Bacterial Chromosome Engineering for Applications in Metabolic Engineering

by

John Heil

A thesis

presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
biology

Waterloo, Ontario, Canada, 2015

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

Abstract

As genetic engineering technologies continue to advance and metabolic engineering projects become more ambitious, the need for chromosome engineering tools is increased. In this thesis, I have described the development of a ΦC31 phage integrase based cassette exchange system. This system is called integrase mediated cassette exchange (IMCE). IMCE allows for the integration of a ΦC31 *attB* flanked DNA cassette, carried on a mobilizable donor vector, into a chromosomal landing pad locus. The landing pad locus, consisting of ΦC31 *attP* sites flanking marker genes, was integrated into the chromosome of three members of the *Rhizobiales*, by homologous recombination to create landing pad strains (LP-strains). The integration of the donor cassette is catalyzed by ΦC31 integrase, expressed from a mobilizable vector. To accomplish IMCE, a tetraparental conjugative mating is carried out, including the *attB* donor vector strain, the ΦC31 integrase expression plasmid strain, the LP-strain, and the mobilizer strain. A variation on IMCE , reverse IMCE, was also accomplished with the addition of a recombination directionality factor gene to the integrase expression plasmid, and the use of *attL* and *attR* sites. Reverse IMCE was used to create a new LP-strain containing the counter-selectable marker gene *sacB*.

A strategy to assemble overlapping DNA cassettes post-IMCE using genetic crosses was developed. An application of this system was to assemble synthetic operons for the metabolic engineering of *Sinorhizobium meliloti* to produce medium chain length (mcl) polyhydroxyalkanoates (PHA). Two homologues each of *phaC* (PHA synthase gene) and *phaG* (3-hydroxydecanoyl-[acyl-carrier-protein]:CoA transacylase gene) were chosen. The genes were successfully combined in all four possible configurations on the chromosome of *S. meliloti*. It was observed that the *phaC* construct did not integrate into the expected location, and therefore its expression was not guaranteed, but promising preliminary results showing the detection of mcl-PHA were obtained.

The *phaC* genes described in this thesis have also been cloned into plasmids, for their expression in strains carrying *phaG* within the landing pad. These strains will provide valuable tools for future work concerning PHA production. The chromosome engineering tools developed in this thesis will be valuable for any future applications requiring stable maintenance of a synthetic construct on the chromosome of these *Rhizobiales* members.

Acknowledgments

I would like to thank everyone around me that has encouraged me and kept me somewhat sane. I would like to thank my parents for their continued support and encouragement. I would like to thank Trevor Charles for taking me on, for affording me the freedom to work independently, yet providing me with much needed guidance. Thank you to my committee members: Marc Aucoin, Josh Neufeld, Bruce Reed for their guidance through the years.

I would like to thank Paul Reginato for working on RIMCE and troubleshooting some of the initial problems with it. I would like to thank my sister Emily for using her microbiology skills to help me out in the lab for a few months. Thanks to Jiujun Cheng for sharing his endless molecular biology knowledge and his reminders with respect to cloning that "You only need one colony!". Thanks to Ricardo Nordeste and Tam Tran for being great collaborators on the PHA work. Thanks to Kathy Lam for commiserating with me about all the molecular biology weirdness we encountered at the bench. Thanks to Maya D'Alessio for cleaning, even if I got annoyed that I couldn't find things after. Thanks to Ariana Marcassa for keeping the music in the lab interesting, even if I was vocal about not liking some of it. Thank you to all those who came through the Charles Lab during my tenure. I'm pretty sure I learned a thing or two from all of you. It was a pleasure working with you all.

Contents

Author's Declaration . . . iii
Abstract . . . v
Acknowledgments . . . vii
List of Tables . . . xiii
List of Figures . . . xv
List of Abbreviations . . . xvii

1 Literature Review **1**
1.1 Introduction . . . 1
1.1.1 Synthetic Biology . . . 2
1.1.2 Chromosome Engineering . . . 2
1.2 Polyhydroxyalkanoates (PHAs) . . . 3
1.2.1 The *Rhizobiales* . . . 5
1.3 Recombination . . . 5
1.3.1 The Application of Recombination to Chromosome Engineering . . . 9
1.4 *In vitro* Chromosome Synthesis . . . 11
1.5 *In vivo* Chromosome Synthesis and Engineering . . . 13
1.5.1 Chromosome Editing . . . 14
1.5.2 Delivery of Large DNA Fragments into the Chromosome . . . 16
1.5.3 *In vivo* Chromosome Synthesis . . . 18
1.5.4 Chromosome Engineering for Combinatorial Studies . . . 21
1.6 This Work . . . 23
1.6.1 Claims of Contributions to Scientific Knowledge . . . 26

2 Materials and Methods **29**
2.1 Bacterial Strains and Plasmids . . . 29
2.2 Oligonucleotides . . . 42
2.3 Bacterial Culture and Microbiological Techniques . . . 52
2.3.1 Bacterial Growth Media . . . 52
2.3.2 Preparation of Bacterial Frozen Stocks . . . 54
2.4 ΦM12 Transductions . . . 54
2.4.1 Preparation of ΦM12 Phage Lysates . . . 54

2.4.2 Determining Titer of ΦM12 Phage Lysates . . . 54
2.4.3 Transduction . . . 55
2.5 Triparental Mating . . . 55
2.6 Plasmid Construction with λ Red Recombineering . . . 56
2.7 Molecular Cloning Techniques . . . 57
2.7.1 DNA Ligation . . . 57
2.7.2 Agarose Gel Electrophoresis . . . 58
2.7.3 Gel Extraction . . . 59
2.7.4 Preparation of $CaCl_2$ Competent Cells . . . 60
2.7.5 $CaCl_2$ Transformation . . . 61
2.7.6 Electroporation . . . 62
2.7.7 Restriction Enzyme Digest of DNA . . . 62
2.7.8 Polymerase Chain Reaction (PCR) . . . 63
2.7.9 Colony PCR . . . 65
2.7.10 Sanger Sequencing . . . 66
2.7.11 Genomic Sanger Sequencing . . . 66
2.8 DNA Purification . . . 67
2.8.1 Purification of DNA Using a Silica Gel Column . . . 67
2.8.2 Ethanol Precipitation of DNA . . . 67
2.8.3 Miniature Preparation of Bacterial Plasmid DNA . . . 68
2.8.4 Large Scale Preparation of Bacterial Genomic DNA for Direct Genomic Sanger Sequencing . . . 69
2.8.5 Miniature Preparation of Bacterial Genomic DNA . . . 71
2.9 Southern Blotting . . . 72
2.9.1 Creation and Quantification of DIG Labeled Probe . . . 73
2.9.2 Transfer of DNA onto Membrane . . . 75
2.9.3 Hybridization of the Probe . . . 76
2.9.4 Detection of the Probe . . . 77
2.10 β-Glucuronidase (GusA) Assay . . . 78
2.11 Sequence Alignments, Genomic and Plasmid Maps . . . 78

3 Integrase Mediated Cassette Exchange **79**
3.1 Introduction . . . 79
3.2 Materials and Methods . . . 84
3.2.1 Construction of SmUW227 LP-Strain . . . 84
3.2.2 Construction of Ow0013 LP-Strain . . . 85
3.2.3 Construction of At11142 LP-Strain . . . 86
3.2.4 Construction of pJC2 ΦC31 Integrase Expression Plasmid . . . 87
3.2.5 Construction of pJH110 IMCE Donor Plasmid . . . 87
3.2.6 Construction of pJH122 IMCE Donor Plasmid . . . 88
3.2.7 Southern Blotting . . . 88

3.2.8 IMCE Tetraparental Mating . . . 89
3.2.9 Colony PCR . . . 92
3.2.10 Visualization of RFP Fluorescence . . . 92
3.3 Results . . . 93
3.3.1 LP-strain Confirmation . . . 93
3.3.2 IMCE Tetraparental Mating Process . . . 98
3.3.3 IMCE Efficiency . . . 105
3.3.4 IMCE Confirmation by PCR . . . 109
3.4 Discussion . . . 111

4 Integrase Mediated Cassette Exchange Tools and Reverse Integrase Mediated Cassette Exchange 119
4.1 Introduction . . . 119
4.2 Materials and Methods . . . 122
4.2.1 Construction of pJH145: Mobilizable *attB* Donor Fosmid . . . 122
4.2.2 Construction of pJH121: Gentamicin *attB* Donor Vector with Gateway . . . 122
4.2.3 Construction of pJH123 and pJH124: Hygromycin *attB* Donor Vectors with Gateway . . . 123
4.2.4 Construction of pJH141 . . . 123
4.2.5 Construction of pJH143 . . . 124
4.2.6 *In vivo* Gateway Recombination . . . 124
4.2.7 RIMCE Tetraparental Mating . . . 125
4.2.8 IMCE using SacB LP-strain . . . 125
4.3 Results . . . 126
4.3.1 RIMCE . . . 126
4.3.2 SacB IMCE . . . 130
4.3.3 Gateway Recombination into IMCE Vector . . . 134
4.3.4 Test IMCE with pJH145 . . . 136
4.4 Discussion . . . 138

5 Recombination Synthesis of PHA Genes 143
5.1 Introduction . . . 143
5.2 Materials and Methods . . . 149
5.2.1 Construction of pJRT8 . . . 149
5.2.2 Construction of pJRT9 . . . 150
5.2.3 Construction of pJRT10 . . . 150
5.2.4 Construction of pJRT12 . . . 151
5.2.5 Construction of pJRT15 . . . 151
5.2.6 Construction of pJRT16 . . . 151
5.2.7 Construction of pJH146 . . . 152

5.2.8 Construction of pJH147 . . . 152
5.2.9 Construction of pJH148, pJH149, and pJH150 . . . 152
5.2.10 Construction of pJH151 . . . 153
5.2.11 SmUW235 ϕ SmUW227 . . . 153
5.2.12 SmUW497 ϕ SmUW227 . . . 153
5.2.13 IMCE . . . 154
5.2.14 RS Triparental Mating . . . 154
5.2.15 RS Transduction . . . 155
5.2.16 Reporter, and Fluorescent Dye Visualization . . . 156
5.2.17 Colony PCR and Sequencing . . . 157
5.2.18 Direct Genomic Sanger Sequencing . . . 157
5.2.19 Transduction Linkage Mapping . . . 158
5.2.20 Cell PHA Content Analysis Gas Chromatography and Mass Spectrometry (CG-MS) . . . 158
5.3 Results . . . 160
5.3.1 RS Transduction . . . 160
5.3.2 RS Conjugation . . . 163
5.3.3 *In vitro* Assembly of PHA Operons . . . 172
5.3.4 IMCE strains Containing Plasmids . . . 174
5.4 Discussion . . . 174

6 General Discussion **181**

Bibliography **191**

Appendix A Supplementary Figures **205**

Appendix B Hyaluronic Acid Project **211**
B.1 Introduction . . . 211
B.2 Materials and Methods . . . 212
B.2.1 Construction of Hyaluronic Acid Operon . . . 212
B.2.2 Oligonucleotides . . . 214
B.2.3 Conclusion . . . 214

List of Tables

2.1 Bacterial strains and plasmids 30
2.2 Oligonucleotides 42
2.3 KOD Xtreme PCR 64
2.4 KOD Hot Start PCR 65

3.1 IMCE selection 91
3.2 IMCE efficiency of each landing pad strain 108

List of Figures

1.1 Integrase mediated cassette exchange 25
1.2 Recombination synthesis assembling two halves of an operon. 25

3.1 ΦC31 integrase mechanism 82
3.2 The ΦC31 integrase mechanism showing only stylized DNA sites. The arm sequences (B,B′, P, P′) surround a core dinucleotide (TT). The base-pairing of the core dinucleotide is required for the *att* sites to be ligated in the recombinant configuration as shown below the arrow. 82
3.3 Integrase mediated cassette exchange 83
3.4 SmUW227 Southern blot 95
3.5 Ow0013 Southern blot 97
3.6 At11142 Southern blot 99
3.7 SmUW227 landing pad locus map 101
3.8 Ow0013 landing pad locus map 101
3.9 At11142 landing pad locus map 101
3.10 pJH110 map 102
3.11 pJH122 map 103
3.12 pJC2 map 104
3.13 IMCE tetraparental mating 106
3.14 SmUW227 IMCE experiment 107
3.15 colony PCR of IMCE derived clones 110

4.1 SacB landing pad map 127
4.2 SmUW246 reverse landing pad map 127
4.3 pJH141 map 128
4.4 RIMCE tetraparental mating 131
4.5 SacB IMCE sucrose selection 133
4.6 pJH121 map 135
4.7 pJH145 map 137

5.1 Recombination Synthesis 146
5.2 mcl-PHA production from acetyl-CoA 147

5.3 RS of PHA half operons 148
5.4 Integrated JRT vectors 161
5.5 *phaC* alignment 161
5.6 *phaG* alignment 162
5.7 RS PCR Amplicons 167
5.8 XylE Reporter Test for pJRT14, pJRT10, and pJRT12 168
5.9 Nile Red Fluorecence Phenotype of IMCE and RS strains 169
5.10 RS colony PCR 170
5.11 GC-MS putative PHO Peak 171
5.12 *In vitro* assembled operon clone displaying GFP fluorescence 173

A.1 At1142 blot measurement 206
A.2 Ow0013 blot measurement 207
A.3 Standard for Ow0013 and At11142 Southern blot 208
A.4 Standard for SmUW227 Southern blot 209

List of Abbreviations

Φ	phage
attB	attachment site bacteria
attL	attachment site left
attP	attachment site phage
attR	attachment site right
att	attachment site
*att*Tn7	transposon seven attachment site
μ	micro
oriT	origin of transfer
rfp	red fluorescent protein gene
sacB	levansucrase gene
in vitro	in glass
in vivo	within the living
BAC	bacterial artificial chromosome
BdhA	3-hydroxybutyrate dehydrogenase
bp	base pair
CAGE	conjugative assembly genome engineering
CAS	CRISPR associated
CAS9	CRISPR associated protein nine
CoA	coenzyme-A

CRISPR	clustered regularly interspaced short palindromic repeats
DAP	diaminopimelic acid
DMF	dimethylformamide
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
dNTPs	deoxynucleoside triphosphates
ds	double stranded
DSB	double strand break
DXP	deoxyxylulose phosphate
ECL	enhanced chemiluminescence
EDTA	ethylenediaminetetraacetic acid
FAME	fatty acid methyl ester
g	gram
GC	gas chromatography
GC-MS	gas chromatorgraphy with mass spectrometry detection
Gm	gentamicin
GusA	β-Glucuronidase
ICE	integrative and conjugative element
IMCE	integrase mediated cassette exchange
Int	integrase
kb	kilobase pair
LB	lysogeny broth
LBmc	lysogeny broth with magnesium and calcium
M	molar
m	milli

MAGE	multiplex automated genome engineering
mb	megabase pair
mcl	medium chain length
mM	millimolar
MS	mass spectrometry
NHEJ	nonhomologous end joining
nm	nanometer
OD	optical density
OD_{600}	optical density at 600 nm
ORF	open reading frame
p	plasmid
PCR	polymerase chain reaction
PFU	plaque forming units
PHA	polyhydroxyalkanoate
PhaC	PHA synthase
PhaG	3-hydroxydecanoyl-[acyl-carrier-protein]:CoA transacylase
PhaZ	polyhydroxyalkanoate depolymerase
PHB	polyhdroxybutyrate
PhbA	β-ketothiolase
PhbB	acetoacetyl-CoA reductase
PhbC	PHA synthase
PHO	polyhydroxyoctanoate
PHV	polyhydroxyvalerate
PNPG	4-nitrophenyl-beta-D-glucopyranoside
RBS	ribosome binding site

RDF	recombination directionality factor
RFP	red fluorescent protein
RIMCE	reverse integrase mediated cassette exchange
RNA	ribonucleic acid
RS	recombination synthesis
S	Svedberg unit
SacB	levansucrase
scl	short chain length
SDS	sodium dodecyl sulphate
Sm	streptomycin
SOB	super optimal broth
SOC	super optimal broth with catabolite repression
Sp	spectinomycin
ss	single stranded
TALEN	transcription activator like effector nuclease
Tn	transposon
TRMR	trackable multiplex recombineering
TY	tryptone yeast extract media
UV	ultraviolet light
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
X-gluc	bromo-4-chloro-3-indolyl-β-D-glucuronide
ZFN	zinc finger nuclease

Chapter 1

Literature Review

1.1 Introduction

Recombinant DNA has provided a foundation for many important fields of research and technologies, including genome sequencing and protein purification. Synthetic DNA technology allowed for the creation of oligonucleotides used for PCR primers, and the synthesis of small genes. Recombinant DNA and synthetic DNA technologies have together formed a foundation for molecular biology in general. The application of molecular cloning technologies has had a positive economic impact (Demain, 2000). An early example of metabolic engineering via recombinant DNA was the production of the dye indigo through the expression of a *Pseudomonas putida* gene in *Escherichia coli* (Ensley *et al.*, 1983). In recent years, the cost of DNA synthesis has fallen dramatically. Coupled with the advancements in cell and molecular biology, this drop in cost has increased the scope of strain construction and design. Although the increased scope allows for more complex strain design, the overwhelming increase in complexity requires intellectual tools to make the system tractable; the intellectual tool-kit is synthetic biology. Synthetic Biology promises to amplify the benefits of synthetic and recombinant DNA technologies through the systematic approach of top-down engi-

neering disciplines where the use of well-characterized components makes for predictable designs (Baker *et al.*, 2006).

1.1.1 Synthetic Biology

Synthetic biologists aim to create cells with specific engineered functions using well characterized parts. Parts are DNA sequences of defined function that can be abstracted to different levels of detail. A part to produce protein would be built up of the required subparts (promoter, ribosome binding site, open reading frame) but it could be simply thought of as a protein generator. Under this paradigm, the genes of whole pathways can be grouped into a single part or device. Such a level of abstraction would allow for more complex genetic designs that are still tractable. Also, new composite parts should be more predictable than naively designed constructs if the component parts are well characterized. The BioBrick standard (Shetty *et al.*, 2008) has been developed to deal with a parts-based paradigm. The BioBrick standard is well suited to plasmid-based engineering and is currently the most widely adopted synthetic biology assembly standard. It lays out a standard restriction site scheme that allows for assembly of two BioBricks into a new BioBrick that may be assembled with every other BioBrick. However, as the synthetic biology paradigm matures and parts become larger and more complex, the chromosome of bacterial cells will provide a more stable platform for synthetic biology than plasmids.

1.1.2 Chromosome Engineering

Chromosome engineering affords stable maintenance of large amounts of heterologous DNA at a consistent copy number. This feature is an advantage over plasmid based systems. Many plasmids are unstable without selective pressure. Moreover cloning genes in plasmids may alter their copy number through inadvertently chang-

ing the expression of plasmid encoded replication factors (del Solar *et al.*, 1998). Chromosome engineering not only includes adding DNA to existing chromosomes, and synthesizing new chromosomes, but also deleting DNA from existing chromosomes and combining parts of synthetic chromosomes. As more becomes known about gene functions and their implications for cell biology it will become possible through synthetic biology to design chromosomes or parts of chromosomes. Methods to synthesize whole chromosomes have been demonstrated (Gibson *et al.*, 2008a,b, 2010), yet our rational design ability is not advanced enough to take advantage of the ability to synthesize whole chromosomes.

What follows is a review of emerging chromosome engineering methods, the advances in DNA synthesis, and the recombination principles that chromosome engineering approaches use to get the synthetic DNA into chromosomes. Also included is a short review of the polyhydroxyalkanoate (PHA) bioplastics, and a short introduction to the *Rhizobiales*. PHA production in members of the *Rhizobiales* is the first application of the chromosome engineering system presented in this thesis.

1.2 Polyhydroxyalkanoates (PHAs)

PHAs are carbon storage polymers that bacteria produce under metabolic stress, such as conditions of excess carbon and limiting nitrogen (Anderson and Dawes, 1990). PHAs accumulate via a variety of pathways. The best studied pathway for PHA accumulation involves 3 enzymes. The enzymes are β-ketothiolase (PhbA), acetoacetyl-CoA reductase (PhbB), that produce hydroxybutyryl-CoA, and then a PHA synthase (PhbC, also known as PhaC), which polymerizes hydroxybutyryl-CoA into polyhydroxybutyrate (PHB). The cycle is completed by a depolymerase (PhaZ) and a dehydrogenase (BdhA) (Aneja *et al.*, 2005) to facilitate the use of the stored PHB by the cell (Anderson and Dawes, 1990; Slater *et al.*, 1988; Tombolini *et al.*, 1995). Other PHA path-

ways use Co-enzyme A (CoA) linked monomers from other areas of cell metabolism to feed a PHA synthase (Anderson and Dawes, 1990). These pathways include β-oxidation and *de novo* fatty acid synthesis. CoA-linked monomers are produced via β-oxidation when cells are fed fatty acids (Chen *et al.*, 2006). CoA-linked monomers may also be produced from *de novo* fatty acid metabolism with a (R)-3-hydroxydecanol-ACP:CoA transacylase gene (*phaG*). Through the expression of *phaG* from *Pseudomonas putida* and broad substrate specificity *pha*Cs from *Pseudomonas stutzeri* (*pha*C_1, *pha*C_2) in *Ralstonia eutropha* mcl-scl co-polymer PHAs were produced (Chen *et al.*, 2006).

The practical uses of PHAs include the use as thermoplastics for moulding and fibres, a source of monomers for fuel additives and chiral chemistry, and drug delivery agents (Chen, 2009). Depending on the specific hydroxyalkanoate monomers incorporated into the chain, bioplastics with different physical properties are produced. There are a variety of PHA polymers that are produced in bacteria. The chain length and the side chain length of the monomer have an impact on the material's physical properties. For example: *Sinorhizobium meliloti* produces mainly polyhydroxybutyrate (PHB) and some polyhydroxyvalerate (PHV); both are short chain length PHAs (Tombolini *et al.*, 1995). Some species of *Pseudomonas* produce medium chain length (C6-C14) PHA (Anderson and Dawes, 1990). PHB can be a brittle material below its glass transition temperature of 10^{o}C whereas a copolymer (PHB-co-HV) containing 20 mol% valerate will remain more pliable down to its glass transition temperature of -1^{o}C (Marchessault and Yu, 2001). Many medium chain length PHAs possess rubber-like properties similar to elastomers (Kellerhals *et al.*, 2000).

Shuffling of orthologues of genes linked to PHA synthesis may result in strains that produce PHAs with different properties. A combinatorial PHA gene expression approach coupled with a screen sensitive to PHAs of different chain lengths could be useful for PHA discovery. A combinatorial assembly approach could create operons

containing different *phaG*, *phaC*$_1$ and *phaC*$_2$ orthologues. The operons could be constructed in different *pha* gene backgrounds, providing varied sources of monomers. This could yield a variety of PHAs containing both short-chain and medium-chain monomers.

1.2.1 The *Rhizobiales*

The *Rhizobiales* are a diverse order of the *α-proteobacteria*. They include the nitrogen fixing nodule forming Rhizobia, such as *S. meliloti*, a natural PHA producer. Many intracellular pathogens, including *Brucella abortus*, count in the ranks of the *Rhizobiales*; *Ochrobactrum anthropi* (Chain *et al.*, 2011) is an opportunistic pathogen very closely related to *Brucella*. The plant pathogen *Agrobacterium tumefaciens*, famous for its use in the genetic modification of plants, is also a member of the *Rhizobiales*. Of note, the *Rhizobiales* also include the methanotroph *Methylosinus trichosporium*. Instances of horizontal gene transfer are very common among the *Rhizobiales* (Carvalho *et al.*, 2010). The presence of megaplasmids or extra auxiliary chromosomes is also common among the *Rhizobiales* (DiCenzo and Finan, 2015; Kaneko *et al.*, 2000; Chain *et al.*, 2011).

Chromosome engineering of the *Rhizobiales* may lead to many unforeseen applications due to the broad functional diversity of this group.

1.3 Recombination

DNA Recombination is simply the joining or un-joining of DNA strands that were previously in a different configuration. Technically speaking, cloning with restriction enzymes and plasmids is a form of recombination, hence the term recombinant DNA. However, *in vitro* cloning, even with a perfect restriction enzyme, is not practical for engineering chromosomal DNA, due to the fragile nature of high-molecular

weight DNA. *In vivo* forms of DNA recombination are quite useful for chromosome engineering. These forms of recombination are usually what is meant when referring to recombination. Functions of these mechanisms include: fostering genetic diversity, DNA repair, and viral infection processes. For the purposes of this review I will focus mostly on forms of recombination that involve recognition of specific unique sequences, either through homology or protein DNA interactions. These forms include general homologous recombination, λ Red recombination, and various phage based site-specific recombination systems. An exception will be the involvement of nonhomologous end joining (NHEJ) in its capacity as a step in some emerging chromosome engineering techniques.

Unlike plasmids, chromosomes cannot be effectively engineered using restriction enzymes, because chromosomes cannot be easily handled *in vitro*, and their large size makes unique restriction sites exceedingly rare. Therefore, chromosomes must be engineered via recombination, usually with an altered plasmid or phage genome. For the purpose of developing an understanding of recombination as it pertains to chromosome engineering, two distinct types of recombination will be discussed: homologous recombination and site-specific recombination. General homologous recombination, reviewed by Otero *et al.* (1995), involves recombination between two homologous strands of double-stranded DNA catalyzed by a host of proteins; a critical one is RecA. In contrast, site specific recombination, reviewed by Grindley *et al.* (2006), is catalyzed by one or a few enzymes that directly recognize specific sites on the DNA to be recombined. There is also the λ Red system (Ellis *et al.*, 2001; Poteete, 2001) which is a homologous recombination system from phage λ. The λ Red system works with small homologous regions (50 bp, compare with $>$200 bp for regular homolous recombination), recombines linear and single stranded DNA, and is RecA independent.

Some form of homologous recombination is active in most if not all bacteria en-

countered, due to its role in DNA repair. Recombination through the RecBCD pathway, a specific mechanism for homologous recombination, involves formation of a single stranded nucleoprotein filament, the presynaptic filament, which invades a homologous strand to form a heteroduplex region. This is followed by migration of the heteroduplux region, formation of a Holliday junction, isomerization of the junction, and resolution of the junction to form recombinant products (Otero *et al.*, 1995). The activity of RecA is very important for the functioning of the pathway since it stabilizes the single stranded DNA in the presynaptic filament and it catalyzes the formation of the heteroduplex region. RecA is also important for other mechanisms of homologous recombination that involve formation of a heteroduplex region, such as the RecF pathway (Otero *et al.*, 1995). Therefore, it is noteworthy that many strains of *Escherichia coli* such as DH5α (Hanahan, 1983) cannot perform homologous recombination due to their *recA* mutations. The mutations are advantageous from a plasmid cloning perspective, but abolish the strains' usefulness for chromosome engineering using wild-type homologous recombination. However, λ Red may still be used in *recA* backgrounds. λ Red relies on three λ phage genes (*exo, bet, gam*) that allow efficient recombination of DNA segments with $\approx$ 40-50 bp of homology. λ Red works well in *E. coli* and has been used in other proteobacteria with reduced efficiency (Lesic and Rahme, 2008). The use of λ Red for chromosome engineering is often called recombineering, a term coined by Ellis *et al.* (2001). It is important to note that for homologous recombination (including λ Red) to occur, the two homologous sites (unless palindromic) must be aligned parallel to each other. This constraint causes the recombination result to be consistent when the orientation of the homologous sites is a certain way. If the sites are parallel to each other on the same duplex, the result of homologous recombination between the two sites will be the excision of the DNA between the two sites and the reduction of the copy number of the site to one. If the sites were anti-

parallel on the same duplex, the recombination result would be an inversion of the DNA between the two sites and conservation of both homologous site copies. If the alignment constraint did not exist (it does not in some cases of site specific recombination) there could be an even mixture of both events. An example of the implications of the alignment constraint of recombination can be seen in work with large serine recombinases (Ghosh *et al.*, 2003). In this study, a particular part of the recombination sites was made palindromic, causing mixtures of excision and inversion results when recombination was carried out.

The site-specific recombinases are split into two families: the tyrosine recombinases (Nunes-Duby *et al.*, 1998) and the serine recombinases (Smith and Thorpe, 2002). Tyrosine recombinases include the λ integrase used in Invitrogen's Gateway technology (Mumm *et al.*, 2006; Katzen, 2007) and the Cre recombinase (Sternberg and Hamilton, 1981). The serine recombinases include the ΦC31 integrase (Kushtoss and Rao, 1991) and the resolvase protein encoded by transposons Tn3 and $\gamma\delta$ (Smith and Thorpe, 2002). The mechanisms of the tyrosine and serine recombinases are markedly different. The tyrosine recombinase mechanism involves the formation of short heteroduplex regions (RecA independent), formation of a Holliday junction, isomerization, and resolution. The serine recombinase mechanism involves double strand breakage, rotation, and re-ligation (Grindley *et al.*, 2006). Regardless of the mechanistic differences between the two families, there are often conceptual parallels between the two with respect to the requirements for recombination, and the recombination site configurations that yield integration, inversion, and excision. For that reason, the nomenclature of the sites involved in the recombination process may be the same for enzymes in both families. ΦC31 integrase and λ integrase both recombine *attB* and *attP* (for attachment bacteria and attachment phage) and produce *attL* and *attR* composite sites; although, the λ *att* sites are completely different than the ΦC31 *att* sites.

On the other hand, the Cre recombinase (tyrosine recombinase) recombines *loxP* sites to form *loxP* sites. Given that the product of recombination is the same as the initial recombination site (*loxP*), DNA integrated with the Crelox system is unstable. The λ system produces stable recombinant products in the absence of its cognate excisionase (Xis) (Mumm *et al.*, 2006), yet integration host factor (IHF) is required, so the λ system is less portable than other systems. The ΦC31 system is particularly well suited to chromosome engineering because its recombination products are stable, and it requires no co-factors. The ΦC31 system works in a variety of hosts including *Drosophila*, human cell lines, bacteria, and yeast (Bateman and Wu, 2008; Dafhnis-Calas *et al.*, 2005; Thomason *et al.*, 2001).

1.3.1 The Application of Recombination to Chromosome Engineering

Homologous recombination is used when precise modifications to a chromosome need to be made. The advantage of homologous recombination is that the target can be very specific and there will be no surrounding mutations or recombinase sites present when the modification is complete. Homologous recombination (including recombineering) may be used to insert constructs, delete genes, to make mutations, and gene fusions. A well used vector system, which takes advantage of homologous recombination in many species, is the SacB system. There are a great deal of vectors containing the *Bacillus subtilus* or *Bacillus amyloliquefaciens sacB* gene, encoding levansucrase, (Schafer *et al.*, 1994; Gay *et al.*, 1985; Bramucci and Nagarajan, 1996) for positive selection of double recombinant clones in many bacteria. This particular levansucrase is toxic to many bacteria when sucrose is present. To insert or delete DNA from the bacterial chromosome, homologous DNA from the target locus is cloned into the vector. In the case of making an insertion, it is cloned flanking the DNA to be inserted.

The vector is introduced into the target host, where the vector cannot replicate, and the host cells are plated on selective media containing the antibiotic for which the vector backbone encodes resistance, thus selecting for vector-chromosome co-integrants. Co-integrant clones are plated on media containing sucrose, thus selecting for the loss or inactivation of *sacB*. Sucrose-sensitive clones are screened for the correct antibiotic sensitivity, to distinguish double recombinants from *sacB* mutants. The double recombinants are then screened phenotypically, with PCR or Southern analysis, to determine if they are wild-type revertants or the desired insertion or deletion mutants.

Site-specific recombination is typically much more efficient than homologous recombination (Mumm *et al.*, 2006; Groth and Calos, 2004; Thomason *et al.*, 2001; Ghosh *et al.*, 2003). The use of site-specific recombinases for chromosome engineering is appealing, however strains containing the DNA sequence specific to the recombinase must be constructed in many cases. Homologous recombination can be used to insert these sequences at specific loci of the chromosome. Some site-specific systems do have existing target sites within certain phyla. The *att*Tn7 sites are present downstream of the *glmS* gene in a number of *γ-proteobacteria* (Choi and Schweizer, 2006b). The Tn7 system is based on a transposase that uses a cut and paste mechanism that is unrelated to the serine or tyrosine recombinases (Biery *et al.*, 2000). Another system that has a similar natural host range to Tn7 is the integrative conjugative element (ICE) *clc*, which is based on a tyrosine recombinase (Miyazaki and van der Meer, 2013). If the goal is to modify organisms that fall outside the *γ-proteobacteria*, the fact remains that recipient strains will have to be engineered. In the long term it makes more sense for recipient strains to be constructed rather than found. This way some of the same tools may be used in divergent organisms.

Before considering how to integrate the proper recombinase sites into the target chromosome, an idea will be needed of how the system will function, what sites to

add, and which selection schemes to use. In the following sections various chromosome engineering schemes illustrating these principles will be discussed.

1.4 *In vitro* Chromosome Synthesis

In vivo recombination is an integral part of chromosome engineering as a whole, but it is important to discuss the advances in *in vitro* DNA synthesis. *In vitro* DNA synthesis provides a means to make a DNA molecule that has no existing template in nature. The method provides the flexibility to codon optimize genes and/or to rearrange genes or sections of genes.

Synthesized genes have become an easily obtainable commodity thanks the advancement of gene synthesis through phosphoramidite-based oligonucleotide synthesis. It is certainly remarkable how accessible synthetic DNA has become considering the first synthesis of a gene segment (77 bp of a yeast transfer RNA gene) by Khorana *et al.* (1972) was such a monumental achievement in its time that the experimental realization of it was released as a series of twelve papers. Since then, the complete chemical synthesis of the polio virus cDNA genome (7.5 kb) (Cello *et al.*, 2002), the bacteriophage ΦX174 genome (5.3 kb) (Smith *et al.*, 2003), and quarter genome sections (144 kb) of *Mycoplasma genitalium* have been synthesized (Gibson *et al.*, 2008a) and assembled *in vitro* using a combination of oligonucleotide synthesis and *in vitro* enzymatic assembly or polymerase extension. The synthesis of the polio cDNA genome (Cello *et al.*, 2002) was quite a large undertaking. Overlapping phosphorylated oligonucleotides of an average length of 69 nucleotides (nt) were assembled using ligation and cloned into vectors as 400-600 bp fragments which overlapped with their to be adjacent fragments. A total of 5-15 clones were sequenced to identify error free clones, or clones that could be readily corrected with PCR-based site-directed mutagenesis. The correct 400-600 bp clones were assembled into two of the three larger fragments (fragments

1 and 3). The authors do not mention how the cloned and sequenced segments were excised from the vector and used in the subsequent assembly of fragments. Fragment 3 was assembled using similar oligonucleotides but used a PCR to generate the 600 bp segments, which were cloned into a T-vector, sequenced, and subsequently assembled by restriction and ligation. The three large fragments were assembled via restriction digest and ligation to yield the full cDNA copy of the genome. The ΦX174 synthesis project (Smith *et al.*, 2003) aimed to reduce the resources needed for quality control of the sequence, so that a correct sequence could be made more quickly and economically. They chose to skip the intermediate cloning and sequencing steps, and instead increase the efficiency of their assembly process. Like Cello *et al.* (2002), they annealed and ligated overlapping phosphorylated oligos, except they performed their ligation at 55°C using *Taq* ligase. The theory is that if the ligation temperature were increased, there would be a stronger selection against mismatches in the overlapping regions, due to higher stringency preventing mutated intermediates from ligating. The entire ΦX174 genome was not assembled by ligation, even though all of the oligos were pooled and included in the ligation reaction. The authors' statistical model predicted that random variation in the concentration of each given oligo of 20% would make the probability of achieving a full length ligation product exceedingly low. As a result, they used a PCR-like reaction called polymerize cycling assembly (PCA), where the ligation products were extended to generate full length product. Synthetic DNA molecules were circularized and introduced into *Escherichia coli*. The DNA from the plaques that arose was sequenced and they found that the frequency of correct genomes was close to what they had estimated. The fidelity of their method would produce 1 kb of DNA with an average of ≈2 errors. In their experiments, phage viability was a selection for correctness of the sequence. Without a selection, DNA sequencing must be done to verify the correctness of the final sequence. As the size of

the synthetic DNA increases, so does the amount of clones that need to be sequenced to find a correct one. Gibson *et al.* (2008a) assembled quarter-length segments of the *M. genitalium* genome via *in vitro* recombination of synthetic fragments. These were commercially synthesized 5-7 kb fragments, presumably made in a similar manner as the ΦX174 genome, that were assembled in parallel to produce intermediate assemblies. Eventually, the last intermediate assemblies were assembled into the final genome, but only 1/4 length segments were achieved *in vitro*. To achieve *in vitro* recombination, each synthetic fragment contained a region of overlap. A custom BAC was made, using PCR, with complementary regions to the ends of the assembly. The BAC was included in the assembly reaction with the fragments. The fragments and the BAC were subjected to a 3' exonuclease digestion, exposing their sticky ends, and allowing them to anneal. The assembly was completed by filling in gaps with *Taq* polymerase and sealing with *Taq* ligase. The final assemblies were married by homologous recombination in *Saccharomyces cerevisiae*. Gibson *et al.* approached the limit of *in vitro* assembly. Blake *et al.* (2010) have proposed a method called pairwise selection assembly (PSA) that was used to assemble a 91-kb chromosome arm from *S. cerevisiae* that relies on *in vitro* reactions, but intermediates are selected *in vivo* through the assemblies' activation of antibiotic resistance gene expression in the carrier plasmid. At this time, it is quite obvious that taking advantage of *in vivo* recombination for synthetic DNA assembly is necessary at some point and, when coupled with the technology to cheaply synthesize shorter ΦX174 genome sized segments, *in vivo* assembly can be quite powerful.

1.5 *In vivo* Chromosome Synthesis and Engineering

DNA in a tube is, after all, just DNA. Therefore it is important to have robust protocols for introducing DNA into cells, and for reasons already mentioned, integrate

the DNA into a chromosome. Depending on the construct, it may also be important to make modifications to existing chromosomes to make the cell work well with synthetic DNA being introduced. Another approach, famously taken by Gibson *et al.* (2010), is to build the whole chromosome elsewhere (*in vivo* or *in vitro* in a different cell) and import the new chromosome into a cell, thereby replacing the cell's existing chromosome. Two main paradigms for chromosome building to facilitate synthetic biology are emerging. One being: building a chromosome in place *in vivo* sequentially; the other being: building a chromosome outside of its intended destination and using the synthetic chromosome to replace the genome of an existing cell, by introducing the synthetic genome. However, chromosome engineering encompasses more than just chromosome building. There are also methods emerging that focus more on directed evolution and combinatorial development of existing chromosomes such as MAGE (Wang *et al.*, 2009) and TRMR (Warner *et al.*, 2010). The chromosome can also be used in the place of a cloning vector for large insert functional screening, or synthetic biology applications.

1.5.1 Chromosome Editing

Many chromosome engineering strategies take advantage of the cell's DNA repair capabilities. One way to increase the cell's DNA repair activity is to damage the DNA, and a particularly dire form of damage is the double strand break (DSB). DSBs can be repaired through non-homologous end joining or through the use of a homologous template. In order to accomplish a chromosome engineering objective, a targeted DSB could be made, and a homologous DNA containing an edit could be supplied to serve as a template in the required DNA repair. The resultant cell then contains the desired chromosomal change included in the homologous DNA. Until recently the above scenario would be nearly impossible except for certain sites targeted by a

homing endonuclease. Now a specific endonuclease can be created that will cut at a desired target site. There are three types of nucleases that allow custom targeting: zinc-finger nucleases (ZFNs) (Carroll, 2011), transcription activator like effector nucleases (TALENs) (Sanjana *et al.*, 2012), and CRISPR associated protein endonucleases (CAS9) (Wiedenheft *et al.*, 2009; Jinek *et al.*, 2012; Jiang *et al.*, 2013). Each type works on a similar principle where a sequence independent endonuclease domain is steered to a DNA target by a custom domain or RNA. In the case of ZFNs and TALENS, it is a custom protein domain; for CAS9 it is a custom guide RNA. The use of ZFNs, TALENS and CAS9 for genome editing has been recently reviewed. (Gaj *et al.*, 2013). Chromosome editing, or genome editing, is the term that has emerged for the general approach of making a DSB with a custom nuclease and then providing a repair template to make desired changes or relying on non-homologous end joining for error prone repair to make mutations.

To date, the primary application of chromosome editing has been the editing of eukaryotic chromosomes. It seems that there is a sufficient base of methods for bacterial chromosome engineering that the new chromosome editing techniques have made less of an impact in the bacterial sphere. For instance, recombineering techniques provide analogous capabilities and results in *Escherichia coli* when compared to chromosome editing. The slow adoption of chromosome editing techniques in the prokaryotic sphere of research does not mean there is not quite a bit of potential in chromosome editing techniques for bacterial chromosome engineering.

Jiang *et al.* (2013) report using CAS9 to edit bacterial chromosomes. They demonstrated markerless mutations in *E. coli* and *Streptococcus pneumoniae*. They demonstrated the use of CAS9 in these distantly related bacterial backgrounds, thereby demonstrating that CAS9 will be useful in a wide range of bacterial phyla. In one of their reported methods using *E. coli*, recombineering was used as the primary means of

genome editing, and CAS9 was used to target the non-edited DNA such that cells that had not undergone successful λ-Red recombination were killed by CAS9 targeting. This CAS9 targeting offered a means to select for successfully recombineered cells without the use of a selectable marker. They also demonstrated genome editing as characterized above. Perhaps the main constraint on the use of CAS9 they discuss is that the edited DNA must be sufficiently different in order to prevent CAS9 targeting, a DSB, and subsequent cell death. There are certain regions of the targeting RNA where a single-nucleotide change will abolish targeting. There are other regions where targeting will be minimally affected by a single nucleotide change. Therefore, making small changes is more difficult than making big changes, like deletions and insertions that encompass whole genes. When single nucleotide changes are to be made, it is imperative that the change be in the protospacer adjacent motif (PAM). The PAM is a 3 bp site (NGG) that is required for CAS9 targeting. The other constraint is that CAS9 targeting is only effective adjacent to NGG sequences. This constraint could be an issue in AT rich regions.

CAS9, and the other custom nucleases, have some drawbacks, but they will undoubtedly offer a powerful set of capabilities to the bacterial chromosome engineering process. The DNA targeting technology also has the potential to contribute to synthetic biology in general by fusing the targeting domains to proteins other than nucleases.

1.5.2 Delivery of Large DNA Fragments into the Chromosome

The state of our understanding of genomes is not to a point where we may rationally design whole genomes, yet some ambitious ideas for the application of synthetic biology have been proposed. In their review, Thaker and Wright (2012) proposed a synthetic biology approach to generating libraries of candidate antibiotics through

mixing and matching genes for synthesis of the scaffold or core molecule with genes for accessorizing the scaffold. Antibiotic synthesis gene clusters can exceed 100 kb in size (Banik and Brady, 2008) and therefore they exceed the stable capacity of cloning vectors, except for BACs. Large gene cluster constructs can benefit from the stability and consistency chromosomal insertion provides. By extension, chromosomal insertion may also lend itself to functional screening of large insert libraries.

A few strategies for inserting segments into bacterial chromosomes have been reported. The aforementioned recombineering strategies could conceivably be used. One drawback to a recombineering approach using λ Red is the specificity of targeting. If λ Red were to be implemented in a general system, a stretch of DNA with sufficient homology occuring in many bacteria would have to be found. A potential difficulty with this approach is the fact that well conserved genes are often essential. Therefore, integration may be more or less host-specific depending on the chosen non-essential target. If future flexibility is desired, and this specificity is not desired. One could insert synthetic sequences into the chromosome, to serve as targets for homologous integration by λ Red, but in that case the use of a site specific recombinase would be preferred.

In an attempt to satisfy the need for a broad host range system, elements from the transposon Tn7 have been used. Tn7 systems are useful because they can insert a DNA segment into a fairly well conserved *att*Tn7 site that is present at the end of a well conserved *glmS* gene (Choi *et al.*, 2005; Choi and Schweizer, 2006b). A BLAST (Altschul *et al.*, 1997) search of the NCBI genome database reveals that the *att*Tn7 site is present in many *γ-proteobacteria*; other classes don't seem to be represented. Therefore, the advantages afforded by the Tn7 system having ready-made hosts do not extend to many well studied bacteria.

Another method using parts of the integrative and conjugative element (ICE) clc

from *Pseudomonas knackmussii* B13 has been reported (Miyazaki and van der Meer, 2013). The reasoning behind the use of an ICE clc based system to insert large DNA sequences is that the natural element moves 103 kb of DNA through conjugation and inserts it into an *attB* site at the end of the $tRNA^{gly}$ gene. The $tRNA^{gly}$ gene is not disrupted by this particular recombination event, due to the sequence of the *attB* and *attP* sites. The recombination sites' 18 bp sequences are identical; the result of integration is a DNA segment surrounded by direct repeats (Gaillard *et al.*, 2006). For this reason, the *att* site nomenclature is fairly misleading in the ICE clc system. As is the case with the Tn7 *att*Tn7 site, a BLAST search of the 18 bp sequence reveals that it is present in many *γ-proteobacteria*. Some problematic behaviours of ICE clc have been observed. ICE clc has been observed to be integrated in two separate loci and a total of up to 7 copies due to tandem integration (Ravatn *et al.*, 1998). Tandem integration is possible due to the identical sequence of the ICE clc *att* sites.

The integration of multiple Tn7 elements at one *att*Tn7 locus has not been reported. That occurrence is unlikely given our understanding of Tn7 transposition. The *att*Tn7 site has a different sequence from each of Tn7L and Tn7R sites that flank the transposon (Biery *et al.*, 2000; Waddell and Craig, 1989; Arciszewska *et al.*, 1989). Furthermore, insertion of Tn7 into random loci is reduced by the exclusion of the *tnsE* gene from the helper plasmid used by Choi *et al.* (2005). TnsE promotes random integration at a low frequency whereas TnsD promotes integration at the *att*Tn7 sequence (Waddell and Craig, 1989). It seems that the Tn7 system is the preferable system for use in the *γ-proteobacteria*.

1.5.3 *In vivo* Chromosome Synthesis

There have been a variety of methods proposed to build chromosomes *in vivo* in an iterative way. Holt *et al.* (2007) proposed that a BAC library be made from partially

digested *Haemophilus influenzae* genomic DNA. Then inserts would be end sequenced and mapped to the *H. influenzae* genome. They selected clones to form a "minimal tiling" that contains the whole genome with minimal overlap between clones. The BAC containing the first clone would be propagated and the second piece of the genome would be added on by digesting the second BAC with *Not*I to excise the fragment, purifying the insert with pulse field gel electrophoresis, transforming into *E. coli* carrying the first BAC and recombination with the first BAC via λ Red recombination. This method was not actually demonstrated, rather it was proposed as a way to go about sequential chromosome building *in vivo*. A similar method was used by Itaya *et al.* (2005) to clone the entire genome of *Synechocystis* PCC 6803 into into *Bacillus subtilus*. The method used was called "inchworm elongation". They used native *B. subtilus* homologous recombination to integrate "landing pad sequences" into the *B. subtilus* genome followed by integrating a *Synechocystis* genomic fragment into the landing pad site via another homologous recombination. Then a new landing pad sequence was integrated that was targeted to the end of the growing *Synechocystis* construct, allowing for the subsequent addition of the next *Synechocystis* fragment targeted to the last landing pad sequence. Holt's method would have an advantage over Itaya's method in that it would produce no scar sequences between pieces. Another iterative *in vivo* approach was demonstrated by Dafhnis-Calas *et al.* (2005). Their approach used the combined activity of two site-specific recombinases (Cre and ΦC31) in DT40 chicken cells to assemble a vertibrate centromeric sequence on a human minichromosome. The first plasmid, contained ΦC31 *attB* sites flanking a Cre *loxP* site and a selectable marker. This first plasmid was integrated into a chromosomal *attP* locus engineered to contain a *loxP* site followed by a selectable marker, followed by the *attP* site. The plasmid contains the synthetic chromosome piece. When this integration occured the action of the ΦC31 integrase removed the duplicated *loxP* site and the initial

chromosomal selectable marker. This leaves the chromosome ready to accept a new synthetic chromosome piece and selectable marker. For each iteration, the two markers are alternated between the chromosome and integrating plasmid. This method takes advantage of the efficiency of site-specific recombinases, but also leaves scars between pieces of the assembly. These methods are scalable, and could be useful for strain development activities, where the goal is not to build an entire chromosome, but rather to introduce a gene cluster that exceeds the capacity of cloning vectors.

