STUDY OF SHORT CHAIN DEHYDROGENASE / REDUCTASES (SDRs) IN *SINORHIZOBIUM MELILOTI*

By

Asha Ivy Jacob, M.Sc.

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Biology

Waterloo, Ontario, Canada, 2007

Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Sinorhizobium meliloti maintains a complex lifestyle, including saprotrophy, rhizophere colonization and root hair infection leading to the formation of root nodules in which the plant provides sustenance in return for nitrogen fixation. *S. meliloti* cells use a variety of carbon substrates for growth; this omnivory probably contributes to competitive ability in the soil. Several candidates for contribution to the catabolic capacity are found within the family of short chain dehydrogenases /reductases (SDR), which catalyze NAD(P)(H) dependent oxidation / reduction reactions.

The 6.7 Mb genome of *S. meliloti* contains 78 SDR-encoding genes distributed on all three replicons. In this work each of these genes were disrupted by single crossover mutagenesis. These mutants were screened for growth on 93 different compounds as carbon source, and phenotypes were found for 17 of the mutants, providing suggestions for potential substrates of the corresponding enzymes. Carbon sources for which phenotype was observed include sugar alcohols, leucine, lysine, ornithine, galactitol, rhamnose, arabinose, mono-methyl succinate and ribono-γ-lactone. In addition, one of the mutants was found to be a proline auxotroph. In several cases, the phenotypes were consistent with the phenotypes of deletion mutants in which large sections of pSymB were absent. Eight of the mutants exhibited symbiotic deficiency after inoculation of alfalfa, while viable cells of three of the mutants could not be isolated from the nodules even though nitrogen fixation occurred. The results suggest that the corresponding SDR enzymes are involved in a pathway that is

required for maintenance of viability by cells throughout infection and nodule development.

This work demonstrates that members of the SDR family contribute to both the catabolic capacity and the symbiotic interactions of *S. meliloti*. Further experiments will address the details of the biochemical pathways involved. Knowledge of the substrate specificities of these enzymes should also prove informative in the description and annotation of orthologs that are identified in other genome sequences.

Claims of contribution to knowledge

1. Identified substrate utilization phenotypes for pSymB deletion mutant strains by screening them on 93 different carbon sources.

2. Identified loci for utilization of ornithine, leucine, γ-aminobutyric acid (GABA), uridine, galactitol, trans-4-hydroxy-L-proline, maltitol and palatinose by functional complementation of deletion mutants of pSymB megaplasmid.

3. Identified 78 genes in the *Sinorhizobium meliloti* genome encoding short chain dehydrogenase/reductases using bioinformatics tools. This is the first report of study of the full complement of this protein in any organism.

4. Mutated 77 of the 78 SDR encoding genes facilitating assay of their physiological and biochemical role in *Sinorhizobium meliloti.* One of the SDRs FabG was recalcitrant to mutation.

5. Determined that 17 of the 77 SDR mutants were involved in the catabolism of one or more substrates tested, implying the critical role played by these enzymes in the metabolic capacity of this organism.

6. Demonstrated that mutation in SMc00778, one of the SDRs in this study, leads to proline auxotrophy.

7. Demonstrated that SDRs are important in symbiosis. 8 of the SDR mutant strains were affected in their ability establish an effective symbiotic relationship with the plant host alfalfa. SMb20492 mutant strain could not form nodules on alfalfa and SMc00326 mutant strain formed nodules devoid of bacteroids. The remaining six mutants were able to infect and form nodules containing bacteria but were unable to fix nitrogen. These results showed the relevance of this group of protein in infection and establishment of a favorable symbiosis.

8. Showed that even though SMc01204, SMc02034, SMc02356 and SMb20750 mutants could successfully infect plants and form healthy nodules, no bacteria could be isolated from the nodules.

9. Demonstrated by microscopic analysis that the nodules formed by all the SDR strains contained bacteria other than those formed by SMc00326 mutant strain. SMc00326 mutants formed empty nodules filled with starch.

10. Developed in-gel activity assay for identifying substrates catabolized by SDRs. Cell extracts from SMc02486, SMb20492, SMb20493 and

SMb21159 mutant strains showed no oxidoreductase activity when succinate was used as a substrate. Similarly, cell extract from SMc01500 mutant did not show activity with sorbitol.

11. 9 of the SDRs showing a carbon utilization phenotype were cloned and over expressed for subsequent analysis. The proteins were of the expected molecular weight.

Acknowledgements

I would like to express my gratitude to Dr. Trevor C. Charles for providing me with the opportunity to work in his lab. He has been a great mentor, a patient supervisor and one of the nicest people I have met during my tenure here. I am really proud to have been a part of his lab. I would also like to thank my committee members, Dr. Owen Ward, Dr. Michael Palmer and Dr. Brendan McConkey for their direction, encouragement and insights into this project. I would also like to thank Dr. Ivan Oresnik and Dr. Turlough Finan for providing the ORFeome clones. Thanks to Dr. Turlough Finan for providing the plasmid integration vector pTH1703 used in this study. Thanks to Dr. Sirin Adham and Dr. Scott Clark for their contribution to this project. A very special thanks to Dale Weber for his valuable assistance with the microscopic work in this thesis.

I wish to thank the Charles lab members past and present for making the time I spent at Waterloo memorable. I would especially like to thank Chunxia Wang, Dave Capstick, Louise Bélanger, Adi Rolider and Keith Walsh. I would like to acknowledge the help from Shahnaz Ajani, Elizabeth Marton and Tara Spence–they have helped me with some of the most laborious stuff in this thesis. A big thanks to Tomo Nikitovic- he was a great help during the formatting of my thesis.

A special thanks to the friends I made in this department- they made Canada feel like home. I would especially like to mention Greg Silsbe, James

Lally, Amanda Hamilton, Hazel Lees, Gabriella Meglei and Piya Lahiry. Thank you guys for being the wonderful people you are.

A special acknowledgement of appreciation goes my parents and sister. Without them none of this would have been possible. Their unconditional love, support and faith have been my strength through some very challenging times. Also, I would like to thank Maya Kurup. She has been a great source of support and strength. I couldn't have asked for a better friend.

Finally, I am especially grateful to Vikram, for encouragement, for believing in me and for putting up with me throughout this work. Through thick and thin he has stood by me and given me love and faith to go on. My world is so much better because of you…….

Table of Contents

List of Figures

List of Tables

List Of Abbreviations

ABC	ATP- binding cassette
adh	alcohol dehydrogenase
Am	ampicillin
ATP	adenosine triphosphate
bp	base pair
Bdh	3-hydroxybutyrate dehydrogenase
CFU	colony forming units
Cm	chloramphenicol
CoA	coenzyme A
DHB	D-3-β-hydroxybutyrate
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DTT	dithiothreitol
ED	Entner Duodroff
EDTA	ethylene diamine tetraacetic acid
EMP	Embden-Mayerhof-Parnas
eg.	for example
Gm	gentamycin
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
idh	isocitrate dehydrogenase
kb	kilobase pair
Km	kanamycin
LHB	L-3-β-hydroxybutyrate
mdh	malate dehydrogenase
MTT	(4.5-dimethyl-2-yl)-2,5-diphenyltetrazolium bromide
NAD	nicotinamide adenine dinucleotide
NADH	reduced nicotinamide adenine dinucleotide
NADP	nicotinamide adenine dinucleotide phosphate
NADPH	reduced nicotinamide adenine dinucleotidephosphate
ng	nanogram
Nm	neomycin
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PMS	phenazine methosulfate
PP	pentose phosphate
SCOP	Structural Classification of Proteins
SDR	short chain dehydrogenase/reductase

SDS	sodium dodecyl sulfate
Sm	streptomycin
Tc	tetracycline
TCA cycle	tricarboxylic acid cycle
Tm	primer annealing temperature
TRAP	Tripartite ATP independent periplasmic transporters
Tris	tris (hydroxymethyl)amino methane
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

CHAPTER 1

Literature Review

1.1 Rhizobia

The Rhizobiales, an order within α- proteobacteria, are best known for symbiotic association with the legume host plants wherein they form root nodules in which they fix atmospheric nitrogen. Members of the Rhizobiales include plant symbionts *Sinorhizobium meliloti*, *Bradyrhizobium japonicum*, *Mesorhizobium loti* and *Rhizobium leguminosarum* and the phytopathogen, *Agrobacterium tumefaciens*. Interestingly, this order also includes mammalian pathogens *Brucella* and *Bartonella*.

The root nodule forming bacteria have been further classified based on their host specificity. The Rhizobiales consists of *S. meliloti* which forms nitrogen fixing nodules on plants of the genera *Medicago*, *Melilotus* and *Trigonella*, *M. loti* which forms nodules on plants of the *Lotus* genus and three biovars of *R. leguminosarum*;bv. *viciae*, bv. *trifolli*, bv. *phaseoli* which form nodules on the plants of the genera *Pisum*, *Trifolium* and *Phaseolus*. The genus *Bradyrhizobium* includes *B. japonicum* whose legume host is soybean.

While *S. meliloti*, *M. loti* and *R. leguminosarum* form symbiotic nitrogen fixing nodules on their respective legume host, *A. tumefaciens* induces tumors on plants, and the disease is known as crown gall. Members of the genus *Brucella* are a serious problem in the cattle industry. Infection by this organism

causes fetus abortion in cattle. Moreover these pathogens can be transmitted from cattle to humans.

1.2 Root nodulation and symbiosis

Symbiosis by definition means "interaction between two different organism living in close physical association, typically to the advantage of both" (www.dictionary.reference.com). In the rhizobia-legume symbiosis, the bacteria fix atmospheric nitrogen in soil by reducing it to ammonia as shown below

$$N_2 + 8e^- + 10H^+ \rightarrow 2NH_4^+ + H_2$$

This allows the legume host to survive in nitrogen deficient soil. Also, this mode of symbiotic nitrogen fixation curtails the use of synthetic nitrogen fertilizers reducing contamination of soil and ground water. This could possibly be a part of the solution to aid in the clean up of the environment.

The process of nodulation and symbiosis involves several complex steps, some of which are yet to be experimentally elucidated (Patriarca *et al.* 2002, Patriarca *et al.* 2004). The plant host secretes flavanoids or isoflavanoids specific to a rhizobial strain, leading to the colonization of bacteria on the roots and the rhizoshere. The bacteria produce specific lipochitooligosaccharide molecules, Nod factor in response to the host flavanoids. This induces root hair curling and root cortex meristem division. Invagination of the primary cell wall of the root hair leads to the formation of the infection thread. A polysaccharide matrix and plant cell wall material now surrounds the invading bacteria. It is

through this growing infection thread that the bacteria invade the host and form root nodules.

Inner cortical cells of the root hairs divide leading to the formation of the nodule primordia. When the growing infection threads reach these dividing cells, they release the bacteria into the cytoplasm. Here the bacteria grow in size and differentiate into pleomorphic nitrogen fixing forms called bacteroids. These bacteroids, surrounded by the peribacteroid membrane form the symbiosome.

Mutations in the plant host or in the bacteria can lead to the abolishment of symbiosis i.e. inability to form nodules (Nod^-) or inability to fix nitrogen though the bacteria is able to infect and form nodules (Fix^-) (Beringer *et.al.* 1980). Studies have shown that in the bacteria mutation of genes such as *nod*, *nif* and *fix* that are directly involved in symbiosis, or mutations in genes involved in general cellular processes like metabolism of substrates and those involved in cell surface topology can result in ineffective symbiosis or total absence of this interaction.

1.3 Physiology of symbiotic nitrogen fixation

During symbiosis, within the developing nodule the bacteroids fix atmospheric nitrogen in the form of ammonia. An enzyme called nitrogenase carries out this process. While the bacteria provide a source of fixed nitrogen to the plant, they in turn get carbon in the form of photosynthate from its host. This reciprocity is the basis of this symbiotic association.

Nitrogenase enzyme is oxygen sensitive, however the bacteria being aerobic require oxygen to carry out their physiological functions. The bacteria have evolved several mechanisms to fix nitrogen under aerobic conditions. It is pertinent that oxygen tension within the nodule be conducive for nitrogenase activity as well as bacterial cellular functions. This fine-tuning is achieved by a molecule called leghaemoglobin present within the nodule tissue that is able to bind oxygen and provide oxygen to the bacteria at low levels (Appleby, 1984). In *S. meliloti* the FixLJK and NifA sense O_2 tension and activate the *fix* and *nif* genes (Dixon and Kahn, 2004). RegSR and FnrN have been shown to play a similar role in other rhizobia (Boesten and Priefer, 2004; Lopez *et al.*, 2001). It has also been shown that high O_2 tensions can directly inactivate NifA (Morett *et al*., 1991). The end result of this complex regulation is that the bacteroids are able to express an active nitrogenase complex under low O_2 tension.

A great deal of research has focused on the cycling of nutrients between the plants and the bacteria. It is believed that the chief photosynthate provided by the plant to the developing nodule is sucrose. However studies have shown that sucrose is not taken up as such by the bacteroid. A broadly accepted hypothesis based on physiological studies (Bergerson and Turner, 1967; Finan *et al.* 1981; Glenn *et al.* 1980; McKay *et al.* 1988) and genetic evidence (Arwas *et al.* 1985; Bolton *et al.* 1986; Engelke *et al.* 1987; Finan *et al.* 1983; Ronson *et al.* 1981 and Watson *et al.* 1988) is that the chief carbon source provided to the bacteroid is a C4- dicarboxylate, malate or succinate. Malate is considered to be a strong candidate. *S. meliloti* C4 dicarboxylate transporter mutants form

ineffective nodules on the alfalfa host, strengthening the hypothesis (Finan *et al.* 1988, Yurgel *et al.* 2005).

Malate is transported into the bacteroid by *dctA*. One molecule of malate is oxidized to oxaloacetate and another molecule of malate is oxidatively decarboxylated to pyruvate. Pyruvate then gets further oxidized to acetyl CoA that gets condensed with oxaloacetate to form citrate, allowing the TCA cycle to proceed. Though small amounts of malate stimulate the nitrogen fixation, even moderate levels of these compounds are inhibitory. In free-living cells the overload of C4 dicarboxylates can be channeled into cellular metabolism. However in non-growing bacteroids, this overload creates an imbalance in the carbon and energy inputs and outputs. This would lead to the probable inhibition of the TCA cycle, if the concentration of malate were not favourable for the bacteroid.

If malate is the sole carbon source utilized by the bacteroid, then it is logical to hypothesize that the TCA cycle must be fully operational. However, the ability of aconitase and 2-ketoglutarate dehydrogenase mutants of *B. japonicum* to fix nitrogen showed that a bypass to the conventional TCA cycle must have been operating in this system (Thöny-Meyer *et al.* 1996). This bypass could be one that involves 2-ketoglutarate decarboxylase or it could be a non-cyclic pathway with intermediates leaving at critical branch points.

Whether the bacteria accumulate the fixed ammonia themselves is a question that needs to addressed. It has been shown that glutamate synthase and glutamine synthase, involved in the major pathway for ammonia assimilation,

are repressed in bacteroids. Mutations in these genes do not affect symbiotic nitrogen fixation (Patriarca *et al.* 2002). It has also been reported that glutamate dehydrogenase activity is absent in rhizobia. This led to the belief that it was alanine and not ammonia that is secreted by the bacteroids at least in soybean nodules (Waters *et al.* 1998). Mutation of *aldA,* an alanine dehydrogenase, involved in alanine biosnythesis did not affect nitrogen fixation in *R. leguminosarum* and *M. loti* (Kumar *et al.* 2005 and Allaway *et al.* 2000). As transamination requires the transfer of an amino group from an existing amino acid to a keto acid, the only possibility for alanine to be secreted as a sole product would be by *de novo* synthesis from pyruvate and ammonia in bacteroid and not by transmination (Prell *et al.* 2006). To this day it is believed that the plant cytosol is where assimilation of ammonia to amino acid takes place.

Two amino acid uptake systems have been identified in *R. leguminosarum*- the *aap* and *bra* systems. They are both members of the ABC transporter family. Mutating either of these individually did not affect nitrogen fixation. However, double *aap-bra* mutants were severely nitrogen starved in symbiosis (Walshaw *et al.* 1996, Hosie *et al.* 2001, Hosie *et al.* 2002 and Lodwig *et al.* 2003). Studies showed that nitrogenase enzyme was active in these mutants but nitrogen assimilation in plants was compromised. The lack of cycling of amino acids in bacteroids seems to have a direct effect on the carbon metabolism. The double *aap-bra* mutants showed accumulation of PHB granules indicating that the excess pyruvate generated was converted to PHB rather than alanine.

It is vital to understand the exchange of amino acids between the plant cytosol and the bacteroid. It is assumed that the plant provides the donor amino acid to the bacteroid which then transaminates a keto acid. The resultant ketoacid produced in the bacteroid as a result of this transamination is not secreted but it is metabolized by the bacteroid.

The identification of amino acids exchanged between the two symbionts is challenging as both *aap* and *bra* systems have broad transporter activity. If it is presumed that one amino acid is exchanged between the two, then a good candidate predicted for the amino acid imported would be glutamate. The by product of this reaction would be 2-ketoglutarate and the accumulation of this molecule will inhibit the TCA cycle. Also, transport studies have shown that glutamate is not efficiently transported across the peribacteriod membrane. On the other hand if imported amino acid transaminates pyruvate to alanine or oxaloacetate to aspartate and the resultant keto acid is not converted to 2-ketoglutarate, a split TCA cycle can operate. As alanine/asparate would be secreted there would be no carbon accumulation. This hypothesis could explain the inhibition of the TCA cycle by C4 dicarboxylates and the lack of a fully functional TCA cycle in bacteroids. Moreover, it would explain the role of transaminases in bacteroids and also the importance of functional amino exporters and importers like the Aap and Bra. A thought provoking hypothesis is that if a highly reduced amino acid like 4-aminobutyrate is imported and a highly oxidized amino acid like alanine is exported there is a large reducing

power provided to the bacteroid that enables it to drive nitrogen fixation (Prell *et al.* 2006).

1.4 Rhizobial genetics

The study of root nodule symbiosis has advanced due the development of genetic manipulation techniques for rhizobial strains. Classical genetic techniques used in enteric bacteria have been adapted or modified for use in this system. Most of these techniques have been well applied to *S. meliloti*, making it an ideal system for genetic analysis.

The identification and isolation of spontaneous and chemically induced antibiotic resistant and auxotrophic mutants was the beginning of the genetic studies in *Rhizobium* (Pain, 1979; Scherrer and Dénarié, 1971). This led to the first chromosomal mapping studies based on Hfr-like method using chromosome mobilizing plasmid RP4 and derivatives. RP4 and derivatives have also been used in the cloning of large contiguous segments of the genome by R prime formation (Cheng *et al.* 1988; Nayadu and Rolfe 1987; Julliot *et al.* 1984; Kondorosi *et al.* 1984 and Kiss *et al.* 1980). The studies on the extended symbiotic region of megaplasmid pSymA in *S. meliloti* was based on the above mentioned observation (Renalier *et al.* 1987).

Transposon mutagenesis is another technique used extensively in rhizobia. This technique is advantageous in that each transposant obtained harbours a genetic lesion; is marked by a resistance marker and has an insertion of defined sequence; and as the insertion results in the inactivation of gene, rates of reversion are much lower than chemically induced mutations. The most

extensively used transposon in this system is the 5.8 kb Tn5 transposon (Berg *et al.* 1975). This transposon has central kanamycin-neomycin, streptomycin, and bleomycin resistances, flanked by 1.5 kb inverted IS50 insertion elements (Berg *et al.* 1989).

Tn5 inserts randomly in the genome and are relatively stable (Meade *et al.* 1982). The transposon is usually conjugated into the recipient cell on a plasmid that is unable to replicate in that recipient cell. Replacing the DNA between the IS50s with DNA encoding alternative antibiotic reisitance and/or plasmid transfer origin (*oriT* or *mob*), plasmid origin of replication (*oriV*), β-galactosidase gene (*lacZ*) or levansucrase gene (*sacB*) (Berg *et al*. 1980; De Vos *et al.* 1986; Furuichi *et al.* 1985; Hirsch *et al.* 1985) has increased the versatility of Tn5 and its derivatives. It is also generally possible to replace one Tn5 derivative insertion with another, by homologous recombination between the IS50s that flank them (De Vos *et al.* 1986). Large scale deletions have been generated on the pSymB megaplasmid using this method (Charles *et al.* 1991). Gene fusion vehicles based on Tn5-derivative transposons that generate fusion to *E. coli lacZ* (β-galactosidase), *phoA* (alkaline phosphatase), *gusA* (β-glucuronidase), *luc* (luciferase) and *tac* promoter, aid in *in vivo* transcriptional and translational fusions allowing monitoring and control of gene expression (Bellofatto et al. 1984; Krebs and Reznikoff 1988, Manoil and Beckwith 1985). The method of gene mapping on the *S. meliloti* chromosome by an HFr-like mobilization was used in *S. meliloti* SU47 (Finan *et al*. 1988; Glazebrook and Walker 1989) and *Rhizobium* sp. str. NGR234 chromosome (Østeras *et al.*

1989). Signature-tagged mutagenesis using mini-Tn5 transposons has aided in the functional anaysis of the sequenced *S. meliloti* genome (Pobigaylo *et al.* 2006)

Development of broad host range, mobilizable and integrating cloning vectors has greatly aided in the transfer of cloned DNA between rhizobia and *E. coli* (Ditta *et al.* 1980, Cowie *et al.* 2006). Using this approach a novel reporter gene fusion library of *S. meliloti* has been generated (Cowie *et al.* 2006), the *S. meliloti* ORFeome was developed (Schroeder *et al.* 2005) and various genes involved in symbiosis have been identified (Luo *et al.* 2005, Zhang *et al.* 2006).

Bacteriophages that use *Rhizobium* as host have been described. In *S. meliloti*, transducing phage ϕM12 and N3 (Finan *et al.* 1984, Martin and Long 1984), coupled with transposon mutagenesis has greatly aided in genetic manipulation of the system.

Microarrays have helped faciltate the study of the complex interacation of these symbionts with their plant host (Barnett *et al.* 2004; Mitra *et al.* 2004). This kind of transcriptome profiling has helped reveal the importance of several genes in symbiosis and osmoadaptation (Domínguez- Ferreras *et al.* 2006, Barnett *et al.* 2004)

1.5 Rhizobial genomics

Advances in scientific technology have resulted in a tremendous amount of information being siphoned out into the research community. Nowhere is this more obvious than in the sequencing of the genomes. The complete genome sequences of several hundreds of bacterial species are now available and many

more continue to be sequenced. This strategy has proven to be effective in better understanding the functioning of genes as a whole in microbes. Not only does sequence information aid in understanding gene function, it also helps in understanding how genes evolved. Better understanding of genetic evolution can be obtained by comparing gene components and organization in the genomes among species, an approach known as comparative genomics. Here the genomes of the plant symbionts *R. leguminosarum*, *R. etli*, *M. loti*, *B. japonicum* will be discussed with more emphasis on *S. meliloti* genomics. The genome of the phytopathogen *A. tumefaciens* will also be reviewed.

Comparison of the genome of Rlv3841 with *M. loti*, *S. meliloti* and *A. tumefaciens* revealed that homologues of 2,253 genes occur in all the 3 genomes, 2272 are absent from all three and 2740 occur in one or another but not in all. Of the latter genes, 546 were present in all the three rhizobia but were missing from *A. tumefaciens*. Also 264 of these genes were present in *B. japonicum*. This set of genes includes genes involved in nitrogen fixation but it has been hypothesized that many of these genes might play an unidentified role in symbiosis. 16% of the predicted genes in *A. tumefaciens* are not found in other members of this genus. The predicted products of these genes include proteins involved in cellulose production, plasmid maintenance, cell growth, transcriptional regulation and cell wall synthesis. Several additional proteins are predicted to catabolize plant cell wall material, sugars and exudates. Linear chromosomes such as the one found in *A. tumefaciens* are only found in two

other prokaryotes, members of *Borrelia* and *Streptomyces* (Young *et al.* 2006, Kaneko *et al.* 2000, Kaneko *et al.* 2002, Wood *et al.* 2001, Galibert *et al.* 2001).

The genomes of the all these organisms have a high GC content. They usually consist of a chromosome and one or more plasmids. Most of the housekeeping genes are maintained on the chromosome whereas genes responsible for functions such as substrate transport and utilization and symbiosis/ pathogenesis are plasmid encoded (Young *et al.* 2006, Kaneko *et al.* 2000, Kaneko *et al.* 2002, Wood *et al.* 2001, Galibert *et al.* 2001). The exception to this is *B. japonicum* that has a very large chromosome but no plasmids.

ABC transporters are widely distributed across these genomes. These transporters aid in sequestering nutrients from the soil providing a competitive advantage to these microorganisms. 183 complete ATP- binding cassette (ABC) operons, corresponding to 11% of the total protein complement have been identified in *R. leguminosarum*, most of which are localized on the plasmids (Young *et. al.* 2006). In the *A. tumefaciens* genome, transporters constitute about 15% of this genome of which 87% are located on the chromosome. There are also about 11 LysE/RhtB aminoacid efflux proteins, twice the number seen in bacteria outside the *Rhizobiaceae*. There are also a large number of ATP independent periplasmic (TRAP) transporters. Like its counterparts *S. meliloti* and *M. loti*, *A. tumefaciens* has an abundance of ABC transporters, constituting 60% of their total transporter complement. Predicted substrates for these ABC transport systems include sugars, amino acids and peptides. A total of 40 genes

encoding the ABC transporter system were identified on *M.loti* plasmid pMLa whereas only 5 of these were found on pMLb. The *R. etli* symbiotic plasmid contains several complete and partial ABC transporters for sugars as well as type III (plays a role in nodulation efficiency) and type IV (plays a role in the delivery of protein or DNA from the prokaryote to the eukaryotic cell) secretion systems (Young *et al.* 2006, Kaneko *et al.* 2000, Kaneko *et al.* 2002, Wood *et al.* 2001, Galibert *et al.* 2001).

Genes for essential metabolic pathways, TCA cycle, glyoxylate cycle and Entner-Doudoroff pathway have been identified in the symbionts and phytopathogen. As they lack phoshofructokinase, the Emden-Meyeroff pathway is absent in these systems (Lodwig *et al*. 2003 and Fuhrer *et al.* 2005). Additionally, genes for the Calvin Benson Bassham cycle, PEP carboxylase gene carbonic anhydrase gene are present *Bradyrhizobium japonicum* and *S. meliloti* (Young *et al.* 2006, Kaneko *et al.* 2000, Kaneko *et al.* 2002, Wood *et al.* 2001, Galibert *et al.* 2001)

A striking finding is the similarity of the circular chromosomes of *A. tumefaciens* and *S. meliloti*, strengthening the belief that these organisms arose from a common ancestor. It has been hypothesized that the *S. meliloti* chromosome was present in a progenitor that later acquired the symbiotic megaplasmids. The mosaic characteristics of the *A. tumefaciens* linear chromosome and plasmids comprising mainly of orthologs found on each of the *S. meliloti* replicons suggests that these organisms diverged after the acquisition of the symbiotic megaplasmids by this progenitor (Galibert *et al.* 2001).

1.5.1 *Sinorhizobium meliloti* Rm1021

S .meliloti is best known for its symbiotic association with its legume host alfalfa wherein it fixes atmospheric dinitrogen in the form of ammonia in specialized structures called root nodules. Apart from this a lot of interest in studying this organism arises due to its close relationship with plant pathogen *A. tumefaciens* and animal pathogen *Brucella*.

The *S. meliloti* genome consists of three replicons – a single chromosome (3.65 Mb) and two megaplamsmids pSymA (1.35 Mb) and pSymB (1.68 Mb). The GC content of the genome is 62.1%. 6204 protein-coding regions have been identified in the genome of which 59.7% have database matches and 8.2% are unique to this organism. The proportion of orphan genes was significantly higher on the megaplasmids than on the chromosome with 11.5% on pSymA and 12.3% on pSymB. Insertion sequences and phage sequences constitute 2.2% of the genome with an overall abundance in pSymA especially near the symbiotic genes, a feature similar to other rhizobial strains (Galibert *et al.* 2001).

pSymA and pSymB bear similarities to plasmids from *Rhizobium* species pNGR234a and *Agrobacterium* Ti and Ri plamids. *repABC* genes have been identified based on sequence similarities and origin of replication has also been inferred. pSymA has putative conjugal transfer genes (*traABCD*) and a putative *oriT* sequence but lacks *traIRMBF* and *trbDJKLFH* found on other rhizobial plasmids. pSymB on the other hand has no transfer genes other than the paralogs of pSymA *traA* and *oriT*. The lower GC content of pSymA and

also the distinct codon usage by this replicon suggest it to be of alien origin. No essential genes could be identified on pSymA consistent with earlier published results wherein curing the *S. meliloti* strain Rm2011 of this plasmid did not affect the viability of the strain. However recent studies have shown single recombinant mutant of SMa2253 (Cowie *et al.* 2006) could not be obtained. On the other hand pSymB carries a number of essential genes, including a tRNA for arginine, *minCDE* required for cell division and two asparagine synthesis genes, one of which should be involved in the growth of this organism on minimal media. Analysis of *S. meliloti* strain Rm1021 carrying defined deletions in pSymB showed that the deletion strains were compromised in their ability to initiate and maintain effective symbiosis. These strains were unable to utilize a wide range of carbon and nitrogen sources (Chapter 2). Taken together, these features suggest that pSymA is plasmid like but pSymB has several chromosomal characteristics (Galibert *et al.* 2001, Barnett *et al.* 2001, Finan *et al.* 2001).

12% of the genome encodes for genes involved in transport, most of which are ABC transporters (200 ABC genes) and the remaining constitute TRAP transporters (15 systems). Most of these transporters, mainly identified to play a role in solute import are on pSymB (17.4%). Rht transporters (hydroxylated amino acid efflux proteins) are particularly abundant in *S. meliloti*. No PTS (phosphoenolpyruvate sugar phosphotransferase) system was identified indicating that the sugars are first transported into the cytoplasm where they are then phosphorylated by sugar kinases encoded by the

chromosome and pSymB (Finn *et al.* 2001, Galibert *et al.* 2001). In a recent study several of these transporters were disrupted by creating reporter gene fusions to identify specific inducers (Mauchline *et al*. 2006). 76 systems constituting ~ 47% of the ABC uptake systems and 53% of the TRAP transporters was shown to be induced by a specific biologically active compound. The abundance of transporters and the wide range of substrates transported by them probably provide a competitive advantage to this organism in the soil by allowing it to scavenge a wide range of nutrients.

8.7% of *S. meliloti* genome is comprised of regulatory elements. The LysR family predominates, particularly on pSymA. Each of the 90 LysR encoding genes was mutated by plasmid insertions to study their role in symbiosis (Luo *et al.* 2005). Two of the *lysR* transcriptional regulators - *lsrA* (SMc00037) and *lsrB* (SMc01225) showed a symbiotic phenotype. GntR class of regulators is found on the megaplasmids whereas in the chromosome AsnC type seems to predominate. 7 members of the σ^{54} dependent transcriptional regulators have been identified. Two quorum sensing systems have been identified (Marketon and Gonzalez 2002). This genome has 36 response regulators and 37 histidine kinases but no serine threonine kinases. *S. meliloti* encodes the most nucleotide kinases of any sequenced bacterial genome. 14 putative RNA polymerase sigma factor genes belonging to the extracytoplasic superfamily has been identified in this genome. *rpoS,* a gene that encodes sigma factor σ^{38} required for the regulation of transcription of several genes expressed

in response to starvation and during the transistion into stationary phase, is absent in this organism (Galibert *et al.* 2001).

S. meliloti exopolysaccharides, lipopolysaccharides, capsular polysaccharides and cyclic β-glucans are synthesized by genes on the chromosome and pSymB. These polysaccharides are essential for successful plant infection. 12% of the genes on pSymB are involved in this biosynthetic process. Genome sequence revealed two new loci for this function on the chromosome and 9 on pSymB (Galibert *et al.* 2001, Finan *et al.* 2001).

Genes for the biosynthesis and transport of Nod factors are located on pSymA. Each megaplasmid carries a copy of *nodPQ*, genes involved in the sulfation of Nod factors. Moreover *nodG* was found to be a paralog of the chromosomal housekeeping gene *fabG*. *nodM* is believed to have arisen from a recent gene duplication of *glmS*, a gene involved in the synthesis of D-glucosamine synthetase. It can be deduced that *S. meliloti nod* genes have two distinct basis of origin, one by horizontal gene transfer of pSymA from an unknown bacterium and secondly by resident gene duplication (Barnett *et al.* 2001).

9 *nif* genes were identified on pSymA (*nif A*, *nifB* , *nif HDKE*, *nifX, nifN* and *nifS*). Probable orthologs of *nifV* and *nifS* were found on the chromosome. No homologues of *nifZ*, *nifW* and *nifQ* were found in *S. meliloti*, though these genes are present in *M. loti* and *Rhizobium* sp. NGR234a. Besides genes for nitrogen fixation, genes for denitrification, nitrate transport and gene encoding glutamate dehydrogenase are located on *pSymA*. *ntrBC*, *glnA*, *glnB*

and *glnT*, *ald* (alanine dehydrogenase), *amtB* (ammonium transporter), regulatory genes *ntrXY*, *glnE*, *glnK* and *glnD*, the GOGAT glutamate synthase genes and glutamine synthesis genes are on the chromosome. pSymB encodes a nitrate reductase (*narB*), two nitrate transporters and a single glutamine synthetase *glnII* (Barnett *et al.* 2001).

Two sets of the duplicated *fixNOQP* cluster on pSymA encode cytochrome c oxidase of the cbb_3 type with high affinity for oxygen. Both pSymA and the chromosome carry a NADH-ubiquinone dehydrogenase cluster that may enhance energy metabolism during symbiosis. pSymA also encodes formate dehydrogenase (Barnett *et al.* 2001) .

The complete and extensively annotated genome sequence of *S. meliloti* allows researchers to better understand the process of biological nitrogen fixation by concentrating on specific metabolic and regulatory pathways. Functional analysis of this genome will provide an insight into this process and also the complex plant - bacterial interactions. Positive functional screens using in vivo expression technology (IVET) aided in the identification of genes involved in early symbiosis. 23 genes were identified based on their response to apigenin, a nod gene inducer. Six of the genes not inducible by apigenin included *lsrA* gene required for symbiosis and *dgkA* involved in cyclic β 1,2-glucan sythesis. In the group of 17 apigenin inducible genes most are previously uncharacterized in *S. meliloti* and none belong to *nod* gene family (Zhang *et al.* 2006). Signature tagged transposon mutagenesis aided in the analysis of survival and competitiveness of several *S. meliloti* mutants in a single

experiment. Results from this study revealed clusters of mutants that had similar growth patterns under different growth conditions and also clones, carrying similar kinds of mutation grouped into the same cluster (Pobigaylo *et al*. 2006). To further understand the complex plant – microbe interaction an ORFeome of *S. meliloti* genome has been constructed. All the 6204 predicted ORFs of *S. meliloti* are now available in a mobilizable plasmid. By using integrase recombination these ORFs can be integrated in vivo or in vitro into other specialized plasmids, further hastening the functional analysis of *S. meliloti* (Schroeder *et al.* 2006).

The Affymetrix GeneChip with the complete genome of *S. meliloti* and ~ 10,000 probes from the plant host *Medicago truncatula* allowed the study of gene expression in both the symbiont and the host plant simultaneously. The microarray data revealed nearly 5000 transcriptome changes in symbiosis and shows defined transcriptional profiles of *S. meliloti* in different environments. Using this chip 46 differentially expressed sequences in plants exposed for 24 hrs to wild type *S. meliloti* or *exoA* mutant (deficient in symbiosis) of *S. meliloti* were identified. These included genes in ribosome biogenesis, transcriptional factors, stress response proteins, defense response proteins, signaling components and several previously uncharacterized genes (Barnett 2004, Mitra *et al.* 2004).

Based on the genome sequence of rhizobia, even though they are phylogenetically related and are symbiotic in most cases, they differ in their genetic content and gene organization. A lot of work needs to be done to

ascertain whether conserved and varying genes relate to adaptations to specific plant rhizospheres, to environmental conditions or conditions yet to be defined.

1.6 Carbon utilization in Rhizobia

The abundance of transporter systems in Rhizobia aids in the scavenging of a wide range of nutrients from their environment, thereby providing them with a competitive advantage in the soil. Carbon is one of the critical elements required for the survival and maintenance of cell physiology in any organism.

Studies done on the utilization of carbohydrate and Kreb cyce intermediates by Rhizobia (Stowers 1985, Graham 1964, Parker and Graham 1964, Gosselin *et al.* 2001) showed that this characteristic could be used as a diagnostic feature for rhizobial taxonomy. The fast growing Rhizobia including *Agrobacterium* could utilize most of the carbon sources tested. However the slow growing ones were more specific in their requirement utilizing only glucose, sodium citrate, xylose, mannitol, arabinose, galactose and fructose. Only six of the 108 strains tested used dextrin.

The central carbon catabolic pathway is the Entner-Doudoroff (ED) pathway and most hexose is metabolized through this pathway (Stowers and Elkan 1983). In some slow growing Rhizobia, the hexose cycle has been shown to simultaneously operate, allowing the conservation of hexoses for biosynthesis from glyceraldehyde-3-phosphate (Gosselin *et al.* 2001). The Embden-Meyerhof-Parnas (EMP) pathway seems to be strain dependent, showing very low levels of activity when present. It has been demonstrated in soybean nodules that nodule cytosol metabolism proceeds via the EMP and pentose

phosphate pathway (PP) (Stowers and Elkan 1983). Key enzymes of the ED and PP pathway are present in *R. leguminosarum* bacteroids whereas *B. japonicum* bacteroids lacked these enzymes but did possess hexose phosphorylating capacities, even though the bacteroids are unable to actively transport and metabolize sugars.

Disaccharides sucrose and lactose are catabolized only by the fast growing Rhizobia. The locus for the catabolism of these disaccharides in *S. meliloti* Rm1021 was mapped to the pSymB megaplasmid (Gage *et al.* 1998). *S. meliloti* grown on lactose and succinate show diauxic growth, suggesting that succinate is preferentially utilized over lactose. Trehalose utilization and transport mutants (*thuA* and *thuB*) of *S. meliloti* were impaired in their ability to colonize alfalfa roots, indicating that this compound may be important in saprophytic colonization of roots. However, these mutants were more competitive for nodule occupancy when compared to the wild type. It has been hypothesized that this increased infectivity and competitiveness is not because of their inability to utilize trehalose *in planta* but due to their ability to mobilize stored trehalose during infection, a scenario observed previously during plant-fungal interaction (Jensen *et al.* 2005).

Respirometric (Stowers and Eaglesham 1984) and enzymatic (Glenn *et al*. 1984) analysis have shown that glucose is catabolized by the ED pathway in most rhizobia, thought there have been reports of the operation of EMP pathway for this carbohydrate. Catabolism of glucose via the EMP pathway has been correlated with the symbiotic efficiency of *B. japonicum* USDA 110 (Mulongoy

and Elkan 1977). Though *R. legumonisarum* bacteroids utilized glucose and fructose at very low rates, at high concentrations (0.5 M) they stimulated bacteroid respiration. Gluconate on the other hand is transported into the cell and phosphorylated to 6-phosphogluconate which is then metabolized to pyruvate and glyceraldehyde-3-phosphate via the ED pathway. An ancillary pathway also operated for the utilization of gluconate whereby it can be dehydrogenated to 2- ketogluconate initiating the ketogluconate pathway (Keele *et al.* 1970). Mannose, another hexose is catabolized by *S. meliloti* via mannose kinase, producing mannose 6- phosphate that in turn is converted to fructose-6-phosphate, which is further metabolized by the ED pathway. Mannose uptake was repressed when the strain was grown on succinate (Arias and Martinez-Drets 1976).

Mannitol, a preferred carbon source for Rhizobia, is metabolized via mannitol dehydrogenase to fructose (Martínez-Drets and Arias 1970). In *S. meliloti* NAD-arabitol dehydrogenase acted on both arabitol and mannitol whereas the NAD- sorbitol dehydrogenase acted specifically on D-sorbitol (Martínez-Drets and Arias 1970). Inositol dehydrogenase (*idhA*) mutants of *S. meliloti* were unable to utilize rhizopine, a symbioisis-specific compound found in some effective alfalfa nodules. The ability to synthesize and catabolize rhizopine gives rhizobial strains a competitive advantage in the rhizosphere (Galbraith *et al.* 1998). In *R. trifolii* five distinct NAD-dependent polyol dehydrogenases are present – inositol dehydrogenase for inositol; ribitol dehydrogenase for ribitol; arabitol dehydrogenase for D-arabitol, D-mannitol,

and D-sorbitol; xylitol dehydrogenase for xylitol; and D-sorbitol dehydrogenase and dulcitol dehydrogease for dulcitol, ribitol, xyliol and D- sorbitol (Primrose and Ronson 1980). Except for inositol and xylitol, all the polyols tested induced more than one polyol dehydrogenase and polyol transport system. Mutants of *R. leguminosarum* defective in inositol catabolism were not able to compete with the wild type strain for nodule occupancy. The wild type strain appears to displace the mutants early on during infection and nodulation (Fry *et al.* 2001). An *idhA* mutant of *S. fredii* USDA191 was drastically affected in its ability to reduce nitrogen and bacteroids within the nodules were senesced prematurely. Under competitive conditions the nodules formed were predominatly occupied by the parent wild type strain (Jiang *et al.* 2001). Similarly, mutants of erythritol catabolism genes of *R. leguminosarum* were compromised in their ability to compete with the wild type strain though they were still able to form nitrogen-fixing nodules (Yost *et al.* 2006). *S. meliloti* catabolizes fructose via the ED pathway and its uptake is inhibited in the presence of succinate. It has been shown that initial enzymes of glucose, fructose and mannitol metabolism and the ED and EMP pathway are not present in succinate grown cells (Stowers and Elkan 1985).

Fast growing rhizobia catabolize L-arabinose resulting in an alkaline reaction. They catabolize L-arabinose to α-ketoglutarate semialdehyde, which is then dehydrogenated to α-ketoglutarate. Slow growers on the other hand convert L-arabinose into pyruvate and acetaldehyde providing an acidic end product. The presence of 2-keto-3-deoxy-L-arabanoate aldolase in slow growers

makes it an important marker for enzymatic distinction of fast and slow growing rhizobia (Duncan and Fraenkel 1979, Pedrosa and Zancan 1974). In *S. meliloti* the genes for transport and catabolism of arabinose have been localized on pSymB. Mutations in this region do not affect the symbiotic capabilities of the strain (Poysti *et al.* 2007). A locus for the utilization of another pentose sugar, rhamnose has been identified in *R. leguminosarum* (Oresnik *et al.* 1998). The pathway for the catabolism of rhamnose in this organism is evidenced to be different from previously described methyl-pentose catabolic pathway of enteric bacteria (Richardson *et al.* 2004).

All C-4 dicarboxylates like succinate and malate are metabolized via the TCA cycle in rhizobia (Keele *et al.* 1970, Stowers and Elkan 1983). Fast growers like *S. meliloti* use succinate to support growth whereas slow growers do not prefer it. Succinate has been shown to support nitrogen fixation in free-living *S. meliloti* cells. Mutants deficient in succinate dehydrogenase and α-ketoglutarate dehydrogenase did not utilize succinate, L-glutamate, acetate, pyruvate or arabinose. They showed a delay in nodulation and formed ineffective nodules (Gardiol *et al.* 1982). A similar symbiotic phenotype was obtained with *S. meliloti* mutants of NAD-dependent malic enzyme (*dme*$^-$) (Driscoll and Finan 1993). Gluconeogenic mutants of *S. meliloti* also did not utilize succinate. The nodules formed by these mutants lacked a distinct symbiotic zone and very few cells were infected. The nitrogen fixation abilities of these mutants were also significantly lower than wild type (Finan *et al.* 1991). Mutations in C-4 dicarboxylate transport (*dct*) genes rendered *S. meliloti*

strains unable to form nitrogen-fixing nodules. Microscopy of the nodules formed revealed that the nodules did not contain any bacteria and were filled with starch granules (Finan *et al*. 1988, Yurgel and Kahn 2004, Yurgel and Kahn, 20005).

Glycerol is utilized by both slow and fast growing rhizobia. Glycerol kinase and glycerophosphate dehydrogenase carry out the breakdown of this compound and this finally results in the formation of pyruvate (Arias and Martinez-Drets 1976, Keele *et al.* 1970, Stowers and Elkan 1983, Poysti *et al.* 2007). Pyruvate is then channeled into the TCA cycle. It can be carboxylated to oxaloacetate or reduced to lactate. Propionate, a C-3 compound like glycerol is metabolized to methylmalonyl-CoA to produce succinate, which then enters the TCA cycle (DeHertogh *et al.* 1964). In *S. meliloti* a cluster of genes, *bhbA-D*, on pSymB responsibe for growth on polyhydroxyalkanoate degradation intermediates 3-hydroxybutyrate and acetoacetate as sole carbon source have been identified. One of the proteins methylmalonyl CoA mutase, an adenosylcobalmine-dependent enzyme catalyzing the interconversion of succinyl-CoA to methylmalonyl-CoA has been purified (Charles and Aneja 1999, Miyamoto *et al.* 2003).

Isocitrate dehydrogenase (Idh), a TCA cycle enzyme, involved in the conversion of isocitrate to α-ketoglutarate was mutated in *B. japonicum* to study its role in symbiotic nitrogen fixation. Results showed that this mutant was a glutamate auxotroph and even though nodulation was slightly delayed, they were able to fix nitrogen as efficiently as wild type (Shah and Emerich 2006).

Citrate synthase mutants of *S. meliloti* are symbiotically compromised and have altered cell surface exopolysaccharides (Mortimer *et al.* 1999)

C-2 compound, acetate is metabolized via the TCA cycle in *S. meliloti* (Duncan and Fraenkel 1979) and *B. japonicum* (Keele *et al*. 1969). Acetaldehyde is metabolized by an NAD- dependent aldehyde dehydrogenase in *B. japonicum* (Peterson and LaRue 1982). This enzyme can also catalyze the breakdown of other substrates like propionaldehyde, butyraldehyde, benzaldehyde and succinic semialdehyde (Peterson and LaRue 1982). These C-2 compounds have shown to increase oxygen uptake in *B. japonicum* bacteroids (Peterson and LaRue 1982). There has been no report of the involvement of the glyoxylate cycle in the metabolism of acetate. However in cells grown on oleate, isocitrate lyase and malate synthethase activity have been reported.

Nutritional diversity of *Rhizobiaceae* was studied using auxanography (Parke and Ornston 1984). An auxanograhic nutritional survey with representative strains revealed that *B. japonicum* and cowpea *Rhizobium* sp. could utilize a relatively large proportion of aromatic and hydroaromatic compounds tested: *R. leguminosarum, R. trifolii* and *Agrobacterium* species showed an intermediate phenotype whereas *S. meliloti* was the most fastidious utilizing only protocatechuate and quinate (Parke and Ornston 1984). Quinate and anthranilate was not toxic any of the strains tested. The β-ketoadipate pathway, a key pathway for the catabolism of protocatechuate has been delineated in *S. meliloti*. The genes involved in this pathway are on the *pSym*B megaplasmid in two transcriptional units designated *pcaDCHGB* and *pcaIJF*

(MacLean *et al.* 2006). Other than protocatechuate, compounds like benzoate, catechol, *p*-hydroxybenzoate, shikimate and quinate are also metabolized via this pathway.

A multienzyme complex similar to xenobiotic degrading systems in other bacteria carries out stachydrine catabolism in *S. meliloti*. This complex includes a putative monooxygenase and a putative NADPH-FMN-reductase. The genes involved in stachydrine catabolism are also required for carnitine metabolism; thus they could be important in the metabolism of a variety of root exudates. The ability of agrobacteria and rhizobia to respond to aromatic compounds may have significance in plant-microbe interaction (Burnet *et al*. 2000).

Poly-3-hydroxybutyrate (PHB) and glycogen are major carbon storage compounds in rhizobia. These carbon storage compounds are believed to provide the bacteria with energy during the process of infection and nodulation. PHB can constitute up to 50% of the dry weight of *B. japonicum* bacteroids (Wong and Evans 1971). Metabolism of both these compounds is well characterized (Wong and Evans 1971). PHB is depolymerized to D-3-β-hydroxybutyrate (DHB) that is dehydrogenated to acetoacetate, which is ultimately oxidized via the TCA cycle (Wong and Evans 1971). The dehydrogenation of DHB and metabolism of acetoacetate provides sufficient reducing power and ATP for nitrogen fixation (Klucas and Evans 1968). Externally supplied DHB can serve as a sole carbon source for *S. meliloti*. However this organism cannot utilize the L-isomer of this substrate unless the

acsA2 (acetoacetyl CoA synthetase) gene is overexpessed (Aneja *et al.* 2002). The overexpression of this gene is accompanied by novel LHB synthetase activity (Aneja *et al.* 2002).. Mutants unable to synthesize PHB showed more delayed nodulation on *M. truncatula* than in *M. sativa*. They seemed to have an effect on N_2 fixation in *M. truncatula* and younger *M. sativa* nodules. Mutation in the glycogen synthase genes resulted in lower N_2 fixation on *M. truncatula* and older *M. sativa* nodules (Wang *et al*. 2007). The glyoxylate cycle in bacteroids is connected with PHB metabolism. The C-4 intermediates of the glyoxylate cycle are important in the biosynthesis of amino acids and could supply carbon skeleton for ammonia assimilation (Wong and Evans 1971). A significant increase in isocitrate lyase activity is observed when PHB utilization is initiated. However the significance or the presence of the glyoxylate cycle in some rhizobia is still unclear. Isocitrate lyase activity could not be detected in *S. meliloti* or *R. leguminosarum* bacteroids and *B. japonicum* bacteroids are reported to have a partial glyoxylate cycle (Johnson et al. 1966, Stovall and Cole 1978). Karr *et al.* found an increase in nitrogen fixation activity correlated with an increase in PHB content, β-ketothiolase, fumarase and β-hydroxybutyrate dehydrogenase activities and a decrease in isocitrate dehydrogenase and acetoacetate-succinyl-CoA transferase. This can generate C-4 dicarboxylates to sustain the physiologically demands of the nitrogen fixing bacteroids. This said the glyoxylate cycle is not operational in all rhizobial bacteroids. No isocitrate lyase activity could be detected in *S. meliloti* and *R.*

leguminosarum bacteroids. *B. japonicum* seems to have only partial functional glyoxylate cycle.

