Sugar alcohol metabolism in the legume endosymbiont *Sinorhizobium meliloti*

by

MacLean G. Kohlmeier

A Thesis Submitted to the Faculty of Graduate Studies of
The University of Manitoba
In partial Fulfillment of the Requirements of the Degree of

Doctor of Philosophy

Department of Microbiology
University of Manitoba
Winnipeg, Manitoba, Canada

“You have learned the lesson by experience.”

— Aeschylus, The Seven Against Thebes

“Back off, man! I’m a scientist!”

— Bill Murray, Ghostbusters

“I’ve already got my degree.”

— Ivan Oresnik, literally every chance he gets

Abstract

Symbiotic nitrogen fixation by diazotrophic bacteria is an environmentally benign alternative to synthetic fertilizer application to agricultural fields. The effectiveness of inoculum strains can be reduced by competition with native strains present in the soil. *Sinorhizobium meliloti* is a soil bacterium that can establish a nitrogen fixing symbiosis with the forage crop *Medicago sativa* and the model legume *Medicago truncatula*. The ability to utilize organic compounds has been shown to be important for competition for nodule occupancy in *S. meliloti* and other symbiotic bacteria. Here, genetic and biochemical techniques are used to examine the metabolism of several substrates in the model organism *S. meliloti*. Genetic loci involved in sugar alcohols galactitol as well as sorbitol, mannitol, and D-arabitol were identified and characterized. Special attention was paid to sorbitol dehydrogenase SmoS and a crystal structure with a resolution of 2.0 Å was generated. These experiments led to the characterization of the metabolism of D-arabinose and L-fucose, which is carried out in part by loci that are involved in mannitol and D-arabitol utilization. Additionally, while *S. meliloti* cannot utilize the sugar alcohol xylitol, it was observed that spontaneous mutations would arise at a modest frequency that permitted growth on xylitol. This mutation increases the metabolic capacity of the organism, and was genetically characterized and determined to be related to the movement of an IS*Rm2011-2* insertion sequence element. These results contribute to our understanding of the metabolism of the model organism *S. meliloti* and further efforts to engineer symbiosis in other organisms.

Acknowledgments

I would like to acknowledge my supervisor, Ivan Oresnik. His guidance and mentorship were invaluable during my studies. Though I suspect we would talk as much about 80's comedy movies as we would science, I would not have been able to achieve any measure of academic success without his support. He manages to curate an environment that is fun but also promotes learning, curiosity, and productivity, and I could not have asked for a more suitable labtoratory to work in.

Thank you to my committee members Brian Mark, Richard Sparling, and Jake Stout, for their time and advice. My committee never failed to make me feel as though I was investigating a new and exciting facet of microbiology, even when I didn't believe it myself.

I would also like to thank the past and present members of the Oresnik lab. Barney, Damien, Harry, and Justin made for an eclectic group of researchers and mentors that eased me into the laboratory. Also, Sabhjeet, Gagan, Patricia, and Derek deserve praise for putting up with me as I grumbled my way through Grad school.

I am thankful for the support of all my friends and family. In particular I would like to thank my friend Drew for his valiant but ultimately unsuccessful attempts to keep me sane.

Finally, I would like to thank all the members of the Department of Microbiology at the University of Manitoba. This department is a friendly and welcoming place and whether you need another pair of hands in the lab or a drink after work, someone is there to help.

Dedication

This work is dedicated to my parents, Glen and Jeannet Kohlmeier. Without the two of them, none of this would be possible. Their contributions cannot be overstated. Thank you both.

Table of Contents

Abstract....i
Acknowledgments....ii
Dedication....iii
Table of Contents....iv
List of Tables....ix
List of Figures....x
List of Abbreviations....xii
Chapter 1: Literature Review....1
1.1 The nitrogen problem....2
1.1.1 Nitrogen fixation....2
1.1.2 Problems associated with agricultural practices....4
1.1.3 Potential alternatives....6
1.2 Symbiotic nitrogen fixation....7
1.2.1 Rhizobia....7
1.2.2 *Sinorhizobium meliloti* Rm1021....10
1.2.3 *Medicago*....12
1.2.4 Symbiosis....18
1.2.5 Extension of N_2 fixation into nonlegumes....32
1.3 Carbon metabolism....34
1.3.1 Free-living metabolism....34
1.3.2 Metabolism in the infection thread....36
1.3.3 Bacteroid metabolism....38
1.4 Sugar alcohols....42
1.4.1 Sorbitol, mannitol, and D-arabitol....42
1.4.2 Galactitol....43
1.4.3 Erythritol, adonitol, and L-arabitol....43
1.4.4 Glycerol....44
1.4.5 Cyclic polyols....44
1.5 Thesis objectives....46

Chapter 2: Galactitol utilization in *Sinorhizobium meliloti* is dependent on a chromosomally encoded sorbitol dehydrogenase and a pSymB encoded operon necessary for tagatose catabolism 48

2.1 Abstract.....49

2.2 Introduction 50

2.3 Materials and methods.....52

2.3.1 Bacterial strains, plasmids, and culture conditions.....52

2.3.2 DNA manipulations and genetic techniques 54

2.3.3 Primer extension assays.....56

2.3.4 Biochemical enzyme assays 57

2.3.5 Growth curve analysis 58

2.3.6 Transport assays 58

2.3.7 Competition for nodule occupancy assays 59

2.3.8 Phylogenetic analysis 59

2.4 Results 60

2.4.1 Identification of an operon necessary for tagatose catabolism.....60

2.4.2 Identification of a negative regulator and the transcriptional start site of *SMb21377*..63

2.4.3 SMb21374 and SMb21373 encode proteins with tagatose kinase and tagatose-6-phosphate epimerase activities 66

2.4.4 SmoS is responsible for galactitol oxidation in *S. meliloti*.....69

2.4.5 Galactitol competes with mannitol for transport.....73

2.4.6 Galactitol mutants are equally competitive for nodule occupancy.....74

2.4.7 *E. coli* GatZ has tagatose-6-phosphate epimerase activity.....76

2.4.8 IPR012062 is a family of tagatose-phosphate epimerases 78

2.5 Discussion.....82

Chapter 3: Characterization of sorbitol dehydrogenase SmoS from *Sinorhizobium meliloti* 1021 87

3.1 Abstract.....88

3.2 Introduction 89

3.3 Materials and methods.....91

3.3.1 Bacterial strains and culture conditions.....91

3.3.2 Overexpression and purification of SmoS.....91

3.3.3 SmoS crystallization 93
3.3.4 X-ray data collection and structure solution 94
3.3.5 Enzyme assays 94
3.3.6 Ligand docking analysis 95
3.4 Results 95
3.4.1 Structural characterization of *Sm*SmoS 95
3.4.2 SmoS has a high pH optimum and a preference for sorbitol 102
3.4.3 SmoS-tagatose complex is predicted to be in a lower energy state than SmoS-fructose complex 106
3.5 Discussion 109
Chapter 4: Physiological characterization of a locus responsible for the metabolism of multiple sugar alcohols in *Sinorhizobium meliloti* 113
4.1 Abstract 114
4.2 Introduction 115
4.3 Materials and methods 117
4.3.1 Bacterial strains, plasmids, and culture conditions 117
4.3.2 DNA manipulations and genetic techniques 117
4.3.3 Protein purification and biochemical enzyme assays 122
4.3.4 Transport assays 123
4.3.5 Fluorescence gene expression analysis 123
4.4 Results 124
4.4.1 *smoS* is responsible for growth with sorbitol and galactitol while *mtlK* is responsible for mannitol and D-arabitol utilization 124
4.4.2 SmoS and MtlK are polyol dehydrogenases 129
4.4.3 Sorbitol and D-arabitol strongly compete with mannitol for transport 131
4.4.4 The *smo* locus is negatively regulated by SmoC 133
4.4.5 *frk* encodes a fructose kinase 136
4.4.6 Phosphoglucose isomerase gene is downstream of the *frc* locus 138
4.4.7 Sorbitol and mannitol can be transported by the *frc* transporter 141
4.4.8 FrcK contributes to fructose uptake and mutations in *frcK* or *frk* permit transport of mannitol under non-inducing conditions 142
4.5 Discussion 146

Chapter 5: Metabolism of D-arabinose and L-fucose in *Sinorhizobium meliloti* 1021 152
5.1 Abstract.....153
5.2 Introduction 154
5.3 Materials and methods.....155
5.3.1 Bacterial strains, plasmids, and culture conditions.....155
5.3.2 DNA manipulations and genetic techniques 157
5.3.3 Biochemical enzyme assays 159
5.3.4 Plant assays.....159
5.4 Results 160
5.4.1 *smo* mutants exhibit reduced growth on D-arabinose which can be restored by complementation with *mtlK*.....160
5.4.2 Mutations at the *SMb21103*-*SMb21113* locus exhibit reduced growth on D-arabinose and no growth on L-fucose.....162
5.4.3 *SMc00680* encodes a D-arabinose reductase.....164
5.4.4 Strains with mutations at both loci are unable to grow on D-arabinose as a sole carbon source.....165
5.4.5 *SMb21103*-*SMb21113* encodes a non-phosphorylative D-arabinose and L-fucose pathway.....165
5.4.6 Discoloration of the growth medium by SRmD622 is linked to medium acidification 167
5.4.7 SMb21109 has D-arabinose and L-fucose dehydrogenase activity.....172
5.4.8 Δ*SMb21111* strain cannot utilize D-arabinose but does grow on L-fucose 173
5.4.9 Plant growth is not significantly altered by inoculation with D-arabinose/L-fucose mutants.....175
5.5 Discussion.....177
Chapter 6: Movement of an insertion sequence element is correlated with increased catabolic activity in *Sinorhizobium meliloti* 1021.....179
6.1 Abstract.....180
6.2 Introduction 181
6.3 Materials and methods.....183
6.3.1 Bacterial strains, plasmids, and culture conditions.....183
6.3.2 DNA manipulations and genetic techniques 183
6.3.3 PacBio sequencing.....186

6.3.4 Dehydrogenase assays 188
6.4 Results 188
6.4.1 Wildtype *S. meliloti* cannot utilize xylitol but spontaneous mutations that permit growth can arise 188
6.4.2 *xlt-1* maps to the chromosome of Rm1021 190
6.4.3 Xylitol is oxidized to xylulose and phosphorylated becoming xylulose-5-phosphate 194
6.4.4 *SMc01991* and *SMc01992* encode xylitol dehydrogenases 199
6.4.5 SRmD268 contains an IS*Rm2011-2* insertion sequence element within *SMc01990* .201
6.4.6 Loss of *SMc1990* is not responsible for the gain of xylitol catabolic function 202
6.4.7 Insertions in *SMc01990* are responsible for the gain of function 204
6.5 Discussion 206
Chapter 7: Conclusions 208
7.1 Thesis conclusions and observations 209
Literature cited 219

List of Tables

Table 2.1 Bacterial strains and plasmids.....53
Table 2.2 Primers used during this study.....55
Table 2.3 Tagatose operon induction.....64
Table 2.4 Carbon phenotypes of *smoS* mutants.....71
Table 3.1 Crystallization and refinement statistics for the SmoS and SmoS-sbt structures.....99
Table 3.2 Kinetic properties of *S. meliloti* SmoS.....105
Table 4.1 Bacterial strains and plasmids.....118
Table 4.2 Primers used during this study.....121
Table 4.3 Complementation analysis of *smo* mutants.....126
Table 4.4 Carbon phenotypes of *smo* and *mpt* transport mutants.....128
Table 4.5 Complementation of putative fructose mutants.....139
Table 4.6 Carbon phenotypes of putative fructose transport mutants.....143
Table 5.1 Bacterial strains and plasmids.....156
Table 5.2 Primers used during this study.....158
Table 5.3 Carbon phenotypes of mutant strains.....161
Table 6.1 Bacterial strains and plasmids.....184
Table 6.2 Primers used during this study.....187
Table 6.3 Frequencies and map distances for strains in which alleles can be linked in transduction.....193
Table 6.4 Carbon phenotypes.....203

List of Figures

Figure 1.1 Global nitrogen fixation 3
Figure 1.2 Phylogeny of selected proteobacteria showing the distribution of rhizobia 8
Figure 1.3 The three components of the *Sinorhizobium meliloti* genome 11
Figure 1.4 Developmental stages of indeterminate and determinate legume nodules. 14
Figure 1.5 Legume and rhizobia conflict over nodule number 16
Figure 1.6 Rhizobia interacting with legumes 19
Figure 1.7 Representative nodulation (Nod) factors 22
Figure 1.8 Predicted kinases that are required for Nod factor signaling 25
Figure 1.9 Nitrogenase structure and genes required for its biosynthesis 30
Figure 1.10 Schematic representation of the nitrogenase Fe protein cycle 31
Figure 1.11 Schematics of carbon and nitrogen metabolic pathways with key enzymes, metabolites, and transporters in determinate nodules and indeterminate nodules 39
Figure 2.1 Locus diagram of *SMb21377-SMb21372* (*tag*) operon housing genes involved in galactitol catabolism. 61
Figure 2.2 Analysis of the *SMb21377* (*tagA*) promoter 65
Figure 2.3 Growth curves comparing *A. tumefaciens* strains grown in defined media with galactitol as a sole carbon source 68
Figure 2.4 Non-denaturing PAGE of galactitol dehydrogenase activity 72
Figure 2.5 Uptake of ^{14}C-mannitol in *S. meliloti* 75
Figure 2.6 Structural comparison of SMb21373 (PDBID: 3TXV) from *S. meliloti* (green) and GatZ (PDBID: 2FIQ) from *E. coli* (blue) 77
Figure 2.7 Growth curves comparing *A. tumefaciens* strains grown in defined media with galactitol as a sole carbon source 80
Figure 2.8 Phylogenetic tree of 2266 amino acid sequences within the D-tagatose-bisphosphate aldolase, non-catalytic subunit GatZ/KbaZ (IPR012062) protein family 81
Figure 2.9 Proposed pathway of galactitol utilization in *S. meliloti* 83
Figure 3.1 Enzymatic reactions catalyzed by SmoS 92
Figure 3.2 Size exclusion chromatography of purified SmoS from a Superdex 75 gel filtration column and analysis by polyacrylamide gel electrophoresis (PAGE) 96
Figure 3.3 Crystal structure of SmoS from *S. meliloti* 1021 98
Figure 3.4 Crystal structure of the SmoS-sbt complex 101
Figure 3.5 Effect of pH on SmoS dehydrogenase activity 103
Figure 3.6 Kinetic analysis of sorbitol and galactitol oxidation by *S. meliloti* SmoS. 104
Figure 3.7 SmoS-fructose and SmoS-tagatose binding complexes predicted by the Ligand docking protocol housed on the ROSIE server 108
Figure 3.8 Comparison of the position and distribution of proline residues 112
Figure 4.1 Locus diagrams of the *smo* and *frc* loci 116
Figure 4.2 Non-denaturing PAGE gel of sorbitol inducible dehydrogenase activity 130
Figure 4.3 Transport rates of ^{14}C-mannitol 132

Figure 4.4 Induction of the *smo* locus measured by GFP relative fluorescence..... 135
Figure 4.5 Fructose kinase activity of partially purified Frk..... 140
Figure 4.6 Transport rates of ^{14}C-fructose and ^{14}C-mannitol..... 145
Figure 4.7 Pathways for the metabolism of sorbitol, mannitol, and D-arabitol..... 147
Figure 5.1 Locus diagram of the pSymB encoded *SMb21101-13* region..... 163
Figure 5.2 Non-phosphorylative metabolism of D-arabinose and L-fucose..... 168
Figure 5.3 Discoloration of agar and broth media by SRmD622 utilizing D-arabinose..... 169
Figure 5.4 Medium acidification and its effect on nodulation kinetics..... 170
Figure 5.5 Non-denaturing PAGE gels showing L-fucose inducible dehydrogenase activity. 174
Figure 5.6 Dry weight of alfalfa plants inoculated with different *S. meliloti* strains..... 176
Figure 6.1 Growth of wildtype Rm1021 and suppressor mutant SRmD268 on LB (Luria-Bertani) complex medium and VMM (Vincent's minimal medium) defined medium supplemented with 15 mM xylitol as a sole carbon source..... 188
Figure 6.2 Schematic diagram featuring the core concept behind the cotransduction mapping strategy..... 190
Figure 6.3 Genetic map of the region containing the suppressor mutation..... 194
Figure 6.4 Xylitol utilization pathway by *Klebsiella pneumonia*..... 195
Figure 6.5 Locus diagram of the *SMc02022-SMc01990* region..... 198
Figure 6.6 Xylitol dehydrogenase activity..... 200
Figure 6.7 Targeted mutagenesis of *SMc01990* with a suicide vector results in a strain with the ability to utilize xylitol as a sole carbon source..... 205
Figure 7.1 Possible models for the regulation of mannitol uptake involving F6P..... 215

List of Abbreviations

3PG	3-Phosphoglycerate
6PG	6-Phosphogluconate
AAT	Aspartate aminotransferase
ABC	ATP-binding cassette
AcCoA	Acetyl coenzyme A
ADP	Adenosine diphosphate
ara	arabinose
AS	Asparagine synthetase
Asn	Asparagine
Asp	Aspartate
ATP	Adenosine triphosphate
BNF	Biological nitrogen fixation
bp	Base pair
Bq	Becquerel
C	Carbon
Ca^{2+}	Calcium
CAPS	N-cyclohexyl-3-aminopropanesulfonic acid
CBB	Calvin Benson Bassham
CCRH	Curled colonized root hair
cDNA	Complementary DNA
Ci	Curies
cm	Centimeters
Cm	Chloramphenicol
CO_2	Carbon dioxide
cpm	Counts per minute
Da	Daltons
DHAP	Dihydroxyacetone phosphate
DNA	Deoxyribonucleic acids
dNTP	Deoxynucleotide triphosphate
e^-	electron
em	emission
ery	erythritol
ex	excitation
ED	Entner-Doudoroff
EDTA	Ethylenediaminetetraacetic acid
EPS	Exopolysaccharide
EPSI	Succinoglycan
EPSII	Galactoglucan
F1,6bisP	Fructose-1,6-bisphosphate
F6P	Fructose-6-phosphate
Fe	Iron
FPLC	Fast protein liquid chromatography
fru	fructose
fuc	fucose

g	Grams
G3P	Glycerol-3-phosphate
G6P	Glucose-6-phosphate
GAP	Glyceraldehyde-3-phosphate
GalNAc	*N*-acetyl-D-galactosamine
GlcNAc	*N*-acetyl-D-glucosamine
glc	Glucose
Gln	Glutamine
Glu	Glutamate
gly	glycerol
Gm	Gentamicin
GNG	Gluconeogenesis
GOGAT	Glutamine oxoglutarate aminotransferase
GS	Glutamine synthetase
gtl	galactitol
h	Hours
H	Hydrogen
H^+	Hydrogen ion
H_2	dihydrogen
H_2O	Water
HPLC	High performance liquid chromatography
IM	Inner membrane
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
IRLC	Inverted repeat lacking clade
IT	Infection thread
kb	Kilobase
KDA	2-keto-3-deoxy-D-arabinoate
kDa	Kilodaltons
KDPG	2-keto-3-deoxy-6-phosphogluconate
KDPgal	2-keto-3-deoxy-6-phosphogalactonate
Km	Kanamycin
L	Liter
LB	Luria-Bertani
lb/in^2	Pounds per square inch
LDH	Lactate dehydrogenase
LPS	Lipopolysaccharide
M	Molar
MBq	Megabecquerel
mCi	Millicurie
MES	2-(N-morpholino)ethanesulfonic acid
mg	Milligram
min	Minute
ML	Maximum likelihood
mL	Milliliter
mM	Millimolar
mmol	Millimole

Mo	Molybdenum
mol	Mole
MOPS	3-(N-morpholino)propanesulfonic acid
mtl	mannitol
N	Nitrogen
N_2	Dinitrogen
Na	Sodium
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NAD^+	Nicotinamide adenine dinucleotide
NADH	Reduced nicotinamide adenine dinucleotide
$NADP^+$	Nicotinamide adenine dinucleotide phosphate
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
NCR	Nodule specific cysteine rich peptides
nd	not determined
NF	Nod factor
NH_3	Ammonia
NH_4^+	Ammonium
Ni-NTA	Nickel nitrilotriacetic acid
nL	nanolitre
Nm	Neomycin
nmol	Nanomole
NO	Nitric oxide
NO_2	Nitrogen dioxide
NO_3^-	Nitrate
NO_x	Nitrogen oxides
NTP	Nucleotide triphosphate
NUE	Nitrogen use efficiency
OAA	oxaloacetate
OM	Outer membrane
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEP	Phosphoenolpyruvate
PEPC	Phosphoenolpyruvate carboxylase
PHB	Polyhydroxybutyrate
P	Phosphate
P_i	Inorganic phosphate
PK	Pyruvate kinase
PP	Pentose phosphate
PQQ	Pyrroloquinoline quinone
PTS	Phosphotransferase system
qRT-PCR	Quantitative reverse transcriptase PCR
R5P	Ribulose-5-phosphate
RAxML	Randomized axelerated maximum likelihood
Rf	Rifampicin

RNA	Ribonucleic acid
RT-PCR	Reverse transcriptase PCR
sbt	sorbitol
s	Seconds
SD	Standard deviation
Sm	Streptomycin
SM	Symbiosome membrane
SMCR	Succinate mediate catabolite repression
SNF	Symbiotic nitrogen fixation
Sp	Spectinomycin
sp	Species
spp	Species (pl)
SS	Sucrose synthase
tag	tagatose
Tc	Tetracycline
TCA	Tricarboxylic acid
TE	Tris-EDTA
Tg	Tetragram
Tn	Transposon
Tris	tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
V	Vanadium
VMM	Vincent's minimal medium
vol	volume
UDP-Glu	Uridine diphosphate glucose
UV	Ultraviolet
VLCFA	Very long chain fatty acid
wt	wildtype
X5P	Xylulose-5-phosphate
xlt	xylitol
yr	year
µCi	Microcurie
µg	Microgram
µL	Microliter
µm	Micrometer
µM	Micromolar
µmol	Micromole

Chapter 1:

Literature Review

1.1 The nitrogen problem

1.1.1 Nitrogen fixation

Nitrogen is an important component of many biomolecules, including photosynthetic chlorophyll, nucleic acids (DNA and RNA), and protein. These molecules can be synthesized intracellularly from base components taken up from the environment; however, most nitrogen is "locked away" in the atmosphere as dinitrogen gas. Approximately 78% of the atmosphere by volume, roughly 4 x 10^9 Tg, is made up of dinitrogen gas (Sorai *et al.*, 2007). Due to the high activation energy required to break the covalent triple bond, N_2 gas is mostly inert and inaccessible to biological systems. The process of liberating nitrogen, and transforming it into a biologically active form, is called nitrogen fixation.

There are three major forms of nitrogen fixation: atmospheric, industrial, and biological. Atmospheric nitrogen fixation occurs primarily due to lightening strikes, which provide the energy required to convert N_2 into nitrogen oxides. Approximately 5 Tg N yr^{-1} are produced by lightening (Fig. 1.1) (Shepon *et al.*, 2007). Industrial nitrogen fixation is facilitated by the Haber-Bosch process, a chemical reaction in which N_2 gas is reduced to ammonia (NH_3) under high temperature and pressure in the presence of an iron catalyst (Erisman *et al.*, 2008). Fritz Haber invented the process in 1908; it was subsequently scaled up for industrial purposes by Carl Bosch. Both men were awarded Nobel prizes for their contributions to the technique (Erisman *et al.*, 2008). Approximately 120 Tg N yr^{-1} are produced by the Haber-Bosch process, of which 80% is used to produce fertilizer and the remaining 20% for feedstock (Fig. 1.1) (Galloway *et al.*, 2008). Biological nitrogen fixation (BNF) is mediated by microorganisms called diazotrophs that are capable of enzymatically converting nitrogen gas into ammonia. BNF is an exclusively prokaryotic process, limited predominantly to bacteria and a few archaea (Boyd & Peters, 2013).

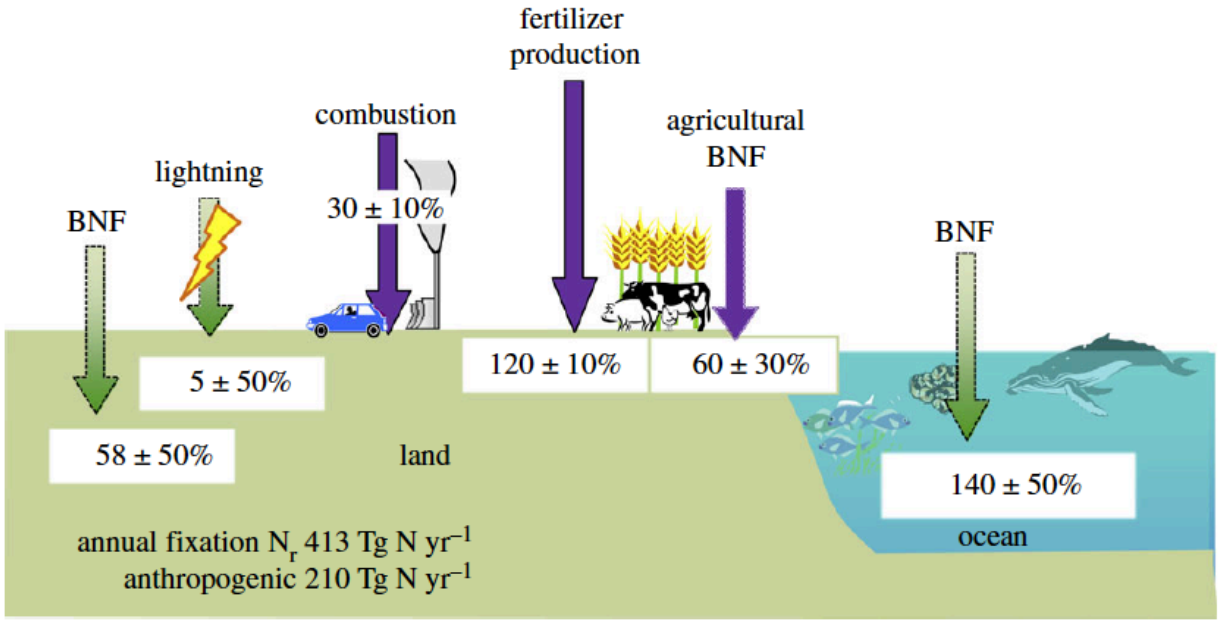


Figure 1.1. Global nitrogen fixation, natural and anthropogenic in both oxidized and reduced forms through combustion, biological fixation, lightning and fertilizer, and industrial production through the Haber–Bosch process for 2010. The arrows indicate a transfer from the atmospheric N_2 reservoir to terrestrial and marine ecosystems, regardless of the subsequent fate of the Nr. Green arrows represent natural sources, purple arrows represent anthropogenic sources.

BNF occurs in both terrestrial and marine ecosystems, with fixation that occurs on land within or in close association with plants referred to as symbiotic nitrogen fixation (SNF).

The nitrogenase enzyme, a unifying feature of diazotrophs, is responsible for catalyzing the N_2 fixation reaction. Nitrogenase expends 16 ATP to reduce N_2 gas into ammonia, and can be represented by the equation:

$$N_2 + 8\ H^+ + 8\ e^- + 16\ ATP \rightarrow 2\ NH_3 + H_2 + 16\ ADP + 16\ P_i$$

Adenosine triphosphate (ATP) can be considered the "energy currency" of biological systems, making the reaction a costly endeavor for the bacterium. Generation of ATP by symbiotic bacteria requires the oxidation of organic compounds, which are secreted into the soil by the host plant. The area of the soil affected by plant root exudate is called the rhizosphere. That the organism would spend so much energy to perform the reaction is indicative of the strength of the N_2 triple bond, but also speaks to the overall importance of generating reduced nitrogen. This reaction can be contrasted with the Haber-Bosch process in that it is performed at ambient temperature and pressure. A total of 260 Tg N yr^{-1} are fixed via BNF, including natural (200 Tg N yr^{-1}) and agricultural sources (60 Tg N yr^{-1}) (Fig. 1.1) (Herridge *et al.*, 2008; Vitousek *et al.*, 2013; Voss *et al.*, 2013).

1.1.2 Problems associated with agricultural practices

The Haber-Bosch process led to an enormous increase in agricultural productivity in most regions of the world (Stewart *et al.*, 2005). Higher agricultural yields have coincided with a growing human population. At the turn of the 20th century the population was approximately 1.5 billion people. It has climbed to 7.2 billion over the last 120 years. It has been estimated that 48% of the population is supported by Haber-Bosch nitrogen (Erisman *et al.*, 2008).

It is indisputable that Haber-Bosch nitrogen has been instrumental in meeting the population's nutritional requirements, but the benefits have come with significant environmental costs. Reactive nitrogen (Nr), defined as all forms of nitrogen excluding N_2, is accumulating in the environment due to the rate of Nr generation greatly exceeding the rate of Nr denitrification into N_2 (Galloway *et al.*, 2003; Sutton *et al.*, 2013). Its been estimated that greater than 50% of the Nr applied to fields is lost to the environment (Lassaletta *et al.*, 2014). Losses of reactive nitrogen typically occur in the form of nitrate (NO_3^-), organic nitrogen species (eg. urea, amines, proteins, and nucleic acids), nitrogen oxides (NO_x, which refers to the sum of nitric oxide (NO) and nitrogen dioxide (NO_2)), nitrous oxide (N_2O), and ammonia (NH_3) (Davidson *et al.*, 2015; Galloway *et al.*, 2003). NO_3^- and organic nitrogen leach into water bodies and contribute to eutrophication and algal blooms in aquatic ecosystems. NO_3^- is a regulated drinking water pollutant, which should be below 10 mg/L according to the United States Environmental Protection Agency and the Ontario Drinking Water Standards. NO_x, N_2O, and NH_3 are leached from soil by emission into the atmosphere. NO_x is a precursor to tropospheric ozone pollution, and additionally, both NO_x and NH_3 contribute to the buildup of particulate matter reducing air quality and increasing N deposition in downwind ecosystems. N_2O is a greenhouse gas that can contribute to climate change, in addition to depletion of stratospheric ozone (Davidson *et al.*, 2015; Galloway *et al.*, 2003). These disruptions to the global nitrogen cycle have exceeded safe operating parameters and crossed a threshold that could lead to unacceptable environmental change (Rockström *et al.*, 2009; Steffen *et al.*, 2015).

Its been estimated that the population will be as high as 9.5 billion people by the year 2050, suggesting that humanity will need to further increase its agricultural productivity. Fertilizers are also required for the generation of bioenergy and biofuels, and while biofuels do

not currently provide a large proportion of the global energy requirement, their production is expected to increase, creating another strain on an already exerted nitrogen budget (Erisman *et al.*, 2008).

1.1.3 Potential alternatives

Nitrogen use efficiency (NUE) is defined as the ratio of nitrogen removed from fields during harvest to the amount of nitrogen added. Improving agricultural practices would mean decreasing N surplus, which could be lost to the environment, while increasing NUE and crop yields. A number of current technologies have great potential to increase NUE including controlled-release fertilizers, urease and nitrification inhibitors, improved irrigation and water management, improved soil and plant testing to match nutrient applications with crop demands, use of winter cover crops, and precision agriculture technologies (Davidson *et al.*, 2015). While these techniques have been used with limited success, there remain significant social and economic barriers in place that reduce their widespread adoption and efficacy (Davidson *et al.*, 2015).

In addition to the implementation of modern agricultural practices, biotechnological strategies are another tactic under investigation to reduce N surplus. In general, regions with a higher contribution of Nr from SNF have a better NUE (Lassaletta *et al.*, 2014). Currently, SNF is largely limited to legume plants, but there is growing interest in extension of SNF to non-legumes. A number of approaches are being utilized to achieve this goal, including direct expression of bacterial nitrogenase genes in plants, engineering non-legume plants to establish symbiosis with diazotrophic bacteria, and the development of novel plant-microbe interactions (Mus *et al.*, 2016).

1.2 Symbiotic nitrogen fixation

1.2.1 Rhizobia

Soil bacteria with the ability to establish N fixing symbiosis with legumes are collectively referred to as rhizobia. These bacteria fall into two orders, α proteobacteria and β proteobacteria, and currently encompass 14 genera and hundreds of species (Fig. 1.2) (Masson-Boivin *et al.*, 2009; Peix *et al.*, 2015). Some of the most well studied members include *Sinorhizobium*, *Rhizobium*, *Bradyrhizobium*, *Mesorhizobium*, and *Azorhizobium*. These bacteria in combination with their hosts are useful model systems for symbiotic research. Common model systems include the *Sinorhizobium meliloti-Medicago truncatula* model (Jones *et al.*, 2007) and the *Mesorhizobium loti-Lotus japonicus* model (Markmann *et al.*, 2012). The only notable exception to the proteobacterial N fixing symbiotic monopoly are actinobacteria of the genus *Frankia*, that are able to fix N while engaged in a symbiotic relationship with actinorhizal plants (Pawlowski & Demchenko, 2012). These bacteria are not nearly as numerous in a taxonomic sense as their rhizobial counterparts (Remigi *et al.*, 2016).

During symbiosis, rhizobia elicit the formation of organs termed nodules on the roots of host plants; the nodules are colonized intracellularly and are the site of N fixation. Apart from the presence of genes necessary for nodulation (*nod*/*nol*/*noe*) and N fixation (*nif*/*fix*/*fdx*), there are few features shared between all rhizobia. Even the *nod* genes may be dispensable, as strains capable of nodulation and symbiosis in the absence of canonical *nod* genes have been reported (Giraud *et al.*, 2007). However, some common features of rhizobia are large (5-10 Mbp) genomes that have high plasticity and a robust capacity for regulation, metabolism, transport, and stress response (Remigi *et al.*, 2016).

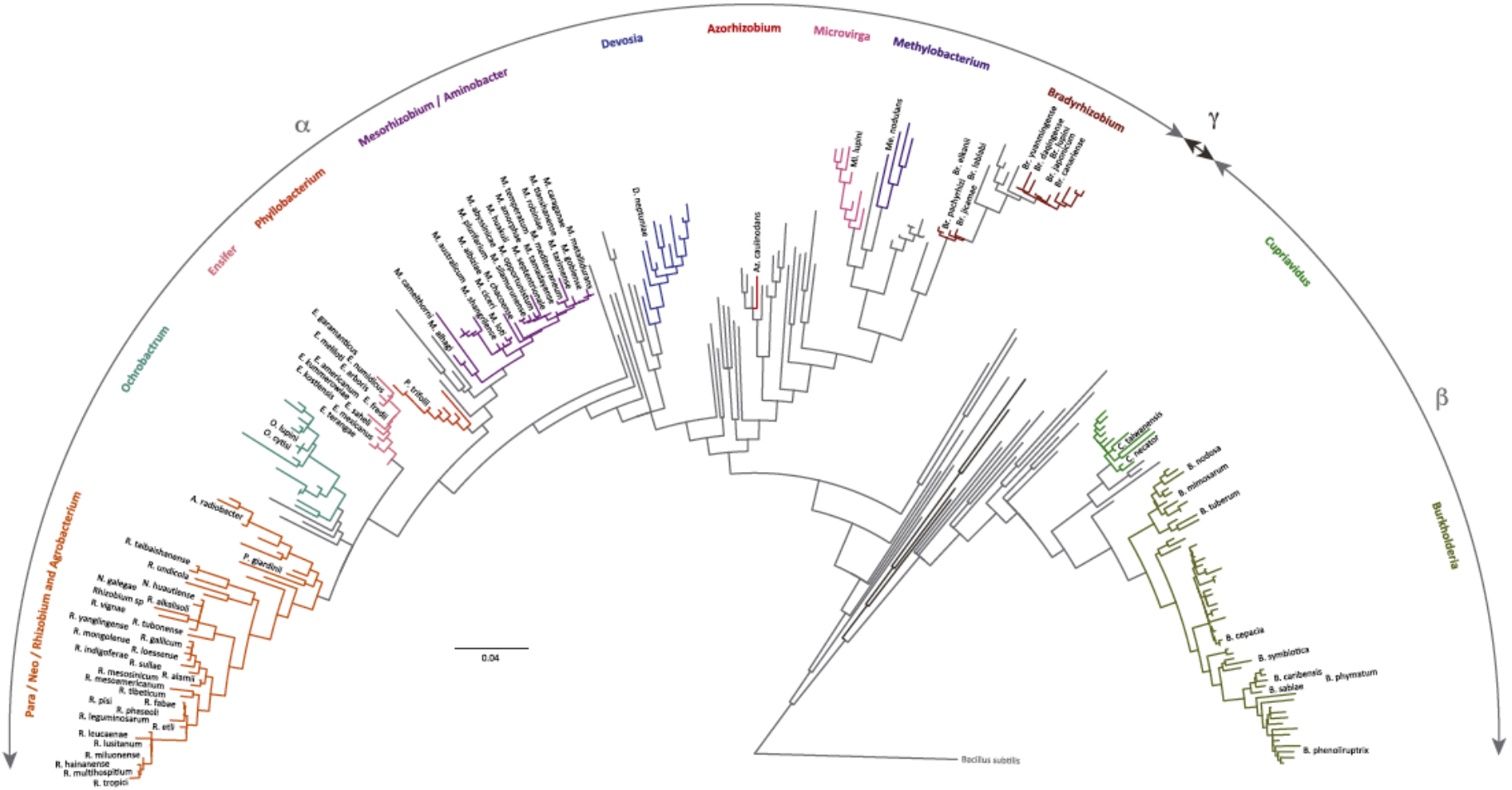


Figure 1.2. Phylogeny of Selected Proteobacteria Showing the Distribution of Rhizobia. Tree constructed from 16S rRNA sequences of type strains from RDP 11.4 (http://rdp.cme.msu.edu/) by Mega 6.06 using 1431 aligned positions and Maximum Likelihood based on the Tamura-Nei model. Taxa/genera currently known to include rhizobia are in color, and rhizobial species are indicated. In addition, other species in the same genus are included, plus a single representative of each genus, family, or order that has no known rhizobia, for the classes Alphaproteobacteria and Betaproteobacteria. Selected representatives of the Gammaproteobacteria are shown.

The genes responsible for nodulation and N fixation are often located on accessory plasmids (Higashi, 1967; Johnston *et al.*, 1978) or symbiosis islands (Kaneko *et al.*, 2002; Sullivan *et al.*, 2002) and are therefor subject to increased mobility and horizontal gene transfer (HGT) events. The distribution of these genes within only 14 genera seems small given the high frequency of HGT events within the rhizosphere (Kroer *et al.*, 1998; van Elsas *et al.*, 2003) as well as the abundance and diversity of soil bacteria (Delgado-Baquerizo *et al.*, 2018). Phylogenetic evidence suggests that three types of transfer events have led to the current distribution of symbiotic genes: frequent transfer within genera, infrequent transfer between genera, and very infrequent, possibly unique (Chen *et al.*, 2003), transfer between α proteobacteria and β proteobacteria (Remigi *et al.*, 2016). This suggests that the transfer of symbiotic genes is not sufficient for the generation of a symbiotic organism, and that the genome of the recipient must be predisposed for receipt of the genes for an effective symbiont to be created. As an example, exopolysaccharide (EPS) production has been shown to be essential for symbiosis in *Sinorhizobium meliloti* (Leigh *et al.*, 1985; Pellock *et al.*, 2000). However, the genes necessary for EPS production in *S. meliloti* are absent in other rhizobial lineages (Tian *et al.*, 2012), suggesting that EPS production may be an innate skill of *Sinorhizobium* that has been adapted for symbiosis. However, each host has unique challenges for a potential symbiotic partner to overcome such as plant immunity (Zipfel, 2014) as well as physiology and metabolism (Udvardi & Poole, 2013), indicating that a long period of co-evolution likely occurs in which the interaction can be fine-tuned (Remigi *et al.*, 2016).

1.2.2 Sinorhizobium meliloti *Rm1021*

S. meliloti Rm1021, also called *Ensifer meliloti* (Fig. 1.2), is a Gram-negative α proteobacterium thatt can be found as a free-living bacterium in the soil or as a participant in an endosymbiotic relationship with legumes of the *Medicago*, *Melilotus*, and *Trigonella* genera. *S. meliloti* was originally isolated from New South Wales, Australia in 1937 and termed SU47, Rm1021 is a streptomycin resistant derivative of the SU47 parental strain (Geddes & Oresnik, 2016; Meade & Signer, 1977; Meade *et al.*, 1982).

Rm1021 has a tripartite genome consisting of a chromosome (3,654,135 bp), a megaplasmid called pSymA (1,354,226 bp), and a chromid called pSymB (1,683,333 bp) (Fig. 1.3) (Capela *et al.*, 2001; diCenzo & Finan, 2017; Downie & Young, 2001; Finan *et al.*, 2001; Galibert *et al.*, 2001). The majority of the essential and housekeeping genes are included on the chromosome (Capela *et al.*, 2001). pSymA contains the genes for nodulation and nitrogen fixation and is often referred to as the symbiotic megaplasmid, but no absolutely essential genes are found on this replicon; in fact it can be removed entirely without major repercussions to the organism (Oresnik *et al.*, 2000). A large proportion of the genes on pSymB are dedicated to solute uptake and polysaccharide synthesis, suggesting that the chromid is important for survival in diverse environments (Finan *et al.*, 2001). The pSymB genes *engA* and $tRNA^{arg}$ were shown to be essential for growth (diCenzo *et al.*, 2013), but if copies of these genes were previously integrated into the chromosome, the entire chromid could be deleted (diCenzo *et al.*, 2014). This realization led to the creation of a pSymAB double cured strain, in which 45% of the genome was removed. This mutant was compromised for growth on a number of carbon, nitrogen, phosphorous, and sulfur sources but was able to grow on minimal media as well as in bulk soil.

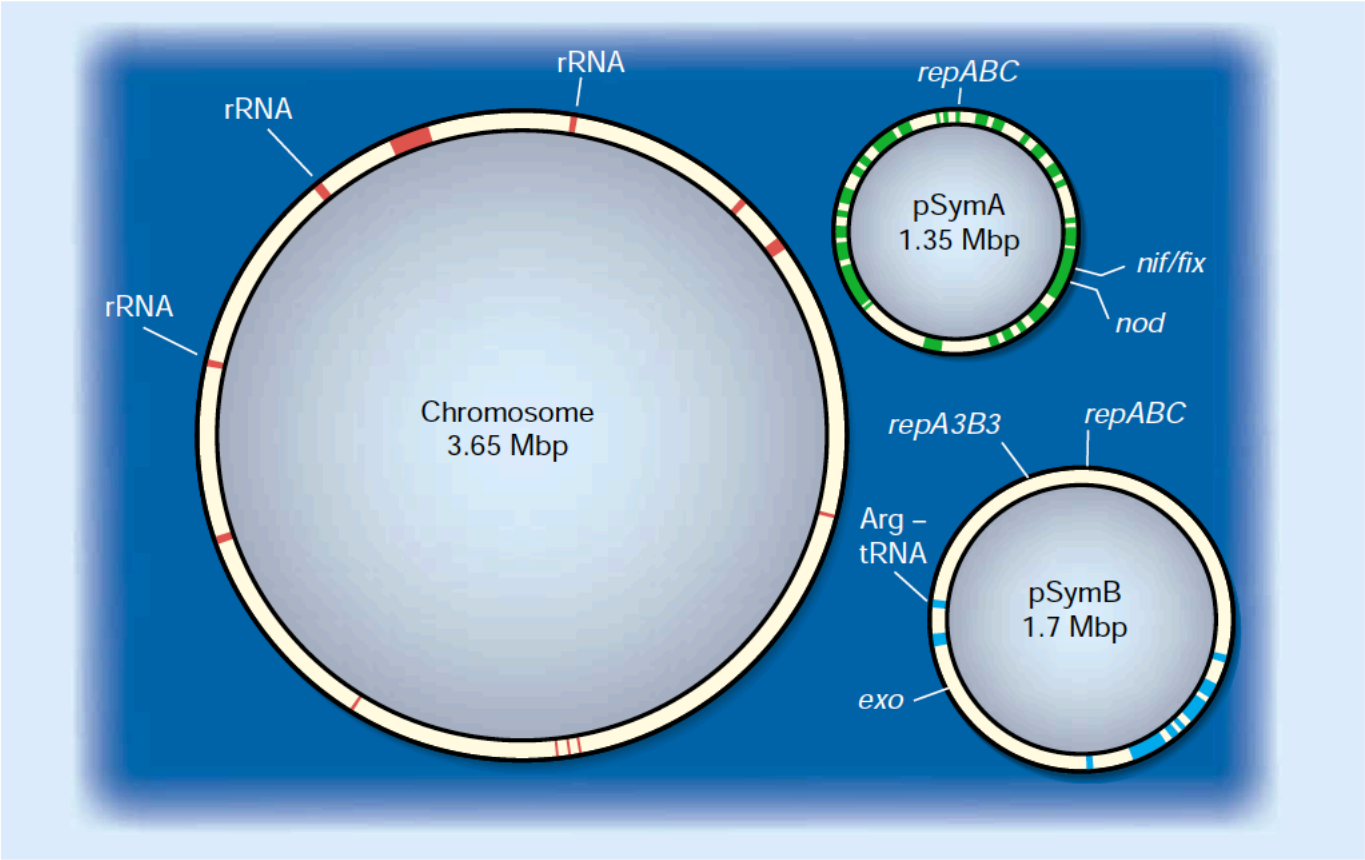


Figure 1.3. The three components of the *Sinorhizobium meliloti* genome: a chromosome and two secondary replicons. Red, green and blue regions have a guanosine and cytidine (G+C) content of less than 60% (averaged over 10-kilobase windows). The positions of some genes are shown, including those needed for the synthesis of ribosomal RNA (rRNA) and for plasmid replication (*rep* genes), as well as the gene encoding the essential transfer RNA (Arg-tRNA) that recognizes the nucleotide triplet CCG. Also shown are the gene regions required for the bacterium to form nodules on the roots of legumes (*nod* genes), for the formation of external polysaccharides (*exo* genes), and for symbiotic nitrogen fixation (*nif/fix* genes).

This suggests that the core chromosome is sufficient for growth in soil and that the accessory plasmids are necessary for survival in more specialized environments such as the rhizosphere or during intracellular interactions (diCenzo *et al.*, 2014).

S. meliloti has emerged as a model organism for the study of bacterial carbon metabolism (Geddes & Oresnik, 2014; Jacob *et al.*, 2008; Mauchline *et al.*, 2006). This is due in part to the organism's ability to utilize a vast number of carbon sources for growth, as well as the distinct metabolic pathways it employs that distinguish it from other established paradigms such as *Escherichia coli* or *Bacillus subtilis*. Recently, constraint based metabolic modeling has become a tool for evaluation of the metabolic characteristics of an organism (Feist *et al.*, 2009; Oberhardt *et al.*, 2009). These models have been used to predict the metabolic behavior of microorganisms under different environmental conditions, as well as simulate the effect of gene deletions and overexpression (Contador *et al.*, 2015; Razmilic *et al.*, 2018). Several models that vary in scope have been developed for *S. meliloti*, including modeling of SNF (Zhao *et al.*, 2012), whole metabolism (diCenzo *et al.*, 2016), and core metabolic network (diCenzo *et al.*, 2018). These models are invaluable resources for making *in silico* observations, genetic manipulations, predictions, as well as for generation of hypotheses.

1.2.3 Medicag*o*

Approximately 90% of the species with the plant family Leguminosae (Fabaceae) can fix atmospheric N into NH_3 through a symbiotic interaction with rhizobia (Liu *et al.*, 2018). Within this family is the genus *Medicago,* which contains approximately 87 species of flowering plants, the most well known being *Medicago sativa* (alfalfa) (Steele *et al.*, 2010). While much investigation of the *Rhizobium*-legume symbiosis has been conducted with alfalfa, *M. truncatula*

(barrel medic) has emerged as one of the main model legumes, along with *Lotus japonicus*, used for research purposes (Barker *et al.*, 1990; Cook, 1999; Handberg & Stougaard, 1992). The importance of having two prominent model legumes is revealed through consideration of the type of nodules formed on each plant during symbiosis. *M. truncatula* forms indeterminate nodules, which are elongated in shape due to their persistent apical meristem. In contrast *L. japonicus* forms determinate nodules, which lack a persistent meristem and are spherical in shape (Fig. 1.4) (Ferguson *et al.*, 2010). *M. truncatula* has several attributes that facilitated its emergence as a model organism, including a sequenced genome (Young *et al.*, 2011), which is diploid (2n=16), small in size (~375 Mbp), and simple in comparison to other legumes. Additionally, it has a short life cycle, is easily transformed with *Agrobacterium* and has high levels of genetic diversity and synteny with other, more complex, crop and forage legumes (Burks *et al.*, 2018; Cañas & Beltrán, 2018).

The host plant exerts a great deal of control over the symbiotic process. For instance, legumes can regulate the number and size of nodules formed on their roots based on their environmental conditions via the autoregulation of nodulation (AON) signalling pathway. For rhizobia, there is little detriment to existing within the nodule, but for the plant nodulation represents a great expenditure of resources and may not be beneficial in every situation (Fig. 1.5). While the AON pathway is not fully understood, key players include peptides MtCLE12 and MtCLE13 (Mortier *et al.*, 2010), receptor MtSUNN (Schnabel *et al.*, 2005), and root factors MtTML1 and MtTML2 (Gautrat *et al.*, 2019), which all act to repress rhizobium-induced nodulation. Legumes can also reduce nodulation in response to N availability in the soil or through the plant hormone ethylene using pathways that are independent of AON pathway (Ferguson *et al.*, 2019).

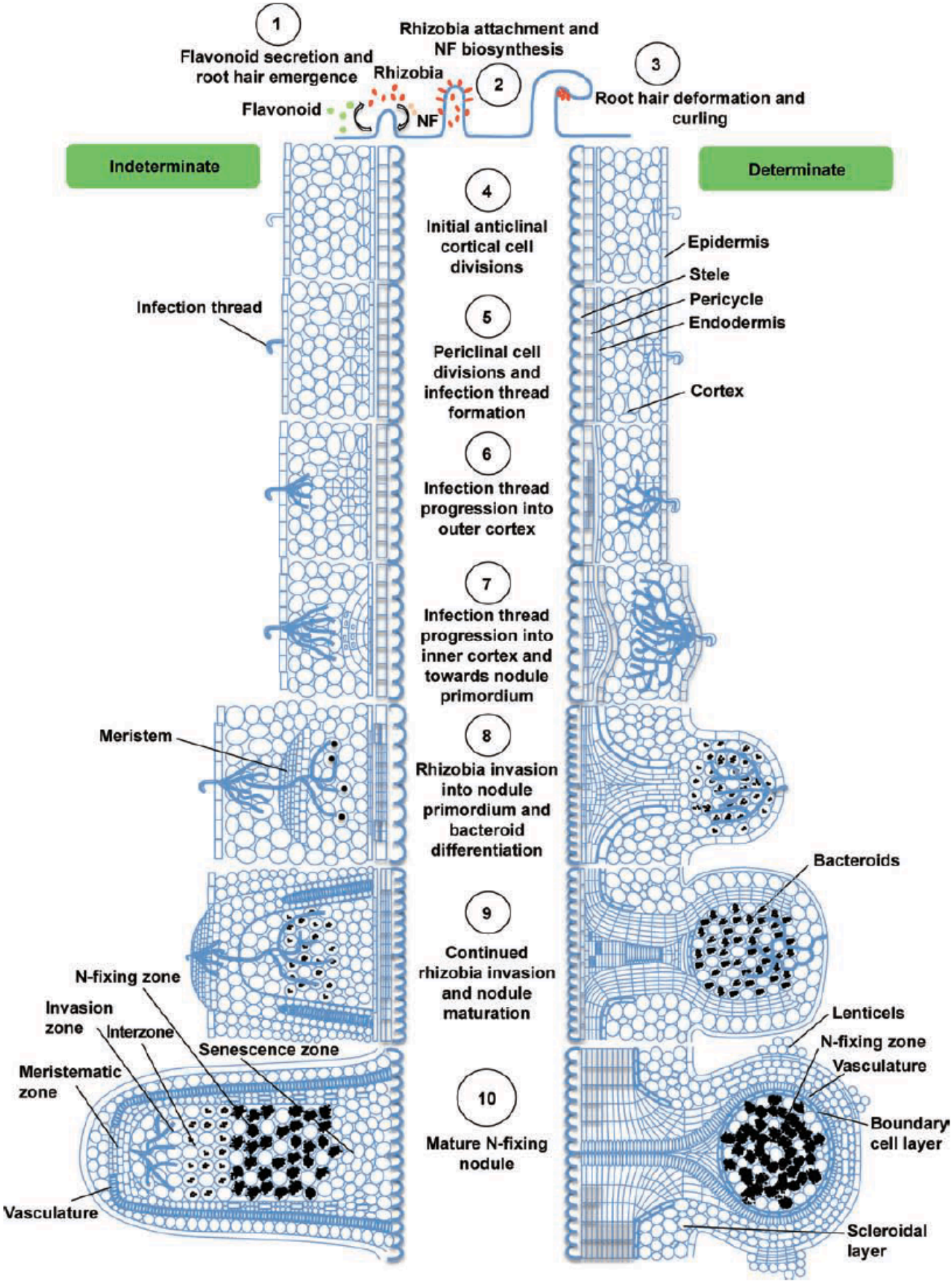

1
Flavonoid secretion and root hair emergence
Rhizobia attachment and NF biosynthesis
Rhizobia
2
3
Root hair deformation and curling
Flavonoid
NF
Indeterminate
Determinate
4
Initial anticlinal cortical cell divisions
Epidermis
Stele
Pericycle
Endodermis
Cortex
Infection thread
5
Periclinal cell divisions and infection thread formation
6
Infection thread progression into outer cortex
7
Infection thread progression into inner cortex and towards nodule primordium
Meristem
8
Rhizobia invasion into nodule primordium and bacteroid differentiation
Bacteroids
9
Continued rhizobia invasion and nodule maturation
N-fixing zone
Invasion zone
Interzone
Senescence zone
Meristematic zone
Lenticels
N-fixing zone
Vasculature
Boundary cell layer
10
Mature N-fixing nodule
Vasculature
Scleroidal layer

Figure 1.4. Developmental stages of indeterminate and determinate legume nodules. Illustrated are the developmental stages of pea (indeterminate; left) and soybean (determinate; right) nodules. Emerging root hairs exude flavonoid compounds, which attract compatible rhizobia and stimulate them to produce nod factors (NF). The root hair deforms and forms a pocket, in which the rhizobia become entrapped. Infection thread structures initiate in the pocket enabling the rhizobia to enter the plant. Cell divisions are first observed in the inner cortex for indeterminate nodules or the sub-epidermal cell layer for determinate nodules. Additional cell layers later divide leading to the formation of the nodule primordium. The infection threads progress towards this primordium and release the rhizobia into infection droplets, in which they differentiate into nitrogen-fixing bacteroids. At the top of the primordium of indeterminate nodules, a meristem develops that continually gives rise to new cells. As these new cells mature, many subsequently become infected, leading to successive zones of rhizobia invasion and differentiation within the nodule. In contrast, determinate nodules do not develop a persistent meristem and hence their invaded cells are all at a similar developmental phase. The various developmental stages, tissue types and nodulation zones are labeled.

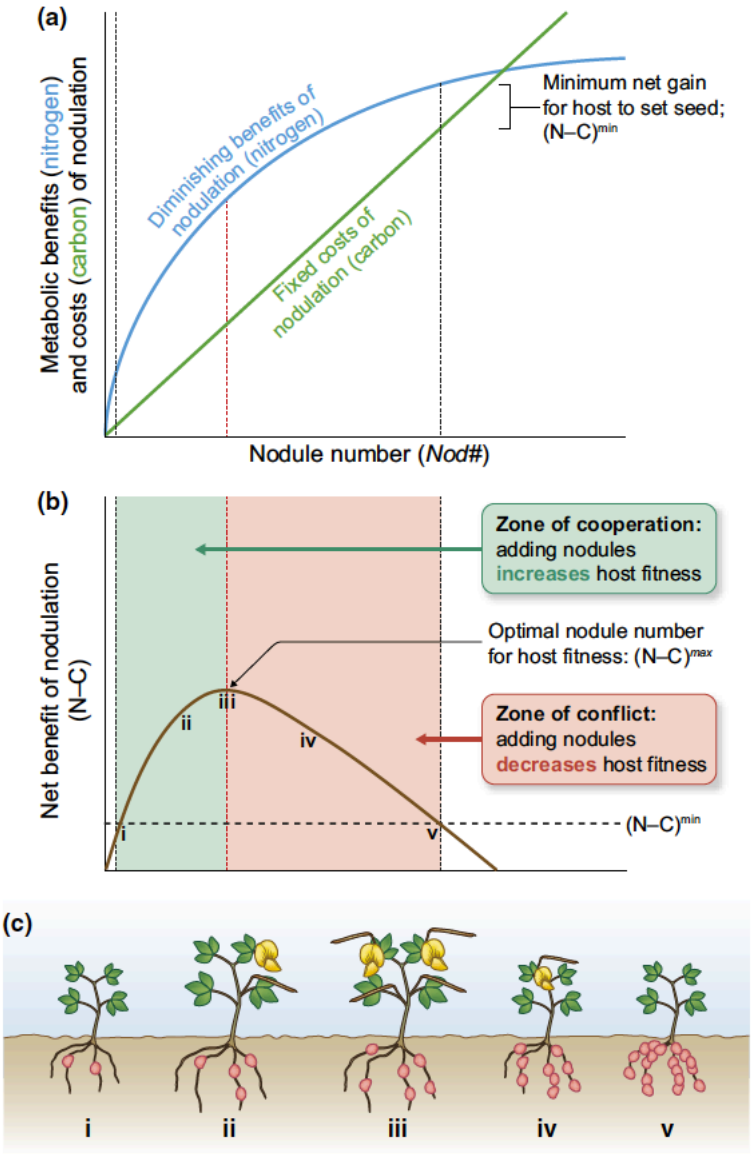

(a)
Metabolic benefits (nitrogen) and costs (carbon) of nodulation
Diminishing benefits of nodulation (nitrogen)
Fixed costs of nodulation (carbon)
Minimum net gain for host to set seed; (N–C)min
Nodule number (Nod#)
(b)
Net benefit of nodulation (N–C)
Zone of cooperation: adding nodules increases host fitness
Optimal nodule number for host fitness: (N–C)max
Zone of conflict: adding nodules decreases host fitness
(N–C)min
i
ii
iii
iv
v
(c)
i
ii
iii
iv
v

Figure 1.5. Legume and rhizobia conflict over nodule number. (a) Costs vs benefits of nodulation are modeled. Plant costs to nodulation (carbon, C) are predicted to be a linear function of the number of nodules formed (Nod#) with a slope of *m* (cost per nodule): f(C) = *m* x Nod#. Plant benefits from nodulation (nitrogen, N) are predicted to be a negative exponential function, $f(N) = \alpha(1-e^{-B \times Nod\#})$, with diminishing returns that reach an asymptote at α and diminish at a rate corresponding to *B*. (b) Net benefits of nodulation can be calculated by subtracting the cost from the benefit functions. The net benefit function for nodulation is unimodal, increasing with the formation of nodules (zone of cooperation) until the optimal number of nodules is reached ($N–C^{max}$), and above which additional nodules reduce the host benefit (zone of conflict). If too few or too many nodules are formed, the host does not acquire the net minimal benefit to set seed (i.e.< $N–C^{min}$). (c) Host fitness (i.e. growth, seed set) varies with the number of nodules formed. *Lotus japonicus* mutants have been generated that form too many nodules compared with wild-type and thus experience reduced fitness (Nishimura *et al.*, 2002).

The plant can also exert control over the post-invasion stages of symbiosis by influencing the differentiation fate of bacteroids, which are the nitrogen fixing form of the bacteria. Bacteroids exist in two forms; terminally differentiated and reversibly differentiated (Oono & Denison, 2010). Terminally differentiated bacteroids are observed in indeterminate nodules such as *M. truncatula*. They are morphologically distinct from free-living cells in that they appear swollen and branched, they exhibit an increased DNA content, and they cannot revert into a free-living state (Mergaert *et al.*, 2006). Reversibly differentiated bacteroids can be found within determinate nodules on *L. japonicus* or *Glycine max*; they are modestly altered but still resemble free-living rhizobia and are able to resume saprophytic growth upon isolation from the nodule. The differentiation state of the bacteroid is under host control (Mergaert *et al.*, 2006) and is influenced by nodule-specific cysteine rich (NCR) plant peptides in *M. truncatula* (Fig. 1.6C) (Van de Velde *et al.*, 2010).

1.2.4 Symbiosis

The *Rhizobium*-legume symbiosis is a host-microbe interaction in which the rhizobia are internalized within host cells, where they fix atmospheric nitrogen into ammonia. The plant uses this reactive nitrogen for growth while providing carbon to the bacteria. Several plant and bacterial genes must be systematically expressed for a functional symbiotic relationship to be established.

Symbiosis begins with an exchange of signal molecules between the bacterium and the host (Fig. 1.6A). The plant secretes flavonoid compounds that are perceived by rhizobia through binding to LysR type transcriptional regulator NodD, which induces the expression of bacterial *nod* genes.

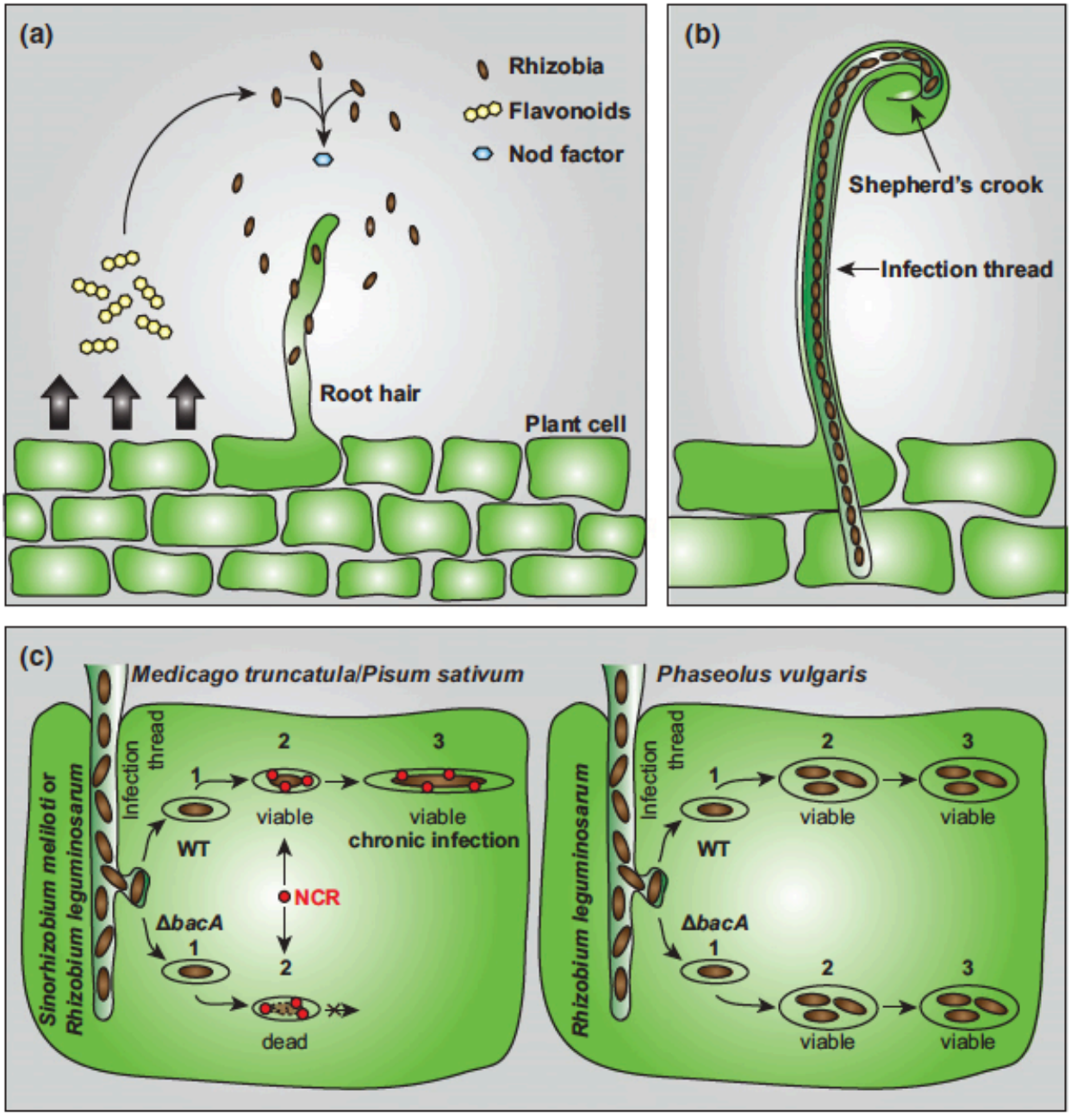

(a)
Rhizobia
Flavonoids
Nod factor
Root hair
Plant cell
(b)
Shepherd's crook
Infection thread
(c)
Medicago truncatula/Pisum sativum
Sinorhizobium meliloti or
Rhizobium leguminosarum
Infection thread
1
2
3
viable
viable
chronic infection
WT
NCR
ΔbacA
1
2
dead
Phaseolus vulgaris
Rhizobium leguminosarum
Infection thread
1
2
3
viable
viable
WT
ΔbacA
1
2
3
viable
viable

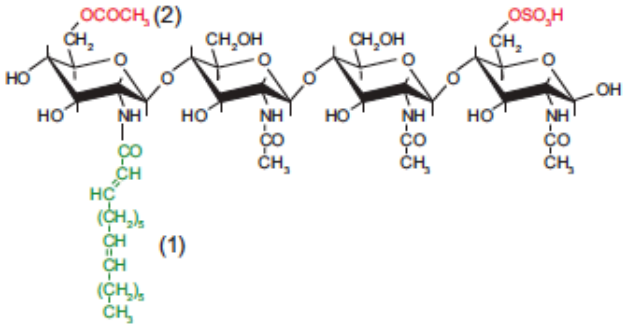
Figure 1.6. Rhizobia interacting with legumes. (a) The legume secretes flavonoids, which induce the rhizobia to produce Nod factors and attract them to the plant root hair cells. (b) Nod-factor signaling triggers a number of developmental changes, including root hair curling, which traps the rhizobia in Shepherd's crooks. Inward growth of the root hair tip results in tubular structures called infection threads, which allow the rhizobia to enter the cortical cell layers of the plant root. (c) The rhizobia escape the infection thread and are taken into the host cell via an endocytosis-like process (1), which encompasses them in a host-derived membrane. These intracellular compartments are known as symbiosomes. In legumes of the IRLC clade such as *Medicago truncatula* and *Pisum sativum*, the rhizobia are challenged with NCR peptides (2) and differentiate into elongated bacteroids (3). The bacterial BacA protein is essential for protecting the rhizobia against the antimicrobial activity of NCR peptides (2). In contrast, BacA is dispensable for rhizobia infecting legumes of the phaseoloid clade that do not produce NCR peptides. In these host plants, rhizobia do not differentiate terminally and often multiple bacteroids can be found inside a single symbiosome membrane. WT, wildtype.