Gibson *et al.* (2008a) managed to assemble very large fragments *in vitro* but relied on *in vivo* assembly via *Saccharomyces cerevisiae* homologous recombination to assemble the four quarters into the complete *Mycoplasma genitalium* chromosome. Shortly after their first publication, they showed that they could assemble the whole *M. genitalium* genome by transforming the ≈5 kb fragments they received from the DNA synthesis companies into *S. cerevisiae* (Gibson *et al.*, 2008b). The synthetic DNA fragments were designed to overlap with each other and the terminal fragments were designed to overlap with a yeast vector. Because *M. genitalium* grew prohibitively slowly the group decided to continue their work using *Mycoplasma capricolum* and *Mycoplasma mycoides* instead. Previous work at the Venter institute had already shown that it was possible to take a purified intact *M. mycoides* genome, transplant that into *M. capricolum* genome and change the species of the recipient cell from *M. capricolum* to *M. mycoides* (Lartigue *et al.*, 2007). They improved methods to purify the synthetic DNA from *S. cerevisiae* and solved the restriction problems that prevented the synthetic DNA from entering the recipient: *M. capricolum*. Recently they managed to combine all of these methods to change the species of a recipient *M. capricolum* to *M. mycoides* with a synthetic *M. mycoides* genome synthesized by homologous recombination of overlapping synthetic framgments in *S. cerevisiae* (Gibson *et al.*, 2010). Their final strategy was to assemble ≈1 kb fragments to make 10 kb intermediates. The 10 kb

intermediates were isolated from *S. cerevisiae* and transformed into *E. coli*. They were then sequenced for quality control. Ten of the neighbouring 10 kb fragments were then assembled in *S. cerevisiae*, and the process was repeated. Three stages of this assembly protocol (10 kb, 100 kb and 1.09 mb) yielded a final assembled genome. There were varying purification and error checking considerations at each stage due to the increasing size of the molecules. The final synthetic genome was properly methylated with *M. capricolum* crude extract and transplanted into *M. capricolum* using the established genome transplantation method (Lartigue *et al.*, 2007). *M. mycoides* cells containing the synthetic genome were selected with tetracycline. The latest development is a combination of previous accomplishments, that allows them to claim the remarkable achievement of changing the species of a cell with a synthetic genome. Although the work of the Venter institute is a huge technical achievement, at the time of writing, there are no applications ready to take advantage of a fully synthetic genome.

1.5.4 Chromosome Engineering for Combinatorial Studies

Methods such as multiplex automated genome engineering (MAGE) (Wang *et al.*, 2009), and trackable multiplex recombineering (TRMR) (Warner *et al.*, 2010) have been developed to alter the *E. coli* genome in a massively parallel way. These methods rely on recombination of single stranded DNA (ssDNA) into the *E. coli* genome via λ Red recombination. MAGE is an automated technique that allows a large group of genes to be targeted for combinatorial mutation to generate a diverse population of cells which can be screened for a desirable phenotype. It involves growing the cells, inducing the λ *bet* gene, washing the cells, and introducing a pool of degenerate oligonucleotides into the cells via electroporation. The *bet* gene allows for the ssDNA to integrate into the chromosomal locus it is mostly homologous to. After many rounds of this, the population of cells is quite diverse. Wang *et al.* (2009) used their method to improve

lycopene production through the DXP pathway. MAGE was subsequently improved by using co-selection. The oligonuceotide pool was spiked with an oligonucleotide that would cause a selectable mutation in the target host's chromosome (prototrophy or antibiotic resistance). Therefore, only cells that had taken up nucleic acids and had expressed the *bet* gene grew into colonies. This improved the efficiency of MAGE (Wang *et al.*, 2012). TRMR is similar to MAGE but the oligos are barcoded so the recombinant genes can be identified on a microarray. The oligos are attached to a functional sequence that serves to either upregulate or down-regulate the target gene by adding a promoter or replacing a ribosome binding site with a poor ribosome binding site. All of the genes are targeted at the same time in a similar protocol as MAGE. Based on isolated phenotypic changes in certain growth conditions, inferences can be made about the up or down-regulated genes' functions, and certain genes can be further targeted (Warner *et al.*, 2010).

A notable addition to the MAGE method is CAGE: conjugative assembly genome engineering (Isaacs *et al.*, 2011). CAGE combined small MAGE edited regions, where TAG stop codons were converted to TAA, into one cell. CAGE was used to change TAG stop codons into TAA stop codons, genome wide. This was done by inserting an *oriT* selectable marker cassette into a MAGE-edited cell line and using conjugation to combine the DNA with another MAGE edited cell line. The selectable markers were chosen wisely in order to continue this process until 32 MAGE edited segments could be combined into one cell using conjugation.

These methods work in *E. coli* but it has yet to be seen how they will work in other bacteria. Because they rely on a λ Red gene, one could speculate that these methods would be applicable in other bacteria, but perhaps at reduced efficiency.

1.6 This Work

The objective of this work was to develop and validate a platform for bacterial chromosome engineering that could be implemented in a broad range of hosts, and to use it for the production of mcl-PHA in *Sinorhizobium meliloti*. This would require a means to assemble large fragments of DNA and a means to incorporate that DNA into target bacteria. The general strategy was to integrate a medium-to-large insert into the chromosome using a phage integrase system, and to subsequently assemble those large integrated inserts by a genetic cross. I hypothesized that the expression of *Pseudomonas putida phaG* and *phaC* genes would allow *S. meliloti* to produce mcl-PHA. It was also hypothesized that the mixing of *phaG* and *phaC* homologues expressed in the cells, would produce variation in the PHA material produced. The chromosome engineering system was employed to test these hypotheses.

To insert relatively large pieces of DNA into the chromosome, as discussed above, I have used the *Streptomyces* phage C31 (ΦC31) integrase system to insert DNA segments into the chromosomes of three *α-proteobacteria* (Heil *et al.*, 2012). ΦC31 integrase is a serine recombinase. It recombines an *attB* site with an *attP* site to create *attL* and *attR* sites. The recombination does not reverse if *gp3*, the gene for recombination directionality factor (RDF) (Khaleel *et al.*, 2011), is not expressed with the integrase gene (*int*). The system is set up with two *attB* sites flanking a cloning site in a donor vector, and two *attP* flanking marker genes in the chromosome (landing pad). When the donor vector is introduced into the recipient host along with a helper plasmid, cassette exchange is catalyzed. In order to use the system in a new host it is necessary to integrate the required *attP* sites and markers, i.e. the landing pad, into the chromosome through another method, thereby creating a *landing pad strain*. A simplified representation of this cassette exchange is shown in Figure 1.1.

The ΦC31 integrase is functional in a broad range of cells across kingdoms of life. It has been used in human cells (Groth *et al.*, 2000), *Drosophila melanogaster* (Bateman and Wu, 2008), and bacteria (Heil *et al.*, 2012). It stands to reason that ΦC31 will work in any sufficiently genetically developed bacterial background no matter its phylum. Once the challenge of creating a landing pad strain is met, the new host becomes compatible with the existing vectors designed to work with the ΦC31 cassette exchange system. There is no risk of multi-copy integration because the Int cannot cause excision on its own. The bacterial ΦC31 system can offer the benefits of the Tn7 system for a broader range of bacterial species.

To assemble overlapping pieces of DNA that are already integrated with IMCE (i.e. *phaC* or *phaG* homologues), a genomic conjugation strategy was attempted. This strategy is called Recombination Synthesis (RS). In the examples of RS demonstrated in this thesis, one IMCE insert contains a gentamicin resistance gene on its left flank (just inside the *att* site) and the corresponding IMCE insert (in an isogenic background) contains a hygromycin resistance gene on its right flank. The gentamicin IMCE insert also contains an R388 origin of transfer (Llosa *et al.*, 1991). R388 conjugation is used to cross the two IMCE strains: The gentamicin strain as a donor and the hygromycin strain as a recipient. Then a homologous recombination event (Figure 1.2) assembling the desired construct is the likely event leading to the dual antibiotic resistance of the transconjugants. The first application of this chromosome engineering system was to create operons to allow *S. meliloti* to make mcl-PHAs. Overlapping half operons were constructed and integrated by IMCE, and crossed using the RS system to complete the assembly.

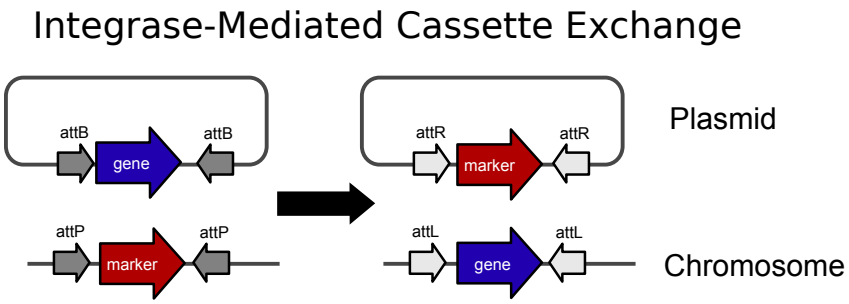

Figure 1.1: **A**: Integrase mediated cassette exchange: In the presence of integrase (not shown), recombination between *attB* and *attP* results in the exchange of marker genes on the chromosome with the "gene(s)" on the plasmid. This leaves *attL* and *attR* sites which are not substrates for recombination.

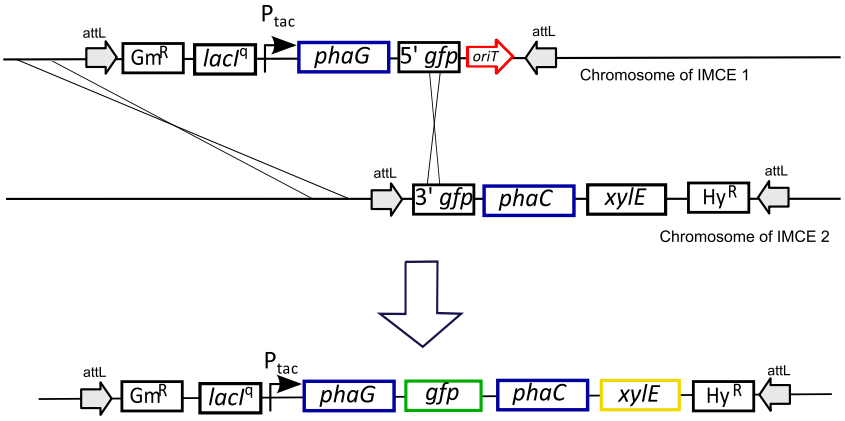

Figure 1.2: Recombination synthesis assembling two halves of an operon.

1.6.1 Claims of Contributions to Scientific Knowledge

- Developed the Integrase Mediated Cassette Exchange (IMCE) system based on the site specific recombinase from the *Streptomyces* phage ΦC31 that allows for the simplified insertion of DNA constructs into engineered host strains, and 3 corresponding host strains (landing pad strains). IMCE applies bacterial conjugation to further simplify the integration procedure.

- Constructed the pJH110 and pJH122 IMCE donor vectors carrying a red fluorescent protein gene, with gentamicin resistance and hygromycin resistance respectively.

- Constructed the pJH143 RIMCE donor vector with *sacB*, *gusA*, and spectinomycin/streptomycin resistance flanked by *attL* sites. It may be used to turn a post-IMCE clone into a SacB landing pad strain.

- Developed the ability to reverse the IMCE reaction direction and its application to change the contents of the landing pad strain to add new IMCE selection functionality.

- Designed and constructed the pJH145 IMCE donor vector fosmid with inducible high copy number, that may be used for single-copy genomic or metagenomic libraries and to increase the utility of IMCE for use in library screening.

- Demonstrated that the conjugative transfer of genomic DNA from *S. meliloti* carrying *phaG* into an isogenic *S. meliloti* recipient results in transconjugants that stably maintain the transferred genetic material.

- Demonstrated that the expression of *phaG* in *S. meliloti* causes the resultant colonies to fluoresce on media containing Nile red, suggesting the accumulation of lipids.

- Demonstrated that the addition of *phaC* from *P. putida* GPo1 and *phaG* from *P. putida* KT2440 into the *S. meliloti* chromosome allows *S. meliloti* to create medium chain length polyhydroxyalkanoates.

Chapter 2

Materials and Methods

2.1 Bacterial Strains and Plasmids

The bacterial strains used in this thesis are shown in Table 2.1. When indicated, strains are saved in the strain collection of Trevor Charles' laboratory. Strains were saved as a permanent stock by mixing overnight culture in equal parts with DMSO diluted to 14% in LB media, to give a final DMSO concentration of 7%, placing the mixture in polypropylene 2 ml vials, and storing at -75°C. For each stock, there is an entry in the Laboratory's strain collection binders. The entry goes into the binder that corresponds to the species being entered. A select group of plasmid sequences has been uploaded to AddGene. In this case a 5 digit ID number is present in the Reference column of Table 2.1. The plasmid information on AddGene may be accessed by adding the 5 digit ID number to the URL: "http://www.addgene.org". For example, the URL for pJH110 is: "http://www.addgene.org/68376/".

Table 2.1: Bacterial strains and plasmids

Strain or Plasmid	Genotype or Description	Freezer position	Reference
E. coli DB3.1	F$^-$ *gyrA462 endA1* Δ(*sr1-recA*) *mcrB mrr hsdS20* (rB- mB-) *supE44 ara-14 galK2 lacY1 proA2 rpsL20* (Sm^r) *xyl-5* Δ*leu mtl-1*	*E. coli* Box Z Vial 10	(Bernard *et al.*, 1993)
E. coli βDH10B	F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) ϕ80d*lacZ*ΔM15 Δ*lacX*74 *endA1 recA1 deoR* Δ(*ara,leu*)7697 *araD139 galU galK nupG rpsL* λ- Δ*dapA*	*E. coli* Box U Vial 1	(Rowe-Magnus, 2009)
E. coli DH5α	F$^-$ *supE44* Δ*lacU169 hsdR17 recA1 endA1 gyrA96 thi-1 relA1* (Φ80*lacZ* ΔM15)	*E. coli* Box M Vial 87	(Hanahan, 1983))
E. coli MT616	MT607 carrying mobilization plasmid pRK600 Cm^R	*E. coli* Box A Vial 25	(Finan *et al.*, 1986)
E. coli DH5α Rif	Spontaneous rifampicin resistant mutant of *E. coli* DH5α	*E. coli* Box O Vial 2	(House *et al.*, 2004)
E. coli EPI300	Epicentre copy control strain contains integrated *trfA*; F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) ϕ80d*lacZ*ΔM15 Δ*lacX74 recA1 endA1 araD*139 Δ(*ara, leu*)7697 *galU galK* λ- *rpsL* (*StrR*) *nupG trfA dhfr*.	*E. coli* Box Y Vial 3	Epicentre, Illumina Inc

Table 2.1: Bacterial strains and plasmids

Strain or Plasmid	Genotype or Description	Freezer position	Reference
S. meliloti Rm1021	SU47 *str-21*	*S. meliloti* Box 5 Vial 48	(Leigh *et al.*, 1985)
S. meliloti SmUW227	SU47 *str-21* SMc04453-SMc01185::LP	*S. meliloti* Box 13 Vial 69	This work
S. meliloti SmUW245	Rm1021 Δ *hsdR*	*S. meliloti* Box 13 Vial 87	This work
S. meliloti SmUW246	SmUW227 :: *rfp* Gm^R	*S. meliloti* Box 13 Vial 87	This work
S. meliloti SmUW235	Rm1021 Δ*phbC*	*S. meliloti* Box 13 Vial 77	Constructed by Ricardo Nordeste; this work
S. meliloti SmUW497	Rm1021 Δ(*phbC phbA phbB phaZ*)	*S. meliloti* Box 16 Vial 95	Constructed by Ricardo Nordeste; this work
S. meliloti SmUW235 ϕSmUW227	Rm1021 Δ*phbC* SMc04453-SMc01185::LP	*S. meliloti* collection	Constructed by Ricardo Nordeste; this work
S. meliloti SmUW497 ϕ SmUW227	Rm1021 Δ(*phbC phbA phbB phaZ*) landing pad strain	*S. meliloti* collection	Constructed by Ricardo Nordeste; this work

Table 2.1: Bacterial strains and plasmids

Strain or Plasmid	Genotype or Description	Freezer position	Reference
S. meliloti JRT15-JRT10 RS	RS strain KT2440 *phaG* GPo1 *phaC*	John Heil IMCE RS box	This work
S. meliloti JRT15-JRT12 RS	RS strain KT2440 *phaG* *phaC*1400	John Heil IMCE RS box	This work
S. meliloti JRT16-JRT10 RS	RS strain UW4 *phaG* GPo1 *phaC*	John Heil IMCE RS box	This work
S. meliloti JRT16-JRT12 RS	RS strain UW4 *phaG* *phaC*1400	John Heil IMCE RS box	This work
S. meliloti SmUW574	SmUW227 with LP replaced with SacB LP	*S. meliloti* Box 18 Vial 6	This work
S. meliloti SmUW579	JRT16 IMCE (SmUW584) with pTH1227	*S. meliloti* Box 18 Vial 22	This work
S. meliloti SmUW580	JRT16 IMCE (SmUW584) with pTH1227-Gpo1-*phaC*	*S. meliloti* Box 18 Vial 23	This work
S. meliloti SmUW581	JRT8 IMCE (SmUW582) with pRK7813-KT2440-*phaC*	*S. meliloti* Box 18 Vial 24	This work
S. meliloti SmUW582	JRT8 IMCE, contains KT2440 *phaG* Δ*phbC* SmUW235 landing pad background	*S. meliloti* Box 18 Vial 25	This work

Table 2.1: Bacterial strains and plasmids

Strain or Plasmid	Genotype or Description	Freezer position	Reference
S. meliloti SmUW583	JRT15 IMCE, contains KT2440 *phaG* with R388 *oriT*, Δ*phbC* SmUW235 landing pad background	*S. meliloti* Box 18 Vial 26	This work
S. meliloti SmUW584	JRT16 IMCE, contains UW4 *phaG* with R388 *oriT*, Δ*phbC* SmUW235 landing pad background	*S. meliloti* Box 18 Vial 27	This work
S. meliloti SmUW585	JRT10 IMCE, contains GPo1 *phaC*, Δ*phbC* SmUW235 landing pad background	*S. meliloti* Box 18 Vial 28	This work
S. meliloti SmUW586	JRT12 IMCE, contains *phaC*1400, Δ*phbC* SmUW235 landing pad background	*S. meliloti* Box 18 Vial 29	This work
S. meliloti SmUW587	JH151 IMCE, contains UW4 *phaG* GPo1 *phaC*, Δ*phbC* SmUW235 landing pad background	*S. meliloti* Box 18 Vial 30	This work,
S. meliloti SmUW588	JH151 IMCE, contains UW4 *phaG* GPo1 *phaC*, Δ(*phbC phbA phbB phaZ*) SmUW497 landing pad background	*S. meliloti* Box 18 Vial 30	This work
O. anthropi ATCC 49188	wild-type isolate	*O. anthropi* Box 1 Vial 3	(Holmes *et al.*, 1988)
O. anthropi Ow0012	ATCC 49188 Rif^R isolate	*O. anthropi* Box 1 Vial 19	This work

Table 2.1: Bacterial strains and plasmids

Strain or Plasmid	Genotype or Description	Freezer position	Reference
O. anthropi Ow0013	Ow0012 Δ*ampC*::LP Ap sensitive Sm^R Sp^R	*O. anthropi* Box 1 Vial 20	This work
O. anthropi Ow0014	JRT15 IMCE	*O. anthropi* Box 1 Vial 21	This work
O. anthropi Ow0015	JRT16 IMCE	*O. anthropi* Box 1 Vial 22	This work
O. anthropi Ow0016	JRT10 IMCE	*O. anthropi* Box 1 Vial 23	This work
O. anthropi Ow0017	JRT12 IMCE	*O. anthropi* Box 1 Vial 24	This work
A. tumefaciens A136	Parental strain of At11142, lacking Ti plasmid Rf^R	*A. tumefaciens* Box 1 Vial 3	(Watson *et al.*, 1975)
A. tumefaciens At11142	A136 Rf^R Δ*tetA* :: LP	*A. tumefaciens* Box 5 Vial 41	This work
pHS62	Source for ΦC31 integrase	*E. coli* Box Q Vial 91	(Thorpe and Smith, 1998)
pRK7813	Broad host range RK2 cosmid vector, with RK2 oriT, Tc resistance	*E. coli* Box A Vial 28	(Jones and Gutterson, 1987)
pJ23118 (part: J23118)	*rfp* with constitutive promoter from registry of standard biological parts	UW iGEM strain Box 1 vial 8	(Shetty *et al.*, 2008) (Campbell *et al.*, 2002)

Table 2.1: Bacterial strains and plasmids

Strain or Plasmid	Genotype or Description	Freezer position	Reference
pMB403	plasmid containing *hph* coding for hygromycin resistance	*E. coli* Box R Vial 91	(Barnett *et al.*, 2000)
pJET	Blunt end PCR cloning kit	commercial kit Thermo Fisher K1231	(Lubys *et al.*, 2007)
pJC2	pRK7813 containing ΦC31 integrase under control of lac promoter	*E. coli* Box R Vial 8	Constructed by Dr. J. Cheng 68381
pK19mobsacB	Used to create integration vectors Km^R	*E. coli* Box K Vial 84	(Schafer *et al.*, 1994)
pK19mob	Used to create integration vectors Km^R	*E. coli* Box L Vial 75	(Schafer *et al.*, 1994)
pHP45Ω	Used as source for spectinomycin resistance gene (*aadA*)	*E. coli* Box A Vial 5	(Fellay *et al.*, 1987)
pPH1JI	Source for gentamicin resistance gene	*E. coli* Box A Vial 29	(Hirsch and Beringer, 1984)
pR751	Source for Trimethoprim resistance gene	*E. coli* Box A Vial 23	(Thorsted *et al.*, 1998)
pTH1227	Used as the source for uidA ORF	*E. coli* Box L Vial 77	(Cheng *et al.*, 2007)
pTH1703	Used as the source for *gfp* reporter	*E. coli* Box M Vial 83	(Cowie *et al.*, 2006)

Table 2.1: Bacterial strains and plasmids

Strain or Plasmid	Genotype or Description	Freezer position	Reference
pTAM	Used as the source for *phaC*1400	*E. coli* collection	(Tran, 2015)
Tn5 XylE	Used as the source for *xylE* reporter	*E. coli* Box D Vial 48	(de Lorenzo *et al.*, 1990)
pMB403	Used as a source of *hph* for Hy resistance	*E. coli* Box R Vial 91	(Barnett *et al.*, 2000)
pMK2016	Source of Gateway Destination for subcloning	*E. coli* Box P Vial 63	(House *et al.*, 2004)
pKD46	λ Red helper plasmid	*E. coli* Box R Vial 27	(Datsenko and Wanner, 2000)
pJC8 lac35 clone	Test clone for Gateway recombination	*E. coli* Box S Vial 39	J Cheng, pJC8 genbank accession KC149513.1
pXINT129	*In vivo* Gateway recombination recombinase expression plasmid	personal box	(House *et al.*, 2004)
pJET-gp3	RDF clone source from Calos Lab	personal box	(Farruggio *et al.*, 2012)
pCC1fos	Fosmid vector with copy control from Epicentre kit	Pure DNA from kit	Illumina Inc.
pSU711	R388 self-mobilizing plasmid: used as mobilizer and R388 *oriT* source	*E. coli* Box T Vial 7	(Francia *et al.*, 1993)

Table 2.1: Bacterial strains and plasmids

Strain or Plasmid	Genotype or Description	Freezer position	Reference
pJH101	pK19mobSacB containing EcoRI-SmaI fragment generated by PCR of Rm1021 intergenic region	intermediate that was not saved	This work
pJH102	pK19mobSacB containing EcoRI-SmaI and SphI-HindIII fragments generated by PCR of Rm1021 intergenic region	*E. coli* Box Q Vial 94	This work
pJH103	pJH102 with cloned *uidA-attP* fragment	intermediate that was not saved	This work
pJH104	pJH103 with cloned *aadA-attP* fragment	*E. coli* Box R Vial 12	This work 68384
pJH106	pK19mobSacB containing *dhfr attB* fragment	*E. coli* Box R Vial 10	This work
pJH107	pK19mobSacB containing *aacC1 attB* fragment	*E. coli* Box R Vial 11	This work
pJH108	pK19mobSacB containing an *attB* fragment	*E. coli* Box R Vial 26	This work
pJH109	pJH106 containing *rfp-attB* fragment	*E. coli* Box R Vial 50	This work
pJH110	pJH107 containing *rfp-attB* fragment	*E. coli* Box R Vial 42	This work 68376

Table 2.1: Bacterial strains and plasmids

Strain or Plasmid	Genotype or Description	Freezer position	Reference
pJH113	pk19mobSacB containing *O. anthropi* Δ*ampC* construct (EcoRI-PstI) *Eco*RI-PstI	*E. coli* Box R Vial 43	This work
pJH115	pJH113 containing Landing Pad sequence from pJH104 (PacI)	*E. coli* Box R Vial 44	This work
pJH116	pUC57 with Gm^R FRT *attB* (side project)	*E. coli* Box R Vial 64	This Work
pJH117	pUC57 with Tp^R FRT *attB* oriT (side project)	*E. coli* Box R Vial 65	This Work
pJH118	pk19mobsacB with *A. tumefaciens* C58 Δ*tetA* (EcoRI-PstI)	*E. coli* Box R Vial 66	This Work
pJH119	pJH118 containing Landing Pad sequence from pJH104 (PacI)	*E. coli* Box R Vial 87	This Work
pJH121	Gm^R *attB* donor vector with Gateway destination	*E. coli* Box S Vial 14	This Work 68378
pJH122	Hy^R *attB* donor vector with *rfp* stuffer	*E. coli* Box S Vial 15	This Work 68377
pJH123	Hy^R *attB* donor vector with Gateway destination	*E. coli* Box S Vial 16	This Work 68379
pJH124	Hy^R *attB* donor vector with Gateway destination in opposite orientation	*E. coli* Box S Vial 17	This Work

Table 2.1: Bacterial strains and plasmids

Strain or Plasmid	Genotype or Description	Freezer position	Reference
pJH127	Nm^R *attB* in pUC57 (IMCE donor vector for RS selection with Nm)	*E. coli* Box S Vial 86	This Work
pJH141 in βDH10B	for rIMCE, integrase and RDF expression plasmid	*E. coli* Box U Vial 50	This work 68383
pJRT1	for PHA gene RS, upstream construct: P_{tac} $lacI^q$, 3′ truncated *gfp* in pk19mobsacB (Schafer *et al.*, 1994)	*E. coli* Box S Vial 42	This work
pJRT2	for PHA gene RS, downstream construct: 5′ truncated *gfp*, *xylE* in pk19mobsacB (Schafer *et al.*, 1994)	*E. coli* Box S Vial 43	This work
pJRT8	JRT1 construct in pJH110 (SphI) with *P. putida* kt2440 *phaG* (PacI–PmeI)	*E. coli* Box S Vial 44	This work 68385
pJRT9	JRT1 construct in pJH110 (SphI) with *P. putida* UW4 *phaG* (PacI–PmeI)	*E. coli* Box S Vial 45	This work 68386
pJRT13	JRT1 markers in pJH110 (SphI)	*E. coli* Box S Vial 46	This work
pJRT14	JRT2 markers in pJH1122 (SphI)	*E. coli* Box S Vial 47	This work

Table 2.1: Bacterial strains and plasmids

Strain or Plasmid	Genotype or Description	Freezer position	Reference
pJRT10	*P. putida* gpo1 *phaC* in pJRT14 (PacI–PmeI)	*E. coli* Box Z Vial 5	This work 68387
pJRT11	*S. meliloti* Rm1021 *phaC* in pJRT14 (PacI–PmeI) note: that this was derived from a synthetic clone, that turned out to be truncated, and therefore was not used	*E. coli* Box Z Vial 6	This work
pJRT12	*phaC* 1400, engineered *phaC* from Tam Tran in pJRT14 (PacI–PmeI)	*E. coli* Box Z Vial 7	This work 68388
pJRT15	R388 *oriT* added to pJRT8 (JRT1 construct in pJH110 (SphI) with *P. putida* kt2440 *phaG* (PacI–PmeI)	*E. coli E. coli* Box Z Vial 8	This work 68389
pJRT16	R388 *oriT* added to pJRT9: JRT1 construct in pJH110 (SphI) with *P. putida* UW4 *phaG* (PacI–PmeI)	*E. coli E. coli* Box Z Vial 9	This work 68390
pk19attlsac	attlsac fragment cloned into pk19mob as a HindIII EcoRI fragment	personal box	This work
pJH145	Mobilizable IMCE fosmid with copy control in EPI300	*E. coli* Box W Vial 75	This work 68380

Table 2.1: Bacterial strains and plasmids

Strain or Plasmid	Genotype or Description	Freezer position	Reference
pJH143	RIMCE donor vector containing containing SacB landing pad sequence	*E. coli* Box Y Vial 98	This work 68382
pJH146	pJET PHA operon clone KT2440 *phaG* GPo1 *phaC*	*E. coli* Box y Vial 99	This work
pJH147	pJET PHA operon clone KT2440 *phaG phaC*1400	*E. coli* Box Y Vial 100	This work
pJH148	pJET PHA operon clone UW4 *phaG* GPo1 *phaC*, clone 1	*E. coli* Box Z Vial 01	This work
pJH149	pJET PHA operon clone UW4 *phaG* GPo1 *phaC*, clone 3	*E. coli* Box Z Vial 02	This work
pJH150	pJET PHA operon clone UW4 *phaG* GPo1 *phaC*, clone displaying GFP	*E. coli* Box Z Vial 03	This work
pJH151	pJH110 PHA operon sub-clone from pJH148	*E. coli* Box Z Vial 4	This work

2.2 Oligonucleotides

The oligonucleotides used in this thesis are shown in Table 2.2. Oligonucleotides were synthesized by various third party suppliers. Upon arrival they were resuspended in 2 mM Tris-HCl pH 8.0 to a concentration of 100 μM. For use they were further diluted to 10 μM in the same solvent. All oligonucleotides were stored at $-20°$C.

Table 2.2: Oligonucleotides

Name	Sequence	Function
L fwd EcoRI	ATAGAATTCCTTTCTCAAGAAACGCATCG	clone L homo for pK19
L-rev 2 PacI SmaI	TCACCCGGGTTAATTAAGTAGGAGCGGC-GATGCTG	clone shortened L-homo
R fwd PacI SphI	AGCGCATGCTTAATTAAGCCAAGCGCAG-GGATGACGATA	clone R homo add *PacI* site
R rev HindIII	TAAAAGCTTCGCAGGTGAGATGGGCATGT	clone R-homo for pK19
GmR fwd	ACTGCATGCCGCACACCGTGGAAACGGAT	for creation of pJH110
GmR rev	AAGAAGCTTGTGCCAGGGCGTGCCCTTGG-GCTCCCCGGGCGCGTGAAATAAGTGGCT-GCGCGC	for creation of pJH110
TpR fwd	ACTGCATGCGAACCCAGTTGACATAAGC-CTGTT	for creation of pJH109
TpR rev	AAGAAGCTTGTGCCAGGGCGTGCCCTTGG-GCTCCCCGGGCGCGGGCTTGAGCGCTGA-GCGGTG	for creation of pJH109

Table 2.2: Oligonucleotides

Name	Sequence	Function
GW DS fwd	AATGCATGCGATTCCCTTTGTCAACAGCAATG	For creation of Gateway compatible ABCDE donor vectors, not used
GW DS rev	ACTGCATGCGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGCAGCAATGGATATCAACCACTTTG	For creation of Gateway compatible ABCDE donor vectors, not used
GW US fwd	AATGCATGCGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGGATTCCCTTTGTCAACAGCAATG	For creation of Gateway compatible ABCDE donor vectors, not used
GW US rev	ACTGCATGCCAGCAATGGATATCAACCACTTTG	For creation of Gateway compatible ABCDE donor vectors, not used
attP-aadA-F SmaI	ATACCCGGGCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCATGCCCGTTCCATACAGAAG	for creation of pJH104
aadAR XbaI	GGGTCTAGATTATTTGCCGACTACCTTGGT	for creation of pJH104
gusA-F SalI	GGGGTCGACGAGGAGTCCCTTATGTTACGTC	for creation of pJH104
gusA-attP-R SphI	ATAGCATGCCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGTCATTGTTTGCCTCCCTGCT	for creation of pJH104

Table 2.2: Oligonucleotides

Name	Sequence	Function
INT-F	ATACTGCAGTGATTGATTGAGGAGAACA-GCTATGGACACGTACGCGGGTGCT	For creation of pJC2 IMCE integrase plasmid
INT-R	GGGGGATCCTTAGGCTAACTAACTAAACC-G	For creation of pJC2 IMCE integrase plasmid
US XbaI RFP attB F	ACTTCTAGAGGTGCCAGGGCGTGCCCTTG-GGCTCCCCGGGCGCGGCATGCTTGACGGC-TAGCTCAGTCCTAG	for construction of pJH110
US PstI RFP R	CATCTGCAGCGATCTACACTAGCACTATC-AGC	for construction of pJH110
DS RFP F	TTGACGGCTAGCTCAGTCCTAG	changed plans, not used
DS RFP attB R	GGTGCCAGGGCGTGCCCTTGGGCTCCCCG-GGCGCGGCATGCCGATCTACACTAGCACT-ATCAGC	changed plans not used
ampC US F EcoRI	ATTGAATTCCCCACTATGCTGGCATTCCG	For creation of pJH115
ampC US R XO	AAAAAAGAACCCCGGAAAATCC	For creation of pJH115
ampC DS F XO	CGGATTTTCCGGGGTTCTTTTTTAATTAAA-GATAAATAAACCTCATGATTATG	For creation of pJH115
ampC DS R PstI	ATTCTGCAGCAGGTCGATGAACGGGTAG	For creation of pJH115
US tetA EcoRI	CATGAATTCTTTCAAGAGCTTTTGCAAGC-AG	For creation of pJH118

Table 2.2: Oligonucleotides

Name	Sequence	Function
US tetA R XO	CGTCCGGAGTCTTAATTAAGGTATGGCATCCCTATAAAGC	For creation of pJH118
DS tetA F XO	GATGCCATACTTAATTAAGACTCCGGACGTCATGCGCC	For creation of pJH118
DS tetA R PstI	CATCTGCAGCGGTATTATCATTGGGAAAC	For creation of pJH118
hph red 1a F	TAACGCTGATAGTGCTAGTGTAGATCGCTGCAGGCATGCGAATTACCACAGCAATGACAAAG	for Λ red recombination of hph into pJH110 to create pJH122
hph red 2a R	GGCTGCGCGCATAAAGCGGGTGATGGGACTAACGCCTGACTAGTAACATAGATGACACC	for Λ red recombination of hph into pJH110 to creat pJH122
hsdR US F EcoRI	ATTGAATTCATGCCCAACCTAAACACAGGTATC	for assembly PCR and cloning homology for *Sinorhizobium meliloti hsdR* deletion
hsdR US R XO	TGGAAGCCAAAGTCTATTCCGCAGCCTCGGCAAAG	for assembly PCR and cloning homology for *S. meliloti hsdR* deletion

Table 2.2: Oligonucleotides

Name	Sequence	Function
hsdR DS F XO	TGCGGAATAGACTTTGGCTTCCAGTCTCCCCG	for assembly PCR and cloning homology for *S. meliloti hsdR* deletion
hsdR DS R PstI	ATTCTGCAGGCAAAGCAGCCAAGCCCGGT	for assembly PCR and cloning homology for *S. meliloti hsdR* deletion
GFPstart-F PmeI XbaI	CATTCTAGAGTTTAAACGATCTGATTGATTGAGGAGATTAATAATG	for PHA half operon cloning
GFPstart-R SphI EcoRI	CATGAATTCGCATGCCTTTCGAAAGGGCAGATTG	for PHA half operon cloning
GFPend-F PstI SphI	CATCTGCAGGCATGCGCACAAATTTTCTGTCAGTGGAG	for PHA half operon cloning
GFPend-R PacI XbaI	CATTCTAGATTAATTAACATCGGATCTCATTATTTGTAGAGC	for PHA half operon cloning
GPO1-F PacI	CATTTAATTAAGCTAGGAGTCCCTTATGTCC	for PHA half operon cloning
GPO1-R PmeI	CATGTTTAAACTCATCAGTGGTGGTGGTGGTG	for PHA half operon cloning
XylE/T-F PmeI XbaI	CATTCTAGAGTTTAAACCATCACCCAGAGCTGTTGGGGGA	for PHA half operon cloning

Table 2.2: Oligonucleotides

Name	Sequence	Function
XylE/T -R SphI EcoRI	CATGAATTCGCATGCCTCTGCAATAAGTCGTACCGGAC	for PHA half operon cloning
lactac-F PstI SphI	CATCTGCAGGCATGCGAATGGTGCAAAACCTTTCGC	for PHA half operon cloning
lactac-R PacI XbaI	CATTCTAGATTAATTAAGCTCGTATAATGTGTGGAATTG	for PHA half operon cloning
phaG-F PacI	CATTTAATTAACTAGGAGTCCCTTATGAGGCCAGAAATCGCTGTAC	for PHA half operon cloning, kt2440 *phaG*
phaG-R PmeI HIS	CATGTTTAAACTCATCAGTGGTGGTGGTGGTGGTGGATGGCAAATGCGTGCTGCCCCTGCTG	for PHA half operon cloning kt2440 *phaG*
phbC-F PacI	CATTTAATTAACAGAGCTAGGAGTCCCTTATG	for PHA half operon cloning
phbC -R PmeI	CATGTTTAAACTCATCAGTGGTGGTGGTGGTG	for PHA half operon cloning
phaGUW4-F PacI	CATTTAATTAACTAGGAGTCCCTTATGAGGCCAGAAATCGCTGTG	for PHA half operon cloning
phaGUW4-R PmeI HIS	CATGTTTAAACTCATCAGTGGTGGTGGTGGTGGTGAAGTGCCAATGCATGGTGGTCC	for PHA half operon cloning
phaC1400-F PacI(2)	CATTTAATTAAAGGAGTCCCTTATGTCGAAC	for PHA half operon cloning
phac1400-R PmeI(2)	CATGTTTAAACCAGTCGTTGCGCTGCAG	for PHA half operon cloning

Table 2.2: Oligonucleotides

Name	Sequence	Function
attLsac F	CATAAGCTTGTGCCAGGGCGTGCCCTTGAGTTCTCTCAGTTGGGGGGGATCGATCCTTTTTAACCCATCAC	For construction of *sacB* landing pad
R388 F attB XbaI	TCTAGAGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGCCGCCTCGTCCTCCAAAAGTG	For cloning the R388 *oriT*
R388 SalI R	ACATGTCGACCTCATTTTCTGCATCATTGTAGCAC	For cloning the R388 *oriT*
attLsac R	CATGAATTCGTGCCAGGGCGTGCCCTTGAGTTCTCTCAGTTGGGGGGTTTAAACGTTGGGCGTCGCTTGGTCGGTC	For construction of *sacB* lanidng pad
gusspec F	CATGTTTAAACCATGCCCGTTCCATACAGAAG	For construction of *sacB* landing pad
gusspec R	CATGTTAAACTCATTGTTTGCCTCCCTGCTG	For construction of *sacB* landing pad
RDF fix F	CATGGATCCAGGAGAACAGCTATGGCGAAGCGTTCGATC	For construction of rIMCE recombination plasmid
RDF fix R	CATGGATCCCTAGTCGGCAATCGCGTCGTTG	For construction of rIMCE recombination plasmid
Gm-rev	CGACCCAAGTACCGCCACC	For colony PCR to verify IMCE and screen for orientation

Table 2.2: Oligonucleotides

Name	Sequence	Function
Hy-fwd	GCGCGCGGTGTCATCTATGTTAC	For colony PCR to verify IMCE and screen for orientation
RDF-fix F	CATGGATCCAGGAGAACAGCTATGGCGAAGCGTTCGATC	For cloning ΦC31 RDF to create pJH141
RDF-fix R	CATGGATCCCTAGTCGGCAATCGCGTCGTTG	For cloning ΦC31 RDF to create pJH141
attlsac F	CATAAGCTTGTGCCAGGGCGTGCCCTTGAGTTCTCTCAGTTGGGGGGGATCGATCCTTTTTAACCCATCAC	For SacB landing pad construction
attLsac R	CATGAATTCGTGCCAGGGCGTGCCCTTGAGTTCTCTCAGTTGGGGGGTTTAAACGTTGGGCGTCGCTTGGTCGGTC	For SacB landing pad construction
gusspec F	CATGTTTAAACCATGCCCGTTCCATACAGAAG	For SacB landing pad construction
gusspec R	CATGTTAAACTCATTGTTTGCCTCCCTGCTG	For SacB landing pad construction
KL12-F	CCTAAGCTTTCGGTCTTGCCTTGCTCGTCGG	For cloning RK2 oriT to construct pJH145
KL13-R	CCTAAGCTTGCGCTTTTCCGCTGCATAACCC	For cloning RK2 oriT to construct pJH145

Table 2.2: Oligonucleotides

Name	Sequence	Function
attBfos-F	GTGACGCGCCCGGGGAGCCCAAGGGCACGCCCTGGCACGGATCCTCTAGAGTCGACCTGCAG	For creating pJH145
attbfos-R	GTGCGCGCCCGGGGAGCCCAAGGGCACGCCCTGGCACGGATCCCCGGGTACCGAGCTCG	For creating pJH145
gfpXO-41-F	GCCCGAAGGTTATGTACAGG	For direct genomic sanger sequencing; designed with primer3 (Untergasser *et al.*, 2012)
gfpXO-313-F	TAGCATCACCTTCACCCTCTC	For direct genomic sanger sequencing; designed with primer3 (Untergasser *et al.*, 2012)
gfpXO-691-R	CTTTTCGTTGGGATCTTTCG	For direct genomic sanger sequencing; designed with primer3 (Untergasser *et al.*, 2012)

Table 2.2: Oligonucleotides

Name	Sequence	Function
gfpXO-755-R	GCTCATCCATGCCATGTGTA	For direct genomic sanger sequencing; designed with primer3 (Untergasser *et al.*, 2012)
gfp-junction-R	GTAAGCTTTCCGTATGTAGCATCA	For direct genomic sanger sequencing; designed with primer3 (Untergasser *et al.*, 2012)
phaG gfp F SphI	GCATGCGAATGGTGCAAAACCTTTCGCGG	For *in vitro* assembly and cloning of PHA half operons
phaG gfpR XO	CTTTCGAAAGGGCAGATTGTGTC	For *in vitro* assembly and cloning of PHA half operons
phaC gfpF XO	GCACAAATTTTCTGTCAGTGGAG	For *in vitro* assembly and cloning of PHA half operons
phaC gfpR SphI	GCATGCCTCTGCAATAAGTCGTACCGGAC	For *in vitro* assembly and cloning of PHA half operons

2.3 Bacterial Culture and Microbiological Techniques

Bacterial strains were cultured in liquid media that was aerated on a shaker at 150 rpm at the desired growth temperature: 37°C for *Escherichia coli* strains and 30°C for *Sinorhizobium meliloti*, *Ochrobactrum anthropi*, and *Agrobacterium tumefaciens*. Bacterial strains were also cultured on plates solidified with 1.5% agar (Bio Basic bacteriological grade). Plates were incubated at the desired temperature: 37°C for *E. coli* strains and 30°C for *S. meliloti*, *O. anthropi*, and *A. tumefaciens*.

Strains containing plasmids were cultured with the appropriate antibiotic(s) in the media to provide selective pressure for plasmid maintenance.

2.3.1 Bacterial Growth Media

Growth media was sterilized by autoclaving at 121°C (103 kPa in addition to atmospheric pressure) for 20 minutes. Heat-labile media components, such as antibiotics were filter sterilized (0.22 μm filter), or sterilized by dissolution in an organic solvent. Heat labile substances were added when media was cool (liquid media), or just before plates were poured (agar media). Agar media was solidified with 1.5% bacteriological grade agar (Bio Basic, Markham ON)

The complex media used for *E. coli*, *A. tumefaciens*, *O. anthropi*, and sometimes *S. meliloti* was lysogeny broth (LB) (Lennox, 1955). *S. meliloti* was often grown on tryptone yeast (TY) complex media (Beringer, 1974). *S. meliloti* was also grown on LB supplemented with 2.5 mM $MgSO_4$ and 2.5 mM $CaCl_2$; this media is referred to as LBmc (Finan *et al.*, 1984). To promote the accumulation of polyhydroxyalkanoates and a mucoid colony appearance, *S. meliloti* was grown on yeast extract mannitol media (Vincent, 1970). For electroporation, and sometimes for $CaCl_2$ transformation, super optimal broth with catabolite repression (SOC) (Hanahan, 1983) was used to revive cells before plating.

Antibiotics were obtained from Bio Basic (Markham ON) as a pure powder. The powder was dissolved in a suitable solvent to make a 1000 times concentrate stock solution.

For *E. coli*, antibiotics were used at the following concentrations in μg/ml in agar solidified media, or half the stated concentration in liquid media: kanamycin sulfate, 25; gentamicin sulfate, 25: hygromycin B, 50; tetracycline hydrochloride, 10; ampicillin sodium salt, 100; rifampicin, 50; spectinomycin dihydrochloride, 50; streptomycin sulfate, 100; chloramphenicol, 12.5 for pCC1fos selection, or 25 otherwise.

For *S. meliloti*, antibiotics were used at the following concentrations in μg/ml in agar solidified media, or half the stated concentration in liquid media: neomycin sulfate, 200; gentamicin sulfate, 25 to 50; Hygromycin B, 50; tetracycline hydrochloride, 10; rifampicin, 50; spectinomycin dihydrochloride, 100; streptomycin sulfate, 200.

For *A. tumefaciens*, antibiotics were used at the following concentrations in μg/ml in agar solidified media, or half the stated concentration in liquid media: gentamicin sulfate, 25, hygromycin B, 200; tetracycline hydrochloride, 10; rifampicin, 50; spectinomycin dihydrochloride, 100; streptomycin sulfate, 200.

For *O. anthropi*, antibiotics were used at the following concentrations in μg/ml in agar solidified media, or half the stated concentration in liquid media: gentamicin sulfate, 25; hygromycin B, 200; rifampicin, 50; spectinomycin dihydrochloride, 100; streptomycin sulfate, 200.

5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) was dissolved at 50 mg/ml in dimethylformamide (DMF) and added to media at one μl per ml to give a final concentration of 50 μg/ml. 5-Bromo-4-chloro-3-indolyl-β-D-glucuronide (X-gluc) was weighed out for the desired amount of media, to give a final concentration of 100-200 μg/ml. The weighed out x-gluc was dissolved in an 0.5-1 ml of dimethyl sulfoxide (DMSO) and all of the solution was added to the agar media. X-gal and X-gluc were

purchased from Gold Biotechnology USA.

2.3.2 Preparation of Bacterial Frozen Stocks

Cultures were grown in the appropriate growth media. The culture was mixed in a 1:1 ratio in a cryogenic storage vial with an LB media solution containing 14% DMSO. The vial was stored at -75°C.

2.4 ΦM12 Transductions

ΦM12 bacteriophage was used to transduce mutations linked to a selectable marker between *Sinorhizobium meliloti* backgrounds (Finan *et al.*, 1984).

2.4.1 Preparation of ΦM12 Phage Lysates

Mutants to be transduced were subcultured in LBmc and incubated at 30°C with shaking to produce actively growing cultures. When the cultures reached early log phase (OD_{600} of 0.4), 50 μl of an Rm1021 lysate was added. The infected cultures were left to incubate under the same conditions for a period of 6 hours to overnight. After infection, the cultures appeared clear indicating cell lysis had occurred. 100 μl of chloroform was added to each lysate to prevent regrowth of surviving bacterial cells. To pellet any cell debris and chloroform, the lysates were centrifuged at 5000 x g in a benchtop centrifuge in 15 ml conical centrifuge tubes. The lysates were either used immediately or stored at 4°C

2.4.2 Determining Titer of ΦM12 Phage Lysates

Ten-fold serial dilutions of the phage lysates were made by diluting 10 μl of phage in 90 μl of TY media. To ensure proper mixing, the liquid was pipetted up and down 20

times following each new dilution. The vortexer was not used, because vortexing can cause destruction of the phage.

Dilutions from 10^{-6} to 10^{-12} were added to a cool but still liquid 0.7% LBmc agar overlay along with 50 μl of saturated Rm1021 culture. The overlay was poured onto LBmc agar plates. The plates were incubated for 1-2 days until plaques were visible.

The titer was determined by multiplying the number of plaques on a plate by the dilution factor of the counted plate. The number of colonies on a 10^{-10} plate would be multiplied by 10^{10}.

