# Discovery and characterisation of tripartite Integrative & Conjugative Elements

**Timothy L Haskett**

**BSc (Hons)**

This thesis is presented for the degree of

Doctor of Philosophy

Of

Murdoch University

2018

I declare that this thesis is my own account of my research and contains as its main content work which has not previously been submitted for a degree at any tertiary education institution.

_____

Timothy Haskett

## Abstract

Bacterial integrative & conjugative elements (ICEs) are chromosomally-integrated DNA islands that excise to form circular molecules capable of horizontal self-transmission via conjugation (cell-to-cell contact). Symbiosis ICEs, such as ICE*Ml*Sym$^{R7A}$ of *Mesorhizobium loti,* are a group of ICEs that carry genes enabling rhizobial bacteria to engage in $N_2$-fixing symbioses with leguminous plants. Transfer of symbiosis ICEs can convert non-symbiotic rhizobia into legume symbionts in a single evolutionary step.

In this thesis, a novel form of "tripartite" ICE (ICE$^3$) is reported that exists as three entirely separated regions of DNA residing in the chromosomes of genetically diverse $N_2$-fixing *Mesorhizobium* spp. These ICE$^3$ regions did not excise independently, rather through multiple recombinations with the host chromosome they formed a single contiguous region of DNA prior to excision and conjugative transfer. Upon integration into a recipient chromosome, the ICE$^3$ recombined the recipient chromosome to disassemble into the tripartite form. These recombination reactions were catalysed by three Integrase proteins IntG, IntM, and IntS, acting on three associated integrase attachment sites. The "excisive" recombination reactions (i.e. assembly and excision) were stimulated by three recombination directionality factors RdfG, RdfM, and RdfS. Expression of ICE$^3$ transfer and conjugation genes were found to be induced by quorum-sensing. Quorum-sensing activated expression of *rdfS*, and in turn RdfS stimulated transcription of both *rdfG* and *rdfM*. Therefore, RdfS acts as a "master controller" of ICE$^3$ assembly and excision. A model for ICE$^3$ recombination and transfer is presented in this thesis.

The conservation of gene content between symbiosis ICE and ICE$^3$ indicated that these elements share a common evolutionary history. However, the persistence of ICE$^3$ structure in diverse mesorhizobia is perplexing due to its seemingly unnecessary complexity. Bioinformatic comparisons of ICE and ICE$^3$ indicated that the tripartite configuration itself may provide selective benefits to the element, including enhanced host range, host stability and resistance to destabilization by tandem insertion of competing integrative elements.

In congruency with ICE*Ml*Sym$^{R7A}$, ICE$^3$ acquisition can convey upon recipients the ability to form $N_2$-fixing symbiosis with the host-legume of the ICE$^3$ donor. Interestingly, the effectiveness of $N_2$-fixation may be impaired. The consequences of the emergence of sub-optimal $N_2$-fixing strains following ICE$^3$ transfer in agriculture is discussed. If ICE$^3$ transfer poses a barrier for future inoculation success, the elucidation of the mechanism of ICE$^3$ assembly, excision, and transfer will allow for the development of strategies for management.

.

# Table of Contents

Abstract ..... iii

Table of Contents ..... v

Acknowledgements ..... ix

Publications arising from this thesis ..... xi

List of commonly used abbreviations ..... xii

Chapter 1 ..... 1

1.1. The symbiosis between rhizobia and legumes ..... 2

1.1.1. Establishment of $N_2$-fixing symbiosis ..... 2

1.1.2. The accessory nature of rhizobial symbiosis genes ..... 5

1.1.3. *Mesorhizobium* ..... 8

1.2. Integrative and conjugative elements (ICEs) ..... 10

1.2.1. Site-specific integration and excision ..... 10

1.2.2. Conjugative transfer ..... 13

1.2.3. Exploitation of conjugation systems by IMEs/CIMEs ..... 16

1.2.4. Maintenance and stability ..... 19

1.2. Symbiosis ICEs ..... 21

1.3.1. Discovery of ICE*Ml*Sym$^{R7A}$ ..... 21

1.3.2. Integration, excision and transfer of ICE*Ml*Sym$^{R7A}$ ..... 21

1.3.3. Regulation of excision and transfer of ICE*Ml*Sym$^{R7A}$ ..... 23

1.3.4. Diversity of symbiosis ICEs ..... 27

1.3.5. Core features of *Mesorhizobium* symbiosis ICEs ..... 28

1.4. Emergence of novel *Mesorhizobium* spp nodulating the pasture legume *Biserrula pelecinus* ..... 30

1.5 Aims of this thesis ..... 32

Chapter 2 ..... 35

2.1. Media and growth conditions ..... 36

2.2. Bacterial strains in this thesis ..... 36

2.3. Construction of plasmids ..... 41

2.4. General molecular techniques ..... 49

2.4.1. Isolation of genomic and plasmid DNA ..... 49

2.4.2. End-point PCR ..... 49
2.4.3. Agarose and Eckhardt gel electrophoresis ..... 50
2.4.4. Restriction endonuclease digestions, ligations and Gibson assembly ..... 51
2.4.5. Preparation and transformation of competent cells ..... 51
2.4.6. Conjugal transfer of plasmids ..... 52
2.5. DNA sequencing and genome assembly ..... 53
2.6. *Mesorhizobium* mating experiments ..... 54
2.7. Quantitative PCR assays for ICE excision ..... 54
2.8. β-galactosidase assays ..... 58
2.9. Melanin deposition assays ..... 58
2.10. Bioassays for detection of AHLs ..... 58
2.10.1. CV026 bioassays ..... 58
2.10.2. AHL inactivation assays ..... 59
2.11. RNA Sequencing ..... 59
2.11.1. Isolation of total RNA ..... 60
2.11.2. cDNA library construction and sequencing ..... 60
2.11.3. Read mapping, read counting and statistical analyses ..... 61
2.12. Glasshouse procedures ..... 62
2.13. General bioinformatics and statistics ..... 63
Chapter 3 ..... 64
3.1. Introduction ..... 65
3.2. Results ..... 66
3.2.1. Three co-transferrable DNA regions in the WSM1271 chromosome ..... 66
3.2.2. Three integrases and three pairs of attachment sites ..... 68
3.2.3. Coordinated formation of three pairs of *attB* and *attP* sites ..... 72
3.2.4. Formation of three pairs of *attP* and *attB* sites requires IntG, IntM and IntS ..... 73
3.2.5. Reconstruction of ICE*Mc*Sym$^{1271}$ integration and disassembly pathways ..... 75
3.2.6. Symbiotic phenotypes of R7ANS ICE*Mc*Sym$^{1271}$ exconjugants ..... 79
3.3. Discussion ..... 80
Chapter 4 ..... 84
4.1. Introduction ..... 85

4.2. Results and discussion ..... 85
4.2.1. ICE$^3$s exist in genetically diverse mesorhizobia ..... 85
4.2.2. Conservation of symbiosis ICE and ICE$^3$-α genes indicates a common evolutionary history ..... 92
4.2.3. ICE$^3$ regions γ and β carry disparate genetic cargo ..... 97
4.2.4. Evolution of a new ICE3-γ integration site through evolutionary recruitment of a serine recombinase ..... 100
4.2.5. Proposed selective advantages of the ICE$^3$ configuration ..... 103
4.3. Summary ..... 108
Chapter 5 ..... 109
5.1. Introduction ..... 110
5.2. Results ..... 113
5.2.1. RdfG and RdfM are required for excisive IntG and IntM-mediated recombination ..... 113
5.2.2. Quorum sensing stimulates all three excisive Int-mediated recombination reactions ..... 116
5.2.3. Dissection of quorum sensing-induced ICE$^3$ excision using RNA deep sequencing ..... 117
5.2.4. *rdfS* is required for all three excisive Int-mediated recombination reactions 124
5.2.5. Overexpression of *rdfS* stimulates expression of *rdfG* and *rdfM* ..... 125
5.3. Discussion ..... 126
Chapter 6 ..... 133
6.1. Introduction ..... 134
6.2. Results ..... 136
6.2.1. Three pairs of LuxR-LuxI QS loci in WSM1271 ..... 136
6.2.2. CV026 bioassays for AHL production by TraI1 and MbrI ..... 138
6.2.3. The α/β-fold family hydrolase AhlI inactivates TraI1-derived AHLs in diverse mesorhizobia ..... 141
6.2.4. TraR1 and TraR2 are activated by TraI1-derived AHLs ..... 146
6.2.5. MbrR regulates melanin biosynthesis ..... 150
6.2.6. MbrR is activated by TraI1 or MbrI-derived AHLs ..... 152
6.3. Discussion ..... 153
Chapter 7 ..... 160

7.1. ICE[3] assembly, excision, integration and dissassembly ..... 161

7.2. Regulation of ICE[3] excision and transfer ..... 162

7.3. Consequences for ICE and ICE[3] transfer in agriculture ..... 167

7.4. Conclusion ..... 171

Bibliography ..... 172

## Acknowledgements

I would like to express my sincerest thanks to my supervisors Dr Jason Terpolilli, Dr Josh Ramsay, and Prof Graham O'Hara. All three have spent considerable time sharing with me their wealth of knowledge and have ensured that my research has remained in focus over the course of my PhD studies. Their habitual and continued willingness to unearth opportunities for me to publish my work, to engage in new collaborations, and to share my research at both national and international conferences will undoubtedly provide some key stepping stones in the future development of my career. I would also like to extend a warm appreciation to my former Honours degree supervisor Dr Wayne Reeve. Much of my understanding of concepts underlying techniques and protocols used in molecular biology were developed under his guidance.

I would like to thank all my colleagues and friends from the Center for Rhizobium Studies, and the Ramsay Lab for providing a stimulating, stress-free, and enjoyable work environment over the past three years. I'll particularly miss the Friday afternoon home-brew sessions with Ron, Brad and Rob. I extend a special thankyou to Karina Yui, Callum Verdonk, and Riley Murphy for various assistance in the laboratory.

Much of work performed in the early publications arising from this thesis was crucially influenced by the input and assistance of Dr. John Sullivan and Prof. Clive Ronson at the University of Otago in New Zealand. For this, I offer my sincerest appreciation and thanks.

I also extend sincere a thanks to Prof Phil Poole for allowing me to undertake some of my PhD experiments at The University of Oxford, and to Dr. Vinoy Ramachandran for sharing with me his expertise during this visit. Thanks also to my friends at the rhizosphere lab for making this experience all the more enjoyable.

I formally thank the Grains Research Development Corporation for the provision of an RTP top-up scholarship for the research presented in this thesis.

Lastly, I would like to thank my family and friends for their continued support. Completing a PhD becomes a far less arduous task knowing that I am surrounded by such fun-loving, relaxed, and generous people.

## Publications arising from this thesis

**1.** Haskett TL, Terpolilli JJ, Ramachandran VK, Verdonk CJ, Poole PS, O'Hara GW, & Ramsay JP (2018) Sequential induction of three recombination directionality factors directs assembly of tripartite integrative and conjugative elements. PLoS Genetics. DOI: 10.1371/journal.pgen.1007292.

**2.** Brewer RJM, Haskett TL, Ramsay JP, O'Hara GW, & Terpolilli JJ (2017) Complete genome sequence of *Mesorhizobium ciceri* bv. biserrulae WSM1497, an efficient nitrogen-fixing microsymbiont of the forage legume *Biserrula pelecinus*. Genome Announcements 5(35):e00902-00917.

**3.** Haskett TL, et al. (2016) Assembly and transfer of tripartite integrative and conjugative genetic elements. PNAS 113(43):12268-122

**4.** Haskett T, et al. (2016) Complete genome sequence of *Mesorhizobium ciceri* bv. biserrulae Strain WSM1284, an efficient nitrogen-fixing microsymbiont of the pasture legume *Biserrula pelecinus*. Genome Announcements 4(3):e00514-00516.

**5.** Haskett TL, et al. (2017) Evolutionary persistence of tripartite integrative and conjugative elements. Plasmid 92:30-36.

# List of commonly used abbreviations

| | |
|---|---|
| 3-oxo-$C_6$-HSL | *N*-(3-oxohexanoyl)-*L*-homoserine lactone |
| 3-oxo-$C_{12}$-HSL | *N*-(3-oxododecanoyl)-*L*-homoserine lactone |
| AHL | *N*-acyl-homoserine lactone |
| Ala | 5-aminolevulinic acid |
| *attB* | Chromosomal ICE integration attachment site |
| *attL* | Integrated ICE attachment site |
| *attP* | Circular ICE attachment site |
| *attR* | Integrated ICE attachment site |
| BLAST | Basic local alignment search tool |
| bp | Base pair |
| Cb | Carbenecillin |
| CIME | *cis*integrative and mobilisable element |
| DSF | Diffusible signal factor |
| DMSO | Dimethyl sulfoxide |
| ERIC | Enterobacterial Repetitive Intergenic Consensus |
| FIS | Factor for inversion stimulation |
| GFP | Green fluroescent protein |
| Gm | Gentamycin |
| GMP | Guanosine mono-phosphate |
| ICE | Integrative and conjugative element |
| $ICE^3$ | Tripartite integrative and conjugative element |
| IHF | Integration host factor |
| IME | Integrative and mobilisable element |
| Int | Integrase |
| IPTG | Isopropyl β-D-1-thiogalactopyranoside |
| Kb | Kilobase pair |
| Km | Kanamycin |
| LB | Luria Burtani |
| LSD | Least significant difference |
| MGE | Mobile genetic element |
| MGT | Mean generation time |

| | |
|---|---|
| MOPS | 3-(N-morpholino)propanesulfonic acid |
| Nm | Neomycin |
| OD | Optical density |
| PCR | Polymerase chain reaction |
| *PtraI1* | *traI1* promoter |
| qPCR | Quantitative PCR |
| QS | Quorum sensing |
| RAPD | Randomly amplified polymorphic DNA |
| RCR | Rolling circle replication |
| RDF | Recombination directionality factor |
| RDM | Rhizobium defined media |
| RHC | Root hair curling |
| rSAP | Alkaline shrimp phosphatase |
| SD | Standard deviation |
| Sm | Streptomycin |
| Sp | Spectinomycin |
| T4CP | Type IV coupling protein |
| T4SS | Type IV protein secretion system |
| TA | Toxin-antitoxin |
| TAE | Tris-acetate buffer |
| TBE | Tris-borate buffer |
| Tc | Tetracycline |
| TPM | transcripts per million reads |
| TY | Tryptone yeast |

# Chapter 1.

## General Introduction and Literature Review

## 1.1. The symbiosis between rhizobia and legumes

Although dinitrogen ($N_2$) is abundant in Earth's atmosphere, the inert nature of this molecule renders it metabolically inaccessible to most organisms. A group of soil-dwelling α and β-proteobacteria termed rhizobia possess the remarkable ability to infect specialised nodule cells that form on the roots of leguminous plants and within these nodules, terminally differentiate into a bacteroid capable of reducing atmospheric $N_2$ into $NH_3$ (5-8). The interaction between rhizobia and legumes is considered symbiotic because the legume host provides the bacteroids with carbon in the form of dicarboxylic acids, and in return, the bacteroids secrete $NH_3$ to the plant root cells where it is assimilated into amino-acids and distributed via the xylem (5, 9).

### *1.1.1. Establishment of $N_2$-fixing symbiosis*

To establish a successful $N_2$-fixing symbiosis, rhizobia must colonise the legume roots and infect the nodule cells. Although multiple modes of rhizobial root infection have been described, the traditional "root hair curling" (RHC) mode of infection is the most common and best characterised (6, 7). RHC infection (Fig 1.1) is initiated when rhizobia recognise the presence of legume-derived chemical signals such as flavonoids, methoxychalcones, aldonic acids and betaines secreted in the root exudate (10-14). Rhizobial NodD receptors detect this signal and respond by directing transcription of various *nod* genes involved in production and secretion of lipochito-oligosaccharide molecules termed Nod factors (15-17). Nod factors trigger several RHC infection-related responses in the legume host, including $Ca^{2+}$ oscillations, $Ca^{2+}$ influx, root-hair curling and infection thread formation (6, 7, 18, 19). Many legumes exhibit

stringent requirements for specific confirmations or concentrations of Nod factors, which serve as a checkpoint for partner-choice in symbiosis (15, 16, 20).

Following plant recognition, rhizobia adhered to the root hair become entrapped when the root hair curls, forming a “Sherpard’s crook”-like structure. These entrapped rhizobia are directed into an invagination in the root hair cell wall where an infection thread begins to develop (6, 7). Rhizobia occupy the infection thread by cell division and are eventually released in an infection droplet into the microaerobic nodule primordium where bacteroid differentiation occurs (8, 21). In most well-studied rhizobia, the low-oxygen environment of the nodule cell triggers the activation of the $N_2$ fixation regulator NifA, which transcriptionally activates a suite of *nif* genes required for the assembly of the nitrogenase enzyme complex. In some strains of rhizobia, FixJ and FixK are additionally activated in response to low oxygen, which in turn up-regulate expression *fix* genes involved in the assembly of bacteroid respiratory systems that fuel nitrogenase (5, 9, 22-26).

Under laboratory conditions where environmental factors can be standardised, legume-rhizobia symbioses may exhibit variable $N_2$-fixation effectiveness outcomes (i.e. amounts of $N_2$ fixed) (27). $N_2$-fixation effectiveness may range from ineffective (nodulation without $N_2$-fixation) to effective (all N demands of the legume are met) $N_2$-fixation. The amount of $N_2$ fixed by rhizobia during a symbiosis is largely dependent on the legume host, the strain of rhizobium and the biological outcome of their interaction (27, 28). Although many strains of rhizobia may nodulate a specific host, only a small subset of strains will fix optimal amounts of $N_2$ (29, 30).

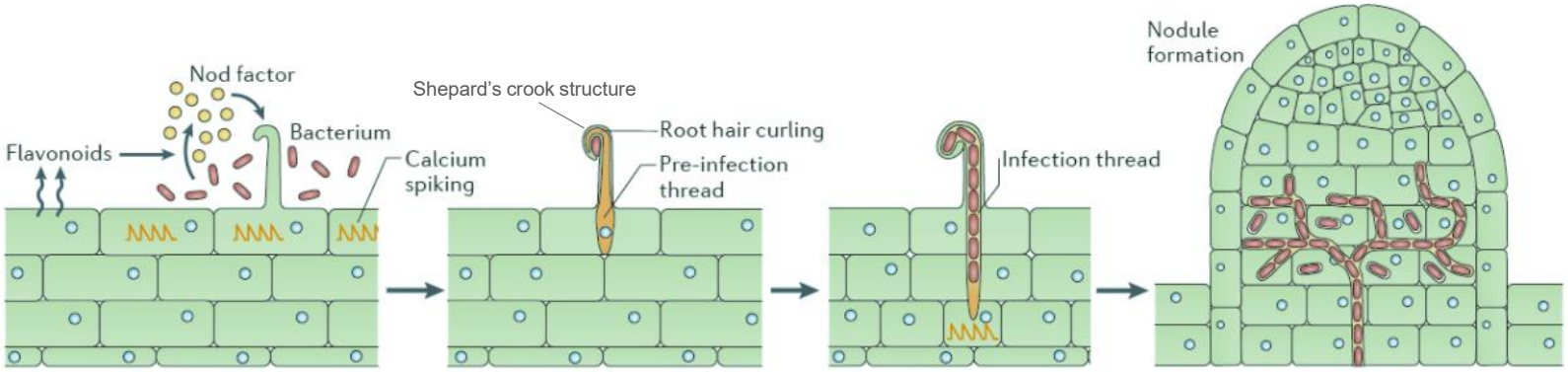


**Figure 1.1. The establishment of rhizobia-legume symbioses.** The classical root hair curling mode of infection in rhizobia-legume symbioses is initiated when rhizobia perceive flavonoid or other signalling molecules secreted into the rhizosphere by the legume. These flavonoids are detected by rhizobial NodD receptors which trigger the expression of nodulation genes required for the biosynthesis and secretion of Nod factors. Nod factors secreted by rhizobia adhered to the legume root-hairs are detected by the legume cell-surface receptors, leading to $Ca^{2+}$-spiking in epidermal and cortical cells. Perception of Nod factors also stimulates root hair curling, trapping rhizobia in a Shepard's crook-like structure and stimulating the development of infection threads. Rhizobia proceed down the developing infection thread via cell division and are released in an infection droplet into developing nodule cells where they differentiate into bacteroids, capable of undergoing $N_2$-fixation. This figure was reproduced from reference (31).

### *1.1.2. The accessory nature of rhizobial symbiosis genes*

Genome sequencing efforts have revealed that strains within a bacterial species typically carry a set of conserved “core” genes that are essential for their growth and a selection of horizontally acquired “accessory” genes that that convey some beneficial adaptive trait (32-35). The core and accessory genes of a bacterial species are collectively termed the pan genome (34).Rhizobial *nod, nif* and *fix* genes (hereby referred to as the symbiosis genes) required to engage in $N_2$-fixing symbioses comprise part of the accessory genome and have probably transferred horizontally to many taxa throughout evolutionary history (33, 36-40). However, the ability to fix $N_2$ with legumes is restricted to only 14 distinct genera of α and β-proteobacteria (41-44), suggesting that transfer of symbiosis genes between distantly related bacteria is rare. In contrast, horizontal transfer of symbiosis genes within rhizobial genera is considered one of the major forces driving genetic diversity and may result in the evolution of new $N_2$-fixing species (45-50). Symbiosis genes transfer freely in rhizobial communities because they are typically carried on mobile genetic DNA elements (MGEs) (51). For most rhizobia, including *Rhizobium* and *Sinorhizobium,* these MGEs are large (>100-kb) plasmids (39, 52, 53) (Fig. 1.2B). Plasmids often encode genes necessary for horizontal self-transmission via conjugation (transfer via cell-to-cell contact), but some plasmids lacking these conjugation genes rely on extrinsic factors for their mobilisation (53). Horizontal transfer of both “conjugative” and “mobilisable” plasmids carrying symbiosis genes has been demonstrated under controlled conditions and conveys upon some recipients the ability to

nodulate and fix $N_2$ in symbiosis with target legume. (54-58). However, the effectiveness of $N_2$fixation of these recipients is commonly impaired compared to the plasmid donor strain.

In contrast to *Rhizobium* and *Sinorhizobium,* rhizobia belonging to the *Mesorhizobium, Bradyrhizobium* and *Azorhizobium* genera typically carry their symbiosis genes on large (>400-kb) chromosomally-integrated genomic DNA islands termed symbiosis islands (59-63) (Fig 1.2A). Symbiosis islands of *Mesorhizobium* and *Azorhizobium* strains can be conjugatively transferred into other symbiotic or non-symbiotic rhizobia and confer on recipients an ability to nodulate or in some cases fix $N_2$ with target legumes (59, 60).

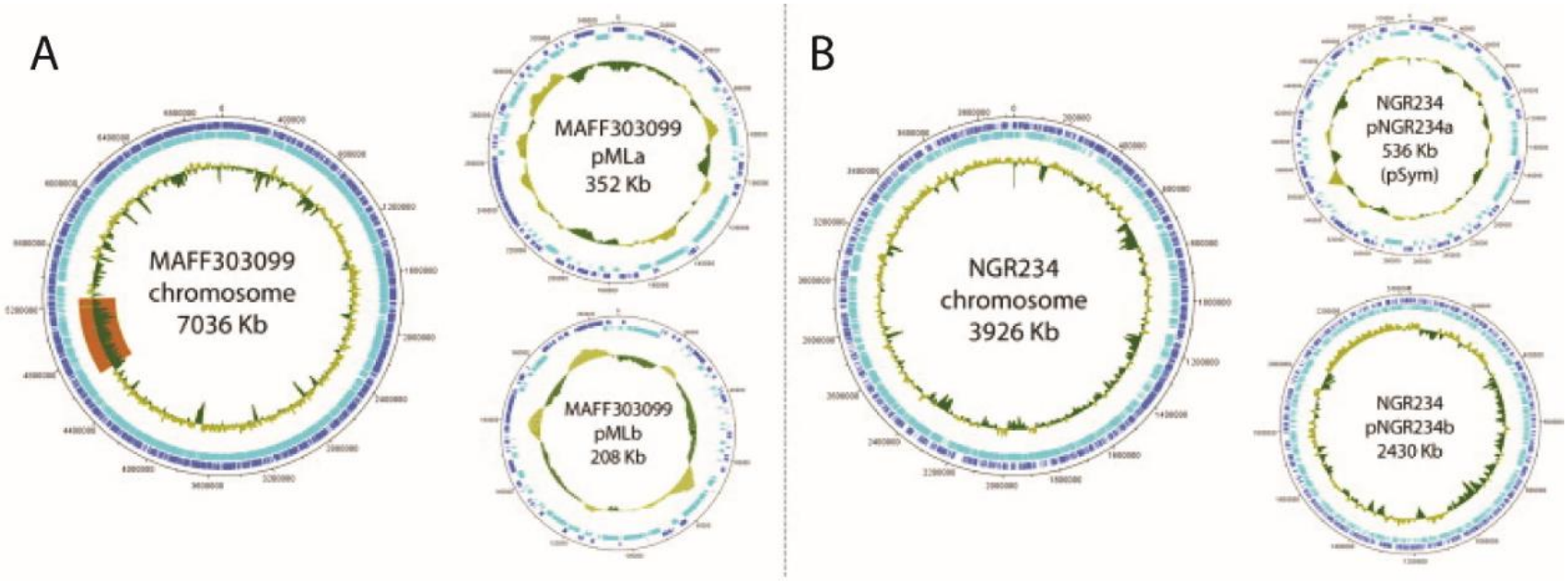


**Figure 1.2. Rhizobial genome architecture.** Inner circles represent GC%, and the second-third most inner circles represent predicted ORFs. The outermost circle details the genome coordinates. (A) The *M. huakii* (formerly *M. loti*) MAFF303099 genome consists of a ~7 mb chromosome and and two extrachromosomal plasmids; pMLa (~352 kb) and pMLb (~208 kb). The symbiosis genes are located on a chromosomally-integrated ~611-kb ICE. (B) The *Sinorhizobium fredii* NGR234 genome is composed of a ~3.9-mb chromosome and two plasmids; pNGR234a (~536 kb), which carries most of the symbiosis genes, and pNGR234b (~2.4 mb). This figure was reproduced from reference (49).

### *1.1.3. Mesorhizobium*

Symbiosis islands were first identified in the *Mesorhizobium* genus which was recognised in 1997 following the reclassification of five former *Rhizobium* species (64). Representatives of the genus are characterised by an intermediate growth rate (MGT 4-8 h) and share significant 16s rDNA homology (65). Based on 16s rDNA sequences, the *Mesorhizobium* genus is the taxonomic intermediate of *Bradyrhizobium* and *Rhizobium* (Fig 1.3). Mesorhizobia establish $N_2$-fixing symbioses with temperate, tropical, sub-tropical and arctic legumes (66). The symbiotic hostrange for *Mesorhizobium* spp. can be broad. *M. loti* NZP2037, for instance, nodulates *Leuceana leucocephala, Carmichaelia flagelliformis, Ornithopus sativus, Clianthus puniceus*, *Vigna* spp. and at least 10 *Lotus* spp. (67-70). Other mesorhizobia exhibit a much narrower range, such as *M. ciceri* CC1192, which has only been confirmed to nodulate *Cicer arietinum* (chickpea) (63, 71). Importantly, not all *Mesorhizobium* spp. carry their symbiosis genes in their chromosomes and some entirely lack symbiosis genes (72, 73). Nevertheless, the existence of chromosomally-encoded symbiosis genes appears to be prevalent in the *Mesorhizobium* genus and extensive characterisation of the paradigm symbiosis island from *M. loti* R7A, ICE*Ml*Sym$^{R7A}$, has revealed that these elements belong to the most abundant class of conjugative MGEs in bacteria, the integrative and conjugative elements (ICEs) (74-76).

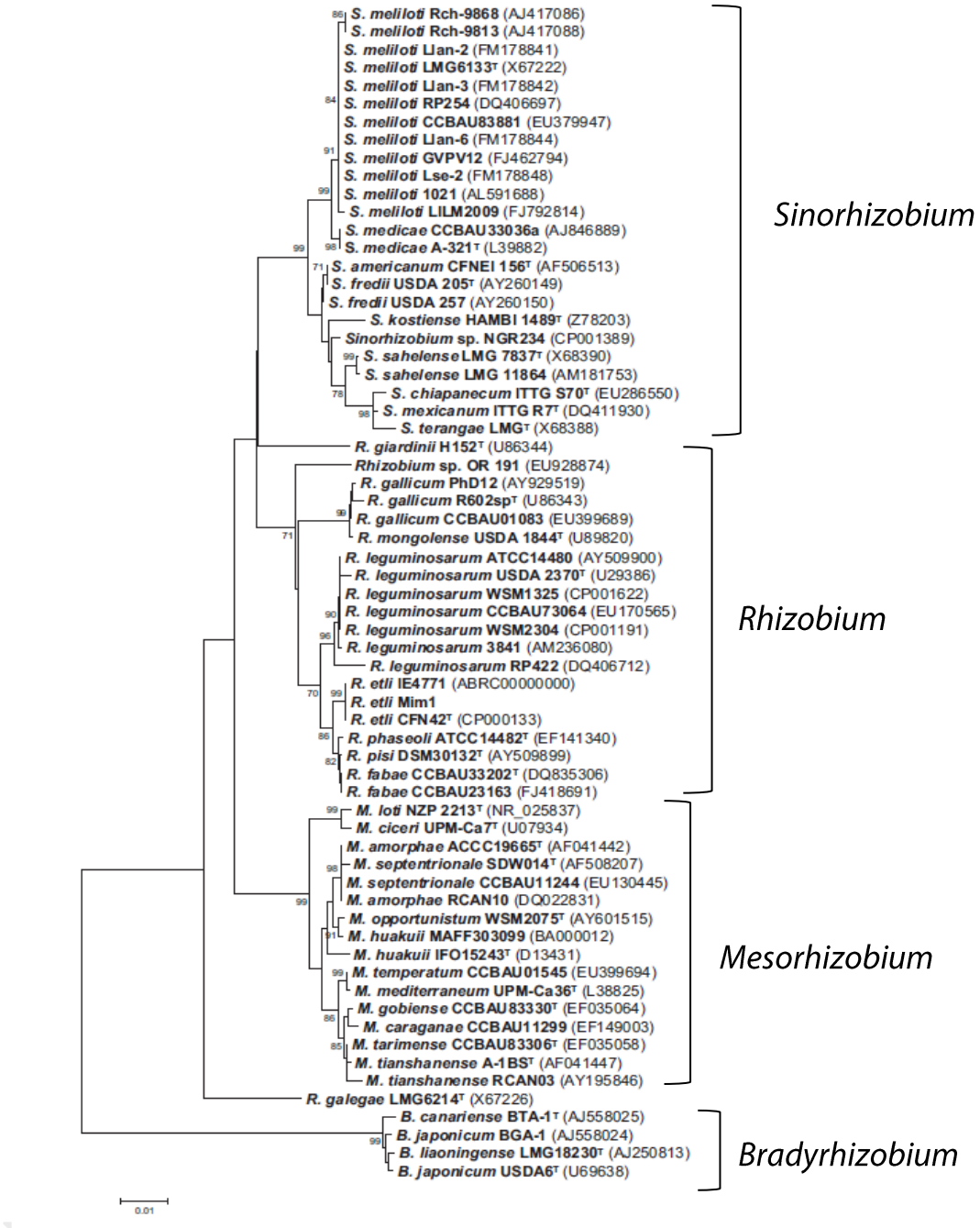


**Fig 1.3. Maximum likelihood phylogeny of four rhizobial genera based on 16s rDNA sequence.** 16s rDNA Phylogenetic analysis places the *Mesorhizobium* genus taxonomically between *Bradyrhizobium* and *Rhizobium* genera. This figure was reproduced from reference (65).

## 1.2. Integrative and conjugative elements (ICEs)

ICEs are regions of DNA that primarily reside integrated within bacterial genomes, but can excise to form a circular entity prior to conjugative transfer (77, 78). Thus, ICEs combine features of bacteriophages, transposons and conjugative plasmids to maximise their propagation by both vertical and horizontal modes of descent (79, 80). ICEs are currently grouped into 28 families (81), each carrying a conserved compliment of "cargo" genes that may convey a fitness benefit upon the host (77, 78). For example, the SXT/R931-family ICEs of *Vibrio cholerae* carry genes conveying multi-drug resistance (82, 83), the ICE*clc*-family of *Pseudomomas putida* carry genes required for chlorocatechol catabolism (84), PAPI-1 ICE of *P. aeruginosa* carries genes required for pathogenicity (85) and ICE*Mc*Sym$^{R7A}$ of *M. loti* R7A carries genes essential to the establishment of $N_2$-fixing symbioses with *Lotus* spp. (60). ICE transfer allows bacteria to acquire complex genetic traits in a single evolutionary step and has a major impact on bacterial adaptation and evolution (80).

### *1.2.1. Site-specific integration and excision*

Most studied ICEs have evolved to conservatively integrate (i.e. integration with no loss nor gain of DNA) into the 3'-end of highly conserved bacterial genes, presumably to maximise their host-range and reduces the fitness cost of integration (86, 87). For example, ICE*Bs1* of *Bacillus subtlis* (88, 89) and PAPI-1 integrate into the amino acyl tRNA genes *leu*-tRNA and *lys*-tRNA,

respectively while the SXT-R391-family ICEs integrate into peptide chain release factor 3 (*prfC*). Some ICEs, such as Tn*916* of *Enterococcus faecalis,* are less stringent in their specificity for integration, targeting AT-rich regions of the genome (90). Regardless of the target site for integration, ICEs typically catalyse their integration and excision from bacterial genomes with the aid of a self-encoded DNA recombinase protein. Recombinase proteins may belong to one of three families; serine recombinases, DDE recombinases, or tyrosine recombinases (also termed integrases). Each family name reflects the conserved amino acid residue(s) in the active site domain (91-93). Serine and DDE-motif-containing recombinases that catalyse excision and integration of ICEs have only been identified in the Firmicutes phylum of bacteria, whereas tyrosine recombinases involved in ICE excision and integration are far more widely distributed (78, 91, 94).

The most extensively characterised tyrosine recombinase is that of bacteriophage λ (λ-Int), which forms a nucleoprotein intasome complex with the *E. coli*-encoded integration host factor IHF and factor for inversion stimulation FIS to catalyse integration of phage λ into the *E. coli* chromosome (95-102). Integration of the phage occurs via a strand-exchange reaction involving the cleavage and formation of a Holliday junction between two short (15 bp) imperfect direct-repeat “core” attachment (*att*) sites located on the circular phage (*attP*) and chromosomal DNA (*attB*) (Fig 1.4). Importantly, the 15-bp core *attP* site located on the phage genome is structurally distinct from the *attB* site and comprises the central region of the complete 240-bp *attP* site containing at least 16 binding sites for proteins involved in recombination (95, 103). Following the strand exchange between *attP* and *attB* sites, the

integrated λ phage (termed a prophage) exists in the chromosome flanked by regions of DNA originating from upstream (for *attL*) or downstream (for *attR*) of the *attP* core sequence in the circular phage. Therefore, although the core *attL, attR*, *attP* and *attB* regions share a perfect or near-perfect core sequence, each region can be distinguished by the sequence of adjacent DNA. It should be noted that although integration of λ phage by Int requires the accessory proteins FIS and IHF (77, 78, 95), many ICEs can integrate into bacterial genomes via the activity of a lone ICE-encoded integrase (74, 104, 105). Nevertheless, the mechanism of integration and excision in most cases appears to follow the λ-phage model for integration and excision (92) and therefore the *att* site nomenclature developed in the study of the *E. coli* phage λ is used widely to describe attachment sites of many integrative elements.

Although integrases such as λ Int catalyse both phage integration (*attP and attB* → *attL and attR*) and excision (*attL and attR* → *attP and attB*) reactions, integration reaction is generally favoured in the absence of additional factors for most well-studied integrases (106). Small proteins termed recombination directionality factors (RDFs, also termed excisionases), may stimulate the excision of ICEs and other integrative elements (107). In the case of λ phage, overexpression of the Xis protein stimulates excision $10^6$-fold and simultaneously inhibits integration by binding a 35-bp regulatory element on the λ *attP* and *attL* site, causing a conformational change in the DNA that facilitates formation of the excisive intasome (108). Most studied RDFs stimulate excision by binding to *att* sites and bending the DNA into a conformation promoting excisive recombination (107), however, a subset of RDFs in the P2-class and 186-class phages termed Cox or ApI proteins,

respectively, are also able to regulate transcription from phage-encoded promoters that control expression of genes involved in lysogeny (109-114). Thus, RDFs can function in both catalytic and regulatory roles to facilitate the excision of various integrative elements.

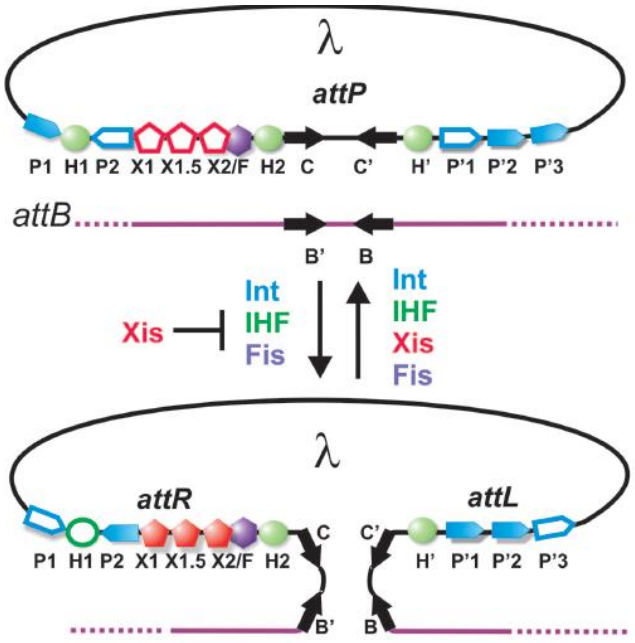


**Figure 1.4. Integration and excision of phage λ.** Integration of λ phage is mediated by recombination between two imperfect direct-repeat attachment (*att*) sites positioned on the circular phage (the core of *attP*) and host chromosome (*attB*). Integration is catalysed by the nucleoprotein intasome complex composed of λ integrase (λ Int), and the accessory proteins factor for inversion stimulation FIS and integration host factor IHF (95-102). λ phage proceeds in the reverse of the integration reaction and is catalysed by the same protein complex, however, the excision reaction is favoured in the presence of the excisionase Xis (108, 115).

### *1.2.2. Conjugative transfer*

Excision and circularisation is an essential prerequisite for conjugative transfer of ICEs. Like conjugative plasmids, ICEs typically encode a VirB/D4-like type IV secretion system (T4SS) (116). Evidence suggests that T4SSs of both classes of elements have moved between ICE and plasmid backbones during

their evolutionary history (116). Conjugative transfer of plasmids has been recently reviewed (117, 118). Conjugation initially requires the formation of the relaxosome, a complex containing a relaxase and various accessory DNA-binding proteins. The relaxosome recognises a cognate double-stranded origin-of-transfer (*oriT*) sequence on the conjugative element, nicking a single-strand of DNA at the *oriT* and forming a covalent phosphotyrosine bond with the 5'-end of the nicked single-strand of DNA (Fig 1.5). The relaxosome-DNA complex (also termed T-DNA) is recruited to the T4SS by a type-IV coupling protein (T4CP) and is translocated into a recipient cell where it is re-circularised at the *oriT* by the relaxosome (117, 118). Because integrase proteins only catalyse recombination between double-stranded DNA molecules (119), it is thought that the single-stranded ICE acts as template for lagging (second) strand synthesis prior to integration into the bacterial genome. In support of this notion, many ICEs encode a single-stranded origin of replication resembling those required by conjugative plasmids for the initiation of rolling-circle replication (RCR) (120, 121). Importantly, not all ICEs follow this traditional mechanism of conjugative transfer. Actinobacterial ICEs such as pSAM2 of *Streptomyces ambofaciens* replicate in the donor cell prior to conjugative transfer and are horizontally transferred as double-stranded molecules by TraB-like translocases (122, 123).

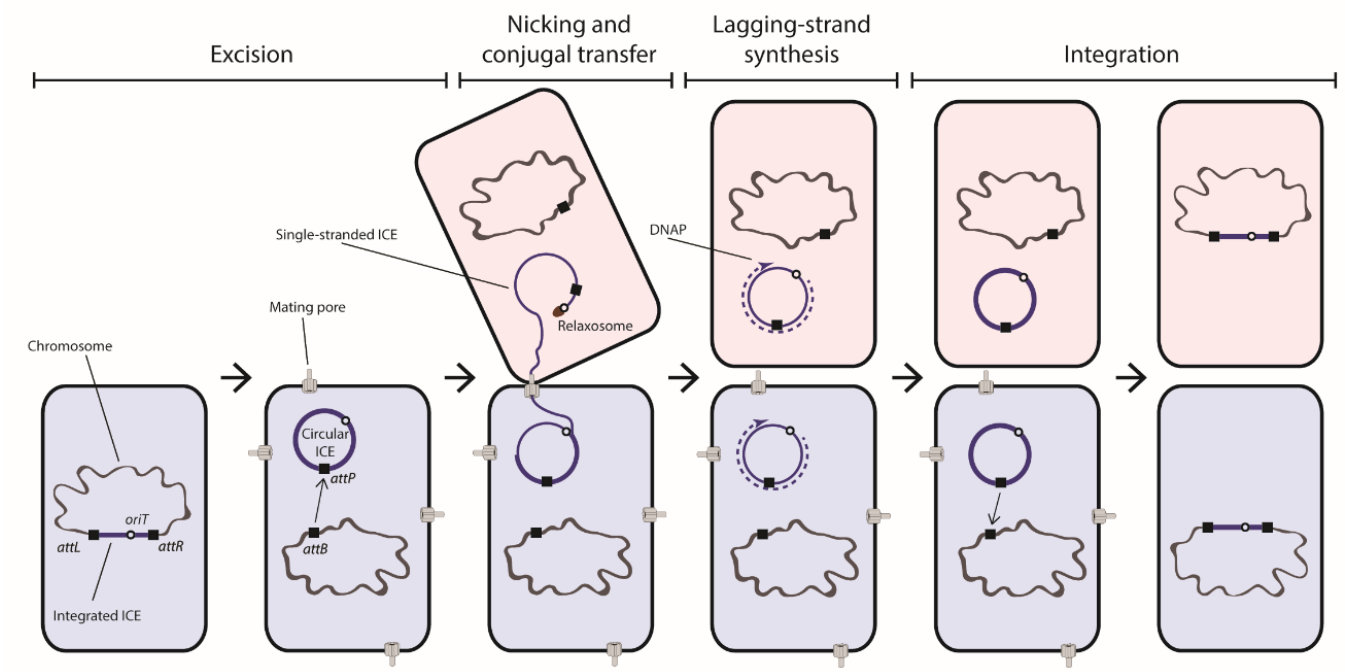


**Figure 1.5. Conjugative transfer of an ICE.** ICEs excise from the chromosome to form a circular plasmid-like entity prior to conjugative transfer (See Fig 1.3). Conjugative transfer is initiated by the relaxosome, a multiprotein complex composed of a relaxase and various accessory DNA-binding proteins. The relaxosome recognises the cognate origin of transfer (*oriT*) of the conjugative element, nicks a single-strand of DNA and forms a covalent bond with the 5'-end of the nicked DNA. The relaxosome-DNA complex is guided to the T4SS by a type IV coupling protein and is translocated into a recipient cell. Within the recipient cytoplasm, the single-stranded ICE is re-circularised by the relaxosome, and acts as template for lagging (second) strand synthesis prior to integration into the bacterial genome (117-119).

### *1.2.3. Exploitation of conjugation systems by IMEs/CIMEs*

Like other MGEs, ICEs display a modular structure in which genes and non-coding regions of DNA involved in similar functions are clustered together (77, 78, 94) (Fig 1.5A). ICEs typically harbour genetic modules required for integration and excision, conjugative transfer and host fitness (124-130). elements exist in bacterial genomes that lack one of more of these modules, but exploit excision and (or) conjugation genes encoded by other MGEs for their excision and transfer (94) (Fig 1.6). Integrative and mobilisable elements (IMEs) carry genes required for integration and excision, an *oriT* sequence and potentially a cognate conjugative relaxase, but lack other genes required for conjugative transfer (94, 126, 131, 132). Nevertheless, these elements can be mobilised by conjugative machinery encoded elsewhere in the bacterial genome. *Cis*-integrative and mobilisable elements (CIMEs) have also been described that carry genes required for integration and excision but may lack an *oriT* or other conjugative factor that cannot be extrinsically provided to permit conjugative transfer (133-136). These elements may “hitchhike” through the conjugal mating pore with invading ICEs following tandem insertion. Tandem insertions occur when invading ICEs integrate into the *attL* or *attR* site of a resident element occupying the cognate *attB* site of both elements. The newly formed tandem element comprises *attL* and *attR* sites derived from each element and a hybrid *attP*-like site derived from *attL* and *attR* of adjacent elements (132-135, 137-139). The outer-most distal *attL* and *attR* may recombine to excise the ICE-CIME composite element, which is then capable of conjugative transfer via the ICE-encoded machinery. ICE-CIME arrays can be highly unstable (133, 137, 138, 140), however, tandem arrays

of integrative elements generated in the laboratory have been shown to recombine resulting in accretion of the two elements and the evolution of a novel ICE (137, 141). The existence of IMEs and CIMEs that pirate bacterial conjugative systems highlights that the bacterial mobilome should be viewed as a DNA ecosystem where MGE are constantly adapting and evolving to compete for an environmental niche (126, 131, 142, 143).

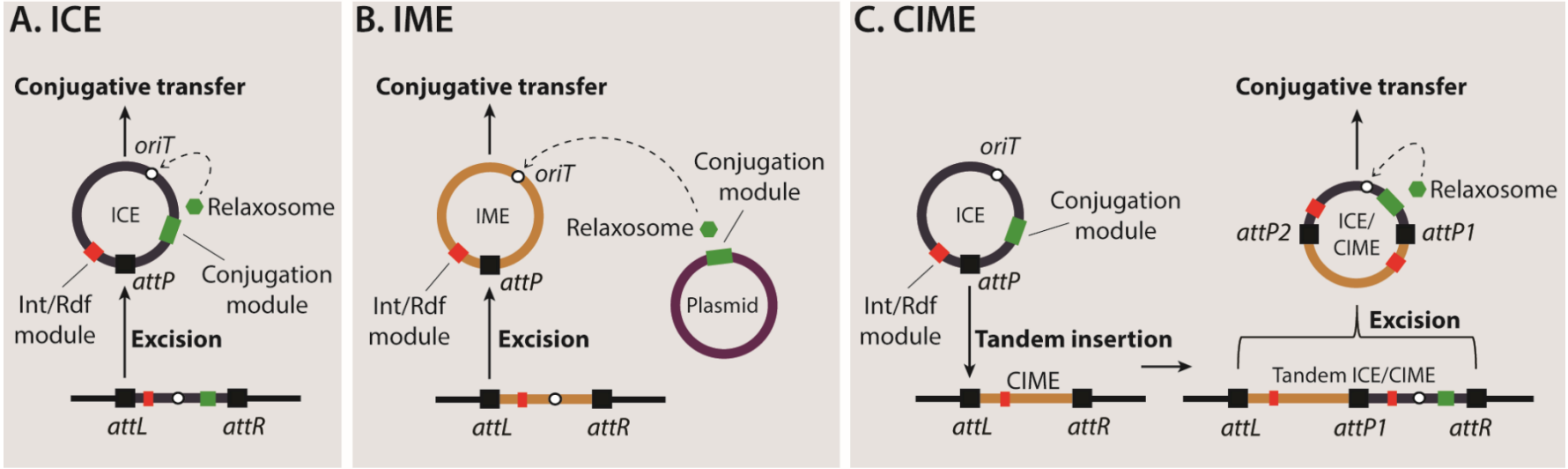


**Figure 1.6. Mobilisation strategies for ICE/IME/CIMEs.** (A) ICEs carry modules required for integration, excision and conjugative transfer, and are therefore self-transmissible. (B) Integrative and mobilizable elements (IMEs) typically carry modules for integration and excision, and some conjugal transfer genes (at least an *oriT*), but lack other genes required for conjugative self-transmission (94, 126, 131, 132). IMEs can be mobilised by conjugative machinery encoded elsewhere in the bacterial genome. (C) *cis*-integrative and mobilizable elements (CIMEs) carry genes required for integration and excision but may lack an *oriT* or other conjugative factor that cannot be extrinsically provided to permit conjugative transfer (133-136). These elements may be *cis*-mobilised by invading ICEs following tandem insertion. (132-135, 137-139).

### *1.2.4. Maintenance and stability*

ICEs exist in bacterial populations in a bistable state. Most ICEs reside integrated within bacterial host genomes, while only a small proportion begin the process of excision and horizontal transfer (144). Because the excised ICE is not passively replicated with the host chromosome there is a risk that it may not be vertically disseminated to progeny cells following division and may therefore be lost from a population (145). Recent studies of diverse ICEs have suggested that excised ICEs may undergo transient autonomous replication and partitioning to circumvent this issue (146, 147). Autonomous replication was first discovered for ICE*Bs1* following the observation that the copy number of ICE*Bs1*-encoded genes increases 2-5 fold in host cells induced for excision (148). Excised ICE*Bs1* was subsequently shown to replicate unidirectionally by RCR initiated at a double-stranded *oriT* by the conjugative relaxase NicK (148). RCR of ICE*Bs1* involves unwinding of the ICE DNA by a chromosomally-encoded helicase (PcrA) commonly associated with conjugative plasmids and a helicase processivity factor (HelP) that is conserved on diverse ICEs (148, 149). Following unwinding and RCR of ICE*Bs1*, the single-stranded ICE is complemented by the initiation of lagging strand synthesis at one of two single stranded origins of replication (*ori*) (120). Strains carrying *nicK* or *oriT* mutations are unable to undertake RCR and show reduced stability of ICE*Bs1* in dividing host-cells induced for ICE*Bs1* excision (120, 148). Because many ICEs carry homologues of PcrA and HelP and all ICEs carry an *oriT* and conjugative relaxase, autonomous replication has been postulated to be a common feature of ICEs (147, 149, 150).

Autonomous replicative mechanisms resembling that of ICE*Bs1* have been described for ICE*St3* of *Streptococcus thermophilus,* Tn*916* and an SXT/R391 family ICE (150-153). Remarkably, the SXT/R391 family ICEs are also able to partition replicated circular ICEs equally between progeny cells following cell division, preventing their loss (150). The loci responsible (*srpMRC*) are homologues of the actin-type ATPase *parMRC* partitioning system described for plasmid R1 (154). Expression of these components is co-regulated with the SXT-encoded integrase, RDF and T4SS by the SetCD regulon, such that excision, autonomous replication and partitioning of circular ICEs are coordinated (150, 155). The *soj* gene carried by PAPI-1 has also been implicated in partitioning and maintenance of this ICE in its circular form, however, its mechanism of action is yet to be elucidated (85).

ICEs may also stabilise themselves in dividing populations by encoding toxin-antitoxin (TA) modules (156). TA modules comprise both a toxic protein that may be lethal to the cell or arrest growth, and a cognate anti-toxin protein which neutralises the toxin. Relative to the labile anti-toxin protein, the toxin protein exhibits a long half-life in the cell. Thus, following loss of the TA module, the stable toxin protein outlasts the anti-toxin and prevents further replication of the cell (156). Three ICE-encoded TA modules have been shown to enhance the stability of SXT/R391 family ICEs; *mosAT, tad-ata* and *hipAB* (150, 157, 158). The TA module *pezTZ* has also been identified on pathogenicity island 1 of *Streptococcus pneumoniae* and on numerous Tn-*5253*-family ICEs, although there is currently no experimental evidence that *pezTZ* enhances ICE stability (128, 159). Overall, the existence and diversity of TA modules on ICEs remains largely unexplored.

## 1.2. Symbiosis ICEs

### *1.3.1. Discovery of ICE*Ml*Sym$^{R7A}$*

The paradigm symbiosis ICE was discovered in the chromosome of *M. loti* ICMP3135 following inoculation of the pasture legume *Lotus corniculatus* with this strain in New Zealand (60, 160). Although indigenous *L. corniculatus*-nodulating rhizobia did not exist in the soil at the time this legume was introduced (161), approximately 81% of *L. corniculatus* nodules sampled 6-years post-inoculation harboured rhizobia that were genetically distinct from the inoculant (160). Further molecular analyses of these novel isolates and a re-isolate of ICMP3153, named *M. loti* R7A, revealed these strains each harboured a contiguous ~502-kb ICE adjacent to the sole chromosomal *phe*-tRNA gene (60, 160, 162, 163). This ICE carries gene modules containing *nod, nif, fix* and other symbiosis-related genes and can be conjugatively transferred to non-symbiotic mesorhizobia in the laboratory, converting them to *Lotus*-nodulating strains (60, 130, 164). In line with naming conventions for previously discovered pathogenicity ICEs, the *M. loti* symbiosis ICE was termed ICE*Ml*Sym$^{R7A}$ (74). *Mesorhizobium* symbiosis ICEs described in this thesis are named using the same convention (ICE *Genus* species-Sym$^{strain\ number}$).

### *1.3.2. Integration, excision and transfer of ICE*Ml*Sym$^{R7A}$*

ICE*Ml*Sym$^{R7A}$ is integrated within the 3' end of the sole *phe*-tRNA gene in the chromosome of *M. loti* R7A, but excises from the chromosome through site-specific recombination between 17-bp core sequences contained within

attachment sites *attL*$_S$ and *attR*$_S$ (subscripts are used to identify attachment sites for distinct recombinases, in this case IntS) (60, 74) (Fig 1.7). Recombination between *attL*$_S$ and *attR*$_S$ produces the new attachment sites *attP*$_S$ on circularised ICE*Ml*Sym$^{R7A}$ and restores the *attB*$_S$ site within the *M. loti* R7A *phe*-tRNA gene. The tyrosine recombinase IntS catalyses both the excisive and integrative reactions. Excision additionally requires the recombination and directionality factor RdfS (74). Like other RDF proteins, RdfS likely stimulates IntS-mediated excision by binding to *att* sites (C. Verdonk, personal communication). Overexpression of *rdfS* cures R7A of ICE*Ml*Sym$^{R7A}$, yielding the non-symbiotic derivative R7ANS (74).

