View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by ScholarWorks@UARK

University of Arkansas, Fayetteville
ScholarWorks@UARK

Theses and Dissertations

8-2014

Identification of Quantitative Trait Loci / Genes for Low-stachyose and High-sucrose Content in Soybean Seeds

Sandra Liliana Florez Palacios
University of Arkansas, Fayetteville

Follow this and additional works at: http://scholarworks.uark.edu/etd

Part of the Agronomy and Crop Sciences Commons, and the Plant Breeding and Genetics Commons

Recommended Citation
Florez Palacios, Sandra Liliana, "Identification of Quantitative Trait Loci / Genes for Low-stachyose and High-sucrose Content in Soybean Seeds" (2014). *Theses and Dissertations*. 2227.
http://scholarworks.uark.edu/etd/2227

This Dissertation is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact scholar@uark.edu, ccmiddle@uark.edu.

Identification of Quantitative Trait Loci / Genes for Low-stachyose and High-sucrose Content in Soybean Seeds

Identification of Quantitative Trait Loci / Genes for
Low-stachyose and High-sucrose Content in Soybean Seeds

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Crop, Soil, and Environmental Sciences

by

Sandra Liliana Florez-Palacios
Universidad Nacional de Colombia
Bachelor of Science in Agronomic Engineering, 2006
University of Arkansas
Master of Science in Crop, Soil, and Environmental Sciences, 2009

August 2014
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____
Dr. Pengyin Chen
Dissertation Director

_____
Dr. John Clark
Committee Member

_____
Dr. Andronikos Mauromoustakos
Committee Member

_____
Dr. Andrew Sharpley
Committee Member

_____
Dr. Ya-Jane Wang
Committee Member

Abstract

Seed carbohydrate content is an important aspect in breeding for food-grade soybeans commercialized in the soyfood market. Sucrose and stachyose are the primary carbohydrates in soybean seed. Sucrose affects the quality and taste of various soyfoods such as tofu, soymilk, and natto; however, consumption of soy-based products with high stachyose concentrations can cause diarrhea and flatulence. A mutant line (V99-5089) with high-sucrose and low-stachyose has been identified. The objectives of this research were: 1) to identify sucrose QTLs in a F_2-derived mapping population; 2) to investigate the genetic relationship between two low-stachyose sources, V99-5089 and the germplasm line PI200508; and 3) to identify stachyose QTLs in a F_2-derived mapping population. An allelism test for low stachyose was performed by using 121 F_2-derived lines from the cross PI200508 x V99-5089 grown in three different environments. Carbohydrate content was determined by a high performance liquid chromatogram system, and lines were classified as high- or low- stachyose. Chi-square analysis was performed to test for goodness-of-fit of observed segregation to the expected genetic ratio. Results showed a 9 high-stachyose : 7 low-stachyose ratio, indicating that two independent recessive genes conferred the low-stachyose trait in the two mutant lines. Additionally, gene dosage effect was observed; however, further study is required in order to confirm its presence. The sucrose and stachyose QTL studies were carried out in 92 $F_{2:7}$ lines derived from the cross V97-3000 x V99-5089. Leaf samples were collected at $F_{2:6}$ for DNA extraction and subsequent molecular analysis using single nucleotide polymorphism (SNP) markers. Seed from $F_{2:7}$ lines grown in two locations, each one with two replications, were analyzed for sugar content. Results showed two sucrose QTLs located on chromosomes 6 and 10, accounting for 17 and 11% of the phenotypic variation, respectively. Additionally, two stachyose QTLs were found on

chromosome 10 and 11, explaining 11 and 46% of the stachyose variation. Results from this research indicate that the two low-stachyose sources may serve as valuable parents in breeding for low-stachyose soybeans. Additionally, V99-5089 may also provide favorable alleles for breeding high-sucrose varieties. The sucrose and stachyose QTLs identified in these studies are stable across environments and will facilitate the marker-assisted selection for both traits.

Acknowledgements

I would never have been able to finish this doctoral dissertation without the guidance, help and support of wonderful people around me.

Above all, I would like to express my deepest thanks to my parents for their endless love and support during these years. Their encouragement when the times got rough are much appreciated and duly noted.

I would like to express my deepest gratitude to my research advisor, Dr. Pengyin Chen. His good advice, knowledge and support, have been invaluable on both an academic and personal level, for which I am extremely grateful.

I would also like to sincerely thank my Dissertation Committee members, Dr. John Clark, Dr. Andrew Sharpley, Dr. Andy Mauromoustakos, and Dr. Ya-Jane Wang, for the learning opportunities provided and for accepting being part of my committee.

My gratitude goes to past and present members of the soybean breeding program at the University of Arkansas. They helped me with the field activities and laboratory analyses, which made the long journeys of work much more enjoyable.

I thank the staff at the Department of Crop, Soil, and Environmental Sciences for their help with assisting me with any paper and office work.

Dedication

I dedicate this dissertation to my parents.

Table of Contents

Chapter I. Introduction

Soybean origin and market 1

Soyfoods 1

Health benefits 3

Seed composition 4

Protein 4

Carbohydrates 6

Oil 7

Mineral and vitamins 8

Other seed components 9

Phytate 9

Isoflavones 9

Lipoxygenases 10

Environmental effect on soybean seed components 10

Development of new cultivars 13

Molecular markers 14

Molecular linkage groups 16

QTLs associated with soybean agronomic traits 16

QTLs associated with main soybean seed carbohydrates 20

Research justification and objectives 23

Implications 24

References 25

Chapter II. Identification of Quantitative Trait Loci / Genes for Sucrose Content in Soybean Seed

Abstract 42

Introduction 43

Materials and methods 46

Population development and field experiment 46

Genotypic data 48

Phenotypic analysis 49

Statistical analysis 50

Results 51

Phenotypic data 51

QTL mapping in F_2-derived populations by SNP markers 52

Discussion 54

Conclusion 59

References 60

Chapter III. Allelism Test for Low-Stachyose Content in Soybean Mutants

Abstract 84

Introduction 85

Materials and methods 86

Parent materials 86

Population development and field experiment 87

Soluble carbohydrate extraction 88

Carbohydrate analysis 88

Statistical analysis 89

Results 90

Discussion 91

Conclusion 93

References 95

Chapter IV. Identification of Quantitative Trait Loci / Genes for Stachyose Content in Soybean Seed

Abstract 105

Introduction 106

Materials and methods 107

Population development and field experiment 107

Genotypic data 109

Phenotypic data 110

Statistical analysis 111

Results 112

Phenotypic data 112

QTL mapping in F_2-derived populations by SNP markers 112

Discussion 113

Conclusion 116

References 117

Overall conclusion 139

List of Tables

Chapter I

Table 1. Soybean linkage groups. 37

Table 2. Molecular markers reportedly associated with soybean seed sucrose content. 39

Table 3. Molecular markers reportedly associated with soybean seed stachyose content. 41

Chapter II

Table 1a. Seed sucrose content (% dry-weight basis) of mapping populations derived from the cross V97-3000 x V99-5089 evaluated across two locations and over three years. 65

Table 1b. Seed sucrose content (% dry-weight basis) of parents V97-3000 and V99-5089 evaluated across two locations and over three years. 66

Table 2. Analysis of variance for seed sucrose content of 92 $F_{2:7}$ population derived from the cross V97-3000 x V99-5089 grown in Fayetteville and Marianna in 2011. 67

Table 3. Summary of SNP markers used for the screening of an $F_{2:6}$ population derived from the cross V97-3000 x V99-5089. 68

Table 4. Single-marker analysis of variance for seed sucrose content in 92 F_2-derived lines from the cross V97-3000 x V99-5089 evaluated in Fayetteville in 2009 and 2010, and in Fayetteville and Marianna in 2011. 70

Table 5. Mean effect of SNP marker alleles on seed sucrose content in 92 F_2-derived population developed from the cross V97-3000 x V99-5089, grown in two locations and over three years. 71

Table 6. Average monthly temperatures (°C) in Fayetteville, AR in 2010, and Fayetteville and Marianna, AR in 2011. 72

Chapter III

Table 1. Mean and range of stachyose content for parents and populations derived from the PI200508 x V99-5089 cross grown in Fayetteville in 2010 and 2011, and Keiser, AR in 2011. 98

Table 2. Analysis of variance for seed stachyose content in $F_{2:7}$ lines derived from PI200508 x V99-5089 grown in two Arkansas locations in 2011. 99

Table 3. Chi-square test for goodness-of-fit to the two-gene model for stachyose content

of F_2-derived lines from the cross PI200508 x V99-5089 grown in Fayetteville, AR in 2010, Fayetteville and Keiser, AR in 2011. 100

Chapter IV

Table 1a. Seed stachyose content (% dry-weight basis) of mapping populations derived from the cross V97-3000 x V99-5089 evaluated across two locations and over three years. 121

Table 1b. Seed stachyose content (% dry-weight basis) of parents V97-3000 and V99-5089 evaluated across two locations and over three years. 122

Table 2. Summary of SNP markers used for the screening of an $F_{2:6}$ population derived from the cross V97-3000 x V99-5089. 123

Table 3. Single-marker analysis of variance for seed stachyose content in 92 F_2-derived lines from the cross V97-3000 x V99-5089 evaluated in Fayetteville in 2009 and 2010, and in Fayetteville and Marianna in 2011. 125

Table 4. Mean effect of SNP marker alleles on seed stachyose content in 92 F_2-derived population developed from the cross V97-3000 x V99-5089 grown in two locations and over three years 126

Table 5. Average monthly temperatures (°C) in Fayetteville, AR in 2010 and Fayetteville and Marianna, AR in 2011. 127

List of Figures

Chapter II

Fig. 1. Frequency distribution of seed sucrose content in populations derived from V97-3000 (P1) x V99-5089 (P2) evaluated in four environments: (a) $F_{2:6}$ lines in Fayetteville, AR in 2010; (b) $F_{2:7}$ lines in Fayetteville, AR in 2011; (c) $F_{2:7}$ lines Marianna, AR in 2011; (d) Combined data of $F_{2:7}$ lines in Fayetteville and Marianna, AR in 2011. 73

Fig. 2a. A genetic map constructed for chromosomes 1, 2, 3, and 4 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes. 74

Fig. 2b. A genetic map constructed for chromosomes 5, 6, 7, and 8 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes. 75

Fig. 2c. A genetic map constructed for chromosomes 9, 10, 11, 12, and 13 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes. 76

Fig. 2d. A genetic map constructed for chromosomes 14, 15, and 16 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes. 77

Fig. 2e. A genetic map constructed for chromosomes 17, 18, 19, and 20 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes. 78

Fig. 3a. Composite interval mapping using SNP markers for seed sucrose QTL on chromosome 6 in 92 F_2-derived lines from the cross V97-3000 x V99-5089, evaluated across locations and years: (a) in Fayetteville, AR in 2009; (b) in Fayetteville, AR in 2010. 79

Fig. 3b. Composite interval mapping using SNP markers for seed sucrose QTL on chromosome 6 in 92 F_2-derived lines from the cross V97-3000 x V99-5089, evaluated across locations and years: (c) in Fayetteville, AR in 2011; (d) in Marianna, AR in 2011. 80

Fig. 3c. Composite interval mapping using SNP markers for seed sucrose QTL on chromosome 6 in 92 F_2-derived lines from the cross V97-3000 x V99-5089 grown in Fayetteville and Marianna in Arkansas in 2011 (Combined data). 81

Fig. 4a. Composite interval mapping using SNP markers for seed sucrose QTL on chromosome 6 in 92 F_2-derived lines from the cross V97-3000 x V99-5089 grown: (1) in Fayetteville, AR in 2009 and (2) in Fayetteville, AR in 2010. 82

Fig. 4b. Composite interval mapping using SNP markers for seed sucrose QTL on

chromosome 6 in 92 F_2-derived lines from the cross V97-3000 x V99-5089 grown in Fayetteville, AR in 2009 and 2010 (Combined data). 83

Chapter III

Fig. 1. Biosynthetic pathway of raffinose family oligosaccharides (RFO). 101

Fig. 2. Frequency distribution of seed stachyose content in F_2-derived populations from PI200508 (P1) x V99-5089 (P2) evaluated in three environments: (a) $F_{2:6}$ lines in Fayetteville, AR in 2010; (b) $F_{2:7}$ lines in Fayetteville, AR in 2011; (c) $F_{2:7}$ lines Keiser, AR in 2011. 102

Fig. 3. Frequency distribution of seed stachyose content in $F_{2:7}$ population derived From the cross PI200508 (P1) x V99-5089 (P2) evaluated in Fayetteville and Keiser, AR in 2011. 103

Fig. 4. Two-gene model for stachyose content of F_2-derived lines from the cross PI200508 x V99-5089. 104

Chapter IV

Fig. 1. Frequency distribution of seed stachyose content in populations derived from V97-3000 (P1) x V99-5089 (P2) evaluated in three environments: (a) $F_{2:6}$ lines in Fayetteville, AR in 2010; (b) $F_{2:7}$ lines in Fayetteville, AR in 2011; (c) $F_{2:7}$ lines Marianna, AR in 2011. 128

Fig. 2a. A genetic map constructed for chromosomes 1, 2, 3, and 4 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes. 129

Fig. 2b. A genetic map constructed for chromosomes 5, 6, 7, and 8 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes. 130

Fig. 2c. A genetic map constructed for chromosomes 9, 10, 11, 12, and 13 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes. 131

Fig. 2d. A genetic map constructed for chromosomes 14, 15, and 16 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes. 132

Fig. 2e. A genetic map constructed for chromosomes 17, 18, 19, and 20 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes. 133

Fig. 3a. Composite interval mapping using SNP markers for seed stachyose QTL on chromosome 10 in 92 F_2-derived lines from the cross V97-3000 x V99-5089, evaluated across locations and years: (1) in Fayetteville, AR in 2009; (2) in Fayetteville, AR in 2010. 134

Fig. 3b. Composite interval mapping using SNP markers for seed stachyose QTL on chromosome 10 in 92 F_2-derived lines from the cross V97-3000 x V99-5089 grown in Fayetteville 2009 and 2010 (Combined data). 135

Fig. 4a. Composite interval mapping using SNP markers for seed stachyose QTL on chromosome 11 in 92 F_2-derived lines from the cross V97-3000 x V99-5089, evaluated across locations and years: (1) in Fayetteville, AR in 2009; (2) in Fayetteville, AR in 2010. 136

Fig. 4b. Composite interval mapping using SNP markers for seed stachyose QTL on chromosome 11 in 92 F_2-derived lines from the cross V97-3000 x V99-5089, evaluated across locations and years: (1) in Fayetteville, AR in 2011; (2) in Marianna, AR in 2011. 137

Fig. 4c. Composite interval mapping using SNP markers for seed stachyose QTL on chromosome 11 in 92 F_2-derived lines from the cross V97-3000 x V99-5089 grown in Fayetteville and Marianna, AR in 2011 (Combined data). 138

I. Introduction

Soybean origin and market

Soybean [*Glycine max* (L.) Merr] is a member of the *Fabaceae* family, whose center of origin and domestication was located in China (Abe et al., 2003; Hymowitz, 1970). After domestication, the area planted with soybean expanded worldwide due to the easy adaptation of the crop and the multipurpose end-uses of the seeds (Johnson et al., 1992). The three major soybean producers in the world in 2013 were: Brazil, accounting for 39% of world production with 83.5 million metric tons; United States, with 37% and 82 million metric tons; and Argentina, with 11% and 51.5 million metric tons (ASA, 2014).

In the United States, the area occupied by soybean ranks second (29%) after corn (35%) (ASA, 2014). In 2013, 31.2 million ha of soybean were planted in the country. Arkansas accounted for 3.2 million ha, which produced a total of 3.7 million metric tons of soybean (ASA, 2014).

Soybeans can be used as raw material for many manufactured goods including oil, whole-bean products, and meal products (Smith and Huyser, 1987). In 2013, oil production in US was 8.6 million metric tons, soybean crush was estimated in 43.9 metric tons, and meal production reached 34.9 million metric tons (ASA, 2014). Additionally, soybeans are used as livestock feed, as an ingredient in the food industry, as fuel when converted into biodiesel, and as the main ingredient of industrial products including lubricants, insulators, paints, glues, crayons, hand cleaners, and shampoos (ASA, 2014).

Soyfoods

Soybeans have been grown for more than 1000 years in Asia for consumption purposes.

However, their consumption in the Western culture commenced only about 30 years ago, when their nutritional value and health benefits began to be publicized (Golbitz, 1995). According to Soyatech (2012), the sales of soyfoods products in U.S. exceeded $5 billion in 2011. These specialty soybeans are grown and commercialized under identity-preserved (IP) contracts (Sonka et al., 2004), which means that growers get paid a premium for producing top-quality soybeans which meet specific standards required for the soyfood industry, including low-saturated fat content, low-linolenic acid content, increased oleic acid content, and an absence of beany flavor.

Soyfoods can be classified as traditional non-fermented foods, which include tofu, edamame, and soymilk*;* and traditionally fermented foods such as miso, natto, soy sauce, and tempeh (Cui et al., 2004). Tofu is curd soybean made by soymilk that has been coagulated by calcium, magnesium salt, or glucono delta-lactone (He and Chen, 2013). Edamame is soybeans harvested early in the vegetable stage which can be commercialized in fresh or frozen food forms. Soymilk is one of the most popular and traditional soyfoods. According to oral historical records, it was developed by the King Liu An in the second century B.C. Soymilk is made by soaking, grinding, and straining soybeans (Shurtleff and Aoyagi, 2013).

Miso is a soybean paste fermented with fungus, bacteria, or yeast. The choice of the microbe used as the fermentation agent depends on the country where the miso is produced. Most of the miso made in Japan involves the *Aspergillus oryzae* fungus while the *Bacillus subtilis* bacteria are the typical fermentation agents of the Chinese miso (Kim et al., 2010). Natto is manufactured from fermentation of soybeans by strains of *Bacillus subtilis*. During its storage, secondary fermentation occurs, which causes ammonia to be released. This gives natto its characteristic odor (Kada et al., 2008). Soy sauce is the most common seasoning in Chinese culture; it is

produced by fermenting a combination of wheat and soybean by *Aspergillus* fungi, lactic acid bacteria, and yeast (Sugiyama, 1984). Tempeh is made of de-hulled fermented soybeans by using *Rhizopus oryzae* mold (Shurtleff and Aoyagi, 2011).

