Gibberellin Biosynthesis in *Bradyrhizobium japonicum* USDA110

by

Ariana Gail Marcassa

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Biology

Waterloo, Ontario, Canada. 2014

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

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

Abstract

Biosynthesis of gibberellin (GA) by bacteria has been studied for nearly 50 years. Early on, bacterial culture extracts were analyzed for the presence of bioactive compounds, thought to be gibberellins. In the 1980s, advances in chromatography allowed specific detection of gibberellins by high performance liquid chromatography combined with gas chromatography and mass spectrophotometry. In the 1990s, the genes involved in bacterial gibberellin biosynthesis were identified for the first time in *Bradyrhizobium japonicum* USDA110. Until 2009, biochemical characterization of the associated gene products had not been completed; at that time, the substrate specificity of two terpene cyclases was worked out. Beginning in 2011, we started to work on determining the role of the remaining enzymes with unknown functions, specifically three cytochrome P450s and one short-chain dehydrogenase/reductase. In 2013 we began a multinational collaboration that resulted in determining the function of two of the cytochrome P450 proteins. More recently, we have probed the origin of the operon in bacteria and using comparative genomics gained insight for a possible biological role of bacterial gibberellin. At this time our hypotheses are untested, but we hope that continued collaboration with researchers around the world will enable us to collectively discover the role of bacterial gibberellin. This thesis is a collection of the methods, materials, and hypotheses crafted and tested over the past three years.

I began my graduate studies by performing a literature review looking at early nodulation transcription control of gibberellin genes in rhizobia which can be found in chapter one. We found that expression of the operon was consistently upregulated during early-stage nodulation across different rhizobia. The second chapter details the methods I used to recombinantly express the *Bradyrhizobium japonicum* USDA110 gibberellin cytochrome P450 genes in *Escherichia coli* and the techniques I used to detect their biochemical activity while visiting a research lab at Iowa State University. The experiments to

create deletion mutants of the cytochrome P450s in the gibberellin operon in three soil dwelling microbes: *Bradyrhizobium japonicum* USDA110, *Sinorhizobium fredii* NGR234, and *Mesorhizobium loti* MAFF303099 are also presented in chapter two. Chapter three summarizes our closer look at the sequences of gibberellin operons and details our hypothesis regarding the role of the genes in bacteria.

The significance of this work stems from the creation of the *B. japonicum* USDA110 deletion mutant strains. Through a multinational collaboration, the strains were used to determine the role of their corresponding gene products in the synthesis of GA in this organism. Additionally, our survey of bacterial gibberellin biosynthesis operons in whole genome sequence data and primary literature has uncovered interesting clues to the biological role of gibberellin. We have developed hypotheses to explain why the gibberellin operon is found in certain bacterial microsymbionts of plants and what the biological role of bacterial gibberellin might be. Although this document is a good basis for research in bacterial gibberellin research, there is still much to learn and many questions to be answered.

Acknowledgments

I would like to thank Dr. Trevor Charles for supervising my research as well as Dr. Bernie Glick and Dr. Frédérique Guinel for sitting on my thesis committee. FG helped me immensely by carefully and meticulously editing my work; she provided me with a huge amount of critical comments and many great insights. Her knowledge on nodule organogenesis and nodule physiology was crucial in the improvement of this document. Trevor trained me become a scientific thinker and allowed me the freedom to pursue my research at my own pace. I deeply appreciate his mentor ship. I must also thank Dr. Hauke Henneke for introducing himself to me at the 22nd North American Symbiotic Nitrogen Fixation Conference at the University of Minnesota in July 2013. His interest in my project and kind words encouraged me to pursue the collaboration that resulted in my major scientific contributions outlined in chapter 2 of the thesis.

I owe a great deal to my past and present lab mates, especially Kathy Lam, Andre Masella, and John Heil who provided endless technical troubleshooting tips along with great friendship and camaraderie during my time in the Charles Lab. The advice of Dr. Juijun Cheng and Katja Engel was also invaluable. Thanks to Tam Tran, Ricardo Nordeste, and Maya D'Alessio for friendship, encouragement and inspiration to achieve my research goals. I would like to acknowledge my funding sources: the province of Ontario through two Ontario Graduate Scholarships as well as the University of Waterloo for additional funding from the President's Scholarship and Graduate Student Entrance Scholarship. I am grateful for my experiences as a teaching assistant for the Department of Biology and I value the memories I made in this role, especially as a TA with WL. I wish him all the best. To my other friends in other labs and those I made as an undergraduate student: thank you! You are all great people and I know you will go on to do wonderful things.

Special thank yous go out to Dr. Reuben Peters and Ryan Nett at Iowa State University for allowing and organizing my two-month visit to their laboratory in Ames, Iowa. It was an invaluable and life-changing experience and resulted in many interesting observations and an ongoing international collaboration. I owe so much to my partner, JK, for supporting me for many months before and after my visit to Ames. I have to thank AS who took care of my cat while I was away.

I extend a sincere thank you to everyone else who helped me along the way, especially Linda Zepf and Dr. Barbara Butler who encouraged my first scholarship applications as an undergraduate student. My achievements at this school are a reflection of my environment and peers. I'm looking forward to starting the next stage of my life, but I will always look back on my time at the University of Waterloo with fondness.

Claims of Contributions to Knowledge

1. We have reviewed the literature published on detection and measurement of bacterial gibberellin and regulation of the putative GA biosynthesis genes during nodule organogenesis and initiation of nitrogen fixation.

2. This is the first study to report creation of cytochrome P450 deletion mutants in the gibberellin biosynthesis operons of *B. japonicum* USDA110, *Sinorhizobium fredii* NGR234, and *Mesorhizobium loti* MAFF303099.

3. Through our collaboration with Dr. Reuben Peters and other researchers in Chile and the United Kingdom, the function of two of the cytochrome P450 enzymes (encoded by *blr2144* and *blr2145*) has been tentatively determined by analyzing bacteroid exudates of our mutant *Bradyrhizobium* strains. Follow up experiments to finalize the preliminary results are ongoing.

4. We propose that there is evidence for lateral transfer of the gibberellin operon between various soil-dwelling bacteria based on our analysis of the GC content and flanking transposon-related ORFs near the gibberellin biosynthesis genes in several bacterial species.

5. The putative gibberellin biosynthesis operon appears to be more widespread in bacteria than was previously thought and may serve different functions in different genetic contexts. We found examples of plant pathogens, symbionts, and endophytic bacteria that harbour homologous putative GA biosynthesis genes.

6. To our knowledge, this is the first report of putative GA biosynthesis operons in *S. meliloti* forming an association with *M. truncatula*. This is significant because *M. truncatula* forms indeterminate nodules harbouring swollen bacteroids and there are very few examples of bacteria with GA operons that form this specific type of plant symbiosis.

Table of Contents

Chapter 3: *In Silico* Analysis of the Bacterial GA Biosynthesis Operon and Surrounding DNA

List of Figures

List of Tables

List of Abbreviations

AG: arabinose-gluconate broth
ALA: amino-levulinic acid
APS: ammonium persulfate
BjFdx: *Bradyrhizobium japonicum* ferredoxin from putative GA operon
BLAST: Basic Local Alignment Sequence Tool
BNF: biological nitrogen fixation
C: Celsius
CFE: cell free extract
CO: carbon monoxide
CPP: copalyl pyrophosphate
CPS: copalyl pyrophosphate synthase
CSPD: Chloro-5-substituted adamantyl-1,2-dioxetane phosphate
CYP: cytochrome P450
DMAPP: dimethylallyl pyrophosphate
DNA: deoxyribonucleic acid
DTT: dithiothreitol
EDTA: ethylenediaminetetraacetic acid
***ent*-CPP**: *ent*-copalyl diphosphate
***ent*-KA**: *ent*-kaurene
***ent*-KAO**: *ent*-kaurenoic acid
EtOH: ethanol
Fdx: ferredoxin
FdxR: ferredoxin reductase
FPP: farnesyl pyrophosphate
FPPS: farnesyl pyrophosphate synhtase
GA: gibberellin
GC/MS: combined gas chromatography/mass spectrophotometry
GGPP: geranylgeranyl pyrophosphate
GGPPS: geranylgeranyl pyrophosphate synthetase
HEPES: 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HPLC: high performance liquid chromatography
IPP: isopentenyl pyrophosphate
IPTG: Isopropyl β-D-1-thiogalactopyranoside
ISU: Iowa State University
KS: kaurene synthases
LB: lysogeny broth
m/z: mass-to-charge ratio
MES: 2-(N-morpholino)ethanesulfonic acid
MSA: multiple sequence alignment
MUSCLE: multiple sequence comparison by log-expectation
NADPH: reduced nicotinamide adenine dinucleotide phosphate
ND: not done
NCBI: National Center for Biotechnology Innovation
OD: optical density

ORF: open reading frame
PCR: polymeras chain reaction
PGPR: plant growth promoting rhizobia
pSym: symbiotic plasmid
SDR: short-chain dehydrogenase
SDS: sodium dodecyl sulphate
SDS-PAGE: sodium dodecyl sulphate polyacrylamide gel electrophoresis
SSC: standard sodium citrate
TAE: Tris-acetate-EDTA buffer
TEMED: tetramethylethylenediamine
TB: Terrific broth
TC: terpene cyclase
TLC: thin layer chromatography
TY: Tryptone yeast extract
UV: ultraviolet
UW: University of Waterloo

Chapter 1

Overview of Gibberellin Biosynthesis in Rhizobia

Comparison of analogous gibberellin biosynthesis genes and how their regulation by an anaerobic promoter. The biochemistry of the putative bacterial biosynthesis is examined. All literature reporting on gibberellin production by bacteria is recapitulated.

This chapter reviews the production of gibberellin by soil-dwelling microbes that are part of the order Rhizobiales, specifically *Bradyrhizobium japonicum* USDA110, *Sinorhizobium fredii* NGR234 and *Mesorhizobium loti* MAFF303099. The Rhizobiales form symbiotic relationships with legumes, living within specialized root organs known as nodules. Within nodules, the bacteria convert atmospheric nitrogen gas (N_2) to ammonium (NH_3) and provide it to the plant in exchange for C4 compounds. This chapter looks at previously published transcriptomic and proteomic data sets to examine links between nodulation, nitrogen fixation and the biosynthesis gibberellin by bacteria. Rationale for study of the gibberellin biosynthesis pathway in microbes is also presented. The current knowledge of gibberellin biosynthesis in bacterial is summarized.

Section 1.1: Introduction and Rationale

This chapter covers several topics related to gibberellin (GA) biosynthesis but focuses specifically on bacterial GA biosynthesis. The genetic regulation of GA biosynthesis in *Bradyrhizobium japonicum* USDA 110, *Sinorhizobium fredii* NGR234 and *Mesorhizobium loti* MAFF303099 were studied with particular attention paid to the transcriptomic and proteomic changes observable between free-living cultures of all three strains and their counterparts living in as nodule microsymbionts. In short, in symbiosis, bacterial GA biosynthesis gene transcription is upregulated, and the corresponding proteins have been found, indicating that the mRNA is translated (Please see review section on this topic.). So far, a reason for increased GA biosynthesis during nodule organogenesis has not been found. Initially, it was thought that the product of the biosynthetic operon would affect legume symbiosis since its gene products were expressed by *B. japonicum* USDA110 bacteroids isolated from *Glycine max* nodules (Tully *et al*., 1993). However, the authors determined a bacterial microsymbiont's ability to synthesize GA is not crucial to the symbiotic relationship when they discovered that a *B. japonicum* USDA110 mutant unable to make GA had successfully colonized functional soybean nodules (Tully *et al*., 1993). Later, analogous biosynthesis operons were found in other Rhizobia (Keister *et al*., 1999) and when the B. japonicum USDA110 genome was sequenced, the GA biosynthesis genes were found on the symbiotic island (Göttfert, 2001). Analogous GA biosynthesis operons were found in the symbiotic island of *M. loti* MAFF303099 and on the symbiotic plasmid of *S. fredii* NGR234 (Kaneko et al., 2000; Pueppke and Broughton, 1999).

Gibberellin biosynthesis genes are also found in the Plantae and Fungi kingdoms. This chapter as a section that briefly compares what is known about GA biosynthesis in *Arabidopsis thaliana* and *Fusarium fujikuroi* to what is known in Bacteria. This is done in order to highlight the differences and similarities between the GA biosynthesis pathways in all three pathways in terms of the genes, gene

products, intermediate metabolites and final GAs made. Although the pathways in *A. thaliana*, *F. fujikuroi*, and bacteria all appear to result in the formation of a GA, the enzymes, intermediate products, and enzymatic mechanisms employed by each organism are different (Bömke and Tudzynski, 2009). In *A. thaliana*, all of the gene products and their functions are known (Bömke and Tudzynski, 2009). In *F. fujikuroi*, the majority of the gene products have functions assigned—only the role of one enzyme remains to be characterized (Bömke and Tudzynski, 2009). Although the pathway has not yet been elucidated at the same level of detail in microorganisms, what is known so far indicates that the biosynthesis pathway in Rhizobia presents a third novel route to the formation of GA (Bömke and Tudzynski, 2009). Lately, research of GA biosynthesis in bacteria has focused on characterizing the enzymes involved in the earlier steps of the pathway and the formation of the intermediate *ent*-kaurene (Morrone *et al*., 2009; Hershey *et al*., 2014). The next challenge to overcome in elucidation of bacterial GA biosynthesis is characterization of the three cytochrome P450 (CYP) enzymes involved in the conversion of ent-kaurene to GA. CYP enzymes often catalyze oxidation reactions and are involved in the synthesis and decomposition of a huge variety of compounds including gibberellin, alcohol, many xenobiotic compounds and environmental contaminants. They are a major constituent of the enzymes found in human liver tissue (Spaggiari *et al*., 2014).

Section 1.2: Gibberellin is a Plant Hormone.

At last count, 136 structurally similar compounds, identified in a variety of plants, fungi, and bacteria, belong to the gibberellin class of plant hormones (Rodrigues *et al*., 2011).[1] Not all gibberellins are bioactive. The majority of GAs are produced as inactive compounds or are deactivated once they have fulfilled their biological function (for a review please see MacMillian, 2002). A number of gibberellins have been identified through analysis of seeds, where they control the expression of amylase in germinating plants (for a review please see Olszewski *et al*., 2002). Current understanding tells us that biosynthesis of gibberellins appears to be species specific—it appears that some organisms produce unique GAs and inactivation products (MacMillian *et al*., 2002; Bömke and Tudzynski, 2009). Ultimately though, this hypothesis may be rejected once techniques to detect GAs improve, and the understanding of plant hormone signaling deepens.

In the current classification system, each gibberellin is assigned a number (GA_1, GA_3, GA_{125}, etc.) based on its order of discovery rather than a trivial name, with only one exception. Only GA_3 is also known by its trivial name: gibberellic acid (MacMillian, 2002). The structure of GAs differs based on their number of carbon atoms and their oxidation and hydroxylation levels (Mander, 2003). One third of gibberellins are based on a 20 carbon *ent*-gibberellane skeleton but nearly all of remaining two thirds of GAs are based on a 19 carbon structure with the main feature being a lactone bridge (Mander, 2003). GA_3 is shown in **Figure 1.1**. GA_3 contains 19 carbon atoms. A 20 carbon GA, GA_{14}, is shown on the right side of the diagram.

[1] A database tracking peer-reviewed studies reporting new forms of GAs is maintained by MacMillian, at http://www.plant-hormones.info/gibberellins.htm.

Gibberellins are synthesized from the isoprenoid precursor compounds isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). Sometimes DMAPP is alternatively named dimethylallyl diphosphate in the literature, because use of the term pyrophosphate to indicate two phosphate groups. There difference in compound is in name only. All GAs belong to an even larger group of compounds known as diterpenoid dicarboxylic acids, so named for their isoprenoid precursors and carboxylic acid groups.

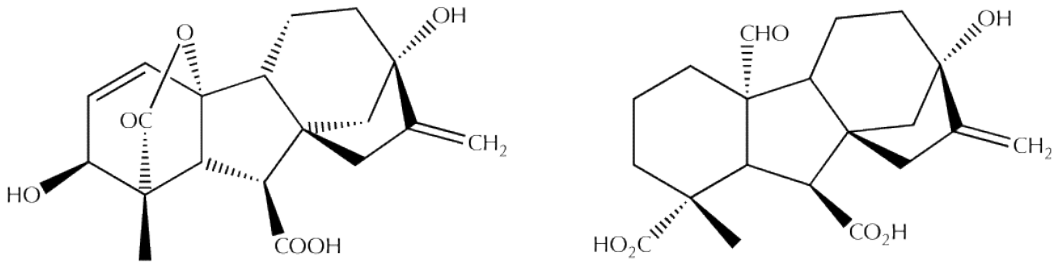


Figure 1: The structure of a C19 gibberellic acid, GA_3 and a C20 GA, GA_{14}.

This figure shows the structure of gibberellic acid (GA_3) on the right and GA14 on the left. GA_{14} is a minor GA biosynthesis pathway intermediate found in *F. fujikuroi* and is not bioactive (for review, please see Bömke and Tudzynski, 2009). So far, GA_{14} has not been found as an intermediate in either plant or bacterial GA biosynthesis (Méndez *et al*., 2014) GA_3 is a bioactive GA found in plants, fungi and bacteria (Bastián *et al*., 1998) and has valuable commercial applications in horticulture and agriculture, as well as brewing (Mander, 2003). Application of exogenous GA_3 is used to grow seedless table grapes, preserve and improve the appearance of citrus rinds, and to quicken malt production when brewing beer. GA_3 is produced for commercial use through industrial fermentation of the fungus *Fusarium fujikuroi* (Darken Jensen, and Shu, 1959; Tudzynski, 1999).

In plants, endogenous gibberellins are involved in regulation of growth and development. They are best known for their role in the induction of bolting, and as the causative agent of the associated disease known as *bakanae* (For a recent review, please see Hwang *et al*., 2013). *Bakanae*, or foolish seedling disease, is a disease of rice caused by exogenous GA_3 produced by the fungus F. *fujikuroi*. *Fusarium fujikuori* was formerly known as *Gibberella fujikuroi,* named as such because GA was initially discovered through research done to determine the cause of *bakanae* (Sawada, 1917; Yabuta 1935; Yabuta and Sumiki, 1938). The high level of GA produced by *F. fujikuroi*, causes excessive stem elongation of the rice seedlings, eventually leading to fatal collapse of the stem (For an overview, please see Ou, 1985). Industrial fermentation using *F. fujikuroi* provides most of the gibberellic acid for commercial and agricultural use (Lale, Jogdand, and Gadre 2006).

Besides bolting, GAs are also involved in the regulation of germination, flower and pollen formation, cell elongation and division, as well as plant senescence though the details of how GA regulates plant growth are beyond the scope of this work (MacMillan, 2002).

In this chapter I want to focus on what is known about bacterial GA so far. The work I did can be categorized in three sections. The *first part* of this chapter introduces the genes and gene products involved in GA biosynthesis in bacteria, comparing and contrasting them with those found in plants (*A. thaliana*) and fungi (*F. fujikuroi*). The second part of this chapter is a review of the literature published on bacterial GA biosynthesis. These efforts revealed that the majority of the work on bacterial GA consists of surveys of GAs isolated from various bacterial species. In many recent publications of this kind, the researchers identified and quantified biosynthesis of particular GAs from bacterial broth cultures. The bacterial species, growth medium and culture conditions, as well as the amount of GA measured and techniques used to quantify and identify the GAs are recapitulated in **Table 1.1** through **Table 1.2c**. My commentary and criticism of these articles follows the table. Details on the growth

conditions and medium recipes used in these studies can be found in the Appendix of this document. The third part of this chapter summarizes a number of large-scale transcriptomic and proteomic studies of *B. japonicum* USDA110, *S. fredii* NGR234 and *M. loti* MAFF303099. Though little work has been done specifically studying the GA genes and their products in bacteria, a good deal of information on the regulation of the operon has come from the supplementary data sections included in these publications. This is because these studies focused on comparing samples from free-living, aerobic cultures and symbiotic bacteroids.

Section 1.3: Bacterial GA biosynthesis

The GA biosynthesis pathway in bacteria has not yet been elucidated at the molecular level. So far, all of the involved genes have been identified in several species as a biosynthetic operon, but some of the corresponding reactions have not been linked back to the enzymes encoded by the operon. The most work to characterize the operon has been completed in the soybean symbiont *Bradyrhizobium japonicum* USDA 110. The relationship between *B. japonicum* and *Glycine max* is of great agricultural significance because soybean is an important crop for the world food supply, and because soybean is a legume that contributes to an overall positive nitrogen balance in agricultural fields because of the legume microsymbiont that performs biological nitrogen fixation (BNF) within the nodules. Interest in BNF is being renewed, and research money is now available from somewhat unexpected sources, including the Bill and Melinda Gates Foundation which has sponsored research grants on the topic since 2009. Efforts to better understand BNF and steps to confer the ability to non-legumes through genetic engineering are underway. Because bacterial GA is believed to be involved in symbiosis, understanding the details of GA biosynthesis in bacteria will contribute greater information on the legume-bacteria symbiosis.

Gibberellin biosynthesis by bacteria may be specifically linked to BNF in *B. japonicum*. There is a known link between the operon's promoter and the regulatory signal cascade in the microbe during

nodule organogenesis and BNF. A schematic diagram showing the operon as it is found in *Bradyrhizobium* is shown below in **Figure 1.2**. It is based off of data collected using NCBI's BLAST tool. The BLAST queries were based on two original research papers and one short communication discussing the operon in *B. japonicum* (Tully and Keister, 1993; Tully *et al*., 1998; Keister *et al*., 1999).

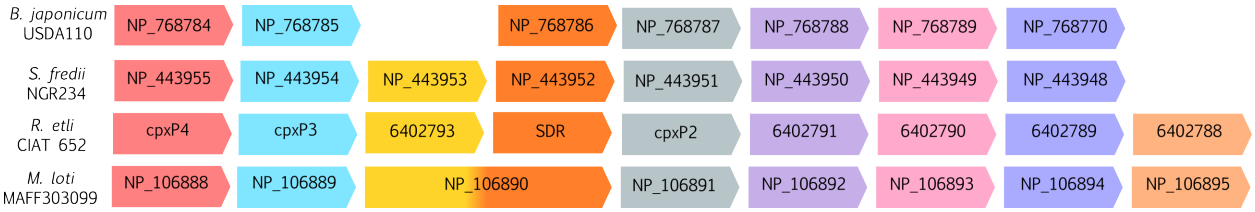


Figure 1.2: Gibberellin Biosynthesis Operon in several species of bacteria
The GA biosynthesis operons from four bacteria are compared here. Each ORF is a box of a different colour. The direction of transcription is indicated by the direction of the arrow. The length of each ORF is not to scale, but represents the order of the genes in each operon. This figure was created from the algorithmic annotations of the genes in each of the four organisms. From left to right, the first pink and blue genes are cytochrome P450s, followed by the golden yellow and orange boxes which represent the ferredoxin and short-chain reductase. The grey ORF is a third cytochrome P450. The purple, pink and periwinkle ORFs are a geranylgeranyl pyrophosphate synthase (GGPPS), class II and class I terpene cyclases, respectively. *Rhizobium etli* CIAT 552 and *M. loti* MAFF303099 appear to have an additional gene at the 5' end of the operon, indicated by the light orange ORF. The function of the corresponding gene product and its role in GA biosynthesis is not known at this time.The current version of the *B. japonicum* USDA110 genome does not have a ferredoxin open reading frame annotated, shown by the lack of a golden yellow ORF which is present in the other three organisms. In *M. loti* MAFF303099, it appears that the ferredoxin is fused with the short-chain dehydrogenase/reductase enzyme, demonstrated by the blending of both ORFs.

Homologous genes with similar organization are found in *M. loti* MAFF303099 and *S. fredii* NGR234 as well (see **Figure 1.2**). Similar operons are also known in other bacterial species, including and *Xanthomonas oryzae* pv. *oryzicola* BLS256 and *Burkholderia* sp. JPY251. These and other ORFs are explored in Chapter 3. As more genomes are sequenced, additional gibberellin operons will certainly be discovered. The GA biosynthesis operons in *S. fredii* NGR234 and *M. loti* MAFF303099 were selected for study early on in this project. *S. fredii* NGR234 was chosen because it is able to nodulate a variety of legumes and can inhabit both indeterminate and determinate nodules, depending on the plant host (Pueppke and Broughton, 1999). The operon in *M. loti* MAFF303099 is interesting because it has additional genes not found in the *B. japonicum* USDA110 and *S. fredii* NGR234 GA biosynthesis operons. *M. loti* MAFF303099 inhabits the determinate nodules formed by *Lotus japonicus* (Uchiumi *et*

al., 2004). Through our review of the literature, we found that the majority of nodulating bacterial microbsymbionts inhabit nodules with determinate meristems. At this time we cannot explain the apparent link between nodule phenotype and bacterial GA biosynthesis.

Section 1.4: Background on the three Rhizobiales species chosen for characterization

B. japonicum USDA110 is a member of the Bradyrhizobiaceae and is best known as the bacterial microsymbiont of *Glycine max* (see Kaneko *et al*., 2002). It has and has a single chromosome that is 9.1 mega basepairs in size and just over 8,300 annotated genes (Kaneko *et al*., 2002). *M. loti* MAFF303099 is the symbiont of *L. japonicus* and is a member of the Phyllobacteriaceae (Kaneko *et al*., 2000). It has three replicons: a chromosome of 7.03 Mbp and two symbiotic plasmids pMLa (351 kbp) and pMLb (200 kbp) (Kaneko *et al*., 2000; Uchiumi *et al*., 2004). The last microbe, *S. fredii* NGR234 is a member of the Rhizobiaceae and is able to form a symbiotic relationship with over one hundred legumes (Schmeisser *et al*., 2009). Like *M. loti* MAFF303099, the genome of *S. fredii* NGR234 has three replicons (Capela *et al*., 2001 Schmeisser *et al*., 2009). *S. fredii* NGR234 has a 3.9 Mbp chromosome, one megaplasmid pNGR234b (2.43 Mbp) and a smaller plasmid pNGR234a (536 Kbp) (Capela *et al*., 2001). In *M. loti* MAFF303099, the GA biosynthesis operon is located on the chromosome, while in *S. fredii* NGR234 it is found on pNGR234a. All three organisms have similar habitats and life styles but genetically they are quiet different. *R. etli* CIAT 522 was not initially selected for study of its GA operon but shown because I came across it while searching the NCBI sequence databases for homologs of translated *B. japonicum* USDA110 CYPs. **Table 1.1** provides more detail on each gene in the GA operons of *B. japonicum* USDA110, *S. fredii* NGR234, and *M. loti* MAFF303099. Most of the genes involved in bacterial GA biosynthesis have not been directly characterized, though they are known because the organism's genome is sequenced. If the genes or the corresponding gene products have been directly studied, that reference is provided instead.

Table 1.1: Gibberellin Biosynthesis Genes in Four Rhizobia

Organism	Locus Tag	Gene Product	Accession Number	Reference
B. japonicum USDA110	*blr2144*	BJ P450-1	NP_768784.1	Tully and Keister, 1993 Tully *et al*., 1998 Kaneko *et al*., 2002 Méndez *et al*., 2014
	blr2145	BJ P450-2	NP_768785.1	
	blr2145.1	ferredoxin	AAC2889.1	
	blr2146	SDR	NP_768786	
	blr2147	BJ P450-4	NP_768787.1	
	blr2148	GGPPS	NP_768788.1	
	blr2149	CPS	NP_768789.1	Morrone *et al*., 2009
	blr2150	KS	NP_768790.1	
S. fredii NGR234	NGR_a02700	P450-1	NP_443955.1	Schmeisser *et al*., 2009
	NGR_a02710	P450-2	NP_443954.1	
	NGR_a02720	ferredoxin	NP_443953.1	
	NGR_a02730	SDR	NP_443952.1	
	NGR_a02740	P450-4	NP_443951.1	
	NGR_a02750	GGPPS	NP_443950.1	Hershey *et al*., 2014
	NGR_a02760	CPS	NP_443949.1	
	NGR_a02770	KS	NP_443948.1	
M. loti MAFF303099	*mlr6364*	P450-1	NP_106888.1	Kaneko *et al*., 2000 Tatsukami *et al*., 2013
	mlr6365	P450-2	NP_106889.1	
	mlr6366	Ferredoxin/SDR fusion	NP_106890.1	
	mlr6367	P450-4	NP_106891.1	
	mlr6368	GGPPS	NP_106892.1	Hershey *et al*., 2014
	mlr6369	CPS	NP_106893.1	
	mlr6370	KS	NP_106894.1	

Organism	Locus Tag	Gene Product	Accession Number	Reference
	mlr6371	IPP isomerase	NP_106895.1	Kaneko *et al*., 2000
	mlr6372	transporter protein	BA000012.4	

So far, only three classes of enzymes involved in bacterial GA biosynthesis have been experimentally characterized (Morrone *et al*., 2009; Hershey *et al*., 2014). These are the GGPPS, class I, and class II terpene cyclases that are involved in the early steps of the pathway: the conversion of the isoprenoid precursors to *ent*-kaurene, considered the first committed step in GA biosynthesis (for a review, please see Bömke and Tudzynski, 2009). The class I and class II terpene cyclase (TC) enzymes have been studied in *B. japonicum* USDA110 (Morrone *et al*., 2009) and the GGPPS, and analogous TC enzymes from *M. loti* 303099 and *S. fredii* NGR234 have also been looked at(Hershey *et al*., 2014). The information for the rest of the genes shown in **Figure 1.2** was compiled from algorithmically annotated sequence data using genetic homology and predicted protein functions.

Although there is obvious similarity between the organization of the bacterial genes and those found in *A. thaliana* and *F. fujikuroi*, there is little amino acid homology between CYPs from all three kingdoms. GA biosynthesis in *A. thaliana* and *F. fujikuroi* utilizes different reactions and enzymes, which results in marked differences between the intermediate metabolites of both pathways (Bömke and Tudzynski, 2009). From preliminary analysis of the bacterial pathways known so far, it seems that the bacterial method differs from the plant and fungal biosynthesis pathways because of the enzymes that are apparently involved and the intermediate GAs formed (Hedden *et al*., 2001; Bömke and Tudzynski, 2009).

All biological GAs are formed from the primary metabolite precursors isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) (Mander, 2003). Though GA is a product of secondary metabolism in plants, fungi, and bacteria, IPP and DMAPP are primary metabolites found in all kingdoms of life and are formed by one of two analogous metabolic pathways (MacMillian, 2002; Bömke and Tudzynski, 2009). In bacteria, three molecules of IPP are joined successively with one DMAPP molecule in a head-to-tail condensation reaction by the GGPPS, forming GGPP. It is believed that the IPP isomerase encoded by the *mlr6371* locus in *M. loti* MAFF303099 converts DMAPP to IPP, which could promote synthesis of GGPP in this bacterium (Reuben Peters, personal communication). Both IPP and DMAPP are also used to form a variety of other secondary metabolites including sterols, triterpenes, and farensylated or geranylated proteins (For a recent review, see Zi, Mafu, and Peters 2014). Besides GA, GGPP is also the precursor of several other important natural products including carotenoids, rebaudiosides and quinones. A single GGPPS gene and enzyme is usually found to be dedicated solely for GA biosynthesis in organisms that make it (Bömke and Tudzynski, 2009). Similarly, carotenoids and rebaudiosides built on the GGPP compound backbone also have dedicated GGPPS genes and products in their biosynthetic pathways (Zi, Mafu, and Peters, 2014). Thereofre, it is believed that the GGPPS enzymes, encoded by *blr2148*, *mlr6368*, and *NGR_a02740* in *B. japonicum* USDA110, *M. loti* MAFF303099, and *S. fredii* NGR234, respectively, synthesize GGPP exclusively for GA biosynthesis (Hershey *et al*., 2014).

In bacteria, GGPP is converted to *ent*-kaurene in a two step reaction catalyzed by the TC enzymes (Morrone *et al*., 2009; Hershey *et al*., 2014). In *A. thaliana*, these conversions are also catalyzed by two enzymes, but in *F. fujikuori*, a single fusion enzyme catalyzes both reactions (Bömke and Tudzynski, 2009). This is the first major piece of evidence that bacterial GA biosynthesis is a third novel, convergent pathway compared to those elucidated in fungus and plants. The class I and II TC

enzymes involved in GA biosynthesis in bacteria are *ent*-copalyl diphosphate synthase (CPS) and *ent*-kaurene synthase (KS). CPS is the class II TC while KS is the class I enzyme. The CPS and KS enzymes of *B. japonicum* USDA110 were been experimentally characterized in 2009 by Morrone and colleagues and in 2014, the same lab published follow-up work in *M. loti* MAFF303099 and *S. fredii* NGR234 (Hershey *et al*., 2014). These genes and products that they characterized are *mlr6369* and *mlr6370* in *M. loti* MAFF303099 and *NGR_a02760* and *NGR_a02770* in *S. fredii* NGR234 which encode the class I and II terpene cyclases.

Ent-kaurene is converted on by a series of cytochrome P450s and one SDR into a GA. The genes corresponding to each enzyme are known, but the specific function of each protein is poorly understood. Unfortunately, amino acid sequence comparison between the bacterial, plant, and fungal CYPs has not yielded much insight into the substrate specificity of the CYPs. Reportedly, only three amino acids are absolutely conserved between different CYPs, making it difficult to infer analogy between CYPs from different organisms (Bömke and Tudzynski, 2009). There is minimal amino acid homology between the *F. fujikuroi* and *A. thaliana* CYPs, so even though the substrate specificity of these GA biosynthesis CYPs are known, that knowledge is not transferable to bacterial systems. Furthermore, *A. thaliana* and other plants use another set of enzymes to perform some of the reactions in the later stages of GA biosynthesis. This class of enzymes are not found in *F. fujikuroi* where all of the reactions are performed by CYPs. Finally, the major intermediate metabolites and final GA products in *F. fujikuroi* and *A. thaliana* are not the same (Bömke and Tudzynski, 2009). The major pathway in *F. fujikuroi* goes through GA_{14}, while in *A. thaliana* the first major GA formed is GA_{12} (Bömke and Tudzynski, 2009). From preliminary analysis of GA biosynthesis in *B. japonicum* USDA110, it appears that GA_{12} is the major GA formed and GA_{14} was not detected (Méndez *et al*., 2014). Interestingly, neither the *F. fujikuroi* nor *A. thaliana* GA biosynthesis gene clusters include a gene encoding an SDR—this appears

to be novel to the bacterial pathway. The two terpene cyclase proteins in bacterial make the first part of the pathway like that in *A. thaliana*. However, the later conversion of *ent*-kaurene to GA is done solely by CYPs, making this part of the pathway in bacteria more like *F. fujikuroi* than *A. thaliana*. It appears that the bacterial GA biosynthesis pathway shares common elements with the fungal and plant pathways, but that it is unique. It is likely that bacterial pathway is a third case of convergent evolution in the production of gibberellin.

Section 1.5: Reports of GA biosynthesis by bacteria.

Table 1.2 summarizes all of the journal articles published on bacterial gibberellin biosynthesis where a GA was found and positively identified. The first column lists the name of the bacterial species and strain studied. The second column briefly lists that growth conditions of the culture and the name of the growth medium used to grow the strain. All of the research studies we found characterized GA from laboratory grown cultures. The particular GA identified and the amount measured are shown in the third column; the method used to identify and quantify the GA is listed in column 4. Finally, the reference for the information is found in the fifth column of the table. The reports are organized in chronological order by year of publication. Beginning in 2004, Korean researchers led by IJ Lee has spearheaded the quantification of GAs from free-living, aerobic bacterial cultures, mostly focusing on characterization of GAs from bacterial endophytes (Joo *et al*., 2004; Joo *et al*., 2005; Joo *et al*., 2009; Kang *et al*., 2009a; Kang *et al*., 2009b; Lee, Ka, and Song, 2012; Kang *et al*., 2012; Kang *et al*., 2013; and Khan *et al*., 2014. Lee completed his PhD thesis in gibberellin metabolism of *Sorghum bicolor* at Texas A&M in 1996. Their results have been largely ignored by the bacterial gibberellin research community, because the articles are difficult to read and the results difficult to interpret because of a lack of transparency regarding the growth conditions and medium recipes. The studies were also mostly published in obscure journals that are not highly cited. Despite these shortcomings, the data are actually consistent with the most recent studies on GA biosynthesis in *B. japonicum* USDA110 produced by Méndez and colleagues

(2014).

A summary of the Korean studies are presented in **Table 1.2a-c** because their studies present a significant increase in the amount of data reported. Up until 2004, investigators had only reported finding bio-active GAs (GA_1, GA_3, or GA_4) in bacterial cultures; the reports issued under the supervision of Dr. In-Jung Lee expand the bacterial GA repertoire to include several intermediate GAs including GA_7, GA_9, GA_{12}, GA_{19}, GA_{20}, GA_{24}, GA_{34}, GA_{36}, GA_{44} and GA_{53}, albeit sometimes at concentrations as low as 10 pg per mL. The characterization of GAs from the bacterial endophytes selected by the Korean researchers provides a solid foundation for understanding the GA biosynthesis data that resulted from analysis of the *B. japonicum* USDA110 wild-type bacteroids (Méndez *et al*., 2014), as well as the Δ*blr2144* and Δ*blr2145* mutant strain bacteroids, as discussed in Chapter 2. For all entries in the upcoming tables, please see the original publications for additional details and medium recipes. The information contained in the tables is to demonstrate the variability of the data from different reports.

Table 1.2: Bacteria that have been reported to synthesize GA

Bacterial Species	Growth Conditions	Gibberellin Concentration	Detection Method	Reference
Azotobacter chorococcum Beji	10 day old, stationary phase culture; Burk's Medium	qualitative	TLC	Vančura, 1961
Arthrobacter globiforis	Sucrose Salts Medium shaken at 26 °C	qualitative	TLC *Avena* coleoptile assay	Katznelson, Sirois and Cole, 1962
Agrobacterium tumefaciens B_6	7 day old aerobic, 100 mL broth culture, standard *Rhizobium* spp. growth medium	GA_3 11 (µg L-1)	initial assessment by TLC and *Avena* coleoptile assay confirmation on: lettuce hypocotyl dwarf pea cucumber hypocotyl dwarf maize	Katznelson and Cole, 1965
A. tumefaciens BNV_6		GA_3 4		
A. radiobacter		GA_3 4		
Rhizobium meliloti		GA_3 4		
R. trifoli		GA_3 1		
R. leguminosarum		GA_3 1		
Ar. citrens		GA_3 2.5		
Ar. pascens		GA_3 7		
Ar. aurescens		GA_3 1		
Ar. tumescens		GA_3 2		
Ar. atrocyaneous		GA_3 2		
Pseudomonas aeruginosa		GA_3 1		
P. fluorescens isolate 1		GA_3 2		
P. fluorescens isolate 2		GA_3 1		
Bacillus cereus		GA_3 14		
Cytophaga sp.		GA_3 2		
Clostridium pasteurianum		GA_3 5		
Az. chroococcum A6	14 day old aerobic broth cultures in Sucrose Salt Medium 26 °C	qualitative	TLC	Brown and Burlingham, 1968
P. fluorescens	100 mL broth culture 95 rpm, 28 °C for 90 hr Medium B	GA_3 30 (ng mL-1)	TLC	Hussain and Vančura, 1970
P. fluorescens isolate 1		GA_3 3		
P. fluorescens isolate 2		GA_3 30		
Bacillus megaterium		GA_3 7		
B. brevis		GA_3 60		

Bacterial Species	Growth Conditions	Gibberellin Concentration	Detection Method	Reference
B. megaterium isolate 2		GA_3 1		
P. fluorescens		GA_3 40		
P. striata		GA_3 30		
P. fragi		GA_3 4		
P. perolens		qualitative	maize coleoptile test	
Pseudomonas. sp.		qualitative		
Chromobacterium violaceum		qualitative		
B. subtilis var. *aterrimus*		qualitative		
Az. brasilense Sp13t SR2	200 mL substrate buffered medium no further information	qualitative	TLC and bioassay	Tien, Gaskins and Hubbell, 1979
Arthrobacter. sp.	2 day old, 50 L fermenter culture YEG broth	qualitative	TLC and bioassay following (Franklin and Wareing 1960)	Grappelli and Rossi, 1981
R. phaseoli 8002 wild isolate	Aerobic culture at 25 °C 12 day old culture Mannitol and Mineral Salts	GA_1 412±100 (ng L^{-1}) GA_4 208±28	HPLC GC/MS and HPLC/EIA-RIA	Atzorn *et al*., 1988
R. phaseoli gRp1 Glasgow field isolate		GA_1 638±78 GA_4 355±110		
R. phaseoli 2671 high efficacy nodules 1		GA_1 142±26 GA_4 224±31		
R. phaseoli 2674 medium efficacy nodules 1		GA_1 341±70 GA_4 221±40		
R. phaseoli 2677 medium efficacy nodules 2		GA_1 202±26 GA_4 43±9		
R. phaseoli 2681 high efficacy nodules 2		GA_1 388±43 GA_4 78±13		
R. phaseoli 2673 ineffective nodules		GA_1 146±19 GA_4 33±4		
R. phaseoli 8401 (8002 derivative; cured pSym)		GA_1 339±38 GA_4 37±18		

Bacterial Species	Growth Conditions	Gibberellin Concentration	Detection Method	Reference
R. phaseoli JL15 (pSym::Tn5 Fix-)		GA_1 379±5 GA_4 228±16		
R. phaseoli JL19 (pSym::Tn5 Fix-)		GA_1 321±76 GA_4 124±21		
Az. lipoferum op 33 (auxin overproducer)	Malic Acid Broth 48 hours at 32 °C until a CFU of 10^9 cells mL^{-1}	qualitative	HPLC GC/MS HPLC GC/SIM	Bottini *et al*., 1989
A. brasilense	250 mL culture shaken at 75 rpm for 7 days at 22 °C in 1 L flasks	qualitative	HPLC GC/MS	Janzen *et al*., 1992
Acetobacter diazotropicus PAL 5	100 mL LGIP broth 32 °C 80 rpm for 5 days	GA_1 1.6 (ng mL^{-1}) GA_3 11.6	HPLC GC-SIM	Bastián *et al*., 1998
Herbospirillum seropedicae Z67	100 mL NFb broth 32 °C shaken at 80 rpm for 2 days	GA_3 12.5 ng mL^{-1}		
B. licheniformis CECT 5106	500 mL nutrient broth (Pronadisa) 24 hours at 28 °C CFU ml^{-1} of 10^8 was achieved	GA_1 129±5 (ng mL^{-1}) GA_3 57±7, GA_4 8±1	HPLC GC/MS	Gutiérrez-Mañero *et al.*, 2001
B. pumilus CECT 5105		GA_1 149±8 ng mL^{-1} GA_3 51±5 GA_4 11±2		
Bradyrhizobium japonicum E109	100 mL YMB broth cultures 250 mL flasks at 30 °C and 80 rpm until $OD_{600} \approx 1$	GA_3 870 ng mL^{-1}	HPLC GC/MS SIM	Boiero *et al*., 2007
B. japonicum SEMIA 5080		GA_3 590		
B. japonicum USDA 110		GA_3 790		
R. radiobacter 204	20 mL aerobic culture Yeast Mannitol Broth (Vincent, 1970)	1.3±0.31 µM	HPLC UV/Vis	Humphry *et al*., 2007
B. aquimaris SU8	Not reported.	109.94 ± 7.0 µg/mg^{-1} protein	spectrophotometry following Paleg (1965)	Upadhyay, Singh, and Saikia, 2009
B. pumilus SU10		54.73 ± 5.0		
Ar. sp. SU18		130.74 ± 9.5		
B. cereus SU24		73.32 ± 5.0		
B. subtilis SU47		7.67 ± 1.2		

Bacterial Species	Growth Conditions	Gibberellin Concentration	Detection Method	Reference
B. subtilis HC8	100 mL Difco Brain-Heart Infusion broth at 28 °C was shaken at 150 rpm for 60 hours	150 ng per 10^9 cells	spectrophotometry	Malfanova *et al*., 2011
B. japonicum USDA110				Méndez *et al*., 2014

Table 1.2a: Characterization of GAs from bacterial endophytes lead by Korean I-J Lee

Bacterial species, identity and Concentration of GA (pg mL^{-1})					
GA	***Bacillus cereus* MJ-1**	***B. pumilus* CJ-29**	***B. macroides* CJ-69**	***Burkholderia* sp. KCTC 11096BP**	***Bu. cepacia* SE4**
GA_1	22 ± 1	35 ± 1	32 ± 1	~ 2 ± 1	ND
GA_3	42 ± 2	27 ± 6	145 ±18	~ 50 ± 10	ND
GA_4	78 ± 16	42 ± 3	76 ± 15	~ 28 ± 2	25.6 ± 0.6
GA_5	ND	ND	61 ± 10	ND	ND
GA_7	56 ± 3	17 ± 1	87 ± 7	ND	2.1 ± 1.1
GA_8	ND	ND	33 ± 1	ND	ND
GA_9	5 ± 0	11 ± 1	60 ± 1	~ 2 ± 1	4.1 ± 1
GA_{12}	1 ± 0	2 ± 0	4 ± 1	~ 2 ± 0	ND
GA_{15}	ND	ND	ND	~ 8 ± 4	ND
GA_{19}	16 ± 1	5 ± 0	5 ± 1	ND	ND
GA_{20}	3 ± 1	2 ± 0	2 ± 0	~ 1 ± 0	ND
GA_{24}	13 ± 1	16 ± 1	16 ± 0	~ 8 ± 2	ND
GA_{34}	15 ± 1	11 ± 1	68 ± 3	ND	ND
GA_{36}	39 ± 1	46 ± 1	7 ± 0	ND	ND
GA_{44}	155 ± 9	3 ± 1	4 ± 0	ND	ND
GA_{53}	1 ± 0	2 ± 1	3 ± 0	ND	ND

The data in **Table 1.2a** were obtained from GC/MS analysis of exudates from *Bacillus* spp. grown in GY medium (Joo *et al*., 2004). The recipe for this medium is can be found in the Appendix. Although the text of the articles does not explicitly state that GY is the medium used, it mentions that culture supernatant was analyzed for GA content, and gives a recipe for solid GY medium. By inference, GY medium was probably used to grow the three *Bacillus* strains for analysis of their GAs biosynthesis capabilities. It is likely that the *Bacillus* spp. were grown as aerobic cultures with shaking at 37 ºC, though only the growth temperature was provided by the authors (Joo et *al*., 2004). Their follow-up paper in 2005 demonstrated that when the *Bacillus* endophytes were used to co-inoculate red pepper plants, they increased the level of endogenous GA produced by the plants (Joo *et al*., 2005). The data reported in both papers was collected by combined HPLC and GC/MS analysis; HPLC was used to concentrate and purify the organic solvent fractions used to extract the GAs from biological tissue and then GC/MS was used for actual detection of the compounds. The researchers compared the GAs in the culture extracts with deuterated authentic GA standards obtained from LN Mander at Australia National University (Joo *et al*., 2004; Joo *et al*., 2005). In 2009, the group followed up with a study of *Burkholderia* sp. KCTC 11096 BP (Joo *et al*., 2009). Although the genome of the *Burkholderia* strain presented above is not yet available, the sequence of *Burkholderia* sp. JPY 251 is available and we found a putative GA biosynthesis operon in that strain. Hopefully, once the genome of *Burkholderia* sp. KCTC 11096 BP becomes available soon, and we can check if a GA operon is present in this organism too.

Table 1.2b: Characterization of GA produced by endophytic *Bacillus aryabhattai* strains.

Bacterial species, identity and Concentration of GA (µg/mg protein)				
GA	***B. aryabhattai* LS9**	***B. aryabhattai* LS11**	***B. aryabhattai* LS12**	***B. aryabhattai* LS15**
GA_3	88.5	53.3	107.1	119.0

A follow up report published in 2012 by Lee, Ka and Song, reported on the plant-growth promotion abilities of the *Bacillus aryabhattai* strains. The researchers characterized GA biosynthesis in four additional *B. aryabhattai* strains. The data presented in Table 1.2b was obtained by growing the *B. aryaabhattai* strains in Difco Nutrient Broth and then separated by a UV detector attached to an HPLC machine (Lee, Ka, and Song, 2012). The study found that growth of *Xanthium italicum* was improved by application of the *B. aryabhattai* strains. There are other reports of GAs produced by *Bacillus* spp.

Table 1.2c: Characterization of GAs in additional endophytic bacteria

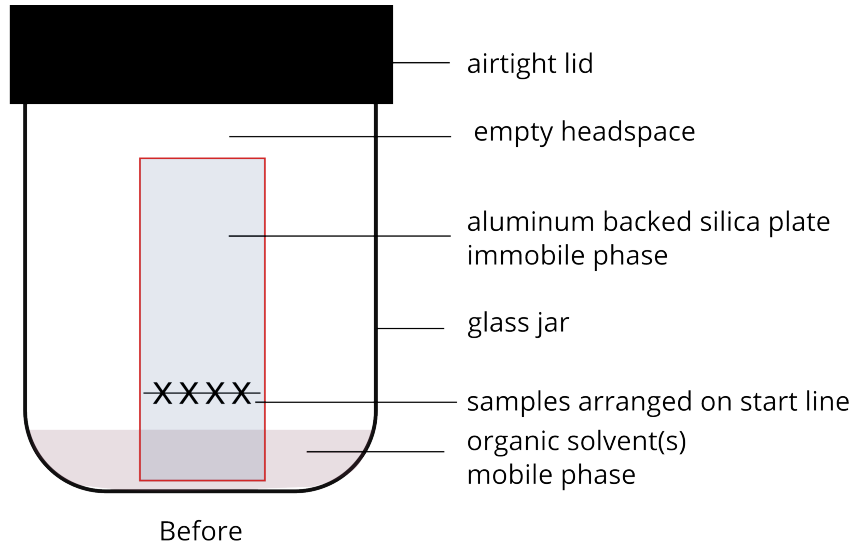
	Bacterial species and amount of GA quantified (pg mL^{-1})			
GA	***Acinetobacter calcoaeticus* SE370**	***Promicromonospora* sp. SE188**	***Leifsonia soli* sp. SE134**	***Sphingomonas* sp. LK11**
GA_1	5 ± 5	9.9 ± 0.3	~ 6 ± 1.5	ND
GA_3	60 ± 5	ND	ND	ND
GA_4	30 ± 4	15.8 ± 1.1	~ 16 ± 2	29.7 ± 1.1
GA_7	ND	ND	~ 5 ± 2	24.1 ± 2.3
GA_8	ND	ND	~ 10 ± 1	ND
GA_9	3 ± 1	ND	~ 4.5 ± 2	9.8 ± 1.5
GA_{12}	1 ± 0	ND	~ 6 ± 3	ND
GA_{15}	10 ± 1	ND	ND	ND
GA_{19}	8 ± 1	ND	~ 8 ± 1.5	ND
GA_{20}	1 ± 0	1.7 ± 0.8	~ 1.8 ± 1	ND
GA_{24}	7 ± 2	5.7 ± 0.7	~ 3.8 ± 1	ND
GA_{34}	ND	2.4 ± 0.3	~ 3.5 ± 1	ND
GA_{53}	2 ± 1	2.9 ± 0.2	~1.7 ± 0.5	ND

The data tabulated in Table 1.2c were collected from 4 separate reports on bacterial endophytes from 4 different bacterial genera. The GA profile of *A. calcoaeticus* SE370 was determined in 2009 (Kang and colleagues). Follow up studies published in 2012 and 2013 by Kang et al., established *Promicromonospora* sp. SE188 and *Leifsonia soli* sp. SE143 as additional plant endophytes that have positive effects on the growth of plants including cucumber, tomato and young radishes. Presumably, the positive effects of the endophytes were contributed to by their ability to produce GA. Most recently, the GAs synthesized by *Sphingomonas* sp. LK11 were profiled and found to have a role in promotion of tomato plat growth (Khan *et al*., 2014). All bacterial species were found to produce GA after growth in aerobic broth culture, and when used to inoculate plant hosts, found to increase endogenous GA production by the plant as well (Kang *et al*., 2009; Kang *et al*., 2012; Kang *et al*., 2013; Khan *et al*., 2014). Additionally, there may be other positive effects from the endophytic bacteria that have not yet been discovered; for example, the researchers also found that *Sphingomonas* sp. LK11 produced another plant hormone (auxin) which improved the drought tolerance of the tomato plants (Khan *et al*., 2014).

Overall, while the results of these studies are interesting, they are disputed by the transcriptomic studies that have shown that the GA biosynthesis genes (in Rhizobia) are regulated by an anaerobic promoter that is only transcriptionally active at low oxygen tension (such as within the nodule) or in anaerobic culture (Tully *et al*., 1993; Chang *et al*., 2007; and others). Additionally, most of the reports published so far challenge this data since they sampled aerobic bacterial culture and apparently detected gibberellins. The promoter of the GA operon in *B. japonicum* USDA110, *S. fredii* NGR234, and *M. loti* MAFF303099 is active at low-oxygen concentrations (see Section 1.6). So far, mRNA transcripts of the GA operon in *B. japonicum* USDA110 have not been found in aerobically grown culture. Only anaerobic culture and symbiotic samples have shown signs of GA operon transcription in *B. japonicum* USDA110 (Tully *et al*., 1993; Chang *et al*., 2007; and others). Similar results for *S. fredii* NGR234 and *M. loti* MAFF303099 have been reported. Therefore, while the results produced by the Korean researchers are interesting because of the GAs that they reported finding, they do not appear to be very credible since their data was obtained from study of aerobic broth culture samples. Additionally, the GA concentrations that they report are very low. In fact, they are so low that one must question whether or not they are above reasonable limits of detection for equipment used to obtain the data. Nevertheless, the reports make up the bulk of the recent experimental studies of of bacterial GA biosynthesis and must be accepted as part of the scientific literature on the topic. Finally, the extreme low concentrations reported in most cases are problematic and raise questions as to whether or not the data is reliable.