In conclusion, progress has been made in the identification of specific metabolic pathways in the infected host cells and *Rhizobium* bacteroids. Infected host cells contain functional EMP and PP pathways as well as enzymes for anaerobic metabolism. *Rhizobium* bacteroids have the abilty to use C-4 dicarboxylates as primary carbon source and are unable to use disaccharides or hexoses. Recent application of genetics to the study of carbon and energy metabolism of *Rhizobium* coupled with biochemical approaches should allow a complete undertanding of the mechanism of energy production for symbiotic nitrogen fixation.

1.7 Short chain dehydrogenases/ reductases (SDRs)

Dehydrogenases/ reductases are divided into 3 distinct groups. The long chain dehydrogenases (LDR) or iron-dependent dehydrogenases with a chain length varying from 380-390 amino acids up to over 900 amino acids as in multifunctional fermentative alcohol dehydrogenase from *Escherichia coli*. The iron moiety seems to have more a structural than a catalytic role. The medium chain alcohol dehydrogenases (MDR) have a chain length varying from 260-380 amino acids and usually have a zinc moeity. They are divided into six classes with a 60% sequence identity - Class I: the typical one, ethanol active horse liver alcohol dehydrogenase, Class III: glutathione dependent formaldehyde dehydrogenase and Class IV: found in the mammalian stomach, having highest affinity towards ethanol. Class II, V and VI are not well studied.

The last group is the short chain dehydrogenases (SDR), the candidate in this research (Danielsson *et al.* 1994). The concept that SDR form an enzyme family was confirmed by the work of Persson *et al.* in 1991. The alignment of 20 enzymes firmly established the existence of a new family of dehydrogenases distantly related to the MDR.

Members of the SDR family are encoded by most genomes. They are one of the largest superfamilies of proteins catalyzing NAD/NADH dependent oxidation-reduction reactions (Jörnvall *et al.* 1995, Kallberg *et al.* 2002). Approximately 3000 different members are known to belong to this family, including species variants, with a broad range of substrate specificity ranging from alcohols, steroids, sugars and aromatic compounds. These enzymes are found in archaea, bacteria and eukarya, suggesting a commom ancestral protein, from which all other members evolved (Filling *et al.* 2002).

SDRs exhibit a great functional diversity and even with only about 15-30% amino acid sequence identity between paralogs, sequence specific motifs are detectable, reflecting a common folding pattern. Most SDRs are formed by subunits of about 250- 350 amino acid residues, frequently containing an N- or C-terminal transmembrane domain or signal peptides or form parts of multienzyme complex. Unlike the MDR where the cofactor binds to the C-terminal end, in SDRs the N-terminal region binds NAD(H) or NADP(H) and the C-terminal binds the substrate. The first reported enzyme was Drosophila alcohol dehydrogenase (ADH), having the same function as horse liver ADH but showing no sequence homology and lacking the Zn^{+} ion, which suggested

that the enzyme may belong to an unrelated family of ADH. 54 crystal structures (SCOP database) are available of which 27 are of bacterial origin (Table 1-1). These structures reveal an almost identical α/β folding pattern and a "Rossman fold" (typical of enzymes binding nucleotide coenzymes) consisting of a central β sheet flanked by α helices, a highly conserved active site consisting of a catalytically important tetrad Ser- Tyr- Lys- Asn residues and a N-terminal T-G-X-X-X-G-X-G motif which forms the part of the nucleotide binding region (Filling *et al.* 2002, Kallberg *et al.* 2002).

By comparing different structures of SDRs in the apo- or holoform or complexed with inhibitors/ substrates, the catalytic reaction mechanism seems to be alike for all of the structures, or at least it shares a common pathway. This fact is supported by extensive mutational studies on 3α, 20β- hydroxysteroid dehydrogenase. The Tyr-152 residue functions as the catalytic base whereas Ser-138 stabiizes the substrate and Lys-155 form hydrogen bonds with nicotinamide ribose moiety and lowers the pK_a of the Tyr-OH to promote proton transfer. Large hydrogen bonded solvent network including water molecule bound by Asn-111 and Lys-155 indicated a proton relay with access to bulk solvent molecules (Jörnvall *et al.* 1995, Kallberg *et al.* 2002, Filling *et al.* 2002).

Recent studies on the conserved motifs have led to the classification of these proteins into five subfamilies. Two of these – the "classical" and the "extended" families known previously have been further divided based on their coenzyme specificities. The classical SDRs have been divided into seven and

extended into three subfamilies. NADP is the preferred coenzyme for "classical" SDRs whereas NAD is the preferred one for the "extended" family. The three other families are "intermediate", "divergent" and "complex" which include short chain alcohol dehydrogenases, enoyl reductases and multifunctional enzymes respectively (Kallberg *et al.* 2002, Persson *et al.* 2003). This assignment has been applied to the genomes of human, mouse, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana* and *Saccharomyces cerevisiae*. In the animal genome extended SDRs constitute 25% or less of the total SDRs, while in *A. thaliana* and *S. cerevisiae* the extended make up about 40% of the SDR forms. The genome of *S. cerevisiae* is smaller than any of the other organisms mentioned, with only 19 SDRs in total and the 7 extended SDRs may reflect the minimum critical SDRs required. In *A. thaliana* genome the extended members are close to half of the total SDRs, perhaps reflecting the different metabolic requirement of plants involving several carbohydrate utilization. Based on coenzyme requirement the number of NAD- and NADP-dependent SDRs are similar in mouse, human and plant. However, the proportion of NAD-dependent enzymes is much lower in fruit fly, worm and yeast. Similar categorization is yet to be reported for archaeal and bacterial species (Filling *et al.* 2002, Jörnvall *et al.* 1995, Kallberg *et al.* 2002, Mulichak *et al.* 2002).

SDRs play an important role in intermediary metabolism. Specific dehydrogenases are often required for degradation and subsequent growth on specific substrates like carbohydrates, organic acids, amino acid and lipids. One

of the most widely studied SDR is 3α, 20β- hydroxysteroid dehydrogenase (Filling *et al.* 2002, Mobus *et al*. 2002). This enzyme in mammals acts as a molecular switch in steroid hormone receptor activation. Though their role in bacteria is unknown, the bacterial homologues show a high degree of similarity to their mammalian counterparts. One of the subunits of the proton-translocating NADH:ubiquinone oxidoreductase located in the inner mitochondrial membrane shows homology to SDRs. D-3-hydroxybutyrate dehydrogenase from *S. meliloti* (Aneja and Charles 2005) and *Pseudomonas fragii* degrades polyhydroxybutyrate granules, a biopolymer which is a reserve energy source in bacteria (Ito *et al*. 2006). This enzyme has also been characterized in humans (Gallego *et al.* 2006, Guo *et al.* 2006). Substrate screening revealed sole NAD^+-dependent conversion of (*R*)-hydroxybutyrate to acetoacetate with K_m values of about 10 mM, consistent with plasma levels of circulating ketone bodies in situations of starvation or ketoacidosis. GDP-fucose synthetase, another member of the SDR family, has been shown to aid in stem development in plants and is associated with immune leukocyte adhesion deficiency type II in humans (Somers *et al.* 1998). Another SDR, tropinone reductase found in plants, constitutes a branching point in the biosynthetic pathway of tropane alkaloids that include medicinally important compounds as hyoscyamine and cocaine (Nakajima *et al.* 1998). SMa2019, an SDR in *S. meliloti* has been reported to be involved in quorum sensing (González and Marketon 2003). Halohydrin dehalogenase from *Agrobacterium radiobacter*, *Arthrobacter* sp. and *Mycobacterium* sp. aid in bioremediation of soil by

converting halogenated aliphatics to epoxides (Vlieg *et al.* 2001). Another protein WWOX isolated from human chromosome 16, containing a region of high homology with SDRs is over-expressed in breast cancer cell lines (Bednarek *et al.* 2000). These enzymes are also involved in neurotransmitter metabolism and serotinin synthesis (Auerbach *et al.* 1997).

1.8 This Work

S. meliloti has two distinct lifestyles – one as a free-living soil microorganism and the other as a symbiont with its plant host alfalfa. This organism is also able to utilize a wide variety of substrates for growth. This project addresses the role of the large number of SDRs in the genome of *S. meliloti* and the role of these proteins in free-living and symbiotic capacity of this organism. Little is known about how these genes function or their preferred substrates. Knowledge about how these genes function would give us an insight into the catabolic diversity of this organism.

This thesis explains the results of experiments undertaken to biochemically and phenotypically characterize the SDRs in *S. meliloti*. The data strongly suggests that these proteins are involved in the catabolism of a wide variety of substrates and affect nitrogen fixation.

Furthermore, pSymB deletion mutant strains have been screened on various carbon sources to help identify loci on the megaplasmid required for carbon catabolism. Whereever applicable the carbon utilization phenotype of these deletion strains was consistent with that observed for SDR mutants in that region of the megaplasmid.

Table 1-1. SDR protein crystal structures on SCOP (Andreeva *et al.* 2004)

1 Uridine diphosphogalactose-4-epimerase (UDP-galactose 4-epimerase)
Escherichia coli
Human (*Homo sapiens*)
Trypanosoma brucei

2 dTDP-glucose 4,6-dehydratase (*RmlB*)
Escherichia coli
Salmonella enterica
Streptococcus suis, serotype 2
Streptomyces venezuelae

3 dTDP-6-deoxy-L-lyxo-4-hexulose reductase (*RmlD*)
Salmonella enterica serovar typhimurium

4 GDP-4-keto-6-deoxy-d-mannose epimerase/reductase (GDP-fucose synthetase)
Escherichia coli

5 GDP-mannose 4,6-dehydratase
Escherichia coli
Pseudomonas aeruginosa
Human (*Homo sapiens*)
Thale-cress (*Arabidopsis thaliana*)

6 ADP-L-glycero-D-mannoheptose 6-epimerase
Escherichia coli

7 CDP-glucose-4,6-dehydratase
Yersinia pseudotuberculosis

8 CDP-tyvelose-2-epimerase
Salmonella typhi

9 Phenylcoumaran benzylic ether reductase
Loblolly pine (*Pinus taeda*)

10 Pinoresinol-lariciresinol reductase
Giant arborvitae (*Thuja plicata*)

11 Sulfolipid biosynthesis protein SQD1
Thale cress (*Arabidopsis thaliana*)

12 Negative transcriptional regulator *NmrA has an UDP-galactose 4-epimerase-like structure but evolved a different function*
Aspergillus nidulans

13 Carbonyl reductase
Mouse (*Mus musculus*)
Human (*Homo sapiens*) *dicarbonyl/L-xylulose reductase*

14 Sepiapterin reductase
Mouse (*Mus musculus*)

15 Dihydropteridin reductase (pteridine reductase)
Rat (*Rattus norvegicus*)
Human (*Homo sapiens*)
Leishmania major
Leishmania tarentolae
Trypanosoma cruzi
Nematode (*Caenorhabditis elegans*)

16 Human estrogenic 17beta-hydroxysteroid dehydrogenase
Human (*Homo sapiens*)

17 7-alpha-hydroxysteroid dehydrogenase
Escherichia coli

18 3-alpha, 20-beta-hydroxysteroid dehydrogenase
Streptomyces hydrogenans

19 3-alpha-hydroxysteroid dehydrogenase
Comamonas testosteroni

20 3beta/17beta hydroxysteroid dehydrogenase
Comamonas testosteroni

21 Putative oxidoreductase Rv2002
Mycobacterium tuberculosis

22 R-specific alcohol dehydrogenase
Lactobacillus brevis

23 Cis-biphenyl-2, 3-dihydrodiol-2, 3-dehydrogenase
Pseudomonas sp., lb400

24 Drosophila alcohol dehydrogenase
Fruit fly (*Drosophila melanogaster*)
Fruit fly (*Drosophila lebanonensis*)

25 Glucose dehydrogenase
Bacillus megaterium

26 Sorbitol dehydrogenase
Rhodobacter sphaeroides

27 meso-2,3-butanediol dehydrogenase
Klebsiella pneumoniae

28 Levodione reductase
Corynebacterium aquaticum

29 Mannitol dehydrogenase
Mushroom (*Agaricus bisporus*)

30 (3R)-hydroxyacyl-CoA dehydrogenase domain of estradiol 17 beta-
Dehydrogenase 4 *peroxisomal multifunctional enzyme type 2 segment swapping involving a C-terminal extension to the family fold*
Rat (*Rattus norvegicus*)

31 beta-keto acyl carrier protein reductase
Oil seed rape (*Brassica napus*)
Escherichia coli
Thermotoga maritima
Thermus thermophilus
Streptomyces coelicolor

32 Enoyl-ACP reductase
Oil seed rape (*Brassica napus*)
Malaria parasite (*Plasmodium falciparum*)
Mycobacterium tuberculosis, TB*, gene InhA*
Escherichia coli
Helicobacter pylori
scop new_sp *Thermus thermophilus*

33 Tropinone reductase
Jimsonweed (*Datura stramonium*), I
Jimsonweed (*Datura stramonium*), II
Thale cress (*Arabidopsis thaliana*)

34 1,3,8-trihydroxynaphtalene reductase (THNR, naphtol reductase)
Rice blast fungus (*Magnaporthe grisea*)

35 1,3,6,8-tetrahydroxynaphthalene reductase
Rice blast fungus (*Magnaporthe grisea*)

36 Biliverdin IX beta reductase *Histidine in the active site instead of the Tyr-Lys dyad*
Human (*Homo sapiens*)

37 Type II 3-hydroxyacyl-CoA dehydrogenase
Rat (*Rattus norvegicus*)
Human (*Homo sapiens*)
Thermus thermophilus

38 Carbonyl reductase/20beta-hydroxysteroid dehydrogenase
Pig (*Sus scrofa*)

39 Halohydrin dehalogenase *HheC haloalcohol dehalogenase: evolved a new activity; lost the NAD-binding site*
Agrobacterium tumefaciens

40 Glucose dehydrogenase
Nematode (*Caenorhabditis elegans*)

41 UDP-N-acetylglucosamine 4-epimerase *WbpP*
Pseudomonas aeruginosa

42 DTDP-4-dehydrorhamnose reductase *RfbD*
Clostridium acetobutylicum

43 Gluconate 5-dehydrogenase
Thermotoga maritima

44 Carbonyl reductase sniffer
Fruit fly (*Drosophila melanogaster*)

45 Hypothetical protein F25D1.5
Caenorhabditis elegans

46 Putative dehydrogenase ARPG836 (MGC4172)
Human (*Homo sapiens*)

47 Aldehyde reductase II
Sporobolomyces salmonicolor

48 Hypothetical protein R05D8.7
Caenorhabditis elegans

49 Hypothetical protein At5g02240 (T7H20_290)
Thale cress (*Arabidopsis thaliana*)

50 Polymyxin resistance protein ArnA (PrmI) *evolved new activities: UDP-glucuronic acid decarboxylase, UDP-4-amino-4-deoxy-L-arabinose formyltransferase*
Escherichia coli

51 2,4-dienoyl-CoA reductase, mitochondrial (DECR)
Human (*Homo sapiens*),

52 11-beta-hydroxysteroid dehydrogenase
Human (*Homo sapiens*)
Guinea pig (*Cavia porcellus*)

53 17-beta-hydroxysteroid dehydrogenase type XI
Human (*Homo sapiens*)

54 D-3-β- hydroxybutyrate dehydrogenase
Pseudomonas fragii

CHAPTER 2

Materials and methods

2.1 Bacterial culture and Microbiological techniques

2.1.1 Bacterial strains, plasmids and phage:

The bacterial strains, plasmids and phage used in this study are listed in Table 2-1. The wild type *S. meliloti* strain for these studies was SmP110 (Sm^r). Plasmids containing the first approx. 600 bp of the SDR gene coding sequence in pGEM T- easy (Promega) are annotated based on their orientation relative to the T7 promoter (TEF^+) or SP6 promoter (TEF^-) of the vector (pSMa/b/cXXXXTEF$^+$ or pSMa/b/cXXXXTEF$^-$). Full length SDR ORFs in pGEM T-easy are described as TEO^+ if they are in T7 orientation and $TEO^–$ if they are in the SP6 orientation (pSMa/b/cXXXTEO$^+$ or pSMa/b/cXXXTEO$^-$). Constructs in which the amino-proximal encoding fragments are cloned in pTH1703 (Fig.2.1.) are indicated by the gene name suffixed by fg or fl (pSMa/b/cXXXXfg or pSMa/b/cXXXXfl). Cointegrate strains in SmP110 are described using the respective gene name followed by fg or fl based on their orientation in pTH1703. If the gene fragment is oriented with gus/rfp, then the mutant is suffixed with fg and if it is in frame with lac/gfp, the mutant is suffixed with fl (SMa/b/cXXXXfg or SMa/b/cXXXXfl). Transductants are denoted as SMa/b/cXXXXfg-110 or SMa/b/cXXXXfl-110. SDR mutant strains complemented with an intact copy of the gene are denoted as SMa/b/cXXXX-37. SDR fragments in pVIK112 are indicated as pSMa/b/cXXXX-112 and

cointegrates as SMa/b/cXXXX -112. SDR ORFs cloned into the pET30 series of expression vectors are described as SMa/b/cXXXX-BL.

Table 2-1: List of strains used in this study

Strains or plasmid	Relevant characteristics	Reference or source
S. meliloti strains		
Rm1021	SU47 *str*-21	Meade *et al.* 1982
SmP110	Rm1021 with changed wild-type *pstC*	Yuan *et al.* 2005
RmG373	ΔΩ5177-5079::Tn*5-oriT*	Charles and Finan 1991
RmF726	ΔΩ5149-5079::Tn*5*-233	Charles and Finan 1991
RmF728	ΔΩ5177-*lac-56*::Tn*5*	Charles and Finan 1991
RmG462	ΔΩ5011-5033::Tn*5-oriT*	Charles and Finan 1991
RmG470	ΔΩ5025-5007::Tn*5*	Charles and Finan 1991
RmG471	ΔΩ5033-5025::Tn*5*	Charles and Finan 1991
RmG472	ΔΩ5026-5007::Tn*5*	Charles and Finan 1991
RmF114	ΔΩ5033-5064::Tn*5*-233	Charles and Finan 1991
RmF909	ΔΩ5085-5047::Tn*V*	Charles and Finan 1991
RmF680	ΔΩ5085-5061::Tn*5*-233	Charles and Finan 1991
RmF693	ΔΩ5085-5142::Tn*5*-233	Charles and Finan 1991
RmF514	ΔΩ5061-5047::Tn*5*-11	Charles and Finan 1991
E. coli strains		
DH5α	*endA1 hsdR17 supE44 thi-1 recA1 gyrA96 relA1* Δ*(argF-lacZYA)*U169 Φ80d*lacZ* ΔM15	Hanrahen 1983
DH5α(pRK600)	Mobilizing strain; Cm^r	Wang *et al*. 2006
HB101	*supE44 hsdS20 recA13 ara-14 proA2 lacY1 galK2 rpsL20 xyl-5 mtl-1*	Laboratory collection
BL21 (DE3)/pLysS	*F−ompT hsdSB (rB− mB−) gal dcm (DE3)/ pLysS* (Cm^r)	Novagen
Rm1021 cosmid library	Rm1021 genomic DNA library cloned in pLAFR1 carried in HB101	Friedman *et al.* 1982
Phage		
ΦM12	*S. meliloti* transducing phage	Finan *et al*. 1984
Plasmids		
pLAFR1	IncP cosmid cloning vector, Tc^r	Friedman *et al.* 1982
pRK600	pRK2013 *npt*::Tn*9*; Cm^r Nm-Km^s	Finan *et al.* 1986
pGEM-Teasy	cloning vector, Amp^r	Promega
pET30 series	bacterial expression vector, Km^r	Novagen
pTH1703	integration vector, Gm^r	Cowie *et al.* 2006
pVIK112	suicide vector, Km^r	Kalogeraki and Winans 1997
pAJ44	pLAFR1 cosmid cloneI complementing RmG470 for galactitol utilization	This study
pAJ45	pLAFR1 cosmid cloneII complementing RmG470 for galactitol utilization	This study

pAJ46	pLAFR1 cosmid cloneIII complementing RmG470 for galactitol utilization	This study
pAJ47	pLAFR1 cosmid cloneIV complementing RmG470 for galactitol utilization	This study
pAJ48	pLAFR1 cosmid cloneI complementing RmG373 for ornithine utilization	This study
pAJ49	pLAFR1 cosmid cloneII complementing RmG373 for ornithine utilization	This study
pAJ50	pLAFR1 cosmid cloneIII complementing RmG373 for ornithine utilization	This study
pAJ51	pLAFR1 cosmid cloneIV complementing RmG373 for ornithine utilization	This study
pAJ52	pLAFR1 cosmid cloneI complementing RmF726 for ornithine utilization	This study
pAJ53	pLAFR1 cosmid cloneII complementing RmF726 for ornithine utilization	This study
pAJ54	pLAFR1 cosmid cloneIII complementing RmF726 for ornithine utilization	This study
pAJ55	pLAFR1 cosmid cloneI complementing RmF728 for ornithine utilization	This study
pAJ56	pLAFR1 cosmid cloneII complementing RmF728 for ornithine utilization	This study
pAJ57	pLAFR1 cosmid cloneIII complementing RmF728 for ornithine utilization	This study
pAJ58	pLAFR1 cosmid cloneIV complementing RmF728 for ornithine utilization	This study
pAJ59	pLAFR1 cosmid cloneV complementing RmF728 for ornithine utilization	This study
pAJ70	pLAFR1 cosmid cloneI complementing RmG506 for ornithine utilization	This study
pAJ71	pLAFR1 cosmid clone II complementing RmG506 for ornithine utilization	This study
pAJ72	pLAFR1 cosmid cloneIII complementing RmG506 for ornithine utilization	This study
pAJ73	pLAFR1 cosmid cloneIV complementing RmG506 for ornithine utilization	This study
pAJ74	pLAFR1 cosmid cloneI complementing RmG373 for GABA utilization	This study
pAJ75	pLAFR1 cosmid cloneII complementing RmG373 for GABA utilization	This study
pAJ76	pLAFR1 cosmid cloneIII complementing RmG373 for GABA utilization	This study
pAJ77	pLAFR1 cosmid cloneIV complementing RmG373 for GABA utilization	This study
pAJ78	pLAFR1 cosmid cloneI complementing RmF726 for GABA utilization	This study
pAJ79	pLAFR1 cosmid cloneII complementing	This study

	RmF726 for GABA utilization	
pAJ80	pLAFR1 cosmid cloneIII complementing RmF726 for GABA utilization	This study
pAJ81	pLAFR1 cosmid cloneIV complementing RmF726 for GABA utilization	This study
pAJ82	pLAFR1 cosmid cloneV complementing RmF726 for GABA utilization	This study
pAJ83	pLAFR1 cosmid cloneI complementing RmF728 for GABA utilization	This study
pAJ84	pLAFR1 cosmid cloneII complementing RmF728 for GABA utilization	This study
pAJ85	pLAFR1 cosmid cloneIII complementing RmF728 for GABA utilization	This study
pAJ86	pLAFR1 cosmid cloneIV complementing RmF728 for GABA utilization	This study
pAJ87	pLAFR1 cosmid cloneI complementing RmG373 for uridine utilization	This study
pAJ88	pLAFR1 cosmid cloneII complementing RmG373 for uridine utilization	This study
pAJ89	pLAFR1 cosmid cloneIII complementing RmG373 for uridine utilization	This study
pAJ90	pLAFR1 cosmid cloneIV complementing RmG373 for uridine utilization	This study
pAJ91	pLAFR1 cosmid cloneV complementing RmG373 for uridine utilization	This study
pAJ92	pLAFR1 cosmid cloneVI complementing RmG373 for uridine utilization	This study
pAJ93	pLAFR1 cosmid cloneVII complementing RmG373 for uridine utilization	This study
pAJ94	pLAFR1 cosmid cloneVIII complementing RmG373 for uridine utilization	This study
pAJ95	pLAFR1 cosmid cloneI complementing RmF726 for uridine utilization	This study
pAJ96	pLAFR1 cosmid cloneII complementing RmF726 for uridine utilization	This study
pAJ97	pLAFR1 cosmid cloneIII complementing RmF726 for uridine utilization	This study
pAJ98	pLAFR1 cosmid cloneIV complementing RmF726 for uridine utilization	This study
pAJ99	pLAFR1 cosmid cloneV complementing RmF726 for uridine utilization	This study
pAJ100	pLAFR1 cosmid cloneVI complementing RmF726 for uridine utilization	This study
pAJ101	pLAFR1 cosmid cloneVII complementing RmF726 for uridine utilization	This study
pAJ102	pLAFR1 cosmid cloneVIII complementing RmF726 for uridine utilization	This study
pAJ103	pLAFR1 cosmid cloneI complementing RmF728 for uridine utilization	This study

pAJ104	pLAFR1 cosmid cloneII complementing RmF728 for uridine utilization	This study
pAJ105	pLAFR1 cosmid cloneIII complementing RmF728 for uridine utilization	This study
pAJ106	pLAFR1 cosmid cloneIV complementing RmF728 for uridine utilization	This study
pAJ107	pLAFR1 cosmid cloneV complementing RmF728 for uridine utilization	This study
pAJ108	pLAFR1 cosmid cloneI complementing RmF728 for leucine utilization	This study
pAJ109	pLAFR1 cosmid cloneII complementing RmF728 for leucine utilization	This study
pAJ110	pLAFR1 cosmid cloneIII complementing RmF728 for leucine utilization	This study
pAJ111	pLAFR1 cosmid cloneIV complementing RmF728 for leucine utilization	This study
pAJ112	pLAFR1 cosmid cloneI complementing RmG373 for leucine utilization	This study
pAJ113	pLAFR1 cosmid cloneII complementing RmG373 for leucine utilization	This study
pAJ114	pLAFR1 cosmid cloneIII complementing RmG373 for leucine utilization	This study
pAJ115	pLAFR1 cosmid cloneIV complementing RmG373 for leucine utilization	This study
pAJ116	pLAFR1 cosmid cloneV complementing RmG373 for leucine utilization	This study
pAJ117	pLAFR1 cosmid cloneVI complementing RmG373 for leucine utilization	This study
pAJ118	pLAFR1 cosmid cloneVII complementing RmG373 for leucine utilization	This study
pAJ119	pLAFR1 cosmid cloneI complementing RmF726 for leucine utilization	This study
pAJ120	pLAFR1 cosmid cloneII complementing RmF726 for leucine utilization	This study
pAJ121	pLAFR1 cosmid cloneIII complementing RmF726 for leucine utilization	This study
pAJ133	pLAFR1 cosmid cloneI complementing RmF909 for D-maltitol utilization	This study
pAJ134	pLAFR1 cosmid cloneII complementing RmF909 for D-maltitol utilization	This study
pAJ135	pLAFR1 cosmid cloneIII complementing RmF909 for D-maltitol utilization	This study
pAJ136	pLAFR1 cosmid cloneI complementing RmF680 for D-maltitol utilization	This study
pAJ137	pLAFR1 cosmid cloneII complementing RmF680 for D-maltitol utilization	This study
pAJ138	pLAFR1 cosmid cloneIII complementing RmF680 for D-maltitol utilization	This study

SDR strains used in this study:

ORF	TEF+	TEF-	TEO+	TEO-	pET30	pTH1703 fg
SMa0059	pSMa0059 TEF+		pSMa0059 TEO+			
SMa0074	pSMa0074 TEF+		pSMa0074 TEO+			
SMa0187	pSMa0187 TEF+	pSMa0187 TEF-				pSMa0187fg
SMa0320	pSMa0320 TEF+	pSMa0320 TEF-				pSMa0320fg
SMa0326						
SMa0329	pSMa0329 TEF+	pSMa0329 TEF-	pSMa0329 TEO+			pSMa0329fg
SMa0335	pSMa0335 TEF+	pSMa0335 TEF-				pSMa0335fg
SMa0339	pSMa0339 TEF+		pSMa0339 TEO+			
SMa0389	pSMa0389 TEF+					
SMa0513	pSMa0513 TEF+		pSMa0513 TEO+			
SMa0719	pSMa0719 TEF+					
SMa0854	pSMa0854 TEF+		pSMa0854 TEO+			pSMa0854fg
SMa0959			pSMa0959 TEO+			pSMa0959fg
SMa1367	pSMa1367 TEF+	pSMa1367 TEF-				pSMa1367fg
SMa1398	pSMa1398 TEF+	pSMa1398 TEF-				pSMa1398fg
SMa1452	pSMa1452 TEF+	pSMa1452 TEF-				pSMa1452fg
SMa1629						
SMa1757	pSMa1757 TEF+	pSMa1757 TEF-				pSMa1757fg
SMa2019	pSMa2019 TEF+	pSMa2019 TEF-				pSMa2019fg
SMa2165	pSMa2165 TEF+	pSMa2165 TEF-				pSMa2165fg
SMa2343	pSMa2343 TEF+	pSMa2343 TEF-				pSMa2343fg
SMb20073	pSMb20073 TEF+	pSMb20073 TEF-	pSMb20073 TEO+			pSMb20073fg
SMb20076	pSMb20076 TEF+		pSMb20076 TEO+	pSMb20076 TEO-	pSMb20076-BL	pSMb20076fg
SMb20210	pSMb20210 TEF+					
SMb20214						pSMb20214fg
SMb20409	pSMb20409 TEF+					
SMb20456	pSMb20456 TEF+		pSMb20456 TEO+			
SMb20492	pSMb20492 TEF+			pSMb20492 TEO-	pSMb20492-BL	pSMb20492fg
SMb20493	pSMb20493 TEF+		pSMb20493 TEO+	pSMb20493 TEO-	pSMb20493-BL	pSMb20493fg
SMb20511	pSMb20511 TEF+					pSMb20511fg
SMb20660	pSMb20660 TEF+					pSMb20660fg
SMb20662			pSMb20662 TEO+			pSMb20662fg
SMb20692		pSMb20692 TEF-	pSMb20692 TEO+			pSMb20692fg
SMb20750	pSMb20750 TEF+	pSMb20750 TEF-	pSMb20750 TEO+			pSMb20750fg
SMb20871	pSMb20871 TEF+		pSMb20871 TEO+			pSMb20871fg
SMb21010						pSMb21010fg
SMb21111	pSMb21111 TEF+		pSMb21111 TEO+			pSMb21111fg
SMb21159	pSMb21159 TEF+	pSMb21159 TEF-	pSMb21159 TEO+		pSMb21159-BL	
SMb21348	pSMb21348 TEF+					pSMb21348fg
SMb21383	pSMb21383 TEF+		pSMb21383 TEO+			
SMb21384	pSMb21384 TEF+					pSMb21384fg
SMb21474	pSMb21474 TEF+					

ORF	pTH1703 fl	pVIK112	SmP110 fg	SmP110 fl	SmP110 pVIK112	Transductants	Complemented strains
SMa0059	pSMa0059fl			SMa0059fl			
SMa0074	pSMa0074fl			SMa0074fl			
SMa0187	pSMa0187fl		SMa0187fg	SMa0187fl			
SMa0320	pSMa0320fl		SMa0320fg	SMa0320fl			
SMa0326	pSMa0326fl			SMa0326fl			
SMa0329	pSMa0329fl		SMa0329fg	SMa0329fl			
SMa0335	pSMa0335fl		SMa0335fg	SMa0335fl			
SMa0339	pSMa0339fl			SMa0339fl			
SMa0389		pSMa0389-112			SMa0389-112		
SMa0513	pSMa0513fl			SMa0513fl			
SMa0719		pSMa0719-112			SMa0719-112	SMa0719-110	
SMa0854			SMa0854fg				
SMa0959			SMa0959fg				
SMa1367	pSMa1367fl			SMa1367fl			
SMa1398	pSMa1398fl		SMa1398fg	SMa1398fl			
SMa1452	pSMa1452fl		SMa1452fg	SMa1452fl			
SMa1629	pSMa1629fl			SMa1629fl			
SMa1757	pSMa1757fl		SMa1757fg	SMa1757fl		SMa1757fl-110	
SMa2019	pSMa2019fl		SMa2019fg	SMa2019fl			
SMa2165	pSMa2165fl		SMa2165fg	SMa2165fl			
SMa2343	pSMa2343fl		SMa2343fg	SMa2343fl			
SMb20073	pSMb20073fl		SMb20073fg			SMb20073fg-110	SMb20073-pCO37
SMb20076	pSMb20076fl		SMb20076fg	SMb20076fl		SMb20076fl-110	SMb20076-pCO37
SMb20210		pSMb20210-112			SMb20210-112	SMb20210-110	
SMb20214			SMb20214fg			SMb20214fg-110	
SMb20409		pSMb20409-112			SMb20409-112		
SMb20456	pSMb20456fl			SMb20456fl			
SMb20492	pSMb20492fl		SMb20492fg	SMb20492fl		SMb20492fg-110	
SMb20493	pSMb20493fl			SMb20493fl		SMb20493fl-110	SMb20493-pCO37
SMb20511	pSMb20511fl		SMb20511fg	SMb20511fl			
SMb20660	pSMb20660fl		SMb20660fg	SMb20660fl			
SMb20662			SMb20662fg				
SMb20692	pSMb20692fl			SMb20692fl			
SMb20750	pSMb20750fl		SMb20750fg			SMb20750fg-110	SMb20750-pCO37
SMb20871			SMb20871fg				
SMb21010			SMb21010fg				
SMb21111	pSMb21111fl		SMb21111fg	SMb21111fl			
SMb21159	pSMb21159fl		SMb21159fl			SMb21159fl-110	SMb21159-pCO37
SMb21348	pSMb21348fl		SMb21348fg	SMb21348fl			
SMb21383		pSMb21383-112			SMb21383-112		
SMb21384	pSMb21384fl		SMb21384fg	SMb21384fl			
SMb21474	pSMb21474fl			SMb21474fl			

ORF	TEF+	TEF-	TEO+	TEO-	pET30	pTH1703 fg
SMc00005	pSMc00005 TEF+					pSMc00005fg
SMc00136	pSMc00136 TEF+			pSMc00136 TEO-		
SMc00165	pSMc00165 TEF+					pSMc00165fg
SMc00260	pSMc00260 TEF+					pSMc00260fg
SMc00264	pSMc00264 TEF+					pSMc00264fg
SMc00268	pSMc00268 TEF+			pSMc00268 TEO-		
SMc00326	pSMc00326 TEF+		pSMc00326 TEO+			pSMc00326fg
SMc00372	pSMc00372 TEF+					pSMc00372fg
SMc00553	pSMc00553 TEF+			pSMc00553 TEO-		pSMc00553fg
SMc00572	pSMc00572 TEF+			pSMc00572 TEO-		pSMc00572fg
SMc00603	pSMc00603 TEF+					pSMc00603fg
SMc00733	pSMc00733 TEF+					pSMc00733fg
SMc00778	pSMc00778 TEF+		pSMc00778 TEO+		pSMc00778-BL	pSMc00778fg
SMc00880	pSMc00880 TEF+					pSMc00880fg
SMc01157						pSMc01157fg
SMc01175	pSMc01175 TEF+			pSMc01175 TEO-		
SMc01204	pSMc01204 TEF+		pSMc01204 TEO+		pSMc01204-BL	pSMc01204fg
SMc01500	pSMc01500 TEF+		pSMc01500 TEO+	pSMc01500 TEO-	pSMc01500-BL	
SMc01571	pSMc01571 TEF+			pSMc01571 TEO-		pSMc01571fg
SMc01635	pSMc01635 TEF+					pSMc01635fg
SMc01698	pSMc01698 TEF+		pSMc01698 TEO+			pSMc01698fg
SMc01955						
SMc01991	pSMc01991 TEF+		pSMc01991 TEO+			pSMc01991fg
SMc02034	pSMc02034 TEF+		pSMc02034 TEO+		pSMc02034-BL	
SMc02037	pSMc02037 TEF+					pSMc02037fg
SMc02039	pSMc02039 TEF+					
SMc02040	pSMc02040 TEF+					pSMc02040fg
SMc02041	pSMc02041 TEF+		pSMc02041 TEO+	pSMc02041 TEO-	pSMc02041-BL	pSMc02041fg
SMc02271	pSMc02271 TEF+			pSMc02271 TEO-		pSMc02271fg
SMc02322	pSMc02322 TEF+		pSMc02322 TEO+		pSMc02322-BL	pSMc02322fg
SMc02336	pSMc02336 TEF+					pSMc02336fg
SMc02339	pSMc02339 TEF+					pSMc02339fg
SMc02356	pSMc02356 TEF+		pSMc02356 TEO+		pSMc02356-BL	pSMc02356fg
SMc02486	pSMc02486 TEF+					pSMc02486fg
SMc03878	pSMc03878 TEF+					pSMc03878fg
SMc04391	pSMc04391 TEF+					pSMc04391fg
SMc02522	pSMc02522 TEF+					pSMc02522fg

ORF	pTH1703 fl	pVIK112	SmP110 fg	SmP110 fl	SmP110 pVIK112	Transductants	Complemented strains
SMc00005	pSMc00005fl		SMc00005fg	SMc00005fl		SMc00005fg-110	SMc00005-pCO37
SMc00136	pSMc00136fl			SMc00136fl			
SMc00165	pSMc00165fl		SMc00165fg	SMc00165fl			
SMc00260	pSMc00260fl		SMc00260fg	SMc00260fl			
SMc00264	pSMc00264fl		SMc00264fg	SMc00264fl			
SMc00268	pSMc00268fl			SMc00268fl		SMc00268fl-110	
SMc00326	pSMc00326fl		SMc00326fg			SMc00326fg-110	SMc00326-pCO37
SMc00372	pSMc00372fl		SMc00372fg	SMc00372fl			
SMc00553			SMc00553fg				
SMc00572							
SMc00603	pSMc00603fl		SMc00603fg	SMc00603fl			
SMc00733	pSMc00733fl		SMc00733fg	SMc00733fl			
SMc00778	pSMc00778fl		SMc00778fg	SMc00778fl		SMc00778fl-110	SMc00778-pCO237
SMc00880	pSMc00880fl		SMc00880fg	SMc00880fl			
SMc01157			SMc01157fg				
SMc01175	pSMc01175fl			SMc01175fl			
SMc01204	pSMc01204fl		SMc01204fg	SMc01204fl		SMc01204fg-110	SMc01204-pCO37
SMc01500	pSMc01500fl			SMc01500fl		SMc01500fl-110	
SMc01571	pSMc01571fl		SMc01571fg	SMc01571fl			
SMc01635	pSMc01635fl		SMc01635fg				
SMc01698	pSMc01698fl		SMc01698fg				
SMc01955	pSMc01955fl			SMc01955fl			
SMc01991			SMc01991fg				
SMc02034	pSMc02034fl			SMc02034fl		SMc02034fl-110	SMc02034-pCO37
SMc02037			SMc02037fg				
SMc02039	pSMc02039fl			SMc02039fl			
SMc02040	pSMc02040fl		SMc02040fg	SMc02040fl			
SMc02041	pSMc02041fl		SMc02041fg	SMc02041fl		SMc02041fg-110	SMc02041-pCO37
SMc02271	pSMc02271fl		SMc02271fg	SMc02271fl			
SMc02322	pSMc02322fl		SMc02322fg	SMc02322fl		SMc02322fg-110	SMc02322-pCO37
SMc02336	pSMc02336fl		SMc02336fg	SMc02336fl			
SMc02339	pSMc02339fl			SMc02339fl			
SMc02356	pSMc02356fl		SMc02356fg	SMc02356fl		SMc02356ffg-110	
SMc02486	pSMc02486fl		SMc02486fg	SMc02486fl		SMc02486fg-110	SMc02486-pCO37
SMc03878	pSMc03878fl		SMc03878fg	SMc03878fl			
SMc04391	pSMc04391fl		SMc04391fg	SMc04391fl			
SMc02522	pSMc02522fl		SMc02522fg	SMc02522fl			

Abbreviations are as follows: Amp, ampicillin; Cm, chloramphenicol; Gm, gentamycin; Km, kanamycin; Nm,
Abbreviations are as follows: Amp, ampicillin; Cm, chloramphenicol; Gm, gentamycin; Km, kanamycin; Nm, neomycin; Str, streptomycin; Tc, tetracycline; *oriT*, origin of transfer; *oriV*, origin of replication; pSMa/b/cXXXX TEF+, SDR fragment in pGEM®-T Easy in T7 orientation in *E.coli* DH5α; pSMa/b/cXXXX TEF-, SDR fragment in pGEM®-T Easy in SP6 orientation in *E.coli* DH5α; pSMa/b/cXXXX TEO+, SDR ORF in pGEM®-T Easy in T7 orientation in *E.coli* DH5α; pSMa/b/cXXXX TEO-, SDR ORF in pGEM®-T Easy in SP6 orientation in *E.coli* DH5α; pSMa/b/cXXXX-BL, SDR ORF in pET 30 series of expression vector in *E.coli* BL21 DE31/pLysS; pSMa/b/cXXXXfg, SDR fragment in pTH1703 in *gus/rfp* orientation in *E.coli* DH5α; pSMa/b/cXXXXfl, SDR fragment in pTH1703 in *lac/gfp* orientation in *E.coli* DH5α; pSMa/b/cXXXX-112, SDR fragment in pVIK112; SMa/b/cXXXXfg, co-integrate in SmP110 with *gusA/rfp* fusion; SMa/b/cXXXXfl, co-integrate in SmP110 with *lac/gfp* fusion; SMa/b/cXXXX-112, co-integrate in SmP110 by homologous recombination with SDR fragment in pVIK112, SMa/b/cXXXXfg/fl-110, tranductants obtained by transducing mutation into fresh SmP110 background; SMa/b/cXXXX-pCO37, SDR mutant strain complemented with an intact copy of the gene of interest in pCO37.

All SDR clones are maintained in replicating plasmids (pGEM T-easy and pTH1703) in *E.coli* DH5α and co-integrate are maintained in *S. meliloti* SmP110.

All strains are preserved as permanent frozen stocks at -80°C in tryptone yeast extract (TY) broth with 16% DMSO for *S. meliloti* strains and Luria broth (LB) with 14% DMSO for *E.coli* strains. Streaking a small amount of sample from these frozen stocks onto appropriate medium revived strains.

2.1.2 Media, antibiotic and growth condition

S. meliloti was grown on TY (Beringer, 1947) or Rhizobium minimal media (RMM) at 30°C. *E. coli* was grown on LB medium (Miller, 1972) at 37°C. The RMM media was composed of 2.05 g/l K_2HPO_4, 1.45 g/l KH_2PO_4, 0.15 g/l NaCl, 0.5 g/l NH_4NO_3, 0.623 g/l $MgSO_4.7H_20$, 0.01 g/l $CaCl_2$. 2 H_20, 0.087 g/l K_2SO_4, 0.247 mg/l H_3BO_3, 0.1 mg/l $CuSO_4.5$ H_20, 0.338 mg/l $MnSO_4$. H_20, 0.288 mg/ml $ZnSO_4.7H_20$, 0.056 mg/l $CoSO_4.7$ H_20, 0.048 mg/l Na_2MoO_4. 2 H_20, 1 mg/l thiamine, 1 mg/l calcium pantothenate and 1 mg/l biotin (Broughton *et al.* 1986). Agar (BioShop) was added to the media at a final concentration of 1.5%.

Unless stated otherwise, all the carbon sources were prepared as 10% stock in distilled water and sterilized by filtration through 0.22 μm filter units. They were added at a final concentration of 0.1% to the media.

Antibiotics were obtained from BioShop or Fisher Biochem. All antibiotics were maintained as sterile stock solutions at 4°C. Water-soluble antibiotics were filter sterilized. For *S. meliloti*, antibiotics were routinely used

at the following concentrations (μg/ml): Gentamicin sulfate, 10-20 (except 60 when selecting for pTH1703); neomycin sulfate, 100-200; streptomycin sulfate, 200-400; tetracycline hydrochloride, 10.

For *E.coli* antibiotics were used at the following concentrations (μg/ml): ampicillin, 50-100; chloramphenicol, 10, kanamycin sulfate, 10-50; tetracycline hydrochloride, 2-10.

For blue white screening, when using pGEM T-easy, 5-bromo-4chloro-3indonyl-β-D-galactopyranoside (X-gal) was used at a concentration of 40 μg/ml for *E. coli*.

2.2 Bacterial genetic techniques

2.2.1 Triparental mating/Conjugation

Triparental mating/conjugation was done using liquid cultures wherein mid- log phase culture of donor, recipient and mobilizer were mixed together, washed 3-4 times with sterile saline and re-suspended in 50 μl of sterile LB broth. This was spotted on TY agar plates, allowed to dry and then incubated at 30°C overnight. The mating spots were re-suspended in sterile 0.85% NaCl and plated on appropriate selective medium.

For mating experiments where *E. coli* was the recipient and *S. meliloti* was the donor, *E. coli* was counter selected by incubation at 37°C and transconjugants identified after 24 h of incubation.

2.2.2 Plasmid insertion mutagenesis

Plasmid insertion mutagenesis was performed as previously described (Sambrook *et al.* 1989). pTH1703 was prepared by alkaline lysis method, digested with NotI or SwaI and purified with GFX^{TM} PCR DNA and Gel Band Purification Kit (Amersham Biosciences, UK). DNA fragments (~ 600 bp) including the start codon of each SDR gene were amplified by PCR from *S. meliloti* SmP110 genomic DNA with specific primers for each gene (forward primer and reverse primer$_{600}$). PCR products were cloned into an intermediary vector pGEM T-easy (Promega, USA) as per manufacturers instruction. The insert was released from pGEM T-easy by digesting with NotI or EcoRI, purified with GFX^{TM} PCR DNA and Gel Band Purification Kit (Amersham Biosciences, UK) and ligated to NotI digested and dephosphorylated pTH1703 or EcoRI cut and dephosphorylated pVIK112, resulting in recombinant plasmids that were transferred to *E. coli* DH5α for insertion mutagenesis. Wherever PCR products could not be obtained, clones from the *S. meliloti* ORFeome were digested with appropriate enzymes and fragments cloned as mentioned above for mutagenesis. Each recombinant plasmid construct was mobilized from DH5α into wild type strain *S. meliloti* SmP110 by triparentental mating using *E.coli* DH5α (pRK600) as helper. The three strains were mixed together and spotted on TY plates without antibiotic and incubated (24 h, 30°C), and the mating spots were used to select for plasmid insertion into SDR genes on TY agar medium containing streptomycin / gentamycin (pTH1703) or streptomycin / neomycin (pVIK112). Single colonies formed were re-streaked

to obtain pure colonies on the same selective medium. Phage lysate of the mutant strains was prepared using ΦM12. This lysate was used to transduce the plasmid insertion in a SDR gene into a fresh SmP110 background. To further reconfirm the insertion, genomic DNA was isolated from the mutant strains and used as a template in a PCR reaction performed with one primer that annealed to the vector (*gus* or *gfp*) and another primer that annealed to the SDR gene (upstream of the start codon for amplifying the SDR gene fragment). DNA fragment of predictable size generated by PCR is indicative of the mutation in the correct locus.

2.2.3 Preparation of ΦM12 lysate

Lysates were prepared from log phase (OD_{600}- 0.3 - 0.4) TY grown SmP110 cultures. 0.1 ml of ϕM12 phage stock propagated on SmP110 was added to 5 ml of SmP110 culture (SDR mutants and wild type). This was incubated with aeration at 30°C for 8 h to overnight until lysis could be detected (clearing of culture). A few drops of chloroform were added and mixed carefully to kill any viable bacterial cells that remained. Lysates were stored at 4°C in screw cap vials.

2.2.4 Transduction

Transductions were carried out by mixing equal volume (usually 0.5 ml) of mid to late log phase *S. meliloti* recipient culture (SmP110) in TY with an appropriately diluted donor lysate (obtained from SDR mutant strains in SmP110 background). The mixture was allowed left at room temperature for 20

- 30 min to allow adsorbion of phage particles to he bacterial cell surface. Then 2-5 ml of 0.85% NaCl was added and cells pelleted at 2000 × g for 3 min. The pellet was washed with 0.85% NaCl to remove unadsorbed phage particles and the pellet resuspended in 0.85% NaCl and plated on appropriate selective LB medium. These plates were incubated at 30°C and the transductants obtained were streaked purified thrice on TY or LB containing the appropriate selection.

2.2.5 Screening of SDR mutant strains on carbon sources

Each of the pSymB deletion strains and SDR mutant strains were screened on 93 different carbon sources. Mutant strains were streaked out on RMM agar containing a single carbon source at a concentration of 0.1% and incubated at 30°C. The plates were scored for growth at 3-day intervals for 4 weeks

2.3 Molecular biology techniques

2.3.1 Plasmid isolation (alkaline lysis)

5 ml LB broth with the appropriate antibiotic was inoculated with a single colony and incubated overnight with aeration. 1.5 ml of this grown culture was pelleted by centrifugation at 10000 rpm in a microcentrifuge (IEC 851) for 2 min and the supernatant discarded. The pellet was resuspended in 100 μl of TEG buffer (50 mM Tris-Cl, pH 8.0; 20 mM EDTA; 1% glucose, (pH 8.0) containing 200 μg/ml RNase (20 μl of 10 mg/ml of RNase stock per 1 ml of TEG). 200 μl of ALS (200 mM NaOH, 1% SDS) was added and the suspension mixed gently. To this, 100 μl of 10 M ammonium acetate was added, cooled on ice for 5 min and centrifuged at 12000 rpm (IEC 851) for 10 min. The supernatant was transferred to a clean tube. 900 μl of ethanol was added to the supernatant, mixed gently and incubated at -80°C for 30 min. This was then centrifuged at 12000 rpm (IEC 851) for 15 min and the pellet washed with 0.5 ml of 70% ethanol. The pellet was air dried and re-suspended in 20-50 μl of autoclaved distilled water.

2.3.2 Isolation of genomic DNA

This method was adapted from Charles *et al.* 1999. Cells grown to saturation in 5 ml TY broth were pelleted by centrifugation and the cell pellet was washed in 5 ml 0.85% NaCl, then in 5 ml TES (10 mM Tris-Cl, 25 mM EDTA, 150 mM NaCl, pH 8.0), and resuspended in 5 ml $T_{10}E_{25}$ in a 15 ml polypropylene tube. To this 0.25 ml 25% SDS, 250 μl 10 mg/ml pronase E

(predigested at 37°C for 2 h), 0.65 ml 5 M NaCl was added, incubated at 68°C for 30 min and cooled to room temperature. 2.5 ml 7.5 M ammonium acetate was added to precipitate the proteins, centrifuged at 8000 rpm for 30 min and the supernatant transferred to a fresh 15ml polypropylene tube. To this 0.8 volume of isopropanol was added, incubated at -80°C for 30min and centrifuged at 8000 rpm for 15min. The pellet was air-dried and re-suspended in 500 µl of T_{10} E_1. This was extracted once with equal volume of phenol : chloroform (1 : 1) and once with equal volume of chloroform. The DNA was precipitated by after the addition of 5 M ammonium acetate to a final concentration of 0.3 M, followed by the addition of 0.8 volume of ice-cold isopropanol. The solution was mixed gently, incubated at -80°C for 30 min. and then centrifuged at 12000 rpm (IEC 851) for 20 min. The pellet was washed with 0.5 ml 70% ethanol, air dried and re-suspended in 200-500 µl sterile distilled water.

