

DEVELOPMENT OF AN *IN VIVO* DNA CLONING PROCEDURE

**DEVELOPMENT OF AN *IN VIVO* DNA CLONING PROCEDURE
FOR BACTERIA**

By

PATRICK S. G. CHAIN

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

©Copyright by Patrick S. G. Chain, December 1998

MASTER OF SCIENCE (1998)
(Biology)

McMaster University
Hamilton, Ontario

TITLE: Development of an *in vivo* DNA Cloning Procedure for Bacteria

AUTHOR: Patrick S. G. Chain, B.Sc. (McMaster University)

SUPERVISOR: Professor T. M. Finan

NUMBER OF PAGES: xi, 129

ABSTRACT

In this thesis, we describe the development of a method to delete and to clone specific large regions from the 1700 kilobase pExo megaplasmid of *Rhizobium meliloti*. In principal, the region to be cloned is flanked by *FRT* sites, which direct site specific recombination by the Flp recombinase. Targeting constructs were designed to include part of the IS50 from Tn5, the *FRT* site, an origin of transfer (*oriT*), and a replication origin from RK2 or the F plasmid. These constructs were directed to known Tn5-derivative insertion sites in the pExo megaplasmid. A plasmid which expresses the Flp recombinase constitutively in *R. meliloti* was made, and the transfer of this plasmid to *FRT*-flanked megaplasmid regions was shown to result in the deletion of the intervening DNA.

We demonstrated that the pExo megaplasmid DNA regions could be captured in *Escherichia coli*, however in this case, the megaplasmid excision event appears to be directed by the *oriT* sites rather than the *FRT* sites. We present strong evidence that a specific 50 kb region contains the *oriV* of the pExo megaplasmid; *R. meliloti* strains deleted for this region could not be isolated, and this region was found to replicate autonomously in *Agrobacterium tumefaciens*. Preliminary sequence analysis has revealed strong homology within this region to genes encoding the RepABC replication proteins of several *Rhizobium* and *Agrobacterium* plasmids.

ACKNOWLEDGMENTS

I would like to thank Dr. Turlough Finan for giving me the opportunity to work in his laboratory for the past few years. I have learned much and have enjoyed myself (well, for the most part). I would like to express my gratitude to the members of the entire Biology Department, especially Marta Boszko and Macarena Busto (the Ukrainian-Uruguayan Super-Combo), Rob Kulathinal (kindred soccer spirit) and the eternally engaged Colin Ross, for their irreplaceable friendship. Special acknowledgments to the members of the Finan lab for their immense help, without which, I would never have survived: Shelley O'Brien, for her love and patience; Mike Mitsch, who was always there to make me laugh (and for his love too); Alison Cowie, for her keen worldly insight and advice about life (and for her silent admiration); Ismael Hernandez-Lucas, who never stops asking questions (the past midnight calls about the location of restriction enzymes), but is a good guy nonetheless, and our former fourth-year student Scott Clark, for his ability to turn every event into a card game.

I would also like to thank my friends outside of the Department for their help in “relaxing” over the weekends. Lastly, I would like to thank my parents for their continuous support and especially my brothers: Benjamin and Jonathan for their love; and Philippe and Frederic for their companionship, talking with you serves as a constant reminder of the hardships others have to endure and overcome, this motivated me to do the best that I can.

TABLE OF CONTENTS

ABSTRACT	iii
ACKNOWLEDGMENTS	iv
TABLE OF CONTENTS	v
LIST OF FIGURES	viii
LIST OF TABLES	x
LIST OF ABBREVIATIONS	xi
CHAPTER 1. INTRODUCTION	1
Rhizobiaceae	1
Symbiotic N ₂ -Fixation.....	1
The <i>Rhizobium meliloti</i> Genome Structure	2
Analysis of the <i>R. meliloti</i> Megaplasmsids.....	3
The pExo Transposon Map	4
The Next Step: Sequencing the pExo Megaplasmid.....	5
Common Whole-Genome Sequencing Methodologies.....	9
Specific Targeted Sequencing Methodologies	11
The Yeast 2 μ m Plasmid Flp/ <i>FRT</i> System of Site-Specific Recombination.....	13
This Work: A Novel Use For the Flp/ <i>FRT</i> Recombination System in <i>R. meliloti</i>	15
CHAPTER 2. MATERIALS AND METHODS	18
Bacterial Strains, Plasmids and Transposons.....	18
Growth Media and Conditions	27
Bacterial Matings	28
Preparation and Testing of Φ M12 Transducing Lysates.....	29
Φ M12-Mediated Transduction.....	29

Transposon Replacements	30
DNA Methodology	30
Preparation of Competent Cells	31
Transformation of Competent Cells	31
Preparation of Electro-Competent Cells	32
Electroporation of Electro-Competent Cells	32
Pulsed-Field Gel Electrophoresis	33
Isolation of Total Genomic DNA from <i>R. meliloti</i>	35
Southern Transfer and Hybridization	35
Polymerase Chain Reaction	36
DNA Sequencing	38
CHAPTER 3. RESULTS	40
3.1: FRT Targeting Vectors	40
Cloning Strategy for the Targeting Plasmids	40
Description of Construction	41
Directing the targeting vectors to specific regions of the pExo Megaplasmid	46
Obtaining Targeting Vectors in Both IS50 Elements	46
Construction of <i>R. meliloti</i> Strains Carrying Two Targeting Vectors	47
3.2: Flp Delivery Vectors	51
Construction of a Flp-testing strain of <i>R. meliloti</i>	51
Construction of the Initial Flp Vector	53
Expression of <i>flp</i> from a <i>R. meliloti</i> inducible promoter	55
3.3: Analysis of Tn5::pTH455 and Tn5-132::pTH456 Combinations	59
Pulsed-Field Gel Electrophoresis—A Method For Isolating the Intervening Regions	59
Flp-mediated deletions in <i>R. meliloti</i>	61
Transfer of pExo <i>R. meliloti</i> DNA into <i>A. tumefaciens</i>	67
Transfer of pExo <i>R. meliloti</i> DNA into <i>E. coli</i>	70
Excision Prior to Transfer of the 50 kb Region Is Not Required	71
Analysis of pTH482, the 60 kb plasmid	73
Excision of Larger pExo Regions	77
3.4: Using BAC Backbones For Recovery of Large pExo Regions	80
Construction of Bac Vector pTH504	80
Deletion and Transfer of pExo Regions Flanked by pTH455 and pTH504	81
Analysis of Regions Retrieved With pTH509 as BAC Vector	86

CHAPTER 4. DISCUSSION	92
Genome Sequencing.....	92
The Flp/ <i>FRT</i> Strategy.....	93
The Flp Delivery Plasmids.....	94
Uses For the Flp/ <i>FRT</i> System in <i>R. meliloti</i>	96
Flp-Independent Transfer of <i>FRT</i> -Flanked pExo DNA	97
Other Features of the Targeting Vectors	99
The Replicator Region of the pExo Megaplasmid	101
Further Sequencing.....	103
Comparative Genomics	103
 APPENDIX A	 105
 APPENDIX B	 108
 REFERENCES	 112

LIST OF FIGURES

Figure 1:	Circular Map of the pExo Megaplasmid	6
Figure 2:	The Flp Recombination Target (<i>FRT</i>) Site.....	14
Figure 3:	Cloning Scheme to Construct the Targeting Vectors (Part 1).....	43
Figure 4:	Cloning Scheme to Construct the Targeting Vectors (Part 2).....	45
Figure 5:	Directing the Targeting Vectors to the IS50 Elements and Determination of Both Possible Recombinations by PCR Analysis.....	48
Figure 6:	Construction of <i>R. meliloti</i> Strains Carrying Two Targeting Vectors.....	50
Figure 7:	Construction of a Test System for Flp in <i>R. meliloti</i>	52
Figure 8:	Cloning Scheme to Construct Flp Delivery Vectors	57
Figure 9:	Pulsed-Field Gel of RmH994 and Wild Type Rm1021 Genomic DNA Digested With <i>I-SceI</i> and <i>SpeI</i>	60
Figure 10:	The Four Possible Combinations of pTH455-targeted Ω 5056::Tn5 with pTH456-targeted Ω 5069::Tn5-132	62
Figure 11:	Diagram of Flp-Mediated Excision of the Two Targeted Transposons	64
Figure 12:	Linear and Circular Maps of pExo Demonstrating the Transposons Targeted in This Study	65
Figure 13:	<i>EcoRI</i> and <i>BamHI</i> Restriction Digests of a Captured Large Plasmid From RmH994.....	69

Figure 14: Diagram of the 60 kb plasmid pTH482 and the Region From Which it Originated.....	72
Figure 15: Southern Blot Indicating that pTH482 Carries <i>R. meliloti</i> DNA	75
Figure 16: Agarose Gel Determination of the Difference in Size of the Plasmids pTH482 and pTH515.....	83
Figure 17: Diagram Demonstrating the Transfer and Subsequent Recombination of the Cassette (From pTH504) Carrying the <i>FRT</i> , <i>oriT</i> , <i>oriV</i> and <i>IS50</i> elements	85
Figure 18: Southern Blot Determination of Both Possible Recombinations Events With the BAC Targeting Vector pTH509	87
Figure 19: Diagram of Loss of Antibiotic Resistance Markers From <i>R. meliloti</i> Strains Carrying a Tn5::pTH455 and a Tn5-132::pTH509	89
Figure 20: Southern Blot Demonstrating that pTH519 Carries <i>R. meliloti</i> DNA.....	91
Figure 21: Multiple Alignment of the Nucleotide Sequence of the Putative pExo Origin of Replication With the <i>oriV</i> sites of several <i>Rhizobium</i> and <i>Agrobacterium</i> Plasmids.....	102
Figure 22: General Representation of the Four Possible Combinations of pTH455-Targeted Tn5 With pTH456-Targeted Tn5-132	107
Figure 23: Nucleotide Sequence of the <i>pcaD/pcaQ</i> Intergenic Region.....	109
Figure 24: Alignment of <i>R. meliloti</i> and <i>A. tumefaciens</i> PcaQ, and the <i>pcaD/pcaQ</i> Intergenic Region.....	110
Figure 25: Alignment of <i>R. meliloti</i> and <i>B. japonicum</i> PcaD, and <i>pcaD</i> promoter region	111

LIST OF TABLES

Table 1: Published Complete Genomes as of December, 1998.....	8
Table 2: Bacterial Strains, Plasmids and Transposons	19
Table 3: Antibiotic Concentrations.....	27
Table 4: PCR and Sequencing Primers and Single Stranded Oligonucleotides	37
Table 5: Level of <i>flp</i> expression, measured in % white colonies	55
Table 6: Highest Matches Obtained Through BLASTX Searches With the <i>EcoRI</i> Fragments of pTH482	76
Table 7: “Sets” of <i>R. meliloti</i> strains carrying two targeting vectors	106

LIST OF ABBREVIATIONS

Ap	ampicillin
BAC	bacterial artificial chromosome
bp	base pair
Cm	chloramphenicol
cTc	chlortetracycline
Dig	digoxygenin
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
<i>FRT</i>	Flp recognition target
Gm	gentamycin
kb	kilobase or kilobase pair
Km	kanamycin
LB	Luria-Bertani
M9	minimal salts media
Nm	neomycin
<i>oriT</i>	origin of conjugal transfer
<i>oriV</i>	origin of vegetative replication
Ot	oxytetracycline
PCR	polymerase chain reaction
PFGE	pulsed-field gel electrophoresis
PFU	plaque forming units
Rf	rifampicin
RNA	ribonucleic acid
rpm	revolutions per minute
Sm	streptomycin
Sp	spectinomycin
Tc	tetracycline
WT	wild type
X-Gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside
$^{\circ}$ C	degree Celcius
Ω Sp	spectinomycin resistant interposon
Φ	phage or lysate
r	resistant or resistance

CHAPTER 1. INTRODUCTION

Rhizobiaceae

Members of the Gram negative family Rhizobiaceae are classified within the alpha subdivision of Proteobacteria (Holt *et al.* 1994). Amongst others, this family includes the genus *Agrobacterium*, responsible for induction of crown gall and hairy root neoplastic diseases, and several genera of nitrogen (N₂)-fixing, root nodule-forming, soil bacteria. The continued study of members of the family Rhizobiaceae furthers general understanding of plant-microbe interactions.

Rhizobium (or *Sinorhizobium*) *meliloti* is one of the fast-growing members of the genus, *Sinorhizobium*. A complex series of interactions between Rhizobia and host leguminous plants results in the formation of N₂-fixing nodules. Rhizobia are, in general, plant species specific, and *R. meliloti* is no exception. *R. meliloti* restricts its nodule formation to some plants of the Medicago (e.g. alfalfa), Melilotus (e.g. sweet clover) and Trigonella genera.

Symbiotic N₂-Fixation

Gaseous nitrogen is an abundant component of the earth's atmosphere (~80%). However, its inert nature makes it unavailable to higher organisms. Although eukaryotic organisms require the reduced form of nitrogen, ammonium ion (NH₄⁺), only some

prokaryotes manifest the intrinsic ability to reduce atmospheric nitrogen to ammonia ($\text{N}_2 \rightarrow \text{NH}_4^+$)—a process known as N_2 -fixation. This key reaction of the biogeochemical nitrogen cycle occurs, in large part, within bacteroids (the terminally differentiated N_2 -fixing bacteria) in root nodules.

This rhizobia-legume symbiosis is beneficial to both microorganism and plant host. To maintain the energy intense-process of N_2 -fixation, a continuous supply of carbohydrate to the nodules is required. C_4 -dicarboxylic acids, such as malate and succinate, appear to be the major carbon source supplied to, and utilized by, the bacteroids (Finan *et al.* 1983; Ronson *et al.* 1981). The advantage to the host plant is of agronomical importance, as the ability to grow under nitrogen deficient conditions reduces the requirement for exogenous nitrogen fertilizer. The commercial and agricultural importance of this symbiotic interaction cannot be overlooked; independence from nitrogenous fertilizers minimizes pollution of lakes and rivers, reduces farming cost, and expands crop production.

The *Rhizobium meliloti* Genome Structure

Much of the genetic material in *Rhizobium*, including most of the symbiotic information, is carried on large plasmids. Large extrachromosomal replicons referred to as megaplasmids (Rosenberg *et al.* 1982), ranging in size from 400 kb to over 1500 kb, are found in a number of bacteria including *Pseudomonas* spp. (Hansen and Olsen 1978) and the lithoautotroph *Alcaligenes eutrophus* (Hogrefe and Friedrich 1984). Megaplasmids are also commonly found in plant-associated bacteria such as *Erwinia herbicola* (Gantotti and Beer 1982), *A. tumefaciens* (Casse *et al.* 1979; Unger *et al.* 1985;

Van Larebeke *et al.* 1974), *A. rhizogenes* (White and Nester 1980), and *P. solanacearum* (Boucher *et al.* 1986; Rosenberg *et al.* 1982).

Most isolates of *R. meliloti* harbour two megaplasms, generally ranging from 1200 to 1700 kb in size (Banfalvi *et al.* 1985; Bromfield *et al.* 1987; Burkhardt and Burkhardt 1984; Burkhardt *et al.* 1987; Charles and Finan 1991; Finan *et al.* 1986; Honeycutt *et al.* 1993; Hynes *et al.* 1986; Rosenberg *et al.* 1982). The genome of *R. meliloti* wild type strain SU47 is composed of a chromosome (3540 kb) and two megaplasms designated pRmeSU47a or pSym (1340 kb) and pRmeSU47b or pExo (1700 kb). The study of genes involved in symbiosis, prompted by the agricultural importance of biological N₂-fixation, has led to the discovery of megaplasmid-located symbiotic clusters.

Analysis of the *R. meliloti* Megaplasms

The two megaplasms carry much of the symbiotic determinants in *R. meliloti* SU47. The pSym megaplasmid carries *nod*, *nif* and *fix* genes which are involved in nodule induction and N₂-fixation (Banfalvi *et al.* 1981; Buikema *et al.* 1983; David *et al.* 1987; Debelle *et al.* 1986; Long *et al.* 1982; Rosenberg *et al.* 1981). The larger pExo megaplasmid, the focus of this study, carries genes required for efficient N₂-fixation, such as genes involved in exopolysaccharide (*exo*, *exp*) and lipopolysaccharide (*lps*) synthesis (Finan *et al.* 1985; Finan *et al.* 1986; Glucksmann *et al.* 1993; Hynes *et al.* 1986; Glazebrook and Walker 1989; Williams *et al.* 1990; Zhan *et al.* 1989), dicarboxylic acid transport (*dct*) (Finan *et al.* 1988; Watson *et al.* 1988), phosphate transport (*pho*) (Bardin *et al.* 1996; Bardin *et al.* 1998; Charles *et al.* 1991; Voegelé *et al.* 1997), as well

as genes involved in pyrroloquinoline quinone (*pqq*) (Chain *et al.* 1998) and thiamine (*thi*) biosynthesis (Finan *et al.* 1986), as well as genes required for lactose (*lac*), dulcitol (*dul*), protochatechuate (*pca*), melibiose (*mel*), and acetoacetate and β -hydroxybutyrate (*bhb*) utilization (Charles and Finan, 1991). Much of the progress that has been made with regards to the location and identification of genes on the pExo megaplasmid has involved the use of the transposon Tn5.

The pExo Transposon Map

The transposon Tn5 and its derivatives have been frequently used for insertion mutagenesis in many gram-negative bacteria (reviewed by Berg and Berg 1983; Reznikoff 1982). Tn5 consists of an inner core region (2750 bp) flanked by two inverted, nearly identical copies of the IS50 insertion element (1534 bp) (Berg *et al.* 1975; Jorgensen *et al.* 1979). The central region varies for each Tn5 derivative, encoding alternative antibiotic resistances and/or functions such as a plasmid origin of replication (TnV) or β -galactosidase gene (Tn5-235) (Berg *et al.* 1980; Berg and Berg 1987; De Vos *et al.* 1986; Furuichi *et al.* 1985; Hirsh *et al.* 1986; Sasakawa and Yoshikawa 1987). It is generally possible to replace a transposon that has been inserted in the *R. meliloti* genome with a different derivative by homologous recombination between the flanking IS50 insertion sequences (De Vos *et al.* 1986; Finan *et al.* 1985).

These transposons insert randomly in the *R. meliloti* genome (Meade *et al.* 1982) and, once inserted, are extremely stable (secondary transposition has not been observed) (De Vos *et al.* 1986). In order to extend the capability for genetic analysis of the pExo megaplasmid, a genetic linkage map was constructed (Charles and Finan 1990). This

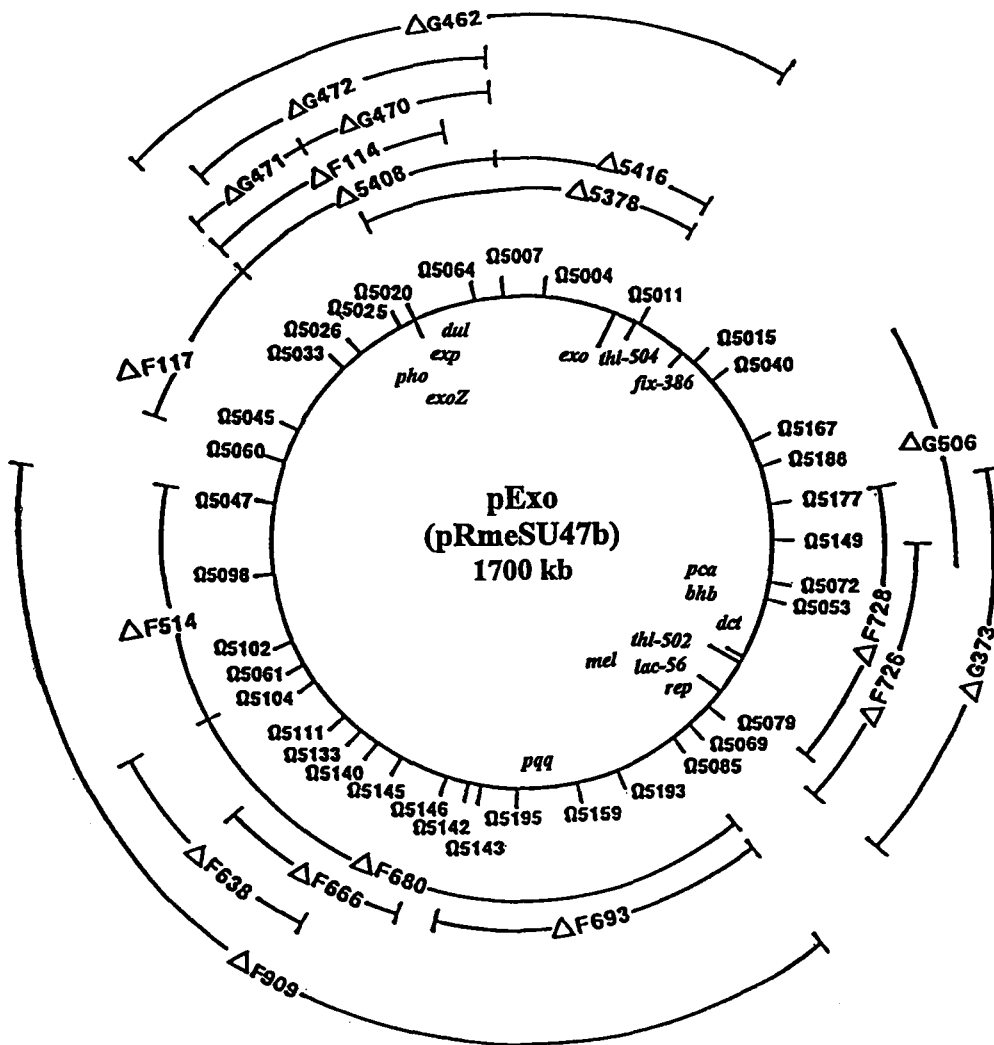
map consists of Tn5 (and Tn5-derivative) transposon insertions which are linked in Φ M12-mediated transduction (see Figure 1), and thus, the position of each insertion relative to any other is known. Using this genetic map and the subsequently generated megaplasmid deletions, several pExo genes (mentioned above) and their functions have been discovered (Charles *et al.* 1991).

The Next Step: Sequencing the pExo Megaplasmid

Despite what has thus far been discovered, over 95% of pExo remains unsequenced and uncharacterized. It appears likely that some of the genes located on this megaplasmid play significant roles in the interaction of *R. meliloti* with its environments, as evidenced by the identification of genes required for symbiosis (*dct*, *exo*, *pho*, amongst others) (Charles and Finan 1991). However, several questions have yet to be answered: what determines competitiveness in the soil; what exactly is involved in establishing host-plant specificity; and, which genes are important to the survival of the bacteria in the soil, laboratory, and symbiotic environments.

Analysis of the nucleotide sequence of the *R. meliloti* genome may go a long way toward answering these questions, and gaining a better overall understanding of what type of genes are present on the pExo megaplasmid, while providing insight into the structure of the *R. meliloti* genome and its evolutionary significance. In the present age, the determination of the entire genome sequence of an organism is claimed to be a prerequisite to understanding its complete biology (Fleischmann *et al.* 1995).

Figure 1: Circular map of the pExo megaplasmid showing the relative locations of representative transposon insertions (Ω). A collection of strains was generated, carrying transductionally linked transposon Tn5 and Tn5-derivative insertions located on the pExo megaplasmid. Defined deletions (Δ) and the known locations of genes or operons are shown. Abbreviations are: *bhb*, β -hydroxy butyrate utilization; *dct*, dicarboxylic acid transport; *dul*, dulcitol utilization; *exo*, *exoZ*, and *exp*, exopolysaccharide biosynthesis; *fix*, N₂-fixation; *lac*, lactose utilization; *mel*, α -galactoside utilization (melibiose utilization); *pca*, protocatechuic acid utilization; *rep*, replication; *thi*, thiamine biosynthesis. This figure was modified from Charles and Finan (1991).



Complete genome sequencing and analysis has ushered in an era of true comparative genomics. Much can be deciphered from genes identified through sequencing by comparing them to the information contained on publicly available databases. The first complete sequence of a rhizobial symbiotic (sym)-megaplasmid, the 536 kb megaplasmid of *Rhizobium* sp. NGR234, is currently available (Freiberg *et al.* 1997) and will facilitate the cataloguing of symbiotic and other genes. Moreover, the daily addition of genetic information into databases and the advent of complete genome sequencing have improved our understanding of genome organization and structure.

There are presently 18 published, complete genomes (see Table 1) and numerous others currently being sequenced in their entirety: more than 45 prokaryotic genomes and 15 eukaryotic genomes are in progress, with over 15 of these scheduled for release by the year 2000. Regularly updated lists of genome sequencing projects are available at World Wide Web sites <http://www.mcs.anl.gov/home/gaasterl/genomes.html> (MAGPIE, Argonne National Laboratory, Illinois, USA), <http://www.tigr.org/tdb/mdb/mdb.html> (TIGR, Gaithersburg, Maryland, USA), and <http://www.genome.ad.jp/kegg/kegg2.html> (KEGG, Institute for Chemical Research, Kyoto University, Japan).

Table 1: Published Complete Genomes as of December, 1998

Category	Species	Genome Size (bp)	Reference:
ARCHAEA			
Euryarchaeota	<i>Archaeoglobus fulgidus</i> DSM4304	2,178,400	Klenk <i>et al.</i> 1997
	<i>Methanobacterium</i> <i>thermoautotrophicum</i> ΔH	1,751,377	Smith <i>et al.</i> 1997
	<i>Methannococcus jannaschii</i> DSM2661	1,664,977	Bult <i>et al.</i> 1996
	<i>Pyrococcus horikoshii</i> OT3	1,738,505	Kawarabayasi <i>et al.</i> 1998
EUBACTERIA			
Cyanobacteria	<i>Synechocystis</i> sp. PCC6803	3,573,470	Kaneko <i>et al.</i> 1996
Gram-positive	<i>Mycobacterium tuberculosis</i> H37Rv	4,411,529	Cole <i>et al.</i> 1998
	<i>Bacillus subtilis</i> 168	4,214,807	Kunst <i>et al.</i> 1997
	<i>Mycoplasma genitalium</i> G-37	580,070	Fraser <i>et al.</i> 1995
	<i>Mycoplasma pneumoniae</i> M129	816,394	Himmelreich <i>et al.</i> 1996
O ₂ -reducing bacteria	<i>Aquifex aeolicus</i> VF5	1,551,335	Deckert <i>et al.</i> 1998
Chlamydiales	<i>Chlamydia trachomatis</i> Serovar D (D/UW-3/Cx)	1,042,519	Stephens <i>et al.</i> 1998
α-Proteobacteria	<i>Rickettsia prowazekii</i> Madrid E	1,111,523	Andersson <i>et al.</i> 1998
γ-Proteobacteria	<i>Escherichia coli</i> K-12 MG1655	4,639,221	Blattner <i>et al.</i> 1997
	<i>Haemophilus influenzae</i> Rd KW20	1,830,137	Fleischmann <i>et al.</i> 1995
ε-Proteobacteria	<i>Helicobacter pylori</i> 26695	1,667,867	Tomb <i>et al.</i> 1997
Spirochaetales	<i>Borrelia burgdorferi</i> B31	1,230,663	Fraser <i>et al.</i> 1997
	<i>Treponema pallidum</i> (Nichols)	1,137,961	Fraser <i>et al.</i> 1998
EUKARYOTAE			
Fungi	<i>Saccharomyces cerevisiae</i> S288C	12,069,313	Goffeau <i>et al.</i> 1997

α, γ, and ε are the alpha, gamma, and epsilon subdivisions of Proteobacteria

A collaborative international effort has been initiated to sequence the entire *R. meliloti* genome (T. M. Finan, personal communication), led by Drs. Jacques Batut and Francis Galibert (chromosome), Dr. Sharon Long (pSym), and Drs. Alf Puhler and Turlough Finan (pExo). The pExo megaplasmid, at 1700 kb, is approximately as large as, or larger than over half of the complete genomes published to date (refer to Table 1). Unlike eukaryotes, most bacteria harbour very little extraneous sequences, hence, sequencing of this megaplasmid is expected to be information rich.

Common Whole-Genome Sequencing Methodologies

Large-scale systematic sequencing has generally depended on the availability of an ordered library of large-insert bacterial genomic clones. The generation of these large insert (50-200 kb) libraries, and the localization of each clone on a genome map, is a laborious and time-consuming process. Although there exist several approaches to genome sequencing, one strategy has been predominantly used in an effort to overcome such problems (Bult *et al.* 1996; Deckert *et al.* 1998; Fleischmann *et al.* 1995; Fraser *et al.* 1995; Fraser *et al.* 1997; Fraser *et al.* 1998; Klenk *et al.* 1997; Kunst *et al.* 1997; Smith *et al.* 1997; Tomb *et al.* 1997), a process known as whole-genome random sequencing (Deininger, 1983). In this procedure, a single random-DNA-fragment (1-2 kb) library is sequenced and assembled in two steps.

In the initial random sequencing or “shotgun” phase, randomly selected fragments are sequenced from both ends and ordered into contiguous sequences (contigs) to produce an almost complete genome. In the directed-sequencing or “finishing” phase, gaps are filled with the use of customized primers: sequence gaps are closed by primer walking

and physical gaps are closed by combinatorial PCR followed by sequencing of the PCR product.

However, this strategy does have its disadvantages, and is not feasible when only one replicon of a multi-replicon genome is to be sequenced, such as pExo. In these cases, the purity of the starting DNA is critical to avoid major problems in assembly (Churcher *et al.* 1997; Fleischmann *et al.* 1995; Gardner *et al.* 1998; Goffeau *et al.* 1997). The currently favored approach for sequencing the replicons of such organisms involves the selection of an overlapping set of bacterial artificial chromosomes (BACs) or cosmids carrying large DNA inserts, randomly shearing this DNA to construct shotgun libraries, and then sequencing many of these fragments from the library (Churcher *et al.* 1997; Cole *et al.* 1998; Dietrich *et al.* 1997; Freiberg *et al.* 1997; Kawarabayasi *et al.* 1998). Like the whole-genome random shotgun approach, this method, entitled "directed random shotgun sequencing", requires highly redundant sequencing (on average, 5-7 fold genome coverage and as high as 10-12.5 fold) to obtain a complete and accurate finished consensus sequence (Bouck *et al.* 1998; Deckert *et al.* 1998; Dietrich *et al.* 1997; Fleischmann *et al.* 1995; Fraser *et al.* 1995; Fraser *et al.* 1997; Fraser *et al.* 1998; Gardner *et al.* 1998; Klenk *et al.* 1997; Smith *et al.* 1997; Voss *et al.* 1995; Wild *et al.* 1996).

A lower redundancy sequence or "limited shotgun" strategy (2-4 fold genome coverage) can still be of use to the scientific community by rapidly generating a genetic blueprint, while postponing the majority of work involved in producing a high quality consensus sequence (Bouck *et al.* 1998). Combined with primer walking, this modified approach has been used to sequence the entire genome of the bacterium *Mycoplasma*

pneumoniae (2.95 average sequencing redundancy) (Himmelreich *et al.* 1996; Voss *et al.* 1993; Voss *et al.* 1995).

The “directed” random sequencing approach may still involve extensive overlap before sequencing even begins, in the initial selection of representative large-insert clones (Churcher *et al.* 1997; Dietrich *et al.* 1997; Wild *et al.* 1996). In order to overcome this initial overlap, an even more directed approach can be used which consists of the integration of highly-specific recognition sequences at known locations in the genome for later excision and purification of the DNA between two such sequences. Two methods following this principle were used in the effort to sequence part of the complete *Escherichia coli* genome (Blattner *et al.* 1997).

Specific Targeted Sequencing Methodologies

One approach to precisely excising DNA of known location was used by Blattner *et al.* (1997) to sequence part of the *E. coli* genome. This approach involved the I-SceI meganuclease, a site-specific intron-encoded endonuclease from *Saccharomyces cerevisiae* mitochondria (Monteilhet *et al.* 1990; Perrin *et al.* 1993). Due to its 18 bp non-palindromic recognition sequence (Colleaux *et al.* 1988), I-SceI meganuclease has an extremely low probability of cutting DNA, even within large genomes (Blattner *et al.* 1997; Bloch *et al.* 1996; Boehringer Mannheim).

An I-SceI site was placed on a transposable element to produce a mapped collection of *E. coli* strains, each with a unique I-SceI site (Bloch *et al.* 1996; Rode *et al.* 1995). Two I-SceI sites, at chosen locations, were combined by P1 transduction, and the intervening sequence excised with I-SceI and purified by pulsed-field gel electrophoresis

(Blattner *et al.* 1997). This method however, has a few disadvantages, namely the time-consuming process of mapping the *I-SceI* insertion sites and the difficulties associated with purification of pulsed-field DNA fragments.

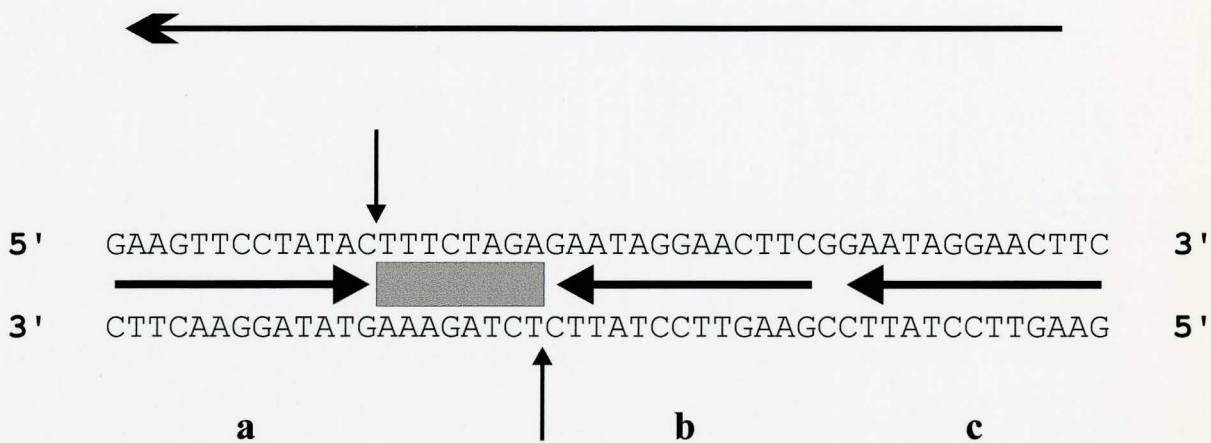
The second approach employed by Blattner *et al.* 1997 to excise particular regions from the *E. coli* genome, entailed the use of Flp (Broach and Hicks 1980), a site-specific recombinase, and its target sequence, *FRT*. Two *FRT* sites (carried on two different vectors) were introduced into the *E. coli* genome at precise locations, in such a way that the addition of Flp *in trans* would excise the intervening region of interest, creating a “pop-out plasmid” (Blattner *et al.* 1997; Pósfai *et al.* 1994; Wild *et al.* 1996). These “targeting” vectors also contained a portion of the target region and were directed to specific locations in the genome via homologous recombination. After Flp-mediated excision, the “pop-out” plasmid was amplified, captured and purified by pulsed-field gel electrophoresis (Blattner *et al.* 1997). The disadvantages in this particular methodology are: firstly, the precluded need to already know the sequence of the sites to which the targeting vectors will be directed; secondly, the necessity of engineering specific targeting vectors for each desired excision event; and lastly, once again, the problems involved in obtaining ultra-pure large DNA fragments using pulsed-field gel electrophoresis. In the work presented in this thesis, a modified strategy involving the Flp/*FRT* system (Pósfai *et al.* 1994; Wild *et al.* 1996) was used to obtain large *R. meliloti* pExo DNA fragments (50-130 kb).

The Yeast 2 μ m Plasmid Flp/*FRT* System of Site-Specific Recombination

Flp is a site-specific recombinase, encoded by the 2 μ m plasmid of the yeast *S. cerevisiae* (Broach and Hicks 1980; Futcher 1986; Reynolds *et al.* 1987; Volkert and Broach 1986). The name *flp* arose from the fact that the gene product (Flp) causes an inversion (flip-flop) of a segment of the 2 μ m plasmid. Flp promotes recombination at 2 distinct sequences within the 2 μ m plasmid, enabling increased copy number by circumventing the host cell-cycle control of DNA replication (reviewed by Broach 1981; Cox 1988; Futcher 1988; Sadowski 1995). Flp is a member of the integrase family of recombinases, which includes the Cre recombinase of the bacteriophage P1 and the Int recombinase of bacteriophage λ , among others. These recombinases share a weak amino acid identity as well as a similar mechanism of recombination (Argos *et al.* 1986; Craig 1988).