Luteolin was determined to be the major flavonoid produced by *M. sativa* and *nodD* is required for induction of the *nod* genes (Mulligan & Long, 1985; Peters *et al.*, 1986). The number of *nodD* genes varies between bacterial species; *S. meliloti* has three *nodD* alleles, *nodD1*, *nodD2*, and *nodD3*, which are all capable of inducing nodulation (Honma & Ausubel, 1987; Honma *et al.*, 1990). NodD binds 47 bp DNA motifs called *nod* boxes that are found within *nod* promoter sequences (Fisher *et al.*, 1988; Rostas *et al.*, 1986).

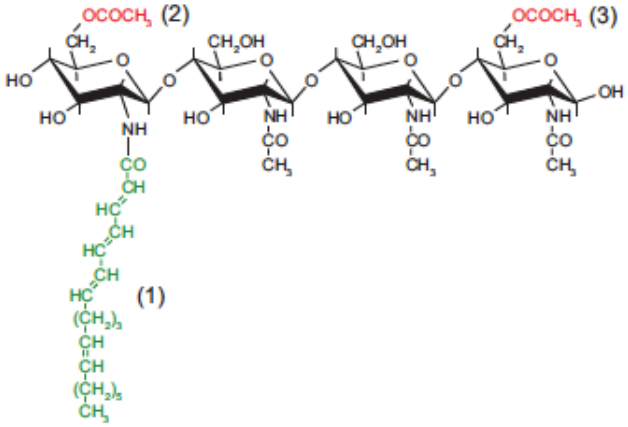
Induction of the *nod* genes leads to expression of proteins capable of Nod factor (NF) synthesis. NFs are lipochitooligosaccharide (LCO) molecules of bacterial origin that are generated in response to plant derived flavonoids. They consist of β-1,4-N-acetyl-D-glucosamine residues in which the acyl moiety is bound to the nonreducing terminal sugar, the number of residues as well as the length of the acyl chains is variable between and within species (Fig. 1.7). The core LCO backbone is generated from the *nodABC* operon; *nodIJ* encode LCO exporters, with species-specific additions such as fucosyl, sulphuryl, acetyl, methyl, carbamoyl, and arabinosyl groups encoded by accessory *nod*, *noe*, or *nol* genes.

NF contributes to the host specificity of a symbiotic bacterium but it is not the sole determinant. For example, different species of bacteria with distinct host ranges can produce identical NF (Cárdenas *et al.*, 1995). The host range of symbiotic bacteria is quite variable. *R. leguminosarum* bv. *trifoli* only nodulates clover (Terpolilli *et al.*, 2014), while *S. fredii* NGR234 is promiscuous and can establish symbiosis with legumes in over one hundred genera by producing a range of NFs (Relic *et al.*, 1993). There are even examples of *Bradyrhizobium* strains establishing symbioses with *Aeschynomene* species via NF independent processes (Okazaki *et al.*, 2016).

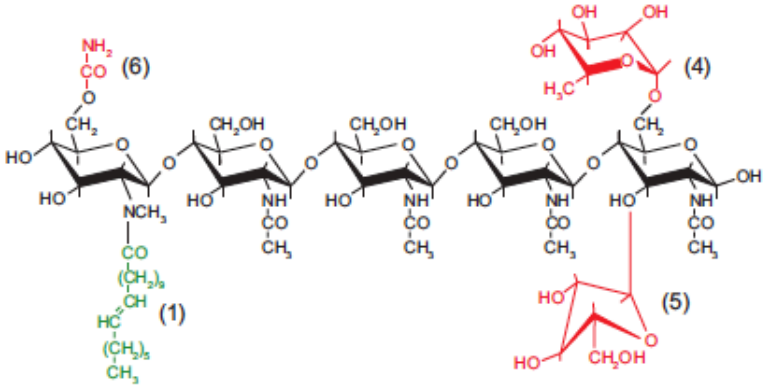
S. meliloti



R. leguminosarum bv. *viciae*



A. caulinodans



M. loti

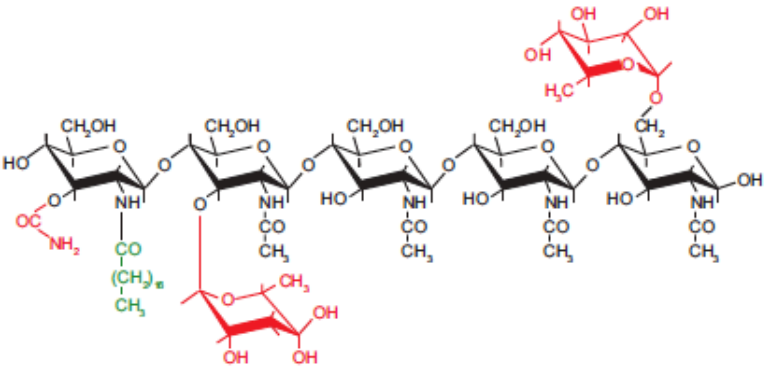


Figure 1.7. Representative nodulation (Nod) factors produced by *Sinorhizobium meliloti*, *Rhizobium leguminosarum* bv. *viciae*, *Azorhizobium caulinodans* and *Mesorhizobium loti* are shown with a backbone of β 1–4-linked N-acetylglucosamine residues (black) with N-linked acyl groups (green) and other host-specific decorations (red). Each species produces multiple Nod factors; e.g., the number of glucosamine residues can be four or five, the acyl chains (1) of the Nod factors from *S. meliloti* and *R. leguminosarum* bv. *viciae* can be C18:1 instead of C16:2 and C18:4 as shown, and not all Nod factors necessarily carry all the host-specific decorations. Mutants of *S. meliloti* and *R. leguminosarum* bv. *viciae* producing Nod factors lacking the acetyl group (2) on the acylated glucosamine and carrying a C18:1 acyl group (owing to mutations affecting *nodL* and *nodFE*, respectively) are defective for infection but can induce signaling responses in legumes. The acetyl group (3) on the *R. leguminosarum* bv. *viciae* Nod factor is attached by NodX and is required for infection of peas carrying the *SYMBIOSIS2* (*SYM2*) allele from the cv. Afghanistan pea. Mutants of *A. caulinodans* lacking the fucosyl (4), arabinosyl (5), and/or carbamoyl groups are defective for root hair infection but not bacterial entry via cracks in the epidermis.

Detection of NFs by legumes results in activation of the common symbiosis (SYM) signalling pathway, which can be trigged by rhizobia or arbuscular mycorrhizae. NFs bind LysM type receptors called Nod factor perception (NFP) receptors. Transgenic *M. truncatula* with mutations to *MtNFP* cannot respond to NF (Amor *et al.*, 2003). Similarly, disruption of two other LysM domain-containing receptor-like kinases (LYK) called *MtLYK3* and *MtLYK4* through RNAi mediated suppression or mutagenesis, reduced invasion by rhizobia (Limpens *et al.*, 2003; Smit *et al.*, 2007). The LysM motif found on these receptors is critical for NF recognition (Bensmihen *et al.*, 2011). MtNFP and MtLYK3 have been suggested to form multimeric complexes to induce plant responses and may be involved in the release of invading rhizobia into plant cells (Fig. 1.8A) (Fliegmann *et al.*, 2016; Moling *et al.*, 2014; Pietraszewska-Bogiel *et al.*, 2013). While the precise roles and interactions of these receptors remains unclear, it seems likely that *M. truncatula* can produce various NF receptors, including, but not limited to, NFP and LYK3, which can form heteromeric complexes and function during initial NF perception as well as at several points following NF recognition.

Rhizobia invade their hosts by penetration of plant tissues through a structure called an infection thread (IT) (Fig. 1.6B). *S. meliloti* mutants producing altered NFs are deficient in the ability to elicit IT formation, suggesting that the chitin backbone of NF may be sufficient for recognition but that entry into host cells requires a more stringent identification provided by the decorative functional groups such as acyl or sulphuryl moieties (Ardourel *et al.*, 1994; Roche *et al.*, 1991). In addition to NF, rhizobia must produce symbiotic exopolysaccharides (EPS) for initiation of and survival within IT. *S. meliloti* can produce EPS in two forms; succinoglycan (EPSI) and galactoglucan (EPSII) (Glazebrook & Walker, 1989; Pellock *et al.*, 2000).

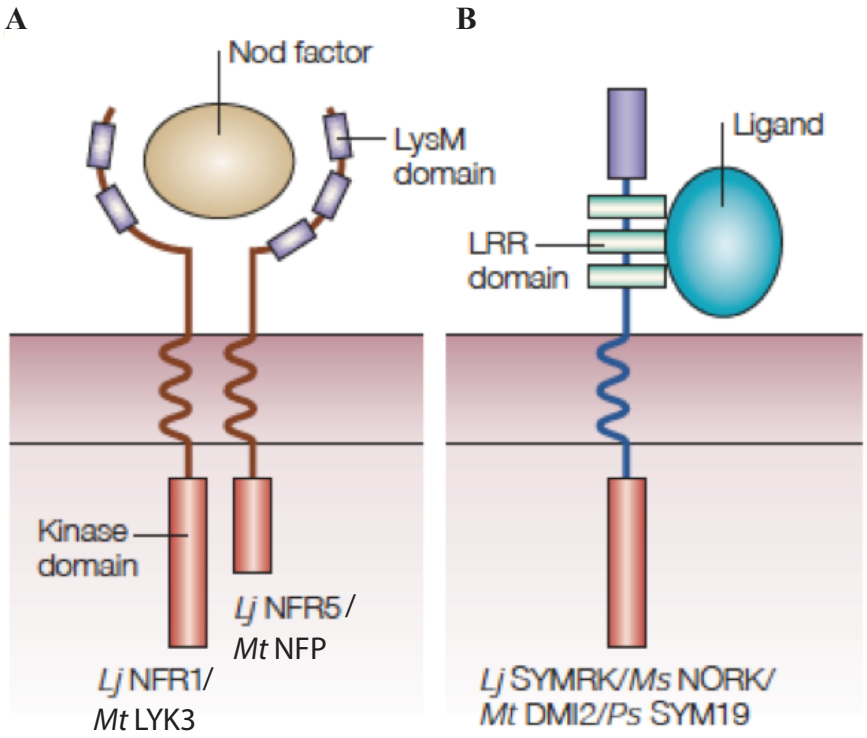


Figure 1.8. Predicted kinases that are required for Nod factor signaling. (A) *Lotus japonicus Lj NFR1* and *Lj NFR5*, which both encode extracellular LysM motifs, are thought to function in nodulation (Nod)-factor binding. The effects of mutations in these genes support a role in Nod-factor recognition. According to a simple model, the Nod-factor receptor is a heterodimer that consists of the two receptor-like kinases, *Lj* NFR1 and *Lj* NFR5. The kinase domains (red) might be involved in signal transduction; whereas *Lj* NFR1 is predicted to have an intact kinase domain, *Lj* NFR5 lacks a kinase-activation loop in the kinase domain. Several closely linked genes that are strongly related to *Lj* NFR1 have been identified in both *L. japonicus* and *Medicago truncatula*, and so it is possible that types of complex other than those shown here could occur. (B) Plants use leucine-rich-repeat (LRR) receptor-like kinases (LRR-RLKs) in various signal-transduction pathways. The LRR-RLKs are related to Toll receptors in *Drosophila melanogaster* and Toll-like receptors in animal cells. The LRR domain is often involved in protein–protein interactions and the kinase domain is involved in protein phosphorylation. The product of *M. truncatula Mt* DMI2 and its orthologues in *M. sativa* (*Ms NORK*), *L. japonicus* (*Lj SYMRK*) and *Pisum sativum* (*Ps SYM19*) belong to this class of proteins. It has been proposed that this protein might interact with an (unidentified) extracellular protein and mediates the phosphorylation of some component that has yet to be identified.

Mutants that cannot produce EPSI fail to initiate IT formation and yield nodules lacking bacteria (Leigh *et al.*, 1985).

Successful binding of appropriate NF-receptor pairs results in an increase in intracellular Ca^{2+} in root hairs (Ca^{2+} flux), followed by oscillation in cytosolic Ca^{2+} (Ca^{2+} spiking), leading to cortical cell division, alterations in the root hair cytoskeleton, and root hair curling. Invading rhizobia become trapped in the curled root hair and from here can begin to penetrate into the root hair cortex through the IT.

Mutations to plant genes *MtDMI1*, *MtDMI2*, or *MtDMI3* (does not make infections) are able to disrupt Ca^{2+} ion polarization patterns and prevent root hair curling and infection. These mutants are also defective for mycorrhizal symbiosis indicating that these genes participate in the establishment of both interactions (Catoira *et al.*, 2000). *DMI1* encodes a ligand-gated ion channel that is associated with the nuclear envelope (Ané *et al.*, 2004; Riely *et al.*, 2007), *DMI2* encodes a leucine-rich-repeat (LRR) receptor-like kinase that binds an as yet unidentified ligand (Fig. 1.8B) (Endre *et al.*, 2002; Esseling *et al.*, 2004), and *DMI3* encodes a calcium and calmodulin-dependent protein kinase (CCaMK) able to interpret complex calcium signatures such as sharp oscillations in Ca^{2+} concentrations and effect gene expression (Lévy *et al.*, 2004; Mitra *et al.*, 2004).

Additionally, several transcription factors have been proposed to function downstream of the *DMI* genes in the SYM pathway. Mutations to legume GRAS-type transcriptional regulators *NSP1* and *NSP2* (nodulation signalling pathway) reduce IT formation and cortical cell division but exhibit wildtype-like induction of the Ca^{2+} response (Kaló *et al.*, 2005; Oldroyd & Long, 2003; Smit *et al.*, 2005; Wais *et al.*, 2000). Similarly, two ERF-type transcription factors *ERN1* and *ERN2* (ERF required for nodulation) are induced by NF (Andriankaja *et al.*, 2007) and

mutations to *ern1* also block IT development (Middleton *et al.*, 2007). Finally, transcription factor *Mt*NIN (nodule inception) is necessary for nodule organogenesis (Marsh *et al.*, 2007).

NF perception triggers localized increases in cytokines, which are recognized by cytokinin receptors such as MtCRE1. Inactivation of MtCRE1 led to cytokinin insensitive roots, which were deficient in IT progression and cortical cell division (Gonzalez-Rizzo *et al.*, 2006). MtCRE1 links early events such as NF binding to invasion and nodulation by increasing expression of transcription factors *MtERN1*, *MtNSP*s, and *MtNIN* (Plet *et al.*, 2011). IT progression towards the plant occurs outside of root hair cells but within a tubule that possesses a plant cell wall (Brewin, 2004). The level of stress experienced by rhizobia within the IT is not well understood. However the IT has been suggested to be an acidic compartment (Geddes *et al.*, 2014). Additionally, rhizobia deficient in the production of catalases are compromised in their ability to invade the host plant, suggesting that reactive oxygen species (ROS) are present in the IT (Jamet *et al.*, 2003; Sigaud *et al.*, 1999).

When rhizobia progress deep enough into the inner plant cortex, they are released into cortical cells via endocytosis and acquire an additional membrane derived from the cortical cell, the entire structure including the new membrane plus an individual bacterium is termed a symbiosome (Brewin, 2004). Plant genes *MtDMI2*, *MtHAP2-1,* and *MtNIP* are required for release of bacteria from the IT. Suppression of these genes often results in failure to release bacteria and aberrant IT formation (Combier *et al.*, 2006; Limpens *et al.*, 2005; Veereshlingam *et al.*, 2004). Additionally, a *S. meliloti hemA* mutant remains trapped in the IT, likely due to the requirement for heme synthesis of several symbiotically important components (Dickstein *et al.*, 1991; Gilles-Gonzalez *et al.*, 1991; Kereszt *et al.*, 1995).

Upon release into host cells rhizobia differentiate into bacteroids, the nitrogen fixing form of the bacteria. Bacteroids found in *M. truncatula* are terminally differentiated, meaning they are elongated, have increased DNA content, and have lost the ability to divide (Mergaert *et al.*, 2006). *S. meliloti bacA* mutants, which produce a modified lipid A component of lipopolysaccharide (LPS), lyse upon endocytosis by cortical cells indicating a role for LPS in bacteroid survival (Fig. 1.6C) (Ferguson *et al.*, 2004; Glazebrook *et al.*, 1993). Consistent with these results, *lpsB* mutants producing an altered LPS core are also deficient for survival after exiting the IT (Campbell *et al.*, 2002). Mutants compromised in transport of ions such as Mn^{2+} and K^{+} form bacteroids that undergo premature senescence and do not fix nitrogen (Davies & Walker, 2007; Putnoky *et al.*, 1998). Additionally, alternative sigma factors encoded by *rpoH1* and *rpoH2* as well as transcription factor *ritA* have roles in symbiosome survivability (Bittner & Oke, 2006; Mitsui *et al.*, 2004; Wells & Long, 2002). The above genes encode proteins that act during the bacterial stress response; that mutants in these genes lead to unviable bacteroids suggests that the symbiosome is a harsh environment in which multiple stresses must be overcome.

After endocytosis and differentiation into bacteroids, N_2 fixation by the bacterial nitrogenase enzyme can begin. Nitrogenase is a two-component metalloenzyme that consists of the iron (Fe) protein, also called nitrogenase reductase, and the molybdenum-iron (MoFe) protein also called dinitrogenase. The Fe protein is a dimer composed of two γ subunits encoded by *nifH* (nitrogen fixation); it contains two MgATP binding sites as well as one metallocluster called the F cluster [Fe_4S_4]. The MoFe protein is a tetramer formed from two α and two β subunits encoded by *nifDK*. It contains two metalloclusters per αβ dimer, the P-cluster [Fe_8S_7], which sits at the interface of the α and β subunits, and the M-cluster or FeMo cofactor [$MoFe_7S_9C$-*R*-

homocitrate], which is the site of N_2 reduction (Fig. 1.9) (Einsle *et al.*, 2002; Hu & Ribbe, 2013; Kirn & Rees, 1992; Lancaster *et al.*, 2011; Spatzal *et al.*, 2011).

Generation of ammonia requires the interplay of two cycles, the Fe protein cycle and the MoFe protein cycle. Initially, the Fe protein receives one e^- from e^- carriers such as ferredoxin or flavodoxin, and binds two molecules of MgATP. Subsequently, binding of the Fe protein to the MoFe protein aligns the F-cluster in close proximity to the P-cluster facilitating e^- flow between the two enzyme components. One e^- is donated from the Fe protein to the MoFe protein at the expense of two MgATP molecules. The e^- is shuttled from the P-cluster to the M-cluster where it can partially reduce N_2. The Fe protein dissociates and the cycles repeat themselves, eight e^- must be transferred in this manner to reduce N_2 into $2NH_3$ (Fig. 1.10). A stoichiometric amount of H_2 gas is released per mole of N_2 reduced by dinitrogenase (Liang & Burris, 1988; Simpson & Burris, 1984).

Both of the nitrogenase components are irreversibly inactivated by oxygen and extreme care must be taken to mitigate the enzyme's oxygen sensitivity. In general, bacteria employ several strategies to limit exposure of nitrogenase to oxygen including anaerobic growth, spatio-temporal separation, consumption of excess oxygen via respiration, or oxygen diffusion barriers. The Rhizobium-legume symbiosis utilizes a strategy in which the nodule cortex acts as an oxygen diffusion barrier, plant derived leghemoglobin facilitates diffusion of oxygen, and a bacterial high affinity terminal oxidase, *cbb*$_3$, permits respiration within bacteroids (Dixon & Kahn, 2004; Pitcher & Watmough, 2004; Preisig *et al.*, 1996; Zufferey *et al.*, 1996).

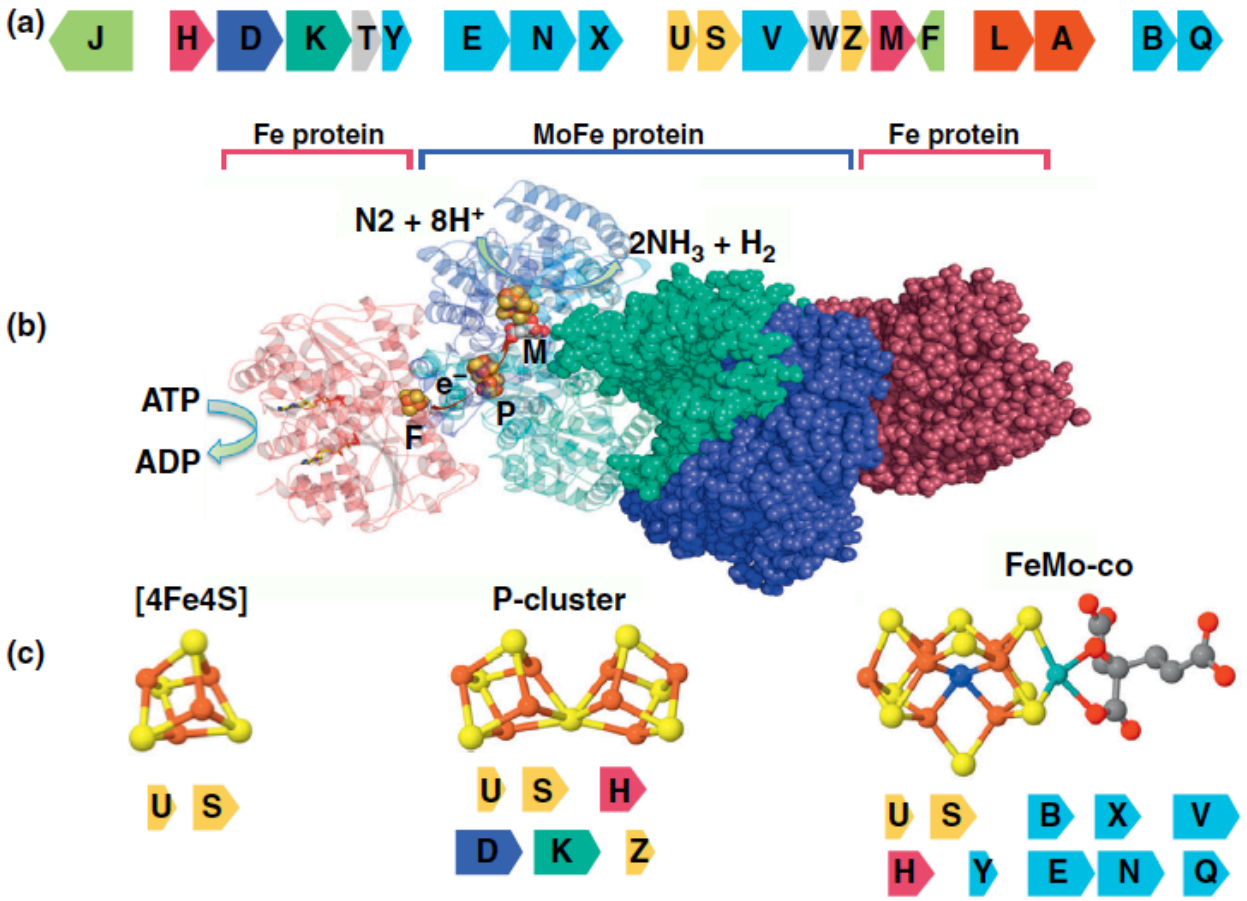


Figure 1.9. Nitrogenase structure and genes required for its biosynthesis. (A) The nitrogen fixation (*nif*) gene cluster from *Klebsiella oxytoca* (formerly *K. pneumoniae*). The nitrogenase structural genes (*nifH*, *nifD* and *nifK*) are coloured according to the crystal structure shown in (B). Remaining genes are colour-coded according to their functions: dark red, Fe protein maturation (*nifM*); light blue, FeMoco biosynthesis (*nifY*, *nifE*, *nifN*, *nifX*, *nifV*, *nifB*, *nifQ*); yellow, Fe–S cluster biosynthesis (*nifU*, *nifS*, *nifZ*); green, electron transport (*nifJ*, *nifF*); orange, transcriptional regulation (*nifL*, *nifA*); grey, unknown function (*nifT*, *nifW*). (B) Structure of the nitrogenase enzyme complex (PDB code 1n2c) showing the MoFe and Fe protein components, with the three metalloclusters revealed on the left-half of the complex (abbreviated as F, [4Fe4S]; P, P cluster and M, FeMo-co, respectively). ATP hydrolysis by the Fe protein, the route of electron transfer to the catalytic site and the enzyme reaction are also illustrated on the left half of the structure. (C) Structures of the three metalloclusters in nitrogenase. Genes required for the biosynthesis of each cluster are illustrated above.

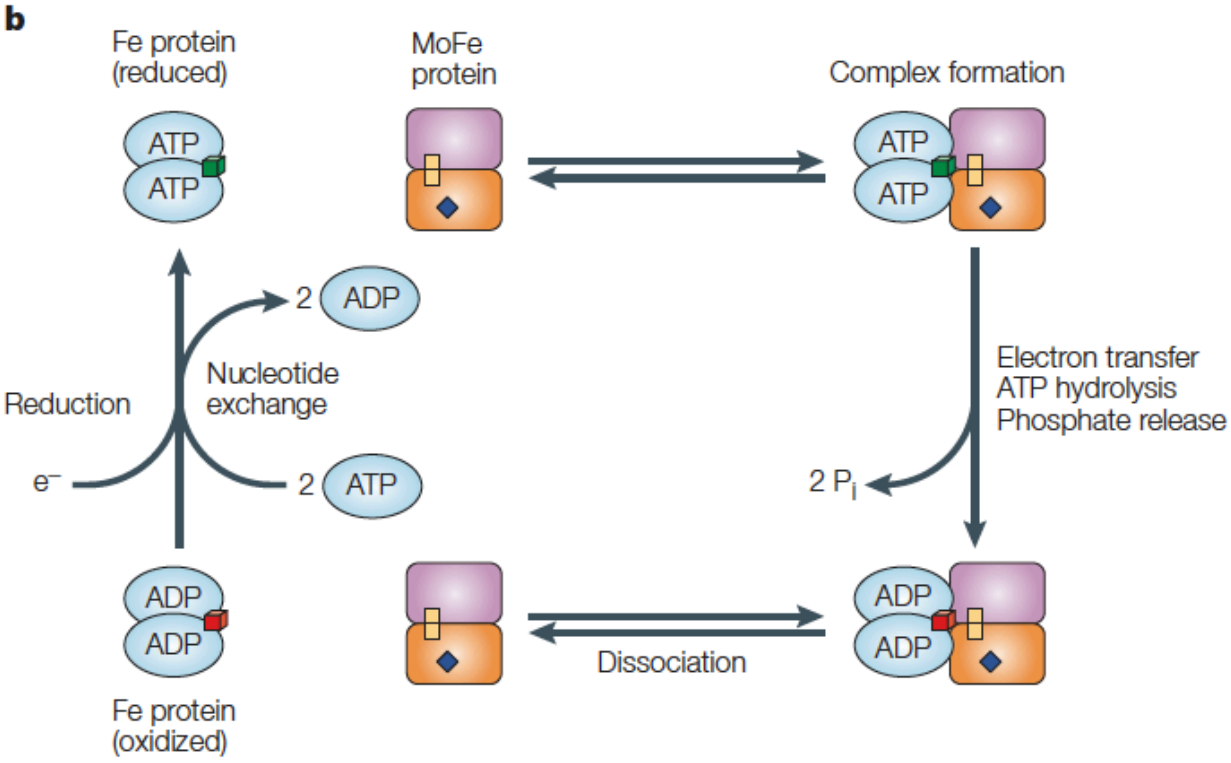


Figure 1.10. Schematic representation of the nitrogenase Fe protein cycle. The Fe protein dimer is shown in light blue with the cube representing the [4Fe–4S] cluster coloured green to indicate the reduced form and red to represent the oxidized form. The α and β subunits of the MoFe protein are depicted as orange and pink, respectively, the yellow squares represent the P cluster and the blue diamond represents the FeMo cofactor. Changes in the oxidation state of the MoFe protein are not shown.

All diazotrophs encode a MoFe nitrogenase, although some organisms are also capable of utilizing Mo-independent components. These alternative nitrogenases, vanadium and iron (VFe) or iron only (FeFe) nitrogenases, are encoded by *vnf* and *anf* genes respectively, and are most often expressed under Mo limiting conditions (Dixon & Kahn, 2004; Eady *et al.*, 1987; Joerger *et al.*, 1988; Shah & Brill, 1977).

1.2.5 Extension of N_2 fixation into nonlegumes

There are three schools of thought regarding the establishment of nonlegumes that fix N_2. These are: direct expression of bacterial nitrogenase in plants, engineering nonlegumes capable of symbiosis and nodulation, and generation of novel associations between N_2 fixing organisms and crop plants. While a core set of genes encoding the functional nitrogenase enzyme has been identified, their transfer to plants is complicated by the enzymes' complex biosynthetic process and sensitivity to oxygen (Rubio & Ludden, 2008). Both mitochondria and plastids are intracellular compartments with the potential to provide the ATP and e^- requirements of nitrogenase as well as a protective microoxic environment (Curatti & Rubio, 2014).

The root-nodulating symbiosis is limited to legumes, but association with arbuscular mycorrhizal fungi is prevalent amongst land plants. These associations utilize a signalling pathway with common components to the SYM pathway, which suggests that components from the AM pathway could be hijacked for use in an engineered symbiosis leading to nodulation and nitrogen fixation in nonlegumes.

Plants can also associate with nonsymbiotic nitrogen fixing bacteria. It has been proposed that these interactions could be tailored to develop a synthetic symbiosis in which crop plants could benefit from an improved association with traditionally nonsymbiotic bacteria. This

method involves manipulation of both participants to ensure proper communication as well as colonization and nitrogen fixation. Typically, the host is made to secrete a specific compound to which only desired microbes are receptive causing a species specific increase in competitiveness. This “biased rhizosphere” approach has been shown to affect microbial populations in the rhizosphere (Mondy *et al.*, 2014; Oger *et al.*, 1997; Savka & Farrand, 1997). Recently, transgenic barley plants have been engineered to synthesize and secrete chemical signals called rhizopines into the rhizosphere at levels that were detectable by a rhizopine biosensor strain. This proof of concept experiment demonstrates that rhizopines could be used to place the symbiotic properties of diazotrophic bacteria under plant control (Geddes *et al.*, 2019).

1.3 Carbon metabolism

1.3.1 Free-living metabolism

Rhizobia exhibit saprophytic growth when not participating in a symbiotic relationship. Utilization of a variety of carbon sources is essential for survival in nutrient limited environments such as bulk soil. The rhizosphere, which is the soil area surrounding the roots, is a distinct environment from bulk soil that is enriched for nutrients by plant root exudate. It has been proposed that acquisition of accessory plasmids facilitates adaptation to distinct environmental niches and that the transition from free living to symbiotic lifestyles is a feature that coincides with large bacterial genomes.

The *S. meliloti* pSymB replicon has a large proportion of genes dedicated to carbon metabolism and transport and has been suggested to have an important role during survival in the soil environment (Finan *et al.*, 2001). Consistent with this idea, metabolic modeling has suggested that pSymB contributes to the fitness of free living cells during growth in bulk soil or the rhizosphere (diCenzo *et al.*, 2016). Additionally, transcriptomic analysis of *R. leguminosarum* across different environmental conditions suggests that genes encoded from plasmid pRL8 are preferentially expressed in the pea rhizosphere (Ramachandran *et al.*, 2011). Plasmid curing experiments also support the concept that accessory plasmids contribute to niche specific growth (Baldani & Weaver, 1992; Baldani *et al.*, 1992; Brom *et al.*, 1992; diCenzo *et al.*, 2014; Hynes & O'Connell, 1990; Hynes & McGregor, 1990; Moënne-Loccoz & Weaver, 1995a; Moënne-Loccoz & Weaver, 1995b; Moënne-Loccoz *et al.*, 1994; Oresnik *et al.*, 2000).

Survival in the soil environment is enhanced by the ability to locate nutrient rich niches such as the rhizosphere. Through a process called chemotaxis, motile bacteria can sense their environment and move in a beneficial direction. Chemotaxis has been best described in *E. coli*,

which has led to identification of similar systems in other organisms. Rhizobia have been shown to migrate towards crude root exudate. While many of the specific components within these exudates have been difficult to determine, some components have been identified. It is generally thought that root exudate contains sugars, organic acids, and amino acids.

S. meliloti has nine chemoreceptors, called methyl accepting chemotaxis proteins (Mcp) that are membrane bound and sense external environmental queues. McpU is a proline and general amino acid receptor (Webb *et al.*, 2014); McpX binds quaternary ammonium compounds (QACs) such as choline, glycine betaine, trigonelline (Webb *et al.*, 2017); and McpV binds short-chain carboxylates such as acetate, propionate, pyruvate, and glycolate (Compton *et al.*, 2018), which all mediate movement of *S. meliloti* towards *M. sativa*. These receptors transmit signals to the core chemotaxis (*che*) genes, of which the majority are chromosomally encoded (Sourjik *et al.*, 1998). *S. meliloti* has two *che* systems, *che1* on the chromosome and *che2* on pSymA (Meier *et al.*, 2007), mutations to *che1* reduce or abolish chemotaxis while *che2* mutants seem unaffected (Meier & Scharf, 2009; Meier *et al.*, 2007). In *R. leguminosarum*, mutations to *che1* as well as *mcpB* or *mcpC* are less able to compete for nodule occupancy (Miller *et al.*, 2007; Yost *et al.*, 1998).

A number of specific carbon sources have been identified which have a presumed importance for survival in the rhizosphere. *S. meliloti* and *R. leguminosarum* mutants unable to catabolize the methylpentose rhamnose exhibit a competitive disadvantage for nodule occupancy (Oresnik *et al.*, 1998). Similarly, *R. leguminosarum* mutants unable to catabolize glycerol, erythritol, or homoserine also show reduced competition for nodule occupancy (Ding *et al.*, 2012; Vanderlinde *et al.*, 2014; Yost *et al.*, 2006). *S. meliloti* proline catabolic mutants nodulate more slowly and are less competitive than wildtype (Jiménez-Zurdo *et al.*, 1995).

Overexpression of proline catabolic genes was able to improve competitiveness in the parental strain (Jiménez-Zurdo *et al.*, 1995; van Dillewijn *et al.*, 2001). Mutations to the pyrroloquiniline quinone (PQQ) linked glucose dehydrogenase (Gcd) were also delayed in nodule emergence and less competitive for nodule occupancy (Bernardelli *et al.*, 2008). However, since this gene is not essential for glucose catabolism in *S. meliloti*, the competition phenotype is likely unrelated to glucose utilization. Possibilities include altered production of EPS or reduced solubilisation of inorganic phosphates by the *gcd* mutant (Bernardelli *et al.*, 2008; de Werra *et al.*, 2009). The inability to catabolize inositol has been associated with reduced competitiveness in *S. meliloti*, *R. leguminosarum*, and *S. fredii* (Fry *et al.*, 2001; Jiang *et al.*, 2001; Kohler *et al.*, 2010). Inositol metabolism is of particular interest due to the rhizopine concept, in which inositol derivatives called rhizopines are synthesized in bacteroids and subsequently catabolized by free-living rhizobia to provide a competitive advantage (Murphy *et al.*, 1995). Despite all that is known about competition in the rhizosphere, there remains a significant knowledge gap regarding how competition for nodulation manifests in the environment. Whether carbon catabolism is most relevant for survival in the rhizosphere or at some point during host invasion remains to be seen.

1.3.2 Metabolism in the infection thread

Continuous generation of NF and EPS is a requirement for symbiosis. Both compounds are complex carbohydrate structures, which would be energetically intensive to synthesize, suggesting that rhizobia are metabolically active during progression through the IT. Poly-β-hydroxy-butarate (PHB) and glycogen are the major polymers used to store carbon in *S. meliloti* (Wang *et al.*, 2007). PHB has been suggested to serve as a carbon source that fuels invasion (Charles *et al.*, 1997). Consistent with this assertion, mutations to PHB synthesis (*phbC*) and

degradation (*bdhA*) genes generated strains that were less competitive for nodule occupancy (Aneja *et al.*, 2005). *S. meliloti* possesses two glycogen synthase genes *glgA1* and *glgA2*. GlgA1 is the functional enzyme while mutations in *glgA2* contained wildtype levels of glycogen (Wang *et al.*, 2007). However, strains with mutations in *glgA2* were reduced in nodulation and nitrogen fixation, as were *phbC* and *glgA1*, suggesting that all three of these genes are involved in symbiosis (Wang *et al.*, 2007). Additionally, *phbC* mutants produce less EPSI, which is required for IT initiation, and could contribute to the competitive disadvantage observed in these mutants (Aneja *et al.*, 2004).

An interesting phenotype relating to EPSI production arose from a mutation to galactose catabolic genes. Strains unable to utilize galactose were shown to be more competitive for nodule occupancy than wildtype (Geddes & Oresnik, 2012a). Subsequently, it was observed that these strains acidified their growth medium and exhibited increased expression of EPSI synthesis genes. It was hypothesized that the increased competitiveness may be the result of enhanced acidification of the environment in response to galactose leading to more EPSI production during invasion (Geddes *et al.*, 2014).

An increase in competition for nodule occupancy of alfalfa was also observed in trehalose catabolic mutants (Jensen *et al.*, 2005). Trehalose is a disaccharide consisting of two α(1,1) linked glucose monomers and can serve as a carbon source as well as an osmotic and stress protectant in many organisms. It has been proposed to be available within IT's due to the induction of catabolic genes during invasion (Jensen *et al.*, 2005). Trehalose transport mutants did not exhibit an increase in competitiveness suggesting that accumulation of trehalose within cells may be the critical factor affecting competition (Jensen *et al.*, 2005). Consistent with this

observation, mutants deficient in the biosynthesis of trehalose were less competitive for nodule occupancy (Domínguez-Ferreras *et al.*, 2009).

1.3.3 Bacteroid metabolism

Upon internalization within host cells, rhizobia are provided carbon by the plant to fuel cellular functions including the nitrogenase reaction. Sucrose derived from photosynthesis is transported via plant tissue called phloem to nodule cells for nourishment of bacteroids. Two enzymes, sucrose synthase and alkaline invertase, cleave sucrose into component hexoses (Fig. 1.11). Antisense repression of sucrose synthase MtSuc1, which converts sucrose to UDP-glucose and fructose, resulted in impaired plant growth and nodulation only under SNF dependent conditions (Baier *et al.*, 2007). *MtInv* was shown to exhibit increased transcription in developing root nodules (Tesfaye *et al.*, 2006).

The hexoses are metabolized via glycolytic and TCA cycle enzymes into dicarboxylic acids, primarily malate, which are the direct carbon source provided to bacteroids (Fig. 1.11) (Miller *et al.*, 1988; Mitsch *et al.*, 2018). Dicarboxylates must cross the SM to be metabolized by bacteroids, while transport of succinate and malate across this membrane has been measured from SM isolated from soybean nodules (Udvardi *et al.*, 1988), a gene encoding this transporter in legumes has yet to be identified. Transport of dicarboxylates across the bacteroid membrane occurs via bacterial dicarboxylate transporters (Dct). Mutations to *S. meliloti dct* genes result in the inability to transport succinate and nodules that are unable to fix nitrogen (Bolton *et al.*, 1986; Finan *et al.*, 1988; Jiang *et al.*, 1989; Watson *et al.*, 1988; Yarosh *et al.*, 1989).

Within bacteroids, malate is converted into acetyl coenzyme A (acetyl-CoA) before being fed into the TCA cycle, facilitated by NAD^+-dependent malic enzyme Dme and pyruvate dehydrogenase PDH.

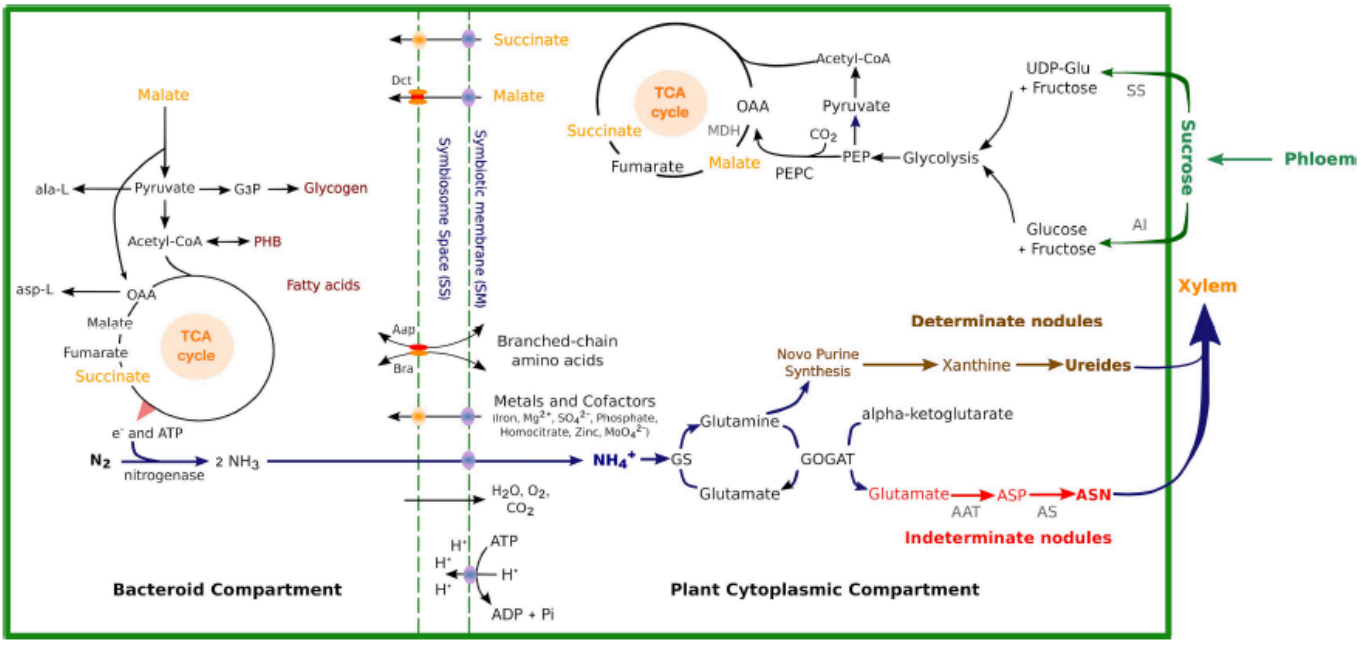


Figure 1.11. Schematics of carbon and nitrogen metabolic pathways with key enzymes, metabolites, and transporters in determinate nodules and indeterminate nodules. Sucrose in the plant cytosol is split into glucose and fructose by AI or UDP-Glc and fructose via SS, which is then catabolized via glycolysis to PEP. Carbon from PEP and carbonic acid is diverted to OAA and then malate by PEPC and the neMDH, respectively. OAA may be further converted to succinate or fumarate. Carbon sources are transported across the peribacteroid and bacteroid membranes and enter the TCA cycle in the bacteroid to be metabolized. Transport of inorganic ions and cofactors required for SNF across the SM is indicated. The ammonia produced by the SNF is transported back to the plant and assimilated by GS and GOGAT into Gln and Glu (blue arrows). In indeterminate nodules, Glu and Gln are further converted to Asp and Asn by AAT and AS, respectively (red arrows). In determinate nodules, Gln enters purine synthesis pathway and is converted to ureides (brown arrows). AI, alkaline invertase; UDP-Glc, UDP-glucose; SS, sucrose synthase; PEP, phosphoenolpyruvate; OAA, oxaloacetate; PEPC, PEP-carboxylase; MDH, malate dehydrogenase; PHB, polyhydroxybutyrate; AAT, aspartate aminotransferase; AS, asparagine synthetase; ASP, aspartate; ASN, asparagine.

Mutations to these genes yield strains that cannot fix nitrogen (Cabanes *et al.*, 2000; Driscoll & Finan, 1993; Soto *et al.*, 2001). *S. meliloti* has a second malic enzyme, encoded by *tme*, with *in vitro* malate dehydrogenase activity, however strains with mutated *tme* alleles are able to fix nitrogen and overexpression of *tme* cannot complement a *dme* mutant's nitrogen fixation deficiency (Driscoll & Finan, 1996; Mitsch *et al.*, 2007). Mutations to PEP carboxylase gene *pck* form nodules morphologically similar to wildtype, but these nodules exhibit approximately 60% fixation efficiency (Finan *et al.*, 1991). However, no Pck activity was detected in wildtype nodules, while Dme was highly expressed, suggesting that Dme is the primary route for entry of carbon into the TCA cycle (Driscoll & Finan, 1996; Finan *et al.*, 1991).

The role of dicarboxylates in symbiosis is further supported by mutations to genes encoding TCA cycle enzymes succinate dehydrogenase (*sdh*), malate dehydrogenase (*mdh*), isocitrase dehydrogenase (*icd*), 2-oxoglutarate dehydrogenase, aconitase (*acnA*), and citrate synthase (*gltA*) which are incapable of N_2 fixation despite forming nodules (Duncan & Fraenkel, 1979; Dymov *et al.*, 2004; Koziol *et al.*, 2009; McDermott & Kahn, 1992; Mortimer *et al.*, 1999).

In addition to the requirement for TCA cycle enzymes, *S. meliloti* bacteroids also appear to be dependent on gluconeogenic enzymes for N_2 fixation. Mutations to enolase (*eno*), glyceraldehyde-3-phosphate dehydrogenase (*gap*), and 3-phosphoglycerate kinase (*pgk*) are essentially Fix^- (Finan *et al.*, 1991), suggesting that carbohydrate synthesis is important for symbiosis.

The ammonia generated from nitrogenase exits the bacteroid via passive diffusion. It is assumed to be protonated in the acidic peribacteroid space becoming ammonium before migration into the cytosol (Day *et al.*, 2001). There it is assimilated into the amino acid

glutamine (Gln) using plant-derived glutamate (Glu) as a scaffold by Gln synthetase (GS) (Fig. 1.11). MtGS1a is the primary contributor to GS activity in *M. truncatula* (Carvalho *et al.*, 1997; Carvalho *et al.*, 2000) and inhibition of GS activity results in significant metabolic changes as well as premature nodule senescence (Seabra *et al.*, 2012).

The amide group is subsequently transferred to α ketoglutarate by Glu oxoglutarate aminotransferase (GOGAT) forming Glu. The Glu produced from this reaction can be recycled for repeated generation of Gln by GS (GS-GOGAT pathway) or converted into aspartate (Asp) and asparagine (Asn), by Asp aminotransferase and Asn synthetase respectively, for uptake (Fig. 1.11). Antisense repression of GOGAT in *M. sativa* resulted in plants exhibiting symptoms of N starvation despite having wildtype N_2 fixation levels (Schoenbeck *et al.*, 2000).

Alternatively, in legumes that form determinate nodules such as soybean or common bean, Gln formed from GS activity can be used for synthesis of ureides for N assimilation (Fig. 1.11) (Tajima *et al.*, 2004).

1.4 Sugar alcohols

1.4.1 Sorbitol, mannitol, and D-arabitol

Mannitol has long been considered a preferred carbon source for rhizobia and is routinely included in culture media for enrichment of soil bacteria (Vincent, 1970). Two distinct NAD^+-dependent sugar alcohol dehydrogenase activities were detected in extracts from *S. meliloti* cells grown on mannitol, one which acted exclusively on sorbitol (glucitol) and another which acted on mannitol and D-arabitol (Martínez De Drets & Arias, 1970). Fructose was determined to be the product of sorbitol and mannitol oxidation while xylulose is the product of D-arabitol oxidation (Martínez De Drets & Arias, 1970). Fructose kinase activity was detected in extracts grown on either sorbitol or mannitol and a mutant lacking the activity is unable to utilize sorbitol, mannitol, or fructose as a sole carbon source (Gardiol *et al.*, 1980; Martínez De Drets & Arias, 1970). A mutant lacking phosphoglucose isomerase activity is deficient for growth on several carbon sources including sorbitol, mannitol, and fructose, suggesting that sorbitol and mannitol enter central metabolism using fructose-6-phosphate and glucose-6-phosphate as intermediates (Arias *et al.*, 1979).

Some of these functions have been tentatively identified as being encoded on the *S. meliloti* chromosome. A putative ABC type transporter (*smoEFGK*) is induced by sorbitol and mannitol (Mauchline *et al.*, 2006) and a mutation to a putative sorbitol dehydrogenase (*smoS*) reduced growth on several polyols including sorbitol, mannitol, and maltitol (Jacob *et al.*, 2008). A mutation to putative xylulose kinase *xylB* abolishes growth on D-arabitol (Geddes & Oresnik, 2012b).

1.4.2 Galactitol

Catabolism of galactitol (dulcitol) has been characterized in *E. coli*, in which it is phosphorylated upon uptake, oxidized, phosphorylated again, and split into glyceraldehyse-3-phosphate and dihydroxyacetone phosphate (Brinkkötter *et al.*, 2002; Lengeler, 1975; Lengeler, 1977; Nobelmann & Lengeler, 1995; Nobelmann & Lengeler, 1996). Galactitol dehydrogenase activity was not detected in extracts from galactitol grown *S. meliloti* cells suggesting that metabolism of galactitol may proceed differently in *S. meliloti* (Martínez De Drets & Arias, 1970). Mutants containing large deletions in pSymB were constructed to facilitate mapping of the pSymB replicon, these experiments identified a region that is essential for growth on galactitol (Charles & Finan, 1990, 1991). Consistent with this observation, an ABC type transporter (*SMb21375-7*) encoded on pSymB is induced by galactitol (Mauchline *et al.*, 2006).

1.4.3 Erythritol, adonitol, and L-arabitol

Metabolism of erythritol, adonitol (ribitol), and L-arabitol proceeds via enzymes encoded in a single locus on the *S. meliloti* chromosome (Geddes & Oresnik, 2012b; Geddes *et al.*, 2010). The three polyols were shown to utilize the ABC transporter MptABCDE (multiple polyol transporter) for uptake (Geddes & Oresnik, 2012b). Subsequently, erythritol is metabolized by products of *eryABCHI* forming erythrose-4-phosphate (Barbier *et al.*, 2014; Hawkins *et al.*, 2018). Similarly, adonitol and L-arabitol are degraded via EryA and also make use of RbtABC, additionally epimerase LalA acts exclusively in the catabolism of L-arabitol, resulting in the generation of ribulose-5-phosphate (Geddes & Oresnik, 2012b). Both erythrose-4-phosphate and ribulose-5-phosphate are intermediates in the pentose phosphate pathway (Hawkins *et al.*, 2018). Mutants deficient in erythritol catabolism were shown to be less competitive for nodule

occupancy in *R. leguminosarum* bv. *viciae* (Yost *et al.*, 2006). However, *S. meliloti* erythritol mutants were not affected with respect to competition for alfalfa nodules (Geddes *et al.*, 2010).

1.4.4 Glycerol

Utilization of glycerol is a common feature among many rhizobia (Arias & Martinez-Drets, 1976; Ding *et al.*, 2012). Glycerol metabolism is not well studied in *S. meliloti*, but in *R. leguminosarum* bv. *viciae* it is mediated by a locus containing an ABC transporter *glpSTPQUV* as well as a glycerol kinase (*glpK*) which acts on glycerol to form glycerol-3-phosphate, and a glycerol-3-phosphate dehydrogenase (*glpD*), producing dihydroxyacetone phosphate (DHAP) (Ding *et al.*, 2012). Mutants at this locus could not grow on glycerol as a sole carbon source and were less competitive for nodule occupancy on peas (Ding *et al.*, 2012).

A homologous operon can be found on the *S. meliloti* chromosome, but *glpK* is absent from this locus and encoded in a region on pSymB (Aneja & Charles, 1999; Ding *et al.*, 2012). Consistent with these determinations, mutations in triose phosphate isomerase (*tpiA*) in *S. meliloti* are unable to grow using glycerol as a sole carbon source (Poysti & Oresnik, 2007). However, *tpiA* mutant strains are not deficient in competition for nodule occupancy (Poysti & Oresnik, 2007).

1.4.5 Cyclic polyols

The sugar alcohols discussed up to this point have been linear compounds that vary in the length of their carbon skeletons, but there are cyclic polyols as well, many of which fall under the general name inositol. There are nine isomers of inositol, seven of these are meso compounds (*cis*, *epi*, *allo*, *neo*, *myo*, *muco*, and *scyllo*) and two are chiral pairs (L-*chiro* and D-*chiro*)

(Loewus & Dickinson, 1982). Bacterial metabolism of inositol is best described in *Bacillus subtilis* in which the genes *iolABCDEFGHIJ* and *iolT* mediate catabolism under the control of negative regulator *iolR* (Yoshida *et al.*, 2008). Inositol is broken down into DHAP and acetyl-CoA.

Metabolism of *myo*, *scyllo*, and D-*chiro*-inositol is thought to proceed via similar route using the genes *iolA* and *iolRCDEB* in *S. meliloti.* Additionally, genes *idhA* and *SMc01163* are thought to be involved in D-*chiro* and *scyllo*-inositol respectively (Kohler *et al.*, 2010). Two ABC transport genes were induced by *myo*-inositol, *ibpA* which is likely associated with neighboring *iatAP*, and *SMb20072* (Mauchline *et al.*, 2006). Mutations to the catabolic genes could not use inositol as a sole carbon source and were less able to compete for nodule occupancy (Kohler *et al.*, 2010).

R. leguminosarum was shown to utilize *iolA* and *iolDEB* for catabolism of *myo*-inositol, and the genes *intABC* for transport. Mutations at these loci exhibited reduced growth on *myo*-inositol and were also less competitive for nodule occupancy on vetch (Fry *et al.*, 2001). That both *S. meliloti* and *R. leguminosarum* mutants are compromised for competition suggests that inositol is a general determinant for nodule occupancy. Inositol derivatives pinitol and ononitol were detected in high amounts in roots, nodule cytosol, and bacteroids of *M. sativa* (Fougère *et al.*, 1991).

1.5 Thesis objectives

The subject of this thesis is to investigate sugar alcohol metabolism in *S. meliloti* Rm1021 with regard to regulation, uptake, catabolic routes, and effect on competition for nodule occupancy. The rationale for conducting research in this area is as follows. *S. meliloti* has been established as a model organism for carbon utilization (Geddes & Oresnik, 2014), but much of the metabolic potential of this bacterium has yet to be described. Further elucidating the nuances of carbon metabolism in this strain increases its value as a model organism. Carbon metabolism has been shown to influence symbiosis and competition for nodule occupancy. Continued study in this area could aid in the development of a robust inoculum strain for use in agriculture, and contribute to our understanding of the *Sinorhizobium-Medicago* symbiotic model. Additionally, if we are to have any hope of engineering nitrogen fixation into crop plants, a thorough understanding of the metabolic mechanisms involved in symbiosis will be absolutely necessary. Finally, in this era of high throughput sequencing and bioinformatics, studies that clearly identify gene function are of paramount importance for validation and curation of data deposits.

The first objective of this thesis was to identify and characterize the genetic locus for galactitol catabolism in *S. meliloti* that was tentatively located during development of a genetic map of the pSymB replicon (Charles & Finan, 1990, 1991). Loss of the ability to utilize galactitol resulted in attenuated virulence of *Salmonella enterica* serovar typhimurium with animal models (Nolle *et al.*, 2017) suggesting that galactitol metabolism may be important for other host-microbe interactions.

The second objective was to characterize sorbitol, mannitol, and D-arabitol metabolism in *S. meliloti*. While the biochemical steps of the pathways for the breakdown of these substrates had been identified, the specific genes encoding these functions were yet to be determined. There

is strong evidence that the chromosomal *smo* locus is integral to metabolism of these polyols (Jacob *et al.*, 2008; Mauchline *et al.*, 2006).

The third objective was to analyze xylitol utilizing suppressor mutants of *S. meliloti* Rm1021. These mutations arose spontaneously following growth of Rm1021 on defined medium with xylitol. The increased catabolic capability of this mutant is of interest from the perspective of metabolic modeling, genetic engineering, and plant-microbe interactions.

Chapter 2:

Galactitol utilization in *Sinorhizobium meliloti* is dependent on a chromosomally encoded sorbitol dehydrogenase and a pSymB encoded operon necessary for tagatose catabolism

This work was carried out by MacLean Kohlmeier in collaboration with Catherine White and Jane Fowler. CW performed the primer extension assay. JF made the mutations at the *tag* locus and collected the *gusA* expression data.

2.1 Abstract

The legume endosymbiont *Sinorhizobium meliloti* can utilize a broad range of carbon compounds to support its growth. The linear, six-carbon polyol galactitol is abundant in vascular plants and is metabolized in *S. meliloti* by the contribution of two loci *SMb21372-SMb21377* and *SMc01495-SMc01503*, which are found on pSymB and the chromosome, respectively. The data suggest that several transport systems, including the chromosomal ATP-binding cassette (ABC) transporter *smoEFGK*, contribute to the uptake of galactitol, while the adjacent gene *smoS* encodes a protein for oxidation of galactitol into tagatose. Subsequently, genes *SMb21374* and *SMb21373* encode proteins that phosphorylate and epimerize tagatose into fructose-6-phosphate, which is further metabolized by the enzymes of the Entner-Doudoroff pathway. Of note, it was found that *SMb21373*, which was annotated as a 1,6-bis-phospho-aldolase, is homologous to the *E. coli* gene *gatZ*, which is annotated as encoding the noncatalytic subunit of a tagatose-1,6-bisphosphate aldolase heterodimer. When either of these genes was introduced into an *Agrobacterium tumefaciens* strain that carries a tagatose-6-phosphate epimerase mutation, they were capable of complementing the galactitol growth deficiency associated with this mutation; strongly suggesting that these genes are both epimerases. Phylogenetic analysis of the protein family (IPR012062) to which these enzymes belong, suggests that this misannotation is systemic throughout the family. *S. meliloti* galactitol catabolic mutants do not exhibit symbiotic deficiencies or the inability to compete for nodule occupancy.

2.2 Introduction

Sinorhizobium meliloti 1021 is a Gram-negative α proteobacterium belonging to the family Rhizobiaceae. It has a tripartite genome consisting of a 3.65 Mb chromosome, a megaplasmid of 1.6 Mb (pSymA), and a chromid of 1.7 Mb (pSymB) (Capela *et al.*, 2001; diCenzo & Finan, 2017; Downie & Young, 2001; Galibert *et al.*, 2001). *S. meliloti* can be found as a free-living soil microbe or as a participant in an endosymbiotic relationship with legumes of the genus *Medicago*, such as *M. sativa* (alfalfa) or *M. truncatula* (barrel medic). During symbiosis, *S. meliloti* fixes biologically inert, atmospheric dinitrogen gas into ammonia, which can be used by the plant for growth (Geddes & Oresnik, 2016).

Nitrogen fixation is an especially important process for legumes as nitrogen is often a major limiting factor in soil (Vitousek *et al.*, 2002). In return for fixed nitrogen, the plant provides a continuous supply of carbon, in the form of dicarboxylic acids, to the bacterium, as well as an enclosed habitat within a plant-derived organ called a nodule (Geddes & Oresnik, 2016; Udvardi & Poole, 2013). The nitrogen fixed by a well nodulated plant can satisfy the entire nitrogen requirement of the legume. In addition to its role in agriculture, *S. meliloti* has become known as a model organism for the study of bacterial carbon metabolism (diCenzo *et al.*, 2016; Geddes & Oresnik, 2014; Jacob *et al.*, 2008; Mauchline *et al.*, 2006). Well-defined model systems are essential for experimental design and testing hypotheses.

Many different Rhizobial strains and species are present in the soil, but these organisms differ in nitrogen fixation efficiency as well as their ability to nodulate their hosts. Strains that are more competitive for nodule occupancy may not be well suited for symbiosis, thereby limiting the interactions of efficient nitrogen fixing organisms with legumes and reducing the effect of an inoculum. This phenomenon is known as the Rhizobium competition problem

(Triplett & Sadowsky, 1992). The ability of a bacterium to catabolize sugar alcohols has been shown to affect its ability to compete for nodule occupancy (Ding *et al.*, 2012; Fry *et al.*, 2001; Yost *et al.*, 2006).

In plants, sugar alcohols can be used to protect plant cells from osmotic stress and are also primary products of photosynthesis (Bieleski, 1982). Sorbitol, mannitol and galactitol (also known as dulcitol) are the most common sugar alcohols found in vascular plants (Williamson *et al.*, 2002). It is estimated that plants can secrete nearly 30% of the carbon allocated to their roots into the rhizosphere (Jones *et al.*, 2009).

Galactitol is a six-carbon polyalcoholic sugar, and oxidation of carbon one yields the aldose sugar galactose. Alternatively, oxidation of carbon two produces the ketose sugar, tagatose. The best-characterized galactitol catabolic system is *E. coli*, in which the *gatYZABCDR* (galactitol) operon is responsible for the majority of the metabolic functions (Nobelmann & Lengeler, 1995; Nobelmann & Lengeler, 1996). Initially, galactitol is taken up and phosphorylated though a phosphoenolpyruvate (PEP)-dependent galactitol phosphotransferase (PTS) system encoded by *gatABC*. The resulting intermediate, galactitol-1-phosphate, is oxidized by the dehydrogenase GatD, forming tagatose-6-phosphate (T6P). Subsequently, T6P is phosphorylated a second time, by kinase PfkA (Brinkkötter *et al.*, 2000), generating tagatose-1,6-bisphosphate (T1,6bisP) before being hydrolyzed into glyceraldegyde-3-phosphate (GAP) and dihydroxyacetone phosphate (DHAP) by a heterodimeric bisphosphoaldolase encoded by *gatYZ* (Lengeler, 1975; Lengeler, 1977). Recently this pathway was also characterized in *Salmonella enterica* serovar *typhimurium* (Nolle *et al.*, 2017).

Initial characterization of the *S. meliloti* pSymB chromid was facilitated by construction of targeted deletions, which showed that the ability to utilize galactitol as a sole carbon source

was dependent on a 60kb region on pSymB adjacent to the EPS II biosynthesis genes (Charles & Finan, 1990, 1991). This locus was referred to as *dul* (dulcitol utilization) on the genetic map (Charles & Finan, 1991). The purpose of this work is to delineate and characterize this region and to determine its contribution to galactitol metabolism in *S. meliloti*.

2.3 Materials and methods

2.3.1 Bacterial strains, plasmids, and culture conditions

Bacterial strains and plasmids used during this work are listed in Table 2.1. *E. coli* strains were grown at 37°C on Luria-Bertani (LB) medium (Cold Spring Harbor Protocols, 2006). *Agrobacterium tumefaciens* strains were grown at 30°C on LB medium as a complex medium or on Vincent's Minimal Media (VMM) as a defined medium (Vincent, 1970). *S. meliloti* strains were grown at 28°C on LB medium as a complex medium, VMM or M9 minimal medium as a defined medium. Carbon sources were filter sterilized and added to defined media to a final concentration of 15 mM. Antibiotics were used in solid media at the following concentrations: chloramphenicol (Cm), 20 µg/ml; gentamicin (Gm), 20 or 60 µg/ml; kanamycin (Km), 20 µg/ml; neomycin (Nm), 200 µg/ml; rifampicin (Rif), 50 or 100 µg/ml; spectinomycin (Sp), 100 or 200 µg/ml; streptomycin (Sm), 200 µg/ml; tetracycline (Tc), 5 µg/ml. If more than one concentration is listed, the larger amount was used on *S. meliloti*, while the lower amount corresponds to *E. coli*. Antibiotic concentrations were generally halved when added to liquid medium.

