto the non-reducing GlcNAc of the Nod factor, and this acetyl-fucosyl group was known to be required for the nodulation
of *L. pedunculatus* (Scott et al., 1996). To test the hypothesis that a polar mutation of *noeKJ* was responsible for the phenotype, the genes *noeKJ* were cloned and used in an attempt to complement the Δ(*nodO*-9.3kb) mutant. In addition, a *noeKJ* deletion mutant was also isolated in an attempt to replicate the defective nodulation phenotype of Δ(*nodO*-9.3kb). The results of these experiments are detailed below.

### 4.2.9.1 Deletion of *noeKJ* in NZP2037

PCR was performed to generate ~500 bp DNA products that flanked DNA regions proximal to *noeKJ* using primer pairs noeKJLL with noeKJLR and noeKJRL with noeKJRR (Table 5) for the 3’ and 5’ end flanks, respectively. The PCR products were then ligated into the MCS of the suicide vector pJQ200SK via Gibson cloning to generate the vector pJQ200SK-noeKJ. The *noeKJ* deletion mutant was then constructed essentially as described above for the *nodO* mutant (section 4.2.8.1). The final mutant was verified by PCR (Figure 4.19) and Sanger sequencing of a PCR product obtained from genomic DNA using the primers NoeKJflankcheckL and NoeKJinternalcheckR (Table 5). The mutant that was confirmed to have lost *noeKJ* was designated as NZP2037 Δ*noeKJ* and was used as inoculum for plant experiments. No defect in nodulation was detected (Figure 4.20).

### 4.2.9.2 *noeKJ* do not complement Δ(*nodO*-9.3kb)

Amplification of *noeKJ* was performed via PCR with the primer pair noeKJcloneL and noeKJcloneR and the resulting PCR product was then ligated into vector pFAJ1700 to form pFAJ1700-noeKJ. The construction and verification of pFAJ1700-noeKJ was performed by Dr. John Sullivan, and I then transferred the plasmid into both wildtype NZP2037 and the Δ(*nodO*-9.3kb) mutant via conjugation and the strains were assessed for nodulation. No complementation was observed (Figure 4.20).
Figure 4.19. PCR confirmation of a markerless deletion mutant of *noeKJ*. L = Ladder with band sizes denoted to the left of the ladder. wt = wildtype control, 1-5 = potential strains numbered 1 through to 5, that were screened for the deletion of *noeKJ*. The primer pairs used for the screen are shown above the potential strains with the primer pair NoeKJflankcheckL and NoeKJflankcheckR on the left side of the gel with a predicted size of 623 bp in the mutant and 3.2 kb in the wildtype. The PCR products for the primer pair NoeKJflankcheckL and NoeKJinternalcheckR produces a 673 bp product only in the wildtype and no product in a successful *noeKJ* deletion mutant. Red boxes indicate expected products (or lack of) in a successful *noeKJ* mutant (in this example, it is mutant #1).
4.2.10 The next target, the Type I secretion system genes *prsDE*

Initially, due to the inability of the construct pFAJ1700-nodO*prsDE* that contained *prsDE* to restore the defective nodulation phenotype of \(\Delta(nodO-9.3kb)\) to wildtype levels (detailed in section 4.2.6), it was thought that the transport system was unlikely to be involved in nodulation. However, there was a possibility that the *prsDE* genes within the plasmid were not expressed properly and/or that effector proteins transported by PrsDE were not represented. Therefore, the genes *prsDE* were targeted for disruption in an attempt to replicate the nodulation phenotype of \(\Delta(nodO-9.3kb)\).

4.2.10.1 Deletion of *prsDE*

PCR was used to generate DNA regions of around 1 kb that flanked the 3’ and 5’ flanks of *prsDE* with the primer pairs prsDEdeletionLL and prsDEdeletionLR and prsDEdeletionRL and prsDEdeletionRR, respectively (Table 5). The two PCR products were then ligated into pJQ200SK via Gibson cloning to generate the plasmid pJQ200SK-
prsDE. The *prsDE* deletion mutant was then constructed essentially as described above for the *nodO* mutant (section 4.2.8.1).

Potential mutants were screened with the primer pair *prsDE* CheckL and *prsDE* CheckR (Table 5) that bound to the DNA regions that flanked the 5’ and 3’ end of *prsDE*, respectively. The expected PCR product produced from the primer pair *prsDE* CheckL and *prsDE* CheckR is a 1000 bp product that is only present within a successful deletion mutant. Another primer pair *prsDE* CheckL and *prsDE* internalCheckR (Table 5) which binds to the 3’ flanking region of *prsDE* and within the start of *prsD*, respectively, produces a 1015 bp product only in the wildtype strain. The results from the PCR screen showed that both PCR products were present (Figure 4.21), and Sanger sequencing confirmed both PCR products. This result suggested that the mutant had lost most of *prsDE* but retained the 3’ end of *prsD*. However, this mutant exhibited a phenotype on *L. pedunculatus* (detailed in Section 4.2.10.2) that was similar to Δ(*nodO*-9.3kb), suggesting that mutation(s) had occurred within the gene(s) responsible for the defective nodulation phenotype. Thus, this mutant was designated as NZP2037 Δ*prsDE* and its genotype was investigated in more detail below.

To elucidate the mutation(s) that had occurred within NZP2037 Δ*prsDE*, a whole genome sequencing approach using the Illumina MiSeq platform was used. The benefits of this approach were that whole genome sequencing would identify all mutations that had occurred within the mutant, and thus was not limited to the loci of *prsDE*. Genomic DNA was purified using the UltraClean® Microbial DNA Isolation Kit and whole genome sequencing was performed by Mr. DNA (Molecular Research LP, Shallowater, Texas USA). Reads obtained from the whole genome sequencing were aligned to the genome.
of NZP2037 with 216 x coverage of the genome and differences between NZP2037 \( \Delta prsDE \) and the wildtype NZP2037 were identified.

**Figure 4.21. PCR results for NZP2037 \( \Delta prsDE \).** Lanes from left to right. L = ladder from HindIII digested \( \lambda \) phage and HaeIII digested \( \phi X174 \). Lanes 1 and 2 represent the wildtype (NZP2037) screened with the primers \( prsDEcheckL + prsDEcheckR \) and \( prsDEcheckL + prsDEinternalcheckR \), respectively. Lanes 3 and 4 represent NZP2037 \( \Delta prsDE \) screened with the primers \( prsDEcheckL + prsDEcheckR \) and \( prsDEcheckL + prsDEinternalcheckR \), respectively. Expected band size for the primer pair \( prsDEcheckL + prsDEcheckR \) in the wildtype is 4019 bp while in a mutant the expected product size is 1000 bp. For the primer pair \( prsDEcheckL + prsDEinternalcheckR \) the expected size for the wildtype is 1015 bp while no product should be present in a markerless deletion mutant.
Genetic differences in NZP2037 ΔprsDE localized to the prsDE locus. No reads aligned with prsE, suggesting that it had been deleted completely in the mutant. However, some reads aligned to the 5’ end of prsD, suggesting that the removal of both target genes prsDE had not occurred. A contig that represents the genetic orientation within the prsDE locus in NZP2037 ΔprsDE was examined in detail and is represented in a schematic (Figure 4.22D).

From the schematic, it was deduced that NZP2037 ΔprsDE contained a truncated prsD, which had lost 341 bp at the 3’ end of the gene; additionally, the whole of the prsE gene was lost. This deletion was then followed by fragments of the suicide vector that flanked the insert in the vector, perhaps due to recombination events within a dimer or from a duplication event followed by a second recombination event. Further downstream, the expected full prsDE deletion was present.

Taken together, the results from the whole genome sequencing correlated with the data obtained from Sanger sequencing and the PCR analysis, with both PCR products represented within the mutant due to the duplication of the 5’ flanking region within NZP2037 ΔprsDE. The mutant strain NZP2037 ΔprsDE was therefore a genuine mutant that was disrupted in prsDE. The whole prsE gene had been lost, and truncation of prsD was likely to have generated a non-functional PrsD. Importantly, the mutant would be unable to revert back to the wildtype genetic orientation from subsequent recombination events.
Figure 4.22. Schematic of the genetic orientation present within NZP2037 ΔprsDE. A) The suicide vector pJQ200-prsDE, the 5’ flank (green box) and the 3’ flank (purple box) of prsDE are located within the MCS of the suicide vector. The red box and yellow boxes represent sequence proximal to the flanks that are unique to the vector. B) Orientation of prsDE within the wildtype, the green box indicates the 5’ flank and the purple box represents the 3’ flank. The green arrow indicates the binding site of the primer prsDEcheckL, the blue arrow indicates binding site of the primer prsDEcheckR. C) An expected ΔprsDE mutant, the green arrow represents the binding site of the primer prsDEcheckL while the purple arrow indicates binding site of the primer prsDEinternalcheckR. D) Genetic orientation within NZP2037 ΔprsDE. The deletion removed around 341 bases off the 3’ end of prsD and the whole of prsE, followed by sequence that corresponds to the regions of the suicide vector ending in the expected markerless deletion.
4.2.10.2 Phenotype of NZP2037 ΔprsDE on *L. pedunculatus*

Strain NZP2037 ΔprsDE showed a similar nodulation on *L. pedunculatus* to the Δ(*nodO*-9.3kb) mutant (Figure 4.23), suggesting that the defective nodulation phenotype of Δ(*nodO*-9.3kb) was due to disruption of *prsDE*.