Health benefits

Soyfoods are considered beneficial to overall health due to their high content of polyunsaturated fats (essential fatty acids and linolenic acid), vitamins, fiber, minerals, and low content of saturated fats. Some studies have reported that soy-based foods may reduce the risk of breast and prostate cancer, osteoporosis, and cardiovascular disease (CVD). In a study conducted by Sacks et al. (2006), isolated soybean protein and isoflavones reduced LDL cholesterol concentrations (~ 3%), as compared with milk and other sources of protein. No undesired effects on HDL cholesterol, triglycerides, lipoprotein, or blood pressure were observed. Similarly, Anderson et al. (1995) studied the relationship between soy protein consumption and serum lipid concentration in humans. They reported that the intake of soy protein significantly reduced serum concentrations of total cholesterol, LDL cholesterol, and triglycerides. Jenkins et al. (2010) reviewed and re-analyzed the two previous studies aiming to determine whether the claims that soy proteins have a beneficial effect on heart health were still maintained. They determined that both studies are well supported by the data collected in the meta-analyses.
The consumption of stearic acid, which is present in soybeans, is desirable in comparison to other saturated fats, because it simulates the effects of some mono-unsaturated fatty acids by not altering HDL cholesterol (Kris-Etherton and Yu, 1997), and by decreasing both LDL cholesterol and the total cholesterol to LDL cholesterol ratio (Hunter et al., 2010).

Additionally, many other studies have assessed the relationship between the intake of vitamins present in soybeans and prevention of some diseases. Results from a vast amount of studies, including animal experiments, epidemiological data, and human trials, suggest that vitamin E offers significant protection against heart disease (Pryor, 2000). Other benefits reported include improvement of cell-mediated immune responses (Meydani et al., 1997; Meydani et al., 1990) and cancer prevention (Yang et al., 2012; Watson and Leonard, 1986). Similarly, intake of vitamin B has been suggested to be associated with prevention of coronary heart disease in women (Rimm et al., 1998) and vascular disease in men (Ubbinik et al., 1993). Additionally, it has been reported that consumption of vitamin K is positively associated with bone mineral density in elderly men (Fujita et al., 2010).

Seed composition

Protein (40%) is the most abundant soybean seed component on a dry-weight basis, followed by carbohydrates (35%), oil (20%), and ashes (5%) (Liu, 2005; Hymowitz and Collins, 1974). Other important seed components include phytate, isoflavones, lipoxygenases, minerals, and vitamins.

Protein

Soybeans are mainly grown for protein and oil. In average, soybean seed has around 42% protein content, however, values ranging between 34 and 57% have been reported (Wilson, 2004). The two major protein fractions in soybean are beta-conglycinin (7S) and glycinin (11S) (Wolf, 1970). The 7S fraction represents one-third of the total soybean protein (37%) and includes the proteins hemagglutinin, lipoxygenase, β-amylase, and 7S globulin. The most abundant protein in the 7S fraction is 7S globulin. The 11S fraction represents another one-third

of the total soybean protein (31%) and 11S globulin is the main protein in this fraction. The main common property to both 7S and 11S globulins is the capacity to form disulfide-linked polymers which causes insolubility (Nash and Wolf, 1967) and increases viscosity (Circle et al., 1964). Soy proteins have emulsifying properties; they decrease the interfacial tension between the water and oil, and help to stabilize the emulsion by forming a physical barrier at the interface (Molina et al., 2001). The 11S-to-7S ratio plays an important role not only in the production of soyfoods but also in food industries where these proteins are used as natural gelling agents. Thus, a high 11S-to-7S ratio is desirable because it increases the firmness and water-holding capacity of the gel (Yagasaki et al., 2000).

Among the soybean essential amino acids, glutamic acid is the most abundant (Van Etten et al., 1959). Glutamic and aspartic acids account for 25% of the total of amino acids, followed by the hydrophobic side amino acids: lysine, arginine, histidine, glycine, alanine, valine, leucine, and isoleucine. The hydrophobic amino acids make up 20% of the total amino acid content. An additional smaller group, constituted by the aromatic amino acids phenylalanine, tyrosine, and tryptophan, account for 9% (Zarkdakas et al.,1993).

A rapid accumulation of protein occurs during the reproductive stage (Kim et al., 2006a), reaching 61% of total protein content at maximum fresh weight stage (MFWS) (Rackis, 1981). The MFWS begins right after R6, about 45 to 50 days after flowering (DAF), when the seeds are still green but the pods start turning yellow (Rackis et al., 1972). High-molecular-weight proteins are the most abundant proteins in mature seeds (Liu, 1997a), achieving their maximum levels at 36 DAF (Hill and Breidenbach, 1974a,b). Conversely, lower-molecular-weight proteins are predominant at the earliest stages of maturity (Liu, 1997a). The amino acid contents such as arginine, serine, glutamic acid, glycine, and leucine increases linearly with seed maturation,

whereas histidine and alanine contents decrease (Yazdi-Samadi et al., 1977). Free amino acid contents are lowered during maturity (Yanagisawa et al., 1997).

There are three major forms of soybean proteins; flours and grits, concentrates, and isolates (Wolf, 1970). Flours and grits are the least refined form because they vary in fat content, particle size, texture, and other properties. Flours are produced by grinding soybean flakes, and then sifting them through a No. 100 mesh screen with an opening of 0.149 mm. Grits are characterized for having a particle size larger than a 100 mesh (Horan, 1966). Protein contents for flours and grits range from 40 to 50% on dry-weight basis. Concentrates are made from defatted flakes or flour, and their protein content is around 70%. Isolates are the most refined soybean proteins, containing 90% or more protein (Meyer, 1966).

Carbohydrates

Carbohydrates are classified into two different groups: soluble and insoluble. In the soyfood industry, soluble carbohydrates are considered the most important desired seed sugars, because they affect the quality and taste of soy-based products (Taira, 1990). Soluble carbohydrates are separated based on their carbon chain length. Thus, glucose and fructose are monosaccharides, followed in size by the disaccharide sucrose, and then the oligosaccharides raffinose and stachyose (Hymowitz et al., 1972; Liu, 1997). Sucrose content accounts for 41.3 - 67.5%, raffinose for 5.2 - 15.8%, and stachyose for 12.1 - 35.2% of the total soluble sugars in soybean seed (Yazdi-Samadi et al., 1977).

High contents of glucose, fructose, and sucrose are desirable; however, low contents of raffinose and stachyose contents are preferred for the production of soyfoods. Monogastric animals lack the α-(1,6)-galactosidase enzyme that converts oligosaccharides into digestible forms such as glucose and fructose (Gitzelmann and Auricchio, 1965). Consequently,

consumption of soybeans with high raffinose and stachyose contents results in low metabolizable energy, flatulence, and diarrhea (Kuriyama and Mendel, 1917; Rackis, 1974; Hawton et al., 1996).

There are several studies that describe the accumulation of soluble carbohydrates during seed formation and development. It has been established that glucose, fructose, and sucrose decrease in the initial stages of seed formation. Subsequently, their contents remain steady for about 3 weeks, and then increase towards the end of seed maturation (Yazdi-Samadi et al., 1977; Min 2008; Obendorf et al., 2009). Similarly, Kim et al. (2006a) reported a comparable pattern of sucrose accumulation during the reproductive stage. Raffinose and stachyose mostly accumulate from 40 to 60 DAF, during the drying phase of seed maturation (Yazdi-Samadi et al., 1977; Min 2008; Obendorf et al., 2009; Saldivar, 2011).

Oil

According to the American Soybean Association (2011), soybeans are the most important oilseed crop worldwide. Average seed oil concentration has been estimated as 20% (dry-weight basis), however, values as low as 8 to 28% have been observed (Wilson, 2004). Oil is accumulated in the seed in the form of triacylglycerol (Hajduch et al., 2005).

Fatty acid content in soybean oil is composed of 110 to 120 g kg^{-1}palmitic acid (C16:0), 40 g kg^{-1} stearic acid (C18:0), 230 to 240 g kg^{-1} oleic acid (C18:1), 530 g kg^{-1} linoleic acid (C18:2), and 70 to 80 g kg^{-1} linolenic acid (C18:3) (Wilson, 2004; Liu, 2005).

Soybean oil has some undesirable characteristics including low oxidative stability, which has a negative effect on its shelf life and durability at high temperatures. Additionally, its low cold-flow capacity can be an obstacle for biodiesel production. However, other components such as monounsaturated fatty acid (i.e. oleic acid) can improve functionality of soybean oil at high

concentrations (>500 g kg^{-1}) (Wilson, 2004). Likewise high palmitate content is desirable because it increases the oil oxidative stability which is beneficial for industrial and food purposes. Palmitate is considered one of the most desired saturated fatty esters in soybean oil (Shen et al., 1997).

Seeds reach 5% oil (dry weight basis) at 25 DAF (Rubel et al., 1972). Then, oil percentage increases rapidly until 40 DAF. At this stage, oil accumulation makes up for 30% of the total oil of the mature seed. The remaining oil is synthesized during the next 40 to 64 DAF, which is also the drying phase of seed maturation (Hajduch et al., 2005). Percentages of palmitic, stearic, and linolenic acid in the oil decrease with seed maturation, whereas percentages of oleic and linoleic acid increase (Rubel et al., 1972).

Mineral and vitamins

Potassium is the most abundant mineral in soybean seed, followed by smaller amounts of P, Ca, Mg, S, Cl, and Na, whose concentrations range between 0.2 and 2%. Minor minerals content range between 0.01 to 140 ppm, and include Si, Zn, Fe, Cu fluorine, Mn, and Cd (Liu, 1997). In general, mineral content remains steady during maturation (Rackis, 1981).

Soybean vitamins can be classified based on their solubility. Soluble vitamins include vitamin B whose main constituents are thiamine (vitamin B_1), riboflavin (vitamin B_2), pyridoxine (vitamin B_6), and niacin (vitamin B_3) (Lebiedzinska and Szefer, 2006). Concentrations of vitamins B_1, B_2, and B_3 are usually 13 to 19 ppm, 3 to 4 ppm, and 20 to 26 ppm, respectively. Other soluble vitamins include pantothenic acid and folic acid whose contents range from 15 to 37 ppm and 1 to 2 ppm, respectively (Burkholder and McVeigh, 1945).

Insoluble or fat-soluble vitamins include vitamin A, whose main component is β-carotene, and vitamin E, which is composed mainly of four tocopherols (α-, β-, γ-, and δ-

tocopherol). These tocopherols are natural chain-breaking antioxidants that can inhibit lipid degradation (Packer and Landvik, 1989; Carrão-Panizzi, 2007). It has been established that β-carotene content decreases with maturation (Rackis, 1981). Vitamin E content in soybean seed varies with values ranging from 10.9 to 28.4, 150 to 191, and 24.6 to 72.5 μg/g (dry basis) for α-, γ-, and δ-tocopherols, respectively (Guzman and Murphy, 1986).

Other seed components

Phytate

In soybean, P is in a complex as phytic acid, or myo-inositol-1, 2, 3, 4, 5, 6-hexakisphosphate or Ins P_6 (salt form: 'phytate'). Phytates include Ca^{+2}, Mg^{2+}, Zn^{2+}, and Fe^{3+} salts of hexaphosphoric acid (Oltmans et al., 2005). Phytate is indigestible to monogastric animals due to the lack of the enzyme *phytase* in their gastrointestinal tract which is necessary to catalyze the breakdown of phytate into digestible forms. Thus, most of the P passes through the intestinal tract of the animal without being used and is then excreted as waste. This manure contributes to environmental pollution when in contact with ground or surface waters by promoting eutrophication (Sharpley et al., 1994; Burkholder et al., 2007; Saghai -Maroof et al., 2009) Moreover, when phytate is consumed, this polianion chelates the cations Ca, Zn, and Fe, which are key minerals in the animals' diet (Raboy et al., 2001). According to Rackis (1981), phytate accumulates during seed maturation and concentration ranges from 1000 to 3000 μg g^{-1}.

Isoflavones

Flavonoids are polyphenolic phytochemicals. These phenolic compounds have low molecular weight and are produced as secondary metabolites. According to the International Union of Pure and Applied Chemistry (IUPAC) (1995), flavonoids are derived from 2-

phenylchromen-4-one, ‘neoflavonoids’ derived from 4-phenylcoumarin, and ‘isoflavonoids’ derived from 3-phenylchromen.

The main isoflavones in soybeans are genistein (4’,5,7-trihydroxyisoflavone) and daidzein (4’,7-dihydroxyisoflavone). Glycitein (4’,7-dihydroxy-6-methoxyisoflavone) is considered a minor isoflavone (Anderson and Wolf, 1995). In general, genistein tends to accumulate in larger quantities in soybeans, compared to daidzein (Murphy et al., 1999). Isoflavone concentrations have been reported to increase with seed maturation from 20 to 22 mg g^{-1} (Rackis, 1981).

Lipoxygenases

Lipoxygenases are enzymes that catalyze the oxygenation of polyunsaturated fatty acids to form monohydroperoxides (Yenofsky et al., 1988). These enzymes are part of several physiological processes, including plant growth, fruit ripening, abscission (Veldink et al., 1977), senescence (Leshem, 1984), and responses to biotic and abiotic stresses (Shin et al., 2012). Soybean seeds possess three lipoxygenases: LOX1, LOX2, and LOX3. Their enzymatic activity causes the characteristic beany flavor of the soyfoods (Kitamura et al., 1985). It has been reported that such activity increases with seed maturation producing a final concentration between 0.84 and 1.39% (Rackis et al., 1972).

Environmental effects on soybean seed composition

The percentages of seed components have been reported to be affected by genotypic and environmental factors (Cicek et al., 2006). Ren et al. (2009) studied the effect of high temperature on soybean seed composition under environment-controlled chamber conditions. Forty plants from the breeding line N98-4445A (mid-high oleic acid content) were grown at a

27/18°C day/night temperature with a 13.5-h day photoperiod of 880 μmol m^{-2} s^{-1}. Once the plants reached the R5 stage, half of the plants were discarded and 10 out of the remaining plants were treated with a 37/30°C day/night temperature with a 13.5-h day photoperiod of 880 μmol m^{-2} s^{-1} from the R5 stage until the R8 stage. Results indicated that high temperature increased total fatty acid concentration, while no significant effect was observed on individual sugar (sucrose, raffinose, and stachyose), protein, and phytic acid concentrations.

In a different study, Gao et al. (2009) studied the effect of the agronomic management systems: tilled with conventional chemical inputs (CT), no-till with conventional chemical inputs (NT), tilled with low chemical inputs (LI), and tilled with no chemical inputs (ORG), on total oil, oleic acid and linolenic acid soybean content over five years. Results indicated that there was no significant effect on total oil content and oil yield on a land-area basis and management systems had negligible differences among treatments. Likewise, no significant treatment effect was observed on oleic acid and linolenic acid content across years.

More recently, VanToai et al. (2012) conducted a study to evaluate the effect of flooding on seed composition of five soybean plant introductions tolerant to this stress and the sensitive cultivar Williams when grown in three different environments. Results showed a significant decrease in linoleic and linolenic acids, daidzein, genistein, and glycitein contents while oleic and stearic acid increased in all the genotypes studied. No significant environmental effects were observed on palmitic and linolenic acid, and glycitein concentrations.

Furthermore, Taira (1990) reported that protein, 7S and 11s globulins, amino acids, linolenic acid, and raffinose and stachyose content were primary affected by genotype, whereas sucrose was mainly influenced by location. He also observed significant yearly variations of oil, and oleic and linolenic acids, and sucrose, total sugar content. Geater et al. (2000) evaluated the

effect of genotype and environment on the seed traits total sugar, sucrose, raffinose, stachyose, protein, oil, protein + oil, fiber, and protein + oil + fiber. Sixteen small-seeded natto lines were grown in three different locations and for two years. Significant differences among the genotypes were observed for one or both years for all traits studied except fiber. Additionally, significant differences among locations and years were reported for all traits except stachyose.

Arslanoglu et al. (2011) evaluated the effects of genotype, environment, and their interaction on oil and protein content in eight soybean cultivars grown at eight different locations in Turkey during two years. The genotype x environment interaction effect was significant on the two traits studied ($p < 0.01$). Likewise, differences among genotypes were significant as well as differences among locations for oil and protein content with values ranging from 18.3 to 23.1% and 29.2 to 38.6%, respectively.

In a bigger study, Carrera et al., (2011) evaluated the effect of climatic variables on amino acid composition during seed filling. Nine soybean genotypes were grown in 31 environments, resulting from combining 13 locations and different planting dates during two years. Climatic variables included average daily temperature, cumulative solar radiation, and precipitation minus potential evapotranspiration. Environment accounted for most of the total variation for all traits studied followed by the genotype x environment interaction. Total amino acid concentration ranged between 31.7 and 49.1% while total essential and non-essential amino acids varied from 12.8 and 19.0% and from 18.9 and 31.2%, respectively.

Cober et al. (1997) evaluated the environment and genotype x environment effects on oil and sugar soybean seed content in two independent experiments. Contrary to Geater et al. (2000) findings, no significant location effects on either of the traits studied were observed, whereas genotype x environment interactions were negligible for both traits but their variance

components had a lesser magnitude than that of the genotype effect alone.

In a separate study, Geater and Fehr (2000) observed significant differences among cultivars for total sugar content. The range of values in the study was greater among cultivars than among locations. As a result, genotype was suggested to have a greater influence than environment on total sugar content. In a different study, Dardanelli et al. (2006) reported that maturity group had a significant effect on seed oil content in soybean. In the same study, maturity group x environment effects were greater for protein and protein and oil content combined than those for oil.

Development of new cultivars

The main objective of plant breeding is to develop new cultivars with desirable traits by using the genetic variability available (Jauhar, 2006; Sleeper and Poehlman, 2006). Traditional plant breeding involves a series of generational advances and subsequent selections until the desired trait(s) become(s) stable and genetically uniform (homozygous). Breeding lines containing the desirable trait(s) are then evaluated for agronomic performance across locations and years. For this reason, in general, the development of new cultivars requires substantial time and resources.

