DEVELOPMENT OF A GENETIC TOOLBOX IN *SINORHIZOBIUM MELILOTI*

DEVELOPMENT OF AN ADVANCED GENETIC TOOLBOX TO ENABLE GENOME SCALE ENGINEERING IN *SINORHIZOBIUM MELILOTI*

By MICHAEL R. MACLEOD, B.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree Master of Science in Biology

McMaster University

MASTER OF SCIENCE (2018) McMaster University

(Biology) Hamilton, Ontario

TITLE: Development of an Advanced Genetic Toolbox to Enable Genome Scale Engineering in S*inorhizobium meliloti*

AUTHOR: Michael R. MacLeod, B. Sc. (McMaster University)

SUPERVISOR: Professor Turlough M. Finan

NUMBER OF PAGES: i-xiv, 1-114

Lay Abstract

Synthetic biology is an emerging field that incorporates principles of molecular biology and engineering for the design and construction of biological systems for application in medicine, agriculture, and industry. Presently, it remains difficult to modify genomes of several organisms due to lack of available techniques. Yeast is currently used for the modification of large DNA pieces, however it is unable to transfer and maintain modified DNA with high G+C content. Here, the bacteria *Sinorhizobium meliloti* was used as a host organism to conduct genetic engineering due to its ability to maintain large DNA pieces with a high G+C content. Characterization experiments were conducted to assess the efficiency of this organism for this task. Using this strain, a proof-of-concept experiment to demonstrate the uptake and maintenance of large, high G+C DNA pieces was completed. This technology may be useful in biotechnology applications for engineering of large DNA pieces from industrially relevant organisms.

Abstract

Synthetic biology has ushered in a new age of molecular biology with the aim towards practical developments in disciplines ranging from medicine, agriculture, and industry. Presently, it remains difficult to manipulate the genomes of many organisms due to lack of genetic tools. These problems can be circumvented by cloning large fragments of DNA into strains where many genetic tools are in place, such as *Saccharomyces cerevisiae*. However, this organism is unable to directly transfer cloned DNA to other organisms and is unable to stably maintain DNA with a G+C content >40%. Many organisms relevant in biotechnology often have G+C content DNA >60%, and therefore are difficult to engineer. Here, the soil bacteria *Sinorhizobium meliloti* was chosen as a host strain to clone and manipulate large fragments of high G+C content DNA. *S. meliloti* is a Gram-negative α-proteobacteria that forms symbiotic relationships with legumes to fix nitrogen. It has a multi-partite genome with a G+C content of 62.7% that includes a chromosome (3.65 Mb), the pSymA (1.35 Mb), and pSymB (1.68 Mb) replicons. A restriction endonuclease *hsdR* mutant strain lacking pSymA and pSymB was created and used in this study. Multi-host shuttle (MHS) vectors were constructed that allow for direct transfer and maintenance of DNA in *E. coli, S. cerevisiae,* and *P. tricornutum.* Characterization of strains was conducted to determine transduction, conjugation, and transformation frequencies, as well as stability of MHS plasmids. Furthermore, a proof-of-concept experiment was conducted to clone large plasmids (70-205 kb) with G+C content >58% via site-specific recombination at a landing pad in the MHS vector, which was then verified using colony PCR. This work demonstrates the usefulness of *S. meliloti* containing a MHS vector for cloning of large fragments with high G+C content DNA, a technology that may be used for several applications in both applied and basic research.

Acknowledgements

I would like to thank Dr. Turlough Finan for giving me the opportunity to work in his laboratory for the last two years, for taking a chance and being patient with me. While working here I have learned a lot and was exposed to many exciting concepts in biology, and for that I am truly grateful. I would like to thank Dr. Brian Golding for being part of my committee and offering advice regarding experiments. I would also like to thank collaborators, Dr. Bogumil Karas and Dr. Trevor Charles along with graduate students Stephanie Brumwell and Ola Matysiakiewicz, whose guidance and weekly meetings helped to keep the project running smoothly.

I am also very grateful for all the work done by previous and present members of the Finan Lab who helped make this research possible. In particular, I would like to thank Dr. George diCenzo, Dr. Maryam Zamani, Jerry Huang, Lisa Situ, Leah Sather, and Jason Kearsley, of which many of their strains and plasmid constructs were used in my study and the results achieved here would not be possible without their hard work.

And of course a special thank you to Dr. Zahed Muhammed and Daniel Hsieh, who helped with the design and troubleshooting of several of my experiments. From talking start-up ideas over coffee and donuts to laughing and making jokes with each other during long days in the lab, their company was always appreciated. I would also like to thank my friends outside the lab who made these last two years enjoyable.

Lastly, I would like to thank my parents, siblings and extended family for their unwavering support and kindness during my graduate studies. It would not have been possible without the contributions of all these people.

Table of Contents

Lay Abstract iii

Abstract iv

Acknowledgements v

Table of Contents vi

List of Figures ix

List of Tables xi

List of Abbreviations and Symbols xii

Declaration of Academic Achievement xiv

1 Literature Review 1

1.1 Introduction 1

1.1.1 Synthetic Biology and Development of Genetic Toolboxes for Genome Scale Manipulation 1

1.1.2 Genome Scale Engineering 3

1.2 *S. meliloti* as a Platform for Manipulation of Large DNA Fragments 7

1.2.1 Multipartite Genome of *Sinorhizobium meliloti* 7

1.2.2 Symbiotic Nitrogen Fixation in *S. meliloti* 8

1.2.3 Development of Minimal *S. meliloti* Cells 9

1.2.4 Restriction Endonuclease System in *S. meliloti* 11

1.3 Development of pGE Plasmids to Enable Large Scale Genomic Engineering 12

1.3.1 Features of pGE Plasmids 12

1.3.2 *In vivo* Cloning Using pGE Plasmids 16

1.4 Site-Specific Recombination Methods 16

1.4.1 Flp/FRT Recombination System 16

1.4.2 *attB/attP* Recombination System 19

1.5 This Work 21

2 Materials and Methods 23

2.1 Bacterial Strains and Plasmids 23

2.2 Oligonucleotides 27

2.3 Bacterial Culturing and Microbiological Techniques 29

2.3.1 Bacterial Growth Media 29
2.3.2 Preparation of Bacterial Frozen Stocks 30
2.4 Bacterial Conjugations 31
2.5 ΦM12 Transductions 31
2.5.1 Preparation and Testing of ΦM12 Transducing Lysates 31
2.5.2 ΦM12-Mediated Transduction 32
2.6 Molecular Biology Techniques 32
2.6.1 Agarose Gel Electrophoresis 32
2.6.2 Preparation of Chemically Competent *E. coli* Cells 33
2.6.3 Transformation of Chemically Competent *E. coli* Cells 33
2.6.4 Preparation of Chemically Competent *S. meliloti* Cells 34
2.6.5 Transformation of Chemically Competent *S. meliloti* Cells 34
2.6.6 Preparation of Electro-Competent Cells 35
2.6.7 Electroporation of Electro-Competent Cells 35
2.6.8 Polymerase Chain Reaction 36
2.6.9 Colony PCR 37
2.6.10 Restriction Enzyme Digestion of DNA 37
2.6.11 Growth Curve Analysis 37
2.7 DNA Purifications 38
2.7.1 Miniature Preparation of Bacterial Plasmid DNA 38
2.7.2 Alkaline Lysis Preparation of Bacterial Plasmid DNA 39
3 Results 39
3.1 Development of Δ*hsdR* Strains 39
3.1.1 Transduction of *hsdR* Deletion in Various Background Strains 39
3.1.2 Colony PCR to Verify Deletion of *hsdR* Deletion Strains 41
3.2 Characterization of Minimal *S. meliloti* Cell 43
3.2.1 Characterization of Growth Rate of Minimal *S. meliloti* Cell 43
3.2.2 Characterizing Transformation Protocols 45
3.2.3 Conjugation of pGE plasmids into Minimal *S. meliloti* Cell 47
3.2.4 Conjugation of Second pGE plasmid into Minimal *S. meliloti* Cell 48
3.2.5 Plasmid Stability Assays 49

3.2.6 Eckhardt Gel Electrophoresis to Observe Possible Recombination of pGE Plasmids 55
3.3 Analysis of Site-Specific Recombination Sites in Useful Background Strains..... 57
3.3.1 Transduction of the wild type *hypRE* locus into Strain with Captured Region from pSymA..... 58
3.3.2 Transduction of Wild Type *hypRE* locus into Strain with Captured pA132 Region.... 60
3.4 Capture of Large Fragments with high G+C Content DNA in pGE Plasmids 61
3.4.1 Cloning Strategy for the Capture of Large High G+C Content Plasmids 61
3.4.2 Construction of pGE Plasmids Harbouring Landing Pad 64
3.4.3 Transfer of the ΦC31 Integrase into the *S. meliloti* Minimal Cell 66
3.4.4 Conjugation of pGE4-LP Plasmid into Background Strains 67
3.4.5 Using Flp-Catalyzed Recombination to Produce Single *FRT* site in Landing Pad..... 68
3.4.6 Conjugation of Large Plasmids into Strains Harbouring pGE4-LP..... 70
3.4.7 Colony PCR to Verify Presence of Large Plasmids in *S. meliloti*..... 72
3.4.8 Colony PCR to Confirm Capture of Large Fragments via Site-Specific Recombination 75
4 General Discussion 80
4.1 Construction and Characterization of the Minimal *S. meliloti* Cell 81
4.2 Cloning of Large Plasmids with High G+C Content DNA into pGE Plasmid 87
4.3 Further Optimization Experiments to Perform 93
4.4 Limitations of pGE Plasmids and Minimal *S. meliloti* Cell..... 93
4.5 Future Prospective Projects 94
4.6 Concluding Statement..... 97
Bibliography..... 97
Appendix..... 105

List of Figures

Figure 1: Circular Maps of pGE Plasmids used in This Study……13

Figure 2: The Flp Recombination Target (*FRT*) Site and Locus of Crossover P1 (*loxP*) Site……17

Figure 3: The Attachment Site on Bacteria (*attB)* and Phage (*attP*) Sites……20

Figure 4: Diagram Depicting Techniques Used in this Study and Diagram of Landing Pad……22

Figure 5: Agarose Gel Demonstrating Deletion of *hsdR* in Genome Reduced Strains……42

Figure 6: Growth Curve Analysis of Genome Reduced Strains……44

Figure 7: Comparison of Transformation Frequency of Strains With and without *hsdR*……47

Figure 8: Plasmid Stability Assay of pGE3 in RmP4258 ΔpSymAB Δ*hsdR*……52

Figure 9: Plasmid Stability Assay of Two pGE Plasmids in pGE3 in RmP4258……53

Figure 10: Second Patching Experiment to Determine Stability of Individual pGE Plasmids……54

Figure 11: Eckhardt Gel Electrophoresis of Strains Harbouring Two pGE Plasmids……56

Figure 12: Diagram of Transduction of Wild Type *hypRE* Region into Strains with 70 kb or 275 kb Regions Putatively Captured at the *hypRE::FRT* Locus……59

Figure 13: Circular Maps of pA132, pTH3255, and pTH3247 Captured in pGE4-LP……63

Figure 14: Schematic Diagram of *In Vivo* Cloning Experiment to Capture Large Plasmids with High G+C content DNA in pGE4-LP……65

Figure 15: Diagnostic Colony PCR to Confirm Presence of ΦC31 integrase gene in RmP4258……67

Figure 16: Diagnostic Restriction Digest of pGE4-LP Following Flp-Catalyzed Recombination……69

Figure 17: Diagnostic Colony PCR to Confirm Presence of pA132 and pTH3255 in RmP4258 and RmP4124.....73

Figure 18: Diagnostic Colony PCR to Confirm Presence of pTH3247 in RmP4314.....74

Figure 19: General Diagram Depicting Binding Locations of Primers Used to Confirm Capture on Landing Pad Locus.....77

Figure 20: Diagnostic Colony PCR to Confirm Capture of 3 Large Plasmids at Landing Pad in pGE4-LP.....78

Figure 21: Alignment of Sequencing Results to Verify Capture of pTH3247 in pGE4-LP at *attB*....90

Figure 22: Diagram Depicting Recombinase-Mediated Cassette Exchange Technique.....96

Figure 23: Percentage of Re-patched Colonies Unable to Grow on LBmc FeCo Media with Single Antibiotic Following Two Plasmid Stability Assay.....105

Figure 24: Colony PCR of *hsdR*::*FRT* Locus to Determine if pA132 or pTH3255 Recombined in RmP4258 or RmP4124 Background.....106

Figure 25: Perl Script to Calculate Generation Times and Standard Deviation of Strains in Growth Curve Analysis.....110

List of Tables

Table 1: Bacterial Strains and Plasmids Used in This Project……24

Table 2: Oligonucleotides Used in This Project……..28

Table 3: Antibiotic Concentrations (µg/mL) Used for *E. coli* and *S. meliloti*……..29

Table 4: Determination of Titer of ΦM12 (RmP110) for Transduction Experiments……..39

Table 5: Transduction Frequency of Neomycin Marker into Reduced Genome Strains……40

Table 6: Conjugation Frequency of pGE Plasmids into RmP4258 ΔpSymAB Δ*hsdR*……..48

Table 7: Conjugation Frequency of a Second pGE Plasmid into RmP4258……49

Table 8: Validation of Large Fragment Integration at the *FRT* site in a Δ*hypRE* Strain……..60

Table 9: Conjugation Frequency of Transfer of 3 Large Plasmids into 3 Background Strains……70

Table 10: Percentage of Patched Colonies that Lost pGE3 plasmid in Stability Assay……..106

Table 11: Percentage of Patched Colonies that Lost two pGE plasmids in Stability Assay……107

Table 12: Percentage of Re-Patched Colonies Unable to Grow Following Two Plasmid Stability Assay……108

List of Abbreviations and Symbols

Δ	deletion
Φ	phage or lysate
λ	lambda
ΩNm	neomycin deletion cassette
°C	degree Celsius
µ	micro
ARS	autonomously replication sequence
attB	attachment site in bacterial genome
attL	attachment site left
attP	attachment site in phage genome
attR	attachment site right
BAC	bacterial artificial chromosome
bp	base pairs
Cas9	CRISPR associated protein 9
CEN	centromere in yeast
CFU	colony forming units
CRISPR	Clustered regularly interspaced short palindromic repeats
DAP	diaminopimelic acid
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
ds	double stranded
DSB	double stranded break
FRT	Flp recognition target
g	gram or genomic
Gm	gentamycin
HGT	horizontal gene transfer
HIS3	histidine biosynthesis gene 3
in vitro	within glass
in vivo	within the living
Int	integrase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	kilobases (10^3)
Km	kanamycin
LB	lysogeny broth
LBmc	lysogeny broth with $MgCl_2$ and $CaCl_2$
loxP	locus of crossover in P1 phage
M	molar
m	milli
M9	minimal salts media
Mb	megabases ($10^{6)}$
mM	millimolar
MCS	multiple cloning site

MHS	multi-host shuttle (vector)
NAT	nourseothricin resistance gene
Nm	neomycin
nm	nanometer
OD_{600}	optical density at 600nm
oriT	origin of conjugal transfer
oriV	origin of vegetative replication
p	plasmid
PCA	protocatechuic acid
PCR	polymerase chain reaction
PEG	polyethylene glycol
PFU	plaque forming units
PHA	polyhydroxyalkanoate
^{r}	resistant
RDF	recombination directionality factor
Rif	rifampicin
RMCE	recombinase mediated cassette exchange
RNA	ribonucleic acid
rpm	revolutions per minute
RT	room temperature
^{s}	sensitive
SDS	sodium dodecyl sulfate
sgRNA	synthetic guide RNA
Sm	streptomycin
SNF	symbiotic nitrogen fixation
Sp	spectinomycin
ss	single stranded
TALEN	transcription activator-like effector nucleases
TAR	transformation associated recombination
TBE	tris-borate-EDTA
Tc	tetracycline
URA3	uracil biosynthesis gene 3
UV	ultraviolet
WT	wildtype
x g	times force of gravity
ZFN	zinc-finger nuclease

Declaration of Academic Achievement

I declare that the research contribution which follows is original work, completed and written by me, with editorial assistance from my supervisor Dr. Turlough M. Finan.

The following individuals contributed to experimentation:

- In Section 1.2.3 – "Development of Minimal *S. meliloti* Cells", the construction of the reduced ΔpSymA/B *S. meliloti* strains was completed by T.M. Finan and G.C. diCenzo.
- In Section 1.3 – "Development of pGE Plasmids to Enable Large Scale Genomic Engineering", the construction and sequence verification of the pGE plasmids was completed by B.J. Karas of Designer Microbes Inc. using TAR cloning techniques.
- In Section 1.3.2 – "*In vivo* Cloning Using pGE Plasmids", the design of this landing pad was completed by J. Kearsley and the construction of the pGE4 plasmid containing the landing pad cassette (pGE4-LP) was completed by P. Janakirama and B.J. Karas of Designer Microbes Inc. using TAR cloning techniques.
- In Section 3.3 – "Analysis of Site-Specific Recombination Sites in Useful Background Strains", the capture of large genomic fragments (70 kb and 275 kb) derived from pSymA at the *hypRE* locus in RmP4254 was completed by L. Sather and L. Situ, respectively.
- In Section 3.2.1 – "Characterization of Growth Rate of Minimal *S. meliloti* Cell", the Perl script used to calculate the generation time and standard deviation of strains used in the experiment was written and provided by G.C. diCenzo.

1 Literature Review

1.1 Introduction

1.1.1 Synthetic Biology and Development of Genetic Toolboxes for Genome Scale Manipulation

Microbes perform many biochemical functions that can be exploited for use in wide-ranging processes, such as producing medicines, feedstocks, biofuels and other bio-products with relative low cost and minimal environmental impact. With continuing advances in genetic engineering and increasing databanks of bioinformatics data, the ability to engineer or optimize microbes to perform certain tasks is becoming more and more commonplace. It is not hard then to imagine a future society where many industrial, medicinal and agricultural processes are completed by the work of engineered microbes to produce renewable and more energy efficient bio-products (Karas, Suzuki, & Weyman, 2015).

The advent of synthetic biology has ushered in a new age of molecular biology with the aim towards practical developments in both commercial and industrial sectors. This emerging field of research combines biological and engineering concepts along with the latest molecular biology technologies towards the study and manipulation of biological organisms. As stated by the European Commission in 2015 (Breitling, Takano, & Gardner, 2015), synthetic biology is “the application of science, technology and engineering to facilitate and accelerate the design, manufacture and/or modification of genetic materials in living organisms.” Experiments in synthetic biology thus aim to use bottom-up approaches and iterative design-build-test-learn

cycles to facilitate building of genetic pathways or introduction of genetic "parts" (such as promoters, terminators, or genes) into useful chassis or host organisms to impart novel biological functions for basic or applied research purposes (Carbonell, et al., 2016).

Research applications of synthetic biology have been geared towards addressing global challenges by means of advances in food production, gene therapies and biofuels or synthetic chemicals. Recent advancements in molecular biology technologies including: improvements in DNA synthesis and sequencing (Kosuri & Church, 2014), use of site-specific genome editing technologies such as CRISPR (Schmidt & Platt, 2017), use of high-throughput and automated procedures (Valente & Fong, 2011), and the development of standard biological parts seen in organizations such as iGEM. Together, these have enabled synthetic biology to become a cost-efficient and independent research field apart from molecular biology or systems biology (Esvelt & Wang, 2013). As these technologies continue to advance and investments in the synthetic biology sector increase, the price to synthesize DNA used to construct genetic pathways or designer genomes will continue to decrease (Baker, 2017).

Despite the possibilities offered by synthetic biology, it remains difficult to manipulate the genomes of many organisms due to a lack of genetic tools. Genome-scale engineering can circumvent many of these problems by cloning large fragments of an organism's genome into more versatile strains where many genetic tools are in place, such as the "workhorse" strains, *Escherichia coli* and *Saccharomyces cerevisiae.* Large fragment cloning (0.1Mb-1Mb) can be achieved by *in vivo* techniques such as transformation-associated recombination (TAR) cloning in *S. cerevisiae* (Noskov, et al., 2012) or by site-specific recombination to capture large regions of genomes as a plasmid (diCenzo & Finan, 2018). *S. cerevisiae* is currently the most common host

used for the capture and manipulation of large DNA fragments, however use of this organism is hindered by its inability to directly transfer cloned DNA to other organisms by conjugation as well as its inability to stably maintain DNA with a G+C content >40% unless yeast origins of replication are added every ~100 kb, which is a tedious process (Noskov, et al., 2012; Karas, et al., 2013). In this study, a soil bacteria *Sinorhizobium meliloti* was chosen as a host strain to clone and manipulate large fragments of high G+C content DNA (see section 1.2).

Many organisms that are commercially or industrially relevant often have genomes with G+C content DNA >60%. One such industrially useful bacteria is *Streptomyces sp.* which has G+C content >65% yet is a useful source for production of antibiotics including tetracycline, kanamycin and gentamicin, among others (de Lima Procopio, et al., 2012). *Pseudomonas sp.*is useful for several industrial and biotechnological purposes including: production of bio-based polymers such as polyhydroxyalkanoates (PHAs), production of pharmaceuticals and agrochemicals, and bioremediation of recalcitrant xenobiotics, and generally have a G+C content of ~62% (Poblete-Castro, et al., 2012; Urgun-Demirtas, Stark, & Pagilla, 2006). *Deinococcus radiodurans*, an extremophile, has recently been considered as a chassis for biotechnology and basic research purposes due to its unparalleled capacity to repair DNA damage following radiation damage (Gerber, et al., 2015). The ability to clone large fragments of high G+C content DNA would enable manipulation of the relevant genomic regions from these and other organisms and allow for genomic engineering for biotechnological purposes.

1.1.2 Genome Scale Engineering

Genome scale engineering can be defined as construction of a genotype that gives rise to a desired phenotype, where a large degree of genetic manipulation is required. The term has

also been used to describe high-throughput techniques used to make single-gene knockout libraries of a given organism (Esvelt & Wang, 2013). Recent technological advances have allowed researchers to overcome previous limitations in genome-scale engineering and allowed unprecedented ability to manipulate whole genomes.

Current methods to conduct genome scale engineering rely on robust protocols for DNA manipulation or assembly of DNA fragments. Current methods can be divided into two main categories: homology-based methods and restriction-based methods (Liu, et al., 2015). Homology-based techniques require the use of homology overlaps between fragments and do not require restriction enzymes. One *in vitro* homology-based technique used widely is Gibson assembly (Gibson, et al., 2008). This has been used to assemble multiple fragments of a bacterial genome up to 166 kb in total size. This technique employs the use of a T5 exonuclease, a DNA polymerase and a DNA ligase in conjunction with homology overlap regions flanking the DNA fragments to be assembled. The T5 exonuclease first chews back the 5' end of the overlap regions for each DNA strand, creating compatible sticky ends which then anneal together. The DNA polymerase is then used to fill in the missing nucleotide gaps in the annealed regions, which are then ligated with DNA ligase. The end result is a single fragment of DNA assembled from smaller DNA fragments containing overlap regions (Gibson, et al., 2009). *In vivo* homology based techniques such as transformation-associated recombination (TAR) cloning in *S. cerevisiae* has also been used to assemble large genomic fragments. The general protocol for TAR cloning involves co-transformation of linearized plasmid and insert DNA to be assembled in competent yeast strains. Each of the DNA fragments contain homology overlaps which are then subsequently recombined together using the naturally occurring yeast recombination systems (Noskov, et al., 2012). Using TAR cloning, 10 or more fragments of DNA can be recombined in a

single experiment to yield large, recombinant plasmids which can then be isolated and transferred into the desired host strain (Noskov, et al., 2012). Other recombination-based methods such as Flp/*FRT*, Cre/*loxP*, and *attB/attP* are frequently used and are described in section 1.4. Other *in vivo* techniques use the λ-red system in *E. coli* to conduct "recombineering", a recombination-based technique first discovered in bacteriophage that allows for homologous recombination of DNA fragments (Esvelt & Wang, 2013). This technique works by the use of the *exo, beta,* and *gam* genes. The Exo protein is a 5' to 3' dsDNA-dependent exonuclease responsible for chewing back DNA to produce ssDNA, which is then able to anneal to the desired insertion site. The Beta protein is responsible for protecting the ssDNA and the Gam protein is required to prevent the endogenous restriction system in *E. coli* from cleaving the incoming DNA (Mosberg, Lajoie, & Church, 2010). Only 35 bp of homology is required to enable recombination between the donor DNA fragment and the recipient fragment, which is less than the usual requirement for recombination in bacterial hosts, which typically requires ~500 bp of homology for recombination to occur (Mosberg, Lajoie, & Church, 2010).

Several restriction-based approaches exist including the *in vitro* technique Golden Gate Assembly, which utilizes Type IIS endonucleases, restriction enzymes that cleave outside of the recognition site. Researchers can design primer sequences with overhang regions that include an overlap region as well as a Type IIS endonuclease site. Cleavage will then generate sticky ends in a sequence-independent manner, and allow two fragments with compatible overlap regions to anneal. A DNA polymerase and a DNA ligase are then used in a similar manner to Gibson assembly to allow for the assembly of recombinant DNA fragments.

In vivo techniques such as zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats) are popular methods for site-specific genome editing. ZFNs and TALENs are considered the predecessors to CRISPR-Cas9 technology, which has revolutionized the ability of researchers to edit genes in a site-specific and efficient manner. ZFNs comprise a linked chain of three zinc-finger domains, which are DNA-binding motifs that recognize a unique 6 bp sequence of DNA, as well as a FokI Type IIS endonuclease domain which cleaves the DNA downstream of the site at which the zinc-finger domains bind. A pair of ZFNs are used together to achieve specificity of >24 bp. The cleaved DNA can then be repaired by providing DNA template during the endogenous cell repair processes of non-homologous end joining (NHEJ) or homology-directed repair (HDR). TALENs are similar to ZFNs in that they consist of a DNA binding domain and a FokI endonuclease domain. The DNA binding domain consists of a transcription activator-like effector (TALE) protein, which was first discovered in the plant bacterial pathogen *Xanthomonas*, where they affect processes in the plant cell and increase its susceptibility to the pathogen (Nemudryi, et al., 2014).TALE enzymes consist of a DNA binding domain, a nuclear localization signal and a domain that activates target gene transcription. Similar to ZFNs, FokI has been translationally fused to the DNA-binding domain of TALE enzymes to allow for cleavage of the desired DNA loci (Nemudryi, et al., 2014).

The field of CRISPR-Cas9 biology continues to rapidly expand since its discovery as a bacterial defense system against invading bacteriophage. When utilized for genome editing, this system includes the nuclease Cas9 (CRISPR-associated protein 9), sgRNA (synthetic guide RNA), and a repair template, and is usually encoded on a single plasmid which can be transformed into the intended organism. The first step is the expression of the sgRNA, an RNA sequence with

homology to the target sequence to be edited. The sgRNA interacts with the Cas9 protein and guides it towards the target sequence where it binds to the target sequence. The Cas9 then induces a double stranded break in the DNA via catalytic cleavage. DNA repair mechanisms such as HDR are recruited, which repair the DNA using a template DNA fragment which is recombined at the target sequence to replace the existing region. Several types of Cas9 proteins and derivations of CRISPR technology exist, including dead Cas9 (Ratner, Sampson, & Weiss, 2016).

Limitations of these technologies include off-target effects which may affect genes fundamental for survival in cells (Gaj, Gersbach, & Barbas III, 2013). Additionally, NHEJ and HDR may introduce deleterious mutations in the target sequence, or may not provide the desired genetic alteration. Despite these limitations, ZFNs, TALENs, and CRISPR-Cas9 are highly efficient tools to enable precise genome editing and modification.

1.2 *S. meliloti* as a Platform for Manipulation of Large DNA Fragments

1.2.1 Multipartite Genome of *Sinorhizobium meliloti*

Sinorhizobium meliloti, the organism used in this study, is a Gram-negative α-proteobacterium that inhabits the soil under free-living conditions and is capable of symbiotically fixing nitrogen within the root of legumes such as alfalfa (*Medicago sativa*) as nitrogen-fixing organisms. It is closely related to other rhizobial species such as the plant pathogen *Agrobacterium,* responsible for crown gall diseases in plants (Galibert, et al., 2001). *S. meliloti* serves as a model organism for rhizobia-legume symbiosis and N_2-fixation, both of which are of major agronomic importance.

The genome of *S. meliloti* is composed of three replicons including a chromosome (3.65 Mb), a pSymA megaplasmid (1.35 Mb), and a pSymB chromid (1.68 Mb) and has an overall G+C

content DNA of 62.7% (Galibert, et al., 2001). Recently, a derivative of this strain was produced which lacked the pSymA and pSymB replicons, resulting in a "minimal genome" with a 45% reduction in size (diCenzo, et al., 2014).

1.2.2 Symbiotic Nitrogen Fixation in *S. meliloti*

The symbiotic relationship between legumes and rhizobia is an important process for both agricultural practices as well as ecologically important processes. Here, the bacteria and plants work together to fix atmospheric (N_2) gas to ammonia (NH_3) for assimilation by the legume host. The symbiosis relationship starts when the plant secretes flavonoids, a type of secondary metabolite, into the surrounding rhizosphere. These molecules induce the expression of *nod* genes in rhizobia, structural genes responsible for expressing and secreting Nod factors known as lipochito-oligosaccharides. These Nod factors are recognized by the plant, and induce cell differentiation in plant cells within the root cortex, which gives rise to the nodule primordium, the basis for the formation of root nodules (Bartsez, et al., 2004). Infection threads allow the rhizobia access into root cortical cells which are then released into the inner nodule cells. Once within the plant cells, rhizobia are then able to undergo differentiation into bacteroids, specialized cells responsible for the reduction of nitrogen into ammonia (Oldryd & Downie, 2008). The gene products responsible for nitrogen fixation include the *nif* and *fix* genes. The *nifD* and *nifK* genes encode the subunits for the nitrogenase enzyme dinitrogenase and the *nifH* encodes the enzyme dinitrogenase reductase, both of which are responsible for the reduction of $N_2$to ammonia. The *fix* genes also play an important role in the nitrogen fixation process; however, the exact function remains unknown for many of these genes. Fix genes are required for N_2-fixation but are not found in free-living diazotrophs such as *Klebsiella* or

Clostridium. Despite having an unknown function, it has been shown that the *fixABCX* operon is essential for nitrogen fixation and is postulated to encode components of the electron transport chain associated with the nitrogenase complex (Torres-Quesada, et al., 2014).

The genomic regions deemed essential for nitrogen fixation have been previously identified by diCenzo et al., by using a systematic deletion approach to delete large regions of pSymA and pSymB and then screening this mutant library for the ability to fix nitrogen and form root nodules in alfalfa (Milunovic, et al., 2014; diCenzo, et al., 2016). By screening the pSymA mutant strains, it was found that four deletions (ΔA116, ΔA117, ΔA118 and ΔA121) were either deficient for fixing $N_2$and/or deficient in the ability to form nodules. Screening the pSymB mutant strains yielding similar results where three deletions (ΔB108, ΔB109 and ΔB123) were found to be deficient for fixing N_2. Genes found in these pSymB regions include *exo* and *dct* genes, which are known to be required for an effective symbiosis (diCenzo, et al., 2016; Finan, et al., 1986). Further work to isolate these essential symbiotic N_2-fixing genomic regions from pSymA and pSymB is underway (Huang, Situ and Finan, 2018, unpublished; Kearsley and Finan, 2018, unpublished).

1.2.3 Development of Minimal *S. meliloti* Cells

A key principal in the field of synthetic biology is genome reduction for “bottom-up” research approaches to better facilitate the integration of synthetic parts and use of recombinant plasmids in a gain-of-function fashion (Zhang, Chang, & Wang, 2010; diCenzo, et al., 2016). The concept of genome reduction can be extrapolated towards the development of a minimal genome comprised of the minimal set of genes necessary and sufficient for survival in an environment free from stress with all nutrients provided (Glass, et al., 2006). A widely used

model organism for genome minimization is *Mycoplasma genitalium*, which has one of the smallest known genomes. Recent studies, including transposon mutagenesis, identified the genes necessary for survival in *M. genitalium.* This organism's genome was reduced from 1079 kb to 531 kb, to produce the smallest known genome in a bacterial species. It was observed that many of the genes deemed essential did not have a known function, highlighting the need to continue studying the minimal set of genes required for life (Hutchison III, et al., 2016). A recent review paper by Esvelt and Wang observed the benefits of reduced-genome *E. coli* strains stating that "the useful traits for a biological chassis would include: 1) those capable of fast growth in minimal media, 2) those capable of fermentation (but may be extrapolated to include any useful biological function), 3) those amenable to genetic manipulation and, 4) those minimally sufficient such that removal of any additional gene negatively affects the other three stated considerations."