![Nodulation kinetics of NZP2037 ΔprsDE on L. pedunculatus.](image)

*Figure 4.23. Nodulation kinetics of NZP2037 ΔprsDE on L. pedunculatus.* Strains used as inocula are indicated in the key. Percentage of plants was calculated from 15 plants per strain and is representative of the two experiments carried out.

4.2.10.3 Complementation of NZP2037 ΔprsDE

The defective nodulation phenotype of Δ(*nodO*-9.3kb) could be restored to wildtype levels by the cosmid 6B1.5 but not by plasmid pFAJ1700-nodOprsDE (section 4.2.6). It was of interest to determine the effect of 6B1.5 and pFAJ1700-nodOprsDE in the NZP2037 ΔprsDE background. Therefore, 6B1.5 and pFAJ1700-nodOprsDE were transferred into NZP2037 ΔprsDE via conjugation and results similar to those in Δ(*nodO*-9.3kb) were observed (Figure 4.24).
Figure 4.24. Complementation of NZP2037 ΔprsDE with the cosmid 6B1.5 and pFAJ1700-nodOprsDE. Strains used as inocula are as indicated by the key. Note that pFAJ1700-nodO represents the plasmid pFAJ1700-nodOprsDE. Percentage of plants was calculated from 15 plants per strain and is representative of the two experiments carried out.

The results support the hypothesis that the phenotype of Δ(nodO-9.3kb) was due to the disruption of prsDE. Additionally, the inability of pFAJ1700-nodOprsDE to rescue the defective nodulation phenotype of NZP2037 ΔprsDE suggested that the plasmid was faulty in gene expression, or that gene(s) downstream of prsDE also contributed to the phenotype.
4.2.10.4 Homologues of PrsDE in NZP2037 and potential effectors

With the knowledge that PrsDE but not NodO was involved in the promotion of nodulation of *L. pedunculatus*, it was of interest to identify potential homologues of PrsDE present within NZP2037, and in turn identify other potential systems that may be unique to NZP2037. PrsD encodes an ABC-transporter, which is commonly associated with proteins that hydrolyse ATP as an energy source. Consequently, if PrsD was used as a search query, matches with multiple different predicted homologues that are most likely not associated with T1SS would be retrieved. PrsE homologues are usually restricted to T1SS and use of PrsE as the search query gave 8 potential protein homologues. Two of these were removed as T1SS candidates as they were not genetically linked to ABC-transporters and therefore, unlikely be part of a functional T1SS. The remaining 6 PrsE homologue candidates are listed in Table 7. A similar search for PrsE homologues in the R7A genome returned 4 homologues, all of which showed 100% protein identity to corresponding NZP2037 counterparts, suggesting that 2 putative T1SS were unique to the NZP2037 genome.

The two T1SS present in NZP2037 but not R7A were also found to be present within strains that belong to the broad host range *M. loti* group (NZP2042, NZP2014, NZP2037 and SU343). These two PrsE homologues were the original PrsE query that had been targeted for mutagenesis in this study, and a conserved hypothetical with the IMG Gene ID: 2723039782. This new T1SS (*mln388/mln399*, which encode the ABC-transporter and MFP, respectively) was located within the symbiosis island, suggesting a potential role in symbiosis. However, the promoter region of this secretion system lacks an obvious *nod*-box and therefore the system is unlikely to be under NodD regulation.
Table 7. Homologues of PrsE within NZP2037 and their corresponding homologue in R7A

<table>
<thead>
<tr>
<th>PrsE homologue in NZP2037</th>
<th>% amino acid identity to PrsE</th>
<th>Homologue in R7A</th>
</tr>
</thead>
<tbody>
<tr>
<td>PrsE</td>
<td>100</td>
<td>none</td>
</tr>
<tr>
<td><em>mlr0589 homologue</em></td>
<td>23</td>
<td>*mlr0589 homologue, IMG gene ID:2512930751. 100% identity. Locus tag: MesloDRAFT_2016. *</td>
</tr>
<tr>
<td>Conserved hypothetical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMG Gene ID: 2723036052*</td>
<td>28</td>
<td>IMG Gene ID: 2512928739, 100% identity. Locus tag: MesloDRAFT_0004. *</td>
</tr>
<tr>
<td><em>mll2582 homologue</em></td>
<td>35</td>
<td>*mll2582 homologue, IMG gene ID: 2512932237. 100% identity. Locus tag MesloDRAFT_3502. *</td>
</tr>
<tr>
<td>Conserved hypothetical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMG Gene ID: 2723039782</td>
<td>24</td>
<td>*mlr6772 homologue, IMG Gene ID:251292873. 100% identity. Locus tag: MesloDRAFT_0238. *</td>
</tr>
<tr>
<td>(mln389)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Encoded on the chromosome

In addition to a search for additional T1SS, a search for potential effectors likely to be secreted by PrsDE was performed. Several proteins were used as search queries. The first search utilised the BLASTP algorithm with the RTX Ca<sup>2+</sup> binding domain from NodO as the search query, as this domain is commonly located within the N-terminal region of proteins that are exported by T1SS. The domain consists of repeats of a nonapeptide Ca<sup>2+</sup> binding motif that is present amongst proteins confirmed to be secreted by T1SS in rhizobia such as NodO and ExsH in *R. leguminosarum* and *S. meliloti*, respectively (Economou *et al.*, 1990, York & Walker, 1997, Krehenbrink & Downie, 2008). The RTX domain is also present in other bacterial Type I effectors such as HlyA in *E. coli*, CyaA in *Bordella pertussis* and RtxA in *Vibrio cholerae* (Welch, 2001).
The results from the BLASTP search identified 3 candidates besides NodO that contained the RTX Ca\(^{2+}\) binding domain (Table 8). The proteins were annotated as probable haemolysin-type calcium binding proteins. Of these 3 proteins, only Mln031 (also known as IMG Gene ID: 2723040143) was genetically linked to a T1SS (prsDE). Interestingly, \textit{mln031} was truncated within pFAJ1700-nodOprsDE, lacking around 60 nucleotides of sequence, and the vector was unable to restore the defective nodulation phenotype of \(\Delta(nodO-9.3kb)\). This observation highlighted Mln031 as a potential effector protein transported by PrsDE involved in the promotion of nodulation on \textit{L. pedunculatus}. Searches using other proteins as queries that are known to be transported by T1SS in rhizobia such as the adhesins RapA1, RapB, RapC involved in biofilm formation (Russo \textit{et al.}, 2006, Mongiardini \textit{et al.}, 2008) were also performed, but no potential homologues of these adhesion proteins were identified within the NZP2037 genome. Similarly, homologues of previously documented glycanases that are secreted by PrsDE, PlyA and PlyB that are involved in the cleavage of EPS and biofilm formation (Finnie \textit{et al.}, 1997), were also not identified. A homologue of another Type I secretion mediated glycanase, ExsH, was found to be encoded within the genomes of both NZP2037 and R7A, suggesting that it was unlikely to play a role in host-specificity of \textit{L. pedunculatus}. 