Section 1.5.1: Measurement of Gibberellin in Bacteria

Biosynthesis of gibberellins by bacteria found in the soil has been studied since the early 1960s with Vančura (1961) and Katznelson, Sirois & Cole (1962) initially reporting on gibberellin production in *Azotobacter chroococcum* Beji and *Arthrobacter globiforis*, respectively. We also found reports of GA-producing endophytic and plant-growth promoting fungi in the literature, but analysis of these articles was beyond the scope of our study. Early on, detection of GA synthesized by bacteria was completed using using thin-layer chromatography (TLC) and bioassays typically using lettuce, carrot or dwarf maize seedlings (Katznelson and Cole, 1965; Hussain and Vančura, 1970; and others). TLC separates compounds based on their polarity in specific solvent systems (for a recent review, please see Dolowy and Pyka, 2014). Bioassays can be used to determine if individual compounds or mixtures have biological activity, typically by increasing the growth rate of a plant, and its efficacy in studying GA has been discussed (Hill and Wimble, 1969). **Figure 1.3** is a simplified diagram that shows how TLC works.



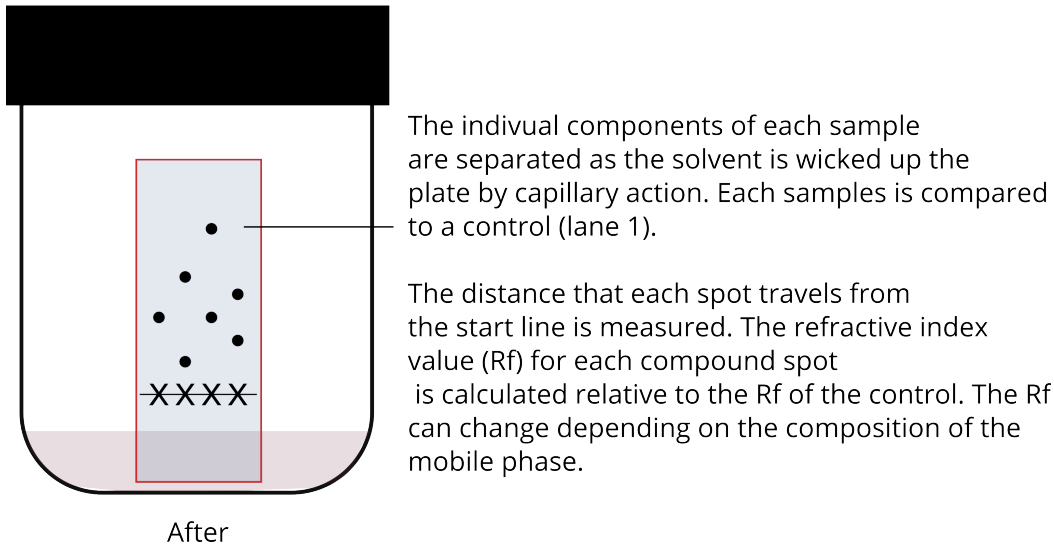


Figure 1.3: Thin layer chromatography is a fast method to analyze GAs

The samples are spotted in a horizontal line on the silica plate with a glass capillary. As the solvent moves up the silica plate, the dissolved samples are drawn up by capillary action. Depending on the properties of the sample compounds, they will move up the plate at different rates, which allows separation of individual compounds from a complex mixture.

TLC is simple, rapid, and reproducible. It is very inexpensive compared to high-performance liquid chromatography (HPLC) but because the column is much longer, it is more effective at separating the components of a mixture. Instead of the carrier solvent being forced through a high-pressure system like it is in HPLC, TLC relies on capillary action to wick the solvent and dissolved samples up an aluminum-backed silica plate that serves as the chromatography column. Individual components of a mixture are recovered from the plate by scraping off the silica at the spot corresponding to an isolated compound. A solution of the isolated compound cab be obtained by re-suspending the silica in solvent, like acetone. It is a similar result as is achieved when fractions are eluted from an HPLC machine. Until the 1980s, TLC was widely used in bacterial study of GA, but now the use of HPLC to purify and fractionate samples prior to analysis with a GC/MS machine has become standard. Though GC/MS is costlier compared to a bioassay, it is faster and has the added benefit of identifying a compound using a specific ion chromatogram. Specific identification of a compound is not possible with a bioassay.

Many of the microbes found to produce GA are plant-growth promoting bacteria. Some of the species listed in **Tables 1.1, 1.2a-c** are also nodule microsymbionts. Bacterial GA may have positive effects on plant growth, but at this time nothing has been experimentally confirmed. There is some indication that GA may be involved in the formation of the nodule primordium, and it may also be relevant once the nodule reaches maturity (Hayashi, Gresshoff, and Ferguson, 2014). As noted in the early surveys by Katznelson and others in 1962 and 1965, there are many factors that can affect the production and detection of GA and GA-like compounds by bacteria. For example, in 1965 Katznelson and Cole documented that GA production apparently changed depending on media type and oxygen concentration. The extraction method itself (solvents used, number of successive extractions, and the specific detection techniques used) may also influence detection of GAs, since some compounds may be lost or altered by the extraction process. It is also possible that contaminating compounds could be

introduced during sample work-up. Loss of sample and false positive identification of GA is an especial concern for the field because the concentrations of GAs reported are sometimes very low.

In light of these concerns, it is unfortunate that steps to standardize or even reproduce previous growth conditions, extraction techniques, and detection methods have not been made by the bacterial GA biosynthesis research community. Because of the variability in the methods to grow, extract and detect bacterial GA reported in the literature, it is difficult to compare and contrast all existing reports quantifying GA production in different bacterial species and prepare general comments on the research. While the differences in culture conditions are understandable because of the diverse requirements of some bacterial strains, an effort should be made to standardize extraction, detection, and reporting of bacterial GA in the future. These steps are necessary so that it is possible to reproduce previously published data. All of the issues that I found while reviewing the literature on bacterial GA are summarized below. Each point is expanded upon on subsequent pages.

1. The concentrations of GAs varies from species to species and report to report. In some cases, the reported values fall below 1 ng per milliliter.
2. Unfortunately, few papers independently reproduce any previously published work and nearly each research group has developed its own method to grow, extract, and measure bacterial GA.
3. Statistical analysis to accompany the reported GA concentrations was not consistent. In some reports it is not clear if the GA concentration is an average of several samples or if the reported number is from a single sample.
4. The units used to report concentration of GA have not been standardized.

Section 1.5.2: Variable, and extremely low concentrations of GA are reported.

The reported concentrations are low overall but vary by orders of magnitude among different species and publications. While this variability could be explained by the variety of organisms tested, culture conditions, measurement techniques, and more, it could also reflect problems with experimental design and measurement techniques. In general, the variability in the data makes it difficult to interpret their meaning. However, overall, the concentration of bacterial GA is consistently much lower than the GA concentrations reported in the fungus *F. fujikuroi* and *Arabidopsis thaliana* (Bömke and Tudzynski, 2009). There is no specific reason to doubt the reports of bacterial GA, but new work confirming previously reported numbers would go a long way to strengthen confidence in these reports, especially because genome sequence data for the majority of the bacterial species studied is not available.

Section 1.5.3: Culture conditions and media are not standardized.

Different culture conditions and media types have been used to study biosynthesis of GA in bacteria. On one hand, this is understandable since the bacteria that have been studied so far come from a variety of genera and may have specific requirements for growth in the lab. On the other hand, non-standardized techniques introduce challenges for data interpretation and comparison between different species, as noted by other researchers. For example, the use of defined or complex media may alter the results with respect to amount of GA detected. Comparison of GA production between different studies is also difficult because each new study tends to focus on finding *new* producers of gibberellin. Several species of *Bacillus, Rhizobium,* and *Pseudomonas* have been found to produce GA, but they are often reported in the same study and not corroborated by other reports. In the few cases where studies of the same species are reported, they are not the same strain and in other cases, the reported units differ or the media type, extraction procedure or detection method varies. For reference, here are some of the culture conditions used for detection of gibberellins in various organisms.

Section 1.5.4: Statistical analysis is not consistently done nor reported.

Failing to report standard deviation reduces the credibility of the data. In some studies it is unclear whether or not the measurements are based on averages of technical replicates, biological replicates or neither. It is important for numerical data to be reported along with the relevant standard deviation or standard error measurements, so that the quality of the measurements can be assessed. This is especially important because of the low concentration of GAs detected in bacterial culture. Studies that report GA production as ng mg^{-1} protein also used complex media including nutrient broth and brain heart infusion and failed to totally disclose their culture techniques. Between the reported GA concentration and lack of information, it is difficult to fit in the information contained these studies with the larger body of literature on GA biosynthesis.

Section 1.5.5: Extraction and detection methods vary.

There is great variability in the extraction and detection methods reported so far. This is with respect to the volumes of the cultures used, volume and type of extraction solvent (ethanol, acetone, etc.), number of washes/extractions, grade of solvent, evaporation, acidification, and more. Each choice adds a new variable that must be accounted for in the final quantification of GAs. In addition, the detection methods themselves also varied, which may have an effect on the reported concentration. For example, some studies report the use of HPLC for quantification and concentration coupled with GC/MS or other methods of detection to identify the specific compounds. However, because of the variability in the extraction and detection process from study to study, it is difficult to compare results between reports, and it is nearly impossible to determine whether or not the data corroborates previous work. As it stands, most of the bacterial GA studies are practically independent data points.

Section 1.5.6: Reports of GA concentration lack context.

Initially, bioassays were used to quantify GA production but because they are not reliable methods of quantification became obsolete once more sensitive and accurate methods became more affordable. Corroboration of measurements by different groups is not often possible because different microbial species are being tested for their ability to produce gibberellin. In the few cases where the same species has been tested, one of the reports has quantification data derived from bioassay results, so it is difficult to compare the two measurements. Variability in GA production data between particular bacterial strains cannot be completely understood because the phylogenetic relationship of most of these organisms remains unknown. Furthermore, assignment of particular isolates to certain genera was usually not based on 16 S rRNA, so if the genomes of the strains are elucidated, their names could change in the future.

To summarize these findings, serious efforts to quantify the production of gibberellin by bacteria have taken place during the past 50 years. The credibility of studies where bioassays were used would be improved through repeating the work with more sensitive techniques. Furthermore, 16s rRNA analysis of the strains reported to make GA as well as sequencing of the corresponding operons (or whole genomes) would be an excellent place to begin building a modern foundation for bacterial GA research. Additionally, a thorough analysis to determine if the use of complex or defined growth media affects the production or detection of GA from bacterial culture could be undertaken. A meta-analysis of this kind would help set a standard for recommended culture conditions for GA biosynthesis research by bacteria. Finally, an effort should be made to reconcile the reports of GA being detected from aerobic broth culture extracts when so far, no GA biosynthesis gene expression has been detected under these conditions. The current state of the literature on bacterial GA biosynthesis is troubling.

Section 1.6: Regulation of the GA biosynthesis genes in bacteria.

Regulation of nodulation and nitrogen fixation-related genes in *B. japonicum* USDA110 is controlled by the FixLJ-$FixK_2$-$FixK_1$ and RegSR-NifA regulons (see below). The Fix regulon controls the initial stages of the symbiotic relationship, as its gene products are involved in inducing the formation of the nodule and differentiation of the bacteroids (for reviews, please see Fischer 1994 and Mesa, Hennecke, and Fischer 2006). FixLJ is a two component regulatory system. FixL is the sensor-kinase and FixJ is the regulator and trans acting factor. Activity of FixL is inhibited by high oxygen concentrations, though at levels of 5% or less oxygen, it phosphorylates FixJ which activates transcription of *$fixK_2$* (Sciotti *et al*., 2003). The $FixK_2$ transcription factor regulates the expression large number of genes involved in nitrogen fixation including *fixGHISNOQP* and the *napEDAC* genes (Torres *et al*., 2011). $FixK_2$ also acts to control expression of *$fixK_1$* and also serves as a direct link to the RegSR-NifA two component regulatory system (Torres *et al*., 2011). $FixK_2$ activates transcription of *$rpoN_1$*, which produces the transcription factor $RpoN_1$, also known as σ^{54} (Torres *et al.*, 2011).

In 2010, Salazar and colleagues reported microarray and RT-PCR evidence showing that expression of analogous GA biosynthesis CYPs and the GGPPS analogs in *Rhizobium elti* CFN42 were down-regulated in NifA mutants in both free-living and symbiotic samples. This work indirectly shows that RpoN is involved in transcription of the GA genes (Salazar *et al*., 2010). This group also determined putative NifA and RpoN binding sites in *R. elti* CFN42, and found that RpoN binding sites are typically found downstream of NifA binding sites in this organism (Salazar *et al*., 2010). The researchers reported a NifA binding consensus sequence of $TGTN_{10}ACA$ and a RpoN binding site $TGGCACN_4TTGCW$ (Salazar *et al*., 2010). In early 2014, Hershey and collaborators reported using the RpoN consensus sequence to find potential RpoN binding sites in the genomes of *B. japonicum* USDA110, *M. loti* MAFF303099, and *S. fredii* NGR234, but failed to report the DNA sequences and location of the sites, though the researchers hinted that they could be important in control of GA biosynthesis.

When we used the RpoN and NifA sequences cited by Hershey and colleagues to search the vicinity of the GA operon in *B. japonicum* USDA110, only matches to the NifA binding site were found. However, using a more degenerate sequence, GG(-24)T(-14)GC(-12), reported by Dong and Mekalanos (2012) we did find putative RpoN binding sties nearby, shown in **Figure 1.4**.

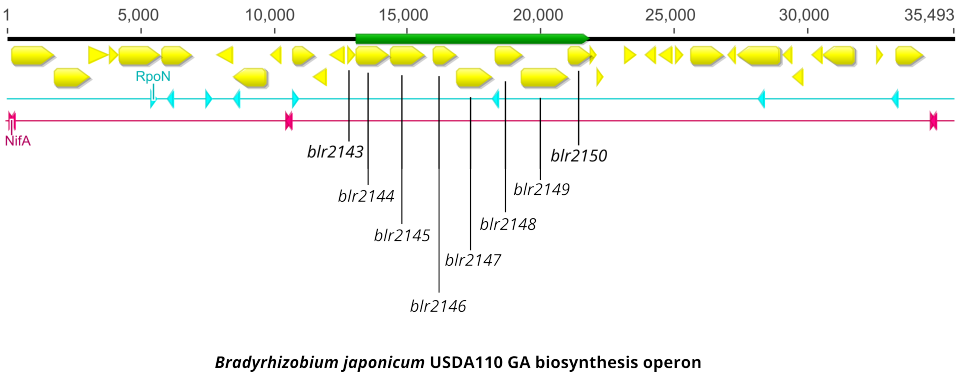


Figure 1.4: RpoN and NifA binding sites near putative GA operon in USDA110
The NifA binding site consensus sequence used was $TGTN_{10}ACA$ (Salazar *et al.*, 2010) The potential NifA binding sites that were found are shown in pink. The RpoN consensus sequence used was based off of the same work with additional insight from Dong and Mekalanos (2012). The query was $TGGN_8TGCW$ which still encompasses the conserved GG(-24)T(-14)GC(-12) RpoN binding motif found in 2010 by Salazar and colleagues. Potential RpoN binding sites are shown in blue. The GA operon is marked in green and all genes are shown in yellow. It appears that there are putative RpoN and NifA binding sites nearby the GA operon in *B. japonicum* USDA110.

RegSR-NifA is also independently activated at very low oxygen tension. At less than 0.5 percent atmospheric oxygen, the NifA receptor upregulates transcription of itself, expanding the signal cascade to include *nif*, *fix*, and other genes related to denitrification. Although RegSR is the top-level controller of the RegSR-NifA regulon, the stimulus for the two component RegSR regulator is not yet known (Torres *et al*., 2011). The variable oxygen tensions required for full diazotrophic activity of *B. japonicum* is thought to reflect the changing oxygen tension that the bacterial cells experience as they move from the rhizosphere to the interior of the plant root. The relationship between $FixK_1$ and NifA is complicated and has not yet been fully explored. While analyzing the transcriptomes of *B. japonicum* USDA110 Δ*fixJ*, Δ*fixK*$_1$, *and* Δ*fixK*$_2$ mutants, (Mesa *et al*., 2008) uncovered a set of genes that appears to

be negatively regulated by $FixK_1$ while also activated by NifA. This regulation pattern could be explained as part of a system to fine-tune gene expression as the oxygen tension decreases from 5 to 0.5 % during nodule organogenesis and bacteroid differentiation. Although FixL is active at an atmospheric oxygen tension of ≤5% it may still repress some genes until the oxygen tension is less ≤ 0.5% through repression by $FixK_1$. Eventually, as the oxygen tension decreases and NifA increases, repression by $FixK_1$ may be overridden by NifA and result in activation of these genes.

All of the GA biosynthesis genes in *B. japonicum* USDA110 are transcribed as part of an operon regulated by an anaerobic promoter (Tully *et al*., 1998, Chang *et al*., 2007, and others). In 2008, Mesa and colleagues found additional evidence to show that the operon is under anaerobic control. Transcription of the cytochrome genes in *B. japonicum* USDA110 appears to be specifically controlled by the FixLJ regulon (Mesa *et al*., 2008). Their study also found that there is an interaction between the Fix and RegSR-NifA regulon through the transcription factor RpoN (Mesa *et al*., 2008). A direct stimulant for the RegSR-NifA two component regulatory system is not yet known (Hauser *et al*., 2007). RpoN binding sites have been reported within the *B. japonicum* USDA110 GA biosynthesis operon. Using in silico modeling, the *M. loti* MAFF303099 and *S. fredii* NGR234 GA biosynthesis operons were found to have RpoN transcription factor binding sites in their promoters (Dombrecht *et al*., 2002). The RpoN-NifA regulon is known to be involved in signaling transcription of symbiotic genes, and RpoN is the transcription best known for enabling transcription of nitrogen fixation genes in low oxygen conditions (Hauser *et al*., 2007). The consensus sequence for RpoN binding is TG**G**CACG-NNNN-TTG**C**W. The bold G and C are located at -24 and -12 relative to the transcription start site (Dombrecht *et al*., 2002). In some rhizobia, including *B. japonicum* USDA110 and *M. loti* MAFF303099, there are two copies of the *rpoN* gene encoding products: $RpoN_1$ and $RpoN_2$ (Kaneko *et al*., 2002; Kaneko *et al*., 2000). In *B. japonicum* the RpoN-NifA regulon is known to as the RegSR-NifA regulon.

In *B. japonicum* USDA110, deletion of *fix*L and fix_{K1} resulted in an increase in transcription of the GA biosynthesis genes *blr2145*, *blr2147*, *blr2148*, and *blr2149* (Hauser *et al*., 2007). However, deletion of $fixK_2$ resulted in markedly decreased transcription of the same genes (Mesa *et al*., 2008). Data showing fold changes in transcription demonstrate how the operon is regulated in both mutant strains and these results are summarized in **Table 1.3**. The data suggest that the $fixk_2$ gene may be under the control of a currently unknown regulator and that transcription factor $FixK_1$ directly represses the GA biosynthesis operon in *B. japonicum* USDA110 (Hauser *et al*., 2007; Mesa *et al*., 2008). Both studies were carried out independently using different gene chips and extraction techniques. My interpretation of these data follows the table. All of the numbers are calculated from comparison to the wild-type strain. A positive integer indicates that transcription of the gene was higher in the mutant strain compared to the wild-type while a negative integer indicates that transcription of the gene was lower in the mutant strain than in the wild-type strain.

Table 1.3: Transcription of GA genes in mutant *B. japonicum* USDA110 strains

	***Bradyrhizobium japonicum* USDA110 mutant strain**				
Locus	***ΔfixJ***	***ΔfixK$_1$***	***ΔfixK$_2$***	***ΔnifA***	***ΔrpoN$_{1/2}$***
blr2143		ND		-111.4	-111.5
blr2144	5.9	6.1	ND	-37.9	-31.3
blr2145	7.9	7.6	2.3	-20.4	-16.8
blr2146		ND		-84.7	-85.5
blr2147	7.5	6.8	2.2	-21.3	-20.0
blr2148	7.5	6.2	2.2	-40.7	-44.6
blr2149	8.8	7.4	2.6	-16.3	-17.3
blr2150		ND		-7.5	-7.4

Data was not detected (ND) for the hypothesized non-functional, pseudo-cytochrome P450 (*blr2143*) or for *ent*-kaurene synthase (*blr2150*) in the *fixJK$_2$K$_1$* deletion mutants strains (Hauser *et al*., 2007). It is possible that the genes were upregulated but that the increase was below the statistical threshold for significance of the data. It is also possible that the sample used did not completely represent the entire transcriptome of *B. japonicum* USDA110. Fortunately the other data is valid and informative. It suggests that transcription of the GA biosynthesis genes are negatively controlled by the Fix trans factors, since deleting them increases expression of the target genes compared to the wild-type strains (Hauser *et al*., 2007). On the other hand, when the genes encoding the primary transfactors of the RegSR-NifA regulon were deleted, transcription of all of the GA biosynthesis genes was greatly decreased in the mutant strains compared to the wild-type. In this case, it appears that the transcription of the GA biosynthesis genes is positively regulated by RpoN and NifA (Mesa *et al*., 2007). Although this data is found in these two publications, the information is not the focus of the paper. My interpretation of the supplementary information allowed me to come to these conclusions. A diagram showing how the FixLJ-FixK$_2$-FixK$_1$ and RegSR-NifA regulons interact to control transcription of the gibberellin operon is shown in **Figure 1.5**.

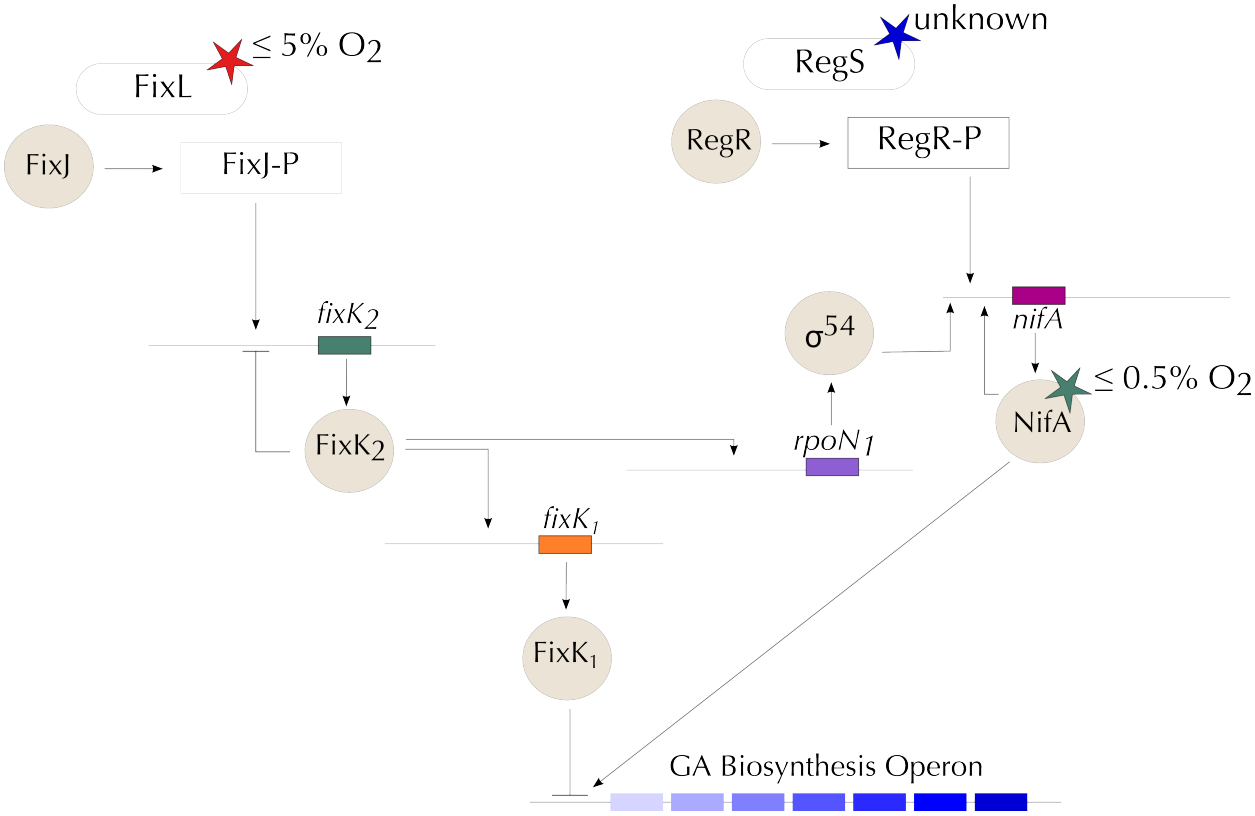


Figure 1.5: Schematic diagram showing interaction of FixLJ-FixK$_2$-FixK$_1$ and RegSR-NifA regulons to control transcription of GA biosynthesis operon in *B. japonicum* USDA110

Each grey circle shows the transcription factors involved in each regulon. There are many other targets of all of the transcription factors shown in this image, but this figure only shows the interactions that are relevant in GA biosynthesis. Genes are shown by various coloured boxes and are labeled with their name. Arrow heads indicate that a transcription factor has a positive effect on transcription of the gene while lines show that the effect of transcription is negative. FixL is the oxygen sensing part of the two component regulatory system and is prevented from phosphorylating the histidine kinase FixJ at oxygen concentrations of greater than 5%. At less than 5% oxygen concentration, phosphorylated FixJ increases transcription of the *fixK$_2$* gene. Transfactor FixK2 autoregulates its own transcription and increases the transcription of *fixK$_1$* and *rpoN$_1$*. FixK1 directly inhibits transcription of the GA biosynthesis operon (shown in blue at the bottom of the diagram). Transcriptional regulation of *rpoN$_1$* by FixK2 is significant because it directly links the two major regulons through the σ^{54} transcription factor. At this time, the stimulus of RegSR is not known, but it is thought to activate transcription of *rpoN1* as well, while NifA is known to be active as a transcription factor at oxygen concentrations of less than 0.5%. The data obtained by Mesa and colleagues using the *rpoN* and *nifA* gene deletion mutants shows that without the RpoN and NifA transfactors, transcription of the GA biosynthesis operon is greatly decreased compared to the wild-type (2008). Overall, when this result is taken together with the work from Hauser *et al*., (2007) showing that FixLJ regulon negatively controls GA biosynthesis gene transcription, a clearer picture of GA gene regulation during nodule organogenesis emerges. It will be important to study these mechanisms in more detail in the future. It is clear that in very low oxygen tension, transcription of the GA biosynthesis genes is most likely increased.

The GA biosynthesis operon of *B. japonicum* USDA110 appears to be negatively controlled by the FixLJ-$FixK_2$-$FixK_1$ regulon transfactors (Mesa *et al*., 2008), but activated by transfactors of the RegSR-NifA regulon (Hauser *et al*., 2007). The relationship between these two regulatory networks is complicated by the negative regulatory role of $FixK_1$ on the GA biosynthesis genes, apparent interplay between FixJ and $FixK_2$, and the uncertainty as to whether there are factors other than FixJ that act to control transcription of the *fixk*$_2$ gene. In any case, $FixK_2$ appears to be important for the regulation of the genes in environments of less than 5% oxygen, where transcription is repressed while RegSR-NifA is more important at lower oxygen tension, where transcription is increased.

Section 1.6.1: Additional studies could help determine a role for bacterial GA

A number of transcriptomic and proteomic studies of *B. japonicum* USDA110 have looked at differential transcription and protein expression profiles in free-living and symbiotic conditions. However, depending on when symbiotic samples are collected, the transcriptome and proteome of the bacteroid may be different. So it can be a challenge to compare the results of different surveys that look at how the genes are transcribed and translated. Issues that arise regarding the steady-state level of protein and stability of transcripts much also be addressed. We believe it is important to sample a range of time points in the nodule life span in order to understand exactly when the bacterial GA genes are active. We hope that this information will lead to a hypothesis to explain the biological role of bacterial GA. With that said, I will now summarize the reports of *B. japonicum* USDA110 GA biosynthesis genes and proteins being detected by next-generation methods.

The first study was a comparative microarray analysis of free living and symbiotic *B. japonicum* USDA110 bacteroids (Chang *et al*., 2007). The researchers found that the transcription of the CYPs involved in GA biosynthesis were upregulated in bacteroids compared to free-living cells (Chang *et al*., 2007). In 2010, Delmotte and colleagues undertook a large-scale analysis of both the transcriptome and

proteome of *B. japonicum* USDA110 cells in free-living, aerobic culture (the control) and symbiotic bacteroids harvested at 21 days post infection. The researchers found many proteins encoded by the GA biosynthesis operon in bacteroids of soybean nodules grown in both the laboratory and in field trials. Delmotte and colleagues focused on identification of proteins using LC-MC/MS analysis, so we can be sure that the transcripts were fully expressed (Delmotte *et al*., 2010). Having protein data to accompany the transcriptomic studies is a great asset because it shows that the GA operon transcripts are actually transcribed *in vivo*. This study is a significant advancement because it confirms the earliest report from Tully and Keister (1993) where CYP112 and CYP114 were isolated from *B. japonicum* USDA110 bacteroids for the first time. In another study, the *B. japonicum* USDA110 bacteroid proteome was sampled during different phases of nodule development (Nomura *et al*., 2010). The authors showed that the GA biosynthesis proteins were more abundant in their samples after the onset of nitrogen fixation at 14-28 days post inoculation (Nomura *et al*., 2010).

Recent studies have shown that oxygen tension is very important for regulation of the GA biosynthesis operon in *Bradyrhizobium japonicum* and isolated stimuli likely present in the rhizosphere during nodule organogenesis have not been found to have an effect on transcription or translation of the genes. For example, two studies examined the expression profile of the *B. japonicum* USDA110 genome in response to stimulants designed to mimic the oxidative burst generated by the initial rhizobial invasion. Jeon and colleagues (2011) looked at the effects of hydrogen peroxide on genome-wide transcription while Donati and others looked at the effect of paraquat (2011). Neither group detected any effect on transcription of the GA biosynthesis operon. The effect of the flavonoid 20 μM coumestrol also had no effect on GA biosynthesis genes, based on analysis of total RNA (Lee et al., 2012). There was no observable change in GA biosynthesis protein expression in *B. japonicum* SEMIA 5079 in response to 1 μM genistein (da Silva Batista and Hungria, 2012). Similarly, auxin does not appear to induce GA

biosynthesis gene expression either (Donati *et al*., 2013).

Inhibition of GA biosynthesis in *B. japonicum* USDA110 by genetic disruption of the loci encoding the P450 enzymes failed to show a phenotypic effect on the host plant as normal, functional nodules were formed (Tully and Keister, 1993). So, although the operon is regulated by the same signaling cascades that regulates nodule organogenesis and nitrogen fixation, and appears to be specifically controlled by oxygen concentration, GA biosynthesis does not appear to be crucial to either nodule organogenesis or nitrogen fixation. Statistical analysis could be done to determine if GA biosynthesis confers any advantage to a bacterial strain; perhaps GA production makes the bacteria more competitive than other strains found in the rhizosphere, or perhaps production of GA allows the strain to form more or larger nodules. At this point, there is still much to learn about bacterial GA biosynthesis and the biological role of the genes.

Section 1.6.2: Studies of GA biosynthesis genes in *Mesorhizobium loti* MAFF303099

Transcriptomic data obtained by microarray analysis of *M. loti* MAFF303099 bacteroids compared to free-living cultures also showed that transcription of the putative GA biosynthesis operon was upregulated in bacteroids compared to free living cells (Uchiumi *et al*., 2004). The authors also identified similarly regulated, analogous genes in *M. loti* R7A and demonstrated that both operons were part of an expression island within the symbiotic island of the *M. loti* genome (Uchiumi *et al*., 2004).

Analysis of the *M. loti* MAFF303099 proteome in free-living and symbiotic conditions also found that the GA biosynthesis enzymes were only expressed during symbiotic growth (Tatsukami *et al*,. 2013). This correlates with the data from *B. japonicum* USDA110 and supports the hypothesis that GA biosynthesis may have a role in nodulation or symbiotic nitrogen fixation. This study established a rapid extraction method for determination of total proteins using LC-MS/MS analysis and the authors used it to compare the proteoemes of free-living, aerobic *M. loti* MAFF303099 culture with the

proteome of MAFF303099 in symbiosis with *L. japonicus*. Many proteins were found to be solely expressed in the symbiotic condition, including the putative GA biosynthesis proteins (Tatsukami *et al*., 2013). The only protein that was not observed in the study was the one corresponding to the *mlr6370* locus, the class I terpene cyclase responsible for converting *ent*-CPS to *ent*-KA (Hershey *et al*., 2014). Interestingly, the authors focused on the expression of Mlr6368, the GGPPS, and hypothesized that the protein could be providing FPP to the plant as a secondary metabolite. The authors stated that no other isoprenoid biosynthesis proteins were found and overlooked the other gibberellin biosynthesis proteins. Regardless of the authors' conclusions, these results provide further evidence the putative GA biosynthesis genes are expressed during times of nodulation and nitrogen fixation, lending more circumstantial evidence to the hypothesis that the operon is regulated by the same signal cascade. An *M. loti* MAFF303099 mutant library constructed using transposon insertions has cataloged insertion mutants in all of the putative GA biosynthesis operon genes, but further characterization of these strains has not yet been reported (Shimoda *et al*., 2008a).

Yeast two-hybrid assays have also been used in an effort to identify protein-protein interactions within *M. loti* MAFF303099 (Shimoda *et al*., 2008b). Yeast two-hybrid screening identifies proteins that interact with each other using two strains of yeast, each one harbouring a different expression vectors. Candidate proteins are expressed as a translational fusion to one part of the *GAL4* gene; GAL4 is the β-galactocidase heterodimer. Bait proteins are also translationally fused to the remaining part of the *GAL4* gene, on the other vector. Usually, the bait vectors are constructed with random library fusions to maximize the size of the analysis that can be performed in a single screen. When the bait and target proteins are co-expressed, the fragments of the *GAL4* gene will also be expressed. When the bait and prey target proteins come in close proximity, the domains of GAL4 are close enough to create a functional enzyme with an observable phenotype. All positive interactions need to be confirmed by

independent methods because the rate of false positives is high (for a review, please see Ferro and Trabalzini, 2013). False positive results are very common since heterodimeric proteins can interact with themselves, or a promiscuous interaction between two proteins under experimental conditions may form but it is not repeatable under biological conditions.

Y2H was done using *mlr6364* and *mlr6370* as candidate (bait) proteins exposed to a random library of target fusions (Shimoda *et al*., 2008b). Other genes of the operon were also selected for testing but failed at the construction step. One interaction between the cytochrome P450 encoded by *mlr6364* was found and it was with the nodulation protein NolX (encoded by locus *mll6337*). Six other protein-protein interactions between the class I terpene cyclase encoded by *mlr6370* were also found. All interactions are summarized in **Table 1.4**.

Table 1.4: Some Putative Protein-Protein Interactions in *M. loti* MAFF303099

Bait	Function	Identified Interaction	Function
mlr6364	CYP	*mll6337*	NolX—root hair formation/elongation
mlr6370	class I terpene cyclase (kaurene synthetase)	*mll4535*	hypothetical protein
		mll5900	hypothetical protein
		mll6123	5-methyltetrahydroteroyltriglutamate-homocysteine methyltransferase
		mlr0590	hypothetical protein
		mlr7120	AgaE: conversion of agropinic acid to mannopinic acid
		mlr7167	acetylonithine deacetylase

The interaction between *mlr6364* and *mll6337* suggests that there may be a relationship between gibberellin biosynthesis and root hair curling or elongation. A NolX homolog in *S. fredii* USDA257 has been localized to root hairs during early nodulation stages and seems to be involved in bacteria and host specificity although its precise role is unknown (Krishnan, 2002). Although the Y2H assay results appear to directly link gibberellin biosynthesis and nodule organogenesis, the interaction between these two proteins should be confirmed by independent analysis.

Section 1.6.3: Studies of GA biosynthesis genes in *Sinorhizobium fredii* NGR234

In *S. fredii* NGR234, a few studies of the pSym transcriptome during symbiosis have been carried out. In the first study of its kind in *S. meliloti* NGR234, RNA was isolated from free-living cultures, cultures treated with daidzein, bacteroids of determinate *Vigna unguiculata* nodules, and bacteroids from indeterminate nodules on *Cajanus cajan* (Perret *et al*., 1999). Daidzein is a soybean flavonoid that induces *nod* gene transcription (Rolfe, 1988). *S. fredii* NGR234 is well known for its ability to nodulate a wide variety of legume hosts. This experiment provided an opportunity to study determinate and indeterminate nodules inhabited by the same bacterial species. The results showed that the GA biosynthesis operon was transcribed in both determinate and indeterminate nodules at the same relative level (Perret *et al*., 1999) However, some other transcripts were only found in the determinate nodules (Perret *et al*., 1999). Additionally, some transcription, (though relatively less than that observed within the nodules) of the first two genes in the operon (y4lD and y4lC corresponding to the first two CYPs) was observed after 1 hour incubation of free-living *S. fredii* NGR234 cultures with daidzein (Perret *et al*., 1999).

Recently, RNA-seq analysis has been used to investigate the transcriptomic differences between free-living and symbiotic *S. fredii* NGR234 in determinate nodules of *V. unguiculata* and indeterminate nodules of *Leucaena leucocephala* (Li *et al*., 2013). The authors reported that over 90% of the genes found on pNGR234a (the symbiotic plasmid) were upregulated in the nodules formed on both plant hosts. RNA corresponding to the GA biosynthesis genes in *S. fredii* NGR234 was found in both indeterminate and determinate nodules, indicating that the operon is expressed under symbiotic conditions in this organism as well (Li *et al*., 2013).

The authors also compared the relative abundance of the transcripts from both types of nodules. They found that some genes, including some in the GA biosynthesis operon, are expressed at a significantly lower level in indeterminate nodules of *L. leucocephala* compared to determinate nodules.

These genes were NGR234_a02720 (encoding a ferredoxin), NGR234_a02750 (encoding a GGPPS), NGR234_a02760 (encoding a class II TC), and NGR234_a02770 (encoding a class I TC) (Li *et al*., 2013). Some of the RNA-seq data was confirmed using qRT-PCR, but unfortunately the genes of the GA biosynthesis pathway were not selected for further analysis.

So far, proteomic studies of *S. fredii* NGR234 have not been reported, but based on comparison of the *S. fredii* NGR234, *M. loti* MAFF303099, and *B. japonicum* USDA110 transcriptomic data, and the similarities between all three datasets, it is likely that proteins corresponding to the putative GA biosynthesis locus in *S. fredii* NGR234 will also be found in bacteroids samples.

Section 1.7: There may be a link between GA biosynthesis and nodule morphology

Soybeans nodules inhabited by *B. japonicum*, and *L. japonicus* nodules inhabited by *M. loti* MAFF303099 have determinate nodules. The data showing that GA biosynthesis genes are not as highly transcribed in indeterminate nodules of *L. leucocephala* inhabited by *S. fredii* NGR234 suggests that there may be a link between nodule shape (specifically determinate nodule shape) and gibberellin biosynthesis by bacteria. Interestingly, *Sinorhizobium meliloti* 1021, the symbiotic partner of alfalfa, does not have any gibberellin biosynthesis genes and it forms indeterminate root nodules. In addition, *Rhizobium leguminosarum* bv. *viciae* 3842, the symbiont of pea, does not have a gibberellin biosynthesis operon, while *R. leguminosarum* bv. *phaseoli*, the symbiont of bean, does. Peas plants form indeterminate nodules while bean plants form determinate nodules. There are several *R. leguminosarum* strains that have been isolated from determinate nodules that have GA operons, but none have been isolated from indeterminate nodules. For a review of nodule morphology, please see Guinel, 2009. This hypothesis is discussed further in Chapter 3.

Section 1.8: Study of GA biosynthesis genes in other rhizobia.

Recent work with the closely related strain *M. loti* R7A confirmed that the GA biosynthesis operon is under RpoN-NifA regulation, and that knockout mutants were still able to form nitrogen-fixing nodules on *L. japonicus* (Sullivan, Brown, and Ronson, 2013). The RpoN-NifA regulon of *Rhizobium etli* CFN42, another relative of *M. loti* and *S. fredii*, has also been studied and found to control homologous GA biosynthesis genes during symbiotic growth (Salazar *et al*., 2010). When the *nifA* gene was deleted in *R. etli* CFN42, transcription of the homologous GA biosynthesis genes were reduced compared to the wild-type strain in symbiosis with *Phaseolus vulgaris* (Salazar *et al*., 2010).

There is much work to be done to determine the role of gibberellin biosynthesis in bacteria. Because rhizobia can effectively inhabit nodules and fix nitrogen without functional GA biosynthesis genes, it is not an essential function for the bacteria nor legume. It's possible that GA biosynthesis by microbes is more useful in non-sterile soils (such as those in an agricultural field). In *Lotus japonicus*, the plant-partner of *M. loti*, application of exogenous GA_3 induced pseudonodule formation by causing cell division within the pericycle of the roots; however, theauthors found that this effect could be totally repressed by addition of 15 mM potassium nitrate or ammonium nitrate (Kawaguchi *et al*., 1996).

Another hypothesis suggests that GA production by bacteroids may delay nodule senescence (Dupont *et al*., 2013). Much work has been done to study the regulation of these genes, but relatively little has been done to answer the question of why bacteria produce gibberellin. Besides attempts to quantify the amount and type of gibberellin produced and noting the nodule and fixation phenotype of mutants, little work has been done to answer this question. I hope that future studies will concentrate on exploring a competitive advantage for bacteria with GA biosynthesis genes and explore this relationship in terms of plant or nodule size, nodule number, or nodule occupancy to begin to answer this question.

Section 1.9: Biosynthesis of gibberellin in *A. thaliana* and *F. fujikuroi*.

Gibberellins are synthesized from isoprenoid precursors formed in one of two primary metabolic pathways: the mevalonic acid pathway and the methylerythritol phosphate pathway (Chang *et al*., 2013). Fungi use the MVA pathway to synthesize the GA building blocks, while bacteria typically use the MEP pathway (Bömke and Tudzynski, 2009). Plants have the ability to use either the MVA pathway (located in the cytoplasm) or the MEP pathway (located in the plastid) pathway, but mostly rely on the MEP pathway (Kasahara *et al*., 2002). The isoprenoid subunits that are condensed to form GGPP are isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP).

In the fungus *F. fujikuroi*, the condensation reaction between IPP and DMAPP is catalyzed first by a farensyl diphosphate syntethase and then by one of two geranylgeranyl pyrophosphate syntethase enzymes in the genome of the organism (Tudzynski, 2005). One of the GGPS enzymes (GGS1) is involved in isoprenoid biosynthesis forming GGPP fated for a variety of terpenoid natural products formed elsewhere in the organism, while the other GGPPS, (GGS2) is specifically designated for GA biosynthesis (Tudzynski, 2005). A similar system with a GA specific GGPPS enzyme has been proposed for *A. thaliana* as well, because that organism contains 12 known GGPPS genes (Hedden and Thomas, 2012). In contrast, in the genome of *B. japonicum* USDA110 there is only one gene, *blr*2148 that encodes a GGPPS and it is found in the GA biosynthesis operon. In *M. loti* MAFF303099 and *S. fredii* NGR234, a homologous gene is known and has been confirmed to produce GGPP from IPP and DMAPP (Hershey *et al*., 2014). In some bacteria, it is possible that there are other GGPPS genes in the chromosome that are used for the biosynthesis of other compounds.

GGPP is cyclized to form *ent*-copalyl diphosphate (also known as ent-copalyl pyrophosphate), then converted to *ent*-kaur-16-ene and finally to *ent*-kaurenoic acid (MacMillian, 2002). In *A. thaliana* and *B. japonicum* USDA110, the conversions from GGPP to *ent*-CPP to *ent*-kaurene are done by two

mono-functional enzymes encoded by two different genes (MacMillan and Beale, 1999; Morrone *et al*., 2009). The *F. fujikuroi* fungus, on the other hand, can catalyze the conversion of GGPP to *ent*-CPP to *ent*-KA using a single bi-functional enzyme (Bömke and Tudzynski, 2009).

In plants, fungi, and likely bacteria, *ent*-kaurene is converted to *ent*-kaurenoic acid (*ent*-KAO) by a cytochrome P450 (Bömke and Tudzynski, 2009). *Ent*-KAO is converted to the first intermediate GA, either GA_{12} in *A. thaliana* or GA_{14} in *F. fujikuroi* and possibly bacteria, by another CYP (Bömke and Tudzynski, 2009). In plants, the cytochrome P450 enzymes that catalyze these reactions are located in the endoplasmic reticulum and plastid membrane and they are called *ent*-kaurene oxidase and *ent*-kaurenoic acid oxidase, respectively. Next, GA_{12} is converted to other gibberellins by 2-oxoglutarate dependent deoxygenates—however, the intermediate GAs are not bioactive (Fischbach and Clardy, 2007). Eventually, GA_3 is synthesized by the plant. In contrast, fungi mainly produce GA_{14} from *ent*-KAO and use only P450s in the conversion of it to other gibberellins (Fischbach and Clardy, 2007). Additionally, although plants and fungi use similar enzymes, they do not appear to be very closely related. The plant and fungal P450s belong to different cytochrome clades: for example the plant kaurene oxidase (KO) belongs to the CYP701A class, while the kaurenoic acid oxidase (KAO) belongs to the CYP88A class. The fungal homologs of KO and KAO respectively belong to the CYP503 and CYP68 clades (Hedden and Thomas, 2012).

We would like to examine the plant and fungal pathways in context with the *B. japonicum* USDA110 GA pathway, and reiterate the similarities and differences found in the three kingdoms. First of all, *B. japonicum* USDA110 has three genes for CYPs *(blr2144, blr2145 and blr2147*) and one SDR (*blr2146*). The SDR could catalyze the conversion of GA_4 to GA_7 similar to the reaction catalyzed by the desaturase protein in fungi. In the bacterial GA biosynthesis operon, there are mono-functional terpene cylcases (like plants), and a lack of homologous soluble oxygenases genes (like in fungus). It appears

that the combination of CYPs and SDR is enough to make GA, also like fungi. Therefore, the genetics of the *B. japonicum* USDA110 GA biosynthesis operon suggests that while the biosynthesis of GA in bacteria is similar to that in plants and fungi, it may represent a third case of convergent evolution for this function. The main questions—why and exactly how are bacteria producing plant hormones—still remain to be answered. A schematic drawing of the GA biosynthesis pathway is shown in **Figure 1.6**.

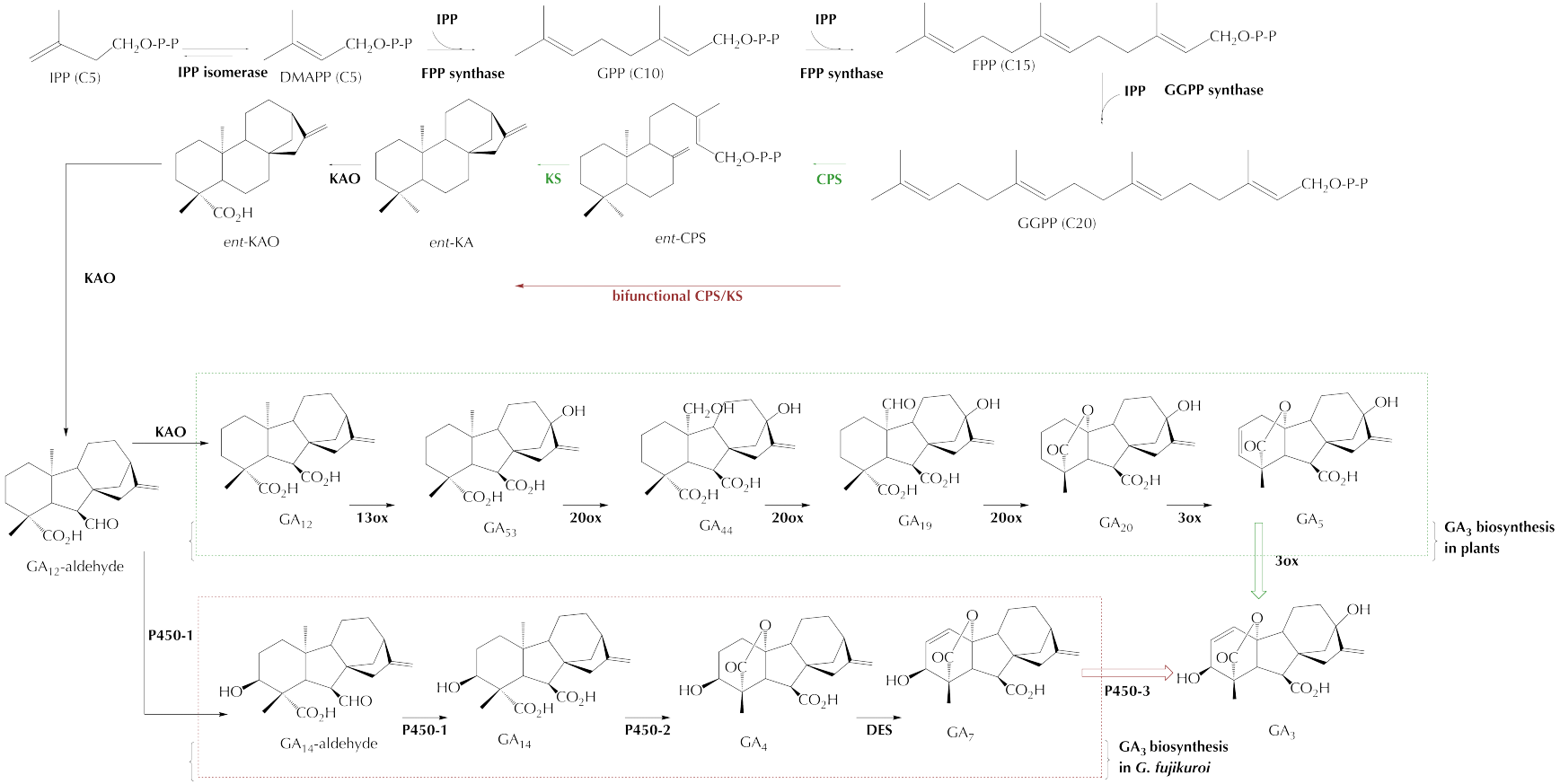


Figure 3: Gibberellin GA_3 biosynthesis in plants and the fungus *G. fujikuri.*

Figure 1.6: Gibberellin GA_3 Biosynthesis in *Arabidopsis thaliana* and *F. fujikuroi*

This figure demonstrates the differences in the GA biosynthesis pathways of plants and the fungus *Fusarium fujikuroi*. It is a simplified schematic of GA= biosynthesis based on figures in the 2009 paper by Bömke and Tudzynski. In plants, like bacteria, there are two mono-functional terpene cyclases that catalyze the conversion of GGPP to *ent*-CPS and then *ent*-KA. In fungi, both reactions are catalyzed by a single bi-functional enzyme. In plants, the intermediate reactions between GA_{12}-aldehyde and GA_3 are catalyzed by soluble oxyenases enzymes In *F. fujikuroi*, three cytochrome P450s and one alcohol dehydrogenase are used instead. In bacteria, the enzymes in this portion of the pathway have not yet been characterized; however, the operon encodes genes for three P450s as well as one SDR. It's very likely that the latter half of the pathway in bacteria is similar to that of *F. fujikuroi*. Overall, the bacterial pathway seems to resemble the plant GA biosynthesis pathway in the initial steps, and the fungal pathway in later steps.

Section 1.10: Chapter Summary

The physiological role of gibberellin synthesized by bacteria remains a mystery. Not only can we not say *why* bacteria make gibberellin, we cannot yet say *how* it is synthesized. The genetic regulation of GA biosynthesis is also complicated, and as of now a clear regulatory network governing these genes during all types of bacterial life has not yet been elucidated—though several excellent studies have created a strong framework for further analysis. This chapter reviewed the current knowledge of how the FixLJ-$FixK_2$-$FixK_1$ and RegSR-NifA signal cascade regulate expression of the GA biosynthesis genes in *B. japonicum* and looked at the genetic and enzymatic differences in the gibberellin biosynthesis pathway in plants, fungi, and bacteria.

The profound differences in the plant and fungal GA biosynthesis pathway at both the genetic and enzymatic level, pose another enigma—why has the pathway developed independently in plants and fungi? Where does the potentially novel bacterial system of GA biosynthesis fit in? Finally, are there links between GA biosynthesis and nodule occupancy, fitness, or nitrogen fixation? So many questions remain unanswered. There is still so much to learn about this plant hormone.

Chapter 2: Expression of Putative P450s Involved in GA Biosynthesis and Deletion of The Corresponding Genes in Their Native Hosts

Analysis of recombinantly expressed cytochrome P450 proteins and metabolic intermediates accumulated by deletion mutants.

Comprising experiments carried out at Iowa State University in Ames, Iowa, USA under the direction of Dr. Reuben Peters and work done in the laboratory of Dr. Trevor Charles at the University of Waterloo in Waterloo, Ontario, Canada. Results from our collaborators are also presented and analyzed.

Section 2.1: Chapter Summary

This chapter summarizes the laboratory experiments performed in fulfillment of this thesis. The cytochrome P450 (CYP) encoding genes from *Bradyrhizobium japonicum* USDA110 were amplified by PCR. The products were cloned into a number of plasmids and heterologously expressed in *Escherichia coli* strains. Attempts to purify N-terminal His-tagged CYPs were made. A small amount of CYP112 was obtained using an expression construct in pET30a(+). Efforts to obtain purified CYP114 and CYP117 using similar methods were unsuccessful. We used a plasmid based metabolic engineering system developed by our collaborators to test the substrate specificity of each CYP. Though we were unable to repeat our results, preliminary data showed that CYP117 is responsible for the conversion of *ent*-kaurene to *ent*-kaurenoic acid. Deletion mutants of several CYP genes were made using homologous recombination of genomic DNA cloned into the integration vector pK19*mobsacB* (Schäfer *et al*., 1994). In *B. japonicum* USDA110, *blr2144* and *blr2145* was deleted. In *Sinorhizobium fredii* NGR234 all three homologous CYP genes were deleted, and the *blr2147* homolog in *Mesorhizobium loti* MAFF303099 was also deleted. The *B. japonicum* deletion strains were used by our collaborators in experiments to determine the substrate specificity and role of the corresponding protein in bacterial gibberellin (GA) biosynthesis.

