CHAPTER 1. LITERATURE REVIEW

Rhizobia

Rhizobia are α- and β- proteobacteria that have the capability of fixing atmospheric nitrogen (N_2) in symbiosis with legumes. The symbiotic relationship is based on exchange of organic acids and amino acids between plant root cells and differentiated N_2-fixing form of bacteria, termed bacteroids. Symbiotic nitrogen fixation is for the most part restricted to a limited number of bacterial groups including the genera *Rhizobium, Mesorhizobium, Sinorhizobium, Bradyrhizobium* and *Azorhizobium*. On roots or stems of host plants, rhizobia form specialized organs called nodules, within which these endosymbionts reduce nitrogen gas to ammonia that is subsequently used by the host plant.

Sinorhizobium meliloti, the organism studied in this thesis, is a gram negative soil bacterium, and is a member of the family Rhizobiaceae in the alpha subdivision of Proteobacteria. *S. meliloti* forms symbiotic associations with plant genera such as *Medicago* including alfalfa (*Medicago sativa*), *Melilotus* (sweet clover *Melilotus alba*) and *Trigonella*. On the roots of these plants *S. meliloti* forms nodules within which the differentiated bacteroids reduce nitrogen gas to ammonia.

The *Rhizobium* Plant Associations

The earliest interaction between legumes and rhizobia is based on release of flavonoids from seed coat or roots. Flavonoids are plant secondary metabolites that are synthesized via the central phenylpropanoid pathway and the acetate – malonate pathway (Forkmann and Heller 1999; Aoki *et al*. 2000). Flavonoids induce the structural *nod* genes *nodA, B, C* as well as *nodI* and *J*, in rhizobia. The first three genes are responsible for Nod factor synthesis, while *nodI* and *nodJ* play a role in the efficiency of secretion of lipochitin oligosaccharides. Approximately 30 different flavonoids have been isolated from nine legume genera growing under axenic conditions (growth of organisms of a single species in the absence of cells or living organisms of any other) (Cooper 2004). In alfalfa (*Medicago sativa*) the most active *nod* inducers are luteolin and 4, 7-dihydroxyflavone (DHF).

Structural *nod* genes are regulated by the LysR-type transcriptional regulator NodD. NodD binds to a conserved DNA sequence, *nod* box, which is part of the promoter region of NodD-induced genes (Fisher and Long 1993). Flavonoides and NodD protein are both required for induction of structural *nod* genes suggesting that a NodD – flavonoid complex interacts with the *nod* box**.** The products of the induced *nodABC* catalyze the synthesis of Nod factors known as lipochito – oligosaccharides, comprising β – 1,4 linked N – acetyl – D – glucosamine residues with a fatty acid chain attached at the nonreducing terminus (Cooper *et al*., 2007). The proteins made by *nodI* and *nodJ* are involved in the secretion of these Nod factors. These Nod factors are recognized by the

plant, inducing cells within the root cortex to become mitotically active, developing into the nodule primordium (Bartsev et al., 2004).

Integration of infection by rhizobium at the root epidermis and the initiation of cell division in the cortex is important for the development of nitrogen-fixing nodules (Gage, 2004). Infection threads are unique plant-made invasive invaginations that are capable of crossing cell boundaries in such way that bacteria can easily enter cortical cells. The nodule primordium is formed as a consequence of mitotic division of cortical cells below the sites of infection. The infection threads grow toward the nodule primordia and rhizobia are released into the inner cells in the nodule. The bacteria, encapsulated within a plant membrane, undergo differentiation into bacteroids which are responsible for reduction of nitrogen gas to ammonia (Oldroyd and Downie, 2008).

Nitrogen fixation in *Rhizobium*

Nitrogen fixation in *Rhizobium* is carried out by the products of the *nif* and *fix* genes. *nif* genes have homologs in free – living nitrogen fixing organism (eg, *Klebsiella pneumoniae*). These genes have been identified through two main approaches: finding homology with the identified *nif* genes of *K. pneumoniae* or by sequence analysis of DNA regions believed to carry *nif* genes because of linkage to known *nif* genes. In *S. meliloti nifH, D, K, N, E, B* were identified using these approaches. In *S. meliloti* bacteroids, nitrogen fixation is carried out by the nitrogenase enzyme complex consisting of two protein complexes. *nifD* and *nifK* encode the α and β subunits of the first, an α2β2 FeMo

protein called dinitrogenase. *nifH* encodes the second, an Fe protein called dinitrogenase reductase. *nifNE* encode the molecular scaffold for assembly of the iron molybdenum cofactor while the gene product for *nifB* is a Fe-S containing precursor of iron molybdenum cofactor. Regulation of *nif* genes in *S. meliloti* depends on NifA protein. Genes that are required for symbiotic nitrogen fixation but whose function is not known to be analogous to a free – living function are called *fix* genes (Long, 1989). It has been postulated that the *fixABCX* operon (essential for nitrogen fixation) produces the proteins responsible for reducing dinitrogenase reductase.

The *Rhizobium* Genome

The complete genome sequences of several rhizobia and related bacteria have been determined, *Mesorhizobium loti* (Kaneko *et al.*, 2000), *Sinorhizobium meliloti* (Galibert *et al*., 2001), *Bradyrhizobium japonicum* (Kaneko *et al.*, 2002), *Rhizobium leguminosarum* bv *viciae* (Young *et al*., 2006), *Rhizobium etli* (Gonzalez *et al*., 2006), *Sinorhizobium fredii* NGR234 (Schmeisser *et al*., 2009) and *Sinorhizobium medicae* WSM419 (Reeve *et al*.,2010). Many rhizobia have a multireplicon genome which might be a result of the complex lifestyle adopted by these plant – associated bacteria. Genome size is influenced by environmental factors and Rhizobia tend to have larger genomes (Bentley and Parkhill, 2004). The largest genome of rhizobia that has been sequenced is 9.2 Mb of *B. japonicum* while the smallest is 6.5 Mb of *R. etli.* Extrachromosomal replicons (plasmids) contribute to the total size of the genome and they are present in all

rhizobia except *B. japonicum*. Most of symbiotic relevant genes (*nod, nif* and *fix)* are located on those extrachromosomal replicons (*S. meliloti*, *R. etli*, *R. leguminosarum* and *S. fredii* NGR234) or within symbiosis islands like in *M. loti* (Kaneko *et al*., 2000) and *B. japonicum* (Kaneko *et al.*, 2002) genome. The presence of many insertion sequence (IS) elements, transposase genes, and related elements within symbiotic regions define rhizobium genomes as a highly dynamic entities. IS elements and transposase genes are abundant in symbiotic islands of *M. loti* (Kaneko *et al.*, 2000) and *B. japonicum* (Kaneko *et al.*, 2002) genome as well as the pSymA megaplasmid of *S. meliloti* (Galibert *et al*., 2001)

The *S. meliloti* genome consists of a 3654 Kb circular chromosome and two megaplasmids called pSymA (1354 Kb) and pSymB (1683 Kb) (Galibert *et al* 2001; Finan *et al*. 2001; Barnett *et al*. 2001). The *S. meliloti* genome has 6204 predicted protein-coding regions and approximately 3000 of these encode Proteins of Unknown Function (PUFs), many of which are on pSymA and pSymB megaplasmids (Galibert *et al*., 2001). pSymA carries most of the genes necessary for nodulation (*nod*), nitrogen fixation (*nif* and *fix*) and nitrogen metabolism as well as genes involved in carbon metabolism, transport, and stress and resistance responses (Banfalvi *et al.* 1981; David *et al.* 1987; Long *et al*. 1982). pSymB carries large numbers of solute transport systems (transport systems represent 18.9% of the pSymB); as well as gene clusters involved in exopolysaccharide synthesis (e*xo,exp*) (Finan *et al.* 1985; Finan *et al.* 1986; Glucksmann *et al.* 1993; Hynes *et al.* 1986), phosphate transport (*pho*) (Bardin *et al.* 1998; Voegele *et al.*1997), (*pst*) (Yuan *et al.*, 2006), thiamine biosynthesis (*thi*) (Finan *et al.* 1986),

lipopolysaccharide synthesis (*lps*), as well as genes involved in ectoine/hydroxyectoine uptake (*ehu*), ectoine utilization (*eut*) (Jebbar *et al*., 2005), protochatechuate (*pca*) (MacLean *et al*., 2006), arabinose (*ara*) (Poysti *et al*., 2007), hydroxyproline (*hyp*) (MacLean *et al*., 2009) dulcitol (*dul*) (White and Finan, unpublished data), lactose (*lac*) (Jelesko and Leigh, 1994), raffinose (*raf*) and melibiose (*mel*) utilization (Gage and Long, 1998).

Megaplasmids

Rosenberg *et al*., 1981 introduced the term megaplasmid into the literature to describe plasmids with a molecular weight greater than 300 x 10^6 (equivalent to about 450 Kb). The first megaplasmid was discovered by examining the plasmid content of *Sinorhizobium meliloti* cells using agarose gel electrophoresis (Rosenberg *et al*., 1981). In the past, the term megaplasmid was used only to describe plasmids that are large in size (> 450 Kb) and not essential for their host. Although there is still no general consensus as to the minimum size at which a plasmid becomes a megaplasmid, some authors suggest that it should be 100 Kb (Schwartz *et al*., 2008). Currently, analysis of megaplasmids' sequences has revealed the presence of genes essential for growth of the organism under most conditions. However, the distinction between chromosomes and megaplasmids harboring essential genes is vague as a single recombinative event can result in a transfer of an essential gene from the chromosome to the megaplasmid. In support of this

statement, many secondary chromosomes carry a plasmid origin of replication (MacLellan *et al*., 2006).

Below I will describe megaplasmids present in different organisms and attempt to identify phenotypes associated with these plasmids (see Table 1.1 for summary).

Many haloarchaea commonly harbor one or more megaplasmids. *Halobacterium* sp NRC-1 has two megaplasmids pNRC100 (191 Kb) and pNRC200 (365 Kb) that are frequently called minichromosomes because of essential genes located on them (Ng *et al*., 1991; Ng and DasSarma 1991). pNRC100 harbors genes that are involved in synthesis of unusual gas-filled vesicles used for cell flotation (DasSarma *et al*., 2004) as well as transcription and replication factors; genes involved in plasmid and chromosomal partitioning, genes that code for proteins similar to subunits I and II of *E. coli* cytochrome *d* oxidase (Miller and Gennis 1983) and genes that code for thioredoxin and thioredoxin reductase, which function in protein reduction (Arner and Holmgren 2000). Megaplasmid pNRC200 has unique regions with genes responsible for synthesis of additional transcriptional factors, a DNA polymerase and the only copy of an arginyl-tRNA synthetase gene (DasSarma *et al*., 2004) (Table 1.1).

In a review by Vedler (2007) thirty seven catabolic megaplasmids involved in the degradation of aromatic compounds by soil bacteria are described. The large size (>100 Kb) of these catabolic plasmids can be explained by the fact that they often contain the full set of plasmid transfer genes as well as the collection of catabolic genes/operons required for (complete) degradation of aromatic compound. Genes for biodegradation of naturally occurring aromatic compounds (phenol, toluene, xylene, naphthalene) seem to

be located on catabolic plasmids as well as genes for degrading xenobiotic compounds (2, 4-D, atrazine) (Top *et al*. 2000). The majority of these catabolic plasmids have been found in *Pseudomonas*. The other most abundant genera are *Rhodococcus*, *Sphingomonas*, *Arthrobacter*, and *Burkholderia*. Many of these bacteria also have a large catabolic plasmid that contains the genes for all or most of the mineralization pathways of polycyclic aromatic compounds and heteroaromatic compounds. It was found that *Pseudomonas putida* G7 harbors a naphthalene catabolic plasmid NAH7 (82 Kb) (Sota *et al*., 2006), *Pseudomonas resinovorans* CA10 has a carbazole catabolic plasmid pCAR1 (199 Kb) (Nojiri *et al*., 2001; Maeda *et al*., 2003) and within *Novosphingobium* sp. KA1 a carbazole catabolic plasmid pCAR3 (254 Kb) was found (Shintani *et al*., 2007) (Table 1.1).

Agrobacterium species are the main cause of crown gall disease and hairy root disease on dicot plants. Most genes essential for pathogenicity are found on large plasmids, approximately 200 Kb, called tumor-inducing (Ti) or root-inducing (Ri) plasmids. Five gene clusters are common on Ti and Ri plasmids: (1) T-DNA, which is transferred to the host plants; (2) the virulence gene (*vir*) region, which directs the recognition of plant phenolic compounds and the processing and transfer of T-DNA; (3) the replication gene (*rep*) region, which is required for the plasmid replication; (4) *tra* and (5) *trb* regions, which direct conjugal transfer of the plasmid (Goodner *et al*. 2001 , Wood *et al*. 2001). Two or more regions that direct uptake and catabolism of opine are also present in each plasmid (Table 1.1).

Shigella species and enteroinvasive *Escherichia coli* have a large (> 200 Kb) virulence plasmid that carries genes necessary for invasion and colonization of the epithelial cell layer of the human gut resulting in dysentery. Analysis of the DNA sequences of the virulence plasmids from different *Shigella* strains, pINV-2457T (serotype 2a strain 2457T), pWR100 and pWR501 (serotype 5 strain M90T), and pCP301 (serotype 2a strain 301), showed that they were largely identical, except for some minor differences in insertion sequence (IS) elements (Wei *et al*., 2003). The genes required for expression of type III secretion systems and many of the effector proteins involved in the early stage of invasion, were located on these virulence plasmids (Table 1.1). The *S. flexneri* virulence plasmid has at least two toxin–antitoxin (TA) systems that are related to well-characterized examples of TA systems. The first of these shows strong DNA sequence similarity to the *ccdA* and *ccdB* system from the F plasmid (Afif *et al*. 2001; Jaffe *et al*. 1985). The second toxin–antitoxin system consists of the *mvpT* and *mvpA* genes (Radnedge *et al*. 1997). The same virulence plasmid has two well defined systems involved in stable inheritance characterized by segregation of plasmid free cells. One of these, *parAB* is related to the partition system of phage/plasmid P1 and consists of the genes *parA*, *parB* and *parS* (Sergueev *et al*. 2005). The second system is important for stable inheritance of a plasmid and it is encoded by the *stbA* and *stbB* genes whose products have significant amino-acid sequence identity to the StbA and StbB proteins of plasmid R100. Virulence plasmids have been also found in the genus *Yersinia.* Three human pathogenic species, *Yersiniae pestis* and the enteropathogenic *Yersiniae pseudotuberculosis* and *Yersiniae enterocolitica*, harbor a 70 Kb virulence plasmid (pYV)

(Huang *et al*., 2006) that is essential for infection of lymph tissues as well as overcoming other host defense mechanisms. Plasmid pYV encodes a type III secretion system and *Yersinia* outer membrane proteins (Table 1.1). Beside pYV, the genome of *Y. pestis* has two additional plasmids pMT1 (100 Kb) (Hu *et al.*, 1998; Lindler *et al*., 1998) and pPCP1 (9 Kb). Plasmid pMT1 carries five important genes that have been associated with virulence, while pPCP1 has been associated with four biochemical activities: biosynthesis of a bacteriocin called pesticin; a protein that provides immunity to pesticin; a plasminogen activator and a coagulase (Table 1.1).

Megaplasmids carrying genes for the response and resistance to heavy metals are commonly found in bacteria belonging to the genus *Cupriavidus*, isolated from metal rich environments (Mergeay, 2000 and Mergeay, 2003). *C. metallidurans* strain CH34 has two megaplasmids, pMOL28 (171 Kb) and pMOL30 (234 Kb) (Janssen *et al*., 2010). The putative genomic island CMGI-28a of pMOL28 contains the *mer* (mercury), *cnr* (resistance to cobalt and nickel) and *chr* (chromium) operons (Liesegang *et al*. 1993). Resistance to Cd(II) (cadmium), Zn(II) (zink), and Co(II) (cobalt) mediated by the *czcCBA* genes is the dominant phenotypic characteristic of pMOL30 plasmid. The *czc* genes are contained in the CMGI-30a putative genomic island which also contains the *pbr* genes responsible for lead resistance (Borremans *et al*. 2001). The presence of metal resistance genes on pMOL28 and pMOL30 allow their host to occupy specific niches characterized by high levels of bioavailable toxic metals (Table 1.1).

The hydrogen oxidizing bacterium *Ralstonia eutropha* H16 and the carbon monoxide oxidizing bacterium *Oligotropha carboxidovorans* OM5 carry key genes for

their respective forms of the lithoautotrophic methabolism on megaplasmids (Paul *et al*., 2010). In *R. eutropha* H16 genes for the H_2–oxidizing system and for CO_2 fixation via the Calvin–Benson–Bassham cycle are located on the 452 Kb megaplasmid pHG1 (Schwartz *et al*., 2003). In addition, pHG1 harbors clusters of genes for denitrification and for degradation of aromatic compounds. The 133 Kb megaplasmid pHCG3 (Fuhrmann *et al*., 2003) harbors genes for CO oxidation in *O. carboxidovorans* OM5. Beside the aerobic CO dehydrogenase, pHCG3 also encodes Calvin cycle enzymes and a dimeric hydrogenase (Schwartz *et al*., 2008) (Table 1.1).

Genes that enable growth of *Arthrobacter nicotinovorans* on nicotine as carbon, nitrogen and energy source were located on the 165 Kb megaplasmid pAO1 (Igloi *et al*., 2003). Plasmid pAO1 carries genes specific for L-nicotine, D-nicotine and γ-*N*–methylaminobutyrate catabolism, for the biosyntheses of the molybdenum cofactor required by the molybdenum enzymes of the pathway, for the two subunits of a small multidrug resistance pump and for gene products for the protection from oxidative stress generated by oxidation of nicotine blue (Table 1.1).

Large genomic deletions

To identify the functional significance of proteins of unknown function, several approaches can be used. First, simple knock-out mutants can be generated, however the probability of identifying the phenotype of these mutants is rather low, as these mutations do not generally produce distinctive phenotypes. A second possibility is to generate

strains in which large regions (~40 Kb and more) are deleted and to screen these strains for altered phenotypes. In addition, large genomic deletions have been used in experimental genome reduction strategies. Attempts of minimization of genomes in the laboratory have been done with *E. coli* (Kolisnychenko *et al*., 2002; Yu *et al*., 2002; Fukiya *et al*., 2004; Hashimoto *et al*., 2005; Posfai *et al*., 2006; Mizoguchi *et al*., 2007) *B. subtilis* (Westers *et al*., 2003; Ara *et al*., 2007; Morimoto *et al*., 2008), *Corynebacterium glutamicum* (Suzuki *et al*., 2005), and *M. genitalium* (Glass *et al*., 2006) (Table 1.2).

Composite transposons. Goryshin *et al*., 2003 have described a Tn5 based deletion technology that uses a composite, linear Tn5 derivative. Preformed transposon complexes were electroporated into the cells (Goryshin *et al*., 2000) and integrated in the genome via external ends (IE). Upon integration, the product of TnpEK/LP gene was expressed and bound to the internal transposon ends (MEs) to catalyze blunt-end cleavage and loss of the donor DNA (Figure 1.1). Double-ended cleavage eliminates the transposase gene (TnpEK/LP) and a selectable marker (Km^r) was used for the initial transposon insertion selection. Tnp-EK/LP then facilitates intramolecular strand transfer into the chromosome. The intramolecular transposition event can create host genome inversions or deletions that begin at the internal end and extend to the point on the chromosome defined by the second transposition event (Goryshin *et al*., 2003). The average deletion size obtained using this approach is 11 Kb but the longest deletion made was 200 Kb (Goryshin *et al*., 2003). Composite transposons strategies for making large genomic deletions eliminate all selectable markers and transposases genes within deletions allowing the accumulation of deletions within the same strain.

λ – Red type mediated recombination method. To construct targeted deletions in *E. coli* strain MG1655, a rapid and straightforward method was developed by Kolisnychenko *et al*., 2002. To delete the chromosomal region between two defined segments, Kolisnychenko *et al*., generated a linear DNA fragment by PCR (Figure 1.2). This fragment, which carrys selectable marker (Cm^r) and which is flanked by two I-*SceI* meganuclease sites, was electroporated into a target cell, where it could be integrated into the chromosome in a presence of the helper plasmid pBADαβγ. Next, I – *SceI* meganuclease from pSTKST was expressed in a target cell and the chromosome was cleaved at 18bp I - *SceI* sites, present on the integrated fragment. The broken chromosomal ends were repaired by intramolecular recombination in a presence of RecA. Recombination sites are homologous regions close to broken ends termini (Figure 1.2.).

The size of deletions made using this strategy was 7 to 82 Kb and the genome of MG1655 was reduced by 8.1% by removing most of the selfish DNA (cryptic prophages, phage remnants, insertion sequences (ISs)) and large sets of genes of unknown function (Kolisnychenko *et al*., 2002).

CRS cassette method. The *red* recombination system of phage lambda, in combination with CRS cassette, is used for construction of markerless deletion mutants. The CRS cassette includes three genes: the chloramphenicol resistant gene (*cat* - Cm^r) and two negative selection markers: 30S ribosomal subunit protein S12 (*rpsL*) and *Bacillus subtilis* levansucrase (*sacB*). In the CRS cassette method two types of strains are constructed for a single deletion mutation (Hashimoto *et al*., 2005). One strain is constructed by replacing a chromosomal region with a CRS cassette by using a positive

marker of this cassette, (chloramphenicol-Cm^r) (Figure 1.3). This inserted cassette is removed using two – negative – selection markers (*sacB* and *rpsL*) to construct the other strain. These two types of strains are used to combine the deletion unit. The deletion unit is transferred by transduction with P1 phages prepared from the former strain (with the CRS cassette) using the positive selectable markers. Next, the cassette is removed by transduction with P1 phage prepared from the latter strain (without CRS cassette) using negative – selection markers (Figure 1.3) (Hashimoto *et al*., 2005). This strategy was used to make 41-300 Kb large genomic deletions in *E. coli* and the authors defined possible phenotypes related to deletions (Hashimoto *et al*., 2005).

Cre/LoxP excision system. The most commonly used method for deleting genes in bacterial genomes are the site – specific recombination systems: Flp/*FRT*, Cre/*loxP* and λ – *red*. Yu *et al*. (2002) employed the Cre/*loxP* recombination system to minimize the *E.coli* genome. They created two independent *E. coli* mutant libraries with transposons TnKloxP and TnCloxP that carried the Km and Cm antibiotic markers, respectively. Further, a pair of mutant strains, one from TnKloxP group and one from TnCloxP group were chosen and double - resistant strains with two loxP sites were constructed using phage P1 transduction (Figure 1.4.). Subsequent expression of Cre recombinase and recombination between the *loxP* sites, resulted in excision of the *loxP* flanked region. Using this approach, mutant strains that are missing genomic fragments of 59 – 117 Kb were obtained. The advantage of this method is being able to use two large *E. coli* transposon libraries with *loxP* sites distributed almost evenly on the chromosome, to construct a variety of deletions (Yu *et al*., 2002).

Flp/FRT recombination system

Flp is a site-specific recombinase, discovered on the 2 μm plasmid of *Saccharomyces cerevisiae* (Sadowski, 1995; Posfai *et al*.,1994; Friesen *et al.*, 1992; Cox M., 1983). Flp belongs to the integrase family of recombinases whose members share a similar mechanism of recombination, although their amino acid identity is relatively weak (Argos *et al*., 1986). Flp recombination target (FRT) sites consist of three 13 bp repeat elements surrounding an 8 bp asymmetrical core that determines the direction of the FRT site. All three repeats appear to be necessary for intermolecular recombination in bacteria, while only two repeats are necessary for intramolecular recombination（Lyznik *et al.*, 1993). The first step in the recombination reaction involves recombinase binding to the recognition target sites followed by cleavage of the phosphodiester bond at the border of the core region by a nucleophilic attack of the active site tyrosine hydroxyl group. The subsequent strand exchange reaction generates a transient Holliday intermediate and the exchange of the second pair of DNA strands completes the recombination reaction. This reaction requires no exogenous high-energy source and is carried out without degradation and synthesis of DNA. Flp-mediated recombination between two FRT sites in direct orientation causes excision of the DNA between them (Fig.1.5.a) while recombination between two inverted FRT sites results in inversion of the intervening DNA (Fig.1.5.b).

The Flp-*FRT* represent a simple two-component (recombinase and its target site) recombination system that is not dependent on added cofactors (Sadowski, 1995). It is one of the most frequently used systems, since it is the least restrictive in terms of host-

range, functioning in bacteria (Cherepanov and Wackernagel, 1995), yeasts (Sadowski, 1995), embryos of the mosquito, *Aedes aegypti* (Morris *et al.*, 1991), *Drosophila melanogaster* (Theodosiou *et al.*, 1998), plants (Lyznik *et al*., 1996) and mammalian cells (Dymecki,1996). This system is used to generate chromosomal translocations and large deletions, produce tissue specific and conditional knockouts and for the precise removal of selectable markers (Buchoiz *et al.* 1996).

Minimal genome

A minimal genome is the minimal set of genes that are necessary and sufficient to sustain a functioning cell under the most favorable conditions, in the presence of unlimited amounts of all essential nutrients and in the absence of any environmental stress (Cho *et al*., 1999; Hutchison *et al*., 1999). Initial model organisms for minimal genome studies were *Hemophilus influenza, Mycoplasma genitalium* and *Mycoplasma pneumonia.* All three bacteria have relatively small genomes. The genus *Mycoplasma* arose by massive genome reduction from a gram – positive clostridial ancestry and for that reason it was hypothesized that many of the genes that were retained during this process are essential for any living cell. To date, several approaches have been used in minimal genome studies: comparative genomics (*in silico*); random transposon mutagenesis, inhibition of gene expression by antisense RNA and biochemical approaches.

Mushegian and Koonin (1996) used the genome sequences of *Hemophilus influenza* and *Mycoplasma genitalium* to determine a set of genes with overlapping functions. A detailed comparison of the protein sets from the two bacteria revealed the minimal set of 256 genes that consisted mostly of orthologues genes with NOD (nonorthologous gene displacement) cases comprising 5% of these genes. NOD means that the same function is performed by unrelated or very distantly related and nonorthologous proteins. Despite all the advantages that a comparative genomic approach offers, the same method involves some uncertainties, as some of the common genes could still be not essential for all living cell, and identification of NOD can result in many false positive. For these reasons experimental approaches to identifying minimal gene sets became extremely important.

According to Koonin, 2000, experimentally constructing a minimal cell can help in verifying comparative genomic results, enhance our understanding of evolution and provide a valuable model system for probing the principles of cell functioning. Large-scale inactivation as an experimental approaches have been performed to define which genes are essential for cell viability in several well-characterized bacterial models, such as *E. coli*, *B. subtilis*, *Staphylococcus aureus*, and *Mycoplasma genitalium*. Strategies such as random transposon mutagenesis, the use of antisense RNA to inhibit gene expression, or the systematic inactivation of each individual gene present in a genome have been employed (Gil *et al*., 2004; Feher *et al*., 2007; Reznikoff and Winterberg, 2008; Moya *et al*., 2009). Transposon mutagenesis is one of the most popular experimental approaches for defining a minimal set of genes, but it has certain

limitations. Transposon mutagenesis might overestimate the minimal set of genes by misclassification of nonessential genes that slow growth without arresting it but can also miss essential genes that tolerate transposon insertions (Gil *et al*., 2004). Transposon mediated knockout mutagenesis in *M. genitalium* and *M. pneumonia* resulted in the identification of 215 nonessential genes out of 480. The number of nonessential genes obtained by this approach suggests a minimal gene set of 265 genes which is close to size produced by the comparative genomic approach. The significant number (38) of viable disruptions within the set of conserved genes predicted by Mushegian and Koonin (1996), is unexpected and could suggest that evolutionary conservation of a gene does not necessarily translate into it being essential under all conditions (Koonin, 2000). For this reason, a minimal genome project makes sense only if conditions under which the respective "minimal organism" should be expected to survive, are defined.

Antisense RNA has been used to inhibit gene expression in *S. aureus*. Essential genes were identified after conditionally expressing random genome fragments and then screening for fragments whose expression blocks growth. The genes targeted by antisense RNA were identified by DNA sequencing and BLAST analysis against the annotated sequence of *S. aureus* (Ji *et al*., 2001). This method identified 658 unique essential genes in *S. aureus* genome. The use of antisense RNA to inhibit gene expression is limited to the genes for which adequate expression of the inhibitory RNA could obtained in the organism under study. Ji *et al*., (2001) showed that the efficiency of this method is several orders of magnitude greater than that of any other conditional growth phenotypic selection procedure.

Beside computational and experimental approaches, Forster and Church (2006) used a biochemical approach to define a minimal set of genes. They searched for all the components of well defined pathways that are important for essential functions and came up with a proposed minimal genome that contains 151 genes; 38 RNA genes plus 113 protein – coding genes involved in essential processes (replication, transcription and translation). This approach is quite different from any computational or experimental approach already described. It is completely composed of genes with well – defined functions.

Why Make Minimal Genome?

According to Kolisnychenko *et al*., (2002) who reduced the genome of *E. coli* strain MG1655 by 8.1%, there are several reasons for size reduction of bacterial genomes. First, making a strain with increased genome stability by removing transposons and IS elements, which are usually the major source of mutations, would be desirable for many scientific experiments (Kolisnychenko *et al*., 2002). Second, a minimal gene set can help us to further understand cell functioning as well as enable us to define the subset of genes that are expected to be essential for life. Starting with cells with a minimal gene set will be preferable for inserting new genetic modules or modifying existing ones to design microorganisms for different purposes (Andrianantoandro *et al*., 2006).

The *E. coli* minimal strain created by Kolisnychenko *et al*., (2002), consists of the back bone (3.7Mb), which includes core functions of *E. coli* that are necessary in its

typical environmental niche. The rest of the genome (0.9 Mb) in genomic islands (toxins, virulence factors, transposable elements, phages, cryptic prophages, pseudogenes and gene remnants) had been deleted (Kolisnychenko *et al*., 2002). Expectations of the minimal *E. coli* strain are that this strain is less sustainable to DNA rearrangements (inversions, duplications and deletions) than its parent and according to the authors this strain can be used to answer the question what horizontal transmissions would it pick up from the environment?

Genome reduction may decrease the redundancy among *E. coli* genes and regulatory elements. DNA sequence repeats mediate inversions, duplications and deletions by homologous recombination events without transposases. To stabilize the genome and metabolic pathways these elements can be deleted and unwanted functions removed (Kolisnychenko *et al*., 2002).

Biotechnological reasons for a reduced genome include unnecessary gene products that are expressed in a production host represent potential contaminants that could drive up the cost of product purification. Deletion of genes is the most reliable and effective way to ensure the complete absence of an unwanted component in a biotechnological product (Kolisnychenko *et al*., 2002).

Toxin Antitoxin Systems in Bacteria

Twenty eight years ago an operon that controls segregation, *ccdBA* (coupled cell division), was identified on the F – plasmid of *E. coli* (Ogura and Hiraga, 1983). It was

found that *ccdB* acts as a toxin by killing cells that become plasmid free. Since then many operons with a similar organization which are capable of stabilizing plasmids have been identified and named 'toxin-antitoxin (TA) systems or modules'. Many TA loci are two-component systems that code for a stable toxin and an unstable antitoxin. According to the antitoxin nature TA systems are divided into two types: type I and type II. In type I the antitoxin is antisense RNA (Gerdes *et al*., 1986; Greenfield *et al*., 2000) and in type II the antitoxin is a protease sensitive protein (Ogura and Hiraga, 1983; Butts *et al*., 2005; Gerdes *et al*., 2005). Stable toxins are neutralized by unstable antisense RNAs or protein antitoxin. In plasmid free cells degradation of unstable molecules (antisense RNAs and antitoxins) is rapid and in the case of antitoxins protein degradation is mediated by the Lon and Clp proteases. At the same time toxic protein or mRNA that will allow synthesis of a toxic protein is present in a plasmid free cell leading to post segregational killing (PSK) or stasis of those cells (Gerdes *et al*., 2005).

Well characterized bacterial toxin antitoxin loci have been classified into eight families (Gerdes *et al*., 2005) (Table 1.3). Most of TA loci have been found on plasmids as well as on chromosomes.

To understand the function of TA loci it is important to know the target(s) of the toxins. So far two targets of TA loci have been identified: DNA gyrase and mRNA. Members of the CcdB (Bernard and Couturier, 1992) and ParE (Jiang *et al*., 2002) families inhibit DNA gyrase, an essential topoisomerase II. Type II topoisomerases are required during cell division to aid in chromosomal segregation and in establishing the proper superhelical density. DNA gyrase is unique in that it is able to catalyze negative

supercoiling in DNA at the expense of ATP. CcdB poisons DNA gyrase by trapping the gyrase in its cleavable complex with DNA, eventually causing double strand breaks in the DNA. During the transition process, it was shown that CcdB binds to the A subunit of gyrase leading to obstruction of DNA and RNA polymerase activity by creating dead-end complex (Critchlow *et al*., 1997).

Toxins from the RelE family mediate cleavage of mRNA in a ribosome dependent manner (Pedersen *et al*., 2003). It was shown that RelE induces RNAse activity of a ribosome (Hayes and Sauer, 2003), thereby affecting the level of protein synthesis. Toxins from MazF and YoeB families are highly specific ribonucleases. MazF specifically cleaves at ACA sequences (Zhang *et al*., 2003), while YoeB is purine – specific ribonuclease that cleaves at adenine tracts (Christensen *et al*., 2004). Experimental data available for the VapC toxin (Anantharaman and Aravind, 2003) suggests the presence of RNAse activity within the toxin molecule based on predicted homology of VapC with PIN domains. PIN domain proteins are ribonucleases involved in nonsense – mediated RNA degradation in eukaryotes.

TA operons are autoregulated at the level of transcription by the complex formed between the toxin and the antitoxin (de Feyter *et al*., 1989). Only antitoxins bind directly to the TA locus promoters, but the toxin can assist as co - repressors of transcription. The N – terminal domains of CcdA and MazE antitoxins are responsible for dimerization and DNA binding, while their C – terminal domains are important for interaction with the dimeric toxin. Toxin dimers can bridge antitoxin dimers bound at two or more distinct sites on the DNA in a way that the interaction becomes cooperative (Dao – Thi *et al*.,

2002). The presence of the toxin enables better binding of the TA complex to the TA locus promoter and more efficient repression of transcription. Results obtained by Afif *et al*., (2001); Johnson *et al*., (1996); Magnuson and Yarmolinsky, (1998), for *ccdA*/*ccdB*, *parD*/*parE* and *phd*/*doc* indicate that the ratio of toxin to antitoxin might have an influence on the transcriptional regulation of TA operon..

Gerdes *et al*., (2005) proposed two models that could explain the activation of *relBE* locus, a simple passive model and an active model. The passive model proposed that induction of *relBE* loci occur passively without involvement of external signal. In this situation *relBE* loci respond to intracellular changes which are caused by starvation. During starvation, the reduced rate of translation of the antitoxin encoding mRNA leads to a reduced level of antitoxin which further leads to derepression of the *relBE* locus promoter. The increased toxin activity might be the result of combined effects of an increased *relBE* operon transcription rate and reduced level of antitoxin (Gerdes *et al*., 2005).

The active model proposed that starvation induces expression of the Lon protease which further degrades the RelB antitoxin. Signals that trigger the Lon protease are not known, but it could be a compound of a low molecular weight, a protein or stalled ribosome (Gerdes *et al*., 2005).

An analysis of all available data regarding function of chromosomal toxin antitoxin systems (Magnuson, 2007) resulted in a list of hypothetical functions.

a) Chromosomal toxin antitoxin systems are genomic junk, acquired from plasmids and/or other sources.

b) Chromosomal toxin antitoxin systems are part of large genomic parasites such as conjugative transposones (Dziewit *et al*., 2007) and temperate bacteriophages (DeShazer, 2004). Suggesting function to prevent elimination by recombination.

c) Some of the toxins act as specific ribonucleases, they can shift gene expression away from site - rich to site - poor messages and in such a way be involved in autoregulation (Ruiz – Echevarria *et al*.,1995), specific (Munoz – Gomez *et al*., 2005) or global regulation (Engelberg – Kulka *et al*., 2006).

d) Many toxin - antitoxin systems are activated by starvation and other stresses (Agguire – Ramirez *et al*., 2006; Buts *et al*., 2005). In most cases toxins, are responsible for bacteriostatic but not bactericidal effects under stressful conditions. The quick arrest of growth in response to starvation, may permit improved survival during starvation or a quicker resumption of growth when conditions improve (Gerdes, 2000).

e) Toxin – antitoxin systems can act as antiphage systems. Bacteriophages can activate toxin - antitoxin systems, which would then limit phage production (Hazan and Engelberg – Kulka, 2004), unless the phage could inhibit host proteases (Engelberg – Kulka *et al*., 1998), neutralize the toxins or exit the cell before the toxins were activated.

f) Toxin - antitoxin systems are also involved in generation of “persisters”, subpopulation of bacterial cells that show low growth and high resistance to some antibiotics (Keren *et al*., 2004; Korch *et al*., 2003).

g) Certain experimental data suggest that chromosomal toxin – antitoxin systems can be involved in programmed cell death (Aizenman *et al*., 1996; Lewis 2000).

This work

The aim of this work was to adapt the Flp/FRT system for use in *S. meliloti* to delete defined regions of the genome. More specifically, we wished to generate multiple gene knock-out mutants using one of the approaches for making large genomic deletions. We believed both *S. meliloti* megaplasmids pSymA and pSymB can be systematically deleted. Although others generated a limited number of large defined deletions in the pSymB (Charles and Finan, 1991) and pSymA (House *et al.*, 2003) megaplasmids of *S. meliloti* using the transposon insertions and Flp/FRT system, respectively as an experimental approach, systematic deletion analysis of both megaplasmids (pSymA and pSymB) of *S. meliloti* had not been done to date. We were able to identify phenotypes associated with deleted genes and identify the essential genes in *S. meliloti* genome.

The major findings were: a) two essential genes, a $tRNA^{arg}$ and *engA* were located on the pSymB megaplasmid of *S. meliloti*, b) new carbohydrate utilization phenotypes, associated with specific deletions were discovered, c) a 40 Kb region on pSymB which harbors a gene involved in symbiotic N_2 - fixation was mapped, d) despite a previous report (Oresnik *et al*., 2000.) describing the curing of the entire pSymA megaplasmid, deletions of the entire pSymA were not obtained, e) three toxin/antitoxin-

like systems on pSymA megaplasmid were identified. These are *sma0471* /*sma0473*, *sma2105* and *sma2230/sma2231*, f) Deletion of the *sma0471/sma0473*, resulted in reduced growth, a loss of cell viability (*sma2105* and *sma2230/sma2231*) and such deletion derivatives were rarely recovered. These phenotypes were complemented when the *sma0471/sma0473*, *sma2105* and *sma2230/sma2231* genes were supplied *in trans*.

The results of this thesis are presented in three chapters.

Chapter 3: Deletion strategy using FRT targeting vector pTH1522 and pTH1937, presented as a part of a manuscript to be published following review by the other author, T. M. Finan, and inclusion of alterations recommended.

Chapter 4: Deletion analysis of the *Sinorhizobium meliloti* pSymB megaplasmid. This chapter was written as a manuscript to be published following review by the other author T. Soh, G. diCenzo, T. M. Finan, and inclusion of alterations recommended.

Chapter 5: Deletion Analysis of the *Sinorhizobium meliloti* pSymA megaplasmid. This chapter was written as a manuscript to be published following review by the other author T. Soh, G. diCenzo, T. M. Finan, and inclusion of alterations recommended.

REFERENCES

Afif, H., N. Allali, M. Couturier, L. Van Melderen. (2001) The ratio between CcdA and CcdB modulates the transcriptional repression of the *ccd* poison-antidote system. Mol. Microbiol. **41**:73-82.

Aguirre-Ramírez, M., J. Ramírez-Santos, L. Van Melderen, M.C. Gómez-Eichelmann. (2006) Expression of the F plasmid *ccd* toxin-antitoxin system in *Escherichia coli* cells under nutritional stress. Can. J. Microbiol. **52**:24-30.

Aizenman, E., H. Engelberg-Kulka, G. Glaser. (1996) An *Escherichia coli* chromosomal "addiction module" regulated by guanosine [corrected] 3',5'-bispyrophosphate: a model for programmed bacterial cell death. Proc. Natl. Acad. Sci. U. S. A. **93**:6059-63.

Anantharaman,V., and L. Aravind. (2003) New connections in the prokaryotic toxin-antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system. Genome. Biol. **4**:R81.

Andrianantoandro, E., S. Basu, D.K. Karig, R. Weiss (2006) Synthetic biology: new engineering rules for an emerging discipline. Mol. Syst. Biol. 2. DOI: 10.1038/msb4100073.

Aoki, T., T. Akashi, S. Ayabe. (2000) Flavonoids of leguminous plants: structure, biological activity, and biosynthesis. J. Plant Res. **113**:475–488.

Ara, K., K. Ozaki, K. Nakamura, K. Yamane, J. Sekiguchi, N. Ogasawara. (2007) *Bacillus* minimum genome factory: effective utilization of microbial genome information. Biotechnol. Appl. Bioc. **46**:169–178.

Arner, E.S., A. Holmgren. (2000) Physiological functions of thioredoxin and thioredoxin reductase. Eur. J. Biochem. **267**: 6102 – 6109.

Banfalvi, Z., V. Sakanyan, C. Koncz, A. Kiss, I. Dusha, 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.D., T.M. Finan. (1998) Regulation of phosphate assimilation in *Rhizobium* (*Sinorhizobium*) *meliloti*. Genetics. **148**:1689-700.

Barnett, M.J., R.F. Fisher, T. Jones, C. Komp, A.P. Abola, F. Barloy-Hubler, L. Bowser, D. Capela, F. Galibert, J. Gouzy, M. Gurjal, A. Hong, L. Huizar, R.W. Hyman, D. Kahn, M.L. Kahn, S. Kalman, D.H. Keating, C. Palm, M.C. Peck, R. Surzycki, D.H. Wells, K.C. Yeh, R.W. Davis, N.A. Federspiel, S.R. Long. (2001) Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. Proc.Natl.Acad.Sci.U.S.A. **98**:9883-8.

Bartsev, A.V., W.J. Deakin, N.M. Boukli, C.B. McAlvin, G. Stacey, P. Malnoë, W.J. Broughton, C. Staehelin. (2004) NopL, an effector protein of *Rhizobium* sp. NGR234, thwarts activation of plant defense reactions. Plant Physiol.**134**:871-9.

Bentley S.D., J. Parkhill. (2004) Comparative genomic structure of prokaryotes. Annu. Rev. Genet. **38:**771-92.

Bernard, P., M. Couturier. (1992) Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. J. Mol. Biol. **226**:735-45.

Borremans B., J.L. Hobman, A. Provoost, N.L. Brown, D. van Der Lelie. (2001) Cloning and functional analysis of the *pbr* lead resistance determinant of *Ralstonia metallidurans* CH34. J.
Bacteriol. **183**:5651 – 5658.

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

Buts, L., J. Lah, M.H. Dao-Thi, L. Wyns, R. Loris. (2005) Toxin-antitoxin modules as bacterial metabolic stress managers. Trends. Biochem. Sci. **30**:672-9.

Capela, D., F. Barloy-Hubler, J. Gouzy, G. Bothe, F. Ampe, J. Batut, P. Boistard, A. Becker, M. Boutry, E. Cadieu. (2001) Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. Proc. Natl. Acad. Sci. U.S.A. **98:** 9877–9882

Chao, L., B.R. Levin. (1981) Structured habitats and the evolution of anticompetitor toxins in bacteria. Proc. Natl. Acad. Sci. U.S.A. **78**:6324-8.

Charles, T.C., T.M. Finan. (1991) Analysis of a 1600-kilobase *Rhizobium meliloti* megaplasmid using defined deletions generated *in vivo*. Genetics **127**:5-20.

Cherepanov, P.P., 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.

Cho, M.K., D. Magnus, A.L. Caplan, D. McGee. (1999) Policy forum: genetics—ethical considerations in synthesizing a minimal genome. Science **286**:2087–90

Christensen, S.K., G. Maenhaut-Michel, N. Mine, S. Gottesman, K. Gerdes, L. Van Melderen. (2004) Overproduction of the Lon protease triggers inhibition of translation in *Escherichia coli*: involvement of the *yefM-yoeB* toxin-antitoxin system Mol. Microbiol. **51**:1705-17.

Cooper, J.E. (2004) Multiple responses of rhizobia to flavonoids during legume root infection. Adv. Bot. Res. **41**:1–62.

Cooper, J.E. (2007) Early interactions between legumes and rhizobia: disclosing complexity in a molecular dialogue. J. Appl. Microbiol. **103**:1355 – 1365.

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

Critchlow, S.E., M.H. O'Dea, A.J. Howells, M. Couturier, M. Gellert, A. Maxwell. (1997) The interaction of the F plasmid killer protein, CcdB, with DNA gyrase: induction of DNA cleavage and blocking of transcription. J. Mol. Biol. **273**:826-39.

Dao-Thi, M.H., D. Charlier, R. Loris, D. Maes, J. Messens, L. Wyns, J. Backmann. (2002) Intricate interactions within the *ccd* plasmid addiction system. J. Biol. Chem. **277**:3733-42.

DasSarma, S. (2004) Genome sequence of an extremely halophilic archaeon. In: Fraser C.M., Read T., Nelson K.E. (eds) Microbial Genomes, Humana, Totowa, NJ pp:383-399

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

de Feyter, R., C. Wallace, D. Lane. (1989) Autoregulation of the *ccd* operon in the F plasmid. Mol. Gen. Genet. **218**:481-6.

DeShazer, D. (2004) Genomic diversity of *Burkholderia pseudomallei* clinical isolates: subtractive hybridization reveals a *Burkholderia mallei*-specific prophage in *B. pseudomallei* 1026b. J. Bacteriol. **186**:3938-50.

Dymecki, S.M. (1996) Flp recombinase promotes site-specific DNA recombination in embryonic stem cells and transgenic mice. Proc. Natl. Acad. Sci. U.S.A. **93**:6191-6.

Dziewit, L., M. Jazurek, L. Drewniak, J. Baj, D. Bartosik. (2007) The SXT conjugative element and linear prophage N15 encode toxin-antitoxin-stabilizing systems homologous to the *tad-ata* module of the *Paracoccus aminophilus* plasmid pAMI2. J. Bacteriol. **189**:1983-97.

Engelberg-Kulka, H., M. Reches, S. Narasimhan, R. Schoulaker-Schwarz, Y. Klemes, E. Aizenman, G. Glaser. (1998) *rexB* of bacteriophage lambda is an anti-cell death gene. Proc. Natl. Acad. Sci. U.S.A. **95**:15481-6.

Engelberg-Kulka, H., S. Amitai, I. Kolodkin-Gal, R. Hazan. (2006) Bacterial programmed cell death and multicellular behavior in bacteria. PLoS Genet. **2**:e135.

Fehér, T., B. Papp, C. Pal, G. Pósfai. (2007) Systematic genome reductions: theoretical and experimental approaches.Chem Rev. **107**:3498-513.

Finan, T.M., A.M. Hirsch, J.A. Leigh, E. Johansen, G.A. Kuldau, S. Deegan, G.C. Walker, 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, E.R. Signer. (1986) Second symbiotic megaplasmid in *Rhizobium meliloti* carrying exopolysaccharide and thiamine synthesis genes. J. Bacteriol. **167**:66-72.

Finan, T.M., S. Weidner, K. Wong, J. Buhrmester, P. Chain, F. J. Vorhölter, I. Hernandez-Lucas, A. Becker, A. Cowie, J. Gouzy, B. Golding, A. Pühler. (2001) The complete sequence of the 1,683-Kb pSymB megaplasmid from the N_2-fixing endosymbiont *Sinorhizobium meliloti*. Proc.Natl.Acad.Sci.U.S.A. **98**:9889-94.

Fisher, R.F., S.R. Long. (1993) Interactions of NodD at the nod Box: NodD binds to two distinct sites on the same face of the helix and induces a bend in the DNA. J. Mol. Biol. **233**:336-48.

Forster, A.C., G.M. Church. (2006) Towards synthesis of a minimal cell. Mol. Syst. Biol. **2**: 45.

Friesen, H., P.D. Sadowski. (1992) Mutagenesis of a conserved region of the gene encoding the FLP recombinase of *Saccaharomyces cerevisiae*. J.Mol.Biol. **225**:313-326.

Fuhrmann S., M. Ferner, T. Jeffke, A. Henne, G. Gottschalk, O. Meyer. (2003) Complete nucleotide sequence of the circular megaplasmid pHCG3 of *Oligotropha carboxidovorans* : function in the chemolithoautotrophic utilization of CO, H_2 and CO_2. Gene **322**:67 – 75.

Fukiya, S., H. Mizoguchi, H. Mori. (2004) An improved method for deleting large regions of *Escherichia coli* K-12 chromosome using a combination of *Cre*/*loxP* and lambda red. FEMS Microbiol. Lett. **234**: 325–331.

Gage, D.J., S.R. Long. (1998) alpha-Galactoside uptake in *Rhizobium meliloti*: isolation and characterization of *agpA*, a gene encoding a periplasmic binding protein required for melibiose and raffinose utilization. J. Bacteriol. **180**:5739-48

Gage, D.J. (2004) Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. Microbiol. Mol. Biol. Rev. **68**:280-300.

Galibert**,** F., T.M. Finan, S.R. Long, A. Pühler, P. Abola, F. Ampe, F. Barloy-Hubler, M.J. Barnett, A. Becker, P. Boistard, G. Bothe, M. Boutry, L. Bowser, J. Buhrmester, E. Cadieu, D. Capela, P. Chain, A. Cowie, R.W. Davis, S. Dréano, N.A. Federspiel, R.F. Fisher, S. Gloux, T. Godrie, A. Goffeau, B. Golding, J. Gouzy, M. Gurjal, I. Hernandez-Lucas, A. Hong, L. Huizar, R.W. Hyman, T. Jones, D. Kahn, M.L. Kahn, S. Kalman, D.H. Keating, E. Kiss, C. Komp, V. Lelaure, D. Masuy, C. Palm, M.C. Peck, T.M Pohl, D. Portetelle, B. Purnelle, U. Ramsperger, R. Surzycki, P. Thébault, M. Vandenbol, F.-J. Vorhölter, S. Weidner, D.H. Wells, K. Wong, K-C. Yeh, J. Batut. (2001) The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science. **293**:668-672.

Gerdes, K. P.B. Rasmussen, S. Molin. (1986) Unique type of plasmid maintenance function, postsegregational killing of plasmid free cells. Proc. Natl. Acad. Sci. U.S.A. **83**:3116-3120.

Gerdes, K. (2000) Toxin-antitoxin modules may regulate synthesis of macromolecules during nutritional stress. J. Bacteriol. **182**:561-72.

Gerdes, K., S.K. Christensen, A. Lobner-Olesen. (2005) Prokaryotic toxin-antitoxin stress response loci. Nat. Rev. Microbiol. **3**:371-82.

Gil, R., F.J. Silva, J. Peretó, A. Moya. (2004) Determination of the core of a minimal bacterial gene set. Microbiol. Mol. Biol. Rev. **68**:518-37.

Glass, J.I., N. Assad-Garcia, N. Alperovich, S. Yooseph, M.R. Lewis, M. Maruf, C.A. III Hutchison, H.O. Smith, J.C. Venter. (2006) Essential genes of a minimal bacterium. Proc. Natl. Acad. Sci. U.S.A.**103**: 425–430.

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

González, V., R.I. Santamaria, P. Bustos, I. Hernández-González, A. Medrano-Soto, G. Moreno-Hagelsieb, S.C. Janga, M.A. Ramírez, V. Jiménez-Jacinto, J. Collado-Vides. (2006) The partitioned *Rhizobium etli* genome: genetic and metabolic redundancy in seven interacting replicons. Proc. Natl. Acad. Sci. U.S.A. **103:** 3834–3839

Goryshin, I.Y., T.A. Naumann, J. Apodaca, W.S. Reznikoff. (2003) Chromosomal deletion formation system based on Tn5 double transposition: use for making minimal genomes and essential gene analysis. Genome Res. **13**:644-53.

Greenfield, T.J., E. Ehli, T. Kirshenmann, T. Franch, K. Gerdes, K.E.Weaver. (2000) The antisense RNA of the *par* locus of pAD1 regulates the expression of a 33-amino-acid toxic peptide by an unusual mechanism. Mol. Microbiol **37**:652-60.

Hashimoto, M., T. Ichimura, H. Mizoguchi (2005) Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. Mol. Microbiol. **55**:137–149.

Hayes, C.S., R.T. Sauer. (2003) Cleavage of the A site mRNA codon during ribosome pausing provides a mechanism for translational quality control. Mol. Cell. **12**:903-11.

Hazan, R., H. Engelberg-Kulka. (2004) *Escherichia coli mazEF*-mediated cell death as a defense mechanism that inhibits the spread of phage P1. Mol. Genet. Genomics. **272**:227-34.

House, B.L., M.W. Mortimer, M.L. Kahn. (2004) New recombination methods for *Sinorhizobium meliloti* genetics. Appl. Environ. Microbiol**. 70:2806-15.**

Hynes, M.F., R. Simon, P. Muller, K. Niehaus, M. Labes, A. Puhler. (1986) The two megaplasmids of *Rhizobium meliloti* are involved in the effective nodulation of alfalfa. Mol. Gen. Genet. **202**:356-362.

Hu P.C., J. Elliott, P. McCready, E. Skowronski, J. Garnes, A. Kobayashi, R.R. Brubaker, E. Garcia. (1998) Structural organization of virulence-associated plasmids of *Yersinia pestis*. J. Bacteriol. **180**: 5192 – 5202.

Huang X.Z., M.P. Nikolich, L.E. Lindler. (2006) Current trends in plague research: from genomics to virulence. Clin. Med. Res. **4**:189 – 199.

Hutchison, C.A., S.N. Peterson, S.R. Gil, R.T. Cline, O. White O. (1999) Global transposon mutagenesis and a minimal *Mycoplasma* genome. Science **286**:2165–69.

Igloi, G.L., R. Brandsch. (2003) Sequence of the 165-kilobase catabolic plasmid pAO1 from *Arthrobacter nicotinovorans* and identification of a pAO1-dependent nicotine uptake system.J. Bacteriol. **185**:1976-86.

Jaffe, A., T. Ogura, S. Hiraga. (1985) Effects of the *ccd* function of the F plasmid on growth. J. Bacteriol. **163**: 841 – 849.

Janssen, P.J., R. Van Houdt, H. Moors, P. P. Monsieurs, N. Morin, A. Michaux, M.A. Benotmane, N. Leys, T. Vallaeys, A. Lapidus, S. Monchy, C. Médigue, S. Taghavi, S. McCorkle, J. Dunn, D. van der Lelie, M. Mergeay. (2010) The complete genome sequence of Cupriavidus metallidurans strain CH34, a master survivalist in harsh and anthropogenic environments. PloS One. **5**(5): e10433.

Jebbar, M., L. Sohn-Bösser, E. Bremer, T. Bernard, C. Blanco. (2005) Ectoine-induced proteins in *Sinorhizobium meliloti* include an Ectoine ABC-type transporter involved in osmoprotection and ectoine catabolism. J. Bacteriol. **187**:1293-304.

Jelesko, J.G., J.A. Leigh. (1994) Genetic characterization of a *Rhizobium meliloti* lactose utilization locus. Mol. Microbiol. **11**:165-73.

Ji, Y., B. Zhang, S.F. Van Horn, P. Warren, G. Woodnutt, M.K. Burnham, M. Rosenberg. (2001) Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. Science. **293**:2266-9.

Jiang, Y., J. Pogliano, D.R. Helinski, I. Konieczny. (2002) ParE toxin encoded by the broad-host-range plasmid RK2 is an inhibitor of *Escherichia coli* gyrase. Mol. Microbiol. **44**:971-9.

Johnson, E.P., A.R. Strom, D.R. Helinski. (1996) Plasmid RK2 toxin protein ParE: purification and interaction with the ParD antitoxin protein. J. Bacteriol. **178**:1420-9.

Kaneko, T., Y. Nakamura, S. Sato, E. Asamizu, T. Kato, S. Sasamoto, A. Watanabe, K. Idesawa, A. Ishikawa, K. Kawashima. (2000) Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. DNA Res. 7**:** 331–338

Kaneko, T., Y. Nakamura, S. Sato, K. Minamisawa, T. Uchiumi, S. Sasamoto, A.Watanabe, K. Idesawa, M. Iriguchi, K. Kawashima. (2002) Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. DNA Res. **9:** 189–197.

Keren, I., D. Shah, A. Spoering, N. Kaldalu, K. Lewis. (2004) Specialized persister cells and the mechanism of multidrug tolerance in *Escherichia coli*. J. Bacteriol. **186**:8172-80.

Kolisnychenko, V., G. Plunkett III, C.D. Herring, T. Fehér, J. Pósfai, F. R. Blattner, and G. Pósfai. (2002) Engineering a reduced *Escherichia coli* genome. Genome Res. **12**:640-7.

Koonin, E.V. (2000) How Many Genes Can Make a Cell: The Minimal-Gene-Set Concept. Annu. Rev. Genomics Hum. Genet. **01**:99-116.

Korch, S.B., T.A. Henderson, T.M. Hill. (2003) Characterization of the hipA7 allele of *Escherichia coli* and evidence that high persistence is governed by (p)ppGpp synthesis. Mol. Microbiol. **50**:1199-213.

Lewis, K. (2000) Programmed death in bacteria. Microbiol. Mol. Biol. Rev. **64**:503-14.

Liesegang H., K. Lemke, R.A. Siddiqui, H.G. Schlegel. (1993) Characterization of the inducible nickel and cobalt resistance determinant *cnr* from pMOL28 of *Alcaligenes eutrophus* CH34. J.
Bacteriol. **175**:767 – 778.

Lindler L.E., G.V. Plano, V. Burland, G.F. Mayhew, F.R. Blattner. (1998) Complete DNA sequence and detailed analysis of the *Yersinia pestis* KIM5 plasmid encoding murine toxin and capsular antigen. Infect. Immun. **66**:5731 – 5742.

Long, S.R. (1989) *Rhizobium* Genetics. Annual Review of Genetics. **23**: 483-506.

Long, S.R., W.J. Buikema, F.M. Ausubel. (1982) Cloning of *Rhizobium meliloti* genes by direct complementation of Nod$^-$ mutants. Nature **298**:485-488.

Lyznik, L.A., K.V. Rao, T.K. Hodges. (1996) FLP-mediated recombination of FRT sites in the maize genome. Nucleic. Acids. Res. **24**:3784-9.

Lyznik, L. A., J. C. Mitchell, L. Hirayama, T. K. Hodges. (1993). Activity of yeast FLP recombinase in maize and rice protoplasts. Nucleic. Acids. Res. **21**:969-75.

MacLean, A.M., G. MacPherson, P. Aneja, T.M. Finan. (2006) Characterization of the beta-ketoadipate pathway in *Sinorhizobium meliloti*. Appl. Environ. Microbiol. **72**:5403-13.

MacLean, A.M., C.E. White, J.E. Fowler, T.M. Finan. (2009) Identification of a hydroxyproline transport system in the legume endosymbiont *Sinorhizobium meliloti*. Mol. Plant Microbe Interact. **22**:1116-27.

MacLellan S.R., R. Zaheer, A.L. Sartor, A.M. MacLean, T.M. Finan. (2006) Identification of a megaplasmid centromere reveals genetic structural diversity within the *repABC* family of basic replicons. Mol. Microbiol. **59**:1559-75.

Maeda, K., H. Nojiri, M. Shintani, T. Yoshida, H. Habe, T. Omori. (2003) Complete nucleotide
sequence of carbazole/dioxin-degrading plasmid pCAR1 in *Pseudomonas resinovorans* strain
CA10 indicates its mosaicity and the presence of large catabolic transposon Tn4676. J. Mol. Biol. **326**:21–33

Magnuson, R., M.B. Yarmolinsky. (1998) Corepression of the P1 addiction operon by Phd and Doc. J. Bacteriol. **180**:6342-51.

Magnuson, R.D. (2007) Hypothetical functions of toxin-antitoxin systems. J. Bacteriol. **189**:6089-92.

McCorkle, J. Dunn, D. van der Lelie, M. Mergeay. (2010) The complete genome sequence of *Cupriavidus metallidurans* strain CH34, a master survivalist in harsh and anthropogenic environments PLoS One. **5**:e10433.

Mergeay M. (2000) Bacteria adapted to industrial biotopes: the metal resistant *Ralstonia*. In:

Hengge-Aronis G. Storz and R. Hengge-Aronis (eds) “Bacterial Stress Responses”. ASM Press, Washington D.C. USA , Chap.26 pp.403 – 414 .

Mergeay M., S. Monchy, T. Vallaeys, V. Auquier, A. Benotmane, P. Bertin, S. Taghavi, J. Dunn, D. van der Lelie, R. Wattiez (2003) *Ralstonia metallidurans*, a bacterium specifically adapted to toxic metals: towards a catalogue of metal-responsive genes. FEMS Microbiol. Rev. **27**: 385 – 410.

Miller, M.J., R.B. Gennis. (1983) The purification and characterization of the cytochrome d terminal oxidase complex of the *Escherichia coli* aerobic respiratory chain. J. Biol. Chem. **258**: 9159 – 9165.

Mizoguchi, H., H. Mori, T. Fujio. (2007) *Escherichia coli* minimum genome factory. Biotechnol. Appl. Bioc. **46**:157–167.

Morimoto, T., R. Kadoya, K. Endo. (2008) Enhanced recombinant protein productivity by genome reduction in *Bacillus subtilis*. DNA Res. **15**:73–81.

Morris, A. C., T. L. Schaub, A. A. James. **(**1991) FLP-mediated recombination in the vector mosquito, *Aedes aegypti*. Nucleic. Acids. Res. **19:**5895-900.

Moya, A., N. Krasnogor, J. Peretó, A. Latorre. (2009) Goethe's dream. Challenges and opportunities for synthetic biology. EMBO Rep. 10 Suppl **1**:S28-32.

Muñoz-Gómez, A.J., M. Lemonnier, S. Santos-Sierra, A. Berzal-Herranz, R. Díaz-Orejas. (2005) RNase/anti-RNase activities of the bacterial *parD* toxin-antitoxin system. J. Bacteriol. **187**:3151-7.

Mushegian, A.R., E.V. Koonin. (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc. Natl. Acad. Sci. U.S.A. **93**:10268-73.

Ng, W-L, S. DasSarma. (1991) Physical and genetic mapping of the unstable gas vesicle plasmid in *Halobacterium halobium* NRC-1. In: Rodriguez-Valera F (ed) Genetic and applied aspects of halophilic microorganisms, vol **201** . Plenum, New York, pp 305 – 311.

Ng, W-L, S. Kothakota, S. DasSarma. (1991) Structure of the gas vesicle plasmid in *Halobacterium halobium* inversion isomers, inverted repeats, and insertion sequence. J. Bacteriol **173**: 1958 – 1964.

Nojiri, H, H. Sekiguchi, K. Maeda, M. Urata, S. Nakai, T. Yoshida, H. Habe, T. Omori. (2001) Genetic characterization and evolutionary implications of a *car* gene cluster in the carbazole degrader *Pseudomonas* sp. strain CA10. J. Bacteriol. **183**: 3663–3679.

Oberhettinger, P., M. Schütz, G. Raddatz, H. Keller, I.B. Autenrieth, D. Linke. (2011) The sequence of the pYV virulence plasmid from *Yersinia enterocolitica* strain WA-314 biogroup 1B serotype O:8 Plasmid. **65**:20-4.

Ogura, T., S. Hiraga. (1983) Mini-F plasmid genes that couple host cell division to plasmid proliferation. Proc. Natl. Acad. Sci. U.S.A. **80**:4784-4788.

Oldroyd, G.E., J.A. Downie. (2008) Coordinating nodule morphogenesis with rhizobial infection in legumes. Annu. Rev. Plant Biol. **59**:519-46.

Oresnik, I.J., S.L. Liu, C.K. Yost, M.F. Hynes. (2000) Megaplasmid pRme2011a of *Sinorhizobium meliloti* is not required for viability. J. Bacteriol. **182**:3582-6**.**

Paul, D, S.M. Bridges, S.C. Burgess, Y.S. Dandass, M.L. Lawrence. (2010) Complete genome and comparative analysis of the chemolithoautotrophic bacterium *Oligotropha carboxidovorans OM5* BMC Genomics. **11**:511.

Pedersen, K., A.V. Zavialov, M.Y. Pavlov, J. Elf, K. Gerdes, M. Ehrenberg. (2003) The bacterial toxin RelE displays codon-specific cleavage of mRNAs in the ribosomal A site. Cell. **112**:131-40.

Porter, M.E., C.J. Dorman. (1997) Differential regulation of the plasmid-encoded genes in the *Shigella flexneri* virulence regulon Mol. Gen. Genet. **256**:93-103

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

Posfai, G., G.III Plunket, T. Feher. (2006) Emergent properties of reduced-genome *Escherichia coli*. Science **312**:1044–1046.

Poysti, N.J., E.D. Loewen, Z. Wang, I.J. Oresnik. (2007) *Sinorhizobium meliloti* pSymB carries genes necessary for arabinose transport and catabolism. Microbiology. **153**:727-36.

Radnedge, L., M.A. David, B. Youngren, S.J. Austin. (1997) Plasmid maintenance functions of the large virulence plasmid of *Shigella flexneri*. J. Bacteriol. **179**:3670 – 3675.

Reznikoff, W.S., K.M. Winterberg. (2008) Transposon-based strategies for the identification of essential bacterial genes. Methods Mol. Biol. **416**:13-26.

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

Ruiz-Echevarría, M.J., G. de la Cueva, R. Díaz-Orejas. (1995) Translational coupling and limited degradation of a polycistronic messenger modulate differential gene expression in the *parD* stability system of plasmid R1.Mol. Gen. Genet. **248**:599-609.

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

Schmeisser C., H. Liesegang, D. Krysciak, N. Bakkou, A. Le Quéré, A. Wollherr, I. HeinemeyerI, B. Morgenstern, A. Pommerening-Röser, M. Flores, R. Palacios, S. Brenner, G. Gottschalk, R.A. Schmitz, W.J. Broughton, X. Perret, A. W. Strittmatter, W.R. Streit. (2009) *Rhizobium* sp. strain NGR234 possesses a remarkable number of secretion systems. Appl. Environ. Microbiol. **75:**4035-45.

Schwartz E., A. Henne, R. Cramm, T. Eitinger, B. Friedrich, G. Gottschalk. (2003) Complete nucleotide sequence of pHG1: a *Ralstonia eutropha H16* megaplasmid encoding key enzymes of H_2-based ithoautotrophy and anaerobiosis. J. Mol. Biol. **332**:369 – 383.

Sergueev, K., A. Dabrazhynetskaya, S. Austin. (2005) Plasmid partition system of the P1 *par* family from the pWR1000 virulence plasmid of *Shigella flexneri*. J. Bacteriol. **187**:3369 – 3373.

Shintani, M., H. Yano, H. Habe, T. Omori, H. Yamane, M. Tsuda, H. Nojiri. (2006) Characterization of the replication, maintenance, and transfer features of the IncP-7 plasmid pCAR1, which carries genes involved in carbazole and dioxin degradation. Appl. Environ. Microbiol. **72**:3206-16.

Sota, M, H. Yano, A. Ono, R. Miyazaki, H. Ishii, H. Genka, E.M. Top, M. Tsuda. (2006) Genomic and functional analysis of the IncP-9 naphthalene-catabolic plasmid NAH7 and

its transposon Tn *4655* suggests catabolic gene spread by a tyrosine recombinase. J. Bacteriol. **188**: 4057–4067.

Suzuki, K., Y. Hattori, M. Uraji, N. Ohta, K. Iwata, K. Murata, A. Kato, K. Yoshida. (2000) Complete nucleotide sequence of a plant tumor-inducing Ti plasmid. Gene. **242**:331-6.

