Regulation of Excision and Transfer of the Mesorhizobium loti R7A Symbiosis Island

Joshua P. Ramsay

BSc (Hons)

A thesis submitted for the degree of

Doctor of Philosophy

University of Otago, Dunedin

New Zealand

February 2008
This thesis is dedicated to Uncle Neil

Allan Neil McDonald

1946-2006
Abstract

Integrative and conjugative elements (ICEs) are a newly appreciated group of mobile elements that contribute immensely to bacterial evolution. The symbiosis island ICEmISym$^{R7A}$ of *Mesorhizobium loti* strain R7A is the largest known mobile ICE (502 kb) and confers the ability to form a nitrogen-fixing symbiosis with plants of the genus *Lotus*.

Transfer of ICEmISym$^{R7A}$ to non-symbiotic bacteria has been demonstrated both in the laboratory and the environment and involves chromosomal integration of ICEmISym$^{R7A}$ adjacent to the phe-tRNA gene in the recipient bacterium. Integration of ICEmISym$^{R7A}$ requires the ICEmISym$^{R7A}$-encoded protein IntS which likely catalyses recombination between the attachment sites attP and attB, present on excised ICEmISym$^{R7A}$ and phe-tRNA gene respectively. In this study it was shown that intS was required for both excision and integration and was required in both the donor and recipient bacterium for efficient transfer. A minimal attP region required for integration was defined and a repetitive sequence motif that likely represents the IntS-binding sequence identified. intS is expressed from two alternate promoters, a weak promoter on attL (left ICEmISym$^{R7A}$-chromosome junction) and a strong promoter formed on attP. Analysis of nested deletions of the attP region demonstrated that an unusually large region of attP, corresponding to that required for integration, was required for full gene expression.

A novel recombination directionality factor RdfS was identified using bioinformatics. A quantitative PCR assay developed to detect the presence of excised ICEmISym$^{R7A}$ in bacterial populations showed that excision of ICEmISym$^{R7A}$ was almost undetectable in an rdfS mutant. Constitutive expression of rdfS resulted in curing of ICEmISym$^{R7A}$ from R7A, creating a non-symbiotic derivative R7ANS.

Bioinformatic analyses of rdfS and associated genes on ICEmISym$^{R7A}$ led to discovery of 26 putative ICEs (ICESym family) in 12 α-proteobacterial species. The ICESym clusters contained homologues of up to 25 ICEmISym$^{R7A}$ proteins, encoding functions involved in mating pore formation, DNA processing and regulation, of which 17 proteins were universally conserved.

Introduction of pJR174 containing a copy of the ICEmISym$^{R7A}$-encoded quorum sensing (QS) gene traR induced excision in 100% of cells and stimulated a 100-fold increase in ICEmISym$^{R7A}$ transfer. Stable maintenance of ICEmISym$^{R7A}$ in these cells required the DNA relaxase-encoding
gene rlxS. Introduction of pJR174 also induced a 1000-fold increase in the production of 3-oxo-C6-homoserine lactone and several other acyl homoserine lactones; this induction required traI1. The induction of QS was accompanied by various growth-inhibitory effects, similar to those observed in other rhizobial QS systems.

The QS system activated ICEMlSymR7A excision through expression of msi172 and msi171, genes that are downstream of QS gene traI2. Expression of the adjacent gene msi170 separately repressed both QS and excision. The expression of msi170 was negatively regulated by Msi169, while msi169 was positively autoregulated. msi169 encodes an Xre family DNA-binding protein and homologues of it were identified on the ICESym clusters and on QS-regulated plasmids, suggesting that they have a conserved role as regulators of transfer- and QS-related genes. Overall, the results suggest that ICEMlSymR7A excision and transfer are subject to both population-density and cell-cycle-dependent regulation.
Acknowledgments

Firstly I would like to thank my supervisors Clive, John and Iain for all your advice and support throughout my years at Otago. I would especially like to thank Clive for always making time for my questions and debate and restraining himself in the face of my admittedly argumentative and sometimes stubborn nature. I thank John also for helpful discussions, fishing trips and the occasional tongue-in-cheek pep talk. All three of you have given me a huge leg-up towards my career for which I will always be in debt. I would also like to thank Helen, Tim, Gabs, Simon, Andree, Pinky, Sieu, Scott, Duncan, everyone else on the 6th floor and the other members of the Microbiology Dept. for friendship, assistance and the legendary rooftop drinking sessions of times gone by. We will no doubt meet again throughout our careers. I would also like to thank my surrogate supervisors in England, Paul, Stephan, Miguel, Jean, Steve and Cathy, as well as fellow students Ras, James, Jim, Marco Christian, Sarah Shelly, Owen, Gen and others for helping me have a productive, yet enjoyable stay in the UK. And thanks especially to Elman for providing the means for my travels. To my family, Margaret, Peter and Luke, I thank you for all your support through my transition from a mischievous teenager to a semi-respectable adult. You have definitely taken the good with the bad and put up with me throughout. Finally I would like to thank my partner Nicole for her support and encouragement and companionship throughout my studies.
# Table of Contents

Abstract ............................................................................................................................................................................. v

Acknowledgments .................................................................................................................................................................... vii

Table of Contents .................................................................................................................................................................. viii

List of Figures ..................................................................................................................................................................... xiii

List of Tables ....................................................................................................................................................................... xv

List of abbreviations ................................................................................................................................................................. xvii

Chapter 1 General Introduction.................................................................................................................................................. 1

1.1 The rhizobium-legume symbiosis ................................................................................................................................. 4

1.2 *Mesorhizobium loti* R7A..................................................................................................................................................... 6

1.2.1 Phylogeny and description .............................................................................................................................................. 6

1.2.2 Discovery and characterisation of ICEMlSym\textsuperscript{R7A} .......................................................................................... 9

1.3 Evolution and genetics of horizontal gene transfer ........................................................................................................ 10

1.3.1 Sex and recombination ................................................................................................................................................... 10

1.3.2 Population-level persistence of MGEs ........................................................................................................................... 11

1.4 Mobile genetic elements...................................................................................................................................................... 12

1.4.1 Regulation of the bacteriophage lifecycle ..................................................................................................................... 13

1.4.1.1 Chromosomal integration ............................................................................................................................................ 14

1.4.1.2 Regulation of lysogeny ................................................................................................................................................. 15

1.4.2 Plasmid maintenance ...................................................................................................................................................... 17

1.4.2.1 Replication and incompatibility .................................................................................................................................. 18

1.4.2.2 Plasmid partition ........................................................................................................................................................... 19

1.4.2.3 Post-segregational killing systems ............................................................................................................................ 19

1.4.3 Conjugative transfer ...................................................................................................................................................... 20

1.4.3.1 Type IV secretion systems ........................................................................................................................................... 20

1.4.3.2 The relaxosome .......................................................................................................................................................... 21

1.4.3.3 The coupling protein .................................................................................................................................................... 21

1.5 Quorum sensing ................................................................................................................................................................. 22

1.5.1 N-acyl homoserine lactone-dependent quorum sensing ............................................................................................ 23

1.5.2 Quorum sensing in rhizobia ........................................................................................................................................ 25

1.5.2.1 Regulation of the *A. tumefaciens* Ti plasmids ............................................................................................................. 25
1.5.2.2 Regulation of transfer of the *R. leguminosarum* plasmid pRL1JI ..............................................................26

1.6 Integrative and conjugative elements (ICEs).................................................................................................................................27
1.6.1 SXT and R391 ................................................................................................................................................................................27
1.6.2 clc and ICE*Hin*1056 .................................................................................................................................................28
1.6.3 Tn4371 .....................................................................................................................................................................................29

1.7 Aims of this study...........................................................................................................................................................................29

Chapter 2 Materials and Methods .................................................................................................................................................31

2.1 Media.........................................................................................................................................................................................33

2.2 Bacterial strains and Plasmids ........................................................................................................................................33

2.3 Storage of Strains ........................................................................................................................................................................36

2.4 Enzymes and Chemicals ............................................................................................................................................................36

2.5 Spectroscopy ..................................................................................................................................................................................36

2.6 DNA Manipulations......................................................................................................................................................................36
2.6.1 Electrophoresis........................................................................................................................................................................36
2.6.2 DNA Isolation............................................................................................................................................................................37
2.6.2.1 Alkaline lysis plasmid extractions..........................................................................................................................................37
2.6.2.2 Prepman Ultra PCR-template Preparations ........................................................................................................................37
2.6.2.3 Commercial plasmid purification kits ................................................................................................................................37
2.6.3 DNA fragment and PCR product purification......................................................................................................................38
2.6.3.1 Commercial DNA fragment purification kits ......................................................................................................................38
2.6.3.2 Ethanol precipitation ...........................................................................................................................................................38
2.6.3.3 Restriction endonuclease digestion ......................................................................................................................................38

2.7 Polymerase Chain Reaction (PCR) .........................................................................................................................................38
2.7.1 Standard PCR protocol ........................................................................................................................................................38
2.7.2 Overlap-extension PCR .......................................................................................................................................................41
2.7.3 Quantitative PCR (QPCR) ....................................................................................................................................................42

2.8 AHL analysis ..................................................................................................................................................................................44
2.8.1 Extraction ...............................................................................................................................................................................44
2.8.2 AHL Detection by *C. violaceum* CV026 ............................................................................................................................44
2.8.2.1 CV026 overlays .................................................................................................................................................................44
2.8.2.2 Plate assays .......................................................................................................................................................................44
2.8.2.3 Thin-layer chromatography ........................................................................................................................................44
2.8.3 LC-MS/MS method ..............................................................................................................................................................45

2.9 Protein induction, purification and analysis ..............................................................................................................................45
2.9.1 SDS-PAGE .............................................................................................................................................................................45
2.10 DNA sequencing.................................................................................................................. 46
2.11 Bacterial transformation........................................................................................................ 46
  2.11.1 Electroporation.................................................................................................................. 46
  2.11.2 Biparental matings............................................................................................................ 47
2.12 In-frame markerless deletion mutant construction ............................................................... 47
2.13 Frozen inocula ....................................................................................................................... 48
2.14 β-galactosidase reporter assays ............................................................................................ 48
2.15 Growth curves and QPCR assays ......................................................................................... 49
2.16 Bioinformatic analysis ......................................................................................................... 49

Chapter 3 The ICEMISymR7A attachment site attP ..................................................................... 50
3.1 The integrase IntS is required for excision, integration and efficient transfer of ICEMISymR7A ........................................................................................................................................ 53
  3.1.1 Complementation of excision in R7AΔintS ..................................................................... 53
  3.1.2 Complementation of transfer in R7AΔintS ..................................................................... 53
  3.1.3 Integration of attP-containing plasmids into R7ANS ....................................................... 54
3.2 Expression of intS from attP .................................................................................................. 56
  3.2.1 Expression of intS is from a strong promoter on attP ..................................................... 56
  3.2.2 Deletion mapping of the attP-intS promoter ................................................................... 57
3.3 Mapping of the minimal attP required for recombination ..................................................... 59
  3.3.1 Identification of a conserved motif on attP ..................................................................... 59
  3.3.2 The repeat motifs are required for efficient integration of attP-containing plasmids ....... 62
3.4 Discussion ............................................................................................................................. 65

Chapter 4 ICEMISymR7A Excision ...................................................................................... 69
4.1 Measurement of excision by QPCR....................................................................................... 73
  4.1.1 Development of a quantitative PCR assay of ICEMISymR7A excision............................ 73
  4.1.2 Excision of ICEMISymR7A increases in stationary-phase cultures ............................. 77
  4.1.3 QPCR assay of complementation of excision in R7AΔintS ............................................. 78
4.2 Identification of RdfS .......................................................................................................... 80
  4.2.1 Bioinformatic identification of msi109 as a possible recombination directionality factor .... 80
  4.2.2 rdfS (msi109) is required for efficient excision and transfer of ICEMISymR7A .......... 81
  4.2.3 rdfS and intS are sufficient to mediate excision of an attP-containing plasmid ................ 82
4.3 Isolation of R7ANS .............................................................................................................. 83
  4.3.1 Constitutive expression of rdfS results in curing of ICEMISymR7A from R7A ............. 83
  4.3.2 Characterisation of R7ANS ............................................................................................ 84
Chapter 7 Concluding discussion ........................................................................................................ 167

7.1 Integration and excision ........................................................................................................... 169

7.2 Regulation of ICEMISym$^{R7A}$ excision and transfer ......................................................... 170

7.3 Evolution of ICEMISym$^{R7A}$ regulation ............................................................................. 173

7.4 Final remarks .......................................................................................................................... 176

Chapter 8 References .................................................................................................................... 177

Chapter 9 Appendices .................................................................................................................... 191
List of Figures

Figure 1-1 Nodule infection by *M. loti* R7A 5
Figure 1-2 Phylogeny of the α-proteobacteria 8
Figure 1-3 Structure of the λ* attP, attB, attL* and *attR* sites. 15
Figure 1-4 Structure of AHL molecules 24
Figure 3-1 Location of primers used to detect ICEMI/Sym<sup>R7A</sup> excision and integration. 52
Figure 3-2 Detection of ICEMI/Sym<sup>R7A</sup> by PCR. 54
Figure 3-3 *attP* regions cloned into pFJX/pFUS2. 56
Figure 3-4 Organisation of arm-type integrase binding sites on various *attP* regions. 60
Figure 3-5 MEME output from a motif search of *attP*-related sequences. 61
Figure 3-6 The nucleotide sequence of the ICEMI/Sym<sup>R7A</sup> attachment site *attP*. 62
Figure 3-7 *attP* fragments cloned into pFUS2 for analysis of integration. 63
Figure 4-1 Location of QPCR primer-probe combination used to detect ICEMI/Sym<sup>R7A</sup> excision. 74
Figure 4-2 QPCR standard curves. 76
Figure 4-3 Excision of ICEMI/Sym<sup>R7A</sup> during growth in broth culture. 77
Figure 4-4 Increased presence of *attPlattB* in stationary-phase cultures. 78
Figure 4-5 The effect of *intS* and *rdfS* mutations on ICEMI/Sym<sup>R7A</sup>-excision. 79
Figure 4-6 Similarity of RdfS to known RDFs. 80
Figure 4-7 Organization of the *msi110–msi106* gene cluster on ICEMI/Sym<sup>R7A</sup>. 81
Figure 4-8 Construction of R7AΔrdfs. 82
Figure 4-9 QPCR and PCR assays used to confirm the isolation of the ICEMI/Sym<sup>R7A</sup>-cured strain R7ANS. 85
Figure 4-10 Isolation of R7ANS cells cured of pJR204. 85
Figure 4-11 Vitamin auxotrophy of R7ANS. 86
Figure 4-12 Model of ICEMI/Sym<sup>R7A</sup> excision, replication and transfer. 90
Figure 5-1 RdfS homologues. 96
Figure 5-2 A comparison of the genetic organisation of ICEMI/Sym<sup>R7A</sup> homologues found on the *Ralstonia oxalatica* (*Cupriavidus oxalaticus*) conjugative transposon Tn4371 and the *Mesorhizobium loti* MAFF303099 plasmid pMlb. 98
Figure 5-3 Comparison of the genetic organisation of 28 ICESym clusters 103
Figure 5-4  Structure of various *oriT* sequences belonging to the R1162 family. 107
Figure 5-5  Conserved motifs within the ICEMISym<sup>R7A</sup> *oriT*. 108
Figure 5-6  Amino acid similarity between Msi172, Msi171, Msi170 and the Tn4371 protein RO00034. 112
Figure 5-7  Distinct Xre-families conserved on ICESym clusters. 115
Figure 5-8  Incongruence of Xre-protein phylogeny with that of Msi170. 116
Figure 5-9  Putative operator structure upstream of *msi169* and *msi168a*. 117
Figure 5-10  Conservation of Xre-family proteins on quorum-sensing regulated plasmids. 119
Figure 5-11  Putative *tra*-boxes upstream of *traI1* and *traI2*. 120
Figure 6-1  Complementation plasmids. 128
Figure 6-2  CV026 bioassay demonstrating induction of AHL production by R7A(pJR206)/R7A(pJR174). 129
Figure 6-3  100% excision of ICEMISym<sup>R7A</sup> in R7A(pJR174). 130
Figure 6-4  Reverse-phase TLC analysis of *M. loti* AHL extracts. 132
Figure 6-5  Construction and PCR confirmation of in-frame deletions in *traR*, *traI2*, *msi172*, *msi171*, *msi170-msi169* and *traI1*. 134
Figure 6-6  The effect of *traR* and *traI1* mutations on ICEMISym<sup>R7A</sup>-excision. 136
Figure 6-7  The effect of *traR* and *traI1* mutations on AHL production. 137
Figure 6-8  The effect of the *traI2* mutation on ICEMISym<sup>R7A</sup> excision. 139
Figure 6-9  The effect of the *traI2* mutation on AHL production. 140
Figure 6-10  The effect of *msi171* and *msi172* mutations on ICEMISym<sup>R7A</sup> excision. 142
Figure 6-11  The effect of *msi172* and *msi171* mutations of AHL production. 143
Figure 6-12  The effect of constitutive *msi170* expression on AHL production. 145
Figure 6-13  The effect of constitutive expression of *msi170* on ICEMISym<sup>R7A</sup>-excision. 146
Figure 6-14  The effect of *msi169* mutations and constitutive *msi170* expression on AHL-production. 148
Figure 6-15  The effect of *msi169* mutations on ICEMISym<sup>R7A</sup> excision. 150
Figure 6-16  Over-expression of TraR in *E. coli*. 154
Figure 6-17  Partial purification of Msi170. 155
Figure 6-18  AHL production by R7A*. 157
Figure 6-19  Transmission electron microscopy of strains R7A, R7A(pJR174) and R7A*Δ*traR*(pJR174) 158
Figure 7-1  Model of ICEMISym<sup>R7A</sup> transfer and its regulation. 171
List of Tables

Table 2-1 Antibiotic concentrations 33
Table 2-2 Bacterial strains and plasmids 33
Table 2-3 Primers used in this study 39
Table 2-4 Primers used to construct mutant alleles using overlap-extension PCR 42
Table 3-1 Expression of lacZ from 5'-nested attP fragments 58
Table 3-2 Expression of lacZ from 3'-nested attP fragments 58
Table 6-1 Acyl-homoserine lactones detected in M. loti by LC-MS/MS* 133
Table 6-2 Expression from the msi169 promoter 151
# List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.a.</td>
<td>amino acid</td>
</tr>
<tr>
<td>ACP</td>
<td>acyl carrier protein</td>
</tr>
<tr>
<td>AHL</td>
<td>acyl homoserine lactone</td>
</tr>
<tr>
<td>ANOVA</td>
<td>analysis of variance</td>
</tr>
<tr>
<td>Ap&lt;sup&gt;R&lt;/sup&gt;</td>
<td>ampicillin resistant</td>
</tr>
<tr>
<td>att</td>
<td>integrase attachment site</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>bv.</td>
<td>biovar</td>
</tr>
<tr>
<td>°C</td>
<td>degrees Celsius</td>
</tr>
<tr>
<td>C</td>
<td>control (as in control protein)</td>
</tr>
<tr>
<td>Ct</td>
<td>cycle at threshold fluorescence</td>
</tr>
<tr>
<td>cfu</td>
<td>colony forming units</td>
</tr>
<tr>
<td>Cm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>chloramphenicol resistant</td>
</tr>
<tr>
<td>CU</td>
<td>codon usage</td>
</tr>
<tr>
<td>cv.</td>
<td>cultivar</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>ddH&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>double distilled H&lt;sub&gt;2&lt;/sub&gt;O</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>deoxynucleotide triphosphate</td>
</tr>
<tr>
<td>dsDNA</td>
<td>double stranded DNA</td>
</tr>
<tr>
<td>Dtr</td>
<td>DNA transfer and replication</td>
</tr>
<tr>
<td>E</td>
<td>QPCR efficiency</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>et al.</td>
<td>et alii</td>
</tr>
<tr>
<td>Fis</td>
<td>factor for inversion stimulation</td>
</tr>
<tr>
<td>F</td>
<td>the F (fertility) plasmid/episome</td>
</tr>
<tr>
<td>g</td>
<td>gram</td>
</tr>
<tr>
<td>g</td>
<td>gravitational force</td>
</tr>
<tr>
<td>GFP</td>
<td>green fluorescent protein</td>
</tr>
<tr>
<td>Gm&lt;sup&gt;R&lt;/sup&gt;</td>
<td>gentamicin resistant</td>
</tr>
<tr>
<td>G/RDM</td>
<td>glucose/RDM</td>
</tr>
<tr>
<td>h</td>
<td>hour</td>
</tr>
<tr>
<td>HGT</td>
<td>horizontal gene transfer</td>
</tr>
<tr>
<td>HSL</td>
<td>homoserine lactone</td>
</tr>
<tr>
<td>HTH</td>
<td>helix-turn-helix</td>
</tr>
<tr>
<td>ICE</td>
<td>integrative conjugative element</td>
</tr>
<tr>
<td>IHF</td>
<td>integration host factor</td>
</tr>
<tr>
<td>Inc</td>
<td>plasmid incompatibility group</td>
</tr>
<tr>
<td>kb</td>
<td>kilo-base pairs</td>
</tr>
<tr>
<td>Km&lt;sup&gt;R&lt;/sup&gt;</td>
<td>kanamycin resistant</td>
</tr>
<tr>
<td>L</td>
<td>litre</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani</td>
</tr>
<tr>
<td>LC</td>
<td>liquid chromatography</td>
</tr>
<tr>
<td>LC-MS/MS</td>
<td>liquid chromatography coupled to hybrid quadrupole-linear ion trap mass spectrometry</td>
</tr>
<tr>
<td>lux</td>
<td>luxICDABEG</td>
</tr>
<tr>
<td>log&lt;sub&gt;10&lt;/sub&gt;</td>
<td>mathematical log, base 10</td>
</tr>
<tr>
<td>M</td>
<td>molar</td>
</tr>
<tr>
<td>Mb</td>
<td>mega-base pairs</td>
</tr>
<tr>
<td>MGE</td>
<td>mobile genetic element</td>
</tr>
<tr>
<td>Milli-Q</td>
<td>filter sterilised ddH&lt;sub&gt;2&lt;/sub&gt;O</td>
</tr>
<tr>
<td>min</td>
<td>minutes</td>
</tr>
<tr>
<td>Mpf</td>
<td>mating pair formation</td>
</tr>
<tr>
<td>MU</td>
<td>Miller units</td>
</tr>
<tr>
<td>MTase</td>
<td>methyltransferase</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>m/z</td>
<td>mass to charge ratio</td>
</tr>
<tr>
<td>NBD</td>
<td>nucleotide binding domain</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>NTP</td>
<td>nucleotide triphosphate</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>ONPG</td>
<td>ortho-nitrophenyl-β-galactoside</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>ori&lt;sup&gt;V&lt;/sup&gt;</td>
<td>origin of replication</td>
</tr>
<tr>
<td>ori&lt;sup&gt;T&lt;/sup&gt;</td>
<td>origin of transfer</td>
</tr>
<tr>
<td>P</td>
<td>promoter</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PSSM</td>
<td>position site scored matrix</td>
</tr>
<tr>
<td>QPCR</td>
<td>quantitative PCR</td>
</tr>
<tr>
<td>QS</td>
<td>quorum sensing</td>
</tr>
</tbody>
</table>
RC  rolling circle
RDF  recombination directionality factor
RDM  rhizobium defined medium
REase  restriction endonuclease
RM  restriction modification
RNA  ribonucleic acid
RNAP  RNA polymerase
rpm  revolutions per minute
rRNA  ribosomal RNA
RU  relative units
s  seconds
sp.  species
SDS  sodium dodecyl sulphate
ssDNA  single stranded DNA
t  time
trb  Mpf-encoding genes
T4SS  Type four secretion system
Tc^8  tetracycline resistant
TLC  thin layer chromatography
tmRNA  transfer messenger RNA
tRNA  transfer RNA
TE  Tris-EDTA
UTR  untranslated region
UV  ultraviolet
V  volt
v/v  volume/volume ratio
w/v  weight to volume ratio
Chapter 1
General
Introduction
A mobile genetic element (MGE) is a DNA molecule able to facilitate its own movement within or – more importantly – between genomes in a population. Available genome sequence data suggest the latter, termed horizontal gene transfer (HGT), is a major driving force in prokaryotic evolution. The movement of MGEs and the genes that they carry into new genomic backgrounds can result in the rapid formation of novel gene combinations, providing bacteria with an evolutionary shortcut to increased fitness (for reviews see references (28, 55)). MGEs most often carry genes with a selective advantage, which allows them to propagate vertically by hitchhiking with the fitness they endow their host (104).

Bacteriophage, plasmids, and conjugative transposons have been implicated in propagating a variety of traits including antibiotic resistance, virulence, toxin production, accessory metabolic pathways and genes required for colonisation of eukaryotic hosts. The more recently appreciated “genomic islands” (also called “fitness islands”, “symbiosis islands” or “pathogenicity islands”) carry a similar spectrum of traits and are ubiquitous in prokaryotes (115, 138).

Genomic islands are large chromosomal gene clusters that have been horizontally acquired and may be self-transmissible (28). They often encode a phage-like integrase and genes for conjugative transfer. As the majority of islands are untested for mobility, they are usually identified in silico by the presence of atypical DNA sequence composition, integration adjacent to tRNA genes, or by their absence in related isolates. The few mobile genomic islands studied excise from the chromosome forming a circular DNA intermediate, and in characterised cases transfer to a recipient either by conjugation or by packaging in phage particles (135). The elements that transfer by conjugation have been collectively termed “integrative and conjugative elements” or ICEs (25).

This thesis reports a study of the mechanism and regulation of HGT used by one such ICE, the symbiosis island ICEMISym$^{R7A}$ of Mesorhizobium loti strain R7A. ICEMISym$^{R7A}$ is the largest known mobile ICE at 502 kb and contains over 400 open reading frames including genes required for a nitrogen-fixing symbiosis with legumes of the genus Lotus (182). The following sections of this chapter give a brief introduction to the symbiosis, Mesorhizobium loti R7A and the discovery and sequence analysis of the ICEMISym$^{R7A}$, followed by more detailed discussion of the evolution of MGEs and common mechanisms used by MGEs for transfer, maintenance and regulation. Finally the genetics of currently characterized ICEs is discussed along with previous studies of ICEMISym$^{R7A}$. 
1.1 The rhizobium-legume symbiosis

The relationship between rhizobia and legumes involves a series of interactions including multiple plant and bacterium-specific signals that ultimately result in the exchange of resources between the bacterium and the legume. The symbiotic bacteria are able to fix atmospheric nitrogen ($N_2$) converting it to ammonia ($NH_3$) which is then assimilated by the plant host. In return the host supplies the microsymbiont within nodules with carbon/energy sources (for reviews see references (60, 91)).

The rhizobium-legume relationship begins with the legume releasing specific diffusible compounds called flavonoids which are perceived by the compatible bacterium through binding to a LysR-type transcriptional regulator NodD. The flavonoid-activated NodD activates expression of bacterial genes that encode for the production of specific lipochitooligosaccharide molecules called Nod-factors (143). (While this is generally the case, recently it has been shown that Bradyrhizobium BTAi1 enters symbiosis through an alternative pathway (63)). The cognate Nod-factor stimulates multiple responses in the plant including cortical cell division and curling of root hairs (53), which serves to encapsulate the bacterial colony. Through this structure the bacteria induce the plant to form an infection thread, within which they migrate to the root cortex (143). The rhizobia are then internalised within the plant cells into structures called symbiosomes, where the bacteria differentiate into nitrogen-fixing cells called bacteroids. In temperate legumes such as *Medicago* and *Trifolium*, nodules grow in size through the supply and infection of new plant cells derived from a persistent meristem (termed indeterminate nodules) (143) whereas in tropical legumes such as *Lotus* the meristem is not persistent and the increase in nodule size is due to expansion and division of infected plant cells (termed determinate nodules). An interesting side-effect of this distinction is that only bacteroids from determinate nodules (such as those formed by *M. loti* R7A on *Lotus*) are likely to survive nodule senescence and return to a free-living lifestyle following the symbiosis (91).
Figure 1-1  Nodule infection by *M. loti* R7A

Fluorescent microscopy (20x) image of a cross-section of a *Lotus corniculatus* root inoculated with *M. loti* R7A containing a plasmid expressing green fluorescent protein (GFP). This photo was taken by Patsarin Rodpothong (2007).

The rhizobial genes required for symbiosis by rhizobia are often encoded on large plasmids (Sym-plasmids) or less frequently on genomic islands. The only completely conserved genes on these elements appear to be those encoding the nitrogenase enzyme and its regulation (*nif* and *fix* genes) and those for Nod-factor synthesis (*nod* genes) (with the exception of *Bradyrhizobium BTAi1*) (65). The symbiotic genes of *Mesorhizobium loti* R7A are carried on a mobile genomic island called the symbiosis island (180), more recently named ICEMlSymR7A (153). In addition to the *nif*, *fix* and *nod* genes, ICEMlSymR7A encodes many genes of symbiotic relevance including a variety of metabolic genes and a VirB/D4 type IV secretion system (T4SS) (182) capable of transporting proteins to eukaryotic cells, which is involved in the specificity of the plant-rhizobium recognition (84).
1.2 *Mesorhizobium loti* R7A

1.2.1 Phylogeny and description

*Mesorhizobium loti* strain R7A is a derivative of *M. loti* strain ICMP3153 and was isolated from a nodule off a field stand of *Lotus corniculatus* (179). It forms effective (nitrogen-fixing) nodules on *L. corniculatus* and *L. japonicus*. The genus *Mesorhizobium* belongs to the order proteobacteria (‘Purple bacteria and others’) of Gram-negative bacteria, which is named after the mythological Greek god Proteus (175) who was able to take on many forms. As the name suggests the proteobacteria are an extremely morphologically and phenotypically-diverse group of bacteria that include phototrophs, heterotrophs, chemolithotrophs as well as pathogens and symbionts of many animals and plants. The proteobacterial group is extremely large and encompasses a major portion of all known bacterial species (40,150 out of 95,493 species on NCBI as of November 2007 (5)). Phylogenetic sequence-signature evidence suggests that the common ancestors of all extant eubacteria have evolved in a sequential (rather than branched) manner, with the proteobacteria being the most recently emerging group. It has been hypothesised that the breadth of the environmental niches occupied by the proteobacteria may reflect success during ancient selective sweeps where they supplanted the majority of ancestral bacterial species (71, 72).

Within the proteobacteria are five named groups, α, β, γ, δ and ε, all of which form distinct clades with the exception of the δ and ε which are now known to be monophyletic (72). The α-proteobacteria includes the phylogenetically distinct groups *Rhodospirillales, Caulobacterales, Sphingomonadales, Rhodobacterales, Rickettsiales* and importantly to this study the *Rhizobiales* (Figure 1-2). The *Rhizobiales* order encompasses the majority of bacteria able to form a nitrogen-fixing symbiosis with legumes, such as *Bradyrhizobium, Sinorhizobium, Rhizobium, Azorhizobium* and *Mesorhizobium* species; however legumes have also been shown to form relationships with members of the β-proteobacteria (126). There are also many non-symbiotic members of the *Rhizobiales* order such as the phototrophic purple bacterium *Rhodopseudomonas palustris*, the *Brucellaceae* whose members are both human and animal pathogens and the nitrite-oxidising species *Nitrobacter winogradskyi*. Also a member of the *Rhizobiales* is the plant pathogen *Agrobacterium tumefaciens* which causes the formation of plant tumours through
interkingdom gene transfer of the Ti plasmid (62), a feature that has been greatly exploited for use in genetic transformation of plants.

The *Mesorhizobium* genus is a relatively new classification for a number of members previously grouped within the genus *Rhizobium*. The species *Rhizobium tianshanense*, *R. ciceri*, *R. huakuii*, *R. mediterraneum*, and *R. loti* were reclassified as a monophyletic group by Jarvis (88) and the genus renamed *Mesorhizobium* due to their intermediate position in both growth-rate and phylogeny between the *Rhizobium* and *Bradyrhizobium*. Non-symbiotic mesorhizobia have also been identified (178).

The mesorhizobia are aerobic chemoorganotrophs able to utilise various carbohydrate (excluding cellulose) and nitrogen sources. Optimal growth conditions are observed between 25-30°C; however plate cultures will grow slowly at 4°C. They are tolerant of a pH range of 4-10 (depending on the species) and produce acid when glucose is used as a sole carbon source. During some growth conditions (e.g. bacteroids) they synthesise carbon reserves in the form of large poly-ß-hydroxybutyrate granules. Mesorhizobia are generally rod-shaped but during stress conditions they can appear oval or more irregularly shaped. In free-living conditions mesorhizobia are motile, being monotrichous or peritrichous (37). Complete genome sequences are available for *M. loti* MAFF303099 (93) and *Mesorhizobium* sp. BNC1 (accession #CP000390). These strains carry single chromosomes of ~ 7 Mb and 4 Mb and carry 2 and 3 plasmids respectively. The *Mesorhizobium loti* R7A genome closely resembles that of MAFF303099 at the DNA sequence level but it is not known to carry any plasmids (182).
Figure 1-2 Phylogeny of the α-proteobacteria

A neighbour-joining bootstrap consensus tree for α-proteobacteria based on 16S rRNA sequences, taken from Gupta (72).
1.2.2 Discovery and characterisation of ICEMlSym\textsuperscript{R7A}

The only \textit{Lotus} species present in New Zealand have been introduced and native rhizobia present in New Zealand soils are unable to nodulate \textit{Lotus} species. In 1986 the agronomic utility of introducing \textit{L. corniculatus} in combination with \textit{M. loti} strain ICMP3153 (also known as NZP2238) into acidic tussock grasslands was tested in Central Otago at a site shown previously to lack rhizobia able to nodulate \textit{Lotus}. Seven years later (1993) isolates from \textit{L. corniculatus} nodules were analysed by 16S rRNA sequencing and restriction fragment length polymorphisms, which revealed them to be a mixture of diverse mesorhizobia related to \textit{M. loti} and \textit{M. huakuii} along with re-isolates, including strain R7A, of the original strain (179). However all the new symbiotic strains had an identical DNA hybridisation profile to that of ICMP3153 when probed with symbiotic genes, indicating that they had acquired their symbiotic DNA from the original inoculant strain (179). Non-symbiotic isolates related to the newly discovered diverse symbionts were also isolated and shown to be auxotrophic for biotin, thiamine, and nicotinate (180).

The chromosomal transfer of symbiotic genes from strain ICMP3153 to the various non-symbiotic isolates was demonstrated in the laboratory (and again in the field site) and the genes were shown to be carried on a ~500-kb conjugative, integrative element (180). The element was termed the symbiosis island, as it encoded a P4 phage-like integrase at its left end and was integrated adjacent to a \textit{phe}-tRNA gene; features common to pathogenicity islands. The DNA sequence of the entire symbiosis island was obtained, which revealed it to be 502 kb in size, encoding over 400 open reading frames (ORFs) and flanked by an identical 17-bp direct repeat sequence (182).

An analysis of the symbiosis island and a comparison to a similar island in \textit{M. loti} MAFF303099 was carried out by Sullivan \textit{et al.} (182) and revealed the conservation of a near-identical (at the nucleotide sequence level) 248-kb core region between the two elements, indicating that they diverged recently. Along with carrying genes likely involved in symbiosis, both elements contain several clusters of genes with a likely role in conjugative transfer. These include a conjugative T4SS (distinct from the VirB/D4-like T4SS), an essential transfer protein TraF and homologues of TraR (Msi174) and TraI (Msi173 and Msi039) which regulate transfer of the \textit{A. tumefaciens} Ti-plasmid by quorum sensing.
1.3 Evolution and genetics of horizontal gene transfer

1.3.1 Sex and recombination

In contrast to sexually-reproducing eukaryotes, bacteria are haploid and reproduce clonally. Early on this led to the perception that bacteria evolved within very distinct and genetically isolated lineages. This view was expressed by Muller (127) in his description of the evolution of the hypothetical genotype ABC from the less fit ancestral genotype abc. A strictly asexually reproducing population is restricted to evolving the ABC genotype sequentially (e.g. from Abc>ABC), with each new mutation occurring in the same lineage as the prior. This strategy will expend an excessive amount of evolutionary time to reach ABC if it does so at all, especially if the intermediary genotypes are no more fit than the original abc. A sexually reproducing species however is able to bring together individuals carrying intermediate genotypes (e.g. Abc and aBC), and through genetic recombination, take a shortcut to generate the fittest genotype.

It is now realised that bacteria are anything but sexually deprived; in fact their genomes most often appear as chimeras of vertically and horizontally acquired regions. For example an analysis of the recently-sequenced genome of Bradyrhizobium sp. BTAi1 revealed the single 8.3-Mb chromosome is host to 29 genomic islands which together make up around one fifth of the chromosomes total size (63). Unlike the sexual cycle in eukaryotes, sex in bacteria occurs through the unidirectional transfer of DNA from one individual to another, including between members of different species. As already mentioned, mobile elements can carry genes conferring a selective fitness advantage. Therefore when these elements arrive in a new host and form an individual with fitness greater than that of the parents, the evolutionary restrictions outlined in Muller’s asexually reproducing population are overcome. Therefore it would appear that HGT benefits the bacterial population as a whole, with MGEs acting as an evolutionary toolbox for the rapid generation of new, fitter individuals (104).

Nevertheless, while MGEs may aid in the evolution of a bacterial community as a whole, selection at this level is strictly indirect (104). When considering selection at the level of the individual bacterium, altruistic donation of potentially beneficial DNA to a fellow bacterium could be both physically and evolutionarily detrimental to the donor’s fitness. However in a
population with dynamic selection pressures, the fittest genotypes will likely be borne of HGT events, leading to the selection for hosts containing MGEs that have (at least recently) maintained their mobility (19, 104). These conflicting selection pressures on MGEs have likely resulted in the complex and often selfish strategies and regulatory mechanisms MGEs employ to ensure their maintenance (and likely their mobility) within populations. For example, numerous MGEs encode post-segregational toxin-antitoxin systems capable of killing the host upon loss of the element (89).

From this perspective, MGEs could be viewed solely as selfish entities, propagating themselves at the expense of the host. Bacteriophages exemplify this notion, as many lack any host-advantageous genes (apart from immunity to super-infecting phage) and their HGT often results in death of the host (10). However many MGE have much lower rates of horizontal transmission than bacteriophages, meaning that a much greater proportion of their success depends on their vertical transmission with the host (104). This has likely led to the carrying of host-advantageous genes by MGEs, which at least temporarily outweighs the disadvantage of carrying and transferring the element. Under this model, HGT is only favoured when the rate of horizontal propagation is greater than the associated reduction in vertical propagation (relative to competing strains) it imposes (49). As is discussed in the following sections, MGEs have evolved elaborate mechanisms to carefully regulate initiation of HGT.

### 1.3.2 Population-level persistence of MGEs

While abundant genomic evidence of HGT is available, there is a paucity of empirical data on the population-level persistence and ecology of MGEs. The particular conditions where they would be predicted to persist also remain unresolved. Bergstrom et al. (19) argue that the long-term persistence and mobility of plasmids requires frequent selective sweeps. Under these conditions MGEs shuttle advantageous genes between emerging chromosomal variants, repeatedly hitchhiking with new successful hosts. Their mathematical model predicts that if the genes carried by the plasmid have a high selective advantage they will eventually be sequestered by the host chromosome, to bypass the burden of carrying and transferring the plasmid. Similarly, an absence of selection for genes carried on the MGE also predicts loss of the element from the population. Lili et al. (107) suggest however that by reducing the stringency of some parameters of the model, even purely parasitic plasmids may persist indefinitely in a population. A study
carried out by Dionisio et al. (44) that investigated the persistence of the R1 plasmid in E. coli in the absence of selection tends to support this notion. Despite R1 conferring a fitness burden on the host, it was never lost within 250 generations. Not only were the plasmids retained, but the fitness cost of the plasmids disappeared in all experiments, through changes in both the plasmid and the host. In summary, it is likely that a dynamically changing environment indirectly increases the prevalence of MGEs in the bacterial community. However the exact nature of the individual selective pressures acting on MGEs even in simple environmental scenarios is complex and nontrivial to model mathematically.