The study of minimal genomes is also studied in the context of the minimal symbiotic genome necessary for nitrogen fixation in *S. meliloti.* Previously, various derivatives of wildtype *S. meliloti* were constructed which lacked the pSymA, pSymB megaplasmids or both (diCenzo, MacLean, Milunovic, Golding, & Finan, 2014). To remove pSymB, essential genes from pSymB (*engA* and $tRNA^{arg}$) were first transferred to the chromosome. Next, an active toxin-antitoxin locus located on pSymB was deleted and finally, a Gm/Sp resistance plasmid (pTH1414) carrying an incompatibility factor, *incA*, identical to the one found on pSymB, was conjugated into the strain and Gm resistance was selected. Two plasmids harbouring the *incA* region cannot co-exist within a cell and thus, cells that had taken up pTH1414 should not contain pSymB (diCenzo, et al., 2014). Using PCR, it was confirmed that the pSymB chromid was indeed lost. The resulting

ΔpSymAB strains lacked 3.04Mb of DNA, representing a 45.4% reduction of the genome relative to wildtype, effectively minimizing the genome (diCenzo, et al., 2014).

In the context of this project, the ΔpSymAB strain was used as the chassis organism to conduct proof-of-concept experiments to clone and further manipulate large fragments of high G+C content DNA. The strain was further modified by deleting the *hsdR* gene (see section 1.2.4) to improve transformation efficiency and optimize transformation protocols. This ΔpSymAB Δ*hsdR* strain is termed the "minimal cell" and can be used as a basis for cloning of large fragments of high G+C content DNA through the use of vectors designed to uptake large DNA fragments (see section 1.3). Here, the term "minimal cell" refers to the ΔpSymAB Δ*hsdR* reduced genome strain that will be used as the host strain for this project.

1.2.4 Restriction Endonuclease System in *S. meliloti*

The *hsdR* gene in *S. meliloti* belongs to a Type I restriction-modification endonuclease system, which encodes for three multifunctional enzymes capable of mediating restriction and modification reactions (Murray, 2000). Nucleotide sequences recognized by these enzymes are asymmetric and consist of two small specific regions of DNA around 3-5 bp, separated by a non-specific spacer sequence 6-8 bp in length (Murray, 2000). The three subunits involved in this system are *hsdR* (restriction), *hsdM* (modification), and *hsdS* (specificity), which together produce the type I restriction endonuclease enzyme. HsdM and HsdS are both necessary and sufficient for methyltransferase activity of self-DNA, whereas HsdR is required for the restriction of foreign DNA molecules, which are unmethylated and subsequently digested (Murray, 2000). Genomic analysis of the *hsdRSM* genes of *S. meliloti* strains showed several transposons

surrounding these genes, suggesting that the operon was acquired via horizontal gene transfer from a vastly different taxonomic group (Ferri, et al., 2010).

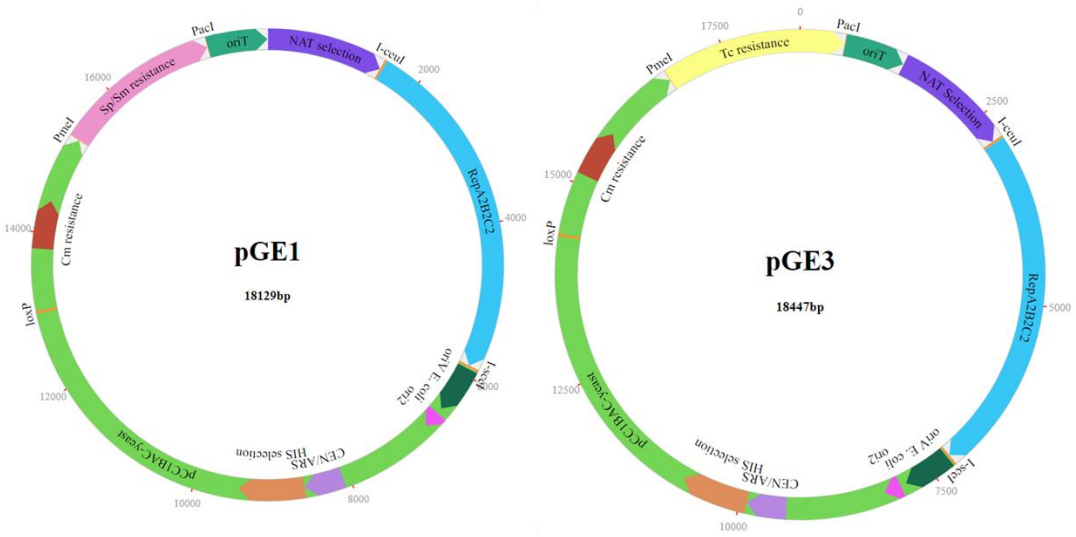
It was reported by Ferri et al., that the deletion of the *hsdR* gene in *S. meliloti* resulted in a 20 to 200-fold increase in electroporation efficiency when the strain was transformed with non-self DNA, compared to the wild-type Rm1021 strain (Ferri, et al., 2010). It is also reported that a 3-fold increase in conjugation efficiency was observed in the Δ*hsdR* strain. In this study, the deletion of *hsdR* was completed in order to improve the transformation efficiency of the host organism to better facilitate its use as a cloning platform host strain.

1.3 Development of pGE Plasmids to Enable Large Scale Genomic Engineering

1.3.1 Features of pGE Plasmids

One of the main features of this project is the use of a multi-host shuttle (MHS) vector to enable cloning and subsequent conjugal transfer of DNA of interest between organisms. These vectors, termed "pGE" plasmids, contain the origins of replication from the pSymA or pSymB replicons along with origins of replication for *E. coli* (*oriV), S. cerevisiae* yeast (CEN/ARS), and the algae *Phaeodactylum tricornutum* (CEN/ARS) to enable replication of this plasmid in each of these organisms. The *oriV* sequence is an RK2 origin of replication was included to allow for TrfA-dependent replication of the large pGE plasmid. TrfA mediates the initiation of replication at the *oriV* site and is responsible for localized unwinding of the DNA to be replicated (Konieczny, et al., 1997). The autonomously replicating sequence (ARS) and centromere (CEN) were included to allow for replication and segregation in *S. cerevisiae* as well as *P. tricornutum* (Karas, et al., 2015). The plasmid also contains an origin of transfer (*oriT*) to facilitate transfer via conjugation to other organisms. Unique restriction enzyme sites were incorporated to enable cloning of DNA

fragments, as well as selective markers to select for the plasmid in each of the organisms listed. Six different versions of the pGE plasmids were constructed previously by the Karas lab (Brumwell et al., 2018, to be published). These plasmids are named: pGE1, pGE3, pGE4, pGE5, pGE6 and pGE7 (see Figure 1) and vary in the *S. meliloti* origin of replication and bacterial antibiotic selection marker found on each plasmid. pGE1, pGE3, and pGE4 contain the *oriV* from pSymA and have spectinomycin/streptomycin, tetracycline, and kanamycin/neomycin resistance markers, respectively. pGE5, pGE6, and pGE7 contain the *oriV* from pSymB and have spectinomycin/streptomycin, tetracycline, and kanamycin/neomycin resistance markers, respectively.



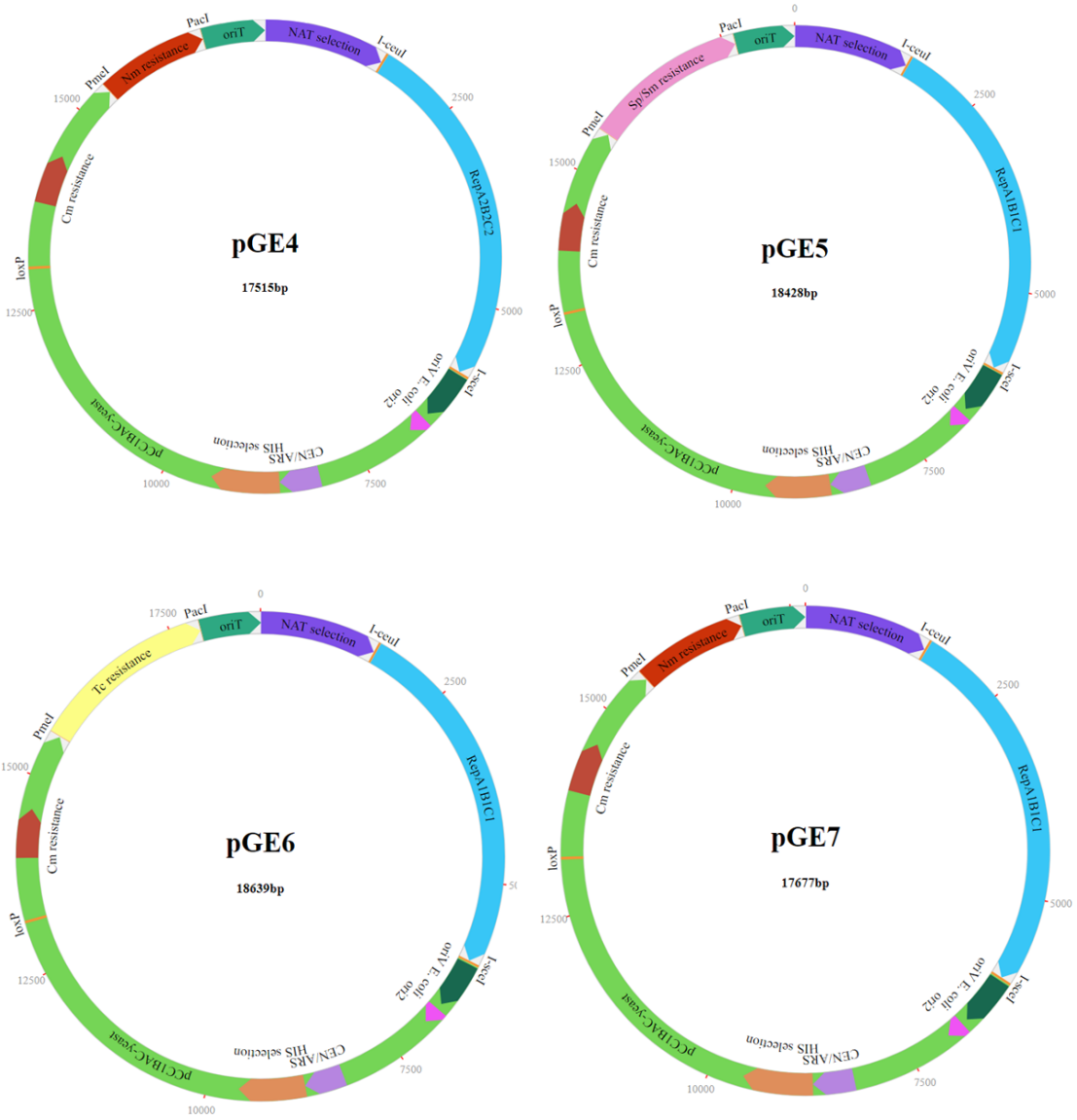


Figure 1: Circular maps of pGE plasmids used in this study which range in size from ~16-18 kb. Each plasmid contains the origin of replication from the pSymA or pSymB replicon (*repA2B2C2* or *repA1B1C1*) along with origins of replication for *E. coli* (*oriV/ori2), S. cerevisiae* yeast (CEN/ARS), and *Phaeodactylum tricornutum* algae (CEN/ARS) to enable replication of this plasmid in each of these organisms. pGE1, pGE3, and pGE4 contain the *repA2B2C2* origin of replication from pSymA, whereas pGE5, pGE6, and pGE7 contain the *repA1B1C1* origin of replication from pSymB. The

centromere (CEN) and autonomously replicating sequence (ARS) were included to allow replication in *S. cerevisiae* as well as *P. tricornutum* (Karas, et al., 2015). The plasmid also contains an origin of transfer (*oriT*) to facilitate transfer via conjugation to other organisms. I-ceuI, I-sceI, PacI and PmeI unique restriction enzyme sites were included to enable cloning of DNA fragments. The nourseothricin (NAT) selective marker was included to allow for selection in *P. tricornutum*, which confers resistance to the broad-use antibiotic nourseothricin. The HIS3 selective marker was included to allow for selection in *S. cerevisiae*, which encodes an enzyme required for the biosynthesis of histidine (Struhl, 1983). An antibiotic selective marker (Sp/Sm, Tc, or Nm) was included to allow for selection in *S. meliloti* and *E. coli.* An additional chloramphenicol antibiotic selective marker is present to allow for selection in *E. coli*.

Due to the *incA* incompatibility region found on the pSymA and pSymB megaplasmids, the pGE plasmid with the corresponding origin of replication will not be able to co-exist within an *S. meliloti* cell containing that replicon (diCenzo, et al., 2014). Incompatibility, here, refers to the presence *cis*- or *trans*-acting elements that interfere with the proper replication and segregation of other plasmids in a cell (Novick, 1987). The plasmids containing a pSymA origin of replication are used in a ΔpSymA background, whereas the plasmids containing a pSymB origin of replication are used in a ΔpSymB background. These origins of replication were used due to their ability to stably maintain and replicate large fragments of DNA in low-copy number, as they permit replication of the pSymA and pSymB megaplasmids. Previous worked performed by Döhlemann et al. (2017), demonstrated the use of the *repABC* origins of replication from *S. meliloti* for use in cloning vectors for synthetic biology purposes, termed pABC (Dohlemann, et al., 2017). Several genetic parts were included in the pABC plasmids, including: *repABC* for

replication in *S. meliloti*, *oriV* for replication in *E. coli*, *oriT* to allow for conjugation of the plasmid, an antibiotic resistance cassette, and a multiple cloning site (MCS) (Dohlemann, et al., 2017). In this present work, one or two pGE plasmids were maintained within a ΔpSymAB Δ*hsdR* minimal strain which will serve as the platform strain to conduct large scale genetic manipulation of large fragments of high G+C content DNA cloned into the pGE plasmids.

1.3.2 *In vivo* Cloning Using pGE Plasmids

To facilitate *in vivo* cloning using the pGE plasmids in the *S. meliloti* host, a "landing pad" sequence was designed and cloned into the PacI site of the pGE4 plasmid (Kearsley and Finan, 2018, unpublished; Designer Microbes Inc., 2018, unpublished). This sequence (see Figure 4B) consists of a minimal *FRT* site, *attB* site, *lox71* site, *EcoRI* restriction enzyme cut site, and an additional minimal *FRT* site and can be used for site-specific recombination (see Section 1.4) of incoming fragments of DNA that contain the corresponding site. Conjugation of a donor plasmid harbouring a site-specific recombination site (either *FRT* or *attP*) can be performed with the ΔpSymAB Δ*hsdR* minimal strain as recipient. In this study, three large plasmids (pA132, pTH3255, and pTH3247) containing regions derived from pSymA and pSymB were conjugated into the ΔpSymAB Δ*hsdR* minimal strain. These plasmids were captured previously in *E. coli*, and these were used as the donor strains.

1.4 Site-Specific Recombination Methods

1.4.1 Flp/FRT Recombination System

Flp recombinase, or flippase, is a 423 amino acid site-specific recombinase belonging to the integrase family of recombinases which works in conjunction with two Flp recognition target (FRT) sites to enable site-specific recombination between fragments of DNA containing these

FRT sites. The system was initially discovered encoded on the 2µm plasmid of *Saccharomyces cerevisiae* where it is used in homologous recombination reactions (Senecoff, Bruckner, & Cox, 1985). The Flp/*FRT* system is a highly versatile tool used by researchers as it is able to function in many types of organisms including bacteria, yeasts, plants, mice and can be used in *in vitro* reactions (Schweizer, 2003; Sadowksi, 1995; Lyznik, Rao, & Hodges, 1996; Vooijs, et al., 1998). FRT sites are 48 bp in length and consist of three 13 bp symmetry elements that surround an 8 bp core which contains an *XbaI* site (see Figure 2).

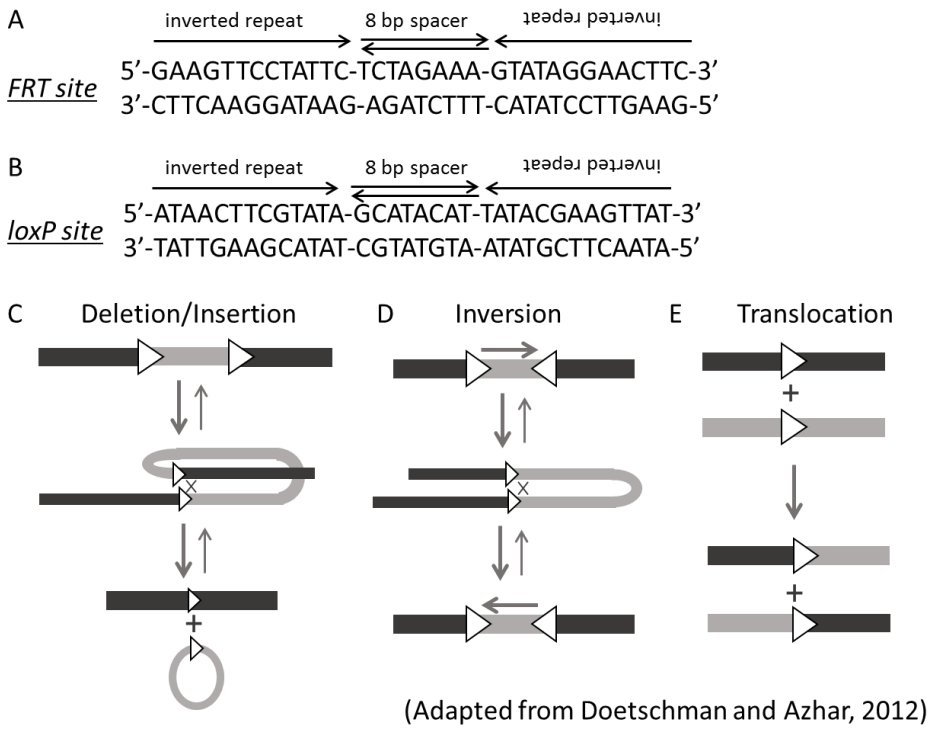


Figure 2: (A) Diagram of the minimal Flp recombination target (*FRT)* site (34 bp) and (B) locus of crossover in P1 bacteriophage (*loxP)* site. Panels (C), (D), and (E) demonstrate the types of site-specific recombination that can occur between pairs of *FRT* or *loxP* sites, which is dependent on

the orientation of the sites. (C) Demonstrates sites that are in direct orientation, (D) demonstrates sites in opposite orientation, and (E) demonstrates sites in direct orientation on different strands of DNA. White triangles denote the directionality of the site-specific recombination site, which is determined by the asymmetry of the core region. The horizontal arrow in (D) denotes the direction of the middle DNA region. Recombination and excision between *FRT* sites is mediated by Flp recombinase and recombination and excision between *loxP* sites is mediated by Cre recombinase. Figure is adapted from Doetschman & Azhar (2012).

The core sequence has a direct effect on the consequence of the Flp-catalyzed reaction. If the core sequences are in direct orientation, the intervening region will be excised during recombination, whereas if the core sequences are in inverted orientation, the sequence will be inverted during the recombination reaction (Schweizer, 2003).

A minimal *FRT* site that lacks one of the 13 bp symmetry elements is often employed by researchers when using the Flp/*FRT* system and is considered a minimal 34 bp *FRT* site. While the Flp recombinase binds to all three symmetry elements, only the two flanking the spacer region are required for cleavage to occur (Andrews, et al., 1985). The first step in the reaction involves binding of the Flp recombinase to the *FRT* site and cleaving of the phosphodiester bond adjacent to the 8 bp core region. This results in the formation of a transient Holliday junction which is then resolved, completing the recombination reaction (Sadowksi, 1995). It has been shown that an increase in size by as little as 2 bp in the core region can cause inactivation of the site (Senecoff, Bruckner, & Cox, 1985).

An analogous system called Cre/*loxP* was discovered in Bacteriophage P1, which works in a similar site-specific manner, although the recombinase only shares a weak amino acid

sequence similarity (Argos, et al., 1986; Gopaul, Van Duyne, & Guo, 1999).Here, the Cre recombinase recognizes the *loxP* site, which is composed of two 13 bp inverted repeats and an 8 bp spacer region, akin to the *FRT* site (see Figure 2 and Araki & Yamamura, 2012). Cre/*loxP* along with Flp/*FRT* systems has been used extensively by researchers to excise, invert, and integrate DNA fragments in both *in vivo* and *in vitro* experiments. They are simple two-component recombination systems that do not require the addition of co-factors (Sadowksi, 1995). Furthermore, mutant sequences of both *FRT* and *loxP* have been identified and can be utilized for unidirectional recombination of DNA fragments. These sequences have mutations in the repeat regions flanking the spacer region allowing unidirectional recombination to occur (Albert, et al., 1995).

1.4.2 *attB/attP* Recombination System

Another site-specific recombination system frequently used is the *attB/attP* (attachment) system. This system consists of an integrase gene belonging to the family of serine recombinase superfamily and allows for site-specific recombination between two recognition sequences, *attB* and *attP,* and was first discovered in the bacteriophage lambda (λ) (Landy & Ross, 1977). Unlike the Cre/*loxP* and Flp/*FRT* systems, where the recombinase enzymes mediate integration and excision recombination events, the *attB/attP* system is a unidirectional system that recombines two non-identical sequences that produces hybrid sites termed *attL* and *attR* following recombination. The first step in this process is the attachment of the integrase enzyme to the *attB* and *attP* sequences, forming a tetrameric complex where cleavage of the DNA strands then occur. Strand exchange then occurs, and new attachment sites are generated (Rutherford & Van Duyne, 2014). The result is the integration of the incoming DNA fragment flanked by *attL* and

attR sites into the recipient DNA fragment (see Figure 3). An advantage of serine integrases is the small size of the recognition sites that do not require additional accessory proteins to allow for proper functioning, unlike those of tyrosine integrases, another family of integrase enzymes (Merrick, Zhao, & Rosser, 2018). In this study, the $attB_{53}$ and $attP_{50}$ sites were used (Groth, et al., 2000).

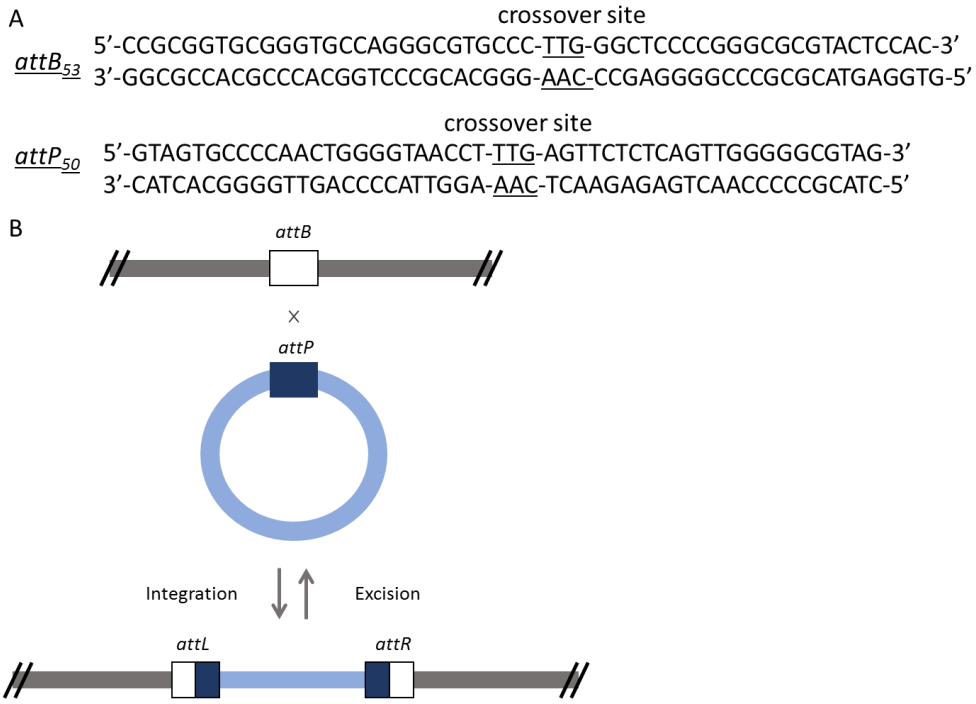


Figure 3: (A) Diagram of the attachment site on bacteria (*attB)* and attachment site on phage, which are site specific recombination sites. (B) Demonstrates the recombination event between an *attB* site located on the chromosome of a bacterial strain and an *attP* site located on a plasmid. The white box denotes the *attB* site and the blue box denotes the *attP* site. Upon recombination, two hybrid sites, *attL* and *attR* are formed and the plasmid DNA is incorporated into the bacterial chromosome. The integration reaction is mediated by the Integrase enzyme and the excision reaction is mediated by a separate Exicase enzyme.

1.5 This Work

The aim of this work is to characterize the use of a minimal *S. meliloti* ΔpSymAB Δ*hsdR* strain as a host for cloning of large fragments (0.1 Mb-1 Mb) of high G+C (>60%) content DNA. A major problem observed in the field of synthetic biology today is the lack of genetic tools to manipulate DNA from organisms with high G+C content DNA, which are relevant in many industrial applications. Current "workhorses" for genetic manipulation include *E. coli* and *S. cerevisiae*. Yeast is used as a host due to its ability to recombine large fragments of DNA, but it lacks the ability to directly transfer DNA fragments to other organism and is unable to replicate DNA with >40% G+C content without additional engineering. *S. meliloti* was chosen here due to its natural ability to maintain large DNA fragments as megaplasmids or chromids and its high G+C content DNA of 62.7%. Use of the *repABC* origins of replication from pSymA and pSymB in the pGE plasmids will facilitate the use of these plasmids for use in site-specific recombination of large DNA fragments of high G+C content DNA. A minimal *S. meliloti* strain lacking the pSymA and pSymB replicons, as well as the Type I restriction endonuclease *hsdR*, was constructed to provide a chassis or host strain to facilitate bottom-up genomic engineering of large DNA fragments. Multi-host shuttle vectors, termed pGE plasmids, were created and designed to replicate and be selected for in several host strains including *S. meliloti, E. coli, S. cerevisiae and P. tricornutum.* Optimized transformation protocols were developed to characterize this strain when pGE plasmids were transformed. Stability assays were then conducted in the *S. meliloti* strain to determine its ability to maintain the plasmids either individually or when two pGE plasmids are present in the same strain (see Figure 4A).

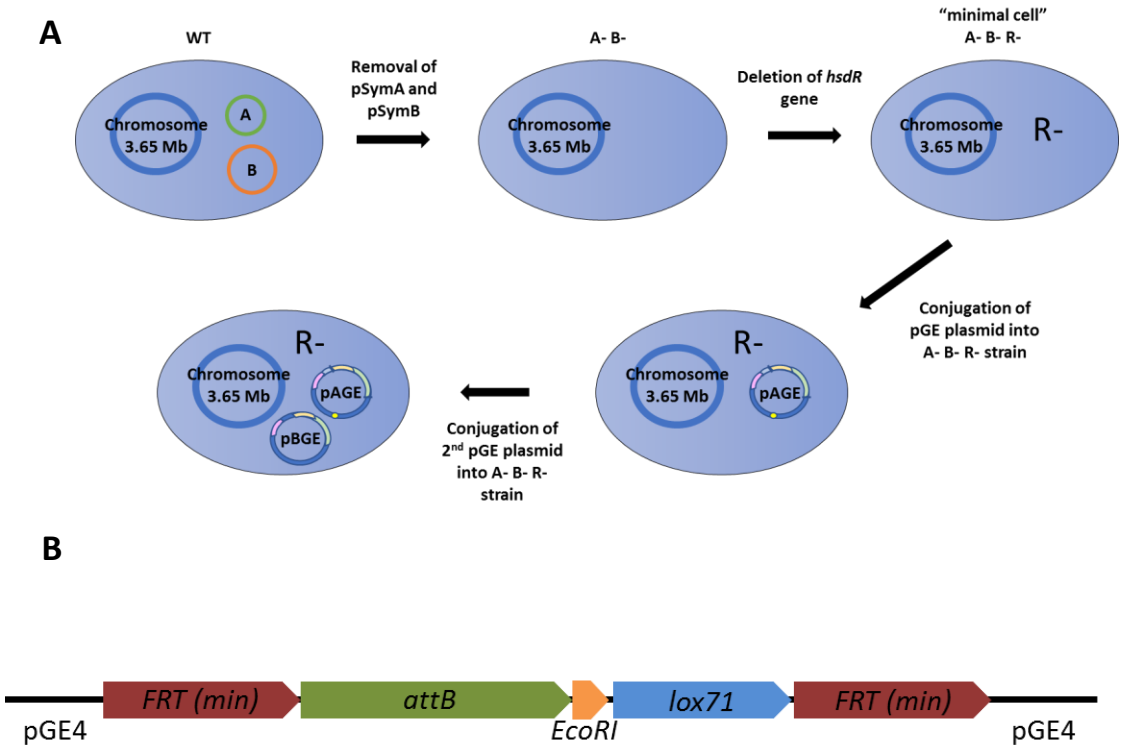


Figure 4: (A) Diagram depicting the techniques used to develop the *S. meliloti* "minimal cell". The minimal cells lack the pSymA and pSymB replicons as well as the endogenous restriction endonuclease *hsdR* gene. Following this, pGE plasmids, large, recombinant plasmids developed to allow for cloning and further manipulation of large fragments of high G+C content DNA, can be conjugated into the strain. These pGE plasmids contain the origin of replication for either pSymA or pSymB (denoted as pAGE or pBGE, respectively). The final strain is ΔpSymAB Δ*hsdR* and contains the pAGE and pBGE plasmids. (B) Diagram depicting the "landing pad" construct cloned into the PacI site of the pGE4 plasmid to enable site-specific recombination between large plasmids with high G+C content DNA that contain *FRT* or *attP* sites. This design was cloned into the PacI site on pGE4 and consists of a minimal *FRT* site, an *attB* site, an EcoRI restriction enzyme recognition site, a *loxP* mutant called *lox71*, and another minimal *FRT* site.

Furthermore, as a proof-of-concept experiment, three large plasmids comprised of DNA derived from the *S. meliloti* pSymA and pSymB replicons, which are ~70 kb, ~170 kb and ~205 kb in size, were conjugated into a minimal strain containing the pGE4 plasmid with a cloned "landing pad" sequence. This landing pad contains site-specific recombination sites such as *FRT, lox71* and *attB* sites to allow for integration of the large fragments into the pGE plasmids (see Figure 4B). Capture of the large fragment plasmids containing either an *FRT* site or an *attP* site into the pGE4 plasmid was performed and verified. This demonstrates the ability of the pGE plasmids to stably maintain large fragments of high G+C content DNA. Using this work as a basis, the *S. meliloti* minimal cell with pGE plasmids can be continued to be used for cloning of high G+C content DNA from commercially and industrially relevant organisms.

2 Materials and Methods

2.1 Bacterial Strains and Plasmids

The bacterial strains and plasmids used in this work are listed in Table 1. The genotype of relevant characteristics as well as the reference, source or methods of construction are indicated for each. Plasmids were stored at -20°C in $T_{10}E_1$ (10mM Tris-HCl, 1mM EDTA pH 8.0) or ddH_20 (sterile, distilled deionized water), while all bacterial strains were stored in glass vials as frozen stocks at -80°C in LB containing 7% dimethylsulfoxide (DMSO). These frozen stocks were made by adding 0.75mL of LBmc liquid bacterial cultures, grown overnight, with an equal volume

of 14% DMSO. Using sterile inoculation sticks, viable cells were recovered from frozen permanent cultures by streaking onto appropriate selective solid media.

Table 1: Bacterial strains and plasmids used in this project.