2.4.3 Transduction

Transduction was carried out by adding 200 μl of mid-log phase recipient culture to 200 μl of phage lysate diluted in LBmc media. The amount of lysate used is based on the titer of the lysate, and the titer of the cells. The dilution is chosen to give a multiplicity of infection (MOI) of 0.5. The recipient culture and the phage are incubated together for 30 minutes so that the phage particles could adsorb to the cells. The cells were then washed 3 times in 1 ml 0.85% NaCl solution to dilute cations and wash away unbound phage. After washing, the cells were plated on the appropriate selective media, which was LB (due to its lack of calcium and magnesium) with the appropriate antibiotic for selection of the marked mutation being transferred.

2.5 Triparental Mating

Preparation of Culture

Cultures of the donor strain, recipient strain, and mobilizer strain were prepared by inoculating a single colony in 3-5 ml of LB media. Antibiotics were added to the media for plasmid maintenance (Table 2.1. Concentrations are found in section 2.3.1. Growth

conditions are found in Section 2.3.

Mating

Once cultures were grown, 1.5 ml of culture was removed and washed 2 times in sterile 0.8% NaCl. The pellet was then resuspended in 100 μl of sterile 0.85% NaCl. To 10 μl of resuspended culture was then spotted on a TY agar plate. These spots served as negative controls. The triparental mating was prepared by adding 40 μl of all three resuspended cultures into a sterile 1.5 ml tube and mixing. Then the mixture was spotted on a TY agar plate. The spots were dried on the plate(s) by leaving the plate open in a laminar flow hood. The plate(s) were incubated at 30°C for 16-20 hours.

Selection

After the incubation, the control spots were streaked on selective media containing antibiotics that were counterselective against donor and recipient, therefore selective for transconjugants. The triparental mating spot was also streaked on the same media. One quarter to one half of the triparental mating spot was resuspended in 0.85% NaCl and a 10 fold dilution series to 10^{-7}. Typically this was done by diluting 10 μl of culture into 90 μl of 0.85% NaCl. 40 μl of each dilution, 10^{-2}-10^{-6}, was spread plated on media selective for transconjugants; 40 μl of each dilution 10^{-4}-10^{-7} on media selective for recipients and transconjugates. Plates were then incubated at the appropriate temperature for the appropriate time (section 2.3).

2.6 Plasmid Construction with λ Red Recombineering

The plasmids pKD46 (Datsenko and Wanner, 2000) (to express λ *exo, bet, gam* genes required for recombination) and the target plasmid were introduced into *Escherichia coli* DH5α by transformation into calcium chloride competent cells. A segment of DNA

for insertion into the target plasmid was amplified by PCR. This insert DNA had 40 base pairs of homology for the target plasmid added to each end. The PCR product was purified by absorption and elution using a silica gel spin column.

The strain containing the target plasmid and pKD46 was grown in 5 ml of LB media containing the relevant antibiotics. The culture was prepared for electroporation by 3 times washing in ice cold sterile 10% glycerol. The cell pellet was then resuspended in 200 μl of 10% glycerol. The purified insert DNA was electroporated into the prepared cells containing the target plasmid and pKD46. Electroporation was done in a 0.1 cm gap electroporation cuvette (Bio-Rad Laboratories Inc.). A voltage of 1 kV was applied using a Bio-Rad micropulser electroporation machine (Bio-Rad Laboratories Inc.). SOC media containing 10 mM L-arabinose was used to rescue cells. Arabinose was used to induce the genes on pKD46. The rescued cells (in SOC 10% L-arabinose) were incubated at 30°C for 2 hours. pKD46 replication is temperature sensitive. The electroporated cultures were then plated on media containing appropriate antibiotics for selection. In the case that the insert was an antibiotic resistance gene: the antibiotics for the backbone resistance and the insert resistance. The plates were incubated for 16-20 hours at 30°C. Colonies were evaluated for the expected phenotype.

2.7 Molecular Cloning Techniques

2.7.1 DNA Ligation

The cut DNA fragments were purified. Usually a gel extracted insert was used with a gel extracted vector, especially for double digest cloning steps. If the concentration of the DNA fragment solutions was below 10 ng/μl, then they were concentrated in a SpeedVac concentrator (Savant, Thermo Fisher Scientific Inc), such that the ligation reaction would fit in a small volume (10-20 μl). 50-100 ng of vector DNA was added to

the tube, a corresponding amount of insert DNA was added. The insert DNA amount was calculated to give the desired molar ratio of insert to vector (commonly 1:1, 3:1, 5:1). Next 1/10 volume of the final volume of 10x ligase buffer was added (Thermo Fisher Scientific Inc). Water was added to bring the volume to 0.5 μl below the final volume. Finally, 0.5 μl of T4 DNA ligase at 5 Weiss units/μl (Thermo Fisher Scientific Inc) was added. The tube was flicked to mix, and quickly centrifuged to collect the liquid at the bottom of the tube. Then the tube was incubated at room temperature for 30 minutes to overnight.

2.7.2 Agarose Gel Electrophoresis

Agarose gel electrophoresis was used to analyze the size and condition of DNA fragments from a variety of processes. Routine agarose gel electrophorsis was done using a 0.8-1% agarose gel, that was made in 1 times Tris-Acetate-EDTA (TAE) buffer. The running buffer was also 1 times TAE. The protocol was adapted from Sambrook and Russell (2001).

The powdered biotechnology grade agarose (Bioshop Canada) was weighed out, and added to the TAE buffer in an erlenmeyer flask. The flask was microwaved at max power until the liquid boiled (usually about 40 seconds). The flask was swirled to distribute the agarose particles and microwaved again for a shorter time. The flask was swirled until there were no more visible granules of undissolved agarose. The agarose was then allowed to cool to 50°C and 1 μl of 1/5 concentrate Gel Red (Biotium Inc.) was added per 10 ml of agararose gel volume. The agarose was then cast in an Owl gel electrophoresis system of the appropriate size. After the gel had solidified the casting gaskets were removed and the running buffer was added. Then the comb was removed.

The samples to be electrophoresed were prepared by adding 1/6 volumes of 6

times concentrated loading dye to the sample.

The electrophoresis system was connected to the power supply with the black lead connected to the terminal behind the wells, and the red lead connected to the terminal in the direction the DNA was to run. The power supply was set at a constant voltage such that the electric field was 7 V/cm e.g. a system that was 10 cm long from electrode to electrode would be run at 70 V. The time the electrophoresis system was run was commonly in the range of 40-50 minutes.

2.7.3 Gel Extraction

This protocol was adapted from Sambrook and Russell (2001).

Gel extraction was used to purify a piece of DNA of a particular size. An agarose gel was prepared as stated above, but the TAE buffer also contained 1 mM guanosine to protect the DNA from UV exposure (Gründemann and Schömig, 1996).

The desired DNA band was excised from the gel using a straight razor blade. The mass of the excised gel was determined by weighing a 1.5 ml centrifuge tube, weighing the tube containing the excised band, and subtracting the weight of the tube. The excised gel was assumed to have the same density of water (1 g/ml) and the volume of the excised gel was calculated.

To dissolve the agarose 3 volumes of binding buffer were added. The composition of the binding buffer was: 140 mM 2-[N-Morpholino] ethanesulfonic acid (MES) pH adjusted to 7.0 with sodium hydroxide; 20 mM ethylenediaminotetraacetic acid (EDTA), and 5.5 M guanidine isothiocyanate (Gu-ITC) (Kim and Morrison, 2009). Then the tubes were incubated at 60°C for 5 minutes.

The dissolved agarose solution was added to a silica gel spin column (Bio Basic Markham ON), and the solution was forced through the column by centrifugation at max speed (typicaly 15 000 x g) for 30 seconds. The flow through was discarded. The

column was washed with 500 μl of wash buffer (80% ethanol, 10 mM Tris-HCl pH 7.5) by forcing the wash buffer through the column by centrifugation at 15 000 x g for 1 minute. Again the flow through was discarded. The wash step was repeated, and to dry the column the centrifugation was repeated again once more without the addition of wash buffer.

To elute the DNA from the column 35 μl of 2 mM tris-HCl pH 8.0 was added to the column and allowed to incubate for 5 minutes. The column was placed in a new collection tube and spun at 8000 x g for 1 minute. The column was discarded and the flow through was the final purified DNA fragment.

2.7.4 Preparation of $CaCl_2$ Competent Cells

This method was adapted from protocol manual by Sambrook and Russell (2001).

A single colony of *Escherichia coli* DH5α was inoculated into 3 ml of LB media in a 16 mm test tube. The tube was incubated at 37°C with constant shaking for 16 hours. 2-9 250 ml baffled Erlenmeyer flasks each containing 50 ml of LB media were inoculated with 200 μl of saturated culture from the previous step. The shaker that was available made using these types and numbers of flasks convenient. The flasks were shaken at 37°C until they reached an OD_{600} of 0.3-0.4, this took approximately 3.5 hours. The culture was decanted into 250 ml polypropylene GSA centrifuge bottle(s) (Nalgene, Thermo Fisher Scientific Inc.). The centrifuge was balanced. The cells were pelleted by centrifugation in a Sorval GSA rotor or equivalent (Sorval, Thermo Fisher Scientific Inc.) at 4000 x g for 10 minutes at 4°C. The supernatant was decanted and the pellet was resuspended in 0.1 M $MgCl_2$ totaling 1/3 the volume of the culture. The cells were centrifuged again with the same settings. Note once divalent cations were added the pellet formed mostly on the side of the bottle and not on the bottom, so care was taken not to lose this more diffuse pellet. The pellet was resuspended in

0.1 M $CaCl_2$ totaling 1/5 the volume of the culture. The bottles were placed on ice and placed in the 4°C fridge for an overnight incubation. The cell suspension was then moved into 40 ml oakridge tubes (Nalgene, Thermo Fisher Scientific Inc.) and centrifuged in an SS34 rotor or equivalent (Sorval, Thermo Fisher Scientific Inc.) at 4000 x g for 10 minutes at 4°C. The supernatant was decanted and the pellet was resuspended in 0.1 M $CaCl_2$ in 15% glycerol totaling 1/50 the volume of the culture. 200 μl aliquots were made by pipetting 200 μl of suspension into pre-chilled ice cold 1.5 ml microcentrifuge tubes. The aliquots were stored on ice, then flash frozen in liquid nitrogen, and then stored at -75°C.

2.7.5 $CaCl_2$ Transformation

1.5 ml microcentrifuge tubes were labeled and chilled on ice. The required number of frozen competent cell aliquots were removed from the freezer and thawed on ice. The DNA in a 10 μl volume or less (20-100 ng for a ligation, 0.01-50 ng for supercoiled plasmid) was added to the chilled tubes, and 100 μl of thawed ice-cold competent cells were added and mixed with the DNA solution. The DNA cell mixture was allowed to incubate on ice for 30 minutes. After incubation the tubes were placed into a 42°C water bath for 60 seconds. After 60 seconds the tubes were promptly removed and placed back on ice for 5 minutes. After which, the cells were pelleted by spinning at max speed in a microcentrifuge for 30 seconds. The supernatant was poured off, leaving approximately 50-100 μl in the tube. The cells were resuspended in the reduced volume and the cells were spread plated on media containing the selective antibiotic appropriate for the plasmid. The plates were incubated at 37°C for 16 hours.

2.7.6 Electroporation

For the electroporation of plasmid or linear DNA, minimization of salts in both the cell suspension and DNA solution was important. To make electrocompetent cells, log phase cells were washed three times in sterile ice-cold 10% glycerol diluted in deionized water. The cells were then resuspended in a volume 1/40 of the original culture in the same ice-cold 10% glycerol. These cells could be used right away or frozen for future use.

DNA (plasmid, linear, ligation mixture) was either ethanol precipitated or purified using a silica gel spin column. 1-50 ng of DNA was mixed with 50 μl of electrocompetent cells and added to pre-chilled 0.1 cm gap electroporation cuvettes (Bio-Rad Laboratories Inc.). Cuvettes were placed one at a time into a Bio-Rad micropulser electroporation unit and a voltage of 1 kV was applied in a short pulse (instrument default for ec1 setting); immediately after the voltage pulse the 1 ml of SOC media was added slowly to the cuvette. The media was gently aspirated and dispensed once more. The media cell suspension was allowed to sit in the cuvette at room temperature until the other samples had been processed and then the cell suspension was transferred to a sterile 1.5 ml centrifuge tube. The tubes were then incubated at 37°C (for *Escherichia coli*) for 1 hour before plating the cells on selective media. If a high efficiency was expected, a portion of the cell suspension could be plated (50 μl of the ml), or all of the cells could be plated by pelleting the cells and resuspending in 50 μl of media prior to spread plating. The plates were incubated for 16-20 hours at 37°C.

2.7.7 Restriction Enzyme Digest of DNA

DNA was restricted using fast digest restriction enzymes (Thermo Fisher Scientific Inc.). For 1 μg of DNA, 1 μl of fast digest restriction enzyme was used. Fast digest restriction enzymes were purchased in a concentration of 1 fast digest unit (FDU) per

μl. An FDU was defined as the enzyme cutting 1 μg of DNA in 5 minutes. To adequately dilute the glycerol contained in the enzyme storage buffer, the final volume of the digest reaction mixture was set at greater than 10 times the volume of restriction enzyme used. Nuclease free water was used to dilute the reaction to the final volume. Fast digest reaction buffer was added to one tenth of the final volume.

The reactions were incubated at 37°C for 20 minutes if large amounts of DNA (>1μg) were being digested. For small restrictions, 5 minutes was sufficient.

2.7.8 Polymerase Chain Reaction (PCR)

High Fidelity PCR using KOD Xtreme Polymerase

The PCR amplification was assembled by adding the components in the following table (Table 2.3) together in a thin-walled 0.2 ml tube (EMD Millipore, 2011b). The tube was placed in a thermocycler and cycled under appropriate conditions. The extension time for KOD Xtreme (EMD Millipore Billerica, MA, USA).

The thermal cycling protocol used was a touchdown protocol where for the first 20 cycles the annealing temperature would decrease 0.5°C from 65°C to 55°C. After that, there were 30 cycles with an annealing temperature of 50°C. Before the first cycle there was a 2 minute heating step at 94°C to start the hot start polymerase. After that a cycle was as follows: 94°C for 30 seconds for denaturing, 50-65°C for 30 seconds for annealing, and 70°C for 1 min/kb amplicon length for extension.

Table 2.3: KOD Xtreme PCR

Volume added (μl)	Reaction component
1	Forward primer
1	Reverse primer
1-5 (depends on source and concentration)	Template
10	2 mM dNTPs
1.25 (optional)	DMSO
25	2X Xtreme buffer
1	KOD Xtreme DNA polymerase
top up to 50	Nuclease free water

High fidelity PCR using KOD Hot Start Polymerase

Using the hot start KOD polymerase was similar to using the Xtreme polymerase with respect to thermal cycling, except the extension time was 30 seconds per kb of amplicon length. The reaction components are listed in Table 2.4, and were taken from the user manual with some modification (EMD Millipore, 2011a).

Table 2.4: KOD Hot Start PCR

Volume added (μl)	Reaction component
1	Forward primer
1	Reverse primer
1-5 (depends on source and concentration)	Template
5	2 mM dNTPs
2.5 (optional)	DMSO
5	10X KOD buffer
1	KOD hot start DNA polymerase
3	25 mM $MgSO_4$
top up to 50	Nuclease free water

2.7.9 Colony PCR

Cells were picked from a single colony and resuspended in 20-30 μl of sterile Milli-Q (Millipore Corporation) ultrapure water. 2 μl of cell suspension was spotted on an appropriate agar plate to save the colony for future use. The cell suspension was heated at 95°C for 20 minutes. The suspension was then centrifuged at max speed for 1 minute to pellet cells. 8 μl of the supernatant was used as the template

The reactions were setup on ice. 8 μl of template was added to a fresh 0.2 ml thin walled PCR tube. 1 μl of each primer (10 pmol) was added, and 10 μl of 2X *Taq* Master Mix (Thermo Fisher Scientific Inc.) was added to give a final reaction volume of 20 μl. The reactions were temperature cycled in a Bio-Rad Mycycler (Bio-Rad Laboratories Inc). The protocol was a touchdown protocol of 50 total cycles with temperatures as

follow: 95°C for 1 minute; 65°C for 30 seconds, dropped 0.5°C each cycle for 20 cycles then goes to 50°C for remaining cycles; 72°C for 1 min/kb of amplicon length; after all cycles a final extension 72°C for 5 minutes; then hold at 10°C.

2.7.10 Sanger Sequencing

Plasmids for sequencing were prepared using the silica gel column protocol for miniature preparation of plasmid DNA. DNA samples were sent to The Centre for Applied Genomics (Toronto Canada) and prepared to their specifications as follows: 200-300 ng of plasmid DNA in a 7 μl volume; if the primer was included, 5 pmol of primer with 200-300 ng of plasmid DNA in a 7.7 μl volume; for PCR products 5 ng of DNA per 100 bp of amplicon length in the same volume and primer molarity as for plasmids.

2.7.11 Genomic Sanger Sequencing

For sequencing using bacterial genomic DNA as the template, samples were sent to the Mobix Lab at McMaster University (Hamilton Canada) and prepared to the their specifications. The samples were prepared as follows: 5 μl of genomic DNA per sample at a concentration of 2 μg/μl; 5 μl of primer per sample was sent at a concentration of 20 μM.

It was important to shear the genomic DNA samples to reduce their viscosity. Sodium acetate instead of ammonium acetate was used for the ethanol precipitation step in the genomic DNA preparation procedure because ammonium ions might have inhibited the DNA polymerase in the sequencing reaction.

2.8 DNA Purification

2.8.1 Purification of DNA Using a Silica Gel Column

To the DNA containing solution there was added three volumes of binding buffer containing 5.5 M guanidine thiocyanate, 140 mM 2-(N-morpholino)ethanesulfonic acid (MES) pH to 7 with NaOH, and 20 mM ethylenediaminetetraacetic acid (EDTA). The DNA binding buffer mixture was added to a silica spin column (Bio Basic, Markham Canada). The column was centrifuged at 15 000 x g for 1 minute. The flow through was discarded and 500 μl of wash buffer, containing 75% ethanol and 10 mM tris-HCl pH 7.5, was added. Centrifugation was repeated with the same settings as above. The flow through was discarded and the wash step was repeated once more. After the flow through of the final wash was discarded, the centrifugation was repeated once or twice to dry the column. 35-50 μl of 2 mM tris-HCl elution buffer was added to the column and allowed to incubate at room temperature for 2 minutes. The column was placed in a 1.5 ml centrifuge tube and centrifuged as above. The flow through of this step was the final product, and it was either used immediately or stored at -20°C.

2.8.2 Ethanol Precipitation of DNA

One tenth volume of 3 M sodium acetate was added to the DNA solution. Then 2 volumes of ice-cold ethanol was added. If DNA shearing was not a problem the samples were then vortexed for 10 seconds. If DNA shearing was to be avoided the samples were inverted gently and DNA was allowed to precipitate on ice for 20 minutes. The samples were then centrifuged for 5-20 minutes at 17 000 x g. DNA that was smaller in length had to be spun longer. The supernatant was discarded, and the pellet was washed in 70% ethanol by pippetting up and down and dislodging the pellet. The centrifugation was repeated. The supernatant was discarded. The pellet was dried until

the edges were translucent and then the pellet was resuspended in 2 mM Tris-HCl pH 8.0.

2.8.3 Miniature Preparation of Bacterial Plasmid DNA

A single colony of plasmid containing *Escherichia coli endA1* strain was inoculated into 3 ml of LB media in a 16 mm test tube containing the appropriate antibiotic (see 2.3.1 for concentrations). The test tube was incubated for 16 hours at 37°C with constant shaking at 200 rpm.

1.5 ml of culture was added to a 1.5 ml microcentrifuge tube. The cells were pelleted by centrifugation at 15 000 x g for 1 minute. The supernatant was decanted. For a greater yield, the remaining 1.5 ml of culture was added and centrifuged as before. The cell pellet was resuspended in 250 μl of buffer P1 (100 mM Tris-HCl ph 8.0, 10 mM EDTA, 100 μg/ml RNAse A). Then to lyse the cells 250 μl of buffer P2 (0.2 N NaOH, 1% SDS) was added. The tube was inverted several times to provide gentle mixing. The cells were allowed to lyse for 1-3 minutes. Then 350 μl of buffer N3 (4.2 M Guanidine thiocyanate, 0.9 M potassium acetate, pH to 4.8 with acetic acid) was added. The tube was immediately inverted serveral times to ensure quick and even neutralization. The tube was then centrifuged for 5 minutes at 15 000 x g to a silica spin column (Bio Basic, Markham Canada). The column was centrifuged at 15 000 x g for 1 minute. The flow through was discarded 500 and μl of wash buffer (75% ethanol and 10 mM tris-HCl pH 7.5) was added. Centrifugation was repeated with the same settings as above. The flow through was discarded and the wash step was repeated once more. After the flow through of the final wash was discarded, the centrifugation was repeated once or twice to dry the column. 35-50 μl of 2 mM tris-HCl elution buffer was added to the column and allowed to incubate at room temperature for 2 minutes. The column was placed in a 1.5 ml centrifuge tube, and centrifuged as above. The

flow through of this step was the final product, and it was either used immediately or stored at -20°C.

2.8.4 Large Scale Preparation of Bacterial Genomic DNA for Direct Genomic Sanger Sequencing

A single colony was inoculated into 3 ml of LB media in a 16 mm test tube. For growth conditions and temperature see section 2.3. After this initial culture was saturated, it was added to 200 ml of LB broth in a 1 l flask. The flask was incubated with shaking. An overnight incubation would usually be sufficient even with more slowly growing species.

After growth, the 200 ml culture was decanted into a 250 ml polypropylene GSA centrifuge bottle (Nalgene, Thermo Fisher Scientific Inc.). The centrifuge was balanced with a second sample or an identical bottle containing 200 ml of water. The cells were pelleted by centrifugation in a Sorval GSA rotor or equivalent (Sorval, Thermo Fisher Scientific Inc.) at 4000 x g for 10 minutes at 20°C. The supernatant was decanted and the pellet was resuspended in 40 ml of 10 mM Tris-HCl pH 8.0 25 mM EDTA. The resuspended cells were transferred into a 50 ml disposable centrifuge tube (various manufacturers: known as "falcon tubes"). The cells were pelleted by centrifugation at 6500 x g in an Eppendorf F35-6-30 rotor (Eppendorf AG). The cells were resuspended in 15 ml of the same Tris-EDTA as above. 1.5 ml of 12.5 % SDS, 50 μl of 19 mg/ml proteinase K, and 1.8 ml of 5 M NaCl were added. The tube was incubated at 65°C for 30 minutes. The tube was cooled to room temperature and 15 ml of a 1:1 phenol chloroform mixture was added. The tube was vortexed for 15 seconds. The layers were separated by centrifugation at 6500 x g for 5 minutes in the Eppendorf f35-6-30 rotor. Then using a 5 ml pipette the phenol-chloroform layer (bottom) was removed. Fresh phenol-chloroform (1:1) was added, the tube vortexed and the centrifugation

repeated. This extra step helps to denature and remove all the protein; there were protein strands in the aqueous layer that interfered with that layer's recovery if this was not done. Next the aqueous layer was removed and placed into a new 50 ml falcon tube. Then 15 ml of chloroform was added, the tube was vortexed, and the centrifugation was repeated. The aqueous layer was then removed and placed into a new 50 ml falcon tube. One tenth volume of 3 M sodium acetate was added (1-1.5 ml depending on recovery). 30 ml of ice cold 95% ethanol was added. The tube was inverted. At this point DNA visibly precipitated. The DNA precipitate was removed from the tube with a 200 μl pipette tip loaded onto a micropipettor. The DNA precipitate was placed in a 1.5 ml microcentrifuge tube. The precipitate was washed 1 x with 1.5 ml of 70% ethanol. The ethanol was removed and the precipitate was allowed to dry until the edges were translucent. The precipitate was resuspended in 400 μl of 2 mM Tris-HCl pH 8.0. 20 μl of 10 mg/ml DNAse free RNAse A was added to the tube. The tube was allowed to incubate at 20°C for 2 hours. The 400 μl of DNA solution was split into 4 100 μl aliquots and each of them was purified as described in section 2.8.1 except the final elution was done with 50 μl of 0.2 mM Tris-HCl pH 8.0. After purification, the 4 tubes containing 50 μl of DNA solution were pooled into 1 tube. The DNA was quantified with a nanodrop spectrophotometer (Thermo Fisher Scientific Inc.). The concentration and purity was noted. The absorbance at 260 nm to 280 nm ratio was typically 1.88, indicating pure DNA with little RNA or protein contamination. The absorbance at 260 nm to 230 nm ratio was greater than 1.6 indicating little contamination with choatropic salts. The DNA was then concentrated by using the DNA speedVac concentrator (Savant, Thermo Fisher Scientific Inc). The volume left after concentrating was estimated by trial and error with a micropipettor and the volume was adjusted by adding 0.2 mM Tris-HCl pH 8.0 to give a final concentration of 2 μg/μl.

2.8.5 Miniature Preparation of Bacterial Genomic DNA

This method was adapted from the method published by Charles and Nester (1993).

A single colony was inoculated into 3 ml of LB media in a 16 mm test tube. For growth conditions and temperature, see section 2.3. A 1.5 ml volume of culture was decanted into a 1.5 ml microcentrifuge tube, and the cells were pelleted by centrifugation at 15 000 x g for 1 minute. The supernatant was discarded and the remaining 1.5 ml of culture was added to the tube and the tube was centrifuged again. The pellet was washed in 1 ml 10 mM Tris-HCl pH 8.0 25 mM EDTA, then the pellet was resuspended in 400 μl of Tris-HCl pH 8.0 with 25 mM EDTA. 40 μl of 12.5 % SDS was added; 10 μl of 19 mg/ml proteinase K was added; 50 μl of 5 M NaCl was added. Then the tube was incubated at 65°C for 30 minutes. 260 μl of 7.5 M ammonium acetate was added, and the tube was incubated on ice for 30 minutes to allow protein to precipitate. The tube was then centrifuged for 15 minutes at 17 000 x g to pellet the protein. The supernatant was saved and added to a new tube. The protein pellet was discarded. 1 volume of chloroform was added to the solution (depending on recovery from protein pellet about 700 μl). The tube was inverted for a minute to extract organic soluble impurities. Then the tube was centrifuged for 5 minutes at 15 000 x g, to separate the phases. The aqueous phase (top layer) was removed and added to a new tube. Then 1 volume of isopropanol was added, the tube was inverted to mix, and the tube was left for 20 minutes for the DNA to fully precipitate. Sometimes the DNA was immediately visible as a string-like precipitate; in those cases the DNA was immediately removed on a pipette tip, dipped in a liberal volume (1.5 ml) of 70 % ethanol and then removed to a clean tube to dry. When the precipitate was not so abundant, the tube was spun at 15 000 x g for 10 minutes; the supernatant was removed; the pellet was washed with 70 % ethanol and allowed to dry after the 70 % ethanol was removed. Once the DNA precipitate was dry the pellet was resuspended in 50-200 μl of 2 mM Tris-HCl pH 8.0.

The DNA could be used at this point for some applications. To remove RNA 1 μl of 10 mg/ml RNAse A was added. After incubation at room temperature for 20 minutes the DNA was re-purified by ethanol precipitation (see section 2.8.2).

2.9 Southern Blotting

Reagents were purchased as part of the Roche DIG High Primer DNA Labeling and Detection Starter Kit II. Protocols from the manual (Roche Applied Science, 2009) were followed and are described below. Where there is a lapse in the Roche protocol, or where it refers to standard methods, the Southern blot protocol from Current Protocols in Molecular Biology (Brown, 2001) was used.

The solutions used were as follows.

Maleic Acid Buffer

0.1 M maleic acid, 0.15 M NaCl, adjust pH to 7.5 with solid NaOH

Washing Buffer

maleic acid buffer with 0.3% v/v Tween 20

Detection Buffer

0.1 M Tris-HCl, 0.1 M NaCl, pH 9.5

Blocking Solution

1:10 dilution of supplied blocking solution in maleic acid buffer

Antibody Solution

1:10 000 dilution, in maleic acid buffer, of the provided anti-DIG-AP-conjugate (provided vial centrifuged at 15 000 x g for 10 minutes prior to use)

Hybridization Buffer

64 ml of water with 1 bottle of DIG Easy Hyb Granules.

20 X SSC Buffer

For 1 L: 175.3 g NaCl, 88.2 g trisodium citrate, adjust to pH 7.0 with HCl, add Milli-Q water to 1 L, sterilized by autoclaving.

Stringency Wash 1 Buffer

2 x SSC, 0.1% SDS

Stringency Wash 2 Buffer

0.5 x SSC, 0.1% SDS

Denaturing Solution

1.5 M NaCl, 0.5 M NaOH

Neutralization solution

1.5 M NaCl, 0.5 M Tris-HCl, pH 7.0 (adjusted with HCl)

2.9.1 Creation and Quantification of DIG Labeled Probe

DNA was random primed labeled with Digoxigenin-11-dUTP by adding 30-100 ng of template DNA to the labeling reaction. First the DNA was denatured by boiling for

10 minutes and quickly chilling in ice-water. The volume of the DNA was adjusted to 16 μl. 4 μl of DIG-High Prime reagent was added. The tube was then incubated at 37°C for 16 hours. The reaction was then stopped by adding 2 μl of 0.2 M EDTA. The concentration of the probe was estimated from the chart provided in the manual at 1000-1500 ng in 20 μl.

Based on the estimated concentration a sample of the probe was diluted to 1 ng/μl. A 10-fold dilution series to 0.01 pg/μl was made. The same dilution series was made with the provided DIG-labeled control DNA. 1 μl of each dilution was spotted on a nylon membrane (Boehringer Mannheim, acquired by Roche). The DNA was cross-linked to the membrane by UV-light exposure in a BioRad GS Gene Linker UV chamber on the damp nylon membrane setting (150 mJoule). The membrane was then incubated in 20 ml of maleic acid buffer on a rotary shaker for 2 minutes at 20°C. The membrane was incubated in 10 ml of blocking solution for 30 minutes. The membrane was incubated in 10 ml of anti-DIG AP conjugate antibody solution for 15 minutes; the solution was drained and then fresh solution was added for another 15 minute wash. The membrane was washed in 10 ml of washing buffer for 15 minutes; the buffer was drained; fresh buffer was added, and the wash was repeated. The membrane was then equilibrated in 10 ml of detection buffer for 5 minutes. The membrane was placed DNA side up on a plastic folder bag, 100 μl of CPSD ready to use solution was evenly spread on the membrane, and the membrane was sealed between two plastic sheets using the heat sealer. The membrane was then placed in a digital enhanced chemiluminescence (ECL) imager. A 15 minute exposure was taken. The brightness of the dots were evaluated by eye and compared to the control to determine the concentration of the probe.

2.9.2 Transfer of DNA onto Membrane

The genomic DNA was isolated as described in section 2.8.5. 1-3 μg of DNA was restricted as described in section 2.7.7. Then the restricted DNA was electrophoresed as described in section 2.7.2 but with care to use a lower voltage than usual (1/4 to 1/2 of usual) to ensure even running. A ruler was included in the photograph of the gel so a standard curve of band migration could be generated.

After electrophoresis was complete, the gel was prepared as follows. First, the gel was rinsed in deionized water. Next, the gel was incubated with shaking in 10 gel-volumes of 0.25 M HCl until 10 minutes after the bromophenol blue dye from the electrophoresis loading dye turned yellow (Brown, 2001). The gel was rinsed with deionized water again. The gel was incubated with shaking in 10 gel volumes of denaturing solution for 20 minutes. The solution was poured off and fresh denaturing solution was added for another 20 minute incubation. Next the gel was incubated with shaking as before but with 10 gel volumes of neutralization solution for 20 minutes. This was repeated with fresh neutralization solution.

The gel was then assembled into the transfer stack. A glass casserole dish was used to hold the liquid and the stack. Next a DNA gel casting tray was placed upside down in the dish for use as a solid support. A piece of Whatman 3 MM chromatography paper that was the same height as the solid support, but wider, was used as a wick. The paper was folded under the solid support on the sides, such that the side of the paper would be immersed in the liquid. Eight pieces of Whatman 3 MM were cut to the exact dimensions of the gel. The Whatman 3 MM pieces were made wet with 10 x SSC (including the wick). A piece of nylon membrane was cut to the exact dimensions of the gel. The nylon membrane was equilibrated in deionized water. Next, four pieces of wet Whatman 3 MM were placed on the wick covered solid support. The prepared gel was placed on top of these. The prepared membrane was placed on top of the

gel. The other 4 pieces of wet Whatman 3 MM were placed on top of the membrane. a 4 cm stack of paper towels, cut to the dimensions of the gel (or slightly smaller), was placed on top of the Whatman 3 MM sheets. A glass plate was placed on top of the paper towels. A weight (usually an Erlenmeyer flask with water in it) was placed on top of the glass plate. The casserole dish was filled about 2 cm deep with 10 x SSC. The stack was wrapped with plastic food wrap to hold it together and prevent excessive evaporation. The stack was allowed to sit for 12 hours. The next day it was disassembled.

The membrane was placed in the DNA cross-linker and the DNA was cross-linked by exposure to 150 mJoule of UV radiation (as above).

2.9.3 Hybridization of the Probe

The appropriate hybridization temperature T_{opt} was determined by the following equations where l is the length of the probe in base pairs.

$$T_m = 49.82 + 0.41(\%GC) - (600/l)$$

$$T_{opt} = T_m - 20$$

These numbers are relatively accurate when T_{opt} is above 25°C.

A volume of hybridization buffer (10 ml/100 cm^2) was preheated to the hybridization temperature (T_{opt}), and added to a preheated VWR hybridization bottle with the membrane. The bottle was placed in the hybridization oven (VWR 5400 oven, VWR International LLC Mississauga ON). The membrane was incubated in the oven with the hybridization buffer for 30 minutes with gentle rotation.

The DIG labeled probe was denatured by boiling for 10 minutes and quickly cool-

ing in ice water. The probe was added to preheated hybridization buffer at a concentration of 25 ng/ml. The probe solution (the probe in hybridization buffer) was mixed well but gently to avoid bubble formation. 3.5 ml of probe solution was used to for 100 cm^2 of membrane area. Before adding the preheated (to T_{opt}) probe solution, the prehybridization solution was poured off. The membrane was incubated with probe solution for hybridization of the probe to the bound DNA for 16 hours at T_{opt}.

After hybridization, The probe solution was poured off. two stringency washes were carried out. In the first wash, 20 ml of stringency wash 1 buffer (an excess amount; more could be used if desired) was added and the membrane was incubated for 5 minutes at 25°C with constant rotation. The buffer was poured off and fresh buffer was added for another 5 minute incubation. The next stringency wash was done in the same way but with stringency wash 2 buffer, a temperature of 68°C, and a time of 15 minutes for each buffer change.

2.9.4 Detection of the Probe

The membrane was rinsed for 1-5 minutes in washing buffer. Then the membrane was incubated for 30 minutes in 100 ml of blocking solution on a rotary shaker at 20°C. The membrane was incubated in 20 ml of anti-DIG AP conjugate antibody solution for 30 minutes. The membrane was washed in 100 ml of washing buffer for 15 minutes. The buffer was drained; fresh buffer was added, and the wash was repeated. The membrane was then equilibrated in 20 ml of detection buffer for 5 minutes. The membrane was placed DNA side up on a plastic folder bag. Then 500 μl of CPSD ready to use solution was evenly spread on the membrane, and the membrane was sealed between two plastic sheets using the heat sealer. The membrane was then placed in a digital enhanced chemiluminescence (ECL) imager. A 15 minute exposure was taken. A ruler was included in the photograph to locate the probed band(s) on a standard curve gen-

erated from the DNA molecular weight standards on the original electrophoresis gel. This was done to provide accurate sizing of the probed band(s).

2.10 β-Glucuronidase (GusA) Assay

A 3-5 ml saturated culture was prepared as described in section 2.3. The OD_{600} was measured in a Spectronic 20D spectrophotometer. 100 μl of culture or 100μl of a diluted culture was used. 400 μl of assay buffer (0.0125% SDS, 50 mM sodium phosphate buffer pH 7.0, 1 mM EDTA, 458 μg/ml PNPG) was added and the reaction was incubated for 30 minutes at 20°C. 500 μl of stop buffer (1 M Na_2CO_3) was added. The reaction was added to a 1 ml cuvette and the absorbance at 405 nm (A_{405}) was measured in a Pharmacia Biotech Ultrospec 2000 UV/visible spectrophotometer.

The GusA activity in Miller units (Miller, 1972) were calculated as follows. Where $t =$ incubation time in minutes, and $v =$ volume in ml.

$$\text{Miller Units} = \frac{1000 \times A_{405}}{t \times OD_{600} \times v}$$

2.11 Sequence Alignments, Genomic and Plasmid Maps

The sequence alignments, plasmid maps, and genomic maps shown in this thesis were created with Geneious R8 software (Kearse *et al.*, 2012).

Chapter 3

Integrase Mediated Cassette Exchange

3.1 Introduction

Integrase mediated cassette exchange (IMCE) takes advantage of the site-specific integration activity of the ΦC31 integrase. This integrase is a member of the family of large serine recombinase proteins (Leschziner *et al.*, 1995; Thorpe and Smith, 1998). The large serine recombinases, like other site-specific recombinases, catalyze a strand exchange reaction between two sites of a specific DNA sequence. In the case of ΦC31 integrase *attB* and *attP* sites (34 bp and 39 bp in length, respectively) are recombined with each other. Such recombination results in *attL* and *attR* sites, which are composites of *attB* and *attP* sites. Even in the presence of integrase, the *attL* and *attR* sites do not recombine. Although the nomenclature used to describe the species required for recombination is similar to the λ phage recombination system used in Gateway technology (Katzen, 2007), the large serine recombinases differ from this system in that they require no cofactors. They work by a concerted strand exchange mechanism and do not go through a Holliday junction intermediate like in λ integrase recombination (Gupta *et al.*, 2007; Mumm *et al.*, 2006). The recombination mechanism is shown in Figure 3.2. They are considered large serine recombinases since they have

a large carboxy-terminal domain not present in other serine recombinases. In the case of ΦC31 integrase the carboxy-terminal domain is responsible for conferring non-permissiveness of recombination between *attL* and *attR* (Rowley *et al.*, 2008). To the best of my knowledge, at the time of writing six enzymes from this family, including ΦC31 integrase, have been characterized (Christiansen *et al.*, 1996; Crellin and Rood, 1997; Kushtoss and Rao, 1991; Matsuura *et al.*, 1996; Popham and Stragier, 1992; Sato *et al.*, 1990; Thorpe *et al.*, 2000).

When the required *att* sites are in single copy, one on each DNA molecule to be recombined, the end result of recombination is a cointegrant. This configuration mirrors the wild type system of the phage where the phage genome integrates into the host genome during lysogeny. When the *att* sites are arranged to flank a sequence, and the corresponding *att* sites are on a separate DNA molecule, the end result of a ΦC31 catalyzed recombination between the two molecules is an exchange of the two sequences in between the *att* sites. This configuration results in cassette exchange (Figure 3.3). The use of a cassette exchange system for the integration of genetic constructs into the chromosome allows for the consistent insertion of sequence in a reliable and repeatable way. It simplifies the process of integrating constructs. The chromosome can be used to maintain these constructs more stably than on a plasmid, without the requirement for selection.

The integration can be done in a similar way, practically speaking, as homologous recombination using conjugative plasmids (Schafer *et al.*, 1994). That is, to use bacterial conjugation to transfer the donor plasmid, containing *attB* sites, into the recipient strain, containing the corresponding *attP* sites. The main distinguishing differences are that the *att* sites are present in both the recipient chromosome and the donor plasmid in the place of homologous DNA and that an integrase enzyme must be expressed. The ΦC31 integrase may be expressed from another plasmid that is

transferred to the recipient at the same time as the donor plasmid.

In this chapter, I describe IMCE as the integration, through double ΦC31 catalyzed recombination, of an *attB* flanked segment of DNA from a plasmid vector such that it replaces an *attP* flanked segment of DNA resident on the chromosome. The strain containing the *attB* flanked DNA segment is referred to as the landing pad strain (LP-strain). The LP-strains had to be constructed, because ΦC31 integrase sites do not naturally occur in the genomes of most bacteria outside the natural *Streptomyces* host of ΦC31. The absence of *att* sites is advantageous because the integration context may be chosen. The vectors containing an *attB* flanked cassette and the vector used to express the ΦC31 integrase (pJC2) also had to be constructed.

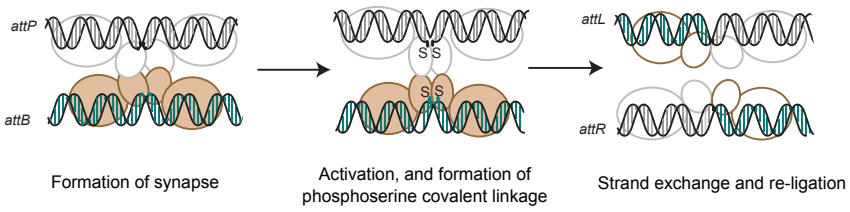

Figure 3.1: The ΦC31 integrase mechanism: Integrase tetramer DNA complex. Figure adapted from (Gupta *et al.*, 2007)

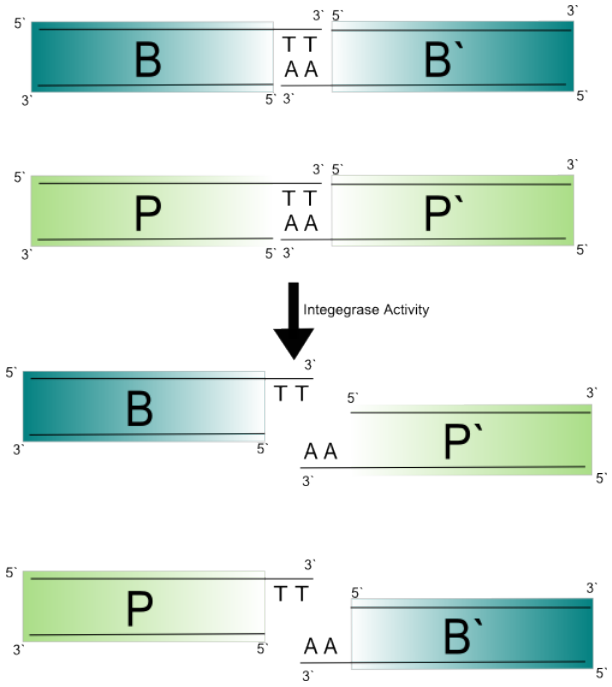

Figure 3.2: The ΦC31 integrase mechanism showing only stylized DNA sites. The arm sequences (B,B′, P, P′) surround a core dinucleotide (TT). The base-pairing of the core dinucleotide is required for the *att* sites to be ligated in the recombinant configuration as shown below the arrow.

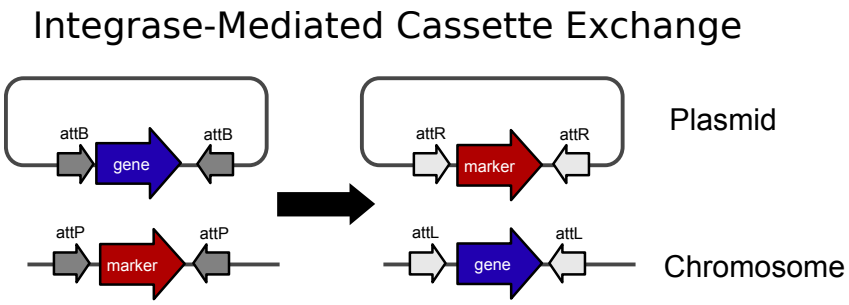

Figure 3.3: **A**: Integrase mediated cassette exchange. In the presence of integrase (not shown), recombination between *attB* (GTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCG) and *attP* (CCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGG) results in the exchange of marker genes on the chromosome with the "gene(s)" on the plasmid. This leaves *attL* (GTGCCAGGGCGTGCCCTTGAGTTCTCTCAGTTGGGGG) and *attR* (CCCCAACTGGGGTAACCTTTGGGCTCCCCGGGCGCG) sites, which are not substrates adequate for recombination by ΦC31 integrase alone.

3.2 Materials and Methods

3.2.1 Construction of SmUW227 LP-Strain

Sinorhizobium meliloti homologous DNA segments were first cloned into pK19mobSacB (Table 2.1). The homologous DNA was chosen to target the landing pad sequence to integrate into a large intergenic region between the ORF: SMc04453 and SMc01885 (Figure 3.4). The fragments were PCR amplified using *S. meliloti* Rm1021 DNA as the template. The primer sets used were, L-fwd-EcoRI with L-rev-2-PacI-SmaI to create the left fragment, and R-fwd-PacI-SphI with R-rev-HindIII (Table 2.2). These two PCR steps created the left and right homologous fragments, respectively. The left fragment was cloned into pK19mobSacB as a EcoRI-SmaI fragment, also inserting a PacI site on the right (or inside) edge of the left fragment, and creating pJH101. The right fragment was cloned into pJH101 as an SphI-HindIII fragment, also inserting a PacI site on the inside edge (left edge of right fragment), to create pJH102. The plasmid pJH102 has both homologous regions flanking two added PacI sites, placed for future subcloning of the landing pad sequence. Next, the landing pad sequence was constructed by cloning two genes to construct a poly-cistron. *attP* sites were added on the ends of these genes such that the resultant construct was flanked by *attP* sites. The first gene to be cloned was *gusA*. It was amplified from the plasmid pTH1227 (2.1), using the primers, gusA-F-SalI with gusA-attP-R-SphI. The *gusA* gene was amplified so that a ribosome binding site sequence was included before the start codon of the open reading frame. The *gusA-attP* fragment was cloned into pJH102 using SalI and SphI to create pJH103. The *aadA* gene (streptomycin-spectinomycin resistance) including its promoter was PCR amplified using pHP45Ω (Table 2.1) as the template. The primers used were attP-aadA-F-SmaI with aadA-R-XbaI (2.2). The resultant *aadA-attP* fragment was cloned into pJH103 using SmaI and XbaI to create pJH104. The *aadA* gene

was cloned upstream of the *gusA* gene so that *gusA* would be transcribed.

To integrate the completed landing pad sequence into *S. meliloti* Rm1021, pJH104 was used as a donor in a triparental mating (see Section 2.5 where *S. meliloti* Rm1021 was used as the recipient, and *Escherichia coli* MT616 was used as the mobilizer). The conjugation mixture was plated on TY agar containing neomycin at 200 μg/ml to select for the integration of the whole plasmid into the *S. meliloti* Rm1021 chromosome, and streptomycin at 200 μg/ml to select against the donor and mobilizer. The resultant cointegrant colonies contained the integrated pJH104. Next, a cointegrant was streaked on TY 10% sucrose containing 100 μg/ml spectinomycin. The resultant colonies were either *sacB* mutants or landing-pad strains (with an integrated landing pad sequence). To differentiate between the two possibilities, colonies were screened on TY agar plates containing 200 μg/ml neomycin. Colonies that were sensitive to neomycin had the vector backbone removed. A neomycin sensitive colony was chosen and streak purified three times. The LP-strain (SmUW227) was later verified by Southern blot analysis.

3.2.2 Construction of Ow0013 LP-Strain

To construct the *Ochrobactrum anthropi* LP-strain, the *ampC* gene (locus tag: OANT RS13415) (ampicillin resistance) was chosen as a replacement target for the landing pad sequence. The integration of the landing pad sequence would cause the deletion of the entire *ampC* ORF. The targeting construct was assembled *in vitro* through a 2 step PCR protocol. First, *O. anthropi* ATCC 49188 (Table 2.1) DNA was used as the template in two PCRs: one contained the primers ampC-US-F-EcoRI and ampC-US-R-XO, and the other containing the primers ampC-DS-F-XO, and ampC-DS-R-PstI (Table 2.2). Subsequently, both PCR products were used as a template with the primers ampC-US-F-EcoRI and ampC-DS-R-PstI. Because of overlapping sequence between the two PCR

products, the PCR results in assembly of the two products. A PacI site in the centre of the upstream and downstream portions was also inserted. This assembled PCR product was cloned into pK19mobSacB as an EcoRI-PstI fragment to create pJH113. The landing pad sequence was then subcloned from pJH104 into pJH113 as a PacI fragment to create pJH115.

To make an *O. anthropi* recipient that could be properly selected in a triparental mating, a rifampicin resistant derivative was isolated by plating 50 μl of *O. anthropi* ATCC 49188 saturated culture onto LB agar containing 100 μg/ml of rifampicin. A colony was chosen and streak purified three times. The culture was then saved and named Ow0012.

To integrate the landing pad sequence into Ow0012 the triparental mating procedure was followed, selecting with kanamycin at 30 μg/ml and rifampicin at 50 μg/ml. Then the double recombinants were selected with sucrose and streptomycin at 200 μg/ml. Kanamycin sensitive colonies were isolated next. Colonies that were streptomycin resistant and sucrose sensitive were putative LP-strain candidates. After isolation of the LP-strain, Ow0013, its ampicillin sensitivity was confirmed.