Table 2.1. Bacterial strains and plasmids

Strain	Relevant Characteristics	Reference
S. meliloti		
Rm1021	SU47 *str-21*	Meade *et al.* (1982)
RmP110	Rm1021 *pstC*$^{+}$	Yuan *et al.* (2006a)
RmP187	Rm1021 *SMb21373*::*gusA*, Nm^{R}	This work
RmP195	Rm1021 *SMb21375*::*gusA*, Nm^{R}	This work
RmP233	Rm1021 *SMb21377*::*gusA*, Nm^{R}	This work
RmP245	Rm1021 *SMb21377*::*gusA*, *SMb21372*ΩSp-Sm	This work
SMc01500	RmP110 *smoS*::pTH1703, Gm^{R}	Jacob *et al.* (2008)
SRmD491	Rm1021 *smoE*::Tn*5*-B20, Nm^{R}	This work
SRmD492	Rm1021 *smoF*::Tn*5*-B20, Nm^{R}	This work
SRmD493	Rm1021 *smoS*::Tn*5*-B20, Nm^{R}	This work
SRmD495	Rm1021 *smoK*::Tn*5*-B20, Nm^{R}	This work
SRmD523	Rm1021 *smoM*::pKNOCK-Gm, Gm^{R}	This work
SRmD615	Rm1021 *SMb20410*::pKNOCK-Tc, Tc^{R}	This work
A. tumefaciens		
C58	Wildtype, nopaline type Ti plasmid, biovar I	Hamilton *et al.* (1971)
Atu1	C58 *agaZ*::pKNOCK-Gm, Gm^{R}	This work
E. coli		
DH5α	F^{-} *supE44 lacU169 hsdR17 recA1 endA1 gyrA96 thi-1 relA1* (80*lacZ*ΔM15)	Hanahan (1983)
DH5αR λpir	λpir lysogen of DH5α	House *et al.* (2004)
MM294A	*pro-82-thi-1 hsdR17 supE44*	Finan *et al.* (1986)
MT607	MM294A *recA56*	Finan *et al.* (1986)
MT616	MT607 (pRK600)	Finan *et al.* (1986)
EcA101	MT607ΩTn*5*-B20	Clark *et al.* (2001)
Plasmids		
pMK28	pRK7813/*SMb21373-4*, Tc^{R}	This work
pMK30	pRK7813/*agaZ*, Tc^{R}	This work
pMK33	pCO37/*SMb21373*, Tc^{R}	This work
pMK34	pCO37/*SMb21374*, Tc^{R}	This work
pMK42	pRK7813/*gatZ*, Tc^{R}	This work
pMK43	pRK7813/*kbaZ*, Tc^{R}	This work
pMK45	pRK7813/*gatZ* codon optimized, Tc^{R}	This work
p*SMc01500*	*SMc01500* expressing plasmid, Tc^{R}	This work
pFL4516	*SMb21377* and upstream promoter region	This work
pRK7813	Broad host range vector, Tc^{R}	Jones *et al.* (1987)
pCO37	Gateway compatible expression vector, Tc^{R}	Jacob *et al.* (2008)
pTH1360	*gusA* reporter suicide vector, Nm^{R}	Yuan *et al.* (2006b)
pKNOCK-Gm	Suicide vector, Gm^{R}	Alexeyev (1999)
pKNOCK-Tc	Suicide vector, Tc^{R}	Alexeyev (1999)
pRK600	pRK2013 *nptI*::Tn*9*, Cm^{R}	Finan *et al.* (1986)
pXINT129	*λint* and *xis* driven by P$_{lac}$, Km^{R}	Platt *et al.* (2000)

2.3.2 DNA manipulations and genetic techniques

Standard techniques were used for plasmid isolation, restriction enzyme digests, ligations, transformations and agarose gel electrophoresis (Sambrook & Russell, 2001). Conjugations and transductions were carried out essentially as previously described (Finan *et al.*, 1984; Finan *et al.*, 1985). Isolation of complementing cosmids, their mutagenesis using Tn*5*-B20, and determining insertion sites was done essentially as previously described (Poysti *et al.*, 2007), resulting in the generation of strain SRmD491-3 and SRmD495. The plasmids pMK33, pMK34, pSMc01500 were constructed using the broad-host-range Gateway compatible vector pCO37 (Jacob *et al.*, 2008), and the *S. meliloti* ORFeome using previously described methods (House *et al.*, 2004; Schroeder *et al.*, 2005). The plasmids pMK28, pMK30, pMK42, pMK43, and pMK45 were generated using the primers *tagDE*_F/R, *agaZ*_F/R, *gatZ*_F/R, *kbaZ*_F/R, and gatZ1_F/R respectively (Table 2.2). These constructs were designed such that the forward primer contained a *Hind*III restriction site and a ribosome-binding site to facilitate translation, and the reverse primer contained an *Eco*RI site. The appropriate template DNA was amplified, and subsequently cloned in pRK7813 such that the open reading frames were constitutively expressed by the P_{lac} promoter from the multiple cloning site. Codon optimization of the *E. coli gatZ* was carried out by GenScript (Piscataway, NJ, USA) and supplied as an insert on pUC57, which was subsequently used as template to generate pMK45. All constructs were confirmed by nucleotide sequencing.

To construct *gusA* reporter gene fusion strains in the tagatose operon, an internal gene fragment from each of *SMb21373*, *SMb21375*, and *SMb21377* was amplified using the primers listed in Table 2.2, cloned into the suicide vector pTH1360 (Yuan *et al.*, 2006b) and mobilized into *S. meliloti* Rm1021 by conjugation. The resulting recombinants were

Table 2.2. Primers used during this study

Name	Sequence 5′→3′
Generation of mutants	
ML1822	GTGCTCAGCAAGACGAGCTTCGG
ML1828	AA**TCTAGA**TCAGCGCTTTGCCTGCTTGGC
ML2467	CCC**AAGCTT**CTTCGTCATCAATCAGATCG
ML2468	AA**TCTAGA**TCAGCGCTTTGCCTGCTTGG
ML3145	ATGAACATGCCCCGCATTCTCAAAACTC
ML3146	AA**TCTAGA**TGCCGATATGCGACGTCACCG
ML7348	CC**AAGCTT**GACGCTTCGACCGTGACG
ML7349	CGC**GGATCC**CGTTGTAGTTCTGCGC
ML7350	GC**GTCGAC**GGATCCGGTGATTGATTGAGC
ML758	TCAAGCTTGCATGCCTGC
*agaZ*int_F	ATAT**GGATCC**CACCGCGCCGCGAAACTC
*agaZ*int_R	ATAT**CTGCAG**TATCCGTCGCGTACCAGTG
*SMb20410*_intF	ATAT**GGATCC**CTGACCGCAAGCCATTCCTG
*SMb20410*_intR	ATAT**GGTACC**AATTCGGTCCGAAATAGGCGT
*smoM*1_F	ATAT**TCTAGA**CTTCGCCCTCAATGCCCGCG
*smoM*1_R	ATAT**CTCGAG**CATGGCATGCACGGTCGGGC
Confirmation of mutants	
*SMb20410*_F	ATGCTGAAACAGATCACTGCAAT
*SMb20410*_R	CTAGCTTTTCAACGCGCTCTC
*smoM*_F	ATGGATCGCCGTTCATTCATCAA
*smoM*_R	TCAGATCTTGCCGTTACGCTGC
Generation of expression constructs	
*agaZ*_F	ATAT**AAGCTT**GGAGATGCATGCATGACCGCCATTTTGGAAAATC
*agaZ*_R	ATAT**GGATCC**TCAGTGGCGGCCTTCGCC
*tagDE*_F	ATAT**AAGCTT**GGAGATGCATGCATGCACAAGATACTGACGATCG
*tagDE*_R	ATAT**GGATCC**CTAGACCCGGCACGCTGC
*gatZ*_F	ATAT**AAGCTT**GGAGATGCATGCATGAAAACGTTAATTGCCCGGC
*gatZ*_R	ATAT**GAATTC**TTATTCCGCACAGCCGTAGC
*kbaZ*_F	ATAT**AAGCTT**GGAGATGCATGCGTGAAACATCTGACAGAAATGG
*kbaZ*_R	ATAT**GGATCC**TTATTGGCCTTCACAGGCTG
*gatZ*1_F	ATAT**AAGCTT**GGAGATGCATGCATGAAGACCCTCATCGCCC
*gatZ*1_R	ATAT**GGATCC**TCACTCGGCGCAGCCATAG
Primer extension	
CWF4	CAGCAATGTGAGCGCGGTGCCGGCG

Restriction sites are in bold.

single colony purified and subsequently transduced into Rm1021. These constructions yielded RmP187, RmP195, and RmP233. RmP245 was constructed by first amplifying and cloning the *SMb21372* open reading frame into vector pUCP30T, which carries a gentamicin resistance marker. The Sm/Sp^R cassette from pHP45 (Prentki & Krisch, 1984) was amplified with primers containing *Sal*I restriction sites on the ends. This fragment was subsequently cloned into two natural sites that occur at bp 466 and 535 of *SMb21372*. The resulting construct was mated into RmP233 and a Sp^R Gm^S recombinant was identified and confirmed.

S. meliloti strains SRmD523 and SRmD615, as well as *A. tumefaciens* strain Atu1 were constructed utilizing pKNOCK suicide vectors (Alexeyev, 1999; MacLean *et al.*, 2006). Briefly, an internal fragment from each gene was amplified and cloned into the appropriate vector. These were then subsequently mobilized into the appropriate wild-type recipient strains by conjugation. Recombinants were verified by generating a PCR product of the junctions, followed by nucleotide sequencing.

2.3.3 Primer extension assays

Assays were carried out as previously described (MacLean *et al.*, 2006). Briefly, RNA isolated from wildtype *S. meliloti* strain RmP110, and RmP110 expressing pFL4516, was used as a template for primer extension reactions. Cells were grown to an OD_{600} of approximately 0.4 in M9 minimal medium containing glycerol or tagatose as a sole carbon source to support growth and induce transcription. The primer CWF4 (Table 2.2) was end labeled using [γ-^{32}P]dATP and T4 polynucleotide kinase. Unincorporated [γ-^{32}P]dATP was removed using a QIAquick nucleotide removal column. Primer extension reactions were carried out by adding $2x10^5$ cpm of radiolabeled primer to 20 µg of RNA template, 20 nmole of dNTP's, and reverse transcriptase

buffer up to a total volume of 16 µl. Extension products and sequencing reactions were separated by electrophoresis on an 8% polyacrylamide gel containing 7.7 M urea. Data were analyzed using a Storm820 phosphorimager.

2.3.4 Biochemical enzyme assays

Dehydrogenase assays were performed as previously described (Geddes & Oresnik, 2012a; Pickering & Oresnik, 2008). Briefly, strains containing an empty vector control or a complementing plasmid were grown in defined medium containing glycerol for two days at 30°C followed by induction with 15 mM sorbitol for six hours. Cell free lysates were prepared and separated by non-denaturing polyacrylamide gel electrophoresis (PAGE). Subsequently, gels were developed in an assay reagent containing p-nitroblue tetrazolium, phenasine methosulfate, NAD^+ as a cofactor, and a substrate of interest.

β glucuronidase assays were performed as previously described (MacLean *et al.*, 2006; MacLean *et al.*, 2009). Briefly, cultures were grown in defined media containing a carbon source of interest to an OD_{600} between 0.3-0.6. Cells were washed with saline and resuspended in GUS buffer. Cultures were further diluted in GUS buffer and a drop of toluene was added to each sample to permeabilize the cells. Samples were vortexed, incubated at 37°C for 30 mins and then equilibrated to room temperature before the addition of p-nitrophenyl β-glucuronide (PNPG). Reactions were allowed to proceed for 20 mins to an hour before being stopped by the addition of sample aliquots to stop buffer solution. Samples were centrifuged to remove cell debris before being measured for optical density at 405 nm. Activity is reported in Miller units and was calculated using the formula:

Activity = $(1000 \times A_{405}) / (A_{600} \times$ reaction time in minutes x volume (ml) of culture use in assay).

Kinase assays were performed as described previously (Geddes & Oresnik, 2012b; Rivers & Oresnik, 2013). Briefly, cultures were grown in defined medium containing glycerol or galactitol and glycerol for two days at 30°C. Cell free extracts were prepared and assayed for kinase activity using a coupled enzyme assay containing 60 mM HEPES pH 7.5, 6 mM $MgCl_2$, 3 mM ATP, 3 mM PEP, 0.3 mM NADH and 1/50 vol pyruvate kinase/lactate dehydrogenase mix. Assays were initiated by the addition of 6 mM hexose and data were collected every ten seconds for ten minutes. Rates were measured at 340 nm from the linear portion of the curves, adjusted for background oxidase activity, and normalized to total protein. Protein concentration was determined using the Lowry protein quantification assay (Stoscheck, 1990).

2.3.5 Growth curve analysis

Overnight cultures were grown in complex medium to a defined OD_{600} and subcultured into fresh media, either complex or defined with galactitol, the following day. Subsequently, 200 µl aliquots were dispensed into a 96-well plate and incubated at 28°C with shaking. Measurements were taken every hour for 72 hrs. Growth (OD_{600}) was measured with a BioTek Synergy 2 plate reader (BioTek Instruments, Inc., Winooski, VT, USA.)

2.3.6 Transport assays

Transport assays were performed as previously described (Poysti *et al.*, 2007). Briefly, cultures were grown in defined media containing a carbon source of interest to an OD_{600} between 0.6-0.8, washed twice and resuspended to an OD_{600} of 0.3 in defined salts media. [^{14}C] mannitol was used as the labeled substrate and was added in the cultures to a final concentration of 2 µM, while unlabeled competing substrates were added to a final concentration of 2 µM or 10 µM. At

appropriate time points, 0.5 ml aliquots were filtered through 0.45 µm Hv filters on a Millipore sampling manifold. Accumulation of radiolabel was quantified using a liquid scintillation counter (Beckman LS6500). Uptake rates were standardized to total protein.

2.3.7 Competition for nodule occupancy assays

Assays were performed as previously described (Geddes & Oresnik, 2012a). Briefly, sterile alfalfa seedlings were planted in nitrogen free medium and, following three days of growth, were inoculated with a mixture of *S. meliloti* strains. Mutant cells were grown overnight and each was mixed with a wildtype culture of equivalent OD_{600}. Initial inoculum ratios were determined by plating and screening for proportion. After 28-35 days of growth, root nodules were harvested, surface sterilized, and finally each nodule was crushed in 50µl sterile water. These extracts were screened for the presence of the antibiotic marker and the inoculum and recovery ratios were determined. Significance was evaluated by using a paired *t* test, assuming that a *P* value of less than 0.05 indicated a significant difference in competitiveness.

2.3.8 Phylogenetic analysis

The amino acid sequences within the protein family IPR012062 were downloaded from the InterPro website as a FASTA file (Finn *et al.*, 2017). The sequences were aligned using the web-based multiple sequence alignment program Clustal Omega on the EBI (European Bioinformatics Institute) webserver using the default settings (Sievers *et al.*, 2011). The alignment was trimmed with TrimAl v. 1.3 using the Automated 1 method and converted to a relaxed PHYLIP format (PHYLIP/PHYLIP 4) using ReadAl v. 1.3; both utilities were accessed through the Phylemon 2.0 webserver (Capella-Gutiérrez *et al.*, 2009; Sánchez *et al.*, 2011). The

phylogeny was built with the RAxML BlackBox v. 8.2.10 mirror site on the CIPRES Science Gateway webserver using a CAT rate of heterogeneity model and a JTT protein substitution matrix (Miller *et al.*, 2010; Stamatakis, 2014). The tree was drawn using FigTree v. 1.4.3 and annotation information was added using a file downloaded from UniProtKB (Pundir *et al.*, 2017).

2.4 Results

2.4.1 Identification of an operon necessary for tagatose catabolism

A global expression map of *S. meliloti* ATP-binding cassette (ABC) and Tripartite ATP-independent periplasmic (TRAP) transporters identified inducing substrates for 76 transport systems, many of which had not been characterized up to that point (Mauchline *et al.*, 2006). The expression map identified several transporters that were induced by galactitol. Of those, the greatest change was an 18-fold induction of *SMb21377*. This gene was also induced by tagatose, as well as galactose (Mauchline *et al.*, 2006). Since the genes necessary for the catabolism of galactose have been described (Geddes & Oresnik, 2012a), this appeared to be an ideal candidate as the *dul* locus that was previously mapped to pSymB (Charles & Finan, 1991).

SMb21377 appears to be within a cluster of genes resembling an operon that consists of a regulator (*SMb21372*), a bisphosphate aldolase (*SMb21373*), a kinase (*SMb21374*) and an ABC transporter (*SMb21375*-*SMb21377*) (Fig. 2.1). We note that there is a 102 bp gap between *SMb21372* and *SMb21373*.

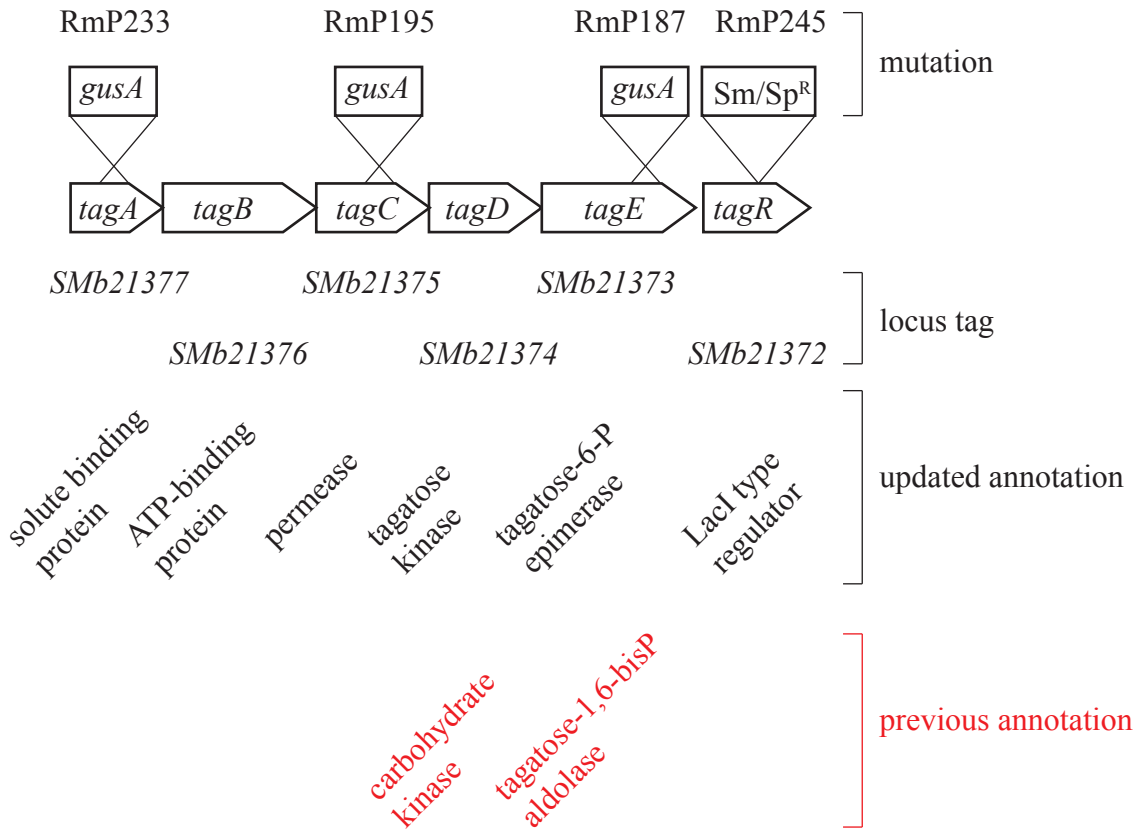


Figure 2.1. Locus diagram of *SMb21377-SMb21372* (*tag*) operon housing genes involved in galactitol catabolism.

Typically, sugar alcohol catabolism proceeds via oxidation of the substrate, subsequent phosphorylation, followed by entry into central metabolism (Mortlock, 1984). However, the gene annotation suggests a pathway in which a substrate is phosphorylated twice, before being split into two three-carbon compounds, reminiscent of the Embden-Meyerhof-Parnas pathway. In that case, the locus appears to be missing two genes, a dehydrogenase capable of galactitol oxidation as well as a second kinase that would be necessary for generation of a bisphosphorylated intermediate. Although this cluster does not contain all of the genes necessary to encode a complete pathway, it was reasoned that it was a likely contributor to galactitol catabolism due to the high induction of the transporter by galactitol and its location within a region that, when deleted, has been shown to abolish growth on galactitol (Fig. 2.1).

Characterization of the region was carried out by generation of series of insertion mutants that would result in a functional knockout of the entire operon (RmP233), or a subset of the metabolic genes (RmP195 and RmP187) (Fig. 2.1). These strains were all capable of growing on defined medium using glucose, galactose, sorbose, and glycerol, whereas they were unable to grow using either galactitol or tagatose as sole carbon sources. Of note, introduction of pMK28, which contains only the catabolic genes *SMb21374-3* (encoding a putative kinase and aldolase) (Fig. 2.1), into each of these strains was able to complement growth on galactitol and tagatose. Together these data suggest that the genes *SMb21377-3* are transcribed as an operon, and that the ABC transporter encoded by *SMb21377-5* is not necessary for growth, or may not be the only transporter in the *S. meliloti* genome that can transport galactitol or tagatose. The inability to utilize both galactitol and tagatose also suggests that galactitol metabolism proceeds using tagatose as an intermediate. The data also suggest a region responsible for the oxidation of galactitol to tagatose is elsewhere in the genome (Fig. 2.1, data not shown).

2.4.2 Identification of a negative regulator and the transcriptional start site of SMb21377

A gene encoding a putative negative regulator belonging to the LacI family lies immediately downstream from *SMb21373*. To determine if this gene (*SMb21372*) affects transcription of *SMb21377* and the genes lying downstream, it was hypothesized that an insertion into *SMb21372* should lead to constitutive transcription of these genes.

It was found that strains RmP233, RmP195, and RmP187, which contain *gusA* transcriptional fusions in *SMb21377*, *SMb21375*, and *SMb21373* respectively, displayed significantly higher expression when grown with galactitol, galactose, or tagatose than with glycerol (Table 2.3). In contrast, strain RmP245, which contains a *gusA* transcriptional fusion in *SMb21377* as well as an insertion interrupting *SMb21372*, showed high constitutive levels of expression regardless of the carbon source that was used for induction (Table 2.3). This suggests that SMb21372 negatively regulated the transcription of this operon at a promoter upstream of *SMb21377*, and that the transcription of *SMb21372* was not dependent on this promoter (Table 2.3).

To determine the transcriptional start site of the operon a primer extension analysis was carried out on RNA from the wild-type as well as from the plasmid pFL4516 which contains the region upstream from *SMb21377*. Whereas a primer extension product was not detected when the RNA from the wild-type grown on glycerol as a sole carbon source was used as a template, if RNA from tagatose-grown cells was used a single product that coincides with a “T” 77 bp upstream from the predicted ATG of *SMb21377* was obtained (Fig. 2.2, lanes 1-2). To corroborate this, RNA was also harvested from RmP1610 (RmP110 carrying pFL4516) from either medium containing glycerol or tagatose (Fig. 2.2, lanes 3-4). The results of the primer extension were consistent with what was found using the wild-type RNA as a template.

Table 2.3. Tagatose operon induction

Strain	Relevant Characteristics	Glycerol	Tagatose[a]	Galactitol	Galactose
Rm1021	Wild-type	14 ± 7	21	19 ± 35	16 ± 2
RmP233	*tagA::gusA*	16± 5	1056	1652±143	653±43
RmP195	*tagC::gusA*	98 ± 13	772	1416 ± 162	724 ± 40
RmP187	*tagE::gusA*	115 ± 5	1138	1312 ± 194	642 ± 63
RmP245	*tagA::gusA, tagR::Sp*^{R}	1658 ± 28	1352	1777 ± 29	985 ± 51

β glucuronidase activity measured in Miller units, data are reported as an average of three independent replicates with standard deviations. [a] Values are from an experiment that contained a single replicate due to availability of tagatose, the result was replicated showing similar induction values.

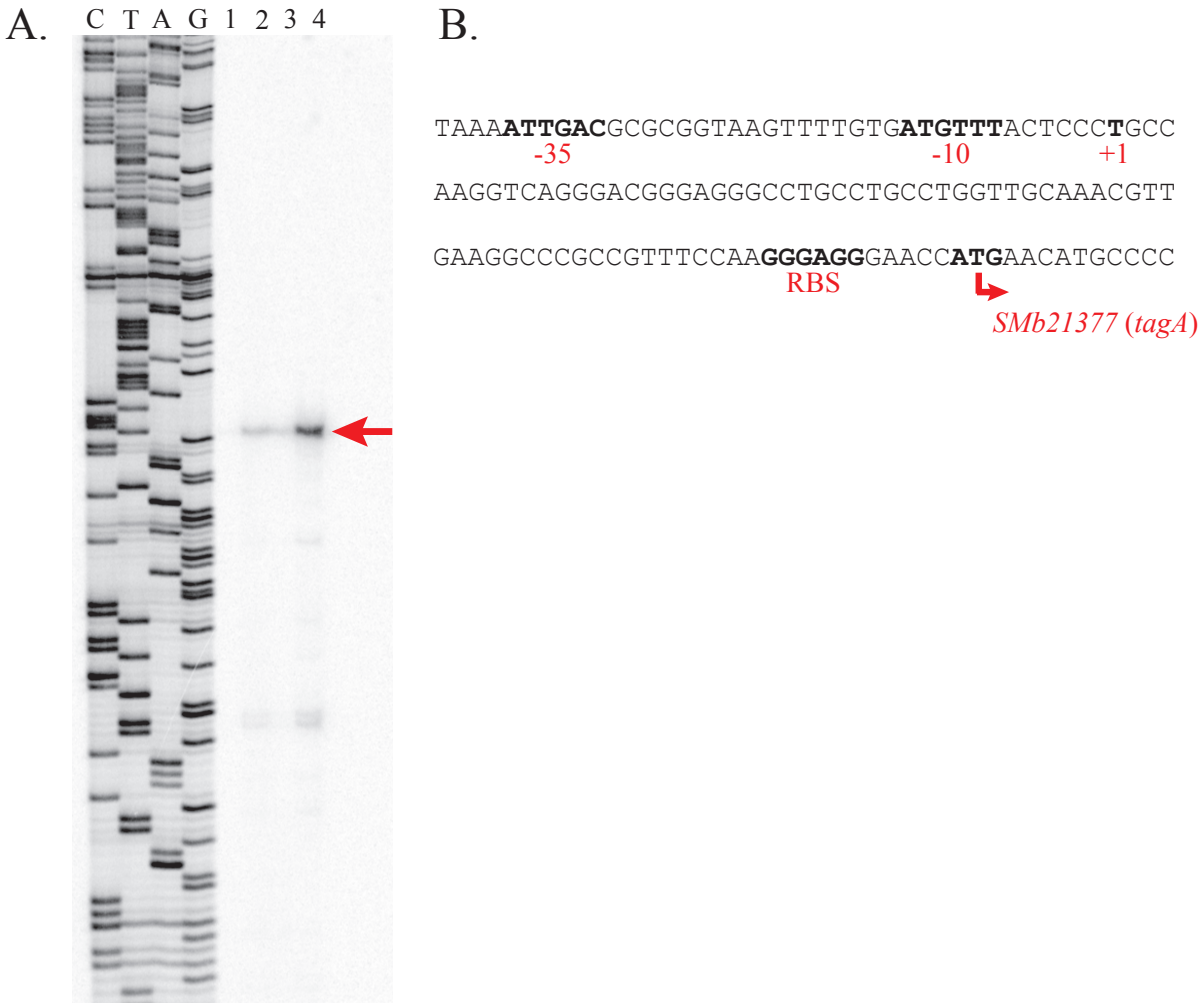


Figure 2.2. Analysis of the *SMb21377* (*tagA*) promoter. (A) Primer extension reactions were performed using RNA isolated from wild-type cells grown in glycerol (lane 1) or tagatose (lane 2) as well as RNA from wild-type cells expressing a *SMb21377* (*tagA*) promoter fragment *in trans* grown in glycerol (lane 3) and tagatose (lane 4). Sequencing reactions are included on the left. The extension products are indicated by an arrow. (B) Architecture of the *SMb21377* promoter. Key regions are in bold. Proposed translational start site is indicated with a bent arrow.

Of note is that the RNA harvested from RmP1610 grown on glycerol also showed a faint band corresponding to 77 bp upstream from the predicted translation start site (Fig. 2.2, lane 3). This is consistent with the operon being negatively regulated, but having excess copies of the promoter relative to the regulator (Fig. 2.2, Table 2.3). We note that there is a four-nucleotide repeat motif CGTT-N_{10}-CGTT that is found five base pairs upstream of the predicted ribosome-binding site for *SMb21377*. A search of the genome for homologous promoter sequences responsible for expression of a potential galactitol dehydrogenase did not reveal any promising candidates.

2.4.3 **SMb21374** *and* **SMb21373** *encode proteins with tagatose kinase and tagatose-6-phosphate epimerase activities*

The conversion of galactitol to tagatose can be carried out by a single redox reaction. Tagatose can enter central metabolism via two possible routes. The first more common route is that tagatose is phosphorylated twice becoming T1,6bisP before being split by an aldolase into GAP and DHAP, as is observed in *E. coli* (Brinkkötter *et al.*, 2002). The second, which occurs in *Agrobacterium tumefaciens*, is that tagatose is phosphorylated only once, becoming T6P, and is subsequently epimerized, to yield fructose-6-phosphate (F6P) (Wichelecki *et al.*, 2015). This is an attractive metabolic route for *S. meliloti* because there is only one annotated kinase at the locus (Fig. 2.1).

Using IMG's orthologous neighborhood viewer (Chen *et al.*, 2016), a search of the *S. meliloti* genome for orthologous genes revealed that *SMb21374* and *SMb21373* are orthologous to *scrK* (*Atu3166*) and *agaZ* (*Atu3167*) from *A. tumefaciens,* which are known to encode proteins with tagatose kinase and T6P epimerase activities respectively (Wichelecki *et al.*, 2015).

Therefore, despite the gene annotation of *SMb21373* as a bisphosphate aldolase, it seemed likely that tagatose catabolism in *S. meliloti* would proceed via a phosphorylation and a subsequent epimerization.

To test this hypothesis, kinase assays were conducted on cell-free extracts from the wildtype, RmP195 (kinase knockout), and RmP195 containing *SMb21374* constitutively expressed from a plasmid (kinase knockout complemented) that were grown with either glycerol, or glycerol with galactitol, as non-inducing and inducing conditions respectively. We were unable to detect tagatose kinase activity that was greater than background NADH oxidation rates in extracts from wildtype or knockout cells. However, extracts from cells that were constitutively expressing the kinase from a plasmid exhibited rates over background of 13.4 μmole/min/mg and 18.2 μmole/min/mg on glycerol and glycerol/galactitol respectively, suggesting that *SMb21374* encodes a tagatose kinase.

To determine if *SMb21373* encoded an epimerase, *agaZ* from *A. tumefaciens* was expressed in RmP187, which carries a mutation in *SMb21373,* in an attempt to heterologously complement the galactitol growth deficiency. Expression of either *agaZ* or *SMb21373 in trans* failed to restore growth on either galactitol or tagatose. Since overexpression of the wildtype allele also failed to complement, we suspected that the growth deficiency could be related to a disproportionate transcription of the epimerase, which may result in a buildup of a toxic metabolite rather than disparate enzyme activities. In an attempt to resolve this, an *agaZ* mutation was constructed by targeted mutagenesis of *A. tumefaciens* C58. Consistent with what has been previously published (Wichelecki *et al.*, 2015), the resultant mutant, Atu1, exhibited poor growth on galactitol as a sole carbon source.

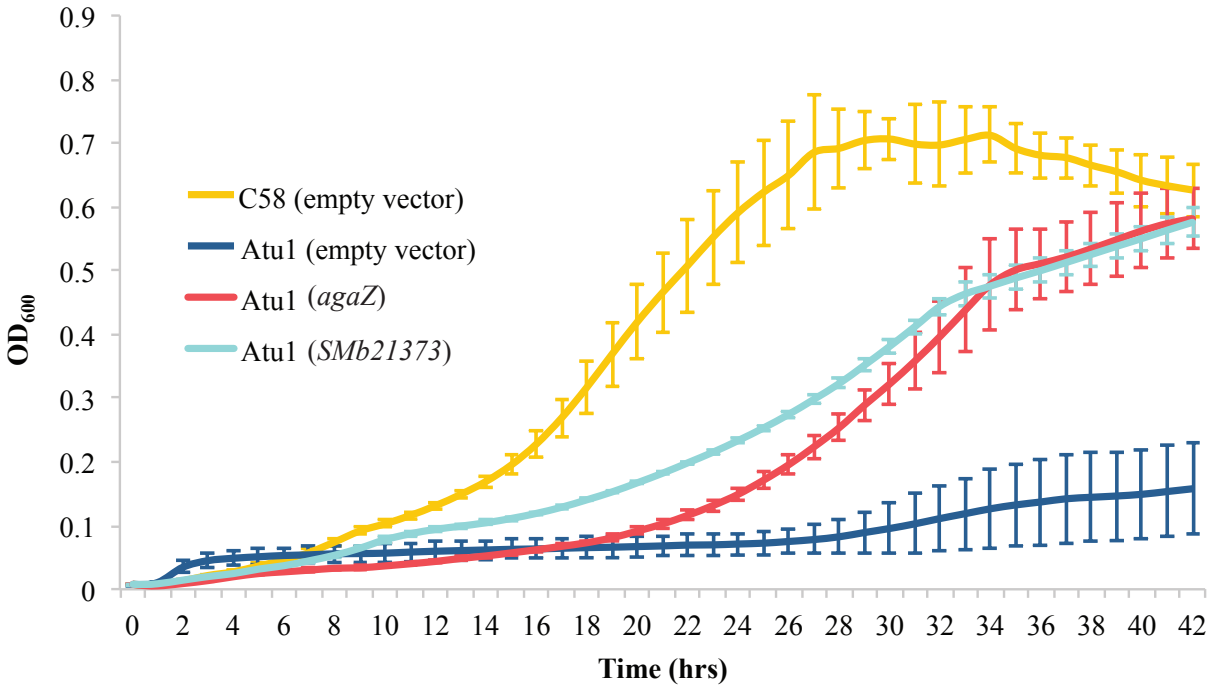


Figure 2.3. Growth curves comparing *A. tumefaciens* strains grown in defined media with galactitol as a sole carbon source. OD_{600} of wildtype, C58 (yellow); epimerase mutant, Atu1 (dark blue); mutant with *agaZ in trans* (red); and mutant with *SMb21373 in trans* (light blue) was measured every hour for 42 hrs. Data are expressed as the mean ± SD of three independent replicates.

Introduction of either *agaZ* or *SMb21373* on a broad host range plasmid into Atu1 resulted in a recovery of the ability to grow using galactitol as a sole carbon source (Fig. 2.3).

Taken together, these data suggest that the pSymB *dul* locus functions in tagatose metabolism with SMb21374 phosphorylating tagatose, and SMb21373 epimerizing T6P to produce F6P. We propose that the region be renamed the *tag* locus, with the genes from *SMb21377-SMb21372* being referred to as *tagABCDER* (Fig. 2.1).

2.4.4 SmoS is responsible for galactitol oxidation in S. meliloti

A large-scale mutagenesis of short-chain dehydrogenase/reductase (SDR) family genes in *S. meliloti* revealed 21 novel carbon utilization or symbiotic deficiencies (Jacob *et al.*, 2008). Of these, two genes relevant to sugar alcohol metabolism, *smoS* and *SMb20409*, were identified. It was determined that a *smoS* mutant could not grow on several sugar alcohols, including sorbitol and mannitol, as a sole carbon source, while a strain with a *SMb20409* mutation exhibited weak growth on galactitol as a sole carbon source (Jacob *et al.*, 2008).

The *smo* operon, putatively responsible for sorbitol/mannitol catabolism, was first described in *Rhodobacter sphaeroides* (Stein *et al.*, 1997). The operon is present on the chromosome in *S. meliloti*, and gene annotation suggests that it contains a transcriptional regulator (*smoC*), an ABC transporter (*smoEFGK*), a sorbitol dehydrogenase (*smoS*), a mannitol dehydrogenase (*mtlK*), a phosphatase (*SMc01502*) and a kinase (*SMc01503*). There is also an orphan solute binding protein, *smoM* (*SMc04251*), at a distant locus on the chromosome. *SMb20409* can be found on the pSymB megaplasmid adjacent to a predicted solute binding protein *SMb20410*. Both *smoS* and *SMb20409* have associated solute binding proteins (SBPs),

smoE and *SMb20410*, which have been shown to be induced by galactitol (Mauchline *et al.*, 2006).

Our carbon utilization screens of a strain carrying a *smoS* mutation indicated that in addition to the previously recorded phenotypes on sorbitol and mannitol (Jacob *et al.*, 2008), the strain was also unable to grow using galactitol as a sole carbon source, while utilization of tagatose is unaffected (Table 2.4). Complementation of a *smoS* mutant with a *smoS* expressing plasmid restored the mutant's ability to grow on sorbitol or galactitol as a sole carbon source (Table 2.4). These data are consistent with the hypothesis that *smoS* encoded the missing dehydrogenase responsible for the oxidation of galactitol into tagatose.

To confirm this hypothesis, cell free extracts were separated using non-denaturing polyacrylamide gel electrophoresis (PAGE) and stained for galactitol dehydrogenase activity. Extracts from wildtype cells showed a single band of activity when galactitol was present as a substrate (Fig. 2.4, lane 1). No bands were visible when extracts from the *smoS* mutant were tested (Fig. 2.4, lane 2). However, extracts from mutant cells with *smoS* complemented on a plasmid showed a band of activity with similar intensity and migration distance to the wildtype band (Fig. 2.4, lane 3). These results indicate that galactitol dehydrogenase activity is present in wildtype *S. meliloti* cells and that SmoS is the enzyme responsible for this activity. Our data do not address the role of SMb20409 in sugar alcohol metabolism although we consider it unlikely to be involved in oxidation of galactitol.

Table 2.4. Carbon phenotypes of *smoS* mutants

Strain	Characteristic	gly	gtl	tag	gtl/gly	tag/gly
Rm1021	wildtype	+	+	+	+	+
SMc01500	*smoS*::pTH1703	+	-	+	-	+
SMc01500	*smoS in trans*	+	+	+	+	+
SRmD495	*smoK*::Tn*5*-B20	+	-	+	nd	nd
SRmD495	*smoS in trans*	+	+	+	+	+
SRmD523	*smoM*::pKNOCK-Gm	+	+	+	+	+
SRmD615	*SMb20410*::pKNOCK-Tc	+	+	+	+	+

Growth scored as follows; +, like wildtype; -, no growth; nd, not determined. Abbreviations are as follows; gly, glycerol; gtl, galactitol; tag, tagatose.

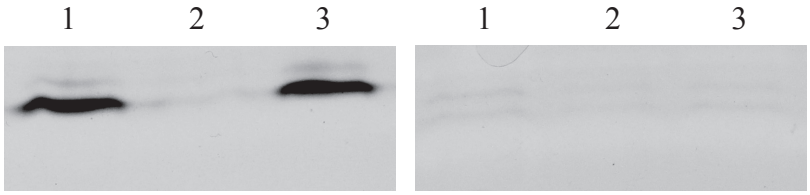


Figure 2.4. Non-denaturing PAGE of galactitol dehydrogenase activity. Extracts from wildtype, Rm1021 (lane 1); *smoS* mutant, SMc01500 (lane 2); and mutant with *smoS in trans*, SMc01500xp*SMc01500* (lane 3) were separated and gels were stained for dehydrogenase activity using an assay reagent containing the substrate listed below each panel

2.4.5 Galactitol competes with mannitol for transport

Based on previously published work, there are five transport systems that may contribute to galactitol uptake. The transporter expression map identified four SBP-encoding genes induced by galactitol; *SMb21647* (*agpA*), *SMb20410*, *SMb21373* (*tagA*), and *SMc01496* (*smoE*) (Mauchline *et al.*, 2006). Additionally, *SMc04251* (*smoM*) is annotated as being a mannitol binding protein.

Mutations of *agpA* have been previously characterized and shown to be involved in the transport of α-galactosides and necessary for growth on raffinose and melibiose (Gage & Long, 1998). It was also found that induction of *agpA* could occur with galactose. Induction by galactitol could be explained by its structural similarity to galactose. Mutagenesis of *SMb20410* or *smoM* had no effect on galactitol utilization (Table 2.4). Therefore, the two most likely candidate galactitol transport systems are *tagABC* and *smoEFGK*, given their induction by galactitol and their locations within operons shown to contribute to galactitol metabolism.

A mutation in *smoK* results in the inability to grow on galactitol as a sole carbon source. However, introduction of a *smoS*-expressing plasmid into a *smoK* mutant strain restores the mutant's ability to utilize galactitol (Table 2.4). Similarly, a galactitol growth deficiency caused by a mutation to *tagC* can be complemented by expression of *tagDE* carried on a plasmid (data in text, see above). This suggests that carbon phenotypes associated with mutations in *smoK* or *tagC* are the result of polar effects on downstream catabolic genes rather than the inability to transport substrate. These data suggest that neither of these transport systems is solely responsible for galactitol transport in *S. meliloti.* However, it is possible that they work in conjunction as low specificity sugar alcohol transporters.

To test this hypothesis, transport assays were conducted using ^{14}C-mannitol as a radiolabelled substrate. Transport rates of labeled mannitol were approximately ten times higher when cells were grown on mannitol as a carbon source rather than glucose (Fig. 2.5A), suggesting that mannitol uptake is inducible in Rm1021. A strain carrying a mutation in *smoK* was not capable of transporting mannitol (data not shown). Mannitol uptake was reduced if ^{14}C-mannitol was competed with a ratio of either 1:1 or 1:5 unlabeled mannitol. The addition of unlabeled galactitol at a 1:1 ratio with the ^{14}C-mannitol did not significantly reduce mannitol uptake rates, whereas initiating the assay with a 5:1 mixture of galactitol to labeled mannitol resulted in a significant decrease in mannitol uptake (Fig. 2.5B). Mannitol transport was essentially unaffected with the addition of either tagatose or glucose (Fig. 2.5B). These data suggest that mannitol and galactitol can compete for the use of the same transporter(s), but that the affinity for galactitol is lower. Whereas tagatose and glucose transport are not dependent upon the mannitol transporter.

2.4.6 Galactitol mutants are equally competitive for nodule occupancy

It has been previously shown that mutants unable to grow using certain sugar alcohols as a sole carbon source are less competitive for nodule occupancy (Ding *et al.*, 2012; Fry *et al.*, 2001; Yost *et al.*, 2006). To determine if galactitol mutants exhibit a competitive phenotype, competition for nodule occupancy assays were performed. Strains RmP233 (*tagA*::*gusA*), SRmD493 (*smoS*::Tn*5*-B20) and SRmD495 (*smoK*::Tn*5*-B20) were competed in approximately equal proportions, as well as in a 10:1 ratio, with wildtype. These mixtures were used to inoculate seedlings, and subsequently, after approximately four weeks of growth, nodules were harvested and the proportion of mutant to wildtype cells was determined.

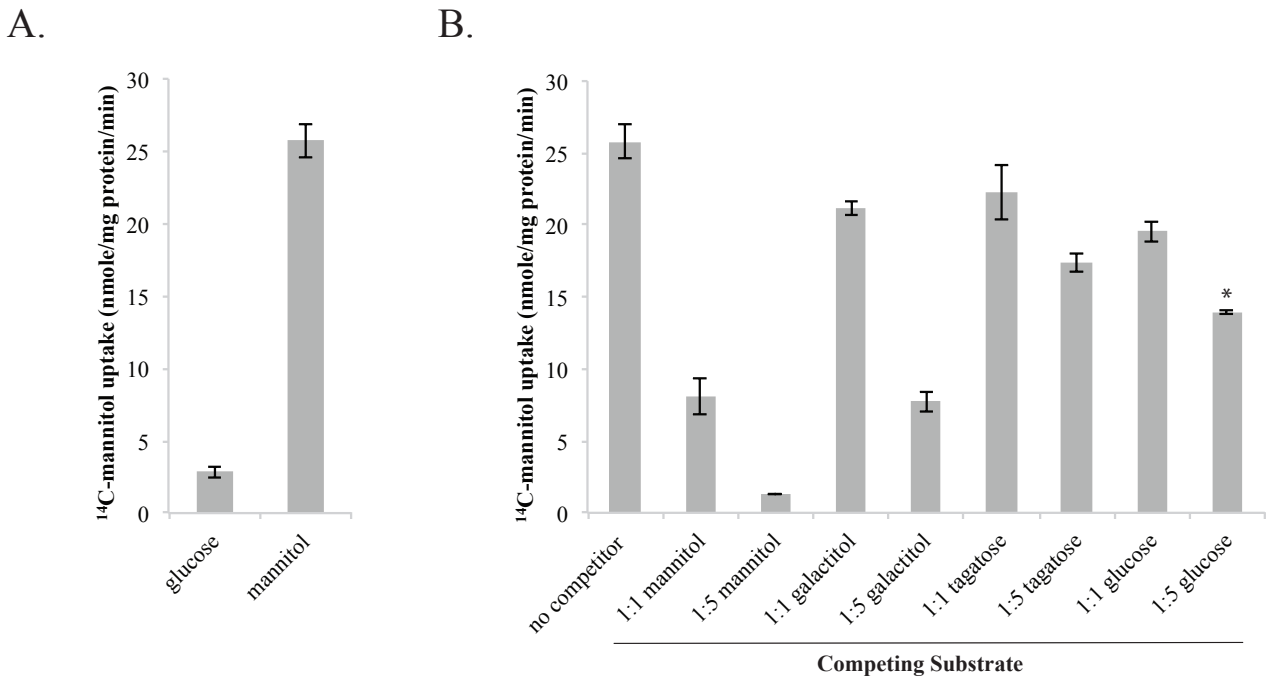


Figure 2.5. Uptake of ^{14}C-mannitol in *S. meliloti*. (A) Transport rates of ^{14}C-mannitol for cells grown on glucose or mannitol as a sole carbon source. (B) Transport rates of ^{14}C-mannitol in competition with unlabelled substrates for cells grown on mannitol as a sole carbon source. 2µM labelled mannitol was competed against either 2µM or 10µM unlabelled substrate. Accumulation of label is measured in nmole/mg protein/min. Data are expressed as the mean ± SD of three independent replicates, * represents data are an average of only two independent replicates.

None of the recovery ratios showed a significant difference from the proportion of mutant cells present in the inoculum suggesting that galactitol mutants are equally as competitive for nodule occupancy as the wildtype cells.

2.4.7 E. coli *GatZ has tagatose-6-phosphate epimerase activity*

The *S. meliloti* epimerase gene *tagE* is annotated as encoding a T1,6bisP aldolase. This designation is largely based on sequence similarity to *gatZ* from *E. coli*, which acts as one component of a *gatYZ* heterodimer responsible for converting T1,6bisP into GAP and DHAP. However, *tagE* is capable of complementing an *A. tumefaciens* T6P epimerase mutant suggesting that it encodes a functional epimerase (Fig. 2). Interestingly, the literature describes *gatZ* as encoding a noncatalytic subunit of the aldolase heterodimer, capable of augmenting the aldolase activity of *gatY* without any activity of its own (Brinkkötter *et al.*, 2002). *E. coli* also contains another gene pair with high similarity to *gatYZ,* termed *kbaYZ* (ketose bisphosphate aldolase), which are involved in galactosamine (GalN or Gam) and N-acetyl-galactosamine (GalNAc or Aga) catabolism (Brinkkötter *et al.*, 2000). Mutations to *gatYZ* can be complemented by overexpression of *kbaYZ* (Lengeler, 1977). Since TagE and AgaZ have high amino acid sequence similarity to GatZ, 49% and 48% respectively, an X-ray crystal structure comparison of TagE (PDBID: 3TXV) and GatZ (PDBID: 2FIQ) was carried out using the jFATCAT_rigid method from the PDB Structure Comparison Tools (Prlić *et al.*, 2010). The program calculated a similarity score of 1076.26 and a RMSD of 1.12 with a *P*-value of 0. A visual representation of both structures shows a similar three-dimensional configuration suggesting that these proteins may have similar activities; leading to the hypothesis that GatZ may have epimerase activity, despite its annotation as a non-catalytic aldolase subunit (Fig. 2.6).

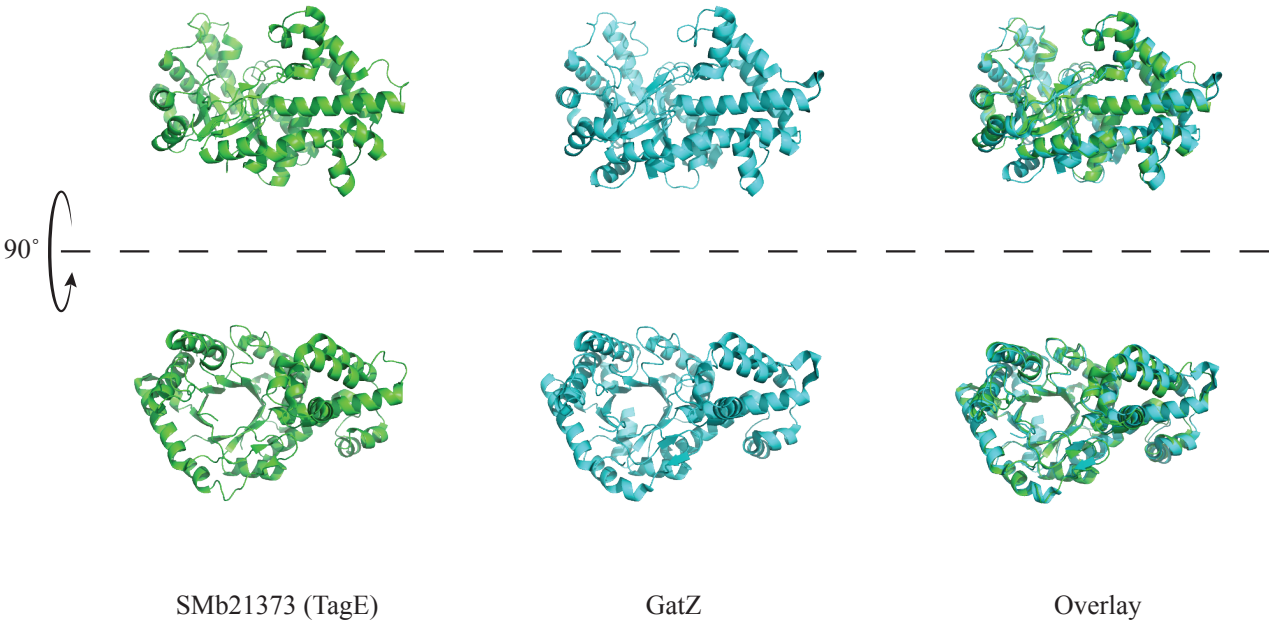


Figure 2.6. Structural comparison of SMb21373 (PDBID: 3TXV) from *S. meliloti* (green) and GatZ (PDBID: 2FIQ) from *E. coli* (blue).

To test this hypothesis, both *gatZ* and *kbaZ* from *E. coli* were separately cloned into pRK7813. These constructs were then conjugated into *A. tumefaciens* epimerase mutant Atu1 and the strains were screened for their ability to grow on galactitol as a sole carbon source. Neither construct permitted growth on galactitol. A sequence comparison of *gatZ* from *E. coli* to *agaZ* from *A. tumefaciens* showed these genes had a 53.48% similarity at the nucleotide level as determined by the Percent Identity Matrix from Clustal Omega. Additionally, *gatZ* has a GC content of 48.9%, while the GC content of *agaZ* is 62.5%. The GC content disparity between the genes suggests that differing codon preferences between the organisms involved could be affecting expression. To address this *gatZ* from *E. coli* was codon optimized for expression in Rhizobiaceae, and was cloned into pRK7813. The construct was conjugated into Atu1 and growth on galactitol as a sole carbon source was evaluated. The results show that codon optimized construct of *gatZ* was as capable as *tagE* at restoring growth of the epimerase mutant Atu1to wild-type levels (Fig. 2.7), suggesting that GatZ can function as an epimerase.

2.4.8 IPR012062 is a family of tagatose-phosphate epimerases

The InterPro family that contains GatZ, AgaZ, and TagE is designated IPR012062. The listed function of the proteins within the family is "D-tagatose-bisphosphate aldolase, non-catalytic subunit GatZ/KbaZ." These three proteins are catalytic and the reaction is an epimerization, not an aldol hydrolysis ((Wichelecki *et al.*, 2015), Fig. 2.3, Fig. 2.7). It was hypothesized that one of two possibilities existed regarding the nature of this family; either the entire family had been misannotated and each member is capable of catalyzing an epimerization reaction, or a small subset of the family had developed a novel catalytic capability.

In order to determine which was true, a phylogenetic analysis of the amino acid sequences in IPR012062 was carried out. At the time of the download from InterPro, the family contained 2266 sequences, and a maximum likelihood tree was constructed from these sequences using RAxML-HPC (Fig. 2.8). It is apparent that the sites representing proteins that have been tested and are known epimerases (highlighted in red) are dispersed across the tree, and not clustered together into a small isolated clade (Fig. 2.8). This pattern is consistent with the hypothesis that epimerase activity is widespread throughout IPR012062, and not an isolated anomaly. We note that the three sequences can be isolated into a large clade containing 1364 sequences, which accounts for approximately 60% of the total sequences. Therefore, we cannot exclude the possibility that the epimerase activity is isolated within this monophyletic group and may not be associated with every member of the family. However, we believe that the data is most consistent with a systemic misannotation and propose that the family be renamed D-tagatose-6-phosphate epimerase TagE.

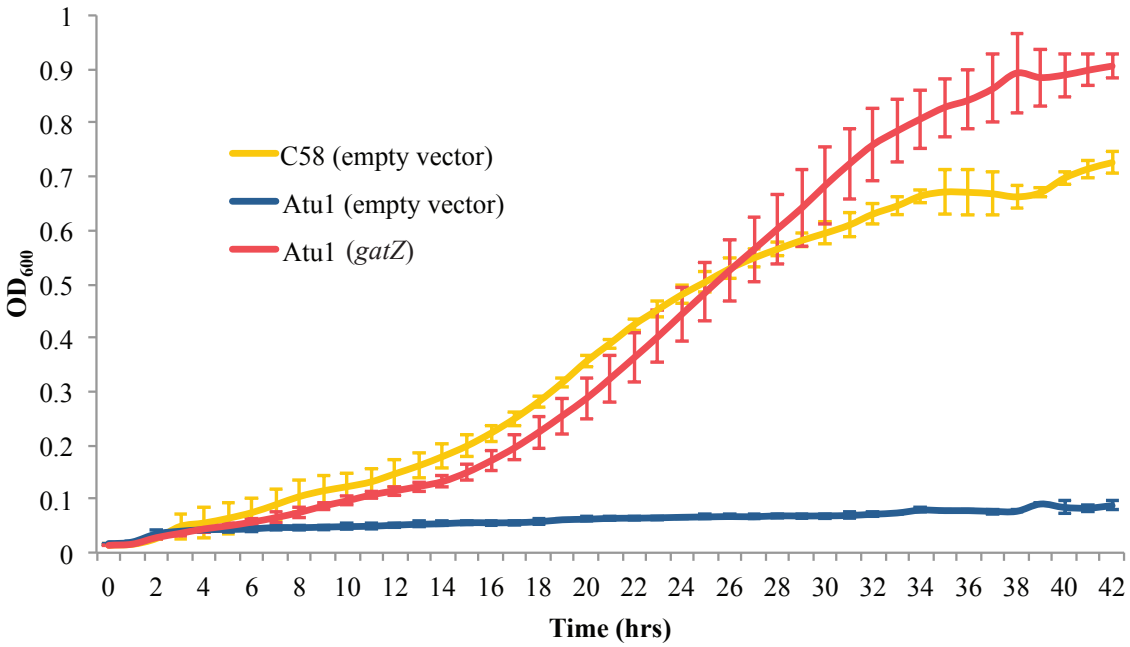


Figure 2.7. Growth curves comparing *A. tumefaciens* strains grown in defined media with galactitol as a sole carbon source. OD_{600} of wildtype, C58 (yellow); epimerase mutant, Atu1 (dark blue); mutant with *gatZ in trans* (red) was measured every hour for 42 hrs. Data are expressed as the mean ± SD of three independent replicates.

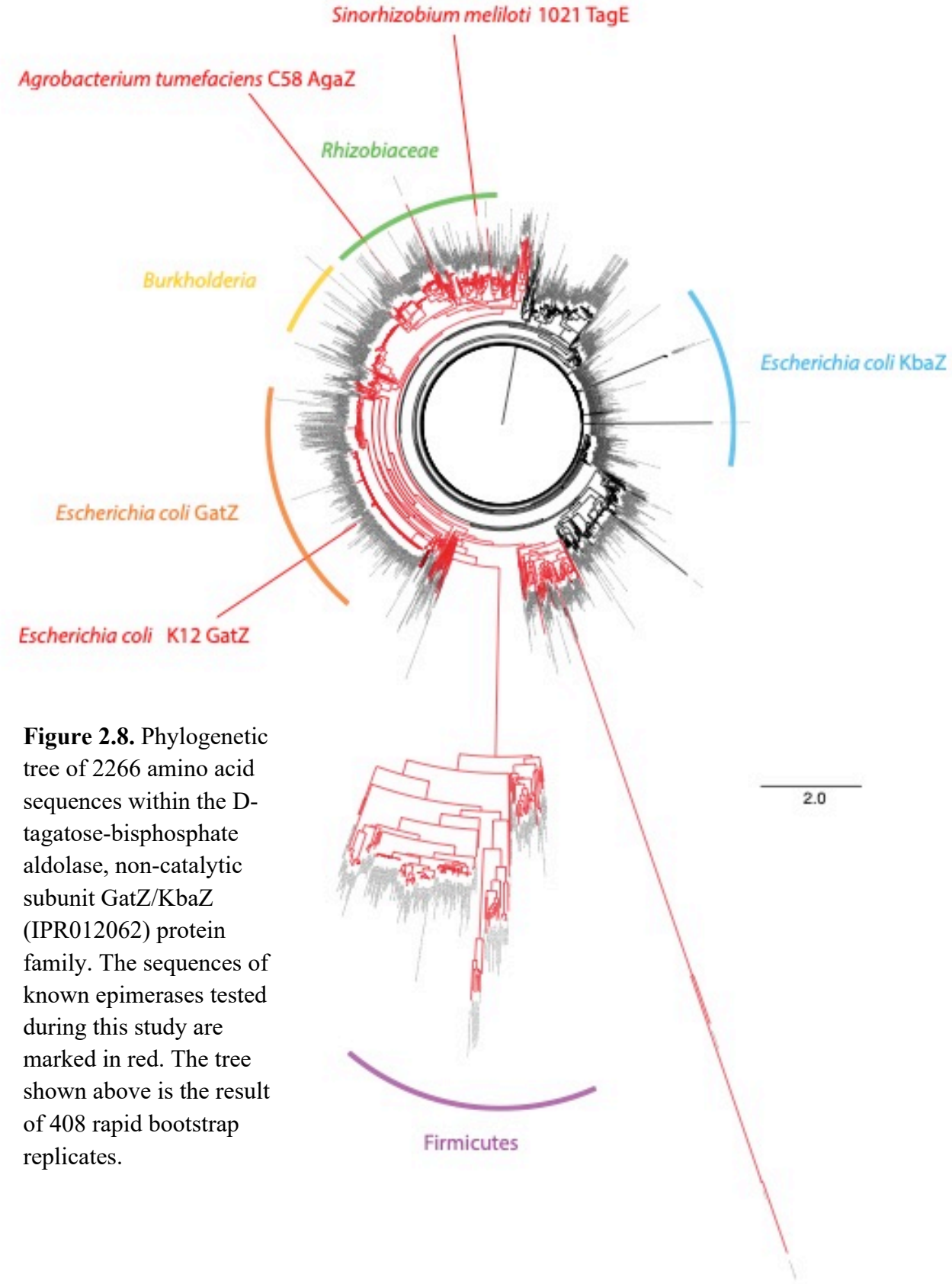


Figure 2.8. Phylogenetic tree of 2266 amino acid sequences within the D-tagatose-bisphosphate aldolase, non-catalytic subunit GatZ/KbaZ (IPR012062) protein family. The sequences of known epimerases tested during this study are marked in red. The tree shown above is the result of 408 rapid bootstrap replicates.

2.5 Discussion

In this work, we have characterized galactitol catabolism in *S. meliloti*. We have shown that the pathway consists of SmoS oxidizing galactitol to tagatose, which is subsequently phosphorylated by TagD (SMb21374), and finally epimerized by TagE (SMb21373) to yield F6P (Fig. 2.9). The transport of galactitol is not dependent on TagABC, and it can weakly compete for the ABC transporter encoded by *smoEFGK*. A mutation in *smoM* did not greatly affect growth on sugar alcohols or sugar alcohol derivatives including galactitol, tagatose, or glycerol. This finding is consistent with other results in the literature that have described SmoM as a tripartite ATP-independent periplasmic (TRAP) transporter with specificity for α-keto acids in *R. sphaeroides* (Gonin *et al.*, 2007; Thomas *et al.*, 2006). This suggests that SmoM does not contribute to sugar alcohol uptake, although it is possible that the protein may contribute in an additive manner in conjunction with the above-mentioned transport systems.

Although tagatose is considered to be a rare sugar, and is not found abundantly in nature, phosphorylated tagatose intermediates are commonly generated within bacterial cells. This occurs predominantly in the catabolism of galactitol (Nobelmann & Lengeler, 1996), as well as galactosamine and GalNAc (Brinkkötter *et al.*, 2000; Leyn *et al.*, 2012). The widespread occurrence of phosphorylated tagatose intermediaries is best exemplified by the number of entries in the IPR012062 protein family, represented by GatZ and KbaZ, which are the most extensively characterized members (Brinkkötter *et al.*, 2002; Brinkkötter *et al.*, 2000; Nobelmann & Lengeler, 1996).

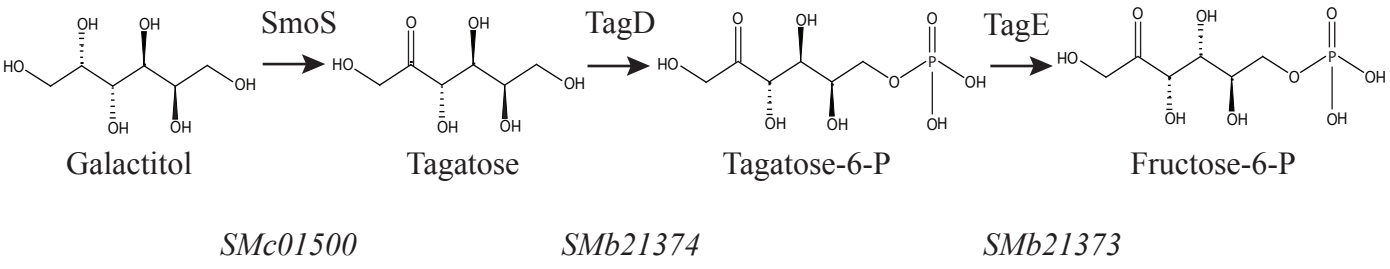


Figure 2.9. Proposed pathway of galactitol utilization in *S. meliloti*.

Initial classification of *gatZ* was difficult due to the lack of available sequences for comparison (Nobelmann & Lengeler, 1995; Nobelmann & Lengeler, 1996). Further complicating matters was the subsequent proposal of an GalNAc catabolic pathway involving a T6P kinase encoded by *agaZ*, a gene highly similar to *gatZ* at the nucleotide level (Reizer *et al.*, 1996). This suggestion was later shown to be false when PfkA was identified as the sole enzyme responsible for T6P phosphorylation in *E. coli* (Brinkkötter *et al.*, 2000). Although this has been mostly corrected in the databases, there are still entries that refer to this protein as a kinase (PDBID: 3TXV, Fig. 2.8). It was eventually concluded that GatZ functioned as a non-catalytic subunit of a GatYZ heterodimer based on the observation that GatZ could enhance the T1,6bisP aldolase activity of GatY, but it did not display aldolase activity of its own (Brinkkötter *et al.*, 2002).

This has led to a great deal of confusion as well as misannotation of genes, which further complicates genome wide studies such as metabolic reconstructions (diCenzo *et al.*, 2016; Leyn *et al.*, 2012). The characterization of the GalNAc pathway in *Shewanella* using a comparative genomic approach was able to predict and confirm a GalNAc kinase, deactylase, as well as isomerase activity. The interpretation, based on comparative genomic analysis, showed that AgaZ was mostly present only in the Enterobacteriales, but missing in other Proteobacteria. This evidence was used to forward the hypothesis that AgaZ must play another essential role in the catabolism of GalNAc. In the published model it was given the role of a T6P kinase (Leyn *et al.*, 2012).

Through the use phylogeny, as well as direct protein comparisons we have shown that TagE, GatZ, and AgaZ have the ability to provide *in vivo* tagatose phosphate epimerase activity. That these proteins are phylogenetically related strongly suggests that this activity may be

present in all of the members of this family. This supports the hypothesis previously put forward that AgaZ has a more central role (Leyn *et al.*, 2012), however our data would suggest that the catabolism of GalNAc utilizes an epimerase to convert T6P to F6P in *Shewanella*. Our data does not address whether the epimerase activity is in place of, or in addition to its role as a non-catalytic subunit for T1,6bisP aldolase.