Table 8. Predicted putative effectors of T1SS containing a RTX Ca\(^{2+}\) binding domain

<table>
<thead>
<tr>
<th>Protein / IMG Gene ID</th>
<th>Amino acid % Identity</th>
<th>Associated Type I secretion system</th>
<th>Conserved domains / Predicted functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>NodO 2723040026</td>
<td>100</td>
<td>PrsDE</td>
<td>probable nodulation protein, NodO</td>
</tr>
<tr>
<td>(mln145)</td>
<td></td>
<td></td>
<td>Hemolysin-type calcium binding protein repeat containing protein</td>
</tr>
<tr>
<td>2723040101* 37</td>
<td></td>
<td>none</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>2723040143* 35</td>
<td></td>
<td>PrsDE</td>
<td>Hemolysin-type calcium binding protein repeat containing protein</td>
</tr>
<tr>
<td>(mln031)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*no corresponding mln number

4.2.11 Complementation of NZP2037 ΔprsDE with a functional mln031

In this study, mln031 was previously annotated as a pseudogene and it was thought to encode a truncated non-functional version of NodO (alignment shown in Figure 4.25) with the two proteins containing the RTX Ca\(^{2+}\) binding domain (residues 41-128). Differences between NodO and Mln031 include NodO having more Ca\(^{2+}\) binding motifs and a larger C-terminal domain. However, with the finding that deletion of nodO had no effect on nodulation of L. pedunculatus and that Mln031 was a candidate that was highly likely to be exported by a T1SS, in addition to its similarities to NodO, it was hypothesised that Mln031 may be the effector of interest transported by PrsDE. To test this hypothesis, a complementation approach was taken using a plasmid that encoded prsDE and mln031 in the hopes that this construct could restore the defective nodulation of NZP2037 ΔprsDE to wildtype levels.
173

Figure 4.25. Protein alignment of Mln031 to NodO. Protein sequence of Mln031 (top) aligned to that of NodO (bottom). Mln031 was thought to be a truncated version of NodO due to the high similarities in the N-terminal domain.

4.2.11.1 Construction of the complementation plasmid pFAJ1700-nodOprsDEmIn031

As the plasmid pFAJ1700-nodOprsDE had previously been DNA sequence verified, it was decided that pFAJ1700-nodOprsDE would be used as the starting point to generate a new complementation vector with a full length mln031. A ligation approach utilising a gBlock (synthetic double stranded DNA fragment) was chosen to restore mln031 to its full length. The method to achieve this was to exploit the exonuclease activity present within the Gibson assembly kit to chew back nucleotides for removal of the HindIII restriction site. Subsequent ligation of the gBlock to the 3’ end of the original pFAJ1700-nodOprsDE insert would then restore the full length mln031 (approach detailed in Figure 4.26).
The plasmid pFAJ1700-nodOprsDE was first linearized with the restriction enzyme *BamHI* and then gel purified. The gBlock Ca2+ bp (Table 6) was then ligated into the plasmid using Gibson assembly. The successful ligation of the gBlock into the vector was screened via PCR with the primer pair Ca2+_check and faj3 (Table 5), that bound within the new insert and to the backbone of the pFAJ1700 vector, respectively. This primer pair produces a PCR product with a size 454 bp (Figure 4.27) in a successful construct that had integrated the gBlock and no product would be generated in the original vector. The insert within the new vector was verified using Sanger sequencing and designated as pFAJ1700-nodOprsDEmln031.

![Figure 4.26. Approach to restore full length *mln031* in the construct pFAJ1700-nodOprsDE.](image)

A) Genetic orientation at the site of interest in pFAJ1700-nodOprsDE. The 3’ end of the truncated *mln031* is shown in green. The *HindIII* and *BamHI* restriction sites are shown in red and purple, respectively. B) Restriction digest with *BamHI*. C) Exonuclease activity removing the *HindIII* cut site. D) Ligation of the gBlock (blue) upon completion of Gibson assembly, restoring the full length *mln031*. 

174
Figure 4.27. PCR screen for a successful ligation of the plasmid pFAJ1700-nodO prsDEmln031. Lanes shown here are; (M) represents Marker made from HindIII digested λ phage and HaeIII digested φX174, (-) represents the negative control and (+) represents the expected 454 bp product in a restored a full length mln031.

4.2.11.2 Complementation of NZP2037 ΔprsDE with pFAJ1700-nodOprsDEmln031

The presence of pFAJ1700-nodOprsDEmln031 restored the nodulation phenotype of NZP2037 ΔprsDE to levels comparable to a strain that harboured the cosmid 6B1.5 (Figure 4.28). The successful restoration of the nodulation phenotype supports the hypothesis that mln031 encoded the effector protein secreted by PrsDE that is involved in the promotion of nodulation of the host L. pedunculatus.
Figure 4.28. Complementation of NZP2037 ΔprsDE with the vector pFAJ1700-nodOprsDEmln031.

Strains used as inocula are as indicated in the key. Note that pFAJ1700-nodO represents the plasmid pFAJ1700-nodOprsDE, while pFAJ1700-nodOCa²⁺ represents the plasmid pFAJ1700-nodOprsDEmln031. Percentage of plants was calculated from 15 plants per strain and is representative of the two experiments carried out.
4.3 Discussion

The deletion of a 9.3-kb region in *M. loti* NZP2037 including the locus containing *nodO* ($\Delta(nodO-9.3kb)$) generated a defective nodulation phenotype on *L. pedunculatus*. The defective nodulation phenotype occurred with $\sim$50% of the plants, which only formed uninfected nodule primordia. The other $\sim$50% of plants formed effective nitrogen fixing nodules after a delay. The deletion of this 9.3-kb region of DNA did not have a major effect on the nodulation of *L. japonicus* Gifu, indicating that gene(s) within the 9.3-kb deleted region were likely involved in the promotion of nodulation of *L. pedunculatus* though they are not essential for the establishment of an effective symbiotic relationship.

Characterisation of the defective nodulation phenotype exhibited by $\Delta(nodO-9.3kb)$ on *L. pedunculatus*, suggested that the block in nodulation occurred at the early stages of infection, likely infection thread initiation. Testing of new roots regenerated from previously nodulated plants showed that natural genetic variation within the plant population did not have a major role on the outcome of nodulation. Additionally, bacteria recovered from effective nodules maintained the $\Delta(nodO-9.3kb)$ nodulation phenotype, indicating that secondary mutations that may have arisen from selection pressures were not responsible for the ability of $\Delta(nodO-9.3kb)$ to form effective nodules. Mutations that caused truncated EPS in NZP2037 did not affect the nodulation of $\Delta(nodO-9.3kb)$ or wildtype on *L. pedunculatus*, suggesting that EPS has a minimal role (if any) in the signalling events that occurs between this host and NZP2037. In addition, an *exoU* $\Delta(nodO-9.3kb)$ double mutant of NZP2037 showed the same phenotype as an *exoU* mutant on *L. japonicus* Gifu, indicating that genes located within $\Delta(nodO-9.3kb)$ region were not essential for circumvention of the negative effect of defective EPS on *L. japonicus* Gifu.
The 9.3-kb deletion mutant was originally made to remove repetitive DNA downstream of the *nodO* locus that gave rise to conflicting sequence assemblies and may have caused future unwanted genetic instability. Eventually, it was found that *nodO* (and by extension *prsDE* which were thought to encode the secretion system for NodO) was not responsible for the nodulation defect observed on *L. pedunculatus*. This shifted the targets towards genes located downstream of the Δ(*nodO*-9.3kb) region that may have been disrupted by the deletion. However, it was found that the downstream genes *noeKJ* were not required for the nodulation of *L. pedunculatus*. This lead to a further search for targets within the 9.3-kb region itself. A mutant of the T1SS PrsDE was found to reproduce the nodulation phenotype of Δ(*nodO*-9.3kb) on *L. pedunculatus* suggesting that PrsDE must secrete an effector that was an important factor for the nodulation of this host. The nodulation phenotype was restored in NZP2037 Δ*prsDE* mutants with the cosmid 6B1.5 and the plasmid pFAJ1700-nodOprsDEmln031, but not the plasmid pFAJ1700-nodOprsDE which encoded a truncated *mln031*. These results showed that the defective nodulation phenotype of Δ(*nodO*-9.3kb) on *L. pedunculatus* was due to disruption of *prsDE* and *mln031* which encode a T1SS and a likely effector, respectively.

The EPS of NZP2037 does not appear to be involved in the signalling events that lead to nodulation on *L. pedunculatus*. This result was not unexpected as transposon generated EPS mutants of NZP2037 had previously been shown to be fully effective on *L. pedunculatus* (Hotter & Scott, 1991). EPS is known to have a role as a signalling molecule in some symbiotic relationships such as in *M. truncatula* (Jones *et al.*, 2007) and *L. japonicus* Gifu (Kawaharada *et al.*, 2015). If we were to assume that EPS were to play a role in *L. pedunculatus* nodulation, it would likely occur downstream of where the
mutation in Δ(nodO-9.3kb) disrupts nodulation as double mutants of exoU and Δ(nodO-9.3kb) displayed a phenotype similar to that of the Δ(nodO-9.3kb) mutant alone. Further discussion on the roles of EPS are presented in Chapter 3, which culminated in a mutant that lacked all unique genes of interest and had a phenotype comparable to a Δ(nodO-9.3kb) deletion alone.