Suzuki, N., H. Nonaka, Y. Tsuge, S. Okayama, M. Inui, H. Yukawa. (2005) Multiple large segment deletion method for *Corynebacterium glutamicum*. Appl. Microbiol. Biotechnol. **69**: 151–161.

Theodosiou, N. A., T. Xu. **(**1998) Use of FLP/FRT system to study *Drosophila* development. Methods **14:**355-65.

Top, E.M., Y. Moenne-Loccoz, T. Pembroke, C.M. Thomas. (2000) Phenotypic traits conferred by plasmids . In: Thomas CM (ed) The horizontal gene pool: bacterial plasmid and gene spread. Harwood, New York , pp 249 – 285

Vedler, E., M. Vahter, A. Heinaru. (2004) The completely sequenced plasmid pEST4011 contains a novel IncP1 bacKbone and a catabolic transposon harboring *tfd* genes for 2,4-dichlorophenoxyacetic acid degradation J. Bacteriol. **186**:7161-74.

Voegele, R.T., S. Bardin, T.M. Finan. (1997) Characterization of the *Rhizobium* (*Sinorhizobium*) *meliloti* high- and low-affinity phosphate uptake systems. J. Bacteriol. **179**:7226-32.

Wei, J., M.B. Goldberg, V. Burland, M.M. Venkatesan, W. Deng, G. Fournier, G.F. Mayhew, G. III Plunkett, D.J. Rose, A. Darling, B. Mau, N.T. Perna, S.M. Payne, L.J. Runyen-Janecky, S. Zhou, D.C. Schwartz, F.R. Blattner. (2003) Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. Infect Immun **71**: 2775 – 2786

Westers, H., R. Dorenbos, J.M. van Dijl. (2003) Genome engineering reveals large dispensable regions in *Bacillus subtilis*. Mol. Biol. Evo. **20**:2076–2090.

Wood D.W., J.C. Setubal, R. Kaul, D.E. Monks, J.P. Kitajima, V.K. Okura, Y. Zhou, L. Chen, G.E. Wood, N.F. Jr Almeida, L. Woo, Y. Chen, I.T. Paulsen, J.A. Eisen, P.D. Karp, D. Sr Bovee, P. Chapman, J. Clendenning, G. Deatherage, W. Gillet, C. Grant, T. Kutyavin, R. Levy, M.J. Li, E. McClelland, A. Palmieri, C. Raymond, G. Rouse, C. Saenphimmachak, Z. Wu, P. Romero, D. Gordon, S. Zhang, H. Yoo, Y. Tao, P. Biddle, M. Jung, W. Krespan, M. Perry, B. Gordon-Kamm, L. Liao, S. Kim, C. Hendrick, Z.Y. Zhao, M. Dolan, F. Chumley, S.V. Tingey, J.F. Tomb, M.P. Gordon, M.V. Olson, E.W.

Nester. (2001) The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58 Science. **294**:2317-23.

Young, J.P., L.C. Crossman, A.W. Johnston, N.R. Thomson, Z.F. Ghazoui, K.H. Hull, M. Wexler, A.R. Curson, J.D. Todd, P.S. Poole. (2006) The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. Genome Biol **7:** R34

Yu, B.J., B.H. Sung, M.D. Koob, C.H. Lee, J.H. Lee, W.S. Lee, M.S. Kim and S.C. Kim. (2002) Minimization of the *Escherichia coli* genome using a Tn5-targeted *Cre*/*loxP* excision system. Nat. Biotechnol. **20**: 1018–1023.

Yuan, Z.C., R. Zaheer, T.M. Finan. (2006) Regulation and properties of PstSCAB, a high-affinity, high-velocity phosphate transport system of *Sinorhizobium meliloti* J. Bacteriol. **188**:1089-102

Zhang, Y., J. Zhang, K.P. Hoeflich, M. Ikura, G. Qing, M. Inouye. (2003) MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in *Escherichia coli*. Mol Cell. **12**:913-23.

Figure 1.1. Composite transposons as deletion strategy. The strategy for deletion formation can be used upon integration of the composite transposon into host's genome. Two deletions result from this transposition event in vivo, the first leading to the removal of the internal part of the transposon (Km^r-Tnp), and the second resulting in the deletion of a portion of the chromosome. Upon transposon integration, the product of TnpEK/LP gene binds to the internal transposon ends (MEs), resulting in blunt-end cleavage and loss of the donor DNA. Tnp-EK/LP then facilitates intramolecular strand transfer into the chromosome. Not all events during strand transfer will result in deletions; some will lead to chromosome inversions. IEs-external transposon ends; MEs-internal transposon ends; Km^r-kanamycin resistance; Cm^r-chloramphenicol resistance; TnpEK/LP- transposase; Δ-region targeted in deletion. (This figure is adopted and modified from Goryshin *et al*., 2003).

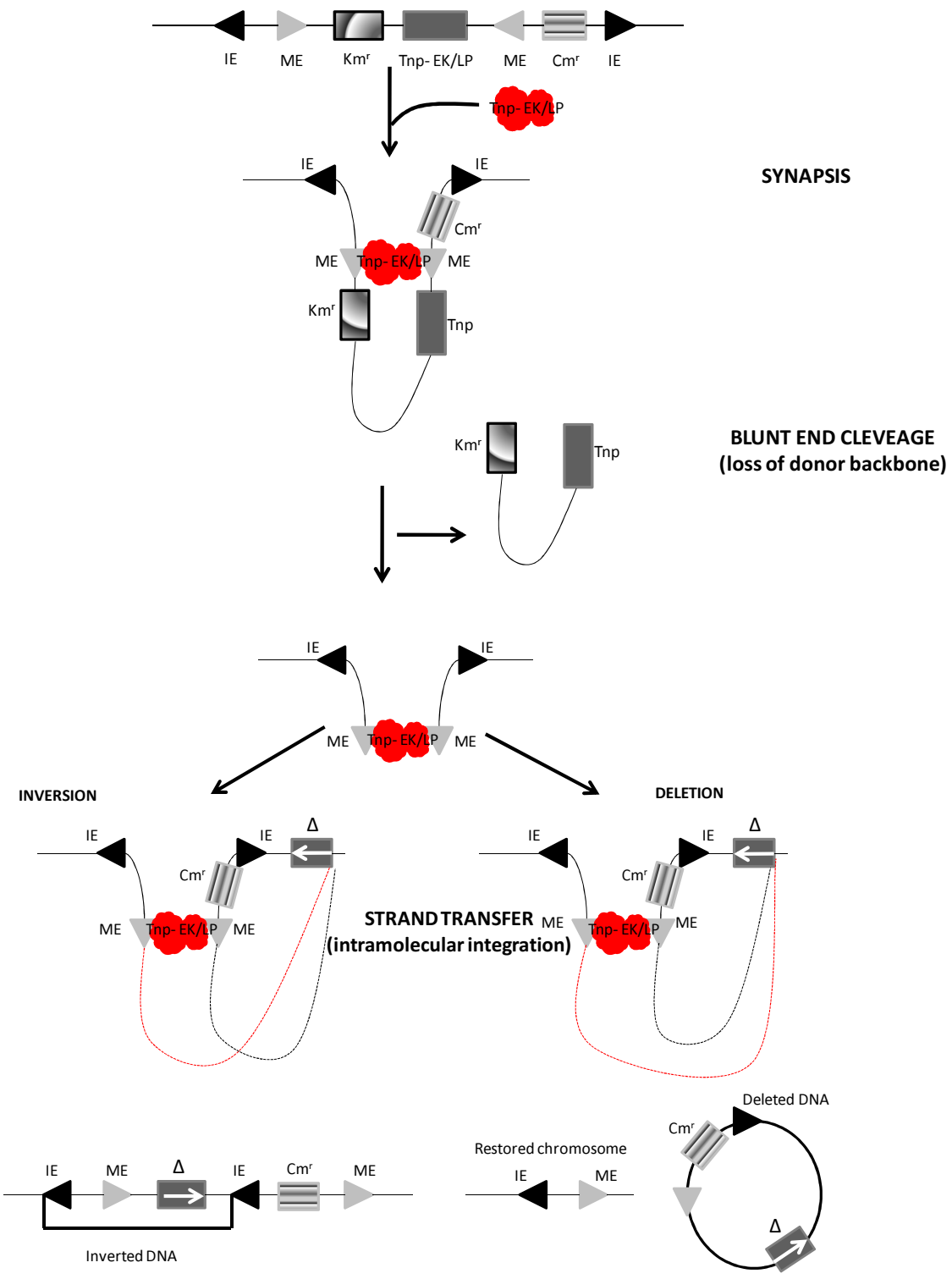

IE
ME
Km^r
Tnp- EK/LP
ME
Cm^r
IE
Tnp- EK/LP
SYNAPSIS
Cm^r
ME
Tnp- EK/LP
ME
Km^r
Tnp
BLUNT END CLEVEAGE
(loss of donor backbone)
Km^r
Tnp
ME
Tnp- EK/LP
ME
INVERSION
DELETION
Δ
Cm^r
STRAND TRANSFER
(intramolecular integration)
Deleted DNA
Restored chromosome
Inverted DNA

Figure 1.2. λ – Red type mediated recombination method. A, B and C are random DNA sequences that have 100% similarity. Polymerase chain reaction primers are labeled as primer1 and primer2. S indicates an I–*SceI* cleavage site. (This figure is adopted and modified from Kolisnychenko *et al*., 2002)

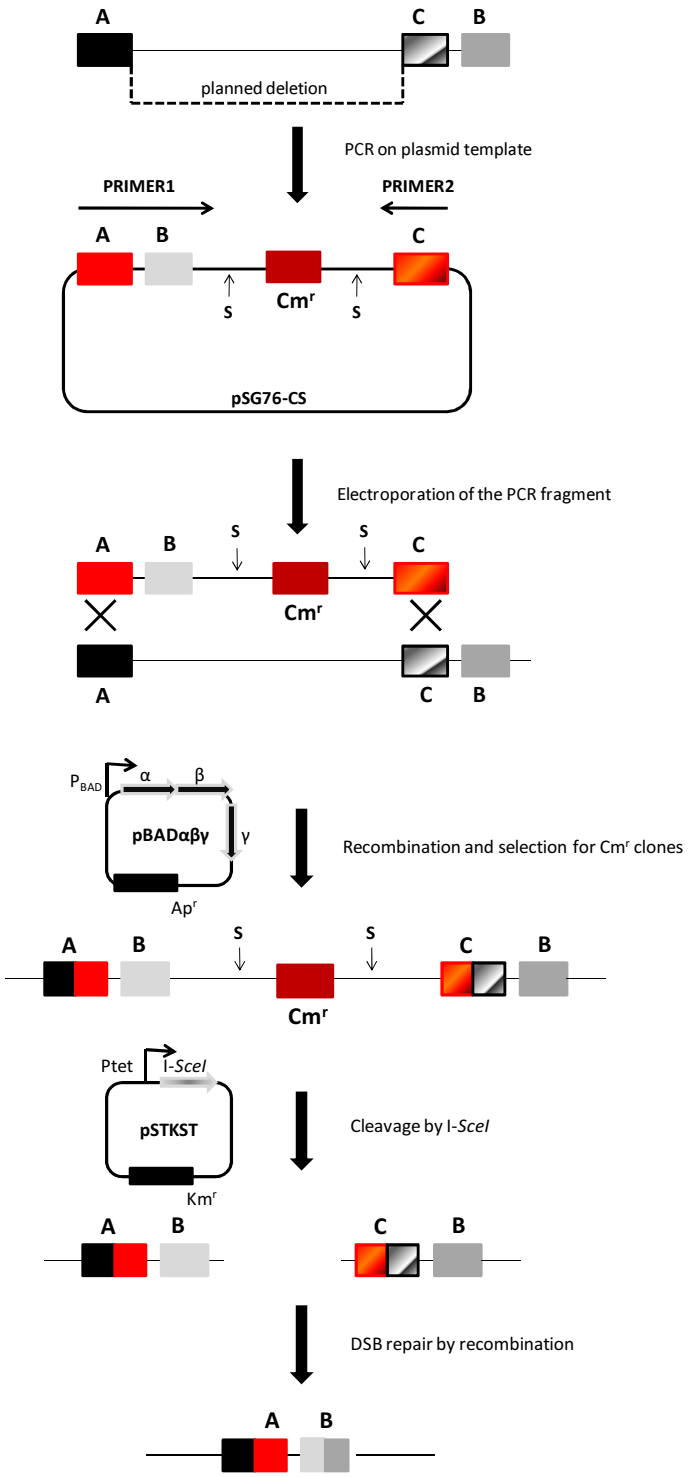
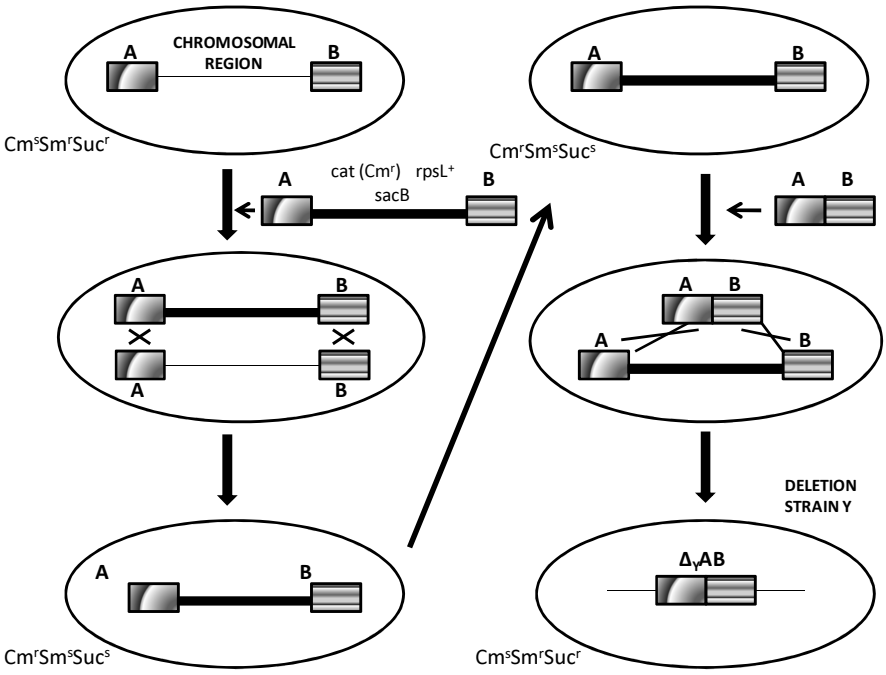

A
C
B
planned deletion
PCR on plasmid template
PRIMER1
PRIMER2
A
B
C
S
Cm^r
S
pSG76-CS
Electroporation of the PCR fragment
A
B
S
S
C
Cm^r
A
C
B
P BAD
α
β
pBADαβγ
γ
Ap^r
Recombination and selection for Cm^r clones
A
B
S
S
C
B
Cm^r
Ptet
I-SceI
pSTKST
Km^r
Cleavage by I-SceI
A
B
C
B
DSB repair by recombination
A
B

Figure 1.3. The construction of deletion units by the CRS cassette method. A) The CRS cassette method consists of two recombination steps. First, λ red recombinase mediates the deletion of the target regions replacement by a PCR generated CRS cassette. The deletion mutant is chloramphenicol resistant. Secondly, the inserted CRS cassette is removed and replaced by a PCR generated DNA fragment constructed by connecting adjacent regions. Markerless deletion mutants Δ_YAB were selected by both sucrose- and streptomycin-resistant phenotype. B) Two generated strains are used to combine the deletion unit, Δ_YAB and Δ_XCD. The deletion unit Δ_YAB is transferred by transduction with the P1 phages prepared from the strain with CRS cassette using the positive selectable markers. Next, the cassette is removed by transduction with P1 phage prepared from the strain without the CRS cassette using negative – selection markers. A, B, C and D are random DNA sequences that have 100% similarity. *cat* - cloramphenicol resistant gene; *rpsL* - 30S ribosomal subunit protein S12; *sacB* - *Bacillus subtilis* levansucrase. Cm^r – chloramphenicol resistant; Cm^s – chloramphenicol sensitive; Sm^r – streptomycin resistant; Sm^s – streptomycin sensitive; Suc^r – sucrose resistant; Suc^s – sucrose sensitive. The solid bar represents the CRS cassette. (This figure is adopted and modified from Hashimoto *et al*., 2005).

A)


A
CHROMOSOMAL REGION
B
Cm^{s}Sm^{r}Suc^r
Cm^{r}Sm^{s}Suc^s
cat (Cm^r) rpsL$^+$ sacB
DELETION STRAIN Y
Δ_YAB
Cm^{r}Sm^{s}Suc^s
Cm^{s}Sm^{r}Suc^r

B)

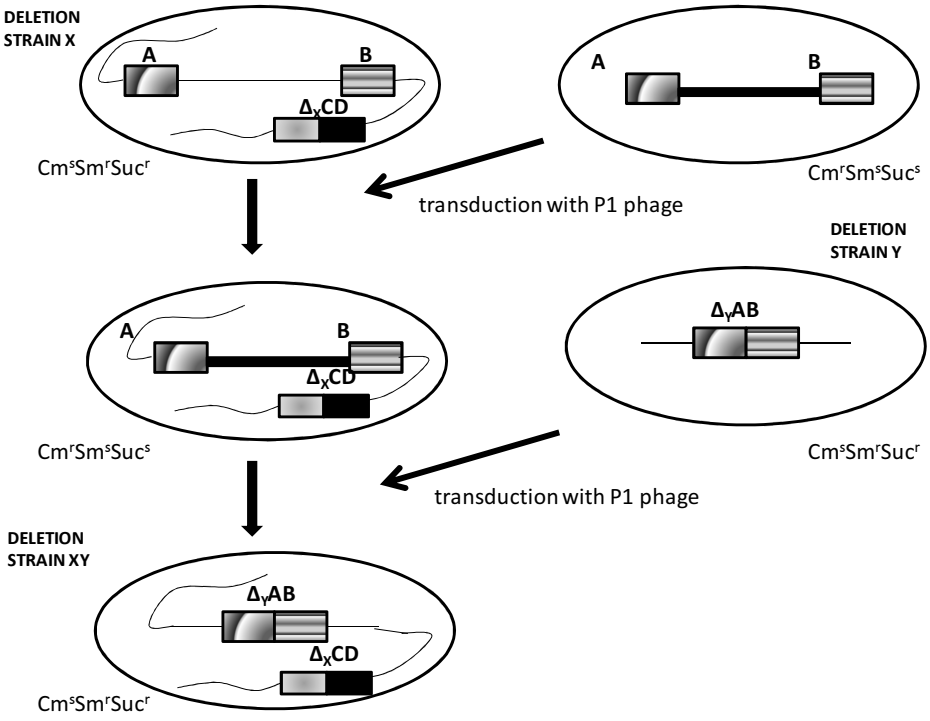

DELETION STRAIN X
A
B
Δ_XCD
Cm^{s}Sm^{r}Suc^r
Cm^{r}Sm^{s}Suc^s
transduction with P1 phage
DELETION STRAIN Y
Δ_YAB
Cm^{r}Sm^{s}Suc^s
Cm^{s}Sm^{r}Suc^r
transduction with P1 phage
DELETION STRAIN XY
Δ_YAB
Δ_XCD
Cm^{s}Sm^{r}Suc^r

Figure 1.4. The Cre/LoxP site-specific recombination system is used in genomic deletions. Two modified Tn-5 transposons are generated: TnKloxP and TnCloxP.Those transposons were used to generate two groups of mutant libraries, by random insertion into *E. coli* chromosome. Two mutant strains, one from each library, with a loxP site in the same orientation are selected and brought in parallel into a single strain by phage P1 transduction. The selected region between the two loxP sites is deleted upon expression of Cre recombinase. OE - outer-end transposase recognition sequence; Km^r-kanamycin resistance; Cm^r-chloramphenicol resistance; loxP- Cre recognition target site. A, B, C, D and E are random DNA sequences. (This figure is adopted and modified from Yu *et al*., 2002).

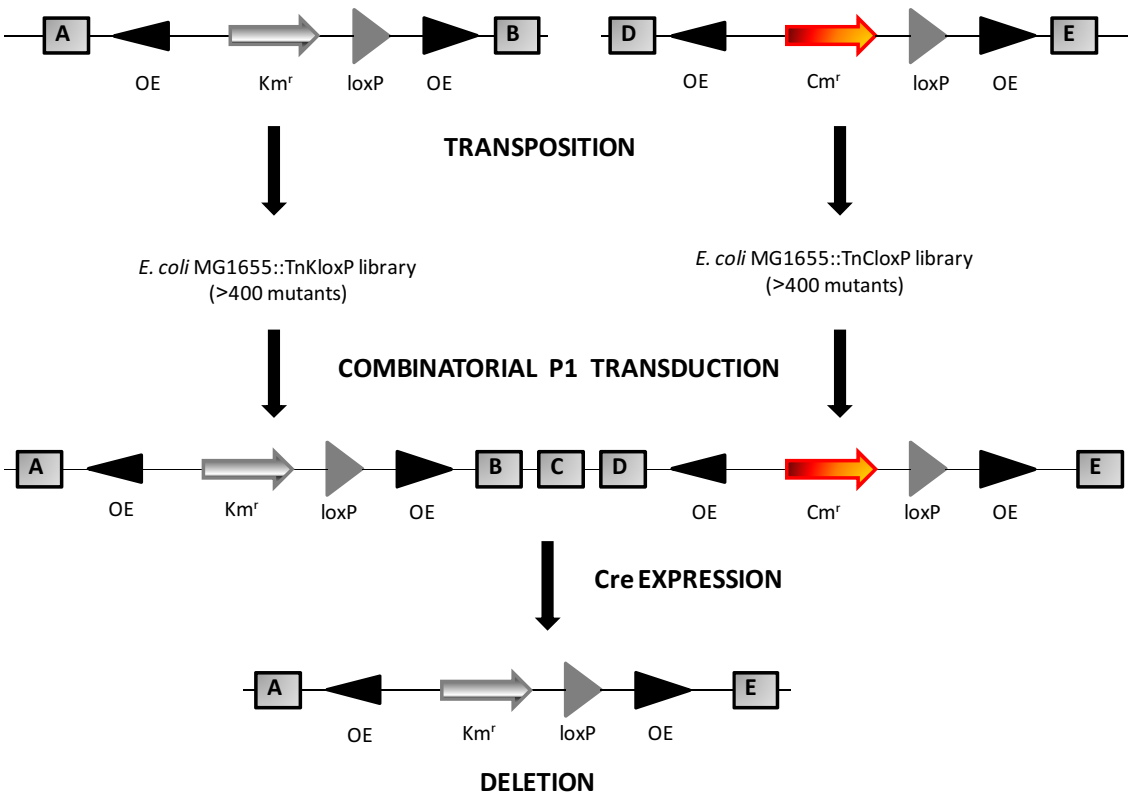
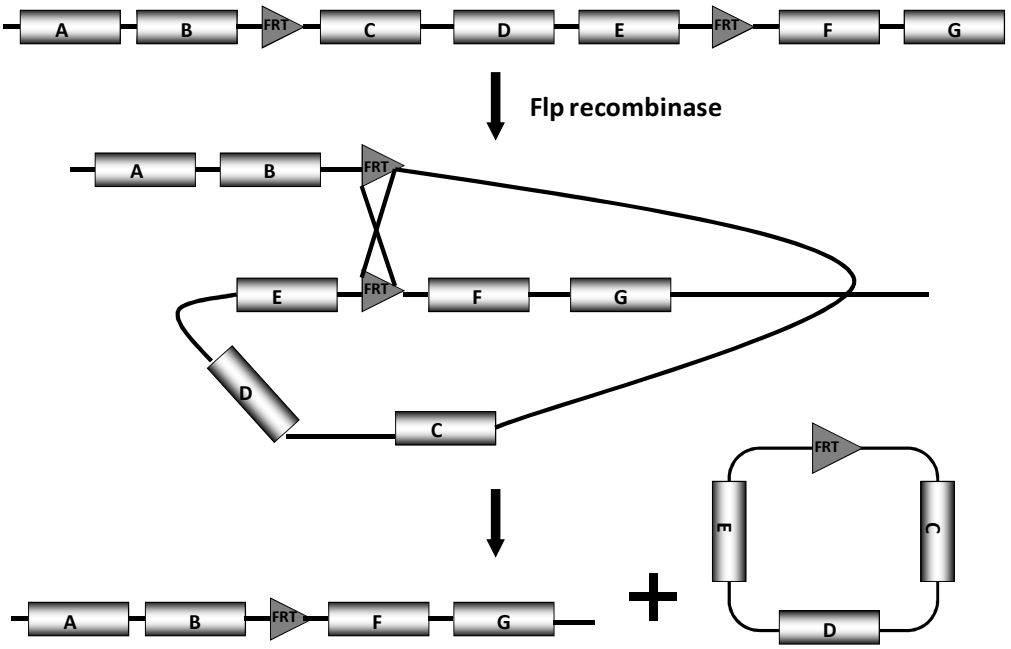

A
B
OE
Km^r
loxP
OE
D
E
OE
Cm^r
loxP
OE
TRANSPOSITION
E. coli MG1655::TnKloxP library
(>400 mutants)
E. coli MG1655::TnCloxP library
(>400 mutants)
COMBINATORIAL P1 TRANSDUCTION
A
B
C
D
E
OE
Km^r
loxP
OE
OE
Cm^r
loxP
OE
Cre EXPRESSION
A
E
OE
Km^r
loxP
OE
DELETION

Figure 1.5 The Flp/FRT site-specific recombination system is used in genomic deletions (a) Two Flp recognition target sites (FRT) are integrated in the genome in a direct orientation. The selected region between two FRT sites is deleted upon expression of Flp recombinase. The Flp/FRT site-specific recombination system can also be used in genomic inversions (b). Two Flp recognition target sites (FRT) are integrated in genome in the opposite orientation. The selected region between two FRT sites is inverted upon expression of Flp recombinase. A, B, C, D, E, F and G are random DNA sequences targeted for deletion.

a) GENOME DELETION - FRT sites are in a direct orientation.



b) GENOME INVERSION – FRT sites are in the opposite orientation.

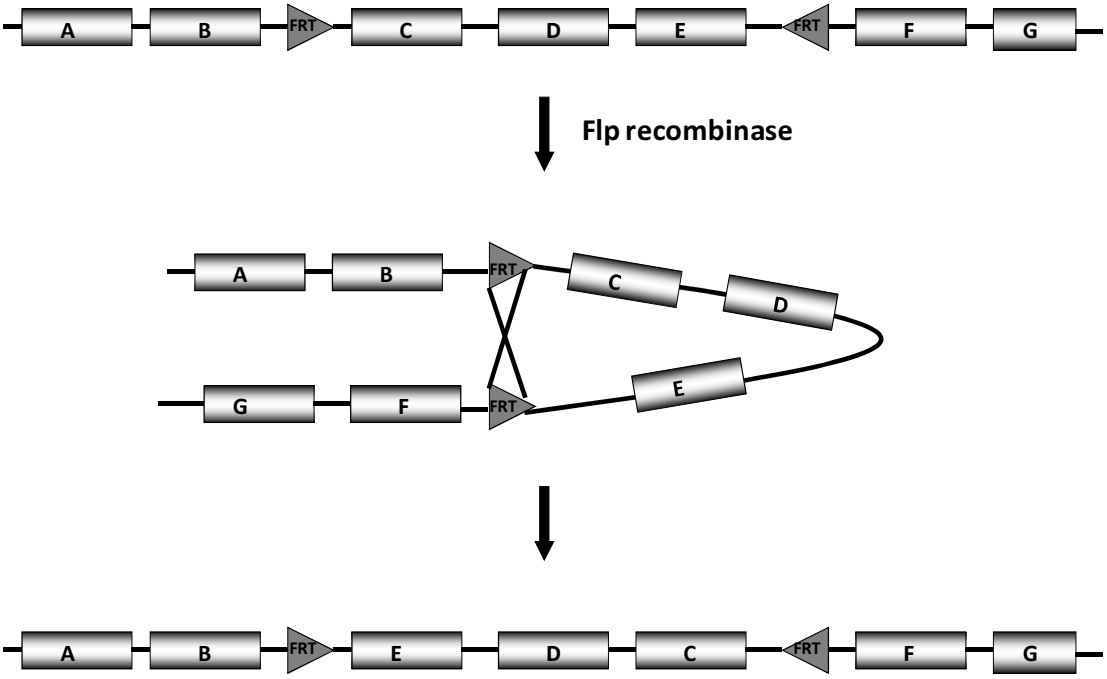


Table 1.1. Megaplasmids present in different organisms and genes associated with these plasmids.

Strain	Plasmid	Plasmid size	Genes present on the plasmid
Halobacterium sp NRC-1	pNRC100	191 Kb	Genes responsible for the synthesis of a unusual gas-filled vesicles, transcription and replication factors, cytochrome oxidase, thioredoxin and thioredoxin reductase
	pNRC200	365 Kb	Additional transcriptional factors, DNA polymerase and the only arginyl-tRNA synthetase
Pseudomonas putida NCIB9869	pRA500	500 Kb	Genes involved in 3,5-Xylenol, *p* -cresol
Variovorax paradoxus TV1	pTV1	200 Kb	Genes involved in degradation of 2,4-dichlorophenoxyacetic acid
Pseudomonas putida G7	NAH7	82 Kb	Genes responsible for naphthalene catabolism
Pseudomonas resinovorans CA10	pCAR1	199 Kb	Genes responsible for carbazole catabolism
Novosphingobium sp. KA1	pCAR3	254 Kb	Genes responsible for carbazole catabolism
Agrobacterium	pTiC58	214 Kb	T-DNA, the virulence gene (*vir*) region, the replication gene (*rep*) region, *tra* and *trb* regions, regions that direct uptake and catabolism of opine
	pRi2659	202 Kb	T-DNA, the virulence gene (*vir*) region, the replication gene (*rep*) region, *tra* and *trb* regions, regions that direct uptake and catabolism of opine
Shigella 2457T	pINV-2457T	218 Kb	Genes required for expression of type III secretion system and many of the effector proteins involved in the early stage of invasion
Yersiniae pestis	pYV	70 Kb	Plasmid pYV encoding a type III secretion system and proteins known as the *Yersinia* outer membrane proteins
	pMT1	101 Kb	Plasmid carries five important genes that have been associated with virulence
	pPCP1	9.6 Kb	Genes involved in biosynthesis of: bacteriocin called pesticin, a protein that provides immunity to pesticin, a plasminogen activator, and a coagulase
Yersiniae pseudotuberculosis	pYV	70 Kb	Plasmid pYV encoding a type III secretion system and proteins known as the *Yersinia* outer membrane proteins

Strain	**Plasmid**	**Plasmid size**	**Genes present on the plasmid**
Yersiniae enterocolitica	pYV	70 Kb	Plasmid pYV encoding a type III secretion system and proteins known as the *Yersinia* outer membrane proteins
Cupriavidus metallidurans CH34	pMOL28	171 Kb	Genes responsible for resistance to mercury cobalt and nickel
	pMOL30	233 Kb	Genes responsible for resistance to cadmium, zinc, cobalt and lead resistance
Ralstonia eutropha H16	pHG1	452 Kb	Genes specific for H_2–oxidizing system and for CO_2 fixation via the Calvin–Benson–Bassham cycle; clusters of genes for denitrification and for degradation of aromatic compounds
Oligotropha carboxidovorans OM5	pHCG3	133 Kb	Genes involve in CO oxidation, CO dehydrogenase as well as Calvin cycle enzymes and a dimeric hydrogenase
Arthrobacter nicotinovorans	pAO1	165 Kb	Genes specific for l-nicotine, d-nicotine and γ-*N*–methylaminobutyrate catabolism

Table 1.2. Summary of results obtained in genome reduction strategies.

Strain	Deletion size	Method for making deletions	Reference
E. coli (Δ20-4)	218.7 Kb (5.6%)	Composite transposons	Goryshin *et al*., 2003
E. coli (CDΔ3456)	313.1 Kb (6.7%)	Cre/LoxP excision system	Yu *et al*., 2002
E. coli MDS12	376.1 Kb (8.1%)	λ – Red type mediated recombination method	Kolisnychenko *et al*., 2002
E. coli MDS43	708.3 Kb (15.3%)	λ – Red type mediated recombination method	Posfai *et al*., 2006
E. coli MGF-01	1.03 Mb (22%)		Mizoguchi *et al*., 2007
E. coli Δ16	1.38 Mb (29.7%)	CRS cassette method	Hashimoto *et al*., 2005
B. subtilis Δ6	320 Kb (7.7%)	Plasmid-based chromosomal integration-excision systems	Westers *et al*., 2003
B. subtilis MG1M	991 Kb (21%)		Ara *et al*., 2007
B. subtilis MGB874	873.5 Kb (20.7%)	The *upp* (encoding uracil-phosphoribosyl-transferase) cassette and 5-fluorouracil (5-FU) selection were applied to remove the drug-resistant markers used to introduce primary deletions	Morimoto *et al*., 2008
C. glutamicum	190 Kb (5.7%)	Cre/loxP system in combination with double-strand break (DSB)-stimulated intramolecular recombination	Suzuki *et al*., 2005
M. genitalium	(20.7%)	Global transposon mutagenesis	Glass *et al*., 2006
S. meliloti	1,230 Kb (18.4%)	Homologous recombination between the insertion elements of flanking transposon insertions	Charles and Finan, 1991
S. meliloti	1,354 Kb (20.2%)	Repetitive rounds of Tn5B12-S mutagenesis with selection for deletion	Oresnik *et al*., 2000

Table 1.3. The eight typical toxin antitoxin (TA) families based on protein homology (Gerdes *et al*., 2005)

TA family	Toxin	Target of toxin	Antitoxin	Protease	Phyletic distribution
ccd	CcdB	Replication through DNA gyrase	CcdA	Lon	Gram negative bacteria
relBE	RelE	Translation through mRNA cleavage	RelB	Lon	Gram negative and gram positive bacteria, Archaea
parDE	ParE	Replication through DNA gyrase	ParD	Unknown	Gram negative and gram positive bacteria
higBA	HigB	Unknown	HigA	Unknown	Gram negative and gram positive bacteria
mazEF	MazF	Translation through mRNA cleavage	MazE	ClpXP/Lon	Gram negative and gram positive bacteria
phd/doc	Doc	Translation	Phd	ClpAP	Gram negative and gram positive bacteria, Archaea
vapBC	VapC	Unknown	VapB	Unknown	Gram negative and gram positive bacteria, Archaea
ωεζ	ζ	Unknown	ε	Unknown	Gram positive bacteria

CHAPTER 2. MATERIAL AND METHODS

Bacterial Strains and Plasmids

Bacterial strains and plasmids used in this study are listed in Table 2.1 and 2.2.

Medium, Antibiotics and Growth Conditions

LB (Luria broth) (per litre: 10 g Difco tryptone, 5 g Difco yeast extract and 5 g NaCl) was used as a complex medium for *E. coli* and supplemented with 2.5 mM $MgSO_4$ and 2.5 mM $CaCl_2$ for *S. meliloti*. M9-medium (per litre: 5.8 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl, 1 g NH_4Cl, 1 μg/ml biotin, 10 ng/ml $CoCl_2$, 1 mM $MgSO_4$ and 0.25 mM $CaCl_2$, 10 mM carbon source) with or without NH_4Cl was used as a minimal medium for all strains. 1/2LB1/2M9 medium was used for transductions consists of equal parts liquid M9 (1x), without additional $MgSO_4$ and $CaCl_2$, and LB containing 3% agar (LB-2x agar). Soft agar used for phage ΦM12 titration consisted of 0.5% agar in LBmc. All carbon and nitrogen sources were filter sterilized and added at the final concentration of 10 mM and 5 mM, respectively.

The final concentrations of antibiotics in LB agar for *S. meliloti* and *E. coli* are presented in Table 2.3. X-gal (5 - Bromo – 4 – chloro – 3 – indolyl – β – D - galactopyranoside) and IPTG (Isopropyl – ß – D - thiogalactopyranosid) were used at the final concentrations of 40 μg/ml and 0.5 mM, respectively. Thiamine auxotrophs were

supplemented with 0.01 mM thiamine. LB medium was supplemented with 2 ng/ml of $CoCl_2$ when needed. *S. meliloti* strains were grown at 30°C while *E. coli* strains were grown at 37°C.

Bacterial Matings

Conjugations were performed from liquid culture by combining equal volumes (20 µl) of previously washed (0.85% NaCl) mid to late-log phase cultures of donor, recipient and mobilizing strains. This mix was spotted directly on LB agar and incubated overnight at 30°C. The mating spots were then suspended in 0.85% NaCl and plated on appropriate selective media.

Preparation of ΦM12 Transducing Lysates

Lysates were prepared from log phase LBmc *S. meliloti* cultures. 0.05 ml of a ΦM12 phage stock propagated on RmP110, with a titre of ca. 10^{10}-10^{11} PFU/ml, was added to 5 ml of *S. meliloti* culture. This was then incubated with aeration for at least 8 h or until lysis could be detected by clearing of the culture. A few drops of chloroform were added to kill any viable cells that remained. Lysates were then stored at 4°C in 13 x 100 mm screwcap test-tubes. Lysate titration was done by mixing 0.1 ml of an appropriate phage lysate dilution (in LBmc) to 0.1 ml of an overnight culture of RmP110 (in LBmc). After 15 min at 30°C (to allow adsorption of the phage particles to the bacterial cell

surface) 2.5 ml of LBmc 0.5% agar, cooled to 50°C, was added. The mixture was immediately poured on the surface of an LBmc agar plate and incubated overnight at 30°C. The titre, in PFU (plaque forming units), was determined from the number of plaques arising in the bacterial lawn after overnight incubation.

ΦM12 General Transduction

Transduction was done by combining equal volumes (0.5 ml) of mid- to late-log phase *S. meliloti* recipient culture in LBmc with an appropriately diluted donor lysate. The donor lysate was diluted in LBmc to give a final MOI (multiplicity of infection) of 0.5. This usually worked out to dilutions in the order of 1/20 to 1/25. The mixture was kept at 30°C for 20 min to allow adsorption of phage particles to the bacterial cell surface. Then 1 ml of 0.85% NaCl was added and the cells were pelleted (16,100 x g, 3 min) in a tabletop centrifuge and the pellet washed with 0.85% NaCl to remove unadsorbed phage particles. The pellet was then suspended in 0.85% saline and plated on appropriate selective medium.

Isolation of Complementing Cosmid Clones

A pLAFR1 cosmid clone bank carrying Rm1021 DNA (Friedman *et al*., 1982) was used to isolate complementing clones. Approximately 1 ml from a frozen permanent

clone bank was added to 10 ml LB with Tc (5 μg/ml) and grown to OD_{600} 1 – 1.2 before mixing with recipient and mobilizer strains. The mating spot was set up using LB medium supplemented with $CoCl_2$ (2 ng/ml). Next day mating spots were plated on a medium selective for complementation ($LBGm^{60}Tc^{5}$). Seven *S. meliloti* colonies from LB selective plate were patched onto LB agar containing Tc to confirm the presence of the cosmid. Spot matings with six of these Cobalt+ colonies were performed with *E. coli* recipient strain DH5α to facilitate cosmid DNA extraction. Cosmid DNA was isolated from the six independent *E. coli* strains and DNA sequencing reactions were performed using primers complementary to the cosmid.

Preparation of Genomic DNA

S. meliloti genomic DNA was prepared from 4 ml of culture grown to saturation in LBmc. All 4 ml of bacteria were collected in 2-ml microcentrifuge tubes by centrifugation, washed once with 0.85% NaCl, and then resuspended in 750 μl 10 mM Tris, 25 mM EDTA (pH 8). Sodium dodecyl sulfate (SDS) was added to 1%, NaCl to 1 M, and proteinase K (Sigma) to 0.5 mg/ml. Samples were mixed gently and incubated at 65°C for 2 h. DNA was extracted once with buffer-saturated phenol, twice with 1:1 phenol-chloroform, and once with chloroform, and then ammonium acetate was added to 0.5 M and nucleic acids precipitated with an equal volume of isopropanol. The pellet was dissolved in 400 μl of 10 mM Tris, 1 mM EDTA with 20 μg/ml RNase A and incubated for 30 min at 37°C. DNA was extracted once with 1:1 phenol-chloroform and once with

chloroform, ammonium acetate was added to 0.5 M, and DNA was precipitated with an equal volume of isopropanol. The optical density at 260 nm (OD260) of the whole sample was measured to determine the DNA concentration. The samples were then lyophilized and dissolved to a final concentration of 2 µg/ml for sequencing.

Confirmation of Deletion Structure by PCR

To investigate whether the FRT flanked regions were deleted and lost from the cell, we used PCR to detect whether a given DNA region was present or absent. PCR reactions were generally run in 25 µl volumes. One bacterial colony or an appropriate amount of template DNA was added to the tube containing the appropriate concentrations of buffers and dNTPs to give a final concentration of 0.2 mM each of dATP, dCTP, dGTP, and dTTP. Each primer was added to a final concentration of 2 pmole/µl. All the primer pairs used in this work are listed in Table 2.4. The reactions were mixed either in ddH_2O or 10% glycerol/5% DMSO. DNA polymerase was added in the initial mix. For most PCR reactions Platinum Taq (Invitrogen) was used.

Plant Growth Conditions

Alfalfa (*Medicago sativa*) seeds were surface sterilized using 95% ethanol (5 min) and 2.5% sodium hypochloride (20 min). Seeds were rinsed with sterile water for 1 hour. Sterilized seeds were evenly spread on water agar plates (1x agar) and allowed to

germinate for two days in the dark. In a meanwhile Leonard assemblies were prepared. Leonard assembly consists of two plastic pots that are connected by cotton wick which extends from top pot into bottom. The top pot was filled with nitrogen free sand/vermiculite (1:1 w.w) mix and 250 ml of 1x Jensen's medium (per litre: 1 g $CaHPO_4$ (7.35 mM), 0.2g K_2HPO_4 (1.15 mM), 0.2 g $MgSO_4$ x $7H_2O$ (0.811 mM), 0.2 g NaCl (3.42 mM), 0.1 g $FeCl_2$ (0.616 mM), 1 ml of 1000 x trace mineral solution). Trace minerals solution was made to a 1000x stock and consisted of (per litre): 1 g H_3BO_3 (16.2 mM), 1 g $ZnSO_4$ $7H_2O$ (3.48 mM), 0.5 g $CuSO_4$ $5H_2O$ (2 mM), 0.5 g $MnCl_2$ $4H_2O$ (2.53 mM), 1 g Na_2MoO_4 $2H_2O$ (4.13 mM), 10 g Na_2EDTA (26.9 mM), 2 g NaFeEDTA (5.45 mM), 0.4 g biotin (1.64 mM). The pH of the Jansen's medium was adjusted to 7 before it was added. The bottom pot was filled with ddH_2O and Jansen's medium. Leonard assemblies were covered with plastic lid and autoclaved. After two days eight germinated seeds were transferred to each pot and placed in a Conviron growth chamber (18h day - 21°C and 6h night - 17°C) with the plastic lids remaining in place. Two days later seeds were inoculated with *S. meliloti* culture (10^7 – 10^8 cells). Plants were watered as required using autoclaved, distilled water and nodules were harvested for expression assays 4 weeks after inoculation and stored at –80°C.

β-Glucuronidase Activity in Root Nodules

β-Glucuronidase activity was quantified in root nodules as follows: First, 10 nodules were placed into a prechilled Eppendorf tube on ice. Then, 750 µl of cold MMS

buffer (40 mM MOPS, 20 mM KOH, 2 mM MgSO4, 0.3 M sucrose; pH 7.0) was added and nodules were crushed with mini pestles. Plant tissue was pelleted via centrifugation at 400 × g for 2 min (6,600 rpm) (Eppendorf Centrifuge 5415D) and 500 µl of supernatant was transferred to a fresh tube on ice. Sodium dodecyl sulfate was added to a final concentration of 0.01% and samples were incubated on ice for 5 min. Lysate was used at 100 µl per assay, and each reaction included 890 µl of buffer (50 mM sodium phosphate, 50 mM DTT, 1 mM EDTA; pH 7.0). Samples were equilibrated in water bath at 37°C for 10 min and the assay was initiated by the addition of 10 µl of 4-nitrophenyl β-D-glucuronide (35 mg/ml) (Sigma-Aldrich). Reactions were terminated by the addition of 200 µl of reaction mixture to 700 µl of 0.46 M Na_2CO_3 and absorbance at 405 nm was read using TCAN Safire Plate Reader. Specific activity was calculated as (absorbance at 405 nm × 1,000) (time [minutes] × milligrams of protein)$^{-1}$. The protein concentration of each lysate was determined by the method established by Bradford (1976) with bovin serum albumin as the protein standard.

Nodulation Kinetics and Isolation of Bacteria from Root Nodules

Nodulation kinetics were examined by performing nodulation assays in test tubes. Assays were done in 18 x 150 mm test tubes containing slants of Jansen's medium solidified with 1% agar. (Jansen's medium per litre: 1 g $CaHPO_4$ (7.35 mM), 0.2 g K_2HPO_4 (1.15 mM), 0.2 g $MgSO_4$ x $7H_2O$ (0.811 mM), 0.2 g NaCl (3.42 mM), 0.1 g $FeCl_2$ (0.616 mM), 1 ml of 1000 x trace mineral solution). Trace minerals solution was

made to a 1000x stock and consisted of (per litre): 1 g H_3BO_3 (16.2 mM), 1 g $ZnSO_4$ $7H_2O$ (3.48 mM), 0.5 g $CuSO_4$ $5H_2O$ (2 mM), 0.5 g $MnCl_2$ $4H_2O$ (2.53 mM), 1 g Na_2MoO_4 $2H_2O$ (4.13 mM), 10 g Na_2EDTA (26.9 mM), 2 g NaFeEDTA (5.45 mM), 0.4 g biotin (1.64 mM). The pH of the Jansen's medium was adjusted to 7.0 before it was added. The tops of the tubes were closed with transluscent polypropylene Kim – Kaps. A single two day old seedling was transfered to each slant using sterile technique. Three days later seedlings were inoculated with *S. meliloti* cultures (10^6 cells). The root systems of these plants were examined every second day and scored for nodules. Data from 25 plants were scored for each strain.

To isolate bacteria from root nodules, nodules removed from the root were rinsed with water and surface sterilized with 1% sodium hypochloride (15 min), washed twice with LB and squashed in 200 µl of LBmc containing 0.3 M sucrose. To determine the number of bacteria per nodule, dilutions were made and plated on LB. To isolate bacteria from nodule the resulting suspension of 0.3 M sucrose and root nodule was streaked on LB plates.

Shoot Dry Weight Determination

Plants were harvested at 28 days post inoculation and cut at the root – shoot junction. All plant tops from each pot were put together in a paper bag (6 – 8 plant tops per bag) and dried at 70°C for at least 1 week. Dry weights were measured using a Mettler PE 600 balance.

DNA Manipulations

Plasmid DNA was extracted from *E. coli* using QIAprep plasmid kits (QIAGEN, Chatsworth, CA). Agarose gel electrophoresis and other recombinant DNA techniques such as restriction analysis, DNA ligation and transformation were done according to Sambrook *et al.,* 1989. DNA fragments were recovered from agarose using QIAquick Gel Extraction Kit (QIAGEN, Chatsworth, CA). DNA sequencing was performed by the MOBIX lab facility at McMaster University, Hamilton, Ontario. DNA and derived protein sequences were analyzed with the BLAST (Altschul *et al.*, 1990) software package accessed at NCBI site.

Preparation of Transformation Competent *E. coli* Cells

A frozen stock of *E. coli* DH5*α* was streaked onto an LB plate and grown overnight at 37°C. The following day, a single colony was inoculated into 3 ml of LB and grown overnight at 37°C. Next morning 100 ml of LB was inoculated to an OD_{600}nm 0.05 and the culture was shaken at 37°C at 275 rpm. The cells were harvested at an absorbance at 600 nm of 0.4 and incubated on ice for 10 min. The cells were pelleted and resuspended in 25 ml ice cold 100 mM $CaCl_2$. The cell suspension was incubated on ice overnight, pelleted, and resuspended in 5 ml of ice cold 100 mM $CaCl_2$ with 15% glycerol. The cell suspension was gently mixed and incubated on ice for 5 min. 200 µl

aliquots were distributed into 1.5 ml microcentrifuge tubes, the tubes were flash frozen in liquid nitrogen and stored at -80°C.

Growth Curves and Viable Cell Counts

Each strain was grown overnight in LBmc medium and the next day cultures were diluted to an OD_{600} of 0.03 to 0.04 and ~ 5 ml pipetted into test tubes. Half of these cultures were grown in the presence of IPTG (0.5 mM final concentration) and growth was measured by monitoring the OD_{600} of the culture every three hours. Cultures were grown and aerated on a roller-drum (New Brunswick Scientific Co. Edison, N.J. USA, Rollordrum^{TM}) in an incubator at 30°C. The number of the colony forming units (CFU) was determined by plating samples on LB medium with or without IPTG. The number of colony forming units was plotted as a function of the time and the results were converted to the fraction of surviving cells by dividing by CFU at t=0.

Table 2.1. Bacterial strains used in this study.

Deletion strain	Derived from:	Phenotype
RmP794 (ΔB102::ΩZA)	RmP727 (pTH1522 (**752548**–753564); pTH1937 (788035–**788984**))	Sm^rTc^r
RmP795 (ΔB103::ΩNG)	RmP728 (pTH1522 (829221-**830108**); pTH1937 (**788035**–788984))	$Sm^rNm^rGm^rTc^r$
RmP793 (ΔB104::ΩZA)	RmP726 (pTH1522 (**829221**–830108); pTH1937 (869642–**870505**))	Sm^rTc^r
RmP792 (ΔB105::ΩNG)	RmP725 (pTH1522 (906915-9**08529**); pTH1937 (**869642**–870505))	$Sm^rNm^rGm^rTc^r$
RmP808 (ΔB106::ΩNG)	RmP807 (pTH1522 (1091104–**1092289**); pTH1937 (**869642**–870505))	$Sm^rNm^rGm^rTc^r$
RmP791 (ΔB107::ΩZA)	RmP724 (pTH1522 (**1091104**–1092289); pTH1937 (1129758–**1131168**))	Sm^rTc^r
RmP790 (ΔB108::ΩNG)	RmP723 (pTH1522 (1169073–**1170466**); pTH1937 (**1129758**–1131168))	$Sm^rNm^rGm^rTc^r$
RmP799 (ΔB109::ΩZA)	RmP733 (pTH1522 (**1169073**–1170466); pTH1937 (1204770–**1207052**))	Sm^rTc^r
RmP734 (B110)	(pTH1522 (1255032–**1256503**); pTH1937 (**1204770**–1207052)	$Sm^rNm^rGm^r$
RmP821 (ΔB111::ΩNG)	RmP820 (pTH1522 (1223871–**1225824**); pTH1937 (**1204770**–1207052))	$Sm^rNm^rGm^rTc^r$
RmP823 (ΔB112::ΩNG)	RmP822 (pTH1522 (1224621–**1226491**); pTH1937 (**1204770**–1207052))	$Sm^rNm^rGm^rTc^r$
RmP824 (B113)	(pTH1522 (1233633–**1235118**); pTH1937 (**1204770**–1207052))	$Sm^rNm^rGm^r$
RmP825 (B114)	(pTH1522 (1241758–**1243525**); pTH1937 (**1204770**–1207052))	$Sm^rNm^rGm^r$

Deletion strain	Derived from:	Phenotype
RmP826 (B115)	(pTH1522 (1243616–**1245558**); pTH1937 (**1204770**–1207052))	$Sm^rNm^rGm^r$
RmP801 (ΔB116::ΩZA)	RmP800 (pTH1522 (**1255032**–1256503); pTH1937 (1307752–**1308912**))	Sm^rTc^r
RmP809 (B117)	(pTH1522 (**1307905**–1308912); pTH1937 (1528150–**1529711**))	$Sm^rNm^rGm^r$
RmP811 (ΔB118::ΩZA)	RmP810 (pTH1522 (**1322226**–1323078); pTH1937 (1528150–**1529711**))	Sm^rTc^r
RmP813 (ΔB119::ΩZA)	RmP812 (pTH1522 (**1332882**–1333550); pTH1937 (1528150–**1529711**))	Sm^rTc^r
RmP815 (ΔB120::ΩZA)	RmP814 (pTH1522 (**1374412**–1375594); pTH1937 (1528150–**1529711**))	Sm^rTc^r
RmP817 (ΔB121::ΩZA)	RmP816 (pTH1522 (**1408135**–1408806); pTH1937 (1528150–**1529711**))	Sm^rTc^r
RmP798 (ΔB122::ΩNG)	RmP732 (pTH1522 (1572422–**1573735**); pTH1937 (**1528150**–1529711))	$Sm^rNm^rGm^rTc^r$
RmP806 (ΔB123::ΩNG)	RmP805 (pTH1522 (1652558–**1654191**); pTH1937 (**1528150**–1529711))	$Sm^rNm^rGm^rTc^r$
RmP803 (ΔB124::ΩZA)	RmP802 (pTH1522 (**1652558**–1654191); pTH1937 (1677882–**1679723**))	Sm^rTc^r
RmP804 (B125)	(pTH1522 (59089–**60148**); pTH1937 (**1677882**–1679723))	$Sm^rNm^rGm^r$
RmP853 (ΔB128::ΩZA)	RmP852 (pTH1522 (**859882**–861539); pTH1937 (869642–**870505**))	Sm^rTc^r
RmP855 (ΔB129::ΩZA)	RmP854 (pTH1522 (**852698**–854051); pTH1937 (869642–**870505**))	Sm^rTc^r
RmP857 (ΔB130::ΩZA)	RmP856 (pTH1522 (**839238**–841005); pTH1937 (869642–**870505**))	Sm^rTc^r

Deletion strain	Derived from:	Phenotype
RmP859 (ΔB131::ΩZA)	RmP858 (pTH1522 (**829255**–830910); pTH1937 (869642–**870505**))	Sm^rTc^r
RmP861 (ΔB132::ΩZA)	RmP860 (pTH1522 (**814652**–816483); pTH1937 (869642–**870505**))	Sm^rTc^r
RmP863 (ΔB133::ΩZA)	RmP862 (pTH1522 (**794204**–795633); pTH1937 (869642–**870505**))	Sm^rTc^r
RmP865 (ΔB134::ΩZA)	RmP864 (pTH1522 (**779439**–780709); pTH1937 (869642–**870505**))	Sm^rTc^r
RmP867 (ΔB135::ΩZA)	RmP866 (pTH1522 (**769971**–771070); pTH1937 (869642–**870505**))	Sm^rTc^r
RmP869 (ΔB136::ΩZA)	RmP868 (pTH1522 (**762942**–764540); pTH1937 (869642–**870505**))	Sm^rTc^r
RmP871 (ΔB137::ΩZA)	RmP870 (pTH1522 (**757101**–758231); pTH1937 (869642–**870505**))	Sm^rTc^r
RmP874 (ΔB139::ΩZA)	RmP873 (pTH1522 (**741497**–743314); pTH1937 (869642–**870505**))	Sm^rTc^r
RmP899 (B140)	(pTH1522 (**64141**–65620); pTH1937 (466499–**467160**))	$Sm^rNm^rGm^r$
RmP876 (ΔB141::ΩZA)	RmP875 (pTH1522 (**100636**–101396); pTH1937 (466499–**467160**))	Sm^rTc^r
RmP878 (ΔB142::ΩZA)	RmP877 (pTH1522 (**121311** – 122108); pTH1937 (466499–**467160**))	Sm^rTc^r
RmP1108 (ΔB143::ΩNG)	RmP1623 (pTH1522 (678812–**679659**); pTH1937 (**635019**–635940))	$Sm^rNm^rGm^rTc^r$
RmP1110 (ΔB144::ΩNG)	RmP1625 (pTH1522 (722036–**723622**); pTH1937 (**635019**–635940))	$Sm^rNm^rGm^rTc^r$
RmP1109 (ΔB145::ΩNG)	RmP1624 (pTH1522 (744320–**745096**); pTH1937 (**635019**–635940))	$Sm^rNm^rGm^rTc^r$

Deletion strain	Derived from:	Phenotype
RmP880 (ΔB146::ΩNG)	RmP879 (pTH1522 (752548–**755260**); pTH1937 (**635019**–635940))	$Sm^rNm^rGm^rTc^r$
RmP882 (ΔB147::ΩNG)	RmP881 (pTH1522 (757101–**758231**); pTH1937 (**635019**–635940))	$Sm^rNm^rGm^rTc^r$
RmP884 (ΔB148::ΩNG)	RmP883 (pTH1522 (762942–**764540**); pTH1937 (**635019**–635940))	$Sm^rNm^rGm^rTc^r$
RmP1681(ΔB151::ΩNG)	RmP902 (pTH1522 (64141–**65620**); pTH1937 (**61240**–62137))	$Sm^rNm^rGm^rTc^r$
RmP904(B152)	pTH1522 (73559–**74302**); pTH1937 (**61240**–62137))	$Sm^rNm^rGm^r$
RmP905(B153)	pTH1522 (88329–**89794**); pTH1937 (**61240**–62137)	$Sm^rNm^rGm^r$
RmP906(B154)	pTH1522 (100636–**101396**); pTH1937 (**61240**–62137)	$Sm^rNm^rGm^r$
RmP903(B155)	pTH1522 (121311–**122108**); pTH1937 (**61240**–62137)	$Sm^rNm^rGm^r$
RmP896(B158)	pTH1522 (1322226–**1323078**); pTH1937 (**1307752**–1308912)	$Sm^rNm^rGm^r$
RmP895(B159)	pTH1522 (1322370–**1323805**); pTH1937 (**1307752**–1308912)	$Sm^rNm^rGm^r$
RmP897(B160)	pTH1522 (1332882–**1333550**); pTH1937 (**1307752**–1308912)	$Sm^rNm^rGm^r$
RmP1055 (ΔB161::ΩZA)	RmP1054 (pTH1522 (**1677882**–1679723); pTH1937 (49523–**51610**))	Sm^rTc^r
RmP1053(B162)	pTH1522 (59089–**60148**); pTH1937 (**49523**–51610)	$Sm^rNm^rGm^r$
RmP926 (A101)	(pTH1522 (9549-**10988**); pTH1937 (**1283082**-1284751))	$Sm^rNm^rGm^r$

Deletion strain	Derived from:	Phenotype
RmP928 (ΔA102::ΩZA)	RmP927 (pTH1522 (**9549**-10988); pTH1937 (47717-**48842**))	Sm^rTc^r
RmP930 (ΔA103::ΩNG)	RmP929 (pTH1522 (90324-**92124**); pTH1937 (**47717**-48842))	$Sm^rNm^rGm^rTc^r$
RmP932 (ΔA104::ΩZA)	RmP931 (pTH1522 (**90324**-92124); pTH1937 (123788–**125128**))	Sm^rTc^r
RmP934 (ΔA105::ΩNG)	RmP933 (pTH1522 (184519-**186200**); pTH1937 (**123788**–125128))	$Sm^rNm^rGm^rTc^r$
RmP936 (ΔA106::ΩZA)	RmP935 (pTH1522 (**184519**-186200); pTH1937 (250917-**251809**))	Sm^rTc^r
RmP981 (ΔA107::ΩNG)	RmP980 (pTH1522 (294016-**295457**); pTH1937 (**250917**-251809))	$Sm^rNm^rGm^rTc^r$
RmP983 (ΔA108::ΩNG)	RmP982 (pTH1522 (306156-**308586**); pTH1937 (**250917**-251809))	$Sm^rNm^rGm^rTc^r$
RmP979 (ΔA109::ΩNG)	RmP978 (pTH1522 (311877-**313654**); pTH1937 (**250917**-251809)	$Sm^rNm^rGm^rTc^r$
RmP975 (ΔA111::ΩZA)	RmP874 (pTH1522 (**294016**-295457); pTH1937 (347789-**348455**))	Sm^rTc^r
RmP977 (ΔA112::ΩZA)	RmP976 (pTH1522 (**306156**-308586); pTH1937 (347789-**348455**))	Sm^rTc^r
RmP973 (ΔA113::ΩZA)	RmP972 (pTH1522 (**311877**-313654); pTH1937 (347789-**348455**))	Sm^rTc^r
RmP987 (ΔA114::ΩZA)	RmP986 (pTH1522 (**294016**-295457); pTH1937 (458916-**459668**))	Sm^rTc^r
RmP989 (ΔA115::ΩZA)	RmP988 (pTH1522 (**306156**-308586); pTH1937 (458916-**459668**))	Sm^rTc^r
RmP991 (ΔA116::ΩZA)	RmP990 (pTH1522 (**311877**-313654); pTH1937 (458916-**459668**))	Sm^rTc^r

Deletion strain	Derived from:	Phenotype
RmP939 (ΔA117::ΩZA)	RmP938 (pTH1522 (**400267**–402136); pTH1937 (458916-**459668**))	Sm^rTc^r
RmP941 (ΔA118::ΩNG)	RmP940 (pTH1522 (505335-**507338**); pTH1937 (**458916**-459668))	$Sm^rNm^rGm^rTc^r$
RmP943 (ΔA119::ΩZA)	RmP942 (pTH1522 (**505335**-507338); pTH1937 (575671-**577241**))	Sm^rTc^r
RmP945 (ΔA120::ΩNG)	RmP944 (pTH1522 (623673–**624863**); pTH1937 (**575671**-577241))	$Sm^rNm^rGm^rTc^r$
RmP947 (ΔA121::ΩZA)	RmP946 (pTH1522 (**623673**–624863); pTH1937 (677157-**678150**))	Sm^rTc^r
RmP949 (ΔA122::ΩNG)	RmP948 (pTH1522 (726673-**727921**); pTH1937 (**677157**-678150))	$Sm^rNm^rGm^rTc^r$
RmP951 (ΔA123::ΩZA)	RmP950 (pTH1522 (**726673**-727921); pTH1937 (774293-**775476**))	Sm^rTc^r
RmP953 (ΔA124::ΩNG)	RmP952 (pTH1522 (828417-**830143**); pTH1937 (**774293**-775476))	$Sm^rNm^rGm^rTc^r$
RmP955 (ΔA125::ΩZA)	RmP954 (pTH1522 (**828417**-830143); pTH1937 (879960-**881169**))	Sm^rTc^r
RmP971 (ΔA126::ΩNG)	RmP970 (pTH1522 (926857-**928282**); pTH1937 (**879960**-881169))	$Sm^rNm^rGm^rTc^r$
RmP969 (ΔA127::ΩNG)	RmP968 (pTH1522 (930000-**930910**); pTH1937 (**879960**-881169))	$Sm^rNm^rGm^rTc^r$
RmP957 (ΔA128::ΩZA)	RmP956 (pTH1522 (**924516**-927009); pTH1937 (1063642-**1064644**))	Sm^rTc^r
RmP959 (ΔA129::ΩNG)	RmP958 (pTH1522 (1122176-**1123504**); pTH1937 (**1063642**-1064644))	$Sm^rNm^rGm^rTc^r$
RmP961 (ΔA130::ΩZA)	RmP960 (pTH1522 (**1122176**-1123504); pTH1937 (1173115-**1173730**))	Sm^rTc^r

Deletion strain	Derived from:	Phenotype
RmP967 (ΔA131::ΩNG)	RmP966 (pTH1522 (1231998-**1232916**); pTH1937 (**1173115**-1173730))	$Sm^rNm^rGm^rTc^r$
RmP965 (ΔA132::ΩZA)	RmP964 (pTH1522 (**1231998**-1232916); pTH1937 (1283082-**1284751**))	Sm^rTc^r
RmP963 (ΔA133::ΩNG)	RmP962 (pTH1522 (1348238-**1349931**); pTH1937 (**1283082**-1284751))	$Sm^rNm^rGm^rTc^r$
RmP1051 (ΔA134)	(pTH1522 (90324-**92124**); pTH1937 (**1283082**-1284751))	$Sm^rNm^rGm^r$
RmP1052 (ΔA135)	(pTH1522 (**1348238**-1349931); pTH1937 (47717-**48842**))	$Sm^rNm^rGm^r$
RmP1032 (ΔB170::ΩNG)	RmP1031 (pTH1522 (1533912–**1535647**); pTH1937 (**1528150**-1529711))	$Sm^rNm^rGm^rTc^r$
RmP1034 (ΔB171::ΩNG)	RmP1033 (pTH1522 (1536036–**1536790**); pTH1937 (**1528150**–1529711))	$Sm^rNm^rGm^rTc^r$
RmP1036 (ΔB172::ΩNG)	RmP1035 (pTH1522 (1542407–**1544317**); pTH1937 (**1528150**–1529711))	$Sm^rNm^rGm^rTc^r$
RmP1038 (ΔB173::ΩNG)	RmP1037 (pTH1522 (1546542–**1547901**); pTH1937 (**1528150**–1529711))	$Sm^rNm^rGm^rTc^r$
RmP1040 (ΔB174::ΩNG)	RmP1039 (pTH1522 (1551630–**1552702**); pTH1937 (**1528150**–1529711))	$Sm^rNm^rGm^rTc^r$
RmP1042 (ΔB175::ΩNG)	RmP1041 (pTH1522 (1551786–**1553747**); pTH1937 (**1528150**–1529711))	$Sm^rNm^rGm^rTc^r$
RmP1046 (ΔB176::ΩNG)	RmP1045 (pTH1522 (1559257–**1553747**); pTH1937 (**1528150**–1529711))	$Sm^rNm^rGm^rTc^r$
RmP1048 (ΔB177::ΩNG)	RmP1047 (pTH1522 (1562886–**1564387**); pTH1937 (**1528150**–1529711))	$Sm^rNm^rGm^rTc^r$
RmP1050 (ΔB178::ΩNG)	RmP1049 (pTH1522 (1568798–**1569454**); pTH1937 (**1528150**–1529711))	$Sm^rNm^rGm^rTc^r$

Deletion strain	Derived from:	Phenotype
RmP1099 (ΔA150::ΩG)	RmP1094 (ΦRmP969 (879960-**930910)** → RmP936 (**184519**–251809))	$Sm^rGm^rTc^r$
RmP1092 (A151)	(ΦRmP959 (1063642-**1123504)**→ RmP936 (**184519**–251809))	$Sm^rGm^rTc^r$
RmP1101	RmP1095 (ΦRmP969 (879960-**930910)** → RmP947 (**623673**–678150))	$Sm^rGm^rTc^r$
RmP1408 (ΔA152::ΩG)	RmP1407 (ΦRmP959 (1063642-**1123504)** → RmP939 (**400267**–459668))	$Sm^rGm^rTc^r$
RmP1523 (A213)	(pTH1522 (306156-**308586**); pTH1937 (**184519**-186200))	$Sm^rNm^rGm^r$
RmP1524 (A214)	(pTH1522 (299677-**301554**); pTH1937 (**184519**-186200))	$Sm^rNm^rGm^r$
RmP1525 (A215)	(pTH1522 (293195-**294862**); pTH1937 (**184519**-186200))	$Sm^rNm^rGm^r$
RmP1526 (A216)	(pTH1522 (289418–**291042**); pTH1937 (**184519**-186200))	$Sm^rNm^rGm^r$
RmP1527 (A217)	(pTH1522 (280175–**281865**); pTH1937 (**184519**-186200))	$Sm^rNm^rGm^r$
RmP1528 (A218)	(pTH1522 (271658–**273493**); pTH1937 (**184519**-186200))	$Sm^rNm^rGm^r$
RmP1529 (A219)	(pTH1522 (265720–**267456**); pTH1937 (**184519**-186200))	$Sm^rNm^rGm^r$
RmP1643(ΔA220::ΩNG)	RmP1530 (pTH1522 (254828–**255655**); pTH1937 (**184519**-186200))	$Sm^rNm^rGm^rTc^r$
RmP1644 (ΔA221::ΩNG)	RmP1531 (pTH1522 (245715–**246848**); pTH1937 (**184519**-186200))	$Sm^rNm^rGm^rTc^r$
RmP1653 (ΔA222::ΩNG)	RmP1540 (pTH1522 (233938-**234774**); pTH1937 (**184519**-186200))	$Sm^rNm^rGm^rTc^r$