Following the introduction of M. loti at the test site in central Otago by Sullivan et al. (179), rapid emergence of new Lotus symbionts arising through transfer of ICEMISym\textsuperscript{R7A} to native bacteria was observed (179). More recently similar studies have been carried out in Western Australia, with the co-introduction of the legume Biserrula pelecinus and the rhizobial symbiont Mesorhizobium ciceri bv. biserrulae (130-132). Like M. loti R7A, the various M. ciceri bv. biserrulae strains carry the symbiotic genes on chromosomal symbiosis islands that are also integrated at the phe-tRNA gene. They also encode homologues of the ICEMISym\textsuperscript{R7A} gene intS (>93.5% nucleotide identity) (130), which is encoded as the first gene adjacent to the phe-tRNA gene. As with the New Zealand study, over a short period of time (5-6 years) the nodules became occupied with diverse mesorhizobia that had acquired the symbiosis island from the inoculating strain. (Interestingly some of the new isolates had near identical 16s rRNA to the non-symbiotic isolate CJ5 identified in New Zealand by Sullivan et al. (178)). An analysis of a commercially-grown plot of B. pelecinus inoculated with M. ciceri bv. biserrulae revealed that after 6 years almost 50% of nodules were occupied by new diverse strains. Most surprisingly, these strains were all found to be less efficient in nitrogen fixation than M. ciceri. bv. biserrulae and several were completely ineffective (131). This suggests that not only are these symbiotic islands promiscuous, but the newly emerging strains are highly competitive and are able to supplant the inoculant strain in these nodules, seemingly regardless of the impact on symbiosis.

1.4 Mobile genetic elements

Movement of DNA between bacteria facilitated by MGE occurs either by transduction or conjugative transfer. Transduction is a bacteriophage-mediated process and involves the packaging of nucleic acid into a diffusible, infectious, non-living particle. Conjugation, as used
by plasmids and ICEs, requires cell to cell contact and involves active pumping of DNA through a needle-like pilus attachment (55). Other MGE such as transposons, satellite bacteriophages and non-conjugative plasmids or genomic islands that are unable to move between bacteria autonomously can sometimes make use of an accessory element. The satellite bacteriophage P4 for example lacks genes required for phage particle formation, but instead packages its DNA within the virion of a helper phage P2 (108). The following section discusses the mechanisms used by bacteriophages and plasmids for maintenance, mobility and regulation.

1.4.1 Regulation of the bacteriophage lifecycle

Bacteriophage (phage) are bacterial viruses, obligate bacterial parasites first discovered through their ability to lyse bacterial cells and form infectious clear plaques on plate cultures (186). Genomes can be single or double-stranded DNA or RNA ranging in size from ~1-500 kb (8). The phage nucleic acid and sometimes accessory proteins are packaged within a protein and/or phospholipid envelope (capsid). Infection involves attachment of the phage to specific receptors on the bacterial cell and delivery of the nucleic acid (and often protein) payload through injection. Packaging of DNA into the capsid usually requires the presence of specific cis-acting sequences. However occasionally with certain types of phage non-specific bacterial DNA may be packaged, which can then be incorporated into the recipient’s chromosome through homologous recombination (generalised transduction). These features have made phage an invaluable tool for genetic manipulation and mapping for several decades.

Following infection by a virulent phage, the phage genome is replicated and new phage particles are rapidly produced at the expense of the host’s resources, often lysing the cell in the process (lytic cycle). Temperate phage such as λ of *Escherichia coli* can enter a lysogenic phase in which they remain dormant in the cell until conditions signal entry into the lytic phase. During the lysogenic phase the phage DNA may persist in the cell either by integrating into the chromosome (commonly called a prophage) or more unusually by replicating as a plasmid. In some cases a lysogenic phage may confer a phenotype, such as immunity to an invading phage or increased virulence in pathogenesis. For example, the Stx phage encodes the shiga toxin produced by the dysentery-causing *E. coli* strain O157:H7 (200).
1.4.1.1 Chromosomal integration

Integration into the host’s chromosome during lysogeny for most phages occurs through site-specific DNA recombination. For dsDNA phages such as λ and P4, this reaction is catalysed by a tyrosine DNA recombinase, also called an integrase, similar to that encoded by the ICEMI/Sym<sup>R7A</sup> gene intS. Integrase catalyze site-specific recombination between short (near)-identical DNA sequences (often part of tRNA genes), resulting in the integration of the circular phage DNA into the chromosome through a single recombination event. The integrase proteins simultaneously bind short “core” DNA sequences at the site of recombination and flanking “arm” sequences collectively referred to as attachment sites (att) (Figure 1-3). Core sequences are found within the attachment site (attP) of the circular phage DNA, the integration site on the host chromosome (attB) and as a direct repeat flanking the integrated element at the prophage-chromosome junctions attL and attR (reviewed in (70)). ICEMI/Sym<sup>R7A</sup> is flanked by a 17-bp direct repeat resembling these core sequences that will hereafter be referred to as the “core”.

Recombination by integrases is highly directional and favours integration (the formation of attL and attR) in the absence of additional factors. In most cases the excision function of an integrase is stimulated by the presence of a recombination directionality factor (RDF), also termed an excisionase, that binds and bends DNA within the attachment sites to promote excisive recombination (162). Many putative and characterised RDFs have now been identified and found to be associated with prophage, genomic island, transposon and even plasmid-encoded integrases. RDFs are small (generally <100 amino acids (a.a.)) proteins and usually have a basic pH. Many contain a helix-turn-helix DNA-binding motif and have been implicated as also having a role as transcriptional regulators. They exhibit considerable variation in amino acid sequence, including between members belonging to the same family (105).
Figure 1-3  Structure of the $\lambda$, $attP$, $attB$, $attL$ and $attR$ sites.  

The Xis-binding sites are shown as pentagon symbols, Int arm-type binding sites are shown as arrows labelled P.  The core region is indicated by inverted arrows labelled C or B.  IHF and Fis-binding sites are shown by circles and hexagons respectively.  Filled symbols indicate sites bound immediately prior to integration (above) or excision (below).  Figure taken from reference (7).

1.4.1.2 Regulation of lysogeny

A prophage can lie dormant and propagate vertically with the host’s genome until such time as the lytic cycle is initiated.  In bacteriophage $\lambda$, this decision is controlled by a complex bistable switch involving multiple DNA-binding transcriptional regulators.  The pair of regulators central to this switch is Cro and cI, which have antagonistic roles in the maintenance of lysogeny.  When cI dominates, lysogeny is maintained while Cro must dominate for entry into the lytic phase (reviewed in references (45, 136, 147)).

The Cro and cI proteins are members of a family of proteins that encode the conserved domain cd00093 (557 members) (116), collectively referred to as Xre-family proteins due to their similarity to the repressor protein encoded by the PBSX prophage (119).  Members of the Xre
family contain the highly conserved helix-turn-helix (HTH) motif required for DNA binding. The mechanism of transcriptional activation by members of this family is likely similar to that of the λ cI protein. Direct interaction of a conserved patch of acidic residues on the surface of cI with a region on RNA polymerase (RNAP) stabilises interactions between the -35 region of the $P_{RM}$ promoter and the $\sigma^{70}$-subunit of RNAP, stimulating isomerisation to an open transcriptional complex. Not all members are necessarily transcriptional activators - for instance the Cro protein lacks the acidic patch and the ability to activate - however an artificial Cro protein containing the acidic patch from cI is able to activate (29). Xre proteins bind as dimers to DNA sequences called operator sequences that are often palindromic and are located near or within the promoter region of the transcriptional start site. Multiple adjacent operators may be present and binding of Xre dimers to these additional operators can augment or silence transcriptional activation (45, 136, 147).

The cI and Cro proteins have an antagonistic effect on each other’s expression, accomplished by binding common operator sequences $O_{R1}$, $O_{R2}$ and $O_{R3}$ with differing affinities. The cI protein activates its own expression and blocks expression of Cro by binding two of the operator sequences between cI and cro ($O_{R1}$ and $O_{R2}$). Binding of Cro to $O_{R1}$, $O_{R2}$ and $O_{R3}$ blocks expression of cI and allows weak transcription from the cro promoter (45, 136).

Entry of λ into the lysogenic phase after infection is controlled by another Xre family protein cII which like cI activates its own expression through positive autoregulation. Accumulation of cII activates the expression of int to stimulate integration and expression of cI (not expressed in the absence of cII or cI) and cI in turn activates its own expression while repressing cro and lytic genes. The cII protein also represses expression of the genes encoding the antiterminator proteins N and Q, that when present allow elongation of transcripts encoding genes required for the lytic cycle. Upon integration of λ into the genome, the cI protein actively maintains lysogeny through binding of additional operator sequences $O_{R3}$, $O_{L1}$, $O_{L2}$ and $O_{L3}$. Distinct from other Xre proteins cI encodes an additional N-terminal domain which allows higher-order oligomerisation of the protein. Looping of λ DNA juxtaposes the cI-bound operator sequences which facilitates the formation of cI tetramers and octomers and blocks the lytic promoters $P_R$ and $P_L$ as well as the cI promoter $P_{RM}$ (45, 136).

Following exposure to DNA-damaging conditions such as ultraviolet light, the SOS response of E. coli induces the expression of the RecA protein. Activated RecA protein induces autocleavage
of CI, resulting in the derepression of cro expression (128). The Cro protein then blocks expression of CI and allows weak expression of the PR and PL, committing λ to the lytic cycle. Following expression from these promoters, the antiterminator N facilitates elongation of transcripts from PR and PL. Genes required for the lytic cycle are then expressed, including xis which encodes Xis, the RDF required for Int-catalysed excision.

The λ regulatory circuit illustrates the complexity that has evolved to control the decision either to propagate with the host or to infect surrounding cells. A primary function of this regulatory circuit is to coordinate a bistable switch either completely activating the lytic pathway or completely repressing it. The cascaded levels of gene expression involved in activation of the lytic cycle allow a minimal footprint while in the lysogenic phase, as nearly all genes are under tight repression until the later stages of the lytic cycle. The circuit is also able to perceive physiological changes in the cell through the action of RecA on CI. Another interesting feature of the system is that the decision made by an infecting phage to enter the lysogenic stage is regulated by the concentration of cII, which is directly dependent on the number of phage infecting an individual cell. In cells with only one infecting phage, the lytic cycle will continue; however if two or more phages are present, the cII concentration is such that formation of a lysogen will occur.

1.4.2 Plasmid maintenance

The term plasmid was originally used by Lederberg (102) to describe all “extrachromosomal hereditary determinants”, while Jacob and Wollman (87) used the term episome to describe a broader group of elements, optionally associated with or even integrated into chromosome. The term plasmid is now used almost exclusively to describe heritable cytoplasmic DNA molecules that are replicated independently from the chromosomal DNA, distinguished from chromosomes by encoding only non-essential or accessory genes. Bacterial plasmids are most commonly covalently-closed circular supercoiled dsDNA molecules, although linear dsDNA plasmids have also been observed in both Gram-positive and Gram-negative bacteria (82). The first-discovered bacterial plasmid was the E. coli fertility factor plasmid F, originally observed in a chromosomally integrated form for its ability to promote transfer of chromosomal markers from Hfr strains to other F- strains via cell to cell contact (198). The plasmid genome database (3) now lists (as of December 18th, 2007) 1069 sequenced plasmids found throughout the three kingdoms
of life. Their sizes vary considerably from the tiny mycobacterial plasmid pBG7AU at 1022 bp up to the 2.1-Mb plasmid pGMI1000MP of *Ralstonia solanacerum*. The following section describes the current knowledge of the mechanisms plasmids use to replicate, partition, and persist.

### 1.4.2.1 Replication and incompatibility

The replication modules of plasmids contain a *cis*-acting region called the *ori* (or *oriV*), which is recognised by the various plasmid- and host-encoded proteins and RNA molecules involved in initiation of replication and its control (42). The control of plasmid copy number is vital to the plasmids survival within the dividing cell - too few will result in the plasmid being lost through segregation while too many could place an unnecessary burden on the cell. The mechanisms of plasmid replication generally fall into one of three classes, theta, strand-displacement and rolling circle (RC). Theta replication is the most common mechanism amongst plasmids studied, while strand displacement and RC replication are less common and mechanistically distinct. In theta and strand-displacement, replication may initiate uni- or bi-directionally from single or sometimes multiple origins (43). In contrast RC replication is unidirectional and involves nicking of a single strand of the plasmid DNA, which is then displaced from the unnicked strand as it is replaced by replication around the circular ssDNA template. The free nicked ssDNA is then recircularised and replicated independently, restoring it to dsDNA. Currently RC replication as a primary mechanism of replication is only known to occur in smaller plasmids (generally less than 10 kb) (94).

Multiple plasmid species may be maintained within a single strain; however each plasmid must have a distinct replication system for stable maintenance. This is because the specific DNA-binding proteins and regulatory elements controlling replication are unable to distinguish between identical *ori* sequences. Therefore the two distinct plasmids with the same *ori* will be controlled as a single replicon, reducing the copy number of both plasmids and allowing either plasmid to be lost simply through random segregation. This situation quickly leads to the loss of one of the plasmids and stable maintenance the other plasmid. This phenomenon was observed early on in plasmid research and used to classify closely related plasmids into incompatibility (Inc) groups (for a review see reference (134)).
1.4.2.2 Plasmid partition

For stable inheritance a plasmid must ensure daughter cells carry at least one copy of the plasmid following cell division. This can be accomplished either by having a high copy-number or by physically positioning plasmid copies into each daughter cell during cell division. The latter is termed plasmid partition and is analogous to the process of pairing and separation of chromosomes in eukaryotes during mitosis/meiosis. As in eukaryotes, plasmid partition involves the pairing through cis-acting centromere-like sites called parS. The parS sites are bound and paired by the DNA-binding protein ParB and then separated to opposite cell poles by the ATPase ParA. Homologues of the ParA and ParB proteins are ubiquitous in bacteria and archaea and include chromosomal partitioning relatives MinD and MinE. Like replication, the specificity of the ParAB interaction with the parS locus precludes compatibility with plasmids containing identical partition apparatus. However the understanding of partition incompatibility and partition itself are less complete than for replication (for a review see reference (22)).

1.4.2.3 Post-segregational killing systems

If the individual copies of a plasmid fail to be correctly partitioned during cell division, one of the daughter cells may fail to inherit the plasmid. This reduces the prospect of long-term persistence of the plasmid, as even a slight fitness advantage for the plasmid-free cell could lead to rapid loss of the plasmid from the population. Presumably in a mechanism to offset any advantage of plasmid loss, several plasmids carry modules conferring a post-segregational killing effect that leads to the death of any plasmid-free cells (also known as plasmid addiction) (89). A well-studied example is the hok-sok module of the R1 plasmid in E. coli, which encodes the Hok toxin, initially borne of a translationally-inactive transcript. Three-prime processing of the hok transcript results in the activation of translation of Hok protein, but its translation is inhibited by the binding of the complementary sok transcript. Upon loss of the plasmid, the hok transcripts quickly gain translational activity and in the absence of a renewable source of sok produce Hok resulting in the death of the cell (184). The majority of other toxin-antitoxin systems studied involve the production of stable toxin proteins along with unstable antitoxin proteins (89).

Another form of post-segregational killing system is the restriction modification module (RM) (75, 97). The RM is composed of a restriction endonuclease (REase) enzyme and a complementary methyltransferase (MTase) that modifies the host and plasmid DNA. Following
entry of the plasmid into a new cell, the MTase is expressed first, methylating the DNA at specific sites, which protects it from digestion by the REase within the same site. However if an established module is lost, replication of the host DNA will produce hemimethylated and eventually unmethylated DNA, which is then vulnerable to attack by the stable REase enzyme. An interesting side effect of the module is that its presence in the cell confers immunity against incoming foreign and possibly harmful unmethylated DNA (e.g. lytic phage) (183). Differential regulation of MEase and RTase expression is often accomplished by an Xre family protein termed the control protein (C) (129).

1.4.3 Conjugative transfer

Conjugative transfer of DNA involves stable cell-cell contact, during which DNA is actively pumped from one cell to another through a pilus. The two major components of the system are the mating pore formation (Mpf) complex and the DNA transfer and replication complex (Dtr). The Mpf is composed of a modified T4SS which is coupled to the Dtr by a specific coupling protein. The Dtr is composed of a cis-acting sequence that is nicked by and covalently attached to a protein called a relaxase. As the relaxase-attached strand is transferred through the T4SS, the remaining ssDNA is replicated, while the transferred strand is recircularised and replicated in the recipient. Conjugative transfer has been more simply described as the secretion of protein-tethered DNA substrate by a T4SS, coupled with RC replication (for a reviews see references (101, 110)).

1.4.3.1 Type IV secretion systems

T4SSs are large membrane-spanning complexes composed of a dozen or more structural proteins, often expressed together from a large polycistronic transcript. The conjugation-associated T4SSs can be phylogenetically divided into three groups, F-like (as found on the F plasmid), P-like and I-like (101). More recently a fourth distinct group has been classified, common to genomic islands in *Haemophilus influenzae* (92). The F-group is the most extensively studied and is found on plasmids from many Inc groups. The F T4SS encodes several proteins distinct from protein-secreting T4SS including those involved in mating-pair stabilisation. The F-pili are long and flexible, extending up to 20 µm. The P-pili are much shorter (< 1 µm) and have a broader host range than the F-pili. The P group includes two distinct members, the VirB system responsible
for transfer of the Ti-plasmid and secretion of protein effectors in Agrobacterium and the RP4/RK2-system found on IncPα plasmids (101).

ICE/MISym$^{R7A}$ encodes two P-like T4SS systems, a VirB-like system and an RP4-like system (182). The VirB system of ICE/MISym$^{R7A}$ is required for transport of effector proteins to eukaryotic cells and is involved in symbiosis (84), but is not found on ICE/MISym$^{MAFF}$ which instead encodes a type III protein-secretion system. The RP4-like system of ICE/MISym$^{R7A}$ is comprised of $trbBCDEJLFGI$ (abbreviated $trb$) and $traF$ encoding the Mpf and $traG$ which encodes the coupling protein (182).

1.4.3.2 The relaxosome

Prior to transfer of the ssDNA strand through the mating pore it first needs to be nicked and covalently attached to the relaxase at a specific DNA site called the origin of transfer ($oriT$). In some systems this process is accomplished by the relaxase alone but more often involves the presence of accessory DNA-binding proteins in addition to the relaxase. Interestingly the particular DNA-binding proteins involved in forming the relaxosome generally do not have a long-standing evolutionary history with relaxases overall; rather a variety of proteins appear to have been recruited for the task in different systems (141). For example the relaxase of the transposon Tn916 uniquely requires the binding of its integrase to the $oriT$ (157). The $oriT$ usually contains a short sequence conserved amongst related elements adjacent to a less conserved inverted repeat sequence and is structurally similar to the $ori$ sequences of RC plasmids. The short conserved sequence is nicked by the relaxase, while the inverted repeat sequence is thought to add specificity for the particular relaxase and may be involved in the priming of rolling-circle replication (for a review see (141)).

1.4.3.3 The coupling protein

The coupling protein is an NTP-binding protein that is actively involved in the recruitment of the DNA and/or protein substrates through the Mpf. ICE/MISym$^{R7A}$ has separate coupling proteins in both the VirB-like and RP4-like systems, encoded by $virD4$ and $traG$ respectively. The coupling protein is absolutely required for conjugation and determines the compatibility between the Mpf and the Dtr (74, 110, 166).
1.5 Quorum sensing

The term quorum sensing (QS) describes a social behaviour observed in bacteria whereby gene regulation is coordinated throughout a population (for a review see reference (190)). Individual cells accomplish QS by secreting a common small diffusible molecule termed an autoinducer that increases in concentration proportionally with population density. Above a threshold concentration the autoinducer is perceived by a receptor that then induces the transcriptional regulation of downstream genes. In most systems the transcriptional regulation includes the upregulation of the autoinducer synthase itself, resulting in further activation and coordination of gene regulation across the population.

One of the earliest discovered examples of QS is that which regulates expression of luminescence in Vibrio fischeri. Nealson et al. (133) observed that a short burst of luminescence occurred during growth in liquid culture as the culture reached high cell-density and more importantly, that a molecule present in the spent medium was able to induce luminescence in cells at low cell density. This molecule was subsequently shown to be an N-acyl homoserine lactone (AHL), N-3-oxohexanoyl homoserine lactone (3-oxo-C6-HSL), synthesised by the product of the luxI gene (48). 3-oxo-C6-HSL activates the LuxR protein which in turn activates transcription of the luxICDABEG luminescence (lux) operon. In addition to a wide range of AHL variants, several other autoinducer molecules have now been identified, such as the furanone compounds collectively termed autoinducer-2, PQS (2-heptyl-3-hydroxy-4-quinolone) of Pseudomonas aeruginosa and the various oligopeptide signals common to Gram-positive bacteria (for a review see reference (31)). QS systems are widespread in bacteria and are involved in the regulation of a plethora of phenotypes including virulence, motility, biofilm formation, siderophore production, secretion, exopolysaccharide production, rhizobium-legume symbiosis, DNA-competence, plasmid transfer and cell growth, just to name a few (for a review see reference (196)).

The concentration of the autoinducer is not always proportional to population density as it can be greatly modulated by environmental conditions that affect diffusion such as the spatial structure and distribution of cells. Some researchers have debated that alternative terms such as ‘diffusion sensing’ or ‘efficiency sensing’ better describe the general phenomenon, especially as the term quorum sensing implies an altruistic cooperative behaviour, which is not necessarily the case.
However for simplicity QS will be used to collectively describe QS, diffusion sensing and efficiency sensing throughout this thesis.

1.5.1 N-acyl homoserine lactone-dependent quorum sensing

Since the discovery of the LuxR system at least 55 additional LuxR and LuxI homologues have been identified (103). The AHL receptor and inducer protein homologues belong to two very distantly related phyla; the major family A includes LuxR and LuxI and its members are distributed throughout the Gram-negative bacteria while the B family members are restricted to the γ-proteobacteria (103).

The receptor proteins of both families are composed of an N-terminal AHL-binding domain and a C-terminal DNA-binding domain containing a HTH motif. In the presence of cognate AHL most LuxR homologues belonging to family A function as DNA-binding transcriptional activators through the binding of short DNA motifs (e.g. lux or tra-boxes) adjacent to promoter regions and interaction with RNAP (11, 117, 193). Binding of the cognate AHL to each monomer usually accompanies dimerisation of the receptor protein to form an active complex. These proteins can similarly act as transcriptional repressors by binding and blocking access of RNAP to a promoter. In contrast, several members of the family B AHL systems are able to dimerise and bind DNA in the absence of cognate AHL. For example the Serratia SmaR protein represses transcription in its dimerised form and transcription is derepressed in the presence of AHL (54). This system therefore achieves some functional equivalence with the activation accomplished by most family A members, albeit through the completely opposite mechanism (for a review see (173)).

It is evident that multiple HGT events have contributed to the wide distribution of family A members (69, 103). Many species have acquired multiple distinct LuxR-LuxI homologues forming hierarchical and/or overlapping regulatory networks that respond to multiple unique AHL molecules (40, 168). The various LuxI homologues produce many distinct AHL molecules with acyl chains ranging from 4-16 carbons (abbreviated as C4-C16) with various modifications, but most commonly substitutions at the third-carbon position (Figure 1-4).
There is much variation in both the range and type of AHLs produced by individual LuxI proteins. The substrates of LuxI proteins are S-adenosylmethionine and acylated acyl carrier protein (acyl-ACP) (90). The type of acyl-ACP incorporated is likely affected by the specificity of the AHL synthase and the fatty-acid pool. A study on the specificity of the *Pantoea stewartii* AHL synthase EsaI revealed that a single threonine restricts the enzyme to incorporate mostly 3-oxo-acyl-ACP derivatives, as mutation at this position allows the enzyme to produce a wider spectrum of AHLs with substituted and unsubstituted acyl chains (67). The same study showed that the *P. aeruginosa* LasI enzyme, while producing predominantly 3-oxo-C12-HSL in its native background, produced a wider spectrum of AHLs including some with acyl chains with an odd number of carbons when expressed in *E. coli*. These factors make the predictions about substrate specificity of LuxI proteins from sequence information tenuous.

There is also much variation in the response to particular AHLs by specific LuxR proteins. For example, in *A. tumefaciens* the TraR protein of the Ti plasmid is activated by 3-oxo-C8-HSL, the major product produced by the TraI protein (86). However the TraR protein is able to be activated by a much wider range of AHLs with acyl chains 6-12C in length including 3-alkanoyl derivatives (34). Similarly *Chromobacterium violaceum* strain CV026 produces the purple pigment violacein in response to AHLs with acyl chains 4-8C, but is inhibited by the presence of AHLs with chains 10-14C in length (118). Several useful biological reporter systems have been created utilising the response of LuxR proteins to exogenously supplied AHLs. These include *lux*, *gfp*, *lacZ* and violacein-based systems that can not only detect AHLs but also tentatively identify molecules when combined with thin-layer chromatography and appropriate molecular standards. However due to variation in sensitivity to specific AHLs these assays cannot replace a more thorough chemical identification (196).
1.5.2 Quorum sensing in rhizobia

The first rhizobial AHL was discovered in *Rhizobium leguminosarum* and initially named bacteriocin *small*, due to a growth inhibitory effect the compound had on specific *R. leguminosarum* strains. This compound was subsequently discovered to be N-(3-hydroxy-7-cis tetradecenoyl)-HSL (3-OH-7-cis-C14:1-HSL) produced by the chromosomally encoded CinI protein (109). Growth sensitivity was found to be conferred by two LuxR-type regulators, BisR and TraR, encoded by the plasmid pRL1JI. The exact mechanism by which these genes and other QS systems in rhizobial species inhibit growth remains unclear; however it has been reported to involve a chromosomal target (78) and likely involves disruption of protein synthesis (194). The role of QS in regulating pRL1JI transfer is discussed in Section 1.5.2.2.

AHL production in rhizobial species is now known to be very common, with at least one AHL-producing species described for most genera (for a review see (164)). A wide variety of AHLs are produced; for example a marine *Mesorhizobium* strain produces the longest AHL molecule yet identified, with an acyl chain of 16C as well as producing novel long chain AHLs with unsaturated acyl chains (98). Interestingly there does not appear to be much conservation in the types of AHLs produced by various strains, as these appear to be almost always distinct even when comparing closely related isolates (164). QS systems in rhizobia have been found to regulate a range of phenotypes including nitrogen fixation, nodulation efficiency, exopolysaccharide production, motility and, as previously mentioned, plasmid transfer. Plasmids that regulate conjugation by QS are found in *A. tumefaciens*, *R. leguminosarum* bv. *viciae*, *Rhizobium etli* strain CFN42, *Sinorhizobium meliloti* strain Rm41 and *Rhizobium* sp. Strain NGR234 (164).

1.5.2.1 Regulation of the *A. tumefaciens* Ti plasmids

Plant tumour-inducing strains such as *A. tumefaciens* C58 carry QS-regulated plasmids, called Ti plasmids. It is these Ti plasmids from which the T-DNA is processed and transferred to plants to induce the formation of crown gall tumours. The QS systems of the Ti plasmids are composed of the LuxR-type TraR protein and LuxI-type 3-oxo-C8-HSL synthase TraI. The expression of *traR* is activated by the presence of molecules called opines which are produced by genes encoded on the Ti plasmid that are only expressed in tumours. Therefore expression of the QS system and
conjugation of the Ti plasmid is likely restricted to this environment (for a review see reference (191)).

The TraR protein of pTiC58 binds 3-oxo-C8-HSL at a 1:1 ratio per subunit and the complex is a homodimer in the active form (187). Interestingly the TraR protein is inherently unstable in the absence of 3-oxo-C8-HSL and is quickly degraded by host proteases. It is likely that binding of 3-oxo-C8-HSL aids in the correct folding of TraR during synthesis or that the molecule causes a conformational change preventing degradation of TraR (202). The active TraR-3-oxo-C8-HSL dimers are able to bind 18-bp palindromic sequences called tra-boxes, found upstream of the traAFBH, traCDG and tral-trb operons that are required for conjugation and AHL synthesis, and the replication genes repABC. Binding of TraR-3-oxo-C8-HSL to the promoter either approximately 63 or 43 bp upstream of the transcriptional start site activates transcription, likely through a direct interaction with RNAP (192, 193).

The QS system of A. tumefaciens is inhibited by the antiactivator TraM, a protein which directly inhibits TraR through a protein-protein interaction. The TraM protein forms a homodimer which is able to bind either free TraR or TraR in complex with DNA, inhibiting its binding or stimulating its release from DNA respectively (36, 148). It has been proposed that the inhibition of the QS system by TraM prevents premature autoinduction of the system through stochastic variation in 3-oxo-C8-HSL concentration and other molecular noise (66). A second type of antiactivator has also been characterised, TrlR which resembles a pseudogene copy of TraR. TrlR likely inhibits the QS system through formation of inactive TraR-TrlR heterodimers (35).

1.5.2.2 Regulation of transfer of the R. leguminosarum plasmid pRL1JI

The Ti plasmids of A. tumefaciens and the tral-trb operon on plasmid pRL1JI of R. leguminosarum are regulated by a TraR protein which is activated by 3-oxo-C8-HSL, produced by TraI (117). The Ti plasmid the system is inhibited by the plasmid-encoded antiactivator TraM (40). Distinctly however, the regulation of plasmid transfer is intimately intertwined with the QS system of the host chromosome, namely the cinR-cinI locus. The CinR protein is activated by 3-OH-7-cis-C14:1-HSL produced by CinI. The pRL1JI plasmid encodes another LuxR-type regulator BisR, which also binds 3-OH-7-cis-C14:1-HSL and has likely evolved through an ancient gene duplication of CinR. In addition to activating the expression of TraR, BisR represses expression of CinI and production of 3-OH-7-cis-C14:1-HSL, possibly through the formation of
inactive CinR-BisR heterodimers. This mechanism effectively shuts off the chromosomal CinR-CinI QS system, but also allows cells to sense the presence of exogenous 3-OH-7-cis-C14:1-HSL through activation of BisR, which in turn activates TraR and conjugation. Therefore cells carrying pRL1J1 are able to sense the presence of cells that lack pRL1JL through their production of 3-OH-7-cis-C14:1-HSL and in response initiate transfer of pRL1J1 (40).

1.6 Integrative and conjugative elements (ICEs)

In 2002 Burrus et al. (25) suggested that a subset of mobile genetic elements including conjugative transposons, integrative plasmids and mobile genomic islands should be grouped under a new classification termed integrative and conjugative elements or ICEs. ICEs are defined as elements that excise from their host chromosome in a site-specific manner, leading to formation of a circularized element that is generally transient. Conjugative transfer initiating from the circularized form is followed by integration of the element into the recipient chromosome. Integration and excision are facilitated by an integrase or recombinase encoded on the element and a tRNA gene is often utilized as the integration target site (25). ICEs appear to be modular and usually consist of conserved blocks containing genes for recombination, conjugation and maintenance, interspersed with more variable genes specific to each element. Some of the early ICEs discovered include the SXT element of *Vibrio cholerae* which encodes resistance to several antibiotics (24, 83), the *clc* element of *Pseudomonas* strain B13 which contains genes required for chlorocatechol degradation (61, 155) and related islands in *Haemophilus influenzae* (125), and the subject of this study, the symbiosis island of *M. loti* R7A ICEML/SymR7A. The number of documented ICEs has expanded massively since the classification; the following sections summarise knowledge of a few of the better known elements relevant to this study.

1.6.1 SXT and R391

The SXT element and R391 are closely related ICEs ~100 kb in length belonging to a family with at least 25 identified members (24). Originally discovered in clinical isolates of *Vibrio cholerae* (83) and *Providencia rettgeri* (24) these ICEs carry genes conferring resistance to antibiotics. Members encode near-identical integrases that weakly resemble those of lambdoid phage and also encode members of a unique family of RDFs. The target of integration is the prfC gene and tandem arrays of SXT and R391 have been observed. The conjugation apparatus distantly
resembles that of the F plasmid. Interestingly integrated tandem arrays of SXT and R391 have been observed to recombine, producing hybrid elements (27).

Like the λ phage, SXT encodes a DNA-binding regulator of the Xre family that represses excision and HGT called SetR. Also like λ, this repressor is sensitive to the SOS response, as activated RecA stimulates cleavage of SetR relieving repression of SXT excision and transfer. As many antibiotics induce the SOS response, they also indirectly stimulate the dissemination of antibiotic resistance genes through induction of SXT (17).

1.6.2 clc and ICEHin1056

The 105-kb clc element was originally discovered in Pseudomonas sp. strain B13 as it conferred the ability to metabolise 3-chlorobenzoate (154). Transmission of clc has been observed between members of the β and γ-proteobacteria and related elements have been documented in these phyla. However sequence evidence suggests that the conjugation apparatus is only very distantly related to that of characterised T4SS (17). Integration of clc into a gly-tRNA gene is catalysed by an unusually long integrase IntB13 with limited similarity to the P4 family of integrases. The intB13 gene is located next to the right clc-chromosome junction attR and is under the control of two alternate promoters (155). A strong constitutive promoter $P_{circ}$ is formed on the excised circular form of clc and initiates from within the integrase attachment site $attP$, and a weaker, possibly regulated, promoter $P_{int}$ is present on the integrated element initiating from attR. The $P_{circ}$ promoter is encoded within the left clc-chromosome junction attL and is juxtaposed with intB13 upon formation of attP. It is believed that the formation of a strong promoter upon excision of this element and others drives integration of the element in recipient strains (170, 171).

ICEHin1056, clc and a number of putative ICEs form a distinct family that encode a unique T4SS designated GI-type. Only 7 of the 24 ORFs identified as likely to contribute to its function have homologues in other systems. The deep evolutionary conservation of this specific T4SS within this family of ICEs illustrates their long ancestry as unique mobile elements, rather than just as mosaics of plasmid and phage-related gene clusters (92).
1.6.3 Tn4371

Originally classed as a conjugative transposon, Tn4371 was discovered following a mating between *Ralstonia* sp. strain A5 and *Ralstonia oxalatica*, as it conferred the ability to degrade biphenyl and 4-chlorobiphenyl (174). DNA sequence of Tn4371 revealed it encoded a 55-kb element, flanked by a short direct repeat and encoding a putative integrase with little similarity to characterized integrases. Of relevance to this study, Tn4371 and related elements were shown to encode a P-like T4SS with identical gene structure to that of ICE*Ml*Sym MAFF (and ICE*Ml*Sym R7A), including homologues of the two hypothetical genes that flank the *trb* cluster, *msi031* and *msi021*. Homologues of the ICE*Ml*Sym R7A genes *msi110-msi106* and *traG* were also identified. Therefore it seems likely that ICE*Ml*Sym R7A and Tn4371 share a distant common ancestor and employ a similar mechanism of conjugative transfer (182, 185).

1.7 Aims of this study

Horizontal gene transfer is the major process by which bacteria evolve the ability to exploit new environmental niches. While plasmids and bacteriophage are well-characterised as conveyers of the horizontal gene pool, ICEs have only recently been recognised as abundant contributors to bacterial adaptation and evolution. Because of the importance of ICEs in microbial evolution, it is imperative to determine the mechanisms of their transfer and the factors that influence their transfer frequency in the environment.

The *M. loti* strain R7A symbiosis island ICE*Ml*Sym R7A integrates into the chromosome in a process mediated by an island-encoded integrase *intS* of the phage P4 family that involves recombination between 17-bp core sites found on the island and at the 3’ end of a *phe*-tRNA gene (180, 182). Current evidence (J. Sullivan and C. Ronson, unpublished data) has demonstrated that the expression of *intS* involves formation of a new *intS* promoter through ICE*Ml*Sym R7A circularisation, as occurs for the *clc* element. The first aim of this study is to characterise the promoter involved and the regions required for IntS-mediated integration.

P4-like integrases such as IntS catalyse both integration and excision of the replicon encoding them, but an excisionase has not yet been identified on ICE*Ml*Sym R7A and how recombination directionality is controlled is not known. The second aim of this study is to identify the factor(s) that favour excision over integration. A quantitative polymerase chain reaction (QPCR) assay will be developed to measure the proportion of cells containing an excised ICE*Ml*Sym R7A in a
cell population. This assay will be used to test the effect on excision of mutation or overexpression of candidate genes identified through bioinformatics as possibly involved in the recombination process.

Despite the importance of ICEs in evolution, the mechanisms that regulate their transfer in the environment are poorly understood. The third aim of this study is to uncover the mechanisms involved in the regulation of ICE\textit{Ml}/Sym\textsuperscript{R7A} excision and transfer. In particular the effects of known and potential regulatory genes identified through bioinformatics (e.g. \textit{traR}, \textit{msi169}) on excision will be investigated.
Chapter 2
Materials and Methods
2.1 Media

*E. coli* strains were cultured at 37°C on Luria-Bertani (LB) agar or in 2YT or TY broths (20, 121, 163). Mesorhizobial strains were grown at 28°C on plates or in broths of TY or rhizobium defined media (RDM) (158). RDM was supplemented either 0.4 % glucose (G/RDM) or 5% sucrose as the sole carbon source. Media were supplemented with antibiotics as required (Table 2-1). Non-symbiotic *M. loti* strains were grown on G/RDM media supplemented with 1mg. L⁻¹ nicotinate, 20 µg.L⁻¹ biotin, and 1 mg.L⁻¹ thiamine HCl. *Chromobacterium violaceum* was cultured at 28°C on LB or TY agar or in LB broth cultures.