The genetic differences between the plasmids pFAJ1700-nodOprsDE and pFAJ1700-nodOprsDEmln031 suggest that the truncation of 20 aa at the C-terminal end of Mln031 was sufficient to eliminate its activity. The loss of the 20 aa may prevent the secretion of Mln031 by PrsDE, as the final 48-60 aa of Type I effectors are known to contain the secretion signal required for Type I secreted proteins (Thomas et al., 2014). It should be noted that there is no universal consensus sequence for the secretion signal of Type I secreted effectors and it is thought that the formation of secondary structures such as amphipathic helices and helix-loop-helix motifs at the C-terminus of these effectors are what is recognised by the T1SS machinery (Welch, 2001, Thomas et al., 2014). Thus, in Mln031 it is hard to pinpoint the exact location of the secretion signal based on the primary sequence alone. What can be inferred from the truncation of Mln031 was that restoration of the 20 aa likely restored the secretion signal and so important secretion signal residues must reside within. A comparison between extracellular protein secretion profiles under conditions in which the nod genes are expressed between prsDE mutant strains that harbour pFAJ1700-nodOprsDE and pFAJ1700-nodOprsDEmln031 should show the secretion of Mln031 in strains that harboured pFAJ1700-nodOprsDEmln031.

Mln031 is predicted to be composed of 208 aa and prediction for conserved protein domains found only the RTX Ca\(^{2+}\) binding domain (Figure 4.29A), while a BLAST search
for homologues found only Ca\(^{2+}\) binding proteins with no clear function in relation to symbiosis (alignment to top homologues shown in Figure 4.29B). Homologues of Mln031 with 99-100% identity are conserved within the *L. pedunculatus* compatible *M. loti* strains NZP2014, NZP2042, and SU343 (represented by NZP2014 in the alignment) and are missing from other *M. loti* strains, suggesting its importance. Interestingly, besides the *M. loti* encoded Mln031, the top homologues found were encoded in *Microvirga vignae* and *Bosea lupini* both of which were initially isolated from root nodules (De Meyer & Willems, 2012, Radl et al., 2014). These observations suggest that Mln031 homologues are likely to be conserved amongst plant associated bacteria and provides support for Mln031 as an effector for plant associated interactions. However, none of these homologues have been studied in detail and their effects on plant-associated interactions have yet to be elucidated. Taken together, clues as to the mechanism of action for Mln031 could not be found based on primary protein sequence or comparison to its closest protein homologues.

The finding in this chapter that a disrupted *nodO* was unable to replicate the defective phenotype of Δ(*nodO*-9.3kb) was unexpected, as *nodO* was the most likely gene target within the 9.3-kb of DNA. NodO is involved in the transfer of cations across membranes and hypothesised to promote ion flux and depolarization of the host cell membrane and, in turn, promote nodulation (Economou et al., 1994, Walker & Downie, 2000, Miwa et al., 2006). In hindsight, it was fortunate that the design of a single *nodO* mutation could not be performed at the start of this study due highly repetitive DNA region downstream of *nodO*. The experiments detailed in this chapter suggest that NodO has minimal effect on the nodulation of *L. pedunculatus* and this lead to the subsequent finding of PrsDE/Mln031 as major components in the determination of *L. pedunculatus* nodulation.
Figure 4.29. Predicted conserved domains of Mln031 and alignment to homologues. A) Domains predicted in Mln031 belong to the RTX Ca^{2+} binding domain consisting of 5 repeat motifs (light blue boxes), the 1<sup>st</sup> repeat begins at residue 41-56, 2<sup>nd</sup> at 59-75, 3<sup>rd</sup> at 76-93, 4<sup>th</sup> at 94-111, and the 5<sup>th</sup> at 113-128. B) Alignment of Mln031 to the top 4 closest homologues. 1. *M. loti* NZP2037, 2. *M. loti* NZP2014, 3. *Microvirga vignae*. 4. *Bosea lupini*.
The observation that a *nodO* single mutation had no effect on nodulation is mirrored in *R. leguminosarum* bv. *viciae*, where NodO was only found to be required when in a double mutant with *nodE* that produces Nod factor that lacks the 18:4 unsaturated acyl chain (Economou *et al.*, 1994). This observation was what lead to the hypothesis that NodO was likely to be involved in the amplification of suboptimal ionic flux that is stimulated by the presence of suboptimal Nod factors during infection thread development, eventually leading to an effective symbiosis (Walker & Downie, 2000, Miwa *et al.*, 2006). In this study, the deletion of the 9.3-kb region was unlikely to have affected the synthesis of Nod factors and the only potential genes involved in Nod factor synthesis were the downstream genes *noeKJ*. Therefore, Δ(*nodO*-9.3kb) should produce Nod factor that is comparable to the wildtype. If Mln031 functions similarly as an amplifier for Nod factor signalling then this may suggest that Nod factors produced by NZP2037 are sub-optimal for the nodulation of *L. pedunculatus*, and the presence of Mln031 was required for efficient nodulation of this host.

Explanations as to why the single *nodO* mutant did not exhibit a defective nodulation phenotype could be simply that NodO has no effect on the nodulation of *L. pedunculatus*. Alternatively, the NodO of NZP2037 may be non-functional, due to an insertion within the RTX domain of the protein that may have affected its tertiary structure (Figure 4.30). The insertion differentiates NodO of NZP2037 from other rhizobial NodO proteins, supporting the suggestion that NodO may be non-functional. However, against this, *nodO* containing this insertion is present in the strains of *M. loti* compatible with *L. pedunculatus* (NZP2037, NZP2014, NZP2042, and SU343), although, NZP2014 also contains a second copy of *nodO* that is a homologue to the *R. leguminosarum nodO*. 

182

The presence of both forms of NodO may suggest that the two have different functions and the presence of both are beneficial to NZP2014. The lack of a common C-terminal region between NodO and Min031 suggests that the two proteins have different mechanisms of action and that NodO and Min031 may have evolved to have different roles. Interestingly, if we assume that NodO and Min031 share the same T1SS, then the
two should share the same secretion signal and the lack of an obvious consensus sequence in the C-terminal (apart from the conserved alanine at -7 and the terminal phenylalanine and valine residues) suggests that the secretion signal only requires a small number of conserved residues and, as discussed above, is based on a more general secondary structure motif.

The observation that ~50% of *L. pedunculatus* plants inoculated with Δ(nodO-9.3kb) still formed effective nitrogen fixing nodules after a delay, combined with the observation that infection thread initiation was affected by the 9.3-kb deletion, suggests that Δ(nodO-9.3kb) was disrupted in its ability to enter the host-root hair. This defective nodulation phenotype of Δ(nodO-9.3kb) is mirrored in a *nodO* transposon mutant in the *R. leguminosarum* bv. *trifolii* background where only ~40% of *Vicia sativa* seedlings nodulated, but no nodulation defective phenotypes were observed for the hosts *Vicia hirsuta* and *Trifolium subterraneum* (de Maagd *et al.*, 1989). The similarities between the phenotypes of the *M. loti* mutant and *R. leguminosarum* bv. *trifolii nodO* mutant supports the hypothesis that Mln031 functions similarly to NodO in the amplification of suboptimal Nod factor signalling. Hence, efficient nodulation of *L. pedunculatus* is achieved through overcoming a concentration threshold in Nod factor signalling. In the context of this model, Mln031 promotes nodulation through amplification of ionic flux signals mediated by Nod factor signalling. In an *mln031* mutant, these ionic flux signals do not reach the threshold in a timely manner at most colonised root tips and as a consequence formation of the infection thread does not occur. In support of this model is the observation that wildtype Nod factor of *S. meliloti* at concentrations of 1 nM was able to elicit the calcium spiking response but no calcium flux was observed, in contrast, 10 nM concentrations were able to elicit the calcium flux response (Shaw & Long, 2003). Similarly, a 100 fold increase of suboptimal Nod factor from *S. meliloti* that lacked an
acetyl group modification on the non-reducing terminal GlcNAc was required to generate a calcium influx response comparable to wildtype Nod factor on *M. truncatula* (Morieri *et al.*, 2013). This requirement for increased concentration correlated with a decreased activity for the suboptimal Nod factor in nodulation and a decreased ability to initiate infection threads, but not in their elongation (Ardourel *et al.*, 1994). This model also explains the delayed nodulation phenotype observed in the ‘escape’ plants inoculated with Δ(*nodO*-9.3kb) that could be interpreted as the time required before the accumulation of Nod factor concentrations overcomes the required threshold for initiation of the infection thread. It could be that continued Nod factor production in the growing infection thread provides enough Nod factor to allow further infection threads and nodules to form on plants once the first nodule is formed. A potential future experiment to test this hypothesis may be the addition of exogenous Nod factor in an attempt to increase the concentration of Nod factor which should restore the nodulation phenotype of a *mln031* mutant to wildtype levels assuming that greater concentrations of Nod factor could overcome the threshold required to achieve the ionic flux response. If the addition of exogenous Nod factor does not restore nodulation then this may suggest that Mln031 functions after entry into the plant cell similar to that observed for NodO (Walker & Downie, 2000). Alternatively, the focus could be on the direct observation of ionic flux into the cytoplasm of root hair cells when treated with *mln031* mutants compared to the wildtype.