ICE*Ml*Sym$^{R7A}$ carries an *oriT* sequence, a conjugative relaxase gene (*rlxS*) and a full suite of type IV conjugation and pilus assembly genes (130). Therefore, conjugal transfer of ICE*Ml*Sym$^{R7A}$ likely occurs in a manner resembling that of bacterial plasmids (53, 165). In R7A cells that are induced for ICE*Mc*Sym$^{R7A}$ excision by overexpressing the quorum-sensing (QS) regulator gene *traR* (discussed in Section 1.3.3), *attP* exists at a ratio of 1.5:1 relative to *attB*, suggesting that like bacterial plasmids, ICE*Ml*Sym$^{R7A}$ may also autonomously replicate in the circular form (74). ICE*Ml*Sym$^{R7A}$ also encodes a homologue of the plasmid partitioning protein ParB (130), however there is no functional evidence as to the role of this protein in partitioning of replicated forms of excised ICE*Ml*Sym$^{R7A}$.

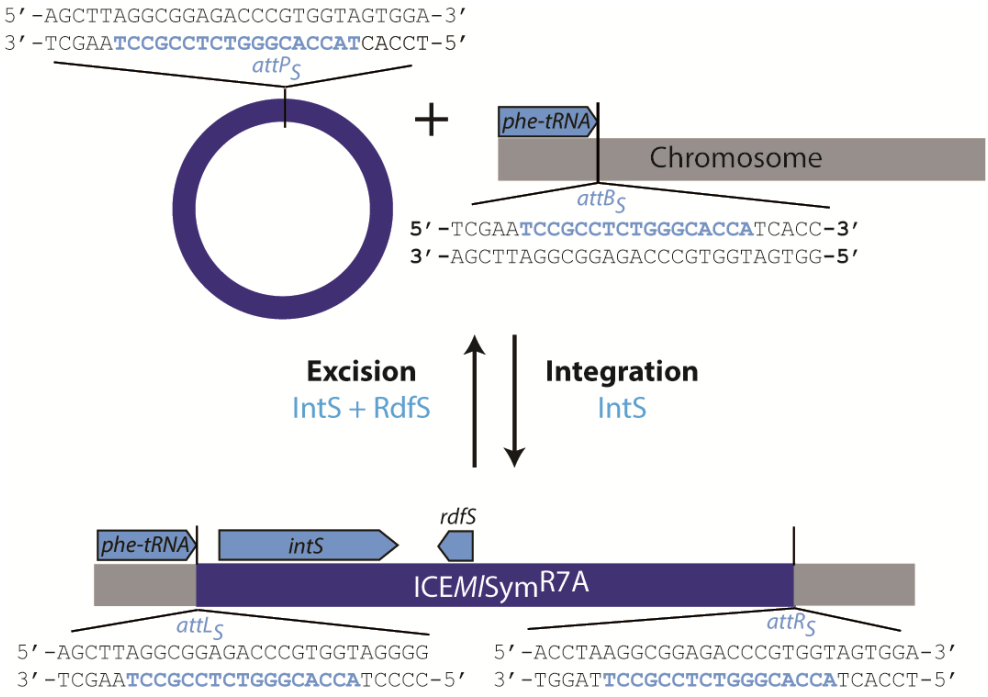


**Figure 1.7. Integration and excision of ICE*Ml*Sym$^{R7A}$.** ICE*Ml*Sym$^{R7A}$ primarily integrated within the chromosome of *M. loti* R7A adjacent to the 3'-end of *phe*-tRNA, but excises from the chromosome to form a circular plasmid-like element via site-specific recombination between the direct repeat sites *attL*$_S$ and *attR*$_S$ positioned at each ICE terminus. This recombination is conservative (there is no loss nor gain of DNA), thus, the direct repeat sequence is preserved within the 5'-end of the *phe*-tRNA gene (*attB*$_S$), and in the excised ICE (*attP*$_S$). Integration occurs via the reverse reaction of excision. Both integration and excision reactions are catalyzed by the integrase IntS, however, the recombination directionality factor RdfS stimulates excision.

### *1.3.3. Regulation of excision and transfer of ICEMlSym$^{R7A}$*

Excision and conjugative transfer of ICE*Ml*Sym$^{R7A}$ are presumably energy-demanding processes and are tightly regulated at multiple levels (75, 76). This regulation is primarily achieved through transcriptional control of *rdfS* and the downstream conjugative transfer genes *traF, msi107* and *rlxS* (Fig 1.8). The *rdfS* operon is activated by a LuxR-LuxI QS system resembling that of the *Agrobacterium tumefaciens* and *R. leguminosarum* plasmids pTi and pRL1JI,

respectively (130, 166-172) (Fig 1.8A). ICE*Ml*Sym$^{R7A}$ encodes the *N*-acyl-homoserine lactone (AHL) synthase TraI1 that catalyses production of the autoinducer molecule *N*-(3-oxohexanoyl)-*L*-homoserine lactone (3-oxo-$C_6$-HSL), which presumably accumulates as a function of population cell density (167, 173-175). At "quorum" concentration, 3-oxo-$C_6$-HSL binds to the LuxR-family transcriptional regulator TraR, activating this protein and allowing it to bind and recruit RNA polymerase at two "*tra*-box" promoters (167). The first *tra*-box is positioned upstream of *traI1,* allowing for positive feedback regulation of AHL production. The second *tra*-box is positioned upstream of a predicted AHL-synthase gene *traI2*. Strains carrying a mutation in *traI2* show no defect in AHL production or ICE*Ml*Sym$^{R7A}$ excision. Furthermore, *M. loti* or *E. coli* strains constitutively expressing *traI2* produce no known detectable AHLs, suggesting that *traI2* may be a pseudogene (167). *traI2* is translationally coupled to two open-reading-frames (ORFs) *msi172* and *msi171,* which undergo ribosomal frameshifting in 4-13% of translational events to produce the transcriptional activator of the *rdfS* operon*,* FseA (170). The existence of QS-regulation of ICE*Ml*Sym$^{R7A}$ excision and transfer suggests that the most prevalent sites of ICE*Ml*Sym$^{R7A}$ transfer in nature are within cell dense areas, such as the rhizosphere, rhizoplane and legume nodule cells.

Although overexpression of TraR in R7A activates ICE*Ml*Sym$^{R7A}$ excision in 100% of cells, TraR mutants exhibit levels of ICE excision resembling that of wild-type cultures (167, 176). This is presumably because activated TraR and FseA are allosterically inactivated by the antiactivator protein QseM (170, 176) (Fig. 1.8B). A divergently-encoded gene *qseC* encodes a regulator that activates its own expression and represses the expression of QseM (176).

Thus, is thought that repression of QS-induced ICE*Ml*Sym$^{R7A}$ excision and transfer by QseM is alleviated in stationary-phase cultures via the action of QseC. QseM likely exists to ensure that the QS-induced transcriptional activation of the *rdfS* operon does not spuriously occur in individual cells within a population that is yet to reach the critical cell density and tightens the expression of the genes involved in the QS-induced cascade of genetic regulation.

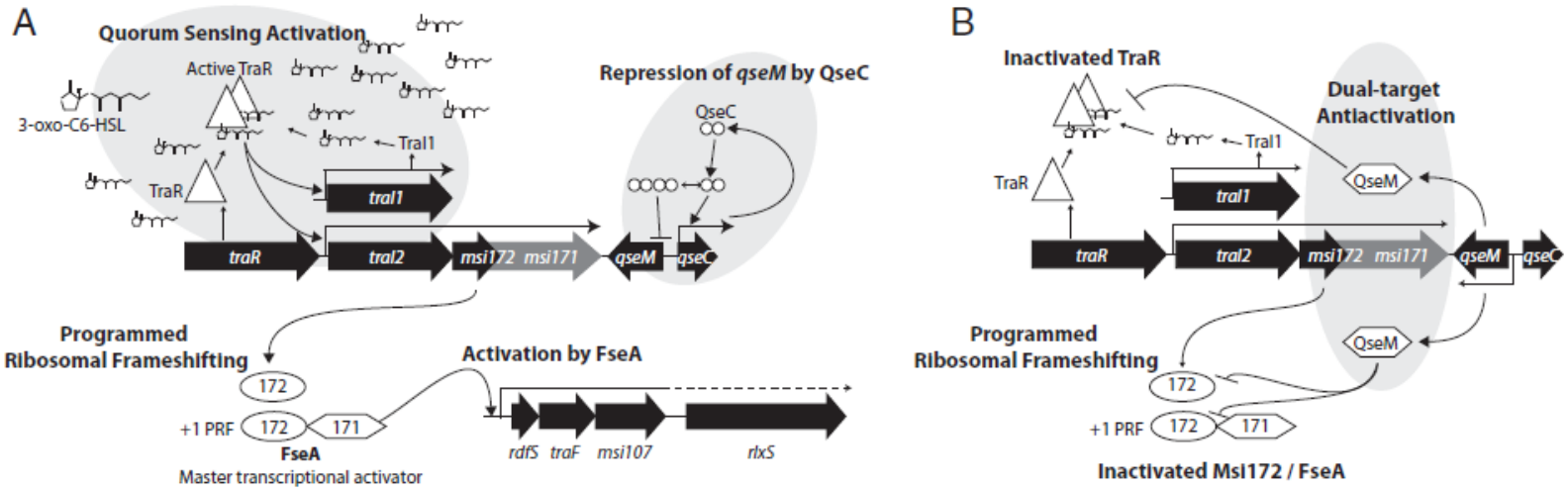


**Figure 1.8. Regulation of ICE*Ml*Sym$^{R7A}$ excision.** (A) The AHL 3-oxo-$C_6$-HSL, produced by TraI1, accumulates in stationary-phase cultures of *M. loti* R7A activating the QS transcriptional regulator TraR. Active TraR stimulates transcription from the *traI1 traI1* promoter completing a positive feedback loop of regulation, and also stimulates transcription from the *traI2* promoter. Although *traI2* does not appear to encode a functional gene, it is encoded as a polycistronic mRNA with *msi172-msi171* which undergoes a programmed ribosomal frameshift in ~4-13% of translational events producing the transcriptional activator of the *rdfS* operon, FseA. Although QseM is an antiactivator of TraR-3-oxo-$C_6$-HSL and FseA, high levels of QseC expression in stationary phase cells repress expression of QseM. (B) In log-phase cultures QseC is lowly expressed, and presumably exists at an inadequate concentration to inhibit QseM expression. Therefore, QseM can allosterically inhibit TraR-3-oxo-$C_6$-HSL and FseA in log-phase. This figure was reproduced from reference (170).

### *1.3.4. Diversity of symbiosis ICEs*

Since the discovery of ICE*Ml*Sym$^{R7A}$, symbiosis ICEs have been functionally or bioinformatically identified in the *Lotus*-nodulating strains; *M. loti* R88B (162), NZP2037 (177, 178) and *M. huakuii* (formerly *M. loti*) MAFF303099 (178, 179); the chickpea-nodulating strain *M. ciceri* CC1192 (63) and the *B. pelecinus*-nodulating strain *M.* sp. AA22 (A. Bekuma, Personal Communication). ICE*Ml*Sym$^{R7A}$ has also been transferred in the laboratory to the non-symbiotic *M. loti* strains CJ3, CJ4 and CJ7 creating CJ3Sym, CJ4Sym and CJ7Sym, respectively (60, 72, 164). Thus far the *Astragalus sinicus* symbiont, *M. huakuii* 7653R is the only *Mesorhizobium* strain to be reported to carry symbiosis genes on plasmids (pMhu7653Ra and pMhu7653Rb) rather than on a chromosomal symbiosis ICE (73). However, it has been argued that these plasmids may in fact have originated from an excised region of the host chromosome following acquisition of a module facilitating the autonomous maintenance and replication of this DNA (73).

Outside of the *Mesorhizobium* genus, an ~86 kb symbiosis ICE (ICE$^{Ac}$) has recently been identified adjacent to a *gly*-tRNA gene in the chromosome of the *Sesbania rostrata* symbiont *A. caulinodans* ORS571 (59) Unlike *Mesorhizobium* spp., *A. caulinodans* does not utilise the traditional RHC mode of legume infection (Fig.1.1) but instead infects the lateral roots and stems of *S. rostrata* via crack entry (180). ICE$^{Ac}$ encodes *nod* genes, but is entirely devoid of *nif* and *fix* genes required for $N_2$ fixation (59). ICE$^{Ac}$ can be conjugally transferred to symbiotic strains of *Sinorhizobium* and *Mesorhizobium* that do not carry a symbiosis ICE and conveys upon them the ability to nodulate *S. rostrata.* ICE$^{Ac}$ exconjugants of *M. huakii* 93 are also able to fix $N_2$ with this

host, presumably through expression of native *nif* and *fix* genes in the *M. huakii* genome. In contrast to ICE*Ml*Sym$^{R7A}$, which regulates excision and transfer through QS, ICE$^{Ac}$ excision and transfer is regulated in response to the plant-derived flavonoid naringenin present in the rhizosphere (59).

Large 860-kb and 681-kb genomic islands carrying *nod, nif* and *fix* genes have also been reported to exist adjacent to a *val*-tRNA gene in the chromosomes of *B. diazoefficiens* USDA110 and *B. japonicum* USDA6, respectively (61, 62). Although there is evidence that symbiosis genes may have horizontally transferred between *Bradyrhizobium* spp. in the field (46, 181, 182), there is currently no empirical data demonstrating mobility for these genomic islands. In-fact, it was reported that the *B. japonicum* USDA6 symbiosis ICE may be fragmented into three regions in the chromosome (61). Laboratory conjugation experiments will be critical to confirm whether these putative symbiosis ICEs remain transmissible.

### *1.3.5. Core features of Mesorhizobium symbiosis ICEs*

The completion of genome sequences for *M. ciceri* CC1192 (63), *M. huakuii* MAFF303099 (179) and *M. loti* strains NZP2037 (177) and R7A (163) has revealed that *Mesorhizobium* symbiosis ICEs carry a conserved compliment of genes. Not surprisingly, this includes *nod, nif, fix* and other genes involved in rhizobia-legume symbiosis (130, 178). A comparison of the R7A, MAFF303099 and NZP2037 symbiosis ICEs revealed that two genes *nodU* and *nolO* whose products are involved in carbamoylation of Nod factors in *Sinorhizobium fredii* NGR234 (183, 184) are present in ICE*Ml*Sym$^{2037}$, but

these are absent from ICE*Ml*Sym$^{R7A}$ and ICE*Ml*Sym$^{303099}$ (178). The presence of *nodU* and *nolO* on ICE*Ml*Sym$^{2037}$ may explain why NZP2037 is able to form nodules with at least six legume genera, whereas R7A and MAFF303099 appear to be restricted to nodulating *Lotus* spp. (68, 69, 185-187). ICE*Ml*Sym$^{R7A}$ and ICE*Ml*Sym$^{2037}$ both carry a VirB1/D4 type-IV protein secretion system (T4SS), whereas ICE*Mh*Sym$^{303099}$ and ICE*Mc*Sym$^{1192}$ carry a type-III secretion system (T3SS) (63, 130, 178, 188). Both have been implicated in the translocation of protein effectors into legume cells (187). Inactivation of the ICE*Ml*Sym$^{R7A}$ T4SS and ICE*Mh*Sym$^{303099}$ T3SS enables R7A and MAFF303099 to nodulate the non-native host *L. leucocephala*, suggesting that there may be a common biological role for both T4SS and T3SS in *Mesorhizobium* symbioses (187). The queuosine biosynthetic genes (QueBCD), involved in hyper-modification of tRNAs (189) and required for functional symbiosis between *S. meliloti* and *Medicago truncatula* (190), also appear to be conserved on *Mesorhizobium* symbiosis ICEs (J Ramsay, Personal Communication). The biological relevance of queuosine biosynthesis in *Mesorhizobium* spp. is yet to be explored.

The conserved core genes of *Mesorhizobium* symbiosis ICEs are not restricted to those involved in symbiosis. The symbiosis ICEs of all four of the fully sequenced *Mesorhizobium* genomes encode biosynthetic modules for production of the essential vitamins biotin, nicotinate and thiamine (63, 130, 178, 191). Symbiotic mesorhizobia are therefore typically prototrophic for biotin, nicotinate and thiamine produciton, whereas symbiosis ICE devoid strains such as *M. loti* R7ANS and *M.* sp. strains N18 and *M.* sp CJ4 are auxotrophic. This feature of symbiosis ICEs has been exploited in the selection

for ICE exconjugants without the need for genetic marking of the symbiosis ICEs (60, 74).

Despite some symbiosis ICEs, such as ICE*Mc*Sym$^{R7A}$ and ICE*Mc*Sym$^{1192}$, existing within different chromosomal locations (*phe*-tRNA and *ser*-tRNA, respectively), nearly all genes involved in regulation of ICE excision (*rdfS , qseC, qseM, traR, traI1, msi172-msi171*) and conjugative transfer (type IV conjugugation genes, *rlxS, traF, msi107*) are conserved in all four of the fully sequenced *Mesorhizobium* symbiosis ICEs (63, 130, 178), suggesting a common mechanism of excision and transfer. However, the *oriT* of ICE*Mh*Sym$^{303099}$ appears to be interrupted by insertion of a transposon indicating that this symbiosis ICE may be non-mobile (130).

Although many of the conserved or unique symbiosis ICE genes in *Mesorhizobium* spp. have been characterised, the vast majority of genes encoded by these elements currently have no assigned function (130).

## *1.4.* Emergence of novel *Mesorhizobium* spp nodulating the pasture legume *Biserrula pelecinus*

The development of agriculture in Australia and New Zealand has been dependent on the introduction of exotic legumes which often lack compatible rhizobia naturally present in the soil (30, 161, 192). In New Zealand, novel *Lotus*-nodulating rhizobia that emerged following transfer of the *M. loti* R7A symbiosis ICE to indigenous soil mesorhizobia is a well-documented and unintended outcome of this practice (60, 160). A similar scenario is likely to have occurred in Western Australia following introduction of the pasture

legume *Biserrula pelecinus,* along with the effective inoculant strains *M. ciceri* bv. biserrulae WSM1271 and WSM1497 (193, 194).

*B. pelecinus* nodules were sampled at an experimental field site six years post-inoculation with WSM1271. Despite the fact that *B. pelecinus*-nodulating rhizobia were absent from the study at the time of inoculation, the Randomly Amplified Polymorphic DNA (RAPD) and Enterobacterial Repetitive Intergenic Consensus (ERIC) profiles from 88 isolates revealed that seven were genotypically distinct from WSM1271 (193). Glasshouse trials indicated effectiveness of $N_2$ fixation of these genetically distinct rhizobia on *B. pelecinus* was suboptimal. Five strains were partially effective (fixed $N_2$ 30-50%of the WSM1271 amount), while two strains nodulated *B. pelecinus* but did not fix $N_2$ (194). Each of these strains encoded an integrase homologous to IntS of ICE*Ml*Sym$^{R7A}$ adjacent to a *phe*-tRNA gene and carried chromosomal *nifH* and *nodA* genes 100% identical to the original inoculant strain WSM1271 (60, 74, 160, 193). PCR and sequencing of 4 isolates also revealed that each harboured a region of DNA homologous to the *attL* junction that demarks the border between ICE*Ml*Sym$^{R7A}$ and the *phe*-tRNA symbiosis ICE insertion site in *M. loti* R7A (193). These four strains were later shown to belong to two entirely new species; *M. opportunistum* WSM2075 (nodulates but does not fix $N_2$) and *M. australicum* strains WSM2073, WSM2074 and WSM2076 (partially effective $N_2$-fixation) (195). Considering these data, it seemed plausible that like R7A, WSM1271 may harbour a symbiosis ICE that had transferred to soil mesorhizobia, converting them to *Biserrula* nodulating strains.

The current commercial inoculant for *B. pelecinus* in Australia is *M. ciceri* WSM1497 (196). As was observed with release of WSM1271, genotyping of

strains isolated from nodules at an experimental *B. pelecinus* field sites seven years after inoculation with WSM1497 revealed 193 out of 387 nodule isolates to be genetically distinct from the inoculant strain (194). Fifty-three genetically distinct isolates were screened for $N_2$-fixation effectiveness in symbiosis with *B. pelecinus* and remarkably, none fixed $N_2$ with effectiveness equal to WSM1497. Rather, 51 were partially effective (fixed $N_2$ at $\leq$ 70% that of WSM1497) and six of these strains nodulated but did not fix $N_2$ in symbiosis with *B. pelecinus*. PCR and sequencing of 12 isolates also revealed that each harboured the region of DNA homologous to the *attL* junction demarkating the border between ICE*Ml*Sym$^{R7A}$ and the *phe*-tRNA symbiosis ICE insertion site in *M. loti* R7A (193). Like WSM1271, it seems likely that WSM1497 also carried a symbiosis ICE that had transferred to the native soil mesorhizobia.

## 1.5 Aims of this thesis

Inoculation of legumes with effective $N_2$-fixing rhizobia is a crucial component of sustainable agriculture in both developed and undeveloped countries. Legumes may be grown in rotation with cereal crops to provide a primary source of N without the requirement for supplementation with energy-expensive fertilisers and provide a source of food for humans and domestic animals (197). In 2016, 225 million hectares of legumes crops were harvested globally, equating to approximately 10% of the planet's arable land (198-200). Legumes account for ~27% of the world's primary crop production and provide at least 33% of humankind's N needs (199, 201).

All legumes used in Australian and New Zealand argiculture have been introduced and due to the lack of compatible naturally occurring populations of rhizobia in the soil, so have their microsymbionts (30, 161, 192). The key to maximising the productivity of these legumes has been matching them with the most elite rhizobial strains that; a) are compatible with the host-legume; b) fix $N_2$ effectively; and c) persist in the soil environment (29, 30, 197, 202, 203). One factor that has been often overlooked in the selection of elite rhizobial strains for Australian New Zealand agriculture is 'symbiotic stability' – i.e. the potential that inoculant rhizobia may transfer genes required to engage in $N_2$-fixing symbioses to other closely related bacteria, converting them to legume nodulating strains (54-60, 160). For *B. pelecinus*, the newly evolved strains were less effective than the original inoculant, suggesting that the mobility of the putative WSM1271 and WSM1497 symbiosis ICEs may produce substantial populations of suboptimally-effective *B. pelecinus*- nodulating mesorhizobia that could out-compete the inoculant for nodulation of the legume, leading to a reduction in agricultural productivity.

Genome sequences for WSM1271 and two sub-optimal $N_2$-fixing putative symbiosis ICE recipients WSM2073 and WSM2075 were recently completed, revealing that each strain carries an identical suite of symbiosis genes (73, 204-206). As previously reported, each of the three strains carries a homologue of the ICE*Ml*Sym$^{R7A}$ integrase IntS adjacent to the ICE*Ml*Sym$^{R7A}$ insertion site (60, 193, 194) suggesting the presence of an ICE*Ml*Sym$^{R7A}$-like element. However, the symbiosis genes in each strain appear to be carried on a region of the chromosome distant to this element, suggesting that they may be carried by a unique form of integrative element. Analysing the potential

existence and assessing mobility of this unique symbiosis element in WSM1271 and its symbiosis ICE recipients WSM2073 and WSM2075 will be critical for explaining how sub-optimal fixing $N_2$-fixing rhizobia evolved following introduction of WSM1271 into Australia and provide insights into the mechaninsms of horizontal gene transfer of these mobile genetic elements.

The aims of this thesis were as follows:

1) Identify the symbiosis ICE of *Biserrula pelecinus*-nodulating strains *M. ciceri* WSM1271, *M. australicum* WSM2073 and *M. opportunistum* WSM2075.

2) Characterise the mechanism of transfer for this symbiosis ICE

3) Elucidate how the transfer of this symbiosis ICE is regulated

# Chapter 2.

## Materials and methods

## 2.1. Media and growth conditions

*Escherichia coli* strains were cultured on Luria-Bertani (LB) media (207) at 37°C. *E. coli* ST18 (a *hemA* mutant of S17-1 which is auxotrophic for 5-aminolevulinic acid) (208) culture medium was supplemented with 60 μg $mL^{-1}$ 5-aminolevulinic acid. *Mesorhizobium* strains were cultured at 28°C on TY (209) or RDM (210) media supplemented with the vitamins: biotin (20 ng $mL^{-1}$); nicotinate (1 μg $mL^{-1}$) and thiamine (1 μg $mL^{-1}$), and either 10 mM D-glucose or 150 mM L-sucrose as a sole carbon source. *Chromobacterium violaceum* CV026 was cultured on LB media at 28°C. Difco grade A agar (1.5% w/v) was used to solidify media where required. Antibiotics (Sigma Aldrich) were added at the following concentrations where appropriate (μg $mL^{-1}$); carbenicillin (Cb) 20; gentamycin (Gm) 20; kanamycin (Km) 50; neomycin (Nm) 250; spectinomycin (Sp) 50 (for *E. coli*) or 250 (for rhizobia); tetracycline (Tc) 10 (for *E. coli*) or 0.5-2.5 (for rhizobia).

## 2.2. Bacterial strains in this thesis

Bacterial strains and plasmids are detailed in Table 2.1. All mutant strains were constructed by replacing or deleting alleles by double crossover homologous recombination following the introduction of a suicide vector containing the levansucrase (*sacB*) gene described in Table 2.1. This was achieved via a two-step process. Single-crossover integration of suicide vectors was initially selected by plating serial dilutions of cells onto RDM agar supplemented with glucose and the appropriate antibiotics and incubating for 6-8 days at 28°C. Single colonies were then cultured in TY broth (without antibiotic selection) to

stationary phase ($OD_{600}$ ~ 2.0) and serial dilutions spread onto RDM plates supplemented with sucrose, to select for double crossover recombinants. Mutants were confirmed by screening for the relevant antibiotic resistance/sensitivity profile on RDM agar, and by Sanger sequencing of PCR products amplified from the deleted/replaced region.

**Table 2.1. Bacterial strains, plasmids and NCBI accessions**

| Strain | [a] Relevant Characteristics and (NCBI accession) | Source |
|---|---|---|
| ***Chromobacterium violaceum*** | | |
| CV026 | $C_4$-$C_8$ *N*-acyl-homoserine lactone biosensor strain | (211) |
| ***Escherichia coli*** | | |
| DH10B | F- *endA1 deoR*[and] *recA1 galE15 galK16 nupG rpsL* Δ*(lac)X74* φ80*lacZΔM15 araD139* Δ*(ara,leu)7697 mcrA* Δ*(mrr-hsdRMS-mcrBC)* SmR $\lambda^-$ | Invitrogen |
| DH5α | F– Φ80*lacZ*Δ*M15* Δ(*lacZYA-argF*) U169 *recA1 endA1 hsdR17* (rK–, mKand) *phoA supE44* λ– *thi*-1 *gyrA96 relA1* | |
| ST18 | S17 Δ*pir* Δ*hemA* | (208) |
| ***Mesorhizobium australicum*** | | |
| WSM2073 | Field-isolated exconjugant of ICE*Mc*Sym[1271] (NC_019973.1) | (206) |
| ***Mesorhizobium ciceri*** | | |
| WSM1271 | Wild-type *Bisserula pelecinus* symbiont, harbours ICE*Mc*Sym[1271] (NC_014923.1) | (204) |
| 1271Δ*intG*::*nptII* | WSM1271 *intG nptII* replacement mutant | This study |
| 1271Δ*intM*::*nptII* | WSM1271 *intM nptII* replacement mutant | This study |
| 1271Δ*intS* | WSM1271 *intS* frameshifted deletion mutant | This study |
| 1271Δ*rdfG*::Ω*aadA* | WSM1271 *rdfG* Ω*aadA* replacement mutant | This study |
| 1271Δ*rdfM*:: Ω*aadA* | WSM1271 *rdfM* Ω*aadA* replacement mutant | This study |
| 1271Δ*rdfS* | WSM1271 *rdfS* in frame deletion mutant | This study |
| 1271Δ*traI1*:: Ω*aadA* | WSM1271 *traI1* Ω*aadA* replacement mutant | This study |
| 1271Δ*traI1*:: Ω*aadA* Δ*ahlI* | WSM1271Δ*traI1*::Ω*aadA ahlI* markerless deletion mutant | This study |
| WSM1497 | Wild-type *B. pelecinus* symbiont isolated from Sardinia (NZ_CP021070) | This study |
| Ca181 | Wild-type *Cicer areninium* (chick-pea) symbiont isolated from India, harbours ICE*Mc*Sym[181] (NZ_CM002796) | (212) |
| WSM4083 | Wild-type *Bituminaria bituminosa* symbiont (JAFG00000000) | [b] G. O'Hara |

| Strain | [a] Relevant Characteristics and (NCBI accession) | Source |
|---|---|---|
| WSM1284 | Wild-type *Biserrula pelecinus* symbiont, harbours ICE*Mc*Sym[1284] (NZ_CP015064.1) | This study |
| ***M. loti*** | | |
| NZP2037 | Wild-type, isolated in New Zealand from *L. divaricatus*, harbours ICE*Ml*Sym[NZP2037] and plasmid pRlo2037 (NZ_KB913026, CP016079, CP016080) | This study, (4, 177) |
| SU343 | Wild-type, *Lotus sp.* symbiont isolated in NSW, Australia, harbours ICE*Ml*Sym[343] (LYTL00000000) | [c] J. Sullivan and C. Ronson |
| WSM1293 | Wild-type *Lotus sp.* symbiont isolated in Greece (AZUV00000000.1) | (213) |
| NZP2042 | Wild-type *Lotus sp.* symbiont isolated in New Zealand (LYTK00000000) | [c] J. Sullivan and C. Ronson |
| R7A | Wild-type field re-isolate of ICMP3153; wild-type symbiotic strain (KI632510.1) | (160) |
| R7ANS | Non-symbiotic derivative of R7A; lacks ICE*Ml*Sym[R7A] | (74) |
| R7ANSxWSM1271 | R7ANS ICE*Mc*Sym[1271] exconjugants carrying pFAJ1708 (LZTK00000000) | This study |
| R7ANSxNZP2037 | R7ANS ICE*Ml*Sym[2037] exconjugants carrying pFAJ1708 (LZTH00000000) | [c] J. Sullivan and C. Ronson |
| R7ANSxNZP2042 | R7ANS ICE*Ml*Sym[2042] exconjugants carrying pFAJ1708 (LZTJ00000000) | [c] J. Sullivan and C. Ronson |
| R7ANSxSU343 | R7ANS ICE*Ml*Sym[343] exconjugants carrying pFAJ1708 (LZTL00000000) | [c] J. Sullivan and C. Ronson |
| R7AMc1 | R7ANS ICE*Mc*Sym[1271] exconjugant cured of all plasmids | This study |
| ***M. metallidurans*** | | |
| STM2683 | Wild-type metal resistant symbiont of *Anthyllis vulneraria* (NZ_CAUM01000099) | (214) |
| ***M. opportunistum*** | | |
| WSM2075 | Wild-type field-isolated exconjugant of ICE*Mc*Sym[1271] (NC_015675.1) | (205) |
| ***M.* sp.** | | |
| AA22 | Wild-type *B. pelecinus* symbiont isolated from Ethiopia (LYTO00000000) | [b] A. Bekuma |
| **Plasmids** | | |
| pJQ200 SK | Suicide vector in *Mesorhizobium*, contains *sacB*, GmR | (215) |
| pEX18Tc | Suicide vector in *Mesorhizobium*, contains *sacB*, TcR | (216) |
| pHP45Ω | Insertional inactivation vector carrying an Ω*aadA* cassette, SmR, SpR | (217) |
| pJET-*aadA1* | pJET 1.2 carrying the Ω*aadA* cassette from pHP45Ω amplified using primers 55 & 56, CbR, SmR, SpR | This study |
| pJET-*nptII* | pJET 1.2 carrying *nptII* amplified from pJP2neo using primers 93 & 94, CbR, NmR | This study |
| pJQΔ*intG* | pJQ200 SK carrying *nptII* from pJET-*nptII* flanked by regions upstream and downstream of *intG,* amplified from WSM1271 using primers 1, 2 & 3, 4, respectively, used to create 1271Δ*intG::nptII,* GmR, NmR | This study |

| Strain | [a] Relevant Characteristics and (NCBI accession) | Source |
|---|---|---|
| pJQΔ*rdfG* | pJQ200 SK carrying Ω-SpR/SpR cassette from pJET-*aadA1* flanked by regions upstream and downstream of *rdfG* amplified using primers 13, 14 & 15,16, respectively, used to create 1271Δ*rdfG*::Ω*aadA*, SmR, SpR, GmR | This study |
| pJQΔ*rdfM* | pJQ200 SK carrying Ω-SpR/SpR cassette from pJET-*aadA1* flanked by regions upstream and downstream of *rdfM* amplified using primers 17, 18 & 19, 20, respectively, used to create 1271Δ*rdfM*::Ω*aadA*, SmR, SpR, GmR | This study |
| pJQΔ*intM* | pJQ200 SK carrying *nptII* amplified from pJET-*nptII* using primers 95 & 96, flanked by regions upstream and downstream of *intM,* amplified from WSM1271 using primers 5, 6 & 7, 8 respectively, used to create 1271Δ*intM*::*nptII*, GmR, NmR | This study |
| pJQΔ*traI1* | pJQ200 SK carrying Ω-SpR/SpR cassette from pJET-*aadA1* flanked by regions upstream and downstream of *traI1* amplified using primers 104, 105 & 106,107, respectively, used to create 1271Δ*traI1*::Ω*aadA*, SmR, SpR, GmR | This study |
| pEXΔ*intS* | pEX18Tc carrying regions flanking *intS* amplified using primers 9, 10 & 11, 12, respectively, used to create WSM1271Δ*intS*, TcR | This study |
| pEXΔ*rdfS* | pEX18Tc carrying regions flanking *rdfS* amplified using primers 18, 19 & 20, 21, respectively, used to create WSM1271Δ*rdfS*, TcR | This study |
| pEXΔ*ahlI* | pEX18Tc carrying regions flanking *ahlI* amplified using primers 99, 100 & 101, 102, respectively, used to create 1271Δ*traI1*::Ω*aadA*Δ*ahlI*, TcR | This study |
| pTH3*attP* | pJQ200 SK carrying ICE*Mc*Sym$^{1271}$ $attP_G$, $attP_M$, $attP_S$ sites amplified from WSM1271 using primers 65 & 66, 61 & 62, 63 & 64, respectively, GmR | This study |
| pJP2 | Stable (contains Par region), low copy number BHR IncP vector, CbR, TcR | (218) |
| pJP2neo | pJP2 carrying *nptII*, CbR, TcR, NmR | [b]J. Terpolilli |
| pJP2-*intG* | pJP2 carrying *intG* from WSM1271 amplified using primers 28 & 29, CbR, TcR | This study |
| pJP2-*intM* | pJP2 carrying *intM* from WSM1271 amplified using primers 30 & 31, CbR, TcR | This study |
| pJP2-*intS* | pJP2 carrying *intS* from WSM1271 amplified using primers 32 & 33, CbR, TcR | This study |
| pJP2-*rdfG* | pJP2 carrying *rdfG* from WSM1271 amplified using primers 34 & 35, CbR, TcR | This study |
| pJP2-*rdfM* | pJP2 carrying *rdfM* from WSM1271 amplified using primers 36 & 37, CbR, TcR | This study |
| pJP2-*rdfS* | pJP2 carrying *rdfS* from WSM1271 amplified using primers 38 & 39, CbR, TcR | This study |
| pPR3 | pPROBE-KT carrying the *nptII* promoter from pFAJ1708, NmR | (219) |
| pPR3-*rdfG* | pPR3 carrying *rdfG* from WSM1271 amplified using primers 40 & 41, NmR | This study |
| pPR3-*traI1* | pPR3 carrying *traI1* from WSM1271 amplified using primers 42 & 43, NmR | This study |
| pPR3-*mbrI* | pPR3 carrying *mbrI* from WSM1271 amplified using primers 91 & 92, NmR | This study |
| pSRKKm | pBBR1MCS-2-derived broad-host-range expression vector containing lac promoter and *lacI*$^q$ , *lacZα*$^{and}$, NmR | (220) |
| pSacB | pSRKKm carrying *sacB* gene amplified from pJQ200 SK amplified using primers 59 & 60, NmR | This study |

| Strain | [a] Relevant Characteristics and (NCBI accession) | Source |
|---|---|---|
| pSacB-*IntG* | pSacB carrying *intG* amplified from WSM1271 using primers 77 & 78, NmR | This study |
| pSacB-*IntM* | pSacB carrying *intM* amplified from WSM1271 using primers 79 & 80, NmR | This study |
| pSacB-*IntS* | pSacB carrying *intS* amplified from WSM1271 using primers 81 & 82, NmR | This study |
| pSacB-*rdfM* | pSacB carrying *rdfM* from WSM1271 amplified using primers 44 & 37, NmR | This study |
| pSacB-*ahlI* | pSacB carrying *ahlI* from WSM1271 amplified using primers 97 & 98, NmR | This study |
| pSDz | BHR plasmid, carries IPTG inducible promoter and promoterless *lacZ,* CbR, TcR | (170) |
| pSDz-*traR1* | pSDz carrying *traR1* from WSM1271 amplified using primers 45 & 46*,* CbR, TcR | This study |
| pSDz-*traR2* | pSDz carrying *traR2* from WSM1271 amplified using primers 87 & 88*,* CbR, TcR | This study |
| pSDz-*mbrR* | pSDz carrying *mbrR* from WSM1271 amplified using primers 89 & 90*,* CbR, TcR | This study |
| pSDz-*msi172171* | pSDz carrying *msi172-msi171* from WSM1271 amplified using primers 47 & 48*,* CbR, TcR | This study |
| pSDz-P*rdfG* | pSDz carrying the *rdfG* promoter from WSM1271 amplified using primers 49 & 50*,* CbR, TcR | This study |
| pSDz-P*rdfM* | pSDz carrying the *rdfM* promoter from WSM1271 amplified using primers 51 & 52*,* CbR, TcR | This study |
| pSDz-P*rdfS* | pSDz carrying the *rdfS* promoter from WSM1271 amplified using primers 53 & 54*,* CbR, TcR | This study |
| pSDz-tb | pSDz carrying the *traI1* promoter from WSM1271 amplified using primers 83 & 84*,* CbR, TcR | This study |
| pSDz-tb*traR1* | pSDz-tb carrying *traR* from WSM1271 amplified using primers 45 & 46*,* CbR, TcR | This study |
| pSDz-tb*traR2* | pSDz-tb carrying *traR2* from WSM1271 amplified using primers 87 & 88*,* CbR, TcR | This study |
| pSDz-tb*mbrR* | pSDz-tb carrying *mbrR* from WSM1271 amplified using primers 89 & 90*,* CbR, TcR | This study |
| pSDz-mb | pSDz carrying the *mbrI* promoter from WSM1271 amplified using primers 85 & 86*,* CbR, TcR | This study |
| pSDz-mb*traR1* | pSDz-mb carrying *traR* from WSM1271 amplified using primers 45 & 46*,* CbR, TcR | This study |
| pSDz-mb*traR2* | pSDz-mb carrying *traR2* from WSM1271 amplified using primers 87 & 88*,* CbR, TcR | This study |
| pSDz-mb*mbrR* | pSDz-mb carrying *mbrR* from WSM1271 amplified using primers 89 & 90*,* CbR, TcR | This study |
| pTHQP-1 | Standard construct for qPCR assays for ICE[3] excision, GmR | This study |
| pJET 1.2. | Commercial blunt cloning vector, CbR | Thermo Fisher Scientific |
| pFUS2 | Suicide vector in *Mesorhizobium,* GmR | (221) |
| pMINI3 | pFUS2 carrying $attP_G$, $attP_M$ and $attP_S$ amplified from pTH3*attP* using primers 67 & 68*,* GmR | This study |

| Strain | [a] Relevant Characteristics and (NCBI accession) | Source |
|---|---|---|
| pTHQP-1 | pTH3*attP* carrying ICE*Mc*Sym$^{1271}$ $attB_G$, $attB_M$ and $attB_S$ sites and a *melR* region amplified from WSM1271 using primers 39 & 40, 41 & 42, 43 & 44, 45& 46 respectively, qPCR standard, GmR | This study |
| pFAJ1708 | Broad-host-range plasmid containing GFP downstream of *nptII* promoter and MCS, CbR, TcR | (222) |

[a] Abbreviations for antibiotic resistance (R) are as follows: Cb, carbenicillin; Gm, gentamycin; Km, kanamycin; Sm, streptomycin; Sp, spectinomycin; Nm, neomycin; Tc, tetracycline. See Table 2.2 for primer details.

[b] Affiliation: Centre for Rhizobium Studies, Murdoch University, Perth, Australia

[c] Affiliation: Department of Microbiology and Immunology, University of Otago, Dunedin 9016, New Zealand

## 2.3. Construction of plasmids

Primers used to construct plasmids in study are detailed in Tables 2.1 and 2.2. Digested plasmids were dephosphorylated with alkaline shrimp phosphatase (rSAP, New England Biolabs) as described in Section 2.4.4. prior to ligation. Ligated plasmids were initially transformed into *E. coli* DH5α or DH10β, then purified and transformed into ST18 for biparental conjugative transfer into *Mesorhizobium* spp. (208) (described in Section 2.4.6). Plasmids were constructed as follows;

**pJET-*aadA1* and pJET-*nptII*.** *aad*A1 and *nptII* were each amplified by PCR from pHP45Ω or pJP2neo plasmid DNA, respectively, and ligated as a blunt fragment into the commercial cloning vector pJET 2.1.

**pJQΔ*intG*.** Regions upstream and downstream of *intG* were amplified by PCR from WSM1271 genomic DNA and digested with SacI/XbaI (for upstream fragment) or XhoI/BamHI (for downstream fragment) and ligated with pJQ200 SK digested with SacI/BamHI, and pJET-*nptII* digested with XbaI/XhoI.

**pEXΔ*intS*, pEXΔ*rdfS*, pEXΔ*ahlI*.** Regions upstream and downstream of *intS, rdfS* and *ahlI* were amplified by PCR from WSM1271 genomic DNA and each

pair of corresponding upstream and downstream regions were cloned into SalI/BamHI digested pEX18Tc using Gibson assembly.

**pJQ∆*rdfG* and pJQ∆*rdfM.*** Regions upstream and downstream of *rdfG* and *rdfM* genes were amplified by PCR from WSM1271 gDNA. Upstream fragments were digested with SacI/XhoI, downstream fragments were digested with XbaI/NotI, and the pJET-*aadA1* plasmid was digested with XhoI/XbaI to release the Ω*aadA* cassette. These three products were ligated with SacI/NotI digested pJQ200 SK and plated onto LB agar supplemented with gentamycin to select for the pJQ200 SK backbone, and spectinomycin to select for Ω*aadA*. The unique arrangement of restriction sites ensured that the final constructs comprised the pJQ200 SK backbone carrying the Ω*aadA* cassette flanked by the upstream and downstream regions of *rdfG* or *rdfM.*

**pJQ∆*traI1.*** Regions upstream and downstream of the *traI1* gene was amplified by PCR from WSM1271 gDNA. The upstream fragment was digested with SacI/KpnI and the downstream fragment was digested with SalI/NcoI. The pJET-*aadA1* plasmid was digested with KpnI/NcoI to release the Ω*aadA* cassette. These three products were ligated with SacI/SalI digested pJQ200 SK and plated onto LB agar supplemented with gentamycin to select for integration of pJQ200 SK into the WSM1271 genome, and spectinomycin to select for Ω*aadA*. The unique arrangement of restriction sites ensured that the final constructs comprised the pJQ200 SK backbone carrying the Ω*aadA* cassette flanked by the upstream and downstream regions of *traI1.*

**pJQ∆*intM.*** The *nptII* gene and regions upstream and downstream of the *intM* gene were amplified by PCR from pJP2neo plasmid DNA (for *nptII*) or

WSM1271 genomic DNA (for *intM* regions), and ligated into SacI/BamHI-digested pJQ200 SK using Gibson assembly. The arrangement of homologous regions ensured that the final constructs comprised the pJQ200 SK backbone carrying the *nptII* gene flanked by the upstream and downstream regions of *intM.*

**pJP2-*intG* and pJP2-*intS*.** The *intG* and *intS* genes and upstream intergenic regions were amplified by PCR from WSM1271 DNA and cloned into pJP2 as BamHI fragments.

**pJP2-*intM*.** *intM* and the upstream intergenic region was amplified by PCR from WSM1271 DNA and cloned into pJP2 as a XhoI fragment.

**pJP2-*rdfG,* pJP2-*rdfM* and pJP2-*rdfS*.** The *rdfG, rdfM* and *rdfS* genes and upstream intergenic regions were amplified by PCR from WSM1271 DNA and cloned into pJP2 as HindIII-XbaI fragments.

**pPR3-*rdfG.*** The *rdfG* gene and its ribosome binding site (RBS) were amplified by PCR from WSM1271 DNA and cloned into pPR3 downstream of the *nptII* promoter as a BamHI-KpnI fragment.

**pPR3-*traI1* and pPR3-*mbrI.*** The *traI1* and *mbrI* genes and artificially introduced RBS sequences were amplified by PCR from WSM1271 DNA and cloned into pPR3 downstream of the *nptII* promoter as a KpnI fragments.

**pSacB-*rdfM*.** The *rdfM* gene and an artificially introduced RBS were amplified by PCR from WSM1271 genomic DNA and cloned into pSacB downstream of the *lac* promoter as a XbaI-SacI fragment.

**pSDz-*traR1* and pSDz-*mbrR*.** The *traR1* and *mbrR* genes and artificially introduced RBS sequences were amplified by PCR from WSM1271 DNA and cloned into pSDz downstream of the IPTG inducible promoter as PstI-XbaI fragments.

**pSDz-*traR2*.** The *traR2* gene and an artificially introduced RBS (Sequence info??) was amplified by PCR from WSM1271 DNA and cloned into pSDz downstream of the *lac* promoter as a SpeI-XbaI fragment.

**pSDz-*msi172171*.** The *msi172-msi171* ORFs and an artificially introduced RBS (same as above?) were amplified by PCR from WSM1271 genomic DNA and cloned downstream of the *lac* promoter of EcoRI/HindIII-digested pSDz using Gibson assembly.

**pSDz-P*rdfG*, pSDz-P*rdfM* and pSDz-P*rdfS*.** Non-coding regions upstream of the *rdfG, rdfM,* and *rdfS* genes (presumably capturing the native promoters) were amplified by PCR from WSM1271 DNA and cloned into pSDz downstream of the IPTG inducible promoter as XhoI fragments.

**pSDz-tb, pSDz-mb, pSDz-tb*traR1*, pSDz-tb*traR2* pSDz-tb*mbrR*, pSDz-mb, pSDz-mb*traR1*, pSDz-mb*traR2* and pSDzmb-*mbrR*.** Non-coding regions upstream of the *traI1* and *mbrI* genes were amplified by PCR from WSM1271 DNA and cloned into pSDz, pSDz-*traR1,* pSDz-*traR2* and pSDz-*mbrR* downstream of the promoterless *lacZ* genes as XhoI-BglII fragments. Plasmids are named pSDz, followed by the *lacZ*-promoter fusion (*tb* = *traI1* promoter, *mb* = *mbrI* promoter), followed by the LuxR-family gene carried by the parent vector.

**pSacB.** The *sacB* gene and promoter were amplified by PCR from pJQ200 SK plasmid DNA and cloned as a XhoI-BamHI fragment into pSRKKm.

**pTH3*attP*.** Regions capturing the *attP*$_G$, *attP*$_M$ and *attP*$_S$ sites of ICE*Mc*Sym$^{1271}$ were amplified by PCR from WSM1271 and cloned in the same orientation as in ICE*Mc*Sym$^{1271}$ into BamHI/NotI digested pJQ200 SK using Gibson assembly.

**pMINI3.** The *attP*$_G$-*attP*$_M$-*attP*$_S$ region was amplified by PCR from pTH3*attP* plasmid DNA and cloned as a KpnI-EcoRI fragment into pFUS2.

**pTHQPS-1.** Regions capturing ICE*Mc*Sym$^{1271}$ *attB*$_G$, *attB*$_M$, *attB*$_S$ sites and a *melR* region were amplified from WSM1271(pSDz-*traR1*) DNA and sequentially cloned into pTH3*attP* as NotI-SacI, SmaI-XbaI, fragments respectively.

**pSacB-*intG*, pSacB-*intM*, and pSacB-*intS*.** The *intG, intM* and *intS* genes and an artificially introduced RBS sequences were amplified from WSM1271 DNA and cloned into pSacB, downstream of the IPTG inducible promoter as SacI fragments.

**pSacB-*ahlI*.** The *ahlI* gene and native RBS were amplified from WSM1271 DNA and cloned into pSacB, downstream of the IPTG inducible promoter as a SacI-XbaI fragment.