The site at which Flp recombinase acts is called the Flp recombination target (*FRT*) site. This site consists of three (elements **a**, **b** and **c**) 13 bp repeats surrounding an 8 bp asymmetrical core (see Figure 2). The 13 bp repeats serve as binding sites for one molecule of Flp protein (Andrews *et al.* 1987; Pan *et al.* 1991). Only the two inverted repeats directly flanking the core (elements **a** and **b**) are required for recombination; the other repeat (element **c**) is dispensable for all known functions of Flp (Andrews *et al.* 1985; Jayaram 1985; Senecoff *et al.* 1985). The Flp recombinase cleaves the *FRT* sites at the boundaries of the core.

Figure 2: Diagram of the Flp recombination target (*FRT*) site. The *FRT* site is characterized by three 13 bp repeat elements **a**, **b**, and **c** (horizontal arrows), surrounding an 8 bp core region (shaded box). Flp binds repeats **a** and **b**, and cleaves the *FRT* site at either end of the core (vertical arrows). The asymmetry of the core region determines directionality of the *FRT* site (indicated by large arrow).



The alignment of two core sequences is one of the factors in determining the course of a Flp-mediated recombination reaction. The 8 bp core region is not symmetrical; this asymmetry dictates the directionality of the reaction. Thus, Flp-mediated recombination between two inverted *FRT* sites causes inversion of the intervening DNA (as in the 2 μ m plasmid), whereas recombination between two *FRT* repeats in direct orientation results in excision of the DNA between them, each new daughter molecule harbouring only one *FRT* site. The length of the core is also important since an increase in size by as little as 2 bp inactivates the *FRT* site (Senecoff and Cox 1986). However, the actual sequence of the bases in the core region of the *FRT* site is not important for recombination, provided the cores of the two participating *FRT* sites are identical (Andrews *et al.* 1986; Senecoff and Cox 1986).

This Work: A Novel Use For the Flp/*FRT* Recombination System in *R. meliloti*

As part of an international effort to sequence the complete genome of *R. meliloti*, the work presented in this thesis was initiated to develop a strategy to excise large (50-200 kb) fragments of the pExo megaplasmid and capture these in *E. coli* for amplification, purification and sequencing steps. The scheme is based on the work described by Pósfai *et al.* (1994) and Wild *et al.* (1996). The goal of this work was to make use of the pExo genetic map (Charles and Finan 1990) (refer to Figure1), by targeting selected Tn5 and Tn5-132 transposons, with vectors containing several elements useful for further manipulation (*FRT* site, I-*Sce*I site, *oriT*, and an origin of

replication). After selectively combining, in one *R. meliloti* strain, two such targeted transposons, the intervening region of interest can be either:

- deleted *in vivo* by supplying Flp *in trans*
- transferred directly (or after excision) into a recipient strain of *E. coli* or *A.*

tumefaciens

- purified by pulsed-field gel electrophoresis after digestion with I-SceI meganuclease

The initial stage was devoted to the construction of the “delivery” and “targeting” vectors. The delivery plasmids were made to consist of a backbone which would replicate in *R. meliloti* and carried *flp* under the control of an inducible promoter. The targeting vectors were constructed to contain a 300 bp portion of the outside end of the IS50 element (used for targeting the Tn5 transposons), an I-SceI site (used for purification purposes), an *FRT* site (used for site-specific excision), an *oriT* (used for transfer of the targeting vectors into recipient *R. meliloti* strains and of the DNA region of interest into *E. coli* recipients), an *oriV* (used for replication of the large pExo region in *E. coli*), and the Ω Sp interposon (used to select for integration of the target constructs and for transfer of the large pExo region to *E. coli*). Two backbones were used to contain these elements: pBluescript (Ap^r) and pBAC (Cm^r). Whereas pBluescript is not known to stably maintain inserts of 50 kb or more (and so, any replication of large inserts would rely solely on the effectiveness of *oriV*), the pBAC backbone has been shown to stably sustain large inserts of up to 350 kb (Shizuya *et al.* 1992; Kim *et al.* 1992; Kim *et al.* 1996a; Kim *et al.* 1996b).

Once the targeting vectors were constructed, they were inserted into selected locations throughout the pExo megaplasmid via homologous recombination at IS50 insertion sequences. Pairs of targeted Tn5 and Tn5-132 insertions were then constructed via Φ M12-mediated transduction. The regions of interest in the resulting strains were either excised and deleted by delivering Flp *in trans*, or transferred as a plasmid to an *E. coli* or *A. tumefaciens* recipient by providing the transfer functions *in trans*. After verification by Southern analysis that the transferred DNA was in fact from *R. meliloti*, *EcoRI* fragments from a 50 kb excised region were subcloned and sequenced from both ends. In addition to confirming that the DNA obtained in this work was from *R. meliloti*, this initial sequencing venture also confirmed that the desired target region had been retrieved. Shotgun sequencing of the excised plasmids is currently underway.

The proposed strategy has many advantages over more random approaches. With defined regions of the pExo megaplasmid transferred and stably maintained in an *E. coli* background, contamination with DNA from the other two replicons is avoided when generating random clones for sequencing. Contamination with *E. coli* DNA can easily be screened out by comparing the sequences to the publicly released *E. coli* genome. With the use of the pExo Tn5 genetic map, specific Tn5 and Tn5-132 transposons can be targeted, the intervening region excised, all the while avoiding the labour-intensive, time-consuming process of ordering randomly generated, overlapping BAC clones.

CHAPTER 2. MATERIALS AND METHODS

Bacterial Strains, Plasmids and Transposons

The bacterial strains, plasmids and transposons used in this work are listed in Table 2. The genotype or relevant characteristics as well as the reference, source or method of construction are indicated for each. All plasmids were stored frozen at -20°C in T₁₀E₁ (10 mM Tris-HCl, 1 mM EDTA pH 8.0) or ddH₂O (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). Such frozen permanents were made by diluting 0.5 ml of LB or LBmc broth bacterial cultures, grown overnight, with an equivalent volume of LB containing 14% DMSO. Using sterile inoculation sticks, viable cells were recovered from frozen permanent cultures by removing ice crystals and streaking onto appropriate selective solid media.

Table 2: Bacterial Strains, Plasmids and Transposons

Strain, plasmid or transposon	Genotype/Relevant characteristics	Source, reference or construction
<i>Rhizobium meliloti</i>		
RCR2011	SU47 wild type	Vincent 1941
Rm1021	SU47, <i>str-21</i>	Meade <i>et al.</i> 1982
Rm5000	SU47, <i>rif-5</i>	Finan <i>et al.</i> 1984
<u>Rm1021 derivatives</u>		
Rm5079	Ω5007::Tn5-132	Finan <i>et al.</i> 1985
Rm5381	Ω5035::Tn5-132	Charles and Finan 1990
Rm5455	<i>lac-56</i> ::Tn5 (Ω5056::Tn5)	Laboratory collection
RmF152	Ω5069::Tn5-132	Laboratory collection
RmF153	<i>dctB12</i> ::Tn5-132	Yarosh <i>et al.</i> 1989
RmF164	Ω5007::TnV	Charles and Finan 1990
RmF232	Ω5085::Tn5-233	I. Oresnik
RmF327	Ω5102::Tn5	Charles and Finan 1990
RmF495	Ω5140::Tn5-132	Charles and Finan 1990
RmF497	Ω5142::Tn5-132	Charles and Finan 1990
RmF532	Ω5143::Tn5-233	Charles and Finan 1990
RmF535	Ω5146::Tn5-233	Charles and Finan 1990
RmF570	Ω5159::Tn5	Charles and Finan 1990
RmF698	Ω5195::Tn5-233	Charles and Finan 1990
RmF951	Ω5079::Tn5-233	Charles and Finan 1990
RmG212	<i>lac</i> ⁻	Jane Glazebrook
RmG497	Ω5033::Tn5-233 (<i>sfx-2</i>)	Oresnik <i>et al.</i> 1994

RmH936	$\Omega 5033::Tn5-235$	Tn5-235 replacement of $\Omega 5033::Tn5-233$ in RmG497, Nm ^r
RmH937	$\Omega 5033::Tn5-235FRT$	Tn5-235FRT replacement of $\Omega 5033::Tn5-233$ in RmG497, Nm ^r
RmH939	<i>lac</i> , $\Omega 5033::Tn5-235$	$\Phi(RmH936) \rightarrow RmG212$, Nm ^r , <i>lac</i> ⁺
RmH940	<i>lac</i> , $\Omega 5033::Tn5-235FRT$	$\Phi(RmH937) \rightarrow RmG212$, Nm ^r , <i>lac</i> ⁺
RmK134	$\Omega 5142::Tn5-132::pTH456$	$\Phi(RmH974) \rightarrow Rm1021$, Sp ^r Ot ^r
RmK135	$\Omega 5142::Tn5-132::pTH456$	$\Phi(RmH975) \rightarrow Rm1021$, Sp ^r Ot ^r
RmK137	$\Omega 5056::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH456$	$\Phi(RmH957) \rightarrow RmK134$, Sp ^r Nm ^r Ot ^r
RmK138	$\Omega 5056::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH456$	$\Phi(RmH957) \rightarrow RmK135$, Sp ^r Nm ^r Ot ^r
RmK139	$\Omega 5056::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH456$	$\Phi(RmH958) \rightarrow RmK134$, Sp ^r Nm ^r Ot ^r
RmK140	$\Omega 5056::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH456$	$\Phi(RmH958) \rightarrow RmK135$, Sp ^r Nm ^r Ot ^r
RmK163	$\Omega 5142::Tn5-132::pTH504$	$\Phi(RmK147) \rightarrow Rm1021$, Sp ^r Ot ^r
RmK164	$\Omega 5142::Tn5-132::pTH504$	$\Phi(RmK148) \rightarrow Rm1021$, Sp ^r Ot ^r
RmK165	$\Omega 5102::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH504$	$\Phi(RmH953) \rightarrow RmK163$, Sp ^r Nm ^r Ot ^r
RmK166	$\Omega 5102::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH504$	$\Phi(RmH953) \rightarrow RmK164$, Sp ^r Nm ^r Ot ^r
RmK167	$\Omega 5102::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH504$	$\Phi(RmH954) \rightarrow RmK163$, Sp ^r Nm ^r Ot ^r
RmK168	$\Omega 5102::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH504$	$\Phi(RmH954) \rightarrow RmK164$, Sp ^r Nm ^r Ot ^r
RmK169	$\Omega 5159::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH504$	$\Phi(RmH955) \rightarrow RmK163$, Sp ^r Nm ^r Ot ^r
RmK170	$\Omega 5159::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH504$	$\Phi(RmH955) \rightarrow RmK164$, Sp ^r Nm ^r Ot ^r
RmK171	$\Omega 5159::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH504$	$\Phi(RmH956) \rightarrow RmK163$, Sp ^r Nm ^r Ot ^r
RmK172	$\Omega 5159::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH504$	$\Phi(RmH956) \rightarrow RmK164$, Sp ^r Nm ^r Ot ^r
RmK173	$\Omega 5056::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH504$	$\Phi(RmH957) \rightarrow RmK163$, Sp ^r Nm ^r Ot ^r
RmK174	$\Omega 5056::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH504$	$\Phi(RmH957) \rightarrow RmK164$, Sp ^r Nm ^r Ot ^r
RmK175	$\Omega 5056::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH504$	$\Phi(RmH958) \rightarrow RmK163$, Sp ^r Nm ^r Ot ^r
RmK176	$\Omega 5056::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH504$	$\Phi(RmH958) \rightarrow RmK164$, Sp ^r Nm ^r Ot ^r
RmK182*	$\Omega 5069::Tn5-132::pTH504$	$\Phi(RmK180) \rightarrow Rm1021$, Sp ^r Ot ^r
RmK183*	$\Omega 5069::Tn5-132::pTH504$	$\Phi(RmK181) \rightarrow Rm1021$, Sp ^r Ot ^r
RmK184*	$\Omega 5102::Tn5::pTH455$ $\Omega 5069::Tn5-132::pTH504$	$\Phi(RmH954) \rightarrow RmK182$, Sp ^r Nm ^r Ot ^r
RmK185*	$\Omega 5102::Tn5::pTH455$ $\Omega 5069::Tn5-132::pTH504$	$\Phi(RmH954) \rightarrow RmK183$, Sp ^r Nm ^r Ot ^r
RmK186*	$\Omega 5159::Tn5::pTH455$ $\Omega 5069::Tn5-132::pTH504$	$\Phi(RmH955) \rightarrow RmK182$, Sp ^r Nm ^r Ot ^r
RmK187*	$\Omega 5159::Tn5::pTH455$ $\Omega 5069::Tn5-132::pTH504$	$\Phi(RmH955) \rightarrow RmK183$, Sp ^r Nm ^r Ot ^r
RmK188*	$\Omega 5056::Tn5::pTH455$ $\Omega 5069::Tn5-132::pTH504$	$\Phi(RmH957) \rightarrow RmK182$, Sp ^r Nm ^r Ot ^r
RmK189*	$\Omega 5056::Tn5::pTH455$ $\Omega 5069::Tn5-132::pTH504$	$\Phi(RmH957) \rightarrow RmK183$, Sp ^r Nm ^r Ot ^r

RmK190*	$\Omega 5056::Tn5::pTH455$	$\Omega 5069::Tn5-132::pTH504$	$\Phi(RmH958) \rightarrow RmK182, Sp^r Nm^r Ot^r$
RmK191*	$\Omega 5056::Tn5::pTH455$	$\Omega 5069::Tn5-132::pTH504$	$\Phi(RmH958) \rightarrow RmK183, Sp^r Nm^r Ot^r$
RmK192	$\Omega 5143::Tn5$		Tn5 replacement of $\Omega 5143::Tn5-233$ in RmF532, Nm ^r
RmK195	$\Omega 5069::Tn5-132::pTH509$		$\Phi(RmK193) \rightarrow Rm1021, Sp^r Ot^r$
RmK196	$\Omega 5069::Tn5-132::pTH509$		$\Phi(RmK194) \rightarrow Rm1021, Sp^r Ot^r$
RmK197	$\Omega 5159::Tn5::pTH455$	$\Omega 5069::Tn5-132::pTH509$	$\Phi(RmH955) \rightarrow RmK195, Sp^r Nm^r Ot^r$
RmK198	$\Omega 5159::Tn5::pTH455$	$\Omega 5069::Tn5-132::pTH509$	$\Phi(RmH955) \rightarrow RmK196, Sp^r Nm^r Ot^r$
RmK199	$\Omega 5159::Tn5::pTH455$	$\Omega 5069::Tn5-132::pTH509$	$\Phi(RmH956) \rightarrow RmK195, Sp^r Nm^r Ot^r$
RmK200	$\Omega 5159::Tn5::pTH455$	$\Omega 5069::Tn5-132::pTH509$	$\Phi(RmH956) \rightarrow RmK196, Sp^r Nm^r Ot^r$
RmK201	$\Omega 5056::Tn5::pTH455$	$\Omega 5069::Tn5-132::pTH509$	$\Phi(RmH957) \rightarrow RmK195, Sp^r Nm^r Ot^r$
RmK202	$\Omega 5056::Tn5::pTH455$	$\Omega 5069::Tn5-132::pTH509$	$\Phi(RmH957) \rightarrow RmK196, Sp^r Nm^r Ot^r$
RmK203	$\Omega 5056::Tn5::pTH455$	$\Omega 5069::Tn5-132::pTH509$	$\Phi(RmH958) \rightarrow RmK195, Sp^r Nm^r Ot^r$
RmK204	$\Omega 5056::Tn5::pTH455$	$\Omega 5069::Tn5-132::pTH509$	$\Phi(RmH958) \rightarrow RmK196, Sp^r Nm^r Ot^r$

Rm5000 derivatives

RmH921	$\Omega 5007::TnV$		$\Phi(RmF164) \rightarrow Rm5000, Nm^r$
RmH922	$\Omega 5102::Tn5$		$\Phi(RmF327) \rightarrow Rm5000, Nm^r$
RmH923	$\Omega 5159::Tn5$		$\Phi(RmF570) \rightarrow Rm5000, Nm^r$
RmH924	$\Omega 5056::Tn5$		$\Phi(Rm5455) \rightarrow Rm5000, Nm^r$
RmH942	$\Omega 5035::Tn5-132$		$\Phi(Rm5381) \rightarrow Rm5000, Ot^r$
RmH943	$\Omega 5069::Tn5-132$		$\Phi(RmF152) \rightarrow Rm5000, Ot^r$
RmH944	$\Omega 5142::Tn5-132$		$\Phi(RmF497) \rightarrow Rm5000, Ot^r$
RmH946	$\Omega 5140::Tn5-132$		$\Phi(RmF495) \rightarrow Rm5000, Ot^r$
RmH951	$\Omega 5007::TnV::pTH455$		pTH455 integration into one IS50 element of RmH921, Sp ^r Nm ^r
RmH953	$\Omega 5102::Tn5::pTH455$		pTH455 integration into one IS50 element of RmH922, Sp ^r Nm ^r
RmH954	$\Omega 5102::Tn5::pTH455$		pTH455 integration into other IS50 element of RmH922, Sp ^r Nm ^r

RmH955	$\Omega 5159::Tn5::pTH455$	pTH455 integration into one IS50 element of RmH923, Sp ^r Nm ^r
RmH956	$\Omega 5159::Tn5::pTH455$	pTH455 integration into other IS50 element of RmH923, Sp ^r Nm ^r
RmH957	$\Omega 5056::Tn5::pTH455$	pTH455 integration into one IS50 element of RmH924, Sp ^r Nm ^r
RmH958	$\Omega 5056::Tn5::pTH455$	pTH455 integration into other IS50 element of RmH924, Sp ^r Nm ^r
RmH959	$\Omega 5035::Tn5-132::pTH456$	pTH456 integration into other IS50 element of RmH942, Sp ^r Nm ^r
RmH960	$\Omega 5035::Tn5-132::pTH456$	pTH455 integration into other IS50 element of RmH942, Sp ^r Nm ^r
RmH969	$\Omega 5069::Tn5-132::pTH456$	pTH455 integration into one IS50 element of RmH943, Sp ^r Ot ^r
RmH970	$\Omega 5069::Tn5-132::pTH456$	pTH455 integration into other IS50 element of RmH943, Sp ^r Ot ^r
RmH974	$\Omega 5142::Tn5-132::pTH456$	pTH455 integration into other IS50 element of RmH944, Sp ^r Ot ^r
RmH975	$\Omega 5142::Tn5-132::pTH456$	pTH455 integration into other IS50 element of RmH944, Sp ^r Ot ^r
RmH990	$\Omega 5007::TnV::pTH455$ $\Omega 5035::Tn5-132::pTH456$	$\Phi(RmH951) \rightarrow RmH959$, Sp ^r Nm ^r Ot ^r
RmH991	$\Omega 5007::TnV::pTH455$ $\Omega 5035::Tn5-132::pTH456$	$\Phi(RmH951) \rightarrow RmH960$, Sp ^r Nm ^r Ot ^r
RmH992	$\Omega 5056::Tn5::pTH455$ $\Omega 5069::Tn5-132::pTH456$	$\Phi(RmH957) \rightarrow RmH969$, Sp ^r Nm ^r Ot ^r
RmH993	$\Omega 5056::Tn5::pTH455$ $\Omega 5069::Tn5-132::pTH456$	$\Phi(RmH957) \rightarrow RmH970$, Sp ^r Nm ^r Ot ^r
RmH994	$\Omega 5056::Tn5::pTH455$ $\Omega 5069::Tn5-132::pTH456$	$\Phi(RmH958) \rightarrow RmH969$, Sp ^r Nm ^r Ot ^r
RmH995	$\Omega 5056::Tn5::pTH455$ $\Omega 5069::Tn5-132::pTH456$	$\Phi(RmH958) \rightarrow RmH970$, Sp ^r Nm ^r Ot ^r
RmK126	$\Omega 5102::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH456$	$\Phi(RmH953) \rightarrow RmH974$, Sp ^r Nm ^r Ot ^r
RmK127	$\Omega 5102::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH456$	$\Phi(RmH953) \rightarrow RmH975$, Sp ^r Nm ^r Ot ^r
RmK128	$\Omega 5102::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH456$	$\Phi(RmH954) \rightarrow RmH974$, Sp ^r Nm ^r Ot ^r
RmK129	$\Omega 5102::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH456$	$\Phi(RmH954) \rightarrow RmH975$, Sp ^r Nm ^r Ot ^r
RmK130	$\Omega 5159::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH456$	$\Phi(RmH955) \rightarrow RmH974$, Sp ^r Nm ^r Ot ^r
RmK131	$\Omega 5159::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH456$	$\Phi(RmH955) \rightarrow RmH975$, Sp ^r Nm ^r Ot ^r
RmK132	$\Omega 5159::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH456$	$\Phi(RmH956) \rightarrow RmH974$, Sp ^r Nm ^r Ot ^r
RmK133	$\Omega 5159::Tn5::pTH455$ $\Omega 5142::Tn5-132::pTH456$	$\Phi(RmH956) \rightarrow RmH975$, Sp ^r Nm ^r Ot ^r

RmK147	Ω5142::Tn5-132::pTH504	pTH504 integration into one IS50 element of RmH944, Sp ^r Ot ^r
RmK148	Ω5142::Tn5-132::pTH504	pTH504 integration into other IS50 element of RmH944, Sp ^r Ot ^r
RmK180	Ω5069::Tn5-132::pTH504	pTH504 integration into one IS50 element of RmH943, Sp ^r Ot ^r
RmK181	Ω5069::Tn5-132::pTH504	pTH504 integration into other IS50 element of RmH943, Sp ^r Ot ^r
RmK193	Ω5069::Tn5-132::pTH509	pTH509 integration into one IS50 element of RmH943, Sp ^r Ot ^r
RmK194	Ω5069::Tn5-132::pTH509	pTH509 integration into other IS50 element of RmH943, Sp ^r Ot ^r
RmK205	<i>dctB12</i> ::Tn5-132	Φ(RmF153)→Rm5000, Ot ^r
RmK207	Ω5195::Tn5-233	Φ(RmF698)→Rm5000, Gm ^r Sp ^r
RmK208	Ω5146::Tn5-233	Φ(RmF535)→Rm5000, Gm ^r Sp ^r
RmK209	Ω5085::Tn5-233	Φ(RmF232)→Rm5000, Gm ^r Sp ^r
RmK210	Ω5079::Tn5-233	Φ(RmF951)→Rm5000, Gm ^r Sp ^r

Agrobacterium tumefaciens

GMI9050	C58C1, Sm ^r Rf pAtC58	Rosenberg and Huguet 1984
GMI9023	GMI9050, cured of pAtC58	Rosenberg and Huguet 1984

Escherichia coli

MM294A	<i>pro-82 thi-1 hsdR17 supE44 endA1</i>	Laboratory collection
MT607	MM294A <i>recA-56</i>	Finan <i>et al.</i> 1986
MT616	MT607 (pRK600)	Finan <i>et al.</i> 1986
MT620	MT607 Rf	T. Finan
DH5α	F ⁻ , <i>endA1</i> , <i>hsdR17</i> (r _k ⁻ , m _k ⁻), <i>supE44</i> , <i>thi-1</i> , <i>recA1</i> , <i>gyrA96</i> , <i>relA1</i> , Δ(<i>argF-lacZYA</i>) U169, φ80 <i>dlacZ</i> , ΔM15	BRL Inc.
JW192	DH5α, <i>trfA278D</i> integrated in the chromosome, Ap ^r	J. Wild

Plasmids

pUC118, pUC119	Ap ^r cloning vectors, ColE1 <i>oriV</i>	Vieira and Messing 1987
pBluescript II (pBS) SK and KS	Ap ^r cloning vectors, ColE1 <i>oriV</i>	Alting-Mees and Short 1989
pBBR1MCS	Cm ^r broad host range, mobilizable cloning vector, <i>lacZα</i>	Kovach <i>et al.</i> 1994
pBBR1MCS-5	Gm ^r broad host range, mobilizable cloning vector, <i>lacZα</i>	Kovach <i>et al.</i> 1995
pBR322	Ap ^r , Tc ^r cloning vector	Bolivar <i>et al.</i> 1977
pRK2013	ColE1 replicon with RK2 transfer region, Nm ^r -Km ^r	Figurski and Helinski 1979
pRK600	pRK2013 <i>npt</i> ::Tn9, Cm ^r Nm ^r -Km ^r	Finan <i>et al.</i> 1986
pRK602	pRK600::Tn5, Cm ^r Nm ^r -Km ^r	Finan <i>et al.</i> 1985
pGS220	Tn5 in deletion derivative of pBR322; Ap ^r , Nm ^r -Km ^r	De Vos <i>et al.</i> 1986
pGS235	Tn5-235 in deletion derivative of pBR322; Ap ^r , Nm ^r -Km ^r	De Vos <i>et al.</i> 1986
pHP45Ω	Ap ^r pBR322 derivative carrying the Sm ^r /Sp ^r Ω fragment	Prentki and Krisch 1984
pLD3	Ap ^r <i>flp</i> expression plasmid, pBR322 derivative	Pan <i>et al.</i> 1991
pJW106	Ap ^r pBR322 derivative carrying <i>tetR-P_{ter}flp-trfA278D</i>	Wild <i>et al.</i> 1996
pMS107-GENO14	Ap ^r pUC derivative carrying <i>FRT-oriV</i> cassette	W. Szybalski
pTC197	Ap ^s pUC19 derivative carrying Gm ^r fragment	T. Charles
pTJS82	pUC7 carrying <i>oriT</i> of RK2, Ap ^r	Schmidhauser and Helinski, 1985
pMP220	IncP, promoterless <i>lacZ</i> , Tc ^r	Spaink <i>et al.</i> 1987
pML31	miniF plasmid carrying <i>oriV</i> and <i>oriS</i> , Km ^r	Lovett and Helinski 1976
pMF21	<i>Bam</i> HI deletion of pML31, <i>oriS</i> , Km ^r	Manis and Kline 1977
pBAC108L	miniF derivative, Cm ^r	Shizuya <i>et al.</i> 1992
pBACe3.6	Ap ^r , Cm ^r pBAC108L derivative, carrying pUC19 and the <i>sacBII</i> gene	E. Frengen
pTH185	~5.3 kb <i>Eco</i> RI fragment carrying the <i>pcaD</i> gene in pUC119	MacPherson 1995
pTH424	~760 bp <i>Eco</i> RI/ <i>Sal</i> I <i>oriT</i> (RK2) cassette from pTJS82 into pBS SK	
pTH428	~900 bp <i>Bam</i> HI <i>oriV</i> / <i>FRT</i> cassette from pMS107-GENO14 into pBS KS (such that the <i>FRT</i> site is closest to <i>P_{lac}</i>)	
pTH430	334 bp <i>Xba</i> I/ <i>Spe</i> I PCR product of IS50 into <i>Xba</i> I site of pBS KS (such that the outside end of the IS50 is closest to <i>P_{lac}</i>)	
pTH431	334 bp <i>Xba</i> I/ <i>Spe</i> I PCR product of IS50 into <i>Xba</i> I site of pBS KS (opposite orientation with respect to pTH430)	
pTH438	~2.0 kb <i>Xma</i> I ΩSp ^r cassette from pHP45Ω into pTH428	

pTH444	~2.9 kb <i>EcoRI/SpeI oriV/FRT-ΩSp^r</i> cassette from pTH438 into pTH424	
pTH455	~3.7 kb <i>SpeI/SalI oriV/FRT-ΩSp^r-oriT</i> cassette from pTH444 into pTH431	
pTH456	~3.7 kb <i>SpeI/SalI oriV/FRT-ΩSp^r-oriT</i> cassette from pTH444 into pTH430	
pTH462	minimal <i>FRT</i> oligonucleotides (in parallel) in the two <i>BglII</i> sites of pGS235	
pTH463	~3.7 kb <i>SalI/SspI P_{tet^rflp/trfA}</i> cassette from pJW106 into pBBR1MCS-5 (<i>SmaI/SalI</i>)	
pTH466	~0.3 kb <i>EcoRI/XbaI</i> PCR product carrying the <i>pcaD</i> promoter into pUC118	
pTH468	~0.3 kb <i>EcoRI/XbaI P_{pcaD}</i> fragment from pTH466 into pMP220	
pTH470	~1.7 kb <i>XbaI/NheI flp</i> fragment from pLD3 into <i>XbaI</i> site of pBBR1MCS-5 (such that <i>flp</i> may be under <i>P_{lac}</i> control)	
pTH471	~1.7 kb <i>XbaI/NheI flp</i> fragment from pLD3 into <i>XbaI</i> site of pBBR1MCS-5 (opposite orientation)	
pTH472	~1.7 kb <i>XbaI/NheI flp</i> fragment from pLD3 into <i>XbaI</i> site of pTH468 (such that <i>flp</i> may be under <i>P_{pcaD}</i> control)	
pTH474	~2.0 kb <i>PstI P_{pcaD}flp</i> cassette from pTH472 into pBBR1MCS-5 (such that <i>flp</i> may not be under <i>P_{lac}</i> control)	
pTH480	~1.4 kb <i>PstI/XhoI</i> PCR product of the <i>flp</i> gene into <i>PstI/SalI</i> of pBS KS	
pTH482	~60 kb vector including the <i>FRT</i> -encompassed region in RmH994, Ap ^r Sp ^r	
pTH483	~260 kb vector including the <i>FRT</i> -encompassed region in RmK137, Ap ^r Sp ^r Nm ^r Tc ^r	
pTH484	~250 kb vector including the <i>FRT</i> -encompassed region in RmK140, Ap ^r Sp ^r	
pTH486	~150 bp <i>BamHI/NlaIII P_{pcaD}</i> PCR product into <i>BamHI/SphI</i> of pTH480	
pTH502	pTH456 with a <i>SacI</i> linker in the <i>SalI</i> -digested and Klenow-treated <i>SalI</i> site	Finan laboratory
pTH504	two ~4.5 kb <i>SacI</i> cassettes (<i>IS50-oriV/FRT-ΩSp^r-oriT</i>) from pTH502 inserted in parallel into pBACe3.6	Finan laboratory
pTH506	~1.55 kb <i>XbaI/XhoI P_{pcaD}flp</i> cassette from pTH486 into pBBR1MCS-5	
pTH509	~4.5 kb <i>SacI IS50-oriV/FRT-ΩSp^r-oriT</i> cassette from pTH502 into pBACe3.6	Finan laboratory
pTH510	~270 kb vector including the <i>FRT</i> -encompassed region in RmK173, Sp ^r	
pTH514	~65 kb vector including the <i>FRT</i> -encompassed region in RmK189, Ap ^r Sp ^r Nm ^r -Km ^r Tc ^r	
pTH515	~55 kb vector including the <i>FRT</i> -encompassed region in RmK190, Sp ^r	
PTH518	~65 kb BAC vector including the <i>FRT</i> encompassed region in RmK203, Cm ^r Sp ^r	
PTH519	~150 kb BAC vector including the <i>FRT</i> encompassed region in RmK198, Cm ^r Sp ^r	

pTH520	~3.6 kb <i>EcoRI</i> fragment from pTH482 into pTC197
pTH521	~7.8 kb <i>EcoRI</i> fragment from pTH482 into pTC197
pTH522	~2.0 kb <i>EcoRI</i> fragment from pTH482 into pTC197
pTH523	~1.3 kb <i>EcoRI</i> fragment from pTH482 into pTC197
pTH524	~1.4 kb <i>EcoRI</i> fragment from pTH482 into pTC197
pTH525	~3.5 kb <i>EcoRI</i> fragment from pTH482 into pTC197
pTH526	~4.8 kb <i>EcoRI</i> fragment from pTH482 into pTC197
pTH527	~0.35 kb <i>EcoRI</i> fragment from pTH482 into pTC197
pTH528	~4.5 kb <i>EcoRI</i> fragment from pTH482 into pTC197
pTH529	~1.45 kb <i>EcoRI</i> fragment from pTH482 into pTC197
pTH530	~6.1 kb <i>EcoRI</i> fragment from pTH482 into pTC197

Transposons

Tn5	Nm ^r -Km ^r , Sm ^r	Berg and Berg 1987
Tn5-132	Ot ^r	Berg and Berg 1987
Tn5-233	Gm ^r /Sp ^r	De Vos <i>et al.</i> 1986
Tn5-235	Tn5 containing <i>E. coli lacZ</i> , Nm ^r -Km ^r	De Vos <i>et al.</i> 1986
Tn5-235 <i>FRT</i>	Tn5-235 with <i>FRT</i> sites inserted in parallel within the <i>BglII</i> sites of the IS50s, Nm ^r -Km ^r	
Tn <i>V</i>	Tn5 containing pSC101 <i>oriV</i> , Nm ^r -Km ^r	Furuichi <i>et al.</i> 1985

Unless indicated otherwise, all strains, plasmids and transposons were constructed in this study. ^r designates resistance, ^s designates sensitivity. The symbol Φ preceding a strain name indicates a ΦM12 transducing lysate. For strain constructions, an arrow (→) indicates transduction from the ΦM12 transducing lysate.

Abbreviation are as follows: Ap, ampicillin; Cm, chloramphenicol; Gm, gentamicin; Km, kanamycin; Nm, neomycin; Ot, oxytetracycline; Rf, rifampicin; Sm, streptomycin; Sp, spectinomycin; Tc, tetracycline; *FRT*, Flp recognition target; *oriT*, origin of transfer; *oriS* and *oriV*, origins of vegetative replication

* The construction of these *R. meliloti* strains was completed with the help of T. M. Finan and I. Hernandez-Lucas

Growth Media and Conditions

Antibiotics were stored as concentrated stock solutions at -20°C in H₂O, ethanol, ethanol/H₂O, methanol or dimethylformamide. Aqueous stock solutions were filter sterilized. The final concentrations for antibiotics in media are given in Table 3. 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) indicator was used in solid media at 50 µg/ml final concentration. Calcofluor-white (Cellufluor; Polysciences, Warrington, Pennsylvania) was used at 0.02% as described in Finan *et al.* (1985).

Table 3: Antibiotic Concentrations

Antibiotic	Concentration (µg/ml) for			
	<i>E. coli</i>		<i>R. meliloti</i>	
	solid media	liquid media	solid media	liquid media
Chloramphenicol	10-20	5-10	2.5-5	--
Chlortetracycline hydrochloride*	5-25	2.5-10	25	2-15
Kanamycin monosulfate	20	10	--	--
Gentamycin sulfate	5-10	2.5-5	20-40	10
Na-Ampicillin	100	50	--	--
Neomycin sulfate	--	--	200-400	100
Oxytetracycline hydrochloride	--	--	0.3-0.5	--
Rifampicin	20	10	20-40	10
Spectinomycin dihydrochloride	50-100	25-50	100	50
Streptomycin sulfate	100	50	200-400	100
Tetracycline hydrochloride	5	2.5	5	2.5

Antibiotic concentrations used for *A. tumefaciens* were identical to *R. meliloti* with one exception: the final concentration of oxytetracycline used in solid media for *A. tumefaciens* was 3 µg/ml.

* The chlortetracycline hydrochloride stock was heat-inactivated (autoclaved) before use.

R. meliloti and *A. tumefaciens* cells were incubated at 30°C while *E. coli* cells were incubated at 30° or 37°C. The complex medium used for the growth of bacteria was LB (Miller 1972): 10 g/l bacto tryptone, 5 gm/l yeast extract, 5 g/l NaCl, with the addition of 4 ml/l (liquid medium) or 1 ml/l (solid medium) of 1N NaOH. Solid media was prepared with the addition of 15g/l Difco bacto Agar. Liquid cultures of *R. meliloti* and *A. tumefaciens* were grown in LBmc: LB broth with 2.5 mM MgSO₄ and 2.5 mM CaCl₂. Minimal media or M9 salts, the defined media, was prepared from a commercially available premixed powder (Gibco-BRL). Carbon sources were added to final concentrations of 15 mM (glucose, succinate, arabinose) or 10 mM (protocatechuate). MgSO₄, CaCl₂ and biotin were also added to final concentrations of 2.5 mM, 1.25 mM and 1µg/ml respectively. For sufficient aeration of 5 ml liquid cultures, these were incubated on a rotary mixer (Rollordrum); for larger volumes in flasks, a rotary shaker was employed.

Bacterial Matings

Conjugations were performed by combining equal (0.5 ml) volumes of mid- to late-log phase cultures of donor, recipient(s) and, if necessary, mobilizing strains. These mobilizing strains carried the helper plasmids pRK2013 or pRK600 which provided the transfer functions *in trans*. The mixture was washed, resuspended in 50 µl LB and spotted onto LB plates which then were incubated overnight at 30°C. The spots were then resuspended in 1 ml of 0.85% NaCl and dilutions were plated onto selective media. Matings into *E. coli* were incubated overnight at 37°C, whereas matings into *A.*

tumefaciens and *R. meliloti* were incubated at 30°C for either 2 days (*A. tumefaciens*) or 3-4 days (*R. meliloti*). The resulting transconjugants were single-colony-purified 3 times by streaking onto selective media.