Furthermore, the nature of inheritance of these desired traits will determine the level of complexity of the breeding process. Correlations among traits must be considered when making selections since selecting one trait may result in another trait being enhanced or diminished.

There are several studies on relationships among soybean traits. Wilcox and Shibles (2001) reported that yield is negatively correlated with seed protein content and positively correlated with oil. Also, a negative correlation between protein and oil has been established (Hartwig et

al., 1997; Filho et al., 2001; Wilcox and Shibles, 2001).

Furthermore, soybean seed Ca content was positively associated with water absorption (Saio, 1976), phytate content (Schaefer and Love, 1992), and seed hardness (Chen et al., 1993). However, other studies have shown no relationship between Ca content and water absorption (Mullins and Xu, 2001; Wei and Chang, 2004). Additionally, negative correlations between seed size and seed hardness (Zhang et al., 2008) and seed size and linoleic acid (Maestri et al., 1998) were observed. Maestri et al. (1998) also indicated that seed size was positively associated with stearic and oleic acids.

Molecular markers

Traits present different levels of complexity which depend upon the number of genes that control them. Moreover, the different correlations, the effects that the environment may have on them, and time invested for their measurement make their study difficult. For these reasons, the use of molecular markers that are linked to the gene(s) of interest has become an effective alternative when breeders want to develop cultivars with specific characteristics in a time-efficient manner.

Quantitative traits are affected by multiple genes and each of these genes has a small effect that adds up to the total effect (Falconer, 1989). Additionally, these traits are greatly affected by the environment and are controlled by quantitative trait loci (QTL). A QTL is either considered the region of the chromosome to which a molecular marker (DNA sequence) is linked or the region defined by two bordering markers that are not necessarily linked to the QTL (Melchinger, 1998).

There are several types of molecular markers including Amplified Fragment Length

Polymorphism (AFLP), Restriction Fragment Length Polymorphism (RFLP), Random Amplified Polymorphic DNA (RAPD), Single Sequence Repeat (SSR), and more recently, Single Nucleotide Polymorphism (SNP).

In general, AFLP and RFLP techniques involve digestion of genomic DNA by restriction enzymes to produce small fragments; fragments are amplified and their size is determined. Both markers allow amplification of specific restriction fragments and distinction of heterozygous loci in the population studied (codominance). However, the procedures require large amounts of DNA and their execution is cumbersome (Rafalski et al., 1991).

With RAPD markers, random segments of genomic DNA are amplified. The main advantage of this technique is that small amounts of DNA are required as compared with RFLP. However, RAPD markers are dominant which means that they only identify one allele at a locus. Thus, they cannot be used to detect heterozygous loci (Rafalski et al., 1991).

Microsatellites, or SSR markers, are stretches of DNA in which the same short nucleotide sequence (2-4 base pairs) is repeated. Polymorphism (different alleles) is determined by the number of times the short sequence repeats. Then, DNA fragments are amplified by Polymerase Chain Reaction (PCR) and separated through gel electrophoresis. These markers are codominant; therefore, it is possible to distinguish between the heterozygote and homozygote.

Alternatively, SNP markers determine differences at the single nucleotide sequence level. Advantages to this technique are that SNPs span the whole genome which provides great coverage. Additionally, sample processing may be completely automated and the use of pre-made chips makes this procedure very time efficient.

Molecular linkage groups (MLG)

Linkage groups are composed by linked genes on the same chromosome. Soybeans have 20 linkage groups and a diploid number of 2n = 40. The number of linkage groups corresponds to the haploid number of chromosomes (Roberts, 1986). The current soybean genetic map was developed through combination of molecular data coming from the five mapping populations 'Minsoy' x 'Noir 1', 'Minsoy' x 'Archer', 'Archer' x 'Noir 1', 'Clark' x 'Harosoy', and A81-356022 x PI468916. The 20 chromosomes represent 2,523.6 cM and contain a total of 1,849 markers among SSRs, RFLPs, RAPDs, AFLPs, morphological traits, isoenzymes, and others (Song et al., 2004).

The nomenclature for 11 linkage groups was designated by Zou et al. (2003). According to Soybase, the remaining nine linkage groups were assigned chromosome numbers in descending order depending on the linkage group genetic length (Table 1).

QTLs for soybean agronomic traits

Several studies have been carried out using molecular markers to identify QTLs associated with the expression of quantitative traits in soybean. Unfortunately, many of these QTLs have proved to be not reproducible across populations with diverse genetic backgrounds. For this reason, it is essential to find QTLs that can be mapped in different populations and, thus, be potentially used as effective tools for marker-assistance selection. This section of the literature review will present some studies on QTL identification for soybean agronomic traits.

Zhang et al. (2004) conducted a study in which 184 recombinant inbred lines (RILs) derived from the cross Kefeng No.1 x Nannong were screened with RFLPs, SSRs, and expressed sequence tags (ESTs) for 10 agronomic traits including days to flowering and maturity, plant

height, number of nodes on main stem, lodging, number of pods per node, protein content, 100-seed weight, and yield. A total of 63 QTLs were found for the traits studied. These QTLs were mapped to 12 chromosomes and accounted for 4.2 to 27.5% of the phenotypic variation; however, most of them were clustered on chromosomes 6 and 11. Additionally, pleiotropism was observed as some QTLs were mapped to the same loci. It was reported that one QTL could control up to five traits (Zhang et al., 2004).

In a different study, Hyten et al. (2004) developed a population consisting of 131 F_6-derived lines derived from the cross Essex x Williams, with the objective of identifying QTLs associated with protein seed content, oil seed content, and seed size. One hundred polymorphic markers were used to screen the population. Four QTLs for protein were found on chromosomes 6 (between Satt277 and Satt202), 13 (between Satt335 and Satt144), 9 (between Satt539 and Satt102), and 7 (between Satt540 and Satt463). These markers accounted for 13.6 to 27.6% of the phenotypic variation. Six QTLs were found to be associated with oil content and were mapped to chromosomes 6 (between Satt277 and Satt460), 1 (between Satt184 and Satt179), 17 (between Satt458 and Satt154), 19 (between Satt166 and Dt1 and between Satt229 and Satt373), and 7 (between Satt540 and Satt463). All of these markers explained between 6.0 and 20.0% of the phenotypic variation.

Panthee et al. (2007) used SSRs to screen 101 F_6-derived RILs from a population developed from the cross N87-984-16 x TN93-99 for five agronomic traits; yield, lodging, plant height, seed filling period, and maturity. Results indicated that there was one QTL for yield (close to Satt076 on chromosome 19), two for lodging (close to Satt225 and Satt593 on chromosome 5), and four for maturity (close to Satt263, Satt293, Satt292, and Satt591 on chromosomes 15, 12, 20, and 5, respectively). The QTLs for lodging explained from 20.6 to

32.7% of phenotypic variation while QTLs for maturity accounted for 12.8 to 25.7% of the phenotypic variation observed. The QTL reported for yield explained 13.8% of the phenotypic variation.

Moreover, Chen et al. (2007) used a population of 154 $F_{2:10}$ RILs derived from the cross Charleston x Dongnong to construct a linkage map for 12 agronomic traits. The progeny was screened with 164 SSRs and the traits evaluated included protein content (P), oil content (O) and P+O contents, pods per plant, seed weight per plant, 100-seed weight, plant height, days to maturity, branches, nod number in main stem, average leaf length, and average leaf width. They reported 68 QTLs for all traits evaluated. Number of QTLs per trait ranged from 3 to 11.

In a more recent study, Xie et al. (2012) used 125 RILs of $F_{5:7}$, $F_{5:8}$, and $F_{5:9}$ generations derived from the cross Hefeng 25 and Dongnong L-5, to identify QTLs associated with linolenic acid content and polyunsaturated fatty acids content. The progeny was screened with 112 SSRs in order to construct a genetic linkage map. Results indicated four QTLs for palmitic acid that were mapped to chromosome 17 (near Satt389), 16 (near Sat_144), 18 (near Sat_164), and 9 (near Satt727) which explained between 2.5 and 15.2% of the phenotypic variation. One QTL mapped to chromosome 2 (close to Satt701) was identified to be associated with stearic acid content and accounted for approximately 13.5% of the phenotypic variation. Four QTLs associated with linoleic acid content were found; one on chromosome 18 (near Sat_164), one chromosome 17 (near Satt389), one on chromosome 16 (near Sat_144), and one chromosome 14 (near Satt726). These markers explained between 3.2 and 27.5% of the phenotypic variation. Six QTLs for linolenic acid content, four of them were mapped to chromosome 14 (close to Satt729, Fad3a-4, Fad3b-1, and Fad3bc-1), one to chromosome 2 (close to Satt701), and one to chromosome 18 (close to Sat_164). These markers accounted for 5.3 - 37.3% of the phenotypic

variation. Additionally, four QTLs were identified for oleic content and mapped to chromosomes 9 (near Satt544), 16 (near Sat_144), 17 (near Satt389), and 18 (near Sat_164). These markers explained between 5.8 and 20.7% of the phenotypic variation for the trait.

Lately, increasing interest in the potential health benefits of soybean consumption has prompted the development of studies to identify QTLs for other seed components. Thus, Primomo et al. (2005) conducted a study with the objective of identifying QTLs associated with individual and total isoflavone contents. In order to do this, a population of 207 $F_{4:6}$ RILs was developed from the cross AC756 x RCAT Angora. A total of 99 SSR polymorphic markers were used to construct the genetic linkage map. Results revealed four QTLs associated with daidzein: Satt538, Satt181, Satt201, and Satt245. The first and second markers were located on chromosomes 8 and 12 while the last two were on chromosome 7; these markers accounted for 4.2 to 17.6% of the phenotypic variation. Four QTLs were found for genistein content: Satt289 on chromosome 6, Satt181 on LG 12, and Satt201 and Satt245 on chromosome 7.These markers explained between 5.2 and 31.4% of the phenotypic variation. Five QTLs were identified to be associated with glycitein content: Satt129 on LG1, Satt547 on LG 16, Satt196 on chromosome 9, and Satt521 on LG 7. These markers accounted for 4.2 to 9.3% of phenotypic variation. Additionally, four QTLs were found to be associated with total isoflavone content and were located close to the markers Satt289 on chromosome 6, Satt181 on chromosome 12, and Satt201and Satt245 on chromosome 7 and explained between 3.9 and 25.5% of the phenotypic variation.

In a different study, Li et al. (2010) developed an F_5-derived F6 RILs population with 144 individuals from the cross OAC Bayfield x Hefeng 25. The first parent is a soybean cultivar which is well known for having high vitamin E content while Hefeng 25 has a low content. The

genetic linkage map was constructed with 107 polymorphic SSR markers. Results showed four QTLs associated with α-tocopherol content; QαB2_1 (Sat_177, LG14), QαC2_1 (Satt376, LG 6), QαD1b_1 (Satt266, LG 2), and QαI_1 (Satt440, LG20). These markers explained between 4.3 and 16.7% of the phenotypic variation. Eight QTLs were reported to be associated with γ-tocopherol, QγC1_1 (Satt 565, LG 4), QγC2_1 (Satt286, LG 6), QγG_1 (Satt199, LG 18), QγD1b-1 (Satt266, LG 2), and QγO-1 (Satt576, LG 10), and accounted for 2.8 to 13.0% of the phenotypic variation. Four QTLs were found to be associated with δ-tocopherol content, QδA2_1 (Sat_383, LG 8), QδD1a_1 (Satt179, LG 1), QδF_1 (Sat_262, LG 13), and QδI_1 (Satt354, LG 20). These markers explained 4.2 to 10.2% of the phenotypic variation. Additionally, five QTLs were found to be associated with total vitamin E content, QTVEC2_1 (Satt376, LG6), QTVEC2-2 (Satt286, LG6), QTVE D1b_1 (Satt172, LG2), QTVEN_1 (Sat_125, LG3), and QTVEO_1 (Satt592, LG10). These markers accounted for 2.9 to 10.9% of the phenotypic variation.

QTLs associated with main soybean seed carbohydrates

Sucrose

Maughan et al. (2000) used RFLP, SSR, and RAPD markers to identify and characterize QTLs controlling sucrose content in an interspecific F_2 soybean population derived from the cross V71-370 x PI 407162. Seventeen markers mapping to seven chromosomes (5, 7, 8, 13, 15, 19, and 20) were significantly associated with sucrose content. Individual markers explained between 6.1 and 12.4% of the total sucrose variation while combined they explained 53% of total phenotypic variation. The major markers accounting for most of the variation were located on chromosome 20 (Table 2).

Later, Feng et al. (2005) developed a population derived from the cross MFS-591 x Camp and used SSR markers to identify different sugars. They reported a QTL for sucrose on chromosome 18, which accounted for 12.9% of the variation in sucrose content (Table 2). Similarly, Kim et al. (2005; 2006b) used SSR markers to screen two populations with a common parent and reported six additional QTLs for sucrose content on chromosomes 2, 11, 12, 16, and 19, that explained between 3.6 and 17.3% of the phenotypic variation (Table 2). More recently, Zeng et al. (2014) used SSRs and SNPs markers on a population derived from MFS-553 x PI 243545 and identified three QTLs for sucrose on chromosomes 5, 9, and 16 that explained 46, 10, and 8% of phenotypic variation, respectively (Table 2).

Stachyose

Sebastian et al. (2000) produced soybean germplasm with low raffinose and stachyose content by using three different approaches. The first approach was to conduct germplasm surveys which required the screening of soybean germplasm collections, the second approach included induction of mutations conferring the desired traits, and the third approach consisted of recombination of the major genes obtained through the two previous approaches. The line LR28 was identified in germplasm collection for having low raffinose and stachyose content. Sebastian et al. (2000) developed F_2 populations derived from crosses between four elite lines and LR28. The segregation ratio for the four populations fit the genetic model of a single recessive gene, which was then designated as *stc1a*.

In the same study, Sebastian et al. (2000) performed an allelism test between LR28 and the low raffinose and stachyose line LR484, developed by mutagenesis. Results indicated the existence of a single recessive gene *stc1b* in LR484, which conferred low raffinose saccharide content and was allelic to *stc1a* in LR28.

Furthermore, Sebastian et al. (2000) evaluated the raffinose saccharide content of line LR33 and compared it with wild-type soybeans. The initial tests showed very little differences; however, after further research, LR33 appeared to be segregating for a mutation that eliminated a high percentage of the raffinose saccharide content. This new mutation, called *mips,* was obtained by chemical mutagenesis. In those soybean lines with the *mips* genotype, the stachyose level was reduced to 5 μmol g^{-1} of seed dry weight and raffinose was reduced to 10 μmol g^{-1}, as compared with 75 μmol g^{-1} of stachyose, and 20 μmol g^{-1} of raffinose in conventional *Mips* lines. The combination of the *mips* gene and *stc1x* genes had additive effects which resulted in a dramatic reduction of raffinose and stachyose content.

Feng et al. (2005) reported the existence of one QTL for stachyose on chromosome 10 that explained 18.8% of the phenotypic variation observed. Additionally, Skoneczka et al. (2009) developed populations from PI 87013 x PI200508 and PI 243545 x PI200508 and reported a QTL for stachyose on chromosome 6 of PI200508 that explained up to 94% of the phenotypic variance. Further study of the genomic region suggested that the low stachyose content in PI200508 was due to a 3 bp deletion in the *galactosyltransferase* gene. This mutation reduces the activity of the enzyme with the same name, which is also involved in the stachyose synthesis pathway.

Furthermore, Maupin et al. (2011) reported that the SSR marker Satt453 on chromosome 11 had 87% selection efficiency in a marker-assisted selection of low stachyose lines (Table 3). Additionally, Jaureguy (2009) identified three SSR markers on chromosome 10 which were linked to stachyose content. He also reported a QTL for this trait flanked between Satt262 and Sat_282 that explained 48% of the stachyose variation observed (Table 3). More recently, Zeng et al. (2012) used SSR and SNP markers on a population derived from the cross Osage x V99-

5089 and identified two QTLs for stachyose on chromosomes 10 and 11, which explained 11 and 81% of the phenotypic variation, respectively. These QTLs were stable across environments (Table 3).

There have been a few studies that aimed to identify the possible correlations between sucrose and stachyose and between these carbohydrates and other soybean traits. Results indicated that sucrose content is positively associated with yield (Cicek et al., 2006); however, no relationship could be established between this last trait and stachyose content (Cicek et al., 2006: Neus et al., 2005). Additionally, a negative correlation between sucrose and stachyose content has been reported (Neus et al., 2005; Hymowitz et al., 1972).

The previously described trait associations seem to indicate that the selection of soybean lines with high sucrose contents could also lead to a concurrent increase in oil content. However, stachyose and protein seed contents may be diminished by these breeding efforts.

Research Justification and Objectives

The soyfoods industry has grown in the last few decades due to the increased demand for these products, resulting from public awareness about their nutritional and health benefits. High sucrose content is a desirable trait because it affects the quality and taste of soyfoods. However, low stachyose content is desired since its intake causes flatulence and diarrhea in monogastric animals. For these reasons, the development of food-grade soybean lines with modified sugar profiles is essential in order to satisfy market demands.

The development of QTL and molecular markers has speeded up the selection process in breeding for certain traits. However, few QTLs and markers have been identified for sucrose and stachyose content. Additionally, several of these markers have proven irreproducible. This

certainly becomes an obstacle when they need to be used to screen populations with different genetic backgrounds.

The overall objective of this research is to investigate the genes/QTLs associated with high-sucrose and low-stachyose content in soybean seed. To address this objective, this research will be divided into three areas of study. The objective of the first area of study is to perform an allelism test in order to confirm the genotype segregation model in a soybean population derived from a low-by-low-stachyose cross, which previously showed the existence of two different recessive genes controlling the low-stachyose content in soybean seed. The second objective is to identify and/or confirm QTLs/genes for stachyose content. Lastly, the third objective is to identify and/or confirm QTLs/genes for sucrose content.

Implications

The two main possible benefits/products of this research are: (1) new reproducible breeding tools that allow the rapid selection of soybean lines with desired sugar profile and (2) confirmed genetic sources of high-sucrose and low-stachyose content that can be used in breeding food-grade soybeans for specialty markets.

References

Abe, J., D.H. Xu, Y. Suzuki, and A. Kanazawa. 2003. Soybean germplasm pools in Asia revealed by nuclear SSRs. Theor. Appl. Genet. 106:445–453.

