

CHARACTERIZATION OF
RAFFINOSE SYNTHASE GENES IN SOYBEAN

A Dissertation
presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

by
EMILY DIERKING
Dr. Kristin Bilyeu, Dissertation Supervisor
May 2009

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation titled:

CHARACTERIZATION OF RAFFINOSE SYNTHASE GENES IN SOYBEAN

presented by Emily C. Dierking

a candidate for the degree of Doctor of Philosophy

and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Kristin Bilyeu

Dr. Dale Blevins

Dr. Michael McMullen

Dr. Emmanuel Liscum

Dr. Thomas Quinn

DEDICATION

To my husband, Ryan, who has always been by my side.

To my family and friends, who for some reason, have never given up on me.

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Dr. Kristin Bilyeu. Through this life-changing experience, you have imparted to me the essential requirements of conducting sound scientific experiments, analyzing the results, and effectively communicating through scientific writing. I not only want to thank you for the incredible professional guidance over my research, but also for giving me personal advice and support when I needed it most. I very much respect you as a scientist and a person, and I thank you for helping me grow through this process.

A special thanks is necessary to the Bilyeu lab members, each of you helped in different ways, both scientific as well as friendly support only those with similar experiences can offer. I would also like to thank my doctoral committee, Drs. Mannie Liscum, Dale Blevins, Mike McMullen, and Tom Quinn. Each of you brought a specific expert perspective to the committee, gave me support, and put up with my unique sense of humor.

I could not have been successful without the support of the following people: Dr. Mark Ellersieck for assistance in statistically analyzing my segregating population and Paul Little for technical assistance and expertise in performing HPLC analysis and screening of the fast neutron mutagenized population. I would like to send a special thanks to Mandy Reinerd and Ashley Tetlow who were undergraduate students in the lab and gave an unending desire to assist the research of the lab and were especially instrumental in designing and carrying out the germination experiments. Thank you both for your technical support and your friendship.

I would like to extend appreciation to Dr. Lin Randall, Angie Lilly, Dr. Chunfeng Mao and other members of the Membrane Laboratory who helped attempt recombinant expression of the raffinose synthase protein. You know more tricks than I thought possible, and even though our efforts were not successful I respect your lab as a group of exceptional scientists.

TABLE OF CONTENTS

TITLE PAGE.....	i
ACKNOWLEDGEMENTS.....	ii
CHAPTER	
1. Literature Review.....	1
Soybean history	
Traditional uses of soybean	
Advantages of a diet high in soy products	
Anti-nutritional components of soy	
Soy foods	
Soy feeds	
Soybean seed composition	
Carbohydrate composition of soybean meal and monogastric animals	
Sources of variation in soybean raffinose and stachyose seed composition	
Soybean (MIPS) mutants	
Plant introduction line with low raffinose	
Genetic diversity screens for desirable seed phenotypes	
Raffinose synthase	
Raffinose synthases in other plant species	
Soybean raffinose synthases	
Soybean stachyose synthases	
Role of raffinose in imbibition	
Role of raffinose and stachyose in seed germination	
Sucrose utilization by the plant	
Complexity of the soybean genome	
Figures.....	19
Literature Cited.....	22
2. Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype.....	27
Summary.....	28
Introduction.....	29
Methods and Materials.....	33
Plant Materials	
Population Development	

Backcrossing PI 200508 RS2 Allele into Williams 82
 Sequencing of Plant Raffinose Synthases
 Expression Analysis of Candidate Genes
 Allele-Specific Molecular Marker Assay Development
 Oligosaccharide Phenotype Determination by HPLC

Results.....	38
Identification of Candidate Genes	
Expression Analysis by Quantitative RT-PCR	
Genomic Structures of Putative Raffinose Synthase Genes	
Identification of Polymorphic Alleles of Candidate Genes	
Association of Candidate Gene Genotype to Seed Raffinose and Stachyose Phenotype	
Analyzing Oligosaccharide Content in a Backcross	
Discussion.....	45
Figures.....	49
Literature Cited.....	58
3. Raffinose and stachyose metabolism are not required for efficient soybean seed germination.....	61
Summary.....	62
Introduction.....	63
Methods and Materials.....	66
Plant Material	
HPLC Oligosaccharide Analysis	
Germination Experiments	
Results and Discussion.....	68
Figures.....	72
Table.....	77
Literature Cited.....	79
4. Candidate soybean raffinose synthase gene RS3 does not associate with altered seed oligosaccharide phenotype.....	81
Summary.....	82

Introduction.....	83
Methods and Materials.....	85
Population Development	
Allele-Specific Molecular Marker Assay	
Oligosaccharide Phenotype Determination by HPLC	
Results.....	87
Discussion.....	89
Figure.....	91
Literature Cited.....	93
5. Reverse genetics technique, TILLING, utilized to identify novel soybean raffinose synthase mutant.....	95
Summary.....	96
Introduction.....	97
Methods and Materials.....	100
Population Development	
Population Screening	
EMS mutagenized population	
Fast Neutron mutagenized population	
Allele-Specific Molecular Marker Assay Development	
Oligosaccharidie Phenotype Determination by HPLC	
Plant Growth Conditions	
Results.....	105
Screening of EMS mutagenized population	
Screening of Fast Neutron mutagenized population	
Discussion.....	108
Figures.....	109
Literature Cited.....	113
APPENDIX	
1. Recombinant Protein Expression.....	115
Recombinant expression in <i>E. Coli</i>	
Constructs	

Strains and induction	
Growth Conditions	
Harvest	
Solubility and purification	
NEB PURExpress	
Recombinant expression in <i>Pichia pastoris</i>	
Stable transformation of <i>Arabidopsis thaliana</i>	
Biochemical assay for enzymatic activity	
Positive control assay	
Recombinant enzyme activity assay	

Figures.....	125
Table.....	140
Literature Cited.....	142
Vita.....	143

LIST OF ILLUSTRATIONS

LIST OF FIGURES

Figure	Page
1.1	Raffinose production is part of the glucose metabolism pathway.....20
1.2	Structure of sucrose, raffinose, and stachyose illustrating similarity between the oligosaccharides.....21
2.1	Boxshade output of a ClustalW sequence alignment of putative <i>Glycine max</i> raffinose synthases..... 50
2.2	Relative expression of putative raffinose synthases for multiple soybean tissues determined by quantitative RT-PCR.....53
2.3	Schematic representation of the genomic structure of putative and known plant raffinose synthases..... 54
2.4	Raffinose synthase amino acid sequence alignments in the regions surrounding polymorphisms in the RS2 gene from PI 200508..... 55
2.5	Phenotype to genotype association of a soybean population segregating for the low raffinose/stachyose phenotype.....56
3.1	Germination of wild-type (WT) and Low raffinose and stachyose (Low RFO) soybean seeds in water (control) or DGJ from 0 to 88 hours after imbibition.....73
3.2	Germination of Williams 82 soybean seeds in water (control), sucrose, sucrose + DGJ, galactose, galactose + DGJ, or DGJ from 0 to 88 hours after imbibition....75
3.3	Germination of Williams 82 seeds in water (control), sucrose, sucrose + DGJ, galactose, galactose + DGJ, and DGJ at 23 hours after imbibition.....76
4.1	Phenotype to genotype association of a soybean population (Williams 82 X PI 200508) set for the RS2 gene and segregating for the RS3 gene..... 92
5.1	Raffinose synthase amino acid sequence alignments in the regions surrounding the induced mutagons in the RS2 gene..... 110

5.2	Phenotype to genotype association of segregating M _{2:3} seeds from line 397.....	111
5.3	Phenotype to genotype association of a homozygous soybean population derived from the mutagenized line 397.....	112
A.1A	Gene sequence for RS2 including restriction sites for directional cloning.....	126
A.1B	Celtek Genes order of RS2 in pBluescript II SK(+) vector.....	128
A.2A	Gene sequence for RS3 including restriction sites for directional cloning.....	129
A.2B	Celtek Genes order of RS3 in pGH vector.....	131
A.3	Vector map and polylinker of <i>E. coli</i> expression vector pET-14b as provided by manufacturer (Novagen).....	132
A.4	Polylinker of pMAL-c4x as provided by manufacturer (NEB).....	133
A.5	Candidate gene RS2 in pPICza-A, <i>Pichia pastoris</i> expression vector (Invitrogen)..	134
A.6	Vector map entry vector pENTR-3C for stable transformation of Arabidopsis (Invitrogen).....	135
A.7	Representative results from positive control assay for raffinose synthase and stachyose synthase activity from developing seeds.....	136
A.8	Representative results from enzymatic activity assay for recombinant RS2 expressed in <i>Pichia pastoris</i>	138

LIST OF TABLES

Table

3.1	Oligosaccharide content of wild-type and low raffinose and stachyose dry seeds	78
A.1	Growth conditions for recombinant protein expression in <i>E. coli</i>	141

CHAPTER 1

Literature Review

Soybean [*Glycine max* (L). Merrill] represents one of the most important economic crops in the United States. It ranks third in grain production behind corn (*Zea mays*) and wheat (*Triticum aestivum*), and second only to corn in value (Sleper and Poehlman, 2006). The oil, protein, and carbohydrate composition of the soybean seed generally determines its use. Seeds of soybean cultivars in the United States have an average composition of 20% oil, 40% protein, and 15% soluble carbohydrates (Hsu et al., 1973). Soybean meal is a major component of the diets of monogastric animals and its usefulness is determined, in part, by the carbohydrate component.

Soybean history: The domestication of the soybean is thought to have occurred as early as the 11th century B.C. in Northeastern China. At that time, the uses for soybean included fresh, fermented, and dried food products. It was not until the late 1700's that this crop was introduced to the United States. At that time, Samuel Bowen, an East India Company seaman, brought soybeans to Savannah, Georgia from China and requested they be grown on the farm of the Surveyor General, Henry Yonge. Shortly after, Bowen began growing soybeans on his farm in nearby Thunderbolt. By 1767, he had been issued a patent for soy sauce and vermicelli (Hymowitz, 1990; Hymowitz and Shurtleff, 2005). However, soybean was not introduced to the Midwest United States until the mid-1800's. For many years, soybean was grown mainly as a forage crop for hay, or as a green fertilizer to improve soil structure and fertility once plowed under (Hymowitz, 1990).

Traditional uses of soybean: The uses of soybean meal and soybean components vary widely. The focus has remained mainly on the protein and oil components of the

seed. Generally, commodity soybeans have a characteristic 40% protein and 20% oil content based on dry weight (Lui, 1997). Following extrusion of the oil component, most soybeans are ground into a meal to be utilized as feed for poultry or swine. Increasingly, soybean products are also being used to enhance food intended for human consumption. This can take many forms, including soy flour, protein, isolates, and concentrates. The oil component of the soybean represents the predominant form of vegetable oil, accounting for nearly 71% of all edible fats and oil production in the US (www.soystats.com). Soybean oil can be found in a wide variety of consumables, including salad and cooking oils, frying fat, and margarine (www.soystats.com). Traditionally, soybeans have been grown in the United States for consistent protein and oil components as well as uniform seed size. Producers generally choose a soybean cultivar in order to maximize seed yield.

Advantages of a diet high in soy products: Soybeans have many common advantages of legumes species such as peas, beans, lentils, and peanuts included in a diet. Legumes are generally high in protein, low in fat, and represent an excellent source of dietary fiber (Messina, 1999). Legumes are also an excellent source of micronutrients such as folate, iron (Fe), zinc (Zn), and calcium (Ca) (Messina, 1999); soybeans are especially rich in Ca, and this is one reason a diet rich in soy has benefits for bone health (Messina, 1999). Non-nutritive components such as trypsin and chymotrypsin inhibitors found especially in soybeans have been hypothesized to be anti-cancer agents (Kennedy, 1995; Kennedy and Szuhaj, 1994; Kennedy, 1998).

Isoflavones, a subclass of flavonoids, present in soybean meal have a number of positive health benefits. A single serving of traditional soy foods provides ~ 25-40 mg of

isoflavones (Coward et al., 1993; Wang and Murphy, 1994). There are three major isoflavones found in soybean: diadzein, genistein, and glycitein. These phytochemicals may play a role in the prevention or treatment of some forms of cancer, osteoporosis, heart disease, and may also relieve some menopausal symptoms due to their weak estrogenic properties (Messina, 1999).

Anti-nutritional components of soy: Phytate or phytic acid is considered a highly important storage molecule in seeds and seedlings; however, it is considered an anti-nutritional component of soybean meal in terms of digestion by monogastric animals. These animals lack the phytase enzyme required to break down phytate to bioavailable phosphorous (P) and release the bound minerals. Therefore, much of the nutritional quality is lost and excreted, leading to P pollution in many watersheds across the Midwest (Cromwell and Coffey, 1991). To offset the lost nutrients, manufacturers supplement the meal with expensive, non-renewable P (Brinch-Pedersen et al., 2002).

Soybean meal contains protease inhibitors, such as the Kunitz inhibitor, and the Bowman-Birk inhibitor (Deak et al., 2008). The Kunitz inhibitor specifically inhibits trypsin (Deak et al., 2008) and represents up to 91% of the trypsin inhibitor activity in the seed (Johnson et al., 1980). The Bowman-Birk inhibitor can inhibit both trypsin and chymotrypsin (Deak et al., 2008), but represents only a small portion of total seed trypsin inhibitor activity. Although these inhibitors have been hypothesized to be involved in cancer-protection (Kennedy, 1995; Kennedy and Szuhaj, 1994; Kennedy, 1998), they restrict protein digestion by monogastric animals; therefore heat inactivation of these inhibitors is necessary during processing.

Perhaps one of the most important measures of the nutritive value of soybean meal is metabolizable energy. There are three major carbohydrates present in soybean meal which contribute to this measure: sucrose, raffinose and stachyose. While sucrose is utilized by monogastric animals as metabolizable energy, raffinose and stachyose cannot be digested due to a lack of α -galactosidase activity in the gut. Removal of raffinose and stachyose from soybean meal has been reported to increase the metabolizable energy of soybean meal (Coon et al., 1990), as well as reduce the adverse side effects produced by the digestion of the oligosaccharide by microbes in the guts of monogastric animals (Zuo et al., 1996).

Soy foods: For some time now, soybean has been a major food source in Asian countries. More recently, it has gained popularity around the world. Due to the higher levels of consumption, soy foods are more easily found in Asian countries, however their availability in the United States is only increasing as positive health benefits are identified. Foods containing soy can take a number of distinctly different forms including soymilk, tofu, edamame, infant formula, miso, natto, soy flour and soy sauce (www.soyfoods.com). Soy protein isolates are used to supplement a variety of processed foods for the purpose of increasing the protein content of the food.

Soymilk is made from soybeans that have been soaked, ground, and strained (www.soyfoods.com). Soymilk is a good substitute for cow's milk because it is lower in fat and contains phyto-estrogens and can be used in cooking. Tofu is a soft cheese-like food made from curdling warm soymilk (www.soyfoods.com). Tofu is rich in protein and thus is often used as a meat substitute in Asian diets or in vegetarian diets; it is also

rich in vitamin B (www.soyfoods.com). Tofu is available in a number of forms which can be used in many different dishes.

The Japanese favor a specific soy product, miso, which is a rich, salty condiment (www.soyfoods.com). Miso is used as flavoring in a number of dishes or sauces including soups and dressings (www.soyfoods.com). Natto is made from fermented, cooked whole soybeans and is more easily digested than whole soybeans (www.soyfoods.com). It is traditionally used as a topping or in soups with miso and vegetables (www.soyfoods.com). Soybeans can also be consumed in a non-processed form, a perfect example is edemame which is soybeans that are harvested green. At this stage, the soybeans are fully expanded, but have not begun to dry down and can be consumed as a high protein, high fiber snack or side dish when steamed in salt water (www.soyfoods.com).

Soy flour is made from roasted soybeans which have been processed into a fine powder and is most often used to boost the protein content of recipes which normally include flour (www.soyfoods.com). One specific advantage of including soy flour in recipes is that it is free from gluten. A more course form of soy flour called soy grits can serve as a substitute for flour, but are more commonly found mixed with rice and other grains to increase the protein content of the food (www.soyfoods.com). Soy flours or soy protein isolate powders are used as a base in a number of soy products including infant formula (www.soyfoods.com).

Perhaps the most widely recognized soy product is soy sauce. Soy sauce can also come in the form of tamari, shoyu, or teriyaki and is a dark brown liquid resulting from soybeans that have been fermented (www.soyfoods.com). Tamari is made solely from

soybeans and is a by-product of miso, while both shoyu and teriyaki include other ingredients (www.soyfoods.com). Shoyu is a mixture of soybean and wheat, while teriyaki includes sugar, vinegar and spices (www.soyfoods.com).

Soy feeds: Soy is a major component of animal feeds in the United States. In most cases this is in the form of soybean meal, although there are advantages to supplementing feed with soy hulls as well. Soy hulls have less crude protein, but similar total digestible nutrients to soybean meal (Rayburn, 1999). Soybean meal serves as an excellent source of protein, fiber, and nutrients in a wide variety of animal feed.

Soybean seed composition: Soybean seeds are generally high in protein and oil; this characteristic of their composition is what gives this crop a long list of varied uses. Protein represents approximately 30 to 50% of the seed weight; up to 80% of the total protein is storage proteins (Murphy, 2008). The storage proteins are comprised of two main classes: vicilins and legumins (Murphy, 2008). These proteins only have one physiological function which is to provide nitrogen in the form of amino acids to the germinating seeds (Murphy, 2008). Soybeans are also high in oil and for this reason can be considered an oil seed. They are unique, however, in that no other oil seeds have as high protein content as soybeans (Johnson, 2008).

The oil, or lipid, component of soybean seeds represents 20% of the total seed weight and is mostly found in the cotyledons (Gerde and White, 2008). The lipid component of soybeans has varied functions such as inclusion in membranes, energy reserve and a solvent medium (Gerde and White, 2008). A number of fatty acids are present in soybeans including palmitic, stearic, oleic, linoleic and linolenic acid (Gerde and White, 2008). These fatty acids have received much focus in terms of making the oil

component better suited for frying (low linolenic; higher saturation) or human health (increased oleic).

The carbohydrate component of soybean seeds represents approximately 30-35% of the seed weight and can even be higher in soybean meal (Middelbos and Fahey, 2008). Carbohydrates are important in terms of cell structure in the plant and metabolizable energy in terms of soybean meal for animal digestion. The carbohydrate component of soybean meal is comprised of two major classes, structural and non-structural carbohydrates (Middelbos and Fahey, 2008). The structural carbohydrates are those collectively called dietary fiber and are the cell wall polysaccharides, noncellulose polysaccharides, and structural nonpolysaccharides such as lignin (Middelbos and Fahey, 2008). These carbohydrates are generally insoluble and are harder to digest or are not digestible.

The soluble portion, or the non-structural carbohydrates, is comprised of three major oligosaccharides: sucrose, raffinose, and stachyose (Openshaw and Hadley, 1978). Of the three oligosaccharides, only sucrose is nutritionally useful and can be fully digested by monogastric animals. Raffinose and stachyose cannot be digested due to the lack of galactosidase activity in the stomach and small intestine of monogastric animals. When consumed, these carbohydrates are instead fermented by microbes present in the gut causing flatulence and discomfort; ultimately leading to poor weight gain and a subsequent loss in market value in livestock operations.

Carbohydrate composition of soybean meal and monogastric animals: Removing the anti-nutritional units, raffinose and stachyose, from soybean meal has been reported to increase the metabolizable energy of the diet by as much as 20% (Coon et al., 1990).

The effects of raffinose and stachyose in diets have been studied in pigs (Smiricky et al., 2002), dogs (Zuo et al., 1996), chickens (Parsons et al., 2000), and humans (Suarez et al., 1999). Raffinose and stachyose present in soybean meal affect the uptake of other nutrients, as well as produce anti-nutritional effects such as flatulence and diarrhea (Zuo et al., 1996). A reduction in these anti-nutritional units has potential to greatly increase the availability of nutrients for digestion in soybean meal by monogastric animals.

Sources of variation in soybean raffinose and stachyose seed composition:

Soybean (MIPS) mutants: In the soybean, *myo*-inositol is responsible for a large number of biological functions including signal transduction, cell-wall biogenesis, growth regulation and phosphate biosynthesis (Irvine and Schell, 2001). *Myo*-inositol, the precursor for all inositol containing compounds, is biosynthesized by the *D*-*myo*-inositol-3-phosphate synthase (MIPS) which converts glucose-6-phosphate to *D*-*myo*-inositol 3-phosphate. Because phytic acid is considered an anti-nutritional component of soybean meal, the MIPS enzyme and its key role in the biosynthesis of phytic acid is of high interest. Phytic acid is also the primary storage component of seed phosphorus and can account for as much as 2% of the dry seed weight (Raboy et al., 1984). In line with reducing the seed phytic acid levels, identifying low phytic acid mutants has been the focus of much research; mutants have been identified in a number of crops (Raboy, 2002). The soybean mutant LR33 is characterized by a 50% reduction in seed phytic acid levels; this line also has elevated sucrose and a reduction in both raffinose and stachyose (Hitz et al., 2002; Kerr and Sebastian, 2000). The key component in both the reduction in phytic acid levels and the altered oligosaccharide phenotype is the molecule *myo*-inositol. *Myo*-inositol is the precursor for galactinol use in both the biosynthesis of raffinose and

stachyose as well as phytic acid. The LR33 phenotype is the result of a missense mutation in one soybean MIPS gene (Hitz et al., 2002). Since LR33 is not available for public breeding programs, attempts have been made to identify or induce similar phenotypic mutants. Recently, two additional MIPS mutants have been identified in soybean which have a 66.6% increase in available phosphorus and a 46.3% decrease in phytic acid levels (Yuan et al., 2007).

Plant introduction line with low raffinose: A plant introduction line, PI 200508, has been identified as having reduced levels of raffinose and elevated levels of sucrose (Kerr and Sebastian, 2000). Generally, plant introduction lines are un-adapted, non-production lines that represent the diversity of a subset of germplasm. These lines have not been cultivated and generally have a number of undesirable physiological traits such as insect susceptibility, inappropriate maturity, pod shattering, low seed weight and poor germination. It is for these reasons that single desirable traits from plant introduction lines are bred into standard cultivars or elite soybean germplasm.

Characterization of PI 200508 was carried out by Hitz et al. (2002) and a physiological analysis of this line was subsequently conducted by Neus et al. (2005). This research indicated that PI 200508 has greatly reduced levels of raffinose and stachyose, elevated levels of sucrose and galactinol, and a 25-fold reduction in raffinose synthase enzyme activity (Hitz et al., 2002). It does not appear that the vigor of PI 200508 is affected as Neus et al. (2005) reported that lines derived from PI 200508 and a high-yielding variety had no significant differences in a number of seed quality traits which included field emergence, seed yield, maturity, and fatty acid content when compared to sibling lines with a standard raffinose phenotype. The LR28 designation is

synonymous with PI 200508, as this plant introduction line was the source of the low raffinose and stachyose phenotype (Kerr and Sebastian, 2000; Kerr and Sebastian, 2003).

In order to further characterize the low raffinose and stachyose seed phenotype, a population was developed by crossing LR28 with another line (X3337), and the result was two low raffinose and stachyose phenotypes (Kerr and Sebastian, 2003). One is the same as LR28, referred to as low1, has reduced raffinose and stachyose; the second phenotype, referred to as low2 has a low raffinose phenotype with a significantly lower stachyose phenotype (Kerr and Sebastian, 2003). The authors attribute the low2 phenotype to the low1 gene phenotype in addition to acting modifiers which further reduce the seed raffinose and stachyose (Kerr and Sebastian, 2003).

Genetic diversity screens for desirable seed phenotypes: Since some of the characterized seed phenotypes are unavailable for public breeding (MIPS), two distinct methods can be used for identifying public sources of the desirable seed phenotypes: forward genetics and reverse genetics. Forward genetics can be used to identify the gene behind a phenotype, while reverse genetics can be utilized to identify the phenotype resulting from a gene mutation.

The MIPS mutation was identified using a forward genetics strategy. The LR33 phenotype (Kerr and Sebastian, 2000) which is characterized by a 50% reduction in phytic acid, an increase in sucrose, and a decrease in both raffinose and stachyose is the result of a single base pair polymorphism resulting in an amino acid substitution in a conserved region of the MIPS gene (Hitz et al., 2002).

Reverse genetics, a relatively new tool, allows the identification of a phenotypic mutant by first identifying a mutation in the gene sequence. Traditionally, genetics or

forward genetics used the genes underlying an observed phenotype. Targeting Induced Local Lesions IN Genomes (TILLING) is a reverse genetics technique that serves as a high throughput method to identify unique mutations within target genes which have potential to change gene expression and/or function (Colbert et al., 2001; McCallum et al., 2000). The final step in mutation identification is confirmation of a mutant phenotype by an association analysis.

Aside from the potential to screen an entire population for a mutation in a gene of interest, a chemically mutagenized population has many benefits. Chemical mutagenesis, by either ethyl-methanesulfonate (EMS) or N-nitroso-N-methylurea (NMU) usually causes single nucleotide polymorphisms (SNPs) which are useful in developing an allelic series as well as novel alleles in gene targets (Cooper et al., 2008). These point mutations are extremely useful in studying gene function as well as their use in crop improvement (Cooper et al., 2008). This is in contrast to irradiation-induced mutagenesis which is characterized by deletion mutations. The radiation-induced mutagenesis can take the form of x-ray, fast neutron, or gamma irradiation. The deletion mutations caused by irradiation represent the opportunity to produce gene knock-outs for use in characterizing gene function or resulting phenotypes. Fast neutron induced deletions from the Arabidopsis population ranged in size from 0.8 kb to 12 kb (Li et al., 2001). In some cases, multiple deletions could be found for a single locus (Li et al., 2001).

Raffinose synthase: The raffinose synthase enzyme belongs to a group of hydrolase family enzymes which mediates the galactosyl transfer from galactinol to sucrose. This transfer produces the three ring molecule raffinose. A by-product formed from this reaction is the molecule *myo*-inositol. The raffinose synthase enzymatic

reaction is also related to glucose metabolism (Figure 1.1). Similarly, stachyose is formed by the action of stachyose synthase which combines raffinose and galactinol. Thus, raffinose synthase and stachyose synthase share one identical substrate, galactinol, and a second similar substrate, sucrose or raffinose, respectively (Figure 1.2).

Raffinose synthases in other plant species: A number of gene sequences have been annotated as raffinose synthases, however complete characterization of these genes is lacking. A search of the GenBank database revealed that raffinose synthase sequences have been annotated for pea (AJ426475), cucumber (AF073744), maize (Q575Z6, Q575Z7, Q575Z8), rice (BAD321.1), grape (AM430487), and *Arabidopsis* (AC007138); but biochemical confirmation and molecular characterization has only been completed for pea (Peterbauer et al., 2001) and rice (Li et al., 2007). Amino acid sequence identity ranges between 50 and 70% among raffinose synthase sequences of the various plant species. The tDNA insertion in the *Arabidopsis thaliana* putative raffinose synthase gene mutant was not extensively characterized; however, leaf raffinose levels were shown to be decreased when compared to wild type (Zuther et al., 2004).

Soybean raffinose synthases: Soybean raffinose synthase activity has been studied by Hitz et al. (2002), but raffinose synthase gene sequences have not been cloned. To date, there are raffinose synthase enzyme activity data from crude lysates from LR 28 seeds taken during the late pod filing stage, which is when raffinose and stachyose accumulate (Hitz et al., 2002). A low raffinose and stachyose phenotype has also been described for the line LR 28 (PI 200608) (Kerr and Sebastian, 2000). Additionally two sequences exist which are annotated as soybean raffinose synthase enzyme on the Genbank database. However, there are no biochemical data to support the annotation.

Some biochemical data exists on pea raffinose synthase; this enzyme is highly specific for the substrates, sucrose and galactinol, and is unable to carry out the synthesis of other oligosaccharides such as stachyose (Peterbauer, 2001).

Soybean stachyose synthases: Until recently, there was no evidence in either the literature or sequence databases that stachyose synthase exists in soybean. The recent release of the Glyma1 assembly contained a sequence which appears to be a putative stachyose synthase. Other published experiments indicate that there is an adzuki bean stachyose synthase which is capable of catalyzing the synthesis of both stachyose and verbascose (Peterbauer et al., 1999; Peterbauer, 2001). This is the only stachyose synthase enzyme which has been biochemically characterized. The adzuki bean stachyose synthase and other candidate stachyose synthase sequences share high similarity with known raffinose synthase sequences, however it seems that the stachyose synthases are distinguishable from raffinose synthases by a characteristic central insertion of 70-80 amino acids (Peterbauer et al., 1999; Peterbauer, 2001); this is also true of the recently identified soybean stachyose synthase.

Role of raffinose in imbibition: Raffinose has previously been thought to be involved in seed protection during the desiccation process (Obendorf, 1997) by stabilization of membranes (Crowe et al., 1987). It was later discovered that raffinose content and seed storability could not be correlated (Bentsink et al., 2000). However, Neus et al. (2005) studied lines derived from a cross between PI 200508 and high-yielding varieties (Kerr and Sebastian, 2000), and found no significant differences in seed quality traits between the derived lines and cultivars with a wild-type raffinose phenotype. These seed quality traits included field emergence, seed yield,

maturity, and fatty acid content. Recently, a low raffinose and stachyose line and a low raffinose, stachyose, and phytin line were tested for their sensitivity to imbibitional chilling (Obendorf et al., 2008). The results indicated that lines low in raffinose and stachyose were similar to control lines, with no sensitivity, while the lines low in raffinose, stachyose, and phytin were sensitive to imbibitional chilling (Obendorf et al., 2008). In addition, the imbibitional chilling tolerance of the low raffinose and stachyose soybeans was speculated to be due to the higher accumulation of cyclitol α -galactosides in the embryo (Obendorf et al., 2008).