Preparation and Testing of Φ M12 Transducing Lysates

Lysates were prepared from log phase *R. meliloti* cultures, to which 50 μ l of Φ M12 phage (10^9 - 10^{11} PFU/ml) was added (Finan *et al.* 1984). This was incubated for 8 hours or overnight, at which time a few drops of chloroform were added to kill any remaining viable cells. Lysates were transferred to screw-cap tubes, centrifuged for a few minutes to pellet cell debris, and stored at 4°C. To determine the titre, in plaque forming units (PFU), lysates were titrated and 100 μ l of various dilutions (10^{-8} - 10^{-10}) were added to an overnight culture (OD_{600} of ~ 0.8) of Rm1021. After an adsorption time of 15 minutes, 2 ml of soft (0.5%) agar, cooled to 50°C, was added. This mixture was poured onto an LB agar plate and incubated overnight at 30°C.

Φ M12-Mediated Transduction

Transductions were performed as described by Finan *et al.* (1984) using bacteriophage Φ M12. 0.5 ml of the *R. meliloti* recipient culture at an OD_{600} of ~ 1.0 was mixed with 0.5 ml of diluted donor phage to give a multiplicity of infection of approximately 0.5. After 20 minutes incubation at 30°C and two washes in 0.85% NaCl, the cells were resuspended in ~ 0.5 ml. The cells were then plated onto appropriate selective medium (selecting for the antibiotic resistance of the marker being transduced) and incubated at 30°C for 3-5 days. The resulting transductants were single-colony-

purified 3 times by streaking onto selective media. For those markers close to one another, the distance between these two markers was estimated by translating the co-transduction frequencies into kb using Wu's formula (Wu 1966): $c = (1-d/l)^3$

where c = co-transduction frequency, d = distance between two markers, and l = length of the transducing particle. The length (l) is 160 kb for Φ M12 (Finan *et al.* 1984).

Transposon Replacements

The replacement of Tn5-233 with Tn5, Tn5-235 or Tn5-235*FRT* was accomplished via double homologous recombination between the flanking IS50 elements. Plasmids pRK602, pGS235 and pTH462 were mated into the *R. meliloti* Tn5-233 recipients. Following selection for neomycin resistant transconjugants, these colonies were then screened for loss of gentamycin and spectinomycin resistance genes, carried by the original Tn5-233 insertion, to distinguish true replacements from cointegrates. The true replacements were single-colony-purified 3 times by streaking onto selective media.

DNA Methodology

Small and large scale preparations of plasmid DNA were performed with the alkaline lysis method as per Sambrook *et al.* (1989). Large scale preparations were further purified using the LiCl, PEG/NaCl precipitation procedure. Restriction endonucleases and other DNA-modifying enzymes were used according to manufacturers recommendations. Following preparation and digestion/modification, DNA samples were electrophoresed through 0.8% to 2% agarose gels in TAE (De Picker *et al.* 1980) buffer. The gels were stained with ethidium bromide and the DNA visualized under

ultraviolet illumination. If any particular bands were desired for subsequent cloning, they were excised and eluted using a GeneClean II kit (BIO101). Such samples were usually resuspended in 10-20 μl of $T_{10}E_1$. Ligations were normally carried out overnight at room temperature in final volumes of 10-20 μl . Constructs were either transformed into chemically-competent cells (Ausubel *et al.* 1989; Sambrook *et al.* 1989) or into electro-competent cells (Sambrook *et al.* 1989).

Preparation of Competent Cells

E. coli cells were made competent by using modified calcium chloride procedures such as the following. The cells of a 200 ml culture (OD_{600} of 0.4) were centrifuged at 4000 rpm for 10 minutes at 4°C, using a Beckman GPR centrifuge, GA-10 rotor. The supernatant was then discarded, the cells resuspended in approximately 25 ml of 50 mM CaCl_2 and 20 mM KAc pH 6.2, and left on ice for 1 hour. After the cells were pelleted at 4000 rpm for 10 minutes at 4°C, the supernatant was discarded. The combined cellular pellet was then resuspended in 10 ml of 50 mM CaCl_2 and 20 mM KAc pH 6.2, 20% glycerol. These competent cells were aliquoted and stored at -80°C.

Transformation of Competent Cells

Between 50 and 100 μl of competent cells were thawed and gently mixed with ≤ 1 μg of DNA in a microfuge tube and placed on ice for 30-60 minutes. The cells were then heat shocked by incubating the tubes at 42°C for 90 seconds, and briefly placed on ice. One ml of LB was then added and the cells returned to 37°C for 1 to 2 hours, allowing expression of plasmid-encoded antibiotic resistance. Aliquots of 100 μl were plated onto

selective medium, while the supplementary cells were centrifuged for 2-3 minutes. After the excess supernatant was removed, the remaining cells were also plated onto the same selective medium and were incubated overnight at 37°C.

Preparation of Electro-Competent Cells

E. coli and *R. meliloti* electro-competent cells were prepared by centrifuging 200 ml cultures, grown to an OD₆₀₀ of 0.4, at 4000 rpm for 10 minutes at 4°C, using a Beckman GPR centrifuge, GA-10 rotor. The harvested cells were washed twice in 25 ml ddH₂O and then once in 25 ml of a 10% glycerol solution. The combined cellular pellet was resuspended in approximately 1 ml of 10% glycerol. These electro-competent cells were aliquoted and stored at -80°C.

Electroporation of Electro-Competent Cells

Between 40 and 70 µl of electro-competent cells were gently mixed with ≤1 µg of DNA in a chilled Gene Pulser/*E. coli* Pulser Cuvette with 0.2 cm electrode gap from Bio-Rad (Catalog No. 165-2086) such that the mixture covered the bottom of the cuvette. These cuvettes were placed in the *E. coli* Pulser Apparatus from Bio-Rad (Catalog No. 165-2101), with a voltage pulse preset at 2.25 V. After the pulse (indicated by a beeping sound), 1 ml of LB was added, and the cells were transferred to a microfuge tube. The cells were incubated for 1 to 2 hours, allowing expression of plasmid-encoded antibiotic resistance, and plated onto selective medium. *E. coli* was incubated overnight at 37°C, while *R. meliloti* was incubated at 30°C for 3 days.

Pulsed-Field Gel Electrophoresis

Pulsed-field gel electrophoresis (PFGE) is a technique for resolving very large (≥ 50 kb) size DNA molecules. PFGE was conducted using the Bio-Rad CHEF-DR II (catalog No. 170-3612 and accompanying cooling module).system based on the CHEF (clamped homogeneous electric fields) technique (Carle and Olson 1984). The methodology used for PFGE was modified from Sobral and Atherly (1989), Sobral *et al.* (1991a), Sobral *et al.* (1991b) and Honeycutt *et al.* (1993), as well as from Bio-Rad and New England Biolabs (NEB) protocols. All solutions were made according to Bio-Rad's CHEF Bacterial Genomic DNA Plug Dit (Catalog No. 170-3592) or NEB's Imbed Kit (Catalog No. 375).

To prepare agarose-embedded DNA (DNA or agarose "plugs"), a 5 ml *R. meliloti* culture, OD₆₀₀ of ~ 1 , was pelleted, washed once with T₁₀E₁ pH 8.0, 0.1% sodium lauryl sarcosine, and again with T₁₀E₁. The pellet was resuspended in cell suspension buffer (10 mM Tris pH 7.2, 20 mM NaCl, 50 mM EDTA) and melted 2% Low Melting Point Agarose (Bio-Rad Catalog No. 170-3594) (cooled to 50°C) in a 1:1 ratio to a final concentration of $\sim 5 \times 10^8$ cells/ml. This was transferred to plug molds and allowed to solidify at 4°C for 10-15 minutes. The agarose plugs were then removed and placed in a 50 ml Falcon tube submerged in Lysozyme solution (10 mM Tris, pH 7.2, 50 mM NaCl, 0.2% sodium deoxycholate, 0.5% sodium lauryl sarcosine, 1 mg/ml lysozyme) at 37°C between 2 hours and overnight. The agarose plugs were then rinsed with sterile water and immersed overnight or up to 2 nights in Proteinase K solution (100 mM EDTA, pH

8.0, 0.2% sodium deoxycholate, 1% sodium lauryl sarcosine, 1mg/ml proteinase K) at 50°C. The plugs were washed once in Wash Buffer ($T_{20}E_{50}$, pH 8.0) and then washed in a phenyl-methanesulfonyl fluoride (PMSF) solution (0.174 g PMSF in 10 ml 100% isopropanol) for 1 hour each. The agarose plugs were washed twice more in Wash Buffer for 1 hour each, for 1 hour in 0.1X Wash Buffer and resuspended in 0.1X Wash Buffer. These plugs can be stored for up to 4 months.

The plugs were then restricted with *I-SceI* (Boehringer Mannheim, Catalog No. 1497 235) and *SpeI*. *I-SceI* digests were conducted as per manufacturer's instructions, whereas *SpeI* digests were carried out as follows. One plug was incubated with 1X enzyme (*SpeI*) buffer and shaken for about 1 hour at room temperature. In 300 μ l fresh 1X enzyme buffer, 50 units enzyme was added for each 100 μ l of agarose plug. This was incubated overnight at the optimal temperature (37°C). The buffer was removed the next day and 1 ml of 1X Wash Buffer was added and incubated with the plug for 30 minutes. This buffer was removed and replaced with 1 ml of the gel running buffer (0.5X TBE). The digested plugs can be stored at 4°C for several days.

Plug samples (between 1/5 and 1/4) were loaded per well on 1% Pulsed-Field Certified Agarose (Bio-Rad, Catalog No. 162-0137) gels and the wells filled with Low Melting Point Agarose. The gels were then submersed in 0.5X TBE buffer. The MidRange I Marker from New England Biolabs (Catalog No. 355-1) was used as size standards. Run conditions were ~6.0 Volts/cm (200 Volts) using a 120° included angle

with a 2-25 second linear switch time ramp (pulse parameters) at 14°C for 24 hours. The gels were then stained with EtBr and treated like other agarose gels.

Isolation of Total Genomic DNA from *R. meliloti*

In a procedure adapted from Meade *et al.* (1982), 3 to 5 ml of overnight *R. meliloti* cultures, once pelleted, were washed once with 0.85% NaCl, once with TES (10 mM Tris-HCl pH 8, 25 mM EDTA, 150 mM NaCl) and were resuspended in 2.5 ml of T₁₀E₂₅. After adding 250 µl of lysozyme solution (2 mg/ml lysozyme in T₁₀E₂₅), the mixture was incubated at 37°C for 15 minutes. 300 µl of sarkosyl-protease solution (5 mg/ml protease E [Sigma] in T₁₀E₂₅, autolysed for 2 hours at 37°C, and 10% sodium sarkosinate) was then added and the mixture incubated at 37°C for 1 hour. The resulting lysate was extracted several times, first with phenol, then phenol/chloroform, and finally chloroform. After the addition of 150 µl of 5 M ammonium acetate, the DNA was precipitated by adding 0.6 volume of isopropanol and mixing by inversion. The precipitated DNA was either collected by centrifugation or picked up using a sterile glass pasteur pipette. The DNA was washed with 70% ethanol, dissolved in 200-700 µl T₂₀E₁, and stored at -20°C.

Southern Transfer and Hybridization

The DNA templates to become probes were labeled with digoxigenin (DIG) using and following Boehringer Mannheim's DIG DNA Labeling and Detection Kit (Catalog No. 1093 657). Plasmid or genomic DNA fragments were separated by agarose gel electrophoresis. After the gel was stained and photographed, it was washed for 10

minutes in a depurination solution of 250 mM HCl, followed by two 15 minute denaturation washes in a 0.5 M NaOH, 1.5 M NaCl solution, and two more 15 minute washes in a neutralization buffer of 0.5 M Tris-HCl pH 7.5, 3 M NaCl, and rinsing with H₂O between each step. The denatured DNA was then transferred (Southern 1975) to nylon membrane by capillary transfer (Sambrook *et al.* 1989) using 20X SSC (3 M NaCl, 300 mM sodium citrate, pH 7.0) for 12 hours or more. The transferred DNA was then cross-linked to the membrane by baking it at 80°C for 1 to 2 hours.

The resulting blot was then submerged for 2 hours or more at 65°C in a prehybridization solution (5X SSC, 0.1% w/v N-lauroyl-sarcosine, 0.02% w/v SDS, 1% blocking reagent). The prehybridization solution was then replaced with hybridization solution (prehybridization solution containing the DIG-labeled probe) and incubated with the blot overnight at 65°C. The membrane then underwent two 5 minute washes in 2X wash solution (2X SSC, 0.1% SDS) at room temperature, followed by two 15 minute washes in 0.5X wash solution (0.5X SSC, 0.1% SDS). Colorimetric detection with NBT (nitroblue tetrazolium) and BCIP (5-bromo-4-chloro-3-indolyl phosphate) was conducted according to Boehringer Mannheim's DIG DNA Labeling and Detection Kit.

Polymerase Chain Reaction

Polymerase chain reaction (PCR) primers were designed using the Gene Runner 3.04 program (Hastings Software, 1994). PCR primers (see Table 4) were obtained from the MOBIX facility and resuspended in T₁₀E₁ to 100 pmoles/μl. Using the GeneAmp PCR Core Reagents (Perkin Elmer), all PCR reactions were performed in a 100 μl final

Table 4: PCR and Sequencing Primers and Single Stranded Oligonucleotides

Name	Nucleotide Sequence (5' to 3')	Complement Region and Added Restriction Sites
AB10170	GCTCTAGAAGCGTCCTGAACGGAACCT TTCC	<i>Xba</i> I site; complement to bp 20 to 42 of IS50
AB10171	GGACTAGTTACGCTAGGGATAACAGG <u>GTAATTGATCGCCTCGGCAGAAACGT</u> TG	<i>Spe</i> I and internal I- <i>Sce</i> I sites; complement to bp 292 to 314 of IS50
AB10352	GATCGAAGTTCCTATACTCTAGTAAG AATAGGAACCTTCGAATTC	<i>Eco</i> RI site; combined with AB10353 makes the minimal <i>FRT</i> site with <i>Bg</i> III sticky ends
AB10353	GATCGAATTCGAAGTTCCTATTCTTAC TAGAGTATAGGAACCTC	<i>Eco</i> RI site; combined with AB10352 makes the minimal <i>FRT</i> site with <i>Bg</i> III sticky ends
AB11511	TGTTGTGCCAGTCATAGCCGAATAG CC	complement to bp 95 to 123 of inside end of Tn5 on the Nm' side
AB11512	ATGTGCTGCAAGGCGATTAAGTTGGG TAAC	complement to bp 130 to 160 from <i>Xba</i> I site in pTH455 on the backbone (pBS) side
AB11513	GCGTGTCTTGGGAGATTGGACGACA GC	complement to bp 40 to 67 of inside end of Tn5 on the Sm' side
AB11514	TTTCTAAGGCAGACCAACCATTGTGA AATCAG	complement to bp 155 to 188 of inside end of Tn5-132 on <i>Eco</i> RI side
AB11515	TTCTCGAACCCTCCCGGCCGCTAACG	complement to bp 110 to 137 from <i>Xba</i> I site in pTH456, in <i>oriV</i>
AB11516	TTCAGTGATCCATTGCTGTTGACAAAG GGAATC	complement to bp 40 to 173 of inside end of Tn5-132 on <i>Xba</i> I side
AB11717	GCTCTAGAACGACATCGCGCCAGATAC GGAAATC	<i>Xba</i> I site; complement to bp 99 to 125 from the predicted <i>pcaD</i> translational start
AB11718	GGAATTCGCATGGCGCAGAAAGACCTC ACCGTAG	<i>Eco</i> RI site; complement to bp -282 to -309 from the predicted <i>pcaD</i> translational start
AB13241	AACTGCAGCATGCCACAATTTGGTATA TTATGTAAAACAC	<i>Pst</i> I and internal <i>Sph</i> I sites; complement to bp 1 to 30 of 1272 bp <i>flp</i> gene on pLD3
AB13242	CCGCTCGAGCACATACAGCTCACTGTT CACGTC	<i>Xho</i> I site; complement to bp 1377 to 1404 downstream of <i>flp</i> on pLD3
AB13243	CCGCATGGCAGCTCTCCTCCGTCGTCA GATTTG	<i>Nla</i> III site; complement to bp -1 to -27 from the predicted <i>pcaD</i> translational start
AB13244	CGGGATCCGCTCTTCTGGCGTGCGACC TCGAC	<i>Bam</i> HI site; complement to bp -135 to -159 from the predicted <i>pcaD</i> translational start
M13 forward	GTA AAAACGACGCCAGT	common sequencing primer
M13 reverse	AACAGCTATGACCATG	common sequencing primer

All restriction sites are italicized; the I-*Sce*I site is underlined; the *Bg*III compatible cohesive ends are in bold.

volume, consisting of 200 μ M each dNTP, 1X PCR Buffer II, 1 μ M each primer and 2.5 units of Taq polymerase with either \sim 1 ng plasmid DNA or \sim 10 ng genomic DNA. A master mix containing PCR buffer, dNTPs, the two primers, Taq polymerase and ddH₂O was prepared and added to PCR tubes containing 5 μ l of template DNA. Because the optimal MgCl₂ concentration varies for each primer-template set, MgCl₂ was then added to final concentrations of 0 to 5 mM in 0.5 mM increments. The *GeneAmp* PCR System 2400 and small modifications of its default incubation cycle were used for all PCR reactions. Between 25 and 35 cycles of amplification were used with a preliminary 5 minute 94°C incubation period and a final 7 minute 72°C incubation period. The cycles consisted of melting (94°C), annealing and extension (72°C) periods of 30 seconds each. The annealing temperature was set between 55°C and 65°C, normally \sim 5°C below the lowest melting temperature (T_m) of a primer pair. The PCR reactions were tested by loading and running 10 μ l of each sample on an agarose gel. If the PCR product was to be cloned, the remaining 90 μ l of the samples were ethanol precipitated and the DNA digested with appropriate restriction endonucleases.

DNA Sequencing

Most of the sequencing was performed using the M13 forward and reverse primers (see Table 4), obtained from MOBIX. All other sequencing primers were resuspended in T₁₀E₁ to 100 pmoles/ μ l. 4 μ l of sequencing primer and 1 μ g of template DNA were used per sequencing reaction. DNA sequencing was performed by the Mobix

facility on an ABI 373 Stretch automatic sequencer using dye terminator chemistry and cycle sequencing.

Gene Runner 3.04 (Hastings Software, 1994) was used for general sequence analysis (e.g. restriction site analysis). The programs ALIGN 2.0 (Myers and Miller 1988) and ClustalW (Thompson et al. 1994) were used to align nucleotide and amino acid sequences. All GenBank searches to compare nucleotide sequences against those in databases at the National Center for Biotechnology Information (NCBI) were accomplished using BLAST 2.0 programs (Altschul et al. 1997) available through the NCBI home page <http://www.ncbi.nlm.nih.gov>.

CHAPTER 3. RESULTS

3.1: *FRT* Targeting Vectors

Cloning Strategy for the Targeting Plasmids

The *in vivo* excision and amplification of large segments of bacterial genomes was suggested by Pósfai *et al.* (1994) and Wild *et al.* (1996). This scheme proposed that a specific region of interest (50-100 kb) be flanked by *F*lp recognition target (*FRT*) sites in direct (parallel) orientation so that excision occurred when the *F*lp recombinase was added *in trans*. In the study presented here, synthetic *FRT* sites were directed, via homologous recombination, to specific locations in the *R. meliloti* pExo megaplasmid. Using the genetic map of the pExo megaplasmid (Charles and Finan 1990) (refer to Figure 1), *FRT* sites can be targeted to specific Tn5-derivative insertions, thus circumventing the problem of having to design new targeting vectors for each pExo region to be excised.

Our strategy required the construction of only two targeting vectors, both of which contained a portion from the IS50 element, the Ω Sp interposon, the *FRT* site, the *oriT* and the *oriV* sites of plasmid RK2. Since these vectors were to be recombined via homologous recombination into Tn5 and Tn5-132 insertions in the pExo megaplasmid, a portion of the IS50 was included. The only difference between the two targeting vectors

is the orientation of the IS50 portion (carrying the I-SceI site) with respect to the rest of the plasmid. The *oriT* (origin of transfer) of RK2 allowed transfer of the targeting vectors into *R. meliloti* strains, and later, allowed the transfer of the pExo region of interest into an *E. coli* or *A. tumefaciens* recipient strain. The Ω Sp interposon, which carries the spectinomycin resistance gene, allowed selection of the initial recombination event in *R. meliloti*, as well as for subsequent selection of the transfer of the pExo region of interest into *E. coli* or *A. tumefaciens*. A cassette carrying the origin of replication of RK2 and the Flp recognition sequence (*oriV/FRT* cassette) from pMS107-GENO14 was kindly provided by J. Wild and W. Szybalski. A site for the extremely rare-cutting I-SceI meganuclease was also cloned into the targeting vectors to permit separation and purification of the desired pExo region from the targeting vector.

The detailed cloning scheme is outlined in Figures 3 and 4. The ColE1 vector pBluescript was chosen as the plasmid backbone for the targeting vectors due to the many restriction sites (18) within the multiple cloning site and also because it cannot replicate in *R. meliloti*. For ease in subsequent clonings, the *oriT*, *oriV/FRT*, and IS50 portion were first individually cloned into pBluescript (pBS KS or pBS SK) to allow for easy screening by blue/white colonies on LB plates supplemented with X-gal.

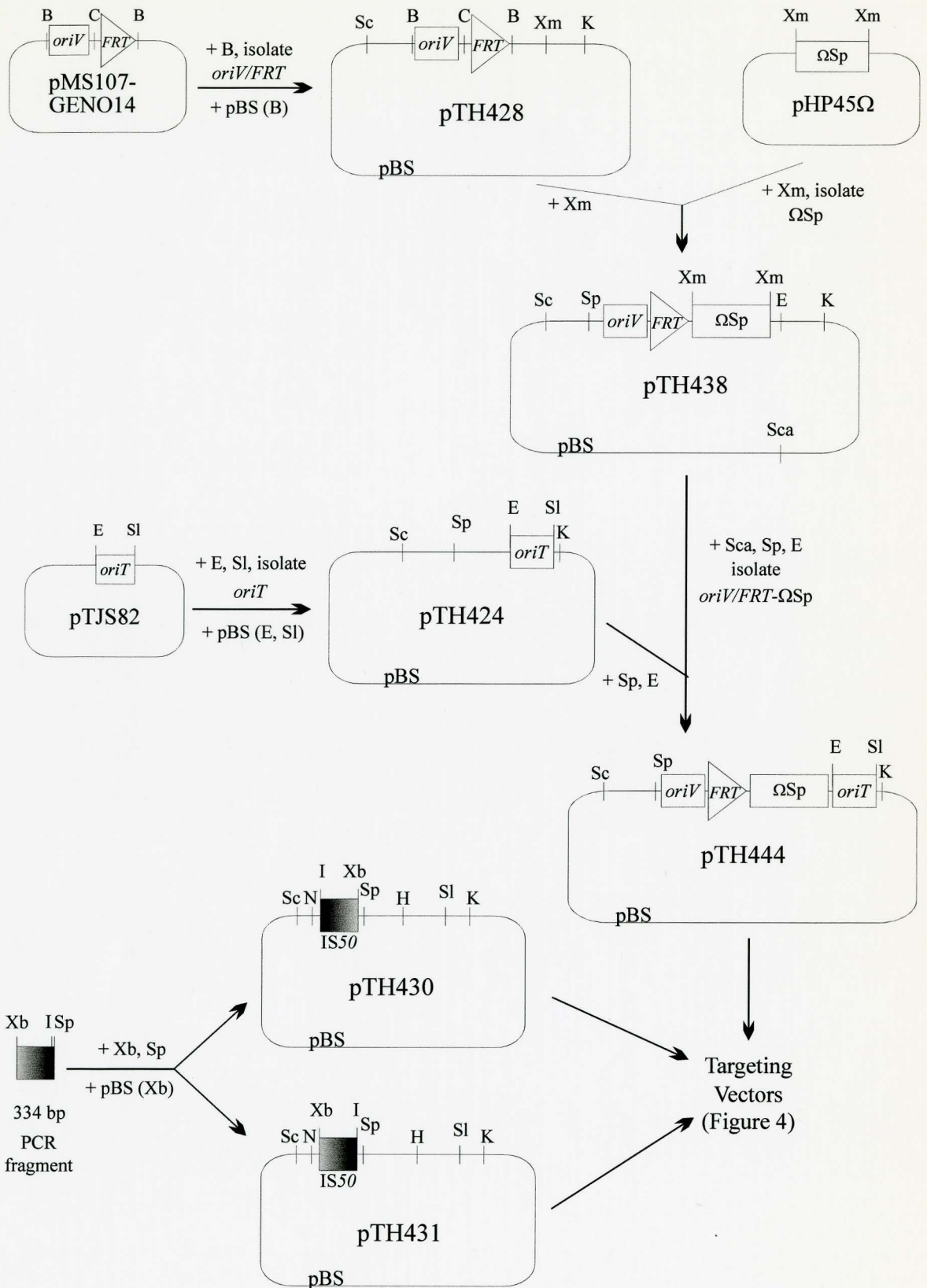
Description of Construction

The 760 bp *EcoRI/SalI* fragment carrying the *oriT* site of RK2 was isolated from pTJS82 (Schmidhauser and Helinski, 1985) and cloned into the same sites of pBS SK giving plasmid pTH424. The *oriV/FRT* cassette was cloned as a 900 bp *BamHI* fragment from plasmid pMS107-GENO14, into the *BamHI* site of pBS KS to give pTH428. The

position of the *FRT* site was determined by *Cla*I digestion and subsequent sequencing (using M13 forward and reverse primers—data not shown) to be adjacent to *Kpn*I site of the pBS multiple cloning site (diagrammed in Figure 3).

A 334 bp *IS50* PCR product, which included base pairs 20 to 314 starting from the outside end of the *IS50* as well as sites for I-*Sce*I, *Spe*I and *Xba*I restriction endonucleases, was synthesized using synthetic primers AB10170 and AB10171. Primer AB10170 (*IS50* nucleotide 20 to 42) was constructed with an *Xba*I site (italicized) at the 5' end: 5'-GCTCTAGAAGCGTCCTGAACGGAACCTTTCC-3'. Primer AB10171 (*IS50* nucleotide 292 to 314) carried the aforementioned I-*Sce*I site (underlined), as well as an *Spe*I site (italicized), both at the 5' end of the primer: 5'-GGACTAGTTAC GCTAGGGATAACAGGGTAATTGATCGCCTCGGCAGAAACGTTG-3'. The PCR template used for these two primers was the Tn5-bearing plasmid pGS220 (De Vos *et al.* 1986). The resulting PCR product was isolated from an agarose gel, digested with *Spe*I/*Xba*I and ligated to *Xba*I-restricted pBS KS. Since *Spe*I and *Xba*I digests leave identical overhangs, the *IS50* PCR product could insert into the vector in both orientations but when the *Spe*I and *Xba*I sites are ligated together, neither site is recreated. The orientation of the PCR fragment could be identified by double digestions with *Xba*I/*Not*I and *Xba*I/*Hind*III and the I-*Sce*I site was tested by digestion with the meganuclease I-*Sce*I. Orientation of the PCR fragment was further confirmed by sequencing using the M13-forward and M13-reverse primers (data not shown). The two resulting plasmids are pTH430 (outside end of the *IS50* is closest to *P_{lac}* and the *Kpn*I site) and pTH431 (opposite orientation to pTH430) (see Figure 3).

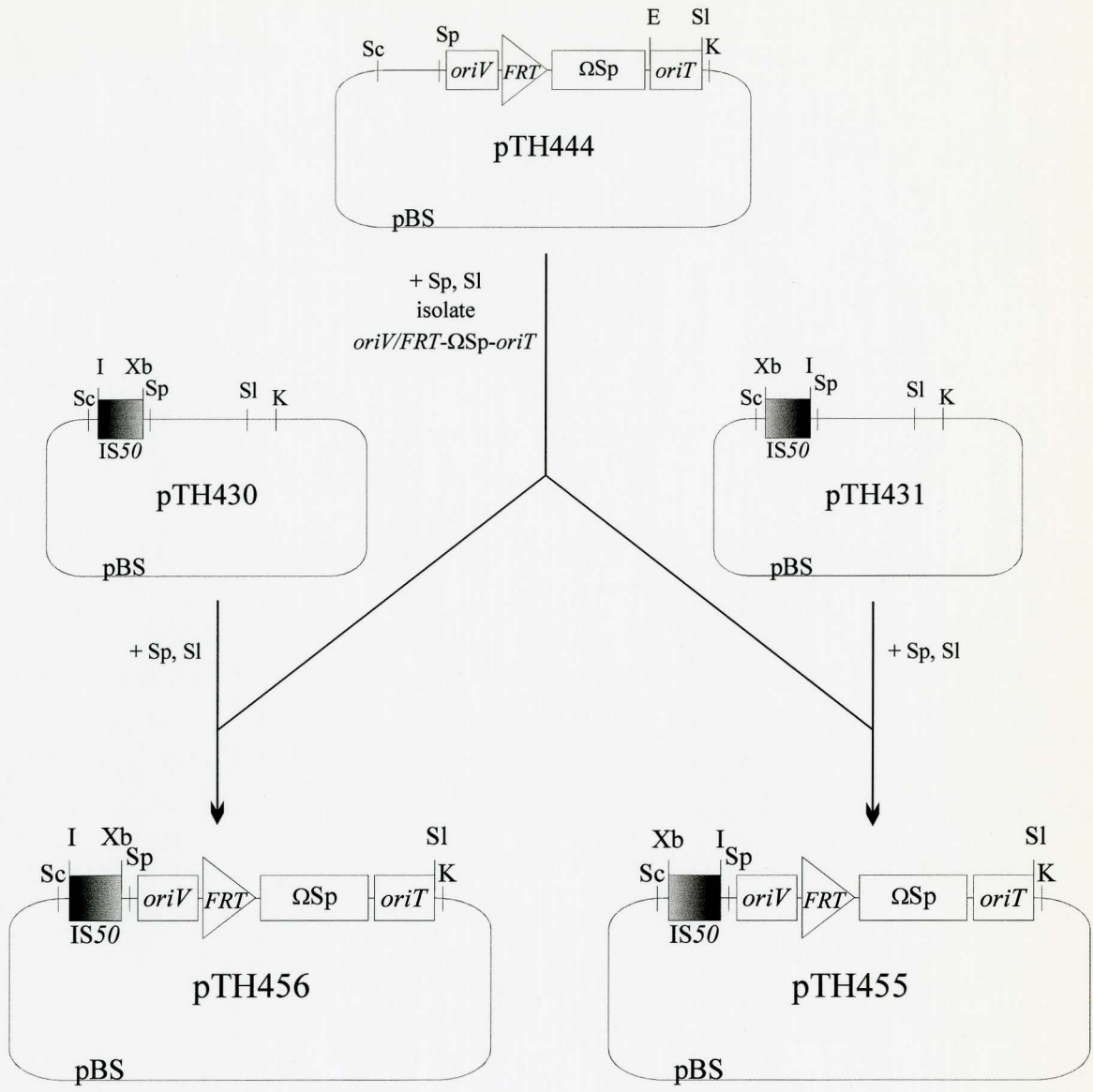
Figure 3: Diagram demonstrating the cloning scheme used to construct the targeting vectors pTH455 and pTH456 (continued in Figure 4). The triangle representing *FRT* indicates directionality. The *IS50* PCR product is shaded white (outside end of the *IS50*) to black (inside end of the *IS50*), indicating directionality. Relevant restriction sites are indicated and steps in the constructions are indicated alongside the arrows. Abbreviations are as follows: B, *Bam*HI; C, *Cl*aI; E, *Eco*RI; H, *Hind*III; I, *I-Sce*I; K, *Kpn*I; N, *Not*I; Sc, *Sac*I; Sca, *Sca*I; Sl, *Sal*I; Sp, *Spe*I; Xb, *Xba*I; Xm, *Xma*I.



The Ω Sp cassette from pHP45 Ω (Prentki and Krisch 1984) was gel-isolated as a 2.0 kb *Xma*I fragment and cloned into the *Xma*I site of pTH428 (the pBS KS plasmid carrying *oriV/FRT*) by selecting for spectinomycin resistance (Sp^r), creating plasmid pTH438. Since the *oriV/FRT*- Ω Sp cassette (2.9 kb) was close in size to the pBS KS backbone (2.95 kb), pTH438 was digested with *Eco*RI/*Spe*I/*Sca*I in order to isolate the *oriV/FRT*- Ω Sp fragment (*Sca*I cleaves the pBS backbone), and the cassette was isolated from an agarose gel. Plasmid pTH444 was constructed by directionally ligating this cassette into the *Eco*RI/*Spe*I-digested *oriT*-bearing plasmid, pTH424. The new 3.7 kb cassette, carrying *oriV/FRT*- Ω Sp-*oriT*, was then excised from pTH444 with a *Sal*I/*Spe*I digest and gel-isolated. This cassette was then cloned into both *Sal*I/*Spe*I-digested pTH431 and pTH430, producing the two targeting vectors pTH455 and pTH456, respectively. These two plasmids differ from one another only in the directionality of the IS50/I-*Sce*I PCR fragment (see Figure 4).

The strategy to build these targeting vectors had taken into account the location and orientation of the various DNA components after their integration into predetermined Tn5 and Tn5-132 insertions on the pExo megaplasmid. The targeting vector, pTH455, was directed to Tn5 insertions, whereas pTH456 was targeted to Tn5-132 insertions. Using this strategy, when two such targeting events are combined on the pExo megaplasmid in the same strain, and the *FRT* sites are recombined in the IS50 sequences closest to one another, these *FRT* sites, as well as the entire plasmid (except the IS50 portion), are in parallel orientation (see Figure 6).

Figure 4: Diagram demonstrating the cloning scheme used to construct the targeting vectors pTH455 and pTH456 (continued from Figure 3). The triangle representing *FRT* indicates directionality. The *IS50* PCR product is shaded white to black indicating directionality. Relevant restriction sites are indicated and steps in the constructions are indicated alongside the arrows. Abbreviations are as follows: E, *EcoRI*; I, *I-SceI*; K, *KpnI*; Sc, *SacI*; Sl, *Sall*; Sp, *SpeI*; Xb, *XbaI*.



Targeting Vectors

Directing the targeting vectors to specific regions of the pExo Megaplasmid

Transposon insertions were chosen for targeting, such that a Tn5 was 50 to 200 kb away from a Tn5-132. The chosen insertions were first transduced into Rm5000, a wild type rifampicin resistant (R^f) strain, to facilitate recipient *R. meliloti* selection and counter-selection against the donor *E. coli* when transferring the targeting vectors.

Since the targeting vectors will not replicate in *R. meliloti*, selection for spectinomycin resistance selects for integration of the targeting plasmid via homologous recombination between the cloned IS50 fragment and the IS50L or IS50R of the targeted transposon (illustrated in Figure 5A). The targeting vectors, pTH455 and pTH456, were mated into Rm5000 derivative strains carrying Tn5 or Tn5-132 transposon insertions respectively. As a negative control, conjugal transfer of the targeting vectors was also performed into Rm5000. The frequency of Sp^f transconjugants in the transposon-carrying strains was moderately low—between 10⁻⁶ and 10⁻⁷ per recipient—which was not entirely unexpected due to the small size (295 bp) of the IS50 homologous sequence. No transconjugants were obtained in wild type Rm5000, suggesting that aberrant integration of the targeting vectors, pTH455 and pTH456, into the *R. meliloti* genome is a rare event.

Obtaining Targeting Vectors in Both IS50 Elements

To ensure that the Sp^f of the targeting vector was 100% linked to the target transposon, phage lysates from selected RmH921-RmH924 and RmH942-RmH944 transconjugants (6 to 10 each) were used to transduce Sp^f into a wild type Rm1021

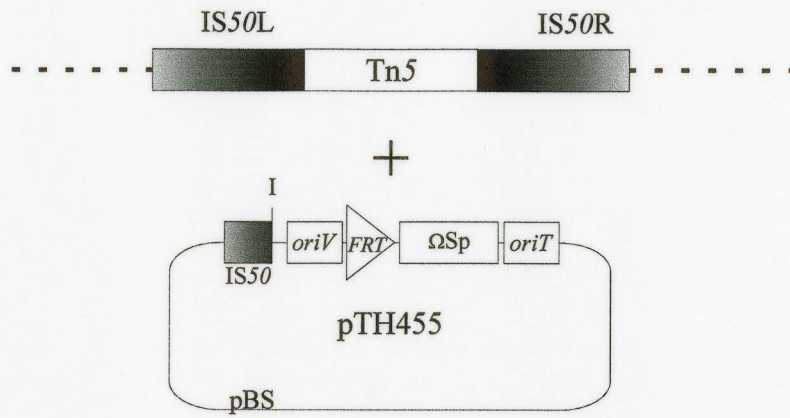
background. In all cases, the Sp^f marker was found to be 100% co-transduced with the target transposon antibiotic resistance marker (Nm^f or Ot^f). This, however, lends no clue as to which of the two *IS50* elements of the transposon (*IS50L* or *IS50R*) have been targeted in each transconjugant. To determine whether recombination into each *IS50* sequence for each transposon occurred, five or six randomly chosen recombinants from each targeting experiment were selected for analysis by PCR.