Sinorhizobium meliloti

Strain	Description	Resistance	Reference
Rm2011	Wild type SU47 *str-3*	Sm^r	(Scherrer & Denarie, 1971)
Rm5000	SU47 *rif-5*	Rif^r	(Finan, et al., 1984)
RmP110	RmP1021, wild type SU47 *str-21* with wild-type *pstC* allele	Sm^r	(Yuan, Zaheer, & Finan, 2006)
RmP1684	RmP110 with ΦC31 integrase on chromosome	Sm^r Nm^r	Fink and Finan, 2010, unpublished
RmP3498*	ΔpSymA P110	Sm^r	(diCenzo, et al., 2014)
RmP3909*	ΔpSymA ΔpSymB with *engA-tRNA*^{arg}*-rmlC* moved to chromosome; *bacA* gene replaced with *bacA* from Rm2011	Sm^r	(diCenzo, et al., 2017)
RmP3952*	RmP3909 $pSymA^+$ ΔpSymB	Sm^r	(diCenzo, et al., 2017)
RmP3953*	RmP3909 ΔpSymA $pSymB^+$	Sm^r	(diCenzo, et al., 2017)
RmP3954*	RmP3909 $pSymA^+$ $pSymB^+$	Sm^r	(diCenzo, et al., 2017)
RmP3975	RmP110 (pTH3144) Δ*hsdR*	Sm^r Nm^r	(Zamani and Finan, 2016, unpublished)
RmP4098*	RmP3950 ΔpSymA *hsdR*::Nm	Sm^r Nm^r	(Finan, 2016, unpublished)
RmP4122*	RmP4098 ΔpSymA *hsdR*::FRT	Sm^r Nm^s	(Finan, 2016, unpublished)
RmP4124	RmP3953 Δ*hsdR*	Sm^r Nm^s	This work
RmP4125	RmP3954 Δ*hsdR*	Sm^r Nm^s	This work
RmP4246	Rm5000 ΔpSymA	Rif^r	(Bindra, Sather, and Finan, 2017, unpublished)
RmP4251	RmP4246 *hypRE*::*FRT*-Km^r-*FRT*	Rif^r Nm^r	(Situ and Finan, 2017, unpublished)
RmP4254	RmP4251 Δ*hypRE*::*FRT* with pTH2505	Sm^r Tc^r Nm^s	(Situ and Finan, 2017, unpublished)
RmP4258	RmP3909 Δ*hsdR*	Sm^r Nm^s	This work
RmP4260	RmP3952 Δ*hsdR*	Sm^r Nm^s	This work
RmP4262	RmP4246 Δ*hsdR*	Rif^r Nm^s	This work
RmP4279	RmP4258 (pGE1)	Sm^r Sp^r	This work
RmP4280	RmP4258 (pGE3)	Sm^r Tc^r	This work
RmP4281	RmP4258 (pGE4)	Sm^r Nm^r	This work
RmP4282	RmP4258 (pGE5)	Sm^r Sp^r	This work

RmP4283	RmP4258 (pGE6)	Sm^r Tc^r	This work
RmP4284	RmP4260 (pGE5)	Sm^r Sp^r	This work
RmP4285	RmP4260 (pGE6)	Sm^r Tc^r	This work
RmP4305	RmP4254 (*FRT*::pA132) at *hypRE* locus	Rif^r Gm^r Nm^r Tc^s	(Sather, Bindra, and Finan, 2017, unpublished)
RmP4306	RmP4254 (*FRT*::[A117-A118-A301-A121]) at *hypRE* locus	Rif^r Gm^r Nm^r	(Situ, and Finan, 2017, unpublished)
RmP4310	RmP4258 (pGE3 and pGE5)	Sm^r Sp^r Tc^r	This work
RmP4311	RmP4258 (pGE4 and pGE5)	Sm^r Sp^r Nm^r	This work
RmP4312	RmP4258 (pGE4 and pGE6)	Sm^r Nm^r Tc^r	This work
RmP4313	RmP4258 (pGE3 and pGE7)	Sm^r Tc^r Nm^r	This work
RmP4314	RmP4258 (ΦC31 integrase on chromosome pGE4-LP)	Sm^r Nm^r	This work
RmP4315	RmP4258 (pGE4-LP and pA132)	Sm^r Nm^r Gm^r	This work
RmP4316	RmP4258 (pGE4-LP and pTH3255)	Sm^r Nm^r Gm^r	This work
RmP4317	RmP4258 (pGE4-LP *attB::*pTH3247)	Sm^r Nm^r	This work

Escherichia coli

Strain	Description	Resistance	Reference
DH5α	F-*endA1 hsdR17 supE44 thi-1 λ- recA1 gyrA96 relA1 Φ80dlacZΔM15*	-	(Hanahan, 1983)
M928	Rif resistant derivative of DH5α	Rif^r	Kahn Laboratory Collection
MT616	MT607 (pRK600); contains RK2 transfer genes	Cm^r	(Finan, et al., 1986)
M1449	DH5α (pTH2505)	Tc^r	(White, et al., 2012)
Epi300	*F⁻ mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ⁻ rpsL nupG trfA dhfr*	Sm^r	(TransforMax™ EPI300™, Lucigen, 2016)
M2593	M928 (pTH3247)	Rif^r Km^r	(Kearsley and Finan, 2018, unpublished)
M2608	DH5α (pTH3255)	Rif^r Gm^r Km^r	(Huang and Finan, 2018, unpublished)

Plasmids

Plasmid	Description	Resistance	Reference
pA132	A132 region from *S. meliloti* Rm2011 flipped out via FRT sites	Km^r Gm^r	(Sather and Finan, 2018, unpublished)
pGE1**	MHS shuttle vector containing origins of replication for *S. meliloti, E. coli, S.*	Cm^r Sm/Sp^r	(Brumwell, et al., 2018, to be published)

	cerevisiae and *P. tricornutum.* Contains *oriT*, Sm/Sp resistance gene, unique cut sites. Contains pSymA origin of replication.		
pGE3**	Derivative of pGE1 containing Tc resistance genes.	Cm^r Tc^r	(Brumwell, et al., 2018, to be published)
pGE4**	Derivative of pGE1 containing Km/Nm resistance genes.	Cm^r Nm^r	(Brumwell, et al., 2018, to be published)
pGE5**	Derivative of pGE1 containing Sm/Sp resistance genes and pSymB origin of replication.	Cm^r Sm/Sp^r	(Brumwell, et al., 2018, to be published)
pGE6**	Derivative of pGE5 containing Tc resistance genes and pSymB origin of replication.	Cm^r Tc^r	(Brumwell, et al., 2018, to be published)
pGE7**	Derivative of pGE5 containing Km/Nm resistance genes and pSymB origin of replication.	Cm^r Nm^r	(Brumwell, et al., 2018, to be published)
pGE4-LP**	Derivative of pGE4 containing a landing pad sequence consisting of *FRT-attB-EcoRI-lox71-FRT* cassette‡	Cm^r Nm^r	(Kearsley and Finan, 2018, unpublished) (Designer Microbes Inc., 2018, unpublished)
pGE4-LPE	Derivative of pGE4-LP where the region between the *FRT* sites was excised via Flp-catalyzed recombination, resulting in a single *FRT* site	Cm^r Nm^r	This work
pRK600	Mobilizing plasmid containing *tra* transfer genes	Cm^r	(Finan, et al., 1986)
pRK7813	Mobilizable cosmid vector	Tc^r	(Jones & Gutterson, 1987)
pTA-Mob	An Rk2- based mobilizer plasmid containing *tra* transfer genes and a non-functional *oriT*	Gm^r	(Strand, et al., 2014)
pTH2505	Expresses *flp recombinase* under control of PCA-inducible promoter	Tc^r	(White, et al., 2012)
pTH2623	pTrc-SC (nt 1248755 – 1249733) via *PacI*	Sm^r Sp^r	(Milunovic, et al., 2014)
pTH3247	B108 B109 B123 regions from *S. meliloti* Rm2011 flipped out via FRT sites	Km/Nm^r	(Kearsley and Finan, 2018, unpublished)
pTH3255	A117 A118 A121 regions flipped out from *S. meliloti* Rm2011 via FRT sites	Gm^r Km/Nm^r	(Huang and Finan, 2018, unpublished)
pTH3264	pGE4-LP *attB::*pTH3247. Further confirmed via sequencing of PCR product	Cm^r Nm^r	This work

Unless stated otherwise, all strains and plasmids were constructed in this study. ^{r} designates resistance, ^{s} designates sensitivity. The symbol Φ preceding a strain name indicates a ΦM12 transducing lysate.

Abbreviations are as follows: Cm, chloramphenicol; Gm, gentamicin; Nm, neomycin; Km, kanamycin; Sm, streptomycin; Rif, rifampicin; Sp, spectinomycin; Tc, tetracycline; *FRT*, Flp recognition target; *oriT*, origin of transfer; *oriV*, origin of replication

*The construction of these *S. meliloti* strains was completed with the help of T.M. Finan and G.C. diCenzo**. The construction of the pGE plasmids was completed by B.J. Karas‡. The design of this landing pad was completed by J. Kearsley and the construction of the pGE4 plasmid containing the landing pad cassette was completed by Designer Microbes Inc. (DMI).

2.2 Oligonucleotides

Primer sequences used in this study are listed in Table 2. All primers were ordered from Integrated DNA Technologies Inc. (IDT) through the Mobix McMaster portal or from Sigma Aldrich through the Mobix McMaster portal. Primers were ordered as lyophilized stocks and resuspended in ddH_2O to a concentration of 100µM. Working stocks were then made to a final concentration of 10µM and used in polymerase chain reaction (PCR) or sequencing procedures.

Table 2: Primers used in this study.

Primer Name	Sequence (5' to 3')	Reference
hsdM U Fwd	CAAAGGACGGCACTTATTCAACGG	(Zamani and Finan, 2016, unpublished)
hsdM U Rev	AGCCGGACGATGGACGAGG	(Zamani and Finan, 2016, unpublished)
hsdR U Fwd	GACATAGCATCCGTATCAGTTGGG	(Zamani and Finan, 2016, unpublished)
hsdR U Rev	CATCCCGCCGTCATCGTCG	(Zamani and Finan, 2016, unpublished)

hsdR D Fwd	ATAATGCCGGAAGCCTGACCG	(Zamani and Finan, 2016, unpublished)
hsdR D Rev	GATCAAGCTACTCGATCAGGTGC	(Zamani and Finan, 2016, unpublished)
*sma2245*_F	CACTGGCAAAGTCGCAATCTATCC	(Muhammed and Finan, 2018, unpublished)
*sma2245*_R	TAGGAAGAGGTTGGGAAGCACC	(Muhammed and Finan, 2018, unpublished)
A118_check-II-F	AGGCTTTCCTCCATGTCCATCAGC	(Huang, Muhammed and Finan, 2018, unpublished)
A121_check-II-R	CGACAGGAACTTCCGGGGATTGG	(Huang, Muhammed and Finan, 2018, unpublished)
1142-IF	ATCGTCTGCCTTGAGGACGG	(Milunovic, 2011)
1142-IR	TTCCAGCAGCAACGAGACCG	(Milunovic, 2011)
1860-OF	GCGCAACGATCTCTTCACCC	(Milunovic, 2011)
1860-OR	CTTGCTGGACGTACTGACGC	(Milunovic, 2011)
ML-07-245	GTGGCGTGAGATCAGACGC	(Fink and Finan, 2010, unpublished)
ML-07-246	CGCTTGTCGTCTACTGACGC	(Fink and Finan, 2010, unpublished)
2611_UF	ACGACAACCTCTCCAACGGC	(Milunovic, 2011)
2611_UR	TCCCAGCCTAGAACCTGACG	(Milunovic, 2011)
HR_1_Fwd	GGATCCCGGATGATGTCGAGCACGG	(Kearsley and Finan, 2017, unpublished)
HR_2_Rev	GGTACCGCACGGTCAGCCATAATTTGCG	(Kearsley and Finan, 2017, unpublished)
916_DF	TTGACCTCTTCGAGCCTCGG	(Milunovic, 2011)
916_DR	GAAACTGATGGCAGCGACGG	(Milunovic, 2011)
pGE4_check_Fwd	TGAGCAGATGGCTGCATTATCC	This work
pGE4_check_Rev	TACGGGGCAACCTCATGTCC	This work
new check_primer_Fwd	ATCCTTGGCGGCAAGAAAGC	This work
pGE1_check_Rev	GGAAAGTCTACACGAACCCTTTGGC	This work

check_primer1_Rev	CTTACAGACAAGCTGTGACCG	This work
check_primer2_Rev	CGGTTCAAAGAGTTGGTAGCTC	This work
check_primer3_Rev	GACGTGCTTGGCAATCACG	This work

2.3 Bacterial Culturing and Microbiological Techniques

Bacterial strains were cultured in liquid media in sterile test tubes in a rolling drum at 200 rpm. *E. coli* strains were grown at 37°C and *S. meliloti* strains were grown at 30°C. Strains were usually grown for 16 hours overnight or up to 30 hours for slow-growing strains. Strains containing plasmids were cultured with the appropriate antibiotics(s) in the media to allow for selection of strains harbouring the plasmid(s).

2.3.1 Bacterial Growth Media

Antibiotics were stored as concentrated stock solutions at -20°C in H_2O, ethanol, ethanol/H_2O, methanol or dimethylformamide. Aqueous stock solutions were filter sterilized. The final concentrations for antibiotics in media are given in Table 3. IPTG was used in solid media at 0.5mM final concentration.

Table 3: Antibiotic concentrations (µg/mL) used for both *E. coli* and *S. meliloti* in both solid and liquid media.

Antibiotic	**Concentration (µg/mL) for *E.coli***		**Concentration (µg/mL) for *S. meliloti***	
	Solid media	**Liquid media**	**Solid media**	**Liquid media**
Chloramphenicol	5	2.5	--	--
Kanamycin monosulphate	25	12.5	--	--
Gentamycin sulfate	10	5	60	30
Neomycin sulfate	--	--	200	100
Rifampicin	50	25	20	10
Spectinomycin	100	50	100	50

dihydrochloride				
Streptomycin sulfate	200	100	200	100
Tetracycline hydrochloride	5	2.5	5	2.5

Complex media used for the growth of bacteria was LB (Miller, 1972)and was made using: bacto tryptone (10g/L), yeast extract (5g/L), NaCl (5g/L). Solid media was prepared with the addition of Difco bacto agar (15g/L). Liquid cultures of *S. meliloti* were grown in LB + 2.5mM $MgSo_4$ + 2.5mM $CaCl_2$ (hereby referred to as LBmc). When growing strains lacking pSymA, 38µM $FeCl_3$ was added to the media and when growing strains lacking pSymB, 2µM $CoCl_2$ was added to the media (hereby referred to as FeCo). Liquid cultures of *E. coli* were grown in LB. Minimal media or defined media was prepared from a commercially available premixed powder (Difco M9 salts). $MgSO_4$, $CaCl_2$, $CoCl_2$, and biotin were added to final concentrations of 2.5mM, 1.25mM, 10ng/mL, 1µg/mL, respectively. $FeCl_3$ was also added at a concentration of 38µM, if required. Carbon sources were added to final concentrations of 10mM (glucose, hydroxyproline, trigonelline). For sufficient aeration, 3-5mL cultures were incubated in autoclaved test tubes on a rotary mixer (New Brunswick Scientific Co.). For larger volumes, cultures were incubated in autoclaved flasks on a floor rotary shaker (New Brunswick Scientific Co.).

2.3.2 Preparation of Bacterial Frozen Stocks

Cultures were grown in the appropriate growth media with antibiotics added, if necessary. The culture was mixed in a 1:1 ratio in a cryogenic glass vial with an LB media containing 14% DMSO for a final DMSO concentration of 7%. The vial was stored at -80°C.

2.4 Bacterial Conjugations

Conjugations were performed by combining equal volumes (25µl) of mid-log phase cultures of donor, recipient and helper or mobilizer strains. The helper strain used in all conjugations was MT616, which carries the Cm^r pRK600 plasmid which provided the transfer gene functions *in trans.* 1mL of each culture was washed twice with 0.85% NaCl and resuspended in 0.5-1mL of 0.85% NaCl and spotted on LBmc and incubated overnight at 30°C. The mating spots were resuspended in 1mL of 0.85% NaCl and dilutions were plated on media selecting for transconjugants. Once plated, cells were incubated in 30°C for 3-4 days for *S. meliloti* or in 37°C for 1 day for *E. coli* transconjugants. Resulting transconjugants were colony purified 3 times by streaking onto selective media. Conjugation frequency was expressed as the number of colony forming units (CFU) per CFU recipient cell.

2.5 ΦM12 Transductions

Bacteriophage ΦM12 was used to transduce mutations from donor *S. meliloti* strains into recipient strains (Finan, et al., 1984).

2.5.1 Preparation and Testing of ΦM12 Transducing Lysates

Lysates were prepared from log-phase *S. meliloti* cultures (3-5mL, OD_{600} of 0.4), to which 100µL of ΦM12 phage (10^9-10^{11} PFU/mL) was added (Finan, et al., 1984). This was incubated on a roller drum overnight, following which 200µL of chloroform was added to kill any remaining viable cells. Lysates were then added to screw-top tubes, centrifuged to pellet any remaining cell debris, and stored at 4°C. The titre of the lysate was determined, in plaque forming units (PFU), by adding 100µL of various dilutions (10^{-8}-10^{-11}) to an overnight culture (OD_{600} of ~0.8) of RmP110. Following adsorption for 15 minutes, 2mL of soft (0.5%) agar was added and the

mixture was poured onto LBmc plates and incubated overnight at 30°C. The number of plaques was counted the next day to determine the titre of the lysate.

2.5.2 ΦM12-Mediated Transduction

Transductions were conducted using ΦM12 (Finan, et al., 1984). 0.5mL of recipient culture grown overnight in LBmc to an OD_{600} of ~1.0 was mixed with an equal volume of 1:25 diluted lysate and incubated for 20 minutes at 30°C. The mixture was then washed two times with 0.85% NaCl and resuspended in 0.5mL of 0.85% NaCl. Recipient alone and lysate alone controls were included. Cells were then plated onto appropriate selective media to select for the antibiotic resistance marker being transduced and incubated at 30°C for 3 days. Resulting transductants were streak purified 3 times onto selective media. Transduction frequency was expressed as the number of CFU observed relative to the number of total CFU recipient cells.

2.6 Molecular Biology Techniques

2.6.1 Agarose Gel Electrophoresis

Agarose gel electrophoresis was used to analyze the size and condition of DNA fragments obtained from various methodologies. 0.8%-1% agarose gels were made and electrophoresis was carried out in 0.5X Tris-Borate-EDTA (TBE) buffer as described by Sambrook and Russell (Sambrook & Russell, 2001). Ethidium bromide was either added to the gel prior to electrophoresis at a concentration of 0.5µg/mL or gels were stained following electrophoresis in a water bath containing ethidium bromide (BioShop Canada) at a concentration 0.1µg/mL for 30 minutes and then washed with ddH_2O for 20 minutes prior to imaging.

Samples were prepared by adding 1/6th volumes of 6X concentrated loading dye to each sample. Loading dye contained bromophenol blue and xylene cyanol dyes, 50% glycerol and ddH_2O. The power supply was set at a constant voltage between 60-100V for 45-60 minutes.

2.6.2 Preparation of Chemically Competent *E. coli* Cells

E. coli cells were made competent using a calcium chloride protocol. *E. coli* DH5α was inoculated overnight and then subcultured into 100mL of LB liquid media to an OD_{600} of ~0.05. Cells were grown on a floor incubator at 37°C until OD_{600} of 0.4 and placed on ice for 10 minutes in pre-chilled conical tubes. Cells were centrifuged at 4400 x g rpm for 10 minutes at 4°C, using a Beckman Coulture Allegra X-22R Centrifuge with SX4250 rotor. The supernatant was then discarded, and the cells were resuspended in 25mL of ice-cold, sterile 100mM $CaCl_2$ solution. Cells were left overnight sitting on ice in a 4°C cold room. The next day, the cells were centrifuged as before and the supernatant discarded, and cell pellet kept on ice. Cells were resuspended in ice-cold, sterile 100mM $CaCl_2$ + 15% glycerol solution and 200µL aliquots were pipetted into microcentrifuge tubes and immediately placed into liquid nitrogen to flash freeze cells. Aliquots were stored at -80°C.

2.6.3 Transformation of Chemically Competent *E. coli* Cells

Competent cells were obtained from the -80°C freezer and mixed with <1µg of DNA and placed on ice for 30 minutes. The cells were then heat shocked by placing the tubes in a 42°C heat block for 45 seconds. Cells were then placed back on ice for an additional 2 minutes. 1mL of LB was then added to the tubes and they were incubated on a rotating drum for 1-2 hours at 37°C, allowing for phenotypic lag. 100µL of diluted transformation mixture was then plated onto selective media. Dependant on the plasmid being transformed, 100µL of undiluted

transformation mixture was also plated onto selective media. Plates were incubated at 37°C overnight. Transformation frequency was expressed as CFU per µg of DNA used.

2.6.4 Preparation of Chemically Competent *S. meliloti* Cells

During this project, an *S. meliloti* PEG-mediated chemical transformation method was developed by the collaborating Karas Lab (Huang and Karas, unpublished). Preparation of chemically competent *S. meliloti* cells were made by first inoculating 5mL cells overnight in LBmc FeCo (if required). The next day, cells were subcultured in 100mL of LBmc FeCo to OD_{600} of ~0.05 and grow at 30°C with shaking until OD_{600} is 0.4. Cells were then transferred to pre-chilled conical tubes and centrifuged at 4400 x g for 10 minutes at 4°C using the abovementioned centrifuge and rotor. Cells were then resuspended in 10mL of ice-cold, sterile 100mM $CaCl_2$ and let sit on ice for 30 minutes. Cells were then centrifuged as before and resuspended in 125µL of ice-cold, sterile 100mM $CaCl_2$ + 15% glycerol. 25µL aliquots were pipetted into microcentrifuge tubes and immediately placed in liquid nitrogen to flash freeze cells. Cells were then thawed immediately to be used for transformation or stored at -80°C until later use.

2.6.5 Transformation of Chemically Competent *S. meliloti* Cells

Cells were first placed on ice to thaw for roughly 20 minutes. Next, 200-500ng of DNA was added to the cells followed immediately by 25µL of 10% PEG 8000. Cells were then incubated on ice for an additional 30 minutes and then heat shocked by placing cells in a 40°C heat block for 8 minutes. 0.5mL of LBmc FeCo was then added to the tubes and incubated in a rolling drum at 30°C for 1.5-2.5 hours, allowing for phenotypic lag. 100µL of transformation mixture was then plated onto selective media. Remaining cells from the transformation mixture were spun down and resuspended in 100µL, which was then plated onto selective media as well.

2.6.6 Preparation of Electro-Competent Cells

S. meliloti cells were prepared according to Ferri et al, (2010). Cells grown overnight in LBmc were subcultured in 100mL of LBmc until the OD_{600} reached 0.6. Cultures were then put on ice for 10-30 minutes and cells were harvested by centrifugation at 4400 x g for 10 minutes at 4°C. The pellet was washed twice with 100mL of ice-cold, sterile ddH_2O and then once with 100mL of ice-cold, sterile 10% glycerol solution, and finally resuspended in 0.5mL of 10% glycerol. 50µL (~10^{10} cells) aliquots were made and pipetted into microcentrifuge tubes and flash frozen with liquid nitrogen and stored at -80°C or used immediately for electroporation.

2.6.7 Electroporation of Electro-Competent Cells

Electrocompetent cells were removed from -80°C freezer and placed on ice to thaw for 30 minutes. 100ng of plasmid DNA was then added to the tube and let sit on ice for 5 minutes. Cells were then quickly added to a pre-chilled, sterile electroporation cuvette (Biorad Canada) with an inter-electrode distance of 0.1cm and electroporated using a Gene Pulser® II Apparatus connected to the Pulse Controller Plus (BioRad Canada) using the following conditions: 25µF, 200Ω, 19-23kV/cm, with a time constant of ~5ms. Immediately following the pulse, 1mL of LB media was added and cells were transferred to a standard sterile test tube and incubated at 30°C with shaking for 2-4 hours, to allow for phenotypic lag and recovery of cells. After incubation, serial dilutions were plated on media selective for the plasmid and incubated at 30°C for 3 days (Ferri, et al., 2010). Transformation frequency was expressed as CFU per µg of DNA used.

2.6.8 Polymerase Chain Reaction

Polymerase chain reaction (PCR) primers were designed using Benchling software, available online (Benchling, 2018). PCR primers (see Table 2) were obtained from the Mobix facility and synthesized by Integrated DNA Technologies, Inc. (IDT). Primers were ordered lyophilized and resuspended in ddH_2O to 100µM. Working stocks of primers were then made by diluting 10X for a final concentration of 10µM. PCR reactions were performed in 10-50µL reaction volumes. Reagents and final concentrations used in the reaction were 200µM each dNTP, 1X PCR Buffer, 50µM $MgCl_2$, 10µM forward and reverse primer, ddH_2O and 2.5U of *Taq* polymerase with either ~1-10ng of plasmid DNA or 10-20ng of genomic DNA (gDNA). A master mix was made prior to each PCR reaction by mixing each of these components to achieve the final concentrations. When available, 2X PCR BestTaq MasterMix (Applied Biological Materials, Inc.) solution was used which contained PCR buffer, dNTPs, $MgCl_2$, and *Taq* polymerase, in which case the BestTaq MasterMix solution was added to primers and ddH_2O, mixed, and then added to PCR tubes containing 5µL of template DNA to make a 1X final concentration. PCR was conducted in the MasterCycler EpCycler S thermocycler (Eppendorf).

PCR programs began with a 5 minute 94°C incubation period and a final 7 minute 72°C incubation period, followed by a hold at 4°C. Each cycle consisted of a 30 second melting set (94°C), a 30 second annealing step and an extension (72°C) step. The annealing temperature was determined based on the melting temperature (T_m) of the primers used in the reaction, normally ~5°C lower than the average primer melting temperature was used. The extension time was determined based on the intended size of the PCR product, which was calculated based on 1 minute/1 kb of PCR product size. PCR reactions were visualized by electrophoresis on a 0.8%

agarose gel alongside a 1 kb ladder (GeneRuler, Thermo Scientific). Loading dye was then added to ~10uL of each PCR sample and was loaded into the gel and agarose gel electrophoresis was performed. Gels were stained with ethidium bromide and visualized with a UV illuminator at 365nm.

2.6.9 Colony PCR

Colony PCR was conducted in the same manner as PCR; however a single colony was used for template DNA instead of purified DNA. A single colony (~1mm in diameter) was resuspended in 20µL of ddH_2O without further dilution. 1uL of resuspended colony was used as template for the PCR reaction. PCR was then performed as stated above and visualized in the same manner.

2.6.10 Restriction Enzyme Digestion of DNA

Restriction endonucleases were used according to manufacturer's recommendations using appropriate buffer (New England BioLabs Inc.). Following restriction enzyme digest, DNA was visualized by electrophoresis on a 0.8% agarose gel in 0.5X TBE buffer. Gels were stained using ethidium bromide and visualized under UV light at 365nm wavelength. If any bands were desired for subsequent cloning or sequencing, bands were excised using a sterile razor blade and DNA was extracted using the EZ-10 Spin Column DNA Gel Extraction Minipreps kit (BioBasic Canada Inc.). Samples were typically resuspended in 20-30µL of Elution Buffer.

2.6.11 Growth Curve Analysis

S. meliloti cells were grown overnight in 5mL at 30°C in LBmc FeCo. The next day, the cells were washed with an equal volume of 0.85% NaCl and then resuspended in 0.85% NaCl. Cultures were then diluted to an OD_{600} of 0.5 in LBmc FeCo. Prior to adding cells to 96-well plate,

15µL of diluted cells was mixed with 135µL of LBmc FeCo in a separate microcentrifuge tube. This mixture was then added to a 96-well plate (Falcon, Corning Brand) and a growth curve protocol was run using the Cytation3 machine (BioTek) using the Gen5 program. Growth curves were then visualized using Gen5 program and data exported to Excel for further analysis.

2.7 DNA Purifications

2.7.1 Miniature Preparation of Bacterial Plasmid DNA

Plasmid DNA was prepared from *S. meliloti* and *E. coli* cells using the EZ-10 Spin Column Plasmid DNA minipreps kit (BioBasic Canada Inc.). 5mL of overnight culture was used in each miniprep procedure. When isolating pGE plasmid DNA from Epi300 *E. coli* cells, 0.1% arabinose was added to overnight cultures to increase the plasmid copy number. Cells were first spun down and resuspended in Resuspension Buffer. Cells were then lysed using Lysis Buffer and let sit for 2 minutes. Neutralization Buffer was then added, and cell debris pelleted at 13000 x g for 5-10 minutes. Supernatant was added to spin columns and spun down at 13000 x g for 2 minutes. Columns were washed 2 times with Wash Buffer containing ethanol and then spun for 2 extra minutes to remove residual ethanol. 30µL of Elution Buffer heated to 60-70°C was then added to the column and let sit for 2 minutes. Elution Buffer was not warmed when isolating small plasmids <10 kb in size. Columns were spun down and eluate added back to the column 2 times to maximize elution of plasmid DNA from the column. DNA concentration was then determined using Gen5 program with the Cytation3 plate imaging spectrophotometer (BioTek) and the concentration was recorded in ng/µL.

2.7.2 Alkaline Lysis Preparation of Bacterial Plasmid DNA

For large scale preparations of plasmid DNA, an alkaline lysis protocol was used to isolate DNA from 50-100mL cultures, as described by Sambrook, et al. (Sambrook, Fritsch, & Maniatis, 1989).

3 Results

3.1 Development of Δ*hsdR* Strains

3.1.1 Transduction of *hsdR* Deletion in Various Background Strains

S. meliloti strain RmP3975 (Δ*hsdR*::Nm^r) was made previously (Zamani and Finan, 2016, unpublished), and here, phage ϕM12 was used to transduce neomycin resistance (i.e. Δ*hsdR*::Nm^r) from RmP3975 into various background strains. Initially the titre of a phage ϕM12 lysate grown on RmP110 (wildtype) was determined to be 1.87 x 10^{10} PFU/mL (see Table 4).

Table 4: Determination of titer of ΦM12 (RmP110) to be used in subsequent transduction experiments.

Dilution of phage (ΦM12 (RmP110))	Number of Plaques Observed
10^{-6}	2030
10^{-7}	187
10^{-8}	21
10^{-9}	3
10^{-10}	0

Using this lysate as inoculums, a phage ϕM12 lysate was then prepared on *S. meliloti* RmP3975 (Δ*hsdR*::Nm^r). This lysate is referred to as ΦM12 (RmP3975) and this was used in subsequent experiments to transduce the neomycin cassette into 5 different strains (Finan, et al., 1984). These strains were RmP3909 (ΔpSymAB), RmP3952 (ΔpSymA pSymB$^+$), RmP3953 (pSymA$^+$ ΔpSymB), RmP3954 (pSymA$^+$ pSymB$^+$), and RmP4246 (ΔpSymA pSymB$^+$ Rif^r). Transductants were plated on LBmc FeCo with neomycin (Nm) (200µg/mL) and enumerated to determine the transduction frequency (see Table 5).

Table 5: Transduction frequency per recipient cell of a neomycin marker into five separate background strains[a]

Donor	Recipient	Transduction frequency
RmP3975	RmP3909 (ΔpSymAB)	5.6 x 10^{-9}
RmP3975	RmP3952 (ΔpSymA pSymB$^+$)	3.3 x 10^{-9}
RmP3975	RmP3953 (pSymA$^+$ ΔpSymB)	2.8 x 10^{-9}
RmP3975	RmP3954 (pSymA$^+$ pSymB$^+$)	3.9 x 10^{-9}
RmP3975	RmP4246 (Rif^r ΔpSymA pSymB$^+$)	4.3 x 10^{-9}

[a] Transductions were performed as described in the Methods section.

$$transduction\ frequency\ per\ recipient = \frac{Number\ of\ transductants/ml}{number\ of\ recipients/ml}$$

$$transduction\ frequency\ per\ recipient = \frac{avg\ \#\ colonies\ X\ DF}{\frac{\#\ colonies\ on\ 10^{-7} plate}{vol.plated}\ x\ DF}$$

$$transduction\ frequency\ per\ recipient = \frac{275\ x\ 10^{0}}{487\ x\ 10^{7}} = 5.6\ x\ 10^{-9}$$

3.1.2 Colony PCR to Verify Deletion of *hsdR* Deletion Strains

Following purification of the neomycin resistant transductants (Δ*hsdR*::Nm^r), the Tc^r plasmid pTH2505 was transferred from *E. coli* to these five strains via a triparental mating. Plasmid pTH2505 is unstable in *S. meliloti* and this plasmid carries the Flp recombinase gene under a protocatechuic acid (PCA) inducible promoter (MacLean, et al., 2011). This was performed in order to excise the neomycin cassette via flanking *FRT* sites, to allow for reuse of the neomycin marker in subsequent experiments. The mating mixture was spotted on LBmc FeCo with 2.5mM PCA. Selection for transconjugants that harbour the pTH2505 plasmid was performed with streptomycin (Sm) (200µg/mL) and tetracycline (Tc) (*5*µg/mL). Transconjugants were then subcultured without antibiotics to lose the unstable pTH2505 plasmid and plated on LBmc FeCo. Colonies were then patched onto LBmc FeCo with Tc (*5*µg/mL) and Tc^s colonies were obtained. All strains were then streak purified 3 times on LBmc FeCo.

To confirm the successful deletion of the *hsdR* gene in the 5 background strains, several diagnostic PCR reactions were performed using 3 separate primer pairs. Figure 5A indicates the primer pairs used together with the structure of the wild type *hsdR* and this locus following excision of the Nm cassette from the Δ*hsdR*::Nm^r strains. The results of the colony PCR can be seen in Figure 5B. The results from the colony PCR reactions supported the successful deletion of the *hsdR* gene.