American Soybean Association. 2014. SoyStats: A reference guide to important soybean facts and Figs. American Soybean Association, St. Louis, MO. http://www.SoyStats.com/2013 (accessed March, 2014).

Anderson, J.W., B.M. Johnstone, and M.E. Cook-Newell. 1995. Meta-analysis of the effects of soy protein intake on serum lipids. N. Engl. J. Med. 333:276-282.

Anderson, R.L. and W.J. Wolf. 1995. Compositional changes in trypsin inhibitors, phytic acid, saponins and isoflavones related to soybean processing. J. Nutr.125:581S-588S.

Arslanoglu, F., S. Aytac, and E.K. Oner. 2011. Effect of genotype and environment interaction on oil and protein content of soybean (*Glycine max* (L.) Merrill) seed. AJB. 10:18409-18417.

ASA [American Soybean Association]. 2012. Soy Stats 2011. Available online at http://www.soystats.com. (verified January 4, 2013).

Burkholder, J., B. Libra, P. Weyer, S. Heathcote, D. Kolpin, and P. Thorne. 2007. Impacts of waste from concentrated animal feeding operations on water quality. Environ. Health Perspect. 115:308-312.

Burkholder, P.R. and I. McVeigh. 1945. Vitamin content of some mature and germinated legume seeds. Plant Physiol. 20:301-306.

Carrão-Panizzi, M.C. and S.Z. Erhan 2007.Environmental and genetic variation of soybean tocopherol content under Brazilian growing conditions. JAOCS**.** 84:921-928.

Carrera, C.S., C.M. Reynoso, G.J. Funes, M.J. Martinez, J. Dardanelli, and S.L. Resnik. 2011. Amino acid composition of soybean seeds as affected by climatic variables. Pesq. Agropec. Bras., Brasilia 46:1579-1587.

Chen, P., G.R., Buss, and K.C. Diehl. 1993. Physical and chemical characteristics associated with hardness of small-seeded soybean for natto. ASA-CSSA-SSSA. International Annual Meetings. Nov. 7-12, Cincinnati, Ohio.

Chen, Q.S., Z.C. Zhang, C.Y. Liu, D.W. Xin, D.P. Shan, H.M. Qiu, and C.Y. Shan. 2007. QTL analysis of major agronomic traits in soybean. Scientia Agricultura Sinica. 40:41-47.

Cicek, M.S., P. Chen, M.A. Saghai-Maroof, and G.R. Buss. 2006. Interrelationships among agronomic and seed quality traits in an interspecific soybean recombinant inbred population. Crop Sci. 46:1253-1259.

Circle, S.J., E.W. Meyer, and R.W. Whitney. 1964. Rheology of soy protein dispersions. Effect of heat and other factors on gelation. Cereal Chem. 41:157-172.

Cober, E.R., J.A. Frégeau-Reid, L.N. Pietrzak, A.R. McElroy, and H.D. Voldeng. 1997. Genotype and environmental effects on natto soybean quality traits. Crop Sci. 37:1151-1154.

Dardanelli, J.L., M. Balzarini, M.J. Martinez, M. Cuniberti, S. Resnik, S.F. Ramunda, R. Herrero, and H. Baigorri. 2006. Soybean maturity groups, environments, and their interaction define mega-environments for seed composition in Argentina. Crop Sci. 46:1939-1947.

De Vries, B.D., W.R. Fehr, G.A. Welke, and R.E. Dewey. 2011. Molecular analysis of mutant alleles for elevated palmitate concentration in soybean. Crop Sci. 51:2554-2560.

Erickson, E.A., J.R. Wilcox, and J.F. Cavins. 1988. Inheritance of altered palmitic acid percentage in two soybean mutants. J. Hered.79:465-468.

Falconer, D.S. 1989. Introduction to quantitative genetics. Longman, New York, 3rd edition. pp. 438.

Fehr, W.R., C.E. Caviness, D.T. Burmood, and J.S. Pennington. 1971. Stage of development descriptions for soybeans, *Glycine max* (L.).Merrill. Crop Sci. 11:929-931.

Feng, C., M. Morsy, E. Giannoccaro, B. Zhang, and P. Chen. 2005. Soybean seed sugar content

and quantitative trait loci mapping. p. 438-439. In C.J. Li et al. (ed.) Plant nutrition for food security, human health, and environmental protection. Tsinghua University Press, Beijing, China.

Fujita, Y., M. Iki, J. Tamaki, K. Kouda, A. Yura, E. Kadowaki, Y. Sato, J.S. Moon, K. Tomiola, N. Okamoto, and N. Kurumatani. 2012. Association between vitamin K intake from fermented soybeans, natto, and bone mineral density in elderly Japanese men: the Fujiwara-kyo Osteoporosis risk in men (Formen) study. Osteoporosis Int. 23:705-714.

Gao, J., X. Hao, K.D. Thelen, and G.P. Robertson. 2009. Agronomic management system and precipitation effects on soybean oil and fatty acid profiles. Crop Sci. 49: 1049-1057.

Geater, C.W. and W.R. Fehr. 2000. Association of total sugar content with other seed traits of diverse soybean cultivars. Crop Sci. 40:1552-1555.

Geater, C.W., W.R. Fehr, and L.A. Wilson. 2000. Association of soybean seed traits with physical properties of natto. Crop Sci. 40:1529-1534.

Gitzelmann, R. and S. Auricchio. 1965. The handling of soya α-galactosidase by a normal and a galactosemic child. Pediatrics 36:231-235.

Guzman, G. and P. Murphy. 1986. Tocopherols of soybean seeds and soybean curd (tofu). J. Agric. Food Chem. 34:791-795.

Hajduch, M., A. Ganapathy, J. W. Stein, and J. J. Thelen. 2005. A systematic proteomic study of seed filling in soybean. Plant Physiol. 137:1397-1419.

Hartwig, E.E., T.M. Kuo, and M.M. Kenty. 1997. Seed protein and its relationship to soluble sugars in soybean. Crop Sci. 37:770-773.

Hawton, J.D., L.J. Johnston, T.M. Salzer, and G.C. Shurson. 1996. Applications of nutritionally improved corn and soybeans for swine. p. 263-304. In Proc. 57th Minnesota Nutrition Conf., University of Minnesota St. Paul, MN.

He, F.-J. and J.-Q. Chen. 2013. Consumption of soybean, soy foods, soy isoflavones and breast

cancer incidence: Differences between Chinese women and women in Western countries and possible mechanisms. Food Science and Human Wellness 2:146-161.

Hill, J.E. and R.W. Breidenbach. 1974a. Proteins of soybeans seeds: I. Isolation and characterization of the major components. Plant Physiol. 53:742-746.

Hill, J.E. and R.W. Breidenbach. 1974b. Protein of soybeans seeds: II. Accumulation of the major protein components during seed development and maturation. Plant Physiol. 53:747-751.

Horan, F.E. Proc. Int. Conf. on Soybean Protein Foods, Peoria, IL, ARS-71-35.U.S.Department of Agriculture, Washington D. C. October.129-141.

Hunter, J.E., J. Zhang, and P.M. Kris-Etherton. 2010. Cardiovascular disease risk of dietary stearic acid compared with *trans*, other saturated, and unsaturated fatty acids: a systematic review. Am. J. Clin. Nutr. 91:46-63.

Hymowitz, T. 1970. On the domestication of the soybean. Econ. Bot. 24:408-421.

Hymowitz, T., F.I. Collins, J. Panczner, and W.M. Walker. 1972. Relationship between the content of oil, protein, and sugar in soybean seed. Agron. J. 64:613-616.

International Union of Pure and Applied Chemistry (IUPAC). 1995.Glossary of class names of organic compounds and reactivity intermediates based on structure. IUPAC Recommendations.doi:10.1351/pac199567081307 Compendium of Chemical Terminology - the Gold Book (verified January 3, 2013).

Jauhar, P.P. 2006. Modern biotechnology as an integral supplement to conventional plant breeding: The prospects and challenges. Crop Sci. 1841-1859.

Jaureguy, L.M. 2009. Identification of molecular markers associated with seed size, protein, and sugar content in soybean. M.S. Thesis, University of Arkansas, Fayetteville, AR.

Jenkins, D.J., A. Mirrahimi, K. Srichaikul, C.E. Berryman, L. Wang, A. Carleton, S. Abdulnour, J.L. Sievenpiper, C.W. Kendall, and P.M. Kris-Etherton. 2010. Soy protein reduces

serum cholesterol by both intrinsic and food displacement mechanisms. J. Nutr. 140:2302S-2311S.

Johnson, L.A., D.J. Myers, and D.J. Burden. 1992. Soy protein's history, prospects in food, feed. Intl. New on Fats, Oils and Related Materials (INFORM), AOCS Press 3:429-444.

Kada, S., M. Yabusaki, T. Kaga, H. Ashida, an K.-I. Yoshida. 2008. Identification of two major ammonia-releasing reactions involved in secondary natto fermentation. Biosci. Biotechnol. Biochem. 72:1869-1876.

Kim, H.K., S.T Kang, J.H. Cho, M.G. Choung, and D.Y. Suh. 2005. Quantitative trait loci associated oligosaccharide and sucrose contents in soybean (*Glycine max* L.). J. Plant Biol. 48:106-112.

Kim S.-L., M.A. Berhow, J.-T. Kim, H.-Y. Chi, S.-J. Lee, and I.-M. Chung. 2006a. Evaluation of soyasaponin, isoflavone, protein, lipid, and free sugar accumulation in developing soybean seeds. J. Agric. Food Chem. 54:10003-10010.

Kim H.K., S.T. Kang, and K.W. Oh. 2006b. Mapping of putative quantitative trait loci controlling the total oligosaccharide and sucrose content of *Glycine max* seeds. J. Plant Res. 119:533-538.

Kim, T.W., J.-H. Lee, M.-H. Park, and H.-Y. Kim. 2010. Analysis of bacterial and fungal communities in Japanese- and Chinese-fermented soybean pastes using nested PCR-DGGE. Curr. Microbiol. 60:315-320.

Kitamura, K., T. Kumagai, and A. Kikuchi. 1985. Inheritance of lipoxygenase-2 and genetic-relationships among genes for lipoxygenase-1, lipoxygenase-2 and lipoxygenase-3 isozymes in soybean seeds. Jpn. J. Breed. 35:413-420

Kris-Etherton, P.M. and S.M. Yu. 1997. Individual fatty acid effects on plasma lipids and lipoproteins: Human studies. Am. J. Clin. Nutr. 65:S1628-S1644.

Kuriyama, S. and L.B. Mendel. 1917. The physiological behavior of raffinose. J. Biol. Chem. 31:125-147.

Lebiedzinska, A. and P. Szefer. 2006. Vitamins B in grain and cereal–grain food, soy-products

and seeds. Food Chem. 95:116-122.

Leshem, Y.Y. 1984. Interaction of cytokinins with lipid-associated oxy free radicals during senescence: a prospective mode of cytokinin action. Can. J. Bot. 62:2943-2949.

Li, H., H. Liu, Y. Han, X. Wu, W. Teng, G. Liu, and W. Li. 2010. Identifications of QTLs underlying vitamin E contents in soybean seed among multiple environments. Theor. Appl. Genet. 120:1405-1413.

Liu, K.S. 2005. Soybeans as a powerhouse of nutrients and phytochemicals. p. 1-22. In: Soybeans as functional foods and ingredients. AOCS Press, Champaign, IL.

Liu, K.S. 1997. Chemistry and nutritional value of soybean components. p. 25-113. In: Soybeans: Chemistry, technology, and utilization. Chapman & Hall, New York, NY.

Maestri, D.M., D.O. Labuckas, C.A. Guzman, and L.M. Giorda. 1998. Correlation between seed size, protein and oil contents, and fatty acid composition in soybean genotypes. Grasas y Aceites. 49(5-6):450-453.

Maughan P.J., M.A. Saghai-Maroof, and G.R. Buss. 2000. Identification of quantitative trait loci controlling sucrose content in soybean *(Glycine max)*. Mol. Breeding 6:105-111.

Marega, F.M., D. Destro, L.A. Miranda, W.A. Spinosa, M.C. Carrão-Panizzi, and R. Montalvan. Relationships among oil content, protein content and seed size in soybeans. 2001. Brazilian Archives of Biology and Technology. 44:23-32.

Maupin, L.M., M.L. Rosso, and K.M. Rainey. 2011. Environmental effects on soybean with modified phosphorus and sugar composition. Crop Sci. 51:642-650.

Melchinger, A.E. 1998. Advances in the analysis of data on quantitative trait loci. In: Proceedings 2nd International Crop Science Congress. New Delhi, India.

Meydani, S.N., M. Meydani, J.B. Blumberg, L.S. Leka, G. Siber, R. Loszewski, C. Thompson, M.C. Pedrosa, R.D. Diamond, and B.D. Stollar. 1997. Vitamin E supplementation

and *in vivo* immune response in healthy elderly subjects. A randomized controlled trial. J. Am. Med. Assoc. 277:1380-1386.

Meydani, S.N., M.P. Barklund, S. Liu, M. Meydani, R.A. Miller, J.G. Cannon, F.D. Morrow, R. Rocklin, and J.B. Blumberg. 1990. Vitamin E supplementation enhances cell-mediated immunity in healthy elderly subjects. Am. J. Clin. Nutr. 52:557-563.

Meyer, E.W. 1966. Proc. Int. Conf. on Soybean Protein Foods, Peoria, IL, ARS-71-95.U.S. Department of Agriculture, Washington D. C. October. 142-154 pp.

Min, X. 2008. Seed chemical composition change during development and post-harvest processing of vegetable soybean. M.S. Thesis. Food Sciences Department. University of Arkansas.

Molina, E., A. Papadopoulou, and D.A. Ledward. 2001. Emulsifying properties of high pressure treated soy protein isolate and 7S and 11S globulins. Food Hydrocolloids 15:263-269.

Mullin, W.J., and W. Xu. 2001. Study of soybean seed coat components and their relationship to water absorption. J. Agric. Food Chem. 49:5331-5335.

Murphy, P., T. Song, G. Buseman, K. Barua, G. Beecher, D. Trainer, and J. Holden. 1999. Isoflavones in retail and institutional soy foods. J. of Ag. and Food Chem. 47:2697-2704.

Narvel, J.M., W.R. Fehr, J. Ininda, G.A. Welke, E.G. Hammond, D.N. Duvick, and S.R. Cianzio. 2000. Inheritance of elevated palmitate in soybean seed oil. Crop Sci. 40:635–639.

Nash, A.M and W.J. Wolf. 1967. Solubility and ultracentrifugal studies on soybean globulins. Cereal Chem. 44:183-192.

Neus, J.D., W.R. Fehr, and S.R. Schnebly. 2005. Agronomic and seed characteristics of soybean with reduced raffinose and stachyose. Crop Sci. 45:589-592.

Obendorf, R.L., A.D. Zimmerman, Q. Zhang, A. Castillo, S.M. Kosina, E.G. Bryant, E.M. Sensenig, J. Wu, and S.R. Schnebly. 2009. Accumulation of soluble carbohydrates

during seed development and maturation of low-raffinose, low-stachyose soybean. Crop Sci. 49:329-341.

Oltmans, S.E., W.R. Fehr, G.A. Welke, V. Raboy, K.L. Peterson. 2005. Agronomic and seed traits of soybean lines with low-phytate phosphorus. Crop Sci. 45:593-598.

Packer, L. and S. Landvik. 1989. Vitamin E: Introduction to biochemistry and health benefits. Annals of The New York Academy of Sciences. 570:1-6.

Panthee, D.R., V.R. Pantalone, A.M. Saxton, D.R. West, and C.E. Sams. 2007. Quantitative trait loci for agronomic traits in soybean. Plant Breeding 126:51-57.

Primomo, V.S., V. Poysa, G.R. Ablett, C.J. Jackson, M. Gijzen, and I. Rajcan. 2005. Mapping QTL for individual and total isoflavone content in soybean seeds. Crop Sci. 45:2454-2464.

Pryor, W.A. 2000. Vitamin E and heart disease: basic science to clinical intervention trials. Free Radical Biology & Medicine. 28:141-164.

Raboy, V., K.A. Young, J.A. Dorsch, and A. Cook. 2001. Genetics and breeding of seed phosphorus and phytic acid. Plant Physiol. 158:489-497.

Rackis, J.J. 1974. Biological and physiological factors in soybeans. J. Am. Oil Chem. Soc. 51:161A-174A**.**

Rackis, J.J., D.H. Honig, D.J. Sessa, and H.A. Moser. 1972. Lipoxygenase and peroxidase activities of soybeans as related to the flavor profile during maturation. Cereal Chem. 49:586-597.

Rackis, J.J. 1981. Comparison of the food value of immature, mature, and germinated soybeans. p. 183-212. In: R. Teranidhi and H. Barrera-Benitez (ed.) Quality of selected fruits and vegetables of North America. A.C.S. Washington, D.C.

Rafalski, J.A., S.V. Tingey, and J.G.K. Williams. 1991. RAPD markers - a new technology for genetic mapping and plant breeding. AgBiotech. News and Information.

Ren, C., K.D. Bilyeu, and P.R. Beuselinck. 2009. Composition, vigor, and proteome of mature soybean seeds developed under high temperature. 49:1010-1022.

Rimm, E.B., W.C. Willett, F.B. Hu, L. Sampson, G.A. Colditz, J.E. Manson, C. Hennekens, and M.J. Stampfer. 1998. Folate and Vitamin B_6 from diet and supplements in relation to risk of coronary heart disease among women. J. Am. Med. Assoc. 279:359-364.

Roberts, M.B.V. 1986. Biology: A Functional Approach, 4th edition. Trans - Atlantic Publications, Incorporated. Pp. 704.

Sacks, F.M., A. Lichtenstein, L. Van Horn, W. Harris, P. Kris-Etherton, and M. Winston. 2006. Soy protein, isoflavones, and cardiovascular health: An American Heart Association Science advisory for professionals from the nutrition committee. Circulation 113:1034-1044.

Saio, K. 1976. Soybean resistance to water absorption. Cereal Foods World. 21:168-173.

Saghai-Maroof, M.A. and G.R. Buss. 2008 Low phytic acid, low stachyose, high sucrose soybean lines. U.S. Patent 12/033,830, filed Feb 19, 2008.