## Table 2.2 Oligonucleotides used in this thesis

| No | Primer | [a] Sequence |
|---|---|---|
| **Cloning primers** | | |
| 1 | Ω*intG*_Up_5'_SacI | ATCAG**GAGCTC**AGGGCGAGTCGGATCCGG |
| 2 | Ω*intG*_Up_3'_XbaI | ATCAG**TCTAGA**AGGAGGAGACGAACTGGCGTAAC |
| 3 | Ω*intG*_Dn_5'_XhoI | ATCAG**CTCGAG**CGGGTCTCGTCTTCCGCG |
| 4 | Ω*intG*_Dn_3'_BamHI | ATCAG**GGATCC**GGCTCTCCATGGGCATGAC |
| 5 | *intM*_Up_5'_gib | **AGGGAACAAAAGCTGGAGCTC**ACATTGTAGGAATTCTCGC |
| 6 | *intM*_Up_3'_gib | **TCGCGCGGCC**TCAAATTGAGTCGGAACAAAC |
| 7 | *intM*_Dn_5'_gib | **TCGCCTTCTTGACGAGTTCTTCTGA**CAACGTTCCTTCCAGACTTTCTCC |
| 8 | *intM*_Dn_3'_gib | **TCGAATTCCTGCAGCCCGGGGGATC**GCGATATTGGGACGGGCTC |
| 9 | Δ*intS*_Up_5'_gib | **AGTGCCAAGCTTGCATGCCTGCAGG**TGCTGCGCCTCGACCGCC |
| 10 | Δ*intS*_Up_3'_gib | **CCCACCATTGCAT**CTCCCAAGGCCATAGGATCGGTAACC |
| 11 | Δ*intS*_Dn_5'_gib | **ATGGCCTTGGGAG**ATGCAATGGTGGGCCGATTATC |
| 12 | Δ*intS*_Dn_3'_gib | **TACGAATTCGAGCTCGGTACCCGGG**ATGGACGCTCTGCATAGGTTG |
| 13 | Ω*rdfG*_Up_5'_SacI | ATCAG**GAGCTC**AAGCAGCGTGACAAGCGGC |
| 14 | Ω*rdfG*_Up_3'_XhoI | ATCAG**CTCGAG**GTCAAATGGGATCGAGGATGACGG |
| 15 | Ω*rdfG*_Dn_5'_XbaI | ATCAG**TCTAGA**AATCCGTCGCGCCTCAATGT |
| 16 | Ω*rdfG*_Dn_3'_NotI | ATCAG**GCGGCC**GCTTGCCCGGCTGGGCCTT |
| 17 | Ω*rdfM*_Up_5'_SacI | ATCAG**GAGCTC**CCACGCAAGCGCAGCG |
| 18 | Ω*rdfM*_Up_3'_XhoI | ATCAG**CTCGAG**ACGCTTGTTGCGTATACGCTGTAGAC |
| 19 | Ω*rdfM*_Dn_5'_XbaI | ATCAG**TCTAGA**GGACGCTGCCTCGGTCCT |
| 20 | Ω*rdfM*_Dn_3'_NotI | ATCAG**GCGGCCGC**GTCACCTGTCAACGATCGGCAAG |
| 21 | Δ*rdfS*_Up_5'_gib | **ACTAAAGGGAACAAAAGCTGGAGCT**CGGCATCGTACCCCGGTCG |
| 22 | Δ*rdfS*_Up_3'_gib | **TGGGTGTGGTTCT**CCTTTTTGGCGCGGGCGG |
| 23 | Δ*rdfS*_Dn_5'_gib | **CGCGCCAAAAAGG**AGAACCACACCCATTCCAACGATG |
| 27 | Δ*rdfS*_Dn_3'_gib | **TTGGGTACCGGGCCCCCCCTCGAGG**TAGCGCTCGGGTCCGGCG |
| 28 | *intG*_5'_BamHI | ATCAG**GGATCC**TCAAATGGGATCGAGGATGACG |
| 29 | *intG*_3'_BamHI | ATCAG**GGATCC**GCGGAATTATTTGGCGGTAGATC |
| 30 | *intM*_5'_XhoI | ATCAG**CTCGAG**GCTCGTGCGGAAGGGATGA |
| 31 | *intM*_3'_XhoI | ATCAG**CTCGAG**TGATTATCTGACGATGCGCAGGT |
| 32 | *intS*_5'_BamHI | ATCAG**GGATCC**TCCCGACACTCCCTTTCGC |
| 33 | *intS*_3'_BamHI | ATCAG**GGATCC**ATGGCGCTTCAATCACTCTTCGC |
| 34 | *rdfG*_5'_HindIII | ATCAG**AAGCTT**GTTCGCCGTCCGCTCAATC |
| 35 | *rdfG*_3'_XbaI | ATCAG**TCTAGA**TCATCCTCGATCCCATTTGACG |
| 36 | *rdfM*_5'_HindIII | ATCAG**AAGCTT**AGCAAGCCTATTCTGGTGGCCG |
| 37 | *rdfM*_3'_XbaI | ATCAG**TCTAGA**TTATCGTTTTTCAACGTCCCGTTTGCT |
| 38 | *rdfS*_5'_HindIII | ATCAG**AAGCTT**GCCGAGGAGCGGCGAAA |
| 39 | *rdfS*_3'_XbaI | ATCAG**TCTAGA**TCATGAGCGGCCTCCATCGT |
| 40 | *rdfG*_5'_BamHI | ATGAC**GGATCC**ACATTGAGGCGCGACGGATT |
| 41 | *rdfG*_3'_KpnI | ATGAC**GGTACC**TCATCCTCGATCCCATTTGACG |
| 42 | *traI1*_5'_KpnI | ATCTA**GGTACC**GGAGGCGACGAATGATGCAGCTAATCACACCTGAGC |
| 43 | *traI1*_3'_KpnI | ATCTA**GGTACC**TTAAGCGTATGCCGGCAGGC |
| 44 | *rdfM*_5'_SacI | ATCAG**GAGCTC**GGAGGCGACGAATGAAGAGTGACGCAATCTCGTATGCC |
| 45 | *traR*_5'_PstI | ATCTA**CTGCAG**GGAGGCGACGAATGCATCGCGTGTTTGAAAATTTC |
| 46 | *traR*_3'_XbaI | ATCTA**TCTAGA**TCAGGATCTCGAATGTCGGGAA |
| 47 | *msi172*_5'_gib | **TAACAATTTCACACATAGCTAACTG**GGAGGCGACGAATGCCTGCAGTTCTCGTG |

| No | Primer | [a] Sequence |
|---|---|---|
| 48 | *msi171*_3'_gib | **CTTTAGATGCCGCTTCTTTTGCAGA**TCAAAGAAGGAAATCCCTGTACCC |
| 49 | P*rdfG*_5'_XhoI | ATGAC**CTCGAG**TGCTCGTGAGCAAGACCTAGGCTT |
| 50 | P*rdfG*_3'_XhoI | ATGAC**CTCGAG**AATCCGTCGCGCCTCAATGT |
| 51 | P*rdfM*_5'_XhoI | ATGAC**CTCGAG**TGGGTCGTTGATCGCCAGC |
| 52 | P*rdfM*_3'_XhoI | ATGAC**CTCGAG**GGACGCTGCCTCGGTCCT |
| 53 | P*rdfS*_5'_XhoI | ATGAC**CTCGAG**TCCGGCCGACCCGAG |
| 54 | P*rdfS*_3'_XhoI | ATGAC**CTCGAG**GATGATCCTCGTTTGGCTTGCG |
| 55 | *aadA1*_5'_Blunt | **ATGCATGTCGAC**GGAGCTGCATGTGTCAGAGGT |
| 56 | *aadA1*_3'_Blunt | **GAGCTCGGTACC**GAGGCCCTTTCGTCTTCAAGA |
| 57 | *rdfS*_5'_NcoI | ATAT**CCATGG**ACAACGAAAACGAACGCG |
| 58 | *rdfS*_3'_HindIII | ATAT**AAGCTT**TTATCATGAGCGGCCTCCATCG |
| 59 | *sacB_5'*_XhoI | ATCAG**CTCGAG**GCCAAAGAGCTACACCGACGAG |
| 60 | *sacB_3'*_BamHI | ATCAG**GGATCC**TAAATTGTCACAACGCCGCG |
| 61 | attP(M)_5'_Gib | TGGAGCTCCACCGCGGTGGCGGCCGCCTCGCTGAATGCAACATC |
| 62 | attP(M)_3'_Gib | CAATCCTAGTGAGAACTGGATGGTGCATG |
| 63 | attP(S)_5'_Gib | ATGCCCAATTCTCACTTTAATGGCTGCGATGAG |
| 64 | attP(S)_3'_Gib | CGAATTCCTGCAGCCCGGGGGATCCACCCAAAGCTGGAGCCCG |
| 65 | attP(G)_5'_Gib | TCCAGTTCTCAATGCCTCCCTCACCATAGC |
| 66 | attP(G)_3'_Gib | TTAAAGTGAGAATTGGGCATTACCCCGC |
| 67 | 3attP_5'_KpnI | ATCAGGGTACCCCTCGCTGAATGCAACATC |
| 68 | 3attP_3'_EcoRI | ATCAG**GAATTC**CCCAAAGCTGGAGCCC |
| 69 | 1271attB(G)_5'_NotI | ATCTA**GCGGCCGC**GAGATCCTGCGCGAAGCC |
| 70 | 1271attB(G)_3'_NotI | ATCTA**GCGGCCGC**TCTGAAATGAACGCTGCTTCATAAAG |
| 71 | 1271attB(M)_5'_SacI | ATCTA**GAGCTC**CGCTTCCGGGACGTTCAG |
| 72 | 1271attB(M)_3'_SacI | ATCTA**GAGCTC**TCGCCCGACACGATGATG |
| 73 | 1271attB(S)_5'_Blunt | TCTAGAGTCGAGAAGTGACACCAGCGG |
| 74 | 1271attB(S)_3'_ Blunt | AAGACATGTGACGGCGTTTCAG |
| 75 | 1271melR_5'_XbaI | ATCTA**TCTAGA**TTTGGGATGGATGTCGGCG |
| 76 | 1271melR_3'_XbaI | ATCTA**TCTAGA**CTGGGGCCAGCAGCGT |
| 77 | *intG_5'*_SacI | ATCAG**GAGCTC**GGAGGCGACGAATGCTCACAGACATCGCACTTAAGA |
| 78 | *intG_3'*_SacI | ATCAG**GAGCTC**TCAAATGGGATCGAGGATGACG |
| 79 | *intM_5'*_SacI | ATCAG**GAGCTC**GGAGGCGACGATGGCTAGGCCCTTTAAGGATGC |
| 80 | *intM_3'*_SacI | ATCAG**GAGCTC**TTATCTGACGATGCGCAGGTTT |
| 81 | *intS_5'*_SacI | ATCAG**GAGCTC**GGAGGCGACGAATGGCCCTTTCCGACGTAAAAT |
| 82 | *intS_3'*_SacI | ATCAG**GAGCTC**TCAATCACTCTTCGCCCTGG |
| 83 | tb_5'_XhoI | ATCTA**CTCGAG**TTGTCGCCTCCGTGCAGG |
| 84 | tb_3'_BglII | ATCTA**AGATCT**CGACATTCGAGATCCTGATTCCTT |
| 85 | mb_5'_XhoI | ATCTA**CTCGAG**GGCGCCCTCCCTTGGTC |
| 86 | mb_3'_BglII | ATCTA**AGATCT**CGCTTTCGATTGTCCGAGGG |
| 87 | *traR2*_5'_SpeI | ATCAG**ACTAGT**GGAGGCGACGAATGACGAGGGACATGCCACTTGT |
| 88 | *traR2*_3'_XbaI | ATCAG**TCTAGA**TCAGAGGATCGAGCTCCCTTGG |
| 89 | *mbrR*_5'_PstI | ATCTA**CTGCAG**GGAGGCGACGAATGATCGATTCGGATGTATTCGAAT |
| 90 | *mbrR*_3'_XbaI | ATCTA**TCTAGA**TTAGGGATGGATCATGCGCC |
| 91 | *mbrI*_5'_KpnI | ATCTA**GGTACC**GGAGGCGACGAATGATAGCGGCTCATGTCGTAAACG |
| 92 | *mbrI*_3'_KpnI | ATCTA**GGTACC**TCATTCGCACATTTGCCGATG |
| 93 | *nptII*_5'_Blunt | GAGCTCGGATCCGAGGTCCAAGGCGGCCG |
| 94 | *nptII*_3'_Blunt | CGCGTCAGACGCCCGTAGCATGCGAATTC |

| No | Primer | [a] Sequence |
|---|---|---|
| 95 | *nptII*_5'_gib | **TCGCCTTCTTGACGAGTTCTTCTGA**CAACGTTCCTTCCAGACTTTCTCCG |
| 96 | *nptII*_3'_gib | **TCGAATTCCTGCAGCCCGGGGGATC**GCGCGAACGCCTGCAAATG |
| 97 | *ahlI*_5'_SacI | ATCAG**GAGCTC**GACGGAGCGGATAATGACGATCTC |
| 98 | *ahlI*_3'_XbaI | ATCAG**TCTAGA**CTACGCTCTGTCGACGCTTGCC |
| 99 | *ahlI*_Up_5'_Gib | **AGTGCCAAGCTTGCATGCCTGCAGG**GAACGCTCGGCGCGTATTG |
| 100 | *ahlI*_Up_3'_Gib | **ACGCTCTGTCGAC**CTTCTGCGAGATCGTCATTATCCG |
| 101 | *ahlI*_Dn_5'_Gib | **GATCTCGCAGAAG**GTCGACAGAGCGTAGGTCCG |
| 103 | *ahlI*_Dn_3'_Gib | **TACGAATTCGAGCTCGGTACCCGGG**ACCGGCCGAATTCGTTGG |
| 104 | Ω*traI1*_Up_5'_SacI | ATCAG**GAGCTC**GGAATGTCACCAATTGGTGCAACA |
| 105 | Ω*traI1*_Up_3'_KpnI | ATCAG**GGTACC**TTGTCGCCTCCGTGCAGG |
| 106 | Ω*traI1*_Dn_5'_NcoI | ATCAG**CCATGG**GATCTTCCCACACTTGAAGGCGTC |
| 107 | Ω*traI1*_Dn_3'_SalI | ATCAG**GTCGAC**TCAGGTTTCGGCTAAGGGCAAG |
| **Attachment site qPCR primers** | | |
| 108 | *attB*(G)F | GCATCAACCGCGTCGTCTA |
| 109 | *attB*(G)R | GAAGTCTCCGGCAGCGAAA |
| 110 | *attB*(M)F | GCTCCAGGTGTGCGTTTCT |
| 111 | *attB*(M)R | TGGGTTGATTTGGGCGATCT |
| 112 | *attB*(S)F | TGTCTTTGGGCTTAGCGTTCT |
| 113 | *attB*(S)R | ACAGGCCCAGATAGCTCAGTT |
| 114 | *attP*(G)F | CAGTCTGCAGCAACGATGAC |
| 115 | *attP*(G)R | CAGTGTGTTGAAATTCCGGTTGA |
| 116 | *attP*(M)F | GACCGTGGTCTTTGCTTTGG |
| 117 | *attP*(M)R | TCTCCGAACGTCCGCAAA |
| 118 | *attP*(S)F | GGAACCGAACCAATCCACAGA |
| 119 | *attP*(S)R | TGCCGAAACAGAAGCGTAGA |
| 120 | *melR*F | CTGATGTCACCAGTGTTGCG |
| 121 | *melR*R | CGCCCAGGTCGAGGTTAATT |
| **Attachment site PCR primers for WSM1271 and R7ANS** | | |
| 122 | Mes-*guaA*F | TGACGGCGGATTTCTACCAC |
| 123 | Mes-*pheR* | TGCTATAACCCACGCGCT |
| 124 | Mes-*metR* | CGTAGAGCGCGATTATGGGT |
| 125 | R7A-*pheF* | TAGTCGCAGGAAACCCTTGG |
| 126 | R7A-*metF* | TGAGACGGACAAGACTGACG |
| 127 | R7A-*guaAR* | ACATAGGCCCTAACCTTCGC |
| **Exconjugant screening primers** | | |
| 128 | ICEMcSym1271-aF | CGAATCACCGGTGCATCAAC |
| 129 | ICEMcSym1271-aR | CTTGATGCAGCAGTGATGGC |
| 130 | ICEMcSym1271-bF | GCAGCGTTCATTCCGACTTG |
| 131 | ICEMcSym1271-bR | TCTGAGGCATCGCTTGGATC |
| 132 | ICEMcSym1271-gF | CATGTGGTTGGAACTGCTGC |
| 133 | ICEMcSym1271-gR | CCGCGCAGTATGAGGAGATT |
| 134 | MesGMCOF | GCCAAATGGTCGACGCTCTA |
| 135 | MesGMCOR | GTCCGACACGAACAGGTTCT |
| 136 | MesHPF | TGACGGCATCGATGATAGGC |
| 137 | MesHPR | GCGATGCAATGACAGGAACG |

[a] Text in bold demarks a restriction site, or overlapping region for Gibson (gib) cloning detailed in the primer name. Underlined text demarks an artificially introduced RBS.

## 2.4. General molecular techniques

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

### *2.4.1. Isolation of genomic and plasmid DNA*

Genomic DNA was extracted for whole-genome sequencing and PCR amplification of cloning products using a phenol:chloroform:isoamylacohol extraction protocol described previously (223). Crude lysates containing genomic DNA were prepared for analytical PCR reactions using the PrepMan Ultra Sample Preparation Reagent (Thermo Scientific) as per the manufactures recommendations. Plasmid DNA was isolated from *E. coli* strains using a Plasmid DNA Extraction Mini Kit (Favorgen Biotech Corp) as per the manufactures recommendations. DNA concentration and purity was analysed using a NanoDrop 1000 (ThermoFisher Scientific) and agarose gel electrophoresis (described in section 2.4.3) where required.

### *2.4.2. End-point PCR*

Primers used for end-point PCRs are listed in Table 2.1. Analytical end-point PCRs were performed in 10-µL volumes containing 500 nM of each primer, 1 µL of genomic or plasmid DNA (1-200 ng $\mu L^{-1}$) or cell lysate, 1 x GoTaq Green

Master Mix (Promega) and PCR grade milliQ water (Fischer Biotech). Cycling conditions were generally as follows 94°C 5 min (x 1); 94°C 30-s, 59°C 30 s, 70°C 60 s per kb (x 30); 70°C 5 min (x 1). PCR amplification of DNA for cloning was performed in 50-uL volumes containing 500 nM of each primer, 3% (v/v) dimethyl sulfoxide, 1 µL genomic or plasmid DNA (1-200 ng $\mu L^{-1}$) and 1 x phusion polymerase high fidelity master mix. Cycling conditions for phusion polymerase reactions were generally as follows; 98°C 30 s (x 1); 98°C 10 s, 72°C 30 s per kb (x 30); 72°C 3 min (x 1). An additional cycle of 98°C 10 s, 58-70°C 30 s, 72°C 30 s per kb (x 5) was also used following initial denaturation for primers which exhibited a Tm $\leq$ 70°C.

### *2.4.3. Agarose and Eckhardt gel electrophoresis*

PCR products and plasmids were electrophoresed at 9 V $cm^{-1}$ for 45 – 60min in 1-1.5% (w/v) agarose (Fisher Biotechnology) dissolved in 1 x TAE buffer (224). HindIII digested λ, 1 kb, or 100-bp DNA Ladders (New England Biolabs) were used as molecular weight markers. Eckhardt gel electrophoresis was performed as previously described (223), however cultures were initially grown to early log phase ($OD_{600nm}$ 0.1 - 0.3) in TY broths. Samples were electrophoresed in Eckhardt gels for 15 h at 4°C. Agarose and Eckhardt gels were post-stained in 1 x TAE or TBE, respectively containing 50 µL $L^{-1}$ ethidium bromide prior to visualization on a GelDoc XRand (BioRad).

### *2.4.4. Restriction endonuclease digestions, ligations and Gibson assembly*

Amplified PCR products used for cloning were purified prior to digestion/ligation using a FavorPrep™ GEL/PCR Purification Kit (Favorgen Biotech Corp) as per the manufactures recommendations. Restriction digestion and ligation of DNA was performed using the buffers, temperatures and incubation times indicated by New England Biolabs. Vectors were dephosphorylated by adding 3 U $\mu g^{-1}$ genomic DNA rSAP directly to the digestion reaction and continuing incubation at 37°C for 30 min. Restriction enzymes and rSAP were inactivated by incubation at 65°C for 20 min, or removed by phenol:chloroform:isoamylacohol extraction and ethanol precipitation (223) where products were digested with heat-tolerant enzymes. Ligation of cohesive end fragments was performed using T4 DNA ligase as per the manufactures recommendations and ligations for blunt fragments was performed as previously described (225). Gibson assembly was performed using Gibson Assembly HiFi Master mix as per the manufacturer's (NEB) recommendations, however, reaction volumes were scaled to 5 µL and incubation at 50°C was extended to 20 min.

### *2.4.5. Preparation and transformation of competent cells*

Chemical transformation or electroporation was used to transform Plasmids were transformed into *E. coli* either chemically or via electroporation. Chemically competent DH5α cells were purchased from Bioline, and chemically competent ST18 and electrocompetent DH10β cells were prepared as previously described (226, 227). For chemical transformations, 50-µL

competent cells were combined with 2 µL of DNA and 1 % (v/v) DMSO incubated on ice for 30 min, heat shocked at 42°C for 90 s and incubated on ice for a further 2 min. For electroporation, 1 µL of DNA was added to 40 uL of electrocompetent cells and transferred into a 0.2-mm gap cuvette (Fisher Biotechnology). A single 5-ms 2.5-kV pulse (25 µF resistance and 200 ohms) was delivered using a Gene Pulser II (BioRad). For both methods, 1 mL SOC recovery media (228) was immediately added directly to the transformation tube or cuvette and the entire contents transferred to a 10-mL falcon tube and incubated at 37°C on a rotary shaker (250 rpm) for 1 h. Transformation reactions were streaked or spread in serial dilutions onto selective LB agar media and incubated overnight at 37°C.

### *2.4.6. Conjugal transfer of plasmids*

*E. coli* plasmid donors and *Mesorhizobium* plasmid recipients were grown in 5-mL broths until mid-log phase ($OD_{600nm}$ 0.1 - 0.8) and late-stationary phase ($OD_{600nm}$ ~2.0), respectively. Cultures were harvested by centrifugation (5,500 x g 5 min), washed once with 1 mL saline (0.89% NaCl in sterile DDI water), and resuspended in 200 µl saline. 50-µl aliquots of donor and recipient, and a 100-µl aliquot of both strains combined (1:1) were spotted onto TY agar supplemented with ALA and incubated overnight at 28°C before streaking, or spreading serial dilutions onto selective media. Single colonies were passaged to remove potential contamination prior to use.

## 2.5. DNA sequencing and genome assembly

Sanger sequencing of PCR amplicons and plasmids was performed by the Australian Genome Research Facility. Whole-genome sequencing was performed by Macrogen (South Korea), using Pacific BioSciences (PacBio) single-molecule real-time cell-sequencing or Illumina HiSeq 2500 technology. Trimmed PacBio sequencing reads (post-filter) were generated as follows; NZP2037, 92,934 reads averaging 8,462-bp; WSM1284, 102,356 reads, average of 11,824-bp; WSM1497, 136,085 reads average of 4,057-bp. Illumina 2 x 100-bp paired-end reads (post-filter) were generated as follows; WSM1284, 21,189,686; WSM1497, 25,226,358. NZP2037 illumina reads were obtained from the Joint Genome Institute (177).

Quality control analyses and genome assemblies were carried out by Dr Joshua Ramsay using an Intel i7- 4790K, (32 Gb DDR4) desktop computer running Ubuntu Linux (v14.04). Raw Illumina reads were analysed using FastQC v0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Illumina sequence adapter contamination was removed using nesoni:clip (v0.132) (https://github.com/ Victorian-Bioinformatics-Consortium/nesoni) and reads were corrected using Lighter (v1.1.1) (47). Genome Filtered Illumina and PacBio reads were assembled *de novo* together with Illumina reads using the SPAdes assembler version 3.6.2 (11), with the number of mismatches and short indels reduced by incurring SPAdes's postprocessing module MismatchCorrector, utilizing the BWAtool (12). Assemblies were scaffolded using SSPACE version 3.0 (13) and annotated using the NCBI Prokaryotic Genome Annotation Pipeline (http://www .ncbi.nlm.nih.gov/genomes/static/Pipeline.html).

## 2.6. *Mesorhizobium* mating experiments

The stable broad host-range plasmids pFAJ1708 or pPR3 were mobilized into R7ANS to create tetracycline or neomycin resistant recipient strains, respectively. Cultures of donor and recipient strains for mating assays were grown in triplicate by inoculating TY broths from single colonies, and growing these cultures to saturation at 28°C. These were used to seed (1/100) fresh TY broth cultures which were incubated for at 28°C for 64-h on a rotary shaker (250 rpm). Cells were harvested by centrifugation (5,500 x g, 5 min) and washed once with 1-mL saline before re-suspending cells in 200 uL saline. Individual 50-µL aliquots and combined 100-µL (1:1) aliquots of donor and recipient strains were spotted onto TY agar and grown for 24-h at 28°C. Mating spots were collected with a sterile inoculation loop, resuspended in saline and spread in 3-fold or 10-fold serial dilutions onto selective G/RDM agar supplemented with thiamine and the appropriate antibiotic. Serial dilutions of the donor strain were also spread onto non-selective media to determine the number of colony forming units in each mating. These plates were grown for 8 - 10 days at 28°C.

## 2.7. Quantitative PCR assays for ICE excision

Cultures for qPCR were grown in triplicate by inoculating TY broths from single colonies and growing these cultures to saturation at 28°C. These were used to seed (1/100) fresh TY broths which were grown for 64 h (or 10-80 h for growth curve experiment), prior to extracting genomic DNA using the PrepMan

Ultra Sample Preparation Reagant (Applied Biosystems) as per the manufacturer's recommendations. PrepMan Ultra genomic DNA samples were stored at -14°C until required. For Fig 3.4, the number of colony forming units in each culture sample was calculated at the time of DNA extraction by serially diluting cells and spreading them onto non-selective G/RDM agar.

The quantitative PCR (qPCR) assay described in reference (74) was adapted to measure the abundance of ICE*Mc*Sym$^{1271}$ *attP* and *attB* sites relative to the chromosomal copy number (*melR* gene), revealing the percentage of cells within a population that had undergone each Int-mediated excisive recombination reaction. The primers for the assay are detailed in Table 2.2 & Fig 3.3. qPCR was performed using an Applied Biosystems ViiA 7 Real-Time PCR System with default cycling conditions. Reactions were carried out in 20-µL volumes containing 10 µL of 2 x SYBR select master mix (Applied Biosystems), 500 nM of each primer and 1-µL of DNA sample prepared using the PrepMan Ultra Sample Preparation Reagent (Thermo Fischer) as per the manufactures recommendations.

The amplification efficiency for each primer pair was initially determined using the qPCR standard construct, pTHQPS-1 (Table 2.1), which carries a region of *melR* and each ICE*Mc*Sym$^{1271}$ *attP* and *attB* site in the same order and orientation as found in the WSM1271 chromosome, as a template. The construct was initially linearized by NcoI-digestion and serially diluted in PCR grade milliQ water. qPCR of serially diluted pTHQPS-1 was used to generate standard curves for each qPCR by plotting relative DNA concentration versus the log($C_t$) value (Fig 2.1). Amplification efficiency ($E$) for each primer pair was calculated using the formula $10^{(-1/C_{tslope})}$ where $C_{tslope}$ is the slope of the $C_t$

values. All qPCR primers had efficiency values between 1.88 and ~2.0, and each assay was accurate over template concentrations spanning approximately six orders of magnitude (12–32 amplification cycles). In qPCR assays for ICE[3] excision, *attP* and *attB* relative abundance values were derived by normalizing results obtained for each PCR reaction against that of *melR,* using the following previously described formula (17):

$$R_{att} = \frac{(E_{att})^{\Delta C_{t(att)}}}{(E_{melR})^{\Delta C_{t(melR)}}}$$

Where: $E_{att}$ represents the efficiency of either *attP* or *attB* PCR reactions for each ICE*Mc*Sym[1271]-encoded *att* site; $E_{melR}$ represents the efficiency of the *melR* PCR reaction; $\Delta C_{t(att)}$ represents the difference in the $C_t$ values between a reference PCR (where linearized pTHQPS-1 is the template), and the relevant *attP* and *attB* PCR; and $\Delta C_{t(melR)}$ represents the difference in the $C_t$ values between a reference PCR and the *melR* PCR reaction.

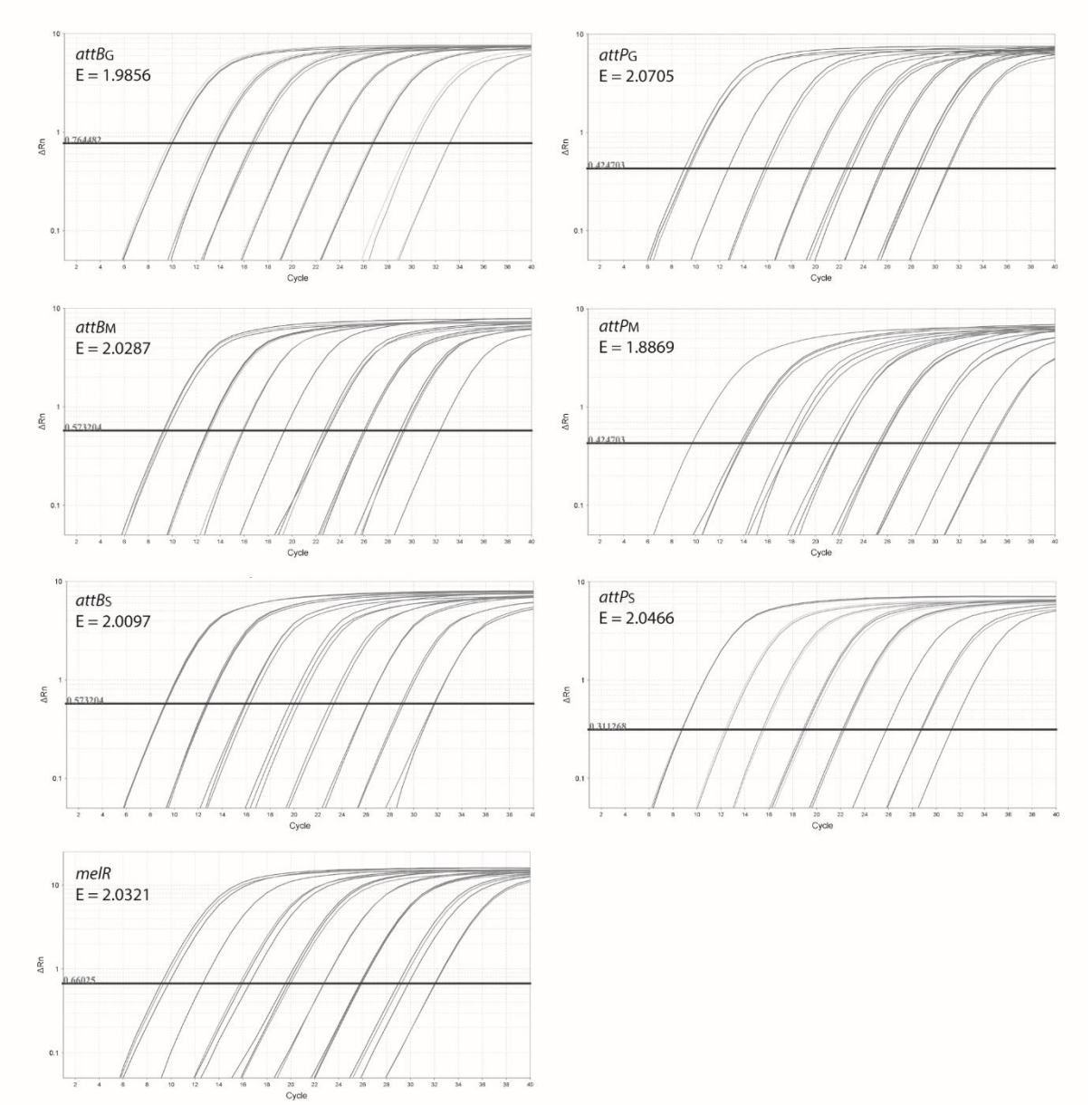


**Figure 2.1. Standard curves for qPCR assays.** Standard curves were derived by performing qPCR on the serially diluted qPCR standard construct, pTHQPS-1, then plotting relative DNA concentration versus the log($C_t$) value. Amplification efficiency ($E$) for each primer pair was calculated using the formula $10^{(-1/\text{slope})}$.

## 2.8. β-galactosidase assays

TY broths were seeded 1/100 with stationary-phase TY broth cultures of *Mesorhizobium* strains and grown for 24 h at 28°C. A 200-µL aliquot of the resulting cultures were transferred into clear-bottom 96-well culture plates and $OD_{600nm}$ was recorded on a Enspire Multimode Plate Reader (PerkinElmer). Samples were frozen at -80°C until required. β-galactosidase assays were performed as previously described (229) with three biological repetitions per treatment, unless otherwise specified.

## 2.9. Melanin deposition assays

For melanin deposition assays (230, 231), 20-µL aliquots of stationary-phase TY broth cultures were spotted onto TY agar supplemented with 600 µg $mL^{-1}$ *L*-tyrosine and 40 µg $mL^{-1}$ $CuSO.5H_2O$ (TYT agar). These plates were incubated at 28°C for 14-days.

## 2.10. Bioassays for detection of AHLs

### *2.10.1. CV026 bioassays*

For CV026 streak bioassays, strains of interest were streaked adjacent to the biosensor strain *C. violaceum* CV026 on LB agar and these plates were incubated for 48 h at 28°C. For CV026 well-diffusion bioassays, supernatants were collected from 50 mM MOPS buffered (pH 6.5) TY or LB broth cultures following centrifugation at 12,000 x g for 5-min. These supernatants were sterilized through a 0.22-µm syringe filter and 100-µL aliquots were loaded into wells bored into 1.5% (w/v) LB agar over-layed with molten 0.3% (w/v) LB agar

seeded 1/100 with a saturated LB broth culture of CV026. These plates were incubated at 28°C for 48 h.

### *2.10.2. AHL inactivation assays*

AHL-inactivation assays were performed using a method adapted from Chan *et al* (232). Cultures were initially grown to stationary-phase in TY or LB broths buffered with 50 mM MOPS (pH 6.5). Cultures were supplemented with 10 µM 3-oxo-$C_6$-HSL and incubated at 28°C for 6 – 12 h. Five-hundred-microlitre samples of sterile supernatant were collected (as described in Section 2.10.1) before and after incubation and the pH of these samples was recorded. One-hundred-microlitre aliquots of the sterile supernatants were subject to CV026 well-diffusion bioassays as described in section 2.10.1. Intracellular AHL-inactivation assays were performed using the same procedure as described above, however broth cultures were sterilized through a 0.22-µM pore syringe filter to remove cells prior to the addition of 3-oxo-$C_6$-HSL.

## 2.11. RNA Sequencing

All equipment and benches were decontaminated with RNase*Zap*™ RNase decontamination solution prior to use. RNA quality and concentration was analysed at various points throughout processing using an Experion™ StdSense or HighSens analysis kit assays (Bio-Rad Technologies).

### *2.11.1. Isolation of total RNA*

Cultures (three biological replicates per treatment) for RNA-Seq libraries were initially grown by streaking single colonies onto TY agar slopes and incubating these cultures for 5-days at 28°C. Two technical replicates of each slope culture was washed off into 50-mL TY broths and incubated for 24 h ($OD_{600}$ ~ 0.8). Twelve-millilitres of each broth culture was added to 24 mL of RNA later solution (Qiagen) and cells were collected by centrifugation (10,000 x g 10-m 4°C). Supernatant was removed by aspiration and the cell-pellet was resuspended in 250 µL of 10 mM Tris-Cl (pH 8.0). This cell-suspension was added to 2-mL lysis tubes filled with: 300 mg silica beads (0.1 mm), 100 mg glass beads (0.1 mm), 350 µL RLT buffer (Qiagen) and 3.5 µL β-mercaptoethanol and mechanically lysed in a FastPrep®-24 instrument (MP biomedicals) at speed 6.5 for 30 s. Total RNA was extracted from the lysate using a RNeasy Mini Kit (Qiagen) as per the manufacturers recommendations. RNA concentration and quality was initially analysed using a NanoDrop 1000 (ThermoFisher Scientific). To remove residual DNA, approximately 3 µg of total RNA was digested with a TURBO DNA-free™ kit (Invitrogen) as per the manufacturers recommendations, and DNA removal was confirmed with a Qubit fluorometer dsDNA BR assay (ThermoFisher Scientific).

### *2.11.2. cDNA library construction and sequencing*

rRNA was depleted from total RNA using a Ribo-Zero rRNA magnetic kit (Illumina) as per manufacturers recommendations, and the resulting RNA was purified using a RNA Clean & Concentrator™ kit (Zymo Research). rRNA-depleted RNA samples were fragmented, hybridised to adapters, reverse

transcribed to cDNA, amplified, barcoded and purified using the Ion Total RNA-Seq kit v2 (Thermo Fisher) as per the manufacturers recommendations. Barcoded cDNA libraries were diluted to 75 pM and pooled for template preparation using an Ion Chef™ instrument (Thermo Fisher). Sequencing was performed using Ion Proton™ system (Thermo Fisher). Read sets from technical repetitions were combined.

### *2.11.3. Read mapping, read counting and statistical analyses*

Adapter sequences were removed using nesoni clip (http://www.vicbioinformatics.com/software.nesoni.shtml). To reduce any potential rRNA/total-RNA abundance biases introduced during rRNA depletion, reads mapping to rRNA genes were removed using FastQ Screen (https://www.bioinformatics.babraham.ac.uk). Reads were mapped to the WSM1271 genome (accession NC_014923) using Bowtie 2 (233) and visualised using Artemis (234). For gene expression analysis, read sets were additionally filtered to remove sequences matching plasmids pPR3-*traI1* and pSDz-*traR1* prior to mapping. An average (per biological replicate) of 14-million (standard deviation (SD) = 3.3-million) QS+ and 8.5-million (SD = 1.5-million) QS- post-filter reads were mapped to WSM1271 with 96.7-98.6% alignment rate. Read counts for gene features were performed using HTSeq (235) with default settings then imported into DESeq2 (236) for identification of differentially expressed genes.

To measure expression from the *traI1* and *traI2* promoter regions, the unfiltered reads were mapped to the WSM1271 chromosome using the

procedures described above and read counting was performed using the "--nonunique all function" on HTSeq so that reads mapping ambiguously to the *traI1* and *traI2* regions and ORFs were counted for both features.

## 2.12. Glasshouse procedures

*Biserrula pelecinus* L. was obtained from Dr Ron Yates (Murdoch University, Western Australia). *B. pelecinus* was grown in free-draining pots containing lawn sand as previously described (192). Pots were steam sterilised for 2.5 h, followed by two flushes with boiling water prior to sowing. Prior to sewing, seeds were lightly scarified with sand-paper and surface sterilised by submersion in 70% (v/v) ethanol, followed by 3% (v/v) NaOCl and then rinsed in sterile DI water 5 times. Plants were initially watered with 20 mL nutrient solution (192) supplemented with 1.5 mM $KNO_3$, then subsequently watered with 20 mL aliquots of N-free nutrient solution once per week. N-fed control plants were supplemented weekly with 5 mL 100 mM $KNO_3$. Seven days after inoculation, sterile alkathene beads were distributed evenly over the surface of the soil to maintain sterility and prevent restriction of plant growth. Plants were grown for 8-weeks and plant shoot dry weights were excised above the cotyledon and individually dried in polypropylene tubes for 2 days at 60°C prior to weighing. The glasshouse experiment was block-randomised with five pot replications, each containing four plants.

## 2.13. General bioinformatics and statistics

For whole-genome BLASTN comparisons in Figs 3.1 & 6.1 BRIG (v0.9.5) (237) was used to produce BLASTN (options: -ungapped, -word_size 2000, upper and lower threshold 99%) alignments of sequence contigs or scaffolds. Nucleotide and amino-acid alignments were performed using the T-Coffee multiple sequence aligner on default settings (238). Synteny were performed using the Artemis Comparison Tool (239) on default settings, and plotted using GenplotR (240). Construction of primers and general sequence analysis was performed using Geneious (v9.1.8) (241). General BLAST searches were carried out using either the NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) or IMG (242) databases as detailed in results. Statistical analyses were performed using Rstudio and are described in figure captions.

# Chapter 3.

## Discovery of tripartite ICEs

## 3.1. Introduction

Horizontal transfer of the *M. loti* R7A symbiosis ICE (ICE*Ml*Sym$^{R7A}$) drives the evolution of novel *Lotus*-nodulating rhizobia in New Zealand soils (74, 160, 162-164, 243). Since these studies, putative symbiosis ICEs have been identified in numerous sequenced *Mesorhizobium* genomes (74-76, 167, 170, 176). Integration of ICE*Ml*Sym$^{R7A}$ is catalysed by the integrase IntS, which recombines the attachment site $attP_S$ located on the circularisedICE*Ml*Sym$^{R7A}$ with $attB_S$ located at the 3′ end of the sole *phe*-tRNA gene present in *Mesorhizobium* genomes. This recombination integrates ICE*Ml*Sym$^{R7A}$, producing the flanking attachment sites *attL* and *attR*. Excision of ICE*Ml*Sym$^{R7A}$ is stimulated by a recombination directionality factor RdfS, which reverses the favoured direction of IntS-mediated recombination toward formation of *attP* and *attB* (12, 17).

Following introduction of the *Biserrula pelecinus* inoculant *Mesorhizobium ciceri* bv. biserrulae WSM1271 (and later WSM1497) into Australian agriculture, the emergence of genetically distinct *B. pelecinus*-nodulating strains was observed (193, 194). Four of these strains (*M. opportunistum* WSM2075 and *M. australicum* strains WSM2073, WSM2074 and WSM2076) were found to carry *nifH* and *nodA* genes 100% identical to the original inoculant WSM1271 and an integrase homologous to ICE*Ml*Sym$^{R7A}$ IntS adjacent to the *phe*-tRNA gene (60, 74, 160, 193). Thus, it was postulated that WSM1271 may carry a symbiosis ICE resembling ICE*Ml*Sym$^{R7A}$ and that transfer of this ICE from WSM1271 into soil *Mesorhizobium* spp. resulted in the evolution of these novel *B. pelecinus*-nodulating organisms (194).

In this chapter, the identity and mechanism of excision/integration and transfer of the *M. ciceri* WSM1271 symbiosis ICE ICE*Mc*Sym$^{1271}$ and two *B. pelecinus*-nodule-isolated strains *M. opportunistum* WSM2073 and *M. australicum* WSM2075 (204-206) were explored.

## 3.2. Results

### *3.2.1. Three co-transferrable DNA regions in the WSM1271 chromosome*

Whole-genome BLASTN comparisons of WSM2073 and WSM2075 with the WSM1271 genome identified three distinct regions each with near-perfect nucleotide identity to regions found in each of the three strains (Fig 1). The first region, denoted α, was 445,220 bp in WSM1271 and was identical in WSM2073 and WSM2075, aside from a point mutation within a single putative transposase gene (Mesci_5575). The second largest region, denoted β, was 22,971 bp, and the smallest region, denoted γ, was 7,760 bp. The β and γ regions were identical in each strain. The chromosome-region junctions in each genome assemby were confirmed by PCR-amplification (Table 2.3) and sequencing, discounting the possibility that the separation of these three regions was an artefact of *de novo* genome assembly.

The near-identical sequence of α, β and γ in both the original inoculant strain WSM1271 and the novel nodule-isolated symbionts strongly suggested that these three regions had been acquired together through horizontal gene transfer and that they were likely involved in the capacity of WSM2073 and WSM2075 to nodulate *B. pelecinus*. To investigate whether α, β and γ transferred independently or in combination, conjugation experiments were

conducted using each of these strains as donors together with the non-symbiotic *M. loti* strain R7ANS (74) carrying pFAJ1708 as a recipient. Genes for nicotinate and biotin biosynthesis were identified on region α in WSM1271, WSM2073 and WSM2075 genomes, providing a mechanism of selection for region α, since R7ANS is auxotrophic for both vitamins (4). Potential R7ANS exconjugants were selected by growth on minimal medium with tetracycline but lacking biotin and nicotinate. Following these mating experiments, tetracycline resistant biotin and nicotinate prototrophs were isolated at a frequency of $4.65 \times 10^{-8} \pm 7.89 \times 10^{-9}$ (SE) in experiments using WSM1271 as a donor, $8.5 \times 10^{-9} \pm 8.5 \times 10^{-10}$ from WSM2073 donors and $3.0 \times 10^{-9} \pm 6.0 \times 10^{-10}$ from WSM2075 donors. Exconjugants from 16 independent conjugation experiments were screened by PCR targeting loci on regions α, β, γ and additionally a PCR specific for the R7ANS chromosome (Table 2.2). Despite only selecting for transfer of biotin and nictotinate genes on α region, all regions α, β and γ were detected in all exconjugants suggesting that all three regions had been acquired in these strains in all experiments. The genome of a one exconjugant (R7ANS×WSM1271) from a mating using the WSM1271 donor was also draft sequenced using Illumina techology. Whole-genome BLASTN comparison of the *de novo*-assembled R7ANS×WSM1271 genome with the WSM1271 genome confirmed complete transfer of all three regions and integration of each of these regions α, β and γ in the same relative position and orientation as found in the genomes of WSM1271, WSM2073 and WSM2075 (Fig 3.1).

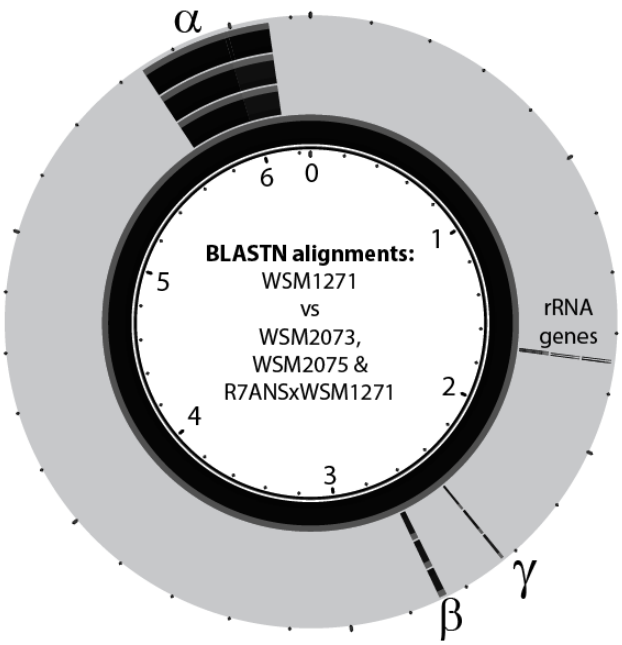


**Figure 3.1 Conservation of three ICE*Mc*Sym$^{1271}$ regions in WSM1271 and exconjugants.** Circular BLASTN alignments carried out using BRIG (237) of WSM1271 with WSM1271, WSM2073, and WSM2075 (204-206) and the laboratory ICE*Mc*Sym$^{1271}$ exconjugant R7ANS×WSM1271. Black regions indicate >99% conserved nucleotide identity.

### *3.2.2. Three integrases and three pairs of attachment sites*

Analysis of gene content of regions α, β and γ revealed *intS* gene identified in previous work was located on region γ downstream of the *phe*-tRNA gene and demarcated one boundary of this region in each strain (Fig. 3.2). Region γ carried a second putative integrase gene *intM* located adjacent to the *met*-tRNA gene, marking the other region γ boundary. Region β was located adjacent to the GMP-synthase gene *guaA* and harboured a third integrase gene *intG*, the product of which resembled integrases associated with MGEs that integrate into *guaA* (87). Region α, despite being the largest region, did not carry an identifiable integrase gene.

Integrase-targeted *attL* sites are composed chromosomal DNA 5′ of the insertion site and ICE DNA 3′ of the insertion site, whereas *attR* sites contain ICE DNA 5′ of the insertion site and chromosomal DNA 3′ insertion site (74, 95, 103-106). Therefore, all four *att* site types are structurally distinct and can be distinguished from each other through inspection of flanking DNA together with the relative orientation of the core sequence. *attL* sites are also commonly located adjacent to the integrase gene which facilitates ICE recombination (244). Because of the association of identified integrase genes *intS*, *intG* and *intM* with *phe*-tRNA, *guaA* and *met*-tRNA genes (respectively), it was predicted that these likely represented the integrase-targeted insertion sites for each region. The 17-bp core sequence associated with the ICE*Ml*Sym$^{R7A}$ integrase IntS is 5′-TCCGCCTCTGGGCACCA-3′. The same sequence was identified at the 3′ end of the γ-region boundary within the WSM1271 *phe*-tRNA gene which was denoted *attL*$_S$ (Fig 3.2). Another copy of the IntS core sequence was identified at the 3′ boundary of the α region and was denoted *attR*$_S$. The conserved core sequence targeted by *guaA*-associated integrases has previously been identified as 5′-GAGTGGGA-3′ (87). Two 11-bp repeat sites (5’-ATCGAGTGGGA-3’) containing this sequence were identified in the WSM1271 chromosome; one within the *guaA* gene at the 5′ end of the β-fragment – here named *attL*$_G$, and one at the 3′ end of the α-fragment – here named *attR*$_G$. Finally, a perfect duplication of the 16 bp sequence 5′-CCCTCCGGGCCCACCA-3′ was identified at the 5′ end of region γ within the end of the *met*-tRNA gene and at the 3′ of region β. These were named *attL*$_M$ and *attR*$_M$, respectively. In summary, regions α, β and γ were each bordered by putative integrase attachment core sites associated with two different

integrases (Figs 3.2 & 3.3) and together the three regions carried three integrases and three pairs of attachment sites.

Excision and circularisation of DNA located between *attL* and *attR* sites requires that their core regions form a directly-orientated repeat. However, *attR*$_S$ on region α was inverted relative to *attL*$_S$, indicating recombination of *attL*$_S$ and *attR*$_S$ to produce *attP*$_S$ and *attB*$_S$ would result in an inversion and juxtaposition of regions α and γ (Fig 3.3 *state ii*). Similarly, IntG-mediated recombination of convergently oriented *attL*$_G$ and *attR*$_G$ would produce *attP*$_G$ and *attB*$_G$ and juxtapose fragments α and β (Fig 3.3 *state iv*). Finally, *attL*$_M$ and *attR*$_M$ were in the same orientation, so their recombination would excise DNA between them, leaving behind *attB*$_M$ and juxtaposing regions β and γ on a circular 248-kb DNA fragment carrying *attP*$_M$ and 218 kb of chromosomal DNA (Fig 3.3 *state iii*).

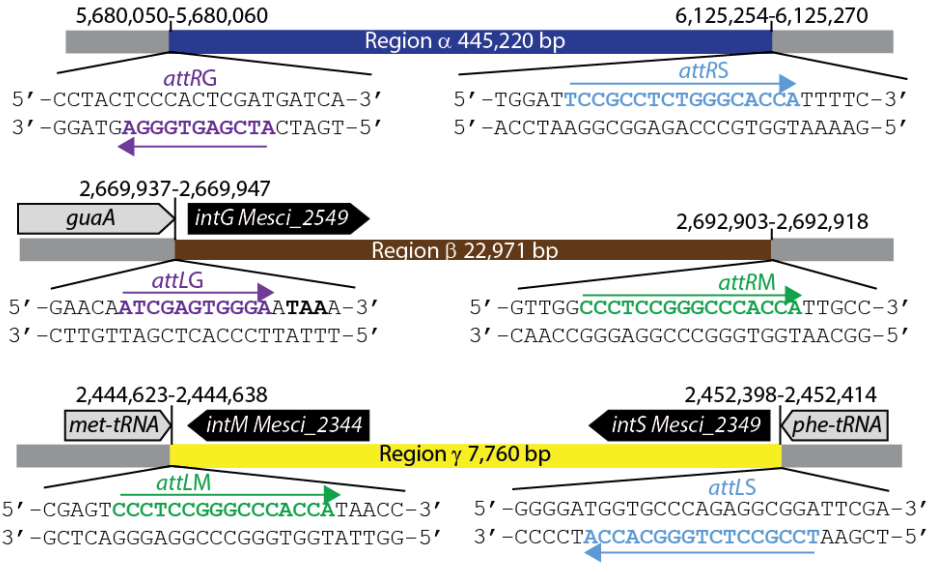


**Figure 3.2 Schematic of ICE*Mc*Sym$^{1271}$ regions α, β and γ, and predicted *att* site core sequences.** Coloured arrows represent orientation of matching *attL* and *attR* site sequences, chromosomal DNA is coloured in grey. Schematics are not to scale. Genome (NC_014923.1) coordinates for each *att* site are provided above each region.

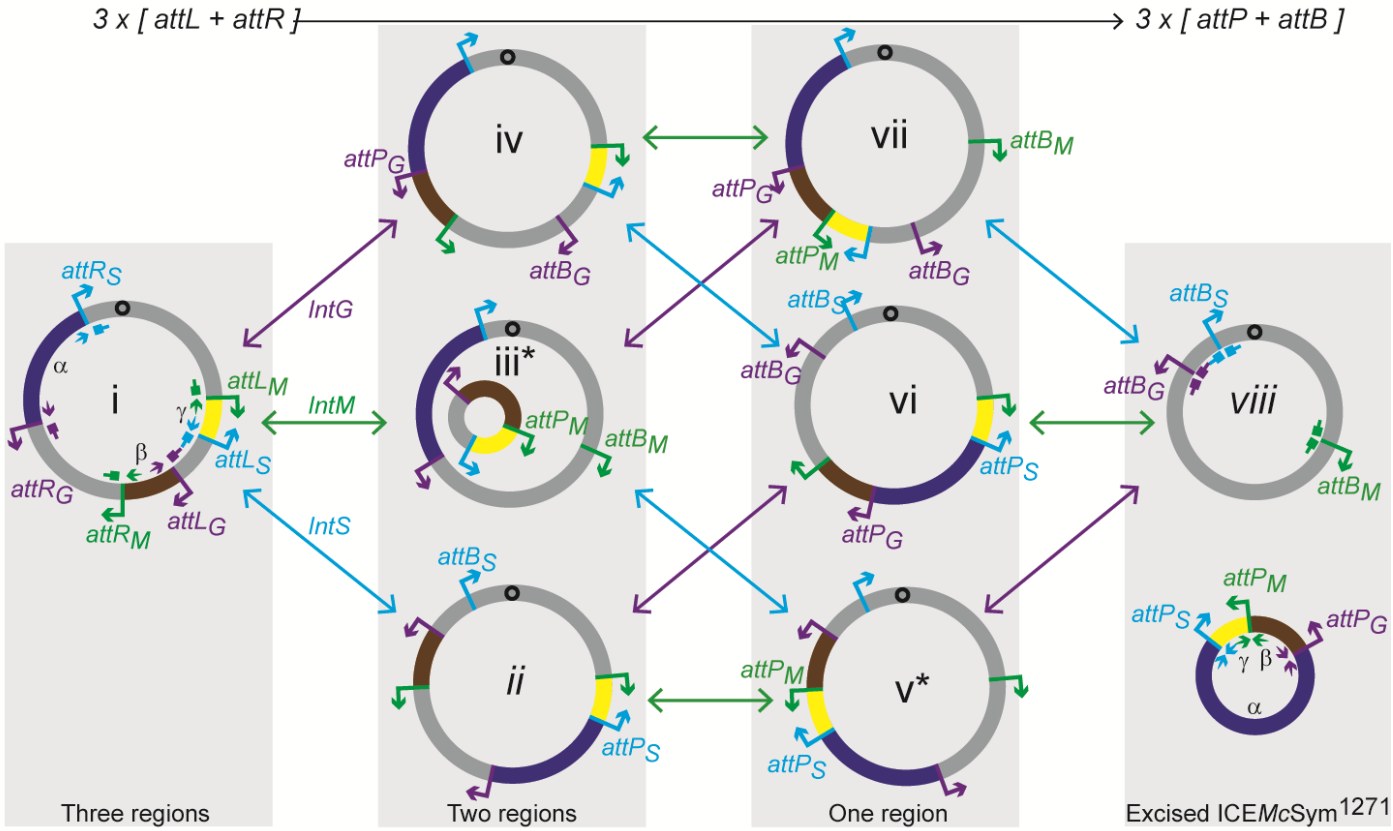


**Figure 3.3. Model of ICE*Mc*Sym$^{1271}$ recombination states i–viii of the WSM1271 chromosome obtained through the actions of IntS, IntG and IntM.** The α, β and γ regions are coloured dark blue, brown and yellow, respectively. IntG and associated *att* sites are coloured magenta, IntM and associated *att* sites are coloured green, IntS and associated *att* sites are coloured cyan. Chromosomal DNA is coloured in grey and is fixed at the WSM1271 origin of replication, indicated by an "O." The orientation of each *att* site is indicated by an arrow (5′–3′ direction) at each region boundary. The binding sites for primers used to amplify each *attB* and *attP* site are shown are displayed on the figure as block-headed arrows and tapered arrows, respectively. Diagrams are not to scale.

### *3.2.3. Coordinated formation of three pairs of attB and attP sites*

The frequency of symbiosis ICE excision in *M. loti* R7A was previously calculated by quantitative PCR (qPCR) measuring *attP* and *attB* abundance relative to the chromosomal gene *melR* (74). Here, this assay was adapted to detect and measure *attP* and *attB* formation in WSM1271. Six pairs of primers for each of the 3 *attP* and 3 *attB* sites were designed along with primers for WSM1271 *melR*. WSM1271 was grown in broth culture and genomic DNA was extracted at 10-h intervals for qPCR (Fig 3.4). All *attP* and *attB* products were detected in DNA extracted from all samples and sequencing of qPCR products confirmed recombination had occurred within each predicted core site. The relative abundance of each *attP* and *attB* pair was comparable for *att* sites of the same type, consistent with the interdependent production of *attP* and *attB* sites from corresponding *attL* and *attR* sites. The *attP* and *attB* sites for IntS and IntG were detected in ~0.01% of cells in log-phase growth (20-h) and this increased to 0.1% of cells in stationary-phase cultures (50 h onwards). While the proportion of IntM attachment sites $attP_M$ and $attB_M$ detected also increased ~10-fold between log phase and stationary phase, the overall abundance of these sites was ~10-fold less than those for IntS and IntG.

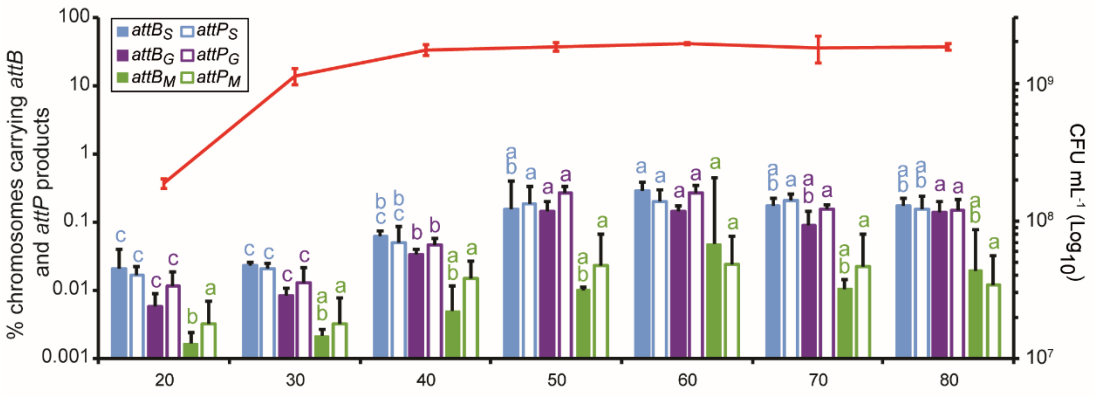


**Figure 3.4. qPCR detection of *attP* and *attB* formation in WSM1271.** Measurements represent the mean percentage of WSM1271 chromosomes in TY batch cultures harbouring each excisive Int-mediated recombination product (*attB*$_S$, *attP*$_S$, *attB*$_G$, *attP*$_G$, *attP*$_M$ *and attP*$_M$) determined by qPCR. Samples of genomic DNA were extracted, and viable cell counts (red line) were performed at 10-h intervals for 80 h. Values for each of the assay types *attB*$_S$, *attP*$_S$, *attB*$_G$, *attP*$_G$, *attP*$_M$ and *attP*$_M$ site were individually compared between time points using ANOVA and Fisher's LSD test controlling for type I error using the Bonferroni adjustment. Groups of values from the same assay type that are not significantly different from each other have the same letter (a, b or c) indicated above.