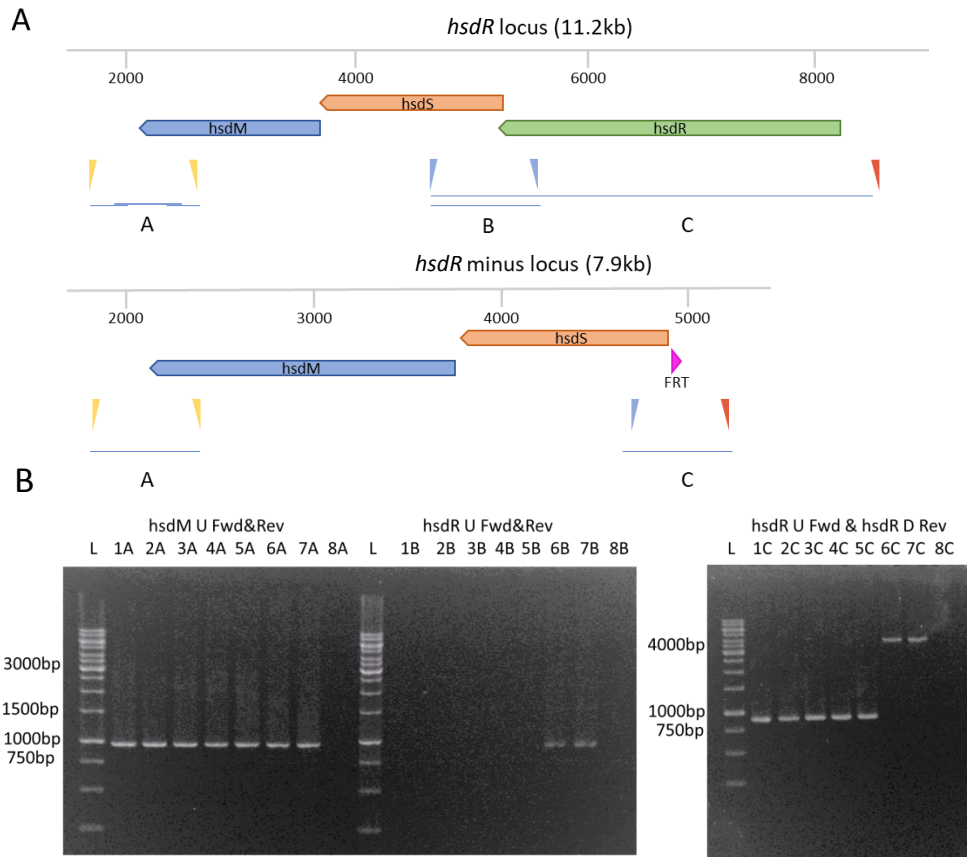


Figure 5: (A) Diagram of *hsdR* locus in the *S. meliloti* genome when *hsdR* is present and when replaced with an Nm cassette flanked by *FRT* sites, which was subsequently excised by Flp recombinase. The pink triangle denotes the *FRT* site following excision of the Nm cassette. Yellow wedges indicate primer set A consisting of hsdM_U_Fwd and hsdM_U_Rev. Blue wedges indicate primer set B consisting of hsdR_U_Fwd and hsdR_U_Rev. Orange wedges indicates reverse primer used in primer set C which was hsdR_D_Rev. (B) Gel image of colony PCR conducted with primer sets A, B, and C on 8 separate samples. These samples are (1) RmP4258, (2) RmP4260, (3) RmP4124, (4) RmP4125, (5) RmP4262, (6) RmP110, (7) isolated gDNA of RmP110 , (8) water alone control. L denotes a 1 kb ladder (GeneRuler 1 kb GTU, ThermoScientific).

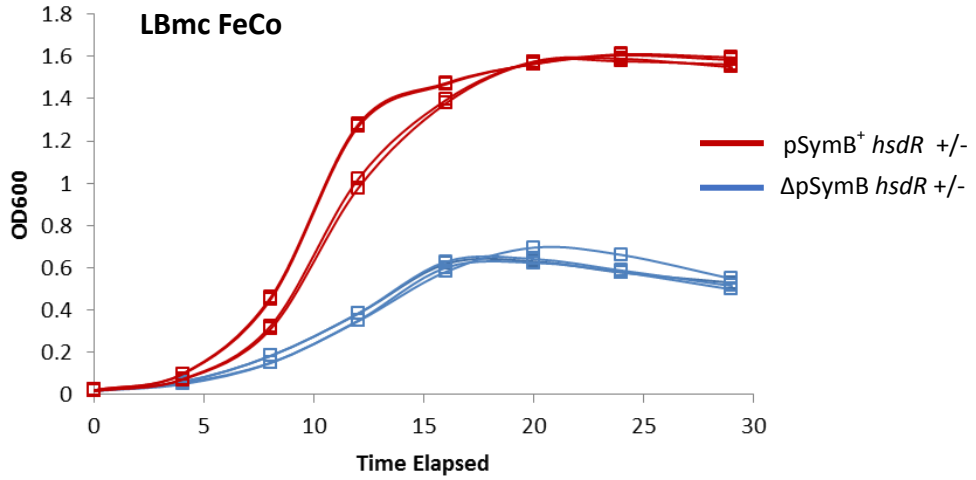
3.2 Characterization of Minimal *S. meliloti* Cell

Following creation of the minimal *S. meliloti* cell, several experiments were conducted to optimize and characterize the use of the strain as a cloning platform to clone large fragments of high G+C content DNA. These experiments involved: determination of growth rate, determination of the transformation and conjugation frequency, Eckhardt gel electrophoresis, and determining the stability of pGE plasmids in RmP4258 minimal cell.

3.2.1 Characterization of Growth Rate of Minimal *S. meliloti* Cell

Previous experiments to determine the growth rate of strains lacking either pSymA, pSymB or both have been conducted by diCenzo et al., (2014). To characterize the growth rate of the strains where the *hsdR* gene was deleted or strains where the pGE plasmids were present, a growth curve was completed, as described in (diCenzo, et al., 2014). Growth curve data was analyzed using the Gen5 program (see Figure 6). The data show that the deletion of *hsdR* does not influence the growth rate of the cultures tested. Generation time was calculated for each strain tested. In Figure 6C and Figure 6D, the data show that strains harbouring the pGE plasmids grow at relatively the same rate when grown in LBmc FeCo, regardless of which pGE plasmid is present, albeit slower than the RmP4258 strain alone. The final OD_{600} reading observed of the RmP4258 strain alone was like that of ΔpSymB strains tested in the previous growth curve experiment.

A

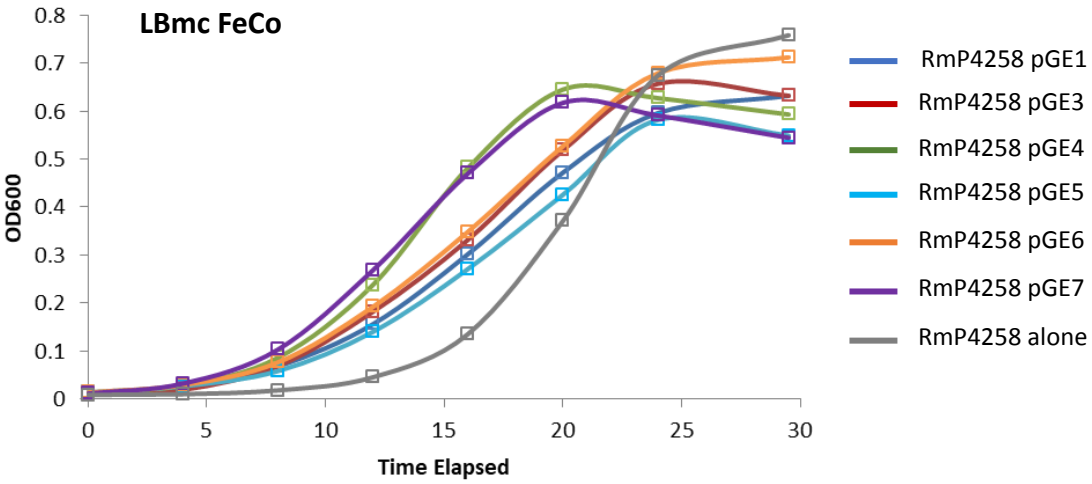


B

Average generation time (hr) (standard deviation)

Medium	ΔpSymAB	ΔpSymAB R⁻	ΔpSymB	ΔpSymB R⁻	ΔpSymA	ΔpSymA R⁻	pSymAB⁺	pSymAB⁺ R⁻
LBmc FeCo	3.6 (0.024)	3.6 (0.005)	3.6 (0.012)	3.6 (0.011)	1.8 (0.005)	1.8 (0.012)	1.9 (0.022)	2.0 (0.012)

C



D

Average generation time (hr) (standard deviation)

Medium	pGE1	pGE3	pGE4	pGE5	pGE6	pGE7	RmP4258 alone
LBmc FeCo	4.8 (0.038)	4.5 (0.045)	3.7 (0.026)	4.8 (0.045)	4.5 (0.040)	3.7 (0.009)	3.2 (0.014)

Figure 6: (A) Growth of *S. meliloti* strains with different combinations of pSymA and *pSymB* with and without the *hsdR* gene present in LBmc FeCo medium. Red lines represent *Δ*pSymB strains with and without *hsdR.* Blue lines represent $pSymB^+$ strains with and without *hsdR*. Data points represent averages from triplicate samples. (B) Average generation times for each sample in (A) with standard deviation. (C) Growth curve analysis of strains harbouring different pGE plasmids relative to a strain alone. (D) Average generation times for each sample in (C) with standard deviation. Generation times were calculated using a Perl script as described by diCenzo et al., 2014, and were calculated during exponential growth phase between OD_{600} of 0.1 and 0.5. Refer to Figure 25 in Appendix.

3.2.2 Characterizing Transformation Protocols

To develop a strain that will be useful as a chassis for cloning large fragments of DNA, development and optimization of high efficient transformation protocols was conducted. Two main transformation protocols were explored: electroporation (Ferri, et al., 2010) and PEG-mediated chemical transformation (Huang and Karas, 2017, unpublished). Electroporation is a transformation technique that utilizes high voltage electrical fields to permit the uptake of DNA into bacterial cells through the cell wall and outer membranes (Ferri, et al., 2010). PEG-mediated chemical transformation is a transformation technique that uses $CaCl_2$, polyethylene glycol (PEG) and a "heat shock" step to permit the uptake of DNA into competent cells. Ca^{2+} ions partially disrupt the cell membrane, which allows DNA to enter the host cell. PEG acts as a crowding agent, and this increases the concentration of DNA in close contact with the cell membranes. With the addition of heat, the cell membrane is disrupted and DNA is able to enter the cell cytoplasm (Chan, et al., 2013).

The electroporation protocol was optimized by independently changing various parameters such as: the preparation of electro-competent cells, the voltage used to electroporate cells, the amount of DNA added to electro-competent cells, the inter-electrode distance of the cuvettes used, and the length of time used to recover cells prior to plating. Despite numerous attempts to electroporate *S. meliloti* cells, very few successful transformants were observed (data not shown), and other transformation techniques were explored.

Following this, a PEG-mediated chemical transformation protocol was developed and optimized by a collaborating lab (Huang and Karas, 2017, unpublished). The optimization of this protocol was done by independently changing various parameters such as: the OD_{600} of cells prior to making cells chemically competent, flash freezing cells prior to transformation with DNA, the temperature and length of time at which the cells were "heat shocked", and the molecular weight and concentration of PEG solution used.

To characterize the transformation efficiency of wild type and *ΔhsdR* mutant cells, plasmids pGE4 or pGE6 were transformed into these cells using the PEG-mediated transformation protocol (see Figure 7). A 5-16-fold increased transformation frequency was observed in the strains where *hsdR* was deleted.

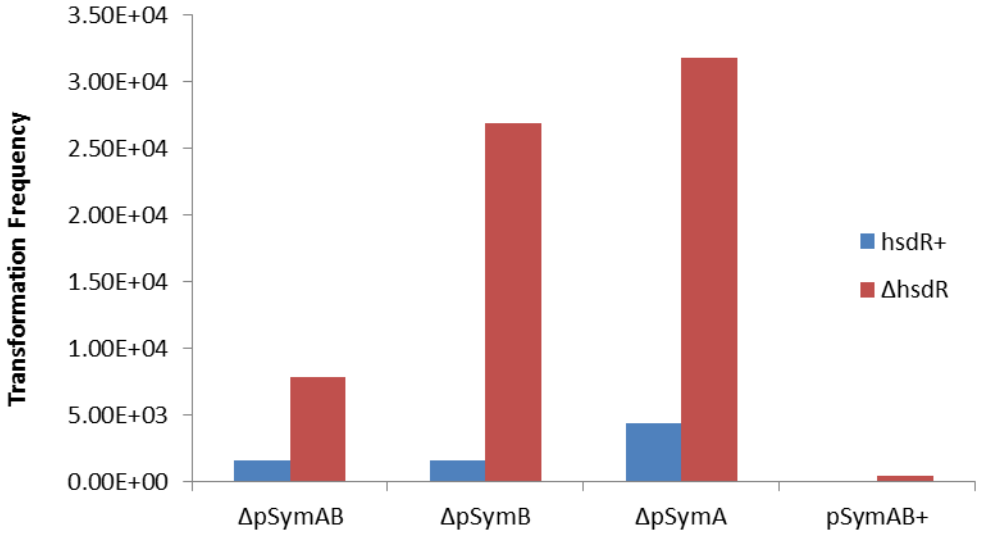


Figure 7: Transformation of strains containing wild type and *ΔhsdR* strains with pGE4 or pGE6 plasmids using a PEG-mediated chemical transformation protocol. Data represents the average of triplicate samples. RmP3909 (ΔpSymAB), RmP4258 (ΔpSymAB *ΔhsdR*), RmP3953 (ΔpSymA pSymB$^+$), RmP4124 (ΔpSymA pSymB$^+$ *ΔhsdR*), RmP3954 (pSymAB$^+$), and RmP4125 (pSymAB$^+$ *ΔhsdR*) were transformed with pGE4 and RmP3952 (pSymA$^+$ ΔpSymB), and RmP4260 (pSymA$^+$ ΔpSymB *ΔhsdR*) were transformed with pGE6.

3.2.3 Conjugation of pGE plasmids into Minimal *S. meliloti* Cell

Aside from transformation and transduction, bacterial conjugation is another method of gene transfer. Conjugation involves direct cell to cell contact to transfer genetic material and was used here to assess the conjugation efficiency of mobilizing the pGE plasmids into minimal *S. meliloti* cells (see Table 6). Triparental matings were performed by mixing cultures of *E. coli* DH5α containing the pGE plasmids. *E. coli* MT616 containing the chloramphenicol (Cm) resistance pRK600 as the mobilizing plasmid and the recipient strain RmP4258. Plasmid pRK600 contains the *tra* genes, which facilitate the transfer of donor plasmids via the *oriT* sequence

(Finan, Kunkel, De Vos, & Signer, 1986). Mating mixtures were spotted on LBmc FeCo and selection for transconjugants that harbour the pGE plasmids was performed with Sm (200µg/mL) and either spectinomycin (Sp) (100µg/mL), Nm (200µg/mL), or Tc (5µg/mL), depending on the pGE plasmid conjugated. Data show that the pGE plasmids transferred to RmP4258 at a frequency of ~10^{-4}/recipient. Selected transconjugants were then colony purified 3 times on LBmc FeCo with the appropriate antibiotic for further experiments.

Table 6: Conjugation frequency of pGE plasmids into RmP4258, a strain lacking pSymA, pSymB and *ΔhsdR.*

Donor Strain	Recipient Strain	Conjugation frequency
DH5α pGE1	RmP4258	2.89×10^{-4}
DH5α pGE3	RmP4258	4.19×10^{-4}
DH5α pGE4	RmP4258	1.06×10^{-3}
DH5α pGE5	RmP4258	2.76×10^{-4}
DH5α pGE6	RmP4258	2.47×10^{-5}
DH5 α pGE7*	RmP4258	6.67×10^{-5}

*conjugation of pGE7 into RmP4258 was completed at a different date than the other conjugation experiments.

3.2.4 Conjugation of Second pGE plasmid into Minimal *S. meliloti* Cell

Derivatives of the pGE plasmids contain the origin of replication (*oriV*) for either pSymA (*repA2B2C2*) or pSymB (*repA1B1C1*). These two *oriVs* allow replication of the two megaplasmids in wild type *S. meliloti* and we wished to investigate how the two pGE plasmids would behave when each was present in a single *S. meliloti* strain. To conjugate a second pGE plasmid into an *S. meliloti* strain already harbouring a pGE plasmid and determine the conjugation frequency (see

Table 7), triparental matings were performed by mixing donor cultures of *E. coli* DH5α containing the pGE plasmids with the *S. meliloti* recipient RmP4258 harbouring either pGE3 or pGE4. Mating mixtures were spotted on LBmc FeCo and selection for transconjugants that harbour two pGE plasmids was performed with Sm (200µg/mL) and two additional antibiotics to select for each of the pGE plasmids present. All strains were then streak purified on LBmc FeCo with two appropriate antibiotics.

Table 7: Conjugation frequency of a second pGE plasmid into RmP4258 harbouring pGE3 or pGE4.

Donor	Recipient	Conjugation frequency
DH5α pGE5	RmP4258 pGE3 (RmP4280)	3.05×10^{-2}
DH5α pGE5	RmP4258 pGE4 (RmP4281)	2.98×10^{-4}
DH5α pGE6	RmP4258 pGE4 (RmP4281)	4.86×10^{-3}
DH5α pGE7	RmP4258 pGE3 (RmP4280)	2.77×10^{-2}

Resulting strains carried two pGE plasmids each containing a different *oriV.* The plasmids pGE3 and pGE4 contain the *repA2B2C2 oriV*, and the plasmids pGE5, pGE6, and pGE7 contain the *repA1B1C1 oriV*.

3.2.5 Plasmid Stability Assays

A desirable trait when using plasmids for genetic engineering is the ability of the plasmid to be stably maintained within the host organism to allow for long-term maintenance of recombinant DNA. This is dependent on several factors, including: growth rate of the cell, plasmid copy number, control of replication, and media in which the host cell is cultured (Imanaka & Aiba, 1981). The *S. meliloti* pSymA and pSymB *repABC* operons are part of the

RepABC family that facilitates replication and partitioning of plasmids. RepC acts as initiator of replication, whereas RepA and RepB act to partition the megaplasmids to daughter cells during cell division (Frage, et al., 2016). Both megaplasmids are maintained within the cell at a low copy number (Finan, et al., 2001).

To characterize the stability and maintenance of the pGE plasmids in RmP4258, two stability assay experiments were conducted. The first experiment examined the stability of the pGE3 plasmid in the strain RmP4280, a derivative of RmP4258 containing pGE3. The second experiment examined four separate strains where two pGE plasmids were present. These strains were RmP4310 (pGE5 and pGE3), RmP4311 (pGE5 and pGE4), RmP4312 (pGE6 and pGE4), and RmP4313 (pGE7 and pGE3), all of which were derivatives of RmP4258.

For each of these experiments, single colonies of strains harbouring the pGE plasmid(s) were inoculated into LBmc FeCo with appropriate antibiotics and the next day, the culture was diluted to 10^{-6} and 100µL was plated on LBmc FeCo non-selective media. After incubation at 30°C for 3 days, 100 randomly chosen colonies were then patched onto LBmc FeCo media with appropriate antibiotics. Colonies unable to grow were deemed to have lost the pGE plasmid(s). Following the initial growth in antibiotics, cultures were then grown exclusively in media without antibiotics. This was done to determine the stability of the pGE plasmid(s) when grown without selective pressure. Each day, cultures were subcultured 1 in 1000 times in LBmc FeCo liquid media (without antibiotics) and grown overnight for 24 hours during which time approximately 10 generations occurred. The number of generations was calculated by measuring the OD_{600} prior to plating cells on non-selective media and following subculturing, and then determining the number of doublings that occurred from the initial OD_{600} value to the final OD_{600} value (see

Figure 8). Prior to each subculture, the cultures were diluted to 10^{-6} and 100μL was plated on LBmc FeCo solid media at let grow at 30°C for 3 days. Following growth, 100 randomly chosen colonies were patched onto solid media with appropriate antibiotics as before to determine the number of colonies unable to grow. These experiments were conducted in triplicate over a 5 day period.

The number of colonies unable to grow on selective solid media was recorded and plotted against the total number of generations cultured (see Figures 8 and 9). Following completion of the second experiment, patched colonies unable to grow on two selective antibiotics were then re-patched onto two separate LBmc FeCo plates, each containing a single antibiotic for each of the pGE plasmids tested. The number of colonies re-patched on the plates with a single antibiotic that were unable to grow was recorded to determine which plasmid was lost during the initial stability assay (see Figure 10).

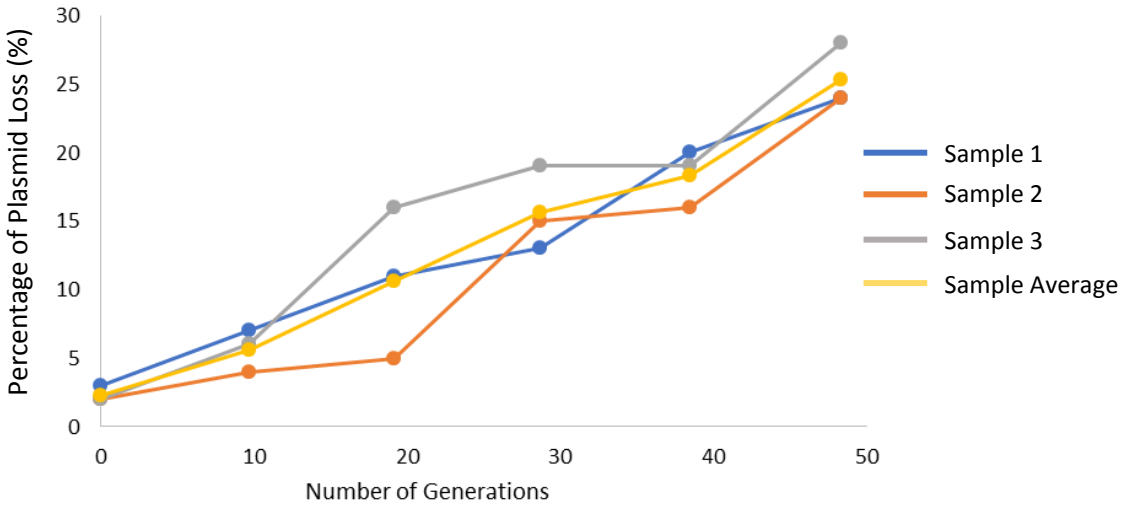


Figure 8: Percentage of cells (colonies) that lost plasmid pGE3 as a function of cell generations. Strain RmP4258 pGE3 (RmP4280), was grown in LBmc FeCo media and number of cells that lost the pGE3 plasmid (Tc^r) was determined by screening colonies on LBmc FeCo media containing Tc (5µg/mL). The experiment was performed in triplicate. Blue, orange and grey lines represent results from each individual sample, and yellow line represents sample averages. Number of generations was calculated using the following equation: $2^n = \frac{Final\ OD_{600}\ measured}{Initial\ OD_{600}\ measured}$, where *n* = number of generations that occurred during each subculturing data point

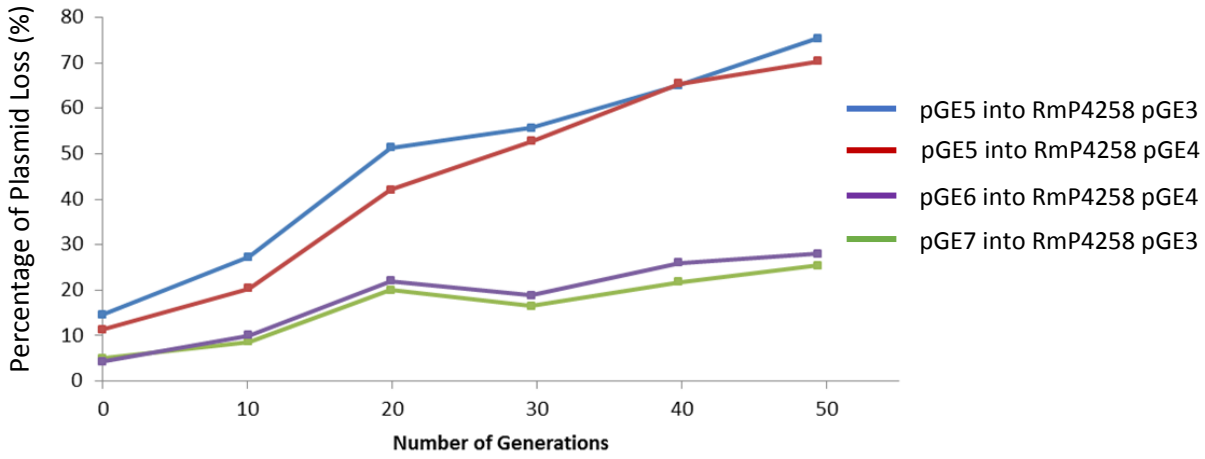


Figure 9: Percentage of patched colonies, RmP4258 with two pGE plasmids, unable to grow on LBmc FeCo with appropriate antibiotics for both pGE plasmids relative to total number of generations of cells growing in non-selective media LBmc FeCo performed. Each line represents the average of triplicate samples. Plasmid stability of four separate strains, each harbouring two pGE plasmids was tested. Colonies unable to grow on selective media were deemed to have lost one or both pGE plasmids. Number of generations was calculated in the same manner as in Figure 8. For colony count data, see Table 11 in Appendix.

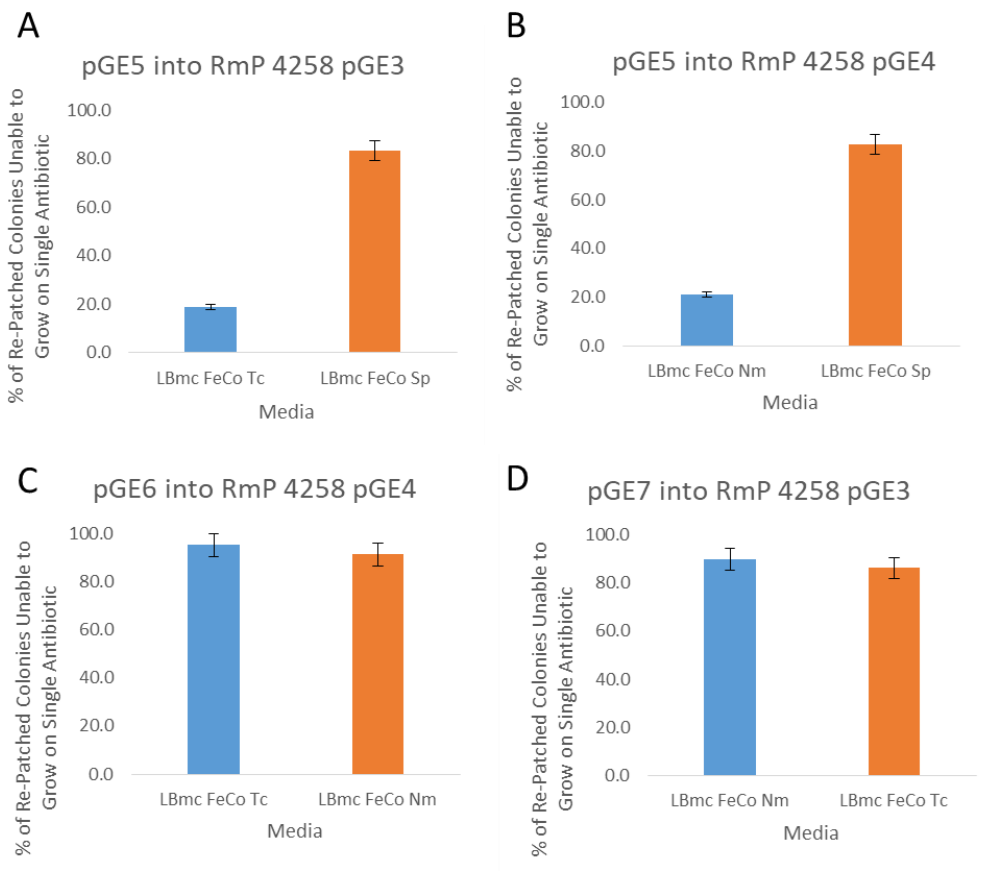


Figure 10: Plasmid stability in cells carrying two plasmids. Colonies from experiment presented in Figure 9, were screen for the presence of the second plasmid by screening for growth on LBmc FeCo media containing the appropriate antibiotic. Patched colonies from Figure 9 unable to grow on media with two antibiotics were then re-patched onto two separate plates each with a single antibiotic to determine which pGE plasmid was lost during initial stability assay. Re-patched colonies unable to grow on media with single antibiotic were considered to have lost corresponding plasmid. Data represents average of triplicate samples. Error bars represent standard deviation. For additional data see Figure 23 and Table 12 in Appendix.

The stability assay data from the second stability experiment suggest that strains harbouring the pGE5 plasmid, RmP4310, RmP4311, lose the plasmid with a greater frequency than the other plasmids tested, where >70% of patched colonies lost pGE5 following ~50 generations of culturing. Furthermore, pGE3, pGE4, pGE6 and pGE7 appear to be moderately stable, where ~25% of patched colonies were unable to grow on selective media following 50 generations of culturing.

3.2.6 Eckhardt Gel Electrophoresis to Observe Possible Recombination of pGE Plasmids

One approach to visualize plasmid DNA is to use Eckhardt gels (Eckhardt, 1978). The Eckhardt gel procedure employs in-well lysis using sodium dodecyl sulfate (SDS) to gently lyse cells, followed by immediate agarose gel electrophoresis. Most chromosomal DNA remains in the well as it is too large to be run through the pores of the agarose gel and only plasmid DNA is visualized (Eckhardt, 1978).

Due to large regions of homology in the backbone vectors of the pGE plasmids, it is possible that double homologous recombination can occur between two pGE plasmids present in the same host strain. High frequency of recombination between two pGE plasmids present in the same cell would be detrimental to the ability to independently harbour and manipulate two large fragments of high G+C content DNA for downstream applications. To determine if homologous recombination occurred in strains carrying two pGE plasmids, Eckhardt gels were conducted using a modified protocol to visualize the relative size of plasmid DNA present in the *S. meliloti* strains listed in Table 7 (Hynes, Simon, & Puhler, 1985). Samples from each of these strains will be compared to a strain harbouring a single pGE plasmid to determine if there is a difference in size. A single band, which runs at a higher size, would suggest the pGE plasmids had

recombined together. A 52 kb plasmid (pTA-MOB) and a ~50 kb plasmid (pTH566) were run as markers to compare relative band sizes. Here, colonies are grown overnight at 30°C in LBmc FeCo with appropriate antibiotics and subcultured the next day to an OD_{600} of 0.2 and then grow to an OD_{600} of 0.45. Cultures were placed on ice and 0.5mL of 0.3% ice-cold, sterile sarkosyl solution was then added to 0.2mL of each culture. Cultures were then centrifuged and the cell pellet resuspended in 20µL of lysis solution (100µg/mL lysozyme, 10µg/mL RNaseA, 10% sucrose in 0.5 times TBE) and immediately added to the wells of a 0.8% agarose gel containing 1% SDS in 0.5 times TBE. Samples were electrophoresed at 10V for 15 minutes and then run at 100V for 2 hours to allow for separation of the plasmid DNA from the cell debris (see Figure 11).

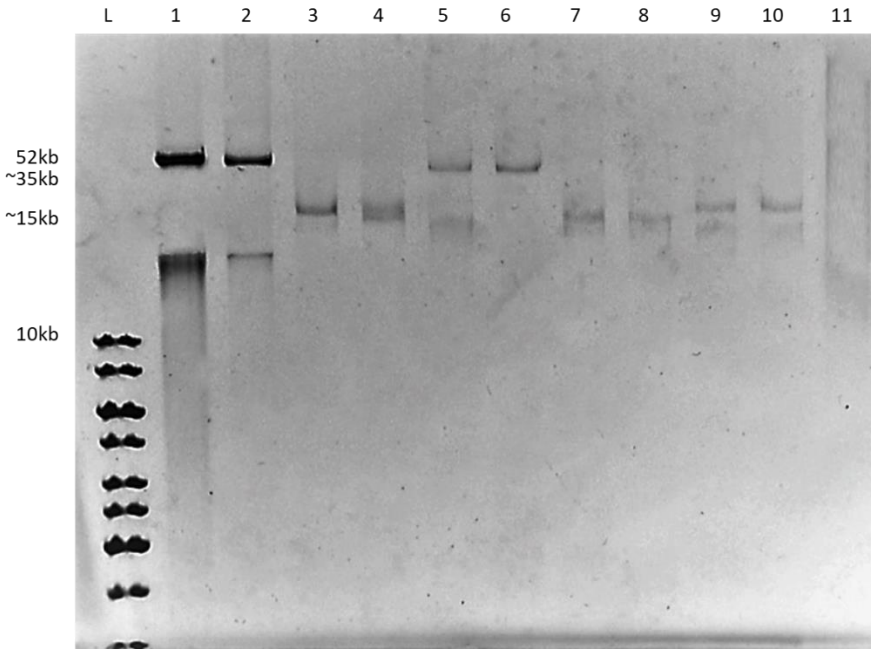


Figure 11: Eckhardt gel electrophoresis on strains harbouring two pGE plasmids. Lane 1 and 2 contain large plasmid markers (pTA-MOB, 52.7 kb and pTH566, ~50 kb) to be used to compare the relative sizes of the tested bands. Lanes 3-6 contain the samples harbouring two pGE plasmids. Lane 3 contains pGE5 in RmP4258 pGE3 (RmP4310). Lane 4 contains pGE5 in RmP4258

pGE4 (RmP4311). Lane 5 contains pGE6 in RmP4258 pGE4 (RmP4312). Lane 6 contains pGE7 in RmP4258 pGE3 (RmP4313). Lanes 7-10 contain the samples harbouring a single pGE plasmid to be used as a point of comparison relative to the samples harbouring two pGE plasmids. Lane 7 contains RmP4258 pGE3 (RmP4280). Lane 8 contains RmP4258 pGE4 (RmP4281). Lane 9 contains RmP4258 pGE5 (RmP4282). Lane 10 contains RmP4258 pGE6 (RmP4283). Lane 11 contains RmP4258 without any pGE plasmid present. L denotes a 1 kb ladder (GeneRuler, ThermoScientific).