Saldivar, X., Y.J. Wang, P. Chen, and A. Hou. 2011. Changes in chemical composition during soybean seed development. Food Chem. 124:1369-1375.

SAS Institute. 2013. SAS/STAT Software Version 12.3. SAS Institute Inc, Cary, NC.

Schaefer, M.J. and J. Love. 1992. Relationships between soybean components and tofu texture. J. of Food Quality 15:53-66.

Schnebly S.R., W.R. Fehr, G.A. Welke, E.G. Hammond, and D.N. Duvick. 1994. Inheritance of reduced and elevated palmitate in mutant lines of soybean. Crop Sci. 34:829–833.

Sebastian, S.A., P.S. Kerr, R.W. Pearlstein, and W.D. Hitz. 2000. Soybean germplasm with novel genes for improved digestibility. p. 56-74. J.K. Drackely (ed.). In: Soy in animal nutrition. Federation of animal Sci. Soc., Savoy, IL.

Skoneczka, J.A., M.A. Saghai-Maroof, C. Shang, and G.R. Buss. 2009. Identification of

candidate gene mutation associated with low stachyose phenotype in soybean line PI200508. Crop Sci. 49:249-255.

Sharpley, A.N., S.C. Chapra, R. Wedepohl, J.T. Sims, T.C. Daniel, and K.R. Reddy. 1994. Managing agricultural phosphorus for protection of surface waters: Issues and options. J. Environ. Qual. 23:437-451.

Shen, N., W.R. Fehr, L. Johnson, and P. White. 1997. Oxidative stabilities of soybean oils with elevated palmitate and reduced linolenate contents. J. Am. Oil Chem. Soc. 74:299–302.

Shin, J.H., Van K., Kim K.D., Lee Y.H., Jun T.H., Lee S.H. 2012.Molecular sequence variations of the *lipoxygenase-2* gene in soybean. Theor. Appl. Genet. 124:613-22.

Shurtleff, W. and A. Aoyagi. 2013. History of soymilk and other non-dairy milks (1226 to 2013). Extensively annotated bibliography and sourcebook. Soyinfo Center, Lafayette, California. http://www.soyinfocenter.com/pdf/166/Milk.pdf (Accessed March 1, 2014)

Shurtleff, W. and A. Aoyagi. 2011. History of tempeh and tempeh products (1811-2011): Extensively annotated bibliography and sourcebook. Soyinfo Center, Lafayette, California. http://www.soyinfocenter.com/pdf/148/Temp.pdf (Accessed March 1, 2014)

Sleeper, D.A., and J.M. Poehlman. 2006. Quantitative inheritance in plant breeding. p. 53-70. *In:* Breeding field crops. Fifth edition, Blackwell Publishing. Ames, Iowa.

Soyatech Inc. and SPINS, Co. 2012. Soyfoods: The U.S. Market. Available online at http://www.soyatech.com/store.htm?catId=21 (verified January 4, 2013).

Soybase and the soybean breeder's tool box. Agricultural Research Service. U.S. Department of agriculture. Iowa State University. http://soybase.org (verified February 5, 2014)

Sugiyama, S. 1984. Selection of micro-organisms for use in the fermentation of soy sauce. Food Microbiol. 1:339-347.

Taira, H. 1990. Quality of soybeans for processed foods in Japan. Japan Agric. Res. Quarterly 24:224-230.

Ubbinik, J.B., W.J. Hayward Vermaak, A. van der Merwe, and P.J. Becker. 1993. Vitamin B-12, vitamin B-6, and folate nutritional status in men with hyperhomocysteinemia. Am. J. Clin. Nutr. 57:47-53.

Van Etten, C.H., J.E. Hubbard, J.M. Mallan, A.K. Smith, and C.W. Blessing. 1954. Amino acid composition of soybean protein fractions. J. Agric. Food Chem. 7:129-131.

VanToai, T.T., J.D. Lee, P.F.P. Goulart, J.G. Shannon, J.D. Alves, H.T. Nguyen, O. Yu, M. Rahman, and R. Islam. 2012. Soybean (*Glycine max* L. Merr.) seed composition response to soil flooding stress. J. Food, Agric. and Env. 10:795-804.

Veldink, G.A., F.G.J. Vliegenthart, and J. Boldingh. 1977. Plant lipoxygenases. Prog. Chem. Fats Other Lipids 15: 131-166.

Watson R.R. and T.K. Leonard. 1986. Selenium and vitamins A, E and C: nutrients with cancer prevention properties. J. Am. Diet. Assoc. 86:505-510.

Wei, Q., and K.C. Chang. 2004. Characteristics of fermented natto products as affected by soybean cultivars. J. of Food Proc. Pres. 28:251-273.

Wilcox, J.R., and R.M. Shibles. 2001. Interrelationships among seed quality attributes in soybean. Crop Sci. 41:11-14.

Wilson, R.F. 2004. Seed composition. p. 621–677. In: H.R. Boerma and J.E. Specht (ed.) Soybeans: Improvement, production, and uses. 3rd ed. Agron. Monogr. 16. ASA, CSSA, and SSSA, Madison, WI.

Wolf, W.J. 1970. Soybean proteins: their functional, chemical, and physical properties. J. Agr. Food Chem. 18:969-976.

Yagasaki, K., F. Kousaka, and K. Kitamura. 2000. Potential improvement of soymilk gelation properties by using soybeans with modified protein subunit compositions. Breed. Sci. 50:101-107.

Yanagisawa, Y., T. Akazawa, T. Abe, and T. Sasahara. 1997. Changes in free amino acid and Kjeldahl N concentrations in seeds from vegetable-type and grain-type soybean cultivars during cropping season. J. Agric. Food Chem. 45:1720-1724.

Yang, C.S., Suh N., and Kong A.T. 2012. Does vitamin E prevent or promote cancer?. Cancer Prev. Res. 5:701-705.

Yazdi-Samadi, B., R.W. Rinne, and R.D. Steif. 1977. Components of developing soybean seeds: Oil, protein, sugars, starch, organic acids, and amino acids. Agron. J. 69:481-486.

Yenofsky, R.L., M. Fine, and C. Liu. 1988. Isolation and characterization of a soybean *(Glycine max)* lipoxygenase-3 gene. Mol. Gen. Genet. 211:215-222.

Zardakas, C.G., Z. Yu, H.D. Voldeng, and A. Minero-Amador. 1993. Assessment of the protein quality of a new high-protein soybean cultivar by amino acid analysis. J. Agric. Food Chem. 41:616-623.

Zeng, A., P. Chen, A. Shi, D. Wang, B. Zhang, M. Orazaly, L. Florez-Palacios, K. Brye, Q. Song, and P. Cregan. 2014. Identification of quantitative trait loci for sucrose content in soybean seed. Crop Sci. 54:1-11.

Zeng, A. 2012. Identification of quantitative trait loci for sucrose and stachyose content in soybean seed. M.S. Thesis, University of Arkansas, Fayetteville, AR.

Zhang, B., M. Tamura, J. Berger-Doyle, and P. Chen. 2008. Comparison of instrumental methods for measuring hardness of food-grade soybean. J. Texture Studies 39:28-39.

Zou, J.J., R.J. Singh, J. Lee, S.J. Xu, P.B. Cregan, and T. Hymowitz. 2003. Assignment of molecular linkage groups to soybean chromosomes by primary trisomics. Theor. Appl. Genet. 107:745-750.

Table 1. Soybean Linkage Groups.

Chr. †	LG‡	cM ♯
1	D1a	98.41
2	D1b	140.63
3	N	99.51
4	C1	112.32
5	A1	86.75
6	C2	136.51
7	M	135.15
8	A2	146.67
9	K	99.60
10	O	132.89
11	B1	124.24
12	H	120.50
13	F	120.03
14	B2	108.18
15	E	99.88
16	J	92.27
17	D2	119.19
18	G	105.00
19	L	101.14

Table 1. Soybean Linkage Groups (Cont.).

Chr. †	LG‡	cM #
20	I	112.77

† Chromosome.
‡ Linkage group.
# Chromosome length in centimorgans.

Table 2. Molecular markers reportedly associated with soybean seed sucrose content.

Marker type	QTL name	Chr.[†]	Marker	cM[‡]	% V[§]	Parents	Reference
RFLP	Sucrose 1-1	5	A487_1	8.8	8.1	PI 407162 x V71-370	Maughan et al. (2000)
SSR	Sucrose 1-9	7	GMSC51	3.1	7.3	PI 407162 x V71-370	Maughan et al. (2000)
RFLP	Sucrose 1-2	8	A136_1	71.4	8.7	PI 407162 x V71-370	Maughan et al. (2000)
RFLP	Sucrose 1-3	8	T153_1	50.4	7.6	PI 407162 x V71-370	Maughan et al. (2000)
RFLP	Sucrose 1-5	13	A186_1	64.9	9.6	PI 407162 x V71-370	Maughan et al. (2000)
RFLP	Sucrose 1-10	15	A963_1	17.1	6.9	PI 407162 x V71-370	Maughan et al. (2000)
RFLP	Sucrose 1-6	19	A023_1	36.7	9.7	PI 407162 x V71-370	Maughan et al. (2000)
RFLP	Sucrose 1-7	19	B164_1	35.3	9.6	PI 407162 x V71-370	Maughan et al. (2000)
RFLP	Sucrose 1-8	19	B162_2	49.4	7.0	PI 407162 x V71-370	Maughan et al. (2000)
RFLP	Sucrose 1-4	20	A144_1	32.4	12.4	PI 407162 x V71-370	Maughan et al. (2000)
RFLP	Sucrose 1-13	8	A486_1	53.2	8.4	PI 407162 x V71-370	Maughan et al. (2000)
SSR	Unassigned	18	Satt324	33.3	12.9	MFS-591 x Camp	Feng et al. (2005)
SSR	Sucrose 2-2	2	Satt546	87.2	6.4	Keunolkong x Iksan10	Kim et al. (2005)

Table 2. Molecular markers reportedly associated with soybean seed sucrose content (Cont.).

Marker type	QTL name	Chr.[†]	Marker	cM[‡]	% V[§]	Parents	Reference
SSR	Sucrose 2-1	11	Satt197	46.4	3.6	Keunolkong x Iksan10	Kim et al. (2005)
SSR	Sucrose 2-3	19	Satt523	29.9	4.1	Keunolkong x Iksan10	Kim et al. (2005)
SSR	Sucrose 2-4	19	Satt278	31.2	17.3	Keunolkong x Iksan10	Kim et al. (2005)
SSR	Sucrose 3-4	12	Satt442	46.9	8.3	Shinpaldalkong x Keunolkong	Kim et al. (2006)
SSR	Sucrose 3-6	16	Sct_065	32.1	8.3	Shinpaldalkong x Keunolkong	Kim et al. (2006)
SNP	Suc1	5	ss245668753	125.1	46	MFS-553 x PI 243545	Zeng et al. (2014)
SNP	Suc2	9	ss246796276	76.1	10	MFS-553 x PI 243545	Zeng et al. (2014)
SNP	Suc3	16	ss249186914	277.3	8	MFS-553 x PI 243545	Zeng et al. (2014)

[†] Chromosome.
[‡] QTL position in centimorgans.
[§] Percentage of phenotypic variation explained by the marker.

Table 3. Molecular markers reportedly associated with soybean seed stachyose content.

Marker type	QTL name	Chr.†	Marker	cM‡	% V§	Parents	Reference
SSR	Unassigned	10	Satt347	42.3	18.8	MFS-591 x Camp	Feng et al. (2005)
SSR	Unassigned	10	Sat_282	59.4	48.0	MFL-552 x R95-1705	Jaureguy et al. (2009)
SSR	Unassigned	10	Satt173	53.4	37.0	MFL-552 x R95-1705	Jaureguy et al. (2009)
SSR	Unassigned	10	Satt262	54.1	40.0	MFL-552 x R95-1705	Jaureguy et al. (2009)
SSR	Unassigned	11	Satt453	108.4	28.0	V99-5089 x Essex	Maupin et al. (2011)
SNP	Sta1	10	Sat_282 – BARC-029531-06209¶	40.3	11.0	Osage x V99-5089	Zeng (2012)
SNP	Sta2	11	BARC-013547-01157 – BARC-018869-03031¶	6.0	79.0	Osage x V99-5089	Zeng (2012)

† Chromosome.
‡ QTL position in centimorgans.
§ Percentage of phenotypic variation explained by the marker.
¶ QTLs were located in the interval between these markers.

II. Identification of Quantitative Trait Loci / Genes for Sucrose Content in Soybean Seed

Abstract

Sucrose is the most abundant soluble carbohydrate in soybean seed. It affects the quality and taste of soyfoods and also increases the metabolizable energy in animal feed. Thus, breeding for high-sucrose soybean cultivars is driven by the demand for livestock feed and the development of food-grade soybeans. The objective of this study was to identify quantitative trait loci (QTL) for sucrose content using single nucleotide polymorphisms (SNP). An F_2-derived mapping population was developed from the cross between a regular sucrose line (V97-3000) and a high-sucrose line (V99-5089). A total of 92 $F_{2:3}$ lines were genotyped with 5361 SNP markers which covered the 20 soybean chromosomes and, from these, 1720 were polymorphic. Seed samples were collected at $F_{2:3}$, $F_{2:6,}$ and $F_{2:7}$ generations, and sucrose analysis was performed by using high performance liquid chromatography. Subsequently, linkage maps were constructed with JoinMap® software and composite interval mapping (CIM) was conducted to locate QTLs associated with sucrose content. Two QTLs for seed sucrose content were identified on chromosomes 6 and 10, accounting for 17 and 11% of the phenotypic variation observed for this trait. Their net contribution was 0.57 and 0.56% to sucrose increase, respectively. The sucrose QTL on chromosome 10 is located on a similar genetic region of another QTL previously reported, which suggests that these two QTL may be the same. The other sucrose QTL on chromosome 6 has not been previously reported, which indicates that this is a novel QTL. Molecular markers tightly linked to these QTLs could be used for marker-assisted selection in breeding soybean lines with high- sucrose profile.

Introduction

Consumption of soybeans has been part of the Asian diet for more than 1000 years. However, soyfood consumption in the United States has only become popular in the last few decades. This increase in demand has largely been due to the advertised health benefits associated with soybean consumption (Golbitz, 1995). Soyfoods can be classified as traditional non-fermented products or traditional fermented products. The first category is mainly dominated by tofu (soybean curd), soymilk (soybeans soaked, finely ground, and strained), and edamame (vegetable soybeans); while the second category include products such as natto (fermented whole soybean), miso (fermented soup-base paste), and tempeh (made of whole cooked soybeans) among others (Cui et al., 2004).

In general, soybean seed is comprised of 40% protein, 35% carbohydrates, 20% oil, and 5% ash on a dry-weight basis (Liu, 1997). Soluble carbohydrates account for up to 47% of the total carbohydrate content (Hymowitz and Collins, 1974). Main soluble sugars are, in order of abundance, sucrose, stachyose, and raffinose; the concentrations of these sugars range between 2.5 and 8.2%, 1.4 and 4.1%, and 0.1 and 0.9%, respectively (Hymowitz et al., 1972). Glucose and fructose are considered minor sugars because they account for less than 1% (Hymowitz and Collins, 1974; Liu, 1997). Among the soluble sugars, sucrose is important in breeding for food-grade soybean, because it influences the sweetness of the products (Rackis, 1975).

Even though genetic variability has been reported for sucrose content in soybean germplasm (Hymowitz et al., 1972; Hou et al., 2009), its content is also affected by possible interrelationships with other seed quality traits. Thus, Wilcox and Shibles (2001) and Jaureguy (2009) proposed that seed protein and sucrose contents are negatively correlated, which increases the challenge of developing soybean breeding lines with high contents of both protein and

sucrose. Additionally, a negative correlation between sucrose and stachyose content has been reported (Hymowitz et al., 1972; Hitz et al., 2002; Florez-Palacios, 2009; Skoneczka et al., 2009). Although food-grade soybeans are required to have high sucrose content and low-oligosaccharide content, specific sugar profiles are needed based on the soy-based product that is being manufactured. For instance, high sucrose content is desired for both tofu and edamame production, whereas low-sucrose and high-stachyose contents are required for natto fermentation (Taira et al., 1990). Moreover, Taira et al. (1990) reported a negative association between sucrose content and the texture of tofu.

Furthermore, accumulation of soluble sugars increases with seed maturation. However, it has been reported that glucose, fructose, and sucrose contents decreased during the initial stages of seed formation. Subsequent to that, their concentrations remained unchanged for about three weeks until the seed began to mature, and then these sugars began to accumulate (Min 2008; Obendorf et al, 2009; Saldivar, 2011).

Measurement of sucrose content is a cumbersome task; apart from being costly, it is very labor intensive. For this reason, breeding programs aiming to develop lines with modified sucrose composition are focused on identifying molecular markers associated to this trait. Studies on inheritance of sucrose have revealed its polygenic nature which typical from a quantitative trait (Maughan et al., 2000; Kim et al., 2006; Zeng et al., 2014). Maughan et al. (2000) developed a breeding population derived from crossing the high sucrose content, large-seeded breeding line, V71-370, with the low sucrose content, small-seeded *G. soja* plant introduction, PI407162. The RLFP, RAPD, and SSR markers were used to identify and characterize QTLs controlling sucrose. Maughan et al. (2000) identified seven QTLs on chromosomes 5, 8, 20, 13, 19, 7, and 15 associated with the trait. The seven QTLs combined

explained 53% of total variation observed.

Feng et al. (2005) used SSR markers to identify QTLs associated with different soluble carbohydrates in a breeding population derived from the cross between MFS-591 (high sugar content) and the cultivar 'Camp' (low sugar content). As a result, one QTL for sucrose on chromosome 18 was reported, and it accounted for 12.9% of the variation in sucrose contents. Additionally, Kim et al. (2005) studied a population derived from crossing the cultivars 'Keunolkong' (early maturation, large-seeded) x 'Iksan10' (late maturation, small-seeded) and identified four other QTLs associated with sucrose content on chromosomes 11, 2, and two regions of chromosome 19. All four QTLs explained 31.4% of the phenotypic variation. A year later, Kim et al. (2006) developed a breeding population from the cross 'Keunolkong' x 'Shinpaldalkong', and reported two additional QTLs on chromosomes 12 and 16. Each of these QTL accounted for 8.3% of the phenotypic variation for sucrose.