### *3.2.4. Formation of three pairs of attP and attB sites requires IntG, IntM and IntS*

To explore the function of the three integrases IntG, IntM and IntS, each gene was inactivated in WSM1271. Both *intG* and *intM* were replaced with *nptII* (creating WSM1271Δ*intG*::*nptII* and WSM1271Δ*intM*::*nptII,* respectively) and a markerless deletion was made in the *intS* gene (creating WSM1271Δ*intS*). qPCR of DNA extracted from stationary-phase cultures of these strains following 64 h incubation (Fig. 3.5) revealed that while each *attP and attB* product was detected in the wild-type control strain, *attP*$_G$, *attB*$_G$, *attP*$_M$, *attB*$_M$, *attP*$_M$ and *attB*$_M$ formation could not be detected in WSM1271Δ*intG*::*nptII,* WSM1271Δ*intM*::*nptII,* or WSM1271Δ*intS*, respectively. In each mutant,

formation of the remaining *attP* and *attB* products remained relatively unaffected. Introduction of cloned wild-type copies of *intG* and *intS* restored *attP* and *attB* detection in WSM1271Δ*intG*::*nptII* and WSM1271Δ*intS*, respectively demonstrating that IntG and IntS, catalyse the excisive recombination reactions producing $attP_G$ and $attB_G$ or $attP_S$ *and* $attB_S$, respectively. In contrast, defective $attP_M$ and $attB_M$ production could not be complemented by introduction of a cloned wild-type copy of *intM*, suggesting that replacement of the *intM* ORF in WSM1271Δ*intM*::*nptII* may have some polar effect on the production of $attP_M$ and $attB_M$.

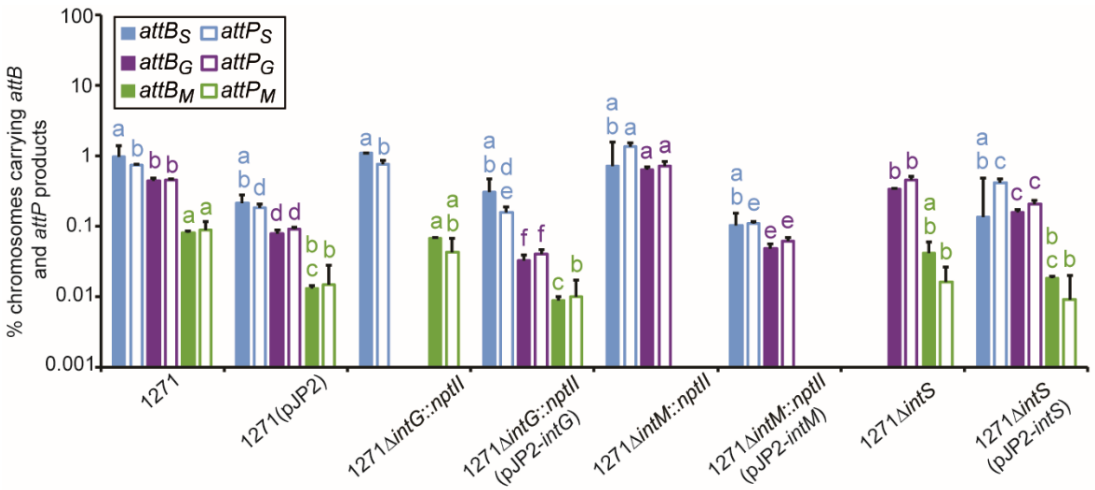


**Figure 3.5. qPCR detection of *attP* and *attB* formation in integrase mutants of WSM1271.** Measurements represent the mean percentage of WSM1271 chromosomes in stationary-phase cultures harbouring each excisive Int-mediated recombination product ($attB_S$, $attP_S$, $attB_G$, $attP_G$, $attP_M$ *and* $attP_M$) determined by qPCR. Where appropriate, plasmids carried by WSM1271 (here abbreviated as 1271) are listed in brackets after the strain name (see Table 1.1 for a description of plasmids). Values for each of the assay types $attB_S$, $attP_S$, $attB_G$, $attP_G$, $attP_M$ and $attP_M$ site were individually compared between strains using ANOVA and Fisher's LSD test controlling for type I error using the Bonferroni adjustment. Groups of values from the same assay type and in the same panel that are not significantly different from each other have the same letter (a, b, c, d or e) indicated above.

### *3.2.5. Reconstruction of ICEMcSym[1271] integration and disassembly pathways*

Given that the symbiosis ICE of WSM1271 harbours three sets of attachement sites and three integrases (*intS*, *intG* and *IntM*) we speculated that recombination of regions α, β and γ leads to the formation of a a single circular "tripartite" ICE (ICE$^3$) in the donor prior to conjugative transfer. To define the potential recombination pathways, a network diagram was created guided by the position and orientation of each *att* site and the predicted products of each recombination (Fig. 3.3). In this network, eight possible recombination states (*states i-viii*) were predicted, with each state able to transition to three other states through the action of one of the three integrases. This model indicated that sequential action of each integrase in any order to form three pairs of *attB* and *attP* sites would result in excision of a single circular ICE$^3$ and restoration of the likely ancestral WSM1271 chromosome. The model also suggested that the three reverse reactions (forming *attL* and *attR* sites) in combination would integrate the putative circular ICE$^3$ and disassemble it back into the tripartite configuration observed in all ICE$^3$ exconjugants. To test this model, a non-replicative mini-ICE$^3$ plasmid (pMINI3) that contained each of the three *attP* sites arranged in the same order and orientation as on the circular ICE$^3$ predicted in state viii (Fig. 3.3A) was constructed. pMINI3 confers gentamicin resistance but does not replicate in *Mesorhizobium*, so recombination with the chromosome is required for it to be maintained. To facilitate this recombination, expression plasmids carrying *intS* (pSacB-*intS*), *intG* (pSacB-*intG*) and *intM* (pSacB-*intM*) downstream of the pSacB *lac* promoter were constructed. pSacB is a derivative of the BHR vector pSRKKm (220) carrying

a copy of the *Bacillus subtilis sacB* gene (245), enabling selection for loss of each pSacB plasmid by exposure to sucrose.

Sequence analysis of R7ANS revealed it carried *attB*$_S$, *attB*$_G$ and *attB*$_M$ in the same relative position and orientation as predicted for the WSM1271 chromosome when cured of the ICE[3] (Fig 3.3 *state viii*) and lacked genes for *intS, intG* and *intM*. Each pSacB plasmid was separately introduced into R7ANS. pMINI3 was then conjugated into each of the three strains and colonies harbouring integrated pMINI3 were selected on medium containing gentamicin. Integration of pMINI3 was observed in each strain carrying an integrase-expressing pSacB plasmid, but not in a strain carrying an empty pSacB vector, confirming dependence of pMINI3 integration on the presence of an integrase gene (Fig 3.6A). PCR of predicted pMINI3-chromosome junctions were used to confirm *attB*$_S$::pMINI3 and *attB*$_M$::pMINI3 insertion occurred in the predicted regions in R7ANS(pSacB-*intS*) and R7ANS(pSacB-*intM*), thus reconstructing recombination states vii and vi, respectively (Fig 3.6A) (Fig 3.6B). Although pMINI3 R7ANS(pSac-*intG*) integrants were isolated, their PCR profiles did not match those predicted for state v, as individual colonies lacked either *attL*$_G$ or *attR*$_G$ (Fig 3.6B), suggesting pMINI3 had integrated elsewhere in the R7ANS chromosome.

*attB*$_M$::pMINI3 and *attB*$_S$::pMINI3 integrants were further manipulated by curing the pSacB-*int* plasmid and separately introducing each of the two other types of pSacB-*int* plasmid. Following IPTG induction, randomly selected single colonies were isolated, cured of the pSac-*int* plasmid and again screened by PCR to confirm the recombination state of each isolate. Using this approach states *iv* and *ii* were derived from *state vi*, and *state iv* was successfully derived

from *state vii*. However, attempts to derive *state iii* from *state vii* were unsuccessful, producing an unexpected *attB*$_G$ PCR product (Fig 3.6B). Moreover, the 248,280-bp region which would presumably exist separated from the chromosome in this state could not be detected by Eckhardt gel electrophoresis (Fig 3.7).

Finally, to stimulate the formation of *state i*, the final tripartite cojnfiguration observed for WSM1271, WSM2073 and WSM2075, each previously unintroduced pSacB-*int* plasmid was introduced into strains in *states ii* and *iv*. Following IPTG induction and curing of pSacB-*int*, PCR screens confirmed the conversion of strains in *states ii* and *iv* to *state i*. Sequencing of PCR amplicons of all *attL* and *attR* junctions amplified from the two independently derived *state i* strains confirmed the predicted pMINI3-chromosome recombination junctions. In summary, this experiment demonstrated that six of the eight predicted recombination states could be isolated solely via the sequential expression of the three ICE*Mc*Sym$^{1271}$ encoded integrases (Fig. 3.6).

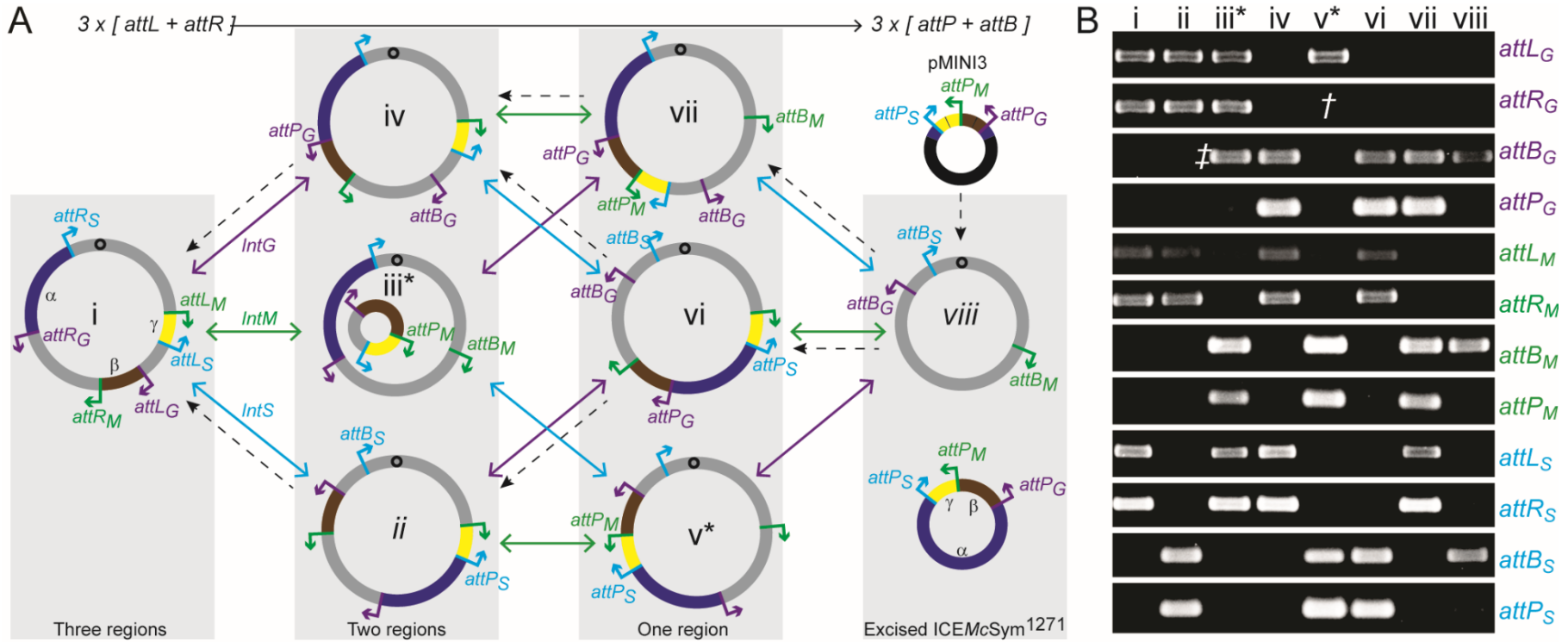


**Figure 3.6. Reconstruction of ICE*Mc*Sym$^{1271}$ Integration and Disassembly Pathways in *M. loti* R7ANS.** (A) Colours correspond to Fig. 3.3." The orientation of each *att* site is indicated by an arrow (5′–3′ direction) at each region boundary. Diagrams are not to scale. (B) DNA gels of PCR products amplified from the 12 *att* sites (rows) in each of the eight possible recombination states (columns) from *M. loti* R7ANS carrying the pMINI3 plasmid depicted in (A). Each dashed-line arrow represents a pathway successfully demonstrated using pMINI3. *Recombination *states iii* and *v* were not obtained. An example PCR profile from a single isolate is shown for each of these recombination states; recombination state iii showed the presence of an unexpected PCR product for *attB*$_G$ (†), whereas recombination *state v* lacked an expected *attR*$_G$ PCR product (‡).

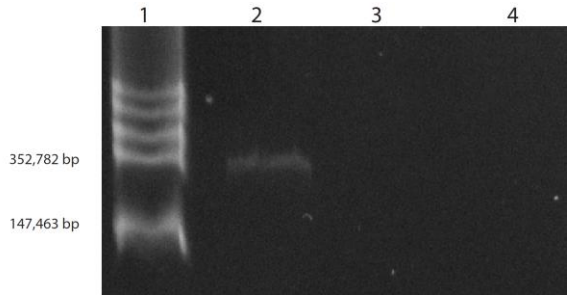


**Figure 3.7. Eckhardt gel electrophoresis of R7ANS and R7ANS-*attB*$_M$::pMINI3(pSac-*IntG*).** Lane 1: *Rhizobium leguminosarum* 3841 (sizes of lower two bands are indicated on left); Lane 2: *M. loti* NZP2037, revealing plasmid pRlo2037; Lane 3: *M. loti* R7ANS; Lane 4: *M. loti* R7ANS carrying the *attB*::pMINI3 insertion following introduction and curing of plasmid pSacIntG. DNA is the same as in PCR profile iii* in Fig. 3.6.B.

### *3.2.6. Symbiotic phenotypes of R7ANS ICEMcSym$^{1271}$ exconjugants*

Strains WSM1271, WSM2073 and WSM2075 all harbour a near identical ICE[3] with an identical compliment of symbiosis genes on this element. However, WSM1271 is an effective $N_2$-fixing microsymbiont of *B. pelecinus*, whereas WSM2073 is only partially effective and WSM2075 nodulates but does not fix $N_2$ with this host (194). To assess the symbiotic properties of R7ANS exconjugants harbouring ICE*Mc*Sym$^{1271}$, *B. pelecinus* was inoculated with nine R7ANS exconjugants, one derived from each of three independent matings with each of WSM1271, WSM2073 and WSM2075. Plants were grown for 8 wk before recording shoot dry weights (Fig. 3.8). All R7ANS exconjugants yielded weights comparable to that of the partially effective strain WSM2073, irrespective of the symbiotic proficiency of the donor strain from which their ICE[3] originated.

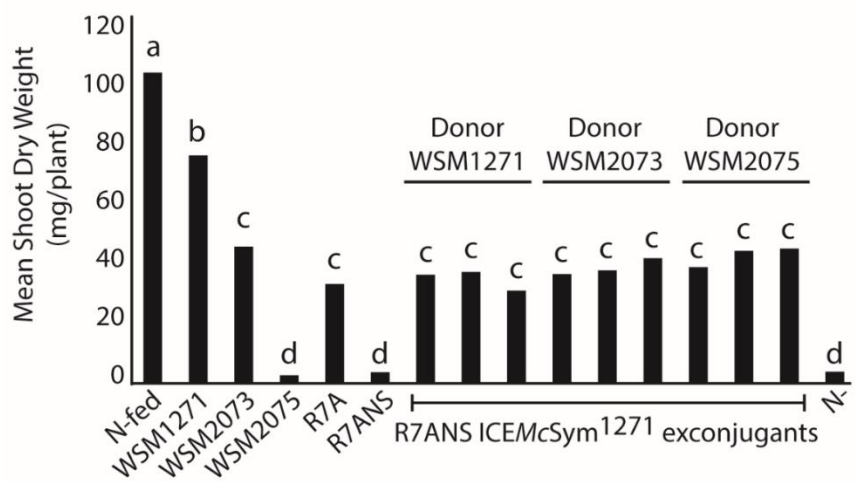


**Figure 3.8. Effectiveness of ICE*Mc*Sym$^{1271}$ exconjugants on *B. pelecinus*.** *B. pelecinus* plants grown in nitrogen-limited conditions were inoculated with indicated strains and grown for 8 wk. Uninoculated and nitrogen-fed (supplied as $KNO_3$) plants were included as negative and positive controls, respectively. Each bar represents the mean dry shoot weight for 20 plants split between five position-randomized pots. Shoot dry weights were compared using one-way ANOVA followed by Tukey's honest significant difference post hoc test at 5% significance. Treatments that share a letter are not significantly different. R7ANS did not nodulate *B. pelecinus*.

## 3.3. Discussion

In this chapter, it was demonstrated that the *Biserrula pelecinus*-nodulating mesorhizobia WSM1271, WSM2073 and WSM2075 each carry a novel symbiosis ICE ICE*Mc*Sym$^{1271}$ composed of three separated chromosomal regions of DNA termed α, β and γ. Each ICE*Mc*Sym$^{1271}$ region is distinct from that of previously characterised symbiosis ICEs (59, 60). Integrases IntS, IntG and IntM each facilitated recombination between specific pairs of *attL* and *attR* sites to form corresponding pairs of *attP* and *attB* sites. Therefore, it is likely that regions α, β and γ assembled into a single transferrable entity via the sequential action of three ICE*Mc*Sym$^{1271}$-encoded integrases prior to transfer. Following transfer, the circular ICE$^3$ is likely able to integrate into any one of the three *attB* sites in a mesorhizobial chromosome and disassemble into the

three regions via the action of the three integrases acting on the three pairs of *attP and attB* sites. Acquisition of this ICE$^{3}$ by R7ANS conferred an ability to nodulate and fix $N_2$ with *B. pelecinus,* albeit partially effectively.

Tyrosine recombinases like the ICE*Ml*Sym$^{R7A}$ integrase, IntS, catalyse the reversible recombination of *attL* + *attR* ←→ *attP* + *attB* (74). For a "single-part" ICE, the forward reaction excises and circularises the ICE, whereas the reverse reaction integrates the ICE (74, 92). For the ICE$^{3}$ ICE*Mc*Sym$^{1271}$, the recombination reaction substrates and products are similar, but the macromolecular rearrangement depends on the relative positions and orientations of three pairs of *att* sites (Fig. 3.3). The action of any

single integrase is inadequate for excision of the ICE*Mc*Sym$^{1271}$, but the combined forward (i.e. *attL* + *attR* → *attP* + *attB*) actions of the three integrases excises this ICE$^{3}$. It is therefore possible that the forward reactions are coregulated. qPCR analysis revealed the abundance of all three pairs of *attP* and *attB* sites increased ~10-fold in stationary-phase cultures (Fig 3.4). A caveat of the qPCR assay is that it averages the ensemble of recombination states in a population, so further single-cell experiments would be necessary to confirm that the three reactions occur concurrently in the same cell. Nevertheless, co-transfer of all three ICE*Mc*Sym$^{1271}$ fragments by conjugation is consistent with all three forward reactions occurring together in single cells to facilitate excision and circularisation of the three ICE$^{3}$ regions prior to conjugal transfer (77, 78).

Using pMINI3 and sequential expression of each integrase, the formation of six of the eight predicted recombination states was demonstrated (Fig 3.6A &

B). However, *states iii* and *v* could not be reproduced. When pMINI3 was introduced into R7ANS(pSacB-IntG) to produce *state v*, isolated colonies had lost both $attP_G$ and $attB_G$, suggesting recombination had occurred as expected. However, individual colonies were positive for either $attL_G$ or $attR_G$, but not both (Fig. 3.6B † symbol). Further inspection of the R7ANS chromosome revealed the presence of an additional copy of the $attB_G$ core sequence within Meslo_RS0109425 (NZ_KI632510). This second $attB_G$ (not present in WSM1271, WSM2073 or WSM2075), together with overexpression of IntG from pSacB-*intG*, may have led to additional IntG-mediated recombination events, destroying one of the $attL_G$ or $attR_G$ sites in each isolate. Interestingly, these secondary recombination events were not apparent in the genomes of the sequenced exconjugants WSM2073, WSM2075 and R7ANS×WSM1271, so this phenomenon could be limited to the artificial system in this study.

Recombination *state iii* is the only state that is predicted to split the chromosome into two parts. The smaller portion (248,280 bp) harbours regions β and γ along with the *guaA* and *phe*-tRNA genes, but appears to lack an origin of replication. Presumably, state iii is not viable, because post-segregational loss of the excised region would result in loss of the sole *phe*-tRNA gene (163). In attempts to recombine pMINI3 from *state ii* to *state vi* using pSacB-*intG*, secondary recombination events mediated by IntG may have reintegrated this fragment into the main chromosome, resulting in the rescue of these recombined cells and the unexpected PCR profile (Fig. 3.6B † symbol). Eckhardt gel DNA electrophoresis did not identify an episomal fragment in the 250-kbp size range. Interestingly, the IntM-mediated excision products $attP_M$

and *attB*$_M$ were the lowest-abundance products detected using the qPCR assay. This finding implies that *attL*$_M$ + *attR*$_M$ → *attP*$_M$ + *attB*$_M$ may be the last or lowest-rate reaction, or that non-viable cells in *state iii* are lost from cell populations. If *attP*$_M$ and *attB*$_M$ formation is the final step in excision of the assembled ICE*Mc*Sym$^{1271}$ (transition *state vi* → *viii* in Fig. 3.3), then recombination *state iii* would be avoided during the recombination pathway that produces circularised ICE*Mc*Sym$^{1271}$.

The data presented in this Chapter strongly support the hypothesis that following agricultural introduction of WSM1271, ICE*Mc*Sym$^{1271}$ was transferred from WSM1271 to the progenitors of WSM2073 and WSM2075, converting them into *B. pelecinus*-nodulating strains (193, 194). Although all three of these strains carry the same ICE$^3$, WSM1271 fixes $N_2$ effectively with *B. pelecinus,* WSM2073 fixes $N_2$ partially effectively and WSM2075 does not fix $N_2$ with this host (194). Transfer of ICE*Mc*Sym$^{1271}$ from any of these donors to *M. loti* R7ANS here converted all recipients to partially effective $N_2$-fixing symbionts of *B. pelecinus*. R7A, which carries ICE*Ml*Sym$^{R7A}$, also exhibits partially effective $N_2$ fixation with *B. pelecinus* relative to WSM1271. Overall, these data suggest that the chromosomal background of symbiosis ICE recipients is a crucial factor governing proficiency for $N_2$ fixation. The evolution of poorly $N_2$-fixing rhizobia may pose a significant problem associated with legume inoculation in agriculture, because ineffective strains may dominate soil populations and reduce crop productivity (46, 160, 193, 194). These experiments provide insight into how ineffective rhizobia can evolve through ICE$^3$ transfer.

# Chapter 4.

## Diversity and evolution of ICE$^{3}$s

## 4.1. Introduction

ICE*Mc*Sym$^{1271}$ comprises a novel tripartite ICE that exists in WSM1271, WSM2073 and WSM2075 as three entirely separated chromosomal DNA regions that recombine the host chromosome and assemble into a contiguous DNA region prior to excision and conjugative transfer. In this chapter, the unique features of ICE*Mc*Sym$^{1271}$ are used as a tool to identify similar ICE$^{3}$s in other mesorhizobia isolated from geographically diverse locations. A bioinformatical analysis of these newly discovered ICE$^{3}$s was undertaken to expand our understanding of the diversity and evolutionary history of these elements.

## 4.2. Results and discussion

### *4.2.1. ICE$^{3}$s exist in genetically diverse mesorhizobia*

BLASTP searches were carried out against sequenced mesorhizobial genomes using the IntS, IntG and IntM amino acid sequences as queries. All three integrase genes were identified ($\underline{\geq}$ 70% amino acid similarity to WSM1271) in the *B. pelecinus* symbionts isolated from Ethiopia and Greece, *M.* sp. AA22 and *M. ciceri* bv. biserrulae WSM1497; the *Anthyllis vulneraria* symbiont *Mesorhizobium metallidurans* STM2683 (28); the *Bituminaria bituminosa* symbiont *M. ciceri* WSM4083; and the *Lotus* spp. symbionts *M. loti* strains WSM1293, NZP2037 (29), NZP2042 and SU343 (Table 4.1).

**Table 4.1. ICE$^{3}$s identified in genetically diverse *Mesorhizobium* spp.**

| Strain | Origin | Genome status[a] | Accession | Legume host(s) | IntG Locus ID[b] | IntM Locus ID[b] | IntS Locus ID[b] | α size (bp) | β size (bp) | γ size (bp) | Total size (bp) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Mesorhizobium ciceri* bv. biserrulae WSM1271 | Bottida, Sardinia, Italy | F | NC_014923.1 | *Biserrula pelecinus* | Mesci_2349 | Mesci_2549 | Mesci_2344 | 445,220 | 22,972 | 7,759 | 475,951 |
| *M. australicum* WSM2073 | Northam, Australia | F | NC_019973.1 | *B. pelecinus* | MESAU_RS11700 | MESAU_RS13120 | MESAU_RS11675 | 445,217 | 22,972 | 7,759 | 475,948 |
| *M. opportunistum* WSM2075 | Northam, Australia | F | NC_015675.1 | *B. pelecinus* | MESOP_RS12425 | MESOP_RS13575 | MESOP_RS12395 | 445,217 | 22,972 | 7,759 | 475,948 |
| *M. loti* NZP2037[c] | New Zealand | F | NZ_CP016079 | *Lotus* sp. | A9174_12560 | A9174_14520 | A9174_12535 | 528,481 | 27,584 | 6,202 | 562,267 |
| *M. ciceri* bv. biserrulae WSM1284[c] | Siniscola, Sardinia, Italy | F | CP015064.1 | *B. pelecinus* | A4R29_21670 | A4R29_20625 | A4R29_21700 | 538,710 | 17,280 | 7,793 | 563,783 |
| *M.* sp. AA22 | Ethiopia | D | LYTO00000000 | *B. pelecinus* | A9K68_RS20060 | A9K68_RS19980 | A9K68_RS19980 | ND[d] | ND | 16,798 | ND |
| *M. ciceri* bv. biserrulae WSM1497[c] | Greece | F | LYTN00000000 | *B. pelecinus* | A9K65_RS20900 | A9K65_RS19910 | A9K65_RS19895 | 443,511 | 19,359 | 5,412 | 468,282 |

| Strain | Origin | Genome status[a] | Accession | Legume host(s) | IntG Locus ID[b] | IntM Locus ID[b] | IntS Locus ID[b] | α size (bp) | β size (bp) | γ size (bp) | Total size (bp) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *M. loti* NZP2042 | New Zealand Palmerston North | D | LYTK00000000 | *Lotus* sp. | A8145_RS28465 | A8145_RS32320 | A8145_RS28430 | ND | ND | 13,384 | ND |
| *M. loti* SU343 | NSW, Australia | D | LYTL00000000 | *Lotus* sp. | A9K72_RS00560 | A9K72_RS34880 | A9K72_RS00585 | ND | ND | 6,202 | ND |
| *M. loti* WSM1293 | Serifos, Greece | D | AZUV00000000.1 | *Lotus* sp. | WP_050596358.1 | WP_027160978.1 | WP_027163438.1 | ND | 31,761 | 11,501 | ND |
| *M. ciceri* WSM4083 | Canary Islands, Spain | D | JAFG00000000.1 | *Bituminaria bituminosa* | MESCI2DRAFT_00027460 | MESCI2DRAFT_00025030 | MESCI2DRAFT_00003130 | ND | 3,672 | ND | ND |
| *M. metallidurans* STM 2683 | Languedoc, France | D | CAUM00000000.1 | *Anthyllis vulneraria* | WP_008877493.1 | WP_040593992.1 | WP_008877522.1 | ND | ND | ND | ND |
| *M. ciceri* Ca181 | India | D | NZ_CM002796.1 | *Cicer arietinum* | M1C_RS32775 | M1C_RS32875* | M1C_RS05150 | ND | 3,356 | 19,340 | ND |

[a]Genome status D indicates a draft sequence, F indicates finished (or completed).

[b] Integrase protein sequences are ≥70% similar to the relevant WSM1271 homologue, except for the *Ca181 IntM protein, for which the coding sequence has undergone several frameshift mutations. Ca181 *intM* nucleotide sequence is 87% (1032/1192) identical to *intM* from WSM1271.

[c] Coordinates for newly discovered ICE[3] regions in complete genomes are as follows; ICE*Mc*Sym[1284]- α, 858,217-1,396,927; ICE*Mc*Sym[1284]-β, 4,374,751-,4,392,030; ICE*Mc*Sym[1284]-γ, 4,618,567-4,626,359; ICE*Mc*Sym[1497]- α, 6,100,975–6,544,486; ICE*Mc*Sym[1497]-β, 2,746,844–2,766,245; ICE*Mc*Sym[1497]-γ, 2,527,429–2,532,841; ICE*Ml*Sym[2037]-α, 6,351,799-6,880,279; ICE*Ml*Sym[2037]-β, 3,031,348-3,058,941; and ICE*Ml*Sym[2037]-γ, 2,577,913-2,584,147. For a full list of ICE[3] *att* site and region coordinates, see reference (4)

[d] ND indicates not determined.

The drought and salt-tolerant chick-pea symbiont, *M. ciceri* Ca181 (212) also carried homologues of IntG and IntS, but not IntM. Nevertheless, a BLASTN search for the nucleotide sequence of *intM* against the Ca181 genome revealed the presence of an *intM* homologue 87% (1032/1192) identical to that of ICE*Mc*Sym[1271] which carried several critical point mutations rendering this allele a likely pseudogene. Thus, *intM* may be a pseudogene in Ca181, but the full complement of ICE[3] integrase genes were present (Table 4.1). In addition to the previously published genome sequences listed above, the genomes of the *B. pelecinus* symbionts *M. ciceri* bv. biserrulae WSM1284 and WSM1497 were sequenced and assembled in this work (CP015064 for WSM1284 and CP021070 for WSM1497). All three ICE[3] integrase proteins were identified by BLASTP on the chromosomes of each of these mesorhizobia.

To delineate the α, β and γ regions of these potential ICE[3], BLASTN searches for the three pairs of *attL* and *attR* core sites corresponding to IntS, IntG and IntM were performed. All *attL* and *attR* core sites were identified in all but three strains. $attR_S$ was not identified in AA22 and STM 2683 and $attR_G$ was not identified in Ca181 (Table 4.1).

Completely assembled genome sequences are available for three strains predicted to carry an ICE[3]: NZP2037, WSM1284 and WSM1497. For WSM1284 and WSM1497, ICE[3] regions α, β and γ regions were identified in the same relative position, order and orientation as observed in WSM1271. However, as previously noted by others, the 7.5 Mbp assembly of the sequenced NZP2037 genome is likely incorrect (177). Specifically, the single-

contig assembly of NZP2037 encompasses pRlo2037, which is an experimentally confirmed extrachromosomal plasmid (56, 67, 177). To amend this issue, SMRT-cell sequencing was performed on genomic DNA extracted from this strain and sequence reads were combined with previously acquired short-read paired-end sequences to create a hybrid *de novo* genome assembly. Two circular contigs were assembled, corresponding to the NZP2037 chromosome (NZ_CP016079) and pRlo2037 (NZ_CP016080), respectively. Using this corrected genome assembly, chromosomal regions corresponding to ICE*Mc*Sym$^{1271}$ regions α, β and γ were identified in the same relative position, order and orientation as located in WSM1271. The coordinates for ICE$^{3}$ regions in WSM1284, WSM1497 and NZP2037 are provided in Table 4.1.

To test the ability of predicted ICE$^{3}$s to transfer, NZP2037, NZP2042 and SU343 were used as donors in conjugation experiments carried out (by Dr John Sullivan at the University of Otago) using R7ANS carrying pFAJ1708 as the recipient, as described in Chapter 3. Putative ICE$^{3}$ exconjugants were isolated from all three matings and were confirmed to nodulate the host legume of the donor strains *L. pedunculatus*. R7ANS did not nodulate this host. Exconjugants were re-isolated from *L. pedunculatus* nodules and draft-sequenced. Whole-genome BLASTN comparisons of the *de novo*-assembled exconjugant genomes R7ANS×NZP2037, R7ANS×NZP2042 and R7ANS×SU343 with the corresponding donor genome sequences confirmed transfer of regions α, β and γ from all three donors (Fig. 4.1).

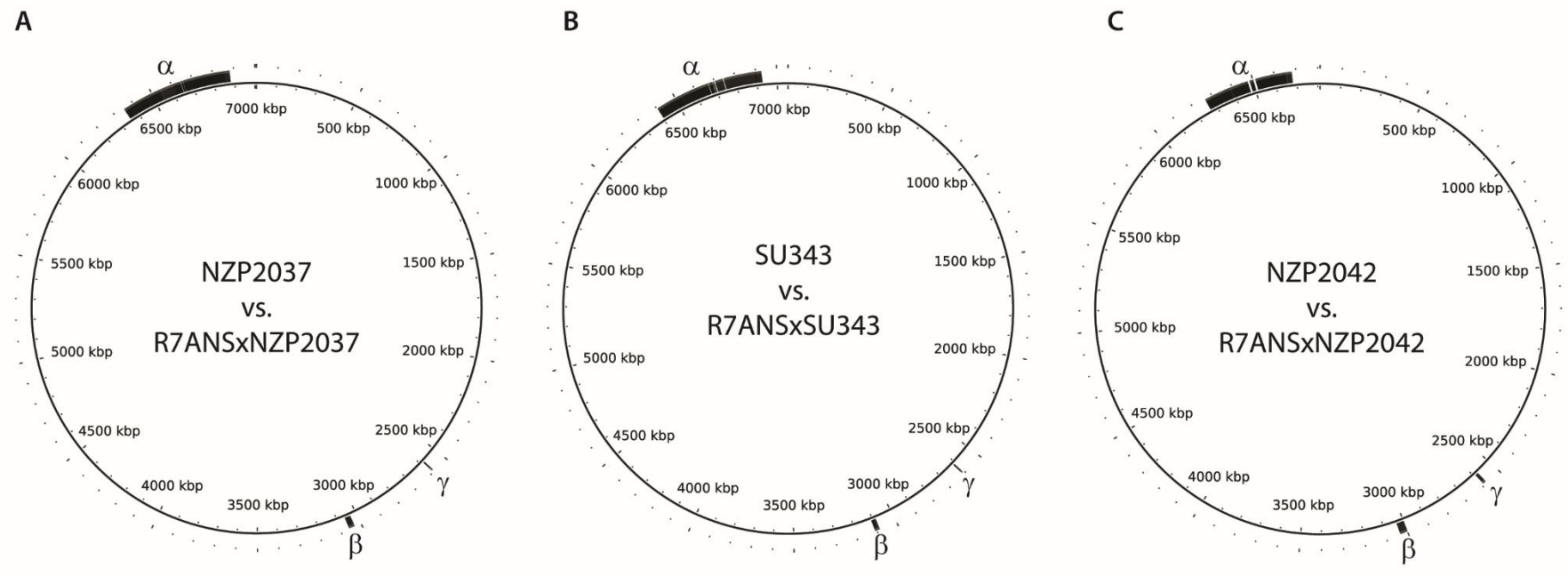


**Figure 4.1. Genome comparisons of ICE$^{3}$ donor and recipients.** BRIG (237) was used to carry out circular ungapped BLASTN alignments of the draft-sequenced exconjugants genomes with the ICE$^{3}$-carrying donor genomes. (A) A comparison of the draft R7ANSxNZP2037 sequence with the complete NZP2037 chromosome. (B) A comparison of the draft R7ANSxSU343 with a draft SU343 sequence scaffold. (C) A comparison of the draft R7ANSxNZP2042 sequence with a draft NZP2042 sequence scaffold. Black regions indicate >99% nucleotide identity. The α, β and γ regions are indicated for each genome comparison.

### *4.2.2. Conservation of symbiosis ICE and ICE$^{3}$-α genes indicates a common evolutionary history*

In the above experiments, 13 putative ICE$^{3}$s were identified and/or functionally confirmed to exist in 13 genetically diverse *Mesorhizobium* spp. isolated from various geographical locations. The genome sequences of these ICE$^{3}$-harbouring strains were compared to identify conserved ICE$^{3}$ genes and to gain insight into the evolutionary history of ICE$^{3}$. The α region of each of the four ICE$^{3}$ identitfied in strains for which complete genome sequences are available (NZP2037, WSM1271, WSM1497 and WSM1284) represents the largest portion (at least 90%) of ICE$^{3}$ DNA ranging from 468.3 kbp in WSM1497 to 563.8 kbp in WSM1284 (Table 4.1). Each ICE$^{3}$-α region carried several gene clusters common to the single-part symbiosis ICEs of *M. loti* R7A and MAFF303099 (130, 163, 178, 179). This included genes associated with nodulation (*nod*) and $N_2$ fixation (*nif*, *fix*) of legume hosts, a type-IV protein-secretion system and genes for biosynthesis of essential vitamins biotin and nicotinate, as well as thiamine for ICE*Ml*Sym$^{2037}$ (130, 187, 191). ICE$^{3}$-α regions also carried genes associated with ICE excision and transfer, including those encoding the conjugative type-IV secretion-system (*traG–trbB-trbI-msi021*), RdfS, RlxS (130) and homologues of QS transcriptional activators related to TraR and the AHL synthase, TraI1 (74, 167, 176) (A comparison of ICE*Mc*Sym$^{1271}$-encoded excision and transfer genes with ICE*Ml*Sym$^{R7A}$ is given in Table 4.2). Like the single-part symbiosis ICEs, ICE$^{3}$-α regions were also littered with transposons, insertion sequences and other MGEs that appear to have undergone significant diversification in genetic regions unrelated to symbiosis or ICE transfer (130, 178).

**Table 4.2. Comparison of selected ICE*Ml*Sym$^{R7A}$ and ICE*Mc*Sym$^{1271}$ genes**

| | R7A | | WSM1271 | | | | | |
|---|---|---|---|---|---|---|---|---|
| Gene name | Locus ID or coordinates | Length (aa) | Locus ID or coordinates | Length (aa) | Identity /coverage | Positives | Domain similarities/predicted function/comments | Associated reference |
| *traR* | MesloDRAFT_6173 | 241 | Mesci_5573 | 241 | 129/239 | 155/239 | Quorum-sensing activator of ICE*Ml*Sym$^{R7A}$ excision & transfer | (167) |
| *traR* | MesloDRAFT_6173 | 241 | Mesci_5676 | 245 | 152/241 | 170/241 | Quorum-sensing activator of ICE*Ml*Sym$^{R7A}$ excision & transfer | (167) |
| *tral1* | MesloDRAFT_6174 | 210 | Mesci_5572 | 211 | 127/206 | 150/206 | N-acyl-L-homoserine lactone synthase | (167) |
| *msi172* | 6289921-6290151 | 77 | 5862428-5862195 | 78 | 37/76 | 49/76 | Encodes N-terminal portion of FseA | (170) |
| *msi171* | 6290150-6290719 | 190 | 5862175-5861624 | 191 | 103/191 | 129/191 | Encodes C-terminal portion of FseA | (170) |
| *qseM* | MesloDRAFT_6177 | 97 | Mesci_5675 | 88 | 62/77 | 69/77 | QseM, dual-target antiactivator of FseA and TraR | (176) |
| *qseC* | MesloDRAFT_6178 | 70 | Mesci_5674 | 72 | 22/68 | 42/68 | DNA-binding activator of *qseC* expression and repressor of *qseM* expression | (176) |
| *msi110* | MesloDRAFT_6247 | 107 | Mesci_5531 | 107 | 97/107 | 105/107 | DUF736, conserved ICE protein | (130) |
| *traF* | MesloDRAFT_6249 | 182 | Mesci_5529 | 182 | 128/182 | 139/182 | TrbC protease TraF | (130) |
| *msi107* | MesloDRAFT_6250 | 227 | Mesci_5528 | 227 | 167/224 | 186/224 | Murein transglycosylase, conserved ICE protein | (130) |
| *rdfS* | MesloDRAFT_6248 | 89 | Mesci_5530 | 79 | 63/67 | 66/67 | RdfS, IntS-associated recombination directionality factor (excisionase) ICE*Ml*Sym$^{R7A}$ | (74) |
| *rlxS* | MesloDRAFT_6251 | 656 | Mesci_5527 | 656 | 526/656 | 591/656 | RlxS relaxase | (74) |
| *virB1* | MesloDRAFT_6305 | 260 | Mesci_5600 | 257 | 229/260 | 238/260 | Protein secretion, involved in symbiosis | (187) |
| *virB2* | MesloDRAFT_6306 | 121 | Mesci_5599 | 121 | 112/121 | 114/121 | Protein secretion, involved in symbiosis | (187) |
| *virB3* | MesloDRAFT_6307 | 108 | Mesci_5598 | 108 | 100/108 | 104/108 | Protein secretion, involved in symbiosis | (187) |
| *ardC* | MesloDRAFT_6200 | 320 | Mesci_5534 | 320 | 253/296 | 272/296 | COG4227 - ArdC antirestriction protein, conserved ICE protein | (130) |

| | R7A | | WSM1271 | | | | | |
|---|---|---|---|---|---|---|---|---|
| Gene name | Locus ID or coordinates | Length (aa) | Locus ID or coordinates | Length (aa) | Identity /coverage | Positives | Domain similarities/predicted function/comments | Associated reference |
| *msi151* | MesloDRAFT_6201 | 139 | Mesci_5533 | 136 | 102/136 | 110/136 | DUF2958 conserved ICE protein | (130) |
| *msi150* | MesloDRAFT_6202 | 576 | Mesci_5532 | 575 | 453/573 | 501/573 | ParB homologue, conserved ICE protein | (130) |
| *virB4* | MesloDRAFT_6308 | 789 | Mesci_5597 | 789 | 754/789 | 773/789 | Type IV Protein secretion system, involved in symbiosis | (246) |
| *virB5* | MesloDRAFT_6309 | 224 | Mesci_5596 | 227 | 170/223 | 185/223 | Type IV Protein secretion system, involved in symbiosis | (187) |
| *virB6* | MesloDRAFT_6310 | 295 | Mesci_5595 | 295 | 240/295 | 261/295 | Type IV Protein secretion system, involved in symbiosis | (187) |
| *virB7* | MesloDRAFT_6311 | 52 | Mesci_5594 | 52 | 45/52 | 49/52 | Type IV Protein secretion system, involved in symbiosis | (187) |
| *virB8* | MesloDRAFT_6312 | 237 | Mesci_5593 | 237 | 219/237 | 228/237 | Type IV Protein secretion system, involved in symbiosis | (187) |
| *virB9* | MesloDRAFT_6313 | 293 | Mesci_5592 | 293 | 265/293 | 280/293 | Type IV Protein secretion system, involved in symbiosis | (187) |
| *virB10* | MesloDRAFT_6314 | 379 | Mesci_5591 | 380 | 325/380 | 346/380 | Type IV Protein secretion system, involved in symbiosis | (187) |
| *virB11* | MesloDRAFT_6315 | 346 | Mesci_5590 | 346 | 328/346 | 337/346 | Type IV Protein secretion system, involved in symbiosis | (187) |
| *virA* | MesloDRAFT_6316 | 846 | Mesci_5589 | 814 | 637/846 | 697/846 | Type IV Protein secretion system, involved in symbiosis | (187) |
| *traG* | MesloDRAFT_6334 | 676 | Mesci_5524 | 676 | 589/676 | 626/676 | Type IV conjugation | (130) |
| *msi031* | MesloDRAFT_6335 | 137 | Mesci_5523 | 137 | 114/129 | 121/129 | Type IV conjugation | (130) |
| *trbB* | MesloDRAFT_6336 | 322 | Mesci_5522 | 322 | 284/311 | 302/311 | Type IV conjugation | (130) |
| *trbC* | MesloDRAFT_6337 | 102 | Mesci_5521 | 102 | 68/77 | 71/77 | Type IV conjugation | (130) |
| *trbD* | MesloDRAFT_6338 | 90 | Mesci_5520 | 90 | 79/90 | 83/90 | Type IV conjugation | (130) |
| *trbE* | MesloDRAFT_6339 | 816 | Mesci_5519 | 813 | 723/813 | 758/813 | Type IV conjugation | (130) |
| *trbF* | MesloDRAFT_6340 | 241 | Mesci_5516 | 243 | 217/237 | 226/237 | Type IV conjugation | (130) |

| | R7A | | WSM1271 | | | | | |
|---|---|---|---|---|---|---|---|---|
| Gene name | Locus ID or coordinates | Length (aa) | Locus ID or coordinates | Length (aa) | Identity /coverage | Positives | Domain similarities/predicted function/comments | Associated reference |
| *trbG* | MesloDRAFT_6341 | 342 | Mesci_5515 | 342 | 297/342 | 318/342 | Type IV conjugation | (130) |
| *trbI* | MesloDRAFT_6342 | 404 | Mesci_5514 | 406 | 348/402 | 373/402 | Type IV conjugation | (130) |
| *trbJ* | MesloDRAFT_6343 | 241 | Mesci_5518 | 241 | 227/241 | 235/241 | Type IV conjugation | (130) |
| *trbL* | MesloDRAFT_6344 | 440 | Mesci_5517 | 440 | 387/440 | 411/440 | Type IV conjugation | (130) |
| *msi021* | MesloDRAFT_6345 | 90 | Mesci_5513 | 90 | 71/90 | 80/90 | Type IV conjugation | (130) |
| *intS* | MesloDRAFT_6365 | 414 | Mesci_2349 | 421 | 366/399 | 379/399 | ICE*Ml*Sym$^{R7A}$ integrase IntS | (60) |
| **Additional genes identified in this study** | | | | | | | | |
| *intG* | | - | Mesci_2549 | 417 | | | *guaA* associated integrase IntG | This study |
| *intM* | - | - | Mesci_2344 | 396 | | | *met-tRNA* gene associated integrase IntM | This study |
| *rdfG* | - | - | Mesci_2550 | 84 | | | putative IntG- associated recombination directionality factor (excisionase) RdfG | This study |
| *rdfM* | | | Mesci_2345 | 83 | | | putative IntM-associated recombination directionality factor (excisionase) RdfM | This study |

The conservation of genes between symbiosis ICE and ICE[3] suggests that these elements share a common evolutionary history. Considering the increased complexity of ICE[3] over single-part ICEs, it seems plausible that ICE[3] may have evolved in a bacterium carrying an ancestral single-part symbiosis ICE and two other integrative elements integrated within the *attB*$_S$, *attB*$_G$ and *attB*$_M$ sites. In such a strain, a genomic inversion between an IntS-associated element and an IntM-associated element, followed by an inversion between a resulting hybrid element and an IntG-associated element, would produce an ICE[3] resembling the structure of ICE*Mc*Sym[1271] (Fig. 4.2). Such inversions could have easily been mediated by one of the numerous transposable elements found on mesorhizobial ICEs (130), either as part of the transposition process or through RecA-mediated recombination between repetitive elements.

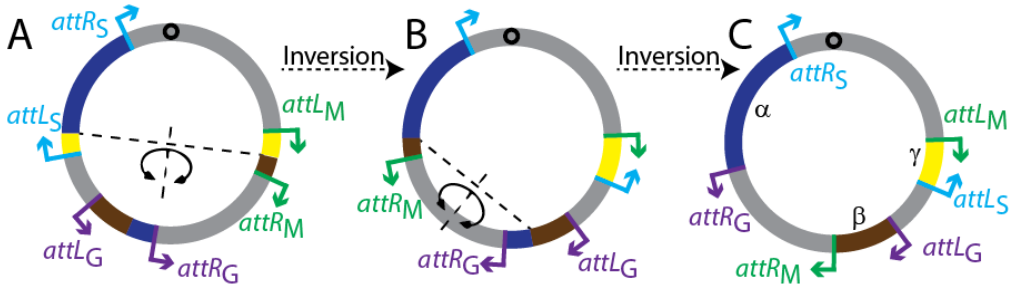


**Fig. 4.2. Model of tripartite ICE evolution.** The arrangement of *att* sites on ICE[3] may have evolved through two chromosomal inversions between three separate elements flanked by distinct *attL* and *attR* sites. Colours correspond to Fig. 3.3. The dashed lines segmenting the chromosome indicate where the inversions may have occurred. (A) The ancestral chromosome configuration. (B) Configuration of the chromosome following the first inversion. (C) The final tripartite ICE structure following the second inversion

### *4.2.3. ICE³ regions γ and β carry disparate genetic cargo*

In addition to the four completely sequenced ICE$^{3}$-harbouring mesorhizobia, sequences of β and γ regions are also available for four additional isolates; *M. loti* strains SU343, NZP2042 and WSM1293 (247) and *M. ciceri* Ca181 (NZ_CM002796). For all ICE$^{3}$s except ICE*Mc*Sym$^{Ca181}$, the β region was larger than γ (Fig 4.3). All ICE$^{3}$-β regions carry the *intG* recombinase gene and RDF gene *rdfG*, located directly downstream of *attL$_G$.* The DNA sequences of the *att* sites and recombination genes were highly conserved, however, disparate genetic cargo was present on the remainder of each β region. For example, ICE*Mc*Sym$^{1497}$ region β carried a large operon encoding ABC-type transporter proteins, ICE*Mc*Sym$^{1271}$-β carried the melanin biosynthesis gene cluster and the ICE*Mc*Sym$^{1293}$-β region carried genes encoding enzymes of the pentose phosphate pathway (Fig 4.3). ICE*Ml*Sym$^{NZP2073}$ and ICE*Ml*Sym$^{SU343}$ shared near-identical β regions and both were closely related to the β region of ICE*Ml*Sym$^{NZP2042}$, however, comparative analysis suggested insertion, deletion and inversion events had occurred in the ICE*Ml*Sym$^{NZP2042}$-β region. The ICE*Ml*Sym$^{NZP2073}$, ICE*Ml*Sym$^{SU343}$ and ICE*Ml*Sym$^{NZP2042}$-β regions each also carried radical SAM (*S*-adenosyl-*L*-methionine)-superfamily genes *hsxABC* similar to those required by NifB for the successful assembly of the nitrogenase molybdenum cofactor (248). The *hsxABC* operon was inverted in ICE*Ml*Sym$^{2042}$-β and present on region α in ICE*Mc*Sym$^{1284}$. ICE*Ml*Sym$^{NZP2073}$, ICE*Ml*Sym$^{SU343}$ and ICE*Ml*Sym$^{NZP2042}$-β regions also carried homologues of the *A. tumefaciens iaaM* gene, which encodes a tryptophan monooxygenase required for synthesis of indole-3-acetamide (IAA), a precursor to the plant auxin hormone indole-3-acetic acid (249-251). *iaaM* was also found on the α

region of the ICE*Mc*Sym[1293]. If IaaM is contributing to IAA production (252, 253), then its presence on these ICE[3]s may indicate a role in symbiosis. Lastly, the ICE[3]-β regions of ICE*Ml*Sym[NZP2073], ICE*Ml*Sym[SU343], ICE*Ml*Sym[NZP2042] and ICE*Mc*Sym[1284] carried genes encoding homologues of the RecD exonuclease protein. RecD is the alpha subunit of the exonuclease V complex, involved in homologous recombination and plasmid maintenance in *E. coli* (254-256). Like the ICE[3]-α regions, many of the ICE[3]-β regions carried recombinase and transposase genes or gene-fragments/pseudogenes, suggesting these regions have been subject to invasion and recombination events mediated by foreign MGEs. In summary, the genetic cargo carried by regions β and γ of genetically diverse *Mesorhizobium* spp. appears to be highly disparate (Fig 4.3).

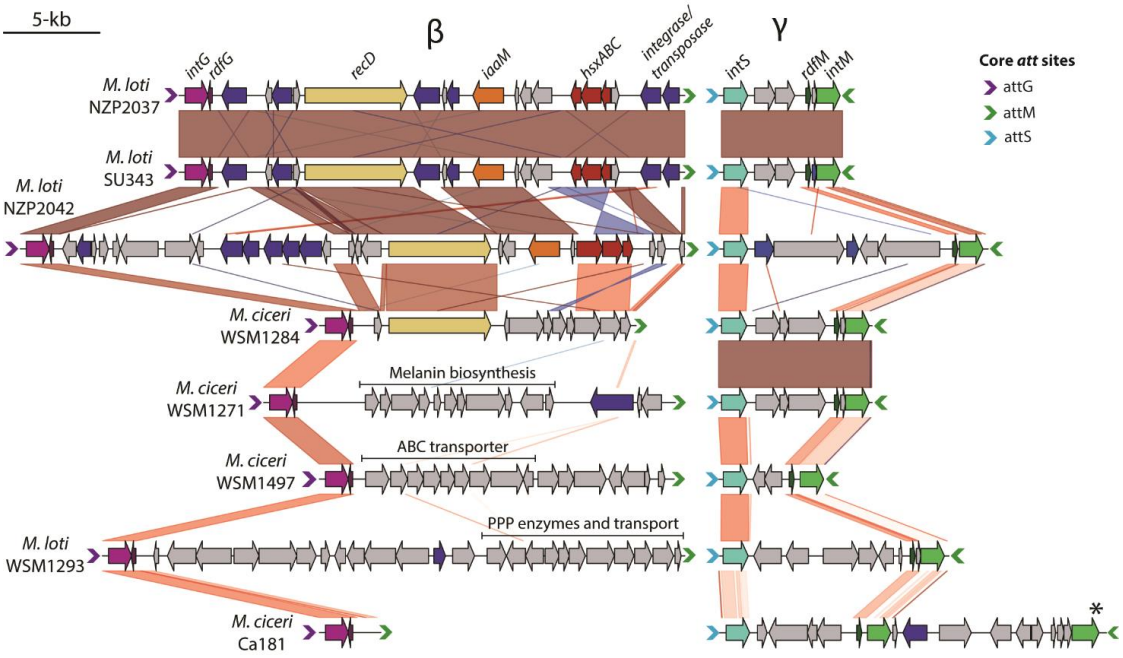


**Figure 4.3. GenoplotR (240) gene map and synteny alignment of ICE$^{3}$ regions beta and gamma.** Gene maps for each of the assembled ICE$^{3}$s β and γ sequences are shown and annotated where possible. Coloured arrows indicate homologues that are also found on other regions, where they are coloured similarly. Grey arrows are unique to each element and may encode hypothetical proteins, or conserved proteins of unknown function unless otherwise specified. Coloured blocks or lines linking each gene represent BLASTN hits (15-bp window) in the same orientation (red) or in an inverted orientation (blue), with increased colour intensity indicating increased similarity. The NZP2042-β was reconstructed using the genome sequences for both wild-type and the ICE$^{3}$ exconjugant (R7ANSxNZP2042) (4). *ICE*Mc*Sym$^{Ca181}$-γ carries an *intM* pseudogene carrying four stop codons, central to the region, and encodes a serine recombinase (*intM*$_{ser}$) adjacent to the site of integration at a distinct *met*-tRNA (M1C_RS0523)

Similar to the β region, the γ region *attL*$_{S}$*-intS* and *rdfM-intM-attL*$_{M}$ sequences were highly conserved, but the genetic cargo between them varied considerably. Most genes located in ICE$^{3}$-γ regions appeared to encode products with unknown function, these regions also contained various transposase sequences or potential remnants of “crash-landed” MGEs, so it is difficult to predict if any of the γ region genes, or regions apart from the *att* sites and recombination genes, have a role in symbiosis or ICE$^{3}$ transfer.