The data from the Eckhardt gel suggest that the samples ran in Lanes 3 and 4 contained two separate plasmids that did not recombine, whereas the samples ran in Lanes 5 and 6 contained a single DNA fragment where the two plasmids had recombined. This was determined based on the comparison of the band sizes in the tested samples relative to the samples where only a single pGE plasmid was present.

3.3 Analysis of Site-Specific Recombination Sites in Useful Background Strains

To investigate the use of site-specific recombination sites as a method for the capture of large genomic fragments, a ΔpSymA strain containing an *FRT* site integrated at the *hypRE* gene locus was examined. Previously, two large fragments of high G+C content DNA, 70 kb and 275 kb in size (Gm^r Nm^r), derived from the pSymA megaplasmid of *S. meliloti* were captured via Flp-catalyzed recombination at the *hypRE* locus on pSymB in a strain lacking pSymA, RmP4254 (Situ and Finan, 2017 unpublished) (Sather, Bindra and Finan, 2017, unpublished). *HypRE* encodes for a hydroxyproline epimerase, an enzyme required to catabolize *trans*-4-hydroxy-L-proline and is located on the pSymB replicon. Mutants lacking the *hypRE* gene are unable to grow on M9 minimal media lacking hydroxyproline and thus, ability to grow on hydroxyproline can be used

for selection (White, et al., 2012). RmP4254 was made by transducing a *hypRE*::*FRT*-Nm^r-*FRT* cassette from the strain RmP2507 into RmP4246, followed by excision of the Nm cassette by Flp-catalyzed recombination of flanking *FRT* sites (Situ and Finan, 2017, unpublished). RmP2507 was made by replacing the *hypRE* gene via double homologous recombination with a *FRT*-Nm^r-*FRT* cassette (White, et al., 2012).

3.3.1 Transduction of the wild type *hypRE* locus into Strain with Captured Region from pSymA

In previous work, a 275 kb region derived from pSymA that also carried gentamycin (Gm) and neomycin (Nm) resistance genes was thought to be captured at the *hypRE* locus in RmP4254. This was completed by first excising the region of interest from RmP4249 via Flp-catalyzed recombination of flanking *FRT* sites followed by simultaneous conjugation into recipient strain RmP4254 to rescue the excised region. This resulted in the isolation of strain RmP4306 (Situ and Finan, 2017, unpublished; diCenzo & Finan, 2018). Strain RmP4306 lacks pSymA and it was assumed that the excised 275 kb region comprising the pSymA-A117, A118, A301, and A121 regions had been rescued via Flp-catalyzed recombination into the *hypRE*:: *FRT* site on pSymB (Situ and Finan, 2017, unpublished). If the latter were the case, then recombining the wild type *hypRE* locus into strain RmP4306 should result in the loss of the 275 kb A117-A118-A301-A121 region.

Here, this was performed by transduction of the wild type *hypRE* region back into RmP4306 (see Table 8). Transductants were recovered at a frequency of 2×10^{-8} on M9 minimal media with 10mM hydroxyproline (M9 Hyp). Strain RmP4306 recipient alone was shown to be unable to grow on this media. To verify that the large genomic fragment had indeed integrated

at the *FRT* site, colonies were screened for resistance to gentamycin (Gm) (60µg/mL) and Nm (200µg/mL). The data show that 100% (50/50 colonies) of the transductants were unable to grow on M9 Hyp with Gm and Nm, demonstrating that the large genomic region was replaced with the wild type *hypRE* regions.

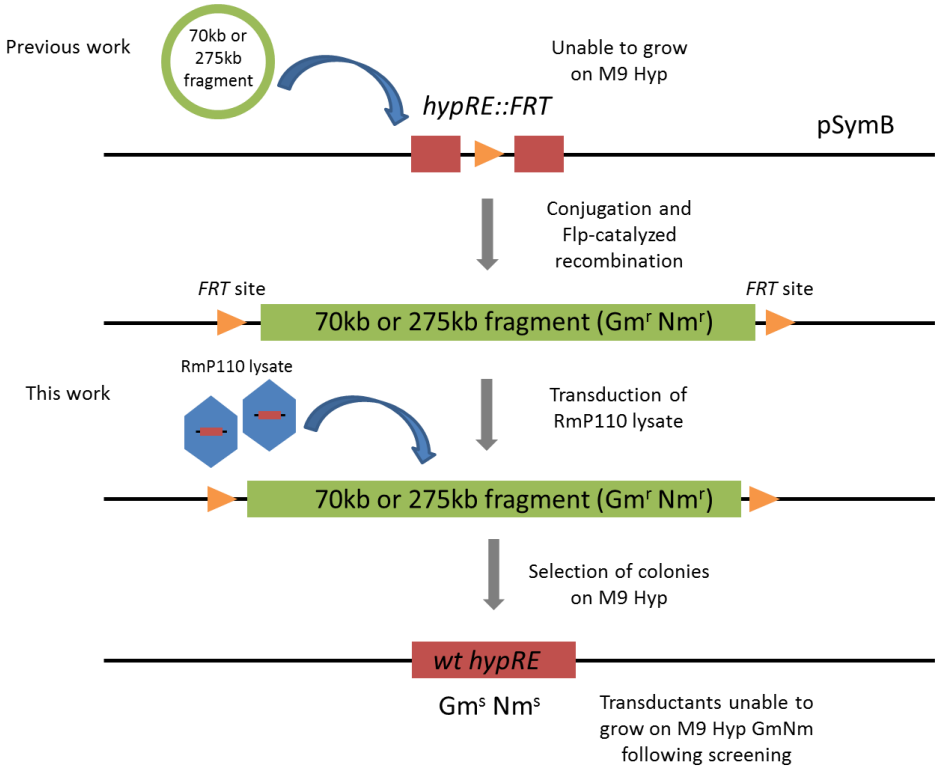


Figure 12: Diagram depicting the transduction of the wild type *hypRE* region back into strains where the 70 kb or 275 kb regions were thought to be captured at the *hypRE::FRT* locus. Colonies selected for on M9 Hyp media were then screened on M9 Hyp Gm (60µg/mL) Nm (200µg/mL) to confirm that the large region was incorporated at the *hypRE::FRT* locus.

Table 8: Validation of large fragment integration at the *FRT* site in a Δ*hypRE* strain.

Experiment	Transduction Frequency	% Patches Unable to Grow on M9 Hyp Gm Nm
RmP4306 with 275 kb plasmid	2.18×10^{-8}	100% (50/50 colonies)
RmP4305 with 70 kb plasmid	5.21×10^{-8}	100% (50/50 colonies)

3.3.2 Transduction of Wild Type *hypRE* locus into Strain with Captured pA132 Region

In an additional previous experiment, the 70 kb fragment comprising the A132 region of pSymA was rescued as a replicating plasmid in *E. coli* (Sather, Bindra and Finan, 2017, unpublished). This fragment was also thought to be captured at the *hypRE*::*FRT* locus in RmP4254 via Flp-catalyzed recombination as described above; resulting in RmP4305 (see Section 3.3.1). To validate the capture of the A132 70 kb fragment at the *FRT* site, a transduction was performed in the same manner as before to transduce the wild type *hypRE* region into the recipient RmP4305. However, it was found that the initial attempts to transduce the wild type *hypRE* gene failed as no transductants were observed. As the pA132 region contains a toxin-antitoxin system, this would be deleted upon transduction of the wild type *hypRE* region into the RmP4205 strain and cause cell death (Milunovic, et al., 2014). Accordingly, in an effort to recover transductants, a spectinomycin resistant (Sp^r) plasmid pTH2623 was conjugated into RmP4305 at a frequency of 4.1×10^{-3}, using methods described previously. This plasmid is a pTH1931 derivative that has had the *sma2253/sma2255*genes cloned into the PacI site, and will express these genes *in trans* (Milunovic, et al., 2014). RmP4305 strains harbouring pTH2623 were then used as a recipient to transduce the wild type *hypRE* region at a frequency of 5.2×10^{-8}. Successful transductants were then recovered on M9 minimal media with 10mM hydroxyproline and then patched on M9 Hyp Gm Nm (see Table 8). The data show that 100% of the patched

colonies were unable to grow on M9 Hyp with Gm and Nm, which demonstrates that the pA132 region was initially captured at the *FRT* site in RmP4305.

3.4 Capture of Large Fragments with high G+C Content DNA in pGE Plasmids

To enable cloning of large fragments in pGE vectors using site-specific recombination-based methods, a derivative of the pGE4 plasmid was made where a "landing pad" was cloned into the pGE4 *PacI* site. This landing pad contains site-specific recombination sites (*FRT, lox71, attB*) to enable capture of incoming DNA fragments containing the corresponding site-specific recombination site (see Figure 4B). The plasmid was designated as pGE4-LP.

3.4.1 Cloning Strategy for the Capture of Large High G+C Content Plasmids

A proof-of-concept experiment was devised to capture 3 large plasmids of high G+C content in pGE4-LP by site-specific recombination (see Figure 14). With this experiment, we wished to demonstrate the ability to conduct *in vivo* cloning of large fragments of DNA into the pGE plasmids. Three large donors plasmids (fragments) were chosen as they represent concepts relevant to synthetic biology, including: 1) a plasmid (pTH3255; ~170 kb) harbouring the putative minimal symbiotic nitrogen fixation regions from pSymA of *S. meliloti*, 2) a plasmid (pTH3247; ~205 kb) harbouring the putative minimal symbiotic nitrogen fixation regions from pSymB of *S. meliloti*, and 3) a plasmid (pA132; ~70 kb) harbouring a putative region involved in bacteriophage resistance in *S. meliloti* (see Figure 13). Plasmids pTH3255 and pA132 were captured in the background strains RmP4258 and RmP4124, whereas pTH3247 was captured in the background strain RmP4314 (see Section 3.4.3).

It should be noted that there exists unwanted regions of homology on pGE4-LP aside from the site-specific recombination sites. These regions include the *nptII* gene that confers neomycin/kanamycin resistance which is also present on all three large fragments to be captured as well as on pGE4-LP. Another region of unwanted homology is the *oriT* sequence present on the large fragments which is also present on pGE4-LP. For the *nptII* gene, there exists 795 bp of homology and or the *oriT* sequence, 110 bp of homology exist. Due to time restrictions of this project, the pGE4-LP plasmid was used in this proof-of-concept experiment. The proper plasmid to be used for this experiment would be a pGE3 plasmid with the landing pad construct cloned into the *PacI* site. This plasmid would not contain the *nptII* gene but rather the *tetR* and *tetA* genes that confer tetracycline resistance. The pGE3 plasmid does contain the *oriT* region; however, it is a small region that would cause lower levels of homologous recombination relative to the homologous regions in the *nptII* plasmid present on both the large fragments and pGE4-LP. As such, an additional screening using colony PCR will be completed to determine if the large fragments successfully recombined at the site-specific recombination site.

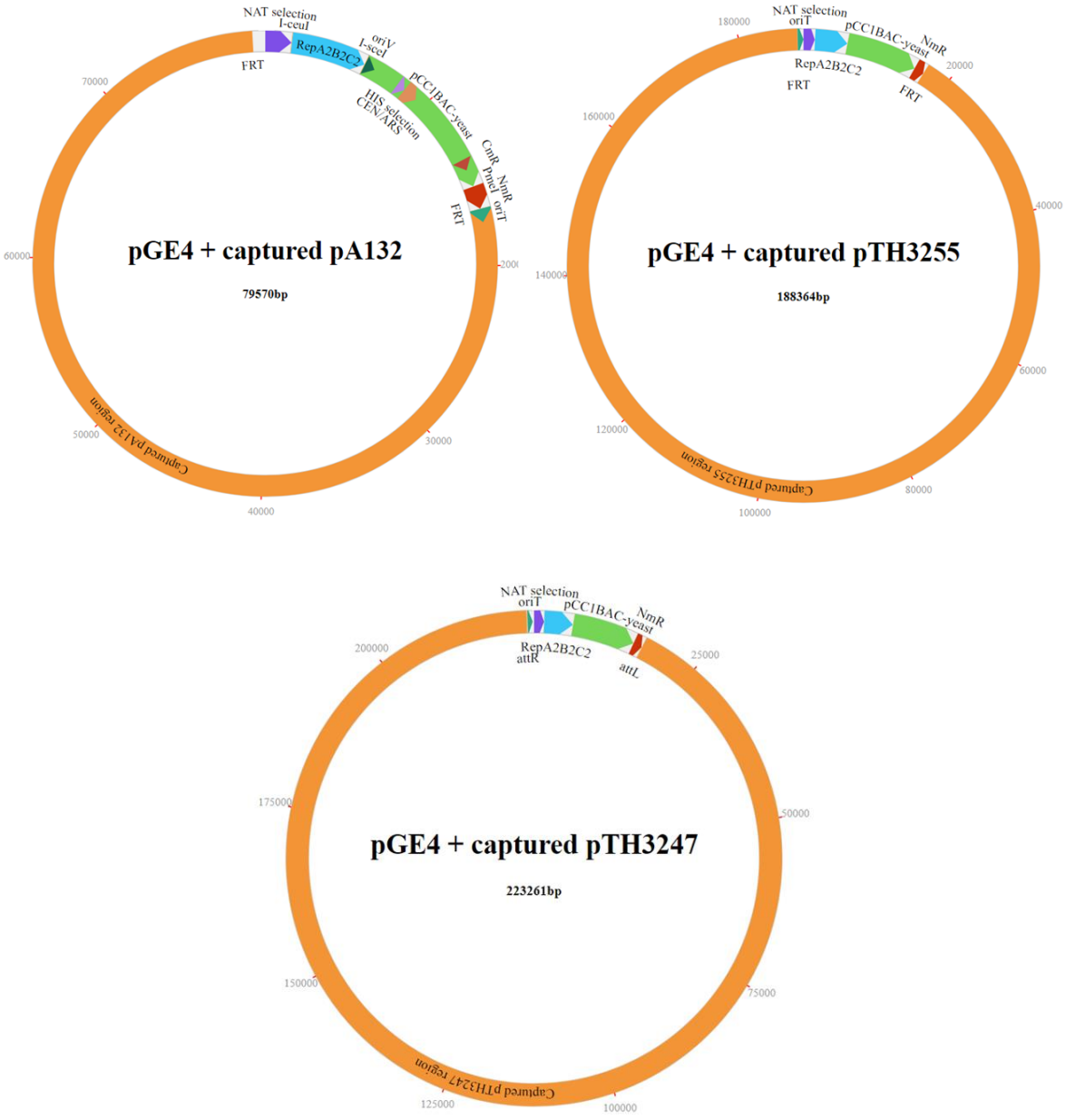


Figure 13: Diagram of three large fragments of high G+C content DNA derived from pSymA and pSymB to be captured in pGE4-LP. (A) Shows pA132 which is a putative bacteriophage resistance region derived from pSymA and has a G+C content of 58.6%. (B) Shows pTH3255 which is the putative minimal symbiotic N_2-fixation region from pSymA and has a G+C content of 59.3%. (C) Shows pTH3247 which is the putative minimal symbiotic N_2-fixation region from pSymB and has

a G+C content of 63.1%. These plasmids were chosen due to their high G+C content DNA (>61%) and their size (70 kb-205 kb) and will be used in a proof-of-concept experiment to demonstrate the ability to clone large DNA fragments into pGE-LP.

3.4.2 Construction of pGE Plasmids Harbouring Landing Pad

The pGE4-LP plasmid was constructed using TAR cloning in yeast and was prepared entirely by P. Janakirama and B. Karas of Designer Microbes Inc. Two primer sets were used to amplify two separate fragments using pGE4 as template. Primers BK416_F and D449R amplified a 7807 bp fragment and primers D450F and BK416_R amplified a 9939 bp fragment. BK416_F and BK416_R are 130 bp oligonucleotides with overhangs each corresponding to approximately one half of the landing pad sequence, with a total overlap region of 60 bp. D450F and D449R are 40 bp oligonucleotides with a total overlap region of 80 bp. Both linear fragments were then co-transformed into competent yeast cells to facilitate homologous recombination during the TAR cloning procedure (see Section 1.1.2) (Designer Microbes Inc., 2018, unpublished).

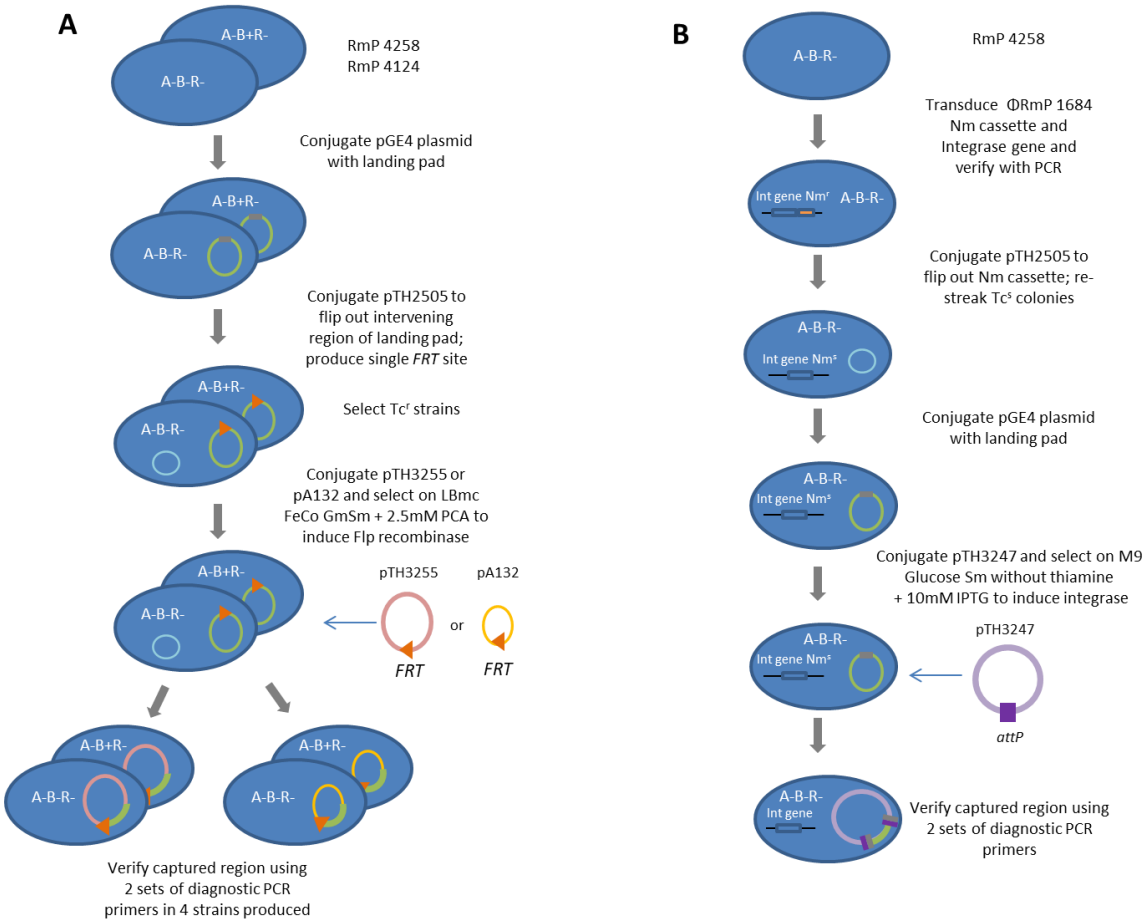

A
A-B+R-
A-B-R-
RmP 4258
RmP 4124
Conjugate pGE4 plasmid with landing pad
Conjugate pTH2505 to flip out intervening region of landing pad; produce single FRT site
Select Tc^r strains
Conjugate pTH3255 or pA132 and select on LBmc FeCo GmSm + 2.5mM PCA to induce Flp recombinase
pTH3255
pA132
or
FRT
FRT
Verify captured region using 2 sets of diagnostic PCR primers in 4 strains produced
B
A-B-R-
RmP 4258
Transduce ΦRmP 1684 Nm cassette and Integrase gene and verify with PCR
Int gene Nm^r
Conjugate pTH2505 to flip out Nm cassette; re-streak Tc^s colonies
Int gene Nm^s
Conjugate pGE4 plasmid with landing pad
Conjugate pTH3247 and select on M9 Glucose Sm without thiamine + 10mM IPTG to induce integrase
pTH3247
attP
Int gene
Verify captured region using 2 sets of diagnostic PCR primers
C
pTH3255 or pA132
Flp-catalyzed recombination between FRT sites to excise intervening region of landing pad
Flp-catalyzed recombination between FRT sites to capture large fragment
pGE4 LP
pGE4 LP
pGE4 LP + Captured pTH3255 or pA132
D
pTH3247
Integrase-mediated recombination between attB and attP to capture large fragment
pGE4 LP
pGE4 LP
pGE4 LP + Captured pTH3247

Figure 14: Schematic diagram of *in vivo* cloning experiment to clone large fragments of high G+C content DNA into the pGE4-LP plasmid. (A) Demonstrates the cloning of pTH3255 or pA132 into RmP4258 and RmP4124 at the *FRT* site and (B) demonstrates the cloning of pTH3247 into RmP4314 at the *attB* site. (C) Demonstrates the intended Flp-catalyzed recombination event to capture large fragments between *FRT* sites on pGE4-LP and pTH3255 or pA132, which were captured using the same method. (D) Demonstrates the intended Integrase-mediated recombination event to capture large fragments between *attB* on pGE4-LP and *attP* on pTH3247. Orange triangles denote *FRT* sites. Violet rectangles denote *attB* sites. Dark purple rectangles denote *attP* sites. Blue triangles denote *lox71* sites. Brown rectangles denote *EcoRI* sites. Adjacent violet and purple squares denote *attL* and *attR* sites.

3.4.3 Transfer of the ΦC31 Integrase into the *S. meliloti* Minimal Cell

For the experiment illustrated in Figure 14B, the first step towards integrating pTH3247 at the *attB* site of pGE4-LP was to transduce the ΦC31 integrase gene into the minimal RmP4258 background strain, to generate strain RmP4314. In previous work, the ΦC31 integrase gene was cloned into the chromosome of RmP110 to generate the strain RmP1684 (Fink and Finan, 2010, unpublished). A phage ΦM12 lysate of RmP1684 was used to transduce the Nm resistance cassette into RmP4258 (transduction frequency was 2.5×10^{-7}). To remove the Nm resistance gene, plasmid pTH2505 carrying the Flp recombinase gene was conjugally transferred to the Nm^r RmP4258 transductant and Nm^s transconjugants were identified upon screening on LBmc FeCo Nm. Colonies were then screened on LBmc FeCo Tc (5µg/mL) to identify cells that lost the pTH2505 plasmid, and the strain was designated RmP4314. Colony PCR was then conducted using diagnostic primers to confirm the presence of the ΦC31 integrase gene in 5 independent

colonies of RmP4314 (see Figure 15). The results showed the presence of the ΦC31 integrase gene in the RmP4314 strain.

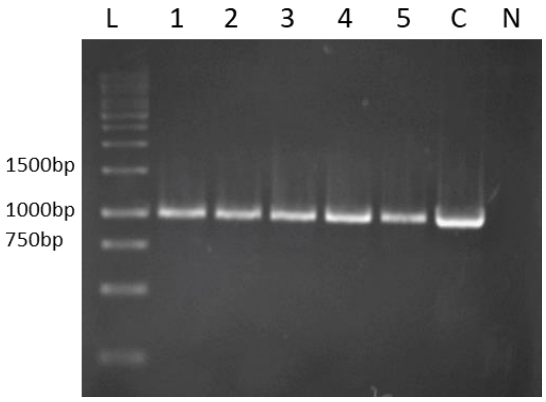


Figure 15: Diagnostic colony PCR to confirm presence of ΦC31 integrase gene in RmP4258. L denotes 1 kb ladder (GeneRuler 1 kb GTU, ThermoScientific), samples 1-5 denote 5 independent transductant colonies of RmP4314, C denotes the control RmP1684 colony, and N denotes the negative control RmP4258 parental strain. A ~1 kb band was observed in each sample except the RmP4258 parental strain, as expected for each other strain should harbour the ΦC31 integrase gene.

3.4.4 Conjugation of pGE4-LP Plasmid into Background Strains

For the three experiments illustrated in Figure 14, the next step was to conjugate the pGE4-LP plasmid into the intended background strains. Three triparental matings were conducted by mixing the donor strain DH5α pGE4-LP with 3 separate recipient strains, RmP4258, RmP4124, and RmP4314, using MT616 harbouring pRK600 as a helper strain. All three matings were spotted onto LBmc FeCo, incubated overnight, and plated on LBmc FeCo Sm (200μg/mL) Nm (200μg/mL) and successful transconjugants were recovered at a frequency of 1.8×10^{-4}, 7.6 x

10^{-4}, and 9.8 x 10^{-3}, respectively. Strains were then streak purified three times on the same media.

3.4.5 Using Flp-Catalyzed Recombination to Produce Single *FRT* site in Landing Pad

For the experiments involving RmP4258 without ΦC31 integrase and RmP4124, the next step was to excise the region between the two *FRT* sites on the landing pad construct in pGE4-LP to produce a single *FRT* site, designated as pGE4-LPE. This was accomplished upon transfer of pTH2505 containing Flp recombinase.

To confirm the intervening region between the *FRT* sites was excised by Flp recombinase, a diagnostic restriction enzyme digestion was completed. The intervening region contains an *EcoRI* restriction enzyme cut site that, when removed, gives a different banding pattern relative to pGE4-LP plasmids without the region removed (see Figure 16). A single RmP4258 or RmP4124 colony was grown overnight in LBmc FeCo Nm (200µg/mL) and pGE4-LP plasmid was then isolated using the EZ-10 Spin Column Plasmid DNA Miniprep kit (BioBasic Canada). A diagnostic restriction enzyme digestion was then prepared by mixing roughly 1µg of DNA with 2 units of *EcoRI-HF* (NEB Canada) in CutSmart buffer (NEB Canada) and let to incubate at 37°C for 2 hours. Digests were then visualized on a 0.8% agarose gel using gel electrophoresis.

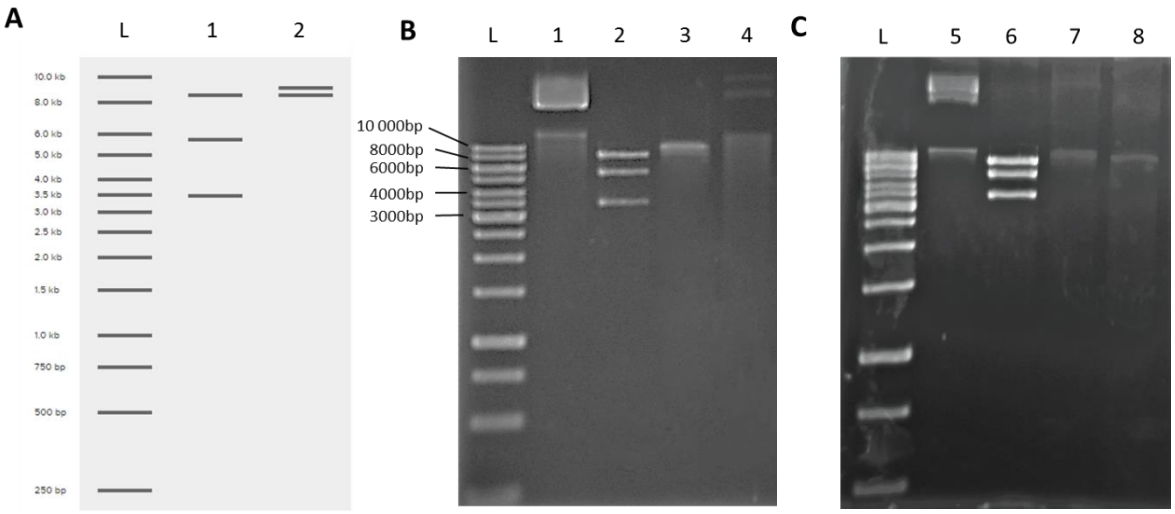


Figure 16: Diagnostic restriction enzyme digest using *EcoRI* to determine if the intervening region between the *FRT* sites in pGE4-LPE was excised via Flp-catalyzed recombination. (A) Shows the expected banding pattern of 1) pGE4-LP without landing pad excised with band sizes of 3472 bp, 5730 bp, and 8474 bp and 2) pGE4-LPE with the landing pad excised with band sizes of 8474 bp and 9072 bp. (B) Shows the results of the restriction enzyme digestion for pGE4-LP/LPE isolated from RmP4124. Lane 1 contains undigested pGE4-LP with the full landing pad region, Lane 2 contains *EcoRI* digested pGE4-LP with full landing pad, Lane 3 contains digested pGE4-LPE with excised region between *FRT* sites, Lane 4 contains undigested pGE4-LPE with excised region between *FRT* sites. (C) Shows the results of the restriction enzyme digest for pGE4-LP/LPE isolated from RmP4258. Lane 5 contains undigested pGE4-LP with the full landing pad region, Lane 6 contains *EcoRI* digested pGE4-LP with full landing pad, Lane 7 contains undigested pGE4-LPE, and Lane 8 contains digested pGE4-LPE with excised region between *FRT* sites. L denotes a 1 kb ladder (GeneRuler 1 kb GTU, ThermoScientific),

The restriction digest data show that the region between the FRT sites was excised in both RmP4258 and RmP4124, based on the expected banding pattern. Additional bands seen in lanes 1, 4 and 7 from panels B and C represent alternative conformations of the undigested pGE4-LP/LPE plasmid and are not seen in the digested samples.

3.4.6 Conjugation of Large Plasmids into Strains Harbouring pGE4-LP

Following completion of the above steps, the three large plasmids were then conjugated into the background strains. Triparental matings were conducted between the donor DH5α Rif^r harbouring each of the large fragment plasmids and the background strains for a total of 5 matings (see Table 9), using MT616 as a helper strain. Mating mixtures for experiments 1-4 were spotted onto LBmc FeCo, incubated overnight, and successful transconjugants were recovered on LBmc FeCo Sm (200µg/mL) Gm (60µg/mL). Colonies were then streak purified 3 times on the same media prior to downstream work.

Table 9: Conjugation experiments performed to transfer 3 large plasmids into 3 different background strains**.**

Donor Strain	Recipient Strain	Conjugation Frequency
1) DH5α Rif^r pA132	RmP4258 pGE4-LP (single *FRT* site) pTH2505	9.94 x 10^{-5}
2) DH5α Rif^r pA132	RmP4124 pGE4-LP (single *FRT* site) pTH2505	5.32 x 10^{-4}
3) DH5α Rif^r pTH3255	RmP4258 pGE4-LP (single *FRT* site) pTH2505	2.39 x 10^{-4}
4) DH5α Rif^r pTH3255	RmP4124 pGE4-LP (single *FRT* site) pTH2505	6.78 x 10^{-4}
5) DH5α Rif^r pTH3247	RmP4258 pGE4-LP (full landing pad)	1.85 x 10^{-3}

When conducting the conjugation of pTH3247 into RmP4314, selection using antibiotics could not be performed, since the pTH3247 205 kb plasmid derived from the putative minimal

symbiotic N_2-fixation region for pSymB carries Nm^r. As such, an alternative selection method was used to select for the uptake of pTH3247 in RmP4315 (Kearsley and Finan, 2018, unpublished).

The pSymB region comprising pTH3247 (B108, B109, B123) includes one of the loci encoding thiamine biosynthesis genes (nt 1633944-1638330 on pSymB). It was observed previously that the thiamine biosynthesis genes could be used to complement *E. coli thi-1* auxotrophs to allow for growth on M9 minimal media without the addition of thiamine (Finan, et al., 1986). Furthermore, *S. meliloti* strains lacking pSymA and pSymB are unable to grow on M9 minimal media lacking thiamine. RmP4314, used as recipient for this experiment, is a ΔpSymA ΔpSymB strain and as such requires added thiamine in media to allow for growth. A triparental mating between the donor DH5α Rif^r pTH3247 and recipient RmP4314 was conducted using MT616 as a helper strain. The mating was spotted into LBmc FeCo and let grow overnight at 30°C. The next day, mating mixtures were plated on M9 minimal media with 10mM glucose (diCenzo, et al., 2016), 38μM $FeCl_3$, Sm (200μg/mL), and 10mM IPTG to induce expression of ΦC31 integrase and let grow at 30°C for 5 days. Initially, high amounts of small background colonies were observed after 5 days. Following an additional 1 day of incubation, two different phenotypes became apparent, larger colonies and much small non-growing colonies. Thiamine is a vitamin and used as a co-factor for carbohydrate and branched-chain amino acid metabolic enzymes and is required in low concentrations to allow for growth of cells (Du, Wang, & Xie, 2011). The small colonies observed here are thought to arise from residual thiamine present from the LBmc FeCo plates where the mating mixtures were spotted. Inefficient washing of cells prior to plating on selective media would cause some amount of thiamine to be present and allow for initial growth of colonies that did not take up the pTH3247 plasmid. As such, a single large colony was streak purified 3 times on the same media for use in downstream experiments.