Role of raffinose and stachyose in seed germination: Although the exact function of raffinose oligosaccharides in germinating seeds is largely unknown, it is clear that seeds require a large amount of energy during germination (Bewley and Black, 1994). This energy is hypothesized to come from stored carbohydrates; sucrose, raffinose and stachyose oligosaccharides are the most abundant of the soluble sugars (Peterbauer et al., 2001) but account for only a small portion of the total carbohydrates present in the seeds (Ziegler, 1995). Additionally, raffinose and stachyose have previously been thought to be involved in seed protection during the desiccation process (Bailly et al., 2001; Black et al., 1996; Obendorf, 1997) by stabilizing membranes (Crowe et al., 1987). However, each of the proposed functions of raffinose and stachyose in seeds has not yet been determined experimentally (Bentsink et al., 2000; Gurusinghe and Bradford, 2001). Since the breakdown of seed components during germination has been studied extensively, it is known that protein, oils and polymeric carbohydrates each break down slowly over a period of several days (Bewley and Black, 1994). This period extends

beyond germination and thus cannot be solely responsible for fulfilling the energy demands of the germinating seed.

Blöchl et al. (2007) directly tested the effect of inhibition of the breakdown of raffinose and stachyose on seed germination in pea. The results of this study indicated that pea seeds exhibited a significant delay in germination when α -galactosidase activity was inhibited by 1-deoxygalactonojirimycin (DGJ). Further, the germination delay in DGJ treated pea seeds could be partially overcome by the addition of galactinol (Blöchl et al., 2007).

Measures, such as germination, field emergence and seed size, have been conducted on other lines with altered seed phenotypes in soybean. Seeds with low phytate (*mips*) and those with transgenically enhanced phytase activity (CAPPA) did not exhibit significant differences in field emergence or seed number per plant (Bilyeu et al., 2008). Alternatively, there was a significant reduction in germination in the line 'CAPPA' as compared to its progenitor line Jack (Bilyeu et al., 2008).

Sucrose utilization by the plant: A reduction in raffinose is accompanied by an increase in seed sucrose and galactinol levels. The conversion of sucrose to raffinose and stachyose works to maintain a sucrose gradient toward the sink tissues. Raffinosaccharides are considered to be a transport molecule in many plants (Ayre et al., 2003) and are included in the polymer trapping model for sucrose transport in the phloem; therefore, it is important to consider any problems which may be associated with disruption of the concentrations of these molecules. However, lines derived from a cross between the low raffinose line, PI 200508 and a high-yielding line studied by Neus et al. (2005) appeared normal through germination and development; therefore soybean plants

must be functioning sufficiently with significantly reduced raffinose oligosaccharides. Seed sucrose content was increased in PI 200508 (Hitz et al., 2002) indicating that whether or not transport to the sink tissue is as efficient as in wild type, sucrose transport is sufficient enough to result in elevated sucrose levels in the seed.

Complexity of the soybean genome: Soybean is a diploid plant which carries the chromosome number $2n=40$. Even though soybean is cross-fertile, it is considered a naturally self-pollinating plant. The genome has likely undergone at least two duplications (Shoemaker et al., 2006; Shoemaker et al., 1996) and it is for this reason that genetic/molecular studies involving soybean are considerably complex. It is likely to find a number of genes which could encode a single enzyme capable of catalyzing one biochemical step. Genome duplications introduce the possibility of genetic redundancy, which in some cases means that a combination of mutant alleles is required to produce a unique phenotype. However, the recent release of the Glyma1 assembly of the soybean genome sequence (Soybean Genome Project, DoE Joint Genome Institute) and the EST collection have made identifying candidate genes far more reasonable.

The purpose of this project is ultimately to increase the value of soybean meal by elevating the metabolizable energy for digestion by monogastric animals at the expense of raffinose and stachyose. We identified four candidate genes; one of the genes completely associated with the low raffinose and stachyose phenotype observed in PI 200508. The contribution of this gene was independently confirmed by the identification of an EMS-induced mutant allele. This discovery represents the first characterized EMS-mutagenized soybean line with an associated phenotype. Perfect molecular markers were

identified for each of the unique alleles of this candidate gene for efficient breeding in to production quality lines. In addition, the long-held assumption that raffinose and stachyose are required for efficient seed germination was tested. It does not appear that raffinose and stachyose are required for soybean seed germination indicating that the desirable seed oligosaccharide phenotype of elevated sucrose and decreased raffinose and stachyose does not appear to interfere with necessary seed performance.

FIGURES

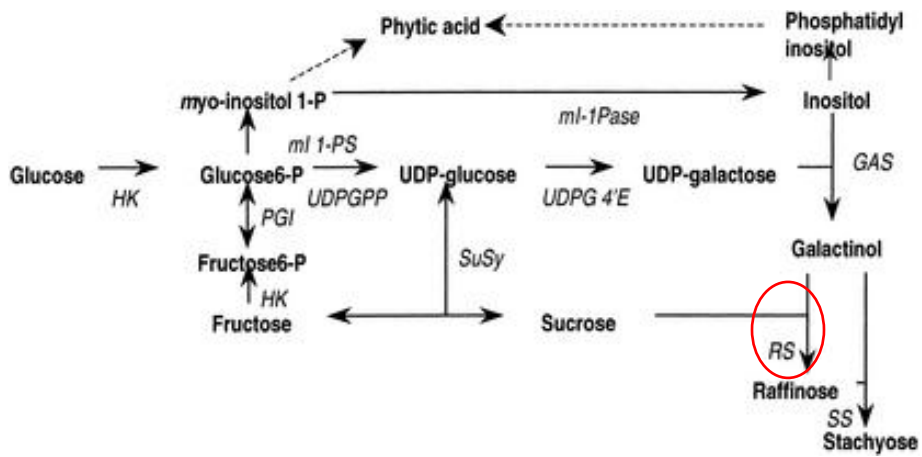


Figure 1.1: Raffinose production is part of the glucose metabolism pathway. Raffinose is formed by a galactosyl transfer from sucrose to galactinol mediated by raffinose synthase (RS) (Hitz et al., 2002). Dashed lines indicated pathways which are not defined.

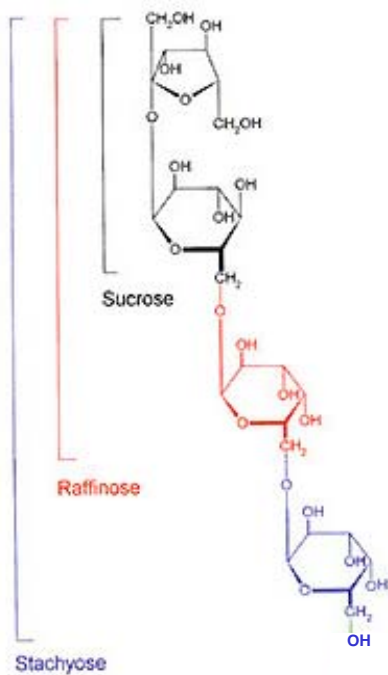


Figure 1.2: Structure of sucrose, raffinose, and stachyose illustrating similarity between the oligosaccharides (modified from <http://www.food-info.net/uk/qa/qa-fp160.htm>).

LITERATURE CITED

- Ayre, B.G., F. Keller, and R. Turgeon. 2003. Symplastic continuity between companion cells and the translocation stream: long-distance transport is controlled by retention and retrieval mechanisms in the phloem. *Plant Physiology* 131:1518-1528.
- Bailly, C., C. Audigier, F. Ladonne, M.H. Wagner, F. Coste, F. Corbineau, and D. Come. 2001. Changes in oligosaccharide content and antioxidant enzyme activities in developing bean seeds as related to acquisition of drying tolerance and seed quality. *Journal of Experimental Botany* 52:701-708.
- Bentsink, L., C. Alonso-Blanco, D. Vreugdenhil, K. Tesnier, S.P. Groot, and M. Koornneef. 2000. Genetic analysis of seed-soluble oligosaccharides in relation to seed storability of *Arabidopsis*. *Plant Physiology* 124:1595-1604.
- Bewley, J.D., and M. Black. 1994. *Seeds: Physiology of Development and Germination*. Second ed. Plenum Press, New York.
- Bilyeu, K.D., P. Zeng, P. Coello, Z.J. Zhang, H.B. Krishnan, A. Bailey, P.R. Beuselinck, and J.C. Polacco. 2008. Quantitative conversion of phytate to inorganic phosphorus in soybean seeds expressing a bacterial phytase. *Plant Physiology* 146:468-477.
- Black, M., F. Corbineau, M. Grzesik, P. Guyi, and D. Come. 1996. Carbohydrate metabolism in the developing and maturing wheat embryo in relation to its desiccation tolerance. *Journal of Experimental Botany* 47:161-169.
- Blöchl, A., T. Peterbauer, and A. Richter. 2007. Inhibition of raffinose oligosaccharide breakdown delays germination of pea seeds. *Journal of Plant Physiology* 164:1093-1096.
- Brinch-Pedersen, H., L.D. Sørensen, and P.B. Holm. 2002. Engineering crop plants: getting a handle on phosphate. *Trends in Plant Science* 7:118-125.
- Colbert, T., B.J. Till, R. Tompa, S. Reynolds, M.N. Steine, A.T. Yeung, C.M. McCallum, L. Comai, and S. Henikoff. 2001. High-throughput screening for induced point mutations. *Plant Physiology* 126:480-484.
- Coon, C.N., K.L. Leske, O. Akavanichan, and T.K. Cheng. 1990. Effect of oligosaccharide-free soybean meal on true metabolizable energy and fiber digestion in adult roosters. *Poultry Science* 69:787-793.
- Cooper, J., B. Till, R. Laport, M. Darlow, J. Kleffner, A. Jamai, T. El-Mellouki, S. Liu, R. Ritchie, N. Nielsen, K. Bilyeu, K. Meksem, L. Comai, and S. Henikoff. 2008. TILLING to detect induced mutations in soybean. *BMC Plant Biology* 8:9.

- Coward, L., N.C. Barnes, K.D.R. Setchell, and S. Barnes. 1993. Genistein, daidzein, and their .beta.-glycoside conjugates: antitumor isoflavones in soybean foods from American and Asian diets. *Journal of Agricultural and Food Chemistry* 41:1961-1967.
- Cromwell, G.L., and R.D. Coffey. 1991. Phosphorus-a key essential nutrient, yet a possible major pollutant - its central role in animal nutrition, *In* T. P. Lyons, ed. *Biotechnology in the Feed Industry*. Alltech Technical Publishers, Nicholasville, Kentucky.
- Crowe, J.H., L.M. Crowe, J.F. Carpenter, and C. Aurell Wistrom. 1987. Stabilization of dry phospholipid bilayers and proteins by sugars. *Biochemistry Journal* 242:1-10.
- Deak, N.A., L.A. Johnson, E.W. Lusas, and K.C. Rhee. 2008. Soy Protein Products, Processing, and Utilization, p. 661-724, *In* L. A. Johnson, et al., eds. *Soybeans: Chemistry, Production Processing, and Utilization*. AOCS Press, Urbana, Illinois.
- Food-info [Online] <http://www.food-info.net/uk/qa/qa-fp160.htm> (verified November 16, 2006).
- Gerde, J.A., and P.J. White. 2008. Lipids, p. 193-228, *In* L. A. Johnson, et al., eds. *Soybeans: Chemistry, Production Processing, and Utilization*. AOCS Press, Urbana, Illinois.
- Gurusinghe, S., and K.J. Bradford. 2001. Galactosyl-sucrose oligosaccharides and potential longevity of primed seeds. *Seed Science Research* 11:121-134.
- Hitz, W.D., T.J. Carlson, P.S. Kerr, and S.A. Sebastian. 2002. Biochemical and molecular characterization of a mutation that confers a decreased raffinose and phytic acid phenotype on soybean seeds. *Plant Physiology* 128:650-660.
- Hsu, S.H., H.H. Hadley, and T. Hymowitz. 1973. Changes in carbohydrate contents of germinating soybean seeds. *Crop Science* 13:407-410.
- Hymowitz, T. 1990. Soybeans: The Success Story, p. 159-163, *In* J. Janick and J. E. Simon, eds. *Advances in new crops*. Timber Press, Portland, OR.
- Hymowitz, T., and W.R. Shurtleff. 2005. Debunking Soybean Myths and Legends in the Historical and Popular Literature. *Crop Science* 45:473-476.
- Irvine, R.F., and M.J. Schell. 2001. Back in the water: the return of the inositol phosphates. *Nature Reviews Molecular Cell Biology* 2:327-338.

- Johnson, L.A. 2008. Oil Recovery from Soybeans, p. 331-376, *In* L. A. Johnson, et al., eds. Soybeans: Chemistry, Production Processing, and Utilization. AOCS Press, Urbana, Illinois.
- Johnson, L.A., C.W. Deyoe, W.J. Hoover, and K.R. Schwenke. 1980. Inactivation of trypsin inhibitors in aqueous extracts of soybeans by direct steam infusion. *Cereal Chemistry* 57:376-379.
- Kennedy, A. 1995. The evidence for soybean products as cancer preventive agents. *Journal of Nutrition* 733-743.
- Kennedy, A., and B. Szuhaj. 1994. Bowman-Birk inhibitor product for use as an anticarcinogenesis agent. United States Patent 5338547. Date Issued: August 16, 1994.
- Kennedy, A.R. 1998. The Bowman-Birk inhibitor from soybeans as an anticarcinogenic agent. *American Journal of Clinical Nutrition* 68:1406-1412.
- Kerr, P.S., and S.A. Sebastian. 2000. Soybean products with improved carbohydrate composition and soybean plants. United States Patent 6147193. Date Issued: November 14, 2000.
- Kerr, P.S., and S.A. Sebastian. 2003. Soybean products with improved carbohydrate composition and soybean plants. United States Patent 6653451. Date Issued: November 25, 2003.
- Li, S., W.D. Kim, S. Kaneko, P.A. Prema, M. Nakajima, and H. Kobayashi. 2007. Expression of rice (*Oryza sativa* L. var. Nipponbare) alpha-galactosidase genes in *Escherichia coli* and characterization. *Bioscience Biotechnology Biochemistry* 71:520-526.
- Li, X., Y. Song, K.S. Century, P. Ronald, X. Dong, M. Lassner, and Y. Zhang. 2001. A fast neutron deletion mutagenesis-based reverse genetics system for plants. *The Plant Journal* 27:235-242.
- Lui, K. 1997. Soybeans: Chemistry, Technology, and Utilization Chapman & Hall, New York.
- McCallum, C.M., L. Comai, E.A. Greene, and S. Henikoff. 2000. Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant Physiology* 123:439-442.
- Messina, M.J. 1999. Legumes and soybeans: overview of their nutritional profiles and health effects. *American Journal of Clinical Nutrition* 70:439-450.

- Middelbos, I.S., and G.C. Fahey, Jr. 2008. Soybean Carbohydrates, p. 269-296, *In* L. A. Johnson, et al., eds. Soybean: Chemistry, Production Processing, and Utilization. AOCS Press, Urbana, Illinois.
- Murphy, P.A. 2008. Soybean Proteins, p. 229-268, *In* L. A. Johnson, et al., eds. Soybeans: Chemistry, Production Processing, and Utilization. AOCS Press, Urbana, Illinois.
- Neus, J.D., W.R. Fehr, and S.R. Schnebly. 2005. Agronomic and seed characteristics of soybean with reduced raffinose and stachyose. *Crop Science* 45:589-592.
- Obendorf, R.L. 1997. Oligosaccharides and galactosyl cyclitols in seed desiccation tolerance. *Seed Science Research* 7:63-74.
- Obendorf, R.L., A.D. Zimmerman, P.A. Ortiz, A.G. Taylor, and S.R. Schnebly. 2008. Imbibitional chilling sensitivity and soluble carbohydrate composition of low raffinose, low stachyose soybean seed. *Crop Science* 48:2396-2403.
- Openshaw, S.J., and H.H. Hadley. 1978. Maternal effects on sugar content in soybean seeds. *Crop Science* 18:581-584.
- Parsons, C.M., Y. Zhang, and M. Araba. 2000. Nutritional evaluation of soybean meals varying in oligosaccharide content. *Poultry Science* 79:1127-1131.
- Peterbauer, T., J. Mucha, U. Mayer, M. Popp, J. Glossl, and A. Richter. 1999. Stachyose synthesis in seeds of adzuki bean (*Vigna angularis*): molecular cloning and functional expression of stachyose synthase. *The Plant Journal* 20:509-518.
- Peterbauer, T., L.B. Lahuta, A. Blochl, J. Mucha, D.A. Jones, C.L. Hedley, R.J. Gorecki, and A. Richter. 2001. Analysis of the raffinose family oligosaccharide pathway in pea seeds with contrasting carbohydrate composition. *Plant Physiology* 127:1764-1772.
- Peterbauer, T. and A. Richter. 2001. Biochemistry and physiology of raffinose family oligosaccharides and galactosyl cyclitols in seeds. *Seed Science Research* 11:185-197.
- Raboy, V. 2002. Progress in breeding low phytate crops. *Journal of Nutrition* 132:503-505.
- Raboy, V., D.B. Dickinson, and F.E. Below. 1984. Variation in seed total phosphorus, phytic acid, zinc, calcium, magnesium, and protein among lines of *Glycine max* and *G. soja*. *Crop Science* 24:431-434.
- Rayburn, E. 1999. Alternative feeds for extending limited feed supplies. West Virginia University Extension Service. [Online]

<http://www.wvu.edu/~agexten/forglvst/alterdrot.htm> (Verified December 4, 2008).

Shoemaker, R.C., J. Schlueter, and J.J. Doyle. 2006. Paleopolyploidy and gene duplication in soybean and other legumes. *Current Opinions in Plant Biology* 9:104-109.

Shoemaker, R.C., K. Polzin, J. Labate, J. Specht, E.C. Brummer, T. Olson, N. Young, V. Concibido, J. Wilcox, J.P. Tamulonis, G. Kochert, and H.R. Boerma. 1996. Genome duplication in soybean (*Glycine* subgenus *soja*). *Genetics* 144:329-338.

Sleper, D.A., and J.M. Poehlman. 2006. *Breeding field crops*. 5th ed. Blackwell Publishing., Ames, Iowa.

Smiricky, M.R., C.M. Grieshop, D.M. Albin, J.E. Wubben, V.M. Gabert, and G.C. Fahey, Jr. 2002. The influence of soy oligosaccharides on apparent and true ileal amino acid digestibilities and fecal consistency in growing pigs. *Journal of Animal Science* 80:2433-2441.

SoyStats [Online]. Available by The American Soybean Association <http://www.soystats.com/> (verified November 10, 2008).

Soyfoods Descriptions [Online]
<http://soyfoods.com/soyfoodsdescriptions/descriptions.html> (verified December 3, 2008).

Suarez, F.L., J. Springfield, J.K. Furne, T.T. Lohrmann, P.S. Kerr, and M.D. Levitt. 1999. Gas production in human ingesting a soybean flour derived from beans naturally low in oligosaccharides. *American Journal of Clinical Nutrition* 69:135-139.

Wang, H., and P.A. Murphy. 1994. Isoflavone content in commercial soybean foods. *Journal of Agricultural and Food Chemistry* 42:1666-1673.

Yuan, F.-J., H.-J. Zhao, X.-L. Ren, S.-L. Zhu, X.-J. Fu, and Q.-Y. Shu. 2007. Generation and characterization of two novel low phytate mutations in soybean (*Glycine max* L. Merr.). *TAG Theoretical and Applied Genetics* 115:945-957.

Ziegler, P. 1995. *Seed development and germination* Marcel Dekker Inc, New York.

Zuo, Y., G.C. Fahey, Jr., N.R. Merchen, and N.L. Bajjalieh. 1996. Digestion responses to low oligosaccharide soybean meal by ileally-cannulated dogs. *Journal of Animal Science* 74:2441-2449.

Zuther, E., K. Buchel, M. Hundertmark, M. Stitt, D.K. Hincha, and A.G. Heyer. 2004. The role of raffinose in the cold acclimation response of *Arabidopsis thaliana*. *FEBS Letters* 576:169-173.

CHAPTER 2

Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype

SUMMARY

Oligosaccharides are an important component of soybean [*Glycine max* (L.) Merrill] meal in terms of metabolizable energy for monogastric animals. Sucrose, raffinose, and stachyose are the three main oligosaccharides present in soybean meal. Of the three, only sucrose is nutritionally useful. When raffinose and stachyose are fermented by microbes present in the gut, the results are flatulence and discomfort which ultimately lead to poor weight gain in livestock. The long term objective of this research is ultimately to increase the nutritional value of soybean meal by elevating the metabolizable energy at the expense of raffinose and stachyose through the manipulation of soybean raffinose synthase, the key enzyme for raffinose and stachyose biosynthesis. The objectives of this research were to develop molecular genetic information about soybean raffinose synthases and to evaluate the candidate raffinose synthase genes in a soybean germplasm accession (PI 200508) that contains low levels of raffinose and stachyose. Our results indicate the soybean genome contains at least two expressed genes similar to other characterized raffinose synthases. A novel allele of one of these putative soybean raffinose synthase genes was discovered from the PI 200508 that completely associates with the low raffinose and stachyose phenotype. Molecular marker assays specific for the PI 200508 allele were developed to allow direct selection for the low raffinose and low stachyose phenotype.

INTRODUCTION

Today, soybean represents one of the most important economic crops in the United States. It is considered similar to corn in acreage and second only to corn in value (Sleper and Poehlman, 2006). The oil, protein, and carbohydrate composition of the soybean seed generally controls its use. Seeds of soybean cultivars in the United States have an average composition of 20% oil, 40% protein, and 15% soluble carbohydrates in dry weights of cotyledons of ungerminated seeds (Hsu et al., 1973). Soybean meal is a major component of the diets of monogastric animals and its usefulness is determined, in part, by the carbohydrate component. The carbohydrate component of soybean meal is comprised of three major oligosaccharides: sucrose, raffinose, and stachyose (Openshaw and Hadley, 1978). Of the three, only sucrose is nutritionally useful and can be fully digested by monogastric animals. Raffinose and stachyose are considered anti-nutritional units because they cannot be digested due to the lack of α -galactosidase activity in the gut of monogastric animals.

Removing raffinose and stachyose from soybean meal has been reported to increase the metabolizable energy of the diet by as much as 20% (Coon et al., 1990). The effects of raffinose and stachyose in diets have been studied in pigs (Smiricky et al., 2002), dogs (Zuo et al., 1996), chickens (Parsons et al., 2000), and humans (Suarez et al., 1999). Generally, raffinose and stachyose are poorly digested by monogastrics; the removal of raffinose-saccharides from soybean meal increases the metabolizable energy of the diet and reduces flatulent production (Coon et al., 1990; Parsons et al., 2000; Suarez et al., 1999).

The key step in raffinose and stachyose biosynthesis is mediated by the enzyme raffinose synthase. The raffinose synthase enzyme belongs to a group of hydrolase family enzymes that execute a galactosyl transfer from galactinol to sucrose. This transfer produces the three ring molecule raffinose; *myo*-inositol is formed as a by-product. Similarly, stachyose is formed by the action of stachyose synthase which combines raffinose and galactinol. Thus, raffinose synthase and stachyose synthase share one identical substrate, galactinol, and a second similar substrate, sucrose or raffinose, respectively. It is not known if raffinose synthase and stachyose synthase have overlapping enzymatic activity in soybean. However, distinct enzyme functions have been identified in adzuki bean (Peterbauer et al., 1999; Peterbauer and Richter 2001).

A plant introduction line, PI 200508, has been identified that has reduced levels of raffinose and stachyose and elevated levels of sucrose (Kerr and Sebastian, 2000). The line PI 200508 is synonymous with the LR 28 abbreviation (Hitz et al., 2002; Kerr and Sebastian, 2000). Hitz et al. (2002) identified this line in a screen for modified carbohydrate composition, and reported initial characterizations of PI 200508. Physiological analyses of derivatives of this line were subsequently examined (Neus et al., 2005). Mature seeds of the PI 200508 had greatly reduced levels of raffinose and stachyose, elevated levels of sucrose and galactinol, and a 25-fold reduction in raffinose synthase enzyme activity in developing seeds (Hitz et al., 2002). It did not appear that the vigor of lines with the low raffinose and stachyose trait was affected as there were no significant differences for low raffinose and stachyose lines in a number of seed quality parameters including field emergence, seed yield, maturity, and fatty acid content as

compared to sibling lines with normal levels of raffinose and stachyose (Neus et al., 2005).

A number of gene sequences have been annotated in GenBank as raffinose synthases; however, complete characterization of these genes is lacking. A search of the GenBank database revealed that raffinose synthase sequences have been annotated for pea (AJ426475), cucumber (AF073744), maize (Q575Z6, Q575Z7, Q575Z8), grape (AM430487), and *Arabidopsis* (At5g40390); but biochemical confirmation and molecular characterization have only been completed for pea (Peterbauer et al., 2001). A rice raffinose synthase (BAD68247) has also been characterized, but to a lesser extent (Li et al., 2007). Amino acid sequence identity ranges between 50 and 70% among putative raffinose synthase sequences of the various plant species.

Little molecular genetic evidence currently exists to link raffinose synthase enzymes with the accumulation of raffinose and stachyose in seeds. In a study concentrating on cold acclimation response in *Arabidopsis thaliana*, a putative raffinose synthase gene T-DNA insertion mutant was used to evaluate leaf freezing tolerance and cold acclimation. Seed raffinose levels were not characterized for the mutant line, although leaf raffinose levels were shown to be decreased when compared to wild type (Zuther et al., 2004). Soybean raffinose synthase enzyme activity was investigated in the low raffinose and stachyose PI 200508, but the associated raffinose synthase gene sequences were not reported in that study (Hitz et al., 2002).

The objective of this work was to develop molecular genetic information about soybean raffinose synthases and to evaluate those raffinose synthase genes for their involvement in the low raffinose and stachyose phenotype from soybean accession PI

200508. New knowledge about oligosaccharide biosynthesis in soybeans can be obtained by utilizing a molecular genetic approach with soybean raffinose synthases as candidate genes. Our hypothesis is that the nutritive value of soybean meal can be increased by disrupting the activity of raffinose synthase to reduce the levels of raffinose and stachyose. This reduced activity will allow the raffinose synthase substrate sucrose to accumulate in the seed; the accumulation of additional sucrose along with the decrease in raffinose and stachyose will increase the metabolizable component of the meal at the expense of anti-nutritional components.

MATERIALS AND METHODS

Plant Materials

Growth chamber settings were 14.5 hour day length and 28 °C/22 °C day/night temperatures. Plants were grown in 6-inch pots in PRO-MIX (Premier Horticulture, Quakertown, PA) medium and fertilized with Osmocote Plus (Scotts, Marysville, OH) per manufacturer's instructions.

The summer 2007 field location was the Bradford Research and Extension Center near Columbia, MO. Seeds were started in germination packets (CYG, Mega International, St. Paul, MN), transplanted to flats after nine days, then transplanted into the field seven days later. In the field, plants were grown approximately 1 foot apart with approximately 2 foot row spacing. Plants were grown to maturity, and irrigated when needed.

Population Development

A 'Williams 82' (Bernard and Cremeens, 1988) X PI 200508 cross was made in the summer of 2005 at Bradford Research and Extension Center near Columbia, Missouri. A set of the resulting F₁ and F₂ plants were grown to maturity in growth chambers. One population consisted of nineteen F₂ plants which were genotyped for RS2 and RS3 alleles as described below. A second population of approximately 84 F₂ plants was grown to maturity at the Bradford Research and Extension Center near Columbia, Missouri in the summer of 2007. The resulting F_{2.3} seeds were phenotyped for seed oligosaccharide content using HPLC. Seeds from all the growth chamber grown plants

were analyzed, while only a subset of the field grown plants were used for the phenotypic analysis (some RS2 heterozygous plants were omitted from phenotyping). For each plant, 8 individual seeds were analyzed.

Backcrossing PI 200508 RS2 Allele into Williams 82

The initial cross of Williams 82 X PI 200508 was the same as for the development of the segregating populations. After F₂ plants were genotyped, two plants homozygous for the PI 200508 RS2 allele were chosen as pollen donors to backcross to Williams 82. BC₁F₁ plants were confirmed by their heterozygous RS2 genotype and grown in a growth chamber to produce BC₁F₂ seeds. Genotype analysis was performed on BC₁F₂ seedlings to select plants homozygous for the PI 200508 RS2 allele. Homozygous plants were grown to maturity in a growth chamber. Four individual BC₁F_{2;3} seeds from two of the independent homozygous PI 200508 RS2 lines were analyzed for seed oligosaccharide phenotype.

Sequencing of Plant Raffinose Synthases

Soybean and pea sequences were obtained using PCR, and either direct sequencing or cloning followed by sequencing. Williams 82 alleles of RS2 and RS3 were sequenced from genomic DNA or BAC DNA. PI 200508 RS2, RS3, and RS4 alleles were sequenced from PCR products which were amplified with primers that were intronic, flanking exonic sequences.