Furthermore, Mozzoni (2009) reported two markers on chromosomes 5 and 8 which increased the sucrose content by 0.3 percent in a breeding population derived from crossing the soybean cultivar 'Ozark' with the breeding line 'V99-5089'. Additionally, Skoneczka et al. (2009) obtained two populations derived from the crosses PI87013 x PI200508 and PI243545 x PI200508, and identified a molecular marker on chromosome 6 that explained 76% of the phenotypic variation in sucrose content. More recently, Zeng et al. (2014) developed a population from the cross 'MFS-553' x 'PI 243545' and reported a major sucrose QTL on chromosome 5 which explained 46% of the phenotypic variation and two minor QTLs on chromosomes 9 and 16 that accounted for 10 and 8% of the trait variation, respectively.

The existence of a number of QTLs / genes controlling sucrose is expected due to the nature of the trait. As with any other quantitative character, sucrose is affected by several genes

with small additive effects, and several markers for sucrose content have been reported (Maughan et al., 2000; Feng et al., 2005; Kim et al., 2005; Kim et al., 2006; Jaureguy, 2009; Mozzoni, 2009; Skoneczka et al., 2009; Zeng et al., 2014). However, results from molecular studies are, in most cases, unrepeatable since the QTLs cannot be confirmed across populations with different genetic backgrounds and grown under different environmental conditions. This study seeks to find common markers that lead to their universal use, unique markers specific for this trait, and new markers that are different from the ones previously reported. Results of this research will provide an approach in looking for molecular markers linked to the sucrose variation.

Materials and methods

Population development and field experiment

An F_2 soybean population segregating for sucrose content was used in this study and both parental lines were developed at Virginia Polytechnic Institute. The regular sucrose (~5.0 %) breeding line, V97-3000, was crossed to the high sucrose (~7.7 %) breeding line, V99-5089, in summer of 2007, at the Arkansas Agricultural Research and Extension Center (AAREC) in Fayetteville, AR. The F_1 plants were grown in Fayetteville during summer of 2008, and morphological markers (leaf shape and seed size) were used to differentiate true hybrids from selfs and outcrosses. For instance, V97-3000 is a small-seeded line (9 g 100 seed^{-1}) with a narrow leaf shape, whereas V99-5089 is a large-seeded line (20 g 100 seed^{-1}) with a broad leaf shape. F_1 plants were then bulked and harvested.

The F_2 population was grown in Fayetteville in 2009, in 3-m rows with 150 seed per row. Soybeans were planted with 1-m rows. Plots were irrigated by furrow irrigation and managed

according to the standard cultural practices for soybean production in Arkansas (Tacker and Voires, 1998). A total of 150 random F_2 plants were tagged and identified with a number. At the end of the season, these tagged plants were harvested individually to form the mapping population. A sample containing 100 seed was sent to a winter nursery for generation advancement in 2009. The remnant of the seed from each $F_{2:3}$line was used for carbohydrate analysis.

The winter nursery is located near Upala city in the northwestern part of Costa Rica at 52 meters above sea level (Mongabay, 2014). There, each $F_{2:3}$ line was grown in a 3-m row with 0.76-m row spacing. A 3-m row spacing of the parental genotypes was planted next to the segregating population. Soil was cultivated before planting and plots were fully irrigated and managed during the growing season using standard cultural practices. After two generation advancements, $F_{2:5}$ seed was brought back to Fayetteville but, unfortunately, the seed quantity was not enough to grow a test with two replications in two locations. For this reason, it was decided to do a seed increase in Fayetteville during the growing season in 2010 and then to establish the test the following year.

In summer of 2011, $F_{2:6}$ plants were grown in a randomized complete block design (RBD) with two replications in two locations. Locations were Fayetteville and Marianna, AR. The soil in Fayetteville is classified as Captina silt loam soil (Fine-silty, siliceous, active, mesic Typic Fragiudults) (Soil Survey Staff, 2013), with very deep, moderately well drained soils, made in a thin mantle of silty material (Soil Series, 2006). The soil in Marianna is classified as a Calloway silt loam soil (Fine-silty, mixed, active, thermic Aquic Fraglossudalfs) (Soil Survey Staff, 2013), and is described as very deep, somewhat poorly drained soil that developed in thick loess or water reworked loess (Soil Series, 2002). Each soybean plot consisted of two 3-m rows,

each with 150 seed and a 1-m row spacing. Both locations were fully irrigated by furrow irrigation and managed based on the standard cultural practices for soybean production in Arkansas (Tacker and Voires, 1998). The youngest fully developed trifoliate of each plant was sampled, and samples belonging to the same row were bulked for DNA extraction. In the fall of 2010, $F_{2:6}$ seed were harvested, a 10-g sample of each plot was used for carbohydrate analysis and the remnant of the seed was saved to be planted the next season.

Genotypic data

Genomic DNA extraction. After collection, leaf tissue samples were stored in the freezer at -80 °C. Subsequently, they were ground with liquid N to a fine powder by using a mortar and a pestle. DNA was extracted using the cetyltrimethylammonium bromide (CTAB) buffer method (Kisha et al., 1997) in which, a buffer containing 5M NaCl, 200mM Tris pH 8.0, 4% (w/v) CTAB, 0.5 mM EDTA, and 6.4 ml β-mercaptoethanol is added to the samples, followed by chloroform:isoamyl alcohol (24:1). After incubation for 60 min at 65°C with occasional gentle mixing, DNA was precipitated, washed with 95% ethanol, and subsequently dissolved in 0.1 x TE buffer. Concentrations were calculated by measuring the absorbance at 260 nm using a BioTek Power Wave XS Microplate Spectrophotometer (BioTek, Winooski, VT). DNA was stored in the freezer at -80°C.

Single nucleotide polymorphism (SNP) genotyping. For genetic map construction, two replications of each parental line and 92 $F_{2:7}$ DNA samples were sent to the Research Technology Support Facility (RTSF) Genomics Core at Michigan State University, East Lansing, MI. There they were genotyped with 5361 SNP markers (dbSNP-NCBI, 2012) using the Illumina Infinium® Genotyping HD BeadChip (652k SNPs) on Illumina iScan (Illumina, San Diego, CA). SNP analysis was performed on 4-μL samples containing a concentration between

50 and 100 ng/µL of DNA. Intensities of the fluorescence were distinguished using the Illumina iScan TM Reader, and alleles for each SNP locus were named using Illumina's Bead Studio TM software (Illumina, San Diego, CA, v3.2.23). For each SNP marker, the genotype data represent three possible genotypes AA (homozygote), AB (heterozygote), and BB (homozygote) (Akond et al., 2013).

Phenotypic data

Soluble carbohydrate analysis. Sugar extraction followed the protocol described by Hou et al. (2009), with some modifications. Briefly, a 10-g seed sample was ground to a fine powder using a coffee bean grinder (Krups®, Shelton, CT). Then, the powder was sifted through a 100-µm stainless steel testing sieve (VWR®, West Chester, PA) in order to obtain a sample with uniform particle size. Afterwards, a 0.15-g sample was weighed, mixed with 1.5 mL of deionized-distilled water (ddH_2O), and transferred into a 2-mL centrifuge tube. Tube was vortexed, shaken horizontally for 20 min at 200 rpm and, centrifuged at 14000 rpm for 10 min. A 500 µL aliquot from the supernatant was placed in a new 2-mL centrifuge tube and 700 µL of acetonitrile (99.9% HPLC grade) (Thermo Fisher Scientific, Inc.) was added. The solution was mixed by inversion and incubated at room temperature for 30 min. Consequently, the tube was centrifuged at 14000 rpm for 10 min and 70 µL of the extract was pushed through a 25 mm Easy-Pressure syringe filter holder (VWR®, West Chester, PA) containing a 0.2 µm filter paper disc (Pall Lifesciences, East Hills, NY). Then, vials containing a 24-µL aliquot of each sample extract diluted in 576 µL of distilled water were used for carbohydrate determination in a HPLC.

The anion-exchange HPLC system (Dionex DX500 HPAEC-PAD) was composed of a GS50 gradient pump, an ED40 pulsed amperometric electrochemical detector, an LC50 chromatographic oven, an AS40 automated sampler with a 25-µL injection loop, and a

Chromeleon Chromatography Management Data System. The mobile phase consisted of a 90 mM NaOH solvent (VWR®, West Chester, PA) at constant flow rate of 1 mL min^{-1}, prepared by diluting carbonate-free HPLC grade 50% (w/w) stock solution in distilled water, filtered with a 0.45-µm membrane, and degassed with compressed nitrogen gas for 30 min before vials were loaded into the auto sampler, as described by Hou et al. (2009). Soluble carbohydrates were separated by an analytical CarboPac PA-10 pellicular anion-exchange resin column (4 x 250 mm) coupled to a CarboPac PA10 guard column (4 x 50 mm) and preceded by AminoTrap column (3 x 30 mm) (Dionex, Sunnyvale, CA). Sugar content was determined from regression curves fitted from a set of standards for sucrose (Sigma-Aldrich, St. Louis, MO) at different concentrations (10, 20, 40, 60, and 80 µg μL^{-1}). Carbohydrate data then converted to milligrams of sugar per gram of seed (mg g^{-1}) on dry weight basis.

Statistical analysis

The Shapiro-Wilk (*w*) normality test from JMP 10.0 (SAS Institute, Cary, NC) was conducted on the sucrose distribution data of the population at $F_{2:3}$, $F_{2:6}$, and $F_{2:7}$ generations. Broad sense heritability (H^2) for sucrose was determined as reported by Nyquist (1991), using the equation:

$$H^2 = \sigma^2_g / [\ \sigma^2_g + (\sigma^2_{gxe} / e) + (\sigma^2 / re)] \qquad [1]$$

Where, σ^2_g is the total genetic variance, σ^2_{gxe} is the genotype by environment interaction, σ^2 is the error variance, r is the number of replications, and e is the number of environments.

Associations between molecular markers and sucrose data were analyzed by single factor analysis of variance (ANOVA) at the 0.05 significant level with the PROC GLM procedure in SAS 12.3 (SAS Institute, 2013). The software JoinMap® 4.1 (Van Ooijen, 2006) was used to construct the linkage maps, and a minimum logarithm of odds (LOD) for linkage group

construction was set to 3.0. A Haldane mapping function (Haldane, 1919) was used to conduct regression mapping algorithm of each chromosome or linkage group. Subsequently, the software Windows QTL Cartographer 2.5 (Wang et al., 2012) and Qgene 4.0 (Joehanes et al., 2008) were used to combine marker positions with phenotypic data, and thus determine the location of the QTL on the linkage group. These programs were also used to perform composite interval mapping (CIM) and multiple interval mapping (MIM) analyses, in order to quantify additive and dominance effects for significant QTL and optimum position and interaction of the QTL, respectively. The empirical significance threshold for CIM was determined by one thousand permutations, with a walk speed of 1cM, and a significance level of $\alpha = 0.05$ (Zhang et al., 2008). MIM was performed based on the model $c(n) = \ln(n)$ with a walk speed of 1 cM. MapChart 2.2 (Voorrips, 2002) was used to create the LOD plots based on the data from JoinMap® 4.0 and Windows QTL Cartographer 2.5.

Results

Phenotypic data

The Shapiro-Wilk test showed that the seed sucrose content in the mapping population derived from V97-3000 x V99-5089 was normally distributed in three generations across all environments studied (Table 1a and Fig. 1). V99-5089 had 2.7 - 3.1% higher sucrose content than V97-3000, as expected, in all environments (location - year) studied (Table 1b and Fig. 1). Results from a t-test was conducted to compare the mean sucrose content in V99-5089 and V97-3000, and showed that V99-5089 (mean = 7.64%, SD = 0.74) had significantly higher sucrose content that V97-3000 (mean = 4.92%, SD = 1.03); $t(6) = 11.44$, $P < .0001$. Some of the lines in the mapping populations had sucrose content higher than the high sucrose parent and lower than

the low sucrose parent, indicating transgressive segregation. The normal distribution of the seed sucrose content and the transgressive segregants in the mapping population confirmed that sucrose is a quantitative trait controlled by multiple genes / QTLs.

ANOVA was conducted on 129 $F_{2:3}$ lines from the cross V97-3000 x V99-5089 and the parents, and means were separated by Least Significant Difference (LSD) at $P < 0.05$ (Table 2). The ANOVA model for seed sucrose was significant with R^2 value of 0.98, indicating that the model chosen for the experiment design and data analysis for this study was appropriate and adequate. Genotype, location, and the genotype x location interaction all were significant sources of variation for sucrose content. However, genotype accounted for the largest proportion of sucrose variation, while the effects of location and genotype x location were relatively small. Sucrose data for the individual lines were consistent among replications and locations, as indicated by non-significant replication effect in the ANOVA (Table 2) and by the similar ranking of the lines tested (data not shown). Based on the variance components for sucrose variation, seed sucrose was a highly heritable trait ($H^2 = 0.79$).

QTL mapping in F_2-derived populations by SNP markers

A total of 5361 random SNP markers covering the 20 soybean chromosomes were used to genotype $F_{2:6}$ lines and 1720 SNP loci (32%) were polymorphic (Table 3). These markers were mapped on 20 chromosomes, representing 663 unique SNP loci in the mapping population (Table 3 and Fig. 2 a-e.). The linkage map spanned 2435.2 cM with an average distance coverage of 3.8 cM per marker.

Results from the single marker analysis showed 10 significant SNP markers on four chromosomes (4, 5, 6, and 10) associated with seed sucrose content (Table 4). Lines with the homozygous genotype carrying the V99-5089 allele consistently had higher sucrose content

compared with those carrying the V97-3000 allele. On average, the V99-5089 allele accounted for an increase of sucrose content between 0.22 and 0.77%. The sucrose variation contributed by the SNP alleles in those four chromosome regions ranged from 2 to 17%. However, the SNP alleles on chromosomes 6 and 10 appeared to have more impact on sucrose content than the ones on chromosomes 4 and 5 (Table 5).

In the composite interval mapping analysis, the empirical significance threshold was computed as a LOD value of 2.5 in the $F_{2:6}$ mapping population in all environments studied; Fayetteville 2009, Fayetteville 2010, Fayetteville 2011 and Marianna 2011. Based on the average seed sucrose content across two locations over 3 years, a major sucrose QTL was identified on chromosome 6, flanked by the SNP markers ss2246102767 and ss2246107039 (Fig. 3a-c), tightly linked to the marker ss246109643 (Table 5 and Fig. 3a-c). This QTL was tentatively named *Suc1*. *Suc1* had a LOD value of 5.8, significantly higher than the threshold, and explained 17% of phenotypic variation for sucrose content on average (Table 5 and Fig. 3c). This QTL was consistently identified in four environments with a LOD value between 3.5 and 7.0 (Figs. 3a-b). Based on the average seed sucrose content across one location in 2 years, a minor sucrose QTL was mapped on chromosome 10 (Figs. 4a and 4b), located on the interval ss247300068 and ss247314578 (Fig. 2c), tightly linked to the marker SS247304924 (Table 5 and Figs. 4a and 4b). This minor sucrose QTL on chromosome 10 accounted for 11% of sucrose variation on average and was tentatively named as *Suc2*. The multiple interval mapping analysis showed the optimum locations of the two QTL identified. *Suc1* was determined to be at 6.7 cM on chromosome 6, while *Suc2* was located at 117.4 cM on chromosome 10.

Discussion

In this study, the high sucrose parent 'V99-5089' consistently produced high sucrose, while the low parent 'V97-3000' produced low sucrose, as expected, in all four environments consisting of two locations and three years (Fayetteville 2009, 2010, 2011, and Marianna 2011). Sucrose content of the F_2-derived lines from the cross V97-3000 x V99-5089 consistently exhibited a normal distribution across four environments. A small proportion of the lines had sucrose content exceeding the range of the low and high parents. The normal distribution of sucrose content and the presence of transgressive segregants in the mapping population confirmed that sucrose is a quantitative trait controlled by multiple genes/QTLs with small effect from the environment.

The average sucrose content for the parents and the F_2-derived lines were higher in Fayetteville in 2010 than in 2011 (Tables 1a and 1b). This can probably be explained by the difference in the average monthly temperature during the growing season for those two years (Table 6). Weather data showed that Fayetteville experienced higher temperatures in 2011 than in 2010. Previous research indicated that high temperatures decreased sucrose content by causing its degradation into more simple carbohydrates (Hou et al., 2009; Zeng et al., 2014). However, when comparing the average seed sucrose content for the parents and the F_2-derived lines for the two Arkansas locations (Fayetteville and Marianna) in 2011, this high-temperature-low-sucrose pattern was observed only for sucrose content of the parent 'V97-3000'. Sucrose level for the parent 'V99-5089' and the F_2-derived lines were slightly higher in Marianna with warmer temperatures (Tables 1a and 1b). This contradicting temperature-sucrose relationship agreed with the observation in a previous study where Ren et al. (2009) evaluated the effect of high temperature on seed composition and found that high temperature did not have a significant

effect on the content of sucrose, raffinose, and stachyose.

Results from the ANOVA showed significant differences among genotypes, locations, and genotype x location effects on sucrose content, which was expected because of the quantitative nature of the trait (Table 2). Therefore, it is important to evaluate genotypes in question at multiple locations and years for the phenotypic assessment of sucrose content. However, genotype accounted for most of the sucrose variation observed (Table 2). Average sucrose data from the two locations were highly correlated and most of the genotypes in the population ranked similarly between the two locations (data not shown). Significant genotypic and environmental effects on seed sucrose content have also been previously reported by other authors (Geater et al., 2000; Cicek et al., 2006; Zeng et al., 2014). The estimated heritability for sucrose content in this study (0.79) was similar to the values previously reported by others (Maughan et al., 2000; Kim et al., 2005; and Zeng et al., 2014), confirming that sucrose is a highly heritable trait.

In this study, the $F_{2:7}$ lines and SNP markers which were randomly selected from the National Center for Biotechnology Information (NCBI) were used to create a unique genetic map for seed sucrose content. It is considered 'unique' because a specific set of SNP markers was used to developed the DNA chip for the genotyping of the $F_{2:7}$ lines, which at this time, has not been reported in the public soybean linkage map (D. Wang, personal communication). For this reason, the comparison of the genetic positions of those SNP markers in relation to the public map is not practical or relevant at this time.