PCR primers were designed to amplify products of a predicted size depending on the *IS50* element (*IS50L* or *IS50R*) into which the targeting vector recombined (example in Figure 5B). These primers (refer to Table 4) are named: a (AB11511), b (AB11512) and c (AB11513), and x (AB11514), y (AB11515) and z (AB11516). PCR primers a, b, and c were used to differentiate between the two possible pTH455 recombinations in Tn5 transposons (see Figure 5B), whereas primers x, y, and z were used to distinguish between the two pTH456 recombination events. Therefore, in pairs (ab and bc; xy and yz), these primers amplify products depending on the location of insertion relative to the inside of the transposon. Based on amplification of a product with one pair, but not the other, the two possible recombinant “types” for each transposon were distinguished and purified.

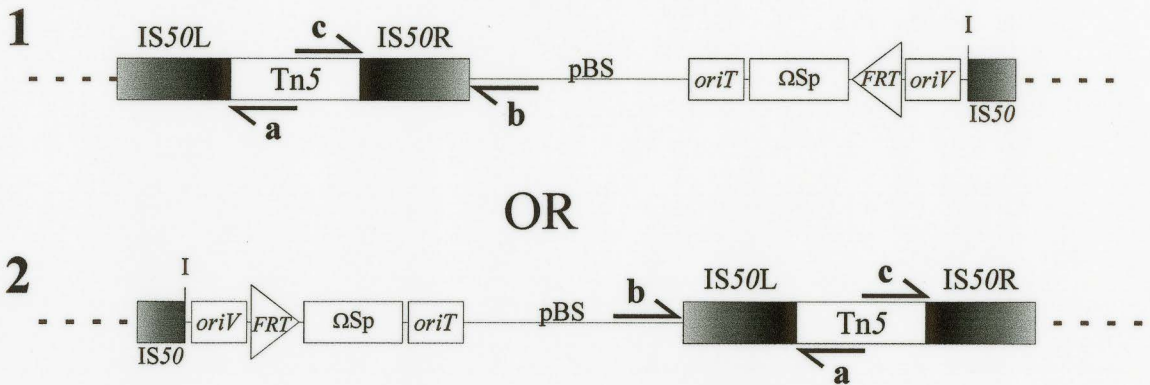
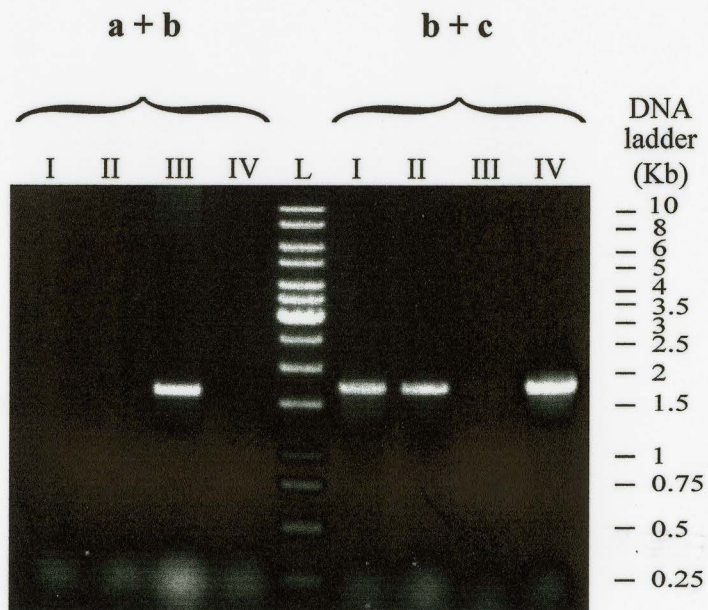
Construction of *R. meliloti* Strains Carrying Two Targeting Vectors

Phage Φ M12 lysates from Tn5::pTH455 strains were used to transduce the Tn5 neomycin resistance gene into Sp^f Tn5-132::pTH456 strains (illustrated in Figure 6). Nm^f transductants were subsequently screened for retention of Ot^f to ensure the Tn5-132

Figure 5: Schematic representation of homologous recombination between the targeting vector pTH455 and a Tn5 transposon, and agarose gel demonstrating both insertional events as determined by PCR. **Panel A.** Directionality of the *IS50* elements is indicated by the white to black shading (outside end to inside end of the *IS50*). Recombination of the targeting vectors can occur at either of the two *IS50* insertion sequences (*IS50R* or *IS50L*); the resulting possibilities are labeled, 1 and 2. The PCR primers are indicated by small half-arrows (**a**, **b**, and **c**). pExo DNA is represented by the dashed line. **Panel B.** Four genomic DNA preparations (I, II, III, IV) were used in PCR with primer sets **ab** and **bc**. When detecting integration of the targeting vectors, amplification of a product will be detected with one primer set only, as seen in Panel A. Sample III carries pTH455 in *IS50R*, whereas the others carry pTH455 in *IS50L*. Determination of pTH456 integration into both *IS50* elements can be accomplished in a similar fashion. **L** indicates the DNA ladder; sizes are indicated at the side of the figure.

A

Homologous recombination
at the IS50 elements

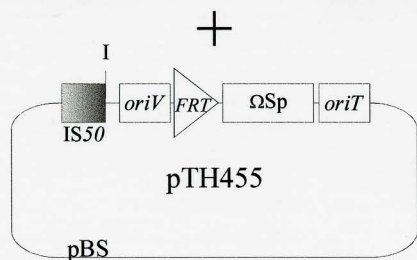
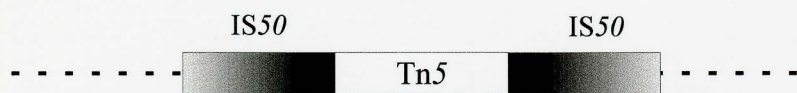
**B**

transposon was not lost by co-transduction. Given that pTH455 is 100% linked to the transposon Tn5 and that pTH456 is 100% linked to the Tn5-132 transposon, the resulting transductants carry two targeting plasmids and with this two copies of the *oriV/FRT-ΩSp-oriT* cassette (refer to Figure 6).

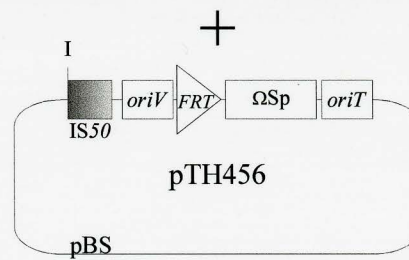
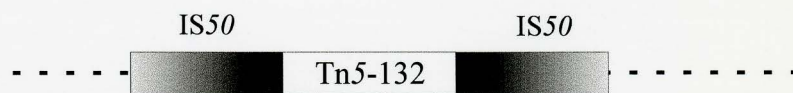
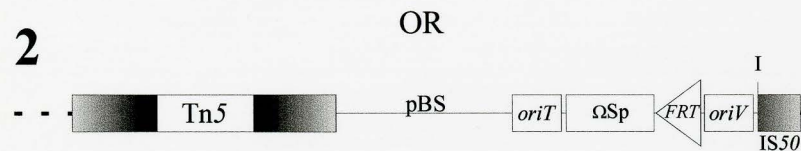
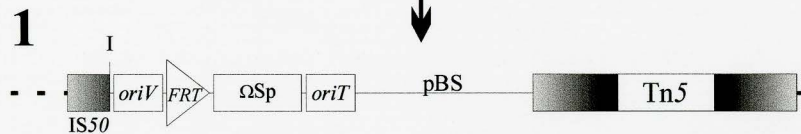
Because one of two possible recombination events can occur at each transposon, at IS50L or IS50R, there are four possible combinations after transduction (each of the two possible insertions in the Tn5 transposon with either of the two possible insertions in the Tn5-132 transposon) (see Appendix A). Most of the preliminary work was conducted on the combinations of $\Omega5056::Tn5::pTH455$ with $\Omega5069::Tn5-132::pTH456$, transposon insertions that lie approximately 50 kb away from one another. The distance between these two insertions was verified, as a co-transduction frequency of 33% was observed when combining the two together.

The four strains carrying combinations of targeted transposon insertions in these particular locations ($\Omega5056$ and $\Omega5069$) were designated RmH992, RmH993, RmH994 and RmH995. The targeting vectors were also directed to other locations on the pExo megaplasmid, and strains carrying various combinations “sets” were constructed (refer to Appendix A). These combinations were used to identify, with Flp-mediated recombination, which of the four strains excised the intervening region from the inside of the transposons, and which excised the intervening region along with the two transposons.

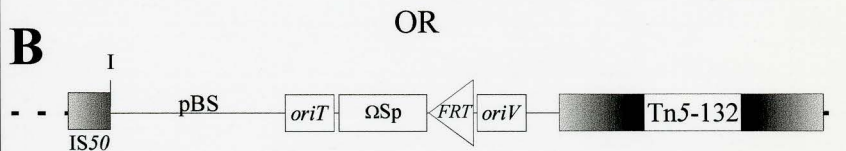
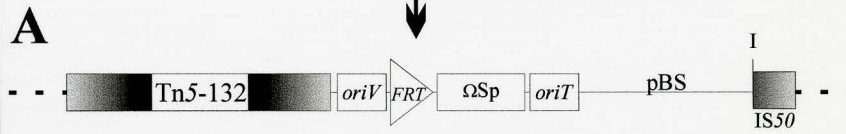
Figure 6: Schematic representation of two independent targeting events, the two possible outcomes for each event, and the combination of a pTH455-targeted Tn5 with a pTH456-targeted Tn5-132. Directionality of the *IS50* elements is indicated by the white to black shading. The two possible recombinations of pTH455 at the *IS50*s of Tn5 are numbered 1 and 2, whereas the two possible recombinations of pTH456 with Tn5-132 are labeled A and B. One of the four possible combination of two targeted transposons is illustrated in this figure by the transduction of $\Phi\text{M12}(2)\rightarrow(\text{B})$ [ie. $\Phi\text{M12}(\text{Tn5}::\text{pTH455})\rightarrow\text{Tn5-132}::\text{pTH456}$]. pExo DNA is represented by the dashed line.



Homologous recombination
at the IS50 elements



Homologous recombination
at the IS50 elements



Φ (2) \rightarrow B



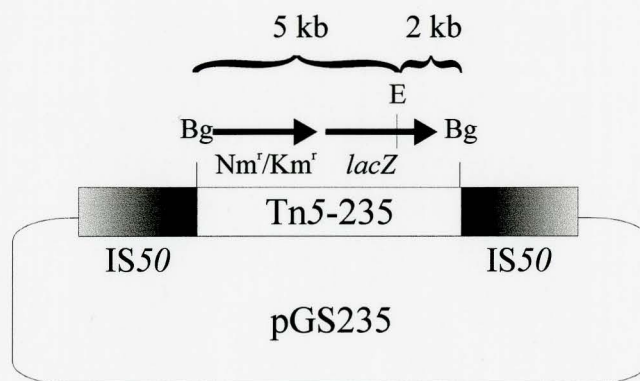
3.2: Flp Delivery Vectors

Construction of a Flp-testing strain of *R. meliloti*

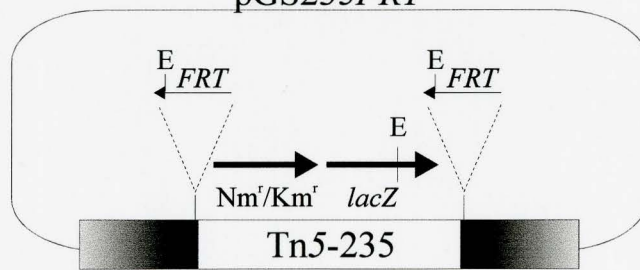
Successful excision of large regions of pExo, and their subsequent transfer to *E. coli*, is dependent upon successful expression and function of the *flp* gene in *R. meliloti*. To address these issues, a *R. meliloti* test system and several Flp vectors were constructed. An appropriate test for this project is a system that allows for relatively simple monitoring of excision from the megaplasmid.

Two single stranded oligonucleotides, AB10352 and AB10353 (kindly provided by F. Graham), were assembled together to make a minimal *FRT* site, flanked by sticky ends compatible with the cohesive ends of *Bam*HI and *Bg*III-digested DNA but that will not regenerate either site (see Figure 7A). Additionally, an *Eco*RI site was added at one of the ends to aid later in detection of the *FRT* site. The minimal *FRT* site was cloned into the two *Bg*III sites of the Tn5 derivative, Tn5-235, flanking the transposon's *Nm*^f and *lacZ* genes (see Figure 7A). This was accomplished by ligating the *FRT* oligonucleotide into *Bg*III-digested pGS235, a pBR322 derivative carrying Tn5-235 (De Vos *et al.* 1986). The ligation mixture was digested with *Bg*III, to cut any re-created *Bg*III sites, transformed into DH5 α , and plated on LB supplemented with ampicillin and X-gal. Plasmid DNA from approximately half of the 40 blue transformants that were selected showed the expected *Eco*RI and *Bg*III digest patterns (no *Bg*III sites remaining and the addition of two *Eco*RI sites—see Figure 7) (data not shown). This however, did not distinguish between the plasmids with parallel *FRT* sites or inverted *FRT* sites.

Figure 7: Diagram demonstrating the cloning of the minimal *FRT* oligonucleotide into the *Bgl*III sites of Tn5-235, and subsequent transfer of this construct into the *R. meliloti* genome by double homologous recombination at Ω 5033::Tn5-233. **Panel A.** The location of the neomycin resistance gene (Nm^r/Km^r) and the β -galactosidase gene (*lacZ*) are indicated by the large arrows. The plasmid pGS235 is about 12.4 kb in size. The minimal *FRT* sites and their orientation are indicated with the small arrows; they have *Bgl*III sticky ends, which do not regenerate a *Bgl*III site (Bg). *Eco*RI (E) and *Bgl*III restriction sites are indicated as well as the distances between them. **Panel B.** The large Xs represent the double recombination between the *IS50* elements of the plasmid-located Tn5-235*FRT* and the *R. meliloti* megaplasmid-located (Ω 5033) Tn5-233. The distance between the *Eco*RI sites present within the *FRT* oligonucleotides and the *Eco*RI site present in *lacZ* are the same as indicated between the *Bgl*III and *Eco*RI sites. The Gm^r and Sp^r resistance genes (large arrows) are replaced with the Nm^r and *lacZ* genes. pExo DNA is indicated with the dashed lines.

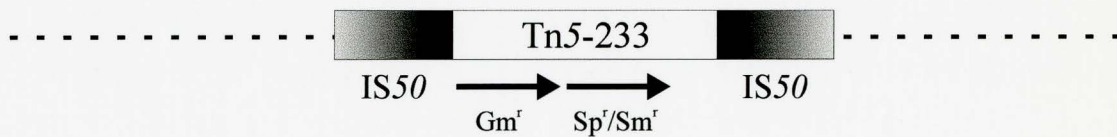
A

+

 $\xleftarrow{E} FRT$ Minimal *FRT* SitepGS235*FRT***B**

IS50

IS50

Homologous
recombination at
both IS50 elements

The Tn5-235*FRT* constructs were introduced into the pExo megaplasmid of the *R. meliloti* genome by transposon replacements at $\Omega 5033::\text{Tn5-233}$ (in RmG497) (see Figure 7B). Ten independent pGS235 constructs containing two *FRT* sites flanking the Tn5-235 and the parental plasmid pGS235 itself were mated into RmG497, selecting for Nm^{r} transconjugants. To distinguish between true replacements and cointegrates, 100 transconjugants were screened for loss of the Sp^{r} and Gm^{r} carried by the Tn5-233. True replacements were then transduced into a *lac*⁻ background (RmG212), so that the loss of the *lacZ* gene could be readily detected. The various Flp expression plasmids (see below) were transferred into these strains, transconjugants were selected, and Flp-mediated excision of the *lacZ* gene was screened for on LB containing gentamycin and X-gal. One of these tester strains, which appeared to have the minimal *FRT* sites in parallel (showed Flp-mediated excision), was saved as RmH940. The corresponding plasmid which gave rise to the strain RmH940, pTH462, was later analyzed by sequencing which confirmed that pTH462 carries one *FRT* site in either of the *Bg*III locations in parallel or “direct” orientation (data not shown).

Construction of the Initial Flp Vector

The synthesis of a functional Flp protein in *R. meliloti* was the primary objective in the construction of a *R. meliloti* Flp-delivery system. The vectors for Flp expression were designed as gentamycin resistant (Gm^{r}) plasmids capable of being transferred into and replicating in *R. meliloti*. The backbone pBBR1MCS-5 (Kovach *et al.* 1995), a Gm^{r} , broad host range, mobilizable cloning vector carrying the *lacZ* α fragment was suitable.

The ideal promoter for *flp* expression is one that would turn on transcription upon addition of an inducing metabolite or chemical. The Flp-inducible system used by Wild *et al.* (1996), in which *flp* and *trfA* are under the control of the tetracycline (*tet*) promoter/repressor, was initially chosen since Flp induction and activity had been proven with this delivery vector in *E. coli*. A 3.7 kb *SalI/SspI* cassette carrying *tetR-P_{tet}flp/trfA* was isolated from pJW106 (Wild *et al.* 1996), and cloned into *SalI/SmaI*-digested pBBR1MCS-5, selecting for white Gm^r colonies. The resulting vector, pTH463, was tested (see Table 5) with and without its inducer, heat-inactivated chlortetracycline (cTc). In these assays, cTc was added at concentrations ranging from 1 to 50 µg/ml as well as at different times during growth (added immediately upon inoculation, at an OD₆₀₀ of 0.2, 0.4, 0.6 and 0.8), and for various lengths of time (from 1 hour to overnight), as described by Pósfai *et al.* (1994) and Wild *et al.* (1996). Addition of plasmid pTH463, regardless of the presence of cTc, appeared to excise the *lacZ* gene in RmH940 at a frequency of less than 10% of transconjugants suggesting that a higher level of *flp* activity was required for efficient excision in *R. meliloti*. Although the Flp protein may not be active in *R. meliloti*, it was thought that the P_{tet} promoter was not functioning and therefore, not expressing *flp* to sufficiently high levels.

Table 5: Level of *flp* expression, measured in % white colonies

Plasmid	Plasmid Description	Number of Blue Transconjugants	Number of White Transconjugants	<i>flp</i> Expression (% White)
pTH463	P _{tet} <i>flp</i>	96*	5*	5 %
pTH471	<i>flp</i>	0	85	100 %
pTH474	P _{pcaD} <i>flp</i> †	0	98	100 %
pTH506	P _{pcaD} <i>flp</i> ‡	113	19	16 %

The results in this table are representative of several experiments. All plasmids were mated into control strains RmG212 (always white transconjugants) and RmH939 (always blue). Transfer frequencies for the Flp plasmids varied between 10⁻³ and 10⁻⁴ (per recipient).

* Many of these colonies were of a “mosaic” type displaying both blue and white sectors, they were classified as white when 50% or more of the colony was white.

† Designates the transcriptional fusion.

‡ Designates the translational fusion at the start site.

Expression of *flp* from a *R. meliloti* inducible promoter

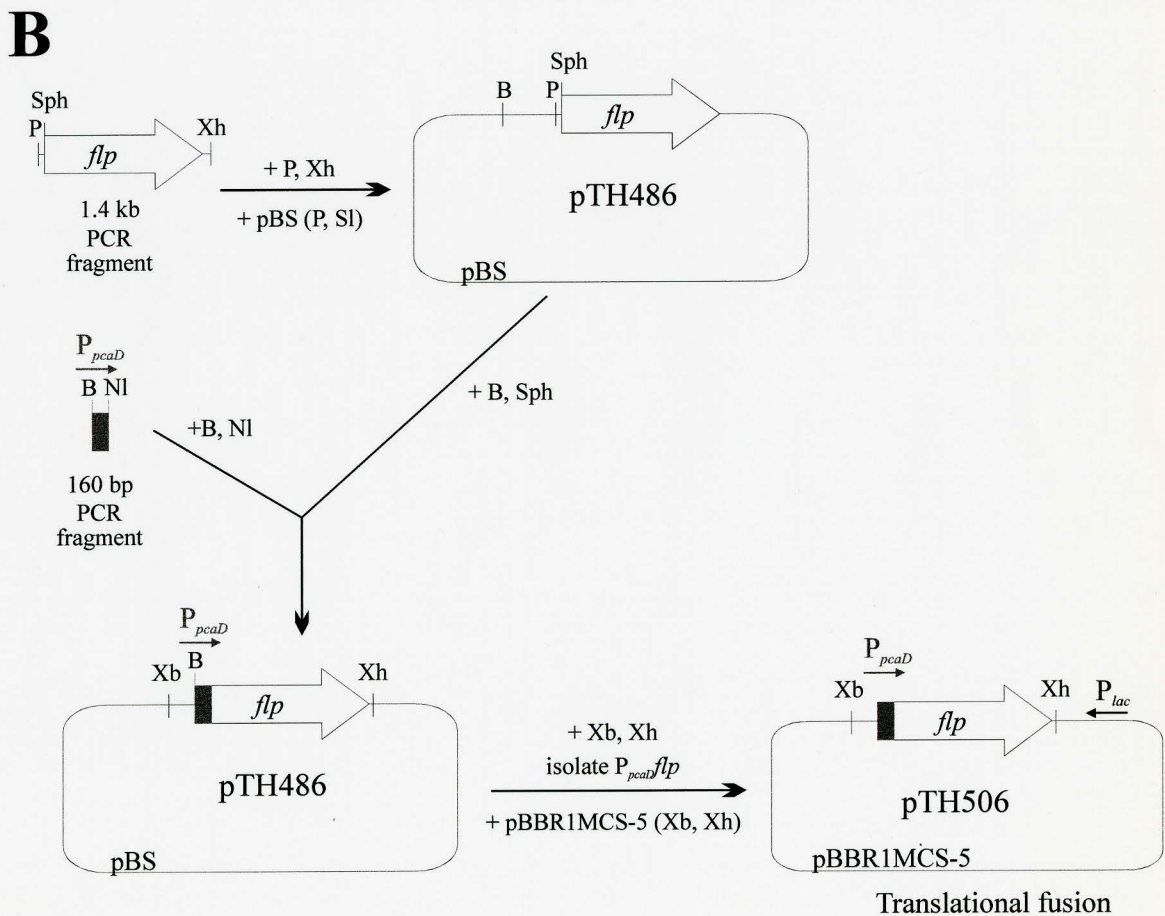
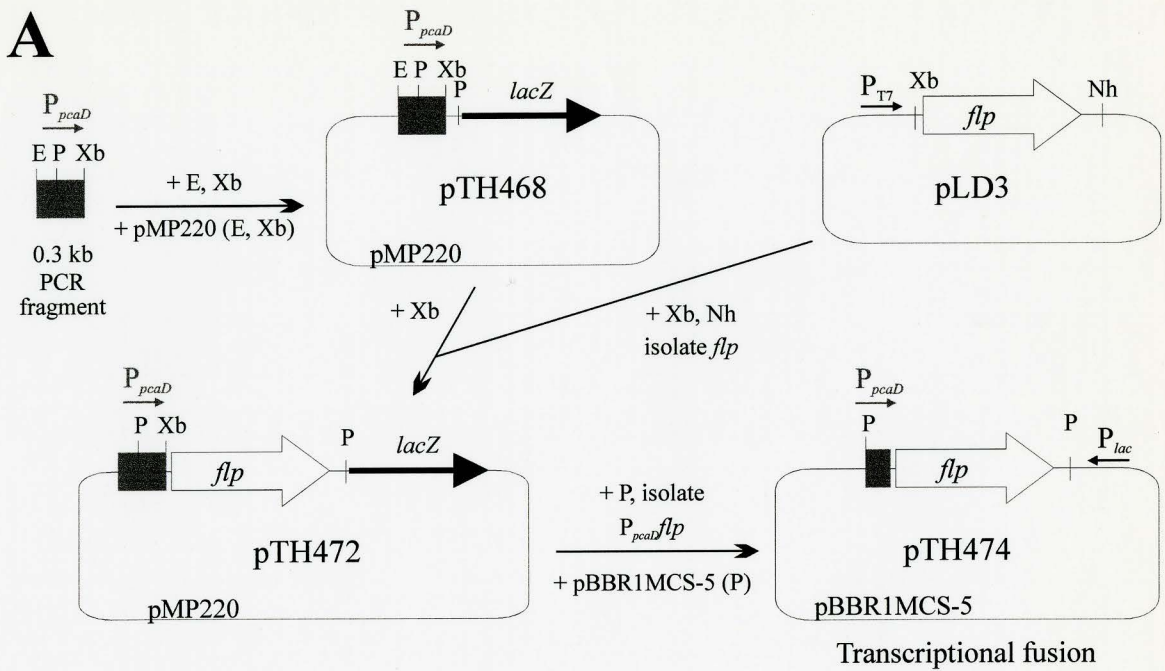
It has recently been shown that the *pcaD* promoter is induced several fold (up to 200 times) upon addition of protocatechuate (Clark 1998; MacPherson 1995). We reasoned that the use of such a *R. meliloti* promoter should allow controlled, high level expression and allow greater Flp activity. Primers AB11717 and AB11718 (refer to Table 4) were synthesized to amplify a fragment containing the *pcaD* promoter region and the beginning of the *pcaD* open reading frame from plasmid pTH185 (MacPherson 1995) (see Appendix B). This *EcoRI/XbaI* PCR product was cloned into pUC118 (for sequencing), and subsequently into pMP220 (Spaink *et al.* 1987). This new vector, pTH468, was used for expression studies (for further details, see Clark 1998), and for further manipulation. A 1.7 kb *XbaI/NheI* fragment from pLD3 (Pan *et al.* 1991), carrying the *flp* gene and 40 bp of T₇ promoter sequence, was cloned into the *XbaI* site of pTH468 such that *flp* expression is driven by the *pcaD* promoter (plasmid pTH472), effectively creating a transcriptional fusion (diagrammed in Figure 8A). As a control for

later experiments, this *flp*-fragment was also cloned into pBBR1MCS-5, such that *flp* is in the opposite orientation to P_{lac} , creating pTH471. The *flp* expression plasmid pTH474 was created by cloning a 2.0 kb *Pst*I fragment containing the $P_{pcaD}flp$ cassette from pTH468 into pBBR1MCS-5 in such a way that *flp* is, once again, in the opposite orientation to P_{lac} , driven only by the *pcaD* promoter (illustrated in Figure 8A).

The ability of pTH474-mediated *flp* expression to excise the *lacZ* gene in RmH940 was also tested (see Table 5). It was found that pTH474 constitutively expresses *flp*; all transconjugants harbouring the Flp vector, lost the *lacZ* gene as viewed on X-gal plates as white colonies. These colonies were confirmed to have lost the region between the minimal *FRT* sequences (refer to Figure 7), by screening for the loss of the Nm^f gene. This high level of expression is likely due to the extra DNA included with the *flp* gene (i.e. part of the T_7 promoter) in the cloning from pLD3, since similar results were found when pTH471 (“promoterless” *flp*) was mated into the tester strain (see Table 5). It therefore appears that the unregulated activity of *flp* is due to the promoter activity from the “leader sequence”, even though only 40 bp of the T_7 promoter are present in front of the *flp* gene.

As Flp synthesis from pTH474 resulted in 100% excision of the *FRT*-flanked *lacZ* gene in *R. meliloti* RmH940 (refer to Table 5), one other construct was made in order to obtain regulated Flp synthesis from the *pcaD* promoter. The primers AB13243 and AB13244 (refer to Table 4) were designed to amplify the *pcaD* promoter using the data from the previous sequencing reactions and alignments to other *pcaD* genes (refer to Appendix B). This cloning strategy (diagrammed in Figure 8B) was designed to create a

Figure 8: Diagram demonstrating the cloning scheme used to construct the ($P_{pcaD}flp$ fusion) Flp-delivery plasmids. The P_{pcaD} PCR products are represented by dark grey boxes, the *flp* gene is represented by the open arrow. The promoterless *lacZ* gene of pMP220 is represented by the dark arrow. The orientation of the promoters is indicated, as are the steps in the constructions (alongside the arrows) and the relevant restriction sites. Abbreviations are as follows: B, *Bam*HI; E, *Eco*RI; Nh, *Nhe*I; NI, *Nla*III; P, *Pst*I; Sph, *Sph*I; Xb, *Xba*I; Xh, *Xho*I. **Panel A.** Diagram demonstrating the construction of the transcriptional $P_{pcaD}flp$ fusion with “constitutive” Flp activity. **Panel B.** Diagram demonstrating the construction of the translational $P_{pcaD}flp$ fusion.



translational fusion at the start site of *pcaD*, with *flp* directly driven by the *pcaD* promoter, thereby eliminating the T₇ promoter sequence found in pTH471 and pTH474. Using PCR primers AB13241 and AB13242, the 1.4 kb product carrying the *flp* gene was amplified from pLD3, digested with *Pst*I/*Xho*I, and ligated to a *Pst*I/*Sal*I-cut pBS KS, creating pTH480. The 160 bp *Bam*HI/*Nla*III PCR P_{*pcaD*} product, amplified from pTH185, was then cloned into the *Bam*HI/*Sph*I-digested pTH480. This vector, pTH486, was then digested with *Xba*I/*Xho*I to cleave out the 1.5 kb P_{*pcaD*}*flp* cassette, and this fragment cloned into pBBR1MCS-5, resulting in plasmid pTH506 (refer to Figure 8B). This vector was tested by mating it into the tester strain RmH940 (see Table 5). Transconjugants consistently displayed low level Flp activity (between 10 and 30% excision) when screened on X-gal media, regardless of the addition of protocatechuate. Therefore, due to the inadequate regulation of Flp activity from this plasmid, pTH474 was used to provide Flp *in trans* in all further experiments.

3.3: Analysis of Tn5::pTH455 and Tn5-132::pTH456 Combinations

Pulsed-Field Gel Electrophoresis—A Method For Isolating the Intervening Regions

In this project, pulsed-field gel electrophoresis (PFGE) was used to determine the IS50-location of the targeting vectors in strains RmH992, RmH993, RmH994 and RmH995 (each containing $\Omega 5056::Tn5::pTH455$ and $\Omega 5069::Tn5-132::pTH456$), where the insertions lie approximately 50 kb apart. After DNA plugs (genomic DNA embedded in agarose) were made from these strains and Rm1021 (as a control), they were digested with *SpeI* and *I-SceI*, separately (refer to Figure 9). These digests were run on pulsed-field-certified agarose gels, and when compared to the marker lane, the Rm1021 control, and each other, several discoveries were made.

The *SpeI* lanes looked very similar to published *SpeI* digests of Rm1021 run on similar pulsed-field gels (Sobral *et al.* 1991b) and the approximate band sizes matched published results (Sobral *et al.* 1991a). The *SpeI* digests of all strains looked similar (example in Figure 9) to the Rm1021 control, even though the targeting vectors add two *SpeI* sites into the pExo megaplasmid. The extra bands can be attributed to the naturally occurring *SpeI* site located in this ~50 kb region (Honeycutt *et al.* 1993) and the two *SpeI* sites introduced on the targeting vectors, and so the difference in band size can not easily be resolved.

The *I-SceI* meganuclease enzyme recognizes a non-palindromic sequence of 18 nucleotides (Colleaux *et al.* 1988) which statistically translates to one recognition site in

Figure 9: Determination of the locations of *FRT* sites (and entire targeting vectors) using pulsed-field gel electrophoresis (PFGE). Genomic DNA from the wild type strain Rm1021 and from a strain carrying two targeting vectors (RmH994) was restricted with *SpeI* and *I-SceI*. The *SpeI* digest patterns of both samples are similar (although not identical—see marked bands) and are in agreement with the known *R. meliloti SpeI* digest pattern (Sobral *et al.* 1991a). A ~50 kb band (marked with the arrow) was observed from RmH994 when digested with *I-SceI*, attributed to the inclusion of its recognition site in the targeting vectors (refer to Figure 10). No such band was observed from wild type DNA.

H994
(Ω 5056::Tn5::pTH455
 Ω 5069::Tn5-132::pTH456)

Rm1021
wild type

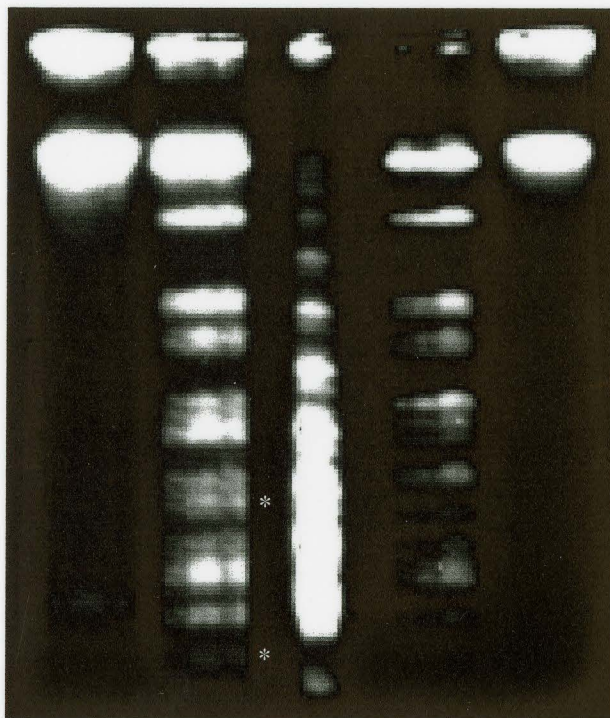
I-SceI

*Spe*I

*Spe*I

I-SceI

~50 kb →



6.9×10^{10} bp. Due to its extremely large recognition sequence, no natural *I-SceI* sites have yet been discovered in any of the genomes analyzed (except yeast) (Boehringer Mannheim 1998, Jumas-Bilak *et al.* 1995). Therefore, as expected, *I-SceI* did not appear to cut Rm1021 DNA. However, in the lanes with *I-SceI*-digested targeted DNA (from strains RmH992-RmH995), there appeared a single band approximately level with the 48.5 kb marker band (for RmH994, refer to Figure 9), as expected because of the inclusion of the *I-SceI* site in both targeting vectors. Upon careful examination, it was clear that the bands varied slightly, but distinctly, in size; although the bands for RmH992 and RmH995 appeared identical in size, for RmH993, the band appeared slightly larger than the others, and for RmH994, the band appeared slightly lower than the others (data not shown). We interpreted these results to mean that the difference in size was due to the inclusion or exclusion of both transposons (see Figure 10). In RmH993, the targeting vectors lie in the outer *IS50s* with respect to the ~50 kb intervening sequence, whereas in RmH994, the vectors were targeted to the inside *IS50s* with respect to the intervening region. From the PFGE results, the directions of the *FRT* and *oriT* sites were deduced (illustrated in Figure 10).

Flp-mediated deletions in *R. meliloti*

The construction of the “constitutive” *flp*-expression vector, pTH474, allowed us to examine the effect of transfer of this plasmid into various *FRT*-containing *R. meliloti* strains. In strains carrying the two targeting vectors on the outside *IS50s* (parallel *FRT*

Figure 10: Schematic representation of the four $\Omega5056::Tn5::pTH455$ $\Omega5069::Tn5-132::pTH456$ combinations. These insertions are approximately 50 kb away from one another. Integration of the targeting vectors into both *IS50* elements of the targeted transposon were identified by PCR (refer to Figure 5). Combinations of the two insertions ($\Omega5056$ and $\Omega5069$) were subsequently made by transduction (refer to Figure 6). The triangle shows *FRT* directionality. The *IS50* elements are shaded white (outside end) to black (inside end), indicating directionality. The locations of the *I-SceI* sites (I) are indicated. pExo DNA is represented by the dashed line.

~50 kb
pExo DNA

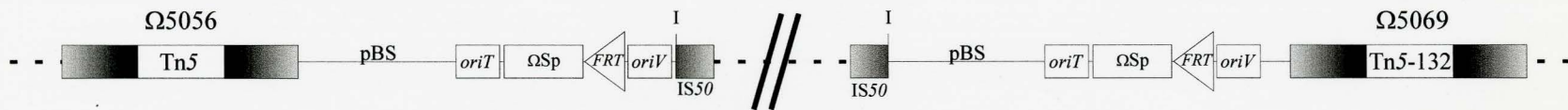
RmH992



RmH993 (outer IS50s)



RmH994 (inner IS50s)



RmH995



sites), excision of both transposons allowed screening for loss of the corresponding Nm^f and Ot^f genes (see Figure 11).