2.3.3 Cosmid DNA isolation

Cosmids were isolated from 10 ml of Terrific Broth (Sambrook *et. al.* 1989) inoculated with pure cultures from LB containing tetracycline. The cultures were centrifuged at 3000 g for 5 min and the cell pellet resuspended in 1 ml TEG buffer (50 mM Tris-Cl pH 8, 20 mM EDTA, 1% glucose). 2 ml of lysis solution (0.2 M NaOH, 1% SDS), 4.5 ml autoclaved distilled water and 1.5 ml of 7.5 M ammonium acetate were added, and the solution mixed gently. The mixture was incubated at -20°C for 15 min, and then centrifuged at 5000 rpm (IEC 21000R) for 20 min. To the supernatant, 4.5 ml of ice-cold isopropanol

was added, the solution incubated at -20°C for 20 min. and then centrifuged at 5000 g for 20 min. The pellet was air dried and resuspended in 500 µl of autoclaved distilled water. To this was added 10 µl of 5 M NaCl, and 1 ml of ice-cold 95% ethanol. DNA was allowed to precipitate overnight at -20°C, centrifuged at 13000 rpm × 20 min, pellet washed with 500 µl of ice-cold 70% ethanol and centrifuged again. The pellet containing the cosmid was resuspended in 50 µl of autoclaved distilled water and 10-15 µl was digested with EcoRI and separated on 0.8% agarose gel (IX TAE buffer was used to cast and run the gel. 50X TAE g/l- 242 g Tris-base, 57.1 ml glacial acetic acid, 18.6 g EDTA, pH should be ~ 8.5), using λ HindIII digested DNA as size marker. Gels were stained with ethidium bromide and visualized under UV light. Fragment sizes were calculated using the MS-DOS program DNASize (imtech.res.in/raghava/dnasize/, Raghava, 2001).

2.3.4 Preparation and transformation of competant *E.coli* DH5α cells

Competent *E.coli* cells were prepared based on the method by Ausubel *et al.* (1987). All the steps were carried out at 4°C and centrifugations were done in a refrigerated centrifuge (IEC 21000R). 100 ml of early log phase culture (OD_{600} of 0.2 - 0.3) of *E.coli* DH5α in LB broth was pelleted by centrifugation (5 min, 5000 rpm, IEC 21000R). The pellet was carefully re-suspended in 100 mM $MgCl_2$, pelleted by centrifugation (10 min, 5000 rpm, IEC 21000R), re-suspended in 4 ml 100 mM $CaCl_2$ and kept on ice for 1 hr. Then the cells were pelleted by centrifugation (10 min, 8000 × g), re-suspended

in 5 ml 100 mM $CaCl_2$ containing 15% glycerol and left overnight on ice at 4°C. The cells were aliquoted and stored at -80°C for subsequent use.

For transformation, up to 50-100 ng of DNA was added to 50 -100 µl of competent cells. 50 µl of ice-cold 100 mM $CaCl_2$ was added to this mix and left on ice for 45 min. Thereafter it was heat-shocked at 42°C for 90 s and left on ice for 2 min. 0.5 ml of sterile LB broth was added and the mixture incubated at 37°C for 1 h, allowing the expression of plasmid encoded antibiotic resistance. The cells were pelleted (5 min, 8000 × g) and re-suspended in 50 µl sterile LB broth. This was then plated on LB agar containing the appropriate selection.

2.3.5 Induction of SDR genes in pET30 series of expression vectors

The recombinant SDR proteins were expressed in *E.coli* BL21 (DE3) strain by autoinduction (non-IPTG expression). The media composition was obtained from http://www.bioc.aecom.yu.edu/labs/blanlab/VETTING/PROTOCOLS and contained (g/l) 6 g Na_2HPO_4, 3 g KH_2PO_4, 20 g tryptone, 5 g yeast extract, 5 g NaCl (pH 7.2). This was sterilized at 121°C, 100 kPa for 20 min. To this, filter sterilized glucose, lactose and glycerol were added to a final concentration of 0.05%, 0.2% and 0.6% respectively. Strains were inoculated into this media (5 ml) with appropriate antibiotics, incubated for 16-24 h and harvested by centrifugation at 6000 × g, 5 min. The cells were lysed by the addition of lysis buffer (50 mM Tris-Cl pH 6.8, 100 mM DTT, 2% w/v SDS, 0.1% bromophenol blue, 10% v/v glycerol), boiling in a water bath for 2 min and cooling on ice for 5 min. The cell extract was run through an SDS polyacrylamide gel using Tris-

glycine buffer (25 mM Tris base, 250 mM glycine pH 8.3, 0.1% w/v SDS) and subsequently analyzed by Western blotting.

2.3.6 Detection of SDRs by western blotting

Two fibre pads and two filter papers were soaked in transfer buffer (192 mM glycine, 25 mM Tris, 20% methanol; pH 8.0 with 1 N NaOH). PVDF membrane cut to size of the gel was wet in 100% methanol (10 s) and left in transfer buffer till needed. After electrophoresis, the gel was equilibrated in transfer buffer for 5 min and the transfer was carried out for 30 min at 200 mA (transfer buffer was kept cold). The membrane was then placed in 100 ml blocking buffer (1X TEN buffer with 5% skim milk. 10 X TEN – 200 ml 1 M Tris-Cl pH 7.5 - 8.0, 280 ml 5 M NaCl and 20 ml 0.5 M EDTA) overnight at 4°C with gentle shaking. The buffer was discarded and the blot was incubated for 1-2 h at room temperature in 10 ml primary antibody solution (His-tag monoclonal antibody; 1: 2000 of a 0.2 mg/ml stock; Novagen), 5 μl in 10 ml blocking buffer). The blot was washed 3 × 10 min in 1X TEN and incubated for 1-2 h in the dark in 10 ml of secondary antibody (AlexaFluor 488 goat anti-mouse IgG antibody; Invitrogen; 1: 3000 of a 2 mg/ml stock i.e. 3.3 μl of antibody in 10 ml blocking buffer). The blot was washed in 3 × 10 min in 1 X TEN and rinsed twice in distilled water and detected on Typhoon 9400.

2.4 DNA manipulation methods

2.4.1 Restriction digestion

Restriction enzymes were purchased from Fermentas or Life Technologies GIBCO/BRL. Routine digestions were performed in a final volume of 25 μl containing reaction buffer at a concentration of 1 X, DNA (200 - 500 ng) and atleast 5 U of the appropriate enzyme. All digestions were carried out at 37°C for 4 - 16 h except SwaI, which was done at room temperature. Enzymes were inactivated by incubation at the recommended temperature for the specific enzyme.

2.4.2 Ligation reaction

The vector and insert DNA were combined at a ratio of 1: 3. Reaction volume did not exceed 10 μl and contained reaction buffer at 1 X final concentration (buffer contains ATP) and ligase (Fermantas) (1 U). This mixture was incubated at 16°C for 3 h to overnight and transformed into competent *E. coli* DH5α.

2.4.3 Dephosphorylation of vector DNA

DNA fragments digested with appropriate restriction enzymes were dephosphorylated using shrimp alkaline phosphatase (Pharmacia Biotech.). The mixture containing the reaction buffer, the digested DNA and enzyme in a final volume of 10 μl was incubated at 37°C for 20 min followed by inactivation of the enzyme at 68°C for 30 min.

2.4.4 Agarose gel electrophoresis

Standard gels were prepared with 1 X TAE buffer with an agarose concentration of 0.8%. For separation of fragment smaller than 500 bp, samples were electrophoresed through 2% agarose gels. DNA was visualized by adding 5 μl of ethidium bromide (10 mg/ml) to 100 ml molten agar before pouring the gel. For estimation of size of fragments the following standard markers were used: Lambda DNA cut with Hind III (Fermentas), 1 kb ladder (Fermentas) and 100 bp ladder (Roche).

2.4.5 DNA amplification by PCR

The primers used in this study are shown in Table 2-2 and were obtained from Sigma or Mobix. Primers were designed using the web-based program Primer3 (Rozen and Skaletsky 2000) to amplify full length SDR ORFs and 600 bp region including the start codon for each of the SDRs. *gfp* and *gus* primer sequences from pTH1703 were obtained from Cowie *et. al* and primer for pVIK112 were designed in house. Universal T7 and SP6 primers (Sigma, USA) were used to check the orientation of cloned fragments in pGEM T-easy and pET30 series of expression vectors.

PCR reactions mixture contained 1 X reaction buffer, 1.5 mM $MgCl_2$, 50 pmole each of the forward and reverse primers, 200 μM dNTPs, Taq DNA polymerase 1 U (Novagen), 50 ng DNA template in a 25 μl reaction volume. Reaction (30 cycles) was carried out in an Eppendorf thermocycler using the recommended 3- step protocol (hot start at 94°C, 1 min; 30 cycles of denaturation, 92°C for 45 s; primer annealing at 60°C, 1 min; extension at 76°C

for 45 sec; 76°C for 10 min). The primer annealing temperature was varied depending on the Tm of the primers used. The products were visualized by staining agarose gels with ethidium bromide.

2.4.6 Cloning of full length and 600 bp fragments of SDR encoding genes

The full-length SDR encoding genes and their corresponding 600 bp fragments were cloned into the intermediary vector, pGEM T-easy (Promega, USA) engineered to ligate to PCR products. The fragment of interest was then cloned from this intermediary vector after restriction digestion with NotI or EcoRI into NotI cut pTH1703 or EcoRI cut pVIK112. Full length ORF were cloned into NotI cut pET30a or pET30b (Novagen, USA) depending on their orientation in pGEM T-easy (Promega, USA)

2.5 Biochemical techniques

2.5.1 Preparation of cell extract

SDR mutant strains and SmP110, grown to saturation in 5 ml TY were harvested by centrifugation at 5000 rpm for 3 min (IEC 815), washed with ice-cold 20 mM Tris pH 8.0 (3 × 3 min) and resuspended in cold 20 mM Tris pH 8.0 (4 ml/g of wet cell weight). DTT was added to a final concentration of 10 mM. To this suspension of cells, equal volume of glass beads (0.1 mm silica/zirconia beads) were added and cell extracts prepared by using Bead-Beater™ at 4°C (4 × 1 min – between each round of bead beating the cells were kept on ice for 1 min to avoid overheating of the extract). This mixture was

spun down at 13000 rpm × 15 min at 4°C to remove cell debris and glass beads. The supernatant was stored at -80°C and used for activity assays.

2.5.2 SDR activity staining

Non-denaturing polyacrylamide gel electrophoresis was used to detect the activity of SDRs. Cell extract (14 μl cell extract + 6 μl 1 X loading dye containing 50 mM Tris-Cl pH 6.8, 100 mM DTT, 0.1% bromophenol blue and 10 % v/v glycerol) was electrophoresed at 4°C through a 8 % native gel using Tris - glycine buffer (25 mM Tris base, 250 mM glycine, pH 8.3) at 200 V for 4 h. The activity of SDRs was detected as previously described by Selander *et al.* after incubation in staining solution containing 0.2 M Tris-Cl (pH 8.0), 2 mg/ml appropriate substrate, 4 mM $MgCl_2$, 0.2 mg/ml NAD or NADP, 0.25 mg/ml MTT (4,5-dimethyl-2,5-diphenyltetrazolium bromide) and 0.1 mg/ml PMS (phenazine methosulphate).

2.6 Symbiotic assays

2.6.1 Surface sterilzation of seed, plant growth and inoculation

Alfalfa seeds (var Iroquois) were surface sterilized in 95% ethanol (5 min) followed by 50% bleach (15 min) and rinsed several times in water (~ 10 times for 2 min each), and germinated in dark on sterile water agar plates. Two day old seedlings were transferred into sterile vermiculite saturated with Jensen's nitrogen free media in Leonard assemblies (Leonard, 1943). Five seedlings were planted per pot and each strain was tested in triplicates. Jensen's media contains (/l): 1g $CaHPO_4$, 0.2 g K_2HPO_4, 0.2 g $MgSO_4.7H_2O$, 0.2 g NaCl, 0.1 g $FeCl_3$, 1 ml 1000 X trace elements, pH 7.2. 1000 X trace elements contained (/l): 0.1 g H_3BO_3, 0.1 g $ZnSO_4.7H_2O$, 0.05 g $CuSO_4.5\ H_2O$, 0.05 g $MnCl_2.4H_2O$, 0.1 g $Na_2MoO_4.2\ H_2O$, 1 g $Na_2EDTA.2H_2O$ and 0.2 g NaFeEDTA. The seedlings were kept in growth chamber (Pericerval, 16 hours of light at 25°C and 8 hours of dark at 20°C, light intensity 300 μmoles $m^{-2}\ s^{-1}$). Two days following transfer into vermiculite, the seedlings were inoculated with 5 ml of a 1:50 dilution (in sterile distilled water) of a saturated TY culture (10^{-8} – 10^{-9}) cells. Plants were grown for 4 - 6 weeks and examined for nodulation phenotype and symptoms of nitrogen starvation indicative of ineffective symbiosis.

Shoot dry weight was determined after separating the root and shoot systems and desiccating at 65°C for at least 10 days. Shoot dry weight of 5

plants from a single pot were measured for each replicate and shoot dry weight was expressed as an mg per five plants.

2.6.2 Isolation of bacteria from nodules

To isolate bacteria from nodules, the nodules were surface sterilized with 1% Na- hypochlorite (15 min), washed twice with sterile TY broth, squashed in a few drops of sterile TY. The resulting suspension was spread on TY agar and incubated at 30°C for 3 days and then scored for growth.

2.7 Microscopic techniques

2.7.1 Confocal laser scanning microscopy

2.7.1.1 Nodule sectioning, staining and visualization

Nodules were harvested into 80 mM HEPES pH 7.0, transferred to 4% formaldehyde in 80 mM HEPES and vacuum infiltrated 3 × 30 s, venting completely between each infilteration. The nodules were fixed with rotation for 45 min at room temperature, rinsed 2 × 5 min with 80 mM HEPES buffer, and transferred to ice cold 80% ethanol. Nodules in ethanol were stored at -20°C for 45 min, then rinsed 2 × 5 min with 80 mM HEPES buffer. Nodules were moved to fresh HEPES buffer and sectioned longitudinally using a double-edged razor blade. Cut nodules were stained in 80 mM HEPES with 1 μl/ml (0.1%) SYTO13 (Molecular Probes, Inc., Eugene, OR, USA) for 15 min.

Stained nodules were transferred to a Laboratory- Tek II chambered #5 coverglass system (Nalge / Nunc International, Naperville, IL, USA) in a small

volume of staining solution, and gently covered with a glass coverslip to minimize sample movement and position the sample closer to the coverslip. Confocal images were acquired on an inverted Zeiss LSM 510 NLO laser-scanning microscope (Carl Zeiss, Inc., Germany). Images of nodules stained with SYTO13 and plant autofluorescence were acquired using 488 and 543 nm wavelength helium Neon Laser lines with 500 - 550 band-pass and 560 long-pass emission filters, respectively. The SYTO fluorescence is depicted in green and the plant autofluorescence in red.

2.7.2 Light microscopy

2.7.2.1 Nodule sectioning, staining and visualization

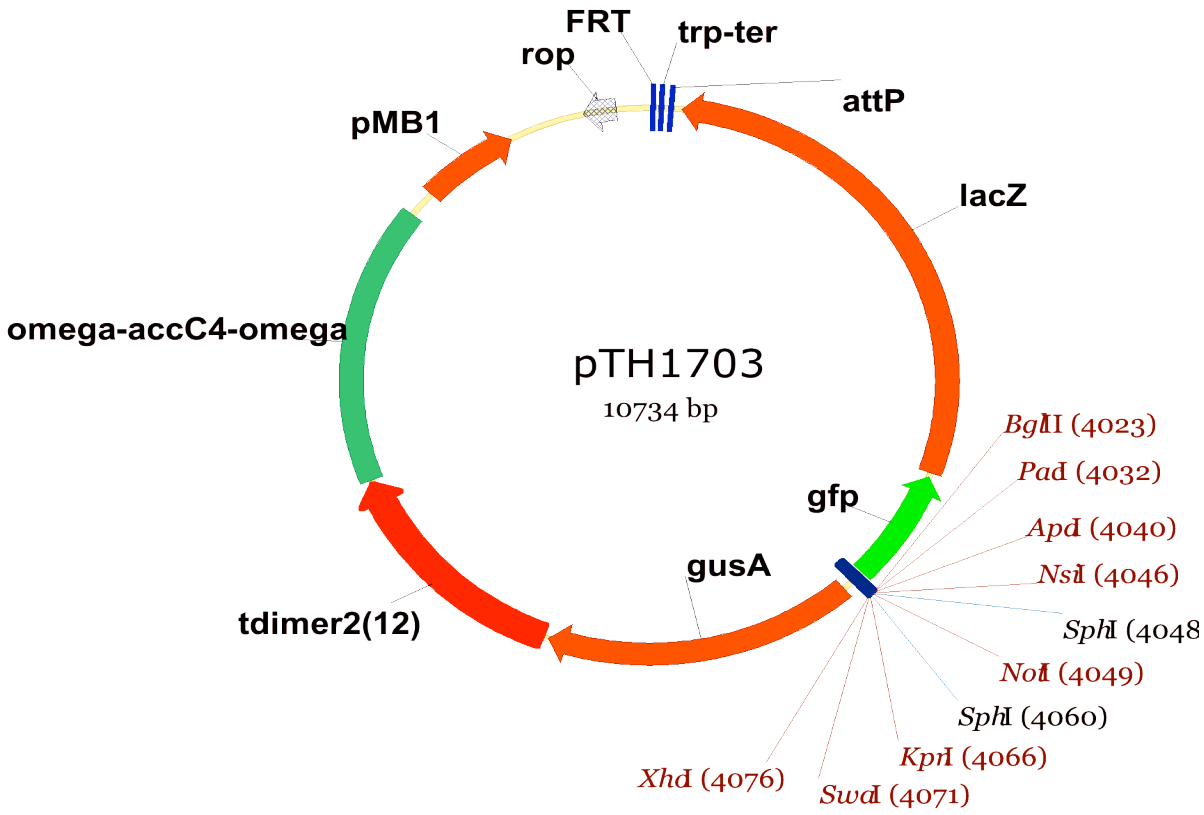
Nodules were harvested in sterile distilled water and fixed in 10% formalin for 24 h and dehydrated by washing in a gradient of ethanol (20% - 1 × 10 min, 50% - 1 × 10 min, 70% - 3 × 10 min, 95% - 2 × 10 min and 100% - 2 × 10 min). Thereafter the nodules were washed in xylene : ethanol 1 : 1 (1 × 10 min), xylene (2 × 10 min), paraffin : xylene 1 : 1 (1 × 10 min), paraffin (2 × 5 min) and embedded in paraffin for sectioning.

5 micron thick ribbons of embedded nodules were obtained using a microtome. The ribbons of paraffin with sections of nodules were placed in water on a slide coated with albumin and dried on a slide dryer at 60°C for 48 hr.

The sections were washed in xylene (1 × 5 min, 1 × 3 min), 100% ethanol (1 × 5 min, 1 × 3 min), 70% ethanol (1 × 5 min), 50% ethanol (1 × 3 min) and distilled water (1 × 1 min). They were stained in toludine blue (pH

4.5) for 20 sec, washed in distilled water (2 × 1 min), rehydrated by washing in a gradient of ethanol (50%- 1 × 3 min,70%- 1 × 5 min, 100%- 1 × 5 min and 1 × 3 min), rinsed in xylene (2 × 5 min) and mounted using permount. The slides were air dried for 48 hr and observed under a light microscope at 40X and 400X magnification.

Fig.2.1. Map of pTH1703



Cowie *et al.* 2006

Table 2-2. Primers used in this study

SDRs	Forward primer 5'-3'	Reverse primer 5'-3'	Reverse primer600 5'-3'	Forward primer (upstream of start codon) 5'-3'
SMa0059	ATGAGCGATTTCAACGGCAAGA	TCAATCTGGAATGACGATGC	GTCCTTTGCCCCAGGTTT	
SMa0074	ATGACAGAGCTTTCCGGAAAG	TCACCAGCCGCCGTCGATCGC	GCGCGCCTTCAGCGACTGGTT	
SMa0187	ATGTATGATGCACCGTTTTACA	TCAATAACCGCCGACGATC	TCTTTGTCCGATGGATTGAG	
SMa0320	ATGACCAGCCTCAACGGAAAG	ACCAACCCTGTTCTACTGGCA	AGGAAGGTCATCTGAGTGGCC	
SMa0326	ATGAGCACGCAAAACCCTAA	GGATTGGTTCAGTTTGCCG	CTGCTGCCAGGAGAGCAT	
SMa0329	ATGTCAAAGCGATTTGATGGAAAAGTAG	TCATTTGCCCCATTCCGGG	TTCATCACTGGCAAGAAATGCGAC	
SMa0335	ATGAGCACCGTAACATTCTCGACG	TCAGATGACGGCGAGCAGACC	CTCGTGCGAAGAAGCTCCGT	
SMa0339	ATGTCCTCGCTTTTCTCAAACA	TCAGAGTTCAAAACCACCGATG	CTGCTGACTGTTTGCCGC	
SMa0389	TTGGCCGAGGCCGGTGC	TCAGATCGAGGCGAGGTAGCCG	CATCGCTGCAGAGGAAAACCGC	
SMa0513	ATGAGCACCGAACTCTTCGACCTG	TCAGAGCGAGGCCGTGATGC	CAGTTCCTCCACTTTGCCCCATC	
SMa0719	GTGACCGGCTCGAGCCGCG	CTAGAAGCCACCCGACAACACGACG	TTCGTCGGGCCTGCCTGCG	ATATTTTCGCTGAAGGGCAGA
SMa0854	ATGTTCGAATTGACCGGGCGCA	TCAAATCATTGCCATACCGCCGTTC	GGATCGGGATCGCCACCATGAT	
SMa0959	ATGAGTCGCGGTTCCACGTCAAC	TCAAAGTGGCACGAATTCTCCGC	ACCGCCGCACGATTGACTAGG	
SMa1367	ATGCGGCTTTTCTTGCACA	CTATGCCAGCAGCAGTCCG	GTCCCGGGTGACAGAAGGA	
SMa1398	GTGAGTAATCCGACCGCGAAA	AAAATCGCGGAACGAGCATG	GATAGGAGTCGATCACCGGAGC	
SMa1452	ATGCAACTCAAGAGCCGCG	CATGCGCAAAGCCCCGTC	TGAACATCCTGAGGCATGCC	
SMa1629	ATGATCGAATTTCTCAACCTCA	CTACGCTGTGGGCACCGTCCCCCC	GCCCTCAAGGTCCGTTCC	
SMa1757	ATGTATCAAGCATCAAAAGAGAGC	TCAGACGTTTGTCCTCCGT	TAGACGTTTCGATTTCGCC	AAGAGGGGCTTGCTTTTCAG
SMa2019	ATGGTTGCCCGAGGATGC	CTACGCCGTGGGCACCGTACCTC	ATCCATGATGATCTTCTTGCCG	
SMa2165	ATGACCAGGTTTACTGGGAA	TCAGTGCATGATCAGTCCG	CGGCACTTGAGGTAATACG	
SMa2343	ATGGACAAGGTCATTCTGATCA	GATGCGGTCGGTCTGATC	GATGTCGGCAGGCTGAAG	CATCGCTCGAGAGCTTGG
SMb20073	ATGGAAAAATTTCCGCACCCTC	TCAAAGGAAAGGCTTGCCGC	GACATTCACTCTTATCCCGCGTTC	
SMb20076	ATGTTGAATTCTCTCGAAGATGC	AAGCGCGCCTGAAGAGAC	CCTGCGACCGGTGTAATTG	
SMb20210	ATGCGTAACCAACAGAAAGTCG	TTACCAGCGTCCGGCGTTCTG	GATCTCGCCCGTGATGAATCCC	
SMb20214	ATGACCCAGCATTCAAAAGG	ACAGTATGGGCAGCGCTATG	AAATGCTCGACAGGTGTGC	
SMb20409	ATGACGGGCAGGCTCAAG	TCACCAAAGTGCGCCGCCCGT	TTCCGTTTCGACCCAGGTCGG	
SMb20456	ATGACATCGATCAGTTTGGCCGGA	CACCGCCGTCGATGGTCAGC	ATACGTTCCCGCTGCGCTTCG	
SMb20492	ATGTCGAAGAAAATCATTCTGATCAC	CAAGGGCCCTGGGTTTTAGT	CGGAATGGGCGAAGTGATT	ATACGACTCCCAAGGCATCA
SMb20493	ATGGAAACGAACAAAGTAGCGA	AGATGATGCCGCCATTGG	GCTATGTCCTGCGGAGTGC	
SMb20511	ATGAGTGACCGGCTTAAGGG	CGTTATAACTCGGGCAATGG	GGAAAGCTGTTCCAGTAGTCG	
SMb20660	ATGCCGATAAAGAGACGTGAT	TCAGACCCGATAATAGCG	CCAATGGGCGTAGTTGTTG	
SMb20662	ATGGCGCAAGGGAAGGGATCGGGC	TTATCCGCGGCCGACCAGCGG	CTCATCCTCGCGGTCATGT	
SMb20692	ATGATGGATGGCGCCTCA	TGATGCCACCGTCGACATAG	GTCTTGAAGTAGCCTGGCGC	
SMb20750	ATGTTCGATCTCAGCGGC	CATCTGCACGGTCATCCC	ATGTCCACGAAGGCTTTGAG	

SMb20871	ATGCACCACCTCAAAGACAAG	CAAGTCAGGGTCATGCCG	AGGAAGCCATGCGGTCTTC	
SMb21010	ATGACCAAGACTGCGGTGA	CTACTGCGCCGTCCAGCCGCC	CACCTGTTCCTCGGTGATG	GCGTCGCAAAAGACCTGT
SMb21111	ATGACAGCGAACCTTGCCG	AAATGGTCCAGCCGCCG	GGCTGGCGGGAGATAAAGG	
SMb21159	ATGCCGAAATCCGTAGCGATCGT	TCAGTAATGGATCGATTGCGAGAAGGT	GTCGACGATACGCTGCAGGGGA	
SMb21348	ATGAGCGTACCGTCGCGCT	CTATCTGGCGAGCCAGCCG	GAGGCGCTGCGTTTTTCGT	
SMb21383	ATGAAAGTGACAGATAACCAGAATGTG	CTACGCAATGCCACCATTCG	ACAGTTCCTGAGCCATTTTCGA	
SMb21384	GTGAAAACGACGGGAAACAC	TCAGCTTGAACCGAGCATCT	CAAGGGCATATAGCCTTCGC	
SMb21474	ATGAACCAGCCGTCGCCCG	GCCGCCGTCGATCCAAACG	GCCGTTTCCTCTTCGTGGCG	
SMc00005	ATGGCTCAGGCATCGGGTCT	TCAACTCTTCAGAACGCTGATGT	CCACTTCAGGATGTAGCGGAAA	GCTGGTGGAGGCAGACAGTA
SMc00136	ATGAGCGAACAGAAGGTGGC	TCAGACGGATCGGGTGATG	ATCAACTCGGCCACCTCCTC	GCTTCAACTGCGCTTCTCTC
SMc00165	ATGAGCACGGACCACGCCC	CGGATGAGGCGAGGCGTCA	TTCGCGCTGGATCCGGTCT	CACGCTCACCAAGGCATATT
SMc00260	GTGAGCGATGGAATGAAATCC	TCATCGTCCTTTCAAAGCCA	GCACGCACGTCATCCGTAT	
SMc00264	GTGGCAGACCGTCTCAAGGGAAAA	TCAGTGTACCCCCATGATCGTGC	GGCGTATTGCGAGGAGATCTGCT	
SMc00268	ATGGGAACCGCTCTACTCGAGG	TCAATGGATCAGCGAGCCG	TTCCTCCAGTTTCTCCGGCG	
SMc00326	ATGAACGGATTGATGAACGG	CGAGGTGATGTTGTAGCCG	GTTCTTCTGCTGCCAGGAGA	GCTCGTCTTCCAGCGCTAT
SMc00372	ATGTCGAGACTGGAAGGTAAA	TCACGCCTTCGTCACCGAAGC	CAATGCATGCAGCCCTTC	
SMc00553	ATGACGCCCAATCTGAAAGACC	CATGCGGTAGTCGACGATCTTC	GGCAGGGTCTCCGGATCTTC	
SMc00572	ATGTTCGATCTTTCAGGCCGCA	TCAGATCATCGCCATGCCGC	GCATCGGGATCGCAGACATGAT	
SMc00603	TTGAAAAGGACACTCAAGGCG	ATTCCACTATCTCCCGAATATCC	GACCTGGGCCTCGAAATCAC	CCACTCGGCATCAGGATT
SMc00733	ATGCCGGATCGCCCCACCAT	TCAGCCGAGGCGACCGATGAT	CATCTGCCGCTCGTATTCCGCAC	
SMc00778	GTGGCCGGTCGATTGCATG	GTCGACGAACAGATGGGCGC	GCACATCCGCCCCTGTTGG	CAAAGGCAGCACTACGACAG
SMc00880	ATGACTTTGCCGAGCAGCCA	GATGATCATGGCCTGTGCGG	TTCCGTGAGCCACAAGCGC	
SMc01157	ATGGATCTCGGCATTAGCGG	TAGAGCATCGAACGGCTGA	GTTTCGACGCTGACGCCTTC	
SMc01175	ATGACAGCTGTATTCACCGCC	CCTGAGCGCTTGTTCTTTGG	CCTCCTGCGTGCTCACCAG	
SMc01204	ATGAGGCCGGGCGATCAAC	TTGCGTGGTGGTTCCGAGC	TGAACAGTGCCACGCGATTG	CTGGAAGGGCTTGCATCC
SMc01500	ATGAAACGTCTTGAAGGAAAGAGC	ATCGACGTTATAGGTCTGCGAGAC	CGGACGGTTCTCGTATCTGG	
SMc01571	ATGGGCAACTATCAGGGAAAGAA	CGAGAGCTTCTGGCCGAGAC	GTCGCCGAGCGTCTTGAAC	
SMc01635	ATGATCCTGAACAACCGGAT	TCAGACCACCGTCTCGCC	CGGATATAGGCTTCGATGG	
SMc01698	ATGAAACTCGGCCTGGAAGGAAAAA	AATGCGCCGTAAAAGCCCCG	CTTCCAGGCGTCGTGTCTTCG	
SMc01955	ATGTTGTCCTACTCCGCTATCGACAAGA	CGGTTTGGTCCTCCCTGCCG	CTTGCGCGGTGAGGGACGTG	
SMc01991	ATGGAATTCTCCGGAAAATCAGTCATCG	GCAAGGAACCCGCCGTCGAC	CGGCGGGATGACGGACCATC	GTACCAATCGCAAACCGACA
SMc02034	ATGTACCTGAGGAAATTCGATCTTT	TTACCAGCAGGTATAGCCGGC	TTGTCCTCCTTGGCGAAAGT	
SMc02037	GTGTACGCTGAGCTTTTCAGA	TCAGTGGATGGTGTAGCCA	GAGCCAGATCTTCGACCATT	
SMc02039	ATGACCGGCAGAAAGCCGCTCGT	TCATCTCCTGGAGATTGACGTCCTGCG	GCATCCAGCCCGCCGTCGAG	
SMc02040	ATGGCGCGATCTCGTCGACAT	TCAGAATTCCTGGCCGCCTGTGTGGGT	GCAGCTGTTTGCGGCCGCTGTC	

SMc02041	ATGACGGAAATTCCCTCCTT	TCAGGTGATCGTGTAACCGC	GCCTTTCTCTCCAGCCCAG	
SMc02271	ATGACACATGTGATCATCAC	TCACAGTCCACTGTTGCGCGCGA	GTTCCCATGATTACGCTTGC	
SMc02322	ATGCTCGACAAGCACCAGGGCG	TCAGCGGGTGAAGGACTGGGCA	CCCAGGTGAAGAGGCCGTGGC	
SMc02336	ATGAAAGACATCATCGATACGTTTCGT	CTAAAAGCCGCCGACCACGA	AACGCTGGGCGATCTTCTCC	
SMc02339	GTGTATCTCGAACGCTTCAGGC	TCACCAGAGCGTGTATCCTCCA	AGCATGGGCCTGTCGTTCTC	
SMc02356	ATGATCAAGACTTCCAGACTGTTT	GCTTAGGCAGGGATTCCGG	AAAAATTGTTGTCGCCGTTC	
SMc02486	ATGAGCGACGGACCCGTGCTTCT	CTACCGCCCGCCGCTGACG	CGGGATCGTGGGTGCGAGGT	ACCGACAGTTCCGATGTCTT
SMc03878	ATGAGCAGGGTAGCACTGGTAAC	CGAAGTACTGGCCGCCAT	GAGAAACACGACGCAGCGT	
SMc04391	ATGACCCAGAAGGCACGCCC	TCACAACCGCCCGATCGAGAG	CCGCTTTCAATCAGACCGTCG	
SMc02522	ATGTATCATCCGGCCTTGTTCAAG	TCAAGCGATGAGCTGCCCA	CAGCTTCTCCGGCGTCGAAT	
SMb21348	ATGAGCGTACCGTCGCGCT	CTATCTGGCGAGCAGCCG	GAGGCGCTGCGTTTTTCGT	
pVIK112		CTGCAGGTCGACCATGGTCATAGC		
gfp (pTH1703)		ACAGTTTTCGCGATCCAGAC		
gus (pTH1703)		CCCTCTCCACTGACAGAAAATTTG		
T7 primer		TAATACGACTCACTATAGGG		
SP6 primer		TATTTAGGTGACACTATAG		

CHAPTER 3

Phenotypic analysis of pSymB deletion strains

3.1 Introduction

Complete sequence of the pSymB megaplasmid revealed that is 1,683 kb in size with very high gene density. It has the only copy of an essential arg-tRNA gene and *minCDE.* Almost 20% of pSymB constitutes genes encoding of solute uptake system and 14% is dedicated to exopolysaccharide biosynthesis. Other gene clusters include many involved in catabolism of substrates like protocatechuate and phosphonates (Finan *et al.* 2001). A 200 kb region in pSymB needed for adaptation to saline shock which contains a high density of osmoregulated genes, was identified by microarray studies (Domínguez-Ferreras, A. *et al.* 2006) . Another interesting feature is the large number of SDRs present on pSymB. 21 SDR encoding genes have been identified on this megaplasmid (Chapter 4), and this probably contributes its catabolic properties.

Before the commencement of this study, defined deletions had been created in the megaplasmid pSymB (Fig.3.1.) by homologous recombination between the insertion elements of flanking transposon insertions in strains generated during the construction of the pSymB genetic linkage map (Charles and Finan 1991). Preliminary studies have been carried out to study the carbon utilization phenotype of these deletion strains. The deletion strains RmG373 (and the overlapping deletion strains RmF726 and RmF728), RmG462 (and overlapping deletion strains RmG470, RmG471, RmG472, Rm5408, Rm5416 and RmF114) and RmF909 (and overlapping deletion strains RmF680,

RmF514 , RmF693, RmF638 and RmF666 were tested on a variety of carbon sources for a phenotype. Compared to the wild type Rm1021, used in this study RmG373 and its subset were unable to utilize β-hydroxybutyrate and acetoacetate, α-galactosides raffinose and melibiose and aromatic compounds protocatechuate and quinate. Subsequent analyses have led to the identification of genetic loci involved in the catabolism of these substrates (Charles *et al*. 1997, Aneja *et al.* 2002, Gage and Long, 1998, MacLean *et al*. 2006). RmG462 and its subset were unable to utilize galactitol. RmF909 and subsets did not show a phenotype of any of the substrates tested. The emphasis of this present study is to test the deletion strains on a broader range of carbon sources and identify loci responsible for carbon utilization phenotypes.

The work presented in this chapter focuses on:

I) Screening of deletion strains on 93 different carbon sources.

II) Phenotypic complementation of the deletion mutants using wild type Rm1021 pLAFR1 cosmid library

III) Genome sequence facilitated restriction analysis of the complementing cosmids to identify loci on pSymB that are involved in the metabolism of the identified substrates.

3.2 Materials and methods

All the procedures done in are as described in Chapter 2. unless otherwise mentioned.

3.3 Results and Discussions

3.3.1 Screening pSymB defined deletion mutants for subtrate utilization

Each of the pSymB deletion strains was screened on a series of RMM agar plates containing single carbon sources at 0.1% concentration. A total of 93 different carbon sources were tested (Table 3-1). Each screen was done in triplicate. The carbon utilization phenotype was tested for each of the deletion mutant strains. Each of the deletion strains grew well on glucose as sole carbon source but each one had at least one deficiency in carbon source utilization (Table 3-2).

3.3.2 Complementation of the pSymB deletion mutant strains with Rm1021 cosmid library and restriction analysis of the complementing clones

The pLAFR1 EcoRI partial digest Rm1021 library was introduced en masse into the mutant strains by triparental conjugation. This resulted in growth of colonies on RMM media supplemented with the corresponding the carbon sources and tetracycline. No growth was observed on control plates containing donor, mobilizer and recipient individually. Following streak purification of individual colonies on RMM medium thrice, they were transferred to TY medium containing streptomycin and tetracycline. The complementation was done with RmG373, RmF726 and RmF728 as recipient on ornithine GABA, uridine and leucine; and Rm909, RmF680 and RmF693 as recipient on D-maltitol, palatinose and hydroxy-L-proline The cosmids were then transferred back into *E. coli* DH5α by triparental conjugation, selecting on LB contain

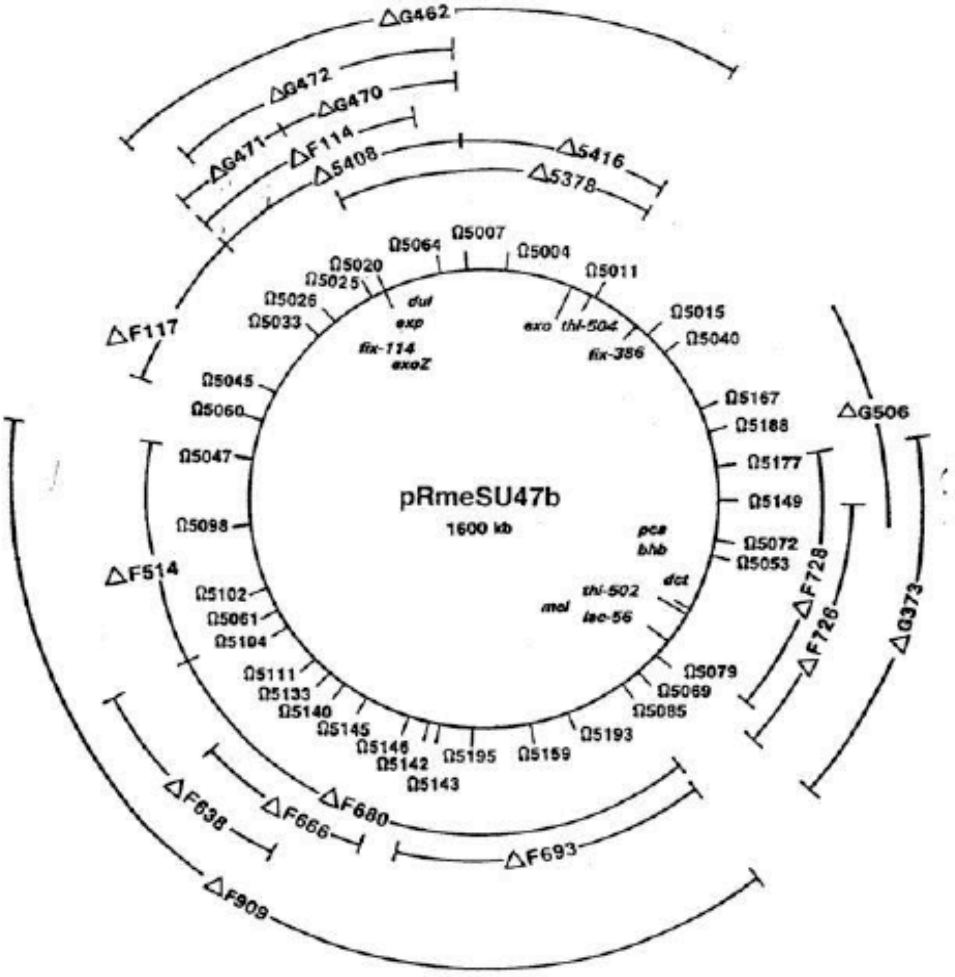
tetracycline at 37°C. 10 ml TB cultures yielded sufficient cosmid DNA for detection by ethidium bromide stained agarose gel electrophoresis following digestion by EcoRI. These cosmids were then conjugated back into the corresponding deletion mutants to reconfirm the phenotypes.

EcoR1 restriction digest analysis established the presence of a single common fragment in each clone corresponding to the 20.4 kb pLAFR1 fragment (Vanbleu *et al.,* 2004). The restriction patterns obtained for each complementation set were analyzed visually and common bands identified in each. These bands defined a maximal complementation region for that particular substrate. The common fragments for each set obtained were compared to the virtual digest of the common deletion region to determine where in the deletion region the complementing cloned fragments lie. Table 3-3 lists the genes present in each complementing region, as annotated on the genome sequence database (http://bioinfo.genopole-toulouse.prd.fr/ annotation/iANT/bacteria/rhime/).

The complementing regions for three amino acids, ornithine, GABA and leucine, contained a 7-kb common region (from 1,534,910 bp- 1,542,313 bp). This region contains genes SMb207460-SMb20751. The region conferring phenotype on D-maltitol and D-palatinose was mapped to the same region by this screen (327604bp-353349bp) containing genes SMb20324-SMb20341. Similarly, region for catabolism of uridine included SMb20594-SMb20608, for hydroxy-L-proline included genes from SMb20257-SMb20277and galactitol included genes SMb21373-SMb21385.

The genes located in the complementation regions (Table 3-2) were compared with the degradation and related pathways (KEGG Pathway Database, http://www.genome.jp/kegg/pathway.html) for each substrate to identify genes that are likely to be responsible for the phenotypes. Several possible genes were identified for each substrate.

Fig.3.1. pSymB deletion map



From Charles & Finan, 1991, Genetics 127:5-20

Table 3-1. Substrates used in the carbon utilization screens of SDR mutant strains

I] Carbohydrates

a) Pentoses and derivatives:
D-ribose
2-deoxy-D-ribose
D-xylose
D-lyxose
D-arabinose
L-arabinose
methyl-β-D-xylopyranoside
D-ribono-γ-lactone

b) hexoses and derivatives
D-glucose
L-glucose
D-galactose
D-fructose
D-mannose
D-tagatose
D-fucose
L-fucose
L-rhamnose
palatinose
arbutin
D-galactono-γ-lactone
methyl-α-D-galactopyranoside
methyl-β-D-galactopyranoside
N-acetyl-D-glucosamine
N-acetyl-D-galactosamine
3-O-methyl-D-glucopyranose

c) Oligosaccharide- α-galactoside
melibiose
raffinose

d) Oligosaccharide- β-galactoside
lactose
lactulose
lactic acid

e) Oligosaccharide- α-glucoside
trehalose
turanose
maltose
sucrose
maltotriose
dextrin
inulin

f) Oligsaccharide- β-glucoside
gentiobiose
cellobiose
salicin

g) Sugar alcohols
sorbitol
adonitol
D-mannitol
meso-erythritol
dulcitol
maltitol
inositol
glycerol
D-arabitol
L-arabitol
ethanol
propanol

h) Sugar phosphates
glucose-1-phosphate
fructose-6-phosphate

II} Carboxylic acid and derivatives
pyruvic acid
aminovaleric acid
formic acid

III] TCA cycle intermediates and derivatives
D-malic acid
L-malic acid
DL-malic acid
fumaric acid
sodium succinate
monomethyl succinate
citric acid
DL-bromosuccinic acid

IV] Amino-acids and derivatives
L-alanine
L-lysine
L-ornithine
L-proline
trans-4-hydroxy-L-proline
glycyl-L-proline
L-serine
L-homoserine

IV] Amino-acids and derivatives
L-leucine
L-isoleucine
L-histidine
L-valine
DL-aspargine
L-aspartic acid
carnitine
L-glutamic acid
glutamine
γ-aminobutyric acid
alanineamide

V] Dipeptides
Gly-Asp
Ala-gly
Glu-glu

VI] Nucleotides and derivatives
uridine
thymidine
adenine hemisulfate

Table 3-2. Carbon utilization phenotypes for (on RMM agar)

I] RmG373*, RmF726✝, RmF728*, RmF117✝		Smallest complementing region on pSymB
Lactose	RmG373, RmF726, RmF728	Ω5079-Ω5149
Lactulose	RmG373, RmF726, RmF728	Ω5079-Ω5149
D-Raffinose	RmG373, RmF726, RmF728	Ω5079-Ω5149
Succinate	RmG373, RmF726, RmF728	Ω5079-Ω5149
Maltotriose	RmG373, RmF726, RmF728	Ω5079-Ω5149
Melibiose	RmG373, RmF726, RmF728	Ω5079-Ω5149
Methyl-α-D-galactopyranoside	RmG373, RmF726, RmF728	Ω5079-Ω5149
Methyl-β-D-galactopyranoside	RmG373, RmF726, RmF728	Ω5079-Ω5149
D-Fucose	RmG373, RmF726, RmF728	Ω5079-Ω5149
Ethanolamine	RmG373, RmF726, RmF117	Ω5079-Ω5149, Ω5060-Ω5026
N-Acetyl-D-galactosamine	RmF117	Ω5060-Ω5026
Thymidine	RmG373, RmF726, RmF728, RmF117	Ω5079-Ω5149, Ω5060-Ω5026
Uridine	RmG373, RmF726, RmF728	Ω5079-Ω5149
Homoserine	RmG373, RmF726, RmF728, RmF117	Ω5079-Ω5149, Ω5060-Ω5026
Aminovaleric acid	RmG373, RmF726, RmF728	Ω5079-Ω5149
γ-Aminobutyric acid	RmG373, RmF726, RmF728	Ω5079-Ω5149
Ornithine	RmG373, RmF726, RmF728	Ω5079-Ω5149
L-Alanine	RmG373, RmF726, RmF728	Ω5079-Ω5149
L-Glutamic acid	RmG373, RmF726, RmF728	Ω5079-Ω5149
Glutamine	RmG373, RmF726, RmF728	Ω5079-Ω5149
Glycyl-L-proline	RmG373, RmF726, RmF728	Ω5079-Ω5149
L-Leucine	RmG373, RmF726, RmF728	Ω5079-Ω5149
II] RmG462★, RmF470★, RmF471★, RmF472★, RmF114		
Succinate	RmG462, RmF470, RmF471, RmF114	Ω5026-Ω5033
D-Tagatose	RmG462, RmF470, RmF472, RmF114	Ω5064-Ω5007
N-Acetyl-D-galactosamine	RmG462, RmF471, RmF114	Ω5026-Ω5033
Aminovaleric acid	RmG462, RmF470, RmF472, RmF114	Ω5020-Ω5007
Alanineamide	RmG462, RmF470, RmF471, RmF472, RmF114	Ω5026-Ω5064

Glutamine	RmG462, RmF471, RmF114	Ω5026-Ω5033
L-Lysine	RmG462, RmF471, RmF472, RmF114	Ω5033-Ω5020

III] RmF909*, RmF514*, RmF680*, RmF693*

Maltotriose	RmF909, RmF680	Ω5143-Ω5104
D-Fucose	RmF909, RmF680, RmF514	Ω5143-Ω5104, Ω5104-Ω5047
Palatinose	RmF909, RmF680	Ω5143-Ω5104
D-Maltitol	RmF909, RmF680	Ω5143-Ω5104
Hydroxy-L-proline	RmF909, RmF680, RmF693	Ω5143-Ω5085
Carnitine	RmF909, RmF514,RmF680, RmF693	Ω5143-Ω5085, Ω5104-Ω5047

*- poor growth on maltotriose, L-glutamic acid and glycyl-L-proline, *- poor growth on maltotriose and L-alanine, + - poor growth on maltotriose, ⋆- poor growth and dispersed colonies on alaninamide, ☆ - poor growth and dispersed colonies on carnitine

Table 3-3. Complementing region identified for substrate utilization loci:

Substrate and complementing region	Gene	Annotation
γ-aminobutyric acid (GABA) 1,534,826-1,561,102 (26,276bp)	SMb20746	hypothetical protein
	SMb20747	hypothetical enzyme,haloacid dehalogenase/ epoxide hydrolase family
	SMb20748 *pssF*	putative glycosyltransferase
	SMb20749 *uxuB*	putative D-mannoate oxidoreductase
	SMb20750	putative dehydrogenase, possibly gluconate-5-dehydrogenase
	SMb20751	putative 3-hydroxyisobutyrate dehydrogenase
	SMb20752	putative enoyl-CoA hydratase
	SMb20753	putaive acyl-CoA dehydrogenase
	SMb20754	conserved hypothetical protein
	SMb20755 *pccB*	putative propionyl-CoA carboxylase β chain
	SMb20756 *pccA*	putative propionyl-CoA carboxylase α chain
	SMb20757 *bhbA*	methylmalonyl-CoA mutase
	SMb20758	putaive transcriptional regulator (GntR/ArsR)
	SMb20759 *phnG-J*	putative carbon-phosphorus lyase component
	SMb20763-64 *phnK-L*	putative phosphonate uptake ABC transporter ATP-binding protein
	SMb20765	putative acetyl transferase
	SMb20766	putative transposase (ISRm22)
	SMb20767 *dak*	putative hydroxyacetone (glycerone) kinase
Ornithine 1,528,768-1,542,213 (13,445bp)	SMb20724	conserved hypothetical exported protein
	SMb20725	hypothetical membrane protein
	SMb20726	conserved hypothetical membrane protein
	SMb20727	hypothetical protein
	SMb20728	hypothetical protein
	SMb20729	hypothetical protein
	SMb20745 *glnII*	glutamine synthetase II
	SMb20746	hypothetical protein
	SMb20747	hypothetical enzyme, haloacid dehalogenase/epoxide hydrolase family
	SMb20748 *pssF*	putative glycosyltransferase
	SMb20749 *uxuB*	putataive D-mannonate oxidoreductase
	SMb20750	putative dehydrogenase,possibly gluconate-5-dehydrogenase
	SMb20751	putative 3-hydroxyisobutyrate dehydrogenase
Leucine 1,534.910-1,546,728	SMb20746	hypothetical protein
	SMb20747	hypothetical enzyme, haloacid dehalogenase/epoxide hydrolase family

(11,818bp)	SMb20748 *pssF*	putative glycosyltransferase
	SMb20749 *uxuB*	putataive D-mannonate oxidoreductase
	SMb20750	putative dehydrogenase,possibly gluconate-5-dehydrogenase
	SMb20751	putative 3-hydroxyisobutyrate dehydrogenase
	SMb20752	putative enoyl-CoA hydratase
	SMb20753	putaive acyl-CoA dehydrogenase
	SMb20754	conserved hypothetical protein
	SMb20755 *pccB*	putative propionyl-CoA carboxylase β chain
Uridine	SMb20594	putaive amicynin precursor
1,607,195-1,622,847	SMb20595	cell division protein FtsK-like protein
	SMb20596	hypothetical protein
(15,652bp)	SMb20597	conserved hypothetical membrane protein
	SMb20598 *repA3*	probable replication protein
	SMb20599 *repB3*	putative replication protein
	SMb20600	hypothetical protein
	SMb21707 and	putative urea/short-chain amide or branched
	SMb20602-5	amino-acid uptake ABC transporter
	SMb20606	hypothetical protein
	SMb20607	conserved hypothetical protein
	SMb20608	putative trancriptional regulator, ArsR family
Hydroxy-L-proline	SMb20257	putative adenylate cyclase protein
262074bp-282645bp)	SMb20258	putative transcriptional regulator protein
	SMb20259	putative dihydropicolinate synthase
(20572bp)	SMb20260	conserved hypothetical protein
	SMb20261	putative malate dehydrogenase
	SMb20262	putative semialdehyde dehydrogenase
	SMb20263	amino acid ABC transporter ATP-binding protein
	SMb20264	amino acid ABC transporter permease protein
	SMb20265	amino acid ABC transporter permease protein
	SMb20266	amino acid ABC transporter ATP-binding protein
	SMb20267	putative amino acid dehydrogenase
	SMb20268	putative proline racemase
	SMb20269	conserved hypothetical protein
	SMb20270	putative proline racemase
	SMb20271	hypothetical protein
	SMb20272	putative transmembrane transporter
	SMb20273	hypothetical protein
	SMb20274	conserved hypothetical protein
	SMb20275	hypothetical protein
	SMb20276	putative transcriptional regulator protein
	SMb20277	putative aminotransferase

Palatinose and D-maltitol 327604bp-353349bp (25741bp)	SMb20324 *thuR*	putative *thuR*, regulatory protein for trehlose/ maltose transport system
	SMb20325 *thuE*	probable trehalose/maltose binding protein
	SMb20326 *thuF*	probable trehalose/maltose transporter permease protein
	SMb20327 *thuG*	probable trehalose/maltose transporter permease protein
	SMb20328 *thuK*	probable trehalose/maltose transporter ATP-binding protein
	SMb20329 *thuA*	probable *thuA* protein, unknown function
	SMb20330 *thuB*	probable trehalose/maltose utilization protein, unknown function
	SMb20331	hypothetical protein
	SMb20332	hypothetical protein
	SMb20333	probable transporter protein
	SMb20334	hypothetical protein
	SMb20335	conserved hypothetical protein
	SMb20336	conserved hypothetical protein
	SMb20337	putative trancriptional regulator protein
	SMb20338	putative transembrane efflux protein
	SMb20339	hypothetical transmembrane protein
	SMb20340	hypothetical protein
	SMb20341	hypothetical protein
	SMb20342	putative aldehyde dehydrogenase protein
	SMb20343	putaive aldehyde dehydrogenase subunit protein
	SMb20344	putative transcriptional regulator protein
	SMb20345	putative transmembrane efflux protein
Galactitol 1037716bp-1049276bp (11,561bp)	SMb21373	putative sugar kinase
	SMb21374	putative sugar kinase, PfkB family
	SMb21375	putative sugar uptake ABC transporter permease protein
	SMb21376	putative sugar uptake ABC transporter ATP-binding protein
	SMb21377	putative sugar uptake ABC transporter periplasmic solute-binding protein
	SMb21378	hypothetical protein
	SMb21379	conserved hypothetical protein
	SMb21380	putative dioxygenase
	SMb21381	hypothetical membrane protein
	SMb21382	hypothetical protein
	SMb21383	putative SDR family dehydrogenase
	SMb21384	putative SDR family dehydrogenase
	SMb21385	putative transcriptional regulator

3.3.3 Amino-acids: ornithine, leucine and GABA

The complementing clones for ornithine, GABA and leucine contain several genes involved in amino acid metabolism. The only gene common to all the three substrates and that identified to be involved in amino acid metabolism codes for a putative 3-hydroxyisobutyrate dehydrogenase (SMb20751). This enzymes is involved in the degradation pathway for valine (KEGG pathway 00280). It catalyses an intermediate step in the eventual conversion to succinyl-CoA, a TCA cycle intermediate. Although this enzyme seems to be involved only in valine catabolism, it could be possible that *S. meliloti* has unique or modified pathways for the degradation of these amino acids with this enzyme being a key link in this process. The gene SMb20750, an SDR, required for the catabolism of the three substrates has been shown to be required for growth on leucine and ornithine (Chapter 4).