3.4.7 Colony PCR to Verify Presence of Large Plasmids in *S. meliloti*

To confirm the presence of the large genomic fragments in each of the 5 recipient strains, diagnostic colony PCR was completed. For each PCR, two sets of diagnostic primers specific to the large region were used. Additionally, one set of primers located outside of the regions of interest were used to ensure only the desired region was transferred.

Colony PCR was conducted on Gm^r Nm^r RmP4258 and Gm^r Nm^r RmP4124 colonies with the putatively captured pTH3255 region using the primers A118_check-II-F and A121_check-II-R (referred to as primer set I) as well as the primers 1142-IF and 1142-IR (referred to as primer set II) (see Figure 17A). Primer set I amplifies a 1778 bp region of the at the border between A118 and A121 regions and primer set II amplifies a 1125 bp region of the *sma0748* and *sma0750* genes. Presence of these bands confirms that the pTH3255 region is present in both the RmP4258 and RmP4124 backgrounds. The primers Sma2245_F and Sma2245_R were also used to ensure only the pTH3255 region was present, and should not amplify a band.

Colony PCR was conducted on Gm^r Nm^r RmP4258 and Gm^r Nm^r RmP4124 colonies with the putatively captured pA132 region using the primers Sma2245_F and Sma2245_R (referred to as primer set III) as well as the primers 1860-OF and 1860-OR (referred to as primer set IV) (see Figure 17B). Primer set III amplifies a 769 bp region of the *sma2245* gene and primer set IV amplifies a 962 bp region of the *sma2301* gene. Presence of these bands confirms that the pA132 region is present in both the RmP4258 and RmP4124 backgrounds. The primers A118_check-II-F and A121_check-II-R were also used to ensure only the pA132 region was present, and should not amplify a band.

For both colony PCR reactions, as additional set of primers, hsdM U Fwd and hsdM U Rev (referred to as primer set V), was used as a positive colony PCR control and amplifies a 953 bp region of the chromosomal *hsdM* gene.

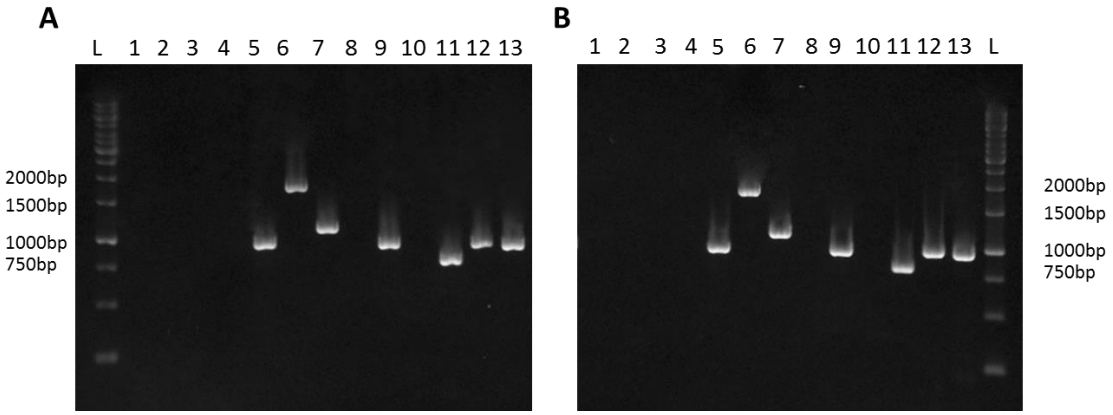


Figure 17: Colony PCR to confirm presence of pA132 and pTH3255 large fragments in RmP4258 and RmP4124. (A) Shows results for PCR reactions performed on RmP4258. (B) Shows results for PCR reactions performed on RmP4124. L denotes 1 kb ladder (GeneRuler 1 kb GTU, ThermoScientific). For both (A) and (B), Lanes 1-5 show the results using RmP4258 pGE4-LPE alone or RmP4124 pGE4-LPE alone as template DNA with primer sets I-V used, respectively. As expected, only Lane 5 contained a band. For both (A) and (B), Lanes 6-9 show the results using RmP4258 pGE4-LPE + pTH3255 or RmP4124 pGE4-LPE + pTH3255 as template DNA with primer sets I, II, III, and V used, respectively. For both (A) and (B), Lanes 10-13 show the results using RmP4258 pGE4-LPE + pA132 or RmP4124 pGE4-LPE + pA132 as template DNA with primer sets I, III, IV, and V used, respectively.

Results from Figure 17 suggest that the pTH3255 and pA132 large fragments are present in the RmP4258 pGE4-LPE and RmP4124 pGE4-LPE strains. The resulting strains in RmP4258 are

designated RmP4316 and RmP4315, respectively. In all lanes where a band is expected, a band matching the expected size is observed. All lanes where no band is expected, no band is seen.

Additionally, colony PCR was conducted on Nm^r RmP4314 colonies with the ability to grow on M9 minimal media without thiamine with the putatively captured pTH3247 plasmid using primer sets 2611_UF and 2611_UR (referred to as primer set VI) and primer set HR_1_Fwd and HR_2_Rev (referred to as primer set VII) (see Figure 18). Primer set VI amplifies an 803 bp region of the *exoB* gene and primer set VII amplifies a 1201 bp region of the border region between B109 and B123 (Kearsley and Finan, 2018, unpublished). Presence of these bands confirms that the pTH3247 region is present in the RmP4314 background. The primers 916_DF and 916_DR were also used to ensure only the pTH3247 region was present, and should not amplify a band (referred to as primer set VIII). As before, the primer set hsdM U Fwd and hsdM U Rev was used as a positive colony PCR control.

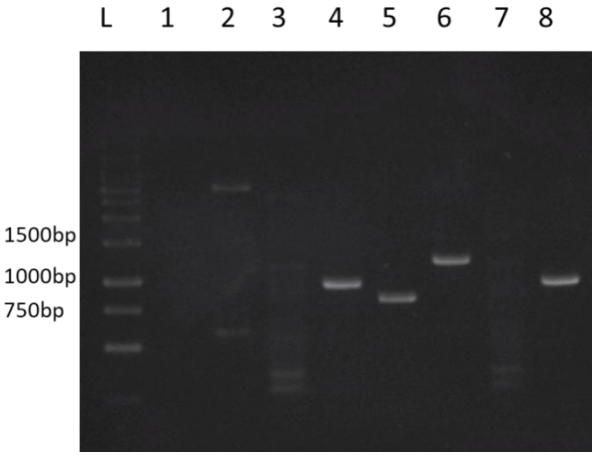


Figure 18: Colony PCR to confirm presence of pTH3247 large fragment in RmP4314. L denotes 1 kb ladder (GeneRuler 1 kb GTU, ThermoScientific). Lanes 1-4 show the results using RmP4314 pGE4-LP alone as template DNA with primer sets VI, VII, VIII, and V used, respectively. Lanes 5-8

show the results using RmP4314 pGE4-LP + pTH3247 as template DNA with primer sets VI, VII, VIII, and V used, respectively.

Results from Figure 18 suggest that the pTH3247 large fragment is present in the RmP4314 strain. This strain was designated RmP4317. Off-target bands are observed in lanes 1-3, but do not match the expected band size if the fragment were present. In lanes 4-6 and 8, a band of the expected size is observed, whereas in Lane 7, no band is observed as expected.

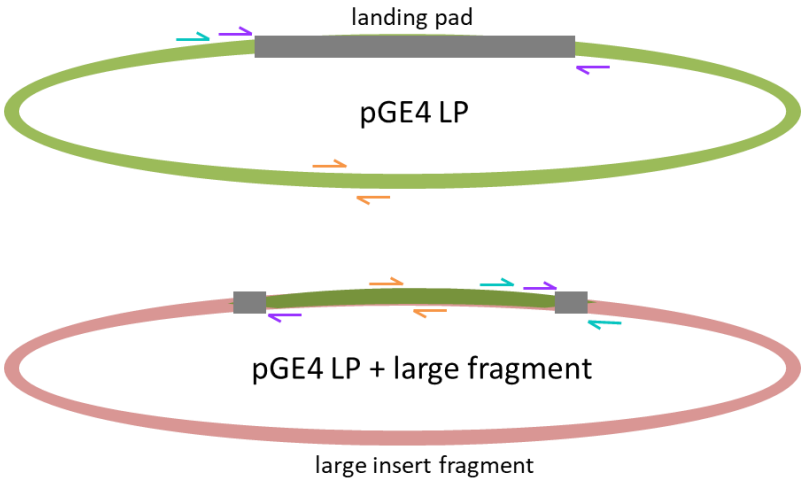
An additional experiment was conducted to further confirm the presence of the pTH3255 plasmid in the RmP4258 and RmP4124 backgrounds. Trigonelline is a secondary metabolite often found in legumes. Catabolism of this molecule is completed by the *trc* genes, which are located on pSymA (Boivin, et al., 1991). *S. meliloti* strains lacking pSymA are thus unable to grow on M9 minimal media with trigonelline as the sole carbon source (Boivin, et al., 1991). Gm^r Nm^r RmP4258 and RmP4124 strains were patched onto M9 minimal media containing 10mM trigonelline (M9 Trig) to determine if the trigonelline catabolism genes, and therefore the pTH3255 plasmid was present. RmP110 wild type was used as a positive control and recipient alone was used as a negative control. Results show 10/10 Gm^r Nm^r patched colonies were able to grow on M9 Trig and the recipient control was unable to grow. This indicates the pTH3255 region was present in the RmP4258 and RmP4124 strains following the conjugation of the large genomic regions.

3.4.8 Colony PCR to Confirm Capture of Large Fragments via Site-Specific Recombination

Following confirmation of the presence of the large fragments, additional colony PCR reactions were conducted to determine if the large fragments were captured via-site specific recombination at the *FRT* sites for pA132 and pTH3255, or at the *attB* site for pTH3247.

Three diagnostic primers sets were used to determine if the large genomic fragment was captured at the landing pad sites. The primers pGE4_check_Fwd and pGE4_check_Rev_new (referred to as primer set IX) were used to confirm the presence of the pGE4-LP/LPE plasmid. Primer set IX amplified a 1371 bp region on the backbone of pGE4-LP/LPE. The primers check_primer_Fwd and pGE1_check_Rev (referred to as primer set X) were used to amplify the region spanning the landing pad region. If the large fragment was incorporated at the landing pad via site-specific recombination, no band would be seen as the intervening region would be too large to PCR amplify. Alternatively, if the large fragment was not incorporated at the landing pad, a 454 bp band would be observed. Lastly, the primers check_primer_Fwd and either check_primer1_Rev for pA132 and pTH3255 or check_primer2_Rev for pTH3247 (referred to as primer set XI and XII, respectively) were used to amplify the region spanning the border where recombination occurred. If the large fragment successfully recombined at the landing pad, a 777 bp, and 833 bp band would be observed, respectively (see Figure 19 and Figure 20).

A



B

Primer Set	Size of PCR product if large fragment captured at landing pad	Size of PCR product if large fragment NOT captured at landing pad
IX	1371 bp	1371 bp
X	no band	454 bp or 580 bp
XI/XII	777 bp or 833 bp	no band

Figure 19: (A) Simplified general diagram depicting the binding location of primer sets IX, X, XI, and XII on pGE4-LP/LPE in cases where large genomic fragments are captured via site-specific recombination. Grey rectangle and squares denotes the landing pad sequence. Coloured arrows denote binding sites of primers used in experiment. Green ellipse denotes pGE4-LP. Red ellipse denotes large genomic fragment. (B) PCR product sizes for different primer sets used under different conditions. Diagram is not to scale.

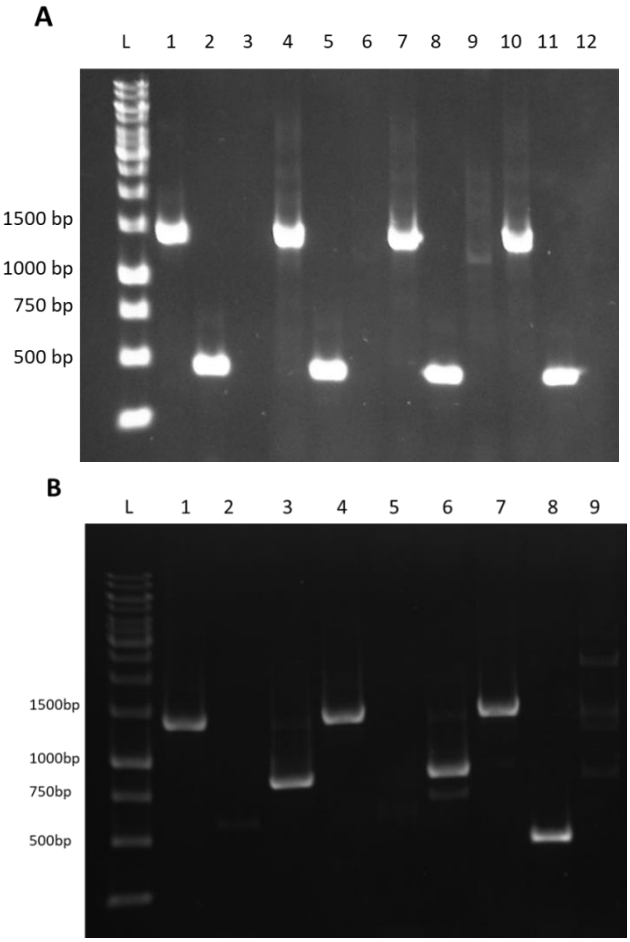


Figure 20: Colony PCR to confirm large fragments were captured at landing pad site. (A) Lanes 1-3 show the results using RmP4258 pGE4-LPE as template with putatively captured pTH3255 with primer sets IX, X, and XI, respectively. Lanes 4-6 show the results using RmP4124 pGE4-LPE as template with putatively captured pTH3255 with primer sets IX, X, and XI, respectively. Lanes 7-9 show the results using RmP4258 pGE4-LPE as template with putatively captured pA132 with primer sets IX, X, and XI, respectively. Lanes 10-12 show the results using RmP4124 pGE4-LPE as template with putatively captured pA132 with primer sets IX, X, and XI, respectively. (B) Lanes 1-3 show the results using RmP4317 with putatively captured pTH3247 in pGE4-LP (pTH3264) colony #1 as template DNA with primer sets IX, X, XII used, respectively. Lanes 4-6 show the

results of the same PCR reaction using the same strain but a second colony using the same primer sets. Lanes 7-9 show the results using RmP4314 alone as template DNA using the same primer sets. L denotes 1 kb ladder (GeneRuler 1 kb GTU, ThermoScientific).

Results for Figure 20A do not show the expected band size of 777 bp when primer set XI is used in lanes 3, 6, 9, or 12. This indicates that the pA132 and pTH3255 plasmids did not recombine at the *FRT* site in the pGE4-LPE plasmid. Furthermore, a band of 454 bp is observed in each of these samples, where no band should have been seen. These plasmids may have recombined at the *nptII* gene in the pGE4-LPE, which shares 795 bp of homology. Alternatively, recombination may have occurred at the *FRT* site at the *hsdR* locus, although no band was observed when a colony PCR to amplify the border region at this locus was performed (see Figure 24 in Appendix).

Results for Figure 20B show a band of approximately 883 bp for lane 3 and lane 6, as well as a lack of band in lane 9. This highly suggests that the large fragment pTH3247 was captured at the *attB* site in pGE4-LP. Furthermore, lane 2 and lane 5 do not show a band spanning the landing pad region, suggesting the pTH3247 region recombined at this site. Lane 8 shows a band as a positive control in the pGE4-LP alone sample.

To demonstrate that the pGE4-LP plasmid harbouring a large fragment of high G+C content DNA could be mobilized, a conjugation experiment was conducted using DH5α Rif^r *E. coli* strain as a recipient and the *S. meliloti* strains harbouring the pGE4-LP plasmid with the large fragment as a donor. MT616 was used as a helper strain. Mating mixtures were spotted onto LBmc FeCo, incubated overnight, and successful transconjugants were recovered on LB Gm (10µg/mL) Rif (50µg/mL) at a frequency of 7.6×10^{-7}, which was noted to be very poor. Gm^r

transconjugants were then streak purified 3 times. Colony PCR was then conducted using primer sets IX, and X to determine if the pGE4-LP plasmid was present in the transconjugants. No bands were observed following this PCR, suggesting the pGE4-LP plasmid was not conjugated but rather that the large fragments had instead recombined out and conjugated into the *E. coli* recipient (data not shown). Due to time constraints, this was not further investigated.

4 General Discussion

This work is centered on the development of a bacterial host system capable of cloning and further allowing for manipulation of large fragments (0.1-1 Mb) of high G+C content DNA. To accomplish this task, a recombinant multi-host shuttle vector (MHS) was created that is able to be conjugated from the host strain, *S. meliloti*, to several recipient strains; including, *E. coli, S. cerevisiae,* and *P. tricornutum*, making it a universal conjugative donor for the MHS vectors developed. These vectors can then be used as genetic tools to transfer and store DNA in direct intra-kingdom or inter-kingdom conjugation experiments.

This work was developed as part of a collaborative project to further develop genetic tools for use in synthetic biology applications. Many organisms relevant to biotechnology and industrial microbiology contain genomes with G+C content >60%. Current “workhorses” for genetic manipulation and cloning include *E. coli* and *S. cerevisiae*, however these strains lack the ability to stably maintain DNA of G+C content >40%, without additional, and cumbersome engineering. *S. meliloti* was chosen due to its high G+C content genome (62.7%) as well as for its ability to harbour large replicons pSymA (1.35 Mb) and pSymB (1.68 Mb). The origin of

replication for these replicons (*repABC)* was used here in the development of the MHS vectors. This portion of the project focused on the characterization of the host strain as well as the development of a proof-of-concept experiment to conduct *in vivo* cloning of large fragments of DNA of high G+C content by use of landing pad site-specific recombination sites. Collaborating labs focused on the optimization of conjugation and transformation protocols to develop robust, high-throughput methodologies to transfer the MHS vector between organisms as well as *in vivo* cloning techniques that did not use site-specific recombination, such as TAR cloning in *S. cerevisiae.*

4.1 Construction and Characterization of the Minimal *S. meliloti* Cell

A key principle in the field of synthetic biology and biological engineering is the reduction or minimization of the genome of the organism of interest. This helps facilitate "bottom-up" approaches towards adding back exogenous DNA for a variety of purposes, whether it be for commercial/industrial applications or for basic research. Here, a minimal *S. meliloti* strain was developed that lacks the pSymA and pSymB megaplasmids as well as the endogenous Type I restriction endonuclease *hsdR* gene. The removal of pSymA and pSymB, performed previously (diCenzo, et al., 2014), allowed for the addition of the MHS (termed pGE) vectors into these strains. The pGE vectors contain one of the origins of replication (*repA2B2C2* or *repA1B1C1*) found on pSymA and pSymB, respectively. An incompatibility factor, *incA*, is encoded within the *repB* and *repC* intergenic region and thus, pSymB cannot be stably co-inherited with another replicon carrying this region, such as the pGE plasmids (diCenzo, et al., 2014).

Deletion of the *hsdR* gene was conducted previously in the wild type RmP110 strain (Zamani and Finan, 2016, unpublished). In this work, the Δ*hsdR* mutant was transferred into background strains lacking either pSymA, pSymB or both, as well as a Rif^r ΔpSymA strain. This generated several minimized genome strains suitable as the host strain for this project. In particular, RmP4258 was generated, which is a ΔpSymA ΔpSymB Δ*hsdR* strain and was used in many of the experiments in this project. RmP4124 is another useful strain generated which is ΔpSymA $pSymB^+$ Δ*hsdR*.

In the development of the minimal *S. meliloti* ΔpSymA ΔpSymB Δ*hsdR* host as part of this project, characterization of this strain was conducted to determine parameters such as the growth rate of strains, the transformation efficiency, conjugation efficiency, and the ability to maintain pGE plasmids without antibiotic selection. As a preliminary experiment, a growth curve of the *hsdR*+ and Δ*hsdR* strains was conducted to determine if the deletion of this gene affected the growth rate of the strain (see Figure 6A). It was shown that the deletion of the *hsdR* gene did not affect the growth rate of the cell, but rather the presence of either pSymA or pSymB affected the rate at which cells grew, which was observed previously by diCenzo et al. (2014). While it was observed that the RmP4258 strain grew at a reduced rate relative to strains harbouring pSymB such as RmP4124, it was still used for downstream applications due to its ability to harbour two separate pGE plasmids, which was subsequently explored in further experiments. Furthermore, as observed in Figure 6C and 6D, it was shown that RmP4258 strains harbouring the pGE plasmids grew at a slower rate than RmP4258 alone. The average growth rate of strains harbouring a pGE plasmid was ~4.3 hour, relative to a 3.3 hour growth rate for the parental strain alone. This demonstrates that the pGE plasmids have a moderate effect on the growth rate of the minimal cell.

The *hsdRSM* operon is involved in restriction of foreign incoming DNA into the cell. It encodes a Type I restriction endonuclease system that cleaves unmethylated DNA that enters the cell. To prevent unwanted restriction of self-DNA, the *hsdM* gene is involved in the methylation of the *S. meliloti* genome, whereas the *hsdR* gene is involved in the restriction of unmethylated DNA (Murray, 2000). Since exogenous DNA transformed into the cell, such as the pGE plasmids, is unmethylated, the deletion of the *hsdR* gene is expected to improve the transformation efficiency of *S. meliloti* strains, as was previously observed by Ferri et al. (2010), where a 20-200 fold increase was observed in the electroporation efficiency of a Δ*hsdR* strain relative to a wild type strain. The mutant strain was observed to have an electroporation efficiency of 10^7 CFU/µg DNA relative to the wild type strain, which was shown to have an electroporation efficiency of 10^3 CFU/µg DNA. One interesting observation made by this team was that the electroporation efficiency of the closely related strain *S. medicae* was seen to be much higher than that of wild type *S. meliloti,* 10^6 CFU/µg DNA vs. 10^3 CFU/µg DNA, respectively. This highlights the need to develop optimized transformation protocols for *S. meliloti* for use in this project due to its naturally lower transformation efficiency. Unfortunately, the high electroporation efficiency observed by Ferri et al. was unable to be replicated here, as only a very low number of transformants were obtained (data not shown).

As a means to explore alternative transformation procedures, a PEG-mediated chemical transformation protocol was optimized and then tested for this project (Huang and Karas, 2017, unpublished). Using the PEG-mediated chemical transformation protocol, there was a 5-16 fold increase in transformation efficiency of pGE4 or pGE6 between *hsdR+* strains relative to Δ*hsdR* strains. Among the strains tested, the RmP4258 strain was shown to have the lowest transformation efficiency of 7.7 x 10^3 CFU/µg DNA and the RmP4124 strain was shown to have

the highest of 3.2 x 10^4 CFU/µg DNA. While this degree of transformation efficiency was sufficient to recover positive colonies harbouring the pGE plasmids, ideally a transformation efficiency of ~10^7 CFU/µg DNA or greater would be desired to develop robust, quick, and high-throughput protocols for use of this system as a genetic toolkit, especially when looking to transform large fragments of DNA greater than 100 kb. To further improve transformation protocols, other techniques should be explored and evaluated such as freeze-thaw techniques (Vincze & Bowra, 2006; Selvaraj & Iyer, 1981; Courtois, Courtois, & Guillaume, 1988), modified chemical transformation techniques (Tu, et al., 2005), or to re-examine the use of electroporation, which has been shown to be useful for highly efficient transformation of other rhizobia species (Garg, Dogra, & Sharma, 1999; Lin, 1994).

Here, conjugal transfer was used to transfer each of the pGE plasmids from an *E. coli* donor to the minimal cell RmP4258 recipient. Following this, a second pGE plasmid with a different *repABC* region was then subsequently conjugated into the same strain. The conjugation frequency of the first experiment was between 10^{-3}-10^{-5}, depending on the plasmid, whereas the conjugation frequency in the second experiment was between 10^{-2}-10^{-4}. Although there appears to be a slight increase in the conjugation frequency when a second plasmid is transferred, the experiment was only completed one time and values may change upon a second experiment. Regardless, these frequencies are consistent, albeit slightly higher with those found in literature (Sibley, MacLellan, & Finan, 2006).

To assess inheritance stability of the pGE plasmids, RmP4258 harbouring one or two of the pGE plasmids was cultured in media without the addition of antibiotics and grown for 5 days with subculturing at the beginning of each day for a total of roughly 55 generations. The results

of Figure 8 showed that in an RmP4258 background, ~25% of cells had lost the pGE3 plasmid after ~50 generations. Results of Figure 9 show that, after ~50 generations, the strains harbouring pGE5 and either pGE3 or pGE4 that 75-80% of patched colonies were unable to grow on LBmc FeCo with appropriate antibiotics. In contrast, results show that strains harbouring either pGE6 or pGE7 and pGE3 or pGE4 that 25-28% of patched colonies were unable to grow on LBmc FeCo with appropriate antibiotics, similar to data demonstrated in Figure 8. As a follow-up experiment to the two plasmid stability assay, patched colonies unable to grow on media with two antibiotics were again re-patched onto two separate media each with one antibiotic corresponding to the two plasmids present initially in the strain. Patched colonies unable to grow on these media were enumerated to determine if that colony had lost one or both plasmids during the initial stability assay (see Figure 10). The data presented here represents the average from the total colonies from each time point (5 in total) that were unable to grow on each antibiotic tested (see Table 10 in Appendix). In Figure 10A and B, it was found that ~80% of the re-patched colonies were unable to grow on Sp (100µg/mL), and ~25% of colonies were unable to grow on either Tc (5µg/mL) or Nm (200µg/mL). This shows the pGE5 plasmid was lost with higher frequency than pGE3 or pGE4. Interestingly, results demonstrated in Figure 10C and D show that roughly ~90-95% of re-patched colonies were unable to grow on either of the antibiotics tested. The fact that both plasmids appear to have been lost in most of these re-patched colonies would suggest that the plasmids had recombined together and were lost concurrently. Due to the high amounts of homology found on the backbones of the pGE plasmids (a very large 9.5 kb), it is very likely that recombination between the two plasmids occurred in these experiments.

The *repABC* operon is present on the large, low-copy replicons pSymA and pSymB, and allows replication and segregation into daughter cells during cell division (Cervantes-Rivera, et al., 2011). Although the *repABC* operon has been shown to provide efficient maintenance of low-copy replicons (MacLellan, et al., 2006), moderate levels of plasmid loss were observed when pGE plasmids were investigated. In the case of the two plasmid stability assay, the high degree of loss of the pGE5 plasmid may be due to a mutation in the *repABC* loci, whereby single point mutations present in either *repA, repB, or repC* have been shown to cause high levels of plasmid instability (MacLellan, et al., 2006; Yip, 2013). Further analysis of the sequence of the pGE5 plasmid should be completed in order to confirm if this may be the case.

The stability assay used in this project could be improved upon by the addition of a reporter gene such as eGFP or *tdimer2* to the pGE plasmids. Such a gene would allow rapid and more quantitative screening of strains that have lost the plasmid(s) of interest by means of fluorescent detection, which may be accomplished using flow cytometry, as described by Döhlemann et al. (2017).

Results of the Eckhardt gel indicate that the sample pGE6 in RmP4258 pGE4 (RmP4312) and sample pGE7 in RmP4258 pGE3 (RmP4313) had a single DNA fragment, instead of two separate plasmids. This suggests that the two pGE plasmids had recombined, which was expected based on the results from the re-patching experiment.

Characterization of a strain harbouring an *FRT* site at the *hypRE* gene (*hypRE*::*FRT*) locus was performed as a means to verify the capture of two large Gm^r Nm^r fragments (pA132 and a plasmid consisting of the A117, A118, A301, and A121 regions) of high G+C content DNA. This was performed to further expand the host strains that may be useful in this project. The wild

type *hypRE* locus was transduced into strains where large plasmids had been putatively captured via site-specific recombination to determine if the large plasmids had properly incorporated at this site. Successful transductants were recovered on M9 minimal media with hydroxyproline as sole carbon source, where recipient alone controls were unable to grow on the same media. These transductants were unable to grow on M9 Hyp Gm (60µg/mL) Nm (200µg/mL) (i.e. Gm^S Nm^S) as expected following recombination of the wild type *hypRE* region and the concomitant loss of the Gm and Nm resistance.

4.2 Cloning of Large Plasmids with High G+C Content DNA into pGE Plasmid

A proof-of-concept experiment was devised to capture large DNA fragments at a landing pad site cloned into the *PacI* site of pGE4. The large fragments used here were plasmids pA132 (~58.6% G+C), pTH3255 (~59.3% G+C), and pTH3247 (~63.1% G+C), which consisted of a putative bacteriophage resistance locus, the pSymA minimal symbiotic nitrogen fixation region (SNF), and the pSymB minimal SNF region, respectively.

To verify the capture of the pA132, pTH3255, and pTH3247 plasmids in pGE4, diagnostic colony PCR was conducted using various sets of PCR primers. Figure 15 demonstrates the incorporation of the ΦC31 integrase in RmP4258 via agarose gel electrophoresis of colony PCR samples. ΦC31 integrase is required to mediate the recombination of pTH3247 via the *attB/attP* system. A ~1 kb band is observed in 5 sample colonies, as expected. Figure 16 demonstrates the Flp-catalyzed excision of the intervening region of the landing pad sequence in pGE4-LPE in RmP4258 and RmP4124. Plasmids from strains with the putatively excised region were isolated and diagnostically digested using *EcoRI*. If the intervening landing pad sequence was excised, two bands of 8474 bp and 9072 bp in size would be observed, whereas if the region was not

excised, three bands of sizes 3472 bp, 5730 bp, and 8474 bp would be observed. Two bands of appropriate size were observed in both the pGE4-LPE samples in RmP4258 and RmP4124.

Colony PCR was then conducted to verify the presence of these large fragments in *S. meliloti*. Figure 17 demonstrates the presence of pA132 and pTH3255 in Gm^r Nm^r strains of RmP4258 and RmP4124. Two diagnostic primer sets were used, which amplified a region specific to the plasmid being tested. An additional primer set, specific to pSymA but outside of the region of interest was included to demonstrate that the entire pSymA was not present. In each experiment, a control set of primers was included as a colony PCR control. In each instance, the expected band sizes were observed for primers specific to pA132 and for primers specific to pTH3255 as well as for the control primers. In each instance where a band was not expected, no band was observed. Figure 18 demonstrates the presence of pTH3247 in RmP4317. Two diagnostic primers specific to the region of interest were used as well as one set specific to pSymB outside of this region. In each instance, the expected band sizes were observed. The PCR reactions shown in Figures 17 and 18 demonstrate that the large fragments of interest are present within the recipient *S. meliloti* strains. These strains were streak purified and used in downstream diagnostic assays.

The pA132, pTH3255, and pTH3247 plasmids are suicide vectors in *S. meliloti* and thus unable to replicate unless integrated into the genome. As mentioned previously, pGE4-LP was the only plasmid constructed containing the landing pad sequence. For use in this *in vivo* cloning experiment, it is therefore not the most suitable vector for the capture of the three large fragments used, as both DNA fragments contain the *nptII* gene and *oriT* sequence, representing roughly 2.1 kb of homology. Furthermore, following the deletion of the *hsdR* gene in the

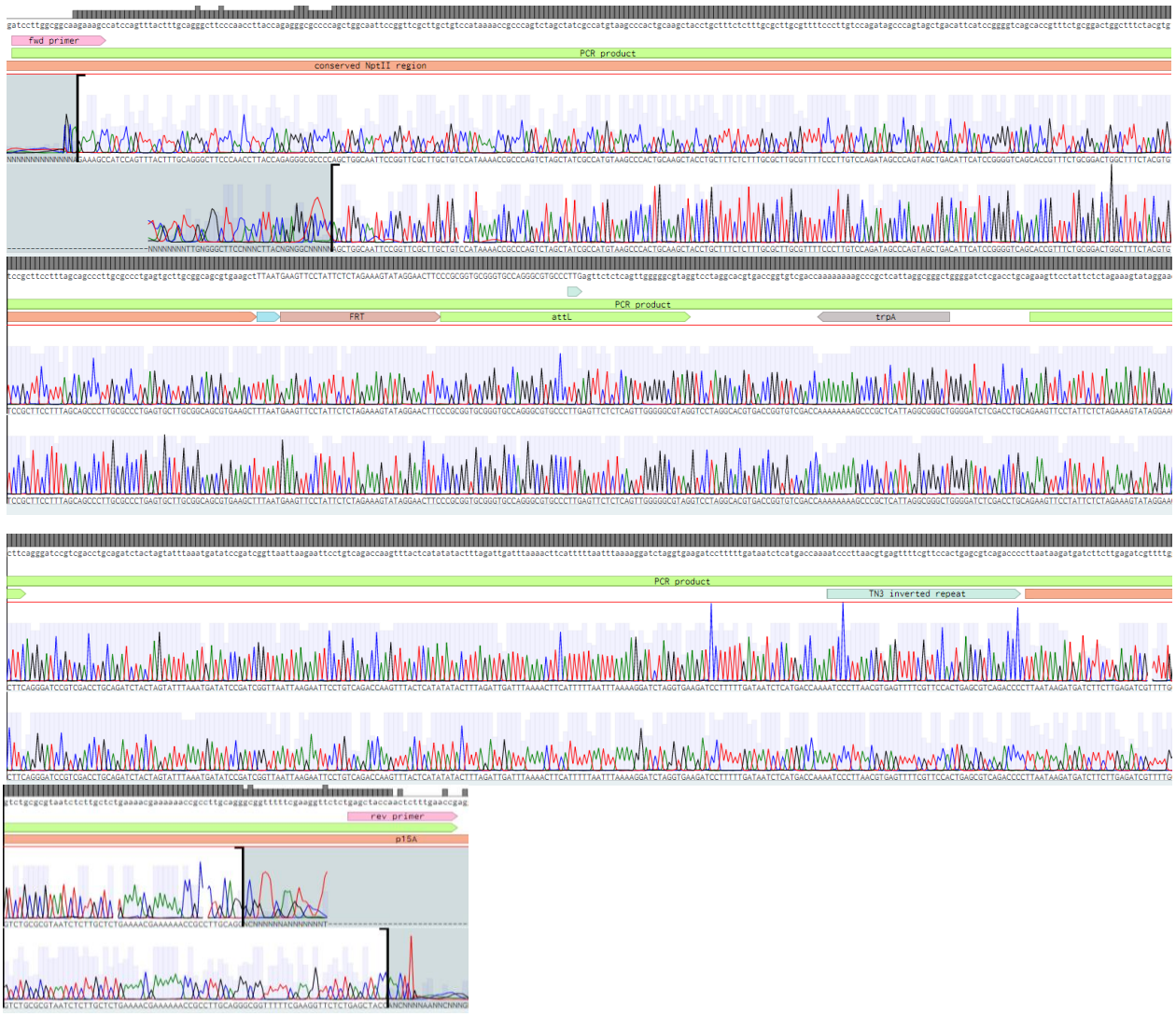
background strains RmP4258 and RmP4124, a single *FRT* site remained. Upon conjugation of the pA132 and pTH3255 plasmids into the background strains harbouring the pGE4-LPE plasmid and pTH2505, recombination may occur at this residual *FRT* site as opposed to the *FRT* site located on the pGE4-LPE plasmid. Therefore, an additional colony PCR screening experiment would be required to determine if the pA132 and pTH3255 plasmids had recombined at this loci.