Two regions were tested in the first deletion experiment, the 50 kb region flanked by $\Omega 5056$ and $\Omega 5069$ in RmH992-RmH995 (refer to Figure 10), and a larger region, ~180 kb, from $\Omega 5007$ to $\Omega 5035$ (see Figure 12 for Ω insertion sites and Appendix A for structural details), encompassing a locus (*exoZ*) involved in the synthesis of the acidic exopolysaccharide, succinoglycan (EPS-I). Although, only two of the combinations of the $\Omega 5007/\Omega 5035$ insertions were available at the time, loss of succinoglycan synthesis can be readily detected as non-fluorescent colonies on LB containing calcofluor (Finan *et al.* 1985, Leigh *et al.* 1985).

The constitutive *flp*-expression vector, pTH474, was mated into the two $\Omega 5007/\Omega 5035$ strains (RmH990 and RmH991), selecting for the Gm^f carried by the plasmid. The transfer frequency, of 10^{-3} Gm^f transconjugants per recipient, was similar to the frequency attained in the transfer of the vector into non-*FRT*-containing *R. meliloti* strains. The transconjugants were screened for loss of the transposon-encoded Nm^f and Ot^f genes (refer to Figure 11) as well as for loss of succinoglycan synthesis. Although the 100 RmH990 transconjugants examined did not appear to have lost any of the markers, all 100 screened RmH991 transconjugants were dark on LB calcofluor plates, suggesting they had lost the 180 kb region, while retaining the transposons. Because these transconjugants retained the transposons but appeared to have lost the intervening region, it was inferred that in RmH991, the *FRT* sites (and thus the entire targeting

Figure 11: Schematic representation of detection of Flp-mediated excision by loss of both transposons and their respective antibiotic resistances. When the *FRT* sites are located in the outer *IS50* elements, Flp-mediated excision results in the loss of both transposons as well as the intervening region. Relevant antibiotic resistances conferred by the elements are indicated. The triangle shows *FRT* directionality. The *IS50* elements are shaded white (outside end) to black (inside end), indicating directionality. The locations of the I-*SceI* sites (I) are indicated. pExo DNA is represented by the dashed line.

FRT sites within the outer *IS50*s

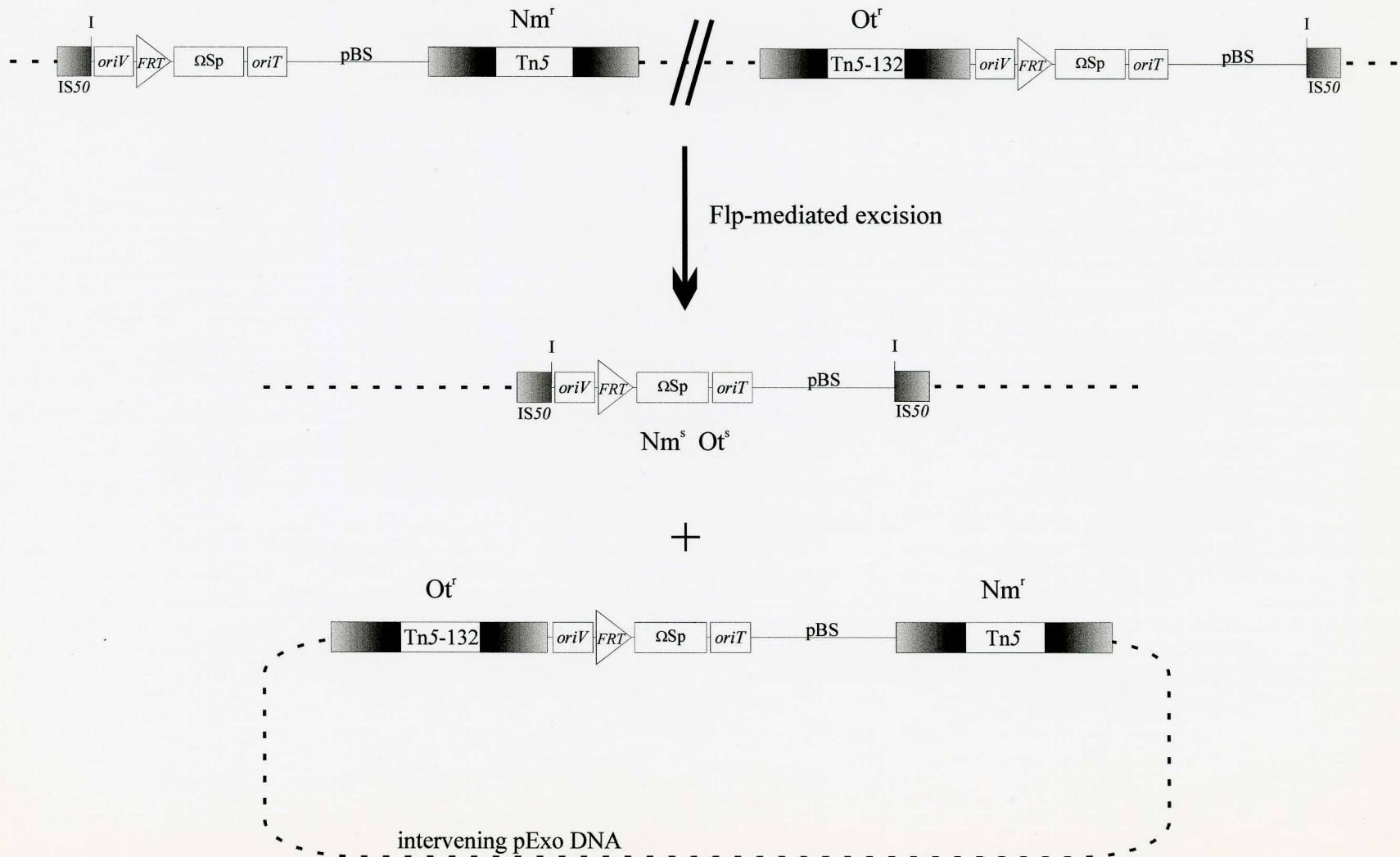
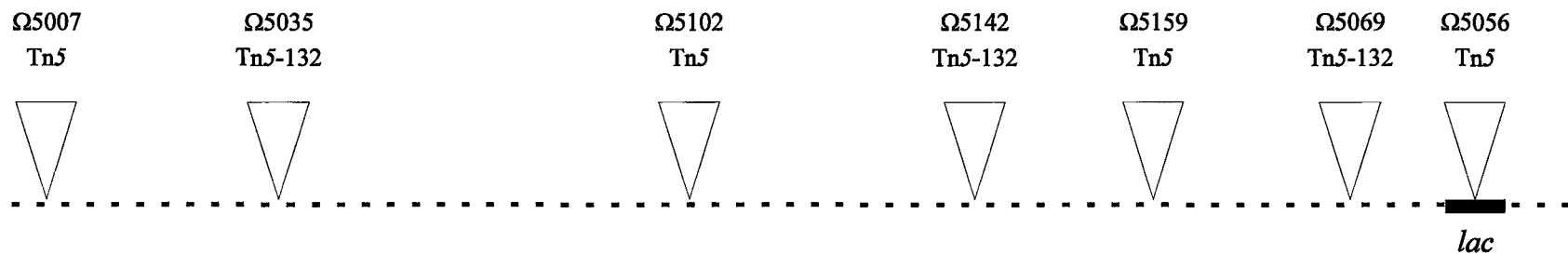
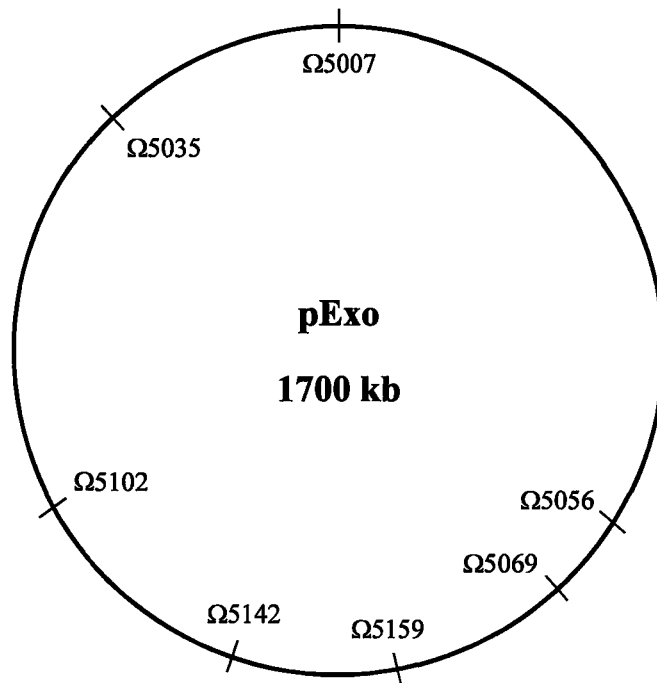


Figure 12: Circular and linear maps of the pExo megaplasmid demonstrating the relative locations of the Ω insertions. The type (Tn5 or Tn5-132) and approximate distance between adjoining insertions (triangles) are indicated. The location of the *lac* gene is also illustrated.



vectors) are located within the inner *IS50* sequences of the transposons, flanking this 180 kb region (refer to Appendix A).

All four variations of $\Omega5056/\Omega5069$ combinations (RmH992-RmH995) were also tested with pTH474. Like the previous experiment, a similar transfer frequency was observed for strains RmH992 and RmH995, however no transconjugants were recovered from matings into RmH993 and RmH994. The RmH992 and RmH995 transconjugants were screened for loss of the transposons but no excision was observed. The absence of any RmH993 and RmH994 transconjugants suggests that transfer of the “constitutive”-Flp plasmid into these strains is lethal. We inferred that Flp mediates excision of the 50 kb region, in strains RmH993 (from the outer *IS50s*) and RmH994 (from the inner *IS50s*), and that excision of this region results in cell death based on the successful deletion of the 180 kb region (above), the PFGE results, and the absence of detectable Flp-mediated excision in RmH992 and RmH995.

After Flp-mediated excision, two distinct molecules remain: the circularized “excised” region and the rest of the pExo megaplasmid. It is assumed that one will not propagate but will be lost from the population due to separation from its replication origin (which is carried on the other molecule—refer to Figure 11). It was therefore postulated that this 50 kb region carries the pExo origin of replication. Consequently, recombination between the two *FRT* sites would separate the pExo megaplasmid from the 50 kb region carrying its origin of replication. In this case, the molecule lost would be the remaining pExo megaplasmid; which may carry functions required for cell viability. A fact that lends support for this idea is that deletion of this particular region of pExo, in

the deletion study conducted by Charles and Finan (1991), was never obtained. It seems likely that genes carried on the rest of the pExo megaplasmid are required for cell survival and thus, removing the origin of replication, by excising the 50 kb fragment, is lethal.

Regions between Ω 5102 and Ω 5142, between Ω 5142 and Ω 5159, and between Ω 5159 and Ω 5069 (see Figure 12 for Ω insertion locations and Appendix A for structural details) are flanked by *FRT*-targeted transposons and have been assayed for deletion with Flp. After selecting for Gm^r transconjugants (carrying the *flp*-expression vector), 50 from each mating were screened for loss of the transposon-encoded antibiotic resistance markers. One of the four strains, for each set of combinations, did indeed lose both resistance markers, indicating excision and loss of the transposons as well as the DNA sequences between them (illustrated in Figure 11). Knowing the location/direction of the *FRT* sites, and thus the targeting vectors, for one transposon insertion facilitates making and screening subsequent, adjoining combinations (refer to Appendix A).

Transfer of pExo *R. meliloti* DNA into *A. tumefaciens*

It was shown by Banfalvi *et al.* (1985) and Finan *et al.* (1986) that the pExo megaplasmid could stably propagate in *A. tumefaciens*. To address the question of whether the region between Ω 5056 and Ω 5069 carries the pExo origin of replication, this region was excised and transferred to *A. tumefaciens*. The strains RmH993 and RmH994 which carry direct *FRT* sites flanking this 50 kb region were used as donors.

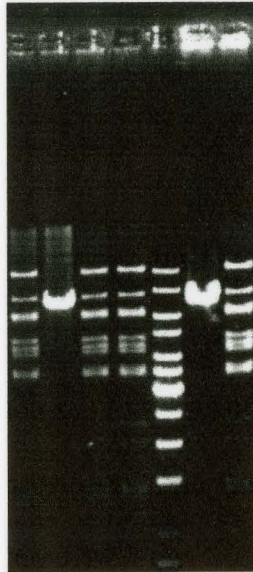
Four-way conjugal bacterial matings were set up to include the donor *R. meliloti* strain, the recipient *A. tumefaciens* strain and the two *E. coli* strains carrying the Flp-delivery and helper plasmids. Transconjugants were selected with 100 µg/ml spectinomycin (selecting for the excised 50 kb plasmid) and 25 µg/ml rifampicin (selecting for the *A. tumefaciens* recipient). Although all four strains (RmH992-RmH995) were used as donors, only two, namely RmH993 and RmH994, transferred Sp^f at a frequency 100-1000 fold higher ($\sim 10^{-5}$ transconjugants per recipient) than the other two strains. 50 transconjugants from each these two mating were screened for resistance to neomycin (200 µg/ml) and oxytetracycline (3 µg/ml). The transconjugants from RmH993 donor were both Nm^r and Ot^r, confirming that the targeting vectors in this donor strain are located in the outer IS50 elements (illustrated in Figure 11). The transconjugants from the RmH994 donor, however, were sensitive to both antibiotics, indicating that the transposons were not transferred along with the Sp^f marker (and the 50 kb region). The replication of the 50 kb region could only be attributed to *R. meliloti* DNA because the *ori*_{V_{RK2}} and ColE1 *ori* of pBluescript do not function in *A. tumefaciens*. Therefore, the 50 kb region excised from strains RmH993 and RmH994 must carry the pExo megaplasmid origin of replication. Plasmid DNA, prepared from 12 of these colonies, was digested with *Eco*RI (see Figure 13A) and *Bam*HI (see Figure 13B) separately. Addition of the resulting fragment sizes totaled approximately 60 kb, consistent with the predicted 50 kb of pExo DNA plus 8 kb of targeting vector DNA.

Figure 13: Determination of the capture of a large plasmid. Following a conjugation with *E. coli* as recipient and *R. meliloti* H994 (carrying two targeting vectors flanking a 50 kb pExo region between Ω 5056 and 5069) as donor, plasmid DNA was prepared from large and small Sp^{f} transconjugants and then digested with *EcoRI* (Panel A) and *BamHI* (Panel B). Six samples are shown; 2 and 5 were obtained from large colonies, whereas 1, 3, 4 and 6 were obtained from small colonies. It appears that the large colonies carry plasmids resembling the ~7.5 kb targeting vectors, whereas the small colonies carry a very large plasmid (~60 kb). In matings where *A. tumefaciens* was used instead of *E. coli* as recipient of *R. meliloti* RmH994 DNA, Sp^{f} transconjugants appeared to sustain a ~60 kb plasmid with similar digest patterns. L indicates the DNA ladder; sizes are indicated at the side of the figure. Several of the lanes from the original scan were omitted from this figure for clarity.

A

*Eco*RI digests

1 2 3 4 L 5 6



DNA
ladder
(Kb)

- 10
- 8
- 6
- 5
- 4
- 3.5
- 3
- 2.5
- 2
- 1.5
- 1
- 0.75

B

*Bam*HI digests

1 2 3 4 L 5 6



DNA
ladder
(Kb)

- 10
- 8
- 6
- 5
- 4
- 3.5
- 3
- 2.5
- 2
- 1.5
- 1
- 0.75

Transfer of pExo *R. meliloti* DNA into *E. coli*

Similar four-way matings were conducted with *E. coli* as the recipient of the 50 kb region of *R. meliloti* pExo DNA. Once again, 100 µg/ml of spectinomycin was used to select for the transfer of the *FRT*-flanked pExo DNA. The *oriV* of RK2 was included in the construction of the targeting vectors because it was hoped it would be sufficient for TrfA-dependent replication of large plasmids. Initiation of replication at the *oriV* site is mediated by a protein called TrfA, which is responsible for localized opening at the *oriV*, enabling replication to proceed (Konieczny *et al.* 1997). The ampicillin resistant (Ap^r) *E. coli* strain JW192 carries a copy of the *trfA* gene integrated in the chromosome, and was used as the recipient strain. After Flp-mediated recombination within an *R. meliloti* strain carrying two *FRT* sites in parallel, the circularized, excised region of pExo (carrying a copy of the targeting vector) can be transferred via the *oriT*, and stably maintained in the *E. coli* recipient as a plasmid via the *oriV*.

The *R. meliloti* donor strains used in the initial experiment were RmH992-RmH995 (refer to Figure 10). After the successful transfer into *A. tumefaciens* of the 50 kb region flanked by *FRT* sites from the same strains, it was expected that matings with RmH993 and RmH994 would exhibit much higher (100 fold) transfer frequencies than RmH992 and RmH995. However, this was not the case; it appears that the targeting vectors are not so stably inserted into the genome, but can be re-transferred via the *oriT*, as the original targeting plasmids, to *E. coli* in these four-way matings. This is expected, as the plasmid cointegrates should resolve at a frequency of approximately 10⁻³. This in turn masks the transfer events of the *R. meliloti* pExo region of interest (in this case, the

50 kb region). However, although there were just as many Sp^f transconjugants in every mating, colonies of two distinct sizes were observed when RmH994 was the donor. Some colonies had typical *E. coli* colony morphology while others were much reduced in size, even after an additional day of incubation. Only typical (large) *E. coli* colonies were observed for all other matings. Plasmid DNA was isolated from representative transconjugants of each mating, including colonies of both sizes from the mating with RmH994 as donor, and digested with *EcoRI* and *BamHI*.

All samples except for those from “small” colonies resembled the digest patterns of the original targeting vectors, pTH455 and pTH456 (see Figure 13). The samples from the small colonies showed more than 10 bands of various sizes (0.5 to 10 kb totaling approximately 60 kb), identical to those seen with the *A. tumefaciens* transconjugant plasmids. Excision from RmH994, which includes a copy of the pTH456 targeting vector along with the intervening DNA (see Figure 14), was expected to generate a plasmid of this size. One such plasmid was named pTH482 and was subjected to further analysis.

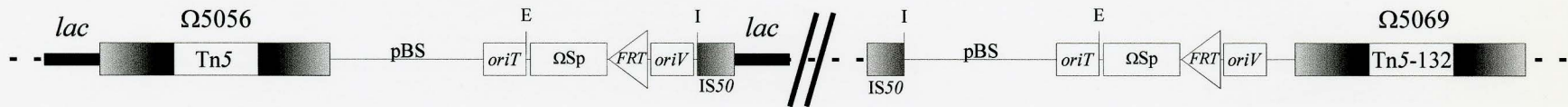
Excision Prior to Transfer of the 50 kb Region Is Not Required

It had not yet been shown whether Flp-mediated excision was required prior to transfer of this large region from *R. meliloti* to the *trfA*-expressing *E. coli* strain. Conjugal DNA transfer occurs in a site (*oriT*) and strand specific (unidirectionally, 5' to 3') manner (reviewed by Willetts and Wilkins 1984). It is possible that transfer originated from one of the *oriT* sites on the pExo megaplasmid and ended at the other

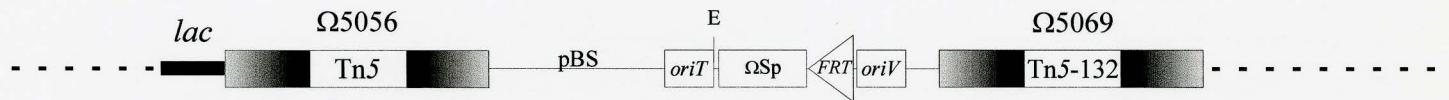
Figure 14: Schematic representation of the Ω 5056- Ω 5069 region and capture of the 50 kb region. The triangle shows *FRT* directionality. The *IS50* elements are shaded white (outside end) to black (inside end), indicating directionality. The locations of the Ω 5056::*Tn5*-interrupted *lac* gene is indicated by a black box. pExo DNA is represented by the dashed line. The *EcoRI* and I-*SceI* sites (I) are indicated.

RmH994 (inner IS50s)

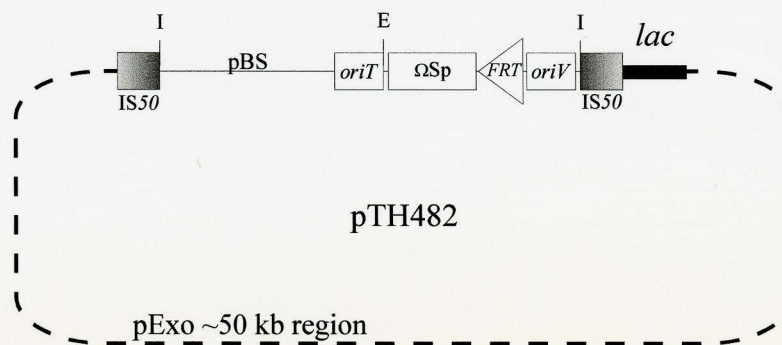
~50 kb
pExo DNA



Flp-mediated excision of the
~50 kb intervening region



+



oriT site, since these sites are, like the *FRT* sites, in parallel orientation. If this is the case, Flp-mediated excision of pExo DNA, prior to transfer, would not be required.

To investigate this possibility, tri-parental matings, without the Flp-delivery plasmid, were compared with the 4-way matings. Similar transfer frequencies and similar colony morphologies were obtained from both types of matings. *EcoRI* and *BamHI* digest patterns of plasmid DNA from the Sp^f transconjugants (RmH994 as donor for both matings) revealed similar ~60 kb plasmids (refer to Figure 13). The 50 kb region flanked by direct *FRT* sites in RmH994 can therefore be transferred into the *trfA*-expressing *E. coli* strain in a Flp-independent fashion.

Analysis of pTH482, the 60 kb plasmid

DNA of the ~60 kb excised plasmid, pTH482, was transformed into DH5 α and JW192 (DH5 α with *trfA* in the chromosome) to test whether the *oriV*/TrfA system was responsible for replication of the plasmid. Both transformations yielded two different types of Sp^f colonies, similar to the large and small *E. coli* colonies described above. Analysis of the plasmid DNA from 10 such colonies revealed that the large colonies harboured constructs resembling the original targeting vectors. The small colony types, however, produced digest patterns identical to pTH482. Since identical results were obtained from both recipient strains (DH5 α and its *trfA*-expressing derivative), it was concluded that the replication of pTH482 was *oriV*/TrfA-independent. Replication is therefore due, at least in part, to the presence of the pBS backbone.

In order to ascertain whether pTH482 contained *R. meliloti* DNA, *R. meliloti* wild type genomic DNA was digested with *Bam*HI and *Eco*RI, run on an agarose gel, and probed by Southern Blot with DIG-labeled pTH482 DNA (see Figure 15). The plasmid pTH482 was run as a positive control. When comparing plasmid and genomic DNA, all but a few fragments appeared identical (indicated in Figure 15). These differences can be attributed to border fragments and vector sequences. It was therefore concluded that the 50 kb pExo region flanked by Ω 5056 and Ω 5069 was carried by pTH482. Final confirmation that pTH482 carried pExo DNA was attained after DNA sequencing (see Table 6).

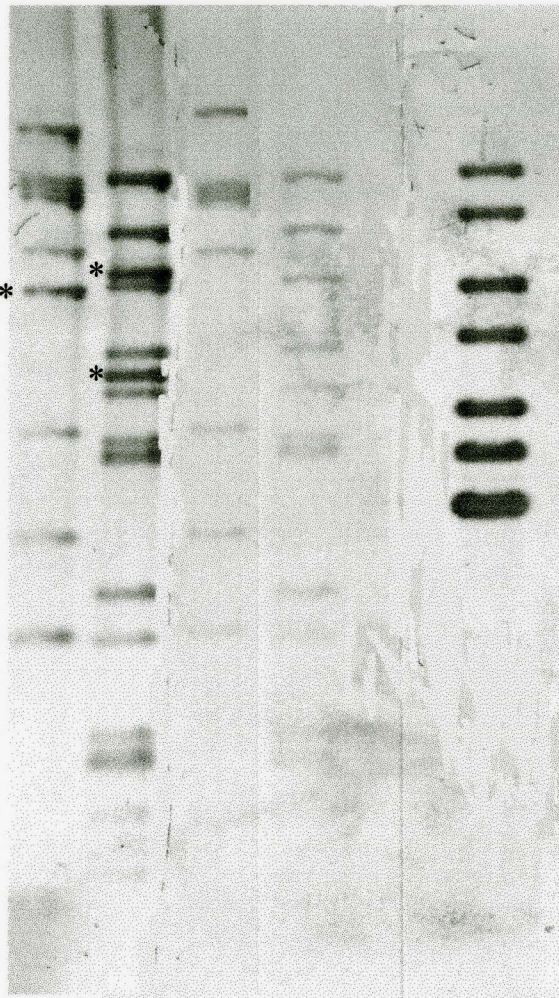
Several of the *Eco*RI fragments from pTH482 were subcloned into a Gm^r pUC19 derivative, pTC197 (in collaboration with I. Hernandez-Lucas). The recovered clones were screened for different sized *Eco*RI fragments by gel electrophoresis. Although not all the *Eco*RI fragments of pTH482 (identified by *Eco*RI digests and by Southern analysis) were obtained, those that were retrieved were sequenced using the M13 forward and reverse primers (see Table 6). The BLASTX program (Altschul *et al.* 1997) was used to translate the resulting sequences into all six possible reading frames and search known or putative open reading frames (ORFs) in GenBank databases. The highest matches for each sequence reaction are summarized in Table 6.

Several of the putative ORFs identified through BLASTX analysis of the sequences from the *Eco*RI fragments support what is known about the structure of pTH482 (illustrated in Figure 14). Particularly, two of the sequences show statistically

Figure 15: Southern blot showing that the DNA carried by pTH482 is *R. meliloti* DNA. Genomic DNA of Rm1021 and the plasmid DNA from pTH482 were probed with DIG-labeled pTH482. Both samples were digested with *EcoRI* (E) and *BamHI* (B). The digest pattern for each pTH482 digest closely resembles those of Rm1021, strongly suggesting that the plasmid pTH482 does indeed carry pExo DNA. The differences between the pTH482 and Rm1021 digests can be attributed to the “border” fragments generated by restricting the vector. L indicates the DNA ladder; sizes are indicated at the side of the figure. Several lanes from the original scan have been omitted from this figure for clarity. Brightness and contrast have also been adjusted to optimize visibility of bands.

pTH482
(~50 kb pExo) Rm1021

B E B E L



DNA
ladder
(kb)

— 10
— 8
— 6
— 5
— 4
— 3.5
— 3

Table 6: Highest Matches Obtained Through BLASTX Searches With the *EcoRI* Fragments of pTH482

Plasmid (Size of <i>EcoRI</i> Fragment)	M13 Primer*	Highest Match (Significant Alignments)	Statistical Significance (Expect Value)†	Putative Gene Product
pTH520 (3.6 kb)	For	hypothetical protein (<i>Chlamydia trachomatis</i>)	0.006	NA
	Rev	hypothetical 36.5 kilodalton protein in GbsA-TlpB intergenic region (<i>Bacillus subtilis</i>)	2e-06	NA
pTH521 (7.8 kb)	For	putative NTP pyrophosphohydrolase (<i>Streptomyces coelicolor</i>)	0.030	NA
	Rev	--	NA	NA
pTH522 (2.0 kb)	For	hypothetical protein of unknown function, similar product in <i>E. coli</i> (<i>Bacillus subtilis</i>)	3e-11	NA
	Rev	Tyrosine aminotransferase (<i>Escherichia coli</i>)	0.014	NA
pTH523 (1.3 kb)	For	--	NA	NA
	Rev	RepA protein (<i>Agrobacterium tumefaciens</i>)	1e-56	RepA
pTH524 (1.4 kb)	For	L-arabinose transport system permease protein AraH (<i>Escherichia coli</i>)	5e-10	membrane transport protein
	Rev	transcriptional repressor CytR (<i>Escherichia coli</i>)	1e-33	transcriptional regulator (LacI family)
pTH525 (3.5 kb)	For	hypothetical protein of unknown function, similar product in <i>E. coli</i> (<i>Bacillus subtilis</i>)	1e-53	NA
	Rev	--	NA	NA
pTH526 (4.8 kb)	For	GstA protein (<i>Rhizobium leguminosarum</i>)	2e-09	NA
	Rev	β -galactosidase (<i>Rhizobium meliloti</i>)	2e-66	β -galactosidase
pTH527 (0.35 kb)	For	elongation factor EF-G (<i>Synechocystis</i> sp.)	8e-11	elongation factor
pTH528 (4.5 kb)	For	(Cloning vector pBSL175)	4e-58	Ω Sp fragment
	Rev	β -galactosidase (<i>Rhizobium meliloti</i>)	2e-69	β -galactosidase
pTH529 (1.45 kb)	For	RepB protein (<i>Rhizobium leguminosarum</i>)	5e-31	RepB
	Rev	RepA protein (<i>Rhizobium leguminosarum</i>)	2e-56	RepA
pTH530 (6.1 kb)	For	hypothetical protein of unknown function (<i>Agrobacterium vitis</i>)	8e-41	NA
	Rev	hypothetical protein Rv0266c (<i>Mycobacterium tuberculosis</i>)	3e-14	NA

NA designates Not Applicable, and two dashes "--" designate no homology found

* M13-forward and M13-reverse primers are designated as "For" and "Rev" respectively

† The Expect value specifies the number of matches, with a given score, that are expected in a search of a database of this size absolutely by chance.

significant homology to the previously identified β -galactosidase gene from *R. meliloti*. This corresponds well with the fact that Ω 5056 insertion generates a Lac⁻ phenotype. Strengthening this argument is the nearly identical match to the Ω Sp interposon on the same fragment as one of the β -galactosidase matches (refer to Table 6). Since the Ω Sp and *oriT* site border the β -galactosidase gene, digestion with an enzyme (like *EcoRI*) would thus leave them on the same fragment (see Figure 14). The replication proteins RepA and RepB from several related microorganisms were also found to be statistically significant matches to three of the sequences obtained from the *EcoRI* fragments (refer to Table 6). This is consistent with the other data showing that this 50 kb region of *R. meliloti* DNA does indeed harbour the origin of replication and associated replication functions for the pExo megaplasmid.

Excision of Larger pExo Regions

In much the same way as described earlier, the targeting vectors pTH455 and pTH456 were directed to transposons (Tn5 and Tn5-132, respectively) between 100 and 300 kb away from one another. These were then combined into one strain by transduction, selecting for the Tn5-encoded Nm^f marker. Once again, there are four possible combinations for each pair of targeted transposons. Because the location of the *FRT* sites (hence the targeting vectors) in the Ω 5056 and Ω 5069 targeted insertions is known, any combination with either of these already had a 50% chance of excising the intervening region. Several “sets” of four combinations were made: RmK126 to RmK129 (targeted transposons flank 240 kb from Ω 5102 to Ω 5142); RmK130 to

RmK133 (targeted transposons flank 115 kb from Ω 5159 to Ω 5142); and RmK137 to RmK140 (targeted transposons flank 295 kb from Ω 5056 to Ω 5142) (refer to Figure 12 for Ω insertion sites and Appendix A for structural details).

All “sets” were subjected to Flp-mediated deletion analysis in order to determine the location of the *FRT* sites with respect to one another. Results observed with the RmK137 to RmK140 “set” was, as expected, similar to those observed with the RmH992 to RmH995 “set”. The insertions in strains RmK137 to RmK140 encompass the 50 kb region carrying the putative origin of replication and replication functions of the pExo megaplasmid. Deletion of this 295 kb region appeared, once again, to result in cell death since no Gm^r transconjugants were recovered. Deletion of the other two regions (between Ω 5159 and Ω 5142 and between Ω 5102 and Ω 5142) did not result in cell death, since similar transconjugation frequencies of $\sim 10^{-3}$ were observed for all four strains of each set. 50 Gm^r transconjugants from each mating were screened for the loss of the transposon-encoded Nm and Ot resistance genes (refer to Figure 11), which indicated that the *FRT* sites had been targeted to the outer IS50s, with respect to the intervening region of interest. The location of the *FRT* targeting vectors in the other three strains of each set were then deduced.

All sets were also included in tri-parental and four way matings into *E. coli* JW192 (*trfA*-expressing strain), selecting with 100 μ g/ml ampicillin and 100 μ g/ml spectinomycin. Although transconjugants were detected for each mating, no “small type” colonies were observed and, after plasmid preparation of 24 representative samples, *EcoRI* and *BamHI* digestions only revealed plasmids which resembled the original 7.5 kb

targeting vectors (data not shown). It therefore appears that the regions larger than 50 kb that have been targeted for excision and transfer, cannot stably replicate in *E. coli* using the *oriV* site, or the ColE1 origin of the pBluescript vector as origin of replication.

Since the insertions of the RmK137 to RmK140 “set” flanked a ~295 kb region which included the putative pExo origin of replication, these strains were mated with *A. tumefaciens* to verify that a region of this size could in fact be transferred and replicate stably as a plasmid in this host. Indeed, Sp^r transconjugants were isolated from matings in which the ~295 kb was flanked by *FRT* sites in direct orientation. Plasmid DNA was prepared from 12 representative transconjugants, and from restriction enzyme digestions, it appeared that a very large plasmid was carried by these strains (data not shown). Replication of this large plasmid is attributed to the origin of replication of the pExo megaplasmid and not to the *oriV* of plasmid RK2, nor to the ColE1 origin of pBluescript.

3.4: Using BAC Backbones For Recovery of Large pExo Regions

Construction of Bac Vector pTH504

The BAC cloning system, based on the *E. coli* F factor, was chosen as an alternative to the *oriV* of RK2, for replicating the large (over 100 kb in size) regions of pExo DNA. In this system, replication and partitioning is strictly controlled, ensuring the stable maintenance of large plasmids. BAC vectors have been successfully used for the cloning of large DNA inserts (Brosch *et al.* 1998; Cai *et al.* 1995; Kim *et al.* 1996a; Kim *et al.* 1996b; Misumi *et al.* 1997; Woo *et al.* 1994; Zimmer *et al.* 1997).

Due to the location of the targeting vector sequences after integration into the pExo megaplasmid, only the pTH456 pBS backbone is excised from the combinations when the *FRT* sites are closest to one another (refer to Appendix A). Therefore, only the targeting vector pTH456 was modified, replacing the pBS KS backbone with a pBAC backbone. pTH456 was digested with *SalI*, treated with Klenow polymerase to fill in the 5' overhang, and ligated to *SacI* linkers. This replaced the *SalI* site with a *SacI* site, effectively creating a 4.5 kb *IS50-oriV/FRT-ΩSp-oriT SacI* cassette. This vector was then cleaved with *SacI* and the cassette ligated to *SacI*-digested pBACe3.6, creating pTH504 (the construction of this plasmid was performed by I. Hernandez-Lucas). The vector pBACe3.6 carries a chloramphenicol resistance (Cm^r) gene, and strains carrying pTH504 are resistant to 10 µg/ml chloramphenicol and 50 µg/ml spectinomycin. The resistance to spectinomycin conferred by pTH504 (50 µg/ml solid media) was markedly

less (approximately half) than that provided by pTH456 (100 µg/ml solid media), presumably due to the lower copy number of the pBAC backbone (1 to 2 copies per cell).

Deletion and Transfer of pExo Regions Flanked by pTH455 and pTH504

The new vector, pTH504, was targeted to the same Tn5-132 locations (Ω5069 and Ω5142) as was pTH456, and similar *R. meliloti* Sp^f transconjugant frequencies (10⁻⁶ to 10⁻⁷ per recipient) were observed. When screened on 10 µg/ml chloramphenicol, none of the transconjugants grew (Cm^s); it was therefore thought that chloramphenicol was not a useful antibiotic for selection in *R. meliloti*. It was not possible to identify (by using PCR) that both IS50 sequences (IS50L and IS50R) were targeted, as no primers were available, so the Tn5::pTH455 insertions in strains RmH953 to RmH958 were transduced into several (4 to 6) of the targeted transconjugants following selection on LB supplemented with Nm (refer to Appendix A). Determination of IS50 location with respect to other insertions was accomplished by Flp-mediated deletion analysis. Similar to the results obtained with pTH456-targeted strains, the combinations (Tn5::pTH455 with Tn5-132::pTH504) which flanked the 50 kb region (between Ω5056 and Ω5069) with direct *FRT* sites were analyzed by the absence of Gm^f Flp transconjugants. As expected, transfer of this plasmid into *A. tumefaciens* was also successful. Combinations, in which regions other than this putative replicator region were flanked by direct *FRT* sites, were screened for loss of the transposons after transfer of the Flp delivery plasmid and presumed excision (refer to Figure 11).