Another potential mechanism for Mln031 in overcoming this threshold requirement may be achieved through intrinsic properties that support nodulation, such as adhesion of the rhizobial cell to the plant cell. Attachment of rhizobia to the root hairs of potential hosts is an important step in symbiosis, as effective attachment allows for the increased localization of rhizobial cell numbers on the root hair and increases the probability of
strains being infective. In addition, attachment promotes nodulation via the localised accumulation of Nod factor, amplifying the effects of Nod factor signalling, such as ionic flux into the root hair (Hirsch, 1999, Oldroyd & Downie, 2008). In *R. leguminosarum*, rhicadhesin is a 14kDa Ca$^{2+}$ binding protein involved in the attachment of bacteria to legume root hairs and is widely distributed amongst almost all rhizobial species (Smit et al., 1989). Three proteins with similarities to rhicadhesin, named RapA1, RapA2 and RapC, were identified as exported by the T1SS in *R. leguminosarum* bv. *trifolii* (Ausmees et al., 2001). Similarities between RapA1 and another adhesin BJ38, showed that both proteins localised to the polar regions of the cell and both had an involvement in cell agglutination (Ausmees et al., 2001). In *Bradyrhizobium*, mutants that lack BJ38 displayed a decreased nodulation ability when compared to the wildtype, suggesting that BJ38 played a role in the early stages of infection (Ho et al., 1994). BLASTP analysis did not identify homologues of RapA1, RapA2 and RapC in NZP2037 and indeed the Rap proteins are limited to *R. leguminosarum* and *R. etli* (Ausmees et al., 2001). It is possible that Mln031 is a functional homologue of a RapA1, as they share many similar qualities such as predicted Ca$^{2+}$ binding, export via T1SS and similar sizes with Mln031 and RapA1 having predicted sizes of ~22 kDa and ~24 kDa, respectively (Ausmees et al., 2001). Therefore, another hypothesis is that Mln031 acts as an adhesin whereby it indirectly acts to increase the localised concentration of Nod factors, thus amplifying the ionic flux leading to the formation of the infection thread. Experiments to test this hypothesis could involve the visualisation of bacterial colonisation on the plant root hair with different bacterial concentrations, where we would expect more bacterial colonisation for the wildtype compared to the *mln031* mutant. Alternatively, an experiment similar to that of Smit et al. (1989) could be used where bacterial suspensions were incubated with plant hair roots and then washed in phosphate buffer to remove non-
specific binding and remaining cells visualised under a microscope

Other future experiments may focus on the purification and characterisation of Mln031 itself. It would be of interest to observe the effects that the addition of purified Mln031 may have on the nodulation of *L. pedunculatus*. If Mln031 functioned as an adhesin it may compete for limited binding sites on the root hair and plants treated with purified Mln031 prior to inoculation may exhibit a decrease in infection due to loss of these binding sites. In contrast, if Mln031 was involved in amplification of the ionic flux, the addition of purified Mln031 directly to the plant may increase the influx signal and could restore nodulation in a prsDE mutant by bypassing the need for the T1SS entirely. In addition, this would provide support that Mln031 acts exogenously.

The cosmids 6B1.5 and 8B4.5 had differing effects on the nodulation of *L. pedunculatus* only in the Δ(*nodO*-9.3kb) background. The presence of 6B1.5 restored an effective nodulation phenotype to Δ(*nodO*-9.3kb) and NZP2037 ΔprsDE. In contrast, the presence of 8B4.5 further decreased the strains already impaired ability to nodulate *L. pedunculatus*. The cosmid 8B4.5 is a chimeric construct that contains *nodO* within its insert, however, 8B4.5 may encode other proteins that may be detrimental for nodulation of *L. pedunculatus*. Alternatively, if we assume that 8B4.5 encoded effectors of PrsDE, the accumulation of these effectors within the cell coupled with the lack of a functional PrsDE may generate adverse effects on nodulation of *L. pedunculatus*. The presence of 8B4.5 within the wildtype does not exhibit a defective phenotype, in contrast to the defective phenotype observed in Δ(*nodO*-9.3kb), and this may suggest that a functional PrsDE could alleviate the negative effects of gene(s) encoded on 8B4.5.
The presence of 6B1.5 was unable to expand the host range of wildtype R7A to include that of *L. pedunculatus*. These results would suggest that other NZP2037 specific factors not encoded within 6B1.5 are involved in the establishment of symbiosis with *L. pedunculatus* and these would be required to increase the host-range of R7A to include *L. pedunculatus*. Perhaps, R7A is unable to survive within the harsh root environment of *L. pedunculatus*, which is known to produce prodelphinidin-rich flavolans (condensed tannins), that have bactericidal effects at high concentrations (Pankhurst *et al.*, 1979, Jones *et al.*, 1987). NZP2037 produces a flavolan binding polysaccharide likely to be covalently linked to the peptidoglycan layer and this provides protection against the bactericidal effects of the flavolan (Jones *et al.*, 1987). It is likely that flavolan concentrations within *L. pedunculatus* are bactericidal to R7A and the genes that encode this flavolan binding polysaccharide are absent within 6B1.5.

Potential T1SS effectors were identified based on the presence of RTX Ca\textsuperscript{2+} domains (detailed in section 4.2.10.4). However, it should be noted that not all effectors of the T1SS contain RTX Ca\textsuperscript{2+} binding domains, with examples such as PlyA/B, and Egl from *A. caulinodans* (Finnie *et al.*, 1998). In addition, genes that encode effectors of the T1SS are not always genetically linked to their respective secretion systems, with examples such as the lipase LipA and the protease PrtA which are both not genetically linked to their T1SS, LipBCD (Kawai *et al.*, 1998). NodO from *R. leguminosarum* is another example of an effector that is not genetically linked to the T1SS that it is associated with (Finnie *et al.*, 1997). Thus, potential effectors that may be secreted by PrsDE within NZP2037 may not have been identified and future experiments may focus on the protein expression profiles of a prsDE mutant and that of the wildtype. It should also be noted that other predicted effector proteins secreted by T1SS in *R. leguminosarum* bv. *viciae*
include: nucleoside diphosphate kinases, glycosyl hydrolase, and parallel beta-helix repeat proteins. However, their roles in relation to symbiosis (if any) have yet to be determined (Krehenbrink & Downie, 2008).

A BLASTP search for homologues of PrsDE in both the R7A and NZP2037 genomes identified two PrsDE homologues unique to the broad host range *Mesorhizobium*. One of these T1SS was the PrsDE described in detail within this study. The other T1SS (*mln388/mln389*) was identified to not contain a nod box, thus, its expression is likely to be independent of flavonoid signalling. The conservation of *mln388/mln389* within the symbiosis islands of strains able to nodulate *L. pedunculatus* suggests that the T1SS that the genes encode may have an important role in symbiosis. It may be possible that Mln388/Mln398 has the ability to partially complement other T1SS, and examples of this have been observed in *R. leguminosarum* where mutants impaired in glycanase secretion via deletion of *prsE* retained some residual glycanase activity, suggesting the presence of another secretion system (Russo *et al.*, 2006). Although the hypothesis of another T1SS compensating for the loss of *prsDE* is attractive, it should be noted Mln388/Mln389 shares only 24% amino acid identity to PrsDE and the lack of similarity between the two systems could mean that they are not functionally related. Regardless, future experiments that explore the nodulation phenotypes of *mln388/mln399* mutants may provide interesting results into other potential T1SS and their roles in symbiosis. Unfortunately, the genes were not discovered in time to be included in the current study.