The *S. meliloti* metabolic model predicted galactosamine catabolism was carried out by the genes *SMb21216-21* (diCenzo *et al.*, 2016). This locus includes an ABC transporter that is induced by galactosamine (Mauchline *et al.*, 2006), as well as a carbohydrate kinase and a deaminase/isomerase in the AgaS superfamily. The latter is likely to convert galactosamine-6-phosphate (GalN6P) to T6P. It was hypothesized that the likely catabolic pathway could use *tagE* (*SMb21373*) and *tagD* (*SMb21374*) to convert galactosamine to a central carbon pathway intermediate (diCenzo *et al.*, 2016). However in our hands, *S. meliloti* galactitol mutants grow as well as wildtype using D-galactosamine as a sole carbon source, suggesting that TagE is unlikely to be involved in the metabolism of this substrate (data not shown). It could be that *tagE* expression is not induced by T6P that is produced from GalN6P, and that another aldolase/epimerase is encoded in the genome and expressed if the strain is grown on D-galactosamine. Alternately, it may be that *S. meliloti* utilizes a novel pathway for the catabolism of D-galactosamine. It is noteworthy that BLAST searches using TagE as a query do not return other closely related proteins from the Rm1021 genome.

The lack of a competition for nodule occupancy phenotype exhibited by galactitol mutants suggests that galactitol is not an important determinant for nodule occupancy in the rhizosphere for this symbiotic model system. Additionally, since these mutants are deficient for growth on sorbitol, mannitol (Jacob *et al.*, 2008), and tagatose, it is unlikely that the ability of *S.*

meliloti to catabolize these substrates is important for nodule occupancy. This is consistent with previous results, which have shown that *Rhizobium leguminosarum* sorbitol catabolic mutants are unaffected with respect to competition for nodule occupancy (Oresnik *et al.*, 1998).

Evidence presented in this study elucidates galactitol metabolism in *S. meliloti*. Also demonstrated is the versatility of *S. meliloti* as a model system for studies in a range of fields including symbiosis, carbon metabolism, and genome evolution. The results also caution against the overreliance on computational predictions without consideration of empirically determined phenotypic data. Our inability to confidently predict the pathway for galactosamine catabolism highlights the need for functional studies to allow us to better understand the metabolic capacity of *S. meliloti*, as well as the direct role played by carbon metabolism during growth in the rhizosphere and as a determinant for nodule occupancy.

Chapter 3:

Characterization of sorbitol dehydrogenase SmoS from *Sinorhizobium meliloti* 1021

This work was carried out by MacLean Kohlmeier in collaboration with Ben Bailey-Elkin. BBE supervised the purification of the protein, the growth of the crystals, and the collection of the X-ray diffraction data, as well as independently solving the crystal structures and contributed to the writing of this chapter.

3.1 Abstract

Sinorhizobium meliloti 1021 is a Gram-negative α proteobacterium with a robust capacity for carbohydrate metabolism. The enzymes that facilitate these reactions assist in the survival of the bacterium across a range of environmental niches, and they may also be suitable for use in industrial processes. SmoS is a dehydrogenase that catalyzes the oxidation of commonly occurring sugar alcohols sorbitol and galactitol into fructose and tagatose respectively using NAD^+ as a cofactor. The main objective of this study was to evaluate SmoS using biochemical techniques. The nucleotide sequence was codon optimized for heterologous expression in *E. coli* BL21 (DE3) GOLD cells, the protein was subsequently overexpressed and purified. Size exclusion chromatography and X-ray diffraction experiments suggest that SmoS is a tetrameric peptide. SmoS was crystallized to 2.1 Å in the absence of substrate and 2.0 Å in complex with sorbitol. SmoS was characterized kinetically and shown to have a preference for sorbitol despite a higher affinity for galactitol. Computational ligand docking experiments suggest that galactitol oxidation proceeds slowly because tagatose binds the protein in a more energetically favorable complex than fructose, and is retained in the active site for a longer time frame following oxidation which reduces the rate of the reaction. These results supplement the inventory of biomolecules with the potential for industrial applications and enhance our understanding of metabolism in the model organism *S. meliloti*.

3.2 Introduction

Sugar alcohols, also called polyols, are carbohydrate compounds that can be formed by the reduction of an aldo or keto sugar. The first polyols were identified from honeydew, a substance secreted by aphids as they feed on plant sap (Bieleski, 1982). The most commonly encountered sugar alcohols in nature are sorbitol, mannitol, and galactitol (also known as dulcitol or melampyrite) (Williamson *et al.*, 2002). These linear, six carbon polyols were named for the higher plants from which they originated; sorbitol from *Sorbus aucuparia*, mannitol from *Fraxinus ornis* or manna ash, and galactitol from *Melampyrum nemorosum* (Bieleski, 1982).

Sugar alcohols and their derivatives have a variety of applications. Sorbitol is commonly included in food products for sweetness, texture, and preservation, and can be present in pharmaceuticals (Rapaille *et al.*, 2003; Silveira & Jonas, 2002). D-tagatose, a product of galactitol oxidation, is classified as a rare sugar and is being considered as a treatment for diabetes due to its insulin independent metabolism in humans and potential to lower blood glucose levels (Ensor *et al.*, 2015; Espinosa & Fogelfeld, 2010; Lu *et al.*, 2008). The concentrations of sugar alcohols in plant tissue are typically too low for chemical extractions to generate sufficient yields, therefore polyols are often synthesized for commercial use via catalytic hydrogenation of more readily available sugars (Rapaille *et al.*, 2003). However, biological enzymes can serve as biocatalysts for the generation of sugar alcohols and related molecules at an industrial scale. Some advantages to biocatalysts include high product selectivity and low environmental or physiological toxicity (Chapman *et al.*, 2018). As an example, galactitol dehydrogenase has been immobilized on gold electrodes for use in electrochemical reactors with the goal of generating precursor molecules for pharmaceuticals via reactions that regenerate reduced cofactors (Gajdzik *et al.*, 2010; Gajdzik *et al.*, 2011; Kornberger *et al.*, 2009).

Enzymes of microbial origin are ideal with respect to industrial applications as they can be produced in an easy, cost effective, and consistent manner (Raveendran *et al.*, 2018). Carbohydrate metabolism in plant associated soil bacteria has been studied in great detail due to the involvement of carbon utilization in symbiotic establishment and efficiency (Geddes & Oresnik, 2014; Udvardi & Poole, 2013). Transport genes responsible for the uptake of sorbitol, mannitol, and galactitol are induced in the rhizosphere (Ramachandran *et al.*, 2011).

In bacteria, the initial step of sugar alcohol metabolism is often oxidation into a keto sugar, followed by phosphorylation (Mortlock, 1984). The root-nodulating bacterium *Sinorhizobium meliloti* has been shown to produce a D-sorbitol specific dehydrogenase (SDH), which uses NAD^+ as a cofactor (Martínez De Drets & Arias, 1970). A mutant lacking fructose kinase activity was unable to grow using sorbitol as a sole carbon source, suggesting that fructose is the product of sorbitol oxidation in *S. meliloti* (Gardiol *et al.*, 1980). A mutation in a gene annotated as a putative sorbitol dehydrogenase *smoS* resulted in a strain with the inability to grow on several sugar alcohols, including sorbitol (Jacob *et al.*, 2008), suggesting that *smoS* encodes the SDH protein. *smoS* was first identified as encoding a SDH in *Rhodobacter sphaeroides,* in which it was described as one gene in a novel polyol metabolic operon, as well as a member of the short-chain dehydrogenase/reductase (SDR) family (Stein *et al.*, 1997).

SDR proteins are typically about 250 amino acids in length and despite having low sequence identity at 20-30%, members of this family share a similar overall three-dimensional structure (Persson & Kallberg, 2013). Currently there are over 230,000 members of the SDR family in the UniProt database, and a recently devised nomenclature system based on Hidden Markov Models placed SmoS within the SDR196C subfamily (Persson *et al.*, 2009; Sola-Carvajal *et al.*, 2012). *Rs*SDH is dependent on NAD^+ as a cofactor and has activity on sorbitol

and galactitol (Fig. 3.1) (Stein *et al.*, 1997). Structural studies on the protein in the absence of bound substrate were some of the first structures of a bacterial SDH in the SDR family (Philippsen *et al.*, 2005). The purpose of this study is to characterize SmoS from *S. meliloti* with respect to its structure, as well as kinetic and physical properties.

3.3 Materials and methods

3.3.1 Bacterial strains and culture conditions

E. coli BL21 (DE3) GOLD cells were grown on Luria Bertani (LB) medium (Cold Spring Harbor Protocols, 2006) at 37°C; when necessary, kanamycin was added to a final concentration of 10 μg/mL in liquid media.

3.3.2 Overexpression and purification of SmoS

S. meliloti smoS is a 774 bp gene with a GC content of 64.5%, the overall GC content of *E. coli* K-12 is 50.8% (Riley *et al.*, 2006). To accommodate this disparity, the *smoS* nucleotide sequence was codon optimized for expression in *E. coli*. Translation of *smoS* is predicted to generate a 257 amino acid sequence with a molecular weight of 27.2 kDa (Gasteiger E., 2005). *smoS* was cloned into overexpression vector pET-28a as a *Bam*HI-*Hind*III fragment (GenScript, Piscataway, NJ, USA) and this construct was transformed into competent *E. coli* BL21 (DE3) GOLD cells.

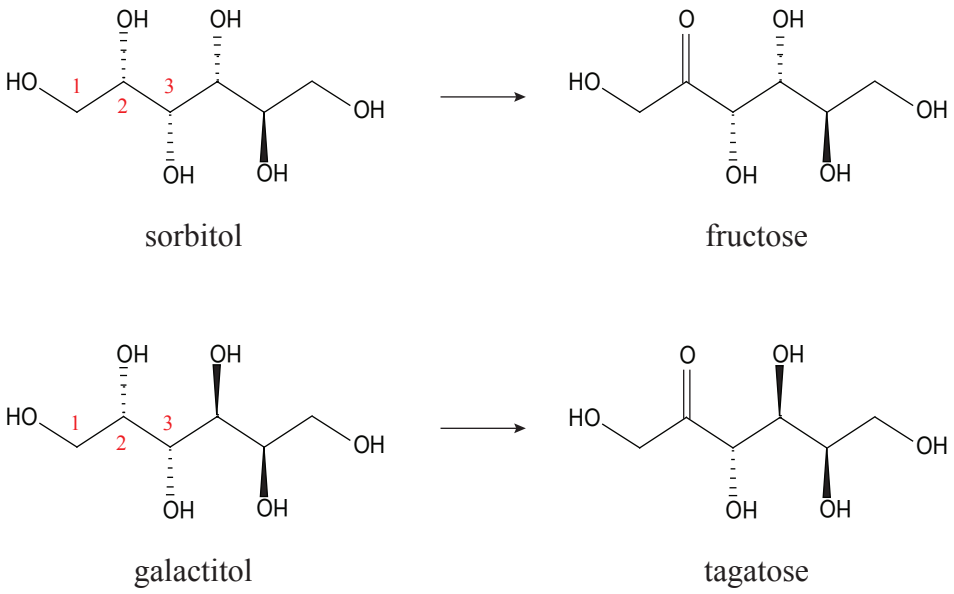


Figure 3.1. Enzymatic reactions catalyzed by SmoS. Sorbitol or galactitol are oxidized at carbon 2, using NAD^+ as a cofactor, producing fructose or tagatose respectively, as well as NADH. Both sugar alcohols are viable substrates for SmoS due to the identical orientation of hydroxyl groups about carbons 1, 2, and 3.

Cultures were grown in 1 L volumes of LB medium at 37°C to an OD_{600} of ~0.6. Induction with 1 mM isopropyl-β-D-galactopyranoside (IPTG) preceded growth overnight, shaking, at 16°C. Cells were pelleted by centrifugation at 10000 rpm for 10 min and stored at -80°C. Pellets were resuspended in 30 mL cold lysis buffer consisting of 50 mM Tris pH 8.0, 300 mM NaCl, 2 mM dithiothreitol (DTT), 10 mM imidazole, and lysed by French Press. Cell debris were removed from extracts by centrifugation at 12000 rpm for 1 hour at 4°C. The cell free lysate was applied to a nickel nitrilotriacetic acid (Ni-NTA) column, which was washed with 10 column volumes of lysis buffer and followed by a second wash with 10 column volumes of lysis buffer with 25 mM imidazole. Final elution was prompted by washing with 3 column volumes of buffer with 500 mM imidazole. Eluted protein was dialyzed against 20 mM HEPES pH 7.5, 150 mM NaCl, 10% (v/v) glycerol, and further purified by gel filtration through a Superdex 75 gel filtration column.

3.3.3 SmoS crystallization

Purified SmoS was concentrated to 10 mg/mL and screened by sitting drop vapour diffusion using a Gryphon (Art Robbins Instruments, Sunnyvale, CA, USA) robotic drop setter. Screening was performed using 600 nL drops containing SmoS and crystallization solution at a 1:1 ratio, equilibrated against 50 µL of reservoir solution. Initial crystallization hits were identified in 100 mM HEPES pH 7.4, 50 mM sodium acetate and 20% PEG 3000, and further optimized by hanging-drop vapour diffusion using 48-well VDX plates. Crystals of apo-SmoS were grown in 100 mM HEPES pH 7.4, 50 mM sodium acetate and 18% PEG 3000 and crystals of the SmoS-sorbitol complex were grown under the same conditions supplemented with 20% sorbitol, galactitol, tagatose, or fructose. Crystallization with galactitol and tagatose was not

pursued due to poor solubility or lack of availability of these respective substrates. Crystals in which sorbitol or fructose were included in the reservoir solution were morphologically indistinguishable from the native crystals.

3.3.4 X-ray data collection and structure solution

X-ray data for individual SmoS crystals were collected at 100K on a Rigaku MicroMax 007-HF equipped with a RAXIS IV++ detector. X-ray diffraction images were integrated and scaled using XDS (Kabsch, 2010), and merged using Aimless (Evans, 2011). Initial phase estimates for apo-SmoS were determined by molecular replacement within Phaser using the deposited structure of *S. meliloti* SmoS (PDB ID: 4E6P) as a search model, and phase estimates for the SmoS-sbt complex were determined using the refined apo-SmoS structure. Structure refinement and model building were performed using REFMAC (Murshudov *et al.*, 1997) and Coot (Emsley *et al.*, 2010), respectively within the ccp4i2 software package. All structure figures were generated using PyMOL (Schrodinger, 2015). The coordinates and structure factors for the apo SmoS and SmoS-sbt structures have been deposited to the Protein Data Bank under PDB ID 6PEI and 6PEJ, respectively.

3.3.5 Enzyme assays

Spectrophotometric dehydrogenase assays were conducted by measuring the reduction of NAD^+ at 340 nm for 60 seconds. Reaction buffer consisted of 20 mM CAPS pH 11, 1.5 mM NAD^+, and increasing concentrations of sorbitol or galactitol, in a total volume of 1 mL. 1µg SmoS was added per reaction. The optimum pH for enzyme activity was determined by measuring dehydrogenase activity using 200 mM MES, MOPS, TRIS, or CAPS to buffer the

reaction mixtures over their appropriate pH ranges. All pH-profiling reactions were initiated with 10 mM sorbitol. Additionally, native gel dehydrogenase assays were performed as previously described (Pickering & Oresnik, 2008). Following elution from the S75 column, fractions were separated by nondenaturing polyacrylamide gel electrophoresis; subsequently the gels were stained for dehydrogenase activity with an assay reagent containing Tris pH 8.0, phenazine methosulfate, nitroblue tetrazolium, NAD^+, and sorbitol.

3.3.6 Ligand docking analysis

D-fructose and D-tagatose model files, in SDF file format, were submitted to the Ligand Docking Protocol on the ROSIE server, found at http://rosie.rosettacommons.org, along with the apo SmoS monomer structure in PDB file format. The ligand SDF files were downloaded from Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) at https://www.rcsb.org. These ligand models were manipulated to within 5 Å of the SmoS substrate binding pocket using PyMOL (Schrodinger, 2015) prior to submission to add coordinate data to the files. The top ten predicted models with the lowest interface delta scores were collected and the distribution of these data sets was compared with box and whisker plots. A paired *t* test was performed on the score arrays, a *P* value of less that 0.01 was considered significant.

3.4 Results

3.4.1 Structural characterization of Sm*SmoS*

Size exclusion chromatography of purified SmoS showed two distinct peaks at elution volumes of 49.22 mL and 55.35 mL (Fig. 3.2A), with the most prominent peak at ~55 mL.

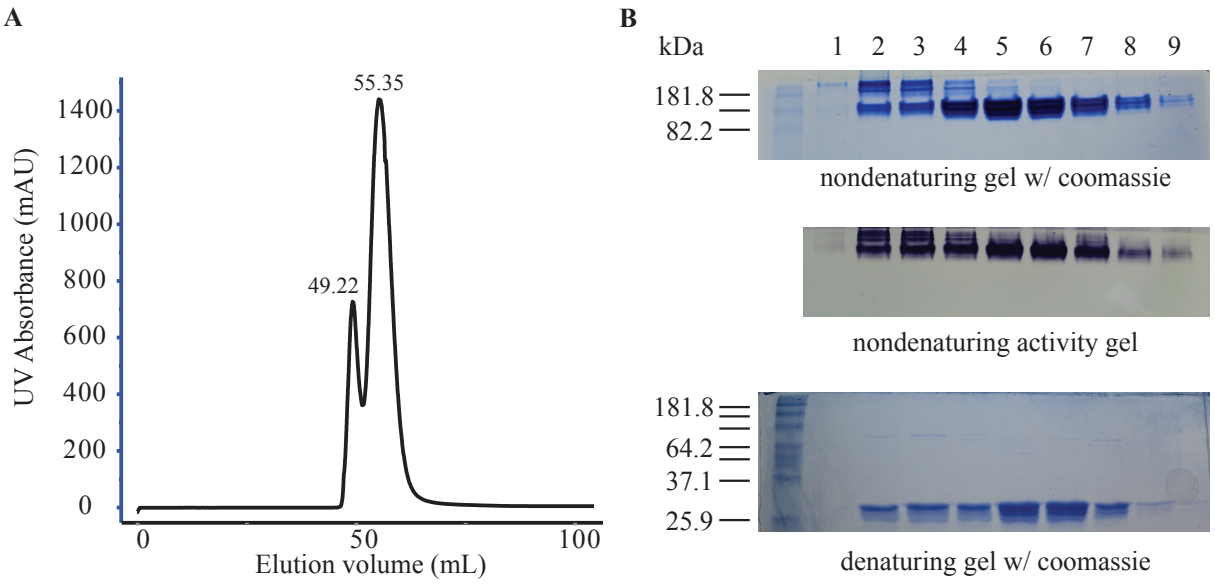


Figure 3.2. Size exclusion chromatography of purified SmoS from a Superdex 75 gel filtration column and analysis by polyacrylamide gel electrophoresis (PAGE). (A) UV trace of elutions from the S75 column displaying two peaks at approximately 49 mL and 55 mL. (B) Elutions separated by nondenaturing PAGE stained with coomassie blue (top), elutions separated by nondenaturing PAGE stained for sorbitol dehydrogenase activity (middle), and elutions separated denaturing PAGE and stained with coomassie blue (bottom).

To corroborate these results, the column fractions were separated by nondenaturing polyacrylamide gel electrophoresis and stained with Coomassie Brilliant Blue. Fractions 2-4 showed two distinct bands, while fractions 5-9 contain a single band, which mimics the migration distance of the lower band from fractions 2-4 (Fig. 3.2B). Both protein bands are capable of sorbitol oxidation when the gel is stained for dehydrogenase activity, and resolve to a molecular weight of 27 kDa when SDS is included in the gel matrix (Fig. 3.2B), suggesting that both bands observed are due to the presence of SmoS.

To further characterize SmoS, the enzyme was crystallized and determined to a resolution of 2.1Å (Fig. 3.3; Table 3.1). Consistent with this observation, SmoS crystallized as a tetramer, with four copies in the asymmetric unit arranged as a dimer of dimers, similarly to a previously determined structure of a *Bradyrhizobium japonicum* D-sorbitol dehydrogenase (Fig. 3.3A) (Fredslund *et al.*, 2016). These results are consistent with SmoS being present in two distinct conformations in solution, with the majority being tetrameric.

The SmoS monomer adopts a structural fold similar to other previously determined Zn-independent SDR enzymes, comprising of an NAD-binding Rossman fold centralized around a core 7-stranded parallel β-sheet, and an extended α-helical clamp-like lobe formed by helices α7 and α8 involved in substrate binding (Philippsen *et al.*, 2005) (Fig. 3.3B). A DALI search (Holm & Laakso, 2016) to identify structural homologues of SmoS identified a previously determined *Sm*SmoS structure (deposited by the New York Structural Genomic Consortium), and a *R. sphaeroides* sorbitol dehydrogenase (81% sequence amino acid identity), which aligned to *Sm*SmoS with an RMSD of 0.7 Å over 256 C_α atoms, and adopted a nearly identical structural fold (Philippsen *et al.*, 2005).

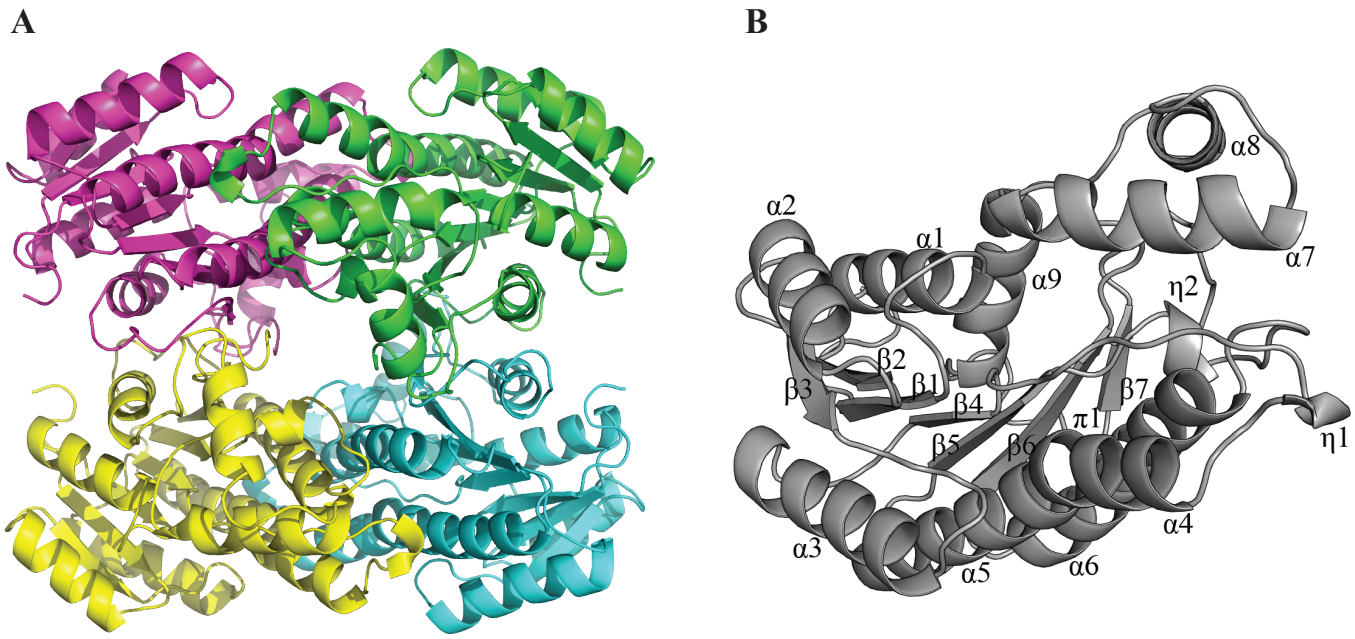


Figure 3.3. Crystal structure of SmoS from *S. meliloti* 1021. (A) Cartoon representation of the SmoS quaternary structure. SmoS forms a homotetramer; the individual monomers are colored magenta, green, blue, and yellow. (B) Cartoon representation of the SmoS monomer (grey). Secondary structure elements are labeled numerically (α, α-helix; β, β strand; π, π helix; η, 3_{10} helix).

Table 3.1. Crystallization and refinement statistics for the SmoS and SmoS-sbt structures

Crystal	SmoS	SmoS-sbt
X-ray source	Rigaku MicroMax-007HF	Rigaku MicroMax-007HF
Crystal geometry		
Space group	P 1 2_1 1	P 1 2_1 1
Unit cell (Å)	*a*=83.30 *b*=88.30 *c*=87.32; α=90.00° β=117.39° γ=90.00°	*a*=83.30 *b*=88.30 *c*=87.32; α=90.00° β=117.39° γ=90.00°
Crystallographic data		
Wavelength (Å)	1.5419	1.5419
Resolution range (Å)	43.70-2.10 (2.15-2.10)*	39.60-2.0 (2.04-2.0)
Total observations	223880 (15515)	224791 (13489)
Unique reflections	65492 (4602)	74395 (4478)
Multiplicity	3.4 (3.4)	3.0 (3.0)
Completeness (%)	99.8 (99.9)	98.0 (99.7)
R_{merge}	0.117 (0.586)	0.139 (0.522)
CC1/2	0.99 (0.71)	0.98 (0.82)
I/σI	8.7 (2.2)	5.4 (2.0)
Wilson B-factor (Å^2)	19.11	20.80
Refinement statistics		
Reflections in test set	3314	3581
Protein atoms	7608	7588
Solvent molecules	737	914
R_{work}/R_{free}	0.192 / 0.252	0.194 / 0.249
RMSDs		
Bond lengths/angles (Å/°)	0.0081 / 1.464	0.0146 / 1.927
Ramachandran plot		
Favored/allowed/disallowed (%)	97.44 / 2.17 / 0.39	97.15 / 2.26 / 0.59
Average B factor (Å^2)		
Macromolecules	25.09	20.64
Solvent	29.46	27.66

*Values in parentheses refer to the highest resolution shell.

In an attempt to uncover the residues involved in substrate binding, SmoS was also crystallized in the presence of sorbitol and a structure determined to 2.0 Å (Table 3.1). Consistent with other described Zn-independent SDR enzymes, conserved active site residues Tyr153, Lys157, Ser140 and Asn111 form the active site (Fig. 3.4A). Residue Asn111 resides on a π-bulge motif formed by an atypical backbone hydrogen bond disrupting helix α4. This deformation allows the backbone carbonyl group of Asn111 to form a hydrogen bond with a water molecule likely to be involved in the formation of a proton relay system similar to what has been described for the *Comamonas testosterone* hydroxysteroid dehydrogenase (Filling *et al.*, 2002; Jornvall *et al.*, 1981). Clear electron density representing sorbitol was visible near the active site of each of the four monomers in the asymmetric unit, with sorbitol coordinated near the active site through a hydrogen-bonding network mediated by SmoS residues Gln141, Glu147, Gly184 and His190 (Fig. 3.4A and B). A comparison of the apo and sorbitol-bound forms of SmoS reveals a slight change in the position of the clamp domain formed by helices α7 and α8, which moves inward during sorbitol binding and allows for the satisfaction of a hydrogen bond between His190 and sorbitol OH1 (Fig. 3.4C).

Interestingly, while clear density for sorbitol was observed in all *Sm*SmoS monomers, the substrate does not appear to be positioned appropriately within in the active site to permit NAD^+-mediated oxidation at C2. In order for the reaction mechanism to proceed as described, the sorbitol C2 hydroxyl group would need to be positioned within hydrogen bonding distance from Tyr153, to allow for Tyr153-mediated proton abstraction and subsequent oxidation of C2 *via* the nicotinamide moiety of NAD^+. In the SmoS-sbt structure, the C2 hydroxyl group is situated ~5.9 Å away from Tyr153, and points away from the active site residue in an arrangement that would not permit the conversion of sorbitol to fructose.

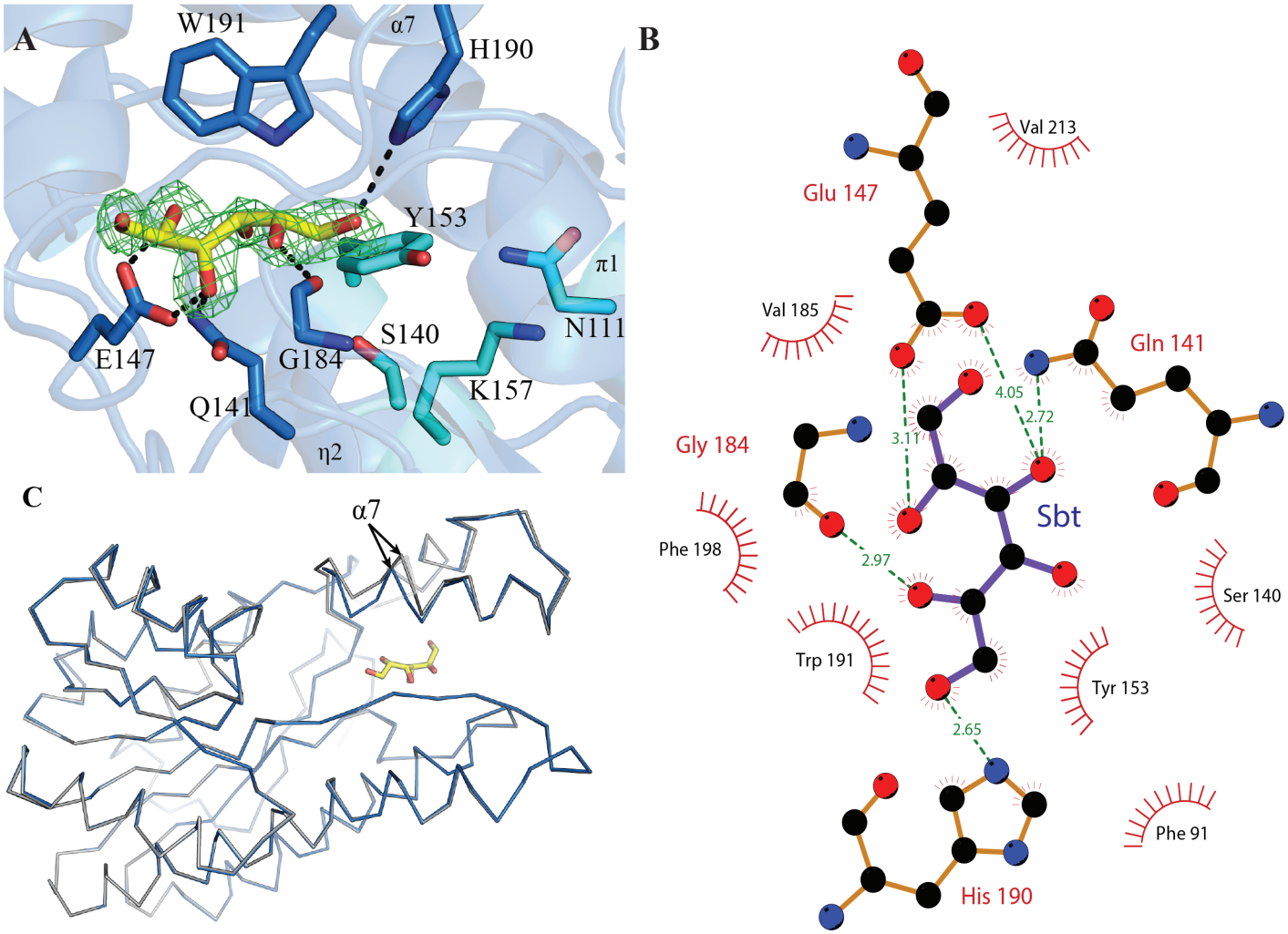


Figure 3.4. Crystal structure of the SmoS-sbt complex. (A) Close up on the active site of sorbitol-bound SmoS. Catalytic residues are shown as blue sticks, and residues involved in the coordination of sorbitol are shown as a cyan sticks. Sorbitol is shown as yellow sticks, surrounded by an mF_o-DF_c omit map generated using phenix.polder ((Liebschner *et al.*, 2017); green mesh) contoured to 3.0σ. (B) Two-dimensional representation of the H-bonding network observed in the SmoS-sbt complex. Carbon atoms are black, oxygen atoms are red, nitrogen atoms are blue, H-bonds are shown as green dashed lines with corresponding bond lengths (Å). Figure was generated using LigPlot (Wallace *et al.*, 1995). (C) Superposition of apo SmoS (grey), and SmoS-sbt (blue) depicted in ribbon diagrams with the movement of helix α7 indicated by arrows.

3.4.2 SmoS has a high pH optimum and a preference for sorbitol

It has been reported that functionally related enzymes to SmoS have optimum activity at alkaline pH levels (Jagtap *et al.*, 2014; Lee *et al.*, 2003). To investigate the pH preference of *S. meliloti* SmoS, sorbitol dehydrogenase assays were conducted across a pH gradient facilitated by several solutions of differed buffering capacities. 1 μg of SmoS was added to the assay mixture along with 10 mM sorbitol and 1.5 mM NAD^+, the buffers included MES, MOPS, TRIS, and CAPS, each at a concentration of 20 mM, which allowed for a pH gradient spanning pH 5.5-12.5. An optimum enzyme activity of 57.8 mM/min/mg was observed at pH 11; activities recorded across the gradient are reported relative to this value (Fig. 3.5). Fifty percent of this activity was found at pH 9.5. All subsequent activity assays were conducted in a solution buffered with 20 mM CAPS pH 11. This result is consistent with observations made in *R. sphaeroides* (Schauder *et al.*, 1995).

Despite the previous inability to detect galactitol dehydrogenase activity (Martínez De Drets & Arias, 1970), recent work has shown that *S. meliloti* is capable of galactitol oxidation and that SmoS is responsible for this activity (Kohlmeier *et al.*, 2019). The ability of SmoS to oxidize sorbitol and galactitol is likely due to the stereochemistry of the functional groups about carbons 1, 2, and 3, which are identical for both substrates (Fig. 3.1). To determine the substrate preference of the enzyme, reaction rates were determined by measurement of NADH accumulation over time in a spectrophotometer at 340 nm. Saturation curves for sorbitol and galactitol dehydrogenase activities were generated along with double reciprocal plots facilitating the determination of Michaelis-Menten reaction constants (Fig. 3.6).

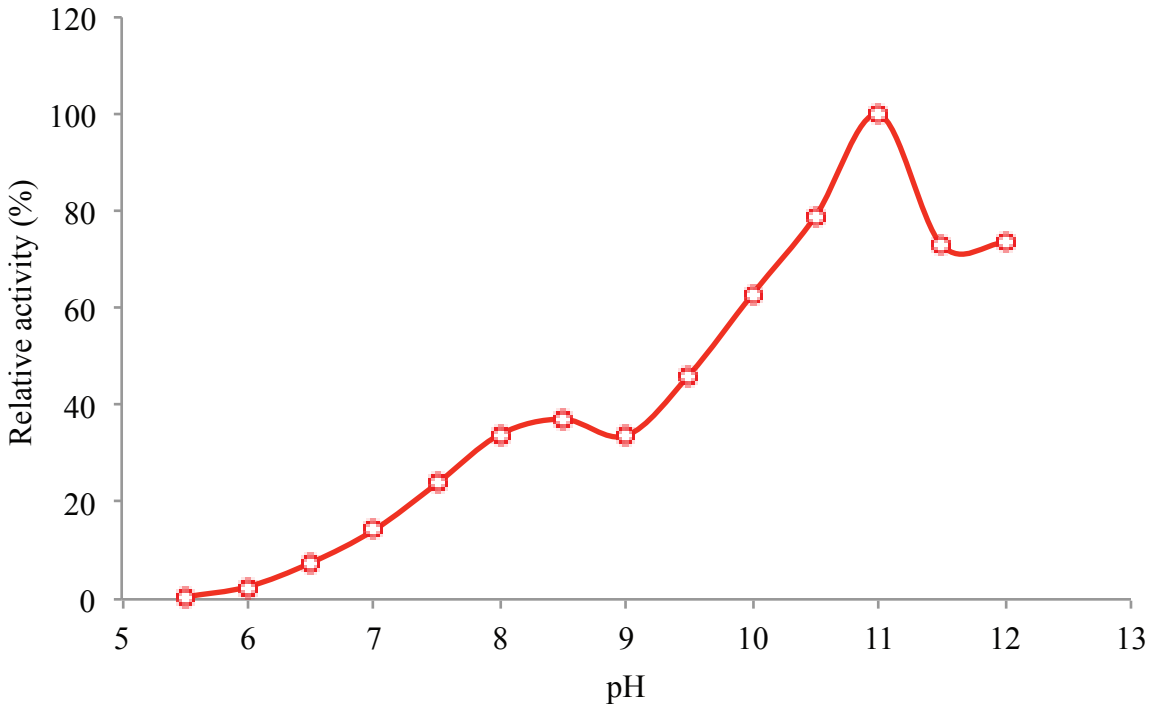


Figure 3.5. Effect of pH on SmoS dehydrogenase activity. Reactions were carried out with 10 mM sorbitol using 200 mM MES, MOPS, TRIS, or CAPS buffers over their appropriate pH ranges. Activity at the optimal pH was defined as 100%.

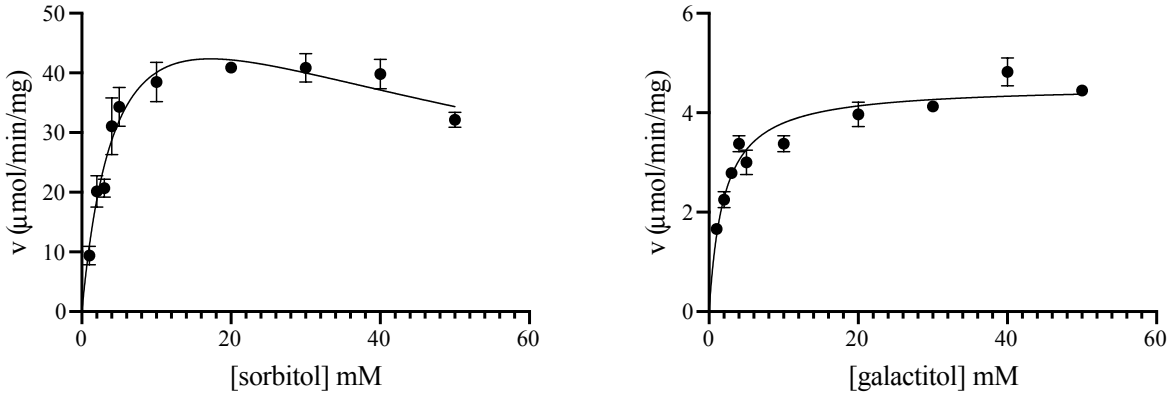


Figure 3.6. Kinetic analysis of sorbitol and galactitol oxidation by *S. meliloti* SmoS.

Table 3.2. Kinetic properties of *S. meliloti* SmoS

Substrate	K_M (mM)	V_{max} (μmol/min)	V_{max}/K_M (min^{-1})
Sorbitol	5.161	67.71	13.12
Galactitol	1.98	4.547	2.30

Sorbitol and galactitol oxidation data were fit to substrate inhibition and Michaelis-Menten models respectively. K_M and V_{max} were determined using GraphPad Prism 8.3.0.

It was determined that SmoS has a K_M of 5.16 mM for sorbitol, and a K_M of 1.98 mM for galactitol (Table 3.2), however, the maximum velocity (V_{max}) of the sorbitol oxidation reaction was calculated to be 67.7 mM/min, while galactitol oxidation proceeded at only 4.5 mM/min (Table 3.2). Despite a higher affinity, the low reaction velocity of galactitol oxidation greatly reduces the overall reaction efficiency (V_{max}/K_M). We note that L-iditol shares hydroxyl group orientation about carbons 1, 2, and 3, with sorbitol and galactitol, however this substrate was not tested due to lack of availability (Philippsen *et al.*, 2005).

3.4.3 SmoS-tagatose complex is predicted to be in a lower energy state than SmoS-fructose complex

Kinetic analysis revealed that galactitol turnover is much less efficient than sorbitol oxidation (Fig. 3.6, Table 3.2). This observation was particularly interesting due to the K_M value of galactitol oxidation, which suggested that the enzyme's affinity for galactitol was higher than for sorbitol (Table 3.2). This led to the hypothesis that tagatose is a poor leaving group in comparison to fructose and the inability of tagatose to quickly leave the active site results in low reaction turnover. This hypothesis is supported by our inability to detect fructose in the active site of SmoS structures determined from crystals grown in the presence of a large concentration (20%) of fructose.

To test this hypothesis, computational ligand docking analysis was conducted using the Rosetta Ligand Docking Protocol on the ROSIE server (Combs *et al.*, 2013; DeLuca *et al.*, 2015; Kothiwale *et al.*, 2015; Lyskov *et al.*, 2013). D-fructose and D-tagatose model files were submitted to the ligand dock protocol along with apo SmoS monomer structure, and the outputs were analyzed for indications of the energy state of the complexes. The server generated 200

docking predictions for each SmoS-ligand complex, which were organized via their interface delta scores. The interface delta score represents the total energy of the complex in isolation subtracted from the total energy of the complex with the substrate bound (Kaufmann & Meiler, 2012). The ten models with the lowest interface delta score from each complex were selected. The scores from the SmoS-fru model complexes were consistently higher than the scores reported for the SmoS-tag complexes, suggesting that the SmoS-tag complex is in a lower energy state with higher stability than the SmoS-fru complex (Fig. 3.7A). The data from each SmoS-ligand complex were analyzed for significance via a student's *t* test, revealing a *P* value of $1.3x10^{-6}$. The entire process from submission to the server through data collection and analysis was repeated independently to evaluate reproducibility; the SmoS-tag complexes were consistently in a lower energy state than the SmoS-fructose complexes. The *P* value for the second trial was $2.4x10^{-6}$.

An examination of the hydrogen bonding interactions that mediate binding reveals that the SmoS-tag complex forms an additional hydrogen bond that is not present in the SmoS-fru complex, which further stabilizes the tagatose bound structure (Fig. 3.7B and C). These data suggest that the SmoS-tag complex is a lower energy and more stable complex than the SmoS-fru complex, and that the predicted interface energies from the SmoS-fru complexes and the SmoS-tag complexes are statistically different. They also support the hypothesis that tagatose is a poor leaving group in comparison with fructose and are consistent with observations of the kinetic properties of the enzyme.

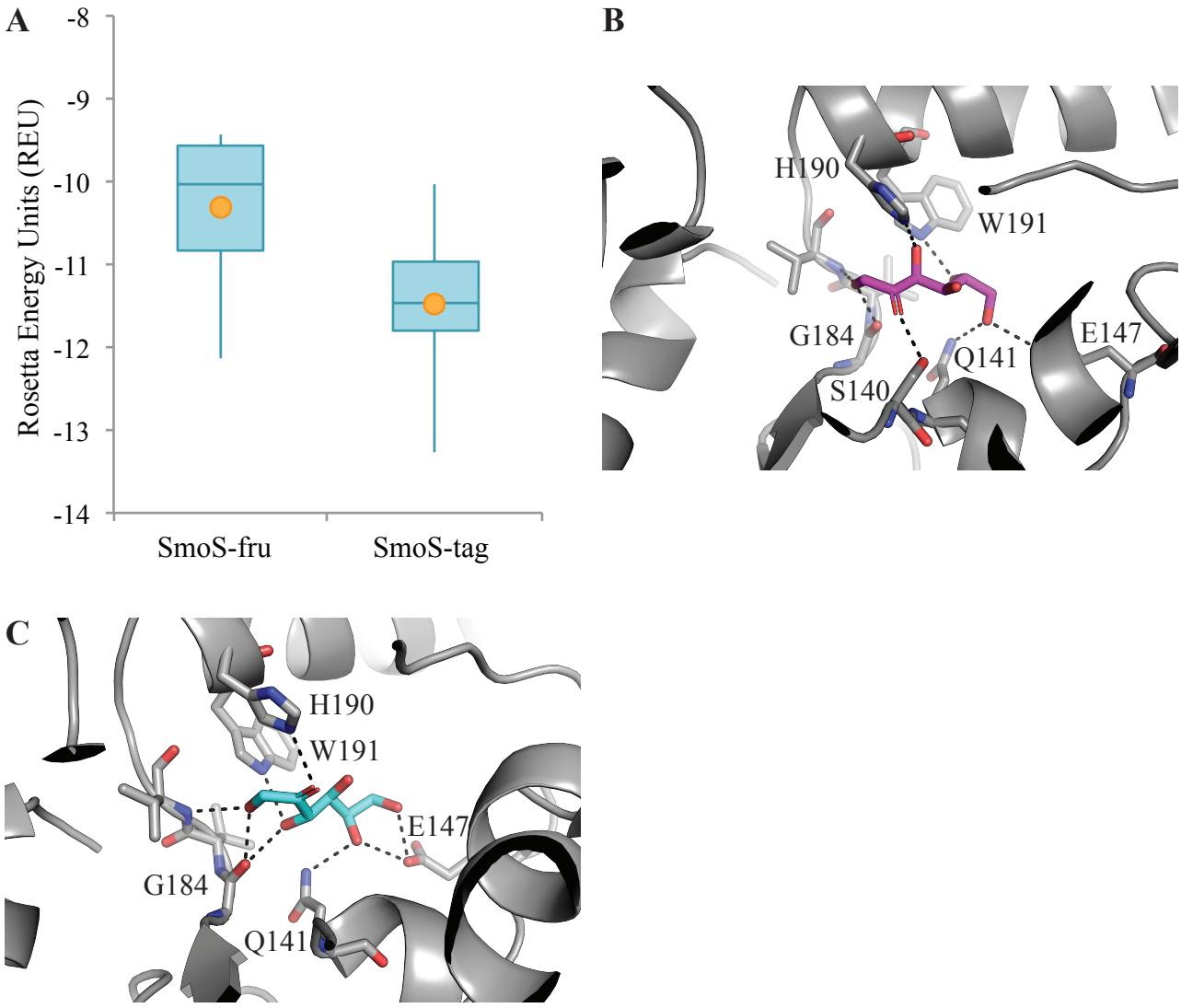


Figure 3.7. SmoS-fructose and SmoS-tagatose binding complexes predicted by the Ligand docking protocol housed on the ROSIE server. (A) The distribution of the top ten interface delta scores displayed as box and whisker plots. The tips of the whiskers represent the maximum and minimum values, the horizontal lines represent the first, second, and third quartiles, and the orange dots represent the averages of the data sets. P value of 1.3x10-6. The lowest energy docking prediction for the SmoS-fructose complex (B) and the SmoS-tagatose complex (C). Fructose is shown in magenta and tagatose in blue.

3.5 Discussion

S. meliloti SmoS appears to be most similar to the sorbitol dehydrogenase from *R. sphaeroides*, these enzymes share kinetic characteristics (Schauder *et al.*, 1995), operon structures (Stein *et al.*, 1997), pH preferences (Schauder *et al.*, 1995), and overall quaternary structure (Philippsen *et al.*, 2005). SmoS can be classified within a group of "high-alkaline enzymes," which are enzymes with a pH optimum from pH 10-11. These enzymes are useful in industry due to their high durability (Fujinami & Fujisawa, 2010). Similar to *Rs*SmoS, *Sm*SmoS was found to have a higher affinity toward galactitol compared to sorbitol, but turned over sorbitol at a faster rate (Schauder *et al.*, 1995).

Most of the crystal structures of SmoS related enzymes have reported tetrameric structures found in the crystal packing (Carius *et al.*, 2010; Fredslund *et al.*, 2016; Philippsen *et al.*, 2005), although reports differ on the structure of the enzyme in solution. *R. sphaeroides* SmoS has been reported as dimeric in solution, on the basis of gel filtration chromatography as well as sucrose gradient centrifugation experiments (Schauder *et al.*, 1995). However, the enzyme was later predicted to function as a tetramer based on predicted surface area exposure (Philippsen *et al.*, 2005), and these results were supported by size exclusion chromatography and light scattering experiments (Carius *et al.*, 2010). *Bj*SDH had been proposed to exist as a trimer in solution (Gauer *et al.*, 2014) but researchers later suggested that a tetramer was more likely (Fredslund *et al.*, 2016). A galactitol dehydrogenase from *Rhizobium leguminosarum* 3841 has also been reported to be tetrameric in solution (Jagtap *et al.*, 2014). The data presented clearly shows that SmoS from *S. meliloti* is present as a tetramer in solution but with a small subset seemingly present as a hexamer or an octamer made up of a dimer of tetramers (Fig. 3.2). Of note, it appears that both the tetrameric as well as the higher oligomeric forms show sorbitol

dehydrogenase activity (Fig. 3.2). Tetrameric configurations are reported most often and likely represent the majority of SDR protein structures in solution (Zhu *et al.*, 2017).

The SmoS-sbt structure shows that the hydroxyl group bonded to C1 of sorbitol associating with catalytic residue Tyr153, and that the structure has a subtle difference from the apo structure in that residues His190 and Trp191 in alpha helix 7 are contorted slightly to accommodate the presence of the substrate (Fig. 3.4C). As well, residues Asn111, Ser140, Tyr153, and Lys157, which have been proposed to be involved in electron transfer, are too distant from the substrate for catalysis (Fig. 3.4A).

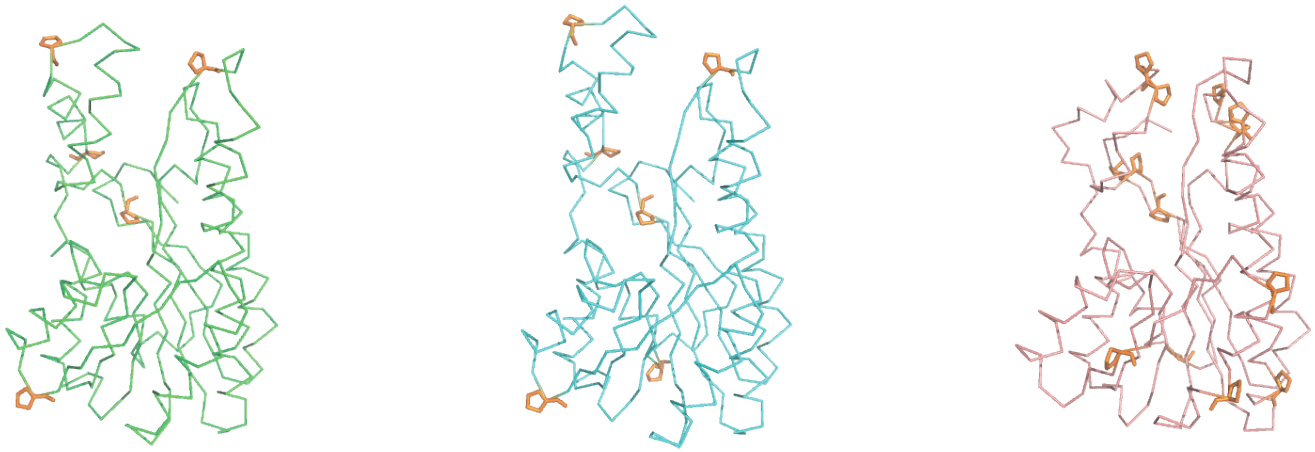
If the positioning of Tyr153 were correct, it would imply that sorbitol should be oxidized to glucose. Based on the available genetic and physiological data it is clear that both sorbitol and galactitol catabolism mediated by SmoS generate fructose and tagatose via an enzymatic reaction in which the hydroxyl group on C2 of the substrate is oxidized forming a planar carbonyl carbon. We also note that enzymes catalyzing the oxidation of sorbitol into glucose are known as sorbitol oxidase (SOX) proteins (Hiraga *et al.*, 1997; Yamaki, 1980). These enzymes are dissimilar to SDH enzymes of the SDR family (Forneris *et al.*, 2008; Heuts *et al.*, 2007).

This anomaly could be due to the absence of NAD^+ in the binding pocket. NAD^+ was left out of the crystallization solution because its presence would result in an enzymatic reaction, which would prevent the capture of a substrate-bound complex. However, SDR reactions proceed with the coenzyme binding first and leaving last (Kavanagh *et al.*, 2008), which may help to explain not only why sorbitol is found in an atypical position, but also why fructose was not found in the active site of the fructose grown crystal structures despite its presence at high concentrations. In addition, modeling of NAD^+ and sorbitol into the *R. sphaeroides* predicted

direct contact and a sandwiching of the C2 carbon of sorbitol between the active site tyrosine, and the nicotinamide ring. Taken together these may explain the observed structure.

Thermal stability of an enzyme can affect its ability to be exploited in industrial processes (Chapman *et al.*, 2018). It has been proposed that the increased thermal stability of SDH is due to the abundance of proline residues and the proline to glycine ratio in its primary amino acid sequence (Fredslund *et al.*, 2016). Proline is a rigid residue with low configurational entropy due to its pyrrolidine ring hindrance, there are several studies that suggest protein thermostability can be influenced by proline content (Matthews *et al.*, 1987; Suzuki, 1989; Suzuki *et al.*, 1987). *Rs*SDH contains 6 proline residues and a Pro/Gly ratio of 0.22, while *Bj*SDH has 13 prolines with a ratio of 0.86. The melting temperatures were found to be 62°C and 47°C respectively (25, 39). The SmoS from *S. meliloti* has 5 proline residues and the Pro/Gly ratio is 0.2, additionally the position of the residues appears to be conserved, indicating that it's thermostability is likely more similar to *Rs*SDH (Fig. 3.8).

The structure and characterization of *S. meliloti* SmoS provides a high quality structure with sorbitol within the active site. In addition, the characterization and determination of its affinities for its substrates provides insight into why the growth rate of the organism on what should be two equivalent carbon substrates shows great differences. This information is invaluable for higher order resolution of metabolism in *S. meliloti*.



*Sm*SmoS *Rs*SmoS *Bj*SDH

Figure 3.8. Comparison of the position and distribution of proline residues for *S. meliloti* SmoS (green), *R. sphaeroides* SmoS (blue; PDB ID: 1K2W), and *B. japonicum* SDH (pink; PDB ID: 5JO9), proline residues are shown in orange.

Chapter 4:

Physiological characterization of a locus responsible for the metabolism of multiple sugar alcohols in *Sinorhizobium meliloti*

4.1 Abstract

The *smo* locus (sorbitol mannitol oxidation) is found on the chromosome of *S. meliloti*'s tripartite genome. Mutations at the *smo* locus reduce or abolish the ability of the bacterium to grow on a number of carbon sources including sorbitol, mannitol, galactitol, D-arabitol, and maltitol. The contribution of the *smo* locus to the metabolism of these compounds as well as its role in symbiosis and competition for nodule occupancy has not been previously investigated.

Genetic complementation of mutant strains revealed that *smoS* is responsible for growth on sorbitol and galactitol, while *mtlK* restores growth on mannitol and D-arabitol. Dehydrogenase assays demonstrate that SmoS and MtlK are NAD^+ dependent dehydrogenases catalyzing the oxidation of their specific substrates. Additionally, it is shown that mutations to *frk* abolish growth on sorbitol, mannitol, as well as fructose, and that Frk has fructose kinase activity. A locus involved in the transport of fructose (*frc*) is implicated in the uptake of sorbitol and mannitol. Transport experiments using a radiolabelled substrate indicate that sorbitol, mannitol, and D-arabitol compete for use of the same transporter(s). Also, evidence is presented suggesting that fructose-6-phosphate is a key metabolite, which regulates the uptake of sugar alcohols. Strains with mutations at the *smo* locus do not exhibit symbiotic deficiencies or the inability to compete for nodule occupancy.

4.2 Introduction

Biochemical enzyme activities for sorbitol dehydrogenase, mannitol dehydrogenase, and D-arabitol dehydrogenase have been demonstrated from protein extracts of *S. meliloti* grown on sugar alcohols (Martínez De Drets & Arias, 1970). Fructose kinase activity has also been observed and mutant strains lacking this activity are unable to grow using sorbitol, mannitol, or fructose as a sole carbon source suggesting that metabolism of these polyols proceeds using fructose as an intermediate (Gardiol *et al.*, 1980). Similarly, phosphoglucose isomerase (*pgi*) mutants are unable to grow using sorbitol, mannitol, or fructose as sole carbon sources (Arias *et al.*, 1979). These mutations have not been precisely identified to a genetic locus.

A genome wide mutagenic screen of short chain dehydrogenase/reductase family genes has identified a putative sorbitol dehydrogenase called *smoS*. Mutations in *smoS* result in a strain which exhibits reduced growth using sorbitol, mannitol, maltitol, myo-inositol, or meso-erythritol, as a sole carbon source (Jacob *et al.*, 2008). The *smo* locus consists of a regulator (*smoC*), an ABC type transporter (*smoEFGK*), a sorbitol dehydrogenase (*smoS*), a mannitol dehydrogenase (*mtlK*), a phosphatase, and a kinase (Fig. 4.1A). The solute-binding protein, *smoE*, has been shown to be induced by sorbitol, mannitol, galactitol, and maltitol (Mauchline *et al.*, 2006).

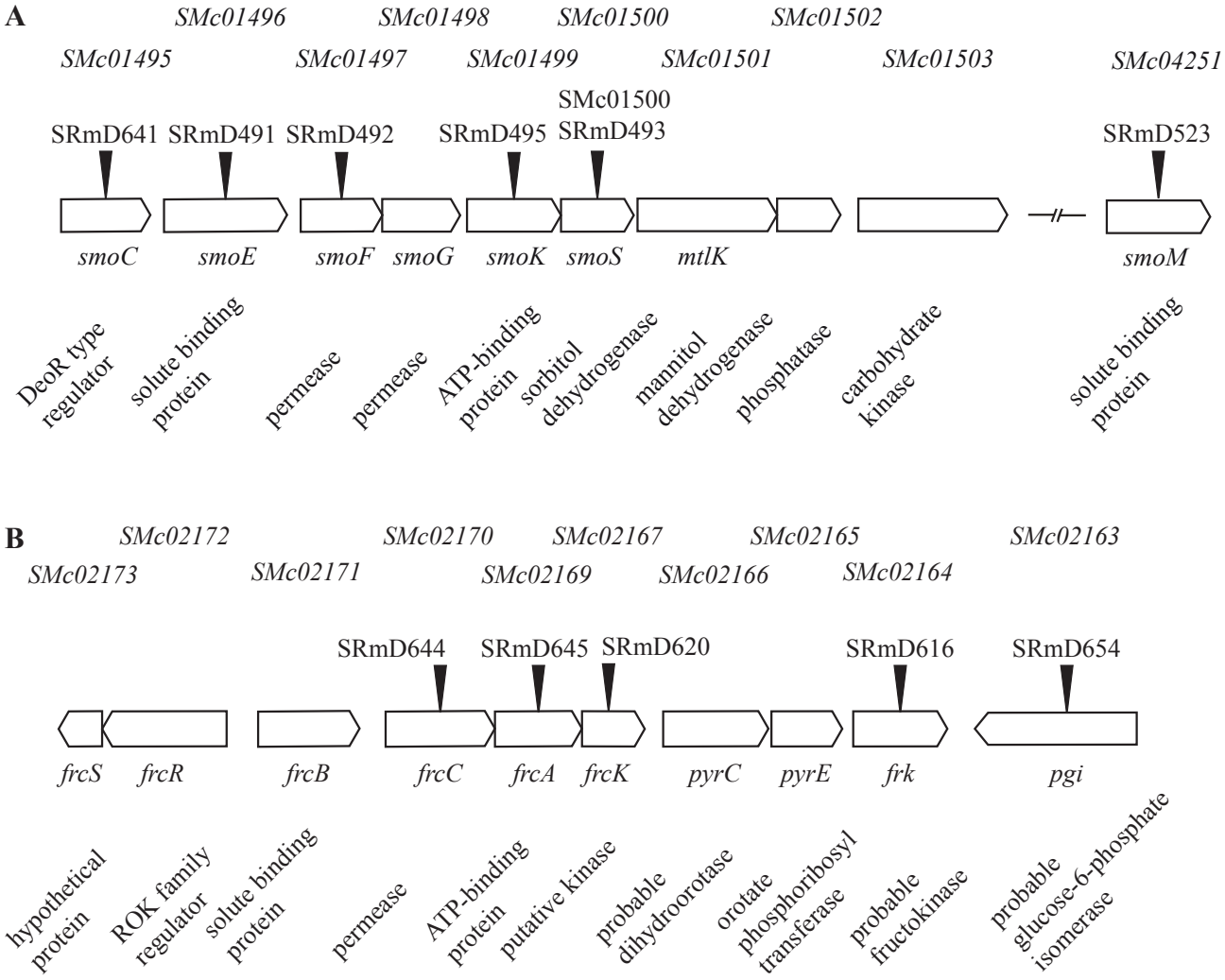


Figure 4.1. Locus diagrams of the *smo* (A) and *frc* (B) loci. Locus tags appear above the genes and annotation information appears below. Black wedges represent the sites of insertional mutagenesis, strain designations are listed above these indicators.

4.3 Materials and methods

4.3.1 Bacterial strains, plasmids, and culture conditions

Strains and plasmids used in this study are listed in Table 4.1. *S. meliloti* strains were grown at 28°C on either Luria-Burtani (LB) (Cold Spring Harbor Protocols, 2006) or LB^{MC} (LB amended with 2.5 mM $MgSO_4$ and 2.5 mM $CaCl_2$ (Finan *et al.*, 1988)) medium as a complex medium or Vincent's minimal medium (VMM) as a defined medium (Vincent, 1970). Filter sterilized carbon sources were generally added to the sterile medium to a final concentration of 15 mM. *E. coli* strains were grown at 37°C on LB medium. Filter sterilized antibiotics were added to the medium as required at the following concentrations: streptomycin (Sm) 200 μg ml^{-1}, neomycin (Nm) 200 μg ml^{-1}, gentamicin (Gm) 20 or 60 μg ml^{-1}, kanamycin (Km) 20 μg ml^{-1}, chloramphenicol (Cm) 20 μg ml^{-1}, tetracycline (Tc) 5 μg ml^{-1}. Strains that were constructed were generally single colony purified prior to use.

4.3.2 DNA manipulations and genetic techniques

Standard techniques were used for plasmid isolation, restriction enzyme digests, ligations, transformations, and agarose gel electrophoresis (Sambrook & Russell, 2001). Conjugations and transductions were carried out essentially as previously described (Finan *et al.*, 1988; Finan *et al.*, 1984). Tn*5* mutagenesis was performed as previously described (Finan *et al.*, 1985), and SRmD620 was isolated as strain unable to utilize fructose as a sole carbon source. The point of insertion was determined by using an arbitrary PCR protocol as previously described (Poysti *et al.*, 2007).

Cosmid pJD06 was isolated by complementation of SMc1500 for growth on sorbitol using cosmid bank CX1 as previously described (Geddes & Oresnik, 2012a; Wang *et al.*, 2006).

Table 4.1. Bacterial strains and plasmids

Strain or plasmid	Relevant Characteristics	Reference
Strains		
S. meliloti		
Rm1021	SU47 *str-21*, Sm^R	Meade *et al.* (1982)
RmP110	Rm1021 *pstC*$^+$, Sm^R	Yuan *et al.* (2006)
SMc01500	RmP110 *smoS*::pTH1703, Gm^R	Jacob *et al.* (2008)
FL4643	RmP110 *mtlK*::pTH1522, Gm^R	Cowie *et al.* (2006)
SRmA723	Rm1021 *SMc01627*::Tn*5*-233, Gm^RSp^R	Geddes *et al.* (2010)
SRmD491	Rm1021 *smoE*::Tn*5*-B20, Nm^R	Kohlmeier *et al.* (2019)
SRmD492	Rm1021 *smoF*::Tn*5*-B20, Nm^R	Kohlmeier *et al.* (2019)
SRmD493	Rm1021 *smoS*::Tn*5*-B20, Nm^R	Kohlmeier *et al.* (2019)
SRmD495	Rm1021 *smoK*::Tn*5*-B20, Nm^R	Kohlmeier *et al.* (2019)
SRmD501	ΦSRmD491→SRmA723	This study
SRmD502	ΦSRmD492→SRmA723	This study
SRmD503	ΦSRmD495→SRmA723	This study
SRmD523	Rm1021 *smoM*::pKNOCK-Gm, Gm^R	Kohlmeier *et al.* (2019)
SRmD524	ΦSRmD491→SRmD523	This study
SRmD525	ΦSRmD492→SRmD523	This study
SRmD526	ΦSRmD495→SRmD523	This study
SRmD616	Rm1021 *frk*::pKNOCK-Gm, Gm^R	This study
SRmD618	Rm1021 *frcK*::pKan, Nm^R	This study
SRmD620	Rm1021 *frcK*::Tn*5*, Nm^R	This study
SRmD641	Rm1021 *smoC*::pKan, Nm^R	This study
SRmD642	ΦSRmD641→FL4643	This study
SRmD654	Rm1021 *pgi*::pKNOCK-Gm, Gm^R	This study
SRmD664	Rm1021 *frcA*::pKNOCK-Gm, Gm^R	This study
SRmD665	Rm1021 *frcC*::pKNOCK-Gm, Gm^R	This study
SRmD666	ΦSRmD495→SRmD664	This study
SRmD667	ΦSRmD495→SRmD665	This study
E. coli		
DH5α	F^- *supE44 lacU169 hsdR17 recA1 endA1 gyrA96 thi-1 relA1* (80*lacZ*ΔM15)	Hanahan (1983)
DH5αRλpir	λpir lysogen of DH5α	House *et al.* (2004)
MM294A	*pro-82-thi-1 hsdR17 supE44*	Finan *et al.* (1986)
MT607	MM294A *recA56*	Finan *et al.* (1986)
MT616	MT607 (pRK600)	Finan *et al.* (1986)
EcA101	MT607ΩTn*5*-B20, Km^R	Clark *et al.* (2001)
Plasmids		
pSMc01500	*smoS*/pCO37, Tc^R	Kohlmeier *et al.* (2019)
pJD02	*mtlK*/pCO37, Tc^R	This work
pJD06	CX1 derived sorbitol/mannitol	This work

	complementing cosmid, Tc^R	
pMK8	pJD06 *smoE*::Tn*5*-B20	This work
pMK12	pJD06 *smoF*::Tn*5*-B20	This work
pMK13	pJD06 *smoS*::Tn*5*-B20	This work
pMK18	pJD06 *smoK*::Tn*5*-B20	This work
pMK38	*frk*/pCO37, Tc^R	This work
pMK39	*frcK*/pCO37, Tc^R	This work
pMK48	6xHis*frk*/pRK7813, Tc^R	This work
pMK62	*pgi*/pCO37, Tc^R	This work
pRK7813	Broad host range vector, Tc^R	Jones and Gutterson (1987)
pCO37	Gateway compatible vector, Tc^R	Jacob *et al.* (2008)
pRK600	pRK2013 *nptI*::Tn*9*, Cm^R	Finan *et al.* (1986)
pRK602	pRK600ΩTn*5*, Cm^R Nm^R	Finan *et al.* (1985)
pPH1JI	IncP plasmid, Gm^R	Beringer *et al.* (1978)
pKan	Suicide vector, pKNOCK-Gm derivative, Km^R	Pickering and Oresnik (2008)
pKNOCK-Gm	Suicide vector, Gm^R	Alexeyev (1999)
pTH1522	Reporter vector, *gfp*$^+$, *lacZ*, *gusA*, *tdimer2(12)*, Gm^R	Cowie *et al.* (2006)

Strains SRmD491, SRmD492, SRmD493 and SRmD495 were generated by mutagenizing with Tn*5*-B20 using the strain EcA101 as previously described (Clark *et al.*, 2001). Briefly, pJD06 was conjugated into EcA101, which carries a chromosomally localized Tn*5*-B20, and subsequently reintroduced into SMc1500 and screening the transconjugants for lack of complementation on sorbitol. Plasmids pMK8, 12, 14, and 17 were found to carry inserts in *smoE*, *smoF*, *smoK*, and *smoS* respectively. These plasmids were then introduced into Rm1021, and allelic exchange was carried out using pPH1JI as previously described (Poysti *et al.*, 2007).