To prove that capture of the same 50 kb pExo region was also possible with the pBAC backbone, the Ω 5056/ Ω 5069 combinations (RmK188 to RmK191) were mated with the same *E. coli* strain expressing *trfA* (JW192) as well as with the rifampicin resistant (Rf^r) *E. coli* strain, MT620, which does not express *trfA*. Once again, similar types of colonies were observed—large Sp^f transconjugant colonies from all matings into MT620 and JW192, and also smaller-type Sp^f JW192 transconjugants from the RmK190 matings (the pTH504-equivalent of RmH994). Oddly, when screened for Cm^r, all colonies, including the small types were Cm^s. Plasmid DNA was prepared from several (10-12) colonies of either type. Only the smaller type colonies gave a large plasmid (pTH515); the others appeared to carry a plasmid of similar size to the original pTH504 targeting vector (data not shown). After digestion of pTH515 with *Bam*HI, the pattern of the DNA bands was identical to pTH482 (~60 kb plasmid) except for one band. Since the pBAC vector is larger than the pBS backbone, a larger-sized plasmid was expected from these matings; however it appeared that one of the bands in these digested samples was 3.0 kb lower than that seen in the pTH482 digest (indicated by arrows in Figure 16).

The pBAC targeting vector, pTH504, was re-examined in search of an explanation for the size discrepancy, the sensitivity of transconjugants to chloramphenicol, and the inability to observe pExo transfer into MT620. Although a *Sac*I digest confirmed that the 4.5 kb cassette carrying *IS50-oriV/FRT- Ω Sp-oriT* was present in a pBACe3.6 backbone, a further digest with *Eco*RI revealed that two *Sac*I cassettes had been ligated to the pBAC backbone in parallel orientation (performed by I. Hernandez-Lucas—data not shown). With both *oriT* sites also in direct orientation,

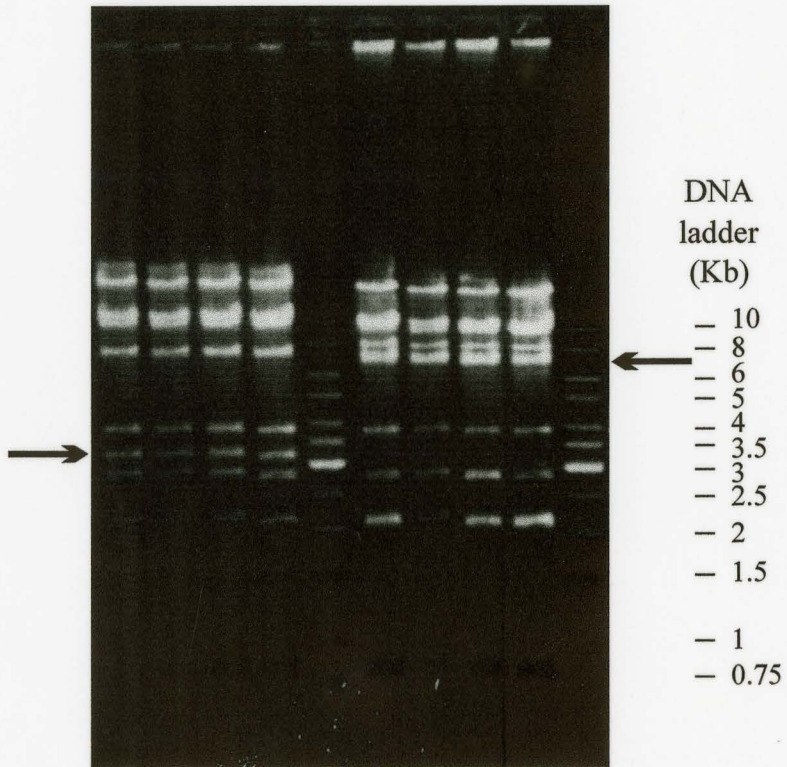
Figure 16: Agarose gel demonstrating the difference between pTH482 and pTH515. Both plasmids carry the same 50 kb pExo region, however, their backbones are different. pTH482 carries the pBS backbone along with the *FRT* cassette. Although pTH515 was first thought to carry the pBAC backbone, this vector appears to have no backbone, only the *FRT* cassette. Four independent clones carrying each vector were digested with *Bam*HI. It appears that pTH515 is missing a ~6.5 kb band (marked by a black arrow) present in pTH482 digests whereas pTH482 lack the ~3.2 kb band (marked by a black arrow) present in pTH515. The pBS backbone is 3 kb in size, which is the approximate size difference between these two indicated bands. L indicates the DNA ladder; sizes are indicated at the side of the figure. Several lanes from the original scan have been omitted from this figure for clarity.

pTH515
50 kb pExo
region without
backbone

pTH482
50 kb pExo
region with
pBS backbone

L

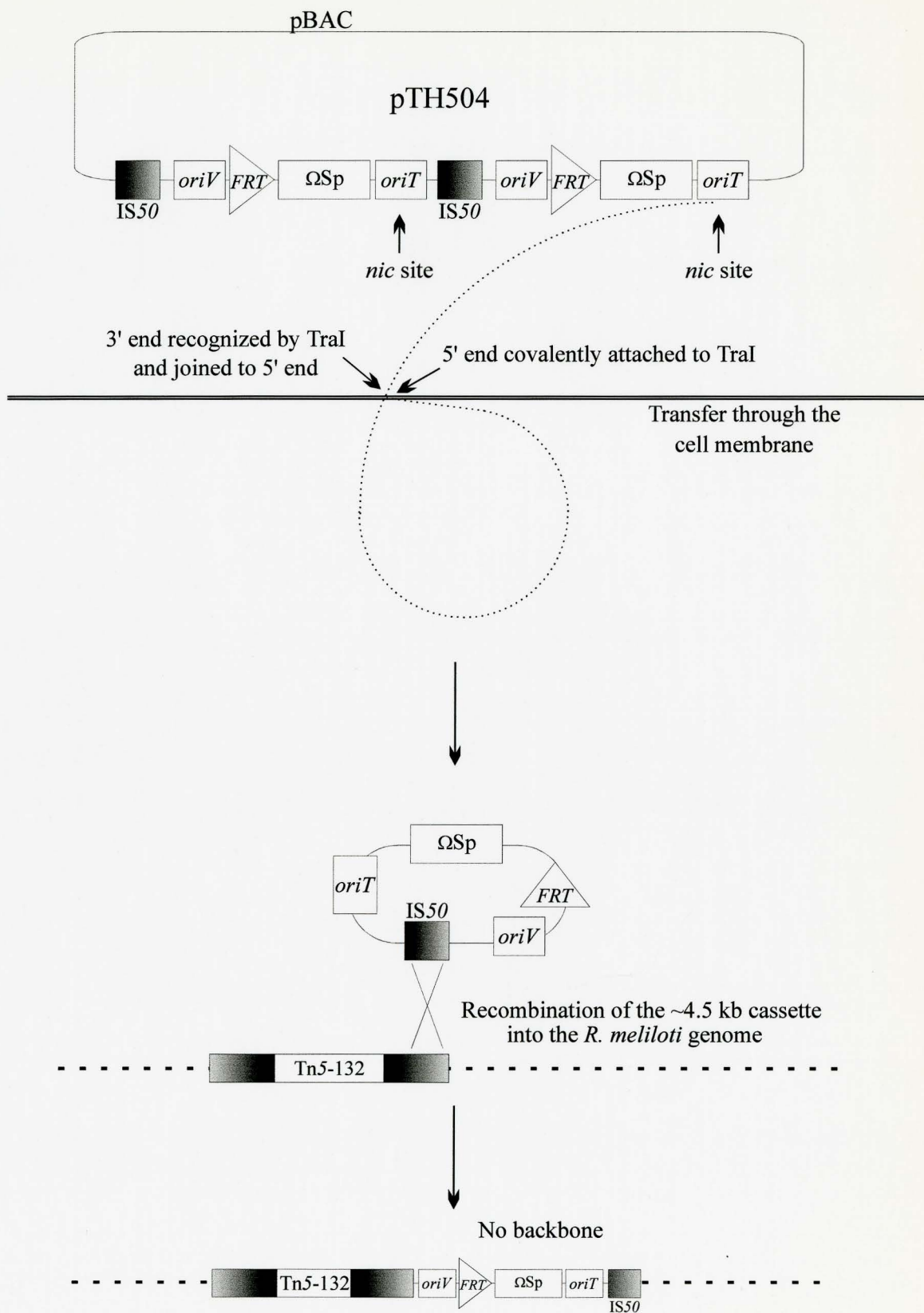
L



transfer and circularization of the *FRT* cassette without the backbone may have occurred (see Figure 17A), similar to the proposed Flp-independent transfer of pExo DNA from pExo-located *oriT* sites into *E. coli* and *A. tumefaciens* (refer to the section entitled “Excision Prior to Transfer of the 50 kb Region Is Not Required”). It was speculated that transfer, and subsequent recombination at the IS50 sequences, of only the 4.5 kb cassette (from *oriT* to *oriT*, lacking any backbone), may have occurred in the initial transfer of pTH504 into *R. meliloti* (diagrammed in Figure 17B). Subsequent transfer from *R. meliloti* strains, carrying two parallel *FRT* sites at Ω 5056 and Ω 5069, into *E. coli* would then result in the 50 kb pExo region being transferred along with only the 4.5 kb cassette, and no backbone (pTH515). This, in turn, explains the 3 kb difference between this plasmid and the larger pTH482: the size of the pBS KS backbone in pTH482. In this hypothesis, the plasmid pTH515 is Cm^s and would only replicate in a TrfA-dependent manner in *E. coli* via the *oriV*, explaining the inability to obtain the 50 kb region in MT620 (TrfA⁻).

To test this hypothesis, pTH515 and pTH482 were transformed into DH5 α and DH5 α expressing *trfA* (JW192). Results from two separate transformation experiments indicated that the plasmid pTH515 does replicate in a TrfA-dependant fashion (no DH5 α Sp^r transformants, ~200 JW192 Sp^r transformants), whereas, as previously tested, pTH482 does not require TrfA to be stably maintained in *E. coli* (~200 Sp^r DH5 α transformants, ~200 Sp^r JW192 transformants). Recovery of this plasmid was still quite useful; having no pBS KS sequences simplified the subcloning of randomly, mechanically sheared (nebulized) fragments into the well characterized and easily

Figure 17: Transfer of the 4.5 kb cassette lacking a backbone, from *oriT* to *oriT*. Two 4.5 kb cassettes were cloned in parallel into pBACe3.6. It appeared that the pBAC backbone was not transferred along with the cassette in conjugations into *R. meliloti* strains. Presented here is a mechanism of transfer, which would explain Cm sensitivity of transconjugants. The *oriT* is nicked and transfer occurs 5' to 3' (dotted line). The 5' end is thought to remain covalently attached at the cell membrane and then ligated to the 3' end of an identical *nic* site (Lanka and Wilkins 1995; Pansegrau *et al.* 1993; Sherman and Matson 1994). The resulting circularized molecule can then undergo recombination at one of the IS50 elements.



manipulated pUC119 plasmid. Initial shotgun sequencing of this 50 kb region has confirmed the previous results obtained from the sequencing of pTH482 *EcoRI* fragments.

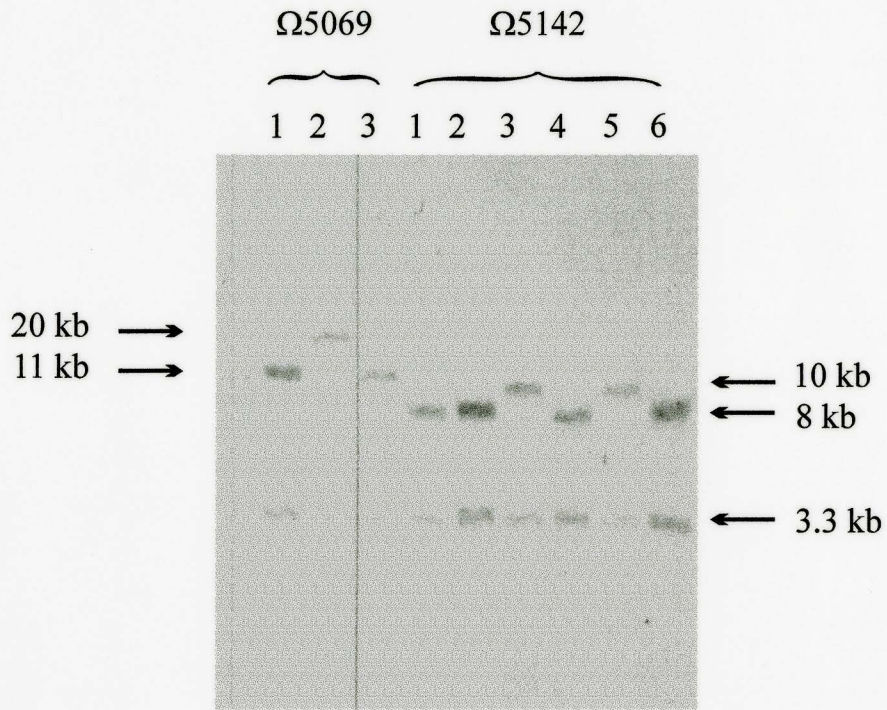
Analysis of Regions Retrieved With pTH509 as BAC Vector

A new pBAC vector, pTH509, was constructed from pBACe3.6 and pTH456 (performed by I. Hernandez-Lucas). pTH509 was verified with several digests to ensure that it carried the 4.5 kb *IS50-oriV/FRT-ΩSp-oriT SacI* cassette from pTH456 as a single copy. This targeting vector, like the others, was directed to the insertions $\Omega5069::Tn5-132$ and $\Omega5142::Tn5-132$. The number of Sp^f transconjugants was approximately 10 fold less (10^{-7} to 10^{-8} per recipient) than the numbers obtained with the other targeting vectors, perhaps due to the larger size of the pBAC backbone. These Sp^f transconjugants were screened for low level Cm^f (resistance to 5 $\mu\text{g/ml}$ chloramphenicol), using non-targeted strains as negative controls. The transconjugants were found to be Cm^f , affirming the presence of the BAC backbone and strengthening the model proposed (mentioned in the previous section) for the mechanism of transfer using pTH504 as targeting vector. Since primers to BACe3.6 were not yet available and the orientation of the 4.5 kb cassette with respect to the backbone was not known, Southern blotting experiments, not PCR, were used to determine whether both *IS50* sequences (*IS50L* and *IS50R*) were targeted (see Figure 18A). By probing with pMF21 (Manis and Kline 1977), which makes up part of the BAC backbone, insertions in the two *IS50* sequences can be distinguished after

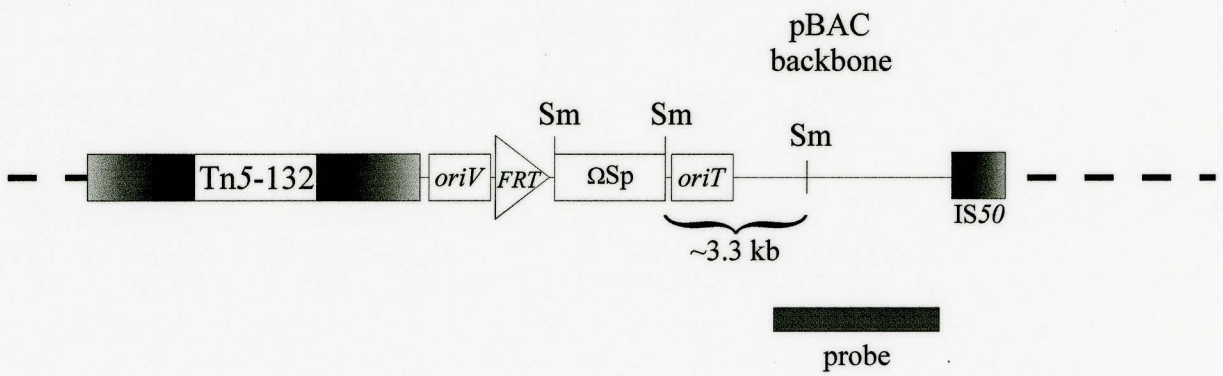
Figure 18: Southern blot indicating recombination of the pBAC targeting vector, pTH509, into both *IS50* elements of insertions Ω 5069 and Ω 5142, and a schematic representation of pTH509 integration. **Panel A:** Two Tn5-132 insertions were targeted with pTH509 and several transconjugants were selected for Southern analysis (three for Ω 5069, six for Ω 5142) after a *Sma*I digestion. Two different band patterns were observed for each insertion site (indicated by black arrows with their corresponding sizes). **Panel B:** Diagram of integration of pTH509 into one of the *IS50*s of a Tn5-132. Directionality of the *IS50* sequences is indicated by shading—white (outside end) to black (inside end). pExo DNA is represented by the dashed line. The approximate location of hybridization with the pMF21 probe (shaded box) is indicated. The locations of *Sma*I (Sm) sites are indicated. Several lanes from the original scan have been omitted from this figure for clarity.

A

Tn5-132 + BAC targeting vector pTH509



B



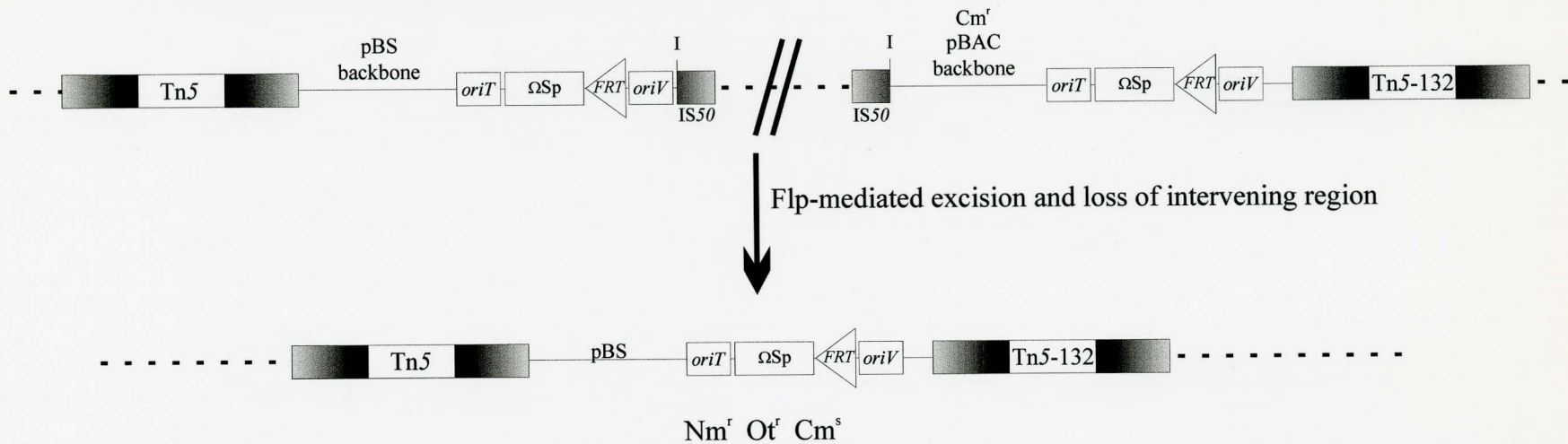
digesting genomic DNA with *SmaI* (see Figure 18B). However, this method will not indicate if the targeting vector recombined within the IS50L or IS50R.

Combinations of Tn5::pTH455 and Tn5-132::pTH509 were made via transduction, and transconjugants carrying the *flp*-expressing vector were analyzed. Two sets of four combinations will be described (refer to Figure 12 for the insertion locations and to Appendix A for structural details): RmK197 to RmK200 (Ω 5159 and Ω 5069, 130 kb apart), and RmK201 to RmK204 (Ω 5056 and Ω 5069, 50 kb apart). Flp-mediated excision of the 50 kb region between Ω 5056 and Ω 5069 once again appeared to be lethal, as determined by the lack of any Gm^r transconjugants. Deletion from the outer IS50 sequences of the 130 kb region resulted in Cm^r, Nm^s, Ot^s strains, due to loss of both transposons, and retention of the pBAC backbone. On the other hand, deletion from the inner IS50 elements retained the two transposons (Nm^r, Ot^r) but lost the Cm^r pBAC backbone (see Figure 19).

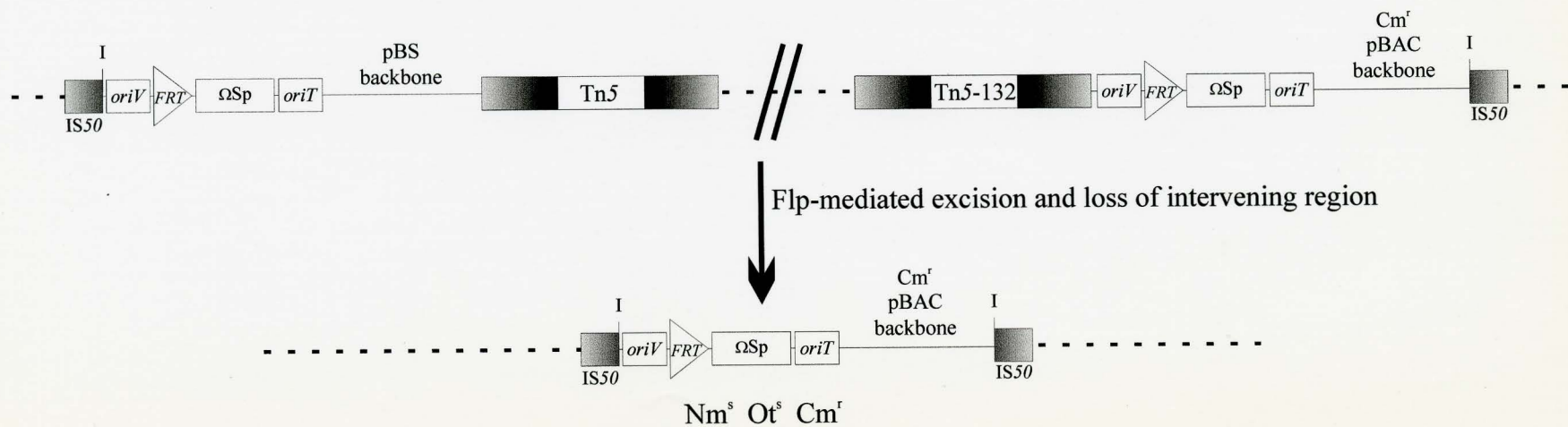
R. meliloti strains carrying the *FRT* sites within the inner IS50 elements were saved (refer to Appendix A). The transfer of the 50 kb and 130 kb regions into *E. coli* strains was attempted with these *R. meliloti* donors. Three recipients were used: JW192 (Ap^r, *trfA*-expressing), MT620 (Rf^r) and DH5 α (where the transferred Cm^r marker itself was used to counter-select *R. meliloti*). Spectinomycin and chloramphenicol were used to select for the transferred DNA. Large and small transconjugant colonies were observed in matings with DH5 α and JW192 as recipients. Matings involving MT620 however, gave 10 fold fewer colonies and all were of the smaller phenotype. In these matings, there also appeared to be a significant difference in transfer frequency between those

Figure 19: Schematic representation of Flp-mediated excision of pExo DNA flanked by Tn5::pTH455 and Tn5-132::pTH509 from the inner and outer *IS50*s. The loss of the Cm, Nm and Ot resistance genes is indicated. The triangle shows *FRT* directionality. The *IS50* elements are shaded white (outside end) to black (inside end), indicating directionality. The locations of the I-SceI sites (I) are indicated. pExo DNA is represented by the dashed line.

***FRT* sites within the inner *IS50*s**



***FRT* sites within the outer *IS50*s**



strains that underwent Flp-mediated excision (*FRT* sites and entire cassette in parallel) (10^{-6} to 10^{-7} per recipient) and those that did not ($\sim 10^{-8}$ per recipient).

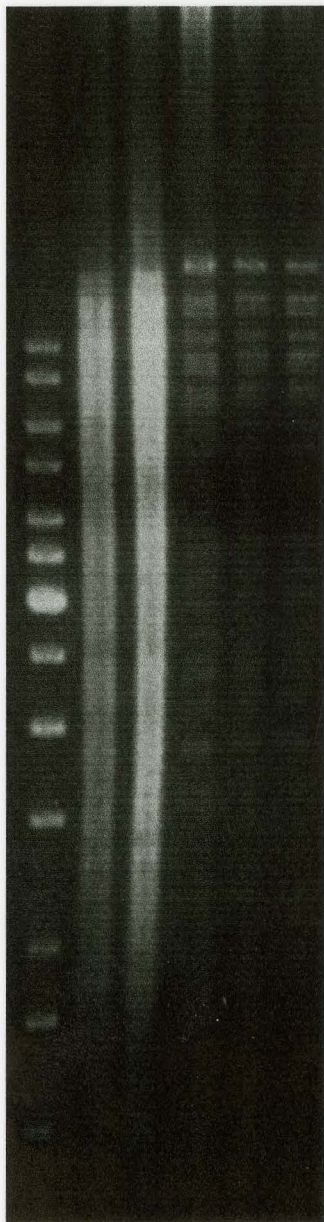
Plasmid DNA from the small colonies from each type of mating was digested with *Bam*HI. After analysis on an agarose gel, it was discovered that the 50 kb region in the form of a ~ 60 kb plasmid, designated pTH518, could be retrieved from all three matings. And, it appears from the many bands (>25 bands ranging from 0.5 kb to over 10 kb in size—see Figure 20) that the 130 kb region of pExo between $\Omega 5069$ and 5159 was isolated from all three matings. This plasmid, called pTH519, appears to be stable in all three strains.

pTH519, which putatively carries 130 kb of pExo DNA, was used to probe *R. meliloti* wild type DNA. The Southern is described in Figure 20. Due to the large number of bands, it is difficult to tell in the positive control (pTH519) which bands carry the targeting vector sequences, but it is clear that most bands are identical to those in the wild type lanes. The BAC vector therefore appears to stably maintain, in *E. coli*, a 130 kb region of pExo megaplasmid DNA. The purification and preparation of a shotgun library from pTH519 is currently under way.

This study establishes that Flp recognizes its *FRT* sites in *R. meliloti* and promotes recombination between them. This study also confirms that with this system, large regions of the pExo megaplasmid can be deleted *in vivo*. Results from this work demonstrate that these large regions of pExo, as big as 130 kb in size, can be transferred into *E. coli* and be stably maintained (if an appropriate origin of replication is present) for purification and sequencing purposes.

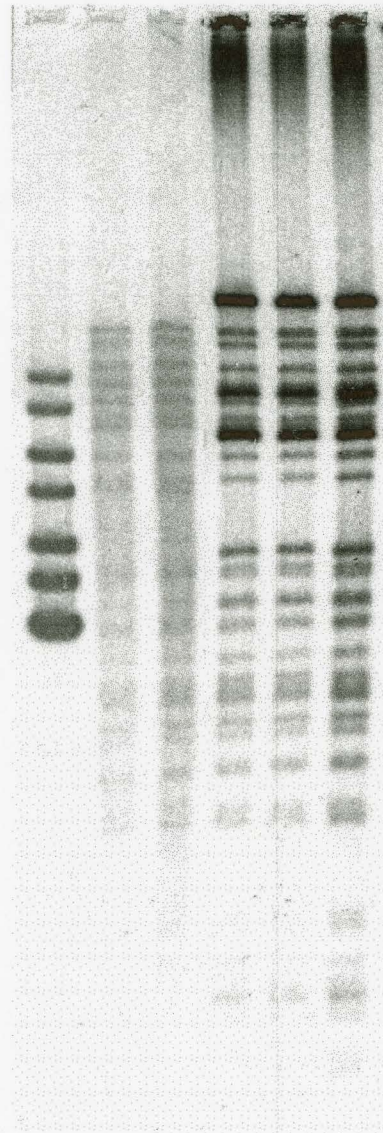
Figure 20: Southern blot showing that the DNA carried by pTH519 is *R. meliloti* DNA. *E. coli* carrying pTH519 derive from matings with *R. meliloti* RmK198, which harbours two targeting vectors in direct orientation flanking a ~130 kb region between Ω 5095 and Ω 5069. Genomic DNA of Rm1021, Rm5000 (both wild type *R. meliloti* strains) and plasmid DNA from the original pTH519 as well as two independent plasmid preparations from transconjugants (1 and 2) were digested with *Bam*HI and probed with DIG-labeled pTH519. The digest pattern for all pTH519 plasmid preparations closely resemble one another as well as those from Rm1021 and Rm5000, which strongly suggests that the plasmid pTH519 does carry a large region of pExo DNA. The dark bands present near the top of the blot in pTH519 digests can be attributed to uncut or single stranded DNA, or to *E. coli* genomic DNA (the probe may not have been very pure since it was made from DNA isolated using the small-scale plasmid preparation procedure). Other differences between the pTH519 with Rm1021 and Rm5000 lanes can be attributed to border fragments. L indicates the DNA ladder; sizes are indicated at the side of the figure. Several lanes from the original scan have been omitted from this figure for clarity. Brightness and contrast have also been adjusted to optimize visibility of bands.

L Rm1021
Rm5000
PTH519 (130 kb pExo DNA)
1 2



L Rm1021
Rm5000
PTH519 (130 kb pExo DNA)
1 2

DNA
ladder
(kb)
10 —
8 —
6 —
5 —
4 —
3.5 —
3 —



CHAPTER 4. DISCUSSION

Genome Sequencing

At the time this work was initiated, the nucleotide sequences of the genomes of only four organisms had been published in their entirety. There are currently (December 1998) 18 published complete genome sequences, several other genomes completed but not published, and many more currently being sequenced with proposed finishing dates before the millenium. Thus far, the entire genomes of 12 prokaryotes, 4 archaeobacteria and 1 eukaryote have been sequenced. Information gained from these genomes has already provided invaluable resources for the detailed analysis of gene function and genome architecture, and will be an essential reference for all biologists.

R. meliloti, an agriculturally important soil microorganism capable of establishing nitrogen-fixing symbiotic associations with certain legumes, has an interesting genome infrastructure consisting of three replicons: one chromosome (3540 kb), and two symbiotic megaplasmids, pSym (1430 kb) and pExo (1700 kb). Given the sizes of these plasmids, genes required for symbiosis, such as *nod* and *nif* genes on pSym, and *exo* and *dct* on pExo, represent only a small fraction of their respective replicons. The fact that no *R. meliloti* strains have been cured of their megaplasmids suggests that these plasmids also encode functions required for viability.

Analysis of the recently sequenced 536 kb symbiotic plasmid of *Rhizobium* sp. NGR234 has revealed, on the basis of sequence homology, 416 predicted ORFs, including new putative symbiotic loci and signaling mechanisms, as well as genes involved in replication and conjugal transfer (Freiberg *et al.* 1997). Similarly, sequencing of the pExo replicon is hoped to significantly contribute to our understanding of the biology of *R. meliloti* and the symbiotic relationship it has with its host. The goal of this work was to devise a novel method for isolating defined regions of the pExo megaplasmid, which could then be sequenced by shotgun cloning into a suitable vector (e.g. pUC).

The work in this thesis involved: the construction of *FRT* targeting constructs; the preparation of a plasmid capable of expressing the Flp recombinase in *R. meliloti*; the use of this site-specific recombination system to generate deletions of large regions of the pExo megaplasmid; the transfer and capture of these regions in *E. coli*, ironically, in a Flp-independent fashion; and finally, the preliminary sequence analysis of one of these captured regions. For the sake of clarity, these latter issues are discussed below, in sequence.

The Flp/*FRT* Strategy

The development of a strategy to obtain large, defined pExo DNA inserts involved the use of the Flp/*FRT* *in vivo* recombination system. By constructing vectors carrying the *FRT* sites and a small region of the Tn5 IS50 element, it was possible to specifically target, via homologous recombination, an *FRT/oriV-ΩSp-oriT* cassette to selected Tn5 insertions. If only one targeting vector was used, the Flp-mediated excision

product would always carry a transposon along with it. For this reason, two nearly identical targeting vectors were made such that when combined in such a way that they are closest to one another, the intervening region would not include a transposon.

For Flp-mediated excision of the region flanked by direct *FRT* sites, it was necessary to express the Flp recombinase from a plasmid that would replicate in *R. meliloti*. Despite its extensive use *in vitro* (Gronostajski and Sadowski 1985; Pan *et al.* 1991; Proteau *et al.* 1986; Waite and Cox 1995), in eukaryotic systems (Broach *et al.* 1982; Golic and Lindquist 1989; Lloyd and Davis 1994; O’Gorman *et al.* 1991) and in *E. coli* (Buchholz *et al.* 1996; Cherepanov and Wackernagel 1995; Cox 1983; Pósfai *et al.* 1994; Pósfai *et al.* 1997; Snaith *et al.* 1996; Wild *et al.* 1996), Flp had not been shown to function in *R. meliloti*. It was therefore necessary to test several Flp-delivery plasmids, with *flp* downstream of various promoters, for Flp activity in *R. meliloti*.

The Flp Delivery Plasmids

To test for activity of the Flp recombinase in *R. meliloti*, strain RmH940, carrying the *E. coli lacZ* gene flanked by direct copies of the *FRT* site was constructed (refer to Figure 7). In initial experiments, little Flp activity was observed using a plasmid in which Flp was under the control of the tetracycline promoter of Tn10. We assume that this result is due to poor expression from this promoter in *R. meliloti*.

A different vector, pTH474, was constructed with a fragment carrying the *flp* gene cloned downstream of the *R. meliloti* inducible *pcaD* promoter. When tested, this vector appeared to express *flp* in a constitutive manner, regardless of the presence of protocatechuate, the inducing metabolite. This plasmid did however prove that the

Flp/*FRT* system was functional in *R. meliloti*. The fragment carrying the *flp* gene also contained 42 bp of the T₇ promoter. Since the same levels of Flp activity were observed from this fragment in the same vector without the *pcaD* promoter, it was reasoned that this small portion of the T₇ promoter was sufficient for “constitutive” *flp* expression. This would explain in turn, the lack of controlled induction of Flp activity in the presence of protocatechuate. The T₇ promoter is known for its tight transcriptional control mechanism and high level of expression (Mertens *et al.* 1995), and has been successfully used in *R. meliloti* GR4 to express nodulation efficiency genes (*nhe*) (Soto *et al.* 1993; Soto *et al.* 1994). It is unclear, however, how the fusion of this short promoter region with the upstream sequences of *pcaD* has created a “constitutive” promoter for the expression of *flp*.

In an effort to obtain controlled expression of *flp*, PCR products of *flp* and the *pcaD* promoter were fused such that the ATG start of *flp* replaced the presumed GTG start codon of the *pcaD* gene, creating a translational fusion at the start site (see Appendix B). This construct was confirmed by sequencing through the *pcaD* promoter into the *flp* gene. However, expression from the *pcaD* promoter has not been well studied and the start codon was deduced from alignments of the *R. meliloti* *pcaD* gene region with the *pcaD* gene from *Bradyrhizobium japonicum* and with the *pcaD* promoter region of *A. tumefaciens* (refer to Appendix B). The plasmid carrying this translational fusion, pTH506, was poorly expressed in *R. meliloti* even when induced, showing low levels of Flp activity throughout. It is possible that the deduced *pcaD* promoter region was mistakenly identified or incomplete, and that the PCR fragment contained a disrupted or

partly eliminated ribosome binding site or activator binding site. It is also possible that Flp activity is particularly susceptible to mutations in *flp*, such as those that may be caused during PCR.

The plasmid pTH474, which provided constitutive Flp activity, was effective in excising the pExo regions flanked by direct *FRT* sites. Although this plasmid proved useful, future study of the Flp/*FRT* system and its use in *R. meliloti* should begin with characterization of Flp expression and activity in *R. meliloti*. For example, it is not clear whether Flp has an unusually long life span and/or unusually high activity in *R. meliloti* once expressed. Further study into inducible promoters endogenous to *R. meliloti* should also be continued, including primer extension analysis and the identification of regulators. RNAase protection assays and Western blotting experiments with Flp antibodies would reveal the level of expression required for Flp to actively recombine two *FRT* sites.

Uses For the Flp/*FRT* System in *R. meliloti*

This work presents strong evidence that the Flp/*FRT* system can be used in *R. meliloti* to generate *in vivo* deletion of DNA sequences flanked by direct *FRT* sites. Although the Flp/*FRT* system was not required for the transfer of pExo DNA to *E. coli*, it was effective for determining orientation of *FRT* sites at different locations. Through this kind of deletion analysis, a 50 kb region that is essential for cell survival was also discovered. In much the same way, this system was used to generate other defined deletions of pExo DNA between two directly oriented *FRT* sites, and thus provides a method for analysis of gene function by deletion instead of transposon insertion.

Deletion analysis of large regions of pExo has already been very informative (Charles and Finan 1991); this system could be used to delete other regions including regions from the pSym megaplasmid and the chromosome.

Now that the Flp/*FRT* system has been shown to function in *R. meliloti*, it can also be used for site-specific integration of any gene or operon, endogenous or foreign, into the *R. meliloti* genome. Several *R. meliloti* strains already harbour *FRT* sites at known locations in the pExo megaplasmid and cloning vectors carrying an *FRT* site could easily be made. This could help immensely in the study of gene expression/gene function in *R. meliloti*. Stable integration into the genome would solve problems of plasmid loss in the absence of selection (such as in plant assays) and of copy number effect on expression. When studying exogenous gene expression, the Flp/*FRT* system of site-specific recombination may prove to be an effective technique of targeting these genes to the genome.