<table>
<thead>
<tr>
<th>Antibiotic</th>
<th>Abbreviation</th>
<th>Concentration (µg.mL⁻¹)</th>
<th>E. coli</th>
<th>Mesorhizobium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampicillin</td>
<td>Ap</td>
<td>100</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tetracycline</td>
<td>Tc</td>
<td>15</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>Neomycin</td>
<td>Nm</td>
<td>-</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>Chloramphenicol</td>
<td>Cm</td>
<td>50</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Kanamycin</td>
<td>Km</td>
<td>50</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gentamicin</td>
<td>Gm</td>
<td>25</td>
<td>50</td>
<td>-</td>
</tr>
</tbody>
</table>

2.2 Bacterial strains and Plasmids

Bacterial strains and plasmids used in this study are listed in Table 2-2

<table>
<thead>
<tr>
<th>Strain/plasmid</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>R7A</td>
<td>Field reisolate of ICMP 3153; wild-type symbiotic strain</td>
<td>(179)</td>
</tr>
<tr>
<td>R7ANS</td>
<td>Non-symbiotic derivative of R7A; lacks ICE/H/Sym</td>
<td>This study, (153)</td>
</tr>
<tr>
<td>CJ4</td>
<td>Nonsymbiotic <em>Mesorhizobium</em> strain; field isolate</td>
<td>(178)</td>
</tr>
<tr>
<td>N18</td>
<td>Nonsymbiotic <em>Mesorhizobium</em> strain; field isolate</td>
<td>This laboratory, (153)</td>
</tr>
<tr>
<td>R7AΔintS</td>
<td>ΔintS::nptII, Nm²; gene replacement deletion mutant of intS</td>
<td>(153)</td>
</tr>
<tr>
<td>R7Amsi109</td>
<td>Δmsi109; markerless in-frame deletion mutant; mutant renamed as R7AΔrdfS when function of Msi109 shown</td>
<td>This study, (153)</td>
</tr>
<tr>
<td>R7Amsi106</td>
<td>pFUS2 insertion mutant of msi106; mutant renamed as R7ArlxS when function of Msi106 shown</td>
<td>(153)</td>
</tr>
<tr>
<td>R7Amsi169:ΩK</td>
<td>msi169 insertional mutant containing Kan² cassette flanked by Ω transcriptional terminators</td>
<td>This laboratory</td>
</tr>
<tr>
<td>R7ArlxS:JacZ</td>
<td>lxs (msi106) insertional duplication containing an integrated copy of pFUS2. Gm² Lac⁺</td>
<td>This laboratory</td>
</tr>
<tr>
<td>R7AΔtraR</td>
<td>traR (msi174) in-frame markerless deletion mutant</td>
<td>This study</td>
</tr>
<tr>
<td>R7AΔrdl2</td>
<td>rdl2 (msi173) markerless deletion mutant</td>
<td>This study</td>
</tr>
<tr>
<td>R7AΔmsi172</td>
<td>msi172 in-frame markerless deletion mutant</td>
<td>This study</td>
</tr>
<tr>
<td>R7AΔmsi171</td>
<td>msi171 in-frame markerless deletion mutant</td>
<td>This study</td>
</tr>
<tr>
<td>R7AΔmsi170</td>
<td>msi170 in-frame markerless deletion mutant</td>
<td>This study</td>
</tr>
<tr>
<td>R7AΔmsi169-170</td>
<td>msi169-msi170 markerless deletion mutant</td>
<td>This study</td>
</tr>
</tbody>
</table>
### Plasmids

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1224</td>
<td>pLAFR1 containing a translational fusion of intS (a 6.1-kb EcoRI fragment also containing melR, phoB, and attL) to lacZ. Unpublished, this laboratory</td>
</tr>
<tr>
<td>p1310</td>
<td>pLAFR1 containing an out-of-frame fusion of intS (derived from attP) to lacZ. Unpublished, this laboratory</td>
</tr>
<tr>
<td>p1311</td>
<td>pLAFR1 containing a translational fusion of intS (derived from attP) to lacZ. Unpublished, this laboratory</td>
</tr>
<tr>
<td>pBSL141</td>
<td>contains QGm cassette</td>
</tr>
<tr>
<td>pFAJ1700</td>
<td>Broad-host-range plasmid, oriT&lt;sup&gt;R&lt;/sup&gt; &lt;sup&gt;Km&lt;/sup&gt; - &lt;sup&gt;Tc&lt;/sup&gt;</td>
</tr>
<tr>
<td>pFAJ1708</td>
<td>pFAJ1700 containing nptII promoter</td>
</tr>
<tr>
<td>pFJX</td>
<td>pFAJ1700 containing the lacZ gene from pFUS2</td>
</tr>
<tr>
<td>pFJX10</td>
<td>pFJX containing a region of attP amplified from pFUS288 using primers IntSTranFusionBamHI and IntSscriptFusion10attp and cloned as a BamHI fragment. F or R denotes the orientation of the fragment relative to attP and the lacZ gene of pFJX.</td>
</tr>
<tr>
<td>pFJX122</td>
<td>pFJX containing a region of attP amplified from pFUS288 using primers IntSTranFusionBamHI and IntSscriptFusion122attp and cloned as a BamHI fragment. F or R denotes the orientation of the fragment relative to attP and the lacZ gene of pFJX.</td>
</tr>
<tr>
<td>pFJX122</td>
<td>pFJX containing a region of attP amplified from pFUS288 using primers IntSTranFusionBamHI and IntSscriptFusion122attp and cloned as a BamHI fragment.</td>
</tr>
<tr>
<td>pFJX288</td>
<td>pFJX containing a region of attP amplified from pFUS288 using primers IntSTranFusionBamHI and IntSscriptFusion288attp and cloned as a BamHI fragment. F or R denotes the orientation of the fragment relative to attP and the lacZ gene of pFJX.</td>
</tr>
<tr>
<td>pFJX53</td>
<td>pFJX containing a region of attP amplified from pFUS288 using primers IntSTranFusionBamHI and IntSscriptFusion53attp and cloned as a BamHI fragment. F or R denotes the orientation of the fragment relative to attP and the lacZ gene of pFJX.</td>
</tr>
<tr>
<td>pFJXattR</td>
<td>pFJX containing a region of attP amplified from pFUS288 using primers IntSTranFusionBamHI and IntSscriptFusionattR and cloned as a BamHI fragment. F or R denotes the orientation of the fragment relative to attP and the lacZ gene of pFJX.</td>
</tr>
<tr>
<td>pFUS2</td>
<td>pFUS2 containing attP amplified from pJJ609 using primers IntSTranFusionBamHI and IntSscriptFusion288attp. This study, (159)</td>
</tr>
<tr>
<td>pFUS288</td>
<td>pFUS2 containing attP amplified from pFUS288 using primers attP-intS and IntSscriptFusion288attp. F or R denotes the orientation of the fragment relative to attP and the lacZ gene of pFJX.</td>
</tr>
<tr>
<td>pFUSP1P4</td>
<td>pFUS2 containing attP amplified from pFUS288 using primers attP-intS and IntSscriptFusion288attp. F or R denotes the orientation of the fragment relative to attP and the lacZ gene of pFJX.</td>
</tr>
<tr>
<td>pFUSP1P5</td>
<td>pFUS2 containing attP amplified from pFUS288 using primers attP-intS and IntSscriptFusion288attp. F or R denotes the orientation of the fragment relative to attP and the lacZ gene of pFJX.</td>
</tr>
<tr>
<td>pFUSP2P4</td>
<td>pFUS2 containing attP amplified from pFUS288 using primers attP-intS and IntSscriptFusion288attp. F or R denotes the orientation of the fragment relative to attP and the lacZ gene of pFJX.</td>
</tr>
<tr>
<td>pFUSP2P5</td>
<td>pFUS2 containing attP amplified from pFUS288 using primers attP-intS and IntSscriptFusion288attp. F or R denotes the orientation of the fragment relative to attP and the lacZ gene of pFJX.</td>
</tr>
<tr>
<td>pJJ609</td>
<td>pFUS2 containing attP-intS BamHI fragment from pJJ608. This study, (159)</td>
</tr>
<tr>
<td>pJJ611</td>
<td>pFAJ1708 containing intS amplified using primers IntCN5 and IntCN3 and cloned as BamHI-EcoRI fragment downstream of nptII promoter. This study, (159)</td>
</tr>
<tr>
<td>pJJ611G</td>
<td>Contains ΩGm cassette from pBSL141 cloned as a blunt fragment into an EcoRV site. This study, (159)</td>
</tr>
<tr>
<td>Vector/Plasmid</td>
<td>Description</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------</td>
</tr>
<tr>
<td>pJK301</td>
<td>pJJ611 containing two PCR products which flank <em>msi109</em> (rdfS) amplified using primer pairs 109LL, 109LR and 109RR</td>
</tr>
<tr>
<td>pJQ200SK</td>
<td>Suicide vector containing sacB gene, Gm&lt;sup&gt;k&lt;/sup&gt;</td>
</tr>
<tr>
<td>pJR039</td>
<td>pFAJ700 containing <em>traI</em> (XbaI fragment), amplified using primers <em>tra1clone5</em> and <em>tra1clone3</em> This study</td>
</tr>
<tr>
<td>pJR169</td>
<td>pFAJ700 containing <em>msi10</em> amplified using primers 169170proS and 169170cloneN, cloned as a BamHI fragment This study</td>
</tr>
<tr>
<td>pJR169170</td>
<td>pFAJ700 containing <em>msi169-msi170</em> amplified using primers 1693clone and 1703clone, cloned as a BamHI fragment This study</td>
</tr>
<tr>
<td>pJR170</td>
<td>pFAJ700 containing <em>msi170</em> amplified using primers 169170proS and 1703clone, cloned as a BamHI fragment This study</td>
</tr>
<tr>
<td>pJR173</td>
<td>pFAJ700 containing the overlap-extensionPCR product used to create pJR174 and the R7A&lt;sub&gt;traR&lt;/sub&gt; cloned as an XbaI fragment. This region contains tra2 and upstream DNA of tra2 and traR</td>
</tr>
<tr>
<td>pJR174</td>
<td>pFAJ700 containing <em>traR</em> (BamHI fragment) amplified by PCR from pJR206 using primers <em>traRIL</em> and <em>traRIR</em> This study</td>
</tr>
<tr>
<td>pJR201</td>
<td>pUC8 containing PCR products spanning <em>attB</em>, <em>attB</em> and a region of <em>melR</em> amplified using primer pairs RE2/LE1, LE2/RE1, and ML/MR respectively This study, (159)</td>
</tr>
<tr>
<td>pJR202</td>
<td>pFAJ700 containing <em>msi109</em> (rdfS) and promoter region amplified using primers 109CN5 and 109CN3 This study, (159)</td>
</tr>
<tr>
<td>pJR204</td>
<td>pFAJ700 containing <em>msi109</em> amplified using primers 109CN5 and 109CN3 This study, (159)</td>
</tr>
<tr>
<td>pJR206</td>
<td>pFAJ700 containing <em>msi109</em> amplified using primers <em>traRIL</em> and <em>traRIR</em> This study, (159)</td>
</tr>
<tr>
<td>pJRNP170</td>
<td>pJRNP2 containing the <em>msi170</em> gene amplified using <em>msi170ATG</em> and <em>msi1703clone</em>, cloned as a BamHI fragment This study</td>
</tr>
<tr>
<td>pJRNP2T2</td>
<td>pPROBE-KT containing nptII promoter amplified from pFAJ708 using primers nptII5 and nptII3 cloned as a HindIII fragment This study</td>
</tr>
<tr>
<td>pJRΔ039</td>
<td>pJQ200SK containing a deleted <em>traI</em> allele and ~2 kb of surrounding DNA (as an XbaI fragment) created using overlap-extensionPCR. This study</td>
</tr>
<tr>
<td>pJRΔ170</td>
<td>pJQ200SK containing a deleted <em>msi170</em> allele and ~2 kb of surrounding DNA (as an XbaI fragment) created using overlap-extensionPCR. This study</td>
</tr>
<tr>
<td>pJRΔ171</td>
<td>pJQ200SK containing a deleted <em>msi171</em> allele and ~2 kb of surrounding DNA (as an XbaI fragment) created using overlap-extensionPCR. This study</td>
</tr>
<tr>
<td>pJRΔ172</td>
<td>pJQ200SK containing a deleted <em>msi172</em> allele and ~2 kb of surrounding DNA (as an XbaI fragment) created using overlap-extensionPCR. This study</td>
</tr>
<tr>
<td>pJRΔ173</td>
<td>pJQ200SK containing a deleted <em>tra2</em> allele and ~2 kb of surrounding DNA (as an XbaI fragment) created using overlap-extensionPCR. This study</td>
</tr>
<tr>
<td>pJRΔ174</td>
<td>pJQ200SK containing a deleted <em>traR</em> allele and ~2 kb of surrounding DNA (as an XbaI fragment) created using overlap-extensionPCR. This study</td>
</tr>
<tr>
<td>pJRΔ174070</td>
<td>pJQ200SK containing 2 kb of DNA surrounding a double deletion of <em>msi169</em> and <em>msi170</em> (as an XbaI fragment) created using overlap-extensionPCR. This study</td>
</tr>
<tr>
<td>pLAPR1</td>
<td>Broad host range plasmid, oriV&lt;sub&gt;RK2&lt;/sub&gt;, Tc&lt;sup&gt;+&lt;/sup&gt;</td>
</tr>
<tr>
<td>pPR691</td>
<td>Low-copy-number cosmid derived from pSC101, Km&lt;sup&gt;+&lt;/sup&gt;, Sp&lt;sup&gt;+&lt;/sup&gt;, Sm&lt;sup&gt;+&lt;/sup&gt;</td>
</tr>
<tr>
<td>pROBE-KT</td>
<td>Km&lt;sup&gt;+&lt;/sup&gt;, pVS1/p15a replicon, contains gfp</td>
</tr>
<tr>
<td>pSK170</td>
<td>pSK&lt;sub&gt;67&lt;/sub&gt; containing PCR product encoding Msi170-(6His) fusion product This study</td>
</tr>
<tr>
<td>pSK67</td>
<td>Protein expression vector. Ap&lt;sup&gt;+&lt;/sup&gt;</td>
</tr>
<tr>
<td>pSKTraR</td>
<td>pSK&lt;sub&gt;67&lt;/sub&gt; containing PCR product encoding TraR-(6His) fusion product This study</td>
</tr>
</tbody>
</table>

*Stephan Heeb, personal communication*
2.3 Storage of Strains

Strains were grown to stationary phase in G/RDM (*Mesorhizobium* spp.) or LB (*E. coli*). Aliquots (800 µL) of the cultures were mixed with 70 µL of dimethylsulfoxide (DMSO) or 15% glycerol in Nalgene cryogenic vials and the vials were stored at -70°C.

2.4 Enzymes and Chemicals

Enzymes were purchased from Roche Molecular Biochemicals and New England Biolabs unless otherwise specified. Antibiotics were purchased from Sigma, dissolved to an appropriate concentration in Milli-Q water and filter-sterilised through a 0.45 µm syringe filter, except for tetracycline and chloramphenicol which were dissolved in 100% methanol and 100% ethanol respectively. All chemicals were analytical grade and solutions were prepared in water unless otherwise indicated. Recipes for solutions and buffers are given in Appendix B.

2.5 Spectroscopy

Optical densities of cultures and absorbance readings of assays solutions were determined using either a Cary UV 50 spectrophotometer or a Jenway 6300 spectrophotometer. Absorbance readings and concentrations of nucleic acids were determined using a Nanodrop ND-1000.

2.6 DNA Manipulations

2.6.1 Electrophoresis

Polymerase chain reaction (PCR) products, restriction enzyme-digested DNA or uncut plasmid DNA were mixed with bromophenol blue tracking dye (Appendix A) and separated on 0.8-2% (w/v) agarose gels made and run in 1× Tris-acetate (TAE) buffer containing 1 µg.mL⁻¹ ethidium bromide. Electrophoresis of DNA was carried out using 1.0-1.5% SeaKem LE agarose (Cambrex Bio Science) gels for 1-2 h at 3.75 V.cm⁻¹ or overnight at 1.6 V.cm⁻¹. Following electrophoresis, DNA was visualised using a UV transilluminator. Gel images were recorded using the Biorad Chemidoc Gel Documentation system. Five µL of Finnzyme’s ready-to-use DNA size standards (F-303SD) was routinely run alongside other samples.
2.6.2 DNA Isolation

2.6.2.1 Alkaline lysis plasmid extractions

*E. coli* cultures were grown overnight in 5-mL LB broths at 37°C with shaking at 200 rpm, after which the cell mass was harvested by centrifugation. The cell pellets were resuspended in 250 µL Buffer P1 (Qiagen; 50 mM Tris pH 8.0, 10 mM EDTA pH 8.0, 100 µg.mL⁻¹ RNAse A). Lysis solution (300 µL of 0.2 M NaOH, 1% w/v SDS) was added to each tube and mixed by inversion. The tubes were incubated at room temperature for 4 min, before 350 µL of neutralisation solution (3 M potassium acetate, 5% w/v formic acid) was added and mixed by inversion. The tubes were then centrifuged at 15110 g for 5 min. To precipitate the plasmid DNA, 800 µL of the resulting supernatant was added to 560 µL (0.7 volumes) of isopropanol, mixed by pipetting, and centrifuged for 5 min at 15110 g. The pelleted DNA was washed with 70% ethanol, air-dried, and resuspended in 50 µL of filter-sterile Milli-Q water.

2.6.2.2 Prepman Ultra PCR-template Preparations

For rapid preparation of DNA for PCR or QPCR, cells from 200 µL – 1000 µL of *M. loti* culture were harvested by centrifugation at 15110 g and resuspended in 40 µL- 200 µL of Prepman Ultra sample reagent, except for preparation of DNA from root nodules where a single nodule was crushed with a sterile pestle and suspended in 200 µL of Prepman Ultra sample reagent. Samples were then placed in a boiling water bath for 10 min and then centrifuged at 15110 g for 8 min. The liquid fraction was then transferred to a new tube. Samples prepared from nodules were ethanol-precipitated and suspended in 50 µL of TE buffer.

2.6.2.3 Commercial plasmid purification kits

Plasmid DNA used for enzymatic digestion and cloning or as a template for DNA sequencing was isolated from *E. coli* cultures using commercial kits according to the manufacturer’s instructions. Plasmid DNA derived from plasmids with a high copy number was prepared using the QIAprep Spin Miniprep kit (Qiagen), while DNA of plasmids with a low copy number was prepared using a Qiagen Plasmid Midi Kit.
2.6.3 DNA fragment and PCR product purification

2.6.3.1 Commercial DNA fragment purification kits

PCR products or small (<5 kb) DNA fragments extracted from agarose gels were purified using a High Pure PCR Product Purification Kit (Roche), with a 50 μL elution volume. Large DNA fragments (>5 kb) were purified after electrophoresis using a Qiaquick Gel Extraction Kit (Qiagen) as per the manufacturer’s instructions.

2.6.3.2 Ethanol precipitation

DNA samples were precipitated by adding 2.5 volumes of ethanol and 0.1 volumes of sodium acetate (3 M, pH 5.2). The tubes were mixed by inversion and centrifuged at 15110 g for 10 min. The pelleted DNA was then washed with 70% ethanol, air-dried, and resuspended in Milli-Q. To aid precipitation of ligation mixtures, 1 μL of co-precipitant, either SeeDNA (GE Healthcare) or Pellet Paint (Merck Biosciences) was added and mixed with samples prior to precipitation.

2.6.3.3 Restriction endonuclease digestion

Plasmid DNA preparations were digested with restriction enzymes using the appropriate buffer, incubation temperature and duration suggested by the manufacturer. Digested vector DNA for ligations was dephosphorylated by the addition of 1 μl of calf intestinal alkaline phosphatase (Roche) to the restriction digest reaction mix, followed by an additional incubation at 37°C for 45 min.

2.7 Polymerase Chain Reaction (PCR)

In this study PCR was used to amplify DNA for cloning, sequencing, screening of plasmids clones and screening of genomic DNA for deleted sequences. PCR was also used to create gene alleles containing in-frame deletions (overlap-extension PCR) for use in a two-step gene replacement strategy to creating genomic mutations. Real-time quantitative PCR (QPCR) was also used to quantify relative abundance of DNA templates in an assay for ICE MlsymR7A excision. Primers used for PCR, QPCR, overlap-extension PCR, and DNA sequencing are listed in Table 2-3.
<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence 5 to 3</th>
<th>use</th>
</tr>
</thead>
<tbody>
<tr>
<td>LE1</td>
<td>GAGCCCGCCACCGTCAGACA</td>
<td>PCR</td>
</tr>
<tr>
<td>LE2</td>
<td>GTGCTATAACCACCGCCTTT</td>
<td>PCR</td>
</tr>
<tr>
<td>RE1</td>
<td>GGTGCAGGTATGCCCCGCAGTCGT</td>
<td>PCR</td>
</tr>
<tr>
<td>RE2</td>
<td>CTCCCTTTTGGTGTCAAGGCGTTCG</td>
<td>PCR</td>
</tr>
<tr>
<td>ML</td>
<td>TAGACAAGCTTCAGATAGACGACGACAAAT</td>
<td>PCR</td>
</tr>
<tr>
<td>MR</td>
<td>ATTAGGATCCACGCTCTTGCAGAAAATCGGA</td>
<td>PCR</td>
</tr>
<tr>
<td>IntSTranFusionBamHI</td>
<td>AGTAATGATCTCTTGCGACTTACGTCCTGT</td>
<td>PCR</td>
</tr>
<tr>
<td>IntSscriptFusion288attp</td>
<td>TGGCTGGAT CCCATTCAATAGACGACTCGTT</td>
<td>PCR</td>
</tr>
<tr>
<td>IntSscriptFusion122attp</td>
<td>TGGTAAGAATCCAGGCGGTTTGGGAAACCA</td>
<td>PCR</td>
</tr>
<tr>
<td>IntSscriptFusion53attp</td>
<td>TGGCAAGAGCCACAGGTGAGGGCGGCAAACA</td>
<td>PCR</td>
</tr>
<tr>
<td>IntSscriptFusion10attp</td>
<td>TGAAGAATCCAGGCGGATCGTTGGGCAAAGAT</td>
<td>PCR</td>
</tr>
<tr>
<td>IntSscriptFusionattR</td>
<td>ATATGGGATCCCTTGTTGGCAGCCGAGGACGAT</td>
<td>PCR</td>
</tr>
<tr>
<td>Int5'UTR+177bp</td>
<td>ATATGGATCCACGATCTCCTAATACCACCCA</td>
<td>PCR</td>
</tr>
<tr>
<td>Int5'UTR+71bp</td>
<td>ATATGGATCCACGATCTCCTAATACCACCCA</td>
<td>PCR</td>
</tr>
<tr>
<td>Int5'UTR+10bp</td>
<td>ATATGGATCCACGATCTCCTAATACCACCCA</td>
<td>PCR</td>
</tr>
<tr>
<td>attPrepeat(-82)</td>
<td>ATATGGGATCCCTTGGGACGACGACTCGGCA</td>
<td>PCR</td>
</tr>
<tr>
<td>attPrepeat(-69)</td>
<td>ATATGGGATCCCTTGGGACGACGACTCGGCA</td>
<td>PCR</td>
</tr>
<tr>
<td>attPrepeat(+187)</td>
<td>ATATGGGATCCCTTGGGACGACGACTCGGCA</td>
<td>PCR</td>
</tr>
<tr>
<td>intS5'UTR+177bp</td>
<td>ATATGGGATCCACGATCTCCTAATACCACCCA</td>
<td>PCR</td>
</tr>
<tr>
<td>intS5'UTR+71bp</td>
<td>ATATGGGATCCACGATCTCCTAATACCACCCA</td>
<td>PCR</td>
</tr>
<tr>
<td>intS5'UTR+10bp</td>
<td>ATATGGGATCCACGATCTCCTAATACCACCCA</td>
<td>PCR</td>
</tr>
<tr>
<td>attPrepeat(-82)</td>
<td>ATATGGGATCCCTTGGGACGACGACTCGGCA</td>
<td>PCR</td>
</tr>
<tr>
<td>attPrepeat(-69)</td>
<td>ATATGGGATCCCTTGGGACGACGACTCGGCA</td>
<td>PCR</td>
</tr>
<tr>
<td>attPrepeat(+187)</td>
<td>ATATGGGATCCCTTGGGACGACGACTCGGCA</td>
<td>PCR</td>
</tr>
<tr>
<td>109RR</td>
<td>ATATTTACTAGTTCCAGCTTGTAAGGCGGCAA</td>
<td>PCR</td>
</tr>
<tr>
<td>109RL</td>
<td>AAATTTCTGCAAGAACC CGGCTCGACGAT</td>
<td>PCR</td>
</tr>
<tr>
<td>109LL</td>
<td>TTAAAACCTGCAGCCGCTCCGGCAACACACATC</td>
<td>PCR</td>
</tr>
<tr>
<td>109LR</td>
<td>TTAAAACCTGCAGCCGCTCCGGCAACACACATC</td>
<td>PCR</td>
</tr>
<tr>
<td>110L</td>
<td>AAATCTGGTAGCCAGCGACACGACACCTA</td>
<td>PCR</td>
</tr>
<tr>
<td>109CN5</td>
<td>GAATACCTGACGACTAGCTAAGGACCGGATCAGGACG</td>
<td>PCR</td>
</tr>
<tr>
<td>109CN3</td>
<td>AATACCTGACGACTAGCTAAGGACCGGATCAGGACG</td>
<td>PCR</td>
</tr>
<tr>
<td>CIB5</td>
<td>AAATTTGGATCCCAACGCTT GCAGCTCCCA</td>
<td>PCR</td>
</tr>
<tr>
<td>CIB3</td>
<td>TTAAAAGGATCCCGGATTCGAGTACGAC</td>
<td>PCR</td>
</tr>
<tr>
<td>INTCN5</td>
<td>AAATTTGGATCCCTAGATGAATAGGAGATGTGATGTTACTGCT</td>
<td>PCR</td>
</tr>
<tr>
<td>INTCN3</td>
<td>AAATTTGGATCCCTAGATGAATAGGAGATGTGATGTTACTGCT</td>
<td>PCR</td>
</tr>
<tr>
<td>TAQATT</td>
<td>TCCGCTCCTGGGCA (Taqman probe)</td>
<td>QPCR</td>
</tr>
<tr>
<td>TAQMR</td>
<td>TTGACGGCATGCTC (Taqman probe)</td>
<td>QPCR</td>
</tr>
<tr>
<td>attBL</td>
<td>GCCTGTATGACGGTGGTTCGA</td>
<td>PCR</td>
</tr>
<tr>
<td>attBR</td>
<td>GCAGCATGAAATCCGGGTATGA</td>
<td>PCR</td>
</tr>
<tr>
<td>attPL</td>
<td>GACATGTTGGAGGCAACAC</td>
<td>PCR</td>
</tr>
<tr>
<td>attPR</td>
<td>CGAAAGAGACTGTCGGGAGAA</td>
<td>PCR</td>
</tr>
<tr>
<td>melRL</td>
<td>CCAAACCGACGACACATTC</td>
<td>PCR</td>
</tr>
<tr>
<td>melRR</td>
<td>AATCCGACAACGACAAATTC</td>
<td>PCR</td>
</tr>
</tbody>
</table>
2.7.1 Standard PCR protocol

For amplification of DNA from plasmid or genomic DNA the Phusion High-Fidelity PCR system (Finnzymes) was used. When resulting PCR products were to be used for subsequent cloning or sequencing, 50 µL reaction volumes were used; otherwise a total reaction volume of 20 µL was used. Reaction mixtures contained 20% v/v Phusion High-Fidelity buffer, 10 nM of dNTP mixture (supplied with kit), 200 nM of each primer, 0.5 units of Phusion High-Fidelity Polymerase and 1-5 µL of DNA template suspension. Milli-Q was added to bring the mixture up to the desired volume. For preparation of multiple PCR reactions, cocktail mixtures were made whenever possible to ensure consistency. Thermal cycling was performed using a Hybaid PCR express thermal cycler. PCR cycling conditions were as follows: one cycle of 98°C for 1 min, thirty cycles of 98°C (10 s), 57-65°C (10-30 s), 72°C (15 s per 1 kb product) and a final cycle of 98°C (10 s), 57-65°C (10-30 s), 72°C (5 min).

2.7.2 Overlap-extension PCR

Overlap-extension PCR was used to generate gene alleles containing an in-frame markerless deletion. The desired deletion was first designed in silico, after which two complementary oligonucleotides (30-40 bp) were designed to evenly span the deletion junction (overlap primers). Each overlap primer was used in a separate PCR reaction with an additional primer positioned 0.9-1.1 kb either side of the deletion junction (flanking primer). The resulting pair of PCR products each contained 0.9-1.1 kb of DNA flanking the desired deletion and an identical 30-40-bp region of DNA spanning the deletion junction. The principle of this method is that the complementary overlap region on each PCR product is able to act as a primer on the other PCR product in a second PCR reaction – ultimately resulting in the extension of the overlap region and synthesis of a new ~2-kb molecule composed of both PCR products (81). For the overlap-extension, 5 µL of each purified PCR product was used in a PCR reaction combined with the two flanking primers, in order to specifically amplify the ~2-kb product created by the overlap extension. The ~2-kb product was then identified by agarose gel electrophoresis and purified.
from the gel for subsequent cloning. Overlap extension PCR primers used to make mutants in this study are listed in Table 2-4 and their sequences are given in Table 2-3.

Table 2-4 Primers used to construct mutant alleles using overlap-extension PCR

<table>
<thead>
<tr>
<th>Mutant alleles</th>
<th>5' flanking primer</th>
<th>Overlap primers</th>
<th>3' flanking primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>traR</td>
<td>traR5F</td>
<td>traR5R</td>
<td>traR3F</td>
</tr>
<tr>
<td>traI2</td>
<td>traI25F</td>
<td>traI25R</td>
<td>traI23F</td>
</tr>
<tr>
<td>msi172</td>
<td>msi1725F</td>
<td>msi1725R</td>
<td>msi1723F</td>
</tr>
<tr>
<td>msi171</td>
<td>msi1715F</td>
<td>msi1715R</td>
<td>msi1713F</td>
</tr>
<tr>
<td>msi170</td>
<td>msi1705F</td>
<td>msi1705R</td>
<td>msi1703F</td>
</tr>
<tr>
<td>msi170-msi169</td>
<td>msi1705F</td>
<td>msi1691705R</td>
<td>msi1691703F</td>
</tr>
<tr>
<td>traI1</td>
<td>traI15F</td>
<td>traI15R</td>
<td>traI13F</td>
</tr>
</tbody>
</table>

2.7.3 Quantitative PCR (QPCR)

In order to detect relative amounts of specific DNA species in *M. loti* genomic DNA preparations, Taqman-probe-based QPCR was used. Primer Express software V. 2.0 (Applied Biosystems) was used to design primers to detect *attP* (attPL, attPR), *attB* (attBL, attBR) and *melR* (melRL, melRR) and to design FAM-labeled minor-groove-binding (MGB) probes for *attP/attB* (TAQATT) and *melR* (TAQMR). The same probe was used for detection of both *attP* and *attB*, as it lies within the 3'-terminal 17-bp of the phetRNA gene that is present within both attachment sites. The Applied Biosystems 7500 Fast System was used for real-time fluorescence detection of PCR products and results were analyzed with Applied Biosystems 7500 Fast System SDS software V 1.3. Reactions were carried out in 10-µL volumes containing 5 µL of TaqMan Fast Universal PCR Master Mix (Applied Biosystems), 900 nM of each primer, 250 nM of probe and 1 µL of template (either Prepman Ultra preparations or Qiagen-mini kit purified DNA). Cycling conditions were 20 s at 95°C then forty cycles of 3 s at 95°C and 30 s at 60°C.

To reduce possible error introduced through pipetting small volumes, cocktails of reaction components were made to minimise the number of pipetting steps. For each component in each cocktail, 10% additional volume was added to account for liquid lost on transfer from plasticware. Each cocktail was mixed thoroughly by vortexing for 5 s. The primer/probe cocktail contained 6 µM of each of the two primers and 1.67 µM of the probe. The template cocktail contained 54 µL of TaqMan Fast Universal PCR Master Mix (Applied Biosystems), 10.8 µL of the template preparation and 27 µL TE buffer. The final reaction mixture contained 28 µL of the template cocktail and 5 µL of primer/probe cocktail, which was used for three 10-µL
QPCR reactions. All reaction plates included both positive (pJR201 DNA linearised with EcoRI) and negative template controls, also in triplicate.

The amplification efficiency of each assay was determined using the plasmid pJR201 (linearised with EcoRI), which contains single copies of attP, attB, and a region of melR. Serial doubling dilutions were prepared and used as templates for QPCR to generate standard curves for each of the three PCR reactions, by plotting relative DNA concentration versus log(Ct) value (Ct is the PCR cycle at which fluorescence rises beyond background levels). Amplification efficiency was determined using Equation 2-1. Relative data from the attB or attP PCR reactions were calculated and normalised to parallel data from the melR reaction using Equation 2-2. To test for significant differences between the strains at each data point, the Student T-test was employed with the assumption of equal variance across groups. Prior to applying the test, the data were log transformed (log base 10) to reduce the positive association between the mean and the variance.

**Equation 2-1 Reaction-efficiency (E) of the QPCR reaction**

\[
E = 10^{-\left(\frac{1}{slope}\right)}
\]

Slope value is derived from standard curves generated from a template dilution series.

**Equation 2-2 Relative quantification and normalisation of the QPCR assay**

\[
R_{att} = \frac{(E_{att})^{\Delta C_t(att)}}{(E_{melR})^{\Delta C_t(melR)}}
\]

\(E_{att}\) represents the efficiency of either the attB or attP PCR reaction, \(E_{melR}\) is the efficiency of the melR PCR reaction, \(\Delta C_t(att)\) represents the difference between the Ct value of the reference att PCR and the test att PCR and \(\Delta C_t(melR)\) represents the difference between the Ct value of the reference melR PCR and the test melR PCR (26, 144, 153)
2.8 AHL analysis

2.8.1 Extraction
Liquid R7A cultures were grown in either 50 mL or 200 mL of G/RDM. Three replicate 8 mL or 50mL aliquots of filtered (0.25µm Millipore) supernatant from each stationary-phase (64 h) culture were extracted twice with one volume of dichloromethane (118). Extracts were then evaporated to dryness in a vacuum centrifuge and resuspended in 50 µL methanol.

2.8.2 AHL Detection by *C. violaceum* CV026

2.8.2.1 CV026 overlays
A 100 mL LB broth was inoculated with CV026 and incubated overnight at 28°C. One hundred mL of molten LB agar was then cooled to ~40°C, mixed with the CV026 culture and quickly applied to the TLC or agar plate using a 10 ml pipette.

2.8.2.2 Plate assays
For routine assays, a loopful of *M. loti* culture was taken from G/RDM plate and spread on half of a TY agar plate and incubated for 48 h at 28°C. A loopful of *C. violaceum* CV026 from an overnight LB agar plate culture was then streaked adjacent to the *M. loti* culture to cover the other half of the TY plate. The resulting plate was then incubated overnight at 28°C.

As an alternative method for presentation, *M. loti* TY broth cultures were grown for 64 h at 28°C, after which 10-20 µL of culture was spotted onto a large (20 cm diameter) TY agar plate and incubated for either 24 h or 48 h. CV026 culture was then either streaked adjacent to the *M. loti* culture or an LB agar overlay was used to cover the culture.

2.8.2.3 Thin-layer chromatography
Reverse-phase thin-layer chromatography was used to separate AHL species extracted from *M. loti* samples or synthetic extracts according reference to (118). Two-µL of AHL extracts (equivalent to ~10 mL of culture supernatant) or 20 mM of synthetic C4-HSL dissolved in acetonitrile were applied 2 cm above the bottom of a Merck RP-18 F254s (20 x 20 cm) plate and allowed to dry. The bottom 1 cm of the plate was then immersed in 70% methanol/water (v/v) in a sealed glass enclosure and left until the methanol/water mix had migrated within 1 cm of the top of the plate by capillary action. The TLC plate was then allowed to dry before applying a CV026 overlay.
2.8.3 LC-MS/MS method

Liquid chromatography and mass-spectrometry of AHL samples was performed by Catherine Ortori. Liquid chromatography was carried out on a Shimadzu series 10AD VP using a Phenomenex Gemini C18, 150x2 mm (5-µm particle size) column at 45°C. Mass spectrometry was conducted using a 4000 QTRAP hybrid triple-quadrupole–linear ion trap mass spectrometer (Applied Biosystems), with a TurboIon ion source. For detailed methods and instrument settings see reference (137).

The information-dependent acquisition feature of the mass spectrometer was used to detect the presence of precursor ions with an \( m/z \) of 102.5 (common HSL moiety) and trigger capture by a linear ion trap. Detailed spectral analysis was then carried out and molecules were identified through comparison to spectra generated from synthetic AHL standards (kindly donated by Paul Williams). Relative quantification was performed by normalising the sample’s peak areas from extracted ion chromatograms (XIC) (obtained in the initial precursor ion scans) to XIC peak areas from synthetic AHL standards, both adjusted for respective molecular weight values. For detailed methods and instrument settings see reference (137).

2.9 Protein induction, purification and analysis

Primers containing appropriate restriction enzyme target sequences and a 3' 6xHis tag were used to amplify genes encoding proteins of interest. The PCR products were then cloned into plasmid pSK67, which contains an IPTG inducible \( \text{tac} \) promoter and an \( E. \) coli consensus ribosome binding site for overexpression. Clones were then introduced into \( E. \) coli strain BL21-DE3(pLys). LB broths (100 mL) were then inoculated with 1 mL of overnight culture containing the expression constructs and grown at 37°C with shaking. At 3 h 1 mM IPTG was added to each culture to induce expression of proteins from pSK67. 1 mL samples were taken each hour from 0-4 h and one further sample was taken after incubation overnight and stored at -80°C prior to analysis by SDS-PAGE. Cells from the remaining overnight culture were then pelleted by centrifugation and stored at -80°C prior to Nickel-NTA column purification.

2.9.1 SDS-PAGE

Cells were recovered by centrifugation at 15110 g for 10 min and then resuspended in 100 mM Tris HCl pH 7.0 to produce final optical density equivalent to OD\text{\textsubscript{600}} = 15. Four µL of SDS
loading dye (50 mM Tris-HCl pH 6.8, 2% SDS w/v, 2% \(\beta\)-mercaptoethanol, 12% glycerol w/v, 0.01% bromophenol blue) was then added to 20 µL of each sample and then boiled for 10 min. Samples were then centrifuged briefly at 15110 \(g\) and placed on ice prior to analysis by SDS-PAGE.

SDS-PAGE of prepared samples was carried out using the buffer system of Laemmli. 16% acrylamide gels were run in SDS running buffer (25 mM Tris-HCl pH 8.3, 192 mM Glycine, 0.1% w/v SDS) at 125 V for ~1 h after which gels were removed and stained with Coomassie brilliant blue staining solution (0.025% Coomassie Brilliant blue R 250, 40% methanol and 7% acetic acid ) for 15-30 min. Gels were then destained overnight in destaining solution (7% v/v acetic acid, 5% v/v methanol) on a platform shaker until background staining was removed and protein bands were easily identified.

2.10 DNA sequencing

Purified PCR or plasmid DNA templates were sequenced by the Allan Wilson Centre Genome Service (ABI3730 Genetic Analyzer, Applied Biosystems; Massey University, Palmerston North, NZ). All plasmids constructed in this thesis were sequenced.

2.11 Bacterial transformation

2.11.1 Electroporation

Plasmids were introduced into \(M. \text{loti}\) strains by electroporation unless otherwise stated. Electrocompetent \(E. \text{coli}\) cells were prepared using a method adapted from Sheng et al. (172). \(E. \text{coli}\) strains were grown to stationary phase; 0.5 mL of this culture was then used to inoculate 500-mL broths of modified SOB medium (Appendix B). Broths were incubated at 37°C with shaking at 200 rpm. The cultures were harvested at \(OD_{600}\) 0.6-0.8 by centrifugation in a chilled (4°C) rotor for 10 min at 5930 \(g\), then washed twice in 500 mL of chilled 10% (v/v) glycerol. Cells were then washed in 30 mL of chilled 10% glycerol, pelleted, resuspended in 1 mL of 10% glycerol, and aliquoted (40 µL) into pre-chilled microcentrifuge tubes. These tubes were snap-frozen in a dry ice/ethanol bath, and stored at -70°C.

Electrocompetent \(Mesorhizobium\) cells were prepared by the method of Kazuhiko Saeki (personal communication). Starter cultures were grown in TY broth until stationary phase, and one mL of
this culture was used to inoculate 200-mL TY broths. Broths were incubated at 28°C with shaking at 180 rpm. Cells were harvested at OD$_{600}$ 0.1-0.3 by centrifugation in a chilled (4°C) rotor at 5930 g, washed once in 200 mL of ice cold 10% (v/v) glycerol and once in 100 mL of 10% glycerol. Cells were washed in 4 mL ice-cold 10% glycerol, pelleted, resuspended in 200 μL 10% glycerol, and aliquoted (40 μL) into pre-chilled microcentrifuge tubes. The tubes were stored at -70°C until use.

Electrocompetent cells were thawed on ice for 10 min. Two μL of ligation or plasmid DNA was added to the tubes and the cells were transferred to a pre-chilled 1 mm-gapped electroporation cuvette (BioRad). Electrottransformation was performed at 1800 V using a Biorad GenePulser Xcell. The cells were immediately transferred to 1 mL of LB or TY broth and incubated at 37°C or 28°C with shaking for 45 min or 2 h for *E. coli* and *Mesorhizobium* spp. respectively. Aliquots were then plated on media containing the appropriate antibiotic(s).