Strains SRmD616, SRmD641, and SRmD654 were generated by targeted mutagenesis using either pKNOCK-Gm (Alexeyev, 1999) or pKan (Pickering & Oresnik, 2008). A fragment from the gene of interest was PCR amplified using primers *frk*_pK_F/R, *smoC*_pK_F/R, and *pgi*_F1/R1 and cloned into pKNOCK-Gm or pKan (Table 4.2). Constructs were conjugated into Rm1021 and recombinants were single colony purified.

Plasmids pJD02, pMK38, and pMK39 were generated using the ORFeome Gateway system (House *et al.*, 2004; Schroeder *et al.*, 2005) using pCO37 as a destination vector as previously described (Geddes & Oresnik, 2012b). To generate pMK48, *frk* was PCR amplified as a *Hind*III/*Eco*RI fragment using primers *frk*_F2/*frk*_R2, which contain a ribosome-binding site and 6xHis tag to facilitate expression and purification respectively (Table 4.2). This fragment was cloned into pRK7813 such that it was constitutively expressed in *S. meliloti* by the P_{lac} promoter (Richardson *et al.*, 2004).

Table 4.2. Primers used during this study

Name	Sequence 5′→3′
Generation of mutants	
*frk*_pK_F	ATAT**GGATCC**AGCTCGTCGACGGGCATGC
*frk*_pK_R	ATAT**CTCGAG**CGTCCGCGCCGCGGGTG
*frcK*_pK_F	ATAT**TCTAGA**GCTTGCCCGCAAGGGGGC
*frcK*_pK_R	ATAT**GGATCC**CGCTTCGTCGAGCAGCAGG
*smoC*_pK_F	ATAT**TCTAGA**TCGACCACCGTGGGCATC
*smoC*_pK_R	ATAT**CCCGGG**TCTTCGCAAAGCGGCGCG
*pgi*_F1	ATAT**TCTAGA**TCGAGACGATGACCAATGCG
*pgi*_R1	ATAT**CCCGGG**GCCGGAAAGCGCGTCAGC
*frcA*_pK_F	ATAT**GGATCC**ACGGTCTATCAGAACCTCGC
*frcA*_pK_R	ATAT**CTCGAG**TGGATGGTCATCAGACCGAG
*frcC*_pK_F2	ATAT**GGATCC**CATCATGGGGCAGTTCACCT
*frcC*_pK_R2	ATAT**CTCGAG**CATAGACATAGCGCCCCCAG
Confirmation of mutants	
*frk*_F	ATGATCGTTTGCTGCGGAGAG
*frk*_R	TCGGAGCCCGAGTTCGTGTC
*frcK*_F	ATGAGCGTCAAATCCCTTGCG
*frcK*_R	CTAAAAGTCCCTGATCACGAC
*smoC*_F	ATGGCACGCAAGGCGGAAAG
*smoC*_R	CTAGACCCGGAGCAGATAGTC
*pgi*_F3	CTTCATCGTCGCCTCCAAGA
*pgi*_R3	ATGACATCGGTTCCCTGGTG
*frcA*_F2	GGCACAGGAACCCATTCTCA
*frcA*_R2	TGGGATTGATGACGCAGAGG
*frcC*_F2	AGCACTTCCTGCATTCGAGC
*frcC*_R2	AGCAAGCCGATCAGGAGATA
Generation of expression constructs	
*frk*_F2	ATAT**AAGCTT**GGAGATGCATGCATGCACCACCACCACCACCA CATCGTTTGCTGCGGAGAGGC
*frk*_R2	ATAT**GAATTC**TCAGAGCCCGAGTTCGTCTC

Restriction sites are in bold.

4.3.3 Protein purification and biochemical enzyme assays

Rm1021 cells expressing pMK48 were grown in 1 L volume of LB, shaking, at 30°C for two days. The cells were then harvested by centrifugation (6000 x g for 10 minutes), resuspended in lysis buffer (50 mM TRIS pH 8.0, 300 mM NaCl, 2 mM DTT, 10 mM imidazole), and lysed using a French Press (16000 lb/in^2). The lysate was cleared by centrifugation (6000 x g for 10 minutes). The cleared lysate was then loaded onto Ni-NTA column, washed with 10 column volumes of lysis buffer and eluted from the column with lysis buffer supplemented with 500 mM imidazole. Eluted fractions were separated by SDS-PAGE and visualized by staining with Coomassie Brilliant Blue. Fractions containing the protein of interest were dialyzed overnight in 2 L dialysis buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 2 mM DTT) prior to being used to determine kinase activity. Fructose kinase assays were conducted essentially as previously described (Anderson & Sapico, 1975). Fructose kinase activity was coupled to the production of NAD^+ by pyruvate kinase (PK) and lactate dehydrogenase (LDH), which was measured at 340 nm for two minutes in a buffer containing 60 mM HEPES pH 7.5, 6 mM $MgCl_2$, 3 mM ATP, 3 mM PEP, 0.3 mM NADH, and 1/50 vol PK/LDH mix (Sigma). The assay was initiated with 6 mM substrate, either fructose or glucose. Measured rates of NADH oxidation were linear over a two-minute period and proportional to the volume of extract that was used in the assay.

Dehydrogenase assays were conducted as previously described (Kohlmeier *et al.*, 2019). Briefly, cell free extracts were separated by nondenaturing polyacrylamide gel electrophoresis and the gels were stained for dehydrogenase activity using an assay reagent containing phenazine methosulfate, nitro blue tetrazolium salts, NAD^+, and a substrate of interest.

4.3.4 Transport assays

Transport rates were determined as previously described (Geddes & Oresnik, 2012a; Geddes & Oresnik, 2012b; Rivers & Oresnik, 2013). Briefly, cultures were grown to an OD_{600} of 0.3 and ^{14}C-mannitol or ^{14}C-fructose were added to cultures to a final concentration of 2µM. Competing substrates were added to a final concentration of 2µM or 10µM. Subsequently, culture aliquots were passed through a 0.45 µm Hv filter on a Millipore sampling manifold at specified time points. Accumulation of radiolabel was quantified using a liquid scintillation counter (Beckman LS6500). Uptake rates were standardized to total protein.

4.3.5 Fluorescence gene expression analysis

Strains were grown overnight in 5 mL cultures at 30°C using LB^{MC} medium with appropriate antibiotics. These cells were subcultured into 3 mL LB medium and induced with 15 mM sorbitol as a positive control, or 10% (v/v) seed exudate, and grown for two days, 200 µg/µL Gm was included in the medium. 100 µL of these cultures were aliquoted into a 96 well black Greiner microplates. Wells were read at OD_{600} for growth as well as 485 nm (excitation) and 510 (emission) for fluorescence using a SpectraMax M2 microplate reader (Molecular Devices, Sunnyvale, CA, USA). Relative fluorescence was determined as (fluorescence at 485/510-background)/OD_{600}. Values obtained from Rm1021 were used as background (Cowie *et al.*, 2006).

4.4 Results

4.4.1 smoS *is responsible for growth with sorbitol and galactitol while* mtlK *is responsible for mannitol and D-arabitol utilization*

The gene encoding SmoS is found within a group of genes that has been annotated as being involved in sorbitol and mannitol catabolism based on its similarity to the *Rhodobacter sphaeroides* (Stein *et al.*, 1997). The region consists of between 8-10 genes and contains a DeoR type regulator (encoded by *smoC*), an ABC transporter consisting of a solute binding protein (encoded by *smoE*), two permeases (encoded by *smoF* and *smoG*), an ATP binding protein (encoded by *smoK*), as well as a sorbitol dehydrogenase (encoded by *smoS*) and a putative mannitol dehydrogenase (encoded by *mtlK*). Two other genes that may be part of this region are *SMc01502* (encoding a putative phosphatase hydrolase) and *SMc01503* (encoding a putative carbohydrate kinase) (Fig 4.1A). It has been shown that mutations in *smoS* result in reduced growth on sorbitol, mannitol, and maltitol (Jacob *et al.*, 2008) as well as galactitol (Kohlmeier *et al.*, 2019). To better characterize this region and its role in polyol catabolism in *S. meliloti*, a cosmid capable of complementing the strain SMc1500 for growth on sorbitol was isolated. The cosmid, pJD06, was mutagenized using Tn*5*-B20 and the insertions were subsequently recombined into the chromosome (Table 4.1). The results of the mutagenesis ultimately yielded strains SRmD491, SRmD492, SRmD493, and SRmD495, which were found to house insertions within *smoE*, *smoF*, *smoS*, and *smoK* respectively (Fig. 4.1A).

When these strains were streaked on defined medium it was found that growth on maltitol is reduced, while growth on sorbitol, mannitol, and D-arabitol is nearly abolished (Table 4.3). Introduction of p*SMc01500*, which carries a wild-type copy of *smoS,* into either SMc01500 or SRmD493, which have insertions in *smoS*, resulted in the restoration of growth on sorbitol,

galactitol, and maltitol. The introduction of pJD02, which carries a wild-type copy of *mtlK*, resulted in the restoration of growth on mannitol and D-arabitol (Table 4.3). Introduction of either pSMc01500 or pJD02 into SRmD491, SRmD492, or SRmD495, which carry mutations in the associated ABC transporter also similarly resulted in the restoration of growth on defined medium. This suggests that SmoEFGK likely transports sorbitol, mannitol, D-arabitol, and galactitol, and that the genes from *smoE* to *mtlK* are part of a monocistronic operon whereby that insertional disruption of *smoEFGK* has polar effects on these downstream metabolic genes (Table 4.3). In addition there must exist an alternate transporter in the genome that is capable of recognizing sorbitol, mannitol, maltitol as well as D-arabitol and providing physiologically relevant transport if the respective dehydrogenases necessary for the catabolism of these substrates is provided on a multicopy plasmid (Table 4.3).

The gene *SMc04251* is annotated as *smoM*, implying that it has a role in sorbitol-mannitol catabolism. SRmD523 was previously constructed to determine if it had a role in galactitol transport (Kohlmeier *et al.*, 2019). To determine if this gene could play a role in the transport of sorbitol/mannitol the Tn*5*-B20 insertions in *smoE*, *smoF*, and *smoK* were transduced into SRmD523 yielding SRmD524, SRmD525, and SRmD526 respectively. It was found that SRmD523 did not have any growth defects on the polyols, and that the addition mutations to the ABC transporter at the *smo* locus did not alter the phenotypes already associated with these genes (Table 4.3).

Table 4.3. Complementation analysis of *smo* mutants

Strain	Characteristic	LB	sbt	mtl	D-atl	fru	gly
Rm1021	wt	+	+	+	+	+	+
SRmD491	*smoE*::Tn*5*B20	+	+/-	+/-	+/-	+	+
smoS	*smoS in trans*	+	+	+/-	+	+	+
mtlK	*mtlK in trans*	+	+/-	+	+	+	+
SRmD492	*smoF*::Tn*5*B20	+	+/-	+/-	+	+	+
smoS	*smoS in trans*	+	+	+/-	+	+	+
mtlK	*mtlK in trans*	+	+/-	+	+	+	+
SRmD493	*smoS*::Tn*5*B20	+	-	-	+/-	+	+
smoS	*smoS in trans*	+	+	-	+/-	+	+
mtlK	*mtlK in trans*	+	-	+	+	+	+
SRmD495	*smoK*::Tn*5*B20	+	+/-	+/-	+	+	+
smoS	*smoS in trans*	+	+	+/-	+	+	+
mtlK	*mtlK in trans*	+	+/-	+	+	+	+
SRmD523	*smoM*::pKGm	+	+	+	+	+	+
smoS	*smoS in trans*	+	+	+	+	+	+
mtlK	*mtlK in trans*	+	+	+	+	+	+
SRmD524	*smoEM*	+	+/-	+/-	+	+	+
smoS	*smoS in trans*	+	+	+/-	+	+	+
mtlK	*mtlK in trans*	+	+/-	+	+	+	+
SRmD525	*smoFM*	+	+/-	+/-	+	+	+
smoS	*smoS in trans*	+	+	+/-	+	+	+
mtlK	*mtlK in trans*	+	+/-	+	+	+	+
SRmD526	*smoKM*	+	+/-	+/-	+	+	+
smoS	*smoS in trans*	+	+	+/-	+	+	+
mtlK	*mtlK in trans*	+	+/-	+	+	+	+

Growth scored as follows; +, like wildtype; -, no growth; +/-, intermediate growth. Abbreviations are as follows; LB, Luria-Bertani; sbt, sorbitol; mtl, mannitol; D-atl, D-arabitol; fru, fructose; gly, glycerol.

We also hypothesized that the range of substrates take up from the multiple polyol transporter (*mpt*) may extend beyond erythritol, L-arabitol, and adonitol (Geddes & Oresnik, 2012b), and may contribute to the uptake of sorbitol, mannitol, and D-arabitol. The mutated *smoE*, *smoF*, and *smoK* alleles were transduced into a strain containing a mutated ABC type transporter permease component (*mptB*) generating SRmD501, SRmD502, and SRmD503. These strains exhibited phenotypes identical to the *smo* mutant parental strain with the exception that they could not utilize erythritol for growth (Table 4.4). Taken together these data suggest that SmoM and MptABCDE do not play a role in sorbitol/mannitol transport.

Table 4.4. Carbon phenotypes of *smo* and *mpt* transport mutants

Strain	Characteristic	LB	sbt	mtl	D-atl	ery	fru	gly
Rm1021	wt	+	+	+	+	+	+	+
SRmA723	*mptB*::Tn*5*-233	+	+	+	+	-	+	+
SRmD491	*smoE*::Tn*5*-B20	+	+/-	+/-	+/-	+	+	+
SRmD492	*smoF*::Tn*5*-B20	+	+/-	+/-	+	+	+	+
SRmD495	*smoK*::Tn*5*-B20	+	+/-	+/-	+	+	+	+
SRmD501	*smoEmptB*	+	+/-	+/-	nd	-	nd	+
SRmD502	*smoFmptB*	+	+/-	+/-	nd	-	nd	+
SRmD503	*smoKmptB*	+	+/-	+/-	nd	-	nd	+

Growth; +, like wildtype; -, no growth; +/-, intermediate growth; nd, not determined.
Abbreviations; LB, Luria-Bertani; sbt, sorbitol; mtl, mannitol; D-atl, D-arabitol; ery, erythritol; fru, fructose; gly, glycerol.

4.4.2 SmoS and MtlK are polyol dehydrogenases

The catabolism of polyols is most commonly initiated by the oxidation at the C2 position to yield a keto-sugar that is subsequently phosphorylated (Kohlmeier *et al.*, 2019; Mortlock, 1984). Sorbitol and mannitol are epimers, oxidation of either of these yields fructose. Although D-arabitol is a pentitol rather than a hexitol, it also shares the same stereochemistry over its terminal three carbons, suggesting that it too can be oxidized in a manner similar to mannitol.

To provide evidence to corroborate this hypothesis wild-type cells were grown in defined medium with glycerol as a sole carbon source and induced with sorbitol for six hours. The cultures were lysed, and the protein extracts were then separated using nondenaturing polyacrylamide gel electrophoresis. Subsequently, the gels were assayed for enzyme activities by staining for dehydrogenase activity and compared with a control gel that did not have exogenous carbon added during the assay (Fig 4.2).

The data show that when the separated wild-type extract is incubated with sorbitol a distinct band is found, whereas when it is incubated with either mannitol or D-arabitol a common band of activity is detected that is separate and distinct from what is observed when incubated with sorbitol (Fig 4.2, lane 1). All of these bands of activity are absent in extracts generated from the *smoS* mutant strain SMc01500 (Fig. 4.2, lane 2). Introduction of either pSMc01500, which carries a wild-type copy of *smoS*, into the mutant strain, or pJD02, which carries a copy of the wild-type *mtlK*, restored the presence of either a sorbitol (Fig. 4.2, lane 3) or mannitol/D-arabitol (Fig. 4.2, lane 4) oxidizing protein within the extract. These bands of activity mimic the intensity and migration distance of the wildtype bands. Taken together, these data suggest that *smoS* encodes an NAD^+ dependent sorbitol and galactitol dehydrogenase while *mtlK* encodes an NAD^+ dependent mannitol and D-arabitol dehydrogenase.

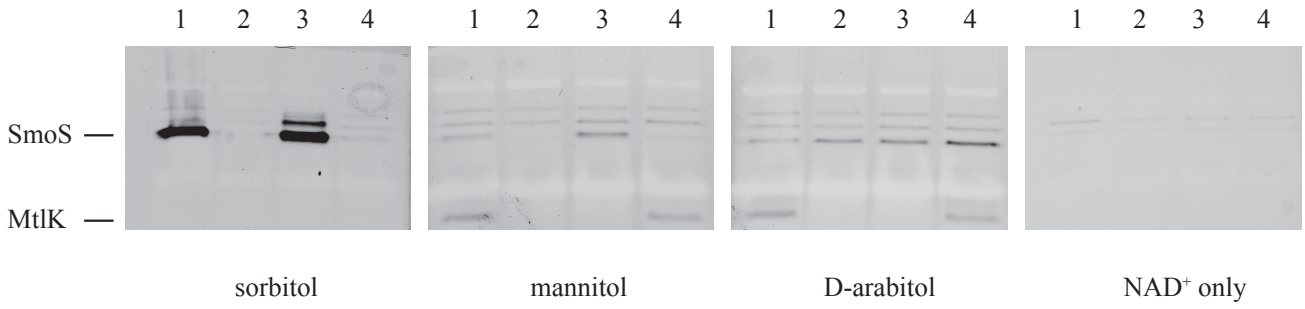


Figure 4.2. Non-denaturing PAGE gel of sorbitol inducible dehydrogenase activity. Extracts from wildtype (lane 1), SMc01500 (lane 2), SMc01500 with *smoS in trans* (lane 3) and SMc01500 with *mtlK in trans* (lane 4) were grown in glycerol and induced with sorbitol for six hours. Gels were stained for dehydrogenase activity using an assay reagent containing the polyol and cofactor listed below each panel.

4.4.3 Sorbitol and D-arabitol strongly compete with mannitol for transport

It has been previously shown that multiple polyols are all capable of being transported by the multiple polyol transporter encoded by *mptABCDE* in *S. meliloti* (Geddes & Oresnik, 2012b). Additionally, it was shown that the transport of mannitol was inducible by mannitol, and that galactitol could directly compete with mannitol transport (Kohlmeier *et al.*, 2019). Our results suggest that the *smo* locus encodes determinants necessary for the transport and the conversion of sorbitol, mannitol, and D-arabitol to central carbon intermediates-namely fructose and D-xylulose respectively.

To determine the extent to which the SmoEFGK transport system contributes to sugar alcohol uptake, we directly tested the ability of SRmD495, a strain carrying an insertion mutation in the ATP-binding protein of the ABC transporter to take up ^{14}C labelled mannitol. The wildtype, Rm1021, as well as the SRmD495 were grown in defined medium with glucose and mannitol as carbon sources. The strains were subsequently assayed for their ability to accumulate radiolabelled mannitol. The results show that SRmD495 cells were severely compromised in their ability to transport mannitol in comparison to the wildtype (Fig. 4.3A), suggesting that the SmoEFGK transport system is involved in the uptake of mannitol.

To determine if sorbitol, and D-arabitol were capable of being taken up by the same transporter, these sugar-alcohols were competed against labelled mannitol in transport assays (Fig. 3B). The data show that when sorbitol or D-arabitol were mixed with ^{14}C-mannitol in equal amounts, labeled mannitol uptake rates were reduced by approximately 50% (Fig. 4.3B). If the concentration of unlabeled sorbitol or D-arabitol is in great excess of the ^{14}C-mannitol concentration (5:1), uptake rates are nearly abolished (Fig. 4.3B).

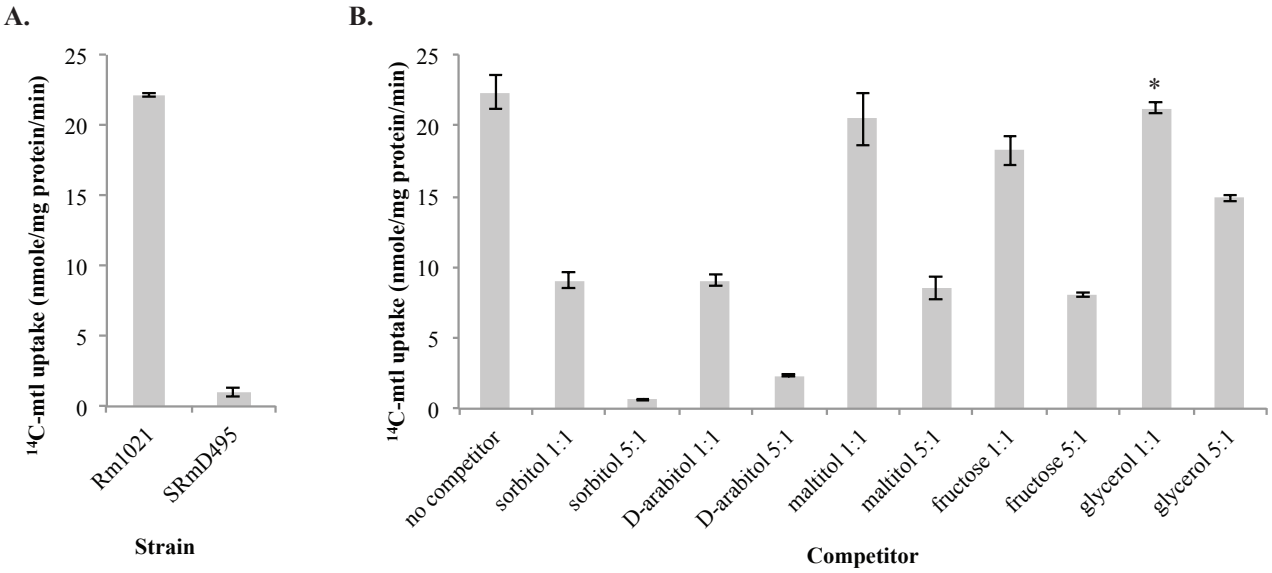


Figure 4.3. Transport rates of ^{14}C-mannitol. (A) Rm1021 (wildtype) and SRmD495 (*smoK*::Tn*5*-B20) grown in mannitol and glucose, 2 µM labeled mannitol was used to initiate the assay. (B) Uptake of ^{14}C-mannitol by Rm1021 in competition with unlabelled substrates. Cells were grown on mannitol as a sole carbon source. 2µM labelled mannitol was competed against either 2µM or 10µM unlabelled substrate. Accumulation of label is shown in nmole/mg protein/min. Data are expressed as the mean ± SD of three independent replicates. * represents data are an average of only two independent replicates.

Maltitol is a disaccharide made up of sorbitol and glucose. Since the loss of *smoS* or the associated transporter resulted in lower growth with maltitol (Table 4.3), it was of interest to determine if maltitol could also compete against labelled mannitol for transport. In addition, since it had been shown that mannitol was capable of inducing the fructose ABC transporter (Lambert *et al.*, 2001), and that both mannitol and sorbitol are ultimately converted to fructose, these substrates, as well as glycerol, which had the same stereochemistry, were also competed against mannitol. The results show that with both maltitol and fructose at a 1:1 ratio there was a slight reduction in mannitol transport. Whereas the maltitol data are not statistically significant, the fructose decrease was significant ($p = 0.03$). There was no visible reduction in transport of mannitol using glycerol at this ratio (Fig. 4.3B). When the ratio of these substrates was increased to be 5:1, it was found that both maltitol and fructose decreased transport rates of mannitol to less than 50% of the wildtype rate, whereas the glycerol reduction was approximately 30% (Fig. 4.3B).

4.4.4 The smo *locus is negatively regulated by SmoC*

The *smo* locus contains a putative regulator annotated as *smoC* (*SMc01495*) that seems likely to be involved in the regulation of the operon (Fig. 4.1A). A comparison of the SmoC amino acid sequence to other proteins encoded by the Rm1021 genome revealed a sequence similarity (E value of $1x10^{-56}$) to EryD, a DeoR type regulator. Regulators of the DeoR family typically negatively regulate the genes under their control (Saxild *et al.*, 1996). However an *S. meliloti* strain carrying an *eryD* mutation was unable to grow using erythritol as a sole carbon source and, based on qRT-PCR data, was consistent with it acting as both a positive and a negative regulator of erythritol catabolism (Geddes *et al.*, 2010).

To determine how SmoC effects the transcription of the *smo* operon, a *smoC* mutant, SRmD641, was constructed. Consistent with the hypothesis that SmoC was a negative regulator, SRmD641 was found to be able to grow using sorbitol, mannitol, and D-arabitol. To determine if it affected the transcription of the *smo* operon, a strain with a reporter gene transcriptionally fused to the operon was identified from an *S. meliloti* fusion library (Cowie *et al.*, 2006). This strain, FL4643, has a *gfp*$^{+}$ gene fused to the 3′ end of *mtlK*. To determine if SmoC could regulate this operon the mutant *smoC* allele was transduced into a FL4643 background, generating strain SRmD642.

The results show that expression of gfp increased when FL4643 cells were grown in the presence of *smo* inducer substrates such as sorbitol (Fig. 4.4), consistent with previous results suggesting that transport was inducible by mannitol (Kohlmeier *et al.*, 2019). *gfp*$^{+}$ expression of SRmD642 was greater than double sorbitol induced FL4643 regardless of the growth conditions (Fig. 4.4). Taken together, the data suggest that SmoC functions as a negative regulator, and that mutations to *smoC* result in constitutive expression of the *smo* operon. Additionally, the data also show that *smoC* is independently transcribed from the rest of the *smo* operon, given that disruption of *smoC* did not have polar effects on the downstream genes (Fig. 4.1A).

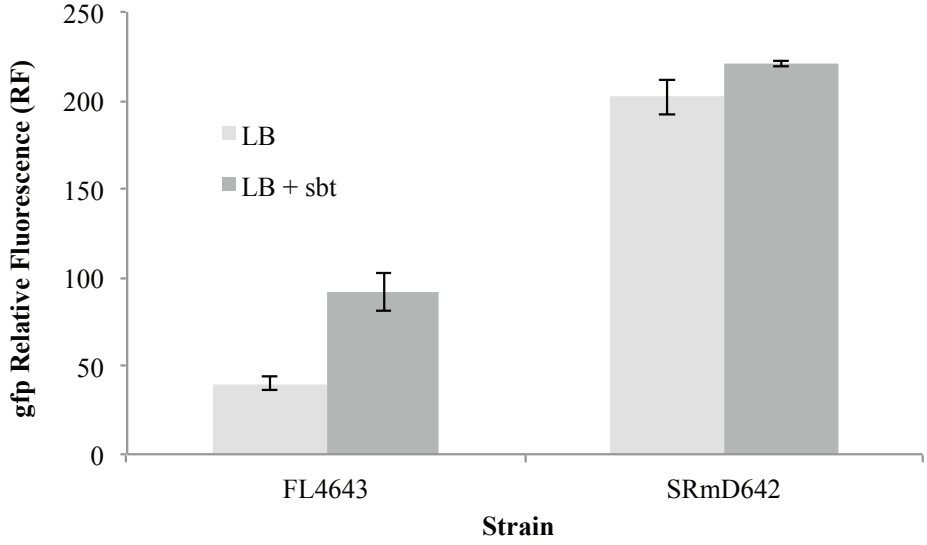


Figure 4.4. Induction of the *smo* locus measured by GFP relative fluorescence (RF). GFP was read at wavelengths of 485 (ex) and 510 (em). RF=(fluorescence-background)/OD600, Rm1021 was used as a background.

4.4.5 frk *encodes a fructose kinase*

The oxidation of sorbitol and mannitol by SmoS and MtlK respectively generate fructose. To enter central carbon metabolism fructose needs to be phosphorylated to become fructose-6-phosphate (F6P), which is then subsequently converted to glucose-6-phosphate in *S. meliloti* by phosphoglucose isomerase (Arias *et al.*, 1979; Gardiol *et al.*, 1980; Geddes & Oresnik, 2014). Whereas enzyme activities and mutants in fructokinase and phosphoglucose isomerase have been previously reported in *S. meliloti* strain L5-30, these were mutants that were generated by chemical mutagenesis prior to the ability to acquire whole genome sequences (Arias *et al.*, 1979; Gardiol *et al.*, 1980). To provide clarity to the genome annotation of *S. meliloti* Rm1021, as well as for characterizing a complete catabolic pathway for sorbitol and mannitol, it was decided to unambiguously ascribe the encoded function to the proper gene(s).

Within the genome annotation there is some ambiguity regarding the gene responsible for encoding the fructose kinase in *S. meliloti*. A region, called the *frc* locus (Fig. 4.1B), has been previously characterized and determined to encode a transporter responsible for the uptake of fructose (Lambert *et al.*, 2001). A gene encoding a putative kinase, *frcK*, was identified immediately adjacent and presumably in the same operon as *frcCA* since its open reading frame overlaps *frcA* by 3 base pairs. It is annotated as a kinase, but was determined to be an unlikely candidate as a fructose kinase based on low sequence similarity to a characterized fructokinase in *R. leguminosarum* (Fennington & Hughes, 1996; Geddes & Oresnik, 2014; Lambert *et al.*, 2001). A second putative kinase, *frk*, is located within 2 kb of the transport genes and *frcK* (Fig. 4.1B).

To resolve the ambiguity a screen for mutants unable to utilize fructose was carried out and a single insertion mutant was identified. The insert was determined to be within *frcK* and the strain carrying this mutation was called SRmD620 (Fig. 4.1B). The *frk* mutation was constructed using pKNOCK-Gm as described, and the strain carrying this mutation was designated SRmD616. The results show that a mutation in *frcK* reduces growth on sorbitol, mannitol, and fructose, but not on glycerol (Table 4.5). However, mutations in *frk* exhibit a more severe growth deficiency, in which growth is completely abolished, similar to a strain carrying a s*moS* mutation with respect to sorbitol and mannitol (Table 4.5). Introduction of pMK38, which carries a wildtype copy of *frk*, into either a strain carrying a mutation in *frk* (SRmD616) or *frcK* (SRmD620) was able to fully restore growth of these mutants on all tested carbon sources to wildtype levels, while introduction of pMK39 (carrying *frcK*) into these mutations did not show a growth improvement over the empty vector control (Table 4.5).

The ability to restore growth to strains carrying either *frcK* or *frk* strongly suggests that *frk* encodes a fructokinase (Table 4.5). Therefore, we initially suspected that the growth deficiency exhibited by *frcK* mutants was the result of polar effects on *frk*. However, the *pyrCE* genes found between *frcK* and *frk* are likely to contribute to pyrimidine biosynthesis and mutations in these genes are reported to be auxotrophic (Fig. 4.1B) (diCenzo *et al.*, 2018; Randhawa & Hassani, 2002; Vineetha *et al.*, 2001). Therefore, it is most likely that *frk* is transcribed separately from *frcK*. To provide evidence for the role of Frk we wished to demonstrate the loss of fructokinase activity in SRmD616. The ability to measure carbohydrate kinase activities in crude extracts can be confounded by high background rates of NADH oxidation (Kohlmeier *et al.*, 2019). As an alternate approach, Frk was N-terminally 6xHis tagged, and expressed from pRK7813 in *S. meliloti* as previously described (Richardson *et al.*,

2008). The cells were then harvested and the protein was partially purified using Ni-affinity column. It was found that when the partially purified Frk was assayed for kinase activity using fructose as a substrate it yielded a rate of 17.1 µmol/min/mg protein (Fig. 4.5). In contrast if glucose was as a substrate to initiate the assay a rate of 2.4 µmol/min/mg were measured which was approximately equivalent to the background rate of NADH oxidation (Fig. 4.5).

4.4.6 Phosphoglucose isomerase gene is downstream of the **frc** *locus*

There are three genes annotated as encoding a glucose-6-phosphate isomerase in the *S. meliloti* genome; *SMc02042* (*pgiA1*), *SMc02163* (*pgi*), and *SMb20857* (*pgiA2*). Of the three, the most likely candidate of these has been assumed to *pgi* (*SMc02163*). This is based on bioinformatic analysis (Geddes & Oresnik, 2014), as well as the results of genome wide Tn-seq guided *in silico* metabolic reconstruction (diCenzo *et al.*, 2018). To provide experimental evidence to support this hypothesis, a strain, SRmD654, carrying an insertional mutation in *pgi* was constructed. A mutation in *pgi* results in a strain that is consistent with previously reported carbon phenotypes (Arias *et al.*, 1979), chiefly it is unable to grow using sorbitol, mannitol, and fructose as a sole carbon source (Table 4.5). The introduction of pMK62, which carries the wild-type *pgi* expressed from a broad-host range plasmid, results in complementation of this mutant (Table 4.5).

Table 4.5. Complementation of putative fructose mutants

Strain	Characteristic	LB	sbt	mtl	fru	gly
Rm1021	wt	+	+	+	+	+
SMc01500	*smoS*::pTH1703	+	-	-	+	+/-
SRmD616	*frk*::pKNOCK-Gm	+	-	-	-	+
pMK38	*frk in trans*	+	+	+	+	+
pMK39	*frcK in trans*	+	-	-	-	+
SRmD620	*frcK*::Tn*5*	+	+/-	+/-	+/-	+
pMK38	*frk in trans*	+	+	+	+	+
pMK39	*frcK in trans*	+	+/-	+/-	+/-	+
SRmD654	*pgi*::pKNOCK-Gm	+	-	-	-	nd
pMK62	*pgi in trans*	+	+	+	+	nd

Growth; +, like wildtype; -, no growth; +/-, intermediate growth. Abbreviations; LB, Luria-Bertani; sbt, sorbitol; mtl, mannitol; fru, fructose; gly, glycerol; nd, not determined. Experiments with the *pgi* mutant used glucose as a neutral carbon source to support growth on defined media.

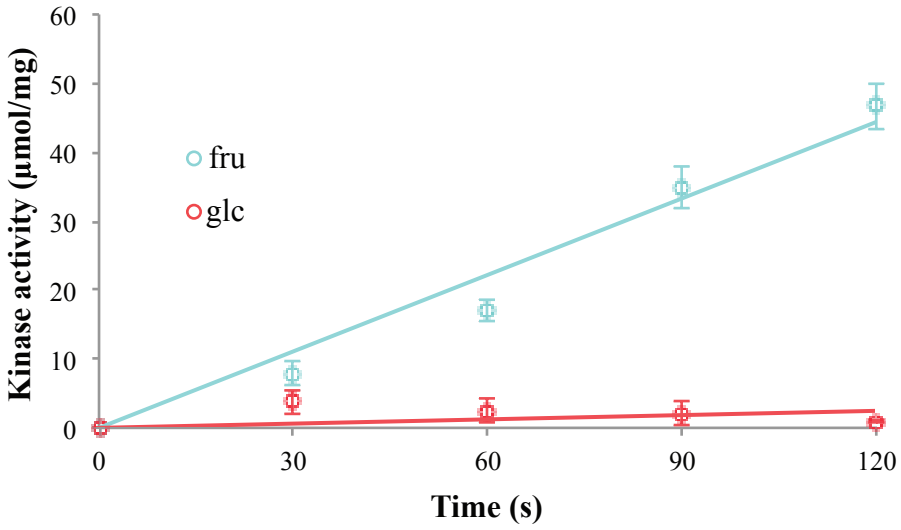


Figure 4.5. Fructose kinase activity of partially purified Frk. Activity was coupled to the production of NAD+ by pyruvate kinase (PK) and lactate dehydrogenase (LDH) and was measured at 340 nm for two minutes in a buffer containing 60 mM HEPES pH 7.5, 6 mM $MgCl_2$, 3 mM ATP, 3 mM PEP, 0.3 mM NADH, and 1/50 vol PK/LDH mix.

4.4.7 Sorbitol and mannitol can be transported by the frc *transporter*

The observation that a strain carrying a mutation in *frcK* affected but did not abolish the growth on defined medium using sorbitol, mannitol, or fructose as a carbon source is puzzling since from the annotation and bioinformatic analysis of *frcK* it is difficult to discern the exact role this gene encodes. *frk* has definitively been shown to have fructose kinase activity (Fig. 4.5), and complementation analysis supports this assertion (Table 4.5). Based on its proximity to transport genes *frcBCA*, we suspected that the growth deficiency exhibited by *frcK* mutants might be the result of reduced ability to transport fructose, sorbitol, and mannitol. That complementation of *smo* transport mutants with *smoS* or *mtlK* can restore growth on sorbitol and mannitol suggests that there is another transporter capable of taking up these substrates (Table 4.3). The inhibition of mannitol transport in the presence of excess fructose supports this hypothesis (Fig. 4.3B). To determine if the fructose transporter plays a role in sorbitol and mannitol uptake mutations were constructed in *frcA* (SRmD644), as well as *frcC* (SRmD645). In addition, the *smoK*::Tn*5*-B20 allele from SRmD495 was transduced into strains carrying either, *frcA* or *frcC* yielding SRmD666 or SRmD667 respectively.

Strains carrying mutations in *frcC*, *frcA*, or *frcK* showed a decrease in growth on fructose, mannitol, and sorbitol (Tables 4.5 and 4.6). A strain carrying a mutation in *smoK* was not impaired for growth using fructose as a carbon source, but was unable to efficiently utilize sorbitol or mannitol (Table 4.6). In contrast, strains carrying mutations in both transporters were completely unable to utilize fructose, sorbitol, or mannitol (Table 4.6). The growth phenotypes suggest that both transporters play a role in the transport of sorbitol, mannitol, and fructose into the cell. Complementation of the double transport mutants with *smoS* or *mtlK* did not permit

growth on sorbitol or mannitol, suggesting that the *frc* transporter is the second transporter capable of sorbitol and mannitol uptake (Table 4.6).

4.4.8 FrcK contributes to fructose uptake and mutations in **frcK** *or* **frk** *permit transport of mannitol under non-inducing conditions*

To directly assess the role of each transporter in taking up either mannitol or fructose a series of uptake assays was carried out. First we sought to determine how these transporters were regulated. Previous results have shown that transport of mannitol is inducible, cells grown in mannitol exhibited ^{14}C-mannitol uptake rates of 25.8 nmoles/mg protein/min compared to 2.8 nmoles/mg protein/min for cells grown in glucose (Kohlmeier *et al.*, 2019). To test for induction by fructose, cells were grown overnight in defined medium with fructose and assayed for uptake of ^{14}C-mannitol. These cells took up labeled mannitol at a rate of 2.0 nmoles/mg protein/min, suggesting that growth on fructose does not induce the transport of mannitol. We subsequently tested for the uptake of ^{14}C-fructose by cells grown in fructose, mannitol, or glucose, which exhibited uptake rates of 111.9 nmoles/mg protein/min, 95.1 nmoles/mg protein/min, and 16.3 nmoles/mg protein/min respectively, indicating that growth in fructose or mannitol induces the uptake of fructose. Induction of *frcC* by mannitol has been previously reported (Lambert *et al.*, 2001).

Table 4.6. Carbon phenotypes of putative fructose transport mutants

Strain	Characteristic	LB	sbt	mtl	fru	glc
Rm1021	wt	+	+	+	+	+
SRmD495	*smoK*::Tn*5*-B20	+	+/-	+/-	+	+
SRmD664	*frcA*::pKNOCK-Gm	+	+/-	+/-	+/-	+
SRmD665	*frcC*::pKNOCK-Gm	+	+/-	+/-	+/-	+
SRmD666	*smoK*::Tn*5*-B20, *frcA*::pKNOCK-Gm	+	-	-	-	+
smoS	*smoS in trans*	+	-	-	-	+
mtlK	*mtlK in trans*	+	-	-	-	+
SRmD667	*smoK*::Tn*5*-B20, *frcC*::pKNOCK-Gm	+	-	-	-	+
smoS	*smoS in trans*	+	-	-	-	+
mtlK	*mtlK in trans*	+	-	-	-	+

Growth scored as follows; +, like wildtype; -, no growth; +/-, intermediate growth. Abbreviations are as follows; LB, Luria-Bertani; sbt, sorbitol; mtl, mannitol; fru, fructose; glc, glucose.

Next, we wondered if any sugar alcohol substrates could compete with fructose for use of the *frc* transporter. Cultures of Rm1021 that were grown in fructose were measured for ^{14}C-fructose uptake in the presence of unlabeled fructose, mannitol, sorbitol, D-arabitol, and glycerol in equivalent amounts or in 5-fold excess. The results show that unlabeled fructose present in a 1:1 ratio with the radiolabel is able to reduce ^{14}C-fructose uptake to approximately 65% of the uncompeted control (Fig. 4.6A). If unlabeled fructose is present at a ratio of 5:1, the uptake of ^{14}C-fructose is reduced to 25% (Fig. 4.6A). However, none of the sugar alcohol substrates tested were able to reduce the accumulation of ^{14}C-fructose regardless of the ratio at which they were present (Fig. 4.6A). This suggests that sugar alcohols do not utilize the *frc* transporter under the conditions of this assay.

Finally, the transport of ^{14}C-fructose and ^{14}C-mannitol by different mutant strains was assayed to ascertain the role that these genes play in the transport of fructose and mannitol. Special attention was paid to *frcK* to determine if that gene encodes a transport function. Many of the mutants tested grow poorly or not at all using fructose as a sole carbon source, so, glucose was added along with fructose to support the growth of these strains. Importantly, transport rates of ^{14}C-fructose by the Rm1021 wildtype were high, indicating that growth in the presence of glucose was not inhibiting the uptake of labeled fructose (Fig. 4.6B). Both the *smoK* and the *frk* mutants are able to transport fructose at wildtype levels, which was consistent given that the *frc* transport genes are unaffected in these strains (Fig. 4.6B). The *frcK* mutant exhibited an approximately 50% reduction in ^{14}C-fructose transport rates, suggesting that FrcK does play a role in transport of fructose (Fig. 4.6B). Additionally, ^{14}C-transport rates of the *frcA* and *smoK/frcA* strains are barely above background, indicating that both strains are compromised in their ability to transport fructose (Fig. 4.6B).

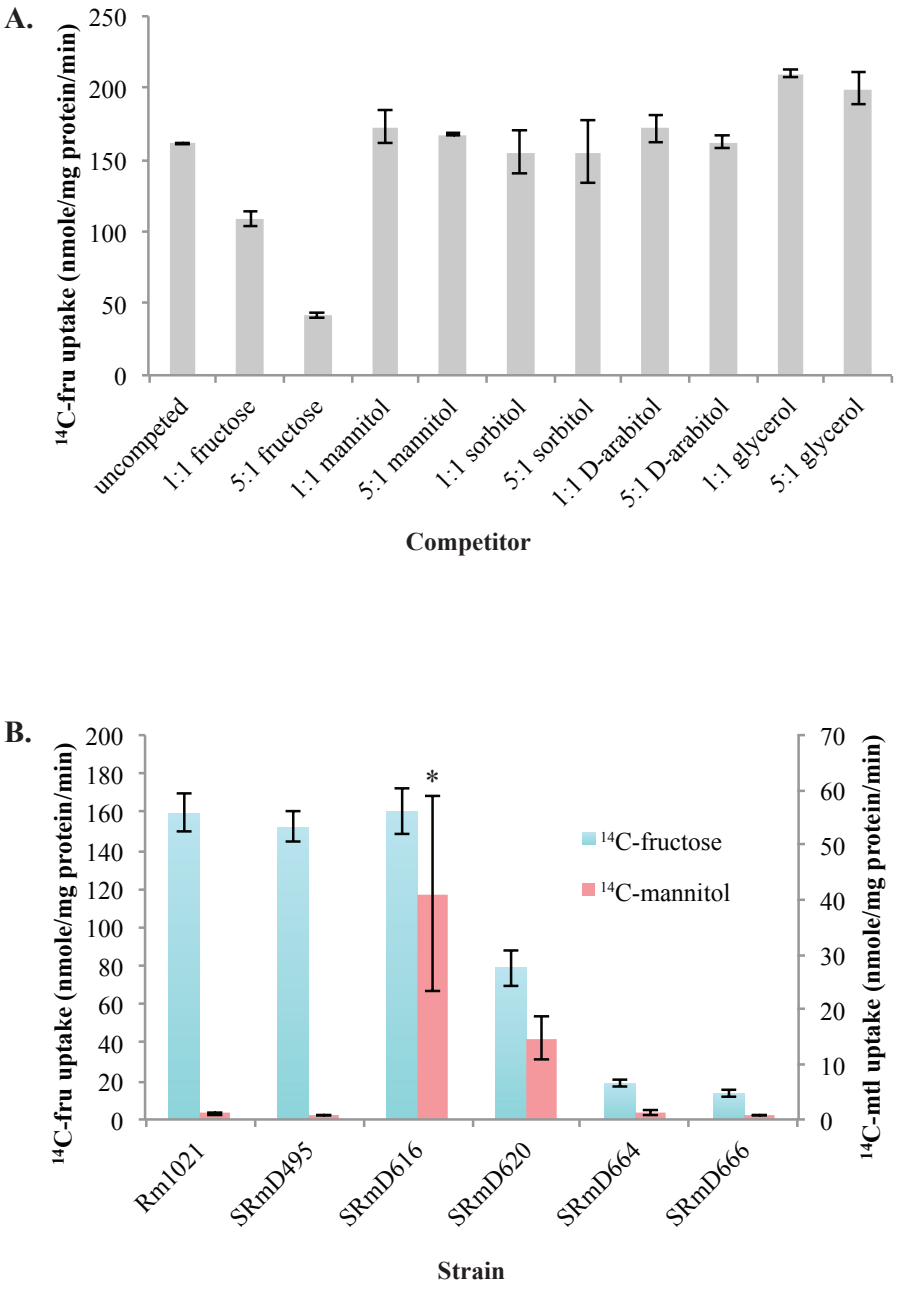


Figure 4.6. Transport rates of ^{14}C-fructose and ^{14}C-mannitol. (A) Rm1021 in competition with unlabelled substrates. Cells were grown on fructose as a sole carbon source. 2µM labelled mannitol was competed against either 2µM or 10µM unlabelled substrate. (B) Uptake of ^{14}C-fructose (blue) or ^{14}C-mannitol (red) by various transport mutants. Cells were grown in fructose and glucose, 2 µM radiolabelled substrate was used to initiate the assay. Accumulation of label is shown in nmole/mg protein/min. Data are expressed as the mean ± SD of three independent replicates. * represents data are an average of only two independent replicates.

Rm1021 exhibited poor transport rates of ^{14}C-mannitol following growth on fructose and glucose (Fig. 4.6B), which is consistent with the lack of transport activity reported earlier suggesting that mannitol uptake is not inducible by fructose. The *smoK*, *frcA*, and *smoK/frcA* mutants showed ^{14}C-mannitol transport rates similar to wildtype (Fig. 4.6B). Interestingly, the *frk* and *frcK* strains showed ^{14}C-mannitol transport rates well over background (Fig. 4.6B), *frk* rates are approximately equivalent to mannitol grown wildtype cells and *frcK* rates are approximately 50% of wildtype. That these strains exhibit mannitol uptake in a non-induced (fructose grown) state indicates that they are able to overcome repression of mannitol transport genes.

4.5 Discussion

In this work we show that *smoS* and *mtlK* encode determinants necessary for the oxidation of sorbitol and mannitol to fructose (Fig. 4.2). In addition, MtlK is shown to be able to use D-arabitol as a substrate yielding D-xylulose (Fig. 4.2). Whereas the fructose that is generated by either SmoS or MtlK is dependent upon the gene products of *frk* and *pgi* to be integrated into central metabolism, D-xylulose is dependent on *xylB* (Fig. 4.7) (Geddes & Oresnik, 2012b). Additionally fructokinase activity has been ascribed to Frk (Fig. 4.5), and experimental evidence has been provided for the *pgi* (Table 4.5).

The catabolic pathways for sorbitol and mannitol are characterized and straight forward, the transport of these compounds is less clear. ABC type transporters are typically characterized as being tripartite in nature containing a periplasmic solute binding protein, permease components, as well as an ATP binding protein. These transporters are generally regarded as high affinity transporters with K_Ms in the micromolar range and consequently often are used to transport a single solute (Walshaw *et al.*, 1997).

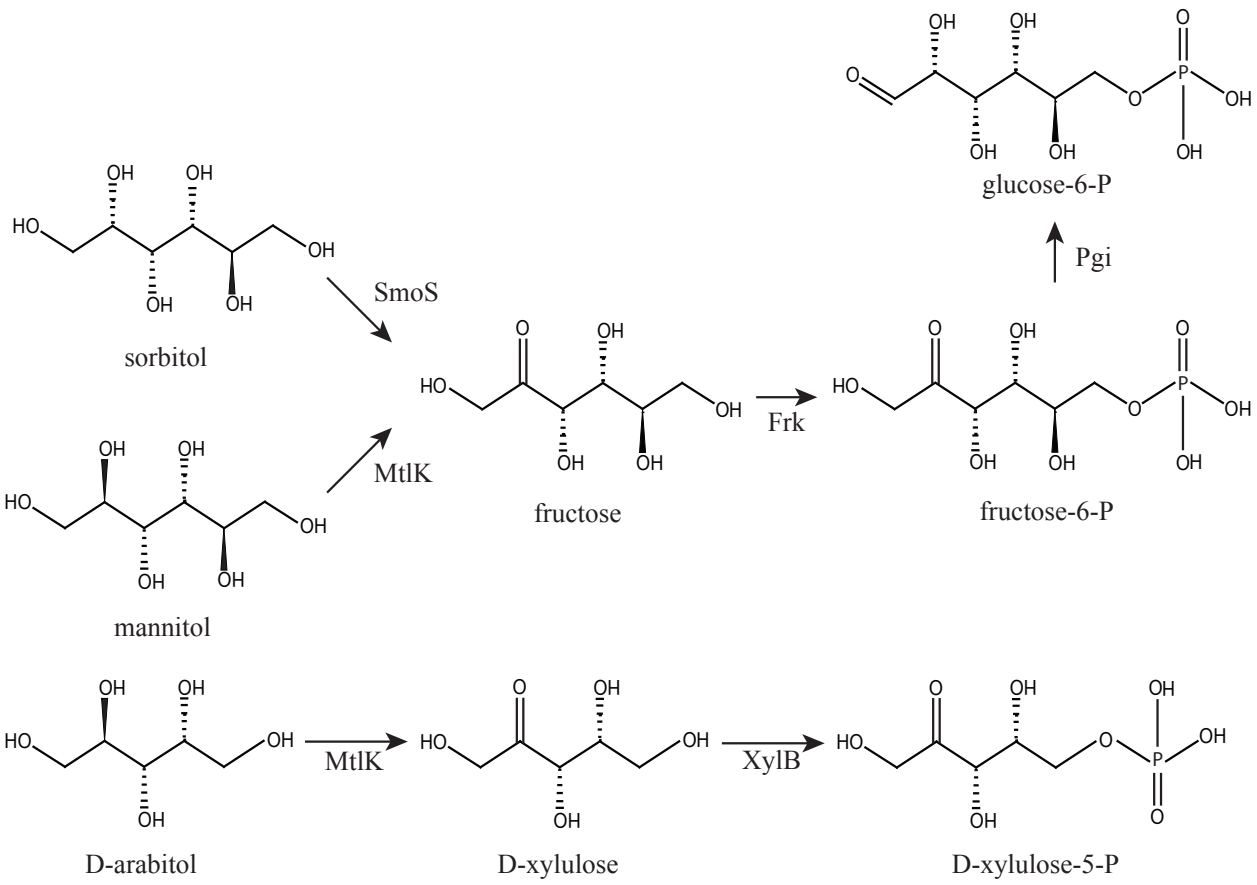


Figure 4.7. Pathways for the metabolism of sorbitol, mannitol, and D-arabitol. Sorbitol and mannitol are oxidized to fructose by SmoS and MtlK respectively, before phosphorylation into fructose-6-phosphate by Frk and isomerization by Pgi. D-arabitol becomes D-xylulose, catalyzed by MtlK, which is subsequently phosphorylated by XylB into D-xylulose-5-phosphate.

Whereas this is often the case, it is noteworthy that several *S. meliloti* transporters have been reported to transport multiple substrates including erythritol, ribitol, and L-arabitol by MptABCDE (Geddes & Oresnik, 2012b), as well as transport of galactose, glucose, and L-arabinose through AraABC (Geddes & Oresnik, 2012a).

The evidence for sorbitol and mannitol utilizing a second transporter for uptake is based on the fact that; if a multicopy plasmid containing either *smoS* or *mtlK* is introduced into a strain carrying a *smo* transporter mutant, it is capable of restoring growth (Table 4.3), a strain carrying a *frcK* mutation exhibits slow growth on fructose, mannitol, and sorbitol that can be reversed by introducing *frk* on a multicopy plasmid (Table 4.5), and finally strains carrying mutations in both the *smo* and *frc* transporters are completely unable to grow on sorbitol, mannitol, or fructose (Table 4.6).

Although taking all of these data together makes a strong case that both transporters can transport all of the listed substrates, we do not believe that they transport all of the substrates with equivalent efficiency. In all cases where a multicopy plasmid was introduced, it is very likely that the overall equilibrium of the pathway might be shifted to favor catabolism, as previously demonstrated (Richardson *et al.*, 2008). In addition the fructose transporter has been shown to have an apparent K_M of 6 μM, which is consistent with it being a high affinity transporter, and although fructose can inhibit mannitol transport, it was far less competitive than sorbitol, mannitol, or D-arabitol (Fig. 4.3B). Finally, a strain carrying a mutation in the *smo* transporter had a growth rate equivalent to that of the wild type when grown on defined medium with fructose (Tables 4.3 and 4.6). These data suggest that although fructose can utilize the *smo* transporter, it is more likely that it behaves as a low affinity fructose transporter that is evident only when conditions are manipulated to allow this to happen.

In contrast the transport of mannitol (probably sorbitol and D-arabitol as well) is more complicated. The fructose locus of *S. meliloti* has been suggested to contain two differentially regulated promoters, one upstream of *frcB* (which encodes the solute binding protein) that is expressed constitutively (Mauchline *et al.*, 2006), and a second upstream of *frcC* (permease) which is inducible by mannitol and fructose (Lambert *et al.*, 2001). In addition, transport of mannitol as well as the transcription of the operon has been demonstrated using mannitol, sorbitol, galactitol, maltitol and D-arabitol (Fig. 4.3) (Mauchline *et al.*, 2006). Taken together with the reduced growth on sorbitol and mannitol of strains carrying mutations in *frcC*, *frcA*, and *frcK* (Table 4.6), and the inability for strains carrying mutations in both the *frc* and *smo* transporters to grow using sorbitol and mannitol (Table 4.6), it is likely that both of these transporters contribute to the uptake and growth on these substrates. However, we observed an inability of sugar alcohols to compete with fructose for transport, which contradicts this theory (Fig. 4.6A). It is noteworthy that the cells assayed for fructose uptake were grown on fructose. Growth on mannitol may eliminate a regulatory hindrance to polyol uptake that is imposed by fructose, and allow for simultaneous transport of fructose with sugar alcohols.

We note that whereas the three independent mutations in the fructose locus that are described in this work consistently exhibit reduced growth on fructose, sorbitol, and mannitol, the insertion at this locus that has been described previously is reported to be completely unable to utilize fructose as a sole carbon source and does not have a deficiency for growth on defined medium containing mannitol (Lambert *et al.*, 2001). Although there are differences in the methods of mutagenesis, the source of the discrepancies between the reported growth phenotypes is not clear.

Mannitol is a traditional carbon source that has been used for the cultivation of Rhizobia, and other soil bacteria, as well as the isolation of exopolysaccharides (Breedveld *et al.*, 1993; Stowers, 1985). Although the catabolism of sorbitol and mannitol differs by only the enzymes that are used to oxidize them to fructose, it has been observed that when Rm1021 is streaked out on defined medium containing mannitol as a sole carbon source, it appears more mucoid than the same culture grown using sorbitol as a carbon source. Since it appears that unlabeled sorbitol is as effective as unlabeled mannitol at competing with ^{14}C-mannitol in transport assays, it is reasonable to assume that the transport of sorbitol and mannitol through the *smo* transporter have approximately equivalent rates (Fig. 4.3B) (Kohlmeier *et al.*, 2019). This implies that the difference in the visible mucoidy phenotype observed may be directly attributable to the kinetic differences of SmoS and MtlK, thus possibly affecting the internal metabolic pool sizes of fructose and/or the hexose phosphates (fructose-6-phosphate and glucose-6-phosphate). This correlates well with previous regression analysis that were carried out with metabolite pools using transaldolase and transketolase mutants that reported that there was a positive relationship between exopolysaccharide production and hexose phosphate pools (Hawkins *et al.*, 2018), as well as the observations made when Rm2011 was grown under acidic conditions (Omar Draghi *et al.*, 2017).

That *frk* and *frcK* mutants exhibit transport of mannitol despite growth on fructose may also support the role of F6P pools in regulation (Fig. 4.6B). While FrcK shares little sequence similarity with known kinases, it is annotated as a kinase and has a motif associated with phosphate binding, known as a P-loop or Walker A motif, encoded in its amino acid sequence. If we assume the FrcK contributes to phosphorylation of fructose, along with Frk, the transport of mannitol by SRmD616 and SRmD620 could be the result of reduced ability to generate F6P.

Therefore, internal F6P pools may inhibit uptake of sugar alcohols in the wildtype when sufficient amounts of the metabolite are available. We showed evidence that FrcK contributes to fructose transport (Fig. 4.6B), and transport associated kinases have been previously described in rhizobia (Rivers & Oresnik, 2013; Rivers & Oresnik, 2015). The affect of metabolites on carbon flux and partitioning has not been systematically studied in *S. meliloti* to date.

Chapter 5:

Metabolism of D-arabinose and L-fucose in *Sinorhizobium meliloti* 1021

This work was carried out by MacLean Kohlmeier in collaboration with Derek D. Kim. DK constructed the Δ*SMb21111* mutant strain.

5.1 Abstract

The ability to utilize different carbon sources is important for diazotrophic bacteria to form an effective symbiosis with their host plants. During symbiosis, soil bacteria invade host tissues forming organs on the plant roots called nodules where they fix nitrogen into a form that can be used by the plant for growth. The bacteria, termed rhizobia, benefit from this relationship by receiving carbon from the plant; in this way rhizobia avoid having to grow saprophytically in the highly competitive soil environment and are nourished by the plant. Bacterial strains that are unable to catabolize certain carbon sources are deficient in their ability to compete for nodule occupancy or establish symbiosis. However, the catabolic pathways for many substrates as well as their relationship to symbiosis have yet to be elucidated. In this study, we use genetic techniques to identify the genes involved in D-arabinose and L-fucose utilization and define the enzymatic reactions that breakdown these substrates in the model organism *Sinorhizobium meliloti*. D-arabinose is metabolized through a forked pathway involving genes found on a chromosomal locus previously associated with polyol metabolism as well as a set of genes found on a distinct replicon called pSymB. The pSymB genes encode proteins that can metabolize L-fucose in addition to D-arabinose using a diketo-hydrolase type pathway. Our results enrich the knowledgebase of the metabolic capacity of the model organism *S. meliloti* and further efforts to engineer symbiosis in other organisms.

5.2 Introduction

Sinorhizobium meliloti is a Gram-negative α proteobacterium with the ability to form a nitrogen fixing symbiosis with leguminous plants. Mutations resulting in the inability of this bacterium to metabolize certain carbon sources have been linked to deficiencies in competition for symbiosis as well as nitrogen fixation. For this reason, pathways for the utilization of carbon compounds have been well studied in *S. meliloti*, however much of this organism's metabolic potential has yet to be fully described.

During investigations into sugar alcohol metabolism it was observed that certain *S. meliloti* mutants unable to utilize sorbitol and mannitol also exhibited reduced growth on medium containing D-arabinose. Curiously, D-arabinose is unlike the other substrates that elicit a growth defect from these mutants in that it contains an aldehyde group, while the other compounds are sugar alcohols with fully reduced hydroxyl groups. Little is known about D-arabinose metabolism in *S. meliloti*, although in other organisms such as *E.coli* and *Aerobacter aerogenes*, D-arabinose is broken down in part or entirely using enzymes involved in L-fucose metabolism (Elsinghorst & Mortlock, 1988; LeBlanc & Mortlock, 1971; Mortlock & Wood, 1964). L-fucose is a methylpentose also known as 6-deoxy-L-galactose. In *E. coli* K-12 it is degraded through the actions of L-fucose isomerase which forms L-fuculose, followed by L-fuculokinase producing L-fuculose-1-phosphate, and lastly L-fuculose-1-phosphate aldolase becoming dihydroxyacetone phosphate (DHAP) and L-lactate (Ghalambor & Heath, 1962; Green & Cohen, 1956; Heath & Ghalambor, 1962). Using the same enzymes, D-arabinose can become D-ribulose, followed by D-ribulose-1-phosphate, culminating in DHAP and L-glycoaldehyde (Elsinghorst & Mortlock, 1988).

In *S. meliloti*, L-fucose metabolism is also poorly understood, but the genes responsible have been tentatively localized on the pSymB chromid, specifically a putative locus designated as *SMb21103-SMb21113* (Geddes & Oresnik, 2014; Jacob *et al.*, 2008; Mauchline *et al.*, 2006; Richardson *et al.*, 2008). The genes at this locus are represented with non-specific annotations, making determination of a metabolic pathway difficult. However there is not a gene annotated, even generally, as a kinase, suggesting that L-fucose is metabolized via a non-phosphorylative pathway (Finan *et al.*, 2001). Additionally, mutations to genes encoding enzymes with triose phosphate isomerase activity are not deficient for growth on L-fucose, suggesting that DHAP is not an intermediate in the metabolism of L-fucose in *S. meliloti* (Poysti & Oresnik, 2007). Taken together, it is unlikely that metabolism of D-arabinose or L-fucose proceed via the pathways described in Enterobacteria. Therefore the purpose of this project is to define the metabolism of D-arabinose and L-fucose in *S. meliloti*.