Flp-Independent Transfer of *FRT*-Flanked pExo DNA

Although the Flp-delivery vector with apparent constitutive Flp activity (pTH474) was to be used for excision of large regions of pExo for subsequent transfer into *E. coli*, it became evident that excision prior to transfer was not required. As long as the *oriT* sites, which accompanied the *FRT* sites on the targeting vectors, were also in direct orientation, transfer of the region flanked by *oriT* and *FRT* sites into *E. coli* could be initiated upon delivery of the *tra* gene-functions. There is a precedent for the conjugal transfer of bacterial plasmid and chromosomal DNA (Simon 1984; Yakobson and Guiney 1984).

Single stranded conjugal transfer is thought to occur via a replicative rolling circle mechanism, beginning and ending at *oriT* sites, the only known *cis*-acting function required for conjugative DNA transfer (reviewed by Willetts and Wilkins 1984). Transfer is initiated by site- and strand-specific nicking of the double stranded DNA at the *oriT*, whose cleaved strand (5' terminus) is covalently attached to the TraI protein, the relaxase (Pansegrau *et al.* 1990). It is believed that the TraI relaxase possesses cleaving-joining activities, and plays an important role in the initiation and termination of conjugative DNA transfer by scanning for the occurrence of a second cleavage site (*oriT*) at the donor-recipient interface (Lanka and Wilkins 1995; Pansegrau *et al.* 1993; Sherman and Matson 1994). In this way, transfer of the desired pExo region would occur just as if excision had already transpired, beginning at one *oriT* and ending at the other.

An alternative explanation may be recombination between these two *oriT* sites after, or upon, transfer of the large pExo megaplasmid regions. It has been previously observed that site-specific recombination occurs between *oriT* (or *oriT*-like) regions of several conjugative plasmids, in a RecA-independent, transfer-dependent fashion (Brasch and Meyer 1987; Broome-Smith 1980; Horowitz and Deonier 1985; Warren and Clark 1980). It is clear that the mechanism of transfer of these large regions of pExo DNA into *E. coli* warrants further investigation.

A natural origin of transfer as well as a cluster of genes homologous to the conjugal transfer genes of *Agrobacterium* Ti plasmids have been recently found on the megaplasmid of *Rhizobium* sp. NGR234 (Freiberg *et al.* 1997). Natural megaplasmid-based origins of conjugal transfer have also been discovered in *R. meliloti* GR4 (Herrera-

Cervera *et al.* 1998), where pRmeGR4a is known to be self-transmissible and is able to mobilize pRmeGR4b *in trans* (Herrera-Cervera *et al.* 1996). Although the pExo megaplasmid of *R. meliloti* SU47 is transmissible (Charles 1990), its transfer frequency ($\sim 8 \times 10^{-9}$) is not sufficient to interfere with the RK2-*oriT* facilitated transfer. Transfer from a pExo *oriT* site to one of the integrated RK2 *oriT* sites would not likely occur since the TraI-mediated cleavage and ligation reactions are sequence specific (Sherman and Matson 1994; Willetts and Wilkins 1984).

Therefore, in this strategy, Flp-mediated excision would likely have been necessary if only one *oriT* site, or none, were used for construction of donor *R. meliloti* strains. However, several other factors had to be considered, such as how to introduce the targeting vectors into *R. meliloti* and how to transfer the large excised pExo region into *E. coli*. Although it has been found that very large plasmids can be efficiently electroporated into *E. coli* (Leonardo and Sedivy 1990), the transformation of the targeting vectors into *R. meliloti* would have to occur at a sufficiently high frequency to observe homologous recombination, and avoid host restriction mechanisms.

Other Features of the Targeting Vectors

Three different *E. coli* strains were used as recipients of the large plasmids carrying pExo DNA: DH5 α , JW192 (DH5 α *trfA*) (Ap^r) and MT620 (Rf^r). Replication of the 50 kb region, without pBS or pBAC sequences, was limited to JW192 (the DH5 α strain supplying TrfA), presumably by TrfA-initiated replication at the *oriV* from RK2. However, when either pBS or pBAC was present, the large ~ 60 kb plasmid could

replicate in any of the three backgrounds. Replication of the 130 kb region in all three *E. coli* strains was observed only with the presence of the pBAC backbone.

The *oriV* of RK2 was first included in the targeting vectors to allow replication of the excised large regions. However, the *oriV* was not essential for the maintenance of the 50 kb region, excised along with the pBS or pBAC backbone, and for the 130 kb region, the *oriV* did not appear to be sufficient for replication. RK2 is a 60 kb plasmid which replicates from the *oriV* at a copy number of 5 to 8 copies per chromosome in *E. coli* (Figurski *et al.* 1979). It was hoped that the *oriV* of RK2 would be able to sustain larger fragments, like the 130 kb region. The pBAC backbone was chosen to replace the *oriV* of RK2 as an origin of replication for large plasmids. The BAC cloning system is based on the *E. coli* mini F plasmid, whose replication and partitioning is stringently controlled at a maximum 2 copies per cell (Timmis *et al.* 1975). This control over copy number reduces the potential for recombination between DNA fragments (Brosch *et al.* 1998), avoids lethal overexpression of foreign genes and ensures the stable maintenance of large inserts.

The targeting vectors were constructed to include the recognition site for the meganuclease I-*SceI*. This site is a very useful tool for the study of genomes and genome organization, and was placed in the targeting vectors in such a way that, after retrieving large *R. meliloti* inserts, the cassette (*oriV/FRT-ΩSp-oriT*) and backbone (pBAC or pBS) could be separated from the pExo region to allow for purification of the insert.

The strategy presented here could be applied in a near-identical fashion to the other replicons in *R. meliloti*. With minor modifications, it may also be sensible to use

this system to obtain large plasmids for sequencing from other organisms in which genetic manipulation with transposons is possible.

The Replicator Region of the pExo Megaplasmid

It was intriguing that no Gm^r pTH474 (“constitutive Flp”) transconjugants were observed for those strains carrying direct *FRT* sites flanking the region between Ω 5056 and Ω 5069. Gm^r transconjugants were obtained from other, even larger, regions that were flanked by direct *FRT* sites. Screening for loss of the transposons indicated that deletions had occurred in these transconjugants. This suggests that Flp-mediated separation of the 50 kb region from the rest of the megaplasmid is lethal. With this evidence and the observed stable maintenance of the 50 kb region in *A. tumefaciens*, it was postulated that this portion carries the replicator region of pExo.

Ongoing sequencing of the 50 kb region, has revealed putative proteins, encoded within this 50 kb region, with strong homology to the replication proteins, RepA, RepB and RepC, of several large plasmids in the related microorganisms *Rhizobium* sp. NGR234 (Freiberg *et al.* 1997), *R. etli* (Ramirez-Romero *et al.* 1997), *R. leguminosarum* (Turner and Young 1995), *A. tumefaciens* (Suzuki *et al.* 1998; Tabata *et al.* 1989) and *A. rhizogenes* (Nishiguchi *et al.* 1987). Replication of the pExo megaplasmid probably begins at *oriV* (see Figure 21), which in all of these replicator regions, is located within the intergenic region between the *repB* and *repC* genes. The *repBC* intergenic region (putative *oriV*) of pExo is highly similar to the conserved predicted origins of replication of *Agrobacterium* and *Rhizobium* plasmids (refer to Figure 21).

Figure 21: A multiple alignment of the nucleotide sequence of large plasmid replication origins. The putative pExo origin of replication, obtained through sequencing of pTH482 (the Ω 5056- Ω 5069 50 kb region), is aligned with known replication origins. Shown are the replication origins of: the Ti plasmid (pTiB6S3) of *A. tumefaciens* B6S3 (Tabata *et al.* 1989); the Ti plasmid (pTi-SAKURA) of *A. tumefaciens* MAFF301001 (Suzuki *et al.* 1998); the pRmeSU47b megaplasmid (pExo) of *R. meliloti* SU47; the Ri plasmid (pRiA4b) of *A. rhizogenes* (Nishiguchi *et al.* 1987); the symbiotic plasmid (p42d) of *R. etli* CFN42 (Ramirez-Romero *et al.* 1997); the symbiotic plasmid (pNGR234) of *R. meliloti* NGR234 (Freiberg *et al.* 1997); and the cryptic plasmid (pRL8JI) of *R. leguminosarum* (Turner and Young 1995). The origins of pNGR234 and pTi-SAKURA were also discovered during the course of DNA sequencing projects (Freiberg *et al.* 1997; Hattori *et al.* 1997; Suzuki *et al.* 1997; Suzuki *et al.* 1998). These two putative replication origins have been determined by sequence homology only. Gaps introduced to give the best sequence alignments are marked with hyphens (-). Conserved nucleotide sequences are in blue; nucleotides identical for all 7 sequences are marked with a star (*).

pTiB6S3	1	CCGCAAAAGAAAAAGGCCCCCGAA--ACG-GCGTTCCGGAAGACCTTCTCTATAGTCTCG
pTi-SAKURA	1	CCGCAAAAGAAAAAGGCCCCCGAA--ACG-GCGTTCCGGAAGACCTTCTCTGTAGTCTCG
pExo	1	GAGCAAAAGAAAAAGGCTTCCGAACGACTAGCGCTGCGGAAACCCCTTCTCT--CGTG TAG
pRiA4b	1	CCGCAAAAGAAAAAGAGCCCCCTCAACGTCGCCGTGCGGAAAGCCCTTCTGTC-TCTCTAG
p42d	1	CCGCAAAAGAAAAAGAGCTCCCTCAACGTCGCCGTGCGGAAAGCCCTTCTGTC-TCTGTAG
pNGR234	1	-CGCAAAAGAAAAAGGCCCCAGACGGTAACCATCGTGGAAGCCTCTCTCA-TCGTTTAG
pRL8JI	1	GCGCAA-GAAAAAGCCCTCCGAA--ACGGTGGTTCAGAAGGCC-TCTCTC-AGTTTGG
		***** ***** * ** * *** * *** * * *

pTiB6S3	58	CAGCTAAGAGAATCGCATTTCAGGAATCCAGTCAAGAGTCCCGTGAGG-AAAGTATCG
pTi-SAKURA	58	CAGCTAAGAGAATCGCATTTCAGGAATCGTAGTCAAGGGTCCCGTAAGGGAAAGCGTCA
pExo	59	CAACTA-GAGAATCCCACTTCCGCGAATCAGTGTC AAGCATTTT-TGACG-----TCG
pRiA4b	60	C GCGAA-CAGAATCGCATTTCCTCGAATCCTCGTCAAGAGT TTT-TAGCG-----CCG
p42d	60	CAAGAA-CAGAATCGCATTTCCTCGAATCCTCGTCAAGAGTCTT-TGGCG-----CCG
pNGR234	59	CAGCCT-GAGAATCGCATTTCACGAATCGCAGTCAAGAGTCTT-TGGCA-----CCGG
pRL8JI	56	TCGCTTAGAGAATCGCATTTCGGAATCACAGTCAAGAGTC----AACG-----CCA
		***** ** ** * ***** ***** *

pTiB6S3	117	TTTCGACGAGCTGATTTCTTTTGCCTAACAAA-GGTACAAGGAA	159
pTi-SAKURA	118	TTTCGACGGCGGATTTCAATTGCCTAACAAAAGGTAAAAGGAA	161
pExo	110	TTTCGGCGAACGGATTTCTTTTGCCTTGGA AAAAGGTGAA GAAAA	155
pRiA4b	111	TTTTGGTGAGCTGATTTCTTTTGCCTGCTGAAAGGTGAAAG---	151
p42d	111	TTTTGGTGAGCGAATTTCTTTTGCCTGCTGAAAGGTGAGAGACA	154
pNGR234	111	AACGGGTGAGCGGACGTCTTTTGCCTGAGGATAGGTGAAAGAAG	154
pRL8JI	105	CACCGGCGTAGCCTTTTC-TTTGCCTTGC GAAAGGTGAAGGAC-	146
		* * ** ***** * ***	

Another putative replication origin from a different region of the pExo megaplasmid has been reported by Margolin and Long (1993). This region was found to be sufficient for conferring replicative functions to normally non-replicative plasmids. However, this region is not required for pExo replication, and the fragment carrying this putative origin does not share high homology with the above-mentioned plasmid origins of replication. Also, this region does not appear to contain the *repBC* genes; the two putative ORFs reported within this region do not reveal homology to any sequence in the GenBank databases as determined by BLAST searches.

Further Sequencing

Sequence from the *EcoRI* fragments has also confirmed the location from which the 50 kb region was obtained. Since the Ω 5056 insertion results in a Lac⁻ phenotype, one of the *EcoRI* fragments was expected to contain the borders of the β -galactosidase gene and the IS50/targeting vector (refer to Figure 14). As reported in Table 7, plasmid pTH528 carries an *EcoRI* fragment containing sequences nearly identical to both the Ω Sp interposon and the previously sequenced 2267 bp β -galactosidase gene of *R. meliloti* (Fanning *et al.* 1994). Several other potential genes have been found through the sequencing of a random shotgun library of the 50 kb region. A library of random 1 to 2 kb fragments of the 130 kb region is in the process of being made for sequencing.

Comparative Genomics

The information obtained from analysis of the nucleotide sequence of pExo will help to elucidate the role of this megaplasmid and may also promote further

understanding of the life of *R. meliloti* in symbiotic and soil environments. This sequence information is also valuable in comparative genomics. Members of the *Brucella* (animal pathogens) and *Agrobacterium* (plant pathogens) genera are phylogenetically related to *R. meliloti* (De Lay *et al.* 1987; Jumas-Bilak *et al.* 1998; Moreno *et al.* 1990; Sola-Landa *et al.* 1998). Whereas *Agrobacterium* spp. are gram-negative plant pathogens which transform their host with transferred DNA (reviewed by Long and Staskawicz 1993), *Brucella* spp. are gram-negative bacteria which cause chronic disease in humans and other animals that results from pathogen proliferation inside macrophages (reviewed by Sangari and Aguero 1996; Smith and Ficht 1990). By comparison with the genomes of members of the *Brucella* and *Agrobacterium* genera, genes such as those required for pathogenicity, host recognition, cell invasion, and intracellular survival, may be identified for further study.

The initial step of this sequencing project has been achieved with the construction of a system whereby large specific regions of the pExo megaplasmid can be excised and/or transferred to *E. coli* for construction of shotgun libraries. However, nucleotide sequence information alone is insufficient to predict gene function; although genome sequencing projects reveal much information, in many ways they are a beginning—much genetic and biochemical research remains to be done.

APPENDIX A

R. meliloti strains carrying two targeting vectors, each in a different Ω insertion location on the pExo megaplasmid are described in Table 7. The structural details, demonstrating the relative locations of the targeting vectors with respect to the transposons, is diagrammed in Figure 22. The locations of the targeted transposons relative to one another is represented on circular and linear maps in Figure 12.

Table 7: "Sets" of *R. meliloti* strains carrying two targeting vectors

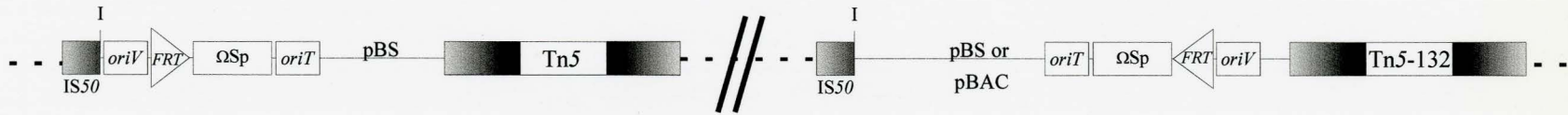
Location of targeted Tn5 and Tn5-132 ("sets") in a Rm5000 or Rm1021 background	Strain Name	Diagram # in Figure 19*
Ω5007::TnV::pTH455 Ω5035::Tn5-132::pTH456 in Rm5000	RmH990	4
	RmH991	3
Ω5056::Tn5::pTH455 Ω5069::Tn5-132::pTH456 in Rm5000	RmH992	1
	RmH993	2
	RmH994	3
	RmH995	4
Ω5102::Tn5::pTH455 Ω5142::Tn5-132::pTH456 in Rm5000	RmK126	4
	RmK127	3
	RmK128	2
	RmK129	1
Ω5159::Tn5::pTH455 Ω5142::Tn5-132::pTH456 in Rm5000	RmK130	2
	RmK131	1
	RmK132	4
	RmK133	3
Ω5056::Tn5::pTH455 Ω5142::Tn5-132::pTH456 in Rm1021	RmK137	2
	RmK138	1
	RmK139	4
	RmK140	3
Ω5102::Tn5::pTH455 Ω5142::Tn5-132::pTH504 in Rm1021	RmK165	1
	RmK166	2
	RmK167	3
	RmK168	4
Ω5159::Tn5::pTH455 Ω5142::Tn5-132::pTH504 in Rm1021	RmK169	2
	RmK170	1
	RmK171	4
	RmK172	3
Ω5056::Tn5::pTH455 Ω5142::Tn5-132::pTH504 in Rm1021	RmK173	2
	RmK174	1
	RmK175	4
	RmK176	3
Ω5102::Tn5::pTH455 Ω5069::Tn5-132::pTH504 in Rm1021	RmK184	4
	RmK185	3
Ω5159::Tn5::pTH455 Ω5069::Tn5-132::pTH504 in Rm1021	RmK186	4
	RmK187	3
Ω5159::Tn5::pTH455 Ω5069::Tn5-132::pTH504 in Rm1021	RmK188	1
	RmK189	2
	RmK190	3
	RmK191	4
Ω5159::Tn5::pTH455 Ω5069::Tn5-132::pTH509 in Rm1021	RmK197	4
	RmK198	3
	RmK199	2
	RmK200	1
Ω5056::Tn5::pTH455 Ω5069::Tn5-132::pTH509 in Rm1021	RmK201	1
	RmK202	2
	RmK203	3
	RmK204	4

* The structural details (including the location and directionality of the *FRT* sites in the *IS50* elements) of the four combinations of two targeted insertions are numbered 1 to 4 in Figure 22.

Figure 22: Schematic representation of the four possible combinations of Tn5::pTH455 with either pTH456, pTH504 or pTH509. The combinations are numbered 1 to 4. Number 1 and 4 carry indirect *FRT* sites. Number 2 carries the *FRT* sites in parallel in the outer *IS50* elements, whereas number 3 carries the *FRT* sites in parallel in the inner *IS50* elements. The triangle shows *FRT* directionality. The *IS50* elements are shaded white (outside end) to black (inside end), indicating directionality. The locations of these elements, the plasmid backbones and the *I-SceI* sites (I) are indicated. pExo DNA is represented by the dashed line.

pExo DNA

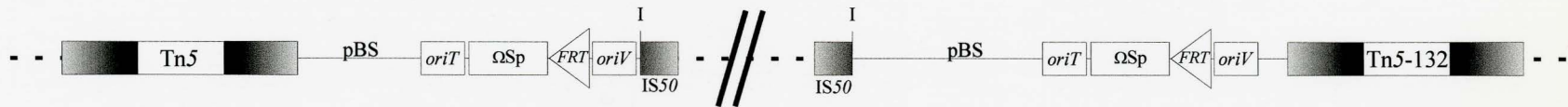
1 inverted *FRT* sites



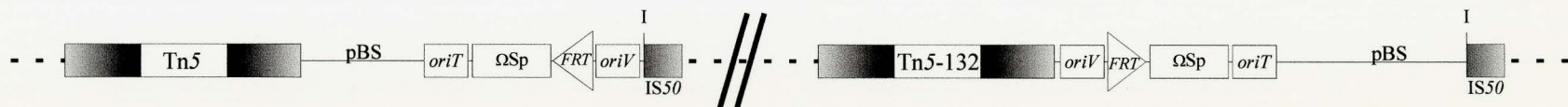
2 parallel *FRT* sites in outer IS50s



3 parallel *FRT* sites in inner IS50s



4 inverted *FRT* sites



APPENDIX B

In an effort to obtain controlled flp expression in *R. meliloti*, two fusions with the *R. meliloti* *pcaD* promoter were constructed. The putative GTG start sites of the divergently expressed *pcaD* and *pcaQ* genes (indicated in Figure 23) were deduced from alignments with the PcaQ amino acid sequence of *A. tumefaciens* (where *pcaD* and *pcaQ* are also divergently expressed) and the PcaD amino acid sequence of *Bradyrhizobium japonicum* (see Figures 24 and 25). Using this information, the primers used in PCR to amplify the *pcaD* promoter were designed (indicated in Figure 23). Although both alignments suggest the same GTG start, it is possible that the chosen *pcaD* promoter in pTH506 may have been mistakenly selected.

Figure 23: DNA sequence of the *pcaD/pcaQ* intergenic region. Putative GTG start sites for *pcaD* (boxed) and *pcaQ* (shaded) are indicated. Annealing sites for PCR primer sets (AB11717 with AB11718, and AB13243 with AB13244) are underlined and labeled (refer to Table 4).

pcaD/pcaQ Intergenic Region

CCGACCGACGAGGT CGAAGATCGCCGATGCCCAATTGTT CGAGAAGCACTGCGTTTTTCACCGG
GGCTGGCTGCTCCAGCTTCTAGCGGCTACGGGTAAACAAGCTCTTCGTGACGCAAAAAGTGGCC

TGACGATCTTGATGCGTGCGCCGGTCTTCTCTTT CAGGAAGAGCGCGATGGCGCGGGCATGA
ACTGCTAGAACTACGCACGCGGCCAGAAGAGAAAAGTCTTCTCGCGCTACCGCGCGCCGTACT

TGCGGGT CGAAAACGGT CGGAAGCGCACCGACGCGGATCGGC GGGCCTTCGCCGGACCGCTCC
ACGCCCAGCTTTGCCAGCCTTCGCGTGGCTGCGCCTAGCCGCCCGGAAGCGGCCTGGCGAGG

AB11718

TGCGAGACGGAATCGAGGCCTTGGCNCAGCGCCGT CAGCGCGGCCAGCATGGCGCAGAAA
ACGCTCTGCCTTAGCTCCGGAACCGNGTCGCGGCAGTCGCGCCGCGGTCGTACCGCGTCTTT

GACCTCACCGTAGCGCGTGATCTTGATGCCACGCCCTCGCGCTCGAAAACCGCGACGCCAG
CTGGAGTGGCATCGCGCACTAGAACTACGGTGC GGGGAGCGCGAGCTTTTGCGCTGCGGGTC

AACCTCTTCGAGTTCCCGGATCGTCTTGGTGACCGCCGGTTGCGTGACATGCAGGAGTTCCGGC
TTGGAGAAGCTCAAGGGCCTAGCAGAACCCTGGCGCCAACGCACTGTACGTCTCAAGCCG

AB13244

CGCCTTACCACGCTCTTCTGGCGTGCGACCTCGACAAATGTCTGCAGATGGCGAAACTTAAC
GCGGAAGTGGTGC GAGAAGACCGCACGCTGGAGCTGTTTACAGACGTCTACCGCTTTGAATTG

pcaQ ←

GCGAGCGTCGATCACGGATCGTATAACCTCCTGGTTAAGGGAAAGCCACGAAATATCATTTTA
CGCTCGCAGCTAGTGCCTAGCATATTGGAGGACCAATTCCTTTTCGGTGCTTTATAGTAAAAT

CCTAACCGGATGAAACATCAAATCTGACGACGGAGGAGAGCTGCC GTG CAATTCACCCGCA
GGATTGGCCTACTTTGTAGGTTTAGACTGCTGCCTCCTCTCGACGG CAC GTTAAGTGGGCGT

AB13243

→ *pcaD*

TCAACGACGTCACGATTCACTATCGCGTGGTTGGCGCGGTCACGGAAAAGCCCGCGCTCGTCT
AGTTGCTGCAGTGCTAAGTGATAGCGACCAACCGCGCCAGTGCCTTTTCGGGCGCGAGCAGA

TCATCAACTCGCTCGGCACGGATTTCCGTATCTGGCGCGATGTGCTCCTCCGGCTGGCCGGCG
AGTAGTTGAGCGAGCCGTGCCTAAAGGCATAGACCGGCTACAGCAGGAGCCGACCGGCCG

AB11717

ATTTTGCCATCGTGCTCTATGACAAGCGTGGCCATGGGC
TAAAACGGTAGCACGAGATACTGTTTCGACCGGTACCCG

Figure 24: Alignment of the PcaQ amino acid sequences and the *pcaD/pcaQ* intergenic regions of *A. tumefaciens* and *R. meliloti*. The putative GTG start sites for *pcaD* (boxed) and *pcaQ* (shaded) are indicated. Identical nucleotides and amino acids are indicated with a star (*), while similar amino acids are indicated with a colon (:).

pcaQ alignment

A. tumefaciens

TGCATGGCATG--CTCCTCCCTGCGAAGACTATTGCAAGCCTGTTCCGGTTAAGTAAAATGACATTAT
**** *

R. meliloti

TG[CAC]GGCAGCTCTCCTCCGTCGTCAGATT--TGGATGTTTTTCATCCGGTTAGGTAAAATGATATT--

A. tumefaciens

M V D Q R I K F R H L
TGATGGAAATCA--TAACCTGTGGATTATATGAGCGATGGTCGATCAGCGTATCAAGTTCCGGCATCTG
* *

R. meliloti

TCGTGGCTTTCCCTTAACCAGGAGTTATACGATCCGTGATCGACGCTCGCGTTAAGTTTCGCCATCTG
M I D A R V K F R H L
: : * * : * * * * * *

A. tumefaciens

Q T F V E V A R Q K S V I R A A E I L H V S Q
CAGACCTTTGTGGAAGTGGCGCGACAGAAAAGCGTCATCCGCGCTGCGGAAATCCTGCATGTCAGCCAG

R. meliloti

CAGACATTTGTGAGGTCGCACGCCAGAAGAGCGTGGTGAAGGCGGCCGAACTCCTGCATGTCACGCAA
Q T F V E V A R Q K S V V K A A E L L H V T Q
* * * * * * * * * * * * * * * : : * * * : * * * * : *

A. tumefaciens

P A V T K T I R E L E D V L G V S L L E R E G V
CCGGCGGTGACGAAGACCATTTCGCGAGCTGGAAGACGTCTCGCGGTGTCGCTGCTGGAGCGGGAGGGCG

R. meliloti

CCGGCGGTACCAAGACGATCCGGGAACTCGAAGAGTTTCTGGGCGTTCGCGGTTTTTCGAGCGCGAGGGGCGT
P A V T K T I R E L E E V L G V A V F E R E G V
* * * * * * * * * * * * * * * : * * * * : : : * * * * *

Figure 25: Alignment of the PcaD amino acid sequences and their corresponding DNA sequences. The putative GTG start sites for *pcaD* (boxed) and *pcaQ* (shaded) are indicated. Identical nucleotides and amino acids are indicated with a star (*), while similar amino acids are indicated with a colon (:).

pcaD alignment

B. japonicum -----GCTCG-ACAGCCTCGAGCGAGATTGATTGTTTGGAGCATGATCCCTTCGGAAAACCGTTCCACACTTTT
* * * * *
R. meliloti CACGGATCGTATAACCTCCTGGTTAAGGGAAAGCCAC-GAAATATCATTTTACCTAACCGGATGAAACA--T

B. japonicum M P M I D A D G C L I N V S V
CCGGATCATGCTCCAGAACGGAGACGCCGCATGCCATGATCGATGCCGACGGTTGCCTGATCAACGTCTCCGTC
** * * * * *
R. meliloti CCAAATCTGACGACGGAGGAGAGCTGCC--GTGCAATTCACCCGCATCAACGACGTACGATCACTATCGCGTG
M Q F T R I N D V T I H Y R V
: : : : * :

B. japonicum E G R D - G G P T L M L S N S L G C T L Q M W E P
GAAGGCCGCGAC---GGCGGGCCGACCTTGATGCTGTCCAATTGCTCGGCTGCACGCTTCAGATGTGGGAGCCG

R. meliloti GTTGGCGCGGTACGGAAAAGCCCGCGCTCGTCTTCATCAACTCGCTCGGCACGGATTTCCGTATCTGGCGCGAT
V G A V T E K P A L V F I N S L G T D F R I W R D
* : * : : * * * * : : : * :

B. japonicum Q M K A L T Q V F R V I R Y D R R G H G
CAGATGAAGGCGCTGACACAGGTGTTCCGCGTCATCCGCTATGACCGCCGCGGCCACGGC

R. meliloti GTCGTCCTCCGGCTGGCCGGCGATTTTGCCATCGTGCTCTATGACAAGCGTGGCCATGGG
V V L R L A G D F A I V L Y D K R G H G
: * : * : : * * : * * * *

REFERENCES

- Alting-Mees, M.A. and J.M. Short.** 1989. pBluescript II: gene mapping vectors. *Nucleic.Acids.Res.* **17**:9494-
- Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic.Acids.Res.* **25**:3389-3402.
- Andersson, S.G., A. Zomorodipour, J.O. Andersson, T. Sicheritz-Ponten, U.C. Alsmark, R.M. Podowski, A.K. Naslund, A.S. Eriksson, H.H. Winkler, and C.G. Kurland.** 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**:133-140.
- Andrews, B.J., L.G. Beatty, and P.D. Sadowski.** 1986. Site-specific recombination of the yeast plasmid two-micron circle: intermediates in the binding process. *Basic.Life Sci.* **40**:407-424.
- Andrews, B.J., L.G. Beatty, and P.D. Sadowski.** 1987. Isolation of intermediates in the binding of the FLP recombinase of the yeast plasmid 2-micron circle to its target sequence. *J.Mol.Biol.* **193**:345-358.
- Andrews, B.J., G.A. Proteau, L.G. Beatty, and P.D. Sadowski.** 1985. The FLP recombinase of the 2 micron circle DNA of yeast: interaction with its target sequences. *Cell* **40**:795-803.
- Argos, P., A. Landy, K. Abremski, J.B. Egan, E. Haggard-Ljungquist, R.H. Hoess, M.L. Kahn, B. Kalionis, S.V. Narayana, and L.S. Pierson.** 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J.* **5**:433-440.
- Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Siedman, J.A. Smith, and K. Struhl.** 1989. Short protocols in molecular biology. Green Publishing Associates and John Wiley & Sons, Inc., New York.
- Banfalvi, Z., E. Kondorosi, and A. Kondorosi.** 1985. *Rhizobium meliloti* carries two megaplasmids. *Plasmid* **13**:129-138.

Banfalvi, Z., V. Sakanyan, C. Koncz, A. Kiss, I. Dusha, and A. Kondorosi. 1981. Location of nodulation and nitrogen fixation genes on a high molecular weight plasmid of *R. meliloti*. *Mol.Gen.Genet.* **184**:318-325.

Bardin, S., S. Dan, M. Osteras, and T.M. Finan. 1996. A phosphate transport system is required for symbiotic nitrogen fixation by *Rhizobium meliloti*. *J.Bacteriol.* **178**:4540-4547.

Bardin, S.D., R.T. Voegele, and T.M. Finan. 1998. Phosphate assimilation in *Rhizobium (Sinorhizobium) meliloti*: identification of a pit-like gene. *J.Bacteriol.* **180**:4219-4226.

Berg, C.M. and D.E. Berg. 1987. Uses of transposable elements and maps of known insertions. pp. 1071-1107. *In*, F.C. Niedhardt, J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaecter, and H.E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D. C.

Berg, D.E. and C.M. Berg. 1983. The procaryotic transposon Tn5. *Bio/Technology* **1**:417-435.

Berg, D.E., J. Davies, D. Allet, and J.-D. Rochaix. 1975. Transposition of R factor genes to bacteriophage . *Proc.Natl.Acad.Sci.U.S.A.* **72**:3628-3632.

Berg, D.E., C. Egner, B.J. Hirschel, J. Howard, L. Johnsrud, R.A. Jorgensen, and T.D. Tlsty. 1980. Insertion, excision and inversion of Tn5. *Cold Spring Harb.Symp.Quant.Biol.* **45**:115-123.

Blattner, F.R., G. Plunkett, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew, J. Gregor, N.W. Davis, H.A. Kirkpatrick, M.A. Goeden, D.J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* **277**:1453-1474.

Bloch, C.A., C.K. Rode, V.H. Obreque, and J. Mahillon. 1996. Purification of *Escherichia coli* chromosomal segments without cloning. *Biochem.Biophys.Res. Commun.* **223**:104-111.

Bolivar, F., R.L. Rodriguez, P.J. Greene, M.C. Betlach, H.L. Heyneker, and H.W. Boyer. 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* **2**:95-113.

Boucher, C.A., A. Martinel, P.A. Barberis, G. Alloing, and C. Zischek. 1986. Virulence genes are carried by a megaplasmid of the plant pathogen *Pseudomonas solanacearum*. *Mol.Gen.Genet.* **205**:270-275.

Bouck, J., W. Miller, J.H. Gorrell, D. Muzny, and R.A. Gibbs. 1998. Analysis of the quality and utility of random shotgun sequencing at low redundancies. *Genome Res.* **8**:1074-1084.

Brasch, M.A. and R.J. Meyer. 1987. A 38 base-pair segment of DNA is required in *cis* for conjugative mobilization of broad host-range plasmid R1162. *J.Mol.Biol.* **198**:361-369.

Broach, J.R. 1981. The yeast plasmid 2 μ circle. pp. 445-470. *In*, J.N. Strathern, E.W. Jones, and J.R. Broach (ed.), *The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance*. Cold Spring Harbor, New York.

Broach, J.R., V.R. Guarascio, and M. Jayaram. 1982. Recombination within the yeast plasmid 2 μ circle is site-specific. *Cell* **29**:227-234.

Broach, J.R. and J.B. Hicks. 1980. Replication and recombination functions associated with the yeast plasmid, 2 μ circle. *Cell* **21**:501-508.

Bromfield, E.S., N.P. Thurman, S.T. Whitwill, and L.R. Barran. 1987. Plasmids and symbiotic effectiveness of representative phage types from two indigenous populations of *Rhizobium meliloti*. *J.Gen.Microbiol.* **133**:3457-3466.

Broome-Smith, J. 1980. RecA independent, site-specific recombination between ColE1 or ColK and a miniplasmid they complement for mobilization and relaxation: implications for the mechanism of DNA transfer during mobilization. *Plasmid* **1**:51-63.

Brosch, R., S.V. Gordon, A. Billault, T. Garnier, K. Eiglmeier, C. Soravito, B.G. Barrell, and S.T. Cole. 1998. Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. *Infect.Immun.* **66**:2221-2229.

Buchholz, F., L. Ringrose, P.O. Angrand, F. Rossi, and A.F. Stewart. 1996. Different thermostabilities of FLP and Cre recombinases: implications for applied site-specific recombination. *Nucleic.Acids.Res.* **24**:4256-4262.

Buikema, W.J., S.R. Long, S.E. Brown, R.C. van den Bos, C. Earl, and F.M. Ausubel. 1983. Physical and genetic characterization of *Rhizobium meliloti* symbiotic mutants. *J.Mol.Appl.Genet.* **2**:249-260.

Bult, C.J., O. White, G.J. Olsen, L. Zhou, R.D. Fleischmann, G.G. Sutton, J.A. Blake, L.M. FitzGerald, R.A. Clayton, J.D. Gocayne, A.R. Kerlavage, B.A. Dougherty, J.F. Tomb, M.D. Adams, C.I. Reich, R. Overbeek, E.F. Kirkness, K.G. Weinstock, J.M. Merrick, A. Glodek, J.L. Scott, N.S.M. Geoghagen, (17 other authors), and J.C. Venter. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**:1058-1073.