Expression Analysis of Candidate Genes

A standard curve was constructed for each gene or pair of gene-specific primers in order to determine the efficiency of each primer pair (Pfaffl, 2001). This efficiency measure serves as a basis to compare the expression levels of different gene transcripts. This data was collected using DNase-treated total RNA template reverse transcribed into cDNA, then amplified by PCR, and was obtained in technical triplicates. Primer sequences for the candidate genes were as follows: RS1: 5'-GCAGGGCAACTCGTGATTCT -3' and 5'-TGGATGCGGGTTTTCTTCTC -3'; RS2: 5'-CTAGGGCCATCTCTGGTGGG -3' and 5'-CGTGTGGGGAGTGCATAGTG -3'; RS3: 5'-CACTGGAGTTCTTGGGGTGT -3' and 5'-GCTTGGCTGAAATACGAAGC -3'. The primers for the housekeeping gene elongation factor 1 α were 5' -CTGTAACAAGATGGATGCCACTAC- 3' and 5' -CAGTCAAGGTTAGTGGACCT- 3' (Czechowski et al., 2005). The RT-PCR reaction was carried out using the QuantiTect® SYBR Green RT-PCR Kit (Qiagen, Valencia, CA) in 20 μ l reactions. The parameters for the one step RT and the PCR were as follows: reverse transcription at 50 °C for 30 minutes followed by 95 °C for 15 minutes, then 35 cycles of 95 °C for 15 seconds, 57 °C for 30 seconds, 72 °C for 30 seconds and an ending hold at 4 °C. Experiments included control reactions lacking the reverse transcriptase enzyme to assess possible genomic DNA contamination.

Allele-Specific Molecular Marker Assay Development

Melting curve analysis molecular marker assays were developed as described by Wang et al. (2005). In order to achieve allele specificity, single base pair mismatches were introduced into the primer sequence to increase the discriminatory power of the

allele-specific primer. These bases and the tails are indicated in lowercase in the primer sequences. Primer sequences for RS2 were: 5'-

gcgggcagggcggcTGGAGCAGGTGTATaTGTGG -3', 5'-

gcgggcTGGAGCAGGTGTATtTGCAC -3', and 5'-

TGGGTCTGACCCCAACCAATAC -3'; for RS3: 5'-

gcgggcagggcggcATGGGTCCAAGCTCGAAGAAAGCgTCAC -3', 5'-

gcgggcATGGGTCCAAGCTCGAAGAAAGCcTCAG -3', 5'-

GAGTTGCAGAGGCTGAAGCCCTTCATG -3'.

Reactions were carried out in 20 μ l; each primer was at 0.5 μ M final concentration in reactions containing template, buffer (40 mM Tricine-KOH (pH 8.0), 16 mM KCl, 3.5 mM MgCl₂, 3.75 μ g ml⁻¹ BSA, 200 μ M dNTPs), 5% DMSO, 0.25X SYBR Green I, and 0.2X Titanium *Taq* polymerase (BD Biosciences, Palo Alto, CA).

PCR parameters for the RS2 assay were as follows: 95 °C for 12 minutes followed by 35 cycles of 95 °C for 20 seconds, 64 °C for 30 seconds, 72 °C for 30 seconds, and then a melting curve from 72 °C to 90 °C. The fluorescence was read after each cycle and every 0.2 °C with a one second hold during the melt. Each genotype produced a product with a characteristic melting profile, as measured by T_m of the negative first derivative of the disappearance of fluorescent signal. The Williams 82 homozygous genotype gave a peak at 82.5 °C, PI 200508 homozygous genotype gave a peak at 81.5 °C, and a heterozygous genotype gave a peak at 82.5 °C with a shoulder at 81.0 °C.

PCR parameters for the RS3 assay were as follows: 95 °C for 12 minutes followed by 35 cycles of 95 °C for 20 seconds, 62 °C for 30 seconds, 72 °C for 20 seconds, and then a melting curve from 72 °C to 90 °C. The fluorescence was read after each cycle and

every 0.2 °C with a one second hold during the melt. Each genotype gave a specific peak during the melt. The Williams 82 homozygous genotype gave a peak at 82.8°C, PI 200508 homozygous genotype gave a peak at 81.0 °C, and a heterozygous genotype gave a peak at 82.5 °C with a shoulder at 80.0 °C.

These assays were carried out with DNA isolated from young leaf tissue using the DNeasy® Plant Mini Kit (Qiagen). DNAs were not quantified, rather 2 µl of a 150 µl elution were used in each reaction.

Oligosaccharide Phenotype Determination by HPLC

Oligosaccharides were determined by high performance ion chromatography with pulsed amperometric detection (PAD) employing an Agilent 1100 series HPLC and an ESA Coulochem III detector (Agilent Technologies, Chesterfield, MO, USA). A 12.5 mg ground seed sample was extracted with 0.5 ml 50% ethanol at 70 °C, 30 min. Samples were then centrifuged 15 min at 16,000g. The supernatant was passed through a 0.2 µm filter. Sugars were separated on a Dionex Carbo Pac PA 10 analytical column (250 mm x 4 mm, 10 µm) connected to a Carbo Pac PA 10 guard column (50 mm x 4 mm). The mobile phase was 90 mM NaOH with flow rate of 1.5 ml min⁻¹, maintained at 30 °C. Detection settings were: time 0, 0.1 v, time 0.41, -2.0 v, time 0.42, 0.6 v, and time 0.44, -0.1 v.

RESULTS

Identification of Candidate Genes

Based on similarity to the pea enzyme sequences, the Arabidopsis genome appears to harbor one raffinose synthase (At5g40390) and one stachyose synthase (AK229121) gene. Due to the nature of the ancestral soybean genome having undergone full genome duplications (Shoemaker et al., 2006; Shoemaker et al., 1996), the existence of multiple raffinose synthase genes is likely. Three putative raffinose synthase genes were identified in soybean at the beginning of this project. Searching the GenBank database resulted in the identification of two annotated raffinose synthase genes: GenBank accession: E25448 (RS1) and E24424 (RS2) from *Glycine max*. An additional putative raffinose synthase (RS3) was reconstructed from overlapping ESTs with high sequence similarity to E24424. After the release of the Glyma1 assembly of the soybean genome sequence (Soybean Genome Project, DoE Joint Genome Institute), a fourth soybean raffinose synthase was identified (RS4). An alignment of the four putative soybean raffinose synthases and the Arabidopsis and pea raffinose synthases illustrates the high degree of sequence conservation (Figure 2.1). RS1 is the least like any of the other sequences, with about 35% identity. Identity between RS2, RS3, RS4, and the Arabidopsis raffinose synthase ranges from 58-61%. The soybean candidate genes RS2, RS3, and RS4 are 72 and 65%, and 66% identical at the amino acid level to the pea raffinose synthase, respectively. The putative soybean raffinose synthase genes are not well represented in the GenBank EST collection. There were no ESTs matching RS4, two matching RS2, seven matching RS3, and 17 matching RS1.

Expression Analysis by Quantitative RT-PCR

In soybean, a history of genome duplication has allowed the specialization of function for some individual members of gene families. Tissue-specific gene expression is one outcome for a subset of duplicated soybean genes (Chappell et al., 2006). We evaluated the steady state expression levels of the three originally identified putative raffinose synthase genes (RS1, RS2, and RS3) in a number of tissues from Williams 82 and PI 200508 (Figure 2.2). Of particular interest was quantitation of gene expression in developing seeds since Hitz et al. (2002) reported a significant decrease in raffinose synthase enzyme activity in developing seeds of line PI 200508 compared to standard lines. Expression of the three putative raffinose synthases was determined using quantitative RT-PCR. The results indicated that the putative raffinose synthase genes have similar transcript levels in all tissues examined as compared to the housekeeping gene elongation factor 1 α (Czechowski et al., 2005). In addition, there was no significant decrease in expression of the raffinose synthase genes in developing seed tissues from PI 200508.

Genomic Structures of Putative Raffinose Synthase Genes

The genomic organization of raffinose synthase genes from *Arabidopsis thaliana*, grape, and rice are known. To determine the genomic arrangement of soybean RS2 and RS3, we sequenced Williams 82 genomic DNA, Williams 82 BAC DNA, and cDNA in those gene regions (Figure 2.3). Comparisons of genomic DNA, BAC DNA and cDNA led to the determination of intron-exon arrangement. Sequences for RS2 and RS3 have

been deposited in the GenBank under accessions EU651888 and EU651889, respectively. During the course of this project, an independent group deposited soybean Williams 82 BAC sequences that contain the RS2 and RS3 genes in GenBank (Schlueter et al., 2007). The RS4 genomic sequence was identified as part of the soybean Glyma1 genome sequence assembly. Both RS3 and RS4 reside on scaffold 35 (linkage group A1, chromosome 05, J. Gillman personal communication) approximately 3.2 Mbp apart. RS2, RS3, and RS4 were predicted to contain five exons interrupted by four introns. The position of the first intron is variable among plant raffinose synthases and the candidate genes. For RS2, the first intron occurs in amino acid 215, while in RS3 it occurs in amino acid 268, and in RS4 it occurs in amino acid 197. In addition, we sequenced the pea raffinose synthase from pea genomic DNA to clarify the relationship between authentic raffinose synthases and genomic organization; previously only cDNA sequence was available for the pea raffinose synthase. The pea raffinose synthase genomic organization is less similar to the soybean candidate raffinose synthase gene organization than the Arabidopsis raffinose synthase. Excluding RS1, the soybean and Arabidopsis raffinose synthases appear to have a characteristic five exon, four intron genomic organization; the other putative raffinose synthases have fewer exons, although some of the exon/intron boundary positions present in the Arabidopsis and soybean raffinose synthases are conserved.

RS1 cDNA sequence was obtained from the GenBank patent database and shows sequence similarity to two hydrolase family enzymes in Arabidopsis (At3g57520 and At1g55740). Recently, trace archives and the Glyma1 assembly of the soybean genome sequencing project have allowed the investigation of the genomic organization of

RS1, and a thirteen exon, twelve intron organization identical to the Arabidopsis hydrolase family gene At3g57520 has been confirmed. Members of this gene family have been shown to be alkaline α -galactosidases (Carmi et al., 2003). No other additional soybean raffinose synthase genes were apparent based on sequence homology to the candidate genes and to the Arabidopsis raffinose synthase (At5g40390).

Identification of Polymorphic Alleles of Candidate Genes

RS2 and RS3 became our primary candidate raffinose synthase genes. RS1 appeared to belong to a related, but distinct gene family of hydrolases and there was no evidence of the existence of RS4 prior to the release of the soybean genome sequence. The soybean raffinose synthase candidate genes, RS2, RS3, and RS4 were sequenced from the low raffinose and stachyose line, PI 200508. No polymorphisms were identified for RS4, while multiple polymorphisms were identified for RS2 and RS3 when compared to the reference Williams 82 sequences. Five single nucleotide polymorphisms (SNPs) and a 3 base pair deletion were identified in the RS2 allele from PI 200508. Two of the SNPs and the deletion resulted in amino acid changes (Figure 2.4). In the first exon a SNP (G512C) resulted in a cysteine to serine (C171S) mutation which does not occur in a conserved region and has a calculated SIFT score (Johnson et al., 2005; Ng and Henikoff, 2003) of 0.25. A three base pair deletion (TGG; beginning at base pair 991) resulted in the absence of a tryptophan (W331-) in a highly conserved region of exon two of RS2 and has a SIFT score of 0.00. This sift score indicates that the loss of this amino acid is likely to be deleterious to protein function (Johnson et al., 2005; Ng and Henikoff, 2003). The second SNP (G1609A) was in exon four and resulted in a valine to isoleucine

mutation. This missense V537I change occurred in a conserved region, although a similar amino acid was substituted; the SIFT score of 1.00 indicates that this amino acid change is unlikely to have an effect on the protein.

Six SNPs were identified in the PI 200508 RS3 sequence (data not shown). Three were silent and three resulted in amino acid changes. Two of the amino acid changes occurred in exon 1, a proline to alanine (C28G; P10A) and a lysine to asparagine (G48T; K16N), neither of which occurred in conserved regions. The third polymorphism (C1105G; R369G) resulted in an arginine to glycine substitution. The PI 200508 glycine was also identified in a number of other soybean lines evaluated for this polymorphism, while the Williams 82 arginine was only present in 'Jack' (data not shown).

Association of Candidate Gene Genotype to Seed Raffinose and Stachyose

Phenotype

In order to facilitate tracking the RS2 and RS3 alleles from the PI 200508, molecular marker assays were developed from a selected polymorphism for each of the PI 200508 RS2 and RS3 alleles. The molecular marker assays, which are single step PCRs utilizing primers containing GC tails (Wang et al., 2005), reliably detected the presence of either a Williams 82 allele or a PI 200508 allele. For the RS2 marker assay, the three base pair deletion in exon two was used to distinguish alleles; for the RS3 marker assay the proline to alanine polymorphism from exon one was used to distinguish alleles.

The molecular marker assays were used to investigate the occurrence of the selected polymorphisms in a subset of ancestors to many modern North American soybean cultivars (Sneller, 1994). The absence of TGG in RS2 may be a rare

polymorphism; it was not detected in any of the 17 ancestral lines assayed (data not shown). For the SNP assayed in RS3, CNS, Illini, Richland, PI 81041, and PI 257435 contained the base in common with Williams 82 (data not shown).

To analyze RS2 and RS3 as candidate genes that influence the low raffinose and low stachyose trait, an association analysis was done on a population of segregating F_2 plants and $F_{2:3}$ seeds of a cross between Williams 82 and PI 200508. The population was divided into two sets, with one set of F_2 plants grown in growth chambers and a second set of F_2 plants grown in a field environment. The molecular marker assays for the PI 200508 RS2 and RS3 alleles were used to genotype F_2 plants of the Williams 82 X PI 200508 segregating populations. The molecular marker assay data from the F_2 plants for both candidate genes and data sets revealed a 1:2:1 ratio (at $p = 0.05$) for Williams 82 alleles: heterozygote: PI 200508 alleles, which would be expected for a segregating population at F_2 (data not shown). There was no evidence of linkage of RS2 and RS3, as the genes segregated independently of each other. RS2 resides on linkage group C2 (chromosome 06) while RS3 and RS4 are separated by approximately 3.2 million base pairs on linkage group A1 (chromosome 05).

The relative oligosaccharide phenotypes of single $F_{2:3}$ seeds were determined after alcohol extraction of ground seed samples. Ion exchange chromatography with electrochemical detection was used to quantitatively measure the relative amounts of sucrose, raffinose, and stachyose. In two data sets, the RS2 genotype strongly associated with the oligosaccharide phenotype (Figure 2.5). Inheritance of the Williams 82 RS2 alleles was sufficient to produce a Williams 82 oligosaccharide phenotype, and inheritance of the PI 200508 RS2 alleles was sufficient to produce a PI 200508

oligosaccharide phenotype ($p < 0.01$). For heterozygous F_2 plants whose F_3 seeds had segregated for the RS2 alleles, we observed two separate phenotypic classes, one similar to the Williams 82 phenotype and one similar to the PI200508 phenotype; there were no seeds with an “intermediate” phenotype. There was no obvious association of oligosaccharide phenotype with the RS3 alleles (data not shown).

Analyzing Oligosaccharide Content in a Backcross

To improve the efficiency of selection of low raffinose and low stachyose soybean lines, the RS2 perfect molecular markers were used in a backcross breeding effort. The molecular markers were applied in the original F_2 generation of the segregants to identify lines for backcrossing, and then in the BC_1F_1 and BC_1F_2 generations. Seeds from the resulting homozygous PI 200508 RS2 allele BC_1F_3 lines were analyzed for the inheritance of the low raffinose and stachyose phenotype. This backcrossed material theoretically contained 75% of the Williams 82 genome along with the PI 200508 RS2 alleles. The seed oligosaccharide phenotype of low raffinose and stachyose was faithfully inherited in the samples analyzed. This further demonstrates the correspondence of the PI 200508 RS2 allele to the seed oligosaccharide phenotype.

DISCUSSION

Significant progress has been made in identifying and sequencing putative soybean raffinose synthase genes, characterizing their expression in multiple tissues, and developing and characterizing the phenotype/genotype associations in a segregating population. The candidate raffinose synthases do not appear to be tissue specific, but highly expressed in all tissues examined. The candidate genes have polymorphic alleles for the standard cultivar, Williams 82, and a low raffinose and stachyose line, PI 200508 (Hitz et al., 2002). The development of molecular marker assays and a segregating population allowed the association of one candidate gene, RS2, with the seed oligosaccharide phenotype. These data provide strong evidence that the candidate gene, RS2, is responsible for raffinose biosynthesis in the soybean seed.

The increasing use of soybeans and soy products in the feed industry has elevated the importance of understanding its nutritional and anti-nutritional components. Here, we describe a soybean raffinose synthase candidate gene which largely controls the amounts of seed sucrose, raffinose, and stachyose. Once integrated into a standard cultivar, this raffinose synthase allele has the potential to greatly improve the digestibility and carbohydrate richness of soybean meal in terms of metabolizable energy for monogastric animals.

A comparison of the soybean candidate genes with other known plant raffinose synthases allowed the identification of the characteristic five exon, four intron genomic arrangement. Candidate genes RS2, RS3, and RS4 are comparable to other plant raffinose synthase genes, while the arrangement of RS1 is not. Therefore, RS2, RS3, and

RS4 remain strong candidates for having raffinose synthase identity, while RS1 appears to be an alkaline α -galactosidase (Carmi et al., 2003).

Similar transcript levels as determined by quantitative RT-PCR, for RS2 and RS3, for Williams 82 and PI 200508 indicated that the difference in oligosaccharide content was unlikely to be a product of disruption in the transcriptional machinery, but rather a polymorphism/mutation which affects folding, function, or both. Sequencing of two of the candidate genes from both the standard cultivar Williams 82 and the low raffinose and stachyose line, PI 200508 allowed the identification of polymorphisms which resulted in amino acid changes. For RS2, two amino acid changes are of interest. The first is the C171S which does not occur in a region of sequence conservation but potentially represents the loss of a cysteine bonding partner in PI 200508. More interesting is the deletion of a tryptophan residue (W331-) from the PI 200508 sequence. This deletion occurs in a region of high conservation; it is likely required for correct protein folding or function. At this time, it is unknown where the catalytic domains are in the protein; therefore it is impossible to definitively determine the role of this tryptophan in enzymatic activity without further biochemical characterization.

For RS3, two interesting polymorphisms were identified. The first is in exon 1 and is a K16N change. Although this residue does not occur in a region of high sequence conservation, the change in charge may result in a disruption of enzyme function. The second polymorphism, R369G occurs in exon 2. It is unknown what affect this change may have on enzymatic activity, but the fact that Williams 82 appears to contain the variant allele is notable.

The identification of polymorphisms between the Williams 82 and PI 200508 alleles of RS2 and RS3 allowed the development of allele specific molecular marker assays. These assays were used to genotype the Williams 82 X PI 200508 segregating population in order to associate the seed oligosaccharide phenotype with the genotype of the candidate genes. The genotype of RS2 completely associated with the seed raffinose phenotype. The complete association indicates that RS2 is largely responsible for the phenotype and is likely a raffinose synthase. Furthermore, based on our analysis of the segregants from the set of RS2 heterozygous F₂ plants, it appears that one copy of the Williams 82 allele of RS2 is sufficient to produce the wild type seed raffinose and stachyose phenotype. Since the desired phenotype is inherited in a recessive manner, the identification of the mutation corresponding to the phenotype will be particularly useful in breeding programs for selecting heterozygous and backcrossed lines that have captured the mutant allele.

The genotype of the candidate gene RS3 showed no association with the seed raffinose and stachyose phenotype. This result indicates that RS3 is either contributing to the raffinose phenotype to a far lesser degree than RS2, is non-functional or is not a raffinose synthase. Since RS3 and RS4 reside on the same chromosome (Soybean Genome Project, DoE Joint Genome Institute), it is also unlikely that RS4 is contributing significantly to the raffinose and stachyose phenotype.

It is notable that while raffinose was almost eliminated in lines containing the PI 200508 RS2 alleles, stachyose was still produced, although at lower levels than in wild-type seeds. Thus it would appear the reduced levels of raffinose are sufficient substrates for seed stachyose synthase activity. Due to the substrate similarity of the two enzymes

(galactinol and sucrose for raffinose synthase, or galactinol and raffinose for stachyose synthase), sequence similarity between the two is expected. The sequence identity for RS2 and RS3 is 65%, for RS2 and RS4 is 70%, and for RS3 and RS4 is 62% at the amino acid level. Both an Arabidopsis raffinose synthase (At5g40390) and a putative stachyose synthase (AK229121) are known, and the amino acid sequence identity is 45% for these sequences. Likewise, the pea raffinose synthase (Q8VWN6) and stachyose synthase (Q93XK2) share 42% identity. A recent screen of the soybean genome revealed a putative stachyose synthase sequence on linkage group L which corresponds to chromosome 19. This sequence has approximately 33% identity with the soybean raffinose synthase gene RS2.

Here, evidence is shown that the candidate gene, RS2, completely associates with the seed raffinose phenotype through a segregating population. Further, the expression of this gene does not appear to be tissue specific, but rather expressed in all tissues examined. The identification of a soybean raffinose synthase candidate gene which completely associated with the low raffinose and stachyose phenotype along with the development of perfect molecular marker assays will enable increased breeding efficiency of this desirable trait into soybean varieties.

FIGURES

PsRS. 1 MAPPSITKTATQQDVIS TVDIGNSPLLSISLDQSRNFLVNGHPFLTQVPPNITTTTTSTP
 GmRS2. 1 -----MAPSISKIVELNSFGLVNGNLP LSIITLEGSNFLANGHPFLTQVPPNIIIVTPSPID
 GmRS4. 1 -----MAPSLTAKQEAALLNANPHLSIKLQSSTFFANNHPILTQVPPNITTTTTPPPH
 GmRS3. 1 -----MGPSKKKASPKSGVTKHMKGFSLCNSTLKVNGQVILSQVPKNVTLTPCTYD
 AtRS. 1 -----MASPCLTKSDSGINGVDFTEKFRLEDSTLLANGQVVLTDPVNVVTLTSSPYL
 GmRS1. 1 -----MTVTPKISVNDGKLVVHGKTLTGVDPDNVLTTPGSGR

PsRS. 61 SPFLDFKSNKDTIANNNNTLQQQGCVFVFNNTT-EAKSHHVVPLGKLGKIKFTSIFRFKVV
 GmRS2. 56 AKSSKNNEDDDVVG-----CFVGFHAD-EPRSRHVASLGKLRGKIFMSIFRFKVV
 GmRS4. 53 DASTTPAG-----CFVGFHAD-EARSRHVISLGKLRGIRFTSIFRFKLV
 GmRS3. 52 THTT-----GCFVGFHAD-SPKSRHVAPLGQKNI SFTSIFRFKVV
 AtRS. 53 VDKDGVPLDVSAGS-----FIGFNLDGEPKSHHVASIGKLNIRFMSIFRFKVV
 GmRS1. 38 GLVTG-----AFVGFHAD-HSKSLHVFPMGVLEGLRFMCCFRFKLV

PsRS. 120 WTHHWGTNGHELEHETQILILDKNISL-----GRPYVLLLPILENSFRSLQPLGND
 GmRS2. 105 WTHHWVGSNGHELEHETQMMLLDKNDQL-----GRPFVLLLPILQASFRASLQPLGDD
 GmRS4. 96 WSTHWSGSGNRDVENETQMMILQNDAVE-----GRPYVLLLPILLEGPFASLQPLGHD
 GmRS3. 92 WTTLWGTSGNRDLETETQFLMLQS-----HPYVLLLPILQPPFRASLQPHSDD
 AtRS. 102 WTHHWVGSNGRDIENETQIIILDQSGSDSGPGSGSGRPYVLLLPILLEGFSRSSFQSGEDD
 GmRS1. 78 WMTQRMGTGCRDVPLETQFMLESKESKESD-GENSPIIYTVLLLPILLEGQFRAVLQGNCKN

PsRS. 173 YVDMVESGSGSTHVTGSTFKACLYLHLSNDPYRLVKEAVKVIQTKLGTFTLEEKTPPSII
 GmRS2. 158 YVDVCMESGSTRVCGSSFGSCLYVHVGHDPYQLREATKVVVRMHLGTFTLLEEKTPAVII
 GmRS4. 149 DVVICMESGSARVTKSRFRSTSVYMHVHDDPFTLIDEALKVIRVYLGTFRLMEEKTPVGI
 GmRS3. 140 NVAVCVESGSSHVTASSFDTVVYLHAGDNPFTLVKEAMRVVRAHLGFSKLLLEEKTPGMV
 AtRS. 162 DVAVCVESGSTEVTGSEFRQIVYVHAGDDPFLKVDAMKVI RVMNTFKLLLEEKSPPGIV
 GmRS1. 137 EIEICLES GDNAVETDQGLHMVYMHAGTNPFEVINQAVKAVEKHMOTFLHREKKRIP SCL

PsRS. 233 EKFGWCTWDAFYLVKHPKGVWEGVKALTDGGCPPGFVIIDDGWQSI SHDDDDPVTERDGM
 GmRS2. 218 DKFGWCTWDAFYLVKHPKGVWEGVKGLVEGGCPPGMVLIDDGWQAI CHDED-PITDQEGM
 GmRS4. 209 DKFGWCTWDAFYLVNHPKGVWEGVKGLVEGGCPPGLVLIDDGWQTF CRDDE-TVSDGGSL
 GmRS3. 200 DKFGWCTWDAFYLVNHPKGVWEGVKGLVDDGGCPPGFVLIDDGWQCISHDSD---PEKEGM
 AtRS. 222 DKFGWCTWDAFYLVNHPKGVWEGVKGLVDDGGCPPGLVLIDDGWQSI GHSDSD--GIDVEGM
 GmRS1. 197 DWFGWCTWDAFYTDVTAEGVEEGLKSLSQGGTTPRFLIIDDGWQO IENKAK-----DATE

PsRS. 293 NRTSAGEQMPCLRLIKYEENYKFREYENGDNGGKKGLVGFVRDLKEEFRSVESVYVWHALC
 GmRS2. 277 KRTSAGEQMPCLRVLKLEENYKFRQYCSGKDS-EKGMGAFVRDLKEEFRSVEQVYVWHALC
 GmRS4. 268 NCSVPGEQMLNRLIKFEENGKFKKEYKCGREG-NKGMGAFVRELKEEFSGLEYVYVWHALC
 GmRS3. 257 NOTVAGEQMPCLRISYEENYKFRSYKE-----GKGLKGFVRELKEEFSGVEYVYVWHALC
 AtRS. 280 NITVAGEQMPCLRLIKFEENHKFKDYVSPKQNDVGMKAFVRDLKDEFSTVDYIYVWHALC
 GmRS1. 252 CLVQEGAQFATRLTGIKENTKFKQKLLQNNEQ-MSGLKHLVHGAKQHHN-VKNVYVWHALA

PsRS. 353 GYWGVRPKVCGMP--EAKVVVVKLSPGVKMTMEDLAVDKIVENGVGLVPPNLAQEMFDG
 GmRS2. 336 GYWGVRPKVPGMP--QAKVVVVKLSPGVKMTMEDLAVDKIVSNGVGLVPPPHLAHLLEYEG
 GmRS4. 327 GYWGVRPKVPGMP--EATVVVVKLSPGAEMTMTDQAVVKIMEIGVGLVPPPHRAHELEYEG
 GmRS3. 312 GYWGVRPGVAGMA--EAAVEKPKLTEGLKGTMEDLAVDKIVNNGVGVVPPPELVGEMYER
 AtRS. 340 GYWGGLRPEAPALP--PSTIIRPELSPGLKLTMEDLAVDKI IETGIGFASPDLAKEFYEG
 GmRS1. 310 GYWGVRKPAATGMEHYDTALAYPVQSPGVGLGNQPDIVMDSI AVHGLGLVHPKVVFNFYNE

PsRS. 411 IHSHLESAGIDGVKVDVIHLLLELLSEEYGGRVELAKAYYKALTSVSNKHKFKGNGVIASME
 GmRS2. 394 LHSRLESAGIDGVKVDVIHLLLEMLSEEYGGRVELAKAYYKALTASVKKHKFKGNGVIASME
 GmRS4. 385 LSHLESVIGIDGVKIDVTHILEMLSEEYGGRVELAKAYYKALTASVRKHKFKGNGVISSMQ
 GmRS3. 370 LHAHLESAGIDGVKVDVIHLLLEMVCEKYGGRVDMAKAYYKALTASVRKHKFKGNGVIASME
 AtRS. 398 LSHLQONAGIDGVKVDVIHILEMLCQKYGGRVDLAKAYFKALTSVSNKHKFNKNGVIASME
 GmRS1. 370 LHAYLASCGVDGVKVDVQNIETLGGAGHGGRVSLTRSYHHALEASIASNFTDNGCIACMC

PsRS. 471 HCNDFFLLGTEAISLGRVGDDFWCCDPSGDPNGTYWLQGCHMVHCAYNLSLWGMNFIHPDW
 GmRS2. 454 HCNDFFLLGTEAIALGRVGDDFWCTDPSGDPNGTYWLQGCHMVHCAYNLSLWGMNFIQPDW
 GmRS4. 445 QCNDFMFLGTETISLGRVGDDFWCTDPAGDPNGTYWLQGCHMVHCAYNLSLWGMNFIHPDW
 GmRS3. 430 HCNDFMFLGTEAISLGRVGDDFWCTDPYGDPNFTFWLQGCHMVHCAYNLSLWGMNFIHPDW
 AtRS. 458 HCNDFMFLGTEAISLGRVGDDFWCTDPSGDPNGTFWLQGCHMVHCAYNLSLWGMNFIQPDW
 GmRS1. 430 HNTDGLYS-AKQTAIVRASDDFYPRDPAS-----HTIHISSVAYNLSLFLGEFMQPDW