Deletion strain	Derived from:	Phenotype
RmP1662 (ΔA228::ΩNG)	RmP1549 (pTH1522 (191837-**193609**); pTH1937 (**184519**-186200))	$Sm^rNm^rGm^rTc^r$
RmP1663 (ΔA223::ΩNG)	RmP1551 (pTH1522 (230534-**231977**); pTH1937 (**184519**-186200))	$Sm^rNm^rGm^rTc^r$
RmP1665 (ΔA224::ΩNG)	RmP1553 (pTH1522 (220726–**222254**); pTH1937 (**184519**-186200))	$Sm^rNm^rGm^rTc^r$
RmP1667 (ΔA225::ΩNG)	RmP1555 (pTH1522 (215333-**216947**); pTH1937 (**184519**-186200))	$Sm^rNm^rGm^rTc^r$
RmP1668 (ΔA226::ΩNG)	RmP1556 (pTH1522 (200708 – **201691**); pTH1937 (**184519**-186200))	$Sm^rNm^rGm^rTc^r$
RmP1669 (ΔA227::ΩNG)	RmP1557 (pTH1522 (193611–**195077**); pTH1937 (**184519**-186200))	$Sm^rNm^rGm^rTc^r$
RmP1948 (A202)	(pTH1522 (**193611**–195077); pTH1937 (306156-**308586**))	$Sm^rNm^rGm^r$
RmP1949 (A205)	(pTH1522 (**230097**-231800); pTH1937 (306156-**308586**))	$Sm^rNm^rGm^r$
RmP1968 (A253)	(pTH1522 (299800-**300876**); pTH1937 (**254828**-255655))	$Sm^rNm^rGm^r$
RmP1969 (A257)	(pTH1522 (276987-**264696**); pTH1937 (**254828**-255655))	$Sm^rNm^rGm^r$
RmP1970 (A254)	(pTH1522 (297414-**298895**); pTH1937 (**254828**-255655))	$Sm^rNm^rGm^r$
RmP1986 (A204)	(pTH1522 (**220726**–222254); pTH1937 (306156-**308586**))	$Sm^rNm^rGm^r$
RmP1987 (A203)	(pTH1522 (**215333**-216947); pTH1937 (306156-**308586**))	$Sm^rNm^rGm^r$
RmP1988 (A201)	(pTH1522 (**191837**-193609); pTH1937 (306156-**308586**))	$Sm^rNm^rGm^r$

Deletion strain	Derived from:	Phenotype
RmP1989 (A252)	(pTH1522 (306156-**308586**); pTH1937 (**254828**-255655))	$Sm^rNm^rGm^r$
RmP1994 (A240)	(pTH1522 (306156-**308586**); pTH1937 (**230097**-231800))	$Sm^rNm^rGm^r$
RmP1995 (A244)	(pTH1522 (280764-**282406**); pTH1937 (**230097**-231800))	$Sm^rNm^rGm^r$
RmP1996 (A243)	(pTH1522 (289418–**291042**); pTH1937 (**230097**-231800))	$Sm^rNm^rGm^r$
RmP2001 (A241)	(pTH1522 (299800-**300876**); pTH1937 (**230097**-231800))	$Sm^rNm^rGm^r$
RmP2003 (A242)	(pTH1522 (297414-**298895**); pTH1937 (**230097**-231800))	$Sm^rNm^rGm^r$
RmP2004 (A245)	(pTH1522 (276987-**279393**); pTH1937 (**230097**-231800))	$Sm^rNm^rGm^r$
RmP2005 (A246)	(pTH1522 (262769-**264696**); pTH1937 (**230097**-231800))	$Sm^rNm^rGm^r$
RmP2007 (A232)	(pTH1522 (280764-**282406**); pTH1937 (**193611**-195077))	$Sm^rNm^rGm^r$
RmP2008 (A231)	(pTH1522 (289418–**291042**); pTH1937 (**193611**-195077))	$Sm^rNm^rGm^r$
RmP2013 (A229)	(pTH1522 (299800-**300876**); pTH1937 (**193611**-195077))	$Sm^rNm^rGm^r$
RmP2015 (A230)	(pTH1522 (297414-**298895**); pTH1937 (**193611**-195077))	$Sm^rNm^rGm^r$
RmP2016 (A233)	(pTH1522 (276987-**279393**); pTH1937 (**193611**-195077))	$Sm^rNm^rGm^r$
RmP2017 (A234)	(pTH1522 (262769-**264696**); pTH1937 (**193611**-195077))	$Sm^rNm^rGm^r$

Deletion strain	Derived from:	Phenotype
RmP2061 (ΔA250::ΩZA)	RmP1997 (pTH1522 (**193611**–195077); pTH1937 (230097-**231800**))	Sm^rTc^r
RmP2062 (ΔA239::ΩNG)	RmP2018 (pTH1522 (200708–**201691**); pTH1937 (**193611**-195077))	$Sm^rNm^rGm^rTc^r$
RmP2063 (ΔA251::ΩZA)	RmP1998 (pTH1522 (**191837**-193609); pTH1937 (230097-**231800**))	Sm^rTc^r
RmP2064 (ΔA248::ΩZA)	RmP1999 (pTH1522 (**220726**–222254); pTH1937 (230097-**231800**))	Sm^rTc^r
RmP2065 (ΔA249::ΩZA)	RmP2000 (pTH1522 (**215333**-216947); pTH1937 (230097-**231800**))	Sm^rTc^r
RmP2066 (A256)	(pTH1522 (280764-**282406**); pTH1937 (**254828**-255655))	$Sm^rNm^rGm^r$
RmP2067 (ΔA236::ΩNG)	RmP2009 (pTH1522 (230097-**231800**); pTH1937 (**193611**-195077))	$Sm^rNm^rGm^rTc^r$
RmP2068 (ΔA237::ΩNG)	RmP2011 (pTH1522 (220726–**222254**); pTH1937 (**193611**-195077))	$Sm^rNm^rGm^rTc^r$
RmP2069 (ΔA238::ΩNG)	RmP2012 (pTH1522 (215333-**216947**); pTH1937 (**193611**-195077))	$Sm^rNm^rGm^rTc^r$
RmP2070 (ΔA235::ΩNG)	RmP2014 (pTH1522 (254828–**255655**); pTH1937 (**193611**-195077))	$Sm^rNm^rGm^rTc^r$
RmP2071 (ΔA259::ΩZA)	RmP1991 (pTH1522 (**220726**–222254); pTH1937 (254828-**255655**))	Sm^rTc^r
RmP2072 (ΔA263::ΩZA)	RmP1990 (pTH1522 (**191837**-193609); pTH1937 (254828-**255655**))	Sm^rTc^r
RmP2073 (ΔA261::ΩZA)	RmP1993 (pTH1522 (**200708**–201691); pTH1937 (254828-**255655**))	Sm^rTc^r
RmP2074 (ΔA262::ΩZA)	RmP1966 (pTH1522 (**193611**–195077); pTH1937 (254828-**255655**))	Sm^rTc^r

Deletion strain	Derived from:	Phenotype
RmP2075 (ΔA247::ΩNG)	RmP2002 (pTH1522 (254828–**255655**); pTH1937 (**230097**-231800))	$Sm^rNm^rGm^rTc^r$
RmP2077 (ΔA211::ΩZA)	RmP1952 (pTH1522 (**297414**-298895); pTH1937 (306156-**308586**))	Sm^rTc^r
RmP2078 (ΔA212::ΩZA)	RmP1950 (pTH1522 (**299800**-300876); pTH1937 (306156-**308586**))	Sm^rTc^r
RmP2079 (ΔA209::ΩZA)	RmP1946 (pTH1522 (**280764**-282406); pTH1937 (306156-**308586**))	Sm^rTc^r
RmP2080 (ΔA210::ΩZA)	RmP1947 (pTH1522 (**289418**–291042); pTH1937 (306156-**308586**))	Sm^rTc^r
RmP2081 (ΔA255::ΩNG)	RmP1965 (pTH1522 (289418–**291042**); pTH1937 (**254828**-255655))	$Sm^rNm^rGm^rTc^r$
RmP2082 (ΔA206::ΩZA)	RmP1951 (pTH1522 (**254828**–255655); pTH1937 (306156-**308586**))	Sm^rTc^r
RmP2083 (ΔA258::ΩZA)	RmP1967 (pTH1522 (**230097**-231800); pTH1937 (254828-**255655**))	Sm^rTc^r
RmP2084 (ΔA207::ΩZA)	RmP1953 (pTH1522 (**262769**-264696); pTH1937 (306156-**308586**))	Sm^rTc^r
RmP1694	RmP896 (pTH1522 (1322226–**1323078**); pTH1937 (**1307752**–1308912)) (pTH2302)	$Sm^rNm^rGm^rTc^rSp^r$
RmP2211	RmP2208 (pTH1522 (1255032–**1256503**); pTH1937 (nt:**1204770**–1207052)) (pTH2027)	$Sm^rNm^rGm^rTc^rSp^r$

Deletion strain	Derived from:	Phenotype
RmP2212	RmP2209 (pTH1522 (1243616–**1245558**); pTH1937 (**1204770**–1207052)) (pTH2027)	$Sm^rNm^rGm^rTc^rSp^r$
RmP2213	RmP2210 (pTH1522 (1241758–**1243525**); pTH1937 (**1204770**–1207052)) (pTH2027)	$Sm^rNm^rGm^rTc^rSp^r$
RmP2233	(pTH1522 (1231998-**1232916**); pTH1937 (**1173115**–1173730)) (pTH1931)	$Sm^rNm^rGm^rSp^r$
RmP2237	(pTH1522 (**1231998**-1232916); pTH1937 (1283082-**1284751**)) (pTH1931)	$Sm^rNm^rGm^rSp^r$
RmP2249	RmP2200 (pTH1522 (**924516**-927009); pTH1937 (1063642-**1064644**)) (pTH2563)	$Sm^rNm^rGm^rTc^rSp^r$
RmP2251	RmP2143 (pTH1522 (9549-**10988**); pTH1937 (**1283082**-1284751)) (pTH2563)	$Sm^rNm^rGm^rTc^rSp^r$
RmP2252	RmP2141 (pTH1522 (299800-**300876**); pTH1937 (**254828**-255655)) (pTH2563)	$Sm^rNm^rGm^rTc^rSp^r$
RmP2253	RmP2195 (pTH1522 (299800-**300876**); pTH1937 (**254828**-255655)) (pTH1931)	$Sm^rNm^rGm^rTc^rSp^r$

Deletion strain	Derived from:	Phenotype
RmP2332	(pTH1522 (nt: **1231998**-1232916); pTH1937 (nt: 1283082-**1284751**)) (pTH2625)	$Sm^rNm^rGm^rSp^r$
RmP2335	(pTH1522 (**1231998**-1232916); pTH1937 (1283082-**1284751**)) (pTH2624)	$Sm^rNm^rGm^rSp^r$
RmP2336	(pTH1522 (1231998-**1232916**); pTH1937 (**1173115**–1173730)) (pTH2622)	$Sm^rNm^rGm^rSp^r$
RmP2352	(pTH1522 (1231998-**1232916**); pTH1937 (**1173115**–1173730)) (pTH2647)	$Sm^rNm^rGm^rSp^r$
RmP2354	RmP2331(pTH1522 (**1231998**-1232916); pTH1937 (1283082-**1284751**)) (pTH2623)	$Sm^rNm^rGm^rTc^rSp^r$
RmP2658	RmP2353 (pTH1522 (1231998-**1232916**); pTH1937 (**1173115**–1173730)) (pTH2646)	$Sm^rNm^rGm^rTc^rSp^r$
RmP2360	RmP2269 (pTH1522 (1322226–**1323078**); pTH1937 (**1307752**–1308912)) (pTH2563)	$Sm^rNm^rGm^rTc^rSp^r$
RmP2369	(pTH1522 (**1173115**–1173730); pTH1937(1231998-**1232916**))	$Sm^rNm^rGm^r$
RmP2377	(pTH1522 (**1173115**–1173730); pTH1937(1231998-**1232916**)) (pTH2623)	$Sm^rNm^rGm^rSp^r$
RmP2378	(pTH1522 (**1173115**–1173730); pTH1937(1231998-**1232916**)) (pTH2625)	$Sm^rNm^rGm^rSp^r$

Deletion strain	Derived from:	Phenotype
RmP2379	(pTH1522 (**1173115**–1173730); pTH1937(1231998-**1232916**)) (pTH2624)	$Sm^rNm^rGm^rSp^r$
RmP2382	(Φ (pTH1522 (73559–**74302**); pTH1937 (**61240**–62137))→RmP110) (pTH2651)	$Sm^rNm^rGm^rTc^r$
RmP2379	(pTH1522 (**1173115**–1173730); pTH1937(1231998-**1232916**)) (pTH2624)	$Sm^rNm^rGm^rSp^r$
RmP2382	(Φ (pTH1522 (73559–**74302**); pTH1937 (**61240**–62137))→RmP110) (pTH2651)	$Sm^rNm^rGm^rTc^r$
RmP2383	(Φ (pTH1522 (73559–**74302**); pTH1937 (**61240**–62137))→RmP110) (pTH2654)	$Sm^rNm^rGm^rTc^r$
RmP2383	(Φ (pTH1522 (73559–**74302**); pTH1937 (**61240**–62137))→RmP110) (pTH2654)	$Sm^rNm^rGm^rTc^r$
RmP2494	(pTH1522 (9549-**10988**); pTH1937 (**1283082**-1284751)) (pTH2646)	$Sm^rNm^rGm^rSp^r$
RmP2495	(pTH1522 (**924516**-927009); pTH1937 (1063642-**1064644**)) (pTH2646)	$Sm^rNm^rGm^rSp^r$
RmP2496	(pTH1522 (1322226–**1323078**); pTH1937 (**1307752**–1308912)) (pTH2646)	$Sm^rNm^rGm^rSp^r$
RmP2659	(pTH1522 (9549-**10988**); pTH1937 (**1283082**-1284751)) (pTH2623)	$Sm^rNm^rGm^rSp^r$

Deletion strain	Derived from:	Phenotype
RmP2660	(pTH1522 (**924516**-927009); pTH1937 (1063642-**1064644**)) (pTH2623)	$Sm^rNm^rGm^rSp^r$
RmP2661	(pTH1522 (1322226–**1323078**); pTH1937 (**1307752** – 1308912)) (pTH2623)	$Sm^rNm^rGm^rSp^r$
RmP2681 ΔA160:: ΩG	ΦΔA105(123788-**186200**)→ΔA102(**9549**-48842**)**	Sm^rGm^r

Abreviations: ^{r} designates resistance; Ap, ampicillin; Cm, chloramphenicol; Gm, gentamicin; Km, kanamycin; Nm, neomycin; Sm, streptomycin; Sp, spectinomycin; Tc, tetracycline. *lac*, lactose utilization genes; *oriT*, origin of transfer; Δ, deletion; ::ΩGN, deletion is marked by gentamicin (G) and neomycin (N) resistance; ::ΩZA, zero antibiotic (ZA) or no antibiotic marker is present within deletion. Bold numbers correspond to the start and stop nucleotide position for each deletion. Where pTH1944 *flp* plasmid transconjugant strains were recovered and the deletion subsequently verified, the region was designated with a Δ (e.g. ΔB118)

Table 2.2. Bacterial strains and plasmids used in this study.

S. meliloti strain	Relevant characteristic(s)	Reference
Rm1021	SU47 str-21 Sm^r (wild type)	Meade *et al*. (1982)
RmG212	*lac*$^-$ derivative of Rm1021	Lab collection
RmP110	Rm1021 with wild-type *pstC*	Yuan *et al.* (2006)
RmH940	RmG212 (*lac*$^-$) derivative with FRT sites flanking Tn5-235 (Nm^r/*lacZ*)	P.Chain, unpublished
RmP1064	ΔA106 without *flp* plasmid Sm^r	This study
RmP1065	ΔA117 without *flp* plasmid Sm^r	This study
RmP1067	ΔA102 without *flp* plasmid Sm^r	This study
RmP1222	RmP110 (pTH2172) Sm^rGm^r; disruption of *pssF*gene	This study
RmP1223	RmP110 (pTH2173) Sm^rGm^r; transcriptional fusion to *pssF* gene	This study
RmP1224	RmP110 (pTH2174) Sm^rGm^r; disruption of *uxuB* gene	This study
RmP1225	RmP110 (pTH2175) Sm^rGm^r; transcriptional fusion to *uxuB* gene	This study
RmP1226	RmP110 (pTH2176) Sm^rGm^r; disruption of *smb20750* gene	This study
RmP1227	RmP110 (pTH2177) Sm^rGm^r; transcriptional fusion to *smb20750* gene	This study

S. meliloti strain	Relevant characteristic(s)	Reference
RmP1228	RmP110 (pTH2178) Sm^rGm^r; disruption of *smb20751* gene	This study
RmP1229	RmP110 (pTH2179) Sm^rGm^r; transcriptional fusion to *smb20751* gene	This study
RmP1230	RmP110 (pTH2180) Sm^rGm^r; disruption of *smb20752* gene	This study
RmP1231	RmP110 (pTH2181) Sm^rGm^r; transcriptional fusion to *smb20752* gene	This study
RmP1232	RmP110 (pTH2182) Sm^rGm^r; disruption of *smb20753* gene	This study
RmP1233	RmP110 (pTH2183) Sm^rGm^r; transcriptional fusion to *smb20753* gene	This study
RmP1234	RmP110 (pTH2184) Sm^rGm^r; disruption of *smb20754* gene	This study
RmP1235	RmP110 (pTH2185) Sm^rGm^r; transcriptional fusion to *smb20754* gene	This study
RmP1810	ΦRmP1408 → RmP110 Sm^rGm^r	This study
RmP2371	RmP110 (pTH2623) Sm^rSp^r	This study
RmP2372	RmP110 (pTH2625) Sm^rSp^r	This study
RmP2380	RmP110 (pTH2624) Sm^rSp^r	This study
RmP2381	RmP110 (pTH1931) Sm^rSp^r	This study
RmP2374	RmP110 (pTH2647) Sm^rSp^r	This study

S. meliloti strain	Relevant characteristic(s)	Reference
RmP2373	RmP110 (pTH2646) Sm^rSp^r	This study
RmP2370	RmP110 (pTH2622) Sm^rSp^r	This study
RmP2451	RmP110 (pTH2563) Sm^rSp^r	This study
RmP2361	Φ (pTH1522 (nt:73559–**74302**); pTH1937(nt:**61240**–62137))→RmP110 Nm^rGm^r	This study
RmP2348	RmP110 (pTH2585) deletions for fix delay Sm^rGm^r	This study
RmP2350	RmP110 (pTH2586) deletions for fix delay Sm^rGm^r	This study
Escherichia coli	Relevant characteristic(s)	Reference
MT616	MT607 (pRK600), mobilizer; Cm^r	Finan *et al.* (1986)
DH5α	φ80d*lacZ* ΔM15	GIBCO BRL
Plasmids	Relevant characteristic(s)	Reference
pMS101	vector containing FRT site within multiple cloning site	Snaith *et al.* (1995)
pTH472	*flp* delivery vector derivative of pMP220; Gm^r	P. Chain, unpablished
pTH1360	pUC119 derivative with *oriT* from RP4 and *gusA* from pFus1; Ap^r/Nm^r	R. Zaheer, unpublished

Plasmids	Relevant characteristic(s)	Reference
pTH1522	reporter vector containing FRT site; Gm^r	Cowie *et al.* (2006)
pTH1703	pTH1522 derivative with unique *Xho*I, *Swa*I, *Kpn*I, *Not*I, *Sph*I, *Nsi*I, *Apa*I *Pac*I, and *Bgl*II sites in the multiple cloning site	Cowie *et al.* (2006)
pTH1919	pBBR MCS-3 derivative with RK2-*tetR-tetA;* Tc^r	J. Cheng, unpublished
pTH1931	expression vector pTrcSC derivative of pTrcStrep; Sm^rSp^r	C. Baron, unpublished
pTH1937	pTH2001 with *Sma*I/*Xho*I fragment from pTH1360 cont. *nptII* from Tn5; Km^r/Nm^r	This study
pTH1938	pTH1937 with *Spe*I/*Eco*RI fragment of pSymB (nt:1129758 – 1131168); Km^r/Nm^r	This study
pTH1939	pTH1937 with *Spe*I/*Eco*RI fragment of pSymB (nt:869642 – 870505); Km^r/Nm^r	This study
pTH1940	pTH1937 with *Spe*I/*Eco*RI fragment of pSymB (nt:788035 – 788984); Km^r/Nm^r	This study

Plasmids	Relevant characteristic(s)	Reference
pTH1942	pTH1937 with *Spe*I/*Eco*RI fragment of pSymB (nt:1528150 – 1529711); Km^r/Nm^r	This study
pTH1943	pTH1937 with *Spe*I/*Eco*RI fragment of pSymB (nt:1204770 – 1207052); Km^r/Nm^r	This study
pTH1944	pTH1919 with 2kb *Pst*I fragment containing *pcaD*, *flp* from pTH472; Tc^r	This study
pTH1992	pTH1937 with *Spe*I/*Eco*RI fragment of pSymB (nt:61240 – 62137); Km^r/Nm^r	This study
pTH1994	pTH1937 with *Spe*I/*Eco*RI fragment of pSymB (nt:466499 – 467160); Km^r/Nm^r	This study
pTH1995	pTH1937 with *Spe*I/*Eco*RI fragment of pSymB (nt:635019 – 635940); Km^r/Nm^r	This study
pTH1996	pTH1703 with 2.1kb *Nsi*I/*Bgl*II fragment including *engA*; Gm^r	This study
pTH1997	pTH1703 with740bp *Xho*I/*Bgl*II fragment internal to *engA*; Gm^r	This study
pTH1998	pACYC177 derivative with *Eco*RI/*Sph*I fragment containing MCS; Km^r/Nm^r	This study

Plasmids	Relevant characteristic(s)	Reference
pTH1999	pTH1998 with *Sph*I/*Bgl*II fragment from pMS101 containing FRT site; Km^r/Nm^r	This study
pTH2000	ΔTn903 inverted repeats in pTH1999; Km^r/Nm^r	This study
pTH2001	pTH2000 with *Nco*I/*Nhe*I fragment from pTH1360 containing the *oriT* site from pRK2; Km^r/Nm^r	This study
pTH2027	pTH1931 with 1431bp *engA* coding region as a *Pac*I fragment; Sm^rSp^r	This study
pTH2038	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:47717 - 48842); Km^r/Nm^r	This study
pTH2039	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:250917 - 251809); Km^r/Nm^r	This study
pTH2040	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:123788 – 125128); Km^r/Nm^r	This study
pTH2041	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:347789 - 348455); Km^r/Nm^r	This study

Plasmids	Relevant characteristic(s)	Reference
pTH2042	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:677157 - 678150); Km^r/Nm^r	This study
pTH2043	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:458916 - 459668); Km^r/Nm^r	This study
pTH2044	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:575671 - 577241); Km^r/Nm^r	This study
pTH2045	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:774293 - 775476); Km^r/Nm^r	This study
pTH2046	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:879960 - 881169); Km^r/Nm^r	This study
pTH2047	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:1063642 - 1064644); Km^r/Nm^r	This study
pTH2048	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:1173115 - 1173730); Km^r/Nm^r	This study

Plasmids	Relevant characteristic(s)	Reference
pTH2049	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:1283082 - 1284751); Km^r/Nm^r	This study
pTH2172	pTH1703 with *Nsi*I/*Blg*II fragment of *smb20748* (nt:1536749 – 1537465); Gm^r	This study
pTH2173	pTH1703 with *Nsi*I/*Xho*I fragment of *smb20748* (nt:1535568 – 1536675); Gm^r	This study
pTH2174	pTH1703 with *Nsi*I/*Blg*II fragment of *smb20749* (nt:1538446 – 1538995); Gm^r	This study
pTH2175	pTH1703 with *Nsi*I/*Blg*II fragment of *smb20749* (nt:1539037 – 1539978); Gm^r	This study
pTH2176	pTH1703 with *Nsi*I/*Blg*II fragment of *smb20750* (nt:1539676 – 1540181); Gm^r	This study
pTH2177	pTH1703 with *Nsi*I/*Xho*I fragment of *smb20750* (nt:1539676 – 1541255); Gm^r	This study

Plasmids	Relevant characteristic(s)	Reference
pTH2178	pTH1703 with *Nsi*I/*Xho*I fragment of *smb20751* (nt:1540506 – 1541058); Gm^r	This study
pTH2179	pTH1703 with *Nsi*I/*Xho*I fragment of *smb20751* (nt:1539555 – 1541058); Gm^r	This study
pTH2180	pTH1703 with *Nsi*I/*Blg*II fragment of *smb20752* (nt:1541626 – 1542348); Gm^r	This study
pTH2181	pTH1703 with *Blg*II/*Nsi*I fragment of *smb20752* (nt:1540831 – 1541532); Gm^r	This study
pTH2182	pTH1703 with *Blg*II/*Nsi*I fragment of *smb20753* (nt:1542450 – 1543168); Gm^r	This study
pTH2183	pTH1703 with *Blg*II/*Nsi*I fragment of *smb20753* (nt:1541770 – 1542524); Gm^r	This study
pTH2184	pTH1703 with *Blg*II/*Nsi*I fragment of *smb20754* (nt:1544334 – 1545004); Gm^r	This study

Plasmids	Relevant characteristic(s)	Reference
pTH2185	pTH1703 with *Blg*II/*Xho*I fragment of *smb20754* (nt:1543155 – 1544138); Gm^r	This study
pTH2288	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:184519 - 186200); Km^r/Nm^r	This study
pTH2289	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:311877 - 313654); Km^r/Nm^r	This study
pTH2290	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:951834 - 953338); Km^r/Nm^r	This study
pTH2302	pTH1931 with 460bp *Pac*I fragment of pSymB (nt: 1314806 – 1315265); Sm^rSp^r	This study
pTH2470	pTH1937 with *Spe*I/*Eco*RI fragment of *thuE* (nt: 328238-328905); Km^r/Nm^r	This study
pTH2519	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:306156 - 308586); Km^r/Nm^r	This study

Plasmids	Relevant characteristic(s)	Reference
pTH2520	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:299800 - 300876); Km^r/Nm^r	This study
pTH2521	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:297414 - 298895); Km^r/Nm^r	This study
pTH2522	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:289418 - 291042); Km^r/Nm^r	This study
pTH2523	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:280764 - 282406); Km^r/Nm^r	This study
pTH2524	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:262769 - 264696); Km^r/Nm^r	This study
pTH2525	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:254828 - 255655); Km^r/Nm^r	This study
pTH2526	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:245715 - 246848); Km^r/Nm^r	This study

Plasmids	Relevant characteristic(s)	Reference
pTH2527	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:230097 - 231800); Km^r/Nm^r	This study
pTH2528	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:200708 - 201691); Km^r/Nm^r	This study
pTH2529	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:193611 - 195077); Km^r/Nm^r	This study
pTH2563	pTH1931 with *Pac*I fragment of *sma0471/sma0473* coding region (nt:257565 - 258837); Sm^rSp^r	This study
pTH2622	pTH1931 with *Pac*I fragment of *sma2151* coding region (nt:1214806 - 1215645); Sm^rSp^r	This study
pTH2623	pTH1931 with *Pac*I fragment of *sma2231* coding region (nt:1248755 - 1249733); Sm^rSp^r	This study
pTH2624	pTH1931 with *Pac*I fragment of *sma2253/sma2255* coding region (nt:1262710 - 1264012); Sm^rSp^r	This study
pTH2625	pTH1931 with *Pac*I fragment of *sma2273/sma2275* coding region (nt:1269637 - 1268234); Sm^rSp^r	This study

Plasmids	Relevant characteristic(s)	Reference
pTH2646	pTH1931 with *Pac*I fragment of *sma2105* coding region (nt:1187240 - 1189099); Sm^rSp^r	This study
pTH2647	pTH1931 with *Pac*I fragment of *sma2133* coding region (nt:1205426 - 1206839); Sm^rSp^r	This study
pTH2648	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:1231998 - 1232916); Km^r/Nm^r	This study
pTH2651	pLAFR1 cosmid clone with *Eco*RI fragment of pSymB (nt:56613-81543) Tc^r	This study
pTH2654	pLAFR1 cosmid clone with *Eco*RI fragment of pSymB (nt:56613–74702) Tc^r	This study
pTH2691	pTH1937 with *Spe*I/*Eco*RI fragment of pSymA (nt:1122176 - 1123504); Km^r/Nm^r	This study

Abreviations: ^{r} designates resistance; Ap, ampicillin; Cm, chloramphenicol; Gm, gentamicin; Km, kanamycin; Nm, neomycin; Sm, streptomycin; Sp, spectinomycin; Tc, tetracycline. *lac*, lactose utilization genes; *oriT*, origin of transfer; Δ, deletion.

Table 2.3. Antibiotic concentrations and stocks

Antibiotic	Concentration for *E. coli* (μg/ml)	Concentration for *S. meliloti* (μg/ml)	Stock (mg/ml)
Ampicillin (Amp)	100	n.a.	100
Kanamycin sulphate (Km)	25	n.a.	20
Tetracycline hydrochloride (Tc)	5	5	5
Streptomycin sulphate (Sm)	n.a.	200	200
Chloramphenicol (Cm)	5	n.a.	20
Gentamycin sulphate (Gm)	10	60	20
Spectinomycin dihydrochloride (Sp)	100	100	100
Neomycin sulphate (Nm)	n.a.	200	100
Rifampin (Rif)	50	20	10

Abreviations: n.a. – not applicable

Table 2.4. Primers used in this study

Primer name	Primer sequence	Gene containing primer	Primer name	Primer sequence	Gene containing primer
1192-UF	CTTCGATCTCCTCGTCCACG	*nodP2*	3669-UR	CAGTTCCGCGTTGAGATCGC	*cyaK*
1192-UR	AAGCCCGTCGAGAAGCAAGG	*nodQ2*	3669-DF	GCCTTATGACGACCAGAGCG	TRm19
1192-DF	CATACCTACCTGCTCGACGG	*nodQ2*	3669-DR	AATCGAGAGGTTCACGCACCG	*smb21701*
1192-DR	TCGGAACTGTTGTGCGAGCG	*smb21225*	678-UF	TCCTGATTCTGGACGAGCCG	*smb21588*
2044-UF	CACTACCTGGATCTCGTGCG	*smb21109*	678-UR	GAATCAAGGCGCTTCCCTCC	*smb21588*
2044-UR	GTCCTTCATCGTGAGCACCG	*smb21111*	678-DF	CCTCATCCAGGGTCTCATCC	*smb21590*
2044-DF	TGTGGCTGAAGGTCAACGGC	*hpaG*	678-DR	TCTGGCGATTGACCCGAAGC	*smb21592*
2044-DR	TGTGATCGAAGCCTGCGTCG	*smb21113*	3394-UF	ATGGCTCTGCTCAACGTCGC	*smb21550*
3949-UF	TCGATTCGAGCAGCCTGTCG	*smb21256*	3394-UR	TCGTCGGGTCGAACACATCG	*smb21551*
3949-UR	CTATGTAGTCGGTCGCTCCC	*smb21257*	3394-DF	TAGCCGTGGTGGTCGAAAGC	*smb21554*
3949-DF	GACCACCGCTATCTCATCGC	*smb21259*	3394-DR	TGCGACAGAACCTCGATCCG	*smb21554 - kefB2*
3949-DR	GCACCGAGAATGTGATCCGC	*smb21260*	2611-UF	ACGACAACCTCTCCAACGGC	*exoB*
778-UF	TCGAAGCCGATGGCAATCGC	*smb21292*	2611-UR	TCCCAGCCTAGAACCTGACG	*exoB*
778-UR	CGTTCCTCCTCGGTAATGCC	*smb21292*	2611-DF	TGCTGCCAGAGATAGGTGCG	*exoQ*
778-DF	AGCTCAGGATGCAGGTTGCC	*gua01*	2611-DR	AGCCTTGTCAGACCTGGAGC	*exoQ*
778-DR	CTCAAGATCGACCTCGTCCG	*smb21295*	4440-UF	TCAGACGGTTTCCGACGACC	*smb20973*
2940-UF	AGAACGCACTTCGACGGAGG	*smb21161*	422-UF	AGTTGCTACCCAGCTCCAGG	*smb21678 - idhA*
2940-UR	TGATGCCGACGACCGATTGG	*smb21162*	4440-UR	TCCCTAAGCATCGCCTACGG	*smb20973*
2940-DF	GTGCAGGTGCTTGTTGTCCG	*hutU*	4440-DF	GACGTGTCCTTGAGGATCGC	*smb20975*
2940-DR	ACCTCAACGAGGACGAGACC	*hutU*	4440-DR	CGCACCAGCATCATCTTCGC	*smb20975*

Primer name	Primer sequence	Gene containing primer	Primer name	Primer sequence	Gene containing primer
422-UR	GCGTAGGCTTCCGTATAGCG	*idhA*	4068-UR	CAACGTCTGGATCTGCCTCG	*smb21445*
422-DF	TTCGGCAGAGCCTTTCGACG	*smb20900*	4068-DF	ATCCGAGCGTCTGATACCCC	*glgX2*
422-DR	CGTTGCCAAGCTGTCTTCGG	*smb21679 - smb20901*	4068-DR	AGACGACATCGAGCAGGACC	*glgB2*
2626-UF	GTCGAGTGGATCAAGGAGCG	*ildD3*	426-UF	CTCTCGCTTCCACTTCTGCG	*smb20988 - smb20989*
2626-UR	GCCGATGTTCTGACCAAGGC	*smb20851*	426-UR	AGAGGACGTGATGGTCGTCG	*nfeD*
2626-DF	GACGAAATCCAGAGCCTCGC	*smb20852*	426-DF	TGGGAGAGCAACTTCGCTGC	*smb20991*
2626-DR	GGCTCGTGGAACAGTTTCGG	*smb20853*	426-DR	ATGCCGAGGAACTGCTCAGG	*smb20992*
2965-UF	TCATCAGCTCCGTTCGCTCG	*agpA*	2659-UF	AGGGAAGGAGCAGGATCAGC	*smb20989*
2965-UR	GACGACATCATCGTGCTGCG	*agpA*	2659-UR	GACGACCTCTTGAAGCAGGC	*nfeD*
2965-DF	GTTGTGCTCGCTCAGATCCG	*agaL1*	2659-DF	AAGACGGCAATCTGCCAGCC	*smb20991*e
2965-DR	TCGGTCCTGATCTCGTAGCC	*agpT*	2659-DR	TGGTCGCTTCAGCCTTGTGC	*smb20992*
916-UF	ATAGAGCTCGTGGTGGTCGC	*smb20631*	2522-UF	AGGATCTCGTCCTTGCTGCG	*smb21284*
916-UR	TGCGACCGTGTCGATCTTCG	*smb20632*	2522-UR	GCCTCCACACCATCGTAACC	*smb21284*
916-DF	TTGACCTCTTCGAGCCTCGG	*smb20634*	2522-DF	GGGTCACGAAGTTCATGCGC	*xdhA1*
916-DR	GAAACTGATGGCAGCGACGG	*smb20634*	2522-DR	TTCCTGGCAACGAGCATCCG	*xdhA1*
3486-UF	TCGAGCAGCAGTTTGGCACC	*smb21682*	5620-UF	CAGATGTCCTGCTTGGTCGC	*nrtA*
3486-UR	CGCTGAAGTGGAACGAAGGC	*smb20914*	5620-UR	CATGAGGTTCACGGTGTCGC	*smb21115*
3486-DF	CGAGCATCTATCACCGTGGC	*aslA1*	5620-DF	CTGGGACTCGTGAACATCGG	*smb21117*
3486-DR	AAGCTCTCCGTCGTCGTTCC	*aslA1*	5620-DR	CACGCCTATCGTCGTGAACG	*smb21120*
4068-UF	CTCGTCACAGGTATCCAGCC	*smb21444 - smb21445*	203-UF	AGAAGGACACGGAACTGCGC	*smb21097*
203-UR	CCGATTCGGAAAGCAGGTCG	*smb21098 - smb21099*	4916-DR	AGACAGTCTGGAGGACGACG	*smb21013*
203-DR	ATCAGCCTGTTCGACACCGG	*smb21102*	608-UF	CGTAAAGGACAGGACCTGCG	*smb20364*
5653-UF	CGCTTCGACGAGTTCAAGGC	*mccA*	608-UR	ATCCAACGGCAGCGATCTGG	*smb20364*

Primer name	Primer sequence	Gene containing primer	Primer name	Primer sequence	Gene containing primer
5653-UR	TCGACTTCCATGAGCCAGGC	*smb21123*	608-DF	CGTCTCCCGTACCTTTACGC	*smb20366*
5653-DF	GCAGTGACAAAGGGACAGGC	*mccB*	608-DR	TAAGAGAGCACGGACACGGC	*smb20366 - smb20367*
5653-DR	CGTTCACGGCAAGCGTATCG	*hmgL*	4180-UF	CGCGTCAACTCCTGCTAACG	*smb20048*
3770-UF	CGAAGCTCACCATCGAAGCG	*smb20452*	4180-UR	TCACTCCGTTCTGGAGAGCG	*fusA2*
3770-UR	TCGACTGCCTGATCTGGAGC	*smb20453*	4180-DF	GAGCATTGACCGACAGCAGG	*fusA2*
3770-DF	GCCGTAGTCAACGAGAAGCG	*smb20454*	4180-DR	CTGGCGAAGTTCGACAAGCC	*cpo*
3770-DR	CAAGGCTTCGACAGTCTCGC	*smb20456*	19-UF	TCGACAGCCAGGTGCAAACC	*smb20062*
3468-UF	GGCGATCAGGAGACAAAGCG	*smb20091*	19-UR	GGCTGACACGTCATACTGGC	*smb20064*
3468-UR	GCAGCGATGAACAGTCTCGC	*smb20093*	19-DF	GGATACGATGTGCGTCCAGG	*smb20065 - smb20066*
3468-DF	GCTTTCGGTCGCTGAGTTCG	*smb20093*	19-DR	GGTTTGATTCCAGGCTCGCC	*smb20067 - smb20068*
3468-DR	GTGCATGACATCCAGGTGGC	*smb20094*	3312-UF	GGGTCTCGATGGAAGATCCG	*smb20050*
3953-UF	CCGCTCGATTCTTCCGAAGC	*smb20109*	3312-UR	ACCATCGCCATCTGGCAACC	*smb20055*
3953-UR	CCTTATGCGTGGATCGAGCC	*smb20109*	3312-DF	AGCGTACCGACAGGCATTCC	*smb20055 - smb20056*
3953-DF	GGAGGAGGTCTTCACTGTCC	*smb20111*	3312-DR	AGTCGCTCATCTGCCTTCCC	*smb20055 - smb20056*
3953-DR	TGGTGCGGATAGCGTTTGGC	*smb20111*	3347-UF	AGGTCGAGCGAGACATTGCC	*smb20722*
4916-UF	CACGAGCTGCTCATTCACGG	*smb20842*	3347-UR	GTCAGCATGTCCTCGTAGCG	*smb20723*
4916-UR	AATGCCGTGGAGCTGATGCG	*algI*	3347-DF	TTCGAGTCCCAGGTGAAGGC	*smb20724*
4916-DF	GTGCTTCATTCCTGCCTCGC	*algI*	3347-DR	CGACAATTGCAGGAGCTTGGC	*smb20725*
3573-UF	TCCAGTCATAGCTGGTGCGC	*repA1*	2356-IR	GGCTCATGGTCTTCTGACGC	*sma0017*
3573-UR	TTCGCCAACGCAAGTGTCGG	*repA1 - smb20047*	3432-OF	ACCACGTTCCATACCGACCC	*sma0089*
3573-DF	AGTGTCATCCTGGCAGAGGG	*smb20048 - fusA2*	3432-OR	TCGAACGGACGATAGCTGCG	*sma0091*
3573-DR	GTATCCGTTCGAGCATGGGG	*fusA2*	3432-IF	GAGGGGCGATATGAAGACCG	*sma0095*
6157-UF	CTCAGGTCAGTAACGACGCG	*smb20078*	3432-IR	ATTGGCAAGGACCAGCGTGC	*sma0095*

Primer name	Primer sequence	Gene containing primer	Primer name	Primer sequence	Gene containing primer
6157-UR	CCAGGATGTTGCTCCAACGC	*smb20078 - smb20079*	4224-OF	CGATCAAGGTTGCGAAGGGC	*pilQ2*
6157-DF	TGCTGCGGAAGATACGCTGC	*smb20079*	4224-OR	GCGTGAAATCCAGACCGACG	*pilQ2*
6157-DR	GTGTAACGGAGACCGATGCG	*smb20079*	4224-IF	CTGTTCGACGCCTCTATGCG	*sma0168*
641-UF	TCCTGATCTGGCGGTTGACG	*glpK*	4224-IR	CCACACCATCCACACAAGGC	*sma0169*
641-UR	CGCCAATCCTTGTGCATGGC	*glpK*	1084-IF	GAGATCACTTCGCCGTGAGG	*sma0224 – sma0226*
641-DF	TGACTGGGAGGAAGGAAGCG	*xdhA2 – xdhB2*	1084-IR	ACCTGTTCCAACTGCCGTGC	*sma0226*
641-DR	ATGGCAAGCAGGATCGCAGG	*xdhA2*	449-OF	TGCCATCATCCGAGCGAACG	*sma0333 – sma0335*
643-UF	AAGCACGGGAACCCAGTTGC	*smb21007*	449-OR	GGATCACCACGATGAGCACC	*sma0337*
643-UR	AGGCGAGTTCCACTTGACCG	*smb21005*	449-IF	TCACTCAAGCAGGTGGACGC	*sma0341*
643-DF	GGAGGAGGAGAGATTGACGG	*bdhA*	449-IR	ACGATGGAGGCACTACTCGC	*sma0346*
643-DR	TCCACCCAGCCTGACTTCGG	*glpK*	3850-OF	TCGGCAGTGAGCGATATGCG	*cyaI2*
2356-OF	TCGGGCAACGCAACTATGCC	*sma0013*	3850-OR	CCTGTTGACTGTGCTAGGCG	*cyaI2*
2356-OR	GAGATGTTCGAGGGTCTGGG	*selB*	3850-IF	TCGACGGCTTCCTCCTTTGC	*cyaI2 – sma0466*
2356-IF	TTTCCGCAGGTCTCTTCCGC	*selB*	3850-IR	CCGAATGGTCCAGAGTCAGC	*sma0466*
4232-OF	GCGCATCGTCCATTCATGGC	*nrtC*	1764-OR	GCAGGAAGAGACTGGAAGGC	*cyaE2*
4232-OR	CGATCAGGGTCTGTTCTGGC	*nrtB*	1764-IF	AGTAGCCCGAGTACGAAGGC	*sma1053*
4232-IF	TCAGTTGGTCGAGCGGAACG	*nrtA – sma0590*	1764-IR	TGGCCGAGCAGAAGAAGTCC	*sma1057*
4232-IR	AGGTCTCGAAGTGGTGGAGC	*sma4002*	1439-OF	GCCCAACTCCTAGCATTCGG	*sma1142*
2291-OF	CAACGGACTGCTACTGTCGC	*sma0649 – sma5027*	1439-OR	ATCCAGCAGTTCTCGGAGCC	*sma1142*
2291-OR	CTGCACAGTCGTTGAGTGCG	*sma5027*	1439-IF	ACCCGTTCTCACATGCTGGC	*sma1146*
2291-IF	GTCGAAGGTGAAGCTCTGCC	*sma0656*	1439-IR	CTATCGAGAGCGACGAAGGC	*sma1146*
2291-IR	CGAGTGTCAACCACGAACGC	*sma0657*	328-OF	CGCTTCAGCTCGTCAATCGC	*napC*
3734-OF	GCAACTCCTGGGACAATCGC	*nifE*	328-OR	GGTCAACACCAACCGATGCC	*napB*

Primer name	Primer sequence	Gene containing primer	Primer name	Primer sequence	Gene containing primer
3734-OR	CCCCACTCATACGGCTATCG	*nifE*	328-IF	TGCCGTCCTTCAGCATTCGG	*napA*
3734-IF	CGTGACAGCTCACAAGTCGC	*nifX*	328-IR	ATGCAGGCTGAGGTGAACCG	*napA*
3734-IR	GACATGAGTCTGGCAGTCCC	*fdxB*	6424-OF	CACAAGCCCTGGATCATCCG	*sma1328*
1142-OF	TGGCTTCGGTGGTGATGAGC	*groEL2*	6424-OR	GAACGCCGAGAATCCAAGCC	*sma1328 – sma1329*
1142-OR	AACACCTCGGACGAAGTCGC	*groEL2*	6424-IF	ATGTACCGTGCATCCTCCGG	*sma1332*
1142-IF	ATCGTCTGCCTTGAGGACGG	*sma0748*	6424-IR	ACCTCTCCGGTTCCTGATCC	*sma1334*
1142-IR	TTCCAGCAGCAACGAGACCG	*sma0750*	2460-OF	ATCGCAGTGAACGGCGTAGC	*ttuD3*
4733-OF	CGGCAAGATCATCGAGACGG	*sma0903*	2460-OR	CTCAAGGAGATGGACCAGGC	*sma1409*
4733-OR	GTCAACGAGCATTGGCTGCC	*sma0907*	2460-IF	TGTCCCTCCAACTCGAACCG	*sma1409*
4733-IF	TCGTCGATGTCACTGGACCG	*sma0909*	2460-IR	GATCCGCTTCATGTCGTGCG	*sma1410*
4733-IR	CAGGTCCTCGTTGAAGTCGC	*sma0911*	1337-OF	GCCAATGACGTAGCTGTCGC	*sma1500*
1764-OF	GGCAGTCAGACACCATACCG	*cyaE2*	1337-OR	CCAAGACCAGATGGCATCCC	*sma1500*
1337-IF	CAGTTCTGCTCATCGCACCC	*sma1503*	5575-OR	TCAGGCGATCTGTTCGAGGG	*sma1971*
1337-IR	GTGGTCGTCCAGCAGATTGC	*sma1505*	5575-IF	TTGTCGCCAAGAACGCTCGC	*sma1976*
1074-OF	CCAGGACATAGCGTACACCG	*cyaF5*	5575-IR	ATCTCCAACACGGACGACGC	*sma1978*
1074-OR	ACGCCACAACGGAGATTCGC	*sma1585 – syrB2*	3976-OF	CGCAGCTTGAACCAGTAGCG	*sma2075*
1074-IF	CTTCACAGGCGATCCAAGCC	*eglC*	3976-OR	TCGAATGGCAGCCGAGTTGC	*sma2075 – sma2077*
1074-IR	GCTTGATGAGGAGCTGGTCC	*eglC – sma1589*	3976-IF	GTCTGGTACACCCACTTGCG	*sma2077*
3625-OF	CTGAAGCCTGAGGTCGAAGC	*sma1662*	3976-IR	TCAAGGGCGAGGACATCACG	*sma2079*
3625-OR	CCATCACGCTTGCGATCACC	*sma1662*	4094-OF	TTGAGCGAGCGAAGTGACGC	*sma2197*
3625-IF	CCGACAGGTTCATTGGCTCG	*sma1664*	4094-OR	TGGACTGTCTCGCCAATGGC	*sma2199*
3625-IR	TCGAGGATGCACAAGCGACG	*sma1664*	4094-IF	ATACACCCGGCTGACAACGG	*sma2201*
2789-OF	ATGTCGAACCTCGTGGAGCG	*sma1652*	4094-IR	CTCATCGCCAATGCTCTCGC	*sma2203*

Primer name	Primer sequence	Gene containing primer	Primer name	Primer sequence	Gene containing primer
2789-OR	TGTCGTGGTGAGCGAACTCG	*sma1653*	1860-OF	GCGCAACGATCTCTTCACCC	*sma2301*
2789-IF	TCCTCGTCGGAATGCACTCC	*sma1660*	1860-OR	CTTGCTGGACGTACTGACGC	*sma2301*
2789-IR	GAACGCGATCCTCATCGTCG	*sma1662*	1860-IF	ATCTCCGCCGAGAGCAATGC	*sma2305*
2840-OF	CGCCCACTATGGTGATTGCG	*sma1869 – sma1871*	1860-IR	CTGAGCCATGTCAGGATCGG	*sma2305*
2840-OR	CGGATGTCCTGAATGGTGGG	*sma1872*	351-UF	GATCCACCGAATGCTGACGG	*sma0516*
2840-IF	CCGTGCAGTCGAATTGACCG	*sma1874*	351-UR	CGTTCGGCAATGAGCATCGG	*sma0520*
2840-IR	GCCTACTCATCCCACCTACC	*sma1874 – sma1875*	351-DF	GGTGATGACAGCGAGACTGG	*sma0523*
5575-OF	TCACCTCGACCTCAAGGACG	*sma1969*	351-DR	CATGAGAACGACCAGCGTCG	*sma0525*
987-UF	CGGGATCAACTCCAGTTCGC	*sma0355 – sma0357*	6403-UR	GTCAGCGGACTCGATAAGCG	*sma0374*
987-UR	GCATGTCAGCCATGACTGGC	*sma0357*	6403-DF	TCTGCCCTCCATGATGTCCC	*sma0376*
987-DF	CATGCGACAAAGTCCGAGGC	*sma0364*	6403-DR	GCGATTGCCGTTGCATTCGG	*sma0380*
987-DR	AGGTTCCTCACGCTCACTGG	*sma0364*	6473-UF	CCAGGATCTGAGACGTCAGG	*sma0431*
1061-UF	GGCATTGTCCTGAACCTCGC	*sma0424*	6473-UR	TCAGCGATGCGATCAGACGC	*sma0433*
1061-UR	TGCTCGATGAGGAAGGACGG	*sma0424*	6473-DF	CTTGAGGCTTCAGGCAAGGC	*sma0436 – sma5022*
1061-DF	GCCTTCTCCCAGAATGTCCG	*sma0429*	6473-DR	AGGAGGTCAACGGAGAAGCG	*sma0436 – sma5022*
1061-DR	GCGGCGAGATCATCGTTTCC	*sma0431*	3213-UF	TACGCGGGTTCTTCGATGCC	*sma0407*
3584-UF	GGAAGAACTCGGCATTGGCC	*sma0563*	3213-UR	CTCAGGTTGGAGATCGTGGG	*sma0407 – sma0412*
3584-UR	AATGCCTGACCGCTATCGGC	*sma0564*	3213-DF	TCTGGGACAGCAAGCATGGC	*sma0414*
3584-DF	GCGATGTCGTCTGGATTCCG	*sma5026*	3213-DR	ATCCGTGGAACTGCGAACCC	*sma0414*
3584-DR	GAGGCGCAACTTCAAGACGG	*sma0567 – cyaF4*	2956-UF	ACGCTGGTGATGAACCTGCG	*sma0403*
5516-UF	AACTCTCCTCCTCGGCAACG	*sma0557*	2956-UR	AAGGACCACCTCTGCAAGCC	*sma0404*
5516-UR	TGGCGAAGGCTTCACTCACC	*sma0559*	2956-DF	AGGTTGCTGCCGTCTATCGC	*sma0407*

Primer name	Primer sequence	Gene containing primer	Primer name	Primer sequence	Gene containing primer
5516-DF	GGAAGAACTCGGCATTGGCC	*sma0563*	2956-DR	CGACAACAGCGTCCTAAGCG	*sma0407 – sma0412*
5516-DR	GTGTTCCCGTCGAATCTGGC	*sma0564*	1689-UF	TGGTCATTTCCGCAGCTCCG	*sma0675*
5920-UF	AGCACGCAAGCGACCTTTGC	*sma0478*	1689-UR	GCTTCGACGATCTCCTCTCC	*sma0675*
5920-UR	TGACGGACTCCACCTGAAGC	*sma5024*	1689-DF	CATCGTGGCAAAGGCAAGCG	*sma0678*
5920-DF	TCCGCTTCGTAGACACTGCG	*thrC2*	1689-DR	GAGTCCTTCAACCGCTTCCC	*sma0680*
5920-DR	CTTCCTGATGCAGGTGGTCG	*thrC2*	54-UF	GGGGTAAACTGCTCCTTCGC	*sma0682*
6403-UF	GATACCTGTCTCGTCGTGCG	*sma0372*	4595UF	CTTCCTTGACGATGCCGAGC	*exoF3*
341UF	TCCAACTCGCGTAGACCAGC	*sma0574*	4595UR	TCTCACGGTCGGCTGCAAGC	*exoF3*
341UR	AGCAAACCGCTGTTCCAGGC	*sma0575*	4595DF	CTTCGAGGCTGCCTATGAGC	*smb21248*
341DF	AAGAACCTGCGCACTGTGGC	*cyaN*	4595DR	CGAGCAACTGCCGCTGAAGC	*smb21248*
341DR	AGTGCCATCGAGACACAGGC	*cyaN*	3938UF	TCATAGGCGTAGACGGTGCC	*smb21229*
54-UR	TCAACACCGAGTCCTGCTGG	*sma0682*	3938UR	GAGCGACGATGAGCTTTGGG	*smb21229*
54-DF	CACCGTGGCAGATTCAAGCC	*sma0684*	3938DF	CCGCTTTCCGAATGTCTCCG	*smb21231*
54-DR	ACGTTGACGCAGCCTGTTGC	*sma0684 – sma0687*	3938DR	ACCTGCGGACCTTGTGGACG	*smb21231*
ML7057	ACGAATTCCTGTCAGACCAAG TTTACTC	n.a.	ML7058	TGGCATGCTGAATACTCATACT CTTCC	n.a.
ML9318	CATGCCATGGTTTTCGCACGA TATACAGG	n.a.	ML9319	CTAGCTAGCTGATAGGTGGGCT GCCCTTC	n.a.
ML9647	CCGCTCGAGCGGTATCTGGAC AAGGGAAAACG	n.a.	ML9648	TCCCCCGGGGGACTCAGAAGA ACTCGTCAAG	n.a.
ML4876	ATAAGGGACTCCTCATTAAGA TAAC	n.a.	ML4875	TTAGGACAACTCCAGTGAAAA GTTC	n.a.
ML11667	TGCATGCATGCATGGATACCT GTGACAGGC	n.a.	ML11660	GAAGATCTTCATCGTCACCCGA CAGCC	n.a.
ML11659	CCGCTCGAGCGGCGACGATCT GCAGGTCTTGC	n.a.	ML11660	GAAGATCTTCATCG TCACCCGACAGCC	n.a.
ML11753	CCTTAATTAAGAGTTTCACCG TCGCCATTATCG	n.a.	ML11754	CCTTAATTAAGCGCCTTTTGCG CGCCTTCG	n.a.

Primer name	Primer sequence	Gene containing primer	Primer name	Primer sequence	Gene containing primer
ML20184	CCTTAATTAAGGCATCTTCCTGATCAGCCACG	n.a.	ML20185	CCTTAATTAAAAAAGACTGGCGCACCCGACAGG	n.a.
ML-07-3615	TGCATGCATCAGGTTCGAGAAGGAGACGG	n.a.	ML-07-3616	GAAGATCTAGGCATAGGGCCAGTTACGC	n.a.
sigma128	CCTCGATCAGCTCTTGCACTCG	n.a.	sigma129	GCAGGTGCTGGCATCGACATTCAGC	n.a.

n.a.-not applicable

CHAPTER 3. DELETION STRATEGY USING FRT TARGETING VECTOR pTH1522 and pTH1937.

ABSTRACT

The *Sinorhizobium meliloti* genome consists of 6204 predicted protein-coding regions and around 2000 of these encode proteins of unknown function (PUFs). To identify the functional significance of *S. meliloti* PUFs, one of several approaches we have adapted is to delete large defined regions using the Flp/FRT site-specific recombination system. To adapt the FRT/Flp system for use in *S. meliloti* we exploited a pTH1522-derived integration gene library of the *S. meliloti* genome (pTH1522 carries a single FRT site). A second FRT site is integrated at defined locations in the genome through integration of pTH1937-derivative plasmid that also carries a single FRT site. Here we outline how this Flp/FRT system was used to delete defined regions and hence generate multiple gene knock-out mutants. We report the construction of over thirty 3-350 kilobase deletions with defined endpoints that cover the entire *S. meliloti* pSymB (1700-kilobase) megaplasmid. Deleting large, regions of the genome helped us to identify phenotypes such as inability to grow on minimal media with fucose, maltotriose, maltitol, trehalose, palatinose, lactulose and galactosamine as sole carbon source. The inability to delete specific regions suggested the presence of essential genes. Using this strategy two essential genes $tRNA^{arg}$ and *engA* located on the pSymB megaplasmid were identified.

INTRODUCTION

The genome of *S. meliloti*, the symbiotic nitrogen-fixing soil bacterium, consists of three replicons: pSymA (1.35 Mb), pSymB (1.68 Mb) and chromosome (3.65 Mb) (Galibert *et al*., 2001). Ninety percent of the *S. meliloti* pSymB nucleotide sequence is predicted as protein coding, consisting of 1,570 ORFs (Finan *et al*., 2001). Insertion elements and regions duplicated elsewhere in the genome are also part of this megaplasmid. Analysis of the annotated pSymB sequence revealed a large number of solute transport systems. Genes that encode for transport systems represent 18.9% of the pSymB open reading frames against 10.4% for the chromosome (Mauchline *et al*., 2006). Genes involved in the biosynthesis and export of surface polysaccharides (essential for nodule invasion and adaptation to rhizosphere environments) and genes involved in nitrate/nitrite reduction are also present on the pSymB. In addition to these, many genes that have potential catabolic activities such as alcohol dehydrogenases were annotated (Finan *et al*., 2001). This replicon allows the bacteria to take up and oxidize many different compounds from the environment, playing an important role in the survival of the bacteria under diverse nutritional living conditions present in the soil and rhizosphere.

Large deletions have been mainly used to define minimal gene sets that contain only those genes that are essential and sufficient to sustain a functioning cell. Processes of minimization of a genome in the laboratory have been done with *E. coli* (Kolisnychenko *et al*., 2002; Yu *et al*., 2002; Fukiya *et al*., 2004; Hashimoto *et al*., 2005; Posfai *et al*., 2006; Mizoguchi *et al*., 2007) *B. subtilis* (Westers *et al*., 2003; Ara *et al*., 2007; Morimoto

et al., 2008), *Corynebacterium glutamicum* (Suzuki *et al*., 2005), *M. genitalium* (Glass *et al*., 2006), and *S. meliloti* (Charles and Finan, 1991, Oresnik *et al*., 2000). In Table 1.2 experimental approaches used for reduction of certain genomes were listed.

A Tn5-derived linkage map of the 1,683 Kb pSymB megaplasmid of *S. meliloti* was employed in the generation of large defined deletions of much of this megaplasmid (Charles and Finan, 1991). Deletion strains were generated by homologoues recombination between the insertion elements of flanking transposon insertions. Examination of the deletion mutants for phenotypes revealed novel loci required for dulcitol, melibiose, raffinose, beta-hydroxybutyrate, acetoacetate, protocatechuate and quinate utilization. Previously unidentified loci required for effective root nodule development and exopolysaccharide synthesis were also found (Charles and Finan, 1991). Oresnik *et al*., 2000, showed that repetitive rounds of Tn5B12-S mutagenesis with selection for deletion can be used successfully to cure the *nod-nif* megaplasmid, pRme2011a, of strain Rm2011.

Chain, 1998 in his MSc thesis project, used the Flp/FRT system in *S. meliloti* for *in vivo* deletion of DNA sequences flanked by direct FRT sites, providing a method for analysis of gene function by deletion of large regions in the genome instead of transposon insertion into a single gene. In this procedure, two suicide plasmids, each possessing an FRT site and IS50 DNA sequence, were mated into *S. meliloti*. Homologous recombination allowed insertion of each plasmid at a Tn5 site within the megaplasmid sequence. In some of the resulting clones, the FRT sites from both plasmids were direct repeats of each other. A *flp* recombinase-expressing plasmid was then introduced by

conjugation, and Flp recombinase was shown to mediate recombination between the FRT sites, excising the intervening DNA sequence. Further, the Flp/FRT system was used to generate several deletions on the symbiotic plasmid (pSymA) of *S. meliloti* (House *et al.*, 2003). In this procedure, two suicide plasmids, pD1 and pD2 both carrying FRT sites, were used to generate double integrant strain Sm1021::(pD1+pD2) with FRT sites in direct orientation. Deletions were obtained upon mating of the *Flp* recombinase-expressing plasmid pBH474 into double integrant strain. Genes that were targeted in those deletions were predicted to code for a denitrification pathway. The largest deletion that was made in this study was not bigger than 9 Kb.

Systematic deletion analysis of both megaplasmids, pSymA and pSymB, of *S. meliloti* has not been done to date and this was the main goal of the work described in this thesis.

MATERIAL AND METHODS

Construction of FRT Targeting Vector, pTH1937

Plasmid pTH1937 (Figure 3.1) was constructed in this study as a derivative of plasmid pACYC177 (Chang, A. and Cohen, S., 1978). pACYC177 is an *E. coli* plasmid cloning vector containing the p15A origin of replication. This allows pACYC177 to coexist in cells with plasmids of the ColE1 compatibility group (e.g., pBR322, pUC19). It

carries ampicilin (*bla*) and kanamycin resistance (*aph* (3')-*la*) genes which are part of Tn3 and Tn903, respectively.

First, the MCS (multiple cloning site) generated through annealing of two complementary oligonucleotides 49 and 57 bp long was ligated into an *Eco*RI/*Sph*I digested PCR product from pACYC177 (ML7057 AC**GAATTC**CTGTCAGACCAAG TTTACTC and ML7058 TG**GCATGC**TGAATACTCATACTCTTCC) (618 to 3709 nt) to give plasmid pTH1998. A 109 bp FRT fragment isolated from plasmid pMS101 (Snaith M., 1995) was cloned as an *Sph*I/*Bgl*II fragment into pTH1998 producing plasmid pTH1999. The Tn903 inverted repeat from pTH1999 was excised using restriction endonucleases *Nhe*I/*Dra*III and the recessed 3' termini of pTH1999 were filled in (Klenow fragment) and ligated to produce plasmid pTH2000. A 2.2 Kb product which includes nt 2735 to 1951 of pTH2000 as well as sites for *Nco*I/*Nhe*I restriction endonucleases, was synthesized by PCR (ML9318 CATG**CCATGG**TTTTCGCACG ATATACAGG and ML9319 CTA**GCTAGC**TGATAGGTGGGCTGCCCTTC) and ligated to the 267 bp *Nco*I/*Nhe*I fragment amplified from pTH1360 (pUC119 derivative (Zaheer, unpublished data)) which contains the *oriT* site from pRK2. Plasmid pTH2001 was obtained as a result of that ligation. The 1.5 Kb product which includes nt 1898 to 980 of pTH2001 as well as sites for *Sma*I and *Xho*I restriction endonucleases, was synthesized by PCR (ML9647 CCG**CTCGAG**CGGTATCTGGACAAGGGAAAACG and ML9648 TCC**CCCGG G**GGACTCAGAAGAACTCGTCAAG) and used in ligation with the neomycin resistance gene from Tn5 (1036 bp *Sma*I/*Xho*I fragment amplified from pTH1360) to produce the 2638bp plasmid pTH1937 (Figure 3.1). Fragments of

genomic *S. meliloti* DNA can be inserted at polylinker region containing the unique restriction sites *Spe*I, *Swa*I, *Eco*RV, *Pvu*I, *Pac*I and *Eco*RI of plasmid pTH1937. These fragments which include sites for *Spe*I and *Eco*RI restriction endonucleases were obtained by PCR using *S. meliloti* DNA fragments in pTH1522 genomic library fusions (Cowie *et al*. 2006) as a PCR template. All constructs were confirmed by DNA sequencing.

Construction of *S. meliloti* Strains Carrying Two FRT Sites

Sequential triparental matings using the *E. coli* mobilizing strain MT616 were performed in order to construct *S. meliloti* strains carrying two targeting vectors. Bacterial matings were performed as described previously (Finan *et al.*, 1986). First, pTH1522 (Figure 3.2) targeting vector, carrying an *S. meliloti* genomic DNA fragment was mated into the wild type strain RmP110 (Yuan *et al.*, 2005), and transconjugants were selected and purified on LB Sm200μg/ml Gm60μg/ml. The identity of recombinant *S. meliloti* strains was confirmed by direct sequencing from genomic DNA with the primers ML4876 ATAAGGGACTCCTCATTAAGATAAC and ML4875 TTAGGACAACTCCAGT GAAAAGTTC (see Chapter 2. Preparation of Genomic DNA). Transconjugant (Sm^rGm^r) strains were further used for the second triparental mating with pTH1937 (Nm^r suicide plasmid) carrying other *S. meliloti* genomic DNA fragments.

Construction of a *flp* Delivery Vector pTH1944

Plasmid pTH1944 was constructed in order to express *flp* recombinase in *S. meliloti*. A 2 Kb *Pst*I fragment containing the *flp* gene driven by *pcaD* promoter was isolated from plasmid pTH472 (Chain *et al*., 1998.). In plasmid pTH472 leaky expression of *flp* gene was observed despite of the fact that *flp* gene was driven by inducible *pcaD* promoter. A 2 Kb fragment containing the *flp* gene was cloned into pTH1919 (Cheng *et al*., unpublished data), to produce pTH1944 (Figure 3.3). The *flp* gene was cloned in pTH1919 in such way that expression of *flp* recombinase in pTH1944 is under control of *lac* promoter (4088 – 4127 nt) which is negatively regulated by *lacIq* (Figure 3.3). We note that pTH1944 does not carry the *lacI* gene, however *lac* operater (*lacO*) is present at 4068 - 4091 nt (Figure 3.3). In some experiments when another plasmid which expressed *lacIq* was used, 0.5 mM IPTG (Isopropyl-ß-D-thiogalactopyranosid) was added to the media to insure induction from the *lac* promoter. The ability of pTH1944-mediated *flp* expression to excise the *lacZ* gene in a *flp*-testing strain of *S. meliloti* RmH940 (Figure 3.4) (Chain *et al*., 1998.) was tested.

RESULTS AND DISCUSSION

Deletion strategy using FRT targeting vector pTH1522 and pTH1937

To delete DNA regions via Flp/FRT directed recombination, it is necessary to insert FRT sites in direct orientation on each side of the region to be excised. For one of the two FRT sites, we made use of the pTH1522 gene fusion library of *S. meliloti*. This library contains 6000 strains in which the FRT-containing plasmid pTH1522 (Figure 3.2) is integrated at known locations throughout the *S. meliloti* genome. The pTH1522 plasmid carries the *aacC4* gene, encoding gentamicin resistance (Gm^r). To insert the second FRT site into a pTH1522 integrant strain, we built a second FRT-containing plasmid designated pTH1937 (Figure 3.1). This plasmid carries the *nptII* gene encoding neomycin/kanamycin resistance (Nm^r) from Tn5, the *oriT* from RK2 to allow it's mobilization from *E. coli* and the *oriV* from pACYC177 that fails to replicate in *S. meliloti*. With the exception of a 103 bp region containing the FRT site, plasmid pTH1937 shares no homology with pTH1522 and hence no recombination is expected between these two plasmids. Consistent with this expectation, in control experiments, no Nm^r transconjugants were detected when pTH1937 was transferred into a *S. meliloti* strain carrying an integrated pTH1522 (frequency of transfer $<10^{-8}$/recipient). To insert a second FRT site at a particular location in the *S. meliloti* genome, the target region was cloned into pTH1937 and the resulting recombinant pTH1937 plasmid was transferred into the appropriate *S. meliloti* ΦpTH1522 integrant strain with selection for Sm^rNm^r

recombinants on LBmc containing Sm200 μg/ml Nm200 μg/ml (transfer frequency 10^{-4} /recipient). Following single colony purification *flp* recombinase was introduced into the *S. meliloti* ΦpTH1522 ΦpTH1937 double integrant strains by transferring the tetracycline resistant (Tc^r) *inc*P delivery plasmid pTH1944 (transfer frequency 10^{-4}/recipient). The ability of pTH1944-mediated *flp* expression to excise the *lacZ* reporter gene and the *nptII* Nm^r gene flanked by FRT sites in direct orientation in the *S. meliloti* RmH940 strain (Figure 3.4) (Chain *et al*., 1998.) was tested. *S. meliloti* strain RmH939 (lacking FRT sites) was used as a control in this experiment. Upon transfer of pTH1944 into RmH940 it was found that all transconjugants harbouring the *flp* plasmid lost the *lacZ* gene as viewed on X-gal plates as white colonies. In total, 100 colonies were screened. These colonies were confirmed to have lost the region between the minimal FRT sequences, by screening for the loss of the Nm^r gene. It was observed that Flp synthesis from pTH1944 resulted in 100% excision of the FRT-flanked *lacZ* gene, *nptII* Nm^r gene in *S. meliloti* RmH940. Since this plasmid expresses Flp constitutively, deletion between the two FRT sites occured in all pTH1944 transconjugant strains.