### 2.11.2 Biparental matings

Plasmids were sometimes introduced into *M. loti* by biparental mating using *E. coli* S17-1 λpir as donor strain. *M. loti* strains were grown in TY broth at 28°C for 48 h and *E. coli* were grown at in TY broth at 37°C overnight. Cell cultures (20 μL) were mixed as a spot on the surface of a TY agar plate. The plate was then dried and incubated at 28°C for 24 h. Bacteria were scraped off the plate surface using an inoculating loop and spread-plated on G/RDM plates supplemented with the appropriate antibiotics.

### 2.12 In-frame markerless deletion mutant construction

The dual positive/negative selection vector pJQ200SK (149) was used to replace various ICEM/Sym$^{R7A}$ genes with mutant alleles containing in-frame markerless deletions. pJQ200SK contains the gentamicin resistance selection marker *aadA* and the *sacB* gene which is lethal when sucrose is used as a carbon source. Mutant alleles were designed such that the central portion of the coding sequence and as much conserved sequence as possible was removed. To avoid disrupting possible DNA regulatory sequences, sequences within 50 bp of the 5’ or 3’ ends of the genes were not deleted. Mutant alleles were created using overlap-extension PCR. The overlap-extension products were then cloned into pJQ200SK and sequenced. The constructs containing the mutant alleles were then transferred into *M. loti* from *E. coli* S17 via biparental spot-mating.
Mating mixtures were transferred to G/RDM medium containing gentamicin to select for *M. loti* that had incorporated pJQ200SK through homologous recombination with the mutant allele. Gm<sup>R</sup> colonies were then single-colony purified twice and screened for the presence of both the mutant and wild-type alleles by PCR, using the flanking primers used to generate overlap-extension products. Successful single-crossover mutants were then used to inoculate 5 mL TY broths which were then grown into stationary phase (2-3 days). The broths were then diluted to 10<sup>-4</sup>-10<sup>-6</sup> in fresh TY and spread-plated on RDM medium containing sucrose as a sole carbon source. Colonies able to grow on sucrose were then single-colony purified twice and screened by PCR for presence of the mutant allele but not the wild-type allele. For successful mutants, the loss of pJQ200SK was confirmed by gentamicin sensitivity.

### 2.13 Frozen inocula

For assay of *M. loti* broth cultures, frozen inocula were used to aid in the consistency and convenience of assay replication. To produce the inocula, a loopful of *M. loti* G/RDM plate culture was used to inoculate a 5-mL broth of the appropriate medium (as to be used in the assay) and grown for 64 h (stationary phase, approximately 3.5 x 10<sup>9</sup> cfu/ml). Glycerol was then added to each broth to final concentration of 15% v/v (882 µl of 100% glycerol per 5 mL broth) and the broth vortexed thoroughly. Five-hundred µL aliquots were then stored at -70°C until use.

### 2.14 β-galactosidase reporter assays

β-galactosidase assays were performed using a modification of the procedure described by Miller (121). The recipes of the buffers and solutions used are given in Appendix B. Cells were harvested by centrifugation (1 min, 15110 g) from 1mL or 100 µl of culture (from log-phase (24 h) or stationary-phase (64 h) cultures respectively). Cells were then resuspended in 1mL of Z buffer and the OD<sub>600</sub> of the suspensions were recorded. Eight hundred mL of the suspension were then added to 200 µl of fresh Z buffer to give a final volume of 1 mL, in 2-mL microfuge tubes. Cells were then permeabilized by addition of 40 µL of chloroform and 20 µL of 0.1% w/v SDS and vortexing for 5 s. After incubation for 10 min at room temperature, the assays were started by the addition of 0.2 mL of ortho-nitrophenyl-β-galactoside (ONPG) to each tube, mixing and then incubating the tubes at room temperature. When a pale yellow colour developed the reactions were stopped by the addition of 0.5mL of 1 M Na<sub>2</sub>CO<sub>3</sub>. Tubes were centrifuged at 15110 g for 3 min to sediment the cells after which the OD<sub>420</sub> of the supernatant was recorded.
2.15 Growth curves and QPCR assays

For the growth of cultures used to sample cell numbers and/or excision frequency by QPCR, 25-mL TY or G/RDM broths were grown in 100-mL Erlenmeyer flasks with shaking. Cultures were inoculated with 125 µL of thawed frozen inoculum (approximately $10^7$ cells).

2.16 Bioinformatic analysis

Alignments, alignment editing and phylogenetic tree construction were carried out using the DAMBE software (199) (the ClustalW a.a. alignments were carried out using the slow/accurate setting) or the Vector NTI Advance software suite (Invitrogen, Version 10.0.1). Gene maps were created using the Vector NTI Advance software suite. Motif searches were carried out using the web interface of the MEME (16) and MAST (15) programs (http://meme.sdsc.edu). In silico estimations of protein weights were carried out using the ExPASy ProtParam tool (http://expasy.org/tools/protparam.html). BLASTP, tBLASTX (9) and PSI-BLAST (10) searches were carried out using the web interface at “http://blast.ncbi.nlm.nih.gov/” or for searches including unfinished microbial genomes at “http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi?”. 
Chapter 3
The ICE\textit{M}l\textit{Sym}^{R7A} attachment site \textit{attP}
ICE\textsubscript{ML} Sym\textsuperscript{R7A} is a 502-kb ICE that is able to integrate into the single-copy phenylalanine-tRNA (\textit{phetRNA}) gene of various \textit{M. loti} isolates. An integrase-encoding gene \textit{intS} is the first ORF located in from the left end of ICE\textsubscript{ML} Sym\textsuperscript{R7A}. Sequence analysis indicates that IntS is a tyrosine recombinase that belongs to the satellite bacteriophage P4 integrase family (51, 180), a family of integrases that often target the 3’ end of tRNA genes for integration (195). Integrases catalyze site-specific recombination between short identical DNA sequences and are responsible for both the integration of the element and excision to form a transient circular DNA intermediate.

Previously it has been shown that ICE\textsubscript{ML} Sym\textsuperscript{R7A} excises to produce a circular form of ICE\textsubscript{ML} Sym\textsuperscript{R7A} and that the host genome is repaired upon its excision. This was demonstrated by PCR, using primers specific for the circular form of ICE\textsubscript{ML} Sym\textsuperscript{R7A} and the chromosomal DNA flanking the island insertion site (Figure 3-1). Sequence analysis of the PCR products obtained for both primer pairs confirmed the expected structures of \textit{attP} and \textit{attB}. However no PCR products were obtained from an \textit{intS} mutant strain R7A\textsuperscript{Δ}\textit{intS} (152, 181). This suggests that \textit{intS} is required for excision of ICE\textsubscript{ML} Sym\textsuperscript{R7A}.

In this chapter the requirement of \textit{intS} for excision and transfer of ICE\textsubscript{ML} Sym\textsuperscript{R7A} was confirmed. The minimal DNA region required at \textit{attP} for \textit{intS} expression and efficient integration was also delineated and a repeat sequence likely comprising the arm-type binding site was identified. Additionally the presence of a strong transcriptional promoter on \textit{attP} was confirmed and a minimal region for expression was defined.
Figure 3-1  Location of primers used to detect ICEMISym$^{R7A}$ excision and integration.

ICEMISym$^{R7A}$ DNA is shown as a black bar or circle, chromosomal DNA is shown by white bars. The core sequences (17-bp repeat sequence) are shown by black or white boxes representing their composite structure. Location of primers used for PCR are shown by small arrows above the target sequence. Genes are shown by thick arrows. Map not to scale.
3.1 The integrase IntS is required for excision, integration and efficient transfer of ICEMISym\textsuperscript{R7A}

3.1.1 Complementation of excision in R7AΔintS

To complement R7AΔintS, a 1.3-kb DNA fragment containing intS and 28-bp of upstream sequence was amplified by PCR using primers INTCN5 and INTCN3 and cloned into pFAJ1708 downstream of the nptII promoter, so as to place intS under constitutive expression. The resulting plasmid pJJ611 was then introduced into R7AΔintS (all DNA transformations in thesis were performed using electroporation unless otherwise stated). PCR products of attP and attB were able to be obtained when using DNA extracted from R7A (Figure 3-2) or R7AΔintS(pJJ611) as a template (data not shown) but not when R7AΔintS DNA was used (Figure 3-2), indicating that expression of intS was required for ICEMISym\textsuperscript{R7A} excision.

3.1.2 Complementation of transfer in R7AΔintS

Transfer of ICEMISym\textsuperscript{R7A} likely requires excision and therefore intS. Transfer of ICEMISym\textsuperscript{R7A} to the non-symbiotic \textit{M. loti} strain N18 containing pFAJ1700 (Tc\textsuperscript{R}) was assayed using R7AΔintS or R7AΔintS(pJJ611) as a donor (conjugation assays carried out by Gabriella Stuart (153)). This revealed that transfer of ICEMISym\textsuperscript{R7A} was indeed reduced from R7AΔintS compared to that from wild-type (1.2 x 10\textsuperscript{-4} to 3.9 x 10\textsuperscript{-9} exconjugants/donor respectively). However complementation of transfer was not observed from R7AΔintS(pJJ611) (data not shown). It seemed likely that intS is also required for integration of ICEMISym\textsuperscript{R7A} in the recipient strain, so complementation of intS in both the donor and the recipient strain was tested. A gentamicin-resistant derivative of pJJ611, pJJ611G (constructed by John Sullivan) was introduced into \textit{M. loti} N18. Transfer of ICEMISym\textsuperscript{R7A} from R7AΔintS(pJJ611) to \textit{M. loti} N18(pJJ611G) was observed at a near-wild type frequency of 6.6 x 10\textsuperscript{-5} (± 3.4 x 10\textsuperscript{-5}), indicating that IntS was required in both the donor and recipient in order to achieve efficient transfer.
Figure 3-2 Detection of ICEM/Sym\textsuperscript{R7A} by PCR.

PCR products for the attachment sites \textit{attP}, \textit{attB}, \textit{attR} and \textit{attL} in various strains. PCR amplicons (indicated left) obtained from wild-type and mutant strains (named above). An \textit{attL} amplicon was not obtained from strain R7A\textDelta\textit{intS} as the \textDelta\textit{intS} deletion removed the priming site for LE1. This figure has been adapted for this thesis from published results (153).

3.1.3 Integration of \textit{attP}-containing plasmids into R7ANS

Previously it was shown that a 1.9-kb DNA region containing \textit{attP} and \textit{intS} is able to direct integration of a suicide vector into the \textit{M. loti} CJ4 (178) \textit{p}he-tRNA gene. This region is composed of 1,464 bp at the left end of ICEM/Sym\textsuperscript{R7A} and 440 bp from the right end (including the 17-bp region believed to contain the core IntS-binding sequence), amplified by PCR from the circular form of ICEM/Sym\textsuperscript{R7A} using primers CIB5 and CIB3. This DNA region was sub-cloned as a BamHI fragment from pJJ608 into plasmid pFUS2 which is unable to replicate in \textit{M. loti}. This construct pJJ609 was tested for the ability to integrate into the \textit{p}he-tRNA gene of the ICEM/Sym\textsuperscript{R7A}-cured derivative of R7A, strain R7ANS (derivation of R7ANS is described in Chapter 4). Following electroporation of pJJ609 into R7ANS, cells were plated on G/RDM containing gentamicin to select for cells maintaining pJJ609. Gentamicin-resistant colonies were
observed at a high frequency (~1,500 colonies on average from 200 ng of DNA electroporated, estimated from a 1/3 dilution of the electroporation mixture), while electroporation of pFUS2 did not give rise to any Gm\(^R\) colonies. DNA from a Gm\(^R\) colony resulting from the pJJ609 electroporation was extracted and analysed by PCR, using primers specific for \textit{attP}, \textit{attB}, \textit{attL}, and \textit{attR} (Figure 3-2). PCR products were obtained for \textit{attL} and \textit{attR}, indicating that pJJ609 had integrated into \textit{attB} adjacent to the \textit{phe}-tRNA gene. The absence of products for \textit{attP} and \textit{attB} suggested pJJ609 did not excise from the chromosome at a frequency detectable by this assay (Figure 3-2).

To demonstrate integration of a vector containing \textit{attP} when IntS was expressed \textit{in trans}, a region containing \textit{attP} (772 bp) consisting of 288 bp upstream of the core, the upstream UTR and the first 287 bp of \textit{intS} (Figure 3-3) was amplified from pJJ609 by PCR (with primers IntSTranFusionBamHI and IntSTranFusion288attp) and cloned into pFUS2 as a BamHI fragment. Electroporation of this construct pFUS288 into R7ANS cells failed to produce any Gm\(^R\)-colonies, indicating that \textit{intS} was required for integration. The plasmid pJJ611 which constitutively expresses \textit{intS} from the \textit{nptII} promoter was then introduced into R7ANS. Similar to the results obtained with pJJ609, the electroporation of pFUS288 into R7ANS(pJJ611) resulted in the appearance of Gm\(^R\) colonies at high frequency and PCR analysis confirmed integration of pFUS288 into \textit{attB} (data not shown).

Taken together, the above results confirm that \textit{intS} is required for excision and efficient transfer of ICE\textit{Ml}/Sym\textit{R7A} and show that it is the only symbiosis island encoded gene required for recombination of the attachment site \textit{attP} with the target integration site \textit{attB}.  


Figure 3-3  *attP* regions cloned into pFJX/pFUS2.

Nested PCR products of *attP* were cloned into pFUS2 or pFJX to demonstrate integration of *attP* into *attB* or measure transcriptional expression from the *attP-intS* promoter. The *attP* region is composed of the right and left ends of *ICE*MISym\(^{R7A}\), joined at the 17-bp region containing the core IntS-binding sequence and is formed as a result of recombination between *attL* and *attR*. Only plasmids pFJX288-(F) and pFJX122-(F) produced maximum levels of expression of *lacZ* (measured by β-galactosidase activity).

### 3.2 Expression of *intS* from *attP*

#### 3.2.1 Expression of *intS* is from a strong promoter on *attP*

Previously it has been demonstrated that *intS* is differentially expressed on *attP* and *attL*; the expression of *intS* translationally fused to *lacZ* (cloned in the vector pLAFR1) is greater from a construct derived from *attP* than *attL* (unpublished data). These assays were confirmed in this study using samples from stationary-phase (64 h) *M. loti* R7A TY-broth cultures. Plasmid p1311 that contains a *lacZ* fusion to *attP* produced 249 (±36) Miller units (MU) of β-galactosidase activity while plasmids containing an out-of-frame *attP-lacZ-intS* fusion (p1310) or a *lacZ* fused to *attB-intS* (p1224) produced less than 4 MU. This suggested that the formation of *attP* results in either the generation or juxtaposition of a promoter not present on *attL*, or alternatively, that *attP* lacks a repressive *cis*-acting element located on *attL* that silences expression of *intS*. To confirm transcriptional activity of *attP*, the *attP*-containing PCR product used to create pFUS288 was cloned into pFJX (122) as a BamHI fragment so as to create a transcriptional fusion with *lacZ*. The pFJX plasmid contains the *lacZ* cassette from pFUS2, which has been constructed such that cloned fragments are positioned upstream of stop codons in all three open reading frames, upstream of a promoterless *lacZ* gene with its own ribosome binding site. The resultant
construct pFJX288-F was introduced into R7A and the strain was assayed for β-galactosidase activity. The strain produced 1148 (±73) MU and 2578 (±74) MU of β-galactosidase activity in log-phase and stationary-phase TY-broth cultures respectively. Results from G/RDM-broth cultures were similar (data not shown). The background expression from the pFJX construct was 117 (±5)MU and 276 (±30) MU, while a pFJX construct containing the attP BamHI-fragment in the reverse orientation (pFJX288-R) provided 16 ± 6MU and 28 (±7) MU (Table 3-1). These data suggest that a strong constitutive promoter exists on attP that drives expression of intS.

### 3.2.2 Deletion mapping of the attP-intS promoter

To delineate the promoter region on attP, a nested series of PCR products was designed to progressively reduce the 5’ end of the attP region present in pFJX288-F. Primers were designed to produce attP fragments containing 122, 53 and 10-bp of DNA upstream of the core in addition to the entire intS upstream UTR (Figure 3-3). The PCR products produced with these primers were subsequently cloned into pFJX. Analysis revealed that the while expression from pFJX122-F was almost identical to that pFJX288-F, constructs containing 53 bp or 10 bp of sequence upstream of the core sequence produced background levels of β-galactosidase activity (Table 3-1) indicating that a crucial promoter element located between 53 and 122 bp upstream of the core sequence. In an attempt to further define the promoter region, a PCR fragment containing only the 122-bp region upstream of the core sequence and the core sequence itself was amplified and cloned into pFJX producing pFJXattR-F. Only a background level of expression was observed from this construct (Table 3-1), indicating that a region downstream of the core sequence is also required for transcriptional activity.

A series of nested PCR products were then designed to progressively reduce the 3’ end of the pFJX288-F attP region. Regions containing 122 bp of DNA upstream of the core and 177, 71 or 10 bp downstream of the core were amplified by PCR and cloned into pFJX (Figure 3-3). β-galactosidase activities from these constructs (assays carried out by Rujirek Noisangiam) were again lower than from pFJX122-F (Table 3-2). This result suggests that a surprisingly large region on attP is required for full expression of intS.
Table 3-1 Expression of lacZ from 5'-nested attP fragments

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>t=24 h</th>
<th>t=64 h</th>
</tr>
</thead>
<tbody>
<tr>
<td>pFJX</td>
<td>117 ± 5</td>
<td>276 ± 30</td>
</tr>
<tr>
<td>pFJX288-F</td>
<td>1148 ± 73</td>
<td>2578 ± 74</td>
</tr>
<tr>
<td>pFJX288-R  a</td>
<td>4 ± 6</td>
<td>17 ± 9</td>
</tr>
<tr>
<td>pFJX122-F</td>
<td>1148 ± 11</td>
<td>2544 ± 86</td>
</tr>
<tr>
<td>pFJX122-R</td>
<td>16 ± 6</td>
<td>28 ± 7</td>
</tr>
<tr>
<td>pFJX53-F</td>
<td>169 ± 16</td>
<td>474 ± 28</td>
</tr>
<tr>
<td>pFJX53-R</td>
<td>32 ± 0</td>
<td>151 ± 128</td>
</tr>
<tr>
<td>pFJX10-F</td>
<td>86 ± 3</td>
<td>219 ± 20</td>
</tr>
<tr>
<td>pFJX10-R</td>
<td>24 ± 3</td>
<td>40 ± 7</td>
</tr>
<tr>
<td>pFJXattR-F</td>
<td>94 ± 19</td>
<td>165 ± 20</td>
</tr>
<tr>
<td>pFJXattR-R</td>
<td>156 ± 5</td>
<td>276 ± 17</td>
</tr>
</tbody>
</table>

*Data presented is the mean of two independent replicates.  a Plasmids with the –R notation contain attP fragments cloned in the reverse orientation and are used as negative controls to determine the background level of expression.

Table 3-2 Expression of lacZ from 3'-nested attP fragments

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>t = 24 h</th>
<th>t = 62 h</th>
</tr>
</thead>
<tbody>
<tr>
<td>pFJXattR-F</td>
<td>142 ± 2*</td>
<td>234 ± 6</td>
</tr>
<tr>
<td>pFJX122-10</td>
<td>231 ± 0</td>
<td>712 ± 9</td>
</tr>
<tr>
<td>pFJX122-71</td>
<td>54 ± 0</td>
<td>69 ± 1</td>
</tr>
<tr>
<td>pFJX122-177</td>
<td>92 ± 3</td>
<td>120 ± 1</td>
</tr>
<tr>
<td>pFJX122-F  a</td>
<td>846 ± 16</td>
<td>1646 ± 23</td>
</tr>
<tr>
<td>pFJX122-R  a</td>
<td>23 ± 0</td>
<td>25 ± 1</td>
</tr>
<tr>
<td>pFJX53-F</td>
<td>254 ± 8</td>
<td>400 ± 7</td>
</tr>
</tbody>
</table>

*Assays were carried out by Rujirek Noisangiam.  a pFJX122-F and pFJX122-R were included as positive and negative controls respectively.  b Data presented is the mean of two independent replicates.
3.3 Mapping of the minimal attP required for recombination

3.3.1 Identification of a conserved motif on attP

Attachment sites of tyrosine recombinases similar to that encoded by intS contain two cis-acting regions, a core binding site at the site of recombination and “arm” binding sites flanking the core. Arm sites are often comprised of a set of short, loosely-conserved direct or inverted repeats, ~50-140-bp from the core (Figure 3-4) (70, 156). To gain an indication of the level of DNA sequence conservation across attP with that of related (putative) attachment sites, the upstream UTRs of closely related integrases were analysed with the bioinformatics motif discovery tool MEME (14). The basic local alignment search tool BLASTP was used to identify proteins similar to IntS excluding those from other M. loti species. The genomic sequences encoding the IntS homologues were then examined for the presence of a tRNA gene directly upstream of the int gene. DNA sequences were then extracted from the UTR between the tRNA gene and the start codon of the intS homologue. Two sequences from Sphingomonas sp. SKA58 (upstream UTR of ZP_01304363 and ZP_0130420) and a sequence from Rhodopseudomonas palustris BisB18 (upstream UTR of YP_530520) were submitted to MEME along with the syntenic region upstream of intS. This revealed the presence of an 11-bp sequence which was most commonly “TGATGGTATCT”, found 4-5 times on each sequence, clustered towards the start of the integrase-encoding ORF (Figure 3-5). After manual investigation of these repeats on the ICEMlSymR7A attP a shorter consensus sequence of “GVTGGTATY” was proposed and one of the original repeats identified in the MEME analysis was then excluded. Further analysis of the attP sequence upstream of the core (not included in the original analysis) revealed the presence of another two repeats. An alignment of the final five sequences produced the 9-bp consensus sequence “GNTGGYAYB”. The individual repeats were labelled P1-P5 inline with common convention (Figure 3-4). Their location within the ICEMlSymR7A attP is shown in Figure 3-6.
Figure 3-4  Organisation of arm-type integrase binding sites on various attP regions.

Comparison of the organisation of arm sites (denoted P) and core site (denoted C) found on the attP regions of various elements encoding tyrosine recombinases. Diagram taken from (156).
Figure 3-5  MEME output from a motif search of attP-related sequences.

Results from a MEME analysis of the ICE/SmR\textsuperscript{R\textalpha} attP region (downstream of the core) with DNA sequences from Sphingomonas sp. SKA58 and Rhodopseudomonas palustris BisB18 located upstream of IntS homologues (ZP_0130420, ZP_01304363 from R. palustrus Bis18 and YP_530520 from S. sp. SKA58). The output shows (a) an alignment of the identified motif with the surrounding DNA context and (b) a diagram of the relative positions and orientations of the motif within the individual sequences. The start site of each sequence shown in (b) is the final nucleotide of the preceding tRNA gene.
Figure 3.6 The nucleotide sequence of the ICEMISym^R7A^ attachment site attP.

The nucleotide sequence of attP coordinates correspond to those published in Sullivan et al. 2002 (182), the first nucleotide after the core represents the +1 position. Primers are shown as arrows above (5’-3’) or below (3’-5’) the sequence (note primer labels often only approximate coordinates). Nucleotides matching the ‘GNTGGYAYB’ motif are shown in bold-italics (all orientated 5’-3’) and labelled P1-P5. The core sequence (17-bp repeat sequence flanking the integrated ICEMISym^R7A^) of attP is shown in bold. Part of the intS sequence is identified by a large arrow above the single-letter code a.a. sequence.

3.3.2 The repeat motifs are required for efficient integration of attP-containing plasmids

To investigate the possible role of the GNTGGYAYB sequence in the IntS-mediated integration of attP, nested PCR products of attP lacking either one or two of the outermost repeats were cloned into pFUS2. The resultant plasmids were then electroporated into R7ANS(pJJ611) and Gm^R^ colonies were selected to detect integration of each plasmid. Primers attPrepeat(-82) and attPrepeat(+187) were used to amplify the region P1-P5 and a second two primers attPrepeat(-69) and intS5’UTR+177bp were used to amplify the region P2-P4. Combinations of the same primers were then used to amplify regions from P1-P4 (attPrepeat(-82) and intS5’UTR+177bp) and P2-P5 (attPrepeat(-69) and attPrepeat(+187)). The PCR products were then cloned into
pFUS2 as BamHI fragments. The orientation of the cloned fragment was verified by PCR using a combination of one primer used to amplify the fragment and the lacZ primer which anneals to pFUS2. Clones containing the PCR fragment in each orientation relative to pFUS2 were both used, to cover for the possibility that the surrounding vector sequence may affect the integration efficiency of the plasmid. DNA was prepared from the eight plasmids containing the four attP regions in each orientation and 200-ng was then electroporated into R7ANS(pJJ611). Results are shown in Figure 3-7. The electroporation of either pFUS2 into R7ANS(pJJ611) or pFUS122-F into R7ANS did not give rise to any GmR colonies (data not shown). Electroporation of plasmids containing the region P1-P5 in either orientation into R7ANS(pJJ611) produced a similar number of colonies to the electroporation of pFUS288, indicating this region contained the cis-acting sequences required for efficient integration. However electroporation of plasmids containing the regions P2-P5 or P1-P4 produced only 60-180 colonies, indicating integration efficiency of these plasmids was impaired. Gentamicin-resistant colonies were not obtained following electroporation of plasmids containing P2-P4, regardless of the fragment’s orientation. These data indicate that the outermost repeat sequences P1 and P5 are both required for efficient IntS-mediated recombination of attP and attB. This suggests that the repeated GNTGGYAAYB motif found on attP forms part of the IntS arm-type binding sites commonly found on the attachment sites of tyrosine recombinases.

Figure 3-7 attP fragments cloned into pFUS2 for analysis of integration.

Various attP fragments were cloned into pFUS2 (in both forward and reverse orientations) were tested for their ability to integrate into attB. Purified plasmid DNA (200 ng) of each construct was electroporated into R7ANS(pJJ611) (using a single batch of electrocompetent cells) which is cured of ICE MlSym R7A but constitutively expresses intS. attP integrants were selected for by plating on media containing gentamicin. Estimated total GmR colonies from the electroporation (3 x plate count from 1/3 dilution) are shown for each cloned fragment. The experiment was repeated and results were similar; data presented are from single plate counts from one experiment.
3.4 Discussion

Experiments carried out in this chapter demonstrated the role of the integrase encoded by \textit{intS} and its cognate attachment site \textit{attP} in excision, integration and transfer of ICEMISym$^{R7A}$. IntS was required for the formation of \textit{attP} and was also able to mediate integration of a plasmid containing \textit{attP} into \textit{attB} in R7ANS \textit{in trans}.

The transfer of ICEMISym$^{R7A}$ to non-symbiotic strains likely requires excision in the donor strain and integration into the recipient’s genome. Therefore it was not surprising that efficient transfer of ICEMISym$^{R7A}$ from R7AΔ\textit{intS} required complementing plasmids containing \textit{intS} in both the donor and the recipient. Interestingly transfer of ICEMISym$^{R7A}$ was not completely abolished in R7AΔ\textit{intS} in the absence of the complementing plasmids. Further work carried out by Gabriella Stuart revealed that this appears to be due to the transfer of part of ICEMISym$^{R7A}$ to the non-symbiotic strain in a process reminiscent of Hfr transfer of the integrated F plasmid in \textit{E. coli} (unpublished data).

Integrases bind a series of short repeat sequences flanking the core region on the attachment sites \textit{attP}, \textit{attL} and \textit{attR} called arm sites. Analysis of the $\lambda$ integrase has revealed that different subsets of the repeats are bound during excision and integration and that they help aid in the conformational changes in the nucleoprotein complex between the two states (150). Additionally these sequences determine the specificity of a particular integrase for its cognate attachment sites \textit{attP}, \textit{attL} and \textit{attR} (32). The bioinformatics analyses in this chapter revealed the presence of the sequence GNTGGYAYB at five sites on \textit{attP} 60-80 bp upstream of the core sequence (P1 and P2) and 146-186 downstream of the core (P3-P5). Similar sequences were found clustered upstream of \textit{intS} homologues identified through a BLASTP search.

An assay of \textit{attP} integration revealed that the outermost arm repeats P1 and P5 were required for efficient IntS-mediated integration. Interestingly low frequencies of pFUSP1P4 and pFUSP2P5 integrants were obtained using this assay, revealing that integration occurs albeit inefficiently without these repeats. An interesting future investigation would be to test if these plasmids and mutated variants are also able to excise, which may give some indication if the repeats required for integration are the same as those required for excision. The integration assay used in this study could also be used to map the IntS-binding region and recombination site within the core,
by introducing site-specific changes within the core and following the migration of the substituted nucleotides after recombination (assuming that recombination was still able to occur).

Several elements express an integrase from alternate promoters on \textit{attP} and \textit{attB}, resulting in increased expression of the integrase when the element is excised (47, 113, 171). ICEMISym\textsuperscript{R7A} shares this feature, expressing \textit{intS} from a strong promoter encoded on \textit{attP}. Experiments designed to delineate the promoter revealed that a region between 53 bp and 122 bp upstream of the core was required for expression, suggesting a promoter could be encoded within this region. Similar to \(\sigma\)\textsuperscript{70} promoters in \textit{E. coli}, studied rhizobial promoters comprise as little as \(~35\) bp of DNA (114), which allows for the possibility that the promoter on \textit{attP} is encoded entirely within the region upstream of the core, as is observed on the \textit{clc} element. However in conflict with this hypothesis was data demonstrating that removal of DNA downstream of the core resulted in a loss in expression from \textit{attP}, indicating further DNA downstream of the core was required.

Analysis of nested deletions of \textit{attP} revealed that a region containing 122 bp upstream and 177 bp of downstream of the core was unable to produce full expression. It is not clear why such a large region on \textit{attP} was required for maximum expression. Interestingly this sequence contains all of the arm repeats except for P5, a sequence also required for efficient IntS-mediated integration of \textit{attP}. Further experiments investigating the coincidence of the sequence requirements for \textit{attP} integration and \textit{intS} expression are underway. On phage P4, the transcriptional start site of \textit{int} is within the second-closest arm-repeat sequence upstream of the start codon and transcription is negatively regulated by binding of Int and the P4 RDF Vis. It is possible that the \textit{intS} transcriptional start site is in an analogous position; however unlike in P4, previous experiments by J. Sullivan (personal communication) showed that expression from \textit{attP} in R7A\textDelta\textit{intS} is similar to that in R7A.

Attempts to map the transcriptional start site of the \textit{intS} promoter(s) using primer extension were made during this study but were unsuccessful, most likely due to difficulties experienced in extracting high enough quantities of RNA from \textit{M. loti}. Further attempts to map the transcript using rapid amplification of cDNA ends (5’-RACE) will be carried out in the near future.

In summary the ICEMISym\textsuperscript{R7A}-encoded IntS is required for excision and efficient transfer as well as being required for integration in the recipient. The integration of a plasmids containing \textit{attP} in R7ANS demonstrated that IntS is the only ICEMISym\textsuperscript{R7A}-encoded protein required for
integration and that catalysis requires repetitive sequences within \textit{attP}, likely representing IntS binding sites. Expression of \textit{intS} is driven from a strong constitutive promoter on \textit{attP}, which likely ensures integration of ICEM/\textit{Sym}^{R7A} following transfer to a new recipient strain. The exact nature of the promoter on \textit{attP} was unable to be elucidated in this study, but was shown to require an unusually large region of DNA composed of regions both up- and downstream of the core on \textit{attP}. Further work is required to map the transcriptional start site(s) of \textit{intS} and confirm the IntS-binding sites.
Chapter 4
ICEMLSym$^{R7A}$
Excision
In the previous chapter it was shown that excision of ICEMlSym\textsuperscript{R7A} to form a circular intermediate is essential for its efficient transfer. As excision likely represents the first step in transfer of ICEMlSym\textsuperscript{R7A}, knowledge of the regulation controlling excision is instrumental in unravelling the transfer strategy employed by ICEMlSym\textsuperscript{R7A}.

While the presence of the integrase protein IntS is a crucial factor for the excision of ICEMlSym\textsuperscript{R7A}, stimulation of excision is likely to require a secondary factor. The recombination reaction catalysed by other characterised integrases has been found to be highly directional, favouring integration in the absence of additional factors. Excision is stimulated by the presence of a recombination directionality factor (RDF), also termed an excisionase, that binds and bends DNA within the attachment sites to promote excisive recombination (162). RDFs are small (generally <100 a.a.) proteins that are usually basic, and some contain a helix-turn-helix DNA-binding motif. They exhibit considerable variation in a.a. sequence, even within members of the same family (105).

In this chapter a quantitative polymerase chain reaction (QPCR) assay was developed to measure the abundance of the excision products \textit{attP} and \textit{attB} in a cell population relative to the number of chromosomes present. Quantitative PCR is a method of estimating the original template concentration in a DNA sample by analysing the number of cycles required to produce a threshold concentration of the PCR product (amplicon) (2, 80). Since PCR involves exponential amplification of the amplicon (at least while all reagents are in excess), the number of cycles (Ct) it takes to reach a threshold amplicon concentration can be used to extrapolate the original template concentration. Amplicon concentration is measured by an increase in a fluorescent signal incorporated into the reaction, and is generally a dsDNA-binding dye such as SYBR green or a DNA sequence-specific fluorescent probe. The total fluorescence is internally normalised to a second dye such as ROX. For final quantitation of DNA within a sample, concentrations are either calculated from a standard curve of known DNA concentration (absolute quantitation) or by normalisation to a second sample template (relative quantitation) (2).

The relative QPCR assay developed was used to demonstrate a growth-phase dependence of excision frequency. In addition a novel ICEMlSym\textsuperscript{R7A} recombination directionality factor, RdfS, was identified and its function confirmed using the QPCR assay.
4.1 Measurement of excision by QPCR

4.1.1 Development of a quantitative PCR assay of ICEMISym\textsuperscript{R7A} excision

To measure the frequency of excision of ICEMISym\textsuperscript{R7A}, a quantitative PCR (QPCR) assay was developed to detect the \textit{attB} (unoccupied integration site) and \textit{attP} (circular form of ICEMISym\textsuperscript{R7A}) sequences in DNA samples (Figure 4-1). Previously a similar assay was developed using SYBR green based chemistry (152). Due to the formation of primer-only generated amplification products, it was decided the non-specific DNA binding activity of SYBR green was unsuitable for this application. Instead an assay utilising Taqman\textregistered FAM-labelled minor-groove-binding (MGB) probes was developed. Taqman\textregistered probes are ssDNA probes designed to anneal within the PCR amplicon. During the template extension stage of PCR the 5’ exonuclease activity of the DNA polymerase cleaves the fluorophore attached to the 5’ end of the probe. When cleaved, the fluorophore is not longer quenched by a molecule attached to the 3’ end of the probe and so generates a fluorescent signal. Probe cleavage and generation of fluorescence only occurs when the probe binds the desired sequence, giving specificity to the assay.

Three primer and probe combinations were designed to detect \textit{attP} (ATTPL, ATTPR and TAQATT (probe)), \textit{attB} (ATTBL, ATTBR and TAQATT (probe)) and \textit{melR} (MELRL, MELRR and TAQMR (probe)) and the resulting amplicons for each QPCR reaction were 100 bp, 105 bp and 99 bp respectively. The \textit{melR} gene is located on the \textit{M. loti} R7A chromosome adjacent to the \textit{phe}-tRNA gene at the site of ICEMISym\textsuperscript{R7A} integration. The detection of \textit{melR} was used as a measure of chromosome copy number for relative quantification of \textit{attP} and \textit{attB}. 

73
Figure 4-1 Location of QPCR primer-probe combination used to detect ICEM/Sym\textsuperscript{R7A} excision.

The positions of primer-probe combinations (see Materials and Methods) are indicated by dumbbells above the target sequence. ICEM/Sym\textsuperscript{R7A} DNA is shown as a black bar or circle, chromosomal DNA is shown by white bars and plasmid DNA is hatched (Maps not to scale). a) Model for ICEM/Sym\textsuperscript{R7A} excision and integration. b) Map of pJR201 used as a standardisation template for the QPCR assay.

To validate the QPCR assay, the individual reactions had to be shown to produce a similar level of relative fluorescence for the template supplied over a wide range of template concentrations. To test this, a plasmid containing a single copy of each of the target amplicons was constructed for use as a control template. Regions containing the QPCR amplicons of \textit{attP} (700 bp), \textit{attB} (500 bp) and \textit{melR} (670 bp) were amplified by PCR using primer pairs RE2/LE1, LE2/RE1, and ML/MR respectively (Table 2-3, Figure 3-1). The \textit{attP} and \textit{melR} PCR products were cloned into pUC8 digested with EcoRI and HindIII, as BamHI-EcoRI and BamHI-HindII fragments respectively. The resulting plasmid was then purified and digested with EcoRI and the \textit{attB} PCR
product was then cloned as an EcoRI fragment to produce pJR201. Purified pJR201 DNA (Figure 4-1) was linearised with BamHI and then purified by agarose gel electrophoresis, so as to avoid possible effects of plasmid supercoiling on amplification efficiency. Several QPCR standard curves were generated using linearised pJR201 as a template at concentrations ranging over 4 orders of magnitude (for an example see Figure 4-2). This revealed that the fluorescent signal was nearly identical for each amplicon when pJR201 was supplied as a template, and that detection of template was accurate except for very low template concentrations (less than ~10 molecules). Amplification efficiencies determined from standard curves (see Materials and Methods) were 1.92, 1.90 and 1.81 for the attB, attP and melR amplicons respectively. These values are close to the optimal efficiency of 2.00. The final calculation of QPCR results included adjustments to compensate for reaction efficiency differences as well as normalisation to pJR201 to account for subtle differences in total signal for each amplicon (see Materials and Methods).
Figure 4-2  QPCR standard curves.
Standard curve of *attP* (red) *attB* (blue) and *melR* (black) amplicons generated by QPCR of performed on a dilution series of linearised pJR201 DNA. DNA was diluted by serial 1 in 4 dilutions; reactions contain between about $1.31 \times 10^5$ and 3.2 copies of pJR201 per reaction (calculated from spectroscopy and molecular weight estimate of pJR201). The y-axis shows the mathematical derivative of the relative fluorescent signal. The x-axis shows the PCR cycle number (Ct). The green line shows a selection of the threshold fluorescence within a period of uniform exponential amplification from which the ct values are taken.
Figure 4-3  Excision of ICEMISym$^{R7A}$ during growth in broth culture.
QPCR was used to determine the percentage of $attP$ and $attB$ sequences resulting from excision of ICEMISym$^{R7A}$. The percentages of cells containing $attP$ are shown as black bars and $attB$ as white bars. The cfu of the growth curve is shown as a continuous line. Data presented was the average of three independent experiments and was log-transformed to remove a positive association of the standard deviation with the mean (error bars represent ± 1 standard deviation). Individual QPCR measurements are the mean of triplicate reactions for each amplicon.

4.1.2 Excision of ICEMISym$^{R7A}$ increases in stationary-phase cultures

The QPCR assay was used to investigate the frequency of excision of ICEMISym$^{R7A}$ throughout growth of *M. loti* R7A in broth culture (for an example QPCR reaction see Figure 4-4). DNA was prepared from broth cultures that were grown with shaking and sampled at 8-h intervals until late-stationary phase. The excision products $attP$ and $attB$ were present at low frequency in exponentially growing cultures (0.08% and 0.04% for $attP$ and $attB$ respectively), but increased approximately 10- to 100-fold during stationary phase (Figure 4-3). The ratio of $attP$ to $attB$ was close to 1:1 indicating that ICEMISym$^{R7A}$ was not being lost from the population or replicating at a rate much higher than that of the chromosome.
Figure 4-4  Increased presence of attPlattB in stationary-phase cultures.