In conclusion, this chapter details the logical experimental approaches that were used to ascertain the genes responsible for a defective nodulation phenotype exhibited by the mutant Δ(*nodO-9.3kb*) on *L. pedunculatus*. The results culminated in the finding that the
original gene of interest *nodO* was not essential for the nodulation of *L. pedunculatus* and that it was the disruption of the T1SS, PrsDE, and another likely effector that it transports, Mln031, that were responsible for this defective nodulation phenotype. These results bring to light the T1SS as an important determinant of host-specificity for the host *L. pedunculatus* and Mln031 as a novel effector that is likely to be secreted by PrsDE. However, the mechanism of action for Mln031 is still up for speculation, though it is hypothesised here to function through the amplification of ionic flux in order to overcome a signalling threshold required for infection thread initiation. Mln031 is proposed to achieve this ionic flux amplification through either i) direct transport of cations into the root hair cell or ii) indirectly via the promotion of adhesion of the rhizobial cell to the root hair surface, the action of which generates an increase in the localised concentration of Nod factors, amplifying their signalling effects and by extension the influx of cations into the plant root hair. Future studies are required to determine the exact mechanism through which Mln031 functions, which in turn should provide insights into the strategies *Mesorhizobium* may utilise to establish a successful symbiosis with certain hosts.
5 Concluding remarks
The aim of this study was to identify the genetic determinants behind the broad host range of a subset *M. loti* strains that are represented in this study by the *M. loti* strain NZP2037. An additional aim was to ascertain if these gene(s) were also involved in the circumvention of defective EPS signalling on *L. japonicus* Gifu and to clarify the role of EPS on *L. pedunculatus*. The results showed that genes associated with incorporating an unsaturated fatty acid moiety on the Nod factor (*nodFEG, nodA2*) were not essential and had little effect on nodulation of *L. japonicus* Gifu in the presence or absence of defective EPS. Additional gene targets pertaining to the T4SS (*virA, virB1-11*) and *mln399* were also not essential for nodulation. Lastly, EPS signalling does not appear to have an important role in the nodulation of *L. pedunculatus*.

One of the more interesting findings was that *nodU* promoted organogenesis of *L. japonicus* Gifu. NodU is a carbamoyl transferase that modifies the non-reducing terminal GlcNAc of the Nod factor via transfer of a carbamoyl group to the C6 position (Jabbouri *et al.*, 1995). A double *nodU exoU* mutant exhibited an inability to induce effective nodules on *L. japonicus* Gifu, suggesting that *nodU* may be involved in circumvention of defective EPS signalling. However, complementation of this defective phenotype could not be achieved and this was attributed to the burden of *nodU* expression compounded with an already delayed nodulation phenotype of the *exoU* mutant. Although restoration of a wildtype nodulation phenotype could not be achieved with *nodU*, it was noted that *exoU* mutants that harboured a *nodU*-containing vector generated more nodule primordia. This suggested that NodU promoted nodulation via the organogenesis pathway, but by itself was unable to overcome the negative signalling of defective EPS. The results support a two separate signalling pathway model that was proposed by Guinzel & Geil (2002), with one signalling pathway involved in development of primordia that is
dependent on Nod factor, while the other pathway is involved in infection at the epidermis and is dependent on the presence of the bacterial cell (in this case EPS produced by the rhizobial cell). These two pathways eventually converge with the infection thread infiltrating the nodule primordia, resulting in effective nodulation and a disruption in either pathway leads to the inability to form an effective symbiosis.

When the nodU mutation was transferred into the R7A\textsuperscript{C}:NZP2037\textsuperscript{I} transconjugant background, the defective nodulation phenotype of the nodU exoU double mutant was not replicated. Additionally, the delayed nodulation phenotype that was observed in exoU mutants of NZP2037 was no longer present within the transconjugant. Potential hypotheses for this observation were that chromosomal genes of NZP2037 were detrimental to the nodulation of \textit{L. japonicus} Gifu. Alternatively, genes with a negative effect on nodulation may be present within the R7A symbiosis island. The successful mobilisation of the NZP2037 island into the R7A genomic background resulted in a transfer of the NZP2037 host-range, suggesting that all genes essential for the nodulation of \textit{L. pedunculatus} are encoded within the symbiosis island of NZP2037.

Ultimately, the gene(s) encoded within the NZP2037 island involved in the circumvention of defective EPS signalling could not be identified. The genes nodU, nodFEG, nodA2, mln399, virB1-11, virA were all expendable for nodulation of \textit{L. japonicus} Gifu and \textit{L. pedunculatus}. The gene nodU was found to be involved in organogenesis but the effect was only observed in exoU mutants on \textit{L. japonicus} Gifu suggesting that negative feedback mechanisms from effective nodulation likely prevents the excess organogenesis effect.
Future work could focus on the identification and/or characterisation of genes that have no known function unique to the islands of the broad host-range *Mesorhizobium*. It is possible that the genes involved in host-specificity are not regulated by NodD, and/or are constitutively expressed, and may have escaped the scope of this study. One such gene of interest that has yet to be identified is the gene responsible for protection against the toxic flavolans of *L. pedunculatus* which is associated with the peptidoglycan layer of the membrane (Pankhurst *et al.*, 1982) and likely to be constitutively expressed. The symbiosis island within R7A may also be of interest, and disruption of genes within this island may produce interesting phenotypes on *L. pedunculatus*. Lastly, the Nod factor produced by a *nodU, nodFEG, nodA2, mln399* mutant is predicted to mirror the Nod factor that is produced by wildtype R7A. However, the mutant remains compatible with *L. pedunculatus*, suggesting that R7A Nod factor is potentially compatible with *L. pedunculatus* and it is some other factor that is required for nodulation of this host.

Investigations into the Nod factor structure produced by NZP2037 Nod factor mutants presented in this study using the more advanced techniques available today may identify modifications that have yet to be identified. The investigation of a Nod factor structure lacking in common modifications and their effects for both *L. pedunculatus* and *L. japonicus* Gifu may also be explored.

The most interesting finding was the characterisation of the *nodO*-9.3kb mutant and its defective nodulation phenotype on *L. pedunculatus*. Initially it was speculated that the defective nodulation phenotype of this mutant was due to the disruption of *nodO*. However, it was concluded that a *nodO* mutant retained an effective nodulation phenotype and therefore was not the gene interest. Instead, the defective nodulation phenotype was attributed to the disruption of *prsDE* and *mln031* which encode a T1SS and its transported
The effector Mln031 represents the second effector secreted by a T1SS to have a functional role in nodulation. The first identified and most well understood rhizobial T1SS with a role in symbiosis is NodO (de Maagd et al., 1989, Walker & Downie, 2000). Mutants of prsDE exhibits a defective nodulation phenotype without the need of a secondary mutation. This is in contrast to NodO of *R. leguminosarum* bv. *viciae* which only exhibits a defective phenotype in a *nodE nodO* double mutant background (Walker & Downie, 2000). The only conserved protein domain identified within Mln031 was the RTX domain, which is common amongst effectors of the T1SS (Welch, 2001, Thomas et al., 2014). No other conserved domains were predicted within the primary sequence. The closest homologues of Mln031 have not been studied in detail. Based on the similarities between NodO and Mln031, it was hypothesised that Mln031 may function through similar mechanisms to that of NodO.

Mln031 was hypothesised to be involved in the amplification of Nod factor signalling via the amplification of the Ca$^{2+}$ influx into the cell required for formation of the infection thread. Alternatively, Mln031 may function through promotion of inherent properties such as adhesion to the host cell, increasing the localised concentration of Nod factor, given that Mln031 shares some similarities to adhesins. Regardless of the exact function of Mln031, this amplification hypothesis is attractive as it provides explanations for the defective and delayed nodulation phenotype of prsDE mutants. Interestingly, this hypothesis may suggest that Nod factor produced by NZP2037 are suboptimal for *L. pedunculatus*, as Mln031 is required for amplification. Therefore, Nod factors that lack modifications to the basic structure may be beneficial and disruption of genes responsible for these modifications would have no observable effects as seen in the *nodA2, nodFEG, nodU, mln399* mutants for both *L. japonicus Gifu* and *L. pedunculatus*. 

195
A search for T1SS present within NZP2037 found 2 systems unique to the broad host-range *M. loti*. The first system was PrsDE (detailed in section 4.2.10). The other T1SS remains an enigma and its role in symbiosis has yet to be ascertained, partly due to the absence of a *nod* box and so it is not under NodD regulation. The T1SS are implicated to have an important role in the establishment of an effective nodulation phenotype on *L. pedunculatus*. Therefore, this other T1SS provides an attractive target for future studies.

In conclusion, complex interactions occur during the establishment of symbiosis that determine the host-specificity. The results within this study provide insights into the broad host range of *M. loti* strain NZP2037 and the ways through which it augments these interactions to favour a successful symbiotic relationship. The major findings of this study were that the majority of unique *nod* genes of NZP2037 were expendable for nodulation of *L. pedunculatus*, and the importance of the T1SS PrsDE and its effector Mln031 on this host. The T1SS presents an interesting frontier for future studies relating to determinants of host-range, and in particular, the importance of Ca\(^{2+}\) influx and its relation to infection thread formation. However, the presence of *prsDE* and *mln031* in R7A were unable to extend the host host-range to include *L. pedunculatus*, suggesting that additional genes within the NZP2037 symbiosis island are required. Taken together, Mln031 represents the second Type I mediated effector identified to have a role in symbiosis and *prsDE* mutants exhibit a defective nodulation phenotype without the requirement of a secondary mutation. Future studies revolving around Mln031 and the T1SS may prove to be a fruitful endeavour in expanding our understanding of the signalling pathways utilised by rhizobia in the establishment of symbiosis.
6 References


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


Economou A., Hamilton W., Johnston A. & Downie J. (1990). The \textit{Rhizobium} nodulation gene \textit{nodO} encodes a \(\text{Ca}^{2+}\)-binding protein that is exported without N-terminal cleavage and is homologous to haemolysin and related proteins. The EMBO journal \textbf{9}: 349-354.