5.3 Materials and methods

5.3.1 Bacterial strains, plasmids, and culture conditions

Bacterial strains and plasmids used in this study are listed in Table 5.1. *S. meliloti* strains were routinely grown on Luria-Bertani (LB) medium (Cold Spring Harbor Protocols, 2006) or Vincent's Minimal Medium (VMM) (Vincent, 1970) at 28°C for approximately three days. Carbon sources were added to a final concentration of 15 mM. When necessary, pH indicator bromocresol purple (BCP) was added to a final concentration of 0.025 g/L. *E. coli* strains were grown on LB medium, overnight, at 37°C. Antibiotics were included at the following concentrations; neomycin (Nm), 200 µg/mL; gentamicin (Gm), 20 or 60 µg/mL, tetracycline (Tc), 5 µg/mL, chloramphenicol (Cm), 20 µg/mL. Antibiotic concentrations were halved when

Table 5.1. Bacterial strains and plasmids

Strain or plasmid	Relevant Characteristic	Reference
Strains		
S. meliloti		
Rm1021	SU47 *str-21*	Meade *et al.* (1982)
RmP110	Rm1021 *pstC*$^{+}$	Yuan *et al.* (2006a)
Rm5000	SU47 *rif-5*	Finan *et al.* (1984)
RmG212	Rm1021 *lacZ*	Glazebrook and Walker (1989)
SMc01500	RmP110 *smoS*::pTH1703	Jacob *et al.* (2008)
RmH582	Rm5000 *gcd*::Tn*5*-233	Geddes *et al.* (2014)
SRmA241	RmG212 *SMb21105*::Tn*5*-B20	This work
SRmA280	Rm1021 *SMb21103*::Tn*5*	This work
SRmD211	Δ*pyc*	Geddes and Oresnik (2012a)
SRmD622	Rm1021 *SMb21112*::Tn*5*	This work
SRmD628	Rm1021 *SMc00680*::pKNOCK-Gm	This work
SRmD629	ΦSMc01500→SRmD622	This work
SRmD630	ΦSRmD622→SRmD628	This work
SRmD653	Rm1021 Δ*SMb21111*	This work
SRmD657	ΦRmH582→SRmD622	This work
E. coli		
DH5α	F^{-} *supE44 lacU169 hsdR17 recA1 endA1 gyrA96 thi-1 relA1* (80*lacZ*ΔM15)	Hanahan (1983)
DH5αR λpir	λpir lysogen of DH5α	House *et al.* (2004)
MM294A	*pro-82-thi-1 hsdR17 supE44*	Finan *et al.* (1986)
MT607	MM294A *recA56*	Finan *et al.* (1986)
MT616	MT607 (pRK600)	Finan *et al.* (1986)
Plasmids		
pJD02	*mtlK*/pCO37, Tc^{R}	Chapter 4
pMK46	*SMb21109*/pCO37, Tc^{R}	This work
pMK47	*SMb21111*/pCO37, Tc^{R}	This work
pMK63	*SMc00680*/pCO37, Tc^{R}	This work
pMK64	*SMb21112*/pCO37, Tc^{R}	This work
pDK28	*SMb21111* flanking regions/pJQ200SK, Gm^{R}	This work
pCO37	Gateway compatible expression vector, Tc^{R}	Jacob *et al.* (2008)
pRK600	pRK2013 *nptI*::Tn*9*, Cm^{R}	Finan *et al.* (1986)
pRK602	pRK600ΩTn*5*, Cm^{R} Nm^{R}	Finan *et al.* (1985)
pKNOCK-Gm	Suicide vector, Gm^{R}	Alexeyev (1999)
pRK7813	Broad host range vector, Tc^{R}	Jones and Gutterson (1987)
pJQ200SK	Gene replacement suicide vector, Gm^{R}	Quandt and Hynes (1993)
pXINT29	*λint* and *xis* driven by P_{lac}, Km^{R}	Platt *et al.* (2000)

added to liquid media. If more than one concentration is listed, the lower value corresponds to *E. coli* while the higher value corresponds to *S. meliloti.*

5.3.2 DNA manipulations and genetic techniques

Standard techniques were used for plasmid isolation, restriction enzyme digests, ligations, transformations, and agarose gel electrophoresis (Sambrook & Russell, 2001). Mutagenesis with transposon Tn*5* was carried out as previously described; generating strains SRmA241, SRmA280, and SRmD622 (Finan *et al.*, 1985). Strains SRmD629, SRmD630, and SRmD657 were constructed via transduction with phage ΦM12 as previously described (Finan *et al.*, 1984). Strain SRmD628 was mutagenized with suicide vector derived from plasmid pKNOCK-Gm (Alexeyev, 1999) containing an internal gene fragment from *SMc00680* PCR amplified using primers *SMc00680*_F/R (Table 5.2). To generate strain SRmD653, regions flanking *SMb21111* were cloned into pJQ200SK (Quandt & Hynes, 1993) using primers dSMb211Flank10F/R and dSMb211Flank12F/R (Table 5.2) with a Gibson assembly kit (New England Biolabs) creating pDK28. pDK28 was mobilized into Rm1021 by conjugation prior to selection for double recombinants. Mutants were screened on selective media and verified by PCR using primers dSMb11ConfF/R (Table 5.2). Plasmids pJD02, pMK46, pMK47, pMK63, and pMK64 were generated using the ORFeome Gateway system (House *et al.*, 2004; Schroeder *et al.*, 2005) using pCO37 as a destination vector as previously described (Geddes & Oresnik, 2012b).

Table 5.2. Primers used during this study

Name	Sequence 5′→3′
Generation of mutants	
*SMc00680*_pKGm_F	ATAT**CCCGGG**CTGCCGGGCCTCGTGGTG
*SMc00680*_pKGm_R	ATAT**CTCGAG**ATGATCAGGTCGAAGCTGCC
dSMb11Flank10F	CTTGATATCGAATTC**CTGCAG**ATGAGCGAGCAGACG ATC
dSMb11Flank10R	GATTGAGCGTTCATTGGATAGTTCCCTCAG
dSMb11Flank12F	TATCCAATGAACGCTCAATCAGGGAAAAGCACATG
dSMb11Flank12R	CGGCCGCTCTAGAACTAGT**GGATCC**TCAGGCGTCGG CGCCGAAC
Confirmation of mutants	
*SMc00680*_F	CATCGTCAGGGCAGTCGC
*SMc00680*_R	GTGCCGGCACCAGCGAG
dSMb11ConfF	ATGAGCGAGCAGACGATC
dSMb11ConfR	TCAGGCGTCGGCGCGAAC
Validation of constructs	
*SMb21109*_F2	ACGATTTCGGGCGGAGGTTC
*SMb21109*_R2	GGCATGGCGTTGCAGATTGG
*SMb21111*_F	CACGATGAAGGACGAGGACC
*SMb21111*_R	TCTCCTACATAATCCGCCGC
*hpaG*_F	TTCGGAAAAGACCGACTGGG
*hpaG*_R	CGAGATGATATCACCGGGGC

Restriction sites are in bold.

5.3.3 Biochemical enzyme assays

Dehydrogenase assays were performed as previously described (Geddes & Oresnik, 2012a; Kohlmeier *et al.*, 2019; Pickering & Oresnik, 2008). Briefly, strains carrying an empty vector control or complementing plasmid were grown to sufficient OD_{600} using defined medium supplemented with glycerol and L-fucose. Cultures were pelleted by centrifugation, resuspended in extraction buffer (100 mM HEPES pH 7, 2 mM $MgCl_2$, 5 mM DTT), lysed via French Press (16000 lb/in^2), and cleared of cell debris by centrifugation. Cell free lysates were separated by nondenaturing polyacrylamide gel electrophoresis and the gels were stained for dehydrogenase activity in an assay reagent containing phenazine methosulfate, *p*-nitroblue tetrazolium, NAD^+, and a substrate of interest.

5.3.4 Plant assays

Plant symbiosis assays were conducted as previously described. Briefly, alfalfa seeds were surface sterilized by washing with 70% ethanol, 1% bleach, and water. Seeds were germinated on water agar plates for two days and subsequently planted in sterile Leonard jar assemblies containing a sand and vermiculite mixture as well as nitrogen free Jensen's medium (Vincent, 1970). Plants were inoculated with strains of equivalent optical density at a ratio of 1 x 10^6 cells per plant. Plants were grown for 3-4 weeks in a growth chamber and subsequently evaluated for dry weight.

5.4 Results

5.4.1 smo *mutants exhibit reduced growth on D-arabinose which can be restored by complementation with* mtlK

It was observed through screening on various carbon sources that a strain with a mutation in *smoS* exhibits reduced growth on D-arabinose as a sole carbon source (Table 5.3). This is noteworthy as D-arabinose is an aldopentose and the *smo* locus has previously only been associated with sugar alcohol catabolism (Jacob *et al.*, 2008; Mauchline *et al.*, 2006). *smoS* mutants cannot express sorbitol dehydrogenase SmoS as well as the downstream mannitol dehydrogenase MtlK and are therefore unable to utilize substrates such as sorbitol, mannitol, galactitol, and D-arabitol for growth. D-arabinose shares significant structural similarity to mannitol and D-arabitol, and both substrates are oxidized by MtlK in *S. meliloti*, suggesting that MtlK may be capable of D-arabinose oxidation as well.

Consistent with this hypothesis, complementation of the *smoS* mutant with *mtlK* expressed from a plasmid restored growth on D-arabinose to wildtype levels, while complementation with *smoS* had no effect. Interestingly, dehydrogenase assays using extracts from sorbitol induced *S. meliloti* did not show D-arabinose dehydrogenase activity (data not shown), suggesting that MtlK cannot utilize D-arabinose as a substrate. However, it is possible that another gene or series of genes may express enzymes that act on D-arabinose to generate a metabolite that induces the *smo* locus, which could explain the observed growth deficiency.

Table 5.3. Carbon phenotypes of mutant strains

Strain	Mutated alleles	LB	D-ara	L-fuc	gly	L-ara	Nm	Gm
Rm1021	wt	+	+	+	+	+	-	-
SMc01500	*smoS*	+	+/-	+	+/-	+	-	+
SRmA241	*SMb21105*	+	+	-	+	+	+	-
SRmA280	*SMb21103*	+	+	-	-	+	+	-
SRmD211	Δ*pyc*	+	+	-	-	+	-	-
SRmD622	*SMb21112*	+	+/-	-	+	+	+	-
SRmD628	*SMc00680*	+	+*	+	+	+	-	+
SRmD629	*smoS*, *SMb21112*	+	-	-	+	+	+	+
SRmD630	*SMc00680*, *SMb21112*	+	-	-	+	+	+	+
SRmD653	Δ*SMb21111*	+	+	-	+	+	-	-

Growth; +, like wildtype; +/-, intermediate growth; -, no growth; +*, strong growth with grey color. Abbreviations; LB, Luria-Burtani; D-ara, D-arabinose; L-fuc, L-fucose; gly, glycerol; L-ara, L-arabinose; Nm, Neomycin; Gm, Gentamicin.

5.4.2 Mutations at the **SMb21103-SMb21113** *locus exhibit reduced growth on D-arabinose and no growth on L-fucose*

To identify other loci that could be involved in D-arabinose catabolism, a Tn*5* mutagenesis was carried out and recombinants were screened for their ability to grow on D-arabinose as a sole carbon source. One mutant, SRmD622, was identified from this screen and exhibited reduced growth on D-arabinose in a manner similar to observations made of *smoS* mutant strains. However, sequencing to identify the locus of the insertion revealed the Tn*5* had transposed into *SMb21112*, a gene found on the pSymB chromid, which is an entirely different replicon from the chromosomally encoded *smo* locus.

SMb21112 is part of a putative operon, *SMb21103-SMb21113*, which has been tentatively associated with the catabolism of L-fucose (Geddes & Oresnik, 2014; Jacob *et al.*, 2008; Mauchline *et al.*, 2006). Consistent with previous research, SRmD622 was unable to grow using L-fucose as a sole carbon source (Table 5.3). This operon contains a putative ABC transporter (*SMb21103-SMb21106*), two putative racemase genes (*SMb21107* and *SMb21113*), a putative mutarotase (*SMb21108*), two putative dehydrogenases (*SMb21109* and *SMb21111*), a putative dehydratase (*SMb21110*), and a hydrolase (*SMb21112*) (Fig. 5.1). While a pathway for the metabolism of D-arabinose and L-fucose was not immediately obvious from these annotations, it was clear that both the *smo* locus and the *SMb21103-SMb21113* locus contribute to the metabolism of D-arabinose.

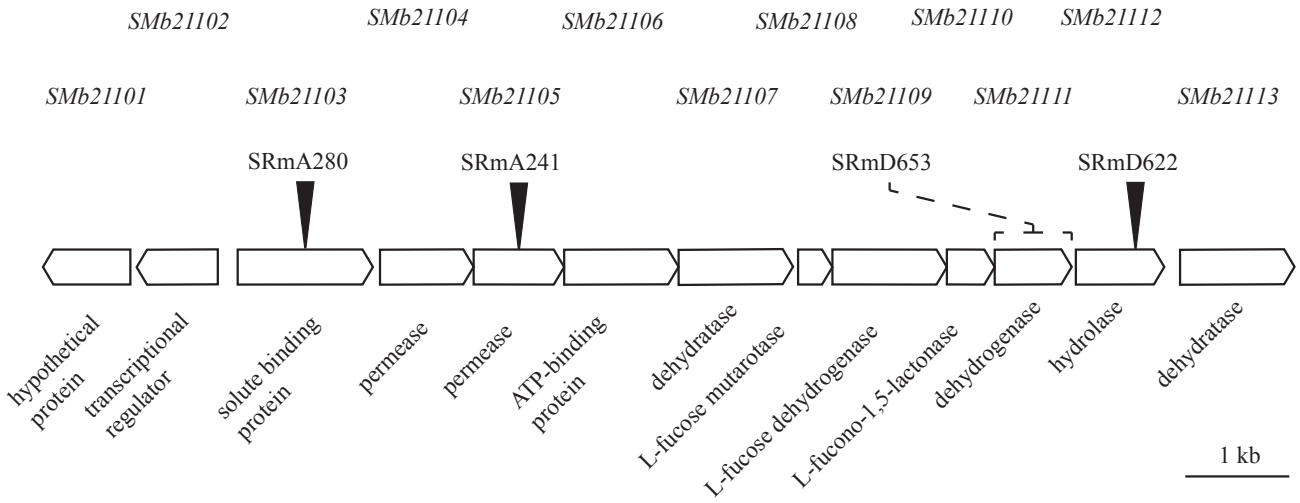


Figure 5.1. Locus diagram of the pSymB encoded *SMb21101-13* region. Locus tags appear above the genes and annotation information appears below. Black wedges represent the sites of insertional mutagenesis and the dashed line displays the removed region of a deletion mutant, strain designations are listed above these indicators.

5.4.3 SMc00680 encodes a D-arabinose reductase

A literature search for potential D-arabinose metabolic pathways revealed a route utilized by *Mycobacterium smegmatis*, in which D-arabinose is reduced to D-arabitol by an NADPH dependent reductase and subsequently oxidized into D-xylulose (Wojtkiewicz *et al.*, 1988). This pathway is an ideal scheme for *S. meliloti* as oxidation of D-arabitol is catalyzed by MtlK, an enzyme encoded from the *smo* locus (Fig. 4.7). The D-arabinose reductase gene was identified as *adhC*, an alcohol dehydrogenase also called Rv3045, from *Mycobacterium tuberculosis* H37Rv (Camus *et al.*, 2002; Cole *et al.*, 1998). A BLAST search against the *S. meliloti* Rm1021 genome using the Rv3045 protein sequence as a query identified a putative $NADP^+$-dependent alcohol dehydrogenase SMc00680 as the top hit, with an expect value of $5x10^{-149}$.

An SMc00680 mutant strain was constructed to evaluate the contribution of SMc00680 to D-arabinose metabolism. This mutant, SRmD628, grew almost as well as wildtype on defined medium with D-arabinose, but the strain seemed to be slightly grey in color (Table 5.3). Assays for D-arabinose reductase activity were conducted on extracts from wildtype, SRmD628, and SRmD628 complemented with *SMc00680* from a plasmid but no D-arabinose reductase activity could be detected from any of the extracts. That there is a visual distinction between the growth of the mutant and growth of the wildtype on defined medium with D-arabinose suggests that SMc00680 is involved in D-arabinose metabolism somehow. It is possible that the enzyme activity is present physiologically but that the assay conditions were not suitable for detection.

5.4.4 Strains with mutations at both loci are unable to grow on D-arabinose as a sole carbon source

Mutations at two distinct loci are able to reduce but not abolish growth on D-arabinose as a sole carbon source. We hypothesized that catabolism of D-arabinose may proceed through two distinct pathways, one route through MtlK encoded by the *smo* locus, and a second route utilizing the enzymes from the *SMb21103-SMb21113* L-fucose locus. If this scheme were true, a strain with mutations at both loci would be unable to grow using D-arabinose as a sole carbon source. To confirm this "metabolic fork" hypothesis, a double mutant strain with lesions at both loci was constructed via transduction. This strain, SRmD629, with mutations in *smoS* and *SMb21112*, was completely unable to utilize D-arabinose (Table 5.3). Similarly, growth on D-arabinose is abolished in an *SMc00680* and *SMb21112* double mutant background, supporting the proposed role of SMc00680 as a D-arabinose reductase (Table 5.3). These data are consistent with the "metabolic fork" hypothesis and suggests that we have identified all genetic loci that contribute to D-arabinose catabolism in *S. meliloti*.

5.4.5 **SMb21103-SMb21113** *encodes a non-phosphorylative D-arabinose and L-fucose pathway*

The catabolism of D-arabinose and L-fucose are linked in *E. coli*, in which those substrates are broken down into dihydroxyacetone-phosphate (DHAP) (Elsinghorst & Mortlock, 1988). However, *S. meliloti tpiA* and *eryH* mutants are not deficient for growth on L-fucose (Poysti & Oresnik, 2007), which would be expected if catabolism proceeded using DHAP as an intermediate. L-fucose cannot support the growth of *S. meliloti* pyruvate carboxylase (*pyc*) mutants, indicating that the ability to form oxaloacetate from pyruvate is important for the

metabolism of L-fucose (Table 5.3). That there is not a gene annotated as a kinase at the *SMb21103-SMb21113* locus, suggests that L-fucose degradation may occur through nonphosphorylated intermediates. A literature search for L-fucose catabolic pathways that meet these criteria identified a route utilized by *Burkholderia multivorans* ATCC 17616, in which L-fucose is oxidized into L-fucono-1,5-lactone before being broken down into pyruvate and lactate (Hobbs *et al.*, 2013). A similar scheme for D-arabinose catabolism was found in *Pseudomonas saccharophila*, with breakdown products of pyruvate and glycolate (Palleroni & Doudoroff, 1957). These two pathways are likely candidates for L-fucose and D-arabinose metabolism encoded by *SMb21103-SMb21113*.

Using the orthologous neighborhood viewer from IMG allowed for a detailed comparison of the genomic regions from each organism (Chen *et al.*, 2016). While the layout of the genes in the *B. multivorans* L-fucose operon is slightly different from *S. meliloti*, most of the genes are common to both organisms. The gene annotations of these regions described by IMG are non-specific, however with the updated annotation from the literature acting as a cipher (Hobbs *et al.*, 2013), the functions of the genes from *B. multivorans* could be ascribed to the analogous genes in the *S. meliloti* genome. The predicted functions are as follows; *SMb21103-SMb21106* encode the components of an ABC type transporter, *SMb21109* encodes an L-fucose/D-arabinose dehydrogenase, *SMb21110* encodes a L-fucono-1,5-lactone/D-arabino-1,4-lactone lactonase, *SMb21111* is a 2-keto-3-deoxy-L-fuconate dehydrogenase, *SMb21112* is a 2,4-diketo-3-deoxy-L-fuconate/2-keto-3-deoxy-D-arabinoate hydrolase, and *SMb21107* as well as *SMb21113* encode dehydratases (Fig. 5.1). *B. multivorans* has only one dehydratase at its L-fucose locus, that *S. meliloti* has two could mean that they have redundant activities or that one dehydratase acts in L-fucose metabolism while the other contributes to D-arabinose utilization. Additionally,

SMb21108 is annotated as a conserved hypothetical protein but shares high sequence identity with a mutarotase from *R. leguminosarum* that uses another methylpentose called L-rhamnose as a substrate (Richardson *et al.*, 2004; Richardson *et al.*, 2008). Mutarotases convert between the α and β anomers of a compound and β-L-fucose is the preferred substrate for L-fucose dehydrogenase in *B. multivorans* (Hobbs *et al.*, 2013), suggesting that SMb21108 makes β-L-fucose, and possibly β-D-arabinose, the predominant form of the substrate during metabolism. The predicted metabolic pathways carried out by these enzymes are outlined in Fig. 5.2.

5.4.6 Discoloration of the growth medium by SRmD622 is linked to medium acidification

It was observed following isolation of SRmD622 that during growth of this strain on defined medium containing D-arabinose, the medium surrounding the streak would take on a brown hue distinct from the medium around the wildtype grown under the same conditions (Fig. 5.3A and B). This discoloration was first observed during growth on a plate in medium containing 1.5% agar, but could also be observed in broth media (Fig. 5.3C). The mutant allele from SRmD622 was transduced into Rm1021 and the discoloration was observed in the novel strain, suggesting that this phenotype is not the result of a secondary mutation. Centrifugation of the broth culture resulted in cleared supernatant that retained its altered coloring, suggesting that the discoloration is due to the secretion of a substance into the culture medium.

SRmD622 has a mutation in *SMb21112*. Based on the predicted pathway, when metabolizing D-arabinose the protein encoded by this gene catalyzes the conversion of 2-keto-3-deoxy-D-arabinoate (D-KDA) into pyruvate and glycolic acid (Fig. 5.2). Interruption of *SMb21112* could result in the buildup of D-KDA especially if the previous reaction is energetically unfavorable in the reverse direction.

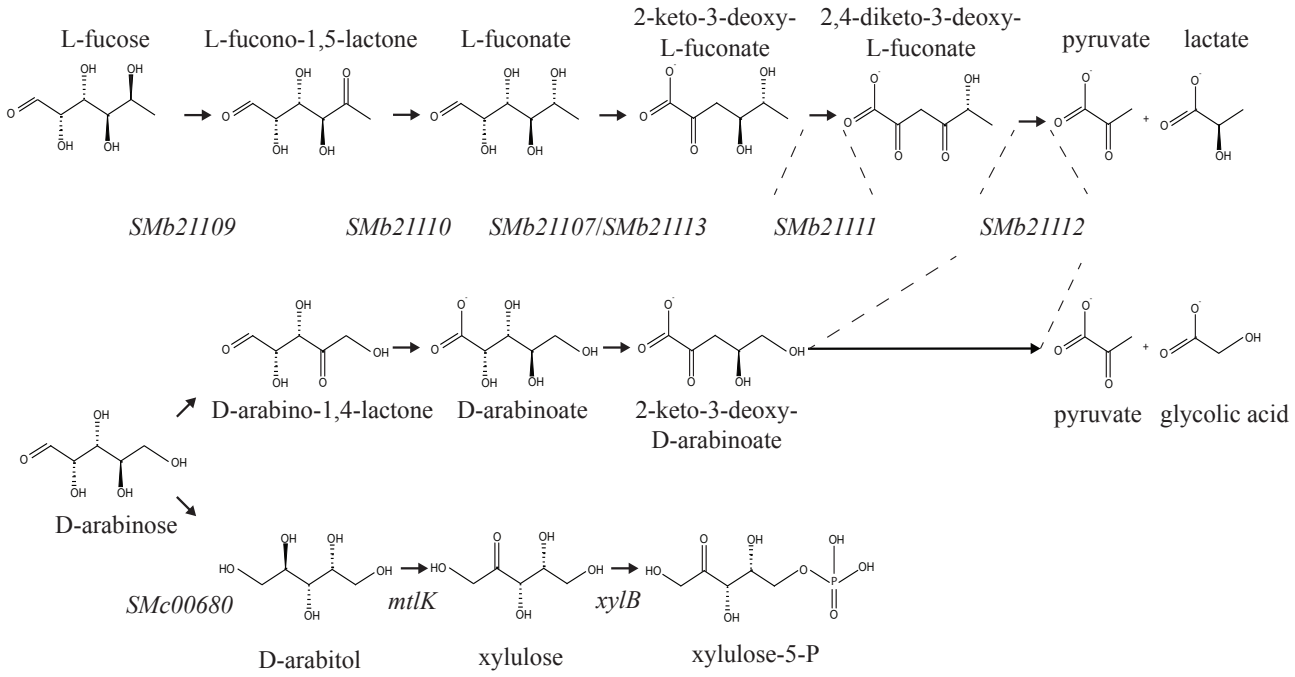


Figure 5.2. Non-phosphorylative metabolism of D-arabinose and L-fucose. L-fucose is broken down into pyruvate and lactate using a diketo hydrolase type pathway. D-arabinose catabolism occurs through a bifurcating pathway in which one route utilizes enzymes associated with sugar alcohol metabolism, resulting in the formation of xylulose-5-P, and the second route involves enzymes that degrade L-fucose which form pyruvate and glycolic acid.

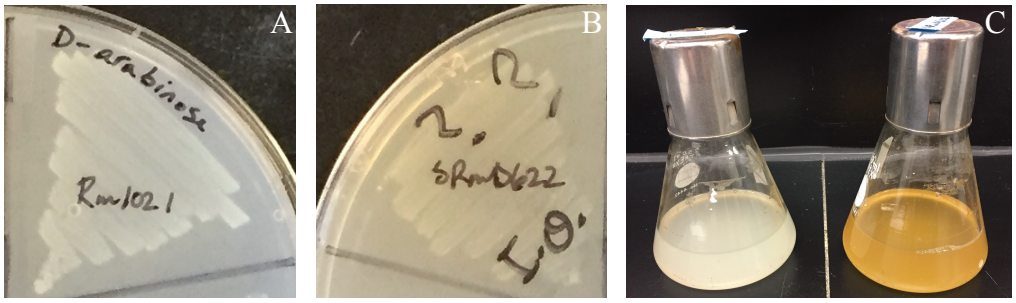


Figure 5.3. Discoloration of agar (A and B) and broth (C) media by SRmD622 utilizing D-arabinose. Strains were grown on Vincent's minimal medium (VMM) supplemented with 15 mM D-arabinose.

We hypothesized that if D-KDA was being secreted, there would be a decrease in the pH of the medium due to the carboxylic acid group on C1 of the compound.

An alternative hypothesis that was briefly explored was that the periplasmic pyrroloquinoline quinone (PQQ)-linked glucose dehydrogenase, *gcd*, which had been suggested to act on D-arabinose (Bernardelli *et al.*, 2008), was responsible for the generation of exogenous D-arabinoate. However, a double *SMb21112*/*gcd* mutant retained the media discoloration phenotype, suggesting that Gcd does not have a role in D-arabinose metabolism in *S. meliloti*.

To test for D-KDA secretion, pH indicator bromocresol purple (BCP), which turns from purple to yellow as the pH drops below 6.8, was included in the plate media along with 15 mM D-arabinose. Following incubation, Rm1021 did not alter the color of the medium while SRmD622 shifted the color from purple to yellow (Fig. 5.4A). To quantify the pH change, 5 mL cultures of Rm1021 and SRmD622 were grown for three days in defined medium with 15 mM D-arabinose and 15 mM glycerol, cleared by centrifugation, and the supernatants were taken and measured for pH using a pH probe. The pH of the wildtype culture was close to neutral at 7.26 ± 0.017 while the mutant was significantly lower at 6.83 ± 0.015 ($p=0.001$) (Fig. 5.4B).

S. meliloti dgoK mutants which acidify their growth medium have been shown to be more competitive for nodule occupancy (Geddes *et al.*, 2014). We hypothesized that SRmD622 could exhibit a similar characteristic when grown in the presence of D-arabinose. To test this, we examined whether the nodulation kinetics of *S. meliloti* inoculated alfalfa would change depending on the strain used to inoculate. Alfalfa seedling were planted in test tube slants containing Jensen's agar supplanted with 0, 2, 5, 10, and 15 mM D-arabinose and evaluated for plant growth based on concerns of growth inhibition by sugar concentration.

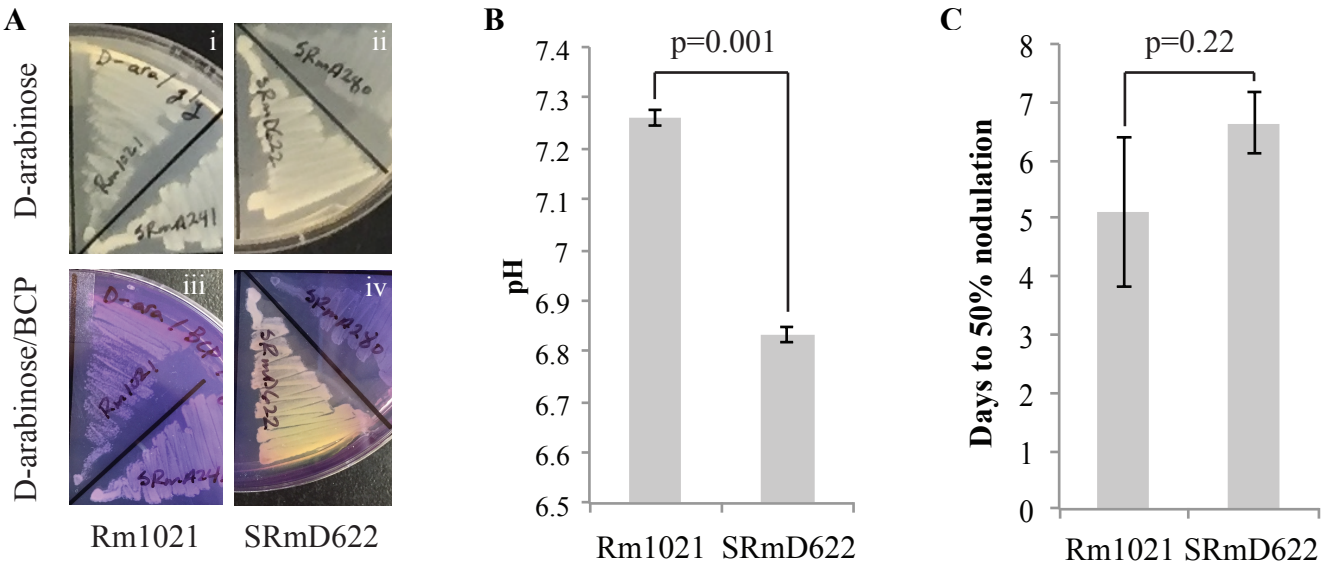


Figure 5.4. Medium acidification and its effect on nodulation kinetics. (A) Inclusion of pH indicator bromocresol purple (BCP) in VMM with 15 mM D-arabinose shows a color change from purple to yellow in the presence of SRmD622, i) Rm1021, ii) SRmD622, iii) Rm1021 with BCP, iv) SRmD622 with BCP. (B) pH quantification following growth in VMM with 15 mM D-arabinose shows a significant pH decrease of SRmD622 culture supernatant. (C) Nodulation kinetics of strains inoculated onto seedlings grown on Jensen's medium agar slants supplemented with 2 mM D-arabinose.

It was determined that 2 mM D-arabinose was the optimum concentration for evaluation of nodulation kinetics without inhibition of plant growth. However, the time to 50% nodulation by plants inoculated with Rm1021 or SRmD622 was not significantly different under these conditions (Fig. 5.4C). It is possible that 2 mM D-arabinose is not a high enough concentration to elicit the medium acidification response, since prior observations were made at a higher concentration (15 mM) of D-arabinose. However, more likely is that the pH change observed during growth of SRmD622 is not severe enough to alter the nodule kinetics of the strain. The *dgoK* mutant is capable of acidifying the medium to approximately a pH of 5, while SRmD622 decreases to only 6.8 units. Regardless the mild pH drop observed during the growth of SRmD622 is consistent with the secretion of D-KDA into the medium.

5.4.7 SMb21109 has D-arabinose and L-fucose dehydrogenase activity

SMb21103-SMb21113 has two predicted dehydrogenases in SMb21109 and SMb21111. We suspected that SMb21109 was a L-fucose/D-arabinose dehydrogenase and that SMb21111 was a 2-keto-3-deoxy-L-fuconate dehydrogenase (Fig. 5.1 and 5.2). To test this, cultures of wildtype Rm1021, mutant strain SRmA280 with an insertion upstream of the dehydrogenases, and SRmA280 complemented with *SMb21109* or *SMb21111* on a plasmid were grown under inducing conditions and lysed via French Press (16000 lb/in^2). The extracts were cleared of cell debris and separated by non-denaturing polyacrylamide gel electrophoresis.

First, we were interested in the coenzyme preference (NAD^+ vs $NADP^+$) of the dehydrogenase, the gels were stained for L-fucose dehydrogenase activity using two assay reagents, one containing NAD^+ and a second containing $NADP^+$. The gel stained with the NAD^+ containing reagent showed two bands of activity, one in the lane loaded with the wildtype

extract, and a second band of similar intensity and migration distance in the *SMb21109* complemented lane, suggesting the SMb21109 is an NAD^+ dependent L-fucose dehydrogenase. No bands were present in the mutant or *SMb21111* complemented lanes (data not shown). Interestingly, the $NADP^+$ stained gel showed a band of activity across all four lanes (data not shown), suggesting that there may be an enzyme that can oxidize L-fucose using $NADP^+$ that is unrelated to L-fucose metabolism.

Next, we wanted to determine if SMb21109 had D-arabinose dehydrogenase activity as well. Extracts from Rm1021, SRmA280, and SRmA280 complemented with *SMb21109* were separated by gel electrophoresis and stained for dehydrogenase activity in an assay reagent containing NAD^+ as well as L-fucose or D-arabinose. The gel stained with D-arabinose showed a band of activity in the lane loaded with wildtype extract as well as a band in the lane loaded with SMb21109 complemented extract (Fig. 5.5). No bands were present in the lane loaded with SRmA280 (Fig. 5.5). Taken together, the data suggest that SMb21109 has NAD^+-dependent L-fucose and D-arabinose dehydrogenase activity, which is consistent with the predicted pathway (Fig. 5.2).

*5.4.8 Δ*SMb21111 *strain cannot utilize D-arabinose but does grow on L-fucose*

The predicted annotation for *SMb21111* is a 2-keto-3-deoxy-L-fuconate dehydrogenase, which converts 2-keto-3-deoxy-L-fuconate into 2,4-diketo-3-deoxy-L-fuconate (Fig. 5.2). This enzyme is not thought to play a role in the metabolism of D-arabinose, therefore a diketo intermediate is not formed during D-arabinose metabolism (Fig. 5.2). If true, we hypothesized that an in-frame deletion of *SMb21111* would result in a strain is unable to use L-fucose for growth, but would grow like wildtype using D-arabinose.

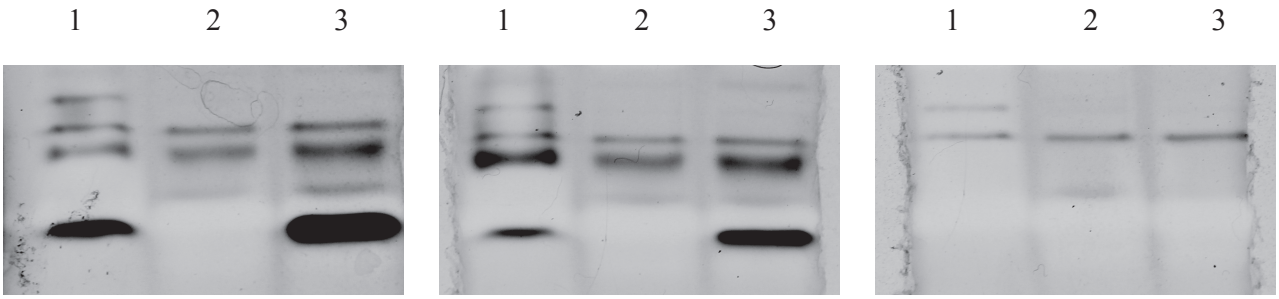


Figure 5.5. Non-denaturing PAGE gels showing L-fucose inducible dehydrogenase activity. Extracts from wildtype (pRK7813) (lane 1), SRmA280 (pRK7813) (lane 2), and SRmA280 (pCO37/*SMb21109*) (lane 3) cultures grown in VMM with glycerol and L-fucose were separated and stained for dehydrogenase activity using an assay reagent containing the substrate and cofactor listed below each panel.

This strain, constructed by conjugation of a suicide vector containing the flanking regions of *SMb21111* into Rm1021, was called SRmD653 and did indeed exhibit no growth on L-fucose and normal growth on D-arabinose (Table 5.3). The growth phenotypes of this strain are consistent with the predicted pathway (Fig. 5.2).

5.4.9 Plant growth is not significantly altered by inoculation with D-arabinose/L-fucose mutants

To investigate whether D-arabinose/L-fucose catabolic mutants were deficient in symbiosis, several strains were used to inoculate alfalfa plants, which were evaluated for generation of biomass after approximately four weeks of growth. These experiments did not reveal a significant growth difference between plants inoculated with wildtype and plants inoculated with any of the mutant strains (Fig. 5.6). However, it is worth noting that the lack of significance may be due to an unusual degree of variation among the Rm1021 inoculated plants. If the dry weights from plants inoculated with SRmA241 were considered equivalent to wildtype, then SRmA280 and SRmD622 inoculated plants do exhibit statistically significant growth deficiencies ($P < 0.05$), which could be the result of a reduced capacity to fix nitrogen. Interestingly, SRmD653 inoculated plants do not display this potential growth defect, which may suggest that the ability to catabolize D-arabinose is important for symbiosis but that L-fucose utilization dispensable (Fig. 5.6). Experiments will need to be repeated to confirm these observations.

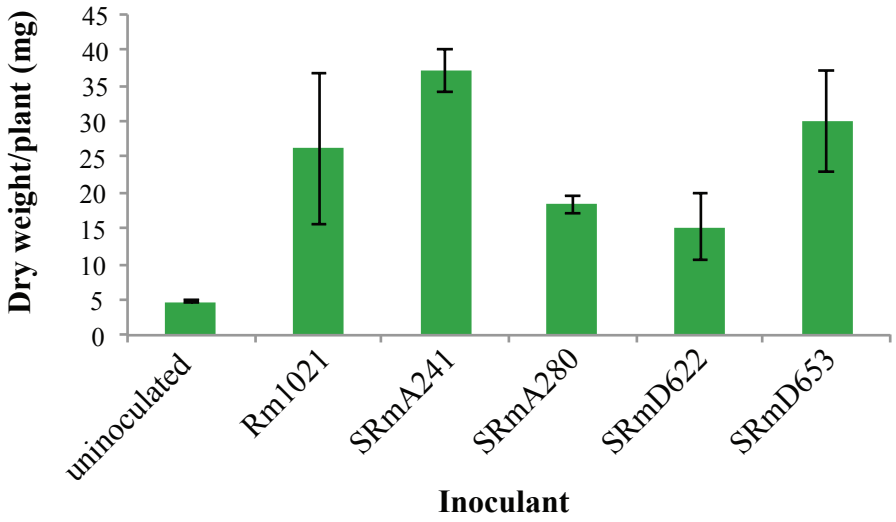


Figure 5.6. Dry weight of alfalfa plants inoculated with different *S. meliloti* strains. Plants were grown in sterile Leonard jar assemblies containing a mixture of sand and vermiculite with nitrogen-free Jensen's medium. Ten seedlings were planted per jar with three jars per treatment.

5.5 Discussion

Both L-fucose and L-rhamnose belong to a class of molecules known as methyl pentoses. L-fucose is also known as 6-deoxy-L-galactose and L-rhamnose is 6-deoxy-L-mannose. Rhizobia that are deficient in their ability to utilize L-rhamnose for growth are also less able to compete for nodule occupancy (Oresnik *et al.*, 1998). The ability to sense and utilize L-fucose has been shown to be important for colonization of the gut by enteric bacteria such as *E. coli* and *Campylobacter jejuni* (Pacheco *et al.*, 2012; Stahl *et al.*, 2011). It is not unreasonable to imagine that L-fucose is present in the rhizosphere. L-fucose is a major component of aerial-root mucilage secreted by maize which is thought to enrich the rhizosphere for plant beneficial bacteria (Van Deynze *et al.*, 2018). Additionally, fucosylation of nod factor (NF) is a common characteristic of species such as *Bradyrhizobium japonicum*, *Azorhizobium caulinodans*, and *Rhizobium* sp. NGR234 (Mergaert *et al.*, 1996; Quesada-Vincens *et al.*, 1997; Stacey, 1995), and NF isolated from *B. japonicum* has been shown to have plant growth promoting properties (Souleimanov *et al.*, 2002). Therefore, fucose is present in the rhizosphere to some extent and may have a role in competition in that environmental niche.

The presence of non-phosphorylative diketo-hydrolase type pathways for the degradation of L-fucose or L-rhamnose has been documented in bacteria and archaea (Reinhardt *et al.*, 2019). Likewise, it's been shown the D-arabinose can be metabolized in tandem with L-fucose via these types of pathways (Wolf *et al.*, 2016). D-arabinose is also metabolized via reduction into D-arabitol in *Mycobacteria* (Wojtkiewicz *et al.*, 1988; Wolucka, 2008). Both routes for the metabolism of D-arabinose have been observed here in *S. meliloti*, we are not aware of any other instances in which both pathways are present in the same organism.

That *pyc* mutants are able to grow using D-arabinose but not L-fucose is an interesting observation (Table 5.3) and suggests that *S. meliloti* is able to utilize a two-carbon intermediate such as glycolate for growth. Glycolic acid was not able to support the growth of the wildtype or the *pyc* mutant when supplied as a sole carbon source (data not shown). However, glycoaldehyde derived from D-arabinose has been shown to enter the TCA cycle via glyoxylate (Nunn *et al.*, 2010). This pathway would bypass the generation of pyruvate negating the requirement for *pyc* and may serve as the route of carbon flow that permits growth of the *S. meliloti pyc* mutant on D-arabinose.

Chapter 6:

Movement of an insertion sequence element is correlated with increased catabolic activity in *Sinorhizobium meliloti* 1021

This work was carried out by MacLean Kohlmeier in collaboration with Barney Geddes and Peter Loewen. BG performed the cotransduction linkage strategy to identify the site of the suppressor mutation. PL assembled the genome sequence of the suppressor mutant.

6.1 Abstract

The legume endosymbiont *Sinorhizobium meliloti* has a robust metabolic capacity; a large proportion of its genome is dedicated to carbon metabolism and transport. Despite its catabolic capability, wildtype *S. meliloti* 1021 is unable to grow using the five-carbon sugar alcohol xylitol as a sole carbon source. However, spontaneous suppressor mutants that are able to use xylitol can arise. Determination of the nature of the mutation that permits growth on xylitol in the suppressor strain is the focus of this project. Generation of a genetic map based on experimentally determined cotransduction frequencies resulted in the identification of a 10 kb chromosomal region. Analysis of the region following whole genome sequencing of the mutant revealed the presence of a 1 kb insertion sequence (IS) element, IS*Rm2011-2*, that is absent from this locus in the wildtype. The IS element is within a gene which is downstream of two candidate dehydrogenases, *SMc01991* and *SMc01992,* and may contribute to xylitol catabolism. Expression of either *SMc01991* or *SMc01992* in a wildtype background is sufficient to allow growth on xylitol as a sole carbon source. Dehydrogenase assays demonstrated that both SMc01991 and SMc01992 have xylitol dehydrogenase activities, which use $NADP^+$ and NAD^+ as coenzymes respectively. Interruption of *SMc01990* with a suicide vector, in a manner that mimics the insertion of IS*Rm2011-2*, allows for the strain to utilize xylitol. However, expression of *SMc01990* in either mutant background does not abolish growth on xylitol, suggesting that expression of *SMc01990* is not directly responsible for the inability to utilize the substrate. Taken together, the data suggest that xylitol utilization is due to a positional effect of an IS*Rm2011-2* element that effects the expression of upstream genes.

6.2 Introduction

Xylitol is a five-carbon sugar alcohol that does not have a chiral center. Although rare, it is found in some plants as well as fungi. Since it has a sweetness close to that of sucrose, and it is poorly utilized by oral bacteria, it has been used commercially as an artificial sweetener (Edelstein *et al.*, 2007; Maguire & Rugg-Gunn, 2003; Ritter *et al.*, 2013). Work on xylitol has mostly been focused on its use as an intermediate during ethanol formation in yeasts, or as novelty pathway generated by early adaptive evolution experiments in enteric bacteria (Mortlock, 1984).

Approximately half of *Sinorhizobium* spp. isolates have the ability to utilize xylitol as a sole carbon source (Bergey's manual of determinative bacteriology (7th ed.), 1964). *Sinorhizobium meliloti* strain Rm1021 does not grow using xylitol as a sole carbon source on defined medium. During the investigation of erythritol, adonitol and L-arabitol catabolism in *S. meliloti* (Geddes & Oresnik, 2012b), spontaneous suppressor mutants arose that were capable of growth utilizing xylitol. Genome sequencing of one such mutant identified an approximately one kb insertion sequence (IS) element, IS*Rm2011-2,* that had increased it's copy number by one in comparison to the wildtype.

Transposable elements (TE) are mobile DNA elements that are capable of "jumping" from locus to locus on a chromosome, or even moving to a separate replicon, in a process called transposition. TE's are found in every kingdom of life, but the simplest examples, IS elements, are found in bacteria. Certain TE's are involved in the mobility of large genetic segments, such as symbiosis islands, between bacterial strains. These segments can transfer between symbiotic and non-symbiotic strains of Rhizobia conferring the ability to establish symbiosis with legumes

(Haskett *et al.*, 2016). TE's play a role in the rearrangements and plasticity of bacterial genomes and act as a mechanism for evolution (Darmon & Leach, 2014).

Bacterial IS elements are small, 0.7 to 2.5 kb in length, DNA segments that only encode functions related to their transposition. This characteristic separates them from other TE's, which usually carry selectable markers. They have short inverted repeats at either end of their nucleotide sequence and often duplicate their target sequence upon insertion. IS elements were first described during characterization of spontaneous galactose and lactose mutants in *E. coli* (Jordan *et al.*, 1968; Malamy, 1970; Shapiro, 1969), but the concept was actually proposed earlier during investigations into genetic variation in maize (McClintock, 1956).

Typically, insertion of an IS element affects gene expression by interrupting a gene or regulatory region and inactivating it. As an example, wildtype *S. meliloti* strain Rm1021 contains an IS*Rm2011-1* element within transcriptional regulator *expR,* leading to an inability to produce symbiotically active polysaccharide, EPSII, under most conditions (Glazebrook & Walker, 1989; Pellock *et al.*, 2002). However, IS elements can also activate neighboring gene expression by inserting upstream of a gene and introducing a new promoter into a locus that is either partially or entirely housed within the element itself.

The increased catabolic capability of the suppressor mutant is of interest due to the effect that carbon utilization can have on symbiotic properties such as nitrogen fixation and competition for nodule occupancy. In order to determine the molecular nature of the mutation, this strain was further studied.

6.3 Materials and methods

6.3.1 Bacterial strains, plasmids, and culture conditions

Bacterial strains and plasmids used in this study are listed in Table 6.1. *S. meliloti* strains were routinely grown at 28°C using Luria-Bertani (LB) (Cold Spring Harbor Protocols, 2006) as a complex medium or Vincent's Minimal Medium as a defined medium (Vincent, 1970). Carbon sources were added to a final concentration of 15 mM. Antibiotics were used in solid media at the following concentrations: chloramphenicol (Cm), 20 µg/ml; gentamicin (Gm), 20 or 60 µg/ml; kanamycin (Km), 20 µg/ml; neomycin (Nm), 200 µg/ml; rifampicin (Rif), 50 or 100 µg/ml; spectinomycin (Sp), 100 or 200 µg/ml; streptomycin (Sm), 200 µg/ml; tetracycline (Tc), 5 µg/ml. If more than one concentration is listed, the higher value was used on Rhizobium species, while the lower value corresponds to *E. coli*. Concentrations were halved when added to liquid media.

6.3.2 DNA manipulations and genetic techniques

Standard techniques were used for plasmid isolation, restriction enzyme digests, ligations, transformations, and agarose gel electrophoresis (Sambrook & Russell, 2001). Conjugations and transductions were carried out essentially as previously described (Finan *et al.*, 1988; Finan *et al.*, 1984). Mutagenesis with Tn*5* was carried out as previously described, resulting in the generation of strains SRmD272-6, SRmD343, SRmD345, SRmD626, and SRmD627 (Finan *et al.*, 1985). The point of insertion was determined through arbitrary PCR and sequencing (Poysti *et al.*, 2007).

Isolation of the suppressor mutant was conducted by streaking the wildtype strain on an agar plate containing VMM supplemented with15 mM xylitol.

Table 6.1. Bacterial strains and plasmids

Strain or plasmid	Relevant characteristics	Reference
S. meliloti		
Rm1021	SU47 *str-21*, Sm^R	Meade *et al.* (1982)
SRmA355	Rm5000 *tpiB1*::pKNOCK-Gm, Gm^R	Poysti and Oresnik (2007)
SRmA449	Rm1021 Δ*tpiA*, Sm^R	Poysti and Oresnik (2007)
SRmA515	Δ*tpiA, tpiB1* Φ(SRmA355)→SRmA449, Gm^R	Poysti and Oresnik (2007)
SRmA737	Δ*tpiA, tpiB1*(SRmA515), *xylB*::Tn*5*, Nm^R	This work
SRmD268	Rm1021 *xlt-1*, Sm^R	This work
SRmD272	SRmD268::Tn*5* Hop bank→Rm1021, xlt^+Nm^R	This work
SRmD273	SRmD268::Tn*5* Hop bank→Rm1021, xlt^+Nm^R	This work
SRmD274	SRmD268::Tn*5* Hop bank→Rm1021, xlt^+Nm^R	This work
SRmD275	SRmD268::Tn*5* Hop bank→Rm1021, xlt^+Nm^R	This work
SRmD276	SRmD268::Tn*5* Hop bank→Rm1021, xlt^+Nm^R	This work
SRmD321	*xylB*::Tn*5*, Φ(SRmA737)→SRmD268, Sm^RNm^R	This work
SRmD343	SRmD268 *SMc02016*::Tn*5*, Nm^R	This work
SRmD345	SRmD268 *SMc02017*::Tn*5*, Nm^R	This work
SRmD617	Rm1021, *xlt-2* (*SMc01990*::pKNOCK-Gm), Gm^R	This work
SRmD625	Rm1021 *SMc02022*::pKan, Nm^R	This work
SRmD626	Rm1021*, xlt-3* (*SMc01990*::Tn*5*), Nm^R	This work
SRmD627	Rm1021, *xlt-4* (*SMc01990*::Tn*5*), Nm^R	This work
E. coli		
DH5α	F^- *supE44 lacU169 hsdR17 recA1 endA1 gyrA96 thi-1 relA1* (80*lacZ*ΔM15)	Hanahan (1983)
DH5αR	DH5α Rif^R	House *et al.* (2004)
DH5α λpir	λpir lysogen of DH5α	House *et al.* (2004)
DH5αR λpir	Rif^R derivative of DH5α λpir	House *et al.* (2004)
MM294A	*pro-82-thi-1 hsdR17 supE44*	Finan *et al.* (1986)
MT607	MM294A *recA56*	Finan *et al.* (1986)
MT616	MT607 (pRK600)	Finan *et al.* (1986)
Plasmids		
pCO37	pRK7813 containing *attB* sites, Gateway-compatible destination vector	Jacob *et al.* (2008)
pKNOCK-Gm	Suicide vector, Gm^R	Alexeyev (1999)
pKan	Suicide vector, pKNOCK-Gm derivative, Km^R	Pickering and Oresnik (2008)

pMK2	*SMc01990*/pCO37, Tc^R	This work
pMK3	*SMc01991*/pCO37, Tc^R	This work
pMK4	*SMc01992*/pCO37, Tc^R	This work
pMK5	*SMc01993*/pCO37, Tc^R	This work
pMK49	*SMc01990*::IS*Rm2011-2*/pRK7813, Tc^R	This work
pMK2014	FRT-*ccdB*-Cm^R-FRT cassette, Pen^R	House *et al.* (2004)
pMK2015	FRT-*ccdB*-Cm^R-FRT cassette, Pen^R	House *et al.* (2004)
pMK2016	*oriV* $oriT_{ColE1}$ with FRT cassette from pMK2014, Sm^R Sp^R	House *et al.* (2004)
pMK2017	$oriV_{R6K}$ $oriT_{RP4}$ with FRT cassette from pMK2015, Tc^R	House *et al.* (2004)
pRK600	pRK2013 *nptI*::Tn*9*, Cm^R	Finan *et al.* (1986)
pRK602	pRK600ΩTn*5*, Cm^R Nm^R	Finan *et al.* (1985)
pRK7813	Broad host range vector, Tc^R	Jones and Gutterson (1987)
pXINT129	*λint* and *xis* driven by P_{lac}, Km^R	Platt *et al.* (2000)

Single colonies were observed after approximately one week, which had spontaneously gained the ability to utilize xylitol. One such colony was picked and designated SRmD268.

Strains SRmD617 and SRmD625 were generated through targeted mutagenesis using suicide vectors pKNOCK-Gm (Alexeyev, 1999) and pKan (Pickering & Oresnik, 2008) respectively. Internal gene fragments from *SMc01990* and *SMc02022* were amplified using primers *SMc01990*_pK_F/R and *SMc02022*_pK_F/R, and cloned into the vectors as *Bam*HI/*Kpn*I and *Xba*I/*Sma*I fragments respectively (Table S1). Constructs were conjugated into Rm1021 and the presence of the inserts was confirmed using primers *SMc01990*_F/R and *SMc02022*_F/R respectively (Table 6.2).

Plasmids pMK2-5 were generated using the ORFeome Gateway system (House *et al.*, 2004; Schroeder *et al.*, 2005) using pCO37 as a destination vector as previously described (Geddes & Oresnik, 2012b). pMK49 was generated by amplifying the whole *SMc01990* gene containing IS*Rm2011-2* as a *Pst*I/*Eco*RI fragment with primers 268_F/R using SRmD268 as a template (Table 6.2). This fragment was cloned into pRK7813 (Jones & Gutterson, 1987) and conjugated into Rm1021.

6.3.3 PacBio sequencing

The genome of strain SRmD268 was isolated using standard techniques. The DNA was sent to Pacific Biosciences for sequencing using SMRT technologies. A *de novo* assembly was performed on the sequence data using the RS_HGAP_Assembly.2 Protocol in the SMRT Portal software. A total of 734 x 10^6 bp were detected in the dataset suggesting approximately 100x coverage. The sequence of strain Rm1021 was used as a reference for an alignment using Mauve (Darling *et al.*, 2004).

Table 6.2. Primers used during this study

Name	Sequence 5′→3′
Generation of mutants	
*SMc01990*_pK_F	ATAT**GGATCC**CGACGACCGTTTCGAGCCC
*SMc01990*_pK_R	ATAT**GGTACC**CCGTACAACCGCCCCCCG
*SMc02022*_pK_F	ATAT**TCTAGA**GACCCGGCCCGAACTGG
*SMc02022*_pK_R	ATAT**CCCGGG**CTTGGACAACTCTTCGCCC
Confirmation of mutants	
*SMc01990*_F	CTAGCCATCCTCCAACTGAATCTG
*SMc01990*_R	ATGATGCCTGAAAGAATCGTGATCG
*SMc02022*_F	ATGAAGCGGAAACTCATTGACG
*SMc02022*_R	TCAGAGACTATCCCTGACGAG
Generation of expression constructs	
268_F	ATAT**CTGCAG**ATGATGCCTGAAAGAATCGTGATCG
268_R	ATAT**GAATTC**CTAGCCATCCTCCAACTGAATCTG
*SMc01990*_F2	ATAT**CTGCAG**GGAGATGCATGCATGATGCCTGAAAGAA TCGTGATCG
*SMc01990*_R2	ATAT**GGATCC**CTAGCCATCCTCCAACTGAATCTG

Restriction sites are in bold.

6.3.4 Dehydrogenase assays

Dehydrogenase assays were performed as previously described (Geddes & Oresnik, 2012a; Kohlmeier *et al.*, 2019; Pickering & Oresnik, 2008). Briefly, strains containing an empty vector control or a complementing plasmid were grown in defined medium containing glycerol and xylitol for two days at 30°C. Cell free lysates were prepared and separated by non-denaturing polyacrylamide gel electrophoresis (PAGE). Subsequently, gels were developed in an assay reagent containing *p*-nitroblue tetrazolium, phenazine methosulfate, NAD^+ as a cofactor, and a substrate of interest.

6.4 Results

6.4.1 Wildtype S. meliloti *cannot utilize xylitol but spontaneous mutations that permit growth can arise*

S. meliloti is unable to utilize xylitol as a sole carbon source. However, when plated on defined media supplemented with xylitol, it was observed that colonies occasionally formed on control plates that contained xylitol. Bacteria from one such colony were isolated, single colony purified three times, and termed SRmD268 (Table 6.1). This allele was termed *xlt-1*. When this strain was re-tested for growth on defined medium with xylitol it was found to have a doubling time of approximately 9 hours; the doubling time of the wildtype strain under the same conditions was determined to be approximately 280 hours, which is essentially negligible (Fig. 6.1). The doubling time for the wild-type on defined medium using a number of other sugars is generally between 7.5-9 hours (Geddes & Oresnik, 2012a). To determine the frequency of this event, four independent cultures of Rm1021 were grown overnight in complex medium, washed, and plated onto defined medium containing xylitol.

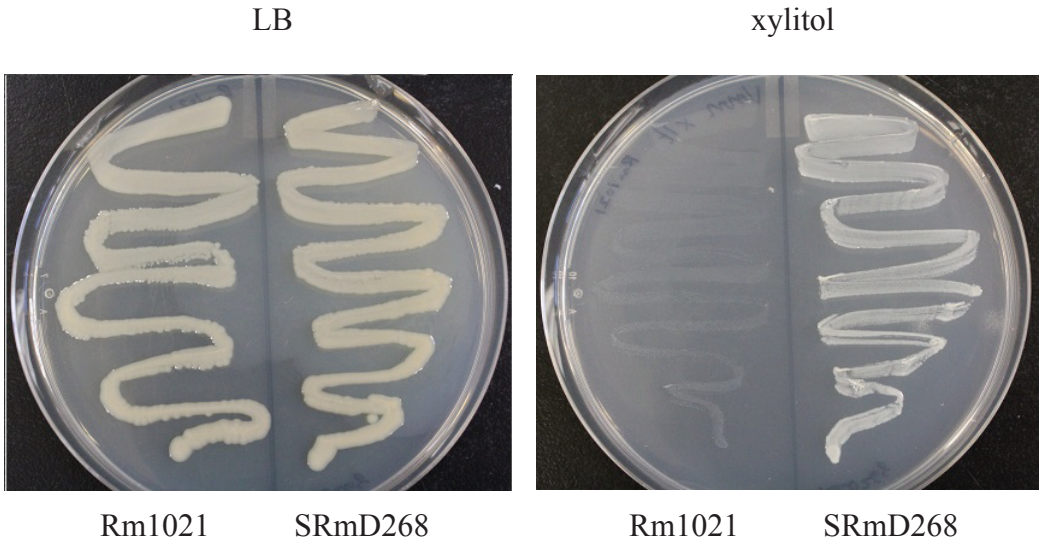


Figure 6.1. Growth of wildtype Rm1021 and suppressor mutant SRmD268 on LB (Luria-Bertani) complex medium and VMM (Vincent's minimal medium) defined medium supplemented with 15 mM xylitol as a sole carbon source. Strains were incubated at 28°C for ~1 week. Strain designation is listed below the panel while media type is listed above.

It was found that the frequency of this type of mutation appeared to be approximately $1x10^{-8}$. It was common to find plates devoid of any spontaneous colonies able to utilize xylitol in some experiments.

6.4.2 **xlt-1** *maps to the chromosome of Rm1021*

A search of the genome for xylitol metabolic genes did not yield any unambiguous candidate genes as being involved in the catabolism of xylitol. To determine that the mutation behaved as a point mutation, as well as to be able to locate it within the genome, a strategy was devised to localize the mutation using a classical bacterial genetics approach (Oresnik *et al.*, 1994). To carry this out SRmD268 carrying *xlt-1* was mutagenized with Tn*5*. The resulting transposon mutants were pooled, and infected with the general transducing phage ΦM12 (Finan *et al.*, 1984). The lysate was then used as a donor in transduction of Rm1021. The neomycin resistant transductants were then screened for the ability to grow on xylitol as a sole carbon source (Fig. 6.2). Five transductants that were neomycin resistant and xlt^+ were isolated and purified. These designated SRmD272, SRmD273, SRmD274, SRmD275, and SRmD276.

To determine the linkage of *xlt-1* with each of the Tn*5* insertions, a lysate of each these strains was grown and used as donor to transduce Rm1021 and to score the co-segregation of *xlt-1* and each Tn*5* insertion (Table 6.3). This frequency was then used to calculate the genetic distance between the markers, which was subsequently converted to an approximate physical distance (Table 6.3). In addition to determining the genetic distance between the insertions and *xlt-1*, the physical position of the Tn*5* transposon was determined using arbitrary PCR and sequencing. The physical distance from the Tn*5* along with its site of insertion would yield the approximate genomic location of the mutation (Fig. 6.2).

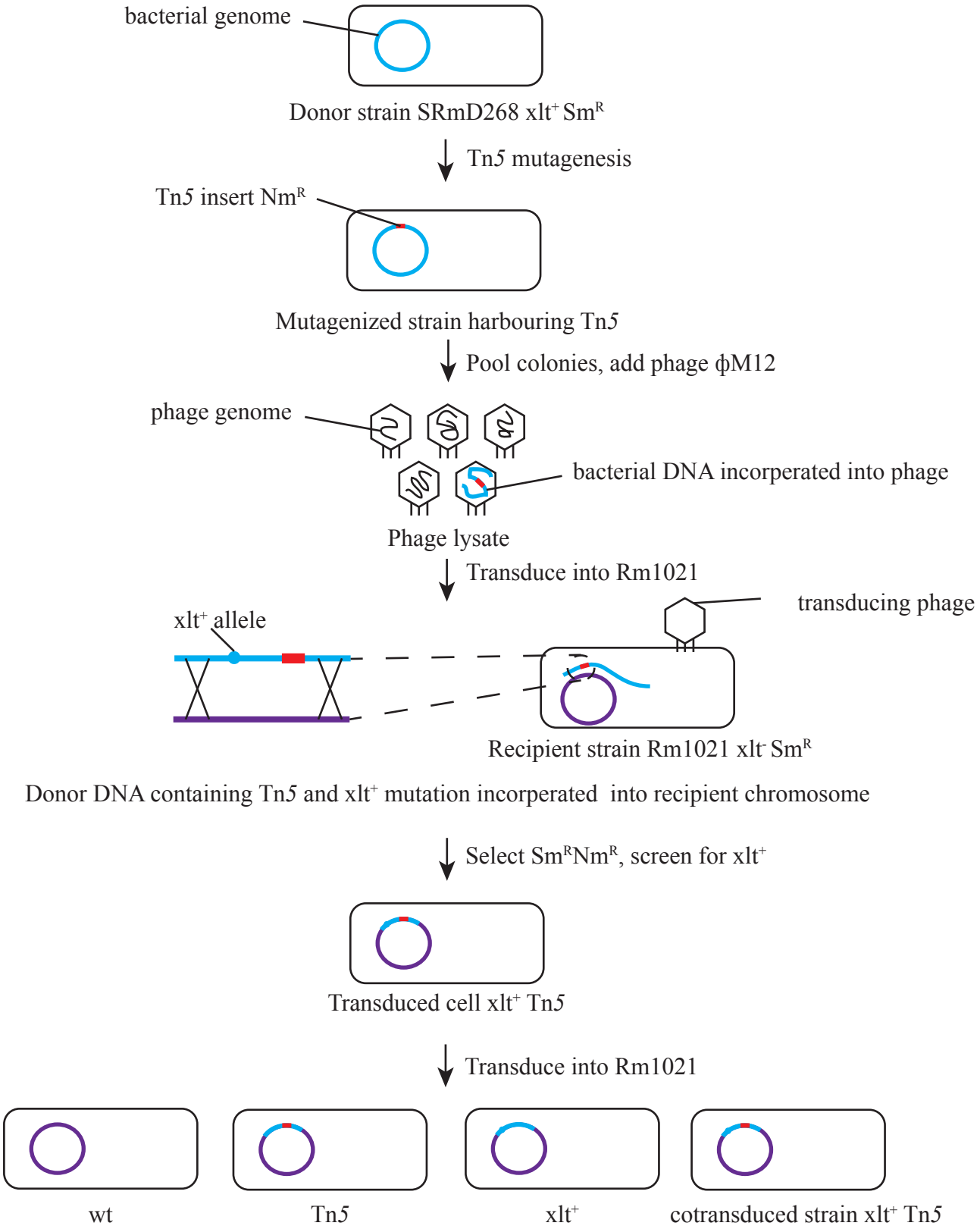

bacterial genome
Donor strain SRmD268 xlt+ SmR
Tn5 mutagenesis
Tn5 insert NmR
Mutagenized strain harbouring Tn5
Pool colonies, add phage ϕM12
phage genome
bacterial DNA incorperated into phage
Phage lysate
Transduce into Rm1021
xlt+ allele
transducing phage
Recipient strain Rm1021 xlt- SmR
Donor DNA containing Tn5 and xlt+ mutation incorperated into recipient chromosome
Select SmRNmR, screen for xlt+
Transduced cell xlt+ Tn5
Transduce into Rm1021
wt
Tn5
xlt+
cotransduced strain xlt+ Tn5

Figure 6.2. Schematic diagram featuring the core concept behind the cotransduction mapping strategy. Rm1021 genome is represented in purple while the SRmD268 donor DNA is in blue. ΦM12 transducing phage was used to make the lysate. The bacterial genome is represented as a single replicon, pSymA and pSymB were omitted for simplicity.

Table 6.3. Frequencies and map distances for strains in which alleles can be linked in transduction

Strain	Tn*5* insert w/in	% cotransduction	Map distance (kb)
SRmD272	*SMc02334*	25	55.5
SRmD273	*SMc01960*	11	78.1
SRmD274	n/a	36	43.3
SRmD275	n/a	8	85.4
SRmD276	*SMc02034*	18	65.3

n/a indicates failure to obtain sequence data. Wu's formula, used to calculate genetic distances from empirically determined cotransduction frequencies, is $F=(1-d/L)^3$ where F is the cotransduction frequency, d is the map distance between loci, and L is the nucleotide packaging size in the phage head.

The linkage data suggested that *xlt-1* was flanked by the Tn*5* insertions that were isolated. Together with the arbitrary PCR data it identified a 10 kb chromosomal region between *SMc01973-SMc02034* as housing the xylitol growth allele (Fig. 6.3).