3.2.3 Construction of At11142 LP-Strain

To construct the *Agrobacterium tumefaciens* LP-strain, the *tetA* gene (locus tag: Atu4206) (cryptic tetracycline resistance) (Luo and Farrand, 1999) was targeted for replacement. The integration of the landing pad sequence was targeted such that the whole *tetA* ORF is deleted upon its integration. The cloning strategy was as described for Ow0013 construction, except the primers used to create pJH119 were US-tetA-F-EcoRI, US-tetA-R-XO, DS-tetA-F-XO, and DS-tetA-R-PstI; respectively substituting for: ampC-US-F-EcoRI, ampc-US-R-XO, ampC-DS-F-XO, and ampC-DS-R-PstI (Table 2.2).

For the same reasons as in the construction of Ow0013, a rifampicin resistant mu-

tant of *A. tumefaciens* was required. Therefore, *A. tumefaciens* A136 was used (Table 2.1). For the integration of the LP-sequence, contained in pJH119, into *A. tumefaciens* A136 the same procedure used for the construction of Ow0013 was used.

3.2.4 Construction of pJC2 ΦC31 Integrase Expression Plasmid

This plasmid was constructed by Dr. Jiujun Cheng.

The ΦC31 integrase gene (*int*) was amplified by PCR with the primers: INT-F and INT-R (table 2.2) and the template: pHS62 (table 2.1). The PCR product was amplified as a PstI-BamHI fragment. The restriction sites were added with 5′ extensions on the PCR primers. The INT-F primer also added stop codons in all three reading frames just downstream of the PstI site. A ribosome binding site was included on the INT-F primer downstream of the stop codons. The PCR product was cloned into pRK7813 as a PstI-BamHI fragment. The *lacI* promoter on the vector drives expression of *int*.

3.2.5 Construction of pJH110 IMCE Donor Plasmid

The primers: GmR-fwd and GmR-rev (Table 2.2) were used to amplify the *aacA* gene using pPH1JI (Table 2.1) as a template. An SphI site was added with the GmR-fwd primer 5′ extension. An *attB* site and HindIII site were added with the GmR-rev primer 5′ extension. This PCR product was cloned into pK19mobSacB as an XbaI-PstI fragment to create pJH107 (Table: 2.1). The primers US-XbaI-RFP-attB-F and US-PstI-RFP-R (Table 2.2) were used to amplify the *rfp* gene with a constitutive promoter, using the J23118 biobrick as a template (Table 2.1). The *attB* site was added by the forward primer, and an SphI site next to the *attB* site was also included. This PCR product was cloned into pJH107 as a XbaI-PstI fragment to create pJH110. Because of the SphI sites added during the vector construction, an SphI digest can be used to replace *rfp* and leave the *attB* sites intact, thereby cloning the desired construct between the two *attB*

sites.

3.2.6 Construction of pJH122 IMCE Donor Plasmid

The primers, hph-red-1a-F and hph-red-2a-R were used to amplify *hph*, including the cauliflower mosaic virus promoter using pMB403 (Table 2.1) as a template. The primers added 40 bp of DNA homology surrounding the *aac1* gene in pJH110. Using this homology, λ Red recombineering (Section 2.6) was used to replace *aac1* in pJH110 to create pJH122.

3.2.7 Southern Blotting

Southern blotting was used to verify the insertion of the landing pad sequences into the LP-strains. The procedure followed was as stated in Section 2.9.

SmUW227

The template for DIG-labeled probe synthesis was the right fragment amplified with the primers: R-fwd-PacI-SphI and R-rev-HindIII (Table 2.2), with pJH104 (Table 2.1) as the template. The *S. meliloti* SmUW227 and Rm1021 (parental control) genomic DNA was cut with SalI.

Ow0013

The template for DIG-labeled probe synthesis was the upstream of *ampC* fragment amplified with the primers: ampC-US-F-EcoRI and ampC-US-R-XO (Table 2.2), with pJH113 (Table 2.1) as the template. The *O. anthropi* OW0013, and Ow0012 (parental control) genomic DNA was cut with EcoRV and HindIII.

3.2.8 IMCE Tetraparental Mating

Preparation of Culture

Cultures of the donor (pJH110, or other in *E. coli*), landing pad recipient (SmUW227 or other see 2.1), integrase helper strain (pJC2 in *E. coli* DH5α), and mobilizer (MT616) were grown in 5 ml of LB media supplemented with antibiotics when required for plasmid maintenance. Typically, those antibiotics were: kanamycin at 10 μg/ml for the donor plasmid, tetracycline at 5 μg/ml for the integrase helper plasmid, and chloramphenicol at 10 μg/ml for the mobilizer plasmid. The *S. meliloti* landing pad strain was incubated at 30 °C with constant shaking at 180 RPM for 2-3 days, until the culture was saturated. Other landing pad strains, that may substitute SmUW227 (*A. tumefaciens*, *O. anthropi*), were incubated under the same conditions for 16-20 hours. The other strains, in *E. coli* backgrounds, were incubated at 37 °C for 16-20 hours with constant shaking at 180 RPM.

Mating

Once cultures were grown, 1.5 ml of culture was removed and washed 2 times in sterile 0.8% NaCl. The pellet was then resuspended in 100 μl of sterile 0.85% NaCl. Then 10 μl of resuspended culture was then spotted on a TY agar plate. These spots served as negative controls. For a no-integrase negative control 40 μl of each resuspended culture, excluding pJC2 was added to a sterile 1.5 ml centrifuge tube and mixed by pippetting up and down. The mixed culture was spotted on a TY agar plate. The tetraparental mating was prepared in the same way except 40 μl of the pJC2 resuspended culture was added. The spots were dried by placing the plates open in a laminar flow hood. The plates were incubated at 30 °C for 16-20 hours.

Selection

After the incubation the control spots were streaked on selective media (see table 3.1) that was selective for transintegrants. The tetraparental mating spot was also streaked on the same media. One quarter to one half of the tetraparental mating spot was resuspended in 0.85% NaCl and a 10 fold dilution series to 10^{-7}. Typically, this was done by diluting 10 μl of culture into 90 μl of 0.85% NaCl. 40 μl of each dilution, 10^{-2}-10^{-5}, was spread plated on media selective for transintegrants; 40 μl of each dilution 10^{-4}-10^{-7} on media selective for recipients and transintegrants. Plates were then incubated at 30 °C for three days in the case of *S. meliloti* landing pad strains, and two days for *O. anthropi* and *A. tumefaciens* landing pad strains.

Table 3.1: Selection of transintegrants from various donor recipient combinations. The construction of pJRT vectors is described in subsequent chapters

Donor	Landing pad strain	Transintegrant selection	Recipient and transintegrant selection
pJH110; pJRT15; pJRT16	SmUW227 and derivatives	Streptomycin 200 μg/ml; Gentamicin 30 μg/ml	Streptomycin 200 μg/ml
pJH122; pJRT10; pJRT11; pJRT12	SmUW227 and derivatives	Streptomycin 200 μg/ml; Hygromycin 50 μg/ml	Streptomycin 200 μg/ml
pJH110; pJRT15; pJRT16	Ow0013; At11142	Rifampicin 50 μg/ml; Gentamicin 30 μg/ml	Rifampicin 50 μg/ml
pJH122; pJRT10; pJRT11; pJRT12	Ow0013; At11142	Rifampicin 50 μg/ml; Hygromycin 200 μg/ml	Rifampicin 50 μg/ml

3.2.9 Colony PCR

Colony PCR, as described in Section 2.7.9, was used to confirm the integration of *attB* flanked cassettes. It was also used to ascertain the orientation of the integrated cassette. The primers used were L-fwd-EcoRI with GmR-rev or R-rev-HindIII with HyR-Fwd (Table 2.2). The template used was the particular IMCE colony derived from *S. meliloti* LP-strain tetraparental matings. The primer set with HyR-Fwd was used with IMCE involving pJH122 and its derivatives; the set with GmR-rev was used with IMCE involving pJH110 and its derivatives. When screening transintegrants derived from Ow0013, the R-rev-HindIII primer was replaced by ampC-DS-R-PstI, and L-fwd-EcoRI was replaced with ampC-US-F-EcoRI. When screening transintegrants derived from At11142, the R-rev-HindIII primer was replaced by DS-tetA-R-PstI, and L-fwd-EcoRI was replaced with US-tetA-F-EcoRI.

3.2.10 Visualization of RFP Fluorescence

Colonies containing RFP on agar Petri plates were visualized by shining white light (from a white phosphor plate on a UV transilluminator) through a LEE filters 139 primary green filter, onto the plate, and then placing a LEE filters 128 bright pink filter on the plate. This assembly was inside a dark box with a camera mounted on top (Heil *et al.*, 2011).

3.3 Results

3.3.1 LP-strain Confirmation

The objective was to construct three LP-strains, one in *Sinorhizobium meliloti*, *Ochrobactrum anthropi*, and *Agrobacterium tumefaciens*. After their construction, the LP-strains were confirmed by phenotypic analysis and Southern Blot analysis. Before moving ahead with the applications of IMCE discussed later in this thesis, it was important to verify the newly constructed landing pad strains. The vectors that go with the strains were verified as part of the conventional cloning process. Later analysis of the resultant colonies from IMCE (colony PCR) also showed consistent results with the analysis discussed in this section.

SmUW227

The *S. meliloti* LP-strain SmUW227 has the expected phenotypes associated with the landing pad sequence integrated into its chromosome: GusA activity and spectinomycin resistance. Rm1021 is streptomycin resistant to begin with, due to a ribosomal mutation. SmUW227 is resistant to spectinomycin at 100 μg/ml. SmUW227 has a GusA activity of 49 Miller units, whereas the wildtype Rm1021 has no GusA activity and is spectinomycin sensitive. The colonies of SmUW227 are also blue, After 5 days of growth on TY plates containing 200 μg/ml of X-gluc (a GusA substrate). Rm1021 has no GusA activity so its colonies never turn blue on this media. SmUW227 is sensitive to kanamycin and resistant to sucrose, as is expected after the sucrose selection step in its construction.

The integration of the landing pad sequence into the Rm1021 chromosome does not delete any genes, only a short stretch of intergenic sequence. Therefore, it was necessary to use Southern blotting to verify the location of the integrated landing pad

sequence in the chromosome. The Southern blotting results (Figure 3.4) show that the expected landing pad sequence is integrated in the targeted locus and that the vector backbone sequence is not present in SmUW227, because a much larger SalI fragment would have been detected had the entire vector been integrated (data not shown). SmUW227 also yields the proper sized amplicon in colony PCR experiments post IMCE (Section 3.3.4). SmUW227 works with IMCE, and the landing pad sequence was integrated in the intended locus.

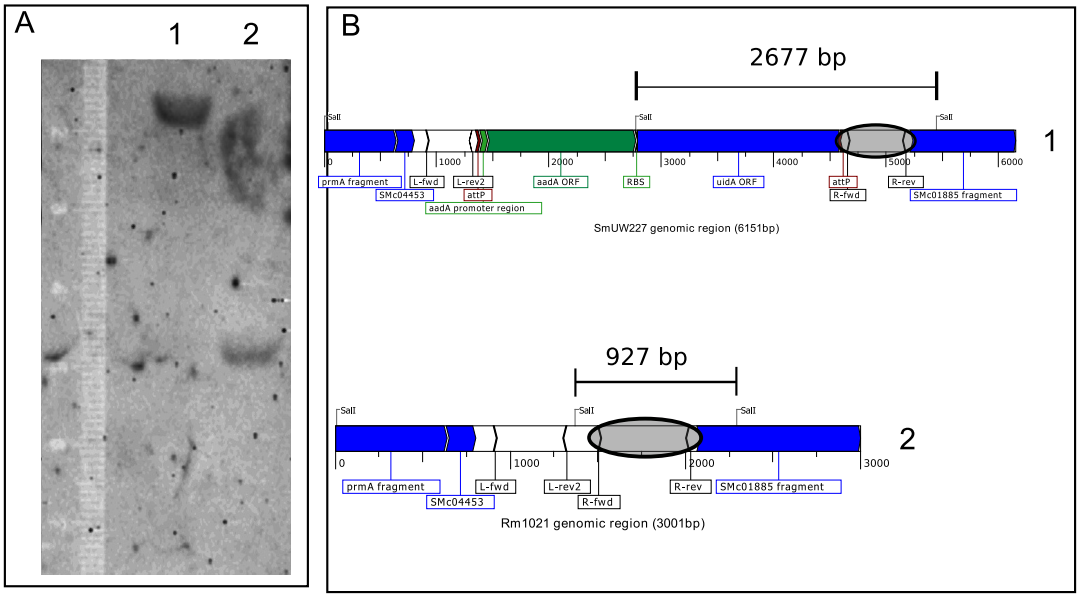

Figure 3.4: *S. meliloti* SmUW227 (LP-strain) Southern blot. An image of the Southern blot is shown in panel A, on the left. Lane A1 contains SmUW227 genomic DNA, and the size of the probed SalI fragment corresponds with map 1 in panel B (expected SmUW227 sequence), on the right. The same is true for lane A2 (the parental control: Rm1021) and map B2. The expected migration of the probed band for the LP-strain (map B1) is 2.3 cm (2.6 kb, lane A1), and ran at 1.9 cm; for the parental strain (map B2) it is 4 cm (0.9 kb, lane A2), and ran at 4.8 cm. The expected sizes are based on the standard curve made from the 1 kb ladder (Fisher Scientific Inc) on the original agarose gel (A.4). These differences from expected migration can be accounted for by the fact that both bands are out of the linear range of agarose resolution.

Ow0013

A similar approach as the one used to verify SmUW227 was used to verify that Ow0013 was constructed correctly. There was an additional phenotype to check, because the integration of the landing pad sequence causes the deletion of *ampC* instead of a small section of non-coding DNA.

The *O. anthropi* LP-strain Ow0013 also has the expected phenotypes associated with integration of the landing pad sequence, GusA activity and spectinomycin/streptomycin resistance. Ow0013 is sensitive to kanamycin and sucrose resistant, indicating the intended loss of the vector backbone sequence. The colonies are blue after 2 days growth on LB plates containing 100 μg/ml of the GusA substrate, X-gluc. Ow0012 has no GusA activity; its colonies never turn blue on this media.

The integration of the landing pad sequence by double recombination into Ow0012 causes the deletion of the *ampC* gene, coding for a broad-substrate range β-lactamase. Ow0012 is ampicillin resistant at 100 μg/ml in LB plates. Ow0013 is completely sensitive to ampicillin at that concentration. The ampicillin sensitivity of OW0013 indicates that the landing pad sequence is integrated in the intended locus, and that *ampC* is deleted. It is extremely unlikely to delete *ampC* without a double recombination event having occurred. A Southern blot (Figure 3.5) shows the expected physical arrangement of the DNA as well. The colony PCR experiments (Section 3.3.4) post IMCE give the proper sized amplicon. Ow0013 works in IMCE, and there is no doubt that the landing pad sequence is integrated in the correct locus.

At11142

As with the previous two LP-strains At11142 was verified using the same methods, except a Southern blot of equivalent quality to the other blots was not obtained. The current Southern result is shown in Figure 3.6. The *A. tumefaciens* LP-strain At11142, like

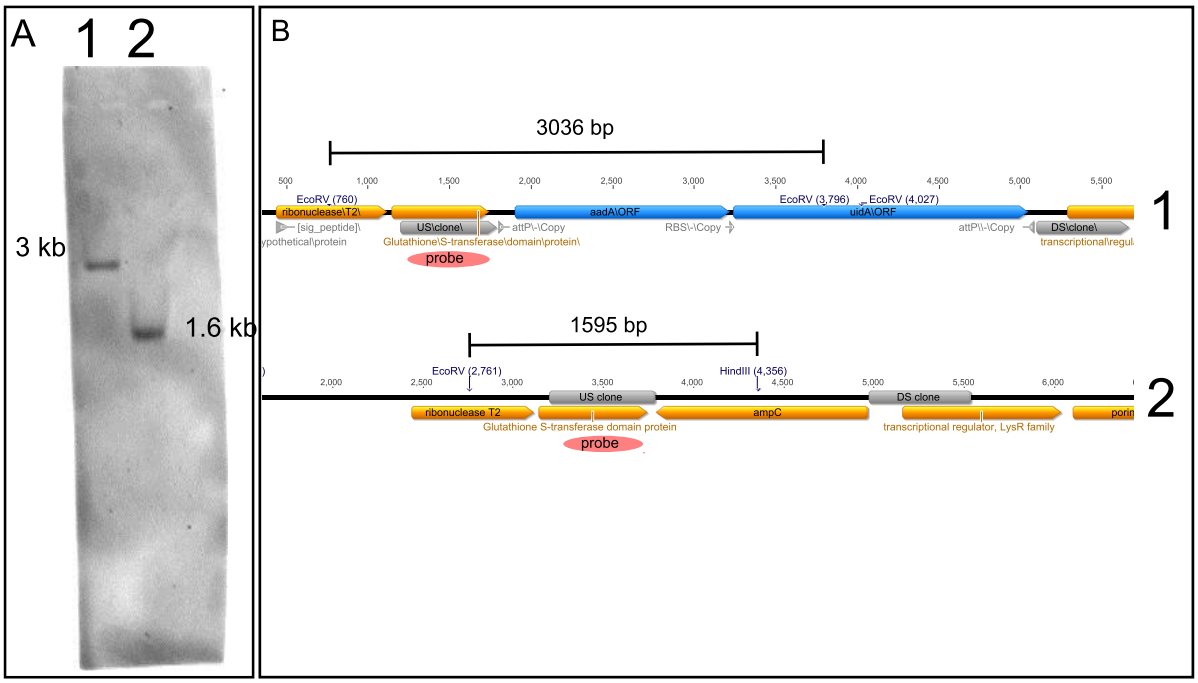

Figure 3.5: *O. anthropi* Ow0013 (LP-strain) Southern blot. An image of the Southern blot is shown in panel A, on the left. Lane A1 is the Ow0013 lane. This lane corresponds with map B1 (expected Ow0013 sequence map) in panel B on the right. The same is true for lane A2 (the parental control: Ow0012) and map B2. The expected migration of the probed band for the LP-strain (map B1) is 2 cm (3 kb, lane A1); for the parental strain (map B2) it is 3 cm (1.6 kb, lane A2). The expected sizes are based on the standard curve made from the 1 kb ladder (Fisher Scientific Inc) on the original agarose gel (Figure A.3).

the other two LP-strains, has the expected phenotypes associated with integration of the landing pad sequence, GusA activity and spectinomycin/streptomycin resistance. At11142 is sensitive to kanamycin and sucrose resistant, indicating the intended loss of the vector backbone sequence. The colonies are blue 2 days after streaking At11142 on LB plates containing 100 μg/ml of the GusA substrate, X-gluc. The parental strain A136 Rif has no GusA activity; its colonies never turn blue on this media.

The integration of the landing pad sequence into the *A. tumefaciens* A136 background by double recombination causes the deletion of the *tetA* gene, which confers cryptic tetracycline resistance in *A. tumefaciens* C58 (Luo and Farrand, 1999). The deletion of *tetA* should cause the elimination of any tetracycline resistance. However, A136 did not display the expected high frequency of tetracycline resistance in my experiments, leaving no room for comparison. The landing pad sequence location is less certain than in other strains. Some more work is needed to be completely confident in the landing pad sequence's location. Direct genomic Sanger sequencing outward from the edge of the landing pad sequence may provide the needed evidence.

3.3.2 IMCE Tetraparental Mating Process

The culmination of the correct assembly of the involved parts (strains and plasmids) is the IMCE tetraparental mating. The three LP-strains, SmUW227 (Figure 3.7), Ow0013 (Figure 3.8), and At11142 (Figure 3.9) were constructed. Two *attB* donor vectors, pJH110 (Figure 3.10) and pJH122 (Figure 3.11), were constructed. The plasmid pJH122 is identical to pJH110 except the gentamicin resistance gene was replaced by the hygromycin resistance gene. The second donor vector was made to accommodate the selection scheme for recombination synthesis (discussed in Chapter 5). The final part that was needed was the ΦC31 integrase expression plasmid pJC2 (Figure 3.12), that was completed with some help from Dr. Jiujun Cheng. All of the constructed plasmids

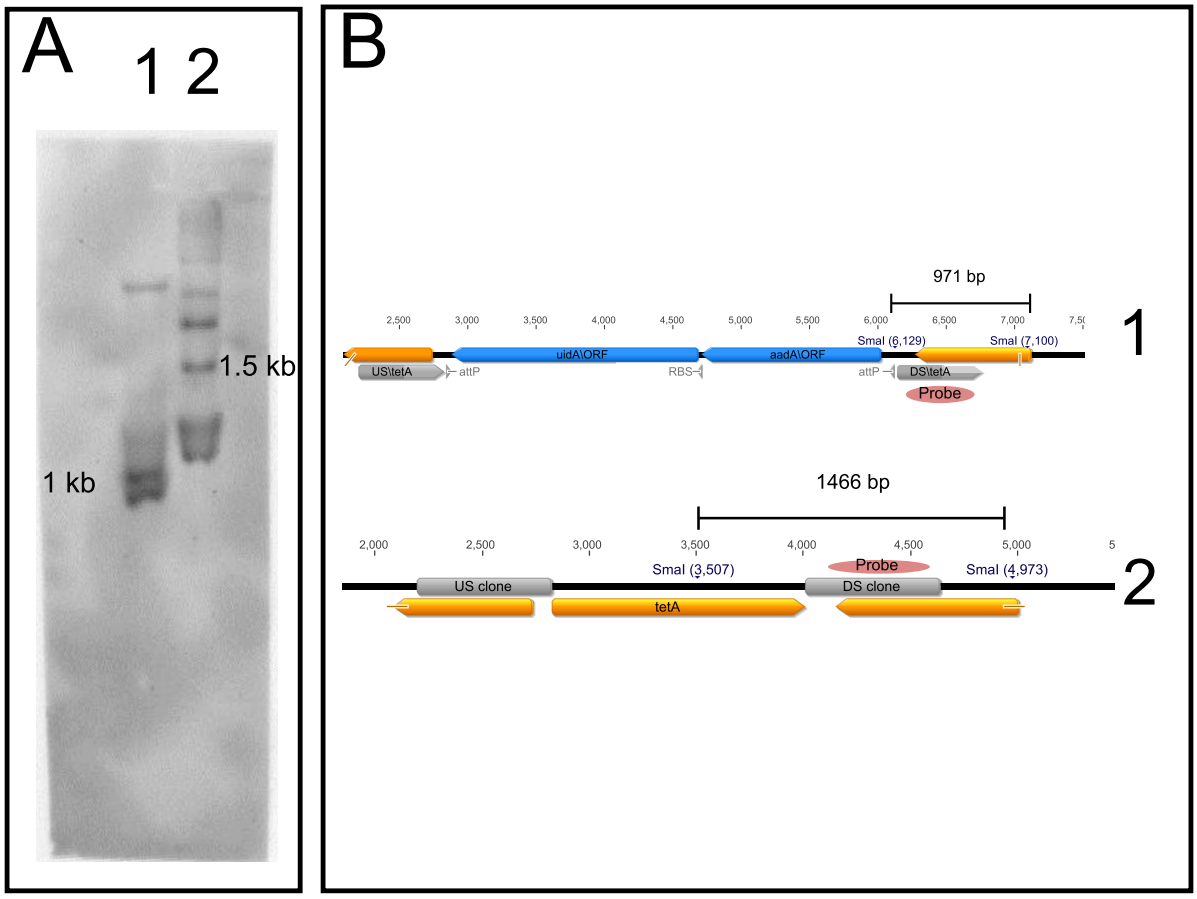

Figure 3.6: At11142 Southern blot. An image of the Southern blot is shown in panel A, on the left. Lane A1 is the At11142 LP-strain lane. This lane corresponds with map B1 (expected At11142 sequence) in panel B, on the right. The same is true for lane A2 (the parental control: A136) and map B2. The expected migration of the probed band for the LP-strain (A1) is 4 cm (1 kb); for the parental strain (A2) it is 3 cm (1.5 kb). The expected sizes are based on the standard curve made from the 1 kb ladder (Fisher Scientific Inc) on the original agarose gel. This blot contains extra bands so it is not fully conclusive, however it should be noted that the expected 1.4 kb band present in the parental control lane, is not expected to be present in the At11142 lane, and this is what is seen on the blot (Figures A.1A.3).

are mobilizable. This made it possible to use the tetraparental mating process.

The establishment of this process formed the foundation for a much of the work discussed later in this thesis. The same general mating procedure was used for variations on the IMCE procedure discussed in Chapter 4. This exact IMCE procedure is the first step in the Recombination Synthesis chromosome engineering process discussed in Chapter 5.

The tetraparental mating process allows for the introduction of the *attB* donor plasmid and the ΦC31 integrase expression plasmid (pJC2) into the recipient LP-strain. The mating process involves four strains. The *attB* donor vector strain, the landing pad recipient, the ΦC31 integrase expression strain (pJC2), and the fourth strain carries the mobilizer plasmid (pRK600). The mobilizer plasmid moves into the donor strains, and once there it allows the expression of the transfer genes required to mobilize the *attB* donor plasmid and pJC2 (Figure 3.13).

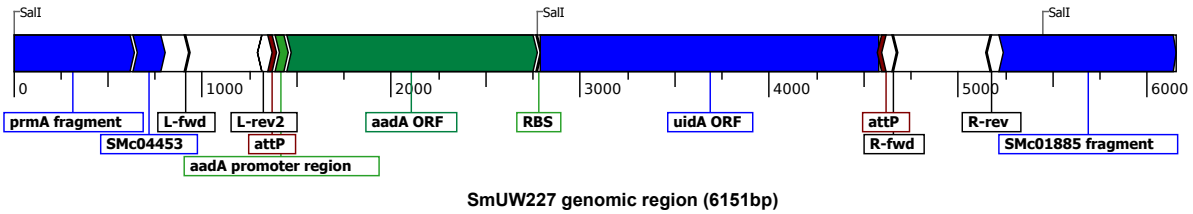

Figure 3.7: *S. meliloti* SmUW227 map: The landing pad locus map of the SmUW227 chromosome is shown.

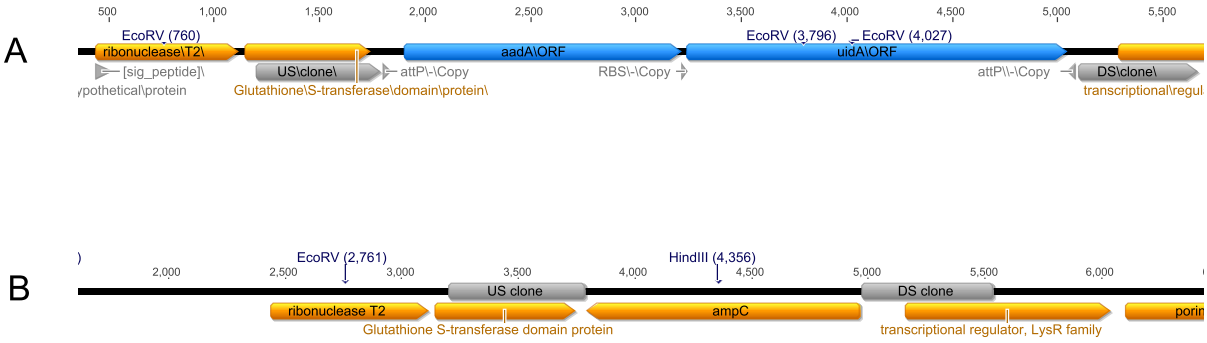

Figure 3.8: *O. anthropi* Ow0013 map: A: The landing pad locus map of the Ow0013 chromosome is shown. B: The parental strain chromosome map at the landing pad locus is shown.

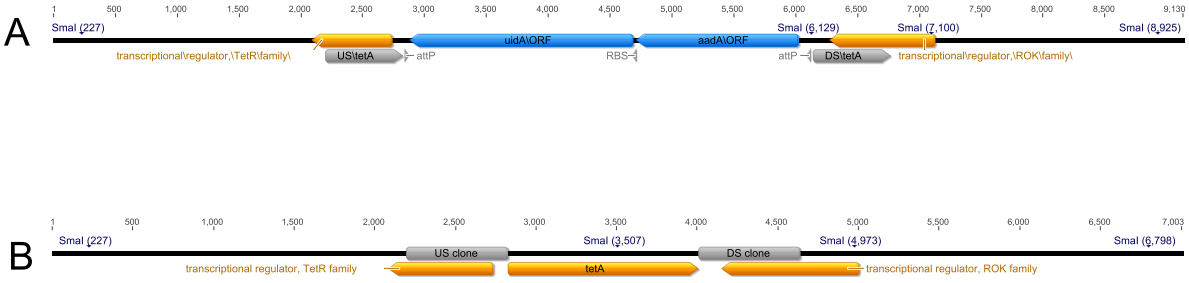

Figure 3.9: *A. tumefaciens* At11142 map: A: The landing pad locus map of the At1142 chromosome is shown. B: The parental strain chromosome map at the landing pad locus is shown.

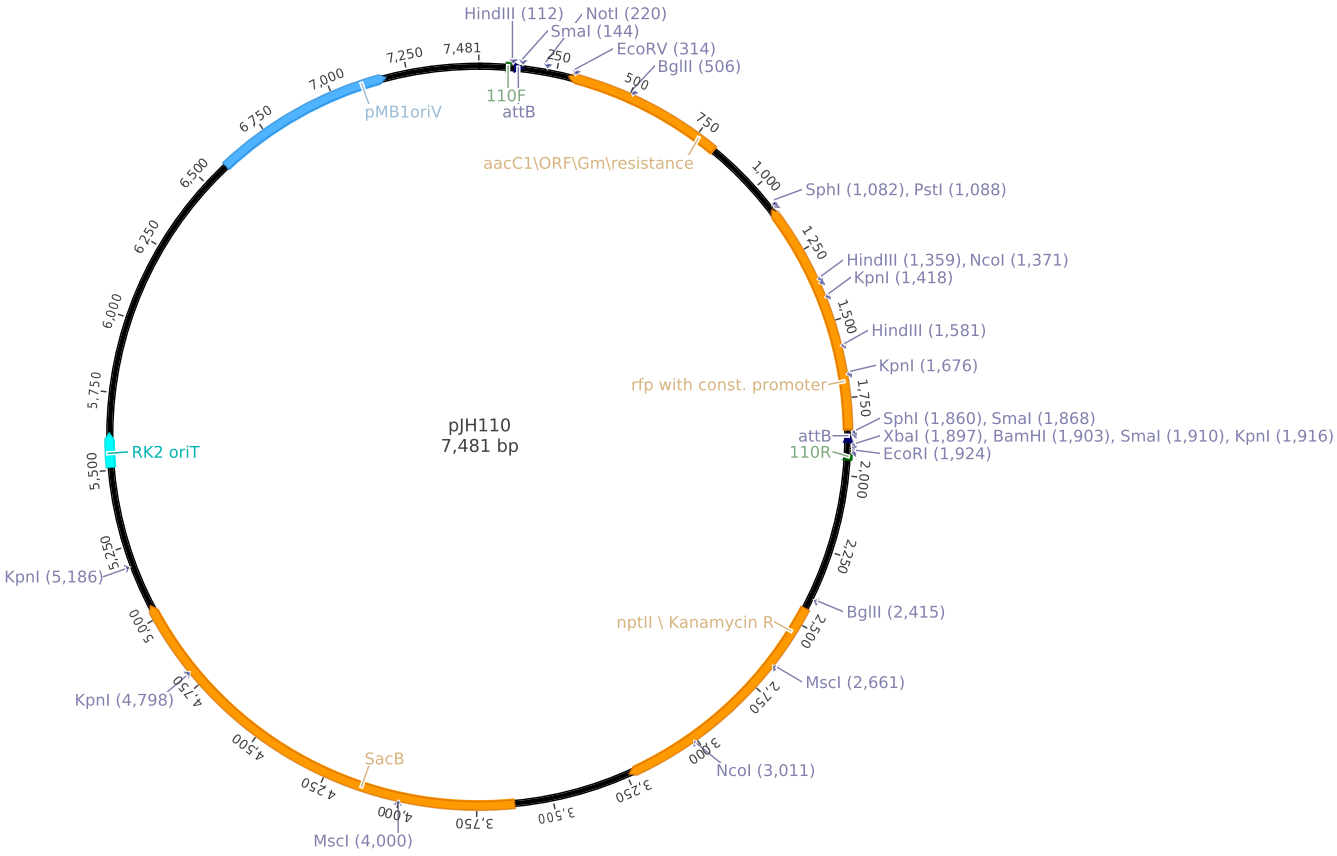

Figure 3.10: The *attB* donor vector pJH110: The pJH110 *attB* donor vector map is shown. It contains the gentamicin resistance gene within the *attB* flanked cassette.

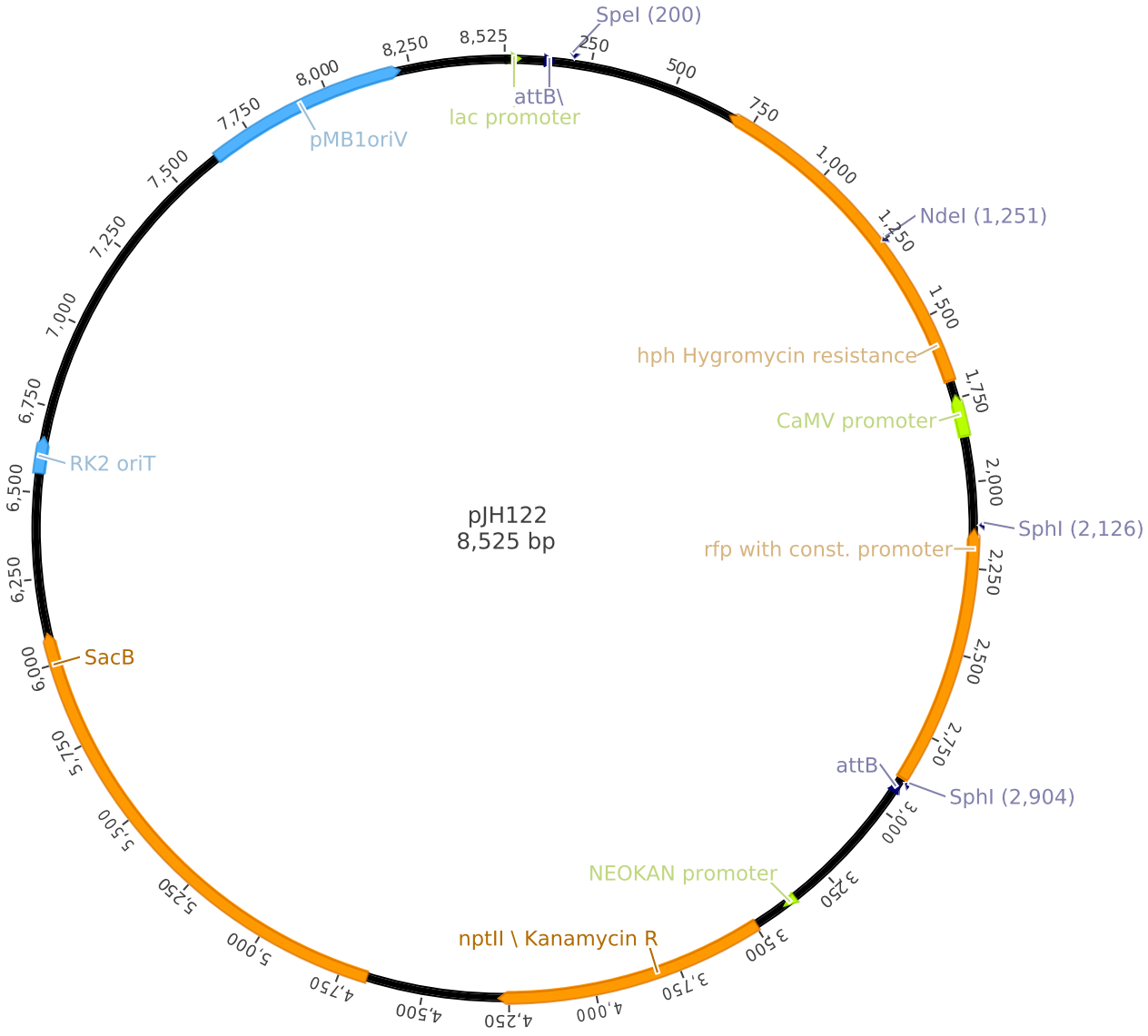

Figure 3.11: The *attB* donor vector pJH122. The pJH122 *attB* donor vector map is shown. It contains the hygromycin resistance gene within the *attB* flanked cassette.

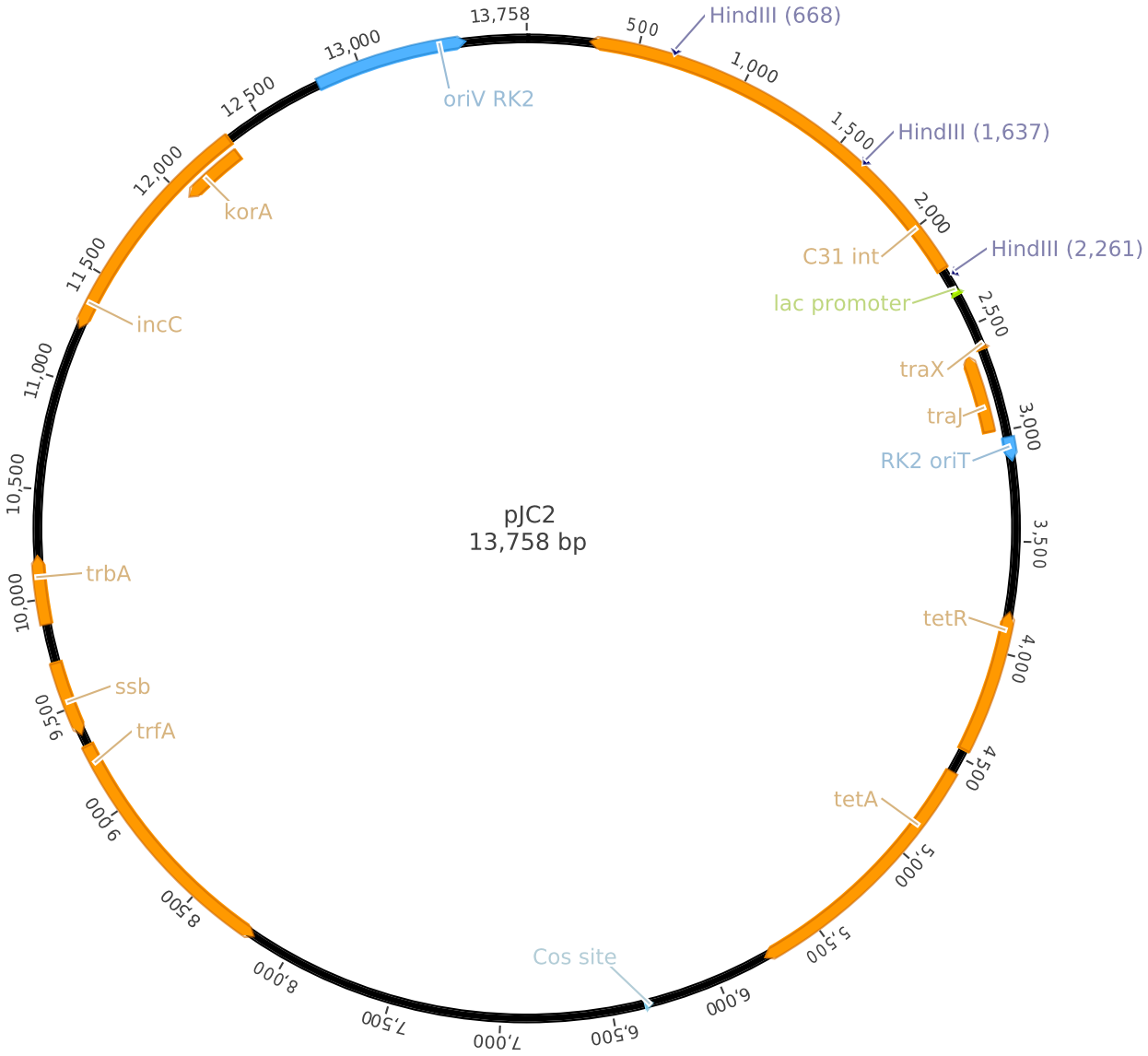

Figure 3.12: The integrase expression plasmid pJC2. The pJC2 integrase expression plasmid map is shown.

3.3.3 IMCE Efficiency

IMCE was carried out with pJH110 as a donor. In this IMCE, a cassette exchange occurs between the JH110 cassette (Gm^R *rfp*) and the landing pad cassette (Sm^R/Sp^R *gusA*). Therefore, the resultant colonies that grow in the initial selection are gentamicin resistant and fluoresce red. These colonies are called transintegrants. A subset of these colonies has undergone true cassette exchange and these colonies will be spectinomycin sensitive and lack GusA activity. In Table 3.2, the efficiencies of the formation of transintegrants is shown. The transintegration efficiency is stated as a proportion of transintegrants to total recipients. Of the transintegrants, those that were found to be spectinomycin sensitive had undergone true cassette exchange. The proportion of these, stated as a proportion relative to transintegrants, is also shown in Table 3.2. The colonies for subsequent routine IMCE experiments were not counted, but there was a great deal of consistency in the results. The plates looked very similar every time. There was some variability in the proportion of transintegrants that turned out to have undergone true cassette exchange. A statistical analysis has not been done because the purpose of knowing these efficiencies is to give a guideline to the user of the IMCE system, not as an indicator of a difference relative to some control experiment. In all IMCE conjugation experiments a no-integrase control was included, where the ΦC31 integrase expression plasmid, pJC2, was left out. There were never any transintegrants obtained from this control mating, indicating that IMCE is indeed dependent on ΦC31 integrase. Photographs of the plates show an example of an IMCE experiment carried out with *S. meliloti* SmUW227 (Figure 3.14).

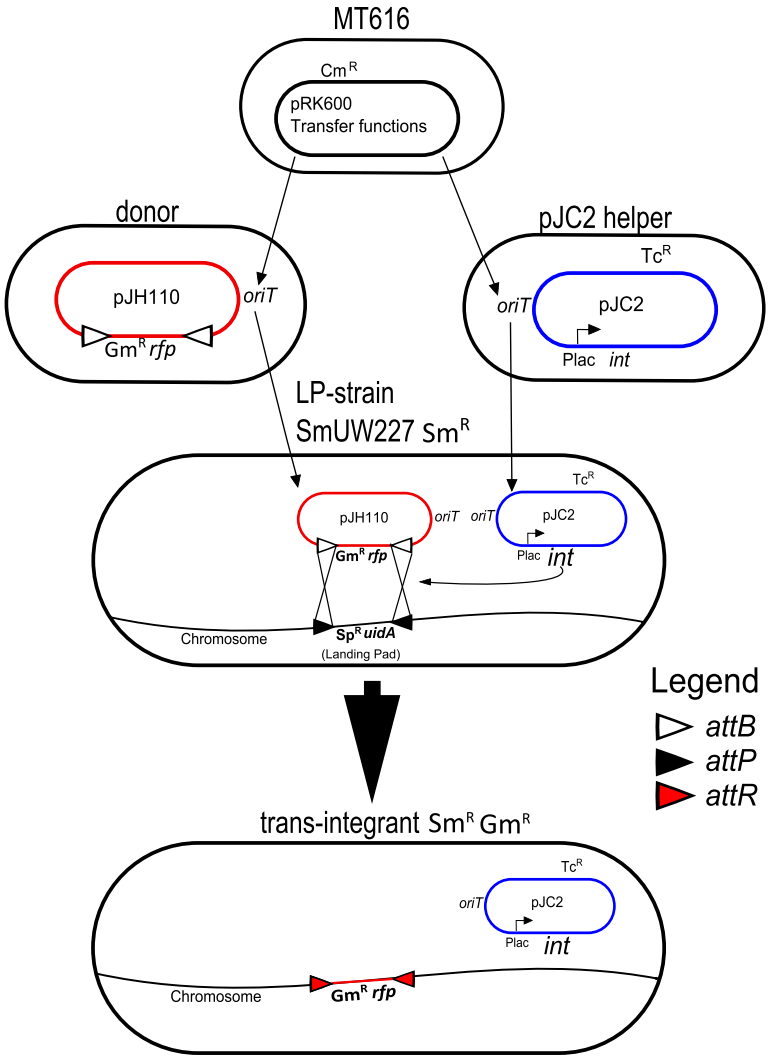

Figure 3.13: IMCE tetraparental mating: an illustration of the conjugation process is shown. The mobilizer plasmid (pRK600) is self-mobilizable and it moves into the strains containing pJC2 (integrase expressing plasmid) and pJH110 (IMCE donor plasmid). Once in these strains, it facilitates the mobilization of these mobilizable plasmids. pJC2 and pJH110 move into the recipient LP-strain (SmUW227) and cassette exchange between the SmUW227 chromosome and pJH110 is facilitated. pJH110 is unable to replicate in any of the LP-strains. Therefore, LP-strains that have undergone cassette exchange may be selected using the antibiotic marker contained in pJH110 between the two *attB* sites. The donor strains are counter-selected, using streptomycin in the case of this figure. Rifampicin may be used in the case of OW0013 and At11142. pJC2 does replicate, but is unstable in these hosts and may be cured by passaging

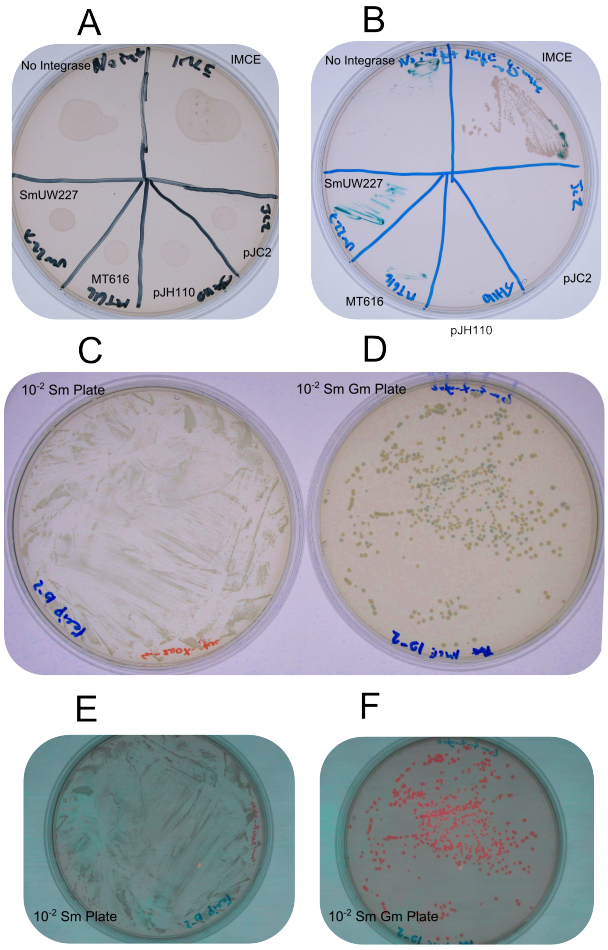

Figure 3.14: A: Mating spot plate on a non-selective TY plate showing dried cell mixtures on agar surface. B: Mating spots streaked on streptomycin-gentamicin-X-gluc plate, from top left clockwise no integrase control, IMCE into S. meliloti, *E. coli* DH5α containing pJC2 control, *E. coli* DH5α containing pJH110 control, *E. coli* MT616 control, and *S. meliloti* SmUW227 control. C: 10^{-2} dilution of mating spot resuspension on TY streptomycin-X-gluc agar. D: 10^{-2} dilution of mating spot resuspension on TY streptomycin-gentamicin-X-gluc agar (blue colonies are single recombinants, white colonies have undergone true cassette exchange). E: 10^{-2} dilution of mating spot resuspension on TY streptomycin-X-gluc agar showing lack of fluorescence. F: 10^{-2} dilution of mating spot resuspension on TY streptomycin-gentamicin-X-gluc agar showing fluorescence. Fluorescent photographs were taken using the method outlined in Heil *et al.* (2011)

Table 3.2: IMCE efficiency (transintegrants as a proportion of total recipients) for each LP-strain, and proportion of transintegrants that are double recombinants

LP-strain	Efficiency (proportion of transintegrants to total recipients)	Proportion true cassette exchange to transintegrants
SmUW227	3×10^{-3}	0.5-0.6
Ow0013	2×10^{-3}	0.5-0.9
At11142	2×10^{-3}	0.5-0.9

3.3.4 IMCE Confirmation by PCR

The ability of IMCE to integrate the *attB* flanked cassette into the landing-pad locus of the LP-strain was verified by colony PCR. Due to the design of the LP-sequence, it is possible for the *attB* sequence to integrate in either orientation with equal probability. For the work discussed in Chapter 5, it was necessary to ensure that the DNA integrated by IMCE was in a particular orientation. For this work, colony PCR was used to ascertain whether a colony was in the arbitrarily chosen desired orientation. The colony PCR used to do this also provided yet more evidence that IMCE was reliably integrating the DNA segments in the landing pad locus.