Expression of *flp* recombinase in pTH1944 is under the control of *lac* promoter which is negatively regulated by *lacIq*. In experiments when another plasmid with *lacIq* gene was used, repression of *lac* operon can be released in the presence of 0.5 mM IPTG. To investigate whether *flp* transcription in pTH1944 is regulated by LacIq, we employed the same testing strain of *S. meliloti*, RmH940. Plasmid pTH1931 carrying *lacIq* was transferred into RmH940 to give RmP2678 and pTH1944 was subsequently transferred into this strain. The ability of pTH1944-mediated *flp* expression in the presence of LacIq

and 0.5 mM IPTG to excise the *lacZ* gene in a *flp*-testing strain of *S. meliloti* RmH940 (Figure 3.4) (Chain *et al*., 1998) was tested. *S. meliloti* strain RmH939 was used as a control in this experiment. It was found that in all transconjugants where the *lac* promoter was repressed by the LacIq repressor, repression was released in the presence of IPTG. This was seen as the loss of the *lacZ* gene in a testing strain RmH940 (viewed on X-gal plates as white colonies). These colonies were confirmed to have lost the region between the minimal FRT sequences, by screening for the loss of the Nm^r gene. In a screen of 100 colonies it was observed that *flp* synthesis from pTH1944 in the presence of IPTG resulted in 100% excision of the FRT-flanked *lacZ* gene and Nm^r gene in *S. meliloti* RmH940. Furthermore, only 13% of colonies lost *lacZ* gene in the absence of the IPTG (100 colonies in total were screened).

The two FRT vectors carry Gm^r and Nm^r, respectively and thus provide phenotypic markers for detection of deletion events. Moreover when expressed, the *lacZ* gene in pTH1522 can also be detected as blue colonies growing on LB containing 5-bromo-4-chloro-3-indolyl-beta-D-galacto-pyranoside (X-Gal). Two different deletion phenotypes result from recombination between the FRT sites depending on whether the deleted region is 5' or 3' relative to the pTH1522 FRT site. In the example illustrated in Figure 3.5 we assume that the circular –A'B-CD'- and –C'D-EF'- regions are unable to replicate and are lost (deleted) from the cell. If pTH1522 integrates into the *S. meliloti* genome downstream from where pTH1937 has integrated the expected phenotype for the deletion mutant will be $Sm^rGm^rNm^r$ and the *lacZ* gene from pTH1522 (if applicable) will be expressed at a lower level in comparison to the parental strain (Figure 3.5). When

pTH1522 integrates into the genome upstream of pTH1937, the expected deletion phenotype is $Sm^{r}Gm^{s}Nm^{s}$ and the pTH1522 *lacZ* gene will be expressed at the same level as in the parental strain (Figure 3.5).

REFERENCE

Ara, K., K. Ozaki, K. Nakamura, K. Yamane, J. Sekiguchi, N. Ogasawara. (2007) *Bacillus* minimum genome factory: effective utilization of microbial genome information. Biotechnol. Appl. Bioc. **46**:169–178.

Chang, A., S.N. Cohen. (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J. Bacteriol. **134** :1141-1156

Charles, T.C., 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., T.M. Finan.(1990) Genetic map of *Rhizobium meliloti* megaplasmid pRmeSU47b. J. Bacteriol. **172**:2469-76.

Chain, P.S.G. (1998) Development of an *in vivo* DNA cloning procedure for bacteria. M.Sc. thesis. McMaster University, Hamilton, Ontario, Canada.

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

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

Finan, T.M., S. Weidner, K. Wong, J. Buhrmester, P. Chain, F. J. Vorhölter, I. Hernandez-Lucas, A. Becker, A. Cowie, J. Gouzy, B. Golding, A. Pühler. (2001) The complete sequence of the 1,683-Kb pSymB megaplasmid from the N_2-fixing endosymbiont *Sinorhizobium meliloti*. Proc.Natl.Acad.Sci.U.S.A. **98**:9889-94.

Fukiya, S., H. Mizoguchi, H. Mori. (2004) An improved method for deleting large regions of *Escherichia coli* K-12 chromosome using a combination of Cre/loxP and lambda red. FEMS Microbiol. Lett. **234**: 325–331.

Galibert, F., T.M. Finan, S.R. Long, A. Pühler, P. Abola, F. Ampe, F. Barloy-Hubler, M.J. Barnett, A. Becker, P. Boistard, G. Bothe, M. Boutry, L. Bowser, J. Buhrmester, E. Cadieu, D. Capela, P. Chain, A. Cowie, R.W. Davis, S. Dréano, N.A. Federspiel, R.F. Fisher, S. Gloux, T. Godrie, A. Goffeau, B. Golding, J. Gouzy, M. Gurjal, I. Hernandez-Lucas, A. Hong, L. Huizar, R.W. Hyman, T. Jones, D. Kahn, M.L. Kahn, S. Kalman,

D.H. Keating, E. Kiss, C. Komp, V. Lelaure, D. Masuy, C. Palm, M.C. Peck, T.M Pohl, D. Portetelle, B. Purnelle, U. Ramsperger, R. Surzycki, P. Thébault, M. Vandenbol, F.-J. Vorhölter, S. Weidner, D.H. Wells, K. Wong, K-C. Yeh, J. Batut. (2001) The composite genome of the legume symbiont *Sinorhizobium meliloti*. Science. **293**:668-672.

Glass, J.I., N. Assad-Garcia, N. Alperovich, S. Yooseph, M.R. Lewis, M. Maruf, C.A. III Hutchison, H.O. Smith, J.C. Venter. (2006) Essential genes of a minimal bacterium. Proc. Natl. Acad. Sci. U.S.A.**103**: 425–430.

Goryshin, I.Y., T.A. Naumann, J. Apodaca, W.S. Reznikoff. (2003) Chromosomal deletion formation system based on Tn5 double transposition: use for making minimal genomes and essential gene analysis. Genome Res. **13**:644-53.

Hashimoto, M., T. Ichimura, H. Mizoguchi (2005) Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. Mol. Microbiol. **55**:137–149.

House, B.L., M.W. Mortimer, M.L. Kahn. (2004) New recombination methods for *Sinorhizobium meliloti* genetics. Appl. Environ. Microbiol. **70**:2806-15.

Kolisnychenko, V., G. Plunkett III, C.D. Herring, T. Fehér, J. Pósfai, F. R. Blattner, G. Pósfai. (2002) Engineering a reduced *Escherichia coli* genome. Genome Res. 12:640-7.

Mauchline, T.H., J.E. Fowler, A.K. East, A.L. Sartor, R. Zaheer, A.H.F. Hosie, P.S. Poole, and T.M. Finan.2006. Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome. Proc.Natl.Acad.Sci.U.S.A. **103**:17933-8.

Mizoguchi, H., H. Mori, T. Fujio. (2007) *Escherichia coli* minimum genome factory. Biotechnol. Appl. Bioc. **46**:157–167.

Morimoto, T., R. Kadoya, K. Endo. (2008) Enhanced recombinant protein productivity by genome reduction in *Bacillus subtilis*. DNA Res. **15**:73–81.

Oresnik, I.J., S.L. Liu, C.K. Yost, M.F. Hynes. (2000) Megaplasmid pRme2011a of *Sinorhizobium meliloti* is not required for viability. J. Bacteriol. **182**:3582-6.

Posfai, G., G.III Plunket, T. Feher. (2006) Emergent properties of reduced-genome *Escherichia coli*. Science **312**:1044–1046.

Snaith M.R., J.A.H. Murray, C.A. Boulter. (1995) Multiple cloning site carrying *loxP* and FRT recognition sites for the Cre and Flp site-specific recombinases. Gene. **166**:173-174.

Suzuki, N., H. Nonaka, Y. Tsuge, S. Okayama, M. Inui, H. Yukawa. (2005) Multiple large segment deletion method for *Corynebacterium glutamicum*. Appl. Microbiol. Biotechnol. **69**: 151–161.

Westers, H., R. Dorenbos, J.M. van Dijl. (2003) Genome engineering reveals large dispensable regions in *Bacillus subtilis*. Mol. Biol. Evo. **20**:2076–2090.

Yu, B.J., B.H. Sung, M.D. Koob, C.H. Lee, J.H. Lee, W.S. Lee, M.S. Kim and S.C. Kim. (2002) Minimization of the *Escherichia coli* genome using a Tn5-targeted Cre/loxP excision system. Nat. Biotechnol. **20**: 1018–1023.

Yuan, Z.C., R. Zaheer, T.M. Finan. (2005) Phosphate limitation induces catalase expression in *Sinorhizobium meliloti*, *Pseudomonas aeruginosa* and *Agrobacterium tumefaciens*. Mol. Microbiol. **58**:877-94.

Figure 3.1. Targeting vector pTH1937 used in this study. Plasmid pTH1937 is a derivative of pACYC177; *E. coli* plasmid cloning vector containing the p15A origin of replication (*oriV*). Plasmid pTH1937 can coexist in cells with plasmids of the ColE1 compatibility group (e.g., pBR322, pUC19). It carries kanamycin resistance *nptII* gene (Kn^r) which is part of Tn5 and origin of transfer (*oriT*) from RK2 plasmid. Also indicated is the position of the FRT site, Tn3 inverted repeat (Tn3 inv rpt) and multiple restriction enzyme sites *Spe*I, *Swa*I, *Eco*RV, *Pvu*I, *Pac*I and *Eco*RI. Plasmid pTH1937 is low copy number plasmid, at about 15 copies per cell.

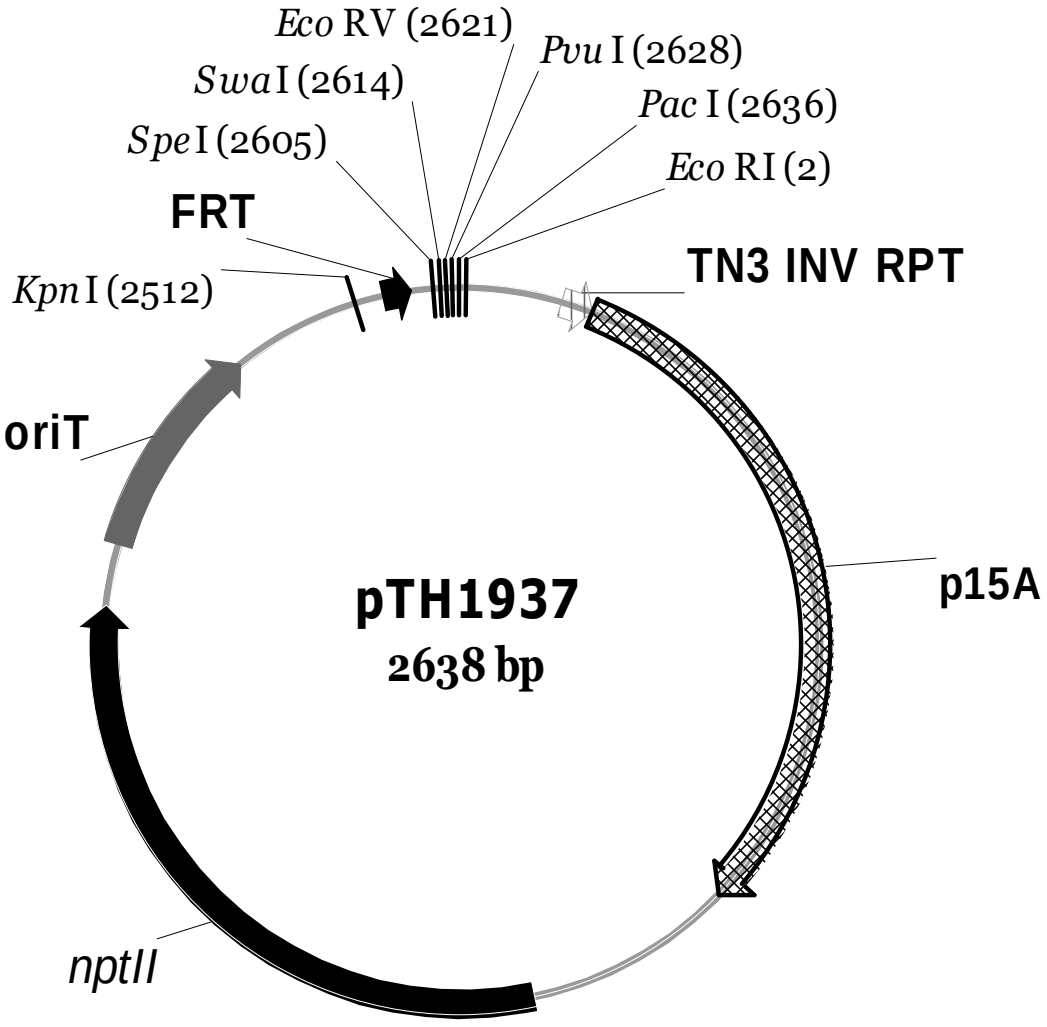

Eco RV (2621)
Pvu I (2628)
Swa I (2614)
Pac I (2636)
Spe I (2605)
Eco RI (2)
FRT
TN3 INV RPT
Kpn I (2512)
oriT
pTH1937
2638 bp
p15A
nptII

Figure 3.2. Targeting vector pTH1522 used in this study (Cowie *et al.*, 2006). Plasmid pTH1522, with reporter genes *gfp*+, *lacZ*, *gusA*, and *tdimer2(12)* shown as divergent operons from the *Xho*I cloning site. Also indicated is the position of the FRT site, *attP* site and gentamicin resistance gene and the locations of stop codons either side of the cloning site to prevent translational readthrough.

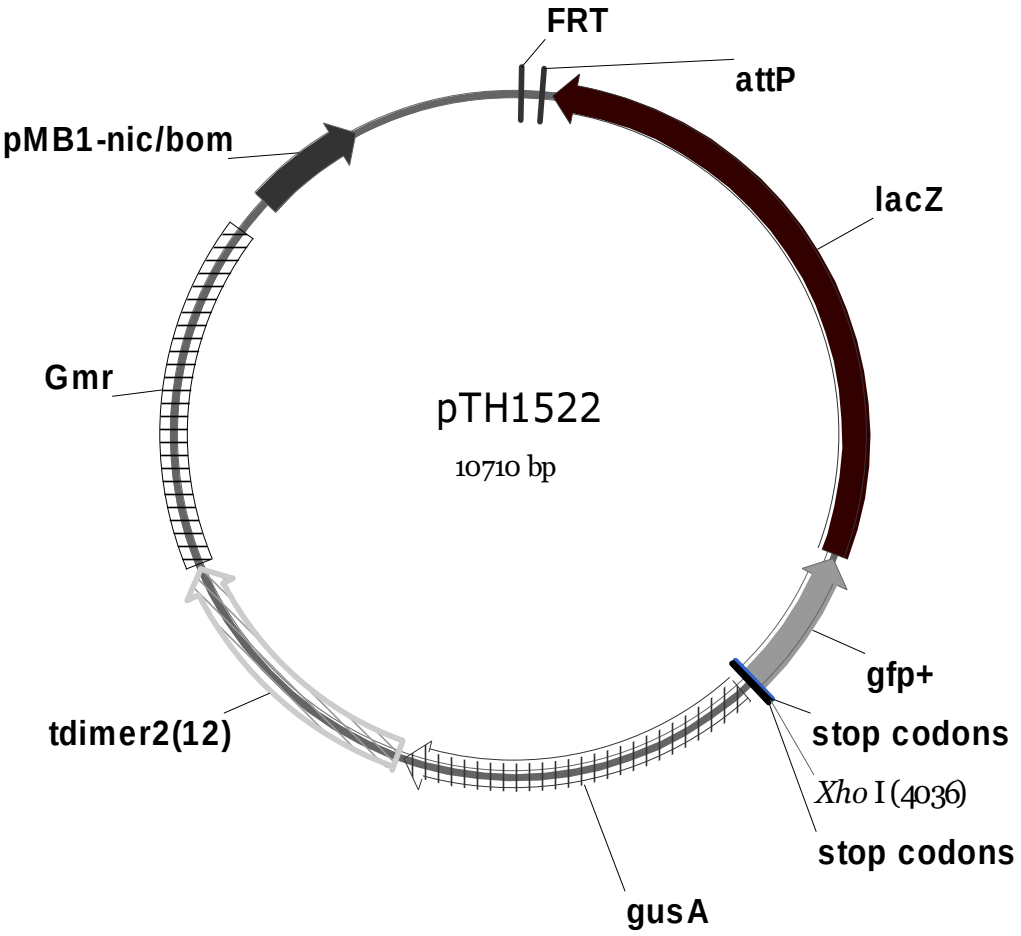

FRT
attP
lacZ
pMB1-nic/bom
Gmr
pTH1522
10710 bp
gfp+
stop codons
Xho I (4036)
stop codons
tdimer2(12)
gusA

Figure 3.3. Flp delivery vector pTH1944. Plasmid pTH1944 is a derivative of pBBR1; *E. coli* plasmid cloning vector containing the pBBR1 origin of replication (*oriV*). It carries tetracycline resistance gene (TetA-TetR) and origin of transfer (*oriT*) from RK2 plasmid. Also indicated is the position of the *flp* recombinase gene and unique restriction enzyme sites *Bgl*II, *Xho*I, *Hind*III, and *BamH*I. Expression of *flp* recombinase in pTH1944 is under control of *lac* promoter (4088 – 4127 nt). The pTH1944 does not carry the *lacI* gene, however *lac* operater (*lacO*) is present at 4068-4091 nt.

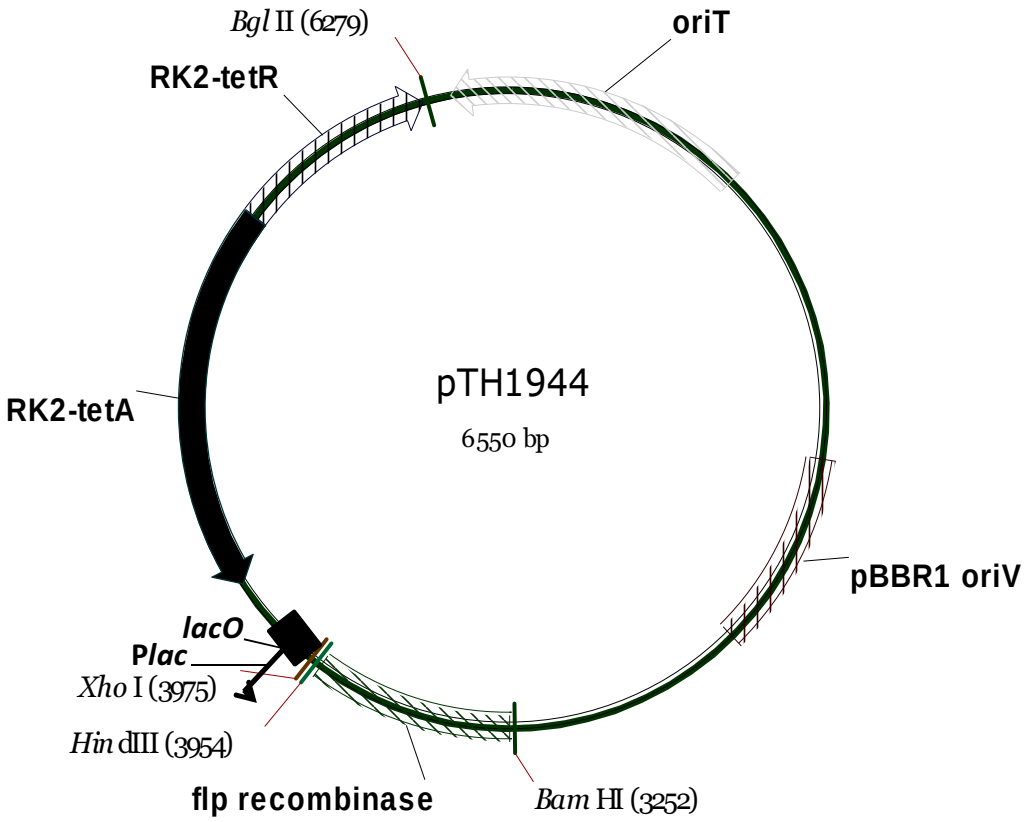

Bgl II (6279)
oriT
RK2-tetR
pTH1944
6550 bp
RK2-tetA
pBBR1 oriV
lacO
Plac
Xho I (3975)
Hin dIII (3954)
flp recombinase
Bam HI (3252)

Figure 3.4. Flp-testing strains of *S. meliloti* RmH940 and RmH939 (Chain *et al*., 1998.) RmH940, has a Lac^- background and inserted into it a Tn5-235 element possessing a neomycin resistance (Nm^r) gene and *lacZ* flanked by direct *flp* recombinase target site (FRT) repeats. When Flp recombinase is expressed, the sequence between the FRT repeats is excised, and the resulting clones are Nm^s and Lac^-. RmH939, has a Lac^- background and inserted into it a Tn5-235 element possessing a neomycin resistance (Nm^r) gene and *lacZ* but lacking FRT repeats. This strain is used as control in all experiments.

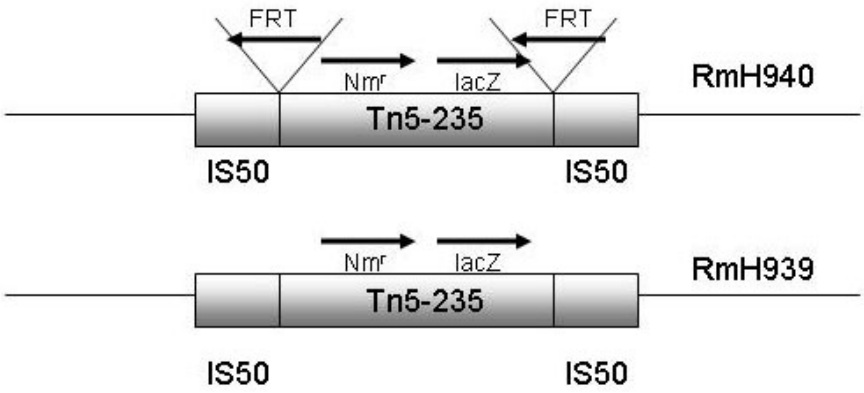
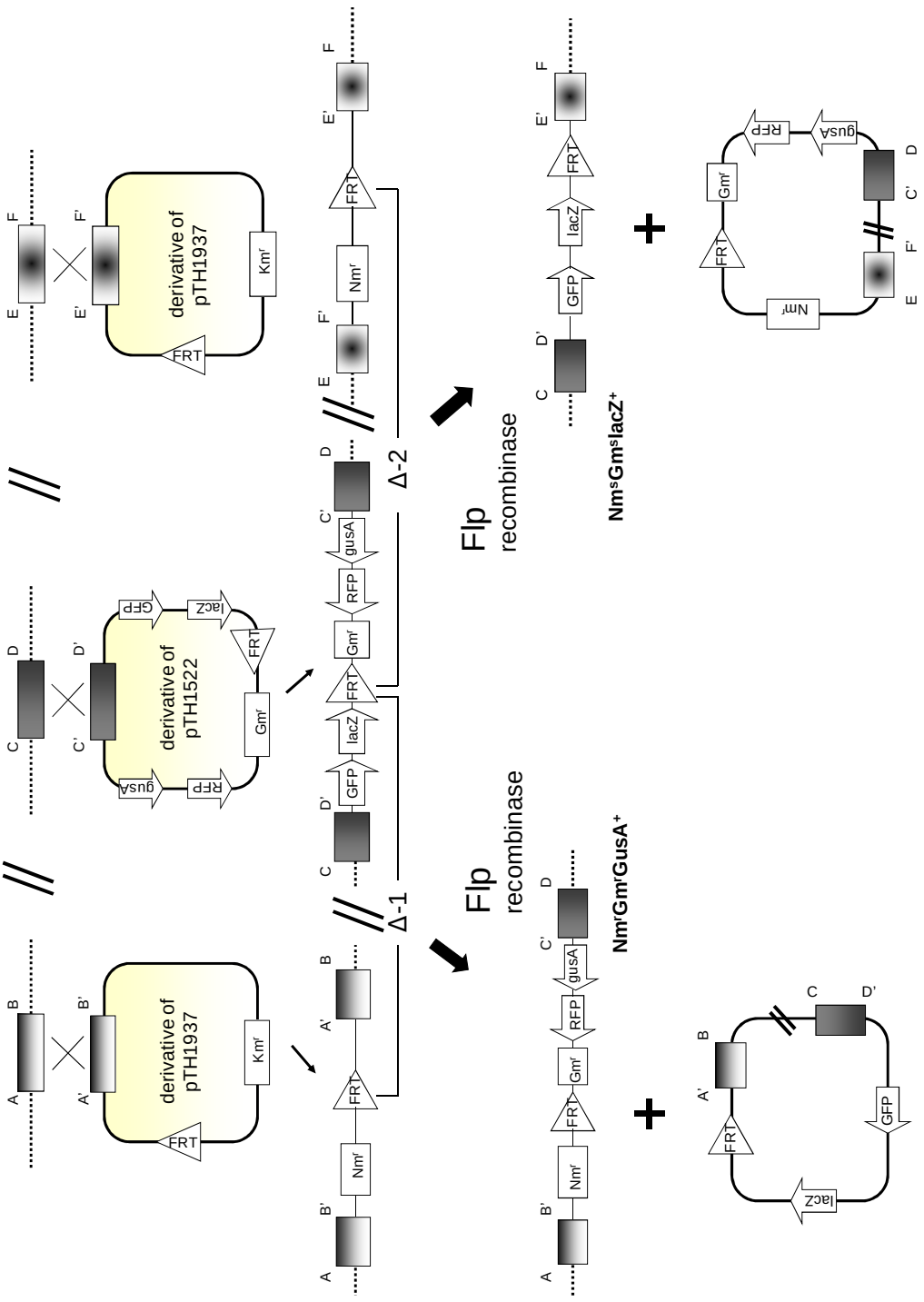

FRT
FRT
Nmr
lacZ
RmH940
Tn5-235
IS50
IS50
Nmr
lacZ
RmH939
Tn5-235
IS50
IS50

Figure 3.5. Two alternate FRT-directed phenotypes/genotypes resulting from pTH1937 integrants on either side of a single pTH1522 integrant in the genome. AB, CD, EF are random DNA sequences that were cloned into targeting vectors. Km^r-kanamycin resistance; Gm^r- gentamicin resistance; Nm^r- neomycin resistance; FRT-*flp* recombinase target site; reporter genes *gusA, lacZ. GFP* (green fluorescence protein); *RFP* (red fluorescence protein); Δ-1-deleted region 1; Δ-2-deleted region 2.


derivative of pTH1937
Km^r
FRT
derivative of pTH1522
lacZ
GFP
FRT
Gm^r
gusA
RFP
Δ-1
Δ-2
Nm^r
GFP
lacZ
RFP
gusA
Flp
recombinase
Nm^sGm^slacZ^+
Nm^rGm^rGusA^+

CHAPTER 4. DELETION ANALYSIS OF THE *Sinorhizobium meliloti* pSymB MEGAPLASMID

MATERIAL AND METHODS

Construction of the pTH1996 and pTH1997

Plasmid pTH1996 was used for the analysis of *engA* expression and was constructed as follows: a 2.1 Kb *Nsi*I/*Bgl*II PCR fragment (ML11667 TGC**ATGCAT**GCATGGATACCTGTGACAGGC and ML11660 GA**AGATCT**TCA TCGTCACCCGACAGCC) including 1246 nt of the *engA* gene to 869 nt downstream of the translational stop for *engA* was cloned into pTH1703 (Cowie *et al*., 2006) to give plasmid pTH1996 (Figure 4.6).

Plasmid pTH1997 (Figure 4.6) was constructed in order to disrupt the *engA* gene by single crossover recombination. Plasmid pTH1997 is pTH1703 with 740 bp *Xho*I/*Bgl*II fragment of the *engA* gene from nt 185 to 925. This fragment was amplified by PCR from RmP110 genomic DNA using the primers ML11659 CCG**CTCGAG**CGGCGACGATC TGCAGGTCTTGC and ML11660 GA**AGATCT**TCATCG TCACCCGACAGCC.

Cloning of *engA* and *smb21712*

A 1.4 Kb *Pac*I fragment containing the *engA* coding region was amplified from RmP110 DNA using the primers (ML11753 CCTTAATTAAGAGTTTCACCGTC GCCATTATCG and ML11754 CCTTAATTAAGCGCCTTTTGCGCGCCTTCG). The product was cloned as a *Pac*I fragment into expression vector pTH1931 to produce plasmid pTH2027. This plasmid was used to provide *engA* expression *in trans*.

A 460 bp *Pac*I fragment (ML20184 CCTTAATTAAGGCATCTTCCTGATC AGCCACG and ML20185 CCTTAATTAAAAAAGACTGGCGCACCCGACAGG) which includes nt 1,314,806 to 1,315,265 of pSymB megaplasmid, was synthesized by PCR and ligated to expression vector pTH1931 to produce plasmid pTH2302. Plasmid pTH2302 was used to provide $tRNA^{arg}$ (*smb21712*) expression *in trans*.

Construction of the pTH2172 – pTH2185 Vectors

Table 4.1 outlines the fragments cloned into the reporter suicide plasmid pTH1703 to generate the pTH2172 – pTH2185 plasmids. Integration of the resulting plasmids into the *S. meliloti* genome generated *gusA* reporter fusions (odd numbered plasmids) or disrupted the target gene (even numbered plasmids) Plasmids pTH2173, pTH2175, pTH2177, pTH2179, pTH2181, pTH2183 and pTH2185 have been used for the promoter expression analysis of the *pssF*, *uxuB*, *smb20750*, *smb20751*, *smb20752*, *smb20753* and *smb20754*, respectively. Each of these seven genes was disrupted by insertion of suicide

plasmid (pTH2172, pTH2174, pTH2176, pTH2178, pTH2180, pTH2182 and pTH2184) in *S. meliloti* strain RmP110. Primer sequences used in this experiment as well as locations of primers on pSymB megaplasmid are listed in Table 4.1.

Construction of the pTH2470 Vector

Plasmid pTH2470 was constructed to knock out *thuE* in *S. meliloti*. It was made by PCR amplifying a 667 bp *Nsi*I/*Bgl*II fragment (ML-07-3615 TGCATGCATCAGG TTCGAGAAGGAGACGG and ML-07-3616GAAGATCTAGGCATAGGGCCAG TTACGC) which includes nt 140 to 454 of the *thuE* gene, using genomic DNA from RmP110 as a template and cloned into pTH1703.

RESULTS

Deletion of the complete pSymB megaplasmid

Using the deletion strategy described in Chapter 3, fifty-six strains that carried FRT sites from pTH1522 and pTH1937 in direct orientation in the pSymB megaplasmid were made. A map of pSymB together with the FRT-flanked regions targeted for deletion is shown in Figure 4.1. The FRT flanked regions of pSymB were designated in ascending numbers from B100 (see Figure 4.1). Upon transfer of the *flp* plasmid pTH1944, the largest deletion recovered was 370 Kb, while the smallest was 3 Kb (Figure 4.1).

Collectively, the double FRT strains encompassed the entire 1678 Kb pSymB megaplasmid of *S. meliloti*. The precise boundaries of the deletion strains are given in Table 2.1. Where pTH1944 *flp* plasmid transconjugant strains were recovered and the deletion subsequently verified, the region was designated with a Δ (e.g. ΔB118). When no pTH1944 *flp* plasmid transconjugant colonies were recovered, no deletion symbol was designated (e.g. B119). No deletions were recovered for four regions of the pSymB megaplasmid and these are discussed further below.

Confirmation of pSymB Deletions

To investigate whether the FRT flanked regions were deleted and lost from the cell, we used PCR to detect whether a given DNA region was present or absent through the amplification, or not, of a predicted DNA product (Figure 4.2). All the primer pairs amplified products of the predicted size from template DNA from the wild-type *S. meliloti* strain P110. As the endpoints for each deletion were precisely defined, we designed four pairs of primers, to analyze the deletion positions (Figure 4.2). Only those deletion strains that were confirmed by PCR are shown in Figure 4.1. The sequence and positions of the primer pairs are given in Table 2.3. (Primers used in this study). For a small number of strains, PCR products were obtained where no product was predicted. This could usually be attributed to a mix-up of strains – as reconstruction of the strains resulted in the predicted deletion phenotypes. An example of PCR results for the deletion strain ΔB123 is shown in Figure 4.2.

Essential genes on pSymB megaplasmid

The pSymB megaplasmid is 1,683,333 bp in size. Megaplasmid pSymB is annotated with nucleotide position 1 as the A of the ATG start codon of the *lacE* gene and the numbering proceeds clockwise from this position on the map as drawn in Figure 4.1. We failed to recover pTH1944 *flp* plasmid transconjugants from four FRT-flanked regions of the pSymB megaplasmid (B125 (60,148 - 1,677,882 nt), B140 (64,141 – 467,160 nt), B110 (1,204,770 - 1,256,503 nt), B117 (1,307,752 - 1,529,711 nt), see Figure 4.1). Since this suggested that deletion of these regions was lethal, we analyzed these regions further for the presence of possible essential genes.

The pSymB *oriV* locus

The *repA1B1C1* genes of pSymB lie at nucleotides 58,760 – 55,090 and the pSymB *oriV* is believed to be located within the *repC* gene at a location yet to be defined (MacLelland *et al* 2005). The B161 region (pTH1522 (1677882–1679723); pTH1937 (49523–51610)) was readily deleted whereas no pTH1944 recombinants were recovered for either B125 (pTH1522 (59089–60148); pTH1937 (1677882–1679723)) or B162 (pTH1522 (59089–60148); pTH1937 (49523–51610)) B125 and B162 but not B161 includes the 50,791 - 60,148 nt region which has nine annotated genes – and most significantly the *repA1B1C1* locus which is capable of autonomous replication and is presumed to contain the *oriV* of the pSymB megaplasmid. Assuming pSymB has no

other *oriV*, excision of the *repABC* locus would result in loss of the remainder of the pSymB megaplasmid (Figure 4.1). Therefore, the failure to recover transconjugants in which the *repA1B1C1* region was deleted suggests that pSymB replicon has only one *oriV* and further that there are genes carried on the rest of the pSymB megaplasmid that are essential for cell survival.

The *smb20055- smb20059* locus

No deletion mutants that removed the 403 Kb B140 region from 64,141 to 467,160 nt were recovered. However the ΔB141 (100,636 – 467,160 nt) and ΔB142 (121,311 – 467,160 nt) regions were readily recovered and this suggested that the region from 64,141 – 100,636 nt carried essential genes. To further localize the region required for growth, five additional FRT double integrant strains were made. Of these, pTH1944-*flp*-plasmid transconjugants were recovered for B151 (61,240 – 65,620 nt) while no transconjugants were recovered from strains carrying B152 (61,240 – 74,302 nt), B153 (61,240 – 89,794 nt), B154 (61,240 – 101,396 nt) and B155 (61,240 – 122,108 nt). These data localized the essential gene/genes within the region 65,620 to 74,302 nt. Within this region lie ten genes: s*mb20055*, *smb20056*, *smb20057*, *smb20058*, *smb20059*, TRm5, *smb20061*, *smb20062*, *smb20063* and *smb20064*. In other work in our laboratory, we characterized the *cbtJKL* (*smb20056-20058*) gene cluster and our data suggest that these genes encode an ABC-transport system for the uptake of cobalt. Interestingly, mutations in these genes result in an inability to form colonies on LB medium unless the medium is

supplemented with cobalt (Cheng *et al*., 2011; see *smb20057* Table 8, Cowie *et al.*, 2006). We tested whether addition of cobalt would allow recovery of pTH1944 transconjugants and hence ΔB140, ΔB152, ΔB153, ΔB154 and ΔB155 deletion strains. Transconjugants were recovered when selected on LB media supplemented with $CoCl_2$ (2 ng/ml). These results confirmed that the inability to recover deletions in this region stemmed from an involvement of genes in this region in cobalt metabolism. Cobalt is known to be required for growth of Rhizobia (Lowe *et al*. 1960; Lowe and Evans, 1962; Watson *et al*. 2001). We screened a *S. meliloti* pLAFR1 clone bank (Friedman *et al*., 1982) for cosmids that allowed ΔB152 to grow on LB medium. Cosmid DNA from five ΔB152 transconjugant strains was examined by sequencing using primers (sigma128 CCTCGATCAGCTCTTGCACTCG and sigma129 GCAGGTGCTGGCATCGACA TTCAGC) from either side of the pLAFR1 *Eco*RI insert site. The insert DNA in four of the plasmids were identical (56,613 to 81,543 nucleotides) whereas cosmid pTH2654 carried 56,613 – 74,702 nt from pSymB together with a fragment from pSymA. The approximately 20-Kb region of *S. meliloti* DNA contained within the all five cosmids (pTH2651 to pTH2655) includes genes annotated to encode a single ABC-type transport system for the uptake of cobalt. The requirement for cobalt in most bacteria is primarily due to its presence in vitamin B_{12} which plays a number of crucial roles in many biological functions. (Martens *et al.*, 2002). The *cobT* gene which is presumed to be required for cobalamine- vitamin B_{12} biosynthesis in *S. meliloti* was previously identified as essential for *S. meliloti* (Cowie *et al*., 2006).

The *engA* gene region: The pSymB *engA* is essential for *S. meliloti*

Flp transconjugants of strains with FRT sites flanking the 52 Kb B110 region (1,204,770-1,256,503 nt) were not recovered. To localize the region that appeared to be essential for cell viability, additional FRT strains were made and no transconjugants were obtained for the B113, B114 and B115, whereas Flp transconjugants were recovered for the B111 (1,204,770 – 1,225,824 nt) and B112 (1,204,770 – 1,226,491 nt) (Figure 4.1). Assuming one "essential" locus, these data suggest that the locus lies in the 8.6 Kb region from 1,226,491-1,235,118 nt. A map of this region is shown in Figure 4.6. This region carries the genes annotated as *smb20991, smb20992, smb21674, smb20993, smb20994, engA, smb20996 and smb20997*. Analysis of these genes with respect to the identification of essential genes in other genomes, suggested that the 1431 bp *engA* gene (1,231,075 - 1,232,505 nt) was a good candidate for the essential gene in this region. *engA* is annotated as a GTP binding protein, a member of the Era subfamily of bacterial GTPases and has been found to be essential in bacteria such as *N. gonorrhoeae, E. coli* and *B. subtilis* (Caldon *et al.*, 2003).

To investigate whether *engA* was essential for *S. meliloti* cell viability, we attempted to construct *engA* null alleles and also to delete the *engA* gene region from pSymB in a strain carrying *engA in trans* on another plasmid. Plasmid pTH1996 (Figure 4.6) carries an 740 bp fragment internal to the *engA* gene cloned in pTH1703, while pTH1997 is the same plasmid except it carries the complete *engA* gene together with flanking DNA (Figure 4.6). Upon transfer and recombination of these plasmids into *S.*

meliloti, no Gm^r transconjugants were recovered for pTH1996 ($<10^{-8}$/recipient) whereas pTH1997 Gm^r transconjugants were readily recovered ($\sim10^{-4}$/recipient). This data suggested that disruption of *engA* gene resulted in a loss of the cell viability.

To provide *engA* expression *in trans*, the *engA* gene was cloned downstream of the *lac* promoter in broad host range Sp^r pBBR derivative to give pTH2027. Plasmid pTH2027 was transferred to strains RmP734 (B110) and RmP826 (B115). Tc^r transconjugants were recovered when the Flp recombinase plasmid pTH1944-Tc^r was transferred into the RmP734 (pTH2027) and RmP826 (pTH2027) strains whereas no Tc^r transconjugants were recovered upon mating into RmP734 and RmP826. This result again suggests that *engA* is essential and that it is the only essential gene in this region of the pSymB megaplasmid. We note that the *bacA* gene lies four genes away from *engA*, and *S. meliloti bacA* mutants are viable (Ichige and Walker, 1997). We note that in the above experiments IPTG was added to the media to insure induction of *flp* recombinase from the *lac* promoter in pTH1944. Presence of a deletion within the *engA* gene region from pSymB in a strain carrying *engA in trans* on another plasmid was checked by PCR using four sets of primers per deletion strain: 2626UFUR; 2626DFDR; 4440UFUR, 4440DFDR for strain RmP734 and 641UFUR; 641DFDR; 4440UFUR, 4440DFDR for strain RmP826 (Table 2.1). Only those deletion derivative strains that were confirmed by PCR were saved as a part of our lab strain collection.

The unique copy of arginine tRNA on pSymB megaplasmid is essential for *S. meliloti.*

No transconjugant colonies were obtained when the pTH1944-*flp* plasmid was transferred to a strain carrying the FRT-flanked B117 region (1,307,905 - 1,528,150 nt). However transconjugant colonies were obtained for strain RmP811 carrying the B118 FRT-flanked region (1,322,226 – 1,529,711 nt) and three other FRT-flanked regions ΔB119 (1,332,882 – 1,529,711 nt) , ΔB120 (1,374,412 – 1,529,711 nt), ΔB121 (1,408,135 –1,529,711 nt). As the *minCDE* genes are located at nucleotides 1,446,108 – 1,448,030, the ΔB119, ΔB120, and ΔB121 strains have lost these genes yet grew like wildtype on LB agar. These data are consistent with our previous findings showing that *minCDE* mutants grow with a similar generation time to the wildtype (Cheng *et al*., 2006). The data for B117 and B118 regions suggested that an essential gene lay in the 14.4 Kb 1,307,905 - 1,322,226 region and consistent with this interpretation, we found that strains carrying the FRT-flanked regions B158 (1,323,078-1,307,752 nt), B159 (1,323,805-1,307,752 nt) and B160 (1,333,550- 1,307,752 nt) failed to yield pTH1944 transconjugants. This region carries a unique copy of a tRNA specific for CCG, the second most frequently used arginine codon. Hence further experiments were performed to demonstrate that the arginine tRNA was the essential gene in the 1,307,905 - 1,322,226 region. A map of this region is shown in Figure 4.7. To confirm that the Arg-tRNA gene was essential for *S. meliloti* cell viability, we attempted to delete the Arg-tRNA gene region from pSymB in a strain carrying Arg-tRNA *in trans* on another plasmid. A 459 bp fragment carrying the

tRNA was amplified and cloned into the replicated plasmid Sp^r pBBR derivative (pTH2302) which was then transferred to RmP896 (B158). Subsequent transfer of pTH1944 (*flp* vector) to this strain generated viable transconjugants carrying ΔB158 deletions. Hence the only essential gene in the 1,307,905-1,528,150 - B117 region is the Arg-tRNA gene. Presence of deletion ΔB158 within the Arg-tRNA gene region from pSymB in a strain RmP896 carrying Arg-tRNA *in trans* on another plasmid was checked by PCR using four sets of primers: 3486UFUR; 3486DFDR; 422UFUR; 422DFDR (Table 2.1).

Carbon and Nitrogen Utilization Phenotypes associated with pSymB deletions

Fucose utilization

In previous work, fucose was found to induce expression of four distinct ABC-transport gene clusters (see Table 1 Mauchline *et al*., 2006). One of these gene clusters, *smb21103* - *smb21106* (744973 – 749228 nt) is located in a region which has been removed from strains carrying the deletion ΔB146, ΔB147, ΔB148 and ΔB139 deletion mutations. We tested these deletions for growth on D-(+)-fucose and found that unlike the wildtype RmP110, all the deletion strains failed to grow with fucose as sole carbon source (Figure 4.8). Thus the *smb21103* - *smb21106* cluster and likely genes downstream of *smb21106* appears to be responsible for the transport and metabolism of D-(+)-fucose.

Hydroxy-proline utilization

In a previous work MacLean *et al*., (2009). showed that putative uptake system *hypMNPQ* (*smb20263-smb20266*) is required for hydroxyproline transport in *S. meliloti*. In this work it was observed that deletion strains ΔB141 (100,636 – 467,160 nt) and ΔB142 (121,311 – 467,160 nt) were unable to metabolize both trans-4-hydroxy-L-proline and cis-4-hydroxy-D-proline, whereas wild-type *S. meliloti* was able to utilize both of these compounds as sole carbon and nitrogen sources (Figure 4.8). This result is consistant with the deletion of the *hypMNPQ* (268,518–271,437 nt) and perhaps other genes required for hydroxyproline catabolism.

Maltotriose, Maltitol, D-(+)-Trehalose and Palatinose utilization

D-(+)-trehalose induces expression of the ABC-transport gene cluster *thuREFGK* (s*mb20325 – smb20328*) (328,098 – 332,311 nt) (see Table 1 Mauchline *et al*., 2006). This gene cluster is known to be involved primarily in the uptake of D-(+)-trehalose and maltose and only to a lesser extent in sucrose uptake (Jensen *et al*., 2002). In this work, deletion strains ΔB141 (100,636 – 467,160 nt) and ΔB142 (121,311 – 467,160 nt) failed to grow on minimal medium with maltotriose, maltitol and palatinose as sole carbon source while growth on D-(+)-trehalose was dramatically reduced (Figure 4.9). It is known from work of Jensen *et al*., (2002) that another trehalose transport system (757,756 – 764,331 nt) is present on chromosome of *S. meliloti* and that can explain why

those deletions were able to grow on trehalose to the some extent. To further specify genes responsible for maltotriose, maltitol, palatinose and D-(+)-trehalose utilization strain RmP2680 carrying a *thuE* plasmid insertion was made. This strain failed to grown on M9 medium with maltotriose, maltitol and palatinose as sole carbon source, but formed small colonies on D-(+)-trehalose This demonstrates that genes *thuEFGK* which are removed in deletions ΔB141 and ΔB142 are involved in maltotriose, maltitol, palatinose and D-(+)-trehalose utilization.

D-(+)-Raffinose and α-D-(+)-Melibiose utilization

The 27 Kb ΔB124 deletion (1,652,558 –1,679,723 nt) includes the ABC-transport cluster *smb21644-smb21647* (1,674,089-1,678,437 nt) reported to be highly induced in minimal medium with D-(+)-raffinose, α-D-(+)-melibiose, galactose, D-(+)-galactosamine, dulcitol and stachyose (Table 1, Mauchline *et al*., 2006). We found that the ΔB124 strain grew like the wildtype on M9-medium supplemented with galactose, D-(+)-galactosamine, dulcitol and stachyose as carbon sources but failed to grow with D-(+)-raffinose and α-D-(+)-melibiose as carbon source (Figure 4.10). Simultaneously, library fusion strain FL357 was checked for growth with D-(+)-raffinose and α-D-(+)-meliobiose and it was shown that integration within the ABC-transport cluster *smb21644-smb21647* enabled strain FL357 to utilize D-(+)-raffinose and α-D-(+)-melibiose. This confirmed that the genes responsible for D-(+)-raffinose and α-D-(+)-melibiose utilization are within the region covered by ΔB124 (Gage *et al*., 1998.).

Thiamine auxotrophs

The deletion strains ΔB116 (1,255,032 – 1,308,912 nt) and ΔB123 (1,528,150 – 1,654,191 nt) are both thiamine auxotrophs which is consistant with the location of thiamine biosynthetic genes (*thiD, thiC, thiO, thiS, thiG* and *thiE*) (Finan *et al.* 1986). The ΔB116 deletion includes the region where the *thiD* (1,191,117 – 1,191,917 nt) gene is located, while ΔB123 includes the region of the *thiC, thiO, thiS, thiG* and *thiE* (1,633,944 – 1,638,330 nt) genes.

Slow growth phenotype

The deletion strains ΔB103 (830,108 - 788,035 nt), ΔB132 (814,652 – 870,505 nt), ΔB133 (794,204 - 870,505 nt**)**, ΔB134 (779,439 – 870,505 nt), ΔB135 (769,971 – 870,505 nt**)**, ΔB136 (762,942 – 870,505 nt), ΔB137 (757,101 – 870,505 nt) and ΔB139 (741,497 - 870,505 nt) show slower growth than the other deletions or wild type strain RmP110. Slower growth of deletions is caused by prolonged *lag* phase of bacterial growth. We showed that the ΔB133 (RmP863) had a prolonged *lag* growth phase compared to the double integrant strain RmP862 from which has been derived and wild type RmP110 (Figure 4.11). Slow growth deletion strains overlap a 14.6 Kb region within which lie eleven genes: *smb21246*, *smb21247*, *smb21248*, *smb21249*, *smb21250*, s*mb21251*, *smb21252*, *smb21253*, *smb21254*, *smb21255* and *smb21256*. The growth phenotypes of strains carrying plasmid single cross-over insertions in each of eleven

genes were examined to identify the specific gene/genes responsible for slow growth phenotype. Insertion of suicide plasmid pTH1703 into *smb21254* (824,431 – 825,261 nt) showed the same slow growth phenotype as the ΔB103-ΔB139 deletion strains. To confirm that *smb21254* was responsible for *S. meliloti* slower growth phenotype, we attempted also to delete the *smb21254* gene region from pSymB in a strain carrying *smb21254 in trans* on another plasmid. To provide *smb21254* expression *in trans*, the *smb21254* gene was cloned downstream of the *lac* promoter of broad host range Sp^r pBBR derivative to give pTH2746. The resulting plasmid, pTH2746, was transferred to strains RmP860 (B132) and RmP862 (B133). Wild type growth phenotype was recovered when the Flp recombinase plasmid pTH1944-Tc^r was transferred into the RmP860 (pTH2746) and RmP862 (pTH2746) strains. This result again suggests that *smb21254* is the gene whose lost is responsible for slow growth. Gene *smb21254* codes for a 276 amino acid protein, which is annotated as conserved hypothetical. Our database search revealed homology to the beta-lactamase domain-containing proteins and to related proteins of various microorganisms (data not shown). We note that in the above experiments IPTG was added to the media to insure induction of *flp* recombinase from the *lac* promoter in the *S. meliloti* cells carrying the pTH2746 (*smb21254*) plasmid. The presence of a deletion within the *smb21254* gene region from pSymB in a strain carrying *smb21254 in trans* on another plasmid was checked by PCR using four sets of primers per deletion strain: 4595UFUR; 4595DFDR; 778UFUR, 778DFDR for strain RmP860 and 3938UFUR; 3938DFDR; 778UFUR, 778DFDR for strain RmP862 (Table 2.1). Only

those deletion derivative strains that were confirmed by PCR are saved as a part of lab strain collection.

α-D-Lactose Monohydrate and Lactulose utilization

Reporter fusions to the gene cluster *smb20929* - *smb20931* (1,152,893 -1,156,054 nt) are highly induced in minimal medium with D-(+)-raffinose, α-D-(+)-melibiose, galactose, α-D-lactose monohydrate and lactulose (Table 1, Mauchline *et al*., 2006). Knowing that the *smb20929* - *smb20931* gene cluster is located within the 41 Kb deletion, ΔB108 (1,129,758 - 1,170,466 nt), we examined the ΔB108 strain for growth on these sugars and found that the ΔB108 strain grew like the wildtype RmP110 on M9-medium supplemented with D-(+)-raffinose, α-D-(+)-melibiose, galactose and α-D-Lactose monohydrate as sole carbon sources but the ΔB108 strain failed to grow with lactulose (4-O-beta-D-Galactopyranosyl-D-fructofuranose) as sole carbon source. A review of the 40 Kb region deleted in ΔB108 suggested that the s*mb20929* - *smb20931* gene cluster maybe required for lactulose utilization in *S. meliloti*.

D-(+)-Galactosamine utilization

Two distinct ABC-transport gene clusters *smb21135-smb21138* (774,396 - 777,407 nt) and *smb21216-smb21221* (779,811 - 783,740 nt) were found to be induced by D-(+)-galactosamine and glucosamine (Mauchline et al., 2006). Both of these gene

clusters are removed in the deletion strain ΔB102 (752,548 - 788,984 nt). The ΔB102 strain and wildtype RmP110 were examined for growth on these sugar amines and results demonstrated that the ΔB102 strain grew like the wildtype on M9-medium supplemented with glucosesamine as sole carbon sources but grew much slower in medium with D-(+)-galactosamine as sole carbon source. This confirmed that the genes responsible for D-(+)-galactosamine utilization are within the region covered by ΔB102 deletion.

Symbiotic phenotypes of pSymB deletions

The symbiotic phenotypes of pSymB deletion strains were assayed by inoculating alfalfa (*Medicago sativa*) seedlings and growing in Leonard assemblies under nitrogen deficient conditions. Plant growth was examined for 28 days post inoculation and shoot dry weights were recorded (Table 4.2). All deletion strains for regions known to be required for effective symbiosis formed nodules that failed to fix N_2, Fix^-. The deletion strain ΔB109 (1,169,073 – 1,207,052 nt) (Figure 4.1) lacking the *exoT*, *exoW*, *exoV*, and *exoU* (1,174,743 – 1,179,394 nt) genes required for exopolysaccharide synthesis (EPS I) which were shown to be required for the formation of Fix^+ nodules in *S meliloti*. (Becker *et al*., 1993). Alfalfa plants inoculated with *dctA* mutants failed to fix nitrogen (Yarosh *et al*., 1989). The deletion strain ΔB123 (1,528,150 – 1,654,191 nt) (Figure 4.1) covers the region with the genes that are involved in C4-dicarboxylate transport (*dctA*, *dctB* and *dctD* (1,626,577 – 1,631,385 nt)) and gave rise to plants with a Fix^- phenotype. The *exoB* (1,168,150 - 1,169,136 nt) gene required for Fix^+ phenotype (Leigh *et al*., 1989; Long *et*

al., 1988) was located within a 40 Kb deletion, ΔB108 (1,129,758 - 1,170,466 nt) (Figure 4.1) and a *S. meliloti* strains carrying this deletion were symbiotically ineffective. In addition, plants inoculated with RmP798 (ΔB122) (Figure 4.1) formed nodules but their N_2-fixing ability appeared delayed relative to wild type nodules.

Fix-delay phenotype on pSymB

Growth of *S. meliloti* strain RmP798 (ΔB122), whose N_2-fixing ability appeared delayed relative to wild type, was monitored in free – living state in three different media; LBmc, M9 – glucose and M9 - succinate. The results presented in Figure 4.12 suggest that the growth rate of deletion strains is comparable to the growth rate of wild type strain RmP110 in all three media indicating that the fix-delay phenotype is not caused by a defect in growth rate of deletion strain.

Plants inoculated with the RmP798 (ΔB122) (1,528,150 – 1,573,735 nt) deletion strain were stunted and yellow in colour at four weeks post inoculation, whereas plants inoculated with wildtype RmP110 were dark green in colour and grew vigorously. Plant shoot dry weights were used as an index of N_2 fixation. The dry weights of plants (averages ± standard errors) inoculated with RmP798 (36.0 ± 21.2 mg/plant) was higher than for the plants inoculated with the Fix^- mutant (10.8 ± 3.1 mg/plant) or the uninoculated control (4.0 ± 0 mg/plant) but lower than shoot dry weight for wild type strain RmP110 (66.9 ± 20.9 mg/plant) (Table 4.2). These results suggest that deletion of the B122 region had an effect on symbiotic N_2 fixation.

The presence of the ΔB122 bacterial strain in the nodules of Fix-delay plants was examined by plating dilutions of the bacterial suspension obtained from nodules on LB medium, followed by selection on LB medium containing 200μg/ml of neomycin and 200μg/ml of gentamicin. Strains reisolated from surface-sterilized nodules retained the expected neomycin, gentamicin antibiotic resistance phenotypes and upon analysis by PCR using four sets of primers 3669UFUR; 3669DFDR; 3347UFUR, 3347DFDR, the nodule's isolates lacked the 1,528,150 – 1,573,735 nt region of the ΔB122 deletion. Five *Rhizobia* cells isolated from the nodules were used for re-inoculation of alfalfa seedlings under nitrogen deficient conditions in Leonard assemblies. For each of five strains used, twenty four plants were examined 28 days post inoculation and they retained the fix-delay phenotype.

To localize the region that appeared to be required for Fix^+ phenotype, nine additional deletions were made within the B122 region. Seven of these deletions (RmP1036, RmP1038, RmP1040, RmP1042, RmP1046, RmP1048 and RmP1050) exhibited the Fix-delay phenotype (Table 4.3). Thus the fix-delay locus appeared to map within 7.5 Kb (1,536,790 - 1,544,317 nt) region (Figure 4.13). In this region *pssF*, *uxuB*, *smb20750*, *smb20751*, *smb20752*, *smb20753*, *smb20754* are located. These genes were annotated as putative glycosyltransferase (*pssF*), putative D-mannonate oxidoreductase (*uxuB*), putative dehydrogense, possibly gluconate 5-dehydrogenase (*smb20750*), putative 3-hydroxyisobutyrate dehydrogenase (s*mb20751*), putative enoyl-CoA hydratase (s*mb20752*), putative acyl-CoA dehydrogenase (*smb20753*) and conserved hypothetical protein (*smb20754*). Each of these seven genes was disrupted by insertion of suicide

plasmid and the resulting derivatives were assayed by inoculating alfalfa seedlings growing under nitrogen deficient conditions. Data obtained in this experiment indicate that loss of the *smb20752* gene resulted in a Fix-delay phenotype. A non polar deletion of *smb20752* was generated in a strain RmP2203 and the same strain was tested by inoculating alfalfa seedlings under nitrogen deficient conditions. Plants were examined 28 days post inoculation and they retained the fix-delay phenotype (Figure 4.14).

A Fix-delay phenotype can be caused by a delay in the process of nodulation or in nitrogen fixation. To investigate these two process causes, the kinetics of nodule formation of the *smb20752* deletion mutant was investigated by performing nodulation assays. Two day old seedlings were transferred to test tubes containing slants of Jensen's medium solidified with 1% agar, inoculated with dilutions of cultures containing the *smb20752* deletion mutant strain, Fix-delay strain ΔB122 or RmP110 wild type strain. Root systems of these plants were examined and scored for nodules every second day. Results were recorded every second day and are presented as a percent of plants nodulated (Figure 4.15) as well as number of nodules formed per plant (Figure 4.15). Nodules were first visible on the roots of alfalfa seedlings 6 days after inoculation with the wild type strain, RmP110. Nodules were visible on seedlings inoculated with the *smb20752* deletion mutant (RmP1230) and Fix-delay strain (RmP798) after a further 2 to 3 days. By 19 days postinoculation with either RmP110 or RmP1230 and RmP798 strain, nodulation as measured by the percentage of plants with visible nodules approached 100%. The average nodule number per plant was similar for all three strains RmP110,

RmP1230 and RmP798. According to these results *smb20752* is not required for the successful process of nodulation.

Expression of all seven genes localized within Fix-delay region was monitored in alfalfa nodules through the use of a transcriptional fusion to the reporter gene *gusA* in *S. meliloti*. As a positive control, alfalfa seedlings were inoculated with strain RmP319 which contains a *gusA* fusion to *nifH*; this gene is known to be expressed in nodules. As negative controls, we used the wild-type strain RmP110 and a strain RmP320 in which *gusA* is integrated into the *S. meliloti* genome in the antisense orientation to a *nifH* promoter. RmP320 was included to reflect the level of β-glucuronidase activity associated with a single copy of the reporter gene in the *S. meliloti* genome. β-Glucuronidase activity was measured in nodule extracts obtained from alfalfa plants 28 days post inoculation. Our results indicate a level of *smb20752* expression that is approximately sevenfold greater (Table 4.4) than that observed in nodules inoculated with the negative controls RmP110 (no *gusA* fusion) or RmP320 (*nifH*::*lacZ*). Nodules harvested from plants inoculated with the positive control (*nifH*::*gusA*) exhibited a higher level of expression (at least ninety fold greater than the controls).

Expression of all seven genes localized within fix-delay region was monitored in free – living state in three different media (LBmc, M9 – glucose and M9 - succinate), through the use of a transcriptional fusion to the reporter gene *gusA* in *S. meliloti*. As a positive control, we used strain RmK990 which contains a *gusA* fusion with *pckA*; this gene is known to be expressed in minimal media with succinate (Osteras *et al*., 1991). As negative controls, we used the wild-type strain RmP110 and a strain RmK991 in which

pckA is fused to the reporter gene *lacZ* in *S. meliloti.* Our results (Table 4.4) indicate a level of *smb20752* expression in all three types of medium that is approximately fivefold greater than that observed with the negative control RmP110 (no *gusA* fusion). The positive control (*pckA*::*gusA*) in minimal medium with succinate exhibited a higher level of expression (at least sixteenfold greater than the control), consistent with previous data obtained regarding expression of this gene in M9 - succinate (Osteras *et al*., 1991).

DISCUSSION

In this study we report the construction of a series of 3-350 kilobase deletions with defined endpoints that cover the entire *S. meliloti* pSymB (1683-kilobase) megaplasmid. We were not able to recover *flp* plasmid transconjugant from four FRT-flanked regions of the megaplasmid. The failure to recover transconjugants in which the *repA1B1C1* region was deleted suggests that the pSymB replicon has only one *oriV* and further that there are genes carried on the rest of the pSymB megaplasmid that are essential for cell survival. This data are in agreement with MacLellan *et al*., 2006., who tested whether the incompatibility region (*inc*) from pSymB, cloned into the broad host range plasmid pBBR1MCS-3, mediated incompatibility in a replicon specific manner and showed essential nature of pSymB megaplasmid. To examine incompatibility MacLellan *et al*. employed a strain of *recA*-deficient *S. meliloti* and demonstrated that the recombinant broad host range plasmid containing the *inc* sequence from pSymB plasmid could not be maintained in a strain that carries pSymB. Inability to obtain *S. meliloti* strain without

pSymB megaplasmid in a presence of broad host range plasmid containing *inc* sequence from pSymB supports our finding that pSymB carries genes that are essential for cell survival. In an earlier deletion analysis of the pSymB megaplasmid by Charles *et al*., 1991 approximately 200 Kb of pSymB was not deleted, leading to speculation that essential genes may reside in undeleted areas.

Using a Flp/FRT deletion strategy two essential genes: $tRNA^{arg}$ and *engA* located on the pSymB megaplasmid were identified. $tRNA^{arg}$ (77 bp) is a unique copy of a tRNA specific for the CCG codon, the second most frequently used arginine codon. The gene for $tRNA^{arg}$ is located on pSymB next to a putative transposase gene (*smb20905*) followed by a 3 Kb region containing eight small ORFs of unknown function (Finan *et al*., 2001). This sequence context indicates possibility that the location of this tRNA gene resulted from a transposition event. In our database search using DNA sequence of *S. meliloti* $tRNA^{arg}$, we found that $tRNA^{arg}$ with 100% identity to *S. meliloti* is present in all members of Rhizobiaceae family; *S. medicae* WSM419 (plasmid pSMED01), *A. tumefaciens* str. C58 circular chromosome, while *Sinorhizobium fredii* NGR234, *Rhizobium leguminosarum bv. trifolii* WSM1325, *Rhizobium leguminosarum bv. viciae* 3841 and *Rhizobium etli* CFN 42 carry $tRNA^{arg}$ on the chromosome. The presence of the $tRNA^{arg}$ gene on the chromosome in a majority of Rhizobiaceae family members allowed us to speculate that the *S. meliloti* $tRNA^{arg}$ gene originated from the chromosome and at some point it was transferred to the pSymB megaplasmid. This idea is supported by the gene organization of the region where $tRNA^{arg}$ lie (Finan *et al*., 2001).This region is abundant with mobile genetic elements such as transposons.

The *engA* gene is annotated as a GTP binding protein in *S. meliloti*. EngA is a broadly conserved bacterial GTPase (a member of the Era subfamily) that lacks a human orthologue and has been shown to be essential for a variety of gram-positive and gram-negative organisms such as *Staphylococcus aureus, Neisseria gonorrhoeae, E. coli* and *B. subtilis* (Caldon *et al.*, 2003). *S. meliloti* EngA is highly similar to hypothetical proteins found in every sequenced genome (data not shown). In our database searches using the EngA sequence, we found that all putative EngA proteins with high identity (> 85%) to *S. meliloti* EngA belong to members of the Rhizobiaceae family (Figure 4.16). The same search revealed that the *S. meliloti engA* gene is located on megaplasmid (pSymB) as well as in *S. medicae* WSM419 (plasmid pSMED01). In *A. tumefaciens* str. C58, the annotated *engA* gene is present on the circular chromosome, while *Sinorhizobium fredii* NGR234, *Rhizobium leguminosarum bv. trifolii* WSM1325, *Rhizobium leguminosarum bv. viciae* 3841 and *Rhizobium etli* CFN 42 carry *engA* gene on the chromosome. EngA is a GTPase with two GTP binding domains arranged in tandem (Caldon *et al.*, 2001). An *engA* mutant of *N. gonorrhoeae* exhibits changes in cellular morphology (Mehr *et al*., 2000). While wild type shows the normal diplococcal shape with one septum-like structure in the center of diplococci, the *engA* mutant has a more rounded cell shape, and often contains more than one septum-like structure per cell (Mehr *et al*., 2000). Morphology changes were also noticed in *engA* mutants of *E. coli* and *B. subtilis* (Hwang and Inouye, 2001; Morimoto *et al*., 2002). The deletion of *engA* was shown to result in filamentous cells with defective chromosomal segregation in *E. coli* (Hwang and Inouye, 2001) and in curved elongated cells with condensed nucleoids in *B. subtilis* (Morimoto *et al*., 2002).

We showed that deletion of *engA*, or disruption of the *engA* gene was lethal, unless *engA* was supplied *in trans* on a replicating plasmid. According to published data by Tomar *et al*., 2009, EngA is likely to play a role in ribosome biogenesis.

Beside two genes *engA* and $tRNA^{arg}$ that are essential in all growth conditions, we identified the *cbtJKL* genes encoding a ABC-type transport system within the B151 region (65,620 - 74,302 nt). These genes are essential for the growth of *S. meliloti* in LB but not in minimal medium. Our deletion analysis showed that pTH1944-*flp*-plasmid transconjugants can be recovered for strain B151 while deletion of 65,620 to 74,302 nt region in B152, B153, B154 and B155 results in lack of transconjugants on LB plates (Figure 4.1). In other work in our laboratory, the *cbtJKL* (*smb20056-20058*) (66,644 – 69,548 nt) gene cluster was shown to encode an ABC-transport system for the uptake of cobalt. *cbtJKL* mutants grew poorly in LB unless the medium was supplemented with $CoCl_2$, while growth in minimal medium was not affected (see *smb20057* Table 8, Cowie *et al.,* 2006). We found that deletions in strains B152, B153, B154 and B155 could only be made if Tc^r transconjugants were selected on LB medium supplemented with 2 ng/ml of cobalt, while growth of deletion strains was not effected on M9-succinate (Figure 4.5).