PsRS. 531 DMFQSTHPCAEFHAASRAISGGPVYVSDCVGNHNFKLLKSFVLPDGSILRCQHYALPTRD
 GmRS2. 514 DMFQSTHPCAEFHAASRAISGGPVYVSDCVGKHNFKLLKSLALPDGTILRCQHYALPTRD
 GmRS4. 505 DMFQSDHACAEFHAASRAISGGPIYVSDSVGKHNFKLLKLLVLPDGSILRCQHYALPTRD
 GmRS3. 490 DMFQSTHPCAAFHAASRAISGGPIYISDTIVGNHNFELLKTLALPDGSILRCEHYALPTRD
 AtRS. 518 DMFQSTHPCAEFHAASRAISGGPIYISDCVGHDFDILLKRLVLPNGSILRCEYALPTRD
 GmRS1. 481 DMFHSLHPAADYHAAARAIGGCPYVSDKPGNHNFDLLKLLVLPDGSVLRALPGRPTRD

PsRS. 591 CLFEDPLHNGKTMLKIWNLNKYAGVLGLFNCQGGGWCPEPTRRNKSASEFSHVAVTCTYASPE
 GmRS2. 574 CLFEDPLHDGKTMLKIWNLNKYTGVLGLFNCQGGGWCPEPTRRNKSASEFSQTVTCLASPQ
 GmRS4. 565 CLFVDPPLHDGKTMLKIWNLNKCSGVLGLFNCQGGGWCPEPTRRNKSSSDYSHSVTCTFASPQ
 GmRS3. 550 CLFADPLHDGKTMLKIWNLNKYTGVLGVFNCQGGGWFREIRSNKCAAEFSHRVSTKTNIK
 AtRS. 578 RLFEDPLHDGKTMLKIWNLNKYTGVI GAFNCQGGGWCPEPTRRNQCFSECVNTLTATTSPK
 GmRS1. 541 SLFVDPARDRTSLKLIWNLNKCSGVVGVFNCQGAGWCKIEKKTRIHDTSPTLTASVCAS

PsRS. 651 DIEWCNGKTPMDIKGVDVFAVYFFKEKKLSLMKCS-DRLEVSLEPFSELMITVSPKLVFS
 GmRS2. 634 DIEWSNGKSPICIKGMNVFAVYLFKDHKLLKMKAS-EKLEVSLEPFTFELLTVSPVIVLS
 GmRS4. 625 DIEWGKPKHPVCIKGVDVFAVYMFKDDKLLKLYT-ESVEVSLEPFSCCELLTVSPVIVLP
 GmRS3. 610 DIEWDSGNPISIEGVQLFASYFSQAKLLILSAPSDDSEEISLEPFNFELITVSPVIVLP
 AtRS. 638 DVEWNSGSSPISIANVEEFALFLSQSKLLLSGLN-DDLELTLEPFKFEKITVSPVVTIE
 GmRS1. 601 DVDLITQVAGAEWLGDITIVYAYRSGEVIRLPGKVS---IPVTLKVLFEFELFHFPCPIQEI

PsRS. 710 KRLIQFAPIGLVNMLNSGGAVQSLEFD--DSASLVKIGVRCGELS VFASEKPVCKIDG
 GmRS2. 693 KKLIIQFAPIGLVNMLNTGGAIQSMEFD--NHIDVVKIGVRCGEMKVFASEKPVSKLDG
 GmRS4. 684 RKSIIQFAPIGLVNMLNSGGSIMSLEFD--QQENLARIGVRGHGEMRVFASEKPEVSKIDG
 GmRS3. 670 GKSIVKFAPIGLVNMLNTGGAVQSLAFD--EGQNLVEVGLRGTGEMRVYASEKPRTCRIDG
 AtRS. 697 GNSVRFAPIGLVNMLNTSGAIRSLVYN--DES--VEVGVFGAGEFRVYASKKPVSLIDG
 GmRS1. 658 P-SISFAAIGLIDMFNTGGAVEQVEIHNRAATKTIALSVRGRGRFGVYSSQRPLKCVVGG

PsRS. 768 VSVEFDYEDKMVRVQILWPGSSTLVVEFLF---
 GmRS2. 751 VVVKFDYEDKMLRVQVPWPSASKLSMVEFLF---
 GmRS4. 742 ESVEFDYVDRITVRLQVSWPCSSRLSVVEYLF---
 GmRS3. 728 KEVDFEYEGSMVNIQVPWPGSSKLSTVQYVF---
 AtRS. 753 EVVEFGYEDSMVMVQVPWSPGPDGLSSIQYLF---
 GmRS1. 717 AETDFNYDSETGLTTFSIPVSPPEMYRWSIEIQV

Figure 2.1: Boxshade output of a ClustalW sequence alignment of putative *Glycine max* raffinose synthases (GmRS1, GmRS2: EU651888, GmRS3: EU651889, and GmRS4), the *Arabidopsis thaliana* raffinose synthase (AtRS: At5g40390), and the pea raffinose synthase (PsRS: Q8VWN6). Identical amino acid residues are highlighted in black while similar amino acid residues are highlighted in gray.

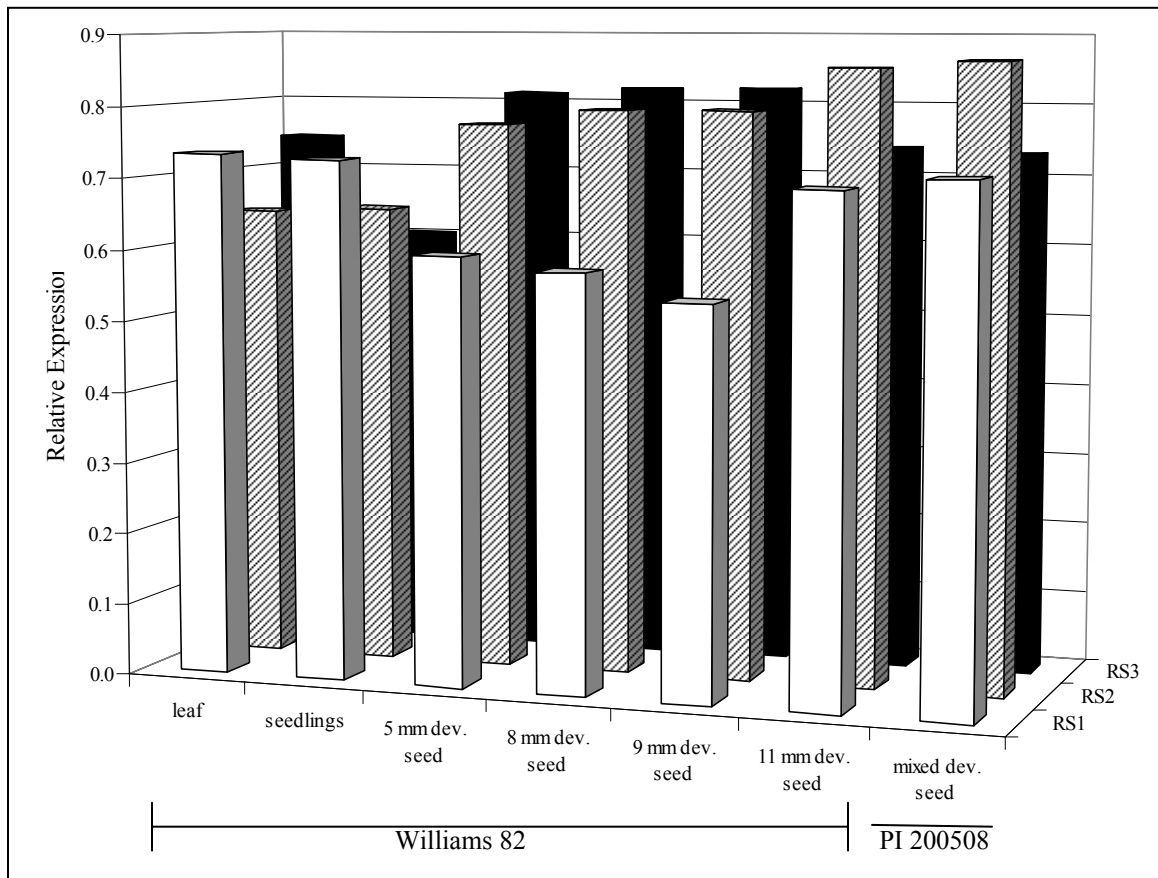


Figure 2.2: Relative expression of putative raffinose synthases for multiple soybean tissues determined by quantitative RT-PCR. Left to right: the first six sets of bars represent expression of the candidate genes in Williams 82 tissue. The last set of bars represent expression in mixed developing seed tissues from PI 200508. The three putative raffinose synthases are along the z-axis. Bar heights represent the average values from three replicates expressed relative to the housekeeping gene (elongation factor 1 α) control.



Figure 2.3: Schematic representation of the genomic structure of putative and known plant raffinose synthases. Exons are represented as boxes separated by lines representing introns, all approximately to scale. *Arabidopsis thaliana* (AtRS) is At5g40390, rice (RiceRS) was derived from GenBank accession BAD68247, pea (PsRS) is GenBank accession AJ426475, and grape (GrapeRS) is GenBank accession AM430487.

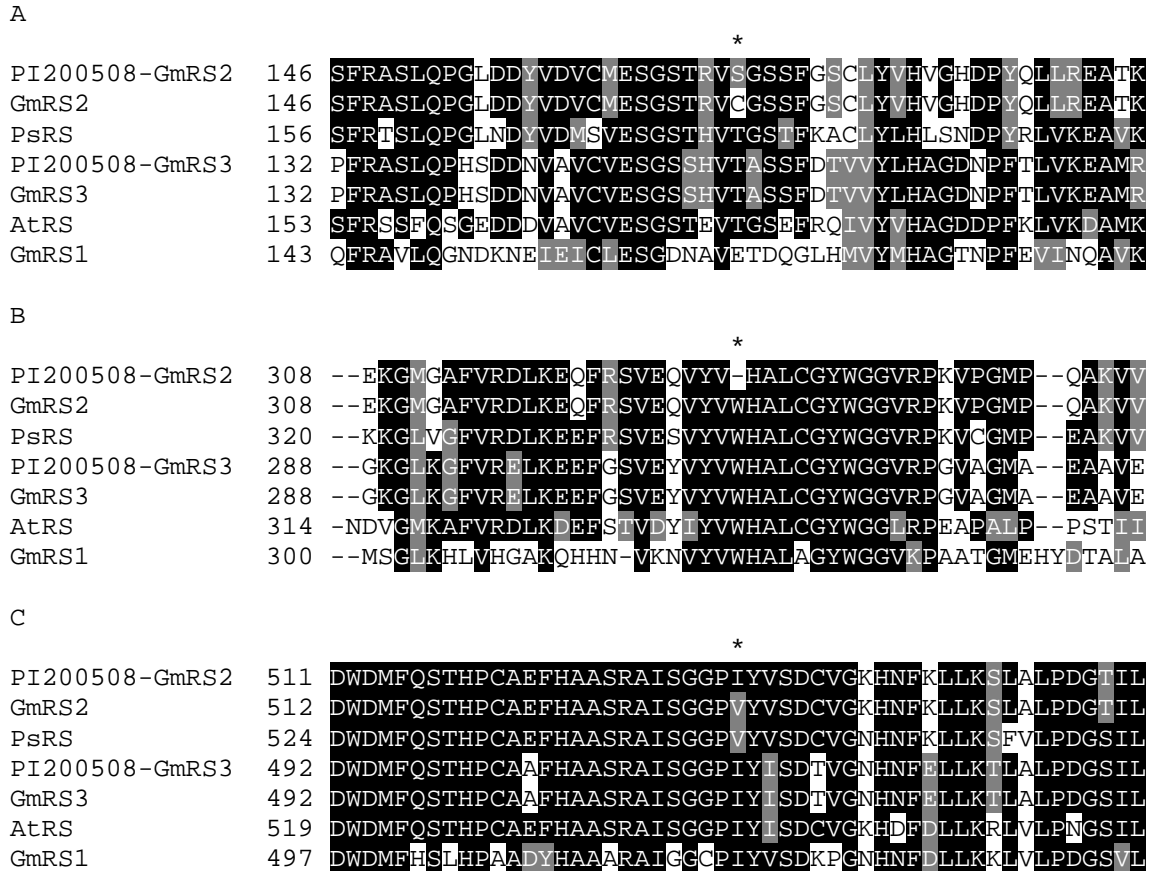
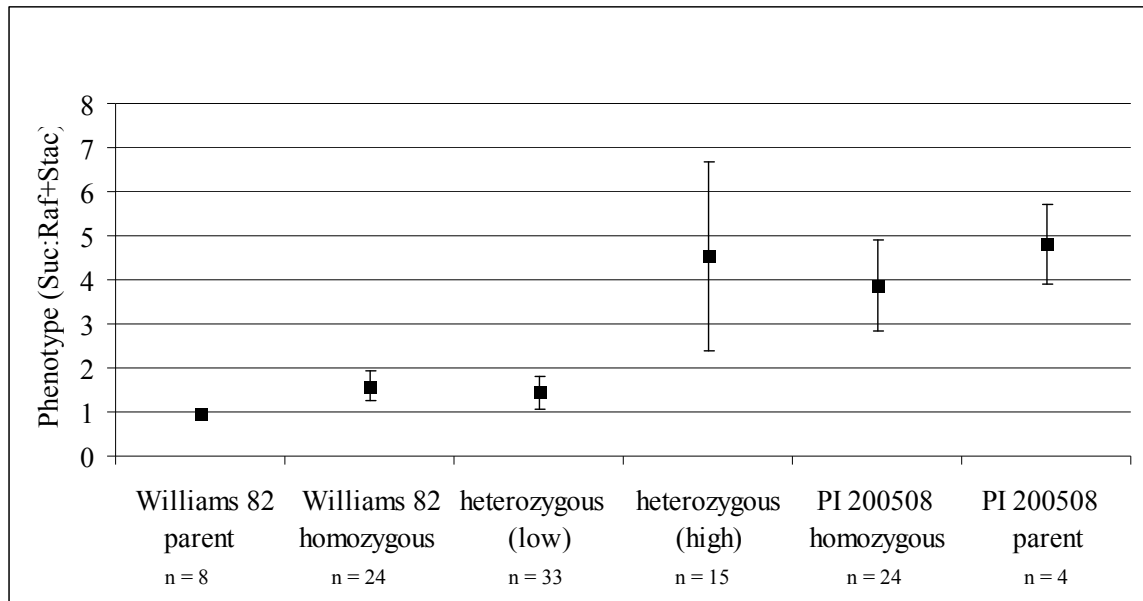


Figure 2.4. Raffinose synthase amino acid sequence alignments in the regions surrounding polymorphisms in the RS2 gene from PI 200508. Amino acid positions are indicated at the beginning of each alignment. The position of the polymorphic amino acid is indicated by an asterisk. Identical amino acid residues are highlighted in black while similar amino acid residues are highlighted in gray. A) The exon one region representing the variant allele of PI 200508. B) The exon two region containing a three base pair deletion the PI 200508 allele that resulted in the absence of a conserved tryptophan (W331-). C) The exon four region representing the PI 200508 variant allele.

A



B

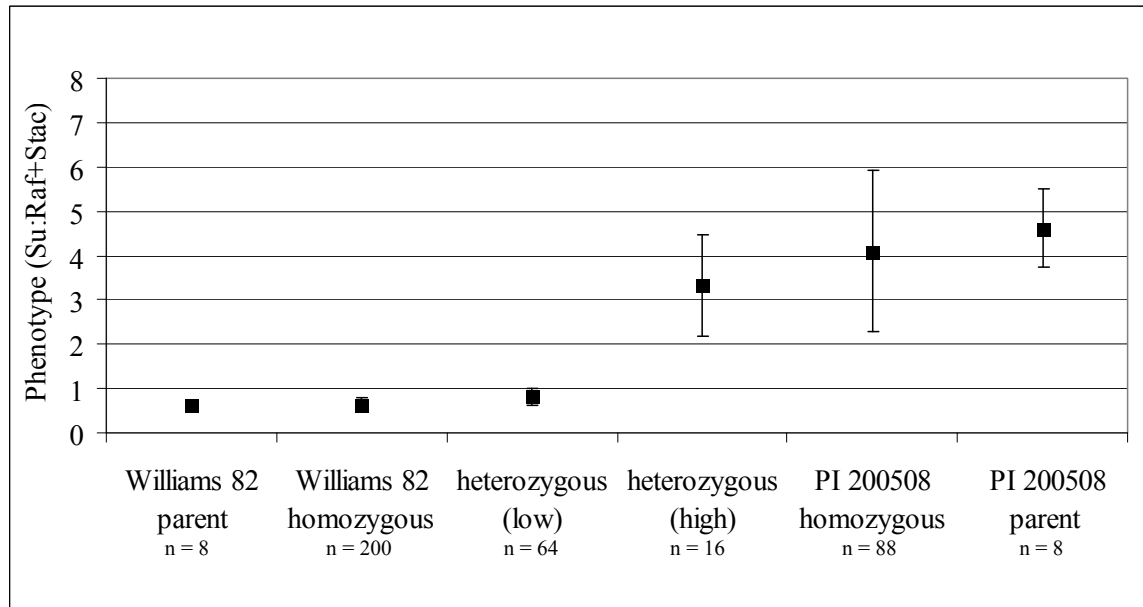


Figure 2.5. Phenotype to genotype association of a soybean population segregating for the low raffinose/stachyose phenotype (Williams 82 X PI 200508). The oligosaccharide phenotype of eight individual F₃ seeds was determined for a subset of the population based on genotype. The data represents the mean of the ratio of extractable seed sucrose to the sum of raffinose and stachyose. Error bars represent plus and minus one standard deviation from the mean. Individual F₂ plants were categorized by their RS2 genotype class and separated on the x-axis. The F₂ RS2 heterozygous class was further separated into two phenotypic classes because no intermediate phenotypes were observed. Parental genotypes represent individual seeds analyzed from either Williams 82 or PI 200508

grown in the same environment. A) Data collected from growth chamber grown F_2 plants. B) Data collected from field grown F_2 plants.

LITERATURE CITED

- Bernard, R.L., and C.R. Cremeens. 1988. Registration of 'Williams 82' soybean. *Crop Science* 28:1027-1028.
- Carmi, N., G. Zhang, M. Petreikov, Z. Gao, Y. Eyal, D. Granot, and A.A. Schaffer. 2003. Cloning and functional expression of alkaline alpha-galactosidase from melon fruit: similarity to plant SIP proteins uncovers a novel family of plant glycosyl hydrolases. *Plant Journal* 33:97-106.
- Chappell, A.S., A.M. Scaboo, X. Wu, H. Nguyen, V.R. Pantalone, and K.D. Bilyeu. 2006. Characterization of the MIPS gene family in *Glycine max*. *Plant Breeding* 125:493-500.
- Coon, C.N., K.L. Leske, O. Akavanichan, and T.K. Cheng. 1990. Effect of oligosaccharide-free soybean meal on true metabolizable energy and fiber digestion in adult roosters. *Poultry Science* 69:787-793.
- Czechowski, T., M. Stitt, T. Altmann, M.K. Udvardi, and W.R. Scheible. 2005. Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiology* 139:5-17.
- Hitz, W.D., T.J. Carlson, P.S. Kerr, and S.A. Sebastian. 2002. Biochemical and molecular characterization of a mutation that confers a decreased raffinose and phytic acid phenotype on soybean seeds. *Plant Physiology* 128:650-660.
- Hsu, S.H., H.H. Hadley, and T. Hymowitz. 1973. Changes in carbohydrate contents of germinating soybean seeds. *Crop Science* 13:407-410.
- Johnson, M.M., J. Houck, and C. Chen. 2005. Screening for deleterious nonsynonymous single-nucleotide polymorphisms in genes involved in steroid hormone metabolism and response. *Cancer Epidemiology Biomarkers and Prevention* 14:1326-1329.
- Kerr, P.S., and S.A. Sebastian. 2000. Soybean products with improved carbohydrate composition and soybean plants. United States Patent 6147193. Date Issued: November 14, 2000.
- Li, S., W.D. Kim, S. Kaneko, P.A. Prema, M. Nakajima, and H. Kobayashi. 2007. Expression of rice (*Oryza sativa* L. var. Nipponbare) alpha-galactosidase genes in *Escherichia coli* and characterization. *Bioscience, Biotechnology, and Biochemistry* 71:520-526.
- Neus, J.D., W.R. Fehr, and S.R. Schnebly. 2005. Agronomic and seed characteristics of soybean with reduced raffinose and stachyose. *Crop Science* 45:589-592.

- Ng, P.C., and S. Henikoff. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research* 31:3812-3814.
- Openshaw, S.J., and H.H. Hadley. 1978. Maternal effects on sugar content in soybean seeds. *Crop Science* 18:581-584.
- Parsons, C.M., Y. Zhang, and M. Araba. 2000. Nutritional evaluation of soybean meals varying in oligosaccharide content. *Poultry Science* 79:1127-1131.
- Peterbauer, T., L. Mach, J. Mucha, and A. Richter. 2002. Functional expression of a cDNA encoding pea (*Pisum sativum* L.) raffinose synthase, partial purification of the enzyme from maturing seeds, and steady-state kinetic analysis of raffinose synthesis. *Planta* 215:839-846.
- Peterbauer, T., L.B. Lahuta, A. Blochl, J. Mucha, D.A. Jones, C.L. Hedley, R.J. Gorecki, and A. Richter. 2001. Analysis of the raffinose family oligosaccharide pathway in pea seeds with contrasting carbohydrate composition. *Plant Physiology* 127:1764-1772.
- Peterbauer, T and A. Richter. 2001. Biochemistry and physiology of raffinose family oligosaccharides and galactosyl cyclitols in seeds. *Seed Science Research* 11:185-197.
- Peterbauer, T., J. Mucha, U. Mayer, M. Popp, J. Glossl, and A. Richter. 1999. Stachyose synthesis in seeds of adzuki bean (*Vigna angularis*): molecular cloning and functional expression of stachyose synthase. *The Plant Journal* 20:509-518.
- Pfaffl, M.W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* 29:e45.
- Schlueter, J.A., J.Y. Lin, S.D. Schlueter, I.F. Vasylenko-Sanders, S. Deshpande, J. Yi, M. O'Bleness, B.A. Roe, R.T. Nelson, B.E. Scheffler, S.A. Jackson, and R.C. Shoemaker. 2007. Gene duplication and paleopolyploidy in soybean and the implications for whole genome sequencing. *BMC Genomics* 8:330.
- Shoemaker, R.C., J. Schlueter, and J.J. Doyle. 2006. Paleopolyploidy and gene duplication in soybean and other legumes. *Current Opinions in Plant Biology* 9:104-109.
- Shoemaker, R.C., K. Polzin, J. Labate, J. Specht, E.C. Brummer, T. Olson, N. Young, V. Concibido, J. Wilcox, J.P. Tamulonis, G. Kochert, and H.R. Boerma. 1996. Genome duplication in soybean (*Glycine* subgenus *soja*). *Genetics* 144:329-338.
- Sleper, D.A., and J.M. Poehlman. 2006. *Breeding field crops*. 5th ed. Blackwell Pub., Ames, Iowa.

- Smiricky, M.R., C.M. Grieshop, D.M. Albin, J.E. Wubben, V.M. Gabert, and G.C. Fahey, Jr. 2002. The influence of soy oligosaccharides on apparent and true ileal amino acid digestibilities and fecal consistency in growing pigs. *Journal of Animal Science* 80:2433-2441.
- Sneller, C.H. 1994. Pedigree analysis of elite soybean lines. *Crop Science* 34:1515-1522.
- Suarez, F.L., J. Springfield, J.K. Furne, T.T. Lohrmann, P.S. Kerr, and M.D. Levitt. 1999. Gas production in human ingesting a soybean flour derived from beans naturally low in oligosaccharides. *American Journal of Clinical Nutrition* 69:135-139.
- Wang, J., K. Chuang, M. Ahluwalia, S. Patel, N. Umblas, D. Mirel, R. Higuchi, and S. Germer. 2005. High-throughput SNP genotyping by single-tube PCR with Tm-shift primers. *Biotechniques* 39:885-893.
- Zuo, Y., G.C. Fahey, Jr., N.R. Merchen, and N.L. Bajjalieh. 1996. Digestion responses to low oligosaccharide soybean meal by ileally-cannulated dogs. *Journal of Animal Science* 74:2441-2449.
- Zuther, E., K. Buchel, M. Hundertmark, M. Stitt, D.K. Hinch, and A.G. Heyer. 2004. The role of raffinose in the cold acclimation response of *Arabidopsis thaliana*. *FEBS Letters* 576:169-173.

CHAPTER 3

Raffinose and stachyose metabolism are not required for efficient soybean seed germination

SUMMARY

Raffinose family oligosaccharides, which include raffinose and stachyose, are thought to be an important source of energy during seed germination. In contrast to their potential for promoting germination, raffinose and stachyose represent anti-nutritional units for monogastric animals when consumed as a component of feed. The exact role for raffinose and stachyose during soybean seed development and germination has not been experimentally determined; but it has been hypothesized that raffinose and stachyose are required for successful germination. Previously, inhibition of raffinose and stachyose breakdown during imbibition and germination was shown to significantly delay germination in pea seeds. The objective of this study was to compare the germination potential for soybean seeds with either wild-type or low raffinose oligosaccharide levels and to examine the role of raffinose and stachyose breakdown in germination of soybean seeds. There was no significant difference in germination between normal and low raffinose and stachyose soybean seeds when imbibed/germinated in water. Similar to the situation in pea, soybean seeds of wild-type carbohydrate composition experienced a delay in germination when treated with a chemical inhibitor of α -galactosidase activity (1-deoxygalactonojirimycin or DGJ) during imbibition. However, low raffinose and stachyose soybean seed germination was not significantly delayed or reduced when treated with DGJ. In contrast to the situation in pea, the inhibitor-induced germination delay in wild-type soybean seeds was not partially overcome by the addition of galactose or sucrose. We conclude that raffinose and stachyose are not an essential source of energy during soybean seed germination.

INTRODUCTION

Although the exact function of raffinose family oligosaccharides in germinating seeds is largely unknown, it is clear that seeds require a large amount of energy during germination (Bewley and Black, 1994). This energy is hypothesized to come from stored carbohydrates; sucrose, raffinose and stachyose are the most abundant of the soluble sugars (Peterbauer et al., 2001) but account for only a small portion of the total carbohydrates present in the seeds (Ziegler, 1995). Raffinose oligosaccharides (raffinose and stachyose) have previously been thought to be involved in seed protection during the desiccation process (Bailly et al., 2001; Black et al., 1996; Obendorf, 1997) by stabilization of membranes (Crowe et al., 1987). Recently, low raffinose and stachyose and low raffinose, stachyose, and phytin lines were tested for their sensitivity to imbibitional chilling (Obendorf et al., 2008). The results indicated that lines low in raffinose and stachyose were similar to control lines, with no sensitivity, while lines low in raffinose, stachyose, and phytin were sensitive to imbibitional chilling (Obendorf et al., 2008). However, many of the proposed functions of raffinose and stachyose in the developing, stored, and germinating seed have not been confirmed experimentally (Bentsink et al., 2000; Gurusinghe and Bradford, 2001), but are generally considered valid. Since the breakdown of seed components during germination has been studied extensively, it is known that protein, oils and polymeric carbohydrates each break down slowly over a period of several days (Bewley and Black, 1994). This period extends beyond germination and thus cannot be solely responsible for fulfilling the energy demands of the germinating seed.

A soybean plant introduction line, PI 200508, was identified as having reduced levels of raffinose and stachyose and elevated levels of sucrose (Kerr and Sebastian, 2000). Initial characterization of soybean line PI 200508 was carried out by Hitz et al. (2002). Seeds of PI 200508, in addition to reduced raffinose and stachyose, had a 25-fold reduction in raffinose synthase enzyme activity (Hitz et al., 2002). Each of these characteristics potentially improves the nutritional quality of the soybean meal and together highlights the soybean line PI 200508 as an important source for the low raffinose oligosaccharide trait. In a subsequent study, Neus et al. (2005), reported that the vigor of backcrossed PI 200508-derived low raffinose and stachyose lines was not affected as there were no significant differences in a number of seed quality traits including field emergence, seed yield, maturity, and fatty acid content as compared to derived lines with a wild-type raffinose oligosaccharide phenotype. We recently determined that the molecular genetic basis for the low raffinose oligosaccharide trait in soybean line PI 200508 was a variant allele of a raffinose synthase gene (Dierking and Bilyeu, 2008). This discovery allowed us to develop closely related soybean germplasm without direct agronomic selection, but with contrasting raffinose and stachyose levels that could be used to produce seeds for germination studies.

Blöchl et al. (2007) directly tested the effect of inhibition of the breakdown of raffinose oligosaccharides on germination in pea. The results of this study indicated that pea seeds exhibited a significant delay in germination when α -galactosidase activity and thus raffinose oligosaccharide metabolism was inhibited by 1-deoxygalactonojirimycin (DGJ). Further, the germination delay in DGJ-treated pea seeds could be partially overcome by the addition of galactinol (Blöchl et al., 2007). Because of the implication

that raffinose and stachyose play critical roles in seed development and germination, the objective of this project was to determine (without any selection for agronomic properties) the germination potential of soybean lines with low raffinose and stachyose as compared to sibling lines with wild-type raffinose and stachyose content. In addition, we sought to examine the effect of blocking raffinose oligosaccharide metabolism in soybean seed germination.

MATERIALS AND METHODS

Plant Material

A 'Williams 82' (Bernard and Cremeens, 1988) X PI 200508 cross was made in the summer of 2005 at Bradford Research and Extension Center near Columbia, Missouri. Successful crosses were carried through F₂. A population of approximately 84 F₂ plants was grown to maturity in the summer of 2007. The resulting F_{2:3} seeds were phenotyped for seed oligosaccharide content using High Performance Liquid Chromatography (HPLC). For each plant, 8 individual seeds were analyzed (Dierking and Bilyeu, 2008). Five lines were selected with wild-type raffinose oligosaccharide phenotypes and five lines with low raffinose oligosaccharide phenotypes. Three hundred F_{2:3} seeds from each line were used for the germination experiments described below. The Williams 82 seeds used were produced in the summer of 2008.