Section 2.2: Introduction and Rational

The putative gibberellin biosynthesis operon in *Bradyrhizobium japonicum* USDA110 has eight ORFs which encode three cytochrome P450s (CYPs), one short-chain dehydrogenase (SDR), a ferredoxin, a geranylgeranylpyrophosphate synthase (GGPPS), an *ent*-kaurene synthase and *ent*-copalyl pyrophosphate synthase. At the beginning of this project, the biochemical role of the *ent*-copalyl pyrophosphate and *ent*-kaurene synthase was known and the role of the GGPPS was implied through indirect evidence that arose from study of the first two enzymes. (Morrone *et al*., 2009). The roles of the CYPs, SDR and ferredoxin were unknown, but from the analogous GA biosynthesis pathways in

Fusarium fujikuroi and *Arabidopsis thaliana*, we knew that the CYPs and SDR would most likely sequentially catalyze the conversion of *ent*-kaurene to GA intermediates.

From studies on fungal GA production, we suspected that the bacterial CYPs could have multiple functions and be able to act on multiple substrates under different reaction conditions. GC/MS analysis of culture exudates of rhizobia found GA_1, GA_3 and GA_4 produced by a variety of microbes (reviewed in Chapter 1). The synthesis pathway of these three GAs is significantly different; if GA_1 or GA_4 is the final GA produced, then it is unlikely that GA_3 is also synthesized due to the nature of the biochemical catalysis carried out by the CYPs. At the time that the project began, there had been one report on the isolation and identification of GA_3 from *B. japonicum* culture (Boiero *et al*., 2007). GA_3 is gibberellic acid and is commercially important because of its use in agriculture and other industries. We were encouraged by this study and undertook our work with the primary goal of determining the substrate specificity of the CYPs and the ferredoxin. By process of elimination, we hoped to determine the function of the SDR as well. We initiated a genetic approach to delete each of the target genes and a method to heterologously express each of the CYPs in *E. coli* hosts. The strains used to do this are listed in **Table 2.1.**

Briefly, our genetic approach to delete the CYPs and SDR genes in the putative GA biosynthesis operons in *B. japonicum* USDA110, *M. loti* MAFF303099 and *S. fredii* NGR324 utilized double homologous recombination with the integration vector pK19*mobsacB* (Schäfer *et al*., 1994). This method is a simple and effective way to completely delete a specific target. Two benefits of using this system are that the resultant mutants are marker-less and the genetic manipulations do not result in polar effects on the surrounding genes once the second recombination event has occurred. To create the targeted gene deletions, we amplified 500 base pairs upstream and downstream of each gene and joined them together using crossover PCR (Sukdeo and Charles, 2002). Once both regions were joined, they were cloned into pK19*mobsacB* and introduced into *B. japonicum*, *M. loti*, and *S. fredii* by conjugation. The tri-parental matings occurred between *E. coli* DH5α harbouring the mobilizer plasmid pRK600, a donor *E. coli* DH5*α* strain harbouring pK19*mobsacB* with the deletion construct and the recipient strain (*B. japonicum* USDA110, *S. fredii* NGR234, or *M. loti* MAFF303099). Integration of the plasmid was positively selected with antibiotics while the second recombination event was negatively selected using the *sacB* gene. pK19*mobsacB* integrates into the recipients' genome at the homologous region either upstream or downstream of the DNA to be deleted. The origin of replication on the plasmid is functional in *E. coli*, but not functional in members of the Rhizobiales. To confer antibiotic resistance, the plasmid must integrate into the genome. The second crossover, to remove the plasmid and ideally deletion target, must occur at the second homologous region. If it occurs at the same site as the insertion, the genotype reverts to wild-type without deletion of any genomic DNA. Using phenotypic screens we were able to isolate putative double crossover mutants which we checked the genotype of using PCR. We confirmed the deletion mutant genotypes with Southern blots.

Because of the potential for the bacterial CYPs to have multiple functions and substrates, we wanted to express and purify His tagged CYPs from the *E. coli* expression vector pET30a(+) so that we could purify them and attempt biochemical assays with different substrates and conditions. We tried several methods to optimize the expression conditions of the CYPs including making truncations of the N-terminus, and changing the expression vector and *E. coli* host strain. One of the last methods that we tried to achieve heterologous expression of the CYPs was a plasmid based metabolic engineering system developed by one of our collaborators. The system is visualized in **Figure 2.1**. For details on the system, please consult the publications by our collaborators (Cyr e*t al*., 2007; Morrone *et al*., 2010).

The system was initially used to study labdane cyclases and other enzymes that make compounds related to the GA intermediates kaurene and kaurenoic acid (Peters *et al*., 2000). Enzymes that produced quantities of two certain enantiomers, specifically *ent*-kaurene and *ent*-kaurenoic acid were found, and the system was expanded to study enzymes which used these compounds as substrates (Harris *et al*., 2005). Our collaborators used the system to begin characterization of the *B. japonicum* USDA110 GA biosynthesis pathway and discovered the two enzymes responsible for the conversion of geranylgeranylpyrophosphate to *ent*-kaurene (Morrone *et al*., 2009).

The main benefit of the plasmid based metabolic engineering system is that the enzymes and their substrates can be synthesized in the same cell which enables direct extraction and analysis of the products. This is beneficial because it circumvents the purification of the recombinant protein and simplifies the process to determine the substrate specificity of the CYPs. Additionally, *ent*-kaurene and *ent*-kaurenoic acid can be purified from the culture and used as substrate for *in vitro* biochemical assays when necessary, for example, when determining enzyme kinetics. The system provides a reliable source of a substrate that is difficult to chemically synthesize (Mander, 2003).

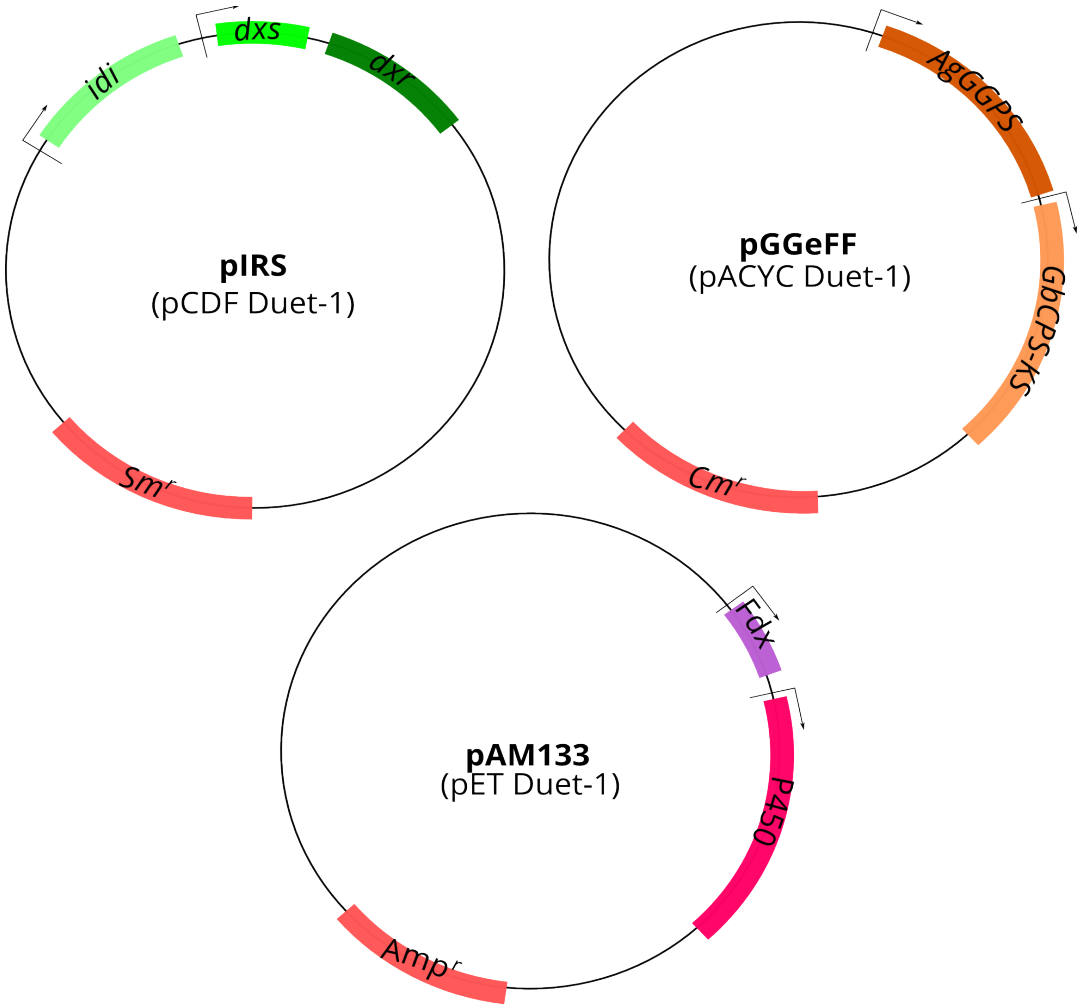


Figure 2.1: Three Co-Expressed Plasmids Used in The Metabolic Engineering System

Briefly, the system uses up to four expression vectors with two multiple cloning sites each. All of the vectors belong to different plasmid incompatibility groups and have different selectable markers. Genes are cloned into the plasmids and co-expressed in *E. coli* C41 cells after induction with IPTG. This figure shows the configuration that we chose to use to test the *B. japonicum* USDA110 CYPs, co-expressed with ferredoxin on plasmid pAM133. The pIRS plasmid encodes enzymes that make IPP and DMAPP, precursors of *ent*-KA and in turn gibberellin. The pGGeFF plasmid has a GGPPS gene from *Abies grandis* (AgGGPPS) and the *F. fujikuroi* CPS/KS gene. When expressed in combination, pIRS and pGGeFF produce *ent*-KA and at least one of the pAM13X series plasmids should be able to convert *ent*-KA to *ent*-KAO, which we planned to detect by GC/MS analysis of hexane extracts.

Section 2.3: Materials and Methods

Table 2.1: List of Bacterial Strains Used in This Study

Strain Name	Genotype	Source
Escherichia coli DH5α	F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsdR17* (rK–, mK+) *pho*A *sup*E44 λ– *thi*-1*gyr*A96 *rel*A1	Hanahan, 1983.
E. coli HiControl 10G	F- *mcrA* Δ(mrr-*hsdRMS*-mcrBC) *endA1 recA1* Φ80d*lacZ*ΔM15 Δ*lacX*74 *ara*D139 Δ(ara, leu)7697 *galU galK rpsL nupG* λ- *tonA*/Mini-F *lacIq*1(Gent^r)	Lucigen.
E. coli BL21 DE3 pLysS	F– *ompT hsdSB*(rB–, mB–) *gal* *dcm* (DE3) pLysS (Cam^R)	Life Technologies.
E. coli C41 DE3	F– *ompT gal dcm hsdSB*(rB-, mB-)(DE3)	Lucigen.
E. coli C43 DE3	F– *ompT gal dcm hsdSB*(rB- mB-)(DE3)	
E. coli C41 DE3 pLysS	F– *ompT gal dcm hsdSB*(rB- mB-) (DE3)pLysS (Cm^r)	
E. coli C43 DE3 pLysS	F– *ompT gal dcm hsdSB*(rB- mB-) (DE3)pLysS (Cm^r)	
Bradyrhizobium japonicum USDA110	Wild type	Kaneko *et al*., 2002
BjUW41	Δ*blr2144*	This work.
BjUW42	Δ*blr2145*	This work.
Sinorhizobium fredii NGR234	Wild type	Pueppke and Broughton, 1999
SfUW1	Rifampicin-resistant derivative.	This work.
SfUW2	Rifr Δ*cpxP*	This work.
SfUW3	Rifr Δ*cpxR*	This work.
SfUW4	Rifr Δ*cpxU*	This work.
Mesorhizobium loti MAFF303099	Wild type	Kaneko *et al*., 2000
MlUW1	Streptomycin-resistant derivative.	This work.
MlUW2	Sm^r, Δ*mlr6367*	This work.

The PCR primers used to construct the expression constructs and deletion mutants are listed in **Table 2.2** and the vectors used in this study are listed in **Table 2.3**. Most of the vectors are commercially available so their maps are easily accessible on the company websites. The exception is pCW Ori+ which we were graciously gifted by Dr. Reuben Peters. We created a new map of the plasmid based on the available reference material and show it in **Figure 2.2**. The PCR primers used to make the expression constructs are listed in **Table 2.2**. The plasmid pRK600 has not been completely sequenced and a map of pK19*mobsacB* can be accessed through the citation. **Figure 2.3** is a work flow diagram outlining the basic details of the planned expression and purification experiments.

Table 2.2: PCR Primers Used to Create Expression And Deletion Mutant Constructs

Name	Primer Sequence (5' to 3')	Use
CYP112 F	GCGGCCCATATGTCCGAACAACAACCCTTGCCGAC	Amplify *blr2144* DNA for cloning into pET30a(+) with a C-terminal His tag using NdeI and XhoI restriction sites
CYP112 R	GATCCTCGAGCCAGAGCACCGGGAACTGCTC	
CYP117 F	GCCGGCCATATGGACATGCTGCTCAACCCGCTGA	Amplify *blr2147* DNA for cloning into pET30a(+)with a C-terminal His tag using NdeI and XhoI restriction sites
CYP117 R	GATCAAAGCTTTGACATCATCACCATCACCACTCATGAGAATCCGATGCGGATAGTC	
CYP112Δ8 F	GATTCATATGGCTCTGTTATTAGCAGTTTTTACGCTGCCGATGTGGCGCGTCGAT	Amplify *blr2144* DNA for cloning into pET30a(+) with a C-terminal His tag using NdeI and XhoI restriction sites; with 8 amino acid truncation from the N-terminal
CYP112Δ33 F	GATTCATATGGCTCTGTTATTAGCAGTTTTTCACCGCGTGCGCTTCCCCTC	As above, with 33 amino acid truncation from the N-terminal
CYP112Δ49 F	GATTCATATGGCTCTGTTATTAGCAGTTTTTTATGACGAGGCCAAGGCGGTGCTG	As above, with 49 amino acid truncation from the N-terminal
CYP112 F1	GATTCATATGTCCGAACAACAACCCTTGCCGA	Amplify *blr2144* DNA for cloning into pET30a(+), pCWOri+ and pETDuet-1 using NdeI and XhoI or SalI restriction sites
CYP112 R1	GATCCTCGAGTCACCAGAGCACCGGGAACTGCTCG	
CYP114 F1	GATTCATATGTTGAGCCGCCACGCCGACATCTTCTGGGCGT	Amplify *blr2145* DNA for cloning into pET30a(+), pCWOri+

Name	Primer Sequence (5' to 3')	Use
CYP114 R1	GATCCTCGAGTCAGCTCCCCTGCGCATGCAGCCGCACCGG CAGCG	and pETDuet-1 using NdeI and XhoI or SalI restriction sites
CYP117 F1	GATTCATATGTTTGGCGGAACGTTGGTCGCCCAGGACGGCA	Amplify *blr2147* DNA for cloning into pET30a(+), pCWOri+ and pETDuet-1 using NdeI and XhoI or SalI restriction sites
CYP117 R1	GATCCTCGAGTCATGAGAATCCGATGCGGATAGTCATGGAC GGAT	
Fdx F	GATCCTCGAGCTAGATTGTGCTGGTCTTTCGCGGC	Amplify the putative ferredoxin ORF DNA for cloning into pET30a(+) and pCWOri+ and pETDuet-1 using NdeI and XhoI or SalI restriction sites
Fdx R	GATTCATATGCGCGTCATGGTCGACCAGGATCTGTGCGGAA	
Fdx F1	AGTATCCATGGATGCGCGTCATGGTCGACCAGGATCTGTGC GGAA	Amplify the putative Fdx ORF for cloning into pETDuet-1 using NcoI and HindIII restriction sites
Fdx R1	TAGTAAGCTTCTAGATTGTGCTGGTCTTTCGCGGC	
Bll4360F	CCTAAGCTTATGGCCACGACCGACACG	PCR positive control resulting in a 1.8 kb amplicon from *B. japonicum* USDA110 genomic DNA
Bll4360R	CGAGGATCCCTATGCGCGAACCCTGAC	
blr2144 F1	GCGGATCCGCACATTCAGCGCGAGAGCAGG	For deletion of *blr2144*; to create the upstream homologous region; BamHI restriction site on F1 primer for cloning
blr2144 R1	CGCGACAGGTGGCACTTCACTCTCTAAG	
blr2144 F2	GAAGTGCCACCTGTCGCGGACGCCACG	For deletion of *blr2144*; to create the downstream homologous region; HindIII restriction site on R2 primer for cloning
blr2144 R2	ACTAAGCTTTGCCGTCCAGGCGCGCTG	
blr2144 CF	TGAAATTTCGCGCAAGGGCTGCCCC	Screening genotype of potential *blr2144* deletion mutants before Southern blotting
blr2144 CR	TCCCCGCGCTCCAGCGCGGGCGTTA	
blr2145 F1	GCGAATTCGTCCCTGGTCGACAAGGCGGTAGAG	For deletion of *blr2145*; to create the upstream homologous region; BamHI restriction site on F1 primer for cloning
blr2145 R1	ACGCGCATC CCTGGTTACCTGTTGGCGGGCTG	
blr2145 F2	GTAACCAGGGATGCGCGTCATGGTCGAC	For deletion of *blr2145*; to create the downstream homologous region; HindIII restriction site on R2 primer for cloning
blr2145 R2	ACTAAGCTTCTGCGCATCGGCGATGTCGAT	
blr2145 CF	AGCGCGAACGCCTGGTCGGCGATCC	Screening genotype of potential *blr2145* deletion mutants before Southern blotting
blr2145 CR	CGTCTCGAACAGCGCTGCCACCGC	

Name	Primer Sequence (5' to 3')	Use
blr2146 F1	GCGAATTCGGGAAGATCATGCTGAGCTTCGG	For deletion of *blr2146*; to create the upstream homologous region; EcoRI restriction site on F1 primer for cloning
blr2146 R1	CGTGTCATCCCGTGTCATCCTAGATTGTGCTGGTC	
blr2146 F2	GGATGACACGTGGACATGCTGCTCAACCCG	For deletion of *blr2146*; to create the downstream homologous region; HindIII restriction site on R2 primer for cloning
blr2146 R2	ACTAAGCTTGATTCCCATGAGACTGAAGATGAGCTTG	
blr2147 F1	GCGAATTCGCGACCAATCTGCGTGGCAC	For deletion of *blr2147*; to create the upstream homologous region; EcoR*I* restriction site on F1 primer for cloning
blr2147 R1	ATGCCAGCCAGCAGCGCGGCGCCG	
blr2147 F2	CGCTGCTGGCTGGCATCGCATGCCCCG	For deletion of *blr2147*; to create the downstream homologous region; HindIII restriction site on R2 primer for cloning
blr2147 R2	ACAAGCTTGATCGTCGTGGACCAACGTACAGGCG	
cpxP F1	ATAAGGATCCCCACGCACAATTCCGCAACGGTCG	For deletion of *cpxP*; to create the upstream homologous region; BamHI restriction site on F1 primer for cloning
cpxP R1	GGCGGCGTCCGCGCAGGTGGCACTCCACTCTCTAAGCGAT	
cpxP F2	CCACCGTGAGGTGGTGTGCGCGGACGCCGCCGGGAATCGC	For deletion of *cpxP*; to create the downstream homologous region; HindIII restriction site on R2 primer for cloning
cpxP R2	TAATAAGCTTCATGCCGTCCAGGCGCGCTGCGACG	
cpxP CF	TTTGAAGCGCAGCGCCCTGTATCGGATGTG	Screening genotype of potential *cpxP* deletion mutants in *S. fredii* NGR234 before Southern blotting
cpxP CR	CCTGTGGCATGCCGAACAGTTCGGCGAAGA	
cpxR F1	ATAAGGATCCCCATCCCTGGTGGACAAGGCGGTGG	For deletion of *cpxR*; to create the upstream homologous region; BamHI restriction site on F1 primer for cloning
cpxR R1	CCTTTTCATGCGGTTCACCAGAGCACCGGGAACTCCTCGA	
cpxR F2	GTGGTCTCGTGGCCCTAACCGCATGAAAAGGGTACAAGGG	For deletion of *cpxR*; to create the downstream homologous region; HindIII restriction site on R2 primer for cloning
cpxR R2	TAATAAGCTTGCCAGCGCATTGCCCGCTTCGGCCG	
cpxR CF	CATGCTGGTGGCAGGACACGAGAGCACCGT	Screening genotype of potential *cpxR* deletion mutants in *S. fredii* NGR234 before Southern blotting
cpxR CR	GCTCGCGTTGTTCACCAGCAGGTCGACCCC	
cpxU F1	ATAAGGATCCCAGACCATGGCGACCAATCTGCGCG	For deletion of *cpxU*; to create the upstream homologous region; BamHI restriction site on F1 primer for cloning
cpxU R1	GCATGCGATGCCAGCCAGCAGCGCGCCGTTTCGGCGGTGT	
cpxU F2	GTCGTCGCGCGGCAAAGCTGGCATCGCATGCCCCGTGTGG	For deletion of *cpxU*; to create the downstream homologous region; HindIII restriction site on R2 primer for cloning
cpxU R2	TAATAAGCTTAGGTCGTCGTGGACCAGGGTACAAG	

Name	Primer Sequence (5' to 3')	Use
cpxU CF	GCACTTCGGTGGGGTCGACCTGCTGGTGAACAACGC	Screening genotype of potential *cpxU* deletion mutants in *S. fredii* NGR234 before Southern blotting
cpxU CR	TGCCTGCAGCGCATCGCCGACCAGGATCGC	
mlr6364 F1	ATAAGGATCCGTCTGCCTCAGGCTAATCCCCTGAA	For deletion of *mlr6364*; to create the upstream homologous region; BamHI restriction site on F1 primer for cloning
*mlr6364 R*1	GGCGGCGTCCGCGCGGGTGGCACTTCACTCTCTAAGCGAT	
mlr6364 F2	CCACCGTGAAGTGAGACGCGCGGACGCCGCCCGGAATCGC	For deletion of *mlr6364*; to create the downstream homologous region; HindIII restriction site on R2 primer for cloning
*mlr6364 R*2	TAATAAGCTTTGCCGTCCAGGCGCGCTGCGACGAT	
mlr6365 F1	ATAAGGATCCATCCCTGGTGGACAAGGCGGTGGAG	For deletion of *mlr6365*; to create the upstream homologous region; BamHI restriction site on F1 primer for cloning
*mlr6365 R*1	ATCACGACGCGCATCCTTGGTTACCTGTTGGCTGGCTGAT	
mlr6365 F2	GAACCAATGGACAACCGAGATGCGCGTCGTGATCGATCAG	For deletion of *mlr6365*; to create the downstream homologous region; HindIII restriction site on R2 primer for cloning
*mlr6365 R*2	TAATAAGCTTACAGCGCTGCCACCGCCTGCGCATC	
mlr6367 F1	ATAAGGATCCGCGACCAATCTGCGCGGCACGCTGC	For deletion of *mlr6367*; to create the upstream homologous region; BamHI restriction site on F1 primer for cloning
mlr6367 R1	GCATGCGATGCCAGCCAGCAGCGCGGCCCCGCGGCCGCTT	
mlr6367 F2	GTCGTCGCGCCGGGGCGGCTGGCATCGCATGCCCCGTGTG	For deletion of *mlr6367*; to create the downstream homologous region; HindIII restriction site on R2 primer for cloning
mlr6367 R2	TAATAAGCTTCGCCGAACTTGACGTGCACGGTCGG	
mlr6367 CF	TGGGGTCGACCTGCTGGTGAACAACGCGAG	Confirm genotype of putative *mlr6367* mutants before Southern blotting
mlr6367 CR	CGCACGAGCGCGATACGGTTGTCGCTCGGC	

Table 2.3 Plasmids Used in This Study

Plasmid	Features	Reference
pRK600	Provides *tra* genes for conjugal transfer of plasmids during bacterial matings; has chloramphenicol resistance; harboured in *E. coli* DH5α	Finan *et al*., 1986.
pGro7	Provides groES and groEL under control of an arabinose inducible promoter; can improve yield of over-expressed protein by increasing presence of folding chaperones	TaKaRa.
pK19*mobSacB*	Integration vector, kanamycin/neomycin resistance, ColE1 replication origin, *mob*, *sacB*, *lacZa* harboured in *E. coli* DH5α	Schäfer *et al*., 1994
pAM101	pK19*mobSacB* with construct for *blr2144* deletion cloned in at BamHI and HindIII sites	This work.
pAM102	pK19*mobSacB* with construct for *blr2145* deletion cloned in at BamHI and HindIII sites	This work.
pAM103	pK19*mobSacB* with construct for *blr2146* deletion cloned in EcoRI and HindIII sites	This work.
pAM104	pK19*mobSacB* with construct for *blr2147* deletion cloned in at EcoRI and HindIII sites	This work.
pAM105	pK19*mobSacB* with construct for *mlr6364* deletion cloned in at BamHI and HindIII sites	This work.
pAM106	pK19*mobSacB* with construct for *mlr6365* deletion cloned in at BamHI and HindIII sites	This work.
pAM107	pK19*mobSacB* with construct for *mlr6367* deletion cloned in at BamHI and *Hind*III sites	This work.
pAM108	pK19*mobSacB* with construct for *cpxP* deletion cloned in at BamHI and HindIII sites	This work.
pAM109	pK19*mobSacB* with construct for *cpxR* deletion cloned in at BamHI and HindIII sites	This work.
pAM110	pK19*mobSacB* with construct for *cpxU* deletion cloned in at BamHI and HindIII sites	This work.
pJET1.2/bunt	Ampicillin resistance, pMB1 origin of replication, blunt end cloning site for cloning PCR products or other small fragments and subsequent sub cloning	Thermo Scientific.
pAM111	*blr2144* blunt-end cloned for sub cloning into pCW Ori+ and pET Duet-1	This work.
pAM112	*blr2145* blunt-end cloned for sub cloning into pCW Ori+ and pET Duet-1	This work.
pAM113	*blr2147* blunt-end cloned for sub cloning into pCW Ori+ and pET Duet-1	This work.
pAM114	*blr2144* blunt-end cloned for sub cloning into pSRK-Km	This work.
pAM115	*blr2145* blunt-end cloned for sub cloning into pSRK-Km	This work.
pAM116	*blr2147* blunt-end cloned for sub cloning into pSRK-Km	This work.

Plasmid	Features	Reference
pCW Ori+	Expression plasmid with *lacI*^{q}, UV5, dual p^{tac} plasmid, multiple cloning site, ampicillin resistance gene, *lacZ*	Fisher *et al*., 1992
pAM117	pCW Ori+ with *blr2144* cloned into NdeI and SalI sites	This work.
pAM118	pCW Ori+ with *blr2145* cloned into NdeI and SalI sites	This work.
pAM119	pCW Ori+ with *blr2147* cloned into NdeI and SalI sites	This work.
pET30a(+)	T7 promoter, option for N or C terminal His tag, kanamycin resistance, pBR322 and f1 origin of replication with *lacI*	Novagen.
pAM120	*blr2144* cloned in at NdeI and XhoI with C-terminal His tag from the vector	This work.
pAM121	*blr2145* cloned in at NdeI and XhoI with C-terminal His tag from the vector	This work.
pAM122	*blr2147* cloned in at NdeI and XhoI with C-terminal His tag from the vector	This work.
pAM123	*B. japonicum* GA operon ferredoxin cloned in at NdeI and XhoI with C-terminal His tag	This work.
pAM124	*blr2144* with C-terminal His tag and 8 amino acid N-terminal truncation cloned into NdeI and XhoI sites	This work.
pAM125	*blr2144* with C-terminal His tag and 16 amino acid N-terminal truncation cloned into Nd*e*I and XhoI sites	This work.
pAM126	*blr2144* with C-terminal His tag and 24 amino acid N-terminal truncation cloned into NdeI and XhoI sites	This work.
pAM127	*blr2144* with C-terminal His tag and 33 amino acid N-terminal truncation cloned into NdeI and XhoI sites	This work.
pAM128	*blr2144* with C-terminal His tag and 49 amino acid N-terminal truncation cloned into NdeI and XhoI sites	This work.
pET Duet-1	T7 promoter, N terminal His tags in first multiple cloning site, C terminal His tag at second MCS, pBR322 and f1 origin of replication; ampicillin resistant	Novagen.
pAM129	*B. japonicum* USDA110 GA operon ferredoxin cloned in at NcoI and HindIII sites	This work.
pAM130	*blr2144* cloned in at NdeI and XhoI sites	This work.
pAM131	*blr2145* cloned in at NdeI and XhoI sites	This work.

Plasmid	Features	Reference
pAM132	*blr2147* cloned in at NdeI and XhoI sites	This work.
pAM133	*blr2144* cloned in at NdeI and XhoI sites with ferredoxin cloned in at NcoI and HindIII sites	This work.
pAM134	*blr2145* cloned in at NdeI and XhoI sites with ferredoxin cloned in at NcoI and HindIII sites	This work.
pAM135	*blr2147* cloned in at NdeI and XhoI sites with ferredoxin cloned in at NcoI and HindIII sites	This work.
pSRK-Km	*LacI*^{q}, kanamycin resistance, *mob*, multiple cloning site, *lacZ* alpha fragment	Khan *et al*., 2008.
pAM136	*blr2144* cloned in at NdeI and XbaI sites	This work.
pAM137	*blr2145* cloned in at NdeI and XbaI sites	This work.
pAM138	*blr2147* cloned in at NdeI and XbaI sites	This work.
pAM139	*blr2144* cloned in at NdeI and XbaI sites with ferredoxin cloned in at BamHI and XhoI sites	This work.
pAM140	*blr2145* cloned in at NdeI and XbaI sites with ferredoxin cloned in at BamHI and XhoI sites	This work.
pAM141	*blr2147* cloned in at NdeI and XbaI sites with ferredoxin cloned in at BamHI and XhoI sites	This work.
pAM142	ferredoxin cloned in at with ferredoxin cloned in at BamHI and XhoI sites	This work.
pET28b(+)/Fdx	pET30a(+) with *fdx* cloned in at unknown sites with a His tag at the N terminus	Peters lab plasmid collection.
pDEST14/sCYP112	pDEST14 with codon optimized *blr2144* cloned in at unknown restriction sites	
pDEST14/sCYP114	pDEST14 with codon optimized *blr2145* cloned in at unknown restriction sites	
pDEST14/sCYP117	pDEST14 with codon optimized *blr2147* cloned in at unknown restriction sites	

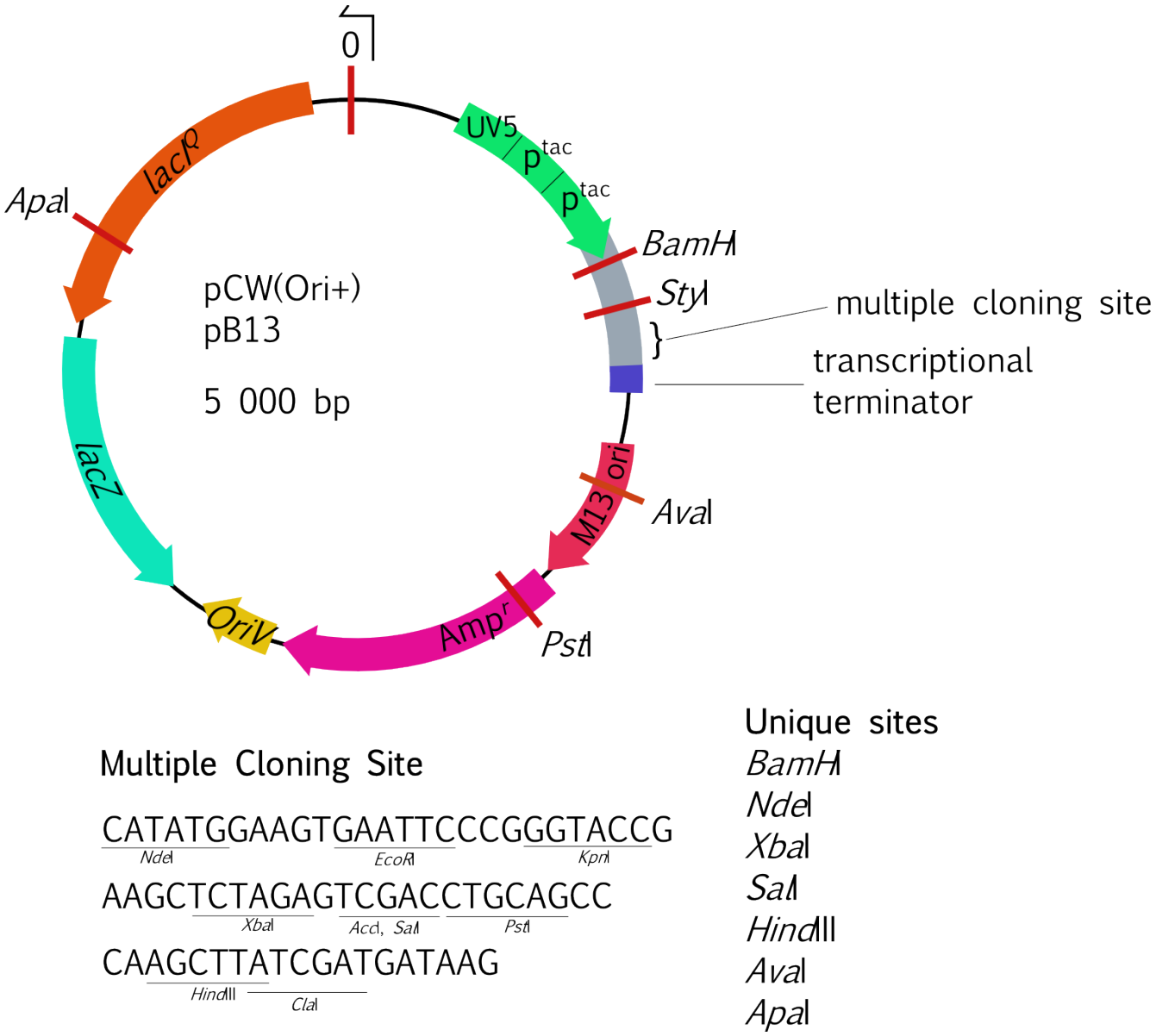


Figure 2.2: The pCW Ori+ Plasmid

The pCW(Ori+) plasmid is an expression vector. It was primarily used for expression of mammalian CYPs during the 1990s. Its primary feature is a double lac promoter (p^{tac}) which is thought to be generally good for expression of CYPs. It is believed that the relatively weaker promoter on pCW(Ori+) is better for producing larger amounts of functional cytochrome P450. Other features of the plasmid include a multiple cloning site with NdeI, EcoRI, KpnI, XbaI, AccI, SalI, PstI, HindIII and ClaI. The marker gene encodes a β-lactamase protein and confers ampicillin resistance.

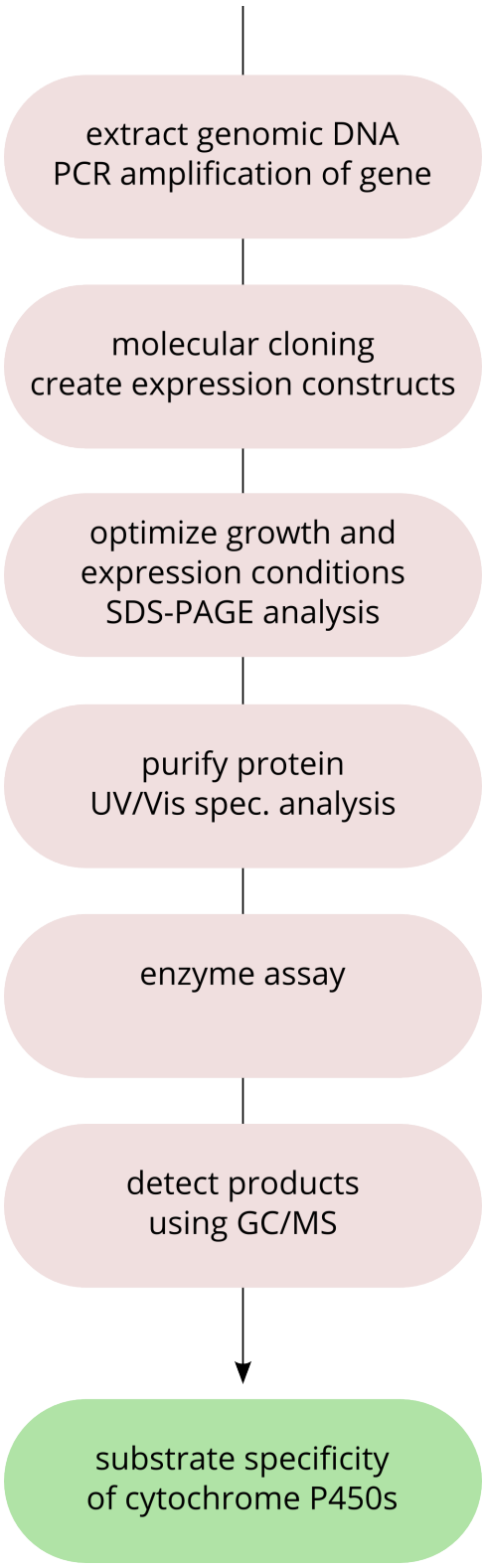


Figure 2.3: Work flow for Expression and Purification of Heterologous CYPs
This diagram shows the main experiments we decided to undertake for this project. The final goal is shown in green. Each pink circle contains a major experiment to reach this goal.

Section 2.4: General Microbiology

All rhizobia cultures were grown at 30 ˚C. *Escherichia coli* cultures were grown at 37 ˚C unless otherwise noted. Cultures were either grown on solid agar in Petri plates, in test tubes, or in shake flasks. Media were made with distilled water in one litre batches unless otherwise noted. Recipes are found in the Appendix at the end of the thesis.

Section 2.4.1: Growth of Rhizobia

The common medium for growth of soil microorganisms is tryptone-yeast extract (TY) broth (Beringer, 1974). It is a low salt, high carbon media with added calcium chloride.. This was the standard media for growth of *S. meliloti* Rm1021, *S. fredii* NGR234 and *M. loti* MAFF303099. Typically, it takes three days for any of these strains to form single, isolated colonies on TY agar plates and at least two days for a single, isolated colony to saturate a 5 mL broth culture. However, *B. japonicum* USDA110 is a slower grower than the other soil organisms used in this study, and takes at least 6 days to form single, isolated colonies on TY agar plates. Therefore, an arabinose-gluconate (AG) medium was used as the predominant medium for growth of *B. japonicum* strains (Sadowsky *et al*., 1987). Using the AG medium, *B. japonicum* generation time is decreased. Both saturated broth cultures and single, isolated colonies arise within three days.

Section 2.4.2: Growth of *Escherichia coli*

Typically, *E. coli* was cultured in a modified version of Lysogeny Broth (Bertani, 1951). For some purposes, such as plasmid DNA extraction or protein expression, a higher cell density than growth in LB could support was required. This was achieved with the use of Terrific Broth which is a richer medium that promotes faster cell growth and division of E. coli(Sambrook and Russell, 2001). Faster replication results in the more rapid accumulation of genetic material. For growth of *E. coli* strains expressing CYPs that require Mg^{2+} ions as a co factor, 1 g of anhydrous $MgSO_4$ was added to the medium prior to autoclaving. Super optimal broth with catabolite repression (SOC) was sometimes used for transformation of ligation reactions during difficult cloning steps. Use of SOC instead of LB to *E. coli* transformation experiments has been reported to result in increased uptake of the plasmid DNA reported to increase transformation efficiency (Hanahan, 1983).

Section 2.4.3: Antibiotic Stock Solutions

Antibiotics were dissolved in either water, 70% ethanol or DMSO depending on the compound. The solvent ethanol was used for tetracycline while DMSO was used for rifampicin. Water was used for all other antibiotics. If the antibiotics were dissolved in water, they were filter-sterilized using a disposable 0.2 µM filter. Typically, antibiotics were prepared, stored at 4 ºC or frozen at -20 ºC for up to one month. Appropriate controls were used to ensure that the antibiotics were functional.

Section 2.4.4: Preparation of $CaCl_2$ Competent *E. coli* cells

Competent cells were prepared according to standardized protocols (Sambrook and Russell, 2001). To make the *E. coli* cells competent (able to take up plasmid or linear DNA through pores during heat shock), cells at early exponential phase were harvested and washed in $MgCl_2$ and $CaCl_2$. For convenience, the cells were left in calcium chloride overnight to increase competency which increases transformation efficiency (Hanahan, 1983) Fifty µL of saturated *E. coli* DH5α broth culture were used to inoculate a 250 ml baffled flask, with 50 mL of LB medium. Several small flasks were used instead of one larger (2 L) flask to keep good aeration in all of the flasks and ensure as close to maximum growth rate of the cultures as possible. We found that using larger flasks resulted in greater lag time which prolonged the culture reaching exponential phase growth. This increased the length of time required to make competent cells by several hours. The flasks were shaken at 220 rpm in a 37 ˚C incubator until the OD_{600} of the cultures reached 0.45 absorbance units. At this point, the cultures were combined into pre-sterilized 250 mL centrifuge bottles and subjected to 10 minutes of centrifugation at 6,000 rpm in a 4 ˚C refrigerated Sorvall RC6+ centrifuge using a F10S-4×1000 rotor with the appropriate adapters. The supernatant was decanted and the pellet resuspended in at least 80 mL of pre-sterilized, pre-chilled 0.1 M $MgCl_2$.

All cell clumps were carefully resuspended and subjected to another round of centrifugation with the same conditions. The supernatant was decanted and the cells were resuspended in 0.1 M $CaCl_2$. The suspension was held in an ice-water bath overnight. The following morning, the cells were collected by centrifugation under the same conditions; this time the pellet was resuspended in 5 mL of sterile, ice-cold 0.1 M $CaCl_2$ with 15% glycerol to preserve the cells during storage. Aliquots of 200 µL volumes were prepared and stored in the -80 ˚C freezer for future use. For quality control and to determine the competency of the batch, heat-shock transformation experiments using 10 pg of supercoiled pUC19 DNA and the appropriate negative transformation controls were done.

Section 2.4.5: Transformation of *Escherichia coli*

E. coli transformation was pioneered in the early 1980s and the technique revolutionized molecular biology (Hanahan, 1983). The protocol used in the Charles Laboratory is based on the method originally developed by Douglas Hanahan (1983). This is my personal protocol and reflects my personal experiences. So while it may differ from other protocols, it achieved the same results using this method. This technique allows virtually free and limitless propagation of DNA *i*n a biological system that is easy to grow, handle and manage. *E. coli* strains can be preserved for decades in freezers, allowing research labs to build up an ever-expanding backup catalog of their work.

Frozen *E. coli* culture can be revived overnight and storage of cells rather than naked DNA is much more robust making loss of important DNA or plasmids unlikely if proper care is taken to maintain the frozen stocks. *E. coli* strains are the most common vehicle for replication of DNA in a laboratory because their generation time is short and their growth requirements are minimal and inexpensive. *E. coli* cultures are cheap, quick and easy to grow in the laboratory. One process for getting DNA into an *E. coli* cell is described below.

A 200 µL aliquot of previously-prepared competent *E. coli* cells was removed from the -80 °C and thawed on ice. One hundred microlitres was transferred to a new sterile microtube. DNA previously prepared for this purpose, typically either a ligation reaction or plasmid DNA from another source, was added up to a volume of 10% of the total aliquot volume. The mixture was left standing on ice for at least 30 minutes. This step is very important because it allows the DNA to adhere to the wall of the bacterial cell. Only DNA that has adhered to the *E. coli* before the next step will be able to enter the cell. Typically, one single DNA molecule enters on *E. coli* cell.

The cell-DNA suspension was then heated to 42 °C for 66 seconds. Heating the cells opens pores in the outer membrane; the DNA enters the cell through these pores. After heat shock, the cells

were returned to ice and chilled for 5 minutes. This step allows the cells to recover after heat shocking. After 5 minutes elapsed, 700 µL of LB or SOC was added to the cell suspensions. The cells were then incubated at 37 °C with shaking for at least one hour before being plated on selective LB agar. Positive clones were streak-purified and then grown in LB broth containing the appropriate antibiotic. The plasmid DNA was isolated and analyzed by restriction mapping and the fragments were subjected to gel electrophoresis. Only colonies from clones that had the desired restriction map were selected for further work. At this point, desired positive *E. coli* strains were saved.

Section 2.4.6: Electroporation of *E. coli*

This protocol is based on standard procedures for molecular biology (Sambrook and Russell, 2001). Ten ml of fresh overnight *E. coli* cultured in non-selective LB broth were spun down in a 4 °C bench centrifuge using a swinging bucket rotor. The samples were spun at 4,000 rpm for 8 minutes. The supernatant was discarded and the pellet was resuspended in 10 mL of 10% ice-cold glycerol. The cells were centrifuged again, the supernatant discarded and the pellet resuspended four times. After the final centrifugation step, the pellets were resuspended in a total of 1 mL of 10% glycerol. The 1 mL suspensions were then partitioned into 100 µL aliquots and immediately chilled on ice until used. At least 75 nanograms of plasmid DNA in one or two µl of 2mM Tris buffer (pH 8.0) was added to a 100 µL aliquot of cells. The cell-DNA suspension was then transferred to a pre-chilled disposable plastic 1 mM cuvette. The metal contacts on the cuvette were wiped dry with a Kim Wipe and then placed in the electroporator (Bio Rad Micro Pulser). The machine was used on setting Ec1. After the pulse, 1 mL of super optimal broth rescue medium, (recipe in Appendix) was immediately added to the cuvette. The entire suspension was then transferred back to the microtube and kept on ice until all samples were finished. The cells were then incubated with shaking at 37 °C for one hour. After incubation the suspension was subjected to centrifugation at 13, 000 rpm using a KL 064 rotor in an Eppendorf 5430 centrifuge. The cell pellet was then resuspended in 100 µL of super optimal broth and then plated on

LB agar containing kanamycin at a final concentration of 30 μg mL^{-1}. The plates were incubated overnight at 37 °C and colonies streak purified. The plasmids were then isolated and restriction mapping was done to confirm the identity of the expression constructs. Strains were saved in the main Charles Laboratory strain collection.

Section 2.4.7: Storage of Strains in The Permanent Collection

Cryopreservation of bacterial strains is crucial to bacterial molecular genetics. The protocol used in the Charles lab is adapted from (Sambrook and Russell, 2001).To save bacterial strains in the permanent collection, 250 μL broth culture aliquots were deposited in sterile cryovials. Depending on the bacterial species, a mixture of 14 % DMSO and LB, TY or AG broth in a total volume of 250 μL was added to the cryovial. Strains were described and annotated in the main strain collection document and then stored in the -80°C freezer.

Section 2.5: Genomic DNA Extraction Protocol

This protocol is adapted from standard protocols (Meade *et al*., 1982; Oresnik, Charles and Finan, 1994).Cells from 5 mL of saturated broth culture were collected by centrifugation at 6,000 rpm for 5 min using an F-35-6-30 rotor in an Eppendorf 5430 centrifuge. The supernatant was discarded and the cells were resuspended in 500 μL of Tris-ethylenediaminetetraacetic acid (Tris-EDTA) buffer concentrated at 10 mM Tris and 1 mM EDTA. Twenty-five μL of 25% sodium dodecyl sulphate (SDS), 62.5 μL of 5 M NaCl and final concentrations of 10 mg/mL of proteinase K and RNase A were added. The solution was incubated at 65 °C until sufficiently clear, which usually required heating for three hours. Occasionally, the solutions were moved to 4 °C overnight for further lysis and re-heated to 65 °C the following day. Next, 600 μL of phenol were added to each sample and mixed using the vortex mixer before centrifugation. Centrifugation for approximately 30 minutes at 13,000 rpm using a KL 064 rotor in the same centrifuge resulted in the best DNA yields. Following the initial extraction, the samples were extracted twice with a 50:50 mix of phenol:chloroform and then twice more with an

equal volume of chloroform. Three M sodium acetate (pH 5.8) or 3 M potassium acetate (pH 4.2) was added at $1/10^{th}$ of the sample volume. Salt such as sodium or potassium acetate is added to dehydrate the DNA in aqueous solution so that when two sample volumes of 95% ethanol are added, the DNA precipitates. The precipitated DNA samples were left at -20 °C for at least 1 hr up to overnight. Precipitated DNA was centrifuged for 5 min at 13,000 rpm in the KL 064 rotor. DNA pellets were washed with at least 100 µL 70% EtOH and then resuspended in 200 to 500 µL of 2 mM Tris (pH 8.0) depending on the size of the pellet. Samples were stored long term at -20 °C or at 4 °C if work was immediate.

Section 2.5.1: Plasmid DNA Extraction Protocol

This protocol is based on the standard technique for isolation of plasmid DNA from *E. coli* cells (Sambrook and Russell, 2001). It is carried out at room temperature. The recipes for the solutions and buffers are found in the Appendix. To collect cells, one to one and a half µL of saturated *E. coli* DH5α culture in LB broth was centrifuged at 13,000 rpm in a DESAGA micro centrifuge for one minute. Supernatant was discarded and the cell pellets were resuspended in 250 µL of Solution I. This solution contains EDTA to chelate divalent ions that may be enzyme co-factors for DNAses (Sambrook and Russell, 2001) The RNAse A degrades the mRNA that would otherwise contaminate the final plasmid DNA sample (Sambrook and Russell, 2001). Addition of 250 µL of Solution II lyses the cells because of its strongly basic pH and detergent concentration. As the cells lyse, the solution slowly clears. The contents of the cell are released into the tube. Addition of 350 µL of Solution III causes the proteins to precipitate but leaves the DNA in solution. It is advised to add Solution III no more than five minutes after addition of Solution II because prolonged exposure to strongly basic pH may damage the plasmid DNA.

The samples were centrifuged at 13,000 rpm for 10 minutes in either the DESAGA or the Eppendorf (KL 064 rotor) micro centrifuge to collect the precipitated protein and cell debris at the bottom of the tube. Eight hundred µL of the DNA-containing supernatant was removed after centrifugation and applied to a micro centrifuge silica column (Bio Basic Catalog #SD5005). The loaded columns were left standing at room temperature for up to five minutes, which is believed to increase binding of the DNA to the column. After incubation, the solution was moved through the column by centrifugation at 10,000 rpm (DESAGA or Eppendorf) for one minute. The DNA remains bound to the silica column.

The column was washed twice with 500 µL of Wash Buffer to remove excess salt and SDS used up until this point. Wash Buffer was passed over the column using the previous centrifuge conditions.

The columns were spun for an additional two minutes to rid the column of supernatant. This is done to eliminate any residual ethanol which is a key ingredient of the Wash Buffer but can interfere with downstream applications that the plasmid DNA may be used for. In some cases, it was necessary to extract plasmid DNA from *E. coli* strains other than DH5α, such as BL21 DE3, C41 and C43 and others. In these situations an extra step was done using PB Wash to denature DNAses and prevent degradation of the plasmid DNA. *E. coli* DH5α is a widely used cloning strain and one of the reasons it is so popular is because it has the genes encoding its DNAses knocked out. The other *E. coli* strains mentioned above do not have these mutations so when the cells are lysed, there is a risk that the DNA could be degraded. The 5 M guanidine thiocyanate in PB Wash neutralizes the threat of the DNAses and makes plasmid isolation from these non-cloning *E. coli* strains possible.

The dried column was moved to a new micro centrifuge tube and 50 µL of elution buffer was added. After an equilibration period of one to two m, the tubes were centrifuged at 10,000 rpm for approximately 45 s to elute the DNA from the silica. A NanoDrop machine (Fisher Scientific) was used to quantify the DNA. The A_{260}/A_{280} and A_{260}/A_{230} ratios, as well as the sample concentration in ng μL^{-1} were recorded. The absorbance spectrum of the sample was also observed. Plasmid DNA was stored at -20 °C and thawed as necessary.

Section 2.5.2: Preparation of Lambda DNA Standards for Gel Electrophoresis

We found quantification of genomic DNA by NanoDrop to be unreliable for our purposes. When it was necessary to quantify the amount of genomic DNA, it was done by comparing the brightness of a particular sample with that of a known amount of bacteriophage Lambda DNA using gel electrophoresis. To prepare the Lambda DNA (Fisher product #FERSD0011) standards, it was was diluted in MilliQ distilled water and 6x loading dye (Fisher product #FERR1151). Standards with total amounts of Lambda DNA corresponding to 5, 10, 25, 50, 75, 100 and 200 nanograms per 12 µL were made and stored at 4 °C. Before use, they were heated for 2 m at 65 °C. For comparison with genomic

DNA samples, 12 μL of each standard was loaded onto an agarose gel and subjected to electrophoresis. After the gel was run, the brightness of the standards was compared to the sample to determine the DNA concentration of the sample. Comparisons of relative brightness could be made by eye using a UV transilluminator, or more exact quantification of the sample could be done using gel imaging software. The procedure for running a gel and taking pictures of it are described in the next section.

Section 2.5.3: Agarose Gel Electrophoresis

Following standard procedures, our protocol for gel electrophoresis was developed (Sambrook and Russell, 2001). Gel electrophoresis was usually done using 1% agarose gels and 1X Tris-acetate EDTA (TAE) buffer (recipe in Appendix). Gels were pre-stained during casting with a volume of GelRed, an alternative to ethidium bromide. The amount of GelRed used was equivalent to a 1:10,000 dilution of the lab stock, already diluted 5x from the commercial solution. To ensure good migration and separation of the DNA, the applied voltage was calculated based on the distance between the two electrodes for any particular gel rig. The distance in centimeters was measured and then multiplied by 7 (John Heil, personal communication). The resultant product was used as the voltage setting. Typically, gels were run for 45 minutes to one hour at the optimal voltage before analysis. Images of gels were captured using a digital camera focused on a UV-transilluminator emitting UV light at 365 nanometers. The software PSRemote (available from http://www.breezesys.com/PSRemote/) was used to take photos and ImageJ (available from http://imagej.nih.gov/ij/index.html) was used to crop and edit the picture files. Images were printed with a Mitsubishi P93 D thermal printer and the files were stored on the lab computer and also in a personal Dropbox account. If the DNA in the gel was destined for gel extraction, the gels were not imaged until after gel extraction was done.

To quantify purified, undiluted genomic DNA, loading dye was mixed in a 1:1 ratio with the DNA sample, and typically a total of 2 µL was loaded per lane. One µL of 1/10 dilutions of the genomic DNA was also run alongside undiluted samples. In these cases, 1 µL of 6x dye was added to 8 µL of sterile, double distilled MilliQ water along with 1 µL of sample genomic DNA and mixed by aspiration with a pipette in a microtitre plate or on a piece of lab parafilm before loading. For extraction of DNA fragments, gel electrophoresis was carried out as normal except guanosine was added to the gel at a concentration of 0.285 grams per litre. Guanosine was added to absorb UV light that was used to expose the gel prior to cutting out the DNA band (John Heil, personal communication). UV light may cause mutations or thymine dimers to form in DNA and can have negative effects when the sample is used for downstream applications such as cloning.