Deleting large, non-essential regions of the genome allowed us to screen for phenotypes associated with the deleted genes. Strains deleted of known loci such as thiamine biosynthesis (*thi*) (Finan *et al.* 1986), as well as genes involved in dulcitol (*dul*) (White and Finan unpublished), lactose (*lac*) (Jelesko and Leigh, 1994), raffinose (*raf*) and melibiose (*mel*) (Gage and Long, 1998) and hydroxy proline utilization (MacLean *et al*., 2009) had the expected phenotypes. Phenotypic analysis of deletion strains revealed

several novel loci on pSymB. Deletion of the B122 region resulted in a Fix-delay phenotype. Genes required for utilization of D-(+)-fucose**,** maltotriose, maltitol, D-(+)-trehalose, palatinose, lactulose, D-(+)-galactosamine and genes responsible for a slow growth phenotype were also localized to pSymB.

In previous work, fucose (D-(+)-fucose and L-(-)-fucose) was found to induce expression of four distinct ABC-transport gene clusters *smc02772* - *smc02774*, *smb21103* - *smb21106*, *smb21587 – smb21590* and *smb20442 – smb20444* (Mauchline *et al*., 2006). Deletion analysis performed in this study indicates that the *smb21103* - *smb21106* cluster and likely genes downstream of *smb21106* appears to be the only responsible for the transport and metabolism of D-(+)-fucose (6-deoxyhexoses). The conclusion that the transported ligand is fucose is based primarily on the finding that strains in which the *smb21103* - *smb21106* genes were deleted did not grow in M9 medium supplemented with D-(+)-fucose, while wild-type *S. meliloti*, as well as deletions that removed the *smb21587 – smb21590* and *smb20442 – smb20444* genes were able to utilize D-(+)-fucose. D-(+)-fucose is found in a limited number of both Gram-negative and Gram-positive bacteria, in which it is a constituent of the cell wall and capsule structures (He and Liu, 2002; Kren and Martinkova, 2001; Power and Jennings, 2003). In *E. coli* and *Campylobacter jejuni*, the uptake of L-fucose, a source of carbon for microorganisms, is mediated by the major facilitator superfamily MFS proton symporter, *fucP* and *cj0486*, respectively (Dang *et al*., 2010, Stahl *et al*., 2011). Similar gene was not identified in *Roseburia inulinivorans* A2–194 by Scott *et al*., 2006 who identified transcriptional regulators, transport systems, and enzymes involved in fucose utilization using the

shotgun microarray analysis. Scott *et al*., 2006 showed high induction of ABC type of sugar transporter when *R. inulinivorans* A2–194 was grown in medium containing fucose. The Smb21103 - Smb21106 proteins appear to be good candidates for typical a solute binding protein-dependent ABC-type transporter, which upon characterization can become one of the first examples of an ABC system for the transport of D-(+)-fucose.

Glenn and Dilworth, 1981 showed that two types of disaccharide transporters exist in *S. meliloti* WU60, *Rhizobium* bv. tifolii WU420 and *Rhizobium leguminosarum* WU163: one responsible for lactose uptake and the other for uptake of sucrose, maltose and trehalose, but not lactose. Similar results were obtained by Gouffi *et al*., 1998 who showed that disaccharide osmoprotectants are taken up via sucrose – disaccharide transporters that cannot mediate the uptake of disaccharides that are not osmoprotective for the strain, such as lactose, lactulose and melibiose in *S. meliloti* strains. Our results confirmed that osmoprotectants like maltotriose, D-(+)-trehalose and palatinose are all transported via the same ABC transport system *thuEFGK* (s*mb20325 – smb20328*). Furthermore maltitol which is not shown to be osmoprotectant for *S. meliloti* cells shares the same ABC transport system with all osmoprotectants mention previously. In our deletion analysis we disrupted an ABC transport system encoded by *thuEFGK* (s*mb20325 – smb20328*) by insertion of a suicide plasmid and confirmed that, maltotriose, maltitol and palatinose are inclusively taken up via this system while D-(+)-trehalose can be also transported via another trehalose transport system (757,756 – 764,331 nt) present on chromosome of *S. meliloti* (Jensen *et al*., 2002).

Mauchline *et al*., 2006 identified two distinct ABC-transport gene clusters *smb21135-smb21138* and *smb21216-smb21221* that were induced by galactosamine and glucosamine. Deletion analysis showed that the absence of the two gene clusters had an effect on *S. meliloti* growth on M9- medium supplemented with D-(+)-galactosamine. The ΔB102 deletion grew like wild type on M9- medium supplemented with glucoseamine. Binkkotter *et al*., 2000 showed that galactosamine in *E. coli* is transported by specific PEP - dependent carbohydrate phosphotransferase (PTSs) proteins encoded by genes *agaBCD*. Involvement of solute binding protein-dependent ABC-type transporter in galactosamine transport in bacteria has not been yet reported.

Results from the deletion analysis done in this study suggests that deletion of *smb21254* resulted in a slow growth phenotype in *S. meliloti*. The slow growth phenotype was rescued when *smb21254* was supplied *in trans* on replicating plasmid. Gene *smb21254* codes for a 276 amino acid protein, which is annotated as conserved hypothetical. Our database search revealed homology to the beta-lactamase domain-containing proteins and to related proteins of various microorganisms (data not shown). Proteins of high identity (> 80%) to Smb21254 are present in all members of Rhizobiaceae family. Smb21254 protein had high homology to the beta - lactamase domain (> 90%), which was described as part of the beta - lactamase enzymes produced by some bacteria and responsible for their resistance to the beta – lactamase antibiotics like penicillins, cephamycins and carbapenems (Finn *et al*., 2010). We showed that the ΔB133 (RmP863) had a prolonged *lag* phase compared to the double integrant strain RmP862 from which has been derived and wild type RmP110 (Figure 4.11). Extended

lag phase of bacterial growth can be caused by many different factors including the size of the inoculum; time necessary to recover from physical damage or shock in the transfer; time required for synthesis of essential coenzymes or division factors; and time required for synthesis of new (inducible) enzymes that are necessary to metabolize the substrates present in the medium We need to confirm that deletion of *smb21254* has an effect on any of previously mentioned factors which can cause prolonged *lag* phase of bacterial growth.

A new locus *smb20752* responsible for the fix delay phenotype was identified within pSymB region B122. Gene *smb20752* encodes 356 amino acid protein annotated as putative enoyl-CoA hydratase. A domain search revealed presence of the enoyl-CoA hydratase/isomerase domain (residues 19 – 190) within Smb20752 protein (Finn *et al*., 2010). The Fix delay phenotype that was identified could be caused by a delay in nodulation or nitrogen fixation. According to our data the *smb20752* gene is not required for nodulation but deletion or interruption of this gene has an effect on nitrogen fixation. Under the plant growth conditions employed in this study, the *smb20752* gene was moderately expressed (at seven fold higher level than wild type RmP110) compared to highly expressed *nifH* (at least ninety fold greater than the controls) in N_2-fixing bacteroids. Expression of the same gene was followed in free living state and three different medium conditions (LBmc, M9 – glucose and M9 – succinate). In all conditions tested, culture containing an *smb20752*::*gusA* fusion strain yielded an average fourfold greater reporter enzyme activity than that observed in the wild type strain RmP110 (no *gusA* fusion).

In the other work in our laboratory a locus important for survival in a soil was also mapped in the same region B122 (Maclean, unpublished results). Further work has demonstrated that the locus responsible for the Fix - delay phenotype of these deletion mutants is distinct from the locus that characterizes soil phenotype (Maclean, unpublished results).

REFERENCE

Bardin, S.D., T.M. Finan. (1998) Regulation of phosphate assimilation in *Rhizobium* (*Sinorhizobium*) *meliloti*. Genetics. **148**:1689-700.

Becker, A., A. Kleickmann, H. Kuster, M. Keller, W. Arnold, and A. Puhler. (1993). Analysis of the *Rhizobium meliloti* genes *exoU*, *exoV*, *exoW*, *exoT*, and *exoI* involved in exopolysaccharide biosynthesis and nodule invasion: exoU and exoW probably encode glucosyltransferases. Mol. Plant. Microbe. Interact. **6**:735-44.

Brinkkötter A, Klöss H, Alpert C, Lengeler JW. (2000) Pathways for the utilization of N-acetyl-galactosamine and galactosamine in *Escherichia coli*. Mol Microbiol. **37**:125-35.

Caldon, C.E., P. Yoong, P.E. March. (2001) Evolution of a molecular switch: universal bacterial GTPases regulate ribosome function Mol Microbiol; **41**:289-97.

Caldon, C. E., and P. E. March, (2003). Function of the universally conserved bacterial GTPases. Curr. Opin. Microbiol. **6**:135-139.

Chain, P.S.G., I**.**Hernandez-Lucas, B., Golding, and T. M. Finan. (2000). *oriT*-directed cloning of defined large regions from bacterial genomes: identification of the *Sinorhizobium meliloti* pExo megaplasmid replicator region. J. Bacteriol. **182**:5486-5494.

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.

Cheng, J., C. D. Sibley, R. Zaheer, T.M. Finan (2007). A *Sinorhizobium meliloti minE* mutant has an altered morphology and exhibits defects in legume symbiosis. .Microbiology.**153**:375-87.

Cheng J., B. Poduska, R.A. Morton, T.M. Finan. (2011) An ABC-type cobalt transport system is essential for growth of *Sinorhizobium meliloti* at trace metal concentrations. J. Bacteriol.

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

Dang, S., L. Sun, Y. Huang, F. Lu, Y. Liu, H. Gong, J. Wang, N. Yan. (2010) Structure of a fucose transporter in an outward-open conformation. Nature. **467**:734-8.

Finan, T.M., A.M. Hirsch, J.A. Leigh, E. Johansen, G.A. Kuldau, S. Deegan, G.C. Walker, 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., S. Weidner, K. Wong, J. Buhrmester, P. Chain, F. J. Vorhölter, I. Hernandez-Lucas, A. Becker, A. Cowie, J. Gouzy, B. Golding, A. Pühler. (2001) The complete sequence of the 1,683-Kb pSymB megaplasmid from the N_2-fixing endosymbiont *Sinorhizobium meliloti*. Proc.Natl.Acad.Sci.U.S.A. **98**:9889-94.

Finn R.D., J. Mistry, J. Tate, P. Coggill, A. Heger, J.E. Pollington, O.L. Gavin, P. Gunesekaran, G. Ceric, K. Forslund, L. Holm, E.L. Sonnhammer, S.R. Eddy, A. Bateman. (2010) The Pfam protein families datebase. Nucl. Acids Res. Database Issue **38**:D211-222.

Friedman, A.M., S.R. Long, S.E. Brown, W.J. Buikema, F.M. Ausubel. (1982) Construction of a broad host range cosmid cloning vector and its used in the genetic analysis of *Rhizobium* mutants. Gene. **18**:289-96.

Gage, D.J. and S.R. Long. (1998) alpha-Galactoside uptake in *Rhizobium meliloti*: isolation and characterization of agpA, a gene encoding a periplasmic binding protein required for melibiose and raffinose utilization. J Bacteriol.; **180**:5739-48.

Glenn, A.R., M.J. Dilworth. (1981) The uptake and hydrolysis of disaccharides by fast- and slow- growing species of *Rhizobium*. Arch. Microbiol. **129**:233-239.

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

Gouffi, K., N. Pica, V. Pichereau, C. Blanco. (1999) Disaccharides as a new class of nonaccumulated osmoprotectant for *Sinorhizobium meliloti*. Appl Environ Microbiol. **65**:1491-500.

He, X.M., H.W. Liu. (2002) Formation of unusual sugars: mechanistic studies and biosynthetic applications. Annu. Rev. Biochem. **71**:701-54.

Hwang, J., M. Inouye. (2001) An essential GTPase, *der*, containing double GTP-binding domains from *Escherichia coli* and *Thermotoga maritime.* J Biol Chem. **276**:31415-21.

Hynes, M.F., R. Simon, P. Muller, K. Niehaus, M. Labes, A. Puhler. (1986) The two megaplasmids of *Rhizobium meliloti* are involved in the effective nodulation of alfalfa. Mol. Gen. Genet. **202**:356-362.

Jelesko, J.G., J.A. Leigh. (1994) Genetic characterization of a *Rhizobium meliloti* lactose utilization locus. Mol. Microbiol. **11**:165-73.

Jensen, J.B., N.K. Peters, and T.V. Bhuvaneswari. (2002) Redundancy in periplasmic binding protein-dependent transport systems for trehalose, sucrose, and maltose in *Sinorhizobium meliloti***.** J Bacteriol. **184**:2978-86.

Kren, V., L. Martínková. (2001) Glycosides in medicine: "The role of glycosidic residues in biological activity" Curr Med Chem.**8:**1303-28

Leigh, J.A., and C.C. Lee. (1988) Characterization of polysaccharides of *Rhizobium meliloti exo* mutants that form ineffective nodules. J. Bacteriol. **170**:3327-32.

Long, S., J.W. Reed, J. Himawan, and G.C. Walker.(1988) Genetic analysis of a cluster of genes required for synthesis of the calcofluor-binding exopolysaccharide of *Rhizobium meliloti*. J. Bacteriol. **170**:4239-48.

Lowe, R. H., H. J. Evans and S. Ahmed, (1960) The effect of cobalt on the growth of *Rhizobium japonicum*. Biochem. Biophys. Res. Commun. **3**:675-8.

Lowe, R. H., and H. J. Evans, (1962) Cobalt requirement for the growth of Rhizobia. J. Bacteriol. **83**: 210–211.

Maclean, A.M., C.E. White, J.E. Fowler, and T.M. Finan. (2009) Identification of a hydroxyproline transport system in the legume endosymbiont *Sinorhizobium meliloti*. Mol Plant Microbe Interact; **22**:1116-27.

MacLellan, S.R., L.A. Smallbone, C.D. Sibley, and T.M. Finan. (2005) The expression of a novel antisense gene mediates incompatibility within the large repABC family of alpha-proteobacterial plasmids. Mol Microbiol. **55**:611-23.

MacLellan S.R., R. Zaheer, A.L. Sartor, A.M. MacLean, T.M. Finan. (2006) Identification of a megaplasmid centromere reveals genetic structural diversity within the *repABC* family of basic replicons. Mol. Microbiol. **59**:1559-75.

Martens, J.H., H. Barg, M.J. Warren, and D. Jahn. (2002) Microbial production of vitamin B12. Appl Microbiol Biotechnol. **58**:275-85.

Mauchline, T.H., J.E. Fowler, A.K. East, A.L. Sartor, R. Zaheer, A.H.F. Hosie, P.S. Poole, and T.M. Finan. (2006) Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome. Proc.Natl.Acad.Sci.U.S.A. **103**:17933-8.

Mehr I.J., C.D.Long, C.D. Serkin, H.S. Seifert. (2000) A homologue of the recombination-dependent growth gene, *rdgC*, is involved in gonococcal pilin antigenic variation. Genetics. **154**:523-32.

Morimoto, T., P.C. Loh, T. Hirai, K. Asai, K. Kobayashi, S. Moriya, N. Ogasawara. (2002) Six GTP-binding proteins of the Era/Obg family are essential for cell growth in *Bacillus subtilis*. Microbiology. **148**:3539-52.

Osterås M., T.M. Finan, J. Stanley. (1991) Site-directed mutagenesis and DNA sequence of *pckA* of *Rhizobium* NGR234, encoding phosphoenolpyruvate carboxykinase: gluconeogenesis and host-dependent symbiotic phenotype. Mol. Gen. Genet. **230**:257-69.

Power, P.M., M.P. Jennings. (2003) The genetics of glycosylation in Gram-negative bacteria FEMS Microbiol Lett. **218**:211-22.

Scott K**.P.**, J.C. Martin, G. Campbell, C.D. Mayer, H.J. Flint. (2006) Whole-genome transcription profiling reveals genes up-regulated by growth on fucose in the human gut bacterium "*Roseburia inulinivorans*". J**.** Bacteriol. **188**:4340-9.

Stahl M., L.M. Friis, H. Nothaft, X. Liu, J. Li, C.M. Szymanski, A. Stintzi. (2011) L-fucose utilization provides *Campylobacter jejuni* with a competitive advantage Proc. Natl. Acad. Sci. USA. **108**:7194-9.

Tomar S.K., N. Dhimole, M. Chatterjee, B. Prakash. (2009) Distinct GDP/GTP bound states of the tandem G-domains of EngA regulate ribosome binding..Nucleic Acids Res. **37**:2359-70.

Watson, R. J., R. Heys, T. Martin and M. Savard, (2001) *Sinorhizobium meliloti* cells require biotin and either cobalt or methionine for growth. Appl. Environ. Microbiol. **67**:3767-70.

Yarosh, O.K., T.C. Charles, and T.M. Finan.(1989) Analysis of C4-dicarboxylate transport genes in *Rhizobium meliloti*. Mol. Microbiol. **3**:813-23.

Table 4.1. Primer sequences used for construction of pTH2172 – pTH2185 vectors as well as locations of primers on pSymB megaplasmid. All plasmids are pTH1703 derivatives.

Plasmid name	Gene	Primer name	Primer sequence	Location of primers on pSymB	PCR product size
pTH2172	pssF-KO	pssFNsiI	TGCATGCATACGCAGAGCGATGCAAGACG	1536749nt	716bp
		pssFBglII	GAAGATCTCGCGAAGGTCGAAGTCATCG	1537465nt	
pTH2173	pssF-F	pssFfNsiI	TGCATGCATCAGATGAACGTCTCGCTGCG	1535568nt	1107bp
		pssFfXhoI	CCGCTCGAGGCAAACTGGATCTCGCACCG	1536675nt	
pTH2174	uxuB-KO	uxuBNsiI	TGCATGCATTGACGATCACCGAGAAGGGC	1538446nt	549bp
		uxuBBglII	GAAGATCTGCGAGATAGCCGAGATAGGC	1538995nt	
pTH2175	uxuB-F	uxuBfBglI	GAAGATCTTGGAAGCGTTGGTCGAAGGG	1539037nt	941bp
		uxuBfNsiI	TGCATGCATCGAGGTCACGAAGATCAGCC	1539978nt	
pTH2176	*smb20750*	y50NsiI	TGCATGCATGGCCAAGGTCTTGAAGAGGC	1539676nt	505bp
	-KO	y50BglII	GAAGATCTATGCGGATGTCCACGAAGGC	1540181nt	
pTH2177	*smb20750*	y50fNsiI	TGCATGCATGGCCAAGGTCTTGAAGAGGC	1539676nt	1579bp
	-F	y50fXhoI	CCGCTCGAGAGTTGAGGAGGAGAGGCAC	1541255nt	
pTH2178	*smb20751*	y51NsiI	TGCATGCATTGCGTCATCACCATGCTGCC	1540506nt	552bp
	-KO	y51XhoI	CCGCTCGAGCTTTCAGCATGAGGCTCGCG	1541058nt	
pTH2179	*smb20751*	y51fNsiI	TGCATGCATTGCGTCATCACCATGCTGCC	1539555nt	951bp
	-F	y51fXhoI	CCGCTCGAGGGCCATGTTCGATCTCAGCG	1540506nt	
pTH2180	*smb20752*	y52NsiI	TGCATGCATAGATGCAGACCACGCTTCCG	1541626nt	703bp
	-KO	y52BglII	GAAGATCTTGAAGGCGAAGCAACGGTCG	1542329nt	
pTH2181	*smb20752*	y52fBglII	GAAGATCTGTCTCAAGCTGACGCTGTCG	1540831nt	701bp
	-F	y52fNsiI	TGCATGCATTCCGACCATGAAGGTCAGCG	1541532nt	
pTH2182	*smb20753*	y53BglII	GAAGATCTCTCACGATGGACGTTCTGGC	1542450nt	718bb
	-KO	y53NsiI	TGCATGCATCGGACAATCTTCTCGACGCC	1543168nt	
pTH2183	*smb20753*	y53fBglII	GAAGATCTTGTCGCCAATGACGCACTGC	1541770nt	754bp
	-F	y53fNsiI	TGCATGCATCGATCATCTCGGAGGGAACG	1542524nt	
pTH2184	s*mb20754*	y54BglII	GAAGATCTGGATCTGAGGGATGGCAAGC	1544334nt	670bp
	-KO	y54NsiI	TGCATGCATTCTGGGCGATCTGCAATGCG	1545004nt	
pTH2185	*smb20754*	y54fBglII	GAAGATCTTGGAGCAGGTCTGTCATCGG	1543155nt	983bp
	-F	y54fXhoI	CCGCTCGAGACGTCCATCGTGAGAAGCGG	1544138nt	

Figure 4.1. Circular map of pSymB megaplasmid of *S. meliloti*, showing FRT-flanked regions, defined on pSymB (B) generated in this study. Green regions – Flp-transconjugants recovered and deletions (Δ) confirmed. Red regions – Flp-transconjugants not recovered.

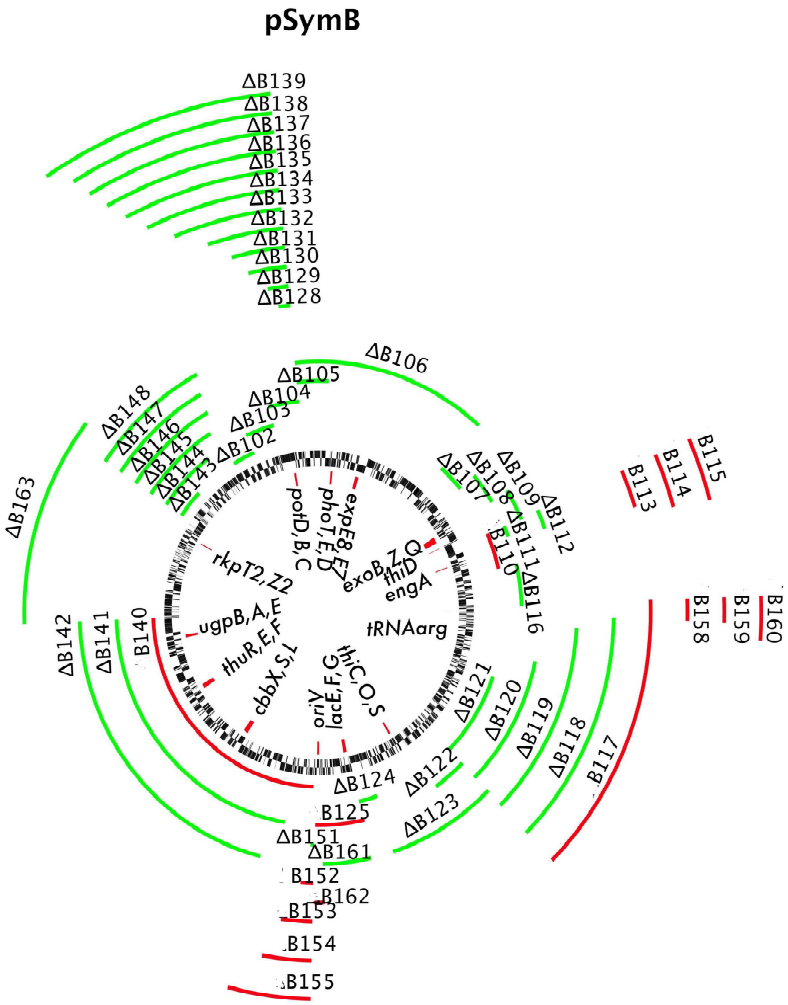

pSymB
ΔB139
ΔB138
ΔB137
ΔB136
ΔB135
ΔB134
ΔB133
ΔB132
ΔB131
ΔB130
ΔB129
ΔB128
ΔB106
ΔB105
ΔB104
ΔB103
ΔB102
ΔB148
ΔB147
ΔB146
ΔB145
ΔB144
ΔB143
ΔB163
ΔB107
ΔB108
ΔB109
ΔB112
B113
B114
B115
B110
ΔB111
ΔB116
B158
B159
B160
ΔB142
ΔB141
B140
ΔB121
ΔB120
ΔB119
ΔB118
B117
ΔB122
ΔB123
ΔB124
B125
ΔB151
ΔB161
B152
B162
B153
B154
B155
potD,B,C
phoT,E,D
expE8,E7
exoB,Z,Q
thiD
engA
tRNAarg
rkpT2,Z2
ugpB,A,E,F
thuR,E,F
cbbX,S,I
oriV
lacI
thiC,O,S,G,F,E

Figure 4.2. Schematic outline of PCR-amplification method used to examine deletion structure. Primer pairs P1 and P4 amplified DNA fragments on either side of the targeted FRT-flanked region. Primer pairs P2 and P3 amplified DNA fragments inside the FRT-targeted region. Amplified products were separated and visualized by agarose gel electrophoresis. Lanes 5, 6, 7 and 8, DNA from wildtype strain RmP110; Lanes 1, 2, 3 and 4, DNA from the ΔB123 strain; M molecular weight marker (GeneRuler™ 1 Kb DNA Ladder, Fermentas). Upper band in lanes1-8 is DNA fragment from the chromosomal region of wildtype strain RmP110 amplified with primer pair P5.

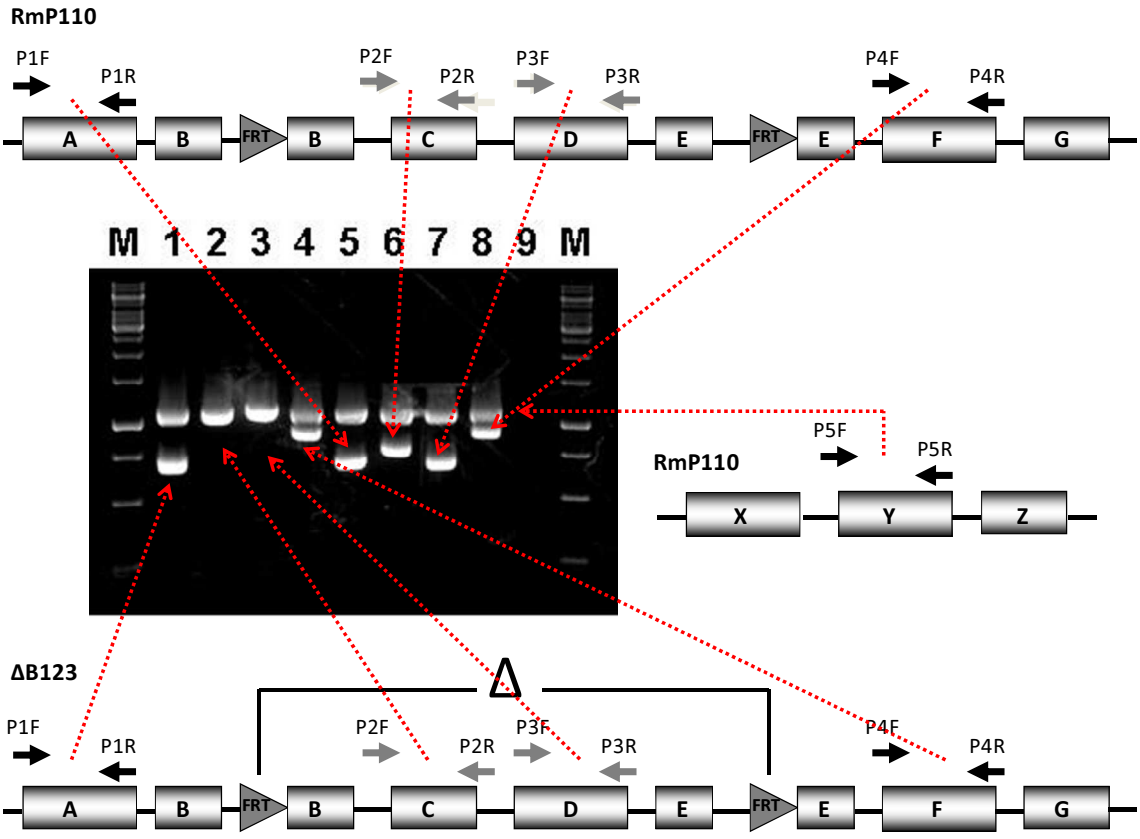

RmP110
P1F
P1R
P2F
P2R
P3F
P3R
P4F
P4R
A
B
FRT
B
C
D
E
FRT
E
F
G
M 1 2 3 4 5 6 7 8 9 M
P5F
P5R
RmP110
X
Y
Z
ΔB123
Δ
P1F
P1R
P2F
P2R
P3F
P3R
P4F
P4R
A
B
FRT
B
C
D
E
FRT
E
F
G

Figure 4.3. The region 49,523 - 60,148 nt on pSymB which harbor origin of replication. Genes present in this region are annotated as: *y20039*-GntR-family transcritional regulator containing an aminotransferase domain; *y20040*-hypothetical protein transmembrane; *y20041*-hypothetical protein; *y20042*-conserved hypothetical protein; *y20043*-hypothetical protein; r*epC1*-probable replication protein C; *repB1*-probable replication protein B; *repA1*-probable replication protein A; *y20047*-hypothetical protein; *y20048*-probable transcriptional regulator; *fusA2*-putative elongation factor G.

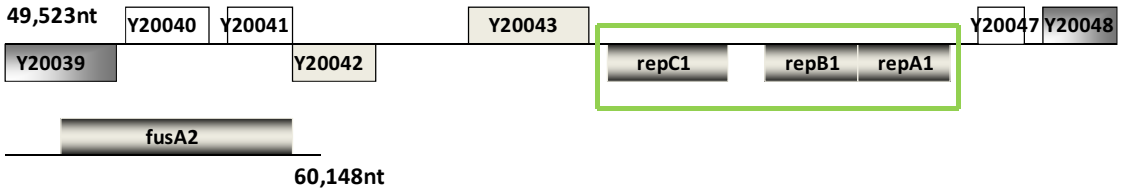
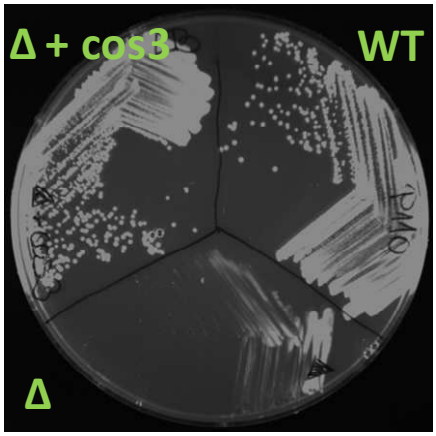

49,523nt
Y20040
Y20041
Y20043
Y20047
Y20048
Y20039
Y20042
repC1
repB1
repA1
fusA2
60,148nt

Figure 4.4. A 8.6 Kb region on the pSymB megaplasmid which harbor an ABC transport system. Genes present in this region are annotated as: *y20056*-ABC transporter, periplasmic solute-binding protein; *y20057*-ABC transporter, permease; *y20058*-ABC transporter, ATP-binding protein; *y20059*-SAM-dependent methyltransferase, may be involved in cobalamin metabolism; TRm5-probable ISRm5 transposase; *y20061*-hypothetical protein; *y20062*-conserved hypothetical protein; *y20063*-conserved hypothetical protein; *y20064*-putative transposase; *y20065*-hypothetical protein; *y20066*-hypothetical protein.

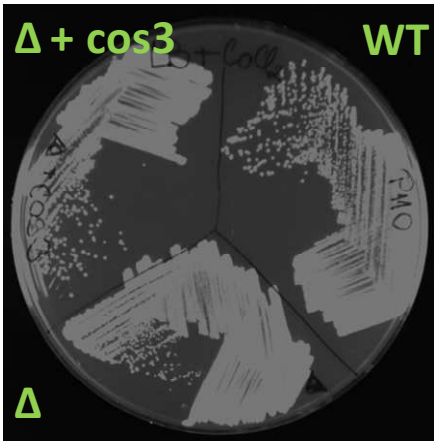

Y20056
Y20057
Y20058
Y20059
TRm5
Y20061
Y20062
Y20063
Y20064
Y20065
65,620
74,302

Figure 4.5. Growth of ΔB152 - Δ, ΔB152 complemented by a cosmid clone pTH2651 – Δ + cos3 and wild type strain RmP110 – P110 on LB medium (A); LB medium supplemented with $CoCl_2$ (2 ng/ml) (B) and minimal medium M9-succinate which contain 2 ng/ml of $CoCl_2$.

A)


LB medium
Δ + cos3
WT
Δ

B)


LB medium + 2ng/ml CoCl2
Δ + cos3
WT
Δ

C)

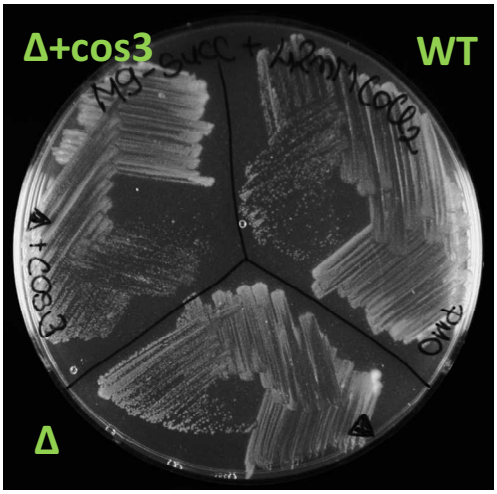

M9-succinate + 2ng/ml CoCl2
Δ+cos3
WT
Δ

Figure 4.6. A 9 Kb region on the pSymB megaplasmid which harbours the *engA* gene. Region cloned in plasmid pTH1996 for the expression analysis of the *engA* promoter – pTH1996. Region cloned in the suicide plasmid to disrupt *engA* function in *S. meliloti* – pTH1997. Genes present in this region are annotated as: *y20992*-conserved hypothetical protein; *y21674*- hypothetical protein; *y20993*-acyl-CoA dehydrogenase; *y20994*-conserved hypothetical protein; *engA*-putative GTP-binding protein; *y20996*-conserved hypothetical protein; *y20997*- conserved hypothetical protein; *y20998*-hypothetical exported protein.

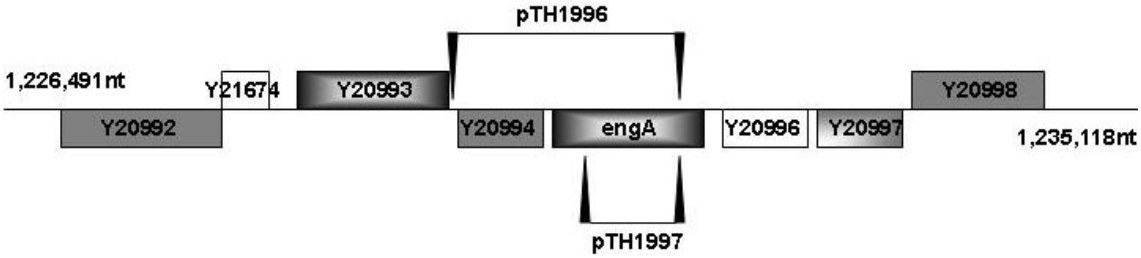

pTH1996
1,226,491nt
Y21674
Y20993
Y20998
Y20992
Y20994
engA
Y20996
Y20997
1,235,118nt
pTH1997

Figure 4.7. A 14.3 Kb region on the pSymB megaplasmid which harbors the arginine tRNA gene. Genes present in this region are annotated as: *y20900*-iguanylate cyclase/phosphodiesterase; *y21679*-hypothetical protein; *y20901*-putative transcriptional regulator; *y20902*-ABC transporter, periplasmic solute-binding protein; *y20903*-ABC sugar transporter, permease component; *y20904*-ABC transporter, ATP-binding protein; arginine tRNA; *y20905*-putative transposase; *y20906*-hypothetical protein; *y20907*-hypothetical protein; *y20908*-hypothetical protein; *y21680*-hypothetical protein; *y21681*-hypothetical protein; *y20909*-hypothetical protein; *y20910*-hypothetical protein; *y20911*-hypothetical protein; *y20912*-putative ATP-dependent DNA ligase; *y20913*-hypothetical protein; *y21682*-hypothetical protein.

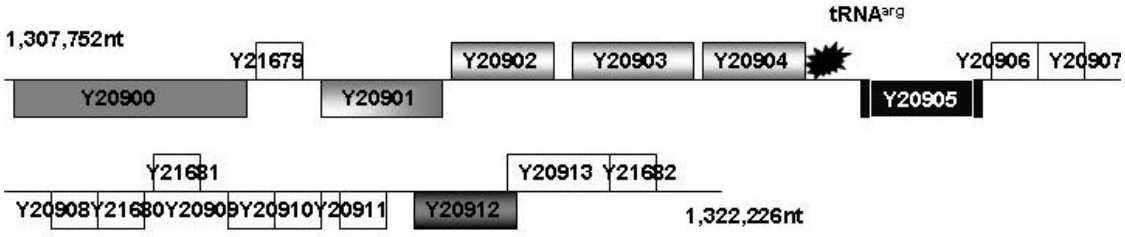
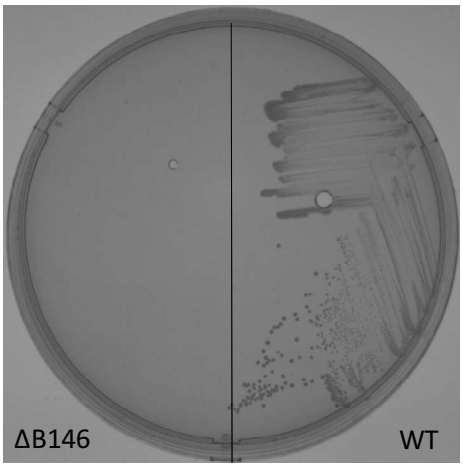

tRNA^arg
1,307,752nt
Y21679
Y20902
Y20903
Y20904
Y20906
Y20907
Y20900
Y20901
Y20905
Y21681
Y20913
Y21682
Y20908
Y21680
Y20909
Y20910
Y20911
Y20912
1,322,226nt

Figure 4.8. A) Growth of deletion strain ΔB146 and wild type strain RmP110 on D-(+)-fucose as sole carbon source. B) Growth of deletion strain ΔB141 and wild type strain RmP110 on trans-4-hydroxy-L-proline as sole carbon source.

A)

D – (+) fucose as a sole carbon source



B)

trans-4-hydroxy-L-proline as sole carbon sources

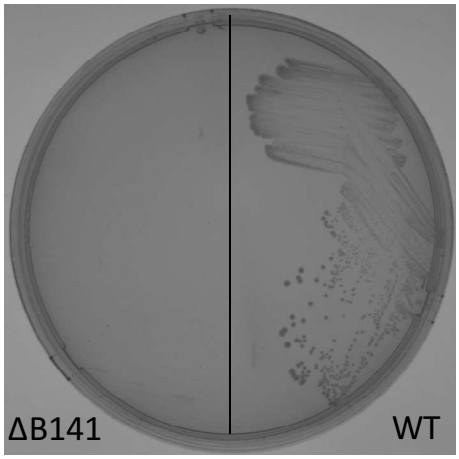
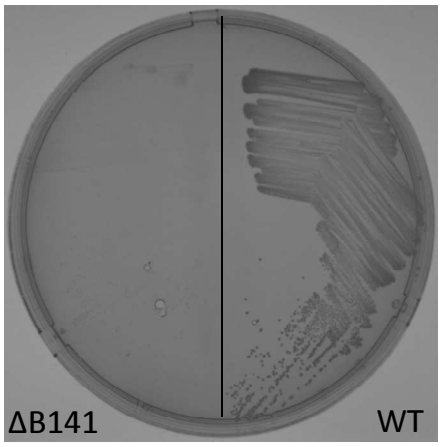


Figure 4.9. Growth of deletion strain ΔB141 and wild type strain RmP110 on maltotriose (A), maltitol (B) and palatinose (C) as sole carbon source.

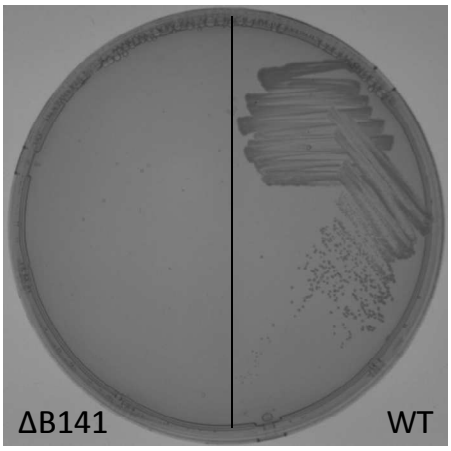
A)

Maltotriose as a sole carbon source



B)

Maltitol as a sole carbon source



C)

Palatinose as a sole carbon source

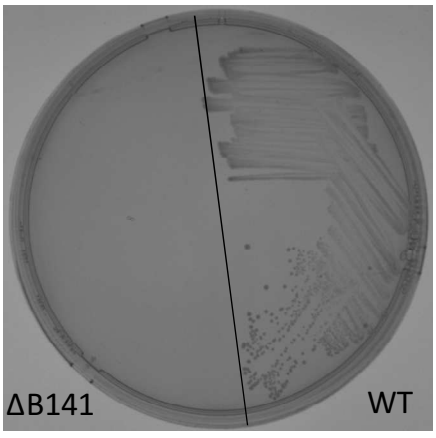
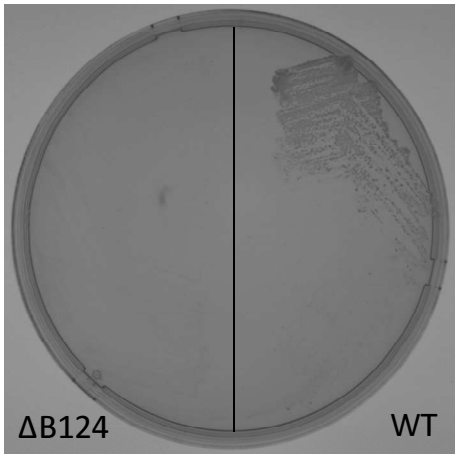


Figure 4.10. Growth of deletion strain ΔB124 and wild type strain RmP110 on D-(+)-raffinose (A) and α-D-(+)-melibiose (B) as sole carbon source.

Raffinose as a sole carbon source



Melibiose as a sole carbon source

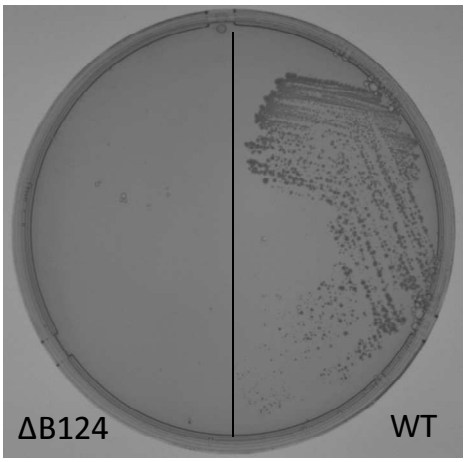


Figure 4.11. Growth of deletion strain RmP863 (ΔB133), double integrant strain RmP864 (B133) and RmP110 (wild type) strains in LBmc.

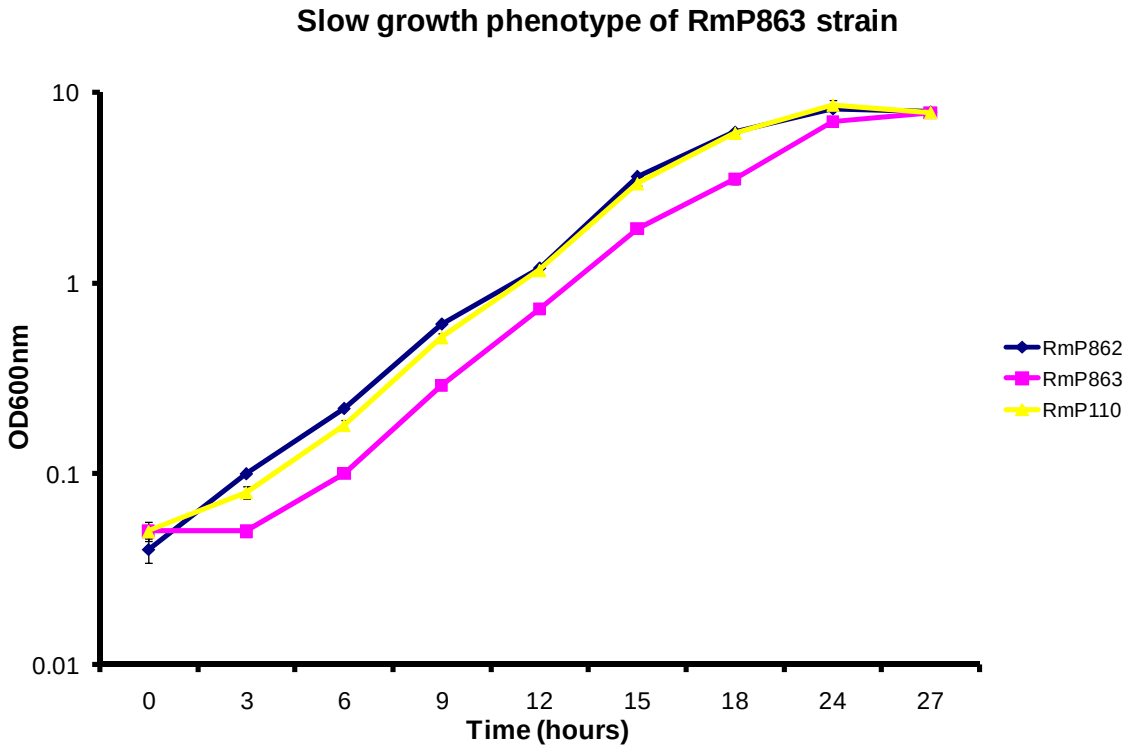

Slow growth phenotype of RmP863 strain
OD600nm
10
1
0.1
0.01
0
3
6
9
12
15
18
24
27
Time (hours)
RmP862
RmP863
RmP110

Table 4.2. Symbiotic phenotype of pSymB deletions

Strain	Shoot dry weight (avg. mg/plant) ± (SD)	% WT	Symbiotic phenotype
P110	66.9 ± (20.9)	100%	Fix$^+$
uninoculated	4.0 ± (0.0)	6%	Fix$^-$
RmG994 (Fix$^-$)	10.8 ± (3.1)	16%	Fix$^-$
ΔB101	54.0 ± (11.1)	81%	Fix$^+$
ΔB104	79.0 ± (16.5)	118%	Fix$^+$
ΔB105	63.5 ± (0.7)	95%	Fix$^+$
ΔB106	47.3 ± (14.2)	71%	Fix$^+$
ΔB108	13.9 ± (10.0)	21%	Fix$^-$
ΔB109	18.3 ± (7.8)	27%	Fix$^-$
ΔB112	82.0 ± (14.1)	123%	Fix$^+$
ΔB116	51.5 ± (7.8)	77%	Fix$^+$
ΔB118	70.1 ± (5.6)	105%	Fix$^+$
ΔB120	55.7 ± (13.8)	83%	Fix$^+$
ΔB121	61.0 ± (13.6)	91%	Fix$^+$
ΔB122	36.0 ± (21.2)	54%	Fix-delay
ΔB123	6.4 ± (0.1)	10%	Fix$^-$
ΔB124	67.0 ± (2.8)	100%	Fix$^+$
ΔB128	54.5 ± (12.1)	81%	Fix$^+$
ΔB129	78.6 ± (15.0)	118%	Fix$^+$
ΔB130	45.7 ± (9.9)	68%	Fix$^+$
ΔB132	49.0 ± (1.4)	73%	Fix$^+$
ΔB133	45.1 ± (14.3)	67%	Fix$^+$
ΔB134	49.7 ± (7.3)	74%	Fix$^+$
ΔB135	66.0 ± (4.2)	99%	Fix$^+$
ΔB136	49.2 ± (17.2)	73%	Fix$^+$
ΔB139	50.5 ± (4.9)	76%	Fix$^+$
ΔB141	53.8 ± (14.6)	80%	Fix$^+$
ΔB142	44.5 ± (10.2)	66%	Fix$^+$
ΔB146	48.0 ± (11.3)	72%	Fix$^+$
ΔB147	62.0 ± (13.9)	93%	Fix$^+$
ΔB148	62.0 ± (2.8)	93%	Fix$^+$

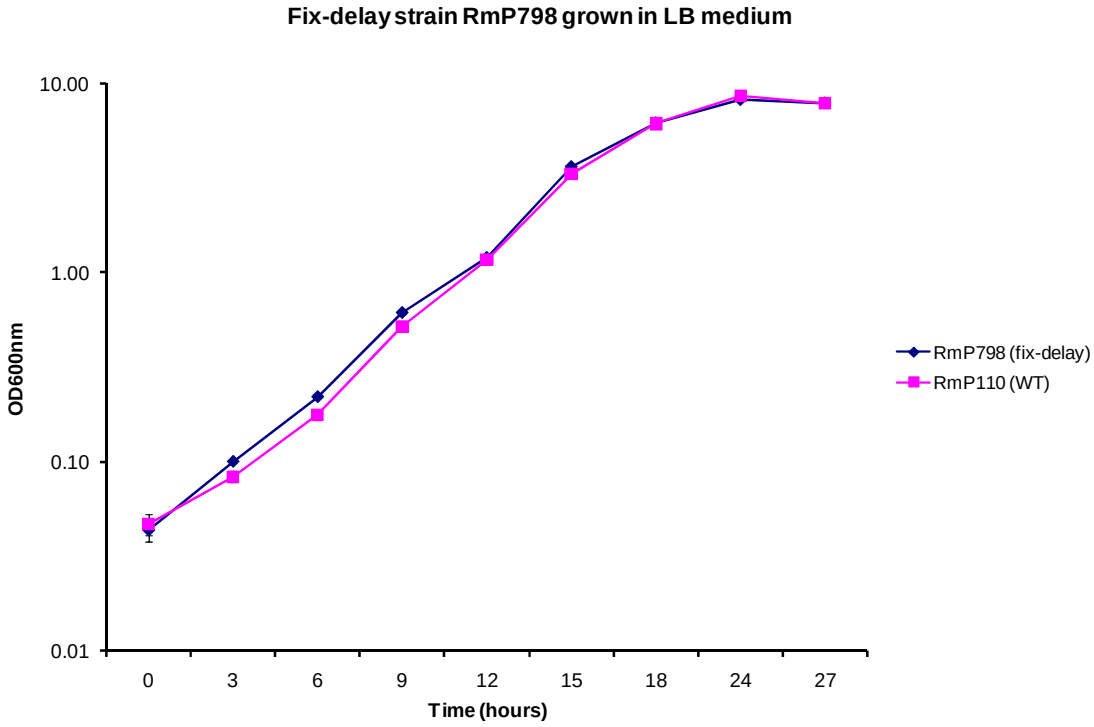
The plant tops were harvested 28 days post inoculation, dried, and shoot dry weights were determined. The values above represent the mean of duplicate samples (each sample consisting of 7-8 shoots) ± standard deviation. The symbiotic phenotype was determined both by examination of the leaves for green (indicating Fix$^+$) or yellow (indicating Fix$^-$) color, and by the shoot dry weight data.

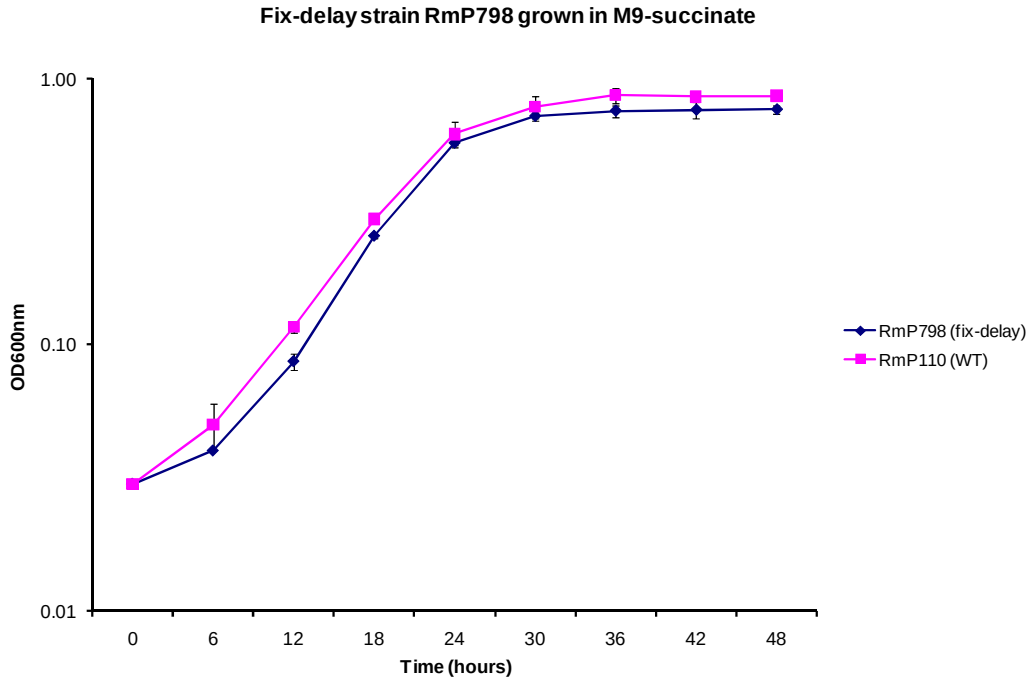
Table 4.3. Symbiotic phenotype of pSymB deletions within region responsible for the Fix delay phenotype.

Strain	Genotype	Shoot dry weight (avg. mg/plant) ± (SD)	% WT	Symbiotic phenotype
RmP11O	wild type	39.5 ± (9.2)	100%	Fix^+
uninoculated		4.6 ± (0.2)	12%	Fix^-
RmG994 (Fix^-)	*dme-3*::Tn5 *tme-4*::ΩSp	5.8 ± (3.2)	15%	Fix^-
RmP1034	ΔB171::ΩNG	50.0 ± (31.1)	127%	Fix^+
RmP1032	ΔB170::ΩNG	48.0 ± (28.3)	122%	Fix^+
RmP1040	ΔB174::ΩNG	15.0 ± (7.1)	38%	Fix - delay
RmP1042	ΔB175::ΩNG	11.5 ± (0.7)	29%	Fix – delay
RmP1036	ΔB172::ΩNG	14.0 ± (5.7)	35%	Fix – delay
RmP1038	ΔB173::ΩNG	11.5 ± (0.7)	29%	Fix – delay
RmP1050	ΔB178::ΩNG	14.1 ± (7.0)	36%	Fix – delay
RmP1048	ΔB177::ΩNG	21.5 ± (4.9)	54%	Fix – delay
RmP1046	ΔB176::ΩNG	14.2 ± (3.8)	36%	Fix – delay
RmP798	ΔB122::ΩNG	10.3 ± (1.1)	26%	Fix - delay

The plant tops were harvested 28 days post inoculation, dried, and dry weights were determined. The values above represent the mean of duplicate samples (each sample consisting of 7-8 shoots) ± standard deviation. The symbiotic phenotype was determined both by examination of the leaves for green (indicating Fix^+) or yellow (indicating Fix^-) color, and by the shoot dry weight data. Δ, deletion; ::ΩGN, deletion is marked by gentamicin (G) and neomycin (N) resistance.

Figure 4.12. Growth of ΔB122 (fix delay) and RmP110 (WT) strains in LBmc, M9 – glucose (15 mM) and M9 – succinate (15 mM).


Fix-delay strain RmP798 grown in LB medium
10.00
1.00
0.10
0.01
OD600nm
0
3
6
9
12
15
18
24
27
Time (hours)
RmP798 (fix-delay)
RmP110 (WT)


Fix-delay strain RmP798 grown in M9-succinate
1.00
0.10
0.01
OD600nm
0
6
12
18
24
30
36
42
48
Time (hours)
RmP798 (fix-delay)
RmP110 (WT)

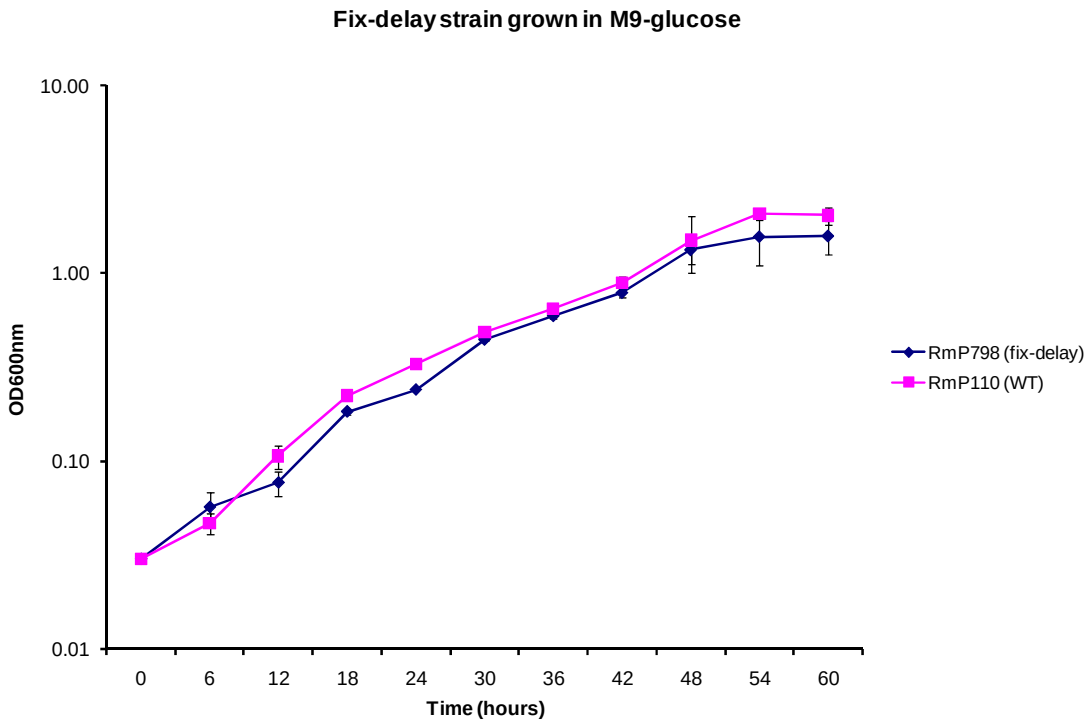

Fix-delay strain grown in M9-glucose
10.00
1.00
0.10
0.01
OD600nm
0
6
12
18
24
30
36
42
48
54
60
Time (hours)
RmP798 (fix-delay)
RmP110 (WT)

Figure 4.13. Delineated 7.5 Kb (1,536,790 - 1,544,317 nt) gene region containing the Fix - locus. Genes present in this region are annotated as: *pssF* - putative glycosyltransferase, *uxuB*-putative D-mannonate oxidoreductase, *smb20750*-putative dehydrogense, possibly gluconate 5-dehydrogenase, s*mb20751*-putative 3-hydroxyisobutyrate dehydrogenase, s*mb20752*-putative enoyl-CoA hydratase, *smb20753*-putative acyl-CoA dehydrogenase and *smb20754*-conserved hypothetical protein (Galibert *et al*., 2001).

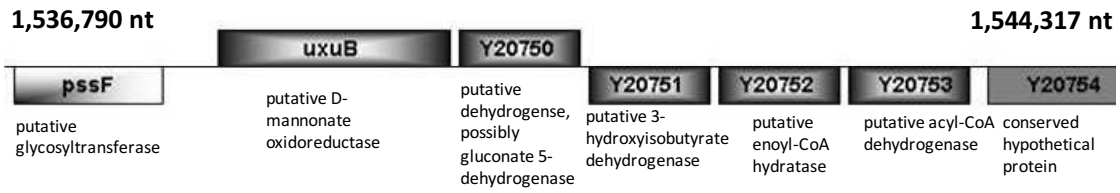

1,536,790 nt
1,544,317 nt
uxuB
Y20750
pssF
Y20751
Y20752
Y20753
Y20754
putative glycosyltransferase
putative D-mannonate oxidoreductase
putative dehydrogense, possibly gluconate 5-dehydrogenase
putative 3-hydroxyisobutyrate dehydrogenase
putative enoyl-CoA hydratase
putative acyl-CoA dehydrogenase
conserved hypothetical protein

Figure 4.14. Fix phenotype of plants grown under nitrogen deficient conditions in Leonard assemblies and examined 28 days post inoculation. Wild type strain RmP110 (A), Fix-delay strain RmP798 (ΔB122) (B), Fix^- strain RmP790 (ΔB108) (C), uninoculated plant (D), RmP2205 (deletion of *smb20753*) (E) and RmP2203 (deletion of *smb20752*).

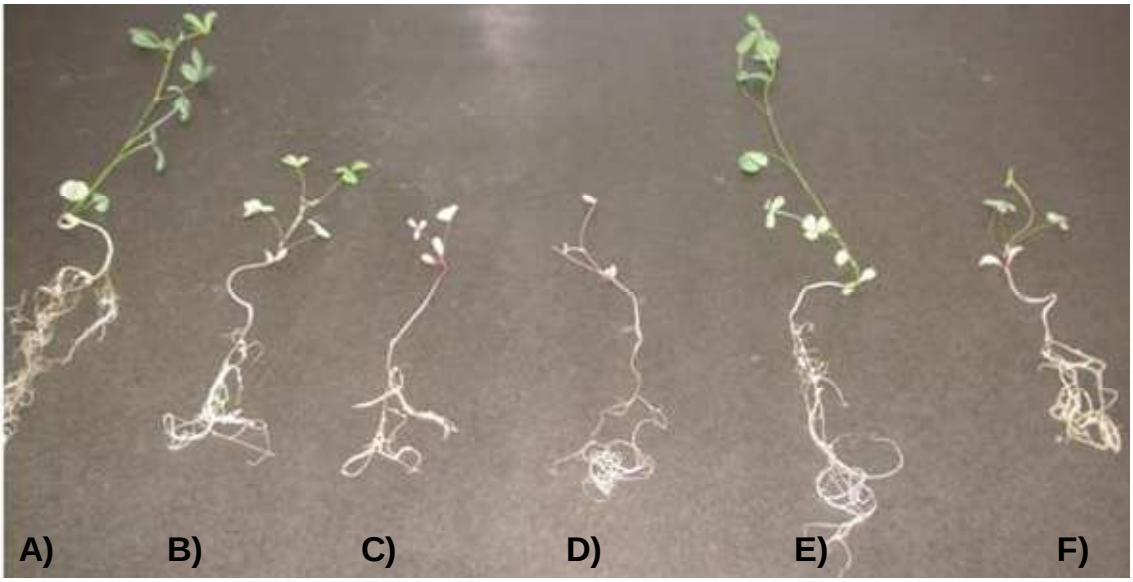
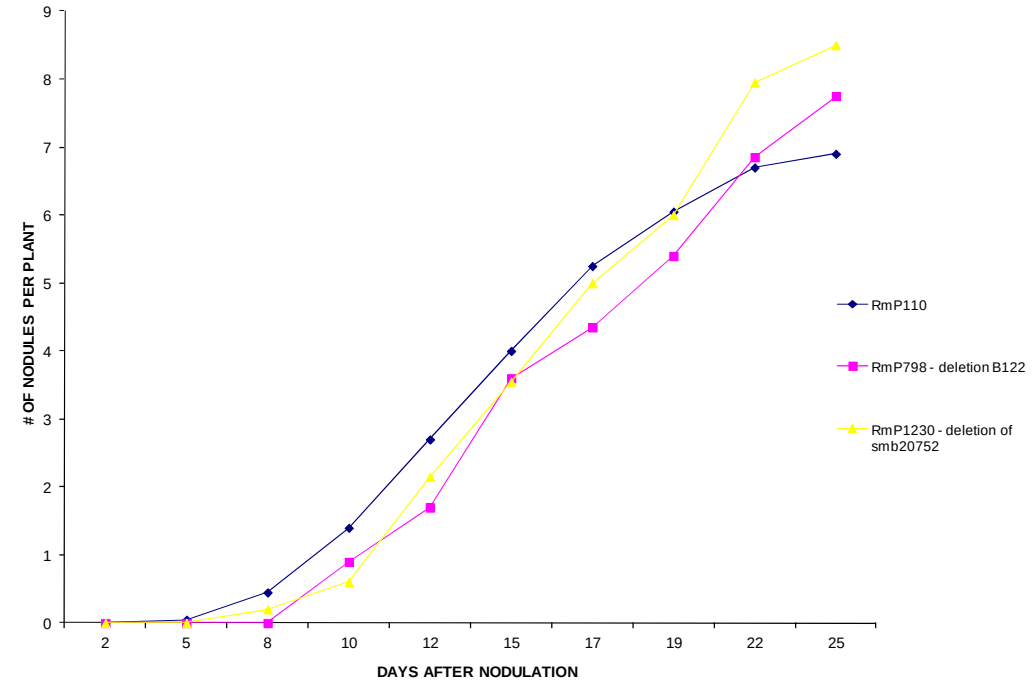

A)
B)
C)
D)
E)
F)

Figure 4.15. Nodulation kinetics for strain RmP798 (ΔB122). A) Comparison of the nodulation kinetics of the RmP1230 (Δ*smb20752*), RmP798 (ΔB122) and the wild-type strain RmP110 on alfalfa seedlings. For each strain, 25 plants were examined every second day for visible nodules. Results were presented as number of nodules formed per plant. B) Comparison of the nodulation kinetics of the RmP1230 (Δ*smb20752*), RmP798 (ΔB122) and the wild-type strain RmP110 on alfalfa seedlings. For each strain, 25 plants were examined every second day for visible nodules. Results were presented as a percentage of plants nodulated every second day

A)


OF NODULES PER PLANT
DAYS AFTER NODULATION
RmP110
RmP798 - deletion B122
RmP1230 - deletion of smb20752

B)

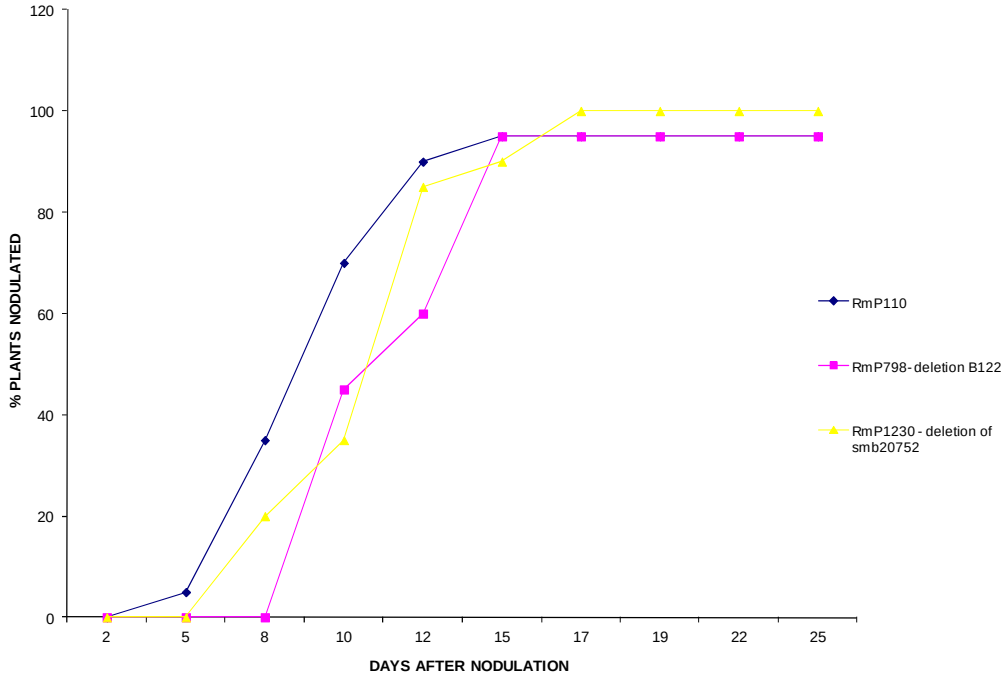

% PLANTS NODULATED
DAYS AFTER NODULATION
RmP110
RmP798- deletion B122
RmP1230 - deletion of smb20752

Table 4.4. Expression of seven genes, localized within Fix-delay region on pSymB megaplasmid, in young alfalfa root nodules and under different free living conditions.