HPLC Oligosaccharide Analysis

Oligosaccharides were separated and quantified by high performance ion chromatography with pulsed amperometric detection (PAD) employing an Agilent 1100 series HPLC and an ESA Coulochem III detector (Agilent Technologies, Chesterfield, MO, USA). A 12.5 mg ground seed sample was extracted with 0.5 ml 50% ethanol at 70 °C, 30 min. Samples were then centrifuged 15 min at 16,000g. The supernatant was passed through a 0.2 µm filter. Sugars were separated on a Dionex Carbo Pac PA 10 analytical column (250 mm x 4 mm, 10 µm) connected to a Carbo Pac PA 10 guard column (50 mm x 4 mm). The mobile phase was 90 mM NaOH with flow rate of 1.5 mL

min⁻¹, maintained at 30 °C. Detection settings were: time 0, 0.1 v, time 0.41, -2.0 v, time 0.42, 0.6 v, and time 0.44, -0.1 v.

Germination Experiments

Germination experiments for wild-type and low raffinose oligosaccharide soybeans included two treatments, using water (control) or DGJ (Industrial Research, Wellington, New Zealand), and were carried out essentially as described by Blöchl et al., (2007). Seeds were imbibed at 25 °C in the dark in Petri dishes (50 seeds per 15 mm Petri dish) with 25 mL of either water (control) or 50 µM DGJ. Seeds were imbibed for 16 hours and then transferred to containers lined with wetted paper towels (DGJ treatment was wetted with 50 µM DGJ). One hundred fifty seeds were used per line, per treatment. The seeds were placed in 25 °C in the dark for germination. Germination was defined as the point at which the radical pierced the seed coat. The number of germinated seeds was counted at six different time intervals: 23, 40, 47, 64, 71, and 88 hours (h); where 0 h is the time at which seeds are introduced to either water or DGJ in the Petri dish. Germination experiments were performed for five lines (F_{2:3} seeds) for each phenotype, either wild-type raffinose oligosaccharide levels or low-raffinose oligosaccharide levels.

Germination experiments for Williams 82 seeds (summer 2008, Bradford Research and Extension Center) included six treatments: water (control), 50 µM DGJ, 25 mM sucrose (Sigma, St. Louis, MO), 50 mM galactose (Sigma, St. Louis, MO), 50 µM DGJ + 25 mM sucrose, and 50 µM DGJ + 50 mM galactose. The experimental time points and data collection were carried out exactly as described above except three replicates of 50 seeds were used for each treatment.

RESULTS AND DISCUSSION

Soybean seeds were produced in a field environment from plants that stably inherited wild-type or low raffinose oligosaccharide phenotypes depending on their inheritance of homozygous variant alleles of a raffinose synthase gene (RS2) from the wild-type (Williams 82) or low raffinose oligosaccharide (PI 200508) parent (Dierking and Bilyeu, 2008). Plants were related F₂ individuals that produced F_{2:3} seeds, but were not selected for any particular agronomic characteristics. Five F_{2:3} lines were selected with wild-type raffinose oligosaccharide profiles and homozygous wild-type RS2 alleles, and five lines were selected with low raffinose oligosaccharide profiles and homozygous variant alleles of RS2 (Table 3.1). In the lines with the low raffinose oligosaccharide profile, sucrose was 165% of the wild-type level, and raffinose and stachyose decreased to 18% and 33% of the wild-type levels, respectively, consistent with a reduction in raffinose synthase enzyme activity in developing seeds (Hitz et al., 2002). Five F_{2:3} lines of each raffinose oligosaccharide profile were chosen to overcome the effect of other unknown segregating genes which might influence germination.

Germination was monitored for the lines with contrasting raffinose oligosaccharide profiles to determine if reduced raffinose and stachyose would have a negative impact on standard germination. Standard germination was efficient for both types of lines, and reached a plateau of approximately 95% germination by 64 h after imbibition (Figure 3.1A). There were no significant differences for standard germination in the control treatment between wild-type and low raffinose oligosaccharide lines.

Given that standard germination was not decreased in low raffinose oligosaccharide lines, the germination potential of soybean seeds treated with the α -galactosidase inhibitor, DGJ, was evaluated. Treatment of imbibing pea seeds with DGJ was previously shown to significantly delay germination; at 42 h after imbibition, only 25% of DGJ-treated pea seeds germinated compared to 90% for the control treatment (Blöchl et al., 2007). Since DGJ inhibits α -galactosidase activity, and the addition of galactose was able to partially overcome the germination delay in pea, we investigated the effect on germination for DGJ-treated soybean seeds. Similar to the situation in pea, soybean lines with wild-type raffinose and stachyose levels exhibited a delay in germination for the DGJ-treatment compared to the control treatment (Figure 3.1B). Although germination was statistically significantly lower for the DGJ-treated seeds for every time point beyond 23 h after imbibition, at 88 h the difference between control treated and DGJ-treated seeds was only 98% to 92%, respectively. The maximum difference in average germination between control and DGJ- treated wild-type raffinose and stachyose soybean seeds was at 40 h and only reduced germination from 82% to 61%, a less drastic reduction than was previously observed for pea. Germination for DGJ-treated low raffinose and stachyose soybeans was not significantly different at any time point (Figure 3.1C). This result indicates that the germination of soybeans with low raffinose oligosaccharide phenotypes was not affected by the α -galactosidase inhibitor; soybeans with wild-type raffinose oligosaccharide phenotypes exhibited a significant but less drastic delay in germination compared to that observed in pea seeds. Soybean seeds appear to germinate efficiently with reduced raffinose and stachyose content, and mobilization of raffinose and stachyose during soybean germination has different

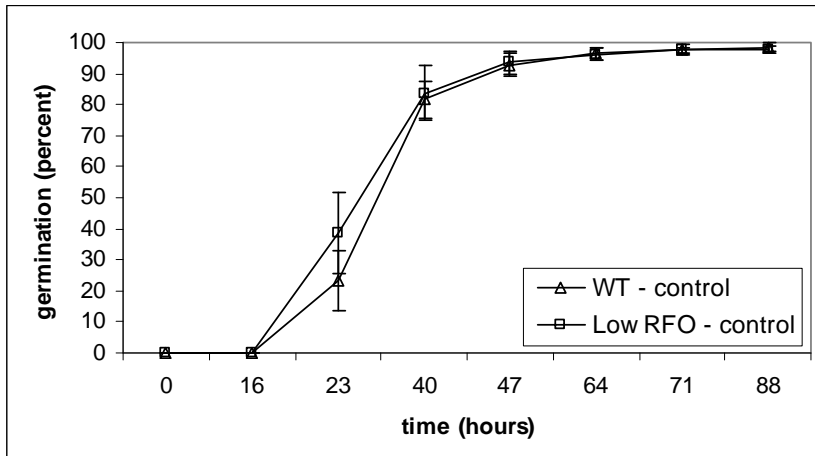
responses to DGJ-treatment than mobilization of raffinose and stachyose during pea germination (Blöchl et al., 2007).

In order to further understand the delay in germination of wild-type raffinose oligosaccharide soybeans treated with DGJ, we carried out additional experiments using seeds of the cultivar Williams 82, the source of the wild-type raffinose oligosaccharide phenotype. The mean oligosaccharide content as determined by HPLC for four Williams 82 seeds is 4.53% of dry weight (DW) sucrose, 0.45% DW raffinose, and 3.40% DW stachyose. Six treatments were employed: water (control), DGJ, sucrose, galactose, DGJ + sucrose, and DGJ + galactose (Figure 3.2). As expected, treatment with DGJ induced a minor reduction in early germination, but by 88 h after imbibition there was no significant difference between any treatment. The maximum difference in average germination between control and DGJ-treated Williams 82 soybean seeds was at 23 h and reduced germination from 42% to 16%. Unlike the situation in pea (Blöchl et al., 2007), the addition of either sucrose or galactose also induced a significant reduction in soybean germination at 23 h, although there was no significant difference as compared to the control treatment from 40 h through 88 h. The addition of sucrose or galactose to DGJ-treated seeds did not overcome the reduction in germination. At 23 h all treatments significantly reduced germination as compared to the control treatment (Figure 3.3). The control treated seeds had significantly higher germination at this time point; there were no significant differences between any of the remaining treatments. This result is in contrast with germination in pea seeds where the addition of sucrose and galactose partially overcame the reduced germination effect of DGJ on germinating seeds (Blöchl et al., 2007).

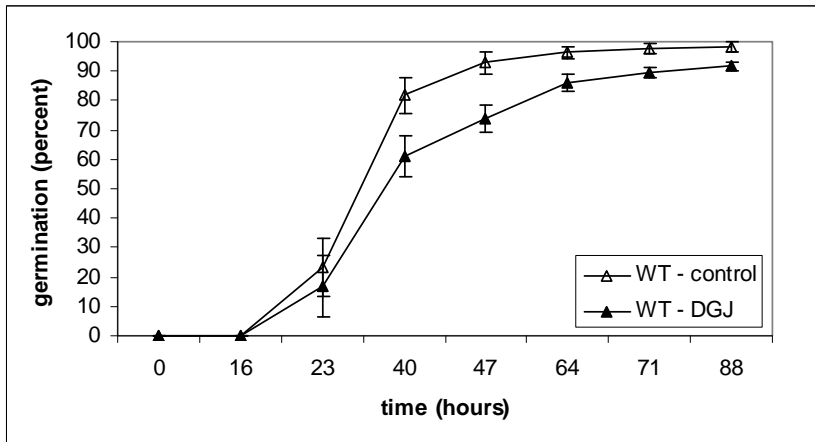
Overall, the results of this study indicate that wild-type levels of raffinose and stachyose are not required for efficient germination of soybean seeds. The suggested necessity of raffinose and stachyose mobilization for efficient seed germination is shown here to be inconsequential, at least for soybeans. There was no significant difference between germination of control treated wild-type raffinose oligosaccharide and low-raffinose oligosaccharide soybeans; the addition of the α -galactosidase inhibitor DGJ delayed the germination of wild-type oligosaccharide soybeans, but DGJ had no effect on soybeans with a reduced raffinose and stachyose oligosaccharide content. Further, the reduction at 23 h in germination of the DGJ-treated Williams 82 seeds could not be overcome by the addition of sucrose or galactose. Our results demonstrate that soybean and pea have a fundamental difference in oligosaccharide requirements during germination. These data refute the long held assumption that accumulated seed raffinose and stachyose are required for successful soybean seed germination. We have shown that reducing raffinose and stachyose in soybean seeds, without selection for agronomic characteristics, does not reduce or delay soybean seed germination. This neutral germination impact provides additional incentive for soybean breeding programs to incorporate the low raffinose oligosaccharide trait into elite varieties to increase the value of soybean meal used for consumption by monogastric animals.

FIGURES

A



B



C

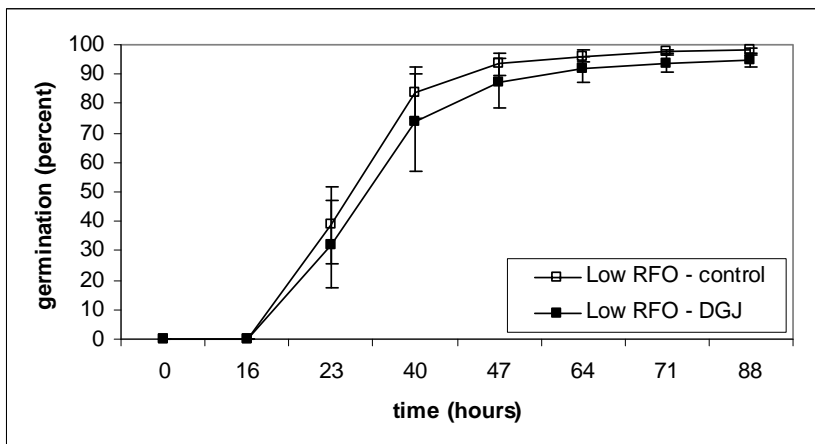
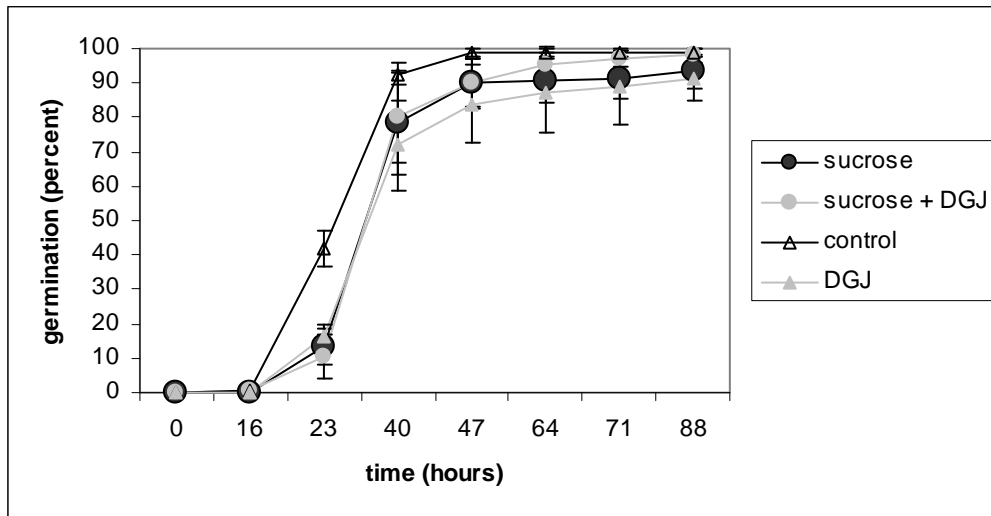


Figure 3.1: Germination of wild-type (WT) and Low raffinose and stachyose (Low RFO) soybean seeds in water (control) or DGJ from 0 to 88 hours after imbibition. Data points represent the means of 5 independent lines, 150 seeds each. Error bars represent plus or minus one standard deviation from the mean. A. Control germination for WT and Low raffinose and stachyose soybean seeds. B. Control and DGJ treatment for WT

oligosaccharide soybean seeds. C. Control and DGJ treatment for Low raffinose and stachyose soybean seeds.

A



B

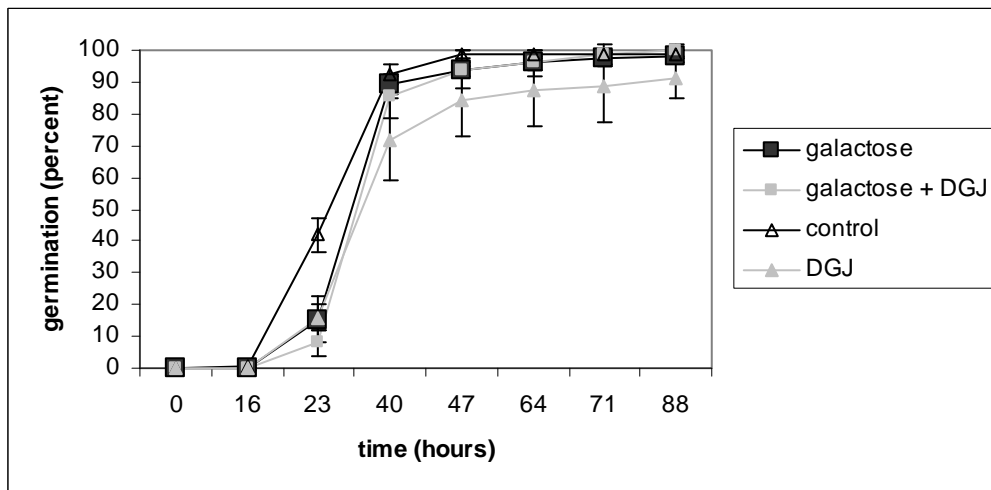


Figure 3.2: Germination of Williams 82 soybean seeds in water (control), sucrose, sucrose + DGJ, galactose, galactose + DGJ, or DGJ from 0 to 88 hours after imbibition. Data points represent the means of three replicates of 50 seeds. Error bars represent \pm STD. A. Germination of Williams 82 soybean seeds in water (control), sucrose, sucrose + DGJ, or DGJ. B. Germination of Williams 82 soybean seeds in water (control), galactose, galactose + DGJ, or DGJ.

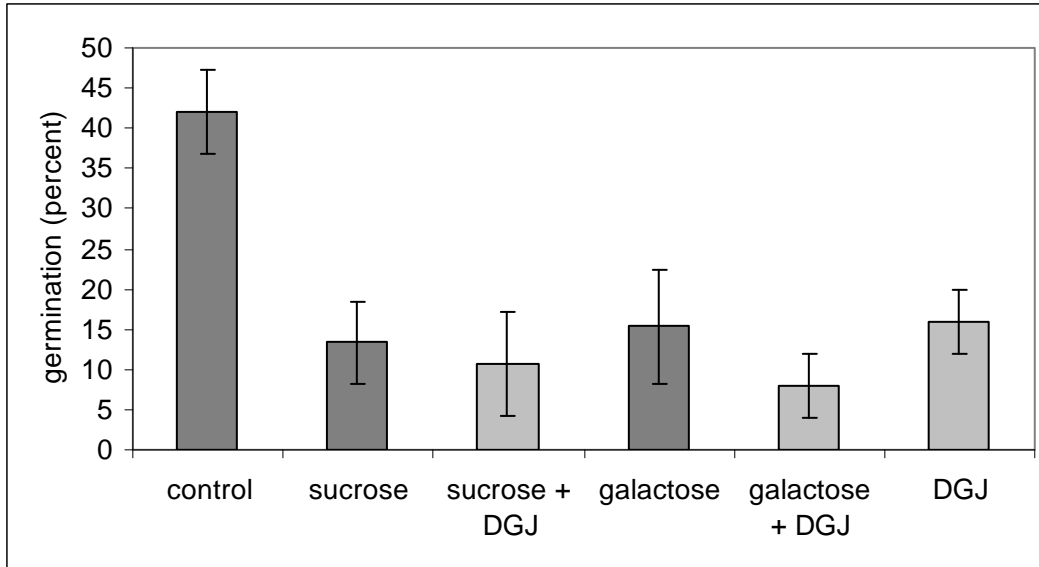


Figure 3.3: Germination of Williams 82 seeds in water (control), sucrose, sucrose + DGJ, galactose, galactose + DGJ, and DGJ at 23 hours after imbibition. Each bar represents the mean of three replicates of 50 seeds each. Error bars represent \pm STD.

TABLE

Table 3.1: Oligosaccharide content of wild-type and low raffinose and stachyose dry seeds.

Line	Genotype	Sucrose (%DW†)		Raffinose (% DW)		Stachyose (% DW)	
		mean	STD‡	mean	STD	mean	STD
1	WT§	4.85	1.76	1.13	0.27	5.99	0.69
2	WT	4.86	0.89	0.99	0.25	5.79	0.43
3	WT	5.05	1.12	1.13	0.13	4.77	0.64
4	WT	3.79	0.54	1.07	0.09	5.96	0.57
5	WT	3.51	0.71	0.85	0.14	4.65	0.81
	mean	4.41	1.00	1.03	0.18	5.43	0.14
6	Low RFO¶	6.06	1.08	0.19	0.04	2.26	0.34
7	Low RFO	6.19	1.36	0.20	0.03	1.47	0.44
8	Low RFO	9.35	1.03	0.18	0.03	1.78	0.24
9	Low RFO	7.74	0.61	0.18	0.07	1.87	0.60
10	Low RFO	7.00	1.40	0.18	0.05	1.58	0.41
	mean	7.27	1.09	0.19	0.04	1.79	0.41

n = 8

† DW, dry weight

‡ STD, standard deviation

§ WT, wild type

¶ RFO, raffinose oligosaccharides, raffinose and stachyose

LITERATURE CITED

- Bailly, C., C. Audigier, F. Ladonne, M.H. Wagner, F. Coste, F. Corbineau, and D. Come. 2001. Changes in oligosaccharide content and antioxidant enzyme activities in developing bean seeds as related to acquisition of drying tolerance and seed quality. *Journal of Experimental Botany* 52:701-708.
- Bentsink, L., C. Alonso-Blanco, D. Vreugdenhil, K. Tesnier, S.P. Groot, and M. Koornneef. 2000. Genetic analysis of seed-soluble oligosaccharides in relation to seed storability of *Arabidopsis*. *Plant Physiology* 124:1595-1604.
- Bernard, R.L., and C.R. Cremeens. 1988. Registration of 'Williams 82' soybean. *Crop Science* 28:1027-1028.
- Bewley, J.D., and M. Black. 1994. *Seeds: Physiology of Development and Germination*. Second ed. Plenum Press, New York.
- Black, M., F. Corbineau, M. Grzesikit, P. Guyi, and D. Côme. 1996. Carbohydrate metabolism in the developing and maturing wheat embryo in relation to its desiccation tolerance. *Journal of Experimental Botany* 47:161-169.
- Blöchl, A., T. Peterbauer, and A. Richter. 2007. Inhibition of raffinose oligosaccharide breakdown delays germination of pea seeds. *Journal of Plant Physiology* 164:1093-1096.
- Crowe, J.H., L.M. Crowe, J.F. Carpenter, and C. Aurell Wistrom. 1987. Stabilization of dry phospholipid bilayers and proteins by sugars. *Biochemistry Journal* 242:1-10.
- Dierking, E.C., and K.D. Bilyeu. 2008. Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. *The Plant Genome* 1:135-145.
- Gurusinghe, S., and K.J. Bradford. 2001. Galactosyl-sucrose oligosaccharides and potential longevity of primed seeds. *Seed Science Research* 11:121-134.
- Hitz, W.D., T.J. Carlson, P.S. Kerr, and S.A. Sebastian. 2002. Biochemical and molecular characterization of a mutation that confers a decreased raffinose and phytic acid phenotype on soybean seeds. *Plant Physiology* 128:650-660.
- Kerr, P.S., and S.A. Sebastian. 2000. Soybean products with improved carbohydrate composition and soybean plants. United States Patent 6147193. Date Issued: November 14, 2000.
- Neus, J.D., W.R. Fehr, and S.R. Schnebly. 2005. Agronomic and seed characteristics of soybean with reduced raffinose and stachyose. *Crop Science* 45:589-592.

- Obendorf, R.L. 1997. Oligosaccharides and galactosyl cyclitols in seed desiccation tolerance. *Seed Science Research* 7:63-74.
- Obendorf, R.L., A.D. Zimmerman, P.A. Ortiz, A.G. Taylor, and S.R. Schnebly. 2008. Imbibitional chilling sensitivity and soluble carbohydrate composition of low raffinose, low stachyose soybean seed. *Crop Science* 48:2396-2403.
- Peterbauer, T. and A. Richter. 2001. Biochemistry and physiology of raffinose family oligosaccharides and galactosyl cyclitols in seeds. *Seed Science Research* 11:185-197.
- Ziegler, P. 1995. *Seed development and germination*. Marcel Dekker Inc, New York.

CHAPTER 4

Candidate soybean raffinose synthase gene RS3 does not associate with altered seed oligosaccharide phenotype

SUMMARY

Oligosaccharides are an important component of soybean (*Glycine max*) meal in terms of metabolizable energy for monogastric animals. Sucrose, raffinose, and stachyose are the three main oligosaccharides present in soybean meal. Of the three, only sucrose is nutritionally useful. When raffinose and stachyose are fermented by microbes present in the gut, the results are flatulence and discomfort which ultimately lead to poor weight gain. Previously the major gene, RS2, underlying a naturally occurring low raffinose and stachyose phenotype in PI 200508 was determined. However, the soybean genome is complex and it is likely that multiple genes contribute to a single phenotype. The objective of this research was to determine the contribution of a second candidate gene, RS3, to the seed oligosaccharide phenotype by measuring the effect of mutant RS3 when the soybean raffinose synthase gene RS2 is set or no longer segregating. The results indicate that RS3 does not contribute to the seed phenotype because the Williams 82 and PI 200508 alleles of this gene segregated independent of seed oligosaccharide phenotype when the genotype of RS2 was set. However, an enhanced oligosaccharide phenotype was observed which had a greater reduction in raffinose and stachyose compared to the PI 200508 phenotype.

INTRODUCTION

Today, soybean represents one of the most important economic crops in the United States. It is considered similar to corn in acreage and second only to corn in value (Sleper and Poehlman, 2006). The oil, protein, and carbohydrate composition of the soybean seed generally controls its use. Seeds of soybean cultivars in the United States have an average composition of 20% oil, 40% protein, and 15% soluble carbohydrates in dry weights of cotyledons of ungerminated seeds (Hsu et al., 1973). Soybean meal is a major component of the diets of monogastric animals and its usefulness is determined, in part, by the carbohydrate component. The carbohydrate component of soybean meal is comprised of three major oligosaccharides: sucrose, raffinose, and stachyose (Openshaw and Hadley, 1978). Of the three, only sucrose is nutritionally useful and can be fully digested by monogastric animals. Raffinose and stachyose are considered anti-nutritional units because they cannot be digested due to the lack of α -galactosidase activity in the gut of monogastric animals.

Removing raffinose and stachyose from soybean meal has been reported to increase the metabolizable energy of the diet by as much as 20% (Coon et al., 1990). The effects of raffinose and stachyose in diets have been studied in pigs (Smiricky et al., 2002), dogs (Zuo et al., 1996), chickens (Parsons et al., 2000), and humans (Suarez et al., 1999). Generally, raffinose and stachyose are poorly digested by monogastrics; the removal of raffinose-saccharides from soybean meal increases the metabolizable energy of the diet and reduces flatulent production (Coon et al., 1990; Parsons et al., 2000; Suarez et al., 1999).

A plant introduction line, PI 200508, was identified that has reduced levels of raffinose and stachyose and elevated levels of sucrose (Kerr and Sebastian, 2000; Kerr and Sebastian, 2003). PI 200508 is synonymous with the LR 28 abbreviation (Hitz et al., 2002; Kerr and Sebastian, 2000; Kerr and Sebastian, 2003) which was identified in a screen for modified carbohydrate composition. Initial characterization of the line PI 200508 was reported by Hitz et al. (2002). Recently, the major gene underlying this phenotype was reported (Dierking and Bilyeu, 2008).

A mutant allele of the soybean raffinose synthase gene, RS2, has been associated with the low raffinose and stachyose phenotype in PI 200508 (Dierking and Bilyeu, 2008). However, due to ancient soybean genome duplications (Schlueter et al., 2007; Shoemaker et al., 2006), it is likely to expect multiple genes to contribute to a single phenotype or enzymatic activity. Therefore, the goal of this research was to identify the contribution, if any, of the soybean candidate gene RS3 to the seed oligosaccharide phenotype.

METHODS AND MATERIALS

Population Development

A 'Williams 82' (Bernard and Cremeens, 1988) X PI 200508 cross was made in the summer of 2005 at Bradford Research and Extension Center near Columbia, Missouri. One hundred seeds of two lines of F₂:F₃ seeds which were homozygous for either parental allele of RS2 were planted in germination packets (CYG, Mega International, St. Paul, MN). The resulting F₄ plants were genotyped for RS3; approximately 50 plants from each of the four RS2, RS3 genotypic classes were transplanted to the field late in the 2008 season and grown to maturity. For each surviving plant, four individual seeds were analyzed by HPLC. The parents of the cross, Williams 82 and PI 200508 were also grown in the field at the same time, however the Williams 82 was planted on May 20 and the PI 200508 was planted June 24 while the experimental lines were germinated in packets on June 25 and transplanted to the field on July 8.

Allele-Specific Molecular Marker Assay

The genotypes of RS2 and RS3 were assayed as described (Dierking and Bilyeu, 2008).

Oligosaccharide Phenotype Determination by HPLC

Oligosaccharides were determined by high performance ion chromatography with pulsed amperometric detection (PAD) employing an Agilent 1100 series HPLC and an

ESA Coulochem III detector (Agilent Technologies, Chesterfield, MO, USA). A 12.5 mg ground seed sample was extracted with 0.5 ml 50% ethanol at 70 °C, 30 min. Samples were then centrifuged 15 min at 16,000g. The supernatant was passed through a 0.2 µm filter. Sugars were separated on a Dionex Carbo Pac PA 10 analytical column (250 mm x 4 mm, 10 µm) connected to a Carbo Pac PA 10 guard column (50 mm x 4 mm). The mobile phase was 90 mM NaOH with flow rate of 1.5 ml min⁻¹, maintained at 30 °C. Detection settings were: time 0, 0.1 v, time 0.41, -2.0 v, time 0.42, 0.6 v, and time 0.44, -0.1 v.

RESULTS

Due to the unique nature of the ancestral soybean genome having undergone full genome duplications (Shoemaker et al., 2006; Shoemaker et al., 1996), the existence of multiple raffinose synthase genes is likely. Previously a soybean gene RS2 has been shown to be completely associated with a soybean seed oligosaccharide phenotype; however, an association with the candidate gene RS3 was not apparent (Dierking and Bilyeu, 2008). A population consisting of four F_{2:3} lines homozygous for wild-type (Williams 82) or mutant (PI 200508) RS2 alleles was developed to study the effects of the gene RS3 on the oligosaccharide phenotype. Since the contribution of RS2 to the seed oligosaccharide phenotype was very large, a homozygous RS2 population allowed any contribution of RS3 to be more easily identified. Two of the four lines were homozygous, wild-type at RS2 and the other two lines were homozygous, mutant at the RS2 locus. The experimental design allowed for the four distinct RS2, RS3 genotypic classes to produce F_{2:4} seed so that the effect of the RS3 genotype on the oligosaccharide phenotype could be more accurately assessed. Four seeds from each of the surviving seven to ten plants from each genotypic class were analyzed by HPLC for oligosaccharide phenotype (Figure 4.1).