Section 2.5.4: The Polymerase Chain Reaction

This protocol for PCR is based on standard techniques (Sambrook and Russell, 2001). Polymerase chain reactions were done in 50 µL volumes composed of 25 µL 2 x KOD Xtreme HotStart DNA polymerase buffer, 10 µL 2 mM dNTPs, 10 mM of each primer, 2 µL of DNA template, 1 µL of KOD Xtreme HotStart DNA polymerase, 5 % DMSO, and 6.5 µL of sterile double distilled water. The reactions were completed in a Bio Rad MyCycler thermal cycler with the following program. All PCRs were done using this program unless otherwise indicated. KOD Xtreme kits were obtained from VWR (Catalog #71975-3). The PCR began with a 2 minute hold at 94 °C to activate the polymerase and denature the DNA template. A touchdown cycle with a 15 second, 94 °C denaturation period followed by a 30 second, 65 °C annealing time and a 90 second, 68 °C elongation period was repeated nine times. In each round of the touchdown cycle the annealing temperature was decreased by 1 °C. Following the touchdown stage, 25 cycles were done, each with a 15 s, 94 °C denaturation period followed by a 30 s, 55 °C annealing period to allow the primers to hybridize with the template. A 90 s extension time at 68 °C was used.

After all 25 cycles were complete, a final 5 min hold at 68 °C was done according to the manufacturer's instructions. After this, the reactions were held infinitely at 20 °C until the machine was switched off. PCR products were typically stored overnight at 4 °C and used the next day.

Section 2.5.5: Purification of PCR Products

To purify amplicons from primers, dNTPs, polymerase and other components of the PCR, 150 µL of Binding Buffer II (recipe in the Appendix) were added to the PCR tube. The reaction was heated for approximately 10 minutes at 65 °C and then the total volume was loaded onto a microcentrifuge size pre-packed silica column (Bio Basic Catalog # SD5005). The tubes were spun at 10,000 rpm for 1 m in an Eppendorf 5430 centrifuge (KL 064 rotor) to filter liquid through the column. The supernatant was discarded and the column was washed twice with 500 µL Wash Buffer. The DNA was eluted using 50 µL of Elution Buffer.

Section 2.5.6: General Restriction Digestion

Routine restriction digestion was typically done in 20 µL volumes. Depending on the DNA concentration of the particular prep, the volume of water and template DNA varied. The reaction mixture contained sterile, double distilled (MilliQ) water, 2 µL 10 X Fast Digest Green Buffer, template DNA and 3 units of each restriction enzyme. Whenever possible, master mixes were used. All ingredients were aspirated to mix and then the reaction mixture was spun down to remove any air bubbles before incubation at 37 °C, with the exception of SmaI and its isoschizomer Cfr91 which were incubated at room temperature. Digestion typically lasted for up to an hour before the reactions products were subjected to gel electrophoresis. To do this, the reaction mixutres were loaded directly onto DNA gels because Fast Digest Green Buffer contains loading dye.

Section 2.5.7: Digestion, Gel Extraction and Cloning

For molecular cloning, the plasmid (vector backbone) that the DNA would be cloned into was often gel purified after digestion with restriction enzymes. This protocol is based on standard

techniques in molecular biology (Sambrook and Russell, 2001). To maximize yield, larger scale digestions were performed and typically consisted of 11 µL sterile double distilled water, 6 µL of 10 X FastDigest Green Buffer, 40 µL DNA, 1 µL (10 units) of FastDigest restriction enzyme 1, 10 units of FastDigest restriction enzyme 2, and 1 µL (10 units) of shrimp alkaline phosphatase (Fermentas #EF0511) or Fast Alkaline Phosphatase (Fermentas # EF0651). The reaction was set up at room-temperature, and initially only 1 µL of one restriction enzyme was added. After incubation at 37 °C for approximately thirty minutes, one microliter of the other enzyme was added and the reactions were incubated for an additional half-hour. Finally, one microlitre of shrimp or fast alkaline phosphatase was added and the digest incubated for another thirty minutes. The reaction mixtures were then loaded onto agarose gels for gel extraction, because the reaction buffer contained loading dye. This was done so that the reaction products could be analyzed. One percent agarose gels made of 1X TAE were supplemented with 0.285 g^{-1} guanosine as described in Section 2.5.3. The standard comb size in the Charles Lab is 15 mm, but a 30 mm comb size was used for this purpose so that larger samples could be loaded. Thirty µL of the reaction mixture was loaded in each lane of the gel. Bands of the desired size were excised using a razor blade and placed into pre-weighed microcentrifuge tubes. The difference in mass of the microcentrifuge tube after the addition of the gel slice was calculated. The volume of Binding Buffer II added to solubilize the gel and bind the DNA was calculated. The calculation is from the instructions that we obtained from a Qiagen kit that we previously used for gel extractions. Though we no longer use the kit and now make our own equivalent Binding Buffer, we have had good success using this method and forumla. This protocol will vary from laboratory to laboratory.

mass of gel slice x 3×1000 = µL of buffer added to gel fragment

Lid locks were added to the tubes and all samples were heated using the 65°C heat block to

melt the gel and release the DNA fragments. The procedure described in Section 2.5.5 to clean up the PCR reaction was repeated. The concentration of the DNA sample after it was isolated from the gel was determined using the NanoDrop 1000 (Thermo Fisher). If necessary, the volume of the DNA samples was reduced by placing them in a Speed Vac centrifuge set to medium heat for 5 to 10 minutes. This evaporated some of the Elution Buffer and concentrated the DNA sample.. An effort was made to prevent the samples from becoming completely dried. After the spin in the Speed Vac, the volume of the sample was empirically calculated by aspirating it with a pipette. The liquid and tip was carefully watched, so that the entire volume was taken into the microtip without any extra air space. It was an empirical measurement because it required trial and error—changing the volume of the plunger—to accommodate the volume of the sample until it was set perfectly. Once the entire volume of the sample could be aspirated into the microtip without any extra air space, we knew the exact volume of the sample since it corresponded to the setting of the pipette. The DNA concentration was determined using the previously recorded Nanodrop values, the previous volume of 50 µL and the new sample volume, determined by empirical pipetting.

Ligations were done in 20 µL volumes with 2 µL of 10 x Ligation Buffer (Fermentas #B69) and 0.5 µL of T4 DNA ligase enzyme (Fermentas #EL0014). The volumes of water, vector and insert were determined empirically based on the concentration of the vector and insert DNA. Typically, 5 µL of vector and 8 µL of insert were added, with the remaining volume made up with MilliQ water. Ligations were left overnight at 16 °C and their products were introduced into $CaCl_2$ competent *E. coli* DH5α cells by transformation (see **Section 2.4.5**). Transformants were isolated on selective LB agar.

Section 2.5.8: DNA Sequencing

Plasmid mini-prep samples with concentrations of 20 ng/µL were sequenced by The Centre for Applied Genomics in Toronto, ON.

Section 2.6: Analysis of Heterologously Expressed Protein

One mL samples of *E. coli* expression cultures were centrifuged and the supernatant was discarded. The cell pellets were stored at -20 °C until analysis, at which time they were thawed on ice.

Section 2.6.1: SDS-PAGE Sample Preparation

Two alternative methods to lyse the cells were used; both protocols are based on standard SDS-PAGE (SDS-PAGE) sample preparation techniques and we did not notice a difference in quality of the samples using either technique. In the first method, 70 µL of sterile MilliQ water was used to resuspend the cell pellets. In the fume hood, 14 µL of 2-mercaptoethanol (BioShop #MER002) was added to each solution. Lid locks fitted on all microcentrifuge tubes and the samples were boiled for 5 to 10 minutes to lyse the cells. This method was the preferred method because it is faster than sonication—which uses high-frequency sound waves to break open the cells. Alternatively, the cell pellet was resuspended in 500 µL of Lysis Buffer (Appendix) and subjected to sonication. Bursts to lyse the cells were for 15 to 20 seconds followed by an equivalent rest period in ice-water. During bursts, the tubes were kept in an ice-water bath so that they stayed cool. Although sonication is a great way to lyse cells, it can cause quite a lot of heat. Normally, sonication is done when the heterologous protein will be purified; using it to prepare samples for SDS-PAGE is somewhat tedious, since SDS-PAGE samples cannot be recovered.

Despite whether the cell pellets were boiled or subjected to sonication for the initial lysis step, the samples were spun down in an Eppendorf 5430 centrifuge with a KL 064 rotor for 8 minutes at 12,000 rpm. The supernatant was aspirated and moved to a new tube and designated the supernatant fraction. This sample is composed of the soluble proteins. The pellet contains cell debris as well as the insoluble proteins, and was designated the pellet fraction. We differentiated between pellet and supernatant fraction because we had an issue with the solubility of CYP112—that is to say, we found that CYP112 was largely insoluble in our expression system Twenty-five µL of 5 X SDS-PAGE sample

buffer (Appendix) was added to the supernatant fraction. The suspension was mixed by aspiration and heated briefly at 65 °C. The pellet fractions were resuspended in 100 µL of 1 X SDS-PAGE running buffer (Appendix) and 50 µL of 5 X sample buffer. These samples were boiled once again for an additional 5 to 10 minutes to solubilize the some of the previously insoluble proteins. After heating both sets of samples, they were all subjected to a quick spin in the centrifuge to collect the debris at the bottom of the tube. This was done to ensure that protein clumps and debris were not loaded into the SDS-PAGE gel wells because this can cause problems with the resolution of the gel due to overloaded wells. After spinning down the samples, both fractions were ready for SDS-PAGE analysis. The samples could be frozen at -20 °C and thawed at 65 °C if they had to be re-run. If multiple gels using the same samples were run in a day, the samples were kept at 4 °C.

Typically, 4 to 4.5 µL of each sample were used for analysis. Eight percent SDS-PAGE gels were cast using either 0.75 or 1.0 mM glass plates using the Mini-Protean III casting system (recipe in Appendix). The gels were were run at 200 V for 65 minutes in 1 X SDS-PAGE running buffer using a BioRad PowerPac 3000 power source and Mini-Protean II gel rig (serial number 125B5 30320) according to standard protocols (Laemmlli, 1970).If His-tag staining was required, it was done prior to Coomassie staining (Appendix). Coomassie staining was either done directly to the gels or after His-tag staining; Coomassie stain interferes with His-tag stain, so His-tag staining must precede Coomassie staining. For a review on Coomassie staining, please see Brunelle and Green, 2014. Pre-stained and unstained protein markers from various vendors were used used to size protein bands. For more accurate protein sizing, unstained markers were used.

Section 2.6.2: His-tag Staining of SDS-PAGE Gels

The protocol shown below is an adaptation of the method recommended by the manufacturer. If the gel was to be His-tag stained, it was rinsed with distilled water and moved to an empty pipette tip box. We found that this container was best for staining because it has high sides to prevent spills, the

plastic can withstand microwaving, and the dimensions of the box allow the gel to move when submerged in stain, but well-fitted so that large volumes of solutions are not required.

In the pipette tip box, the gel was immersed in an ample volume of fixative solution (Appendix) and microwaved on high for 25 seconds, following the manufacturer's recommendations for His-tag staining. The gel was then moved to a rotary shaker set at low speed for at least thirty minutes. The gel was rinsed thrice with MilliQ water and then left in MilliQ water for 5 to 10 minutes with shaking. The water was decanted and approximately 20 mL of His-tag stain was applied to the gel. The gel was microwaved with the stain for 25 s and then gently agitated at room-temperature for one hour. After one hour, the stain was poured off and the gel was immersed in ample 20 mM phosphate buffer for destaining. The container was shaken at room-temperature on an orbital shaker. After five minutes the destain was decanted and fresh buffer was added. The gel was immediately imaged using a transilluminator system (AlphaImager). The His-tag stain was detected using a UV-transilluminator set to the 365 nm wavelength. Gels were placed directly on the box and exposed for up to thirty seconds, depending on the empirical conditions of the experiment. Typically, the aperture of the camera was set between 1.5 and 2.0 mm, the focus of the lens was 20.0 and the zoom was between 14 and 15. Larger aperture sizes let in more light so a shorter exposure time is needed. Images were captured using AlphaImager software (Available from http://www.proteinsimple.com/alphaimager_hp.html). After applying the His-tag in gel stain, a Coomassie stain was done to visualize the total protein profile of the samples (Appendix).If His-tag staining was not done, Coomassie staining was done directly after the gel was finished running.

Section 2.6.3: Coomassie Staining of SDS-PAGE Gels

Coomassie stain is used to visualize the size marker (if it was not prestained) and all protein samples after electrophoresis (Brunelle and Green, 2014). A volume of Coomassie stain sufficient to completely cover the gel was added to the container holding the gel. The gel and stain were

microwaved on high for 45 seconds. Steam and vapour were allowed to diffuse from the container in the fume hood. The gel was then shaken in the Coomassie stain for at least 10 min on a bench top rotary shaker set to low speed. The stain was decanted and reused two to three times without a loss in potency of the dye. Stained gels were rinsed with distilled water to rinse off excess stain and then destained with Coomassie Destain Solution (Appendix).The gel immersed in Destain was also microwaved to quicken the process. Knotted Kim Wipes were added to the container to absorb excess dye. After a room temperature incubation of approximately 10 min on a low-speed rotary shaker, the destain solution was refreshed. Sometimes, additional destaining was done overnight with MilliQ water. A white-light light box was used to immediately visualize protein bands. Photographs of the gel were recorded using a UV to white light adapter. Images were captured using a digital camera and PSRemote software.

Section 2.7: Optimized Conditions for Expression of CYP112 from pET30a(+)

Small scale expression trials were conducted to determine the optimal expression conditions for the full-length CYP constructs. Media type, growth and induction temperature, induction time and isopropyl β-D-1-thiogalactopyranoside (IPTG) concentration were optimized to achieve the highest expression of the protein possible, based on SDS-PAGE analysis of crude extracts. The optimal expression conditions for expression of CYP112 from the pET30a(+) vector with a C-terminal His-tag used one litre of Terrific broth (TB) supplemented with 34 μg mL^{-1} chloramphenicol and 35 μg mL^{-1} kanamycin. The culture was grown in a 6 l flask to maximize aeration of the culture during the growth phase. The TB was inoculated with 10 mL of overnight *E. coli* BL21.DE3 pLysS (pET30a(+)) culture and incubated at 30 °C on an orbital shaker set to 200 rpm. The culture was incubated until an OD_{600} of 0.6 to 0.8 was reached, typically after 3 to 4 hours o growth. At this point, the culture was moved to a room-temperature shaker set to 90 rpm and 1.0 mM of 5-amino-levulinic acid (ALA) and 0.5 mM $FeCl_3 \bullet 6\ H_2O$ was added (Fisher *et al*., 1994). ALA and iron chloride are precursor compounds for

heme, a key component of CYPs (Fisher *et al*., 1994). Addition of heme precursors promotes heme biosynthesis which can be a limiting compound in heterologous CYP expression . Addition of ALA and iron chloride prior to the induction of CYP expression (using IPTG) is thought to promote folding and functional capacity of the CYPs (Fisher *et al*., 1994).

The culture was agitated for an additional 20 min to allow heme biosynthesis to occur then 1 mM of IPTG (BioShop Catalog #IPT001) was added to induce expression of the CYP. Incubation and protein expression continued for up to 48 h during optimization, but it was found that 24 h yielded high amounts of CYP112 with minimal degradation. One mL samples of the culture were taken prior to IPTG induction and at 2, 6, 18, 24 and 48 h post-induction in order to monitor protein expression and find an optimal time to harvest the culture. Five *E. coli* expression strains were tested, including *E. coli* C41 and C43 with and without the pLysS plasmid. Co-expression of the CYPs with pGro7 was also attempted in *E. coli* C41 and C43. Expression cultures with the pCW (Ori+) constructs harboured in *E. coli* DH5α were also tried. The best expression was achieved using *E. coli* BL21 DE3 pLysS, but expression using *E. coli* C41 and *E. coli* C43 was nearly equivalent for the pET30a(+) constructs. Unfortunately, CYP112 was largely insoluble and yields of purified protein were low.

Section 2.7.1: Expression Conditions for Codon Optimized CYPs and Ferredoxin

Our collaborators at Iowa State University (ISU) optimized the codon use of the *B. japonicum* USDA110 CYP and ferredoxin gene for expression in *E. coli* (Reuben Peters, personal communication). During my research trip to Ames, Iowa I was granted access to the constructs. Unfortunately, the lab had not kept a record of the optimized sequence. We attempted to derive it by sequencing the constructs from the pUC57 cloning vector that they were provided in from the synthesis company as well as the pDEST-14 expression vector that lab personnel had cloned them into (Appendix). We were successfully obtained the sequence of the CYP112 and CYP114 genes but were unable to elucidate the sequence of the CYP117 and ferredoxin gene.

Despite not knowing the genetic sequence of the constructs, we were assured that they were correct. I was encouraged to use these constructs rather than my own to perform experiments at ISU. Unfortunately, our collaborators were also challenged by a lack of familiarity with microbiology and bacteriology and they had not optimized the expression conditions or expression protocols for these constructs prior to the visit. Although the constructs had apparently been used for several years by previous lab members, records were not kept and old protocols were not available to me. I set to work attempting to optimize the expression conditions for these constructs, but I was not able to complete the experiments to an appropriate level of detail. The work presented below is a good start, but additional effort must be made to find the best expression conditions for these proteins.

Two mL LB broth cultures supplemented with the appropriate antibiotic(s) were inoculated from fresh streak plates of *E. coli* C41 harbouring the expression constructs. Kanamycin was used at a working concentration of 25 µg/mL, carbenicillin at 50 µg/mL and chloramphenicol at 34 µg/mL. Seed cultures were grown overnight at 37 °C with shaking at 220 rpm. The following morning, the seed cultures were used to inoculate modified TB (mTB) at 1/100th of the expression medium volume

(recipe in Appendix). Typically, large scale expression was carried out in 2.8 L baffled Fernbach flasks filled with 450 mL of mTB while small scale expression was done in unbaffled, 250 mL flasks filled with 45 mL of broth. After autoclaving the medium, sterile phosphate buffer (pH 7.5) was added at 1/10th of the mTB volume. Antibiotics were added at the same concentrations used to select for the expression plasmids in the seed cultures. The expression cultures were grown for 4.5 hours at 250 rpm in a 37 °C incubator. After 4.5 h had elapsed, 0.5 mM ALA and 0.5 mM $FeCl_3$•6 H_2O was added to the broth, as was done in Section 2.7.0.The cultures were incubated for 20 min to allow synthesis of heme prior to induction of the expression plasmid by the addition of 1 mM IPTG. When IPTG was added, the cultures were moved to a 16 °C incubator. The cultures were shaken at 200 rpm for an additional 18 to 20 hours. The cells were harvested by centrifugation and the pellets stored at -80 °C..

Section 2.7.2: Cloning and Expression of Ferredoxin

The ferredoxin gene (locus *blr2145.1*) was amplified from B. japonicum USDA110 genomic DNA. The PCR primers were amended with NdeI and XhoI restriction endonuclease recognition sites for cloning into the pET30a(+) expression vector. The same cloning procedure used to clone the CYP genes was used to clone the ferredoxin gene. The protocols were modified from the manufacturer's instructions and standard molecular biology techniques (Sambrook and Russell, 2001). Sanger sequencing of the insert in pET30a(+) verified the construct. The plasmid was expressed in the same five *E. coli* strains as the CYP constructs but ultimately *E. coli* C41 was chosen as the expression strain since *E. coli* C41 was also used to express the CYPs at ISU.

Expression cultures of ferredoxin were grown in parallel with expression cultures of the codon-optimized CYPs (Section 2.7.2) and wild type *E. coli* C41 at 37 °C with 220 rpm shaking. Ferredoxin expression from pET30a(+) was induced by addition of 1 mM IPTG and the cultures were incubated at 16 °C for 20 hours. The cells were harvested by centrifugation and the pellets stored at -80 °C for up to one week. A spectrophotometric assay can be done to determine the presence of ferredoxin in cell-free extracts (Schenkman and Jansson, 2006). A cell-free extract (CFE) of the ferredoxin culture was created using the technique presented in Section 2.7.4. The CFE was empirically examined for turbidity and diluted 50% in P450 Lysis Buffer if necessary (Appendix). A 1 mL sample of CFE or diluted CFE was used to zero the spectrophotometer and establish a baseline absorbance reading from 600 to 300 nm. The assay measures a difference in absorbance between these wavelengths when the heme co-factor in ferredoxin is reduced by sodium dithionite. The addition of a few crystals will reduce the heme center of ferredoxin and produce a shift in the absorbance spectrum of the protein between these wavelengths (Schenkman and Jansson, 2006). After sodium dithionite crystals were added to the sample, another spectrum scan from 600 to 300 nm was recorded. The data was saved and plotted using spreadsheet software.

This type of assay is a powerful tool to learn if a recombinant enzyme that contains heme is present at a high concentration and whether or not it may have catalytic activity (Estabrook, 2003). The heme must be correctly inserted and coordinated with the ferredoxin amino acids for the protein to be active; if it isn't, the protein will mostly likely not function nor will the spectrum match the theoretical ideal (Schenkman and Jansson, 2006). Unfortunately, this type of assay can be complicated by other hemoproteinspresent in the CFE; however, f the concentration of ferredoxin in the CFE is high enough, its signal should be apparent (Schenkman and Jannson, 2006). Therefore, the assay tests efficiency of expression as well as heme insertion and indicates whether or not the heterologous hemoprotein may

function. SDS-PAGE analysis of the CFE was also done and while it is a useful technique, it cannot inform us on the status of the heme.

Section 2.7.3: Purification of Ferredoxin Cloned by Our Collaborators

The commercial vector pET28b(+) can be used to include either an N-terminal or C-terminal His-tag on a recombinant protein, according to the manufacturer's instructions. Our collaborators codon optimized the ferredoxin ORF from *B. japonicum* USDA110 and cloned it into pET28b(+). The construct was graciously provided to me during my visit to ISU. Unfortunately, the information on the synthetic sequence, cloning sites and choice of terminal tag had been lost (Reuben Peters, personal communication). Although we tried to have the plasmid sequenced through the multiple cloning site, to rediscover this information, the reactions were unsuccessful.

E. coli C41 transformants were grown overnight in LB broth with 10 µg/mL kanamycin selection. 1/100th volumes of the saturated seed cultures were used to inoculate 50 or 500 mL expression cultures the next morning. Expression cultures were grown for approximately three hours at 250 rpm in a 37 °C incubator before supplementation with 1 mM $FeCl_3 \bullet 6H_2O$. Expression of ferredoxin was induced by the addition of 1 mM IPTG. Protein expression was carried out at 16 °C for 20 h. The cells were harvested by centrifugation in a J2-HS machine using either a JA 17 or JA 10 rotor as appropriate to accommodate the culture volume (Beckman-Coulter). Pellets were stored at -80 °C until purification of the protein.

Pellets were rapidly thawed by the addition of Lysis Buffer (Appendix) and homogenized thrice with the pressure set between 10,000 and 15,000 kPa in the high-pressure homogenizer. The CFE was clarified by centrifugation in the J2-HS centrifuge chilled to 4 °C. The rpm was set to 14,500 and the JA 17 rotor was used. The supernatant was applied to 4 mL cobalt resin slurry (Thermo Fisher Scientific # 89965) that had been equilibrated with 10 X column volume of Lysis Buffer. The slurry

and supernatant were allowed to bind for one hour at 4 °C. During this time the mixture was kept in 50 mL plastic centrifuge tubes attached to a LabQuake rotisserie apparatus by elastics (Thermo Fisher Scientific #400110Q). After this, the mixture was poured into a plastic column. The flow through was re-applied to the column before washing the resin 6 times with 5X bed volume of Wash Buffer (Appendix). Protein was eluted from the column using 10 mL of elution buffer (Appendix) and fractions were collected in 200 µL volumes. Samples were analyzed by SDS-PAGE and spectrophotometry.

Section 2.7.4: Production of CFE from *E. coli* culture Expressing CYPs

Expression cultures of *E. coli* C41 harbouring the pDEST-14 plasmids with codon-optimized CYP genes were prepared as described (Section 2.7.1). The cells were harvested by centrifugation using the JA 10 rotor fitted on the J2-HS centrifuge manufactured by Beckman-Coulter. The 10 minute spin was conducted at 4,000 rpm and 4 °C. The supernatant was discarded and the cells were resuspended in P450 Lysis/Activity Buffer (Appendix) at 1/10th of the original culture volume. To assist in lysis of the cell wall, lysozyme was added to a final concentration of 0.5 mg/mL to assist in digestion of the cell walls (Juijun Cheng, personal communication). The suspensions were incubated in ice for up to one hour with occasional mixing by aspiration with a plastic transfer pipette. Sonication or pneumatic homogenization was used to decrease the viscosity lysed cells. Sonication was carried out in an ice-water bath using a Branson Sonifier 450 fit with a sonicator horn. The instrument was used on a duty cycle of 70 with the output control set to 4. A 20 s sonication period followed by 40 s of rest was repeated 8 times for each sample. Sonication was done on samples of 5 mL or less. If larger volumes were used, one pass through the homogenizer at 10,000 to 15,000 kPa was sufficient to reduce the viscosity of the mixture. Following sonication or homogenization, the paste was centrifuged for 20 min at 12,000 rpm using the JA 17 rotor in the J2-HS centrifuge to produce CFE. The temperature of the machine was set to 4 °C. This spin served to clarify the supernatant by pelleting debris and unlysed

cells. The resultant supernatant was the CFE.

Section 2.8: Difference Spectra of CYPs.

We based our protocol on established spectrophotometric methods for detection of CYPs (Guengerich *et al*., 2009; Schenkman and Jansson, 2006; Morrone *et al*., 2010). Immediately following clarification of the lysed cells (described in **Section 2.7.4**), the supernatant was decanted and the pellet discarded. Empirical observations of the CFE determined if the supernatant was still too turbid for spectral analysis. If the supernatant was found to be too turbid, it was diluted by 50% in P450 Lysis/Activity Buffer (Appendix) for spectrophotometric analysis. Dilution of the supernatant reduced interference in the spectral scans. When this dilution was made, the fresh DTT added by the P450 Lysis/Activity Buffer reduced whatever proteins (including CYPs) were present in the homogenate. Therefore, the addition of sodium dithionite had less impact on the spectral shifts and the results of the scan were less dramatic. Fresh CFE was prepared each day that spectral scans were performed.

Modified P450 difference spectra were recorded using a dual beam Cary50 UV/Vis spectrophotometer set to wavelength scan with baseline correction from 525 to 385 nm. A sodium dithionite solution of 0.01 g/mL in P450 Lysis/Activity Buffer (Appendix) was prepared daily and stored on ice prior to use. A 10 µL volume was added to a 1 mL sample of clarified CFE in a disposable plastic cuvette. The cuvette was covered with parafilm and inverted several times to mix, or the liquid was aspirated with a pipette to achieve the same result. After 30 seconds, a baseline scan of the sample was recorded. The cuvette was removed from the machine and CO gas was bubbled in at a rate of 1 bubble per s for 40 s (Schenkman and Jansson, 2006). A scan between 525 and 385 nm was recorded.

Section 2.8.1: Enzyme Assays using CFEs

Although the CFE samples were immediately used to prepare enzyme assays, time course spectral analysis demonstrated that the CYPCFEs were stable for up to two hours after lysis. After two hours, the peak between 450 and 490 nm decreased and the CFEs were considered catalytically

inactive; despite this, enzyme assays were allowed to proceed for 24 hours on the advice of our collaborators (Reuben Peters, personal communication). Spectral comparison between sonication and homogenization did not show a difference between sample preparation techniques, although there was an internal debate among lab members over which method was the best. Enzyme assays were performed in 15 mL screw-cap glass tubes. Five hundred μL of CYP CFE were added to the tube first, followed by NADPH and authentic substrate. The reaction mixture was allowed to incubate on ice while the CYP spectra were recorded from another CFE sample. To prepare the ferredoxin for addition to the reaction mixture, sodium dithionite was added. Ten μL of a 0.01 g/mL solution was used on the advice of our collaborators (Reuben Peters, personal communication). After 10 minutes of reduction on ice, five hundred mL of the ferredoxin CFE was added to the mixture and the reactions were incubated at 30 °C overnight. Despite the fact that sodium dithionite renders the CYPs useless, this practice was encouraged by our collaborators. I later learned that it was was an untested protocol and had never worked for the lab before, although I was led to believe that it was standard practice in their laboratory. This should be revisited if protocols are to be developed in the future.

Organic extractions and derivatization of the products were carried out in a working fume hood. Derivatization of the products is necessary so that they can be detected by gas chromatography/mass spectrophotometry (GC/MS) because all of the potential reaction products are very polar compounds. The following morning, 1 mL of n-hexane was overlaid onto the aqueous phase of the enzyme assay mixture described in the previous paragraph. The tube was gently inverted several times and centrifuged at half speed in a bench top clinical centrifuge to reduce bubbles and clarify the phase interface. The hexane was removed using a 9 inch Pasteur pipette and aliquoted into a 16 x 10 mm glass test tube. An additional mL of n-hexane followed by two 1 mL ethyl acetate extractions of the aqueous fraction were also done. The ethyl acetate was aliquoted into another glass test tube. The

hexane and ethyl acetate fractions were evaporated to dryness with N_2 gas at ambient temperature and then resuspended in a small volume (~5 drops, approximately 80 µL) of n-hexane. A single drop of hexane saturated diazomethane was added to the test tube to derivatize the reaction products (For a review of the use of diazomethane in GC/MS, please see Müller, Düchting and Weiler, 2006). After approximately 10 min the hexane-diazomethane solution was evaporated under nitrogen and the compounds were resuspended in 4 drops of n-hexane. The solution was transferred to autosampler-compatible GC vials with screw top closures fitted with small-volume adapters for GC/MS analysis. This protocol was adapted from previously established GC/MS procedures used by other members of the lab (Wilderman and Peters, 2007).

Section 2.9: GC/MS Analysis

At Iowa State University, GC/MS analysis was done with assistance from members of the Peters laboratory. We used a pre-established GC/MS method (Wilderman and Peters, 2007). Briefly, samples of 1.0 µL were auto injected into a HP-1MS column set to splitless mode. Resultant mass spectra from 50 to 500 *m/z* were collected over a period of 30 min and compared to standards and experimental controls to determine if compounds of interest were generated by the CYP enzymes. Peaks of interest were computationally isolated and the resultant ions manually examined for characteristics of diterpenoid carboxcylic acids. At UW, I developed a GC/MS protocol for the detection of isoprenoid compounds with the assistance of Dr. Richard Smith of the University of Waterloo Spectrophotometry Facility to continue the work. We used synthetic standards obtained from Dr. Lewis Mander at Australia National University to test and validate our GC/MS method. We tested the plasmid based metabolic engineering system developed at ISU and graciously provided by Reuben Peters for its usefulness for synthesizing kaurene as an authentic substrate for our CYP assays.

Extractions from cultures of the system were done following the ISU method described by Wilderman and Peters (2007), but the GC/MS settings were changed in accordance with the mechanics and protocols for the UW equipment, which were different. Unfortunately, due to regulatory challenges, we were unable to obtain Diazald (Sigma #D28000) required for safe synthesis of diazomethane. Diazomethane, despite being a very useful compound for gas chromatography, is dangerous for someone untrained in organic synthesis to make unless Diazald is used in a controlled synthesis apparatus, according to manufacturer's instructions. Diazald is considered a hazardous good by the Canadian government and the regulations on its import are difficult to get around, however given time they may be bypassed. On our time line, we were not able to do so, so we could not continue our study of the CYPs involved in GA biosynthesis. However, we were able to detect *ent*-kaurene generated by the plasmid-based metabolic engineering system, so, if this research were to be continued, a solid foundation at UW already exists to further the collaboration**.**

For GC/MS analysis at UW, A 1.0 µL sample was autoinjected into an Agilent 5975B gas chromatography machine in split mode maintained under high vacuum at 7.34 e^{-6} Torr. A 3 min solvent delay was enabled because we found that the carrier solvent, hexane, eluted from the column at 1.8 min. The solvent delay protects the sensors from damage because a much larger concentration of solvent is used in the analysis. If the sensor is left on while the solvent passed over it, it would cause damage to the sensor (Richard Smith, personal communication). An initial oven temperature of 50 °C that ramped to 350 °C at a rate of 20 °C per min was used. The temperature was held at 350 °C for 3 min. Scans were taken in the 50 to 500 m/z range after the three min solvent delay. Using the HP-1 MS column, we found that *ent*-kaurene eluted from the GC column at precisely 15.9 min.

Section 2.10: The Plasmid-Based Metabolic Engineering System

We made six new expression constructs using the pET Duet-1 plasmid to test the substrate specificity of each of the CYPs with and without the ferredoxin gene expressed. Each of the pAM13X plasmids have one *B. japonicum* USDA110 CYP cloned into the second multiple cloning site (MCS) of pET Duet-1. Three of the plasmids also have the ferredoxin cloned into the first MCS; this was done so that we could test whether or not the ferredoxin was required for the catalytic function of the CYPs. We tested the expression of the CYPs in *E. coli* C41 by SDS-PAGE. We also measured the production of *ent*-kaurene by GC/MS. Unfortunately we were not able to test the complete system using pIRS, pGGeFF and the pAM13X vectors because of time constraints and technical limitations in our laboratory.

Section 2.11: Deletion Mutant Construction: Molecular Biology

To create the deletion mutants, a two-step PCR reaction was used to fuse the upstream and downstream DNA flanking each target ORF together (Sukdeo and Charles, 2002). The amplicon was cloned into an integration vector and through a two-step selection process with kanamycin and then sucrose, potential deletion mutant colonies were identified (Schäfer *et al*., 1994). Using standard phenotypic screening methods, and then genotyping by PCR and Southern blotting deletion mutants were confirmed. For a recent review on Southern blotting, please see (Brown, 2001).

Genomic DNA was isolated from the wild-type strains of three Rhizobiales: *B. japonicum* USDA110, *M. loti* MAFF303099, and *S. fredii* NGR234. This DNA was used as PCR template. The primers used are shown in **Table 2.2**. The integration plasmids used to generate the deletion mutants were made by direct molecular cloning of the PCR amplicons. To analyze the plasmids before using them to delete the CYP genes, the, plasmid DNA was isolated from *E. coli* DH5α cells, subjected to restriction digestion and then gel electrophoresis. Confirmed clones were mated into the rhizobial strains (*B. japonicum* USDA110, *M. loti* MAFF303099, and *S. fredii* NGR234) and deletion mutants obtained through a two-step selection. Our protocol was based on previous work (Schäfer *et al*., 1994).

Please refer to **Figure 2.4** for a visual description of this experiment. **Figure 2.5** has a more in-depth description of the early stages of deleting a gene. The PCR primers were designed with a 15 to 20 base pair overlap in accordance with protocols developed in our lab (Sukdeo and Charles, 2002). This allows for a second PCR to join the original two upstream and downstream products from the first round of amplification. Restriction endonuclease recognition sites were added to the ends of the F1 and R2 primers to facilitate molecular cloning into the vector pK19*mobsacB* when necessary (Schäfer *et al*., 1994). PCRs were set up at room temperature, but enzymes and buffers were kept in -20 °C bench top coolers during set up. The thermocycler was programmed with settings previously optimized for amplification of *B. japonicum* USDA110 genomic DNA and involved use of a touchdown cycle

(Trainer, 2008).After gel electrophoresis analysis to confirm the expected amplicon size and check the experimental controls for the PCR, 1 μL of each original PCR product, each one corresponding to the upstream or downstream DNA flanking the gene to be deleted, was added as template for the second PCR. The same reaction volumes and thermocycler settings were used for the second round of PCR (Trainer, 2008). The PCR product was gel extracted and cloned into pK19*mobsacB* using the protocol in **Section 2.5.7**. The resultant plasmids are listed in **Table 2.3**. Restriction digestion and gel electrophoresis were used to confirm that the PCR amplicon was successfully cloned into the integration vector as described in **Section 2.5.6**.

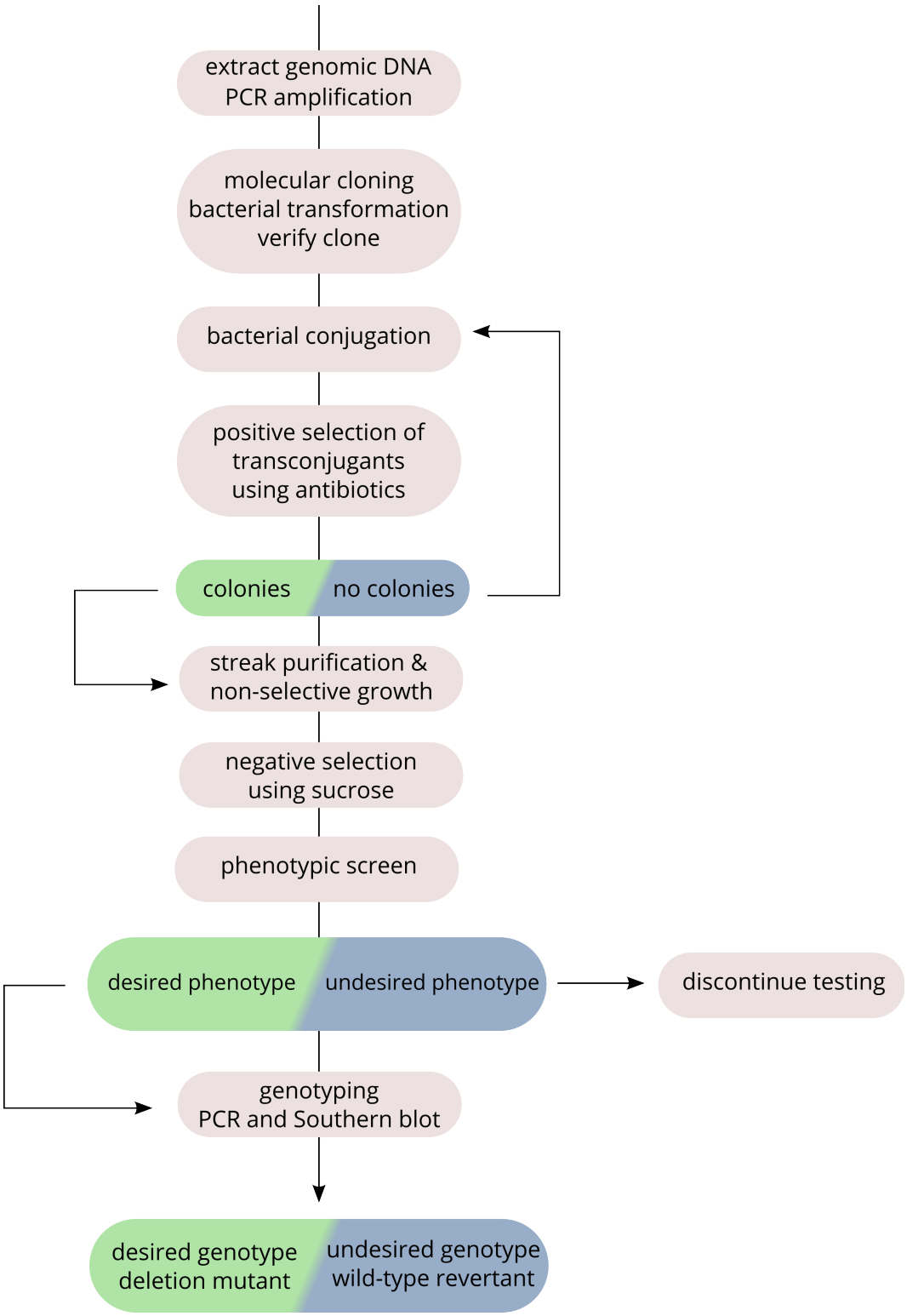


Figure 2.4: Work flow Diagram: Construction of Deletion Mutant Strains

Details for each step in the process are found later on in this chapter. The major steps in design and production of a bacterial deletion mutant strain are shown in grey ovals. Steps where a result decides if the process is continued or not are shown by green and blue split ovals. The green portion is the desired positive result and the blue section is the undesired negative outcome. When working with *B. japonicum* USDA110, or other rhizobia, the entire process may take several months because of the need to restart at certain steps and the relatively long growth phase of the organisms.

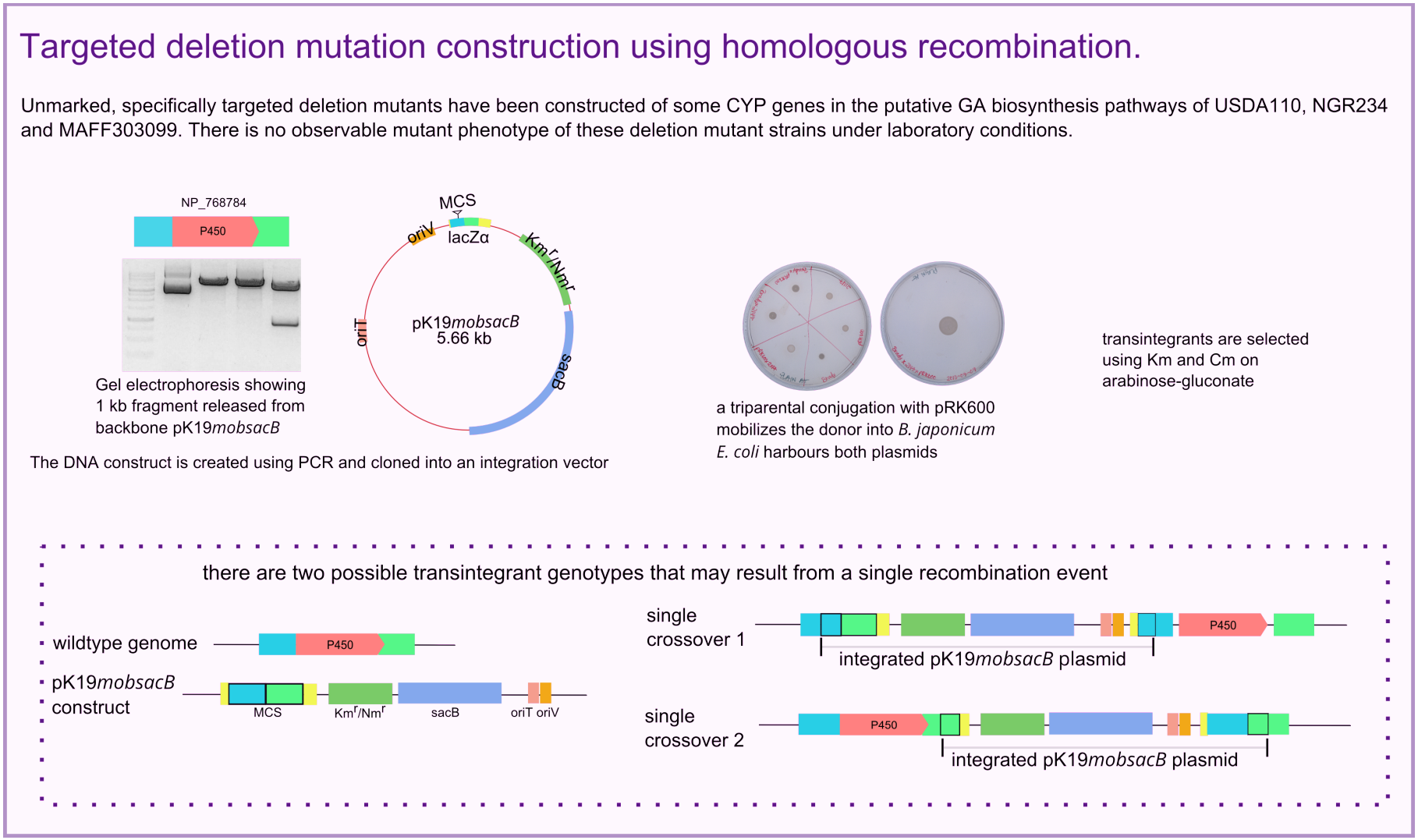


Figure 2.5: Two transintegrant genotypes are possible with homologous recombination.

This figure shows the key steps involved in creation of a deletion mutant. The gel image shows how the cloned fragment is checked using restriction digestion. The reaction products are visualized using gel electrophoresis and the resultant integration vector (pK19*mobsacB*). Bacterial conjugation is shown on the Petri plates. The lower panel shows the two possible genotypes that can arise from integration of the homologous fragments into the genome.

Section 2.11.2: Deletion Mutant Construction: Microbiology

To move the integration plasmids from the *E. coli* donor strain (the one harbouring the integration plasmid) to the rhizobia, tri-parental mating with an *E. coli* mobilizer strain (harbouring plasmid pRK600) was performed (Finan *et al*., 1986). Our protocol is based on the work referenced here. The cells from 5 mL of saturated *B. japonicum* USDA110, *M. loti* MAFF303099 or *S. fredii* NGR234 cultures were collected by centrifugation at 6,000 rpm for 5 minutes in an Eppendorf 5430 centrifuge using an F-35-6-30 rotor. These were the recipient bacterial strains. One mL of both donor and mobilizer *E. coli* strains (harbouring pAM101-pAM110 or pRK600, respectively)were subjected to centrifugation in a KL 064 for 1 minute at 13,000 rpm to collect the cells. The cell pellets were washed with 1 mL of sterile saline (8 g/L NaCl) and resuspended in 50 µL saline. A 2 µL sample was removed from each tube and spotted on solid AG or TY agar as control mating spots. The following control mating spots were generated: single spots of recipient, donor and mobilizer strains; bi-parental matings between the recipient and donor strains; recipient and mobilizer strains; and donor and mobilizer strains. After creation of the control spots, all strains were mixed together by gentle aspiration with a pipette and spotted on solid agar. Plates were incubated at 30 °C overnight. The following day all spots were scraped from the plates and resuspended in sterile TY or AG broth, depending on the recipient species. Control mating spots were plated directly while the experimental (tri-parental) mating spots were serially diluted to 10^{-6} and then plated. The agar plates were impregnated with antibiotics as required for the particular mating. A sample of the undiluted mating spot was also streaked on the same medium. The medium, antibiotics, and the concentrations that were used for selection of transconjugants are tabulated in **Table 2.4**. All colonies that arose were assumed to be single integrant/recombinant strains but checks and controls were employed to determine if any were spontaneous antibiotic resistance mutants.

Table 2.4: Antibiotics Used to Create Mutant Strains

Recipient Strain	Media Used	Antibiotic Concentration ($\mu g\ mL^{-1}$)
B. japonicum USDA 110	Arabinose-gluconate	Tetracycline 200 Kanamycin 50
M. loti MAFF303099	Tryptone-yeast	Streptomycin 200 Neomycin 200
S. fredii NGR234	Tryptone-yeast	Rifampicin 25 Neomycin 200

Any colonies that arose on the selection plates were streak-purified on agar with antibiotics appropriate for the host strain. Single isolated colonies that had been streak-purified twice were used to inoculate two test tubes of broth, one with and one without antibiotic selection. The cultures under antibiotic selection were used to make frozen stocks while the non-selective cultures were serially diluted to a final dilution of 10^{-6} in broth and then plated on medium with 10% (w/v) sucrose for negative selection of the second recombination event.

Section 2.11.3: Negative Selection Using *sacB* and Phenotypic Screening

Any colonies arising on sucrose plates were streak-purified twice on the same medium. Then, the sucrose and antibiotic resistance phenotype of all colonies were screened to distinguish their implied genotypes. Colonies that were sucrose-resistant and kanamycin/neomycin resistant were assumed to have acquired *sacB* mutations. Colonies that were resistant to sucrose but unable to grow on antibiotic plates were either mutants or their genome had reverted to the wild-type state. Using phenotypic screening we were able to find colonies with sacB mutations and discount them from further analysis. To distinguish between mutants and wild-type revertants, genotypic with PCR or Southern blotting was used.

To screen a colony one was picked from the sucrose streak plate and then patched successively onto an agar plate with sucrose, one plate impregnated with the antibiotic for background host selection (either tetracycline or streptomycin depending on the recipient strain, see **Table 2.4**) and then a third

plate with either kanamycin or neomycin. All plates were incubated at 30 °C for at least three days.

Section 2.11.4: PCR Screen of Putative Mutants

Once colonies with the desired phenotype (sucrose resistant, kanamycin/neomycin sensitive) were found, they were used to inoculate non-selective broth cultures and incubated with shaking at 30 °C until they reached saturation. Genomic DNA extraction was done as previously described (**Section 2.5**). The strains were also saved at this point and stored at -80 °C. PCR analysis was done using the reaction conditions previously described to determine the local genotype at the desired deletion locus for each candidate strain (Trainer, 2008). New primers were designed and their sequences are shown in **Table 2.2.** If a mutant genotype was detected following the gel electrophoresis of the PCR products performed to determine their size, Southern blot analysis was carried out on that DNA to confirm the genotype of the strain. Our Southern blot protocol was based on the one found in Short Protocols in Molecular Biology (Ausubel *et al*., 1995) and instruction from other lab members (Trevor Charles, personal communication).

Section 2.11.5: Southern Blot Protocol

Southern blot probe was synthesized from PCR product using a kit and the protocol was carried out according to the manufacturer's instructions (Roche # 11 093 657 910).Depending on the specific gene and species, either the upstream or downstream PCR product was used as template. The probes are synthesized by Klenow polyerase using random hexanucleotide primers and incorporate digoxigen-conjugated UTPs. The digoxigen-UTP is an antigenic hapten. This feature is used to detect the location probe once it has bound to the genomic DNA that is cross-linked to the Southern blot membrane. An antibody that recognizes digoxigen and is conjugated to horse radish peroxidase is used for this purpose. The peroxidase degrades a chemiluminescent substrate and the light that is emitted is detected by a machine.

Genomic DNA was quantified using Lambda DNA standards as previously described (**Section**

2.5.2). Approximately one µg of DNA was digested for each sample using the appropriate restriction enzyme. The recipe and reaction conditions for the digestion were standard, but scaled up to a total of 60 µL (**Section 2.5.6**). Typically, 40 µL of the mixture was composed of DNA sample and at least 1 µL (10 U) of the total volume was enzyme. Digestion was done for at least one hour at 37 °C. Occasionally, 5 U of restriction enzyme was added after 30 min to ensure that the DNA sample was completely digested. The entire reaction volume was loaded onto the gel; combs that made wells that could accommodate up to 60 µL volumes were used to cast these gels. Although it is recommended to use minimally thick gels for Southern blotting, the gels we cast were typically slightly thicker than normal to ease transfer of DNA from gel to membrane (Ausubel *et al*., 1995). We found that casting slightly thicker gels helped promote transfer of DNA from gel to the membrane by preventing the paper layers from touching the sponge during transfer. The samples were run on a 0.8% or 1 % agarose gel at 45 V for 4.5 hours. To visualize the DNA after electrophoresis, the gel was submerged in a solution of 1X TAE and GelRed. 25 µL of GelRed was used per 250 mL of 1X TAE. The container holding the gel was covered with aluminum foil and gently shaken at room temperature for 30 min. After this incubation period, the gel was imaged and then mounted for overnight transfer following an amalgamation of standard protocols outlined in two molecular protocol reference books (Sambrook and Russell, 2001; Ausubel *et al*., 1995).

After staining and imaging, the gel was rinsed with about 200 mL of distilled water and then immersed in 200 mL of 0.25 M HCl for 30 min. The gel was agitated at room-temperature on a rotary shaker at a low speed. Once thirty minutes had elapsed, the gel was rinsed once again in distilled water and then immersed in 200 mL Gel Denaturation Solution (Appendix)and kept on a low-speed shaker for 20 min. This was repeated twice then the gel was rinsed with distilled water. Finally, the gel was submerged in 200 mL of Neutralization Solution (Appendix) and gently agitated for 20 min. This was repeated. After the second wash in Neutralization Solution, the gel was set up on the transfer bed as

described and left overnight (Ausubel *et al.*, 1995).

The transfer reservoir contained enough 20X Standard Sodium Citrate (Appendix) to cover half of the sponge so that there was enough liquid to allow transfer of the DNA from the gel to the positively charged nylon membrane. Either Bio-Rad ZetaProbe (Bio Rad #162-0165) or Amersham Hybond-N+ (GE Healthcare Life Sciences #RPN119B) membranes were used for Southern blotting. The top left corner of the membrane was cut off so that tracking of the "DNA side" could be done once the membrane was removed from the sponge. We had best success incubating the membrane DNA side up when hybridizing and detecting the probe.

Immediately after disassembling the transfer stack, the DNA was permanently fixed to the membrane by UV cross linking using a GS Gene Chamber UV Linker (Bio Rad) on program C3. First, the program was run with the DNA side facing up, and then once again with the DNA side facing down. Appearance of DNA loading dye on the membrane was considered a positive sign of DNA transfer. Once cross linked, the membrane was immediately immersed in either a proprietary hybridization buffer (Roche # 11 093 657 910) or 50 mL Homemade Hybridization Buffer (Appendix) The gel was then re-stained with GelRed in 1X TAE and imaged. The lack of DNA visible in the gel was considered the hallmark of DNA transfer.

Pre-hybridization was carried out for 30 min to 3 h at 37 °C in the hybridization oven. The buffer was decanted and fresh hybridization buffer, with probe added at an approximate final concentration of 25 ng mL^{-1}, was added to the cylinder. The membrane was gently agitated overnight in the hybridization oven. The hybridization temperature was calculated according to the formula in the product manual. The formula is reproduced below, but for more detail, please review the manual (Roche # 11 093 657 910).

T_m = 49.82 + 0.41 (%G+C) – (600/l) [l = length of hybrid in base pairs]

$T_{opt.}$ = Tm – 22˚C.