This graph shows overlaid QPCR assay results (reactions are in triplicate) from t = 24 h and t = 64 h samples from one biological replicate of the growth curve shown in Figure 4-3. The y-axis shows the mathematical derivative of the relative fluorescent signal. The x-axis shows the PCR cycle number. The green line shows a selection of the threshold fluorescence within a period of uniform exponential amplification from which the Ct values are taken. This result demonstrates that while the concentration of melR (shown in black) is equal for both samples, the attP (red lines) and attB (blue lines) show lower ct values (and therefore higher DNA concentrations) at t = 64 h than at t = 24 h.

4.1.3 QPCR assay of complementation of excision in R7AΔintS

In order to confirm the previous result that excision absolutely required intS, DNA was extracted from cultures of R7AΔintS and analysed by QPCR (Figure 4-5). As expected, amplification products were not produced for attP or attB. The effect of constitutive production of IntS on excision was examined by introducing pJJ611 into strain R7A. Strain R7A(pJJ611) showed levels of excision in exponentially-growing and stationary-phase cells similar to those of R7A or R7A containing the pFAJ1700 vector alone. Surprisingly the levels of excision in R7AΔintS(pJJ611) were about 10-fold higher than in R7A in both phases of growth. This result confirms that excision was restored in R7AΔintS by providing intS in trans and also suggests that the ΔintS mutation may have some regulatory effect on excision.
Figure 4-5 The effect of intS and rdfS mutations on ICEMISymR7A-excision.

Triplicate measurements were obtained for each sample, and the mean ± standard deviation is shown for each assay. Percentage attP and attB at 24 h (exponential growth) are shown by black and white bars respectively, while percentage attP and attB at 64 h (stationary phase) are shown by black bars/white hatching and white bars/black hatching. Asterisks indicate samples with significant deviation from the wild-type data (* P < 0.05, ** P < 0.01). No target was detected in R7AΔintS, and no attB was detected in R7AΔrdfS. At t=24 for R7AΔrdfS, attP was detected in only a single sample and so no standard deviation or significance value was obtained. Data presented is the average of three independent experiments (error bars represent ± 1 standard deviation). Individual QPCR measurements are the mean of triplicate reactions for each amplicon. pJJ611 contains intS expressed from the nptII promoter, pJR202 contains rdfS and pFAJ1700 was included as a plasmid control.
4.2 Identification of RdfS

4.2.1 Bioinformatic identification of *msi109* as a possible recombination directionality factor

Most integrases require the presence of a RDF to stimulate excision (70). In order to find candidate RDFs on ICE*Ml*Sym<sup>R7A</sup>, BLAST searches of the symbiosis island sequence were carried out using the sequence of the bacteriophage P4 RDF, Vis (30), as the query. Vis is a member of the SLP1 subfamily of RDFs, all of which are associated with P4-family integrases (105). No obvious candidates were found using this approach so PSI-BLAST (10) was used to identify more distantly related RDF homologues from the NCBI database. The Vis sequence was used as the query, and sequences of between 70 and 150 residues were selected for submission in iterations. After the third iteration, the search results included the ICE*Ml*Sym<sup>R7A</sup> ORF *msi109* that encodes an 89 amino-acid protein with limited similarity to Vis and related RDFs (Figure 4-6).

**Figure 4-6** Similarity of RdfS to known RDFs.

ClustalW alignment of the RdfS amino-acid sequence with similar RDFs that interact with P4-family integrases. Rox (gi:40949919) from the *Shigella flexneri* 2a she pathogenicity island, Vis (gi:1589630) of bacteriophage P4 and Hef (Yp43) (gi:4106643) from HPI of *Yersinia pestis*. Bold font indicates identical a.a.

*msi109* is a member of a gene cluster *msi110 – msi106* (Figure 4-7) that Sullivan *et al.* (182) speculated was involved in conjugative transfer on the basis of the similarity of *msi108* (*traF*) and *msi107* to genes known to have a conjugative role. A near-identical cluster is present on the putative symbiosis island of *M. loti* strain MAFF303099 except for the presence of a transposon inserted between the *msi107* and *msi106* orthologues in MAFF303099 (93, 182)
Figure 4-7  Organization of the msi110–msi106 gene cluster on ICEMISymR7A. msi109 was renamed as rdfS during this work (data from Sullivan et al., 2002 (182)).

4.2.2 rdfS (msi109) is required for efficient excision and transfer of ICEMISymR7A

To determine if msi109 had a role in excision of ICEMISymR7A, an in-frame markerless deletion mutant R7AΔmsi109 was constructed (Figure 4-8). QPCR analysis of the mutant showed that attP was only present sporadically at low frequencies (<0.003%) in independently-derived templates extracted from exponential or stationary-phase cultures, while attB was not detected (Figure 4-5). To complement R7AΔmsi109, the DNA encoding msi109 and the preceding 210-bp region amplified by PCR using primers 109L and 109CN3 and cloned into pFAJ1700 to produce pJR202. When pJR202 was introduced into R7A the excision frequency was not affected in comparison to R7A or R7A(pFAJ1700). Surprisingly excision frequencies approximately 100-fold higher were observed when pJR202 was used to complement R7AΔmsi109, in both exponential and stationary phases of growth (Figure 4-5). Based on these results, msi109 was renamed as rdfS to reflect its role as the ICEMISymR7A recombination directionality factor.

To test whether R7AΔrdfS was able to transfer ICEMISymR7A to non-symbiotic mesorhizobia, conjugation experiments were carried out using the non-symbiotic Mesorhizobium sp. strain N18 containing pFAJ1700 (TcR) as recipient (experiments carried out by Gabriella Stuart). Wild-type R7A transferred the island at a frequency of $1.2 \times 10^{-4}$ ($\pm 1.0 \times 10^{-4}$) exconjugants/donor, while no transfer of the island was detected from R7AΔrdfS (transfer frequency less than $10^{-10}$ exconjugants/donor).
Figure 4-8 Construction of R7AΔrdfS.

Plasmid pJK301, which contains an in-frame markerless 136-bp deletion within rdfS, was used to construct R7AΔrdfS (see Section 2.12). A) The a.a. sequence encoded by of the mutant rdfS allele in R7AΔrdfS and the deleted region (shown above). Sequence shown in bold is the a.a. sequence corresponding to a cloned rdfS gene in pJR204 isolated after electroporation of this plasmid into R7A. B) An agarose gel of PCR products generated using primers 110L and 109CN3 from DNA extracts prepared from R7A, R7AΔintS and R7AΔrdfS (gel photo has been adapted for this thesis from published results (153)).

4.2.3 rdfS and intS are sufficient to mediate excision of an attP-containing plasmid

To investigate if intS and rdfS were the only ICEM1SymR7A genes required for efficient excision, excision of the integrated attP-intS-containing plasmid pJJ609 in R7ANS::pJJ609 was analysed using QPCR. QPCR of DNA extracted from stationary-phase broth cultures of R7ANS::pJJ609 revealed that excision was only observed sporadically in a low number of cells with attP being present in less than 0.01% of cells, and attB in less than 0.003%. To test if constitutive expression of intS induced excision of pJJ609, pJJ611 was introduced into R7ANS::pJJ609. QPCR analysis revealed that pJJ611 increased the average presence of to attP to 0.12%, but attB was detected on average at 9.1%, suggesting that some loss of the mini-island was occurring in these cells. To test if the constitutive expression of rdfS in trans induced excision of pJJ609, the rdfS coding sequence was amplified by PCR using primers 109CN5 and 109CN3 and cloned into pFAJ1708 to give pJR204. pJR204 was then introduced into R7ANS::pJJ609 cells and the cells were plated on medium containing tetracycline, with or without gentamicin. Colonies did not form on plates containing both antibiotics, but colonies were obtained on plates containing only
tetracycline. DNA was extracted from one of these isolates and analysed by PCR. The absence of a PCR product specific for attL (using primers LE2 and LE1) and presence of a PCR product for attB (using primers LE2 and RE1) showed that the strain had lost pJJ609 (identical to the result shown for R7ANS in Figure 4-9). Therefore ICEMISymR7A genes other than rdfS and intS are not required for island excision, confirming that RdfS is the symbiosis island RDF. Introduction of pJR202 into R7ANS::pJJ609 failed to induce increased excision (attP and attB were only observed sporadically at less than 0.02%), suggesting that expression of rdfS from its native promoter may require activation by other genes encoded on ICEMISymR7A.

4.3 Isolation of R7ANS

4.3.1 Constitutive expression of rdfS results in curing of ICEMISymR7A from R7A

The effect of constitutive expression of rdfS in trans on the excision of ICEMISymR7A was investigated by introducing pJR204 into strain R7A. Relatively few TcR colonies appeared following electroporation and the majority of these took 2 days longer to grow to 1-mm diameter on GRDM plates than colonies maintaining pFAJ1708 (which grew at the same rates as R7A colonies), suggesting that constitutive expression of rdfS caused a reduced growth rate. However a small number of colonies (≤3 per plate) grew at the same rate as R7A(pFAJ1708) colonies from a control electroporation. A similar result was observed when pJR204 was introduced into R7AΔrdfS. Introduction of pJR204 into R7AΔintS by either electroporation or conjugation resulted in only a small number of colonies that grew at the R7A rate and no slower-growing colonies. Since colonies with wild-type growth rate were in the minority in each electroporation with pJR204 (excluding R7AΔintS), the possibility that these cells contained mutated copies of pJR204 not expressing rdfS was tested. Primers FAJ5 and FAJ3 which flank the nptII promoter and MCS in pFAJ1708, were used to amplify DNA from six R7A(pJR204) isolates. PCR products were only able to be obtained from 2 out of the 6 isolates, while control PCR reactions for attL (using primers LE2 and LE1) produced products for all 6 (data not shown). The two PCR products amplified from pJR204 were then sequenced, which revealed that one of the isolates contained a deletion which removed part of the nptII promoter and the 5’ end of rdfS (data not shown), while the other contained a 12-bp in-frame deletion within the rdfS coding sequence (a.a. sequence of deletion is shown in bold in Figure 4-8).
The above results suggested that the introduction of an intact pJR204 caused a reduction in growth in wild-type or ΔrdfS cells, but was lethal in the R7AΔintS background. It seemed possible that the lethality of pJR204 in R7AΔintS cells was due to the inability of the strain to excise ICEMISymR7A and thus that the slow-growing R7A(pJR204) and R7AΔrdfS(pJR204) isolates were potentially cured of ICEMISymR7A. This hypothesis was tested by QPCR analysis of DNA isolated from four slow-growing isolates. Unoccupied attB sites were present on 100% of chromosomes ([attB] = [melR]) and attP was undetectable in all four isolates, suggesting that these isolates had lost ICEMISymR7A (Figure 4-9). One of these isolates was used in proceeding experiments and was renamed R7ANS(pJR204).

4.3.2 Characterisation of R7ANS

In order to find isolates of R7ANS(pJR204) that were cured of pJR204, the strain was passaged through 3 TY 64 h broth cultures. When this culture was spread on G/RDM without tetracycline, both slow and fast growing colonies were apparent (Figure 4-10). DNA was extracted from four of the fast-growing isolates and was then analysed by PCR using primers FAJ5 and FAJ3 (specific for pJR204), which gave a negative result for 3 out of the 4 colonies. These three isolates were found to be tetracycline-sensitive, further indicating they had lost pJR204. This suggests that the reduction in growth rate of R7ANS(pJR204) was due to the presence of pJR204 rather than the loss of ICEMISymR7A. DNA derived from one isolate (named R7ANS) was screened by PCR for the presence of ICEMISymR7A genes nodB, thiC (data not shown), rdfS, intS, attP and the ICEMISymR7A–chromosome junctions attL and attR. None of these assays produced PCR products, whereas PCR products of the expected size were obtained for fixJ (not shown), melR and attB amplifications (Figure 4-9). Phenotypic studies showed that R7ANS, in contrast to wild-type R7A, was unable to nodulate Lotus corniculatus or utilize succinate as a sole carbon source (data not shown), and was auxotrophic for thiamine, biotin (Figure 4-11) and nicotinate (data not shown). Genes essential for the synthesis of these three vitamins and the dct genes required for transport of C4-dicarboxylic acids are located on ICEMISymR7A (177, 180, 182). These results confirmed that R7ANS was cured of ICEMISymR7A.
Figure 4-9  QPCR and PCR assays used to confirm the isolation of the ICEMISymR7A-cured strain R7ANS.
a) QPCR assay of R7ANS(pJR204) demonstrating an absence of detectible attP DNA and that the concentration of melR (black lines) equals attB (blue lines), indicating ICEMISymR7A is not present (diagram shows results of triplicate reactions from a single sample). b) PCR analysis (amplicons labelled on right) of R7A and R7ANS showing an absence of ICEMISymR7A-specific DNA (gel photo has been adapted for this thesis from published results (153)).

Figure 4-10  Isolation of R7ANS cells cured of pJR204.
G/RDM agar plate showing the presence of fast and slow-growing isolates from R7ANS(pJR204) following passaging in serial TY broth cultures (3 x 64 h cultures). Fast growing colonies were found to have lost pJR204, indicating that this plasmid was responsible for the reduction in growth rate observed in R7ANS(pJR204).
Figure 4-11 Vitamin auxotrophy of R7ANS.
Phenotype of R7ANS and R7A on G/RDM or G/RDM lacking thiamine or biotin. Both strains grow similarly on media supplemented with all three vitamins, while R7ANS has a reduced growth rate and less-mucoid appearance plates missing either thiamine or biotin.
The development of a QPCR assay using Taqman probes during this work allowed the precise measurement of the proportion of cells in the population containing $\text{attP}$ and $\text{attB}$ sites. Using this technique, it was found that during liquid broth culture only a small proportion of cells in the population contained excised ICE$_{Ml}$Sym$^{R7A}$ and that the proportion increased roughly 100-fold in stationary-phase. Excision was dependent on a novel RDF encoded by $msi109$, which was renamed $rdfS$. Constitutive expression of $rdfS$ in trans cured strain R7A of ICE$_{Ml}$Sym$^{R7A}$, creating the new non-symbiotic strain R7ANS. Finally $rdfS$ and $intS$ were shown to be the only ICE$_{Ml}$Sym$^{R7A}$ genes required for excision of a plasmid containing $\text{attP}$.

The excision of ICE$_{Ml}$Sym$^{R7A}$ to form $\text{attP}$ was confirmed by QPCR to require IntS and its formation coincided with the formation of $\text{attB}$. The ratio of $\text{attP}$ to $\text{attB}$ was not significantly higher than 1:1, suggesting that ICE$_{Ml}$Sym$^{R7A}$ was not replicating above the level of chromosome copy number, but neither was it being lost from the population. This is consistent with the observed stability of ICE$_{Ml}$Sym$^{R7A}$ in laboratory cultures of R7A.

The QPCR assay gives an estimate of the total proportion of cells with excised islands at a particular point in time. The observation that ICE$_{Ml}$Sym$^{R7A}$ is only excised in a proportion of cells could suggest that excision is only transient and is occurring dynamically throughout the population, so that all cells are equally likely to have an excised ICE$_{Ml}$Sym$^{R7A}$ at any particular time. Immediate reintegration of ICE$_{Ml}$Sym$^{R7A}$ following excision (and likely transfer) could reduce the chance of ICE$_{Ml}$Sym$^{R7A}$ being lost through segregation during cell division, explaining its stability. Alternatively the QPCR data could be explained by sustained excision of ICE$_{Ml}$Sym$^{R7A}$ in only a small proportion of the population. If this were the case, then it would suggest that the island has evolved additional mechanisms to prevent its loss through segregation while it is excised in these cells.

The QPCR assay was used to investigate the function of $rdfS$; it was shown that deletion of $rdfS$ abolished both excision and transfer of ICE$_{Ml}$Sym$^{R7A}$ consistent with it encoding an RDF. Constitutive expression of $rdfS$ in R7ANS containing the integrated $intS$- and $\text{attP}$-encoding plasmid pJJ609 caused loss of pJJ609, demonstrating that $rdfS$ and $intS$ are the only ICE$_{Ml}$Sym$^{R7A}$-encoded genes required for excision. If the function of the RdfS is similar to that
of characterised RDFs, it is likely that it binds to one or several sites with attP and stimulates excision through DNA binding and bending within the attP-IntS nucleoprotein complex. On bacteriophage P4 for instance, the RDF Vis binds two sites found between the core and the arm-type repeats encoded upstream of int (30). It is also possible that other host-encoded proteins are involved in this reaction.

rdfS falls within a cluster of genes (msi110 - msi106) on ICEMISymR7A (182). Sullivan et al. (2002) previously commented that this region is likely to be involved in transfer due to the similarity of Msi108 to the TrbC protease TraF and the similarity of Msi107 to murein lytic transglycosylases found associated with some T4SS (182). Parallel experiments carried out during this work (197) revealed that the msi106 gene encodes a DNA relaxase (gene renamed rlxS) (153) and that the intergenic region between rlxS and msi107 when cloned is able to mobilise transfer of a plasmid from M. loti R7A to E. coli, indicating that this region contains the ICEMISymR7A origin of transfer (oriT) (197). The plasmid pJR202 containing rdfS and upstream intergenic DNA complemented excision in strain R7A∆rdfS suggesting that the rdfS promoter was immediately upstream of the gene. The 166-bp intergenic region between msi110 and rdfS provides sufficient space for a promoter. In contrast, the rdfS stop codon overlaps with the traF start codon, indicating that rdfS is the first gene in the rdfS-traF-msi107 operon and hence is co-regulated with genes required for conjugative transfer. This provides an elegant mechanism by which ICEMISymR7A excision and transfer may be coordinately regulated.

The position of rdfS upstream of Mpf and Dtr genes is to the author’s knowledge a unique position for an RDF gene. RDFs are commonly encoded adjacent or even overlapping with the cognate integrase gene and most evidence suggests that the two genes closely coevolve. The alignment in Figure 4-6 shows that the RDFs of P4, HPI and she share only 11% a.a. (out of 64 a.a.) with RdfS, while their cognate integrases share 25% a.a. (out of 390 a.a., data not shown) with IntS. In contrast, an alignment of the P4, HPI and she RDFs alone reveals that they share 24% a.a. (out of 64 a.a., data not shown). This demonstrates that RdfS has diverged significantly from characterised RDFs of elements encoding P4-like integrases. This could suggest that RdfS may have adapted to functions required by ICEMISymR7A through coevolution with the transfer system, or it may simply be that there are more synonymous a.a. substitutions available in RDFs than in integrases. Another possibility is that RdfS has coevolved primarily with its DNA-binding site rather than with the IntS protein. A similar divergence of RDFs from their cognate
integrase families has been observed for the LC3 family of integrases, which are often associated separately with two distinct families of RDFs (Tn5276 and L54a) which lack almost any similarity (105).

A model of excision and transfer of ICEMISym$^{R7A}$ is presented in Figure 4-12. In summary, activation of transfer is likely associated with induction of expression of the rdfS-rlxS and trb operons. RdfS in concert with IntS stimulates excision while RlxS, TraG and the T4SS couple excision with conjugation. Upon entry into a new cell it is likely that RlxS mediates recircularisation and initiates RC replication of ICEMISym$^{R7A}$ ssDNA to form dsDNA, after which strong expression of intS from attP drives integration into the genome.

The result that ICEMISym$^{R7A}$ excision is increased in stationary-phase cultures suggests that transfer may be under growth-phase or cell-density dependent regulation. Several bacteriophages show increased rates of transition to lytic phase during entry into stationary phase, usually reflecting host stress such as the SOS response (111, 151, 160). Another interesting possibility is that the increase in ICEMISym$^{R7A}$ excision is a cell-density response mediated by QS. As mentioned previously, ICEMISym$^{R7A}$ encodes a LuxR homologue and two LuxI homologues.

The loss of ICEMISym$^{R7A}$ upon introduction of pJR204 may be due to RdfS acting to force excisive integrase-mediated recombination and to inhibit integration, resulting in loss of the island in the absence of expression of genes required for rolling-circle replication. Unexpectedly, constitutive expression of rdfS was lethal in a ΔintS strain and caused a reduction in growth rate of the non-symbiotic strain R7ANS. RDFs such as the Cox protein of bacteriophage HP1 and the Vis protein of bacteriophage P4 also act as transcriptional regulators (51, 52, 146, 161). In both these systems the RDF is a repressor; in HP1 Cox represses genes involved in lysogeny, however in P4 Vis represses IntS expression both transcriptionally and post-transcriptionally, by binding to both attP and intS mRNA respectively (145). Hence it is possible that rdfS may have a regulatory role additional to its role as a RDF and constitutive expression may lead to the expression or repression of ICEMISym$^{R7A}$ and/or chromosomal genes which leads to this effect.
Figure 4-12 Model of ICEMIsym$^{R7A}$ excision, replication and transfer.
Attachment sites attP, attB, attL and attR are shown by black bars and DNA is shown by a single line. The broken line labelled oriT represents the initiation of RC replication. Arrows represents various transitions in the process of excision, replication and transfer.

In summary, ICEMIsym$^{R7A}$ requires the expression of the novel RDF encoded by rdfS, which along with IntS is sufficient to mediate excision in the absence of other ICEMIsym$^{R7A}$ genes. It is likely that the induction of transfer involves expression of the rdfS-rlxS and trb operons, resulting in the coordination of ICEMIsym$^{R7A}$ excision with expression of Dtr and Mpf functions. Further work is needed to unravel this regulatory circuit, including investigating the possible additional role of RdfS as a transcriptional regulator. As RdfS is highly likely to bind attP, DNA footprinting experiments involving RdfS would not only further knowledge about the IntS-attP-RdfS complex but also enable putative identification of further RdfS DNA targets. The proportion of cells containing an excised ICEMIsym$^{R7A}$ increased in stationary-phase cultures, suggesting a growth-phase or population-density dependency of excision. The QS homologue-encoding genes traR, tral2 and tral1 are obvious candidates genes for investigation into this effect.
Chapter 5
Bioinformatics of ICEMLSym^{R_{7A}}-related Genomic islands
The *msi110-rlxS* cluster shows similarities to genes found on the conjugative transposon Tn4371 of *Ralstonia oxalatica* (species recently renamed *Cupriavidus oxalaticus*) and the plasmid pMlb of *M. loti* MAFF303099 (182). A comparative analysis of Tn4371-related elements with the symbiosis island of *M. loti* MAFF303099 by Toussaint et al. (185) published in 2003 revealed similarity of several Tn4371-genes to *M. loti* MAFF303099 genes that correspond to *msi110*, *rdfS*, *traF*, *msi107* and *rlxS* on ICEMISymR7A. A second region common to both ICEMISymR7A and Tn4371 is the *trbBCDEJLFGI* operon encoding a T4SS that is unique in that it is missing *trbK* and *trbH*, while encoding two hypothetical genes *msi021* and *msi031* which flank the cluster (182). These observations suggest that these gene clusters are part of a conserved transfer system sharing a common ancestor with that of ICEMISymR7A.

In this chapter a bioinformatic analysis revealed that genes encoded in the *msi110-rlxS* cluster are conserved in 40 putative genomic islands found in a variety of proteobacteria. In a second analysis, two additional conserved gene clusters were identified on ICEMISymR7A which include genes distinct from those found in the Tn4371 family of elements. These genes were used to identify and annotate 26 new genomic clusters found in 13 α-proteobacterial species, representing a new family of putative genomic islands that we named the ICESym family. Finally, bioinformatic analyses and review of both the genetic organisation and content of the conserved ICESym gene set are presented.
5.1 Results and Discussion

5.1.1 Conservation of rdfS on putative genomic islands

Potential orthologues of RdfS revealed in a BLASTP search (conducted in June 2006) were screened for ORFs with similar genetic context to that found in ICEMISym$^{R7A}$. Nucleotide sources of the 50 matching protein sequences with an E-value less than $10^{-3}$ were examined for the presence of homologues of $traF$ and $rlxS$ within a 6-kb region. Homologues of both genes were found downstream of and in the same orientation as the rdfS orthologue in 40 cases. Of the remaining ten excluded proteins, six were part of contigs from incomplete genomes that contained homologues of conserved hypotheticals found in either ICEMISym$^{R7A}$ or Tn4371, while two homologues were found directly downstream of predicted P4-family integrases (gi:99078558, gi:84687547).

The 40 RdfS orthologues were aligned using ClustalW and the alignment was used to create a bootstrapped neighbour-joining tree (Figure 5-1). The tree grouped RdfS with sequences from the North Atlantic Ocean isolate Sphingomonas SKA58 and Sargasso Sea isolate Parvularcule bermudensis HTCC2503 and these sequences formed a distinct clade. RdfS orthologues from the putative genomic islands previously identified by their similarity to the conjugative transposon Tn4371 (185) formed a tight cluster together with several new additions (Figure 5-1). The group of sequences ranged in length from 75-123 a.a. (average 93 a.a.) and all had a predicted basic isoelectric point (average 10.70 ± 0.72), consistent with the properties of other characterized and predicted RDFs.
Neighbour-joining tree based on a ClustalW amino-acid sequence alignment of RdfS homologues. The percentage of trees supporting each branch are indicated (1000 trees total).

5.1.2 Identification of conserved ICEMl/Sym\textsuperscript{R7A}-family genomic island backbone

The similarities of the rdfS, msi106 and msi031-trbBCDEJFGI-msi021 genes to those found on Tn4371 prompted a search for other homologues of Tn4371 genes on ICEMl/Sym\textsuperscript{R7A}. BLASTP searches were carried out for each annotated a.a. sequence from Tn4371 (185) on the ICEMl/Sym\textsuperscript{R7A} sequence (1). In addition to previously identified homologues this search revealed homologues of the Tn4371 proteins RO00010, RO00014, RO00033 and RO00030 to proteins encoded by msi151 (60% a.a. ID), msi150 (27% a.a. ID), msi151 (second copy 46% a.a. ID) and traG (69% a.a. ID) respectively as well as a single gene RO00034 that showed similarity to both msi172 and msi171 (36% and 40% a.a. ID respectively) in a concatenated form. The same
analysis was then applied to genes on the \textit{M. loti} MAFF303099 plasmid pMlb, which revealed the presence of genes with similarity to \textit{ ardC, msi150, msi172, msi171, msi110} (two copies), \textit{ rdfS, msi107, rlxS} and \textit{ traG} along with previously identified homologues of the \textit{trbBCDEJLFGI} cluster and \textit{ msi021}. A comparison of the genetic organisation of the ICE\textit{MISym}^{R7A} homologues on Tn4371 and pMlb is presented in Figure 5-2.

Following the original BLASTP search and analysis of \textit{ rdfS} homologues detailed in section 1.1, several new matches to \textit{ rdfS} appeared in BLAST searches of the NCBI database. At the present time of writing (October 2007), the closest match to the RdfS sequence was YP_001413092 (77\% a.a. ID) of \textit{Parvibaculum lavamentivorans} DS-1\textsuperscript{T}, an \textalpha{}-proteobacterium capable of degrading the commercial surfactant linear alkylbenzenesulfonate (165). The genomic sequence of this strain was analysed for the presence of ICE\textit{MISym}^{R7A} homologues. Amazingly, homologues of \textit{ msi170, msi169, ardC, msi150, msi110-msi106, traG, msi031, trbBCDEJLFGI} and \textit{ msi021} were found as a single contiguous segment, syntenic with ICE\textit{MISym}^{R7A} except for the absence of \textit{msi172-msi171}, the absence of \textit{msi151} between \textit{ ardC} and \textit{msi150} and the presence of a second \textit{msi169} homologue adjacent to the first. Suspecting that homologues of \textit{msi172, msi171} and \textit{msi151} remained unannotated on \textit{P. lavamentivorans} DS-1\textsuperscript{T}, tBLASTX searches were carried out using the sequence syntenic to \textit{msi172-msi171} and \textit{msi151}. This revealed the presence of unannotated \textit{msi172} and \textit{msi171} homologues directly upstream of the \textit{msi170} homologue; however no \textit{msi151} homologue was identified. In a parallel analysis carried out by Clive Ronson (personal communication), an unannotated \textit{msi169} homologue (named \textit{msi168a}) was identified on ICE\textit{MISym}^{R7A}, encoded downstream of \textit{msi169} in a pattern similar to that found on \textit{P. lavamentivorans} DS-1\textsuperscript{T} and \textit{M. loti} MAFF303099. This large region of homology in \textit{P. lavamentivorans} DS-1\textsuperscript{T} therefore appeared to further delineate a core gene-set common to this family of genomic islands (hereafter referred to as ICE\textit{Sym}) present as three gene clusters on ICE\textit{MISym}^{R7A}. 

97
Figure 5-2  A comparison of the genetic organisation of ICEM!SymR7A homologues found on the Ralstonia oxalatica (Cupriavidus oxalaticus) conjugative transposon Tn4371 and the Mesorhizobium loti MAFF303099 plasmid pMlb.

Genes are shown as arrow symbols and are to scale; colours specific for each ICEM!SymR7A gene are used to indicate genes on Tn4371 and pMlb that encode similar proteins (white indicates no homology). Coordinates of ICEM!SymR7A refer to positions annotated in Sullivan et al (2002) (182) and coordinates of Tn4371 and pMlb refer to the annotated database sequences available from NCBI (accession numbers AJ536756 and AP003017 respectively). Relative positions of msi021, msi031 and the trbBCDEJLFGI (trb) cluster are indicated between < and > symbols.
In order to identify further genomic islands of the ICESym family, a search for the combined presence of the *msi172-msi169* and *msi110-msi106* clusters was carried out using the presence of genes encoding the RdfS and Msi170 a.a. sequences as markers for these loci. The Msi170 protein sequence was chosen as a marker for the *msi172-msi169* cluster as it produced a large number of database hits in BLASTP searches and also because it was absent from Tn4371 and so possibly specific to closer relatives of ICE*MLISym*<sup>R7A</sup>. PSIBLAST searches of Msi170 and RdfS sequences were carried out on a database of microbial genomes (4) (including unfinished sequences) and run through four iterations with default sequences and settings selected. Sequences from strains containing matches to both sequences were then collected. A total of 86 Msi170 and 59 RdfS homologues were tallied in 33 species with a total of 55 Msi170-RdfS pairs (i.e. a copy of an Msi170 homologue was present for each RdfS homologue). Upon further inspection of the matches to the Msi170 sequence it became apparent that many were proteins with closer similarity to Msi171 or a concatenated Msi172-Msi171 protein similar to that found on Tn4371. The significance of this is addressed further in section 5.1.3.6. Removal of these spurious matches resulted in a final tally of 37 proteins with similarity to Msi170 and a total of 30 Msi170-RdfS pairs. Interestingly none of the RdfS sequences from these Msi170-RdfS pairs grouped with the large clade of sequences encoded by the Tn4371-like elements in Figure 5-1. Instead they grouped with the ICE*MLISym*<sup>R7A</sup> RdfS sequence or the second main group of sequences, thus indicating a common divergence of these RdfS homologues from the Tn4371-encoded sequences.
The genomic sequences corresponding to the 29 newly identified Msi170-RdfS pairs were then manually inspected for the presence of nearby genes encoding proteins similar to Msi172, Msi171, Msi169, Msi168a, ArdC, Msi151, Msi150, Msi110, TraF, Msi107, RlxS, TraG, Msi031, TrbBCDEJLFGI and Msi021. This resulted in the discovery of 25 additional gene clusters resembling those on ICE/MISymR7A in 12 α-proteobacterial species. Of the remaining 4 Msi170-RdfS pairs, two were from unfinished genomes at the end of contigs that resembled ICE/MISymR7A clusters (*Caulobacter* sp. K31 and *Magnetospirillum magnetotacticum* MS-1), one was located in an unrelated position (*Nitrobacter* sp. Nb-311A) and a final homologue was within a ICE/MISymR7A-like cluster, but an abundance of transposon related sequences made annotation too difficult (*Nitrobacter winogradskyi* Nb-255). All 28 ICESym clusters covering 13 species were annotated and are presented in Figure 5-3. In regions of synteny to ICE/MISymR7A that lacked obvious homologues (e.g. Msi172, Msi171 and Msi169 homologues were often unannotated but present adjacent to Msi170) tBLASTx searches were carried out of the regions against ICE/MISymR7A and the genes were then annotated.
Figure 5-3 Comparison of the genetic organisation of 28 ICESym clusters

Abbreviations used for each cluster in this thesis are shown in bold to the right of each cluster, while species names and accession numbers are shown in bold to the left of each cluster. Genes are shown as arrow symbols and are to scale; colours specific for each ICESymR7A gene are used to indicate genes that encode similar proteins in other clusters. Black indicates the presence of a recognized transposon-related sequence while white indicates a lack of homology to any ICESymR7A gene. Coordinates (numbers flanking clusters) of ICESymR7A refer to positions annotated in Sullivan et al (2002) (182) and coordinates of other clusters refer to positions within the annotated database sequences available from NCBI. Slanted double lines indicate where the maps have been abridged and display the length of the DNA segment removed. Relative positions of msi021, msi031 and the trbBCDEJLFGI (trb) cluster are indicated between < and > symbols. An asterisk indicates genes were annotated during this study. Yellow lollipops indicate the presence of a putative cis-acting sequence identified in this study while purple lollipops indicate putative oriTs. Numbers under Msi169 (Xre-family) homologues (yellow arrows) indicate which of the phylogenetic groups (defined in Figure 5-7) each sequence belongs to.
5.1.3 Comparative analysis of ICESym-family genes

The 28 ICESym clusters presented in Figure 5-3 demonstrate a considerable conservation of genetic content and structural organisation as well as areas of apparent diversification. The \( \text{traG msi031-trbBCDEILFGI-msi021} \) gene cluster was invariably present nearby to other ICESym genes, downstream of \( \text{rlxS} \) in all cases except in \( \text{Sphingomonas sp. SKA56} \) where it was encoded 5’ of \( \text{msi172} \). The \( \text{rdfS} \) homologues were consistently found in the same orientation and likely in the same operon as homologues of \( \text{msi107, traF and rlxS} \), reaffirming the close association and likely co-regulation of the expression of \( \text{rdfS} \) with transfer-related genes in ICESym clusters. The presence of the \( \text{luxR and luxI} \) homologues \( \text{traR (msi174)} \) and \( \text{traI2 (msi173)} \) appeared to be restricted to ICE\( \text{Ml} \)Sym\( \text{R7A} \), ICE\( \text{Ml} \)Sym\( \text{MAFF} \) and ICESymS.SKA58.1. The similarity of the protein sequences encoded by the \( \text{traR, traI2 and msi170} \) homologues in the SKA58 cluster with the ICE\( \text{Ml} \)Sym\( \text{R7A} \) proteins was low for all three gene products (27%, 26% and 28% a.a. ID respectively), so it is unclear whether or not the positioning of the \( \text{luxRI} \) homologues in both clusters is the result of a single ancestral event.

The conservation of an \( \text{msi169} \) homologue adjacent to \( \text{msi170} \) on all clusters was an especially intriguing feature. A BLASTP search of the Msi169 protein sequence revealed it was related to Xre family of transcriptional regulators and in particular showed similarity to several C proteins of RM cassettes (e.g. C.PvuII and C.AhdI). Also closely linked with the \( \text{msi170-msi169} \) genes were homologues of \( \text{msi172 and/or msi171} \), always encoded in that order. Homologues of \( \text{msi110} \) were found on all clusters, often within the \( \text{rdfS-rlxS} \) cluster and/or adjacent to the \( \text{msi172-msi169} \) cluster.

Contrastingly the conservation of \( \text{ardC, msi151 and msi150} \) was relatively low, with these genes only being found in 16, 8 and 21 of the 28 ICESym clusters respectively. This could indicate these genes have non-essential roles or that their roles are specific to the elements they are encoded on. Alternatively homologues may have been missed in during this annotation, which is likely as these genes in particular were often encoded a long distance from the other clusters.

The ICESym clusters identified in this section (excluding the pMlb cluster) were all chromosomally-located and are potentially integrated and possibly mobile genomic islands. A BLASTP search for a.a. sequences similar to IntS revealed the presence of homologues in all strains containing ICESym clusters. Indeed there were many more IntS-related sequences per
strain than ICESym clusters. The four ICESym clusters found in *Bradyrhizobium* BTAi1 during this study have been previously identified as genomic islands through GC-content and comparative sequence analyses (63). These regions correspond to island numbers 10, 6, 29 and 28 (designated in reference (63), indicated in Figure 5-3) and encode genes implicated in a variety of features such as type I secretion, multidrug and metal cation efflux, chemotaxis, aromatic compound degradation and various other metabolic processes (63).

### 5.1.3.1 *traF, msi107, ardC, msi031 and msi021*

The likely functions of *traF, msi107* and *ardC* have been discussed previously by Sullivan *et al.* (182). The *traF* gene likely encodes an enzyme with a protease/peptidyl-transferase activity which catalyses cyclisation of the pilin subunit TrbC of the T4SS and is essential for conjugation (50, 73). The *msi107* gene likely encodes a lytic transglycosylase (muramidase) capable of local peptidoglycan degradation (201). The *ardC* gene encodes a protein with similarity to the antirestriction protein ArdC of the plasmid pSa. This protein binds ssDNA and protects against ssDNA restriction and is possibly transported with DNA through the mating pore into the recipient bacterium (18).

The T4SS-encoding operon *trbBCDEJLFGI* on ICE*Ml*SymR7A is unique in that it lacks *trbK* and *trbH* but encodes two hypothetical genes *msi031* and *msi021* which flank the cluster (182). For transfer of the *Agrobacterium tumefaciens* strain C58 Ti plasmid, *trbK* appears to be non-essential but has a role in entry exclusion of closely related plasmids, while mutation of *trbH* appears cause a reduction in transfer efficiency (106). A BLASTP search of the 146 a.a. sequence of Msi031 produced a large number of matches to proteins annotated as putative CopG-like DNA-binding proteins. The CopG protein (accession number P13920) is a 45 a.a DNA-binding protein containing a ribbon-helix-helix domain (64) and controls copy number of the streptococcal rolling-circle plasmid pMV158 by repressing its own expression and that of the *repB* gene (41). A direct alignment of CopG with Msi031 showed Msi031 shared 22% a.a. ID over the length of CopG with the N-terminal portion of Msi031. This could suggest that Msi031 has a similar role in the regulation of the *trb* operon or other aspects of transfer. A BLASTP analysis of the Msi021 sequence revealed that this protein was highly conserved but lacked similarity to any characterised proteins.
5.1.3.2 rlxS and the ICEMlSymR7A origin of transfer (oriT)

The ICEMlSymR7A gene msi106 (rlxS) was originally annotated as encoding a possible relaxase required for initiation of the rolling-circle replication involved in conjugative transfer, based on very limited similarity (25% identity over 22% of Msi106; E = 0.16) of its gene product to VirD2 from A. tumefaciens (139, 182), the relaxase required for T-DNA transfer (139). Subsequent experiments carried out by Gabriella Stuart have shown that mutation of rlxS abolishes transfer of ICEMlSymR7A (153). Furthermore, cloning of the intergenic DNA upstream of rlxS into a normally non-mobile plasmid enables the transfer of the plasmid from M. loti R7A to E. coli by conjugation and this requires a functional copy of rlxS (197). The position of the oriT within this region was suspected because many known plasmid oriTs are located adjacent to the cognate relaxase gene (141).

oriT DNA regions are composed of a short, highly conserved core region and an adjacent inverted repeat. Only the core-side arm of the inverted repeat sequence is required for transfer (141) (Figure 5-4). Nicking of the core region by the relaxase results in the covalent attachment of the protein to the 5'-end of the DNA adjacent to the core sequence (140, 141). A search for a conserved sequence resembling an oriT was carried out within the intergenic sequences located upstream of the rlxS homologues on the ICESym clusters. The 28 sequences were extracted and submitted to the MEME motif-finding program (14). A highly conserved 22-bp sequence resembling a core region was found encoded in the same orientation on all sequences, and is presented as a consensus logo in Figure 5-5. The positions of the oriT in each sequence are indicated in Figure 5-3. Directly adjacent to this region on the side closer to the rlxS gene a poorly conserved AT-rich region was identified by the MEME program; however a conserved inverted repeat was not obvious (data not shown). High AT content within oriT sequences is common and has been suggested to aid in local denaturation of the oriT DNA prior to nicking (100, 141). It is common for oriT sequences of the same family to contain a highly conserved core sequence while having unrelated inverted-repeat sequences, and this is believed provide specificity of the relaxase for its cognate oriT and/or binding of specific accessory factors (141) (Figure 5-4). Manual examination of the rlxS upstream sequence alone revealed the presence of a 19-bp imperfect inverted repeat, separated by a 4-bp spacer and overlapping the previously identified core-like sequence by 13 bp (Figure 5-5).
Figure 5-4 Structure of various oriT sequences belonging to the R1162 family.