The leucine complementing clones contain two genes, encoding enoyl-CoA hydratase (SMb20752) and acyl-CoA dehydrogenase (SMb20753), which code for enzymes involved in the degradation of leucine. Another gene, SMb21121 (putative isovaleryl-CoA dehydrogenase), which is located in the region deleted from RmG373, RmF726 and RmF728 also belongs to the acyl-CoA dehydrogenase family. This gene could probably carry out the role of SMb20753. Therefore, SMb20753 may not be essential for leucine degradation by the mutants used in this study. SMb20752 is not directly involved in the degradation of acetyl-CoA, although it is involved in the metabolism of one of the intermediates, 3-methylbutenoyl-CoA.

Methylmalonyl-CoA mutase (*bhbA*) is required for the conversion of methylmalonyl-CoA, a product from the metabolism of valine and isoleucine to succinyl-CoA. However this gene does not appear to be directly involved in the degradation of the three amino acids of interest. An intermediate in this pathway is propionyl-CoA, which enters the propanoate metabolism pathway (KEGG pathway 00640), where methylmalonyl-CoA mutase is involved. This enzyme is required for growth on 3-hydroxybutyrate and acetoacetate (Miyamoto *et. al.* 2003), which are intermediates of butanoate and propanoate metabolism. The locus containing putative 3-hydroxyisobutyrate dehydrogenase, enoyl-CoA hydratase, acyl-CoA dehydrogenase, propionyl CoA carboxylase and methylmalonyl CoA mutase (SMb20751-SMb20757) has been implicated to play a role in the catabolism of branched chain amino acids (Miyamoto, *et. al.* 2003).

Glutamine synthetase, the product of *glnII* (SMb20745) found in the complementing regions for ornithine and GABA, converts glutamate to glutamine by incorporation of an amino group. This enzyme may be required for GABA metabolism by a single-step conversion between GABA and glutamate (KEGG pathway 00251). The link to ornithine may be due to the central involvement of ornithine in the urea cycle. Glutamate is converted to ornithine by two different pathways in the urea cycle (KEGG pathway 00220). Two genes encoding enoyl-CoA hydratase (SMb20752) and an acetyltransferase (SMb20765), are both involved in the degradation of GABA through the butanoate pathway (KEGG pathway 00650), ultimately resulting in

the conversion of GABA to acetyl-CoA. These genes should be further investigated to determine whether they are essential for GABA utilization in *S. meliloti*. A pathway for the utilization of ornithine, GABA and L-leucine has been proposed (Fig.3.2., Fig.3.3.). Individual mutations in genes in these loci would help ascertain which of the genes are common to the pathways for degradation of the amino acids.

3.3.4 Uridine

The complementing region identified for uridine codes for two replication protein, RepA3 and RepB3, which are involved in plasmid segregation (Finan *et al*., 2001). Although these may be connected with uridine catabolism for the purpose of DNA replication, they are not a convincing link to the utilization of uridine as a sole carbon source. The genes encoding the four uncharacterized proteins in this region may be worth analyzing further.

3.3.5 Galactitol (also referred to as dulcitol)

The complementing region identified for galactitol utilization carries several genes identical to the one involved in the utilization of this substrate in *E.coli* (KEGG pathway 00052). Previously this region was mapped to lie been 979918bp-1044125bp (Charles and Finan, 1991). Phenotypic complementation has narrowed the region down to 1037716bp-1049276bp. Two of the genes (SMb21373 and SMb21374) have been annotated as sugar kinases either of which may be involved in the first step of the reaction wherein galactitol is converted to galactitol-1-phosphate. Thereafter by the action of a

dehydrogenase (SMb21380 or SMb21383), it could be converted to D-tagatose-6-phosphate that is then probably converted to D-tagatose 1,6-bisphosphate by a kinase (SMb21373 or SMb21374) and then gets converted to glyceraldehyde-3-phophosphate and dihydroxyacteone phosphate that then finally enters the glycolytic cycle. SMb21383, an SDR, did not show a phenotype on galactitol but SMb20409, another SDR showed poor growth on this substrate. There is s strong possibility that the role of SMb21380 and/or SMb21383 may be carried out by SMb20409. By putting together all the genes identified in the complementing clones a pathway has been delineated for galactitol utilization in *S. meliloti* (Fig.3.4.).

3.3.6 D-palatinose and D-maltitol

The region for the catabolism of D-palatinose and D-maltitol were narrowed down to the deletion in RmF693 based on the phenotypic analysis. The complementing regions for both the substrates appear to be identical. This region was compared with the palatinose utilization region of phytobacterium *Erwinia rhapontici* and *Agrobacterium tumefaciens* and a few similarities were noted (Börnke *et al.* 2001 and DeCosta *et al.* 2003). The ABC transporter system (SMb20325- SMb20328) has been annotated to transports trehalose /maltose. However these transporters could also transport isomers of these compounds; D-palatinose (isomaltulose-, 6-O-alpha -D-glucopyranosyl-D-fructose) or trehulaose (1-O-alpha-D-glucopyranosyl-D-fructose). These transporters have shown an identity of approximately 78% to those identified in *A. tumefaciens* involved for palatinose transport. Mutation in the ATP binding

domain of the transporter in this organism resulted in no uptake of palatinose (De Costa *et al.* 2003). A strong possibility is that the transport function might be missing in the RmF693 deletion strain, rendering it incapable of using palatinose as a carbon source. Also worth mentioning is the fact that though this deletion strain grew on RMM media supplemented with trehalose, it was unable to grow on M9 media supplemented with the same. As the pathway for the utilization of palatinose and trehalose seem to overlap, it is a fair assumption that this region is probably involved in the catabolism of both these carbon compounds.

Similarly D-maltitol a sugar alcohol derived by the hydrogenation of maltose by a oxidoreductase could be further hydrolysed to sorbitol, which could then be acted upon by a dehydrogenase (SMb20342 or SMb20343) and get converted to fructose which could the enter the glycolytic cycle (KEGG pathway 00051).

3.3.7 Trans-4-hydroxy-L-proline

The function of the genes encoded in the complementing region for the utilization of trans-4-hydroxy-L-proline were compared to the KEGG database (KEGG pathway 00330). This region consists of an amino acid ABC transporter (SMb20263-SMb20266) and an amino acid dehydrogenase (SMb20267) which could aid in the conversion of trans-4-hydroxy-L-proline to L-1-pyrroline-3-hydroxy-5-carboxylate which may be acted upon by one of the following dehydrogenases (SMb20261, SMb20262 or SMb20267) and be converted to L-4-hydroxy-glutmate semialdehyde. L-4-hydroxy-glutmate semialdehyde could

be converted to l-erythro-4-hydroxy-glutamste by a semialdehyde dehydrogenase (SMb20262), which could then be acted upon by an aminotranferase (SMb20277) to yield D-4-hydroxy-2-oxoglutarate. D-4-hydroxy-2-oxoglutarate can be converted to pyruvate and glyoxylate by an aldolase. A pathway has been delineated based on the genes identified in this region (Fig.3.5.)

3.3.8 Future work

Analysis of each of the complementing regions can provide an insight into the genes responsible for conferring a phenotype to particular deletion strain. However experiments need to be done to confirm the essentiality or non-essentiality of the genes identified. A possible approach could be mutating individual genes in the complementing region in the wild type strain and checking for its inability to use a particular substrate as a sole carbon source. It is also worthwhile to confirm if the complementing regions are in fact the regions identified by restriction analysis. This can be done by using individual library clones to complement mutants lacking one or a few of the genes expected to be on the insert. For example, mutant of SMb20751, a gene common to the complementing clones of all the three amino acid, should be complemented by all clones isolated from the ornithine, GABA, leucine utilization screens and should not grow on minimal media without complementation. Sequencing a few inserts could also aid in the accurate determination of the complementing regions. Furthermore biochemical analysis

of the genes involved to identify their roles in the said pathways is also required.

3.4 Conclusion

Attempt was made to identify regions on pSymB megaplasmid responsible for the metabolism of methyl-α-D-galactopyranoside, ornithine, GABA, leucine, uridine, D-palatinose, galactitol, D-maltitol and trans-4-hydroxy-L-proline by functional phenotypic complementation. Individual genes within the identified complementing regions were compared to current annotations and also examined against the carbon metabolism pathways to identity genes most likely to be involved in the utilization of the substrate.

Catabolism of amino-acids ornithine, GABA and leucine appear to require some common genes, which may include a putative hydroxyisobutyrate dehydrogenase and a putative gluconate-5- dehydrogenase (consistent with observation in Chapter 4). Other possible genes might include *glnII*, encoding glutmate synthetase for ornithine utilization and *bhbA* encoding methylmalonyl CoA mutase and an acetyltransferase for GABA. ABC transporters probably involved in the uptake of trans-4-hydroxy-L proline; D-maltitol and D-palatinose and galactitol have been identified. Previous work on solute binding proteins in *S. meliloti* has shown other transporters for these substrates. The study reports that SMb20320 is induced by hydroxy-L-proline, SMb20316 by erythritol, a sugar alcohol like maltitol, SMb20325 is induced by trehalose and SMb20410 is induced by galactitol (Mauchline *et al.* 2006). The transporters identified in this study might be involved in transport of multiple substrates, a

phenomenon not uncommon in soil bacteria. Also the loci identified for trans-4-hydroxy-L proline utilization contains several dehydrogenases, 2 gene annotated as proline racemase (for the conversion of trans-4-hydrox-L-proline to cis-4-hydroxy-D-proline) and an aminotransferase (for the conversion of L-erythro-4-hydroxyglutamate to D-4-hydroxy-2-oxoglutarate) (KEGG pathway 00330), all of which have been shown to required for its catabolism.

The complementing region for galactitol contains ABC transporter genes probably required for its uptake, kinases and dehydrogenases and also show an operon arrangement similar to the region involved in the utilization of this compound in *E. coli*. SDR mutant strain SMb20409, not identified in this screen was unable grow utilize galactitol. This dehydrogenase could probably complement the function of SMb21380/SMb21383. The complementing clones for uridine, D-palatinose and D-maltitol did not provide convincing results. However, these regions contain several uncharacterized hypothetical proteins and knowledge about their function could aid in understanding the pathway involved in the catabolism of these compounds.

Fig.3.2. Pathway linking ornithine, GABA and L-leucine catabolism in *S. meliloti*

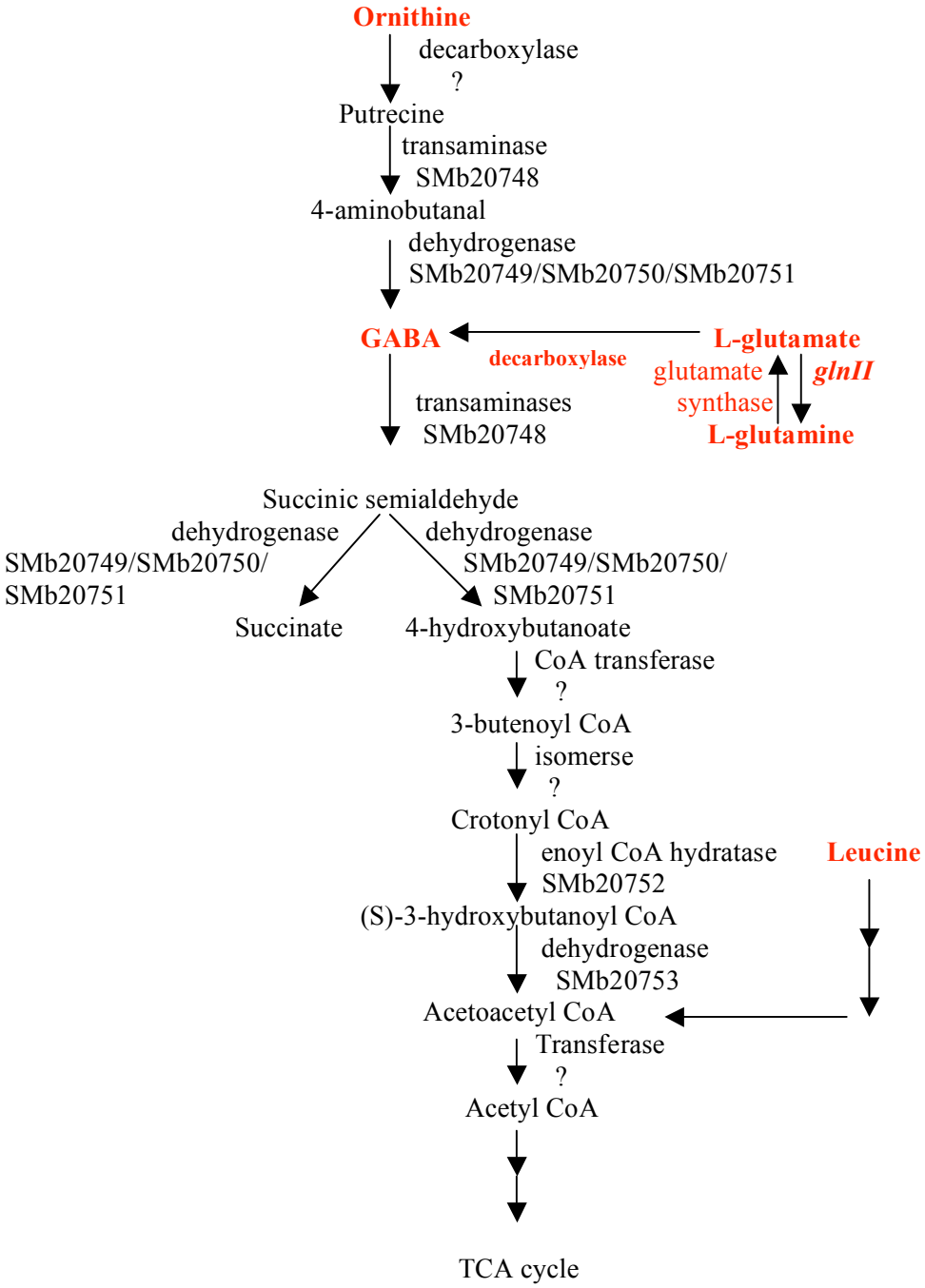

Ornithine
decarboxylase
?
Putrecine
transaminase
SMb20748
4-aminobutanal
dehydrogenase
SMb20749/SMb20750/SMb20751
GABA
decarboxylase
L-glutamate
glutamate
synthase
glnII
L-glutamine
transaminases
SMb20748
Succinic semialdehyde
dehydrogenase
SMb20749/SMb20750/
SMb20751
dehydrogenase
SMb20749/SMb20750/
SMb20751
Succinate
4-hydroxybutanoate
CoA transferase
?
3-butenoyl CoA
isomerse
?
Crotonyl CoA
enoyl CoA hydratase
SMb20752
Leucine
(S)-3-hydroxybutanoyl CoA
dehydrogenase
SMb20753
Acetoacetyl CoA
Transferase
?
Acetyl CoA
TCA cycle

Fig.3.3. Pathway for L- leucine utilization in *S. meliloti*

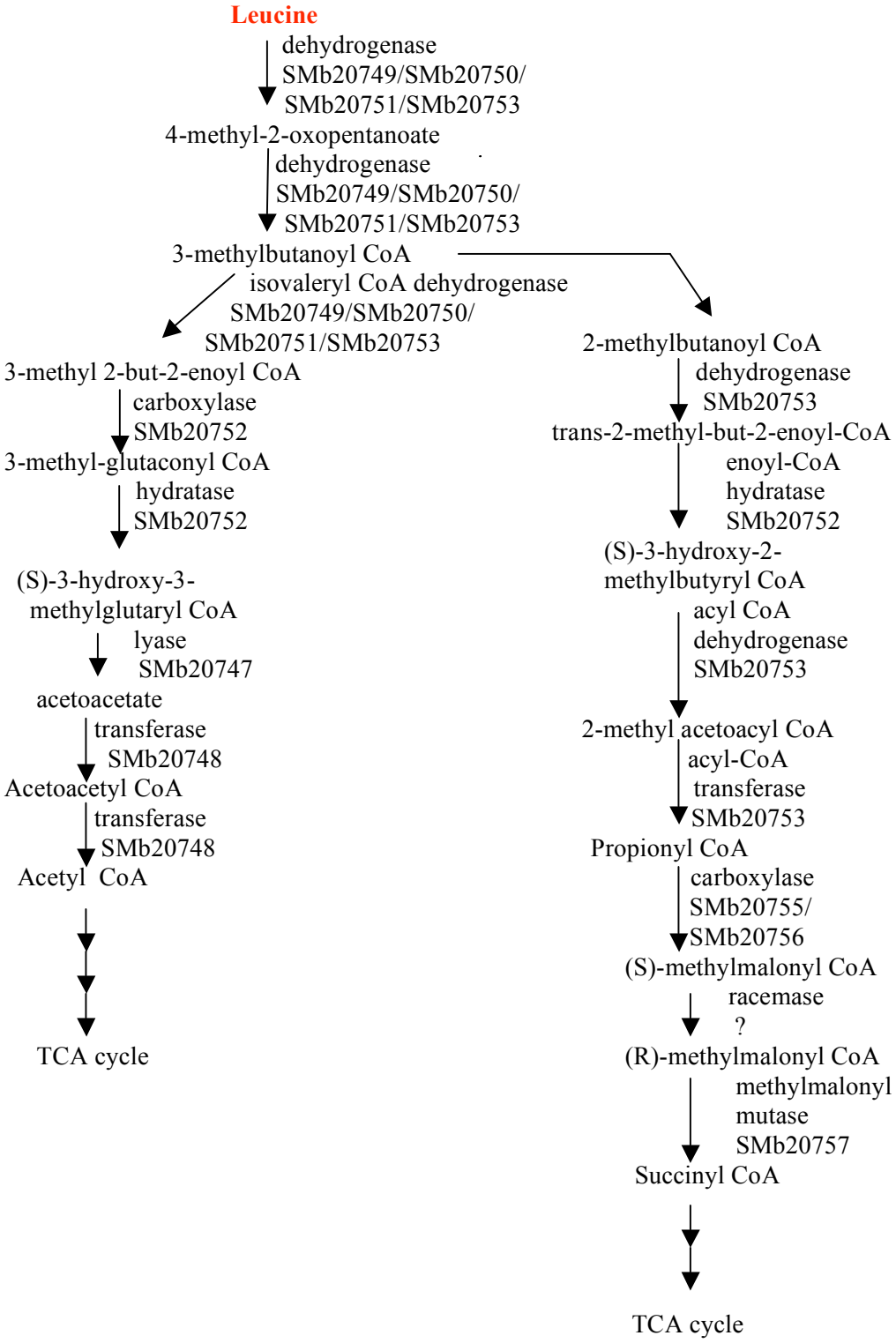
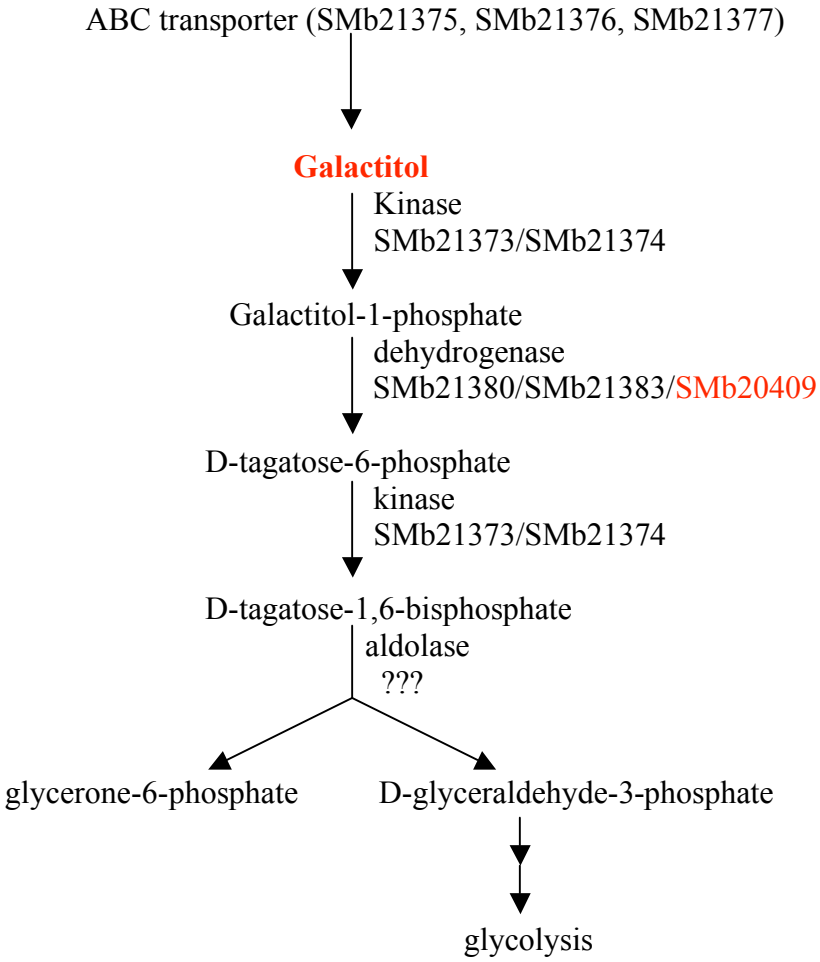
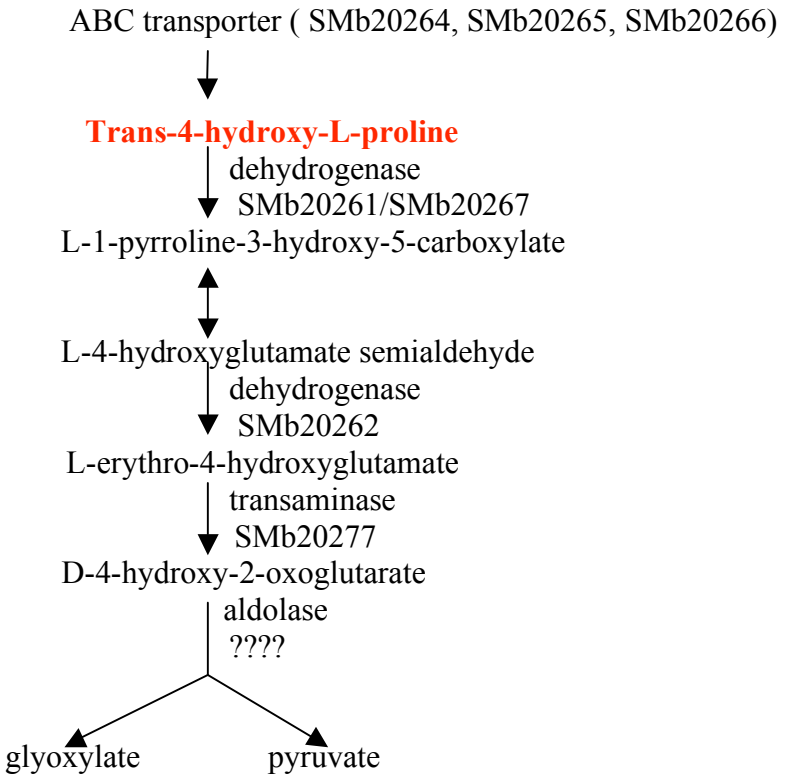

Leucine
dehydrogenase
SMb20749/SMb20750/
SMb20751/SMb20753
4-methyl-2-oxopentanoate
dehydrogenase
SMb20749/SMb20750/
SMb20751/SMb20753
3-methylbutanoyl CoA
isovaleryl CoA dehydrogenase
SMb20749/SMb20750/
SMb20751/SMb20753
3-methyl 2-but-2-enoyl CoA
carboxylase
SMb20752
3-methyl-glutaconyl CoA
hydratase
SMb20752
(S)-3-hydroxy-3-
methylglutaryl CoA
lyase
SMb20747
acetoacetate
transferase
SMb20748
Acetoacetyl CoA
transferase
SMb20748
Acetyl CoA
TCA cycle
2-methylbutanoyl CoA
dehydrogenase
SMb20753
trans-2-methyl-but-2-enoyl-CoA
enoyl-CoA
hydratase
SMb20752
(S)-3-hydroxy-2-
methylbutyryl CoA
acyl CoA
dehydrogenase
SMb20753
2-methyl acetoacyl CoA
acyl-CoA
transferase
SMb20753
Propionyl CoA
carboxylase
SMb20755/
SMb20756
(S)-methylmalonyl CoA
racemase
?
(R)-methylmalonyl CoA
methylmalonyl
mutase
SMb20757
Succinyl CoA
TCA cycle

Fig.3.4. Pathway showing galactitol utilization in *S. meliloti*



Upstream of SMb21373 is a putative lacI type transcriptional regulator (SMb21372)

Fig.3.5. Pathway delineating trans-4-hydroxy-L-proline catabolism in *S. meliloti*



Two putative transcriptional regulators: SMb20258 and SMb20259

CHAPTER 4

Analysis of SDRs in *S. meliloti*

The *Sinorhizobium meliloti* genome consists of 6204 genes (Galibert *et al.* 2002). Of these the functions of approximately 40% are unknown and this includes a family of proteins called Short Chain Dehydrogenases/Reductases (SDRs). Specific dehydrogenases are often required for the degradation and subsequent growth on specific substrates, such as carbohydrates, organic acids, amino-acids and lipids. The *S. meliloti* genome is rich in dehydrogenases/ oxidoreductases, reflecting the broad catabolic capacity of the organism and the nutritional richness of the soil and rhizosphere environments. Chapters 4 of this thesis focuses on the use of a functional genomics approach to characterize the SDR encoding genes to establish their importance to the survival of the bacterium and in symbiosis. The approach taken was to generate mutants of each of the SDR encoding gene by creating reporter gene fusions and screening these mutants for substrate utilization and symbiotic phenotype. It is hypothesized that by ascribing functions to these proteins in *S. meliloti* will facilitate better understanding of the complex plant–microbe interactions and also provide an insight into physiological and biochemical relevance of this group of proteins in other organisms including symbionts and soil borne plant or animal pathogens. Moreover, as most dehydrogenases are recognized by sequence similarities, they maybe novel and their substrate specificity is unknown. This will be of predictive value for other bacterial genome analyses

with respect to the putative substrate and/or phenotype associated with these proteins.

RESULTS

4.1 Identification of SDR-encoding genes in the *S. meliloti* genome

Given the moderately high level of sequence conservation within the SDR family, primary screening of the 6.7 Mb genome by BLASTP (Altschul *et al.* 1990) was possible. The previously well characterized SDR D-3-hydroxybutyrate dehydrogenase (BdhA) of *S. meliloti* was used as query. Eighty genes encoding predicted products with greater than 23% sequence identity to BdhA were considered as putative members of the SDR family. Further screening for the SDR- specific ADH_SHORT PROSITE consensus motif [LIVSPADNK] - x(9) - {P} - x(2) - Y - [PSTAGNCV] - [STAGNQCIVM] - [STAGC] - K - {PC} - [SAGFYR] - [LIVMSTAGD] - x - {K} - [LIVMFYW] - {D} - x - {YR} - [LIVMFYWGAPTHQ] - [GSACQRHM] (Hula et al. 2006) using PatScan identified a subset of 56 of these genes while the Pfam ADH_SHORT family (Finn *et al.* 2006) based on this same motif identified 77 genes (Tables 4-1, 4-2, 4-3). While PatScan (Dsouza *et al.* 1997) identifies a protein family based on the exact pattern of the PROSITE consensus sequence (and will not deviate from the pattern), Pfam identification uses Hidden Makov Model and is based on the probability of a residue in the consensus sequence being at a particular position (it allows for substitutions of residue). This likely accounts for the discrepancy between results obtained using the two methods.

Although SMa2019, SMc02040 and SMc02356 were not identified in the Pfam subset they were included in the study as they showed the presence of the PROSITE consensus sequence. SMc02273 and SMb20974 identified as SDR by Pfam were not included in this study.

SMc02273 appears to be a composite of proteins that contains multiple domains of which one is an SDR. These domains are usually found encoded by individual genes organized as an operon in other organisms. If SMc02273 were functional, it would appear to be a multifunctional enzyme and not exclusively an SDR. SMb20974 does not have the PROSITE consensus sequence and appears to have a truncated domain. Interestingly SMc02356 has been annotated as an ABC transporter but has the motif characteristic of an SDR. A total of 22 SDRs predicted by Pfam did not contain the PROSITE consensus sequence (Tables 4-1, 4-2, 4-3). These were all included in this study. Out of the 80 SDRs identified by BLAST, two – SMa1967 and SMb20974 were not included in the study. SMa1967 was not identified as an SDR in any of the other analyses (Pfam and PatScan). SMb20974, also identified as an SDR by Pfam was excluded from the study due to reasons discussed above. Based on these analyses 78 SDRs in the *S. meliloti* genome were identified for this study.

Protein structure studies have revealed the importance of a glycine rich T-G-X_3-G-X-G motif and the N-N-A-G motif for the maintenance of the structure of the central β-sheet involved in coenzyme binding and in determining the direction of reaction (Filling *et al.* 2002, Jörnvall *et al.* 1995). These motifs are highly conserved and are not a part of the PROSITE consensus

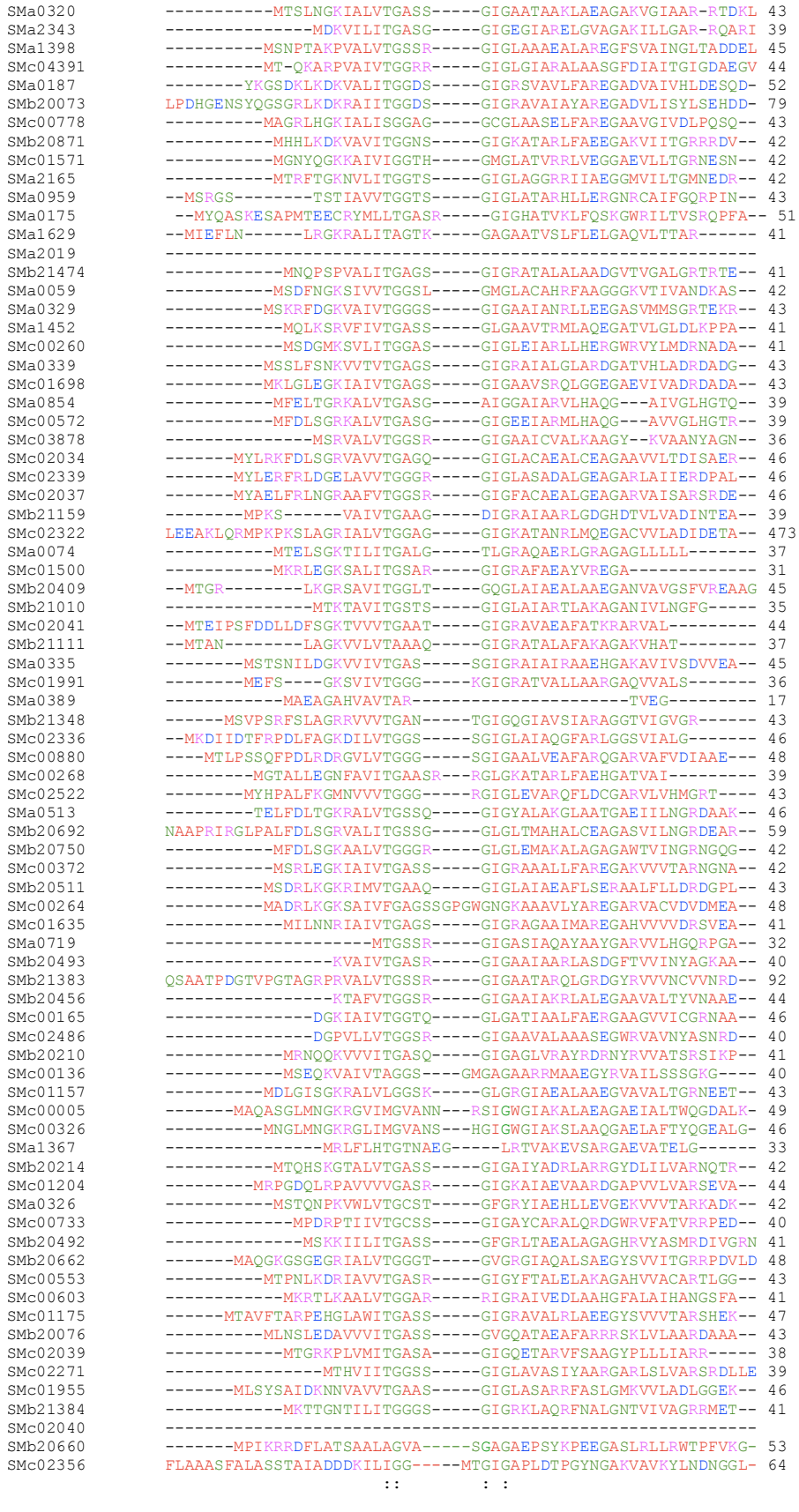
sequence. Moreover, the active site tetrad constituting the residues N-S-Y-K (part of the PROSITE consensus) is conserved in this group of proteins. The T-G-X_3-G-X-G motif was present in each predicted SDR except in SMa1367, SMa0389, SMc00005, SMc00326 and SMc02040. Also, the NNAG motif was absent in SMa2019, SMa1629, SMc00326, SMc00005 and SMc02040. The active site residue aspargine was not found in SMa2019 (Fig. 4.1.). SMa1629 was identified by Pfam and also has the PROSITE consensus motif.

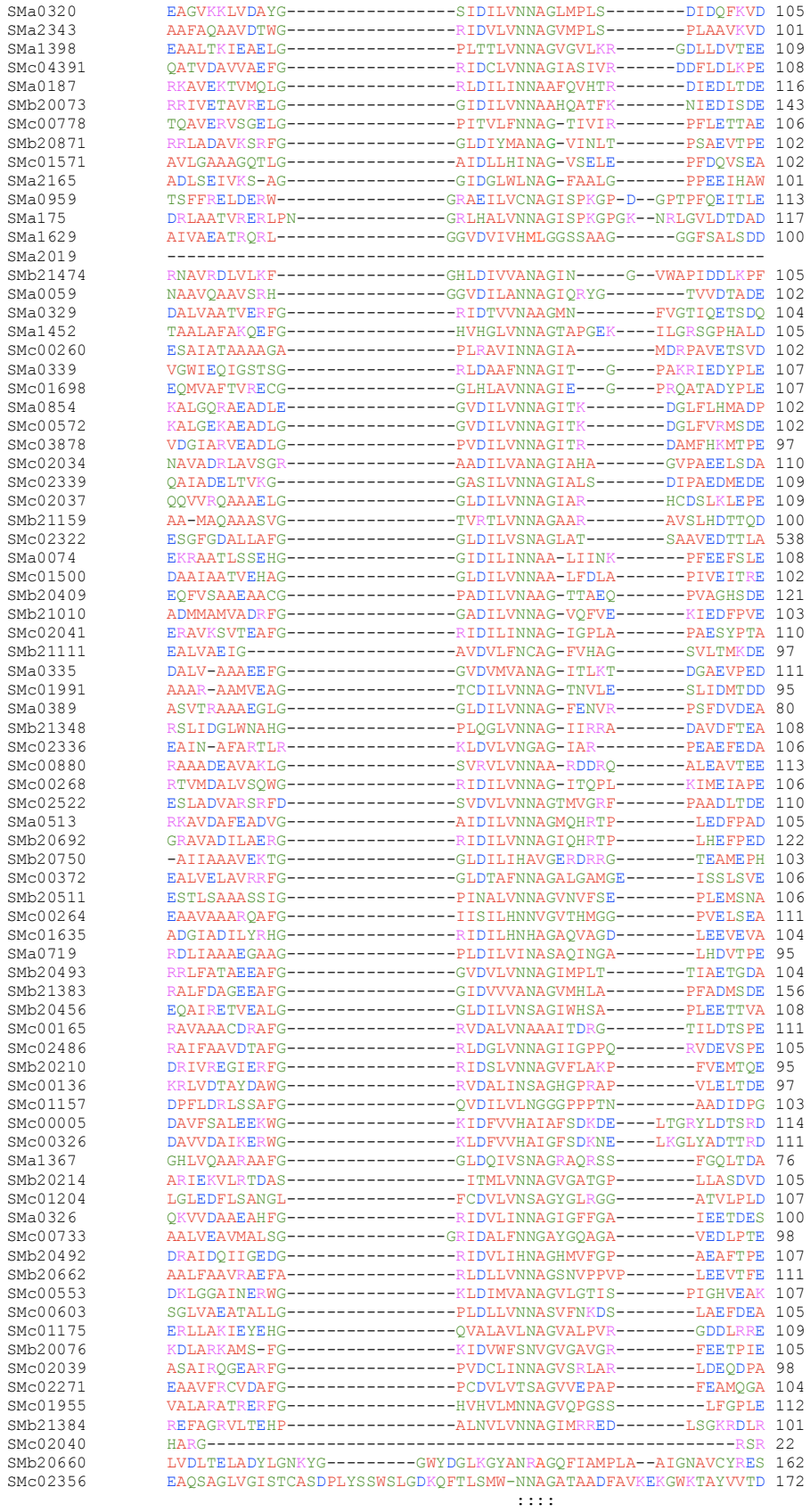
The SDRs are evenly distributed in the genome, 21 on each of the megaplasmids pSymA and pSymB, and 36 on the chromosome (Tables 4-1, 4-2, 4-3, Fig. 4.2., Fig.4.3., Fig.4.4.). Substrates for most of these SDRs have yet to be determined. The predicted proteins range from 136-699 amino acids in length. The smallest SDR identified, SMa2019, has much of the N-terminal region absent when compared to other SDRs. The longest SDR identified, SMc02322, is believed to have three domains – an aldolase domain, an SDR domain and a domain corresponding to epimerase (Finn *et al.* 2006). The orthologue of SMc02322 has been studied in *R. leguminosarum* as part of the rhamnose utilization operon (Richardson *et al.* 2004). Three of the remaining SDRs identified have been previously characterized in *S. meliloti*: *phbB*, *nodG* and *bdhA*.

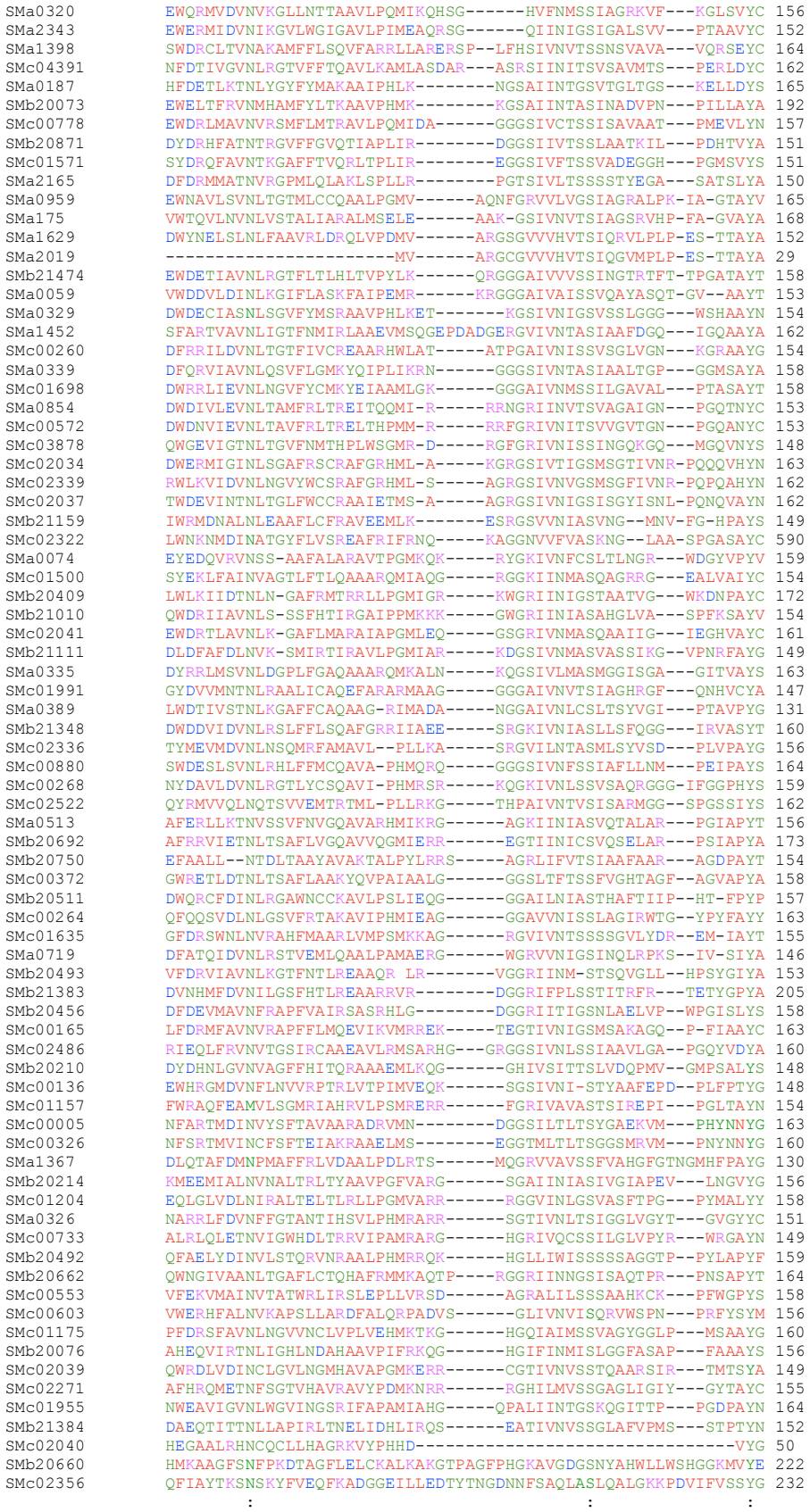
An analysis was done on other sequenced proteobacterial genomes in the Pfam database to see the distribution of this group of protein. *Bradyrhizobium japonicum*, another member of the α –proteobacteria, is reported to have 111 SDR in its genome. This organism has a free-living and symbiotic lifestyle

similar to *S. meliloti* and hence the large number of SDRs. On the other hand genomes with only a single SDR in their genome have been reported (Finn *et al.* 2006). These include the endosymbiont *Buchnera aphidicola* and human pathogen *Neorickettsia sennetsu*. Both of these organisms have very small genomes, but as most of their life cycle is within their eukaryotic host, they do not have to sequester nutrients by competing against a large micro-flora as that found in the soil ecosystem. Most of their genome would be geared towards aiding survival within the host.