- Burkardt, B. and H.J. Burkardt.** 1984. Visualization and exact molecular weight determination of a *Rhizobium meliloti* megaplasmid. *J.Mol.Biol.* **175**:213-218.
- Burkardt, B., D. Schillik, and A. Puhler.** 1987. Physical characterization of *Rhizobium meliloti* megaplasms. *Plasmid* **17**:13-25.
- Cai, L., J.F. Taylor, R.A. Wing, D.S. Gallagher, S.S. Woo, and S.K. Davis.** 1995. Construction and characterization of a bovine bacterial artificial chromosome library. *Genomics* **29**:413-425.
- Carle, G.F. and M.V. Olson.** 1984. Separation of chromosomal DNA molecules from yeast by orthogonal-field- alternation gel electrophoresis. *Nucleic.Acids.Res.* **12**:5647-5664.
- Casse, F., C. Boucher, J.S. Julliot, M. Michel, and J. Dénarié.** 1979. Identification and characterization of large plasmids in *Rhizobium meliloti* using agarose gel electrophoresis. *J.Gen.Microbiol.* **113**:229-242.
- Chain, P.S.G., K. Napper, and T.M. Finan.** 1997. Mineral phosphate solubilization by *Rhizobium meliloti*. pp. 504. *In*, C. Elmerich, A. Kondorosi, and W. I. Newton (ed.), *Biological Nitrogen Fixation for the 21st century*. Kluwer Academic Publishers, The Netherlands.
- Charles, T.C.** 1990. Construction of a genetic linkage map of the *Rhizobium meliloti* 1600 kilobase megaplasmid pRmeSU47b, generation of defined megaplasmid deletions, and study of megaplasmid-borne genes. McMaster University, Ph.D. thesis.
- Charles, T.C. and T.M. Finan.** 1990. Genetic map of *Rhizobium meliloti* megaplasmid pRmeSU47b. *J.Bacteriol.* **172**:2469-2476.
- Charles, T.C. and T.M. Finan.** 1991. Analysis of a 1600-kilobase *Rhizobium meliloti* megaplasmid using defined deletions generated *in vivo*. *Genetics* **127**:5-20.
- Charles, T.C., W. Newcomb, and T.M. Finan.** 1991. *ndvF*, a novel locus located on megaplasmid pRmeSU47b (pEXO) of *Rhizobium meliloti*, is required for normal nodule development. *J.Bacteriol.* **173**:3981-3992.
- Cherepanov, P.P. and W. Wackernagel.** 1995. Gene disruption in *Escherichia coli*: Tc^R and Km^R cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158**:9-14.

Churcher, C., S. Bowman, K. Badcock, A. Bankier, D. Brown, T. Chillingworth, R. Connor, K. Devlin, S. Gentles, N. Hamlin, D. Harris, T. Horsnell, S. Hunt, K. Jagels, M. Jones, G. Lye, S. Moule, C. Odell, D. Pearson, M. Rajandream, P. Rice, N. Rowley, J. Skelton, V. Smith, S. Walsh, S. Whitehead, and B. Barrell. 1997. The nucleotide sequence of *Saccharomyces cerevisiae* chromosome IX. *Nature* **387(suppl.):84-87.**

Clark, S. 1998. Construction of a vector for delivery of a Flp-recognition target (*FRT*) to the *pca* region of *Rhizobium meliloti* and the assessment of the inducibility of the *pcaD* and *pcaQ* promoters by protocatechuate. McMaster University, 4th Year, B.Sc. thesis.

Cole, S.T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry, F. Tekaiia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, (17 other authors), and B.G. Barrell. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393:537-544.**

Colleaux, L., L. D'Auriol, F. Galibert, and B. Dujon. 1988. Recognition and cleavage site of the intron-encoded omega transposase. *Proc.Natl.Acad.Sci.U.S.A.* **85:6022-6026.**

Cox, M.M. 1983. The FLP protein of the yeast 2 μ plasmid: expression of a eukaryotic genetic recombination system in *Escherichia coli*. *Proc.Natl.Acad.Sci.U.S.A.* **80:4223-4227.**

Cox, M.M. 1988. FLP site-specific recombination system of *Saccharomyces cerevisiae*. pp. 429-443. *In*, R. Kucherlapati and G.R. Smith (ed.), Genetic Recombination. American Society for Microbiology, Washington, D. C.

Craig, N.L. 1988. The mechanism of conservative site-specific recombination. *Annu. Rev.Genet.* **22:77-105.**

David, M., O. Domergue, P. Pognonec, and D. Kahn. 1987. Transcription patterns of *Rhizobium meliloti* symbiotic plasmid pSym: identification of *nifA*-independent fix genes. *J.Bacteriol.* **169:2239-2244.**

De Lay, J., W. Mannheim, P. Segers, A. Lievens, M. Denijn, M. Vanhoucke, and M. Gillis. 1987. Ribosomal ribonucleic acid cistron similarities and taxonomic neighborhood of *Brucella* and CDC group Vd. *Int.J.Syst.Bacteriol.* **37:35-42.**

De Picker, A., M. De Wilde, G.F. De Vos, M. De Vos, M. Van Montagu, and J. Schell. 1980. Molecular cloning of overlapping segments of the nopaline Ti-plasmid pTiC58 as a means to restriction enzyme analysis. *Plasmid* **3:193-211.**

De Vos, G.F., G.C. Walker, and E.R. Signer. 1986. Genetic manipulations in *Rhizobium meliloti* utilizing two new transposon Tn5 derivatives. *Mol.Gen.Genet.* **204**:485-491.

Debellé, F., C. Rosenberg, J. Vasse, F. Maillet, E. Martinez, J. Dénarié, and G. Truchet. 1986. Assignment of symbiotic developmental phenotypes to common and specific nodulation (nod) genetic loci of *Rhizobium meliloti*. *J.Bacteriol.* **168**:1075-1086.

Deckert, G., P.V. Warren, T. Gaasterland, W.G. Young, A.L. Lenox, D.E. Graham, R. Overbeek, M.A. Snead, M. Keller, M. Aujay, R. Huber, R.A. Feldman, J.M. Short, G.J. Olsen, and R.V. Swanson. 1998. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392**:353-358.

Deininger, P.L. 2-15-1983. Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. *Anal.Biochem.* **129**:216-223.

Dietrich, F.S., J. Mulligan, K. Hennessy, M.A. Yelton, E. Allen, R. Araujo, E. Aviles, A. Berno, T. Brennan, J. Carpenter, E. Chen, J.M. Cherry, E. Chung, M. Duncan, E. Guzman, G. Hartzell, S. Hunicke-Smith, R.W. Hyman, A. Kayser, C. Komp, D. Lashkari, H. Lew, D. Lin, D. Mosedale, (14 other authors), and R.W. Davis. 1997. The nucleotide sequence of *Saccharomyces cerevisiae* chromosome V. *Nature* **387**(suppl.):78-81.

Fanning, S., M. Leahy, and D. Sheehan. 1994. Nucleotide and deduced amino acid sequences of *Rhizobium meliloti* 102F34 lacZ gene: comparison with prokaryotic beta-galactosidases and human beta-glucuronidase. *Gene* **141**:91-96.

Figurski, D.H. and D.R. Helinski. 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided *in trans*. *Proc.Natl.Acad.Sci. U.S.A.* **76**:1648-1652.

Finan, T.M., E. Hartweig, K. LeMieux, K. Bergman, G.C. Walker, and E.R. Signer. 1984. General transduction in *Rhizobium meliloti*. *J.Bacteriol.* **159**:120-124.

Finan, T.M., A.M. Hirsch, J.A. Leigh, E. Johansen, G.A. Kuldau, S. Deegan, G.C. Walker, and E.R. Signer. 1985. Symbiotic mutants of *Rhizobium meliloti* that uncouple plant from bacterial differentiation. *Cell* **40**:869-877.

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

Finan, T.M., I. Oresnik, and A. Bottacin. 1988. Mutants of *Rhizobium meliloti* defective in succinate metabolism. *J.Bacteriol.* **170**:3396-3403.

Finan, T.M., J.M. Wood, and D.C. Jordan. 1983. Symbiotic properties of C₄-dicarboxylic acid transport mutants of *Rhizobium leguminosarum*. *J.Bacteriol.* **154**:1403-1413.

Fleischmann, R.D., M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J.F. Tomb, B.A. Dougherty, J.M. Merrick, (29 other authors), and J.C. Venter. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**:496-512.

Fraser, C.M., S. Casjens, W.M. Huang, G.G. Sutton, R. Clayton, R. Lathigra, O. White, K.A. Ketchum, R. Dodson, E.K. Hickey, M. Gwinn, B. Dougherty, J.F. Tomb, R.D. Fleischmann, D. Richardson, J. Peterson, A.R. Kerlavage, J. Quackenbush, S. Salzberg, M. Hanson, R. van Vugt, N. Palmer, M.D. Adams, J. Gocayne, (13 other authors), and J.C. Venter. 1997. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* **390**:580-586.

Fraser, C.M., J.D. Gocayne, O. White, M.D. Adams, R.A. Clayton, R.D. Fleischmann, C.J. Bult, A.R. Kerlavage, G. Sutton, J.M. Kelley, (18 other authors), and J.C. Venter. 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**:397-403.

Fraser, C.M., S.J. Norris, G.M. Weinstock, O. White, G.G. Sutton, R. Dodson, M. Gwinn, E.K. Hickey, R. Clayton, K.A. Ketchum, E. Sodergren, J.M. Hardham, M.P. McLeod, S. Salzberg, J. Peterson, H. Khalak, D. Richardson, J.K. Howell, M. Chidambaram, T. Utterback, L. McDonald, P. Artiach, C. Bowman, M.D. Cotton, (8 other authors), and J.C. Venter. 1998. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* **281**:375-388.

Freiberg, C., R. Fellay, A. Bairoch, W.J. Broughton, A. Rosenthal, and X. Perret. 1997. Molecular basis of symbiosis between *Rhizobium* and legumes. *Nature* **387**:394-401.

Furuichi, T., M. Inouye, and S. Inouye. 1985. Novel one-step cloning vector with a transposable element: application to the *Myxococcus xanthus* genome. *J.Bacteriol.* **164**:270-275.

Futcher, A.B. 1986. Copy number amplification of the 2 micron circle plasmid of *Saccharomyces cerevisiae*. *J.Theor.Biol.* **119**:197-204.

Futcher, A.B. 1988. The 2 micron circle plasmid of *Saccharomyces cerevisiae*. *Yeast* **4**:27-40.

Gantotti, B. and S.V. Beer. 1982. Plasmid-borne determinants of pigmentation and thiamine prototrophy in *Erwinia herbicola*. *J.Bacteriol.* **151**:1627-1629.

Gardner, M.J., H. Tettelin, D.J. Carucci, L.M. Cummings, L. Aravind, E.V. Koonin, S. Shallom, T. Mason, K. Yu, C. Fujii, J. Pederson, K. Shen, J. Jing, C. Aston, Z. Lai, D.C. Schwartz, M. Pertea, S. Salzberg, L. Zhou, G.G. Sutton, R. Clayton, O. White, H.O. Smith, C.M. Fraser, M.D. Adams, J.C. Venter, and S.L. Hoffman. 1998. Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* **282**:1126-1132.

Glazebrook, J. and G.C. Walker. 1989. A novel exopolysaccharide can function in place of the calcofluor-binding exopolysaccharide in nodulation of alfalfa by *Rhizobium meliloti*. *Cell* **56**:661-672.

Glucksmann, M.A., T.L. Reuber, and G.C. Walker. 1993. Family of glycosyl transferases needed for the synthesis of succinoglycan by *Rhizobium meliloti*. *J.Bacteriol.* **175**:7033-7044.

Goffeau, A. et al. (632 other authors) 1997. The yeast genome directory. *Nature* **387**(suppl.):1-105.

Golic, K.G. and S. Lindquist. 1989. The FLP recombinase of yeast catalyzes site-specific recombination in the *Drosophila* genome. *Cell* **59**:499-509.

Gronostajski, R.M. and P.D. Sadowski. 1985. Determination of DNA sequences essential for FLP-mediated recombination by a novel method. *J.Biol.Chem.* **260**:12320-12327.

Hansen, J.B. and R.H. Olsen. 1978. Isolation of large bacterial plasmids and characterization of the P2 incompatibility group plasmids pMG1 and pMG5. *J.Bacteriol.* **135**:227-238.

Hattori, Y., K. Suzuki, M. Uraji, N. Ohta, A. Katoh, and K. Yoshida. 1997. Genome structure of pTi-SAKURA (I): strategy for DNA sequencing of a Japanese cherry-Ti plasmid. *Nucleic.Acids.Symp.Ser.* **37**:159-160.

Herrera-Cervera, J.A., J. Olivares, and J. Sanjuan. 1996. Ammonia inhibition of plasmid pRmeGR4a conjugal transfer between *Rhizobium meliloti* strains. *Appl. Environ.Microbiol.* **62**:1145-1150.

Herrera-Cervera, J.A., J.M. Sanjuan-Pinilla, J. Olivares, and J. Sanjuan. 1998. Cloning and identification of conjugative transfer origins in the *Rhizobium meliloti* genome. *J.Bacteriol.* **180**:4583-4590.

Himmelreich, R., H. Hilbert, H. Plagens, E. Pirkl, B.C. Li, and R. Herrmann. 11-15-1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic.Acids.Res.* **24**:4420-4449.

- Hirsh, P.R., C.L. Wang, and M.J. Woodward.** 1986. Construction of a Tn5 derivative determining resistance to gentamicin and spectinomycin using a fragment cloned from R1033. *Gene* **48**:203-209.
- Hogrefe, C. and B. Friedrich.** 1984. Isolation and characterization of megaplasmid DNA from lithoautotrophic bacteria. *Plasmid* **12**:161-169.
- Holt, J.G., N.R. Krieg, P.H.A. Sneath, J.T. Staley, and S.T. Williams.** 1994. *In*, Bergey's manual of determinative biology. Ninth Edition. Williams and Wilkins.
- Honeycutt, R.J., M. McClelland, and B.W. Sobral.** 1993. Physical map of the genome of *Rhizobium meliloti* 1021. *J.Bacteriol.* **175**:6945-6952.
- Horowitz, B. and R.C. Deonier.** 1985. Formation of Δtra F' plasmids: specific recombination at *oriT*. *J.Mol.Biol.* **186**:267-274.
- Hynes, M.F., R. Simon, P. Muller, K. Niehaus, M. Labes, and A. Puhler.** 1986. The two megaplasmids of *Rhizobium meliloti* are involved in the effective nodulation of alfalfa. *Mol.Gen.Genet.* **202**:356-362.
- Jayaram, M.** 1985. Two-micrometer circle site-specific recombination: the minimal substrate and the possible role of flanking sequences. *Proc.Natl.Acad.Sci.U.S.A.* **82**:5875-5879.
- Jorgensen, R.A., S.J. Rothstein, and W.S. Reznikoff.** 1979. A restriction enzyme cleavage map of Tn5 and location of a region encoding neomycin resistance. *Mol.Gen. Genet.* **177**:65-72.
- Jumas-Bilak, E., C. Maugard, S. Michaux-Charachon, A. Allardet-Servent, A. Perrin, D. O'Callaghan, and M. Ramuz.** 1995. Study of the organization of the genomes of *Escherichia coli*, *Brucella melitensis* and *Agrobacterium tumefaciens* by insertion of a unique restriction site. *Microbiology.* **141**:2425-2432.
- Jumas-Bilak, E., S. Michaux-Charachon, G. Bourg, M. Ramuz, and A. Allardet-Servent.** 1998. Unconventional genomic organization in the alpha subgroup of the Proteobacteria. *J.Bacteriol.* **180**:2749-2755.
- Kaneko, T., S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirosawa, M. Sugiura, S. Sasamoto, T. Kimura, T. Hosouchi, A. Matsuno, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda, and S. Tabata.** 1996. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**:109-136.

Kawarabayasi, Y., M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, Y. Nakamura, T.F. Robb, K.Horikoshi, Y. Masuchi, H. Shizuya, and H. Kikuchi. 1998. Complete sequence and gene organization of the genome of a hyper- thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3. *DNA Res.* **5**:55-76.

Kim, U.J., B.W. Birren, T. Slepak, V. Mancino, C. Boysen, H.L. Kang, M.I. Simon, and H. Shizuya. 1996. Construction and characterization of a human bacterial artificial chromosome library. *Genomics* **34**:213-218.

Kim, U.J., H. Shizuya, P.J. de Jong, B. Birren, and M.I. Simon. 1992. Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids.Res.* **20**:1083-1085.

Kim, U.J., H. Shizuya, H.L. Kang, S.S. Choi, C.L. Garrett, L.J. Smink, B.W. Birren, J.R. Korenberg, I. Dunham, and M.I. Simon. 1996. A bacterial artificial chromosome-based framework contig map of human chromosome 22q. *Proc.Natl.Acad. Sci.U.S.A.* **93**:6297-6301.

Klenk, H.P., R.A. Clayton, J.F. Tomb, O. White, K.E. Nelson, K.A. Ketchum, R.J. Dodson, M. Gwinn, E.K. Hickey, J.D. Peterson, D.L. Richardson, A.R. Kerlavage, D.E. Graham, N.C. Kyrpides, R.D. Fleischmann, J. Quackenbush, N.H. Lee, G.G. Sutton, S. Gill, E.F. Kirkness, B.A. Dougherty, K. McKenney, M.D. Adams, B. Loftus, (26 other authors), and J.C. Venter. 1997. The complete genome sequence of the hyperthermophilic, sulphate- reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390**:364-370.

Konieczny, I., K.S. Doran, D.R. Helinski, and A. Blasina. 1997. Role of TrfA and DnaA proteins in origin opening during initiation of DNA replication of the broad host range plasmid RK2. *J.Biol.Chem.* **272**:20173-20178.

Kovach, M.E., P.H. Elzer, D.S. Hill, G.T. Robertson, M.A. Farris, R.M. Roop, and K.M. Peterson. 1995. Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene* **166**:175-176.

Kovach, M.E., R.W. Phillips, P.H. Elzer, R.M. Roop, and K.M. Peterson. 1994. pBBR1MCS: a broad-host-range cloning vector. *Biotechniques* **16**:800-802.

- Kunst, F., N. Ogasawara, I. Moszer, A.M. Albertini, G. Alloni, V. Azevedo, M.G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S.C. Brignell, S. Bron, S. Brouillet, C.V. Bruschi, B. Caldwell, V. Capuano, N.M. Carter, S.K. Choi, J.J. Codani, I.F. Connerton, (126 other authors), and A. Danchin.** 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390**:249-256.
- Lanka, E. and B.M. Wilkins.** 1995. DNA processing reactions in bacterial conjugation. *Annu.Rev.Biochem.* **64**:141-169.
- Leigh, J.A., E.R. Signer, and G.C. Walker.** 1985. Exopolysaccharide-deficient mutants of *Rhizobium meliloti* that form ineffective nodules. *Proc.Natl.Acad.Sci.U.S.A.* **82**:6231-6235.
- Leonardo, E.D. and J.M. Sedivy.** 1990. A new vector for cloning large eukaryotic DNA segments in *Escherichia coli*. *Biotechnology (N.Y.)* **8**:841-844.
- Lloyd, A.M. and R.W. Davis.** 1994. Functional expression of the yeast FLP/*FRT* site-specific recombination system in *Nicotiana tabacum*. *Mol.Gen.Genet.* **242**:653-657.
- Long, S.R., W.J. Buikema, and F.M. Ausubel.** 1982. Cloning of *Rhizobium meliloti* genes by direct complementation of Nod⁻ mutants. *Nature* **298**:485-488.
- Long, S.R. and B.J. Staskawicz.** 1993. Prokaryotic plant parasites. *Cell* **73**:921-935.
- Lovett, M.A. and D.R. Helinski.** 1976. Method for the isolation of the replication region of a bacterial replicon: construction of a mini-*F'**km* plasmid. *J.Bacteriol.* **127**:982-987.
- MacPherson, G.** 1995. The β -ketoacid pathway in *R. meliloti*. McMaster University, M. Sc. thesis.
- Manis, J.J. and B.C. Kline.** 1977. Restriction endonuclease mapping and mutagenesis of the F sex factor replication region. *Mol.Gen.Genet.* **152**:175-182.
- Margolin, W. and S.R. Long.** 1993. Isolation and characterization of a DNA replication origin from the 1,700-kilobase-pair symbiotic megaplasmid pSym-b of *Rhizobium meliloti*. *J.Bacteriol.* **175**:6553-6561.
- Meade, H.M., S.R. Long, G.B. Ruvkun, S.E. Brown, and F.M. Ausubel.** 1982. Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn5 mutagenesis. *J.Bacteriol.* **149**:114-122.

- Mertens, N., E. Remaut, and W. Fiers.** 1995. Tight transcriptional control mechanism ensures stable high-level expression from T7 promoter-based expression plasmids. *Biotechnology (N.Y.)* **13**:175-179.
- Miller, J.H.** 1972. Experiments in molecular genetics. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, N.Y.
- Misumi, D.J., D.L. Nagle, S.H. McGrail, B.J.J. Dussault, J.S. Smutko, H. Chen, O. Charlat, G.M. Duyk, C. Ebeling, L. Baldini, G.A. Carlson, and K.J. Moore.** 1997. The physical and genetic map surrounding the Lyst gene on mouse chromosome 13. *Genomics* **40**:147-150.
- Monteilhet, C., A. Perrin, A. Thierry, L. Colleaux, and B. Dujon.** 1990. Purification and characterization of the *in vitro* activity of I-Sce I, a novel and highly specific endonuclease encoded by a group I intron. *Nucleic Acids Res.* **18**:1407-1413.
- Moreno, E., E. Stackebrandt, M. Dorsch, J. Wolters, M. Busch, and H. Mayer.** 1990. *Brucella abortus* 16S rRNA and lipid A reveal a phylogenetic relationship with members of the alpha-2 subdivision of the class Proteobacteria. *J. Bacteriol.* **172**:3569-3576.
- Myers, E.W. and W. Miller.** 1988. Optimal alignments in linear space. *Comput. Appl. Biosci.* **4**:11-17.
- Nishiguchi, R., M. Takanami, and A. Oka.** 1987. Characterization and sequence determination of the replicator region in the hairy-root-inducing plasmid pRiA4b. *Mol. Gen. Genet.* **206**:1-8.
- O'Gorman, S., D.T. Fox, and G.M. Wahl.** 1991. Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science* **251**:1351-1355.
- Oresnik, I.J., T.C. Charles, and T.M. Finan.** 1994. Second site mutations specifically suppress the Fix- phenotype of *Rhizobium meliloti ndvF* mutations on alfalfa: identification of a conditional *ndvF*-dependent mucoid colony phenotype. *Genetics* **136**:1233-1243.
- Pan, H., D. Clary, and P.D. Sadowski.** 1991. Identification of the DNA-binding domain of the FLP recombinase. *J. Biol. Chem.* **266**:11347-11354.
- Pansegrau, W., W. Schroder, and E. Lanka.** 1993. Relaxase (TraI) of IncP alpha plasmid RP4 catalyzes a site-specific cleaving-joining reaction of single-stranded DNA. *Proc. Natl. Acad. Sci. U.S.A.* **90**:2925-2929.

Pansegrau, W., G. Ziegelin, and E. Lanka. 1990. Covalent association of the *traI* gene product of plasmid RP4 with the 5'-terminal nucleotide at the relaxation nick site. *J.Biol.Chem.* **265**:10637-10644.

Perrin, A., M. Buckle, and B. Dujon. 1993. Asymmetrical recognition and activity of the I-*SceI* endonuclease on its site and on intron-exon junctions. *EMBO J.* **12**:2939-2947.

Pósfai, G., M. Koob, Z. Hradecná, N. Hasan, M. Filutowicz, and W. Szybalski. 1994. *In vivo* excision and amplification of large segments of the *Escherichia coli* genome. *Nucleic.Acids.Res.* **22**:2392-2398.

Pósfai, G., M.D. Koob, H.A. Kirkpatrick, and F.R. Blattner. 1997. Versatile insertion plasmids for targeted genome manipulations in bacteria: isolation, deletion, and rescue of the pathogenicity island LEE of the *Escherichia coli* O157:H7 genome. *J.Bacteriol.* **179**:4426-4428.

Prentki, P. and H.M. Krisch. 1984. *In vitro* insertional mutagenesis with a selectable DNA fragment. *Gene* **29**:303-313.

Proteau, G., D. Sidenberg, and P. Sadowski. 1986. The minimal duplex DNA sequence required for site-specific recombination promoted by the FLP protein of yeast *in vitro*. *Nucleic.Acids.Res.* **14**:4787-4802.

Ramirez-Romero, M.A., P. Bustos, L. Girard, O. Rodriguez, M.A. Cevallos, and G. Davila. 1997. Sequence, localization and characteristics of the replicator region of the symbiotic plasmid of *Rhizobium etli*. *Microbiology.* **143**:2825-2831.

Reynolds, A.E., A.W. Murray, and J.W. Szostak. 1987. Roles of the 2 μ gene products in stable maintenance of the 2 μ plasmid of *Saccharomyces cerevisiae*. *Mol.Cell Biol.* **7**:3566-3573.

Reznikoff, W.S. 1982. Tn5 transposition and its regulation. *Cell* **31**:307-308.

Rode, C.K., V.H. Obreque, and C.A. Bloch. 1995. New tools for integrated genetic and physical analyses of the *Escherichia coli* chromosome. *Gene* **166**:1-9.

Ronson, C.W., P. Lyttleton, and J.G. Robertson. 1981. C₄-dicarboxylate transport mutants of *Rhizobium trifolii* form ineffective nodules on *Trifolium repens*. *Proc.Natl. Acad.Sci.U.S.A.* **78**:4284-4288.

Rosenberg, C., P. Boistard, J. Dénarié, and F. Casse-Delbart. 1981. Genes controlling early and late functions in symbiosis are located on a megaplasmid in *Rhizobium meliloti*. *Mol.Gen.Genet.* **184**:326-333.

- Rosenberg, C., F. Casse-Delbart, I. Dusha, M. David, and C. Boucher.** 1982. Megaplasmids in the plant-associated bacteria *Rhizobium meliloti* and *Pseudomonas solanacearum*. *J.Bacteriol.* **150**: 402-406.
- Rosenberg, C. and T. Huguet.** 1984. The pAtC58 plasmid is not essential for tumour induction. *Mol.Gen.Genet.* **196**:533-536.
- Sadowski, P.D.** 1995. The Flp recombinase of the 2- μ plasmid of *Saccharomyces cerevisiae*. *Prog.Nucleic.Acid.Res.Mol.Biol.* **51**:53-91.
- Sambrook, J., E.F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. 2nd. Cold Spring Harbour Laboratory Press, Cold Spring Harbour,N.Y.
- Sangari, F.J. and J. Aguero.** 1996. Molecular basis of *Brucella* pathogenicity: an update. *Microbiologia.* **12**: 207-218.
- Sasakawa, C. and M. Yoshikawa.** 1987. A series of Tn5 variants with various drug-resistance markers and suicide vector for transposon mutagenesis. *Gene* **56**:283-288.
- Schmidhauser, T.J. and D.R. Helinski.** 1985. Regions of broad-host-range plasmid RK2 involved in replication and stable maintenance in nine species of gram-negative bacteria. *J.Bacteriol.* **164**:446-455.
- Senecoff, J.F., R.C. Bruckner, and M.M. Cox.** 1985. The FLP recombinase of the yeast 2-micron plasmid: characterization of its recombination site. *Proc.Natl.Acad. Sci.U.S.A.* **82**:7270-7274.
- Senecoff, J.F. and M.M. Cox.** 1986. Directionality in FLP protein-promoted site-specific recombination is mediated by DNA-DNA pairing. *J.Biol.Chem.* **261**:7380-7386.
- Sherman, J.A. and S.W. Matson.** 1994. *Escherichia coli* DNA helicase I catalyzes a sequence-specific cleavage/ligation reaction at the F plasmid origin of transfer. *J.Biol.Chem.* **269**:26220-26226.
- Shizuya, H., B. Birren, U.J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and M. Simon.** 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc.Natl.Acad.Sci.U.S.A.* **89**:8794-8797.
- Simon, R.** 1984. High frequency mobilization of gram-negative bacterial replicas by the *in vitro* constructed Tn5-Mob transposon. *Mol.Gen.Genet.* **196**:413-420.

- Smith, D.R., L.A. Doucette-Stamm, C. Deloughery, H. Lee, J. Dubois, T. Aldredge, R. Bashirzadeh, D. Blakely, R. Cook, K. Gilbert, D. Harrison, L. Hoang, P. Keagle, W. Lumm, B. Pothier, D. Qiu, R. Spadafora, R. Vicaire, Y. Wang, J. Wierzbowski, R. Gibson, N. Jiwani, A. Caruso, D. Bush, (12 other authors), and J.N. Reeve.** 1997. Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J.Bacteriol.* **179**:7135-7155.
- Smith, L.D. and T.A. Ficht.** 1990. Pathogenesis of *Brucella*. *Crit.Rev.Microbiol.* **17**:209-229.
- Snaith, M.R., N.J. Kilby, and J.A. Murray.** 1996. An *Escherichia coli* system for assay of F1p site-specific recombination on substrate plasmids. *Gene* **180**:225-227.
- Sobral, B.W., R.J. Honeycutt, and A.G. Atherly.** 1991. The genomes of the family Rhizobiaceae: size, stability, and rarely cutting restriction endonucleases. *J.Bacteriol.* **173**:704-709.
- Sobral, B.W., R.J. Honeycutt, A.G. Atherly, and M. McClelland.** 1991. Electrophoretic separation of the three *Rhizobium meliloti* replicons. *J.Bacteriol.* **173**:5173-5180.
- Sola-Landa, A., J. Pizarro-Cerda, M.J. Grillo, E. Moreno, I. Moriyon, J.M. Blasco, J.P. Gorvel, and I. Lopez-Goni.** 1998. A two-component regulatory system playing a critical role in plant pathogens and endosymbionts is present in *Brucella abortus* and controls cell invasion and virulence. *Mol.Microbiol.* **29**:125-138.
- Soto, M.J., A. Zorzano, F.M. Garcia-Rodriguez, J. Mercado-Blanco, I.M. Lopez-Lara, J. Olivares, and N. Toro.** 1994. Identification of a novel *Rhizobium meliloti* nodulation efficiency *nfe* gene homolog of *Agrobacterium* ornithine cyclodeaminase. *Mol.Plant Microbe.Interact.* **7**:703-707.
- Soto, M.J., A. Zorzano, J. Mercado-Blanco, V. Lepek, J. Olivares, and N. Toro.** 1993. Nucleotide sequence and characterization of *Rhizobium meliloti* nodulation competitiveness genes *nfe*. *J.Mol.Biol.* **229**:570-576.
- Southern, E.M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J.Mol.Biol.* **24**:865-868.
- Spaink, H.P., R.J. Okker, C.A. Wijffelman, E. Pees, and B.J. Lugtenberg.** 1987. Promoters in the nodulation region of the *Rhizobium leguminosarum* Sym plasmid pRL1JI. *Plant Mol.Biol.* **9**:27-39.

Stephens, R.S., S. Kalman, C. Lammel, J. Fan, R. Marathe, L. Aravind, W. Mitchell, L. Olinger, R.L. Tatusov, Q. Zhao, E.V. Koonin, and R.W. Davis. 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* **282**:754-759.

Suzuki, K., Y. Hattori, M. Uraji, N. Ohta, A. Katoh, and K. Yoshida. 1997. Genome structure of pTi-SAKURA (II): genetic map constructed by complete DNA sequencing. *Nucleic.Acids.Symp.Ser.* 161-162.

Suzuki, K., N. Ohta, Y. Hattori, M. Uraji, A. Kato, and K. Yoshida. 1998. Novel structural difference between nopaline- and octopine-type *trbJ* genes: construction of genetic and physical map and sequencing of *trb/tral* and *rep* gene clusters of a new Ti plasmid pTi-SAKURA. *Biochim.Biophys.Acta* **1396**:1-7.

Tabata, S., P.J. Hooykaas, and A. Oka. 1989. Sequence determination and characterization of the replicator region in the tumor-inducing plasmid pTiB6S3. *J.Bacteriol.* **171**:1665-1672.

Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic.Acids.Res.* **22**:4673-4680.

Timmis, K., F. Cabello, and S. Cohen. 1975. Cloning, isolation, and characterization of replication region of complex plasmid genomes. *Proc.Natl.Acad.Sci.U.S.A.* **72**:2242-2246.

Tomb, J.F., O. White, A.R. Kerlavage, R.A. Clayton, G.G. Sutton, R.D. Fleischmann, K.A. Ketchum, H.P. Klenk, S. Gill, B.A. Dougherty, K. Nelson, J. Quackenbush, L. Zhou, E.F. Kirkness, S. Peterson, B. Loftus, D. Richardson, R. Dodson, H.G. Khalak, A. Glodek, K. McKenney, L.M. Fitzgerald, N. Lee, M.D. Adams, (17 other authors), and J.C. Venter. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**:539-547.

Turner, S.L. and J.P. Young. 1995. The replicator region of the *Rhizobium leguminosarum* cryptic plasmid pRL8JI. *FEMS Microbiol.Lett.* **133**:53-58.

Unger, L., S.F. Ziegler, G.A. Huffman, V.C. Knauf, R. Peet, L.W. Moore, M.P. Gordon, and E.W. Nester. 1985. New class of limited-host-range *Agrobacterium* mega-tumor-inducing plasmids lacking homology to the transferred DNA of a wide-host-range, tumor-inducing plasmid. *J.Bacteriol.* **164**:723-730.

Van Larebeke, N., G. Engler, M. Holsters, D. van den Elsack, I. Zaenen, R.A. Schilperoort, and J. Schell. 1974. Large plasmid in *Agrobacterium tumefaciens* essential for crown-gall inducing activity. *Nature* **255**:742-743.

- Vieira, J. and J. Messing.** 1987. Production of single-stranded plasmid DNA. *Methods Enzymol.* **153:3-11:3-11.**
- Vincent, J.M.** 1941. Serological properties of the root-nodule bacteria. I. Strains of *Rhizobium meliloti*. *Proc.Linn.Soc.N.S.W.* **66:145- 154.**
- Voegelé, R.T., S. Bardin, and T.M. Finan.** 1997. Characterization of the *Rhizobium (Sinorhizobium) meliloti* high- and low-affinity phosphate uptake systems. *J.Bacteriol.* **179:7226-7232.**
- Volkert, F.C. and J.R. Broach.** 1986. Site-specific recombination promotes plasmid amplification in yeast. *Cell* **46:541-550.**
- Voss, H., C. Schwager, S. Wiemann, J. Zimmermann, J. Stegemann, H. Erfle, A.M. Voie, H. Drzonek, and W. Ansorge.** 1995. Efficient low redundancy large-scale DNA sequencing at EMBL. *J.Biotechnol.* **41:121-129.**
- Voss, H., S. Wiemann, D. Grothues, C. Sensen, J. Zimmermann, C. Schwager, J. Stegemann, H. Erfle, T. Rupp, and W. Ansorge.** 1993. Automated low-redundancy large-scale DNA sequencing by primer walking. *Biotechniques* **15:714-721.**
- Waite, L.L. and M.M. Cox.** 1995. A protein dissociation step limits turnover in FLP recombinase-mediated site-specific recombination. *J.Biol.Chem.* **270:23409- 23414.**
- Warren, G.J. and A.J. Clark.** 1980. Sequence-specific recombination of plasmid ColE1. *Proc.Natl.Acad.Sci.U.S.A.* **11:6724-6728.**
- Watson, R.J., Y.K. Chan, R. Wheatcroft, A.F. Yang, and S.H. Han.** 1988. *Rhizobium meliloti* genes required for C₄-dicarboxylate transport and symbiotic nitrogen fixation are located on a megaplasmid. *J.Bacteriol.* **170:927-934.**
- White, F.F. and E.W. Nester.** 1980. Relationship of plasmids responsible for hairy root and crown gall tumorigenicity. *J.Bacteriol.* **144:710-720.**
- Wild, J., Z. Hradecná, G. Pósfai, and W. Szybalski.** 1996. A broad-host-range *in vivo* pop-out and amplification system for generating large quantities of 50- to 100-kb genomic fragments for direct DNA sequencing. *Gene* **179:181-188.**
- Willetts, N. and B.M. Wilkins.** 1984. Processing of plasmid DNA during bacterial conjugation. *Microbiol.Rev.* **48:24-41.**
- Williams, M.N., R.I. Hollingsworth, S. Klein, and E.R. Signer.** 1990. The symbiotic defect of *Rhizobium meliloti* exopolysaccharide mutants is suppressed by *lpsZ*⁺, a gene involved in lipopolysaccharide biosynthesis. *J.Bacteriol.* **172:2622-2632.**

Woo, S.S., J. Jiang, B.S. Gill, A.H. Paterson, and R.A. Wing. 1994. Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*. *Nucleic Acids Res.* **22**:4922-4931.

Wu, T.T. 1966. A model for three-point analysis of random general transduction. *Genetics* **54**:405-410.

Yakobson, E.A. and D.G. Guiney. 1984. Conjugal transfer of bacterial chromosomes mediated by the RK2 plasmid transfer origin cloned into transposon Tn5. *J.Bacteriol.* **160**:451-453.

Yarosh, O.K., T.C. Charles, and T.M. Finan. 1989. Analysis of C₄-dicarboxylate transport genes in *Rhizobium meliloti*. *Mol.Microbiol.* **3**:813-823.

Zhan, H.J., S.B. Levery, C.C. Lee, and J.A. Leigh. 1989. A second exopolysaccharide of *Rhizobium meliloti* strain SU47 that can function in root nodule invasion. *Proc.Natl. Acad.Sci.U.S.A.* **86**:3055-3059.

Zimmer, R. and G.A. Verrinder. 1997. Construction and characterization of a large-fragment chicken bacterial artificial chromosome library. *Genomics* **42**:217-226.