Single marker analysis showed three significant SNP markers (ss246109643, ss246056190, and ss246037023) in an approximately 18.5 cM region on chromosome 6 that are associated with sucrose content with 0.31 to 0.57% net contribution to sucrose (Table 5).

Composite interval mapping analysis with individual and combined data from four environments further confirmed the existence of a major sucrose QTL linked to ss246109643 with additional two flanking markers nearby ss246107039 and ss246102767. These three SNP markers are in a 2.2 cM region on chromosome 6 (Fig. 3a-c). However, ss246109643 at 6.7 cM accounted for 17% of the sucrose variation, while ss246037023 at 25.2 cM accounted for 12% of the sucrose variation (Table 5). It is possible that there are two QTLs in this region as a minor non-significant peak was obtained in the composite interval mapping analysis (Fig. 3a-c). This region deserves more attention in future research of sucrose QTL discovery and confirmation. Similarly, four significant SNP markers with 0.40 to 0.56% sucrose content contribution were identified on a 39.3 cM region on chromosome 10; ss247292289, ss247300068, ss247304924, and ss247330645 (Table 4). The composite interval mapping analysis showed a sucrose QTL linked to ss247304924 and near ss247300068 in a 5.9 cM region on chromosome 10. Evidently this chromosome region is rather large and the SNP markers identified accounted for relatively small percent of variation in sucrose as compared to the ones on chromosome 6. Further study is needed to confirm the minor sucrose QTL in this region and possibly find additional sucrose QTL in this region.

Although the single marker analysis indicated a large chromosome region on chromosome 6 and another large region on chromosome 10 that are associated with sucrose content, the composite interval mapping analysis clearly showed two QTLs, a major QTL on chromosome 6 and a minor QTL on chromosome 10. The major sucrose QTL is linked to ss246109643 and located at 48,100,279 bp on chromosome 6. The minor sucrose QTL is linked to ss247304924 and located at 45,826,662 bp on chromosome 10.

Although results from the composite interval analysis did not show QTLs on

chromosomes 4 or 5, the single marker analysis suggested a region on chromosome 4 and another region on chromosome 5 that are highly associated with sucrose content. SNP markers on chromosome 4 showed between 0.33 and 0.50% net contribution to sucrose, while a marker on chromosome 5 showed 0.54% net contribution to sucrose (Table 5). It is possible these two regions may contain QTLs for sucrose; therefore these regions deserve research attention in the future. While there is not a previous report of a sucrose QTL on chromosome 4, a couple of studies have identified markers linked to sucrose QTLs on chromosome 5 (Zeng et al., 2014; Maughan et al., 2000).

Previous research has identified 19 QTLs associated with seed sucrose content; these are located on 12 different soybean chromosomes. Only few sucrose QTLs have been mapped to the same chromosome, i.e. two QTLs were identified on chromosome 5, three QTLs on chromosome 8, two QTLs on chromosome 16, and four QTLs on chromosome 19 (Zeng et al., 2014; Kim et al., 2006; Maughan et al., 2000). Although, previous research done by Maughan et al. (2000) showed several QTLs associated with sucrose content, their results were obtained by single marker analysis, which is not an adequate analysis for QTL discovery because it only suggests association between the trait and the marker and does not account for the interaction effect of the marker alleles. Moreover, Kim et al. (2006) identified sucrose QTLs using data from only one location which does not provide enough information about the stability of the trait.

The QTL located on chromosome 6 that was identified in this study, is a novel sucrose QTL because no previous reference to this chromosomal region has been made. On the contrary, the other sucrose QTL identified in this study and located on chromosome 10 is in agreement with what was previously reported by Saghai-Maroof and Buss (2008). They indicated that chromosome 10 contained a major QTL for sucrose in V99-5089 close to the marker Satt453 (at

38,360,612 - 38,360,653 bp of Gm11 in the Williams 82 physical map). This marker is closely located upstream of the sucrose QTL found in my study (at 45,826,662 bp of Gm11, Williams 82).

Conclusion

Two sucrose QTLs were identified in this study. A new major sucrose QTL was located on chromosome 6 and linked to ss2246109643 with additional two flanking markers ss246107039 and ss246102767. This QTL explained 17% of the phenotypic variation for the trait and had an average net contribution of 0.57% to sucrose content. The physical proximity of these three SNP markers (2.2 cM) suggests that they all can be used for marker-assisted selection. Additionally, a minor sucrose QTL was identified on chromosome 10, linked to SS247304924 and near ss247300068. This QTL accounted for 11% of the sucrose variation observed and had an average net contribution of 0.56% to sucrose content. Stability of these QTLs across environments makes them reliable tools to be used in molecular breeding for soybeans with improved sucrose profile.

References

Akond, M., S. Liu, L. Schoener, J.A. Anderson, S.K. Kantartzi, K. Meksem, Q. Song, D. Wang, Z. Wen, D.A. Lightfoot, and M.A. Kassem. 2013. A SNP-based genetic linkage map of soybean using the SoyS-NP6K Illumina Infinium BeadChip genotyping array. Journal of Plant Genome Sciences 1:80-89.

Cicek, M.S., P. Chen, M.A. Saghai-Maroof, and G.R. Buss. 2006. Interrelationships among agronomic and seed quality traits in an interspecific soybean recombinant inbred population. Crop Sci. 46:1253-1259.

Cui, Z.L, A.T. James, S. Miyazaki, R.F. Wilson, and T.E. Carter, Jr. 2004.Breeding specialty soybeans for traditional and new soyfoods. p. 264-322, In: K.S. Liu (ed.) Soybeans as Functional Foods and Ingredients. AOCS Press. Champaign, Illinois, USA.

Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine. http://www.ncbi.nlm.nih.gov/snp (accessed February, 2014).

Feng, C., M. Morsy, E. Giannoccaro, B. Zhang, and P. Chen. 2005. Soybean seed sugar content and quantitative trait loci mapping. p. 438-439. In C.J. Li et al. (ed.) Plant nutrition for food security, human health, and environmental protection. Tsinghua University Press, Beijing, China.

Florez-Palacios, S.L. 2009. Genetic study of seed sugar content in diverse soybean germplasm. M.S. Thesis, University of Arkansas, Fayetteville, AR.

Geater, C.W., W.R. Fehr, and L.A. Wilson. 2000. Association of soybean seed traits with physical properties of natto. Crop. Sci. 40:1529-1534.

Golbitz P. 1995. Traditional soyfoods: Processing and products. J. Nutr. 125:570-572.

Haldane, J.B.S. 1919. The combination of linkage values, and the calculation of distance between the loci of linked factors. J. Genet. 8:299-309.

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 raffinosaccharide and phytic acid phenotype on soybean seed. Plant Physiology. 128:650-660.

Hou, A., P. Chen, J. Alloatti, D. Li, L. Mozzoni, B. Zhang, and A. Shi. 2009. Genetic variability of seed sugar content in worldwide soybean germplasm collections. Crop Sci. 49:903-912.

Hymowitz, T., and F.I. Collins. 1974. Variability of sugar content in seed of Glycine max (L.) Merrill and G. soja Sieb and Zucc. Agron. J. 66:239-240.

Hymowitz, T., F.I. Collins, J. Panczner, and W.M. Walker. 1972. Relationship between the content of oil, protein, and sugar in soybean seed. Agron. J. 64:613-616.

Jaureguy, L.M. 2009. Identification of molecular markers associated with seed size, protein, and sugar content in soybean. M.S. Thesis, University of Arkansas, Fayetteville, AR.

JMP®, version 10.0. 2012. SAS Institute Inc., Cary, NC.

Joehanes, R., and J.C. Nelson, 2008.QGene 4.0, an extensible Java QTL-analysis platform. Bioinformatics 24:2788-2789.

Kim, H.K., S.T. Kang, and K.W. Oh. 2006. Mapping of putative quantitative trait loci controlling the total oligosaccharide and sucrose content of Glycine max seed. J. Plant Res. 119:533-538.

Kim, H.K., S.T Kang, J.H. Cho, M.G. Choung, and D.Y. Suh. 2005. Quantitative trait loci associated oligosaccharide and sucrose contents in soybean (*Glycine max* L.). J. Plant Biol. 48:106-112.

Kisha, T., C.H. Sneller, and B.W. Diers. 1997. Relationship between genetic distance among parents and genetic variance in populations of soybean. Crop. Sci. 37:1317-1325.

Liu, K.S (ed.) 1997. Chemistry and nutritional value of soybean components. p. 25-113. In: Soybeans: Chemistry, technology, and utilization. Chapman & Hall, New York, NY.

Maughan P.J., M.A. Saghai-Maroof, and G.R. Buss. 2000. Identification of quantitative trait loci

controlling sucrose content in soybean *(Glycine max)*. Mol. Breeding 6:105-111.

Min, X. 2008.Seed chemical composition change during development and post-harvest processing of vegetable soybean. M.S. Thesis. Food Sciences Department. University of Arkansas.

Mongabay. 2014. http://population.mongabay.com/population/costa-rica/3621153/upala (accessed February 26, 2014).

Mozzoni, L.A. 2009. Quality attributes, processing, and selection of vegetable soybean. Doctoral dissertation, University of Arkansas, Fayetteville, AR.

Nyquist, W.E. 1991. Estimation of heritability and prediction of selection response in plant populations. Crit. Rev. Plant Sci. 10:235-322.

Obendorf, R.L. A.D. Zimmerman, Q. Zhang, A. Castillo, S.M. Kosina, E.G. Bryant, E.M. Sensenig, J. Wu, and S.R. Schnebly. 2009. Accumulation of soluble carbohydrates during seed development and maturation of low-raffinose, low-stachyose soybean. Crop Sci. 49:329-341.

Rackis, J.J. 1975. Oligosaccharides of food legumes: alpha-galactosidase activity and flatus problem, in physiological effects of food carbohydrates, Jeanes, A. and J. Hodge, Ed., pp. 207-222, American Chemical Society, Washington, D.C., USA.

Ren, C., K.D. Bilyeu, and P.R. Beuselinck. 2009. Composition, vigor, and proteome of mature soybean seeds developed under high temperature. Crop Sci. 49:1010-1022.

Saldivar, X., Y.J. Wang, P. Chen, and A. Hou. 2011. Changes in chemical composition during soybean seed development. Food Chem. 124:1369-1375.

SAS Institute. 2013. SAS/STAT Software version 12.3. SAS Institute Inc., Cary, NC.

Skoneczka, J.A., M.A. Saghai-Maroof, C. Shang, and G.R. Buss. 2009. Identification of candidate gene mutation associated with low stachyose phenotype in soybean line PI200508. Crop Sci. 49:247-255.

Soil Series. 2006. Captina Series. Rev. LBW/RLT. National Cooperative Soil Survey. U.S.A. https://soilseries.sc.egov.usda.gov/OSD_Docs/C/CAPTINA.html (accessed February 6, 2014).

Soil Series. 2002. Calloway Series. Rev. WAC:WMK:RBH:MAV. National Cooperative Soil Survey. U.S.A. https://soilseries.sc.egov.usda.gov/OSD_Docs/C/CALLOWAY.html (accessed February 5, 2014).

Soil Survey Staff. 2013. Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. http://websoilsurvey.nrcs.usda.gov (accessed February 5, 2014).

Tacker, P., and E. Voires. 1998. Irrigation, p. 42-49 In: Arkansas soybean handbook. University of Arkansas. Cooperative Extension Service.

Taira, H., H. Tanaka, and M. Saito. 1990. Effect of cultivar, seed size, and crop year on total and free sugar contents of domestic soybeans. J. Jap. Soc. Food Sci. & Tech. 37:203-213.

Van Ooijen, J.W. 2006. JoinMap® 4, software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, Netherlands.

Voorrips, R.E. 2002. MapChart: Software for the graphical presentation of linkage maps and QTLs. J. Hered. 93:77-78.

Wang Shengchu, C.J. Basten, and Z.-B. Zeng. 2012. Windows QTL Cartographer V2.5. Statistical Genetics & Bioinformatics. North Carolina State University. http://statgen.ncsu.edu/qtlcart/WQTLCart.htm (accessed April, 2013).

Wilcox, J.R., and R.M. Shibles. 2001. Interrelationships among seed quality attributes in soybean. Crop Sci. 41:11-14.

Zeng, A., P. Chen, A. Shi, D. Wang, B. Zhang, M. Orazaly, L. Florez-Palacios, K. Brye, Q. Song, and P. Cregan. 2014. Identification of quantitative trait loci for sucrose content in soybean seed. Crop. Sci. 54:1-11.

Zhang, B., P. Chen, C.Y. Chen, D. Wang, A. Shi, A. Hou, and T. Ishibashi. 2008 Quantitative trait loci mapping of seed hardness in soybean. Crop. Sci. 48:1341-1349.

Table 1a. Seed sucrose content (% dry-weight basis) of mapping populations derived from the cross V97-3000 x V99-5089 evaluated across two locations and over three years.

Year - Location	Generation	No. Lines	Mean	SD†	Range	Prob < W‡
2009 Fayetteville	$F_{2:3}$	129	5.84	1.66	3.27 - 9.47	0.975
2010 Fayetteville	$F_{2:6}$	92	5.81	1.93	2.71 - 9.76	0.983
2011 Fayetteville	$F_{2:7}$	92	5.55	1.64	3.20 - 8.91	0.981
2011 Marianna	$F_{2:7}$	92	5.64	1.72	3.03 - 9.18	0.979

† Standard deviation.
‡ Shapiro-Wilk normality test.

Table 1b. Seed sucrose content (% dry-weight basis) of parents V97-3000 and V99-5089 evaluated across two locations and over three years.

Year-Location	V97-3000		V99-5089	
	Mean	SD†	Mean	SD†
2009 Fayetteville	4.69	1.16	7.40	1.05
2010 Fayetteville	5.12	1.43	8.26	0.84
2011 Fayetteville	5.07	0.90	7.32	0.72
2011 Marianna	4.82	0.64	7.57	0.36
Overall‡	4.92	1.03	7.64	0.74

† Standard deviation.
‡ Sucrose mean and standard deviation across four environments.

Table 2. Analysis of variance for seed sucrose content of 92 $F_{2:7}$ population derived from the cross V97-3000 x V99-5089 grown in Fayetteville and Marianna in 2011.

Source	Degree of freedom[†]	Mean square	Variance components[‡]	Percent Variation[§]	P-value	R^2
Model	185	3.368	-	-	<0.0001	0.98
Location	1	0.893	0.004	0.260	<0.0001	
Replication (Location)	2	0.010	<0.001	0.001	0.5075	
Genotype	91	6.765	1.673	97.198	<0.0001	
Genotype x Location	91	0.073	0.029	1.707	<0.0001	
Error	182	0.014	1.721	0.835		

[†] Analysis of variance, using PROC GLM in SAS 12.3, of 92 $F_{2:7}$ lines grown in Fayetteville and Marianna, AR in 2011.
[‡] Estimate of variance components, as calculated using PROC VARCOMP in SAS 9.3.
[§] Percent of variation explained by each term in the ANOVA model.

Table 3. Summary of SNP markers used for the screening of an F_2-derived population from the cross V97-3000 x V99-5089.

Chr.[†]	LG[‡]	Length (cM)[§]	No. SNP mapped	No. SNP locus located	Average distance (cM) between SNP loci
1	D1a	29.8	23	15	2.0
2	D1b	167.0	49	26	6.4
3	N	23.8	99	23	1.0
4	C1	136.4	77	34	4.0
5	A1	82.4	56	14	5.9
6	C2	152.5	100	49	3.1
7	M	149.1	51	27	5.5
8	A2	135.6	90	51	2.6
9	K	23.1	90	14	1.6
10	O	139.5	59	32	4.4
11	B1	134.0	141	50	2.7
12	H	150.4	37	27	5.6
13	F	110.6	121	43	2.6
14	B2	48.0	33	16	3.0
15	E	127.1	157	62	2.1
16	J	118.2	90	34	3.5

Table 3. Summary of SNP markers used for the screening of an F_2-derived population from the cross V97-3000 x V99-5089 (Cont.).

Chr.[†]	LG[‡]	Length (cM)[§]	No. SNP mapped	No. SNP locus located	Average distance (cM) between SNP loci
17	D2	164.9	124	44	3.7
18	G	133.9	68	44	3
19	L	309.6	158	40	7.7
20	I	99.3	97	18	5.5
Average		121.8	86	33	3.8
Total		2435.2	1720	663	

[†]Chromosome.
[‡]Linkage group.
[§]Chromosome length in centimorgans.

Table 4. Single marker analysis of variance for seed sucrose content in 92 F_2-derived lines from the cross V97-3000 x V99-5089 evaluated in Fayetteville (Fay) in 2009 and 2010, and in Fayetteville and Marianna (Mar) in 2011.

Chr.[†]	SNP Marker	Position (cM)	P-value				
			2009	2010	2011		Combined[‡]
			Fay	Fay	Fay	Mar	
4	ss245330356	61.3	0.0818	0.0229	0.0482	0.0334	0.0382
4	ss245333397	62.4	0.0755	0.0234	0.0457	0.0327	0.0368
5	ss245693942	14.9	0.0033	0.0189	0.0123	0.0067	0.0077
6	ss246109643	6.7	0.0010	0.0001	0.0003	0.0001	0.0002
6	ss246056190	22.1	0.0547	0.0202	0.0194	0.0149	0.0216
6	ss246037023	25.2	0.0488	0.0379	0.0432	0.0297	0.0365
10	ss247292289	98.5	0.0358	0.0440	0.0718	0.0408	0.0425
10	ss247300068	111.5	0.0116	0.0091	0.0150	0.0156	0.0109
10	ss247304924	117.4	0.0089	0.0021	0.0100	0.0054	0.0048
10	ss247330645	137.8	0.0272	0.0109	0.0301	0.0212	0.0184

[†] Chromosome.
[‡] Pooled data from all environments studied.

Table 5. Mean effect of SNP marker alleles on seed sucrose content in 92 F_2-derived population developed from the cross V97-3000 x V99-5089 grown in two locations and over three years.