Taken from Parker et al. (141).
Figure 5-5 Conserved motifs within the ICE\(\text{Ml}\text{Sym}^{R7A}\) ori\(T\).

a) DNA logo consensus diagram (created using Weblogo (38)) of the conserved 22-bp sequence discovered using the MEME motif analysis of sequences from the 28 ICESym clusters.  

b) A diagram of the ICE\(\text{Ml}\text{Sym}^{R7A}\) ori\(T\) aligned with the same region on the ICE\(\text{Ml}\text{Sym}^{\text{MAFF}}\) sequence. The inverted repeat is indicated by arrows (consensus of the repeat shown above) and the 22-bp sequence from (a) is shown in bold. The presence of a transposon sequence in ICE\(\text{Ml}\text{Sym}^{\text{MAFF}}\) is indicated by italics.

It seemed likely that the conserved 22-bp sequence identified above represents the ori\(T\) core site which is bound to and nicked by RlxS. In other ori\(T\) sequences examined the nick site of the relaxase is generally at the end of the core region most distant from the inverted repeat (141). Interestingly, this precise region on the \(M.\text{loti}\) MAFF303099 symbiosis island contains a transposon insertion (also commented on in Sullivan \textit{et al} 2002 (182)) indicating it may be inactive in transfer (Figure 5-5).

The position of the putative nick site suggests that the right end of the symbiosis island (coordinates 133.5-501.8 kb) would be transferred first during conjugation, followed by att\(P\) on the circularised island and then the left end of the ICE\(\text{Ml}\text{Sym}^{R7A}\) (coordinates 0.0-133.5 kb). This is consistent with the observed bias for mating pore genes (\(trb\)) to be encoded towards the end of the transferred strand (100), as the \(trb\) operon is encoded to the left of \(rlxS\) on ICE\(\text{Ml}\text{Sym}^{R7A}\) and almost all ICESym clusters (Figure 5-3). In experiments documented in section 3.1.2, transfer of ICE\(\text{Ml}\text{Sym}^{R7A}\) from R7A\(\Delta\text{intS}\) to N18(pFAJ1700) was shown to be highly reduced compared to transfer from R7A, but was not totally abolished. In a follow-up experiment carried out by Gabriella Stuart (unpublished data) this was shown to be due to
chromosomal transfer from the ICEMISymR7A oriT in an Hfr-like manner. PCR analysis of DNA extracted from ex-conjugants from these experiments all contained attR and attB and sequences 3’ of and often including msi107 (but not 5’ of msi107), indicating transfer was likely occurring from the oriT in the direction of attR, as predicted by the above bioinformatic analysis.

5.1.3.3 msi110

A BLASTP search of Msi110 revealed it to be a widely conserved hypothetical protein found on all ICESym clusters and Tn4371-related elements. A conserved-domain search revealed membership in the COG5489 group of conserved proteins of unknown function. Interestingly another member of this group (NP_396633.3) was found on the Ti-plasmid of Agrobacterium tumefaciens strain C58, and encoded nearby were proteins with similarity to ICEMISymR7A proteins ArdC, Msi150 and TraG. This suggests this gene has a conserved role in a variety of transfer systems.

The ICESym clusters had one or two copies of the msi110 gene, encoded next to the msi172-msi171 and/or rdfS-rlxS regions. The presence of msi110 homologues in these two distinct positions could suggest that separate elements evolved with the msi110 gene in each position and have in some cases recombined to form a single element.

5.1.3.4 msi150

A search of the protein family hidden Markov model database (6) revealed that the N-terminal portion of Msi150 (first 186 of 576 a.a.) contains the ParB partitioning protein domain (TIGR00180) common to plasmid and chromosome partitioning proteins (E = 2.5 x 10^{-11}). These proteins segregate plasmids during cell division by binding and spreading along centromere-like parS sites on plasmid DNA. Plasmid pairing then forms through multiple parS-ParB dimer complexes that are then separated and directed to daughter cells through interaction with the Walker-A type ATPase ParA (for a review see ref. (167)). Msi150 differs in its C-terminal region to characterised ParB proteins in that it is noticeably longer; however a similar protein was also found on Tn4371 (RO00014, Figure 5-2) (note Toussaint et al (2003) (185) incorrectly annotated RO00038 as ParB), along with proteins with similarity to the ParA and the DNA helicase RepA (similar to that encoded by the pVS1 replicon). During annotation of the ICESym clusters the ParA and RepA proteins were commonly observed near the rdfS-rlxS region on
clusters other than ICEMISym<sup>R7A</sup> (data not shown). This conservation of these genes could suggest the ICESym family have arisen from an ancient plasmid ancestor and may continue to have a role in stability, maintenance and regulation of these elements. For instance the Msi150 and plasmid-related genes in other ICESym clusters could have a role in partitioning or replication during ICEMISym<sup>R7A</sup> excision. Alternatively these proteins could be involved in conjugation; recently Atmakuri et al. (12) demonstrated that the ParA-like VirC1 protein of the Agrobacterium tumefaciens Ti plasmid actively recruits the VirD2 relaxase and T-DNA to the cell poles and the T4SS. Other ParB homologues such as KorB have been shown to have roles as transcriptional repressors that bind and spread in a cooperative manner, repressing transcription over large regions of DNA (21).

5.1.3.5 <i>msi151</i>

<i>msi151</i> encodes a conserved hypothetical 138-a.a. protein that has no significant similarity to any characterised proteins. A BLASTP search resulted in matches to many proteins annotated as having similarity to a variety of proteins including a class-2 dihyroorotate dehydrogenase and COG0608 which includes proteins with similarity to single-stranded DNA exonucleases. However upon further investigation no convincing similarity to either of these groups was found.

5.1.3.6 <i>msi172, msi171 and msi170</i>

The hypothetical ORFs <i>msi172</i> and <i>msi171</i> (COG5419) were found to be poorly conserved at the a.a. sequence level (data not shown), although they were present in various forms on the ICESym clusters and Tn<sub>4371</sub>-family elements. On Tn<sub>4371</sub> and all related elements identified by Toussaint et al. (185), <i>msi172</i> and <i>msi171</i> are present as a single ORF (RO00034) (with the exception of ICEMISym<sup>MAFF</sup> which resembles ICEMISym<sup>R7A</sup>). All of the ICESym clusters encoded either or both <i>msi172</i> and <i>msi171</i> as two separate ORFS except for ICESymS.SKA58.1. During the process of annotation the <i>msi172</i> and <i>msi171</i> genes were observed to have poorly conserved start sites and so it is possible that further <i>msi172</i> and <i>msi171</i> homologues were present on these clusters but were not annotated during this analysis. These genes presented a peculiar case as they appeared to be very poorly conserved but were nonetheless found in a similar position in each cluster and always encoded in the same order as the concatenated protein encoded by RO00034 in Tn<sub>4371</sub>. 
The hypothetical ORF *msi170* was found next to *msi171* and/or *msi172* in all but two ICESym clusters and was in every case encoded adjacent to an Xre-family gene. There appears to be no direct orthologue of *msi170* on Tn4371 or related elements. However as previously mentioned, during the PSIBLAST search for sequences similar to Msi170, more than half of the matches retrieved (46 from 83 in strains that also had RdfS) were to proteins with more similarity to the product of *msi171* or Tn4371 RO00034-like genes. An alignment of RO00034, *msi170*, *msi172* and *msi171* products is presented in Figure 5-6, demonstrating the similarity between the four proteins. The weak similarity of Msi170 to Msi171 (19 % a.a. identity over 74 a.a.) and the absence of an *msi170* orthologue on Tn4371 may suggest that *msi170* is paralogous to *msi171* and could have evolved a similar but differentiated function.
Figure 5-6  Amino acid similarity between Msi172, Msi171, Msi170 and the Tn4371 protein RO00034.
A ClustalW alignment of the RO00034 protein sequence from the conjugative transposon Tn4371 with the conserved ICESym-encoded proteins Msi172, Msi171 and Msi170. Conserved and/or similar a.a. are shown in white text on a black background.
5.1.3.7 *msi169* and Xre-related transcriptional regulators

As mentioned above, the *msi169* gene encodes an Xre-family transcriptional regulator protein similar to the C proteins of RM cassettes. Members of this family including Xre, \( \lambda \)-cI and C proteins such as C.AhdI (120), C.PvuII (96, 119), C.BamHI (23) and C.EcoRV (169) positively regulate their own expression by binding to operator sequences found upstream of the genes encoding them. Transcriptional activation is believed to involve direct interaction with \( \sigma^{70} \) in a mechanism similar to that of \( \lambda \) cI protein. In the case of RM systems, the C protein is usually encoded upstream of and co-regulated with the expression of the restriction endonuclease (REase). This configuration allows a delayed expression of the REase during entry of the RM cassette into a new host, as the initial absence of C-protein results in very low expression of both C and the REase, allowing initial expression of the MTase before the expression of the REase. The genes encoding the MTase and C proteins are often divergently orientated and have overlapping promoter regions (23, 169, 188). It has been proposed that this allows the simultaneous repression of MTase expression and activation of C and REase expression by the C protein (169) but to this author’s knowledge this has never been experimentally demonstrated. The close association of genes encoding Xre-family proteins with the *msi170* homologues on all ICESym clusters closely resembles the organisation of the REase and genes found in many RM cassettes. Indeed 18 of the 28 ICESym clusters contain divergent *msi169* and *msi170* homologues (others are in tandem with a small gap between the two sequences), which may indicate that these gene-pairs are regulated in a similar way to the RM cassettes.

During the annotation of the Xre-family genes on the ICESym clusters, it was apparent that the similarity between the various Xre protein sequences varied greatly, and that this variation did not appear congruent with the presumed relatedness of the elements. For instance, the ICE*M*Sym\textsuperscript{R7A}, ICE*M*Sym\textsuperscript{MAFF} and ICESymPl.DS-1.1 clusters show very similar genetic organisation and encode two Xre homologues in tandem. A comparison of the a.a. sequences of these proteins revealed that Msi169 sequence was almost identical to the protein encoded on ICE*M*Sym\textsuperscript{MAFF} (95% a.a. ID) but only 35% identical to the protein encoded by the syntenic gene on ICESymPl.DS-1.1. More surprisingly however was that the Msi168a sequence (the second Xre protein on ICE*M*Sym\textsuperscript{R7A}) was far more similar to the protein encoded in the syntenic position in ICESymPl.DS-1.1 (68% a.a. ID) than it is to that encoded on ICE*M*Sym\textsuperscript{MAFF} (15% a.a. ID). To gain a clearer picture of the relatedness of the Xre proteins encoded on all ICESym
clusters the 31 a.a. sequences were aligned using ClustalW and this alignment was used to create a bootstrapped neighbour-joining tree, rooted using the C.AhdI sequence as an outgroup (Figure 5-7). This tree clustered the sequences into 6 distinct groups (numbered 1 to 6). Membership of each sequence to each group is indicated in Figure 5-3. Interestingly the 9 members of group 5 are the only $xre$ genes encoded in the same orientation as $msi170$ with the exception of the $xre$ on ICESymPd.PD1222.1.

This analysis again indicated incongruence in the genetic organisation of each cluster with the relatedness of the Xre proteins. For instance, ICESymXaPy2.2 and ICESymXaPy2.3 show almost identical gene content and organisation, but encode Xre proteins from different groups. Conversely ICESymB.BTAi1.6 and ICESymPd.PD1222.1 both show distinct genetic organisation patterns compared to ICEMISymR7A but encode Xre proteins in the same group as Msi169. For a more thorough comparison, a ClustalW alignment of all 28 Msi170 homologues was carried out and used to create a bootstrapped neighbour-joining tree. This tree again demonstrated that with the exception of the tree’s extremities, the Xre-family groups identified do not generally co-segregate with the Msi170 tree branches. This may suggest that several recombinational events have occurred between several of these or similar elements to separate the $xre$-$msi170$ pairs or that these $xre$ genes may have been disseminated by some unknown mechanism specific for these genes.
Figure 5-7  Distinct Xre-families conserved on ICESym clusters.

A ClustalW alignment of the Xre-family proteins encoded on the ICESym clusters was used to created a neighbour-joining tree (bootstrapped 1000 times, percentages of trees supporting each branch are shown if >80%). The C.AhdI sequence was used as an outgroup. Distinct family groups are numbered 1-6 and alongside each group is an alignment of the conserved DNA motif found upstream of each gene (asterisks indicate identical nucleotides). Proteins from elements with two XRE genes are labelled as XRE1 or XRE2 based on the order that they are encoded.
The upstream intergenic sequences of each *xre* gene on the ICESym clusters were aligned using ClustalW. This failed to produce any significant consensus sequence so the sequences were split into groups 1-6 based on the phylogeny of the protein sequences inferred in the previous analysis.
Separate alignments of these sequences revealed the presence of highly-conserved DNA motifs upstream of each \textit{xre} gene that were specific for each phylogenetic group, with the exception of the \textit{xre} gene from cluster 1 of \textit{Paracoccus denitrificans} PD1222 which lacked an identifiable conserved sequence (ICESymPd.PD1222.1, Figure 5-7). Further manual inspection of the conserved sequences upstream of \textit{msi169} and \textit{msi168a} revealed they contained regions of dyad symmetry, resembling a pair of 14-bp and 15-bp operator sequences centred 36 bp and 54 bp upstream of each genes start codon respectively (Figure 5-9). This organisation is similar to the inverted-repeat sequences found upstream of the genes encoding C.PvuII and C.AhdI, which both encode a pair of operator sequences centred 36 bp upstream of the start codon. It seems likely therefore that the conserved sequences found upstream of the \textit{xre} genes in this study contain operator sequences that are specific for each Xre protein.

\textbf{Figure 5-9} Putative operator structure upstream of \textit{msi169} and \textit{msi168a}.

(a) A map of the \textit{msi170} to \textit{msi168a} region is shown (1,054 bp), genes are shown as arrows, putative operator sequences are shown as open boxes. (b) DNA sequence found upstream of \textit{msi169} and \textit{msi168a} (start codon shown in bold italics). Arrows and bold font indicate regions of dyad symmetry within each operator sequence.

A search for the group 1 and 2 operator sequences (associated with \textit{msi169} and \textit{msi168a}) on the ICEMI/Sym\textsuperscript{R7A} sequence was carried out by searching for the motifs (motifs were created using the MEME program (14)) using the MAST motif search program (15). This approach failed to find a significant match to the \textit{msi168a} motif but was able to find a match to the pair of \textit{msi169} operator sequences (matching 24 bp over a 27 bp stretch at 160.8 kb). However this was within a region containing only transposon sequences and small decayed gene fragments. This suggests that the operator sequences identified upstream of each \textit{xre} gene are likely to be the only
functional Xre DNA targets on ICEM1Sym\textsuperscript{R7A}. Therefore the position of these operator sequences and the close association of the xre genes with msi170 suggest that these Xre proteins most probably regulate the expression of themselves and msi170.

To investigate if members of the Xre families identified in this study were associated with other genes in different elements, a BLASTP search was carried out using the Msi169 sequence as query. This identified Msi169 homologues present on the plasmids pRi1724 of Agrobacterium rhizogenes (64% a.a. ID), pRiA4b of Agrobacterium tumefaciens (60% a.a. ID), p42a of Rhizobium etli CFN 42 (39% a.a. ID), pRL1JI of Rhizobium leguminosarum bv. viciae (44% a.a. ID), pNGR234a of Rhizobium sp. NGR234 (41% a.a. ID) and pSMED02 of Sinorhizobium medicae WSM419 (63% a.a. ID). In these plasmids the Msi169 homologues are all encoded adjacent to and divergently operated to traM, which itself is encoded adjacent to and convergently orientated with traR (Figure 5-10). This configuration is remarkably similar to the traR-msi169 configuration encoded on ICEM1Sym\textsuperscript{R7A}, where the msi170 gene is present in place of traM.

The plasmid-encoded Xre sequences were then added to the ClustalW alignment used to create Figure 5-7 (data not shown). This revealed that while the sequences from pRi1724, pRiA4b and pSMED02 clustered with group 1 (as does Msi169), the pRL1JI and pNGR234a sequences clustered in group 6 and the sequence from p42a fell into group 3. Analysis of the traM-msi169 intergenic regions on each plasmid also revealed the presence of conserved motifs shown to be associated with each respective group in Figure 5-7 (data not shown). This again suggested the likely movement of these genes between elements; in this case the Xre sequences of pRi1724, pRiA4b and pSMED02 were more closely related to Msi169 from ICEM1Sym\textsuperscript{R7A} than the Xre sequences on the more obviously related plasmid clusters found on pRL1JI, pNGR234a and p42a.
The genetic organisation plasmids encoding \textit{traR}, \textit{traM} and \textit{msi169}-homologues compared with the similar clusters on ICE\textit{Ml\ Sym}\textsuperscript{R7A}. Numbers underneath the \textit{Msi169}/Xre-family homologues (yellow) indicate the clustering of each sequence with the groups identified in Figure 5-7.

5.1.3.8 \textit{traR}, \textit{traI1} and \textit{traI2}

The ICE\textit{Sym} clusters ICE\textit{Ml\ Sym}\textsuperscript{R7A}, ICE\textit{Ml\ Sym}\textsuperscript{MAFF} and ICE\textit{Sym}\textsuperscript{S.SKA58.1} encode \textit{traR} and \textit{traI} homologues upstream of \textit{msi172}. In ICE\textit{Sym}\textsuperscript{S.SKA58.1} the \textit{traI} gene is separated from the \textit{msi172} homologue by a gene encoding a putative phytanoyl-CoA dioxygenase. In ICE\textit{Ml\ Sym}\textsuperscript{R7A} and ICE\textit{Ml\ Sym}\textsuperscript{MAFF} a second \textit{traI} gene (\textit{traI1}) is encoded as a single-gene operon at 44.5 kb on ICE\textit{Ml\ Sym}\textsuperscript{R7A}.

As mentioned earlier, LuxR homologues are divided into two distinct phylogenetic families (103). A BLASTP search of the ICE\textit{Ml\ Sym}\textsuperscript{R7A} TraR sequence revealed the closest characterised protein match was to TraR on the Ti plasmid of \textit{A. tumefaciens} strain C58 (26\% a.a. ID) which suggested that the ICE\textit{Ml\ Sym}\textsuperscript{R7A} protein was a member of family A. The closest matches to putative TraR proteins (excluding ICE\textit{Ml\ Sym}\textsuperscript{R7A} and ICE\textit{Ml\ Sym}\textsuperscript{MAFF}) were to two proteins in \textit{Bradyrhizobium} BTAi1 with a.a. ID of 38\% and 34\%. The latter copy is encoded on the plasmid pBBya01 and was divergently oriented with a \textit{traI} homologue and upstream of a putative T4SS.
The sensitivity of particular LuxR homologues to specific AHLs varies greatly, as does the breadth and type of molecules produced by individual LuxI-type proteins. An alignment of the TraI1 and TraI2 sequences encoded by ICEMISym<sup>R7A</sup> revealed that these proteins shared 63% a.a. ID. An alignment with the corresponding ICEMISym<sup>MAFF</sup> sequences revealed they shared 99% and 91% a.a. ID to TraI1 and TraI2 respectively. Similarly to the previous BLASTP search of TraR, a BLASTP search of TraI2 revealed that the best match was to TraI on the plasmid pBBya01 of Bradyrhizobium BTai1 (54% a.a. ID). The next 3 matches were to the plasmids p42a of Rhizobium etli CFN 42 (47% a.a. ID), pNGR234a of Rhizobium sp. NGR234 (47% a.a. ID) the Ti plasmid of Agrobacterium tumefaciens strain C58 (45% a.a. ID) and pSMED02 of Sinorhizobium medicae WSM419 (45% a.a. ID). A BLASTP analysis of TraI1 demonstrated similar results.

Genes that are transcriptionally activated by TraR proteins of the Ti plasmid and plasmids pNGR234a of Rhizobium sp. NGR234 and pRL1JI of R leguminosarum bv. viciae share a semi-conserved dyad repeat sequence called a tra-box centred ~63 or 43 bp upstream of the transcriptional start site (57, 77, 117, 193). Manual investigation of the sequences upstream of the traI1 and traI2 genes of ICEMISym<sup>R7A</sup> revealed the presence of an inverted repeat weakly resembling a tra-box-like sequence centred 69 and 67 bp upstream of each start codon respectively (Figure 5-11).

Figure 5-11 Putative tra-boxes upstream of traI1 and traI2.

An alignment of the ICEMISym<sup>R7A</sup> traI1 and traI2 UTR sequences showing the presence of an inverted repeat (indicated by arrows) upstream of the start codon (italics).

### 5.1.4 Summary

In this chapter it was revealed that ICEMISym<sup>R7A</sup> is a member of an extensive family of chromosomally located elements found throughout the proteobacteria. A total of 27 genomic regions (named ICESym clusters) in 13 α-proteobacterial species were found to encode homologues of both RdfS and Msi170 in close association with genes encoding homologues of up to 25 other ICEMISym<sup>R7A</sup> proteins. Seventeen of these proteins were found to be encoded by 120
all elements identified, which included those with functions involved in T4SS, Mpf, Dtr and transcriptional regulation. Additionally each element was found to encode a region resembling an oriT upstream of rlxS, further confirming a conserved Dtr system for these elements. The less frequent conservation of proteins with plasmid associated functions such as Msi150 suggests that the ICESym elements may have evolved from an ancient plasmid ancestor and/or have similar mechanisms of maintenance to those of plasmids.

The lack of quorum sensing genes on all ICESym clusters except for ICEMlSymR7A, and ICEMlSymMAFF and ICESymS.SKA58.1 suggests that these genes may have been recently acquired at this locus. The position of traI2 and a putative tra box upstream of msi172 and msi171 could suggest the induction of QS could regulate expression of these genes also. An interesting future investigation would be to inspect the genes encoded upstream of msi172 and msi171 homologues on other ICESym clusters to see if other regulatory systems have been acquired in front of these genes in other systems.

Genes encoding Xre family transcriptional regulators were found to be conserved on all ICESym clusters indicating that these proteins have an evolutionarily conserved role on these elements. The Xre-encoding genes were positioned adjacent to msi170 and were most often divergently orientated, resembling the genetic organisation of RM cassettes, possibly indicating the clusters share a similar mechanism of regulation. Interestingly the Xre proteins sequences identified on ICESym clusters and several QS-regulated plasmids were able to be grouped phylogenetically into 6 distinct families, each with distinct operator sequences. This may suggest a selective pressure is acting to promote variation at this locus, possibly through negative frequency dependent selection.

The presence of the genes encoding Xre proteins on QS-regulated plasmids adjacent to traM could suggest that traM is regulated in a similar way to msi170. Likewise, the similarity between the genetic organisation of the plasmid-encoded and ICEMlSymR7A-encoded QS genes could suggest that Msi170 has an analogous role in regulation to that of TraM, which acts as a negative regulator of QS and transfer.
Chapter 6
Regulation of ICEMlSym$^{R7A}$
excision
Many MGE have evolved elaborate regulatory mechanisms to control the initiation of HGT, often in response to environmental and/or physiological cues. These functions are most often controlled by DNA-binding proteins that can act as simple genetic switches or combine with other regulators to form complex hierarchical networks that integrate signals from multiple sources (147). The previous chapter outlined the conservation in the ICESym clusters of genes encoding small DNA-binding proteins related to the XRE-family of transcriptional regulators, members of which are well known for regulating phage and RE-cassette gene expression. The XRE proteins are found encoded adjacent to homologues of msi170 and in most cases msi172 and msi171. In addition, ICEMISymR7A, ICEMISymMAFF and ICESymS.SKA58.1 encode homologues of the quorum sensing regulatory protein TraR upstream of msi172, such that the traR and msi169/msi168a are flanking the msi172-msi170 cluster. A very similar organisation is also observed on known regulatory centres of QS-regulated plasmids, where the TraR antiactivator protein TraM is encoded in place of the msi172-msi170 genes.

In this chapter the roles of quorum sensing and msi169-172 in the regulation of island excision were investigated.
6.1 Excision and transfer are regulated by quorum sensing

6.1.1 A plasmid containing **traR** is able to induce AHL production in R7A

The ICEMlSym\textsuperscript{R7A} genes **traR**, **traI2**, and **traI1** show similarity to the **traR** and **traI** quorum-sensing genes of *Agrobacterium tumefaciens* that regulate Ti plasmid transfer in response to population density (58, 182). As the frequency of island excision was increased at high population density, it seemed possible that these genes were involved in the regulation of excision. To test this, the **traR** and **traI2** genes and the preceding 696 bp were amplified by PCR using primers traRI2BamHI5 and traI2PstI3 and cloned into pFAJ1700 as a BamHI-PstI fragment to create pJR206 (Figure 6-1, Table 2-2). pJR206 was then introduced into R7A and the resultant strain caused intense purple colouration (due to violacein production) of the N-acyl homoserine lactone-sensitive indicator strain *C. violaceum* CV026 (118) when plated adjacent to it, whereas strain R7A caused little to no induction (Figure 6-2). This suggested that strain R7A(pJR206) overproduced an N-acyl homoserine lactone compared to strain R7A, as might be expected if the expression of the **traI1** and/or **traI2** genes were under positive autoregulation by TraR. To test if this effect required the copy of **traI2** on pJR206, a second plasmid containing only **traR** was constructed (pJR174) by amplifying **traR** from pJR206 using primers traRI2BamHI5 and traRBamHI3 and cloning this region into pFAJ1700 as a BamHI fragment (Figure 6-1, Table 2-2). Strain R7A(pJR174) induced the same response in CV026 as R7A(pJR206) (Figure 6-13) indicating that it was the presence of **traR** on these plasmids that was responsible for AHL induction. The ICEMlSym\textsuperscript{R7A}-cured strain R7ANS or derivatives of it containing pJR206 did not induce violacein production in CV026, indicating either that additional genes were required for AHL production or that expression of the **traR** and/or **traI2** genes was not induced in this background (Figure 6-2).
Figure 6-1  Complementation plasmids.

A map of regions cloned in the various plasmids constructed and/or used for complementation or expression studies in this chapter. Fragments are shown as lines with inward facing arrow ends to indicate the end of the fragment or outward facing arrow ends to indicate the fragment extends beyond the limits of the map. The dotted line below pJR173 indicates the deleted region within the overlap extension PCR product. Plasmid names are shown in bold above each fragment, primer names and total length of the cloned fragment are shown below.
Figure 6-2  CV026 bioassay demonstrating induction of AHL production by R7A(pJR206)/R7A(pJR174).

Twenty-µL aliquots of stationary phase M. loti TY cultures were spotted onto a large Petri dish (top and bottom of figure) and grown for two days at 28°C. The plate was then inoculated with CV026 from a fresh overnight TY plate by streaking the culture adjacent to the M. loti cultures without making contact. The purple colouration of CV026 indicates where AHL molecules have diffused from the M. loti culture through the agar and induced CV026 to produce violacein. pJR206 contains \(\text{traR}\) and \(\text{traI2}\) and pJR174 contains \(\text{traR}\) alone.

6.1.2 pJR174 induces excision and transfer of ICE\(\text{MISym}^{\text{R7A}}\)

The induction of violacein production in CV026 by R7A(pJR206) and R7A(pJR174) indicated that AHL production may be under autoregulation by TraR and that these strains were in an induced state. To test if R7A(pJR174) was affected in the excision of ICE\(\text{MISym}^{\text{R7A}}\), QPCR assays of DNA from exponential and stationary-phase cultures of strain R7A(pJR174) were carried out. Results from these experiments showed that the island was excised in 100% of cells ([\(\text{attB}\) = [\(\text{melR}\)]] in exponentially-growing or stationary-phase cultures and that \(\text{attP}\) was present at a ratio of approximately 1.5 to 1 to \(\text{attB}\) (Figure 6-3). To investigate if the transfer of ICE\(\text{MISym}^{\text{R7A}}\) was also increased, transfer of ICE\(\text{MISym}^{\text{R7A}}\) to the non-symbiotic \(M. \text{loti}\) strain N18 containing pFAJ1700 (Tet\(^R\)) was assayed using R7A(pJR206) as a donor (carried out by Gabriella Stuart (153)). Transfer from R7A(pJR206) was over 100-fold higher than from to R7A (3.6 x 10\(^{-2}\) (± 2.6 x 10\(^{-3}\)) versus 1.3 x 10\(^{-4}\) (±2.2 x 10\(^{-5}\)).
Figure 6-3 100% excision of ICEM/Sym\textsuperscript{R7A} in R7A(pJR174).

An example of a QPCR reaction (3 assay replicates shown) using DNA extracted from a log-phase (24 h) culture of R7A(pJR174) (pJR174 contains traR). Reactions of the attP, attB and melR amplicons are shown by red, blue and black respectively. The y-axis shows the mathematical derivative of the relative fluorescent signal. The x-axis shows the PCR cycle number (ct). The green line shows a selection of the threshold fluorescence within a period of uniform exponential amplification from which the ct values were taken. This demonstrates that the relative concentration of excised ICEM/Sym\textsuperscript{R7A} is 100%. This experiment was repeated at least 3 times from both log-phase and stationary-phase (64 h) cultures.

6.1.3 pJR174-induced AHL production requires \textit{traI1} but not \textit{traI2}

R7A(pJR174) and R7A(pJR206) induced violacein production in CV026, indicating the strains produced AHLs compatible with induction of this strain, namely those with an acyl chain 4-8C in length (118). In a preliminary analysis \textit{M. loti} supernatant extracts were separated by reverse-phase thin-layer chromatography (TLC) and the TLC plate was then overlayed with a thin agar film containing CV026 (Figure 6-4). Following overnight incubation the plate exhibited four distinct spots where the R7R(pJR206) extract induced violacein production in CV026, indicating at least four distinct AHLs species were present. In order to more comprehensively characterise the spectrum of AHLs produced by \textit{M. loti} strains, liquid chromatography coupled to hybrid quadrupole-linear ion trap mass spectrometry (LC-MS/MS) was used to identify and semi-quantitate AHL molecules present in extracts. This technique was used to specifically detect 130
molecules with acyl chains with or without 3-oxo- or 3-hydroxy substituent and lengths of 4, 6, 8, 10, 12 or 14 carbons. Media-only extraction controls were carried out in parallel and showed the presence of 3-oxo-C6-HSL (one sample 2.96 x 10^4 relative units (RU)) and some lower levels of other AHLs (< 2.59 x 10^3 RU), despite attempts made to reduce cross contamination by using all new glassware in all extraction steps. This suggests that results at this level or lower may be attributable to experimental noise or environmental contamination. Therefore a cut-off of 1 x 10^4 RU was used below which data were ignored, as were data on AHLs only detected in a single sample.

Analysis of R7A by LC-MS/MS revealed that the strain produced low but consistently detectable levels of 3-oxo-C6-HSL and 3-oxo-C12-HSL (log-transformed data shown in Table 6-1). R7A(pJR174) extracts however contained over 1000-fold more 3-oxo-C6-HSL than R7A as well as lesser amounts of C4-HSL, 3-oxo-C4-HSL, C6-HSL, 3-oxo-C6-HSL, 3-hydroxy-C6-HSL, C8-HSL, 3-oxo-C8-HSL, 3-hydroxy-C8-HSL, 3-oxo-C10-HSL, and 3-hydroxy-C10-HSL. The amount of 3-oxo-C12-HSL in R7A(pJR174) appeared similar to that of R7A. To establish the level of AHL production in the absence of traR or other ICEMISym^{R7A} genes, a traR mutant R7AΔtraR was constructed (see Figure 6-5 and Materials and Methods) and analysed along with the ICEMISym^{R7A}-cured strain R7ANS. While extracts from R7AΔtraR produced contained an identical AHL profile to that of R7A, extracts derived from R7ANS contained only low levels of 3-oxo-C12-HSL and 3-hydroxy-C10-HSL. This may suggest that the relatively low amounts of 3-oxo-C6-HSL observed in R7A are representative of an uninduced state and that the longer-chain AHLs produced in R7ANS and possibly R7A are synthesised by an unknown chromosomally-encoded AHL synthase(s).
Figure 6-4 Reverse-phase TLC analysis of *M. loti* AHL extracts

Two-μL of AHL extracts (equivalent to ~10 mL of culture supernatant) dissolved in acetonitrile were spotted at the base of the plate and allowed to dry before running the plate in 70/30 methanol/water. A synthetic preparation of C4-HSL was used as a control. An overlay of LB containing 50% v/v of an overnight CV026 culture and 0.5% w/v agar was then applied and incubated overnight. pJR206 contains *traR* and *traI2*.
Table 6-1  Acyl-homoserine lactones detected in *M. loti* by LC-MS/MS*

<table>
<thead>
<tr>
<th></th>
<th>C4</th>
<th>3=O C4</th>
<th>3-OH C4</th>
<th>C6</th>
<th>3=O C6</th>
<th>3-OH C6</th>
<th>C8</th>
<th>3=O C8</th>
<th>3-OH C8</th>
<th>3=O C10</th>
<th>3-OH C10</th>
<th>3=O C12</th>
</tr>
</thead>
<tbody>
<tr>
<td>R7ANS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R7A</td>
<td></td>
<td>4.3 ± 0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R7A (pJR174)</td>
<td>4.3 ± 0.2</td>
<td>4.5 ± 0.2</td>
<td>6.5 ± 0.2</td>
<td>7.4 ± 0.1</td>
<td>5.5 ± 0.2</td>
<td>6.6 ± 0.1</td>
<td>6.6 ± 0.3</td>
<td>5.1 ± 0.1</td>
<td>5.1 ± 0.7</td>
<td>4.6 ± 0.2</td>
<td>4.4 ± 0.3</td>
<td></td>
</tr>
<tr>
<td>R7AΔtraI1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R7AΔtraI1(pJR174)</td>
<td>4.6 ± 0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R7AΔtraI2</td>
<td></td>
<td>5.6 ± 0.1</td>
<td>5.0 ± 0.8</td>
<td></td>
<td>5.3 ± 0.0</td>
<td>4.9 ± 0.1</td>
<td></td>
<td>4.7 ± 0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R7AΔtraI2(pJR174)</td>
<td>4.6 ± 0.1</td>
<td>5.4 ± 0.2</td>
<td>6.7 ± 0.1</td>
<td>7.7 ± 0.0</td>
<td>6.4 ± 0.3</td>
<td>7.0 ± 0.1</td>
<td>7.3 ± 0.1</td>
<td>6.0 ± 0.1</td>
<td>5.5 ± 0.5</td>
<td>4.8 ± 0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R7AΔmsi170</td>
<td></td>
<td>4.1 ± 0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R7AΔmsi170</td>
<td>4.5 ± 0.4</td>
<td>4.1 ± 0.4</td>
<td>6.5 ± 0.4</td>
<td>7.1 ± 0.2</td>
<td>5.7 ± 0.3</td>
<td>6.7 ± 0.2</td>
<td>6.2 ± 0.5</td>
<td>5.2 ± 0.3</td>
<td>4.4 ± 0.1</td>
<td>4.7 ± 0.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Values represent the mean of log-transformed relative unit (RU) data from 6 replicates derived from 2 biological samples (data were log-transformed to remove positive association of the standard deviation with the mean). Error values represent ± 1 standard deviation.
Figure 6-5 Construction and PCR confirmation of in-frame deletions in traR, traI2, msi172, msi171, msi170, msi170-msi169, and traI1.

a) A map of constructed gene deletions is shown with the length of each deletion indicated above. The positions of the primer-pairs (1-5) used to confirm deletions are shown below the genetic map. b) Agarose-gel electrophoresis of PCR products obtained using primer-pairs 1-5 with DNA template extracted from wild-type and deletion-mutant strains.
To investigate if either of the putative AHL synthases TraI1 or TraI2 were involved in the synthesis of the above mentioned AHLs, *traI1* and *traI2* deletion mutants were constructed (see Figure 6-5 and Materials and Methods). Both mutants were initially intended to contain in-frame deletions within the coding region of each gene however due to a design error (an incorrect *traI2* start codon was used) the *traI2* mutant also contained a deletion of 22 bp preceding the start codon. Analysis of AHLs from the *traI1* mutant by LC-MS-MS revealed that it produced a similar low level of 3-oxo-C12-HSL to R7A but that in contrast to R7A it lacked 3-oxo-C6-HSL (Table 6-1). The R7AΔ*traI1*(pJR174) strain produced an AHL spectrum almost identical to R7AΔ*traI1* although a low level of 3-oxo-C6-HSL was present. As R7AΔ*traI1*(pJR174) lacked the high levels of 3-oxo-C6-HSL and broad spectrum of AHLs produced by R7A(pJR174), these data indicated that TraI1 was required for the majority of AHLs produced by R7A(pJR174), or was at least required for the positive autoregulation. Analysis of the *traI2* mutant revealed that surprisingly it had slightly increased production of 3-oxo-C6-HSL over that of R7A. Similarly, *traI2*(pJR174) showed slightly increased production of all produced AHLs relative to R7A(pJR174), except for 3-oxo-C12-HSL, which only in this strain was not detected (Table 6-1).

### 6.1.4 Excision is positively regulated by *traR*

As pJR174 induced excision in 100% of cells and as *traI1* was required for the pJR174-dependent induction in AHL production, it seemed likely that both *traR* and *traI1* were involved in the regulation of both AHL production and ICEMISymR7A excision in a population-density dependent manner. However, analysis of excision of ICEMISymR7A in R7AΔ*traR* revealed only a minor reduction in *attPlattB* abundance, and surprisingly, an increase in excision in stationary phase was still observed (Figure 6-6). The phenotype of R7AΔ*traR*(pJR174) appeared identical to R7A(pJR174) with respect to violacein production in CV026 (Figure 6-7) and excision of ICEMISymR7A (Figure 6-6). Similarly the *traI1* mutant showed only a slightly reduced level of excision. To complement R7AΔ*traI1*, the plasmid pJR039 that contains a copy of *traI1* and 704 bp of upstream sequence was constructed by amplifying the region by PCR using primers traI1clone5 and traI1clone3 and cloning it into pFAJ1700 as an XbaI fragment (Figure 6-1, Table 2-2). Initially after introduction of pJR039 into R7AΔ*traI1* the strain very weakly induced violacein production in CV026, however assays using inocula from subsequent platings showed no induction. Excision in R7AΔ*traI1*(pJR039) was increased relative to that of R7A (Figure 6-6), indicating that the plasmid complemented and likely overcompensated for the effect of the
traI1 mutation on excision. These results indicated that while the traR and traI1 genes were able to upregulate excision, other factors were contributing to the dependence of excision on growth-phase.