Finn R. D., Attwood T. K., Babbitt P. C., Bateman A., Bork P., Bridge A. J., Chang


bound polysaccharide associated with resistance of *Rhizobium loti* strain NZP2037 to *Lotus pedunculatus* root flavolan. Microbiology **133**: 2617-2629.


Laus M., Van Brussel A. & Kijne J. (2005). 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.


Priefer U. B. (1989). Genes involved in lipopolysaccharide production and symbiosis are


**Roche P., Debellé F., Maillet F., Lerouge P., Faucher C., Truchet G., Dénarié J. &**

212


specificity-determining NodE protein. The EMBO journal 8: 2811-2818.


215


7 Appendix
### Recipe for LB

<table>
<thead>
<tr>
<th>Compound</th>
<th>g/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacto-tryptone</td>
<td>10</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>5</td>
</tr>
<tr>
<td>NaCl</td>
<td>5</td>
</tr>
</tbody>
</table>

### Recipe for SOB

<table>
<thead>
<tr>
<th>Compound</th>
<th>g/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacto-tryptone</td>
<td>20</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>5</td>
</tr>
<tr>
<td>NaCl</td>
<td>0.5</td>
</tr>
<tr>
<td>0.25 M KCl</td>
<td>10 mL</td>
</tr>
</tbody>
</table>

pH adjusted to 7.0 with 10 M NaOH

### Recipe for TY

<table>
<thead>
<tr>
<th>Compound</th>
<th>g/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacto-tryptone</td>
<td>10</td>
</tr>
<tr>
<td>Yeast extract</td>
<td>5</td>
</tr>
<tr>
<td>CaCl$_2$·6 H$_2$O</td>
<td>1.3</td>
</tr>
</tbody>
</table>
Recipe for RDM (G/RDM and S/RDM)

<table>
<thead>
<tr>
<th>Stock solutions</th>
<th>mL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salts solution</td>
<td>10</td>
</tr>
<tr>
<td>BTB</td>
<td>10</td>
</tr>
<tr>
<td>NH₄Cl</td>
<td>6</td>
</tr>
<tr>
<td>Trace elements</td>
<td>1</td>
</tr>
<tr>
<td>L-histindine</td>
<td>100 mg</td>
</tr>
<tr>
<td>After autoclave</td>
<td></td>
</tr>
<tr>
<td>Phosphate solution</td>
<td>10</td>
</tr>
<tr>
<td>20% Glucose/20% Sucrose</td>
<td>5/250</td>
</tr>
</tbody>
</table>

Stock solution recipes

<table>
<thead>
<tr>
<th>Compound</th>
<th>g/L</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phosphate solution</strong></td>
<td></td>
</tr>
<tr>
<td>K₂HPO₄</td>
<td>100</td>
</tr>
<tr>
<td>KH₂PO₄</td>
<td>100</td>
</tr>
<tr>
<td><strong>Salts solution</strong></td>
<td></td>
</tr>
<tr>
<td>MgSO₄·7 H₂O</td>
<td>25</td>
</tr>
<tr>
<td>CaCl₂·2 H₂O</td>
<td>2</td>
</tr>
<tr>
<td>Fe·EDTA</td>
<td>1.5</td>
</tr>
<tr>
<td>NaCl</td>
<td>20</td>
</tr>
<tr>
<td><strong>BTB (bromothymol blue)</strong></td>
<td></td>
</tr>
<tr>
<td>Bromothymol blue</td>
<td>2</td>
</tr>
</tbody>
</table>
### Trace elements (mg per 200 mL)

<table>
<thead>
<tr>
<th>Compound</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZnSO$_4$$\cdot$7 H$_2$O</td>
<td>3</td>
</tr>
<tr>
<td>Na$_2$MoO$_4$$\cdot$2 H$_2$O</td>
<td>40</td>
</tr>
<tr>
<td>H$_3$BO$_3$</td>
<td>50</td>
</tr>
<tr>
<td>MnSO$_4$$\cdot$H$_2$O</td>
<td>40</td>
</tr>
<tr>
<td>CuSO$_4$$\cdot$5 H$_2$O</td>
<td>4</td>
</tr>
<tr>
<td>CoCl$_2$$\cdot$6 H$_2$O (0.2 g/L)</td>
<td>1 mL</td>
</tr>
</tbody>
</table>

### Hoagland’s growth solution

<table>
<thead>
<tr>
<th>Compound</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>K$_2$HPO$_4$</td>
<td>0.19</td>
</tr>
<tr>
<td>KCl</td>
<td>0.29</td>
</tr>
<tr>
<td>CaCl$_2$$\cdot$2H$_2$O</td>
<td>0.73</td>
</tr>
<tr>
<td>MgSO$_4$$\cdot$7H$_2$O</td>
<td>0.37</td>
</tr>
<tr>
<td>Micronutrients (see below)</td>
<td>1 mL</td>
</tr>
<tr>
<td>Fe EDTA (see below)</td>
<td>1 mL</td>
</tr>
<tr>
<td>K$_2$HPO$_4$</td>
<td>0.19</td>
</tr>
</tbody>
</table>
### Hoagland’s micronutrients and Fe·EDTA

<table>
<thead>
<tr>
<th>Micronutrients</th>
<th>g/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZnSO$_4$·7H$_2$O</td>
<td>0.22</td>
</tr>
<tr>
<td>CuSO$_4$·5H$_2$O</td>
<td>0.08</td>
</tr>
<tr>
<td>MnCl$_2$·4H$_2$O</td>
<td>1.81</td>
</tr>
<tr>
<td>H$_3$BO$_3$</td>
<td>2.86</td>
</tr>
<tr>
<td>Na$_2$MoO$_4$·4H$_2$O</td>
<td>0.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fe EDTA</th>
<th>g/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>FeSO$_4$·7H$_2$O</td>
<td>17.2</td>
</tr>
<tr>
<td>1 M KOH</td>
<td>250 mL</td>
</tr>
</tbody>
</table>

Mix thoroughly for 5 minutes before adding EDTA

### Hornum’s growth solution

<table>
<thead>
<tr>
<th>Compound</th>
<th>g/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH$_4$NO$_3$</td>
<td>40 g</td>
</tr>
<tr>
<td>KNO$_3$</td>
<td>30 g</td>
</tr>
<tr>
<td>MgSO$_4$·7H$_2$O</td>
<td>30 g</td>
</tr>
<tr>
<td>NaH$_2$PO$_4$·2H$_2$O</td>
<td>10 g</td>
</tr>
<tr>
<td>Fe EDTA (9%)</td>
<td>2 g</td>
</tr>
<tr>
<td>MnSO$_4$·H$_2$O</td>
<td>120 mg</td>
</tr>
<tr>
<td>H$_3$BO$_3$</td>
<td>120 mg</td>
</tr>
<tr>
<td>CuSO$_4$·5H$_2$O</td>
<td>40 mg</td>
</tr>
<tr>
<td>ZnSO$_4$·7H$_2$O</td>
<td>40 mg</td>
</tr>
<tr>
<td>Na$_2$MoO$_4$·2H$_2$O</td>
<td>8 mg</td>
</tr>
</tbody>
</table>
List of genes identified to be unique to the NZP2037 symbiosis island. Genes that were targeted for deletion in this study are indicated by an * (except, mln452 and mln454 which are predicted effectors of the Type IV secretion system).