Colony PCR was used with one primer annealing to the wild-type LP-strain sequence adjacent to the landing pad sequence and the other primer annealing to the antibiotic resistance gene that was proximal to the first primer's annealing site. A frequency of PCR success of around 50% was observed as expected (Figure 3.2.9). The size of the amplified product was correct (600 bp for Gm amplicon, 650 bp for Hy amplicon in SmUW227; slightly larger for other LP-strains). This screen was done for the *S. meliloti* LP-strain SmUW235 with SmUW227 incorporated by transduction (SmUW235 ϕ SmUW227) (Figure 3.15), *O. anthropi* Ow0013 (data not shown), and *A. tumefaciens* At11142 (data not shown). The expected results were observed using SmUW235 ϕ SmUW227, and Ow0013 templates. Colony PCR with with At11142 as a template did not produce any amplified product. The failure of the colony PCR using At11142 derived IMCE strains could have been because of a failure of those particular primers, or the presence of PCR inhibitors.

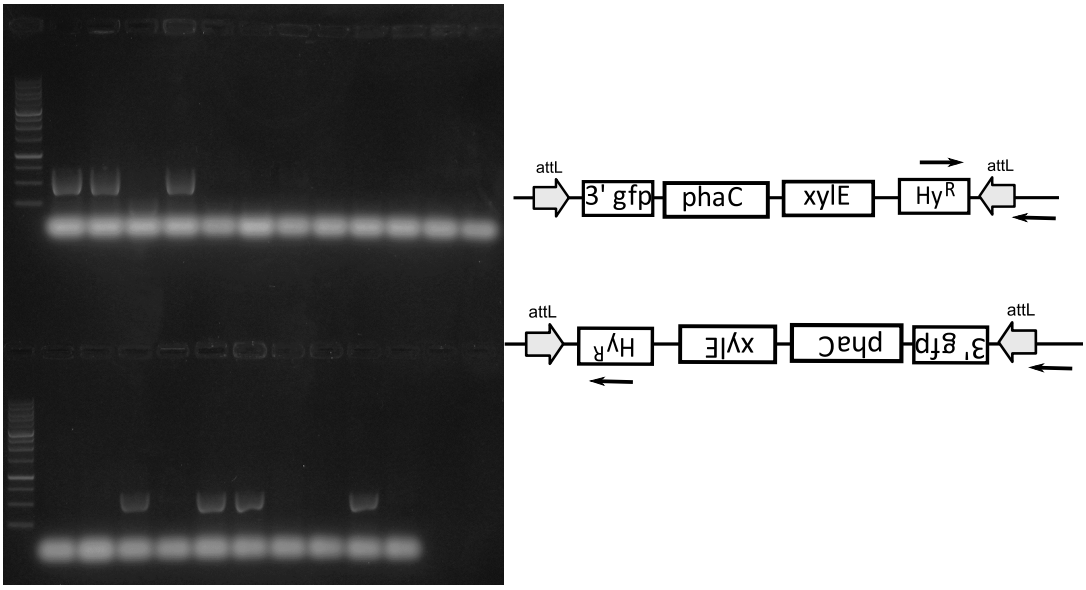

Figure 3.15: In the left panel, a photograph of the gel electrophoresis of screened clones shows 7 out of 20 PCRs produced the proper size DNA molecule (the size marker is the Fisher 1 kb ladder and the second band from the bottom is 500 bp). The right panel shows the locations of the primer annealing in the case of both possible IMCE double recombinant transintegrants.

3.4 Discussion

In an effort to construct a chromosome engineering system that allows the reliable integration of DNA cassettes into the bacterial chromosome, the integrase mediated cassette exchange (IMCE) system was successfully designed and implemented. The IMCE system required an *attP* landing pad strain (LP-strain), a ΦC31 integrase expression plasmid, a mobilizer plasmid, and an *attB* donor vector to accomplish IMCE through a tetraparental mating procedure. All of these, except the mobilizer plasmid, had to be constructed. For this thesis, three LP-strains in the species *Agrobacterium tumefaciens*, *Ochrobactrum anthropi*, and *Sinorhizobium meliloti* were constructed. Experiments were performed to verify each LP-strain. The *O. anthropi* and *S. meliloti* LP-strains were confirmed; the *A. tumefaciens* LP-strain requires additional experiments to demonstrate that the LP-sequence is integrated in the intended location. The IMCE system was used to integrate test constructs containing the gentamicin resistance and RFP genes. These tests were used to establish the frequency of colonies having undergone successful IMCE versus the number of available total recipients. In later work, IMCE was used to integrate PHA half operons. At this stage, colony PCR was used to verify their integration. Colony PCR showed the expected result, that the intended sequence was integrated at the landing pad locus.

Nomenclature

The term transintegrant has been used because the recipient cells that are altered have received the donor plasmid via conjugation, and thus are transconjugants, but have also integrated the plasmid (single recombination) or the donor cassette (double recombination), so they are also integrants. The term transintegrant refers to colonies having undergone either double or single recombinations.

Confirmation of the LP-strains

All three LP-strains have proven very reliable. Each LP-strain produces the expected results when used in the IMCE protocol. Multiple approaches were used to ensure the adherence of the LP-strains the specifications. For the *O. anthropi* and *S. meliloti* LP-strains, the Southern blotting provides very strong confirmation of those LP-strains. For the *A. tumefaciens* strain, the Southern blotting was done; there was a difference between the parental strain and the landing pad strain. The expected bands were present, but some extraneous bands were also present. The probe hybridization conditions would need to be adjusted to get a more conclusive result. However, direct genomic Sanger sequencing from the edge of the landing pad sequence may provide a better alternative. The behavior of all three LP-strains in the IMCE protocol is also strong evidence for their adherence to the design specification. It is highly unlikely that colonies with the particular phenotypes would arise in similar repeatable frequencies by chance. In the case of the *O. anthropi* LP-strain, there is good phenotypic evidence that the landing pad sequence is integrated in the intended locus. In that case an antibiotic resistance gene was deleted. The physical evidence of successful IMCE given by the colony PCR results is also in agreement with the Southern blotting. It is clear that the *S. meliloti* and *O. anthropi* LP-strains have been validated very stringently.

Design Considerations

There were a few design considerations made before the construction of the IMCE system. Various markers used in IMCE and markers intended for use in downstream applications were considered to avoid conflicts. GusA (also referred to as UidA) was chosen as the blue white screen marker because it is not present in the any of the genomes of the three species backgrounds used for the LP-strains described in this

chapter. It was later noticed that the production of a blue colour on X-gluc containing media is delayed in the *S. meliloti* background. The blue colour first appears in the centre of the colonies when they are 5 days old. In the other backgrounds the blue colour develops quite quickly. This may be because *S. meliloti* may lack a transporter in its cell envelope capable of importing X-gluc. The other major design consideration was choosing the topology of the *att* sites. The *attP* sites flanking the landing pad sequence are in opposite orientations, facing inwards. In other words, the *att* sites are in an antiparallel configuration. This configuration of *att* sites was chosen so the backbone of the donor vector would not be integrated. The *attP* and *attB* sites must synapse in parallel with each other for recombination to occur (Gupta *et al.*, 2007). If the landing pad sequence and the donor vectors had been designed with parallel *att* sites, the *attB* and *attP* sites could align in parallel no matter which particular *attB* site aligned with which particular *attP* site. Therefore there would be no topological distinction between the backbone portion of the vector and the cassette intended for integration, relative to the *attB* sites. With an antiparallel *att* site design, each *attP* site must synapse with its intended partner *attB* site in order to satisfy the conditions for successful recombination. Therefore, the antiparallel design ensures only the intended cassette is transferred in the IMCE process. The disadvantage of this design is that there is no topological difference, as far as the recombination is concerned, between both orientations of the donor cassette. That is why IMCE transintegrant clones are found in both orientations. A design using parallel *att* sites would require a selection against integration of the backbone, or extra screening for the presence of the intended cassette. The parallel design would, however, remove uncertainty about the orientation of the integrated cassette.

IMCE Efficiency

The IMCE system provides a reliable and rapid method to integrate *attB* flanked DNA into the chromosome of the LP-strains described above. The protocol produces hundreds to thousands of colonies in one conjugation step, without the need for further selection steps. In fact, the efficiency of IMCE is such that a clone could easily be isolated by the GusA blue-white screen alone, thereby producing a markerless clone. The same blue-white screen allows for the differentiation of double recombinants (having undergone true cassette exchange) and colonies that are merely transintegrants. The presence of the streptomycin/spectinomycin resistance gene also allows a second marker confirmation of the double recombination event.

The high efficiency of IMCE is quite impressive when compared with the conventional means of integrating DNA into bacterial chromosomes. Homologous recombination, accomplished by introducing a plasmid containing ≈ 500 bp of host chromosome homology into a RecA positive host by triparental mating, gives a much lower frequency of ten to one hundred times less. The observed transconjugant frequency does vary when recombination is involved. It varies because of features and length of the homologous DNA. These features include the proximity to χ-sequences (Otero *et al.*, 1995), and mutation frequency fluctuation. Mutation frequency fluctuation is due to the fact that in some cases a recombination event will happen in the first few generations of growth, during transfer, or from subsequent transfers in later generations (Luria and Delbrück, 1943; Lea and Coulson, 1949). Though it is difficult to make a rigorous comparison of absolute transintegration frequencies between conventional homologous recombination and IMCE, practically speaking, from my experience IMCE will yield approximately five hundred times the double recombinants than homologous recombination will yield single recombinants (data not shown). The frequency of transintegrants that are double recombinants in the homologous recom-

bination case is usually around the square of the transintegrant frequency. This means that a two step selection is used to obtain double recombinants(Schafer *et al.*, 1994). The fact that the proportion of double recombinants among the transintegrants in IMCE is so high suggests that the bottleneck is the rate of DNA transfer. If the bottleneck were the recombination event, the double recombinant frequency would be close to the square of the overall transintegrant frequency ($\approx 9 \times 10^{-6}$), but the observed frequency ($\approx 2 \times 10^{-3}$) is three orders of magnitude higher than that. Given this inference, IMCE efficiency could be further improved if the integrase expression plasmid (pJC2) were maintained in the LP-strain, making the tetraparental mating into a triparental mating. This would increase the number of cells having acquired all the plasmids required for IMCE, because every recipient cell would contain pJC2.

The Advantages of Phage C31 Integrase

An important advantage of developing a system based on ΦC31 integrase is the simplicity of implementing the ΦC31 integrase. Only one enzyme (the integrase) is required for the initial recombination. There are no factors, like integration host factor, needed by the tyrosine recombinase from λ phage (Mumm *et al.*, 2006). A system like the Tn7 site-specific integration system (Choi *et al.*, 2005, 2006b), where many bacteria naturally have the required *att* sites, provides a quick route to chromosome engineering in a particular organism, but can be complicated by species variations such as *att* site copy number. The number and location of the *att*Tn7 sites differs between genomes (Choi and Schweizer, 2006b; Choi *et al.*, 2006a; Choi and Schweizer, 2006a). In the case of the Tn7 system, the *att* site is present only within the *γ-proteobacteria*. The limitations of the Tn7 system are apparent, considering IMCE has been demonstrated here in three *α-proteobacteria*, in which Tn7 integration is not possible. The ΦC31 *att* sites do not occur naturally in any bacterial genomes so far investigated (*Synechocys-*

tis PCC 6803, *Methylosinus trichosporium OB3b*, and *Bacillus subtilus*, in addition to the discussed LP-strains). This is not a major disadvantage, considering the limited nature of even the most broad host range natural systems. Instead this can afford design flexibility when creating the landing pad strain.

Overcoming Limitations of IMCE

The size of the construct to be integrated by IMCE is limited by the size of DNA that can be successfully cloned into the donor vector via ligation. Donor vectors containing Gateway destinations (Katzen, 2007) can be used, and are especially useful for large insert fosmid/cosmid libraries. Gateway donor vectors have been constructed as part of this work and are described in subsequent chapters of this thesis. The antibiotic resistance markers in the donor vectors may also be used to select for the assembly of overlapping DNA fragments, post IMCE, by crossing two different trans-integrants via transduction, conjugation, or genomic electroporation. In this thesis, the assembly of already integrated segments post-IMCE is called recombination synthesis (RS) and its implementation is discussed in subsequent chapters.

Application of IMCE to Functional Metagenomics

In functional metagenomics, screens are based on the functions of the gene expressed from the library clone (Simon and Daniel, 2011). The identification of clones of a particular function is limited by the ability of the screening host to express the genes on the clone, as well as a suitable selectable or screenable phenotype. Given a metagenomic library may contain DNA from a diverse community of bacteria, *Escherichia coli* may not be able to express many of the genes present in the library. Therefore, alternate screening hosts from different phyla are useful in functional metagenomics (Craig *et al.*, 2010). A metagenomic library made in an IMCE compatible vector would be

compatible with any host where a landing pad strain has been, or can be constructed.

Chapter 4

Integrase Mediated Cassette Exchange Tools and Reverse Integrase Mediated Cassette Exchange

4.1 Introduction

To use the IMCE system, the desired DNA must first be cloned into the *attB* donor vector. If an antibiotic resistance marker is included in the sequence between the *attB* sites this simplifies the IMCE protocol. IMCE is efficient enough that the blue-white screen may be used to isolate double recombinants for the largest DNA constructs yet tried (still less than 10 kb). However, it is desirable to have a selection that is not reliant on the contents of the *attB* donor cassette. It is also desirable to have *attB* donor vectors that allow for Gateway cloning (Katzen, 2007), which may streamline the subcloning of inserts into the *attB* donor cassette. Gateway cloning is a plasmid-based cassette exchange system, very similar to IMCE. It uses the λ phage integrase and excisionase. ΦC31 *attB* donor vectors containing Gateway destinations (equivalent to a landing pad sequence) were constructed.

In functional metagenomics, a factor that limits the availability of screening hosts is the ability of plasmids to replicate in the host. At the time of library construction, a broad host range vector may be used. However, there will be potential hosts that will not support its replication. Conversely, if IMCE is used for functional screens, plasmid replication is not a factor. A LP-strain in the desired host may be created after library construction has been completed. For use of the IMCE system involving functional metagenomics, a fosmid vector was created.

The LP-strains were constructed using a two step homologous recombination protocol. Double recombinants were selected in the second step by plating on sucrose, thereby selecting for the loss of *sacB* (Gay *et al.*, 1985). Because *sacB* is on the backbone of the vector, the vector sequences are lost, as was the goal of the selection. The *sacB* gene is a useful counterselectable marker, and including it within the landing pad sequence is desirable. However, including *sacB* on the landing pad sequence precludes its use on the vector backbone. The strategy used for the initial construction of the LP-strains depends on the *sacB* gene contained on the vector backbone. Therefore, a strategy using reverse IMCE to remodel LP-strains after their initial construction was used to make a SacB LP-strain. To allow for reverse IMCE, the directionality of ΦC31 recombination must be reversed.

There are two aspects of ΦC31 integrase recombination that one may think of when considering the directionality of recombination. There is the directionality of the synapses of the two *att* sites during recombination. If there is no specificity for synapse direction, a mixture of excision and inversion of the DNA between an *attP* and an *attB* will result from recombination. The other directional consideration is the direction of the recombination reaction itself: from *attP-attB* to *attL-attR* or the other way (Ghosh *et al.*, 2008; Khaleel *et al.*, 2011).

The ΦC31 integrase recombines *attP* and *attB* sequences when they synapse in

a parallel orientation. This parallel specificity is governed by the core dinucleotide found in both *att* sites (Gupta *et al.*, 2007). The core dinucleotide of *attP* must base pair with the dinucleotide in *attB*. In the case of the ΦC31 *att* sites, the core dinucleotide is the non-palindromic TT, so base-pairing may only be accomplished when the *att* sites synapse in parallel. If the dinucleotide in the serine recombinases is mutated so that it is a palindrome (eg. TA), then the selectivity for parallel synapses is lost (Ghosh *et al.*, 2003). An even mixture of excisions and inversions is observed. In the context of chromosome engineering, this fact can be used to create different classes of *att* sites to increase the specificity of recombination reactions (Colloms *et al.*, 2013).

The other directional aspect of ΦC31 recombination is the reaction direction. This is usually what is referred to by the term ''directionality" in the literature. Because the sequences flanking the dinucleotide are inverted imperfect repeats, a recombination of *attP* and *attB* sites gives rise to two distinct composite sites: *attR* and *attL*. The ΦC31 integrase can distinguish between *attP-attB* and *attL-attR* (Rowley *et al.*, 2008). The integrase will only catalyze recombination in the *attP-attB* to *attL-attR* direction. A protein called the recombination directionality factor (RDF) is responsible for reversing this direction. It does so by binding to the integrase protein and changing the specificity by which the integrase recognizes the *att* sites (Ghosh *et al.*, 2008; Khaleel *et al.*, 2011). Expression of the RDF is how the integrated prophage would initiate lytic growth. In the context of chromosome engineering, the use of the reverse reaction can make the existing IMCE system more powerful. The reverse reaction can be used to do reverse IMCE after IMCE has been completed. This reverse IMCE (RIMCE) was used to create the SacB LP-strain.

4.2 Materials and Methods

4.2.1 Construction of pJH145: Mobilizable *attB* Donor Fosmid

The primers attbfos-F and attbfos-R (Table 2.2) were used to amplify the pCC1fos vector. The primers annealed adjacent to the Eco72I cloning site of the vector, causing the deletion of the original cloning site. ΦC31 *attB* sites were added flanking the vector, and on each end of the PCR product, 1/2 of an Eco72I site was added. The PCR product was amplified with a proofreading polymerase (KOD Xtreme); it had blunt ends. The PCR product was ligated, forming a complete Eco72I site flanked by *attB* sites. The ligated plasmid was transformed into *Escherichia coli* EPI300 (Table 2.1). The intermediate vector was called pCC1fosattB. The RK2 origin of transfer (*oriT*) was amplified from pJC8 (Table 2.1) using the primers KL12-F and KL13-R (designed by Kathy Lam) (Table 2.2). The PCR product was cloned into pCC1fosattB as a HindIII fragment to create pCC1fosattBoriT. The Eco72I fragment from pJC8 (Table 2.1) was subcloned into pCC1fosattBoriT to create pJH145.

4.2.2 Construction of pJH121: Gentamicin *attB* Donor Vector with Gateway

pMK2016 (Table 2.1) was cut with SphI. The 1.9 kb fragment, the Gateway destination sequence, was extracted from an agarose gel. It was subcloned into pJH110 as an SphI fragment, replacing *rfp*. The orientation of the insert was mapped with EcoRI. The clone with the orientation placing *attR2* proximal to *aac1* was saved. The plasmid was transformed, isolated, and stored in the *ccdB* resistant *E. coli* DB3.1 because the Gateway destination contains the *ccdB* gene.

4.2.3 Construction of pJH123 and pJH124: Hygromycin *attB* Donor Vectors with Gateway

These vectors were constructed in the same way as pJH121 except the 1.9 kb fragment was cloned into pJH122 as an SphI fragment. An EcoRI digest was completed and the orientation of the insert was mapped. Clones of both orientations were saved. pJH123 was determined to have the insert in the same orientation as pJH121 (*attR2* proximal to *hph*), and pJH124 was determined to contain the insert in the opposite orientation. The plasmid was isolated and stored in *E. coli* DB3.1. PJH123 restriction with EcoRI produced the following expected sizes: 7013 bp, 1667 bp, 669 bp, and 407 bp. The same restriction of pJH124 produced the expected sizes: 7013 bp, 1476 bp, 860 bp, and 407 bp.

4.2.4 Construction of pJH141

The primers RDF-fix-F and RDF-fix-R (Table 2.2) were used to amplify the RDF gene (*gp3*) from the template, pJET-gp3 (Table 2.1). BamHI sites and a ribosome binding site were added by the primers. The PCR product was cloned into pJC2 as a BamHI fragment to create a polycistron with *int*. The resultant clones were screened for the proper orientation (same direction of transcription as *int*) by restriction mapping with HindIII. The HindIII restriction of the expected clone produced the band sizes: 900 bp, 650 bp, 300 bp, and 750 bp (excluding much larger backbone), but there was an extra 450 bp band. Sequencing revealed that the clone was a dimer in the direct orientation. The restriction map was consistent with that sequence. The clone was saved and named pJH141, and every subsequent attempt to repeat this construction resulted in dimeric clones.

4.2.5 Construction of pJH143

The *sacB* gene, including its promoter, was amplified from the template pK19mobSacB (Table 2.1) using the primers attLsac-F and attlsac-R (Table 2.2). The primers added ΦC31 *attL* sites flanking *sacB*. The primers also added a HindIII site on the 5′ end and a EcoRI site on the 3′ end. A PmeI site was added on the inside of the downstream *attL* site. The PCR product was cloned into pK19mob (Table 2.1) as an HindIII-EcoRI fragment, producing the intermediate plasmid: pK19attlsac. The LP-sequence excluding the *attB*sites was amplified from pJH104 (Table 2.1) using the primers gusspec-F and gusspec-R (Table 2.2). These primers added PmeI sites on both ends. This PCR product was cloned into pK19attlsac as a PmeI fragment to create pJH143.

4.2.6 *In vivo* Gateway Recombination

The *in vivo* Gateway recombination protocol was described previously (House *et al.*, 2004). The general conjugation procedure was similar to that of a triparental mating (section2.5), except there are 5 strains. The following strains (Table 2.1) were grown with the appropriate antibiotic for plasmid maintenance (section 2.3): pJH121 (Gateway destination) with kanamycin in *E. coli* DB3.1, pJC8 lac clone (Gateway entry clone) with tetracycline, *Escherichia coli* DH5*α* Rif (final recipient), *E. coli* MT616 (mobilizer), pXINT129 in *E. coli* DH5*α* with ampicillin (recombinase expression). The cultures were washed with sterile saline. The cultures were mixed and spotted on plain LB agar with 1 mM IPTG to induce recombinase expression. The spots were dried down and incubated for 16 hours at 37°C. The transconjugants having undergone Gateway recombination were selected on LB agar with rifampicin at 50 μg/ml and gentamicin at 10 μg/ml.

4.2.7 RIMCE Tetraparental Mating

The RIMCE protocol was nearly identical to the IMCE protocol (section 3.2.8) with a few key differences. The pJC2 plasmid was replaced by pJH141, which contains the ΦC31 RDF gene (*gp3*).The donor *attB* donor plasmid was replaced by the *attL* donor plasmid pJH143. The LP-strain was replaced with the reverse LP-strain (rLP strain). The rLP strain was 246, which contains an integrated cassette from pJH110. pJH143 contains the original landing pad sequence with *sacB* added, flanked by *attL* sites. The final difference was that all *E. coli* strains were in a β-DH10B (Rowe-Magnus, 2009) background, which requires diaminopimelic acid (DAP) for growth. All *E. coli* strains were grown in LB broth supplemented with DAP at 60 μg/ml, and supplemented with the appropriate antibiotics for plasmid maintenance. The selection against donors was the omission of DAP from the final selection plates. The selection for reverse-transintegrants was TY media containing spectinomycin at 100 μg/ml.

For integrated clone recovery onto a plasmid, a pentaparental mating was used. A final recipient strain, *E. coli* DH5α Rif, was added. The selection for a plasmid containing the pJH110 cassette was LB media containing rifampicin at 50 μg/ml and gentamicin at 10 μg/ml, or 10% sucrose to substitute for the gentamicin, or both sucrose and gentamicin with the stated concentrations.

4.2.8 IMCE using SacB LP-strain

This protocol was identical to the IMCE protocol (section 3.2.8), except the LP-strain was the SacB LP-strain derived from the RIMCE, using pJH143 as a donor. The donor vector was pJH110. The selection for transintegrants was TY containing 10% sucrose and streptomycin at 200 μg/ml. The 10^{-3} and 10^{-4} dilutions of the resuspended mating spot and no-integrase control spot were plated. The mating spot and all of the controls (individual spots, and no-*int*,*gp3* control) were also plated on TY gentamicin

at 30 μg/ml, streptomycin at 200 μg/ml. This was done as a positive control for the integrity of the SacB LP-strain.

4.3 Results

4.3.1 RIMCE

The reverse IMCE (RIMCE) process allows for the replacement of the landing pad sequence. It was developed to increase the flexibility of the IMCE system and to allow for the creation of the SacB LP-strain. The parts required for RIMCE were constructed. They were a donor vector with an *attL* flanked cassette (Figure 4.1), a reverse LP-strain (Figure 4.2), and a plasmid to express both ΦC31 integrase and RDF (Figure 4.3). Fortunately, the reverse LP-strain was made as a consequence of doing IMCE using pJH110, the resultant strain (SmUW246) was a suitable reverse LP-strain.

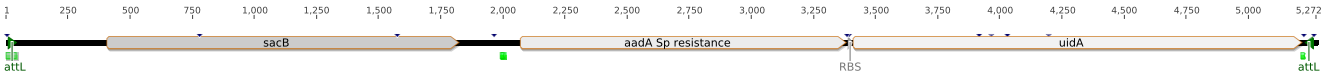

Figure 4.1: The SacB landing pad. The SacB landing pad sequence map is shown. It exists on the plasmid pJH143, which has a pK19mob (Schafer *et al.*, 1994) vector backbone. The *attL* flanked cassette contains *sacB*, in addition to the *aadA*, and *uidA* (also known as *gusA*) genes present in the original landing pad sequence.

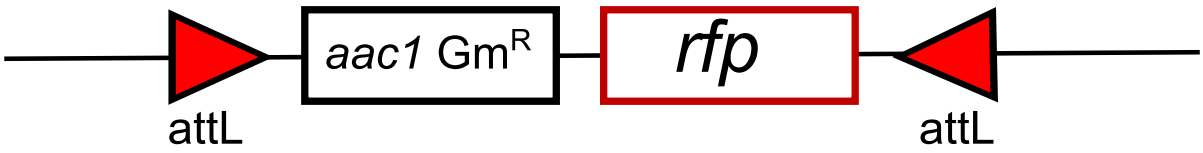

Figure 4.2: The reverse landing pad. The reverse landing pad map is shown. It is present at the *S. meliloti* landing pad locus between SMc04453 and SMc01885.

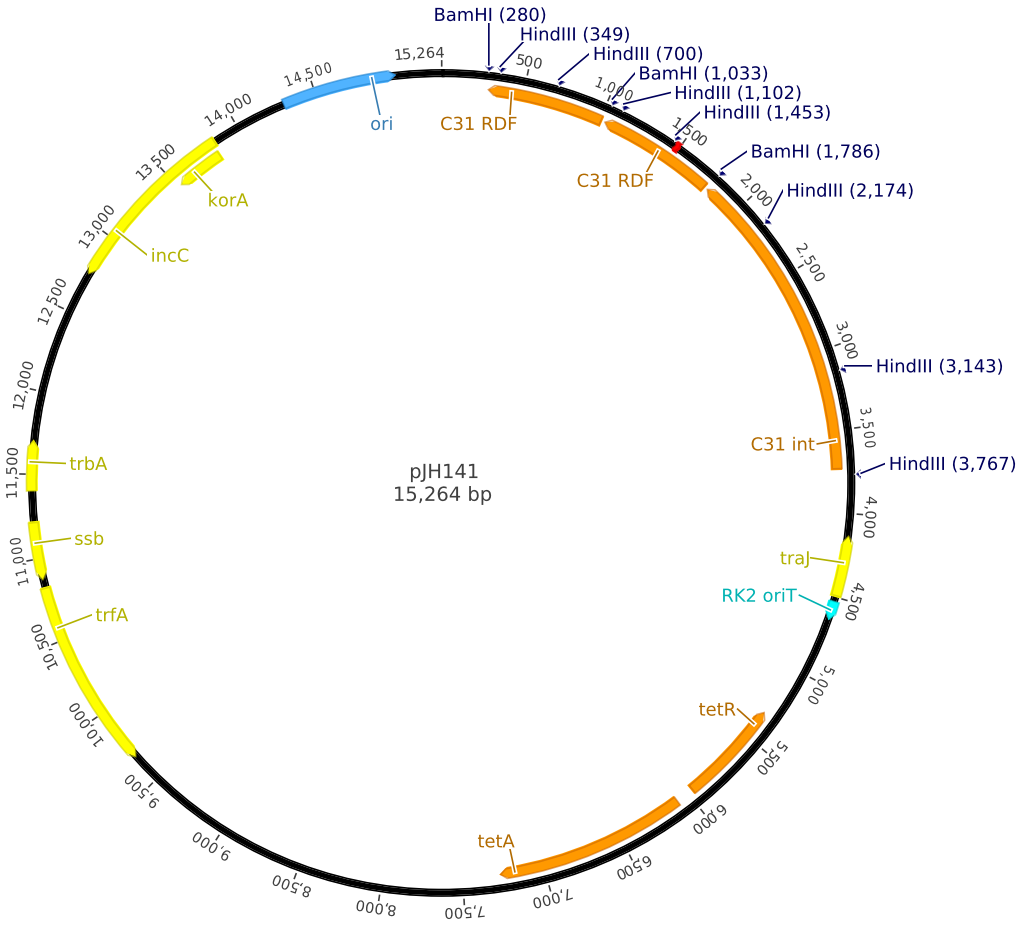

Figure 4.3: RIMCE integrase expression plasmid. The map of pJH141 is shown. It contains the ΦC31 *int* and *gp3* (RDF gene) genes. Transcription of the genes is driven by the lac promoter included in the multiple cloning site of the original vector: pRK7813.

The RIMCE process is a tetraparental mating process that is very similar to IMCE. The key differences are the *att* sites and the inclusion of the *gp3* gene on the integrase expression plasmid (Figure 4.4). Also, the background *Escherichia coli* strain was changed to βDH10B (Rowe-Magnus, 2009). In the tetraparental mating spot, the self-mobilizable pRK600 was transferred into both βDH10B containing pJH143 and βDH10B containing pJH141, where it mobilized the pJH143 and pJH141 plasmids. The plasmids were transferred into the rLP-strain (SmUW246) where the RIMCE was catalyzed by the integrase and RDF. *E. coli* βDH10B, a DAP auxotroph, was used, because of the potential selection conflict with streptomycin. Rm1021 (SmUW246 is one) derived strains are streptomycin resistant, and that is the usual counterselection against the donor. In this case, the streptomycin counterselection does not work, because the donor plasmid replicates in *E. coli* and confers streptomycin resistance. βDH10B requires DAP (a component of peptidoglycan) to grow. Because production of DAP is limited to bacteria, DAP is not present in common complex media based on casein digests and yeast extract. This makes the use of βDH10B (Rowe-Magnus, 2009) a convenient way to select against donors in the RIMCE tetraparental mating, and the use of streptomycin is avoided.

RIMCE was demonstrated in the *S. meliloti* background using SmUW246 reverse LP-strain as a recipient. SmUW246 is the strain that resulted from IMCE into SmUW227 using pJH110 as a donor. At 3×10^{-5} as a proportion of recipients, the frequency of transintegrants was much lower than what was obtained with IMCE. Most of the transintegrants were single recombinants. 1 in 20 transintegrants were double recombinants. The double recombinants were easy to identify, because they had lost the RFP fluorescence of SmUW246. The result of RIMCE was the creation of the SacB LP-strain: SmUW574.

To recover integrated clones an RIMCE conjugation with the addition of a final *Escherichia coli* DH5α Rif recipient was added. The RIMCE cassette exchange between pJH143 and SmUW246 reconstructs a pJH110 like plasmid (that lacks *sacB* on its backbone). This plasmid was to be selected using LB media containing rifampicin and sucrose, or rifampicin and gentamicin. The recovery of the integrated pJH110 cassette was attempted numerous times, but no recovered clones were obtained.

Although Figure 4.4 shows pJH141 as having a copy of the *gp3*, pJH141 actually contains a dimer of *gp3*. All attempts to isolate a clone with a monomeric *gp3* resulted in isolating only dimers. No monomeric clones in the proper orientation were found after screening many colonies, and the dimeric construct is in fact very stable.

4.3.2 SacB IMCE

Th SacB IMCE process is nearly the same as IMCE, except that the LP-strain differs in having *sacB* in the LP-sequence. The selection was tested by doing a tetraparental mating using pJH110 as the *attB* donor, MT616 as the mobilizer strain, pJC2 as the integrase expressing donor, and SmUW574 as the LP-strain. A gentamicin selection was done with this mating as well, and the gentamicin selection gave identical results to IMCE using SmUW227. The SacB selection allowed for more background growth than the gentamicin selection did, but the SacB selection was sufficient to achieve the desired results (Figure 4.5). Transintegrants were easily isolated. I observed that placing a high density of recipient cells on the sucrose selection plate tended to result in an excessive background growth level. In Figure 4.5, the "headstreak" of SmUW574 (SacB LP-strain) on sucrose media shows growth at the start and then clears as the cell density drops near the end of the streak. The efficiency of SacB IMCE was similar to that of IMCE. Single recombinants were counter-selected because single recombinants still contain *sacB*. Of the colonies that arose on the sucrose plate, approximately

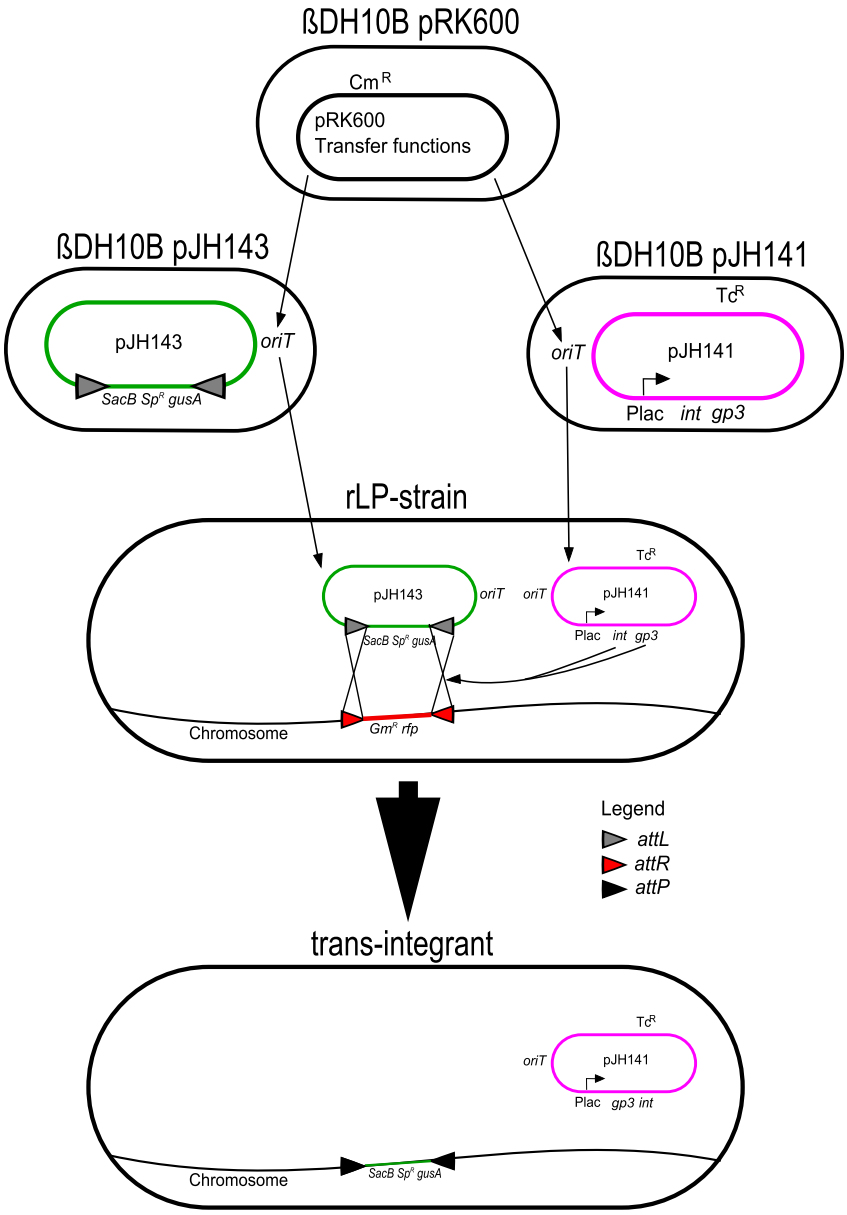

Figure 4.4: RIMCE tetraparental mating: an illustration of the conjugation process is shown. The mobilizer plasmid (pRK600) is self-mobilizable and it moves into the strains containing pJH141 (integrase RDF expressing plasmid) and pJH143 (RIMCE donor plasmid). Once in these strains, it facilitates the mobilization of these mobilizable plasmids. pJC2 and pJH110 move into the recipient rLP-strain (SmUW246) and cassette exchange between the SmUW246 chromosome and pJH143 is facilitated. pJH143 is unable to replicate in SmUW246. Therefore, LP-strains that have undergone cassette exchange may be selected using the antibiotic marker contained in pJH143 between the two *attL* sites. The donor strains are counter-selected by omission of DAP in the media. The plasmid pJH141 may be cured by passaging.

half were still spectinomycin resistant, which would indicate that they had a mutation in *sacB*. The other half were sucrose resistant and spectinomycin sensitive, indicating that they were double recombinants.

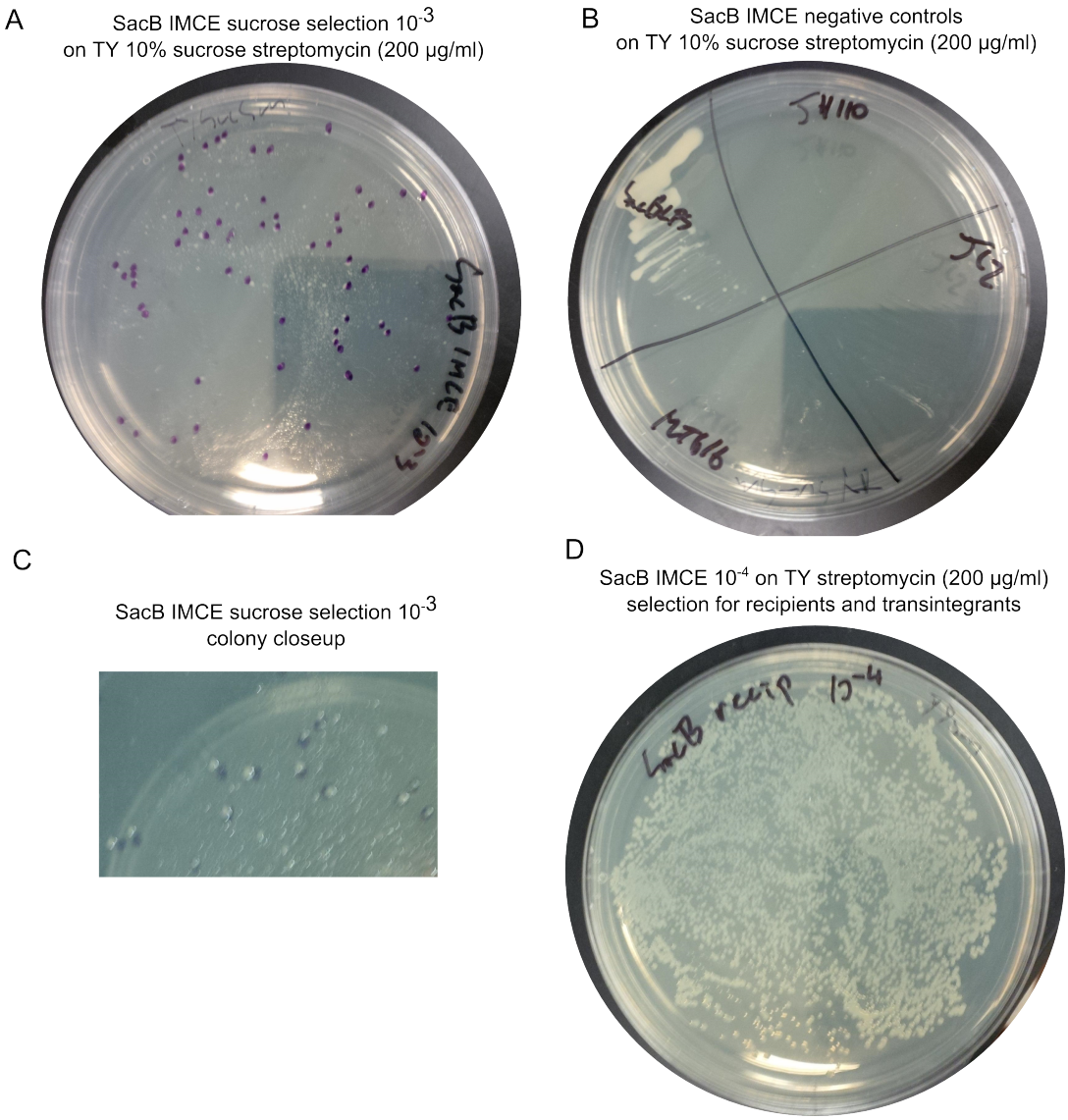

Figure 4.5: SacB IMCE selection plates. Panel A: The sucrose selection plate with large healthy colonies growing among some background; 10^{-3} dilution. Panel B: The individual strains, that were not mated, streaked on the same sucrose selection media (only a "headstreak"); SmUW574 (SacB LP-strain) can grow if the cell density is high, but it will not form healthy isolated colonies. Panel C: a closeup of the sucrose resistant colonies from panel A. Panel D: The mating spot plated on TY streptomycin with no sucrose; this level of growth at the 10^{-4} dilution relative to the growth on the sucrose 10^{-3} dilution shows that the sucrose selection worked.

4.3.3 Gateway Recombination into IMCE Vector

The Gateway IMCE donor vectors were confirmed by restriction mapping. Because they were constructed by subcloning the Gateway destination, a detailed restriction map is sufficient for their confirmation. To test the Gateway recombination, a cosmid clone (lac35) in pJC8 was used as a Gateway entry clone in the pentaparental mating procedure using pJH121 (Figure 4.6) as the destination vector. The pentaparental mating yielded colonies that were gentamicin resistant (selection for pJH121 plasmid backbone) and rifampicin resistant (selection for final *ccdB* sensitive recipient). These colonies arose at a frequency relative to recipient of approximately 1×10^{-6}, which is relatively low. House *et al.* (2004) indicated they obtained hundreds to thousands per 5 mm spot of mating mixture. A larger volume was used in the current work. Given the differences in the results are of orders of magnitude, two or four fold differences in initial cell numbers do not account for that frequency difference. There were no colonies when the pXINT129 strain was excluded from the mating. It is likely that the low numbers of colonies obtained were not merely spontaneous *ccdB* resistant mutants, or rifampicin resistant mutants. Culture was grown from three colonies and the plasmid DNA was isolated. The plasmid DNA was cleaved with EcoRI and BamHI. The size of the plasmids was too small to contain the cosmid clone, but the restriction pattern was different than that of pJH121. pJH121 gives bands of 6.6 kb, 1 kb, 0.7 kb, and 0.4 kb. The plasmids isolated from the Gateway recombination all gave 0.1 kb, 2 kb, and 8 kb. The proper clone was not isolated but the a recombination did occur.

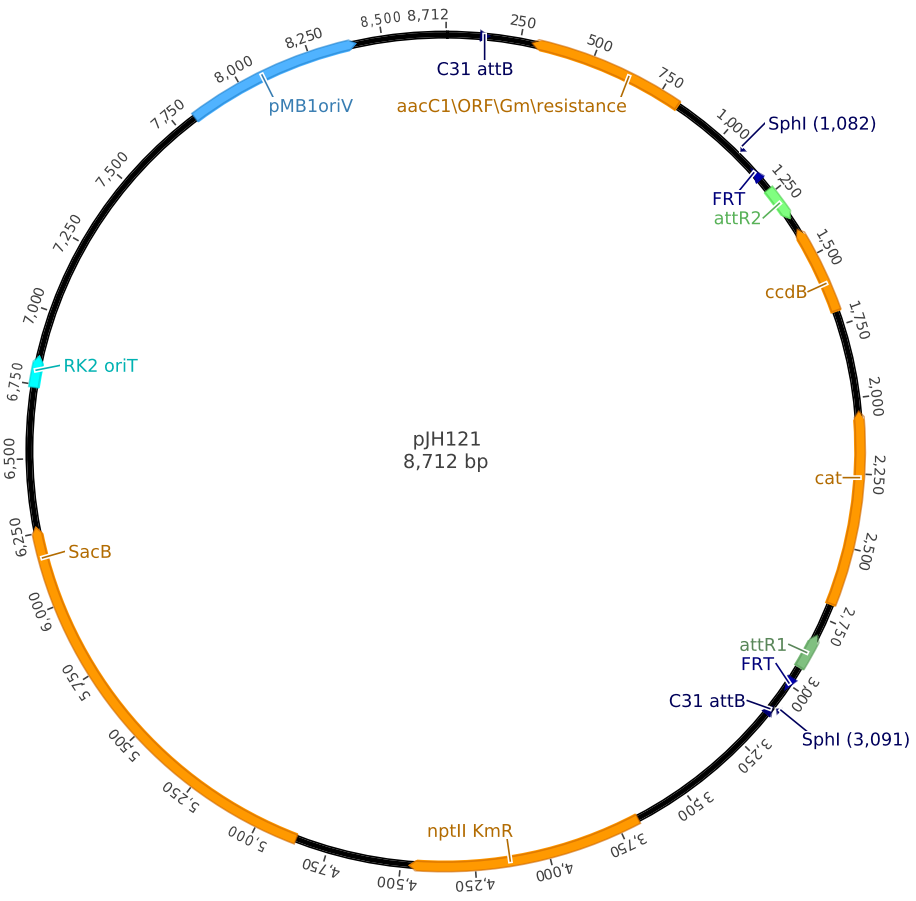

Figure 4.6: Gateway IMCE donor vector. The map of pJH121 is shown. It contains a Gateway destination flanked by ΦC31 *attB* sites. A gentamicin resistance gene is also included in the *attB* flanked cassette.

4.3.4 Test IMCE with pJH145

The fosmid vector pJH145 was constructed to allow for integration, via IMCE, of cosmid library inserts. The vector was tested to make sure the *attB* sites were functional. To test pJH145 an IMCE mating was done. The gentamicin resistance gene cloned in the Eco72I site occurs in between the two *attB* sites, the location of an insert had a library been made with the vector. Therefore, a gentamicin selection could be used for the IMCE. The IMCE was carried out with the identical procedure to the IMCE with pJH110 into SmUW227, except the donor was pJH145 instead of pJH110. IMCE also requires the donor vector to be mobilizable. The IMCE test therefore tested both the integrity of the *attB* sites and the functionality of the RK2 *oriT*. This IMCE worked with the same efficiency and as previous IMCE procedures. Double recombinants that were spectinomycin sensitive were isolated at the expected frequency (50%).