Figure 6-6 The effect of traR and traI1 mutations on ICEMISymR7A-excision.
Percentage attP and attB at 24 h (exponential growth) are shown by black and white bars respectively, while percentage attP and attB at 64 h (stationary phase) are shown by black bars/white hatching and white bars/black hatching respectively. Data presented is the average of three independent experiments (error bars represent ± 1 standard deviation). Individual QPCR measurements are the mean of triplicate reactions for each amplicon. Asterisks indicate samples with significant deviation from the wild-type data (* P < 0.05, ** P < 0.01). pJR174 contains traR, pJR206 contains traR and traI2, pJR039 contains traI1.
Figure 6-7 The effect of traR and traI1 mutations on AHL production.
CV026 bioassay plate assay of various *M. loti* strains (photo taken from beneath plate). Ten-µL of stationary phase *M. loti* TY culture was spotted (marked by X) onto a large Petri dish containing TY agar and grown for 24 h at 28°C. An overlay of LB containing 50% v/v of an overnight CV026 culture and 0.5% w/v agar was then applied and incubated overnight. The purple colouration of CV026 indicates where AHL molecules have diffused from the *M. loti* culture into the overlaid culture and induced CV026 to produce violacein. pJR174 contains *traR*, pJR039 contains *traI1*. 
The *traI2* gene encodes a putative AHL synthase but its role in AHL production remained unclear. The LC-MS/MS analysis of R7AΔ*traI2* extracts suggested that *traI2* is not required for the pJR174-induced production of AHLS seen in R7A(pJR174); in fact R7AΔ*traI2* and R7AΔ*traI2*(pJR174) exhibited slightly increased AHL production over their non-mutant counterparts. To investigate a possible role of *traI2* in excision, DNA extracted from R7AΔ*traI2* was examined by QPCR, which surprisingly revealed a marked reduction in excision (Figure 6-8). In an attempt to complement this strain, the plasmid pJR173 containing *traI2* (the overlap-extension PCR product used to create the *traR* deletion was cloned into pFAJ1700 as it contains both *traR* and *traI2* upstream intergenic regions) was constructed and introduced into R7AΔ*traI2* (Figure 6-1, Table 2-2). This construct did not reverse either the increased AHL production (Figure 6-9) or the decreased excision phenotype (Figure 6-8). Introduction of pJR174 or pJR206 (which also contains a copy of *traI2*) increased excision to between ~0.1-10% in both log and stationary phase cultures, but as the level of excision did not match that of R7A(pJR206)/R7A(pJR174), this suggested that the plasmids were partially masking the effect of the *traI2* mutation rather than fully reversing its effect.

A third attempt was then made to complement R7AΔ*traI2* by introducing the cosmid pUT11G that contained a 22.8-kb ICEMISym<sup>R7A</sup> fragment encoding the entire *traR-msi168a* region (ICEMISym<sup>R7A</sup> coordinates 198.9-221.7 kb (182)) (Figure 6-1). The R7A(pUT11G) strain was first analysed by QPCR and CV026 bioassay, which revealed that it elicited a weak induction of violacein production in CV026 (Figure 6-9) similar to that of R7AΔ*traI2* and that excision was increased to >10% in log-phase cultures and 100% in stationary-phase cultures (Figure 6-8). Analysis of R7AΔ*traI2*(pUT11G) in a CV026 bioassay revealed it induced a greater response than R7AΔ*traI2* or R7A(pUT11G), indicating that the induction resulting from the *traI2* mutation had an additive effect with that caused by introduction of the plasmid, as was also observed with R7AΔ*traI2*(pJR174) (Figure 6-9). The frequency of excision of ICEMISym<sup>R7A</sup> in R7AΔ*traI2*(pUT11G) was nearly identical to that of R7A(pUT11G), indicating this cosmid was able to complement the *traI2* mutation for excision (Figure 6-8). These results suggest that the *traI2* mutation had a pleiotropic effect on ICEMISym<sup>R7A</sup> excision and AHL production and that the latter was not able to be complemented by supplying the region *in trans*. 
Figure 6-8 The effect of the $traI_2$ mutation on ICEMISym$^{R7_A}$ excision.

Percentage $attP$ and $attB$ at 24 h (exponential growth) are shown by black and white bars respectively, while percentage $attP$ and $attB$ at 64 h (stationary phase) are shown by black bars/white hatching and white bars/black hatching respectively. Data presented are the average of at least two independent experiments (error bars represent ± 1 standard deviation). Individual QPCR measurements are the mean of triplicate reactions for each amplicon. Asterisks indicate samples with significant deviation from the wild-type data (* $P < 0.05$, ** $P < 0.01$). The $attB$ amplicon was only detected in a single sample at $t = 24$ and so no standard deviation or significance value was obtained. pUT11G is a cosmid containing 22.8 kb of ICEMISym$^{R7_A}$ DNA spanning the $traR$-$msi168a$ region. pJR173 contains $traI_2$, pJR174 contains $traR$ and pJR206 contains $traR$ and $traI_2$. 

$pUT11G$ is a cosmid containing 22.8 kb of ICEMISym$^{R7_A}$ DNA spanning the $traR$-$msi168a$ region. pJR173 contains $traI_2$, pJR174 contains $traR$ and pJR206 contains $traR$ and $traI_2$. 

139
**Figure 6-9** The effect of the *traI2* mutation on AHL production.

CV026 bioassay plate assay of various *M. loti* strains (photo taken from beneath plate). Ten-µL aliquots of stationary phase *M. loti* TY cultures were spotted (marked by X) onto a large Petri dish containing TY agar and grown for 24 h at 28° C. An overlay of LB containing 50% v/v of an overnight CV026 culture and 0.5% w/v agar was then applied and incubated overnight. The purple colouration of CV026 indicates where AHL molecules have diffused from the *M. loti* culture into the overlaid culture and induced CV026 to produce violacein. Movement of the *M. loti* culture during application of the overlay resulted in an irregular shaped zone of induction for some strains.

pUT11G is a cosmid containing 22.8 kb of ICE*Mll*Sym<sup>R7A</sup> DNA spanning the *traR-msi168a* region. pJR173 contains *traI2* and pJR174 contains *traR*. See Figure 6-7 for R7A control.

### 6.1.5 Regulation of excision is coupled to quorum sensing through *msi171* and *msi172*

The finding that pJR173, pJR174 or pJR206 (that carry *traI2*, *traR* and *traR-traI2* respectively) failed to complement the reduced excision seen in R7A*ΔtraI2* suggested that the mutation may have had a polar effect on downstream genes. Analysis of the *msi172* and *msi171* region revealed these genes were located directly downstream of *traI2*, with little room for a promoter region, indicating that they likely formed an operon (Figure 6-5). An analysis in the previous chapter revealed the presence of a *tra*-box-like motif centred 67 bp upstream of *traI2*. Currently characterised *tra*-boxes are found centred approximately either 63 or 43 nucleotides upstream of the transcriptional start site, which for *traI2* would predict the transcriptional start site near 24 or
4 nucleotides upstream of the start codon. Therefore a plausible explanation for the reduction in excision in tral2 was that the inadvertent deletion of 22 bp upstream of the tral2 start codon in R7AΔtral2 disrupted a tral2 promoter and therefore reduced expression of msi172 and msi171.

To test if msi172 or msi171 had a role in excision, in-frame deletion mutants of msi172 and msi172 were constructed (see Figure 6-5 and Materials and Methods) and analysed by QPCR. The attP/attB amplicons were only detected sporadically near the lower limits of detection in DNA extracted from either R7AΔmsi172 or R7AΔmsi171 in log or stationary-phase cultures, demonstrating that mutation of these genes nearly completely abolished excision (Figure 6-10). Next pJR174 was introduced into these strains and the resulting strains were then analysed by QPCR and CV026 bioassay. Both R7AΔmsi172(pJR174) and R7AΔmsi171(pJR174) induced violacein production in CV026 at an identical level to that of R7A(pJR174) (Figure 6-11); however ICEMISymR7A excision products were again unable to be consistently detected.

The pUT11G cosmid was then introduced into R7AΔmsi172 and R7AΔmsi171. Analysis of R7AΔmsi172(pUT11G) and R7AΔmsi171(pUT11G) revealed that pUT11G complemented excision in these strains to a level similar to that of R7A(pUT11G). Therefore the mutation of either msi172 or msi171 appeared to uncouple the regulation of excision from AHL production, indicating that the QS system may induce excision indirectly through activation of msi171-msi172. The reduction in excision seen in R7AΔtral2 therefore could have been a consequence of the disrupted tral2 promoter region and reduction in transcription of msi172 and msi171, rather than an effect from the loss of tral2 itself. However the mechanism resulting in the increased AHL production observed in the R7AΔtral2 strain still remained unclear.
Figure 6-10 The effect of *msi171* and *msi172* mutations on ICEMISym<sup>R7A</sup> excision.

Percentage *attP* and *attB* at 24 h (exponential growth) are shown by black and white bars respectively. Percentage *attP* and *attB* at 64 h (stationary phase) are shown by black bars/white hatching and white bars/black hatching respectively. Data presented are the average of at least two independent experiments (error bars represent ± 1 standard deviation). Individual QPCR measurements are the mean of triplicate reactions for each amplicon. Asterisks indicate samples with significant deviation from the wild-type data (* P < 0.05, ** P < 0.01). The *attPlattB* amplicons for samples R7AΔ*msi172*, R7AΔ*msi172*(pJR174), R7AΔ*msi171* and R7AΔ*msi171*(pJR174) were only detected sporadically (i.e. only in a single biological replicate and in either one or two assay replicates) and so the standard deviation and significance could not be calculated. pUT11G is a cosmid containing 22.8 kb of ICEMISym<sup>R7A</sup> DNA spanning the *traR-msi168a* region. pJR174 contains *traR*. 
Figure 6-11 The effect of msi172 and msi171 mutations of AHL production.

CV026 bioassay plate assay of various M. loti strains (photo taken from beneath plate). Ten-µL aliquots of stationary phase M. loti TY cultures were spotted (marked by X) onto a large Petri dish containing TY agar and grown for 24 h at 28° C. An overlay of LB containing 50% v/v of an overnight CV026 culture and 0.5% w/v agar was then applied and incubated overnight. The purple colouration of CV026 indicates where AHL molecules have diffused from the M. loti culture into the overlaid culture and induced CV026 to produce violacein. Movement of the M. loti culture during application of the overlay resulted in an irregular shaped zone of induction for some strains. pJR174 contains traR.

6.2 Regulation of quorum sensing and excision by msi169 and msi170

6.2.1 Excision and QS are repressed by Msi170

The conserved ICESym-family gene msi170 was found to be consistently genetically linked with an allele of Xre family similar to that encoded by msi169 and msi168a. Alleles of the same Xre-families were also found adjacent to traM on several QS-regulated plasmids, which led to the speculation that msi170 may function in an analogous way to traM in regulation of QS and transfer on ICEMlSymR7A. An in-frame deletion was constructed in msi170 (see Figure 6-5 and Materials and Methods) creating R7AΔmsi170. Analysis of AHL production in this strain by LC-MS/MS (Table 6-1) and CV026 bioassay (Figure 6-12) revealed that it produced a similar
spectrum and quantity of AHLs to the induced strain R7A(pJR174). Excision in this strain was then analysed, again revealing a similar profile to R7A(pJR174), with attPlattB present in ~40% of cells in log-phase cultures and 100% of cells in stationary-phase cultures. To complement R7AΔmsi170, DNA containing the msi170 gene and 160-bp upstream was amplified by PCR using primers 169170pro3 and msi1703clone and was then cloned into pFAJ1700 as a BamHI fragment (Figure 6-1, Table 2-2). This plasmid pJR170 was introduced into R7AΔmsi170 and following introduction the resulting strain was only able to very weakly induce CV026 (weak to no-induction, which varied plate to plate). R7AΔmsi170(pJR170) also displayed excision of ICEM/ SymR7A at almost wild-type levels. These data suggest that Msi170 is able to repress both excision and AHL production.

The TraM antiactivator protein encoded by the Ti plasmid is a direct inhibitor of the TraR protein (56, 85, 112, 148). Bioinformatic analysis in the previous chapter revealed that the traR and traI homologues were present in only 3 of the 27 ICESym clusters. This suggested Msi170 has likely co-evolved for a longer period of time with more conserved ICESym proteins and so likely has a target(s) other than TraR. If this was the case, increasing the expression of msi170 would likely repress excision even in the absence of TraR (i.e. in R7AΔtraR). To this end, the constitutive expression plasmid pJRNPT2 was created by amplifying the nptII promoter from pFAJ1708 using primers nptII5 and nptII3 (stop codons in all three frames were incorporated the 3’ primer), and cloning it into pPROBE-KT as a HindII fragment (Table 2-2). The region encoding msi170 and the upstream 47 bp (not including the putative msi169 operator sequence identified in the previous chapter) was then amplified by PCR using primers msi170ATG and msi1703clone and cloned into pJRNPT2 as a BamHI fragment creating pJRN170 (Figure 6-1, Table 2-2). This construct was then introduced into R7A and R7AΔtraR. Analysis of R7A(pJRN170) and R7AΔtraR(pJRN170) revealed that excised islands were present on average in fewer than 0.03% and 0.02% of cells respectively, demonstrating that repression of excision of ICEM/ SymR7A by Msi170 was able to occur independently of TraR (Figure 6-13). As a control, the pJRNPT2 plasmid was introduced into R7A and R7A(pJR174) and experiments with these strains showed that the vector alone had no effect on excision (Figure 6-13) or AHL production (Figure 6-12).
Figure 6-12 The effect of constitutive msi170 expression on AHL production.

CV026 bioassay plate assay of various *M. loti* strains (photo taken from beneath plate). Ten-µL of stationary phase *M. loti* TY culture was spotted (marked by an X) onto a large Petri dish containing TY agar and incubated for 24 h at 28°C. An overlay of LB containing 50% v/v of an overnight CV026 culture and 0.5% w/v agar was then applied and incubated overnight. The purple colouration of CV026 indicates where AHL molecules have diffused from the *M. loti* culture into the overlaid culture and induced CV026 to produce violacein. pJR174 contains *traR*, pJR170 contains *msi170*, pJRNPT170 contains the *msi170* gene expressed from the *nptII* promoter. Strain R7A(pJR174)(pJRNPT2) was included as a pJRNPT2 plasmid control.
Figure 6-13 The effect of constitutive expression of msi170 on ICE\textit{Ml}\textit{Sym}^{R7A}-excision.

Percentage \textit{attP} and \textit{attB} at 24 h (exponential growth) are shown by black and white bars respectively. Percentage \textit{attP} and \textit{attB} at 64 h (stationary phase) are shown by black bars/white hatching and white bars/black hatching respectively. Data presented are the average of at least two independent experiments (error bars represent $\pm$ 1 standard deviation). Individual QPCR measurements are the mean of triplicate reactions for each amplicon. Asterisks indicate samples with significant deviation from the wild-type data (* $P < 0.05$, ** $P < 0.01$). The \textit{attP}/\textit{attB} amplicons from R7A\textit{ΔtraR}(pJRNP170) samples at t= 24 h were only detected sporadically (only in a single biological replicate and in either one or two assay replicates) and so the standard deviation and significance could not be calculated. pJR170 contains \textit{msi170}, pJRNP170 contains \textit{msi170} expressed from the \textit{nptII} promoter, pJR174 contains \textit{traR} and pJRNPT2 was included as a plasmid control.

The increase in AHL production in R7A\textit{Δmsi170} suggests that the Msi170 protein must block the autoinduction of AHL production as well as the induction of excision. To test if pJRNP170 reduced AHL production and/or excision in the presence of pJR174, it was introduced into R7A(pJR174). Analysis of R7A(pJR174)(pJRNP170) revealed that it did not induce violacein production in CV026, while R7A(pJR174)(pJRNPT2) induced violacein production at a level similar to that of R7A(pJR174) (Figure 6-12). Analysis of excision revealed that R7A(pJR174)(pJRNP170) showed an intermediary level between R7A(pJR174) and R7A(pJRNPT2), with between 1% and 10% of cells containing an excised ICE\textit{Ml}\textit{Sym}^{R7A} at 24 hours.
or 64 h (Figure 6-13). These data therefore showed that the \textit{traR} and \textit{msi170} genes had opposite effects on both excision and AHL production; however the factors linking the action of the TraR and Msi170 proteins actions remained unclear.

6.2.2 Msi170 represses QS independently of \textit{msi172} and \textit{msi171}

Prior experiments demonstrated that both the \textit{msi172} and \textit{msi171} genes have a critical role in activation of ICEM/ISym\textsuperscript{R7A} excision, independent of AHL production, while \textit{msi170} appeared to repress both excision and QS. These data and the weak a.a. similarity between Msi170 and Msi171 (section 5.1.3.6) led to the speculation that the Msi170 protein may interact with either Msi171 or Msi172. Therefore the pJRNP170 plasmid was introduced into R7A\textDelta\textit{msi172}(pJR174) and R7A\textDelta\textit{msi171}(pJR174), to test if this construct was able to repress the AHL production induced by pJR174 in the absence of either \textit{msi172} or \textit{msi171}. Analysis of these strains by CV026 bioassay indicated the repression of QS by Msi170 was independent of \textit{msi172} or \textit{msi171} (Figure 6-14).
Figure 6-14 The effect of msi169 mutations and constitutive msi170 expression on AHL-production.
Three CV026 bioassay plate assays of various M. loti strains (photos taken from beneath plate – not to scale). Ten-µL of stationary phase M. loti TY culture was spotted (marked by X) onto a large Petri dish containing TY agar and incubated for 24 h at 28° C. An overlay of LB containing 50% v/v of an overnight CV026 culture and 0.5% w/v agar was then applied and incubated overnight. The purple colouration of CV026 indicates where AHL molecules have diffused from the M. loti culture into the overlaid culture and induced CV026 to produce violacein. Movement of the M. loti culture during application of the overlay resulted in an irregular shaped zone of induction for some strains. pJR174 contains traR, pJR170 contains msi170, pJR169170 contains msi169-msi170 and pJRNP170 contains msi170 expressed from the nptII promoter.

6.2.3 Msi169 acts as a positive regulator of excision and QS

The DNA-binding proteins encoded by msi169 and msi168a seemed likely candidates as regulators of msi170 expression and, by proxy, QS and excision of ICEMISymR7A. To investigate this possibility an msi169 insertional mutant containing a KanR cassette flanked by Ω-transcriptional terminators (R7Amsi169::ΩKan, constructed by Helen McIntyre) was analysed by QPCR. This revealed that R7Amsi169::ΩKan exhibited reduced excision of ICEMISymR7A.
relative to R7A, suggesting that msi169 may have a positive effect on excision (Figure 6-15). It was possible that Msi169 had a direct role in activation of excision. To test if deletion of msi169 still reduced excision in the absence of msi170, a strain containing a deletion of both genes and the intergenic sequence between them was constructed (see Figure 6-5 and Materials and Methods). This strain R7AΔmsi169msi170 exhibited an identical AHL-overproduction phenotype to R7AΔmsi170 and this phenotype was able to be complemented by introduction of pJR170, pJRNPNP170 or the plasmid pJR169170 which contained msi169 and msi170 (Figure 6-14) (pJR169170 was produced by amplifying the msi169-msi170 region by PCR using primers msi1693clone and msi1703clone and cloning it into pFAJ1700 as a BamHI fragment (Figure 6-1, Table 2-2)). The excision of ICEMI/SymR7A in R7AΔmsi169msi170 was also identical to that in R7AΔmsi170 (Figure 6-15). This indicated that Msi169 was unlikely to have a direct positive role (at least not a stronger action than the repression by Msi170) in the regulation of excision but rather that it was repressing expression of msi170. To complement the R7Amsi169::ΩKan strain, a region containing msi169 and 116 bp of upstream DNA was amplified using primers msi1693clone169170pro5 and cloned into pFAJ1700 as a BamHI fragment (Figure 6-1, Table 2-2). This plasmid pJR169 was then introduced into R7Amsi169::ΩKan and the resultant strain was found to at least be partially complemented for excision (Figure 6-15).

To test if increased expression of msi169 affected excision of ICEMI/SymR7A or AHL production, the msi169 ORF and 15 bp upstream of the start codon was amplified using primers msi169ATG and msi1693clone and cloned into pJRNPT2 as a BamHI fragment downstream of the nptII promoter (Figure 6-1, Table 2-2). This plasmid pJRNPNP169 was then introduced into R7A. Analysis of R7A(pJRNPNP169) by QPCR revealed that this plasmid had no effect on excision of ICEMI/SymR7A (Figure 6-15) and analysis by CV026 bioassay revealed that like R7A it did not induce violacein production (data not shown). During construction of pJRNPNP169 it was evident that transcription through the msi169 ORF was indeed occurring at least in E. coli, as the cell cultures appeared fluorescent green due to the expression of the promoterless GFP cassette encoded downstream of the cloned msi169. The DNA sequence obtained from this construct did not reveal any abnormalities.

It seemed likely that strain R7Amsi169::ΩKan expressed msi170 at an increased level and that this resulted in repression of excision by Msi170. Since the increased expression of msi170 provided by introduction of pJRNPNP170 was shown to quench the AHL production and excision...
induced by pJR174, R7Amsi169::ΩKan was tested to see if the msi169 mutation in this strain had a similar effect. The pJR174 plasmid was then introduced into R7Amsi169::ΩKan. The strain R7Amsi169::ΩKan(pJR174) showed almost no induction of CV026 (very-weak to none, this varied between plates) and excision frequencies of 0.6% in log-phase cultures and 40% in stationary-phase cultures. These data are consistent with msi169 having a positive role in excision and AHL production through repression of msi170 expression.

**Figure 6-15** The effect of msi169 mutations on ICEMISymR7A excision.

Percentage attP and attB at 24 h (exponential growth) are shown by black and white bars respectively. Percentage attP and attB at 64 h (stationary phase) are shown by black bars/white hatching and white bars/black hatching respectively. Data presented are the average of at least two independent experiments (error bars represent ± 1 standard deviation). Individual QPCR measurements are the mean of triplicate reactions for each amplicon. Asterisks indicate samples with significant deviation from the wild-type data (* P < 0.05, ** P < 0.01). pJR174 contains traR, pJR169 contains msi169 and pJRNP169 contains msi169 expressed from the nptII promoter.
### 6.2.4 Positive autoregulation of Msi169

To investigate the expression of *msi169*, the *msi169* promoter region was amplified by PCR using primers 169170pro5 and 169170pro3 and cloned upstream of the promoterless lacZ gene in pFJX as a BamHI fragment (Figure 6-1, Table 2-2). This plasmid pFJX169 was then introduced into R7A, R7ANS, R7A*msi169*::ΩKan, R7AΔ*msi170* and R7AΔ*msi169msi170*. Analysis of expression in R7A(pFJX169) revealed 400 ± 13 MU of β-galactosidase activity from log-phase cultures (t = 24) and 683 ± 73 MU in stationary-phase cultures (t = 64). Strain R7ANS(pFJX169) strain however only produced ~100 MU regardless of growth-phase (Table 6-2). This suggested that the presence of Msi169 might be required for full expression from the *msi169* promoter. Analysis of the other three strains confirmed this, as the strains lacking a copy of *msi169* (R7A*msi169*::ΩKan(pFJX169) and R7AΔ*msi169msi170*(pFJX169)) showed expression similar to R7ANS(pFJX169), while R7AΔ*msi170*(pFJX169) showed similar expression to R7A(pFJX169) (Table 6-2). These data suggested that Msi169 positively autoregulates its expression from the *msi169* promoter and that this expression may be increased in stationary-phase cultures.

#### Table 6-2 Expression from the *msi169* promoter

<table>
<thead>
<tr>
<th>Strain</th>
<th>β-galactosidase activity (MU) ± standard deviation*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t=24 h</td>
</tr>
<tr>
<td>R7A(pFJX169)</td>
<td>400 ± 13</td>
</tr>
<tr>
<td>R7ANS(pFJX169)</td>
<td>102 ± 3</td>
</tr>
<tr>
<td>R7A<em>msi169</em>(pFJX169)</td>
<td>83 ± 2</td>
</tr>
<tr>
<td>R7AΔ<em>msi170</em>(pFJX169)</td>
<td>337 ± 71</td>
</tr>
<tr>
<td>R7AΔ<em>msi169msi170</em>(pFJX169)</td>
<td>69 ± 26</td>
</tr>
</tbody>
</table>

*Data presented is the mean of two independent biological replicates.

### 6.3 ICEMISymR7A may replicate by rolling circle replication in its excised form

The excision of ICEMISymR7A in 100% of cells in strains such as R7AΔ*msi170*, R7A(pJR174), R7A(pJR206) and R7A(pUT11G) indicates that it is likely to replicate autonomously in its excised form. This was surprising as the island lacks obvious plasmid replication genes (182). To determine if *rlxS* was required for the maintenance of ICEMISymR7A in its excised state, pJR206
was electroporated into strain R7ArlxS::lacZ, an insertional mutant containing an integrated copy of the plasmid pFUS2 (mutant constructed by G. Stuart). Following electroporation and plating on media selecting for both ICEMISymR7ArlxS::lacZ (GmR) and pJR206 (TcR), small colonies were obtained but these were unable to be subcultured. A number of normal-sized colonies were obtained on medium with tetracycline alone and these were able to be subcultured on media containing tetracycline but not on media containing gentamicin. Analysis of one such TcR isolate by PCR revealed that the intS gene was unable to be amplified and the strain was also auxotrophic for nicotinate and thiamine (data not shown) indicating loss of ICEMISymR7A. These results suggest that maintenance of ICEMISymR7A in its excised circular form in requires the rlxS gene product, consistent with RlxS being required for extrachromosomal replication of the island.

6.4 Excision of ICEMISymR7A in bacteroids

To investigate if ICEMISymR7A was excised from the chromosome in bacteroids, Lotus corniculatus seedlings were inoculated with M. loti R7A. At 6 weeks post-inoculation, nodules containing R7A were crushed and DNA was extracted. QPCR of DNA extracted from 6 independent plant nodules revealed that attP was present at an average of 4% (95% confidence interval between 3% and 15%) and attB was present in 3% (95% confidence interval between 3% and 10%) of cells, similar to the level of excision previously observed in stationary-phase broth cultures.

6.5 ICEMISymR7A excision and quorum sensing are not required for symbiosis of Lotus corniculatus

Quorum sensing has been implicated in regulation at various stages of nodule infection/formation and symbiosis. To test the possibility that the QS system and/or the excision of ICEMISymR7A from the chromosome was required for symbiosis, R7AΔtraR, R7AΔtraI2, R7AΔmsi172, R7AΔmsi171, R7AΔmsi170, R7AΔmsi169msi170, R7AΔtraI1, R7AΔintS and R7AΔrdfs cultures were used as inocula for L. corniculatus seedlings (three plant replicates per strain). After 6 weeks, all plants had formed between 1 and 6 nodules. This suggested that the QS system and the ability of ICEMISymR7A to excise were not required for symbiosis. However
this does not discount more subtle effects which may only be revealed in more specific circumstances, such as in competition with non-mutant strains.

6.6 Attempted purification of TraR and Msi170

In order to facilitate future biochemical experiments with the ICEMlSym$^{R7A}$-encoded TraR and Msi170, the proteins were over-expressed in *E. coli* with the intention of subsequent purification. The *traR* ORF was amplified by PCR using primers traRpurEco5 and traRpurKpn3 and the *msi170* ORF was amplified using 170purEco5 and 170purKpn3. The 3’ primer for each product introduced a 6xHis sequence and a stop codon to each ORF to allow nickel-NTA column purification of the proteins. The PCR products were then cloned into the IPTG-inducible expression vector pSK67, which contains a synthetic ribosome-binding site centred 9 bp upstream of the single EcoRI site. The *msi170* PCR product was cloned as an EcoRI-KpnI fragment (producing pSK170) and the *traR* PCR product was cloned in two parts as EcoRI-BglII and BglII-KpnI fragments produced from separate digestions of the PCR product (producing pSKTraR). The plasmids pSK67, pSK170 and pSKTraR were then introduced into *E. coli* strain BL21(DE3)(pLysS), producing strains SK67, SK170 and SKTraR.

6.6.1 Heterologous expression of TraR in *E. coli*

To express the TraR-(6xHis) fusion protein from pSKTraR, 100-mL LB broths were inoculated with 1-mL from 37°C overnight cultures of SK67 and SKTraR. When the OD$_{600}$ of the cultures reached 0.5, IPTG was added to a final concentration of 1mM. The TraR protein of *A. tumefaciens* requires the presence of its cognate AHL for correct folding, stability and dimerisation. To account for the possibility the ICEMlSym$^{R7A}$-encoded TraR had the same requirements, the total AHL extracts from two 100-mL TY broth cultures of stationary-phase R7A(pJR174) was added to parallel SK67 and SKTraR cultures at the time of IPTG induction. Samples taken from overnight cultures showed that both the AHL(+) and AHL(-) extracts of SKTraR (but not SK67) contained a protein band corresponding to the predicted TraR-(6xHis) proteins molecular weight of 27 kDa (Figure 6-16). Interestingly, at 3-hours post-inoculation, the culture supplemented with AHL extracts from R7A(pJR174) appeared to contain a relatively denser 27-kDa band than the overnight cultures, while the same band in the AHL(-) sample was not visible. This could suggest that the ICEMlSym$^{R7A}$-encoded TraR may also require AHL for
stability and/or folding in *E. coli*. However as the total protein concentration of the SKTraR-AHL(-) sample was particularly low in the SDS-PAGE analysis the result will require further confirmation.

Figure 6-16 Over-expression of TraR in *E. coli*.
A 16% acrylamide SDS-PAGE gel of total protein extracts from SK67 and SKTraR cultures sampled before (T = 0) and after induction with IPTG (at 3 h post induction (T = 3) or overnight). Lanes containing protein from cultures supplemented with AHLs extracted from *M. loti* R7A(pJR174) are indicated while others lack added AHL. The left-hand lane contains protein standards and the molecular weights are shown in kDa. The position of the expected 27kDa band corresponding to the TraR-(6xHis) fusion is shown by an arrow.

6.6.2 Purification of Msi170

Similar to the induction of SKTraR, a 100-mL LB broth of SK170 was inoculated from an overnight culture and induced with 1 mM IPTG at an OD$_{600}$ of 0.5. Preliminary SDS-PAGE analysis indicated that a 12-kDa band corresponding to the Msi170-(6xHis) protein was produced in this strain and was present in highest quantity in overnight cultures (data not shown). A nickel-column purification of Msi170-(6xHis) was attempted (Figure 6-17). Analysis of the insoluble protein fraction from the total protein extract revealed a large amount of the 12-kDa band was present. Likewise the column flow-through following incubation of the soluble fraction with the nickel-NTA slurry also contained a dense 12-kDa band. The nickel-NTA bound 154
Msi170-(6xHis) was then eluted by washing the column with increasing concentrations of imidazole, which resulted in a small amount of the purified protein being eluted in the 40mM and 100mM concentrations. These results indicated that in the current protocol the majority of the Msi170-(6xHis) protein was not bound to the column, suggesting that the 6xHis tag may have been inaccessible to the nickel-NTA, possibly due to denaturation of the Msi170-(6xHis) protein and/or formation of insoluble inclusion bodies during broth culture or protein extraction.

![Figure 6-17](image)

**Figure 6-17** Partial purification of Msi170. A 16%-acrylamide SDS-PAGE gel of protein extracts from SK67 and SK170 samples taken during attempted purification of Msi170-(6xHis). The outermost lanes contain protein standards ranging in size between 212 and 2.3-kDa. The position of the expected 12kDa band corresponding to the Msi170-(6xHis) fusion is shown by an arrow.

### 6.7 Growth effects of pJR206 and pJR174

During the course of this study several observations made it apparent that expression of the QS system of ICE*Sym*<sup>R7A</sup> was having an inhibitory effect on growth, namely a reduction in stationary-phase cell density (herein termed growth reduction) in R7A(pJR206)/ R7A(pJR174) and a total inhibition of growth (herein termed growth inhibition) following introduction of pJR206 or pJR174 into specific mutant strains. For instance the cfu of strain R7A(pJR206) in stationary-phase culture (64 h) was lower than that of R7A (1.1 x 10<sup>9</sup> ± 2.2 x 10<sup>8</sup> vs. 3.6 x 10<sup>9</sup> ±
This effect was not observed in R7AΔtraII(pJR206) cells indicating it was associated with increased AHL production and possibly excision.

The introduction of pJR206 or pJR174 into particular strains resulted in the formation of growth-inhibited colonies that grew to between 0.5-2 mm in diameter, but failed to grow any further, suggesting a cell-density effect was preventing growth. Two of the strains this phenomenon occurred in were R7AΔrdfS and R7AΔintS. Interestingly when colonies resulting from electroporations of pJR174 into R7AΔrdfS or R7AΔintS were picked and restreaked as soon as they became visible, the isolates grew normally for 2-3 days before growth would cease. After several days of continued incubation of the original electroporation or restreaked plates, colonies with wild-type growth would appear in the primary-streak. These were able to be recultured on media containing tetracycline and grew at a wild-type rate, indicating that a mutation may have occurred allowing the colony to overcome the growth inhibition. Further work is needed to investigate the nature of these isolates.

Prior to the construction of the in-frame deletion mutant strains R7AΔtraII, R7AΔtraR, R7AΔmsi172, R7AΔmsi171, R7AΔmsi170 and R7AΔmsi169msi170, the same mutations were constructed in a derivative of strain R7A (renamed R7A*), that was subsequently found to exhibit increased AHL production (Figure 6-18) and excision (~100-fold increase, data not shown). The mutants previously described in this chapter were subsequently constructed from the original frozen stock culture of R7A that was shown to have low AHL production. The strain R7A* had been passaged on G/RDM plates in the lab over a 5 month period, during which the cultured strain had somehow altered. DNA regions of R7A* containing traR-msi168a, traI1, msi110-rlxS, traG and upstream intergenic sequences were amplified by PCR and sequenced, to investigate the possibility that a mutation had occurred in these genes. However no differences from the sequence reported by Sullivan et al. (182) were discovered. Suspecting that the R7A* strain was a constructed mutant mistaken to be R7A, R7A* was tested for antibiotic resistance to gentamicin, neomycin and tetracycline; however it was sensitive to all three antibiotics. Storage of R7A* at -70º C for 1 month, or passaging through several TY broth cultures also had no effect on the strain’s altered phenotypes.
CV026 bioassay of R7A* (left) and R7A (right). M. loti strains were inoculated from G/RDM plate cultures onto TY plates and then grown for 48 h. CV026 was then streaked adjacent to the M. loti strains and incubated overnight. The purple colouration of CV026 indicates where AHL molecules have diffused from the M. loti culture and induced CV026 to produce violacein.

The R7A* in-frame deletion mutants that were constructed all exhibited increased AHL production over their R7A counterparts, except for traR and traI1 mutants which lacked any observable AHL production. Additionally, introduction of pJR206 or pJR174 into R7A*ΔtraR or R7A*Δmsi170 resulted in the formation of growth-inhibited colonies, similar to those formed when these plasmids were introduced into R7AΔrdfS or R7AΔintS. (It was this result that originally drew attention to R7A*). Interestingly when the pJR174 plasmid was introduced into the remade R7AΔmsi170 mutant, growth-inhibited colonies were again produced, indicating that this effect was not a result of the R7A* background for this particular mutant. In contrast R7AΔtraR(pJR174) is viable. Light microscopy was used to analyse the growth-inhibited R7A*ΔtraR(pJR174) colonies shortly after they appeared on plates following electroporation. This revealed that the cells were larger and irregularly shaped compared to R7A cells (data not shown). To gain a clearer picture of the morphology of these cells, colonies arising on plates following electroporation of pJR174 into R7A and R7AΔtraR and a mock (no DNA) electroporation of R7A were scraped off and fixed for analysis by transmission electron microscopy (TEM) (Figure 6-19 and Appendix D). Similarly to the view seen under light microscopy, the R7A*ΔtraR(pJR174) cells appeared much larger and irregularly shaped in comparison to the R7A or R7A(pJR174) cells. Some of the R7A*ΔtraR(pJR174) cells also appeared to contain excessive numbers of white granules likely composed of poly-β-hydroxybutyrate. Due to time constraints, further investigation into the causes of these growth effects was not possible and further investigation is clearly needed.
Figure 6-19 Transmission electron microscopy of strains R7A, R7A(pJR174) and R7A*ΔtraR(pJR174)

Transmission electron microscopy of *M. loti* R7A, R7A(pJR174) and R7A*ΔtraR(pJR174) cells scraped from G/RDM plate cultures 4 days after electroporation with either H₂O or purified pJR174 DNA. Scale bars are shown in the bottom-right of each picture. For additional photos see Appendix D.
6.8 Discussion

In this chapter it was found that the introduction of various plasmids containing \( \text{traR} \) led to excision and stable maintenance of ICE\( Ml\text{Sym}^{R7A} \) in 100% of cells and that this was associated with a large increase in AHL production. The increase in excision and AHL production were dependent on \( \text{traI1} \). Excision of ICE\( Ml\text{Sym}^{R7A} \) also required \( msi172 \) and \( msi171 \) that are likely co-transcribed with \( \text{traI2} \). Both AHL production and excision were independently repressed by expression of \( msi170 \), and \( msi170 \) itself was negatively regulated by \( msi169 \).

The introduction of pJR174 (\( \text{traR} \)) or pJR206 (\( \text{traR-traI2} \)) resulted in a massive increase in the amount of AHLs produced by \( M. \text{loti} \) R7A. QS systems often encode positive feedback loops regulating the production of autoinducer, so introduction of either of these plasmids may cause a similar autoregulation. It is likely that introduction of these plasmids increases the copy number of the TraR protein relative to a negative regulator such as Msi170, allowing the autoregulation to occur in the presence of a much lower AHL concentrations than that which occurs in the wild-type. The result of introduction of pUT11G at least partially supports this theory, as this cosmid contains both \( \text{traR} \) and \( msi170 \) and strains carrying it elicited a much weaker induction of AHL production compared to strains carrying either pJR174 or pJR206, which carry \( \text{traR} \) but not \( msi170 \).

The most abundant AHLs produced by R7A(pJR174) were 3-oxo-C6-HSL, 3-oxo-C8-HSL, C8-HSL and C6-HSL, ordered highest to lowest. As the increased AHL production induced by pJR174 required \( \text{traI1} \) but not \( \text{traI2} \), it seems likely that TraI1 produces these species. The 3-oxo-C6-HSL was produced at a ~6-fold higher amount than the next most abundant species and was the only AHL detectible in R7A apart from 3-oxo-C12-HSL. The 3-oxo-C6-HSL was undetectable in an R7A\( \Delta \text{traI1} \) mutant, indicating that this was the major species produced by TraI1 in R7A.

The LC-MS/MS technique used in this chapter detected common AHL species likely to be present in samples and confirmed their presence by comparison with synthetic standards. AHL species other than those specifically tested are not quantified using this approach, so it is possible that \( M. \text{loti} \) R7A may produce other less common or unique AHL species. For instance a preliminary search for AHL species resembling the unique long-chain AHLs identified in a
marine *Mesorhizobium* sp. isolate (98) revealed the presence (in extracts from R7A and R7A(pJR174)) of a molecule containing the HSL moiety and having an m/z of 282, which resembles 5-cis-C12-HSL (see Appendix C). However due to the lack of a molecular standard this could not be confirmed nor quantified. In future studies the development of AHL reporter systems specific for *M. loti* R7A-encoded QS regulators should allow detection of all AHL species with biological relevance.

It is possible that TraI2 contributes to production of the AHLS identified but that its presence is not required for autoregulation in R7A(pJR174). The TraI1 and TraI2 proteins are their own closest relatives (62% a.a. identity), suggesting that they could have arisen by gene duplication and so may produce a similar spectrum of AHLS. Surprisingly, deletion of *traI2* actually increased the production of the same AHLS that are increased in R7A(pJR174). Mutagenesis studies carried out on the AHL synthases LuxI, RhlI and EsaI have identified several residues critical for activity (67, 76, 142). All the critical residues identified in these studies are conserved in the TraI2 sequence and there are no obvious a.a. substitutions that would render the protein inactive (analysis not shown). Future experiments designed to resolve this issue will involve determination of the AHLS produced following constitutive expression of the *traI1* and *traI2* genes from plasmids in R7ANS and in heterologous backgrounds such as *A. tumefaciens* and *E. coli*.