The expected wild-type oligosaccharide phenotype was observed for the lines which carried the Williams 82 allele of RS2 regardless of RS3 genotype (Figure 4.1). This is the expected outcome for seeds which carry the homozygous, wild-type allele for the gene RS2 as previous data have suggested that the inheritance of a single wild-type allele of RS2 is sufficient for a wild-type seed oligosaccharide phenotype (Dierking and Bilyeu, 2008).

For those lines that carried the mutant PI 200508 alleles at RS2, a mutant phenotype of reduced raffinose and stachyose was observed (Figure 4.1). This phenotype is characterized by significantly less raffinose and stachyose compared to the PI 200508 parent seeds grown in the same environment. Previously, the phenotype observed for plants which were homozygous for the PI 200508 allele at RS2 were not significantly different than PI 200508 seed produced in the same environment (Dierking and Bilyeu, 2008). A similar phenotype has been observed by another group, which describes a greater reduction in raffinose and stachyose due to one mutant gene in addition to modifier genes present in the genome (Kerr and Sebastian, 2003). It does not appear that RS3 is the causative factor, or modifier gene, involved in the enhanced mutant phenotype as RS3 segregates independently of the altered seed oligosaccharide phenotype.

DISCUSSION

Due to the fact that the soybean genome is redundant, it is likely to expect multiple genes to contribute to a single phenotype or to expect that genes with similar functions will have localized or specialized function (Chappell et al., 2006). This does not appear to be the case in terms of soybean raffinose synthase genes given that RS3 has high transcript levels across a number of tissues including leaves, seedlings, and multiple stages of developing seeds (Dierking and Bilyeu, 2008); however, the effect of RS3 on oligosaccharide accumulation is not distinguishable between the Williams 82 or PI 200508 allele.

There was no association between the genotype of the candidate raffinose synthase gene RS3 and the seed oligosaccharide phenotype in this population. However, an exaggerated mutant phenotype characterized by a significant decrease in raffinose and stachyose occurred which had not been observed previously in our lines. The lines that contained the PI 200508 allele of RS2 had a higher sucrose:raffinose + stachyose ratio than the parent PI 200508 seed produced in the same environment. This result indicates that there must be an additional factor contributing to this unique phenotype. Since the parent PI 200508 seeds were planted earlier than the segregating lines, it is possible that environment is playing a role in raffinose oligosaccharide accumulation in the seed. Previous research pertaining to the effect of planting date on seed oligosaccharide accumulation conducted on tofu and natto cultivars indicated that seeds produced from plants with later planting dates accumulated more sucrose and less raffinose than those with earlier planting dates (Brown, 2006). An alternative explanation of the enhanced phenotype is that additional genes are involved in this

phenotype. The observation of an enhanced low raffinose oligosaccharide phenotype was reported once before (Kerr and Sebastian, 2003). This report included oligosaccharide data for a “low1” phenotype (Kerr and Sebastian, 2003) which is highly similar to the phenotypes associated with the inheritance of the PI 200508 allele of RS2 (Dierking and Bilyeu, 2008). An additional phenotype was described “low2” which refers to an even lower raffinose and stachyose seed phenotype which resulted from the combination of LR 28 or LR 484 and modifier genes from other germplasm lines (Kerr and Sebastian, 2003). It is possible the “low2” phenotype is potentially the same phenotype that was observed in this population and the result of modifier genes present in the soybean genome which remain uncharacterized, however such a phenotype has not previously been observed in seeds with a Williams 82 background.

FIGURE

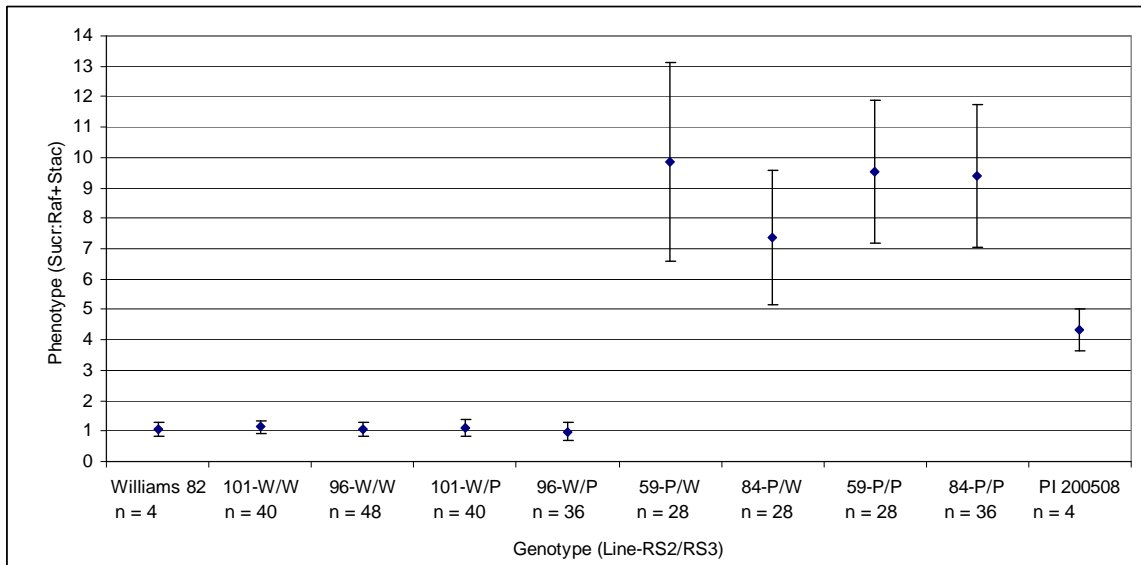


Figure 4.1: Phenotype to genotype association of a soybean population (Williams 82 X PI 200508) set for the RS2 gene and segregating for the RS3 gene. The oligosaccharide phenotype of four individual F₄ seeds was determined for a subset of the population based on homozygous RS3 genotype. The data represents the mean of the ratio of extractable seed sucrose to the sum of raffinose and stachyose. Individual F₃ plants were categorized by their RS2 and RS3 genotype class and separated on the x-axis. Parental genotypes represent individual seeds analyzed from either Williams 82 or PI 200508 grown in the same environment. Error bars represent plus and minus one standard deviation from the mean.

LITERATURE CITED

- Bernard, R.L., and C.R. Cremeens. 1988. Registration of 'Williams 82' soybean. *Crop Science* 28:1027-1028.
- Brown, S.L. 2006. The effect of environment on seed composition of tofu and natto soybean cultivars. M.S. Thesis. University of Missouri, Columbia, MO.
- Chappell, A.S., A.M. Scaboo, X. Wu, H. Nguyen, V.R. Pantalone, and K.D. Bilyeu. 2006. Characterization of the MIPS gene family in *Glycine max*. *Plant Breeding* 125:493-500.
- Coon, C.N., K.L. Leske, O. Akavanichan, and T.K. Cheng. 1990. Effect of oligosaccharide-free soybean meal on true metabolizable energy and fiber digestion in adult roosters. *Poultry Science* 69:787-793.
- Dierking, E.C., and K.D. Bilyeu. 2008. Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. *The Plant Genome* 1:135-145.
- Hitz, W.D., T.J. Carlson, P.S. Kerr, and S.A. Sebastian. 2002. Biochemical and molecular characterization of a mutation that confers a decreased raffinose and phytic acid phenotype on soybean seeds. *Plant Physiology* 128:650-660.
- Hsu, S.H., H.H. Hadley, and T. Hymowitz. 1973. Changes in carbohydrate contents of germinating soybean seeds. *Crop Science* 13:407-410.
- Kerr, P.S., and S.A. Sebastian. 2000. Soybean products with improved carbohydrate composition and soybean plants. United States Patent 6147193. Date Issued: November 14, 2000.
- Kerr, P.S., and S.A. Sebastian. 2003. Soybean products with improved carbohydrate composition and soybean Plants. United States Patent 6653451. Date Issued: November 25, 2003.
- Openshaw, S.J., and H.H. Hadley. 1978. Maternal effects on sugar content in soybean seeds. *Crop Science* 18:581-584.
- Parsons, C.M., Y. Zhang, and M. Araba. 2000. Nutritional evaluation of soybean meals varying in oligosaccharide content. *Poultry Science* 79:1127-1131.
- Schlueter, J.A., J.Y. Lin, S.D. Schlueter, I.F. Vasylenko-Sanders, S. Deshpande, J. Yi, M. O'Bleness, B.A. Roe, R.T. Nelson, B.E. Scheffler, S.A. Jackson, and R.C. Shoemaker. 2007. Gene duplication and paleopolyploidy in soybean and the implications for whole genome sequencing. *BMC Genomics* 8:330.

- Shoemaker, R.C., J. Schlueter, and J.J. Doyle. 2006. Paleopolyploidy and gene duplication in soybean and other legumes. *Current Opinions in Plant Biology* 9:104-109.
- Shoemaker, R.C., K. Polzin, J. Labate, J. Specht, E.C. Brummer, T. Olson, N. Young, V. Concibido, J. Wilcox, J.P. Tamulonis, G. Kochert, and H.R. Boerma. 1996. Genome duplication in soybean (*Glycine* subgenus *soja*). *Genetics* 144:329-338.
- Sleper, D.A., and J.M. Poehlman. 2006. *Breeding field crops*. 5th ed. Blackwell Pub., Ames, Iowa.
- Smiricky, M.R., C.M. Grieshop, D.M. Albin, J.E. Wubben, V.M. Gabert, and G.C. Fahey, Jr. 2002. The influence of soy oligosaccharides on apparent and true ileal amino acid digestibilities and fecal consistency in growing pigs. *Journal of Animal Science*. 80:2433-2441.
- Suarez, F.L., J. Springfield, J.K. Furne, T.T. Lohrmann, P.S. Kerr, and M.D. Levitt. 1999. Gas production in human ingesting a soybean flour derived from beans naturally low in oligosaccharides. *American Journal of Clinical Nutrition* 69:135-139.
- Zuo, Y., G.C. Fahey, Jr., N.R. Merchen, and N.L. Bajjalieh. 1996. Digestion responses to low oligosaccharide soybean meal by ileally-cannulated dogs. *Journal of Animal Science* 74:2441-2449.

CHAPTER 5

Reverse genetics technique, TILLING, utilized to identify novel soybean raffinose synthase mutant

SUMMARY

Although several techniques are available to study gene function, many are less than ideal for soybean. Reverse genetics, a relatively new approach, can be utilized to identify novel mutations in candidate genes; this technique has not yet been utilized to identify a mutation which has a confirmed phenotype in soybean. Using our molecular genetic resources for the raffinose synthase genes in soybean developed as part of this research, four mutations in independent lines were identified in the raffinose synthase gene RS2; two mutations resulted in amino acid mutations and one had an altered seed oligosaccharide phenotype. The mutation which resulted in a seed phenotype was a C to T transition which resulted in a T107I amino acid change in RS2. This missense mutation is in a semi-conserved region of plant raffinose synthase sequences. The phenotype was an increase in seed sucrose levels as well as a decrease in both raffinose and stachyose seed oligosaccharide levels. This is similar to previously observed seed oligosaccharide phenotypes in RS2-mutant (PI 200508) allele containing lines. Due to the anti-nutritional characteristics of raffinose and stachyose, this represents a positive change in seed composition. A molecular marker assay was developed to reliably detect the inheritance of the mutant allele and can be used in efficient breeding of this seed phenotype. The work serves as the first demonstration of the identification of a soybean mutant through the reverse genetics technique TILLING.

INTRODUCTION

Few tools are available to characterize soybean genetically despite the fact that it represents an important agronomic crop in terms of protein and oil. To date a number of techniques which were first proven in other crops have been applied to soybean with limited success. Transformation of soybean with either *Agrobacterium tumefaciens* or *A. rhizogenes* is not very efficient (Olhoft et al., 1996) and can be limited by genotype specificity (Somers et al., 2003). Another challenge that exists with *A. rhizogenes* transformation is that the mutant phenotype is not heritable (Cho et al., 2000; Subramanian et al., 2005) making it impossible to use transformation with an ultimate goal of breeding the newly identified mutant alleles into elite germplasm. Alternatively, it is possible to use particle bombardment with soybean, but this method can cause multiple copies of DNA to be introduced resulting in co-suppression (Hadi et al., 1996), which generally complicates any effort to genetically characterize a novel phenotype. A complete loss in gene function, gene knockout, makes identifying the function of a particular gene much more straight forward while avoiding co-suppression and creating a heritable phenotype. RNAi, for gene knockout, has been demonstrated in soybean in a few cases (Flores et al., 2008; Li et al., 2006; Subramanian et al., 2005), but relies on transformation which generally has a low success rate in soybean. The development of mutagenized soybean populations either by chemical mutagenesis or by bombardment with fast neutrons represents the opportunity to create novel, heritable, and breedable alleles with desirable phenotypes.

Reverse genetics can be used to identify novel, induced mutations in candidate genes. This is in contrast to forward genetics, a more traditional approach, which relies

on the identification of a mutant phenotype followed by the investigation of the causative gene. Utilizing reverse genetics allows us to take advantage of genes characterized in other plant genomes and use this knowledge to create a pool of candidate genes in soybean which can then be screened using reverse genetics to identify the resulting phenotype and therefore function of the gene. Alternatively this technique can be applied to genes with known function to create allelic series.

Currently there are two distinct methods of creating mutagenized populations in soybean, chemical mutagenesis and radiation-induced mutagenesis. Chemical mutagenesis, either EMS (ethyl-methanesulfonate) or NMU (N-nitroso-N-methylurea) typically induces single nucleotide polymorphisms (Cooper et al., 2008). These point mutations are extremely useful in studying gene function as well as their potential use in crop improvement (Cooper et al., 2008). The mutants are generally characterized by knocked-down or altered gene function rather than a knock-out; for the population used in this study, the mutation distribution is 66% missense, 34% silent, and 0% truncation (Cooper et al., 2008). Targeting Induced Local Lesions IN Genomes (TILLING) is a technique that serves as a high throughput method to identify unique, chemically induced mutations within target genes which have potential to change gene expression and/or function (Colbert et al., 2001; McCallum et al., 2000a; McCallum et al., 2000b).

Radiation-induced mutation can take the form of fast neutron, x-ray, or gamma irradiation. For this work, a mutagenized population was created by bombardment with fast neutrons in a nuclear reactor. Fast neutrons create deletion mutations rather than C to T or G to A transitions as in chemical mutagenesis (Li et al., 2001). The size of the deletion is variable, but for an *Arabidopsis* population detected deletions ranged from 0 to

12 kb with the most common deletion ranging from 2 to 4 kb (Li et al., 2001). Clearly identifying a deletion mutation has powerful implications in studying gene function since most deletions will completely knock-out gene function. However, screening for such mutations is challenging with current PCR methods (Li et al., 2001).

The purpose of this research is to demonstrate the usefulness of utilizing mutagenized populations to identify novel mutations in soybean in order to study gene function. Specifically, the objective was to identify additional mutations in the soybean raffinose synthase gene, RS2, as well as identify novel mutations in the candidate raffinose synthase gene, RS3. Raffinose synthase catalyzes the biochemical reaction to produce raffinose from sucrose and galactinol. Stachyose is formed in a stepwise reaction utilizing raffinose and galactinol as substrates. Both raffinose and stachyose are indigestible by monogastric animals and are therefore considered anti-nutritional components of soybean meal. Previously, the PI 200508 allele of RS2 was associated with the low raffinose and stachyose seed phenotype (Dierking and Bilyeu, 2008). Screening the EMS-mutagenized populations represents the opportunity to find additional raffinose synthase mutants in RS2 or mutations in the candidate gene RS3 for breeding into elite germplasm as well as demonstrate TILLING as a useful genetic tool in soybean.

METHODS AND MATERIALS

Population Development

The 'Williams 82' (Bernard and Cremeens, 1988) EMS mutagenized population screened in this study was described by Cooper et al. (2008). The population screened was exposed to 40 mM EMS and was developed by Dr. Kristin Bilyeu. M₁ plants were advanced to M₂ families, leaf tissue was collected and DNA prepared from a single M₂ plant from each family. M_{2,3} seeds were catalogued for storage.

The fast neutron mutagenized population screened in this study was developed by exposing seeds to 10 Gy of fast neutron radiation in a nuclear reactor. As for the EMS population, M₁ plants were advanced to M₂ families; leaf tissue was collected and DNA prepared from plot row and column pools of single M₂ plants from each family. M_{2,3} seeds were catalogued for storage.

Population Screening

EMS mutagenized population

A portion of each RS2 and RS3 were screened for EMS induced mutations. Exon 1 of RS2 was screened using labeled primers:

5'-GAGTCTCATATTGTACATGGTAG-3' and 5'-

GCAATTCGATGCTTCTTATGAG-3'. Exons 3 and 4 of RS3 were screened using a

single primer pair: 5'-GGTCGGCACGTTATTTCTTTTT-3' and 5'-

CAAGTGCTCATATATAAGGTTATC-3'. Standard TILLING PCR parameters were as

follows. One cycle of 95 °C for 2 minutes and 94 °C for 20 seconds followed by 56 cycles of 94 °C for 20 seconds, 56 °C for 30 seconds, and 72 °C for 1 minute. The next

step in the PCR was 72 °C for 5 minutes, followed by 99 °C for 10 minutes followed by a 70 °C to 0 °C melt. The PCR reaction was then held at 10 °C.

The PCR products were then run on a TILLING gel and visualized using both the IR700 and IR800 channels (Colbert et al., 2001; McCallum et al., 2000a, McCallum et al., 2000b). Lanes where bands could be visualized indicating an induced mutation or heteroduplex mismatch were deconvoluted by separating the pools into individual plant DNA samples for sequencing in order to identify the line containing the mutation (Colbert et al., 2001; McCallum et al., 2000a; McCallum et al., 2000b). The location of the mutation as well as the zygoty could then be verified. Screening of the EMS mutagenized population was carried out at Purdue University in West Lafayette, IN by Cliff Weil and Rita Monde.

Fast Neutron mutagenized population

This population was screened for deletion mutations as described (Li et al., 2001). Primers for screening RS2 were: 5'-CACTTAATGAACCTAGACAGATAC-3' and 5'-GGTGGCACCAGTCCAACCTCC-3'; 5'-GTACATGGTAGTTTGACTTTGACAC-3' and 5'-CTTGGACAAAGAATACGCATGGG-3'. Primers for screening RS3: 5'-GCAGTAAGACGTACCTTTCTCTTG-3' and 5'-CCAAACTCCTCCTTCAATTCTCTCAC-3'; CAATCCTAACACGTGGACAGAATC-3' and 5'-GACAATATCACTGATCATTGGAGG-3'.

Reactions were carried out in 20 µl; each primer was at 0.5 µM final concentration in reactions containing template, buffer (40 mM Tricine-KOH (pH 8.0), 16

mM KCl, 3.5 mM MgCl₂, 3.75 µg ml⁻¹ BSA, 200 µM dNTPs), 5% DMSO, and 0.2X Titanium *Taq* polymerase (BD Biosciences, Palo Alto, CA). The PCR parameters were one cycle of 95 °C for 5 minutes followed by 35 cycles of 95 °C for 20 seconds, 60 °C for 20 seconds, and 72 °C for 3 minutes. Following the PCR, 0.25X SYBR Green I was added to each reaction which was subject to a melt from 75 °C to 90 °C.

Allele-Specific Molecular Marker Assay Development

An allele specific molecular marker assay was developed for the mutation identified in line 397 of the EMS mutagenized population. This assay was utilized to discriminate between wild-type Williams 82 or mutant alleles of the RS2 gene. The assay was designed as described (Wang et al., 2005). In order to achieve allele specificity, single base pair mismatches were introduced into the primer sequence to increase the discriminatory power of the allele-specific primer. These bases and the tails are indicated in lowercase in the primer sequences. Primer sequences were: 5'-gcgggcGTTGCTACCGACCCAGtGAA -3', 5'-gcgggcagggcggcGTTGCTACCGACCCAGcGAG -3', and a common forward primer 5'-CAGAGGAATAAAATTCATGAGCATA -3'.

Reactions were carried out in 20 µl; each primer was at 0.5 µM final concentration in reactions containing template, buffer (40 mM Tricine-KOH (pH 8.0), 16 mM KCl, 3.5 mM MgCl₂, 3.75 µg ml⁻¹ BSA, 200 µM dNTPs), 5% DMSO, 0.25X SYBR Green I, and 0.2X Titanium *Taq* polymerase (BD Biosciences, Palo Alto, CA).

PCR parameters for the RS2 assay were as follows: 95 °C for 12 minutes followed by 35 cycles of 95 °C for 20 seconds, 60 °C for 30 seconds, 72 °C for 30 seconds, and

then a melting curve from 72 °C to 90 °C. The fluorescence was read after each cycle and every 0.2 °C with a one second hold during the melt. Each genotype produced a product with a characteristic melting profile, as measured by T_m of the negative first derivative of the disappearance of fluorescent signal. The Williams 82 homozygous genotype gave a peak at 83.6 °C, mutant homozygous genotype gave a peak at 79.2 °C, and a heterozygous genotype gave a peak at 83.6 °C with a shoulder at 79.0 °C.

Oligosaccharide Phenotype Determination by HPLC

Oligosaccharides were determined by high performance ion chromatography with pulsed amperometric detection (PAD) employing an Agilent 1100 series HPLC and an ESA Coulochem III detector (Agilent Technologies, Chesterfield, MO, USA). A 12.5 mg ground seed sample from either a whole seed or a chipped seed (seed were cut with a razor blade so that approximately $\frac{1}{2}$ could be used for oligosaccharide extract and the remaining half could be germinated) was extracted with 0.5 ml 50% ethanol at 70 °C, 30 min. Samples were then centrifuged 15 min at 16,000g. The supernatant was passed through a 0.2 μm filter. Sugars were separated on a Dionex Carbo Pac PA 10 analytical column (250 mm x 4 mm, 10 μm) connected to a Carbo Pac PA 10 guard column (50 mm x 4 mm). The mobile phase was 90 mM NaOH with flow rate of 1.5 ml min^{-1} , maintained at 30 °C. Detection settings were: time 0, 0.1 v, time 0.41, -2.0 v, time 0.42, 0.6 v, and time 0.44, -0.1 v.

Plant Growth Conditions

Seven homozygous wild-type and nine homozygous mutant RS2 M_{2:3} plants from the mutagenized line 397 were grown to maturity in a growth chamber with 13 hour day length. The dark temperature was 22 °C and the light temperature was 28 °C. Plants were grown, three per 3-gallon pot, in PRO-MIX (Premier Horticulture) medium and fertilized with Osmocote Plus (Scotts) per manufacturer's instructions.

RESULTS

Screening of EMS mutagenized population

Reverse genetics screening the EMS mutagenized population created the potential to find additional soybean raffinose synthase mutants for further characterization of a unique seed oligosaccharide phenotype such as in PI 200508. Four lines were identified which contained single nucleotide polymorphisms in the soybean raffinose synthase gene RS2. Two of the base pair mutations did not result in amino acid changes and therefore were not considered candidates for phenotypic mutants. The other two lines, 165 and 397, contained base pair mutations which resulted in amino acid changes (Figure 5.1).

Line 165 resulted in a homozygous C to T transition and a S150F amino acid change. This mutation lies in a semi-conserved region of plant raffinose synthase genes (Figure 5.1). Seeds from the homozygous S150F line 165 did not have an obvious oligosaccharide phenotype as determined by quantitatively measuring sucrose, raffinose and stachyose of $M_{2:3}$ seeds and comparing them to wild-type seeds (data not shown).

The mutagenized M_2 line 397 was characterized by a heterozygous C to T transition which resulted in a T107I amino acid change. This induced mutation lies in a semi-conserved region of plant raffinose synthase gene sequences (Figure 5.1). To investigate the inheritance of this novel allele and its subsequent effect on seed oligosaccharide content, thirty-seven individual $M_{2:3}$ seeds were chipped and analyzed by HPLC (Figure 5.2). The remaining portion of the seed was germinated and genotyped by the developed allele specific molecular marker assay. The genotype/phenotype association results on $M_{2:3}$ seeds reveal alterations in oligosaccharide content (Figure 5.2). Furthermore, one wild-type allele of RS2 appears sufficient to produce the wild-

type oligosaccharide seed phenotype, which is consistent with previous results (Figure 5.2) (Dierking and Bilyeu, 2008).

An $M_{2:3}$ population consisting of contrasting RS2 genotypes was then developed from the plants that contained either homozygous wild-type (Williams 82) or homozygous mutant alleles at the RS2 locus in order to further characterize the phenotype resulting from the novel allele. Seven independent wild-type RS2 and nine independent mutant RS2 lines were selected to negate the action of unidentified genes that may contribute to the oligosaccharide content. Four seeds from each of the plants of the homozygous population were analyzed by HPLC for oligosaccharide content. For the plants that contained the T107I mutation, the average $M_{3:4}$ seed sucrose was increased by 28%, raffinose was reduced by 37% and stachyose was reduced by approximately 24% as compared to 397 lines which carried the wild-type allele of RS2 (Figures 5.3). This phenotype is similar to the previously described RS2 mutations in PI 200508 (Dierking and Bilyeu, 2008).

In all cases where RS2 is homozygous mutant a statistically significant difference in sucrose:raffinose + stachyose ratio is observed as compared to either RS2 homozygous wild-type alleles or a heterozygous state. However, some range in ratio is observed in the homozygous mutant classes: for the $M_{2:3}$ seed the ratio was approximately 3.0, while for the $M_{3:4}$ seeds the ratio was approximately 1.3. The difference in oligosaccharide ratio could be due to the production of these seed in different environments: the $M_{2:3}$ seeds were produced in a field environment, while the $M_{3:4}$ seeds were produced in the growth chamber. The growth chamber settings in this experiment were also different than those previously used (Chapter 2). Furthermore, although chipped seeds used in HPLC

analysis for oligosaccharide content gave indistinguishable results from whole seeds; it is notable that the M_{2:3} seeds were chipped, while the M_{3:4} seeds that were analyzed were whole seeds.

There were no induced mutations identified in the candidate soybean raffinose synthase gene RS3.

Screening of Fast Neutron mutagenized population

Screening the fast neutron mutagenized population presented the unique opportunity to identify a deletion mutation in either RS2 or RS3. For either gene this would represent a novel mutation and potentially a novel phenotype because a knock-out of a soybean raffinose synthase gene sequence has not been previously identified. Deletion mutations are inherently difficult in terms of screening using the PCR method described by Li et al. (2001). Essentially the primers are placed far enough apart on the target sequence that PCR is unsuccessful unless a deletion has occurred; the pooled DNA would be deconvoluted to identify the individual DNA displaying the deletion, sequencing of the amplicon would follow to characterize the exact location of the deletion. However, if a primer is designed for a deleted region a false negative occurs making it difficult to discriminate from the absence of a deletion. To combat understanding the two negative results, multiple primer pairs which spanned the target region were utilized. No deletion mutations were identified in the fast neutron mutagenized population for either of the two genes, RS2 or RS3.

DISCUSSION

The identification of an EMS induced mutation in RS2 serves as an additional confirmation of the contribution of this gene to the seed raffinose and stachyose content (Dierking and Bilyeu, 2008). Similar to the PI 200508 allele of RS2, the mutagenized line 397 has reduced raffinose and stachyose as well as an increase in seed sucrose content. It appears that inheritance of a single wild-type allele of RS2 is sufficient for a wild-type oligosaccharide seed phenotype which is consistent with previously characterized RS2 alleles (Dierking and Bilyeu, 2008).

The discovery of a second, novel mutation in the soybean raffinose synthase gene RS2 serves as an independent confirmation of the involvement of RS2 in soybean seed oligosaccharide phenotypes. Furthermore, since similar phenotypes resulted from two independent mutations in RS2, it may indicate that the mutations in the gene are deleterious to enzyme function and the detected raffinose in these lines is the result of an additional soybean raffinose synthase gene.

Alternatively, the discovery of an EMS induced mutation in the RS2 gene demonstrates that a chemically induced mutation, identified by the reverse genetics approach TILLING, can be associated with a phenotype in soybean. The developed molecular marker assay will allow the efficient incorporation of this novel allele into elite germplasm. Clearly, the identification of mutations in genes has wide applications and now TILLING has been demonstrated as a very powerful genetics tool for soybean.

FIGURES

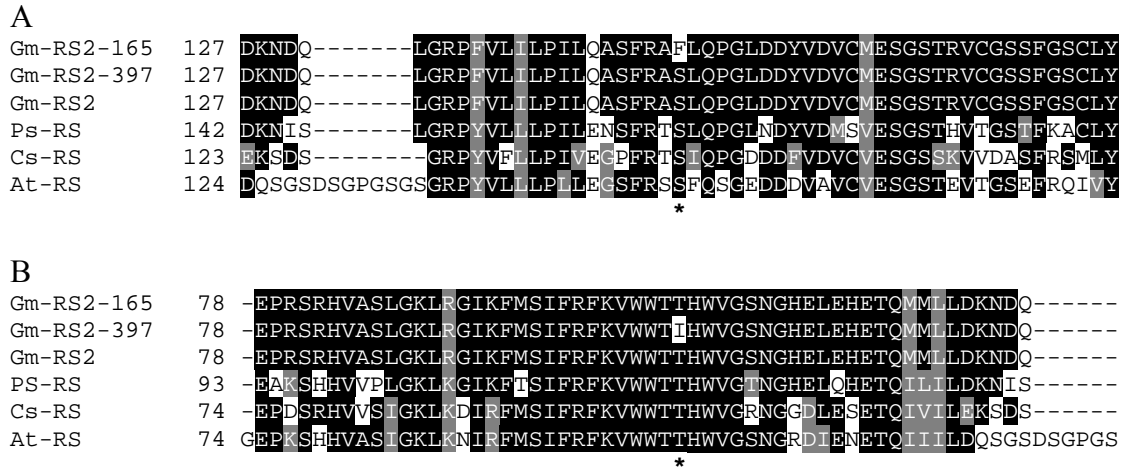


Figure 5.1: Raffinose synthase amino acid sequence alignments in the regions surrounding the induced mutations in the RS2 gene. Amino acid positions are indicated at the beginning of each alignment. The position of the polymorphic amino acid is indicated by an asterisk. Identical amino acid residues are highlighted in black while similar amino acid residues are highlighted in gray. A. The exon one region containing the induced mutation in line 165 which resulted in S150F. B. The exon one region containing the induced mutation in line 397 which resulted in T107L.