The l variable, the length of hybridized region, was taken to be 500 bp. The variable for % GC changed depending on the target genome. For *B. japonicum* USDA110 Southern blot calculations, a % GC of 64.1 was used (Kaneko *et al*., 2002). When using the formula to calculate the optimal hybridization temperature for the *S. fredii* NGR234 Southern blots, a % GC content of 61.2 was used (Viprey *et al*., 2000). After an overnight hybridization, the probe solution was decanted and the membrane was washed twice with 100 mL of 2X Standard Sodium Citrate (SSC) and 0.1% SDS at room-temperature. Then, the membrane was washed twice with 100 mL of 0.1X SSC and 0.1% SDS that had been heated in a water bath to 68 °C. Further washes were carried out according to the manufacturer's instructions but are summarized here (Roche # 11 093 657 910). Most of the solutions were based on Maleic Acid Buffer (Appendix) The membrane was rinsed in Washing Buffer (Appendix) for 1 to 5 min then incubated with gentle shaking at room temperature in 100 mL Blocking Solution (Appendix) for 30 min. The Blocking Solution was discarded and replaced with 20 mL antibody solution (Appendix)) for an additional 30 min. The membrane was washed twice in 100 mL Washing Buffer for 15 min each time. Then it was incubated in 20 mL room temperature Detection Buffer (Appendix) for 2 to 5 min. Finally, the membrane was moved to a heat-sealable plastic bag and fresh detection buffer was added. To detect bound probe, a chemiluminescent substrate, chloro-5-substituted adamantyl-1,2-dioxetane phosphate (CSPD), was added. The 100 X CSPD stock solution was diluted to a working concentration of 1 X by adding 20 µL of CSPD per 1.8 mL of Detection Buffer (Roche # 11 093 657 910). The bag was sealed with a heat sealer and the chemiluminescent reaction was activated by incubation at 37 °C for 10 min before an image was captured.

Section 2.11.6: Southern Blot Detection of CYP Deletion Mutants

Detection of bound probes was done using the BIS303PC fluorescent imaging machine and the ECL tray in the University of Waterloo Department of Biology Molecular Biology Core Facility. The software was preset to detect large western blots. The probe and antibody system we selected to detect the Southern blots utilizes horse radish peroxidase conjugated to primary antibody. Please review the manufacturer's reference material for more information on the technique. This system does not use any radioactive material which simplifies the process since it does not require additional training, nor exposure and development of film. It is also relatively safer for the operator and does not generate radioactive materials. Light is emitted when the peroxidase catalyzes the breakdown of the chemiluminescent substrate CSPD. The antibody itself binds to the DNA probes which were chemically synthesized *in vitro* and contain digoxigen conjugated to a uracil nucleoside. Digoxigen is a strongly antigenic hapten which means that antibodies become strongly associated to it and by extension the DNA probe. The horse radish peroxidase is conjugated to the antibody and allows for detection of the bound probe over a time-lapse exposure as it breaks down CSPD. Light is released as a reaction product and is detected by the machine. We found that a time-lapse exposure that combined 4 images captured 5 mins apart was adequate for detecting the bound probe. Once an appropriate image was obtained, the blot was discarded.

Section 2.12: Results

Section 2.12.1: Initial Expression of CYP112

Attempts were made to obtain CYP112 by over-expression using *E. coli* as the heterologous host. The majority of the protein was insoluble, as shown by the picture of the SDS-PAGE gel in Image A of **Figure 2.6.** A small amount of purified protein could be obtained using methods described in Section X, though the yield was low. One litre of culture yielded approximately 10 g of cells and we calculated that the entire purification experiment yielded approximately 200 µg of protein. The purified CYP112 from this *E. coli* strain was very clean and the protein sample did not have extraneous or contaminating protein bands, as is obvious from the second SDS-PAGE gel. Ultimately, we decided that the yield was too low to pursue the enzyme assays, so we focused on the methods to optimize the expression conditions with the goal of increasing the solubility of CYP112 and subsequently the yield of purified protein. This included varying the expression conditions temperature, pH, concentration of IPTG used to induce the culture, rpm of the expression culture, media recipe, addition of supplements and more as well as changing the *E. coli* strain. We also tried expressing pAM101 in *E. coli* C41 and C43 both with and without pLysS, but did not see an increase in the solubility of the protein.

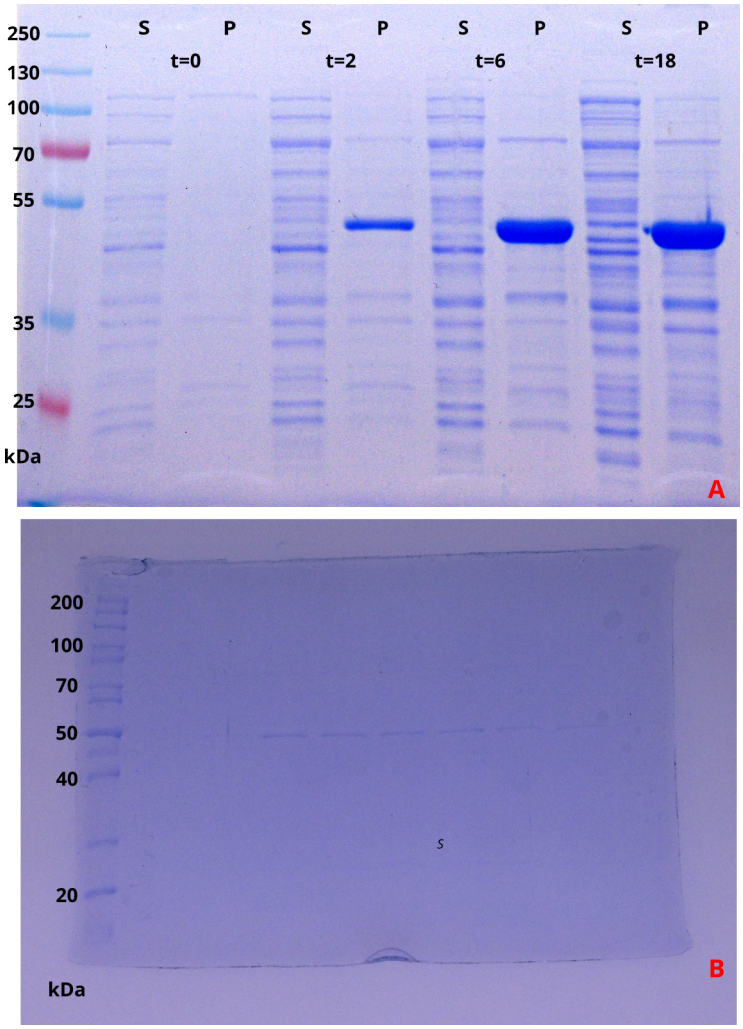


Figure 2.6: Expression and Purification of CYP112 from *E. coli* BL21.DE3 pLysS Cells

Using the His-tag staining technique described in detail in the appendix of this chapter, we were able to visualize the C-terminal His-tag of CYP112 when it was expressed from pAM101 using *E. coli* BL21.DE3 pLysS as the heterologous host (data not shown). Image A is a representative SDS-PAGE gel image stained with Coomassie. It shows the insolubility of CYP112 when cloned into pET30a(+) with a C-terminal His-tag. Samples were taken at time of induction (t=0) and then at 2, 6 and 18 hours post-induction with IPTG. The soluble fraction of each sample is marked with an S above the lane and the pellet (insoluble) fraction is marked with a P. The large bands just under the 55 kDa marker in the insoluble fractions are corresponding to CYP112. The vector used in to generate this image was named pAM101. The pAM101 plasmid was harboured by *E. coli* BL21.DE3 pLysS. Image B shows CYP112 purified from the soluble fraction of the CFE. The eluent from the cobalt resin was collected in approximately 200 µL fractions and 10 µL samples from every third fraction were used for this analysis. Relatively little CYP112 was purified overall.

Our collaborators provided us with a codon-optimized *B. japonicum* USDA110 ferredoxin construct in pET28b(+) that had an N-terminal His-tag. While visiting their laboratory, we attempted to purify the ferredoxin from CFE of *E. coli* C41 culture and then analyze it using similar spectrophotometric methods as in the analysis of the CYPs. The results of this purification are shown below in **Figure 2.7** on an SDS-PAGE gel. We collected 200 µL fractions and observed that beginning with fraction 10, the fractions were clear and brown as they eluted from the cobalt resin. The intensity of the colour decreased by fraction 19 and over two days at 4 ºC, the brown colour of the fractions faded completely.

SDS-PAGE analysis was done immediately after elution of the fractions from the column. Using 15 % denaturing non-reducing gels, we found that the fractions contained contaminating proteins, as evidenced by the extraneous bands shown in **Figure 2.7**. Additionally, the major bands appeared to correspond to 75, 25 and 15 kDa (highlighted in red ink on the Figure) while we expected the molecular weight of the His-tagged ferredoxin to be closed to 10 kDa. Eight µL of protein size marker (lane 1, first from the left) Precision Plus Protein All Blue Standard (BioRad #161-0373) was used.

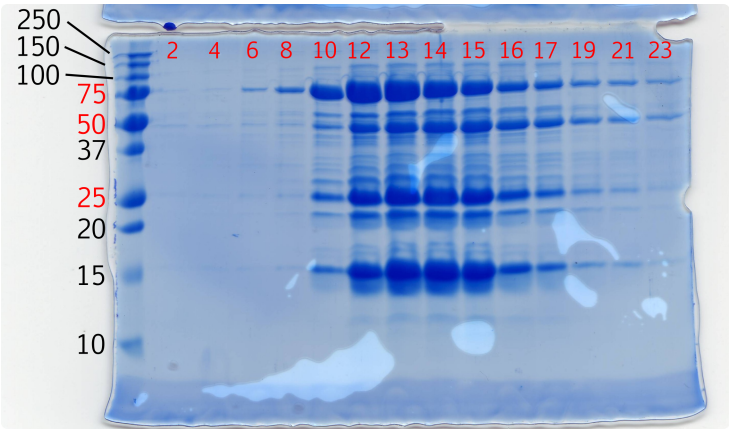


Figure 2.7: Purification of Ferredoxin from Whole Cell Free Extract
Many extraneous bands were observed on the SDS-PAGE gel. Analysis of the resulted was complicated because the major bands (at 75, 25 and 15 kDa) are not the size that we predicted the ferredoxin would be (10 kDa).

I tried to produce reduction spectra of the semi-purified ferredoxin but unfortunately, the results of this experiment were inconclusive. I was not able to produce a good spectrum for the protein so we decided not to use the purified ferredoxin for our future experiments. Instead, I decided to use CFE from separate cultures expressing both the CYP and ferredoxin after consulting with Dr. Peters at ISU. Unfortunately, I was also unable to produce reduction spectra for the CYP homogenates. I performed the enzyme assays anyway, and was largely unsuccessful in this venture as well. However, I did manage to obtain one piece of data which appeared to indicate that CYP117 was responsible for catalyzing the conversion of *ent*-kaurene to *ent*-kaurenoic acid. I was not able to reproduce this result during the research trip. Once I returned to Waterloo, I was unable to continue the experiments because of technical limitations in my research space and with my equipment and resources.

With the tools available to us, we were able to use the *E. coli* plasmid-based metabolic engineering system developed by our collaborators to produce and detect *ent*-kaurene by GC/MS. Roughly 95% of the product produced by the cells was actually a byproduct—indole—rather than *ent*-kaurene. Our collaborators has assured us that *ent*-kaurene would be the main metabolite produced under the expression conditions used, so we were very surprised to discover that in fact it was a minor product. We did not attempt to optimize the system to increase the yield of *ent*-kaurene. A 5% yield of *ent*-kaurene could not justify the use of litres and litres of culture medium, organic solvents and GC/MS costs. The yield of *ent*-kaurene from 1 L of culture was less that 50 μg in our hands, while our collaborators reported yields over 1 g/litre (Ryan Nett, personal communication). Although the issue of indole overproduction has not been documented by our collaborators, a brief literature search revealed that production of indole rather than the desired isoprenoid compound is a known problem with similar plasmid-based metabolic engineering systems (Ajikumar *et al*., 2010; Boghigian *et al*., 2012).

Section 2.12.2: Summary of Heterologous Expression Results

To summarize, we were able to express CYP112 at a high level, though it was mostly insoluble. Our attempts to increase its solubility were unsuccessful. We were able to purify small amounts of C-terminal His-tagged CYP112 from the soluble fraction of *E. coli* CFEs, but the yield was rather low. Ultimately, we did not use the purified protein for any enzyme assays or spectrophotometry. We were unable to completely purify the putative GA ferredoxin from *B. japonicum* USDA110, but did obtain some promising preliminary results indicating that it was expressed although we could not see it on SDS-PAGE gels of crude extracts. Unfortunately, we were not able to obtain good spectral data to indicate that the ferredoxin was present and catalytically active in the sample. The use of CFEs from *E. coli* cultures harbouring codon-optimized CYPs from our collaborators afforded us the opportunity to attempt to record CYP spectra and conduct enzyme assays to test for reactivity of the CYPs with *ent*-kaurene. Preliminary data suggested that CYP117 was able to catalyze the conversion of *ent*-kaurene to *ent*-kaurenoic acid.

We were able to use the *E. coli* metabolic engineering system developed by our collaborators to produce *ent*-kaurene, although we found that in our hands the major metabolite produced appeared to be the undesired by-product indole. Only a small amount of *ent*-kaurene synthesized. With time, we should be able to optimize the *E. coli* growth and protein expression conditions so that the formation of indole is reduced and the production of ent-kaurene is maximized. Additionally, although we could not test them, we were able to construct a set of vectors using the pET Duet-1 dual expression plasmid for co-expression of the ferredoxin in combination with each CYP. These vectors are compatible with the plasmid-based metabolic engineering system. Our collaborators should be able to make use of them in the future as their research on the CYPs progresses.

Because of the challenges of expressing and working with these CYPs in *E. coli*, our collaborators attempted their heterologous expression in an unconventional host: *Sinorhizobium meliloti* Rm1021. They were able to re-tool their plasmid-based metabolic engineering system for expression in *S. meliloti* using pLAFR5 vectors and using this system they achieved some very promising results for each of the CYPs; Furthermore, their results suggest a role for the ferredoxin as well as the SDR in GA biosynthesis (Reuben Peters, personal communication). Although they expressed the putative GA CYPs from *S. fredii* NGR234 rather than *B. japonicum* USDA110, these appear to be analogous systems (Reuben Peters, personal communication). Our collaborators' work with the CYP117 analog, CpxU, corroborated our preliminary results for CYP117, the analog CpxU (the gene product of *cpxU*) appears to be responsible for converting *ent*-kaurene to *ent*-kaurenoic acid (Ryan Nett, personal communication). The other major result that our collaborators obtained was a role for the SDR in GA biosynthesis. At this time, it appears that the SDR is responsible for the conversion of GA_{12}-aldehyde to GA_{12}, and possibly other reactions as well. All together, the data allowed for a putative role to be determined for all of the CYPs as well as the SDR and ferredoxin.

Section 2.12.3 CYP Gene Deletion Mutant Strains

Our major success and contribution to the GA project remains the CYP deletion mutants that we made and shared with our collaborators. The preliminary data that have arisen from the analysis of these strains in the laboratories of Dr. Cecilia Rojas and Peter Hedden are very promising. Through this collaboration, we believe that the metabolic roles of each of the CYPs involved in GA production in *B. japonicum* USDA110, and other bacteria, will soon be confirmed. We hope to continue the collaboration and look forward to the forthcoming discoveries made by our peers. We present data related to the construction of the mutants and their analysis on the following pages, beginning with **Figure 2.8**.

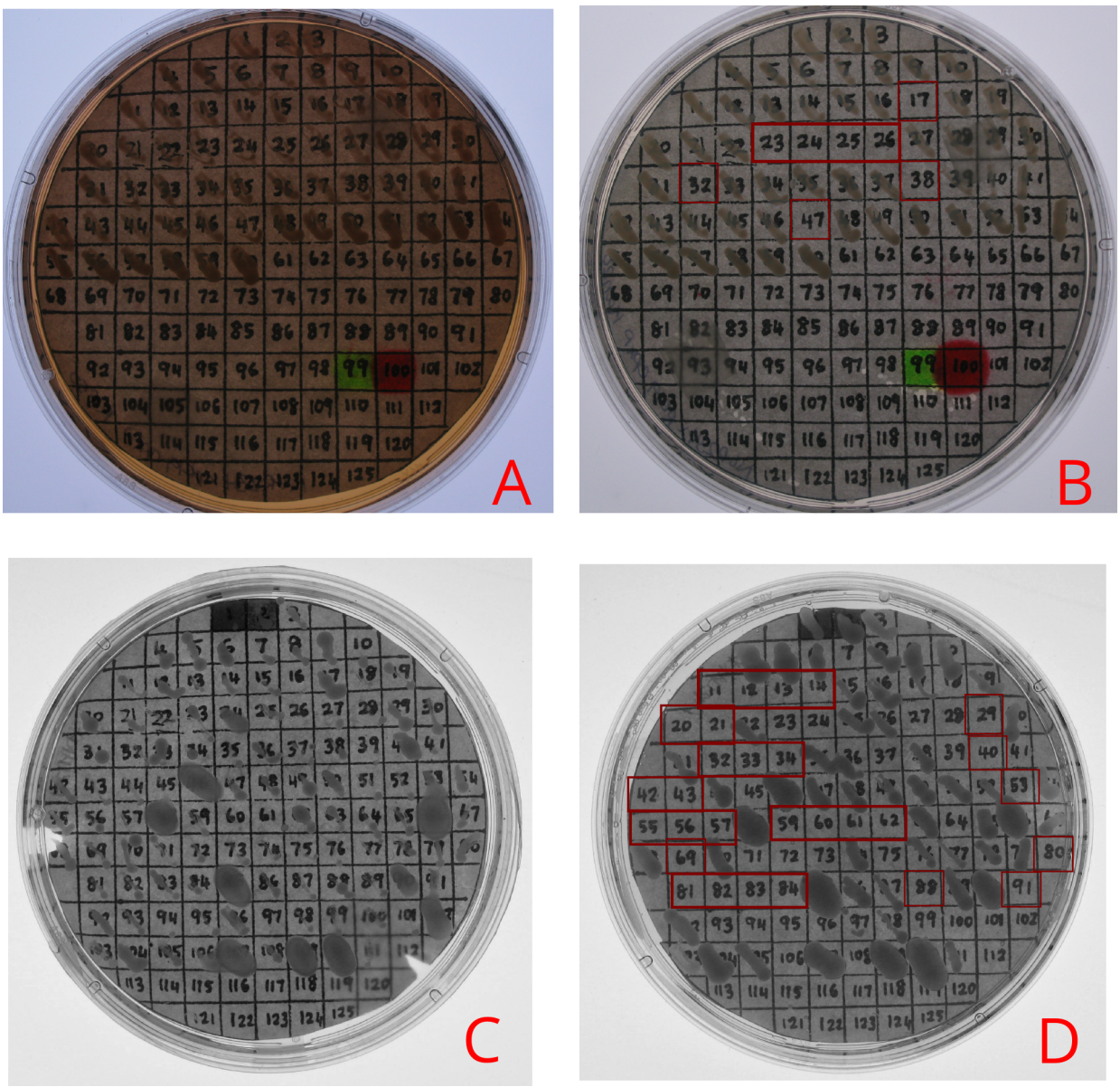


Figure 2.8: Screening Antibiotic Phenotypes To Infer Genotype

Picture A shows a TY agar plate with 50 µg/mL rifampicin. The corresponding plate is shown in image B. This is a TY plate with 200 µg/mL neomycin impregnated in the medium. The patches that didn't grow on plate B represent putative *cpxP* deletion mutants in *Sinorhizobium fredii* NGR324. Plate C and D contain 200 µg/mL tetracycline and 50 µg/mL kanamycin, respectively. Plates C and D were used to screen putative *Bradyrhizobium japonicum* USDA110 *blr2144* deletion mutants. Putative mutants are marked with red boxes in the images. Although neomycin and kanamycin resistance are encoded by the same gene, *Bradyrhizobium* is naturally resistant to neomycin but susceptible to kanamycin. So, kanamycin must be used for selection in *Bradyrhizobium*. Neomycin is used for selection in *Sinorhizobium* because that genus is naturally resistant to kanamycin.

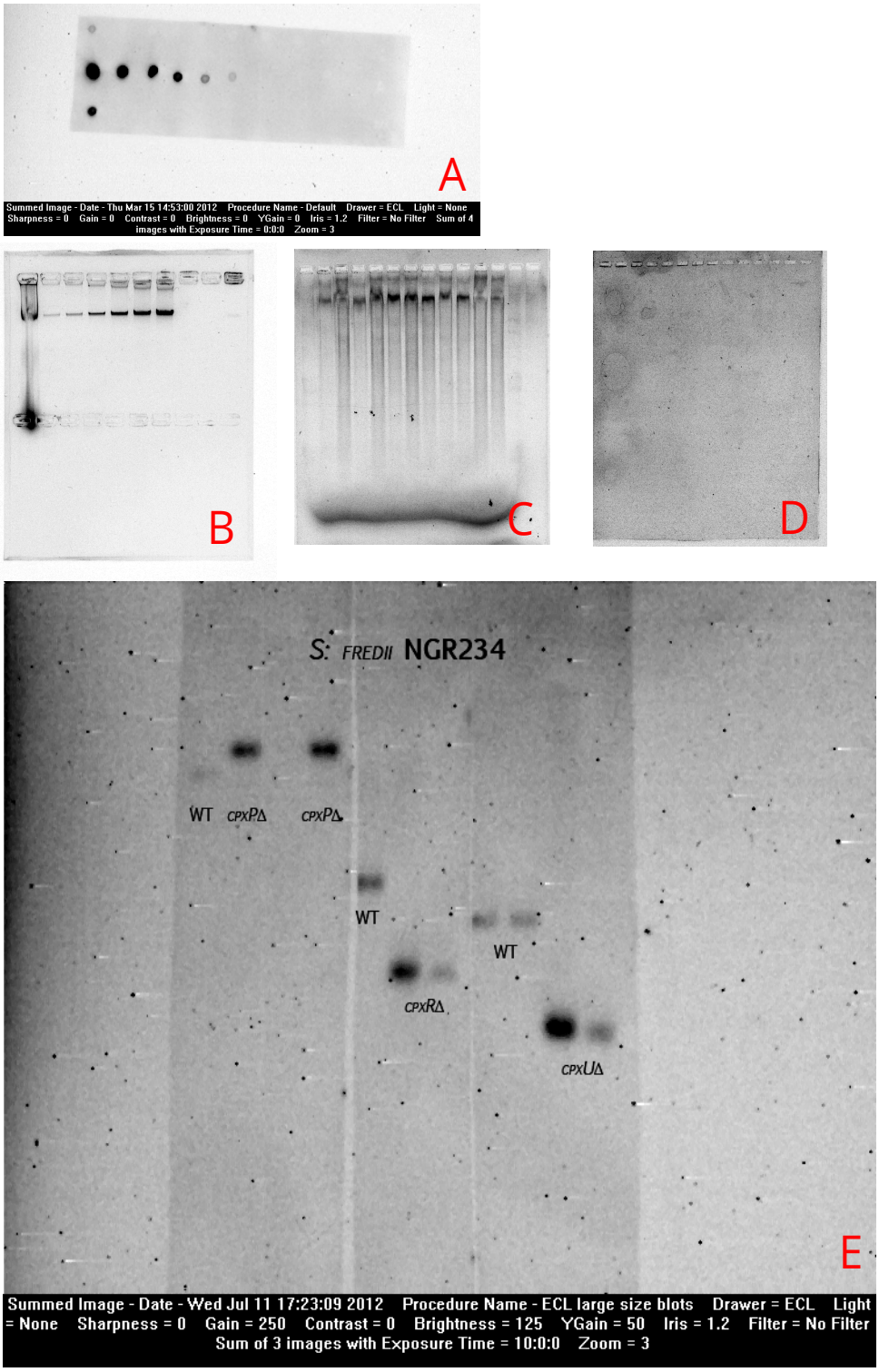


Figure 2.9: Southern blot probe detection, post-transfer, and blot detection for *S. fredii* NGR234 Δ*cpxP*, Δ*cpxR*, and Δ*cpxU* mutant strains

Figure 2.9 shows some important stages in the Southern blot process including the final detection. Image A shows the detection of the probe diluted to the suggested concentration indicating that it can be used according to the manufacturer's recommendations. Its presence was detected using the antibody and following the instructions provided by the manufacturer. Image B shows the lambda DNA quantification of a representative genomic DNA sample before digestion. Lane 1 (first from the left) contains 1 µL of the undigested genomic DNA sample and each subsequent lane contains 12 µL of Lambda DNA. Image C shows each genomic DNA sample after digestion with restriction enzymes. The samples are shown after running on the agarose gel for approximately 5 hr at 50 v. Image D is the gel after the DNA has been transferred overnight to the positively charged Nylon membrane. Image B, C, and D were stained with Gel Red as described in Section X. After transfer, the agarose gel becomes concentrated because it has lost water which is why Image D appears darker than Image B and C. Image E is a high-contrast picture showing the results of the experiment. A deletion mutant for each gene (*cpxP, cpxR,* and *cpxU*) was found and confirmed using this technique.

Deletion constructs for all 9 CYP genes from three bacterial strains were created using the vector pK19*mobsacB.* Integration mutants were created for all 9 target genes. We also created a single-crossover integrant strain for deletion of the SDR gene, *blr2146*, in *B. japonicum* USDA110 but were unable to obtain a deletion mutant. Deletions of 6 of the 9 CYPs were confirmed by Southern blotting. These strains were made available to our collaborators and have been used in preliminary experiments to discover the role of the affected genes in the GA biosynthesis pathway. We were unable to obtain knockouts of two *Mesorhizobium loti* MAFF303099 genes, *mlr6364* and *mlr6365 (*analogs of CYP112 and CYP114 respectively*)*, and the *blr2147* gene, encoding CYP117, in *B. japonicum* USDA110.

Section 2.12.4: Preliminary Results of CYP Substrate Specificity

The *blr2144* and *blr2145* deletion mutants (where CYP112 and CYP114 had been respectively deleted) were sent to Dr. Cecilia Rojas through our collaborators at Iowa State University. Her lab isolated the metabolites of extracted bacteroids fed with radioactively-labeled precursors, as described in their recent publication (Méndez *et al*., 2014). Deletion mutant bacteroids were incubated with ^{14}C-kaurenoic acid, ^{14}C-7b-OH kaurenoic acid, ^{14}C-GA$_{12}$-aldehyde and ^{14}C-GA$_{12}$, ^{14}C-GA$_{15}$ or ^{14}C-GA$_{24}$ depending on the strain. Feeding ^{14}C-KA to the *blr2144* deletion bacteroids resulted in accumulation of ^{14}C-7b-OHKA and ^{14}C-GA$_{12}$ aldehyde. A low conversion of ^{14}C-GA$_{15}$-ol to ^{14}C-GA$_{24}$ and ^{14}C-GA$_{9}$ was also observed. ^{14}C-GA$_{24}$ was not converted, nor was ^{14}C-GA$_{12}$. (Rojas, personal communication).

The reaction products from the *blr2145* incubation are still preliminary and must be completely identified by additional analysis in the laboratory of Peter Hedden. Therefore, the results presented are preliminary and based only on GC/MS analysis by comparison of the retention time and compound chromatograms with known standards. Bacteroids of the *blr2145* mutant were able to convert ^{14}C-KA, ^{14}C-7b-OHKA, ^{14}C-GA$_{12}$ aldehyde and ^{14}C-GA$_{12}$ to a variety of products, some with a high degree of efficiency (Rojas, personal communication). The most relevant results concern the lack of conversion of *ent*-kaurenoic acid into any GA derivatives—indicating a blockage in the metabolic pathway at this step. Conversion of downstream products, namely ^{14}C-GA$_{12}$ aldehyde to GA$_{9}$, GA$_{24}$ and GA$_{12}$,was highly efficient. ^{14}C-GA$_{12}$ gave the usual profile of the wild-type strain: conversion to GA$_{9}$ with intermediates of GA$_{24}$ and C-GA$_{15}$ also detected. The findings are summarized in **Figure 2.10** and **Figure 2.11**.

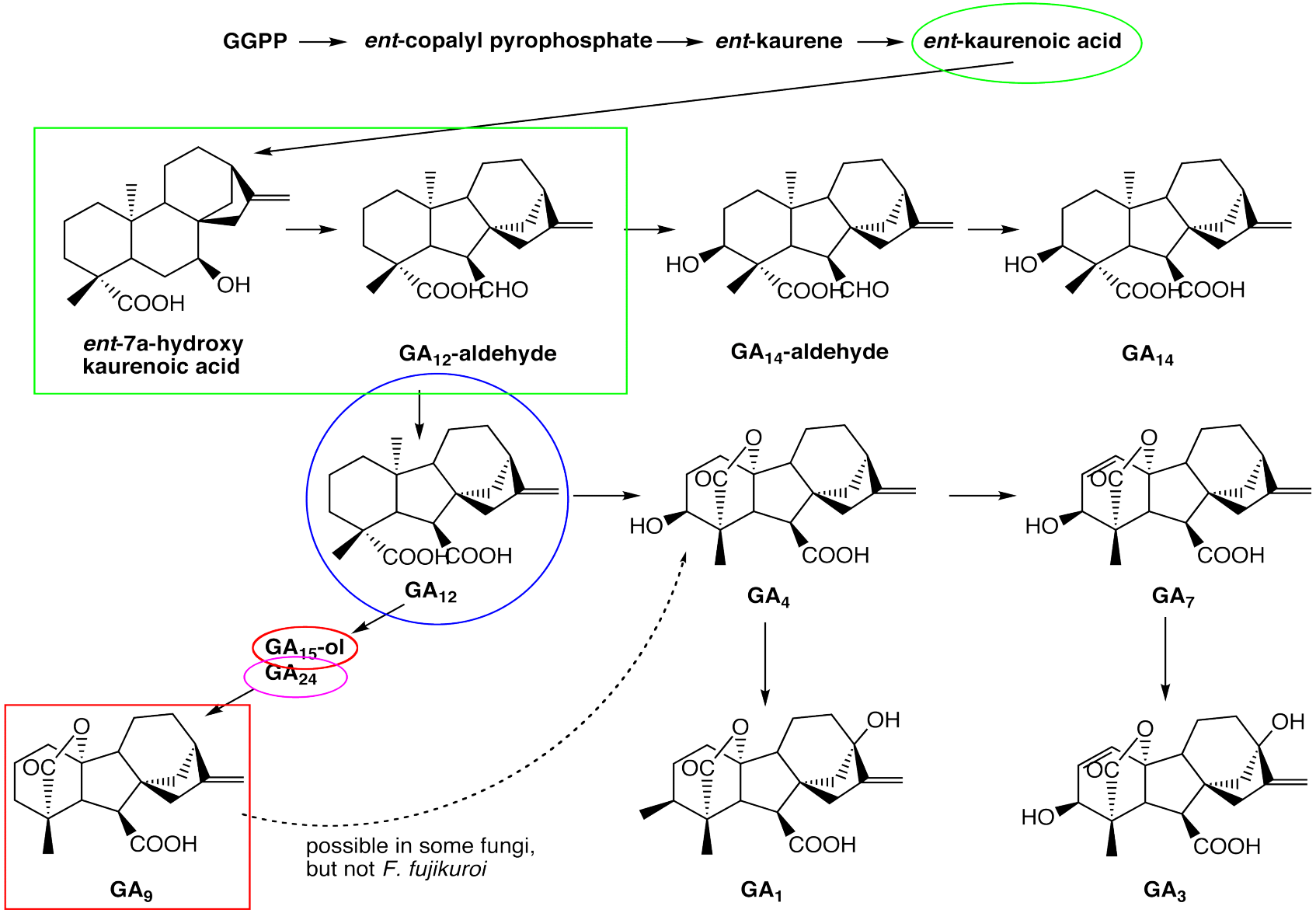


Figure 2.10: Metabolite Profile of *B. japonicum* USDA110 *blr2144* (CYP112) Deletion Mutant

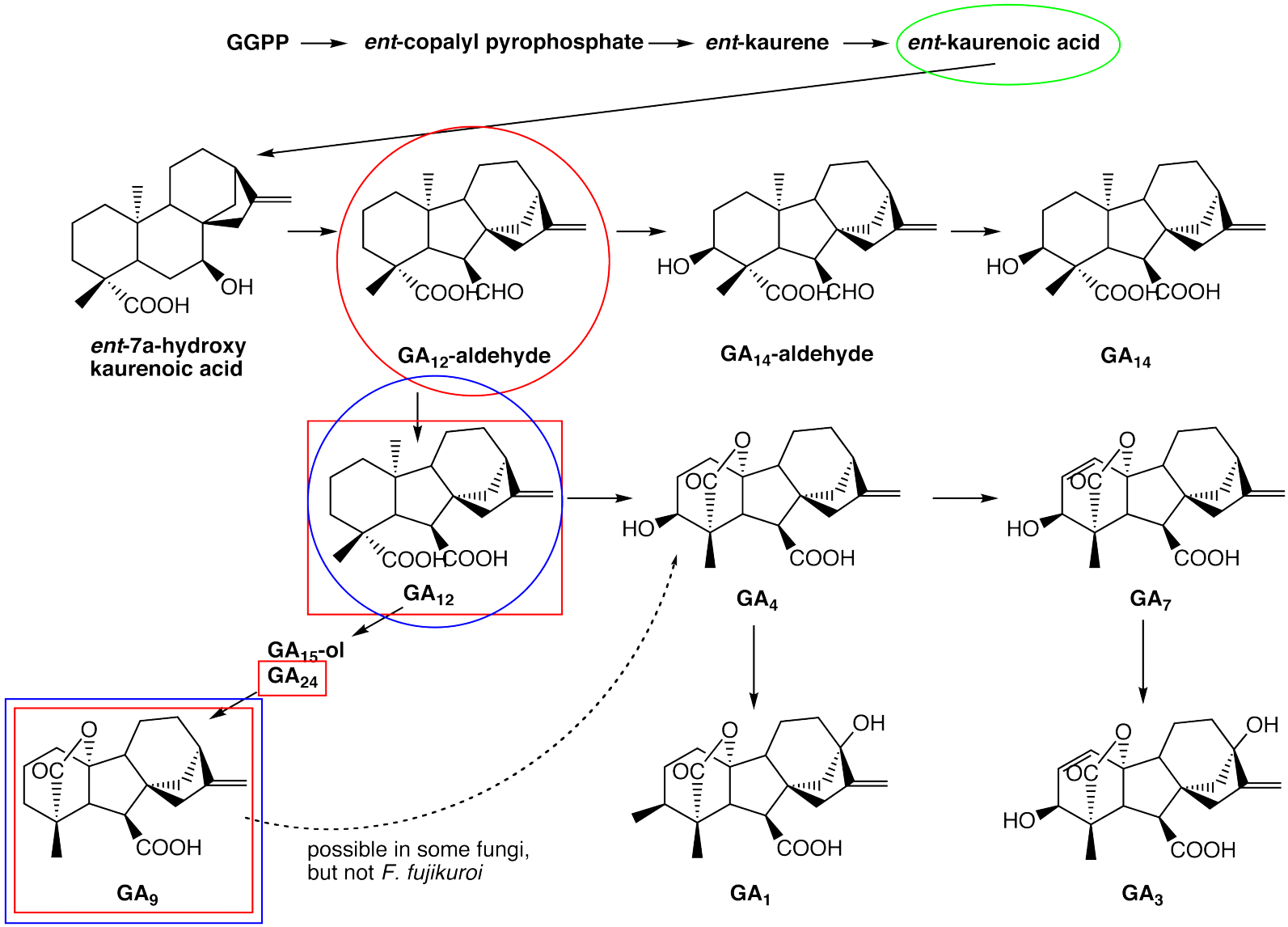


Figure 2.11: Metabolite Profile of *B. japonicum* USDA110 *blr2145* (CYP114) Deletion Mutant

Each strain was used to inoculate soybean plants and bacteroids were extracted from both samples (Méndez *et al*., 2014). **Figure 2.10** shows the results obtained from feeding Δ*blr2144* bacteroids. Coloured circles indicate the individual substrates fed to the bacteroids and corresponding coloured boxes indicate the identified products. It appears that the Δ*blr2144* mutant was largely unable to convert GA_{12} and GA_{24}. In this figure, the green circle indicates that the bacteroids fed ent-kaurene converted it to *ent*-7α-hydroxy kaurenoic acid and GA_{12}-aldehyde. When GA_{12}, indicated by the blue circle, and GA_{24}, indicated by the purple circle, were fed to the bacteroids, no products were detected. Feeding the substrate GA_{15}-ol, circled in red, resulted in detection of GA_9. Because GA_{12} and GA_{24} were not converted by the Δ*blr2144* bacteroids, the corresponding gene product, CYP112, was identified to catalyze these reactions. The results for the Δ*blr2145* mutant are shown in **Figure 2.11**. When *ent*-kaurenoic acid (circled in green) was fed to these bacteroids, no GA products were detected; however, when GA_{12}-aldehyde (circled in purple), GA_{12} (circled in blue), were fed to the Δ*blr2145* bacteroids, GA-products that match the wild-type product profile were found. These results show that the CYP114 protein, encoded by *blr2145* is responsible for the conversion of *ent*-kaurenoic acid to GA_{12}-aldehyde. At this time, some of the products generated by the fed bacteroids remain unidentified and are awaiting confirmation by NMR analysis (Cecilia Rojas, personal communication). Those products that have been tentatively identified by GC-MS are shown using the same colour-coded system. Products that were detected after feeding a particular substrate are indicated by coloured boxes that match the colour of the substrate. At this time, the identity of products generated by feeding the Δ*blr2145* strain *ent*-kaurenoic acid are not known—all that is known so far is that they are not any known gibberellin compounds.

Section 2.13: Discussion

Our research goal was to determine the substrate specificity of the cytochrome P450 enzymes involved in gibberellin biosynthesis in *Bradyrhizobium japonicum* USDA110 and we approached this question with two plans. The first plan was to express and purify the CYPs from *E. coli* cultures, perform enzyme assays, and analyze the products by GC/MS. The second plan was to create deletion mutants then analyze the metabolites of those strains compared to those made by the wild-type. Admittedly, when we began the genetic approach we did not have the expertise to analyze the mutants; however, we were confident that the resultant strains would prove useful in the future.

We hypothesized that by using different mutant strains, we would be able to detect where the pathway was blocked by the deletion of an enzyme by determining the build up of intermediate metabolites. Ultimately, some of mutants that we constructed as part of the genetic approach were used to determine the substrate specificity of the cytochromes in the pathway. Through our collaborators, the mutant strains were provided to a research group in Chile where a majority of the analysis was completed. That group collaborated with researchers in the United Kingdom to confirm the results and identify the compounds. The Chilean scientists isolated metabolites from soybean nodule bacteroids formed by the mutant strains and forwarded them to the British group where they are being identified by NMR analysis. The deletion strains were critical in determining the role of the cytochromes and the work done so far would not have been possible without them. Our collaborators in Iowa are continuing with the project and plan to delete additional genes in the operon so that the pathway can be completely characterized with this method. Although we dedicated a great deal of time and resources to heterologous expression of the CYPs in *E. coli*, our methods were largely unsuccessful. Fortunately, our efforts were not in vain because our collaborators have assimilated the knowledge from our failed attempts and continue to make progress with this venture using an alternative expression host.

Section 2.13.1: Heterologous Protein Expression

There were some significant challenges associated with heterologous expression of the CYPs in *E. coli*. Initially, we decided to clone the CYP ORFs from *B. japonicum* USDA110 into pET vectors based on recent reports characterizing CYPs from *Streptomyces, Mycobacterium* and *Pseudomonas* species (Lamb *et al*., 2002; Sielaff and Andreesen, 2005; Olaofe *et al*., 2013). N and C terminal His-tags have been used to purify CYPs, so we decided to try C-terminal His-tags with our constructs (Kempf *et al*., 1995; Lamb *et al*., 2002; and others)

Encouraged by the very good expression of CYP112 (and poor expression of the other CYPs), we decided to focus our efforts on the expression and purification of CYP112. Although expression of CYP112 was high, approximately 90% of CYP112 appeared to be insoluble. We hypothesized that CYP112 was most likely localized in an inclusion body (Williams *et al*., 1982). But, we thought it could be bound to the membrane and located in the pellet fraction of the CFE that we used for SDS-PAGE analysis. Unfortunately, our attempts to optimize the expression conditions did not result in increased solubility of CYP112, despite trying several methods.

We knew that our choice of expression vector could have been contributing to the insolubility of CYP112. The pET vectors have strong T7 promoters, a feature known to be problematic for the heterologous expression of some P450s (Barnes, 1996). Although pET vectors had recently been successfully used to express other bacterial CYPs, they did not appear immediately compatible with these CYPs. Still though, we did not want to re-clone the ORF unless it appeared to be absolutely necessary. Therefore, we tried to alleviate the solubility issue by focusing on creating conditions that could promote protein folding and increase solubility. We tried five different *E. coli* expression strains with no success. In compatible strains, those without a pLysS plasmid, we co-expressed our construct with pGro7 which encodes the GroES and GroEL chaperone proteins (Nishihara *et al*., 1998). We

supplemented the expression cultures with heme biosynthesis precursors, iron chloride heptahydrate and amino-levulinic acid, but did not achieve the desired results. Folding of CYPs is dependent on insertion of the N-terminal amino acids into the cell membrane and this is taxing on the cell in heterologous systems with very strong promoters, because more polypeptide is made (Fisher *et al*., 1992). Although CYP112 is not predicted to have trans-membrane domains, we decided to truncate the N-terminal amino acids in an effort to increase the solubility of the protein following previously published reports on eukaryotic CYPs (Barnes *et al*., 1991). Unfortunately, the truncated versions were not suitable for our purposes. We were forced to reevaluate our vector choice since all three of the CYPs constructs appeared fruitless. We decided to revisit our use of the C-terminal His-tag when we re-cloned the CYPs as well. We decided to re-clone the complete ORF with a His-tag into a well-tested CYP expression vector, pCW Ori+ (Barnes *et al*., 1991). Unfortunately, we were unable to find suitable expression conditions for these constructs as well despite many attempts to optimize the conditions as we had done with the pET vectors.

We turned to two other approaches. First, we decided to use a plasmid-based metabolic engineering system previously developed by our collaborators. From study of the analogous GA biosynthesis pathways in *Arabidopsis thaliana* and *Fusarium fujikuroi* (formerly *Gibberella fujikuroi*) we hypothesized that *ent*-KA, *ent*-KAO, and a gibberellin aldehyde (either GA_{12}-aldehyde or GA_{14}-aldehyde) were the most likely substrates for the CYPs (Bömke and Tudzynski, 2009). Again, we were disappointed by the expression results from these plasmids, and unfortunately technical limitations prevented us from carrying out the metabolic engineering experiments at the University of Waterloo. We anticipate that our collaborators will be able to use these plasmids in their future work.

The other approach was to conduct *in vitro* assays with codon-optimized CYPs and ferredoxin that our collaborators had previously synthesized. Although we still did not see any sign of the

ferredoxin or CYPs on SDS-PAGE gels from CFEs of cultures harbouring the expression vectors, we attempted to record difference P450 difference spectra. The recordings we obtained appeared to show some features of cytochrome P450 and ferredoxin spectra, but they were internally inconsistent. We encountered several frustrating technical and mechanical issues that prevented exhaustive optimization of the expression conditions and creation of the CFE while on the research trip.

The purification initially appeared to be successful because of the colour of the eluent (Sielaff and Andreesen, 2005). Although our collaborators had previously purified the ferredoxin and confirmed our observations regarding the colour of the eluent, they had not analyzed an aliquot of their sample using SDS-PAGE. They did preform a similar spectrophotometric technique, but the results were not available for comparison. It is difficult to explain rationally the results of our purification. It is possible that the ferredoxin could be a tetramer or trimer which would explain the band sizes and colour we observed. Additionally, it is possible that the resin had been previously used without our knowledge, or the resin bed volume was too great which resulted in co-purification of extraneous protein. The experiment must be repeated, but because of time constraints with the visit it was not possible.

With encouragement from our collaborators, we decided to try the enzyme assays. We chose to test all three CYPs for their ability to transform *ent*-kaurene to *ent*-kaurenoic acid. Because we could not obtain high quality purified ferredoxin, we decided to use CFE from cultures expressing the ferredoxin and mix it with CFE of *E. coli* cultures expressing each CYP. Our collaborators had had previous success with similar assays for other CYPs. After incubating the CFEs with *ent*-KA and NADPH, the products were analyzed by GC/MS. One data set suggested that *ent*-kaurene was partially converted to *ent*-kaurenoic by CYP117, but we were unable to reproduce successfully the experiment while visiting our collaborators. Eventually, they were able to replicate our results using a similar system and are working towards confirming their findings. From our collaborators' analysis of the

CYP117's analog in *S. fredii* NGR234, CpxU, it appears that this enzyme is responsible for the conversion of *ent*-kaurene to *ent*-kaurenoic acid.

Section 2.13.2: Analysis of Deletion Mutants

The *B. japonicum* USDA110 deletion mutants strains that I constructed were critical to the work done by Dr. Cecila Rojas and her group in their ongoing experiments to determine directly the catalytic roles of CYP112 and CYP114 and indirectly the roles of CYP117 and the SDR. Our collaborators are continuing my work to create two more *B. japonicum* USDA110 deletion mutants which will be critical in confirming the roles of CYP117 and the SDR. They have also undertaken a project to delete the ferredoxin (Ryan Nett, personal communication).

Using previously reported methods, the Chilean research team led by Dr. Rojas was able to isolate gibberellin intermediates, from the bacteroids in soybean nodules, formed by the mutant USDA110 strains (Méndez *et al*., 2014). The identity of the products from the CYP112 mutant nodule extracts were confirmed by Dr. Peter Hedden at Rothamptom Research Station and the CYP114 extracts are currently under study by the same group. Although complete analysis of the bacteroid extracts has yet to be completed, the preliminary data (provided by Cecilia Rojas, personal communication) suggests that CYP112 is likely responsible for the conversion of GA_{12} to GA_9 while CYP114 is responsible for the conversion of *ent*-kaurenoic acid to GA_{12}-aldehyde. Conversion of *ent*-kaurenoic acid to GA products was normal in the Δ*blr2144* deletion strain; however, the inability of the *B. japonicum blr2145* deletion mutant to catalyze the same reactions (conversion of ent-kaurenoic acid to GA products) indicates that CYP114 is responsible for step in the biosynthesis pathway.

In all of the experiments, the final GA detected was GA_9. Our collaborators have speculated that GA_9 may actually be the final product of the GA biosynthesis pathway in *B. japonicum* USDA110

(Reuben Peters, personal communication). They believe that GA_9 is transported across the peribacteroid membrane and further transformed by soybean enzymes located outside of the nodule. This theory challenges all other reports on bacterial GA biosynthesis (reviewed in Chapter 1) with the exception of the most recent work on *B. japonicum* USDA110 by Dr. Rojas' research group (Méndez *et al*., 2014). Although we noted the shortcomings of the studies reporting on bacterial GA_1, GA_3 and GA_4, we are not convinced that they are completely misguided or incorrect. We present some thoughts to rebut the theory that GA_9 is the final GA end product in *B. japonicum* USDA110 and maintain that it is crucial to keep an open mind regarding bacterial GA biosynthesis.

The fungus *F. fujikuroi* has one major and at least two minor routes that result in the production of different GA products and all of the reactions are catalyzed by four versatile and multifunctional CYPs (Tudzynski, 2005). Currently we do not know when or why the alternative GA biosynthesis routes are relevant in fungal GA biosynthesis, but they could provide a layer of control on top of transcription and protein expression. Although little work has been done to tackle this question, what has been accomplished is intriguing and could possibly be extended to the bacterial GA pathway with further experiments. In *F. fujikuroi*, deletion of an NADPH cytochrome P450 oxidoreductase was found to be associated with two of the CYPs involved in GA biosynthesis: P450-1 and P450-2. These enzymes are responsible for the production of GA_{14} and GA_4, respectively (Troncoso *et al*., 2008). Interestingly, in the *F. fujikuroi* deletion mutant strain, the major products formed by both enzymes changed from GA_{14} and GA_4 to minor GAs (Troncoso *et al*., 2008). The researchers showed that the CYPs were able to adopt another redox partner, cytochrome b5:NADH cytochrome b5 reductase, and in the process converted from being NADPH-dependent to NADH-dependent enzymes (Troncoso *et al*., 2008). This could be important to the GA biosynthesis pathway in bacteria because of the environment, i. e. the nodule, that it occurs in. At this point we do not understand enough about the

biochemistry of the bacteroid to understand how the gibberellin biosynthesis pathway could be controlled or altered by the nodule environment. However, because the redox potential of the nodule changes over time throughout the lifespan of the nodule, first functioning in nitrogen fixation, before senescesing and finally decaying, it is possible that different gibberellins could be formed to correspond with redox states throughout the life span of the bacteroids, nodule and plant. Furthermore, although a cytochrome P450 reductase associated with the GA biosynthesis operon in bacteria has not yet been identified, presumably such an enzyme is involved in the process. Many bacterial CYPs known so far have an associated redox partner—either as a domain fused to the cytochrome itself or as a separate enzyme. However, the discovery that different biosynthesis reactions are catalyzed when different redox partners and electron transporters are paired with a CYP, contributes additional questions to the complex problem of substrate specificity of the GA CYPs in *B. japonicum* USDA110. We appreciate that GA_9 may have appeared to be the final product detected in the most recent work under certain conditions but we postulate that under different *in vitro* reaction conditions, and even perhaps under different biological conditions such as a different stage of nodule life, it may not be a final or major GA formed by *B. japonicum* USDA110. It is important that follow-up experiments are done to tease out the intricacies of the CYPs involved in bacterial GA biosynthesis like the studies that have been done with the fungal systems (please see Chapter 1 for a review of some of these).

At this time nothing is known about the enzyme mechanisms involved in bacterial GA biosynthesis. While it is extremely tempting to speculate about the final pathway products, role of the enzymes in the pathway, and even possible biological roles of the products detected in the wild-type and mutant *B. japonicum* USDA110 strains, there is simply not enough information to draw coherent conclusions. We cannot form a coherent hypothesis to explain the results obtained so far. Unfortunately, we must refrain from prematurely assigning roles to CYP112, CYP114, and even CYP117, though the

preliminary results with the deletion mutant strains do appear to directly identify the roles of CYP112 and CYP114.. More experiments are clearly required to reach a complete understanding of the phenomenon. The *B. japonicum* USDA110 deletion mutant strains and the corresponding deletion mutants in *S. fredii* NGR234 and *M. loti* MAFF303099 will be helpful in determining the roles of the enzymes in additional experiments.

Understanding the role of each CYP in this biosynthesis pathway could open up unknown pathways of research. Although the biological role of bacterial gibberellin is still unknown, once it is determined, it may become useful to be able to control GA production. Understanding which CYPs are involved with which specific steps, and their molecular mechanisms could make them targets for modulation of the pathway if necessary. Additionally, the gibberellin biosynthesis pathway and its enzymes could have as of yet unknown applications to the pharmaceutical and natural products industries—a better understanding of the enzymes and mechanisms available in nature could one day inform better drug solutions for humans as well. Finally, the importance of nodulating bacteria to global agriculture simply cannot be understated. A deeper understanding of the partners in this relationship could help inform and develop new tools to enhance and harness the power of the symbiosis. Although our experiments heterologously expressing the CYPs yielded negative results, it was still a valuable contribution to the overall project. Obviously, there will be many more discoveries made regarding bacterial GA biosynthesis and we look forward to following the progress that will be made on the topic in the next few years.

Section 2.14: Contributions to Knowledge

- I provided *E. coli* expression strains and deletion mutants to the Peters lab at Iowa State University in Ames, IA. Most notably, I provided *B. japonicum* USDA110 strains where *blr2144* and *blr2145* have been completely removed from the chromosome. I also provided the Peters lab with *S. fredii* NGR234 strains where *cpxP, cpxR* and *cpxU* have been deleted and also an *M. loti* MAFF303099 *mlr6367* deletion mutant strain. These were forwarded to Dr. Cecilia Rojas at the University of Chile. Dr. Rojas is collaborating with Dr. Peter Hedden to study the *B. japonicum* USDA110 mutants. In the future, the other mutant strains may be studied too.
- I expressed and purified N-terminally His-tagged ferredoxin. The gene was cloned from *B. japonicum* USDA110 ferredoxin and put into the expression vector pET28b(+). The heterologous protein was purified using a cobalt resin.
- I attempted to express and detect the CYPs using SDS-PAGE. I tried to record their spectra and determine their substrates by analyzing enzyme assay products using GC/MS.
- New expression constructs have been made to allow co-expression of the *B. japonicum* USDA110 ferredoxin and CYPs in concert with the metabolic engineering plasmids pGGeFF and pIRS. These have not yet been tested.
- The full-length CYPs were amplified from the *B. japonicum* USDA110 genome and cloned into pCW (Ori+) without His-tags for expression trials in six *E. coli* strains, though the results have not yet demonstrated high expression of the corresponding proteins.

Chapter 3

In Silico Analysis of the Bacterial Gibberellin Biosynthesis Operon and Surrounding DNA

Insights into the genomic context of the genes. This chapter includes a look at the putative gibberellin biosynthesis operon in several bacterial species including *B. japonicum* USDA110 and various nodulating microsymbionts. Some strains of *Sinorhizobium meliloti* and *S. fredii* as well as in two plant pathogens are also examined.

Section 3.1: Chapter Summary

The substrate specificity of two *Bradyrhizobium japonicum* USDA110 cytochrome P450 proteins involved in gibberellin (GA) biosynthesis was recently elucidated by the research group of Dr. Cecilia Rojas at Universidad de Chile in Santiago, Chile. Their laboratory achieved this by analyzing the metabolites in bacteroids formed by corresponding *B. japonicum* USDA110 deletion mutant strains described in Chapter 2 of this thesis. Once the capacity for production of bacterial GAs was confirmed by these data, we began thinking about the origin of the bacterial gibberellin biosynthesis genes. Once again, the question of the biological role of bacterial GA arose.