Strain	Gene fusion	Days post inoculation.	β-Glucuronidase activity (SD) Nodules[1]	β-Glucuronidase activity (SD)[a] LBmc	Glucose	Succinate
RmP1223	*pssF* :: *gusA*	28	506 (±14)	477 (±54)	395 (±13)	431 (±90)
RmP1225	*uxuB* :: *gusA*	28	394 (±7)	519 (±72)	658 (±43)	534 (±58)
RmP1227	*smb20750* :: *gusA*	28	130 (±20)	259 (±38)	190 (±9)	209 (±10)
RmP1229	*smb20751* :: *gusA*	28	122 (±8)	252 (±50)	183 (±5)	197 (±22)
RmP1231	*smb20752* :: *gusA*	28	455 (±23)	1039 (±29)	859 (±43)	402 (±76)
RmP1233	*smb20753* :: *gusA*	28	576 (±3)	1108 (±22)	881 (±55)	492 (±52)
RmP1235	*smb20754* :: *gusA*	28	368 (±15)	800 (±17)	464 (±28)	340 (±59)
RmK990	*pckA* :: *gusA*	n.a.	n.a.	847 (±24)	473 (±23)	3118 (±180)
RmPK991	*pckA* :: *lacZ*	n.a.	n.a.	891 (±25)	956 (±39)	818 (±45)
RmP110	none	28	64 (±2)	208 (±28)	178 (±7)	188 (±10)
RmP319	*nifH* :: *gusA*	28	6324 (±124)	n.a.	n.a.	n.a.
RmP320[2]	*nifH* :: *lacZ*	28	85 (±3)	n.a.	n.a.	n.a.

[1] Results are representative of two independent experiments using nodules obtained from a minimum of 14 alfalfa plants inoculated per strain; Data represents mean values of triplicate assays of samples ± standard deviation (SD).
[2] RmP320 contains a *gusA* fusion in the opposite orientation with respect to *PnifH* and serves as a negative control for endogenous β-glucuronidase activity.
[a] Data represents mean values of triplicate assays of samples ± standard deviation (SD).
Miller units of β-glucuronidase activity was determined for cells grown in LBmc, or in minimal media with glucose or succinate as sole carbon source, as indicated.

Figure 4.16. Alignments of the EngA homologues found by BLAST searches.

Alignments of the eleven predicted EngA homologues. The homologues have the following Entrez accession numbers (percentage of identity): *Sinorhizobium meliloti* 1021 – NP_437661 (100%), *Sinorhizobium medicae* WSM419 – YP_001313247 (96%), *Sinorhizobium fredii* NGR234 – YP_002826885 (95%), *Agrobacterium tumefaciens* str. C58 – NP_355257 (85%), *Rhizobium leguminosarum bv. trifolii* WSM1325 – YP_002976910 (86%), *Rhizobium leguminosarum bv. viciae* 3841 – YP_769133 (86%), *Rhizobium etli* CFN 42 – YP_470601 (85%), *Escherichia coli* str. K-12 – NP_417006 (41%), *Neisseria gonorrhoeae* FA 1090 – YP_207579 (36%), *Staphylococcus aureus* subsp. aureus D139 – ZP_06324500 (37%), *Bacillus subtilis* subsp. spizizenii ATCC 6633 – ZP_06875435 (37%).

```
S.meliloti1021                  -MSFTVAIIGRPNVGKSTLFNRLVGKKLALVDDTPGVTRDRRPGDAKLVD 49
S.medicaeWSM419                 -MSFTVAIIGRPNVGKSTLFNRLVGKKLALVDDTPGVTRDRRPGDAKLVD 49
S.frediiNGR234                  -MSFTVAIVGRPNVGKSTLFNRLVGKKLALVDDTPGVTRDRRPGDAKLVD 49
A.tumefaciens.str.C58           -MSFTVAIVGRPNVGKSTLFNRLVGKKLALVDDTPGVTRDRRPGDAKLID 49
R.leguminosarum.bv.trifoliiWSM  -MSFTVAIVGRPNVGKSTLFNRLVGKKLALVDDTPGVTRDRRPGDARLMG 49
R.leguminosarum.bv.viciae3841   -MSFTVAIVGRPNVGKSTLFNRLVGKKLALVDDTPGVTRDRRPGDARLMG 49
R.etliCFN42                     -MSFTVAIVGRPNVGKSTLFNRLVGKKLALVDDTPGVTRDRRPGDARLMG 49
E.coli.str.K-12                 -MVPVVALVGRPNVGKSTLFNRLTRTRDALVADFPGLTRDRKYGRAEIEG 49
N.gonorrhoeae                   -MKPTIALIGRPNVGKSTLFNRLTRTKDALVHDLPGLTRDRHYGHGKVGS 49
S.aureus                        MTKPIVAIVGRPNVGKSTIFNRIVGERVSIVEDTPGVTRDRIYSSGEWLT 50
B.subtilis                      MGKPVVAIVGRPNVGKSTIFNRIAGERISIVEDTPGVTRDRIYSSAEWLN 50
                                     :*::*********:***:.  : ::* * **:****  . ..

S.meliloti1021                  LKFRIIDTAGLEESSPDSLQGRMWAQTEAAIDEADLSLFVVDAKAGLTPA 99
S.medicaeWSM419                 LKFRIIDTAGLEQSSPDSLQGRMWAQTEAAIDEADLSLFVVDAKAGLTPA 99
S.frediiNGR234                  LKFRIIDTAGLEQSAPDSLQGRMWAQTEQAIDEADLSLFVIDAKAGLTPA 99
A.tumefaciens.str.C58           LRFTIIDTAGLEQSGPETLQGRMWAQTEAAIDEADVTLFVIDAKAGLTPA 99
R.leguminosarum.bv.trifoliiWSM  LTFTIIDTAGLEEADAESLQGRMRAQTEAAIDEADLSLFVVDAKSGLTPV 99
R.leguminosarum.bv.viciae3841   LTFTIIDTAGLEEADAESLQGRMRAQTEAAIDEADLSLFVVDAKSGLTPV 99
R.etliCFN42                     LTFTIIDTAGLEEADEESLQGRMRAQTEAAIDEADLSLFVVDAKNGLTPV 99
E.coli.str.K-12                 REFICIDTGGIDGTED-GVETRMAEQSLLAIEEADVVLFMVDARAGLMPA 98
N.gonorrhoeae                   KPYFVIDTGGFEPVVDSGILHEMAKQTLQAVDEADAVVFLVDGRTGLTPQ 99
S.aureus                        HDFNIIDTGGIEIGDA-PFQTQIRAQAEIAIDEADVIIFMVNVREGLTQS 99
B.subtilis                      YDFNLIDTGGIDIGDE-PFLAQIRQQAEIAMDEADVIIFMVNGREGVTAA 99
                                  :  ***.*::      .  .:  *:  *::***  :*::: : *:
```

```
S.meliloti1021                  DQTLAEMLRRRGKPVVVVANKSEARGSEGGFYDAFTLGLGEPCPISAEHG 149
S.medicaeWSM419                 DKTLGEMLRRRGKPVVVVANKSEARGSEGGFYDAFTLGLGEPCPISAEHG 149
S.frediiNGR234                  DETLAEMLRRRGKPVIVVANKSEARGSDGGFYDAFTLGLGEPCPISAEHG 149
A.tumefaciens.str.C58           DETLGEMLRRRGKPVVLVANKSEARGSDAGFYDAFTLGLGEPCPVSAEHG 149
R.leguminosarum.bv.trifoliiWSM  DTDLAEMLRRRGKPVVLVANKSEARGSDSGFYDAYTLGLGEPTPISAEHG 149
R.leguminosarum.bv.viciae3841   DTDLAEMLRRRGKPVVLVANKSEARGSDSGFYDAYTLGLGEPTPISAEHG 149
R.etliCFN42                     DTALAEMLRRRGKPVVLVANKSEARGSDSGFYDAYTLGLGEPTPISAEHG 149
E.coli.str.K-12                 DEAIAKHLRSREKPTFLVANKTDGLDPDQAVVDFYSLGLGEIYPIAASHG 148
N.gonorrhoeae                   DKIIADRLRQSPRPVYLAVNKGEGGDRAVLAAEFYELALGEPHVISGAHG 149
S.aureus                        DEMVAQILYKSKKPVVLAVNKVDNMEMRTDVYDFYSLGFGEPYPISGSHG 149
B.subtilis                      DEEVAKILYRTKKPVVLAVNKLDNTEMRANIYDFYSLGFGEPYPISGTHG 149
                                *  :.. *    :*. :..** :         : : *.:**   ::. **

S.meliloti1021                  QGMLDLRDAIVAALGEERAFPPAEDVAETNVDIRPVAG-EGTEDEEVEPA 198
S.medicaeWSM419                 QGMLDLRDAIVAALGEERAFPPAEDVAETNVDIRADIGSEGPGEEEIEPV 199
S.frediiNGR234                  QGMLDLRDAIVAALGEERAFPPREDVAETDVDIRPAGA-GGGEDEESEPA 198
A.tumefaciens.str.C58           QGMIDLRDAIVEAIGEDMAFPPDVDEAETDIVLPRTEP-GSEEEEDEEPV 198
R.leguminosarum.bv.trifoliiWSM  QGMLDLRDAIVAAIGEDRAYPTKEDVAVTDVDIPQTEGEDSDADE--EPA 197
R.leguminosarum.bv.viciae3841   QGMLDLRDAIVAAIGEDRAYPEKEDVAVTDVDIPQS-SDEGDEDE--EPI 196
R.etliCFN42                     QGMLDLRDAIVEAIGKDRAY-AKEDVAVTDVDIPPSENEADGEDE--EPA 196
E.coli.str.K-12                 RGVLSLLEHVLLPWMEDLAPQEEVDEDAE--YWAQFEAEENGEEEEEDDF 196
N.gonorrhoeae                   DGVYYLIEEILENFPE------------------------PEAEEADAK 174
S.aureus                        LGLGDLLDAVVSHFGE------------------------EEEDPYDE 173
B.subtilis                      LGLGDLLDAVAEHFKN------------------------IPETKYDE 173
                                 *:  * : :    :                          *
```

```
S.meliloti1021                  YDETKPLRVAIVGRPNAGKSTLINRFLGEDRLLTGPEAGITRDSISVEWD 248
S.medicaeWSM419                 YDETKPLRVAIVGRPNAGKSTLINRFLGEDRLLTGPEAGITRDSISVEWD 249
S.frediiNGR234                  YDKTKPLRVAIVGRPNAGKSTLINRFLGEDRLLTGPEAGITRDSISVEWE 248
A.tumefaciens.str.C58           YDETKPLRVAIIGRPNAGKSTLINRFLGEDRLLTGPEAGITRDSISVEWD 248
R.leguminosarum.bv.trifoliiWSM  YDETKPLRVAIVGRPNAGKSTLINRFLGEDRLLTGPEAGITRDSISVEWD 247
R.leguminosarum.bv.viciae3841   YDDTKPLRVAIVGRPNAGKSTLINRFLGEDRLLTGPEAGITRDSISVEWD 246
R.etliCFN42                     YDETKPLRVAIVGRPNAGKSTLINRFLGEDRLLTGPEAGITRDSISVEWD 246
E.coli.str.K-12                 DPQSLPIKLAIVGRPNVGKSTLTNRILGEERVVVYDMPGTTRDSIYIPME 246
N.gonorrhoeae                   HP-----VFAVIGRPNVGKSTLVNAILGEKRVIAFDMAGTTRDSIHIDFE 219
S.aureus                        DT----IRLSIIGRPNVGKSSLVNAILGEDRVIVSNVAGTTRDAIDTEYS 219
B.subtilis                      EV----VQFCLIGRPNVGKSSLVNAMLGEERVIVSNVAGTTRDAVDTAFT 219
                                        .::****.***:* * :***.*::.   .* ***::

S.meliloti1021                  WRGRTIKMFDTAGMRRKAKVQEKLEKLSVADALRAIRFAETVVIVFDATI 298
S.medicaeWSM419                 WRGRTIKMFDTAGMRRKAKVQEKLEKLSVADALRAIRFAETVVIVFDATI 299
S.frediiNGR234                  WRGRTIKMFDTAGMRRKAKVQEKLEKLSVADALRAIRFAETVVIIFDATI 298
A.tumefaciens.str.C58           WRGRTIKMFDTAGMRRKAKVTEKLEKLSVADSLRSIRFAETVVIVFDSTI 298
R.leguminosarum.bv.trifoliiWSM  WRGRTIKMFDTAGMRRKARVIEKLEKLSVADALRAIRFAETVVIVFDATI 297
R.leguminosarum.bv.viciae3841   WRGRTIKMFDTAGMRRKARVIEKLEKLSVADALRAIRFAETVVIVFDATI 296
R.etliCFN42                     WRGRTIKMFDTAGMRRKARVTEKLEKLSVADALRAIRFAETVVIVFDATI 296
E.coli.str.K-12                 RDGREYVLIDTAGVRKRGKITDAVEKFSVIKTLQAIEDANVVMLVIDARE 296
N.gonorrhoeae                   REGKPFTIIDTAGVRRRGKVDEAVEKFSVIKAMQAVEAANVAVLVLDAQQ 269
S.aureus                        YEGQDYVLIDTAGMRKKGKVYESTEKYSVLRALKAIERSNVVLVVIDAEQ 269
B.subtilis                      YNQQEFVIVDTAGMRKKGKVYETTEKYSVLRALKAIDRSEVVAVVLDGEE 269
                                   :   :.****:*::.:: :  ** **  :::::  ::.. :::*.
```

```
S.meliloti1021                  PFEKQDLQIVDLVLREGRAAVLAFNKWDLVENWQALLVDLREKTERLLPQ 348
S.medicaeWSM419                 PFEKQDLQIVDLVIREGRAAVLAFNKWDLVENWQELLADLREKTERLLPQ 349
S.frediiNGR234                  PFEKQDLQIVDLVLREGRAAVLAFNKWDLVEDWQAVLADLREKTERLLPQ 348
A.tumefaciens.str.C58           PFEKQDLQLVDLVIREGRAAVLAFNKWDLVEDPQAYLADLREKTERLLPQ 348
R.leguminosarum.bv.trifoliiWSM  PFEKQDLQIVDLVLREGRAAVLAFNKWDMIEDRQAVLADLREKTDRLLPQ 347
R.leguminosarum.bv.viciae3841   PFEKQDLQIVDLVLREGRAAVLAFNKWDMIEDRQAVLADLREKTDRLLPQ 346
R.etliCFN42                     PFEKQDLQIVDLVLREGRAAVLAFNKWDMIEDRQAVLADLREKTDRLLPQ 346
E.coli.str.K-12                 GISDQDLSLLGFILNSGRSLVIVVNKWDGLSQE--VKEQVKETLDFRLGF 344
N.gonorrhoeae                   DIADQDATIAGFALEAGRALVVAVNKWDGISEE--RREQVKRDISRKLYF 317
S.aureus                        GIIEQDKRVAGYAHEQGKAVVIVVNKWDTVEKDSKTMKKFEDEVRKEFQF 319
B.subtilis                      GIIEQDKRIAGYAHEAGKAVVIVVNKWDAVDKDESTMKEFEENIRDHFQF 319
                                 : .**  : .   . *:: *:.**** :..      ...      :

S.meliloti1021                  ARGIRAVPISGHTGYGLDRLMQAIIETDKVWNRRISTARLNRWLESQQVQ 398
S.medicaeWSM419                 ARGIRAVPISGHTGYGLDRLMQAIIETDKVWNRRISTARLNRWLESQQVQ 399
S.frediiNGR234                  ARGIRAVPISGHTGYGLDRLMQAIIETDKVWNRRISTARLNRWLESQQVQ 398
A.tumefaciens.str.C58           ARGIRAVPMSGQTGYGLDRLMQSIIDTDKTWNRRISTAKLNRWLDAQTTQ 398
R.leguminosarum.bv.trifoliiWSM  ARGIRAVPISGQTGWGLDKLMQAIIDTDKVWNKRISTARLNRWLETQQIQ 397
R.leguminosarum.bv.viciae3841   ARGIRAVPISGQTGWGLDKLMQAIIDTDKVWNKRISTARLNRWLETQQVQ 396
R.etliCFN42                     ARGIRAVPISGQTGWGLDKLMQSIIDTDRVWNKRISTARLNRWLETQQIQ 396
E.coli.str.K-12                 IDFARVHFISALHGSGVGNLFESVREAYDSSTRRVGTSMLTRIMTMAVED 394
N.gonorrhoeae                   LDFAKFHFISALKERGIDGLFESIQAAYNAAMIKMPTPKITRVLQTAVER 367
S.aureus                        LDYAQIAFVSAKERTRLRTLFPYINEASENHKKRVQSSTLNEVVTDAISM 369
B.subtilis                      LDYAPILFMSALTTKRIHTLMPAIIKASENHSLRVQTNVLNDVIMDAVAM 369
                                       :*.     :  *:  :  :      :: :  :.  :
```

```
S.meliloti1021                  HPPPAVSGRRLKLKYMTQVKARPPGFMISCTRPEAVPESYTRYLINGLRN 448
S.medicaeWSM419                 HPPPAVSGRRLKLKYMTQVKARPPGFMISCTRPEAVPESYTRYLINGLRN 449
S.frediiNGR234                  HPPPAVSGRRLKLKYMTQVKARPPGFMISCTRPEAVPESYVRYLINGLRN 448
A.tumefaciens.str.C58           HPPPAVSGRRLKLKYMTQVKARPPAFMISCTRPEAIPESYTRYLVNGLRK 448
R.leguminosarum.bv.trifoliiWSM  HPPPAVSGRRIKLKYMTQVKARPPAFMISCTRSDALPESYTRYLINGLRA 447
R.leguminosarum.bv.viciae3841   HPPPAVSGRRIKLKYMTQVKARPPAFMISCTRSDALPESYTRYLINGLRA 446
R.etliCFN42                     HPPPAVSGRRIKLKYMTQVKARPPAFMISCTRSDALPESYTRYLINGLRA 446
E.coli.str.K-12                 HQPPLVRGRRVKLKYAHAGGYNPPIVVIHGNQVKDLPDSYKRYLMNYFRK 444
N.gonorrhoeae                   QQPPRAGLVRPKMRYAHQGGMNPPVIVVHGNSLHAISDSYTRYLTQTFRK 417
S.aureus                        NPTPTDKGRRLNVFYATQVAIEPPTFVVFVNDVELMHFSYKRYLENQIRA 419
B.subtilis                      NPTPTHNGSRLKIYYATQVSVKPPSFVVFVNDPELMHFSYERFLENRIRD 419
                                : .*     * :: *      .** .::  .  . :  ** *:* : :*

S.meliloti1021                  DFDLPGVPIRVHFRASENPFESKARKRR---------------------- 476
S.medicaeWSM419                 DFDLAGVPIRIHFRASENPYESKARKKR---------------------- 477
S.frediiNGR234                  DFDMPGVPIRVHFRAADNPFESKARKRR---------------------- 476
A.tumefaciens.str.C58           DFDMPGVPIRVHYRGSDNPFESKAKKRR---------------------- 476
R.leguminosarum.bv.trifoliiWSM  DFDMPSVPIRIHFRSPENPFEGKKRRT--------------------- 474
R.leguminosarum.bv.viciae3841   DFDMPSVPIRIHFRSPDNPFEGKKRRT--------------------- 473
R.etliCFN42                     DFDMPSVPIRIHFRSPDNPFESKKKRT--------------------- 473
E.coli.str.K-12                 SLDVMGSPIRIQFKEGENPYANKRNTLTPTQMRKR-------KRLMKHIK 487
N.gonorrhoeae                   AFNLQGTPLRIQYNVSENPYENAEDKPKKKPLRRVSLSNRIEKREGRKEE 467
S.aureus                        AFGFEGTPIHIIARKRN--------------------------------- 436
B.subtilis                      AFGFEGTPIKIFARARK--------------------------------- 436
                                 :.. . *:::  .  .
```

```
S.meliloti1021                  ------------------
S.medicaeWSM419                 ------------------
S.frediiNGR234                  ------------------
A.tumefaciens.str.C58           ------------------
R.leguminosarum.bv.trifoliiWSM  ------------------
R.leguminosarum.bv.viciae3841   ------------------
R.etliCFN42                     ------------------
E.coli.str.K-12                 KNK--------------- 490
N.gonorrhoeae                   KNRFKKKTKVSVKKQFSK 485
S.aureus                        ------------------
B.subtilis                      ------------------
```

CHAPTER 5. DELETION ANALYSIS OF THE *Sinorhizobium meliloti* pSymA MEGAPLASMID

ABSTRACT

The genome of the soil bacterium *Sinorhizobium meliloti* contains 6204 predicted protein-coding regions and approximately 2000 of these encode proteins of unknown function (PUFs). To identify functional significance of *S.meliloti* PUFs, we have deleted large defined regions of the *S. meliloti* genome using the FRT/flp recombination system. We constructed strains with FRT sites flanking 43–166 kilobase regions that cover the entire *S. meliloti* pSymA (1354-kilobase) megaplasmid. Introduction of Flp recombinase into these strains led to the identification of three toxin/antitoxin-like loci (*sma0471/sma0473; sma2105, sma2230/sma2231*) whose deletion resulted in reduced growth or a loss in colony forming ability. The introduction of these genes *in trans* allowed their deletion from pSymA without loss of cell viability. These loci are therefore active toxin/antitoxin-like systems whereas 47 other systems predicted to be located on pSymA can be removed without apparent loss of cell viability.

INTRODUCTION

Twenty eight years ago an operon that controls segregation of plasmid *ccdBA* (coupled cell division) was identified on the F – plasmid of *E. coli* (Ogura and Hiraga, 1983). It was found that *ccdB* acts as a toxin by killing cells that become plasmid free. Since then many operons with a similar organization which are capable of stabilizing plasmids have been identified and named 'toxin-antitoxin (TA) systems or modules'. TA loci are two-component systems that code for a stable toxin and an unstable antitoxin. According to the antitoxin nature TA systems are distinguished in two types: type I and type II. In type I the antitoxin is antisense RNA (Gerdes *et al*., 1986; Greenfield *et al*., 2000) and in type II the antitoxin is a protease sensitive protein (Ogura and Hiraga, 1983; Butts *et al*., 2005; Gerdes *et al*., 2005). Stable toxins are neutralized by unstable antisense RNAs or protein antitoxins. In plasmid free cells degradation of unstable molecules (antisense RNAs and proteins) is rapid and in the case of antitoxin proteins, protein degradation is mediated by the Lon and Clp proteases. At the same time toxic protein or mRNA that will be translated into a toxic protein is present in a plasmid free cell leading to post segregational killing (PSK) or stasis of those cells (Gerdes *et al*., 2005).

Chromosomally encoded TA operons have been identified by a range of bioinformatics techniques and are present in a very wide range of bacteria and archaea (Arcus *et al*., 2011). It has been found that the chromosomal TAs potentially play biological roles associated with growth arrest under conditions of nutritional stress

(Gerdes, 2000), challenge by antibiotics (Sat *et al*., 2001), DNA damage or genomic rearrangement (Hazan *et al*., 2004).

Well characterized bacterial toxin antitoxin loci have been classified into eight families (Gerdes *et al*., 2005) using amino sequence homology as the main criteria (see Chapter 1. Table 1.1). These families are briefly described below.

The *ccd* locus of the F plasmid. The *ccd* (coupled cell division) locus consist of *ccdB* (toxin) and *ccdA* (antitoxin) genes which are used by the F plasmid as a method of ensuring its maintenance in the host cell. The *ccd* locus increases plasmid stability by inhibiting the growth of plasmid free daughter cells (Jaffe *et al*., 1995). Lon protease was shown to degrade CcdA antitoxin and the free toxin CcdB acts on DNA gyrase (an essential topoisomerase II), by trapping the gyrase in its cleavable complex with DNA, eventually causing double strand breaks in the DNA. Chromosomal *ccd* loci are rare.

The *relBE* locus of *Escherichia coli*. The *relBE* locus of *E. coli* encodes an mRNA interferase RelE that cleaves mRNA positioned at the ribosomal A-site (Pedersen *et al*., 2003) and the antitoxin protein RelB counteracts this activity. The *relBE* operon is autoregulated by RelB. The RelBE protein complex represses transcription more efficiently than RelB alone; thus, RelE functions as a co-repressor of transcription (Gotfredsen *et al*., 2003; Overgaard *et al*., 2008). During steady-state cell growth, *relBE* transcription is repressed by the RelBE complex but conditions such as amino acid starvation induce *relBE* transcription and activate RelE. Overexpression of *relE* induces a bacteriostasis from which cells can be rescued. Gerdes *et al*., 2005 showed that *relBE* is a stress response module.

The *parDE* locus of the RK2 plasmid. The 9 kDa ParD protein and the 12 kDa ParE protein function as a post segregational killing (PSK) system, in which the antitoxin ParD protein protects the bacterial cell by forming a complex with the ParE toxin protein (Johnson *et al*., 1996). ParD exists as a dimer and binds to the dimer form of the ParE toxin to form a tetrameric complex (Johnson *et al*., 1996; Oberer *et al*., 1999). ParE inhibits DNA replication *in vitro* and this inhibition results from inactivation of *E. coli* DNA gyrase by the toxin. Furthermore, the inactivation of gyrase by ParE can be reversed by the addition of the antitoxin ParD (Jiang *et al*., 1992). The biological significance of *parDE* was demonstrated in *Vibrio cholerae* by Yuan *et al*., 2011, who shown that *parDE* aids in the maintenance of the integrity of the *V. cholerae* superintegron and in ensuring the inheritance of chromosome II.

The *higBA* locus of the RtsI plasmid. The *higBA* locus was first identified in a temperature-sensitive plasmid (Rts1) of *Proteus vulgaris* (Tian *et al.*, 1996). *higBA* loci differ from other characterized TA loci in that the toxin-encoding gene (*higB*) lies upstream of the antitoxin-encoding gene (*higA*); however, as with other TA loci, it appears that the antitoxin represses transcription of the operon. It was shown by Hurley *et al*., 2009 that HigB associates with the 50 S ribosomal subunit in a complex that cleaves within mRNA coding regions at all AAA triplet sequences.

The *mazEF* locus. The *mazEF* encodes the stable toxin MazF, and the antitoxin MazE which is degraded by the ATP-dependent ClpAP serine protease. MazF is an endoribonuclease which specifically cleaves mRNAs at ACA sequences (Zhang *et al*., 2003) in a ribosome-independent manner. The *mazEF* system, was the first TA system

described as responsible for bacterial programmed cell death. *mazEF*-mediated cell death requires a quorum-sensing factor called EDF (extracellular death factor). Under stressful conditions that inhibit *mazEF* expression, the *de novo* synthesis of both MazE and MazF is prevented: because MazE is much more labile than MazF, the cellular amount of MazE decreases faster than that of MazF, allowing MazF to cause cell death.

The *phd/doc* locus of the P1 plasmid. The *phd* (prevent host death)*/doc* (death on curing) operon encodes a TA module responsible for the maintenance of the plasmid-prophage P1 in *Escherichia coli* (Lehnherr *et al*., 1993). Doc is an inhibitor of translation elongation through its association with the 30 S ribosomal subunit (Liu *et al*., 2008). The C-terminal domain of Phd protein harbors an interaction site with Doc important for the prevention of Doc-mediated growth arrest. The N-terminal region of Phd is a DNA-binding domain that binds to the operator site of the *phd/doc* operon (McKinley *et al.*, 2005). Phd forms a heterotrimeric complex with Doc (Gazit *et al*., 1999).

The *vapBC* locus. The *vapBC* (virulence associated protein) locus was identified on a *Salmonella dublin* virulence plasmid (Pullinger *et al*., 1992). The *vapBC* operons form the largest family of TA systems in prokaryotes. These are grouped together based on their toxin components VapC belonging to the PilT N-terminal domain (PIN domain) family of proteins that are thought to act as ribonucleases. Thus, the *vapBC* operons encode for a toxic ribonuclease whose upstream antitoxin is a transcription factor that binds tightly to the ribonuclease and inhibits its activity. The biological roles of a *vapBC* operon in the pathogenic bacterium *Neisseria gonorrhoeae* have been experimentally investigated (Wilbur *et* al., 2005; Mattison *et* al., 2006). The operon is called *fitAB* (fast

intracellular trafficking locus) and the role of *fitB* (the *vapC* PIN *domain* toxin) is to slow bacterial growth in the intracellular environment. A *vapBC* operon in *Sinorhizobium meliloti* (*ntrPR*) has been characterised, and the *ntrPR* knockout strain shows increased rates of nitrogen fixation and biomass production in the host plant (Bodogai *et al*., 2006). Similar to *fitAB* from *N. gonorrhoeae*, the function of *ntrPR* is to regulate the global level of translation in specific environments (Bodogai *et al*., 2006). In *Leptospira interrogans*, the biological role of *vapBC* is not well defined, but expression of *L. interrogans* VapC in *E. coli* arrests bacterial growth, and this can be rescued by co-expression of the antitoxin VapB (Zhang *et al*., 2004). In addition, in *L. interrogans* VapBC stabilize unstable plasmids, suggesting that this module has characteristics similar to those of other plasmid maintenance TA systems.

The ω-ε-ζ locus of plasmid pSM19035. The ω-ε-ζ locus of plasmid pSM19035 from *Streptococcus pyogenes* encodes three components of a TA system. The ω repressor autoregulates transcription of the ω-ε-ζ operon, while ε and ζ encodes an antitoxin and a toxin, respectively. The biological role of a ω-ε-ζ operon is maintenance of an unstable plasmid in the host cell by killing or preventing the growth of plasmid free cells (Meinhart *et al*., 2003). Similar operon was identified in *Enterococcus faecium* on the plasmid pVEF3 (Sletvold *et al*., 2008).

Very few free-living organisms do not have identified TA loci. Examples include *Lactococcus lactis*, that thrives in rich medium only (milk), the pathogens *Haemophilus ducreyi* 35000HP and *Pasteurella multocida*, one phototroph *Thermosynechococcus elongatus* and the deeply branching thermophile *Maritima thermotoga* (Pandey *et al*.,

2005). Thus, almost all free-living bacteria that live in very different environments have TA loci. *Nitrosomonas europaea* is an obligate chemolithotrophic soil organism with 43 intact TA loci and 2 single toxin genes; *Mycobacterium tuberculosis* (*Mtb*) H37Rv and CDC1551 have 38 and 36 TA loci, respectively, while *Mycobacterium leprae* did not have a single intact TA locus (Pandey *et al*., 2005). *M. leprae* is an obligate intracellular pathogen, whereas *Mtb* has an extra- and an intracellular growth phase. This example supports the notion that obligate host-associated organisms do not retain TA loci while they are beneficial to free-living organisms. A similar pattern was seen in spirochetes: the obligate parasitic spirochetes *Treponema pallidum* and *Borrelia burgdorferi* have no TA loci whereas the free-living spirochete *Leptospira interrogans* has five. In a survey of genome sequences published in 2009, *S. meliloti* was among the top twenty bacteria with a genome highly abundant with TA loci (Makarova *et al*., 2009).

The total length of the *S. meliloti* pSymA megaplasmid is 1,354,226 bp. Eighty four percent of the nucleotide sequence is predicted as protein coding, consisting of 1,293 ORFs (Barnett *et al*., 2001; Beker *et al*., 2009). pSymA proteins linked to nitrogen metabolism are organized in a 53-kb segment encoding a complete pathway for denitrification. Nitrous oxide reductase (*nos*), as well as putative genes for nitrate transport (*nrtAB*) are also located on this replicon (Barnett *et al*., 2001). The pSymA megaplasmid encodes two tRNAs: one with anticodon for methionine for which another copy is present on the chromosome. The other tRNA with a UCA anticodon specifies selenocysteine and is unique in the genome. Housekeeping functions such as: DnaJ/CpbA-like chaperonin, UvrD2, DnaE3, RpoE6, *groESL* operons, DNA ligase, and

DNA-damage inducible protein, are also encoded on pSymA but they are redundant. A number of pSymA ORFs may be involved in stress responses (three cold shock proteins, one heat shock protein, a hydroperoxidase, and haloperoxidases) (Barnett *et al*., 2001). pSymA genes involved in nodulation and nitrogen fixation are clustered within a 275 kb region that includes *nod* genes required for synthesis of Nod factor as well as the *nol* and *noe* genes (Schlaman, 1998). In the same 275 kb region, previously discovered *nif* and *fix* genes for symbiotic nitrogen fixation were localized (Kaminski *et al*., 1988).

A search of the *S. meliloti* chromosome for the presence of toxin-antitoxin loci using standard BLASTP (Pandey *et al*., 2005) identified 12 TA systems (2 *relBE*-like, 3 *higBA*-like, and 7 *vapBC*-like). In 2007 Sevin *et al*. used *S. meliloti* as model for the validation of RASTA-bacteria, a web-based tool for identifying toxin-antitoxin loci in prokaryotes. RASTA bacteria is based on the typical TA loci organization, length of toxin and antitoxin, and the presence of conserved functional domains. The search of the *S. meliloti* genome predicted 25 TA loci on the chromosome, 17 loci on pSymA and 11 on pSymB. 53 TA loci of *S. meliloti* included 95 genes of which 18 were newly identified. RASTA-bacteria identified all TA loci previously predicted by Pandey *et al*. (2005) except for one locus (*higBA*-2, GI: 15965582-15965583). The RASTA method showed a random distribution of TA loci across the various replicons. The genetic organization of the TA loci was diverse: 12 genes occured in 4 triplets, 68 genes in 34 pairs and 15 individual genes (12 antitoxins and 3 toxins).

Makarova *et al*., (2009) did a comprehensive comparative-genomic analysis of Type II toxin-antitoxin systems in prokaryotes. They used sensitive methods for distant

sequence similarity searchs, genome context analysis and an approach for the identification of two-component systems. This search revealed numerous previously unidentified protein families that were homologous to toxins and antitoxins of known type II toxin antitoxin systems. In addition, 12 new families of toxins and 13 families of antitoxins were predicted. Seven hundred fifty complete genomes of archaea and bacteria were analyzed. This analysis revealed 211 TA loci in *S. meliloti*: 113 loci on the chromosome, 50 loci on the pSymA megaplasmid and 48 loci on the pSymB megaplasmid (Table (5.1, 5.2, 5.3)). All predicted loci were classified into known and newly created TA families. All the TA loci that were predicted using RASTA bacteria as a search method were confirmed using this new approach.

More recently Shao *et al*., (2010) published TADB: a web-based resource for type II toxin–antitoxin loci in bacteria and archaea. TADB was derived from computationally predicted data sets and/or reports of experimentally validated TA genes. TADB contains 10,753 type II TA gene pairs identified within 1,240 genome sequences representative of 962 strains of phylogenetically diverse bacteria and archaea (Shao *et al*., 2010). Initially the BLASTP-identified 921 TA loci present in 147 sequenced genomes reported by Pandey *et al*., 2005 were incorporated into TADB. Next, the 5806 TA loci found in 604 genomes which had been assigned to 44 conserved TA domain pairs by Makarova *et al*., 2009 were archived in TADB. The database was then further complemented with the data set of RASTA-Bacteria predicted TA loci identified in 883 annotated genomes (Sevin *et al*., 2007). Subsequently, the three data sets mentioned above were compared to ensure that only unique entries were included in TADB. In addition, further examples of known

or putative TA loci were identified by searching PubMed using the search terms ‘toxin and antitoxin’ and manually inspecting all PubMed hits. Supplementary TA loci identified by this literature search strategy included six TA pairs, absent from the first three data sets mentioned above. In TADB only results obtained for the *S. meliloti* chromosome was considered.

The first characterized functional TA system in *Rhizobiaceae* is the chromosomal *ntrPR* operon of *S. meliloti* (Bodogai *et al*., 2006). Based on protein homologies, the *ntrPR* operon belongs to the *vapBC* family of TA systems. The operon is negatively autoregulated by the NtrPNtrR complex. The N-terminal part of NtrP is responsible for the interaction with promoter DNA, whereas the C-terminal part is required for protein-protein interactions (Bodogai *et al*., 2006). It was shown that NtrR expression results in the inhibition of cell growth and colony formation. According to results obtained by Bodogai and co-workers the *ntrPR* may be important regulators of metabolic rates required for bacterial survival under stressful conditions and/or for the transition from freeliving to symbiotic life style.

MATERIAL AND METHODS

Construction of RmP1099 (ΔA150), RmP1092 (A151), RmP1408 (ΔA152) RmP2681 (ΔA160)

Strains RmP1094 (A150), RmP1092 (A151), RmP1407 (A152), and RmP2671 (A160) were made by combining two existing deletions in a single strain. Experimentally, this was accomplished by transduction of antibiotic resistance from a deletion marked by Gm^rNm^r into a strain carrying a deletion with no antibiotic marker (Figure 5.1) New deletion mutants, RmP1099 (ΔA150), RmP1408 (ΔA152) and RmP2681 (ΔA160) were generated upon the introduction of Flp recombinase encoded on the broad – host range vector pTH1944-Tc^r into the *S. meliloti* strains carrying the two existing deletions, each harboring one FRT site. The donor deletion carries Gm^rNm^r and thus provides markers for detection of deletion events. A deletion mutant using A151 as parental strain could not be recovered. End points as well as deletion sizes of newly made deletion mutants were listed in Table 5.4.

Cloning of *sma0471/sma0473, sma2151, sma2231, sma2253/sma2255, sma2273/sma2275, sma2105* and *sma2133*

*Pac*I fragments containing the *sma0471/sma0473*, *sma2151*, *sma2231*, *sma2253/sma2255*, *sma2273/sma2275*, *sma2105* and *sma2133* coding regions and their predicted promoter regions were amplified from RmP110 DNA using the primers listed in the Table 5.5. The products were cloned as a *Pac*I fragments into the Sp^r broad host range expression vector pTH1931 to produce plasmids pTH2563, pTH2622, pTH2623, pTH2624, pTH2625, pTH2646 and pTH2647 (Table 5.5)

RESULTS

Deletions encompassing the complete pSymA megaplasmid

Using the deletion strategy described in Chapter 3, thirty-three strains that carried FRT sites from pTH1522 and pTH1937 in direct orientation in the pSymA megaplasmid were made. A map of pSymA together with the FRT-flanked regions is shown in Figure 5.2. The FRT flanked regions of pSymA were designated in ascending numbers from A100 (see Figure 5.2). Collectively, the double FRT strains encompassed the entire 1,354,226 bp pSymA megaplasmid of *S. meliloti*. Upon transfer of the *flp* plasmid pTH1944, the largest deletion recovered was 166 kb, while the smallest was 43 kb (Figure 5.2). The precise boundaries of the deletion strains are given in Table 2.1 Chapter

2. Where pTH1944 *flp* plasmid transconjugant strains were recovered and the deletion subsequently verified, the region was designated with a Δ (e.g. ΔA104). No deletions were recovered for the A101 region of the pSymA megaplasmid and this is discussed further below.

Confirmation of pSymA Deletions

We used PCR to detect whether a given DNA region was present or absent in putative deletion strains (described in detail in Chapter 4). The sequence and positions of the primer pairs are given in Table 2.3. Only those deletion strains that were confirmed by PCR are shown in Figure 5.2.

Essential genes on pSymA megaplasmid

The pSymA megaplasmid is 1,354,226 bp in size. The numbering of deletion strains proceeds clockwise from position 1 (*oriV*) on the map as drawn in Figure 5.2. We failed to recover pTH1944 *flp* plasmid transconjugant colonies from the FRT-flanked region A101 (1,283,082 – 10,988 nt), A134 (1,283,082 - 92,124) and A135 (1,348,238 - 48,842) (Figure 5.2).

Flp transconjugants were readily detected for A133 (1,283,082 - 1,349,931) and A102 (9,549 - 48,842). A101, A134 and A135 but not A133 and A102 include the 1,349,931-9,549 nt region which has nine annotated genes – and most significantly the *repA2B2C2*

locus (1,350,001-1,353,535 nt) which is presumed to contain the *oriV* of the pSymA megaplasmid (MacLelland *et al* 2005). Assuming pSymA has no other *oriV*, excision of the *repABC* locus would result in loss of the remainder of the pSymA megaplasmid (Figure 5.2). Therefore, the failure to recover transconjugants in which the *repA2B2C2* region is deleted suggests that the pSymA replicon has genes that are essential for cell survival.

Combining deletions of the pSymA megaplasmid

Although the original deletion series (Figure 5.2) did not show the presence of genes essential for growth in LB complex medium, some pSymA deletions were observed to occur at a lower frequency than others. In additional experiments, we wished to combine two existing deletions and their single FRT sites into single strains. In the case of FRT site in direct orientation, the pSymA region flanked by the FRT sites could then be deleted upon introduction of Flp recombinase. Experimentally, this was accomplished by transduction of antibiotic resistance from a deletion marked by Gm^rNm^r into a strain carrying a deletion with no antibiotic marker (Figure 5.1) (Chapter 5; Material and Methods). Strains that carry deletions with no antibiotic marker harbor the *flp* recombinase plasmid pTH1944. To successfully combine the deletions it was important to cure this plasmid from the recipient cells. This was done by sequential subculturing. Deletion mutants were generated upon the introduction of *flp* recombinase encoded on the broad – host range vector pTH1944-Tc^r into the *S. meliloti* strains carrying the two

existing deletions Using transduction several double deletion were combined; ΔA150 (ΦΔA127 (Gm^rNm^r) → ΔA106 (184,519 nt – 930,910 nt)), ΔA152 (ΦΔA129 (Gm^rNm^r) → ΔA117 (400,267 nt – 1,123,504 nt) and ΔA160 (ΦΔA105 (Gm^rNm^r) → ΔA102 (9,549 nt – 186,200 nt)). However, pTH1944-*flp*-plasmid transconjugants were not recovered from strains carrying A151 (ΦΔA129 (Gm^rNm^r) → ΔA106 (184,519 nt – 1,123,504 nt) (Figure 5.3); indicating presence of essential gene/genes within 184,519 nt – 1,123,504 region flanked by FRT sites.

Next, we employed transduction as an experimental approach to identify genes that affect cell survival. Phage lysates prepared on strains carrying the ΔA150, ΔA152, ΔA160, ΔA130, ΔA131, ΔA132 and ΔA133 were employed to transduce Gm^r into the wild type RmP110. Successful transduction indicates absence of genes essential for cell survival in a deleted region. A reduced transduction frequency or failure to obtain transductants suggested the presence of genes important for cell survival. In control experiments, the *S. meliloti* ΦpTH1522 single integrant strains that were used to generate the deletions were also transduced into the wild type RmP110 (experimental approach is presented in Figure 5.4). No transductants were obtained for the donor strains ΔA131 (1,173,115 – 1,232,916 nt), ΔA132 (1,231,998 – 1, 284,751) and ΔA150 (184,519 nt – 930,910) suggesting the presence of essential genes in the deleted regions (Table 5.6). The *S. meliloti* ΦpTH1522 single integrant strains that were used to generate the deletions were transduced into wild type strain RmP110 with the transductional frequency per recipient greater than 10^{-8} (Table 5.6).

The *sma0471* - *sma0473* locus

No transductants were obtained for ΔA150 (184,519 nt – 930,910), but transductant were readily recovered for strain ΔA152 (400,267 nt – 1,123,504 nt) (Table 5.6) indicating presence of essential gene/genes within the region 193,611 to 400,267 nt. This region is unique for the deletion strain ΔA150. In the region 184,519 to 308,586 nt five sets of double integrant strains with large overlapping regions were made with the purpose of narrowing down the region with the possible gene/genes that effect cell survival. Upon the conjugation of *flp* recombinase encoded on the broad – host range vector pTH1944-Tc^r into double integrant strains of *S. meliloti*, Tc^r transconjugants were recovered for the certain number of strains within each set (deletions colored in green) (Figure 5.5 A-E). These data allowed us to narrow down the region of pSymA that may carried gene(s) essential for growth in LB complex medium. A 9 kb region (255,655 - 264,696 nt) (Figure 5.6) carries the genes annotated as *sma0469, sma0470, sma0471, sma0473, sma0475, sma0476, sma0478, sma5024* and *sma0483* was delineated. *In silico* analysis of these genes suggested that the 309 bp *sma0471* and 294 bp *sma0473* genes (258,225 nt -258,823 nt) are annotated as encoding a RelB antitoxin and addiction module toxin, RelE/StbE, respectively. Makarova *et al.*, (2009) predicted *Sma0471* and *sma0473* to be members of RelE/RelB toxin antitoxin family. These genes were also identified by Sevin *et al*., (2007).

To determine whether *sma0471* and *sma0473* are essential for *S. meliloti* cell viability, we attempted to delete the *sma0471* and *sma0473* genes (and the upstream

region including the promoter) from pSymA in a strain carrying these genes *in trans* on a replicating plasmid pTH1931. A 1272 bp fragment carrying *sma0471* and *sma0473* was amplified and cloned into the Sp^r pTH1931 vector to form pTH1931-*sma0471sma0473* (pTH2563) which was then transferred to the *S. meliloti* strains: A257 (region flanked by FRT sites, harbors *sma0471* and *sma0473*), A101 (region flanked by FRT sites harbors pSymA *oriV*); A128 (region on pSymA flanked by FRT sites that previously was successfully deleted) and B158 (region flanked by FRT sites harbors essential gene for $tRNA^{arg}$). At the same time, the Sp^r pTH1931 (empty vector) was also transferred to the above strains as positive controls for this experiment. The *flp* plasmid pTH1944 was conjugated into each of the strains and Tc^r transconjugants were selected on LBTc medium with or without IPTG. Knowing that the *lacIq* gene is carried on the pTH1931 vector, IPTG was added to the media to insure induction of *flp* recombinase from the *lac* promoter in pTH1944. Conjugation frequencies for both types of medium are listed in a Table 5.7. The *flp* conjugation frequency for the strain A257 *sma0471/73* in a presence of IPTG (A257 region is deleted) of $1x10^{-2}$ is comparable to the conjugational frequency $1.1x10^{-2}$ of control strain ΔA128, indicating that loss of cell viability is counteracted by the *sma0471/sma0473* genes presence on the replicating plasmid. Tc^r transconjugants were recovered at ~ 100 fold lower frequency when the Flp recombinase was expressed in strains: A257 (pTH1931-), A101 (pTH1931 - *sma0471/73*) and B158 (pTH1931 - *sma0471/73*). This result again suggests that the pTH1931 - *Sma0471Sma0473* plasmid can complement deletion of A257 region with predicted toxin/antitoxin genes, but not the deletion of the entire pSymA megaplasmid (A101 (pTH1931 - *sma0471/73*)). s*ma0471*

and/or *sma0473* are essential and they are the only essential genes in A257 region of the pSymA megaplasmid, but not the only essential genes in the entire pSymA megaplasmid. The deletion of the A257 region from pSymA in a strain carrying *Sma0471Sma0473 in trans* on another plasmid was confirmed by PCR using four sets of primers per deletion strain: 4058UFUR; 4058DFDR; 5889UFUR, 5889DFDR for strain A257 (Table 2.1). Only those deletion derivative strains that were confirmed by PCR were saved as a part of the lab strain collection.

The role of the *sma0471/sma0473* genes in transduction of the Gm^rNm^r marked ΔA257 deletion was also examined (Table 5.8). ΔA257 transductants were readily obtained only when the wild type recipient RmP110 carried the *sma0471/sma0473* genes with corresponding promoter region *in trans* on replicating plasmid. In this experiment a transductional frequency of 10^{-7}/ recipient was obtained (Table 5.8).

Effect of *sma0471/sma0473* on growth of *S.meliloti*. Strains A257 (pTH1931 - *sma0471/73*), A257 (pTH1931-), A101 (pTH1931 - *sma0471/73*) and B158 (pTH1931 - *sma0471/73*) (Table 5.7) were streak purified and used to check an effect of *sma0471/sma0473* on growth of *S. meliloti*. For these experiments, overnight cultures growing in LBmc were diluted to an OD_{600} of 0.03 to 0.04 pipetted into test tubes (5 ml per tube), half of which was supplemented with IPTG (0.5 mM final concentration) to induce deletion formation. Growth was measured by monitoring the OD_{600} of the cultures. During the first six hours, the growth rate for all used strains was the same. Growth of: A257 (pTH1931-), A101 (pTH1931 - *sma0471/73*) and B158 (pTH1931 - *sma0471/73*) start to decline after six hours indicating that deletion of the *sma0471/sma0473* genes had

a bacteriostatic if not bacteriocidal effects on *S. meliloti* (Figure 5.8). To investigate cell viability, the number of colony forming units (CFU) was determined by plating samples on LB medium. The number of CFU was plotted as a function of the time and the results indicated the fraction of surviving cells (Figure 5.8). It was apparent that the number of viable cells initially decreased for strains A101 (pTH1931 - *sma0471/73*) and B158 (pTH1931 - *sma0471/73*), while the number of viable cells for the strain A257 (pTH1931-) stayed the same during the time suggesting a bacteriostatic effect resulting from the deletion of Sma0471/Sma0473. In the A257 strain in which the *sma0471/sma0473* genes and the corresponding promoter region were present *in trans* on another plasmid, the numbers of viable cells were similar to the uninduced A257 cultures A257 (pTH1931-) and A257 (pTH1931 - *sma0471/73*). Thus the *sma0471/sma0473* genes appear to completely complement the ΔA257 deletion.

The s*ma2105* loci.

Results from conjugation and transduction experiment suggested that the 60 kb region, A131 (1,173,115 - 1,232,916 nt) carries genes essential for growth in LB complex media (Figure.5.9). *In silico* analysis of this region with respect to the identification of potential toxin antitoxin systems, identified the 1221bp *sma2105* gene (1,187,704 – 1,188,924 nt); the 792bp *sma2133* gene (1,205,622 – 1,206,413 nt) and the 324bp *sma2151* gene (1,215,301 – 1,215624nt). These genes are predicted members of toxin/antitoxin families (s*ma2105* (Fic), *sma2133* (Xre) and *sma2151* (Xre)) (Makarova

et al., 2009) and are annotated as conserved hypothetical, hypothetical and DNA – binding protein, respectively.

Conditional deletion of the A131 region. To determine whether *sma2105*, *sma2133* or *sma2151* is essential for *S. meliloti* cell viability, we attempted to delete the A131 region from pSymA in strains carrying *sma2105* or *sma2133* or *sma2151* with their corresponding promoter region *in trans* on the Sp^r pTH1931 vector, (pTH2646, pTH2647 and pTH2622, respectively). These were transferred to the *S. meliloti* strains: A131 (region flanked by FRT sites harbors *sma2105, sma2133* and *sma2151*), A128 (region on pSymA that previously was successfully deleted) and B158 (region flanked by FRT sites harbors essential gene for $tRNA^{arg}$). The Sp^r pTrc (empty vector) was also transferred to these strains as controls in these experiments. Upon transfer of the *flp* plasmid pTH1944 to each of these strains, transconjugants were plated on LB media with or without IPTG. Conjugation frequencies for both types of media are presented in a Table 5.9. Strain A131 (pTH1931 - *sma2105*) has similar value for the conjugational frequency $5.2x10^{-2}$ in a presence of IPTG as control strain ΔA128 ($1.1x10^{-2}$) (Table 5.9.) This result suggests that complementation was successful only in a presence of *sma2105* gene. Tc^r transconjugants were recovered at a lower frequency when the Flp recombinase was expressed in strains: A131 (pTH1931-) ($1.1x10^{-4}$), A131 (pTH1931 - *sma2133*) ($1.2x10^{-3}$), A131 (pTH1931 - *sma2151*) ($1.2x10^{-4}$) and B158 (pTH1931-) ($2.1x10^{-4}$). This result again suggests that *Sma2105* is essential and that it is the only essential gene in this region of the pSymA megaplasmid.

The role of the *sma2105*, *sma2133* or *sma2151* genes in transduction of the Gm^rNm^r marked ΔA131 deletion was also examined (Table 5.10). ΔA131 transductants were readily obtained only when the wild type recipient RmP110 carried the *sma2105* gene with corresponding promoter region *in trans* on replicating plasmid pTH1931. No transductants were obtained when ΔA131 was transduced into RmP110 carrying *sma2133* or *sma2151* gene, whereas ΦFL4094 was readily transduced into these strains. ΦFL4094 is single integrant strain that was used to generate ΔA131. Transduction frequency obtained when ΔA131 was transduced into RmP110 with *sma2105* gene on replicating plasmid, was $4.32x10^{-7}$/ recipient. This frequency is comparable to one for control strain, $3.66x10^{-7}$/ recipient (Table 5.10).

Results from conjugation and transduction experiment suggested that the *sma2105* gene deletion in A131 region effected the growth of *S. meliloti*. Sma2105 protein had high homology to the Fic domain (filamentation induced by cAMP) present in proteins involved in cell division and the synthesis of PAB (p-aminobenzoate) or folate, indicating that the Fic protein and cAMP are involved in a regulatory mechanism of cell division via folate metabolism (Komano *et al*., 1991). Fic domain spans from 144 to 249 amino acids. Analysis of the Sma2105 protein revealed a second domain, MarR - like which spans from 336 to 382 amino acids (Figure 5.14).

Effect of *sma2105* on growth of *S.meliloti*. We examined the affect of *sma2105* on the growth and cell viability of cells in which the deletion of A131 was controlled via the IPTG-dependent induction of Flp on plasmid pTH1944 (strains from Table 5.9) Strains grown overnight in LBmc medium were diluted to an OD_{600} of 0.03 to 0.04 and half of

the resulting cultures were grown with IPTG (0.5 mM final concentration) (Figure 5.10). Growth of strains: A131 (pTH1931-), A101 (pTH1931 – *sma2105*) and B158 (pTH1931 – *sma2105*) started to decline after six hours of growth indicating that deletion of *sma2105* in a strain A131 (pTH1931-) had a negative effect on the *S. meliloti* cells. To examine the influence of *sma2105* on cell viability the colony forming units (CFU) were determined by plating samples on LB medium with or without IPTG (Figure 5.10). The number of viable cells initially decreased for strains A131 (pTH1931-), A101 (pTH1931 – *sma2105*) and B158 (pTH1931 – *sma2105*), suggesting deletion of *sma2105* had a bactericidal effect.

Complementation with the pTH1931-*sma2105* plasmid was successful for the deletion of A131 region but not for the deletion of the entire pSymA megaplasmid (A101 (pTH1931 - *sma2105*)). This suggests that while *sma2105* is the only essential gene in the A131 region of pSymA, other essential genes are present on the pSymA megaplasmid.

The s*ma2231* loci

The third region on the pSymA megaplasmid that may carry gene(s) essential for growth in LB complex media, is within a 52.7 kb A132 region(1,231,998 - 1,284,751 nt) (Figure 5.11). *In silico* analysis of this region with respect to possible toxin antitoxin like systems, identified that the 384bp *sma2231* gene (1,249,316 – 1,249,699 nt); the 378bp *sma2253* gene (1,263,295 – 1,263,672 nt) and the 339bp *sma2255* gene (1,263,675 – 1,264,013 nt); the 351bp *sma2273* gene (1,268,471 – 1,268,821 nt) and 276bp *sma2275*

gene (1,268,802 – 1,269,077 nt) are good candidates for the "essential genes" in this region. These genes are predicted members of toxin/antitoxin families (Makarova *et al.*, 2009). *sma2231*, *sma2253/sma2255* and *sma2275/sma2273* are annotated as conserved hypothetical; conserved hypothetical/hypothetical and conserved hypothetical/conserved hypothetical protein, respectively.

To investigate the influence of these genes on *S. meliloti* cell viability, we attempted to delete the A132 region from pSymA in strains carrying s*ma2231*, *sma2253/sma2255* or *sma2275/sma2273* genes with their corresponding promoter regions *in trans* on another plasmid. The cloning of *sma2253/sma2255* with the corresponding promoter region in *E. coli* strain DH5α appeared to be lethal. Next, assuming that lethality was caused by the toxin, the gene predicted to be the antitoxin together with its promoter region was cloned and used in further experiments. Each of the genes and promoter regions were amplified and cloned into the Sp^r pTH1931 vector to form pTH1931-*sma2231* (pTH2623), pTH1931-*sma2253* (pTH2624) and pTH1931- *sma2275/sma2273* (pTH2625), respectively. These plasmids and the Sp^r pTH1931 (empty vector) were transferred to the *S. meliloti* strains A132 (region flanked by FRT sites harbors *sma2231*, *sma2253/55* and *sma2275/73*), A128 (region on pSymA that previously was successfully deleted) and B158 (region flanked by FRT sites harbors essential gene for $tRNA^{arg}$).

Conditional deletion of the A132 region. Upon transfer of the *flp* plasmid pTH1944 to each of the *S. meliloti* strains, transconjugants were plated on LB media with or without IPTG. Conjugation frequencies for both types of media are presented in a Table 5.11. The *flp* conjugation frequency $2.2x10^{-3}$ for the strain A132 (pTH1931 -

sma2231) in the presence of IPTG was similar to the conjugational frequency $1.1x10^{-2}$ of the control strain ΔA132 indicating that the deletion was made in the presence of *sma2231* on replicating plasmid. Tc^r transconjugants were recovered at a ten times lower frequency when the Flp recombinase was expressed in strains: A132 (pTH1931-) ($1.8x10^{-4}$), A132 (pTH1931 - *sma2253*) ($2.5x10^{-4}$), A132 (pTH1931 - *sma2275/sma2273*) ($2.1x10^{-4}$) and B158 (pTH1931-) ($2.1x10^{-4}$). This result suggested that *sma2231* is important for *S. meliloti* cell viability and that it is the only essential gene in this region of the pSymA megaplasmid.

ΔA132 from the original deletion series was not marked by any antibiotic marker. For the purpose of a next experiment we attempted to make deletions that covered the same region but would be marked by antibiotic resistance GmNm. The region of interest was designated with an A136. All our attempts to delete A136 region were unsuccessful; we could not obtain viable Tc^r transconjugants. Next, we changed the experimental approach and introduced the pTH1931 plasmid carrying the *sma2231* gene to A136 strain. Upon transfer of the *flp* plasmid pTH1944, viable transconjugants were obtained at the expected conjugational frequency ($\sim10^{-2}$).

The role of the *sma2231* gene in transduction of the Gm^rNm^r marked ΔA136 deletion was also examined (Table 5.12). ΔA136 transductants were readily obtained only when the wild type recipient RmP110 carried the *sma2231* genes with corresponding promoter region *in trans* on a replicating plasmid. In this experiment transductional frequency of $1.01x10^{-7}$/recipient was obtained (Table 5.12). These results were comparable to data obtained for donor control transductional frequency $2.3x10^{-7}$/

recipient. No transductants were obtained when we attempted to transduce ΔA136 into RmP110 carrying *sma2253* or *sma2275/sma2273*. Complementation was successful only in a presence of *sma2231* gene on replicating plasmid (Table 5.12).

The Sma2231 protein showed similarity to the VapC protein family and to related proteins of various microorganisms (Figure 5.12). It was found that Sma2231 protein had high homology to the PIN domain, which is described as the N-terminal domain of the PilT protein (Wall and Kaiser 1999; Wolfgang *et al*. 2000). PIN domain has been identified in many bacteria, *Archaea*, and eukaryota (Makarova *et al*. 1999). Despite the fact that the annotation of the *S. meliloti* genome did not reveal gene directly upstream of *sma2231*, we found an open reading frame that is homologous to antitoxin genes of *Cupriavidus metallidurans* CH34, *Rothia dentocariosa* M567 and *Mycobacterium tuberculosis* KZN 4207 in the *sma2229-sma2231* intergenic region. The newly identified open reading frame which is localized upstream of *sma2231* was designated *sma2230.* The *sma2230* gene encodes a 66 amino acid protein. Domain analysis within Sma2230 revealed the presence of a domain of unknown function 2291 between amino acids 4 and 47 (Figure 5.14). Most characterized *vapB* antitoxins contain a SpoVT/AbrB domain, which is a DNA binding domain, and, as such, belongs to the super family of transcriptional regulators of the same name. The presence of a SpoVT/AbrB domain within Sma2230 was not confirmed in our analysis.

Effect of *sma2230/sma2231* on growth of *S.meliloti*. We examined the effect of *sma2230/sma2231* on the growth and cell viability of cells in which the deletion of A132 was controlled via the IPTG-dependent induction of Flp on plasmid pTH1944 (strains

from Table 5.11). For these experiments, overnight *S. meliloti* cultures growing in LBmc were diluted to an OD_{600} of 0.03 to 0.04 pipetted into test tubes (5 ml per tube), half of which was supplemented with IPTG (0.5 mM final concentration) to induce deletion formation and growth was measured by monitoring the OD_{600} of the cultures. Growth of strains: A132 (pTH1931-), A101 (pTH1931 – *sma2230/sma2231*) and B158 (pTH1931 – *sma2230/sma2231*) started to decline after six hours of growth indicating that deletion of *sma2230/sma2231* in a strain A132 (pTH1931-) had a negative effect on the *S. meliloti* cells (Figure 5.13). To examine the influence of *sma2230/sma2231* on cell viability, the colony forming units (CFU) were determined by plating samples on LB medium with (to induce *flp* recombinase) or without IPTG (Figure 5.13). The number of viable cells initially decreased for strains A132 (pTH1931-), A101 (pTH1931 – *sma2230/sma2231*) and B158 (pTH1931 – *sma2230/sma2231*), suggesting deletion of *sma2230/sma2231* had a bactericidal effect.

Complementation with the pTH1931- *sma2230/sma2231* plasmid was successful for the deletion of the A132 region but not for the deletion of the entire pSymA megaplasmid (A101 (pTH1931 - *sma2230/sma2231*)). This suggests that while *sma2230/sma223* are the only essential genes in A132 region of pSymA, other essential genes are present on the pSymA megaplasmid.

Carbon and Nitrogen Utilization Phenotypes associated with pSymA deletions

Malonic acid utilization

In a previous work Chen *et al*., 2010., showed that putative uptake system *matPQMAB* (*sma0150-sma0157*) is required for malonate transport in *S. meliloti*. The same gene cluster *sma0150-sma0157* (82,374 – 88,287 nt) was found to be induced by 10 mM malonic acid (Mauchline *et al*., 2006). This gene cluster was removed in the deletion strain ΔA103 (47,717 – 92,124 nt). The ΔA103 strain and wildtype RmP110 were examined for growth on malonic acid and the result demonstrated that the ΔA103 strain grew slower than the wildtype on M9-medium supplemented with 5mM malonic acid as sole carbon sources. This confirmed that the genes responsible for malonic acid utilization are within the region covered by the ΔA103 deletion.

Symbiotic phenotypes of pSymA deletions

The symbiotic phenotypes of pSymA deletion strains were assayed by inoculating alfalfa (*Medicago sativa*) seedlings and growing in Leonard assemblies under nitrogen deficient conditions. Plant growth was examined 28 days post inoculation and shoot dry weights were determined (Table 5.13). All deletion strains for regions known to be required for effective symbiosis formed nodules that failed to fix N_2, Fix^-. The deletion strain ΔA117 (400,267 – 459,668 nt) (Figure 5.2) lacked the *nodD2, fixT2K2N2D2O2P2,*

nodL, noeAB, fixU, nifBA, fixXCBA and *nifHDKEX* genes required for the formation of Fix^+ nodules (Honma *et al*, 1987; Ardourel *et al*, 1995; Hirsch *et al*., 1986; Ruvkun *et al*.,1982; Earl *et al*., 1987). The same set of genes was missing in deletion strains ΔA114 (294,016 – 459,668 nt), ΔA115 (306,156 – 459,668 nt) and ΔA116 (311,877 – 459,668 nt), but also genes *fixN3O3Q3P3*. Strains ΔA114, ΔA115, ΔA116 and ΔA117 can form nodules but can not fix nitrogen due to the deletion of various *nif* and *fix* genes. The deletion strain ΔA118 (458,916–507,338 nt) (Figure 5.2) includes the region with the genes that are involved in nodulation (*nodD3HFEGP1Q1JICBAD1NM*) as well as nitrogen fixation (*nifEXN*) and gave rise to plants without root nodules (Honma *et al*, 1987). Alfalfa plants inoculated with deletion strain ΔA121 (623,673-678,150 nt) were symbiotically ineffective. This deletion strain lacks genes important for efficient nitrogen fixation *fixS1I1HGP1Q1O1N1MK1T1JL* (Kahn *et al*., 1989).

DISCUSSION

We report the construction of a series of 43-166 kilobase deletions with defined endpoints that cover the entire *S. meliloti* pSymA (1354 -kilobase) megaplasmid. The failure to recover A101 *flp* plasmid transconjugants in which the *repA2B2C2* region was deleted suggests that pSymA replicon has only one *oriV* and further that there are genes carried on the rest of the pSymA megaplasmid that are essential for cell survival. Using the Flp/FRT deletion strategy combined with transduction, three toxin antitoxins like

systems, *sma0471/sma0473*, *sma2105* and *sma2231*, located on the pSymA megaplasmid were identified.

Toxin antitoxin systems are of interest in developing experimental tools in molecular biology. Two of the well-described TA systems, *ccdAB* and *mazEF* have already found application. The *ccdAB* is used as a factor for positive selection of transformants, primarily in *E. coli* strains (Bernard *et al.*, 1994). Commercially available systems (StabyCloningTM and StabyExpressTM, Delphi Genetics SA), are based on CcdB toxicity against gyrase and allow one-step selection of transformants (Bukowski *et al*., 2011). The *mazEF* system has been adapted for single protein production systems. The MazF toxin is used to trigger bacteriostasis and to shut down endogenous protein synthesis. The recombinant gene lacks the ACA sequences, recognized by the MazF interferase. Therefore upon induction of MazF expression the production of the recombinant protein of interest is continued almost exclusively (Bukowski *et al*., 2011). Identification of new toxin antitoxin families like the one on the pSymA megaplasmid may allow the development of new experimental tools. Application of newly found toxins and antitoxins in antibiotic therapy have potential as it is known that TA systems induce bacteriostasis, whose prolongation results in bacterial cell death (Bukowski *et al*., 2011).

Makarova *et al*., 2009 analyzed 750 complete genomes of prokaryotes for the presence of solo toxin and antitoxin genes as well as TA pairs. According to their *in silico* analysis, *Sinorhizobium meliloti* is among the top twenty bacteria whose genome is highly abundant with TA loci. Makarova *et al*., 2009, predicted 50 loci on the pSymA megaplasmid including *sma0471/sma0473*, *sma2105* and *sma2231* as potential toxin

antitoxin like systems. Two years earlier using RASTA-bacteria, a web-based tool for identifying toxin-antitoxin loci in prokaryotes Sevin *et al*. identified 17 toxin antitoxin loci on pSymA megaplasmid including *sma0471/sma0473* and *sma2231*.

The *sma0471/sma0473* genes were proposed to be members of *relB/relE* toxin antitoxin family. Database searches using the BLAST program (Altschul *et al*., 1990) revealed a large number of *relBE* homologous gene systems in a wide variety of prokaryotic organisms. Thus, many Gram-negative and Gram-positive bacteria contain *relBE* homologous gene systems. The *relBE* homologous gene systems are also abundant in Archaea (*A. fulgidus* and *M. jannaschii*). The organization of the *relBE* operon is typical for bacterial TA systems. The two genes encode two small proteins. In a case of *sma0471*/*sma0473* antitoxin and toxin proteins are 102 and 97 amino acids long. The start codon of *sma0473* overlaps with the stop codon of *sma0471*, perhaps indicating translational coupling between the genes. In all cases, when both genes are identified the *relB* antitoxin homolog (*sma0471*) are located upstream of the toxin-encoding gene *relE* (*sma0473*). Deletion analysis of *sma0471/sma0473* resulted in reduced growth and a loss in colony forming ability (Figure 5.8). In our database search we found putative *relBE* toxins antitoxins with high identity to Sma0471/Sma0473 belong to members of Rhizobiaceae family as well as *Methylobacterium nodulans* ORS 2060, *Nitrobacter hamburgensis* X14, *Escherichia coli* str. K-12 and *Pseudomonas syringae pv. phaseolicola* 1448A. Some putative and well characterized members of the *relBE* TA family, are aligned in Figure 5.7 A and B. The degree of identity ranges from 49% to 84% among the RelB homologues and from 50% to 100% among the RelE homologues. The

alignment of the RelB homologues shows that these proteins are considerably more divergent than the RelE homologues (Figure 5.7 A, B). The *relB/relE* toxin antitoxin systems are localized on bacterial plasmids as well as chromosomes. The systems that are encoded by bacterial plasmids are considered as mechanisms to ensure plasmid maintenance (Jensen and Gerdes, 1995).