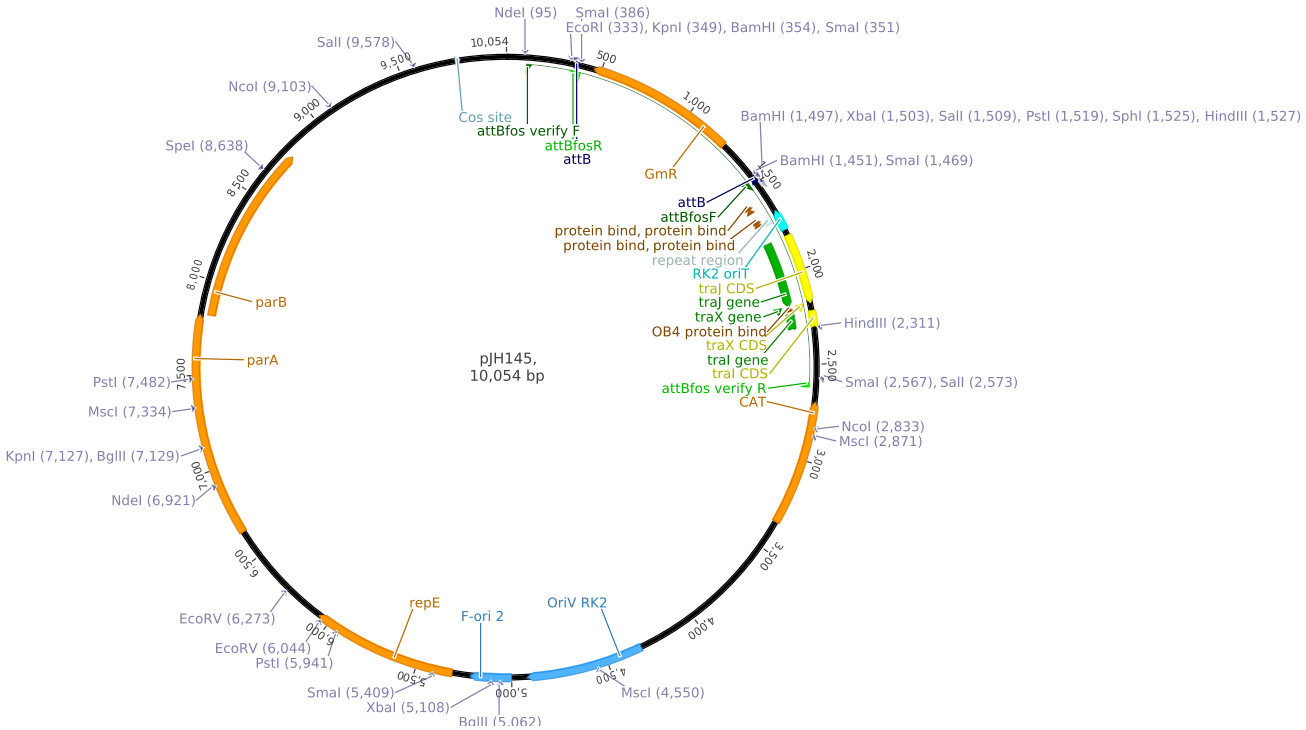

Figure 4.7: IMCE donor cosmid vector. The map of pJH145 is shown. It contains a blunt cloning site (Eco72I sites flanking Gm^R) flanked by *attB* sites. The vector is single copy, but is copy-up inducible in the EPI300 strain using L-arabinose

4.4 Discussion

To increase the utility of the IMCE system, the reverse IMCE system (RIMCE) was developed and it was used to create a new landing pad strain that increases the utility of IMCE by allowing for the integration of marker-less cassettes more easily, through the use of *sacB* counterselection. This SacB LP-strain is suited to be used as a recipient screening host in a functional screen. To accomplish this in the future, the vector pJH145 was created to allow for the creation of cosmid libraries that are compatible with IMCE. The vectors pJH121, pJH123, and pJH124 allow for Gateway entry clones to be integrated into IMCE *attB* donor vectors, further expanding the compatibility of the IMCE system with other systems.

RIMCE

In the beginning of this work the ΦC31 RDF had not yet been discovered. The RDFs for a number of other large serine recombinases had been discovered (Breuner *et al.*, 1999; Bibb and Hatfull, 2002; Ghosh *et al.*, 2006). The RDFs of the different phage systems were not related to each other in sequence, so each had to be isolated through a functional study. The ΦC31 RDF was isolated midway through the project (Khaleel *et al.*, 2011). The discovery of the RDF allowed me to take this work in several new directions.

The discovery of the ΦC31 RDF allowed for RIMCE to be accomplished. RIMCE was used to integrate a new landing pad marker cassette into SmUW246 to create the SacB LP-strain. This demonstrated that the landing pad sequence could be altered without having to make a new LP-strain from scratch. RIMCE was particularly useful when creating the SacB LP-strain because it circumvented the problem that creating a SacB LP-strain using homologous recombination posed. Creating a SacB LP-strain using homologous recombination would eliminate the utility of the SacB selection when

integrating the landing pad sequence into the chromosome. RIMCE has the potential to allow for the recovery of integrated inserts back onto plasmids. It seems the frequncy of *attL-attR* recombination is too low to accomplish recovery of the integrated clones back into *Escherichia coli*. The integrase-RDF expression plasmid (pJH141) was transferred into SmUW246 so that step would be eliminated in the clone recovery conjugation, but this did not boost the frequency enough to obtain a clone of the recovered pJH110 cassette in *E. coli*. For now, recovery of the integrated cassettes has not been accomplished. Because the frequency of the *attP-attB* reaction is higher, the conjugation conditions for clone recovery could be explored using IMCE through an *attB-attP* reaction. The landing pad markers could be recovered on a plasmid into an *E. coli* background. Another potential way to improve the chances of RIMCE clone recovery is to address the expression of the RDF. Increasing RDF expression could increase the frequency of the *attL-attR* reaction.

The lower frequency of the *attL-attR* reaction relative to the *attB-attP* reaction is evident in the lower frequency of transintegrants given by RIMCE relative to IMCE. That lower frequency can also explain the lower proportion of double recombinants in RIMCE. It stands to reason that the limiting step in IMCE is the conjugation because around 50% of transintegrants are double recombinants. If the overall frequency was limited by recombination you would expect to see that the frequency of double recombination was the square of the frequency of single-recombination. For IMCE, the frequency of double transintegrants is much higher than 10^{-6}. Using this logic, it is probable that the the reduction of the recombination frequency would reduce the proportion of double recombinants relative to transintegrants. This is what is seen in RIMCE.

A dimer of the RDF gene (*gp3*) in pJH141 could be contributing to the lower *attL-attR* reaction frequency. It is very strange that a monomeric clone could not be ob-

tained. For the purposes of creating the SacB LP-strain, pJH141 performed well. An earlier version of pJH141 was constructed using a synthesized clone of *gp3*, however the DNA sequence of *gp3*, that the synthesized clone was based on, turned out to be incorrect. Paul Reginato, an undergraduate thesis student, discovered this mistake when a BLAST search turned up a new revision of the *gp3* sequence. It was safe to assume that the synthetic *gp3* sequence was correct as it was specified by me, but it was not safe to assume it was correct as specified by nature. Interestingly the incorrect *gp3* clone that changed a number of amino acids, still yielded a few colonies. No double recombinants were ever found, but this was later found out to be because of a mutation in one of the *attL* sites on the earlier version of pJH143 (pJH129). Because of the low frequency of recombination with the incorrect *gp3*, double recombinants would be hard to find. Though the incorrect *gp3* is not useful for RIMCE, it may be worth noting its partial activity for the study of the RDF's activity.

SacB selection

The SacB selection in IMCE allows for more backgroud growth than the antibiotic selections, but it still provides a good positive-selection for double-recombination. When the first experiments were done, they were discounted because of the growth of the SacB LP-strain (SmUW574) on plates containing sucrose. Because there was far less growth on the sucrose plates when compared with the streptomycin only plates that SmUW574 grew on (data not shown), it became apparent that SmUW574 did not form isolated colonies on media containing sucrose, but only confluent growth when the cell density initially plated was extremely high. For this reason, if the frequency of the recombination were to drop too low, then it would be difficult to isolate colonies from background growth on sucrose containing media. It is conceivable that a larger *attB* donor cassette could cause the recombination frequency to drop. If that effect

turns out to be strong, the SacB selection may be insufficient for very large donor cassettes. However, that limitation remains to be seen, and to date the SacB selection has performed sufficiently.

Gateway ΦC31 Vectors

The Gateway vectors provide another tool to make IMCE compatible with more existing clones. Gateway cloning and Gateway entry clones are common in research labs. For a lab that uses *in vitro* Gateway cloning, the use of the Gateway ΦC31 *attB* vectors would be simplified compared to their use in the pentaparental *in vivo* Gateway recombination mating described in this chapter. One potential problem with the vectors (pJH121, pJH123, pJH124) is that they are high copy number. A large insert could be unstable in these vectors. In the test Gateway recombination, a large cosmid insert was used in the *in vivo* Gateway recombination. The resultant isolated plasmid was not pJH121, although it had the backbone of pJH121. Therefore, Gateway recombination is likely to have occurred. It could have been that the large insert that was stored on the pJC8 cosmid (a low copy number plasmid) was unstable in the high copy number plasmid. In this event, the only stable Gateway recombined clones were ones having undergone internal recombinations that resulted in a lessening of the insert size. A recombination event is not very likely in the *E. coli* strains used, because they are *recA* mutants. However, there are RecA independent pathways in *E. coli* that are usually repressed by single-strand exonucleases, by elimination of the ssDNA required for strand invasion (Dutra *et al.*, 2007; Lovett *et al.*, 2002). The *in vivo* Gateway recombination procedure is based on conjugation. Conjugation does increase the level of transient ssDNA in the *E. coli* cell (Baharoglu *et al.*, 2010; Bialowska-Hobrzanska and Kunicki-Goldfinger, 1977). This could cause the competitive inhibition of the aforementioned ssDNA exonucleaes. Therefore, the level of recombination through RecA

independent pathways could increase during conjugation. It is also worth noting that the number of Gateway clones recovered was quite low. The amount of Gateway recombinants reported by House *et al.* (2004) was hundreds to thousands of colonies for a 5 mm diameter mating spot. THE mating spots in this research were approximately 20 mm in diameter with less than 10 colonies isolated. For larger sized inserts, a low copy number Gateway IMCE vector would be preferable to the high copy number vectors that were constructed.

ΦC31 *attB* Fosmid

The IMCE donor vector fosmid, pJH145, is a single-copy (inducible to medium copy) cosmid vector that only replicates in *E. coli*. It is mobilizable for use in IMCE. A higher copy number can be induced when pJH145 is in the *E. coli* EPI300 host when grown on media containing 10 mM L-arabinose. The IMCE test worked as expected, and the blunt cloning site (for cosmid library construction) restricts as expected. The gentamicin resistance gene was cloned into the blunt site because it had been noted that the efficiency of the restriction was improved relative to having the single restriction site (J. Cheng, personal communication). The pJH145 vector can also function as a bacterial artificial chromosome (BAC). The vector is much better suited for maintaining large fragments of DNA in *E. coli* for integration into the chromosome of other strains by IMCE.

Chapter 5

Recombination Synthesis of PHA Genes

5.1 Introduction

To increase the utility of the overall chromosome engineering system, a homologous recombination based system, called recombination synthesis was devised. Recombination synthesis (RS) is designed to allow overlapping IMCE clones to be combined so that the resulting construct is present in the IMCE locus. This combination of the two IMCE inserts is based on a specific homologous recombination event that is selected by two antibiotic resistance genes flanking the entire goal construct (Figure 5.1). The design of RS allows for either the assembly of smaller overlapping IMCE inserts into a larger construct, or the combinatorial assembly of a library of IMCE inserts with common overlaps.

With the goal of developing strains for polyhydroxyalkanoate (PHA) synthesis, the application and proof of concept of RS was to make medium chain length (mcl) PHA synthesis operons. Mcl-PHAs have quite different physical properties from

short chain length (scl) PHAs. The most common scl PHA, PHB, has properties close to isotactic polypropylene (crystalline polypropylene with the side chains all in the same orientation) (Griffin, 1994). PHB does have some limitations when compared to polypropylene. It is more fragile than polypropylene and is thermally unstable at temperatures above its melting temperature. This can be mitigated by mixing PHB with other polymers (Pachekoski *et al.*, 2009). Creating various co-polymers comprised of mcl-scl mixes, or different mcls, has the potential to produce a material that overcomes some of the limitations of scl-PHAs and opens up new applications of PHAs. Currently, mcl PHAs are most commonly produced in *Pseudomonas*. Producing mcl-PHAs in natural scl producers has the potential to open the door to the aforementioned new materials. I hypothesized that co-expressing *phaG* and *phaC* originating from mcl producers in *Sinorhizobium meliloti*, which does not naturally produce mcl-PHA, would confer on *S. meliloti* the ability to produce mcl-PHA.

The operons for the proof of concept contain both a *phaG* gene and a *phaC* gene. The *phaG* gene encodes (R)-3-hydroxydecanol-ACP:CoA transacylase (PhaG). PhaG is responsible for removing the acyl carrier protein (ACP) from 3-hydroxyacyl intermediates of fatty acid synthesis and replacing ACP with a coenzyme-A (CoA) molecule (Fiedler *et al.*, 2000). These fatty acid CoA thioesters are then imbued with the chemical energy to power PHA synthase (PhaC). The nature of the PHA produced may depend on the substrate specificity of the PhaC, the substrate specificity of the PhaG, which would affect what PhaC substrates are available, and the background metabolism of the cell (various rates of metabolic steps). The construction of PHA synthesis operons containing orthologues of *phaG* and *phaC* was undertaken. The *egfp* (Cormack *et al.*, 1996) gene was used as the common overlap, and a *xylE* reporter (Ingram *et al.*, 1989) was placed after *phaC*. To produce these operons, half operons were designed so

that when inserted by IMCE, RS would fall into place (Figure 5.3). The DNA transfer method chosen was first transduction, and was later changed to conjugation.

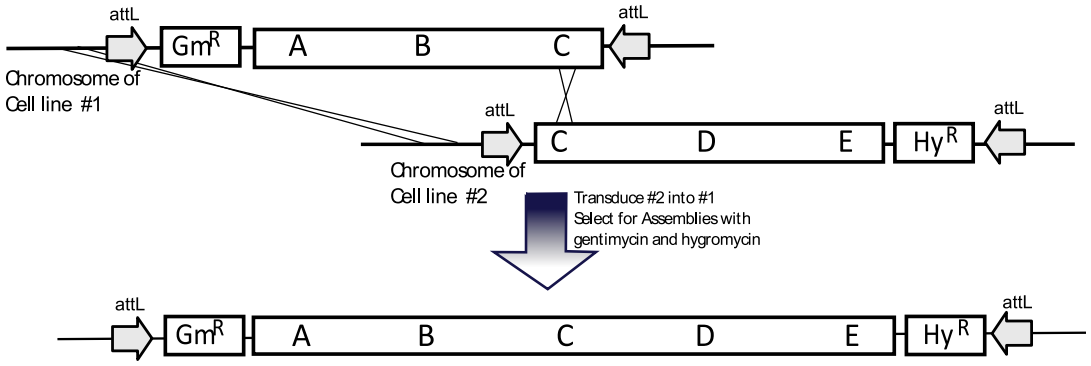

Figure 5.1: Recombination synthesis: DNA from one IMCE clone is transferred into another IMCE clone. There is overlapping sequence (C) that is a substrate for homologous recombination cross-over. On the left end there is a gentamicin resistance gene, and on the right end a hygromycin resistance gene. The desired homologous recombination event will yield cells resistant to both antibiotics. The DNA transfer method suggested in this figure is transduction. The use of another DNA transfer method may be necessary

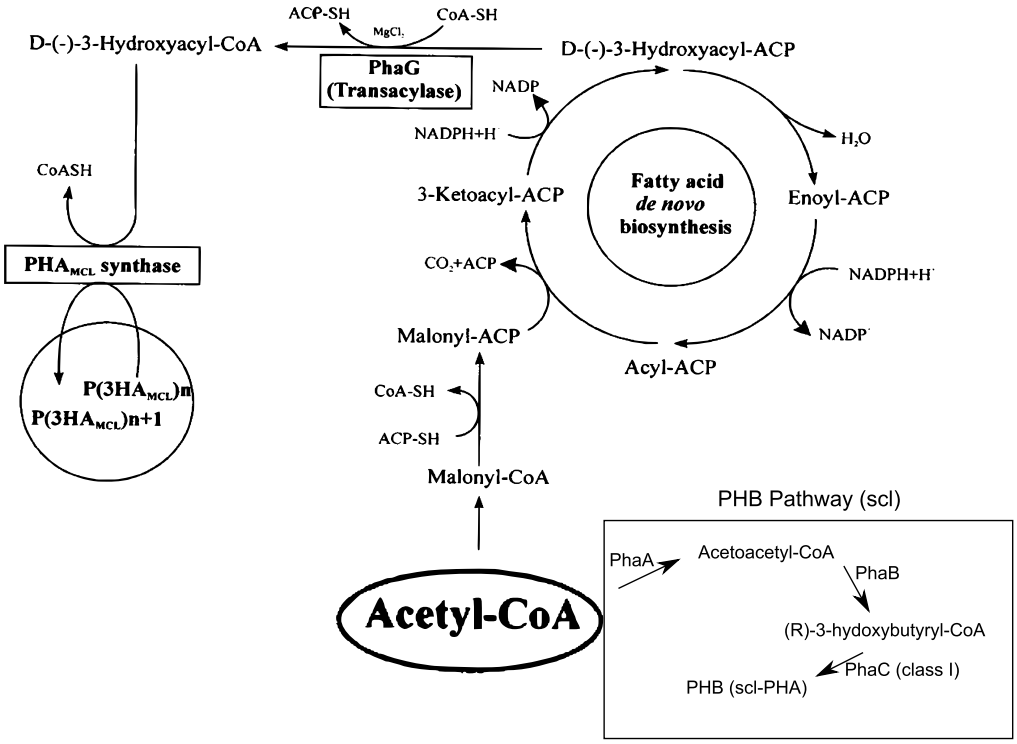

Figure 5.2: PHA pathways. For PHB production, a short chain length (scl) PHA, acetyl-CoA enters the PHB pathway. For medium chain length (mcl) PHA production, acetyl-CoA is transformed into malonyl-ACP were it enters the *de novo* fatty acid biosynthesis cycle. From there, the 3-hydroxyacyl intermediate is shunted by PhaG into PHA synthesis through PhaC (class II synthase), shown on the left. Adapted from Fiedler *et al.* (2000)

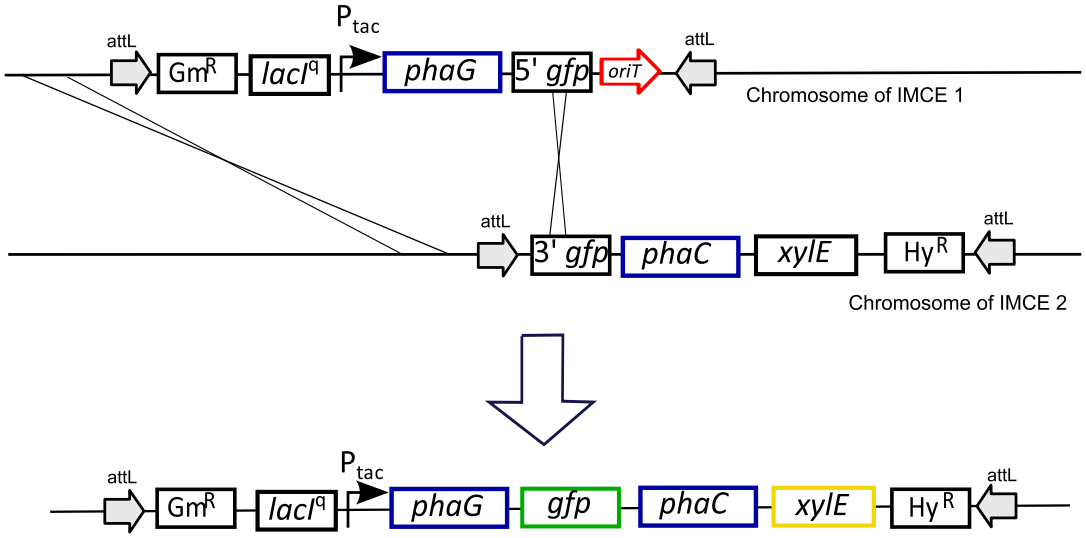

Figure 5.3: Above the arrow, the transfer of the *phaG* construct by conjugation into the *phaC* construct IMCE clone brings the two segments of DNA into proximity. The recombination that is drawn will result in the assembled operon shown below the arrow. This recombination event is selected using gentamicin and hygromycin. The *oriT* is recombined out in the bottom panel.

5.2 Materials and Methods

5.2.1 Construction of pJRT8

The plasmid pJRT8 was used to integrate, via IMCE, the upstream half operon containing *phaG* from *Pseudomonas putida* KT2440.

The primers GFPstart-F-PmeI-XbaI and GFPstart-R-SphI-EcoRI (Table 2.2) were used with the template pTH1703 (Table 2.1) to amplify a 5‘ truncated segment of *gfp*, including its ribosome binding site (RBS). A PmeI site was added for the future cloning of *phaG*. An SphI site was added for the future subcloning of the construct into pJH110. The PCR product was ligated into pK19mobSacB as an XbaI-EcoRI fragment to create pk19gfpstart. The primers lactac-F-PstI-SphI and lactac-R-PacI-XbaI were used with the template pTH1227 to amplify the *lacI*^{q} P_{lac} construct from pTH1227. An SphI site was added for future subcloning of the entire construct into pJH110, and the PacI site was added for cloning of *phaG*. The PCR product was cloned as a PstI-XbaI fragment into pk19gfpstart to create pJRT1 (Table 2.1). The pJRT1 construct was subcloned into pJH110 as an SphI fragment to create pJRT13. The desired orientation was verified by restriction mapping with HindIII. pJRT13 gave HindIII fragments of 2.5 kb and 6 kb as opposed to the plasmid with the opposite insert orientation that gave HindIII fragments of 0.5 kb and 8 kb. The primers phaG-F-PacI and phaG-R-PmeI-HIS were used to amplify *phaG* using *P. putida* KT2440 genomic DNA as the template. A C-terminal his-tag was added to *phaG*. The *phaG* PCR product was cloned as a PacI-PmeI fragment into pJRT13 to create pJRT8. Ricardo Nordeste and Tam Tran aided in the construction of this vector, by contributing to the cloning strategy and cloning in the *phaG* gene.

5.2.2 Construction of pJRT9

The plasmid pJRT9 was used to integrate, via IMCE, the upstream half operon containing *phaG* from *P. putida* UW4.

The primers phaGUW4-F-PacI and phaGuW4-R-PmeI-HIS were used to amplify *phaG* using *P. putida* UW4 genomic DNA as a template. The PCR product was cloned into pJRT13 as a PacI-PmeI fragment to create pJRT9.

5.2.3 Construction of pJRT10

The plasmid pJRT10 was used to integrate, via IMCE, the downstream half operon containing *Sinorhizobium meliloti* codon optimized *phaC* from *P. putida* GPo1.

The primers GFPend-F-PstI-SphI and GFPend-R-PacI-XbaI were used to amplify a 3′ *gfp* using pTH1703 as a template. The PacI site was added for the future cloning of *phaC*. The SphI site was added for the future subcloning of the marker construct into pJH122. The PCR product was cloned as a PstI-XbaI fragment into pK19mobSacB to create pK19gfpend. The primers XylE/T-F-PmeI-XbaI and XylE/T-R-SphI-EcoRI were used to amplify the overlapping *xylE* and *xylT* open reading frames (ORF) including their RBSs, using mini-Tn5-XylE (de Lorenzo *et al.*, 1990) as the template. The PCR product was cloned as an XbaI-EcoRI fragment into pk19gfpend to create pJRT2. The SphI fragment from pJRT2 was subcloned into pJH122 to create pJRT14. The orientation of the SphI fragment in pJRT14 was verified with with a HindIII digest, that yielded HindII fragments of 4 kb and 5.5 kb. The primers GPO1-F-PacI and GPO1-R-PmeI were used to amplify *phaC*, adding a C-terminal his-tag, using a synthetic gene codon optimized for *S. meliloti* (optimizer program) as the template (Puigbò *et al.*, 2007). This PCR product was cloned into pJRT14 as a PacI-PmeI fragment to create pJRT10.

5.2.4 Construction of pJRT12

The plasmid pJRT12 was used to integrate, via IMCE, the downstream half operon containing *S. meliloti* codon optimized *phaC* from *P. putida* pTAM (*pha*C1400) (Tran, 2015).

The primers phaC1400-F-PacI(2) and phaC1400-R-PmeI(2) were used to amplify the *phaC1400* ORF using the codon optimized synthetic DNA construct containing *phaC1400* (designed by Tam Tran, contained on pTAM) as the template (Tran, 2015). The PCR product was cloned into pJRT14 as a PacI-PmeI fragment to create pJRT12.

5.2.5 Construction of pJRT15

The plasmid pJRT15 was used in the same way as pJRT8, except it contained the R388 *oriT* in the IMCE cassette, allowing for the resultant IMCE strain's use as a donor in conjugative RS.

The primers R388-F-attB-XbaI and R388-SalI-R were used to amplify the R388 *oriT*, with the ΦC31 *attB* site included. The template used was pSU711. The PCR product was cloned into pJRT8 as an XbaI-SalI fragment, replacing *attB* and a small segment of the already truncated *gfp* with the R388 *oriT-attB* construct. This created pJRT15.

5.2.6 Construction of pJRT16

The plasmid pJRT16 was used in the same way as pJRT9, except it contained the R388 *oriT* in the IMCE cassette, allowing for the resultant IMCE strain's use as a donor in conjugative RS.

pJRT16 was created in the same way as pJRT15 except the XbaI-SalI fragment was cloned into pJRT9 to create pJRT16.

5.2.7 Construction of pJH146

The plasmid pJH146 is a pJET clone of one of the *in vitro* PCR assembled PHA operons. To construct the plasmid, the primers phaG-gfp-F-SphI and phaG-gfpR-XO (Table 2.2) were used to amplify the *P. putida* KT2440 *phaG*, including the $lacI^q$, P_{tac} and truncated *gfp*, using pJRT8 (Table 2.1) as a template. The primers phaC-gfpF-XO and phaC-gfpR-SphI (Table 2.2) were used to amplify the *P. putida* GPo1 *phaC* including the truncated *gfp* and *xylE, xylT* genes (note *xylE* reporter is in overlapping ORF with *xylT*), using pJRT10 (Table 2.1) as a template. These two PCR products were gel purified and used as a template in a following PCR reaction using the primers phaG-gfp-F-SphI and phaC-gfpR-SphI. This second PCR step allows the overlap between the truncated *gfp*s on both templates to act as the internal primers thereby, assembling the two templates into one amplicon and assembling the operon. This large assembled PCR product was gel purified and ligated into the pJET vector using the cloneJET PCR cloning kit (Fisher Scientific Inc.) to create pJH146

5.2.8 Construction of pJH147

The plasmid pJH147 is a pJET clone of one of the *in vitro* PCR assembled PHA operons. The same strategy and primers were used to create pJH147 as was used to create pJH146 except that the initial templates in the first round of PCR were pJRT8 and pJRT12, to give the assembly containing *P. putida* KT2440 *phaG* and *phaC* 1400.

5.2.9 Construction of pJH148, pJH149, and pJH150

The plasmids pJH148, pJH149, and pJH150 are pJET clones of one of the *in vitro* PCR assembled PHA operons. Again the same strategy and primers were used as in the construction of pJH146. The templates used for the initial round of PCR were pJRT9

and pJRT10, to give the assembly containing *P. putida* UW4 *phaG* and *P. putida* GPo1 *phaC*. From the initial screen a few clones that had the correct restriction pattern were saved. pJH148 and pJH149 were ostensibly equivalent. pJH150 colonies fluoresced under GFP excitation conditions.

5.2.10 Construction of pJH151

The plasmid pJH151 is a subclone of the operon from pJH148 that could be integrated into a LP-strain with IMCE. To construct pJH151, the insert from pJH148 was subcloned into pJH110 as an SphI fragment (replacing *rfp*), so that the completely assembled operon could be integrated into LP-strains with IMCE.

5.2.11 SmUW235 ϕ SmUW227

Ricardo Nordeste helped construct this strain. Using transduction (section 2.4), a ΦM12 lysate of SmUW227 was used to infect SmUW235 and transductants were selected with spectinomycin at 100 μg/ml. The strain is a *phbC* deletion strain that contains the landing pad sequence. *phbC* is the *S. meliloti* PHA synthase. The deletion was made using the standard techniques involving cloning homologous DNA flanking the gene into pK19mobSacB (Schafer *et al.*, 1994), and using a two step selection method. This selection method was described in Section 3.2.1.

5.2.12 SmUW497 ϕ SmUW227

As with SmUW235 ϕ SmUW227, the landing pad sequence was added to SmUW497 by transduction. Ricardo Nordeste constructed the strain. SmUW497 has all 3 PHB synthesis genes deleted (*phbC, phbA, phbB*), and has the PHA depolymerase gene deleted (*phaZ*). Deletions were made using the standard techniques involving cloning

homologous DNA flanking the gene into pK19mobSacB (Schafer *et al.*, 1994), and using a two step selection method. This selection method was described in Section 3.2.1.

5.2.13 IMCE

IMCE (section 3.2.8) was used to integrate into *S. meliloti* SmUW235 ϕ SmUW227, the PHA half operon contained between the two *attB* sites in the donor vectors: pJRT8, pJRT9, pJRT10, pJRT12, pJRT15, pJRT16. The orientation and presence of the half operons was verified by colony PCR (sections 2.7.9 and 3.2.9). The same was done for *Ochrobactrum anthropi* Ow0013 and *Agrobacterium tumefaciens* At11142. The colony PCR was used to verify the orientation of Ow0013 IMCE inserts.

IMCE was also used to integrate the *in vitro* cloned operon from pJH151 into *S. meliloti* SmUW235 ϕ SmUW227, and SmUW497 ϕ SmUW227.

5.2.14 RS Triparental Mating

These matings were carried out with both the *S. meliloti* and *O. anthropi* IMCE clones. The matings were done between clones of the same species. IMCE clones in the *S. meliloti* background were paired with the other IMCE clones of the same background. The same procedure was carried out for the IMCE clones in the *O. anthropi* background.

Preparation of Culture

Cultures of the donor (JRT15 IMCE or JRT16 IMCE), recipient (JRT10 IMCE or JRT12 IMCE), and mobilizer strain (pSU711 in *Escherichia coli* DH5α) were prepared by inoculating a single colony in 5 ml of LB media. Kanamycin was added to the mobilizer culture at 12.5 μg/ml for plasmid maintenance. The *S. meliloti* landing pad strain was incubated at 30°C with constant shaking at 180 RPM for 2-3 days, until the culture was

saturated. The mobilizer strain, in *E. coli* was incubated at 37°C for 16-20 hours with constant shaking at 180 RPM.

Mating

Once cultures were grown, the procedure stated in section 2.5 was followed. The cultures were combined so that the following combinations were present: JRT15-JRT10, JRT15-JRT12, JRT16-JRT10, JRT16-JRT12. Those combinations contained Kt2440 *phaG* and GpO1 *phaC*, Kt2440 *phaG* and 1400 *phaC*, UW4 *phaG* and GpO1 *phaC*, UW4 *phaG* and 1400 *phaC*, respectively.

Selection

After the incubation the control spots were streaked on selective media containing hygromycin at 50 μg/ml and gentamicin at 25 μg/ml, that was counterselective against donor and recipient, therefore selective for transconjugants. The triparental mating spot was also streaked on the same media. One quarter to one half of the triparental mating spot was resuspended in 0.85% NaCl and a 10 fold dilution series to 10^{-7}. Typically this was done by diluting 10 μl of culture into 90 μl of 0.85% NaCl. 40 μl of each dilution, 10^{-1}-10^{-4}, was spread plated on media selective for transconjugants; 40 μl of each dilution 10^{-4}-10^{-7} on media selective for recipients and transconjugants. Plates were then incubated at 30°C for three days in the case of *S. meliloti* background strains, and 2 days for *O. anthropi* background strains. *O. anthropi* backgrounds required a higher amount of hygromycin for selection: 100-200 μg/ml.

5.2.15 RS Transduction

For the *S. meliloti* IMCE clones, transductions were carried out as described in section 2.4. Lysates of each IMCE strain (JRT8, JRT9, JRT10, JRT12) were made. The

same IMCE strains were used as recipients. JRT8 and JRT9 lysates were used to infect JRT10 and JRT12 cultures, and RS transductants were selected on LB agar containing 50 μg/ml gentamicin and 50 μg/ml hygromycin. As a positive control general transductants were selected on LB agar containing 50 μg/ml gentamicin. Conversely, JRT10 and JRT12 lysates were used to infect JRT8 and JRT9 cultures. RS transductants were selected on LB agar containing 50 μg/ml gentamicin and 50 μg/ml hygromycin. As a positive control general transductants were selected on LB agar containing 50 μg/ml of hygromycin.

5.2.16 Reporter, and Fluorescent Dye Visualization

XylE

An aqueous solution of 0.5 M catechol was sprayed directly onto the colonies. Colonies turned yellow in 5 to 60 seconds (Ingram *et al.*, 1989).

GFP fluorescence

Colonies containing GFP on agar Petri plates were visualized by shining white light (from a white phosphor plate on a UV transilluminator) through a LEE Filters 120 deep blue filter, onto the plate, and then placing a LEE Filters 139 primary green filter on the plate. This assembly was inside a dark box with a camera mounted on top. Before the camera there was also a UV filter that was used for DNA gel imaging. This generic filter was of a yellow colour and helped cut the background green in the images, increasing contrast. This method was derived from the RFP visualization method for plates that I developed (Heil *et al.*, 2011).

Nile Red Fluorescence

Colonies suspected to make PHAs were grown on yeast mannitol agar (YMA) containing 0.5 μg/ml Nile Red (Spiekermann *et al.*, 1999), with and without 0.5 mM IPTG. The fluorescence was visualized in the same was as above, with the excitation filter being Lee Filters 139 Primary Green and the emission filter being Lee Filters 128 Bright Pink (Heil *et al.*, 2011).

5.2.17 Colony PCR and Sequencing

Colony PCR was done using the protocol from section 2.7.9. The left border of the hypothetical construct was targeted with the primers: L-fwd-EcoRI and Gm-rev. The right border was targeted with: Hy-Fwd and R-rev-HindIII. The central gfp region was targeted with: GFPstart-F-PmeI-XbaI and GFPend-R-PacI-XbaI. These primer sets were used on colonies of each RS combination.

5.2.18 Direct Genomic Sanger Sequencing

Direct genomic Sanger sequencing was carried out at the McMaster Mobix Lab (Hamilton ON). Samples of *S. meliloti* genomic DNA were sent at a concentration of 2 μg/μl. 5 μl of DNA was used for each read. The primers were sent at a concentration of 20 μM. To obtain large amounts of concentrated DNA, a large scale genomic DNA preparation protocol was used (section 2.8.4)

The genomic DNA template for the JRT15-JRT12 combination was sequenced with the primers: gfpXO-41-F, gfpXO-313-F, gfpXO-162-R, gfpXO-691-R, gfpXO-755-R. The other three combinations were sequenced with the primers: gfpXO-755-R, and gfp-junction-R (Table 2.2).

5.2.19 Transduction Linkage Mapping

Lysates of the JRT15-JRT12 RS strain, and the JRT16-JRT10 RS strain were made according to the method described in section 2.4. The lysates were used to infect *S. meliloti* Rm1021. Transductants were selected on LB plates containing 50 μg/ml hygromycin. The plates were incubated for three days at 30°C. The transductants that arose were checked for gentamicin resistance on LB plates containing 50 μg/ml gentamicin. Transductants that were also gentamicin resistant were co-transductants. A graph of a linkage distance as a function of cotransductant frequency based on Wu's formula was used to estimate the linkage distance (Charles, 1990).

5.2.20 Cell PHA Content Analysis Gas Chromatography and Mass Spectrometry (CG-MS)

GC-MS analysis was used to detect PHAs present in the bacterial strains reported in this thesis. The GC-MS method was first reported by Braunegg *et al.* (1978).

For the GC-MS analysis of cultures 50 ml of culture was prepared using the methods in section 2.3. For *S. meliloti* the 50 ml culture was grown for 3 days at 30°C with constant shaking. 0.5 mM IPTG, for induction of the putative PHA operon, was included in the media.

The clones used were: JRT15-JRT10, JRT15-JRT12, JRT16-JRT10, JRT16-JRT12, SmUW235 ϕ SmUW227, Rm1021, JRT15 IMCE, JRT16 IMCE, JRT10 IMCE, and JRT12 IMCE. Pure polymer standards were run by Ricardo Nordeste previously using the same protocol, on the same instrument.

Derivatization of Samples for Lipid Analysis

The culture was decanted into a 50 ml centrifuge tube and the cells were pelleted by centrifugation at 7500 x g for 5 minutes. The supernatant was discarded. The cells were resuspended in 45 ml of deionized water and centrifuged again with the same settings. The supernatant was discarded. This wash was repeated three more times. It was important to note that once the pellet was suspended in water it did not reform a very strong pellet, so care was taken not to lose the pellet.The pellet was dried in a 60°C incubator over a period of 1-2 days.

Next the pellet was carefully placed into a pre-weighed KIMAX 5 ml screw cap test tube. The dry weight of the pellet was recorded. Then 2 ml of chloroform was added to the tube. Then 1 ml of 15 % 18 M sulphuric acid in methanol was added. The tube was tightly capped. The tube was placed on a 100°C heat block for 6 hours. The tube was allowed to cool to room temperature and then 1 ml of deionized water was added. The tube was vortexed for 30 seconds. The phases were allowed to separate and then approximately 1 ml of the organic layer (bottom layer) was removed with a pasteur pipette and placed in a GC-vial.

GC-MS protocol

An Agilent GC-MS system with a 6890N GC, 5975B mass spectrometer and a 7693 autosampler was used (Agilent Technologies, Santa Clara, California). The column used was a capillary HP–5MS 5% phenyl methyl siloxane column (30 m x 0.25 mm, film thickness 0.25 μm, Agilent 19091S-433). Injector syringe wash vials were washed and filled with chloroform. A 10 μl syringe was used. The injection volume was 1 μl. Activation of the electron ionization filament was delayed by 3 minutes after injection to protect the filament from the solvent front. The solvent front came through at 1.9 minutes. After injection the oven temperature was held at 50°C for 5 minutes. Then

the oven temperature was increased at 20°C per minute for 11.5 minutes to reach a temperature of 280°C. 280°C was held for 10 minutes.

5.3 Results

5.3.1 RS Transduction

To assemble PHA operons on the *Sinorhizobium meliloti* chromosome, the half operons were assembled on plasmids and integrated by IMCE for a subsequent genetic cross that would accomplish RS. The half-operon plasmids: pJRT8, pJRT9, pJRT10, and pJRT12 were constructed such that, after IMCE (Figure 5.4), the resultant strains could be crossed, achieving a crossover event as seen in Figure 5.1. A sequence alignment of the two homologues of each PHA gene included in these constructs was done to highlight their similarities and differences (Figures 5.5, 5.6). The first method tried for achieving this genetic cross to assemble the half operons was transduction (Finan *et al.*, 1986). The IMCE strains of *S. meliloti* were used in the transductions. The crossover event was selected with the two antibiotics, gentamicin and hygromycin. No colonies grew in these selections. The titer of the phage in the lysates was adequate, ranging from 7×10^9 to 5×10^{10} PFU per ml. The control transductions used a less stringent selection, only selecting for the resistance contained on the DNA in the transducing particle. For pJRT8 and pJRT9 lysates, that was gentamicin; for pJRT10 and pJRT12 lysates, that was hygromycin. These control transductions yielded 5-10 colonies consistently. With that frequency of successful transductants, it was unlikely that transductants would be present that satisfied the more specific double selection. No RS transductants were isolated.

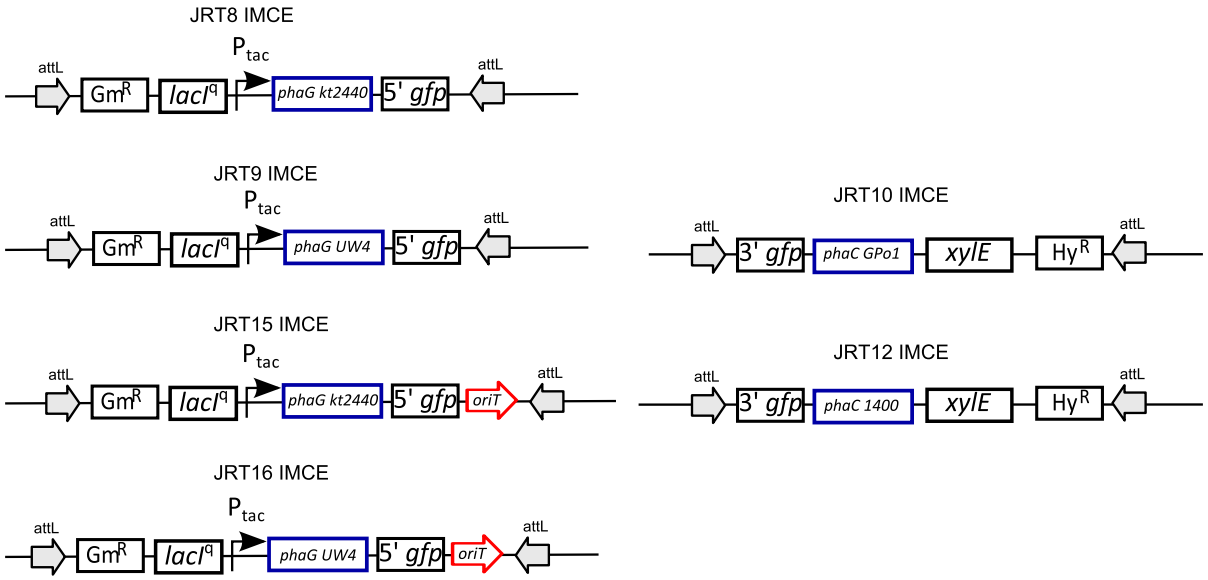

Figure 5.4: PHA half operons. In the left column are the IMCE JRT variants that contain a *phaG* gene; *phaG*Kt2440 is from *P. putida* Kt2440; *phaG* UW4 is from *P. putida* UW4. In the right column are the *phaC* containing variants; *phaC* GPo1 is from *P. putida* GPo1; *phaC* 1400 is a *phaC* isolated by in through mutation experiments (Yang *et al.*, 2010)

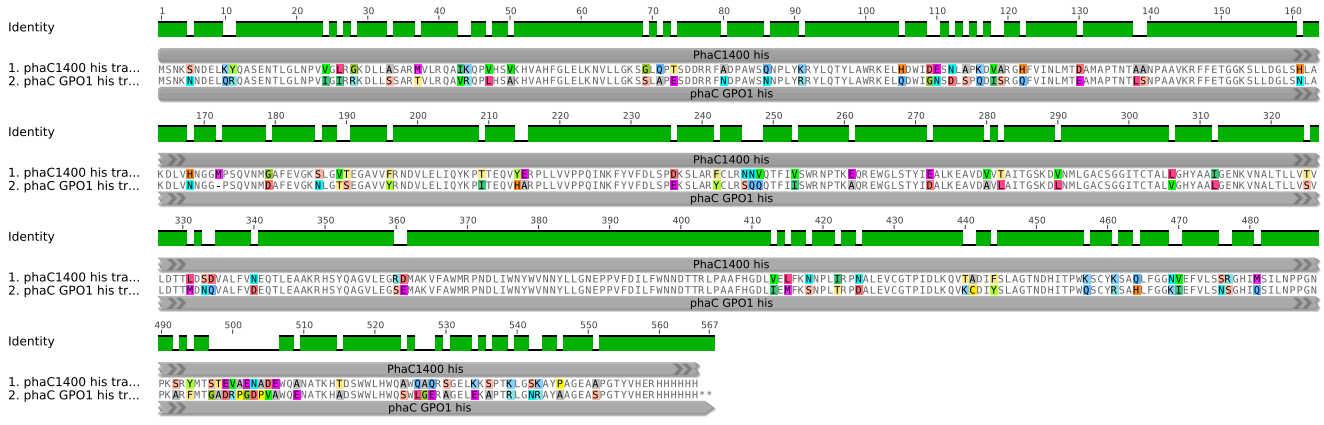

Figure 5.5: PhaC sequence alignment. A global pairwise sequence alignment of the amino acid sequence of *phaC*1400 and *phaC* from GPo1 is shown. Mismatches are highlighted.

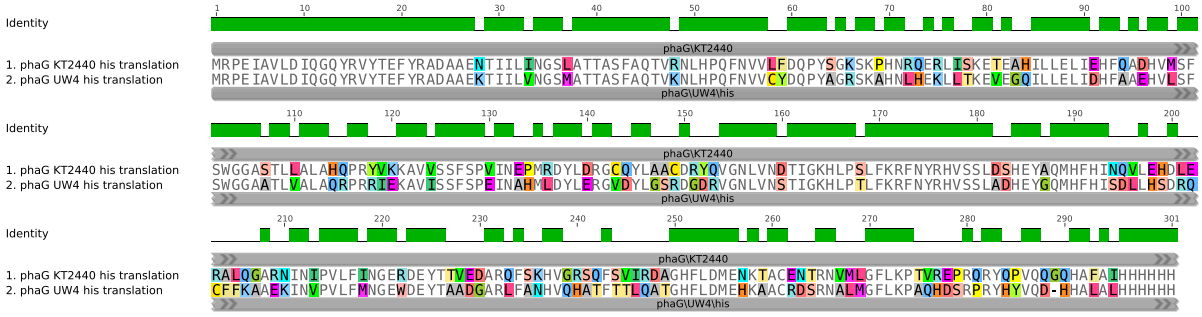

Figure 5.6: PhaG sequence alignment. A global pairwise sequence alignment of the amino acid sequence of *phaG* from KT2440 and *phaG* from UW4 is shown. Mismatches are highlighted.

5.3.2 RS Conjugation

After the failure of transduction as an RS genetic cross method, conjugation was chosen. It was hypothesized that a more specific region of DNA (i.e. the target region) could would be transferred, leading to a higher likelihood of successful RS

Because the IMCE protocol already made use of the RK2 conjugation system, the R388 conjugation system was chosen for use with RS. If the RK2 system had been used the donor plasmids would have carried two RK2 *oriT*s. One in the *attB* flanked cassette and one on the vector backbone. This may have caused conjugation of a truncated vector segment between the two *oriT*s (Chain, 1998). The R388 conjugation system was chosen because it was another broad host range conjugation system that had been used successfully (Demarre *et al.*, 2005). If the R388 *oriT* was recognized by the RK2 transfer proteins IMCE may have yielded only single-recombinants. IMCE worked as it did before, indicating that there was no problem caused to the addition of the R388 *oriT*.

For RS using the R388 conjugation system the IMCE strains containing pJRT15 and pJRT16 (Figure 5.3) were used as donors. These strains are similar to JRT8 and JRT9 IMCE strains, except that they contain an R388 *oriT*. Transconjugants were isolated at a frequency of 10^{-4}. The isolated colonies did not show any GFP fluorescence, nor did they show any *xylE* yellow colour. *Escherichia coli*, carrying pJRT14, did turn yellow when sprayed with catechol (Figure 5.8). *E. coli* carrying the same vector but with *phaC* cloned before *xylE* (pJRT10) did not turn yellow. IMCE was done with pJRT14 and RS was done with JRT14 IMCE as a recipient. The resultant colonies still did not turn yellow when sprayed with catechol. No GFP was detected either. The same was done for the *Ochrobactrum anthropi* LP-strain, Ow0013. The frequency of colonies for *O. anthropi* Ow0013 RS was lower at 10^{-5}. The Ow0013 RS colonies did not display the reporter phenotypes either. The Ow0013 IMCE transintegrant clones used for RS

were added to the strain collection. The resultant RS clones were not added.

The *S. meliloti* RS clones were examined more closely. The clones did have the gentamicin resistance phenotype of the donor and the hygromycin resistance phenotype of the recipient. The IMCE strains containing *phaG* (JRT15 and JRT16) fluoresced on media containing Nile red dye. The fluorescence was stronger on plates containing IPTG and Nile red, indicating that the induction of the P_{tac} promoter is working (Figure 5.9). The RS strains also have this phenotype, where the IMCE JRT10 and JRT12 strains (recipients in RS conjugation) do not have this phenotype. It is apparent from the transfer of the Nile red phenotype that the genes in the landing pad locus of the donor have been transferred and integrated into the recipient somewhere.

The phenotype of the cells was not as expected; expression of the reporter genes could not be detected. Because of the possibility of poor expression of the reporter genes due to a variety of factors, I chose to physically verify that the operons were indeed properly assembled. This was done with the *S. meliloti* RS variants only. Colony PCR to amplify the actual point of crossing over (the *gfp* gene) and the two borders of the landing pad sequence (Figure 5.7). Colony PCR successfully amplified the correct size amplicon for all but RS JRT16-JRT10 with the left border primer set (Figure 5.10). Each PCR product derived from the *gfp* primer set was sequenced. The sequence for each PCR product indicated a fully intact *gfp* ORF. Since before the RS conjugation the IMCE strains had contained truncated *gfp* ORFs, this was indicative of a successful recombination. The fact that the left and right borders were amplified (except for one case) was also indicative of a properly assembled operon. However, due to the fact that the *gfp* halves overlapped, even separated *gfp* halves could produce the amplified PCR product that was observed. Direct Sanger sequencing using a genomic template was used to give the definitive answer. This technique is not vulnerable to *in vitro* assembly of two actually separated overlapping segments of DNA. The genomic

Sanger sequencing revealed that the *gfp* segments were not actually assembled, and that the downstream half operon (containing *phaC*) was actually in the landing pad locus independently. The sequence of the ΦC31 *attL* site appears directly upstream of the point of 5′ truncation of the downstream *gfp* segment. Further upstream of the *attL* site, the *S. meliloti* genomic sequence is found. This was the case for all four RS strains sequenced. Given the positive PCR results for the left and right borders, where there should be no overlapping sequence, it is reasonable to hypothesize that the landing pad locus of the donor had inserted somewhere randomly, thereby duplicating the some sequence surrounding the locus. This would be enough sequence to provide the template for the left border of the intended operon while the operon was not actually assembled.