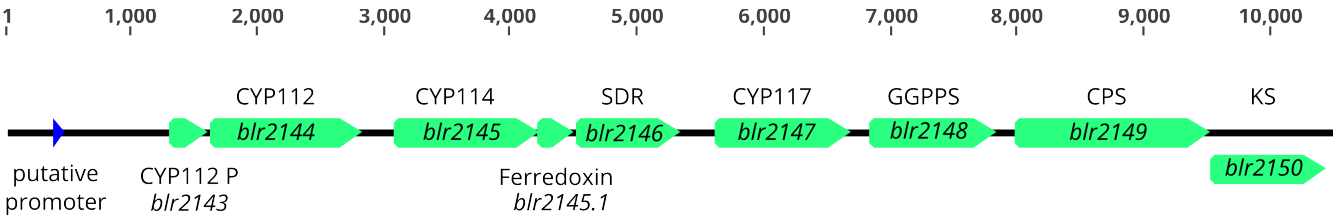
We attempted to form hypotheses to explain the presence of the operon as well as the biological role of bacterial GA. In pursuit of common ancestors of the *B. japonicum* USDA110 gibberellin genes, we searched for and found a number of analogous operons in published whole genome sequences deposited in sequence databases. These matches were from a diverse range of Proteobacteria genera including *Sinorhizobium*, *Xanthomonas*, and *Erwinia*. We examined the DNA flanking the putative GA operons in some of the identified α-and β proteobacteria. We found open reading frames that were predicted to encode transposases, integrases, and DNA repair enzymes. We noted that the %GC was different between the GA operon genes and the immediate genomic context. We also found putative gibberellin operons in eight *Sinorhizobium meliloti* genomes and performed a similar survey. The S. *meliloti* matches are interesting because they were recovered from indeterminate nodules—until now, all other nodulating endosymbionts with GA genes have been isolated from determinate nodules.

In all organisms, we found that the GC content of the GA biosynthesis operon was higher than that of the immediately surrounding genomic regions, and that in many genomes the annotated genes within 10 kb on either side of the operon are different in different organisms, suggesting that the GA operons may be a product of horizontal gene transfer rather than genetic inheritance.

In *B. japonicum* USDA110, the operon is found within the 400 kb symbiotic island which confers the ability to nodulate soybean roots and fix atmospheric nitrogen; similarly, the operon is found in the symbiotic island of *M. loti* MAFF303099. In *Sinorhizobium fredii* NGR234, the operon is found on the symbiotic plasmid. In the Rhizobiales we focused on with this work—USDA110, NGR234, and MAFF303099—the gibberellin operon has been shown to be part of the NifA-RegSR and FixLJ-$FixK_2$-$FixK_1$ master symbiosis regulons. Transcriptomic studies that have examined this relationship in these three strains are reviewed in Chapter 1.

Section 3.2: Introduction and Rationale

Gibberellins are, along with ethylene, cytokinin, auxin, and abscisic acid, one of the major phytohormones that regulate growth and development of vascular plants. So far, 136 different gibberellins have been discovered. (For information about their structure and biosynthesis, please see Chapter 1.) The majority of the gibberellins are biologically inactive and are either intermediate metabolites, inactivated versions, or synthetic derivatives (Rodrigues *et al*., 2011). Over fifty years ago, pioneering work with *Azotobacter* spp. demonstrated that soil-dwelling microbes are able to synthesize gibberellin (Vančura, 1961). For the past 50 years, gibberellins have been detected in various bacterial culture exudates but the bacterial genes and proteins involved in the biosynthesis are only just being understood. A diagram of the USDA110 operon is shown in **Figure 3.1**.



Bradyrhizobium japonicum USDA110
putative GA biosynthesis operon

Figure 3.1: The eight-gene putative GA biosynthesis operon in *B. japonicum* USDA110.

The authors of the first report on the GA biosynthesis operon in *B. japonicum* USDA110 isolated two CYPs and knocked out the function of one corresponding gene (Tully and Keister, 1993). To do this, the authors extracted 350 g of bacteroids from 2.5 kg of soybean nodules and then purified native CYP112 and CYP114 from the bacteroids. Polyclonal antibodies were raised against both CYPs and using the antibodies, western blots were performed to screen *B. japonicum* culture for CYP112 and CYP114 expression. Free-living aerobic and anaerobic cultures as well as wild-type bacteroid extracts were screened; results showed that expression of the CYPs was highest in the bacteroids, detectable in the anaerobic broth culture, and undetectable in the bacteria grown under aerobic conditions. To find the genes encoding the CYPs, a lambda gt11 cosmid library of *Bradyrhizobium japonicum* USDA110 was expressed in *Escherichia coli* and screened with additional western blots. Clones expressing CYP112 and CYP114 were found. DNA fragments corresponding to each gene were sub-cloned from the expression vector into pUC18 for further characterization. Targeted disruption of the *blr2144* gene (corresponding to CYP112) in *B. japonicum* USDA110 was achieved with a Tn5 transposon insertion into the gene. Subsequent soybean plant tests failed to find a symbiotic phenotype associated with the mutant strain, as the plants were capable of nodulating, and their nodules were functional (Tully and Keister, 1993).

Later on, the same group sequenced the operon and found that it had three cytochrome P450 genes: *blr2144*, *blr2145*, and *blr2147* (Tully *et al*., 1998). These encode CYP112, CYP114, and CYP117, respectively. The authors hypothesized that the *blr2143* locus is a non-functional pseudo-cytochrome P450 gene fragment. The operon, depicted in **Figure 3.1**, has five other open reading frames with sequence similarity to other known proteins (Tully *et al*., 1998). The locus annotated as *blr2145.1* was found to share similarity with a [3Fe-4S] ferredoxin from *Streptomyces griseolus*, and the *blr2146* gene was predicted to encode a short-chain dehydrogenase/reductase (SDR) protein based

on conserved domain analysis (Tully *et al*., 1998). The *blr2148* gene is assumed to encode a GGPPS protein based on sequence homology to *Rhodobacter* GGPPS genes and experimental evidence from biochemical analysis of homologous proteins in the related rhizobia *Mesorhizobium loti* MAFF303099 and *Sinorhizobium fredii* NGR234 (Hershey *et al*., 2014). The *blr2149* and *blr2150* genes, respectively predicted to encode Class II and Class I terpene cyclase proteins, have been shown to be responsible for conversion of GGPP to *ent*-kaurene in recombinant strains (Morrone *et al*., 2009).

After comparing the putative GA biosynthesis operons in *B. japonicum* USDA110, *M. loti* MAFF303099 and *S. fredii* NGR234, we hypothesized that the gene order would most likely be well conserved among analogous GA operons, but we knew to expect single nucleotide and amino acid substitutions as well as ORF fusions, gains and even losses. For example, *M loti* MAFF303099 has two additional genes with unknown functions at the 3' end of the operon that are not found in *B. japonicum* USDA110 or *S. fredii* NGR234. With this in mind, we decided to search sequence databases for additional bacterial GA operons since so many (20+) reports of bacterial gibberellin biosynthesis have been published, but little genetic and biochemical characterization of the pathway has been completed.

Results

Section 3.3: Comparative Genomics: *In Silico* Search for Bacterial GA Operons

Since the bacterial GA operons' initial discovery, exponential developments in high-throughput sequencing technology have allowed for the rapid accumulation of high quality bacterial genome sequences. We searched the NCBI sequence database for evidence of putative bacterial GA biosynthesis gene clusters based on sequence homology to *B. japonicum* USDA110 genes or *S. fredii* NGR234 genes. We used those hits to perform iterative searches for more distantly-related genes. When possible, we searched the genomes of bacteria that had been reported to synthesize GA in the primary literature to look for analogous GA operons. We used the conserved domain tools from NCBI to find additional candidates operons (Marchler-Bauer *et al*., 2013).

Section 3.4: Querying Sequence Databases

Initially, we used NCBI's BLASTp algorithm to identify homologs of the translated *B. japonicum* USDA110 gene *blr2144* in other organisms using the NR database (Altschul *et al*., 1997). We found several matches with CYP analogs from a diverse variety of proteobacteria including *Burkholderia*, *Xanthomonas* and *Erwinia.* The top hits were from *Rhizobium*, *Sinorhizobium*, and *Mesorhizobium* species. We also inspected the genome sequences of some recently sequenced *Rhizobium* and *Mesorhizobium* species using the Nucleotide database on NCBI and found hypothetical GA biosynthesis operons with ORFs that appeared functionally similar to the USDA110 GA operon. We downloaded the sequence data and used Geneious to compare the operons and the surrounding genetic context (Geneious version 6.0.6. Available from www.biomatters.com). **Figure 3.2** presents the operons in detail and shows their similar structure and gene order.

Using similar search strategies, we used the *B. japonicum* USDA110 Blr2149 and Blr2150 protein sequences to perform a tBLASTx search of 48 new *Sinorhizobium* draft genomes published by Sugawara and colleagues in 2013 (Sugawara *et al*., 2013). The draft genome sequences are stored on the Genoscope website (Vallenet *et al.*, 2013). We chose these two proteins because they work sequentially to perform the first committed step in bacterial GA biosynthesis: the conversion of geranylgeranyl pyrophosphate to *ent*-kaurene (Morrone et al., 2009). And from empirical data, we knew that these were more highly conserved at the amino acid level than the CYPs. In total, we found 9 matches, 8 of which had complete GA operon analogs. Most matches were in various *Sinorhizobium* species, but four were found in *S. meliloti* strains. These are shown in **Figure 3.3**.

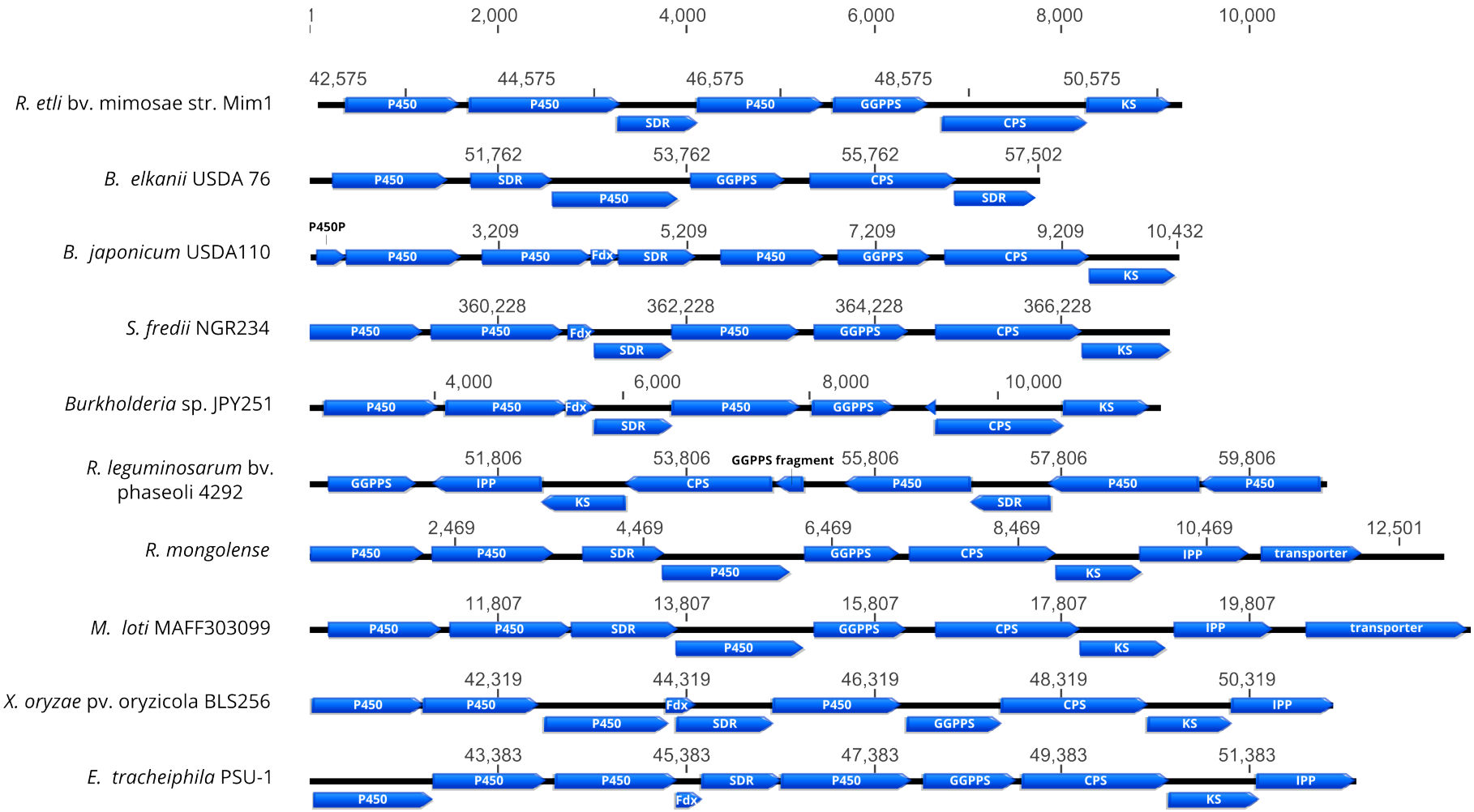


Figure 3.2: Gibberellin Biosynthesis Operons In Some Proteobacteria

A collection of gibberellin biosynthesis operons found through tBLASTx searching of the NR database with the *blr2144* query and examination of recently sequenced genomes. We noticed that some of the operons (notably those in *Xanthomonas* and *Erwinia*) had an extra P450 gene compared to the *B. japonicum* USDA110, *S. fredii* NGR234, and *M. loti* MAFF303099 operons. The 5' ends of the operon are on the left hand of the figure. Overall, the operons are very similar in terms of organization as well as ORF length.

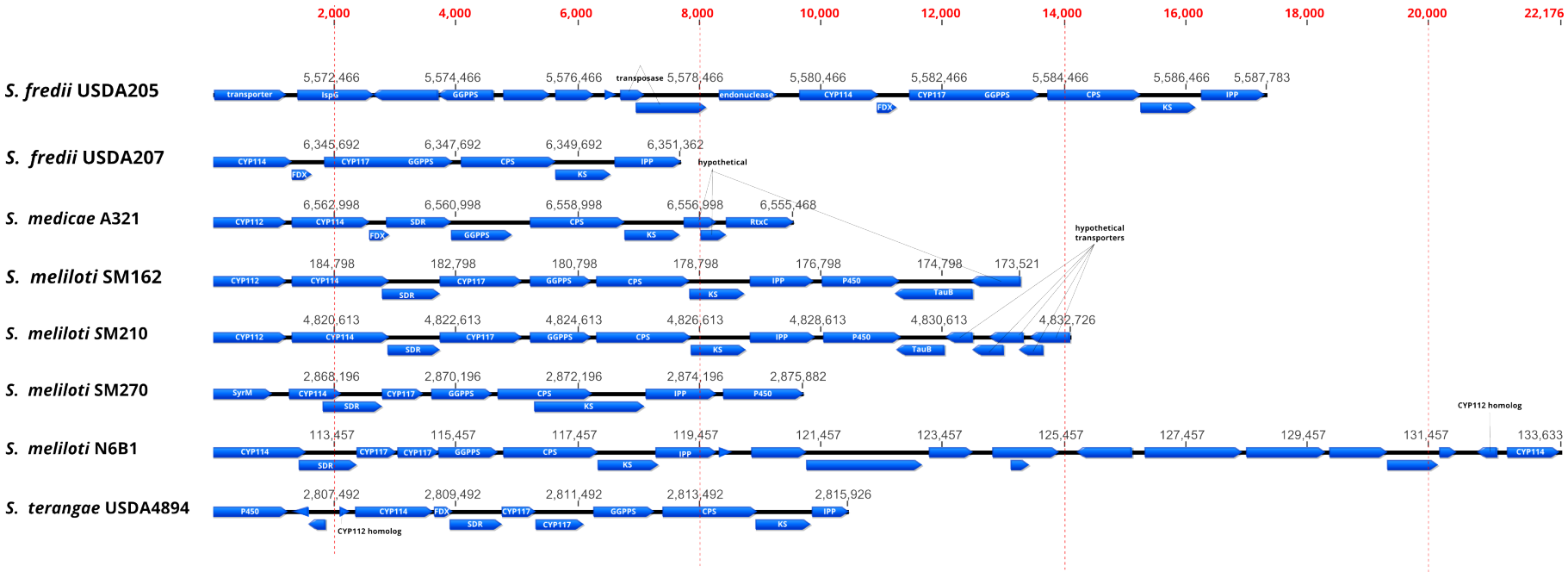


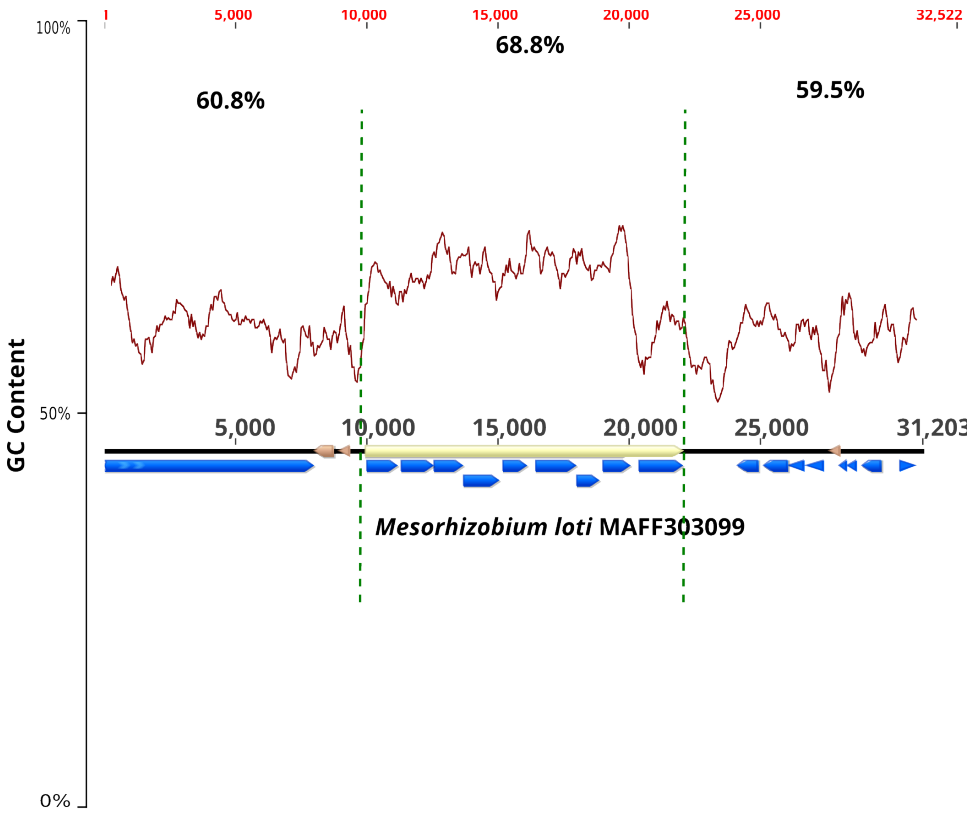
Figure 3.3: Organization of Putative GA Biosynthesis Operons in *Sinorhizobium* spp.

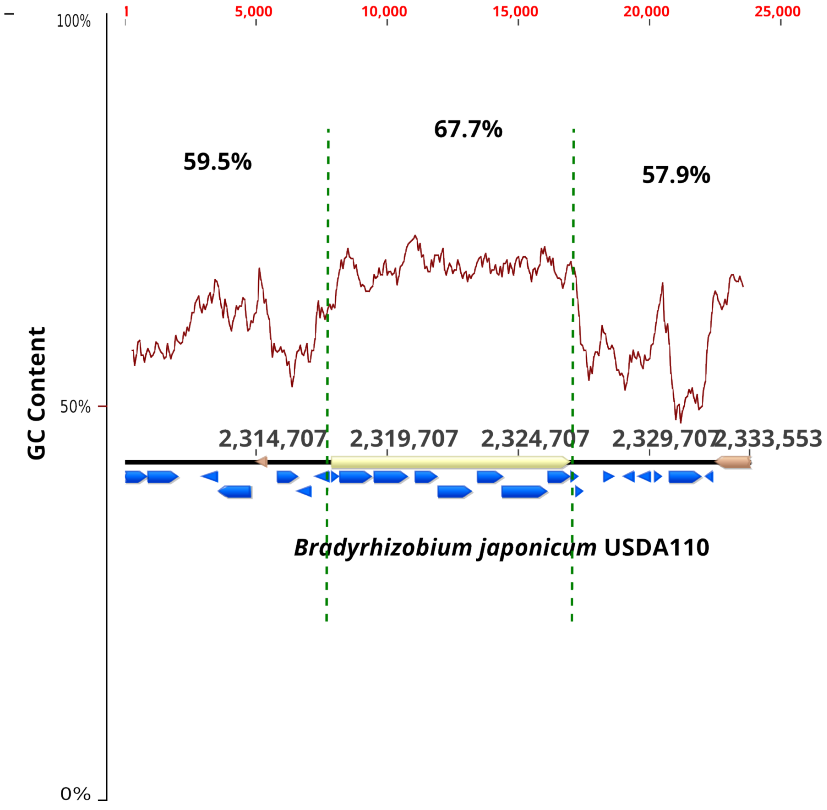
This figure shows the GA operons from 8 *Sinorhizobium* species isolated from determinate and indeterminate plant nodules. Each gene is shown and is annotated according to homology with other analogous GA operons. The length of each operon in base pairs can be determined by comparing tits relative length to the markers at the top of the figure.

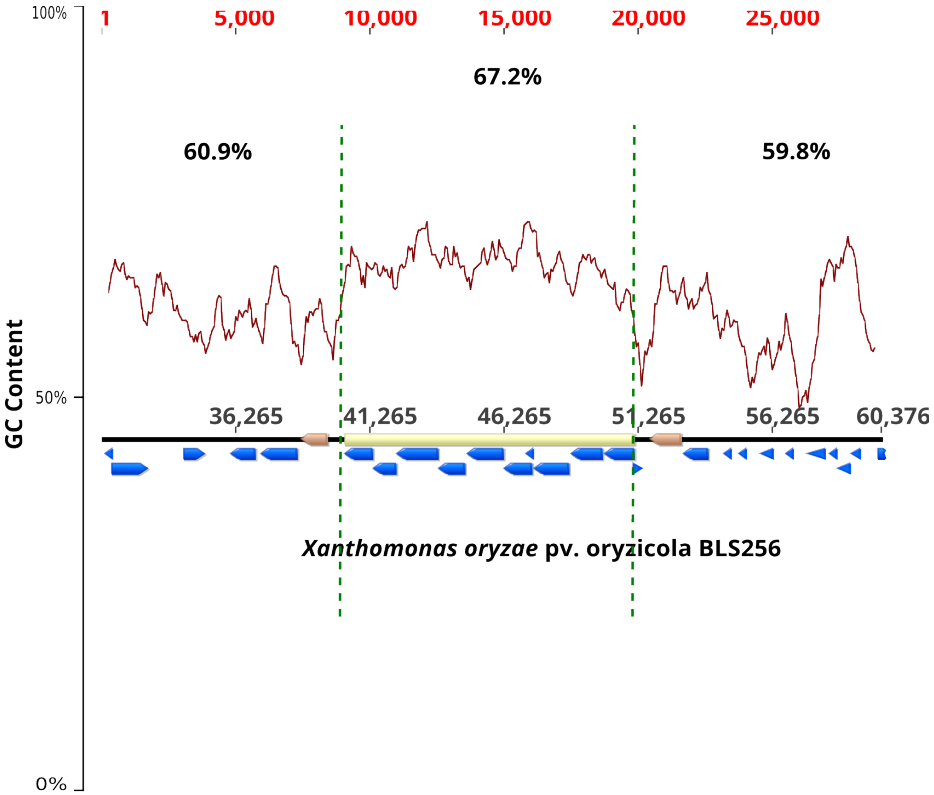
We searched the publicly available data from a large-scale sequencing survey of *Sinorhizobium* spp. microsymbionts isolated from legume root nodules (Sugawara *et al*., 2013). We found eight putative GA operons. Like those found in other Proteobacteria, these have a similar number of open reading frames organized in the same order. Notably, the *S. fredii* USDA207 operon appears shorter than the others, and appears to lack one of the CYPs and the SDR. It may not have the same functional capacity as the other bacterial GA biosynthesis operons. The operon in *S. meliloti* N6B1 is interesting because of the large insertion between the IPP gene and the CYP112 and CYP114 analogs at the 3' end of the operon. *S. meliloti* SM160 and *S. meliloti* SM270 appear to harbour one additional CYP compared to other GA operons found in rhizobia.

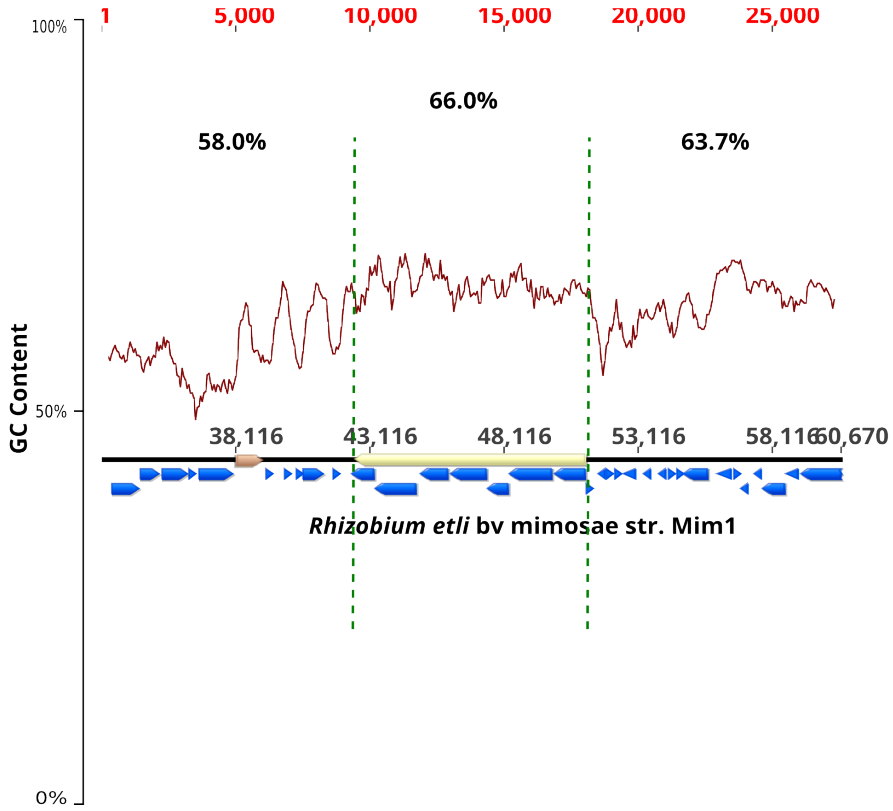
Using Geneious we calculated the %GC content of the putative GA operons and compared it to that of the surrounding DNA. To do this, we used windows of 500 base pairs and selected either 10 kb upstream or downstream of the operon along with the operon itself. The software calculated the % GC for each region and we recorded it. We created representative diagrams from the output and added annotations highlighting the % GC differences corresponding to the operon and flanking DNA. A clear GC content difference between the operon and the surrounding DNA is shown in **Figure 3.4.**We were able to make this comparison for all operons tested in **Figure 3.2** and **Figure 3.3** except for *Burkholderia* sp. JPY 251 and *R. mongolense* where incomplete contig assemblies made it impossible to compare the flanking DNA.

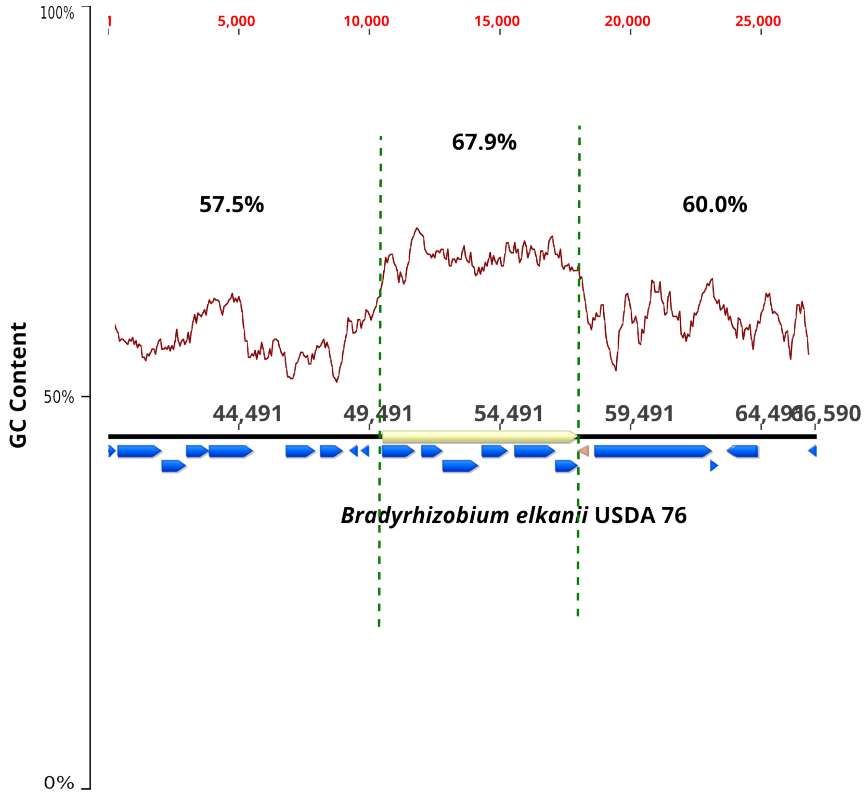
We observed tRNA genes and regions annotated as frame shift mutations near some of the GA operons we examined (data not shown). We were surprised to see that some of the ORFs flanking the GA operons were annotated as transposons, integrases or other DNA repair proteins. Although some of the genomes (including *B. japonicum* USDA110) did not have any transposon-related annotations near the putative GA operon, translated DNA queries of the DNA directly flanking the GA operon did result in matches to transposon-related genes (data not shown). We also noted that the genes closest to the putative GA operons in various species were not the same, which meant that their location in various genome was most likely not conserved. Taking these points together, we hypothesize that the gibberellin operon in Proteobacteria represents a mobile gene cluster; the operon may be horizontally transferred between microorganisms that live in, around, and on plants.

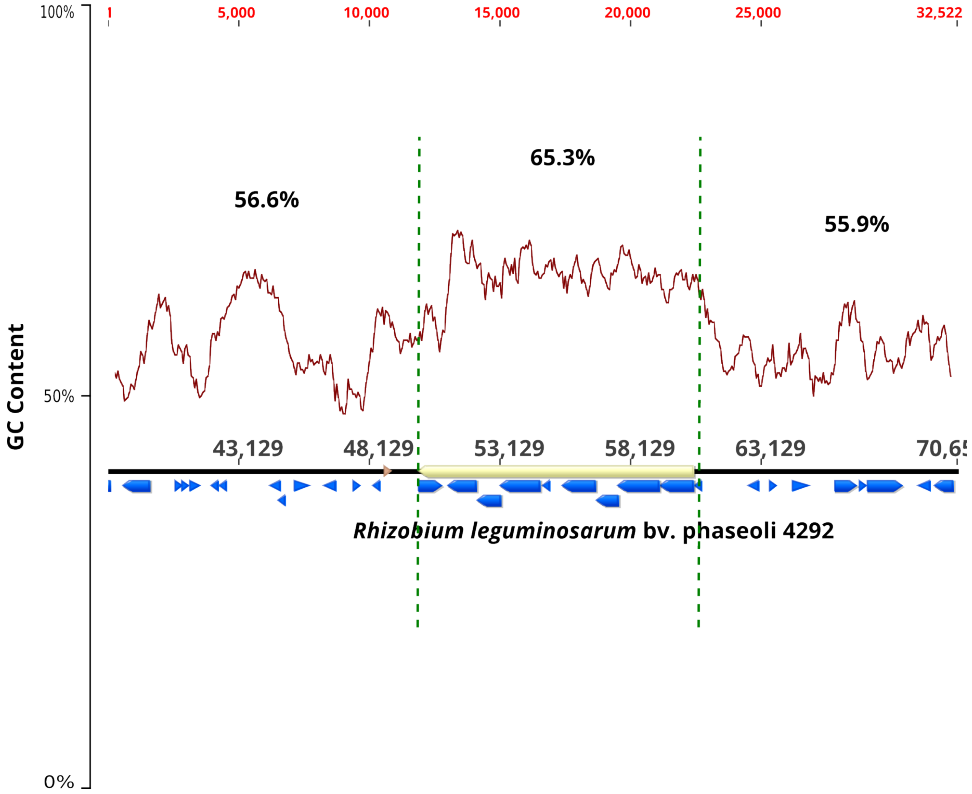

1
5,000
10,000
15,000
20,000
25,000
32,522
100%
50%
0%
GC Content
68.8%
60.8%
59.5%
5,000
10,000
15,000
20,000
25,000
31,203
Mesorhizobium loti MAFF303099

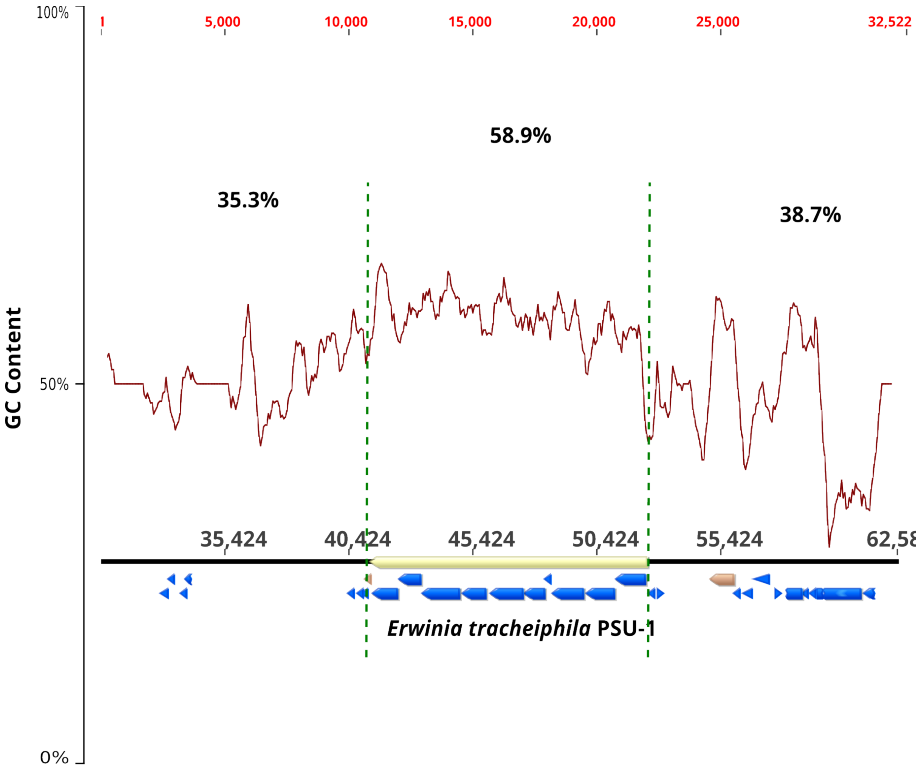

1
5,000
10,000
15,000
20,000
25,000
100%
50%
0%
GC Content
67.7%
59.5%
57.9%
2,314,707
2,319,707
2,324,707
2,329,707
2,333,553
Bradyrhizobium japonicum USDA110


100%
50%
0%
GC Content
1
5,000
10,000
15,000
20,000
25,000
67.2%
60.9%
59.8%
36,265
41,265
46,265
51,265
56,265
60,376
Xanthomonas oryzae pv. oryzicola BLS256


100%
50%
0%
GC Content
1
5,000
10,000
15,000
20,000
25,000
66.0%
58.0%
63.7%
38,116
43,116
48,116
53,116
58,116
60,670
Rhizobium etli bv mimosae str. Mim1


100%
5,000
10,000
15,000
20,000
25,000
67.9%
57.5%
60.0%
GC Content
50%
44,491
49,491
54,491
59,491
64,491
66,590
0%
Bradyrhizobium elkanii USDA 76


100%
5,000
10,000
15,000
20,000
25,000
32,522
65.3%
56.6%
55.9%
GC Content
50%
43,129
48,129
53,129
58,129
63,129
70,65
0%
Rhizobium leguminosarum bv. phaseoli 4292


100%
1
5,000
10,000
15,000
20,000
25,000
32,522
58.9%
35.3%
38.7%
GC Content
50%
35,424
40,424
45,424
50,424
55,424
62,58
Erwinia tracheiphila PSU-1
0%

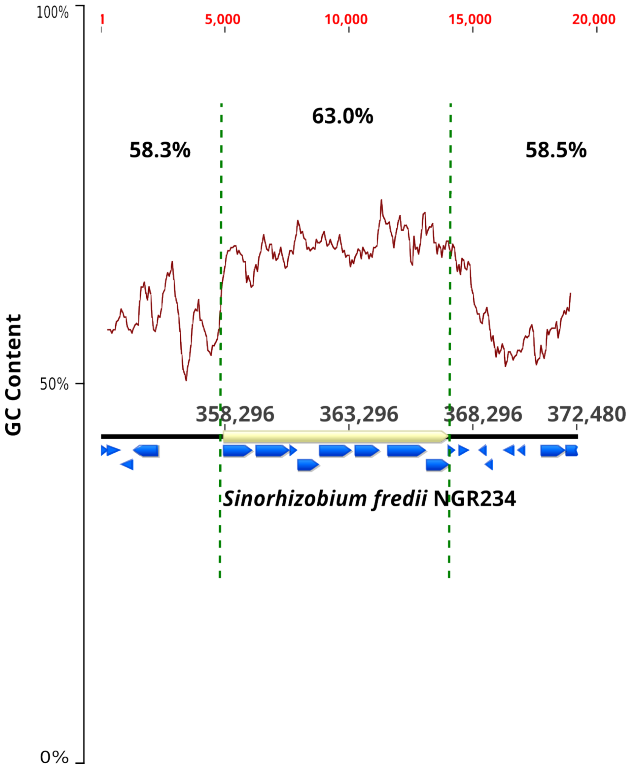

100%
1
5,000
10,000
15,000
20,000
63.0%
58.3%
58.5%
GC Content
50%
358,296
363,296
368,296
372,480
Sinorhizobium fredii NGR234
0%

Figure 3.4: % GC Content Comparison in Some Putative GA Operons
The % GC of the GA operon (marked with a yellow bar and delineated with the dashed lines in Figure 3.4) is higher than the surrounding DNA in all species that we examined. Using the algorithmic annotations, we scanned the surrounding 10 kb for transposon-related ORFs. The nearest transposition related ORF was tagged with an orange annotation. In some strains, particularly the genomes that were sequenced and annotated over a decade ago (namely B. *japonicum* USDA110 and *M. loti* MAFF303099), the annotations were not as robust. We were unable to find transposon-related ORFs without resorting to manual BLAST analysis of the flanking DNA. However, overall, this figure shows that the gibberellin genes have a higher GC content than the DNA found on either side. We were also able to find several examples of transposons flanking the putative GA operon. Perhaps the most interesting result was the matches to *Xanthomonas* and *Erwinia* because they are plant pathogens. The other species are either endophytes or nodulating bacteria.

Overall, the most striking example of differential GC content was in *Erwinia tracheiphila* PSU-1. In this strain, we observed a difference of over twenty percentage points between the gibberellin operon and its genomic surroundings. Another obvious difference was observed in *Xanthomonas oryzae* pv. *oryzicola* BLS256 which had very obvious flanking transposons as well as a marked difference in %GC content between the operon and nearby genome. Both the *Erwinia* and *Xanthomonas* operons have a gene encoding a putative isopentenyl pyrophosphate (IPP) isomerase at the 3' end of the operon; this gene is also found in the *M. loti* MAFF303099 putative GA biosynthesis operon but not in *S. fredii* NGR234 or *B. japonicum* USDA110. The IPP gene product is hypothesized to function in the conversion of DMAPP to IPP. This reaction is believed to drive formation of GGPP and could result in increased production of GAs (Reuben Peters, personal communication). The extra CYP found at the 5' end of both the *Erwinia* and *Xanthomonas* operons appears to be rather unique as well, although we found that *S. meliloti* SM162, *S. meliloti* SM21, and *S. meliloti* SM270 also have one more CYP compared to other putative GA operons found in Proteobacteria. At this time a role for the corresponding gene product has not been assigned. Overall, we noticed that the gene order was very well conserved among the different putative GA operons.

Finally, while the predicted transporter found at the 3' end of the operon in MAFF303099 appeared to be unique to *Mesorhizobium* species, TauB transporter genes were also found in *S. meliloti* SM162 and *S. meliloti* SM210. BLASTp searches of the NR database using the *M. loti* MAFF303099 transporter query only matched other *Mesorhizobium* strains that do not appear to have GA biosynthesis operons. The putative TauB transporters from the *S. meliloti* strains were not studied. The catalytic roles of the enzymes or GAs made by these organisms is not known at this time.

Section 3.5: The Location of The GA Operon in the Genome

The Sugawara *Sinorhizobium* genomes are still draft assemblies, so replicons have not yet been determined for each strain. One of the defining characteristics of *S. meliloti* is its three replicons: one single circular chromosome and two megaplasmids pSymA and pSymB (Capela *et al*., 2001). We expect that once the genome assemblies of the newly sequenced *Sinorhizobium* strains are completed, a similar genome architecture will become apparent for most, if not all, of the *S. meliloti* strains. The genome organization of *S. fredii* USDA205 and USDA207 should be similar to *S. fredii* NGR234, while *S. medicae* A321 should reflect the genome of type strain *S. medicae* WSM 419 which was recently reported (Reeve *et al*., 2010). Currently, *S. terangae* USDA4894 is somewhat of a phylogenetic orphan—the genome of its closest relative (determined by 16 S rRNA sequence analysis), *S. chiapanecum* ITTG S70, has not yet been sequenced (Rincón-Rosales *et al*., 2009). However even though the assemblies were incomplete and the replicons not known, we wanted to see if we could determine exactly where in the genomes of these *Sinorhizobium* spp. the putative GA operons were found.

We thought it might be possible to compare the location of the GA operon in the new strains to sequenced species that also have the gibberellin biosynthesis genes, namely *B. japonicum* USDA110, *M. loti* MAFF303099, and *S. fredii* NGR234. In these organisms the GA genes are found either on the chromosomal symbiotic island (*B. japonicum* USDA110 *and M. loti* MAFF303099) or on a symbiotic plasmid (pNGR234a in *S. fredii* NGR234). For reasons unclear at this time, *S. meliloti* Rm1021 does not have a putative gibberellin biosynthesis operon. In an attempt to map the *Sinorhizobium* GA operons to a symbiotic island or plasmid, we used BLAST to search for key symbiosis genes (*nodABC, nifAB, fixABC)* and mega plasmid maintenance genes (*repABC*) in *B. japonicum* USDA110, *M. loti* MAFF303099, and *S. fredii* NGR234. We constructed maps with the relative position of these genes in all three organisms, shown in **Figure 3.5**, **Figure 3.6**, and **Figure 3.7**.

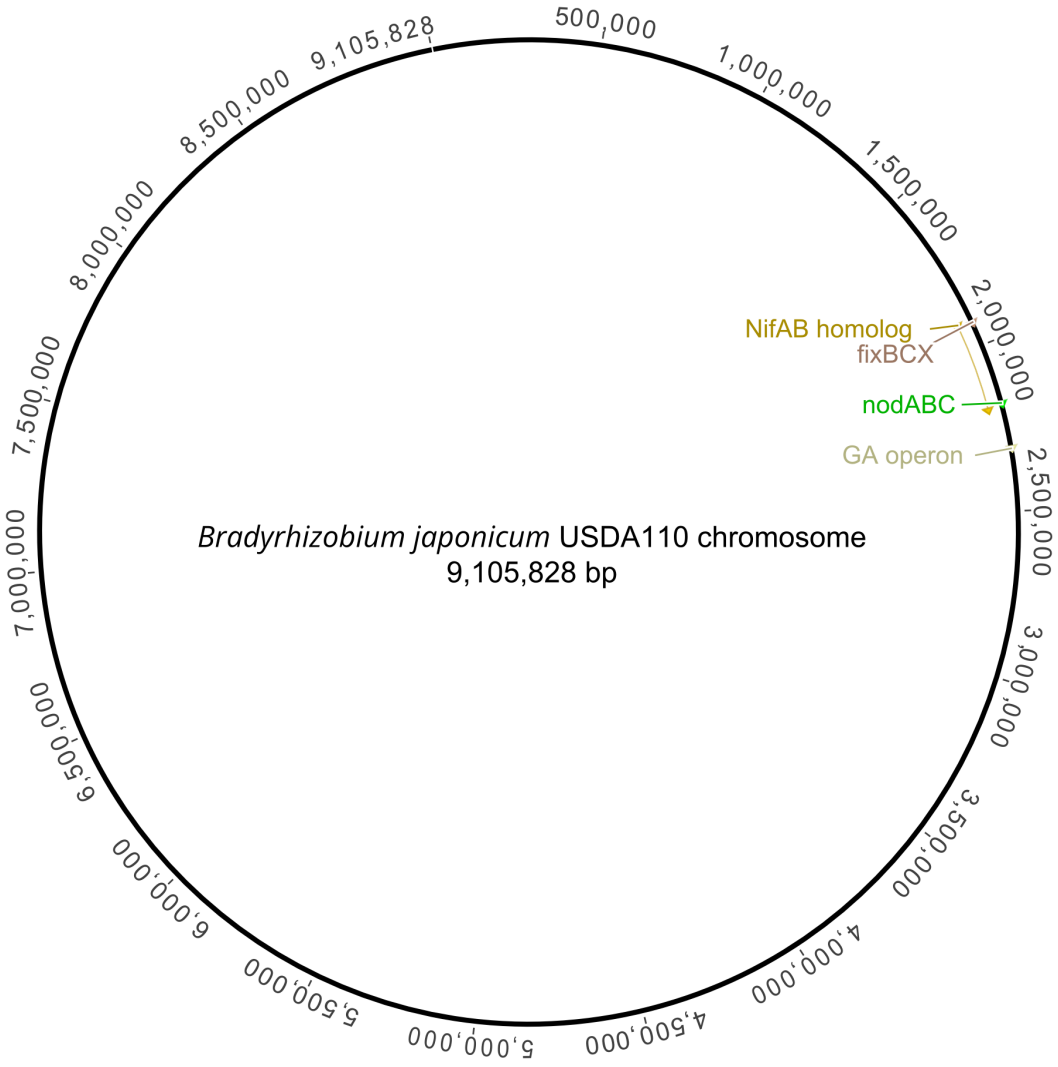


Figure 3.5: The *Bradyrhizobium japonicum* USDA110 genome.
Bradyrhizobium japonicum USDA110 has a single replicon of 9.10 Mbp (Kaneko *et al*., 2002). The gibberellin biosynthesis genes are found within the symbiotic island. The order and organization of the *nifAB*, *fixBCX* and *nodABC* genes in *B. japonicum* USDA110 appears to be similar to those in *M. loti* MAFF303099. In both organisms the symbiotic genes as well as the gibberellin operon are found on the chromosome. *B. japonicum* USDA110 does not have *repABC* homologs because it does not harbour a symbiotic plasmid.

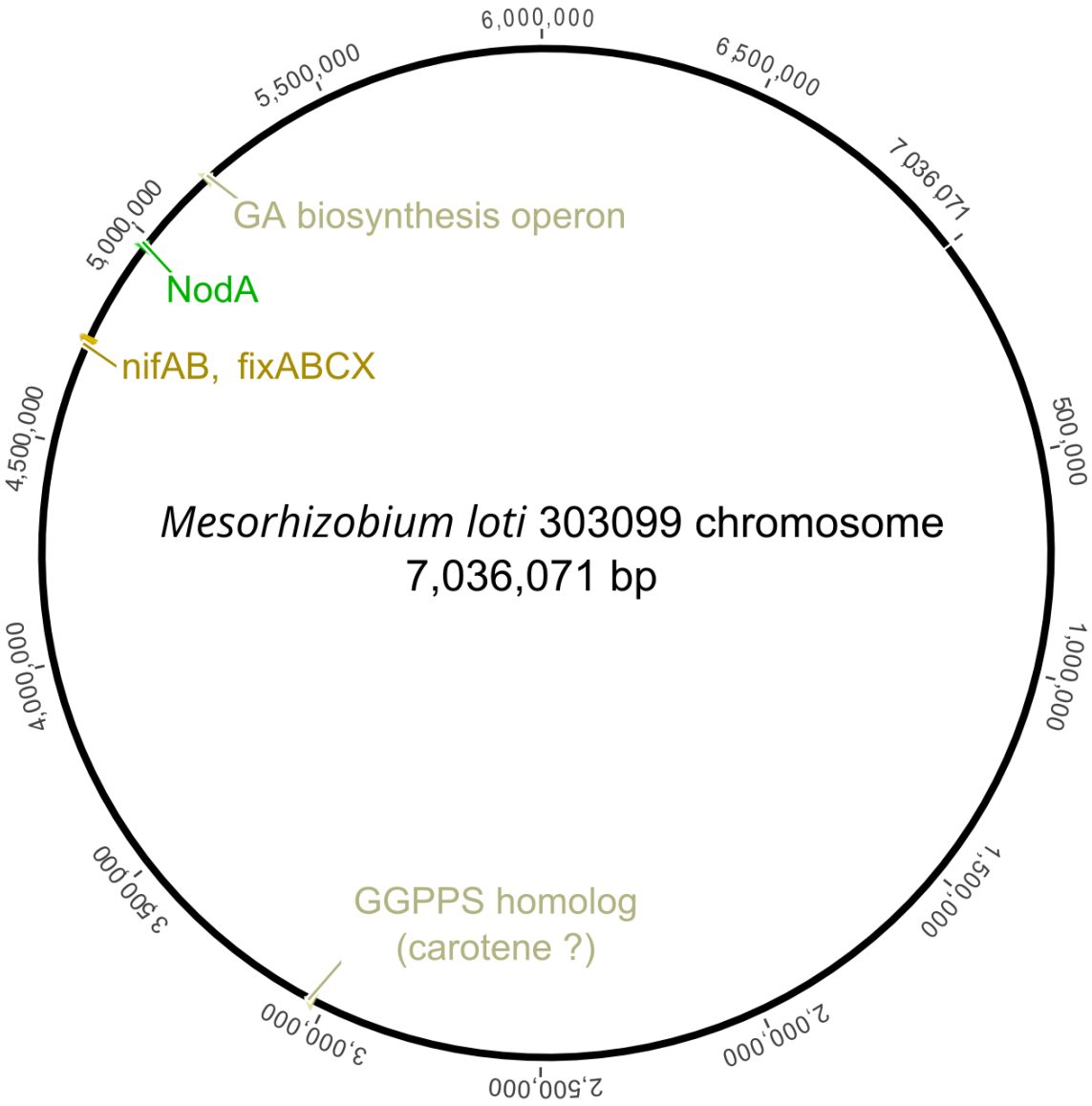


Figure 3.6: The *Mesorhizobium loti* MAFF303099 chromosome.
In *Mesorhizobium loti* MAFF303099, there is a large chromosome of 7.03 mega base pairs and two plasmids pMLa (351,911 bp) and pMLb (208,315 bp). The symbiotic genes are found on the chromosome, in a symbiotic island (Kaneko *et al*., 2000; Uchiumi *et al*., 2004).

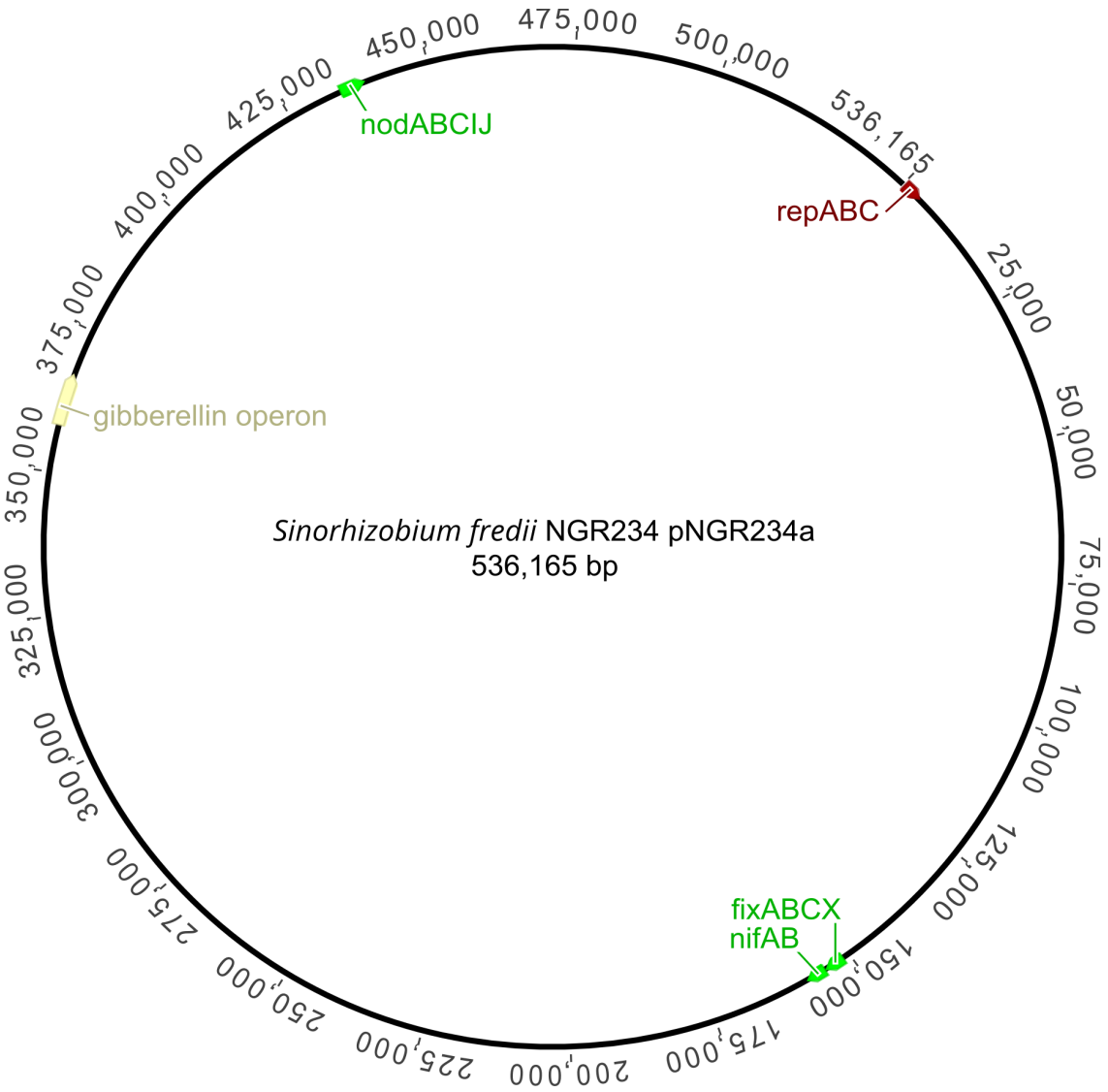


Figure 3.7: The symbiotic plasmid of *Sinorhizobium fredii* NGR234, pNGR234a.
Sinorhizobium fredii NGR234 has three replicons: the chromosome and two plasmids sized at 3.93 Mbp, 536,135 bp, and 2.43 Mbp respectively. The symbiotic plasmid, pNGR234a, harbours the symbiotic genes as well as the gibberellin operon. The location of the GA operon relative to the *nod*, *nif* and *fix* genes that we selected arbitrary marker genes is quite different in this organism compared to *B. japonicum* USDA110 and *M loti* MAFF303099.