We showed that deletion of *sma0471/sma0473* caused bacteriostasis from which the cells can be recovered. Bacteriostasis is counteracted by the *sma0471/sma0473* gene products in a strain carrying *sma0471/sma0473 in trans* on another plasmid. The phenotype can be explained if the RelB (*sma0471*) antitoxin is more unstable than the RelE (*sma0473*) toxin: cells that lose the *sma0471/sma0473*-carrying region (ΔA257) on the pSymA plasmid experience decay of the antitoxin, which thus leads to activation of the toxin and killing of the cells or leading cells toward bacteriostasis. In several cases, this simple model has been shown to be valid for the plasmid stabilization systems (Jensen and Gerdes, 1995). Toxicity of *Eco*RelE protein is due to cleavage of translating mRNA at the ribosomal A site (Neubauer *et al*., 2009). Other RelE homologues seem to cleave RNA in a similar manner, and thus cells exposed to the toxin show a drastic growth arrest (Christensen *et al*., 2003; Nieto *et al*., 2006). High percentage of homologues among Sma0473 and RelE proteins from other microorganisms allow us to predict the same mode of action for Sma0473 toxin, but this still needs to be experimentally confirmed.

According to Makarova *et al*., 2009 gene *sma2105* from pSymA megaplasmid of *S. meliloti* is predicted to be a toxin component of TA system with the protein domain

similar to Fic/Doc toxins. Our search for conserved domains in Sma2105 protein (406 amino acids) revealed presence of a Fic domain within this protein (residue 144 - 249) (Finn *et al*., 2010). The Fic domain is present in three protein families: Doc toxin (death on curing component of *phd/doc* TA family), the bacterial cAMP-induced filamentation protein (Fic) and the eukaryotic Huntingtin Yeast Protein E (HYPE) protein (Anantharaman *et al*., 2003). All Doc homologues are 51 – 331 amino acids proteins with a conserved central motif of 9 residues HPFXXGNG that is shared with two other protein families. The central HPFXXGNG motif was also identified in Sma2105. The presence of MarR - like domain (Finn *et al*., 2010) (residue 336 – 382) indicates the possible involvement of this protein in transcriptional regulation which is a major characteristic of antitoxins in TA pairs. The *phd/doc* operon encodes a TA module aiding the maintenance of the plasmid-prophage P1 in *E. coli* (Lehnherr *et al*., 1993). Doc is an inhibitor of translation elongation through its association with the 30 S ribosomal subunit. Deletion analysis showed that deletion of A131 region, which contains *sma2105*, has a bacteristatic if not bactericidal effect on *S. meliloti* cells (Figure 5.10). This is seen as a decrease in the number of viable cells upon deletion of A131. These data are in agreement with Garcia – Pino *et al.*, 2008 who showed that induction of Doc leads to growth arrest of *E. coli*. They observed that growth arrest is reversible as cells are capable of colony formation upon plating on selectable medium. Bacteristatic/bactericidal effect of ΔA131 to the *S. meliloti* cells can be explained using the established model of TA action: if *sma2105* acts as a toxin, then upon deletion of A131 region cells that lose the *sma2105* on the pSymA plasmid experience decay of yet unidentified antitoxin, which thus leads to activation of

toxin and killing of the cells. Further, this phenotype can be counteracted by the *sma2105* gene product in a strain carrying *sma2105 in trans* on another plasmid. According to *in silico* data (Makarova *et al*., 2009) *sma2105* is the only gene among all predicted TA systems in *S. meliloti* genome with a Fic domain.

In prokaryotes, the vast majority of PIN-domain proteins are the toxic components of toxin–antitoxin (TA) system, *vapBC*. *S. meliloti* is one of the organisms with a greatly expanded number of *vapBC* TA loci with 21 predicted members. These are evenly spread among chromosome and two megaplasmids. According to the work of Makarova *et al*., 2009, gene *sma2231* from pSymA megaplasmid is a predicted member of the PIN-domain protein family. Presence of a PIN domain within a toxin is a distinctive feature of the toxins from the VapC and ChpK families (Arcus *et al*., 2005; Miallau *et al*., 2009). The PIN domains (homologues of the pilT N-terminal domain) are small protein domains of 130 amino acids (Arcus *et al*., 2005). In eukaryotes, PIN domain proteins function as ribonucleases with activity linked to RNAi and nonsense-mediated RNA degradation (Clissold & Ponting, 2000). In prokaryotes, the majority of PIN-domain proteins are the toxic components of chromosomally encoded TA operons (Arcus *et al*., 2005). We need to show endoribonuclease activity of *sma2231* which is present in the other VapC homologues and which specifically blocks protein translation via mRNA cleavage (Ramage *et al*., 2009). Despite the fact that annotation of *S. meliloti* genome did not reveal a gene upstream of *sma2231*, we were able to identify an open reading frame within the intergenic region *sma2229-sma2231* that is homologous to antitoxin genes of *Cupriavidus metallidurans* CH34, *Rothia dentocariosa* M567 and *Mycobacterium*

tuberculosis KZN 4207. The gene that lies within *sma2229-sma2231* intergenic region was designated *sma2230*. Further, Sma2230 was searched for conserved domains to elucidate its function. Domain analysis revealed the presence of a domain of unknown function (DUF 2291) (Finn *et al*., 2010) between amino acids 4 and 47 within Sma2230. The organization of the *sma2230/sma2231* operon identified in *S. meliloti* exhibits remarkable similarities to bacterial TA systems. The genes encode two small proteins (66 and 127 amino acids, *sma2230* and *sma2231*, respectively). The domain organization resembled one family of TA systems, the presence of the domain of the unknown function in the first and a PIN domain in the second protein (Anantharaman and Aravind 2003). Furthermore, based on the homology to the VapC protein described earlier (Oláh *et al*. 2001), the *sma2231* and *sma2230* pair shows a certain level of similarity to the most highly abundant class of TA systems, the *vapBC* family (Gerdes *et al*. 2005). If *sma2230* and *sma2231* are components of toxin antitoxin system one would expect that deletion of both genes will have certain effects on cell viability. Our deletion analysis results show that deletion of *sma2231* and the upstream gene *sma2230* is lethal for the *S. meliloti* cells. This phenotype is expected from deletion of both components of the TA system simultaneously. We can support our statement with an already established model where antitoxin is more unstable than the toxin: cells that lose the *sma2230/sma2231*-carrying region (ΔA132) on the pSymA plasmid experience decay of the antitoxin (*sma2230*), which thus leads to activation of the *sma2231* toxin and killing of the cells.

According to *in silico* analysis done by Makarova *et al*., 2009, eight genes on the pSymA megaplasmid contain a PIN domain and are predicted to be a members of *vapBC*

toxin-antitoxin family. Beside these eight genes, 42 genes located elsewhere on pSymA were predicted to be members of other toxin antitoxin families. It is interesting to note that several TA systems, or multiple copies of one TA system, maybe present in one genome. The presence of multiple TA systems in one particular genome, such as *S. meliloti*, may imply the complexity of regulation during the transition of metabolism (Zhang *et al*., 2004). The organisms that have many TA loci are slow - growing and free - living, consistent with the hypothesis that TA loci function in quality control of gene expression. By reducing the production of faulty proteins, the loci might help the cells to survive nutritional stress.

In our work we found that deletion of only three out of 50 toxin antitoxin like systems, on pSymA has an effect on *S. meliloti* viability. This finding raises a question what is role of the other 47 predicted toxin antitoxin genes localized on the same megaplasmid? Results obtained in this study are in agreement with data from several groups that have tested single-TA-deletion mutants and could not observe TA-dependent program cell death under various stress conditions. Single-TA-deletion mutants were tested in *E. coli* (Christensen *et al.*, 2003), *Vibrio cholerae* (Budde *et al*., 2007) and *S. mutans* (Lemos *et al*., 2005) under amino acid starvation, antibiotic treatments and long-term starvation, respectively. Our data showed absence of a synergistic effect on *S. meliloti* cell viability when more than one toxin antitoxin like system is deleted simultaneously. Similar results were obtained by Tsilibaris *et al*., 2007 who analyzed the behavior of an *E. coli* mutant strain deleted for all five known TA systems under a variety of stress conditions (amino acid starvation, rifampin treatment, acidic stress, and

nutritional downshift) and during the poststress recovery phase. Under all the conditions tested, the wild-type and mutant strains showed similar reversible growth inhibitions, indicating that neither the growth inhibition nor the poststress recovery phase was dependent on any of the five TA systems (Tsilibaris *et al*., 2007).

The apparent lack of a phenotype resulting from the deletion of TA like systems can be explained by having conditions under which specific factors directly modulate the regulation of TA systems. Gerdes *et al*., proposed the existence of a cofactor modulating the specificity of the Lon ATP-dependent protease which might provide a molecular explanation for the specific activation of *yefM-yoeB* under Lon overproduction conditions (Christensen *et al*., 2001, 2004).

Another explanation is based on the hypothesis that TA systems might have different functions depending on their genomic locations. It was shown that TA systems located in the superintegron of *Vibrionaceae* promote genomic stability either by stabilizing the superintegron structure (Christensen-Dalsgaard *et al*., 2006, Szekeres *et al*., 2007) or by preserving the integrity of an entire chromosome (chromosome II of *V. cholerae*) (Yamaichi *et al*., 2011). In *E. coli*, *relBE* is actually located in the cryptic Qin prophage while the other four systems are part of the genomic core (Perna *et al*., 2002). According to Tsilibaris *et al*., 2007 the ability to stabilize genomes might be characteristic of TA systems localized on or recently originated from mobile genetic elements. However, the stabilization potential of some of these systems might fade and eventually be lost once the systems are integrated into the genomic core. Such TA systems could

have adapted to their new chromosomal environments by evolving toward a new function that could provide benefit to their host under specific conditions (Tsilibaris *et al*., 2007).

In 2000 Oresnik *et al*., showed that repetitive rounds of Tn5B12-S mutagenesis with selection for deletion can be used successfully to cure the *nod-nif* megaplasmid, pRme2011a, of strain Rm2011. Our explanation for this success is that there is very low probability of this event and that this was successfully achieved in their work. MacLellan *et al*. 2006., tested whether the incompatibility region (incγ) from pSymA, cloned into the broad host range plasmid pBBR1MCS-3 mediated incompatibility in a replicon specific manner. To examine incompatibility, MacLellan *et al*. employed a strain of *recA*-deficient *S. meliloti* and demonstrated that the recombinant broad host range plasmid containing the *inc* sequence from pSymA plasmid can not be maintained in strain that carries pSymA. Inability to obtain *S. meliloti* strains without the pSymA megaplasmid in a presence of broad host range plasmid containing *inc* sequence from pSymA supports our finding that pSymA carries genes that are essential for cell survival.

REFERENCES

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, D.J. Lipman. (1990) Basic local alignment search tool. J. Mol. Biol. **215**:403-10.

Anantharaman, V., L. Aravind. (2003) New connections in the prokaryotic toxin-antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system Genome Biol. **4**:R81.

Arcus, V.L., P.B. Rainey, S.J. Turner. (2005) The PIN-domain toxin-antitoxin array in mycobacteria. Trends Microbiol. **13**:360-5.

Ardourel, M., G. Lortet, F. Maillet, P. Roche, G. Truchet, J.C. Promé, C. Rosenberg. (1995) In *Rhizobium meliloti*, the operon associated with the *nod* box n5 comprises *nodL*, *noeA* and *noeB*, three host-range genes specifically required for the nodulation of particular Medicago species. Mol.Microbiol.**17**:687-99.

Barnett, M.J., R.F. Fisher, T. Jones, C. Komp, A.P. Abola, F. Barloy-Hubler, L. Bowser, D. Capela, F. Galibert, J. Gouzy, M. Gurjal, A. Hong, L. Huizar, R.W. Hyman, D. Kahn, M.L. Kahn, S. Kalman, D.H. Keating, C. Palm, M.C. Peck, R. Surzycki, D.H. Wells, K.C. Yeh, R.W. Davis, N.A. Federspiel, S.R. Long. (2001) Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. Proc.Natl.Acad.Sci.U.S.A. **98**:9883-8.

Becker, A., M.J. Barnett, D. Capela, M. Dondrup, P.B. Kamp, E. Krol, B. Linke, S. Rüberg, K. Runte, B.K. Schroeder, S. Weidner, S.N. Yurgel, J. Batut, S.R. Long, A. Pühler, A. Goesmann. (2009) A portal for rhizobial genomes: RhizoGATE integrates a *Sinorhizobium meliloti* genome annotation update with postgenome data. J. Biotechnol. **140:**45-50.

Bernard, P., P. Gabant, E.M. Bahassi, M. Couturier. (1994) Positive-selection vectors using the F plasmid *ccdB* killer gene. Gene **148**: 71–74.

Bodogai, M., S. Ferenczi, D. Bashtovyy, P. Miclea, P. Papp, I. Dusha. (2006) The *ntrPR* operon of *Sinorhizobium meliloti* is organized and functions as a toxin–antitoxin module. Mol. Plant-Microbe Interact. **19**:811–822.

Budde, P.P., B.M. Davis, J. Yuan, M.K. Waldor. (2007) Characterization of a *higBA* toxin-antitoxin locus in *Vibrio cholera* J. Bacteriol. **189**:491-500.

Bukowski, M., A. Rojowska, B. Wladyka. (2011) Prokaryotic toxin-antitoxin systems--the role in bacterial physiology and application in molecular biology. Acta Biochim. Pol. **58**:1-9.

Chen, A.M., Y.B. Wang, S. Jie, A.Y. Yu, L. Luo, G.Q. Yu, J.B. Zhu, Y.Z. Wang. (2010) Identification of a TRAP transporter for malonate transport and its expression regulated by GtrA from *Sinorhizobium meliloti*. Res. Microbiol. **161**:556-64.

Christensen-Dalsgaard**,** M**.**, K. Gerdes. (2006) Two *higBA* loci in the *Vibrio cholerae* superintegron encode mRNA cleaving enzymes and can stabilize plasmids. Mol**.** Microbiol. **62**:397-411**.**

Christensen, S.K., G. Maenhaut-Michel, N. Mine, S. Gottesman, K. Gerdes, L. Van Melderen. (2004) Overproduction of the Lon protease triggers inhibition of translation in *Escherichia coli*: involvement of the *yefM-yoeB* toxin-antitoxin system. Mol. Microbiol. **51**:1705-17.

Christensen, S.K., K. Pedersen, F.G. Hansen, K. Gerdes. (2003) Toxin-antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBK cleave translated RNAs and are counteracted by tmRNA. J. Mol. Biol. **332**:809-19.

Christensen, S.K., M. Mikkelsen, K. Pedersen, K. Gerdes. (2001) RelE, a global inhibitor of translation, is activated during nutritional stress Proc. Natl. Acad. Sci. U.S.A. **98**:14328-33.

Clissold, P.M., C.P. Ponting. (2000) PIN domains in nonsense-mediated mRNA decay and RNAi. Curr. Biol. **10**:R888-90.

Earl, C.D., C.W. Ronson, F.M. Ausubel. (1987) Genetic and structural analysis of the *Rhizobium meliloti fixA, fixB, fixC,* and *fixX* genes J. Bacteriol. **169**:1127-36.

Finn R.D., J. Mistry, J. Tate, P. Coggill, A. Heger, J.E. Pollington, O.L. Gavin, P. Gunesekaran, G. Ceric, K. Forslund, L. Holm, E.L. Sonnhammer, S.R. Eddy, A. Bateman. (2010) The Pfam protein families database. Nucl. Acids Res. Database Issue **38**:D211-222.

Garcia-Pino, A., M. Christensen-Dalsgaard, L. Wyns, M. Yarmolinsky, R.D. Magnuson, K. Gerdes, R. Loris. (2008) Doc of prophage P1 is inhibited by its antitoxin partner Phd through fold complementation. J. Biol. Chem. **283**:30821-7.

Gazit, E., R.T. Sauer. (1999) The Doc toxin and Phd antidote proteins of the bacteriophage P1 plasmid addiction system form a heterotrimeric complex J. Biol. Chem. **274**:16813-16818.

Gerdes, K. P.B. Rasmussen, S. Molin. (1986) Unique type of plasmid maintenance function, postsegregational killing of plasmid free cells. Proc. Natl. Acad. Sci. U.S.A. **83**:3116-3120.

Gerdes, K. (2000) Toxin-antitoxin modules may regulate synthesis of macromolecules during nutritional stress. J. Bacteriol. **182**:561-72.

Gerdes, K., S.K. Christensen, A. Lobner-Olesen. (2005) Prokaryotic toxin-antitoxin stress response loci. Nat. Rev. Microbiol. **3**:371-82.

Gotfredsen, M., K. Gerdes. (1998) The *Escherichia coli relBE* genes belong to a new toxin–antitoxin gene family. Mol. Microbiol. **29**:1065–1076.

Greenfield, T.J., E. Ehli, T. Kirshenmann, T. Franch, K. Gerdes, K.E.Weaver. (2000) The antisense RNA of the *par* locus of pAD1 regulates the expression of a 33-amino-acid toxic peptide by an unusual mechanism. Mol. Microbiol **37**:652-60.

Hazan, R., H. Engelberg-Kulka. (2004) *Escherichia coli mazEF*-mediated cell death as a defense mechanism that inhibits the spread of phage P1. Mol. Genet. Genomics. **272**:227-34.

Honma, M.A., F.M. Ausubel. (1987) *Rhizobium meliloti* has three functional copies of the *nodD* symbiotic regulatory gene. Proc. Natl. Acad. Sci. U.S.A. **84**:8558-62.

Hurley, J.M., N.A. Woychik. (2009) Bacterial toxin HigB associates with ribosomes and mediates translation-dependent mRNA cleavage at A-rich sites. J. Biol. Chem. **284**:18605-13.

Jaffe, A., T. Ogura, S. Hiraga. (1985) Effects of the *ccd* function of the F plasmid on growth. J. Bacteriol. **163**: 841 – 849.

Jensen, R.B., K. Gerdes. (1995) Programmed cell death in bacteria: proteic plasmid stabilization systems. Mol. Microbiol. **17**:205-10.

Jiang Y, J. Pogliano, D.R. Helinski, I. Konieczny. (2002) ParE toxin encoded by the broad-host-range plasmid RK2 is an inhibitor of *Escherichia coli* gyrase. Mol. Microbiol. **44**:971-9.

Johnson, E.P., A.R. Strom, D.R. Helinski. (1996) Plasmid RK2 toxin protein ParE: purification and interaction with the ParD antitoxin protein. J. Bacteriol. **178**:1420-29.

Kahn, D., M. David, O. Domergue, M.L, Daveran, J. Ghai, P.R. Hirsch, J. Batut. (1989) *Rhizobium meliloti fixGHI* sequence predicts involvement of a specific cation pump in symbiotic nitrogen fixation. J. Bacteriol. **171**:929-39.

Kaminski, P.A., F. Norel, N. Desnoues, A. Kush, G. Salzano, C. Elmerich. (1988) Characterization of the *fixABC* region of *Azorhizobium caulinodans* ORS571 and identification of a new nitrogen fixation gene. Mol Gen Genet. **214**:496-502.

Komano, T., R. Utsumi, M. Kawamukai. (1991). Functional analysis of the *fic* gene involved in regulation of cell division**.** Res. Microbiol. **142:**269-277.

Lehnherr, H., E. Maguin, S. Jafri, M.B. Yarmolinsky. (1993) Plasmid addiction genes of bacteriophage P1: *doc*, which causes cell death on curing of prophage, and *phd*, which prevents host death when prophage is retained. J. Mol. Biol. **233**: 414-428.

Lemos, J.A., T.A. Brown, J. Abranches, R.A. Burne. (2005) Characteristics of *Streptococcus mutans* strains lacking the MazEF and RelBE toxin-antitoxin modules. FEMS Microbiol. Lett. **253**:251-7.

Liu, M., Y. Zhang, M. Inouye, N.A. Woychik. (2008) Bacterial addiction module toxin Doc inhibits translation elongation through its association with the 30S ribosomal subunit. Proc. Natl. Acad. Sci. U. S. A. **105**:5885-5890.

MacLellan, S.R., L.A. Smallbone, C.D. Sibley, and T.M. Finan. (2005). The expression of a novel antisense gene mediates incompatibility within the large *repABC* family of alpha-proteobacterial plasmids. Mol Microbiol. **55**:611-23.

MacLellan S.R., R. Zaheer, A.L. Sartor, A.M. MacLean, T.M. Finan. (2006) Identification of a megaplasmid centromere reveals genetic structural diversity within the *repABC* family of basic replicons. Mol. Microbiol. **59**:1559-75.

Makarova, K.S., Y.I. Wolf, E.V. Koonin. (2009) Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. Biol Direct. 3;4:19.

Mattison, K., J.S. Wilbur, M. So, R.G. Brennan. (2006) Structure of FitAB from *Neisseria gonorrhoeae* bound to DNA reveals a tetramer of toxin–antitoxin heterodimers containing PIN-domains and ribbon–helix–helix motifs. J. Biol. Chem. **281**:37942–37951.

Mauchline, T.H., J.E. Fowler, A.K. East, A.L. Sartor, R. Zaheer, A.H.F. Hosie, P.S. Poole, and T.M. Finan.(2006). Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome. Proc.Natl.Acad.Sci.U.S.A. **103**:17933-8.

McKinley, J. E., R.D. Magnuson. (2005) Characterization of the *Phd* repressor-antitoxin boundary. J. Bacteriol. **187**:765-770.

Meinhart, A., J.C. Alonso, N. Sträter, W. Saenger. (2003) Crystal structure of the plasmid maintenance system epsilon/zeta: functional mechanism of toxin zeta and inactivation by epsilon 2 zeta 2 complex formation. Proc. Natl. Acad. Sci. USA. **100**:1661-6.

Miallau, L., M. Faller, J. Chiang, M. Arbing, F. Guo, D. Cascio, D. Eisenberg. (2009) Structure and proposed activity of a member of the VapBC family of toxin-antitoxin systems. VapBC-5 from *Mycobacterium tuberculosis* J. Biol. Chem. **284**:276-83.

Neubauer, C., Y.G. Gao, K.R. Andersen, C.M. Dunham, A.C. Kelley, J. Hentschel, K. Gerdes, V. Ramakrishnan, D.E. Brodersen. (2009) The structural basis for mRNA recognition and cleavage by the ribosome-dependent endonuclease RelE. Cell. **139**:1084-95.

Nieto, C., T. Pellicer, D. Balsa, S.K. Christensen, K. Gerdes, M. Espinosa. (2006) The chromosomal *relBE2* toxin-antitoxin locus of *Streptococcus pneumoniae*: characterization and use of a bioluminescence resonance energy transfer assay to detect toxin-antitoxin interaction. Mol. Microbiol. **59**:1280-96.

Oberer, M., H. Lindner, O. Glatter, C. Kratky, W. Keller. (1999) Thermodynamic properties and DNA binding of the ParD protein from the broad host-range plasmid RK2/RP4 killing system. Biol Chem. **380**:1413-20.

Ogura, T., S. Hiraga. (1983) Mini-F plasmidgenes that couple host cell division to plasmid proliferation. Proc. Natl. Acad. Sci. U.S.A. **80**:4784-4788.

Oláh, B., E. Kiss, Z. Györgypál, J. Borzi, G. Cinege, G. Csanádi, J. Batut, A. Kondorosi, I. Dusha. (2001) Mutation in the *ntrR* gene, a member of the *vap* gene family, increases the symbiotic efficiency of *Sinorhizobium meliloti*. Mol.Plant. Microbe. Interact. **14**:887-94.

Oresnik, I.J., S.L. Liu, C.K. Yost, M.F. Hynes. (2000) Megaplasmid pRme2011a of *Sinorhizobium meliloti* is not required for viability. J. Bacteriol. **182**:3582-6**.**

Overgaard, M., J. Borch, M.G. Jorgensen, K. Gerdes. (2008) Messenger RNA interferase RelE controls *relBE* transcription by conditional cooperativity. Mol. Microbiol. **69**:841–857.

Pandey, D.P., K. Gerdes. (2005) Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. Nucleic Acids Res. **33**:966-76.

Pedersen, K., A.V. Zavialov, M.Y. Pavlov, J. Elf, K. Gerdes, M. Ehrenberg. (2003) The bacterial toxin *RelE* displays codon-specific cleavage of mRNAs in the ribosomal A site. Cell. **112**:131-40.

Pullinger, G.D., A.J. Lax. (1992) A *Salmonella dublin* virulence plasmid locus that affects bacterial growth under nutrient-limited conditions. Mol Microbiol. **6**:1631-43.

Ramage, H.R., L.E. Connolly, J.S. Cox. (2009) Comprehensive functional analysis of *Mycobacterium tuberculosis* toxin-antitoxin systems: implications for pathogenesis, stress responses, and evolution. PLoS Genet. **5**:e1000767

Ruvkun, G.B., V. Sundaresan, F.M. Ausubel. (1982) Directed transposon Tn5 mutagenesis and complementation analysis of *Rhizobium meliloti* symbiotic nitrogen fixation genes Cell. **29**:551-9.

Sat, B., R. Hazan, T. Fisher, H. Khaner, G. Glaser, H.J. Engelberg-Kulka. (2001) Programmed cell death in *Escherichia coli*: some antibiotics can trigger *mazEF* lethality. Bacteriol. **183**:2041-5.

Schlaman, H.L., D.A. Phillips, E. Kondorosi. (1998*)* In *The Rhizobiaceae, eds* Spaink H P, Kondorosi A, Hooykaas P J J *(*Kluwer, Dordrecht, The Netherlands*)*, pp:361–386.

Sevin, E.W., and F. Barloy-Huber. (2007). RASTA-Bacteria: a web-based tool for identifying toxin-antitoxin loci in prokaryotes. Genome Biol. **8**:R155.

Shao, Y., E.M. Harrison, D. Bi, C. Tai, X. He, H.Y. Ou, K. Rajakumar, Z. Deng. (2011) TADB: a web-based resource for Type 2 toxin-antitoxin loci in bacteria and archaea. Nucleic Acids Res. 39 (Database issue):D606-11.

Sletvold, H., P.J. Johnsen, I. Hamre, G.S. Simonsen, A. Sundsfjord, K.M. Nielsen. (2008) Complete sequence of *Enterococcus faecium* pVEF3 and the detection of an omega-epsilon-zeta toxin-antitoxin module and an ABC transporter. Plasmid. **60**:75-85.

Szekeres**,** S**.**, Dauti**,** M**.**, C. Wilde, D. Mazel, D.A. Rowe-Magnus. (2007) Chromosomal toxin-antitoxin loci can diminish large-scale genome reductions in the absence of selection. Mol**.** Microbiol. **63**:1588-605.

Tian, Q. B., M. Ohnishi, A. Tabuchi, Y. Terawaki. (1996) A new plasmid-encoded proteic killer gene system: cloning, sequencing, and analyzing *hig* locus of plasmid Rts1. Biochem. Biophys. Res. Commun. **220:**280-284.

Tsilibaris, V., G. Maenhaut-Michel, N. Mine, L. Van Melderen. (2007) What is the benefit to *Escherichia coli* of having multiple toxin-antitoxin systems in its genome? J. Bacteriol. **189**:6101-8

Wilbur, J.S., P.T. Chivers, K. Mattison, L. Potter, R.G. Brennan, M. So. (2005) *Neisseria gonorrhoeae* FitA interacts with FitB to bind DNA through its ribbon–helix–helix motif. Biochemistry **44**:12515–12524.

Wolfgang, M., J.P. van Putten, S.F. Hayes, D. Dorward, M. Koomey. (2000) Components and dynamics of fiber formation define a ubiquitous biogenesis pathway for bacterial pili. EMBO J. **19**:6408-18.

Yuan, J., Y. Yamaichi, M.K. Waldor. (2011) The three *vibrio cholerae* chromosome II-encoded *ParE* toxins degrade chromosome I following loss of chromosome II. J Bacteriol. **193**:611-9.

Zhang, Y.X., X.K. Guo, C. Wu, B. Bi, S.X. Ren, C.F. Wu, G.P. Zhao. (2004) Characterization of a novel toxin–antitoxin module, VapBC, encoded by *Leptospira interrogans* chromosome. Cell Res. **14**:208–216.

Table 5.1. The list of predicted solo toxin and antitoxin genes as well as TA pairs on the *Sinorhizobium meliloti* chromosome identified by Makarova *et al.*, 2009.

Gene	Previously characterized and new candidateTA systems	Gene	Previously characterized and new candidateTA systems
SMc00036	COG3832 (toxin)		
SMc00075	GNAT (toxin)		
SMc00080	ArsR (antitoxin)		
SMc00089	Xre (antitoxin)		
SMc00097	COG3832 (toxin)		
SMc00178	RelE (toxin)		
SMc00348	RHH (antitoxin)		
SMc00351	Xre (antitoxin)	SMc00352	GNAT (toxin)
SMc00377	Xre (antitoxin)		
SMc00383	COG3832 (toxin)		
SMc00392	PHD (antitoxin)	SMc00393	PIN (toxin)
SMc00397	GNAT (toxin)		
SMc00407	COG3832 (toxin)		
SMc00429	Xre (antitoxin)		
SMc00479	Xre (antitoxin)		
SMc00558	Xre (antitoxin)		
SMc00605	COG3832 (toxin)		
SMc00687	PIN (toxin)	SMc00686	AbrB (antitoxin)
SMc00694	RelE (toxin)	SMc00693	RHH (antitoxin)
SMc00740	RHH (antitoxin)	SMc00739	GNAT (toxin)
SMc00769	COG2856		
SMc00769	Xre (antitoxin)		
SMc00822	RelE (toxin)	SMc00823	Xre (antitoxin)
SMc00900	PIN (toxin)	SMc00899	RHH (antitoxin)
SMc01110	Xre (antitoxin)		
SMc01132	COG3832 (toxin)		
SMc01149	COG3832 (toxin)		
SMc01177	PIN (toxin)	SMc01176	AbrB (antitoxin)
SMc01210	HEPN (antitoxin)		
SMc01226	ArsR (antitoxin)		
SMc01229	GNAT (toxin)		
SMc01238	COG3832 (toxin)		
SMc01269	Xre (antitoxin)		
SMc01452	RHH (antitoxin)		
SMc01462	GNAT (toxin)		
SMc01610	Xre (antitoxin)		
SMc01636	Xre (antitoxin)		

Gene	Previously characterized and new candidateTA systems	Gene	Previously characterized and new candidateTA systems
SMc01748	PHD (antitoxin)		
SMc01762	ArsR (antitoxin)		
SMc01842	ArsR (antitoxin)		
SMc01857	AbrB (antitoxin)		
SMc01932	PIN (toxin)	SMc01933	RHH (antitoxin)
SMc01957	Xre (antitoxin)		
SMc01995	RelE (toxin)	SMc01836	Xre (antitoxin)
SMc02151	Xre (antitoxin)		
SMc02184	Xre (antitoxin)		
SMc02223	Xre (antitoxin)		
SMc02226	COG3832 (toxin)		
SMc02313	HicB (toxin)		
SMc02394	Xre (antitoxin)	SMc02461	RelE (toxin)
SMc02477	Xre (antitoxin)	SMc02476	RelE (toxin)
SMc02609	Xre (antitoxin)		
SMc02620	Xre (antitoxin)		
SMc02638	COG3832 (toxin)		
SMc02639	ArsR (antitoxin)		
SMc02647	ArsR (antitoxin)		
SMc02658	AbrB (antitoxin)	SMc02657	PIN (toxin)
SMc02710	Fic (toxin)		
SMc02715	PIN (toxin)	SMc02716	RHH (antitoxin)
SMc02757	HipA (antitoxin)		
SMc02783	Xre (antitoxin)		
SMc02865	GNAT (toxin)		
SMc02887	RHH (antitoxin)		
SMc02888	Xre (antitoxin)		
SMc02982	GNAT (toxin)		
SMc02987	PIN (toxin)	SMc02988	RHH (antitoxin)
SMc03137	PIN (toxin)	SMc03136	PHD (antitoxin)
SMc03151	GNAT (toxin)		
SMc03170	Xre (antitoxin)		
SMc03206	COG3832 (toxin)		
SMc03840	GNAT (toxin)		
SMc03842	COG3832 (toxin)	SMc03843	ArsR (antitoxin)
SMc03949	AbrB (antitoxin)	SMc01521	PIN (toxin)
SMc03998	ArsR (antitoxin)		
SMc04161	COG3832 (toxin)	SMc04162	ArsR (antitoxin)
SMc04169	Xre (antitoxin)		
SMc04198	Xre (antitoxin)		
SMc04220	ArsR (antitoxin)		

Gene	Previously characterized and new candidateTA systems	Gene	Previously characterized and new candidateTA systems
SMc04221	GNAT (toxin)	SMc04222	RHH (antitoxin)
SMc04312	RHH (antitoxin)	SMc04313	RelE (toxin)
SMc04321	COG3832 (toxin)		
SMc04387	Xre (antitoxin)		
SMc04431	COG2929 (toxin)		
SMc04436	RelE (toxin)		
SMc04441	HicA (antitoxin)	SMc04269	HicB (toxin)
SMc04881	PIN (toxin)	SMc04408	AbrB (antitoxin)
SMc05001	HicB (toxin)		
SMc05006	AbrB (antitoxin)		
SMc05009	ArsR (antitoxin)		

Genes labelled in red were also predicted using RASTA method (Sevin *et al*., 2007). Genes labelled in green were also predicted using TADB (toxin-antitoxin database approach) (Shao *et al*., 2010).

Table 5.2. The list of predicted solo toxin and antitoxin genes as well as TA pairs on the *Sinorhizobium meliloti* pSymA megaplasmid identified by Makarova *et al*., 2009.

Gene	Previously characterized and new candidateTA systems	Gene	Previously characterized and new candidateTA systems
SMa0151	PIN (toxin)		
SMa0191	COG3832 (toxin)	SMa0193	ArsR (antitoxin)
SMa0285	PIN (toxin)	SMa0286	COG2442 (antitoxin)
SMa0319	Xre (antitoxin)		
SMa0453	PIN (toxin)		
SMa0471	RHH (antitoxin)	SMa0473	RelE (toxin)
SMa0545	PIN (toxin)	SMa0548	RHH (antitoxin)
SMa0572	RelE (toxin)		
SMa0592	HipA (toxin)	SMa0594	Xre (antitoxin)
SMa0917	MazF (toxin)		
SMa0967	HEPN (antitoxin)		
SMa0981	PIN (toxin)		
SMa1056	ArsR (antitoxin)		
SMa1076	Xre (antitoxin)		
SMa1253	MNT (toxin)		
SMa1413	RHH (antitoxin)		
SMa1455	COG5654 (antitoxin)	SMa1456	COG5642 (toxin)
SMa1476	Xre (antitoxin)		
SMa1497	COG3832 (toxin)		
SMa1725	Xre (antitoxin)		
SMa1749	Xre (antitoxin)		
SMa1770	RelE (toxin)		
SMa1822	Xre (antitoxin)		
SMa1823	Xre (antitoxin)		
SMa1825	Xre (antitoxin)		
SMa1924	Xre (antitoxin)		
SMa1990	RHH (antitoxin)		
SMa2105	Fic (toxin)		
SMa2133	Xre (antitoxin)		
SMa2151	Xre (antitoxin)		
SMa2163	Xre (antitoxin)		
SMa2231	PIN (toxin)		
SMa2253	PIN (toxin)	SMa2255	PHD (antitoxin)
SMa2273	RelE (toxin)	SMa2275	RHH (antitoxin)
SMa2279	COG5654 (antitoxin)	SMa2281	COG5642 (toxin)
SMa2315	MNT (toxin)		
SMa2319	COG3832 (toxin)		

Gene	Previously characterized and new candidateTA systems	Gene	Previously characterized and new candidateTA systems
SMa5006	PIN (toxin)	SMa5007	Xre (antitoxin)
SMa5008	RelE (toxin)		

Genes labelled in red were also predicted using RASTA method (Sevin *et al.*, 2007).

Table 5.3. The list of predicted solo toxin and antitoxin genes as well as TA pairs or triplets on the *Sinorhizobium meliloti* pSymB megaplasmid identified by Makarova *et al*., 2009.

Gene	Previously characterized and new candidateTA systems	Gene	Previously characterized and new candidateTA systems	Gene	Previously characterized and new candidate TA systems
SMb20005	COG3832 (toxin)				
SMb20062	RHH (antitoxin)	SMb20063	RelE (toxin)		
SMb20215	Xre (antitoxin)				
SMb20222	Xre (antitoxin)				
SMb20256	COG3832 (toxin)				
SMb20411	PIN (toxin)				
SMb20412	PIN (toxin)	SMb20413	RHH (antitoxin)		
SMb20420	COG3832 (toxin)				
SMb20607	COG3832 (toxin)	SMb20608	ArsR (antitoxin)		
SMb20626	Xre (antitoxin)				
SMb20627	RelE (toxin)	SMb20628	RHH (antitoxin)	SMb20629	RHH (antit.
SMb20695	COG5654 (antitoxin)	SMb20696	COG5642 (toxin)		
SMb20754	COG2856 (antitoxin)				
SMb20754	Xre (antitoxin)				
SMb20835	HEPN (antitoxin)				
SMb20835	MNT (toxin)				
SMb20859	Xre (antitoxin)				
SMb20935	Xre (antitoxin)				
SMb21007	COG3832 (toxin)	SMb21008	ArsR (antitoxin)		
SMb21035	Xre (antitoxin)				
SMb21117	Xre (antitoxin)				
SMb21127	COG5642 (antitoxin)	SMb21128	COG5654 (toxin)		
SMb21153	GNAT (toxin)				
SMb21169	ArsR (antitoxin)				
SMb21187	Xre (antitoxin)				
SMb21336	Xre (antitoxin)				
SMb21419	Xre (antitoxin)				
SMb21475	MazE (antitoxin)	SMb21476	MazF (toxin)		
SMb21509	RHH (antitoxin)	SMb21510	PIN (toxin)	SMb21511	PIN (toxin)
SMb21559	Xre (antitoxin)				
SMb21576	ArsR (antitoxin)				
SMb21649	Xre (antitoxin)				
SMb21651	PIN (toxin)				
SMb21670	COG3832 (toxin)				
SMb22004	RelE (toxin)	SMb20121	RHH (antitoxin)		

Gene	Previously characterized and new candidateTA systems	Gene	Previously characterized and new candidateTA systems	Gene	Previously characterized and new candidate TA systems
SMb22021	GNAT (toxin)				

Genes labelled in red were also predicted using RASTA method (Sevin *et al*., 2007).

Table 5.4. Construction of RmP1099 (ΔA150), RmP1092 (A151), RmP1408 (ΔA152) and RmP2681 (ΔA160) strains.

Paren strain	Deletions used to generate parental strain	End points of potential deletion strain	Deletion strain	Deletion size
A150	ΦΔA127(Gm^rNm^r)→ΔA106	184,519 nt – 930,910 nt	RmP1099 (ΔA150)	746,391 nt
A151	ΦΔA129(Gm^rNm^r)→ΔA106	184,519 nt – 1,123,504 nt	n.g.	n.a.
A152	ΦΔA129(Gm^rNm^r)→ΔA117	400,267 nt – 1,123,504 nt	RmP1408 (ΔA152)	723,237 nt
A160	ΦΔA105(Gm^rNm^r)→ΔA102	9,549 nt – 186,200 nt	RmP2681 (ΔA160)	176,651 nt

n.g – not recovered
n.a. – not applicable

Figure 5.1. FRT-directed phenotype/genotype resulting from deletion of ΔIJ into the *S. meliloti* genome between AB and KL. AB, CD, EF, GH, IJ, KL are random DNA sequences that were cloned into targeting vectors. Gm^r- gentamicin resistance; Nm^r- neomycin resistance; FRT-*flp* recombinase target site; reporter genes *gusA*, *lacZ*, *GFP* (green fluorescence protein); *RFP* (red fluorescence protein).

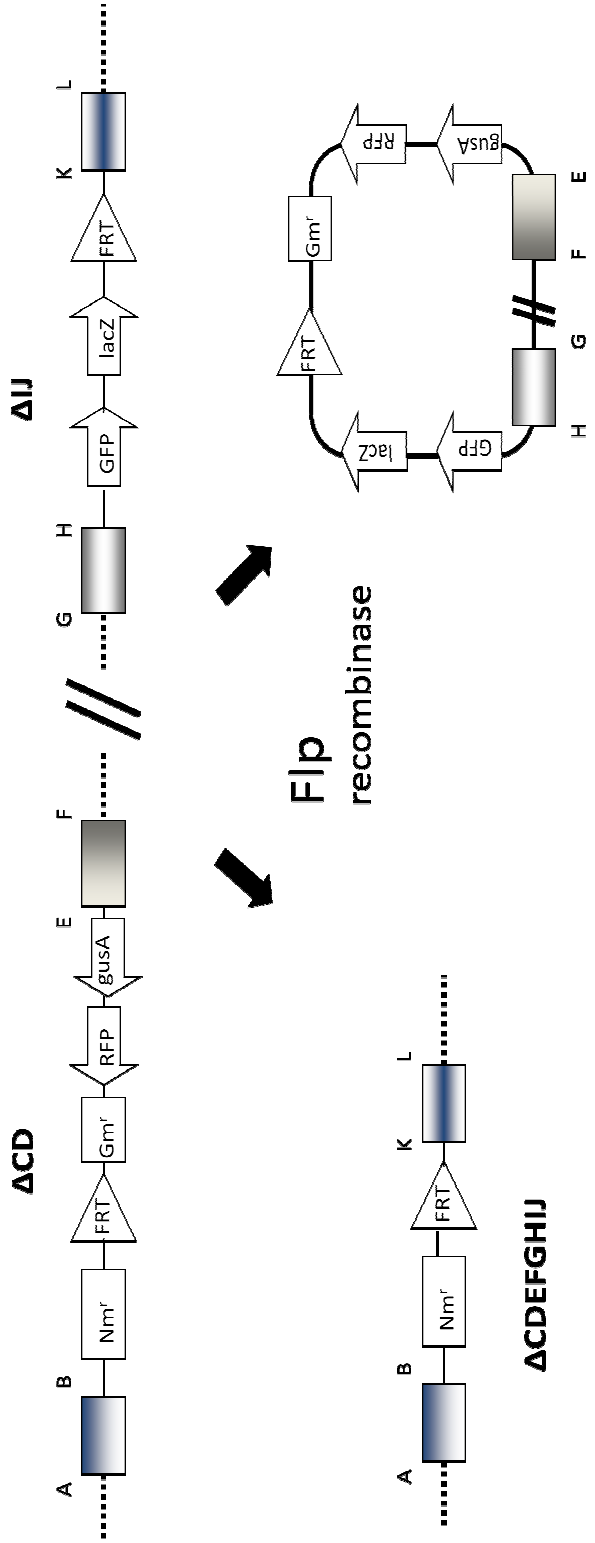

ΔCD
A
B
Nmr
FRT
Gmr
RFP
gusA
E
F
ΔIJ
G
H
GFP
lacZ
FRT
K
L
Flp
recombinase
Gmr
FRT
lacZ
GFP
gusA
RFP
E
F
G
H
ΔCDEFGHIJ

Table 5.5. Primer sequences used for construction of pTH2563, pTH2622, pTH2623, pTH2624, pTH2625, pTH2646 and pTH2647 plasmids as well as locations of primers on pSymA megaplasmid.

Plasmid name	Gene	Primer name and sequence		Location of primers on pSymB	PCR prod. size
pTH2563	*sma0471/ sma0473*	TApacF	CCTTAATTAACAACTGTCCACCGCGCTTCG	257565nt	1272bp
		TApacR	CCTTAATTAAGAGTGAGGCTGCTTCTACCC	258837nt	
pTH2622	*sma2151*	sma2151F	CCTTAATTAACCGGAGATGATTGCCAA	1214806nt	839bp
		sma2151R	CCTTAATTAATCCTTCTGCGGCAGTGTCC	1215645nt	
pTH2623	*sma2231*	sma2231F	CCTTAATTAACGCTCTATCCGGTCGA	1248755nt	978bp
		sma2231R	CCTTAATTAAGCAAGAGCAACGGCCATGG	1249733nt	
pTH2624	*sma2253/ sma2255*	sma53/5F	CCTTAATTAATCGCGCTGCACAAACAA	1262710nt	1303bp
		sma53/5R	CCTTAATTAACATCGCGGCAGCCTGTCTTCG	1264012nt	
pTH2625	*sma2273/ sma2275*	sma73/5F	CCCAAGCTTGAAAGCACAACTGGACGCGG	1269637nt	1403bp
		sma73/5R	CCTTAATTAAGTCCTGGTCGTCACCATCC	1268234nt	
pTH2646	*sma2105*	sma105F	CCTTAATTAAAACAGCATCGTGCCCT	1187240nt	1859bp
		sma105R	CCTTAATTAACGTTCGTTGACCTCCTA	1189099nt	
pTH2647	*sma2133*	sma133F	CCTTAATTAAGGCAAGGATGCGGAAG	1206839nt	1414bp
		sma133R	CCTTAATTAATCGGGTCGATGTGACT	1205425nt	

All plasmids are pTH1931 (Sp^r) derivatives.

Figure 5.2. Circular map of pSymA megaplasmid of *S. meliloti*, showing FRT-flanked regions, defined on pSymA (A) generated in this study. Green regions – Flp-transconjugants recovered and deletions (Δ) confirmed. Red regions – Flp-transconjugants not recovered.

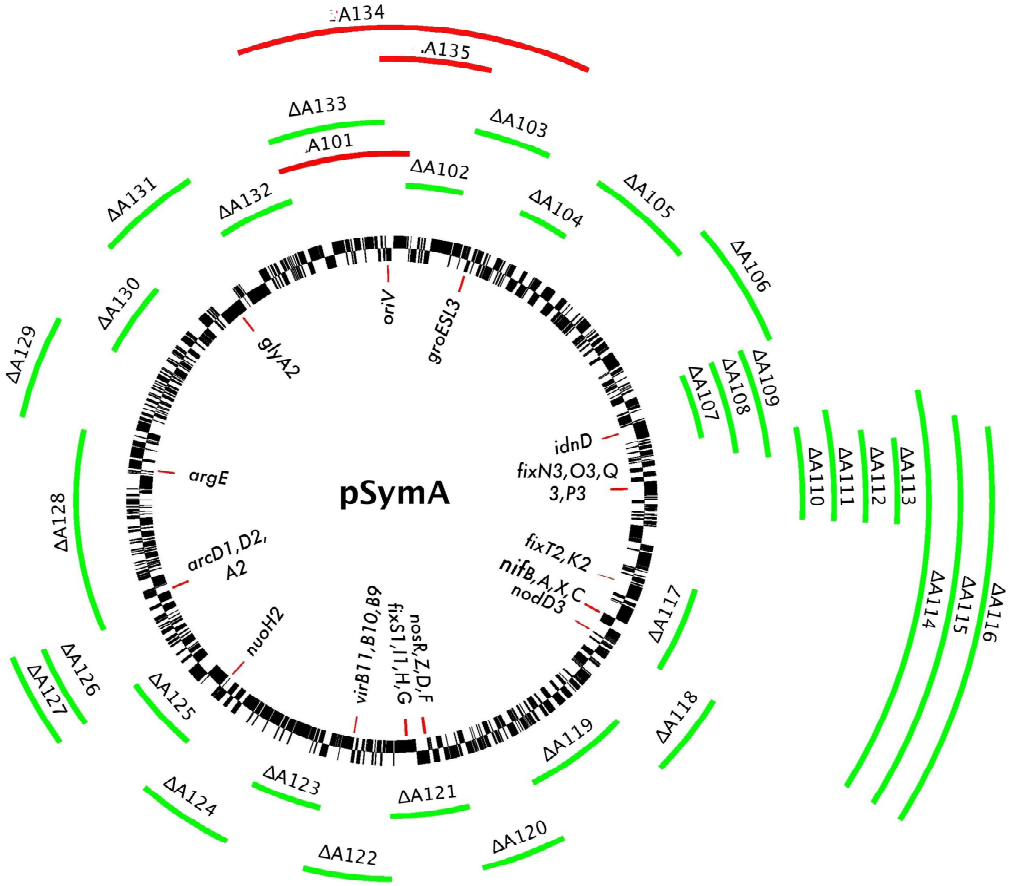

pSymA
oriV
groESL3
glyA2
idnD
fixN3,O3,Q
3,P3
fixT2,K2
nifB,A,X,C
nodD3
nosR,Z,D,F
fixS1,I1,H,G
virB11,B10,B9
nuoH2
arcD1,D2,
A2
argE
ΔA101
ΔA102
ΔA103
ΔA104
ΔA105
ΔA106
ΔA107
ΔA108
ΔA109
ΔA110
ΔA111
ΔA112
ΔA113
ΔA114
ΔA115
ΔA116
ΔA117
ΔA118
ΔA119
ΔA120
ΔA121
ΔA122
ΔA123
ΔA124
ΔA125
ΔA126
ΔA127
ΔA128
ΔA129
ΔA130
ΔA131
ΔA132
ΔA133
ΔA134
ΔA135

Figure 5.3. Circular map of pSymA megaplasmid of *S. meliloti*, showing FRT-flanked regions, defined on pSymA (A) generated in this study. Green regions – Flp-transconjugants recovered and deletions (Δ) confirmed. Red regions – Flp-transconjugants not recovered. Yellow regions – deletions and their single FRT sites (black dots) that were used to combine into single strains Deletions presented here were used to identify presence of genes on pSymA megaplasmid essential for cell survival.

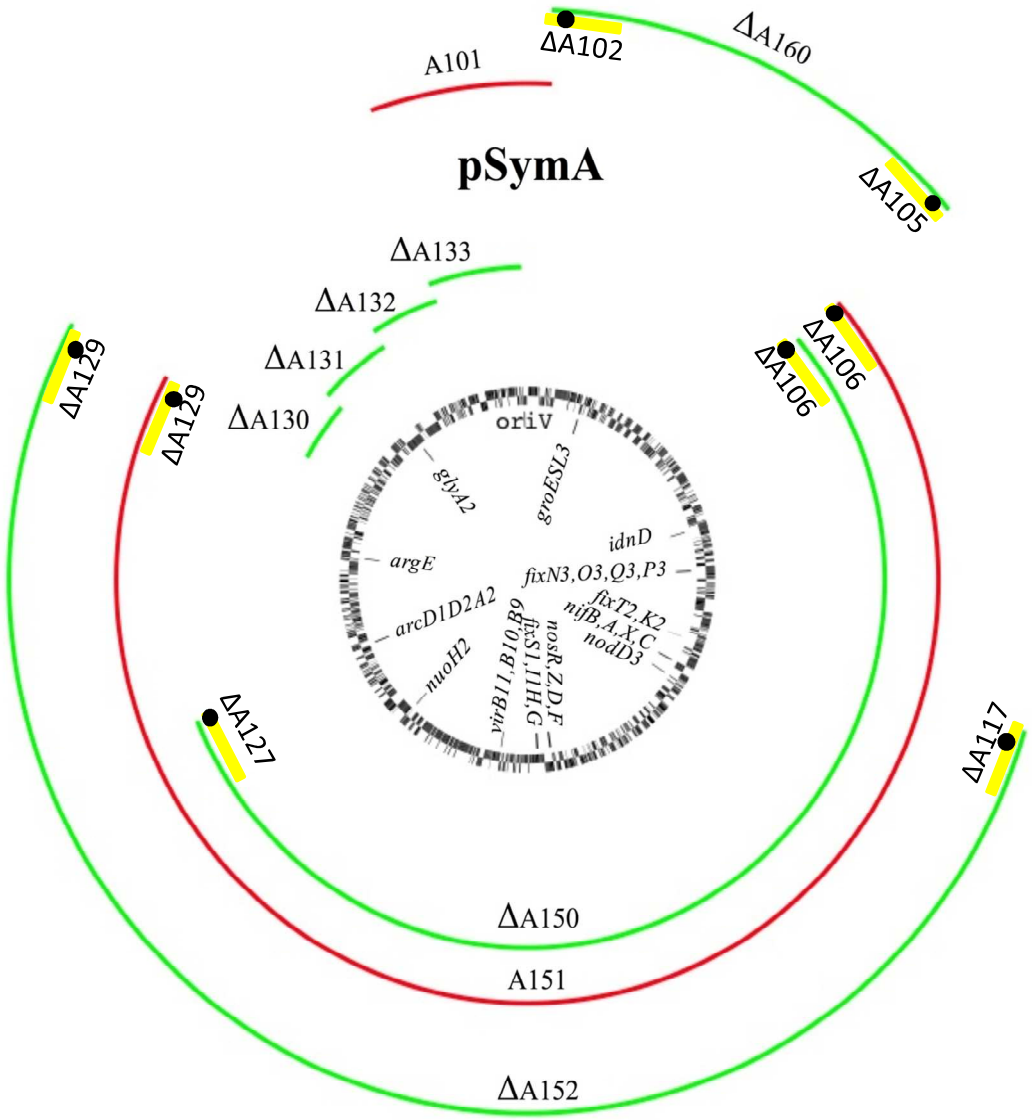

pSymA
A101
ΔA102
ΔA160
ΔA105
ΔA133
ΔA132
ΔA131
ΔA130
ΔA129
ΔA129
ΔA106
ΔA106
oriV
groESL3
idnD
fixN3,O3,Q3,P3
fixT2,K2
nifB,A,X,C
nodD3
fixS1,I1,H1,G
nosR,Z,D,F
virB11,B10,B9
nuoH2
arcD1D2A2
argE
glyA2
ΔA127
ΔA117
ΔA150
A151
ΔA152

Figure 5.4. Transduction of deletions (ΔA) from transconjugant deletion strains into the wild type (WT) recipient using phage ΦM12. Successful transduction indicated absence of toxin/antitoxin like systems in the deleted region. An inability to transduce some deletions suggested the possible presence of toxin/antitoxin (TA) like systems.

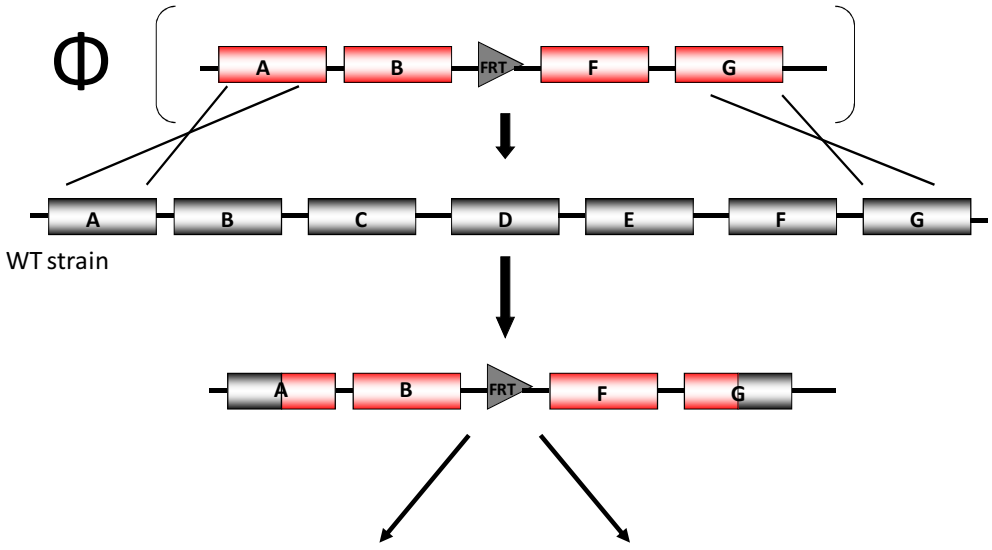

Φ
A
B
FRT
F
G
A
B
C
D
E
F
G
WT strain
A
B
FRT
F
G
Reduced transduction frequency - possible presence of toxin/antitoxin like systems in a deleted region
Wild type transduction frequency– absence of toxin/antitoxin like systems in a deleted region

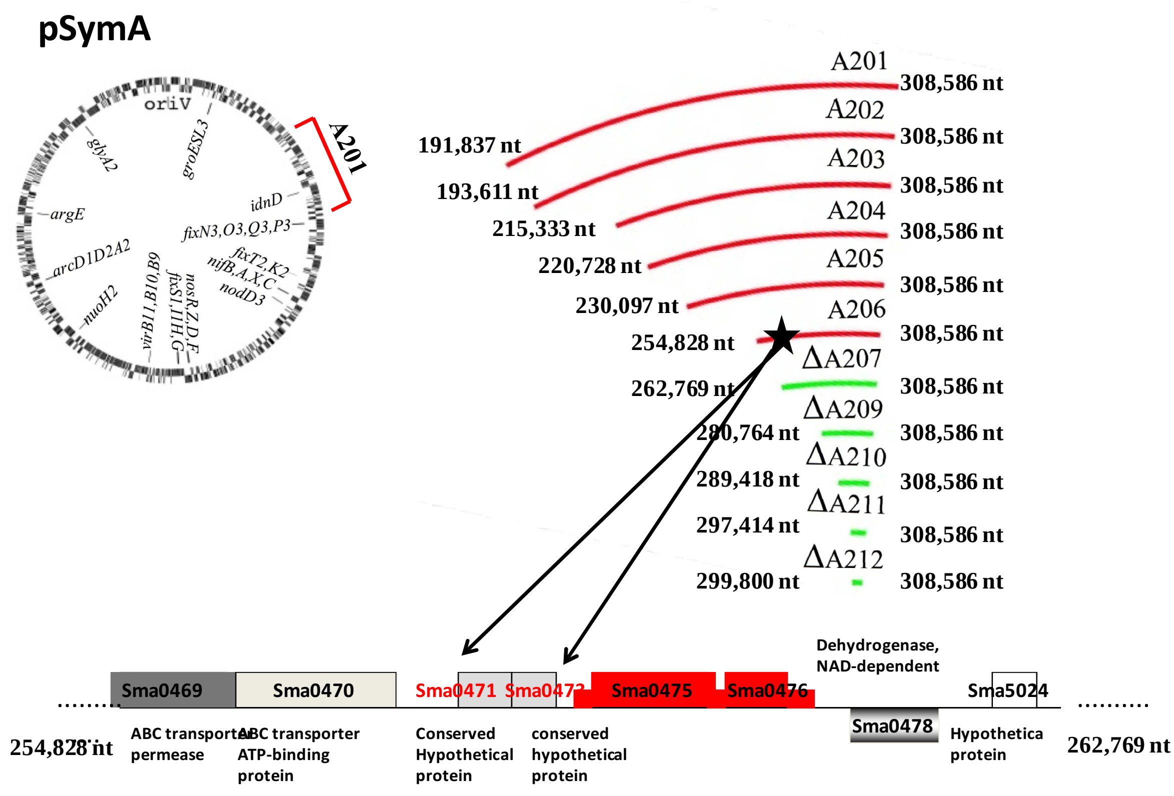
Table 5.6. Transduction of ΔA130, ΔA131, ΔA132, ΔA133 ΔA150, ΔA152 and ΔA160 into wild type (WT) *S. meliloti* P110 using phage ΦM12.

	Transductant per plate (transductional frquency per recipient)
ΦΔA130→ WT	39 ($6.19x10^{-8}$)
ΦFL3976 → WT	52 ($8.25x10^{-8}$)
ΦΔA131 → WT	0 ($< 10^{-9}$)
ΦFL 4094→ WT	69 ($1.09x10^{-7}$)
ΦΔA132 → WT	0 ($< 10^{-9}$)
ΦFL 4094→ WT	60 ($9.52x10^{-8}$)
ΦΔA133 → WT	18 ($2.85x10^{-8}$)
ΦFL4393→ WT	113 ($1.8x10^{-7}$)
ΦΔA150 → WT	0 ($< 10^{-9}$)
ΦFL449→ WT	84 ($1.33x10^{-7}$)
ΦΔA152 → WT	19 ($3.01x10^{-8}$)
ΦFL5575→ WT	102 ($1.61x10^{-7}$)
ΦΔA160→ WT	23 ($3.65x10^{-8}$)
ΦFL449→ WT	84 ($1.33x10^{-7}$)

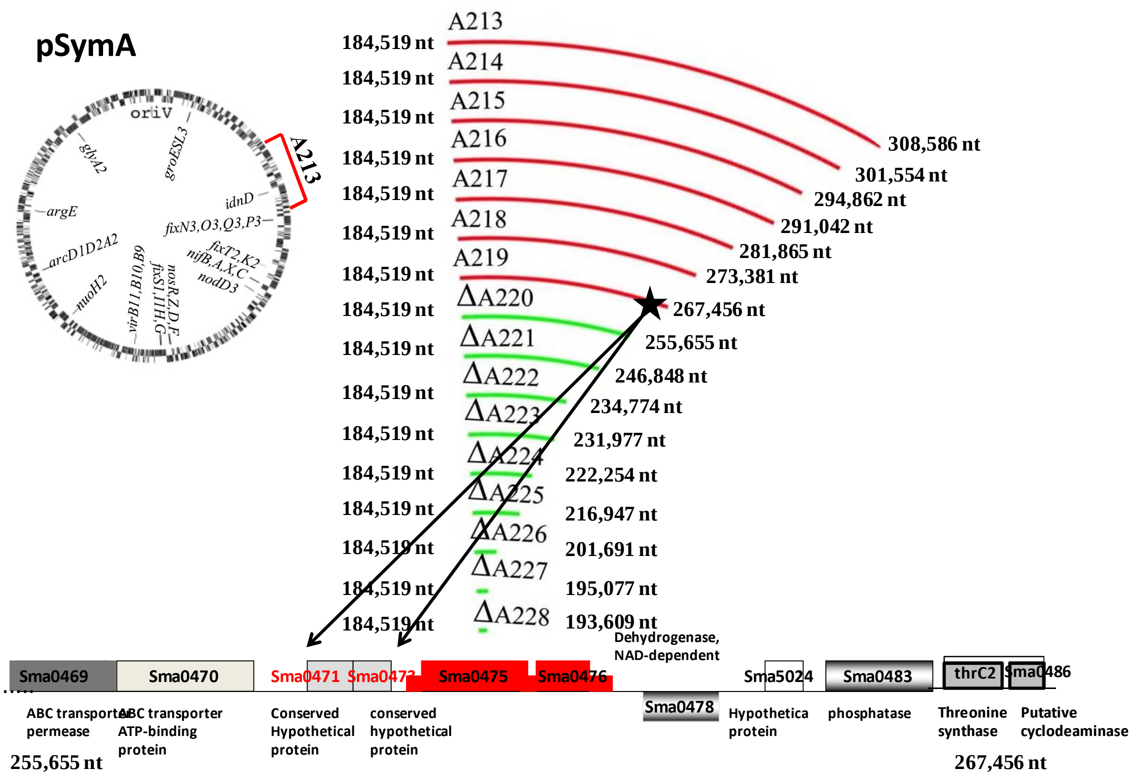
Strain FL3976, FL4094, FL4393, FL449, FL5575 carrying one of the ΔA130, ΔA131, ΔA132, ΔA133, ΔA150, ΔA152 and ΔA160 FRT borders were used as a transduction controls. Data represents mean values of four plates.

Figure 5.5. Circular maps of pSymA megaplasmid of *S. meliloti*, showing FRT-flanked regions, defined on pSymA (A) generated in this study. Green regions – Flp-transconjugants recovered and deletions (Δ) confirmed. Red regions – Flp-transconjugants not recovered. Within the region 184,519 to 308,586 nt, five sets of deletions were made to narrow down the region with the possible gene/genes which has effect on *S. meliloti* cell survival. Deletion sets C), D) and E) indicate that region responsible for the cell viability lies within 255,655 – 264,432 nt. A) deletion set narrow down essential region to 254,921 – 262,769 while according to deletion set B) essential gene/genes are within 255,655 – 267,456 nt region.