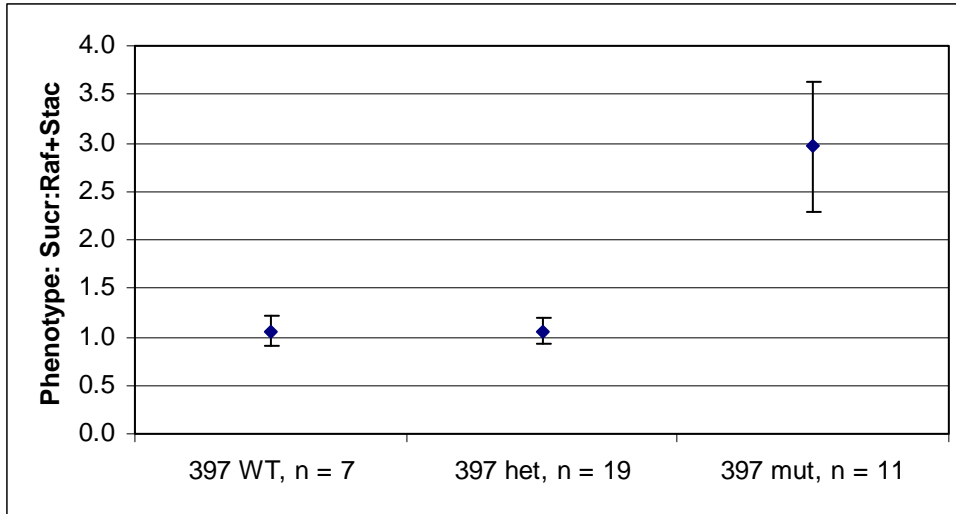


Figure 5.2: Phenotype to genotype association of segregating $M_{2:3}$ seeds from line 397. The oligosaccharide phenotype of 37 individual $M_{2:3}$ seeds was determined. The data represents the mean of the ratio of extractable seed sucrose to the sum of raffinose and stachyose. Error bars represent plus and minus one standard deviation from the mean.

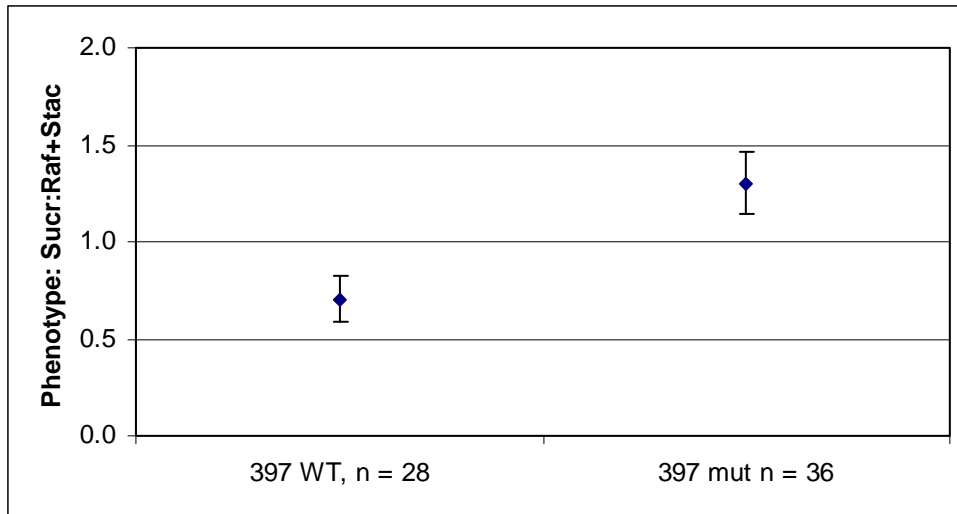


Figure 5.3: Phenotype to genotype association of a homozygous soybean population derived from the mutagenized soybean line 397. The oligosaccharide phenotype of four individual $M_{3:4}$ seeds from each plant was determined. The data represents the mean of the ratio of extractable seed sucrose to the sum of raffinose and stachyose. Error bars represent plus and minus one standard deviation from the mean. Individual $M_{2:3}$ plants were categorized by their RS2 genotype class and separated on the x-axis.

LITERATURE CITED

- Bernard, R.L., and C.R. Cremeens. 1988. Registration of 'Williams 82' soybean. *Crop Science* 28:1027-1028.
- Cho, H.-J., S.K. Farrand, G.R. Noel, and J.M. Widholm. 2000. High-efficiency induction of soybean hairy roots and propagation of the soybean cyst nematode. *Planta* 210:195-204.
- Colbert, T., B.J. Till, R. Tompa, S. Reynolds, M.N. Steine, A.T. Yeung, C.M. McCallum, L. Comai, and S. Henikoff. 2001. High-throughput screening for induced point mutations. *Plant Physiology* 126:480-484.
- Cooper, J., B. Till, R. Laport, M. Darlow, J. Kleffner, A. Jamai, T. El-Mellouki, S. Liu, R. Ritchie, N. Nielsen, K. Bilyeu, K. Meksem, L. Comai, and S. Henikoff. 2008. TILLING to detect induced mutations in soybean. *BMC Plant Biology* 8:9.
- Dierking, E.C., and K.D. Bilyeu. 2008. Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. *The Plant Genome* 1:135-145.
- Flores, T., O. Karpova, X. Su, P. Zeng, K. Bilyeu, D. Sleper, H. Nguyen, and Z. Zhang. 2008. Silencing of Gm FAD3 gene by siRNA leads to low α -linolenic acids (18:3) of fad3 -mutant phenotype in soybean [*Glycine max* (Merr.)]. *Transgenic Research* 17:839-850.
- Hadi, M.Z., M.D. McMullen, and J.J. Finer. 1996. Transformation of 12 different plasmids into soybean via particle bombardment. *Plant Cell Reports* 15:500-505.
- Li, X.P., R. Gan, P.L. Li, Y.Y. Ma, L.W. Zhang, R. Zhang, Y. Wang, and N. Wang. 2006. Identification and functional characterization of a leucine-rich repeat receptor-like kinase gene that is involved in regulation of soybean leaf senescence. *Plant Molecular Biology* 61:829-844.
- Li, X., Y. Song, K.S. Century, Shelly, P. Ronald, X. Dong, M. Lassner, and Y. Zhang. 2001. A fast neutron deletion mutagenesis-based reverse genetics system for plants. *The Plant Journal* 27:235-242.
- McCallum, C.M., L. Comai, E.A. Greene, and S. Henikoff. 2000. Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant Physiology* 123:439-442.
- McCallum, C.M., L. Comai, E. A. Greene, and S. Henikoff. 2000. Targeted screening for induced mutations. *Nature Biotechnology* 18:455-457.

- Olhoft, P., C. Donovan, and D. Somers. 1996. Soybean (*Glycine max*) transformation using immature cotyledonary node explants, p. 385-396 *Agrobacterium Protocols*, Vol. 343 2nd edition. Humana Press, Totowa, NJ.
- Somers, D.A., D.A. Samac, and P.M. Olhoft. 2003. Recent advances in legume transformation. *Plant Physiology* 131:892-899.
- Subramanian, S., M.Y. Graham, O. Yu, and T.L. Graham. 2005. RNA interference of soybean isoflavone synthase genes leads to silencing in tissues distal to the transformation site and to enhanced susceptibility to *Phytophthora sojae*. *Plant Physiology* 137:1345-1353.
- Wang, J., K. Chuang, M. Ahluwalia, S. Patel, N. Umblas, D. Mirel, R. Higuchi, and S. Germer. 2005. High-throughput SNP genotyping by single-tube PCR with Tm-shift primers. *Biotechniques* 39:885-93.

APPENDIX 1

Recombinant Protein Expression

Recombinant expression of soybean candidate raffinose synthase genes represented the opportunity to biochemically characterize the genes. There are a number of questions that could be answered: first, does the candidate protein catalyze the expected reaction, second, does the candidate protein have the ability to catalyze both the raffinose synthase and stachyose synthase enzymatic reactions, and finally, what is the biochemical difference between the alleles (Williams 82 or PI 200508) of each of the genes.

Numerous unsuccessful attempts to isolate full length coding sequence of the candidate genes RS2 and RS3 were made. It is unclear whether the problem lies in isolating high quality, undegraded RNA or if the reverse transcriptase step was itself unsuccessful. Both RS2 and RS3 were designed with specific restriction sites at the 5' and 3' ends for expression in either *E. coli* or *Pichia pastoris* and ordered from Celtek Genes (Figures A.1 and A.2; <http://www.celtek-genes.com/>). The restriction sites specifically designed at the 5' and 3' ends of the gene were removed from the internal gene sequences by changing single base pairs which did not cause changes in the amino acid sequence. In addition, restriction sites, either naturally occurring or inserted, were specifically designed in the internal gene sequence for the purpose of switching from the Williams 82 to the PI 200508 allele of the gene.

There is limited information available in the literature about plant raffinose synthase biochemical activity. The catalytic domains are unknown as well as whether or not the activity is enhanced by any cofactors. Protein expression has been assayed for crude lysates from developing soybean seed (Hitz et al., 2002); pea raffinose synthase has been recombinantly expressed in insect cells (Peterbauer et al., 2002); adzuki bean

stachyose synthase has been purified and characterized (Peterbauer and Richter, 1998) and the rice raffinose synthase recombinantly expressed using *E. coli* (Li et al., 2007).

Although many constructs, strains, growing conditions and other methods were employed, our challenge was in obtaining soluble, active protein. Only in a few cases were extremely small amounts of soluble protein were detected by SDS-PAGE; when tested using a biochemical activity assay, the recombinant protein was determined to be inactive. In each expression or biochemical activity assay, a empty vector control was used as well as a zero timepoint control.

*All restriction enzymes were obtained from New England Biolabs (NEB) and used as per manufacturers instructions.

Recombinant expression in *E. coli*:

Expression in *E. coli* was the first choice because of its' inherent simplicity and quick growth period (Krogdram et al., 2003). Expression in *E. coli* often results in very high yields of up to 50% of the total cell protein (Krogdram et al., 2003). The expression of proteins in *E. coli* is most often intracellular; however, some systems exist which can localize recombinant proteins to the periplasm or secrete the proteins to the surrounding medium (Krogdram et al., 2003). The expression systems used here are designed for intracellular expression and are HIS-tagged for affinity purification. Obvious drawbacks are that *E. coli* is a prokaryote while the protein to be expressed is from a eukaryote. *E. coli* is unable to provide any splicing or post-translational modifications which may be necessary to produce an active eukaryotic protein. It is unknown if these eukaryote-

specific post-translational modifications are required for soybean raffinose synthase enzyme activity.

Constructs: The pET expression system (Novagen Cat. No. 69660-3) was utilized for most of the recombinant expression experiments in *E. coli*. This system includes hybrid promoters, multiple cloning sites, and protease cleavage sites (Sørensen and Mortensen, 2005). The specific expression vector pET-14b was used for three of the constructs tested in *E. coli*. The pET-14b vector map and polylinker can be seen in Figure A.3. Initially, expression was attempted by directional cloning of the candidate RS2 sequence by digestion with NdeI and XhoI, and the construct was sequence verified (N-terminus: 5'-GGAGACCACAACGGGTTTCC-3', C-terminus: 5'-GCCACGTGTCGGCTTCTGG-3'). Two other modifications of this construct were also attempted. The thrombin site was removed and the entire N-terminal tag removed by Pfu-based PCR (primers for removal of thrombin site: 5'-CATCATCACAGCAGCGGCATGGCTCCAAGCATAAGC-3' and 5'-GCTTATGCTTGGAGCCATGCCGCTGCTGTGATGATG-3'; for N-terminal tag removal: 5'-CTTTAAGAAGGAGATATACCATGGCTCCAAGCATAAGC-3' and 5'-GCTTATGCTTGGAGCCATGGTATATGTCCTTCTTAAAG-3'). Each of these new constructs was verified by sequencing with the same N-terminal and C-terminal primers as stated above. The objective of each modified construct was to remove sequence that could potentially form a secondary structure which could interfere with initiation of translation of the target protein.

Low levels of expression could be detected by SDS-PAGE with the second of the modified pET-14b constructs; however, the protein produced was insoluble. To combat this problem, the coding sequence of the candidate gene RS2 was ligated into the pMAL-c4x vector (New England Biolabs, Cat No. N8107S; Figure A.4) using EcoRI and BamHI and verified by sequencing (N-terminus: 5'-GTACTGCGGTGATCAACGCC-3'). This vector fuses the candidate gene sequence downstream of the *malE* gene from *E. coli* (NEB, pMAL Protein Fusion and Purification, 2007). The *malE* gene encodes the maltose binding protein (MBP) which is expressed cytoplasmically and can be cleaved from the target protein by Factor Xa (NEB Cat. No. P8010S). MBP is characteristically very stable and thus may help to retain the solubility of the target protein. Expression experiments were carried out as described by the manufacturer. Recombinant protein was expressed at low levels and was insoluble.

One additional construct has been prepared which includes the PI 200508 allele of RS2 in the pET-14b expression vector (Novagen Cat. No. 69660-3). The switch from the Williams 82 allele of RS2 to the PI 200508 allele occurred in the Celtek Genes vector. The Celtek Genes vector and a TOPO TA Cloning vector (Invitrogen Cat. No. 45-0641) containing exon two of PI 200508 allele (obtained by PCR and cloning per manufacturers instructions; primers: 5'-CACAATATTCACCTAATCTTAATGAGA-3' and 5'-GCTGAAGCCTATAAACACCATA-3') of RS2 was digested using MluI and SphI. The entire gene was then excised using NdeI and XhoI and ligated into the pET-14b vector. This construct has not been tested for expression.

Strains and induction: Three different *E. coli* strains were utilized for recombinant expression: BL21(DE3) (Novagen Cat. No. 69449-3), the standard

expression strain, BL21-CodonPlus(DE3)-RIL (Stratagene Cat. No. 230280), and Rosetta-gami (Novagen Cat. No 71045-3). BL21 is the most widely used *E. coli* strain for pET expression recombinants (Novagen TB009, 2004). DE3 refers to the host being a lysogen from λ DE3; thus includes a copy of the T7 RNA polymerase which is compatible with pET recombinants and induction with IPTG. BL21-CodonPlus(DE3)-RIL is a strain which is optimized for the use of codons which are rarely use in *E. coli*, but may be more common in eukaryotic genes. Rosetta-gami host-strains are derived from Origami and combine enhanced di-sulfide bond formation with enhanced expression of proteins that contain codons rarely used by *E. coli* (Novagen TB009, 2004). All three strains were tested for the original (unmodified) pET-14b construct containing RS2. Both the modified pET-14b constructs were tested using BL21(DE3) and BL21-CodonPlus(DE3)-RIL. The pMAL-c4x construct was only tested with BL21(DE3).

Levels of IPTG used for induction were experimentally determined for each of the constructs tested. Initially 50 μ M and 500 μ M were tested by plotting a growth curve. A marked difference in the level of induction and production of protein could not be observed. Therefore, 50 μ M was used for all constructs, except the pMAL-c4x construct where 300 μ M was used per manufacturers instruction.

Growth conditions: Four distinct growing conditions were utilized. A summary of the growth conditions can be seen in Table A.1. Initial growth of an individual construct was always tested at what is considered to be “normal conditions”. Normal conditions are 37 °C and 200 RPM for sufficient aeration: a 1 to 8, media to flask volume ratio (v,v). Low temperature refers to room temperature or 25 °C. Low aeration has two components; a slower rate on the orbital shaker of 125 RPM and a 1 to 2, media to flask

volume ratio (v,v). The lower temperature and lower aeration were both intended to slow the *E. coli* growth which in turn would produce protein more slowly in effort to retain solubility.

Harvest: Both sonication and French Press were used for harvest of the recombinant protein expressed in the pET-14b vectors. A distinct difference in protein yield or solubility between the two methods was not apparent. However, French Press is thought to be better for proteins high in Cys, because oxidation is less of a threat than with sonication. Therefore, only French Press was used with the pMAL-c4x construct. The protein was kept under reducing conditions during each of these procedures by using DTT at a concentration ranging from 2 mM to 20 mM.

Solubility and purification: If expression could be detected, the recombinant protein was in the insoluble fraction. Both common methods of regaining solubility were attempted: Guinidinumn-Hydrochloric Acid (3 M) and Urea (6 M). Neither was successful. Two methods of purification were employed, HIS-mag magnetic beads (PopCulture His-Mag Purification Kit, Novagen Cat. No. 71114-3REF) for the original pET-14b construct containing RS2, and a Ni²⁺ column (HisTrap HP, GE Healthcare Cat. No. 17-5247) was used for the modified pET-14b construct which had the entire N-terminal tag removed. In most cases, recombinant RS2 protein expression levels were low to non-detectible and purification was not attempted.

NEB PURExpress:

NEB PURExpress (NEB Cat. No. E6800S) is a cell-free expression system that utilizes components isolated and purified from *E. coli*. The advantages of this system are

that it is free of exonucleases, RNases and proteases, all of which would interfere with recombinant expression, and it has the potential to produce protein within one hour. This expression kit was used with the RS2 gene sequence as template as described by the manufacturer, however, recombinant protein could not be visualized by SDS-PAGE.

Recombinant expression in *Pichia pastoris*:

The yeast expression host, *Pichia pastoris*, was utilized as a eukaryotic system for recombinant protein expression. Eukaryotic systems offer the ability to post-translationally modify target proteins which may be required for their enzymatic function. The candidate genes RS2 and RS3 were inserted into the expression vector pPIC α -A (Figure A.5) using directional cloning with EcoRI at the 5' end and XbaI at the 3'. The constructs were verified by sequencing (N-terminus: 5'-CTACTATTGCCAGCATTGCTGC-3' and C-terminus: 5'-ATGATGATGGTCGACGGCGC-3') and molecular marker assay for presence of Williams 82, RS2 sequence (Dierking and Bilyeu, 2008). Only the construct containing RS2 was tested for expression.

The expression vector pPIC α -A is characterized by the α secretion factor which targets the fused recombinant protein for transport into the medium. This is useful since only a small number of native *Pichia* proteins are secreted; the resulting recombinant protein should be relatively pure. Expression in *Pichia pastoris* was carried out as described by the manufacturer (Invitrogen) using X-33 cells and collected after one day or six days of feeding with methanol. The recombinant protein was concentrated in 55

mM HEPES pH 7.0, 5.5mM KCl by a factor of 200 prior to conducting biochemical activity assays.

Stable transformation of *Arabidopsis thaliana*:

Two constructs were designed in anticipation of stable transformation of *Arabidopsis*. The Williams 82 allele of RS2 and the Williams 82 allele of RS3 were ligated into the entry vector pENTR-3C (Invitrogen Cat. No. A10464; Figure A.6) using EcoRI and XbaI. The constructs were verified by molecular marker assay (Dierking and Bilyeu, 2008), prepared for the clonase reaction (Gateway LR Clonase II Enzyme Mix, Invitrogen Cat. No. 11791-020) and are appropriate for delivery into a destination vectors using Gateway technology from Invitrogen, which has an N-terminal tag (CD3-696, pN-TAPa). The destination vector was obtained from The Arabidopsis Information Resource (www.arabidopsis.org).

Biochemical assay for enzymatic activity:

Positive control assay: A positive control assay for developing soybean seeds was derived from Hitz et al. (2002). This raffinose synthase assay was carried out exactly as described, except the supernatant was not dried, but was extracted with 50% ethanol, filtered, and applied directly to the HPLC column as described (Dierking and Bilyeu, 2008). The products raffinose and stachyose could be reliably detected using this assay (Figure A.7). It is noteworthy that a compound which a retention time similar to stachyose appeared in the crude lysate prior to access to substrate indicating that some stachyose has already accumulated in developing soybean seeds.

Recombinant enzyme activity assay: The activity of recombinantly expressed raffinose synthase was assessed by a biochemical activity assay (Hitz et al., 2002; Peterbauer and Richter, 1998). Protein was either dialyzed or concentrated with 55 mM HEPES pH 7.0, 5.5 mM KCl protein buffer and represented 90% of the reaction mix. A substrate mix of 200 mM galactinol and 500 mM sucrose (raffinose synthase activity) or raffinose (stachyose synthase activity) was used as 10% of the reaction mix. Six 100 μ l reactions were prepared and incubated at 30 °C for 0 minutes, 30 minutes, 1 hour, 2 hours, 8 hours, or 16 hours. Four-hundred microliters of water was then added to each reaction, centrifuged to clear, the supernatant was then analyzed directly by HPLC (Dierking and Bilyeu, 2008). Neither raffinose nor stachyose could be detected following the biochemical activity assay (Figure A.8) indicating that either the recombinant protein was inactive (folding incorrectly) or that no recombinant protein was expressed.

FIGURES

ctttcaagCATATGatccttgaattcagtc^{ccc}ATGGCTCCAAGCATAAGCAAAACTGTG
GAACTAAATTCATTTGGTCTTGTCAACGGTAATTTGCCTTTGTCCATAACCCTAGAAGG
ATCAAATTTCTCGCCAACGGCCACCCTTTTCTCACGGAAGTTCCCGAAAACATAATAG
TCACCCCTTCACCCATCGACGCCAAGAGTAGTAAGAACAACGAGGACGACGACGTGCGTA
GGTTGCTTCGTGGGCTTCCACGCGGACGAGCCCAGAAGCCGACACGTGGCTTCCCTGGG
GAAGCTCAGAGGAATAAAATTCATGAGCATATTC^{CGGTTTAAGGTGTGGTGGACCACTC}
ACTGGGTCGGTAGCAACGGACACGAACTGGAGCACGAGACACAGATGATGCTTCTCGAC
AAAAACGACCAGCTCGGACGCCCTTTGTGTTGATTCTCCCGATCCTCCAAGCCTCGTT
CCGAGCCTCCCTGCAACCCGGTTTGGATGATTACGTGGACGTTTGCATGGAGAGCGGGT
CGACACGTGTCTGTGGCTCCAGCTTCGGGAGCTGCTTATACGTCCACGTTGGCCATGAC
CCGTATCAGTTGCTTAGAGAAGCAACTAAAGTCGTTAGGATGCATTTGGGGACGTTCAA
GCTTCT^gGAGGAGAAAACCGCGCCAGTGATCATAGACAAGTTTGGTTGGTGTACATGGG
ACGCGTTTTACTTGAAGGTGCATCCCTCAGGTGTGTGGGAAGGGGTGAAAGGGTTGGTG
GAGGGAGGGTGCCCTCCAGGGATGGTCTAATCGACGACGGGTGGCAAGCCATTTGTCA
CGACGAGGACCCATAACGGACCAAGAGGGTATGAAGCGAACCTCCGCAGGGGAGCAAA
TGCCATGCAGGTTGGTGAAGTTGGAGGAAAATTACAAGTTCAGACAGTATTGTAGTGGA
AAGGATTCTGAGAAGGGTATGGGTGCCTTTGTTAGGGACTTGAAGGAACAGTTTAGGAG
CGTGGAGCAGGTGTATGTGTGGCACGCGCTTTGTGGGTATTGGGGTGGGGTCCAGACCCA
AGGTTCCGGGCATGCCCCAGGCTAAGGTTGTCACTCCGAAGCTGTCCAATGGACTAAAA
TTGACAATGAAGGATTTAGCGGTGGATAAGATCGTCAGTAACGGAGTTGGACTGGTGCC
ACCACACCTGGCTCACCTTTTGTACGAAGGGCTCCACTCCCGTTTGGAAATCTGCGGGTA
TTGACGGTGTTAAGGTTGACGTTATACACTTGCT^gGAGATGCTATCCGAGGAATACGGT
GGCCGTGTTGAGCTAGCCAAAGCTTATTACAAAGCGCTCACTGCTTCGGTGAAGAAGCA
TTTCAAAGGCAATGGGGTCATTGCGAGCATGGAGCATTGTAATGACTTCTTTCTCCTTG
GTACCGAAGCCATAGCCCTTGGGCGCGTAGGAGATGATTTTTGGTGCCTGATCCCTCT
GGAGATCCAAATGGCACGTATTGGCTCCAAGGGTGTACATGGTGCCTGTGCCTACAA
CAGCTTGTGGATGGGGAATTTTATTACGCCGATTGGGACATGTTCCAGTCCACTCACC
CTTGTGCCGAgTTCCATGCAGCCTCTAGGGCCATCTCTGGTGGACCAGTTTACGTTAGT
GATTGTGTTGGAAAGCACAACTTCAAGTTGCTCAAGAGCCTCGCTTTGCCTGATGGGAC
GATTTTGCCTTGTCAACACTATGCACTCCCCACACGAGACTGTTTGTGTTGAAGACCCCT
TGCATGATGGGAAGACAATGCTCAAAATTTGGAAATCTCAACAAATATACAGGTGTTTTG
GGTCTATTTAATTGCCAAGGAGGTGGGTGGTGTCCCGTAACTAGGAGAAACAAGAGTGC

CTCTGAATTTTCACAAACTGTGACATGCTTAGCGAGTCCTCAAGACATTGAATGGAGCA
ATGGGAAAAGCCCAATATGCATAAAAGGGATGAATGTGTTTGCTGTATATTTGTTCAAG
GACCACAAACTAAAGCTCATGAAGGCATCAGAGAAATTGGAAGTTTCACTTGAGCCATT
TACTTTTGAGCTATTGACAGTGTCTCCAGTGATTGTGCTGTCAAAAAAGTTAATTCAAT
TTGCTCCAATTGGATTAGTGAACATGCTTAACACTGGTGGTGCCATTCAGTCgATGGAG
TTTGACAACCACATAGATGTGGTCAAAATTGGGGTTAGGGGTTGTGGGGAGATGAAGGT
GTTTGCATCAGAGAAACCAGTTAGTTGCAAACCTAGATGGGGTAGTTGTAAAATTTGATT
ATGAGGATAAAATGCTGAGAGTGCAAGTTCCTGGCCTAGTGCTTCAAATTTGTCAATG
GTTGAGTTTTTATTTTGAccccCTCGAGttttTCTAGActct

Figure A.1A: Gene sequence for RS2 including restriction sites for directional cloning. Coding sequence is represented by uppercase letters. Non-coding sequence or bases changed within the coding sequence to remove restriction sites are represented by lowercase letters. Restriction sites of interest are underlined. The 5' restriction sites are NdeI, EcoRI, and NcoI. The 3' restriction sites are XhoI and XbaI. Internally, MluI and SphI are available for switching from the Williams 82 allele to the PI 200508 allele.