In some of the *Sinorhizobium* spp. genomes, we found a putative IPP isomerase and one additional putative cytochrome P450 gene in addition to the previously described ORFs. In the majority of the *Sinorhizobium* operons, the fourth cytochrome was located at the 3' end of the operon, following the IPP isomerase. However, *S. terangae* USDA4894 had an additional CYP gene at the beginning of the operon, like the extra CYPs found in the putative *Erwinia* and *Xanthomonas* operons.

Using *S. meliloti* SM162's additional cytochrome as the query, we performed a BLASTp search of the NR database and hit upon conserved amino acid domains found in the cytochrome P450 responsible for oxidation of *ent*-kaurenoic acid plants. Using the other cytochromes does not result in the same match, making this result quite interesting, though we did not pursue it beyond noting that it is evidence that the genes are likely involved in gibberellin biosynthesis. A putative IPP isomerase homolog was found in all of the operons except *S. medicae* A321.

Like the genome of *B. japonicum* USDA110, and other genomes that we had previously examined, we noticed that there were transposons flanking the GA genes in most of the *Sinorhizobium* genomes. We used the RefSeq annotations as guides to find the nearest mobile elements flanking the GA genes and then used BLAST analysis to confirm the annotations by comparing them with similar transposons already categorized in the database. Transposon or integrase related ORFs were found nearby in all genomes except for *S. meliloti* SM210. We tabulated data in **Table 3.1** to enable quantitative comparisons between the different *Sinorhizobium* species and strains. Overall, the GA biosynthesis operon had a higher % GC content than the nearby DNA in all strains. We also found that in general, the DNA segment downstream of the GA operon had a lower % GC content than the DNA preceding the operon. **Figure 3.8** shows some representative operons, their % GC and immediate surrounding ORFs.

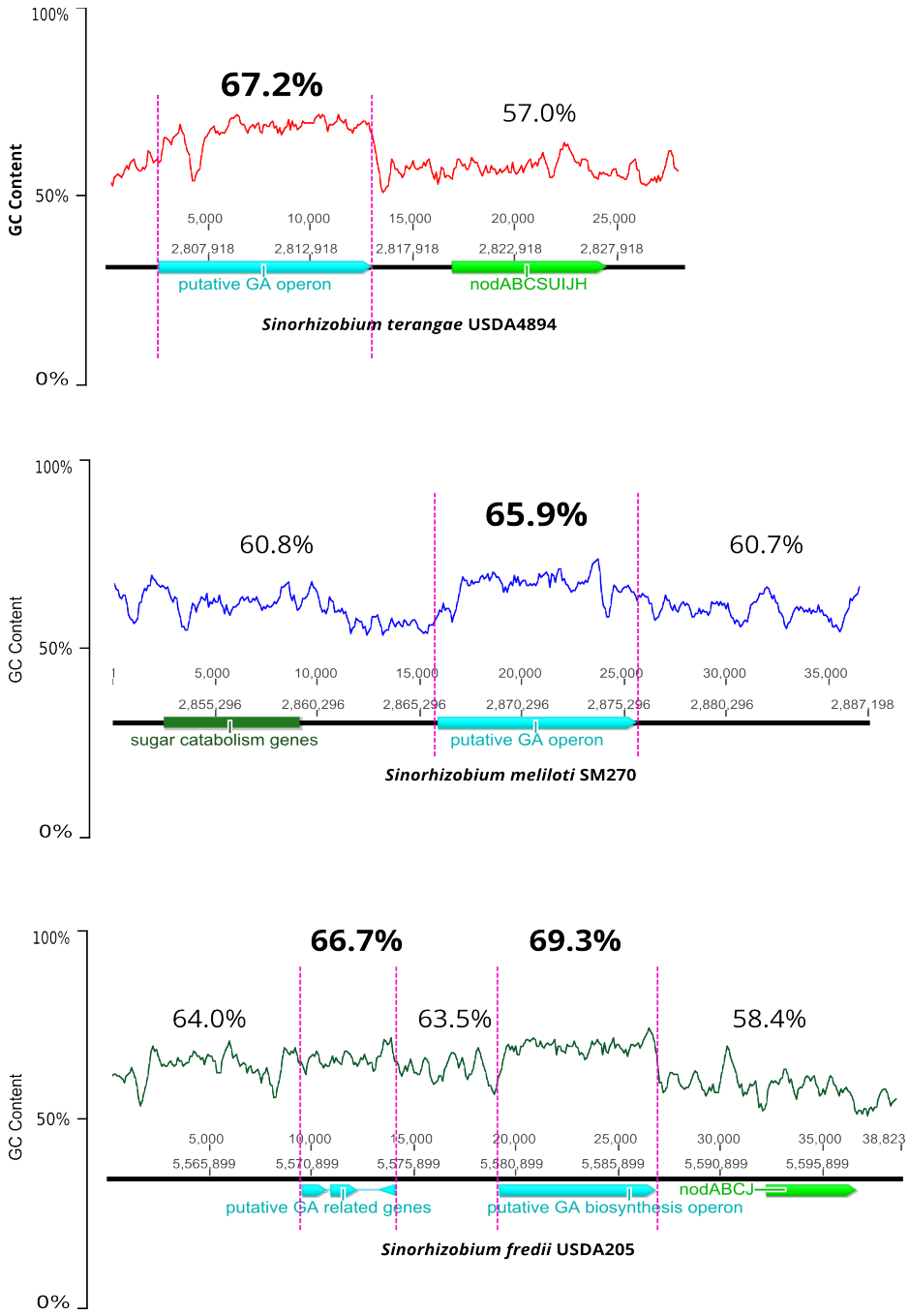


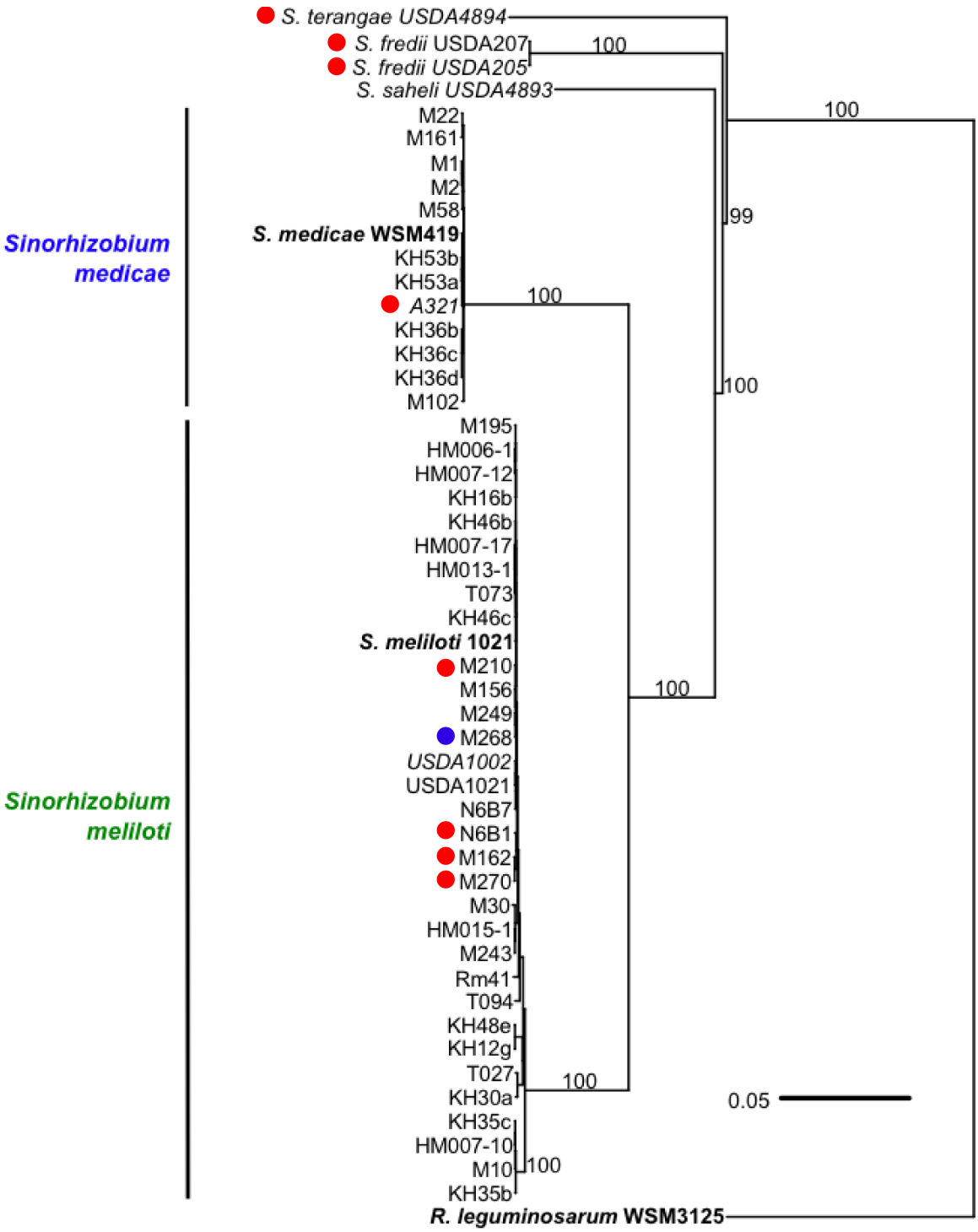
Figure 3.8: Comparison of GA Operons in Three *Sinorhizobium* Strains

This figure shows the % GC content of three of the *Sinorhizobium* spp. GA biosynthesis operons identified in the genome sequences produced by Sugawara *et al*., (2013). The putative GA operons are highlighted in light blue. The location of the reference genes selected to map the relative location of the GA operons on the chromosome are show in green. The *nod* gene homologs are show in lime green; two such matches are shown in *S. terangae* USDA4894 and *S. fredii* USDA205. In *S. meliloti* SM270, the selected *nod* gene homologs were not found nearby the GA operon. In this organism, sugar catabolism genes were found. These are shown in dark green. Genes that belong to the putative GA operons are marked by pink, vertical, hatched lines in addition to their blue colour. The % GC is represented on the vertical axis by the dark green line which corresponds to the individual genes found below it. Bold percentage points indicate the % GC of the putative GA biosynthesis genes.

Table 3.1: Features of the GA Operons in Eight *Sinorhizobium* Strains

Strain	5' Mobile Element	3' Mobile Element	Operon Length (bp)	upstream GC (%)	Operon GC (%)	downstream GC (%)
S. medicae A321	IS5	ISL3 transposon	14,499	57.6	62.4	57.1
S. meliloti N6B1	Mutator family transposase	IS240/6100 family transposase	8,276	56.6	61.6	58
S. meliloti SM162	DDE3 super family transposase	Not found	13,227	63.0	66.4	57.8
S. meliloti SM210	Not found	Not found	11,316	62.4	67.3	59.6
S. meliloti SM270	IS4 family transposase	IS66 transposase	9,689	59.8	66.6	57.9
S. fredii USDA205	IS4/IS5	Phage Xer family integrase	7,670	64.8	69.3	58.2
S. fredii USDA207	IS66 transposon	Phage Xer family integrase	7,670	60.7	69.3	59.3
S. terangae USDA4894	IS3	IS66 transposon	10,434	57.7	67.2	57.9

Phylogenetic trees showing the relationships of the Sugawara *Sinorhizobium* strains were available (Sugawara *et al*., 2013). We annotated the strains where a putative GA operon was found and show them in **Figure 3.9** and **Figure 3.10.** Each of the *Sinorhizobium meliloti* strains appear to be very closely related, however, not all of them harbour a putative GA operon. The grouping of strains is similar in both phylogenetic trees. There is not much difference except the relationship between the two *S. fredii* USDA205 and USDA207 strains which became more defined when the comparison was based on 16S rRNA instead of on the 645 protein coding genes. All of the other *S. meliloti* strains remain closely related, as expected. The strain marked with the blue dot, *S. meliloti* SM268 has a *blr2149* homolog gene, but it does not have any other gibberellin biosynthesis genes and so was ignored in the other analyses that we did.



- appears to have complete GA biosynthesis operon
- appears to have only *blr2149* analog

Figure 3.9: Phylogenetic tree of Sugawara *Sinorhizobium* strains made from alignment of 645 concatenated protein sequences with putative GA operons identified.

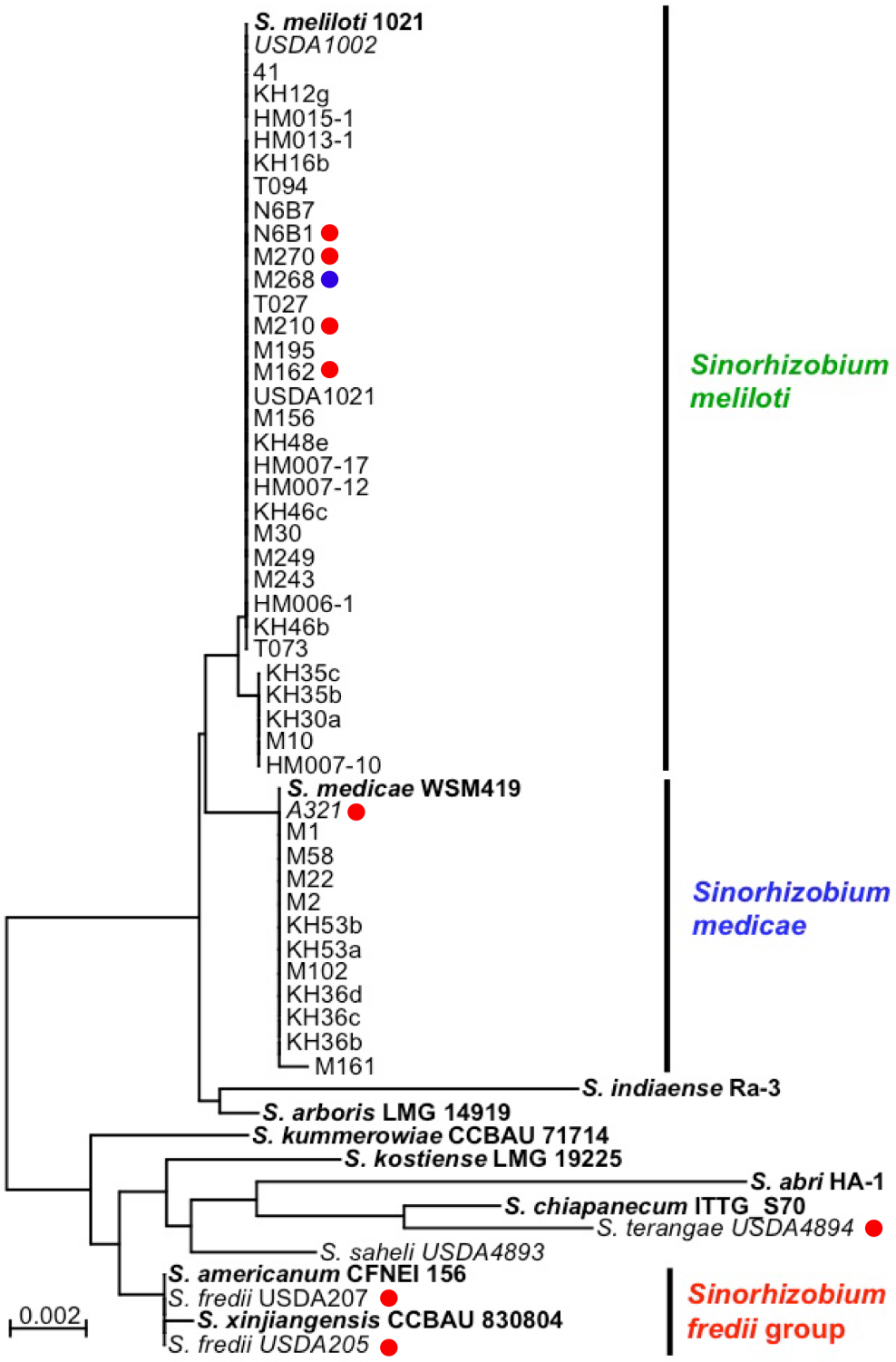


Figure 3.10: Phylogenetic tree based on the 16S rRNA gene sequences for the Sugawara *Sinorhizobium* strains with putative GA operons identified.

Section 3.6: tBLASTx Searches of Taxid 382 (*Sinorhizobium meliloti* group) on NCBI

Additional *Sinorhizobium meliloti* strains have been sequenced; the genomes are completely assembled and available on NCBI. So far, gap-less *S. meliloti* genomes for strains BL225c, SM11, AK83, GR4, 2011, and Rm41 exist. Gapped scaffold genomes of WSM4191, CIAM1775, RRI128, Mlalz-1, B021CC, AK58, and WSM1022 are also available. In light of finding gibberellin biosynthesis operons in other *S. meliloti* strains, we decided to check all of the available *S. meliloti* genome sequences for gibberellin biosynthesis genes. Using *blr2149* and *blr2150* as queries, we were unable to find any matches in the *Sinorhizobium meliloti* group. We decided to try *blr2147* (CYP117) since the enzyme catalyzes the next reaction in the pathway: the conversion of *ent*-kaurene to *ent*-kaurenoic acid. We did not find good matches with this query either. We searched for GGPPS homologs using the *blr2148* gene from USDA110 and we did get some matches. When we looked at the context of the hits, we saw that they were not located near cytochromes at all, but were found in close proximity to peptidoglycan and cell wall biosynthesis genes. It is unlikely that they are involved in GA biosynthesis. We searched for each of the remaining GA genes in the 382 *Sinorhizobium meliloti* taxid group.

The top 5 hits matching *blr2144* were in *S. meliloti* AK83 (max/total score 848/6162), *S. melilot*i SM11 plasmid pSmeSM11c (843/5142), *S. meliloti* BL225C plasmid pSINMEB01 (841/7730), *S. meliloti* GR4 plasmid pRmeGR4b (840/5282) and *S. meliloti* plasmid pHRC017 (718/5292). The query coverage for each was 99% and the E value was 0.0 for each match. The top five matching replicons were downloaded from NCBI and loaded into Geneious for comparative analysis and annotation. Two example matches are shown in **Figure 3.11.** We noticed that the cytochrome P450 homologs were found near transposon-related ORFs and had higher % GC than the surrounding DNA, like the other operons we had already examined. We did not look for tRNA genes or transposition related ORFs like we did with complete GA operons.

In all of the *Sinorhizobium* strains isolated by Sugawara *et al*., (2013), we observed a putative multi-copper oxidase gene directly following the 3' end of the *blr2145* homolog gene. We used tBLASTx to look for a homologous protein in the Sugawara *Sinorhizobium* strains, again using the Micro Scope tools hosted on the Genoscope website. We found two matches. The first was in *S. meliloti* N6B1, directly downstream of two GA-related cytochrome P450 genes, proximity located from the rest of the gibberellin biosynthesis genes. The score of the match was 1862, with an E value of 0. Query coverage was 98.87 percent. The other match was in *S. medicae* A321, but was not located near any of the GA biosynthesis genes; in this case, the score of the match was 1106, E value 0, and query coverage 93.97 percent.

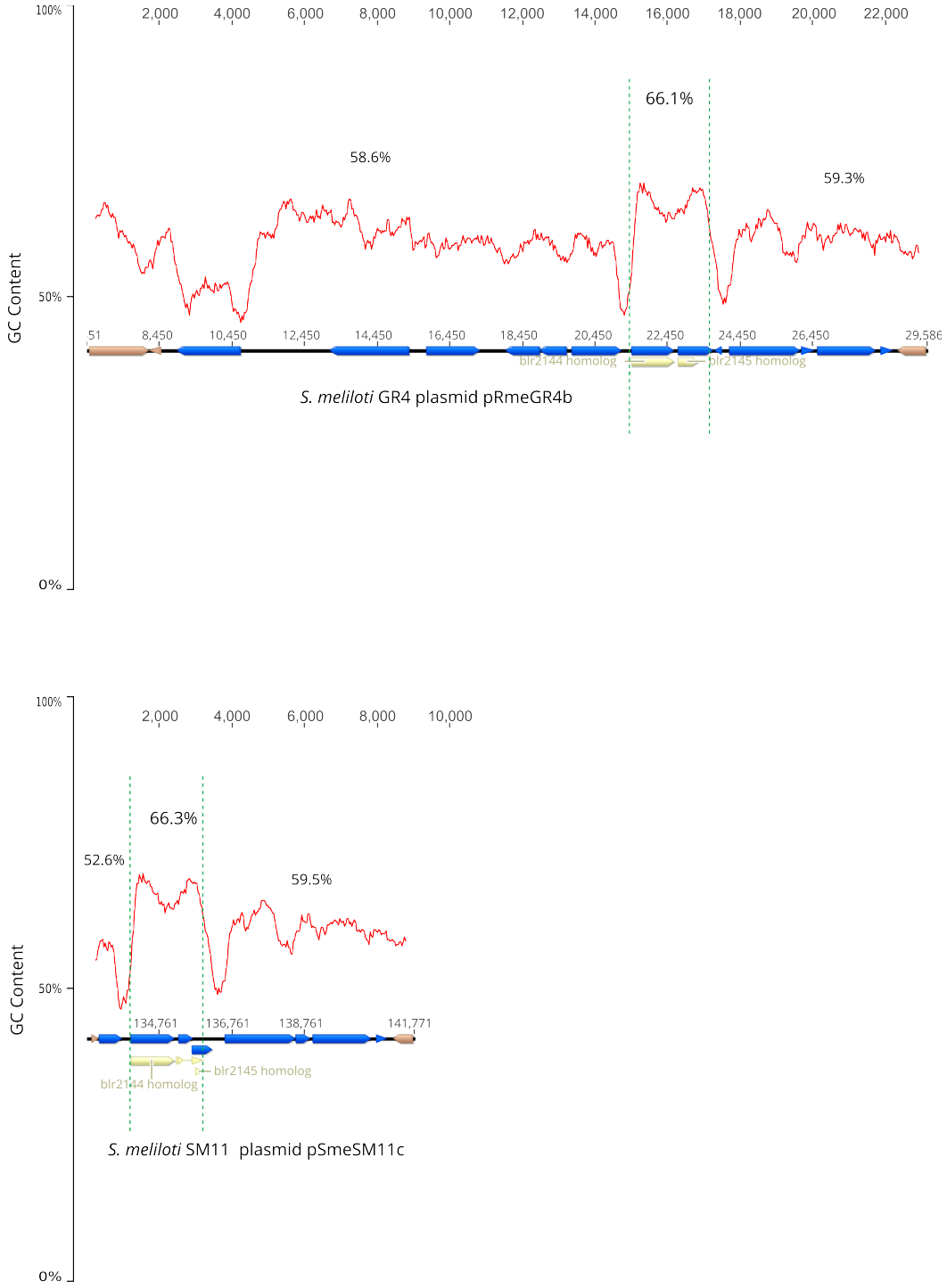


Figure 3.11: GC Content in The Region Near The *blr2144* and *blr2145* Homologs in two *Sinorhizobium* strains that do not have complete GA biosynthesis operons.

The yellow annotations show the homologous cytochrome genes. Orange open reading frames show the nearest transposon-related genes, found by searching the automated annotations. The % GC is noted above the ORFs. These strains do not appear to harbour any other genes related to gibberellin biosynthesis.

Section 3.7: Additional Gibberellin Biosynthesis Operons

In 2011, Tian and colleagues collected over 1,100 soybean symbionts from four main soil types in China (Tian *et al*., 2011). Their goal was sampling soybean nodule microsymbionts from diverse growing in diverse soil conditions. The authors focused on sampling where the soil is mainly dry and has a neutral oralkaline pH or where soil is humid and the pH is acidic (Tian *et al*., 2011). The specific sampling locations were selected based on their previous work that found soybean nodules in alkaline soil were dominated by *Sinorhizobium* spp. symbionts while those in acidic soil were predominantly colonized by *Bradyrhizobium* sp. (Man *et al*., 2008; Vinuesa *et al*., 2008; Han *et al*., 2009; Zhang *et al*, 2011). From their collection, 26 representative strains were selected for whole genome sequencing using an Illumina HiSeq2000 machine. The sequence data is available on NCBI under at project accession PRJNA77219.

In total, 9 *Sinorhizobium* and 18 *Bradyrhizobium* species were found based on 16 S gene analysis. As they predicted, the majority of the *Sinorhizobium* spp. symbionts were isolated from nodules grown in dry, or semi-arid alkaline soils found in either Xinjiang or Huang-Huai-Hai province (Tian *et al*., 2011; Supplementary Figure 1). We compared the *B. japonicum* USDA110 gibberellin biosynthesis open reading frames with the whole genome sequencing (WGS) project using the tBLASTn algorithm on NCBI. We found good matches for putative GA biosynthesis genes in 25 of the 26 strains where the E value was 0.0 using peptide sequence derived from the *blr2144* or *blr2150* genes in *B. japonicum* USDA110.

We chose *blr2144* and *blr2150* as the query sequences because they encode the first and last gene in the putative GA biosynthesis operon. Additionally, *blr2144* encodes a cytochrome P450 the sequence of which is well conserved in other α-proteobacteria that we studied (data not shown) and the reaction catalyzed by the *blr2150* gene product (formation of *ent*-kaurene) represents the first committed step in gibberellin biosynthesis. Only one strain, *B. yuanmingense* CCBAU 25021 did not

have a strong matches to either query; however, a hit did appear in the search results of *blr2144.* This match had an E value of 2e-33 but no hits were found using the *blr2150* query. We also noticed that most of the *Sinorhizobium* species had relatively worse matches to the *blr2144* amino acid sequence (E values greater than zero but less than 3e-23) compared to the *Bradyrhizobium* species, most of which matched with E values equal to zero. Interestingly, we noticed an opposite trend using the *blr2150* query where the majority of the *Sinorhizobium* species hit with E values of 0.0 while some of the *Bradyrhizobium* strains that had *blr2144* matches with perfect E values of 0.0 had similar E values for *blr2150*. At this time, the genomes are still in very rough draft phase and have not been assembled into complete contigs; furthermore gene, open reading frame, and coding sequence predictions are not available within the project accession files—this lack of information makes comparison of the operons themselves impossible at a level of detail compared to other operons examined in this work.

Section 3.8: The Geranylgeranyl Pyrophosphate Synthase

For several years, the annotation of *blr2148* in the *B. japonicum* USDA110 genome on NCBI has been troubling. The gene is annotated as an farnesyl pyrophosphate synthase (FPPS) instead of a geranylgeranyl pyrophosphate synthase (GGPPS) protein. Though both are similar proteins that catalyze similar reactions, an FPPS enzyme cannot make the GGPP precursor compound required for GA biosynthesis. It was important for us to distinguish between them for the sake of clarity. We knew from our literature search that there were conserved motifs and residues involved in determining the chain length synthesis capabilities of the prenyltransferases, the class of enzymes to which FPPS and GGPPS belong. Essentially, the number of carbon atoms in the molecule is what differentiates FPP from GGPP and its subunit IPP. We are unaware of any published reports stating that *blr2148* does encode a GGPPS so we wanted to find evidence to confirm this. Indirectly, showing that the prenyltransferase in the putative GA operon of *B. japonicum* USDA110 is a GGPPS would bolster the evidence that the pathway products are in fact gibberellins.

When we examined the MSA comparing the *blr2148* query with experimentally characterized or crystalized FPPS and GGPPS proteins, we found congruence of the query sequence with the previously reported conserved regions in GGPPS enzymes (Cantera *et al*., 2002). Importantly, the homology included the chain length determination region of the GGPPS genes (Ohnuma et *al.*, 1997). Previously, the authors had identified a specific position where FPPS enzymes reliably have a tyrosine (Y) residue while GGPPS enzymes have an alanine (A). The *B. japonicum* USDA110 bl*r2148 g*ene product has an A at that position. The MSA demonstrating this is shown in **Figure 3.12**. Interestingly, instead of a tyrosine or alanine residue at the chain length determination position in some GGPPS analogs, we observed a cysteine (C). At this time, we do not know the biological significance of this; additionally, the GA biosynthesis operon in organisms with this residue have not been studied.

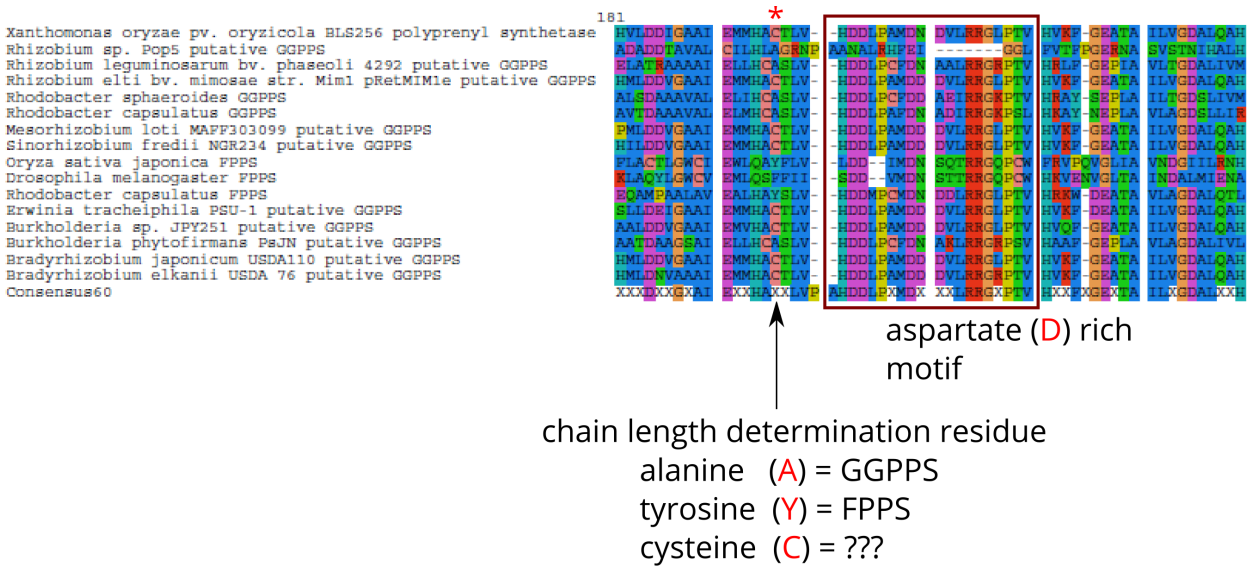


Figure 3.12: Amino Acid Alignments of FPPS and (putative) GGPPS genes.

Once the *blr2148* gene product was decidedly predicted to be a GGPPS rather than an FPPS, we attempted to use it derive an evolutionary hypothesis for the gibberellin biosynthesis genes in the *Bradyrhizobium* taxid (374) on NCBI. Following up on an early report that used the FPPS protein sequence as a phylogenetic marker, we queried the NR database with the translated amino acid sequence of the putative GGPPS homolog (the *blr2148* gene) from the USDA110 GA operon (Cantera *et al*., 2002). We selected 80 top hits with e values greater than 4e-04 which represented protein sequences annotated as polyprenyl synthases, FPPS and GGPPS proteins. We created a MSA with MUSCLE and a phylogenetic tree (data not shown) using the PhyML algorithm using Seaview 4.4.3 (Guoy *et al*., 2010). Four main clusters were formed. Based on the annotations of the genes, the groups appear to represent prenyl transferases with different biological functions. It appears that the other GGPPS gene products are involved in forming products of different carbon chain lengths or making GGPP as a precursor for different biosynthesis pathways.

Analysis by Ohnuma, Cantera and others has shown that GGPPS enzymes may have given rise to both farnesyl synthases and polyprenyl synthases which create prenyl chains of different lengths. FPPS enzymes form FPP, a 15-carbon alkene, from sequential condensation of two IPP units with one DMAPP molecule, while GGPPS condenses three IPP subunits with one DMAPP making the 20-carbon compound GGPP instead. Polyprenyltransferases create even longer final products by incorporating more of the 5-carbon IPP units. The biological fates of the FPP, GGPP and other compounds are different. For example, FPP is a precursor for cholesterol, steroids, and heme, while GGPP is a precursor for gibberellins, carotenoids, bacteriochlorophylls and ether-linked lipids in the cytoplasmic membrane of Archaea (Ohnuma *et al*., 1997). We looked specifically for GGPPS genes that were flanked by GGPPS or isoprenoid cyclase genes similar to the rhizobial GA operon organization.

The largest clade consisted of query hits labeled as polyprenyl transferases. The nearby genes did not appear to be cytochromes or isoprenoid cyclases. The second largest cluster appeared to have GGPPS genes near peptidoglycan biosynthesis genes and we could not guess at a biological function. The third cluster had GGPPS matches that appeared to be involved in carotenoid biosynthesis, specifically biosynthesis of spirilloxanthin, and bacteriochlorophyll. Cluster 2 had GGPPS genes near cytochromes and isoprenoid cyclase genes. These GGPPS genes appear to be part of putative GA operons. We were intrigued by the GGPPS genes in Cluster 2 and 3; however, we did not pursue characterization of the additional putative GA operons. We were most interested in the GGPPS genes found in cluster 3 which were involved in carotenoid and bacteriochlorophyll biosynthesis and if they were related in any way to the gibberellin biosynthesis genes in *B. japonicum* USDA110.

Carotenoid and bacteriochlorophyll biosynthesis has been studied in *Bradyrhizobium* sp. ORS278 (Hannibal *et al*., 2000; Giraud *et al*., 2004). We downloaded the genome sequence of ORS278 and annotated the spirilloxanthin and canthaxanthin biosynthesis operons, shown in **Figure 3.13**. Both biosynthesis operons have GGPPS genes, but we were especially intrigued by the spirilloxanthin biosynthesis operon because of its close proximity to the bacteriochlorophyll oxidoreductase genes. We thought that there could be a link between these genes and the gibberellin CYPs.

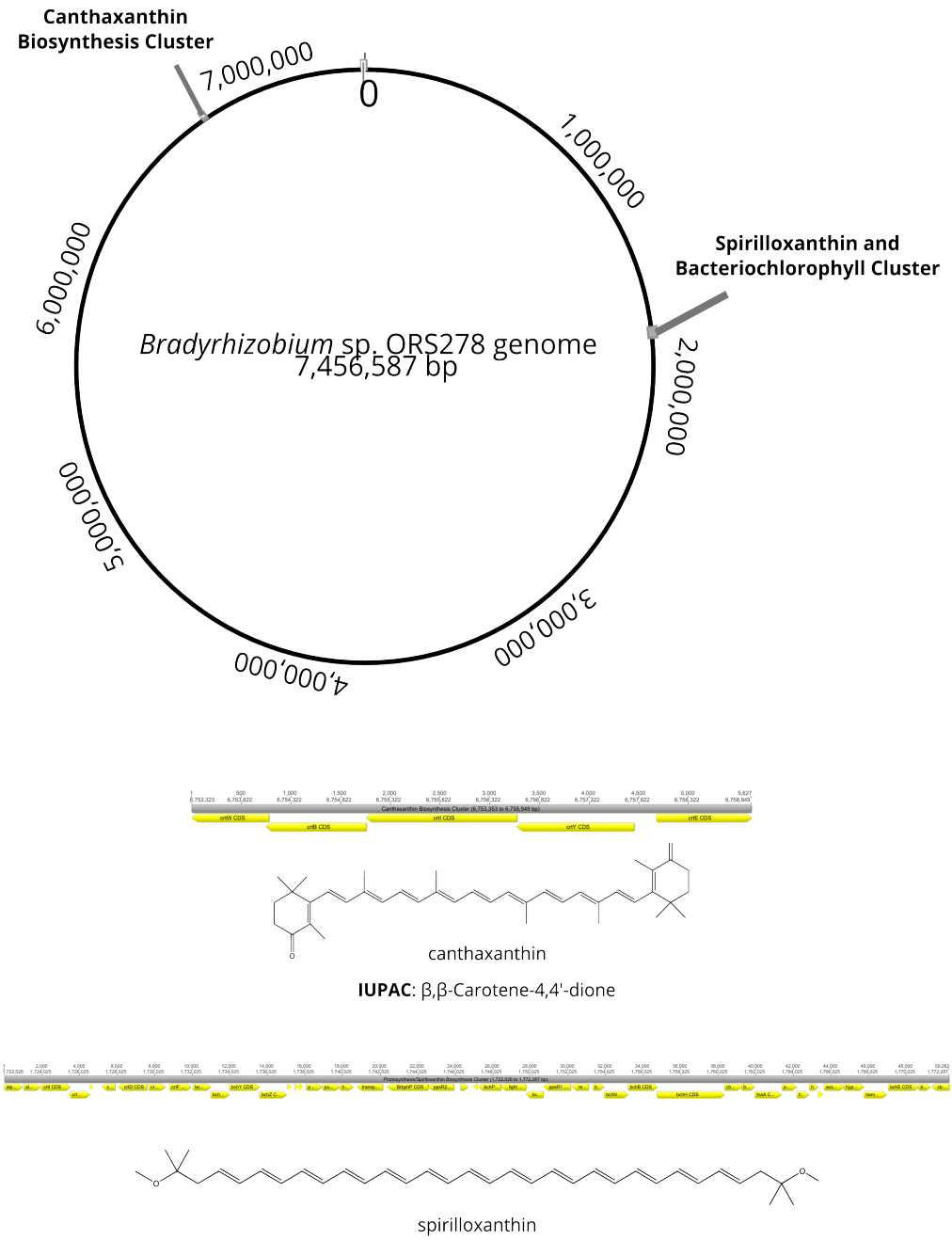


Figure 3.13: The spirilloxanthin/bacteriochlorophyll, and canthaxanthin biosynthesis operons in *Bradyrhizobium* sp. ORS278 and their relative locations on the chromosome.
The genome of *Bradyrhizobium* sp. ORS278 is shown above. The location of the canthaxanthin and spirilloxanthin biosynthesis gene clusters are marked. The structure of both carotenoids and the operons responsible for their synthesis are d. Both operons contain a GGPPS gene. The spirilloxanthin cluster is found directly beside genes involved in bacteriochlorophyll biosynthesis. *Bradyrhizobium* sp. ORS278 inhabits the stem nodules of of aquatic legumes including *Sesbania rostrata*. Spirilloxanthin and bacteriochlorophyll are involved in photosynthesis performed by the bacteroids.

Ultimately, although we did not find any direct evidence to suggest a relationship between these two biosynthetic pathways, we did discover some other interesting information that could be applicable to the study of the gibberellin biosynthesis pathway. Unlike *B. japonicum* USDA110 which inhabits legume root nodules, *Bradyrhizobium* sp. ORS278 colonizes stem nodules of the aquatic plant *Sesbania rostrata* (Lievens *et al*., 2005). Interestingly, endogenous GA is involved in nodulation of *S. rostrata (*Lievens et al., 2005). Bradyrhizobium sp. ORS278 does not have a gibberellin biosynthesis operon, according to the search we conducted. While we did not find any significant sequence homology between the reductases in the ORS278 spirilloxanthin cluster and the *B. japonicum* USDA110 GA operon, we did find a transcription factor binding site within both GGPPS genes.

Previous work found that the transcription factor is a repressor known as PpsR; a binding site for PpsR is not found in the canthaxanthin GGPPS gene. The PpsR repressor is is regulated by iron concentration and far red light (Giraud *et al*., 2004). An increase in the intracellular iron concentration causes PpsR to unbind from the DNA activating transcription of the downstream genes. Far red light also indirectly controls PpsR binding through a two-component regulatory system (Giraud *et al*., 2004). Presumably, the sensor protein (BbrbphP) detects the light within stem nodules which activates the biosynthesis of bacteriochlorophyll by de-repressing transcription of the GGPPS gene and activating transcription of the downstream ORFs. GGPP is a precursor for spirilloxanthin as well as being required for geranylation, and thus activation, of bacteriochlorophyll (Mizoguchi *et al*., 2012). The PpsR transcription factor, first identified in photosynthetic *Rhodobacter,* binds to DNA at palindromic sites with a degenerate binding site of $TGTN_{12}ACA$ (Ponnampalam *et al*., 1995). It prevents transcription of the GGPPS gene unless the iron concentration increases above a certain threshold or its conformation is changed by interaction with BbrbphP (Giraud *et al*., 2004). Although we hoped to make a stronger connection between the ORS278 GGPPS and the USDA110 GGPPS, at this time we

only have the PpsR binding site. Interestingly, we also found the PpsR binding site in the MAFF303099, NGR234, *Xanthomonas*, and some of the *Sinorhizobium meliloti* GGPPS genes (data not shown).

Discussion of Results

Section 3.9: Plant Pathogens with Putative GA Biosynthesis Genes

Using *blr2144* as the BLAST query allowed us to find the putative gibberellin biosynthesis operons in *Xanthomonas* and *Erwinia*. *Xanthomonas* sp. are pathogens that threaten crop yields of rice, peaches and other plants by causing blight—necrotic patches—on their leaves and fruits. *X. oryzae* specifically attacks rice plants. Initially, we were puzzled by our finding of the GA operon in a *Xanthomonas* sp. since we had always thought of the bacterial GA gene products as having beneficial or neutral effect on the plant partner. We had been focused on the rhizobial GA operons where the GA genes appear to be linked to symbiosis for several years. After some time, we came to the realization that *Xanthomonas* and *Erwinia* species may use gibberellin to counteract the plant response to infection and enhance the pathogenicity of its bacterial attack, much like how the *Fujikuroi* fungus uses GA to cause plant disease (Sawada 1917; for a recent review please see Albermann *et al*., 2013). Although we had considered *Xanthomonas* infection the causative agent of plant blight, it is more accurate to attribute blight to a defensive response by the plant. Tissue necrosis is the plant's version of an autoimmune response to the bacterial infection since it reduces the resources available to the pathogen and is targeted at reducing the spread of the infection (Swings *et al*., 1990; Mew *et al*., 1993). Recent transcriptomic analysis of the interaction demonstrated that gibberellin signaling in infected peach plants was significantly down-regulated compared to wild-type controls uninfected by *X. arboricola* pv. *pruni* (Socquet-Juglard *et al*., 2013*)*. The authors suggested that this response could be indicative of the plant's aim to reduce cell elongation and division within the area of infection. Results from our collaborators at Iowa State University appear to confirm the hypothesis. Using *Xanthomonas oryzae* pv.

oryzicola strains unable to synthesize GA, they found that the blighted areas of infected plants are larger than those found on plants infected by wild-type control (Reuben Peters, personal communication). Although we are awaiting publication of this data, it appears that there is experimental evidence to support the hypothesis that bacterial GA contributes to pathogenicity.

Section 3.10: Could Iron Regulate GA Biosynthesis in Late Stage Nodules?

While the best known effects of gibberellins are as plant cell growth promoters, GA can also inhibit tissue senescence in roots and leaves (For a recent review, please see Gupta and Chakrabarty, 2013). In *Bradyrhizobium* sp. ORS278, a high intracellular iron concentration causes PpsR-mediated repression of the GGPPS gene to stop. High iron concentrations can be dangerous to the cell and free iron must be sequestered to prevent damage to cellular components (For a review on this topic, please see Rodrigo-Moreno, Poschenrieder, and Shabala 2013). Therefore, transcription of genes that encode proteins with iron co-factors (such as reductases and ferredoxins) seem to be a plausible strategy to mitigate free iron and prevent damage. The additional layer of transcriptional control for bacteriochlorophyll biosynthesis by a light-sensing two-component regulatory system adds another layer of control to the operon. We believe that the central location of the GGPPS gene in the spirilloxanthin and bacteriochlorophyll clusters, as well as in the putative GA operon, supports the idea that regulation of the intracellular iron concentration could be one of the main roles of bacterial GA biosynthesis. Because *B. japonicum* USDA110 does not photosynthesize and colonizes root rather than stem nodules, we do not expect to find a light-sensing two-component regulatory system involved in GA biosynthesis regulation in this strain.

We were inspired by recent results showing that the *blr2149* and *blr2150* genes of USDA110 are upregulated in 10-week old nodules (Méndez *et al*., 2014). We speculate that during very late stage nodulation, transcription of the gibberellin biosynthesis operon may be controlled by iron concentration

through PpsR. This dual level of control could explain the transcriptomic data showing that the CYPs are upregulated early during nodule organogenesis (Chang *et al*., 2007 and others) as well as the more recent data showing that the *blr2149* and *blr2150* genes are also transcribed during the final stages of the nodule life cycle—early senescence even. During nodule senescence, the degradation of leghemoglobin and nitrogenase releases heme into the cytoplasm; acidic conditions cause free iron to dissociate from the heme, creating toxic intracellular conditions. Biosynthesis of GA could help both the bacteria and plant cells since free iron from the breakdown of the nitrogen-fixing protein would be sequestered by the heme in the bacterial CYPs, and death of the plant cells could be staved off for some time, providing both organisms with a competitive advantage. Although nodules do have a finite life span and eventually die, prolonging the time that the bacteroids spend within the nodule may give the bacteroids a better chance at recolonizing the rhizosphere for the next generation of plant host. Furthermore, a portion of the iron within the nodule is transported to the shoot, and likely the seed once the nodule begins to senescence (Rodríguez-Haas *et al*., 2013). If nodule tissue is prevented from becoming necrotic for a longer period, more iron may be transferred to the seed. Finally, gibberellin may be a protective compound for the bacteroids themselves, though a mechanism for this is not known. The bacteria may more successfully survive the harsh conditions of nodule senescence, facilitating more efficient re-population of the rhizosphere, benefiting the next generation of plant host. and the plant host may be able to produce more seed, resulting in a direct advantage for the plant in the immediate future. Production of more seed could lend an additional indirect to the bacterium in subsequent years.

We propose that there could be at least two levels of transcriptional control governing the gibberellin biosynthesis operon in *Bradyrhizobium japonicum* USDA110: the first being during early nodulation where expression is controlled by oxygen concentration (reviewed in Chapter 1), and the

second during late stage nodulation where gene expression is induced by increased iron concentration. This hypothesis has not yet been tested. It could be tested by studying the fates of radio-labeled gibberellin precursors. Presumably, because plants and bacterial enzymes synthesize gibberellin using different enzymatic mechanisms and catalyze reactions are different carbon atoms in different orders, differential labeling and monitoring of the products should be able to distinguish bacterial and plant gibberellins. This may be an experiment that our collaborators led by Dr. Rojas in Chile could be equipped to perform. In either case, since nodule formation is a complex process and crucial to both the bacterium and plant, it makes sense that multiple levels of control govern expression of related genes.

Section 3.11: Comments on Overall Operon Organization

Both the *Xanthomonas* and *Erwinia* GA operons have the IPP isomerase as well as the additional cytochrome P450 compared to the *B. japonicum* USDA110 gibberellin operon. Although at this time we do not know the function of the putative IPP isomerase, the additional transporter in *M. loti* MAFF303099, or the extra cytochromes P450s found in the *Xanthomonas, Erwinia*, or *Sinorhizobium* species, we infer that these additional genes would contribute additional biochemical reactions that could result in different GAs made in these strains compared to the GA made by *B. japonicum* USDA110. An in-depth characterization of the gibberellin operons, including their gene, proteins and intermediate metabolites in *Xanthomonas* and *Erwinia* is definitely warranted and should allow us to pinpoint the roles of IPP isomerase and the extra P450 when the data are compared with information obtained from study of the *B. japonicum* USDA110 mutants. The other GA operons in the endophytes and symbiotic rhizobia that we found appear to be physically different—they do not have the putative IPP isomerase gene and also lack the extra cytochrome P450. At first, we thought that the spirilloxanthin and bacteriochlorophyll biosynthesis genes could be evolutionary ancestors of the gibberellin operon in other rhizobia because they share similar GGPPS genes, but we were unable to find any evidence to support this hypothesis.

We hypothesize that the GA operon will be found on a symbiotic island in *S. terangae* USDA4894 because we could not find *repABC* analogs in this strain; however, because the genome assembly is still a draft, we recognize that the DNA sequence currently available might not be a true representation of the complete genome. Unfortunately, the genome sequence and replicon information of the closest relative of USDA4894 (*S. chiapanecum* ITTG S70) is not available.

There are transposase and integrase related ORFs surrounding most of the GA operons. The GC content of the putative GA operons is higher than that of the surrounding DNA. It appears that analogs of two *Bradyrhizobium japonicum* USDA110 cytochrome P450 genes involved in gibberellin synthesis are found in five *Sinorhizobium meliloti* strains (*S. meliloti* plasmid pHRC017, AK83 chromosome 3, BL225C pSNMEB01, pRmeGR4b, pSmeSM11c) without any other accompanying GA operon genes. We looked at the GC content of those two genes compared to the surrounding DNA and found that it was also higher than that of the surrounding DNA. We also found transposon-related sequences flanking the cytochrome homologs in these strains. Interestingly, we observed a multi-copper oxidase ORF following the 3' end of the *blr2145* homolog in all five of these *Sinorhizobium meliloti* strains. When we looked for a homolog of the multi-copper oxidase in the Sugawara *Sinorhizobium* species with the GA operon, we found one match in *S. meliloti* N6B1 and another in *S. medicae* A321. In *S. meliloti* N6B1, the multi-copper oxidase gene is near cytochrome P450s thought to be involved in GA biosynthesis, but in *S. medicae* A321, the multi-copper oxidase is not found near any gibberellin biosynthesis genes.

Section 3.12: Nodule Morphology and Gibberellin Biosynthesis

It appears that gibberellin biosynthesis is linked to bacterial species that inhabit legumes with determinate nodules. I assert this because so far, the operon has been found only in bacteria that have symbiotic relationships with plants with determinate nodules with only a few exceptions. For example, a strain of Bradyrhizobium isolated from lupin was recently found to have GA biosynthesis genes when

the genome was sequenced (Reeve *et al*., 2013) and some of the Sinorhizobium strains isolated and sequenced by Sugawara and colleagues (2013) were also from plants with indeterminate nodules of *Medicago* spp. plants. Interestingly, while biovars of *Rhizobium leguminosarum* that interact with pea and bean are quite similar, strains that nodulate pea, a plant that forms indeterminate nodules, do not have the GA biosynthesis genes. In the future we hope to perform a comprehensive analysis of the sequenced bacterial genomes available and compile the meta data on the strain isolates—including nodule morphology—in order to form a stronger hypothesis to support this observation. At this time, I simply wanted to include this paragraph in the thesis to reflect the ongoing research in our lab.

Section 3.13: Conclusions

New putative gibberellin biosynthesis operons were identified by BLAST analysis in two pathogenic bacteria, some endophytes and several nodulating rhizobia including *Sinorhizobium meliloti* strains. This is the first report of putative GA biosynthesis operons in *S. meliloti* strains. We were able to compare the organization of the operons in each organism and saw that most of them shared a high degree of similar genes in the same order. We also noticed that the % GC content of the putative GA biosynthesis genes was higher than that of the surrounding DNA, and in most cases transposon-related ORFs were found flanking the GA biosynthesis genes. We hypothesize that the gibberellin operon is a small mobile genetic region most likely found within mobile, chromosomal symbiotic islands (in the case of *B. japonicum* USDA110 or *M. loti* MAFF303099) or on megaplasmids like in *S. fredii* NGR234. Once the genome assemblies of the Sugawara *Sinorhizobium* strains are completed, we will be able to determine where on the chromosome the operon is found and whether or not it is located within a symbiotic island or plasmid, as we suspect it will be. Although the presence of the putative GA operon in *Xanthomonas* and *Erwinia* strains was unexpected, given what is known about fungal pathogens' use of GA in foolish seedling disease, it is easy to imagine that bacterial GA could also have a role in pathogenicity in some strains. More work to understand transcription of the GA genes during

late-stage nodulation is necessary since it appears that an iron-dependent transcription factor binding site may be located within the GGPPS gene in some species. Finally, more time could be spent examining the transposon and integrase sequences, GC content and codon usage bias to construct an evolutionary history for the GA operon within the bacteria that have it.

Section 3.14: Future Directions

Expansion of the foundational knowledge established by this survey of some newly identified gibberellin biosynthesis operons will be useful in furthering our understanding of bacterial GA. As more bacterial genomes are sequenced, it is very likely that additional GA operons will be discovered. As more GA operons are found, it will be important to catalog them and the GAs that they produce. As the underlying molecular mechanisms become better understood, it will become easier to characterize the new operons. The role of horizontal gene transfer in the evolution of the bacterial GA genes may become more apparent as more work is done to support this hypothesis. For example, codon useage bias between the GA operons and the flanking DNA could be easily calculated; this could provide clues to determine the origin of the operon. Another facet of GA biosynthesis in nodulating bacteria may be nodule and bacteroid morphology. Cataloging the plant species, bacteroid and nodule morphology for all new rhizobial isolates with putative GA operons would provide an additional level of detail to help understand the biological role of bacterial GA. Finally, a targeted study of the GA pathway in *Xanthomonas* and *Erwinia* could increase our understanding of the role of bacterial GA in pathogenic relationships. Finally, now that GA biosynthesis genes have been found in *S. meliloti*, it is important to characterize those pathways as well; many of them are inherently interesting because of the organization of their operons and the additional CYP gene compared to *B. japonicum* USDA110.

Section 3.15: Chapter Summary

1. We propose that there is evidence for lateral transfer of the gibberellin operon between various soil-dwelling bacteria based on our analysis of the GC content and flanking transposon-related ORFs near the gibberellin biosynthesis genes in several bacterial species.

2. The putative gibberellin biosynthesis operon appears to be more widespread in bacteria than was previously thought and may serve different functions in different genetic contexts. We found examples of plant pathogens, symbionts, and endophytic bacteria that harbour homologous putative GA biosynthesis genes.

3. To our knowledge, this is the first report of putative GA biosynthesis operons in *S. meliloti* forming an association with *M. truncatula*. This is significant because *M. truncatula* forms indeterminate nodules harbouring swollen bacteroids and there are very few examples of bacteria with GA operons that form this specific type of plant symbiosis.