An alternative strategy that was employed simultaneously was a Tn*5* mutagenesis of SRmD268 which carried the *xlt-1* allele followed by screening for the loss of the ability to grow on xylitol to identify genes involved in xylitol metabolism. This mutagenesis resulted in the isolation of two strains, SRmD343 and SRmD345, which were found to have insertions within *SMc02016* and *SMc02017* respectively. It is noteworthy that these genes are located approximately at the midpoint of the region identified by the genetic map, meaning that two independent strategies have identified the same region (Fig. 6.3).

6.4.3 Xylitol is oxidized to xylulose and phosphorylated becoming xylulose-5-phosphate

A typical strategy for sugar alcohol utilization is the formation of a keto-sugar, followed by phosphorylation, before entering central metabolism (Mortlock, 1984). Using this scheme, xylitol would be oxidized into D-xylulose before phosphorylation into xylulose-5-phosphate (X5P) (Fig. 6.4). To test if this route is the pathway utilized by the suppressor mutant, a mutated copy of *xylB*, a known xylulose kinase gene in *S. meliloti*, was transduced into SRmD268 resulting in the generation of SRmD321. This strain is unable to utilize xylitol for growth suggesting that XylB is involved in the metabolism of xylitol.

Since the utilization of xylitol by Rm1021 was dependent on XylB, xylitol only needed to undergo a single oxidation reaction to become a usable carbon source. An examination of the region that was delineated through mapping and mutagenesis experiments contains a putative operon, which consists of 11 genes spanning *SMc02022-SMc01990*.

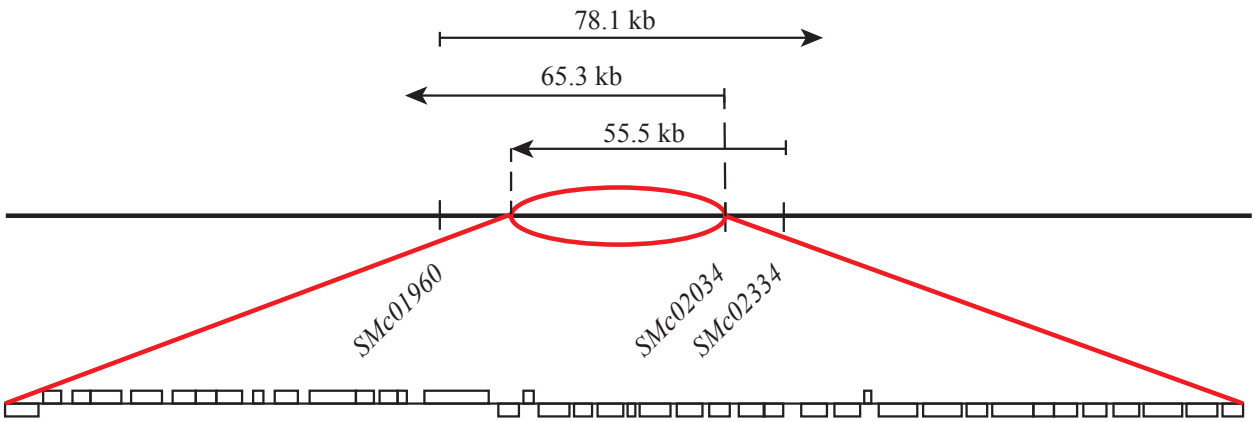


Figure 6.3. Genetic map of the region containing the suppressor mutation. The loci of the Tn5 insertions serve as anchor points, from which overlay of the genetic distances reveals an area common to all three segments. An ~10 kb chromosomal region can be seen through expansion of this area.

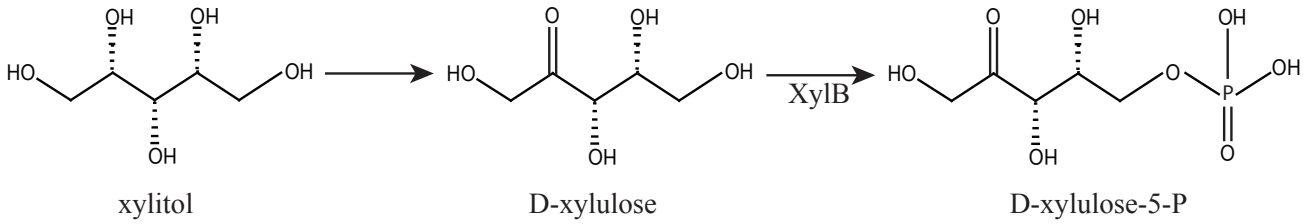


Figure 6.4. Xylitol utilization pathway by *Klebsiella pneumonia*. Xylitol is oxidized into D-xylulose before phosphorylation into xylulose-5-phosphate. Mutation of the *S. meliloti* xylulose kinase, *xylB*, in the suppressor mutant background abolished growth on xylitol as a sole carbon source, suggesting that this is the route employed by *S. meliloti* to catabolize xylitol.

The putative operon includes a LacI type transcriptional regulator (*SMc02022*), a solute binding protein (*SMc02021*), an ATP binding component (*SMc02020*), a permease (*SMc02019*), a hypothetical protein (*SMc02018*), a sterol desaturase (*SMc02017*), a putative ferredoxin reductase (*SMc02016*), a dioxygenase ferredoxin (*SMc01993*), D-xylulose reductase (*SMc01992*), dicarbonyl reductase (*SMc01991*), and a ferredoxin reductase (*SMc01990*) (Fig. 6.5). From these annotations, four of these genes appear to be involved in redox reactions, thus making them promising candidates for encoding a xylitol dehydrogenase.

To test if these genes were involved in xylitol metabolism the open reading frames of *SMc01990*, *SMc01991*, *SMc01992*, and *SMc01993* were cloned into pCO37 yielding pMK2, pMK3, pMK4 and pMK5 respectively. These were then conjugated into the wildtype and screened for their ability to grow on xylitol as a sole carbon source. The results showed that whereas pMK2 (containing *SMc01990*) and pMK5 (containing *SMc01993*) did not confer the ability of Rm1021 to utilize xylitol, the introduction of pMK3 (containing *SMc01991*) and pMK4 (containing *SMc01992*) were able to imbue growth on xylitol. We note that although the introduction of either *SMc01991* or *SMc01992* on a plasmid allowed the wild-type to grow on defined medium containing xylitol as a sole carbons source, the growth of the wild-type carrying of pMK4 (*SMc01992*) appeared to be more robust on agar plates.

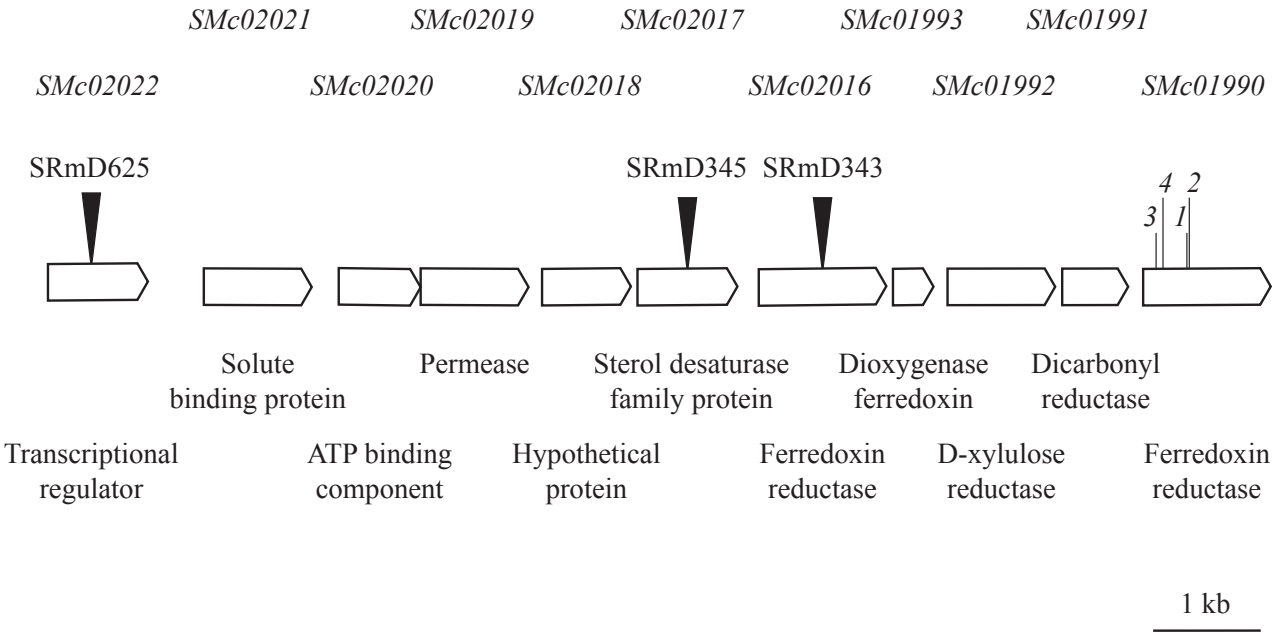


Figure 6.5. Locus diagram of the *SMc02022-SMc01990* region, genes are represented in the 5' to 3' direction. Black wedges represent the sites of insertional mutagenesis, vertical lines indicate different xylitol growth alleles numbered 1-4, locus tags are displayed above and annotation information can be seen below.

6.4.4 SMc01991 *and* SMc01992 *encode xylitol dehydrogenases*

Based on the ability of SMc01991 and SMc01992 to confer the ability of the wild-type to grow on xylitol, and that growth of the strain carrying the *xlt-1* allele are dependent on *xylB*, it was hypothesized that these genes encode xylitol dehydrogenases, and that one, or both, of these were expressed in SRmD268.

To determine if these proteins were expressed in the suppressor mutant and had xylitol dehydrogenase activity cultures of Rm1021, SRmD268, Rm1021 carrying pMK3, and Rm1021 carrying pMK4 were grown in defined medium with glycerol and xylitol. The cultures were resuspended in lysis buffer, lysed via French Press, and cleared of cell debris. These extracts were separated by nondenaturing polyacrylamide gel electrophoresis and the gels were stained for xylitol dehydrogenase activity.

Extracts from the wildtype did not exhibit xylitol dehydrogenase activity, whereas SRmD268 displayed two bands of activity; one that was dependent on $NADP^+$ and migrated far down the gel, and a second that was dependent on NAD^+ which did not migrate well into the gel (Fig. 6.6, lanes 1 and 2). The *SMc01991* expression extract showed a band of activity in the gel stained with $NADP^+$ and xylitol, which corresponded to the lower band from SRmD268 (Fig. 6.6, lane 3). The *SMc01992* expression extract also showed a band of activity, this one corresponded to the higher band that was dependent on NAD^+ (Fig. 6.6, lane 4). Taken together the data suggest that both SMc01991 and SMc01992 have $NADP^+$ and NAD^+ dependent xylitol dehydrogenase activity respectively, and that SRmD268 expresses both proteins, either of which can permit growth on xylitol in a wildtype background.

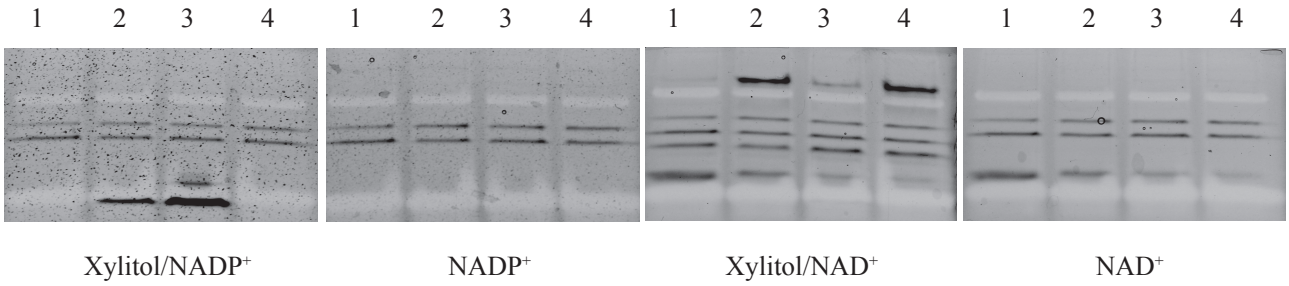


Figure 6.6. Xylitol dehydrogenase activity. Extracts from wildtype (lane 1), SRmD268 (lane 2), Rm1021 with *SMc01991 in trans* (lane 3) and Rm1021 with *SMc01992 in trans* (lane 4) separated by nondenaturing polyacrylamide gel electrophoresis. Gels were stained for dehydrogenase activity using an assay reagent containing the substrate(s) listed below each panel.

6.4.5 *SRmD268 contains an IS*Rm2011-2 *insertion sequence element within* SMc01990

To identify the molecular nature of the mutation, we attempted to sequence the region by generating a series of overlapping PCR products spanning the putative operon. However, preliminary sequencing of the region failed to provide adequate coverage. Therefor, the entire genome of SRmD268 was sequenced using Pacific Biosciences SMRT platform. The 10 kb region identified by genetic mapping was compared to the corresponding wildtype sequence. The only difference was an approximately one kb insertion within *SMc01990*, a putative ferredoxin reductase, which was absent from the same region in the wildtype (Fig. 6.5).

A BLAST search using the insert sequence as a query returned a number of hits to different *S. meliloti* strains including Rm1021, the sequence specifically matched an insertion sequence (IS) element called IS*Rm2011-2*. IS elements are short, mobile genetic elements that only encode components necessary for their own transposition. IS elements and other types of mobile genetic elements are common constituents of prokaryotic genomes (Burrus & Waldor, 2004; Dobrindt *et al.*, 2004). IS*Rm2011-2* elements belong to the IS630 family, which are characterized by target site duplication and a cut and paste transposition mechanism (Siguier *et al.*, 2014).

To determine the origin of the IS element, the Rm1021 and SRmD268 genomes were analyzed using ISfinder, a dedicated database of IS elements located at https://www-is.biotoul.fr (Siguier *et al.*, 2006). ISfinder revealed six IS*Rm2011-2* elements in the Rm1021 genome; interestingly, seven of these elements were identified in SRmD268. Six of these elements were in the same position as that in Rm1021 whereas the seventh was within *SMc01990*.

6.4.6 Loss of SMc1990 *is not responsible for the gain of xylitol catabolic function*

Typically, insertion of an IS element affects gene expression by interrupting a gene or regulatory region resulting in loss of function mutations which either directly or indirectly lead to a phenotypic change. We hypothesized that the a novel growth phenotype observed might be due to the loss of negative regulation (presumably by the loss of the closely associated LacI type regulator encoded by *SMc02022*), and that its binding activity might be responding to a metabolite that was a product of the reaction that would be encoded by the *SMc01990* (Fig. 6.5).

Two experiments were carried out to address this hypothesis. The first, *SMc01990* was introduced into SRmD268 which carries the *xlt-1* allele, the second, to introduce a mutation in *SMc02022* in a wild-type background. The results show that the introduction of pMK2, which carries *SMc01990*, does not revert the ability of SRmD268 to utilize xylitol, nor does it change the inability of the wildtype to utilize xylitol. In addition, mutation of the negative regulator encoded by *SMc02022* did not result in a gain of function (Table 6.4). These results taken together support the hypothesis that the ability of SRmD268 to express the xylitol dehydrogenase activities encoded by *SMc01991* and *SMc01992* is not brought about by a loss of function of *SMc01990* and that *SMc02022* does not appear to be responsible for the negative regulation of these genes (Table 6.4). The data also suggest that the insertion defining the *xlt-1* allele appears to function as a *cis* acting element.

Table 6.4. Carbon phenotypes

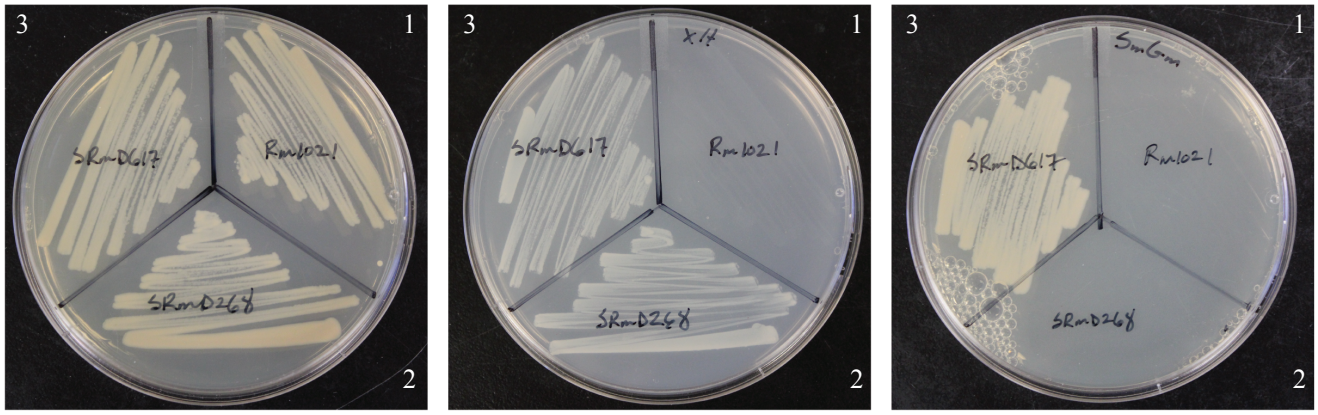
Strain	Characteristic	LB	xlt	gly	L-ara
Rm1021	wt	+	-	+	+
SRmD268	*SMc01990*::IS*Rm2011-2*	+	+	+	nd
SRmD617	*SMc01990*::pKNOCK-Gm	+	+	nd	+
SRmD625	*SMc02022*::pKan	+	-	+	nd
SRmD626	*SMc01990*::Tn*5*	+	+	+	nd
SRmD627	*SMc01990*::Tn*5*	+	+	+	nd
Rm1021 (pMK2)	*SMc01990 in trans*	+	-	nd	nd
SRmD268 (pMK2)	*SMc01990 in trans*	+	+	nd	nd
SRmD617 (pMK2)	*SMc01990 in trans*	+	+	nd	nd
Rm1021 (pMK49)	*SMc01990*::IS*Rm2011-2 in trans*	+	-	nd	+

Growth; +, like wildtype; -, no growth; nd, not determined. Abbreviations; LB, Luria-Bertani; xlt, xylitol; gly, glycerol; L-ara, L-arabinose.

6.4.7 Insertions in SMc01990 are responsible for the gain of function

Based in the simplicity of the IS*Rm2011-2* element, we suspected that the xylitol growth phenotype was not a property of the IS element itself, but a product of the position of the insertion. If true, an insertion of a different type at the same locus should generate a strain with the ability to utilize xylitol. To test this hypothesis, an internal fragment of *SMc01990* was PCR amplified and cloned into the suicide vector pKNOCK-Gm (Alexeyev, 1999), and subsequently introduced into Rm1021 to generate SRmD617 (carrying allele *xlt-2*), ostensibly mimicking the insertion of the IS element in SRmD268. When SRmD617 was tested, it was found that it was able to grow on xylitol as well as the suppressor mutant (Fig. 6.7, Table 6.4), suggesting that xylitol utilization is not a property of the IS element itself.

To determine if there are any other sites of insertion that may prompt a xylitol growth phenotype, a random Tn*5* mutagenesis was carried out on Rm1021. Two independent mutagenesis representing approximately 10,000 transposition events were directly plated onto defined medium containing xylitol. Although fifty-six colonies were found, only two of these were found to truly be able to grow on xylitol. These two colonies were subsequently purified and designated SRmD626 and SRmD627, carrying alleles *xlt-3* and *xlt-4* respectively. The sites of insertion were identified and both transposons were within *SMc01990*. Interestingly, both transposons were found near the 5′ end of *SMc01990*, in close proximity to the insert site of the IS element. In fact, all of the xylitol-growing mutants generated in this study have an insert within the first 330 bp of *SMc01990,* suggesting that this region may be a hotspot for insertions that permit xylitol utilization (Fig. 6.5).



LB xylitol Gm

Figure 6.7. Targeted mutagenesis of *SMc01990* with a suicide vector results in a strain with the ability to utilize xylitol as a sole carbon source. Clockwise from top right; 1, Rm1021; 2, SRmD268; 3, SRmD617. Growth media from left; LB; VMM with 15 mM xylitol; LB with gentamicin (Gm).

We hypothesized that xylitol utilization was attributed to positional insertion within *SMc01990*, meaning that the insertion was functioning as a *cis* acting element. As a control experiment, the entire sequence of *SMc01990* containing the IS element was PCR amplified and cloned into expression vector pRK7813 (Jones & Gutterson, 1987), generating pMK49. This construct was conjugated into the wildtype background and the strain was screened for growth. It was found that expression of the interrupted gene from a plasmid does not permit growth on xylitol, suggesting that the IS element must be inserted into the chromosome to be effective (Table 6.4).

6.5 Discussion

In this chapter it has been shown that *SMc01991* or *SMc01992* both encode xylitol dehydrogenases and that the generated xylulose is subsequently phosphorylated by XylB yielding X5P, a metabolite associated with central metabolism. The mechanism of up-regulation of these two genes is currently unknown. It may be that the insertion of an element within a multi-cistronic mRNA can lead to increased transcript stability; thus allowing an increased probability of having *SMc01991* and *SMc01992* transcribed, or it may be that an insertion changes the overall regional DNA packing which leads to increased transcription.

To our knowledge IS elements have not previously been associated with increased catabolic activity in *S. meliloti*. However, there are examples of IS elements activating silent operons during exposure to starvation conditions in *E. coli* (Hall, 2000). The ability of TEs to modulate gene expression has caused many mutations leading to improved adaptation to environmental conditions (Casacuberta & González, 2013; McClintock, 1984; Miousse *et al.*, 2015). It is noteworthy that the IS*Rm2011-2* element described herein has inserted downstream

of the genes which are being modulated, which differentiates this situation from other instances described in the literature.

Genomic rearrangements have been observed in cultures of *S. meliloti* including large changes resulting in alterations in the number of replicons in the genome. An IS*Rm11* element was determined to be at the site of cointegration of the two megaplasmids resulting in a two-replicon *S. meliloti* strain, which may indicate that IS elements have an important role in determination of genome architecture (Guo *et al.*, 2003).

The movement of IS elements has also associated with improved symbiotic efficiency of *Sinorhizobium* with soybean. It was shown that strains with IS element insertions within a type 3 secretion system (T3SS) gene cluster could induce the formation of effective nodules on *Glycine max*. This observation became the basis of an adaptive evolution strategy, which was able to improve nodulation only under selection pressure from *G. max*. Interestingly, the number of IS elements was shown to increase during adaptive evolution (Zhao *et al.*, 2017). The precise mechanism for this phenomenon will be the focus of future studies.

Chapter 7:

Conclusions

7.1 Thesis conclusions and observations

The objectives for this thesis as they were outlined in the first chapter were: to identify and characterize genetic loci involved in metabolism of galactitol, to identify the genes involved in the metabolism of sorbitol, mannitol, and D-arabitol, and to characterize a suppressor mutation that permits growth of *S. meliloti* on xylitol. The completion of these objectives, as well as additional projects that arose from them, is described in the previous chapters. Conclusions derived from these works and prospects for future studies are outlined in the following section.

The first objective for this thesis was to identify and characterize a genetic locus involved in galactitol metabolism that had been tentatively localized to the pSymB chromid (Charles & Finan, 1990, 1991). It was shown that galactitol is metabolized using components from both the chromosome and pSymB, galactitol is first oxidized by SmoS (*SMc01500*) generating tagatose, then phosphorylated by TagD (*SMb21374*) into T6P, and subsequently epimerized by TagE (*SMb21373*) into F6P (Fig. 2.9).

TagE had been previously annotated as an aldolase, but bioinformatic analysis and heterologous complementation experiments demonstrated that TagE could substitute for a known tagatose epimerase from *A. tumefaciens* (Fig.2.2). Additionally, we discovered that the *E. coli* gene on which the aldolase annotation was based, *gatZ*, could also complement an epimerase mutant when expressed from a plasmid (Fig. 2.7), suggesting that the basis of the assigned annotation is incorrect. A phylogenetic analysis of more than 2000 protein sequences suggested that the misannotation was not isolated to *S. meliloti*, *A. tumefaciens*, and *E. coli*, but was systemic throughout the InterPro protein family (Fig. 2.8). There are likely many datasets containing epimerase genes masquerading as aldolases due to inaccurate gene annotations. This example highlights the dangers of overreliance on predictions generated from *in silico* analysis,

and researchers should be aware that current methods for prediction of gene function are not error free and instances of misannotation are high (Schnoes *et al.*, 2009).

The galactitol catabolic genes in *E. coli* are encoded by *gatYZABCDER*, and had been previously thought to encode a heterodimeric bisphosphoaldolase from *gatYZ* (Nobelmann & Lengeler, 1995; Nobelmann & Lengeler, 1996). It was suggested that GatY hydrolyzes T1,6bisP into GAP and DHAP and that GatZ enhances the activity of GatY without any activity of its own (Lengeler, 1975; Lengeler, 1977). The discovery that GatZ can complement for an *A. tumefaciens* epimerase mutation (Fig. 2.7) raises the question, why does *E. coli* encode a T6P epimerase and a T1,6BP aldolase? Having two metabolic routes to degrade T6P seems counterintuitive. However, metabolic redundancy and plasticity has been suggested to be an essential feature of many organisms. Benefits to repetitive or redundant genes include resistance to both environmental changes and detrimental mutations (Güell *et al.*, 2014).

It was shown that galactitol could compete with mannitol for transport, albeit at a seemingly low affinity (Fig. 2.5B). Since mannitol uptake is severely reduced in a *smoK* mutant (Fig. 4.3A), the SmoEFGK transporter is likely responsible for high affinity transport of mannitol, as well as low affinity transport of galactitol. It seems probable that a higher affinity galactitol transporter exists as well, possibly encoded by the ABC type transporter system, *tagABC*, which is adjacent to the metabolic genes (Fig. 2.1). It would be interesting to confirm this hypothesis by measuring the uptake of radiolabelled galactitol by *tag* and *smo* transporter mutants, as well as to test for uptake inhibition by tagatose.

The second objective for this thesis was to characterize the metabolism of sorbitol, mannitol, and D-arabitol in *S. meliloti*. Two approaches were utilized to facilitate completion of

this objective; biochemical characterization of SmoS and genetic characterization of the *smo* locus.

Biochemical and kinetic characterization of SmoS revealed some notable insights into the protein's behavior. Kinetic analysis showed that sorbitol was the preferred substrate for SmoS, although the enzyme has a higher affinity for galactitol (Table 3.2). The enzyme's low reaction velocity of galactitol oxidation was attributed to the binding energy of the SmoS-tagatose complex, which was shown to be more stable than the SmoS-fructose complex (Fig. 3.7).

To characterize SmoS, crystal structures of unbound SmoS and sorbitol bound SmoS were generated (Table 3.1). A confounding feature of the SmoS-sorbitol structure was that sorbitol was determined to be in a non-reactive position within the active site, meaning that it was too distant and in the wrong orientation to permit oxidation by catalytic residue Tyr153 (Fig. 3.4B). We speculated that this may be due to the absence of NAD^+ in the binding pocket. Inclusion of NAD^+ may stabilize and orient sorbitol in the active site, but it would also prompt a reaction, which would prevent the capture of a substrate-bound complex.

To rectify this, a non-catalytic mutation of SmoS could be made which would allow for capture of both NAD^+ and a polyol substrate in their binding sites without prompting a reaction. An innocuous mutation that would cause minimal structural disruption to the protein would be Tyr153Phe, while also allowing for substrate capture in a catalytic position.

The pathway for sorbitol, mannitol, and D-arabitol utilization in *S. meliloti* had been determined previously from biochemical assays from strains carrying unmarked catabolic mutations (Arias *et al.*, 1979; Gardiol *et al.*, 1980; Martínez De Drets & Arias, 1970). We have furthered the understanding of this route through identification of the genes responsible for encoding sorbitol and mannitol dehydrogenase activity (*smoS* and *mtlK*), fructose kinase activity

(*frk*), and phosphoglucose isomerase activity (*pgi*). The genes required for sorbitol, mannitol, and D-arabitol transport and oxidation are encoded from the *smo* locus (Fig. 4.1A), while fructose kinase and phosphoglucose isomerase genes are found downstream of the *frc* locus involved in fructose utilization (Fig. 4.1B) (Lambert *et al.*, 2001).

Its been shown that the substrate binding protein of the *smo* transporter, SmoE, is induced by sorbitol, mannitol, D-arabitol, and maltitol (Mauchline *et al.*, 2006). Sorbitol and D-arabitol compete with mannitol for use of the *smo* transporter, but maltitol exhibits low affinity transport in conjunction with mannitol (Fig. 4.3). Maltitol is a disaccharide consisting of glucose and sorbitol. That growth on maltitol is reduced in a *smoS* mutant background suggests that SmoS is involved in the catabolism of maltitol. We hypothesized that maltitol is being hydrolysed by an unknown glucosidase into glucose and sorbitol, with the glucose monomer able to support the growth of a *smoS* mutant. Attempts at identification of this glucosidase via Tn*5* mutagenesis have been so far unsuccessful. However, a BLAST search of the maltitol degrading α-glucosidase PalH sequence from *Erwinia rhapontici* (Börnke *et al.*, 2001) against the *S. meliloti* genome identified two genes, *agaL2* and *melA*, as likely to encode a similar protein. Additionally, *S. meliloti* trehalose transport, *thuEFGK*, and catabolic genes, *thuAB*, have been implicated in the utilization of maltitol via a non-specific catabolic route (Ampomah *et al.*, 2013). These genes make compelling targets for future work on maltitol metabolism in *S. meliloti*.

Fructose transport at the *frc* locus has been extensively characterized, but our data are not entirely consistent with previously reported results. Prior fructose transport mutations (*frcC*) were reported to abolish growth on fructose with no effect on mannitol utilization or transport (Lambert *et al.*, 2001). Our data indicate that *frc* mutants exhibit reduced growth on sorbitol,

mannitol, and fructose, and that transport of these substrates is shared between the *smo* and *frc* transporters (Table 4.6). We note that the strains and method of mutagenesis differ between previous work and our own, though how these differences could account the phenotypic discrepancies is not clear.

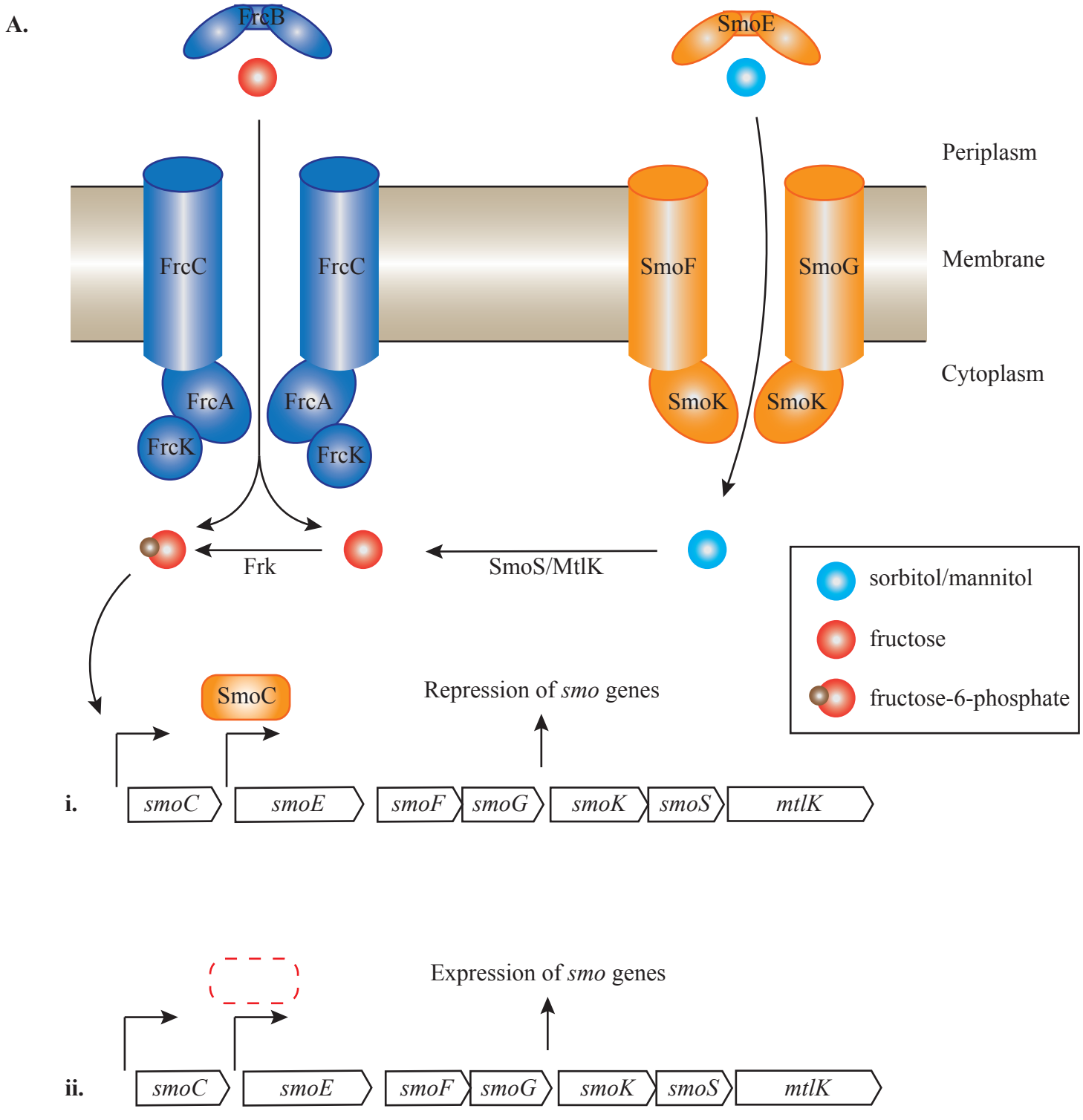
Consistent with previous results, we determined that growth on fructose and mannitol induces the uptake of fructose (Lambert *et al.*, 2001). Additionally, we showed that growth on fructose does not induce transport of mannitol, but that mannitol transport could be permitted following growth on fructose in *frcK* or *frk* mutant backgrounds (Fig. 4.6B). While Frk is definitively a fructose kinase (Fig. 4.5), the role of FrcK is less clear. Strains with mutations in *frcK* exhibit a reduced ability to transport fructose (Fig. 4.6B), suggesting that it has a role in fructose uptake. However, *frcK* mutants also exhibit reduced growth on sorbitol and mannitol, which can be restored by complementation with *frk* (Table 4.5), suggesting that FrcK has fructose kinase activity. Additionally, *frcK* mutants permit the transport of mannitol in a manner similar to *frk* mutants indicating that they have a related function (Fig. 4.6B). Therefore, FrcK exhibits properties of both a fructose transport protein and a fructose kinase. Taken together, the data suggest that the ability to make F6P is important for the proper regulation of sugar alcohol transport. The idea that metabolite pools influence gene regulation has been suggested previously (Hawkins *et al.*, 2018), but has not yet been subject to a thorough investigation.

We showed that SmoC is a negative regulator that represses the expression of the *smo* locus in non-inducing conditions (Fig. 4.4). It's possible that F6P could interact with SmoC, directly or indirectly, to repress transcription (Fig. 7.1A). Alternatively, its been shown that ABC type transporters can be regulated via direct phosphorylation of regulatory domains that can be fused to either the transmembrane domain or the nucleotide binding domain (Biemans-

Oldehinkel *et al.*, 2006). Therefore, it is also possible that SmoEFGK could be regulated by direct phosphorylation of a regulatory domain mediated by F6P independently of SmoC (Fig. 7.1B).

To test these possibilities, transport assays of fructose grown *smoC* mutant strain SRmD641 using labeled mannitol might provide some insight into the involvement of SmoC in this process. Additionally, ^{14}C-mannitol transport assays using strain with a mutation to *pgi*, which converts F6P into G6P, may reveal if F6P really is the key intermediate participating in transport regulation or if it is a downstream metabolite such as G6P. If it was determined that direct phosphorylation of the transporter is affecting regulation, linker-scanning mutagenesis of the core transporter could identify the site of phosphorylation (Rivers & Oresnik, 2015). There are many different methods employed by organisms to regulate uptake of substrates by ABC transporters, including concentration dependent binding or dissociation of the substrate-binding protein from the core ABC transporter (Bao & Duong, 2012), or global regulation of ABC transporters by phosphotransferase system components (Prell *et al.*, 2012; Untiet *et al.*, 2013). More experiments are necessary to elucidate the regulatory mechanisms at play during transport of these substrates in *S. meliloti*.

Analysis of the *smo* operon led to the observation that a *smoS* mutant exhibits reduced growth when using D-arabinose as a sole carbon source (Table 5.3). This was noted as a peculiarity since every substrate shown to induce or be metabolized at this locus was a sugar alcohol substrate (Jacob *et al.*, 2008; Mauchline *et al.*, 2006). It was shown that D-arabinose is first reduced by a D-arabinose reductase encoded by *SMc00680* into D-arabitol, which is than acted on by MtlK and XylB to enter central metabolism (Fig. 5.2).


A.
FrcB
SmoE
Periplasm
Membrane
Cytoplasm
FrcC
FrcC
SmoF
SmoG
FrcA
FrcA
SmoK
SmoK
FrcK
FrcK
Frk
SmoS/MtlK
sorbitol/mannitol
fructose
fructose-6-phosphate
SmoC
Repression of smo genes
i.
smoC
smoE
smoF
smoG
smoK
smoS
mtlK
Expression of smo genes
ii.
smoC
smoE
smoF
smoG
smoK
smoS
mtlK

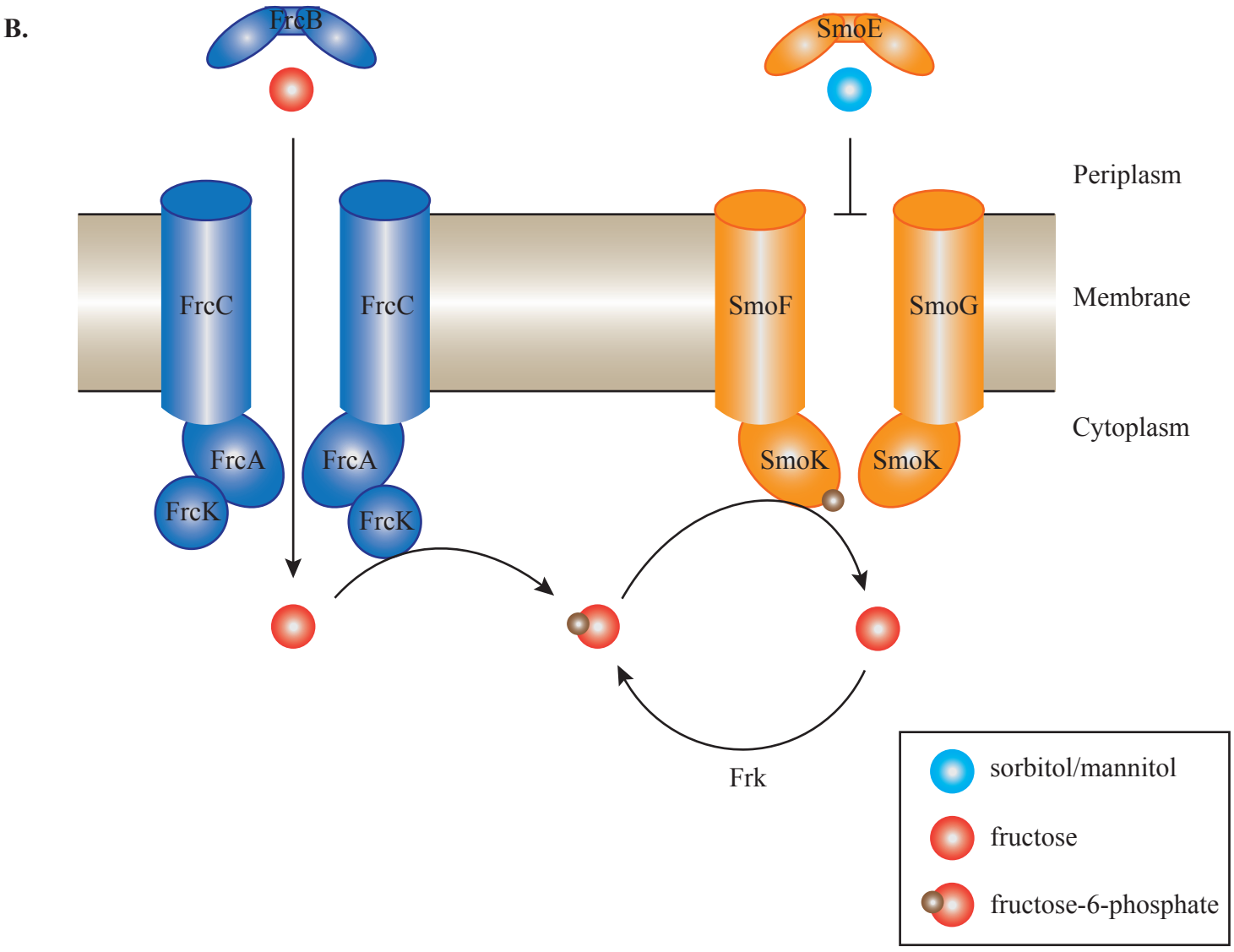


Figure 7.1 Possible models for the regulation of mannitol uptake involving F6P. (A) Fructose is internalized and phosphorylated by FrcK or Frk. (i) The presence of F6P activates expression of the SmoC negative regulator, which represses transcription of the *smo* operon and inhibits transport. (ii) The absence of F6P reduces expression of SmoC, which permits expression of the *smo* operon and allows for transport of sugar alcohols. (B) Fructose is internalized and phosphorylated by FrcK. Subsequently, F6P phosphorylates a regulatory region on the core transporter and inhibits transport. The regulatory domain is depicted as fused to the nucleotide-binding domain (NBD; SmoK), however it could be found on the transmembrane domain (TMD; SmoFG).

However, D-arabinose can also be degraded via an entirely separate, non-phosphorylative pathway that it shares with L-fucose (Fig. 5.2).

L-fucose is a methylpentose, which has been suggested to influence the ability of enteric bacteria to colonize the gut (Pacheco *et al.*, 2012; Stahl *et al.*, 2011). Another methylpentose, L-rhamnose has been shown to be a determinate of competition for nodule occupancy in *S. meliloti* and *R. leguminosrum* (Oresnik *et al.*, 1998). While the results were somewhat ambiguous, we showed that there was a difference in the dry weights of alfalfa plants inoculated with *S. meliloti* L-fucose catabolic mutants suggesting that they may be deficient in the ability to fix nitrogen (Fig. 5.6). Future experiments should attempt to confirm these results as well as determine if these mutants are compromised for the ability to compete for nodule occupancy.

The final objective for this thesis was to analyze a spontaneously generated xylitol utilizing suppressor mutant of *S. meliloti*. We determined that there are two xylitol dehydrogenases, SMc01991 and SMc01992, which are differentially expressed in the mutant (Fig. 6.6). Additionally it was shown that this expression could be correlated with the presence of an IS*Rm2011-2* IS element that has inserted downstream from the dehydrogenase genes (Fig. 6.5).

We showed that the insertion permitting utilization of xylitol behaved as a *cis* acting element (Table 6.4) and hypothesized that this may be in part due to an alteration of chromosome packaging arising from insertion of the IS element at that particular locus. Chromosome packaging has been shown to effect gene expression in bacteria (Le *et al.*, 2013). It was hoped that mutagenesis with Tn*5* would reveal other loci that were packed in close proximity with *SMc01990*, in which insertions would yield a similar growth phenotype. However, no other loci were detected via this method.

An alternative hypothesis is that the IS element is increasing the stability of the mRNA transcribed from the region, which is allowing for increased expression of SMc01991 and SMc01992. An interesting experiment would be to measure the abundance of mRNA isolated from the wildtype and the suppressor mutant with qRT-PCR at various timepoints to determine if the degradation rate of the transcript differed between the strains.

Literature cited

Alexeyev, M. F. (1999). The pKNOCK series of broad-host-range mobilizable suicide vectors for gene knockout and targeted DNA insertion into the chromosome of gram-negative bacteria. *Biotechniques* 26, 824-828.

Amor, B. B., Shaw, S. L., Oldroyd, G. E. D., Maillet, F., Penmetsa, R. V., Cook, D., Long, S. R., Dénarié, J. & Gough, C. (2003). The NFP locus of *Medicago truncatula* controls an early step of Nod factor signal transduction upstream of a rapid calcium flux and root hair deformation. *The Plant Journal* 34, 495-506.

Ampomah, O. Y., Avetisyan, A., Hansen, E., Svenson, J., Huser, T., Jensen, J. B. & Bhuvaneswari, T. V. (2013). The *thuEFGKAB* operon of Rhizobia and *Agrobacterium tumefaciens* codes for transport of trehalose, maltitol, and isomers of sucrose and their assimilation through the formation of their 3-keto derivatives. *J Bacteriol* 195, 3797.

Anderson, R. L. & Sapico, V. L. (1975). D-fructose (D-mannose) kinase. *Methods Enzymol* 42, 39-43.

Andriankaja, A., Boisson-Dernier, A., Frances, L., Sauviac, L., Jauneau, A., Barker, D. G. & de Carvalho-Niebel, F. (2007). AP2-ERF transcription factors mediate Nod Factor–dependent *MtENOD11*activation in root hairs via a novel *cis*-regulatory motif. *The Plant Cell* 19, 2866.

Ané, J.-M., Kiss, G. B., Riely, B. K., Penmetsa, R. V., Oldroyd, G. E. D., Ayax, C., Lévy, J., Debellé, F., Baek, J.-M. & other authors (2004). *Medicago truncatula DMI1* required for bacterial and fungal symbioses in legumes. *Science* 303, 1364.

Aneja, P. & Charles, T. C. (1999). Poly-3-hydroxybutyrate degradation in *Rhizobium* (*Sinorhizobium*) *meliloti*: Isolation and characterization of a gene encoding 3-hydroxybutyrate dehydrogenase. *J Bacteriol* 181, 849-857.

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

Aneja, P., Dai, M., Lacorre, D. A., Pillon, B. & Charles, T. C. (2004). Heterologous complementation of the exopolysaccharide synthesis and carbon utilization phenotypes of *Sinorhizobium meliloti* Rm1021 polyhydroxyalkanoate synthesis mutants. *FEMS Microbiol Lett* 239, 277-283.

Ardourel, M., Demont, N., Debellé, F., Maillet, F., de Billy, F., Promé, J. C., Dénarié, J. & Truchet, G. (1994). *Rhizobium meliloti* lipooligosaccharide nodulation factors: different structural requirements for bacterial entry into target root hair cells and induction of plant symbiotic developmental responses. *The Plant cell* 6, 1357-1374.

Arias, A. & Martinez-Drets, G. (1976). Glycerol metabolism in *Rhizobium*. *Can J Microbiol* 22, 150-153.

Arias, A., Cerveńansky, C., Gardiol, A. & Martínez-Drets, G. (1979). Phosphoglucose isomerase mutant of *Rhizobium meliloti*. *J Bacteriol* 137, 409-414.

Baier, M. C., Barsch, A., Küster, H. & Hohnjec, N. (2007). Antisense repression of the *Medicago truncatula* nodule-enhanced sucrose synthase leads to a handicapped nitrogen fixation mirrored by specific alterations in the symbiotic transcriptome and metabolome. *Plant Physiol* 145, 1600-1618.

Baldani, J. I. & Weaver, R. W. (1992). Survival of clover rhizobia and their plasmid-cured derivatives in soil under heat and drought stress. *Soil Biology and Biochemistry* 24, 737-742.

Baldani, J. I., Weaver, R. W., Hynes, M. F. & Eardly, B. D. (1992). Utilization of carbon substrates, electrophoretic enzyme patterns, and symbiotic performance of plasmid-cured clover rhizobia. *Appl Environ Microb* 58, 2308-2314.

Bao, H. & Duong, F. (2012). Discovery of an auto-regulation mechanism for the maltose ABC transporter MalFGK2. *PloS One* 7, e34836.

Barbier, T., Collard, F., Zúñiga-Ripa, A., Moriyón, I., Godard, T., Becker, J., Wittmann, C., Van Schaftingen, E. & Letesson, J.-J. (2014). Erythritol feeds the pentose phosphate pathway via three new isomerases leading to D-erythrose-4-phosphate in *Brucella*. *Proc Natl Acad Sci* 111, 17815.

Barker, D. G., Bianchi, S., Blondon, F., Dattée, Y., Duc, G., Essad, S., Flament, P., Gallusci, P., Génier, G. & other authors (1990). *Medicago truncatula*, a model plant for studying the molecular genetics of the *Rhizobium*-legume symbiosis. *Plant Molecular Biology Reporter* 8, 40-49.

Bensmihen, S., de Billy, F. & Gough, C. (2011). Contribution of NFP LysM domains to the recognition of Nod Factors during the *Medicago truncatula/Sinorhizobium meliloti* symbiosis. *PloS One* 6, e26114.

Bergey's manual of determinative bacteriology (7th ed.). (1964). *Am J Public Health Nations Health* 54, 544-544.

Bernardelli, C. E., Luna, M. F., Galar, M. L. & Boiardi, J. L. (2008). Symbiotic phenotype of a membrane-bound glucose dehydrogenase mutant of *Sinorhizobium meliloti*. *Plant Soil* 313, 217.

Bieleski, R. L. (1982). Sugar Alcohols. In *Plant Carbohydrates I: Intracellular Carbohydrates*, pp. 158-192. Edited by F. A. Loewus & W. Tanner. Berlin, Heidelberg: Springer Berlin Heidelberg.

Biemans-Oldehinkel, E., Doeven, M. K. & Poolman, B. (2006). ABC transporter architecture and regulatory roles of accessory domains. *FEBS Letters* 580, 1023-1035.

Bittner, A. N. & Oke, V. (2006). Multiple *groESL* operons are not key targets of RpoH1 and RpoH2 in *Sinorhizobium meliloti*. *J Bacteriol* 188, 3507.

Bolton, E., Higgisson, B., Harrington, A. & O'Gara, F. (1986). Dicarboxylic acid transport in *Rhizobium meliloti*: isolation of mutants and cloning of dicarboxylic acid transport genes. *Arch Microbiol* 144, 142-146.

Börnke, F., Hajirezaei, M. & Sonnewald, U. (2001). Cloning and characterization of the gene cluster for palatinose metabolism from the phytopathogenic bacterium *Erwinia rhapontici*. *J Bacteriol* 183, 2425.

Boyd, E. S. & Peters, J. W. (2013). New insights into the evolutionary history of biological nitrogen fixation. *Frontiers in microbiology* 4, 201-201.

Breedveld, M. W., Cremers, H. C., Batley, M., Posthumus, M. A., Zevenhuizen, L. P., Wijffelman, C. A. & Zehnder, A. J. (1993). Polysaccharide synthesis in relation to nodulation behavior of *Rhizobium leguminosarum*. *J Bacteriol* 175, 750.

Brewin, N. J. (2004). Plant Cell Wall Remodelling in the Rhizobium–Legume Symbiosis. *Critical Reviews in Plant Sciences* 23, 293-316.

Brinkkötter, A., Shakeri-Garakani, A. & Lengeler, J. W. (2002). Two class II D-tagatose-bisphosphate aldolases from enteric bacteria. *Arch Microbiol* 177, 410-419.

Brinkkötter, A., Klöß, H., Alpert, C. A. & Lengeler, J. W. (2000). Pathways for the utilization of N-acetyl-galactosamine and galactosamine in *Escherichia coli*. *Mol Microbiol* 37, 125-135.

Brom, S., García de los Santos, A., Stepkowsky, T., Flores, M., Dávila, G., Romero, D. & Palacios, R. (1992). Different plasmids of *Rhizobium leguminosarum* bv. phaseoli are required for optimal symbiotic performance. *J Bacteriol* 174, 5183-5189.

Burks, D., Azad, R., Wen, J. & Dickstein, R. (2018). The *Medicago truncatula* Genome: Genomic Data Availability. In *Functional Genomics in Medicago truncatula: Methods and Protocols*, pp. 39-59. Edited by L. A. Cañas & J. P. Beltrán. New York, NY: Springer New York.

Burrus, V. & Waldor, M. K. (2004). Shaping bacterial genomes with integrative and conjugative elements. *Research in Microbiology* 155, 376-386.

Cabanes, D., Boistard, P. & Batut, J. (2000). Symbiotic induction of pyruvate dehydrogenase genes from *Sinorhizobium meliloti*. *Mol Plant Microbe In* 13, 483-493.

Campbell, G. R. O., Reuhs, B. L. & Walker, G. C. (2002). Chronic intracellular infection of alfalfa nodules by *Sinorhizobium meliloti* requires correct lipopolysaccharide core. *Proc Natl Acad Sci USA* 99, 3938-3943.

Camus, J. C., Pryor, M. J., Medigue, C. & Cole, S. T. (2002). Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* 148, 2967-2973.

Cañas, L. A. & Beltrán, J. P. (2018). Model legumes: functional genomics tools in *Medicago truncatula*. In *Functional Genomics in Medicago truncatula: Methods and Protocols*, pp. 11-37. Edited by L. A. Cañas & J. P. Beltrán. New York, NY: Springer New York.

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

Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics (Oxford, England)* 25, 1972-1973.

Cárdenas, L., Dominguez, J., Quinto, C., Lopez-Lara, I. M., Lugtenberg, B. J., Spaink, H. P., Rademaker, G. J., Haverkamp, J. & Thomas-Oates, J. E. (1995). Isolation, chemical structures and biological activity of the lipo-chitin oligosaccharide nodulation signals from *Rhizobium etli*. *Plant molecular biology* 29, 453-464.

Carius, Y., Christian, H., Faust, A., Zander, U., Klink, B. U., Kornberger, P., Kohring, G.-W., Giffhorn, F. & Scheidig, A. J. (2010). Structural insight into substrate differentiation of the sugar-metabolizing enzyme galactitol dehydrogenase from *Rhodobacter sphaeroides D*. *J Biol Chem* 285, 20006-20014.

Carvalho, H., Sunkel, C., Salema, R. & Cullimore, J. V. (1997). Heteromeric assembly of the cytosolic glutamine synthetase polypeptides of *Medicago truncatula*: complementation of a *glnA Escherichia coli* mutant with a plant domain-swapped enzyme. *Plant molecular biology* 35, 623-632.

Carvalho, H., Lescure, N., de Billy, F., Chabaud, M., Lima, L., Salema, R. & Cullimore, J. (2000). Cellular expression and regulation of the *Medicago truncatula* cytosolic glutamine synthetase genes in root nodules. *Plant molecular biology* 42, 741-756.

Casacuberta, E. & González, J. (2013). The impact of transposable elements in environmental adaptation. *Molecular Ecology* 22, 1503-1517.

Catoira, R., Galera, C., de Billy, F., Penmetsa, R. V., Journet, E. P., Maillet, F., Rosenberg, C., Cook, D., Gough, C. & other authors (2000). Four genes of *Medicago truncatula* controlling components of a Nod Factor transduction pathway. *The Plant cell* 12, 1647-1666.

Chapman, J., Ismail, A. E. & Dinu, C. Z. (2018). Industrial applications of enzymes: Recent advances, techniques, and outlooks. *Catalysts* 8, 238.

Charles, T. C. & Finan, T. M. (1990). Genetic map of *Rhizobium meliloti* megaplasmid pRmeSU47b. *J Bacteriol* 172, 2469-2476.

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

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

Chen, I.-M. A., Markowitz, V. M., Palaniappan, K., Szeto, E., Chu, K., Huang, J., Ratner, A., Pillay, M., Hadjithomas, M. & other authors (2016). Supporting community annotation and user collaboration in the integrated microbial genomes (IMG) system. *BMC Genomics* 17, 307.

Chen, W.-M., Moulin, L., Bontemps, C., Vandamme, P., Béna, G. & Boivin-Masson, C. (2003). Legume Symbiotic Nitrogen Fixation by β-Proteobacteria Is Widespread in Nature. *J Bacteriol* 185, 7266.

Clark, S. R. D., Oresnik, I. J. & Hynes, M. F. (2001). RpoN of *Rhizobium leguminosarum* bv. *viciae* strain VF39SM plays a central role in FnrN-dependent microaerobic regulation of genes involved in nitrogen fixation. *Mol Gen Genet* 264, 623-633.

Cold Spring Harbor Protocols (2006). LB (Luria-Bertani) liquid medium. *Cold Spring Harb Protoc* 2006, pdb.rec8141.

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S. & other authors (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537-544.

Combier, J.-P., Frugier, F., de Billy, F., Boualem, A., El-Yahyaoui, F., Moreau, S., Vernié, T., Ott, T., Gamas, P. & other authors (2006). *MtHAP2-1* is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. *Genes & Development* 20, 3084-3088.

Combs, S. A., DeLuca, S. L., DeLuca, S. H., Lemmon, G. H., Nannemann, D. P., Nguyen, E. D., Willis, J. R., Sheehan, J. H. & Meiler, J. (2013). Small-molecule ligand docking into comparative models with Rosetta. *Nat Protoc* 8, 1277.

Compton, K. K., Hildreth, S. B., Helm, R. F. & Scharf, B. E. (2018). *Sinorhizobium meliloti* chemoreceptor McpV senses short-chain carboxylates via direct binding. *J Bacteriol* 200, e00519-00518.

Contador, C. A., Rodriguez, V., Andrews, B. A. & Asenjo, J. A. (2015). Genome-scale reconstruction of *Salinispora tropica* CNB-440 metabolism to study strain-specific adaptation. *Antonie van Leeuwenhoek* 108, 1075-1090.

Cook, D. R. (1999). *Medicago truncatula* — a model in the making! *Current Opinion in Plant Biology* 2, 301-304.

Cowie, A., Cheng, J., Sibley, C. D., Fong, Y., Zaheer, R., Patten, C. L., Morton, R. M., Golding, G. B. & Finan, T. M. (2006). An integrated approach to functional genomics:

construction of a novel reporter gene fusion library for *Sinorhizobium meliloti*. *Appl Environ Microb* 72, 7156-7167.

Curatti, L. & Rubio, L. M. (2014). Challenges to develop nitrogen-fixing cereals by direct *nif*-gene transfer. *Plant Science* 225, 130-137.

Darling, A. C., Mau, B., Blattner, F. R. & Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome research* 14.

Darmon, E. & Leach, D. R. F. (2014). Bacterial Genome Instability. *Microbiol Mol Biol Rev* 78, 1-39.

Davidson, E. A., Suddick, E. C., Rice, C. W. & Prokopy, L. S. (2015). More Food, Low Pollution (Mo Fo Lo Po): A Grand Challenge for the 21st Century. *Journal of Environmental Quality* 44, 305-311.

Davies, B. W. & Walker, G. C. (2007). Disruption of *sitA* compromises *Sinorhizobium meliloti* for manganese uptake required for protection against oxidative stress. *J Bacteriol* 189, 2101.

Day, D. A., Poole, P. S., Tyerman, S. D. & Rosendahl, L. (2001). Ammonia and amino acid transport across symbiotic membranes in nitrogen-fixing legume nodules. *Cellular and Molecular Life Sciences CMLS* 58, 61-71.

de Werra, P., Péchy-Tarr, M., Keel, C. & Maurhofer, M. (2009). Role of gluconic acid production in the regulation of biocontrol traits of *Pseudomonas fluorescens* CHA0. *Appl Environ Microb* 75, 4162-4174.

Delgado-Baquerizo, M., Oliverio, A. M., Brewer, T. E., Benavent-González, A., Eldridge, D. J., Bardgett, R. D., Maestre, F. T., Singh, B. K. & Fierer, N. (2018). A global atlas of the dominant bacteria found in soil. *Science* 359, 320.

DeLuca, S., Khar, K. & Meiler, J. (2015). Fully Flexible Docking of Medium Sized Ligand Libraries with RosettaLigand. *PloS One* 10, e0132508.

diCenzo, G., Milunovic, B., Cheng, J. & Finan, T. M. (2013). The tRNAarg Gene and *engA* Are Essential Genes on the 1.7-Mb pSymB Megaplasmid of *Sinorhizobium meliloti* and Were Translocated Together from the Chromosome in an Ancestral Strain. *J Bacteriol* 195, 202-212.

diCenzo, G. C. & Finan, T. M. (2017). The divided bacterial genome: Structure, function, and evolution. *Microbiol Mol Biol Rev* 81, e00019-00017.

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

diCenzo, G. C., Benedict, A. B., Fondi, M., Walker, G. C., Finan, T. M., Mengoni, A. & Griffitts, J. S. (2018). Robustness encoded across essential and accessory replicons of the ecologically versatile bacterium Sinorhizobium meliloti. *PLOS Genetics* 14, e1007357.

diCenzo, G. C., Checcucci, A., Bazzicalupo, M., Mengoni, A., Viti, C., Dziewit, L., Finan, T. M., Galardini, M. & Fondi, M. (2016). Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. *Nat Commun* 7, 12219.

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

Ding, H., Yip, C. B., Geddes, B. A., Oresnik, I. J. & Hynes, M. F. (2012). Glycerol utilization by *Rhizobium leguminosarum* requires an ABC transporter and affects competition for nodulation. *Microbiology* 158, 1369-1378.

Dixon, R. & Kahn, D. (2004). Genetic regulation of biological nitrogen fixation. *Nat Rev Micro* 2, 621-631.

Dobrindt, U., Hochhut, B., Hentschel, U. & Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* 2, 414-424.

Domínguez-Ferreras, A., Soto, M. J., Pérez-Arnedo, R., Olivares, J. & Sanjuán, J. (2009). Importance of trehalose biosynthesis for *Sinorhizobium meliloti* osmotolerance and nodulation of alfalfa roots. *J Bacteriol* 191, 7490-7499.

Downie, J. A. & Young, J. P. W. (2001). The ABC of symbiosis. *Nature* 412, 597-598.

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

Driscoll, B. T. & Finan, T. M. (1996). NADP+ -dependent malic enzyme of *Rhizobium meliloti*. *J Bacteriol* 178, 2224-2231.

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

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

Eady, R. R., Robson, R. L., Richardson, T. H., Miller, R. W. & Hawkins, M. (1987). The vanadium nitrogenase of *Azotobacter chroococcum*. Purification and properties of the VFe protein. *Biochemical Journal* 244, 197.

Edelstein, S., Smith, K., Worthington, A., Gillis, N., Bruen, D., Kang, S. H., Ho, W. L., Gilpin, K., Ackerman, J. & other authors (2007). Comparisons of six new artificial sweetener

gradation ratios with sucrose in conventional-method cupcakes resulting in best percentage substitution ratios. *Journal of Culinary Science & Technology* 5, 61-74.

Einsle, O., Tezcan, F. A., Andrade, S. L. A., Schmid, B., Yoshida, M., Howard, J. B. & Rees, D. C. (2002). Nitrogenase MoFe-Protein at 1.16 Å Resolution: A Central Ligand in the FeMo-Cofactor. *Science* 297, 1696.

Elsinghorst, E. A. & Mortlock, R. P. (1988). D-arabinose metabolism in *Escherichia coli* B: induction and cotransductional mapping of the L-fucose-D-arabinose pathway enzymes. *J Bacteriol* 170, 5423-5432.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66, 486-501.

Endre, G., Kereszt, A., Kevei, Z., Mihacea, S., Kaló, P. & Kiss, G. B. (2002). A receptor kinase gene regulating symbiotic nodule development. *Nature* 417, 962-966.

Ensor, M., Banfield, A. B., Smith, R. R., Williams, J. & Lodder, R. A. (2015). Safety and efficacy of D-tagatose in glycemic control in subjects with type 2 diabetes. *J Endocrinol Diabetes Obes* 3, 1065.

Erisman, J. W., Sutton, M. A., Galloway, J., Klimont, Z. & Winiwarter, W. (2008). How a century of ammonia synthesis changed the world. *Nat Geosci* 1, 636.

Espinosa, I. & Fogelfeld, L. (2010). Tagatose: from a sweetener to a new diabetic medication? *Expert Opin Investig Drugs* 19, 285-294.

Esseling, J. J., Lhuissier, F. G. P. & Emons, A. M. C. (2004). A nonsymbiotic root hair tip growth phenotype in *NORK*-mutated legumes: Implications for Nodulation Factor–induced signaling and formation of a multifaceted root hair pocket for bacteria. *The Plant Cell* 16, 933.

Evans, P. R. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr* 67, 282-292.

Feist, A. M., Herrgard, M. J., Thiele, I., Reed, J. L. & Palsson, B. O. (2009). Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* 7, 129-143.

Fennington, G. J., Jr. & Hughes, T. A. (1996). The fructokinase from *Rhizobium leguminosarum* biovar *trifolii* belongs to group I fructokinase enzymes and is encoded separately from other carbohydrate metabolism enzymes. *Microbiology* 142 (Pt 2), 321-330.

Ferguson, B. J., Indrasumunar, A., Hayashi, S., Lin, M. H., Lin, Y. H., Reid, D. E. & Gresshoff, P. M. (2010). Molecular analysis of legume nodule development and autoregulation. *Journal of integrative plant biology* 52, 61-76.

Ferguson, B. J., Mens, C., Hastwell, A. H., Zhang, M., Su, H., Jones, C. H., Chu, X. & Gresshoff, P. M. (2019). Legume nodulation: The host controls the party. *Plant, cell & environment* 42, 41-51.

Ferguson, G. P., Datta, A., Baumgartner, J., Roop, R. M., Carlson, R. W. & Walker, G. C. (2004). Similarity to peroxisomal-membrane protein family reveals that *Sinorhizobium* and *Brucella* BacA affect lipid-A fatty acids. *Proc Natl Acad Sci USA* 101, 5012.