Transduction experiments were done to determine the genetic linkage of the gentamicin resistance gene, near the left border, and the hygromycin resistance gene, near the left border. The experiments were done with the JRT16-JRT10 and JRT15-JRT12 RS clones. The JRT15-JRT12 experiment yielded no cotransductants for the 33 colonies that were transductants. The JRT16-JRT10 clone yielded 2 cotransductants of 34 transductants, giving a cotransduction frequency of less than 10 %. This cotransduction frequency corresponds to a distance of around 85-95 kb between the two markers (Charles, 1990). It is clear from these transduction experiments that the PHA operon was not assembled as intended by RS. The expected distance between the two markers is approximately 7 kb, which would give a cotransduction frequency of approximately 90 %. The transduction experiments supported the hypothesis of random integration post-conjugation.

Because the PCR results implied the correct assembly of the PHA operons, despite the absence of any reporter gene phenotypes, GC-MS analysis was done on the RS clones. In one clone (RS JRT16-JRT10) some polyhydroxyoctanoate (PHO), an mcl-

PHA, was detected (Figure 5.11). This result needs to be replicated (technical issues with the instrument arose) to make certain that the detected peak was not merely contamination from previous experiments. Though, it is unlikely that this result is due to contamination from other samples, because the peak in question did not appear in the control samples (data not shown), and the smaller peak, present in the pure polymer sample, was not present in the RS JRT16-JRT10 sample. However, it is curious that there should be any polymer made, because it is unknown how the *phaC* is expressed. Although, it is plausible that *phaC* would be expressed through transcription read-through from adjacent genes. The same experiment was tried with no IPTG in the media, and clones containing JRT12 (*pha*C1400) did show a small PHB peak (data not shown). It is unclear why this small PHB peak was not present in the experiment with IPTG media. The Rm1021 control gave a very large PHB peak (data not shown). The PHO peak was not present in the experiment that excluded IPTG (data not shown). Because of the Nile red fluorescence phenotype of the *phaG*-containing strains, the GC-MS chromatograms of the empty LP-strain and JRT15 and JRT16 IMCE were compared. There were no peaks that appeared in the two IMCE strains' chromatograms that did not appear in the empty LP-strain chromatogram. There was some difference in the intensity of certain peaks between strains, but there was no discernible pattern.

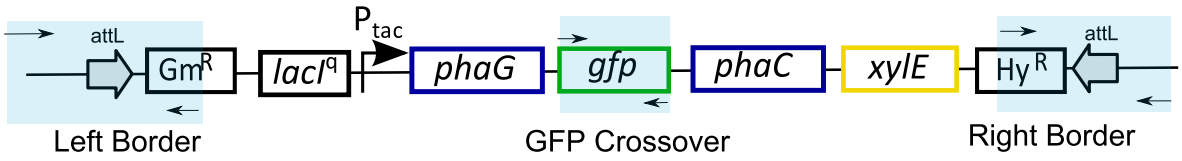

Figure 5.7: PCR evaluation of RS strains. The location of the amplicons that were targeted in colony PCR experiments to ascertain the genotype of the colony. The map is not to scale. A blue box indicates the amplicon. The amplicon is named below the blue box.

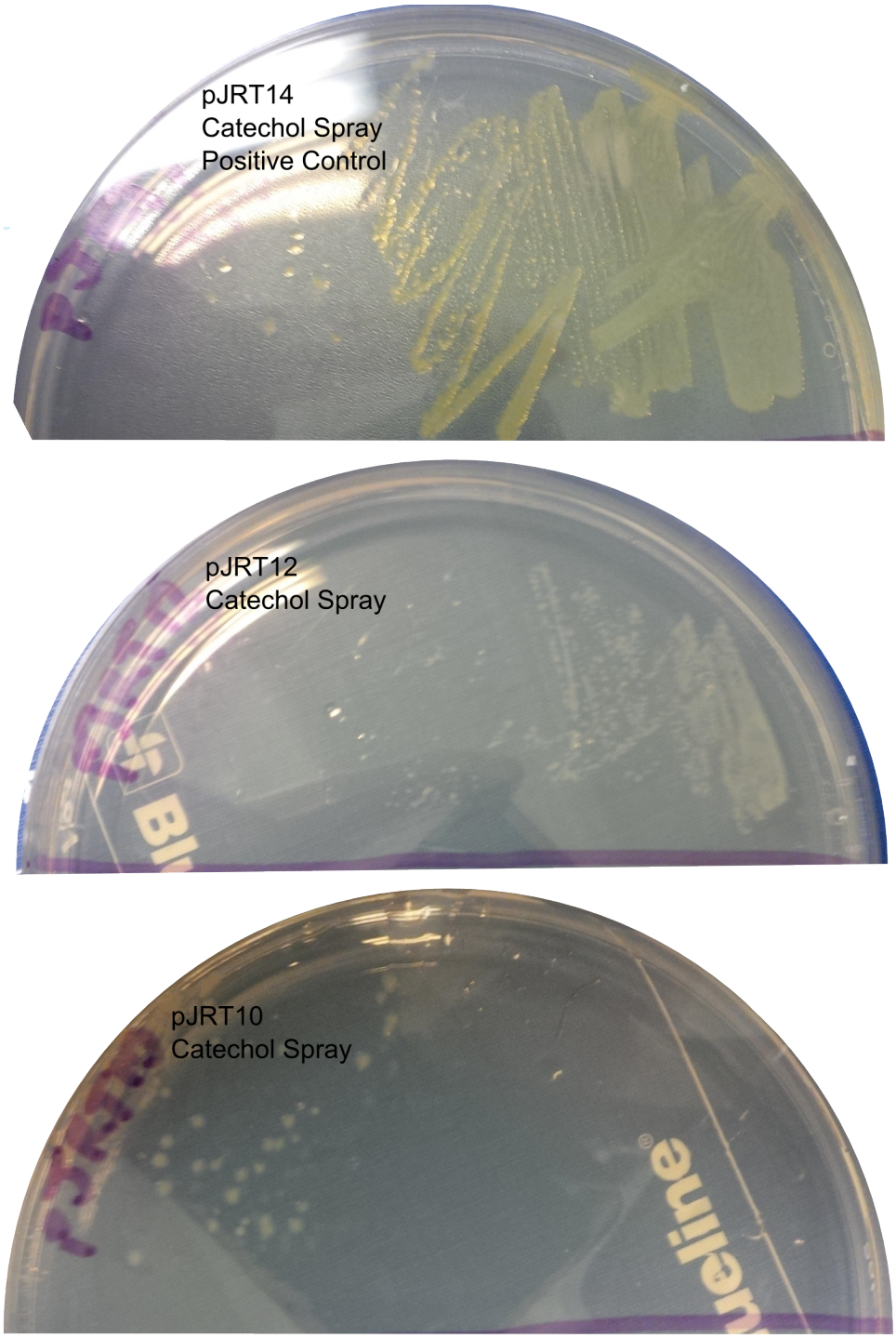

Figure 5.8: XylE reporter test. The colonies on the streak plates were sprayed with 0.5 mM catechol. The expected yellow colour developed on the pJRT14 colonies. It did not develop on pJRT10 and pJRT12 which are derivatives of pJRT14. The pictures were taken on a blue backgroud to accentuate the yellow colour.

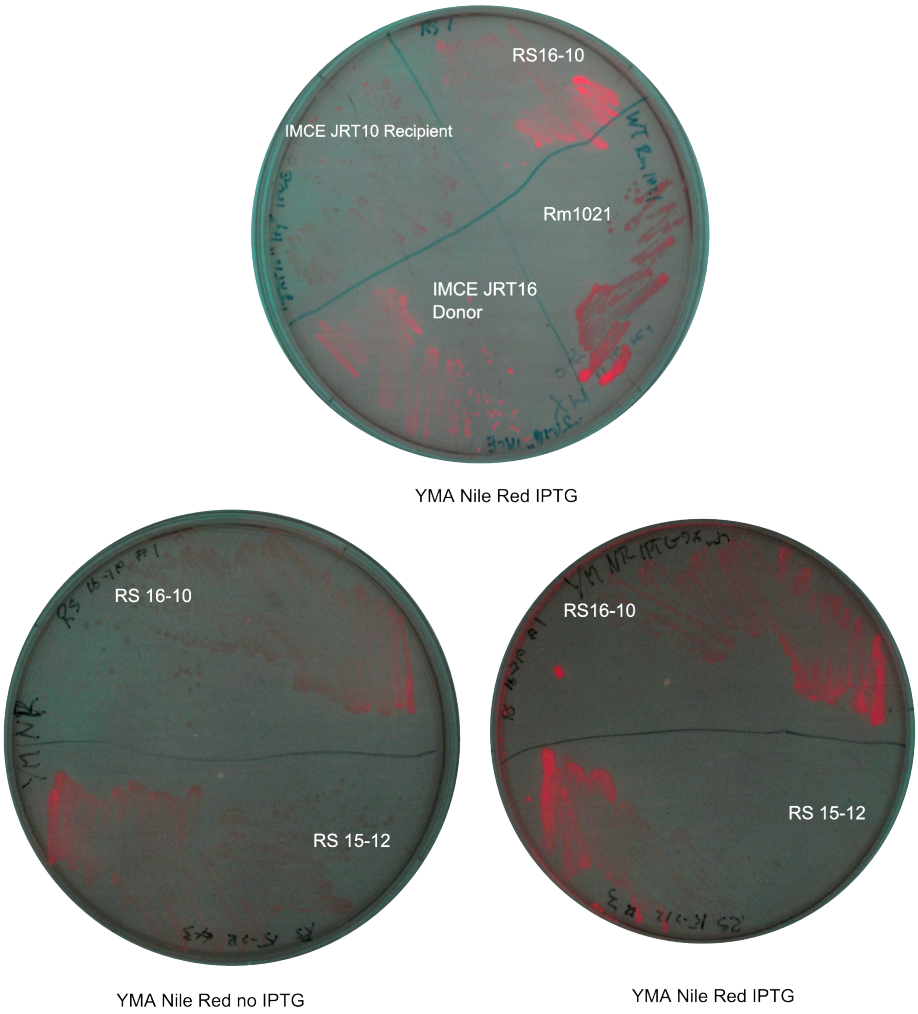

Figure 5.9: Nile red fluorescence in RS strains. The IMCE and select RS strains streaked on YMA Nile Red plates, with and without IPTG. The plate on the top panel shows that the fluorescent phenotype is transferred into the RS strain by the donor, since the recipient did not display the phenotype. The bottom two plates show the increase in fluorescence intensity when IPTG is added versus when it is not added.

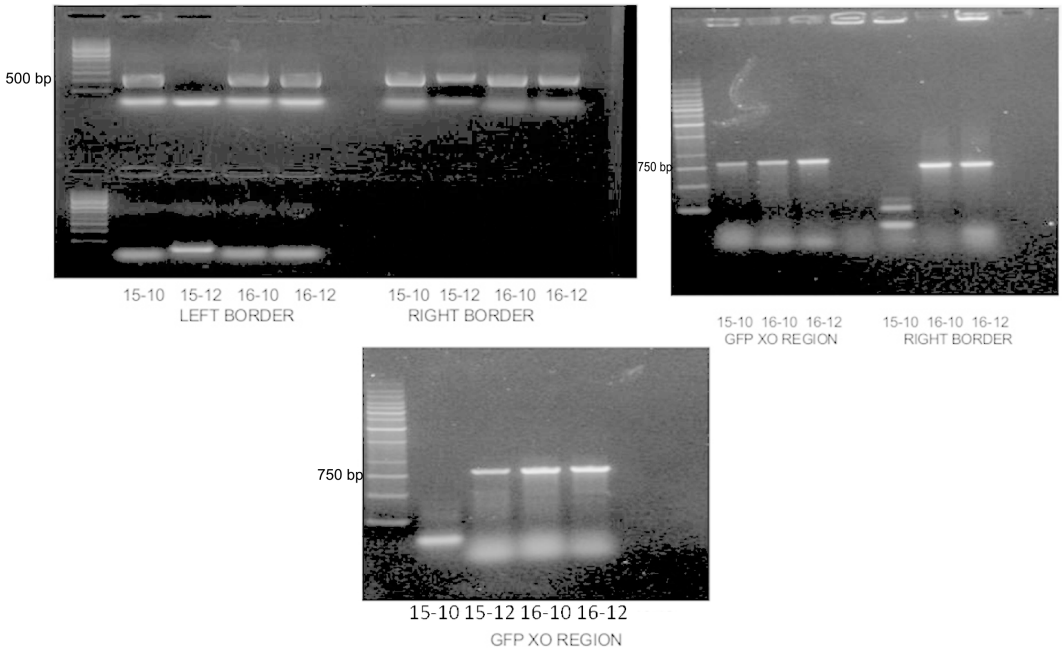

Figure 5.10: RS PCR results. The labels refer to the top row on the top right gel photograph. The other gels contain 1 row which the lane labels refer to. The expected size of the left border is 600 bp, the right border is 650 bp and the gfp XO region is 700 bp. The molecular weight standard used was the Fisher GeneRuler 1 kb ladder. The band sizes are from the bottom up in bp: 250, 500, 750, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 5000, 6000, 8000, 10000. The PCR produced all amplicons successfully for all clones except for 15-12 template with the right border primers. Primer dimers can be seen in all lanes as the smallest band.

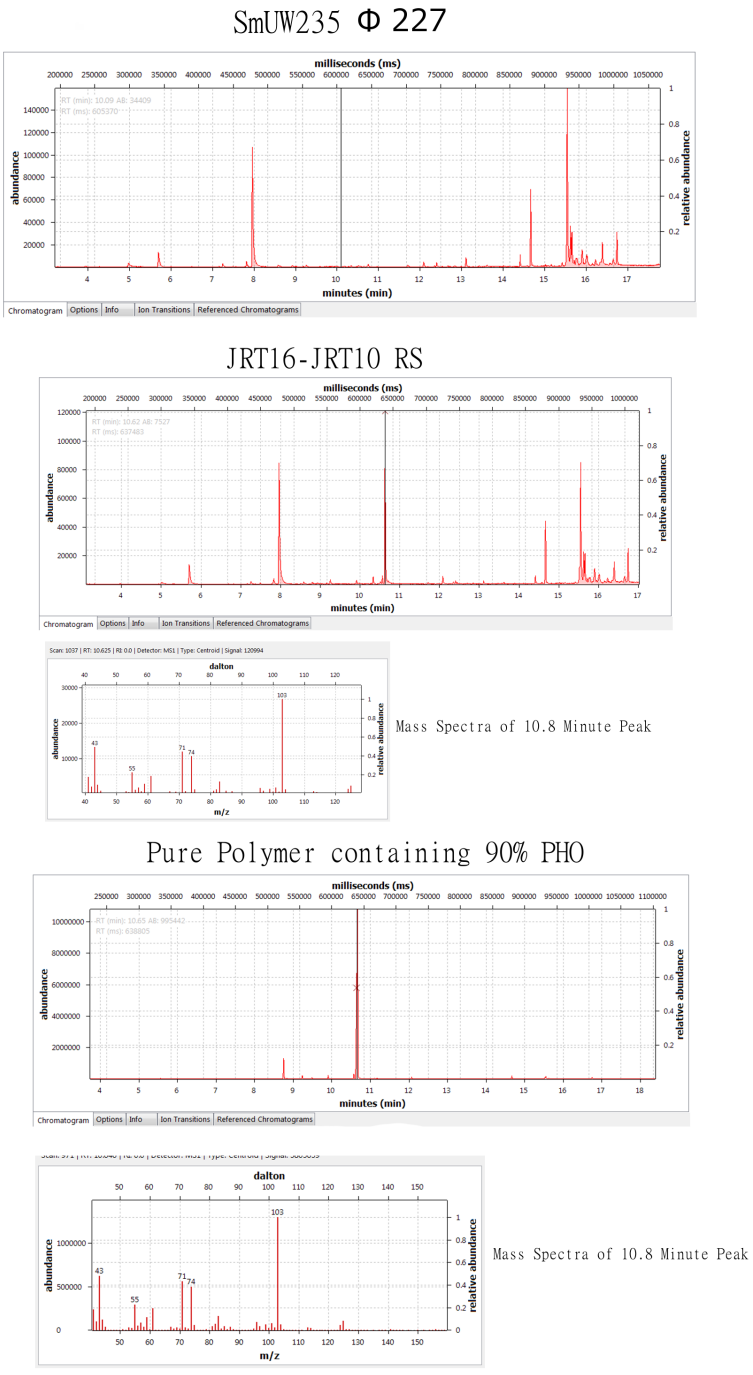

Figure 5.11: GC-MS results. The red trace line is the chromatogram. The black vertical line is the cursor. The empty LP-strain shows no peak at 10.8 minutes, while the RS strain does show a 10.8 minute peak at a height of 120 000 (partially overlapped by black cursor line). The peak in the RS strain matches PHO from pure polymer. The IMCE strains did not contain the 10.8 minute peak either (data not shown).

5.3.3 *In vitro* Assembly of PHA Operons

After it was apparent that RS had not assembled the operons correctly, another strategy was used to assemble the operons prior to IMCE. Because there was already an overlapping section of *gfp* contained on both the *phaG* and *phaC* half operon clones, a two step assembly PCR method was chosen. The PCR products were the proper length. The products were blunt end cloned into pJET. A clone for three of four operons was isolated. For one of the configurations three clones were saved. The configuration of a single operon was the primary focus because others were more difficult to subclone due to internal SphI sites. The chosen operon contained the UW4 *phaG* and GPo1 *phaC* genes. One of the three clones displayed GFP fluorescence. A first round of Sanger sequencing was done using vector primers and primers internal to *gfp* the read outwards in both directions. This sequencing revealed that all three clones contained all the expected genes in the correct order. Therefore, the PCR to assemble the PHA operon did completely assemble the operon as intended. These results rule out that there was a flaw in the sequence of the half operons that caused the RS strains to not be properly assembled. The green fluorescent colony that was isolated was called pJH150, the other two were called pJH148 and pJH149. The clone that was eventually subcloned was pJH148. Only pJH150 displayed green fluorescence (Figure 5.12), regardless of the addition of IPTG. More sequencing reads were acquired for pJH149 that showed a mutation in pJH149, which changed a single amino acid in the *gfp* sequence. Not enough sequencing coverage was gained to rule out the possibility of mutations in all of the genes. The pJH148 insert was subcloned. The pJH148 clone insert was subcloned into pJH110. This subclone (pJH151) was used in IMCE to integrate the assembled operon into two deletion mutant LP-strains: a *phbC* mutant and *phbC*, *phbA*, *phbB*, *phaZ* mutant. These IMCE strains did gain some slight Nile red fluorescence when compared to the empty LP-strain backgrounds, although it was not

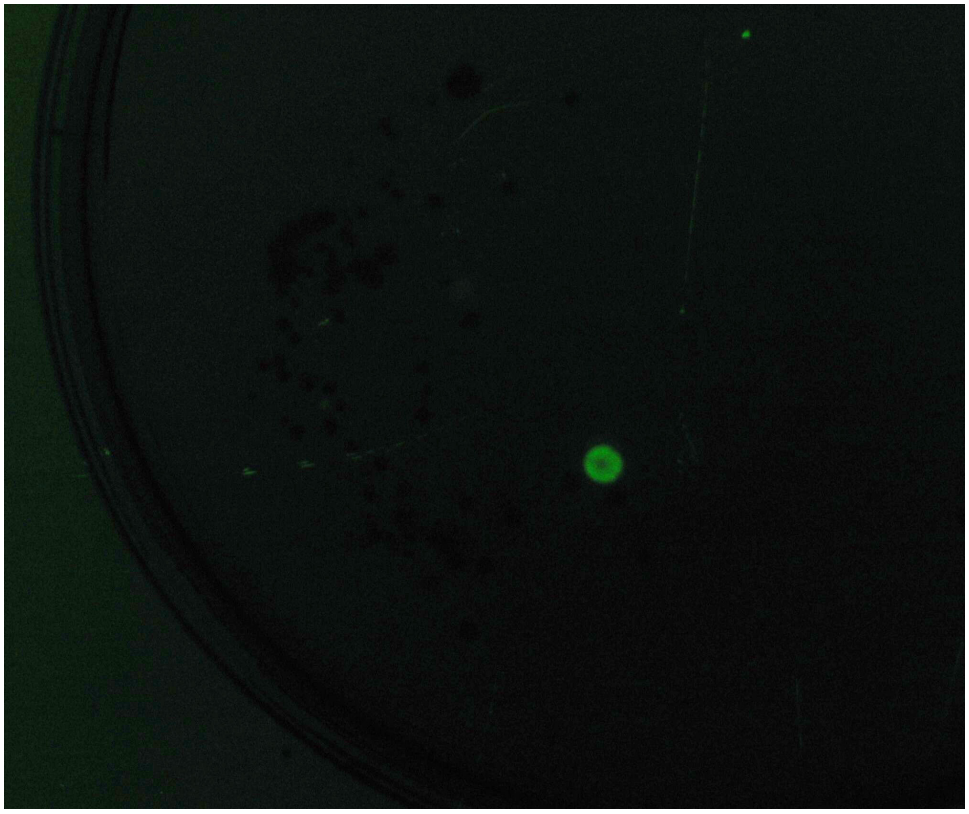

Figure 5.12: Colony displaying GFP fluorescence. This is the original selection plate for the cloning of the assembled PHA operon PCR product. A single colony displaying GFP fluorescence was detected. This clone was saved as pJH150.

as strong as the fluorescence observed with the previous clones containing *phaG*. Both IMCE clones were run on GC-MS and there were no fatty acid compounds present that were not present in the empty LP-strains. No GFP fluorescence was seen in these IMCE strains. It would be worthwhile to isolate and verify a mutation free clone, and look for PHA production again.

5.3.4 IMCE strains Containing Plasmids

The expression and function of *phaG* was apparent from the Nile red fluorescence phenotype of the *phaG* IMCE strains. Therefore, the addition of the *phaC* genes on plasmids was completed. The plasmids had already been constructed by Ricardo Nordeste and Jiujun Cheng. Ricardo had seen some promising GC-MS results using the IMCE strain containing KT2440 *phaG* and a KT2440 *phaC* clone on pRK7813 (constructed by Jiujun Cheng)(personal communication). The GPo1 *phaC* had been subloned into pTH1227 by Ricardo. I moved that plasmid into the JRT16 IMCE strain containing UW4 *phaG*. There are many more *phaC* clones that have been isolated by Jiujun Cheng that have been moved into the *phaG* containing IMCE clones.

5.4 Discussion

Nile Red Phenotype of *phaG* IMCE Clones

It was intriguing that the strains containing *phaG* displayed fluorescence when plated on YM agar containing Nile red. This phenotype was intensified by the addition of IPTG, making it clear that the integrated *phaG* construct was responsible. Nile red is a fluorescent dye that binds to lipids in general. It is used as a dye to visualize intracellular lipid droplets (Greenspan *et al.*, 1985). PHA granules are in a sense lipid droplets. That makes Nile red a good choice for visualizing PHA. However, there can

be no PHA granules, in the case of the IMCE strains containing only *phaG* that lack a polymerase gene (*phaC*).

When the native PHA polymerase gene of *Sinorhizobium meliloti* (*phbC*: a more specific name for *phaC*) is deleted, the Nile red fluorescence of the wild type strain is abolished. The wild type pathway for short-chain PHA production produces PHB through the polymerization of 3-hydroxybutyryl-CoA. This monomer substrate is produced starting from acetyl-CoA, by the activity of β-ketothiolase (PhbA) and acetoacetyl-CoA reductase (PhbB) (Tombolini *et al.*, 1995). In the case of a *phbC* deletion mutant, the monomer substrate is likely still produced. It is likely broken down back into acetyl-CoA through β-oxidation. In the mcl-PHA pathway, that this work aimed to reconstruct, the monomers are 3-hydroxyacyl-CoAs derived from *de novo* fatty acid synthesis through the activity of PhaG. These 3-hydroxyacyl-CoAs are of a longer chain length, and they are also intermediates in the β-oxidation pathway. These 3-hydroxyacyl-CoAs may enter the β-oxidation cycle midway through, and would be catabolized into a shorter acyl-CoA, and one acetyl-CoA. The shortened acyl-CoA would re-enter the pathway and be shortened once more. Longer chain fatty acids would require more cycles to be fully catabolized. It is therefore plausible that, by the addition of *phaG*, fatty acids are being shunted into β-oxidation causing the cells to accumulate lipid droplets in the form of acyl-CoAs.

The β-oxidation build up hypothesis is also supported by the fact that there was no difference seen between the empty LP-strain and the *phaG* containing strain when GC-MS was used to analyse fatty acid methyl esters (FAMEs). Yet the difference in Nile red fluorescence was apparent. The chemical reaction used to derivatize lipids in the cell pellet is essentially the same reaction used to make biodiesel: an acid catalyzed transesterification reaction, using methanol. It breaks the ester linkage and leaves a methyl group, producing methyl-esters.This reaction works well for detecting

PHAs, because PHAs are polyesters. The GC-MS is detecting FAMEs, and anything not derivatized will not be detected. In Acyl-CoAs the carbon chain is linked to the CoA molecule through a thioester linkage. There are derivatization methods for the anlysis of acyl-CoA thioesters (Tamvakopoulos and Anderson, 1992), but the method used in this work will not allow them to be detected.

RS Transfer

In the case of the transduction mediated genetic cross there is a limited but large amount of DNA contained in the transducing particle, approximately 160 kb (Finan *et al.*, 1984). There is a large amount of homologous DNA that can recombine into the recipient genome. In RS, a recombination event at the *gfp* crossover region was required. This region is only a small portion of all the homologous DNA available. It is therefore very likely that the recombination will happen elsewhere, which would give rise to a transductant cell resistant to the single antibiotic corresponding to the gene carried by the transducing particle. Only the recombination event that occurred at the *gfp* region would yield a transductant resistant to both antibiotics. Because the frequencey of transduction was low, it was unlikely an RS transductant would be among the transductants. After the failure of the RS transductions, conjugation was chosen as a method to accomplish the genetic cross that the RS method depended on.

I thought that placing an *oriT* right next to the *gfp* crossover region and using conjugation to specifically transfer that DNA would increase the likelihood of a cross-over event occurring within the *gfp* region. Colonies that were resistant to both antibiotics were successfully obtained. The Nile red fluorescence phenotype of *phaG*, from the donor strain, was transferred. Moreover, there was evidence of the integration of the donor DNA from PCRs, and it seemed to be in the right place. It was revealed by the direct genomic Sanger sequencing that the donor DNA was not integrated in the

correct place. To find out exactly where it was integrated was not trivial, so it was not attempted. The PCR results taken together with the direct genomic sequencing results suggest that the donor DNA integrated randomly. The donor DNA also included a segment of the adjacent chromosomal DNA, as one would expect, some sequence adjacent to the landing pad locus was duplicated. It would seem that integration of the donor DNA at the intended location would have been more likely than a random integration of the donor DNA, but there could be some mechanisms in place that may explain the outcome. It has been shown that conjugation causes the induction of the SOS response in *Escherichia coli* and *Vibrio cholerae* (Baharoglu *et al.*, 2010). The SOS response causes the activation of a host of genes involved in recombination and DNA replication and repair. Error prone DNA polymerases are expressed, so the replication fork may pass through damaged DNA regions. It has been shown that the SOS response can cause an increase in duplication of sequence (Dimpfl and Echols, 1989). In relation to this issue, horizontal gene transfer has been associated with an increase in rearrangements of the transferred genes (Hao and Golding, 2009). Perhaps some of these effects were significant enough in *S. meliloti* to cause the donor DNA to integrate the way it did, instead of the, ostensibly more likely, intended location.

PHO production in RS strain

Even though RS did not accomplish the integration of the donor DNA in the correct location, the *phaG* Nile red phenotype was transferred, indicationg that the donor DNA was indeed functioning. The recipients contained the *phaC* gene already integrated at the landing pad locus. The *phaC* is promoterless, but it is possible that a small amount of transcription could be happening. To make a detectable amount of PHA, a large amount of PhaC enzyme should not be required. It is possible that the result showing the PHO production was due to contamination of the sample with the standard

pure polymer, though the accompanying peak present in the standard is not present in the RS sample. In any case the strain itself is not easily described, because the location of the *phaG* insertion is unknown. More work is needed to construct a strain where the PHA genes are surely expressed, and the detection of PHO is repeatable. The detection of PHO in the RS strain does offer some promise of success.

PCR False Positive

PCR is now commonly used as a method to detect physical changes in genomes. When compared with Southern blot analysis, PCR is much more rapid and easily accomplished. One potential problem of using PCR analysis is the amplification of non-specific products. However, this problem is easily circumvented by sequencing of PCR products, and also simple verification that the PCR products are the correct size. In the case of the RS strains, initially PCR was used to evaluate the physical makeup of the chromosome at the locus of the putative assembly. The PCR product for the *gfp* junction, where the amplicon straddled the truncation in the initial *gfp* pieces, was of the correct size for full-length *gfp*. Sequencing of the *gfp* amplicon gave the full intact sequence of un-truncated *gfp*. At the time this result seemed like very strong evidence for the correct assembly of the PHA half operons by RS. Later it was revealed by direct genomic Sanger sequencing that this was not the case. The direct genomic sequencing was done because the expected GFP fluorescence was absent in the RS strains, yet the sequence of the PCR product showed full length *gfp* without one mutation. I was aware that the PCR product may have assembled *in vitro* if the overlapping truncated templates were available. The overlapping DNA would serve as a primer for its overlapping counterpart. Direct genomic sequencing was chosen, because it was not susceptible to being confounded by the presence of overlapping sequence. The sequencing revealed that the truncated *gfp* remained truncated. The successful

PCR amplification must have been successfully amplified due to *in vitro*assembly of the overlapping parts. It is important to take this possibility into consideration when designing PCR experiments to evaluate the physical arrangement of the chromosome.

PHA Metabolic Engineering Future Work

RS was successful in transferring the genes of the donor into the recipient despite the fact that they did not integrate in the precise location to cause the desired assembly of the full operon. Though the RS strains were not the desired strains, the GC-MS results showing the detection of PHO in the JRT16-JRT10 RS strain are encouraging. When *phaC* is supplied on a plasmid the need for RS is circumvented, and further analysis of the strains containing *phaC* on a plasmid is needed.

Chapter 6

General Discussion

Brief Summary of This Work

This work is centered on the development of chromosome engineering tools for general use in a wide range of bacteria. In this thesis, the work has been confined to three members of the *Rhizobiales*: *S. meliloti*, *Agrobacterium tumefaciens*, and *Ochrobactrum anthropi*. Because of the demonstrated functionality of ΦC31 integrase *in vitro* and across kingdoms of life, (Groth *et al.*, 2000; Thomason *et al.*, 2001; Bateman and Wu, 2008; Kushtoss and Rao, 1991), and the broad host range of RK2 conjugation (Thomas, 1981), the methods developed here are broadly applicable to many diverse bacterial species. The work is also focused on the more specific goal of expanding the types of PHAs that *S. meliloti* and other members of the *Rhizobiales* can make.

The integrase mediated cassette exchange (IMCE) system and its variations (rIMCE, SacB IMCE) were successfully developed and they have been shown to be very reliable. However, the recombination synthesis (RS) system to assemble already integrated overlapping DNA segments proved to be more challenging. Alternatives to RS that have been developed in recent years will be discussed in this context. The PHA operons were successfully assembled through more conventional means, because they

were not quite big enough to preclude the use of traditional cloning techniques. With one of these operons, *S. meliloti* did make medium chain length polyhydroxy alkanoates (mcl-PHA).

More specific discussions of the content of the previous chapters: 3, 4, and 5 are present at the end of those chapters. This discussion aims to integrate the results of the chapters to inform the future application of the contributions made by this thesis.

The Chromosome Engineering Strategy: Comparison to Other Strategies

The overall goal of the chromosome engineering system was to allow for the assembly of large synthetic DNA constructs on the chromosome. The end result being a heavily modified strain of the host bacteria with a large volume of integrated DNA. As a proof of concept of the methods that were conceived to do this, a smaller amount of DNA that would be functional was to be assembled. In the end, it is irrelevant whether the DNA is assembled after it is introduced into the chromosome or whether it is assembled first and integrated later. The rationale behind the design of the system in this thesis is that it is difficult to manipulate and clone large pieces of DNA before integration. Therefore, the system described in this thesis first uses IMCE to reliably integrate the DNA into a common locus, and then uses RS to take the already integrated overlapping fragments and assemble them through a genetic cross. The DNA pieces have to be designed carefully before IMCE to ensure proper conditions for RS to take place.

Around the time this work was started, there had been the announcement by the Venter Institute of a synthetic genome (Gibson *et al.*, 2008a) where most of the assembly was done *in vitro* but the final assembly was done by homologous recombination in the yeast *Saccharomyces cerevisiae*. Later the *S. cerevisiae* assembly method was pushed to its full potential (Gibson *et al.*, 2008b), so that the whole genome of *Mycoplasma genitalium* was assembled in yeast in one step. Though what was coming out of the Venter

Institute was certainly impressive, for DNA constructs significantly shorter than a fully synthetic cell, it seemed that the RS-based method had a niche to fill. In the end we desired bacterial strains containing the engineered DNA: not a fully synthetic cell. Members of the George Church group developed conjugation methods similar to RS to do chromosome engineering (Isaacs *et al.*, 2011), a method called MAGE/CAGE. In MAGE/CAGE the stop codons of every gene in *E. coli* were changed by first making the mutations with λ-Red recombination, and then combining all those mutants into one strain by conjugating overlapping segments of the genome into another mutant. The underlying theory of the conjugation step is very similar to RS. The main difference between them was that, MAGE/CAGE was used to combine many point mutations into a single strain, whereas RS would be used to combine many medium to large insertions at a single locus.

In theory, the advantage of a method similar to RS, is that the assembly is done in the desired host, therefore the transfer to a new host and the establishment of a large piece of DNA in that host is not necessary. The Gibson yeast methods do require the transfer of the finally constructed DNA to its desired host. The DNA transfer method was developed using natural DNA (Lartigue *et al.*, 2007) and then applied to a synthetic *Mycoplasma mycoides* genome some time later (Gibson *et al.*, 2010). This method is similar in concept to a $CaCl_2$ transformation but with specialized gentle methods for handling the large DNA. It was developed specifically for *Mycoplasma capricolum*. If RS had worked well, it would have maintained the advantage of the DNA transfer happening during the assembly process itself, and the underlying methods of RS are already developed for most proteobacteria. Unlike the transformation based method for installing the synthetic DNA into the intended host, a conjugation based method does not require a host that is, restriction enzyme deficient and recombination deficient. The underlying methods used by the Venter Institute were new and being

developed for *M. capricolum* and *M. mycoides*, and they had to solve the restriction and recombination problems.

In the years since the invention of the yeast based assembly, other genomes have been assembled and cloned in yeast (Karas *et al.*, 2013b,a; Noskov *et al.*, 2012). The technical problems that were solved concern the toxicity of certain genes to yeast, and replication problems due to high GC content causing there to be a reduced frequency of randomly occurring yeast origins of replication. Still the focus has been mostly on assembling and cloning the genomes in yeast, not the installation of the synthetic DNA into the intended host. However, the yeast assembly method is proving to be very powerful indeed. It is conceivable that an assembly could be made in yeast that could be a substrate for IMCE. It would be transferred into the landing pad strain (LP-strain) from yeast, probably through a transformation based method similar to the method described by Lartigue *et al.* (2007). The disadvantage being that in addition to developing an LP-strain, a transformation method would also have to be developed. Conjugation is a much more general method for a variety of reasons. One reason is that during conjugation a single strand of DNA is used as a template for DNA synthesis in the recipient strain during transfer. In many cases this reduces the barrier posed by restriction modification systems, because the incoming DNA is sufficiently modified to avoid restriction (Langella and Chopin, 1989). This protection is not afforded to the naked DNA taken up during transformation. To address this problem in the context of transformation, strains devoid of restriction modification systems must be developed, or the donor DNA must be incubated with appropriate modification enzymes (Gibson *et al.*, 2010). The development of transformation methods is therefore non-trivial. To circumvent the need to develop transformation methods, the yeast assembly could be transformed into *E. coli* and then conjugated into the LP-strain through a regular IMCE protocol. For sub-genome sized assemblies this may

work, and the *in vitro* DNA handling requirements would be less stringent, than for full genomes.

Although I have established the RS methodology, it still requires some refinement for reliable and routine use. With future investigation, RS could possibly be improved so that it works reliably. Given the acceleration of the development of the yeast based assembly methods, and the apparent power of those methods (assembly of tens or hundreds of fragments at once), it would probably be a better investment to develop the link between the yeast based assembly and IMCE. That way sub-genome sized constructs could quickly be assembled and integrated. That type of system would allow for quick prototyping or screening of constructs for the intended metabolic engineering, or synthetic biology application, when a whole new synthetic genome is more than what is needed.

A method that could be used for chromosome engineering in the intended host is inchworm elongation (Itaya *et al.*, 2008). Essentially this method relies on homologous recombination to sequentially assemble a chromosomal insert. At the end of every insert is inserted homologous DNA to act as a substrate for the next insertion by homologous recombination. In this respect it is somewhat similar in concept to RS. RS, in its original conception, would be an iterative process as well. At each iteration the antibiotic resistance gene at the growing end of the construct that selects for the RS transconjugants would be alternated. Three markers would be required, one at the stationary end, and two for alternate use at the growing end. It is conceivable that the RS process could be used in a construction tree arrangement where with each iteration the construct grows exponentially larger. However, with the limitations of the RS system, even if it had proved reliable, this would be impractical. A large number of selectable markers would be needed to avoid conflicts in the construction tree, unless a very careful design, avoided the pairing of growing constructs flanked by the same

markers. In any case, during this project, only enough markers were available for the pure iterative process, and this thesis describes the attempt at the first assembly iteration. The advantage of combining IMCE coupled with yeast DNA assembly is that the whole construct could be made at once and integrated into the intended host. The assembly in yeast works well enough that multiple iterations are not necessary. The all-at-once yeast approach may turn out to be more rapid than an iterative approach.

The application of RS that was the focus of this thesis is its ability to create constructs in a combinatorial way. Such was the case with the various PHA operons, each containing a different combination of *phaG*, and *phaC* orthologues, with the same overlapping homology between them. In this thesis, the number of possible combinations was small, so the different combinations were made deliberately. If the number of combinations was large, such as combining two libraries of hundreds of *phaC* and *phaG* genes, the combinations would not be deliberate, but random. Sufficient numbers of colonies would have to be screened to ensure a reasonable probability that the combinatorial space was adequately covered. It would be important to have an efficient system to be able to deliver the number of colonies required. In this example (PHA production), it would also be important to employ an efficient phenotypic screen, since using GC-MS on that number of colonies would be impractical. However, that is a separate issue. The yeast system has the potential for use in combinatorial assembly in this way (Gibson *et al.*, 2008b). Judging from its ability to take up and assemble many DNA fragments. Since the assembly principle in yeast is also based on DNA homology, like in RS, a slight modification of the existing RS constructs to add homology to a yeast vector on the outside ends would be all that is needed to adapt the system. This would be all that is needed after getting all the other pieces working, and successfully linking the yeast assembly system to IMCE.

PHA Production: Why Try Different Combinations of Genes?

The plastics industry produces a staggering array of different materials with various physical properties. There are myriad physical properties that can be measured: tensile strength, stiffness, Young's modulus (a measure of the force needed to stretch or compress a material), deformation before breaking, glass transition temperature, melting temperature, density, and more. A material with certain properties makes it well suited for a certain application. In the case of bioplastics, expanding the diversity of the materials available will serve to increase the diversity of applications suited to their use. These physical properties are influenced by the molecular structure of the polymer and packing of that molecular structure in crystalline or amorphous ways (van Krevelen and te Nijenhuis, 2009). Even two polymers with the exact same monomer composition can vary in physical properties. For example, polypropylene may be isotactic, syndiotactic, or are atactic. Tacticity refers to the orientation of the methyl side groups on the polymer. An isotactic polymer has all the methyl groups in the same orientation, syndiotactic alternates the orientation. Atactic polymers have random orientations of side-chains. The tacticity of the polymer affects its physical properties (Jones *et al.*, 2002). Natural PHB is exclusively isotactic, having all of its monomers in the R stereochemical orientation (Timmins *et al.*, 1996). So this is not a factor in the structure of natural PHAs, but perhaps there is a novel PHA synthase out there in the world that would incorporate S-3-hydroxyalkanoates as well as the R-3-hydroxyalkanoates. Even though, this point does not concern as yet discovered natural PHAs it does illustrate how the finer aspects of polymer structure can affect the physical properties of the material. PHAs are a polyester with 3 carbons between linkages, the rest of the carbons make up the side chain. PHB is a 4 carbon monomer so has a methyl group as the side chain. Mcl-PHAs have longer side chains. Any polymer containing monomers longer than 5 carbons is considered an mcl-PHA. Var-

ious homopolymers and co-polymers containing scl monomers, mcl monomers, or a mixture of scl and mcl monomers would cover a range of physical properties. The arrangement of the monomers in a co-polymer also matters. There are random copolymers and block copolymers. Many examples of PHA copolymers in the literature are random copolymers (Kato *et al.*, 1996).

This thesis discussed the engineering of *S. meliloti* to produce mcl-PHA (6-12 carbons long), where *S. meliloti* naturally produces scl-PHA (3-5 carbons long). To my knowledge, mcl-PHAs have not been produced in bacteria other than *Pseudomonas*. The activity of the *phaG* should provide mcl substrates for polymerization. The chain length of the monomer substrate should depend on the underlying metabolism of the cell and the substrate specificity of the *phaG*. On top of that, the substrate specificity of the *phaC* would determine which of the available monomers get incorporated into the polymer. There have been examples of PHAs that are mcl-scl copolymers reported (Kato *et al.*, 1996). In this case, the *phaC* had a broad substrate specificity and the strain in which it was operating was producing a broad range of monomers appropriate for incorporation. It could even be possible that a block copolymer would be made given the right constellation of substrate selectivity of either enzyme, where it would exhaust one substrate before moving onto the other substrate of a different chain length. Given all of the interacting factors that would affect the final physical properties of the bioplastic, it makes sense that a combinatorial approach would be worthwhile. A rapid screening method to differentiate between colonies containing PHAs varying in chain length would be a great complement to the genetic engineering techniques that allow for the generation of a combinatorial library. The Nile red screen method published by Wu *et al.* (2003) does allow the differentiation of mcl and scl PHA by identifying a shift in the Nile red fluorescence emission wavelength. A refinement of this screening method or something similar, the ability to create the combinatorial li-

brary of PHA producing strains, and the discovery of new PHA metabolism genes all taken together could provide a powerful system for novel PHA discovery.

Concluding Statement

The focus of this thesis work was on developing the chromosome engineering systems of IMCE and RS. The secondary focus was applying these systems to metabolic engineering to produce mcl-PHAs in *S. meliloti*. The IMCE system worked very well and turned out to be very robust. The RS system was implemented, yet the donor DNA did not integrate in the intended location. In the future, it makes sense to integrate IMCE into other emerging and proven chromosome engineering methods, such as assembly in yeast, rather than troubleshooting RS. The IMCE system provides a very robust way of creating a marker free strain with stably integrated genes.

Some promising yet preliminary results were obtained showing mcl-PHA production in *S. meliloti*. Using plasmids to express the *phaC* genes for the purpose of screening and prototyping in IMCE strains containing *phaG* should be more productive than attempting to troubleshoot RS. In the future the IMCE strains containing *phaG* will provide a useful test-bed for any plasmids expressing *phaC*. The use of the *phaG* IMCE strains to test *phaC* clones isolated from the environment may prove fruitful in the search for interesting PHAs.

Bibliography

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402.

Anderson, A. J. and E. A. Dawes, 1990. Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. Microbiol Rev **54**:450–472.

Aneja, P., A. Zachertowska, and T. C. Charles, 2005. Comparison of the symbiotic and competition phenotypes of *Sinorhizobium meliloti* PHB synthesis and degradation pathway mutants. Can J Microbiol **51**:599–604.

Arciszewska, L. K., D. Drake, and N. L. Craig, 1989. Transposon Tn7. J Mol Biol **207**:35–52.

Baharoglu, Z., D. Bikard, and D. Mazel, 2010. Conjugative DNA transfer induces the bacterial SOS response and promotes antibiotic resistance development through integron activation. PLoS Genet. **6**:e1001165.

Baker, D., G. Church, J. Collins, D. Endy, J. Jacobson, J. Keasling, P. Modrich, C. Smolke, and R. Weiss, 2006. Engineering life: building a fab for biology. Sci Am **294**:44–51.

Banik, J. J. and S. F. Brady, 2008. Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. Proc Natl Acad Sci USA **105**:17273–7.

Barnett, M. J., V. Oke, and S. R. Long, 2000. New genetic tools for use in the *Rhizobiaceae* and other bacteria. Biotechniques **29**:240–242,244–245.

Bateman, J. R. and C. T. Wu, 2008. A simple polymerase chain reaction-based method for the construction of recombinase-mediated cassette exchange donor vectors. Genetics **180**:1763–1766.

Beringer, J. E., 1974. R factor transfer in *Rhizobium leguminosarum*. J Gen Microbiol **84**:188–98.

Bernard, P., K. E. Kezdy, L. Van Melderen, J. Steyaert, L. Wyns, M. L. Pato, P. N. Higgins, and M. Couturier, 1993. The F plasmid CcdB protein induces efficient ATP-dependent DNA cleavage by gyrase. J Mol Biol **234**:534–541.

Bhavsar, A. P., X. Zhao, and E. D. Brown, 2001. Development and characterization of a xylose-dependent system for expression of cloned genes in *Bacillus subtilis*: conditional complementation of a teichoic acid mutant. Appl Environ Microbiol **67**:403–10.

Bialowska-Hobrzanska, H. and W. J. H. Kunicki-Goldfinger, 1977. Mechanism of conjugation and recombination in bacteria XVI. MGG Mol. Gen. Genet. **151**:319–326.

Bibb, L. A. and G. F. Hatfull, 2002. Integration and excision of the *Mycobacterium tuberculosis* prophage-like element, ϕRv1. Mol Microbiol **45**:1515–1526.

Biery, M. C., M. Lopata, and N. L. Craig, 2000. A minimal system for Tn7 transposition: the transposon-encoded proteins TnsA and TnsB can execute DNA breakage and joining reactions that generate circularized Tn7 species. J Mol Biol **297**:25–37.

Blake, W. J., B. A. Chapman, A. Zindal, M. E. Lee, S. M. Lippow, and B. M. Baynes, 2010. Pairwise selection assembly for sequence-independent construction of long-length DNA. Nucl Acids Res **38**:2594–2602.

Bramucci, M. G. and V. Nagarajan, 1996. Direct selection of cloned DNA in *Bacillus subtilis* based on sucrose-induced lethality. Appl Environ Microbiol **62**:3948–3953.

Braunegg, G., B. Sonnleitner, and R. M. Lafferty, 1978. A rapid gas chromatographic method for the determination of poly-β-hydroxybutyric acid in microbial biomass. Eur J Appl Microbiol Biotechnol **6**:29–37.

Breuner, A., L. Brondsted, and K. Hammer, 1999. Novel organization of genes involved in prophage excision identified in the temperate lactococcal bacteriophage TP901-1. J Bacteriol **181**:7291–7297.

Brown, T., 2001. Southern blotting. Curr Protoc Mol Biol **Chapter 2**:Unit2.9A.

Campbell, R. E., O. Tour, A. E. Palmer, P. A. Steinbach, G. S. Baird, D. A. Zacharias, and R. Y. Tsien, 2002. A monomeric red fluorescent protein. Proc Natl Acad Sci USA **99**:7877–7882.

Carroll, D., 2011. Genome engineering with zinc-finger nucleases. Genetics **188**:773–82.

Carvalho, F. M., R. C. Souza, F. G. Barcellos, M. Hungria, and A. T. R. Vasconcelos, 2010. Genomic and evolutionary comparisons of diazotrophic and pathogenic bacteria of the order *Rhizobiales*. BMC Microbiol. **10**:37.

Cello, J., A. V. Paul, and E. Wimmer, 2002. Chemical synthesis of poliovirus cDNA: generation of infectious Virus in the absence of natural template. Science **297**:1016–1018.

Chain, P. S. G., 1998. Development of an in vivo DNA cloning procedure for bacteria. Ph.D. thesis, McMaster University.

Chain, P. S. G., D. M. Lang, D. J. Comerci, S. A. Malfatti, L. M. Vergez, M. Shin, R. A. Ugalde, E. Garcia, and M. E. Tolmasky, 2011. Genome of *Ochrobactrum anthropi* ATCC 49188 T, a versatile opportunistic pathogen and symbiont of several eukaryotic hosts. J Bacteriol **193**:4274–5.