As such, homologous recombination may occur at sites other than the site-specific recombination sites, and thus colonies were screened using diagnostic colony PCR to determine if recombination had occurred at the intended landing pad locus using 3 primer sets, which can be seen in Figure 19. Figure 20A demonstrates that the pA132 and pTH3255 plasmids were not captured at the *FRT* site in both RmP4258 and RmP4124. Here, a 450 bp band was observed using a primer set that spans the landing pad region, which would only be observed if no recombination took place at the *FRT* site. Furthermore, a primer set designed to amplify the border region of the large fragment and the pGE4-LPE fragment did not produce the expected band size. Despite this, a primer set designed to amplify a region from pGE4-LPE yielded the correct band size, demonstrating that the plasmid was present. When taking into consideration the previous results showing the presence of the large fragments, the data suggest that the large fragments had recombined in the pGE4-LPE plasmid, which may have occurred at the *nptII* gene or the *oriT* region on the pGE4-LPE plasmid or possibly at a loci in the chromosome.

Figure 20B demonstrates that the pTH3247 plasmid was captured at the *attB* site in the landing pad of pGE4-LP, and the resulting plasmid was designated pTH3264 in strain RmP4317. No band was observed using the primer set spanning the landing pad region, suggesting recombination had occurred preventing amplification of a band. Furthermore, a band of the

expected size was observed using primers spanning the border region where recombination should occur as well as a band indicating the presence of pGE4-LP. To further confirm the region was captured, sequencing of this PCR product was conducted (see Figure 21). The data suggest that the pTH3247 plasmid was successfully captured at the *attB* site in the pGE4-LP vector, as the PCR primer aligns with 100% match to the expected recombinant plasmid.

A



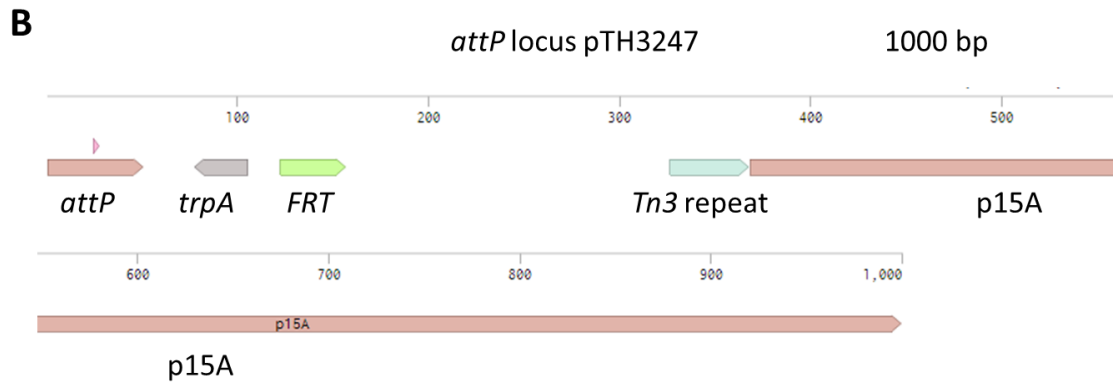


Figure 21: (A) Alignment of sequencing results for the PCR product of the border region where pTH3247 was captured in pGE4-LP (pTH3264). Sequencing was performed using primers check_primer_Fwd and check_primer2_Rev and results were aligned to an *in silico* construct pGE4-LP with pTH3247 captured at *attB* site. (B) Plasmid map of site 1 kb immediately downstream of *attP* recombination site on pTH3247 plasmid. Results were shown to align as expected.

This proof-of-concept experiment demonstrated the ability of the pGE4-LP plasmid to capture a large plasmid pTH3247 that was 205 kb in size and >63% G+C content. As such, this *in vivo* cloning strategy was deemed successful and may be used in other experiments to capture large genomic fragments in this system. Here, the *attB/attP* sites were used to capture the fragment of interest. An advantage of using the *attB* and *attP* sites is the directionality of use. The integrase-mediated recombination is non-reversible in the presence of the serine integrase alone, and as such offers a cloning system with reduced chance to recombine out. Excision of these sites may only occur with the addition of an excisionase (excicase) enzyme or recombination directionality factors (RDFs) (Fogg, et al., 2014).

The capture of the pTH3247 plasmid is significant as it comprises the minimal SNF genetic regions found on pSymB (Kearsley and Finan, 2018, unpublished). With the pTH3264 plasmid, further manipulation of this region can be performed in a streamlined manner to further characterize the genes essential for symbiotic nitrogen fixation, and to eliminate and further reduce genes deemed unnecessary in this process. The ability of the pGE plasmids to be conjugated to other organisms will further increase the genetic tools available to study the minimal SNF region. Once the minimal SNF region is determined, the resulting strain may be used as a suitable platform for identification of SNF-promoting genes through the screening of metagenomics libraries, which may also be facilitated by the pGE plasmids (see Section 4.4) (diCenzo, et al., 2016).

To further demonstrate the usefulness of the pGE plasmids to transfer large DNA fragments between organisms, the conjugation experiments of pTH3264 plasmid between *E. coli* and *S. meliloti* should be re-done. Despite recovering Rif^r Gm^r transconjugants (indicating transfer of the pTH3247 plasmid) colony PCR to detect the presence of the pGE4-LP plasmid failed. This suggests pTH3247 had recombined out of pGE4-LP prior to the conjugation event and was then conjugated into the *E. coli* Rif^r strain while the pGE4-LP plasmid was not transferred. This experiment should be completed again to further characterize the ability to mobilize this captured fragment between different species of bacteria.

As mentioned earlier, it should be noted that this work should be re-done using a pGE3-LP plasmid to minimize unwanted off-target homologous recombination events to take place following successful recombination in this experiment as a stable final construct is desired for downstream work.

4.3 Further Optimization Experiments to Perform

As noted in Figure 6B, the growth rate of the background minimal ΔpSymAB Δ*hsdR* strain, RmP4258, used for this project was roughly 2 times slower than that of strains containing pSymB. While it was observed that the pSymB^{+} strain RmP4124 was able to grow at a faster rate, this strain is limited in its ability to harbour more than one pGE plasmid at the current time. Two separate strategies may be used to circumvent this issue. The first would be to use other *repABC* operons to expand the combinability of the pGE plasmids. Döhlemann et al., (2017), used three different *repABC* modules to make three separate cloning vectors for use *in vivo* cloning experiments using site-specific recombination sites. Rhizobia species, such as *Rhizobium etli*, have been shown to harbour up to 6 independent replicons aside from the main chromosome (Gonzalez, et al., 2006). With the introduction of alternative *repABC* operons, two or more pGE plasmids could be harboured in the RmP4124 strain.

A second method to improve the use of this system would be to increase the growth rate of the RmP4258 (ΔpSymAB Δ*hsdR*) strain. A pLAFR1 cosmid library (Friedman, et al., 1982) carrying wild type *S. meliloti* DNA can be transferred into RmP4258 and colonies demonstrating an increased growth rate could be selected and verified.

Following the successful capture of pTH3247 in RmP4317, a stability assay should be conducted to determine the ability of the host strain to maintain pTH3264.

4.4 Limitations of pGE Plasmids and Minimal *S. meliloti* Cell

To better facilitate the use of two pGE plasmids in the same host strain for future experiments and to limit the amount of recombination that may occur between the plasmids, a re-design of the backbone plasmid must be undertaken. This may include the substitution of

alternative selective markers for *S. cerevisiae* or *P. tricornutum*, which currently utilize a biosynthesis gene and NAT resistance gene, in a similar manner done for the antibiotic selective markers used for selection in bacterial strains. Other selective markers may include the uracil biosynthesis (URA3) gene for selection in *S. cerevisiae* and zeocin resistance gene in *P. tricornutum* (Karas, et al., 2015). As reported by Döhlemann, et al. (2016) in a paper regarding the creation of recombinant plasmids (pABC) with the *repABC* origins of replication, similar to the pGE plasmids, it was noted that parts of the pABC plasmids were designed "on the premise of avoiding high and extended sequence identity" to avoid unwanted recombination between plasmids present in the same host strain. Following similar logic, this must be applied to the design of the pGE plasmids as well in future downstream work.

Furthermore, to improve the stability of the pGE plasmids, other *repABC* origins of replication may be used that have been shown to have improved stability. Sequencing of the *repABC* origins of replication currently present in the pGE plasmids should be completed as well to determine if any mutations exist in these regions which may lead to the instability observed in this project. As mentioned previously, the growth rate of the minimal *S. meliloti* cell should be optimized as well to improve the stream-lined use of protocols when working with this system. This may be achieved by screening with a pLAFR1 library containing the *S. meliloti* genome to find regions that may improve the growth of this strain.

4.5 Future Prospective Projects

The pGE plasmids may be used to capture a wide-range of large fragments of high G+C content DNA. Cloning of genomic regions of interest for the organism *Deinococcus radiodurans* is currently being conducted by the collaborating Karas lab, an important organism in the study of

chromosome repair systems and may be a useful chassis strain for biotechnology endeavors (Gerber, et al., 2015). pGE plasmids may also be useful as backbone vectors for the construction of single organism genomic or metagenomic libraries derived from soil microorganisms due to its ability to maintain large fragments of high G+C content DNA as well as its ability to be conjugated in an intra- and inter-kingdom manner (Brumwell et al., 2018, to be published). This would facilitate functional metagenomics or bioprospecting and allow for transfer of useful DNA fragments into multiple organisms. One use of a metagenomic library is based on the successful expression of heterologous genes in the host strain harbouring the vector, and as such, the ability of pGE plasmids to be transferred and then replicated in several hosts offers an advantage in terms of functional metagenomics analysis (Neufeld, et al., 2011; Dohlemann, et al., 2017).

Alternative *in vivo* cloning strategies include the use of recombinase mediated cassette exchange (RMCE). This strategy involves the direct transfer and exchange of two cassette sequences via the double homologous recombination of site-specific recombination sites flanking the region of interest and is mediated by a recombinase (see Figure 22). The design of the landing pad used in this study would enable the use of RMCE to exchange DNA fragments between an incoming donor plasmid and a recipient plasmid. By cloning a selectable marker into the *attB* or *lox71* site of the landing pad in pGE4-LP, RMCE could then be used to exchange an incoming plasmid with the region of interest being flanked by mutant *FRT* sites in the presence of Flp recombinase. The *FRT* sites would be mutated version such as *FRT G* or *FRT H* (Takata, et al., 2011) to minimize downstream recombination upon exchange of cassettes.

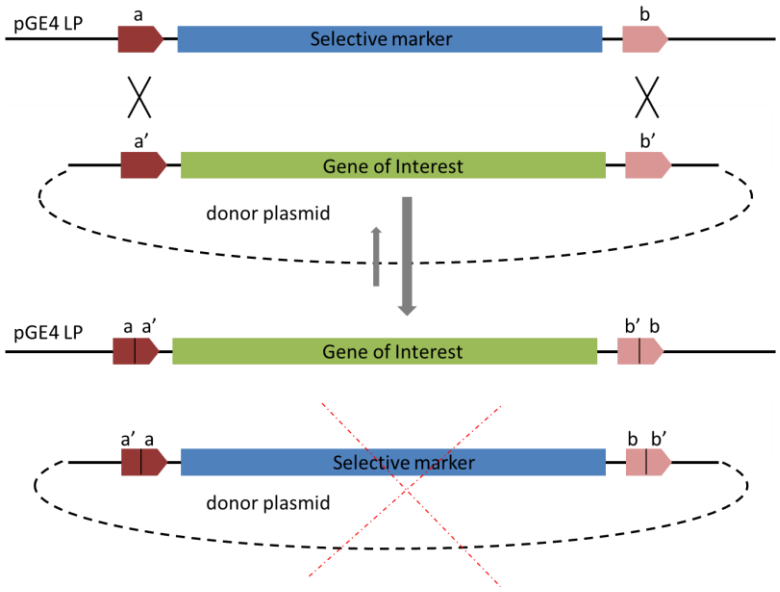


Figure 22: Diagram depicting recombinase-mediated cassette exchange (RMCE) that may be employed using the pGE plasmids. Red and pink arrows denote mutant *FRT* sites flanking a selective marker (blue rectangle) or a genomic region of interest (green rectangle). If the donor plasmid is unable to replicate in the *S. meliloti* host strain, it will be lost following exchange of DNA cassettes.

To develop a robust system for use in cloning experiments, a stable MHS vector would be required. For use in commercial applications, a plasmid that relies on nonselective maintenance would be desired. One method to accomplish this would be through the integration of toxin-antitoxin maintenance systems, gene pairs that are found throughout the genome of *S. meliloti* (Milunovic, et al., 2014). Following the cloning of large DNA fragments in the MHS vectors, several methods of downstream engineering could then be used to make changes in the cloned DNA fragment, including CRISPR/Cas9.

4.6 Concluding Statement

Overall, the primary goal of this project aimed to characterize a minimal *S. meliloti* ΔpSymA ΔpSymB Δ*hsdR* strain to be used as a cloning platform in synthetic biology applications. These include the *in vivo* cloning of large fragments of high G+C content DNA using a multi-host shuttle vector. A secondary focus was to perform a proof-of-concept experiment to clone three large plasmids DNA regions with high G+C content DNA (58-63%) into a landing pad sequence located on pGE4-LP/LPE plasmid. Characterization of the strain was completed and showed reasonable conjugation and transformation efficiencies. The efficiency of transformation can continue to be improved by continued optimization of the PEG-mediated protocol used or by exploring other protocols such as freeze-thaw or electroporation. Stability assays of the pGE plasmids showed moderate stability of pGE3, pGE4, pGE6, and pGE7. However, pGE5 showed high levels of instability and this should be investigated further.

Bibliography

Albert, H., Dale, E. C., Lee, E., & Ow, D. W. (1995). Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. *The Plant Journal, 7*(4), 649-659.

Andrews, B. J., Proteau, G. A., Beatty, L. G., & Sadowksi, P. D. (1985). The FLP Recombinase of the 2 Micron Circle DNA of Yeast: Interaction with its Target Sequences. *Cell, 40*, 795-803.

Araki, K., & Yamamura, K.-i. (2012). Genetic Manipulations Using Cre and Mutant LoxP Sites. In A. Morozov, *Neuromethods* (pp. 29-45). Totowa, NJ: Humana Press.

Argos, P., Landy, A., Abremski, K., Egan, J. B., Haggard-Ljungquist, E., Hoess, R. H., et al. (1986). The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO Journal, 5*, 433-440.

Baker, B. (2017). Synthetic Biology and the Marketplace: Building the new bioeconomy. *BioScience, 67*(10), 877-883.

Bartsez, A. V., Deakin, W. J., Boukli, N. M., McAlvin, C. B., Stacey, G., Malnoe, P., et al. (2004). NopL, an effector protein of Rhizobium sp. NGR234, thwarts activation of plant defense reactions. *Plant Physiology, 134*, 871-879.

Boivin, C., Barran, L. R., Malpica, C. A., & Rosenberg, C. (1991). Genetic Analysis of a Region of the Rhizobium meliloti pSym Plasmid Specifying Catabolism of Trigonelline, a Secondary Metabolite Present in Legumes. *Journal of Bacteriology, 173*(9), 2809-2817.

Breitling, R., Takano, E., & Gardner, T. S. (2015). Judging synthetic biology risks. *Science, 347*(6218), 107.

Carbonell, P., Currin, A., Jervis, A. J., Rattray, N. J., Swainston, N., Yan, C., et al. (2016). Bioinformatics for the synthetic biology of natural products: integrating across the Design–Build–Test cycle. *Natural Product Reports, 33*(8), 925-932.

Cervantes-Rivera, R., Pedraza-Lopez, F., Perez-Segura, G., & Cevallos, M. A. (2011). The replication origin of a repABC plasmid. *BMC Microbiology, 11*(158), 1-14.

Chan, W.-T., Verma, C. S., Lane, D. P., & Gan, S. K.-E. (2013). A comparison and optimization of methods and factors affecting the transformation of Escherichia coli. *Bioscience Reports, 33*(6), e00086.

Courtois, J., Courtois, B., & Guillaume, J. (1988). High-Frequency Transformation of Rhizobium meliloti. *Journal of Bacteriology, 170*(12), 5925-5927.

de Lima Procopio, R. E., da Silva, I. R., Martins, M. K., de Azevedo, J. L., & de Araujo, J. M. (2012). Antibiotics produced by Streptomyces. *The Brazilian Journal of Infectious Diseases, 16*(5), 466-471.

diCenzo, G. C., & Finan, T. M. (2018). Techniques for Large-Scale Bacterial Genome Manipulation and Characterization of the Mutants with Respect to In Silico Metabolic Reconstructions. In M. Fondi, *Metabolic Network Reconstruction and Modeling. Methods in Molecular Biology* (Vol. 1716, p. 1). New York, NY: Humana Press.

diCenzo, G. C., MacLean, A. M., Milunovic, B., Golding, G. B., & Finan, T. M. (2014). Examination of Prokaryotic Multipartite Genome Evolution through Experimental Genome Reduction. *PloS Genetics, 10*(10), e1004742.

diCenzo, G. C., Zamani, M., Ludwig, H. N., & Finan, T. M. (2017). Heterologous Complementation Reveals a Specialized Activity for BacA in the Medicago-Sinorhizobium meliloti Symbiosis. *Molecular Plant-Microbe Interactions, 30*(4), 312-324.

diCenzo, G. C., Zamani, M., Milunovic, B., & Finan, T. M. (2016). Genomic resources for identification of the minimal N2-fixing symbiotic genome. *Environmental Microbiology*, *18*(8), 1-18.

Dohlemann, J., Brennecke, M., & Becker, A. (2016). Cloning-free genome engineering in Sinorhizobium meliloti advances applications of Cre/loxP site-specific recombination. *Journal of Biotechnology, 223*, 160-170.

Dohlemann, J., Wagner, M., Happel, C., Carrillo, M., Sobetzko, P., Erb, T. J., et al. (2017). A Family of Single Copy repABC-Type Shuttle Vectors Stably Maintained in the Alpha-Proteobacterium Sinorhizobium meliloti. *ACS Synthetic Biology, 6*, 968-984.

Du, Q., Wang, H., & Xie, J. (2011). Thiamin (Vitamin B1) Biosynthesis and Regulation: A Rich Source of Antimicrobial Drug Targets? *International Journal of Biological Sciences, 7*(1), 41-52.

Eckhardt, T. (1978). A Rapid Method for the Identification of Plasmid Desoxyribonucleic Acid in Bacteria. *Plasmid, 1*, 584-588.

Esvelt, K. M., & Wang, H. H. (2013). Genome-scale engineering for systems and synthetic biology. *Molecular Systems Biology, 9*, 641.

Ferri, L., Gori, A., Biondi, E. G., Mengoni, A., & Bazzicalupo, M. (2010). Plasmid electroporation of Sinorhizobium strains: The role of the restriction gene hsdR in type strain Rm1021. *Plasmid, 63*, 128-135.

Finan, T. M., Hartwieg, E., Lemiuex, K., Bergman, K., Walker, G. C., & Signer, E. R. (1984). General transduction in rhizobium meliloti. *Journal of Bacteriology, 159*(1), 120-124.

Finan, T. M., Kunkel, B., De Vos, G. F., & Signer, E. R. (1986). Second symbiotic megaplasmid in Rhizobium meliloti carrying exopolysaccharide and thiamine synthesis genes. *Journal of Bacteriology, 167*(1), 66-72.

Finan, T. M., O'Brian, M. R., Layzell, D. B., Vessey, J. K., & Newton, W. (2001). *Nitrogen Fixation: Global Perspectives.* New York, NY: CABI Publishing.

Fogg, P. C., Colloms, S., Rosser, S., Stark, M., & Smith, M. C. (2014). New Applications for Phage Integrases. *Journal of Molecular Biology, 426*(15), 2703-2716.

Frage, B., Dohleman, J., Robledo, M., Lucena, D., Sobetzko, P., Graumann, P. L., & Becker, A. (2016). Spatiotemporal choreography of chromosome and megaplasmids in the Sinorhizobium meliloti cell cycle. *Molecular Microbiology, 100*(5), 808-823.

Friedman, A. M., Long, S. R., Brown, S. E., Buikema, W. J., & Ausubel, F. M. (1982). Construction of a broad host range cosmid cloning vector and its use in the genetic analysis of Rhizobium mutants. *Gene, 18*(3), 289-296.

Gaj, T., Gersbach, C. A., & Barbas III, C. F. (2013). ZFN, TALEN and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology, 31*(7), 397-405.

Galibert, F., Finan, T. M., Long, S. R., Chain, P., Cowie, A., Hernandez-Lucas, I., et al. (2001). The composite genome of the legume symbiont Sinorhizobium meliloti. *Science, 293*(5530), 668-672.

Garg, B., Dogra, R., & Sharma, P. K. (1999). High-Efficiency Transformation of Rhizobium leguminosarum by Electroporation. *Applied and Environmental Microbiology, 65*(6), 2802-2804.

Gerber, E., Bernard, R., Castang, S., Chabot, N., Coze, F., Dreux-Zigha, A., et al. (2015). Deinococcus as new chassis for industrial biotechnology: biology, physiology and tools. *Journal of Applied Microbiology, 119*, 1-10.

Gibson, D. G., Benders, G. A., Adrews-Pfannkoch, C., Denisova, E. A., Baden-Tillson, H., Zaveri, J., et al. (2008). Complete Chemical Synthesis, Assembly, and Cloning of a Mycoplasma genitalium Genome. *Science, 319*(5867), 1215-1220.

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison III, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods, 6*, 343-345.

Glass, J. I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M. R., Maruf, M.,et al. (2006). Essential genes of a minimal bacterium. *PNAS, 103*(2), 425-430.

Gonzalez, V., Satamaria, R. I., Bustos, P., Hernandez-Gonzalez, I., Medrano-Soto, A., Moreno-Hagelsieb, G., et al. (2006). The partitioned Rhizobium etli genome: Genetic and metabolic redundancy in seven interacting replicons. *Proceedings of the National Academy of Sciences of the United States of America, 103*(10), 3834-3839.

Gopaul, D. N., Van Duyne, G. D., & Guo, F. (1999). Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse. *Proc Natl Acad Sci, 96*, 7143-7148.

Groth, A. C., Olivares, E. C., Thyagarajan, B., & Calos, M. P. (2000). A phage integrase directs efficient site-specific integration in human cells. *Proceedings of the National Academy of Sciences USA, 97*(11), 5995-6000.

Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. *Journal of Molecular Biology, 166*(4), 557-580.

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

Hutchison III, C. A., Chuang, R.-Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., et al. (2016). Design and synthesis of a minimal bacterial genome. *Science, 351*(6280), 1414-1422.

Hynes, M. F., Simon, R., & Puhler, A. (1985). The Development of Plasmid-free Strains of Agrobacterium tumefaciens by Using Incompatibility with a Rhizobium meliloti Plasmid to Eliminate pAtC58. *Plasmid, 13*, 99-105.

Imanaka, T., & Aiba, S. (1981). A Perspective on the Application of Genetic Engineering: Stability of Recombinant Plasmid. *Annals New York Academy of Sciences*, 1-14.

Jones, J. D., & Gutterson, N. (1987). An efficient mobilizable cosmid vector, pRK7813, and its use in a rapid method for marker exchange in Pseudomonas fluorescens strain HV37a. *Gene, 61*(3), 299-306.

Karas, B. J., Diner, R. E., Lefebvre, S. C., McQuaid, J., Phillips, A. P., Noddings, C. M., et al. (2015). Designer diatom episomes delivered by bacterial conjugation. *Nature Communications, 6*, 1-10.

Karas, B. J., Molparia, B., Jablonvic, J., Hermann, W. J., Lin, Y.-C., Dupont, C. L., et al. (2013). Assembly of eukaryotic algal chromosomes. *Journal of Biological Engineering, 7*(30), 1-12.

Karas, B. J., Suzuki, Y., & Weyman, P. D. (2015). Strategies for cloning and manipulating natural and synthetic chromosomes. *Chromosome Research, 23*(1), 57-68.

Keravala, A., & Calos, M. P. (2008). Site-Specific Chromosomal Integration Mediated by φC31 Integrase. In J. M. Walker, *Chromosomal Mutagenesis* (pp. 165-234). Totowa, NJ: Humana Press.

Kolisnychenko, V., Plunkett III, G., Herring, C. D., Feher, T., Posfai, J., Blattner, F. R., & Posfai, G. (2002). Engineering a reduced Escherichia coli genome. *Genome Resources, 12*, 640-647.

Konieczny, I., Doran, K. S., Helinski, D. R., & Blasina, A. (1997). Role of TrfA and DnaA proteins in origin opening during initiation of DNA replication of the broad host range plasmid RK2. *Journal of Biological Chemistry, 272*, 20173-20178.

Kosuri, S., & Church, G. M. (2014). Large-scale de novo DNA synthesis: technologies and applications. *Nature Methods, 11*(5), 499-507.

Landy, A., & Ross, W. (1977). Viral Integration and Excision: Structure of the Lambda att Sites. *Science, 197*(4309), 1147-1160.

Lin, J.-J. (1994). Optimization of the transformation efficiency of Agrobacterium tumefaciens cells using electroporation. *Plant Science, 101*, 11-15.

Liu, R., Bassalo, M. C., Zeitoun, R. I., & Gill, R. T. (2015). Genome scale engineering techniques for metabolic engineering. *Metabolic Engineering, 32*, 143-154.

Lyznik, L. A., Rao, K. V., & Hodges, T. K. (1996). FLP-mediated recombination of FRT sites in the maize genome. *Nucleic Acids Research, 24*, 3784-3789.

MacLean, A. M., Haerty, W., Golding, G. B., & Finan, T. M. (2011). The LysR-type PcaQ protein regulates expression of a protocatechouate-inducible ABC-type transport system in Sinorhizobium meliloti. *Microbiology, 157*, 2522-2533.

MacLellan, S. R., Zaheer, R., Sartor, A. L., MacLean, A. M., & Finan, T. M. (2006). Identification of a megaplasmid centromere reveals genetic structural diversity within the repABC family of basic replicons. *Molecular Microbiology, 59*(5), 1559-1575.

Merrick, C. A., Zhao, J., & Rosser, S. J. (2018). Serine Integrases: Advancing Synthetic Biology. *ACS Synthetic Biology, 7*, 299-310.

Miller, J. H. (1972). *Experiments in molecular genetics.* Cold Spring Harbour, NY: Cold Spring Harbour Laboratory Press.

Milunovic, B. (2011). *Deletion Analysis of the Sinorhizobium meliloti Genome.* Hamilton, ON: McMaster University.

Milunovic, B., diCenzo, G. C., Morton, R. A., & Finan, T. M. (2014). Cell Growth Inhibition upon Deletion of Four Toxin-Antitoxin Loci from the Megaplasmids of Sinorhizobium meliloti. *Journal of Bacteriology, 196*(4), 811-824.

Mosberg, J. A., Lajoie, M. J., & Church, G. M. (2010). Lambda Red Recombineering in Escherichia coli Occurs Through a Fully Single-Stranded Intermediate. *Genetics, 186*(3), 791-799.

Murray, N. E. (2000). Type I RestrictionSystems: Sophisticated Molecular Machines. *Microbiology and Molecular Biology Reviews, 64*(2), 412-434.

Nemudryi, A. A., Valetdinova, K. R., Medvedev, S. P., & Zakian, S. M. (2014). TALEN and CRISPR/Cas Genome Editing Systems: Tools of Discovery. *ActaNaturae, 6*(3), 19-40.

Neufeld, J. D., Engel, K., Cheng, J., Moreno-Hagelsieb, G., Rose, D. R., & Charles, T. C. (2011). Open resource metagenomics: a model for sharing metagenomic libraries. *Standards in Genomic Sciences, 5*(2), 203-210.

Noskov, V. N., Karas, B. J., Young, L., Chuang, R.-Y., Gibson, D. G., Lin, Y.-C., et al. (2012). Assembly of Large, High G+C Bacterial DNA Fragments in Yeast. *ACS Synthetic Biology, 1*(7), 267-273.

Novick, R. P. (1987). Plasmid Incompatibility. *Microbiological Reviews, 51*(4), 381-395.

Oldryd, G. E., & Downie, J. A. (2008). Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annual Review of Plant biology, 59*, 519-546.

Poblete-Castro, I., Becker, J., Dohnt, K., dos Santos, V. M., & Wittmann, C. (2012). Industrial biotechnology of Pseudomonas putida and related species. *Applied Microbiology and Biotechnology, 93*, 2279-2290.

Ratner, H. K., Sampson, T. R., & Weiss, D. S. (2016). Overview of CRISPR–Cas9 Biology. *Cold Spring Harbour Protocols, 12*, 1-12.

Rutherford, K., & Van Duyne, G. V. (2014). The ins and outs of serine integrase site-specific recombination. *Current Opinions in Structural Biology, 24*, 125-131.

Sadowksi, P. D. (1995). The Flp recombinase of the 2um plasmid of Saccharomyces cerevisiae. *Progress in Nuclieic Acid Research and Molecular Biology, 51*, 53-91.

Sambrook, J., & Russell, D. W. (2001). Agarose Gel Electrophoresis. In J. Sambrook, & D. W. Russell, *Molecular Cloning 3rd Ed.* (pp. 1-12). Cold Spring Harbour, NY: Cold Spring Harbour Laboratory Press.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: a laboratory manual, 2nd Ed.* Cold Spring Harbour, NY: Cold Spring Harbour Laboratory Press.

Scherrer, A., & Denarie, J. (1971). Symbiotic properties of some auxotrophic mutants of Rhizobium meliloti and of their prototrophic revertants. *Plant Soil, 35*(1), 39-45.

Schmidt, F., & Platt, R. J. (2017). Applications of CRISPR-Cas for synthetic biology and genetic recording. *Current Opinion in Systems Biology, 5*, 9-15.

Schweizer, H. P. (2003). Applications of the Saccharomyces cerevisiae Flp-FRT system in bacterial genetics. *Journal of Molecular Microbiology and Biotechnology, 5*, 67-77.

Selvaraj, G., & Iyer, V. N. (1981). Genetic transformation of Rhizobium meliloti by plasmid DNA. *Gene, 15*, 279-283.

Senecoff, J. F., Bruckner, R. C., & Cox, M. M. (1985). The FLP recombinase of the yeast 2-nxm plasmid: Characterization of its recombination site. *Proc Natl Acad Sci, 82*, 7270-7274.

Sibley, C. D., MacLellan, S. R., & Finan, T. (2006). The Sinorhizobium meliloti chromosomal origin of replication. *Microbiology, 152*, 443-455.

Strand, T. A., Lale, R., Degnes, K. F., Lando, M., & Valla, S. (2014). A New and Improved Host-Independent Plasmid System for RK2-Based Conjugal Transfer. *PLoS One, 9*(3), e90372.

Struhl, K. (1983). Promoter Elements, Regulatory Elements, and Chromatin Structure of the Yeast his3 Gene. *Cold Spring Harbor Symposia on Quantitative Biology, 47*, 901-910.

Takata, Y., Kondo, S., Goda, N., Kanegae, Y., & Saito, I. (2011). Comparison of efficiency between FLPe and Cre forrecombinase-mediated cassette exchange in vitro and in adenovirus vector production. *Genes to Cells, 16*, 765-777.