Fig.4.1. Portion of the ClustalW multiple sequence alignment showing conserved residues in 78 SDRs identified in *S. meliloti*


SMa0320 ----------MTSLNGKIALVTGASS-----GIGAATAAKLAEAGAKVGIAAR-RTDKL 43
SMa2343 ----------MDKVILITGASG-----GIGEGIARELGVAGAKILLGAR-RQARI 39
SMa1398 ----------MSNPTAKPVALVTGSSR-----GIGLAAAEALAREGFSVAINGLTADDEL 45
SMc04391 ----------MT-QKARPVAIVTGGRR-----GIGLGIARALAASGFDIAITGIGDAEGV 44
SMa0187 --------YKGSDKLKDKVALITGGDS-----GIGRSVAVLFAREGADVAIVHLDESQD- 52
SMb20073 LPDHGENSYQGSGRLKDKRAIITGGDS-----GIGRAVAIAYAREGADVLISYLSEHDD- 79
SMc00778 ----------MAGRLHGKIALISGGAG-----GCGLAAASELFAREGAAVGIVDLPQSQ-- 43
SMb20871 ----------MHHLKDKVAVITGGNS-----GIGKATARLFAEEGAKVIITGRRRDV-- 42
SMc01571 ----------MGNYQGKKAIVIGGTH-----GMGLATVRRLVEGGAEVLLTGRNESN-- 42
SMa2165 ----------MTRFTGKNVLITGGTS-----GIGLAGGRRIIAEGGMVILTGMNEDR-- 42
SMa0959 --MSRGS---------TSTIAVVTGGTS-----GIGLATARHLLERGNRCAIFGQRPIN-- 43
SMa0175 --MYQASKESAPMTEECRYMLLTGASR-----GIGHATVKLFQSKGWRILTVSRQPFA-- 51
SMa1629 --MIEFLN-------LRGKRALITAGTK-----GAGAATVSLFLELGAQVLTTAR------ 41
SMa2019 ----------
SMb21474 ----------MNQPSPVALITGAGS-----GIGRATALALAADGVTVGALGRTRTE-- 41
SMa0059 ----------MSDFNGKSIVVTGGSL-----GMGLACAHRFAAGGGKVTIVANDKAS-- 42
SMa0329 ----------MSKRFDGKVAIVTGGGS-----GIGAAIANRLLEEGASVMMSGRTEKR-- 43
SMa1452 ----------MQLKSRVFIVTGASS-----GLGAAVTRMLAQEGATVLGLDLKPPA-- 41
SMc00260 ----------MSDGMKSVLITGGAS-----GIGLEIARLLHERGWRVYLMDRNADA-- 41
SMa0339 ----------MSSLFSNKVVTVTGAGS-----GIGRAIALGLARDGATVHLADRDADG-- 43
SMc01698 ----------MKLGLEGKIAIVTGAGS-----GIGAAVSRQLGGEGAEVIVADRDADA-- 43
SMa0854 ----------MFELTGRKALVTGASG-----AIGGAIARVLHAQG---AIVGLHGTQ-- 39
SMc00572 ----------MFDLSGRKALVTGASG-----GIGEEIARMLHAQG---AVVGLHGTR-- 39
SMc03878 ----------MSRVALVTGGSR-----GIGAAICVALKAAGY--KVAANYAGN-- 36
SMc02034 -------MYLRKFDLSGRVAVVTGAGQ-----GIGLACAEALCEAGAAVVLTDISAER-- 46
SMc02339 -------MYLERFRLDGELAVVTGGGR-----GIGLASADALGEAGARLAIIERDPAL-- 46
SMc02037 -------MYAELFRLNGRAAFVTGGSR-----GIGFACAEALGEAGARVAISARSRDE-- 46
SMb21159 --------MPKS------VAIVTGAAG-----DIGRAIAARLGDGHDTVLVADINTEA-- 39
SMc02322 LEEAKLQRMPKPKSLAGRIALVTGGAG-----GIGKATANRLMQEGACVVLADIDETA-- 473
SMa0074 ----------MTELSGKTILITGALG-----TLGRAQAERLGRAGAGLLLLL------- 37
SMc01500 ----------MKRLEGKSALITGSAR-----GIGRAFAEAYVREGA------------ 31
SMb20409 --MTGR--------LKGRSAVITGGLT-----GQGLAIAEALAAEGANVAVGSFVREAAG 45
SMb21010 ----------MTKTAVITGSTS-----GIGLAIARTLAKAGANIVLNGFG----- 35
SMc02041 --MTEIPSFDDLLDFSGKTVVVTGAAT-----GIGRAVAEAFATKRARVAL--------- 44
SMb21111 --MTAN--------LAGKVVLVTAAAQ-----GIGRATALAFAKAGAKVHAT-------- 37
SMa0335 --------MSTSNILDGKVVIVTGAS-----SGIGRAIAIRAAEHGAKAVIVSDVVEA-- 45
SMc01991 --------MEFS----GKSVIVTGGG-----KGIGRATVALLAARGAQVVALS------- 36
SMa0389 ----------MAEAGAHVAVTAR----------TVEG--------- 17
SMb21348 ------MSVPSRFSLAGRRVVVTGAN-----TGIGQGIAVSIARAGGTVIGVGR------ 43
SMc02336 --MKDIIDTFRPDLFAGKDILVTGGS-----SGIGLAIAQGFARLGGSVIALG------- 46
SMc00880 ----MTLPSSQFPDLRDRGVLVTGGG-----SGIGAALVEAFARQGARVAFVDIAAE--- 48
SMc00268 ---------MGTALLEGNFAVITGAASR---RGLGKATARLFAEHGATVAI--------- 39
SMc02522 --------MYHPALFKGMNVVVTGGG-----RGIGLEVARQFLDCGARVLVHMGRT---- 43
SMa0513 ---------TELFDLTGKRALVTGSSQ-----GIGYALAKGLAATGAEIILNGRDAAK-- 46
SMb20692 NAAPRIRGLPALFDLSGRVALITGSSG-----GLGLTMAHALCEAGASVILNGRDEAR-- 59
SMb20750 ----------MFDLSGKAALVTGGGR-----GLGLEMAKALAGAGAWTVINGRNGQG-- 42
SMc00372 ----------MSRLEGKIAIVTGASS-----GIGRAAALLFAREGAKVVVTARNGNA-- 42
SMb20511 ----------MSDRLKGKRIMVTGAAQ-----GIGLAIAEAFLSERAALFLLDRDGPL-- 43
SMc00264 ----------MADRLKGKSAIVFGAGSSGPGWGNGKAAAVLYAREGARVACVDVDMEA-- 48
SMc01635 ----------MILNNRIAIVTGAGS-----GIGRAGAAIMAREGAHVVVVDRSVEA-- 41
SMa0719 --------------------MTGSSR-----GIGASIAQAYAAYGARVVLHGQRPGA-- 32
SMb20493 -----------------KVAIVTGASR-----GIGAAIAARLASDGFTVVINYAGKAA-- 40
SMb21383 QSAATPDGTVPGTAGRPRVALVTGSSR-----GIGAATARQLGRDGYRVVVNCVVNRD-- 92
SMb20456 -----------------KTAFVTGGSR-----GIGAAIAKRLALEGAAVALTYVNAAE-- 44
SMc00165 ---------------DGKIAIVTGGTQ-----GLGATIAALFAERGAAGVVICGRNAA-- 46
SMc02486 ---------------DGPVLLVTGGSR-----GIGAAVALAAASEGWRVAVNYASNRD-- 40
SMb20210 ----------MRNQQKVVVITGASQ-----GIGAGLVRAYRDRNYRVVATSRSIKP-- 41
SMc00136 ----------MSEQKVAIVTAGGS----GMGAGAARRMAAEGYRVAILSSSGKG---- 40
SMc01157 ----------MDLGISGKRALVLGGSK-----GLGRGIAEALAAEGVAVALTGRNEET-- 43
SMc00005 -------MAQASGLMNGKRGVIMGVANN---RSIGWGIAKALAEAGAEIALTWQGDALK- 49
SMc00326 ----------MNGLMNGKRGLIMGVANS---HGIGWGIAKSLAAQGAELAFTYQGEALG- 46
SMa1367 -----------------MRLFLHTGTNAEG-----LRTVAKEVSARGAEVATELG------ 33
SMb20214 ----------MTQHSKGTALVTGASS-----GIGAIYADRLARRGYDLILVARNQTR-- 42
SMc01204 ---------MRPGDQLRPAVVVVGASR-----GIGKAIAEVAARDGAPVVLVARSEVA-- 44
SMa0326 ----------MSTQNPKVWLVTGCST-----GFGRYIAEHLLEVGEKVVVTARKADK-- 42
SMc00733 -------------MPDRPTIIVTGCSS-----GIGAYCARALQRDGWRVFATVRRPED-- 40
SMb20492 --------------MSKKIILITGASS-----GFGRLTAEALAGAGHRVYASMRDIVGRN 41
SMb20662 -------MAQGKGSGEGRIALVTGGGT-----GVGRGIAQALSAEGYSVVITGRRPDVLD 48
SMc00553 ----------MTPNLKDRIAVVTGASR-----GIGYFTALELAKAGAHVVACARTLGG-- 43
SMc00603 ----------MKRTLKAALVTGGAR-----RIGRAIVEDLAAHGFALAIHANGSFA-- 41
SMc01175 ------MTAVFTARPEHGLAWITGASS-----GIGRAVALRLAEEGYSVVVTARSHEK-- 47
SMb20076 ----------MLNSLEDAVVVITGASS-----GVGQATAEAFARRRSKLVLAARDAAA-- 43
SMc02039 ----------MTGRKPLVMITGASA-----GIGQETARVFSAAGYPLLLIARR----- 38
SMc02271 ----------------MTHVIITGGSS-----GIGLAVASIYAARGARLSLVARSRDLLE 39
SMc01955 -------MLSYSAIDKNNVAVVTGAAS-----GIGLASARRFASLGMKVVLADLGGEK-- 46
SMb21384 ----------MKTTGNTILITGGGS-----GIGRKLAQRFNALGNTVIVAGRRMET-- 41
SMc02040 ----------
SMb20660 -------MPIKRRDFLATSAALAGVA-----SGAGAEPSYKPEEGASLRLLRWTPFVKG- 53
SMc02356 FLAAASFALASSTAIADDDKILIGG----MTGIGAPLDTPGYNGAKVAVKYLNDNGGL- 64
:: : :


SMa0320 EAGVKKLVDAYG----------SIDILVNNAGLMPLS--------DIDQFKVD 105
SMa2343 AAFAQAAVDTWG----------RIDVLVNNAGVMPLS--------PLAAVKVD 101
SMa1398 EAALTKIEAELG----------PLTTLVNNAGVGVLKR------GDLLDVTEE 109
SMc04391 QATVDAVVAEFG----------RIDCLVNNAGIASIVR------DDFLDLKPE 108
SMa0187 RKAVEKTVMQLG----------RLDILINNAAFQVHTR-------DIEDLTDE 116
SMb20073 RRIVETAVRELG----------GIDILVNNAAHQATFK-------NIEDISDE 143
SMc00778 TQAVERVSGELG----------PITVLFNNAG-TIVIR-------PFLETTAE 106
SMb20871 RRLADAVKSRFG----------GLDIYMANAG-VINLT-------PSAEVTPE 102
SMc01571 AVLGAAAGQTLG----------AIDLLHINAG-VSELE-------PFDQVSEA 102
SMa2165 ADLSEIVKS-AG----------GIDGLWLNAG-FAALG-------PPEEIHAW 101
SMa0959 TSFFRELDERW----------GRAEILVCNAGISPKGP-D--GPTPFQEITLE 113
SMa175 DRLAATVRERLPN----------GRLHALVNNAGISPKGPGK--NRLGVLDTDAD 117
SMa1629 AIVAEATRQRL----------GGVDVIVHMLGGSSAAG------GGFSALSDD 100
SMa2019 ----------
SMb21474 RNAVRDLVLKF----------GHLDIVVANAGIN-----G--VWAPIDDLKPF 105
SMa0059 NAAVQAAVSRH----------GGVDILANNAGIQRYG--------TVVDTADE 102
SMa0329 DALVAATVERFG----------RIDTVVNAAGMN--------FVGTIQETSDQ 104
SMa1452 TAALAFAKQEFG----------HVHGLVNNAGTAPGEK----ILGRSGPHALD 105
SMc00260 ESAIATAAAAGA----------PLRAVINNAGIA--------MDRPAVETSVD 102
SMa0339 VGWIEQIGSTSG----------RLDAAFNNAGIT---G----PAKRIEDYPLE 107
SMc01698 EQMVAFTVRECG----------GLHLAVNNAGIE---G----PRQATADYPLE 107
SMa0854 KALGQRAEADLE----------GVDILVNNAGITK--------DGLFLHMADP 102
SMc00572 KALGEKAEADLG----------GVDILVNNAGITK--------DGLFVRMSDE 102
SMc03878 VDGIARVEADLG----------PVDILVNNAGITR--------DAMFHKMTPE 97
SMc02034 NAVADRLAVSGR----------AADILVANAGIAHA-------GVPAEELSDA 110
SMc02339 QAIADELTVKG----------GASILVNNAGIALS-------DIPAEDMEDE 109
SMc02037 QQVVRQAAAELG----------GLDILVNNAGIAR--------HCDSLKLEPE 109
SMb21159 AA-MAQAAASVG----------TVRTLVNNAGAAR--------AVSLHDTTQD 100
SMc02322 ESGFGDALLAFG----------GLDILVSNAGLAT--------SAAVEDTTLA 538
SMa0074 EKRAATLSSEHG----------GIDILINNAA-LIINK-------PFEEFSLE 108
SMc01500 DAAIAATVEHAG----------GLDILVNNAA-LFDLA-------PIVEITRE 102
SMb20409 EQFVSAAEAACG----------PADILVNAAG-TTAEQ-------PVAGHSDE 121
SMb21010 ADMMAMVADRFG----------GADILVNNAG-VQFVE-------KIEDFPVE 103
SMc02041 ERAVKSVTEAFG----------RIDILINNAG-IGPLA-------PAESYPTA 110
SMb21111 EALVAEIG----------AVDVLFNCAG-FVHAG-------SVLTMKDE 97
SMa0335 DALV-AAAEEFG----------GVDVMVANAG-ITLKT-------DGAEVPED 111
SMc01991 AAAR-AAMVEAG----------TCDILVNNAG-TNVLE-------SLIDMTDD 95
SMa0389 ASVTRAAAEGLG----------GLDILVNNAG-FENVR-------PSFDVDEA 80
SMb21348 RSLIDGLWNAHG----------PLQGLVNNAG-IIRRA-------DAVDFTEA 108
SMc02336 EAIN-AFARTLR----------KLDVLVNGAG-IAR---------PEAEFEDA 106
SMc00880 RAAADEAVAKLG----------SVRVLVNNAA-RDDRQ-------ALEAVTEE 113
SMc00268 RTVMDALVSQWG----------RIDILVNNAG-ITQPL-------KIMEIAPE 106
SMc02522 ESLADVARSRFD----------SVDVLVNNAGTMVGRF-------PAADLTDE 110
SMa0513 RKAVDAFEADVG----------AIDILVNNAGMQHRTP--------LEDFPAD 105
SMb20692 GRAVADILAERG----------RIDILVNNAGIQHRTP--------LHEFPED 122
SMb20750 -AIIAAAVEKTG----------GLDILIHAVGERDRRG--------TEAMEPH 103
SMc00372 EALVELAVRRFG----------GLDTAFNNAGALGAMGE-------ISSLSVE 106
SMb20511 ESTLSAAASSIG----------PINALVNNAGVNVFSE--------PLEMSNA 106
SMc00264 EAAVAAARQAFG----------IISILHNNVGVTHMGG--------PVELSEA 111
SMc01635 ADGIADILYRHG----------RIDILHNHAGAQVAGD--------LEEVEVA 104
SMa0719 RDLIAAAEGAAG----------PLDILVINASAQINGA--------LHDVTPE 95
SMb20493 RRLFATAEEAFG----------GVDVLVNNAGIMPLT--------TIAETGDA 104
SMb21383 RALFDAGEEAFG----------GIDVVVANAGVMHLA--------PFADMSDE 156
SMb20456 EQAIRETVEALG----------GLDILVNSAGIWHSA--------PLEETTVA 108
SMc00165 RAVAAACDRAFG----------RVDALVNAAAITDRG--------TILDTSPE 111
SMc02486 RAIFAAVDTAFG----------RLDGLVNNAGIIGPPQ-------RVDEVSPE 105
SMb20210 DRIVREGIERFG----------RIDSLVNNAGVFLAKP--------FVEMTQE 95
SMc00136 KRLVDTAYDAWG----------RVDALINSAGHGPRAP--------VLELTDE 97
SMc01157 DPFLDRLSSAFG----------QVDILVLNGGGPPPTN--------AADIDPG 103
SMc00005 DAVFSALEEKWG----------KIDFVVHAIAFSDKDE----LTGRYLDTSRD 114
SMc00326 DAVVDAIKERWG----------KLDFVVHAIGFSDKNE----LKGLYADTTRD 111
SMa1367 GHLVQAARAAFG----------GLDQIVSNAGRAQRSS--------FGQLTDA 76
SMb20214 ARIEKVLRTDAS----------ITMLVNNAGVGATGP--------LLASDVD 105
SMc01204 LGLEDFLSANGL----------FCDVLVNSAGYGLRGG--------ATVLPLD 107
SMa0326 QKVVDAAEAHFG----------RIDVLINNAGIGFFGA--------IEETDES 100
SMc00733 AALVEAVMALSG----------GRIDALFNNGAYGQAGA--------VEDLPTE 98
SMb20492 DRAIDQIIGEDG----------RIDVLIHNAGHMVFGP--------AEAFTPE 107
SMb20662 AALFAAVRAEFA----------RLDLLVNNAGSNVPPVP-------LEEVTFE 111
SMc00553 DKLGGAINERWG----------KLDIMVANAGVLGTIS-------PIGHVEAK 107
SMc00603 SGLVAEATALLG----------PLDLLVNNASVFNKDS--------LAEFDEA 105
SMc01175 ERLLAKIEYEHG----------QVALAVLNAGVALPVR--------GDDLRRE 109
SMb20076 KDLARKAMS-FG----------KIDVWFSNVGVGAVGR--------FEETPIE 105
SMc02039 ASAIRQGEARFG----------PVDCLINNAGVSRLAR--------LDEQDPA 98
SMc02271 EAAVFRCVDAFG----------PCDVLVTSAGVVEPAP--------FEAMQGA 104
SMc01955 VALARATRERFG----------HVHVLMNNAGVQPGSS---------LFGPLE 112
SMb21384 REFAGRVLTEHP----------ALNVLVNNAGIMRRED-------LSGKRDLR 101
SMc02040 HARG----------RSR 22
SMb20660 LVDLTELADYLGNKYG---------GWYDGLKGYANRAGQFIAMPLA--AIGNAVCYRES 162
SMc02356 EAQSAGLVGISTCASDPLYSSWSLGDKQFTLSMW-NNAGATAADFAVKEKGWKTAYVVTD 172
:::::


SMa0320 EWQRMVDVNVKGLLNTTAAVLPQMIKQHSG------HVFNMSSIAGRKVF---KGLSVYC 156
SMa2343 EWERMIDVNIKGVLWGIGAVLPIMEAQRSG------QIINIGSIGALSVV---PTAAVYC 152
SMa1398 SWDRCLTVNAKAMFFLSQVFARRLLARERSP--LFHSIVNVTSSNSVAVA---VQRSEYC 164
SMc04391 NFDTIVGVNLRGTVFFTQAVLKAMLASDAR---ASRSIINITSVSAVMTS---PERLDYC 162
SMa0187 HFDETLKTNLYGYFYMAKAAIPHLK--------NGSAIINTGSVTGLTGS---KELLDYS 165
SMb20073 EWELTFRVNMHAMFYLTKAAVPHMK--------KGSAIINTASINADVPN---PILLAYA 192
SMc00778 EWDRLMAVNVRSMFLMTRAVLPQMIDA------GGGSIVCTSSISAVAAT---PMEVLYN 157
SMb20871 DYDRHFATNTRGVFFGVQTIAPLIR--------DGGSIIVTSSLAATKIL---PDHTVYA 151
SMc01571 SYDRQFAVNTKGAFFTVQRLTPLIR--------EGGSIVFTSSVADEGGH---PGMSVYS 151
SMa2165 DFDRMMATNVRGPMLQLAKLSPLLR--------PGTSIVLTSSSSTYEGA---SATSLYA 150
SMa0959 EWNAVLSVNLTGTMLCCQAALPGMV------AQNFGRVVLVGSIAGRALPK-IA-GTAYV 165
SMa175 VWTQVLNVNLVSTALIARALMSELE------AAK-GSIVNVTSIAGSRVHP-FA-GVAYA 168
SMa1629 DWYNELSLNLFAAVRLDRQLVPDMV------ARGSGVVVHVTSIQRVLPLP-ES-TTAYA 152
SMa2019 ----------MV------ARGCGVVVHVTSIQGVMPLP-ES-TTAYA 29
SMb21474 EWDETIAVNLRGTFLTLHLTVPYLK------QRGGGAIVVVSSINGTRTFT-TPGATAYT 158
SMa0059 VWDDVLDINLKGIFLASKFAIPEMR------KRGGGAIVAISSVQAYASQT-GV--AAYT 153
SMa0329 DWDECIASNLSGVFYMSRAAVPHLKET-------KGSIVNIGSVSSLGGG---WSHAAYN 154
SMa1452 SFARTVAVNLIGTFNMIRLAAEVMSQGEPDADGERGVIVNTASIAAFDGQ---IGQAAYA 162
SMc00260 DFRRILDVNLTGTFIVCREAARHWLAT-----ATPGAIVNISSVSGLVGN---KGRAAYG 154
SMa0339 DFQRVIAVNLQSVFLGMKYQIPLIKRN------GGGSIVNTASIAALTGP---GGMSAYA 158
SMc01698 DWRRLIEVNLNGVFYCMKYEIAAMLGK------GGGAIVNMSSILGAVAL---PTASAYT 158
SMa0854 DWDIVLEVNLTAMFRLTREITQQMI-R-----RRNGRIINVTSVAGAIGN---PGQTNYC 153
SMc00572 DWDNVIEVNLTAVFRLTRELTHPMM-R-----RRFGRIVNITSVVGVTGN---PGQANYC 153
SMc03878 QWGEVIGTNLTGVFNMTHPLWSGMR-D-----RGFGRIVNISSINGQKGQ---MGQVNYS 148
SMc02034 DWERMIGINLSGAFRSCRAFGRHML-A-----KGRGSIVTIGSMSGTIVNR-PQQQVHYN 163
SMc02339 RWLKVIDVNLNGVYWCSRAFGRHML-S-----AGRGSIVNVGSMSGFIVNR-PQPQAHYN 162
SMc02037 TWDEVINTNLTGLFWCCRAAIETMS-A-----AGRGSIVNIGSISGYISNL-PQNQVAYN 162
SMb21159 IWRMDNALNLEAAFLCFRAVEEMLK-------ESRGSVVNIASVNG--MNV-FG-HPAYS 149
SMc02322 LWNKNMDINATGYFLVSREAFRIFRNQ------KAGGNVVFVASKNG--LAA-SPGASAYC 590
SMa0074 EYEDQVRVNSS-AAFALARAVTPGMKQK-----RYGKIVNFCSLTLNGR---WDGYVPYV 159
SMc01500 SYEKLFAINVAGTLFTLQAAARQMIAQG-----RGGKIINMASQAGRRG---EALVAIYC 154
SMb20409 LWLKIIDTNLN-GAFRMTRRLLPGMIGR-----KWGRIINIGSTAATVG---WKDNPAYC 172
SMb21010 QWDRIIAVNLS-SSFHTIRGAIPPMKKK-----GWGRIINIASAHGLVA---SPFKSAYV 154
SMc02041 EWDRTLAVNLK-GAFLMARAIAPGMLEQ-----GSGRIVNMASQAAIIG---IEGHVAYC 161
SMb21111 DLDFAFDLNVK-SMIRTIRAVLPGMIAR-----KDGSIVNMASVASSIKG--VPNRFAYG 149
SMa0335 DYRRLMSVNLDGPLFGAQAAARQMKALN-----KQGSIVLMASMGGISGA---GITVAYS 163
SMc01991 GYDVVMNTNLRAALICAQEFARARMAAG-----GGGAIVNVTSIAGHRGF---QNHVCYA 147
SMa0389 LWDTIVSTNLKGAFFCAQAAG-RIMADA-----NGGAIVNLCSLTSYVGI---PTAVPYG 131
SMb21348 DWDDVIDVNLRSLFFLSQAFGRRIIAEE-----SRGKIVNIASLLSFQGG---IRVASYT 160
SMc02336 TYMEVMDVNLNSQMRFAMAVL--PLLKA-----SRGVILNTASMLSYVSD---PLVPAYG 156
SMc00880 SWDESLSVNLRHLFFMCQAVA-PHMQRQ-----GGGSIVNFSSIAFLLNM---PEIPAYS 164
SMc00268 NYDAVLDVNLRGTLYCSQAVI-PHMRSR-----KQGKIVNLSSVSAQRGGG-IFGGPHYS 159
SMc02522 QYRMVVQLNQTSVVEMTRTML-PLLRKG-----THPAIVNTVSISARMGG--SPGSSIYS 162
SMa0513 AFERLLKTNVSSVFNVGQAVARHMIKRG------AGKIINIASVQTALAR---PGIAPYT 156
SMb20692 AFRRVIETNLTSAFLVGQAVVQGMIERR------EGTIINICSVQSELAR---PSIAPYA 173
SMb20750 EFAALL--NTDLTAAYAVAKTALPYLRRS-----AGRLIFVTSIAAFAAR---AGDPAYT 154
SMc00372 GWRETLDTNLTSAFLAAKYQVPAIAALG------GGSLTFTSSFVGHTAGF--AGVAPYA 158
SMb20511 DWQRCFDINLRGAWNCCKAVLPSLIEQG------GGAILNIASTHAFTIIP--HT-FPYP 157
SMc00264 QFQQSVDLNLGSVFRTAKAVIPHMIEAG------GGAVVNISSLAGIRWTG--YPYFAYY 163
SMc01635 GFDRSWNLNVRAHFMAARLVMPSMKKAG------RGVIVNTSSSSGVLYDR--EM-IAYT 155
SMa0719 DFATQIDVNLRSTVEMLQAALPAMAERG------WGRVVNIGSINQLRPKS--IV-SIYA 146
SMb20493 VFDRVIAVNLKGTFNTLREAAQR LR-------VGGRIINM-STSQVGLL--HPSYGIYA 153
SMb21383 DVNHMFDVNILGSFHTLREAARRVR--------DGGRIFPLSSTITRFR---TETYGPYA 205
SMb20456 DFDEVMAVNFRAPFVAIRSASRHLG--------DGGRIITIGSNLAELVP--WPGISLYS 158
SMc00165 LFDRMFAVNVRAPFFLMQEVIKVMRREK-----TEGTIVNIGSMSAKAGQ--P-FIAAYC 163
SMc02486 RIEQLFRVNVTGSIRCAAEAVLRMSARHG---GRGGSIVNLSSIAAVLGA--PGQYVDYA 160
SMb20210 DYDHNLGVNVAGFFHITQRAAAEMLKQG------GHIVSITTSLVDQPMV--GMPSALYS 148
SMc00136 EWHRGMDVNFLNVVRPTRLVTPIMVEQK------SGSIVNI-STYAAFEPD--PLFPTYG 148
SMc01157 FWRAQFEAMVLSGMRIAHRVLPSMRERR------FGRIVAVASTSIREPI---PGLTAYN 154
SMc00005 NFARTMDINVYSFTAVAARADRVMN--------DGGSILTLTSYGAEKVM---PHYNNYG 163
SMc00326 NFSRTMVINCFSFTEIAKRAAELMS--------EGGTMLTLTSGGSMRVM---PNYNNYG 160
SMa1367 DLQTAFDMNPMAFFRLVDAALPDLRTS------MQGRVVAVSSFVAHGFGTNGMHFPAYG 130
SMb20214 KMEEMIALNVNALTRLTYAAVPGFVARG------SGAIINIASIVGIAPEV---LNGVYG 156
SMc01204 EQLGLVDLNIRALTELTLRLLPGMVARR------RGGVINLGSVASFTPG---PYMALYY 158
SMa0326 NARRLFDVNFFGTANTIHSVLPHMRARR------SGTIVNLTSIGGLVGYT---GVGYYC 151
SMc00733 ALRLQLETNVIGWHDLTRRVIPAMRARG------HGRIVQCSSILGLVPYR---WRGAYN 149
SMb20492 QFAELYDINVLSTQRVNRAALPHMRRQK------HGLLIWISSSSSAGGTP--PYLAPYF 159
SMb20662 QWNGIVAANLTGAFLCTQHAFRMMKAQTP----RGGRIINNGSISAQTPR---PNSAPYT 164
SMc00553 VFEKVMAINVTATWRLIRSLEPLLVRSD------AGRALILSSSAAHKCK---PFWGPYS 158
SMc00603 VWERHFALNVKAPSLLARDFALQRPADVS------GLIVNVISQRVWSPN---PRFYSYM 156
SMc01175 PFDRSFAVNLNGVVNCLVPLVEHMKTKG------HGQIAIMSSVAGYGGLP---MSAAYG 160
SMb20076 AHEQVIRTNLIGHLNDAHAAVPIFRKQG------HGIFINMISLGGFASAP---FAAAYS 156
SMc02039 QWRDLVDINCLGVLNGMHAVAPGMKERR------CGTIVNVSSTQAARSIR---TMTSYA 149
SMc02271 AFHRQMETNFSGTVHAVRAVYPDMKNRR------RGHILMVSSGAGLIGIY---GYTAYC 155
SMc01955 NWEAVIGVNLWGVINGSRIFAPAMIAHG-----QPALIINTGSKQGITTP---PGDPAYN 164
SMb21384 DAEQTITTNLLAPIRLTNELIDHLIRQS------EATIVNVSSGLAFVPMS---STPTYN 152
SMc02040 HEGAALRHNCQCLLHAGRKVYPHHD----------VYG 50
SMb20660 HMKAAGFSNFPKDTAGFLELCKALKAKGTPAGFPHGKAVGDGSNYAHWLLWSHGGKMVYE 222
SMc02356 QFIAYTKSNSKYFVEQFKADGGEILLEDTYTNGDNNFSAQLASLQALGKKPDVIFVSSYG 232

```
SMa0320         ATKHAVTAFSDGLRMEVGQ----------KHGIRVT------CIQPGAVATELYDHITD 199
SMa2343         ATKFAVRAISDGLRQES-------------TNIRVT------CVNPGVVESELAGTITH 192
SMa1398         ASKAAASMVSKALAVRLG------------RENVAVY------DVQPGLIATEMTAPVID 206
SMc04391        MSKAGLAAFSQGLALRLA------------ETGIAVF------EVRPGIIRSDMTAAVSG 204
SMa0187         MTKGGIHAFTRALSGHLV------------PKGIRVN------AVAPGPVWTPLNPSDKE 207
SMb20073        TTKGAIHNFSAGLAQMLA------------ERGIRVN------VVAPGPIWTPLIPSTMP 234
SMc00778        TTKGACHMFARSVAVEFR------------DRNIRCN------AVCPGFIQTAHGQREVE 199
SMb20871        GSKAATAAFAKNWALEFK------------SRRIRVN------ILSPGPVETEILAKLGV 193
SMc01571        ASKAALVSFASVLAAELL------------PRGIRVN------SVSPGFIDTPTKGVAGI 193
SMa2165         ATKGAVLAMSRSWASAMA------------PRGIRVN------VLVPGPIESNLRSFL-- 190
SMa0959         ASKAALAGFARSLIARYA------------GQGITVN------VVAPGRIATEMAGPRDS 207
SMa175          ASKAALASLTREMAHEFG------------QRGVRAN------AIAPGEIETSILSPG-- 208
SMa1629         AAKAALSTYSKAMSKEVS------------PKGVRVV------RVSPGWIETEASVRLAE 194
SMa2019         AAKAALSTYGKSIAKEIS------------SKGIRVV------RVSPGWIATEASVRLAE 71
SMb21474        ATKAAQVAIVQQLALELG------------KHHIRVN------AVCPGAIETNISDNTKL 200
SMa0059         ASKGAINALVRAMALDHA------------GDNITVN------AVCPASIDTPMLRWAAD 195
SMa0329         AAKGGVANLTRSAACDLG------------KFGVRAN------TVAPGLTVTG-------- 189
SMa1452         ASKGGVAALTLPAARELA------------RFGIRVV------TIAPGIFDTP-------- 197
SMc00260        ASKGAVNLLTYILATELG------------QDGIRVN------AIAPGAIDTP-------- 189
SMa0339         ASKHGVQGLTRVVAMENA------------AHGIRVN------AIAPGWTETP-------- 193
SMc01698        AAKHGVVGLTKAAGIEYA------------RMGIRIN------AVGPGWIETP-------- 193
SMa0854         ASKAGMIGFSKSLAQEIA------------TRNITVN------CVAPGFIES-------- 187
SMc00572        ASKAGMIGFSKSLAQEIA------------TRNVTVN------CVAPGFIES-------- 187
SMc03878        AAKAGDLGLTKALAQEGA------------AKGITVN------AICPGYIGT-------- 182
SMc02034        AAKAGVHHLTRSLAAEWA------------ARGVRVN------SVAPTYIDT-------- 197
SMc02339        ASKAAVHHLTKSLAAEWA------------PRGVRVN------AVAPTYIET-------- 196
SMc02037        ASKAGVHMLTKSLAGEFA------------KSNIRIN------AVAPGYIET-------- 196
SMb21159        AAKAGLLHLTRLIAVEYG------------KFGIRAN------AVAPATVR-------TQ 184
SMc02322        TAKAAEIHLARCLALEGA------------SAQIRVN------VVNPDAVLRGSKIWTGE 632
SMa0074         ASKGAMLGLTKALARELG------------PHGVRVN------AVSPGAVVSEAEERVFA 201
SMc01500        ATKAAVISLTQSAGLDLI------------KHRINVN------AIAPGVVDGEHWDGVDA 196
SMb20409        ASKSGLLGLTRCVALEGA------------PHGVTCV------MISPTWVETELMRRNVQ 214
SMb21010        AAKHGIMGLTKTVALEVA------------ESGVTVN------SICPGYVLTPLVEKQIP 196
SMc02041        ASKAGIIGMTNCMALEWG------------PRGVTVN------AVSPTVVETELG---LT 200
SMb21111        VTKAAVIGLTKAVAADYV------------GDGIRCN------AICPGTVESPSLESRMR 191
SMa0335         TSKGGVVLMAKSLADALG------------PDGIRVN------AVAPGTIDTELLRTSPG 205
SMc01991        ASKAGLEGATRVMAKELG------------PYGIRVV------AIAPTITMTELAAAAWA 189
SMa0389         ASKSGLLGVTRALATEWA------------AHNIRVN------AIAPGYFRTAMTAGFYE 173
SMb21348        ASKHGALGITRLLACEWA------------AKGINVN------AVAPGYVVTNNTEALRA 202
SMc02336        ASKSGVLGLTRHLAHAFG------------PDGIRVN------AISPGYHKTDMTKGLWT 198
SMc00880        TAKAGIIGLTKSLAGKLG------------PDNIRVN------AILPGMIVTERQRRLWL 206
SMc00268        AAKAGVLGLTKAMARELA------------PDNVRVN------AICPGFIATDITGGLLT 201
SMc02522        ATKAFVATYSKALARELA------------PDGIRVN------CVSPGTITTDFHERYST 204
SMa0513         ATKGAVGNLTKGMATDWA------------RYGLQCN------AIAPGYFDTPLNAALVA 198
SMb20692        ASKGGLKMLTKGMALDWG------------RYGIRVN------GLAPGYFKTELNSALVS 215
SMb20750        AAKGGLAALTRSLAVELG------------ADNLTVN------AIAPGWFATETN-AHLA 195
SMc00372        ASKAGLIGLVQALAVELG------------ARGIRVN------ALLPGGTDTPANFANLP 200
SMb20511        VAKHALIGMTKALGIEYA------------SKGVRVN------ALAPGYVLTQKAYDYWN 199
SMc00264        ATKAAVNQATVAIAAQYA------------PHGIRAN------CVVPGLIDTPLIYKQIS 205
SMc01635        TTKHAVIAMTRQMAGDYA------------KYGVRVN------ALCPGWVDTPFNEPFID 197
SMa0719         ATKAAQHNLIQSLARDYA------------SRGVLLN------TLAPGLIDTDRNAARRD 188
SMb20493        AAKAGVEAMTHVLSKELR------------GRDITVN------AVAPGPTAT------DL 189
SMb21383        ASKMAQELYATVLAKELA------------GRNISVN------AIAPGVVNT------TL 241
SMb20456        ASKAALAGLTKGLARDLG------------PRGITVN------IVHPGSTDT------DM 194
SMc00165        ASKGALETLTKNTAYAVL------------RNRIRVN------GLNIGWMASEGEDRIQR 205
SMc02486        ATKGAIDSFTVGLAREVA------------TEGIRVN------AVRPGIIDT------EI 196
SMb20210        LTKGGLNAVTRSLAMEFS------------RSGVRVN------AVSPGVIKTP-------- 183
SMc00136        VFRSGLAAFTKLFADRYA------------ADNIRMN------NVLPG--------FIDS 182
SMc01157        ALRSALAGWMKTLAREVA------------ADGVTVN------MLLPGRLATDRTLGFDR 196
SMc00005        VAKAALEASVRYLAVDLG------------NRGIRVN------AISAGPIKTLAASGIGD 205
SMc00326        VAKAALEASVRYLAADYG------------SRGIRVN------AISAGPIRTLAGAGISD 202
SMa1367         AAKAALEALAKSLAAQLA------------PVGVTVN------CVAPGFTRKDTGGHAAT 172
SMb20214        GSKAFVLAFTLSLQKELE------------GRNIRIQ------AVLPGATATDFWGIAGT 198
SMc01204        ASKGFVRSFSEALHQELR------------HTGVTVT------CVAPGPVSTEFLATSGA 200
SMa0326         ATKFAVEGLSDTLRNEVA------------PLGINVM------TVEPSAFRTEWAGSSNE 193
SMc00733        ASKFALEALSLTLRMELA------------GSGIGVS------LIEPGPIASKFTTNAIA 191
SMb20492        AAKAAMDAIAVQYARELS------------RWGIETS------IIVPGAFTSGTNHFAHS 201
SMb20662        ATKHAITGLTKSTALDGR------------MHDIACG------QIDIGNAATDMTARMST 206
SMc00553        ASKAAVEALARTWAYETQ------------RLPLRIL------SVDPGATRTAMRAQAVP 200
SMc00603        LSKSALWTATQTMAQALA------------PDIRVN------GIGPGPTLPNERQDPRD 197
SMc01175        ASKAALINMAESLKFDLD------------RIGIRLQ------VICPGFVDTQAAPKGGF 202
SMb20076        ASKFGLRGFSEALRAELAD-----------ERDIHIC------DVYPTFMDTPGVGHGAN 199
SMc02039        APSISCMPSAKACARPCR--------------HIASA------SWSSRPVLRSRRSTRPS 189
SMc02271        ASKFALNGFAQALRSEAR------------AHNVGIS------ICFPPDTETPQFKRELA 197
SMc01955        LSKAGVKVFTEALQHELRN---------TEDCKVNAH------LLIPGFVYTSLTAQGRT 209
SMb21384        ATKAAIHSYTVSLREQLKG------------KVEVI------ELAPPAVQTELTPGQST 193
SMc02040        GTKHFVHAVSEGMRQTMS------------SYRVRVM------VISPGVAKSEIDKTITH 92
SMb20660        SGKVTINSPETLAAVNYAKSL-YETFIPGTESWLDINN---NRAFLAGQVSLTANGVSLY 278
SMc02356        TDKGVIIRALREVGYDAPVLGGDAYDDPAMHEALGEKFGNDVYFVTHTWMGPEAHPDMPK 292
                  :
```

Table 4-1. SDRs on megaplasmid pSymA

SDRs	Location (bp)	Protein size (aa)	BLAST % identity to BdhA	Pfam	PROSITE adh_short [LIVSPADNK] - x(9) - {P} - x(2) - Y - [PSTAGNCV] - [STAGNQCIVM] - [STAGC] - K - {PC} - [SAGFYR] - [LIVMSTAGD] - x - {K} - [LIVMFYW] - {D} - x - {YR} - [LIVMFYWGAPTHQ]- [GSACQRHM]	Functional SDR motifs
SMa0059	31488-30703	261	33	adh_short	-	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMa0074	39072-39830	252	29	adh_short	S LTLNGRWDG Y VP Y V A S K G A M L G L T K A L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMa0187	102832-103602	256	28	adh_short	S VTGLTGSKE L LD Y S M T K G G I H A F T R A L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMa0320	179119-179868	249	25	adh_short	S IAGRKVFKG L SV Y C A T K H A V T A F S D G L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMa0326	181233-182072	279	32	adh_short	S IGGLVGYTG V GY Y C A T K F A V E G L S D T L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMa0329	182145-182906	253	33	adh_short	S VSSLGGGWS H AA Y N A A K G G V A N L T R S A	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMa0335	184352-183576	258	28	adh_short	S MGGISGAGI T VA Y S T S K G G V V L M A K S L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMa0339	185620-184847	257	34	adh_short	S IAALTGPGG M SA Y A A S K H G V Q G L T R V V	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMa0389	210350-211030	226	33	adh_short	S LTSYVGIPT A VP Y G A S K S G L L G V T R A L	NNAG, N-S-Y-K
SMa0513	277846-278601	251	34	adh_short	S VQTALARPG I AP Y T A T K G A V G N L T K G M	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMa0719	391032-391751	239	30	adh_short	-	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMa0854	471615-472352	245	30	adh_short	S VAGAIGNPG Q TN Y C A S K A G M I G F S K S L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMa0959	537885-538661	258	27	adh_short	-	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMa1367	751855-751181	224	26	adh_short	-	NNAG, N-S-Y-K
SMa1398	769735-768956	259	26	adh_short	S SNSVAVAVQ R SE Y C A S K A A A S M V S K A L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMa1452	801177-800410	255	25	adh_short	S IAAFDGQIG Q AA Y A A S K G G V A A L T L P A	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMa1629	907592-906813	259	26	adh_short	I QRVLPLPES T TA Y A A A K A A L S T Y S K A M	T-G-X_3-G-X-G, N-S-Y-K
SMa1757	999841-999071	256	25	adh_short	I AGSRVHPFA G VA Y A A S K A A L A S L T R E M	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMa2019	1143193-1143603	136	27	-	I QGVMPLPES T TA Y A A A K A A L S T Y G K S I	T-G-X_3-G-X-G, S-Y-K
SMa2165	1219615-1220355	246	27	adh_short	S SSTYEGASA T SL Y A A T K G A V L A M S R S W	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMa2343	1320285-1321007	240	25	adh_short	S IGALSVVPT A AV Y C A T K F A V R A I S D G L	T-G-X_3-G-X-G, NNAG, N-S-Y-K

Fig.4.2. Distribution of SDRs on pSymA

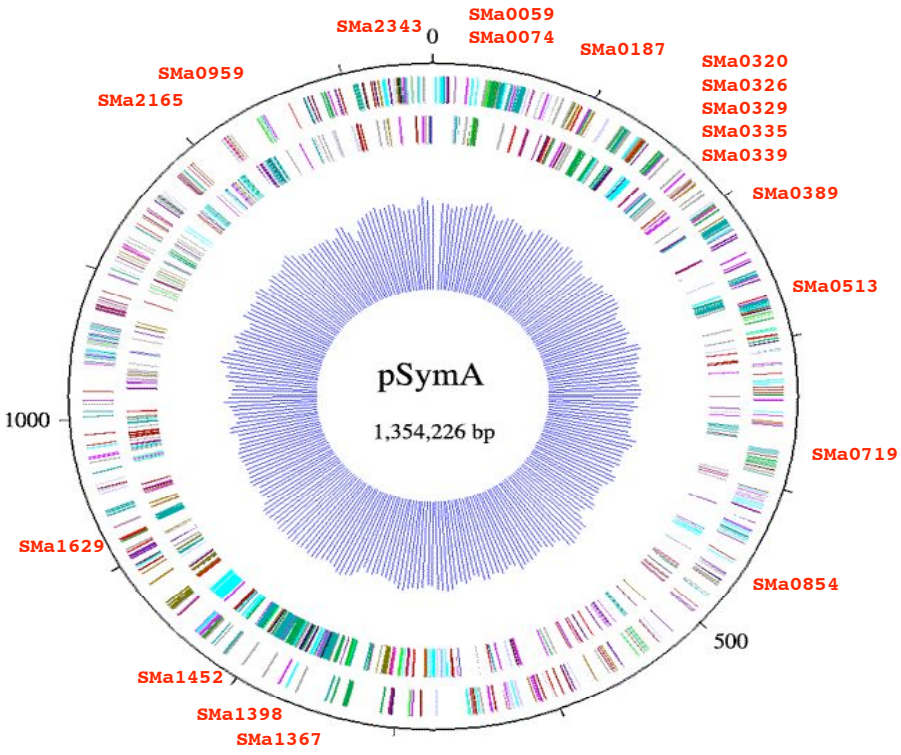

SMa2343
0
SMa0059
SMa0074
SMa0187
SMa0320
SMa0326
SMa0329
SMa0335
SMa0339
SMa0389
SMa0513
SMa0719
SMa0854
500
SMa1367
SMa1398
SMa1452
SMa1629
1000
SMa2165
SMa0959
pSymA
1,354,226 bp

Table 4-2. SDRs on megaplasmid pSymB

SDRs	Location (bp)	Protein size (aa)	BLAST % identity to BdhA	Pfam	PROSITE adh_short [LIVSPADNK] - x(9) - {P} - x(2) - Y - [PSTAGNCV] - [STAGNQCIVM] - [STAGC] - K - {PC} - [SAGFYR] - [LIVMSTAGD] - x - {K} - [LIVMFYW] - {D} - x - {YR} - [LIVMFYWGAPTHQ]- [GSACQRHM]	Functional SDR motifs
SMb20073	83188-82331	285	31	adh_short	S INADVPNPI L LA Y A T T K G A I H N F S A G L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb20076	83965-84981	338	30	adh_short	S LGGFASAPF A AA Y S A S K F G L R G F S E A L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb20210	218000-217287	237	27	adh_short	–	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb20214	222127-222930	267	26	adh_short	S IVGIAPEVL N GV Y G G S K A F V L A F T L S L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb20409	424841-425671	276	33	adh_short	S TAATVGWKD N PA Y C A S K S G L L G L T R C V	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb20456	468077-468823	248	32	adh_short	N LAELVPWPG I SL Y S A S K A A L A G L T K G L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb20492	510469-509567	300	32	adh_short	–	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb20493	511318-510584	244	31	adh_short	–	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb20511	533617-534396	259	30	adh_short	S THAFTIIPH T FP Y P V A K H A L I G M T K A L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb20660	1462236-1460932	434	29	adh_short	–	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb20662	1464299-1463523	258	30	adh_short	S ISAQTPRPN S AP Y T A T K H A I T G L T K S T	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb20692	1494932-1494126	268	32	adh_short	S VQSELARPS I AP Y A A S K G G L K M L T K G M	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb20750	1539559-1540308	249	27	adh_short	S IAAFAARAG D PA Y T A A K G G L A A L T R S L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb20871	1278191-1278940	249	30	adh_short	S LAATKILPD H TV Y A G S K A A T A A F A K N W	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb21010	1244108-1244884	258	100	adh_short	S AHGLVASPF K SA Y V A A K H G I M G L T K T V	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb21111	752219-752953	244	32	adh_short	V ASSIKGVPN R FA Y G V T K A A V I G L T K A V	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb21159	902521-903297	258	30	adh_short	A SVNGMNVFG H PA Y S A A K A G L L H L T R L I	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb21348	1014811-1014047	254	30	adh_short	S LLSFQGGIR V AS Y T A S K H G A L G I T R L L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb21383	1047259-1047259	296	28	adh_short	–	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb21384	1048106-1047363	247	25	adh_short	S GLAFVPMSS T PT Y N A T K A A I H S Y T V S L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMb21474	1399297-1400079	260	31	adh_short	-	T-G-X_3-G-X-G, NNAG, N-S-Y-K

Fig.4.3. Distribution of SDRs on pSymB

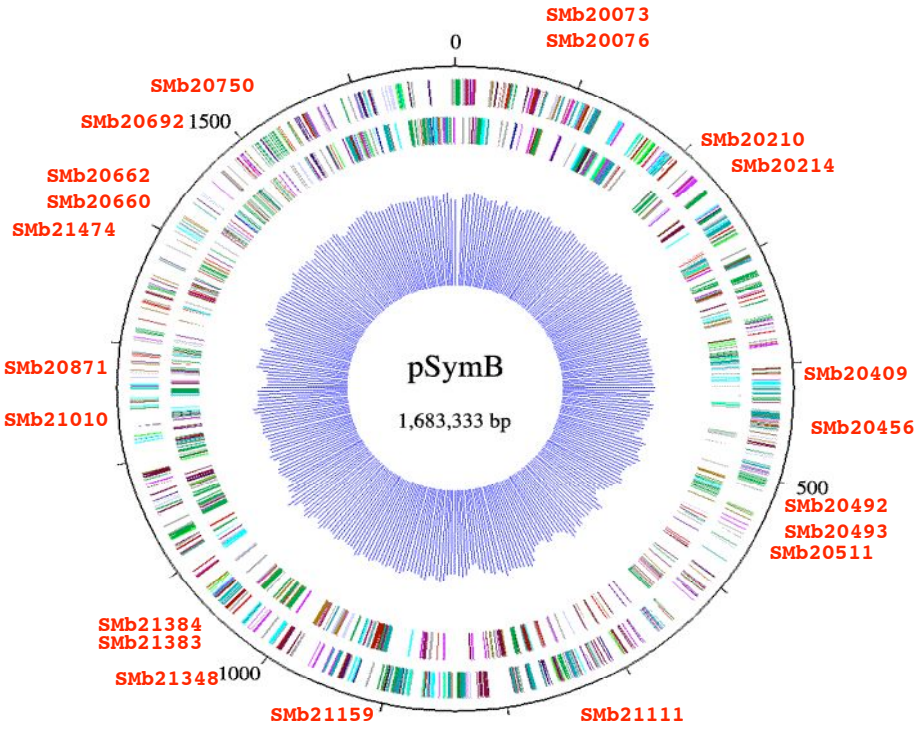

SMb20073
SMb20076
0
SMb20750
SMb20692
1500
SMb20210
SMb20214
SMb20662
SMb20660
SMb21474
SMb20871
SMb21010
pSymB
1,683,333 bp
SMb20409
SMb20456
500
SMb20492
SMb20493
SMb20511
SMb21384
SMb21383
SMb21348
1000
SMb21159
SMb21111

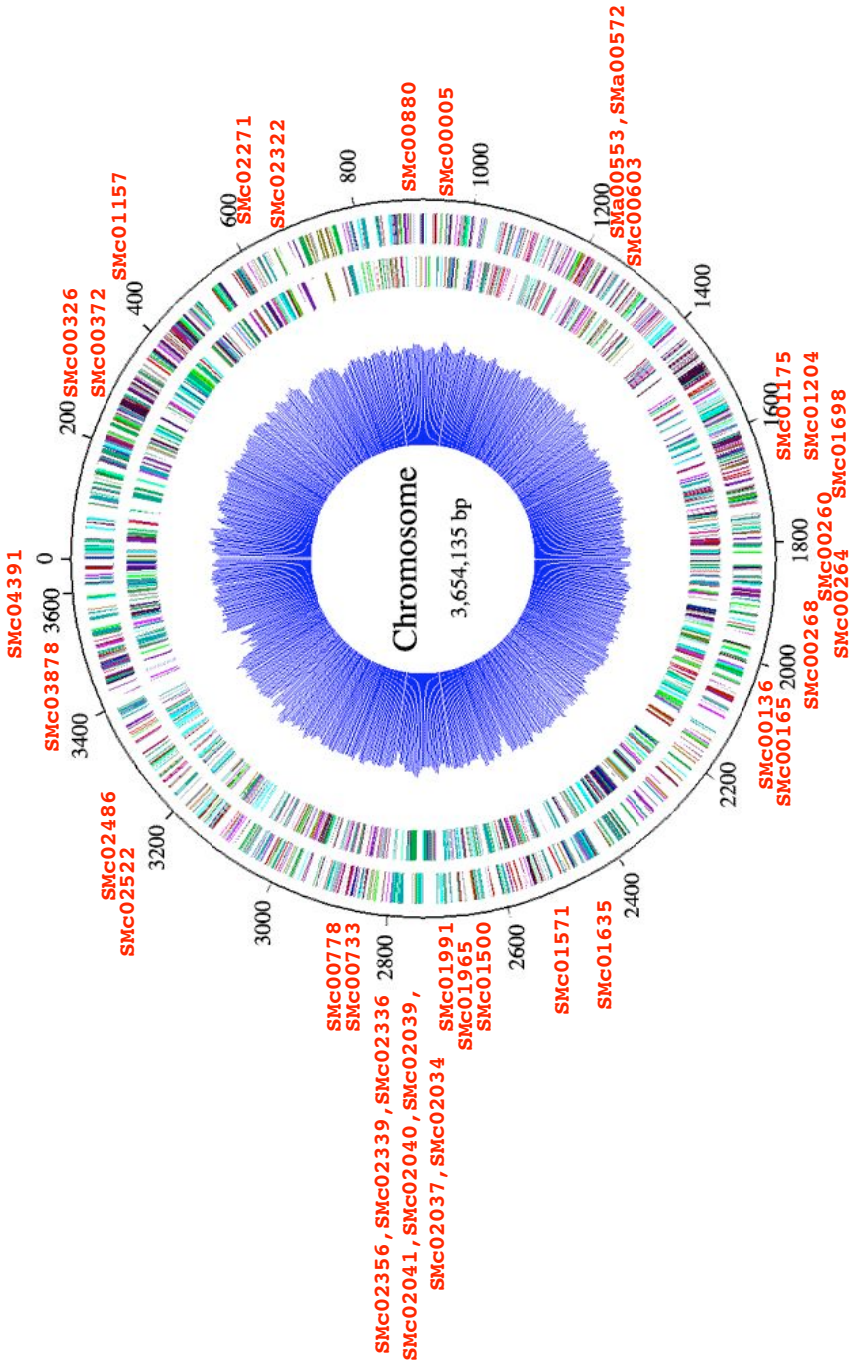
Table 4-3: SDRs on chromosome:

SDRs	Location (bp)	Protein size (aa)	BLAST- % identity to BdhA	Pfam	PROSITE- adh_short [LIVSPADNK] - x(9) - {P} - x(2) - Y - [PSTAGNCV] - [STAGNQCIVM] - [STAGC] - K - {PC} - [SAGFYR] - [LIVMSTAGD] - x - {K} - [LIVMFYW] - {D} - x - {YR} - [LIVMFYWGAPTHQ]- [GSACQRHM]	Functional SDR motifs
SMc00005	986489-987307	272	23	adh_short	–	N-S-Y-K
SMc00136	2034496-2033792	234	23	adh_short	–	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc00165	2000628-1999807	273	29	adh_short	S MSAKAGQPF I AA Y C A S K G A L E T L T K N T	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc00260	1828790-1828026	254	27	adh_short	S VSGLVGNKG R AA Y G A S K G A V N L L T Y I L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc00264	1833254-1832451	267	23	adh_short	–	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc00268	1838204-1837446	252	32	adh_short	S AQRGGGIFG G PH Y S A A K A G V L G L T K A M	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc00326	281601-280795	268	23	adh_short	–	N-S-Y-K
SMc00372	326479-327249	256	31	adh_short	–	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc00553	1223651-1223651	246	29	adh_short	S SAAHKCKPF W GP Y S A S K A A V E A L A R T W	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc00572	1241639-1242376	245	31	adh_short	S VVGVTGNPG Q AN Y C A S K A G M I G F S K S L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc00603	1275094-1274330	254	23	adh_short	–	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc00733	2847909-2848742	277	25	adh_short	S ILGLVPYRW R GA Y N A S K F A L E A L S L T L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc00778	798230-798997	255	27	adh_short	–	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc00880	907031-906261	256	29	adh_short	S IAFLLNMPE I PA Y S T A K A G I I G L T K S L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc01157	393585-392806	259	28	adh_short	–	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc01175	1755390-1754542	282	23	adh_short	–	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc01204	1725822-1724938	294	24	adh_short	–	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc01500	2630149-2629376	257	27	adh_short	S QAGRRGEAL V AI Y C A T K A A V I S L T Q S A	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc01571	2514076-2514837	253	30	adh_short	S VADEGGHPG M SV Y S A S K A A L V S F A S V L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc01635	2427523-2428284	253	33	adh_short	S SSGVLYDRE M IA Y T T T K H A V I A M T R Q M	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc01698	1729161-1729916	251	35	adh_short	S ILGAVALPT A SA Y T A A K H G V V G L T K A A	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc01955	2676633-2677508	291	25	adh_short	–	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc01991	2717557-2716832	241	30	adh_short	–	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc02034	2741143-2740370	257	34	adh_short	S GTIVNRPQQ Q VH Y N A A K A G V H H L T R S L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc02037	2743512-2744282	256	32	adh_short	S GYISNLPQN Q VA Y N A S K A G V H M L T K S L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc02039	2745430-2746128	232	23	adh_short	–	T-G-X_3-G-X-G, NNAG, N-S-Y-K

Table 4-3. SDRs on chromosome

SDRs	Location (bp)	Protein size (aa)	BLAST % identity to BdhA	Pfam	PROSITE adh_short [LIVSPADNK] - x(9) - {P} - x(2) - Y - [PSTAGNCV] - [STAGNQCIVM] - [STAGC] - K - {PC} - [SAGFYR] - [LIVMSTAGD] - x - {K} - [LIVMFYW] - {D} - x - {YR} - [LIVMFYWGAPTHQ]- [GSACQRHM]	Functional SDR motifs
SMc02040	2745726-2746157	143	25	-	L HAGRKVYPH H DV Y G G T K H F V H A V S E G M	N-S-Y-K
SMc02041	2746159-2746923	254	35	adh_short	S QAAIIGIEG H VA Y C A S K A G I I G M T N C M	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc02271	639120-638317	267	28	adh_short	S GAGLIGIYG Y TA Y C A S K F A L N G F A Q A L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc02322	693957-691858	699	24	adh_short, epimerase, aldolase	-	T-G-X_3-G-X-G, NNAG, N-S-Y-K T-G-X_3-G-X-G, NNAG, N-S-Y-K T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc02336	2755753-2756511	252	30	adh_short	S MLSYVSDPL V PA Y G A S K S G V L G L T R H L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc02339	2761053-2760283	256	32	adh_short	S GFIVNRPQP Q AH Y N A S K A A V H H L T K S L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc02356	2777494-2778666	390	23	-	S MTGIQAPLD T PG Y N G A K V A V K Y L N D N G	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc02486	3310309-3309557	250	28	adh_short	I AAVLGAPGQ Y VD Y A A T K G A I D S F T V G L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc03878	3546426-3545701	241	34	adh_short	S INGQKGQMG Q VN Y S A A K A G D L G L T K A L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc04391	3595211-3594438	257	28	adh_short	S VSAVMTSPE R LD Y C M S K A G L A A F S Q G L	T-G-X_3-G-X-G, NNAG, N-S-Y-K
SMc02522	3270206-3269436	256	27	adh_short	I SARMGGSPG S SI Y S A T K A F V A T Y S K A L	T-G-X_3-G-X-G, NNAG, N-S-Y-K

Fig.4.4. Distribution of SDRs on the chromosome


Chromosome
3,654,135 bp
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
2600
2800
3000
3200
3400
3600
SMc04391
SMc03878
SMc02486
SMc02522
SMc02356,SMc02339,SMc02336
SMc02041,SMc02040,SMc02039,
SMc02037,SMc02034
SMc00778
SMc00733
SMc01991
SMc01965
SMc01500
SMc01571
SMc01635
SMc00136
SMc00165
SMc00268
SMc00260
SMc00264
SMc01175
SMc01204
SMc01698
SMa00553,SMa00572
SMc00603
SMc00880
SMc00005
SMc02271
SMc02322
SMc01157
SMc00326
SMc00372

4.2 Construction of SDR mutations

Phenotypic analysis is a potentially very powerful approach to determine gene function. In order to understand the role of the SDR-encoding genes in *S. meliloti*, it was important that these genes be mutated. SmP110, a derivative of Rm1021 in which a frameshift mutation in *pstC* is corrected (Yuan *et al*. 2005) was used as the parental strain. The goal was to create reporter gene fusions that inactivate each of the SDR encoding genes. A segment of the gene of interest was used to facilitate single homologous recombination event, thereby disrupting the gene of interest.