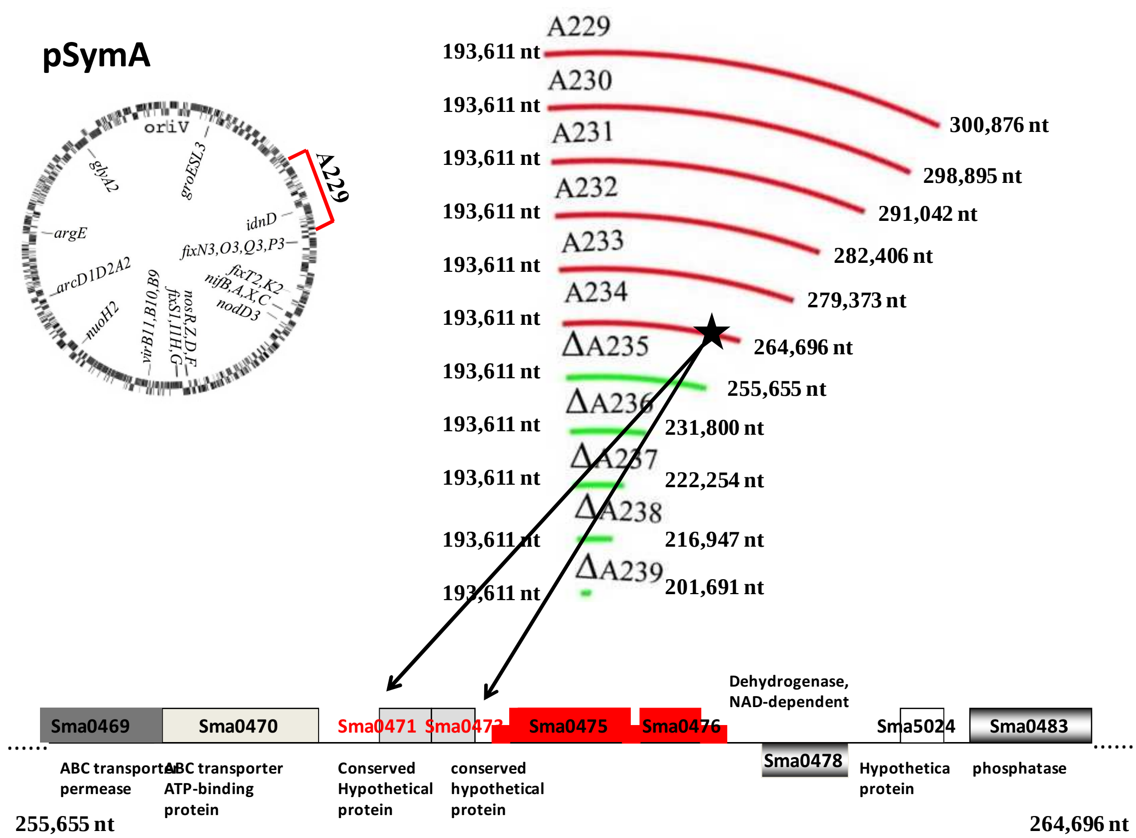
A)


pSymA
oriV
groESL3
glyA2
argE
idnD
fixN3,O3,Q3,P3
arcD1D2A2
fixT2,K2
nifB,A,X,C
nodD3
nuoH2
virB11,B10,B9
fixS1,I1H,G
nosR,Z,D,F
A201
A201
191,837 nt
308,586 nt
A202
193,611 nt
308,586 nt
A203
215,333 nt
308,586 nt
A204
220,728 nt
308,586 nt
A205
230,097 nt
308,586 nt
A206
254,828 nt
308,586 nt
ΔA207
262,769 nt
308,586 nt
ΔA209
280,764 nt
308,586 nt
ΔA210
289,418 nt
308,586 nt
ΔA211
297,414 nt
308,586 nt
ΔA212
299,800 nt
308,586 nt
Sma0469
Sma0470
Sma0471
Sma0473
Sma0475
Sma0476
Sma0478
Sma5024
Dehydrogenase, NAD-dependent
254,828 nt
ABC transporter permease
ABC transporter ATP-binding protein
Conserved Hypothetical protein
conserved hypothetical protein
Hypothetica protein
262,769 nt

B)

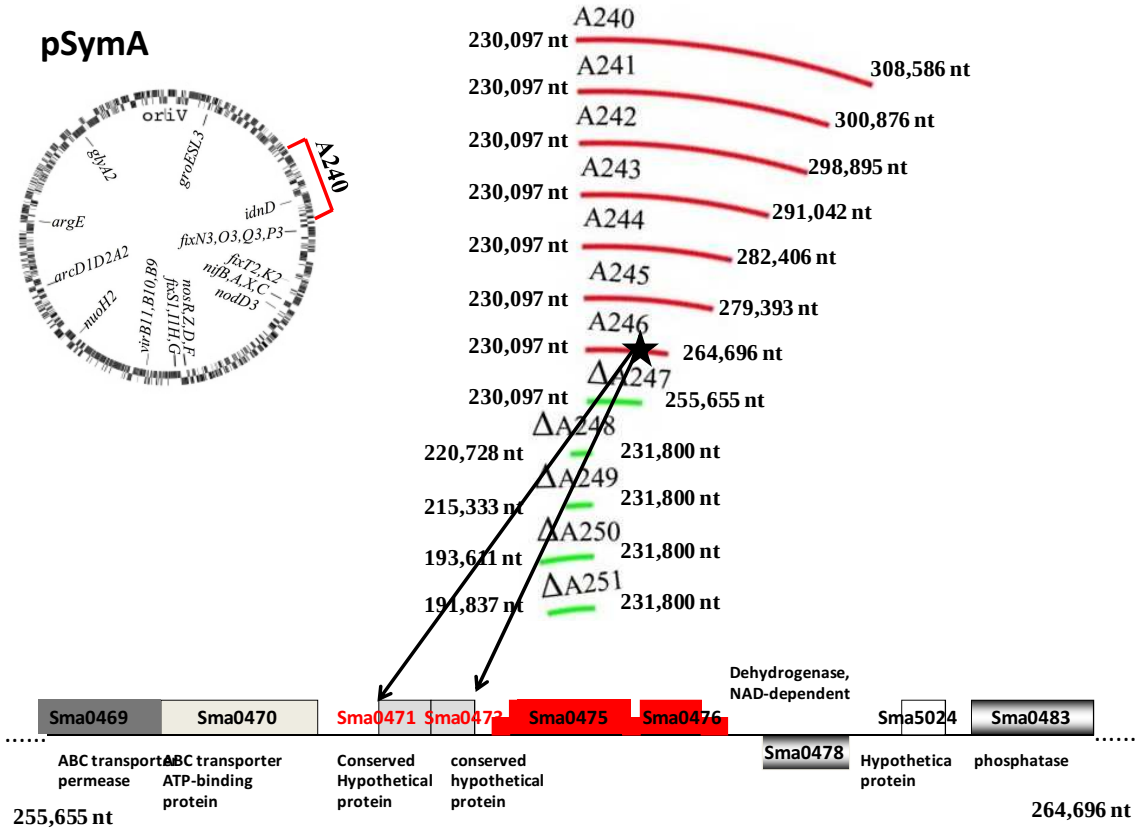

pSymA
oriV
A213
184,519 nt
A214
A215
A216
A217
A218
A219
ΔA220
ΔA221
ΔA222
ΔA223
ΔA224
ΔA225
ΔA226
ΔA227
ΔA228
308,586 nt
301,554 nt
294,862 nt
291,042 nt
281,865 nt
273,381 nt
267,456 nt
255,655 nt
246,848 nt
234,774 nt
231,977 nt
222,254 nt
216,947 nt
201,691 nt
195,077 nt
193,609 nt
Sma0469
Sma0470
Sma0471
Sma0475
Sma0476
Dehydrogenase, NAD-dependent
Sma0478
Sma0483
thrC2
ABC transporter permease
ABC transporter ATP-binding protein
Conserved Hypothetical protein
conserved hypothetical protein
phosphatase
Threonine synthase
Putative cyclodeaminase
255,655 nt
267,456 nt

C)



pSymA
oriV
groESL3
glyA2
argE
idnD
fixN3,O3,Q3,P3
arcD1D2A2
fixT2,K2
nifB,A,X,C
nodD3
nuoH2
virB11,B10,B9
fixS1,I1H,G
nosR,Z,D,F
A229
A229
A230
A231
A232
A233
A234
ΔA235
ΔA236
ΔA237
ΔA238
ΔA239
193,611 nt
300,876 nt
298,895 nt
291,042 nt
282,406 nt
279,373 nt
264,696 nt
255,655 nt
231,800 nt
222,254 nt
216,947 nt
201,691 nt
Sma0469
Sma0470
Sma0471
Sma0472
Sma0475
Sma0476
Sma0478
Sma5024
Sma0483
ABC transporter permease
ABC transporter ATP-binding protein
Conserved Hypothetical protein
conserved hypothetical protein
Dehydrogenase, NAD-dependent
Hypothetica protein
phosphatase
255,655 nt
264,696 nt

D)


pSymA
oriV
A240
gltA2
groESL3
argE
idnD
fixN3,O3,Q3,P3
arcD1D2A2
fixT2,K2
nifB,A,X,C
nodD3
nuoH2
virB11,B10,B9
fixS1,I1,H1,G1
nosR,Z,D,F
A240
230,097 nt
308,586 nt
A241
230,097 nt
300,876 nt
A242
230,097 nt
298,895 nt
A243
230,097 nt
291,042 nt
A244
230,097 nt
282,406 nt
A245
230,097 nt
279,393 nt
A246
230,097 nt
264,696 nt
ΔA247
230,097 nt
255,655 nt
ΔA248
220,728 nt
231,800 nt
ΔA249
215,333 nt
231,800 nt
ΔA250
193,611 nt
231,800 nt
ΔA251
191,837 nt
231,800 nt
Dehydrogenase, NAD-dependent
Sma0469
Sma0470
Sma0471
Sma0472
Sma0475
Sma0476
Sma5024
Sma0483
Sma0478
ABC transporter permease
ABC transporter ATP-binding protein
Conserved Hypothetical protein
conserved hypothetical protein
Hypothetica protein
phosphatase
255,655 nt
264,696 nt

E)

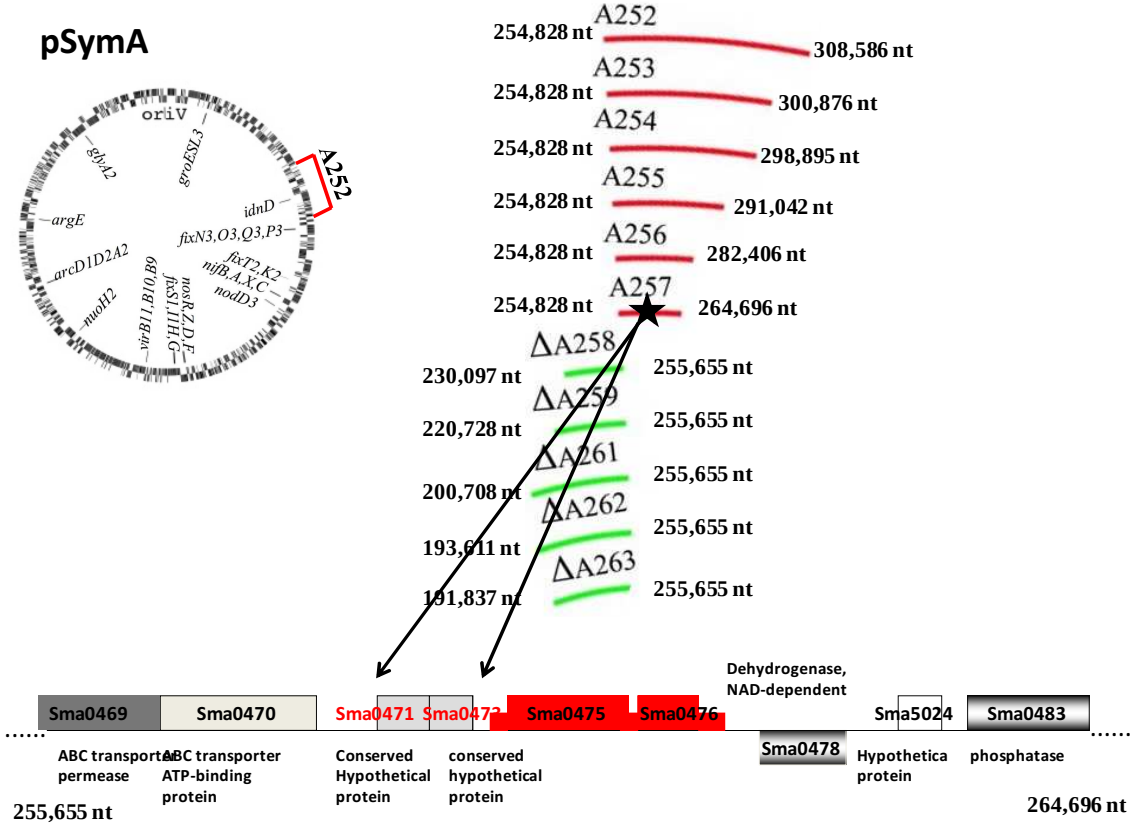

pSymA
oriV
A252
254,828 nt
308,586 nt
A253
254,828 nt
300,876 nt
A254
254,828 nt
298,895 nt
A255
254,828 nt
291,042 nt
A256
254,828 nt
282,406 nt
A257
254,828 nt
264,696 nt
ΔA258
230,097 nt
255,655 nt
ΔA259
220,728 nt
255,655 nt
ΔA261
200,708 nt
255,655 nt
ΔA262
193,611 nt
255,655 nt
ΔA263
191,837 nt
255,655 nt
Dehydrogenase, NAD-dependent
Sma0469
Sma0470
Sma0471
Sma0472
Sma0475
Sma0476
Sma0478
Sma5024
Sma0483
ABC transporter permease
ABC transporter ATP-binding protein
Conserved Hypothetical protein
conserved hypothetical protein
Hypothetica protein
phosphatase
255,655 nt
264,696 nt

Figure 5.6. 9 kb region from the pSymA region with the predicted toxin/ antitoxin gene pair *sma0471*/*sma0473*. Genes present in this region are annotated as: *sma0469*-ABC transporter, permease; *sma0470*-ABC transporter, ATP-binding protein; *sma0471*-putative antitoxin; *sma0473*-putative plasmid stabilization system; *sma0475*-TRm17a transposase; *sma0476*-TRm17b transposase; *sma0478*-putative formate dehydrogenase; *sma5024*- hypothetical protein; *sma0483*- phosphatase (Galibert *et al*., 2001).

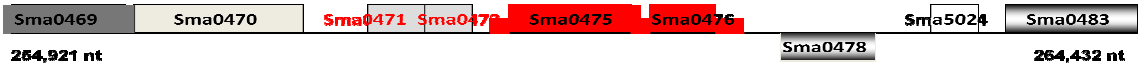

Sma0469
Sma0470
Sma0471
Sma0472
Sma0475
Sma0476
Sma0478
Sma5024
Sma0483
264,921 nt
264,432 nt

Figure 5.7. Alignments of the RelB and RelE homologues found by BLAST searches.

A) Alignments of the ten predicted RelB homologues. The homologues have the following Entrez accession numbers (percentage of identity): *Rhizobium etli* CNPAF512 – EGE_61323.1 (89%), *Sinorhizobium medicae* WSM419 – YP_001328891.1 (79%), *Methylobacterium nodulans* ORS 2060 – YP_002490079.1 (83%), *Nitrobacter hamburgensis* X14 – YP_572052.1 (80%), *Sinorhizobium meliloti* 1021 – NP_435493.1 (100%), *Rhizobium leguminosarum bv. viciae* 3841 – YP_768107.1 (87%), *Agrobacterium tumefaciens* str. C58 – NP_353955.1 (59%), *Sinorhizobium fredii* NGR234 – YP_002823109.1 (45%), *Pseudomonas syringae pv. phaseolicola* 1448A – YP_272311.1 (66%), *Escherichia coli* str. K-12 – NP_414761.1 (57%).

B) Alignments of the nine predicted RelE homologues that have the following Entrez accession numbers (percentage of identity): *Rhizobium etli* CFN 42 – YP_472199 (43%), *Rhizobium leguminosarum bv. viciae* 3841 – YP_002985102 (45%), *Escherichia coli* E1167 – EFZ_76192 (42%), *Methylobacterium nodulans* ORS 2060 – YP_002490080 (66%), *Nitrobacter hamburgensis* X14 – YP_572053 (64%), *Sinorhizobium meliloti* 1021 – NP_435494 (100%), *Agrobacterium tumefaciens* str. C58 – NP_356195 (49%), *Sinorhizobium medicae* WSM419 YP_001328890 (57%), *Sinorhizobium fredii* NGR234 – NP_443945 (46%)

A)

```
R.etliCNPAF512                  ----------MAANALVQTRIDAEVRDRASAVLENMGLTVSDAVRILLTR 40
S.medicaeWSM419                 ----------MAANALVQTRINAGVRDRASAVLESMGLTVSDVVRILLTR 40
M.nodulansORS2060               ----------MPANALVQTRIDADVRDRASAVLENMGLTVSDAVRILLTR 40
N.hamburgensisX14               ----------MAANALVQTRIDADVRDRASAVLEGLGLTVSDAVRILLTR 40
S.meliloti1021                  ----------MASNALVQTRIDAEVKERATAVLENMGLTVSDAVRILLTR 40
R.leguminosarumbv.viciae3841    ----------MASNALVQTRIDAAVKDRATAVLENMGLTVSDVVRILLTR 40
A.tumefaciensC58                ----------MTANAYVRARIDQTLKDDATAVLDRLGLTVSDVMRMMLTR 40
S.frediiNGR234                  ----------MAANAYVRARIEPSVKDSAALVLDSLGLTTSDIIRIVLTR 40
P.syringae1448A                 MWLTPLEEEIMSANAVVRARIDEHIKEEATAVLAAMGLTVSDAFRIMLTR 50
E.coliK-12                      ----------MAANAFVRARIDEDLKNQAADVLAGMGLTISDLVRITLTK 40
                                          *.:** *::**:  ::: *: **  :*** ** .*: **:

R.etliCNPAF512                  TANEGALPLELFSGSEAHDAWFRTKVLEALNDTRPDVSDDEVEAQFAKRR 90
S.medicaeWSM419                 TANEGVLPLELVTGSEAHDAWFRAKVLQALNDTRPDVPDHEAEAHFAEPR 90
M.nodulansORS2060               TANEGALPLELVTNSEAHDAWFRAKVLEALDDTRPDVPDDQAEAHFAQRR 90
N.hamburgensisX14               TANEGALPLELVTNGEAHDAWFRAKVQEALEDTRPDVSDRQADAYFASRR 90
S.meliloti1021                  TANEGALPLELFSHSEAHDAWFRAKVLRALEDTRPDVDDADADAHFRERR 90
R.leguminosarumbv.viciae3841    TANEGALPLELISNSDAYDAWFRGKVLEALHDTRPDVDDTEVETGFEKRR 90
A.tumefaciensC58                IAREKALPIELTQPN--------AETLAAIEEARAIAAAGRNRFGTSEAL 82
S.frediiNGR234                  IARDKALPVELTRPN--------AKTIAAMEEARAIKAGHGKHFDTAGEL 82
P.syringae1448A                 VAREKALPFEPLVPN--------ATTIEAMKEAR---RGGLKSFASVEDL 89
E.coliK-12                      VAREKALPFDLREPN--------QLTIQSIKNSE--AGIDVHKAKDADDL 80
                                 *.: .**.:    .          .  ::.::.

R.etliCNPAF512                  AAARLNAGSRKS 102
S.medicaeWSM419                 AAARHKAGNQNS 102
M.nodulansORS2060               AAARRRAGELKS 102
N.hamburgensisX14               AAARKKADAKS- 101
S.meliloti1021                  AAALRKAAAGDR 102
R.leguminosarumbv.viciae3841    EAALRKSQVDRS 102
A.tumefaciensC58                FEALDAGKR--- 91
S.frediiNGR234                  FESLEGHKVVK- 93
P.syringae1448A                 MADLNADD---- 97
E.coliK-12                      FDKLGI------ 86
```

B)

```
R.etliCNPAF512                  ---MIVSWLQQTLLDREHQLRHVFAQNPKAAIALDDVIRHQAKMLADHPE 47
R.leguminosarumbv.viciae3841    ---MIVKWLQQALLDREGQIRHIFAQNPKAAIALDDVIRHQAKMLADHPE 47
E.coliE1167                     --MMEIFWTILASQDRKCIRGYITEQNLMAAIELDERIGYSASSLAGQPY 48
M.nodulansORS2060               ---MKLVWSAFALSDRGGIFTHIEADNPAAAIVIDERIVAATRRLRDFPE 47
N.hamburgensisX14               ---MKLLWSAFALADRDGIFTHFEADNPRAAIVIDERIAAAVRLLVDFPE 47
S.meliloti1021                  ---MRLVWARYALDDRDTIFSYIERENPRAAVHVDEEIVSAVRRLLDFPE 47
A.tumefaciensC58                MSDRRIRWTLRALRRLDEIGAHIEQDNPAAAARVISRIVSAADMLVEQPA 50
S.medicaeWSM419                 ---MIVEFSDEAESDLEQIADYIAKDNPRLALSFVQELRAKCERLGDTP- 46
S.frediiNGR234                  ---MKLEWTSKAVSDIGRLYDFLAPVNRQAAVRTVQSLTSAPTRLLEQPR 47
                                     : :   :         ..   *   *    . :      *   *

R.etliCNPAF512                  AGRS------GRLEGTRELVIPRTAFLLIYRIDKKAQRVEILRLLHGAQH 91
R.leguminosarumbv.viciae3841    AGRG------GRLGGTRELVIPRTAFLLIYRIDRKAQRVEILRLLHGAQQ 91
E.coliE1167                     KGRN------GRVEGIRELVIY-PHFVLVYEIDSQWGKVYILRVLHTAQK 91
M.nodulansORS2060               SGRP------GRIAGTRELVVTGTPYIAAYQVTGET--VRILRVLHGAQQ 89
N.hamburgensisX14               SGRLGRLGRLGRVVGTRELVINDTPYIAAYIVTENA--IRILRVLHGTQR 95
S.meliloti1021                  SGRP------GRIAGTRELVIPRTPYIAAYMVMEDR--IRILRVLHGAQK 89
A.tumefaciensC58                IGRV------GRIKGTREAVLSDISYIIAYRVGRD---IEILTIIHTSRR 91
S.medicaeWSM419                 NGFP----LVPRYEEYGIRRRVHSSYLIFYRVEGEQ--IVIVHVLHGAMN 90
S.frediiNGR234                  IGER-----LDEFDPREVRRILIGHYEMRYEIRQST--IYVLRLWHTREE 90
                                 *         .            :   * :  .   : :: : *   .

R.etliCNPAF512                  WPPKR--- 96
R.leguminosarumbv.viciae3841    WPPKR--- 96
E.coliE1167                     WP------ 93
M.nodulansORS2060               WPDDLPEG 97
N.hamburgensisX14               WPDELPEG 103
S.meliloti1021                  WPSELDDG 97
A.tumefaciensC58                WPSAL--- 96
S.medicaeWSM419                 YAAILFES 98
S.frediiNGR234                  R------- 91
```

Table 5.7. FRT- directed deletion of the A257 region in the presence of putative toxin - antitoxin genes *sma0471/sma0473*.

	Conjugational frequency LB - IPTG	Conjugational frequency LB + IPTG	Complementation
A257 pTH1931	$4.4x\ 10^{-3}$	$2.3x10^{-4}$	n.a
A257 *sma0471/73*	$5.5x10^{-3}$	$1.0x10^{-2}$	+
A101 *sma0471/73*	$1.3x10^{-2}$	$1.0x10^{-4}$	-
B158 pTH1931	$1.1x10^{-2}$	$2.1x10^{-4}$	n.a
A128pTH1931	$1.4x10^{-2}$	$1.1x10^{-2}$	n.a

We attempted to make deletions using *S. meliloti* strains: A257 (region flanked by FRT sites, harbors *sma0471* and *sma0473*); A257 (region flanked by FRT sites, harbors *sma0471* and *sma0473*) complemented with *sma0471/73*, A101 (region flanked by FRT sites harbors pSymA *oriV*) complemented with *sma0471/73*; A128 (region on pSymA flanked by FRT sites that previously was successfully deleted) and B158 (region flanked by FRT sites harbors essential gene for $tRNA^{arg}$). Transconjugants were selected on LB media with (+) and without (-) IPTG. Successful complementation was indicated with +, lack of complementation was indicated with-.
n.a. - not applicable.

Table 5.8. Transduction of ΔA257 into wild type (WT) *S. meliloti* P110 with or without *sma0471/ sma0473* genes *in trans* on the replicating pTH1931 plasmid.

	Transductant per plate (transductional frquency per recipient)	Complementation
ΦΔA257 → WT (pTH1931 – *sma0471/73*)	120 ($1.44x10^{-7}$)	+
ΦΔA257 → WT (pTH1931 –)	2 ($1.9x10^{-9}$)	n.a.
ΦFL5920 → WT (pTH1931 – *sma0471/73*)	105 ($1.43x10^{-7}$)	n.a.
ΦFL5920 → WT (pTH1931 –)	112 ($1.71x10^{-7}$)	n.a.

Strain FL4094 carrying one of the A257 FRT borders was used as a transduction control. +, Complementation; -, No Complementation; not applicable, n.a. Data represents mean values of four plates.

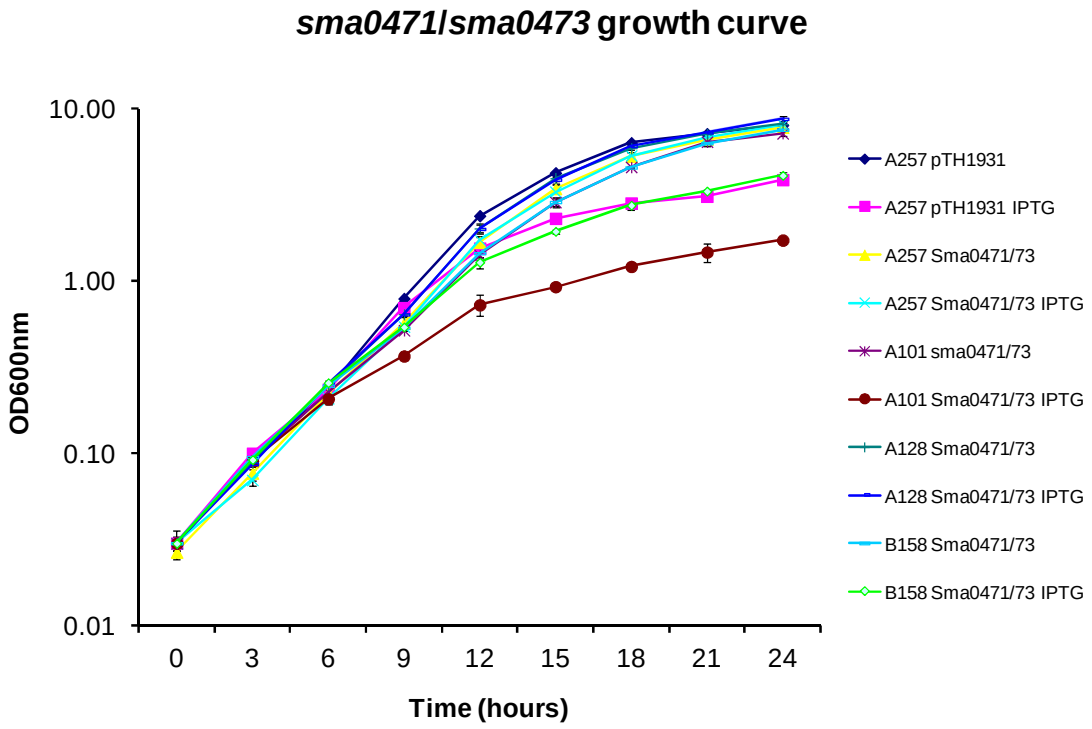
Figure 5.8. The *sma0471/sma0473* genes of *S. meliloti* constitute a TA locus.

S. meliloti strain A257 carries FRT sites flanking a 24.5 kb region that includes the predicted toxin/antitoxin pair *sma0471*/s*ma0473. sma0471/sma0473* on pTH1931 were introduced *in trans* into A257 and deletion of the 24.5 kb region was achieved upon induction of *flp* gene by 0.5 mM IPTG.

A) Growth of strains A257 (pTH1931), A257 (pTH1931-*sma0471/sma0473*) and three control strains in LB medium with or without IPTG. In strain A101 the pSymA *oriV* is flanked by FRT sites. In strain A128 FRT sites flank a non-essential pSymA region. Strain B158 carries FRT sites flanking an essential $tRNA^{Arg}$ gene on pSymB.

B) Viable counts (CFU/ml) of cultures with or without the addition of IPTG.

A)


sma0471/sma0473 growth curve
OD600nm
10.00
1.00
0.10
0.01
0
3
6
9
12
15
18
21
24
Time (hours)
A257 pTH1931
A257 pTH1931 IPTG
A257 Sma0471/73
A257 Sma0471/73 IPTG
A101 sma0471/73
A101 Sma0471/73 IPTG
A128 Sma0471/73
A128 Sma0471/73 IPTG
B158 Sma0471/73
B158 Sma0471/73 IPTG

B)

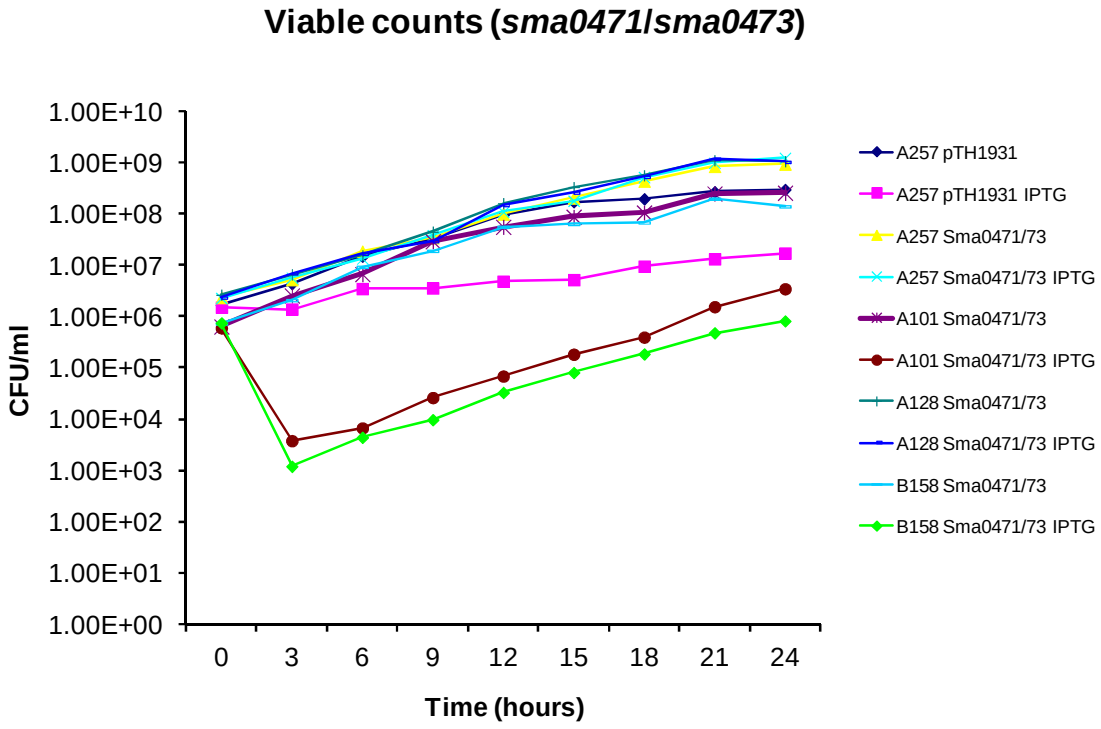

Viable counts (sma0471/sma0473)
CFU/ml
1.00E+10
1.00E+09
1.00E+08
1.00E+07
1.00E+06
1.00E+05
1.00E+04
1.00E+03
1.00E+02
1.00E+01
1.00E+00
0
3
6
9
12
15
18
21
24
Time (hours)
A257 pTH1931
A257 pTH1931 IPTG
A257 Sma0471/73
A257 Sma0471/73 IPTG
A101 Sma0471/73
A101 Sma0471/73 IPTG
A128 Sma0471/73
A128 Sma0471/73 IPTG
B158 Sma0471/73
B158 Sma0471/73 IPTG

Figure 5.9. Part of the 40 kb A131 region harboring the predicted toxin *sma2105*. Genes present in this region are annotated as: *sma2101*-nitrilotriacetate monooxygenase component A; *sma2103*-oxidoreductase; *sma2105*-conserved hypothetical protein; *sma2107*-transcriptional regulator; *sma2109*-conserved hypothetical protein; *sma2111*-hypothetical protein; *adhC2*- glutathione-dependent dehydrogenase; *sma2125*-ABC transporter, permease; *sma2127*-ABC transporter, ATP-binding protein; *sma2129*-conserved hypothetical protein; *sma2131*-conserved hypothetical protein; *sma2133*-hypothetical protein; *glyA2*-serine hydroxymethyltransferase; *sma2137*-dehydrogenase; *sgaA*-serine-glyoxylate aminotransferase; *sma2145*-aminomethyltransferase; *sma2147*-hypothetical protein; *sma2151*-putative DNA-binding protein; *sma2157*-oxidoreductase; *sma2159*-conserved hypothetical protein; *sma2163*-transcriptional regulator; *sma2165*-short chain alcohol dehydrogenase (Galibert *et al*., 2001).

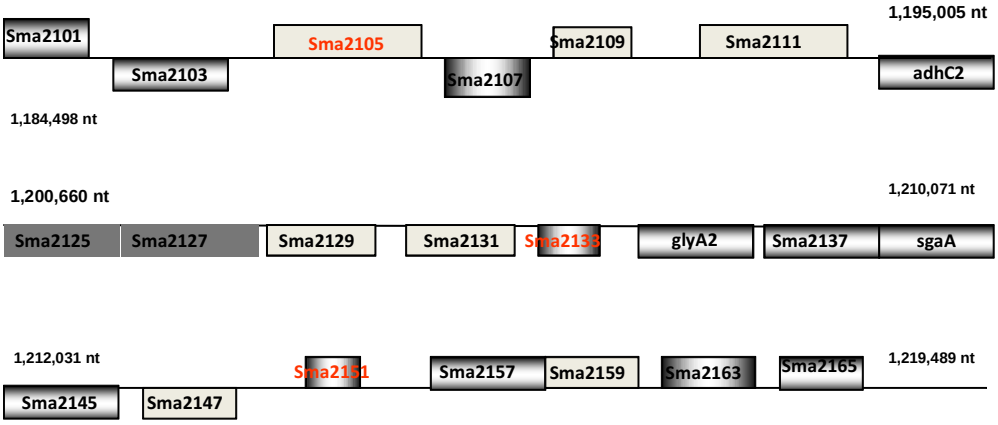

1,195,005 nt
Sma2101
Sma2105
Sma2109
Sma2111
Sma2103
Sma2107
adhC2
1,184,498 nt
1,200,660 nt
1,210,071 nt
Sma2125
Sma2127
Sma2129
Sma2131
Sma2133
glyA2
Sma2137
sgaA
1,212,031 nt
Sma2151
Sma2157
Sma2159
Sma2163
Sma2165
1,219,489 nt
Sma2145
Sma2147

Table 5.9. FRT- directed deletion of the A131 region in the presence of putative toxin - antitoxin genes *sma2105*, *sma2133* and *sma2151*.

	Conjugational frequency LB - IPTG	Conjugational frequency LB +IPTG	Complementation
A131 pTH1931	$4.1x10^{-2}$	$1.1x10^{-4}$	n.a.
A131 *sma2105*	$4.1x10^{-2}$	$5.2x10^{-2}$	+
A131 *sma2133*	$1.3x10^{-2}$	$1.2x10^{-3}$	-
A131 *sma2151*	$3.2x10^{-2}$	$1.2x10^{-4}$	-
B158 pTH1931	$1.1x10^{-2}$	$2.1x10^{-4}$	n.a.
A128 pTH 1931	$1.4x10^{-2}$	$1.1x10^{-2}$	n.a.

A131 is defined by FRT sites located 1,173,115 - 1,232,916 nt on pSymA. The region includes genes *sma2105*, *sma2133* and sma2151. A128 (region on pSymA flanked by FRT sites that previously was successfully deleted) and B158 (region flanked by FRT sites harbors essential gene for $tRNA^{arg}$). Transconjugants were selected on LB media with (+) and without (-) IPTG. Successful complementation was indicated with +, lack of complementation was indicated with - n.a. - not applicable.

Table 5.10. Transduction of ΔA131 into *S. meliloti* P110 (WT, wild type) with plasmid pTH1931 containing *sma2105,* or *sma2133* or *sma2151*. These genes are predicted members of toxin/antitoxin families (Makarova *et al.*, 2009).

	Transductant per plate (transductional frquency per recipient)	Complementation
ΦΔA131 → WT (pTH1931 – *sma2105*)	99 (4.32×10^{-7})	+
ΦΔA131 → WT (pTH1931 – *sma2133*)	0 ($< 10^{-9}$)	-
ΦΔA131 → WT (pTH1931 – *sma2151*)	0 ($< 10^{-9}$)	-
ΦΔA131 → WT (pTH1931 –)	0 ($< 10^{-9}$)	n.a.
ΦFL4094 → WT (pTH1931 – *sma2105*)	84 (3.66×10^{-7})	n.a.
ΦFL4094 → WT (pTH1931 – *sma2133*)	80 (8.7×10^{-8})	n.a.
ΦFL4094 → WT (pTH1931 – *sma2151*)	54 (4.5×10^{-8})	n.a.
ΦFL4094 → WT (pTH1931 –)	54 (5.25×10^{-8})	n.a.

Strain FL4094, containing one of the FRT borders used to make ΔA131 was a control donor. +, Complementation; -, No Complementation , n.a., not applicable. Data represents mean values of four plates.

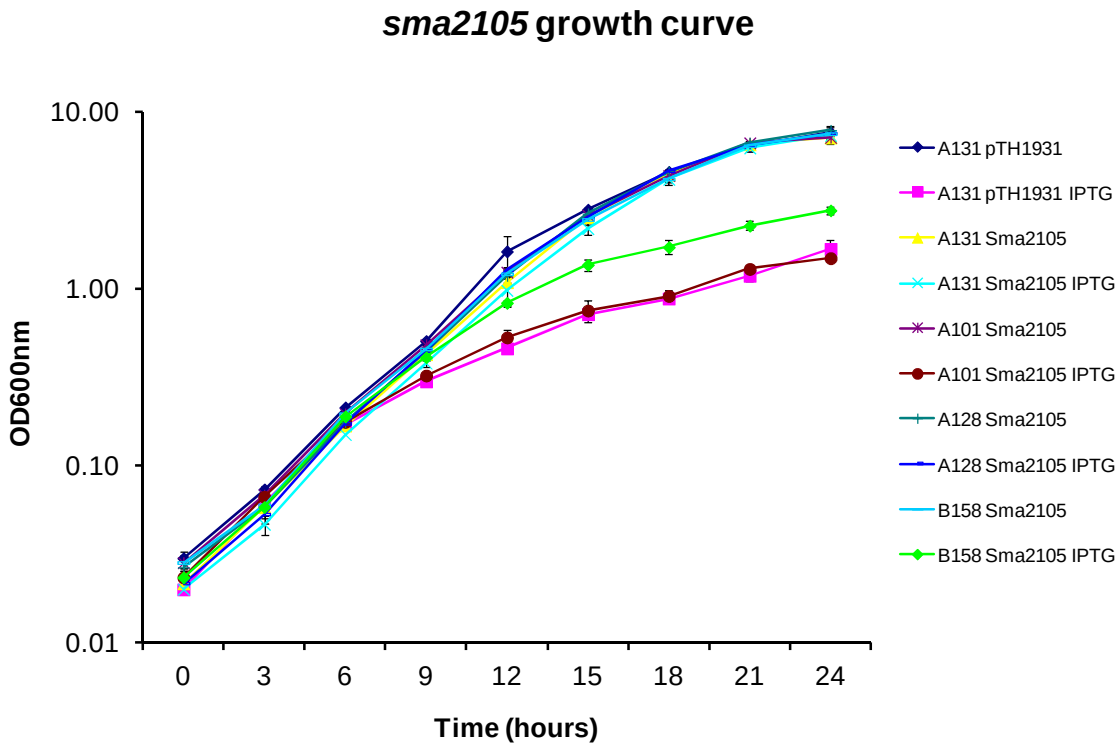
Figure 5.10. The *sma2105* gene of *S. meliloti* encodes a toxin.

The *sma2105* gene was cloned in pTH1931 and conjugated into strain A131 in which FRT sites flank a 40 kb region that includes *sma2105*. Deletion of the 40 kb region was achieved upon induction of the *flp* gene by 0.5 mM IPTG.

A) Growth of A131 (pTH1931),), A131 (pTH1931-*sma2105*), and three control strains in LB medium with or without IPTG. In strain A101 the pSymA *oriV* is flanked by FRT sites. In strain A128 FRT sites flank a non-essential pSymA region. Strain B158 carries FRT sites flanking an essential $tRNA^{Arg}$ gene on pSymB.

B) Viable cell counts (CFU/ml) in all cultures, with or without addition of IPTG.

A)


sma2105 growth curve
10.00
1.00
0.10
0.01
OD600nm
0
3
6
9
12
15
18
21
24
Time (hours)
A131 pTH1931
A131 pTH1931 IPTG
A131 Sma2105
A131 Sma2105 IPTG
A101 Sma2105
A101 Sma2105 IPTG
A128 Sma2105
A128 Sma2105 IPTG
B158 Sma2105
B158 Sma2105 IPTG

B)

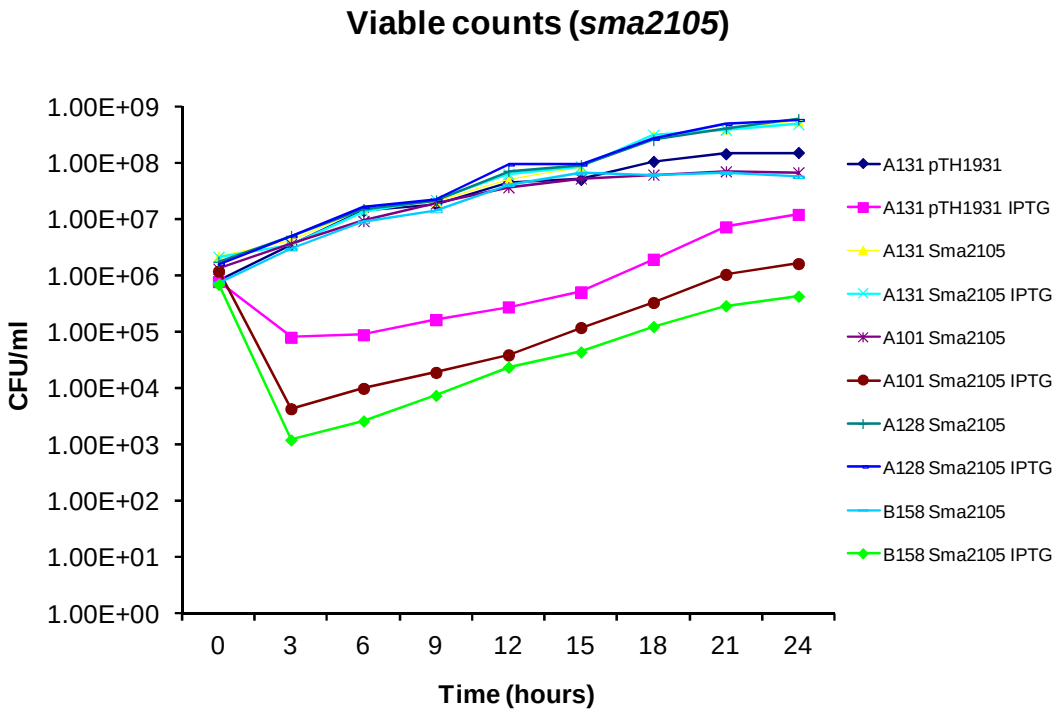

Viable counts (*sma2105*)
1.00E+09
1.00E+08
1.00E+07
1.00E+06
1.00E+05
1.00E+04
1.00E+03
1.00E+02
1.00E+01
1.00E+00
CFU/ml
0
3
6
9
12
15
18
21
24
Time (hours)
A131 pTH1931
A131 pTH1931 IPTG
A131 Sma2105
A131 Sma2105 IPTG
A101 Sma2105
A101 Sma2105 IPTG
A128 Sma2105
A128 Sma2105 IPTG
B158 Sma2105
B158 Sma2105 IPTG

Figure 5.11. Part of the 40 kb A132 region harboring the predicted antitoxin/toxin pair *sma2230*/*sma2231*. Genes present in this region are annotated as: *sma2229*-conserved hypothetical protein; *sma2230*- newly identified open reading frame within the intergenic region *sma2229*-*sma2231* that is homologues to antitoxin genes; *sma2231*-conserved hypothetical protein; *sma2233*-hypothetical protein; *sma2235*-hypothetical protein; *sma2237*-hypothetical protein; *sma2239*- conserved hypothetical protein; *sma2241*-conserved hypothetical protein; *sma2245*-conserved hypothetical protein; *sma2249*-hypothetical protein; *sma2251*-hypothetical protein; *sma2253*-conserved hypothetical protein; *sma2255*-hypothetical protein; *sma2257*-integrase/recombinase; *sma2259*-conserved hypothetical protein; *sma2263*- hypothetical protein; *sma2265*-hypothetical protein; *sma2267*-conserved hypothetical protein; *sma2269*-hypothetical protein; *sma2273*-conserved hypothetical protein; *sma2275*-conserved hypothetical protein; *sma2279*-conserved hypothetical protein; *sma2281*-conserved hypothetical protein; *sma2285*-hypothetical protein; *sma2287*-transcriptional regulator; *sma2289*-conserved hypothetical protein; *sma2291*-dehydrogenase (Galibert *et al*., 2001).

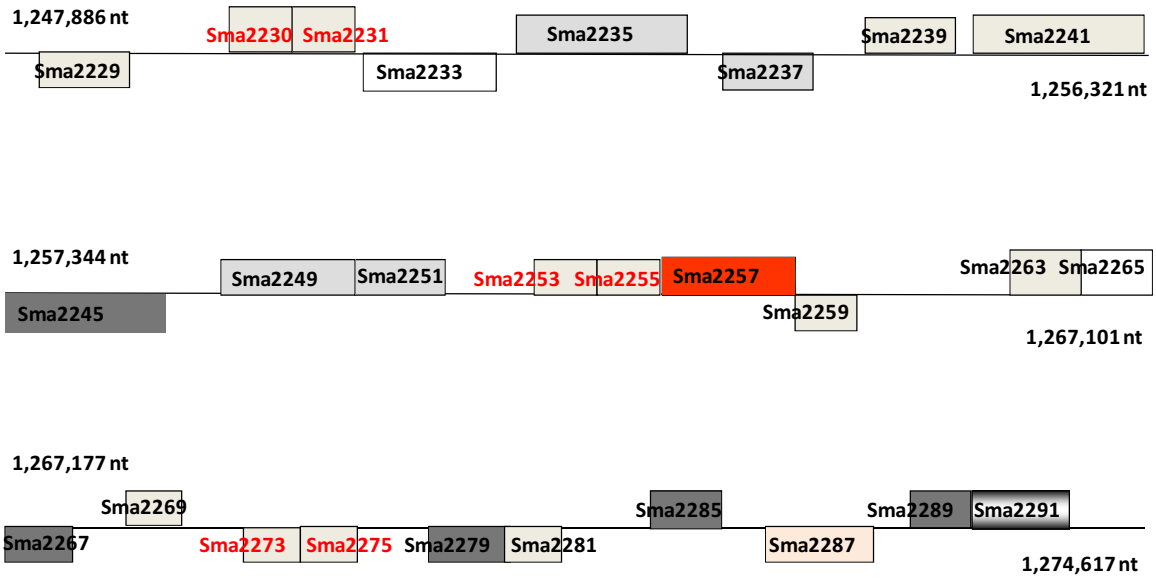

1,247,886 nt
Sma2230
Sma2231
Sma2235
Sma2239
Sma2241
Sma2229
Sma2233
Sma2237
1,256,321 nt
1,257,344 nt
Sma2249
Sma2251
Sma2253
Sma2255
Sma2257
Sma2263
Sma2265
Sma2245
Sma2259
1,267,101 nt
1,267,177 nt
Sma2269
Sma2285
Sma2289
Sma2291
Sma2267
Sma2273
Sma2275
Sma2279
Sma2281
Sma2287
1,274,617 nt

Table 5.11. FRT - directed deletion of the A132 region in the presence of putative toxin - antitoxin genes *sma2230/31*, *sma2253* and *sma2275/sma2273*.

	Conjugational frequency LB - IPTG	Conjugational frequency LB + IPTG	Complementation
A132 pTH1931	$1.7x10^{-3}$	$1.8x\ 10^{-4}$	n.a
A132 *sma2230/31*	$2.7x10^{-3}$	$2.2x10^{-3}$	+
A132 *sma2253*	$1.2x10^{-3}$	$2.5x10^{-4}$	-
A132 *sma2275/73*	$4.6x10^{-3}$	$2.1x10^{-4}$	-
B158 pTH1931	$1.1x10^{-2}$	$2.1x10^{-4}$	n.a
A128 pTH1931	$1.4x10^{-2}$	$1.1x10^{-2}$	n.a

A132 is defined by FRT sites located 1,231,998 - 1,284,751 nt on pSymA. The region includes genes *sma2230/31*, *sma2253* and *sma2275/sma2273*. A128 (region on pSymA flanked by FRT sites that previously was successfully deleted) and B158 (region flanked by FRT sites harbors essential gene for $tRNA^{arg}$). Transconjugants were selected on LB media with (+) and without (-) IPTG. Successful complementation was indicated with +, lack of complementation was indicated with -n.a. - not applicable.

Table 5.12. Transduction of ΔA132 into *S. meliloti* P110 (WT, wild type) with plasmid pTH1931 containing *sma2230/31,* or *sma2253* or *sma2275/73.* These genes are predicted members of toxin/antitoxin families (Makarova *et al.*, 2009).

	Transductant per plate (transductional frquency per recipient)	Complementation
ΦΔA132 → WT (pTH1931 – *sma2230/31*)	30 ($1.01x10^{-7}$)	+
ΦΔA132 → WT (pTH1931 – *sma2253*)	0 ($< 10^{-9}$)	-
ΦΔA132 → WT (pTH1931 – *sma2275/73*)	0 ($< 10^{-9}$)	-
ΦΔA132 → WT (pTH1931 –)	0 ($< 10^{-9}$)	n.a.
ΦFL4094 → WT (pTH1931 – *sma2230/31*)	60 ($2.3x10^{-7}$)	n.a.
ΦFL4094 → WT (pTH1931 – *sma2253*)	45 ($3.1x10^{-7}$)	n.a.
ΦFL4094 → WT (pTH1931 – *sma2275/73*)	34 ($3.01x10^{-7}$)	n.a.
ΦFL4094 → WT (pTH1931 –)	44 ($7.6x10^{-8}$)	n.a.

Strain FL4094, containing one of the FRT borders used to make ΔA132 was a control donor. +, Complementation; -, No Complementation , n.a., not applicable. Data represents mean values of four plates.

Figure 5.12. Alignments of the Sma2230 and Sma2231 homologues found by BLAST searches.

A) Alignments of the nine predicted Sma2230 homologues. The homologues have the following Entrez accession numbers (percentages of identity): annotation is missing, *Sinorhizobium meliloti* AK83 - YP_004557944.1 (100%), *Sinorhizobium fredii* NGR234 - YP_002822682.1 (96%), *Rhizobium etli* 8C-3 - ZP_03510656.1 (94%), *Rhizobium leguminosarum bv. trifolii* WSM2304 - YP_002279073.1 (94%), *Agrobacterium vitis* S4 - YP_002548505.1 (94%), *Mycobacterium tuberculosis* H37Rv - NP_215629.1 (58%), Rothia dentocariosa M567 - ZP_07071973.1 (54%), Cupriavidus metallidurans CH34 - YP_581968.1 (50%).

B) Alignments of the seven predicted Sma2231 homologues that have the following Entrez accession numbers (percentages of identity): *Rhizobium leguminosarum bv. trifolii* WSM2304 - YP_002279072.1 (88%), *Sinorhizobium fredii* NGR234 - YP_002822621.1 (87%), *Rhizobium etli* CIAT 652 - YP_001985720.1 (87%), *Sinorhizobium meliloti* 1021 - NP_436449.1 (100%), *Agrobacterium vitis* S4 - YP_002548504.1 (75%), *Pseudomonas syringae pv. phaseolicola* 1448A - YP_272157.1 (42%), *Leptospira biflexa serovar Patoc* - YP_001840262.1 (32%).

A)

```
S.meliloti1021                  MRSTINLDDTLMERAKSLTGTKETAALVRQALETLVRVESGKRLIALGGT 50
S.melilotiAK83                  MRSTINLDDTLMERAKSLTGTKETAALVRQALETLVRVESGKRLIALGGT 50
S.frediiNGR234                  MRSTINLDDTLLERARSLTGTKETAALVRQALETLVRVESGKRLIALGGT 50
R.etli8C-3                      MRSTINLDDTLMERARFLTGTKETATLVRQALETLVRVESGKRLIALGGT 50
R.leguminosarum.bv.trifoliiWSM  MRSTINLDDTLMERARSLTGTKETATLVRQALETLVRVESGKRLVALGGT 50
A.vitisS4                       MRSTINLDDNLMERAKSLTGTKETAALVRQALETLVRVESGKRLIALGGS 50
M.tuberculosisKZN4207           ----MTVDDALLAKAAELTGVKEKSTLLREGLQTLVRVESARRLAALGGT 46
R.dentocariosaM567              MRTTVTLDEEIVAKAAKLTGIEEHSSLVRAALETLVRVENGRLLATLGGS 50
C.metalliduransCH34             MRTTITLDDELLAKARAYTGLEEKTALVREALKALIQREAAKRLANLGGS 50
                                    :.:*: :: :*   ** :* ::*:* .*::*:: * .: *  ***:

S.meliloti1021                  MPDAEAAPRRRSAAAK------------ 66
S.melilotiAK83                  MPDAEAAPRRRSAAAK------------ 66
S.frediiNGR234                  MPDAEAAPRRRSAAAK------------ 66
R.etli8C-3                      MPDAEAAPRRRSAARK------------ 66
R.leguminosarum.bv.trifoliiWSM  MPDAEAAPRRRSAAAK------------ 66
A.vitisS4                       MSDAEAAPRRRSEVTK------------ 66
M.tuberculosisKZN4207           DPQATAAPRRRTSPR------------- 61
R.dentocariosaM567              EPQAQAAPRHQDT--------------- 63
C.metalliduransCH34             QPGIEGAPRRRSDDDENRIGEPPDGGSD 78
                                 .   .***::
```

B)

```
R.leguminosarum.bv.trifoliiWSM      --MILADTSIWIDHFRHAD---SELRRIIEDDRLLCHPAVIGELALGSLR 45
S.frediiNGR234                      --MILADTSIWIDHFRHAD---SELRRIIEDDRLFCHPAVIGELALGSLR 45
R.etliCIAT652                       --MILADTSIWIDHFRHAD---SELRRIIEDDRLLCHPAVIGELALGRLR 45
S.meliloti1021                      --MILADTSIWIDHFRHTD---AELRRIIEDDRLLCHPAVIGELALGSLR 45
A.vitisS4                           --MILVDTSIWIDHFRHGD---VELCKIIEDDRLLCHPFVIGELALGSLR 45
P.syringae1448A                     MKGVLVDTSVWVEHFRNNS---PELVNLLSQDRVLIHPMVIGELACGTPP 47
L.biflexa                           --MILVDTSVWIEFFRGKEPYFSKLVGLVESSDVIAHEVVFGELLQG--C 46
                                       :*.***:*::.**  .    :*  ::... :: *  *:***  *

R.leguminosarum.bv.trifoliiWSM      DRGSVMAFLAAQREAFVAAHDEVMMMIDRHAIFS-----MGIGYTDAHLM 90
S.frediiNGR234                      DRSNVIAFLAAQREAFVATHDEVMMMIDRHAIFS-----MGIGYTDAHLM 90
R.etliCIAT652                       DRSNVIAFLAAQRQALVATHDEVMMMIDRHAIFS-----MGIGYTDAHLI 90
S.meliloti1021                      ERSSVIAFLMAQREALVATHQEVMMMIDRHAIFS-----MGIGYTDAHLL 90
A.vitisS4                           EREAVIGFLAAQREAAIATHAEVMTVIDRYSIFS-----MGIGYTDAHLL 90
P.syringae1448A                     DRSNTLTDLGDLRGAQQPTVSEVIAFLNTHKLYG-----LGCGLVDMTLL 92
L.biflexa                           KNKSEIAFVLDYWESLNNIFSNGMFVKAGKLSFENKHLEIGIGIIDSILI 96
                                    ..   :  :    :     : : .     :     :* *  *  *:

R.leguminosarum.bv.trifoliiWSM      ASVLLDQRAALWTRDKRLRAAAEKAGASLHTPDNSRN 127
S.frediiNGR234                      ASVLLDQRAALWTRDKRLRAAAERVGAPLHTPANARN 127
R.etliCIAT652                       ASVLLDRRAALWTRDKRLRAAAEKAGASLHTPDNARS 127
S.meliloti1021                      ASVLLDQRMALWTRDKRLQAAAEKAGASLHTPAHTRN 127
A.vitisS4                           TSTLLDRRVSLWTRDKRLAAAAQKVGATVHASANSPH 127
P.syringae1448A                     ASALLSG-TALWTLDKRLERLASRMAVSYQPPTH--- 125
L.biflexa                           S-ETKQRNLKLWTLDKKILKVLQAQHIYEI------- 125
                                    :    .    *** **::    .
```

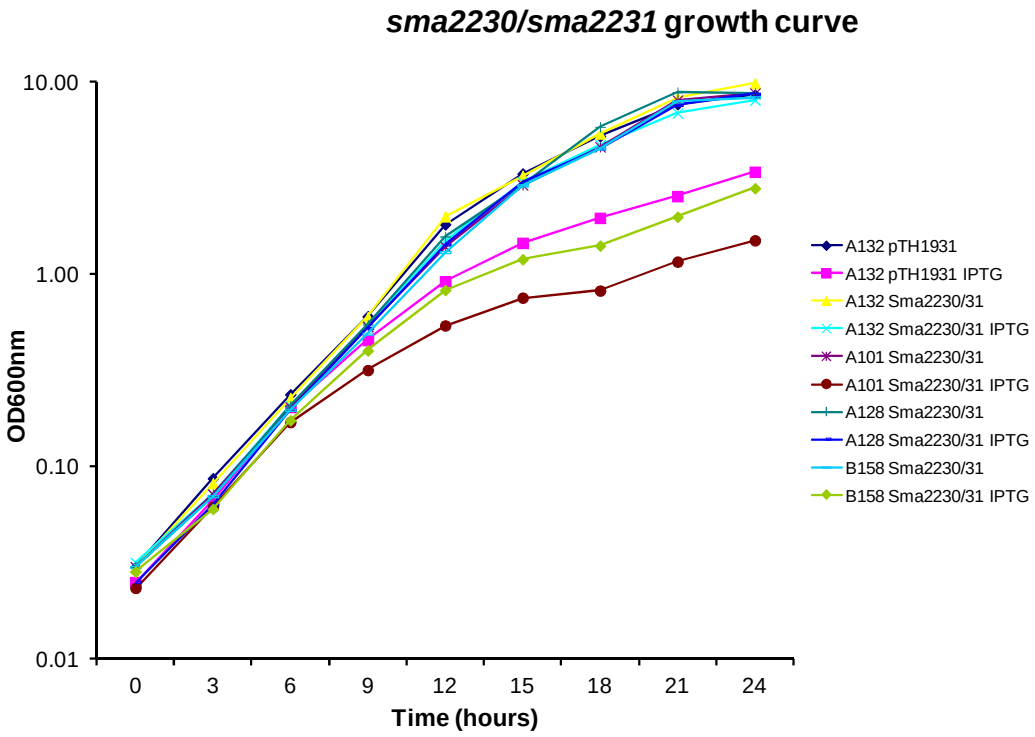
Figure 5.13. The *sma2230/sma2231* genes of *S. meliloti* constitute a TA locus

The *sma2230/sma2231* genes were cloned in pTH1931 and conjugated into strain A132 in which FRT sites flank a 40 kb region that includes *sma2230/sma2231*. Deletion of the 40 kb region was achieved upon induction of the *flp* gene by 0.5 mM IPTG.

A) Growth of A132 (pTH1931),), A132 (pTH1931- *sma2230/sma2231*), and three control strains in LB medium with or without IPTG. In strain A101 the pSymA *oriV* is flanked by FRT sites. In strain A128 FRT sites flank a non-essential pSymA region. Strain B158 carries FRT sites flanking an essential $tRNA^{Arg}$ gene on pSymB.

B) Viable cell counts (CFU/ml) in all cultures, with or without addition of IPTG.

A)


sma2230/sma2231 growth curve
OD600nm
10.00
1.00
0.10
0.01
0
3
6
9
12
15
18
21
24
Time (hours)
A132 pTH1931
A132 pTH1931 IPTG
A132 Sma2230/31
A132 Sma2230/31 IPTG
A101 Sma2230/31
A101 Sma2230/31 IPTG
A128 Sma2230/31
A128 Sma2230/31 IPTG
B158 Sma2230/31
B158 Sma2230/31 IPTG

B)

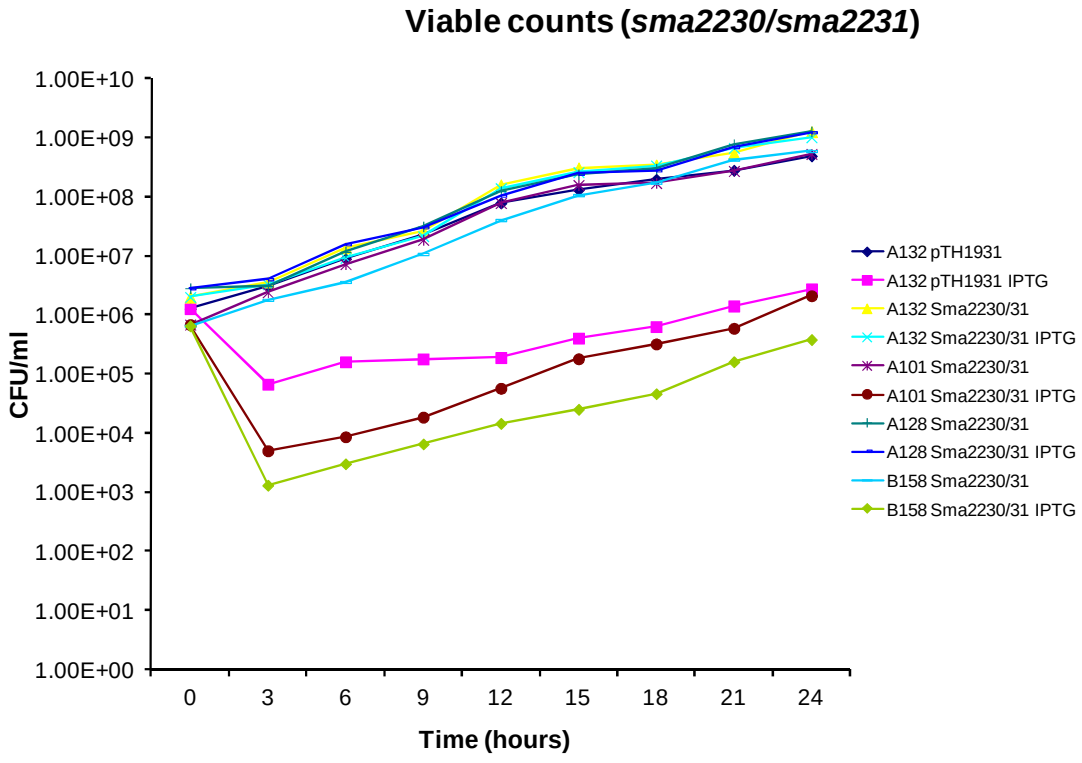

Viable counts (*sma2230/sma2231*)
CFU/ml
1.00E+10
1.00E+09
1.00E+08
1.00E+07
1.00E+06
1.00E+05
1.00E+04
1.00E+03
1.00E+02
1.00E+01
1.00E+00
0
3
6
9
12
15
18
21
24
Time (hours)
A132 pTH1931
A132 pTH1931 IPTG
A132 Sma2230/31
A132 Sma2230/31 IPTG
A101 Sma2230/31
A101 Sma2230/31 IPTG
A128 Sma2230/31
A128 Sma2230/31 IPTG
B158 Sma2230/31
B158 Sma2230/31 IPTG

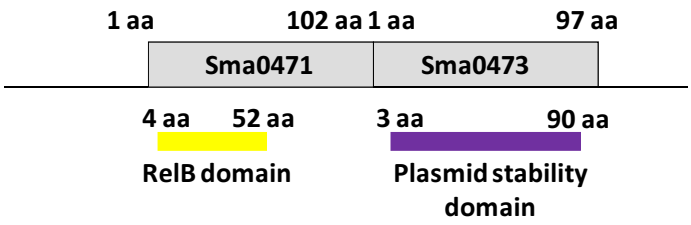
Table 5.13. Symbiotic phenotype of pSymA deletions

Strain	Shoot dry weight (avg. mg/plant)(SD)	% WT	Symbiotic phenotype
P110	66.9±(20.9)	100%	Fix$^+$
uninoculated	4.0±(0.0)	6%	Fix$^-$
RmG994 (Fix$^-$)	6.4±(13.1)	10%	Fix$^-$
ΔA102	43.5±(19.4)	65%	Fix$^+$
ΔA103	37.4±(20.8)	56%	Fix$^+$
ΔA104	54.6±(19.2)	82%	Fix$^+$
ΔA105	84.0±(18.6)	126%	Fix$^+$
ΔA106	79.0±(20.5)	118%	Fix$^+$
ΔA107	36.1±(16.5)	54%	Fix$^+$
ΔA108	95.9±(5.3)	143%	Fix$^+$
ΔA109	89.1±(6.4)	133%	Fix$^+$
ΔA111	74.3±(13.4)	111%	Fix$^+$
ΔA112	68.2±(13.2)	102%	Fix$^+$
ΔA113	50.7±(10.4)	76%	Fix$^+$
ΔA114	4.8±(0.1)	7%	Fix$^-$
ΔA115	4.9±(0.3)	7%	Fix$^-$
ΔA116	4.2±(0.7)	6%	Fix$^-$
ΔA117	15.8±(4.9)	24%	Fix$^-$
ΔA118	7.8±(4.8)	12%	Fix$^-$Nod$^-$
ΔA120	83.3±(10.0)	124%	Fix$^+$
ΔA121	4.4±(0.2)	7%	Fix$^-$
ΔA122	86.6±(17.9)	129%	Fix$^+$
ΔA124	68.8±(14.7)	103%	Fix$^+$
ΔA125	63.7±(15.6)	95%	Fix$^+$
ΔA126	48.4±(17.2)	72%	Fix$^+$
ΔA128	39.7±(9.5)	59%	Fix$^+$
ΔA129	48.4±(8.9)	72%	Fix$^+$
ΔA130	116.2±(11.8)	174%	Fix$^+$
ΔA132	113.6±(15.7)	170%	Fix$^+$
ΔA133	78.1±(1.0)	117%	Fix$^+$

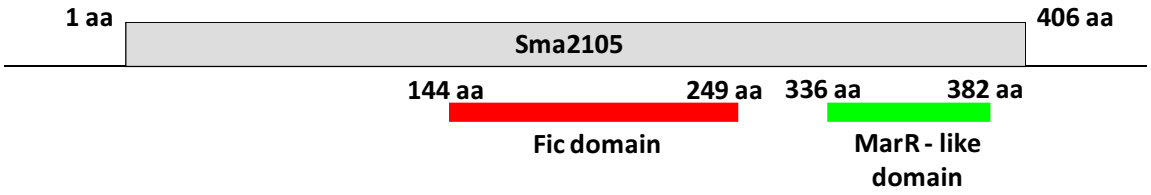
The plant tops were harvested 28 days post inoculation, dried, and shoot dry weight were determined. The values above represent the mean of duplicate samples (each sample consisting of 7-8 shoots) ± standard deviation. The symbiotic phenotype was determined both by examination of the leaves for green (indicating Fix$^+$) or yellow (indicating Fix$^-$) color, and from the shoot dry weight data.

Figure 5.14. Domains organization within predicted toxins and antitoxins like systems *sma0471*/*sma0473* (A), *sma2105* (B) and *sma2230*/*sma2231* (C).

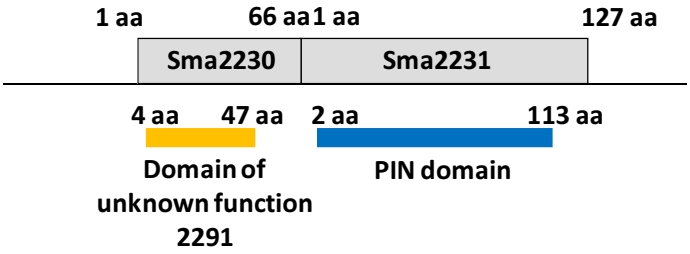
A)


1 aa
102 aa
1 aa
97 aa
Sma0471
Sma0473
4 aa
52 aa
3 aa
90 aa
RelB domain
Plasmid stability domain

B)


1 aa
406 aa
Sma2105
144 aa
249 aa
336 aa
382 aa
Fic domain
MarR - like domain

C)


1 aa
66 aa
1 aa
127 aa
Sma2230
Sma2231
4 aa
47 aa
2 aa
113 aa
Domain of unknown function 2291
PIN domain

CONCLUSION

This thesis describes an effective strategy for deleting large regions of the *Sinorhizobium meliloti* genome using Flp/FRT site-specific recombination. These deletions allowed us to identify phenotypes associated with deleted regions of the *S. meliloti* pSymA and pSymB megaplasmids. The pSymB megaplasmid contains two genes, $tRNA^{arg}$ and *engA*, that are essential for viability, as well as many other loci controlling the uptake and utilization of diverse nutrients. The pSymA deletions revealed three toxin-antitoxin (TA) systems that contribute to pSymA maintenance. This work opens the possibility for further characterization of identified loci and better understanding of biological significance of toxin antitoxin systems which are abundant in *S. meliloti* genome.

Results of deletion analysis can help us to better understand the evolution, plasticity and maintenance of *S. meliloti* megaplasmids. Each *S. meliloti* megaplasmid is somewhat self-interested because each has arrived at a distinct strategy for survival: pSymA utilizes potent TA systems, while pSymB utilizes essential functions transferred from the chromosome.

Using results from deletion analysis a clear strategy for producing a megaplasmid-free *S. meliloti* strain can be made. We will be using the ϕC31 integrase system of *Streptomyces lividans* to move the essential pSymB genes ($tRNA^{arg}$ and *engA*) and three pSymA toxin-antitoxin (TA) systems (*sma0471*/*sma0473*, *sma2105* and *sma2230*/*sma2231*) to the chromosome and then systematically delete the entire megaplasmids by combining already existing deletions using transduction as an

experimental approach. By producing a megaplasmid-free *S. meliloti* strain, the genome of this bacterium will be reduced by approximately 45% which will be significant contribution toward minimization of the *S. meliloti* genome.