### *4.2.4. Evolution of a new ICE3-γ integration site through evolutionary recruitment of a serine recombinase*

Because the ICE$^3$ excision process naturally requires the assembly of the three ICE$^3$ regions into a single contiguous region of DNA, there is evolutionary potential for ICE$^3$ to revert into a single-part ICE. Why then do these elements persist in nature? One possibility is that the cotransfer of three regions is more beneficial than transfer of any single ICE$^3$ region. Implicit in this reasoning is that genes contributed by each distinct region are of significance to the long-term persistence of ICE$^3$, however, the dispartity of genetic cargo carried by diverse ICE$^3$ β and γ regions suggests that this is not the case. Rather, ICE$^3$ may persist because the tripartite configuration itself provides some selective advantages. The putative ICE$^3$ of *M. ciceri* Ca181 exhibited distinct features that strongly supported this notion. The genome of *M. ciceri* Ca181 was found to carry a putative ICE$^3$ including α, β and γ regions and the expected recombinase genes *intS*, *intG* and *intM* (Fig 4.4). However, the *intM* gene (M1C_RS32875) contained several stop codons, suggesting *intM* has become a pseudogene (Fig 4.4). Adjacent to the *intM* pseudogene was a sequence matching the IntM-associated *att* core sequence not associated with the canonical ICE$^3$ *met*-tRNA integration site (M1C_RS10995 in Ca181). A serine recombinase gene *intM*$_{ser}$ (M1C_RS05230) was identified ~9-kb downstream of the *intM* pseudogene positioned adjacent to a distinct *met*-tRNA gene (M1C_RS05235) (not the normal ICE$^3$ integration site). Homologues of *intM*$_{ser}$ were also identified encoded adjacent to *met*-tRNA genes in several other organisms including Gdia_1616 in *Gluconacetobacter diazotrophicus* PAl 5 and BUE85_RS01630 in *Ochrobactrum pituitosum* strain AA2. As expected,

the *intG* gene was located adjacent to the *guaA* gene on the ICE*Mc*Sym$^{Ca181}$-β region. Although the IntM *att* core site was not present near this region, an exact 14-bp duplicate of the 3' end of the *intM$_{ser}$*-associated *met*-tRNA gene sequence 5'-TGGTTGCGGGGACA-3' was identified directly downstream of *intG* and *rdfG*. Thus, it appears that for ICE*Mc*Sym$^{Ca181}$, the IntM recombinase has been replaced by IntM$_{ser}$ and ICE*Mc*Sym$^{Ca181}$ has adopted the new IntM$_{ser}$-associated *met*-tRNA *attB* site for integration. Remarkably, the replacement of *intM* with *intM$_{ser}$* in Ca181 has preserved the arrangement and orientation of each of the attachment sites such that recombination of the three ICE$^3$ regions would still be expected to form a single contiguous element for transfer (excluding the *attR$_G$* site which remains unidentified). Therefore, the putative ICE$^3$ of Ca181 appears to be an instance of a potentially recent replacement event of an *attB* site and its cognate recombinase with a new recombinase and *attB* site. This clearly indicates that the tripartite structure of this element has been maintainted even when one of the recombinases and regions has been completely replaced.

.

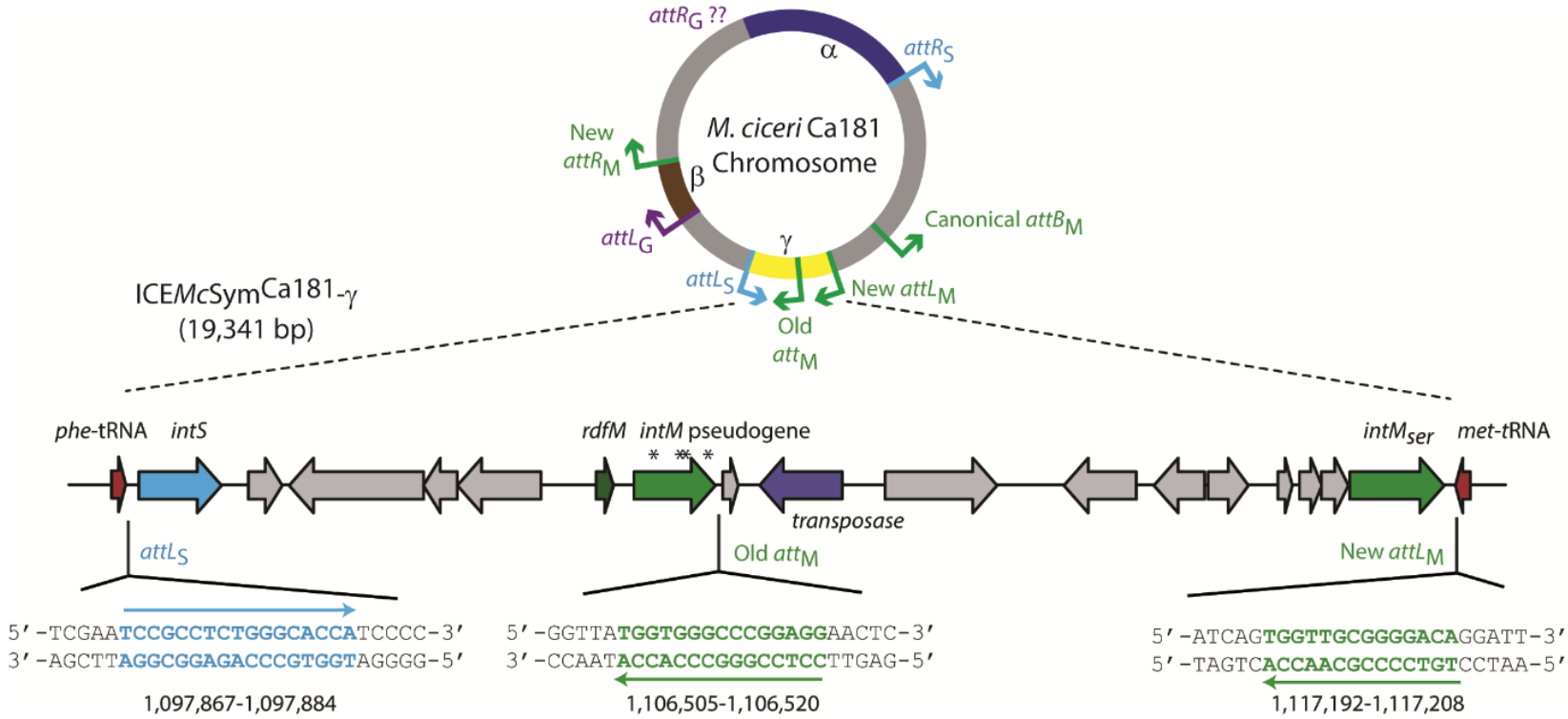


**Figure 4.4.** ICE*Mc*Sym$^{Ca181}$ region γ recruitment of a new *met*-tRNA integration site and serine-recombinase gene. ICE*Mc*Sym$^{Ca181}$-γ is integrated in the *M. ciceri* Ca181 genome between convergently orientated *phe*-tRNA and *met*-tRNA genes. However, the *met*-tRNA gene is distinct to that in which the other identified ICE$^{3}$-γ regions are integrated. ICE*Mc*Sym$^{Ca181}$-γ also carries a new serine recombinase (*intM*$_{ser}$) adjacent to the *met*-tRNA gene. Therefore, ICE*Mc*Sym$^{Ca181}$ likely utilises a new 14 bp IntM$_{ser}$ core sequence present within a distinct *met*-tRNA gene (M1C_RS05235), forming the *attL*$_M$ sequence, and at the end of the β region, forming *attR*$_M$. An *intM* pseudogene containing several stop codons (*) is located in the centre of the γ region.

### *4.2.5. Proposed selective advantages of the ICE³ configuration*

The observed preservation of ICE$^3$ structure on ICE*Mc*Sym[181] indicates that there may be selectable benefits associated with the ICE$^3$ structure. Four proposed scenarios that might explain the evolutionary success of ICE$^3$ are discussed below.

a) Increased host range afforded by multiple attachment sites

ICEs site-specifically integrate into bacterial chromosomes at *attB* sites, usually within conserved genes such as tRNAs (86, 257), *guaA* (87) and *prfC* (258). There is selective pressure to maintain this specificity for integration, because non-specific ICE integration results in reduced viability and frequency of ICE transfer in recipients (259). Given the enormous diversity and abundance of ICEs in bacteria (116), competition for available *attB* sites in bacterial chromosomes would be expected. The configuration of ICE$^3$ may therefore be advantageous because the ability to integrate at three distinct *attB* sites maximizes the potential for chromosomal integration, even if one or more of the cognate *attB* sites are not present or are not perfectly conserved in the recipient (Fig 4.5A).

b) Passive stabilization

Toxin anti-toxin modules such as *mosAT, tad-ata* and *hipAB* enhance the stability of spuriously excised SXT-family ICEs by post-segregational killing or growth arrest following loss of the element (150, 158, 260). However, in the absence of such active stabilization modules, spurious excision of ICEs from their host chromosomes may lead to their loss (145). Therefore, any mechanisms that can reduce or prevent spurious excision likely stabilize ICEs

in the long term. Work on the regulation of transfer of ICE*Ml*Sym$^{R7A}$ has demonstrated that numerous layers of transcriptional, translational and post-translational regulation are present that likely prevent spurious excision of ICE*Ml*Sym$^{R7A}$ from the *M. loti* chromosome (75). Although there may be unidentified TA modules or other genes facilitating the active stabilization of symbiosis ICEs and ICE$^{3}$s, it is plausible that the tripartite configuration of ICE$^{3}$ is intrinsically resistant to spurious excision and loss because it requires three separate recombination events for excision (Fig 4.5B).

c) Genome stability and competitiveness in an ICE/IME-rich environment

It has been observed that elements with the same integration site and similar recombinases may integrate in tandem at a single site (132-135, 137-139). Tandem arrays of integrative elements are formed when one or more invading ICE(s) integrate site-specifically at the *attL* or *attR* site of a resident ICE/IME occupying its cognate *attB* site in the bacterial chromosome (Fig 4.5B). The result is a composite ICE carrying distal *attL* and *attR* sites derived from the most outer elements, and one or more central hybrid *attP*-like site(s) derived from *attL* and *attR* of adjacent elements. The *Streptomyces scabiei* ICE resembles a tandem ICE/IME array comprised of two “toxigenic regions” (TR1 and TR2) which can each excise independently through recombination of a distal *attL* or *attR* site with the central *attP*-like site*,* or which can excise as a composite element through recombination of the distal *attL* and *attR* sites (261). Arrays of SXT and R391-family ICEs (138), and ICE-CIME (*cis*-integrative and mobilizable element) arrays of ICE*St3* and CIME*L$_3$catR$_3$* (133) also excise as individual units or composite elements in the same manner. However, tandem arrays can be highly unstable, even in RecA$^{-}$ backgrounds

(133, 137, 138, 140). This is probably because most tyrosine recombinases exhibit a strong directional preference for integrative recombination, i.e. *attP* + *attB* → *attL* + *attR* (107), and thus formation of the central *attP*-like site in tandem ICE arrays facilitates excision and loss of one or more adjacent elements in the array carrying a DR *attL* or *attR* site (133, 138). However, the instability caused by tandem ICE/IME insertion may not affect ICE$^3$ elements in this manner, since no individual region of an ICE$^3$ carries any directly repeated *att* sites. Tandem insertion of an invading ICE into *attL* or *attR* sites of any ICE$^3$ region would result in the formation of a composite element where only the invading ICE/IME carries a direct repeat of compatible *attL* and *attR*. Thus, an ICE$^3$ is likely stable following a tandem insertion event by an invading element because none of its regions can be excised by a single recombination event (Fig 4.5B). Moreover, following transfer of the ICE$^3$ into a strain occupied by a resident ICE/IME, ICE$^3$ integration within the *attL* or *attR* sites of the resident element should not affect the ICE$^3$ integration process and tripartite separation, but may induce instability in the resident element. Therefore, ICE$^3$s may be competitively superior in their ability to occupy *attB* sites and usurp resident integrated elements in an ICE/IME-rich environment.

d) Increased opportunity for gene capture

Although tandem arrays of SXT and R391-family ICEs are highly unstable and have never been found in natural isolates, tandem arrays of these ICEs generated in the laboratory recombine to facilitate the evolution of hybrid elements (137, 141) (Fig 4.5C). Rearrangement of genes in tandem arrays has also been observed for ICE*St3* and CIME*L$_3$catR$_3$* (133). Therefore, ICE *attL* and *attR* sites represent hotpots for the capture and generation of novel

MGEs through tandem integration and accretion of associated genes. It seems possible that ICE$^{3}$ structure could be advantageous because stably occupying three *attB* sites might provide ICE$^{3}$ with an increased opportunity to capture and stockpile genes from invading ICEs/IMEs, enabling more rapid evolution and adaptation of the ICE$^{3}$ gene content (Fig 4.5C). Accretion of an invading ICE/IME carrying *IntM*$_{ser}$ likely explains the evolution of the distinct γ region of ICE*Mc*Sym$^{Ca181}$ (Fig 4.4).

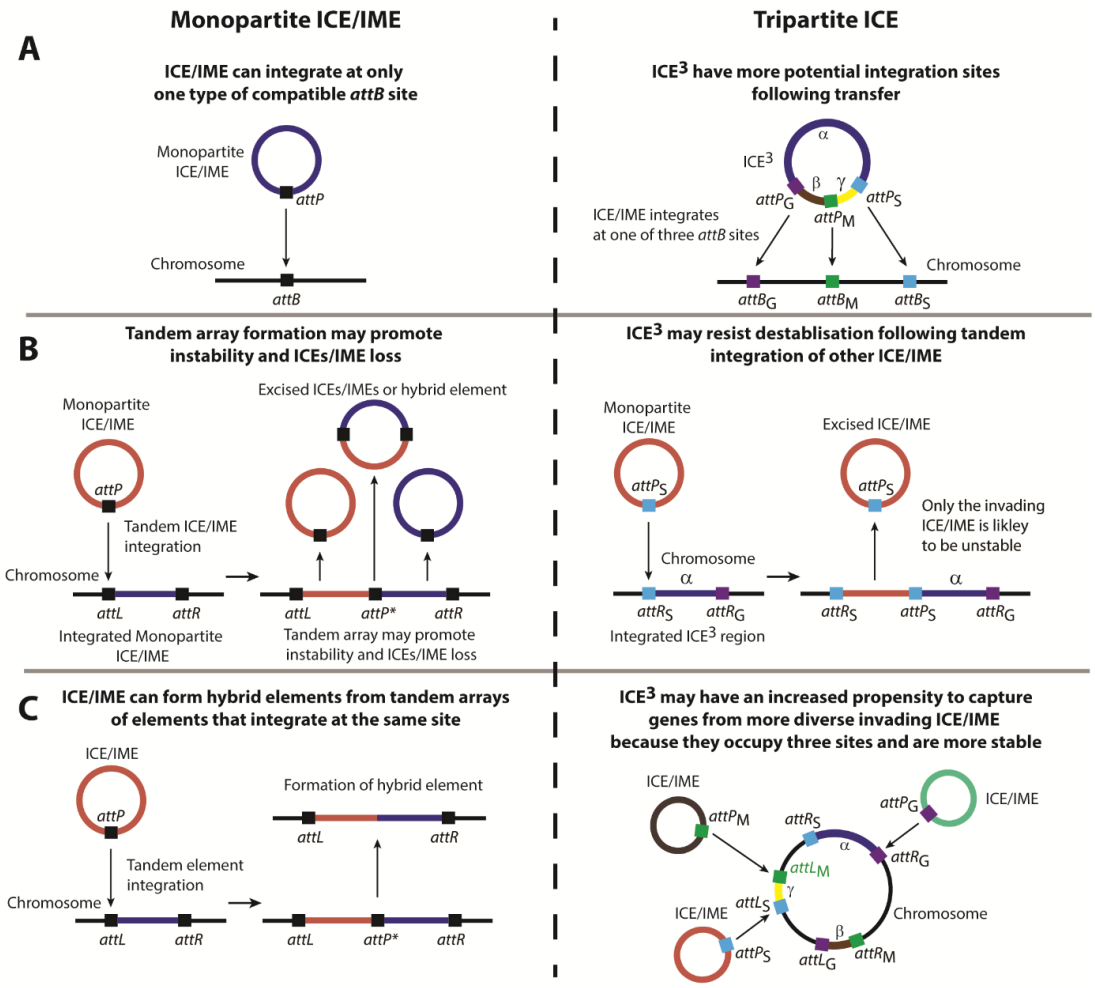


**Figure 4.5. Evolutionary advantages of ICE$^3$.** (A) Carrying three *attP* sites in the circular form enables the ICE$^3$ the option of integrating at 3 different *attB* sites, increasing its chances of becoming stably integrated within the chromosome following transfer. (B) Tandem ICE/IME arrays may in some cases produce an unstable arrangement in which one or both ICE/IME are lost. Because no region of the ICE$^3$ carries DR *att* sites and no single recombination event causes excision of the ICE$^3$, insertion of an ICE/IME at any of the sites already occupied by the ICE$^3$ cannot destabilise the ICE$^3$. Moreover, when the ICE$^3$ integrates into the *attL* or *attR* site of a resident element and disassembles into it tripartite form, only this resident element carries a DR of the newly formed *attP* site on the now composite ICE$^3$ region. Thus, recombination of these sites could drive the excision and potentially loss of the resident element from its cognate *attB* site, while the ICE$^3$ remains stably integrated. ICE$^3$s may therefore be competitively superior in their occupation of *attB* sites. (C) By stably occupying three *attB* sites in the bacterial chromosome, the ICE$^3$ has increased opportunity to become associated with other ICE/IME elements at the same sites via tandem integration, potentially facilitating increased propensity for gene capture through acquisition of genes from adjacent elements.

## 4.3. Summary

In this Chapter, ICE$^3$s were identified in 13 genetically diverse *Mesorhizobium* spp. originally isolated from various geographical locations. Analysis of the ICE$^3$-α regions revealed that ICE$^3$ share several clusters of genes common to single-part symbiosis ICEs of *M. loti* R7A and MAFF303099, suggesting a common evolutionary history for these elements. It was proposed that symbiosis ICE$^3$s probably evolved following recombination between a single-part symbiosis ICE and two other integrative elements in an ancestral bacterium. As to why ICE$^3$s have persisted in nature is a more complex question to answer. Here, it was proposed that the tripartite structure itself may provide four selective benefits for ICE$^3$;

a) Increased host range afforded by multiple attachment sites
b) Passive stabilization
c) Genome stability and competitiveness in an ICE/IME-rich environment
d) Increased opportunity for gene capture

Although the four proposed advantages for the ICE$^3$ configuration are yet to be experimentally tested, it seems likely that such benefits may be important in environments where integrative elements are abundant and there may be competition for commodities such as available *attB* integration sites.

In the next Chapter, the information uncovered regarding the genetic content, evolutionary history, and mechanism of recombination for ICE$^3$ are integrated to explore the regulatory control of ICE$^3$ excision and transfer.

# Chapter 5.

## Regulation of ICE*Mc*Sym$^{1271}$ assembly, excision and transfer

## 5.1. Introduction

The direction of recombination catalysed by an integrase is often determined by a recombination directionality factor, also known as an excisionase (106, 107, 262). Excisionases are non-catalytic DNA-binding proteins that promote formation of *attP* and *attB*. In R7A, ICE*Ml*Sym$^{R7A}$ excision is stimulated by RdfS and in its absence, IntS activity favours formation of *attL* and *attR* (74). Expression of IntS, IntG and IntM stimulated recombination of the pMINI3 *attP* sites with each cognate *attB* site in R7ANS, producing *attL* and *attR*, suggesting equilibrium reactions favour *attL* and *attR* production in the absence of other ICE*Mc*Sym$^{1271}$ genes for all three ICE*Mc*Sym$^{1271}$ integrases. ICE*Mc*Sym$^{1271}$ encodes a homologue of *rdfS* and two other putative AlpA-family excisionase (263) genes located adjacent to *intG* and *intM*, termed *rdfG* and *rdfM,* respectively (Table 3.1). It seems likely that expression of the *rdfS, rdfG* and *rdfM* genes is coregulated to promote excision of ICE*Mc*Sym$^{1271}$.

For the tripartite ICE*Mc*Sym$^{1271}$, the increased complexity introduced by the three separate recombination reactions required for ICE*Mc*Sym$^{1271}$ integration and excision leads to the potential formation of eight distinct chromosomal recombination states (Figure 5.1A). The arrival at any particular state depends on the prior order and direction of the Int-mediated recombination reactions. In the synthetic "mini-ICE*Mc*Sym$^{1271}$" experiments presented in Chapter 3, not all eight states were reconstructed, suggesting some states are non-viable. Specifically, the model for ICE*Mc*Sym$^{1271}$ excision (Fig 5.1A) indicates that if the first excisive reaction is catalysed by IntM, i.e. $attL_M + attR_M > attP_M + attB_M$, then the chromosome is split into two parts, one part lacking the likely essential *phe* and *his*-tRNA genes and the other part an origin-of-replication (ICE$^3$

reactions producing *attP* and *attB* do not necessarily result in excision *per se* for ICE[3], but for simplicity recombinations producing *attP* and *attB* will be referred as 'excisive'). qPCR assays measuring IntM-mediated formation of $attP_M$ and $attB_M$ indicate the excisive IntM reaction occurs at the lowest frequency of the three integrase-mediated reactions, suggesting evolved regulatory control mechanisms might prevent IntM-mediated excisive recombination occurring before other reactions, precluding formation of the non-viable chromosome state.

In this chapter, the role of the three predicted RDFs RdfS, RdfG and RdfM during ICE*Mc*Sym[1271] assembly and excision are explored through a combination of mutation analysis qPCR, RNAseq and reporter assays, with the aim of producing a robust model for the coordination and regulation of ICE*Mc*Sym[1271] assembly and excision (Figure 5.1).

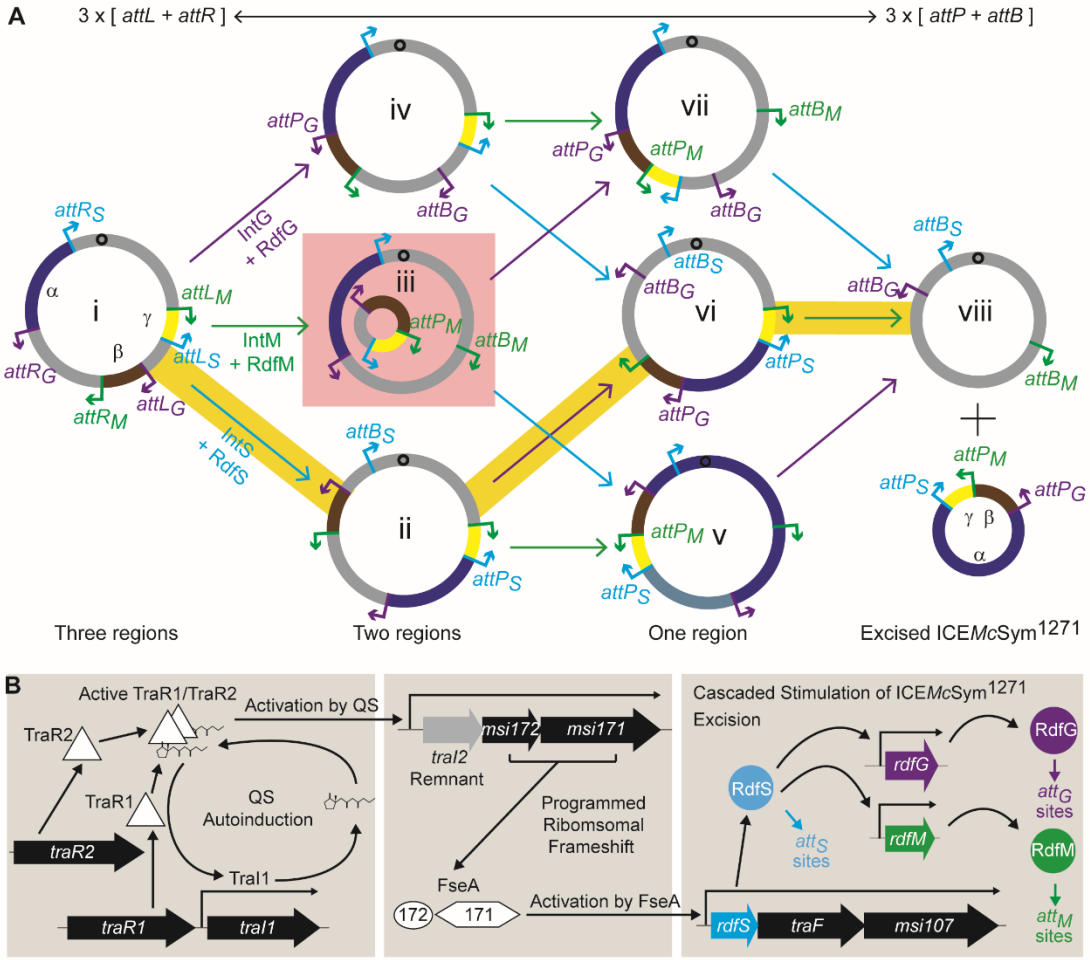


**Fig 5.1. ICE*Mc*Sym$^{1271}$ assembly, excision and regulation.** (A) Schematic representing excisive chromosomal recombinations leading to excised ICE*Mc*Sym$^{1271}$ or formation of a non-viable state (pink box). RdfS stimulates the IntS-mediated excisive reaction *attL*$_S$ and *attR*$_S$ > *attP*$_S$ and *attB*$_S$ delineated by cyan arrows; RdfG stimulates the IntG-mediated excisive reaction *attL*$_G$ and *attR*$_G$ > *attP*$_G$ and *attB*$_G$ delineated by purple arrows; RdfM stimulates the IntM-mediated excisive reaction *attL*$_M$ and *attR*$_M$ > *attP*$_G$ and *attB*$_M$ delineated by green arrows. The combined data in this chapter support the hypothesis that the excisive reactions likely occur in the order IntS > IntG > IntM (highlighted in yellow) during ICE*Mc*Sym$^{1271}$ excision. (B) Following the model of QS-mediated induction of excision for ICE*Ml*Sym$^{R7A}$ and data presented here, TraR1 and TraR2 bind AHLs produced by TraI1 and TraR1/R2-AHL complexes activate transcription from the *traI1* and *traI2* promoters. This results in autoinduction of *traI1* and activation of *traI2-msi172-msi171* expression. The programmed ribosomal frameshift site within the 3' end of *msi172* fuses the translational reading frames of *msi172-msi171* producing FseA. FseA then activates transcription of the *rdfS* operon. RdfS stimulates excisive IntS-mediated recombination and promotes expression of RdfG and RdfM which subsequently stimulates the excisive IntG and IntM-mediated recombination reactions, respectively.

## 5.2. Results

### *5.2.1. RdfG and RdfM are required for excisive IntG and IntM-mediated recombination*

The tripartite ICE*Mc*Sym[1271] encodes three predicted excisionase genes *rdfS* (Mesci_5530), *rdfG* (Mesci_2550) and *rdfM* (Mesci_2345). *rdfG* is oriented convergently with *intG* on ICE*Mc*Sym[1271] region β and *rdfM* is encoded directly upstream of *intM* on ICE*Mc*Sym[1271] region γ. Like RdfS, RdfG and RdfM are MerR superfamily proteins with a predicted winged helix-turn-helix secondary structure (Fig 5.2). Each of these genes was replaced with an Ω*aadA* cassette producing strains 1271Δ*rdfG*::Ω*aadA* and 1271Δ*rdfM*::Ω*aadA,* respectively. qPCR assays were performed with each strain to assess the affects of these insertions on ICE[3] assembly/excision. In wild-type WSM1271, $attP_G$ and $attB_G$ and $attP_S$ and $attB_S$ sites were detected at a frequency of 0.1-1% per chromosome and $attP_M$ and $attB_M$ sites were detected at 0.01-0.1% (Fig 5.3A). In contrast, $attP_G$ and $attB_G$ sites were undetectable in 1271Δ*rdfG*::Ω*aadA* and $attP_M$ and $attB_M$ sites were undetectable in 1271Δ*rdfM*::Ω*aadA*. The abundance of the two remaining *attP* and *attB* sites in each of these mutant strains was similar to that of WSM1271. Complementation of 1271Δ*rdfG*::Ω*aadA* with a cloned copy of *rdfG* and its native promoter partially restored $attP_G$ and $attB_G$ formation and complementation of 1271Δ*rdfM*::Ω*aadA* with a cloned copy of *rdfM* and its native promoter restored $attP_M$ and $attB_M$ production. These experiments therefore confirmed the roles of RdfG and RdfM in excisive IntG and IntM reactions, respectively.

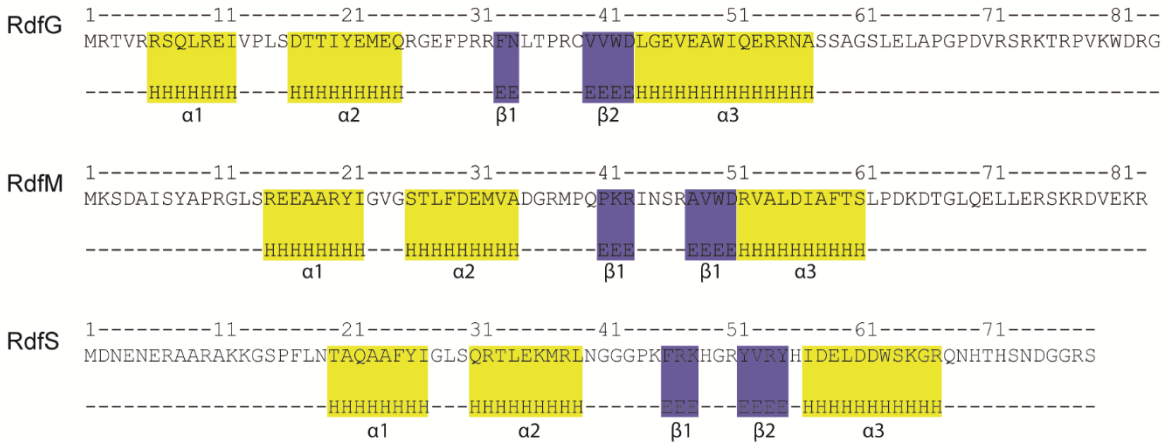


**Fig 5.2. Predicted secondary structures of RdfG, RdfM and RdfS.** Secondary structures were predicted using Jpred(v4) (264). α-helices are highlighted in yellow, β-sheets are highlighted in blue. All three proteins carry a predicted two stranded MerR-family winged helix-turn-helix motif characteristic of RDFs (107, 265).

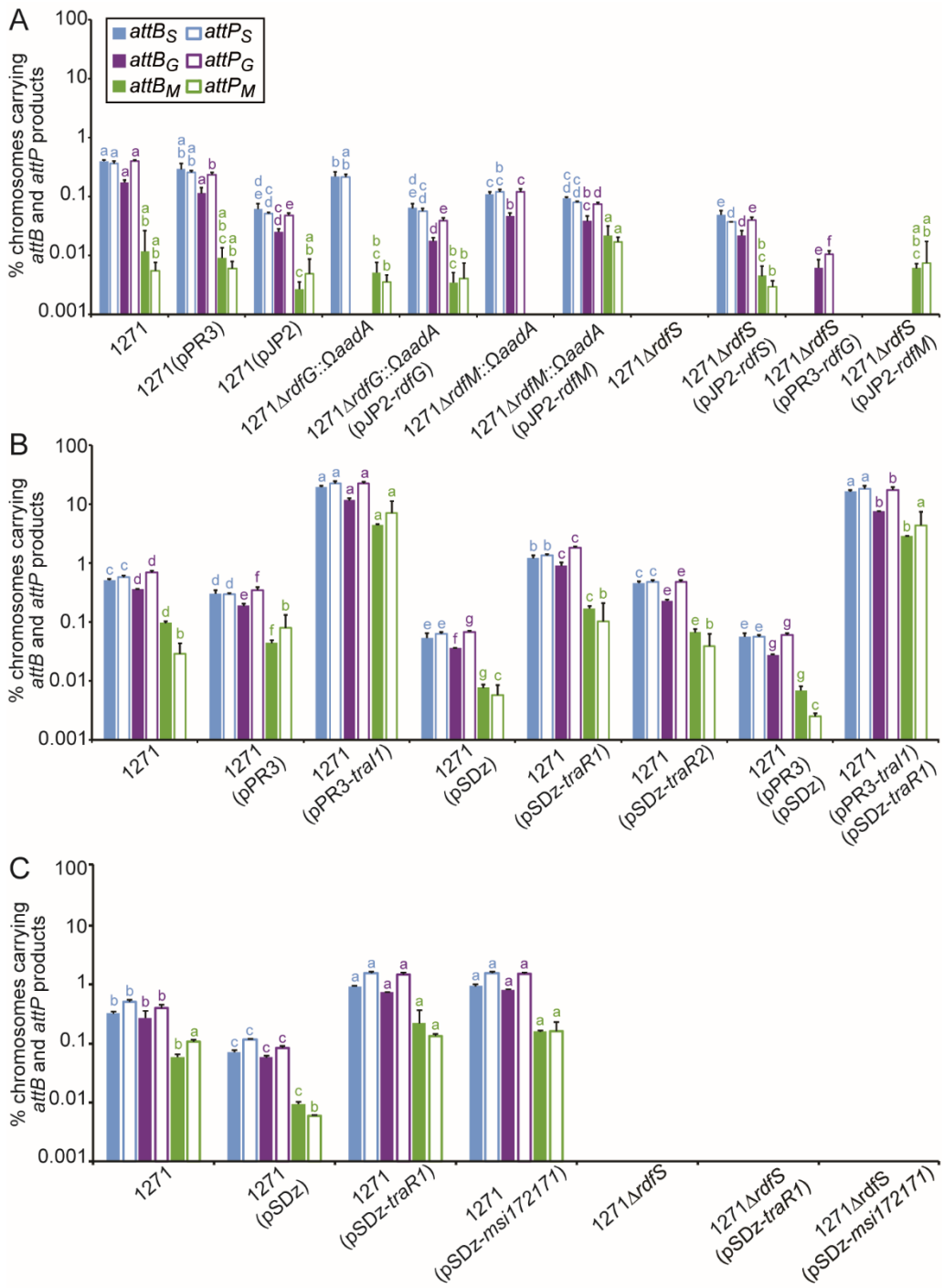


**Fig 5.3. qPCR measurement of excisive ICE*Mc*Sym$^{1271}$ recombination.** Measurements represent the mean percentage of WSM1271 chromosomes in stationary-phase cultures harbouring each excisive Int-mediated recombination product (*attB$_S$, attP$_S$, attB$_G$, attP$_G$, attP$_M$ and attP$_M$*) determined by qPCR. Where appropriate, plasmids carried by WSM1271 (here abbreviated as 1271) are listed in brackets after the strain name (see Table 1.1 for a description of plasmids). Values for each of the assay types *attB$_S$, attP$_S$, attB$_G$, attP$_G$, attP$_M$* and *attP$_M$* site were individually compared between strains within the same panel (panel A, B, or C) using ANOVA and Fisher's LSD test controlling for type I error using the Bonferroni adjustment. Groups of values from the same assay type and in the same panel that are not significantly different from each other

have the same letter (a, b, c, d, e, f or g) indicated above. Expression from the IPTG inducible promoter of pSDz constructs were not induced with IPTG as they exhibit leaky expression without induction in TY medium used for assays. (A) Involvement of *rdfG* and *rdfM* in excisive recombination. (B) Quorum sensing induction of excisive recombination. (C) Involvement of *rdfS* in excisive recombination

### *5.2.2. Quorum sensing stimulates all three excisive Int-mediated recombination reactions*

ICE*Mc*Sym$^{1271}$ carries two homologues of ICE*Ml*Sym$^{R7A}$ *traR,* hereby termed *traR1* (Mesci_5573) and *traR2* (Mesci_5676) and a homologue of ICE*Ml*Sym$^{R7A}$ *traI1* (Mesci_5572). ICE*Mc*Sym$^{1271}$ *traI1, traR1* and *traR2* were each individually overexpressed in WSM1271 on plasmids and ICE$^{3}$ excision was measured by qPCR (Fig 5.3B). Constitutive expression of *traI1* from the *nptII* promoter stimulated a 10-100-fold increase in abundance of all three *attP* and *attB* sites relative to vector-only controls. Non-induced *lac* promoter-driven expression of *traR1* or *traR2* only stimulated a modest increase in *att* site abundance relative to WSM1271, however, the vector-only control exhibited ~10-fold reduced excision frequencies, so relative to this background overexpression of the *traR1/2* genes each induced a 10-100-fold increase for all *attP* and *attB* sites. Overexpression of *traI1* and *traR1* in the same background stimulated ~1000-fold increase in abundance of all three *attP* and *attB* sites relative to the vector-only control strain. To investigate effects of the QS genes on conjugative transfer, strains overexpressing *traR1, traR2* and *traI1* were each used as donors in mating assays where *M. loti* R7ANS carrying pPR3 or pFAJ1708 was the recipient (Table 5.1). The pattern of fold-changes in conjugation frequencies for each donor strain largely mirrored

excision frequency changes observed in qPCR assays (Fig 5.3B) confirming that *traI1*, *traR1* and *traR2* also stimulated conjugative transfer.

**Table 5.1. Quorum sensing induced ICE*Mc*Sym$^{1271}$ conjugative transfer.**

| [a]Donor | Recipient | Exconjugants (per donor) | Standard deviation | [b]Fold-change |
|---|---|---|---|---|
| WSM1271 | R7ANS(pPR3) | $8.02 \times 10^{-8}$ | $1.82 \times 10^{-8}$ | - |
| WSM1271(pSDz) | R7ANS(pPR3) | $2.22 \times 10^{-8}$ | $9.12 \times 10^{-9}$ | - |
| WSM1271(pSDz-*traR1*) | R7ANS(pPR3) | $4.69 \times 10^{-7}$ | $1.11 \times 10^{-7}$ | 21 |
| WSM1271(pSDz-*traR2*) | R7ANS(pPR3) | $5.97 \times 10^{-7}$ | $1.66 \times 10^{-7}$ | 27 |
| WSM1271(pSDz-*msi172171*) | R7ANS(pPR3) | $8.49 \times 10^{-7}$ | $8.23 \times 10^{-8}$ | 38 |
| WSM1271 | R7ANS(pFAJ1708) | $8.35 \times 10^{-8}$ | $4.87 \times 10^{-8}$ | - |
| WSM1271(pPR3) | R7ANS(pFAJ1708) | $8.74 \times 10^{-8}$ | $3.89 \times 10^{-8}$ | - |
| WSM1271(pPR3-*traI1*) | R7ANS(pFAJ1708) | $1.04 \times 10^{-5}$ | $1.50 \times 10^{-6}$ | 119 |

[a] Plasmids carried by WSM1271 are listed in brackets after the strain name and are named according to the parent vector (pPR3 or pSDz) and the gene carried. Genes cloned into pPR3 and pSDz vectors are under transcriptional control from the constitutive *nptII* promoter, or an inducible IPTG promoter, respectively. Expression from the IPTG inducible promoter of pSDz constructs was not induced as they exhibit leaky expression without induction in TY medium used for assays.

[b] Fold-change is relative to control strains carrying the appropriate pPR3 or pSDz parent vector.

### *5.2.3. Dissection of quorum sensing-induced ICE$^{3}$ excision using RNA deep sequencing*

To explore the regulation of genes downstream of *traR1, traR2* and *traI1,* transcriptome sequencing (RNAseq) was carried out for a QS-induced (QS+) strain, carrying plasmid-borne copies of *traR1* and *traI1* and an uninduced strain (QS-) carrying the parent vectors pSDz and pPR3. Overall, 187 significantly differentially-expressed genes (adjusted *P*-value < 0.05) were

identified and although ICE*Mc*Sym$^{1271}$ comprised only ~7.6% of the chromosome, 15.5% (29 genes) of the differentially-expressed genes were located on ICE*Mc*Sym$^{1271}$. Genes likely involved in activation of excision and conjugation, including *msi172-msi171, rdfS*, *rlxS* and the type-IV conjugative pilus gene cluster *msi031-trbBCDEJLFGI-msi021,* were all significantly induced (Fig 5.1B and Table 5.2). The full list of differentially expressed genes has been uploaded to NCBI GEO database accession GSE108732.

An alignment of the *traI1* promoter regions ($P_{traI1}$) from ICE*Ml*Sym$^{R7A}$ and ICE*Mc*Sym$^{1271}$ revealed a *tra*-box sequence centred 69-bp upstream of the ICE*Mc*Sym$^{1271}$ *traI1* start codon (Fig 5.4A). The reads mapping to the *traI1* coding sequence were filtered from the RNAseq libraries prior to differential expression analyses (Table 5.2) because they were also present on the introduced plasmid, however, a secondary comparison of the unfiltered RNAseq reads mapping to the $P_{traI1}$ region in the QS+ transcriptome libraries relative to the QS- libraries revealed a sharp 121-fold increase in mapped reads beginning ~44-bp downstream from *tra*-box centre and 26-bp upstream of the *traI1* start codon (Table 5.2 & Fig 5.4A).

**Table 5.2. QS induced/repressed ICE*Mc*Sym$^{1271}$-encoded genes**

| Gene | Locus ID | [a]Fold-change | SE (+ 1) |
|---|---|---|---|
| **Region-α** | | | |
| *rdfS* | Mesci_5530 | 19.74 | 1.20 |
| *traF* | Mesci_5529 | 29.21 | 1.20 |
| *msi107* | Mesci_5528 | 41.10 | 1.19 |
| *rlxS* | Mesci_5527 | 58.14 | 1.17 |
| $P_{traI1}$ | - | 121.45 | 1.16 |
| $P_{traI2}$ | - | 37.54 | 1.18 |
| *traI2* | - | 141.41 | 1.16 |
| *msi172* | - | 61.71 | 1.18 |
| *msi171* | - | 156.99 | 1.16 |
| *msi021* | Mesci_5513 | 8.28 | 1.19 |
| *trbI* | Mesci_5514 | 10.58 | 1.17 |
| *trbG* | Mesci_5515 | 18.07 | 1.19 |
| *trbF* | Mesci_5516 | 14.48 | 1.19 |
| *trbL* | Mesci_5517 | 19.35 | 1.19 |
| *trbJ* | Mesci_5518 | 42.31 | 1.18 |
| *trbE* | Mesci_5519 | 64.16 | 1.17 |
| *trbD* | Mesci_5520 | 14.43 | 1.20 |
| *trbC* | Mesci_5521 | 9.71 | 1.20 |
| *trbB* | Mesci_5522 | 5.39 | 1.21 |
| *msi031* | Mesci_5523 | 13.88 | 1.20 |
| *traG* | Mesci_5524 | 2.75 | 1.16 |
| *queD* | Mesci_5560 | -2.35 | 0.83 |
| *queC* | Mesci_5561 | -2.29 | 0.82 |
| *queB* | Mesci_5562 | -2.34 | 0.83 |
| hypothetical | Mesci_5526 | 1.90 | 1.18 |
| **Region-β** | | | |
| *cbb3*-type COx (SI) | Mesci_5510 | 1.92 | 1.16 |
| Nicotinate biosynthesis | Mesci_5579 | -1.85 | 0.83 |
| *rdfG* | Mesci_2550 | 2.46 | 1.18 |
| Hypothetical | Mesci_2555 | 2.03 | 1.19 |
| **Region-γ** | | | |
| *intS* | Mesci_2349 | 2.85 | 1.15 |

[a] Differentially expressed genes (adjusted two-sided *P*-value of < 0.05) were identified using the DESeq2 package (236). Since introduced plasmids carried copies of the *traI1* and *traR* ORFs (not including promoter regions), reads mapping to these sequences were of an ambiguous origin and were therefore filtered and removed prior to mapping reads. Differential expression analysis of the *traI1* and *traI2* untranslated mRNA promoter regions, $P_{traI1}$ and $P_{traI2}$, was carried out prior to filtering – as these reads were able to be distinguished from plasmid-borne mRNAs. Reads mapping to the plasmid backbones and rRNA genes were removed prior to mapping reads for both analyses. The full list of differentially expressed genes has been uploaded to NCBI GEO database accession GSE108732.

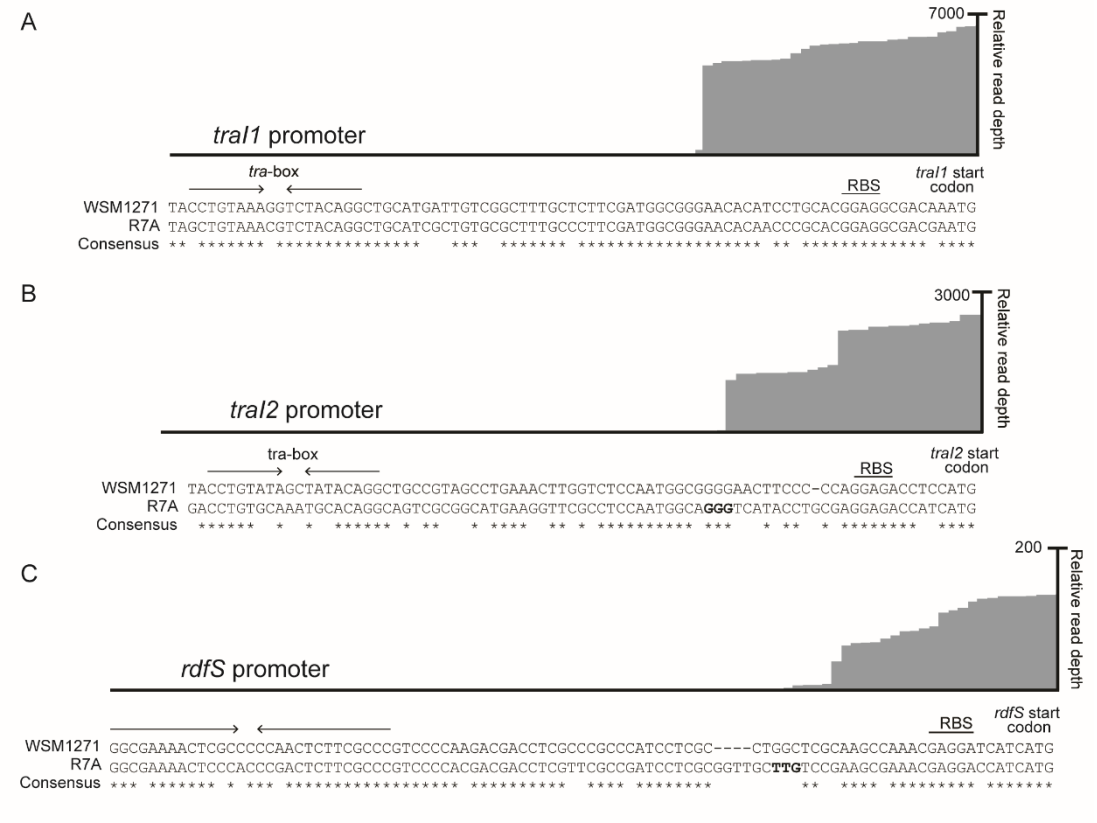


**Fig 5.4. RNA-Seq mapping of the transcriptional start sites for *traI1*, *traI2* and *rdfS*.** The promoter regions of (A) *traI1*, (B) *traI2* and (C) *rdfS* genes from WSM1271 were identified based on similarity with homologous regions in R7A. Nucleotide alignments were performed using the T-Coffee multiple sequence aligner (238). Transcriptional start sites for R7A genes previously mapped by 5'-RACE are shown in bold (167, 170). Relative read depth (or sequencing depth) plots represent a standardised value for the mean number of reads mapped to the positive strand of the regions shown in this figure from the three unfiltered QS+ transcriptome libraries of WSM1271.These plots were produced using Integrated Genome Browser (266). QS+ strains were induced for QS by overexpressing both *traI1* and *traR1* from the plasmids pPR3-*traI1* and pSDz-*traR1*, respectively Mean values of 2196.16 $\pm$ (SD) 434.70 TPM unfiltered reads and 660.88 $\pm$ 276.84 TPM unfiltered reads were mapped to the non-coding regions between the transcriptional start sites and start codons for *traI1* and *traI2,* respectively. A students t-test revealed that this difference was significant ($P$ = 0.01).

Homologues of *msi172* and *msi171* are present on ICE*Mc*Sym$^{1271}$ (4) (Fig 5.1B) but initial sequence analaysis did not identify an ICE*Ml*Sym$^{R7A}$ *traI2* homologue positioned upstream of these genes. *traI2* of ICE*Ml*Sym$^{R7A}$ appears to encode an AHL synthase paralogous with TraI1, however, mutation of *traI2* has no effect on ICE*Ml*Sym$^{R7A}$ excision and no identifiable AHL products are produced by TraI2 (167). Further inspection of the ICE*Mc*Sym$^{1271}$ *msi172-msi171* region (Figs 5.4B & 5.5A) revealed the presence of a potential *tra*-box sequence centred 398-bp upstream of the *msi172* start codon. A nucleotide alignment with the corresponding ICE*Ml*Sym$^{R7A}$ region revealed this *tra*-box was also centred 66-bp upstream of an internally-truncated *traI2* gene remnant (Figs 5.4B & 5.5). This *traI2* pseudogene overlapped the start codon of *msi172*, as does *traI2* on ICE*Ml*Sym$^{R7A}$. Interestingly, inspection of *traI2-msi172* regions in *M. loti* USDA 3471 and *M. ciceri* strains WSM4083, WSM1497 and WSM1284 revealed a similar situation; the *traI2* gene in each case was present as a potential protein-coding pseudogene upstream of *msi172* and overlapping the *msi172* start codon (Fig 5.5B). Therefore, although *traI2* has likely become a pseudogene on ICE*Mc*Sym$^{1271}$ and other symbiosis ICEs, the transcriptional coupling of the *tra*-box and translational coupling of the TraI2 and Msi172 coding sequences has been maintained. In the RNAseq experiments, *traI2, msi172* and *msi171* reads were increased ~60-160-fold in QS+ cells (Table 5.2). A sharp increase in relative read depth was observed at the *traI2* promoter 44-bp downstream of the *tra*-box centre and 21-bp upstream of the *traI2* start codon (Fig 5.4B) which spanned the entire *traI2-msi172-msi171* operon (Fig 5.6A). The likely transcription start site for *traI2* observed from RNAseq reads was consistent with the previously

mapped ICE*Ml*Sym$^{R7A}$ *traI2* promoter (Fig 5.4B) (167). Interestingly, comparison of the number of unfiltered transcripts mapping to the *traI1* and *traI2* promoter regions revealed that QS-induced expression from the *traI1* promoter (2196.16 ± [SD] 434.70 TPM) is ~3-fold stronger than that of *traI2* (660.88 ± 276.84 TPM) (Figs 5.4A-B). A similar ratio of *traI1*:*traI2* expression is also observed for ICE*Ml*Sym$^{R7A}$ (167).

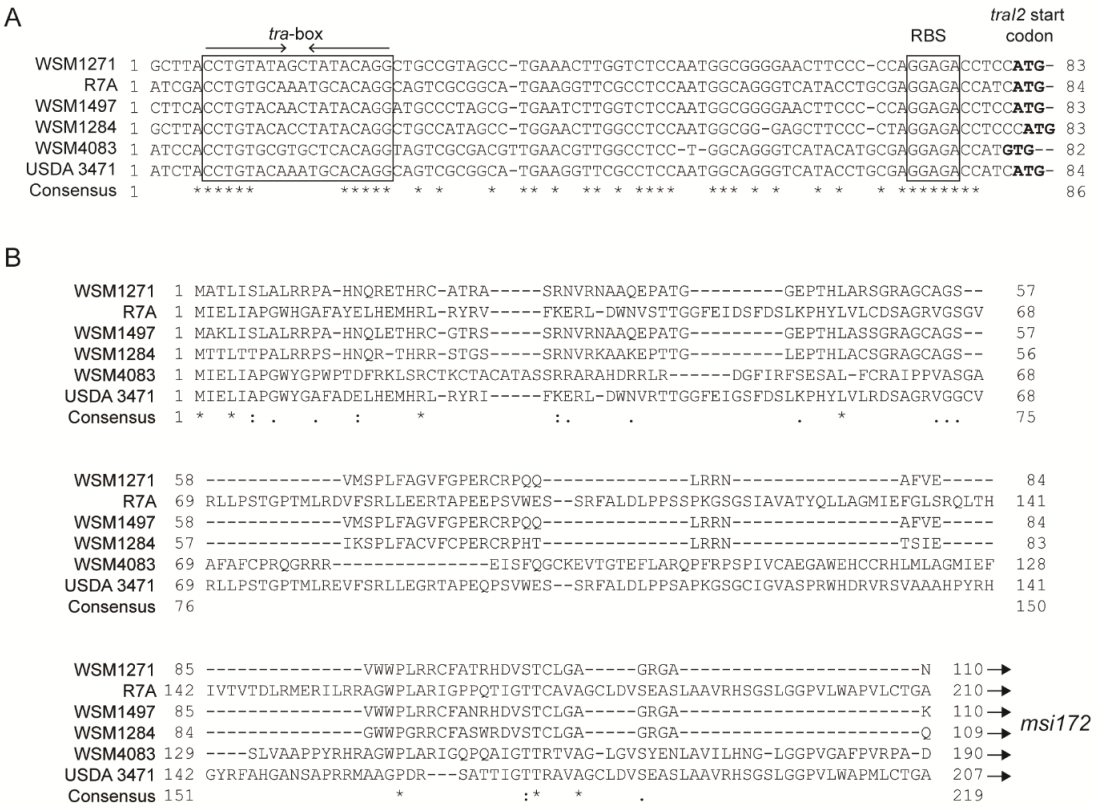


**Fig 5.5. Alignment of *traI2* promoter regions and TraI2 protein sequences in diverse *Mesorhizobium* spp.** (A) The nucleotide sequence of *traI2* promoters and (B) the TraI2 amino acid sequences from six *Mesorhizobium* strains were aligned using the T-coffee multiple sequence aligner (238).