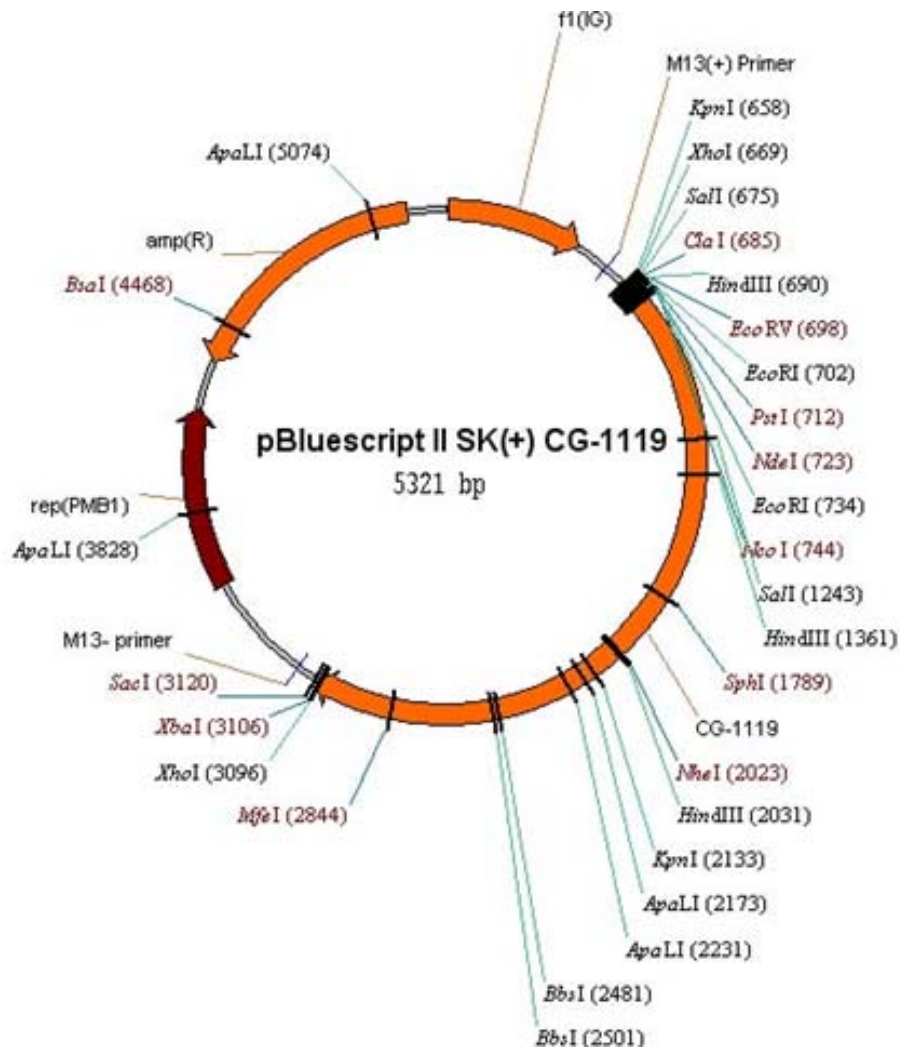


Figure A.1B: Celtek Genes order of RS2 in pBluescript II SK(+) vector. Vector map provided by manufacturer (Celtek Genes: www.celtekgenes.com). CG-1119 refers to RS2. RS2 was ordered with a 5' NdeI, EcoRI, and NcoI restriction sites, at the 3' end are XhoI and XbaI. These restriction sites were removed from the internal gene sequence by changing a base pair which did not result in an amino acid change. Restriction sites MluI and SphI were introduced in order to switch the Williams 82 allele or RS2 to the PI 200508 allele.

ctttcaagCATATGatccttgaattcagtc~~ccc~~ATGGGTCCAAGCTCGAAGAAAGCTTCA
CCTAAATCAGGTGTGACAAAGCACATGAAGGGCTTCAGCCTCTGCAACTCAACCCTAAA
AGTAAATGGGCAAGTCATCCTCTCCAAGTCCCCAAGAACGTAACCCTCACCCCATGCA
CCTACGACACTCACACCACCGGATGCTTCCTCGGTTTCCACGCCACCTCCCCAAAATCC
CGCCACGTGGCACCCCTTAGGACAGCTTAAAAACATAAGCTTCACTTCCATCTTCCGGTT
CAAGGTTTGGTGGACCACTCTCTGGACCGGCTCCAACGGCCGCGACCTGGAAACCGAAA
CCCAATTCCTCATGCTCCAATCCCACCCTTATGTTCTCTTCTTACCCTCCTCCAACCC
CCATTTTCGCGCCTCGCTGCAGCCTCACTCAGACGACAACGTTGCGGTGTGTGTGGAGAG
CGGCTCCAGCCACGTAACAGCCTCATCATTCGACACTGTCGTCTACTTGCACGCAGGGG
ACAACCCCTTCACGCTGGTCAAGGAAGCCATGCGCGTCGTCCGGGCCCACTTGGGGAGC
TTCAAGCTTCTGGAAGAGAAAACCGTTCCGGGAATGGTGGACAAGTTCGGTTGGTGCAC
GTGGGACGCCTTTTACCTGACGGTGCACCCTGAGGGCGTCAGAGAGGGCGTGAAGGGCC
TGGTTGACGGCGGTTGTCCTCCGGGATTCGTCTGATCGACGACGGCTGGCAGTGCATC
AGCCACGATTCCGATCCGGAGAAGGAGGGGATGAATCAGACGGTGGCCGGGGAGCAAAT
GCCCTGCAGGTTGATTAGTTACGAGGAGAATTACAAGTTTAGGAGCTATAAGGAAGGGA
AGGGGTTGAAGGGGTTTGTGAGAGAATTGAAGGAGGAGTTTGGGTTCGGTGGAGTACGTG
TACGTGTGGCAGCGGtTGTGCGGGTATTGGGGAGGGGTGAGGCCGGGGGTGGCGGGGAT
GGCGGAGGCGGCGGTGGAGAAGCCAAAGCTGACGGAGGGGTGAAGGGAACGATGGAGG
ATCTGGCGGTGGACAAGATTGTGAATAATGGGGTCCGGGTGGTGCCGCCGGAGCTGGTG
GGGAAATGTACGAGCGCCTTCACGCGCACTTGGAGAGTGCGGGTATTGATGGGGTCAA
AGTTGAcGTCATCCACTTGCTAGAAATGGTGTGTGAGAAATATGGAGGGCGAGTGGATA
TGCGGAAAGCATATTACAAAGCTCTCACTGCTTCCGTGAGAAAACATTTTAAGGGCAAC
GGCGTCATTGCCAGCATGGAGCATTGCAACGATTTTCATGTTGCTGGGAACTGAAGCAAT
ATCCCTTGGTCGTGTTGGGGATGATTTCTGGTGCCTGACCCTTATGGTGATCCAAATG
GTACATTTTGGCTACAAGGGTGTACATGGTGCATTGTGCATACAACAGCTTGTGGATG
GGCAATTTTCATCCACCAGATTGGGACATGTTCCAATCTACTCATCCTTGTGCTGCCTT
CCATGCTGCCTCAAGAGCCATATCTGGTGGCCCCATTTACATCAGTGACACAGTTGGGA
ACCACAACCTTTGAGCTGCTTAAGACCTTGGCCTTGCCAGATGGGTCCATCCTCAGATGT
GAGCACTATGCACTCCCAACCAGGGACTGTCTCTTTGCTGACCCTCTCCATGATGGCAA
ACAATGCTCAAGATATGGAACCTCAACAAGTACACTGGAGTTCTTGGGGTGTTTAACT
GCCAGGGAGGAGGTTGGTTCCGTGAGATTAGGTCCAACAAATGTGCTGCTGAGTTTTCT
CATAGGGTATCAACCAAGACCAATATCAAAGACATTGAATGGGATAGTGGAAAGAATCC

AATTTCCATTGAAGGGGTGCAACTTTTCGCTTCGTATTTTCAGCCAAGCCAAGAAACTCA
TCCTCTCAGCACCATCTGATGACAGTGAAGAGATTTCTTGGAGCCATTCAATTTGAG
CTTATAACAGTTTCCCCTGTGACTGTCTTGCCTGGCAAGTCAGTGAAGTTTGCTCCTAT
TGGTTTGGTGAATATGCTAAACACTGGTGGAGCAGTCCAGTCTTTAGCTTTTGATGAGG
GTCAGAAATTTGGTTGAAGTTGGTTTAAGAGGCACTGGGGAGATGAGAGTCTATGCCTCA
GAGAAACCAAGAACTTGTAGAATTGATGGCAAAGAAGTTGATTTTGAATATGAAGGGTC
TATGGTCAACATTCAAGTACCgTGGCCTGGTTCTTCAAAATTGTCCACTGTTTCAGTATG
TATTTTAAccccCTCGAGttttTCTAGActct

Figure A.2A: Gene sequence for RS3 including restriction sites for directional cloning. Restriction sites of interest are underlined. Coding sequence is represented by uppercase letters. Non-coding sequence or bases changed within the coding sequence to remove restriction sites are represented by lowercase letters. Restriction sites of interest are underlined. The 5' restriction sites are NdeI, EcoRI, and NcoI. The 3' restriction sites are XhoI and XbaI.

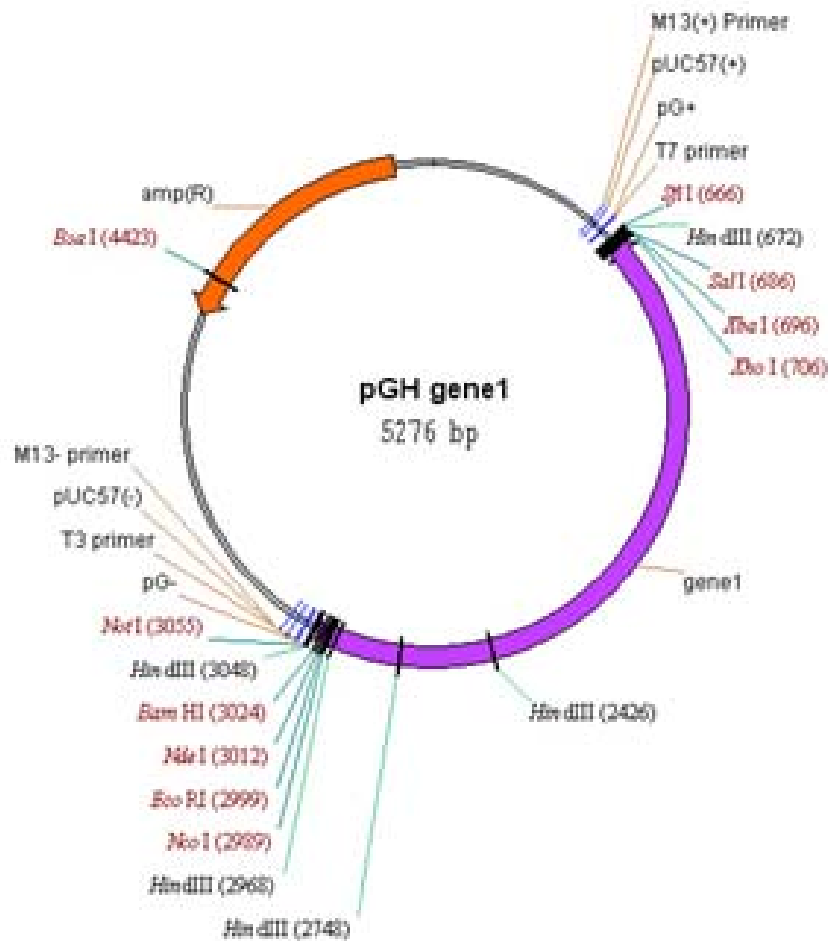
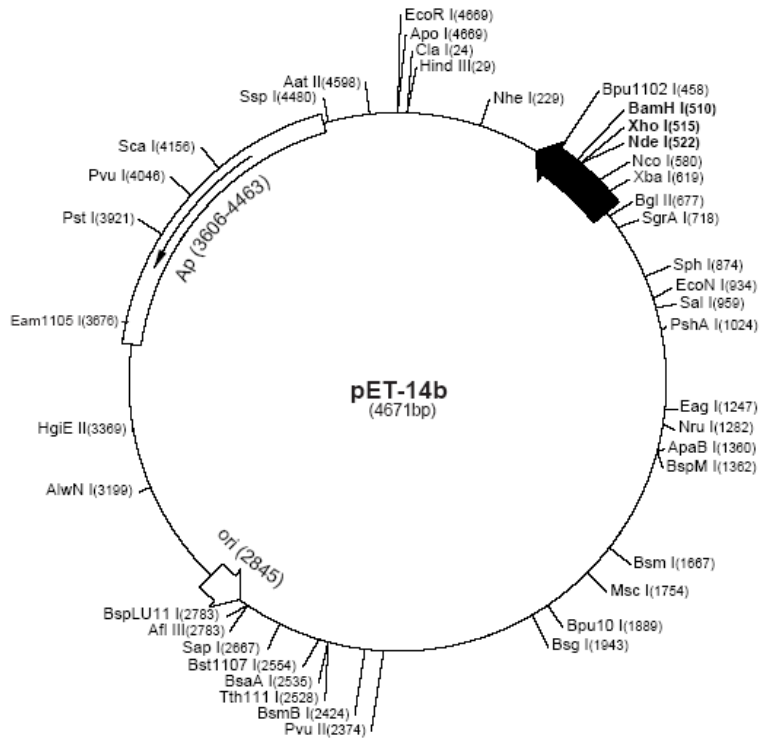


Figure A.2B: Celtek Genes order of RS3 in pGH vector. Vector map provided by manufacturer(Celtek Genes: www.celtekgenes.com). Gene1 refers to RS3. RS3 was ordered with a 5' NdeI, Ecor-1, and NcoI restriction sites, at the 3' end are XhoI and XbaI. These restriction sites were removed from the internal gene sequence by changing a base pair which did not result in an amino acid change.

pET-14b sequence landmarks

T7 promoter	646-662
T7 transcription start	645
His*Tag coding sequence	554-571
Multiple cloning sites	
(<i>Nde</i> I - <i>Bam</i> H I)	510-526
T7 terminator	404-450
pBR322 origin	2845
<i>bla</i> coding sequence	3606-4463



pET-14b cloning/expression region

Figure A.3: Vector map and polylinker of *E. coli* expression vector pET-14b as provided by manufacturer (Novagen).

pMAL-c4X Polylinker:

```

                    SacI
5' ma]E...TCG AGC TCG (AAC)n AAT AAC AAT (AAC)n CTC GGG ATC GAG GGA AGG ATT TCA
                                     Ile Glu Gly Arg
EcoRI  BamHI  XbaI  SalI  PstI  HindIII
GAA TTC GGA TCC TCT AGA GTC GAC  CTG CAG GCA AGC TTG ... lacZα
```

Figure A.4: Polylinker of pMAL-c4x as provided by manufacturer (NEB). No vector map available.

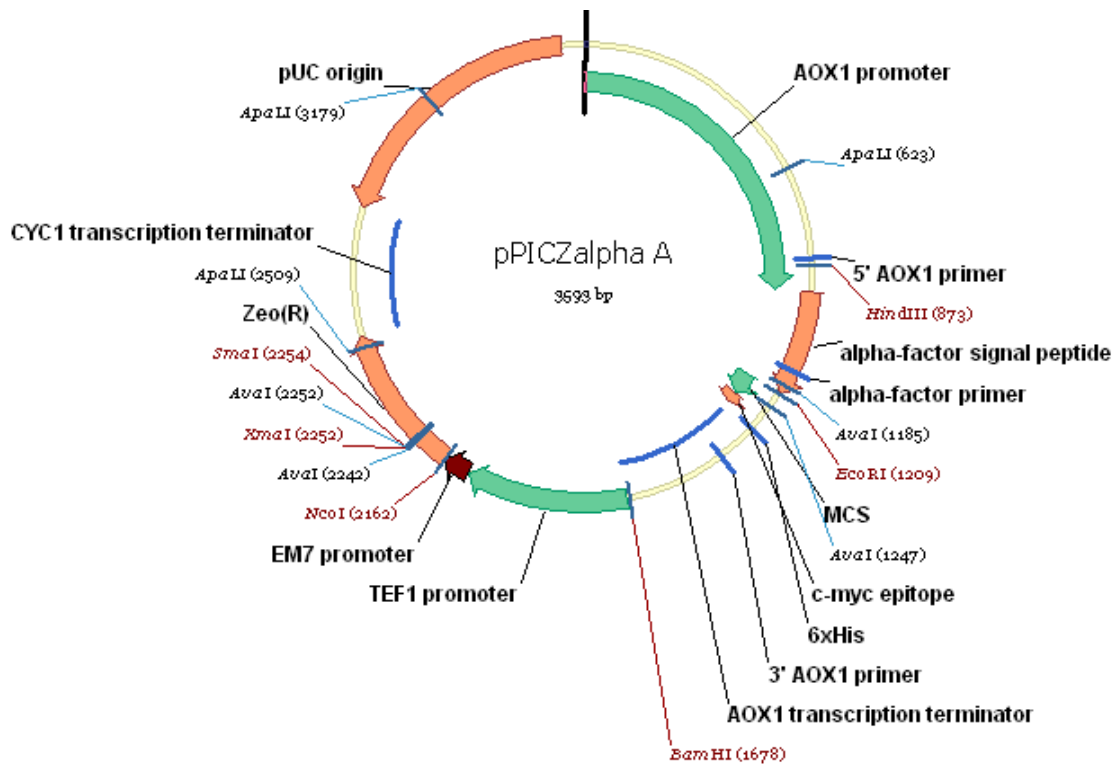
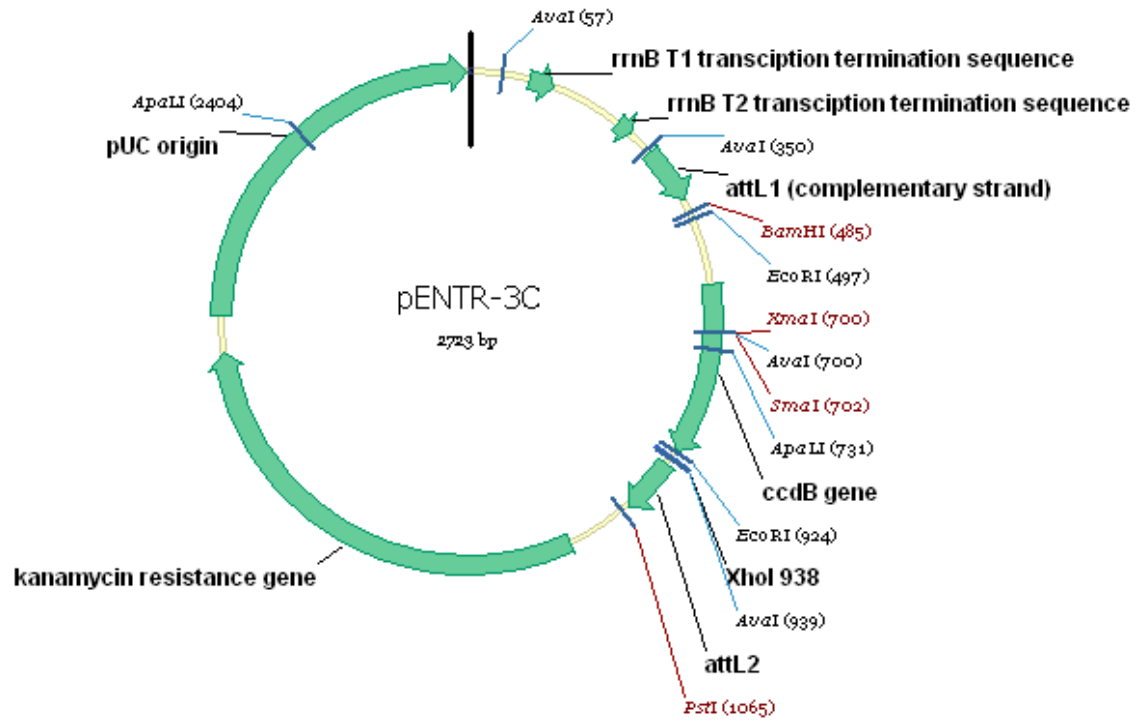


Figure A.5: Candidate gene RS2 in pPIC α -A, *Pichia pastoris* expression vector (Invitrogen). This vector exhibits zeocin resistance. The gene sequence is fused at the 5' end to the α -secretion factor; both are under the control of the AOX1 promoter which is induced by methanol.



Figures A.6: Vector map entry vector pENTR-3C for stable transformation of Arabidopsis (Invitrogen). The insertion of RS2 replaced the *ccdB* gene. This vector exhibited kanamycin resistance.

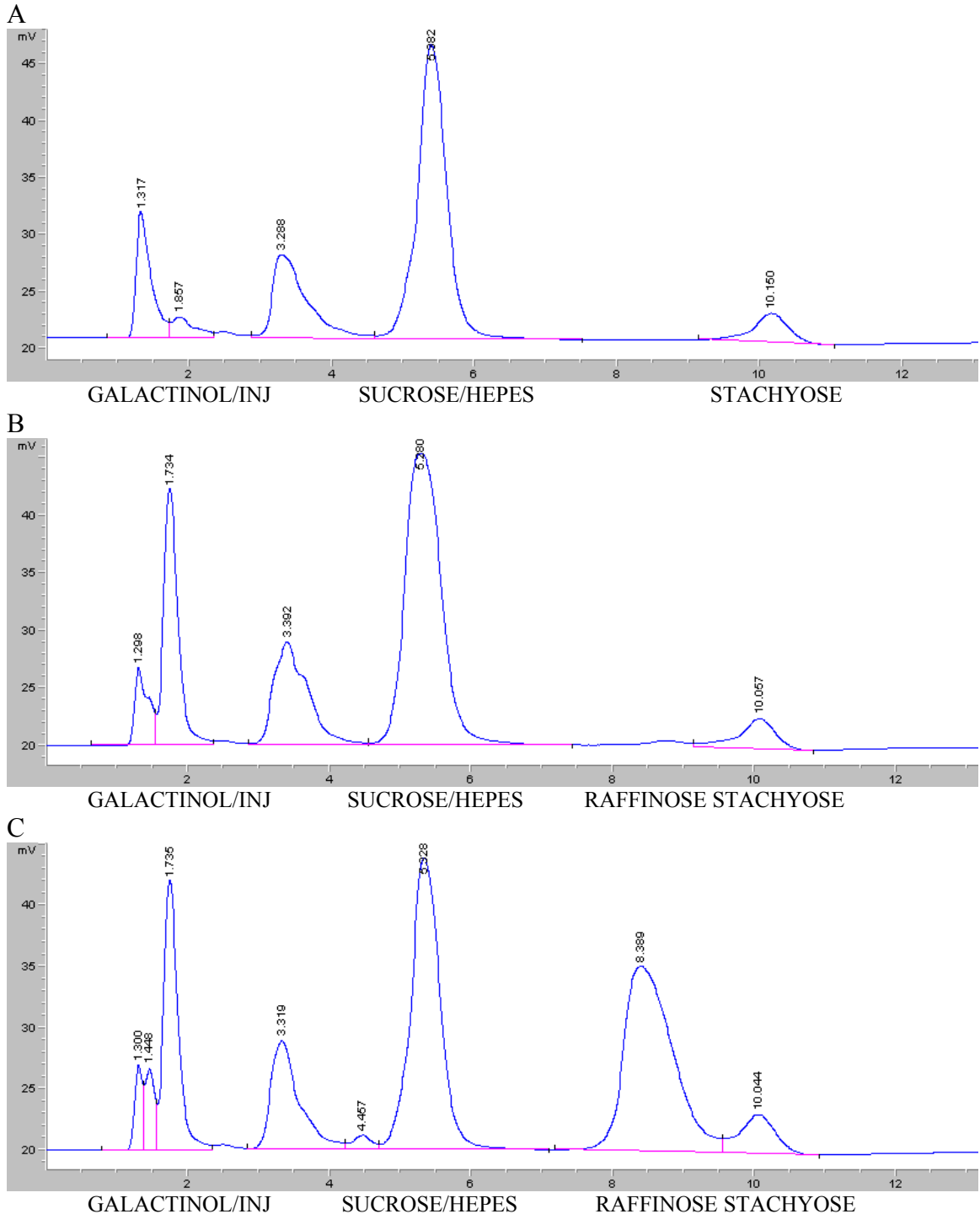


Figure A.7: Representative results from positive control assay for raffinose synthase and stachyose synthase activity from developing seeds. Inj refers to the injection peak which overlays galactinol elution. A. Time point zero for the activity assay demonstrates the initial amount of substrates. This HPLC profile is from an assay which included the substrates for the raffinose synthase enzyme. Both galactinol and sucrose can be observed. B. Timepoint 16 hours for raffinose synthase activity. HPLC profile indicates a small amount of raffinose has been produced. C. Time point 16 hours for stachyose

synthase activity. This HPLC profile has peaks for galactinol and raffinose, the substrates for the enzyme stachyose synthase. We have demonstrated that the peak near the sucrose retention time is the buffer component HEPES. There is no additional accumulation of stachyose indicating that the recombinant enzyme did not catalyze the synthesis of stachyose.

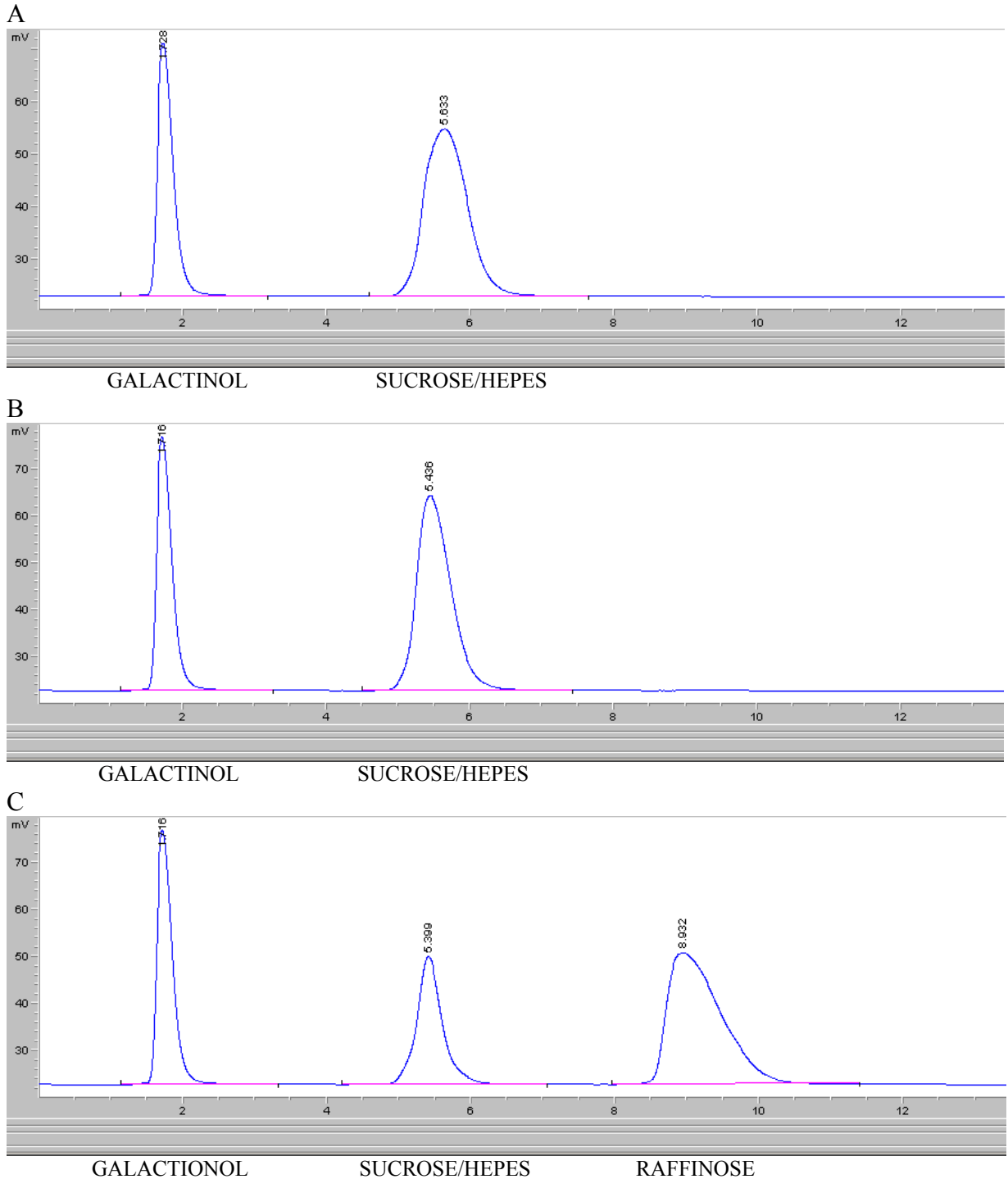


Figure A.8: Representative results from enzymatic activity assay for recombinant RS2 expressed in *Pichia pastoris*. A. Time point zero for the activity assay demonstrates the initial amount of substrates. This HPLC profile is from an assay which included the substrates for the raffinose synthase enzyme. Both galactinol and sucrose can be observed. B. Timepoint 16 hours for raffinose synthase activity. HPLC profile appears identical to time point zero indicating that no reaction has taken place which results in the

accumulation of raffinose. C. Time point 16 hours for stachyose synthase activity. This HPLC profile has peaks for galactinol and raffinose, the substrates for the enzyme stachyose synthase. We have demonstrated that the peak near the sucrose retention time is the buffer component HEPES. There is no accumulation of stachyose indicating that the recombinant enzyme did not catalyze the synthesis of stachyose.

TABLE

Table A.1: Growth conditions for recombinant protein expression in *E. coli*.

	temperature (° C)	aeration (RPM)	media/flask volume
normal conditions	37	200	1/8
low temperature	25	200	1/8
low aeration	37	125	1/2
low temperature, low aeration	25	125	1/2

LITERATURE CITED

- Dierking, E.C., and K.D. Bilyeu. 2008. Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. *The Plant Genome* 1:135-145.
- Hitz, W.D., T.J. Carlson, P.S. Kerr, and S.A. Sebastian. 2002. Biochemical and molecular characterization of a mutation that confers a decreased raffinose and phytic acid phenotype on soybean seeds. *Plant Physiology* 128:650-660.
- Krogdram, A.-M., K. Kristiansen, and J. Nøhr. 2003. Expression in *E. coli* Systems, p. 103-109, Vol. 232.
- Li, S., W.D. Kim, S. Kaneko, P.A. Prema, M. Nakajima, and H. Kobayashi. 2007. Expression of rice (*Oryza sativa* L. var. Nipponbare) alpha-galactosidase genes in *Escherichia coli* and characterization. *Bioscience, Biotechnology, and Biochemistry* 71:520-526.
- Peterbauer, T., and A. Richter. 1998. Galactosylononitol and stachyose synthesis in seeds of adzuki bean. *Plant Physiology*. 117:165-172.
- Peterbauer, T., L. Mach, J. Mucha, and A. Richter. 2002. Functional expression of a cDNA encoding pea (*Pisum sativum* L.) raffinose synthase, partial purification of the enzyme from maturing seeds, and steady-state kinetic analysis of raffinose synthesis. *Planta* 215:839-846.
- Sørensen, H.P., and K.K. Mortensen. 2005. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *Journal of Biotechnology* 115:113-128.

VITA

Emily Catherine Dunn Dierking was born January 14, 1983 along with her twin sister in Town and Country, Missouri and grew up on a farm in Bourbon, Missouri. She graduated salutatorian from Bourbon High School in 2001 and moved to Kirksville, Missouri to study biology at Truman State University. Emily earned her bachelors of science in biology in 2005, magna cum laude, with biology honors. She then moved to Columbia, Missouri to complete her PhD at the University of Missouri-Columbia in the Division of Plant Science under the advisement of Dr. Kristin Bilyeu. She is married to Ryan and has one dog, Fisher.