Charles, T. C., 1990. Construction of a Genetic Linkage Map of the Rhizobium meliloti 1600 Kilobase Megaplasmid pRmeSU47b, Generation of Defined Megaplasmid Deletions, and Study of Megaplasmid-Borne Genes. Ph.D. thesis, McMaster University.

Charles, T. C. and E. W. Nester, 1993. A chromosomally encoded two-component sensory transduction system is required for virulence of *Agrobacterium tumefaciens*. J Bacteriol **175**:6614–25.

Chen, G. Q., 2009. A microbial polyhydroxyalkanoates (PHA) based bio- and materials industry. Chem Soc Rev **38**:2434–2446.

Chen, J.-Y., G. Song, and G.-Q. Chen, 2006. A lower specificity PhaC2 synthase from Pseudomonas stutzeri catalyses the production of copolyesters consisting of short-chain-length and medium-chain-length 3-hydroxyalkanoates. Antonie Van Leeuwenhoek **89**:157–167.

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

Choi, K.-H., D. DeShazer, and H. P. Schweizer, 2006a. mini-Tn7 insertion in bacteria with multiple *glmS*-linked *attTn7* sites: example *Burkholderia mallei* ATCC 23344. Nat Protoc **1**:162–9.

Choi, K.-H., J. B. Gaynor, K. G. White, C. Lopez, C. M. Bosio, R. R. Karkhoff-schweizer, and H. P. Schweizer, 2005. A Tn7-based broad-range bacterial cloning and expression system. Nat Methods **2**:443–448.

Choi, K.-H., A. Kumar, and H. P. Schweizer, 2006b. A 10-min method for preparation of highly electrocompetent *Pseudomonas aeruginosa* cells: Application for DNA fragment transfer between chromosomes and plasmid transformation. J Microbiol Methods **64**:391–397.

Choi, K.-H. and H. P. Schweizer, 2006a. mini-Tn7 insertion in bacteria with secondary, non-glmS-linked *attTn7* sites: example *Proteus mirabilis* HI4320. Nat Protoc **1**:170–8.

Choi, K.-H. and H. P. Schweizer, 2006b. Mini-Tn7 insertion in bacteria with single *attTn7* sites: example *Pseudomonas aeruginosa*. Nat Protoc **1**:153–61.

Christiansen, B., L. Brondsted, F. K. Vogensen, and K. Hammer, 1996. A resolvase-like protein is required for the site-specific integration of the temperate lactococcal bacteriophage TP901-1. J Bacteriol **178**:5164–5173.

Colloms, S. D., C. A. Merrick, F. J. Olorunniji, W. M. Stark, M. C. M. Smith, A. Osbourn, J. D. Keasling, and S. J. Rosser, 2013. Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. Nucleic Acids Res. .

Cormack, B. P., R. H. Valdivia, and S. Falkow, 1996. FACS-optimized mutants of the green fluorescent protein (GFP). Gene **173**:33–8.

Cowie, A., J. Cheng, C. D. Sibley, Y. Fong, R. Zaheer, C. L. Patten, R. M. Morton, G. B. Golding, and T. M. Finan, 2006. An integrated approach to functional genomics: construction of a novel reporter gene fusion library for *Sinorhizobium meliloti*. Appl Environ Microbiol **72**:7156–67.

Craig, J. W., F.-Y. Chang, J. H. Kim, S. C. Obiajulu, and S. F. Brady, 2010. Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. Appl Environ Microbiol **76**:1633–41.

Crellin, P. K. and J. I. Rood, 1997. The resolvase/invertase domain of the site-specific recombinase TnpX is functional and recognizes a target sequence that resembles the junction of the circular form of the *Clostridium perfringens* transposon Tn4451. J Bacteriol **179**:5148–5156.

Dafhnis-Calas, F., Z. Xu, S. Haines, S. K. Malla, M. C. M. Smith, and W. R. A. Brown, 2005. Iterative *in vivo* assembly of large and complex transgenes by combining the activities of ΦC31 integrase and Cre recombinase. Nucleic Acids Res. **33**:e189.

Datsenko, K. A. and B. L. Wanner, 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci USA **97**:6640–6645.

de Lorenzo, V., M. Herrero, U. Jakubzik, and K. N. Timmis, 1990. Mini-Tn5 transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in Gram-negative eubacteria. J Bacteriol **172**:6568–6572.

del Solar, G., R. Giraldo, M. J. Ruiz-Echevarria, M. Espinosa, and R. Diaz-Orejas, 1998. Replication and control of circular bacterial plasmids. Microbiol Mol Biol Rev **62**:434–464.

Demain, A. L., 2000. Small bugs, big business: the economic power of the microbe. Biotechnol Adv **18**:499–514.

Demarre, G., A.-M. Guérout, C. Matsumoto-Mashimo, D. A. Rowe-Magnus, P. Marlière, and D. Mazel, 2005. A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPα) conjugative machineries and their cognate *Escherichia coli* host strains. Res Microbiol **156**:245–255.

DiCenzo, G. C. and T. M. Finan, 2015. Genetic redundancy is prevalent within the 6.7 mb *Sinorhizobium meliloti* genome. Mol Genet Gen .

Dimpfl, J. and H. Echols, 1989. Duplication mutation as an SOS response in *Escherichia coli*: enhanced duplication formation by a constitutively activated RecA. Genetics **123**:255–60.

Dutra, B. E., V. A. Sutera, and S. T. Lovett, 2007. RecA-independent recombination is efficient but limited by exonucleases. Proc Natl Acad Sci USA **104**:216–21.

Ellis, H. M., D. Yu, T. DiTizio, and D. L. Court, 2001. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. Proc Natl Acad Sci USA **98**:6742–6.

EMD Millipore, 2011a. KOD hot start DNA polymerase user protocol.

EMD Millipore, 2011b. KOD xtreme user protocol.

Ensley, B. D., B. J. Ratzkin, T. D. Osslund, M. J. Simon, L. P. Wackett, and D. T. Gibson, 1983. Expression of naphthalene oxidation genes in *Escherichia coli* results in the biosynthesis of indigo. Science. **222**:167–169.

Farruggio, A. P., C. L. Chavez, C. L. Mikell, and M. P. Calos, 2012. Efficient reversal of ΦC31 integrase recombination in mammalian cells. Biotechnol. J. **7**:1332–6.

Fellay, R., J. Frey, and H. Krisch, 1987. Interposon mutagenesis of soil and water bacteria: a family of DNA fragments designed for in vitro insertional mutagenesis of Gram-negative bacteria. Gene **52**:147–154.

Fiedler, S., A. Steinbuchel, and B. H. A. Rehm, 2000. PhaG-mediated synthesis of poly(3-hydroxyalkanoates) consisting of medium-chain-length constituents from nonrelated carbon sources in recombinant *Pseudomonas fragi*. Appl Environ Microbiol **66**:2117–2124.

Finan, T. M., E. Hartweig, K. LeMieux, K. Bergman, G. C. Walker, and E. R. Signer, 1984. General transduction in *Rhizobium meliloti*. J Bacteriol **159**:120–4.

Finan, T. M., B. Kunkel, G. F. De Vos, and E. R. Signer, 1986. Second symbiotic megaplasmid in *Rhizobium meliloti* carrying exopolysaccharide and thiamine synthesis genes. J Bacteriol **167**:66–72.

Francia, M. V., F. de la Cruz, and J. M. Lobo, 1993. Secondary sites for integration mediated by the Tn21 integrase. Mol Microbiol **10**:823–828.

Gaillard, M., T. Vallaeys, F. J. Vorhölter, M. Minoia, C. Werlen, V. Sentchilo, A. Pühler, and J. R. van der Meer, 2006. The clc element of *Pseudomonas* sp. strain B13, a genomic island with various catabolic properties. J Bacteriol **188**:1999–2013.

Gaj, T., C. A. Gersbach, and C. F. Barbas, 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. **31**:397–405.

Gay, P., D. Le Coq, M. Steinmetz, T. Berkelman, and C. I. Kado, 1985. Positive selection procedure for entrapment of insertion sequence elements in Gram-negative bacteria. J Bacteriol **164**:918–921.

Ghosh, P., L. a. Bibb, and G. F. Hatfull, 2008. Two-step site selection for serine-integrase-mediated excision: DNA-directed integrase conformation and central dinucleotide proofreading. Proc Natl Acad Sci USA **105**:3238–43.

Ghosh, P., A. I. Kim, and G. F. Hatfull, 2003. The orientation of mycobacteriophage bxb1 integration is solely dependent on the central dinucleotide of *attP* and *attB*. Mol Cell **12**:1101–1111.

Ghosh, P., L. R. Wasil, and G. F. Hatfull, 2006. Control of phage Bxb1 excision by a novel recombination directionality factor. PLoS Biol. **4**:e186.

Gibson, D. G., G. A. Benders, C. Andrews-Pfannkoch, E. A. Denisova, H. Baden-Tillson, J. Zaveri, T. B. Stockwell, A. Brownley, D. W. Thomas, M. A. Algire, C. Merryman, L. Young, V. N. Nosskov, J. I. Glass, J. C. Venter, C. A. Hutchisson III, and H. O. Smith, 2008a. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. Science. **319**:1215–1219.

Gibson, D. G., G. A. Benders, K. C. Axelrod, J. Zaveri, M. A. Algire, M. Moodie, M. G. Montague, J. C. Venter, H. O. Smith, C. A. Hutchison 3rd, and r. Hutchison C. A., 2008b. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. Proc Natl Acad Sci U S A **105**:20404–20409.

Gibson, D. G., J. I. Glass, C. Lartigue, V. N. Noskov, R.-Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z.-Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P. P. Parmar, I. I. I. Hutchison Clyde A., H. O. Smith, J. C. Venter, and C. A. Hutchison III, 2010. Creation of a bacterial cell controlled by a chemically synthesized genome. Science **329**:52–56.

Greenspan, P., E. P. Mayer, and S. D. Fowler, 1985. Nile red: a selective fluorescent stain for intracellular lipid droplets. J Cell Biol **100**:965–73.

Griffin, G. J. L., 1994. Chemistry and technology of biodegradable polymers. Springer Netherlands.

Grindley, N. D. F., K. L. Whiteson, and P. A. Rice, 2006. Mechanisms of site-specific recombination. Annu. Rev. Biochem. **75**:567–605.

Groth, A. C. and M. P. Calos, 2004. Phage integrases: biology and applications. J Mol Biol **335**:667–678.

Groth, A. C., E. C. Olivares, B. Thyagarajan, and M. P. Calos, 2000. A phage integrase directs efficient site-specific integration in human cells. Proc Natl Acad Sci U S A **97**:5995–6000.

Gründemann, D. and E. Schömig, 1996. Protection of DNA during preparative agarose gel electrophoresis against damage induced by ultraviolet light. Biotechniques **21**:898–903.

Gupta, M., R. Till, and M. C. Smith, 2007. Sequences in *attB* that affect the ability of ΦC31 integrase to synapse and to activate DNA cleavage. Nucleic Acids Res **35**:3407–3419.

Hanahan, D., 1983. Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol **166**:557–580.

Hao, W. and G. B. Golding, 2009. Does gene translocation accelerate the evolution of laterally transferred genes? Genetics **182**:1365–75.

Heil, J. R., J. Cheng, and T. C. Charles, 2012. Site-specific bacterial chromosome engineering: ΦC31 integrase mediated cassette exchange (IMCE). J Vis Exp :1–5.

Heil, J. R., R. F. Nordeste, and T. C. Charles, 2011. The fluorescence theatre: a cost-effective device using theatre gels for fluorescent protein and dye screening. Can J Microbiol **57**:339–342.

Hirsch, P. R. and J. E. Beringer, 1984. A physical map of pPH1JI and pJB4JI. Plasmid **12**:139–141.

Holmes, B., M. Popoff, M. Kiredjian, and K. Kersters, 1988. *Ochrobactrum anthropi* gen. nov., sp. nov. from human clinical specimens and previously known as group vd. Int J Syst Bacteriol **38**:406–416.

Holt, R. A., R. Warren, S. Flibotte, P. I. Missirlis, and D. E. Smailus, 2007. Rebuilding microbial genomes. BioEssays **29**:580–590.

House, B. L., M. W. Mortimer, and M. L. Kahn, 2004. New Recombination Methods for Sinorhizobium meliloti Genetics. Apple Env Microbiol **70**:2806–2815.

Ingram, C., M. Brawner, P. Youngman, and J. Westpheling, 1989. *xylE* functions as an efficient reporter gene in *Streptomyces* spp.: use for the study of *galP1*, a catabolite-controlled promoter. J Bacteriol **171**:6617–24.

Isaacs, F. J., P. A. Carr, H. H. Wang, M. J. Lajoie, B. Sterling, L. Kraal, A. C. Tolonen, T. A. Gianoulis, D. B. Goodman, N. B. Reppas, C. J. Emig, D. Bang, S. J. Hwang, M. C. Jewett, J. M. Jacobson, and G. M. Church, 2011. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. Science **333**:348–53.

Itaya, M., K. Fujita, A. Kuroki, and K. Tsuge, 2008. Bottom-up genome assembly using the *Bacillus subtilis* genome vector. Nat Methods **5**:41–43.

Itaya, M., K. Tsuge, M. Koizumi, and K. Fujita, 2005. Combining two genomes in one cell: stable cloning of the *Synechocystis* PCC6803 genome in the *Bacillus subtilis* 168 genome. Proc Natl Acad Sci USA **102**:15971–15976.

Jiang, W., D. Bikard, D. Cox, F. Zhang, and L. A. Marraffini, 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. **31**:233–9.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science **337**:816–21.

Jones, J. D. and N. Gutterson, 1987. An efficient mobilizable cosmid vector, pRK7813, and its use in a rapid method for marker exchange in *Pseudomonas fluorescens* strain HV37a. Gene **61**:299–306.

Jones, T. D., K. A. Chaffin, F. S. Bates, B. K. Annis, E. W. Hagaman, M.-H. Kim, G. D. Wignall, W. Fan, and R. Waymouth, 2002. Effect of tacticity on coil dimensions and thermodynamic properties of polypropylene. Macromolecules **35**:5061–5068.

Kaneko, T., Y. Nakamura, S. Sato, E. Asamizu, T. Kato, S. Sasamoto, A. Watanabe, K. Idesawa, A. Ishikawa, K. Kawashima, T. Kimura, Y. Kishida, C. Kiyokawa, M. Kohara, M. Matsumoto, A. Matsuno, Y. Mochizuki, S. Nakayama, N. Nakazaki, S. Shimpo, M. Sugimoto, C. Takeuchi, M. Yamada, and S. Tabata, 2000. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. DNA Res. **7**:331–8.

Karas, B. J., J. Jablanovic, L. Sun, L. Ma, G. M. Goldgof, J. Stam, A. Ramon, M. J. Manary, E. A. Winzeler, J. C. Venter, P. D. Weyman, D. G. Gibson, J. I. Glass, C. A. Hutchison, H. O. Smith, and Y. Suzuki, 2013a. Direct transfer of whole genomes from bacteria to yeast. Nat Methods **10**:410–2.

Karas, B. J., B. Molparia, J. Jablanovic, W. J. Hermann, Y.-C. Lin, C. L. Dupont, C. Tagwerker, I. T. Yonemoto, V. N. Noskov, R.-Y. Chuang, A. E. Allen, J. I. Glass,

C. A. Hutchison, H. O. Smith, J. C. Venter, and P. D. Weyman, 2013b. Assembly of eukaryotic algal chromosomes in yeast. J Biol Eng **7**:30.

Kato, M., H. J. Bao, C.-K. Kang, T. Fukui, and Y. Doi, 1996. Production of a novel copolyester of 3-hydroxybutyric acid and medium-chain-length 3-hydroxyalkanoic acids by *Pseudomonas* sp. 61-3 from sugars. Appl Microbiol Biotechnol **45**:363–370.

Katzen, F., 2007. Gateway recombinational cloning: a biological operating system. Expert Opin Drug Discov :571–586.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, and A. Drummond, 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics **28**:1647–9.

Kellerhals, M. B., B. Kessler, B. Witholt, A. Tchouboukov, and H. Brandl, 2000. Renewable long-chain fatty acids for production of biodegradable medium-chain-length polyhydroxyalkanoates (mcl-PHAs) at laboratory and pilot plant scales. Macromolecules **33**:4690–4698.

Khaleel, T., E. Younger, A. R. McEwan, A. S. Varghese, and M. C. M. Smith, 2011. A phage protein that binds ΦC31 integrase to switch its directionality. Mol Microbiol **80**:1450–63.

Khorana, H. G., K. L. Agarwal, H. Buchi, M. H. Caruther, N. K. Gupta, K. Kleppe, A. Kumar, E. Ohtsuka, U. L. Rajbhand, V. J. H, V. Sgaramel, T. Terao, H. Weber, and T. Yamada, 1972. Studies on polynucleotides 103 total synthesis of structural gene for an alanine transfer ribonucleic acid from yeast. J Mol Biol **72**:209–217.

Kim, Y.-C. and S. L. Morrison, 2009. A rapid and economic in-house DNA purification method using glass syringe filters. PLoS One **4**:e7750.

Kushtoss, S. and R. N. Rao, 1991. Analysis of the integration function of the Streptomycete Bacteriophage ΦC3 1. J Mol Biol **222**:897–908.

Langella, P. and A. Chopin, 1989. Effect of restriction-modification systems on transfer of foreign DNA into *Lactococcus lactis* subsp. lactis. FEMS Microbiol Lett **59**:301–305.

Lartigue, C., J. I. Glass, N. Alperovich, R. Pieper, P. P. Parmar, I. I. I. Hutchison Clyde A., H. O. Smith, J. C. Venter, C. A. Hutchison III, and C. A. Hutchison, 2007. Genome transplantation in bacteria: changing one species to another. Science. **317**:632–638.

Lea, D. E. and C. A. Coulson, 1949. The distribution of the numbers of mutants in bacterial populations. J Genet **49**:264–285.

Leigh, J. A., E. R. Signer, and G. C. Walker, 1985. Exopolysaccharide-deficient mutants of *Rhizobium meliloti* that form ineffective nodules. Proc Natl Acad Sci U S A **82**:6231–6235.

Lennox, E. S., 1955. Transduction of linked genetic characters of the host by bacteriophage P1. Virology **1**:190–206.

Leschziner, A. E., M. R. Boocock, and N. D. Grindley, 1995. The tyrosine-6 hydroxyl of gamma delta resolvase is not required for the DNA cleavage and rejoining reactions. Mol Microbiol **15**:865–870.

Lesic, B. and L. G. Rahme, 2008. Use of the lambda Red recombinase system to rapidly generate mutants in Pseudomonas aeruginosa. BMC Mol Biol **9**:20.

Llosa, M., S. Bolland, and F. De La Cruz, 1991. Structural and functional analysis of the origin of conjugal transfer of the broad-host-range IncW plasmid R388 and comparison with the related IncN plasmid R46. Mol Gen Genet **226**:473–483.

Lovett, S. T., R. L. Hurley, V. A. Sutera, R. H. Aubuchon, and M. A. Lebedeva, 2002. Crossing over between regions of limited homology in *Escherichia coli* RecA-dependent and RecA-independent pathways. Genetics **160**:851–9.

Lubys, A., J. Bagdanaviciute, and L. Zakareviciene, 2007. Improved positive selection cloning vector.

Luo, Z. Q. and S. K. Farrand, 1999. Cloning and characterization of a tetracycline resistance determinant present in *Agrobacterium tumefaciens* C58. J Bacteriol **181**:618–26.

Luria, S. E. and M. Delbrück, 1943. Mutations of bacteria from virus sensitivity to virus resistance. Genetics **28**:491–511.

Marchessault, R. H. and G.-e. Yu, 2001. Crystallization and material properties of polyhydroxyalkanoates, vol. 3b: Polyes. 1st ed. Wiley-VCH, Weinheim.

Matsuura, M., T. Noguchi, D. Yamaguchi, T. Aida, M. Asayama, H. Takahashi, and M. Shirai, 1996. The *sre* gene (ORF469) encodes a site-specific recombinase responsible for integration of the R4 phage genome. J Bacteriol **178**:3374–3376.

Meddeb-Mouelhi, F., C. Dulcey, and M. Beauregard, 2012. High transformation efficiency of *Bacillus subtilis* with integrative DNA using glycine betaine as osmoprotectant. Anal Biochem **424**:127–9.

Miller, J. H., 1972. Experiments in molecular genetics. Cold Spring Harbour Press, Cold Spring Harbour, New York.

Miyazaki, R. and J. R. van der Meer, 2013. A new large-DNA-fragment delivery system based on integrase activity from an integrative and conjugative element. Appl Environ Microbiol **79**:4440–7.

Mumm, J. P., A. Landy, and J. Gelles, 2006. Viewing single lambda site-specific recombination events from start to finish. EMBO J. **25**:4586–4595.

Noskov, V. N., B. J. Karas, L. Young, R.-Y. Chuang, D. G. Gibson, Y.-C. Lin, J. Stam, I. T. Yonemoto, Y. Suzuki, C. Andrews-Pfannkoch, J. I. Glass, H. O. Smith, C. A. Hutchison, J. C. Venter, and P. D. Weyman, 2012. Assembly of large, high G+C bacterial DNA fragments in yeast. ACS Synth Biol **1**:267–73.

Nunes-Duby, S. E., H. J. Kwon, R. S. Tirumalai, T. Ellenberger, and A. Landy, 1998. Similarities and differences among 105 members of the Int family of site-specific recombinases. Nucl Acids Res **26**:391–406.

Otero, R. D. C., P. Hsieh, and R. D. Camerini-Otero, 1995. Homologous recombination proteins in prokaryotes and eukaryotes. Annu. Rev. Genet. **29**:509–552.

Pachekoski, W. M., J. A. M. Agnelli, and L. P. Belem, 2009. Thermal, mechanical and morphological properties of poly (hydroxybutyrate) and polypropylene blends after processing. Mater. Res. **12**:159–164.

Popham, D. L. and P. Stragier, 1992. Binding of the *Bacillus subtilis spoIVCA* product to the recombination sites of the element interrupting the sigma K-encoding gene. Proc Natl Acad Sci USA **89**:5991–5995.

Poteete, A. R., 2001. What makes the bacteriophage lambda Red system useful for genetic engineering: molecular mechanism and biological function. FEMS Microbiol Lett **201**:9–14.

Puigbò, P., E. Guzmán, A. Romeu, and S. Garcia-Vallvé, 2007. Optimizer: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res. **35**:W126–31.

Ravatn, R., S. Studer, D. Springael, A. J. Zehnder, and J. R. van der Meer, 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**:4360–9.

Roche Applied Science, 2009. DIG high primer DNA labeling and detection starter kit II.

Rowe-Magnus, D. A., 2009. Integrase-directed recovery of functional genes from genomic libraries. Nucleic Acids Res. **37**:e118–e118.

Rowley, P. A., M. C. Smith, and E. Younger, 2008. A motif in the C-terminal domain of ΦC31 integrase controls the directionality of recombination. Nucleic Acids Res **36**:3879–3891.

Sambrook, J. and D. W. Russell, 2001. Molecular cloning a laboratory manual. Third edit ed. Cold Spring Harbour Press, Cold Spring Harbour, New York.

Sanjana, N. E., L. Cong, Y. Zhou, M. M. Cunniff, G. Feng, and F. Zhang, 2012. A transcription activator-like effector toolbox for genome engineering. Nat Protoc **7**:171–92.

Sato, T., Y. Samori, and Y. Kobayashi, 1990. The *cisA* cistron of *Bacillus subtilis* sporulation gene *spoIVC* encodes a protein homologous to a site-specific recombinase. J Bacteriol **172**:1092–1098.

Schafer, A., A. Tauch, W. Jager, J. Kalinowski, G. Thierbach, and A. Puhler, 1994. Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutumicum. Gene **145**:69–73.

Shetty, R. P., D. Endy, T. F. Knight, and J. Knight T. F., 2008. Engineering BioBrick vectors from BioBrick parts. J Biol Eng **2**:5.

Simon, C. and R. Daniel, 2011. Metagenomic Analyses: past and future trends. Appl Environ Microbiol **77**:1153–1161.

Slater, S. C., W. H. Voige, and D. E. Dennis, 1988. Cloning and expression in *Escherichia coli* of the *Alcaligenes eutrophus* H16 poly-beta-hydroxybutyrate biosynthetic pathway. J Bacteriol **170**:4431–4436.

Smith, H. O., r. Hutchison C. A., C. Pfannkoch, J. C. Venter, and C. A. Hutchison 3rd, 2003. Generating a synthetic genome by whole genome assembly: ϕX174 bacteriophage from synthetic oligonucleotides. Proc Natl Acad Sci U S A **100**:15440–15445.

Smith, M. C. M. and H. M. Thorpe, 2002. Diversity in the serine recombinases. Mol Microbiol **44**:299–307.

Spiekermann, P., B. H. A. Rehm, R. Kalscheuer, D. Baumeister, and A. Steinbüchel, 1999. A sensitive, viable-colony staining method using Nile red for direct screening of bacteria that accumulate polyhydroxyalkanoic acids and other lipid storage compounds. Arch Microbiol **171**:73–80.

Sternberg, N. and D. Hamilton, 1981. Bacteriophage P1 site-specific recombination : I. recombination between *loxP* sites. J Mol Biol **150**:467–486.

Tamvakopoulos, C. S. and V. E. Anderson, 1992. Detection of acyl-coenzyme a thioester intermediates of fatty acid β-oxidation as the N-acylglycines by negative-ion chemical ionization gas chromatography?mass spectrometry. Anal Biochem **200**:381–387.

Thaker, M. N. and G. D. Wright, 2012. Opportunities for synthetic biology in antibiotics: expanding glycopeptide chemical diversity. ACS Synth Biol **4**:195–206.

Thomas, C. M., 1981. Molecular genetics of broad host range plasmid RK2. Plasmid **5**:10–19.

Thomason, L. C., R. Calendar, and D. W. Ow, 2001. Gene insertion and replacement in *Schizosaccharomyces pombe* mediated by the *Streptomyces* bacteriophage ΦC31 site-specific recombination system. Mol. Genet. Genomics **265**:1031–1038.

Thorpe, H. M. and M. C. Smith, 1998. In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. Proc Natl Acad Sci USA **95**:5505–5510.

Thorpe, H. M., S. E. Wilson, and M. C. M. Smith, 2000. Control of directionality in the site-specific recombination system of the *Streptomyces* phage ΦC31. Mol Microbiol **38**:232–241.

Thorsted, P. B., D. P. Macartney, P. Akhtar, A. S. Haines, N. Ali, P. Davidson, T. Stafford, M. J. Pocklington, W. Pansegrau, B. M. Wilkins, E. Lanka, and C. M. Thomas, 1998. Complete sequence of the IncPbeta plasmid R751: implications for evolution and organisation of the IncP backbone. J Mol Biol **282**:969–990.

Timmins, M. R., R. W. Lenz, P. J. Hocking, R. H. Marchessault, and R. C. Fuller, 1996. Effect of tacticity on enzymatic degradability of poly(β-hydroxybutyrate). Macromol. Chem. Phys. **197**:1193–1215.

Tombolini, R., S. Povolo, A. Buson, A. Squartini, and M. P. Nuti, 1995. Poly-beta-hydroxybutyrate (PHB) biosynthetic genes in *Rhizobium meliloti* 41. Microbiology **141**:2553–2559.

Tran, T., 2015. Metabolic engineering *Sinorhizobium meliloti* and *Pseudomonas putida* for novel polymer production. Ph.D. thesis, University of Waterloo, Waterloo Ontario.

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm, and S. G. Rozen, 2012. Primer3–new capabilities and interfaces. Nucleic Acids Res. **40**:e115.

van Krevelen, D. and K. te Nijenhuis, 2009. Properties of polymers: their correlation with chemical structure; their numerical estimation and prediction from additive group contributions. 4th ed. Elsevier, Oxford.

Vincent, J. M., 1970. A manual for the practical study of root-nodule bacteria. [Published for the] International Biological Programme [by] Blackwell Scientific.

Waddell, C. S. and N. L. Craig, 1989. Tn7 transposition: recognition of the attTn7 target sequence. Proc Natl Acad Sci USA **86**:3958–62.

Wang, H. H., F. J. Isaacs, P. A. Carr, Z. Z. Sun, G. Xu, C. R. Forest, and G. M. Church, 2009. Programming cells by multiplex genome engineering and accelerated evolution. Nature **460**:894–898.

Wang, H. H., H. Kim, L. Cong, J. Jeong, D. Bang, and G. M. Church, 2012. Genome-scale promoter engineering by coselection MAGE. Nat Methods **6**:591–3.

Warner, J. R., P. J. Reeder, A. Karimpour-Fard, L. B. A. Woodruff, and R. T. Gill, 2010. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. Nat Biotech **28**:856–862.

Watson, B., T. C. Currier, M. P. Gordon, M. D. Chilton, and E. W. Nester, 1975. Plasmid required for virulence of *Agrobacterium tumefaciens*. J Bacteriol **123**:255–264.

Widner, B., R. Behr, S. Von Dollen, M. Tang, T. Heu, A. Sloma, D. Sternberg, P. L. Deangelis, P. H. Weigel, and S. Brown, 2005. Hyaluronic acid production in *Bacillus subtilis*. Appl Environ Microbiol **71**:3747–52.

Wiedenheft, B., K. Zhou, M. Jinek, S. M. Coyle, W. Ma, and J. A. Doudna, 2009. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. Structure **17**:904–12.

Wu, H.-A., D.-S. Sheu, and C.-Y. Lee, 2003. Rapid differentiation between short-chain-length and medium-chain-length polyhydroxyalkanoate-accumulating bacteria with spectrofluorometry. J Microbiol Methods **53**:131–135.

Yang, T. H., T. W. Kim, H. O. Kang, S.-H. Lee, E. J. Lee, S.-C. Lim, S. O. Oh, A.-J. Song, S. J. Park, and S. Y. Lee, 2010. Biosynthesis of polylactic acid and its copolymers using evolved propionate CoA transferase and PHA synthase. Biotechnol. Bioeng. **105**:150–60.

Yasbin, R. E., G. A. Wilson, and F. E. Young, 1975. Transformation and transfection in lysogenic strains of *Bacillus subtilis*: evidence for selective induction of prophage in competent cells. J Bacteriol **121**:296–304.

Appendix A

Supplementary Figures

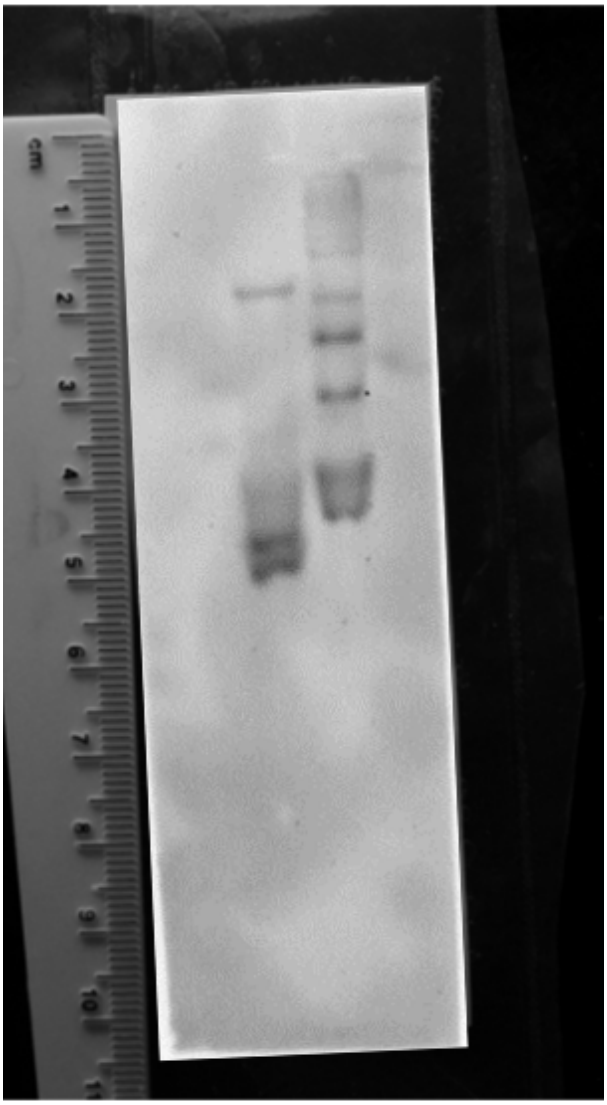

Figure A.1: At1142 blot measurement: shown is the blot ECL bitmap superimposed on the regular photograph of the blot with a ruler included

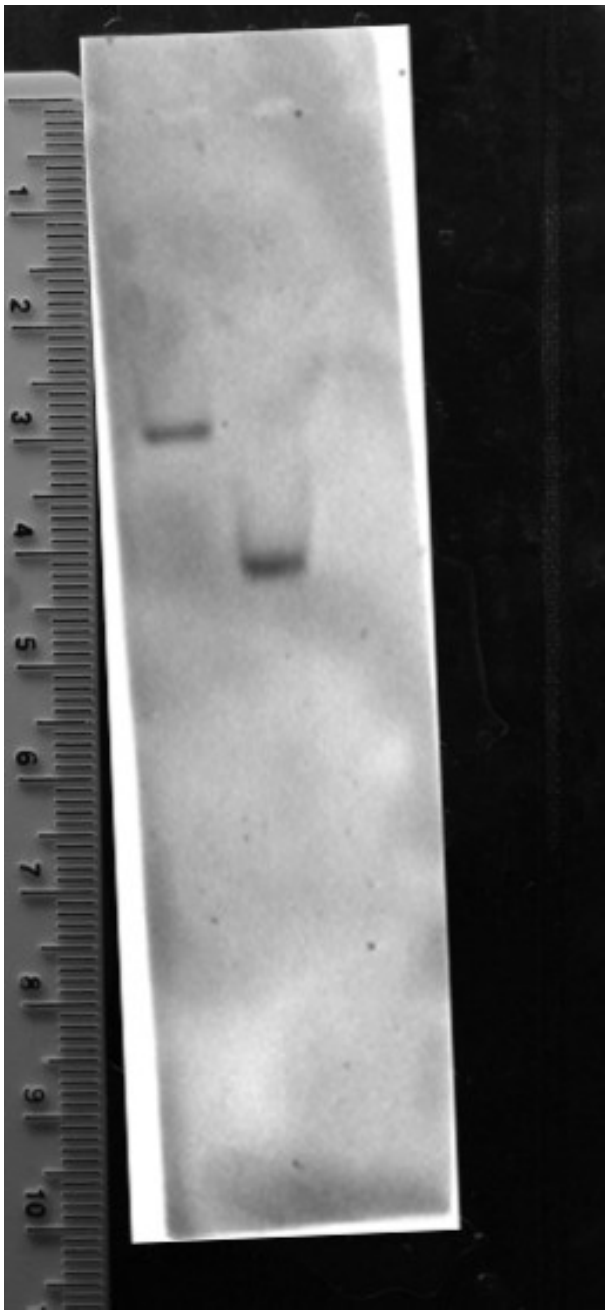

Figure A.2: Ow0013 blot measurement: shown is the blot ECL bitmap superimposed on the regular photograph of the blot with a ruler included

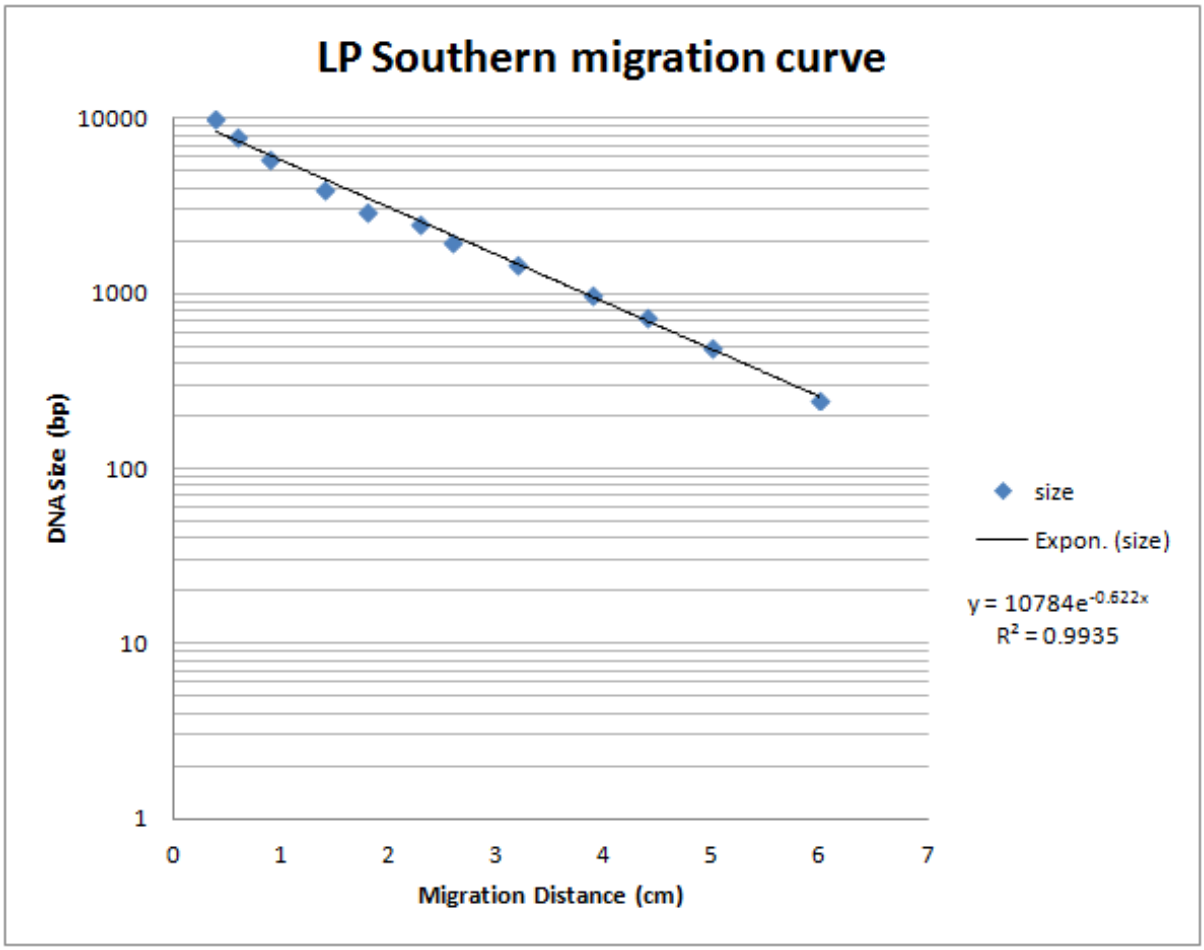

Figure A.3: Standard for Ow0013 and At11142 Southern blot. The standard curve for DNA migration as a function of size is shown. The equation $y = 10784e^{-0.622x}$ was used to determine DNA fragment size (y) from the distance migrated on the agarose gel (x).

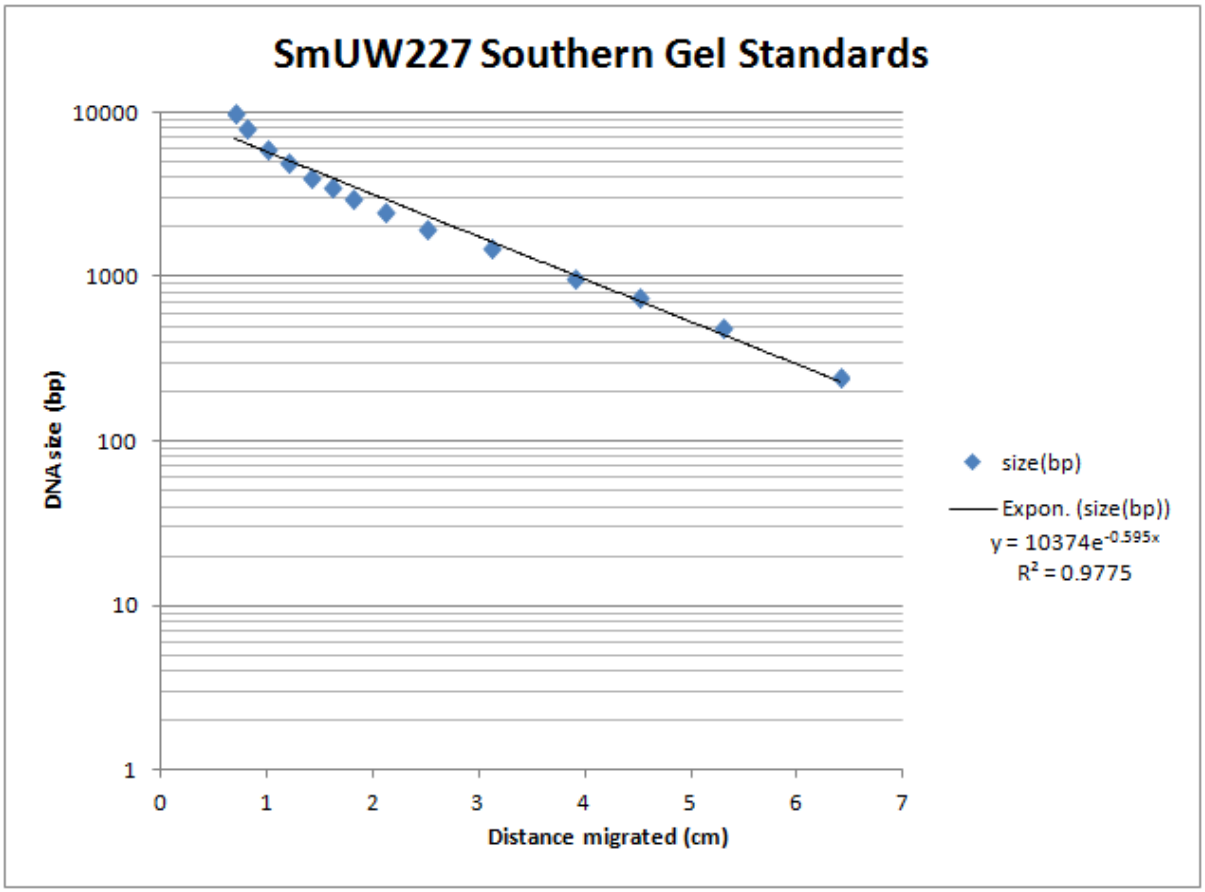

Figure A.4: Standard for SmUW227 Southern blot. The standard curve for DNA migration as a function of size is shown. The equation $y = 10374e^{-0.595x}$ was used to determine DNA fragment size (y) from the distance migrated on the agarose gel (x).

Appendix B

Hyaluronic Acid Project

B.1 Introduction

This project that I worked on was a collaboration with a company who wanted to make high molecular weight hyaluronic acid in *Bacillus subtilus*. Hyaluronic acid production had been shown in *B. subtilus* before (Widner *et al.*, 2005). The company representative thought it would make the difference to use a hyaluronic acid synthase from *Rattus norvegicus* instead of the synthase from *Streptococcus equisimilis*. Some genes (*tuaD, gtaB, gcaD*) from *B. subtilus* were also cloned such that their copy number would be increased, as Widner *et al.* (2005) has shown that increases production. The *S. equisimilis* synthase was also cloned to serve as a positive control. The clones were given to the company and another lab who received subsequent funding. They are also stored in the Charles lab collection. We did not have access to the *B. subtilus* strains used by Widner *et al.* (2005) due to their protection by patents.

B.2 Materials and Methods

B.2.1 Construction of Hyaluronic Acid Operon

What follows is a step by step description of the strain construction for the Hylauronic Acid *Bacillus* project: (All PCR used B. Subtilis 168 genomic DNA as a template). PCR primers also added a ribosome binding site to the 5′ end of the product. The native ribosome binding site (RBS) was used for *gtaB*, *gcaD*, and *tuaD*. The ribosome binding site used in rnHAS2 and spHAS was the native ribosome binding site from *gcaD*. This RBS turned out to be unnecessary for the use in pHT01 (MoBiTec), since pHT01 contains an RBS.

The gene *gtaB* was amplified adding a His tag and adding SmaI-EcoRI sites to the end using the primers: gtaB F SmaI, gtaB R EcoRI his. The gene *gtaB* was cloned into pK19mob (Schafer *et al.*, 1994) using SmaI EcoRI restriction sites, creating pJH131. This step was verified with sequencing. The *gcaD* gene was amplified adding a His tag and adding SphI-XbaI sites using the primers: gcaD F SphI, gcaD R XbaI his. Cut sites (PacI,SalI, XhoI) for future cloning steps were also added in forward primer (gcaD F).The *gcaD* gene was cloned into pJH131 using SphI-XbaI restriction sites to create pJH132. This step was verified by restriction mapping. The *tuaD* gene was amplified adding a His tag and adding XbaI-BamHI sites to the ends using the primers: tuaD F XbaI, tuaD R BamHI his. A stuffer (2 kb long Sp resistance gene) was cloned into the BamHI site of pJH132, to allow for better digestion of the vector. The stuffer was was removed as part of the subsequent cloning steps. Its absence was verified by restriction mapping. The *tuaD* gene was cloned, using the XbaI and BamHI restriction sites, into pJH132 to create pJH133. Here the construction splits into two streams (spHAS and rnHAS2). Codon optimized rnHAS2 was synthesized by biobasic (BioBasic Inc.Markham ON) with SalI-XhoI ends.The codons were optimized with

Optimizer (Puigbò *et al.*, 2007) for expression in *Bacillus subtilus*.The restriction enzymes SalI-XhoI produce compatible ends, so the SalI-XhoI cut insert was cloned into SalI cut pJH133 to create pJH136. The orientation and presence of the insert was verified by two independent restriction maps. Codon optimized spHAS was synthesized in the same way but with PacI-SalI ends. PacI SalI cut SpHAS was cloned into PacI-SalI cut pJH136 to replace rnHAS2 thereby creating pJH135. At this stage a HindIII digest was done on both pJH136 and pJH135 to verify the proper banding pattern (4 bands with 2 differing).The spectinomycin resistance stuffer was used again, but this time to carry new restriction sites (PacI-PmeI) into the Bacillus vectors (pHT01 and pSWEET). pSWEET (Bhavsar *et al.*, 2001) required the addition of just a PmeI site, where pHT01 (MoBiTec) required the addition of PacI and PmeI sites. The full length construct from pJH135 was cloned into both pSWEET and pHT01 (modified versions) using PacI-PmeI. This step created pJH137 (pHT01 based) and pJH139 (pSWEET based). The full length construct from pJH136 was subcloned into both pSWEET and pHT01 (modified versions) using PacI-PmeI. This step created pJH138 (pHT01 based) and pJH140 (pSWEET based).The pHT01 based vectors were transformed into *B. subtilus* 168 using electroporation (Meddeb-Mouelhi *et al.*, 2012). The resultant colonies had the correct resistance and they are stored in the Charles Lab *Bacillus* strain collection. The pSWEET based vectors were transformed into *B. subtilus* 168 using *Bacillus* induced natural competence (Yasbin *et al.*, 1975). A clone potentially containing the construct from pJH140 was isolated. It displayed the correct antibiotic resistance phenotype: chloramphenicol resistance. The pSWEET vector has recombination sites for homologous recombination into the *amyE* gene by double homologous recombination. A screen for the inability to use starch can be done to verify integration.

B.2.2 Oligonucleotides

- gtaB F SmaI: CATCCCGGGAAGGAAGGTGCCTTTTAAATGAAAAAAGTACG-TAAAGCCATAATTC
- gtaB R EcoRI his: CATGAATTCGTTTAAACTTAGTGATGGTGATGGTGTGGATTTCTTC-TTTGTTTAGTAAACC
- gcaD F SphI: CATGCATGCTTAATTAAGTCGACCTCGAGATATTGGAGGCCAATAAATG-GATAAGCGGTTTGCAGTTG
- gcaD R XbaI his: CATTCTAGATTAGTGATGGTGATGGTGATGTTTTTTATGAATATTTTTCA-CATAATCGTC
- TuaD F XbaI: CATTCTAGATACAGGAGAGGGTTGAGCGCATGAAAAAAATAGCT-GTCATTGGAAC
- TuaD R BamHI his: CATGGATCCTTAGTGATGGTGATGGTGATGTAAATTGACGCTTC-CCAAGTCTTTAGC
- PmePacF pHT01 Bam: GATCCTTAATTAAATCGGCTACTGTTTAAAC
- PmePacR pHT01 blunt: GTTTAAACAGTAGCCGATTTAATTAAG
- PmeF pSweet Bam: GATCCATCGGCTACTGTTTAAACG
- PmeR pSweetBam HI: GATCCGTTTAAACAGTAGCCGATG

B.2.3 Conclusion

The clones were made as specified in the project proposal in the time that was allowed. The integrated strain may have the plasmid integrated in one of the cloned *B. subtilus* genes (*tuaD, gtaB, gcaD*) instead of the targeted *amyE* locus.