Appendix

Synthetic DNA Sequences

Codon optimized CYPs for expression in *E. coli* C41. These sequences were created by our collaborators at Iowa State University.

>sCYP112

ATGAGCGAACAGCAGCCGCTGCCGACCCTGCCGATGTGGCGTGTGGATCATATTGAACCGA
GCCCGGAAATGCTGGCGCTGCGTGCGAACGGCCCGATTCATCGTGTGCGTTTTCCGAGCGG
CCATGAAGGTTGGTGGGTGACCGGCTATGATGAAGCAAAAGCAGTGCTGAGCGATGCAGC
ATTTCGTCCGGCCGGTATGCCGCCGGCAGCATTTACCCCGGATAGCGTGATTCTGGGCAGCC
CGGGTTGGCTGGTTAGCCATGAAGGCCGTGAACATGCGCGTCTGCGTGCGATTGTGGCGCC
GGCGTTTAGCGATCGTCGTGTGAAACTGCTGGTGCAGCAGGTGGAAGCGATTGCGGCGCAT
CTGTTTGAAACCCTGGCAGCACAGCCGCAGCCGGCGGATCTGCGTCGTCATCTGAGCTTTC
CGCTGCCGGCGATGGTGATTAGCGCGCTGATGGGCGTGCTGTATGAAGATCATGCATTTTTC
GCGGGCCTGAGCGATGAAGTGATGACCCATCAGCATGAAAGCGGCCCGCGTAGCGCGAGC
CGTCTGGCGTGGGAAGAACTGCGTGCGTATATTCGTGGCAAAATGCGTGATAAACGTCAGG
ACCCAGATGATAACCTGCTGACCGATCTGCTGGCAGCAGTGGATCAGGGTAAAGCGAGCGA
AGAAGAAGCAGTGGGTCTGGCAGCAGGCATGCTGGTTGCAGGCCATGAAAGCACCGTGGC
GCAGATTGAATTTGGCCTGCTGGCGATGTTTCGTCATCCGCAGCAGCGTGAACGTCTGGTG
GGCGATCCGAGCCTGGTGGATAAAGCGGTGGAAGAAATTCTGCGTATGTATCCGCCGGGCG
CGGGCTGGGATGGCATTATGCGTTATCCGCGTACCGATGTGACCATTGCGGGCGAACATATT
CCGGCGGAAAGCAAAGTTCTGGTNGGGTCTGCCGGCGACCAGCTTTGATCCGCATCATTTT
GATGATCCGGAAATTTTTGATATTGAACGTCAGGAAAAACCGCATCTGGCGTTTAGCTATGG
CCCGCATGCGTGCATTGGCGTGGCGCTGGCGCGTCTGGAACTGAAAGTGGTGTTTGGCAGC
ATTTTTCAGCGTCTGCCGGCGCTGCGTCTGGCGGTGGCGCCGGAACAGCTGAAACTGCGTA
AAGAAATTATTACCGGCGGCTTTGAACAGTTTCCGGTGCTGTGGTAA

>sCYP114
TTGGTTAGCTGCCCTGCGCATGCAGACGCACCGGCAGGGTACGAAAGGTACGCAGAAACG
CGCTCGGTTCACGGGTCGGCTGGCCCGCAAACGCCAGGGTCGGAAAACGTGCCTGAATAC
GCGGCAGGCTTTCTGCCAGCTGCACACGTGCCAGCTGTGCACCCAGGCAAAAGTGGATTC
CATGACCAAAGCTCAGCATAATTTTGCCATCGGTGCTCATGCCCGGGCTGGTGCCATAAAAA
CGCGCCGGATCAAAACGATCTGGATCTGCAAACGCATCCGGATCACGGTTGCCGCTCGCAA
TCAGCACACGCACATCCGCGTTTTTCGGAATCACCACGCCGCCCAGTTCAATATCACGCTGC
GCAATACGCGGAATGCTGCTAAACATCGCCGGCGCATCGCAACGCAGCACTTCTTCCACAA
ACGCACGCACACGCGCCGCATCCGCCTGCAGCCAATGACGCTGTTCCGGATACGCCAGCAT
CGCCAGCACCGCATGATCAATGCTCGCCGCGGTGGTCACAAAGCCGCCCAGCAGCATGCCC
CACAGCATGCTAATCAGTTCCGCATCGCTCAGGGTATCCGCATCATCATCATGCGCACCAAC
CAGCATGCTCACAATATCACGACGCGGATCGGTACGTTTACGCTGAATCAGATCGCCAAAAT
ACGCCTGCACACGTGCGCTTGCTGCATCTGCTGCTGCCAGCTGCGGATCGCTTGCATGCGG
ACCCAGANCTTCCAGAATGGTGCCAATGCCCGCCGCCAGTTCAAACATATCATCCTGCGGC
ATGCCAAACAGTTCCGCAAACACCAGCATCGGCAGCGCCAGCGCAAATTCACGATGCANAT
CCACCGCTTCNCCACGTTCCAGCGCCGGGGTAATGCCATCCAGACGCGCCGCCACAATACG
CGCAATGCTCGGACGCAGGTTATCAATCTGACCCATCGTAAAATCACGGCTAATCAGACGAC
GCAGACGGGTATGGGTCGGCGGATCTTTCATCGCCAGGGTGCTCGCCAGCAGGTTCAGGCT
CGGGCTGGTTGCTGCACGGCTAAAATAACGTGCCAGTTCACCCGGTGCCGGACCACGAAAT
GCATCACCGGTTGCTTTAAATGCCCAGAAGATATCCGCATGACGGCTCAGCAGAAACAGGC
CGCTCGCCGCACGATGCACCGGATCATGTTCACGCAGCCAACGCATAAACGGATACGGATC
ATGAATGCACGCCGGGCTCGCCAGTTCCGCAAACGCATCACGGCACGCCGCGGTGGTTTCC
TGCACATCCATATTGG

>sCYP117
Could not be sequenced completely either at ISU or by TCAG in Toronto, Canada.

>Fdx in pET28b(+)
Could not be sequenced completely either at ISU or by TCAG in Toronto, Canada.

Growth Media and Solutions for Heterologous Expression of CYPs

Tryptone-Yeast Extract Medium

5 g tryptone
3 g yeast extract
0.44 g of $CaCl_2$ 2 H_2O.
To make solid medium, 17 g of agar was added to 1 L of broth.

Arabinose-Gluconate Medium

To prepare one litre of AG, the sugar, gluconic acid and yeast extract are autoclaved together. Sterilized buffer and mineral solutions are added separately after the basal medium has sufficiently cooled.

1 g L-arabinose
1 g gluconic acid
1 g yeast extract

10 mL of the following solutions:
HEPES-MES buffer (65 g HEPES, 55 g MES, pH 6.6 to 6.9 with NaOH in 500 mL dH_2O)
0.67 g/L $FeCl_3 \cdot 6\ H_2O$
18 g/L $MgSO_4 \cdot 7\ H_2O$
1.3 g/L $CaCl_2 \cdot 2\ H_2O$
25.0 g/L Na_2SO_4
32.0 g/L NH_4Cl
12.5 g/L Na_2HPO_4
To make solid media, 17 g of agar was added per litre.

Lysogeny Broth Medium

tryptone.....5 g
yeast extract.....10 g
sodium chloride.....5 g
Use 10 M sodium hydroxide to adjust pH to 7.4 before autoclaving.
To make solid LB, add agar to at least 17 grams per litre.

Terrific Broth Medium (0.9 L)

tryptone.....12 g
yeast extract.....24 g
glycerol.....4 mL
Add 100 mL phosphate buffer pH 7.5 after autoclaving.

Modified Terrific Broth (0.9 L)

casein hydrolysate.....24 g
yeast extract.....48 g
glycerol.....20 mL
magnesium sulfate.....1 g
Add 100 mL phosphate buffer pH 7.5 after autoclaving.

Super Optimal Broth (with Catabolite Repression) Medium

0.5 % yeast extract
2 % tryptone
10 mM NaCl
2.5 mM KCl
10 mM anhydrous $MgCl_2$
10 mM anhydrous $MgSO_4$

The pH was adjusted to 7 before sterilization. After autoclaving, 20 mM glucose was added. If glucose is not added, the medium does not have catabolite repression and is simply called super optimal broth.

Miscellaneous Solutions for Molecular Biology

TAE (1L)

242 g Tris
32.1 g EDTA
57.1 mL glacial acetic acid
distilled water

Binding Buffer II

5.3 M guanidine thiocyanate
15 mM EDTA
120 mM MES (pH 7)
sterile MilliQ water

Isolation of Plasmid DNA

Solution I
50 mM Tris (pH 8.0)
10 mM EDTA
1 mg/mL RNAse A
MilliQ water
Store at 4 ºC when not in use to protect RNAse A from degradation.

Solution II
200 mM NaOH
1% SDS in distilled water
MilliQ water

Solution III
4 M guanosine thiocyanate
0.9 M potassium acetate (pH 4.8)
MilliQ water

Wash Buffer (Isolation of Plasmid DNA)
10 mM Tris (pH 7.4)
80% EtOH
MilliQ water

PB Wash
5 M guanidine thiocyanate
80% ethanol
sterile MilliQ water

Elution Buffer
sterile 2 mM Tris-base (pH 8.0)
sterile MilliQ water

Protein Purification

Lysis Buffer
50 mM phosphate buffer
300 mM NaCl10 mM imidazole
pH 7.0

Protein Purification Lysis Buffer
NaH_2PO_4.....50 mM
NaCl.....300 mM
imidazole.....10 mM
pH 8.0

P450 Lysis/Activity Buffer
Tris.....20 mM
EDTA.....0.5 mM
DTT.....2 mM
pH 7.5

Protein Purification Wash Buffer
NaH_2PO_4.....50 mM
NaCl.....300 mM
imidazole.....20 mM
pH 8.0

Protein Purification Elution Buffer
NaH_2PO_4.....50 mM
NaCl.....300 mM
imidazole.....300 mM
pH ~6.5

Southern blot Solutions

Gel Denaturation Solution
1.5 M NaCl
0.5 M NaOH

Gel Neutralization Solution
1.5 M NaCl
0.5 M Tris-HCl (pH 7.0)

Standard Sodium Citrate
3 M NaCl
300 mM trisodium citrate
(pH 7.0)

Homemade Hybridization Buffer
5X SSC
0.1% (w/v) N-laurylsarcosine (Sarkosyl)
0.02% (w/v) SDS
MilliQ water to 50 mL

Maleic Acid Buffer
0.1 M maleic acid
0.15 M NaCl
pH 7.5.Adjust with solid NaOH pellets.

Prepared as a 10X stock and autoclaved for long-term storage at room temperature.

Washing Buffer
0.1 M maleic acid
0.15 M NaCl (pH 7.5)
0.3% (v/v) Tween 20

Made as 5X stock. Autoclaved and stored at room temperature for up to 8 months. and autoclaved for long-term storage at room temperature. Add Tween 20 after pH is adjusted.

Blocking Solution
Dilute 10X commercial stock in Maleic Acid Buffer to 1X. Use fresh; prepare daily as needed.

Antibody Solution
Supplement 1X Blocking Solution with 75 mU/mL antibody fragments (Roche # 11 093 657 910).

Detection Buffer
0.1 M Tris-base
0.1 M NaCl
pH 9.5

Detection buffer was stored in the dark for up to three months without autoclaving and it maintained its function. The pH can be adjusted after both ingredients are added. A 1.0 M Tris-base solution (pH 8.0) was used as the Tris stock solution.

Protein Gel Electrophoresis

SDS-PAGE Separating Buffer
0.75 M Tris pH 8.8
0.2 % SDS

SDS-PAGE Stacking Buffer
0.5 M Tris pH 6.8
0.4 % SDS

Acrylamide Solution
Biobasic Catalog Number: A0007
Acrylamide/Bis-Acrylamide 40% (29:1)
Stored at 4 ºC Celsius.

10 % Ammonium Persulfate Solution
0.1 g ammonium persulfate in 1 mL distilled water
APS must be prepared fresh every few days. The powder and solution are stored at 4 ºC.

TEMED
BioRad Catalog Number: 161-0801
Stored at four degrees Celsius.

10x SDS-PAGE Running Buffer
0.025 M Tris
0.192 M glycine
0.1% SDS

10X running buffer was diluted to a working concentration of 1X and typically 900 millilitres were used for each electrophoresis run. Running buffer (10X and 1X) may be stored indefinitely at room temperature.

5X SDS-PAGE Sample Buffer
0.3 M Tris pH 6.8
10 % SDS
50 % glycerol
0.125 % (w/v) bromophenol blue
25 % (v/v) 2-mercaptoethanol
This buffer is stored at -20 °C. When needed, it is thawed on ice or at 4 °C.

Separating Gel (This recipe makes 2 mini-gels.)

Separating gel buffer.....5 mL
40 % Acrylamide/Bis (29:1).....2.7 mL
MilliQ distilled water.....2.3 mL
10 % APS.....50 µL
TEMED.....5 µL

These compounds were mixed together in the order listed and mixed before immediately pipetting the solution between secured glass plates. Seventy or eighty percent ethanol (diluted in distilled water) was overlaid on top of the gel solution to facilitate polymerization of the acrylamide. The separating gel was left to stand for a minimum of thirty minutes and a maximum of one hour. After the gel polymerized, distilled water was used to rinse off the ethanol solution and strips of Whatmann 3MM paper were used to dry the top of the gels before the stacking gel was cast.

Stacking Gel (This recipe makes 2 mini-gels.)

stacking gel buffer.....1 mL
40 % Acrylamide/Bis (29:1).....500 µL
MilliQ distilled water.....2.5 mL
10% APS.....25 µL
TEMED.....5 µL

Solutions were mixed in the order listed and immediately pipetted on top of the solidified separating gels. Combs appropriate for the chosen gel width (0.75 or 1.0 millimeter) were inserted with care to avoid trapping air bubbles in the stacking gel. Gels were left to harden for thirty minutes to one hour and either used immediately or stored at 4 °C for up to four days. For this short-term storage, the combs were removed and the plates were wrapped in paper towel soaked in distilled water and kept in a plastic container in the fridge.

In-Gel His-tag Stain

This stain was used to visualize proteins with His-tags after samples were run on SDS-PAGE gels. It is a rapid alternative to western blotting for detection of His-tagged proteins. Once the gel was done running, it was removed from the rig and glass plates and then rinsed with distilled water to remove most of the running buffer. After that the following solutions and buffers were used as described to visualize the His-tagged proteins on a UV transilluminator.

His-tag Fixing Buffer

Methanol.....500 mL
Distilled water.....400 mL
Glacial acetic acid.....100 mL

InVisionTM His-tag In-Gel Stain

Invitrogen catalog number: LC6030

His-tag Destain Buffer

20 mM phosphate buffer 7.8 mL
distilled water

Coomassie Stain

Milli-Q dH_2O.....500 mL
methanol.....400 mL
glacial acetic acid.....100 mL
Coomassie R-250 dye.....1 gram

Coomassie Destain

Distilled water.....500 mL
Methanol.....400 mL
Glacial acetic acid.....100 mL

Mix stain and destain ingredients together in the order listed. Take care to avoid staining surrounding surfaces. Stain can be reused at least twice without filtering and more if it is filtered after use to remove stray pieces of gel. Destain cannot be reused and should not be left to incubate with an SDS-PAGE gel overnight, otherwise too much destaining will occur and the gel must be re-stained.

Bibliography

Ajikumar P. K., Xiao W-H., Tyo K. E. J., Wang Y., Simeon F., Leonard E., Mucha O., Phon T. H., Pfeifer B., Stephanopoulos G. 2010. Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in *Escherichia coli*. *Science*. **330**(6000):70-74.

Albermann S., Elter T., Teubner A., Krischke W., Hirth T., Tudzynski B. 2013. Characterization of novel mutants with an altered gibberellin spectrum in comparison to different wild-type strains of *Fusarium fujikuroi*. *Applied Microbiology and Biotechnology.* **97**:7779-7790.

Altschul S. F., Madden T. L., Schäffer A., A., Zhang J., Miller W., Lipman D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acid Research*. **25**:3389-3402.

Atzorn R., Crozier A., Wheeler C. T., Sandberg G. 1988. Production of gibberellins and indole-3-acetic acid by *Rhizobium phaseoli* in relation to nodulation of *Phaseolus vulgaris* roots. *Planta*. **175**:532-538.

Ausubel F., Brent R., Kingston RE *et al*. 1995. Short Protocols in Molecular Biology. Third Edition. John Wiley and Sons. Chichester.

Barnes HJ. 1996. Maximizing Expression of Eukaryotic Cytochrome P450s in *Escherichia coli*. *Methods in Enzymology.* Cytochrome P450 Part B. 272: 3-15. Edited by Eric F. Johnson and Michael R. Waterman.

Barnes H., Arlotto M., Waterman M. 1991. Expression and enzymatic activity of recombinant cytochrome P450 17α-hydroxylase in *Escherichia coli*. *Proceedings of the National Academy of Sciences*. **88**(13):5597-5601.

Bastián F., Cohen A., Piccoli P., Luna V., Baraldi R. Bottini R. 1998. Production of indole-3-acetic acid and gibberellins A_1 and A_3 by *Acetobacter diazotrophicus* and *Herbaspirillum seropedicae* in chemically-defined culture media. *Plant Growth Regulation* **24**:7-11.

Beringer J. E. 1974. R factor transfer in *Rhizobium leguminosarum*. *Journal of General Microbiology.* **84**:188-198.

Bertani G. 1951. Studies on Lysogenesis I. The Mode of Phage Liberation by Lysogenic *Escherichia coli*. *Journal of Bacteriology.* **62**(3):293-300.

Boghigian B. A., Salas D., Ajikuar P. K., Stephanopoulos G., Pfeifer B. A. 2012. Analysis of heterologous taxadiene production in K- and B-derived *Escherichia coli*. *Applied Microbial and Cell Physiology.* **93**:1651-1661.

Boiero L., Perrig D., Masciarelli O., Penna C., Cassán F., Luna V. 2006. Phytohormone production by three strains of *Bradyrhizobium japonicum* and possible physiological and technological implications. *Applied Microbiology and Biotechnology.* **78**: 874-880.

Bömke C., Tudzynski B. 2009. Diversity, regulation, and the evolution of the gibberellin biosynthetic pathway in fungi compared to plants and bacteria. *Phytochemistry.* **70**:1876-1893.

Bottini R., Fulchieri M., Pearce D., Pharis R. P. 1989. Identification of gibberellins A_1, A_3 and iso-A_3 in cultures of *Azospirillum lipoferum. Plant Physiology.* **90**: 45-47.

Brown M. E., Burlingham S. K. 1968. Production of plant growth substances by *Azotobacter chroococcum*. *Journal of General Microbiology.* **53**:135-144.

Brown C. M., Dilworth M. J. 1975. Ammonia assimilation by *Rhizobium* cultures and bacteroids. *Journal of General Microbiology.* **86**: 39-48.

Brown T. 2001. Southern Blotting. *Current Protocols in Immunology.* Chapter 10 Unit 10.6A.

Brunelle J. L., Green R. 2014. Coomassie blue staining. *Methods in Enzymology.* **541**:161-170.

Cantera J. J. L., Kawasaki H., Seki T. 2002. Farnesyl diphosphate synthase gene of three phototrophic bacteria and its use as a phylogenetic marker. *IUMS*. **52**:1953-1960.

Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J., Boistard P., Becker A., Boutry M., Cadieu E., Dréano S., Gloux S., Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D., Pohl T., Portetelle D., Pühler A., Purnelle B., Ramsperger U., Renard C., Thébault P., Vandenbol M., Weidner S., Galibert F. 2001. Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. *Proceedings of the National Academy of Science USA*. **1498**(17):9877-9882.

Chang W. C., Song H., Liu H. W., Liu P. 2013. Current development in isoprenoid precursor biosynthesis and regulation. *Current Opinion in Chemical Biology.* **17**(4):571-579.

Chang W-S., Franck W. L., Cytryn E., Jeong S., Joshi T., Emerich D. W., Sadowsky M. J., Xu D., Stacey G. 2007. An oligonucleotide microarray resource for transcriptional profiling of *Bradyrhizobium japonicum*. *Molecular Plant Microbe Interactions*. **20**:1298-1307.

Cyr A., Wilderman PR., Determan M., Peters RJ. 2007. A modular approach for facile biosynthesis of labdane-related diterpenes. *Journal of the American Chemistry Society.* **129**:6684-6685.

Darken M. A., Jensen A. L., Shu P. 1959. Production of gibberellic acid by fermentation. *Appl. Microbiol.* **5**:301-303.

da Silva Batista J. S., Hungria M. 2012. Proteomics reveals differential expression of proteins related to a variety of metabolic pathways by genistein-induced *Bradyrhizobium japonicum* strains. *Journal of Proteomics*. **75**:1211-1219.

Delmotte N., Ahrens C. H., Knief C., Qeli E., Koch M., Fischer H-M., Vorholt J. A., Hennecke H., Pessi G. 2010. An integrated proteomics and transcriptomics reference data set provides new insights into the *Bradyrhizobium japonicum* bacteroid metabolism in soybean root nodules. *Proteomics*.

10:1391-1400.

Dolowy M., Pyka A. 2014. Application of TLC, HPLC and GC methods to the study of amino acid and peptide enantiomers: a review. *Biomedical Chromatography.* **28**(1):84-101.

Dombrecht, B., Marchal K., Vabderleyden J., Michiels J. 2002. Prediction and overview of the RpoN-regulon in closely related species of Rhizobiales. *Genome Biology.* **3**:1-11.

Donati A. J., Jeon J-M., Sangurdekar D., So J-S., Chang W-S. 2011. Genome-wide transcriptional and physiological responses of *Bradyrhizobium japonicum* to paraquat-mediated oxidative stress. *Applied and Environmental Microbiology.* **77**:3633-3643.

Donati A. J., Lee H-I., Leveau J. H., Chang W-S. 2013. Effects of indole-3-acetic acid on the transcriptional activities and stress tolerance of *Bradyrhizobium japonicum*. *PLos One*. **8**Le76559.

Dupont L., Alloing G., Pierre O., El Msehli S., Hopkins J., Hérouart D., Frendo P. 2013. The legume root nodule: from symbiotic nitrogen fixation to senescence.

Ernstsen A., Sandberg G., Crozier A., Wheeler C. T. 1987. Endogenous indoles and the biosynthesis and metabolism of indole-3-acetic acid in cultures of *Rhizobium phaseoli*. *Planta*. **171**:422-428.

Ferre E., Trabalzini L. 2013. The yeast two-hybrid and related methods as powerful tools to study plant cell signalling. 83(4-5):287-301.

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

Fischbach M. A., Clardy J. 2007. One pathway, many products. *Nature Chemical Biology.* 3:353-355.

Fischer H. M. 1994. Genetic regulation of nitrogen fixation in rhizobia. *Microbiology Reviews*. 58(3):352-386.

Fisher C. W., Caudle D. L., Martin-Wixtrom C., Quattrochi L., C., Tukey R. H., Waterman M. B., Estabrook R. W. 1992. High-level expression of functional cytochrome P450 1A2 in *Escherichia coli*. *The Journal of the Federation of American Societies for Experimental Biology.* **6**:759-764.

Franklind B., Wareing P. F. 1960. Effect of gibberellic acid on hypocotyl growth of lettuce seedlings. *Nature*. **185**:255-256.

Grappelli A., Rossi W. 1981. The effect of phytohormones produced by *Arthrobacter* sp. on the phosphatase activity in plant roots. *Folia Microbiol*. **26**:137-141.

Giraud E., Hannibal L., Fardoux J,. Jaubert M., Jourand P., Dreyfus B., Sturgis J. N., Verméglio A. 2004. Two Distinct *crt* Gene Clusters for Two Different Functional Classes of Carotenoid in *Bradyrhizobium*. *Journal of Biological Chemistry.* **279**:15076-15083.

Giraud E, Moulin L, Vallenet D, Barbe V, Cytryn E, Avarre JC, Jaubert M, Simon D, Cartieaux F, Prin Y., et al. 2007. Legumes symbioses: absence of Nod genes in photosynthetic *Bradyrhizobia*. *Science.* **316**: 1307–1312.

Göttfert M., Röthlisberger S., Kündig C., Beck C., Marty R., Hennecke H. 2001. Potential symbiosis-specific genes uncovered by sequencing a 410-kilo base DNA region of the *Bradyrhizobium japonicum* chromosome. *Journal of Bacteriology.* **183**:1405-1412.

Guengerich F. P., Martin M. V., Sohl C. D., Cheng Q. Measurement of cytochrome P450 and NADPH-cytochrome P450 reductase. *Nature Protocols*. **4**(9):1245-1251.

Guinel F. C. 2009. Getting around the legume nodule: I.The structure of the peripheral zone in four nodule types. *Botany.* **87**(12):1117-1138.

Guoy M., Guindon S., Gascuel O. 2010. SeaView Version 4: Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution*. 27(2):221-224.

Gupta R., Chakrabarty S. K. 2013. Gibberellic acid in plant: still a mystery unresolved. *Plant Signal Behaviour.* **8**(9):pii e25504

Gutiérrez-Mañero F. J., Ramos-Solano B., Probanza A., Mehouachi J., Tadeo F. R., Talon M. 2001. The plant-growth-promoting rhizobacteria *Bacillus pumilus* and *Bacillus licheniformis* produce high amounts of physiologically active gibberellins. *Physiologia Plantarum* **111**:206-211.

Han L. L. *et al*., 2009. Unique community structure and biogeography of soybean rhizobia in the saline-alkaline soils of Xinjiang, China. *Plant Soil*. **324**:291-305.

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

Hannibal L., Lorquin J., Angles D'Ortoli N., Garcia N., Chaintreuil C., Masson-Boivin C., Dreyfus B., Giraud E. 2000. Isolation and Characterization of Canthaxanthin Biosynthesis Genes from the Photosynthesis Bacterium *Bradyrhizobium* sp. Strain ORS278. *J. Bacteriology* **182**(13):3850-3853.

Harris L. J., Saparno A., Johnston A., Prisic S., Xu M., Peters R.J. 2005. The maize *An2* gene is induced by *Fusarium* attack and encodes an *ent*-copalyl diphosphate synthase. *Plant Molecular Biology.* **59**:881-894.

Hauser F., Pessi G., Friberg M. Weber C., Rusca N., Lindemann A., Fischer H-M., Hennecke H. 2007. Dissection of the Bradyrhizobium japonicum NifA+sigma54 regulon, and identification of a ferredoxin gene (fdxN) for symbiotic nitrogen fixation. *Molecular Genetics and Genomics*. **278**:255-271.

Hayashi S., Gresshoff P. M., Fergusom B. J. 2014. Mechanistic action of gibberellins in legume nodulation. *Journal of Integrative Plant Biology.* **56**(10)971-978.

Hedden P., Phillips A. L., Rojas M. C., Carrera E., Tudzynski B. 2001. Gibberellin Biosynthesis in Plants and Fungi: A Case of Convergent Evolution? *Journal of Plant Growth Regulation*. **20**(4):319-331.

Hedden P., Thomas S. G. 2012. Gibberellin biosynthesis and its regulation. *Biochemical Journal.* **444**:11-25.

Hershey D. M., Lu X., Zi J., Peters R. J. 2014. Functional Conservation of the Capacity for *ent*-Kaurene Biosynthesis and an Associated Operon in Certain Rhizobia. *Journal of Bacteriology.* **196**(1):100-106.

Hill T. A., Wimble R. H. 1969. A Note on the Precision of Estimates of Gibberellin Concentration from Regression Lines Calculated from Bioassay Data. *Planta*. 87:20-25.

Hussain, Vančura, V. 1970. Formation of biologically active substances by rhizosphere bacteria and their effect on plant growth. *Folia Microbiol*. **15**:468-478.

Humphry D. R., Andrews M., Santos S. R., James E. K., Vinogradova L. V., Perin L., Reis V. M., Cummings S. P. 2007. Phylogenetic assignment and mechanism of action of a crop of growth promoting *Rhizobium radiobacter* strain used as a biofertiliser on graminaceous crops in Russia. *Antonie van Leeuwenhoek*. **91**:105-113.

Hwang I. S., Kang W. R., Hwang D. J., Bae S. C., Yun S. H., Ahn I. P. 2013. Evaluation of bakane disease progression caused by *Fusarium fujikuroi* in *Oryza sativa* L. *Journal of Microbiology.* **51**(6):858-865.

Janzen R. A., Rood S. B., Dormar J. F., McGill W. B. 1992. *Azospirillum brasilense* produces gibberellin in pure culture on chemically-defined medium and in co-culture on straw. *Soil Biol. Biochem*. **10**:1061-1064.

Jeon J-M., Lee H-I., Donati A., So J-S., Emerich D. W., Chang W-S. 2011. Whole-genome expression profiling of *Bradyrhizobium japonicum* in response to hydrogen peroxide. *Molecular Plant Microbe Interactions*. **24**:1472-1481.

Joo G-J., Kim Y-M., Lee I-J., Song K-S., Rhee I-K. 2004. Growth promotion of red prepper plug seedlings and the production of gibberellins by *Bacillus cereus*, *Bacillus macroides* and *Bacillus pumilus*. *Biotechnology Letters*. **26**:487-491.

Joo G-J., Kim Y-M., Kim J-T., Rhee I-K., Kim J-H., Lee I-J. 2005. Gibberellins-Producing Rhizobacteria Increase Endogenous Gibberellins Content and Promote Growth of Red Peppers. The Journal of Microbiology. 43(6):510-515.

Joo G-L., Hamayun M., Kim S-K., Na C-I., Shin D-H., Lee I-J. 2009. *Burkholderia* sp. KCTC 11096BP as a newly isolated gibberellin producing bacterium. *The Journal of Microbiology.* **47**:167-171.

Jung S. T., Lauchli R., Arnold F. H. 2011. Cytochrome P450: taming a wild type enzyme. Current Opinion in Biotechnology. **22**:809-817.

Kaneko T., Nakamura Y., Sato S., Asamizu E., Sasamoto S., Watanabe A., Idesawa K., Kawashima K., Kimura T., Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Mochizuki Y., Nakayama N., Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S. 2000. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Research.* **7**:331-338.

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

Kang S-M., Joo G-J., Hamayun M., Na C-I., Shin D-H., Kim H. K., Hong J-K., Lee I-J. 2009. Gibberellin production and phosphate solubilization by newly isolated strain of *Acinetobacter calcoaceticus* and its effect on plant growth. *Biotechnology Letters.* **31**:277-281.

Kang S-M., Khan A. L., Hamayun M., Hussain J., Joo G-J., You Y-H., Kim J-G., Lee I-J. 2012. Gibberellin producing *Promicromonospora* sp. SE188 improves *Solanum lycopersicum* plant growth and influences endogenous plant hormones. *The Journal of Microbiology.* **50**: 902-909.

Kang S-M., Khan A. L., You Y-H., Kim J-G., Kamran M., Lee I-J. 2013. Gibberellins production by newly isolated strain of *Leifsonia soli* SE143 and its potential for plant growth promotion. *Journal of Microbiology and Biotechnology.* Online edition.

Khan A. L., Waqas M., Kang S. M., Al-Harrasi A., Hussain J., Al-Rawahi A., Al-Khiziri S., Ullah I., Ali L., Jung H. Y., Lee I. J. 2014. Bacterial endophyte *Sphingomonas* sp. LK11 produces gibberellins and IAA and promotes tomato plant growth. *Journal of Microbiology.* **52**(8):689-695.

Kasahara H., Hanada A., Kuzuyama T., Takagi M., Kamiya Y., Yamaguchi S. 2002. Contribution of the mevalonate and methylerythritol phosphate pathways to the biosynthesis of gibberellins in *Arabidopsis*. *Journal of Biological Chemistry.* 277:45188-45194.

Katznelson H., Cole S. E. 1965. Production of gibberellin-like substances by bacteria and actinomycetes. *Canadian Journal of Microbiology.* **11**: 733-741.

Katznelson H., Sirois J. C., Cole S. E. 1962. Production of a gibberellin-like substance by *Arthrobacter gloiformis*. *Nature*. **196**: 1012-1013.

Kawaguchi M., Imaizumi-Anraku H., Fukai S., Syōno K. 1996. Unusual branching in the seedlings of *Lotus japonicus*—gibberellins reveal the nitrogen-sensitive cell divisions within the pericycle on roots. *Plant Cell Physiology.* **37**:461-470.

Keister D. L., Tully R. E., van Berkum P. 1995. Short communication: A cytochrome P450 gene cluster in the Rhizobiaceae. *The Journal of Applied and General Microbiology.* **45**:301-303.

Kelly S. L., Kelly D. E. 2013. Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us? Philosophical

Transactions of the Royal Society B: Biological Sciences. **368**:20120476.

Kempf A. C., Zanger U. M., Meyer U. A. 1995. Truncated human P450 2D6: expression in *Escherichia coli*, Ni2+ chelate affinity purification, and characterization of solubility and aggregation. *Archives of Biochemistry and Biophysics.* **321**:277-288.

Krishnan H. B. 2002. NolX of *Sinorhizobium fredii* USDA257, a type III-secreted protein involved in host range determination, is localized in the infection threads of cowpea (Vigna unguiculata [L.] Walp) and soybean (Glycine max [L.] Merr.) nodules. *Journal of Bacteriology.* **184**(3):831-839.

Laemmli U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*. **227**(5259):680-685.

Lale G., Jogdand V. V., Gadre R. V. 2006. Morphological mutants of *Gibberella fujikuroi* for enhanced production of gibberellic acid. *Journal of Applied Microbiology.* **100**(1):65-72.

Lamb D. C., Skaug T., Song H-L., Jackson C. L., Podust L. M., Waterman M. R., Kell D. B., Kelly D., E., Kelly S. L. 2002. The Cytochrome P450 Complement (CYPome) of *Streptomyces coelicolor* A3(2). *The Journal of Biological Chemistry.* **277**:24000-24005.

Lee I-J., Foster K. R., Morgan P. W. 1998. Photoperiod control of gibberellin levels and flowering in Sorghum. *Plant Physiology.* **116**:1003-1011.

Lee, Ka, Song

Lee H-I., Lee J-H., Park K-H., Sangurdekar D., Chang W-S. 2012. Effect of soybean coumestrol on *Bradyrhizobium japonicum* nodulation ability, biofilm formation and transcriptional profile. *Applied and Environmental Microbiology.* **78**:2896-2903.

Li Y., Tian C. F., Chen W. F., Wang L., Sui X H., Chen W. X. 2013. High-resolution transcriptomic analysis of *Sinorhizobium* sp. NGR234 bacteroids in determinate nodules of *Vigna unguiculata* and indeterminate nodules of *Leucaena leucocephala*. *PLOS One.* **8**:e70531.

Lievens S., Goormachtig S., Den Herder J., Capoen W., Mathis R., Hedden P., Holsters M. 2005. Gibberellins are involved in nodulation of *Sesbania rostrata*. *Plant Physiology.* **139**(3):1366-1379.

MacMillian J. 2002. Occurrence of gibberellins in vascular plants, fungi, and bacteria. *Journal of Plant Growth Regulation*. **20**:387-442.

Macmillan J., Beale M. H. 1999. Diterpene biosynthesis. In Isoprenoids including Carotenoids and Steroids (Cane D. E., ed), pp. 217-243, Elsevier, Amsterdam.

Malfanova N., Kamilova F., Validov S., Shcherbakov A., Chebotar V., Tikhonovich I., Lugtenberg B. 2011. Characterization of *Bacillus subtilis* HC8, a novel plant-beneficial endophytic strain from giant hogweed. *Microbial Biotechnology.* **4**:523-532.

Man C. X. *et al*. 2008. Diverse rhizobia associated with soybean grown in the sub-tropical and tropical

region of China. *Plant Soil*. 310:77-87.

Mander L. N. 2003. Twenty years of gibberellin research. *Natural Product Reports*. **20**:49-69.

Marchler-Bauer A., Zheng C., Chitsas F., Derbyshire M. K., Geer R. C., Gonzalez N. R., Gwadz M., Hurwitz D. I., Lanxzycki C. J., Lu F., Lu S., Marchler G. H., Song J. S., Thanki N, Yamashita R. A., Zhang D., Bryant S. H. 2013. CDD: Conserved domains and protein three-dimensional structure. *Nucleic Acids Research*. **41**(D1):D348-352.

Meade H. M., Long S. R., 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 Tn5 mutagenesis. *Journal of Bacteriology.* **149**(1):114-122.

Méndez C., Baginsky C., Hedden P., Gong F., Carú M., Rojas M. C. 2014. Gibberellin oxidase activities in Bradyrhizobium japonicum bacteroids. *Phytochemistry.* **98**:101-109.

Mesa S., Hauser F., Friberg M., Malaguti E. Fischer H-M., Hennecke H. 2008. Comprehensive assessment of the regulons controlled by the FixLJ-$FixK_2$-$FixK_1$ cascade in *Bradyrhizobium japonicum*. *Journal of Bacteriology.* **190**:6568-6579.

Mesa S., Hennecke H., Fischer H. M. 2006. A multitude of CRP/FNR-like transcripton proteins in *Bradyrhizobium japonicum*. *Biochemical Society Transactions*. **34**(1):156-159.

Mew TW, Alvarez AM, Leach JE, Swings J. 1993 Focus on bacterial blight of rice. *Plant Disease.* 77:5-12.

Mizoguchi T., Isaji M., Harada J., Tamiaki H. 2012. Isolation and pigment composition of the reaction centers from purple photosynthetic bacterium *Rhodopseudomonas palustris* species. *Biochimica et Biophysica Acta*. **1817:**395-400.

Morrone D., Chambers J., Lowry L., Kim G., Anterola A., Bender K., Peters R. J. 2009. Gibberellin biosynthesis in bacteria: separate *ent*-copalyl diphosphate and *ent*-kaurene synthases in *Bradyrhizobium japonicum*. *FEBS Letters*. **583**:475-480.

Morrone D., Lowry L., Determan MK., Hershey DM., Xu M., Peters RJ. 2010. Increasing diterpene yield with a modular metabolic engineering system in *E. coli*: comparison of MEV and MEP isoprenoid precursor pathway engineering. *Applied Microbial Biotechnology.* **85**:1893-1906.

Müller A., Düchting P., Weiler E. W. 2006. Hormone profiling in *Arabidopsis*. *Methods in Molecular Biology.* **323**:449-457.

Nelson D. R. 2009. The Cytochrome P450 Homepage. *Human Genomics*. **4**:59-65.

Newton J. W., Wilson P. W., Burris R. H. 1953. Direct demonstration of ammonia as an intermediate in nitrogen fixation by *Azotobacter. Journal of Biological Chemistry* **204**:445-451.

Nishihara K., Kanemori M., Kitagawa M., Yanagi H., Yura T. 1998. Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. *Applied Environmental Microbiology.* **64**:1694–1699.

Nomura M., Arunothayanan H., Van Dao T., Le H., T-P., Kaneko T., Sato S., Tabata S., Tajima S. 2010. Differential protein profiles of *Bradyrhizobium japonicum* USDA110 bacteroid during soybean nodule development. *Soil Science and Plant Nutrition*. **56**:579-590.

Ohnuma S-I., Hirooka K., Ohto C., Nishino T. 1997. Conversion from Archaeal Geranylgeranyl Diphosphate Synthase to Farnesyl Diphosphate Synthase: two Amino Acids Before The First Aspartate-rich Motif Solely Determine Eukaryotic Farnesyl Diphosphate Synthase Activity. *Journal of Biological Chemistry.* **272**:5192-5198.

Olszewski N., Sun T. P., Gubler F. 2002. Gibberellin signaling: biosynthesis, catabolism and response pathways. *Plant Cell* (Suppl) **14**:S61-S80

Oluwafemi A. O., Fenner C. J., Gudiminchi R. K., Smit M. S., Harrison STL. 2013. The influence of microbial physiology on biocatalyst activity and efficiency in the terminal hydroxylation of n-octane using *Escherichia coli* expressing the alkane hydroxylase, CYP153A6. *Microbial Cell Factories*. 12(8):1-12.

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**(4):1233-1243.

Ou S. H. 1985. Rice Diseases. Commonwealth Mycological Institute, Kew, England.

Paleg L. G. 1965. Physiological effects of gibberellins. *Annual Reviews in Plant Physiology.* **16**:291-322.

Perret X., Freiberg C., Rosenthal A., Broughton W. J., Fellay R. 1999. High-resolution transcriptional analysis of the symbiotic plasmid of *Rhizobium* sp. NGR234. *Molecular Microbiology.* **32**:415-425.

Peters R. J., Flory J. E., Jetter R., Ravn M. M., Lee H-J et al., 2000. Abietadiene synthase from grand fir (*Abies grandis*): characterization and mechanism of action of the “pseudomature” recombinant enzyme. *Biochemistry.* **39**:15592-15602.

Ponnampalam S. N., Buggy J. J., Bauer C. E. 1995. Characterization of an aerobic repressor that coordinately regulates bacteriochlorophyll, carotenoid, and light harvesting-II expression in *Rhodobacter capsulatus*. *Journal of Bacteriology.* **177**(11):2990-2997.4

Poulos TL., Finzel BC., Howard AJ. 1987. High-resolution crystral structure of cytochrome P450cam. *Journal of Molecular Biology.* **195**(3):687-700.

Pueppke S. G., Broughton W. J. 1999. *Rhizobium* sp. Strain NGR234 and *R. fredii* USDA257 Share

Exceptionally Broad, Nested Host Ranges. *Molecular Plant-Microbe Interactions*. **12**(4)293-318.

Rodrigo-Morena A., Poschenrieder C., Shabala S. 2013. Transition metals: a double edge sward in ROS generation and signaling. *Plant Signal Behaviour.* **8**(3):e23425.

Rolfe B. G. 1988. Flavones and isoflavones as inducing substances of legume nodulation. *Biofactors*. **1**(1):3-10.

Reeve W. G., O'Hara G., Chain P., Ardley J., Bräu L., Nandesena K. *et al*. 2010. Complete genome sequence of *Rhizobium leguminosarum* bv. *trifolii* strain WSM1325, an effective microsymbiont of annual Mediterranean clovers. *Standards in Genomic Science*. **2**:347-356.

Reeve W., Terpolilli J., Melino V., Ardley J., Tian R., De Meyer S., Tiwari R., Yates R., O'Hara G., Howieson J., Teshima H., Bruce D., Detter C., Tapia R., Han C., Wei C. L., Huntemamm M., Han J., Chem I. M., Mavrommatis K., Markowitz V., Ivanova N., Ovchinnikova G., Pagani I., Pati A., Goodwin L., Peters L., Woyke T., Kyrpides N. 2013. Genome sequence of the lupin-nodulating *Bradyrhizobium* sp. strain WSM 1417. *Standards in Genomic Science*. **9**(2):273-282.

Rincón-Rosales R., Lloret L., Ponce E., Martínez-Romero E. 2009. Rhizobia with differet symbiotic efficiencies nodulate Acaciella angustissima in Mexico, including *Sinorhizobium chiapanecum* sp. nov. which has common symbiotic genes with *Sinorhizobium mexicanum*. *FEMS Microbial Ecology.* (1):103-107.

Rodrigues C., de Souza Vandenberghe L. P., de Oliveira J., Soccol C. R. 2011. New perspectives of gibberellic acid production: a review. *Critical Reviews in Biotechnology.* Online Edition. 1-11.

Rodríguez-Haas B., Finney L., Vogt S., González-Melendi P., Imperial J., González-Guerrero M. 2013. Iron distribution through the developmental stages of *Medicago truncatula* nodules. *Metallomics*. **5**(9):1247-1253.

Sadowsky M. J., Tully R. E., Cregan P. B., Keyser H. H. 1987. Genetic diversity in *Bradyrhizobium japonicum* serogroup 123 and its relation to genotype-specific nodulation of soybean. *Applied Environmental Microbiology.* **53**:2624-2630.

Salazar E., Diaz-Mejia J. J., Moreno-Hagelsieb G., Martinez-Batallar G., Mora Y., Mora J., Encarnación S. 2010. Characterization of the NifA-RpoN regulon in *Rhizobium etli* in free life and in symbiosis with *Phaseolus vulgaris*. *Applied and Environmental Microbiology.* **76**:4510-4520.

Sambrook JF., Russell DW. 2001. Molecular Cloning: A Laboratory Manual. 3Rd Edition. Volume 1, 2, and 3. Cold Spring Harbor Laboratory Press.

Sawada K. 1917. Contributions on Formosan fungi. *Transactions of the National Historical Society.* **7**:131-133.

Schäfer A, Tauch A, Jäger W, Kalinowski J., Thierbach G., Pühler A. 1994. Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of

defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene*. **145**:69-73.

Schenkman J. B., Jansson I. 2006. Spectral analyses of cytochromes P450. *Methods in Molecular Biology: Cytochrome P450 Protocols*. 320:11-18.

Schmeisser C., Liesegang H., Krysciak D., Bakkou N., Quéré A. L., Wollherr A., Heinemeyer I., Morgenstern B., Pommerening-Röser A., Flores M., Palacios R. Brenner S., Gottschalk G., Schmitz R. A., Broughton W. J., Perret X., Strittmatter A. W., Streit W. R. 2009. *Rhizobium* sp. strain NGR234 possesses a remarkable number of secretion systems. *Applied Environmental Microbiology.* **75**:4035-4045.

Sciotti M-A., Chanfon A. Hennecke H., Fischer H-M. 2003. Disparate oxygen responsiveness of two regulatory cascades that control expression of symbiotic genes in *Bradyrhizobium japonicum*. *Journal of Bacteriology.* **185**: 5639-5642.

Sielaff B., Andreesen J. R. 2005. Kinetic and binding studies with purified recombinant proteins ferredoxin reductase, ferredoxin and cytochrome P450 comprising the morpholine mono-oxygenase from *Mycobacterium* sp. strain HE5. *The FEBS Journal*. **272**(5):1148-1159.

Shimoda Y., Mitsui H., Kamimatsuse H., Minamisawa K., Nishiyama E., Ohtsubo Y., Nagata Y., Tsuda M., Shinpo S., Watanabe A., Kohara M., Yamada M., Nakamura Y., Tabata S., Sato S. 2008a. Construction of signature-tagged mutant library in *Mesorhizobium loti* as a powerful tool for functional genomics. *DNA Research.* **15**:297-308.

Shimoda Y., Shinpo S., Kohara M., Nakamura Y., Tabata S., Sato S. 2008b. A large scale analysis of protein-protein interactions in the nitrogen-fixing bacterium *Mesorhizobium loti*. *DNA Research.* **15**:13-23.

Snell E. E., Strong F. M. 1939. A microbiological assay for riboflavin. *Industrial & Engineering Chemistry Analytical Edition.* **11**:346-350.

Socquet-Juglard D., Kamber T., Pothier J. F., Christen D., Gessler C., Duffy B., Patocchi A. 2013. Comparative RNA-Seq Analysis of Early-Infected Peach Leaves by the Invasive Phytopathogen *Xanthomonas arboricola* pv. *pruni*. *PLOS one*. **8**(1):e54196.

Spaggiari D., Geiser L., Daali Y., Rudaz S. 2014. A cocktail approach for assessing the in vitro activity of human cytochrome P450s: An overview of current methodologies. *Journal of Pharmaceutical and Biomedical Analysis*. pii:S0731-7085(14)00145-9.

Sugawara M., Epstein B., Badgley B. D., Unno T., X L., Reese J., Gyaneshwar P., Denny R., Mudge J., Bharti A. K., Farmer A. D., May G. D., Woodward J. E., Medigue C., Vallenet D., Lajus A., Rouy Z., Martinez-Vaz B., Tiffin P., Young N. D., Sadowsky M. J. 2013. Comparative genomics of the core and accessory genomes of 48 *Sinorhizobium* strains comprising five genospecies. *Genome Biology.* 14:**R17** 1-19.

Sukdeo N. and Charles T. C. 2003. Application of crossover-PCR-mediated deletion-insertion

mutagenesis to analysis of the *bdhA-xdhA-xdhB2* mixed-function operon of *Sinorhizobium meliloti*. *Archives of Microbiology.* **179**(4):301-304.

Sullivan J. T., Brown S. D., Ronson C. W. 2013. The NifA-RpoN regulon of *Mesorhizobium loti* strain R7A and its symbiotic activation by a novel LacI/GalR family regulator. *PLOS One*. **8**:e53762.

Swings J., Van Den Mooter M., Vauterin L., Hoste B., Gillis M., Mew TW et al. 1990. Reckassification of the Causal Agents of Bacterial Blight (*Xanthomonas campestris* pv. *oryzae*) and Bacterial Leaf Streak (*Xanthomonas campestris* pv. *oryzicola*) of Rice as Pathovars of *Xanthomonas oryzae* (ex Ishiyama 1922) sp. nov., nom. rev. *International Journal of Systematic Bacteriology.* **40**:309-311.

Tatsukami Y., Nambu M., Morisaka H., Kuroda K., Ueda M. 2013. Disclosure of the differences of Mesorhizobium loti under the free-living and symbiotic conditions by comparative proteome analysis without bacteroid isolation. *BMC Microbiology.* **13**:180-189.

Taylor C. B. 1951. The nutritional requirements of predominant bacterial flora of the soil. *Proceedings of the Society of Applied Bacteriology.* **14:**101-111.

Tian C. F., Zhou Y. J., Zhang Y. M., Li Q. Q., Zhang Y. Z., Li D. F., Wang S., W J., Gilbert L. B., Li Y. R., Chen W. X. 2011. Comparative genomics of rhizobia nodulating soybean suggests extensive recruitment of lineage-specific genes in adaptations. *Proceedings of the National Academy of Sciences Early Edition*. **109**(22):1-6.

Tien T. M., Gaskina M. H., Hubbell D. H. 1979. Plant growth substances produced by *Azospirillum brasilense* and their effect on the growth of Pearl Millet (*Pennisetum americanum* L.) *Applied and Environmental Microbiology.* **37**:1016-1024.

Torres M. J., Bueno E., Mesa S., Bedmar E. J., Delgado M. J. Emerging complexity in the denitrification regulatory network of *Bradyrhizobium japonicum*. *Biochemical Society Transactions*. **39**:284-288.

Trainer M. A. 2008. Carbon Metabolism and Desiccation Tolerance in the Nitrogen-Fixing Rhizobia *Bradyrhizobium japonicum* and *Sinorhizobium meliloti*. PhD Thesis. University of Waterloo.

Troncoso C., Cárcamo J., Hedden P., Tudzynski B., Rojas MC. 2008. Influence of electron transport proteins on the reactions catalyzed by *Fusarium fujikuroi* gibberellin monooxygenases. *Phytochemistry.* **69**(3):672-683.

Tudzynski B. 1999. Biosynthesis of gibberellins in *Gibberella fujikuroi*: biomolecular aspects. *Applied Microbial Biotechnology.* **52**:298-310.

Tudzynski B. 2005. Gibberellin biosynthesis in fungi: genes, enzymes, evolution, and impact on biotechnology. *Applied Microbial Biotechnology.* **66**:597-611.

Tully R. E., Keister D. L. 1993. Cloning and mutagenesis of a cytochrome P-450 locus from *Bradyrhizobium japonicum* that is expressed anaerobically and symbiotically. *Applied and Environmental Microbiology.* **59**:4236-4142.

Tully R. E., van Berkum P., Lovins K. W., Keister D. L. 1998. Identification and sequencing of a cytochrome P450 gene cluster in *Bradyrhizobium japonicum*. *Biochimica et Biohysica Acta*. **1398**:243-255.

Uchiumi T., Ohwada T., Itakura M., Mitsui H., Nukui N., Dawadi P., Kaneko T., Tabata S. Yokoyama T., Tejima K., Saeki K., Omori H., Hayashi M., Maekawa T., Sriprang R., Murooka Y., Takima S., Simomura K., Nomura M., Suzuki A., Shimoda Y., Sioya K., Abe M., Minamisawa K. 2004. Expression islands clustered on the symbiosis island of the *Mesorhizobium loti* genome. *Journal of Bacteriology.* **186**: 2439-2448.

Upadhyay S. K., Singh D. P., Saikia R. 2009. Genetic diversity of plant growth promoting rhizobacteria isolated from the rhizospheric soil of wheat under saline condition. *Current Microbiology.* **59**:489-496.

Vágnerová K., Macura J., Čatská V. 1960. Rhizosphere microflora of wheat. I. Composition of properties of bacterial flora during the first stages of growth. *Fol. Microbiol*. **5**:298-310.

Vallenet D., Belda E., Calteau A., Cruveiller S., Engelen S., Lajus A., Le Fèvre F., Longin C., Mornico D., Roche D., Rouy Z., Salvignol G., Scarpelli G., Thil Smith AA., Weiman M., Médigue C. 2013. MicroScope—an integrated microbial resource for curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Research*. **41**:D636-647.

Vančura V. 1961. Detection of gibberellic acid in *Azotobacter* cultures. *Nature*. **192**:88-89.

Vinuesa P. *et al*. 2008. Multilocus sequence analysis for assessment of the biogeography and evolutionary genetics of four *Bradyrhizobium* species that nodulate soybeans on the asiatic continent. *Applied Environmental Microbiology.* **74**:6987-6996.

Viprey V., Rosenthal A., Broughton W. J., Perret X. 2000. Genetic snapshots of the *Rhizobium* species NGR234 genome. *Genome Biology.* **1**(6):Research0014 (Epub).

Wilderman P. R., Peters R. J. 2007. A Single Residue Switch Converts Abietadiene Synthase into a Pimaradiene Specific Cyclase. *Journal of The American Chemical Society.* **129**(51): 15736-15737.

Williams D. C., Van Frank R. M., Muth W. L., Burnett J. P. 1982. Cytoplasmic inclusion bodies in *Escherichia coli* producing biosynthetic human insulin proteins. *Science*. **215**:687-689.

Yabuta T. 1935. Biochemistry of the *bakanae* fungus of rice. *Agriculture and Horticulture*. **10**:17–22.

Yabuta T., Sumiki Y. 1938. The crystallization of gibberellins A and B. *Journal of the Agricultural Chemical Society of Japan*. **14**:1526.

Zhang Y. M. *et al*. 2011. Biodiversity and biogeography of rhizobia associated with soybean plants grown in the North China Plain. *Applied Environmental Microbiology.* **77**:6331-6342.

Zi J., Mafu S., Peters R. J. 2014. To gibberellins and beyond! Surveying the evolution of (di)terpenoid metabolism. *Annual Reviews in Plant Biology.* 65:259-286.