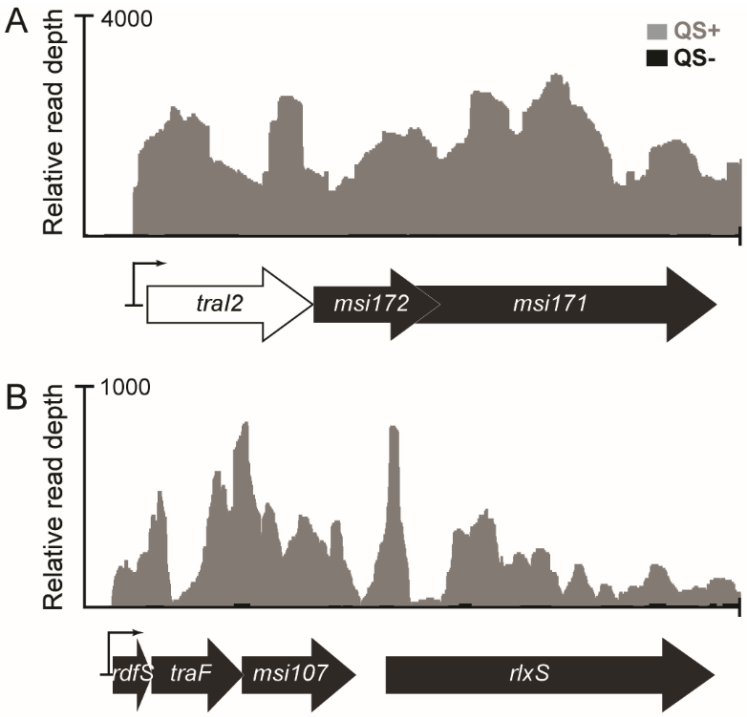


**Fig 5.6. Quorum-sensing activation of ICE*Mc*Sym$^{1271}$ promoters.** Overlayed relative read coverage (or sequencing depth) plots represent standardised values for the mean number of reads mapped to the positive strand of the regions shown in this figure from the three unfiltered QS+ (grey) and QS- (black) transcriptome libraries of WSM1271. QS+ strains were induced for QS by overexpressing both *traI1* and *traR1* from the plasmids pPR3-*traI1* and pSDz-*traR1*, respectively, whereas the QS- control strains carried the parent vectors pPR3 and pSDz. The mean read depth (or sequencing coverage) in the (A) *traI2-msi172-msi171* and (B) *rdfS-traF-msi107* and *rlxS* regions of ICE*Mc*Sym$^{1271}$ for QS- transcriptome libraries were almost non-existent relative to that of the QS+ strain. A magnified view of reads mapping to the promoter region and the DNA sequence is shown in Fig 5.4. These plots were produced using Integrated Genome Browser (266).

For ICE*Ml*Sym$^{R7A}$, FseA stimulates expression from an operon containing *rdfS*, *traF* and *msi107* (130, 170) (Fig 5.6B). The same gene cluster is present on ICE*Mc*Sym$^{1271}$ and the RNAseq read depth for the corresponding ICE*Mc*Sym$^{1271}$ homologues was increased 20-58-fold in QS+ cells (Table 5.2). A distinct read depth increase was observed 25 bp upstream of the *rdfS* start codon corresponding closely with the mapped transcriptional start site for ICE*Ml*Sym$^{R7A}$ *rdfS* (Fig 5.4C) (170). In summary, despite several genetic

rearrangements, the QS regulon of ICE*Mc*Sym$^{1271}$ appears functionally analogous to that of ICE*Ml*Sym$^{R7A}$ and importantly, QS induces the expression of *msi172*, *msi171* and *rdfS*.

### *5.2.4. rdfS is required for all three excisive Int-mediated recombination reactions*

To explore the involvement of RdfS in ICE$^{3}$ assembly and excision, a markerless deletion in *rdfS* was constructed. As expected, no $attP_S$ and $attB_S$ products were detected in this strain, but interestingly $attP_G$ and $attB_G$ and $attP_M$ and $attB_M$ products were also undetectable (Fig 5.3B). Introduction of *rdfS* expressed from its native promoter restored *attP* and *attB* production at all three sites, albeit at lower levels than wild-type WSM1271. Plasmid-based overexpression of *traR1* or *msi172-msi171* in the *rdfS* mutant did not induce excision, however, the same plasmids did induce excision and conjugative transfer in the wild-type WSM1271 (Fig 5.3C, & Table 5.2). Together these data confirmed that the stimulation and coordination of all three excision reactions by QS and *msi172-msi171* was dependent on *rdfS*.

RdfS could act either by directly stimulating excisive recombination at $att_G$ and $att_M$ sites or by up-regulating *rdfG* and *rdfM* expression. To explore these possibilities, *rdfG* and *rdfM* were overexpressed in the *rdfS* mutant to observe whether it would restore the formation of $attP_G$ and $attB_G$ and $attP_M$ and $attB_M$ sites, respectively. *rdfG* was cloned downstream of the strong constitutive *nptII* promoter and *rdfM* was cloned downstream of the *lac* promoter. Interestingly, introduction of *lac* driven *rdfM* resulted in growth arrest even in the absence of

IPTG inducer and in the presence of glucose to repress *lac* expression. This was consistent with the model for ICE[3] excision (Fig 5.1A), in which expression of *rdfM* alone splits the chromosome, presumably resulting in loss of viability. Constitutive expression of *rdfG* in the *rdfS* mutant resulted in the restored detection of $attP_G$ and $attB_G$ products in approximately 0.01% of cells (Fig 5.3A) while the other two sites remained undetectable. In contrast to *lac*-driven expression, introduction of the cloned copy of *rdfM* downstream of its native promoter restored the production of $attP_M$ and $attB_M$ sites in 0.001-0.01% of cells. Therefore, it was clear that *attP* and *attB* formation was abolished in the *rdfS* mutant but RdfS was not directly essential for excisive IntG and IntM recombination. The observation that artificially increased levels of *rdfG* or *rdfM* compensated for the loss of *rdfS* implied RdfG and RdfM expression was abolished in the *rdfS* mutant.

### *5.2.5. Overexpression of rdfS stimulates expression of rdfG and rdfM*

Inspection of RNAseq data revealed *rdfG* mRNA abundance was ~2.5-fold higher in QS+ cells (Table 5.1). *rdfM* was very weakly expressed in both QS+ and QS- cells and while there was ~2-fold more *rdfM* reads in QS+ cells, this difference was not statistically significant ($P > 0.05$). To clarify the potential role for RdfS in activation of the *rdfG* and *rdfM* promoters, the non-coding regions present upstream of each gene were cloned upstream of the promoterless *lacZ* gene of pSDz. Constructs carrying this fusion were introduced into WSM1271 carrying a constitutively expressed copy of *rdfS* (Fig 5.7A). β-galactosidase expression from the *rdfG* and *rdfM* promoters was induced ~4.5 and ~8-fold, respectively, in the presence of constitutively

expressed *rdfS*. Consistent with RNAseq data, *rdfM* expression was much lower than *rdfG* expression and almost undetectable in the absence of *rdfS*. To discount the possibility that RdfS induced expression indirectly through other factors on ICE*Mc*Sym[1271], the same set of experiments were repeated using the heterologous *M. loti* R7ANS background, which lacks all ICE genes (Fig 5.7B). These assays produced comparable results to those carried out in WSM1271, supporting the hypothesis that the transcriptional activation of *rdfG* and *rdfM* promoters by RdfS was likely direct.

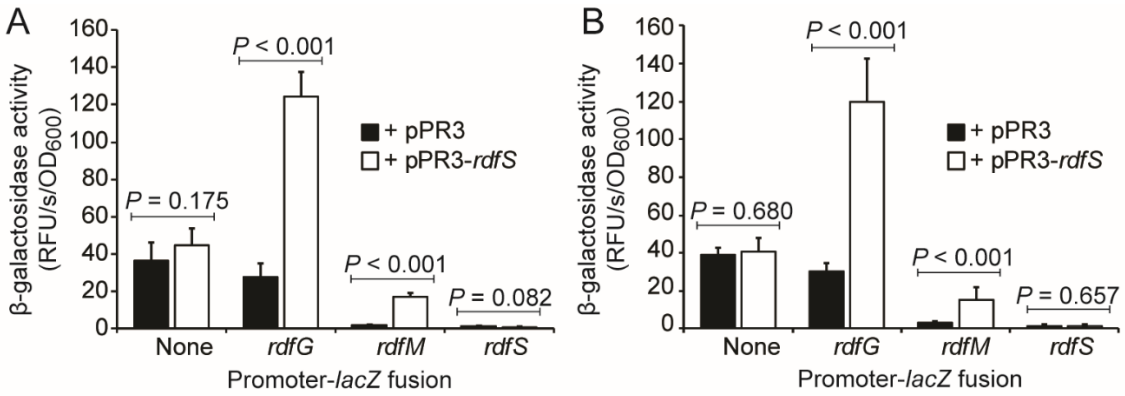


**Fig 5.7. Transcriptional regulation of *rdfG* and *rdfM* by RdfS.** β-galactosidase assays (229) were performed for (A) WSM1271 and (B) R7ANS carrying either control vector pPR3 or pPR3-*rdfS* (constitutively expressing *rdfS*) together with one of three RDF promoter-*lacZ* fusion constructs cloned into the pSDz vector. Assays were performed with six biological replicates and mean β-galactosidase activity values Relative Fluorescent Units (RFU)/s/$OD_{600}$) were compared by Bonferroni adjusted student's t-tests. SD is denoted by error bars.

## 5.3. Discussion

Excision and circularisation is an essential prerequisite for conjugative transfer of ICEs. Integrase proteins of ICEs and temperate phages generally catalyze

both the excision and integration reactions, but integrative recombination is generally favoured in the absence of a cognate RDF (107). Unlike most ICEs that excise following a single integrase-mediated recombination, ICE*Mc*Sym$^{1271}$ requires three Int-mediated reactions to excise (4). This Chapter demonstrated that RdfG, RdfM and RdfS are required for the ICE*Mc*Sym$^{1271}$ excisive IntG, IntM and IntS-mediated recombination reactions, respectively. It was also demonstrated that overexpression of the QS sensors TraR1 and TraR2 or autoinducer synthase TraI1 in WSM1271 simultaneously increased the proportion of cells in a population undergoing all three ICE*Mc*Sym$^{1271}$ excision reactions 10-100-fold. QS significantly induced mRNA abundance for the WSM1271 *traI1*, *traI2-msi172-msi171*, *rdfS* and *rdfG* genes, as well as those for conjugative pilus formation (74, 167, 170). In addition to stimulating the ICE*Mc*Sym$^{1271}$ IntS-mediated excisive recombination, RdfS was shown to transcriptionally activate the *rdfG* and *rdfM* genes. Therefore, RdfS acts as the master regulator for ICE*Mc*Sym$^{1271}$ excision.

The model for assembly and excision of ICE*Mc*Sym$^{1271}$ indicates that if the first excisive reaction is catalysed by IntM, then the chromosome is split into two inviable parts (Fig 5.1A). However, transcription of *rdfM* and *rdfG* is dependent on RdfS and thus excisive IntS-mediated recombination probably occurs prior to that of IntM and IntG in WSM1271 cells induced for ICE*Mc*Sym$^{1271}$ assembly an excision. This hierarchical genetic regulation of the three RDFs has likely evolved to minimise the potential for formation of the non-viable split chromosome configuration following spurious *rdfM* expression. In wild-type WSM1271 or QS-induced WSM1271 cells, the frequency $attP_M$ and $attB_M$ site formation was also significantly less than either $attP_G$ and $attB_G$

or *attP*$_S$ and *attB*$_S$, as was expression of *rdfM* relative to *rdfG* and *rdfS*. Moreover, introduction of a plasmid-borne copy of *rdfM* under the control of the relatively weak *lacI* promoter on pSacB (220) resulted in arrested growth of 1271Δ*rdfS* cells suggesting that even a low level of RdfM expression in the absence of RdfS and RdfG is deleterious. It is possible that the *rdfM* promoter, in addition to evolving transcriptional dependency on RdfS, has evolved to promote only subtle levels of *rdfM* expression to further reduce the likelihood of the formation of a non-viable chromosomal state. Considering the data, it seems probable that the *in situ* excisive recombination pathway of ICE*Mc*Sym$^{1271}$ follows the sequence IntS > IntG > IntM (Fig 5.1A).

In addition to RdfS, several other bacteriophage excisionases exist that act as both RDFs and transcriptional regulators (109-111, 113, 114, 267-271). Phage-P2 Cox and the coliphage-186 Apl excisionases bind and bend *attP* and *attL* DNA to promote prophage excision, but they also stimulate induction of the lytic cycle by blocking transcription of repressor genes *cI* and *c,* respectively (109-111, 113, 114, 267-270). The Cox protein additionally stimulates derepression of neighbouring P4 prophages by activating transcription from the late P4-phage promoter (269, 272). Cox-bound promoter and *attP* regions each contain six or more repeats of a “cox-box” consensus sequence that may vary in direction or percentage identity between different binding targets, and may be bound with variable affinity (111, 269, 270). A protein sharing structural homology with excisionases has recently been shown to be essential for relaxasome processing of the conjugative plasmid pIP501 (273). These examples and the findings in this Chapter emphasise that

RDFs/excisionases have evolved differential and evolutionarily flexible roles in the control of MGE dissemination.

As described in Chapter 3, the formation of ICE*Mc*Sym$^{1271}$ may have occurred following only two chromosomal inversions between three single-part ICEs or non-conjugative integrating elements (Fig 3.10). The RdfS proteins of ICE*Mc*Sym$^{1271}$ and ICE*Ml*Sym$^{R7A}$ are almost identical at the amino-acid level apart from the extreme C-terminus (Fig 5.2). Therefore, it is possible that the *rdfG* and *rdfM* promoter regions could have evolved DNA-binding targets that respond to RdfS, rather than RdfS having evolved specific new functions associated with ICE$^{3}$. Preliminary analysis of the *attL* and *attP$_S$* or the *rdfG* or *rdfM* promoter regions did not reveal any clearly conserved DNA sequence motifs for RdfS binding. However, excisionase binding sites are often poorly conserved at the DNA-sequence level and for most the mode of site recognition is not well understood. Most characterised RDFs have a winged-helix-turn-helix structure that contacts both major and minor DNA grooves, therefore overall DNA topology is believed to be especially critical for recognition (274). Given that RdfS presumably binds multiple distinct sites on ICE*Mc*Sym$^{1271}$, further work characterising the excisionase-DNA recognition characteristics of this protein could reveal the multifaceted roles of excisionases in stimulating horizontal transfer of diverse MGE.

It seems likely that the regulatory control of RdfS over *rdfG* and *rdfM* transcription could have pre-existed ICE$^{3}$ on the ancestral single-part constituents from which ICE$^{3}$ putatively evolved. Several putative symbiosis ICEs carry *rdfS* but lack an associated IntS gene and instead carry a unique integrase and distinct *attL* site within one of five serine tRNA genes

(*Mesorhizobium* spp. strains CC1192 (63); WSM3873 (NZ_LYTM00000000.1), AA23 (NZ_LYTP00000000.1) and WSM3859 (NZ_NSGG00000000.1)). Moreover, numerous more distantly related putative ICEs in the α-proteobacteria carry a homologue of *rdfS* but lack an obvious *intS* homologue (176). The conservation of *rdfS* but lack of conservation of *intS* on these ICEs suggests that RdfS homologues may be able stimulate excisive recombination through interactions with multiple distinct recombination systems. With this view in mind, the evolution of ICE$^{3}$ and capture of unique ICE genes (3) potentially involves recombination between groups of distinct ICE$^{3}$, single-part ICEs and non-conjugative integrative elements that already share common regulatory control elements.

ICE*Mc*Sym$^{1271}$-α carries two functional QS-sensor genes, *traR1* and *traR2*. Sequence comparisons of the ICE*Ml*Sym$^{R7A}$ and ICE*Mc*Sym$^{1271}$ QS loci suggest that the ICE*Mc*Sym$^{1271}$-derived TraR2 protein is the more immediate orthologue of R7A-derived TraR. Broader comparisons of the QS loci organisation between these ICEs suggest that each ICE may have evolved from an ancestral ICE carrying two complete sets of *traR-traI* loci (Fig 5.8). The DNA sequence upstream of *traI1* on ICE*Ml*Sym$^{R7A}$ lacks a *traR1* homologue but does contain sequence homologous to the 3' end of *traR1* from ICE*Mc*Sym$^{1271}$, suggesting deletion of an ancestral copy of *traR1* has occurred in R7A. The *traI2* gene on ICE*Mc*Sym$^{1271}$ appears to have become a pseudogene with several internal truncations, but a truncated seemingly nonsense open-reading-frame remains that has retained both its position relative to the upstream *tra* box and translational overlap with *msi172*, as is the case on other related ICEs (Fig 5.5). On ICE*Ml*Sym$^{R7A}$, *traI2* is a complete

and potentially functional gene, but ICE*Ml*Sym$^{R7A}$ excision or transfer is unaffected for a markerless deletion *traI2* mutant, suggesting it too may be in the early stages of pseudogenisation.

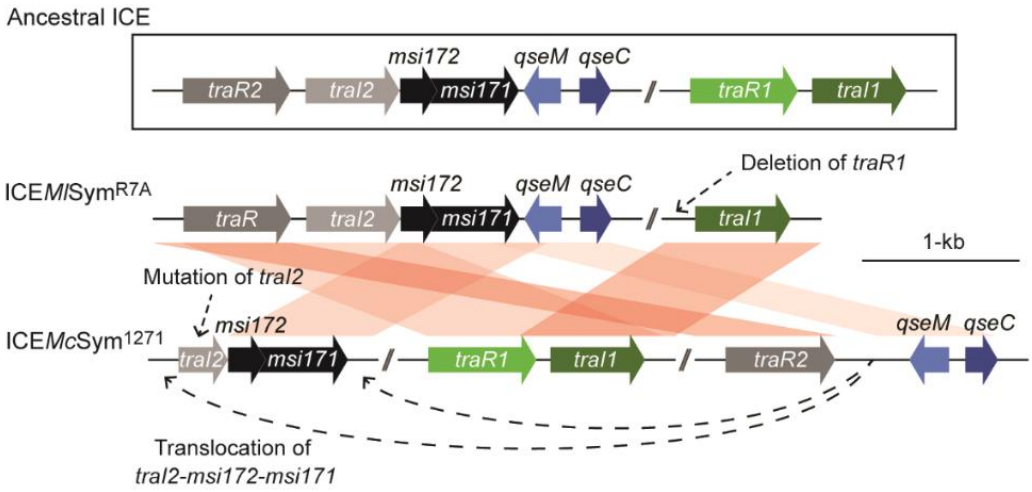


**Fig 5.8. Possible evolution of QS loci on ICE*Ml*Sym$^{R7A}$ and ICE*Mc*Sym$^{1271}$.** On ICE*Ml*Sym$^{R7A}$, *traR* is encoded upstream of an operon encoding the likely non-functional AHL-synthase gene *traI2*, *msi172-msi171* and *qseM-qseC*. The functional AHL synthase TraI1 is encoded at a separate location. ICE*Mc*Sym$^{1271}$ carries *traR2* upstream of *qseM-qseC*, however, the *traI2-msi172-msi171* region has been translocated to a different position and *traI2* has become internally truncated. ICE*Ml*Sym$^{1271}$ carries a second *traR* gene *traR1* paired with the *traI1* gene. It is likely that ICE*Ml*Sym$^{R7A}$ originally had a *traR1* gene that has subsequently been deleted. Consistent with this notion, the 100 bp upstream of *traI1* closely resembles the 3' end of *traR1*. Thus, it seems likely that an ancestral ICE carried an operon comprising *traR2-traI2-msi172-msi171* upstream of divergent *qseC* and *qseM* genes and a second QS locus containing *traR1-traI1*. Synteny comparisons were performed using the Artemis Comparison Tool (239) and plotted with genoplotR (240).

For both ICE*Mc*Sym$^{1271}$ and ICE*Ml*Sym$^{R7A}$ the functional AHL-synthase *traI1* and the apparent *traI2* pseudogene that is translationally coupled to *msi172-msi171* are proceeded by a *tra*-box sequence allowing for transcriptional

control by TraR. ICE*Ml*Sym$^{R7A}$ is sensitive to overexpression of *msi172-msi171* or *rdfS*, which cause growth inhibition and loss of ICE*Ml*Sym$^{R7A}$ respectively (74, 76, 170). In the presence of AHLs, expression from *traI2-msi172-msi171* in R7A is lower than that observed for *traI1* (167). The RNAseq data presented in this chapter similarly indicates that that expression from the ICE*Mc*Sym$^{1271}$ *traI1* promoter is stronger than from the *traI2-msi172-msi171* promoter (Table 5.2 & Figs 5.4A-B). As previously speculated (167), this separation of QS-activated genes involved in stimulation of excision (*msi172-msi171*) and AHL-production (*traI1*) has likely facilitated independent adjustment of expression levels from each QS-activated promoter. This type of genetic uncoupling of AHL synthase genes from other QS-activated genes could in some instances explain the presence of orphan – or solo - QS regulators and AHL synthase genes frequently identified throughout gram-negative bacteria (275, 276).

ICE$^{3}$s are a novel and unexpected form of MGE that exhibit a complex three-integrase system with eight separate theoretical recombination states, some of which may be inviable (Fig 5.1A). This chapter demonstrated that the activity of RdfS as a master regulator of ICE$^{3}$ excision greatly simplifies the pathway to excision. With RdfS in control, the excisive recombination reactions are induced in a predetermined order to excise ICE*Mc*Sym$^{1271}$. Like the single-part ICEs, expression of *rdfS* and excision and conjugative transfer of ICE*Mc*Sym$^{1271}$ are under QS-control. However, the ICE*Mc*Sym$^{1271}$ QS system encodes an addition LuxR-family regulator that has been lost from ICE*Ml*Sym$^{R7A}$. In the next chapter, the ICE*Mc*Sym$^{1271}$ QS-systems are explored in greater detail.

# Chapter 6.

## ICE*Mc*Sym$^{1271}$ quorum-sensing systems

## 6.1. Introduction

Quorum sensing (QS) is a form of bacterial cell-to-cell communication that involves the production of diffusible signalling molecules, termed autoinducers, that accumulate as a function of increasing population density. Bacteria perceive this signal and modulate their gene expression accordingly (168, 169). Biological functions that are controlled via QS include the production of virulence factors, biofilm formation, swarming motility, bioluminescence and horizontal transfer of conjugative MGEs (167, 174, 277-281).

Quorum sensing systems are present in both Gram-positive and Gram-negative bacteria and may be facilitated by a numerous distinct autoinducer signalling molecules including *N*-acyl-homoserine lactones (AHLs), diketopiperazines, 4-hydroxy-2-alkylquinolines, diffusible signal factors (DSF), autoinducer-2 (AI-2) and others (282). *N*-acyl-homoserine lactone-induced QS systems are undoubtedly the most common class found in gram-negative bacteria. *N*-acyl-homoserine lactones are small neutral lipid molecules that are synthesised by a LuxI-family AHL synthase through coupling of an acyl carrier protein (ACP) with *S*-adenosyl-*L*-methionine (SAM). Specific AHL-synthases produce AHLs that may vary by acyl chain length, 3-oxo or 3-hydroxy substituents, degree of unsaturation and terminal methyl branches (282-285). However, the conformation of AHLs produced by an AHL-synthase is also dependent on the availability of ACP molecules in the bacterial lipid pool (286-288). Thus, competition for available ACP substrate molecules influences the functionality of AHL-dependent bacterial QS systems.

The LuxR-family of cytosplasmic receptors perceive AHL signals (289, 290). These receptors are structurally composed of two domains; the first is an *N*-terminal acyl-binding pocket, which facilitates binding of AHLs required for the stabilisation, activation, or in some cases repression, of activity (276, 291-297); the second is a C-terminal domain helix-turn-helix motif that facilitates promoter binding required for LuxR-mediated transcriptional activation (276, 289, 290). AHLs together with LuxR-family regulators typically up-regulate cognate AHL-synthase gene expression, either through transcriptional activation or derepression mechanisms (168, 169, 277). LuxR proteins exhibit varying degrees of specificity in their response to different AHL types (275, 291, 292). Some LuxR regulators, such as AbaR of *Acinetobacter baumannii,* are highly selective, binding only a specific AHL molecule (291), whereas other LuxR regulators, such as LasR of *P. aeruginosa*, bind AHLs far more promiscuously to over 15 unique AHLs, permitting cross-talk between distinct QS systems that may exist within a single cell or within the microbial community (291, 298).

While some bacteria “eavesdrop” on AHL-signal production in neighbouring organisms via QS-cross-talk, other bacteria may attenuate or “quench” QS-regulation in neighbouring organisms through the activity of AHL-inactivating enzymes (299-303). AHL-inactivating enzymes may be beneficial for competitiveness, exogenous genetic regulation or resource scavenging (303). However, the biological relevance of these proteins is rarely fully understood. AHL-inactivating enzymes are broadly classified into three families; acylases, lactonases, or oxido-reductases, based on their mechanism of action (301). Acylases hydrolyse the amide bond between the acyl chain and the

homoserine lactone ring, whereas lactonases hyrdolyse an ester bond in the AHL opening the homoserine lactone ring. Lastly oxidoreductases modify AHLs by oxidizing or reducing the acyl chain at the third or distal carbon without cleaving the AHLs.

In Chapter 4, it was demonstrated that QS regulation of assembly, excision and conjugative transfer of the ICE[3] ICE*Mc*Sym[1271] closely resembles that described for ICE*Ml*Sym[R7A] (74, 75, 167, 170, 176). However, unlike ICE*Ml*Sym[R7A] which encodes a single copy of TraR, ICE*Mc*Sym[1271] encoded two distinct TraR homologues (TraR1 and TraR2), indicating that the QS regulation of ICE[3] excision and transfer may be more complex in this system. Therefore, in this chapter, the activity of the QS-systems of ICE*Mc*Sym[1271] were further explored.

## 6.2. Results

### *6.2.1. Three pairs of LuxR-LuxI QS loci in WSM1271*

Interrogation of the WSM1271 genome for the AHL-synthase pfam domain pfam00765 revealed that in addition to the TraI1 (Fig 6.1), a second AHL-synthase domain protein Mesci_2559 was encoded within the putative melanin biosynthesis gene-cluster carried on ICE*Mc*Sym[1271] region β (Fig 6.1B). A LuxR-family regulator Mesci_2554 harbouring the characteristic autoinducer-binding domain pfam03472 and helix-turn-helix motif (289) was encoded ~3.5 kb upstream of Mesci_2559. Mesci_2559 and Mesci_2554 were later shown to be involved in the regulation of melanin biosynthesis. These loci were renamed melanin biosynthesis regulator (*mbr*)*I* and *mbrR,* respectively.

A third AHL-synthase domain containing protein (Mesci_5594) was also encoded on the chromosome of WSM1271 (Fig 6.1C) that was 98% identical to Mlr1 of *M. loti* DSM 2626 (formerly NZP 2213). Mlr1 catalyses the synthesis of $C_{12}$ AHLs when expressed in *E. coli* (304), and DSM 2626 Mlr1 mutants are defective in the production of $C_{12}$ AHLs. A putative LuxR-family transcriptional regulator Mesci_5995 was identified directly upstream of Mesci_5594 (Fig 6.1C). BLASTN searches revealed that the Mesci_5594-5 loci are conserved across *Mesorhizobium* spp., and in reflection of this, these loci were denoted *Mesorhizobium* quorum-sensing loci (*mqs*)*R* and *mqsI*, respectively.

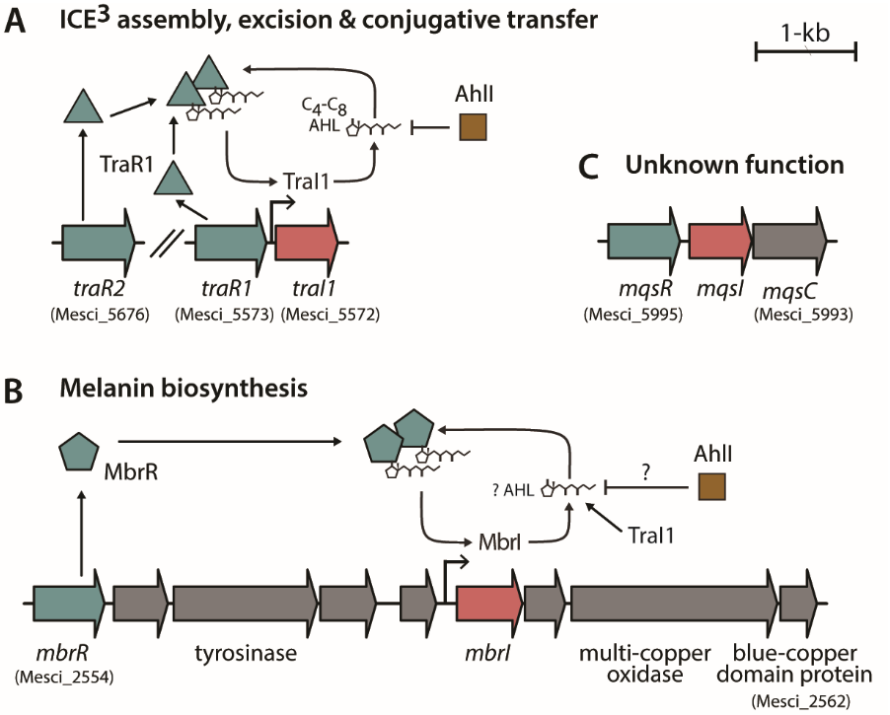


**Figure 6.1. Organization of quorum-sensing loci in WSM1271**. (A) *traR1-traI1* and *traR2* are encoded on ICE*Mc*Sym$^{1271}$-α and each gene stimulates QS-induced assembly, excision and transfer of ICE*Ml*Sym$^{1271}$ when overexpressed in WSM1271. Data presented in this chapter revealed that TraR1 and TraR2 promote expression from the *traI1* promoter and are activated by TraI1-derived AHLs. (B) *mbrR-mbrI* are encoded on ICE*Mc*Sym$^{1271}$-β. MbrR promotes transcription from the *mbrI* promoter and stimulates the biosynthesis of melanin, presumably through transcriptional activation of the downstream multi-copper oxidase and blue-copper domain containing protein which together likely encode a laccase protein (305-309). MbrR is activated by TraI1 or MbrI-derived AHLs. (C) The *mqsR-mqsI-mqsC* genes are conserved in *Mesorhizobium* chromosomes, however, their biological function and regulation has not yet been elucidated. All QS systems are may be partially repressed in WSM1271 through the inactivation of AHLs by AhlI.

### *6.2.2. CV026 bioassays for AHL production by TraI1 and MbrI*

To functionally characterize the QS-genes identified in ICE*Mc*Sym$^{1271}$, *Chromobacterium violaceum* CV026 bioassays (211) were used to detect the production of short chain ($C_4$-$C_8$) AHLs in *E. coli* DH5α ectopically expressing

*traI1* and *mbrI* following introduction of the plasmids pPR3-*traI1* and pPR3-*mbrI* (a vector carrying *mbrI* downstream of the *nptII* promoter), respectively. Violacein production was induced in CV026 streaked adjacent to DH5α overexpressing *traI1,* but not when streaked next to DH5α overexpressing *mbrI* or carrying the empty vector pPR3 (Fig 6.2A), confirming that like its homologue in R7A, TraI1 likely produced $C_4$-$C_8$ AHLs (167), whereas MbrI may produce a different molecule (Fig 6.2A). pPR3-*traI1* was then introduced into R7ANS. Violacein production was induced in CV026 by supernatant collected from this strain, but not by supernatants from the control strain R7ANS, confirming that *traI1* catalysed the synthesis of $C_4$-$C_8$ AHLs in R7ANS (Fig 6.2C).

Next, CV026 well-diffusion bioassays were performed on supernatants from cultures of WSM1271 and WSM1271 carrying pPR3-*traI1*. Unexpectedly, violacein production in CV026 was not induced by any of these supernatants (Fig 6.2B,C). Therefore, despite pPR3-*traI1* conferring the capacity for CV026 violacein production in both *E. coli* and R7ANS, the same construct was unable to induce the production of CV026-detectable AHLs in WSM1271. This suggested that TraI1-dependent short-chain AHL production might be either defective in this background, or that the AHLs produced might be actively degraded.

To test whether AHL production from ICE*Mc*Sym$^{1271}$ could be detected when the ICE is expressed in different *Mesorhizoibum* backgrounds, pPR3-*traI1* was introduced into three other ICE*Mc*Sym$^{1271}$-harbouring strains; *M. australicum* WSM2073, *M. opportunistum* WSM2075 and R7Mc1 (an R7ANS exconjugant

of ICE*Mc*Sym[1271] cured of all introduced plasmids). Supernatants of WSM2073 and WSM2075 failed to induce violacein production in CV026, with or without pPR3-*traI1* (Fig 6.2C). In contrast, supernatants of R7Mc1 induced violacein synthesis in CV026 with or without pPR3-*traI1,* demonstrating that the existence of the ICE*Mc*Sym[1271]-derived AHL-synthase was able produce $C_4$-$C_8$ AHLs at a concentration detectable by CV026 bioassays in the R7ANS background. These observations together with the previously presented RNAseq and qPCR data (Chapter 4) suggest that *traI1* is functional, but the detection of TraI1-derived AHLs in WSM2073, WSM2075 and WSM1271 is suppressed.

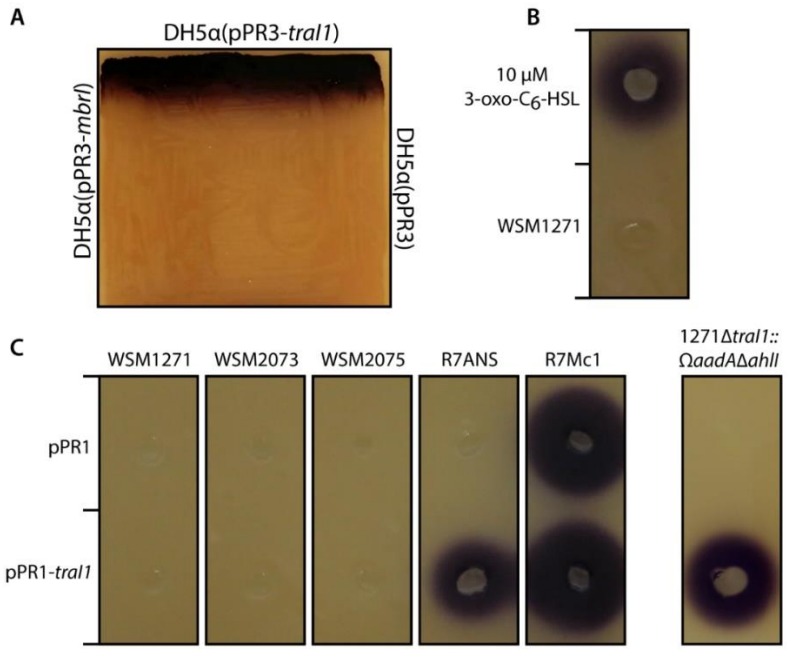


**Figure 6.2. AHL production in various bacterial strains ectopically expressing the WSM1271 derived AHL synthases *traI1* or *mbrI*.** (A) The *C. violaceum* CV026 biosensor strain (211) was streaked adjacent to *E. coli* DH5α carrying constitutively expressed plasmid borne copies of *traI1* (pPR3-*traI1*), *mbrI* (pPR3-*mbrI*), or the empty vector pPR3. (B & C) Spent supernatants of pH 6.5 buffered TY cultures of *Mesorhizobium* strains were loaded into 10 mm diameter wells bored into LB agar that had been overlayed with a molten agar culture of CV026. Ten micro-molar 3-oxo-$C_6$-HSL was loaded as a positive control where relevant. The production of the purple pigment, violacein, indicates detection of $C_4$-$C_8$ AHLs.

### *6.2.3. The α/β-fold family hydrolase AhlI inactivates TraI1-derived AHLs in diverse mesorhizobia*

To explore whether WSM1271, WSM2073 and WSM2075 actively degraded AHLs, 10 µM 3-oxo-$C_6$-HSL was added into pH 6.5 buffered stationary-phase broth cultures of these strains and the cultures were incubated for 12 h at 28°C, prior to collecting sterile supernatants for detection of AHLs by CV026 bioassays (232). Supernatants collected from each culture immediately following the addition of 3-oxo-$C_6$-HSL (0 h incubation) induced violacein production in CV026 (Fig 6.3A). Following 12 h incubation, supernatants collected from the R7ANS culture still induced CV026 violacein production, whereas supernatants collected from culture of WSM1271, WSM2073 and WSM2075 showed no induction. The pH of samples after 12 h incubation was confirmed to be between 6.5 and 7.0, seemingly ruling out the possibility of alkaline pH-dependent AHL lactonolysis (310). Therefore, it seemed likely that 3-oxo-$C_6$-HSL had been degraded in the cultures of WSM1271, WSM2073 and WSM2075.

To elucidate whether 3-oxo-$C_6$-HSL inactivation required WSM1271, WSM2073 and WSM2075 cells or could occur with exposure to supernatant alone, the previous experiment was modified by adding 10 µM 3-oxo-$C_6$-HSL to the filter sterilised (cell-free) broth culture supernatants of these strains, prior to incubation and detection of AHLs by CV026 bioassays. Supernatants collected from all cultures at both 0 and 12 h incubation induced violacein synthesis in CV026 (Fig 6.3B), indicating that inactivation of 3-oxo-$C_6$-HSL does not occur in the in supernatants. Therefore, inactivation was likely to occur through some intracellular degradation mechanism.

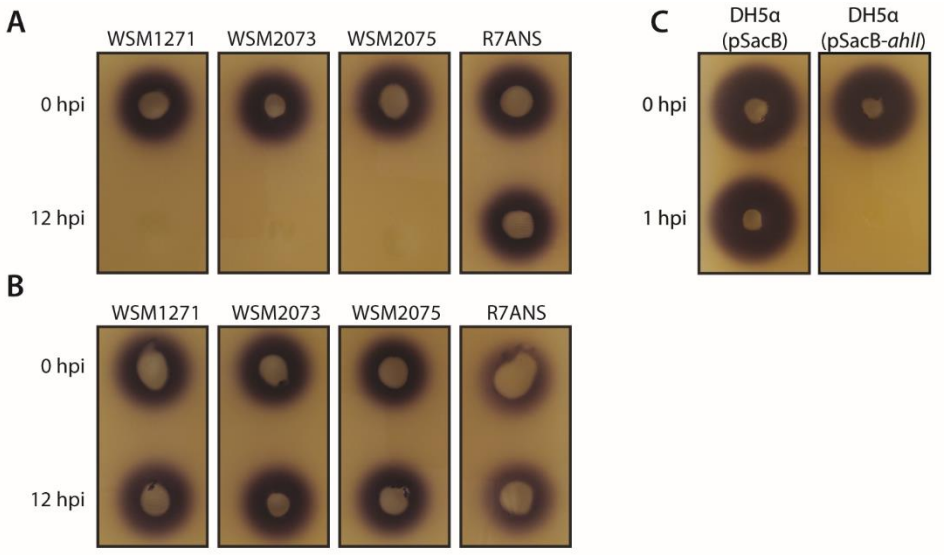


**Figure 6.3. *aidH*-dependent inactivation of 3-oxo-$C_6$-HSL.** *C. violaceum* CV026 well-diffusion bioassays (211) were used to detect C4-C8 AHLs in samples of cell-free supernatant collected from stationary-phase (A&C) cell suspensions; or (*B*) culture supernatants that had been incubated in the presence of 10 μM 3-oxo-$C_6$-HSL. Production of the purple pigment violacein at 0 h post incubation (hpi) but not at later time points indicates that the concentration of 3-oxo-$C_6$-HSL had fallen below detectable levels by CV026 bioassays.

To identify potential AHL-inactivating enzymes encoded by WSM1271, WSM2073 and WSM2075, the amino-acid sequences of diverse bacterial and archaeal AHL-inactivating enzymes listed in reference (301) were queried against these bacterial genomes using BLASTP. Based on a cutoff value of 70% amino-acid similarity, only homologues of the *Ochrobactrum* sp. T63 α/β-hydrolase fold family $C_4$-$C_{10}$ AHL-lactonase AidH were identified (311). AidH was 81% similar to Mesci_2383 in WSM1271, 80% similar to Mesop_2525 in WSM2075, 79% similar to Mesau_02412 in WSM2073, with the protein absent in R7A (Table 6.1). An alignment of these AidH homologues revealed that each carried the Ser(102)/His(248)/Glu(219) catalytic triad typical of the α/β-hydrolase fold family proteins (312), and each also carried the Gly100-X-Ser102-X-Gly104 motif required for AidH lactonolysis activity (311, 313) (Fig

6.4). Using a BLASTP query for the *Ochrobactrum* sp. T63 AidH amino acid sequence against the 113 fully or partially sequenced *Mesorhizobium* genomes on the NCBI (https://www.ncbi.nlm.nih.gov/) and IMG (242) databases, AidH homologues were identified in an additional 17 diverse *Mesorhizobium* strains (Table 6.1). None of the identified homologues harbored a type III or type IV secretion signal peptide sequence suggesting that these proteins were not secreted from the cell. Moreover, AidH was not encoded on a symbiosis ICE in any of the mesorhizobia with completed genomes sequences or present in the R7ANS genome. Considering these observations, the *Mesorhizobium* AidH homologue appeared a good candidate for further investigation of the observed inactivation AHLs in strains WSM1271, WSM2073 and WSM2075.


**Figure 6.4. Alignment of AidH homolgoues.** Amino acid sequences of AidH (sequence ID ACZ73823.1) from *O.* sp. T63, AiiO from *O.* sp. A44 (sequence ID WP_095447712.1), and homologues in WSM1271 (Locus ID Mesci_2383), 2073 (Locus ID Mesau_02412) and WSM2075 (Locus ID Mesop_2525) were aligned using T-COFFEE multiple sequence aligner (238). Each protein carries the Ser(102)/His(248)/Glu(219) catalytic triad (highlighted in grey) typical of α/β-fold family hydrolase proteins (312), and the Gly100-X-Ser102-X-Gly104 motif

which has shown to be essential for AidH activity in *O.* sp. T63 (311, 313). Catalytic residues are shown in bold.

**Table 6.1. BLASTP results for *Ochrobactrum* sp. T63 AidH homologues in diverse *Mesorhizobium* spp.**

| Strain | Locus ID [a] | Length | Identities | Positives |
|---|---|---|---|---|
| *Mesorhizobium australicum* B5P | Ga0048943_5523 | 275 | 154/267 | 189/267 |
| *M. australicum* WSM2073 | Mesau_02412 | 272 | 186/271 | 215/271 |
| *M. ciceri* bv. biserrulae WSM1271 | Mesci_2383 | 272 | 182/266 | 217/266 |
| *M. ciceri* bv. biserrulae WSM1497 | Ga0199033_11 | 272 | 182/266 | 216/266 |
| *M. ciceri* CC1192 | Ga0133496_126007 | 272 | 182/266 | 217/266 |
| *M. ciceri* CMG6 | MescicDRAFT_00051400 | 272 | 180/266 | 216/266 |
| *M. ciceri* WSM1284 | Ga0133321_122410 | 272 | 182/266 | 217/266 |
| *M. ciceri* WSM4083 | MESCI2DRAFT_00027250 | 272 | 180/266 | 214/266 |
| *M. loti* DSM 2626 | Ga0215673_11147 | 272 | 176/266 | 214/266 |
| *M. loti* TONO | - | 272 | 183/266 | 212/266 |
| *M. loti* WSM1293 | MesloDRAFT_00041470 | 272 | 180/266 | 215/266 |
| *M. mediterraneum* USDA 3392 | - | 274 | 164/268 | 197/268 |
| *M. opportunistum* WSM2075 | Mesop_2525 | 272 | 178/271 | 216/271 |
| *M.* sp. L2C084A000 | Ga0123922_101761 | 272 | 178/266 | 215/266 |
| *M.* sp. LNHC221B00 | Ga0123916_103246 | 244 | 152/226 | 181/226 |
| *M.* sp. LNHC232B00 | Ga0123914_114216 | 272 | 178/266 | 215/266 |
| *M.* sp. Root172 | Ga0124814_10272 | 272 | 184/266 | 214/266 |
| *M.* sp. STM 4661 | Ga0035947_04202 | 274 | 163/267 | 201/267 |
| *M.* sp. URHC0008 | N549DRAFT_05317 | 272 | 185/266 | 216/266 |
| *M.* sp. YR577 | Ga0115469_10271 | 272 | 177/271 | 215/271 |

[a] A dash (-) indicates a sequence that has not been denoted a Locus ID. Proteins were considered homologous if above 70% similarity (positives).

To assess whether Mesci_2383 encoded an AHL-inactivating protein, this ORF was cloned from WSM1271 downstream of the IPTG inducible promoter of pSacB (creating pSacB-*ahlI*) and introduced into *E. coli* DH5α for AHL inactivation assays (as described in Fig 6.3). Culture supernatants of both DH5α overexpressing Mesci_2382 and the control strain DH5α carrying pSacB were collected immediately following the addition of 3-oxo-$C_6$-HSL (0 h incubation). These supernatants induced violacein production in CV026 (Fig 6.3C). However, although supernatants of DH5α carrying pPR3 collected at 6 h incubation induced violacein production in CV026, the supernatant of DH5α

overexpressing Mesci_2383 did not, indicating that the 3-oxo-$C_6$-HSL in this sample had been inactivated. Owing to the observed AHL-inactivating activity of this enzyme, Mesci_2383 was subsequently termed *N*-acyl-homoserine lactone inactivator (*ahlI*).

To further characterize *ahlI*, a markerless deletion of WSM1271 was attempted using the two-step gene deletion protocol described in Section 2.2. Although strains were readily isolated carrying the *ahlI* alelle deletion vector pEX∆*ahlI* integrated via single crossover adjacent to *ahlI*, screening of all colonies following SacB-mediated selection of double crossover *ahlI* deletion mutants revealed that the plasmid integrants had reverted to wild-type genotype during final recombination step. Overexpression of the QS system in *M. loti* R7A, or overexpression of *msi172-msi171*, can cause growth-inhibitory effects due to deregulated overexpression of *rdfS* and potentially other genes involved in ICE excision and transfer (167, 170). It seemed possible that deletion of *ahlI* might have similar effects in WSM1271 through the increased stability of AHLs produced from *traI1* and the resulting positive-feedback loop with TraR1/TraR2. To test this hypothesis, the *traI1* ORF in WSM1271 was replaced with an Ω*aadA* cassette and *ahlI* mutant construction was then attempted in the resulting strain 1271∆*traI1*::Ω*aadA*. The *ahlI* deletion was successfully constructed in the 1271∆*traI1*::Ω*aadA* background in the first attempt, producing strain 1271∆*traI1*::Ω*aadA*∆*ahlI.* Introduction of pPR3-*traI1* into this strain stimulated CV026-detectable AHL production in supernatants, confirming that *ahlI* was responsible for the inability to detect AHLs in wild-type WSM1271 (Fig 6.2C).

### *6.2.4. TraR1 and TraR2 are activated by TraI1-derived AHLs*

The RNAseq data presented in Chapter 4 demonstrates that expression from the *traI1* promoter ($P_{traI1}$) is induced in WSM1271 cells co-overexpressing *traR1* and *traI1*. To confirm that TraR1 and TraR2 activate expression from $P_{traI1}$, the 112-bp DNA region upstream of *traI1* was fused to the *lacZ* genes of pSDz, pSDz-*traR1* and pSDz-*traR2,* creating pSDz-tb, pSDz-tb*traR1* and pSDz-tb*traR2*, respectively. The resulting constructs were mobilized into WSM1271 and expression from the plasmid borne $P_{traI1}$ of each strain monitored by β-galactosidase assays. Expression from $P_{traI1}$ was measured in the negative-control strain WSM1271 carrying the pSDz-tb at 100.69 $\pm$ 9.07 (SD) relative fluorescent units (RFU)/s/$OD_{600}$, and this was increased a further 3.74-fold to 386.32 $\pm$ 32.74 in the presence of cloned *traR1* (without addition of IPTG), confirming that TraR1 stimulates expression from $P_{traI1}$. In contrast, the presence of cloned *traR2* (120.56 $\pm$ 22.31 RFU/s/OD600) did not induce $P_{traI1}$.

The effect of overexpressing various combinations of ICE*Mc*Sym$^{1271}$-derived QS genes on the induction of $P_{traI1}$ was next investigated in the background of *M. loti* R7ANS, which lacks QS genes apart from the chromosomal *mqsRI* locus present in all *Mesorhizobium* spp. To facilitate this, a set of R7ANS strains carrying pSDztb, pSDz-tb*traR1,* or pSDz-tb*traR2*, in combination with pPR3, pPR3-*traI1* and pPR3-*mbrI,* were generated, and β-galactosidase assays were performed.

Expression from $P_{traI1}$ was measured in the negative-control strain R7ANS carrying pSDz-tb and pPR3 at 28.50 $\pm$ 9.22 RFU/s/$OD_{600}$ (Fig 6.5). Overexpression of *traI1* or *mbrI* in the absence of *traR1* or *traR2* did not induce

$P_{traI1}$. In contrast, overexpression of *traR1* in the absence of *traI1* and *mbrI* induced $P_{traI1}$ 17.75-fold, and this was increased a further 2.16-fold by co-overexpressing *traI1* with *traR1*, demonstrating that TraR1-derived AHLs activated TraR1. Interestingly, co-overexpression of *mbrI* with *traR1* resulted in $P_{traI1}$ induction ~half that observed when *traR1* was overexpressed in the absence of *traI1* or *mbrI,* indicating that MbrI-derived AHLs may inhibit TraR1. Although both overexpression of *traR2* in the absence of *traI1* and *mbrI,* or co-overexpression of *traR2* with *mbrI* did not induce $P_{traI1}$, co-overexpression of *traR2* with *traI1* induced $P_{traI1}$ ~44-fold more than that of the control strain R7ANS carrying pSDz-tb and pPR3. Therefore, like TraR1, TraR2 also promotes expression from the *traI1* promoter and requires TraI1-derived AHLs for its activation. It is possible TraR2-dependent activation of the *traI1* promoter was masked our previous experiments for WSM1271 as a consequence of AhlL-mediated AHL-inactivation in this strain.

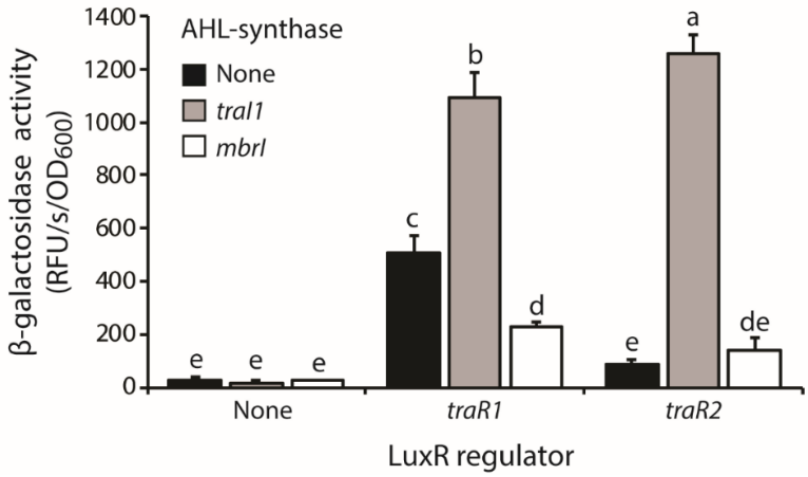


**Figure 6.5. Activation of TraR1 and TraR2 by TraI1-derived AHLs.** A set of 9 R7ANS strains carrying pSDztb, pSDz-tb*traR1,* or pSDz-tb*traR2*, in combination with pPR3, pPR3-*traI1* and pPR3-*mbrI,* were generated to monitor expression from plasmid borne *traI1* promoter-*lacZ* fusions when *traR1* or *traR2* was overexpressed alone, or in combination with *traI1* or *mbrI*. β-galactosidase assays were used to monitor expression. Values for relative fluorescent units (RFU)/s/$OD_{600}$ are the mean and SD (denoted by error bars) of 3 biological replicates. Mean values were compared using Fisher's LSD test controlling for type I error using the Bonferroni adjustment. Matching letters above bars indicate no significant difference between mean values.

If TraI1-derived AHLs were required for the activation of TraR1 or TraR2 in WSM1271, then overexpression of *traR1* or *traR2* in 1271Δ*traI1*::Ω*aadA* would presumably not induce ICE*Mc*Sym$^{1271}$ excision. To test this, pSDz-*traR1* and pSDz-*traR2* were mobilised into 1271Δ*traI1*::Ω*aadA* and analysed by qPCR. In wild-type WSM1271, $attB_G/P_G$ and $attB_S/P_S$ sites were detected in 0.1-1% of cells and $attB_M/P_M$ sites in 0.01-0.1% of cells (Fig 6.6). Deletion of *traI1* in WSM1271 did not have a major effect on the abundance of any ICE*Mc*Sym$^{1271}$ *attP* and *attB* sites, demonstrating that this gene was not essential for ICE*Mc*Sym$^{1271}$ assembly and excision. Although overexpression of *traR1* in

WSM1271 elevated the abundance of all *attP* and *attB* sites 10-100-fold relative to the control strain WSM1271 carrying pSDz, overexpression of *traR1* in 1271Δ*traI1*::Ω*aadA* only elevated the abundance of *attP and attB* sites ~2-10 that of the same control strain WSM1271(pSDz). The same trend was observed when *traR2* was overexpressed in WSM1271 and 1271Δ*traI1*::Ω*aadA*. Therefore, *traI1*-produced AHLs were not essential for the activation of TraR1 or TraR2, but were required for maximum activation of ICE*Mc*Sym$^{1271}$ assembly and excision by either *traR* gene. This situation mirrors observations for ICE*Ml*Sym$^{R7A}$, where deletion of *traI1* does not reduce excision but is required for activation of excision by TraR.

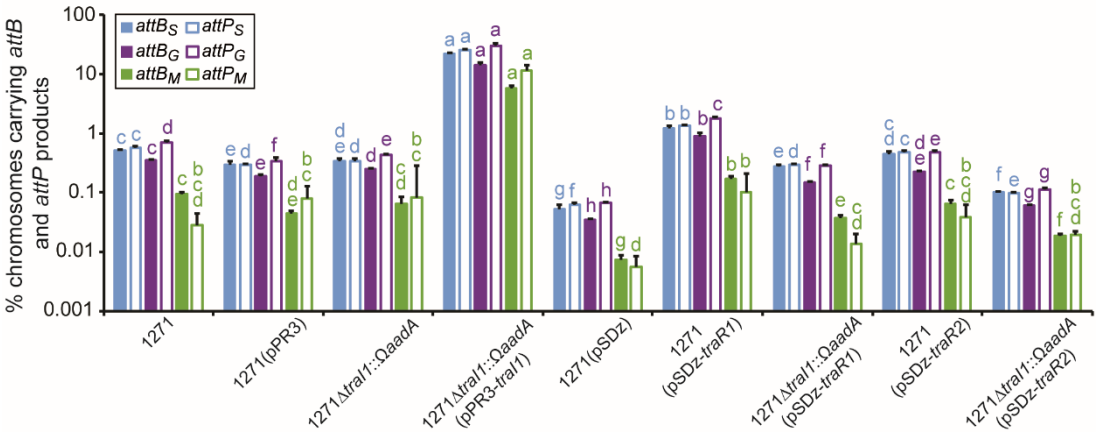


**Figure 6.6. Involvement of *traI1* for TraR1/TraR2-induced excision of ICE*Mc*Sym$^{1271}$.** Measurements represent the mean percentage of WSM1271 chromosomes in stationary-phase cultures harbouring each excisive Int-mediated recombination product (*attB$_S$, attP$_S$, attB$_G$, attP$_G$, attP$_M$ and attP$_M$*) determined by qPCR. Where appropriate, plasmids carried by WSM1271 (here abbreviated as 1271) are listed in brackets after the strain name (see Table 1.1 for a description of plasmids). Values for each of the assay types *attB$_S$, attP$_S$, attB$_G$, attP$_G$, attP$_M$* and *attP$_M$* site were individually compared between strains using ANOVA and Fisher's LSD test controlling for type I error using the Bonferroni adjustment. Groups of values from the same assay type and in the same panel that are not significantly different from each other have the same letter (a, b, c, d, e, f, g or h) indicated above.