The finding that *R7AΔtraI2* exhibited increased AHL production suggests that the TraI2 protein could have a negative effect on AHL production. However as plasmids carrying *traI2* did not complement the increased AHL production phenotype of *R7AΔtraI2*, this seems unlikely. Furthermore the increase in AHLS in *R7AΔtraI2* appeared to be additive with that induced by introduction of the *traI2*-carrying plasmids pJR206 and pUT11G. Another possibility is that the increased AHL production seen in *R7AΔtraI2* is an artifactual effect of the deletion of sequence upstream of the *traI2* start codon. The *R7AΔtraI2* strain also showed reduced excision of ICE Ml SymR7A, likely due to a polar effect on msi172 and msi171. This could also be responsible for the increased AHL production in *R7AΔtraI2*, but this is not consistent with the phenotypes of *R7AΔmsi172* and *R7AΔmsi171* which exhibited wild-type-like AHL production. A third possibility is that the deletion of the 20 bp upstream of the start codon of *traI2* prevents or reduces binding of TraR at the putative *tra*-box upstream of *traI2*. This could lead to an excess of free TraR available to bind at the second *tra*-box identified in front of *traI1*, resulting in an
increased level of TraI1 and AHL synthesis. A future reconstruction of a mutant with an in-frame deletion entirely within the traI2 ORF and the construction of a traI1-traI2 double mutant should shed light on the likelihood of each of these possibilities.

A higher than wild-type level of AHL production consistently correlated with a large increase in ICEMISym<sup>R7A</sup> excision, with the exception of strains affected in the expression of traI2, msi172 or msi171. This highlights a link between the two phenotypes; the QS system positively regulates expression of traI2, msi172 and msi171 and the latter two genes are critical for excision of ICEMISym<sup>R7A</sup>. As the in-frame deletion of either msi172 or msi171 had an identical negative effect on excision, the two genes may have a closely related function. This is supported by the bioinformatic analysis in the previous chapter which showed that msi172 and msi171 when present are always encoded in the same order and on some elements are even parts of the same ORF. One possibility is that they are required for activation or derepression of transcription of the rdfS-rlxS and trb operons.

The action of Msi170 was shown to be repression, both of QS and excision. Constitutive expression of msi170 reduced excision even in the absence of traR (e.g. reduced excision was observed in R7AΔtraR(pJRNP170)) and also repressed QS in the absence of excision (AHL production was suppressed in R7AΔmsi172(pJR174)(pJRNP170) and R7AΔmsi171(pJR174)(pJRNP170)), indicating that repression of these phenotypes occurred independently. The weak similarity of the Msi170 and Msi171 proteins outlined in the Chapter 5 is suggestive of the proteins having arisen from gene duplication. The experimental data presented in this chapter showed that the two genes have opposite actions. It is conceivable that one of the proteins has maintained the same target of action as the ancestral protein, while the other has evolved to take on a structurally inhibitive or competitive role antagonistic to the other. As Msi170 can repress QS in the absence of msi172 or msi171, it must act independently of Msi172/Msi171. Furthermore the strains R7AΔmsi172(pJR174) and R7AΔmsi171(pJR174) exhibited similar AHL production to that of R7A(pJR174). If the role of Msi172/Msi171 was to antagonise Msi170, it might be expected that these strains would show reduced AHL production in comparison to R7A(pJR174).

The bioinformatics analyses support the existence of an ancestor common to msi171 and msi170 and the experimental data best support a model whereby Msi170 and Msi172/Msi171 are able to
independently repress and activate ICEMISym\textsuperscript{R7A} excision respectively. It is possible that the Msi170 and Msi171 proteins have a similar structure and/or mechanism of action, but different targets, or that they share the same target but affect it differently. As these proteins have no characterised homologues, protein-protein interaction experiments such as a yeast or bacterial two-hybrid system may be sensible first steps in elucidating their function. As it is likely that other proteins conserved on the ICESym clusters have coevolved with Msi172, Msi171 and Msi170, they would be sensible candidates to initially test for interaction.

In Chapter 5 it was found that, on all ICESym clusters identified, the msi170 homologues were closely linked with genes encoding members of the DNA-binding Xre family such as msi169. In this chapter it was shown that an msi169 insertion mutant had significantly reduced ICEMISym\textsuperscript{R7A} excision and was also unable to be induced to overproduce AHLs through introduction of pJR174. Both these phenotypes are similar to those observed in strains containing msi170 constitutively expressed from plasmid pJRNP170, which suggests that Msi169 represses msi170. A strain containing a deletion of both msi169 and msi170 exhibited AHL and ICEMISym\textsuperscript{R7A} excision phenotypes identical to those of R7A\textsuperscript{-}msi170, consistent with this hypothesis. The similarity of the organisation of RM cassettes with msi169-msi170 indicates that msi170 may show a similar pattern of expression to MTase-encoding genes. The regulation of expression of MTase proteins from RM cassettes has not yet been comprehensively defined and is probably complex. The transcription of pvul\textsubscript{IM} for example initiates from at least two sites within the beginning of the pvul\textsubscript{IC} ORF itself and strangely requires DNA sequence within the -10 region of pvul\textsubscript{IC} (between +33 and +48 relative to pvul\textsubscript{IM} start site). Therefore analysis of the regulation of msi170 would benefit from transcriptional start site mapping to guide design of expression analysis experiments.

Expression analysis of the msi169 upstream intergenic region showed that msi169 expression was positively autoregulated. In comparison the C.PvuII protein both positively and negatively autoregulates its expression through binding one or two operator sequences as a dimer (by analogy an Msi169 dimer would likely bind to each of the two inverted repeat upstream of msi169). pvul\textsubscript{IC} is only weakly expressed in the absence of C.PvuII and binding of a C.PvuII dimer to the first operator results in strong activation of expression. The second operator sequence has a lower affinity for C.PvuII and therefore a higher concentration of protein is required for it to become occupied. Binding of a C.PvuII dimer to the second operator results in
strong attenuation of pvulIC expression. In this way the concentration of C.PvuII is maintained in equilibrium with replication, so that REase and MTase expression are carefully controlled (96, 188).

As previously mentioned, a double inverted repeat sequence is centred 36 bp upstream of the msi169 start codon, identical in position to the operator sequence observed upstream of pvulIC. The pvulIC promoter is unusual in that it appears to lack a -35 hexamer, a region that is completely replaced by the pvulIC operator sequence. Also unusual is that it produces a leaderless mRNA, initiating from the start codon of pvulIC (96, 188). Like pvulIC, msi169 lacks an obvious ribosome-binding site. Experiments in this chapter showed that the reduced excision phenotype of R7Amisi169::ΩKan could be complemented by introduction of pJR169 (msi169 downstream of native promoter) but not pJRN170 (msi169 downstream of nptII promoter). The translation of Msi169 from a leaderless mRNA could provide an explanation for this result i.e. expression of msi169 from the nptII promoter would produce an artificially long mRNA from which Msi169 may not be efficiently translated.

A surprising result was that mutation of traR or traI1 resulted in only a minor reduction in ICEMISymR7A excision and that growth-phase dependence of excision was still observed. An increase in msi169 expression was also observed in stationary-phase cultures. This could explain the stationary-phase increase in excision, if the increased msi169 expression was accompanied by a decrease in msi170 expression. However the interpretation of the β-galactosidase assay results is complicated by the accumulation of β-galactosidase protein over time, so direct measurement of msi170 mRNA levels is required to provide more convincing evidence.

The observation that in QS-induced strains ICEMISymR7A was excised in 100% of cells was surprising, as it suggests ICEMISymR7A is able to replicate independently of the chromosome. One possibility is that the RC replication component of conjugation also facilitates replication of ICEMISymR7A within the same cell. Consistent with this hypothesis was the result that stability of ICEMISymR7A upon introduction of pJR174 (containing traR) required rlxS, which encodes the relaxase likely responsible for initiation and completion of RC replication. Conjugation-independent replication requiring a relaxase has been previously observed on the plasmid R64. The NikB relaxase is able to mediate RC replication-dependent recombination of a plasmid containing tandemly repeated oriT sequences, resulting in the deletion of one of the oriT
sequences (59). A similar experiment could be undertaken to test if this is the case in *M. loti* R7A, by constructing a tandem duplication of the ICE*MiSym*<sup>R7A</sup> *oriT* on a plasmid and introducing it into an Mpf-deficient strain.

In Chapter 4 it was shown that constitutive expression of *rdfS* was lethal in R7A and growth inhibitory in R7ANS. Several growth-inhibition effects were observed in this chapter that were associated with induction of the QS system. As *rdfS* is most likely expressed at high levels during induction of the QS system, it seemed possible that this was responsible for the growth reduction seen in R7A(pJR174) and R7A(pJR206). However introduction of pJR174 into R7AΔ*rdfS* resulted in complete inhibition of growth, indicating that *rdfS* expression was not required for the growth inhibition. This suggests that the growth effects observed in this chapter are more likely to be an effect of the QS system itself, in line with the bacteriocin-like effects of QS systems in other rhizobia (78). A further experiment to test this hypothesis would be to compare the stationary-phase culture density of R7A(pR174) with those of R7AΔ*msi171*(pJR174) and R7AΔ*msi172*(pJR174), as these strains produce a similar amount of AHL but the latter strains do not excise ICE*MiSym*<sup>R7A</sup>.

Regulation of growth by QS has been documented in *R. leguminosarum*, *R. etli*, and *Rhizobium* sp. strain NGR234, however a common mechanism responsible (if there is one) has yet to be described. As mentioned in Chapter 1, the pRL1JI plasmid of *R. leguminosarum* confers sensitivity to 3-OH-7-cis-C14:1-HSL-induced growth inhibition, mediated by the plasmid-encoded genes *bisR* and *traR* and 3-oxo-C8-HSL. Maximum growth inhibition requires the presence of both AHL species and is associated with an increase in the production of the translational elongation factor EF-Ts (40, 194). In contrast, the QS systems *cinR-cinI* and *raiR-raiI* of *Rhizobium etli* CNPAF512 are required for maximum growth rates (39). The pNGR234 plasmid of *Rhizobium* sp. strain NGR234, like pRL1JI, produces a TraR protein that responds to 3-oxo-C8-HSL. The *traR* gene of pNGR234 was shown to be the only plasmid-encoded gene responsible for sensitivity to 3-oxo-C8-HSL (77). It has been reported that a chromosomal target of TraR has been identified in *Rhizobium* sp. strain NGR234, however these data remain to be published (78).

The growth inhibition phenotype observed upon introduction of pJR174/pJR206 into strains R7AΔ*intS*, R7AΔ*rdfS*, R7AΔ*msi170*, R7A*Δ*msi170* and R7A*Δ*traR*, appeared to be delayed,
suggesting that the accumulation of AHLs was required before inhibition occurred. This is consistent with the result that early subculture of these strains produced individual colonies which would again grow to 0.2-2.0 mm. Interestingly if the inhibited plates were incubated further, single colonies arose with normal growth rate. This could indicate that a mutation on ICEMISym\textsuperscript{R7A}, pJR174 or elsewhere in the R7A genome may have occurred, resulting in either the silencing of QS or the suppression of the growth inhibition. A screen for isolates which continue to produce AHLs but are able to grow at normal rates could lead to discovery of genes involved in the growth inhibition.

The inadvertent isolation of R7A*, a strain with increased AHL production and ICEMISym\textsuperscript{R7A} excision, suggests that during passaging and storage of the strain a mutant derivative of R7A was selected. R7A* appeared not to be the outcome of a strain mix-up or mutations within the QS genes on ICEMISym\textsuperscript{R7A}. Therefore it remains unclear exactly what caused the observed changes in this strain. It is possible that mutations or even epigenetic changes may have occurred on either ICEMISym\textsuperscript{R7A} or elsewhere in the genome. Isolation and analysis of an exconjugant containing ICEMISym\textsuperscript{R7A} derived from a mating between R7A* to R7ANS could clarify where the change has occurred, by separating ICEMISym\textsuperscript{R7A} in R7A* from its chromosomal background. A complementary experiment would be to cure R7A* of ICEMISym\textsuperscript{R7A} through introduction of pJR204 and then transfer ICEMISym\textsuperscript{R7A} to this new non-symbiotic strain from R7A and analyse the resulting strain.

The total growth inhibition as seen in strains R7A\textDelta intS(pJR174) and R7A\textDelta rdfS(pJR174) indicates that induction of the QS system has a much stronger effect in these strains than in R7A. A possible explanation for this is that because both these strains are deficient in ICEMISym\textsuperscript{R7A} excision, the expression of genes that act in response to QS to activate excision (i.e. msi172 and msi171) is growth inhibitory in the absence of excision. Consistent with this hypothesis, growth inhibition was not observed in either of the AHL-induced strains R7A\textDelta msi171(pJR174) or R7A\textDelta msi172(pJR174), even though these strains do not excise ICEMISym\textsuperscript{R7A}.

An alternative hypothesis for the growth inhibition seen in some strains is that it is related to a higher level of AHL production. Both the R7A* and R7A\textDelta msi170 strains showed increased AHL production in the absence of pJR174. Introduction of pJR174 into R7A\textDelta msi170 or derivatives of R7A* (with the exception of R7A*\textDelta traII) resulted in growth inhibition. If the
AHL production in these strains is additive with that induced by pJR174, they would be expected to produce a much greater level of AHL. As strain R7AΔtraI2(pJR174) produced higher levels of AHL than strain R7AΔ(pJR174), a comparison of stationary-phase culture densities of these strains could test this hypothesis. It may that the growth inhibition observed upon introduction of pJR174 into R7AΔintS and R7AΔrdfS is also be due to increased AHL production in these strains, as the growth inhibition prevented their being tested for AHL production. For instance a secondary or pre-existing mutation in these strains could have induced AHL production similar to that of R7A*. However the presence of a pre-existing mutation in these strains seems unlikely as they are derivatives of the laboratory R7A strain which at the time demonstrated levels of ICEM/SymR7A excision similar to that of the stock R7A strain.

The growth inhibition phenotype observed upon introduction of pJR174/pJR206 into strains R7AΔintS, R7AΔrdfS, R7AΔmsi170, R7A*Δmsi170 and R7A*ΔtraR, appeared to be delayed, suggesting that the accumulation of AHLs was required before inhibition occurred. This is consistent with the result that early subculture of these strains produced individual colonies which would again grow to 0.2-2.0 mm. Interestingly if the inhibited plates were incubated further, single colonies arose with normal growth rate. This could indicate that a mutation on ICEM/SymR7A, pJR174 or elsewhere in the R7A genome may have occurred, resulting in either the silencing of QS or the suppression of the growth inhibition. A screen for isolates which continue to produce AHLs but are able to grow at normal rates could lead to discovery of genes involved in the growth inhibition.

In summary, the results presented in this chapter suggest that ICEM/SymR7A excision and transfer are positively regulated by quorum sensing mediated through the expression of positive regulators Msi171 and Msi172, and negatively regulated by Msi170. Msi170 production is itself negatively regulated by msi169, perhaps in response to the cell cycle. A model for the regulation of ICEM/SymR7A excision is presented in Chapter 7.
Chapter 7
Concluding discussion
7.1 Integration and excision

The studies carried out in this thesis have shown that ICEMISym\textsuperscript{R7A} is capable of mediating its excision and integration using a unique configuration of a conserved apparatus. Like many phage and transposons, an integrase (IntS) catalyses excision of ICEMISym\textsuperscript{R7A} likely through the binding of core and arm binding sites encoded on \textit{attL} and \textit{attR}. IntS is required for both excision of ICEMISym\textsuperscript{R7A} and integration into the new host cell, as complementation of an \textit{intS} mutant required expression of \textit{intS} in both the donor and the recipient.

The \textit{intS} gene was expressed highly from \textit{attP}, a situation resembling the expression of the P4 integrase-encoding gene \textit{intb13} on the ICE \textit{clc} (171). High expression from \textit{attP} likely insures rapid integration of ICEMISym\textsuperscript{R7A} and \textit{clc} upon entry into a new host and is possibly a conserved evolutionary strategy of these and other elements. The \textit{attP} promoter of \textit{intB13} is encoded entirely within a region found upstream of the core and is juxtaposed in front of the \textit{intB13} gene upon circularisation of \textit{clc} (171). In contrast the promoter of the resolvase-encoding gene \textit{tnpX} of the Tn4451 and Tn4453 transposons straddles the recombination site, so that the -35 and -10 regions of the promoter are only brought together upon circularisation of the elements (113). Interestingly the expression from \textit{attP} required an unusually large region encompassing the same region required for integration. Therefore the position of the promoter on ICEMISym\textsuperscript{R7A} \textit{attP} could be similar to that of either \textit{intB13} or \textit{tnpX}. The requirement of such a large region on \textit{attP} suggests that there is more involved than just the formation of a promoter, i.e. that there are possibly other important regions on \textit{attP} required for full expression. One possibility is that IntS, RdfS and/or other ICEMISym\textsuperscript{R7A} or chromosomally-encoded proteins bind this region and regulate expression. The expression of the P4 integrase for example is negatively regulated by both Int and its cognate RDF Vis at both the transcriptional and post-transcriptional level (145). Future transcript mapping and expression studies in various genetic backgrounds (e.g. R7ANS, \textit{rdfS} and \textit{intS}) should shed light on the likelihood of these various possibilities.

Excision of ICEMISym\textsuperscript{R7A} was found require a novel RDF that is found to be both genetically and possibly functionally associated (discussed in 7.2) with the conjugation system, a unique position for a gene traditionally associated with phage. Like \textit{intS}, \textit{rdfS} is required for efficient excision and transfer of ICEMISym\textsuperscript{R7A} and expression of these two genes is sufficient to mediate recombination of \textit{attL} and \textit{attR} in the absence of other ICEMISym\textsuperscript{R7A} genes. RdfS appears to be
only weakly related to RDFs of other P4 family integrases but BLASTP searches revealed it is closely related to a large family of genes conserved on elements related to Tn4371 and the ICESym elements identified in this study. Therefore it appears the RdfS and its relatives represent a unique family of RDFs associated with ICEs.

Interestingly not only did overexpression of rdfS cure R7A of ICE\textit{Ml}Sym\textsuperscript{R7A}, it was lethal in an \textit{intS} mutant and reduced the growth rate of the non-symbiotic strain R7ANS. This suggests that RdfS has targets elsewhere in the genome. Another example of an RDF having an adverse effect on growth has been documented in \textit{E. coli}. AlpA, the RDF of a cryptic CP4 prophage that encodes a P4-like integrase, indirectly induces the expression of a protease phenotype as prophage excision creates a single nucleotide deletion within the tmRNA gene (95). Work in this laboratory has shown that an integrase closely related to IntS, found on the adhesin island IMEMI/Adh\textsubscript{R88B} (which interestingly encodes a homologue of AlpA) of \textit{M. loti} strain R88B, also leaves a single nucleotide deletion within the single-copy \textit{phe}tRNA gene integration site upon excision (33). It is possible that interaction of RdfS with integrases similar to this in R7A could be responsible for the growth inhibitory effects.

**7.2 Regulation of ICE\textit{Ml}Sym\textsuperscript{R7A} excision and transfer**

A proposed global model of ICE\textit{Ml}Sym\textsuperscript{R7A} regulation based on data presented in this thesis is shown in Figure 7-1. In summary, excision and transfer of ICE\textit{Ml}Sym\textsuperscript{R7A} are regulated by both QS and at least one Xre-family regulator, coupled to excision through the expression of \textit{msi172}, \textit{msi171} and \textit{msi170}. TraR is likely activated by binding one or some of the major AHL species induced in R7A(pJR174) such as 3-oxo-C6-HSL. TraR-3-oxo-C6-HSL likely binds the putative \textit{tra}-boxes upstream of \textit{traI1} and \textit{traI2}, inducing further production of 3-oxo-C6-HSL and expression of \textit{msi172} and \textit{msi171}. Msi172 and Msi171 then activate ICE\textit{Ml}Sym\textsuperscript{R7A} excision through an unknown mechanism. Both QS and excision are repressed by Msi170 through an unknown mechanism and \textit{msi170} expression is repressed by Msi169. The initial absence of Msi169 in a new recipient likely allows expression of \textit{msi170}, preventing expression of QS and excision-stimulating genes.
Figure 7-1 Model of ICEISym\textsuperscript{R7A} transfer and its regulation.

(a) A map of the regulatory network controlling excision and transfer is shown. Genes are shown as filled arrows, proteins are shown as filled or open squares, rectangles, circles and ellipses. AHLs are indicated by an open triangle. Lines with arrows indicate a positive interaction while lines ending in a perpendicular line indicate inhibition or repression. (b) A map of transfer and excision-related genes on ICEISym\textsuperscript{R7A}. Sites of possible transcriptional regulation are indicated by arrows. Note that the gene map in part (b) is presented in a smaller scale than that presented in part (a).

The expression of rdf\textsuperscript{S} is likely coupled with that of \textit{traF} and possibly the DNA-relaxase-encoding gene \textit{rlxS}. This genetic configuration likely facilitates coordinated induction of excision, Mpf and Dtr functions through expression of a single transcript. So far it remains to be discovered how the expression of the \textit{traG-trb} operon is regulated, but it seems likely that it is
also coupled to expression of the aforementioned genes. A distinct possibility is that the same mechanisms that connect Msi172, Msi171, and Msi170 with activation or repression of excision and QS, also act on this operon. Another possibility is that RdfS or the CopG-like protein Msi031 act as transcriptional regulators connecting the expression of these and/or other ICEMISymR7A genes. On the QS-regulated Ti and pRL1JI plasmids, TraR directly activates the expression of the trb operon, by binding a tra-box upstream of tral1-trb (40, 192, 193). However the lack of a sequence resembling a tra-box upstream of the traG-trb operon on ICEMISymR7A and the absence of QS genes on the other ICESym clusters makes this possibility less likely. As the regulatory system controlling ICEMISymR7A excision and transfer appears to be complex, microarray experiments in QS-induced and excision-induced strains or protein-protein interaction experiments (e.g. yeast/bacterial two-hybrid system) would likely help elucidate interactions in this network.

Data presented in this thesis and from experiments carried out in parallel indicate that RdfS may have a role in transfer other than that as an RDF. Firstly, partial transfer of ICEMISymR7A from strain R7AΔintS to N18 was observed at a low frequency through an Hfr type mechanism, but transfer from R7AΔrdfS was never observed. Secondly, transfer of an oriT-containing plasmid to E. coli is observed at 1 x10^{-7} ± 0.4 exconjugants/donor from wild-type M. loti, but transfer was not observed when R7AΔrdfS was used as the plasmid-carrying donor. This may suggest that RdfS has a structural or regulatory role in transfer. For example, RdfS could act as an accessory protein in the Dtr nucleoprotein complex in conjunction with RlxS. A variety of DNA-binding proteins have been recruited from diverse origins throughout the evolution of specific relaxase systems, so this is not at all unlikely. Alternatively, RdfS could act as a transcriptional regulator, activating expression of operons such as rdfS-rlxS, trb, or even traR or tral2-msi171. Either of these possibilities could lead to explanations for anomalous results such as the growth reduction and lethality seen in R7ANS(pJR204) and R7AΔintS(pJR204) respectively, or the conservation of an RdfS homologue on the M. loti strain MAFF303099 plasmid pMlb, which would not seemingly require an RDF.

An intriguing finding in this study was that ICEMISymR7A is often excised only in a proportion of cells but is able to be maintained indefinitely in the excised state. The presence of excised ICEMISymR7A in only a proportion of cells could indicate either that excision is only transient and occurring throughout the population, or that a small proportion of cells maintain an excised
ICE\textsubscript{Ml}\textsubscript{Sym\textsuperscript{R7A}} the majority of the time (or somewhere between these two extremes). The latter scenario resembles the outcome of bistability, a phenomenon whereby a transcriptional switch (usually involving positive autoinduction) is stable only when completely on or completely off, leading to a population of genetically identical cells expressing different phenotypes (46). At the population level, this allows a colony to benefit from both behaviours, or even prepare for multiple common environmental outcomes simultaneously. With respect to ICE\textsubscript{Ml}\textsubscript{Sym\textsuperscript{R7A}}, the switching on of HGT in a small proportion of cells could allow it to better manage the potential risks and benefits of transfer on fitness and propagation rate, by inducing HGT in a only a small fraction of cells. It is tempting to speculate that QS, commonly associated with cooperative social behaviours (190), is regulating the factors governing the likelihood of switching from a non-mobile to an actively-transferring state. In this way not only could QS induce transfer in only a proportion of cells, but it could adjust the proportion at the population level to suit environmental conditions. The theory that a small proportion of cells perpetually maintain an excised ICE\textsubscript{Ml}\textsubscript{Sym\textsuperscript{R7A}} could be tested through the use of fluorescent microscopy over time of living cells expressing markers indicative of the presence of an excised ICE\textsubscript{Ml}\textsubscript{Sym\textsuperscript{R7A}}. For example, it may be possible to couple expression from \textit{attP}, which encodes a strong constitutive promoter (see Chapter 3), to an unstable GFP variant, so that the presence of \textit{attP} coincides with expression of fluorescence.

### 7.3 Evolution of ICE\textsubscript{Ml}\textsubscript{Sym\textsuperscript{R7A}} regulation

Central to the model presented in Figure 7-1 are the genes \textit{msi172}, \textit{msi171} and \textit{msi170}. The alignment of the proteins produced by these genes with the RO00034 protein from Tn\textsubscript{4371} (see Chapter 6) showed that these four genes may have a common ancestor. The RO00034 protein resembles a concatenated version of Msi172 and Msi171 suggesting these proteins may have a related function. This is consistent with the finding that both Msi172 and Msi171 were required for ICE\textsubscript{Ml}\textsubscript{Sym\textsuperscript{R7A}} excision. The Msi170 sequence also showed weak similarity to the C-terminal portions of both RO00034 and Msi171. Again, the function of Msi170 appears to be related to that of Msi171 and Msi172, albeit with the opposite outcome. Extrapolation of these results provides support for a model of evolution of Msi172, Msi171 and Msi170 from an ancestor similar to RO00034. This ancestor was likely a two-domain protein, required for activation of transfer of the ancestral element, as RO00034 more closely resembles Msi172-Msi171 than Msi170. Duplication of the gene encoding this protein, followed by divergent evolution, then
resulted in one of the proteins evolving a function antagonistic to the other, involving the C-terminal portion of both proteins. The expression of the truncated repressor protein (proto-Msi170) then came under control of an Xre regulator, completing the regulatory switch present on the ICESym clusters. In this way, the regulatory centre could have evolved from a single RO00034-like gene. Likewise, the control of the activators Msi172 and Msi171 could have recently come under control of QS, through acquisition of the traR and traI2 genes upstream of msi172. In a recent review, Ptashne (147) argues that most complex genetic regulatory circuits throughout biological systems have likely evolved from simpler components such λ-cI, by combining with other regulators in a modular way. Recent experimental work in both λ and QS systems supports this theme (13, 54). If the functions of RO00034, Msi172, Msi171 and Msi170 are as predicted here, then this provides an excellent example of the evolution of a complex regulatory switch from more simple components.

The bioinformatic analysis in Chapter 5 revealed that the ICESym clusters and several plasmids carry alleles of the Xre family. These Xre proteins were grouped by sequence similarity into 6 distinct families, each also having a distinct conserved DNA motif upstream of the genes encoding them. Incongruence was observed between the inheritance of the Xre alleles and Msi170 homologues on the ICESym clusters, suggesting that the Xre alleles were being frequently shuffled or replaced with other family members through the locus history. This could suggest a strong selection pressure on this locus, acting to promote variation amongst elements. This type of variation could be the result of a negative frequency dependent selection that favours rare Xre alleles in the population of mobile elements.

A similar hypothesis has been proposed to explain the breadth of sequence variation seen in the operator specificities of the C proteins of RM cassettes, as well as variation of REase and MTase site specificities themselves (97, 99, 129). Successful introduction of a C-regulated RM cassette into a new host requires that the initial C protein concentration is low to prevent premature induction of C and REase expression. Therefore the presence of another element regulated by a C protein can exclude entry into the host of another cassette if they recognise the same operator sequence. This has been demonstrated for the BamHI and PvuII RM cassettes, which encode REases with different site specificities but C proteins with the same operator specificity. Transfer of one of these cassettes into a strain containing the other cassette, or even just a plasmid expressing either of the C proteins, results in highly reduced transfer efficiency due to the
premature expression of the incoming REase and cleavage of the host genome. This effect has been termed “apoptotic mutual exclusion” as it lethally restricts the presence of members of the same C-protein family in a cell (129).

While none of the Xre proteins encoded on the ICESym clusters appear to regulate REase genes, the selection pressures imposed by exclusion or incompatibility could explain the apparent selection for variation in the Xre proteins encoded on these elements. For instance, transfer of ICE\textit{Ml}\textsc{Sym}^{\text{R7A}} into a strain expressing an Xre that recognizes the \textit{msi169} operator sequence could lead to premature expression of excision and transfer genes, preventing integration and establishment in the new host. This model would predict that multiple mobile elements frequently coinhabit the same cells and must adapt to maintain their individual regulation. As each of the distinct \textit{xre} genes identified in this study was associated with a conserved operator sequence, these regions could be used as markers (detectible by PCR, hybridisation etc.) for the presence of \textit{xre} sequences in environmental isolates from the rhizosphere and hence be used to test this hypothesis.

The gene clusters on ICE\textit{Ml}\textsc{Sym}^{\text{R7A}}, ICE\textit{Ml}\textsc{Sym}^{\text{MAFF}} and ICESym.Pl.DS1.1 are unusual in that they encode two Xre alleles in tandem adjacent to \textit{msi170}. An obvious question is do these second alleles have a functional role or are they merely fossils of evolution at the locus? As discussed, there appears to have been frequent shuffling of the Xre alleles between the elements, which could leave scars in the form of pseudogenes at the locus. Consistent with this, tBLASTX analysis in these regions on ICE\textit{Ml}\textsc{Sym}^{\text{R7A}} and ICE\textit{Ml}\textsc{Sym}^{\text{MAFF}} indicate the presence of small gene fragments with similarities to various Xre alleles (data not shown). If the evolutionary divergence of ICE\textit{Ml}\textsc{Sym}^{\text{R7A}} and ICE\textit{Ml}\textsc{Sym}^{\text{MAFF}} are considered, the most parsimonious progression resulting in this configuration would require replacement of the second Xre allele after the ancestral \textit{msi169}-\textit{msi170} configuration had already formed. This would suggest selection has acted to replace the second Xre allele, implying that it confers a phenotype. As mentioned above, evolution of complex regulatory circuits may occur through modular addition of regulators. This could be an example of the addition of yet another regulatory component to this switch. Another distinct possibility to these hypotheses is that these Xre alleles have some inbuilt mechanism of promoting their movement between elements at particular DNA hotspots and that these second alleles are merely a result of this. To investigate these various possibilities it would be interesting to see if transfer of ICE\textit{Ml}\textsc{Sym}^{\text{R7A}} into a strain expressing Msi169 and/or
Msi168a is reduced. Also, as evidence indicates the Xre alleles can be replaced by other members, it would be interesting to see if artificially replacing \textit{msi169} or \textit{msi168a} with other Xre alleles and their operator sequences would produce a functional complementation.

### 7.4 Final remarks

\textit{ICE\textit{M}}/\textit{Sym}^{R7A} appears a member of an abundant class of MGE that are likely at least as instrumental in bacterial evolution as are their plasmid counterparts. Likewise, regulation of these elements appears to be similarly complex, involving multiple regulators, growth effects, social communication, lifecycle stages and interactions with other MGE. In this study the full induction of transfer was only obtained through genetic manipulation. It seems likely then that the specific environmental conditions inducing transfer are yet to be revealed. The regulatory model presented in this thesis will hopefully provide a framework for further investigation into the regulatory cascade resulting in HGT of these important MGE.
Chapter 8
References


Chapter 9
Appendices
A

Growth Media

Media were made up in distilled water unless otherwise stated, and autoclaved at 121 C (15 psi) for 15 min on a liquid cycle. When solid culture media was required, GIBCO-GRL bacteriological agar was added to broth to give a final concentration of 1.5% before autoclaving.

Growth media for *E. coli*

**Luria-Bertani (LB)**
(121)
To 1000 mL water add: 
10 g bacto-tryptone 
5 g yeast extract 
5 g NaCl

**2YT**
(163)
To 1000 mL water add: 
16 g bacto-tryptone 
10 g yeast extract 
5 g NaCl

**SOB**
(163)
To 1000 mL water add: 
20 g bacto-tryptone 
5 g yeast extract 
0.5 g NaCl

Do not add MgCl₂
Add 10 ml of 250 mM KCl.
Growth media for *Mesorhizobium*

*Rhizobium defined medium* (RDM)

To 700 mL of water add:

- 10 mL salts (25 g MgSO$_4$.7H$_2$O, 2 g CaCl$_2$, 1.5 g FeEDTA, 20 g NaCl per litre)
- 10 mL Bromothymol blue (2 mg.mL$^{-1}$)
- 6 mL NH$_4$Cl (18 g.L$^{-1}$)
- 1 mL Trace elements (3 mg ZnSO$_4$.7H$_2$O, 40 mg Na$_2$MoO$_4$.2H$_2$O, 50 mg H$_3$BO$_3$, 40 mg MnSO$_4$.4H$_2$O, 4 mg CuSO$_4$.5H$_2$O), 1 mL CoCl$_2$ [0.2 g.L$^{-1}$])
- 100 mg L-Histidine
- 10 g MES (S/RDM only)

For G/RDM: adjust pH to 6.5-7.0 with 2 M KOH
For S/RDM: adjust pH to 6.1-6.4 with solid KOH

Adjust volume to 1000 mL or for sucrose RDM adjust the volume to 750 mL

After the medium has been autoclaved, allow to cool and then add:

Carbon source

- For RDM containing glucose as the sole carbon source (G/RDM):
  - 20 mL of 20% (w/v) glucose; sterilised by autoclaving
- For RDM containing sucrose as the sole carbon source:
  - 250 mL of 20% (w/v) sucrose; sterilised by autoclaving

Phosphates
For glucose, glutamate, and sucrose RDM:
10 mL of stock solution containing 10% (w/v) K$_2$HPO$_4$ and 10% (w/v) KH$_2$PO$_4$

For succinate G/RDM:
4 mL of stock solution containing 10% (w/v) K$_2$HPO$_4$ and 10% (w/v) KH$_2$PO$_4$

Vitamins
1 mL of stock solution containing 50 mg nicotiamide, 50 mg thiamine HCl and 1 mL of Biotin (1 mg.ml$^{-1}$) per 250 mL of water; filter sterilised.

Antibiotics/colourmetric substrates/antifungals/ferrichrome/EDDHA/root and seed exudates/FeEDTA added if required.

**TY**
(20)
To 1000 mL water add:
- 5 g Bacto-tryptone
- 3 g yeast extract
- 2 mL CaCl$_2$ (0.65 mg.mL$^{-1}$ stock)

**Nitrogen-free medium for growth of Lotus seedlings in plant tubes**

**Jensen’s seedling agar**
(189)
(per litre)
To 500 mL of water add:
- 1 g CaPO$_4$
- 0.2 g K2JPO$_4$
- 0.2 g MgSO$_4$.&H$_2$O
- 0.1 g NaCl
- 0.1 g FeCl$_3$

Heat to dissolve solutes.
To 500 mL of distilled H₂O add 12 g GIBCO-BRL agar and autoclave to dissolve agar.

Add molten agar to the warm salts solution and dispense 8 mL into 150 mm × 19 mm tubes. Cap tubes with a foam stopper. After autoclaving place the tubes on an angle so a slope forms when the agar sets.
Buffers and Solutions

Unless otherwise stated the recipes are from Sambrook et al. (1989)

**Alkaline lysis buffer**

(S. Tran, pers. comm)

[per 5 mL of 1% (w/v) SDS, 0.2 M NaOH]

To 4.4 mL Milli-Q H$_2$O add: 100 µL 10 M NaOH
500 µL of 10% (w/v) SDS

Make fresh as required.

**0.5 M EDTA (pH 8.0)**

Dissolve 186.1 g disodium ethylenediaminetetraacetate.2H$_2$O in 800 mL water. Adjust pH to 8 with solid NaOH, and adjust volume to 1000 mL.

**IPTG (24 mg.mL$^{-1}$)**

Dissolve 240 mg Isopropylthio-β-D-galactosidase (IPTG, Gold Biotechnology, cat# 12481C) in 10 mL water. Sterilise by filtration through a 0.45 µm syringe filter and store -20 °C. Used at final concentration of 24 µg.mL$^{-1}$ in LB media.

**5 M NaCl**

Dissolve 292.2 g NaCl in 800 mL H$_2$O. Adjust volume to 1000 mL.

**Neutralisation buffer**

(S. Tran, pers. comm)
[per 100 mL of 3 M Potassium acetate, 5% (v/v) Formic acid]

To 80 mL distilled water add:  
29.44 g CH$_3$COOK  
5 mL Formic acid

Adjust volume to 100 mL and autoclave.
ONPG (o-Nitrophenyl-β-D-Galactopyranoside)

Dissolve 20 mg of ONPG (Sigma cat# N-1127) in 5 mL of Z buffer. Prepare fresh on the day of the assay. Protect from light.

0.9% Saline

(per 100 mL) 0.9 g NaCl

99.5 mL distilled H₂O

10% (w/v) SDS

Add 100 g of Sodium dodecyl sulfate (SDS) to 800 mL water, heat to 68 °C to dissolve. Cool and then adjust pH to 7.2 with concentrated HCl, adjust volume to 1000 mL.

3 M Sodium acetate (pH 5.2)

Dissolve 246.09 g CH₃COONa in 800 mL distilled water. Adjust pH to 5.2 with glacial acetic acid. Adjust volume to 1000 mL.

1 M Sodium carbonate

Dissolve 21.2 g Na₂CO₃ in 150 mL distilled water. Autoclave and store at room temperature.

0.46 M Sodium carbonate

Dissolve 9.75 g Na₂CO₃ in 200 mL distilled water. Autoclave and store at room temperature.
**20× SSC (pH 7)**

(1× SSC is 0.15 M NaCl, 15 mM Sodium citrate [pH 7.0])

To 800 mL of water add:  
175.3 g NaCl  
88.2 g Tri-Sodium citrate

Adjust to pH 7.0 with concentrated HCl, adjust volume to 1000 mL, autoclave.

**Tracking dye**

(per 9 mL)

To 6 mL of Milli-Q H$_2$O add:  
30 mg Bromothymol blue  
3 mL glycerol

**1 M Tris HCl (pH 8.0)**

Add 121.1 g Tris (hydroxylmethly)aminomethane to 800 mL water, add approximately 42 mL 12 M HCl to adjust to pH 8, adjust volume 1000 mL.

**50× Tris-acetate buffer (TAE)**

(1× TAE is 40 mM Tris-acetate, 1 mM EDTA)

(per litre)

242 g Tris base  
57.1 mL Glacial acetic acid  
25 mL of 0.5 M EDTA

**X-gal (20 mg.mL$^{-1}$)**

Dissolve 200 mg 5-Bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal, Gold Biotechnolgy, cat# X428C) in 10 mL dimethylformamide. Wrap tube in tinfoil to prevent light damage and store at -20 °C. Used at a final concentration of 40 µg.mL$^{-1}$ in LB, TY, and G/RDM media.

200
**Z Buffer (β-galactosidase assay buffer)**

(121)

(per litre)

To 950 mL of distilled H2O add:

- 8.6 g Na₂HPO₄
- 5.4 g NaH₂PO₄
- 0.750 g KCl
- 0.246 g MgSO₄·7H₂O

Adjust pH to 7.0, adjust volume to 1000 mL. Do not autoclave. Store at 4 °C.
C

Additional AHL Data

see accompanying CD /Appendix C/Appendix C.ppt
D

Additional transmission electron microscopy

R7A

see accompanying CD /Appendix D/R7A

R7A(pJR174)

see accompanying CD /Appendix D/R7A(pJR174)

R7A*ΔtraR(pJR174)

see accompanying CD /Appendix D/R7AΔtraR(pJR174)