Torres-Quesada, O., Reinkensmeier, J., Schluter, J.-P., Robledo, M., Peregrina, A., Giegerich, R., et al. (2014). Genome-wide profiling of Hfq-binding RNAs uncovers extensive post-transcriptional rewiring of major stress response and symbiotic regulons in Sinorhizobium meliloti. *RNA Biology, 11*(5), 563-579.

Tu, Z., He, G., Li, K., Chen, M., Chang, J., Chen, L., et al. (2005). An improved system for competent cell preparation and high efficiency plasmid transformation using different Escherichia coli strains. *Electronic Journal of Biotechnology, 4*, 815.

Urgun-Demirtas, M., Stark, B., & Pagilla, K. (2006). Use of Genetically Engineered Microorganisms (GEMs) for the Bioremediation of Contaminants. *Critical Reviews in Biotechnology, 26*, 145-164.

Valente, A. X., & Fong, S. S. (2011). High-Throughput Biologically Optimized Search Engineering Approach to Synthetic Biology. *Quantitative Biology*, 1-17.

Vincze, E., & Bowra, S. (2006). Transformation of Rhizobia with Broad-Host-Range Plasmids by Using a Freeze-Thaw Method. *Applied and Environmental Microbiology, 72*(3), 2290-2293.

Vooijs, M., van der Valk, M., te Riele, H., & Berns, A. (1998). Flp-mediated tissue-specific inactivation of the retinoblastoma tumor suppressor gene in the mouse. *Oncogene, 17*, 1-12.

White, C. E., Gavina, J. M., Morton, R., Britz-McKibbin, P., & Finan, T. M. (2012). Control of hydroxyproline catabolism in Sinorhizobium meliloti. *Molecular Microbiology, 85*(6), 1133-1147.

Wilkins, B. M. (1990). Factors influencing the dissemination of DNA by bacterial conjugation. In J. C. Fry, & M. J. Day, *Bacterial Genetics in Natural Environments* (pp. 22-30). New York, NY: Chapman and Hall.

Yamaguchi, Y., Park, J. H., & Inouye, M. (2011). Toxin-antitoxin systems in bacteria and archaea. *Annual Reviews Genetics, 45*, 61-79.

Yip, C. (2013). *Replication, stability, and incompatibility determinants of Rhizobium leguminosarum plasmids.* University of Calgary, Department of Biological Sciences. Calgary, AL: University of Calgary.

Yuan, Z.-C., Zaheer, R., & Finan, T. M. (2006). Regulation and Properties of PstSCAB, a High-Affinity, High-Velocity Phosphate Transport System of Sinorhizobium meliloti. *Journal of Bacteriology, 188*(3), 1089-1102.

Zhang, L.-y., Chang, S.-h., & Wang, J. (2010). How to make a minimal genome for synthetic minimal cell. *Protein & Cell, 1*(5), 427-434.

Appendix

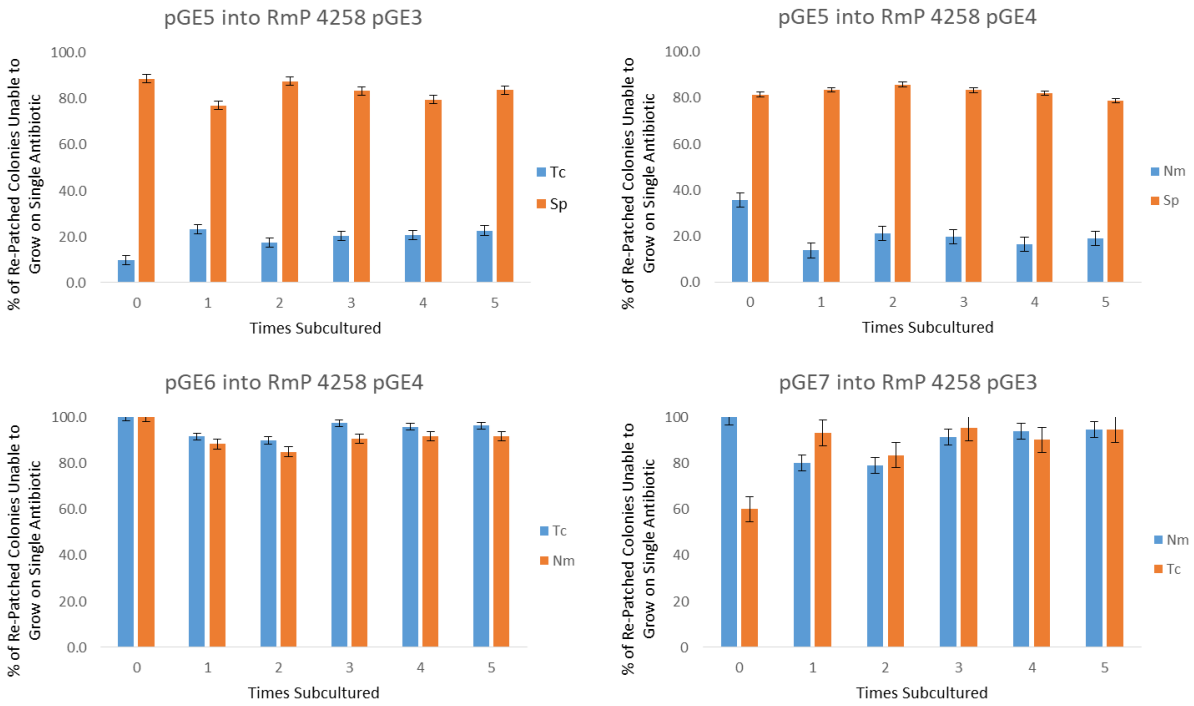


Figure 23: Percentage of re-patched colonies unable to grow on LBmc FeCo media with single antibiotic following two plasmid stability assay. Bars represent average of triplicate samples. Data is represented for 5 time points.

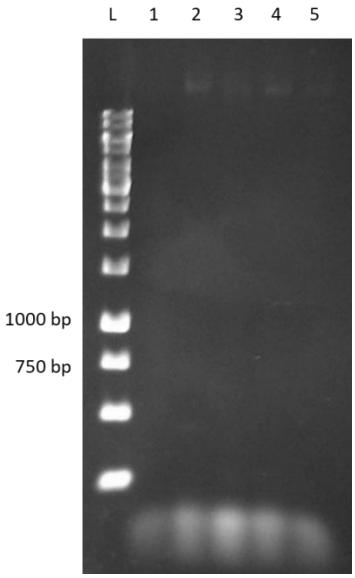


Figure 24: Colony PCR to amplify border region if the pA132 or pTH3255 plasmids recombined at the *FRT* site at the *hsdR* locus. This was conducted using the primer set hsdR U Fwd and check_primer1_Rev and a roughly 979 bp band would be observed. Lane 1 used RmP4258 with pGE4-LP and pTH3255 as template. Lane 2 used RmP4124 with pGE4-LP and pTH3255 as template. Lane 3 used RmP4258 with pGE4-LP and pA132 as template. Lane 4 used RmP4124 with pGE4-LP and pA132 as template. Lane 5 used pGE4-LP alone in RmP4124 as template. No bands were observed in any of the samples suggesting that the large plasmids did not recombine at the *hsdR* locus and instead may have recombined at another region of the pGE4-LP plasmid.

Table 10: Percentage of colonies unable to grow on media LBmc FeCo Tc (5µg/mL). 100 randomly chosen colonies were patched for each sample.

No. Generations	Sample 1	Sample 2	Sample 3	Sample Avg
0	3	2	2	2.3
9.6	7	4	6	5.6
19.2	11	5	16	10.6
28.7	13	15	19	15.6
38.5	20	16	19	18.3
48.3	24	24	28	25.3

Table 11: Percentage of colonies unable to grow on media with two antibiotics to select for each pGE plasmid present. Data represents triplicate samples. 100 randomly chosen colonies were patched for each sample. - denotes no colonies patched. (*) denotes where <100 colonies were patched.

pGE5 into RmP4258 pGE3 (RmP4310)

Media: LBmc FeCo Tc (5µg/mL) Sp (100µg/mL)

Sample	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5
1A	19	29	48	59	64	69
1B	13	28	52	57	63	73
1C	12	25	54	51	68	84
Avg	14.6	27.3	51.3	55.6	65	75.3

pGE5 into RmP4258 pGE4 (RmP4311)

Media: LBmc FeCo Nm (200µg/mL) Sp (100µg/mL)

Sample	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5
2A	14	26	43	59	68*	80*
2B	9	12	42	47	63	79
2C	11	23	41	51	65	82
Avg	11.3	20.3	42	52.5	65.3	80.4

pGE6 into RmP4258 pGE4 (RmP4312)

Media: LBmc FeCo Tc (5µg/mL) Nm (200µg/mL)

Sample	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5
3A	5	8	20	18*	26	27*
3B	5	10	-	13	27*	24
3C	5	8	-	19	12	24
Avg	5	8.6	20	16.5	21.7	24.6

pGE7 into RmP4258 pGE3 (RmP4313)

Media: LBmc FeCo Nm (200µg/mL) Tc (5µg/mL)

Sample	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5
4A	5	9	24	22	24	28
5B	3	10	20	16	28	28
6C	5	11	-	-	-	-
Avg	4.3	10	22	19	26	28

Table 12: Percentage of re-patched colonies unable to grow on media with single antibiotic for each sample tested for triplicate samples. (-) denotes no colonies patched.

pGE5 into RmP4258 pGE3 (RmP4310)

Media: LBmc FeCo Tc (5µg/mL) or LBmc FeCo Sp (100µg/mL)

Sample	Day 0		Day 1		Day 2		Day 3		Day 4		Day 5	
	Tc	Sp	Tc	Sp	Tc	Sp	Tc	Sp	Tc	Sp	Tc	Sp
1A	5.3	89.5	27.6	72.4	20.0	86.0	18.6	84.7	25.0	76.6	24.6	84.1
1B	15.4	84.6	21.4	78.6	22.9	89.6	22.8	80.7	19.0	82.5	20.5	83.6
1C	8.3	91.7	20.0	80.0	9.4	86.8	19.6	84.3	17.6	79.4	22.6	83.3
Avg	9.7	88.6	23.0	77.0	17.5	87.5	20.4	83.3	20.6	79.5	22.6	83.7

pGE5 into RmP4258 pGE4 (RmP4311)

Media: LBmc FeCo Nm (200µg/mL) or LBmc FeCo Sp (100µg/mL)

Sample	Day 0		Day 1		Day 2		Day 3		Day 4		Day 5	
	Nm	Sp	Nm	Sp	Nm	Sp	Nm	Sp	Nm	Sp	Nm	Sp
2A	35.7	85.7	11.5	76.9	15.0	90.0	22.4	81.6	20.6	76.5	22.6	75.5
2B	44.4	77.8	16.7	83.3	22.2	80.6	21.3	83.0	14.3	87.3	19.0	78.5
2C	27.3	81.8	13.6	90.9	26.8	87.8	15.7	86.3	14.7	83.1	15.9	82.9
Avg	35.8	81.8	13.9	83.7	21.4	86.1	19.8	83.6	16.5	82.3	19.2	79.0

pGE6 into RmP4258 pGE4 (RmP4312)

Media: LBmc FeCo Tc (5µg/mL) or LBmc FeCo Nm (200µg/mL)

Sample	Day 0		Day 1		Day 2		Day 3		Day 4		Day 5	
	Tc	Nm	Tc	Nm	Tc	Nm	Tc	Nm	Tc	Nm	Tc	Nm
3A	100.0	100.0	100.0	100.0	90.0	85.0	100.0	100.0	96.2	92.3	88.9	94.4
3B	100.0	100.0	100.0	90.0	-	-	92.3	76.9	91.3	100.0	100.0	90.9
3C	100.0	100.0	75.0	75.0	-	-	100.0	94.7	100.0	83.3	100.0	90.0
Avg	100.0	100.0	91.7	88.3	90.0	85.0	97.4	90.6	95.8	91.9	96.3	91.8

pGE7 into RmP4258 pGE3 (RmP4313)
Media: LBmc FeCo Tc (5µg/mL) or LBmc FeCo Nm (200µg/mL)

Sample	Day 0		Day 1		Day 2		Day 3		Day 4		Day 5	
	Nm	Tc	Nm	Tc	Nm	Tc	Nm	Tc	Nm	Tc	Nm	Tc
4A	100.0	60.0	88.9	88.9	78.3	87.0	95.2	90.5	91.7	87.5	96.4	100.0
4B	100.0	100.0	70.0	100.0	80.0	80.0	87.5	100.0	96.4	92.9	92.9	89.3
4C	100.0	20.0	81.8	90.9	-	-	-	-	-	-	-	-
Avg	100.0	60.0	75.9	95.5	80.0	80.0	87.5	100.0	96.4	92.9	92.9	89.3

```
#!/usr/bin/perl
use 5.010;

#temporary file list.
$input = 'input.txt';
$tempfile = 'temp_file.txt';
$tempblank = 'temp_blank.txt';
$gsample = 'gsample.txt';
$generation1 = 'generation_time1.txt';
$generation2 = 'generation_time2.txt';
$generation3 = 'generation_time3.txt';
$generation4 = 'generation_time4.txt';
$generation_times = 'generation_time_all.txt';
$generation_times2 = 'generation_times_all2.txt';

#transfer input file to a temporary file.
while(<>) { #while reading the input file, store it as a variable.
                $input1 = $input1 . $_;
}
open $in, '>', $input; #save the variable as a temporary file
print $in ( "$input1" );
close $in;

#replace time in seconds with time in minutes.
&time( $tempfile, $input );

#subtract the blank from each sample.
&blank( $tempfile, $tempblank );

#calculate and print averages and standard deviations of each set of triplicates, with the approriate headers.
&headers;
&average_and_stdev( $tempblank );

#print headers for the generation times
print("\n\n\nGeneration Times,\n");
&headers;

#calculate and print average generation times and standard deviations for each set of triplicates, and state density range that the generation time is
calculated for.
&generation( $tempblank, '0.100', '0.500' ); #calculate generation time between two numbers listed.
print('0.100-0.500'); #print lower and upper boundary range for generation time calculation.
&average_and_stdev( $generation_times2 ); #Calculate average generation times and standard deviation.
#to calculate the generation time between two (or more) density ranges, repeat the last three lines of code as many times as desired, adjusting the
upper and lower range as required.


#remove temporary files
&clear;

#Subroutine to replace time in seconds with time in minutes, and change all sample values to having 2 digits before the decimal and 3 after the
decimal.
sub time {
                my ($tempfile, $input) = @_; #set variables for the subroutine.
                open $input1, $input;#open file containing the input data.
                $time = '0.00'; #Variable that will be used to set time in minutes for each line of data.
                #a loop to store the entire input file in the $file variable, with each line preceeded by a number, with the number increasing by
0.25 each line.
                #The number represents the time in minutes
                while(<$input1>)  { #while reading the input file,
                                if( /[0-9]s/ ) { #if a 's' is found (all lines containing OD600 readings will have a number followed by 's',
while none of the others do),
                                                $csv = $csv . "$time $_"; #add the line containing the 's' to a variable that will hold
all the lines containing a 's'
                                                $time = $time + 0.25; #increase the time variable by 0.25.
                                                $time = sprintf '%.2f', $time; #fix the time variable to 2 decimal places.
                                }
                }
                close $input1; #close temp file holding input data.
                #remove all the time in seconds values from the variable.
                $csv =~ s/ \d+s//g;
                #fix all sample values and time points to have 2 digits before the decimal by adding a 0 at the start of values below 10.
                $csv =~ s/\b0\b/00/g;
                $csv =~ s/\b1\b/01/g;
                $csv =~ s/\b2\b/02/g;
                $csv =~ s/\b3\b/03/g;
                $csv =~ s/\b4\b/04/g;
                $csv =~ s/\b5\b/05/g;
                $csv =~ s/\b6\b/06/g;
                $csv =~ s/\b7\b/07/g;
                $csv =~ s/\b8\b/08/g;
                $csv =~ s/\b9\b/09/g;
                #fix all time points to having 2 digits before the decimal by adding a 0 at the start of time points below 10.
                #save the modified input file as a temporary file
                open $temp ,'>', $tempfile; #open the temporary file
                print $temp ( "$csv \n" ); #print the input file with the time in seconds replaced with time in minutes, any unnecessary lines
removed, and all sample values and time points having 2 digits prior to the decimal.
                close $temp; #close the temporary file
}
```

```
#Subroutine to substract the blank value from each of the samples.
sub blank {
		my ($tempfile, $tempblank) = @_; #Set subroutine variables.
		#Loop to blank each of the samples
		open $blank, $tempfile; #open temp file containing the modified input file
		open $tblank, '>', $tempblank; #open temp file which will hold the blanked data
		while ( <$blank> ) { #while reading the modified input file,
				print $tblank substr( $_, 0, 6); #print the time point at the start of the row
				$blank1 = substr( $_, -22, 6); #variable holding first replicate of the blank sample
				$blank2 = substr( $_, -15, 6); #variable holding second replicate of the blank sample
				$blank3 = substr( $_, -8, 6); #variable holding third replicate of the blank sample
				$blankavg = ( $blank1 + $blank2 + $blank3 ) / 3; #variable holding the average of the blank samples
				for( $sample = 6; $sample <= 650 ; $sample = $sample + 7) { #for each data point at this time point,
						$datum = substr( $_, $sample, 6); #variable holding the data point to be blanked
						$blanked =  $datum - $blankavg; #variable holding the data point with the blank
subracted from it
						if( $blanked > 0 ) { #if the blanked data point is greater than 0,
								if( $blanked < 10 ) { #if the blanked data point is less than 10
										print $tblank ( "0" ); #add a 0 before
the data point to once again have 2 digits before the decimal
								}
								printf $tblank '%.3f', $blanked; #print blanked data point to
the temp file with 3 decimal places
								print $tblank ( "," ); #print a comma after the data point to
indicate the end of the cell in Excell
						}
						elsif( $blanked <= 0) { #if the blanked data point is not greater than 0
								print $tblank ('00.000,'); #print 00.000 followed by a comma
						}
				} #repeat for each of the 93 data points at each time point
				print $tblank ("\n"); #go to new line for the next time point
		} #repeat for each of the time points in the growth curve
		close $blank; #close file containing modified input
		close $tblank; #close file containing blanked data
}

#Subroutine to print the headers for the file.
sub headers {
		$header = "Averages,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,Standard  Deviations,\n,A1-3,A4-6,A7-9,A10-12,B1-3,B4-6,B7-9,B10-12,C1-3,C4-6,C7-
9,C10-12,D1-3,D4-6,D7-9,D10-12,E1-3,E4-6,E7-9,E10-12,F1-3,F4-6,F7-9,F10-12,G1-3,G4-6,G7-9,G10-12,H1-3,H4-6,H7-9,,,,A1-3,A4-6,A7-9,A10-12,B1-
3,B4-6,B7-9,B10-12,C1-3,C4-6,C7-9,C10-12,D1-3,D4-6,D7-9,D10-12,E1-3,E4-6,E7-9,E10-12,F1-3,F4-6,F7-9,F10-12,G1-3,G4-6,G7-9,G10-12,H1-3,H4-6,H7-
9,\n"; #print a line at top of Excell with Averages in one cell, and Standard Deviations several cells later, followed by a row indicating cells that were
averaged
		print $header;
}

#Subroutine to calculate averages and standard deviations of each set of tiplicate
sub average_and_stdev {
		my ($input) = @_; #set variable for subroutine
		open $values, $input; #open file holding blanked data points
		while( <$values> ) { #while reading the file contianing the blanked data
				print substr( $_, 0, 6 );
				#Loop for averages
				for( $sample2 = 6; $sample2 <=636; $sample2 = $sample2 + 21) { #for each sample, do
						$areplicate1 = substr( $_, $sample2, 7); #variable holding first replicate
						$areplicate2 = substr( $_, $sample2 + 7, 7); #variable holding second replicate
						$areplicate3 = substr( $_, $sample2 + 14, 7); #variable holding third replicate
						$avgreplicate =  ($areplicate1 + $areplicate2 + $areplicate3) / 3; #calculate the
average of the blanked samples for the triplicates
						printf '%.3f', "$avgreplicate" ; #print average of the triplicate to three decimal
places
						print (","); #print a comma after the average of the triplicate to indicate end of cell
				} #repeat for each of the 31 sets of triplicates
				print ( ",,,"); #separate the averages and standard deviations tables by 5 columns in Excell
				#Loop for standard deviations
				#The standard deviations and the averages for a specific time point are placed on the same line, but as
part of separate tables
				for( $sample3 = 6; $sample3 <=636; $sample3 = $sample3 + 21) { #for each sample, do
						$sreplicate1 = substr( $_, $sample3, 7); #variable holding first replicate
						$sreplicate2 = substr( $_, $sample3 + 7, 7); #variable holding second replicate
						$sreplicate3 = substr( $_, $sample3 + 14, 7); #variable holding third replicate
						$avgreplicate2 = ( $sreplicate1 + $sreplicate2 + $sreplicate3 ) / 3; #variable holding
the average of the triplicates
						$stdev = sqrt( ( (($sreplicate1 - $avgreplicate2)**2) + (($sreplicate2 -
$avgreplicate2)**2) + (($sreplicate3 - $avgreplicate2)**2) ) / 3 ); #standard deviation
						printf '%.3f', "$stdev"; #print standard deviation of the triplicate to three decimal
places
						print (","); #print a comma after the standard deviation
				} #repeat for each of the 31 sets of triplicates
				print ("\n"); #go to new line for the next time point
		} #repeat for each of the time points in the growth curve
		close $values; #close temporary file holding the blanked data
}
```

```
#subroutine to calculate generation time for each sample.
sub generation {
	my( $tempblank, $lower, $upper ) = @_; #set variables for the subroutine.  $lower is the lower boundary of density for generation time to be calculate for, while $upper is the upper boundary
	#Setting variables for range of where to look for reading closest to each border.  Looks +/- 40% of the value of the border; thus as long as doubling time is more than11.9 minutes, a value for each border should be found.
	$lowerl = 0.6 * $lower; #lower boundary of lower border
	$lowerl = sprintf '%.3f', ( "$lowerl" );
	$lowerl = "0$lowerl";
	$lowerh = 1.4 * $lower; #upper boundary of lower border
	$lowerh = sprintf '%.3f', ( "$lowerh" );
	$lowerh = "0$lowerh";
	$upperl = 0.6 * $upper; #lower boundary of higher border
	$upperl = sprintf '%.3f', ( "$upperl" );
	$upperl = "0$upperl";
	$upperh = 1.4 * $upper; #upper boundary of higher border
	$upperh = sprintf '%.3f', ( "$upperh" );
	$upperh = "0$upperh";
	open $ti, '>', $generation_times; #open temporary file where generation times are to be held
	print $ti ("    ,"); #print 5 spaces (to keep spacing constant with the $tempblank spacing), followed by a comma so everything will print starting in cell 2 in Excell
	close $ti; #close temp file
	for( $gensample = 1 ; $gensample <= 93 ; $gensample = $gensample + 1 ) {#for each of the sample,
		#clear all temporary files which hold values for calculation of generation time.
		open $clear1, '>', $generation1;
		open $clear2, '>', $generation2;
		open $clear3, '>', $generation3;
		open $clear4, '>', $generation4;
		print $clear1 ("");
		print $clear2 ("");
		print $clear3 ("");
		print $clear4 ("");
		close $clear1;
		close $clear2;
		close $clear3;
		close $clear4;
		#clear all variables which hold values involved in calculation of generation time.
		$g1 = '';
		$g2 = '';
		$g3 = '';
		$g4 = '';
		open $extract, $tempblank; #open temp file holding the blanked data
		open $value1, '>', $gsample; #open temp file which will hold one sample (and the time point) for all time points
		while ( <$extract> ) { #while reading temp file holding blanked data
			print $value1 substr( $_, 0, 6); #print the time point
			say $value1 substr( $_, 6 + $gdata, 7); #print the values for one sample
		}
		close $extract; #close temp file holding blanked data
		close $value1; #close temp file holding the data for one sample
		for( $lowerlv = $lowerl; $lowerlv < $lower; $lowerlv = $lowerlv + 0.001 ) { #for numbers between the lower boundary and the lower border,
			$lowerlv = sprintf '%.3f', ( "$lowerlv" ); #print the value to 3 decimal places
			open $border1, $gsample; #open the temp file holding the one sample
			while( <$border1> ) { #while reading the temp file
				if( /$lowerlv/ ) { #if the line contains the value,
					open $gen1, '>', $generation1; #open temp file
					print $gen1 ( $_ ); #print the line containing the value (holds both the time point and the OD600 reading)
					close $gen1; #close the temp file
				}
			}
			close $border1; #close the temp file holding the entire sample
		} #repeat for each value between the lower boundary and lower border, starting at the lowest number, and the temp file holds only the value closest to the lower border (not the lower boundary).
		for( $lowerhv = $lowerh; $lowerhv > $lower; $lowerhv = $lowerhv - 0.001 ) { #same as above, except start at the higher boundary of the lower border.
			$lowerhv = sprintf '%.3f', ( "$lowerhv" );
			open $border2, $gsample;
			while( <$border2> ) {
				if( /$lowerhv/ ) {
					open $gen2, '>', $generation2;
					print $gen2 ( $_ );
					close $gen2;
				}
			}
			close $border2;
		}
		open $border12, $gsample; #open the temp file holding the sample values
		while( <$border12> ) { #while reading the temp file
			if (/$lower/) { #if a time point has a reading identical to the lower boundary
				open $gen1, '>', $generation1; #open both temp files holding the values just above and below the lower boundary
				open $gen2, '>', $generation2;
				print $gen1 ( $_ ); #replace the values in the temp files with the line containing the OD600 reading that exactly matches lower boundary
				print $gen2 ( $_ );
				close $gen1; #close both temporary files.
				close $gen2;
			}
```

```
} #between this loop, and the previous two 'for' loops, two temp files were created.  One contains the
OD600 reading and time point closest to, but below, the lower boundary, and the other contains the OD600 reading and time point closest to, but
above, the lower boundary; unless there is a time point where the OD600 reading is equal to the lower boundary, in which case both temp files hold
this OD600 reading and the corresponding time point.
close $border12; #close the temp file holding the sample values
open $gen1v, $generation1; #open file containing the value closest to, but below, the lower boundary
while( <$gen1v> ) { #while reading file
        $g1 = substr( $_, 6, 6); #set a variable equal to the OD600 reading
        $g1t = substr( $_, 0, 5); #set a variable equal to the time point
}
close $gen1v; #close the temp file
open $gen2v, $generation2; #same as above, but for the value closest to, but above, the lower boundary
while( <$gen2v> ) {
        $g2 = substr( $_, 6, 6);
        $g2t = substr( $_, 0, 5)
}
close $gen2v;
if( ($g2 - $lower) - ($lower - $g1) < 0) { #if the value above the lower boundary is closer to the lower
boundary than the value below it,
        $ds = $g2; #set the initial density equal to the value above the boundary
        $ts = $g2t; #set the initial time equal to the value above the boundary
}
elsif( ($g2 - $lower) - ($lower - $g1) > 0) { #if the value below the lower boundary is closer to the lower
boundary than the value above it,
        $ds = $g1; #set the initial density equal to the value below the boundary
        $ts = $g1t; #set the initial time equal to the value below the boundary
}
elsif( ($g2 - $lower) - ($lower - $g1) == 0) { #if both values are equally close to lower boundary.
        $ds = $g2; #set the initial density equal to the value above the boundary
        $ts = $g2t; #set the inital time equal to the value above the boundary
}
if( $g2 eq '' ) { #if there is no value between the lower boundary and the higher border,
        $ds = $g1; #set the initial density equal to the value below the boundary
        $ts = $g1t; #set the initial time equal to the vlaue below the boundary
}
if( $g1 eq '' ) { #if there is no value between the lower boundary and the lower border,
        $ds = $g2; #set the initial density equal to the value above the boundary
        $ts = $g2t; #set the initial time equal to the value above the boundary
}

for( $upperlv = $upperl; $upperlv < $upper; $upperlv = $upperlv + 0.001 ) { #loop same as above, but for
between upper boundary and lower border
        $upperlv = sprintf '%.3f', ( "$upperlv");
        open $border3, $gsample;
        while( <$border3> ) {
                if( /$upperlv/ ) {
                        open $gen3, '>', $generation3;
                        print $gen3 ( $_ );
                        close $gen3;
                }
        }
        close $border3;
}
for( $upperhv = $upperh; $upperhv > $upper; $upperhv = $upperhv - 0.001 ) { #loop same as above, but
for between upper boundary and higher border
        $upperhv = sprintf '%.3f', ( "$upperhv" );
        open $border4, $gsample;
        while( <$border4> ) {
                if( /$upperhv/ ) {
                        open $gen4, '>', $generation4;
                        print $gen4 ( $_ );
                        close $gen4;
                }
        }
        close $border4;
}
open $border34, $gsample;
while( <$border34> ) { #loop same as above, but for the upper boundary
        if (/$upper/) {
                open $gen3, '>', $generation3;
                open $gen4, '>', $generation4;
                print $gen3 ( $_ );
                print $gen4 ( $_ );
                close $gen3;
                close $gen4;
        }
}
close $border34;
open $gen3v, $generation3; #from here to line 344, script is the same as for lines 243-274, except for the
upper boundary
while( <$gen3v> ) {
        $g3 = substr( $_, 6, 6);
        $g3t = substr( $_, 0, 5);
}
close $gen3v;
open $genv4, $generation4;
while( <$genv4> ) {
        $g4 = substr( $_, 6, 6);
        $g4t = substr( $_, 0, 5);
}
```

```
                                        close $genv4;
                                        if( ($g4 - $upper) - ($upper - $g3) > 0) {
                                                $df = $g3;
                                                $tf = $g3t;
                                        }
                                        elsif( ($g4 - $upper) - ($upper - $g3) < 0) {
                                                $df = $g4;
                                                $tf = $g4t;
                                        }
                                        elsif( ($g4 - $upper) - ($upper - $g3) == 0) { #if both values are equally far from the boundary, pick the
lower value
                                                $df = $g3;
                                                $tf = $g3t;
                                        }
                                        if( $g4 eq '' ) {
                                                $df = $g3;
                                                $tf = $g3t;
                                        }
                                        if( $g3 eq '' ) {
                                                $df = $g4;
                                                $tf = $g4t;
                                        } #end of part of script that is similar to before
                                        open $times, '>>', $generation_times; #open, to append, a file to hold all the generation times
                                        if ( $ds ne '' and  $ds > $lowerl and $ds < $lowerh and $df > $upperl and $df < $upperh) { #if the initial
density reading is not empty, the upper and lower density readings fall withing the appropriate range,
                                                $final = int( 100 * (( $tf - $ts ) / ( log( $df / $ds ) / log( 2 ) ) ) / 100; #a variable that
calculates and holds the generation time for the sample
                                                if( $final < 10 ) { #if the generation time is less than 10 hours
                                                        print $times ("0"); #print a 0
                                                        printf $times '%.3f', ( "$final" ); #then print the generation
time (so there are 2 digits before the decimal)
                                                        print $times (","); #print a comma to indicate end of the cell
in Excell
                                                }
                                                else { #else if the generation time is >= 10,
                                                        printf $times '%.3f', ("$final"); #print the generation time
(there are 2 digits before the decimal
                                                        print $times( "," ); #print a comma
                                                }
                                        }
                                        else { #else if at least one of the conditions in the if statement 12 lines ago (just prior to the calculation of
generation time), meaning that the culcure did not grow,
                                                print $times ( '00.000,' ); #print 00.000
                                        }
                                        close $times; #close the temp file holding the generation time
                                        $gdata = $gdata + 7; #increase this variable by 7, so that in the next time through the loop, the generation
time is calculated for the next sample.
                        }
                        $gdata = 0; #reset this variable to 0, so that if generation times are calculated between multiple OD600 ranges, this subroutine
will always start with the first sample
                        open $fixit, '>>', $generation_times; #open to append the file holding the generation time.
                        print $fixit ("\n"); #print a carriage return
                        close $fixit; #close the file
                        open $replace, $generation_times; #open the file holding the generation times to read
                        open $gen_times, '>', $generation_times2; #open a second temporary file to transfer the generation times to
                        while( <$replace> ) { #while reading the file holding the generation times
                                print $gen_times ( $_ ); #print the file to the second temporary file
                        }
                        close $replace; #close both temporary files
                        close $gen_times;
                        open $empty, '>', $generation_times; #open the initial temp file holding the generation times
                        print $empty (""); #clear it so that it is empty if generation times are to be calculated for a new OD600 range
                        close $empty; #close the file
}

#subroutine to remove the temporary files
sub clear {
                        unlink $input;
                        unlink $tempfile;
                        unlink $tempblank;
                        unlink $gsample;
                        unlink $generation1;
                        unlink $generation2;
                        unlink $generation3;
                        unlink $generation4;
                        unlink $generation_times;
                        unlink $generation_times2;
}
```

Figure 25: Perl script used to calculate the generation time and standard deviations of samples used in growth curve analysis. Generation time was calculated using values from exponential phase between OD_{600} 0.1-0.5. Perl Script provided by G.C. diCenzo, 2018.