For each of the SDR-encoding genes, a 600-bp region of the coding sequence defined on one end by the start codon was amplified and cloned into pGEM T-easy (Promega, USA). The orientation of the fragments with respect to T7 or SP6 promoter in pGEM T-easy was determined by PCR (TEF^+ or TEF^- respectively). In most cases, these fragments were then released as Not1 fragments and ligated into the multiple cloning site of the vector pTH1703 (Cowie *et al.* 2006). pTH1703 has a pMB1 replicon (unable to funcion in *S. meliloti*) and confers resistance to gentamicin. Its multiple cloning site is flanked on one side by *gusA* (β-glucuronidase) and tdimer2 (RFP), and on the other side by *gfp*+ (GFP) and *lacZ* (β-galactosidase). The 600 bp fragments of SMa0389, SMa0719,SMb20210 and SMb21383 were cloned as EcoR1 fragments into pVIK112 (Winans *et al.* 1997), a suicide vector with R6K replicon and kanamycin / neomycin as selectable marker. A 600 bp region from six of the SDRs, SMa0326, SMa0959, SMa1629, SMb20662, SMc01157 and

SMc01955 could not be PCR amplified. To create single recombinants for these genes the *S. meliloti* ORFeome (Schroeder *et al.* 2005) was used. The *S. meliloti* ORFeome consists of each of the 6204 ORF identified in the genome, cloned into pMK2010 (Schroeder *et al.* 2005). Using appropriate restriction enzymes internal fragments of these genes were cloned into pTH1703. A 393 bp fragment of HaeIII digested SMa1629, a 343 bp fragment of HincII digested SMa0959, a 687 bp fragment of AluI digested SMa0326 and a 329 bp fragment of BmgBI cut SMc01955 were ligated to SwaI cut, dephosphorylated pTH1703. Similarly, 616 bp and 285 bp fragments of EagI digested SMb20662 and SMc01157 respectively, were ligated to NotI cut, dephosphorylated pTH1703 (Fig.4.7.). The orientation of the fragments in each of the pTH1703 clones was tested by PCR.

To construct the mutants, pTH1703 or pVIK112 carrying the ORF fragments were mobilized into SmP110 by triparental mating and transconjugants counter-selected using streptomycin / gentamicin for pTH1703 and streptomycin / neomycin for pVIK112 (Fig.4.5.). Cointegrates resulting, as a single recombination event will have two copies of the gene of interest. One copy would be the intact promoterless copy of the gene and the second copy would be the truncated version of the gene with the promoter and fusion to a reporter gene. Recombinant strains in the *gus/rfp* orientation were designated as SMa/b/c XXXXfg, while those in the *lac/gfp* orientation were designated as SMa/b/cXXXXfl, where XXXX refers to the annotated gene number.

Three of the 78 genes identified as SDRs, *bdhA* (SMb21010), *nodG* (SMa0854) and *phbB* (SMc03878) have previously been characterized. Of the identified SDRs, *fabG* (SMc00572) was recalcitrant to mutation. Thus 77 of the 78 genes identified as SDRs have been mutated. Fourteen of the mutants were screened by PCR to confirm if recombination resulting in the mutation had occurred in the gene of interest. For this purpose a primer upstream from start codon of the gene of interest (forward primer) was used in a PCR reaction with a primer designed from *gus* or *gfp* on pTH1703 (reverse primer). These PCR results confirmed that the mutations were in the desired locus (Fig.4.6.). The 77 mutants were further characterized for their role in substrate utilization and symbiosis.

Fig.4.5. Construction of *sdr*-reporter gene fusion:

600 bp fragments from the genes of interest were cloned into NotI or SwaI digested pTH1703 or EcoRI digested pVIK112. Transconjugants resulting as a result of a single crossover event was selected on antibiotic markers (Sm and Gm), streak purified thrice and used to screen for a metabolic and symbiotic phenotype.

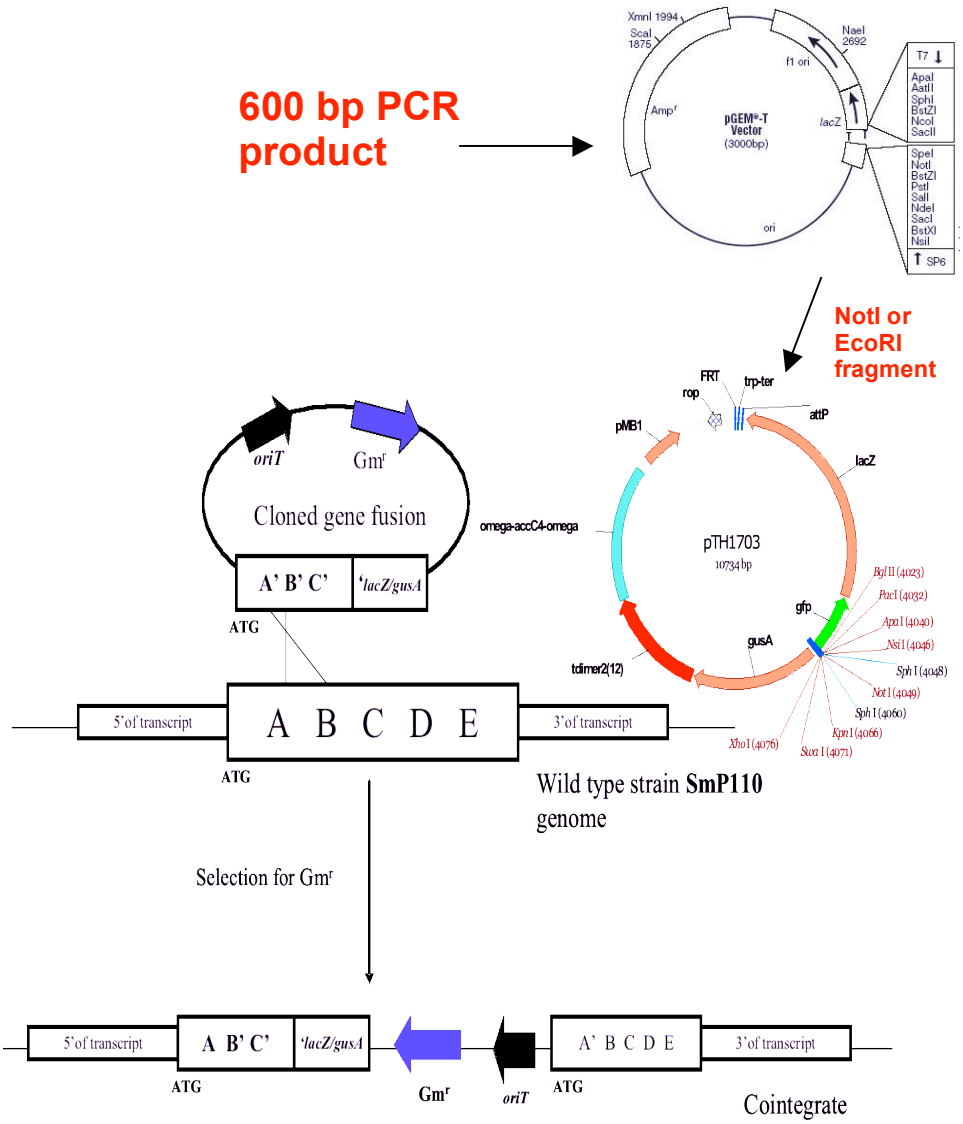
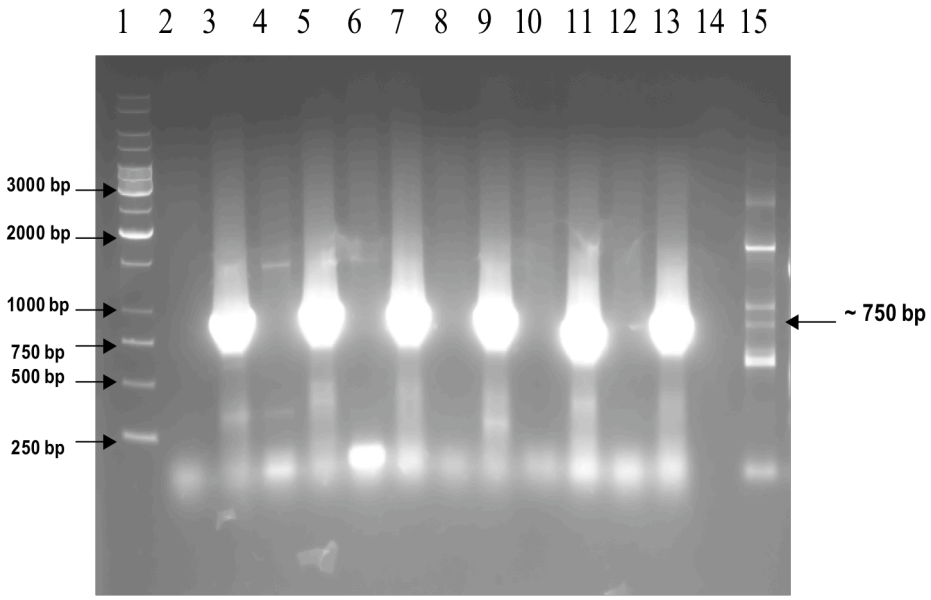

600 bp PCR product
pGEM®-T Vector (3000bp)
NotI or EcoRI fragment
oriT
Gmr
Cloned gene fusion
A' B' C'
'lacZ/gusA
ATG
5'of transcript
A B C D E
3'of transcript
ATG
pTH1703
10734 bp
omega-accC4-omega
pMB1
FRT
trp-ter
rop
attP
lacZ
gfp
gusA
tdimer2(12)
Wild type strain SmP110 genome
Selection for Gmr
A' B C D E
Gmr
oriT
Cointegrate

Fig.4.6. PCR showing that recombination has occurred in the desired locus in the correct orientation:

Genomic DNA was isolated from each of the recombinant strains indicated. PCR was done using primers generated from the start codon of the gene of interest and primers from the vector (either *gfp* or *gus*). The expected result would be a PCR product approximately 750 bp size.

Lane1: 1kb ladder, Lane 2: SMc00165fg with *gfp* primer, Lane 3: SMc00165fg with *gus* primer, Lane 4: SMc00778fl with *gus* primer, Lane 5: SMc00778fl with *gfp* primer,Lane 6: SMb20456fl with *gus* primer, Lane 7: SMb20456fl with *gfp* primer, Lane 8: SMc00603fg with *gfp* primer, Lane 9: SMc00603fg with *gus* primer, Lane10: SMc01991fg with *gfp* primer, Lane11: SMc01991fg with *gus* primer, Lane 12: SMc00136fl with *gus* primer, Lane 13: SMc00136fl with *gfp* primer, Lane 14: SMc00778fg with *gfp* primer and Lane 15: SMc00778fg with *gus* primer, Lane 16: 1kb ladder, Lane 17: SMa1757fg with *gus* primer, Lane 18: SMa1757fg with *gfp* primer, Lane 19: SMa2343fl with *gfp* primer, Lane 20: SMa2343fl with *gus* primer, Lane 21: SMa2343fg with *gus* primer, Lane 22: SMa2343fg with *gfp* primer


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
3000 bp
2000 bp
1000 bp
750 bp
500 bp
250 bp
~ 750 bp

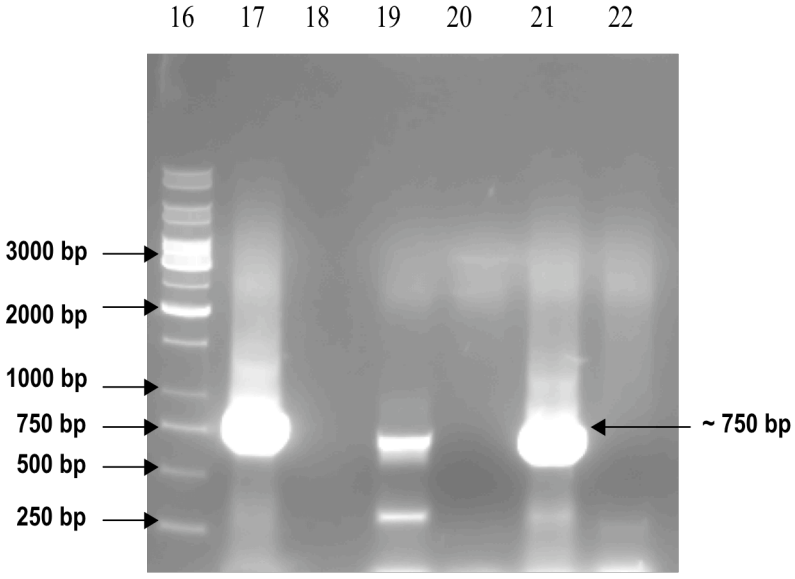
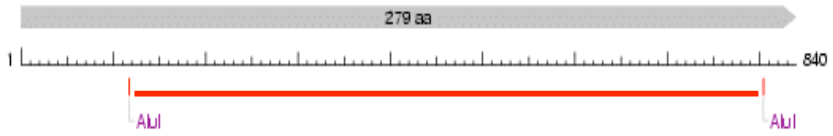

16 17 18 19 20 21 22
3000 bp
2000 bp
1000 bp
750 bp
500 bp
250 bp
~ 750 bp

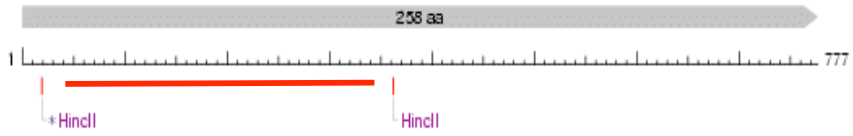
Fig.4.7. Cloning of SDR fragments from the ORFeome:

SMa0326 (I-687 bp AluI fragment), SMa0959 (II-343 bp HincII fragment), SMa1629 (III-393 bp HaeIII fragment), SMc01955 (IV-329 bp BmgBI fragment), SMb20662 (V-616 bp EagI fragment), SMc01157 (VI- 285 bp EagI fragment) from *S. meliloti* ORFeome (Schroeder *et al.* 2005). The region depicted in red indicates the region cloned.

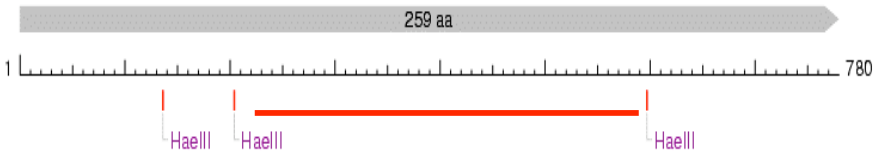
I



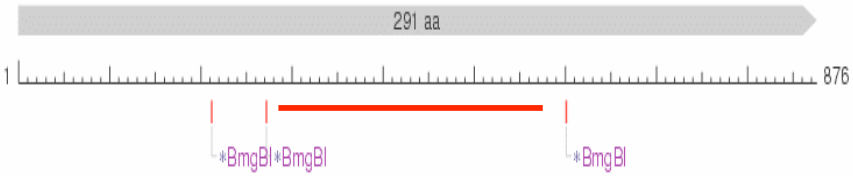
II



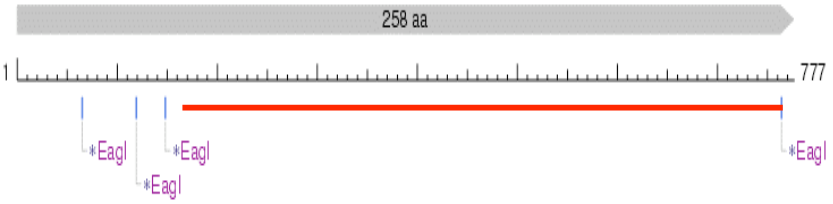
III



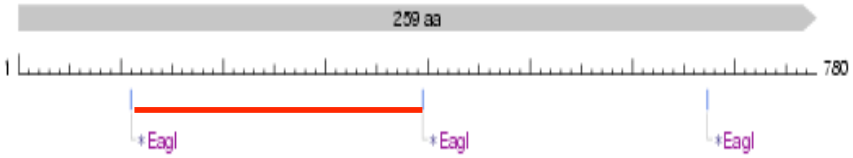
IV



V



VI



4.3 Screening the SDR mutant strains for carbon utilization phenotype

Each of the SDR mutants was screened on 93 different substrates as sole carbon source (Table 3-1) on RMM agar. Whenever available, mutants constructed in each orientation in the integration vector were screened. Each carbon source was supplemented at a concentration of 0.1%. The mutants were streaked for single colonies, and the plates were incubated at 30°C for 4 weeks and scored for growth by comparison to the growth of SmP110. If the mutated SDR is involved in the catabolism of a particular substrate, then the mutant should exhibit altered growth. Each screen was performed three times, and each strain exhibiting a phenotype was further confirmed by testing transductants for the same phenotype.

SmP110 and all of the mutants except the SMc00778 mutants could grow on glucose as sole carbon source. The SMc00778 mutant appears to be auxotrophic for proline. This mutant was unable to utilize any the carbon sources unless the media was supplemented with proline. As shown in the Table 4-4 17 SDR mutants showed a phenotype with respect to carbon utilization. SMc02322 mutant is not able to utilize rhamnose and from previously published results it is known that this region is equivalent to the region responsible for rhamnose catabolism in *R. leguminosarum* (Richardson *et al*. 2004). The inability of SMc01500 to utilize sugar alcohols further strengthens the validity of our screening as this SDR lies in a region annotated to be involved in sorbitol metabolism. The SMb20210 mutant shows very weak growth on most substrates and lies in the region of pSymB annotated to be

involved in the Calvin cycle. The inability of SMc01204, SMc02034 and SMc02356 to utilize ribonic-γ-lactone is a significant find and needs to be further investigated. Very little information is available about the catabolism of this substrate.

SMb20492 and SMb20493 lie adjacent to each other in the genome and are transcribed in the same direction. Downstream of these SDRs are genes annotated to be involved on sugar transport, a carbohydrate kinase and a D-fucose phosphate aldolase (*fucA2*). From the substrate utilization screens it was determined that SMb20492 and SMb20493 mutants are unable to utilize fucose. This observation taken together with their genetic context strengthens the validity of our screen. Similarly SMb20073 and SMb20076 show similar phenotypes. SMb20072, an ABC transporter was induced by myo-inositol (Mauchline *et al.* 2007) but our screens did not reveal a phenotype on myo-inositol for these two SDRs. SMb20750 mutant is unable to grow on ornithine and leucine. A region involved in the catabolism of ornithine, GABA and leucine (see Chapter 3) contains the SDR SMb20750, further confirming this phenotype. This gene has been annotated as a putative gluconate-5-dehydrogenase based on sequence homology. The result of the phenotypic screens of the SDR mutants on pSymB is consistent with a study done in our lab (Chapter 3) wherein deletion of the region surrounding and including these genes from pSymB resulted in similar phenotypes. RmF909 carries a deletion in the pSymB megaplasmid that includes the SDR encoding genes SMb20492, SMb20493, SMb20409, SMb20210 and SMb20214. This deletion strain is

unable to utilize fucose and succinate as its sole carbon source, a phenotype consistent with SMb20492, SMb20493 and SMb20409. However this deletion strain does not exhibit a poor growth phenotype as that exhibited by SMb20210.

A few of the SDR mutants (SMc02486, SMb20210 and SMa0719) showed general poor growth phenotype on all carbon sources. SMc02486 lies in a region that includes several genes encoding enzymes involved in the TCA cycle (Dymov *et al.* 2004). Genes annotated to be involved in the Calvin cycle, an essential pathway for carbon fixation, flank SMb20210. Arginine metabolism genes flank SMa0719. The SMb20409 mutant showed very weak growth on galactitol and D-mannose. SMb20410, an ABC transporter is induced by galactitol and D-mannose, supporting this observation. Mutation in SMb21111 resulted in weak growth on D- and L- fucose.

Table 4-4. Free-living and symbiotic phenotypes of SDR mutant strains:

ORF	Gene	Carbon utilization deficiency	Symbiotic phenotype
SmP110		none	
SMa0059		none	Fix^+
SMa0074		none	Fix^+
SMa0187		none	Fix^+
SMa0320		none	Fix^+
SMa0326		none	Fix^+
SMa0329		none	Fix^+
SMa0335		none	Fix^+
SMa0339		none	Fix^+
SMa0389		none	Fix^+
SMa0513		none	Fix^+
SMa0719[b]		all except glucose, DL-malic acid, succinate, fructose, D-lyxose, D-xylose	$Fix^{+/-}$
SMa0854	*nodG*	none	Fix^+
SMa0959		none	Fix^+
SMa1367		none	Fix^+
SMa1398		none	Fix^+
SMa1452		none	Fix^+
SMa1629		none	Fix^+
SMa1757		none	Fix^+
SMa2019		none	Fix^+
SMa2165		none	Fix^+
SMa2343		none	Fix^+
SMb20073[a]		D-fucose, lysine, leucine, proline	Fix^+
SMb20076[a]		D-fucose,lysine, leucine, proline	Fix^+
SMb20210[b]		all except glucose, DL-malic acid, succinate, fructose, D-lyxose, D-xylose, D-ribose	Fix^-
SMb20214		none	$Fix^{+/-}$
SMb20409[c]		galactitol, D-mannose	Fix^+
SMb20456		none	Fix^+
SMb20492[a]		D-fucose, succinate, monomethyl-succinate, leucine, lysine	Nod^-
SMb20493[a]		D-fucose, succinate, monomethyl-succinate, leucine, lysine	$Fix^{+/-}$
SMb20511		none	Fix^+
SMb20660		none	Fix^+
SMb20662		none	Fix^+
SMb20692		none	Fix^+
SMb20750[a]		D-fucose, glucose-1-phosphate, leucine, ornithine	Fix^+
SMb20871		none	Fix^+
SMb21010	*bdhA*	none	Fix^+
SMb21111[c]		D-fucose and L-fucose	Fix^+
SMb21159[a]		succinate, glucose-1-phosphate	Fix^+

ORF	Gene	Carbon utilization deficiency	Symbiotic phenotype
SMb21348		none	Fix$^+$
SMb21383		none	Fix$^+$
SMb21384		none	Fix$^+$
SMb21474		none	Fix$^+$
SMc00005	*fabI1*	none	Fix$^-$
SMc00136		none	Fix$^+$
SMc00165		none	Fix$^+$
SMc00260		none	Fix$^+$
SMc00264		none	Fix$^+$
SMc00268		none	Fix$^+$
SMc00326	*fabI2*	none	Fix$^-$
SMc00372		none	Fix$^+$
SMc00553		none	Fix$^+$
SMc00572	*fabG*	not tested	not tested
SMc00603		none	Fix$^+$
SMc00733		none	Fix$^+$
SMc00778[b]		proline auxotroph, all except L-proline, trans-4-hydroxy-L-proline, glycyl-L-proline	Fix$^+$
SMc00880		none	Fix$^+$
SMc01157		none	Fix$^+$
SMc01175		none	Fix$^+$
SMc01204[a]		ribonic-γ-lactone	Fix$^+$
SMc01500[a]	*smoS*	mannitol, maltitol, sorbitol, myo-inositol, meso-erythritol	Fix$^+$
SMc01571		none	Fix$^+$
SMc01635		none	Fix$^+$
SMc01698		none	Fix$^+$
SMc01955		none	Fix$^+$
SMc01991		none	Fix$^+$
SMc02034[a]		ribonic-γ-lactone	Fix$^+$
SMc02037		none	Fix$^+$
SMc02039		none	Fix$^+$
SMc02040		none	Fix$^+$
SMc02041[a]		monomethyl succinate	Fix$^+$
SMc02271		none	Fix$^+$
SMc02322[a]		L-rhamnose	Fix$^+$
SMc02336		none	Fix$^+$
SMc02339		none	Fix$^+$
SMc02356[a]		ribonic-γ-lactone, monomethyl succinate	Fix$^+$
SMc02486[b]		all except glucose, D-lyxose, L-leucine, N-acetyl-D-glucosamine, N-acetyl-D-galactosamine, sorbitol, meso-erythritol, maltitol, mannitol, inositol	Fix$^-$
SMc02522		none	Fix$^+$
SMc03878	*phbB*	none	Fix$^+$
SMc04391		none	Fix$^+$

[a]- Grows well on all substrates except those indicated. [b]- Generally shows poor growth on most substrates. [c]- Grows well on all substrates but weakly on those indicated. All growth phenotypes are compared to the wild type strain SmP110. SMc00572 was recalcitrant to mutation and hence not tested

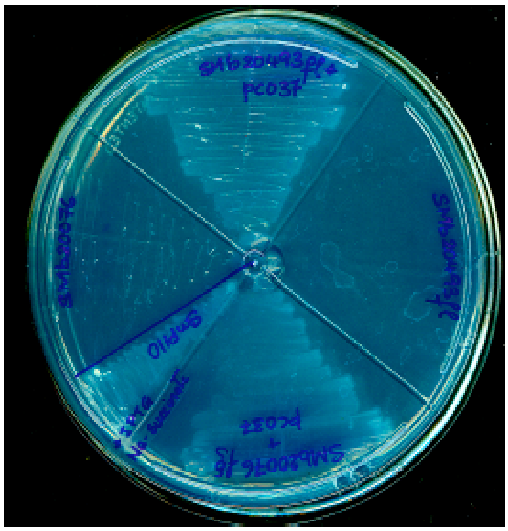
4.4 Complementation of carbon source utilization phenotype of free-living SDR mutant strains

All the ORFs of *S. meliloti* have been cloned into pMK2010 (Schroeder *et al.* 2005) and the 78 SDR containing clones were obtained. The SDRs from these clones were transferred into the Gateway-compatible expression vector pCO37 (Clark *et al*., unpublished). This tetracycline resistance-conferring vector has a *tac* promoter and generates an N-terminal strep-tag. pCO37 clones containing the ORF of SDRs showing a carbon utilization phenotype were introduced into the corresponding mutant background to see if the intact gene could complement the loss of function. All phenotypes could be complemented in this manner (Fig.4.8.).

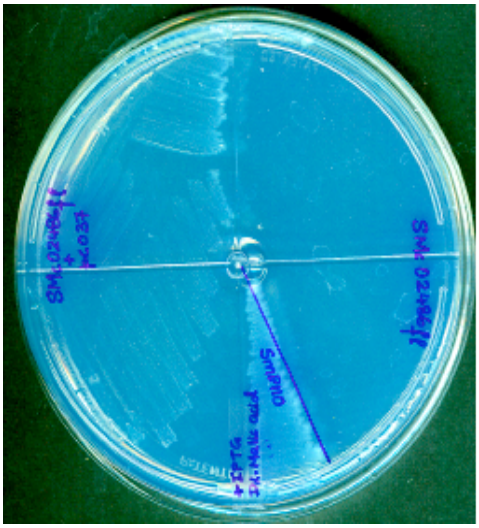
Fig.4.8. RMM agar plates showing phenotypic complementation:

The SDR mutant strains and the mutant complemented with the wild type gene were plated on RMM containing appropriate carbon source. The mutant did not show growth on the media, however the complemented strains could grow as well as the wild type SmP110. SmP110 is used as a positive control in all plates.

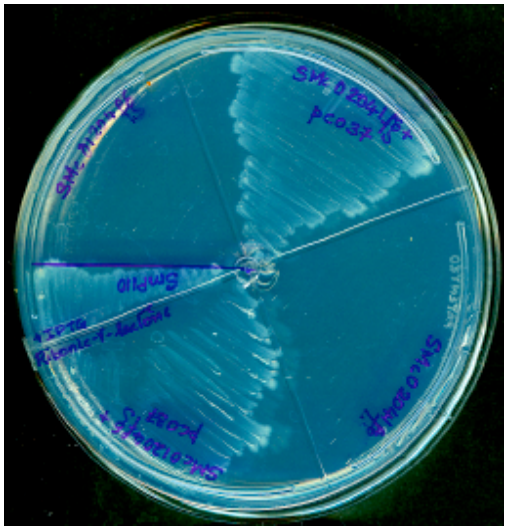
A. SMb20076 and SMb20493 mutants strains and corresponding complemented strains on RMM + sodium succinate.
B. SMc02486fl and SMc02486fl complemented strain on RMM + DL- malic acid.
C. SMc01204fg and SMc02041fg and the corresponding complemented strains on RMM + D-ribonic-γ-lactone.
D. SMc00778fl and SMc02486fl and the corresponding complemented strains on RMM + D- arabinose.



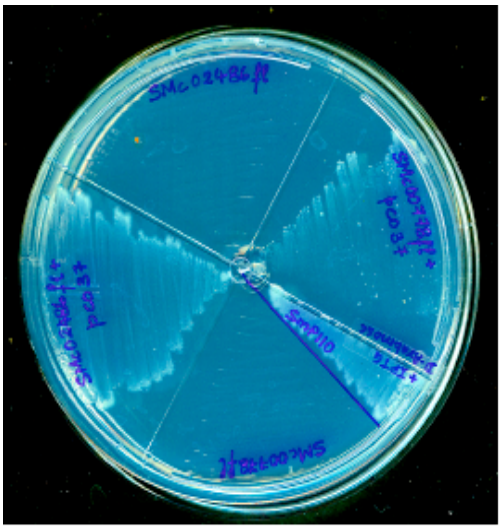
A



B



C



D

4.5 Characterization of symbiotic phenotype of SDR mutant strains

The SDR mutants were screened on alfalfa plants in vermiculite containing Leonard assemblies and agar slants to determine whether any of these mutations affect symbiosis. Seedlings were inoculated with mutant strains and incubated in a growth chamber (Percival Scientific, 25°C, 16 hr light and 8 hr dark cycle) for 4 weeks. The shoot dry weights were measured and compared to the wild type and uninoculated control. Furthermore, the roots were inspected for the presence of nodules. Nodules were surface sterilized, crushed and plated onto TY plates to see if bacteria could be recovered from the nodules.

Plants inoculated with the wild type strain were healthy and produced pink nodules (Fig.4.9.A). The uninoculated control plants were dry and chlorotic with no root nodules (Fig.4.9.B). Most of the SDR mutants screened were able to produce healthy pink nodules. All the SDRs mutants that showed a carbon utilization phenotype were Fix^+ except the SMb20492 and SMc02486 mutants (Table 4-5). Though the SMc02486 mutant formed a large number of white nodules, the plants inoculated with this strain were dry and stunted, almost comparable to the un-inoculated control (Fig.4.9.G). No nodules formed on plants inoculated with the SMb20492 mutant, and hence Nod^- (Fig.4.9.C). The shoot dry weights of plants inoculated with these mutant strains were comparable to that of the un-inoculated control (Table 4-5). SMb20493 showed an intermediate phenotype wherein the inoculated plants were not as healthy as those inoculated with SmP110 neither were they as unhealthy as the un-

inoculated control (Fig.4.9.J). Even though the SMc00778 mutant appears to be a proline auxotroph, it was Fix^+ (Fig.4.9.K). SMb20219 mutant does not grow well on most of the substrates tested and showed a Fix^- phenotype on alfalafa plants.

Mutants that did not have a carbon utilization phenotype were also tested on plants to determine if the SDR mutations resulted in a loss of symbiotic phenotype without the loss of carbon utilization phenotype. The SMc00005 (*fabI1*) and SMc00326 (*fabI2*) mutants also showed a Fix^- phenotype (Fig.4.9.D, Fig.4.9.E, Fig.4.9.F) that was reflected in the reduced shoot dry weights (Table 4-5). SMc00005 (*fabI1*) and SMc00326 (*fabI2*) mutants formed nodules that were very small, white and irregularly shaped. These genes have been annotated to be involved in fatty acid biosynthesis. Plants infected with SMc00005 mutant strains showed abnormal root development. The roots appeared to be stumpy and very few lateral roots were present. The shoot dry weights of these mutant strains were similar to that of the un-inoculated control (Table 4-5). Plants infected by the SMa0719 and SMb20214 mutant strains showed an intermediate phenotype between wild type and un-inoculated control (Fig.4.9.H, Fig.4.9.I). The nodules formed by this strain were white and elongated. The shoot dry weights of plants inoculated with these mutants were significantly lower than wild type but higher than the un-inoculated control (Table 4-5).

As further confirmation of the above results each mutation was transduced into a fresh SmP110 background and these transductants tested on

plants. In each case, the same results were obtained, thus confirming the previous results. The plant assays have revealed that this group of proteins has a critical role in the symbiotic capabilities of this organism. Five of the SDRs, SMa0719, SMb20210, SMb20492 and SMb20493 that showed a symbiotic phenotype also showed a carbon utilization phenotype (Table 4-4). Three of them, SMb20214, SMc00005 (*fabI1*) and SMc00326 (*fabI2*), seemed to be exclusively involved in symbiosis.

Fig.4.9. Screening the SDR mutant strains for a symbiotic phenotype:

The 76 SDR mutant strains generated in this study were tested on alfalfa to test for a symbiotic phenotype. The plants were tested 4 weeks post-inoculation and compared to the positive control (alfalfa inoculated with SmP110) and negative control (uninoculated alfalfa seedlings).

A. Alfalfa seedling inoculated with wild type strain SmP110. Plants are green, healthy and form normal root and shoot systems.

B. Uninoculated alfalfa seedlings. The growth of the plants is stunted and chlorosis is evident in the stem and leaves.

A

B

C. Alfalfa seedling inoculated with SMb20492fg SDR mutant strain. It shows a Nod^- phenotype. The plant is comparable to the un-inoculated control.

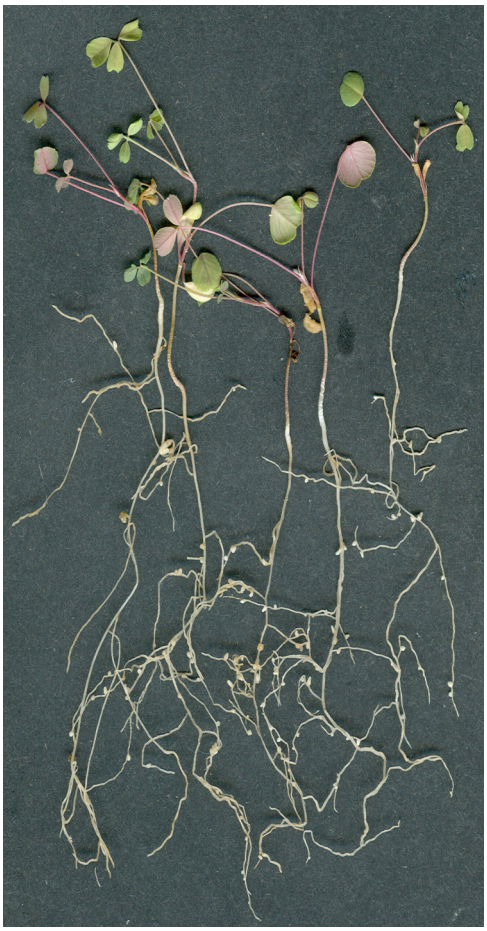
D. Alfalfa seedling inoculated with SMb20210 SDR mutant strain. It shows a Fix^- phenotype. The plants are stunted and show chlorosis.

C D

E. Alfalfa seedlings inoculated with SMc00005fg SDR mutant strain. It is Fix^- with only a few nodules formed and the plants are severely stunted and chlorotic. The lateral root formation seems to be affected and the roots appear to be knobby and stunted.

F. Alfalfa seedling inoculated with SMc00326fg SDR mutant strain. It is Fix^-. The plants are comparable to the un-inoculated control. The nodules formed are of abnormal morphology and not like the wild type (pink, elongated)

E



F

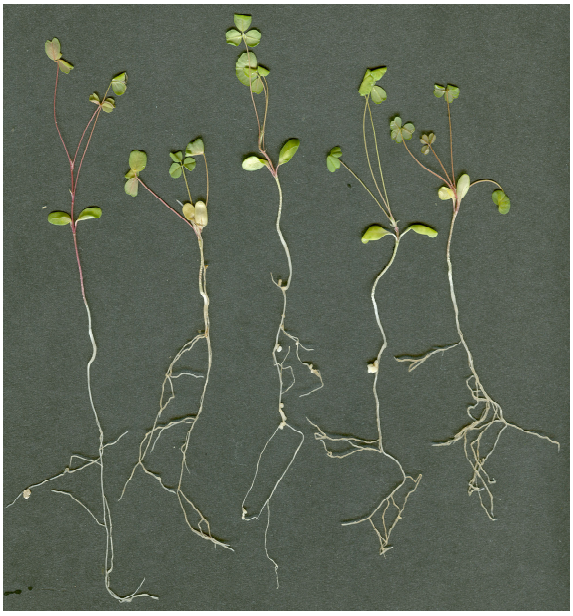
G. Alfalfa seedling inoculated with SMc02486fg SDR mutant strain. This strain is Fix^-.The plants are stunted and chlorotic. The nodules formed do not appear to be fixing nitrogen.

H. Alfalfa seedlings inoculated with SMa0719 SDR mutant strain. This strain appears to fix nitrogen but the plants are not healthy as the wild type. They grow considerably better than the uninoculated control, and thus show an intermediate phenotype.

I. Alfalfa seedlings inoculated with SMb20214fg SDR mutant strain. This shows an intermediate phenotype similar to SMa0719 SDR mutant.

G

H



I

J. Alfalfa seedlings inoculated with SMb20493fl mutant strain. It shows an intermediate phenotype similar to SMa0719 and SMb20214 mutants.

K. Alfalfa seedlings inoculated with SMc00778fl mutant strain. It is Fix^+ and comparable to wild type in growth even though it appears to be a proline auxotroph.

J

K

Table 4-5. Shoot dry weight of alfalfa plants inoculated with SDR mutant strains showing a Fix^- or $Fix^{+/-}$ phenotype.

Strains	Avg. Shoot Dry Weight (g)
Uninoculated	0.048 ± 0.005
SmP110	0.451 ± 0.012
SMa0719	0.231 ± 0.015
SMb20492	0.045 ± 0.008
SMb20493	0.243 ± 0.014
SMb20210	0.048 ± 0.004
SMb20214	0.287 ± 0.011
SMc00005	0.045 ± 0.005
SMc00326	0.052 ± 0.005
SMc02486	0.055 ± 0.009

4.6 Nodule occupancy of the SDR mutant strains:

A reason for lack of nitrogen fixation by the SDR mutants could have been due to the fact that the infection was blocked, resulting in empty nodules (devoid of bacteroids) or that the bacteria could infect, differentiate and form nodules but could not fix nitrogen. Two approaches were taken to answer this, one was to recover bacteria from nodules by plating crushed nodules and the second approach was microsopy (confocal laser scanning and light microscopy).

To isolate bacteria from nodules, the nodules were removed from roots, surface sterilized, crushed aseptically in TY medium and then plated onto TY agar plates. These plates were incubated at 30°C for 2-3 days and the colonies counted. This was done for 3 nodules per mutant. Nodules formed by SmP110 and most other mutant strains with a Fix^+ phenotype were healthy and filled with bacteroids and yielded greater than 250 CFU per nodule.

The CFU per nodule for the mutant SMc02486 was similar to SmP110, indicating that even though the mutant cannot fix nitrogen it is able to infect the plant host and form nodules filled with bacteria. Contrary to this, SMc01204, SMc02356, SMc02041 and SMb20750 mutants though forming pink nitrogen fixing nodules, did not yield viable bacteria (Fig. 4-9). From our phenotypic screens it was determined that SMc01204, SMc02356 and SMc02041 mutants are unable to utilize ribonic-γ-lactone as the sole carbon source. SMb20750 is unable to utilize D-fucose, glucose -1-phosphate, leucine and ornithine. We hypothesize that these genes might play a role in the ability of the bacteroids to repopulate the soil. Further experiments need to be done to

confirm this conclusion. SMc00326 mutants formed about 1 or 2 stumpy irregular shaped nodules per plant. No viable bacteria could be isolated from these nodules.

To ascertain that the nodules contained bacteroids confocal microscopy and light microscopy was done on nodules infected by mutants showing a symbiotic phenotype. The microscopic analysis showed nodules formed by SMc02486, SMc01204, SMc02356, SMc02041 and SMb20750 mutants were filled with bacteroids. Though SMc02486 mutant was able to carry out successful infection of the plant and form nodules, mutation in this gene rendered the organism incapable to fix nitrogen (Fig.4.10., Fig.4.11.). In case of SMc01204, SMc02041, SMc02356 and SMb20750 mutants though these strains were able to infect the plant and fix nitrogen, they somehow could not be isolated from the nodule. Sections from nodules formed by SMc00326 were full of starch and showed no bacteria within them (Fig.4.10.).

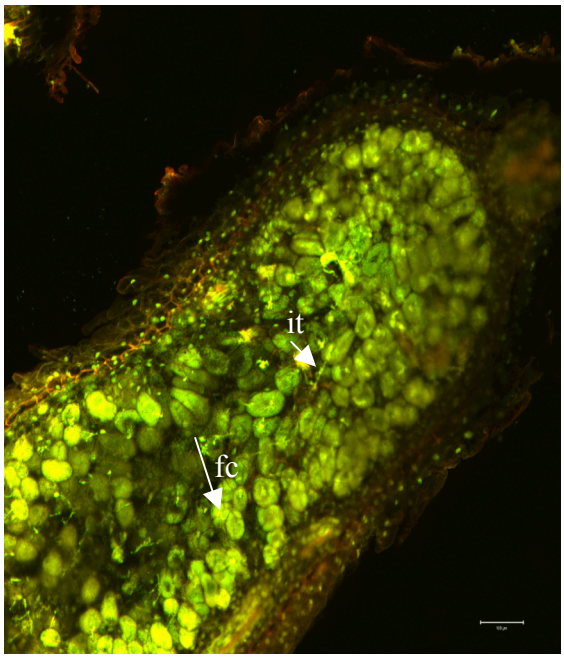
Fig.4.10. Confocal laser scanning microscopy:

Nodules infected by SDR mutant strains were stained by SYBR and observed under Zeiss confocal microscope. SYBR stains nucleic acids resulting in green fluorescence. The size bar is 10 μ

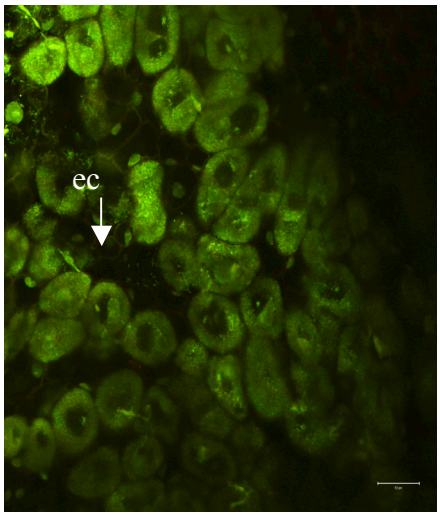
A. Wild type strain SmP110 nodule: The entire nodule in shown (100 μm scale). The cells filled with bacteria (fc) fluoresce bright green. Infection threads (it) are visible throughout the nodule.

Aa. Section of the wild type SmP110 nodule (50 μm scale): Infected cells and empty plant cells (ec) can be distinctly visualized. The empty plant cell nucleus stains green.