### *6.2.5. MbrR regulates melanin biosynthesis*

*mbrR* is encoded upstream of a predicted melanin biosynthesis gene cluster of ICE*Mc*Sym[1271], therefore it seemed likely that this gene may be involved in the regulation of melanin biosynthesis. To explore this possibility, *mbrR* was overexpressed in WSM1271 by introducing the plasmid pSDz-*mbrR* which carries the *mbrR* ORF fused to an IPTG inducible promoter and melanin deposition assays were performed (230, 314). An intense purple-brown pigment was produced by WSM1271 induced for *mbrR* expression with IPTG following 14 d incubation at 28°C, however, no such pigmentation was observed for the control strains WSM1271 or WSM1271 carrying pSDz (Fig 6.7A). Thus, MbrR appears to function as a regulator of melanin biosynthesis in WSM1271. Melanin deposition assays were also performed on WSM1271 overexpressing *traR1* and *traR2*, however both strains failed to produce the melanin-like pigment (Fig 6.7A).

The potential role of MbrR in ICE*Mc*Sym[1271] assembly and excision was also explored by performing qPCR on WSM1271 overexpressing *mbrR*. Even when *mbrR* expression was induced with IPTG, *attP* and *attB* abundance was no different from the control strain WSM1271 carrying pSDz (Fig 6.7B). Therefore, it was unlikely that MbrR stimulated ICE*Mc*Sym[1271] excision.

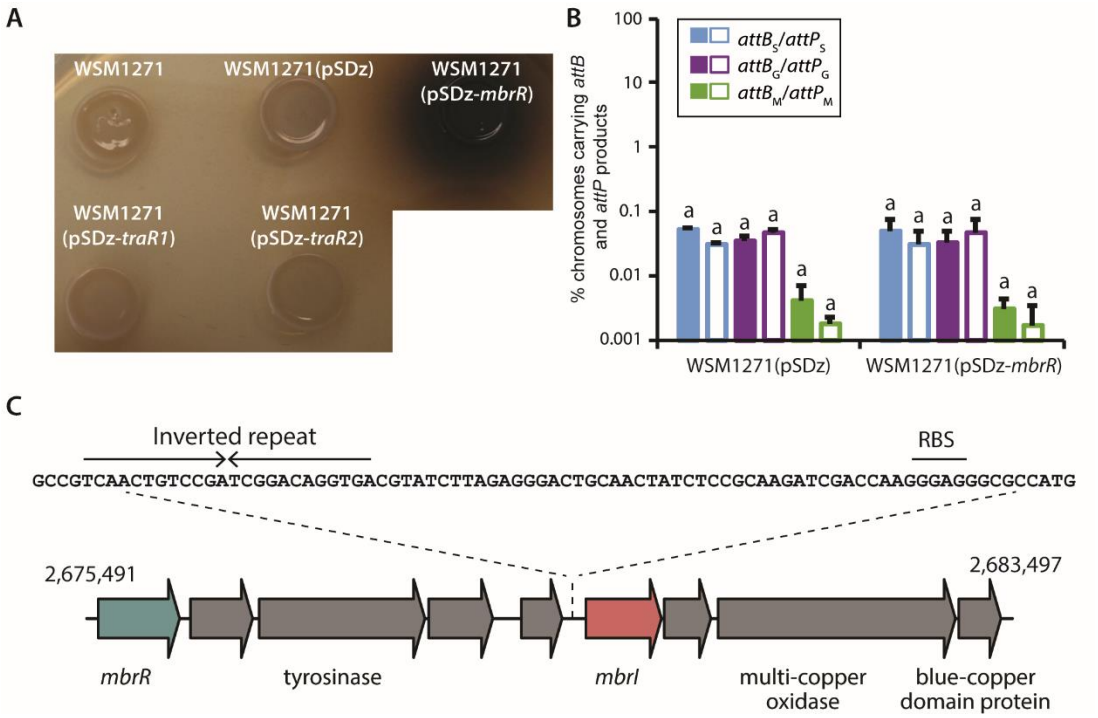


**Figure 6.7. Role of *mbrR* in melanin biosynthesis.** (A) Melanin production was monitored visually by spotting 20-µL aliquots of stationary-phase *Mesorhizobium* broth cultures onto TYT agar supplemented with 1 µM IPTG and incubating for 14 days at 28°C (230, 314). (B) Measurements represent the mean percentage of WSM1271 chromosomes in stationary-phase cultures harbouring each excisive Int-mediated recombination product ($attB_S$, $attP_S$, $attB_G$, $attP_G$, $attP_M$ *and* $attP_M$) determined by qPCR. Values for each of the assay types $attB_S$, $attP_S$, $attB_G$, $attP_G$, $attP_M$ and $attP_M$ site were individually compared between strains using ANOVA and Fisher's LSD test controlling for type I error using the Bonferroni adjustment. Groups of values from the same assay type and in the same panel that are not significantly different from each other have the same letter (a) indicated above.

Inspection of the predicted melanin biosynthesis cluster of ICE*Mc*Sym[1271] revealed the presence of a 12-bp inverted repeat centred 69-bp upstream from the *mbrI* start codon that may comprise a binding site for MbrR (Fig 6.7C). To explore whether MbrR stimulated expression from this putative promoter, the 193-bp region of DNA upstream of *mbrI* was fused to the *lacZ* gene of the plasmids pSDz and pSDz-*mbrR* (creating pSDz-mb and pSDz-mb*mbrR*, respectively) and the resulting plasmids were mobilised into WSM1271. β-galactosidase assays were used to monitor expression from the putative *mbrI*

promoter region ($P_{mbrI}$) in the resulting strains. Expression from $P_{mbrI}$ was measured in the negative-control strain WSM1271 carrying pSDz-mb at 5.11 ± 0.69 RFU/s/$OD_{600}$, and this was increased a further 28.8-fold to 147.14 ± 16.11 RFU/s/$OD_{600}$ when *mbrR* was overexpressed without IPTG induction. Thus, it seemed likely that MbrR induced transcription from $P_{mbrI}$ in WSM1271.

### *6.2.6. MbrR is activated by TraI1 or MbrI-derived AHLs*

To explore which ICE*Mc*Sym$^{1271}$-encoded AHL-synthase(s) produced AHLs that activated MbrR, a set of 6 R7ANS strains carrying pSDzmb, or pSDz-mb*mbrR*, in combination with pPR3, pPR3-*traI1* and pPR3-*mbrI* were generated. β-galactosidase assays were used to monitor the induction of expression from the plasmid-borne *mbrI* promoter $P_{mbrI}$ in each of the newly generated strains where *mbrR* was overexpressed alone, or in combination with *traI1* or *mbrI.*

Expression from $P_{mbrI}$ was measured in the negative-control strain R7ANS carrying pSDz-mb and pPR3 at 3.87 ± 1.34 RFU/s/$OD_{600}$ (Fig 6.8). Overexpression of *traI1* or *mbrI* in the absence of *mbrR* did not induce $P_{mbrI}$, but surprisingly, overexpression of *mbrR* alone induced $P_{mbrI}$ 118-fold indicating that this regulator maintains its ability to induce $P_{mbrI}$ in the absence of *traI1* and *mbrI*.

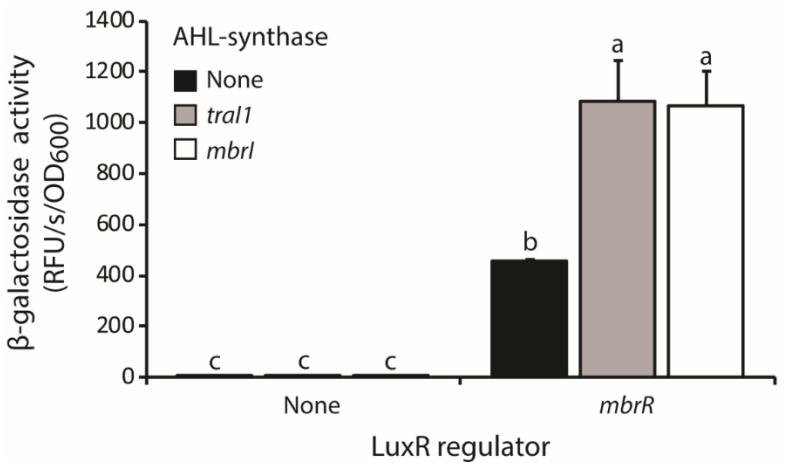


**Figure 6.8. Activation of MbrR by TraI1 and MbrI-derived AHLs.** A set of 6 R7ANS strains carrying pSDzmb, or pSDz-mb*mbrR*, in combination with pPR3, pPR3-*traI1* and pPR3-*mbrI,* were generated to monitor expression from plasmid borne *mbrI* promoter-*lacZ* fusions when *mbrR* was overexpressed alone, or in combination with *traI1* or *mbrI*. β-galactosidase assays were used to monitor expression. Values for relative fluorescent units (RFU)/s/$OD_{600}$ are the mean and SD (denoted by error bars) of 3 biological replicates. Mean values were compared using Fisher's LSD test controlling for type I error using the Bonferroni adjustment. Matching letters above bars indicate no significant difference between mean values.

## 6.3. Discussion

WSM1271 encodes three sets of LuxR-LuxI QS-loci, two of which were located on ICE*Mc*Sym[1271]. TraR1/TraR2-TraI, is encoded on region α of ICE*Mc*Sym[1271] and regulates assembly, excision and transfer of ICE*Mc*Sym[1271], and at least 187 chromosomally-encoded genes (Chapter 4). Both TraR1 and TraR2 appeared to activate the *traI1* promoter in the absence of *traI1*, but expression was greatly increased in the presence of *traI1*. This suggests that the TraR molecules may have weak capacity for transcriptional activation in the absence of AHLs, or that they can be weakly activated by AHLs produced from the conserved chromosomal QS locus. MbrR-MbrI were encoded within the melanin biosynthesis gene cluster of ICE*Mc*Sym[1271] region β, and overexpression of MbrR stimulated production of a melanin-like

pigment in WSM1271. MbrR was activated by either TraI1 or MbrI-derived AHLs and engaged in positive feedback regulation with *mbrI*. WSM1271 and 19 other *Mesorhizobium* sp. were found to encode an AHL inactivating enzyme AhlI homologous to the $Mn^{2+}$-dependent $C_4$-$C_{10}$ AHL lactonase AidH, similar to that described in *Ochrobactrum* sp. T63 (311). AhlI was shown to inactivate TraI1-derived AHLs in WSM1271, however the full range of AHLs inactivated by this protein was not explored. Nevertheless, it seems likely that QS regulation may be partially repressed in these strains.

Most LuxR family of proteins become functional only after interacting with AHLs (276, 289-292). Some LuxR proteins require the cognate AHL to remain stable (295, 296), whereas others can stably exist in the cell, but require a cognate AHL for biological activity (297). Even though *M. loti* R7ANS does not carry the typical ICE/$ICE^3$-AHL-synthase genes, overexpression of *traR1* or *mbrR* here partially induced expression from target promoters. R7ANS does carry the chromosomally-encoded AHL-synthase *mqsI* (167), therefore it was possible that TraR1 and MbrR may have been partially activated by non-cognate MqsI-derived AHLs. Such activation of LuxR proteins by promiscuous AHLs has been reported for LasR of *P. aeruginosa* (291), which is activated by $\geq$ 15 AHL species and for CepR of *Burkholderia cepacia*, which is activated by $C_4$-$C_{12}$ AHLs with or without 3-oxo constituents (315). Alternatively, the activity of TraR1 and MbrR in the absence of cognate AHLs may be a consequence of overexpressing *traR1* and *mbrR* from the IPTG inducible promoter of pSDz at artificially high levels. More fine-tuned control of expression for the LuxR-family regulators may be required to more accurately analyse the LuxR-autoinducer interactions.

Although TraR1 and TraR2 were shown to be activated by TraI1-derived AHLs and each induced $P_{traI1}$ when coexpressed in R7ANS with *traI1*, TraR2 failed to induce $P_{traI1}$ in WSM1271. It is possible that there may have been differences in the availability of ACP within the lipid pools of WSM1271 and R7ANS that may explain the differential function of TraR2 in these strains (286-288). However, it seemed more likely that the AHL-inactivating enzyme AhlI present in WSM1271 was responsible for this result. The CV026 bioassays performed in this chapter demonstrated that AhlI restricts the concentration of TraI1-derived AHLs in WSM1271, so perhaps TraR2 requires a higher concentration of TraI1-derived AHLs for its activation. TraR2 also exhibited lower levels of activity than TraR1 when overexpressed in the R7ANS background which lacks TraI1-derived AHLs. The AHL "quorum" concentration for activation of different LuxR receptors is known to vary between biosensor strains that respond to the same AHLs (316). Considering these observations, clarification of the interactions between the QS-sensor proteins and AHL-synthases encoded by WSM1271 will require the engineering of AHL-synthase free strains of WSM1271 and R7ANS that do not degrade AHLs. WSM1271Ω*traI1*Δ*ahlI* or R7ANS may serve as ideal parent strains in which to undertake this work.

Melanin is a common bacterial secondary metabolite produced by several plant-associative bacteria of the *Rhizobium* and *Sinorhizobium* genera (230, 306, 308, 314, 317, 318). Microbial melanin production is the result of the oxidative polymerisation of phenolic compounds by two main polyphenol multi-copper oxidases: a tyrosinase and a laccase (230, 314, 319, 320), however the exact biosynthetic pathway in rhizobia is yet to be functionally or

genetically characterised. Nevertheless. it is known that most rhizobia encode the melanogenic tyrosinase and (or) laccase protein on plasmids (230, 314, 317) and that in *R. leguminosarum* 8008, the tyrosinase is under the regulatory control of *nifA* (318, 321). *nifA*-dependent regulation of the 8008 tyrosinase implies that melanin biosynthesis may be induced during symbiosis, however, melanogenic mutant strains show no obvious defects in nodulation or $N_2$ fixation (322). In contrast to all other melanogenic rhizobia, the tyrosinase and laccase genes of WSM1271 are encoded on region β of ICE*Mc*Sym$^{1271}$, and melanin production in WSM1271 is induced by the QS-regulator MbrR. To our knowledge, WSM1271 is the first bacterium in which QS-regulation of melanin production has been observed, however, melanin production is controlled by QS in the dual-lifestyle pathogenic yeast *Cryptococcus neoformans* that causes severe central nervous system infections in immune-compromised humans (323). Melanisation in *C. neoformans* acts as a major virulence factor allowing the yeast to thrive within the host at high cell-density (305, 307, 323). Evidence has been presented suggesting that melanisation may also protect free-living *C. neoformans* cells from ultraviolet light (324), temperature fluctuations (325), heavy metal toxicity (326) and cell wall-degrading enzymes such as those that may be produced by fungal predators (327). Such biological functions may explain why the melanin biosynthesis genes have persisted in WSM1271.

The genes involved in QS-regulation of melanin biosynthesis in *C. neoformans* are yet to be elucidated. In WSM1271, it is also unclear which genes are activated by MbrR to induce the production of melanin, however it was here demonstrated that MbrR activates expression from the *mbrI* promoter. A

hypothetical protein, a multi-copper oxidase and a blue-copper domain containing protein are encoded in close proximity downstream of *mbrI* (Fig 6.1B). Thus, it is possible that these genes may be transcribed as a polycistronic mRNA under MbrR-dependent QS induction. The tyrosinase gene is encoded upstream of *mbrI* within the melanin biosynthesis gene cluster and visual inspection of this revealed that there is no inverted repeat resembling that present in the *mbrI* promoter region (Fig 6.7A). Therefore, if this gene is regulated via QS, then this regulation may be indirect.

WSM1271 and a large cohort of related *Mesorhizobium* sp. encode a homologue of the *Ochrobactrum.* sp. T63 α/β-fold family C4-C10 AHL-lactonase AidH (311, 313) on their chromosomes that was shown to inactivate TraI1-derived AHLs. Interestingly, a $C_4$-$C_{14}$ AHL-acylase termed AiiO was recently discovered in *O.* sp. A44 (328) that is 95.9% identical to AidH (Fig 6.4) and ~80% identical to AhlI in WSM1271, WSM2073 and WSM2075. Therefore, it is difficult to speculate as to whether AhlI comprises an AHL-lactonase or acylase. Given that both AidH and AiiO inactivate a broad range of AHLs (311, 313, 328, 329), it is possible that this may also be true of AhlI. AhlI may even inactivate all AHLs produced by TraI1, MbrI and MqsI in WSM1271 supressing or partially supressing QS in this strain.

It seems counterintuitive that WSM1271 and other QS bacteria would benefit from inactivating endogenously synthesised AHLs. However, in *A. tumefaciens* spp. it has been proposed that inactivation of AHLs by the AHL-lactonase *blcC* (formerly termed *attM*) may be an important component of genetic regulation, allowing for rapid exit from the QS-dependent pTi conjugal transfer state (330, 331). If the primary role of BlcC or AhlI-mediated AHL-

inactivation was to attenuate QS-induced transfer of pTi or symbiosis ICEs, one might expect that the *blcC* and *ahlI* genes would have evolved to be transcribed in response to increasing concentrations of AHLs. Although *blcC* expression is strongly induced in the stationary phase of growth, this is not a response to the accumulation of AHLs (330, 332, 333). The RNAseq experiments presented in Chapter 4 also revealed that *ahlI* was not differentially expressed in QS+ cells of WSM1271 overexpressing *traI1* and *traR1*, relative to the QS- cells (fold change = 1.067 $\pm$ 1.18). Therefore, like *blcC,* transcription of *ahlI* is probably not induced in response to increasing concentrations of AHLs. Indeed, it was recently demonstrated that *A. tumefaciens* mutants for the AHL-lactonase *blcC* exhibit wild-type frequencies of Ti plasmid conjugative transfer, even in stationary phase, suggesting that *blcC* may have an alternative primary role (332). *blcC* was subsequently shown to comprise the third gene encoded in the *blcABC* operon that functions in the catabolism of γ-butyrolactone to succinate, which may be fully oxidisied in the tricarboxylic acid cycle to produce energy (332, 334). Alternatively, to resource scavenging or exogenous genetic regulation, AhlI may function in WSM1271 to attenuate QS induction of genes in neighbouring organisms. This has been proposed as the role for the closely related AHL-acylase AiiO encoded by *O.* sp. A44 (328), which has been shown to attenuate the QS-dependent maceration of potato tissue by the bacterial pathogen *Pectobacterium carotovorum* (329). Further characterisation of AhlI will be crucial for the determination of its biological function and may provide useful insight regarding attenuation of QS-regulation in pathogenic bacteria.

The data presented in this chapter illustrate the complexity of QS-regulation in bacteria. WSM1271 encodes three QS systems; MbrR-MbrI, which controls melanin biosynthesis; TraR1/2-TraI1, which controls ICE$^{3}$ assembly, excision, transfer and the expression of at least 187 chromosomal genes; and lastly MqsR-MqsI, for which the biological function is unknown. Although TraR1 and TraR2 regulate the same biological process, these regulators probably require different concentrations of TraI1-derived AHLs for their activation. Within single cells, there may be competition between AHLs for binding various TraR proteins, and there may even be crosstalk between QS systems. In WSM1271 and other mesorhizobia listed in the chapter, QS may even be partially supressed by the AHL-inactivating enzyme AhlI. Because ICE$^{3}$ excision and transfer are regulated by QS, these above factors may have a profound influence on the dynamics of ICE transfer in different environments.

# Chapter 7.

# Concluding discussion

## 7.1. ICE$^{3}$ assembly, excision, integration and dissassembly

ICEs have traditionally been regarded as comprising single regions of contiguous DNA integrated within bacterial genomes, capable of excision and horizontal transfer via conjugation (77, 78). In this thesis, a unique family of symbiosis ICEs, termed ICE$^{3}$s, were identified in 13 diverse *Mesorhizobium* spp., existing as three entirely separated chromosomal DNA regions (α, β and γ). Detailed analysis of the first identified ICE$^{3}$ ICE*Mc*Sym$^{1271}$ of *Mesorhizobium ciceri* bv. biserrulae WSM1271 revealed that these three regions do not excise independently, but rather, recombine in the host chromosome to form a single contiguous DNA element prior to excision and conjugative transfer. Following transfer, ICE*Mc*Sym$^{1271}$ integrates within a recipient chromosome at one of three insertion locations and reconfigures the chromosome to disassemble back into the tripartite configuration. Acquisition of ICE*Mc*Sym$^{1271}$ conveys upon recipient's nodulation proficiency with the legume-host of the donor strain *B. pelecinus*, however, $N_2$ fixation effectiveness is commonly impaired. Given the structural similarity of the 13 identified ICE$^{3}$s, it is highly probable that each shares the same mechanism of assembly, excision, integration and transfer.

A model for the mechanism of assembly, excision, integration and dissasembly of ICE*Mc*Sym$^{1271}$ was here proposed. The three regions of ICE*Mc*Sym$^{1271}$ collectively carry three distinct *attL* and *attR* sites at their termini and encode three associated Int proteins. The arrangement and orientation of the three pairs of *attL* and *attR* sites across ICE*Mc*Sym$^{1271}$ regions α, β and γ is fundamental for the assembly mechanism prior to excision. Overall, the complete assembly and excision of ICE*Mc*Sym$^{1271}$ prior

to conjugative transfer requires the concerted action of IntS, IntG and IntM acting on their associated *attL* and *attR* sites in any sequential order, except for those that result in non-viable segregation of the chromosome. Following conjugative transfer of ICE*Mc*Sym$^{1271}$, IntS, IntG and IntM catalyse integration of the ICE$^{3}$ into *attB* sites nested in the 3'-ends of the chromosomal *phe*-tRNA, *guaA,* or *met*-tRNA genes*,* respectively, and the concerted action of the three Int proteins reverses the assembly and excision process dispersing the ICE$^{3}$ back into the tripartite configuration.

Although ICE*Mc*Sym$^{1271}$ is the only element discovered which obligatorily requires chromosomal inversions to facilitate excision and transfer, other integrative elements have been found to harbour multiple sets of *att* sites capable of site-specific inversion (335). Given the diversity and abundance of ICEs in bacterial genomes (116), it is plausible that "multipartite" elements resembling the structure of ICE*Mc*Sym$^{1271}$ have been overlooked in other organisms. It may even be that many presumed immobile genetic elements identified in diverse organisms could in-fact be mobile. The discovery and characterisation of ICE$^{3}$s provides the foundation for the discovery of increasingly diverse and complex multipartite ICEs in other bacterial species and genera.

## 7.2. Regulation of ICE$^{3}$ excision and transfer

Following the discovery of ICE*Mc*Sym$^{1271}$ and the elucidation of the mechanism of excision and integration, ICE$^{3}$s were identified in a total of 13 genetically diverse *Mesorhizobium* spp. originally isolated from various

geographical locations. ICE$^{3}$-α regions of diverse mesorhizobia carry genes required for symbiosis, vitamin biosynthesis, and all the genes necessary for QS-induction of *rdfS* and ICE*Ml*Sym$^{R7A}$ excision and transfer in *M. loti* R7A (74-76, 130, 167, 170, 176). It was proposed that ICE$^{3}$s may have evolved from single-part symbiosis ICEs following recombination between two other integrative elements in an ancestral bacterium.

Considering the common evolutionary history of ICE and ICE$^{3}$, it is not surprising that ICE$^{3}$ excision and transfer is regulated via QS. ICE*Mc*Sym$^{1271}$ encoded dual copies of the LuxR-family transcriptional regulator TraR (termed TraR1 and TraR2). Overexpression of *traR1* or *traR2* in WSM1271 activated expression from the *traI1* promoter, but maximum expression from this promoter was only achieved in the presence of the AHL-synthase gene *traI1,* suggesting that TraI1 produced the autoinducer of TraR1 and TraR2. This was consistent with results observed for R7A TraR (167). Induction of QS in WSM1271 by overexpression of *traR1* and *traI1* stimulated the transcription of genes involved in ICE*Ml*Sym$^{R7A}$ excision (*rdfS, intS* and *msi172-msi171*) and conjugative transfer (*traF, msi107, rlxS, traG* and *msi031-trbBCDEJLFGI-msi021*), and stimulated ICE*Mc*Sym$^{1271}$ assembly, excision and conjugative transfer ~10-100-fold. The cascade of genetic regulation leading to the activation of *rdfS* was shown to closely resembled that described for *M. loti* R7A (74, 75, 167, 170).

In contrast to single-part symbiosis ICEs that encode a single RDF (75), ICE*Mc*Sym$^{1271}$ encodes three RDFs RdfS, RdfG and RdfM, which were shown to be required for the excisive IntS, IntG and IntM-mediated recombination reactions, respectively. Transcription of *rdfG* and *rdfM* is dependent on RdfS.

QS activation of dual function RDF and transcriptional activator *rdfS* suggests a hierarchical order of expression for the three RDFs. *rdfS* is expressed first in the hierarchy, and therefore likely catalyses excisive IntS-mediated recombination prior to the occurrence of excisive IntG and IntM-mediated recombination, avoiding the formation the inviable chromosomal state (Fig 4.1A. *state iii*). These combined data demonstrated that *rdfM* is always the lowest expressed RDF gene, and that excisive IntM-mediated recombination is always ~10-fold less prevalent than the other Int-mediated excision reactions. Thus, excisive IntM-mediated recombination is probably the final recombination reaction to occur following QS-induction of ICE*Mc*Sym$^{1271}$ assembly and excision, and the entirely assembly/excision of ICE*Mc*Sym$^{1271}$ probrably occurs via the sequence of excisive Int-mediated recombination reactions IntS -> IntG -> IntM.

Interestingly, ICE*Mc*Sym$^{1271}$ encoded a second QS-system *mbrR-mbrI* that was absent from all other identified ICE$^{3}$s. *mbrR* was shown to activate *mbrI* expression in a *mbrI*-dependent manner, and also stimulated the production of a melanin-like pigment. However, the *mbrR-mbrI* QS-system had little effect on the QS-regulation of ICE*Mc*Sym$^{1271}$ excision. A third QS-system *mqsR-mqsI* was found to be encoded on the chromosomes of all analysed mesorhizobial genomes, however, it is currently unclear as whether this QS-system imparts an influence on ICE or ICE$^{3}$ excision and transfer. Although the biological role of the conserved chromosomal QS-loci *mqsRI* is yet to be discerned, it has been demonstrated that MqsI catalyses the synthesis of $C_{12}$ AHLs when its gene is overexpressed in *E. coli*, and DSM 2626 *mqsI* mutants fail to produce $C_{12}$ AHLs (304). Moreover, both *mqsR* and *mqsI* are required

to induce expression from a plasmid borne *mqsI* promoter in *M. loti* DSM 2626, indicating that MqsR is probably activated by MqsI-derived AHLs, and engages in positive feedback regulation of *mqsI* (304). Interestingly, comparison of the *mqsRI* regions of diverse *Mesorhizobium* spp. revealed the presence of a conserved crotonase-family gene *mqsC* that exists downstream of *mqsI* (Fig 5.1C). Although the function of this gene has not been elucidated, a bi-functional crotonase homologue Bcam0581 has been shown to catalyse biosynthesis of the *B. cenocepacia* QS DSF *cis*-2-dodecenoic that controls virulence in this bacterium (336-338). Bcam0581 exhibits both dehydratase activity, introducing a double bond at the $C_2$ position of the fatty acid intermediate substrate molecule hydroxy-dodecanoyl-ACP, and thioesterase activity, cleaving the thioester bond of this molecule to leave the free unsaturated fatty acid (336). A similar mechanism of action has been reported for the *X. campestris* DSF synthase RpfF (339). Interestingly, the marine *Mesorhizobium* sp. R8-Ret-T53-13d carries the conserved *mqsRIC* loci and produces two novel DSF-like unsaturated AHL signalling molecules 5-*cis*-3-oxo-$C_{12}$-HSL and 5-*cis*-$C_{12}$-HSL which activate *V. fischeri* LuxR and *P. aeruginosa* LasR regulators in *E. coli* bioassays (340). MqsC in *Mesorhizobium* spp. may therefore function to introduce a double bond at the $C_5$ position of $C_{12}$ AHLs produced by MqsI. Modification of AHLs in *Mesorhizobium* is an area which requires further further exploration, and could reveal new insight into the QS-regulation of ICE and $ICE^3$ excision and transfer.

In this thesis, it was discovered that WSM1271 and other ICE harbouring mesorhizobia encode an AHL inactivating enzyme AhlI, homologous to the

$Mn^{2+}$-dependent $C_4$-$C_{10}$ AHL lactonase AidH encoded by *Ochrobactrum* sp. T63 (311), and the $C_4$-$C_{14}$ AHL-acylase AiiO encoded by *O.* sp. A44 (328) that degrades TraI1-derived AHLs. It is not yet clear as to the range of AHLs inactivated by AhlI, or the biological function of endogenous AHL-inactivation mesorhizobia, however, it is likey that QS-regulation may be partially supressed in these strains. This may influence the dynamics of ICE and $ICE^3$ transfer in this genus.

The elucidation of $ICE^3$ recombination and regulation of assembly/excision highlights the complex nature of $ICE^3$s. Considering that $ICE^3$s naturally assembly into a single contiguous element prior to conjugative transfer, it is not entirely clear as to why $ICE^3$s have not simply reverted into single-part ICEs. The observation that ICE*Mc*Sym[181] has maintainted its tripartite configuration following the replacement of a recombinase and associated *att* site indicated that there may be some selective advantages associated with this tripartite form. Four possible selective advantages associated with $ICE^3$ configuration were proposed;

a) By being able to integrate into three distinct *attB* sites, $ICE^3$s maximize their potential for host-integration and potentially broaden their host range.

b) in the tripartite configuration, the $ICE^3$ is more resistant to loss following spurious recombinase-mediated recombination events, because it requires three recombination events to excise, rather than one.

c) Incoming ICE/IME that integrate in tandem at any of the $ICE^3$ *att* sites cannot alone stimulate excision of the $ICE^3$, thus $ICE^3$s likely avoid the destabilization associated with formation of tandem ICE/IME arrays.

d) By occupying three *attB* sites in the fully integrated form, ICE$^3$s likely have an increased propensity to acquire and accumulate genes from invading ICEs/IMEs that target the same sites.

The benefits described would likely be most advantageous in an environment where ICE, IME and other integrative elements are abundant and there is fierce competition for a limited number of *attB* integration sites. If there are similar complex multipartite ICEs present in the genomes of bacteria other than mesorhizobia, then they may be most prevalent in genomes with these characteristics.

## 7.3. Consequences for ICE and ICE$^3$ transfer in agriculture

Acquisition of a symbiosis ICE may convey upon recipient's proficiency for nodulation. For example, transfer of symbiosis ICE$^3$s from *M. loti* NZP2037, NZP2042 and SU343 converted R7ANS to a *L. pedunculatus* nodulating strain, and transfer of ICE*Mc*Sym$^{1271}$ from WSM1271, WSM2073 or WSM2075 converted R7ANS to *B. pelecinus* nodulating strain. However, acquisition of a symbiosis ICE$^3$ may be insufficient to convert recipients into effective $N_2$-fixing symbionts. In glasshouse trials, all R7ANS ICE*Mc*Sym$^{1271}$ recipients fixed $N_2$ partially effectively relative to WSM1271, regardless of whether the ICE was donated from the effective $N_2$-fixing strain WSM1271, the partially effective $N_2$-fixing strains WSM2073, or non-$N_2$-fixing strain WSM2075. Other genetic factors are likely important in determining this outome. For example, genes essential to symbiotic $N_2$ fixation may be harboured on the chromosome of WSM1271, which are absent from the chromosomes of WSM2073 or

WSM2075. Alternatively, there may be mis-regulation or inadequate expression patterns of ICE-encoded $N_2$-fixation genes in these new genetic backgrounds. Introduction of new gene clusters into cells from diverse genetic sources often results in poor gene expression and the disruption of existing metabolic pathways in the recipient (341-343).

Based on the data presented in this thesis, it seems highly probable that the genetically diverse, sub-optimal $N_2$-fixing rhizobia found to occupy *B. pelecinus* nodules in Western Australian field sites inoculated with WSM1271 or WSM1497 (193, 194) may comprise symbiosis ICE$^3$ recipients. It could also be speculated that these newly evolved symbiotic strains frequently develop into stable persistent soil populations because they are well-adapted to the soil in which they exist, possessing unique chromosomal genes conferring beneficial adaptations which the inoculant strain may lack (344). This raises key questions regarding the consequences for agriculture. Pointedly, how prevalent is ICE/ICE$^3$ transfer; what are the impacts on biological $N_2$ fixation by *B. pelecinus* in the field and does this influence agricultural productivity in these systems?

The prevalence of symbiosis ICE transfer at *B. pelecinus* field sites is likely a function of the abundance of potential ICE/ICE$^3$ recipients in the rhizosphere. Little is known regarding the identity of potential ICE/ICE$^3$ recipients. As is the case for most HGT events, symbiosis ICE transfer and integration from donor cell to recipient seems most likely to occur between closely related species that share a common genetic framework (59, 345, 346). However, even closely related rhizobia that carry a compatible *attB* sequence for ICE/ICE$^3$ integration may not be capable of aquiring a symbiosis ICE (59). In the *Vibrio, Bacillus,*

*Streptomyces* and *Streptococcus* genera, a range of barriers to the horizontal acquisition of ICEs and other MGEs have been idenfied. These include DNA restriction-modification or CRISPR systems in the recipient genome which may digest the incoming ICE prior to integration (347-350); ICE incompatibility or exclusion determinants which may prevent acquisition of a closely related element (139, 351, 352); or the inability of a recipient cell to express ICE recombination genes to allow integration of the element (341). Considering these barriers to ICE/ICE$^3$ acquisition, it seems most probable that ICE/ICE$^3$ recipients are commonly non-symbiotic mesrhizobia, or perhaps mesorhizobia that carry symbiosis plasmids rather than ICEs. Non-symbiotic mesrhizobia were previously isolated from the rhizosphere of a *L. corniculatus* stand in New Zealand (72), however, it is unclear as to their abundance in the soil at these sites. Identifying precisely what comprises a potential ICE/ICE$^3$ recipient, and how abundant these strains are in the rhizosphere of field sites, will be essential to producing estimations of the prevalence of symbiosis ICE/ICE$^3$ transfer in agricultural systems.

It could be speculated that even in a field site where ICE/ICE$^3$ recipients are abundant in the soil, and ICE/ICE$^3$ transfer is prevalent, inputs of fixed $N_2$ may not be significantly altered if newly evolved symbionts fix $N_2$ with a similar efficacy to the inoculant strain. Of 59 genetically diverse strains that putatively acquired symbiosis ICE$^3$s from WSM1271 or WSM1497 tested for $N_2$-fixation, none fixed $N_2$ with equal or improved effectiveness relative to the original inoculant strain (194). However, these putative symbiosis ICE$^3$ recipient strains were isolated from *B. pelecinus* nodules over a relatively small geographical range. Thus, it is not clear as to the true proportion of sub-optimal

$N_2$-fixing strains that emerge following ICE$^3$ transfer. In a more recent study, it was speculated that the symbiosis ICE carried by the Australian commercial inoculant for *Cicer arietinum* (chickpea) inoculant *M. ciceri* CC1192 had transferred to indigenous soil rhizobia, converting them to chick-pea nodulating strains (353). Approximately 53% of nodules sampled over a small geographical range in New South Wales were found to be occupied by rhizobia other than CC1192 between 1 and 10 years after inoculation (353). Interestingly, of 29 strains tested for $N_2$-fixation effectiveness with *C. arietinum*, approximately half fixed $N_2$ with equal effectiveness relative to the inoculant CC1192, while the other half were less effective. In this study, there was no evidence that the novel mesorhizobia in the soils compromised $N_2$ fixation or *C. arietinum* productivity (353). More comprehensive nodule sampling experiments will be necessary to gauge the proportion of suboptimal $N_2$-fixing strains that arise following ICE/ICE$^3$ transfer at both *C. arietinum* and *B. pelecinus* field sites.

If ICE$^3$ transfer poses a barrier for inoculation success in the future, the work performed in this thesis elucidating the mechanism of ICE$^3$ assembly, excision and transfer will provide multiple genetic targets for engineering inoculant strains in which symbiosis ICE$^3$ transfer has been neutered. For example, even in the rare event that the inoculant acquires an exogenous copy of the cognate relaxase on an invading MGE. Alternatively, transfer could be managed by targeting repression of the QS-system that controls ICE$^3$ assembly, excision and transfer. This could, for example, be achieved by co-inoculating field sites with *B. pelecinus* symbionts paired with quorum-quenching bacteria.

## 7.4. Conclusion

In this thesis, it has been demonstrated that a diversity of *Mesorhizobium* spp. isolated from various geographical locations carry an entirely novel “tripartite” symbiosis ICE composed of three co-transferrable regions that convey upon recipients the ability to nodulate and fix $N_2$ with target legumes. An elaborate mechanism for ICE$^3$ assembly, excision, integration, disassembly and regulation of these processes was elucidated. In the field, transfer of ICE$^3$s commonly results in the emergence of sub-optmimal $N_2$-fixing strains that may compete with the original inoculant for legume nodulation. However, the prevalence of transfer, proportion of suboptimal $N_2$-fixing strains emerged, and the overall effects on agricultural productivity are poorly understood. The data presented in this thesis now provides a crucial framework to further explore these questions and develop effective management strategies for agriculture

# Bibliography

1. Brewer RJM, Haskett TL, Ramsay JP, O'Hara GW, & Terpolilli JJ (2017) Complete genome sequence of *Mesorhizobium ciceri* bv. biserrulae WSM1497, an efficient nitrogen-fixing microsymbiont of the forage legume *Biserrula pelecinus*. *Genome Announcements* 5(35):e00902-00917.

2. Haskett T*, et al.* (2016) Complete genome sequence of *Mesorhizobium ciceri* bv. biserrulae Strain WSM1284, an efficient nitrogen-fixing microsymbiont of the pasture legume *Biserrula pelecinus*. *Genome Announcements* 4(3):e00514-00516.

3. Haskett TL*, et al.* (2017) Evolutionary persistence of tripartite integrative and conjugative elements. *Plasmid* 92:30-36.

4. Haskett TL*, et al.* (2016) Assembly and transfer of tripartite integrative and conjugative genetic elements. *PNAS* 113(43):12268-12273.

5. Oldroyd GE, Murray JD, Poole PS, & Downie JA (2011) The rules of engagement in the legume–rhizobial symbiosis. *Annual Review of Genetics* 45:119-144.

6. Jones KM, Kobayashi H, Davies BW, Taga ME, & Walker GC (2007) How rhizobial symbionts invade plants: the *Sinorhizobium–Medicago* model. *Nat. Rev. Microbiol.* 5:619.

7. Murray JD (2011) Invasion by invitation: rhizobial infection in legumes. *Molecular Plant-Microbe Interactions* 24(6):631-639.

8. Alunni B & Gourion B (2016) Terminal bacteroid differentiation in the legume-rhizobium symbiosis: nodule-specific cysteine-rich peptides and beyond. *New Phytolologist* 211(2):411-417.

9. Udvardi M & Poole PS (2013) Transport and metabolism in legume-rhizobia symbioses. *Annual review of plant biology* 64:781-805.

10. Gagnon H & Ibrahim RK (1998) Aldonic acids: A novel family of nod gene inducers of *Mesorhizobium loti, Rhizobium lupini,* and *Sinorhizobium meliloti*. *Molecular plant-microbe interactions : MPMI* 11:988-998.

11. Maxwell CA, Hartwig UA, Joseph CM, & Phillips DA (1989) A chalcone and two related flavonoids released from alfalfa roots induce *nod* genes of *Rhizobium meliloti. Plant Physiology* 91(3):842-847.

12. Dakora FD, Joseph CM, & Phillips DA (1993) Alfalfa (*Medicago sativa* L.) root exudates contain isoflavonoids in the presence of *Rhizobium meliloti*. *Plant Physiology* 101(3):819-824.

13. Kamboj DV, Bhatia R, Pathak DV, & Sharma PK (2010) Role of *nodD* gene product and flavonoid interactions in induction of nodulation genes in *Mesorhizobium ciceri*. *Physiology and Molecular Biology of Plants* 16(1):69-77.

14. Liu C-W & Murray JD (2016) The role of flavonoids in nodulation host-range specificity: an update. *Plants* 5(3):33.

15. D'Haeze W & Holsters M (2002) Nod factor structures, responses, and perception during initiation of nodule development. *Glycobiology* 12(6):79R-105R.

16. Mergaert P, Van Montagu M, & Holsters M (1997) Molecular mechanisms of Nod factor diversity. *Mol. Microbiol.* 25(5):811-817.

17. Geurts R & Bisseling T (2002) *Rhizobium* nod factor perception and signalling. *The Plant Cell* 14(Suppl):s239-s249.

18. Wais RJ, Keating DH, & Long SR (2002) Structure-function analysis of Nod factor-induced root hair calcium spiking in *Rhizobium*-legume symbiosis. *Plant physiology* 129(1):211-224.

19. Via VD, Zanetti ME, & Blanco F (2016) How legumes recognize rhizobia. *Plant Signaling & Behavior* 11(2):e1120396.

20. Wang D, Shengming Y, Tang F, & Zhu H (2012) Symbiosis specificity in the legume – rhizobial mutualism. *Cellular Microbiology* 14(3):334-342.

21. Fournier J*, et al.* (2008) Mechanism of infection thread elongation in root hairs of *Medicago truncatula* and dynamic interplay with associated rhizobial colonization. *Plant physiology* 148(4):1985-1995.

22. Gallon JR (The oxygen sensitivity of nitrogenase: a problem for biochemists and micro-organisms. *Trends in Biochemical Sciences* 6:19-23.

23. Tsoy OV, Ravcheev DA, Čuklina J, & Gelfand MS (2016) Nitrogen fixation and molecular oxygen: comparative genomic reconstruction of transcription regulation in alphaproteobacteria. *Front Microbiol.* 7:1343.

24. Fischer HM (1994) Genetic regulation of nitrogen fixation in rhizobia. *Microbiology Reviews* 58(3):352-386.

25. Foussard M*, et al.* (1998) Regulation of Nitrogen Fixation Gene Expression in Rhizobia: An Overview. *Biological Nitrogen Fixation for the 21st Century: Proceedings of the 11th International Congress on Nitrogen Fixation, Institut Pasteur, Paris, France, July 20–25, 1997*, eds Elmerich C, Kondorosi A, & Newton WE (Springer Netherlands, Dordrecht), pp 101-106.

26. Sciotti MA, Chanfon A, Hennecke H, & Fischer HM (2003) Disparate oxygen responsiveness of two regulatory cascades that control expression of symbiotic genes in *Bradyrhizobium japonicum*. *J. Bacteriol.* 185(18):5639-5642.

27. Terpolilli JJ, Hood GA, & Poole PS (2012) What determines the efficiency of N(2)-fixing *Rhizobium*-legume symbioses? *Advances in microbial physiology* 60:325-389.

28. Terpolilli JJ, O'Hara GW, Tiwari RP, Dilworth MJ, & Howieson JG (2008) The model legume *Medicago truncatula* A17 is poorly matched for N2 fixation with the sequenced microsymbiont *Sinorhizobium meliloti* 1021. *New Phytologist* 179(1):62-66.

29. Howieson JG, Malden, J., Yates, R.J. and O'Hara, G.W. (2000) Techniques for the selection and development of elite inoculant strains of *Rhizobium leguminosarum* in southern Australia. *Symbiosis* 28(1):33-48.

30. Howieson JG, Nutt B, & Evans P (2000) Estimation of host-strain compatibility for symbiotic N-fixation between *Rhizobium meliloti*, several annual species of *Medicago* and *Medicago sativa*. *Plant and Soil* 219(1):49-55.

31. Oldroyd GED (2013) Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol.* 11:252.

32. Sim SH*, et al.* (2008) The core and accessory genomes of *Burkholderia pseudomallei:* implications for human melioidosis. *PLoS Pathogens* 4(10):e1000178.

33. Sugawara M*, et al.* (2013) Comparative genomics of the core and accessory genomes of 48 *Sinorhizobium* strains comprising five genospecies. *Genome Biol.* 14(2):R17.

34. Tettelin H*, et al.* (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *PNAS* 102(39):13950-13955.

35. Segerman B (2012) The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. *Frontiers in Cellular and Infection Microbiology* 2(116).

36. Ochman H & Moran NA (2001) Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* 292(5519):1096-1099.

37. Lemaire B*, et al.* (2015) Recombination and horizontal transfer of nodulation and ACC deaminase (*acdS*) genes within Alpha- and Betaproteobacteria nodulating legumes of the Cape Fynbos biome. *FEMS Microbiology and Ecolology* 91(11).

38. Zheng JZ*, et al.* (2017) The structure and evolution of beta-rhizobial symbiotic genes deduced from their complete genomes. *Immunome Research* 13(2).

39. Gonzalez V*, et al.* (2003) The mosaic structure of the symbiotic plasmid of *Rhizobium etli* CFN42 and its relation to other symbiotic genome compartments. *Genome Biol.* 4(6):R36.

40. Laguerre G*, et al.* (2001) Classification of rhizobia based on *nodC* and *nifH* gene analysis reveals a close phylogenetic relationship among *Phaseolus vulgaris* symbionts. *Microbiology* 147(Pt 4):981-993.

41. Masson-Boivin C, Giraud E, Perret X, & Batut J (2009) Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? *Trends in Microbiology* 17(10):458-466.

42. Werner GD, Cornwell WK, Sprent JI, Kattge J, & Kiers ET (2014) A single evolutionary innovation drives the deep evolution of symbiotic N2-fixation in angiosperms. *Nature communications* 5:4087.

43. Peix A, Ramírez-Bahena MH, Velázquez E, & Bedmar EJ (2015) Bacterial associations with legumes. *Critical Reviews in Plant Sciences* 34(1-3):17-42.

44. Pawlowski K & Demchenko KN (2012) The diversity of actinorhizal symbiosis. *Protoplasma* 249(4):967-979.

45. Gnat S*, et al.* (2015) Phylogeny of symbiotic genes and the symbiotic properties of rhizobia specific to *Astragalus glycyphyllos* L. *PLoS ONE* 10(10):e0141504.

46. Barcellos FG, Menna P, da Silva Batista JS, & Hungria M (2007) Evidence of horizontal transfer of symbiotic genes from a *Bradyrhizobium japonicum* inoculant strain to indigenous diazotrophs *Sinorhizobium* (*Ensifer*) *fredii* and *Bradyrhizobium elkanii* in a Brazilian savannah soil. *Applied and environmental microbiology* 73(8):2635-2643.

47. Menna P & Hungria M (2011) Phylogeny of nodulation and nitrogen-fixation genes in *Bradyrhizobium*: supporting evidence for the theory of monophyletic origin, and spread and maintenance by both horizontal and vertical transfer. *International Journal of Systematic and Evololutionary Microbiology* 61(Pt 12):3052-3067.

48. Bailly X, Olivieri I, Brunel B, Cleyet-Marel J-C, & Béna G (2007) Horizontal gene transfer and homologous recombination drive the evolution of the nitrogen-fixing symbionts of *Medicago* Species. *J. Bacteriol.* 189(14):5223-5236.

49. Laranjo M, Alexandre A, & Oliveira S (2014) Legume growth-promoting rhizobia: an overview on the *Mesorhizobium* genus. *Microbiological Research* 169(1):2-17.

50. Remigi P, Zhu J, Young JPW, & Masson-Boivin C (2016) Symbiosis within symbiosis: evolving nitrogen-fixing legume symbionts. *Trends in Microbiology* 24(1):63-75.

51. MacLean AM, Finan TM, & Sadowsky MJ (2007) Genomes of the symbiotic nitrogen-fixing bacteria of legumes. *Plant physiology* 144(2):615-622.

52. Barnett MJ*, et al.* (2001) Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. *PNAS* 98(17):9883-9888.

53. Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EPC, & de la Cruz F (2010) Mobility of plasmids. *Microbiology and Molecular Biology Reviews* 74(3):434-452.

54. Rao JR, Fenton M, & Jarvis BDW (1994) Symbiotic plasmid transfer in *Rhizobium leguminosarum* biovar *trifolii* and competition between the inoculant strain ICMP2163 and transconjugant soil bacteria. *Soil Biology and Biochemistry* 26(3):339-351.

55. Kondorosi A, Kondorosi E, Pankhurst CE, Broughton WJ, & Banfalvi Z (1982) Mobilization of a *Rhizobium meliloti* megaplasmid carrying nodulation and nitrogen fixation genes into other rhizobia and *Agrobacterium*. *Molecular and General Genetics* 188(3):433-439.

56. Pankhurst CE, Broughton WJ, & Wieneke U (1983) Transfer of an indigenous plasmid of *Rhizobium loti* to other rhizobia and *Agrobacterium tumefaciens*. *J Gen Microbiol* 129(8):2535-2543.

57. Marchetti M*, et al.* (2010) Experimental evolution of a plant pathogen into a legume symbiont. *PLoS Biology* 8(1):e1000280.

58. Hirsch AM*, et al.* (1984) *Rhizobium meliloti* nodulation genes allow *Agrobacterium tumefaciens* and *Escherichia coli* to form pseudonodules on alfalfa. *J. Bacteriol.* 158(3):1133-1143.

59. Ling J*, et al.* (2016) Plant nodulation inducers enhance horizontal gene transfer of *Azorhizobium caulinodans* symbiosis island. *PNAS* 113(48):13875-13880.

60. Sullivan JT & Ronson CW (1998) Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a *phe*-tRNA gene. *PNAS* 95(9):5145-5149.

61. Kaneko T*, et al.* (2011) Complete genome sequence of the soybean symbiont *Bradyrhizobium japonicum* strain USDA6(T). *Genes* 2(4):763-787.

62. Kaneko T*, et al.* (2002) Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Research* 9(6):189-197.

63. Haskett T*, et al.* (2016) Complete genome sequence of *Mesorhizobium ciceri* strain CC1192, an efficient nitrogen-fixing microsymbiont of *Cicer arietinum*. *Genome Announcements* 4(3).

64. Jarvis BDW*, et al.* (1997) Transfer of *Rhizobium loti*, *Rhizobium huakuii*, *Rhizobium ciceri*, *Rhizobium mediterraneum*, and *Rhizobium tianshanense* to *Mesorhizobium* gen. nov. *International journal of systematic and evolutionary microbiology* 47(3):895-898.

65. Rogel MA, Ormeno-Orrillo E, & Martinez Romero E (2011) Symbiovars in rhizobia reflect bacterial adaptation to legumes. *Systematic And Applied Microbiology* 34(2):96-104.

66. Chen WX, Wang ET, & Kuykendall LD (2005) The Proteobacteria. *Bergeys manual of systematic bacteriology*, (Springer, New York), Vol 2, pp 403-408.

67. Pankhurst CE, Macdonald PE, & Reeves JM (1986) Enhanced nitrogen fixation and competitiveness for nodulation of *Lotus pedunculatus* by a plasmid-cured derivative of *Rhizobium loti*. *Microbiology* 132(8):2321-2328.

68. Pankhurst CE, Hopcroft DH, & Jones WT (1987) Comparative morphology and flavolan content of *Rhizobium loti* induced effective and ineffective root nodules on *Lotus* species, *Leuceana leucocephala, Carmichaelia flagelliformis, Ornithopus sativus*, and *Clianthus puniceus*. *Canadian Journal of Botany* 65(12):2676-2685.

69. Chua KY*, et al.* (1985) Isolation and characterization of transposon *Tn*5-induced symbiotic mutants of *Rhizobium loti*. *J. Bacteriol.* 162(1):335-343.

70. Pueppke SG & Broughton WJ (1999) *Rhizobium* sp. strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges. *Molecular Plant-Microbe Interactions* 12(4):293-318.

71. Nandasena KG*, et al.* (2004) Symbiotic relationships and root nodule ultrastructure of the pasture legume *Biserrula pelecinus* L.—a new legume in agriculture. *Soil Biology and Biochemistry* 36(8):1309-1317.

72. Sullivan JT, Eardly BD, van Berkum P, & Ronson CW (1996) Four unnamed species of nonsymbiotic rhizobia isolated from the rhizosphere of *Lotus corniculatus*. *Appl. Environ. Microbiol.* 62(8):2818-2825.

73. Wang S*, et al.* (2014) Whole-genome sequencing of *Mesorhizobium huakuii* 7653R provides molecular insights into host specificity and symbiosis island dynamics. *BMC genomics* 15:440.

74. Ramsay JP, Sullivan JT, Stuart GS, Lamont IL, & Ronson CW (2006) Excision and transfer of the *Mesorhizobium loti* R7A symbiosis island requires an integrase IntS, a novel recombination directionality factor RdfS, and a putative relaxase RlxS. *Mol. Microbiol.* 62(3):723-734.

75. Ramsay JP & Ronson CW (2015) Genetic Regulation of Symbiosis Island Transfer in Mesorhizobium loti. *Biological Nitrogen Fixation*, ed de Bruijn FJ (John Wiley & Sons, Inc), Vol 1, pp 217-224.

76. Ramsay JP & Ronson CW (2015) Silencing quorum sensing and ICE mobility through antiactivation and ribosomal frameshifting. *Mob Genet Elements* 5(6):103-108.

77. Wozniak RA & Waldor MK (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat. Rev. Microbiol.* 8(8):552-563.

78. Johnson CM & Grossman AD (2015) Integrative and conjugative elements (ICEs): what they do and how they work. *Annu. Rev. Genet.* 49:577-601.

79. Rankin DJ, Rocha EPC, & Brown SP (2011) What traits are carried on mobile genetic elements, and why? *Heredity* 106(1):1-10.

80. Frost LS, Leplae R, Summers AO, & Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* 3(9):722-732.

81. Bi D*, et al.* (2012) ICEberg: a web-based resource for integrative and conjugative elements found in Bacteria. *Nucleic Acids Res* 40(Database issue):D621-D626.

82. Hochhut B*, et al.* (2001) Molecular analysis of antibiotic resistance gene clusters in *Vibrio cholerae* O139 and O1 SXT constins. *Antimicrob. Agents Chemother.* 45(11):2991-3000.