<table>
<thead>
<tr>
<th>Gene/Locus tag</th>
<th>p-value</th>
<th>Predicted function</th>
</tr>
</thead>
<tbody>
<tr>
<td>virB5*</td>
<td>0.03571429</td>
<td>Type IV secretion system</td>
</tr>
<tr>
<td>mln452*</td>
<td>0.03571429</td>
<td>Hypothetical Type IV effector</td>
</tr>
<tr>
<td>group_5687</td>
<td>0.03571429</td>
<td>RNA polymerase sigma-54 factor</td>
</tr>
<tr>
<td>group_5682</td>
<td>0.03571429</td>
<td>glycosyl transferase</td>
</tr>
<tr>
<td>group_6998</td>
<td>0.03571429</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_729</td>
<td>0.03571429</td>
<td>peptidase C39</td>
</tr>
<tr>
<td>group_5676</td>
<td>0.03571429</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_6994</td>
<td>0.03571429</td>
<td>phosphatase</td>
</tr>
<tr>
<td>group_5675</td>
<td>0.03571429</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_5674</td>
<td>0.03571429</td>
<td>IS110 family transposase</td>
</tr>
<tr>
<td>group_4246</td>
<td>0.03571429</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_5673</td>
<td>0.03571429</td>
<td>transposase</td>
</tr>
<tr>
<td>gabD_3</td>
<td>0.03571429</td>
<td>NAD-dependent succinate-semialdehyde dehydrogenase</td>
</tr>
<tr>
<td>group_4309</td>
<td>0.03571429</td>
<td>2OG-Fe(II) oxygenase</td>
</tr>
<tr>
<td>group_7042</td>
<td>0.03571429</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_212</td>
<td>0.03571429</td>
<td>membrane protein</td>
</tr>
<tr>
<td>group_7040</td>
<td>0.03571429</td>
<td>integrase</td>
</tr>
<tr>
<td>group_7039</td>
<td>0.03571429</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_7038</td>
<td>0.03571429</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_7036</td>
<td>0.03571429</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_5747</td>
<td>0.03571429</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_4295</td>
<td>0.03571429</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_7035</td>
<td>0.03571429</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_7034</td>
<td>0.03571429</td>
<td>ABC transporter ATP-binding protein</td>
</tr>
<tr>
<td>group_2683</td>
<td>0.03571429</td>
<td>ABC transporter ATP-binding protein</td>
</tr>
<tr>
<td>group_7031</td>
<td>0.03571429</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_5785</td>
<td>0.03571429</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_7175</td>
<td>0.03571429</td>
<td>3-methyl-2-oxobutanoate hydohxymethyltransferase</td>
</tr>
<tr>
<td>nodG*</td>
<td>0.03571429</td>
<td>beta-ketoacyl-ACP reductase</td>
</tr>
<tr>
<td>group_4319</td>
<td>0.03571429</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_5761</td>
<td>0.03571429</td>
<td>IS110 family transposase</td>
</tr>
<tr>
<td>aprA</td>
<td>0.03571429</td>
<td>Metallopeptidase AprA</td>
</tr>
<tr>
<td>group_5762</td>
<td>0.03571429</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_7096</td>
<td>0.03571429</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>mln031*</td>
<td>0.03571429</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>prsE*</td>
<td>0.03571429</td>
<td>HlyD family type I secretion periplasmic adaptor subunit</td>
</tr>
<tr>
<td>prsD*</td>
<td>0.03571429</td>
<td>type I secretion system permease/ATPase</td>
</tr>
<tr>
<td>group_5665</td>
<td>0.03571429</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_2645</td>
<td>0.03571429</td>
<td>integrase</td>
</tr>
<tr>
<td>group_5664</td>
<td>0.03571429</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_5663</td>
<td>0.03571429</td>
<td>IS110 family transposase</td>
</tr>
<tr>
<td>group_711</td>
<td>0.03571429</td>
<td>cold-shock protein</td>
</tr>
<tr>
<td>Gene/Locus tag</td>
<td>p-value</td>
<td>Predicted function</td>
</tr>
<tr>
<td>---------------</td>
<td>---------</td>
<td>--------------------</td>
</tr>
<tr>
<td>group_5660</td>
<td>0.03571429</td>
<td>helix-turn-helix domain-containing protein</td>
</tr>
<tr>
<td>group_5659</td>
<td>0.03571429</td>
<td>IS110 family transposase</td>
</tr>
<tr>
<td>group_6982</td>
<td>0.03571429</td>
<td>outer membrane protease</td>
</tr>
<tr>
<td>nodU*</td>
<td>0.03571429</td>
<td>nodulation protein NodU</td>
</tr>
<tr>
<td>group_7001</td>
<td>0.10714286</td>
<td>chromosome partitioning protein ParB</td>
</tr>
<tr>
<td>nodA*</td>
<td>0.10714286</td>
<td>NodA family N-acyltransferase</td>
</tr>
<tr>
<td>hlyA</td>
<td>0.10714286</td>
<td>Hemolysin, chromosomal</td>
</tr>
<tr>
<td>group_7092</td>
<td>0.10714286</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>nodO*</td>
<td>0.10714286</td>
<td>Nodulation protein O</td>
</tr>
<tr>
<td>group_6983</td>
<td>0.10714286</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_5779</td>
<td>0.42857143</td>
<td>integrase</td>
</tr>
<tr>
<td>group_5778</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_7127</td>
<td>0.42857143</td>
<td>integrase</td>
</tr>
<tr>
<td>group_277</td>
<td>0.42857143</td>
<td>ISNCY family transposase</td>
</tr>
<tr>
<td>group_7128</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_7129</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_5770</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_2701</td>
<td>0.42857143</td>
<td>DNA helicase</td>
</tr>
<tr>
<td>group_5814</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_7272</td>
<td>0.42857143</td>
<td>IS110 family transposase</td>
</tr>
<tr>
<td>group_7271</td>
<td>0.42857143</td>
<td>NAD(P)/FAD-dependent oxidoreductase</td>
</tr>
<tr>
<td>cbb3-type cytochrome oxidase assembly protein CcoS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>group_302</td>
<td>0.42857143</td>
<td>cytochrome oxidase</td>
</tr>
<tr>
<td>group_130</td>
<td>0.42857143</td>
<td>oxygen-independent coproporphyrinogen III oxidase</td>
</tr>
<tr>
<td>group_7147</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>mln452*</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_5705</td>
<td>0.42857143</td>
<td>IS1380 family transposase</td>
</tr>
<tr>
<td>group_5775</td>
<td>0.42857143</td>
<td>His-Xaa-Ser repeat protein HxsA</td>
</tr>
<tr>
<td>group_5773</td>
<td>0.42857143</td>
<td>His-Xaa-Ser system radical SAM maturase HxsC</td>
</tr>
<tr>
<td>group_4345</td>
<td>0.42857143</td>
<td>His-Xaa-Ser system radical SAM maturase HxsB</td>
</tr>
<tr>
<td>group_7243</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_5802</td>
<td>0.42857143</td>
<td>IS1380 family transposase</td>
</tr>
<tr>
<td>group_7244</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_7245</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_7246</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_741</td>
<td>0.42857143</td>
<td>exopolysaccharide repressor protein</td>
</tr>
<tr>
<td>group_110</td>
<td>0.42857143</td>
<td>ferredoxin III, nif-specific</td>
</tr>
<tr>
<td>group_6999</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_7247</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_4248</td>
<td>0.42857143</td>
<td>hemolysin secretion protein D</td>
</tr>
<tr>
<td>group_7193</td>
<td>0.42857143</td>
<td>tyrosine recombinase XerD</td>
</tr>
<tr>
<td>group_7195</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_7197</td>
<td>0.42857143</td>
<td>replication primase</td>
</tr>
<tr>
<td>group_166</td>
<td>0.42857143</td>
<td>antirestriction protein ArdC</td>
</tr>
<tr>
<td>Gene/Locus tag</td>
<td>$p$-value</td>
<td>Predicted function</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>group_5679</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_5794</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_5678</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_7123</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_5769</td>
<td>0.42857143</td>
<td>IS5 family transposase</td>
</tr>
<tr>
<td>group_5809</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_759</td>
<td>0.42857143</td>
<td>transposase</td>
</tr>
<tr>
<td>group_2710</td>
<td>0.42857143</td>
<td>IS66 family transposase</td>
</tr>
<tr>
<td>group_5808</td>
<td>0.42857143</td>
<td>YecA family protein</td>
</tr>
<tr>
<td>group_5767</td>
<td>0.42857143</td>
<td>aspartate aminotransferase family protein</td>
</tr>
<tr>
<td>group_5766</td>
<td>0.42857143</td>
<td>NAD-dependent succinate-semialdehyde dehydrogenase</td>
</tr>
<tr>
<td>group_5753</td>
<td>0.42857143</td>
<td>IS110 family transposase</td>
</tr>
<tr>
<td>group_5805</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_5804</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_7174</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
<tr>
<td>group_2661</td>
<td>0.42857143</td>
<td>conjugal transfer protein TraI</td>
</tr>
<tr>
<td>group_5759</td>
<td>0.42857143</td>
<td>HlyD family secretion protein</td>
</tr>
<tr>
<td>group_709</td>
<td>0.42857143</td>
<td>type I secretion system permease/ATPase</td>
</tr>
<tr>
<td>group_7260</td>
<td>0.42857143</td>
<td>transposase</td>
</tr>
<tr>
<td>group_7171</td>
<td>0.42857143</td>
<td>hypothetical protein</td>
</tr>
</tbody>
</table>