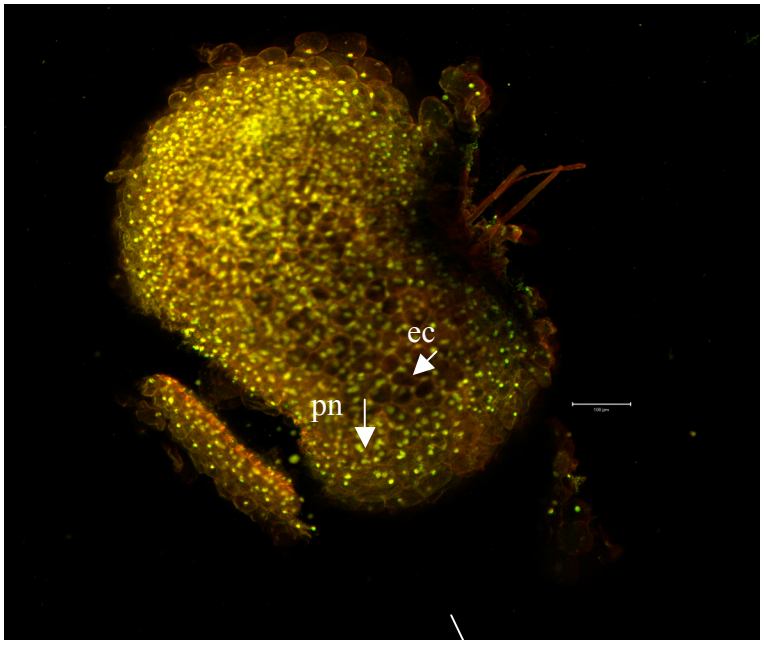
B. Nodule formed by Rm7055 (*exoY* mutant, 100 μm scale): This strain produces ineffective nodules. The nodules are empty (ec) and only the plant nucleus (pn) stains green.



A



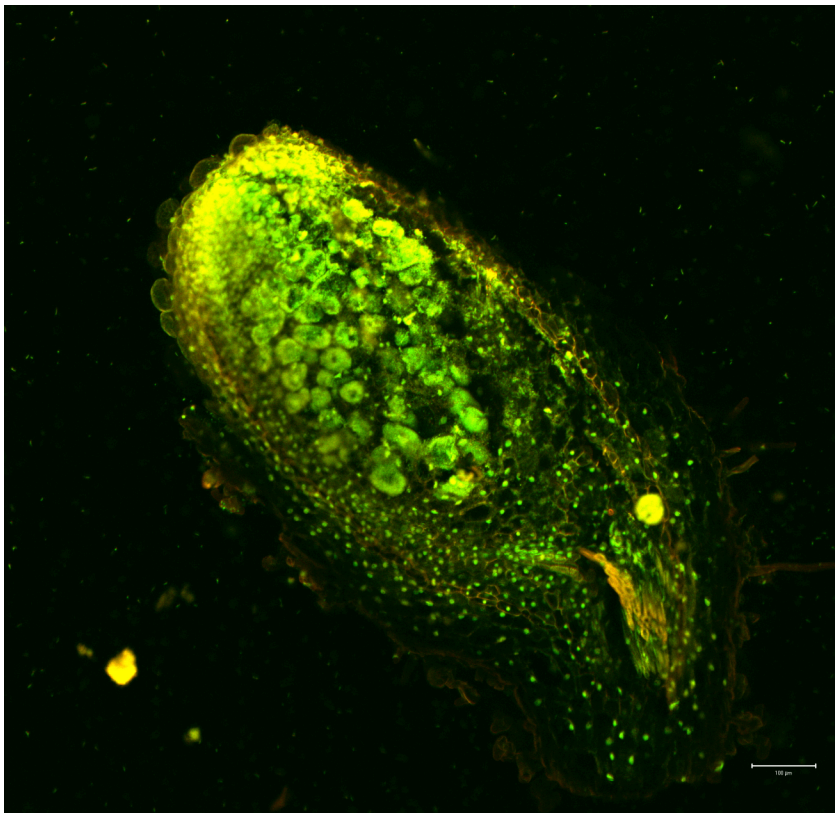
Aa



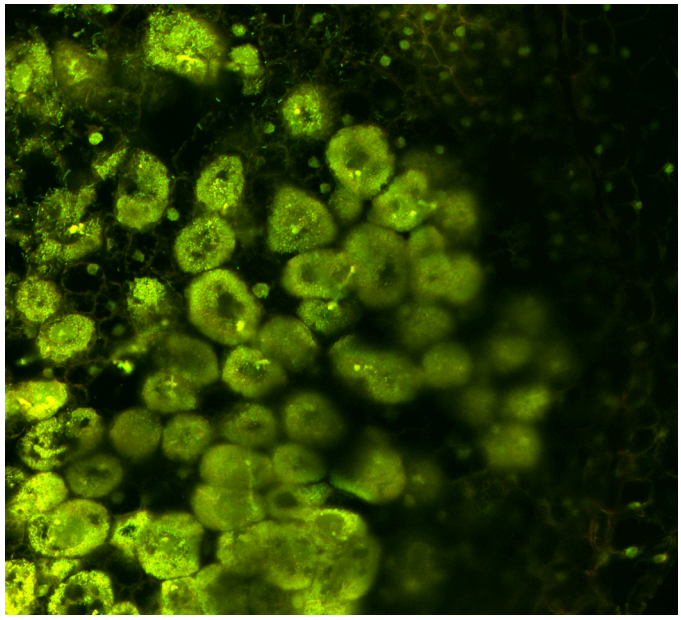
B

C. Nodule formed by SMc02486fg SDR mutant strain (100 μm scale): The nodule is filled with bacteria even though this strain is unable to fix nitrogen.

Cc. Section of the nodule induced by SMc02486fg (50 μm scale).



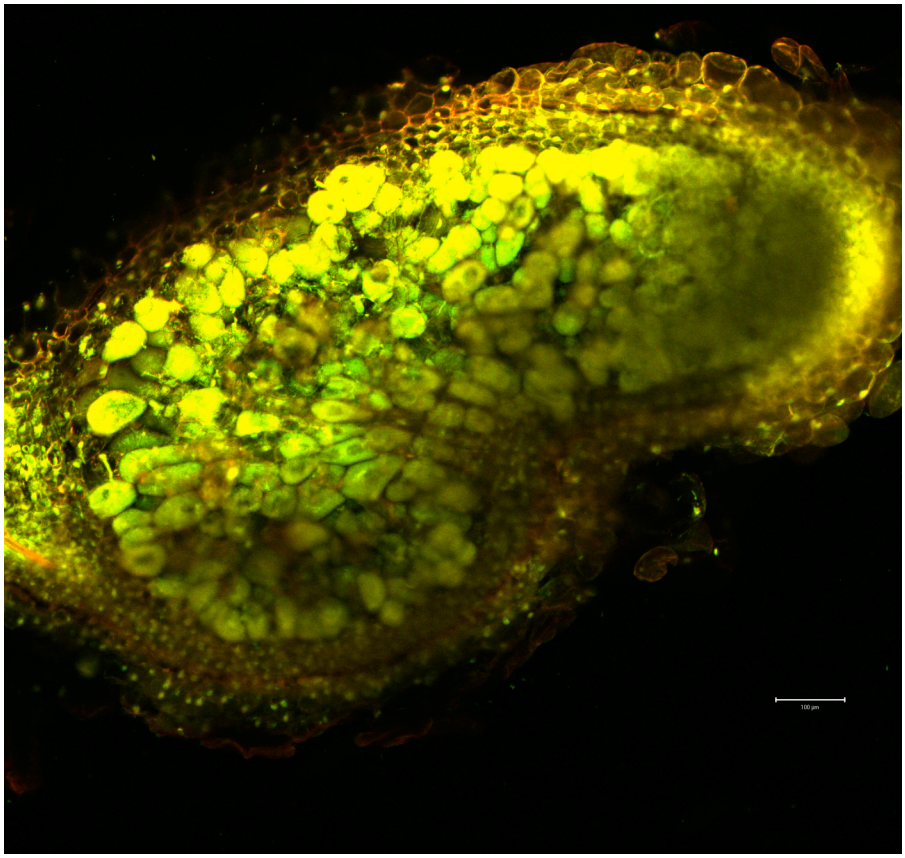
C



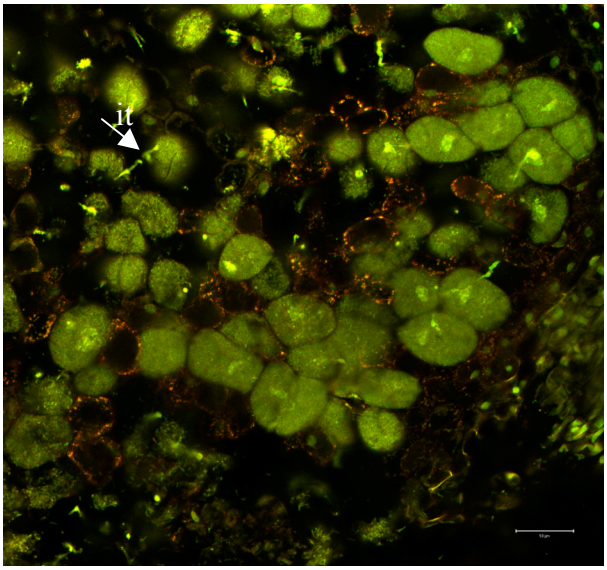
Cc

D. Nodule formed by SMc01204fg SDR mutant strain (100 μm scale): The nodule is filled with bacteria even though none could be isolated on crushing the nodules.

Dd. Section of a nodule induced formed by SMc01204fg (50 μm scale): Infection threads are visible and cells are filled with bacteria.



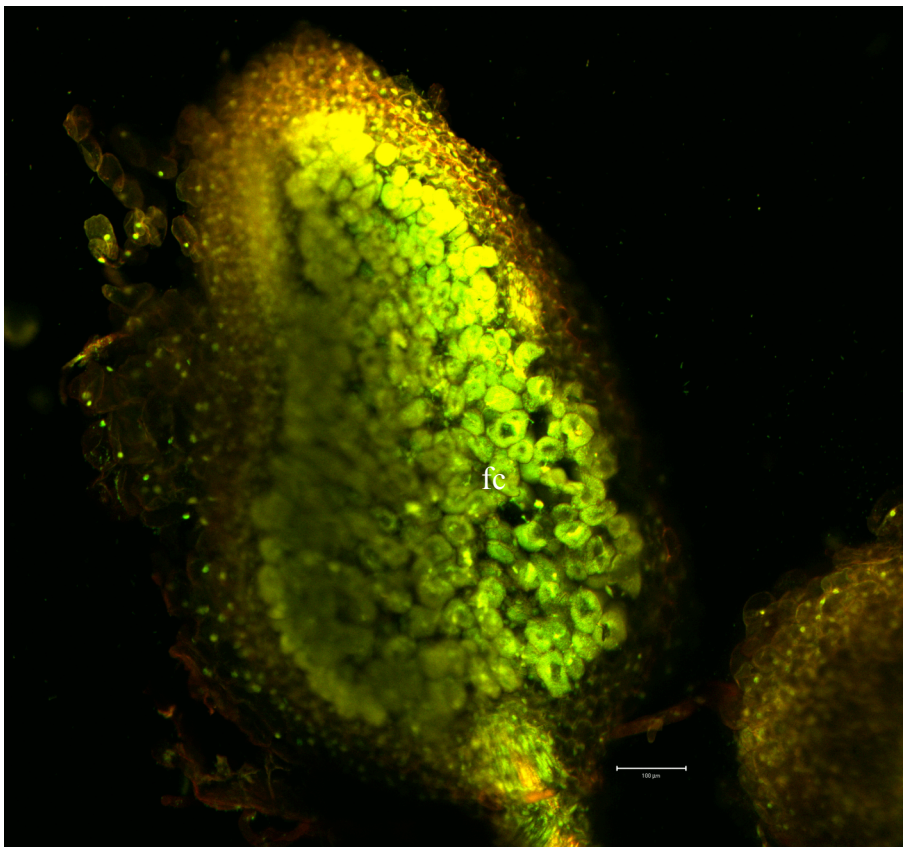
D



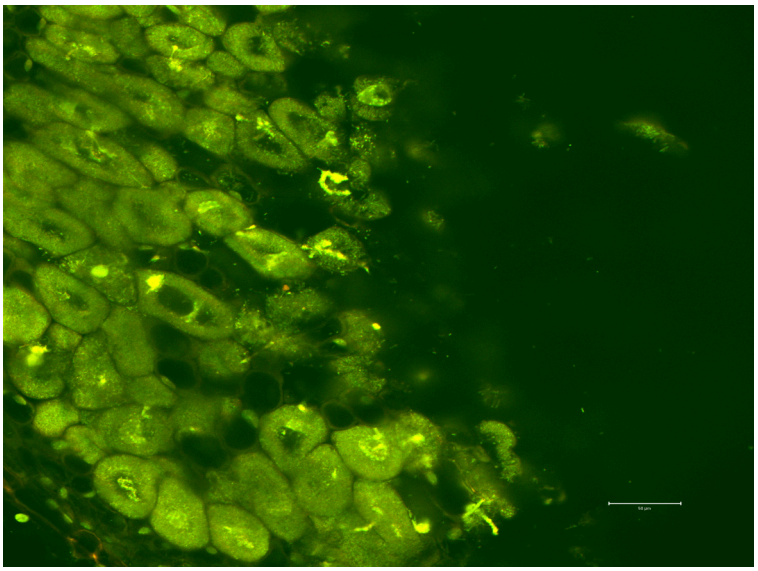
Dd

E. Nodules formed by SMc02356fg mutant strain (100 μm scale): The cells are filled with bacteria though no bacteria could be recovered from nodules formed by this strain.

Ee. Section of the nodule formed by SMc02356fg mutant strain (50 μm scale).



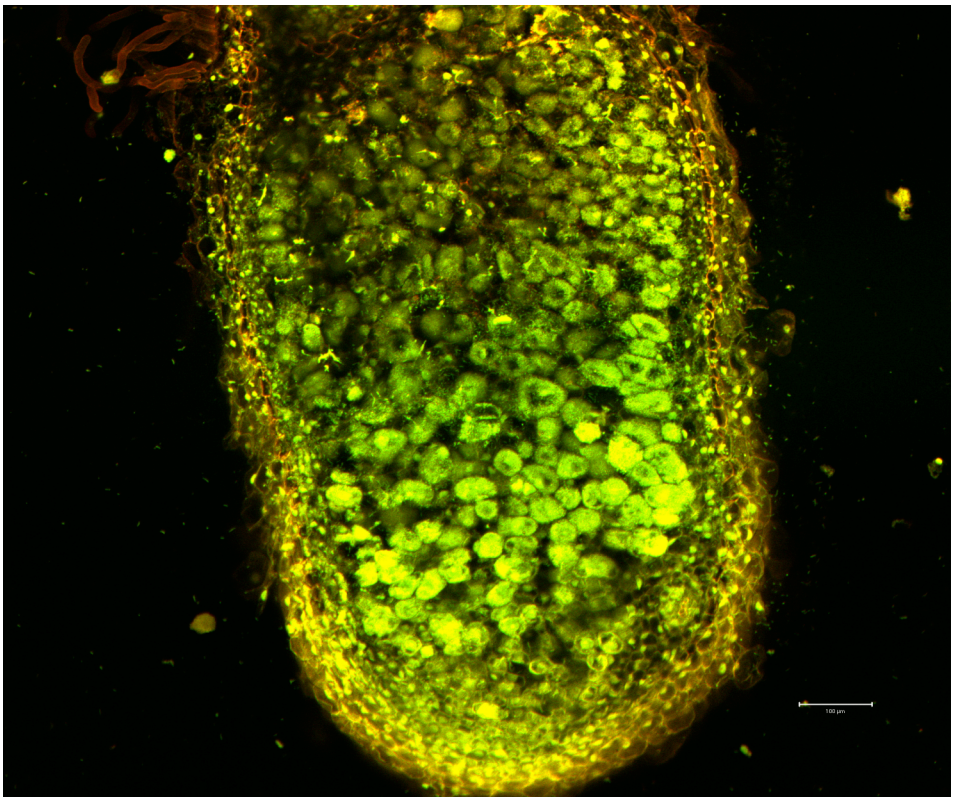
E



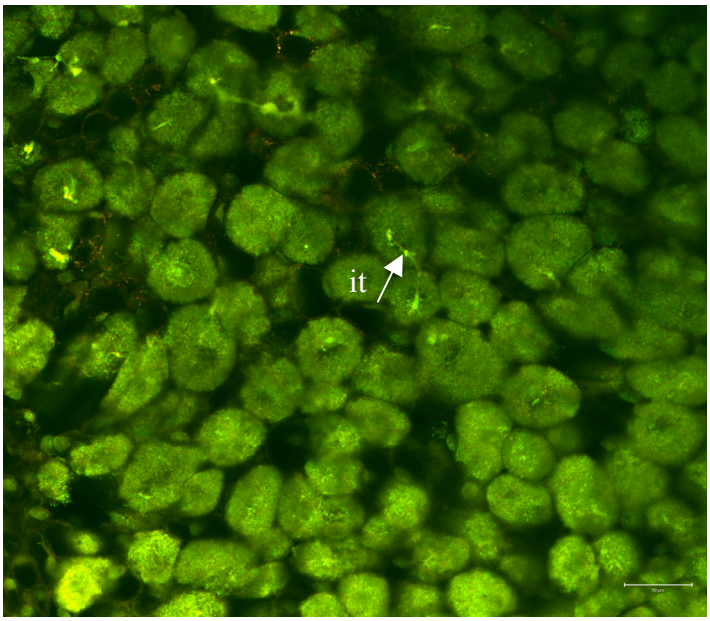
Ee

F. Nodule formed by SMb20750fg mutant strain (100 μm scale): The nodule is filled with bacteria. However no bacteria could be recovered from the nodules.

Ff. Section of the nodule formed by SMb20750fg SDR mutant strain (50 μm scale): Infection threads are clearly visible.



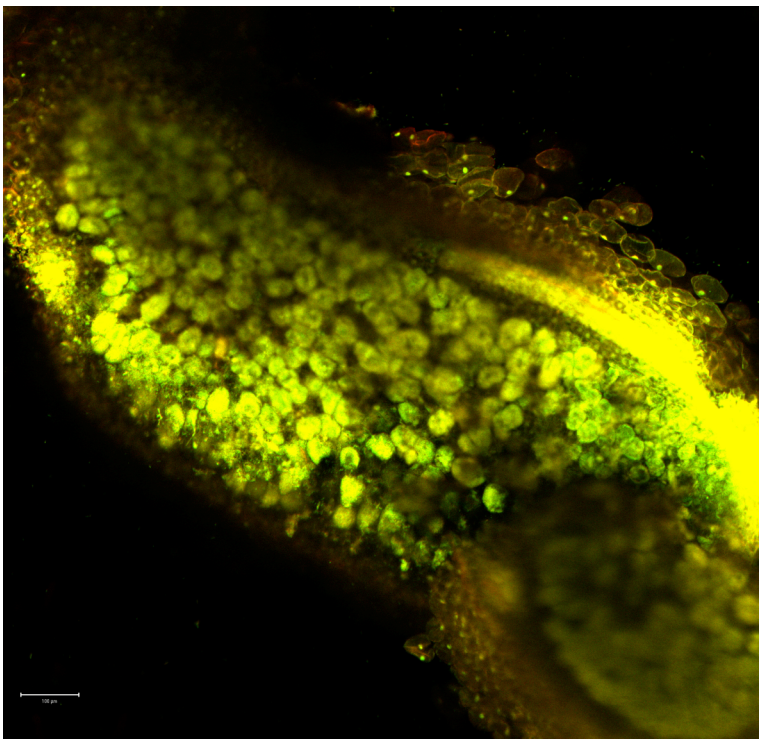
F



Ff

G. Nodule formed by SMc00778fl mutant strain (100 μm scale): Even though the strain appears to be a proline auxotroph, is able to infect the plant and form nitrogen fixing nodules.

Gg. Section of the nodules formed by SMc00778fl (50 μm scale).

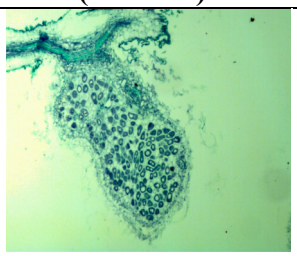
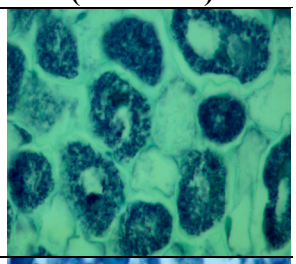
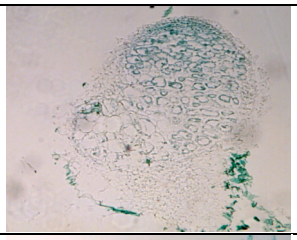
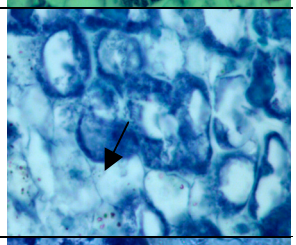
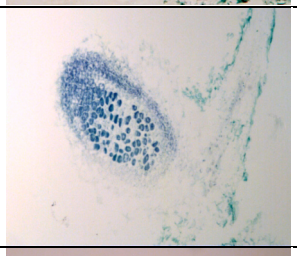
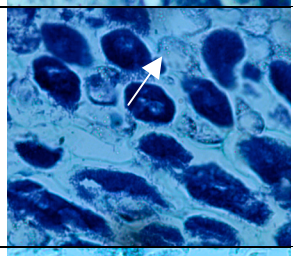
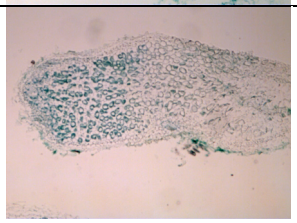
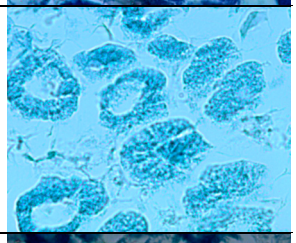
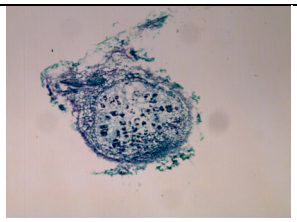
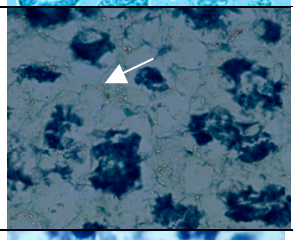
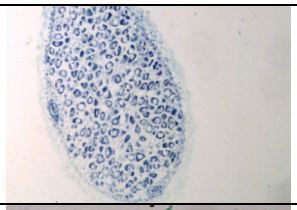
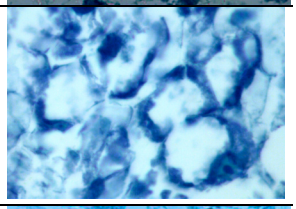
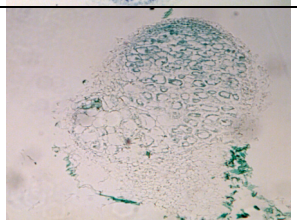
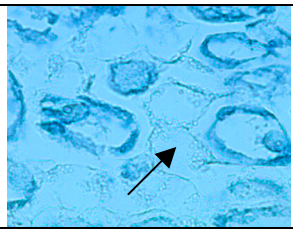


G

Gg

Fig.4.11. Light microscopy of nodules formed by SDR mutant strains showing a symbiotic phenotype:

The nodules sections were stained with toluidine blue at an acidic pH. This stains nucleic acid. The bacterial cells and the plant nucleic acid stain blue. In most Fix^- strains starch granules are visible. The arrows show accumulated starch granules.

SDR Mutant Strain	Phenotype	Photograph (40x TM)	Photograph (400x TM)
SmP110 (WildType Strain)	Fix^{+}		
SMc00005	Fix^{-}		
SMb20493	$Fix^{+/-}$		
SMa0719	$Fix^{+/-}$		
SMc02486	Fix^{-}		
SMb20210	Fix^{-}		
SMc00326	Fix^{-}		

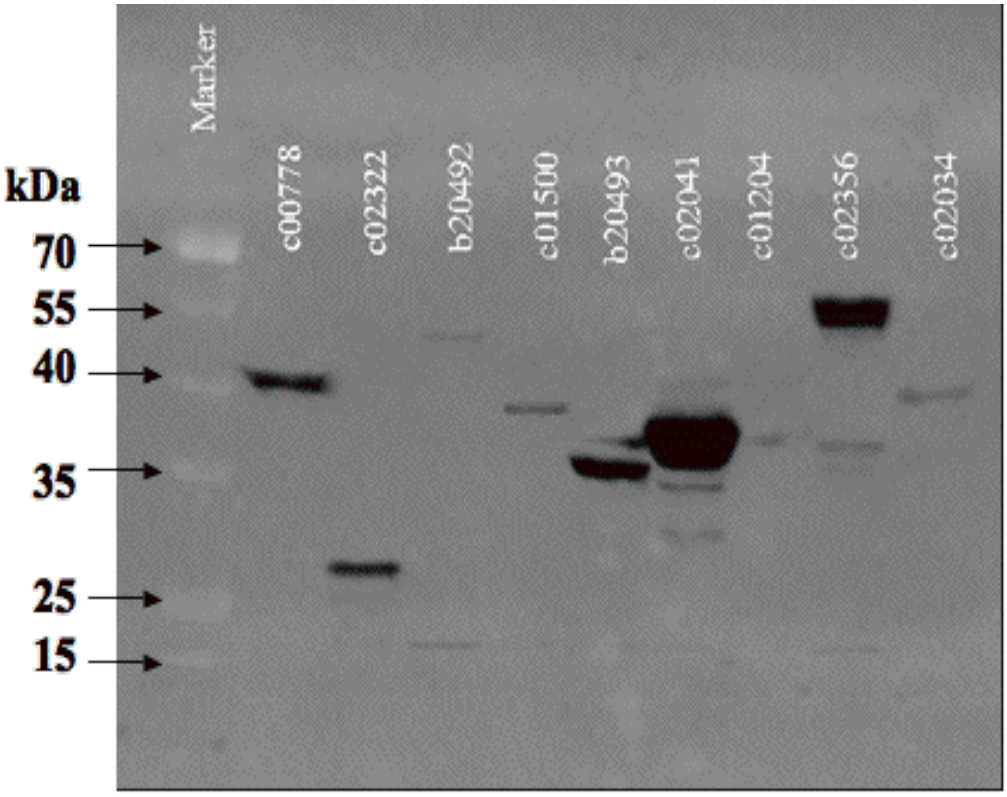
4.7 Overexpression of SDRs showing a carbon utilization phenotype:

SDR's were overexpressed to detect their size and to aid in subsequent purification of the protein for further characterization. For this purpose, complete ORF sequence of SMc00778, SMc01204, SMc02034, SMc02322, SMc02356 and SMb20076 (T7 orientation) from pGEM-Teasy were cloned as Not1 fragment into dephosphorylated NotI digested pET30a. Similarly, ORFs of SMb20492, SMb20493, SMb21159, SMc01500 and SMc2041 (SP6 orientation) were cloned as Not1 fragments from pGEM T-easy into dephosphorylated NotI digested pET30b. The proteins generated carry an N-terminal His-tag. The induction of the protein in *E.coli* BL21 background was carried out in autoinduction medium (Chapter 2) and protein expression was confirmed by Western blot with detection facilitated by the His-tag (Fig.4.12.).

The molecular weight of each of the proteins corresponded to the predicted except for SMc02322. The size ranged from approximately 38kD to about 49kD. The expected size of the SDR encoded by SMc02322 was 83kD, however that observed was about 29kD. Autocleavage or cleavage by endogenous cellular proteases of the SMc02322 protein could be the reason the observed band of low molecular weight was observed. The fragment observed must be the N-terminal sequence.

Fig.4.12. Western blot detecting over-expressed SDRs:

The blot was detected using an anti-His tag antibody. The bands for all, except SMc02322, are of the expected molecular weight, indicating that the desired protein is being expressed.


kDa
70
55
40
35
25
15
Marker
c00778
c02322
b20492
c01500
b20493
c02041
c01204
c02356
c02034

4.8 Assaying the activity of SDR

Once a carbon utilization phenotype was identified for an SDR, the next step was to determine if the substrate or one of its intermediate is the direct target for this protein i.e. the goal is to identify the exact substrate for the SDR. To address this cell extracts of wild type SmP110 and SDR mutant strains (SMc01500, SMc02322, SMc02486, SMb20492, SMb20493 and SMb21159) were run on native polyacrylamide gel. The gels were then stained for activity using either NAD or NADP as a cofactor and the potential carbon substrate. In all cases activity was observed when NAD was used as a cofactor and not NADP The presence of an activity band in the wild type and its absence in the mutant strain would indicate that the mutant was deficient in the enzyme required for the catabolism of that substrate (Fig.4.13.ii – Fig. 4.13.vi). As a control all the extracts were assayed for β-hydroxybutyrate dehydrogenase activity (Fig.4.13.i). This indicated whether the proteins in the cell extract were active and also gives an indication of equal gel loading.

Extracts obtained from SMc01500 mutant strain did not show any activity with D-sorbitol, however the wild type cell extract showed a clear band. Similarly, no activity was seen for extracts from SMb20492fg, SMb20493fl, SMb21159fl and SMc02486fg with sodium succinate, even though the wild type showed strong activity in the presence of this substrate and NAD. Where sodium succinate was used as a substrate, activity band obtained for the wild type was the same in all four cases. Whether these four SDRs have a common function needs to be further investigated. In case of SMc02322, a very faint

band was observed in the wild type that is absent in the mutant in the presence of L-rhamnose (data not shown).

Fig.4.13. Activity staining gels for SDR mutant strains

A. β- hydroxybutyrate dehydrogenase staining of cell extracts with NAD as cofactor

B. Activity staining for SMc01500 with D-sorbitol + NAD

C. Activity staining for SMb20492 with Na-succinate + NAD

D. Activity staining for SMb20493 with Na-succinate + NAD

E. Activity staining for SMb21159 with Na-succinate + NAD

F. Activity staining for SMc02486 with Na-succinate + NAD


SmP110
SMc02486
SMb21159
SMb20492
SMb20493
SMc01500

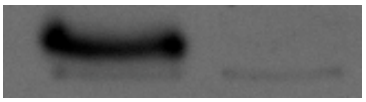
A


SmP110
SMc01500fl

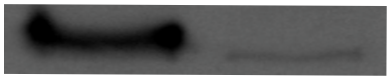
B


SmP110
SMb20492fg

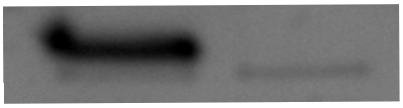
C


SmP110
SMb20493fl

D


SmP110
SMb21159fl

E


SmP110
SMc02486fg

F

CHAPTER 5

Discussion

In this study 78 SDR encoding genes were identified and 77 of these were mutated and further investigated for a phenotype under free-living and symbiotic conditions. To the best of our knowledge this is the first report of functional analysis of the full complement of SDRs in any organism. This study shows that this group of proteins plays an essential role in the free-living and symbiotic lifestyle of *S. meliloti*.

An interesting observation was that 85% of the SDRs were either flanked by or were in close proximity to genes predicted to encode ABC transporter (ATP- binding cassette transporters) or TRAP transporters (tripartite ATP-independent periplasmic transporters). In bacterial systems, both these types of transporters aid in the uptake of substrates including carbohydrates, proteins, amino acids and peptides (Davidson and Chen 2004; Kelly and Thomas 2001). The proximity of SDR encoding genes to the ABC and TRAP transporters would imply that these SDRs are involved in the catabolism of solutes transported by these systems. A recent study (Mauchline *et al.* 2006) on solute binding transporters in *S. meliloti* supports the carbon utilization phenotype of SMb20409 SDR mutant strain. ABC-T SMb20410 is induced by galactitol and D-mannose. Mutation in the SDR encoding gene SMb20409, adjacent to this transporter also resulted in poor growth on these carbon sources. For most of the other SDRs with a carbon utilization phenotype, the transporters did not show any specific inducers. In this case, whether the substrate acted

upon by the SDR is actually transported by the proximal transporters needs to be addressed. Moreover, it has been shown that ABC transporters may transport more than one solute; consequently another transporter could transport the substrate acted upon by an SDR.

Genetic context can also provide important insights into possible function of a gene. If the SDRs are in an operon then there is a strong probability that it will be involved in one of the reactions carried out by this operon. For example, SMc01500 is in an operon involved in the metabolism of sugar alcohols. Mutants of this gene are unable to utilize sugar alcohols as sole carbon source. Similarly SMc00326 (*fabI2*) is in an operon with *fabB* and *fabA*, and SMc00572 (*fabG*) seems to be in an operon with *fabD*. The disruption of *fabI2* could lead to lesions in the fatty acid biosynthesis pathway. As fatty acids are important constituents of Nod factors and N-acyl homoserine lactones and are of great significance in the physiology and functioning of the bacterial system, it is not surprising that the *fabI2* mutant is symbiotically challenged. Another good example of this is SMc02486. It seems to be a part of an operon encoding TCA cycle enzymes. Like most of its counterparts, mutation in this gene results in poor growth and a symbiotically deficient phenotype. This being said, there seem to be exceptions to this rule. SMb20409 is in an operon annotated to be involved in hydantoin metabolism, SMb21111 seems to be a part of an mixed function operon involved in degradation of cyclic aromatic compounds like benzoate/toluene/xylene and amino acid tyrosine, SMb20750 is adjacent to a gene annotated to be a D-mannonate oxidoreductase and

SMc02356 seems to be a part of an operon involved in aspartate metabolism. However mutation of these SDRs did not give an obvious phenotype based on the operon context. Whether the phenotypes of these SDR encoding genes are a result of downstream effects is something worth considering.

5.1 SDRs with characterized orthologs in other organisms

SMc01500, annotated as *smoS* based on sequence homology to *E. coli* sorbitol dehydrogenase, is probably involved in the oxidation of sugar alcohols to their respective ketones. Enzyme assay with whole cell extracts of mutant strains shows the inability of the mutant to catabolize the sugar alcohol tested (Fig. 4-12B). An ABC transporter SMc01496, upstream of SMc01500, is induced by sugar alcohols (Mauchline *et al.* 2006).

Studies on the rhamnose operon in *R. leguminosarum* have led to the speculation that the substrate for RhaD (rhamnose dehydrogenase) in this system is probably phosphorylated rhamnose (Richardson *et al*. 2004). This study showed that phosphorylation of rhamnose occurs prior to isomerization or oxidation. Mutation of its ortholog SMc02322 in *S. meliloti* renders the strain incapable of utilizing rhamnose as the sole carbon source. Moreover, gene expression of the transporter system SMc02324 is strongly induced by rhamnose and erythritol (Mauchline *et al.* 2006). Whether the substrate for SMc02322 is the same as that of its counterpart in *R. leguminosarum* still needs to be ascertained. The genetic context of *rhaD* in *R. leguminosarum* and SMc02322 is very similar, making it a strong possibility that phosphorylated rhamnose is the substrate for this enzyme in *S. meliloti*.

5.2 SDRs involved in the catabolism of amino acids and sugars

In this study 5 SDRs SMb20073, SMb20076, SMb20492, SMb20493 and SMb20750, on the pSymB megaplasmid were found to be unable to use one or more of the following amino acids, leucine, lysine, proline, ornithine and carbohydrate D-fucose and/or succinate, monomethyl succinate and glucose 1-phosphate. Two of these SMb20492 and SMb20493, show a symbiotic phenotype. The hypothesis made in the analysis is that all these SDRs are probably linked biochemically via common intermediate/s and disruption of any of them might result in a block in the metabolic pathway resulting in the lack of that key intermediate. Similar genetic and biochemical studies have led to better understanding and elucidation of the several critical pathways involved in substrate utilization and symbiosis in *S. meliloti* like the pathway for glutamate degradation (Fitzmaurice *et al.* 1991), the PHB cycle (Cai *et al.* 2000), TCA cycle (Driscoll *et al.* 1990, Mortimer *et al.* 1999, Finan et al. 1988, Duncan *et al.* 1979, Gardiol *et al.* 1982) and the β-ketoadipate pathway in *S. meliloti* (MacLean *et al.* 2006). Studies on *S. meliloti* auxotrophic mutants of leucine, isoleucine-valine have also aided in better understanding of the amino acid degradation pathway (Hassani *et al.* 2002). *S. meliloti* strains showing auxotrophy for leucine have been reported to be Fix^- and supplementation of leucine did not restore this phenotype. Microscopic analysis has shown that nodules formed by the leucine auxotrophs are underdeveloped (Hassani *et al.* 2002, Dickstein *et al.* 1991, Kerppola and Kahn 1988).

Figure 5.1. shows the link between the catabolism of the amino acids L-proline, L-leucine, L-lysine and ornithine. In the cases of SMb20073, SMb20076, SMb20492, SMb20493 and SMb20750 the key intermediate seems to be glutarate semialdehyde that is subsequently converted to acetyl CoA. These dehydrogenases could be involved in the conversion of glutarate semialdehyde to glutarate or in the subsequent conversion of glutaryl CoA to crotonyl CoA or 3-but-2-enoyl CoA. As the SMb20750 mutant is unable to catabolize ornithine it could also be involved in the interconversion of glutamate-L-semialdehyde to L-1-pyrroline-5-carboxylate. Enzyme assays using either glutarate semialdehyde, glutaryl-CoA or glutamate-L-semialdehyde would give conclusive evidence if these genes are involved in the said pathway. An interesting observation along with the inability of these mutant strains to use amino acids is their inability to utilize sugars like D-fucose and/or succinate, monomethyl succinate and glucose-1-phosphate. Fucose has been shown to play an important role in communication of *Bacteroides thetaiotaomicron*, a normal inhabitant of the distal small intestine with intestinal epithelial cells of the human gut, forging a mutually beneficial relationship (Hooper *et al.* 1999). Disruption in genes required for the catabolism of this substrate by the bacteria results in inefficient colonization.

Upstream of SMb20492 and SMb20493 are genes involved in sugar uptake and fucose metabolism. Activity assays using succinate as the substrate confirmed that these mutant strains are unable to catabolize this substrate (Fig 4-12C, Fig. 4-12D). Whether carbon utilization phenotypes of SMb20073,

SMb20076, SMb20492, SMb20493 and SMb20750; and symbiotic phenotypes of SMb20492 and SMb20493 observed are due to the inability to utilize the carbon source per se or due to a down regulation of the transporters for these sugars or an indirect regulatory effect needs to be further assessed. Another rationale for the observed phenotype in the above mentioned SDRs could be that as they are involved in the metabolism of sugars like fucose and succinate, disruption of these genes would affect the carbon flux as observed in *R. leguminosarum* (Walshaw *et al.* 1997) and inhibit the uptake of amino acids leucine, lysine, proline and ornithine. Amino acid uptake via the Aap of *R. leguminosarum* was strongly inhibited in *suc* mutants (TCA cycle mutants) and these strains are Fix^-, even though the transcription level of the *aap* operon was the same as the wild type (Walshaw *et al.* 1997). A similar regulatory effect might be occurring when these SDR encoding genes are mutated.

A key observation was that even though the SMb20750 mutant formed nitrogen-fixing nodules, no bacteria could be isolated from these nodules. Whether the SDR activity encoded by this gene is essential for the switch from symbiotic to free-living lifestyle is worth exploring. Another key point to note is that RmF680, an Rm1021 pSymB deletion strain in which SMb20492 and SMb20493 are deleted, does not grow on D-fucose. Similarly deletion strain RmG373 in which SMb20750 is deleted is unable to grow on ornithine and leucine (Chapter 3). However, deletion strains of pSymB including SMb20492 and SMb20493 did not show the same phenotype on plants as this study i.e Nod^- phenotype for SMb20492 mutant and intermediate phenotype for

SMb20493 mutant (Charles *et al.* 1991). This could be because deletion of a single gene in a pathway could trigger a cascade of events that leads to ineffective symbiosis but deletion of genes encoding entire pathways might not.

5.3 SDR with a probable role in the TCA cycle

An operating TCA cycle is essential for beneficial host – microbe interaction. Mutants of the TCA cycle enzymes of *S. meliloti* are unable to utilize a wide variety of carbon sources. Citrate synthase mutants of *S. meliloti* are unable to grow on succinate, malate, pyruvate and glucose. These mutants required arabinose and glutamate for growth (Mortimer *et al.* 1999) and are ineffective in symbiosis as are mutants of succinyl-CoA synthetase and 2-oxoglutarate dehydrogenase (Mortimer *et al.* 1999, Walshaw *et al.* 1997, Dymov *et al.* 2004). Succinate dehydrogenase mutants were unable to catabolize pyruvate, succinate, acetate, glutamate or arabinose but could grow on glucose, fructose and sucrose (Gardiol *et al.* 1982, Finan *et al.* 1988). C4-dicarboxylates such as malate, TCA cycle intermediates are essential for effective symbiosis, as they are utilized by the nitrogen fixing bacteroids as energy source. Mutants of *S. meliloti* unable to transport these dicarboxylates are Fix$^-$. Rhizobial strains unable to utilize succinate as carbon source are compromised in symbiosis (Gardiol *et al.* 1982, Finan *et al.* 1988). Mutation in malate dehydrogenase encoding gene, another TCA cycle enzyme, resulted in poor growth on glutamate and arabinose and resulted in a Fix$^-$ symbiotic phenotype. This strain also showed poor growth on glucose and succinate but

the growth phenotype was better than that observed on glutamate and arabinose (Driscoll and Finan 1993, Dymov *et al.* 2004).

SMc02486 lies in the region on the chromosome carrying genes encoding the TCA cycle enzymes malate dehydrogenase (*mdh*), succinyl-CoA synthetase (*sucCD*) and 2-oxoglutarate dehydrogenase (*sucAB* and *lpdA*) previously identified by Dymov *et al*. and is upstream of *lpdA* in the *lpdA-sucBA-sucDC-mdh* operon. This mutant strain grows very slowly on complex and minimal media and the colonies formed are smaller than the wild type. It shows poor growth on most carbon sources. This would indicate that this enzyme is a key component of the TCA cycle. Activity assay using succinate also indicates that this mutant does not have the capacity to catabolize this substrate. In this case, it seems more like a regulatory effect rather than succinate being the direct substrate for this enzyme. Further characterization would be required to ascertain whether a reaction in the TCA cycle is affected by this dehydrogenase. Moreover these mutants were able to form nodules on alfalfa but unable to fix nitrogen. The nodules formed were white. Microscopic data show that the nodules contain bacteroids indicating successful infection. These mutants seem to be unable to effectively fix nitrogen.

5.4 SDR with a probable role in the carbon fixation

Calvin cycle (reductive pentose pathway) is an essential pathway for carbon fixation in several organisms, which ultimately leads to the synthesis of carbohydrates required by the organism to grow. SMb20210 lies in the region of pSymB megaplasmid annotated to be involved in this pathway. Genes involved

in pyroloquinoline quinone synthesis (PQQ), an essential cofactor for several dehydrogenases that are involved in key metabolic pathways are also in close proximity to this gene. The high probability that SMb20210 encodes a key enzyme in either pathway may be the reason for the poor growth of this mutant strain on most of the carbon sources tested. Also, disruption of genes like SMb20210 may result in a deficiency of nutrients required for effective symbiosis, leading to a Fix^- phenotype. Previous studies on large-scale deletion of pSymB megaplasmid including this gene did not show similar carbon utilization or symbiotic phenotype (Charles *et al.* 1991). A possible explanation could be that when large regions comprising of genes encoding for entire pathways are deleted no deleterious effects are observed. However the deletion of a single gene constituting the pathway may lead to an effect similar to that observed in toxin-antitoxin systems.

5.5 SDRs involved in monomethyl succinate and ribonic-γ-lactone utilization

Thus far to the best of our knowledge there is no literature on the catabolism of monomethyl succinate and ribonic-γ-lactone in rhizobial systems. SMc02041 and SMc02356 may be involved at some point in the conversion of the methyl ester of succinic acid to glucose or its conversion into succinate to enter the TCA cycle. Hexoses such as galactose (KEGG pathway 00052) and glucose (KEGG pathway 00030) and methyl pentose like rhamnose and pentose like xylose (KEGG pathways 00040 and 00051) are converted to their corresponding lactones by the activity of dehydrogenases. These lactones are

further catabolised by the action of lactonases and other enzymes into acetyl CoA , which can then enter the Krebs cycle. Whether SMc01204, SMc02034 and SMc02356 are dehydrogenases that aid in the lactone ring cleavage remains to be investigated.

Furthermore, even though SMc01204, SMc02034 and SMc02356 mutant strains could fix nitrogen, no bacteria could be isolated from their nodules. Microscopic analyses of the nodules formed showed that the nodules are filled with bacteria, indicating successful infection. Whether the common lack of ability to catabolize ribonic-γ-lactone affects the ability of the strain to repopulate the soil is something that needs to be further addressed.

5.6 SDRs involved in fatty acid biosynthesis

Disruption of the fatty acid biosynthesis gene *fabI* (enoyl ACP-reductase) involved in the first reductive step in fatty acid elongation, affected the cell wall composition of *Mycobacterium tuberculosis* strains and conferred a temperature sensitive phenotype to *E. coli* strains (Neidhardt *et al.* 1996), indicating that this is an essential gene. Fatty acids are also important constitutents of Nod factors and N-acyl homoserine lactones, essential for host-bacterial signaling during symbiosis in rhizobia (Lopez-Lara and Geiger 2001). The inability of SMc00005 (*fabI1*) and SMc00326 (*fabI2*) mutants to fix nitrogen could be attributed to changes in cell wall composition and/or permeability or to the fact that they are unable to synthesize appropriate Nod factors. Nodules formed by the SMc00326 mutant strain appear to be devoid of bacteria and those formed by the SMc00005 mutants have very few bacteria in

them. This is indicative of the fact that these mutants are not able to carry out effective infection.

Of the 78 SDRs identified in this study, only one, encoding *fabG* was recalcitrant to mutation. FabG reduces the β-ketoester from the condensation of malonyl-CoA with a growing acyl chain (Cronan & Rock, 1996). In *E. coli* and *Salmonella enterica* this gene has been shown to be essential for viability, with only temperature-sensitive mutants having been isolated (Lai & Cronan, 2004; Zhang & Cronan, 1998). NodG, a 3-oxoacyl-acyl carrier protein reductase has the same biochemical function as FabG, and the high degree of similarity at the protein and DNA level suggest that *nodG* is a duplication of the housekeeping gene *fabG*. *nodG* has been well characterized and in our study did not show a carbon utilization phenotype or symbiotic phenotype consistent with previous studies (Lopez-Lara and Gieger et al. 2001., Cloutier et al. 1997). It has been reported that NodG can be functionally complemented by FabG. Though *fabG* could not be mutated in this study, a transposon insertion mutant of this gene was isolated during the study of salt-sensitive mutants of *S. meliloti* 1021 (Miller-Williams *et al.* 2006). This mutant is compromised in nodule occupancy. Whether the lethality of *fabG* mutation depends on the background strain used or the conditions used for selection is worth exploring.

5.7 SDRs with an intermediate symbiotic phenotype

Three of the SDR mutants, SMa0719, SMb20214 and SMb20493 showed a partial fix phenotype in symbiosis. Additionally, SMa0719 showed poor growth on most substrates tested in this study. The reason for this

phenotype is not clear from its genetic context in the genome. This SDR is in close promimity to genes annotated to be involved in the arginine iminase pathway. Dihydropicolinate synthase, an enzyme unique to the biosynthesis of lysine, is upstream of this gene.

All three mutant strains seemed to be able to fix nitrogen but at lower levels than the wild type strain. Previous reports have shown that *S. meliloti* mutants deficient in the transport of C4-dicarboxylates induce the formation of ineffective nodules, and that that these metabolites are generally considered to be the main energy-bearing compound supplied by the plant to the bacteroids in the nodules (Vasse *et al.* 1990, Yarosh *et al*. 1989). Also, in some biosynthetic pathways, intermediates, rather than end products, may be required for bacteria to elicit nodule development or infection e.g. studies with *S. meliloti* tryptophan auxotrophs suggested that anthranilate synthesis was important for bacteroid development and symbiosis with alfalfa (Barsomian *et al.* 1992). Mutation of genes involved in the arginine deiminase pathway resulted in an intermediate symbiotic phenotype in *R. etli* (Hooghe *et al.* 1997). Genes involved in this pathway are in close proximity to SMa0719. So we speculate that symbiotic deficiency elicited by SMa0719, SMb20214 and SMb20493 mutant strains may be associated with blockage of some key metabolic pathway. However, further investigation of the mechanisms by which these SDR-encoding genes mediate nodulation needs to be addressed.

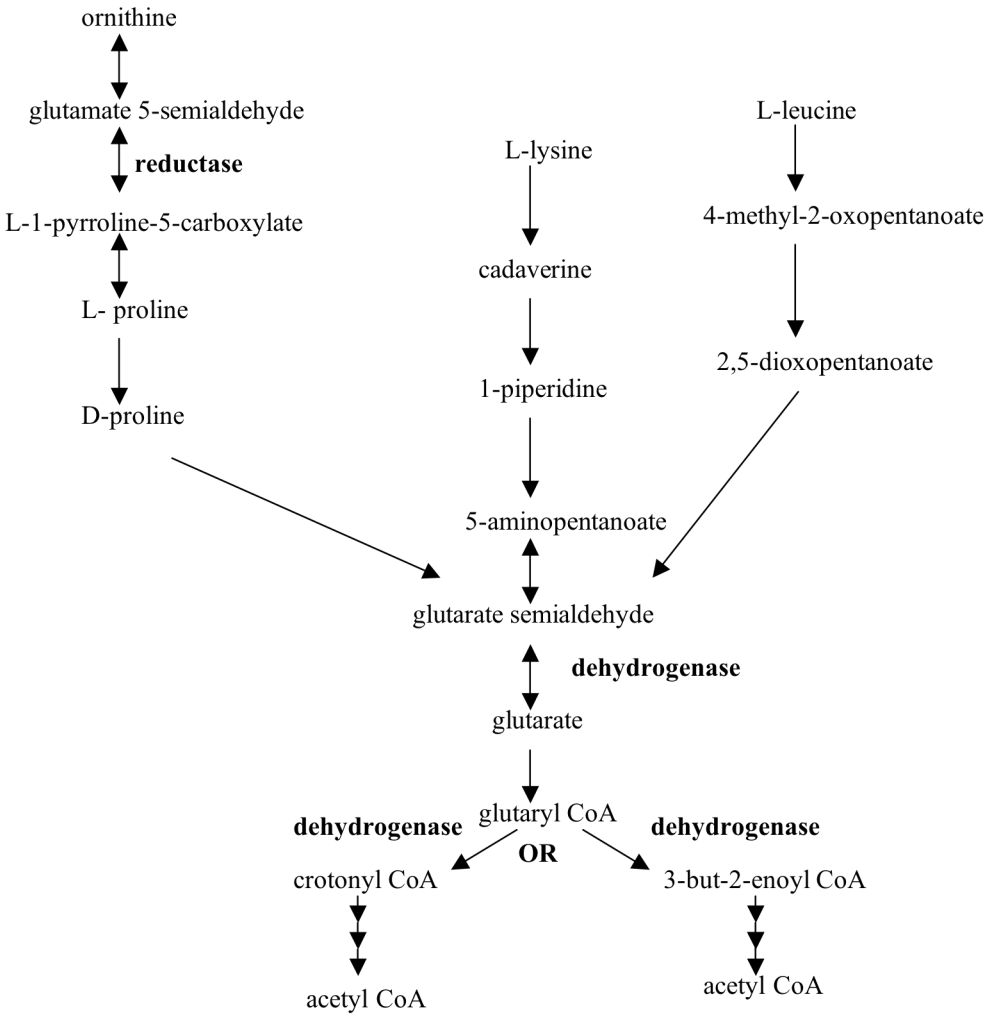
Although of the 77 SDR mutants screened only eight showed a symbiotic phenotype, it is still plausible to assume that these genes might have a role in

providing a competitive advantage to the bacteria by scavenging and metabolizing different nutrients from the soil. It has been shown that deficiency in metabolic pathway can cause a decrease in nodule occupancy, while still forming effective nitrogen fixing nodules (Fry *et al.* 2001). Also, previous work has demonstrated that while rhizobial mutants, unable to utilize different hexoses, formed effective nodules. But hexose import may be critical for competition for nodule occupancy (Glenn *et al.* 1984; Lambert *et al.* 2001). This suggests that further studies on the SDR mutants are necessary to provide an insight into the physiology of nodule development.