83. Wozniak RAF*, et al.* (2009) Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. *PLOS GENET* 5(12):e1000786.

84. Ravatn R, Studer S, Springael D, Zehnder AJ, & van der Meer JR (1998) Chromosomal integration, tandem amplification, and deamplification in *Pseudomonas putida* F1 of a 105-kilobase genetic element containing the chlorocatechol degradative genes from *Pseudomonas* sp. Strain B13. *J. Bacteriol.* 180(17):4360-4369.

85. Qiu X, Gurkar AU, & Lory S (2006) Interstrain transfer of the large pathogenicity island (PAPI-1) of *Pseudomonas aeruginosa*. *PNAS* 103(52):19830-19835.

86. Reiter WD, Palm P, & Yeats S (1989) Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Res* 17(5):1907-1914.

87. Song L, Pan Y, Chen S, & Zhang X (2012) Structural characteristics of genomic islands associated with GMP synthases as integration hotspot among sequenced microbial genomes. *Computational Biololgy and Chemistry* 36:62-70.

88. Auchtung JM, Aleksanyan N, Bulku A, & Berkmen MB (2016) Biology of ICE*Bs*1, an integrative and conjugative element in *Bacillus subtilis*. *Plasmid* 86:14-25.

89. Auchtung JM, Lee CA, Monson RE, Lehman AP, & Grossman AD (2005) Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. *PNAS* 102(35):12554-12559.

90. Roberts AP & Mullany P (2011) Tn*916*-like genetic elements: a diverse group of modular mobile elements conferring antibiotic resistance. *FEMS Microbiol Rev* 35(5):856-871.

91. Hickman AB, Chandler M, & Dyda F (2010) Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. *Critical reviews in biochemistry and molecular biology* 45(1):50-69.

92. Grindley ND, Whiteson KL, & Rice PA (2006) Mechanisms of site-specific recombination. *Annual review of biochemistry* 75:567-605.

93. Nesmelova IV & Hackett PB (2010) DDE transposases: structural similarity and diversity. *Advanced drug delivery reviews* 62(12):1187-1195.

94. Bellanger X, Payot S, Leblond-Bourget N, & Guedon G (2014) Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiol Rev* 38(4):720-760.

95. Casjens SR & Hendrix RW (2015) Bacteriophage lambda: early pioneer and still relevant. *Virology* 0:310-330.

96. Papagiannis CV*, et al.* (2007) Fis targets assembly of the Xis nucleoprotein filament to promote excisive recombination by phage lambda. *J. Mol. Biol* 367(2):328-343.

97. Esposito D & Gerard GF (2003) The Escherichia coli Fis protein stimulates bacteriophage lambda integrative recombination in vitro. *J. Bacteriol.* 185(10):3076-3080.

98. Rice PA, Yang S, Mizuuchi K, & Nash HA (1996) Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* 87(7):1295-1306.

99. Mendelson I, Gottesman M, & Oppenheim AB (1991) HU and integration host factor function as auxiliary proteins in cleavage of phage lambda cohesive ends by terminase. *J. Bacteriol.* 173(5):1670-1676.

100. Craig NL & Nash HA (1984) *E. coli* integration host factor binds to specific sites in DNA. *Cell* 39(3 Pt 2):707-716.

101. Kikuchi Y & Nash HA (1979) Nicking-closing activity associated with bacteriophage lambda *int* gene product. *PNAS* 76(8):3760-3764.

102. Miller HI, Kikuchi A, Nash HA, Weisberg RA, & Friedman DI (1979) Site-specific recombination of bacteriophage lambda: the role of host gene products. *Cold Spring Harbor symposia on quantitative biology* 43(2):1121-1126.

103. Nash HA (1981) Integration and excision of bacteriophage λ: the mechanism of conservative site specific recombination. *Annu. Rev. Genet.* 15(1):143-167.

104. Burrus V & Waldor MK (2003) Control of SXT integration and excision. *J. Bacteriol.* 185(17):5045-5054.

105. Lee CA, Auchtung JM, Monson RE, & Grossman AD (2007) Identification and characterization of int (integrase), xis (excisionase) and chromosomal attachment sites of the integrative and conjugative element ICE*Bs1* of *Bacillus subtilis*. *Mol. Microbiol.* 66(6):1356-1369.

106. Echols H & Guameros G (1983) Control of integration and excision. *Lambda II*, eds Hendrix R, Roberts J, Stahl F, & Wiesberg R (Cold Spring Harbor Lab Press, Cold Spring Harbor, New York), pp 77-93.

107. Lewis JA & Hatfull GF (2001) Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res* 29(11):2205-2216.

108. Abbani MA*, et al.* (2007) Structure of the cooperative Xis–DNA complex reveals a micronucleoprotein filament that regulates phage lambda intasome assembly. *PNAS* 104(7):2109-2114.

109. Dodd IB, Reed MR, & Egan JB (1993) The Cro-like Apl repressor of coliphage 186 is required for prophage excision and binds near the phage attachment site. *Mol. Microbiol.* 10(5):1139-1150.

110. Reed MR, Shearwin KE, Pell LM, & Egan JB (1997) The dual role of Apl in prophage induction of coliphage 186. *Mol. Microbiol.* 23(4):669-681.

111. Saha S, Haggard-Ljungquist E, & Nordstrom K (1987) The cox protein of bacteriophage P2 inhibits the formation of the repressor protein and autoregulates the early operon. *EMBO J* 6(10):3191-3199.

112. Haggard-Ljungquist E, Kockum K, & Bertani LE (1987) DNA sequences of bacteriophage P2 early genes cox and B and their regulatory sites. *Molecular Genetics and Genomics* 208(1-2):52-56.

113. Yu A & Haggard-Ljungquist E (1993) The Cox protein is a modulator of directionality in bacteriophage P2 site-specific recombination. *J. Bacteriol.* 175(24):7848-7855.

114. Esposito D, Wilson JC, & Scocca JJ (1997) Reciprocal regulation of the early promoter region of bacteriophage HP1 by the Cox and Cl proteins. *Virology* 234(2):267-276.

115. Yin S, Bushman W, & Landy A (1985) Interaction of the lambda site-specific recombination protein Xis with attachment site DNA. *PNAS* 82(4):1040-1044.

116. Guglielmini J, Quintais L, Garcillan-Barcia MP, de la Cruz F, & Rocha EP (2011) The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *PLOS GENET* 7(8):e1002222.

117. Ilangovan A, Connery S, & Waksman G (2015) Structural biology of the Gram-negative bacterial conjugation systems. *Trends in Microbiology* 23(5):301-310.

118. de la Cruz F, Frost LS, Meyer RJ, & Zechner EL (2010) Conjugative DNA metabolism in Gram-negative bacteria. *FEMS Microbiol Rev* 34(1):18-40.

119. Rajeev L, Malanowska K, & Gardner JF (2009) Challenging a paradigm: the role of DNA homology in tyrosine recombinase reactions. *Microbiology and Molecular Biology Reviews : MMBR* 73(2):300-309.

120. Wright LD, Johnson CM, & Grossman AD (2015) Identification of a single strand origin of replication in the integrative and conjugative element ICE*Bs*1 of *Bacillus subtilis*. *PLOS GENET* 11(10):e1005556.

121. Khan SA (2005) Plasmid rolling-circle replication: highlights of two decades of research. *Plasmid* 53(2):126-136.

122. Ghinet MG*, et al.* (2011) Uncovering the prevalence and diversity of integrating conjugative elements in actinobacteria. *PLoS ONE* 6(11):e27846.

123. te Poele EM, Bolhuis H, & Dijkhuizen L (2008) Actinomycete integrative and conjugative elements. *Antonie van Leeuwenhoek* 94(1):127-143.

124. Ambroset C*, et al.* (2015) New insights into the classification and integration specificity of *Streptococcus* integrative conjugative elements through extensive genome exploration. *Front Microbiol.* 6:1483.

125. Burrus V, Pavlovic G, Decaris B, & Guedon G (2002) The ICE*St*1 element of *Streptococcus thermophilus* belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration. *Plasmid* 48(2):77-97.

126. Coluzzi C*, et al.* (2017) A glimpse into the world of integrative and mobilizable elements in streptococci reveals an unexpected diversity and novel families of mobilization proteins. *Front Microbiol.* 8:443.

127. Guerillot R, Da Cunha V, Sauvage E, Bouchier C, & Glaser P (2013) Modular evolution of TnGBSs, a new family of integrative and conjugative elements associating insertion sequence transposition, plasmid replication, and conjugation for their spreading. *J. Bacteriol.* 195(9):1979-1990.

128. Mingoia M*, et al.* (2014) Tn*5253* family integrative and conjugative elements carrying *mef*(I) and *catQ* determinants in *Streptococcus pneumoniae* and *Streptococcus pyogenes*. *Antimicrob. Agents Chemother.* 58(10):5886-5893.

129. Dahmane N*, et al.* (2017) Diversity of integrative and conjugative elements of *Streptococcus salivarius* and their Intra- and interspecies transfer. *Appl. Environ. Microbiol.* 83(13):e00337-00317.

130. Sullivan JT*, et al.* (2002) Comparative sequence analysis of the symbiosis island of *Mesorhizobium loti* strain R7A. *J. Bacteriol.* 184(11):3086-3095.

131. Carraro N, Rivard N, Burrus V, & Ceccarelli D (2017) Mobilizable genomic islands, different strategies for the dissemination of multidrug resistance and other adaptive traits. *Mob Genet Elements* 7(2):1-6.

132. Doublet B, Golding GR, Mulvey MR, & Cloeckaert A (2008) Secondary chromosomal attachment site and tandem integration of the mobilizable *Salmonella* genomic island 1. *PLoS One* 3(4):e2060.

133. Bellanger X*, et al.* (2011) Site-specific accretion of an integrative conjugative element together with a related genomic island leads to *cis* mobilization and gene capture. *Mol. Microbiol.* 81(4):912-925.

134. Pavlovic G, Burrus V, Gintz B, Decaris B, & Guédon G (2004) Evolution of genomic islands by deletion and tandem accretion by site-specific recombination: ICE*St1*-related elements from *Streptococcus thermophilus*. *Microbiology (Reading, England)* 150(4):759-774.

135. Puymège A, Bertin S, Chuzeville S, Guédon G, & Payot S (2013) Conjugative transfer and *cis*-mobilization of a genomic island by an integrative and conjugative element of *Streptococcus agalactiae*. *J. Bacteriol.* 195(6):1142-1151.

136. Wang P*, et al.* (2017) Dissemination and loss of a biofilm-related genomic island in marine *Pseudoalteromonas* mediated by integrative and conjugative elements. *Environmental microbiology* 19(11):4620-4637.

137. Burrus V & Waldor MK (2004) Formation of SXT tandem arrays and SXT-R391 hybrids. *Journal of bacteriology* 186(9):2636-2645.

138. Hochhut B, Beaber JW, Woodgate R, & Waldor MK (2001) Formation of chromosomal tandem arrays of the SXT element and R391, two conjugative chromosomally integrating eements that share an attachment site. *Journal of bacteriology* 183(4):1124-1132.

139. Possoz C, Ribard C, Gagnat J, Pernodet JL, & Guerineau M (2001) The integrative element pSAM2 from *Streptomyces*: kinetics and mode of conjugal transfer. *Mol. Microbiol.* 42(1):159-166.

140. Ravatn R, Studer S, Springael D, Zehnder AJB, & van der Meer JR (1998) Chromosomal integration, tandem amplification, and deamplification in *Pseudomonas putida* F1 of a 105-Kilobase genetic element containing the chlorocatechol degradative genes from *Pseudomonas* sp. Strain B13. *Journal of bacteriology* 180(17):4360-4369.

141. Garriss G, Waldor MK, & Burrus V (2009) Mobile antibiotic resistance encoding elements promote their own diversity. *PLoS genetics* 5(12):e1000775.

142. Ramsay JP*, et al.* (2016) An updated view of plasmid conjugation and mobilization in *Staphylococcus*. *Mob Genet Elements* 6(4):e1208317.

143. Guédon G, Libante V, Coluzzi C, Payot S, & Leblond-Bourget N (2017) The obscure world of integrative and mobilizable elements, highly widespread elements that pirate bacterial conjugative systems. *Genes* 8(11):337.

144. Delavat F, Miyazaki R, Carraro N, Pradervand N, & van der Meer JR (2017) The hidden life of integrative and conjugative elements. *FEMS Microbiol Rev* 41(4):512-537.

145. Burrus V (2017) Mechanisms of stabilization of integrative and conjugative elements. *Current Opinion in Microbiology* 38(Supplement C):44-50.

146. Carraro N & Burrus V (2015) The dualistic nature of integrative and conjugative elements. *Mob Genet Elements* 5(6):98-102.

147. Grohmann E (2010) Autonomous plasmid-like replication of Bacillus ICE*Bs1*: a general feature of integrative conjugative elements? *Mol. Microbiol.* 75(2):261-263.

148. Lee CA, Babic A, & Grossman AD (2010) Autonomous plasmid-like replication of a conjugative transposon. *Mol. Microbiol.* 75(2):268-279.

149. Thomas J, Lee CA, & Grossman AD (2013) A conserved helicase processivity factor is needed for conjugation and replication of an integrative and conjugative element. *cPLoS Genetics* 9(1):e1003198.

150. Carraro N, Poulin D, & Burrus V (2015) Replication and active partition of integrative and conjugative elements (ICEs) of the SXT/R391 family: the line between ICEs and conjugative plasmids is getting thinner. *PLOS GENET* 11(6):e1005298.

151. Wright LD & Grossman AD (2016) Autonomous replication of the conjugative transposon Tn*916*. *J. Bacteriol.* 198(24):3355-3366.

152. Carraro N*, et al.* (2016) Plasmid-like replication of a minimal streptococcal integrative and conjugative element. *Microbiology* 162(4):622-632.

153. Carraro N*, et al.* (2011) Differential regulation of two closely related integrative and conjugative elements from *Streptococcus thermophilus*. *BMC microbiology* 11:238.

154. Salje J, Gayathri P, & Lowe J (2010) The ParMRC system: molecular mechanisms of plasmid segregation by actin-like filaments. *Nat. Rev. Microbiol.* 8(10):683-692.

155. Poulin-Laprade D, Matteau D, Jacques PE, Rodrigue S, & Burrus V (2015) Transfer activation of SXT/R391 integrative and conjugative elements: unraveling the SetCD regulon. *Nucleic Acids Res* 43(4):2045-2056.

156. Unterholzner SJ, Poppenberger B, & Rozhon W (2013) Toxin–antitoxin systems: biology, identification, and application. *Mob Genet Elements* 3(5):e26219.

157. Wozniak RA & Waldor MK (2009) A toxin-antitoxin system promotes the maintenance of an integrative conjugative element. *PLOS GENET* 5(3):e1000439.

158. Dziewit L, Jazurek M, Drewniak L, Baj J, & Bartosik D (2007) The SXT conjugative element and linear prophage N15 encode toxin-Antitoxin-stabilizing systems homologous to the *tad-ata* module of the *Paracoccus aminophilus* plasmid pAMI2. *J. Bacteriol.* 189(5):1983-1997.

159. Khoo SK*, et al.* (2007) Molecular and structural characterization of the PezAT chromosomal toxin-antitoxin system of the human pathogen *Streptococcus* pneumoniae. *Journal of Biological Chemistry* 282(27):19606-19618.

160. Sullivan JT, Patrick HN, Lowther WL, Scott DB, & Ronson CW (1995) Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. *PNAS* 92(19):8985-8989.

161. Chapman HM, Lowther WL, & Trainor KD (1990) Some factors limiting the success of Lotus corniculatus in hill and high country. *Proceedings of the New Zealand Grassland Association* 51:147-150.

162. Reeve W*, et al.* (2014) Genome sequence of the Lotus corniculatus microsymbiont *Mesorhizobium loti* strain R88B. *Standards in genomic sciences* 9:3.

163. Kelly S*, et al.* (2014) Genome sequence of the *Lotus* spp. microsymbiont *Mesorhizobium loti* strain R7A. *Standards in genomic sciences* 9:6-6.

164. Reeve W*, et al.* (2015) High-Quality draft genome sequence of the *Lotus* spp. microsymbiont *Mesorhizobium loti* strain CJ3Sym. *Standards in genomic sciences* 10:54.

165. Ding H & Hynes MF (2009) Plasmid transfer systems in the rhizobia. *Can J Microbiol* 55(8):917-927.

166. Frederix M & Downie JA (2011) Chapter 2 - Quorum Sensing: Regulating the Regulators. *Advances in microbial physiology*, ed Poole RK (Academic Press), Vol 58, pp 23-80.

167. Ramsay JP*, et al.* (2009) A LuxRI-family regulatory system controls excision and transfer of the *Mesorhizobium loti* strain R7A symbiosis island by activating expression of two conserved hypothetical genes. *Mol. Microbiol.* 73(6):1141-1155.

168. Waters CM & Bassler BL (2005) Quorum sensing: cell-to-cell communication in bacteria. *Annual Review of Cell and Developmental Biology* 21(1):319-346.

169. Papenfort K & Bassler B (2016) Quorum-sensing signal-response systems in gram-negative bacteria. *Nat. Rev. Microbiol.* 14(9):576-588.

170. Ramsay JP*, et al.* (2015) Ribosomal frameshifting and dual-target antiactivation restrict quorum-sensing–activated transfer of a mobile genetic element. *PNAS* 112(13):4104-4109.

171. McAnulla C, Edwards A, Sanchez-Contreras M, Sawers RG, & Downie JA (2007) Quorum-sensing-regulated transcriptional initiation of plasmid transfer and replication genes in *Rhizobium leguminosarum* biovar viciae. *Microbiology* 153(Pt 7):2074-2082.

172. Wisniewski-Dye F & Downie JA (2002) Quorum-sensing in *Rhizobium*. *Antonie Van Leeuwenhoek* 81(1-4):397-407.

173. White CE & Winans SC (2007) Cell–cell communication in the plant pathogen *Agrobacterium tumefaciens*. *Philosophical Transactions of the Royal Society B: Biological Sciences* 362(1483):1135-1148.

174. Fuqua WC & Winans SC (1994) A LuxR-LuxI type regulatory system activates *Agrobacterium* Ti plasmid conjugal transfer in the presence of a plant tumor metabolite. *J. Bacteriol.* 176(10):2796-2806.

175. Hwang I*, et al.* (1994) TraI, a LuxI homologue, is responsible for production of conjugation factor, the Ti plasmid *N*-acylhomoserine lactone autoinducer. *PNAS* 91(11):4639-4643.

176. Ramsay JP*, et al.* (2013) A widely conserved molecular switch controls quorum sensing and symbiosis island transfer in *Mesorhizobium loti* through expression of a novel antiactivator. *Molecular microbiology* 87(1):1-13.

177. Kelly S*, et al.* (2014) Genome sequence of the *Lotus* spp. microsymbiont *Mesorhizobium loti* strain NZP2037. *Standards in genomic sciences* 9:7.

178. Kasai-Maita H*, et al.* (2013) Commonalities and differences among symbiosis islands of three *Mesorhizobium loti* strains. *Microbes Environ.* 28(2):275-278.

179. Kaneko T*, et al.* (2000) Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Research* 7(6):331-338.

180. O'Callaghan KJ, Davey MR, & Cocking EC (1998) Crack Entry Invasion of *Sesbania rostrata* by *Azorhizobium caulinodans* ORS571 is *Nod* Gene-Independent. *Biological Nitrogen Fixation for the 21st Century: Proceedings of the 11th International Congress on Nitrogen Fixation, Institut Pasteur, Paris, France, July 20–25, 1997*, eds Elmerich C, Kondorosi A, & Newton WE (Springer Netherlands, Dordrecht), pp 266-266.

181. Batista JS, Hungria M, Barcellos FG, Ferreira MC, & Mendes IC (2007) Variability in *Bradyrhizobium japonicum* and *B. elkanii* seven years after introduction of both the exotic microsymbiont and the soybean host in a cerrados soil. *Microbial ecology* 53(2):270-284.

182. Ferreira MC & Hungria M (2002) Recovery of soybean inoculant strains from uncropped soils in Brazil. *Field Crops Research* 79(2):139-152.

183. Jabbouri S*, et al.* (1995) Involvement of *nodS* in *N*-methylation and *nodU* in 6-*O*-carbamoylation of *Rhizobium* sp. NGR234 nod factors. *Journal of Biological Chemistry* 270(39):22968-22973.

184. Jabbouri S*, et al.* (1998) *nolO* and *noeI* (HsnIII) of *Rhizobium* sp. NGR234 are involved in 3-*O*-carbamoylation and 2-*O*-methylation of Nod factors. *Journal of Biological Chemistry* 273(20):12047-12055.

185. Pankhurst CE & Jones WT (1979) Effectiveness of Lotus Root Nodules. effect of combined nitrogen on nodule effectiveness and flavolan synthesis in plant roots. *Journal of Experimental Botany* 30(6):1109-1118.

186. Jarvis BDW, Pankhurst CE, & Patel JJ (1982) *Rhizobium loti*, a new species of legume root nodule bacteria. *International journal of systematic and evolutionary microbiology* 32(3):378-380.

187. Hubber A, Vergunst AC, Sullivan JT, Hooykaas PJ, & Ronson CW (2004) Symbiotic phenotypes and translocated effector proteins of the *Mesorhizobium loti* strain R7A VirB/D4 type IV secretion system. *Mol. Microbiol.* 54(2):561-574.

188. Sanchez C, Iannino F, Deakin WJ, Ugalde RA, & Lepek VC (2009) Characterization of the *Mesorhizobium loti* MAFF303099 type-three protein secretion system. *Molecular Plant-Microbe Interactions* 22(5):519-528.

189. Vinayak M & Pathak C (2009) Queuosine modification of tRNA: its divergent role in cellular machinery. *Bioscience reports* 30(2):135-148.

190. Marchetti M*, et al.* (2013) Queuosine biosynthesis is required for *Sinorhizobium meliloti* induced cytoskeletal modifications on HeLa cells and symbiosis with *Medicago truncatula*. *PLoS ONE* 8(2):e56043.

191. Sullivan JT, Brown SD, Yocum RR, & Ronson CW (2001) The bio operon on the acquired symbiosis island of *Mesorhizobium* sp. strain R7A includes a novel gene involved in pimeloyl-CoA synthesis. *Microbiology* 147(Pt 5):1315-1322.

192. Howieson JG, Loi A, & Carr SJ (1995) *Biserrula pelecinus* L.-a legume pasture species with potential for acid, duplex soils which is nodulated by unique root-nodule bacteria. *Australian Journal of Experimental Agriculture* 46:997-1009.

193. Nandasena KG, O'Hara GW, Tiwari RP, & Howieson JG (2006) Rapid on situ evolution of nodulating strains for *Biserrula pelecinus* L. through lateral transfer of a symbiosis island from the original mesorhizobial inoculant. *Appl. Environ. Microbiol.* 72(11):7365-7367.

194. Nandasena KG, O'Hara GW, Tiwari RP, Sezmis E, & Howieson JG (2007) *In situ* lateral transfer of symbiosis islands results in rapid evolution of diverse competitive strains of mesorhizobia suboptimal in symbiotic nitrogen fixation on the pasture legume *Biserrula pelecinus* L. *Environmental microbiology* 9(10):2496-2511.

195. Nandasena KG, O'Hara GW, Tiwari RP, Willems A, & Howieson JG (2009) *Mesorhizobium australicum* sp. nov. and *Mesorhizobium opportunistum* sp. nov., isolated from *Biserrula pelecinus* L. in Australia. *International Journal of Systematic and Evolutionary Microbiology* 59(Pt 9):2140-2147.

196. Bullard GK, Roughley RJ, & Pulsford DJ (2005) The legume inoculant industry and inoculant quality control in Australia: 19532003. *Australian Journal of Experimental Agriculture* 45(3):127-140.

197. Drew E, Herridge, D., Ballard, R., O'Hara, G., Deaker, R., Denton, M., Yates, R., Gemell, G., Hartley, E., Phillips, L., Seymour, N., Howieson, J., Ballard, N. (2012) *Inoculating legumes: a practical guide. Grains Research and Development Corporation* (Grains Research and Development Organisation, Kingston, Australia).

198. van Kessel C & Hartley C (2000) Agricultural management of grain legumes: has it led to an increase in nitrogen fixation? *Field Crops Research* 65(2):165-181.

199. FAOSTAT (2018) (FAO Statistics Division).

200. Kinzig AP & Socolow RH (1994) Human impacts on the nitrogen cycle. *Physics Today* 47(11):24-31.

201. Vance CP, Graham PH, & Allan DL (2000) Biological Nitrogen Fixation: Phosphorus - A Critical Future Need? *Nitrogen Fixation: From Molecules to Crop Productivity*, eds Pedrosa FO, Hungria M, Yates G, & Newton WE (Springer Netherlands, Dordrecht), pp 509-514.

202. Slattery J, Pearce, D. (2002) Development of Elite Inoculant Rhizobium Strains in Southeastern Australia. *Inoculants and Nitrogen Fixation of Legumes in Vietnam*, ed Herridge D (ACIAR, Australia), pp 86-94.

203. Slattery JF, Pearce DJ, & Slattery WJ (2004) Effects of resident rhizobial communities and soil type on the effective nodulation of pulse legumes. *Soil Biology and Biochemistry* 36(8):1339-1346.

204. Nandasena K*, et al.* (2014) Complete genome sequence of *Mesorhizobium ciceri* bv. biserrulae type strain (WSM1271(T)). *Standards in genomic sciences* 9(3):462-472.

205. Reeve W*, et al.* (2013) Complete genome sequence of *Mesorhizobium opportunistum* type strain WSM2075(T.). *Standards in genomic sciences* 9(2):294-303.

206. Reeve W*, et al.* (2013) Complete genome sequence of *Mesorhizobium australicum* type strain (WSM2073(T)). *Standards in genomic sciences* 9(2):410-419.

207. Miller JH (1972) *Experiments in molecular genetics* (Cold Spring Harbour Laboratory Press, New York) p pdb.rec8141.

208. Thoma S & Schobert M (2009) An improved *Escherichia coli* donor strain for diparental mating. *FEMS Microbiol Lett* 294(2):127-132.

209. Beringer JE (1974) R factor transfer in *Rhizobium leguminosarum*. *J Gen Microbiol* 84(1):188-198.

210. Ronson CW, Nixon BT, Albright LM, & Ausubel FM (1987) *Rhizobium meliloti ntrA* (*rpoN*) gene is required for diverse metabolic functions. *J. Bacteriol.* 169(6):2424-2431.

211. McClean KH*, et al.* (1997) Quorum sensing and *Chromobacterium violaceum*: exploitation of violacein production and inhibition for the detection of N-acylhomoserine lactones. *Microbiology* 143 ( Pt 12):3703-3711.

212. Goel A, Sindhu S, & Dadarwal K (2002) Stimulation of nodulation and plant growth of chickpea (*Cicer arietinum* L.) by *Pseudomonas* spp. antagonistic to fungal pathogens. *Biology and Fertility of Soils* 36(6):391-396.

213. Reeve W*, et al.* (2015) A genomic encyclopedia of the root nodule bacteria: assessing genetic diversity through a systematic biogeographic survey. *Standards in genomic sciences* 10:14-14.

214. Vidal C*, et al.* (2009) *Mesorhizobium metallidurans* sp. nov., a metal-resistant symbiont of *Anthyllis vulneraria* growing on metallicolous soil in Languedoc, France. *International journal of systematic and evolutionary microbiology* 59(Pt 4):850-855.

215. Quandt J & Hynes MF (1993) Versatile suicide vectors which allow direct selection for gene replacement in gram-negative bacteria. *Gene* 127(1):15-21.

216. Hoang TT, Karkhoff-Schweizer RR, Kutchma AJ, & Schweizer HP (1998) A broad-host-range Flp-FRT recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked *Pseudomonas aeruginosa* mutants. *Gene* 212(1):77-86.

217. Prentki P & Krisch HM (1984) *In vitro* insertional mutagenesis with a selectable DNA fragment. *Gene* 29(3):303-313.

218. Prell J, Boesten B, Poole P, & Priefer UB (2002) The *Rhizobium leguminosarum* bv. viciae VF39 gamma-aminobutyrate (GABA) aminotransferase gene (*gabT*) is induced by GABA and highly expressed in bacteroids. *Microbiology* 148(Pt 2):615-623.

219. Rodpothong P*, et al.* (2009) Nodulation gene mutants of *Mesorhizobium loti* R7A-*nodZ* and *nolL* mutants have host-specific phenotypes on *Lotus* spp. *Molecular Plant-Microbe Interactions* 22(12):1546-1554.

220. Khan SR, Gaines J, Roop RM, & Farrand SK (2008) Broad-host-range expression vectors with tightly regulated promoters and their use to examine the influence of TraR and TraM expression on Ti plasmid quorum sensing. *Appl. Environ. Microbiol.* 74(16):5053-5062.

221. Antoine R*, et al.* (2000) New virulence-activated and virulence-repressed genes identified by systematic gene inactivation and generation of transcriptional fusions in *Bordetella pertussis*. *J. Bacteriol.* 182(20):5902-5905.

222. Dombrecht B, Vanderleyden J, & Michiels J (2001) Stable RK2-derived cloning vectors for the analysis of gene expression and gene function in gram-negative bacteria. *Molecular Plant-Microbe Interactions* 14(3):426-430.

223. Reeve WG, Tiwari RP, Melino V, & Poole PS (2016) Fundamental molecular techniques for rhizobia. *Working with rhizobia*, ed Howieson JG, Dilworth, M.J. (Australian Centre for International Agricultural Research, Canberra), pp 221–244.

224. Sambrook J, Russell, R.W (2001) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York) 3 Ed.

225. Ma J, Wang, Y., Wang, Y. (2013) A simple, fast and efficient method for cloning blunt DNA fragments. *Afr. J. Biotechnol.* 12(26):4094-4097.

226. Wu N, Matand, K., Kebede, B., Acquaah, G., Williams, S. (2010) Enhancing DNA electrotransformation efficiency in *Escherichia coli* DH10B electrocompetent cells. *Electronic Journal of Biotechnology* 13(5).

227. Li X*, et al.* (2010) An improved calcium chloride method preparation and transformation of competent cells. *Afr. J. Biotechnol.* 9(50):8549-8554.

228. Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol* 166(4):557-580.

229. Ramsay J (2013) High-throughput β-galactosidase and β-glucuronidase assays using fluorogenic substrates. *Bio-Protocol* 3(14):e827.

230. Cubo MT, Buendia-Claveria AM, Beringer JE, & Ruiz-Sainz JE (1988) Melanin production by *Rhizobium* strains. *Appl. Environ. Microbiol.* 54(7):1812-1817.

231. Johnston AWB, Hombrecher G, Brewin NJ, & Cooper R, M. C. (1982) Two transmissible plasmids in *Rhizobium leguminosarum* strain 300. *Microbiology* 128(1):85-93.

232. Chan KG*, et al.* (2011) Characterization of N-acylhomoserine lactone-degrading bacteria associated with the *Zingiber officinale* (ginger) rhizosphere: co-existence of quorum quenching and quorum sensing in *Acinetobacter* and *Burkholderia*. *BMC microbiology* 11:51.

233. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357-359.

234. Rutherford K*, et al.* (2000) Artemis: sequence visualization and annotation. *Bioinformatics* 16(10):944-945.

235. Anders S, Pyl PT, & Huber W (2015) HTSeq - A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31.

236. Love MI, Huber W, & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15(12):550.

237. Alikhan NF, Petty NK, Ben Zakour NL, & Beatson SA (2011) BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC genomics* 12:402.

238. Notredame C, Higgins DG, & Heringa J (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol* 302(1):205-217.

239. Carver TJ*, et al.* (2005) ACT: the Artemis Comparison Tool. *Bioinformatics* 21(16):3422-3423.

240. Guy L, Roat Kultima J, & Andersson SGE (2010) genoPlotR: comparative gene and genome visualization in R. *Bioinformatics* 26(18):2334-2335.

241. Kearse M*, et al.* (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647-1649.

242. Markowitz VM*, et al.* (2012) IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res* 40(D1):D115-D122.

243. Sullivan J & Ronson C (1998) Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a *phe*-tRNA gene. *Proceedings of the National Academy of Sciences* 95:5145 - 5149.

244. Groth AC & Calos MP (2004) Phage integrases: biology and applications. *J. Mol. Biol* 335(3):667-678.

245. Hynes MF, Quandt J, O'Connell MP, & Puhler A (1989) Direct selection for curing and deletion of *Rhizobium* plasmids using transposons carrying the *Bacillus subtilis sacB* gene. *Gene* 78(1):111-120.

246. Hubber AM, Vergunst AC, Sullivan JT, Hooykaas PJJ, & Ronson CW (2004) Symbiotic phenotypes and translocated effector proteins of the *Mesorhizobium loti* strain R7A VirB/D4 type IV secretion system. *Molecular microbiology* 54(2):561-574.

247. Haskett TL*, et al.* (2016) Assembly and transfer of tripartite integrative and conjugative genetic elements. *Proceedings of the National Academy of Sciences* 113(43):12268-12273.

248. Byer AS, Shepard EM, Peters JW, & Broderick JB (2015) Radical S-adenosyl-L-methionine chemistry in the synthesis of hydrogenase and nitrogenase metal cofactors. *Journal of Biological Chemistry* 290(7):3987-3994.

249. Lehmann T, Hoffmann M, Hentrich M, & Pollmann S (2010) Indole-3-acetamide-dependent auxin biosynthesis: A widely distributed way of indole-3-acetic acid production? *European Journal of Cell Biology* 89(12):895-905.

250. Spaepen S & Vanderleyden J (2011) Auxin and plant-microbe interactions. *Cold Spring Harbor Perspectives in Biology* 3(4):a001438.

251. Oetiker JH, Lee DH, & Kato A (1999) Molecular analysis of a tryptophan-2-monooxygenase gene (IaaM) of *Agrobacterium vitis*. *DNA Sequence* 10(4-5):349-354.

252. Camerini S*, et al.* (2008) Introduction of a novel pathway for IAA biosynthesis to rhizobia alters vetch root nodule development. *Archives of microbiology* 190(1):67-77.

253. Bianco C, Senatore B, Arbucci S, Pieraccini G, & Defez R (2014) Modulation of endogenous indole-3-acetic acid biosynthesis in bacteroids within *Medicago sativa* nodules. *Appl. Environ. Microbiol.* 80(14):4286-4293.

254. Amundsen SK, Taylor AF, Chaudhury AM, & Smith GR (1986) *recD*: the gene for an essential third subunit of exonuclease V. *PNAS* 83(15):5558-5562.

255. Biek DP & Cohen SN (1986) Identification and characterization of *recD*, a gene affecting plasmid maintenance and recombination in *Escherichia coli*. *Journal of bacteriology* 167(2):594-603.

256. Lovett ST, Luisi-DeLuca C, & Kolodner RD (1988) The genetic dependence of recombination in *recD* mutants of *Escherichia coli*. *Genetics* 120(1):37-45.

257. Williams KP (2002) Integration sites for genetic elements in prokaryotic *tRNA* and *tmRNA* genes: sublocation preference of *integrase* subfamilies. *Nucleic Acids Research* 30(4):866-875.

258. Hochhut B & Waldor MK (1999) Site-specific integration of the conjugal *Vibrio cholerae* SXT element into *prfC*. *Molecular microbiology* 32:99-110.

259. Menard KL & Grossman AD (2013) Selective pressures to maintain attachment site specificity of integrative and conjugative elements. *PLoS genetics* 9(7):e1003623.

260. Wozniak RAF & Waldor MK (2009) A toxin–antitoxin system promotes the maintenance of an integrative conjugative element. *PLoS genetics* 5(3):e1000439.

261. Chapleau M*, et al.* (2015) Identification of genetic and environmental factors stimulating excision from *Streptomyces scabiei* chromosome of the toxicogenic region responsible for pathogenicity. *Molecular plant pathology*.

262. Weisberg RA & Gottesman ME (1971) The stability of Int and Xis functions. *The bacteriophage lambda*, ed Hershey AD (Cold Spring Harbor Lab Press, Cold Spring Harbor, New York), pp 489–500.

263. Trempy JE, Kirby JE, & Gottesman S (1994) Alp suppression of Lon: dependence on the *slpA* gene. *J. Bacteriol.* 176(7):2061-2067.

264. Drozdetskiy A, Cole C, Procter J, & Barton GJ (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* 43(W1):W389-W394.

265. Aravind L, Anantharaman V, Balaji S, Babu MM, & Iyer LM (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev* 29(2):231-262.

266. Freese NH, Norris DC, & Loraine AE (2016) Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* 32(14):2089-2095.

267. Lundqvist B & Bertani G (1984) Immunity repressor of bacteriophage P2. Identification and DNA-binding activity. *J. Mol. Biol* 178(3):629-651.

268. Dodd IB, Kalionis B, & Egan JB (1990) Control of gene expression in the temperate coliphage 186. VIII. Control of lysis and lysogeny by a transcriptional switch involving face-to-face promoters. *J. Mol. Biol* 214(1):27-37.

269. Saha S, Haggård-Ljungquist E, & Nordström K (1989) Activation of prophage P4 by the P2 Cox protein and the sites of action of the Cox protein on the two phage genomes. *PNAS* 86(11):3973-3977.

270. Ahlgren-Berg A*, et al.* (2009) A comparative analysis of the bifunctional Cox proteins of two heteroimmune P2-like phages with different host integration sites. *Virology* 385(2):303-312.

271. Piazzolla D*, et al.* (2006) Expression of phage P4 integrase is regulated negatively by both Int and Vis. *J Gen Virol* 87(Pt 8):2423-2431.

272. Six EW & Lindqvist BH (1978) Mutual derepression in the P2-P4 bacteriophage system. *Virology* 87(2):217-230.

273. Goessweiner-Mohr N*, et al.* (2014) Structure of the double-stranded DNA-binding type IV secretion protein TraN from *Enterococcus*. *Acta crystallographica. Section D, Biological crystallography* 70(Pt 9):2376-2389.

274. Abbani M, Iwahara M, & Clubb RT (2005) The structure of the excisionase (Xis) protein from conjugative transposon Tn*916* provides insights into the regulation of heterobivalent tyrosine recombinases. *J Mol Biol* 347(1):11-25.

275. Patel HK*, et al.* (2013) Bacterial LuxR solos have evolved to respond to different molecules including signals from plants. *Front Plant Sci.* 4:447.

276. Patankar AV & Gonzalez JE (2009) Orphan LuxR regulators of quorum sensing. *FEMS Microbiol Rev* 33(4):739-756.

277. Verma SC & Miyashiro T (2013) Quorum sensing in the squid-*Vibrio* symbiosis. *International Journal of Molecular Sciences* 14(8):16386-16401.

278. Daniels R, Vanderleyden J, & Michiels J (2004) Quorum sensing and swarming migration in bacteria. *FEMS Microbiol Rev* 28(3):261-289.

279. Li Y-H & Tian X (2012) Quorum sensing and bacterial social interactions in biofilms. *Sensors (Basel, Switzerland)* 12(3):2519-2538.

280. Rutherford ST & Bassler BL (2012) Bacterial quorum sensing: its role in virulence and possibilities for its control. *Cold Spring Harbor perspectives in medicine* 2(11).

281. Antunes LC, Ferreira RB, Buckner MM, & Finlay BB (2010) Quorum sensing in bacterial virulence. *Microbiology* 156(Pt 8):2271-2282.

282. Rajput A, Kaur K, & Kumar M (2016) SigMol: repertoire of quorum sensing signaling molecules in prokaryotes. *Nucleic Acids Res* 44(D1):D634-D639.

283. Thiel V, Kunze B, Verma P, Wagner-Dobler I, & Schulz S (2009) New structural variants of homoserine lactones in bacteria. *Chembiochem : a European journal of chemical biology* 10(11):1861-1868.

284. Schaefer AL, Val DL, Hanzelka BL, Cronan JE, Jr., & Greenberg EP (1996) Generation of cell-to-cell signals in quorum sensing: acyl homoserine lactone synthase activity of a purified *Vibrio fischeri* LuxI protein. *PNAS* 93(18):9505-9509.

285. Churchill MEA & Chen L (2011) Structural basis of acyl-homoserine lactone-dependent signaling. *Chemical Reviews* 111(1):68-85.

286. Gould TA, Herman J, Krank J, Murphy RC, & Churchill MEA (2006) Specificity of acyl-homoserine lactone synthases examined by mass spectrometry. *J. Bacteriol.* 188(2):773-783.

287. Hoang TT, Sullivan SA, Cusick JK, & Schweizer HP (2002) Beta-ketoacyl acyl carrier protein reductase (FabG) activity of the fatty acid biosynthetic pathway is a determining factor of 3-oxo-homoserine lactone acyl chain lengths. *Microbiology* 148(Pt 12):3849-3856.

288. Raychaudhuri A, Jerga A, & Tipton PA (2005) Chemical mechanism and substrate specificity of RhlI, an acylhomoserine lactone synthase from *Pseudomonas aeruginosa*. *Biochemistry* 44(8):2974-2981.

289. Chen J & Xie J (2011) Role and regulation of bacterial LuxR-like regulators. *Journal of cellular biochemistry* 112(10):2694-2702.

290. Fuqua C, Winans SC, & Greenberg EP (1996) Census and consensus in bacterial ecosystems: the LuxR-LuxI family of quorum-sensing transcriptional regulators. *Annual review of microbiology* 50:727-751.

291. Gerdt JP*, et al.* (2017) Chemical interrogation of LuxR-type quorum sensing receptors reveals new insights into receptor selectivity and the potential for interspecies bacterial signaling. *ACS chemical biology* 12(9):2457-2464.

292. Koch B*, et al.* (2005) The LuxR receptor: the sites of interaction with quorum-sensing signals and inhibitors. *Microbiology* 151(Pt 11):3589-3602.

293. Minogue TD, Carlier AL, Koutsoudis MD, & Von Bodman SB (2005) The cell density-dependent expression of stewartan exopolysaccharide in *Pantoea stewartii* ssp. stewartii is a function of EsaR-mediated repression of the *rcsA* gene. *Mol. Microbiol.* 56(1):189-203.

294. von Bodman SB, Majerczak DR, & Coplin DL (1998) A negative regulator mediates quorum-sensing control of exopolysaccharide production in *Pantoea stewartii* subsp. *stewartii*. *PNAS* 95(13):7687-7692.

295. Zhu J & Winans SC (2001) The quorum-sensing transcriptional regulator TraR requires its cognate signaling ligand for protein folding, protease resistance, and dimerization. *PNAS* 98(4):1507-1512.

296. Urbanowski ML, Lostroh CP, & Greenberg EP (2004) Reversible acyl-homoserine lactone binding to purified *Vibrio fischeri* LuxR protein. *J. Bacteriol.* 186(3):631-637.

297. Yang M, Giel JL, Cai T, Zhong Z, & Zhu J (2009) The LuxR family quorum-sensing activator MrtR requires its cognate autoinducer for dimerization and activation but not for protein folding. *J. Bacteriol.* 191(1):434-438.

298. Pérez PD, Weiss JT, & Hagen SJ (2011) Noise and crosstalk in two quorum-sensing inputs of *Vibrio fischeri*. *BMC Systems Biology* 5(1):153.

299. Czajkowski R & Jafra S (2009) Quenching of acyl-homoserine lactone-dependent quorum sensing by enzymatic disruption of signal molecules. *Acta biochimica Polonica* 56(1):1-16.

300. Uroz S, Dessaux Y, & Oger P (2009) Quorum sensing and quorum quenching: the yin and yang of bacterial communication. *Chembiochem : a European journal of chemical biology* 10(2):205-216.

301. Chen F, Gao Y, Chen X, Yu Z, & Li X (2013) Quorum quenching enzymes and their application in degrading signal molecules to block quorum sensing-dependent infection. *International Journal of Molecular Sciences* 14(9):17477.

302. Utari PD, Vogel J, & Quax WJ (2017) Deciphering physiological functions of AHL quorum quenching acylases. *Front Microbiol.* 8:1123.

303. Grandclement C, Tannieres M, Morera S, Dessaux Y, & Faure D (2016) Quorum quenching: role in nature and applied developments. *FEMS Microbiol Rev* 40(1):86-116.

304. Yang M*, et al.* (2009) Functional analysis of three AHL autoinducer synthase genes in *Mesorhizobium loti* reveals the important role of quorum sensing in symbiotic nodulation. *Can J Microbiol* 55:210+.

305. Williamson PR, Wakamatsu K, & Ito S (1998) Melanin biosynthesis in *Cryptococcus neoformans*. *J. Bacteriol.* 180(6):1570-1572.

306. Castro-Sowinski S, Martinez-Drets G, & Okon Y (2002) Laccase activity in melanin-producing strains of *Sinorhizobium meliloti*. *FEMS Microbiol Lett* 209(1):119-125.

307. Zhu X & Williamson PR (2004) Role of laccase in the biology and virulence of *Cryptococcus neoformans*. *FEMS yeast research* 5(1):1-10.

308. Plonka PM & Grabacka M (2006) Melanin synthesis in microorganisms-biotechnological and medical aspects. *Acta biochimica Polonica* 53(3):429-443.

309. Reiss R*, et al.* (2013) Laccase versus laccase-like multi-copper oxidase: a comparative study of similar enzymes with diverse substrate spectra. *PLoS ONE* 8(6):e65633.

310. Yates EA*, et al.* (2002) N-acylhomoserine lactones undergo lactonolysis in a pH-, temperature-, and acyl chain length-dependent manner during growth of *Yersinia pseudotuberculosis* and *Pseudomonas aeruginosa*. *Infection and immunity* 70(10):5635-5646.

311. Mei G-Y, Yan X-X, Turak A, Luo Z-Q, & Zhang L-Q (2010) AidH, an alpha/beta-hydrolase fold family member from an *Ochrobactrum* sp. strain, Is

a novel *N*-acylhomoserine lactonase. *Appl. Environ. Microbiol.* 76(15):4933-4942.

312. Holmquist M (2000) Alpha/beta-hydrolase fold enzymes: structures, functions and mechanisms. *Current protein & peptide science* 1(2):209-235.

313. Gao A*, et al.* (2013) High-resolution structures of AidH complexes provide insights into a novel catalytic mechanism for *N*-acyl homoserine lactonase. *Acta crystallographica. Section D, Biological crystallography* 69(Pt 1):82-91.

314. Mercado-Blanco J, García F, Fernández-López M, & Olivares J (1993) Melanin production by *Rhizobium meliloti* GR4 is linked to nonsymbiotic plasmid pRmeGR4b: cloning, sequencing, and expression of the tyrosinase gene *mepA*. *J. Bacteriol.* 175(17):5403-5410.

315. Riedel K*, et al.* (2001) *N*-acylhomoserine-lactone-mediated communication between *Pseudomonas aeruginosa* and *Burkholderia cepacia* in mixed biofilms. *Microbiology* 147(Pt 12):3249-3262.

316. Steindler L & Venturi V (2007) Detection of quorum-sensing *N*-acyl homoserine lactone signal molecules by bacterial biosensors. *FEMS Microbiol Lett* 266(1):1-9.

317. Hynes MF, Brucksch K, & Priefer U (1988) Melanin production encoded by a cryptic plasmid in a *Rhizobium leguminosarum* strain. *Archives of microbiology* 150(4):326-332.

318. Borthakur D, Lamb JW, & Johnston AW (1987) Identification of two classes of *Rhizobium phaseoli* genes required for melanin synthesis, one of which is required for nitrogen fixation and activates the transcription of the other. *Molecular Genetics and Genomics* 207(1):155-160.

319. Bell AA, Wheeler, M.H. (1986) Biosynthesis and functions of fungal melanins. *Annual Review of Phytopathology* 24(1):411-451.

320. Pinero S*, et al.* (2007) Tyrosinase from *Rhizobium etli* is involved in nodulation efficiency and symbiosis-associated stress resistance. *Journal of molecular microbiology and biotechnology* 13(1-3):35-44.

321. Hawkins FK & Johnston AW (1988) Transcription of a *Rhizobium leguminosarum* biovar phaseoli gene needed for melanin synthesis is activated by *nifA* of *Rhizobium* and *Klebsiella pneumoniae*. *Mol. Microbiol.* 2(3):331-337.

322. Lamb JW, Hombrecher G, & Johnston AWB (1982) Plasmid-determined nodulation and nitrogen-fixation abilities in *Rhizobium phaseoli*. *Molecular and General Genetics* 186(3):449-452.

323. Albuquerque P*, et al.* (2014) Quorum sensing-mediated, cell density-dependent regulation of growth and virulence in *Cryptococcus neoformans*. *mBio* 5(1).

324. Wang Y & Casadevall A (1994) Decreased susceptibility of melanized *Cryptococcus neoformans* to UV light. *Appl. Environ. Microbiol.* 60(10):3864-3866.

325. Rosas AL & Casadevall A (1997) Melanization affects susceptibility of *Cryptococcus neoformans* to heat and cold. *FEMS Microbiol Lett* 153(2):265-272.

326. Garcia-Rivera J & Casadevall A (2001) Melanization of *Cryptococcus neoformans* reduces its susceptibility to the antimicrobial effects of silver nitrate. *Medical mycology* 39(4):353-357.

327. Rosas AL & Casadevall A (2001) Melanization decreases the susceptibility of *Cryptococcus neoformans* to enzymatic degradation. *Mycopathologia* 151(2):53-56.

328. Czajkowski R*, et al.* (2011) Inactivation of AHLs by *Ochrobactrum* sp. A44 depends on the activity of a novel class of AHL acylase. *Environmental Microbiology Reports* 3(1):59-68.

329. Jafra S*, et al.* (2006) Detection and characterization of bacteria from the potato rhizosphere degrading *N*-acyl-homoserine lactone. *Can J Microbiol* 52(10):1006-1015.

330. Zhang H-B, Wang L-H, & Zhang L-H (2002) Genetic control of quorum-sensing signal turnover in *Agrobacterium tumefaciens*. *PNAS* 99(7):4638-4643.

331. Carlier A*, et al.* (2003) The Ti Plasmid of *Agrobacterium tumefaciens* Harbors an *attM*-paralogous Gene, *aiiB*, also encoding *N*-acyl homoserine lactonase activity. *Appl. Environ. Microbiol.* 69(8):4989-4993.

332. Khan SR & Farrand SK (2009) The BlcC (AttM) lactonase of *Agrobacterium tumefaciens* does not quench the quorum-Sensing system that regulates Ti plasmid conjugative transfer. *J. Bacteriol.* 191(4):1320-1329.

333. Zhang HB, Wang C, & Zhang LH (2004) The quormone degradation system of *Agrobacterium tumefaciens* is regulated by starvation signal and stress alarmone (p)ppGpp. *Mol. Microbiol.* 52(5):1389-1401.

334. Chai Y, Tsai CS, Cho H, & Winans SC (2007) Reconstitution of the biochemical activities of the AttJ repressor and the AttK, AttL, and AttM catabolic enzymes of *Agrobacterium tumefaciens*. *J. Bacteriol.* 189(9):3674-3679.

335. Dorgai L, Oberto J, & Weisberg RA (1993) Xis and Fis proteins prevent site-specific DNA inversion in lysogens of phage HK022. *J. Bacteriol.* 175(3):693-700.

336. Bi H, Christensen QH, Feng Y, Wang H, & Cronan JE (2012) The *Burkholderia cenocepacia* BDSF quorum sensing fatty acid is synthesized by a bifunctional crotonase homologue having both dehydratase and thioesterase activities. *Mol. Microbiol.* 83(4):840-855.

337. Deng Y, Wu J, Eberl L, & Zhang LH (2010) Structural and functional characterization of diffusible signal factor family quorum-sensing signals produced by members of the *Burkholderia cepacia* complex. *Applied Environmental Microbiology* 76(14):4675-4683.

338. Ryan RP, McCarthy Y, Watt SA, Niehaus K, & Dow JM (2009) Intraspecies signaling involving the diffusible signal factor BDSF (*cis*-2-dodecenoic acid) influences virulence in *Burkholderia cenocepacia*. *J. Bacteriol.* 191(15):5013-5019.

339. Zhou L*, et al.* (2015) The multiple DSF-family QS signals are synthesized from carbohydrate and branched-chain amino acids via the FAS elongation cycle. *Scientific Reports* 5:13294.

340. Krick A*, et al.* (2007) A marine *Mesorhizobium* sp. produces structurally novel long-chain *N*-acyl-*L*-homoserine lactones. *Appl. Environ. Microbiol.* 73(11):3587-3594.

341. Wilson JW & Nickerson CA (2006) A new experimental approach for studying bacterial genomic island evolution identifies island genes with bacterial host-specific expression patterns. *BMC Evolutionary Biology* 6(1):2.

342. Wang L*, et al.* (2013) A minimal nitrogen fixation gene cluster from *Paenibacillus sp.* WLY78 enables expression of active nitrogenase in *Escherichia coli*. *PLoS genetics* 9(10):e1003865.

343. Kim J & Copley SD (2012) Inhibitory cross-talk upon introduction of a new metabolic pathway into an existing metabolic network. *Proceedings of the National Academy of Sciences* 109(42):E2856-2864.

344. Kumar N*, et al.* (2015) Bacterial genospecies that are not ecologically coherent: population genomics of *Rhizobium leguminosarum*. *Open biology* 5(1):140133.

345. Williams D, Gogarten JP, & Papke RT (2012) Quantifying homologous replacement of loci between haloarchaeal species. *Genome biology and evolution* 4(12):1223-1244.

346. Coscolla M, Comas I, & Gonzalez-Candelas F (2011) Quantifying nonvertical inheritance in the evolution of *Legionella pneumophila*. *Molecular biology and evolution* 28(2):985-1001.

347. Gormley NA, Watson MA, & Halford SE (2001) Bacterial restriction–modification systems. *eLS*, (John Wiley & Sons, Ltd).

348. Thomas CM & Nielsen KM (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology* 3(9):711-721.

349. Lopez-Sanchez MJ*, et al.* (2012) The highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome. *Molecular microbiology* 85(6):1057-1071.

350. Zhang Q, Rho M, Tang H, Doak TG, & Ye Y (2013) CRISPR-Cas systems target a diverse collection of invasive mobile genetic elements in human microbiomes. *Genome Biology* 14(4):R40.

351. Marrero J & Waldor MK (2007) The SXT/R391 family of integrative conjugative elements is composed of two exclusion groups. *Journal of bacteriology* 189(8):3302-3305.

352. Auchtung JM, Lee CA, Garrison KL, & Grossman AD (2007) Identification and characterization of the immunity repressor (ImmR) that controls the mobile genetic element ICE*Bs1* of *Bacillus subtilis*. *Molecular microbiology* 64(6):1515-1528.

353. Elias NV & Herridge DF (2015) Naturalised populations of mesorhizobia in chickpea (*Cicer arietinum* L.) cropping soils: effects on nodule occupancy and productivity of commercial chickpea. *Plant and Soil* 387(1):233-249.