Filling, C., Berndt, K. D., Benach, J., Knapp, S., Prozorovski, T., Nordling, E., Ladenstein, R., Jornvall, H. & Oppermann, U. (2002). Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. *J Biol Chem* 277, 25677-25684.

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

Finan, T. M., Kunkel, B., De Vos, G. F. & Singer, E. R. (1986). Second Symbiotic Megaplasmid in *Rhizobium meliloti* Carrying Exopolysaccharide and Thiamine Synthesis Genes. *J Bacteriol* 167, 66-72.

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

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

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

Finan, T. M., Weidner, S., Wong, K., Buhrmester, J., Chain, P., Vorhölter, F. J., Hernandez-Lucas, I., Becker, A., Cowie, A. & other authors (2001). The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endosymbiont *Sinorhizobium meliloti*. *Proc Natl Acad Sci* 98, 9889-9894.

Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., Chang, H.-Y., Dosztányi, Z., El-Gebali, S. & other authors (2017). InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res* 45, D190-D199.

Fisher, R. F., Egelhoff, T. T., Mulligan, J. T. & Long, S. R. (1988). Specific binding of proteins from *Rhizobium meliloti* cell-free extracts containing NodD to DNA sequences upstream of inducible nodulation genes. *Genes & Development* 2, 282-293.

Fliegmann, J., Jauneau, A., Pichereaux, C., Rosenberg, C., Gasciolli, V., Timmers, A. C. J., Burlet-Schiltz, O., Cullimore, J. & Bono, J.-J. (2016). LYR3, a high-affinity LCO-binding protein of *Medicago truncatula*, interacts with LYK3, a key symbiotic receptor. *FEBS Letters* 590, 1477-1487.

Forneris, F., Heuts, D. P. H. M., Delvecchio, M., Rovida, S., Fraaije, M. W. & Mattevi, A. (2008). Structural Analysis of the Catalytic Mechanism and Stereoselectivity in Streptomyces coelicolor Alditol Oxidase. *Biochemistry* 47, 978-985.

Fougère, F., Le Rudulier, D. & Streeter, J. G. (1991). Effects of salt stress on amino acid, organic acid, and carbohydrate composition of roots, bacteroids, and cytosol of alfalfa (*Medicago sativa* L.). *Plant Physiol* 96, 1228-1236.

Fowler, D., Coyle, M., Skiba, U., Sutton, M. A., Cape, J. N., Reis, S., Sheppard, L. J., Jenkins, A., Grizzetti, B. & other authors (2013). The global nitrogen cycle in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368, 20130164.

Fredslund, F., Otten, H., Gemperlein, S., Poulsen, J.-C. N., Carius, Y., Kohring, G.-W. & Lo Leggio, L. (2016). Structural characterization of the thermostable *Bradyrhizobium japonicum* D-sorbitol dehydrogenase. *Acta Crystallogr F* 72, 846-852.

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

Fujinami, S. & Fujisawa, M. (2010). Industrial applications of alkaliphiles and their enzymes--past, present and future. *Environ Technol* 31, 845-856.

Gage, D. J. & Long, S. R. (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. *J Bacteriol* 180, 5739-5748.

Gajdzik, J., Lenz, J., Natter, H., Kohring, G.-W., Giffhorn, F., Wenz, G. & Hempelmann, R. (2010). Directed immobilisation of modificated galactitol-dehydrogenase on gold electrodes for electrochemical cofactor regeneration. *ECS Trans* 25, 13-20.

Gajdzik, J., Lenz, J., Natter, H., Walcarius, A., Kohring, G. W., Giffhorn, F., Göllü, M., Demir, A. S. & Hempelmann, R. (2011). Electrochemical screening of redox mediators for electrochemical regeneration of NADH. *J Electrochem Soc* 159, F10-F16.

Galibert, F., Finan, T. M., Long, S. R., Pühler, A., Abola, P., Ampe, F., Barloy-Hubler, F., Barnett, M. J., Becker, A. & other authors (2001). The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* 293, 668-672.

Galloway, J. N., Aber, J. D., Erisman, J. W., Seitzinger, S. P., Howarth, R. W., Cowling, E. B. & Cosby, B. J. (2003). The Nitrogen Cascade. *BioScience* 53, 341.

Galloway, J. N., Townsend, A. R., Erisman, J. W., Bekunda, M., Cai, Z., Freney, J. R., Martinelli, L. A., Seitzinger, S. P. & Sutton, M. A. (2008). Transformation of the nitrogen cycle: recent trends, questions, and potential solutions. *Science* 320, 889-892.

Gardiol, A., Arias, A., Cerveñansky, C., Gaggero, C. & Martínez-Drets, G. (1980). Biochemical characterization of a fructokinase mutant of *Rhizobium meliloti*. *J Bacteriol* 144, 12-16.

Gasteiger E., H. C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, pp. 571-607. Edited by J. M. Walker: Humana Press.

Gauer, S., Wang, Z., Otten, H., Etienne, M., Bjerrum, M. J., Lo Leggio, L., Walcarius, A., Giffhorn, F. & Kohring, G.-W. (2014). An L-glucitol oxidizing dehydrogenase from *Bradyrhizobium japonicum* USDA 110 for production of D-sorbose with enzymatic or electrochemical cofactor regeneration. *Appl Microbiol Biot* 98, 3023-3032.

Gautrat, P., Mortier, V., Laffont, C., De Keyser, A., Fromentin, J., Frugier, F. & Goormachtig, S. (2019). Unraveling new molecular players involved in the autoregulation of nodulation in *Medicago truncatula*. *J Exp Bot*.

Geddes, B. A. & Oresnik, I. J. (2012a). Inability to catabolize galactose leads to increased ability to compete for nodule occupancy in *Sinorhizobium meliloti*. *J Bacteriol* 194, 5044-5053.

Geddes, B. A. & Oresnik, I. J. (2012b). Genetic characterization of a complex locus necessary for the transport and catabolism of erythritol, adonitol and L-arabitol in *Sinorhizobium meliloti*. *Microbiology* 158, 2180-2191.

Geddes, B. A. & Oresnik, I. J. (2014). Physiology, genetics, and biochemistry of carbon metabolism in the alphaproteobacterium *Sinorhizobium meliloti*. *Can J Microbiol* 60, 491-507.

Geddes, B. A. & Oresnik, I. J. (2016). The mechanism of symbiotic nitrogen fixation. In *The Mechanistic Benefits of Microbial Symbionts* (Advances in Environmental Microbiology), 1 edn, pp. 69-97. Edited by C. J. Hurst: Springer International Publishing.

Geddes, B. A., González, J. E. & Oresnik, I. J. (2014). Exopolysaccharide production in response to medium acidification is correlated with an increase in competition for nodule occupancy. *Mol Plant Microbe In* 27, 1307-1317.

Geddes, B. A., Pickering, B. S., Poysti, N. J., Collins, H., Yudistira, H. & Oresnik, I. J. (2010). A locus necessary for the transport and catabolism of erythritol in *Sinorhizobium meliloti*. *Microbiology* 156, 2970-2981.

Geddes, B. A., Paramasivan, P., Joffrin, A., Thompson, A. L., Christensen, K., Jorrin, B., Brett, P., Conway, S. J., Oldroyd, G. E. D. & other authors (2019). Engineering transkingdom signalling in plants to control gene expression in rhizosphere bacteria. *Nat Commun* 10, 3430.

Ghalambor, M. A. & Heath, E. C. (1962). The metabolism of L-fucose: II. The enzymatic cleavage of L-fuculose 1-phosphate. *J Biol Chem* 237, 2427-2433.

Gilles-Gonzalez, M. A., Ditta, G. S. & Helinski, D. R. (1991). A haemoprotein with kinase activity encoded by the oxygen sensor of *Rhizobium meliloti*. *Nature* 350, 170-172.

Giraud, E., Moulin, L., Vallenet, D., Barbe, V., Cytryn, E., Avarre, J.-C., Jaubert, M., Simon, D., Cartieaux, F. & other authors (2007). Legumes symbioses: absence of *nod* genes in photosynthetic bradyrhizobia. *Science* 316, 1307.

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

Glazebrook, J., Ichige, A. & Walker, G. C. (1993). A *Rhizobium meliloti* homolog of the *Escherichia coli* peptide-antibiotic transport protein SbmA is essential for bacteroid development. *Genes & Development* 7, 1485-1497.

Gonin, S., Arnoux, P., Pierru, B., Lavergne, J., Alonso, B., Sabaty, M. & Pignol, D. (2007). Crystal structures of an extracytoplasmic solute receptor from a TRAP transporter in its open and closed forms reveal a helix-swapped dimer requiring a cation for α-keto acid binding. *BMC Struct Biol* 7, 1-14.

Gonzalez-Rizzo, S., Crespi, M. & Frugier, F. (2006). The *Medicago truncatula* CRE1 cytokinin receptor regulates lateral root development and early symbiotic interaction with *Sinorhizobium meliloti*. *The Plant Cell* 18, 2680.

Green, M. & Cohen, S. S. (1956). Enzymatic conversion of L-fucose to L-fuculose. *J Biol Chem* 219, 557-568.

Güell, O., Sagués, F. & Serrano, M. Á. (2014). Essential plasticity and redundancy of metabolism unveiled by synthetic lethality analysis. *PLOS Computational Biology* 10, e1003637.

Guo, X., Flores, M., Mavingui, P., Fuentes, S. I., Hernández, G., Dávila, G. & Palacios, R. (2003). Natural genomic design in *Sinorhizobium meliloti*: Novel genomic architectures. *Genome research* 13, 1810-1817.

Haag, A. F., Arnold, M. F. F., Myka, K. K., Kerscher, B., Dall'Angelo, S., Zanda, M., Mergaert, P. & Ferguson, G. P. (2013). Molecular insights into bacteroid development during *Rhizobium*–legume symbiosis. *FEMS Microbiology Reviews* 37, 364-383.

Hall, B. G. (2000). Transposable elements as activators of cryptic genes in *E. coli*. *Genetica* 107, 181-187.

Hanahan, D. (1983). Studies on Transformation of *Eschericia coli* with Plasmids. *J Mol Biol* 166, 557-580.

Handberg, K. & Stougaard, J. (1992). *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. *The Plant Journal* 2, 487-496.

Haskett, T. L., Terpolilli, J. J., Bekuma, A., O'Hara, G. W., Sullivan, J. T., Wang, P., Ronson, C. W. & Ramsay, J. P. (2016). Assembly and transfer of tripartite integrative and conjugative genetic elements. *Proc Natl Acad Sci* 113, 12268-12273.

Hawkins, J. P., Ordonez, P. A. & Oresnik, I. J. (2018). Characterization of mutations that affect the nonoxidative pentose phosphate pathway in *Sinorhizobium meliloti*. *J Bacteriol* 200, e00436-00417.

Heath, E. C. & Ghalambor, M. A. (1962). The metabolism of L-fucose: I. The purification and properties of L-fuculose kinase. *J Biol Chem* 237, 2423-2426.

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

Heuts, D. P. H. M., van Hellemond, E. W., Janssen, D. B. & Fraaije, M. W. (2007). Discovery, Characterization, and Kinetic Analysis of an Alditol Oxidase from *Streptomyces coelicolor*. *J Biol Chem* 282, 20283-20291.

Higashi, S. (1967). Transfer of clover infectivity of *Rhizobium trifolii* to *Rhizobium phaseoli* as mediated by an episomic factor. *J Gen Appl Microbiol* 13, 391-403.

Hiraga, K., Kitazawa, M., Kaneko, N. & Oda, K. (1997). Isolation and Some Properties of Sorbitol Oxidase from *Streptomyces* sp. H-7775. *Biosci Biotech Bioch* 61, 1699-1704.

Hobbs, M. E., Vetting, M., Williams, H. J., Narindoshvili, T., Kebodeaux, D. M., Hillerich, B., Seidel, R. D., Almo, S. C. & Raushel, F. M. (2013). Discovery of an L-fucono-1,5-lactonase from cog3618 of the amidohydrolase superfamily. *Biochemistry* 52, 239-253.

Holm, L. & Laakso, L. M. (2016). Dali server update. *Nucleic Acids Res* 44, W351-355.

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

Honma, M. A., Asomaning, M. & Ausubel, F. M. (1990). *Rhizobium meliloti nodD* genes mediate host-specific activation of nodABC. *J Bacteriol* 172, 901.

House, B. L., Mortimer, M. W. & Kahn, M. L. (2004). New recombination methods for *Sinorhizobium meliloti* genetics. *Appl Environ Microb* 70, 2806-2815.

Hu, Y. & Ribbe, M. W. (2013). Nitrogenase assembly. *Biochimica et biophysica acta* 1827, 1112-1122.

Hynes, M. F. & O'Connell, M. P. (1990). Host plant effect on competition among strains of *Rhizobium leguminosarum*. *Can J Microbiol* 36, 864-869.

Hynes, M. F. & McGregor, N. F. (1990). Two plasmids other than the nodulation plasmid are necessary for formation of nitrogen-fixing nodules by *Rhizobium leguminosarum*. *Mol Microbiol* 4, 567-574.

Jacob, A. I., Adham, S. A. I., Capstick, D. S., Clark, S. R. D., Spence, T. & Charles, T. C. (2008). Mutational analysis of the *Sinorhizobium meliloti* short-chain dehydrogenase/reductase family reveals substantial contribution to symbiosis and catabolic diversity. *Mol Plant Microbe In* 21, 979-987.

Jagtap, S. S., Singh, R., Kang, Y. C., Zhao, H. & Lee, J.-K. (2014). Cloning and characterization of a galactitol 2-dehydrogenase from *Rhizobium legumenosarum* and its application in D-tagatose production. *Enzyme Microb Tech* 58-59, 44-51.

Jamet, A., Sigaud, S., Van de Sype, G., Puppo, A. & Hérouart, D. (2003). Expression of the bacterial catalase genes during Sinorhizobium meliloti-Medicago sativa symbiosis and their crucial role during the infection process. *Mol Plant Microbe In* 16, 217-225.

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

Jiang, G., Krishnan, A. H., Kim, Y.-W., Wacek, T. J. & Krishnan, H. B. (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.). *J Bacteriol* 183, 2595-2604.

Jiang, J., Gu, B. H., Albright, L. M. & Nixon, B. T. (1989). Conservation between coding and regulatory elements of *Rhizobium meliloti* and *Rhizobium leguminosarum dct* genes. *J Bacteriol* 171, 5244-5253.

Jiménez-Zurdo, J. I., van Dillewijn, P., Soto, M. J., de Felipe, M. R., Olivares, J. & Toro, N. (1995). Characterization of a *Rhizobium meliloti* proline dehydrogenase mutant altered in nodulation efficiency and competitiveness on alfalfa roots. *Molecular plant-microbe interactions : MPMI* 8, 492-498.

Joerger, R. D., Bishop, P. E. & Evans, H. J. (1988). Bacterial alternative nitrogen fixation systems. *CRC Critical Reviews in Microbiology* 16, 1-14.

Johnston, A. W. B., Beynon, J. L., Buchanan-Wollaston, A. V., Setchell, S. M., Hirsch, P. R. & Beringer, J. E. (1978). High frequency transfer of nodulating ability between strains and species of *Rhizobium*. *Nature* 276, 634-636.

Jones, D. L., Nguyen, C. & Finlay, R. D. (2009). Carbon flow in the rhizosphere: carbon trading at the soil–root interface. *Plant Soil* 321, 5-33.

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

Jones, K. M., Kobayashi, H., Davies, B. W., Taga, M. E. & Walker, G. C. (2007). How rhizobial symbionts invade plants: the *Sinorhizobium-Medicago* model. *Nat Rev Microbiol* 5, 619-633.

Jordan, E., Saedler, H. & Starlinger, P. (1968). O^{o} and strong-polar mutations in the *gal* operon and insertions. *Mol Gen Genet* 102, 353-363.

Jornvall, H., Persson, M. & Jeffery, J. (1981). Alcohol and polyol dehydrogenases are both divided into two protein types, and structural properties cross-relate the different enzyme activities within each type. *Proc Natl Acad Sci U S A* 78, 4226-4230.

Kabsch, W. (2010). XDS. *Acta Crystallogr D Biol Crystallogr* 66, 125-132.

Kaló, P., Gleason, C., Edwards, A., Marsh, J., Mitra, R. M., Hirsch, S., Jakab, J., Sims, S., Long, S. R. & other authors (2005). Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. *Science* 308, 1786.

Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M. & other authors (2002). Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA research : an international journal for rapid publication of reports on genes and genomes* 9.

Kaufmann, K. W. & Meiler, J. (2012). Using RosettaLigand for small molecule docking into comparative models. *PloS One* 7, e50769-e50769.

Kavanagh, K., Jörnvall, H., Persson, B. & Oppermann, U. (2008). The SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell Mol Life Sci* 65, 3895-3906.

Kereszt, A., Slaska-Kiss, K., Putnoky, P., Banfalvi, Z. & Kondorosi, A. (1995). The *cycHJKL* genes of *Rhizobium meliloti* involved in cytochrome c biogenesis are required for "respiratory" nitrate reduction *ex planta* and for nitrogen fixation during symbiosis. *Mol Gen Genet* 247, 39-47.

Kirn, J. & Rees, D. C. (1992). Crystallographic structure and functional implications of the nitrogenase molybdenum–iron protein from *Azotobacter vinelandii*. *Nature* 360, 553-560.

Kohler, P. R. A., Zheng, J. Y., Schoffers, E. & Rossbach, S. (2010). Inositol catabolism, a key pathway in *Sinorhizobium meliloti* for competitive host nodulation. *Appl Environ Microb* 76, 7972.

Kohlmeier, M. G., White, C. E., Fowler, J. E., Finan, T. M. & Oresnik, I. J. (2019). Galactitol catabolism in *Sinorhizobium meliloti* is dependent on a chromosomally encoded sorbitol dehydrogenase and a pSymB-encoded operon necessary for tagatose catabolism. *Mol Genet Genomics*.

Kornberger, P., Gajdzik, J., Natter, H., Wenz, G., Giffhorn, F., Kohring, G. W. & Hempelmann, R. (2009). Modification of galactitol dehydrogenase from *Rhodobacter sphaeroides* D for immobilization on polycrystalline gold surfaces. *Langmuir* 25, 12380-12386.

Kothiwale, S., Mendenhall, J. L. & Meiler, J. (2015). BCL::Conf: small molecule conformational sampling using a knowledge based rotamer library. *J Cheminformatics* 7, 47.

Koziol, U., Hannibal, L., Rodríguez, M. C., Fabiano, E., Kahn, M. L. & Noya, F. (2009). Deletion of citrate synthase restores growth of *Sinorhizobium meliloti* 1021 aconitase mutants. *J Bacteriol* 191, 7581-7586.

Kroer, N., Weber, D., Barkay, T. & Sørensen, S. (1998). Effect of root exudates and bacterial metabolic activity on conjugal gene transfer in the rhizosphere of a marsh plant. *FEMS Microbiol Ecol* 25, 375-384.

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

Lancaster, K. M., Roemelt, M., Ettenhuber, P., Hu, Y., Ribbe, M. W., Neese, F., Bergmann, U. & DeBeer, S. (2011). X-ray emission spectroscopy evidences a central carbon in the nitrogenase iron-molybdenum cofactor. *Science* 334, 974.

Lassaletta, L., Billen, G., Grizzetti, B., Anglade, J. & Garnier, J. (2014). 50 year trends in nitrogen use efficiency of world cropping systems: the relationship between yield and nitrogen input to cropland. *Environ Res Lett* 9, 105011.

Le, T. B. K., Imakaev, M. V., Mirny, L. A. & Laub, M. T. (2013). High-resolution mapping of the spatial organization of a bacterial chromosome. *Science* 342, 731-734.

LeBlanc, D. J. & Mortlock, R. P. (1971). Metabolism of D-arabinose: a new pathway in *Escherichia coli*. *J Bacteriol* 106, 90-96.

Lee, J.-K., Koo, B.-S. & Kim, S.-Y. (2003). Cloning and characterization of the *xyl1* gene, encoding an NADH-preferring xylose reductase from *Candida parapsilosis*, and Its functional expression in *Candida tropicalis*. *Appl Environ Microb* 69, 6179-6188.

Leigh, J. A., Signer, E. R. & Walker, G. C. (1985). Exopolysaccharide-deficient mutants of *Rhizobium meliloti* that form ineffective nodules. *Proc Natl Acad Sci USA* 82, 6231-6235.

Lengeler, J. (1975). Mutations affecting transport of the hexitols D-mannitol, D-glucitol, and galactitol in *Escherichia coli* K-12: isolation and mapping. *J Bacteriol* 124, 26-38.

Lengeler, J. (1977). Analysis of mutations affecting the dissimilation of galactitol (dulcitol) in *Escherichia coli* K12. *Mol Gen Genet* 152, 83-91.

Lévy, J., Bres, C., Geurts, R., Chalhoub, B., Kulikova, O., Duc, G., Journet, E.-P., Ané, J.-M., Lauber, E. & other authors (2004). A putative Ca2+ and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. *Science* 303, 1361.

Leyn, S. A., Gao, F., Yang, C. & Rodionov, D. A. (2012). N-acetylgalactosamine utilization pathway and regulon in proteobacteria: genomic reconstruction and experimental characterization in *Shewanella*. *J Biol Chem* 287, 28047-28056.

Liang, J. & Burris, R. H. (1988). Hydrogen burst associated with nitrogenase-catalyzed reactions. *Proc Natl Acad Sci* 85, 9446.

Liebschner, D., Afonine, P. V., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C. & Adams, P. D. (2017). Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr D Struct Biol* 73, 148-157.

Limpens, E., Franken, C., Smit, P., Willemse, J., Bisseling, T. & Geurts, R. (2003). LysM domain receptor kinases regulating rhizobial Nod Factor-induced infection. *Science* 302, 630.

Limpens, E., Mirabella, R., Fedorova, E., Franken, C., Franssen, H., Bisseling, T. & Geurts, R. (2005). Formation of organelle-like N2-fixing symbiosomes in legume root nodules is controlled by *DMI2*. *Proc Natl Acad Sci USA* 102, 10375.

Liu, A., Contador, C. A., Fan, K. & Lam, H.-M. (2018). Interaction and regulation of carbon, nitrogen, and phosphorus metabolisms in root nodules of legumes. *Front Plant Sci* 9.

Loewus, F. A. & Dickinson, D. B. (1982). Cyclitols. In *Plant Carbohydrates I: Intracellular Carbohydrates*, pp. 193-216. Edited by F. A. Loewus & W. Tanner. Berlin, Heidelberg: Springer Berlin Heidelberg.

Lu, Y., Levin, G. V. & Donner, T. W. (2008). Tagatose, a new antidiabetic and obesity control drug. *Diabetes Obes Metab* 10, 109-134.

Lyskov, S., Chou, F.-C., Conchúir, S. Ó., Der, B. S., Drew, K., Kuroda, D., Xu, J., Weitzner, B. D., Renfrew, P. D. & other authors (2013). Serverification of Molecular Modeling Applications: The Rosetta Online Server That Includes Everyone (ROSIE). *PloS One* 8, e63906.

MacLean, A. M., MacPherson, G., Aneja, P. & Finan, T. M. (2006). Characterization of the β-ketoadipate pathway in *Sinorhizobium meliloti*. *Appl Environ Microb* 72, 5403-5413.

MacLean, A. M., White, C. E., Fowler, J. E. & Finan, T. M. (2009). Identification of a hydroxyproline transport system in the legume endosymbiont *Sinorhizobium meliloti*. *Mol Plant Microbe In* 22, 1116-1127.

Maguire, A. & Rugg-Gunn, A. J. (2003). Xylitol and caries prevention — is it a magic bullet? *British Dental Journal* 194, 429-436.

Malamy, M. H. (1970). Some properties of insertion mutations in the *lac* operon. In *The lactose operon*, vol. 1, p. 359. Edited by J. R. Beckwith & D. Zipser: Cold Spring Harbor Laboratory Cold Spring Harbor, New York:.

Markmann, K., Radutoiu, S. & Stougaard, J. (2012). Infection of *Lotus japonicus* Roots by *Mesorhizobium loti*. In *Signaling and Communication in Plant Symbiosis*, pp. 31-50. Edited by S. Perotto & F. Baluška. Berlin, Heidelberg: Springer Berlin Heidelberg.

Marsh, J. F., Rakocevic, A., Mitra, R. M., Brocard, L., Sun, J., Eschstruth, A., Long, S. R., Schultze, M., Ratet, P. & other authors (2007). *Medicago truncatula NIN* is essential for rhizobial-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase. *Plant Physiol* 144, 324.

Martínez De Drets, G. & Arias, A. (1970). Metabolism of some polyols by *Rhizobium meliloti*. *J Bacteriol* 103, 97-103.

Masson-Boivin, C., Giraud, E., Perret, X. & Batut, J. (2009). Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? *Trends in Microbiology* 17, 458-466.

Matthews, B. W., Nicholson, H. & Becktel, W. J. (1987). Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc Natl Acad Sci* 84, 6663-6667.

Mauchline, T. H., Fowler, J. E., East, A. K., Sartor, A. L., Zaheer, R., Hosie, A. H. F., Poole, P. S. & Finan, T. M. (2006). Mapping the *Sinorhizobium meliloti* 1021 solute-binding protein-dependent transportome. *Proc Natl Acad Sci* 103, 17933-17938.

McClintock, B. (1956). Controlling elements and the gene. In *Cold Spring Harbor symposia on quantitative biology*, vol. 21, pp. 197-216.

McClintock, B. (1984). The significance of responses of the genome to challenge. *Science* 226, 792-801.

McDermott, T. R. & Kahn, M. L. (1992). Cloning and mutagenesis of the *Rhizobium meliloti* isocitrate dehydrogenase gene. *J Bacteriol* 174, 4790-4797.

Meade, H. M. & Signer, E. R. (1977). Genetic mapping of *Rhizobium meliloti*. *Proc Natl Acad Sci* 74, 2076-2078.

Meade, H. M., Long, R. S., Ruvkun, G. B., Brown, S. E. & Ausubel, F. M. (1982). Physical and genetic characterization of symbiotic and auxotrophic mutants of *Rhizobium meliloti* induced by transposon Tn*5* mutagenesis. *J Bacteriol* 149, 114-122.

Meier, V. M. & Scharf, B. E. (2009). Cellular localization of predicted transmembrane and soluble chemoreceptors in *Sinorhizobium meliloti*. *J Bacteriol* 191, 5724.

Meier, V. M., Muschler, P. & Scharf, B. E. (2007). Functional analysis of nine putative chemoreceptor proteins in *Sinorhizobium meliloti*. *J Bacteriol* 189, 1816.

Mergaert, P., D'Haeze, W., Fernández-López, M., Geelen, D., Goethals, K., Promé, J.-C., Van Montagu, M. & Holsters, M. (1996). Fucosylation and arabinosylation of Nod factors in

Azorhizobium caulinodans: involvement of *nolK*, *nodZ* as well as *noeC* and/or downstream genes. *Mol Microbiol* 21, 409-419.

Mergaert, P., Uchiumi, T., Alunni, B., Evanno, G., Cheron, A., Catrice, O., Mausset, A.-E., Barloy-Hubler, F., Galibert, F. & other authors (2006). Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*–legume symbiosis. *Proc Natl Acad Sci USA* 103, 5230.

Middleton, P. H., Jakab, J., Penmetsa, R. V., Starker, C. G., Doll, J., Kaló, P., Prabhu, R., Marsh, J. F., Mitra, R. M. & other authors (2007). An ERF transcription factor in *Medicago truncatula* that is essential for Nod Factor signal transduction. *The Plant Cell* 19, 1221.

Miller, L. D., Yost, C. K., Hynes, M. F. & Alexandre, G. (2007). The major chemotaxis gene cluster of *Rhizobium leguminosarum* bv. *viciae* is essential for competitive nodulation. *Mol Microbiol* 63, 348-362.

Miller, M. A., Pfeiffer, W. & Schwartz, T. (2010). Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. In *SC10 Workshop on Gateway Computing Environments (GCE10)*.

Miller, R. W., McRae, D. G., Al-Jobore, A. & Berndt, W. B. (1988). Respiration supported nitrogenase activity of isolated *Rhizobium meliloti* bacteroids. *Journal of Cellular Biochemistry* 38, 35-49.

Miousse, I. R., Chalbot, M.-C. G., Lumen, A., Ferguson, A., Kavouras, I. G. & Koturbash, I. (2015). Response of transposable elements to environmental stressors. *Mutation Research/Reviews in Mutation Research* 765, 19-39.

Mitra, R. M., Gleason, C. A., Edwards, A., Hadfield, J., Downie, J. A., Oldroyd, G. E. D. & Long, S. R. (2004). A Ca2+/calmodulin-dependent protein kinase required for symbiotic nodule development: Gene identification by transcript-based cloning. *Proc Natl Acad Sci USA* 101, 4701.

Mitsch, M. J., Cowie, A. & Finan, T. M. (2007). Malic enzyme cofactor and domain requirements for symbiotic N2 fixation by *Sinorhizobium meliloti*. *J Bacteriol* 189, 160-168.

Mitsch, M. J., diCenzo, G. C., Cowie, A. & Finan, T. M. (2018). Succinate transport is not essential for symbiotic nitrogen fixation by *Sinorhizobium meliloti* or *Rhizobium leguminosarum*. *Appl Environ Microb* 84, e01561-01517.

Mitsui, H., Sato, T., Sato, Y., Ito, N. & Minamisawa, K. (2004). *Sinorhizobium meliloti* RpoH1 is required for effective nitrogen-fixing symbiosis with alfalfa. *Mol Genet Genomics* 271, 416-425.

Moënne-Loccoz, Y. & Weaver, R. W. (1995a). Plasmids and saprophytic growth of *Rhizobium leguminosarum* bv. *trifolii* W14-2 in soil. *FEMS Microbiol Ecol* 18, 139-144.

Moënne-Loccoz, Y. & Weaver, R. W. (1995b). Plasmids influence growth of rhizobia in the rhizosphere of clover. *Soil Biology and Biochemistry* 27, 1001-1004.

Moënne-Loccoz, Y., Sen, D., Krause, E. S. & Weaver, R. W. (1994). Plasmid profiles of rhizobia used in inoculants and isolated from clover fields. *Agron* 86, 117-121.

Moling, S., Pietraszewska-Bogiel, A., Postma, M., Fedorova, E., Hink, M. A., Limpens, E., Gadella, T. W. J. & Bisseling, T. (2014). Nod Factor Receptors form heteromeric complexes and are essential for intracellular infection in *Medicago* nodules. *The Plant Cell* 26, 4188.

Mondy, S., Lenglet, A., Beury-Cirou, A., Libanga, C., Ratet, P., Faure, D. & Dessaux, Y. (2014). An increasing opine carbon bias in artificial exudation systems and genetically modified plant rhizospheres leads to an increasing reshaping of bacterial populations. *Molecular Ecology* 23, 4846-4861.

Mortier, V., Den Herder, G., Whitford, R., Van de Velde, W., Rombauts, S., D'Haeseleer, K., Holsters, M. & Goormachtig, S. (2010). CLE peptides control *Medicago truncatula* nodulation locally and systemically. *Plant Physiol* 153, 222-237.

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

Mortlock, R. P., (editor) (1984). *Microorganisms as Model Systems for Studying Evolution*. New York: Plenum Press.

Mortlock, R. P. & Wood, W. A. (1964). Metabolism of pentoses and pentitols by *Aerobacter aerogenes*. I. Demonstration of pentose isomerase, pentulokinase, and pentitol dehydrogenase enzyme families. *J Bacteriol* 88, 838-844.

Mulligan, J. T. & Long, S. R. (1985). Induction of *Rhizobium meliloti nodC* expression by plant exudate requires *nodD*. *Proc Natl Acad Sci USA* 82, 6609-6613.

Murphy, P. J., Wexler, W., Grzemski, W., Rao, J. P. & Gordon, D. (1995). Rhizopines—Their role in symbiosis and competition. *Soil Biology and Biochemistry* 27, 525-529.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53, 240-255.

Mus, F., Crook, M. B., Garcia, K., Garcia Costas, A., Geddes, B. A., Kouri, E. D., Paramasivan, P., Ryu, M.-H., Oldroyd, G. E. D. & other authors (2016). Symbiotic nitrogen fixation and the challenges to its extension to nonlegumes. *Appl Environ Microb* 82, 3698-3710.

Nishimura, R., Hayashi, M., Wu, G.-J., Kouchi, H., Imaizumi-Anraku, H., Murakami, Y., Kawasaki, S., Akao, S., Ohmori, M. & other authors (2002). HAR1 mediates systemic regulation of symbiotic organ development. *Nature* 420, 426-429.

Nobelmann, B. & Lengeler, J. W. (1995). Sequence of the *gat* operon for galactitol utilization from a wild-type strain EC3132 of *Escherichia coli*. *BBA-Gene Struct Expr* 1262, 69-72.

Nobelmann, B. & Lengeler, J. W. (1996). Molecular analysis of the *gat* genes from *Escherichia coli* and of their roles in galactitol transport and metabolism. *J Bacteriol* 178, 6790-6795.

Nolle, N., Felsl, A., Heermann, R. & Fuchs, T. M. (2017). Genetic characterization of the galactitol utilization pathway of *Salmonella enterica* serovar typhimurium. *J Bacteriol* 199, e00595-00516.

Nunn, C. E. M., Johnsen, U., Schönheit, P., Fuhrer, T., Sauer, U., Hough, D. W. & Danson, M. J. (2010). Metabolism of pentose sugars in the hyperthermophilic archaea *Sulfolobus solfataricus* and *Sulfolobus acidocaldarius*. *J Biol Chem* 285, 33701-33709.

Oberhardt, M. A., Palsson, B. Ø. & Papin, J. A. (2009). Applications of genome-scale metabolic reconstructions. *Molecular systems biology* 5, 320-320.

Oger, P., Petit, A. & Dessaux, Y. (1997). Genetically engineered plants producing opines alter their biological environment. *Nature Biotechnology* 15, 369-372.

Okazaki, S., Tittabutr, P., Teulet, A., Thouin, J., Fardoux, J., Chaintreuil, C., Gully, D., Arrighi, J.-F., Furuta, N. & other authors (2016). Rhizobium-legume symbiosis in the absence of Nod factors: two possible scenarios with or without the T3SS. *The ISME journal* 10, 64-74.

Oldroyd, G. E. D. & Long, S. R. (2003). Identification and characterization of *nodulation-signaling pathway 2*, a gene of *Medicago truncatula*; involved in Nod Factor signaling. *Plant Physiol* 131, 1027.

Oldroyd, G. E. D. & Downie, J. A. (2004). Calcium, kinases and nodulation signalling in legumes. *Nat Rev Mol Cell Biol* 5, 566-576.

Oldroyd, G. E. D. & Downie, J. A. (2008). Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu Rev Plant Biol* 59, 519-546.

Oldroyd, G. E. D. & Dixon, R. (2014). Biotechnological solutions to the nitrogen problem. *Curr Opin Biotech* 26, 19-24.

Omar Draghi, W., Florencia Del Papa, M., Barsch, A., Albicoro, F. J., Lozano, M. J., Pühler, A., Niehaus, K. & Lagares, A. (2017). A metabolomic approach to characterize the acid-tolerance response in *Sinorhizobium meliloti*. *Metabolomics* 13.

Oono, R. & Denison, R. F. (2010). Comparing symbiotic efficiency between swollen versus nonswollen rhizobial bacteroids. *Plant Physiol* 154, 1541-1548.

Oresnik, I. J., Charles, T. C. & Finan, T. M. (1994). Second site mutations specifically suppress the Fix- phenotype of *Rhizobium meliloti ndvF* mutations on alfalfa: identification of a conditional *ndvF*-dependent mucoid colony phenotype. *Genetics* 136, 1233-1243.

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

Oresnik, I. J., Pacarynuk, L. A., O'Brien, S. A. P., Yost, C. K. & Hynes, M. F. (1998). Plasmid-encoded catabolic genes in *Rhizobium leguminosarum* bv. *trifolii*: Evidence for a plant-inducible rhamnose locus involved in competition for nodulation. *Mol Plant Microbe In* 11, 1175-1185.

Pacheco, A. R., Curtis, M. M., Ritchie, J. M., Munera, D., Waldor, M. K., Moreira, C. G. & Sperandio, V. (2012). Fucose sensing regulates bacterial intestinal colonization. *Nature* 492, 113-117.

Palleroni, N. J. & Doudoroff, M. (1957). Metabolism of carbohydrates by *Pseudomonas saccharophila*: III. Oxidation of D-arabinose. *J Bacteriol* 74, 180-185.

Pawlowski, K. & Demchenko, K. N. (2012). The diversity of actinorhizal symbiosis. *Protoplasma* 249, 967-979.

Peix, A., Ramírez-Bahena, M. H., Velázquez, E. & Bedmar, E. J. (2015). Bacterial associations with legumes. *Critical Reviews in Plant Sciences* 34, 17-42.

Pellock, B. J., Cheng, H.-P. & Walker, G. C. (2000). Alfalfa root nodule invasion efficiency Is dependent on *Sinorhizobium meliloti* polysaccharides. *J Bacteriol* 182, 4310-4318.

Pellock, B. J., Teplitski, M., Boinay, R. P., Bauer, W. D. & Walker, G. C. (2002). A LuxR Homolog Controls Production of Symbiotically Active Extracellular Polysaccharide II by Sinorhizobium meliloti. *J Bacteriol* 184, 5067-5076.

Persson, B. & Kallberg, Y. (2013). Classification and nomenclature of the superfamily of short-chain dehydrogenases/reductases (SDRs). *Chem-Biol Interact* 202, 111-115.

Persson, B., Kallberg, Y., Bray, J. E., Bruford, E., Dellaporta, S. L., Favia, A. D., Duarte, R. G., Jörnvall, H., Kavanagh, K. L. & other authors (2009). The SDR (short-chain dehydrogenase/reductase and related enzymes) nomenclature initiative. *Chem-Biol Interact* 178, 94-98.

Peters, N. K., Frost, J. W. & Long, S. R. (1986). A plant flavone, luteolin, induces expression of *Rhizobium meliloti* nodulation genes. *Science* 233, 977.

Philippsen, A., Schirmer, T., Stein, M. A., Giffhorn, F. & Stetefeld, J. (2005). Structure of zinc-independent sorbitol dehydrogenase from *Rhodobacter sphaeroides* at 2.4 A resolution. *Acta Crystallogr D* 61, 374-379.

Pickering, B. S. & Oresnik, I. J. (2008). Formate-dependent autotrophic growth in *Sinorhizobium meliloti*. *J Bacteriol* 190, 6409-6418.

Pietraszewska-Bogiel, A., Lefebvre, B., Koini, M. A., Klaus-Heisen, D., Takken, F. L. W., Geurts, R., Cullimore, J. V. & Gadella, T. W. J. (2013). Interaction of *Medicago*

truncatula Lysin Motif Receptor-Like Kinases, NFP and LYK3, produced in *Nicotiana benthamiana* induces defence-like responses. *PloS One* 8, e65055.

Pitcher, R. S. & Watmough, N. J. (2004). The bacterial cytochrome cbb3 oxidases. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1655, 388-399.

Platt, R., Drescher, C., Park, S. K. & Phillips, G. J. (2000). Genetic system for reversible integration of DNA constructs and *lacZ* gene fusions into the *Escherichia coli* chromosome. *Plasmid* 43, 12-23.

Plet, J., Wasson, A., Ariel, F., Signor, C. L., Baker, D., Mathesius, U., Crespi, M. & Frugier, F. (2011). MtCRE1-dependent cytokinin signaling integrates bacterial and plant cues to coordinate symbiotic nodule organogenesis in Medicago truncatula. *The Plant Journal* 65, 622-633.

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

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

Preisig, O., Zufferey, R., Thöny-Meyer, L., Appleby, C. A. & Hennecke, H. (1996). A high-affinity cbb3-type cytochrome oxidase terminates the symbiosis-specific respiratory chain of *Bradyrhizobium japonicum*. *J Bacteriol* 178, 1532.

Prell, J., Mulley, G., Haufe, F., White, J. P., Williams, A., Karunakaran, R., Downie, J. A. & Poole, P. S. (2012). The PTSNtr system globally regulates ATP-dependent transporters in *Rhizobium leguminosarum*. *Mol Microbiol* 84, 117-129.

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

Prlić, A., Bliven, S., Rose, P. W., Bluhm, W. F., Bizon, C., Godzik, A. & Bourne, P. E. (2010). Pre-calculated protein structure alignments at the RCSB PDB website. *Bioinformatics* 26, 2983-2985.

Pundir, S., Martin, M. J. & O'Donovan, C. (2017). UniProt Protein Knowledgebase. In *Protein Bioinformatics: From Protein Modifications and Networks to Proteomics*, pp. 41-55. Edited by C. H. Wu, C. N. Arighi & K. E. Ross. New York, NY: Springer New York.

Putnoky, P., Kereszt, A., Nakamura, T., Endre, G., Grosskopf, E., Kiss, P. & Kondorosi, Á. (1998). The *pha* gene cluster of *Rhizobium meliloti* involved in pH adaptation and symbiosis encodes a novel type of K+ efflux system. *Mol Microbiol* 28, 1091-1101.

Quandt, J. & Hynes, M. F. (1993). Versatile suicide vectors which allow direct selection for gene replacement in Gram-negative bacteria. *Gene* 127, 15-21.

Quesada-Vincens, D., Fellay, R., Nasim, T., Viprey, V., Burger, U., Prome, J. C., Broughton, W. J. & Jabbouri, S. (1997). *Rhizobium* sp. strain NGR234 NodZ protein is a fucosyltransferase. *J Bacteriol* 179, 5087-5093.

Ramachandran, V. K., East, A. K., Karunakaran, R., Downie, J. A. & Poole, P. S. (2011). Adaptation of *Rhizobium leguminosarum* to pea, alfalfa and sugar beet rhizospheres investigated by comparative transcriptomics. *Genome Biol* 12, R106.

Randhawa, G. S. & Hassani, R. (2002). Role of rhizobial biosynthetic pathways of amino acids, nucleotide bases and vitamins in symbiosis. *Indian journal of experimental biology* 40, 755-764.

Rapaille, A., Goosens, J. & Heume, M. (2003). SUGAR ALCOHOLS. In *Encyclopedia of Food Sciences and Nutrition (Second Edition)*, pp. 5665-5671. Edited by B. Caballero. Oxford: Academic Press.

Raveendran, S., Parameswaran, B., Ummalyma, S. B., Abraham, A., Mathew, A. K., Madhavan, A., Rebello, S. & Pandey, A. (2018). Applications of microbial enzymes in food industry. *Food Technol Biotechnol* 56, 16-30.

Razmilic, V., Castro, J. F., Andrews, B. & Asenjo, J. A. (2018). Analysis of metabolic networks of *Streptomyces leeuwenhoekii* C34 by means of a genome scale model: Prediction of modifications that enhance the production of specialized metabolites. *Biotechnology and bioengineering* 115, 1815-1828.

Reinhardt, A., Johnsen, U. & Schonheit, P. (2019). L-Rhamnose catabolism in archaea. *Mol Microbiol*.

Reizer, J., Ramseier, T. M., Reizer, A., Charbit, A. & Saier, M. H. (1996). Novel phosphotransferase genes revealed by bacterial genome sequencing: a gene cluster encoding a putative N-acetylgalactosamine metabolic pathway in *Escherichia coli*. *Microbiology* 142, 231-250.

Relic, B., Talmont, F., Kopcinska, J., Golinowski, W., Prome, J. C. & Broughton, W. J. (1993). Biological activity of *Rhizobium* sp. NGR234 Nod-factors on *Macroptilium atropurpureum*. *Molecular plant-microbe interactions : MPMI* 6, 764-774.

Remigi, P., Zhu, J., Young, J. P. W. & Masson-Boivin, C. (2016). Symbiosis within Symbiosis: Evolving Nitrogen-Fixing Legume Symbionts. *Trends Microbiol* 24, 63-75.

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

Richardson, J. S., Carpena, X., Switala, J., Perez-Luque, R., Donald, L. J., Loewen, P. C. & Oresnik, I. J. (2008). RhaU of *Rhizobium leguminosarum* is a rhamnose mutarotase. *J Bacteriol* 190, 2903-2910.

Riely, B. K., Lougnon, G., Ané, J.-M. & Cook, D. R. (2007). The symbiotic ion channel homolog DMI1 is localized in the nuclear membrane of *Medicago truncatula* roots. *The Plant Journal* 49, 208-216.

Riley, M., Abe, T., Arnaud, M. B., Berlyn, M. K. B., Blattner, F. R., Chaudhuri, R. R., Glasner, J. D., Horiuchi, T., Keseler, I. M. & other authors (2006). *Escherichia coli* K-12: a cooperatively developed annotation snapshot—2005. *Nucleic Acids Res* 34, 1-9.

Ritter, A. V., Bader, J. D., Leo, M. C., Preisser, J. S., Shugars, D. A., Vollmer, W. M., Amaechi, B. T. & Holland, J. C. (2013). Tooth-surface-specific effects of xylitol:Randomized trial results. *Journal of Dental Research* 92, 512-517.

Rivers, D. & Oresnik, I. J. (2013). Carbohydrate kinase (RhaK)-dependent ABC transport of rhamnose in *Rhizobium leguminosarum* demonstrates genetic separation of kinase and transport activities. *J Bacteriol* 195, 3424-3432.

Rivers, D. M. R. & Oresnik, I. J. (2015). The sugar kinase that is necessary for the catabolism of rhamnose in *Rhizobium leguminosarum* directly interacts with the ABC transporter necessary for rhamnose transport. *J Bacteriol* 197, 3812.

Roche, P., Debellé, F., Maillet, F., Lerouge, P., Faucher, C., Truchet, G., Dénarié, J. & Promé, J.-C. (1991). Molecular basis of symbiotic host specificity in *Rhizobium meliloti*: *nodH* and *nodPQ* genes encode the sulfation of lipo-oligosaccharide signals. *Cell* 67, 1131-1143.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin III, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C. & other authors (2009). A safe operating space for humanity. *Nature* 461, 472-475.

Rostas, K., Kondorosi, E., Horvath, B., Simoncsits, A. & Kondorosi, A. (1986). Conservation of extended promoter regions of nodulation genes in *Rhizobium*. *Proc Natl Acad Sci* 83, 1757.

Rubio, L. M. & Ludden, P. W. (2008). Biosynthesis of the iron-molybdenum cofactor of nitrogenase. *Annual Review of Microbiology* 62, 93-111.

Sachs, J. L., Quides, K. W. & Wendlandt, C. E. (2018). Legumes versus rhizobia: a model for ongoing conflict in symbiosis. *New Phytologist* 219, 1199-1206.

Sambrook, J. & Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual*, Third edn. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Sánchez, R., Serra, F., Tárraga, J., Medina, I., Carbonell, J., Pulido, L., de María, A., Capella-Gutíerrez, S., Huerta-Cepas, J. & other authors (2011). Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. *Nucleic Acids Res* 39, W470-W474.

Savka, M. A. & Farrand, S. K. (1997). Modification of rhizobacterial populations by engineering bacterium utilization of a novel plant-produced resource. *Nature Biotechnology* 15, 363-368.

Saxild, H. H., Andersen, L. N. & Hammer, K. (1996). Dra-nupC-pdp operon of *Bacillus subtilis*: nucleotide sequence, induction by deoxyribonucleosides, and transcriptional regulation by the deoR-encoded DeoR repressor protein. *J Bacteriol* 178, 424-434.

Schauder, S., Schneider, K.-H. & Giffhorn, F. (1995). Polyol metabolism of *Rhodobacter sphaeroides*: biochemical characterization of a short-chain sorbitol dehydrogenase. *Microbiology* 141, 1857-1863.

Schnabel, E., Journet, E. P., de Carvalho-Niebel, F., Duc, G. & Frugoli, J. (2005). The *Medicago truncatula SUNN* gene encodes a *CLV1*-like leucine-rich repeat receptor kinase that regulates nodule number and root length. *Plant molecular biology* 58, 809-822.

Schnoes, A. M., Brown, S. D., Dodevski, I. & Babbitt, P. C. (2009). Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies. *PLoS Comput Biol* 5, e1000605.

Schoenbeck, M. A., Temple, S. J., Trepp, G. B., Blumenthal, J. M., Samac, D. A., Gantt, J. S., Hernandez, G. & Vance, C. P. (2000). Decreased NADH glutamate synthase activity in nodules and flowers of alfalfa (*Medicago sativa* L.) transformed with an antisense glutamate synthase transgene. *Journal of Experimental Botany* 51, 29-39.

Schrodinger, L. (2015). The PyMOL Molecular Graphics System, Version 1.8.

Schroeder, B. K., House, B. L., Mortimer, M. W., Yurgel, S. N., Maloney, S. C., Ward, K. L. & Kahn, M. L. (2005). Development of a functional genomics platform for *Sinorhizobium meliloti*: construction of an ORFeome. *Appl Environ Microb* 71, 5858-5864.

Seabra, A. R., Pereira, P. A., Becker, J. D. & Carvalho, H. G. (2012). Inhibition of glutamine synthetase by phosphinothricin leads to transcriptome reprograming in root nodules of *Medicago truncatula*. *Mol Plant Microbe In* 25, 976-992.

Shah, V. K. & Brill, W. J. (1977). Isolation of an iron-molybdenum cofactor from nitrogenase. *Proc Natl Acad Sci USA* 74, 3249-3253.

Shapiro, J. A. (1969). Mutations caused by the insertion of genetic material into the galactose operon of Escherichia coli. *J Mol Biol* 40, 93-105.

Shepon, A., Gildor, H., Labrador, L. J., Butler, T., Ganzeveld, L. N. & Lawrence, M. G. (2007). Global reactive nitrogen deposition from lightning NOx. *J Geophys Res-Atmos* 112.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M. & other authors (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7.

Sigaud, S., Becquet, V., Frendo, P., Puppo, A. & Hérouart, D. (1999). Differential regulation of two divergent *Sinorhizobium meliloti* genes for HPII-like catalases during free-living growth and protective role of both catalases during symbiosis. *J Bacteriol* 181, 2634.

Siguier, P., Gourbeyre, E. & Chandler, M. (2014). Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiology Reviews* 38, 865-891.

Siguier, P., Perochon, J., Lestrade, L., Mahillon, J. & Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* 34, D32-D36.

Silveira, M. & Jonas, R. (2002). The biotechnological production of sorbitol. *Appl Microbiol Biot* 59, 400-408.

Simpson, F. B. & Burris, R. H. (1984). A nitrogen pressure of 50 atmospheres does not prevent evolution of hydrogen by nitrogenase. *Science* 224, 1095.

Smit, P., Raedts, J., Portyanko, V., Debellé, F., Gough, C., Bisseling, T. & Geurts, R. (2005). NSP1 of the GRAS protein family is essential for rhizobial Nod Factor-induced transcription. *Science* 308, 1789.

Smit, P., Limpens, E., Geurts, R., Fedorova, E., Dolgikh, E., Gough, C. & Bisseling, T. (2007). Medicago LYK3, an entry receptor in rhizobial nodulation factor signaling. *Plant Physiol* 145, 183.

Sola-Carvajal, A., García-García, M. I., García-Carmona, F. & Sánchez-Ferrer, Á. (2012). Insights into the evolution of sorbitol metabolism: phylogenetic analysis of SDR196C family. *BMC Evol Biol* 12, 147.

Sorai, M., Yoshida, N. & Ishikawa, M. (2007). Biogeochemical simulation of nitrous oxide cycle based on the major nitrogen processes. *J Geophys Res-Biogeo* 112.

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

Souleimanov, A., Prithiviraj, B. & Smith, D. L. (2002). The major Nod factor of *Bradyrhizobium japonicum* promotes early growth of soybean and corn. *Journal of Experimental Botany* 53, 1929-1934.

Sourjik, V., Sterr, W., Platzer, J., Bos, I., Haslbeck, M. & Schmitt, R. (1998). Mapping of 41 chemotaxis, flagellar and motility genes to a single region of the *Sinorhizobium meliloti* chromosome. *Gene* 223, 283-290.

Spatzal, T., Aksoyoglu, M., Zhang, L., Andrade, S. L. A., Schleicher, E., Weber, S., Rees, D. C. & Einsle, O. (2011). Evidence for interstitial carbon in nitrogenase FeMo cofactor. *Science* 334, 940.

Stacey, G. (1995). *Bradyrhizobium japonicum* nodulation genetics. *FEMS Microbiol Lett* 127, 1-9.

Stahl, M., Friis, L. M., Nothaft, H., Liu, X., Li, J., Szymanski, C. M. & Stintzi, A. (2011). L-fucose utilization provides *Campylobacter jejuni* with a competitive advantage. *Proc Natl Acad Sci* 108, 7194-7199.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312-1313.

Steele, K. P., Ickert-Bond, S. M., Zarre, S. & Wojciechowski, M. F. (2010). Phylogeny and character evolution in *Medicago* (Leguminosae): Evidence from analyses of plastid *trnK/matK* and nuclear *GA3ox1* sequences. *American Journal of Botany* 97, 1142-1155.

Steffen, W., Richardson, K., Rockström, J., Cornell, S. E., Fetzer, I., Bennett, E. M., Biggs, R., Carpenter, S. R., de Vries, W. & other authors (2015). Planetary boundaries: Guiding human development on a changing planet. *Science* 347, 1259855.

Stein, M. A., Schäfer, A. & Giffhorn, F. (1997). Cloning, nucleotide sequence, and overexpression of *smoS*, a component of a novel operon encoding an ABC transporter and polyol dehydrogenases of *Rhodobacter sphaeroides* Si4. *J Bacteriol* 179, 6335-6340.

Stewart, W. M., Dibb, D. W., Johnston, A. E. & Smyth, T. J. (2005). The contribution of commercial fertilizer nutrients to food production. *Agron* 97, 1-6.

Stoscheck, C. M. (1990). Quantitation of protein. In *Methods in Enzymology*, vol. 182, pp. 50-68. Edited by M. P. Deutscher: Academic Press.

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

Sullivan, J. T., Trzebiatowski, J. R., Cruickshank, R. W., Gouzy, J., Brown, S. D., Elliot, R. M., Fleetwood, D. J., McCallum, N. G., Rossbach, U. & other authors (2002). Comparative sequence analysis of the symbiosis island of *Mesorhizobium loti* strain R7A. *J Bacteriol* 184, 3086.

Sutton, M. A., Bleeker, A., Howard, C. M., Erisman, J. W., Abrol, Y. P., Bekunda, M., Datta, A., Davidson, E., Vries, W. d. & other authors (2013). Our nutrient world. The challenge to produce more food & energy with less pollution. Edinburgh: Centre for Ecology & Hydrology.

Suzuki, Y. (1989). A General Principle of Increasing Protein Thermostability. *P Jpn Acad B-Phys* 65, 146-148.

Suzuki, Y., Oishi, K., Nakano, H. & Nagayama, T. (1987). A strong correlation between the increase in number of proline residues and the rise in thermostability of five *Bacillus* oligo-1,6-glucosidases. *Appl Microbiol Biot* 26, 546-551.

Tajima, S., Nomura, M. & Kouchi, H. (2004). Ureide biosynthesis in legume nodules. *Frontiers in bioscience* 9, 1374-1381.

Terpolilli, J., Rui, T., Yates, R., Howieson, J., Poole, P., Munk, C., Tapia, R., Han, C., Markowitz, V. & other authors (2014). Genome sequence of *Rhizobium leguminosarum* bv *trifolii* strain WSM1689, the microsymbiont of the one flowered clover *Trifolium uniflorum*. *Standards in Genomic Sciences* 9, 527-539.

Tesfaye, M., Samac, D. A. & Vance, C. P. (2006). Insights into Symbiotic Nitrogen Fixation in *Medicago truncatula*. *Mol Plant Microbe In* 19, 330-341.

Thomas, G. H., Southworth, T., León-Kempis, M. R., Leech, A. & Kelly, D. J. (2006). Novel ligands for the extracellular solute receptors of two bacterial TRAP transporters. *Microbiology* 152, 187-198.

Tian, C. F., Zhou, Y. J., Zhang, Y. M., Li, Q. Q., Zhang, Y. Z., Li, D. F., Wang, S., Wang, J., Gilbert, L. B. & other authors (2012). Comparative genomics of rhizobia nodulating soybean suggests extensive recruitment of lineage-specific genes in adaptations. *Proc Natl Acad Sci* 109, 8629.

Triplett, E. W. & Sadowsky, M. J. (1992). Genetics of competition for nodulation of legumes. *Annu Rev Microbiol* 46, 399-422.

Udvardi, M. & Poole, P. S. (2013). Transport and Metabolism in Legume-Rhizobia Symbioses. *Annu Rev Plant Biol* 64, 781-805.

Udvardi, M. K., Price, G. D., Gresshoff, P. M. & Day, D. A. (1988). A dicarboxylate transporter on the peribacteroid membrane of soybean nodules. *FEBS Letters* 231, 36-40.

Untiet, V., Karunakaran, R., Krämer, M., Poole, P., Priefer, U. & Prell, J. (2013). ABC transport is inactivated by the PTSNtr under potassium limitation in *Rhizobium leguminosarum* 3841. *PloS One* 8, e64682.

Van de Velde, W., Zehirov, G., Szatmari, A., Debreczeny, M., Ishihara, H., Kevei, Z., Farkas, A., Mikulass, K., Nagy, A. & other authors (2010). Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science* 327, 1122.

Van Deynze, A., Zamora, P., Delaux, P.-M., Heitmann, C., Jayaraman, D., Rajasekar, S., Graham, D., Maeda, J., Gibson, D. & other authors (2018). Nitrogen fixation in a landrace of maize is supported by a mucilage-associated diazotrophic microbiota. *PLOS Biology* 16, e2006352.

van Dillewijn, P., Soto, M. a. J., Villadas, P. J. & Toro, N. (2001). Construction and environmental release of a *Sinorhizobium meliloti* strain genetically modified to be more competitive for alfalfa nodulation. *Appl Environ Microb* 67, 3860-3865.

van Elsas, J. D., Turner, S. & Bailey, M. J. (2003). Horizontal gene transfer in the phytosphere. *New Phytologist* 157, 525-537.

Vanderlinde, E. M., Hynes, M. F. & Yost, C. K. (2014). Homoserine catabolism by *Rhizobium leguminosarum* bv. *viciae* 3841 requires a plasmid-borne gene cluster that also affects competitiveness for nodulation. *Environmental Microbiology* 16, 205-217.

Veereshlingam, H., Haynes, J. G., Penmetsa, R. V., Cook, D. R., Sherrier, D. J. & Dickstein, R. (2004). *nip*, a symbiotic *Medicago truncatula* mutant that forms root nodules with aberrant infection threads and plant defense-like response. *Plant Physiol* 136, 3692.

Vincent, J. M. (1970). *A Manual for the Practical Study of Root-Nodule Bacteria*. United Kingdom: Blackwell Scientific Publications.

Vineetha, K. E., Vij, N., Prasad, C. K., Hassani, R. & Randhawa, G. S. (2001). Ultrastructural studies on nodules induced by pyrimidine auxotrophs of *Sinorhizobium meliloti*. *Indian journal of experimental biology* 39, 371-377.

Vitousek, P. M., Hättenschwiler, S., Olander, L. & Allison, S. (2002). Nitrogen and Nature. *AMBIO* 31, 97-101.

Vitousek, P. M., Menge, D. N. L., Reed, S. C. & Cleveland, C. C. (2013). Biological nitrogen fixation: rates, patterns and ecological controls in terrestrial ecosystems. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 368, 20130119-20130119.

Voss, M., Bange, H. W., Dippner, J. W., Middelburg, J. J., Montoya, J. P. & Ward, B. (2013). The marine nitrogen cycle: recent discoveries, uncertainties and the potential relevance of climate change. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 368, 20130121-20130121.

Wais, R. J., Galera, C., Oldroyd, G., Catoira, R., Penmetsa, R. V., Cook, D., Gough, C., Dénarié, J. & Long, S. R. (2000). Genetic analysis of calcium spiking responses in nodulation mutants of *Medicago truncatula*. *Proc Natl Acad Sci* 97, 13407.

Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein engineering* 8, 127-134.

Walshaw, D. L., Reid, C. J. & Poole, P. S. (1997). The general amino acid permease of *Rhizobium leguminosarum* strain 3841 is negatively regulated by the Ntr system. *FEMS Microbiol Lett* 152, 57-64.

Wang, C., Meek, D. J., Panchal, P., Boruvka, N., Archibald, F. S., Driscoll, B. T. & Charles, T. C. (2006). Isolation of poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants. *Appl Environ Microb* 72, 384-391.

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

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

Webb, B. A., Hildreth, S., Helm, R. F. & Scharf, B. E. (2014). *Sinorhizobium meliloti* chemoreceptor McpU mediates chemotaxis toward host plant exudates through direct proline sensing. *Appl Environ Microb* 80, 3404.

Webb, B. A., Karl Compton, K., Castañeda Saldaña, R., Arapov, T. D., Keith Ray, W., Helm, R. F. & Scharf, B. E. (2017). *Sinorhizobium meliloti* chemotaxis to quaternary ammonium compounds is mediated by the chemoreceptor McpX. *Mol Microbiol* 103, 333-346.

Wells, D. H. & Long, S. R. (2002). The *Sinorhizobium meliloti* stringent response affects multiple aspects of symbiosis. *Mol Microbiol* 43, 1115-1127.

Wichelecki, D. J., Vetting, M. W., Chou, L., Al-Obaidi, N., Bouvier, J. T., Almo, S. C. & Gerlt, J. A. (2015). ATP-Binding Cassette (ABC) transport system solute-binding protein-guided identification of novel D-altritol and galactitol catabolic pathways in *Agrobacterium tumefaciens* C58. *J Biol Chem* 290, 28963-28976.

Williamson, J. D., Jennings, D. B., Guo, W.-W., Pharr, D. M. & Ehrenshaft, M. (2002). Sugar Alcohols, Salt Stress, and Fungal Resistance: Polyols—Multifunctional Plant Protection? *J Am Soc Hortic Sci* 127, 467-473.

Wojtkiewicz, B., Szmidzinski, R., Jezierska, A. & Cocito, C. (1988). Identification of a salvage pathway for D-arabinose in *Mycobacterium smegmatis*. *Eur J Biochem* 172, 197-203.

Wolf, J., Stark, H., Fafenrot, K., Albersmeier, A., Pham, T. K., Müller, K. B., Meyer, B. H., Hoffmann, L., Shen, L. & other authors (2016). A systems biology approach reveals major metabolic changes in the thermoacidophilic archaeon *Sulfolobus solfataricus* in response to the carbon source L-fucose versus D-glucose. *Mol Microbiol* 102, 882-908.

Wolucka, B. A. (2008). Biosynthesis of D-arabinose in mycobacteria – a novel bacterial pathway with implications for antimycobacterial therapy. *FEBS Journal* 275, 2691-2711.

Yamaki, S. (1980). A sorbitol oxidase that converts sorbitol to glucose in apple leaf. *Plant Cell Physiol* 21, 591-599.

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

Yoshida, K.-i., Yamaguchi, M., Morinaga, T., Kinehara, M., Ikeuchi, M., Ashida, H. & Fujita, Y. (2008). myo-Inositol catabolism in *Bacillus subtilis*. *J Biol Chem* 283, 10415-10424.

Yost, C. K., Rochepeau, P. & Hynes, M. F. (1998). *Rhizobium leguminosarum* contains a group of genes that appear to code for methyl-accepting chemotaxis proteins. *Microbiology* 144, 1945-1956.

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

Young, N. D., Debelle, F., Oldroyd, G. E. D., Geurts, R., Cannon, S. B., Udvardi, M. K., Benedito, V. A., Mayer, K. F. X., Gouzy, J. & other authors (2011). The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480, 520-524.

Yuan, Z.-C., Zaheer, R. & Finan, T. M. (2006a). Regulation and properties of PstSCAB, a high-affinity, high-velocity phosphate transport system of *Sinorhizobium meliloti*. *J Bacteriol* 188, 1089-1102.

Yuan, Z.-C., Zaheer, R., Morton, R. & Finan, T. M. (2006b). Genome prediction of PhoB regulated promoters in *Sinorhizobium meliloti* and twelve proteobacteria. *Nucleic Acids Res* 34, 2686-2697.

Zhao, H., Li, M., Fang, K., Chen, W. & Wang, J. (2012). *In silico* insights into the symbiotic nitrogen fixation in *Sinorhizobium meliloti* via metabolic reconstruction. *PloS One* 7, e31287.

Zhao, R., Liu, L. X., Zhang, Y. Z., Jiao, J., Cui, W. J., Zhang, B., Wang, X. L., Li, M. L., Chen, Y. & other authors (2017). Adaptive evolution of rhizobial symbiotic compatibility mediated by co-evolved insertion sequences. *The Isme Journal* 12, 101.

Zhu, L., Wang, S., Tian, W., Zhang, Y., Song, Y., Zhang, J., Mu, B., Peng, C., Deng, Z. & other authors (2017). Stabilization of Multimeric Proteins via Intersubunit Cyclization. *Appl Environ Microb* 83, e01239-01217.

Zipfel, C. (2014). Plant pattern-recognition receptors. *Trends in immunology* 35, 345-351.

Zufferey, R., Preisig, O., Hennecke, H. & Thöny-Meyer, L. (1996). Assembly and function of the cytochrome cbb oxidase subunits in *Bradyrhizobium japonicum*. *J Biol Chem* 271, 9114-9119.