SDRs showing a phenotype have been over expressed for purification and characterization with respect to activity and structure. All the SDRs were of the expected size except SMc02322, which was considerably smaller than expected. As protease inhibitors were not added to the cell extracts, it is possible that the protein shows autocleavage or was degraded by proteases. In-gel activity assays showed that the preferred cofactor for reactions dependent on SDRs encoded by SMb20492, SMb20493, SMb21150, SMc01500 and SMc02486 is NAD and not NADP.

In conclusion, this study has demonstrated the importance of the SDR family in the metabolic and symbiotic capabilities of the plant symbiont *S. meliloti*. Competiton assays and further characterization of pathways utilizing this group of enzymes will provide further insight into the host–symbiont interaction and catabolic diversity of this organism.

Fig.5.1. Metabolic link between ornithine, L-proline, L-leucine and L-lysine degradation pathway


ornithine
glutamate 5-semialdehyde
reductase
L-1-pyrroline-5-carboxylate
L- proline
D-proline
L-lysine
cadaverine
1-piperidine
L-leucine
4-methyl-2-oxopentanoate
2,5-dioxopentanoate
5-aminopentanoate
glutarate semialdehyde
dehydrogenase
glutarate
glutaryl CoA
dehydrogenase
OR
dehydrogenase
crotonyl CoA
3-but-2-enoyl CoA
acetyl CoA
acetyl CoA

References

Allaway, D., E.M. Lodwig, L.A. Crompton, M. Wood, R. Parsons, T.R. Wheeler, *et al.* (2000). Identification of alanine dehydrogenase and its role in mixed secretion of ammonium and alanine by pea bacteroids. *Molecular Microbiology 36*, 508-515.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers & D.J. Lipman. (1990). Basic local alignment search tool. *Journal of Molecular Biology 215*, 403-410.

Andreeva, A., D. Howorth, S.E. Brenner, T.J.P. Hubbard, C. Chothia & A.G. Murzin. (2004). SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acid Research 32, 226-229*

Aneja, P. & T.C. Charles. (2005). Characterization of *bdhA*, encoding the enzyme D-3-hydroxybutyrate dehydrogenase, from *Sinorhizobium* sp. strain NGR234. *FEMS Microbiology Letters 242*, 87-94.

Aneja, P., R. Dziak, G.Q. Cai & T.C. Charles. (2002). Identification of an acetoacetyl coenzyme A synthetase-dependent pathway for utilization of L-(+)-3-hydroxybutyrate in *Sinorhizobium meliloti. Journal of Bacteriology 184*, 1571-1577.

Aneja, P., A. Zachertowska & T.C. Charles. (2005). Comparison of the symbiotic and competition phenotypes of *Sinorhizobium meliloti* PHB synthesis and degradation pathway mutants. *Canadian Journal of Microbiology 51*, 599-604.

Appleby, C.A. (1984). Leghemoglobin and *Rhizobium* respiration. *Annual Review of Plant Physiology 35*, 443-478.

Arias, A. & G.Martinez-Drets (1976). Glycerol metabolism in Rhizobium. *Canadian Journal of Microbiology 22*, 150-153.

Arwas, R., I.A. McKay, F.R.P. Rowney, M.J. Dilworth & A.R. Glenn. (1985). Properties of organic acid utilization mutants of *Rhizobium leguminosarum* strain 300. *Journal of General Microbiology 131*, 2059-2066.

Auerbach, G., A. Herrmann, M. Gutlich, M. Fischer, U. Jacob, A. Bacher, *et al*. (1997). The 1.25 A crystal structure of sepiapterin reductase reveals its binding mode to pterins and brain neurotransmitters. *The EMBO Journal 16*, 7219-7230.

Barnett, M.J., R.F. Fisher, T. Jones, C. Komp, A.P. Abola, F. Barloy-Hubler, *et al.* (2001). Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. *Proceedings of the National Academy of Sciences of the United States of America 98*, 9883-9888.

Barnett, M.J., C.J. Toman, R.F. Fisher & S.R. Long. (2004). A dual-genome Symbiosis Chip for coordinate study of signal exchange and development in a prokaryote–host interaction. *Proceedings of the National Academy of Science USA 101*, 16636-16641.

Barsomian, G.D., A. Urzainqui, K. Lohman & G.C. Walker. (1992). *Rhizobium meliloti* mutants unable to synthesize anthranilate display a novel symbiotic phenotype. *Journal of Bacteriology 174*, 4416-4426.

Baudouin, E., P. Frendo, M. Le Gleuher & A. Puppo. (2004). A *Medicago sativa* haem oxygenase gene is preferentially expressed in root nodules. *Journal of Experimental Botany 55*, 43-47.

Bednarek, A., K. Laflin, R. Daniel, Q. Liao, K. Hawkins & C. Aldaz. (2000). WWOX, a novel WW domain containing protein mapping to the human chromosome 16q23.3-24.1, a region frequently affected in breast cancer. *Cancer Research 60*, 2140-2145.

Bellafonto, V., L. Shapiro, Hodgson & D.A. (1984). Generation of a Tn*5* promoter probe and its use in the study of gene expression in *Caulobacter crescentus*. *Proceedings of the National Academy of Science USA 81*, 1035-1039.

Benach, J., C. Filling, U.C. Oppermann, P. Roversi, G. Bricogne, K.D. Berndt, *et al.* (2002). Structure of bacterial 3β/17β-hydroxysteroid dehydrogenase at 1.2 A resolution: a model for multiple steroid recognition. *Biochemistry 41*, 14659-14668.

Berg, D.E. (1980). Control of gene expression by a mobile recombinational switch. *Proceedings of the National Academy of Science USA 77*, 4880-4884.

Berg, D.E., J. Davies, B. Allet & J.D. Rochaix. (1975). Transposition of R factor genes to bacteriophage lambda. *Proceedings of the National Academy of Science USA 72*, 3628-3632.

Berg, D.E. (1989). Transposon Tn5. *Mobile DNA, Edited by D. E. BERG and M. M. HOWE. American Society for Microbiology, Washington, D.C. 1*, 185-210.

Bergersen, F.J. & G.L. Turner. (1967). Nitrogen fixation by the bacteroid fraction of breis of soybean root nodules. *Biochimica Et Biophysica Acta 141*, 507-515.

Beringer, J.E. (1980). The developement of *Rhizobium* genetics. *Journal of General Microbiology 116*, 1-7.

Boesten, B. & U.B. Priefer. (2004). The C-terminal receiver domain of the *Rhizobium leguminosarum* bv. *viciae* FixL protein is required for free-living microaerobic induction of the *fnrN* promoter. *Microbiology 150*, 3703-3713.

Börnke, F., M. Hajirezai & U. Sonnewald (2001). Cloning and characterization of the gene cluster for palatinose metabolism from the phytopathogenic bacterium *Erwinia rhapontici*. *Journal of Bacteriology 183*, 2425-2433.

Bringhurst, R.M. & D.J. Gage. (2002). Control of inducer accumulation plays a key role in succinate-mediated catabolite repression in *Sinorhizobium meliloti*. *Journal of Bacteriology 184*, 5385-5392.

Broughton, W.J., Wong, C.H., Lewin, A., Samrey, U., Myint, H., Meyer, H.A., Dowling, D.N. & Simon, R. (1986). Identification of *Rhizobium* plasmid sequences involved in recognition of *Psophocarpus*, *Vigna*, and other legumes. *Journal of Cell Biology 102*, 1173-1182

Burnet, M.W., A. Goldmann, B. Message, R. Drong, A. El Amrani, O. Loreau, et al. (2000). The stachydrine catabolism region in *Sinorhizobium meliloti* encodes a multi-enzyme complex similar to the xenobiotic degrading systems in other bacteria. *Gene 244*, 151-161.

Cai, G.Q., B.T. Driscoll & T.C. Charles. (2000). Requirement for the enzymes acetoacetyl coenzyme A synthetase and poly-3-hydroxybutyrate (PHB) synthase for growth of *Sinorhizobium meliloti* on PHB cycle intermediates. *Journal of Bacteriology 182*, 2113-2118.

Capela, D., F. Barloy-Hubler, J. Gouzy, G. Bothe, F. Ampe, J. Batut, *et al*. (2001). Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. *Proceedings of the National Academy of Sciences of the United States of America 98*, 9877-9882.

Chancy, S., D. Wood, E. Pierson & L. Pierson. (2002). Survival of GacS/GacC mutants of the biological control bacterium *Pseudomonas aureofaciens* 30-38 in the wheat rhizosphere. *Applied and Environmental Microbiology 68*, 3308-3314.

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

Charles, T.C., Cai. (1997). Megaplasmid and chromosomal loci for the PHB degradation pathway in *Rhizobium (Sinorhizobium) meliloti*. *Genetics 146*, 1211-1220.

Charles, T.C. & P. Aneja. (1999). Methylmalonyl-CoA mutase encoding gene of *Sinorhizobium meliloti*. *Gene 226*, 121-127.

Cheng, H.P. & G.C. Walker. (1988). Succinoglycan is required for initiation and elongation of infection threads during nodulation of alfalfa by *Rhizobium meliloti*. *Journal of Bacteriology 180*, 5183-5191.

Cloutier, J., S. Laberge & H. Antoun. (1997). Sequence and mutational analysis of the 6.7-kb region containing *nodAFE*G genes of *Rhizobium* sp. strain N33: evidence of DNA rearrangements. *Molecular Plant-Microbe Interactions 10*, 401-406.

Cowie, A., J. Cheng, C.D. Sibley, Y. Fong, R. Zaheer, C.L. Patten, *et al.* (2006). An integrated aproach to functional genomics: Construction of a novel reporter gene fusion libraray for *Sinorhizobium meliloti*. *Applied and Environmental Microbiology 72*, 7156-7167.

Cronan Jr. J.E. & C.O. Rock. (1996). Biosynthesis of membrane lipids. In: F.C. Neidhardt, R. Curtis, C.A. Gross, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W. Reznikoff, M. Riley, M. Schaechter and H.E. Umbarger, Editors, *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology. *American Society for Microbiology, Washington, DC 1*, 612-636.

Danielsson, O., S. Atrian, T. Luque, L. Hjelmqvist, R. Gonzalez-Duarte & H. Jornvall. (1994). Fundamental molecular differences between alcohol

dehydrogenase classes. *Proceedings of the National Academy of Sciences of the United States of America 91*, 4980-4984.

Davidson, A.L. & J. Chen (2004). ATP-binding cassette transporters in bacteria. *Annual Review of Biochemistry 73*, 241-268.

De Costa, D.M., K. Suzuki & K. Yoshida. (2003). Structural and functional analysis of a putative gene cluster for palatinose transport on the linear chomosome of *Agrobacterium tumefaciens* MAFF301001. *Journal of bacteriology 185, 2369-2373*

De Jong, R.M., J.J. Tiesinga, H.J. Rozeboom, K.H. Kalk, L. Tang, D.B. Janssen, *et al*. (2003). Structure and mechanism of a bacterial haloalcohol dehalogenase: a new variation of the short-chain dehydrogenase/reductase fold without an NAD(P)H binding site. *The EMBO Journal 22*, 4933-4944.

De Vos, G., G. Walker & E. Signer. (1986). Genetic manipulations in *Rhizobium meliloti* utilizing two new transposon Tn5 derivatives. *Molecular and General Genetics 204*, 485-491.

DeHertough, A.A., P.A. Mayeux & H.J. Evans. (1964). Relationship of cobalt requirement to propionate metabolism in *Rhizobium*. *Journal of Biological Chemistry 239*, 2446-2453.

D'Hooghe, I., C. Vander Wauven, J. Michiels, C. Tricot, P. de Wilde, J. Vanderleyden, *et al*. (1997). The arginine deiminase pathway in *Rhizobium etli*: DNA sequence analysis and functional study of the *arcABC* genes. *Journal of Bacteriology 179*, 7403-7409.

Dickstein, R., D.C. Scheirer, W.H. Fowle & F.M. Ausubel. (1991). Nodules elicited by *Rhizobium meliloti* heme mutants are arrested at an early stage of development. *Molecular & General Genetics : MGG 230*, 423-432.

Ditta, G., Stanfield, S., Corbin, D. & Helinski, D. R. (1980). Broad host range DNA cloning system for Gram-negative bacteria: Construction of a gene bank of *Rhizobium meliloti. Proceedings of the National Academy of Science USA 77*, 7347-7351.

Dixon, R., Kahn, D. (2004). Genetic regulation of biological nitrogen fixation. *Nature Review of Microbiology 2*, 621-631.

Djordjevic, M.A. (2004). *Sinorhizobium meliloti* metabolism in the root nodule: a proteomic perspective. *Proteomics 4*, 1859-1872.

Djordjevic, M.A., H.C. Chen, S. Natera, G. Van Noorden, C. Menzel, S. Taylor, *et al*. (2003). A global analysis of protein expression profiles in *Sinorhizobium meliloti*: discovery of new genes for nodule occupancy and stress adaptation. *Molecular Plant-Microbe Interactions : MPMI 16*, 508-524.

Domínguez-Ferreras,A., Pérez-Arnedo,R., Becker,A., Olivares,J., Soto,M.J., Sanjuán,J. (2006). Transcriptome profiling reveals the importance of plasmid pSymB for osmoadaptation of *Sinorhizobium meliloti*. *Journal of Bacteriology 188*, 7617-7625.

Driscoll, B.T., Finan. (1993). NAD^+-dependent malic enzyme of *Rhizobium meliloti* is required for symbiotic nitrogen fixation. *Molecular Microbiology 7*, 865-873.

Dsouza M, Larsen N, Overbeek R. (1997). Searching for patterns in genomic data. *Trends in Genetics 13*, 497-498.

Duncan, M.J. & D.G. Fraenkel. (1979). α-Ketoglutarate dehydrogenase mutant of *Rhizobium meliloti*. *Journal of Bacteriology 137*, 415-419.

Dunn, M.F. (1998). Tricarboxylic acid cycle and anaplerotic enzymes in rhizobia. *FEMS Microbiology Reviews 22*, 105-123.

Dunn, M.F., G. Araiza & T.M. Finan. (2001). Cloning and characterization of the pyruvate carboxylase from *Sinorhizobium meliloti* Rm1021. *Archives of Microbiology 176*, 355-363.

Dymov, S.I., D.J. Meek, B. Steven & B.T. Driscoll. (2004). Insertion of transposon Tn*5tac1* in the *Sinorhizobium meliloti* malate dehydrogenase (*mdh*) gene results in conditional polar effects on downstream TCA cycle genes. *Molecular Plant-Microbe Interactions : MPMI 17*, 1318-1327.

Fellay, R., X. Perret, V. Viprey, W.J. Broughton & S. Brenner. (1995). Organization of host-inducible transcripts on the symbiotic plasmid of *Rhizobium* sp. NGR234. *Molecular Microbiology 16*, 657-667.

Feller, C., R. Gunther, H.J. Hofmann & M. Grunow. (2006). Molecular basis of substrate recognition in D-3-hydroxybutyrate dehydrogenase from *Pseudomonas putida*. *Chembiochem : A European Journal of Chemical Biology 7*, 1410-1418.

Filling, C., K. Berndt, J. Benach, S. Knapp, T. Prozorovski, E. Nordling, et al. (2002). Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. *Journal of Biological Chemistry 277*, 25677-25684.

Finan, T.M., J.M. Wood & D.C. Jordan. (1983). Symbiotic properties of C4-dicarboxylic acid transport mutants of *Rhizobium leguminosarum*. *Journal of Bacteriology 154*, 1403-1413.

Finan,T.M., Wood, J.M., Jordan, D.C. (1981). Succinate transport in *Rhizobium leguminosarum*. *Journal of Bacteriology 148*, 193-202.

Finan, T.M., E. McWhinnie, B. Driscoll & R.J. Watson. (1991). Complex symbiotic phenotypes result from gluconeogenic mutations in *Rhizobium meliloti*. *Molecular Plant-Microbe Interaction 4*, 386-392.

Finan, T.M., I. Oresnik & A. Bottacin. (1988). Mutants of *Rhizobium meliloti* defective in succinate metabolism. *Journal of Bacteriology 170*, 3396-3403.

Finan, T.M., S. Weidner, K. Wong, J. Buhrmester, P. Chain, F.J. Vorholter, *et al.* (2001). The complete sequence of the 1,683-kb pSymB megaplasmid from the N_2-fixing endosymbiont *Sinorhizobium meliloti*. *Proceedings of the National Academy of Sciences of the United States of America 98*, 9889-9894.

Finn, R.D., J. Mistry, B. Schuster-Böckler, S. Griffiths-Jones, V. Hollich, T. Lassmann, *et al.* (2006). Pfam: clans, web tools and services *Nucleic Acids Research 34*, 247-251.

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 use in the genetic analysis of *Rhizobium* mutants. *Gene 18*, 289-296.

Fry, J., M. Wood & P.S. Poole. (2001). Investigation of myo-inositol catabolism in *Rhizobium leguminosarum* bv. *viciae* and its effect on nodulation competitiveness. *Molecular Plant-Microbe Interaction 14*, 1016-1025.

Furuichi, T., M. Inouye & S. Inouye. (1986). Novel one-step cloning vector with a transposable element: application to the *Myxococcus xanthus* genome. *Journal of Bacteriology 164*, 270-275.

Gage, D.J. & S.R. Long. (1998). α-Galactoside uptake in *Rhizobium meliloti*: isolation and characterization of *agpA*, a gene encoding a periplasmic binding protein required for melibiose and raffinose utilization. *Journal of Bacteriology 180*, 5739-5748.

Galbraith, M.P., S.F. Feng, J. Borneman, E.W. Triplett, F.J. deBruijn & S. Rossbach. (1998). A functional myo-inositol catbolism pathway is essential

for rhizopine utilization by *Sinorhizobium meliloti*. *Microbiology 144*, 2915-2924.

Galibert, F., T.M. Finan, S.R. Long, A. Puhler, P. Abola, F. Ampe, *et al*. (2001). The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science 293*, 668-672.

Gallego, O., O.V. Belyaeva, S. Porte, F.X. Ruiz, A.V. Stetsenko, E.V. Shabrova, *et al.* (2006). Comparative functional analysis of human medium-chain dehydrogenases, short-chain dehydrogenases/reductases and aldo-keto reductases with retinoids. *The Biochemical Journal 399*, 101-109.

Garcia-de los Santos, A., A. Morales, L. Baldoma, S.R. Clark, S. Brom, C.K. Yost, *et al.* (2002). The *glcB* locus of *Rhizobium leguminosarum* VF39 encodes an arabinose-inducible malate synthase. *Canadian Journal of Microbiology 48*, 922-932.

Gardiol, A., A. Arias, C. Cervenansky & G. Martinez-Drets. (1982). Succinate dehydrogenase mutant of *Sinorhizobium meliloti*. *Journal of Bacteriology 151*, 1621-1623.

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

Glenn, A.R., I.A. McKay, R. Arwas & M.J. Dilworth. (1984). Sugar metabolism and the symbiotic properties of carbohydrate mutants of *Rhizobium leguminosarum*. *Journal of General Microbiology 130*, 4-6.

Glenn, A.R., P.S. Poole & J.F. Hudman. (1980). Succinate uptake by free-living and bacteroid forms of *Rhizobium leguminosarum*. *Journal of General Microbiology 119*, 267-271.

González, J.E. & M.M. Marketon. (2003). Quorum sensing in nitrogen-fixing Rhizobia. *Journal of Bacteriology 67*, 574-592.

Gonzalez, V., P. Bustos, M.A. Ramirez-Romero, A. Medrano-Soto, H. Salgado, I. Hernandez-Gonzalez, *et al.* (2003). The mosaic structure of the symbiotic plasmid of *Rhizobium etli* CFN42 and its relation to other symbiotic genome compartments. *Genome Biology 4*, R36.

Gosselin, I., O. Wattraint, D. Riboul, J. Barbotin & J. Portais. (2001). A deeper investigation on carbohydrate cycling in *Sinorhizobium meliloti*. *FEBS Letters 499*, 45-49.

Graham, P.H. (1964). Studies on the utilization of carbohydrates and Krebs cycle intermediates by Rhizobia, using agar plate method. *Antonie Van Leeuwenhoek 30*, 68-72.

Grimm, C., E. Maser, E. Mobus, G. Klebe, K. Reuter & R. Ficner. (2000). The crystal structure of 3alpha -hydroxysteroid dehydrogenase/carbonyl reductase from *Comamonas testosteroni* shows a novel oligomerization pattern within the short chain dehydrogenase/reductase family. *The Journal of Biological Chemistry 275*, 41333-41339.

Guo, K., P. Lukacik, E. Papagrigoriou, M. Meier, W.H. Lee, J. Adamski*, et al*. (2006). Characterization of human DHRS6, an orphan short chain dehydrogenase/reductase enzyme: a novel, cytosolic type 2 R-beta-hydroxybutyrate dehydrogenase. *The Journal of Biological Chemistry 281*, 10291-10297.

Hassani, R., C.K. Prasad, K.E. Vineetha, N. Vij, P. Singh, R. Sud, *et al.* (2002). Symbiotic characterization of isoleucine+valine and leucine auxotrophs of *Sinorhizobium meliloti*. *Indian Journal of Experimental Biology 40*, 1110-1120.

Hirsch, A.M., D. Drake, T.W. Jacobs & S.R. Long. (1985). Nodules are induced on alfalfa roots by *Agrobacterium tumefaciens* and *Rhizobium trifolii* containing small segments of the Rhizobium meliloti nodulation region. *Journal of Bacteriology 161*, 223-230.

Hooper, L.V., J. Xu, P.G. Falk, T. Midtvedt & J.I. Gordon. (1999). A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem. *Proceedings of the National Academy of Science USA 96*, 9833-9838.

Hosie, A. H. F., D. Allaway, C.S. Galloway, H.A. Dunsby & P.S. Poole. (2002). *Rhizobium leguminosarum* has a second general amino acid permease with unusually broad substrate specificity and high similarity to branched-chain amino acid transporters (Bra/Liv) of the ABC Family. *Journal of Bacteriology 184*, 4071-4080.

Hosie, A. H. F., D. Allaway, M.A. Jones, D.L. Walshaw, Johnston, A. W. B. & P.S. Poole. (2001). Solute-binding protein-dependent ABC transporters are responsible for solute efflux in addition to solute uptake. *Molecular Microbiology 40*, 1449-1459.

Hulo N., Bairoch A., Bulliard V., Cerutti L., De Castro E., Langendijk-Genevaux P.S., *et al*. (2006). The PROSITE database. *Nucleic Acids Research 34*, 227-230.

Irigoyen, J.J., M. Sanchez-Diaz & D.W. Emerich. (1990). Carbon metabolism enzymes of *Rhizobium meliloti* cultures and bacteroids and their distribution within alfalfa nodules. *Applied and Environmental Microbiology 56*, 2587-2589.

Ito, K., Y. Nakajima, E. Ichihara, K. Ogawa, N. Katayama & K. Nakashima, (2006). D-3-hydroxybutyrate dehydrogenase from *Pseudomonas fragi*:

molecular cloning of the enzyme gene and crystal structure of the enzyme. *Journal of Molecular Biology 355*, 722-733.

Jelesko, J.G. & J.A. Leigh. (1994). Genetic characterization of a *Rhizobium meliloti* lactose utilization locus. *Molecular Microbiology 11*, 165-173.

Jensen, J.B., O.Y. Ampomah, R. Darrah, N.K. Peters & T.V. Bhuvaneswari. (2005). Role of trehalose transport and utilization in *Sinorhizobium meliloti*--alfalfa interactions. *Molecular Plant-Microbe Interactions 18*, 694-702.

Jiang, G., A.H. Krishnan, Y. Kim, T.J. Wacek & H.B. Krishnan. (2001). A functional myo-inositol dehydrogenase gene is required for efficient nitrogen fixation and competitiveness of *Sinorhizobium fredii* USDA191 to nodulate soybean (*Glycine max* [L.] *Merr*. *Journal of Bacteriology 183*, 2595-2604.

Jimenez-Zurdo, J.I., F.M. Garcia-Rodriguez & N. Toro. (1997). The *Rhizobium meliloti putA* gene: its role in the establishment of the symbiotic interaction with alfalfa. *Molecular Microbiology 23*, 85-93.

Johansson, K., M. El-Ahmad, C. Kaiser, H. Jornvall, H. Eklund & J. Hoog, (2001). Crystal structure of sorbitol dehydrogenase. *Chemico-Biological Interactions 130-132*, 351-358.

Johnson, G.V., H.J. Evans & T. Ching. (1966). Enzymes of the glyoxylate cycle in rhizobia and nodules of legumes. *Plant Physiology 41*, 1330-1336.

Jörnnvall, H., B. Persson, M. Krook, S. Atrian, R. Duarte, J. Jeffery, *et al*. (1995). Short-chain dehydrogenases/reductases (SDR). *Biochemistry 34*, 6003-6013.

Jörnvall, H., J.O. Hoog & B. Persson. (1999). SDR and MDR: completed genome sequences show these protein families to be large, of old origin, and of complex nature. *FEBS Letters 445*, 261-264.

Julliot, J.S., I. Dusha, M.H. Renalier, B. Terzaghi, A.M. Garnerone & P. Boistard. (1984). An RP4-prime containing a 285 kb fragment of *Rhizobium meliloti* pSym megaplasmid: structural characterisation and utilisation for genetic studies of symbiotic functions controlled by pSym. *Journal of Bacteriology 193*, 17-26.

Kallberg, Y., U. Oppermann, H. Jörnvall & B. Persson. (2002). Short-chain dehydrogenases/reductases (SDRs)-Coenzyme based functional assignments in the completed genomes. *European Journal of Biochemistry 269*, 4409-4417.

Kallberg, Y., U. Oppermann, H. Jörnvall & B. Persson. (2002). Short-chain dehydrogenase/reductase (SDR) relationships: a large family with eight clusters common to human, animal, and plant genomes. *Protein Science: A Publication of the Protein Society 11*, 636-641.

Kallberg, Y., U. Oppermann, H. Jörnvall & B. Persson. (2002). Short-chain dehydrogenases/reductases (SDRs). *European Journal of Biochemistry / FEBS 269*, 4409-4417.

Kallberg, Y. & B. Persson. (2006). Prediction of coenzyme specificity in dehydrogenases/reductases. A hidden Markov model-based method and its application on complete genomes. *The FEBS Journal 273*, 1177-1184.

Kalogeraki, V.S. & S.C. Winans. (1997). Suicide plasmids containing promoterless reporter genes can simultaneously disrupt and create fusions to target genes of diverse bacteria. *Gene 188*, 69-75.

Kaneko, T., Y. Nakamura, S. Sato, E. Asamizu, T. Kato, S. Sasamoto, *et al*. (2000). Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Research 7*, 331-338.

Kaneko, T., Y. Nakamura, S. Sato, E. Asamizu, T. Kato, S. Sasamoto, *et al*. (2000). Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti* (supplement). *DNA Research 7*, 381-406.

Kaneko, T., Y. Nakamura, S. Sato, K. Minamisawa, T. Uchiumi, S. Sasamoto, *et al.* (2002). Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Research 9*, 189-197.

Kaneko, T., Y. Nakamura, S. Sato, K. Minamisawa, T. Uchiumi, S. Sasamoto, *et al.* (2002). Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110 (supplement). *DNA Research 9*, 225-256.

Karr, D.B., J.K. Waters, F. Suzuki & D.W. Emerich. (1984). Enzymes of the Poly-β-Hydroxybutyrate and Citric Acid Cycles of *Rhizobium japonicum* Bacteroids . *Plant Physiology 75*, 1158-1162.

Katznelson, H. & A.C. Zagallo. (1957). Metabolism in Rhizobium in relation to effectiveness. *Canadian Journal of Microbiology 3*, 879-884.

Keele, B.B., P.B. Hamilton & Elkan. G.H. (1970). Gluconate catabolism in *Rhizobium japonicum. Journal of Bacteriology 101*, 698-704.

Keller, B., A. Volkmann, T. Wilckens, G. Moeller & J. Adamski. (2006). Bioinformatic identification and characterization of new members of short-chain dehydrogenase/reductase superfamily. *Molecular and Cellular Endocrinology 248*, 56-60.

Kelly, D.J. & G.H. Thomas. (2001). The tripartite ATP-independent periplasmic (TRAP) transporters of bacteria and archaea. *FEMS Microbiology Reviews 25*, 405-424.

Kerppola, T.K. & M.L. Kahn. (1988). Symbiotic phenotypes of auxotrophic mutants of *Rhizobium meliloti* 104A14. *Journal of General Microbiology 134*, 913-919.

Kho, R., J.V. Newman, R.M. Jack, H.O. Villar & M.R. Hansen. (2003). Genome-wide profile of oxidoreductases in viruses, prokaryotes, and eukaryotes. *Journal of Proteome Research 2*, 626-632.

Kiss, G.B., K. Dobo, I. Dusha, A. Breznovits, L. Orosz, E. Vincze, et al. (1980). Isolation and characterization of an R-prime plasmid from *Rhizobium meliloti*. *Journal of Bacteriology 141*, 121-128.

Klucas, R.V. & H.J. Evans. (1968). An electron donor system for nitrogenase-dependent acetylene reduction by extracts of soybean nodules. *Plant Physiology 42*, 1458-1460.

Kohl, D.H., P.F. Struab & G. Shearer. (1994). Does proline play a special role in bacteroid metabolism? *Plant, Cell & Environment 17*, 1257-1262.

Kohl, D.H., K.R. Schubert, M.B. Carter, C.H. Hagedorn & G. Shearer. (1988). Proline metabolism in N_2-fixing root nodules: energy transfer and regulation of purine synthesis. *Proceedings of the National Academy of Sciences of the United States of America 85*, 2036-2040.

Kondorosi, E., Z. Banfalvi & A. Kondorosi. (1984). Physical and genetic analysis of a symbiotic region of *Rhizobium meliloti*: identification of nodulation genes. *Molecular and General Genetics 193*, 445-452.

Koropatkin, N.M. & H.M. Holden. (2005). Structure of CDP-D-glucose 4,6-dehydratase from *Salmonella typhi* complexed with CDP-D-xylose. *Acta Crystallographica.Section D, Biological Crystallography 61*, 365-373.

Krebs, M.P. & Reznikoff WS. (1988). Use of a Tn*5* derivative that creates *lacZ* translational fusions to obtain a transposition mutant. *Gene 63*, 277-285.

Kumar, S., A. Bourdès & P.S. Poole. (2005). De novo alanine synthesis by bacteroids of *Mesorhizobium loti* is not required for nitrogen transfer in the determinate nodules of *Lotus corniculatus*. *Journal of Bacteriology 187*, 5493-5495.

Lai, C. & J.E. Cronan. (2004). Isolation and characterization of beta-ketoacyl-carrier protein reductase (*fabG*) mutants of *Escherichia coli* and *Salmonella enterica* serovar typhimurium. *Journal of Bacteriology 186*, 1869-1878.

Lambert, A., M. Osteras, K. Mandon, M.C. Poggi & D. Le Rudulier. (2001). Fructose uptake in *Sinorhizobium meliloti* is mediated by a high-affinity ATP-binding cassette transport system. *Journal of Bacteriology 183*, 4709-4717.

Lodwig, E.M., A.H. Hosie, A. Bourdes, K. Findlay, D. Allaway, R. Karunakaran, et al. (2003). Amino-acid cycling drives nitrogen fixation in the legume-*Rhizobium* symbiosis. *Nature 422*, 722-726.

Lopez, O. (2001). Regulation of gene expression in response to oxygen in *Rhizobium etli*: role of FnrN in *fixNOQP* expression and in symbiotic nitrogen fixation. *Journal of Bacteriology 183*, 6999-7006.

Lopez-Lara, I.M. & O. Geiger. (2001). The nodulation protein NodG shows the enzymatic activity of an 3-oxoacyl-acyl carrier protein reductase. *Molecular Plant-Microbe Interactions : MPMI 14*, 349-357.

Luo, L., S. Yao, A. Becker, S. Ruberg, G. Yu, J. Zhu, *et al*. (2005). Two new *Sinorhizobium meliloti* LysR-type transcriptional regulators required for nodulation. *Journal of Bacteriology 187*, 4562-4572.

MacLean, A.M., G. MacPherson, P. Aneja & T.M. Finan. (2006). Characterization of the beta-ketoadipate pathway in *Sinorhizobium meliloti*. *Applied and Environmental Microbiology 72*, 5403-5413.

Marketon, M.M. & J.E. González. (2002). Identification of two quorum-sensing systems in *Sinorhizobium meliloti*. *Journal of Bacteriology 184*, 3466-3475.

Martin, M.O. & S.R. Long. (1984). Generalized transduction in *Rhizobium meliloti*. *Journal of Bacteriology 159*, 125-129.

Martinez-De Drets, G. & A. Arias. (1972). Enzymatic basis for the differenciation of *Rhizobium* into fast and slow growing groups. *Journal of Bacteriology 109*, 467-470.

Martinez-Rodriguez, S., J.M. Clemente-Jimenez, F. Rodriguez-Vico & F.J. Las Heras-Vazquez. (2005). Molecular cloning and biochemical characterization of L-N-carbamoylase from *Sinorhizobium meliloti* CECT4114. *Journal of Molecular Microbiology and Biotechnology 9*, 16-25.

Mauchline, T.H., J.E. Fowler, A.K. East, A.L. Sartor, A. Cowie, Hosie, A. H. F., *et al.* (2006). Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome. *Proceedings of the National Academy of Science USA 103*, 17933-17938.

McKay, I.A., M.J. Dilworth & A.R. Glenn. (1988). C4-dicarboxylate metabolism in free-living and bacteroid forms of *Rhizobium*

leguminosarum MNF3841. *Journal of General Microbiology 134*, 1433-1440.

Meade, H.M., S.R. Long, G.B. Ruvkun, S.E. Brown & F.M. Ausubel. (1982). Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn*5* mutagenesis. *Journal of Bacteriology 149*, 114-1220.

Miller-Williams, M., P.C. Loewen & I.J. Oresnik. (2006). Isolation of salt-sensitive mutants of *Sinorhizobium meliloti* strain Rm1021. *Microbiology 152*, 2049-2059.

Miyamoto, E., F. Watanabe, T.C. Charles, R. Yamaji, H. Inui & Y. Nakano. (2003). Purification and characterization of homodimeric methylmalonyl-CoA mutase from *Sinorhizobium meliloti*. *Archives of Microbiology 180*, 151-154.

Morett, E., H.M. Fischer & H. Hennecke. (1991). Influence of oxygen on DNA binding, positive control, and stability of the *Bradyrhizobium japonicum* NifA regulatory protein. *Journal of Bacteriology 173*, 3478-3487.

Mortimer, M.W., T.R. McDermott, G.M. York, G.C. Walker & M.L. Kahn. (1999). Citrate synthase mutants of *Sinorhizobium meliloti* are ineffective and have altered cell surface polysaccharides. *Journal of Bacteriology 181*, 7608-7613.

Mulichak, A.M., C.P. Bonin, W.D. Reiter & R.M. Garavito. (2002). Structure of the MUR1 GDP-mannose 4,6-dehydratase from *Arabidopsis thaliana*: implications for ligand binding and specificity. *Biochemistry 41*, 15578-15589.

Mulongoy, K. & G.H. Elkan. (1977). Glucose catabolism in two derivatives of a *Rhizobium japonicum* strain differing in nitrogen-fixing efficiency. *Journal of Bacteriology 131*, 179-1878.

Nakajima, K., A. Yamashita, H. Akama, T. Nakatsu, H. Kato & T. Hashimoto (1998). Crystal structures of two tropinone reductases: different reaction stereospecificities in the same protein fold. *Proceedings of the National Academy of Sciences of the United States of America 95*, 4876-4881.

Natera, S.H., N. Guerreiro & M.A. Djordjevic. (2000). Proteome analysis of differentially displayed proteins as a tool for the investigation of symbiosis. *Molecular Plant-Microbe Interactions 13*, 995-1009.

Neidhardt, F.C., R. Curtiss III, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, *et al*. (1996). Biosynthesis of lipids. *Escherichia coli and Salmonella Cellular and Molecular Biology 1*, 612-637.

Oldenburg, G., K. Nishina & G. Stephanopoulos. (2000). Identification and analysis of the polyhydroxyalkanoate-specific beta-ketothiolase and acetoactyl-CoA reductase genes in cyanobacterium *Synechocystis* sp. strain PCC 6803. *Applied and Environmental Microbiology 66*, 4440-4448.

Østerås, M., J. Stanley, W.J. Broughton & D.N. Dowling. (1989). A chromosomal genetic map of *Rhizobium* sp. NGR234 generated with Tn*5*-Mob. *Molecular and General Genetics 220*, 157-160.

Ovtsyna, A.O., M. Schultze, I.A. Tikhonovich, H.P. Spaink, E. Kondorosi, A. Kondorosi, et al. (2000). Nod factors of *Rhizobium leguminosarum* bv. *viciae* and their fucosylated derivatives stimulate a nod factor cleaving activity in pea roots and are hydrolyzed in vitro by plant chitinases at different rates. *Molecular Plant-Microbe Interactions 13*, 799-807.

Pain, A.N. (1979). Symbiotic properties of antibiotic-resistant and auxotrophic mutants of *Rhizobium leguminosarum*. *Journal of Applied Microbiology 47*, 53-64.

Parke, D. & L.N. Ornston. (1984). Nutritional diversity of Rhizobiaceae revealed by auxanography. *Journal of General Microbiology 130*, 1743-1750.

Parker, C.A. & P.H. Graham. (1964). Diagnostic feature in the characterization of he root-nodule bacteria of legumes. *Plant and Soil 3*, 383-395.

Patriarca, E.J., R. Tate, S. Ferraioli & M. Iaccarino. (2004). Organogenesis of legume root nodules. *International Review of Cytology 234*, 201-262.

Patriarca, E.J., R. Tate & M. Iaccarino. (2002). Key role of bacterial NH4+ metabolism in *Rhizobium*-plant symbiosis. *Microbiology and Molecular Biology Reviews 66*, 203-222.

Pedrosa, F.O. & G.T. Zancan. (1974). L-Arabinose Metabolism in *Rhizobium japonicum*. *Journal of Bacteriology 119*, 336-338.

Persson, B., Y. Kallberg, U. Oppermann & H. Jornvall. (2003). Coenzyme-based functional assignments of short-chain dehydrogenases/reductases (SDRs). *Chemico-Biological Interactions 143-144*, 271-278.

Peterson, J.B. & T.A. LaRue. (1982). Soluble aldehyde dehydrogenase and metabolism of aldehydes by soybean bacteroids. *Journal of Bacteriology 151*, 1473-1484.

Pobigaylo, N., D. Wetter, S. Szymczak, U. Schiller, S. Kurtz, F. Meyer, et al. (2006). Construction of a large signature-tagged mini-Tn5 transposon library and its application to mutagenesis of *Sinorhizobium meliloti*. *Applied and Environmental Microbiology 72*, 4329-4337.

Povolo, S. & S. Casella. (2000). A critical role for *aniA* in energy-carbon flux and symbiotic nitrogen fixation in *Sinorhizobium meliloti*. *Archives of Microbiology 174*, 42-49.

Poysti, N.J., Loewen, E. D. M., Z. Wang & I.J. Oresnik. (2007). *Sinorhizobium meliloti* pSymB encodes genes necessary for arabinose transport and catabolism. *Microbiology 153*, 727-736.

Poysti, N.J. & I.J. Oresnik. (2007). Characterization of *Sinorhizobium meliloti* triose phosphate isomerase genes. *Journal of Bacteriology 189*, 3445-3451.

Prell, J. & P. Poole. (2006). Metabolic changes of rhizobia in legume nodules. *Trends in Microbiology 14*, 161-168.

Primrose, S.B. & C.W. Ronson. (1980). Polyol metabolism in *Rhizobium trifolii*. *Journal of Bacteriology 149*, 1109-1111.

Raghava, G.P.S. (2001). A web server for computing the size of DNA/protein fragments using a graphical method. *Biotech Software & Internet Report 2*, 198-200.

Ramu, C., H. Sugawara, T. Koike, R. Lopez, T.J. Gibson, D.G. Higgins, *et al.* (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research 31*, 3497-3500.

Renalier, M.H., J. Batut, J. Ghai, B. Terzaghi, M. Gherardi, M. David, *et al.* (1987). A new symbiotic cluster on the pSym megaplasmid of *Rhizobium meliloti* 2011 carries a functional *fix* gene repeat and a *nod* locus. *Journal of Bacteriology 169*, 2231-2238.

Richardson, J.S., M.F. Hynes & I.J. Oresnik. (2004). A genetic locus necessary for rhamnose uptake and catabolism in *Rhizobium leguminosarum* bv. *trifolii*. *Journal of Bacteriology 186*, 8433-8442.

Ronson, C.W., P. Lyttleton & J.G. Robertson. (1981). C4-dicarboxylate transport mutants of *Rhizobium trifolii* form ineffective nodules on *Trifolium repens*. *Proceedings of the National Academy of Science U S A. 78*, 4284-4288.

Rosen, S. & H.J. Staletsky. (2000). Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ 1*, 365-386.

Sambrook, J. & D.W. Russell. (1989). Molecular cloning: A laboratory manual. *Cold Spring Harbor Laboratory Press 3*, A2.2-A2.4.

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

Schroeder, B.K., B.L. House, M.W. Mortimer, S.N. Yurgel, S.C. Maloney, K.L. Ward, *et al*. (2005). Development of a functional genomics platform for *Sinorhizobium meliloti*: construction of an ORFeome. *Applied and Environmental Microbiology 71*, 5858-5864.

Selander, R.K., D.A. Caugant, H. Ochman, J.M. Musser, M.N. Gilmour & T.S. Whittam. (1986). Methods of multilocus enzyme electrophoresis for bacterial genetics and systematics. *Applied and Environmental Microbiology 51*, 873-884.

Shah, R. & D.W. Emerich. (2006). Isocitrate dehydrogenase of *Bradyrhizobium japonicum* is not required for symbiotic nitrogen fixation with soybean. *2006 188*, 7600-7608.

Somers, W.S., M.L. Stahl & F.X. Sullivan. (1998). GDP-fucose synthetase from *Escherichia coli*: structure of a unique member of the short-chain

dehydrogenase/reductase family that catalyzes two distinct reactions at the same active site. *Structure 6*, 1601-1612.

Soto, M.J., J. Sanjuan & J. Olivares. (2001). The disruption of a gene encoding a putative arylesterase impairs pyruvate dehydrogenase complex activity and nitrogen fixation in *Sinorhizobium meliloti*. *Molecular Plant-Microbe Interactions 14*, 811-815.

Spaink, H.P. (2000). Root nodulation and infection factors produced by rhizobial bacteria. *Annual Review of Microbiology 54*, 257-288.

Stacey, G., M. Libault, L. Brechenmacher, J. Wan & G.D. May. (2006). Genetics and functional genomics of legume nodulation. *Current Opinion in Plant Biology 9*, 110-121.

Stovall, I. & M. Cole. (1978). Organic acid metabolism by isolated *Rhizobium japonicum* bacteroids. *Plant Physiology 61*, 787-790.

Stowers, M.D. (1985). Carbon metabolism in *Rhizobium* species. *Annual Review of Microbiology 39*, 89-107.

Stowers, M.D. & A.R.J. Eaglesham. (1984). Physiological and symbiotic characteristics of fast-growing *Rhizobium japonicum* strains. *Plant and Soil 77*, 3-1459.

Stowers, M.D. & G.H. Elkan. (1985). Regulation of hexose catabolism in *Rhizobium* sp. 32H1. *FEMS Microbiology Letters 26*, 45-4884.

Stowers, M.D. & G.H. Elkan (1983). The transport and metabolism of glucose in cowpea rhizobia. *Canadian Journal of Microbiology 29*, 398-406.

Stowers, M.D. (1985). Carbon metabolism in *Rhizobium* species. *Annual Review of Microbiology 39*, 89-108.

Stowers, M.D. & G.H. Elkan. (1985). Regulation of hexose catabolism in *Rhizobium* species. *FEMS Microbiology Letters 26*, 45-48.

Tesfaye, M., D.A. Samac & C.P. Vance. (2006). Insights into symbiotic nitrogen fixation in *Medicago truncatula*. *Molecular Plant-Microbe Interactions 19*, 330-341.

Thony-Meyer, L. & P. Kunzler. (1996). The *Bradyrhizobium japonicum* aconitase gene (*acnA*) is important for free-living growth but not for inefective root-nodule symbiosis. *Journal of Bacteriology 178*, 6166-6172.

Trainer, M.A. & T.C. Charles. (2006). The role of PHB metabolism in the symbiosis of Rhizobia with legumes. *Applied Microbiology and Biotechnology 71*, 377-386.

Van Hylckama Vlieg, J.E., L. Tang, J.H. Lutje Spelberg, T. Smilda, G.J. Poelarends, T. Bosma, *et al*. (2001). Halohydrin dehalogenases are structurally and mechanistically related to short-chain dehydrogenases/reductases. *Journal of Bacteriology 183*, 5058-5066.

Vanbleu, E., K. Marchal & J. Vanderleyden. (2004). Genetic and physical map of the pLAFR1 vector. *DNA Sequence 15*, 225-227.

Walshaw, D., A. Wilkinson, M. Mundy, M. Smith & P.S. Poole. (1997). Regulation of the TCA cycle and the general amino acid permease by overflow metabolism in *Rhizobium leguminosarum*. *Microbiology 143*, 2201-2221.

Wang, C., D.J. Meek, P. Panchal, N. Boruvka, F.S. Archibald, B.T. Driscoll, et al. (2006). Isolation of poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants. *Applied and Environmental Microbiology 72*, 384-391.

Wang, C., M. Saldanha, X. Sheng, K.J. Shelswell, K.T. Walsh, B.W.S. Sobral, et al. (2007). Roles of poly-3-hydroxybutyrate (PHB) and glycogen in symbiosis of *Sinorhizobium meliloti* with *Medicago* sp. *Microbiology 153*, 388-398.

Waters, J.K., B.L. Hughes II, L.C. Purcell, K.O. Gerhardt, T.P. Mawhinney & D.W. Emerich. (1998). Alanine, not ammonia is excreted from N2-fixing soybean nodule bacteroids. *Proceedings of the National Academy of Science U S A. 95*, 12038-12042.

Watson, R.J., Y.K. Chan, R. Wheatcroft, A.F. Yang & Han S. (1988). *Rhizobium meliloti* genes required for C4-dicarboxylate transport and symbiotic nitrogen fixation are located on a megaplasmid. *Journal of Bacteriology 170*, 927-934.

Wong, P.P. & H.J. Evans. (1971). Poly-3-hydroxybutyrate utilization by soybean (Glycine max Merr.) nodules and assessment of its role in maintenance of nitrogenase activity. *Plant Physiology 47*, 750-755.

Wood, D.W., J.C. Setubal, R. Kaul, D.E. Monks, J.P. Kitajima, V.K. Okura, *et al.* (2001). The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science 294*, 2317-2323.

Yarosh, O.K., T.C. Charles & T.M. Finan. (1989). Analysis of C4-dicarboxylic acid transport genes in *Rhizobium meliloti*. *Molecular Microbiology 3*, 813-823.

Yost, C.K., A.M. Rath, T.C. Noel & M.F. Hynes. (2006). Characterization of genes involved in erythritol catabolism in *Rhizobium leguminosarum* bv. *viciae*. *Microbiology 152*, 2061-2074.

Young, J.P., L.C. Crossman, A.W. Johnston, N.R. Thomson, Z.F. Ghazoui, K.H. Hull, et al. (2006). The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biology 7*, R34.

Yurgel, S. & M.L. Kahn. (2005). *Sinorhizobium meliloti dctA* mutants with partial ability to transport dicarboxylic acids. *Journal of Bacteriology 187*, 1161-1172.

Yurgel, S., M.W. Mortimer, K.N. Rogers & M.L. Kahn. (2000). New substrates for the dicarboxylate transport system of *Sinorhizobium meliloti*. *Journal of Bacteriology 182*, 4216-4221.

Yurgel, S.N. & M.L. Kahn. (2004). Dicarboxylate transport by rhizobia. *FEMS Microbiology Reviews 28*, 489-501.

Zhang, X.S. & H.P. Cheng. (2006). Identification of *Sinorhizobium meliloti* early symbiotic genes by use of a positive functional screen. *Applied and Environmental Microbiology 72*, 2738-2748.

Zhang, Y. & J.E. Cronan. (1998). Transcriptional analysis of the essential genes in the fatty acid biosynthesis gene cluster by functional replacement with the analogous *Salmonella typhimurium* gene cluster. *Journal of Bacteriology 180*, 3295-3303.

Zhu, Y., G. Shearer & D.H. Kohl. (2000). Proline fed to intact soybean plants influences acetylene reducing activity and content and metabolism of proline in bacteroids. *Plant Physiology 98*, 1020-1028.