Chr.	SNP	Position (cM)	2009			2010			2011†			Combined‡			R^2
			P1§	P2¶	Diff.#	P1	P2	Diff.	P1	P2	Diff.	P1	P2	Diff.	
4	ss245330356	61.3	6.93	7.15	0.22	7.14	7.53	0.39	7.12	7.51	0.39	7.06	7.40	0.33	0.04
4	ss245333397	62.4	6.88	7.21	0.33	6.94	7.46	0.52	6.92	7.57	0.65	6.91	7.41	0.50	0.09
5	ss245693942	14.9	6.91	7.34	0.43	6.99	7.51	0.52	6.93	7.59	0.66	6.94	7.48	0.54	0.09
6	ss246109643	6.7	6.92	7.44	0.52	7.12	7.71	0.59	7.09	7.69	0.60	7.04	7.61	0.57	0.17
6	ss246056190	22.1	7.03	7.80	0.77	6.15	6.40	0.25	6.44	6.87	0.43	6.54	7.02	0.48	0.06
6	ss246037023	25.2	6.88	7.13	0.25	7.04	7.36	0.32	7.04	7.41	0.37	6.99	7.30	0.31	0.02
10	ss247292289	98.5	7.03	7.55	0.52	6.10	6.40	0.30	6.54	6.94	0.37	6.56	6.96	0.40	0.08
10	ss247300068	111.5	7.06	7.58	0.52	6.17	6.47	0.30	6.72	7.11	0.41	6.65	7.05	0.40	0.05
10	ss247304924	117.4	6.88	7.26	0.38	6.99	7.58	0.59	6.91	7.63	0.72	6.93	7.49	0.56	0.11
10	ss247330645	137.8	7.22	7.78	0.56	6.76	7.04	0.28	6.39	6.99	0.60	6.79	7.27	0.48	0.04

† Pooled data from Fayetteville and Marianna in 2011.
‡ Pooled data from Fayetteville in 2009 and 2010, and Fayetteville and Marianna in 2011.
§ Allelic effect of P1 = V97-3000.
¶ Allelic effect of P2 =V99-5089.
# Allelic difference.

Table 6. Average monthly temperatures (°C) in Fayetteville, AR in 2010 and Fayetteville and Marianna, AR in 2011.

Year	Location	Month											
		Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
2010	Fayetteville	0.6	0.6	7.8	15.0	19.4	25.6	26.1	27.8	21.7	14.4	8.9	2.2
2011	Fayetteville	0.6	3.3	10.0	15.0	17.8	25.6	28.9	27.8	18.3	14.4	10.0	4.4
	Marianna	3.3	5.6	11.7	18.3	21.1	27.2	28.3	27.2	21.1	15.6	12.2	7.2

Information extracted from http://www.wunderground.com (accessed March, 2014).

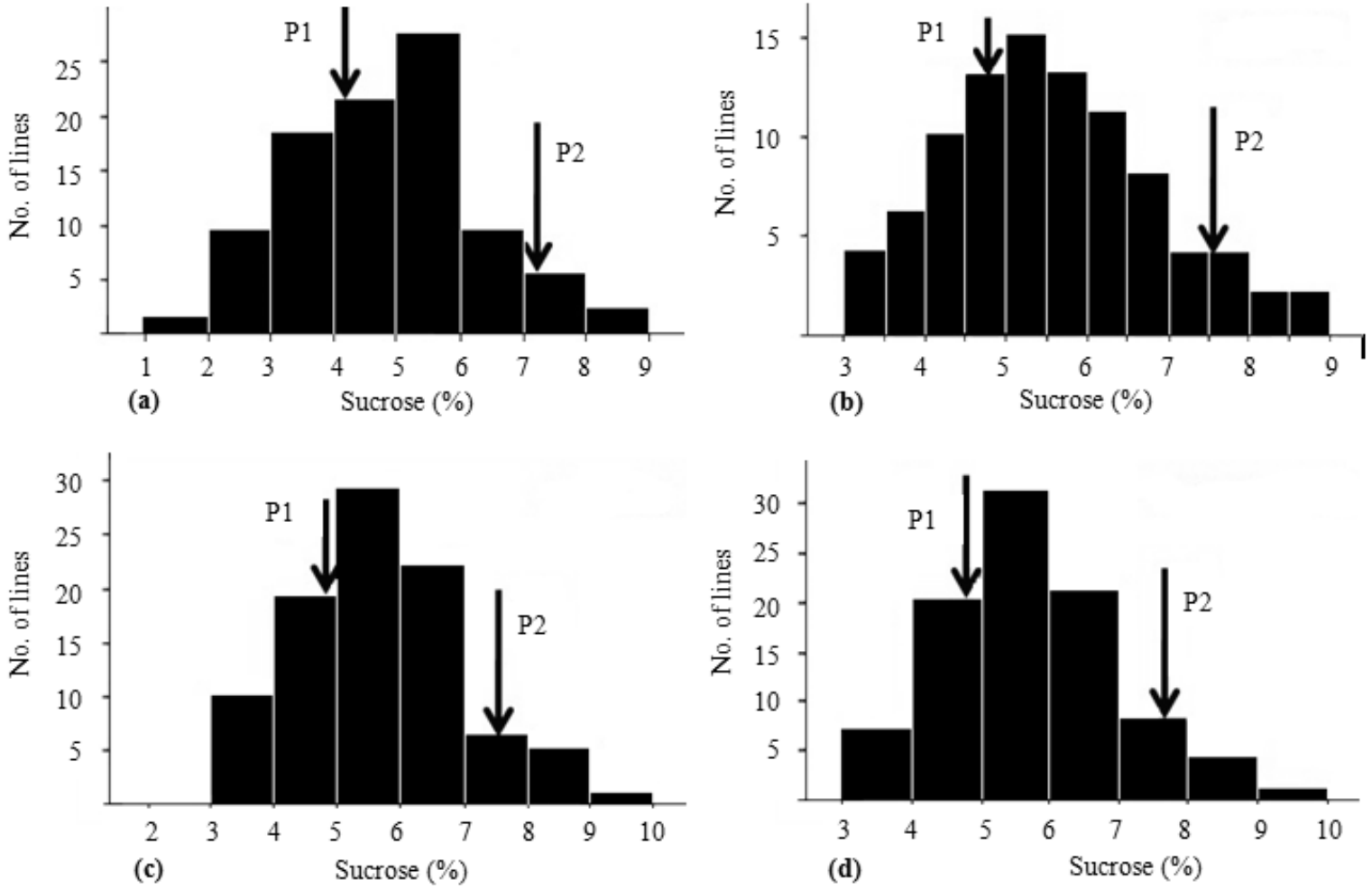

Fig. 1. Frequency distribution of seed sucrose content in populations derived from V97-3000 (P1) x V99-5089 (P2) evaluated in four environments: (a) $F_{2:6}$ lines in Fayetteville, AR in 2010; (b) $F_{2:7}$ lines in Fayetteville, AR in 2011; (c) $F_{2:7}$ lines Marianna, AR in 2011; (d) Combined data of $F_{2:7}$ lines in Fayetteville and Marianna, AR in 2011.

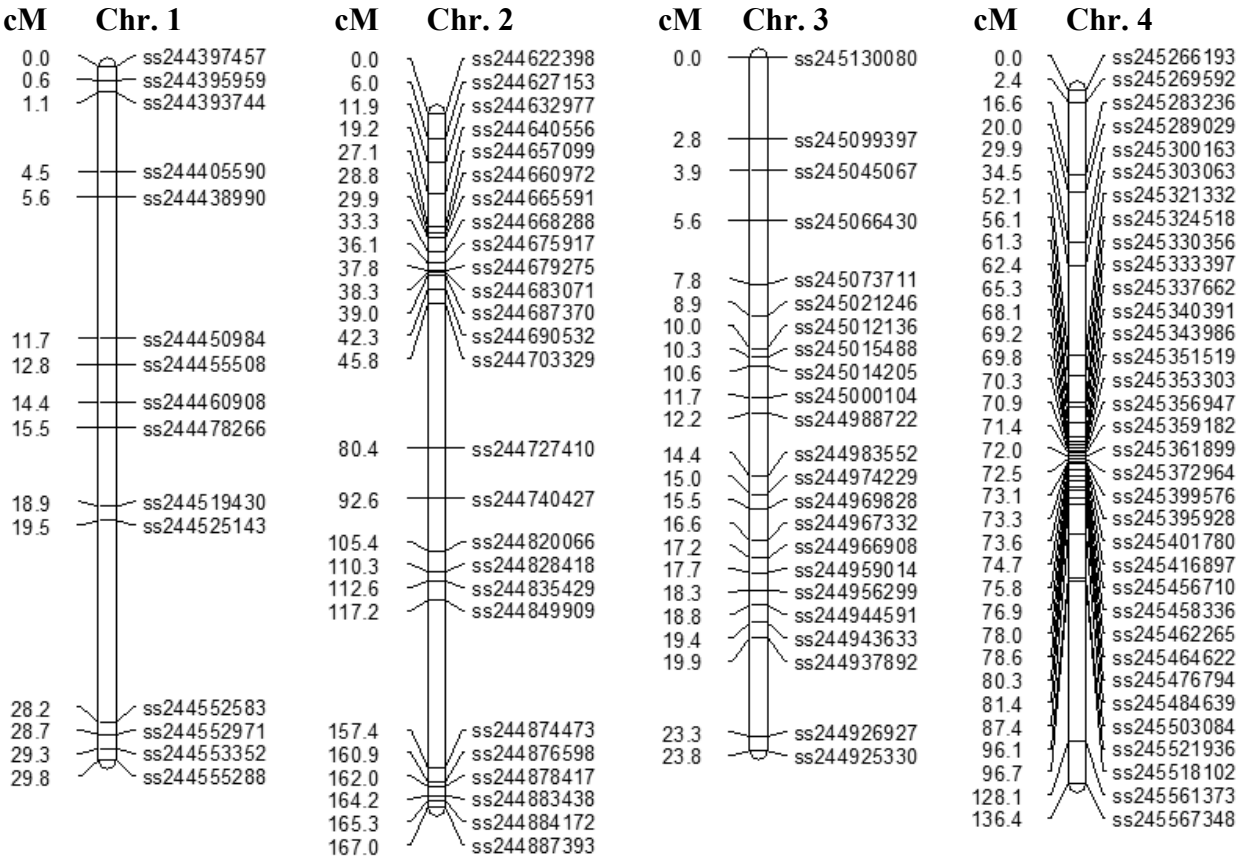

Fig. 2a. A genetic map constructed for chromosomes 1, 2, 3, and 4 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes.

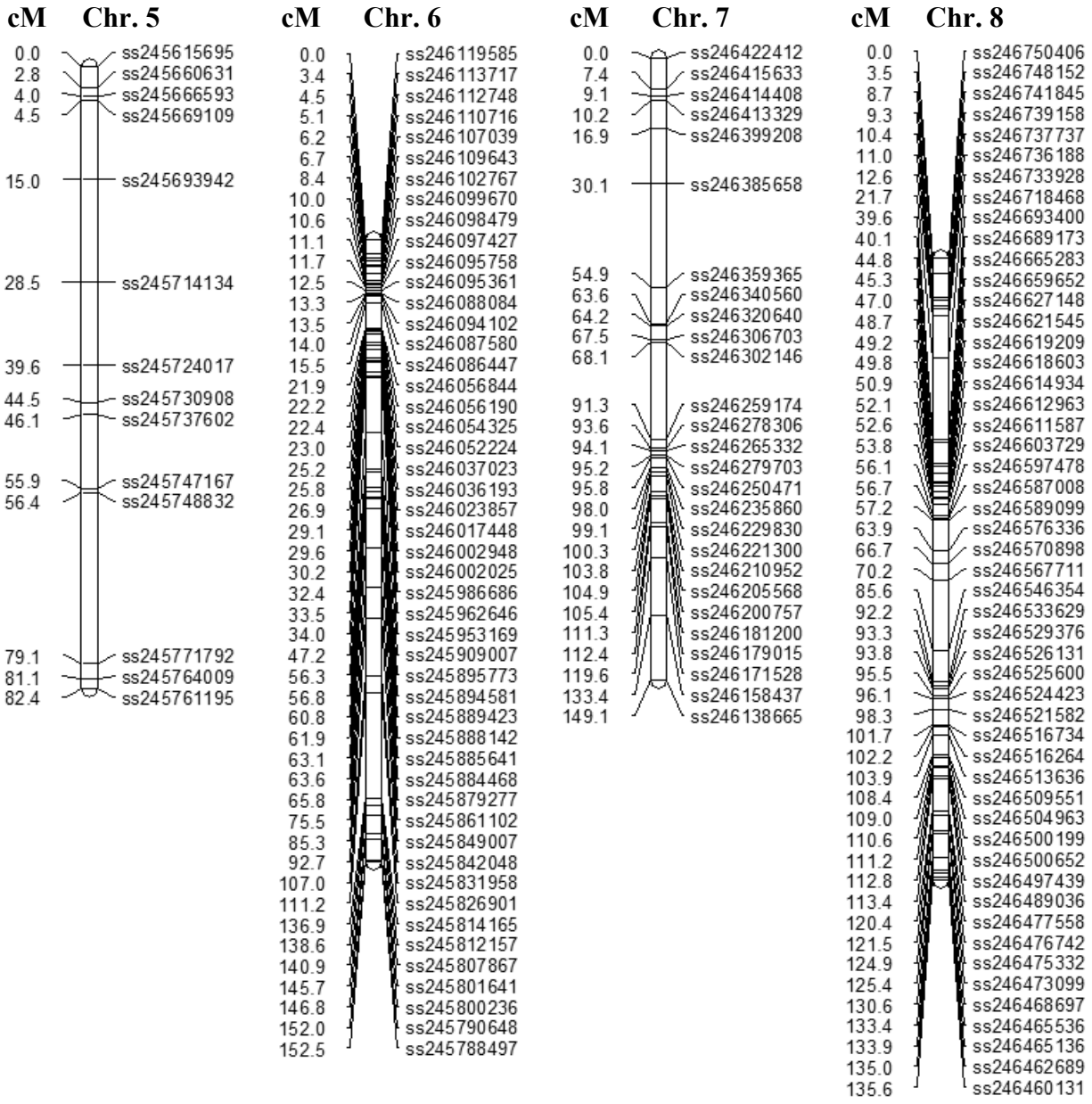

Fig. 2b. A genetic map constructed for chromosomes 5, 6, 7, and 8 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes.

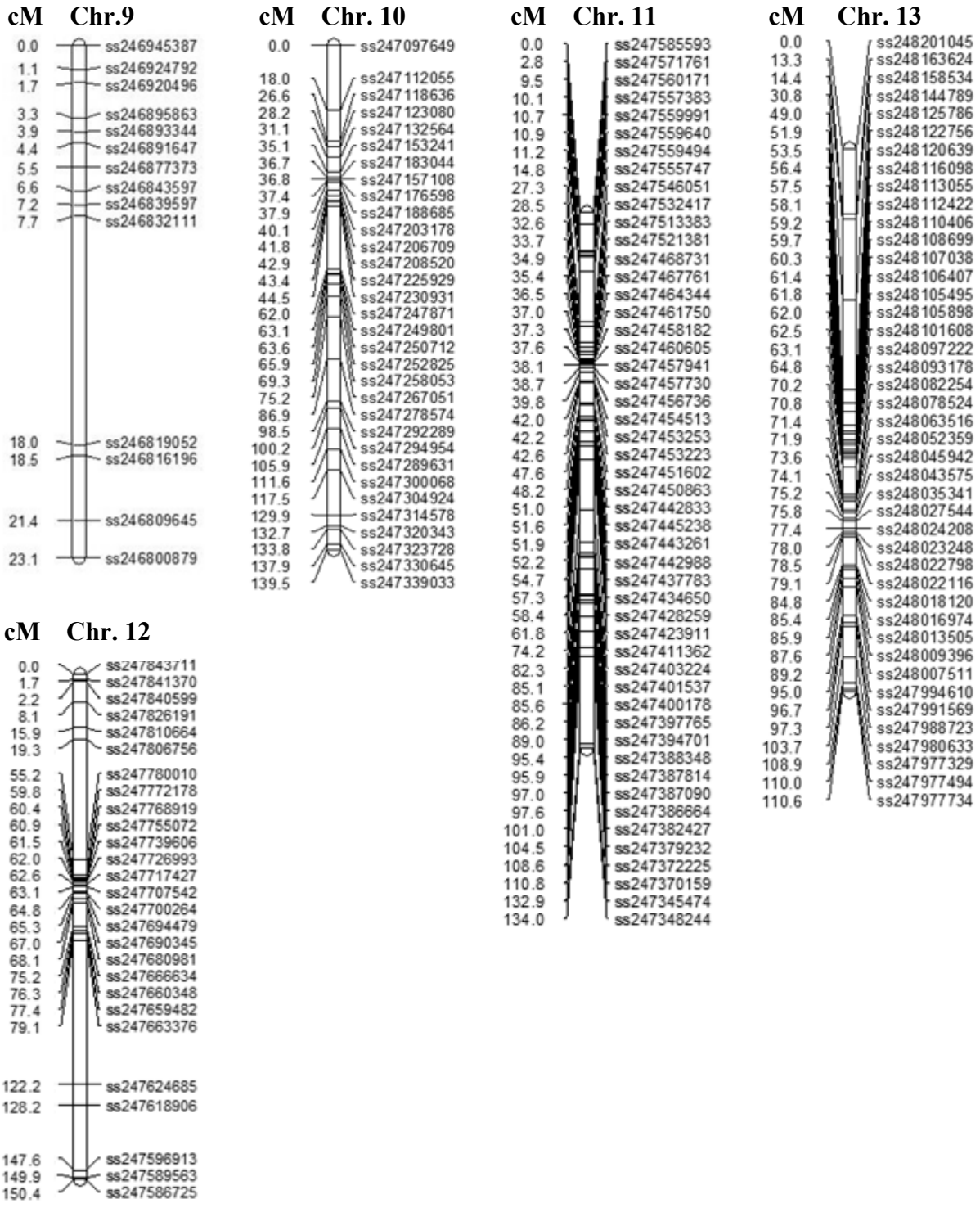

Fig. 2c. A genetic map constructed for chromosomes 9, 10, 11, 12, and 13 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes.

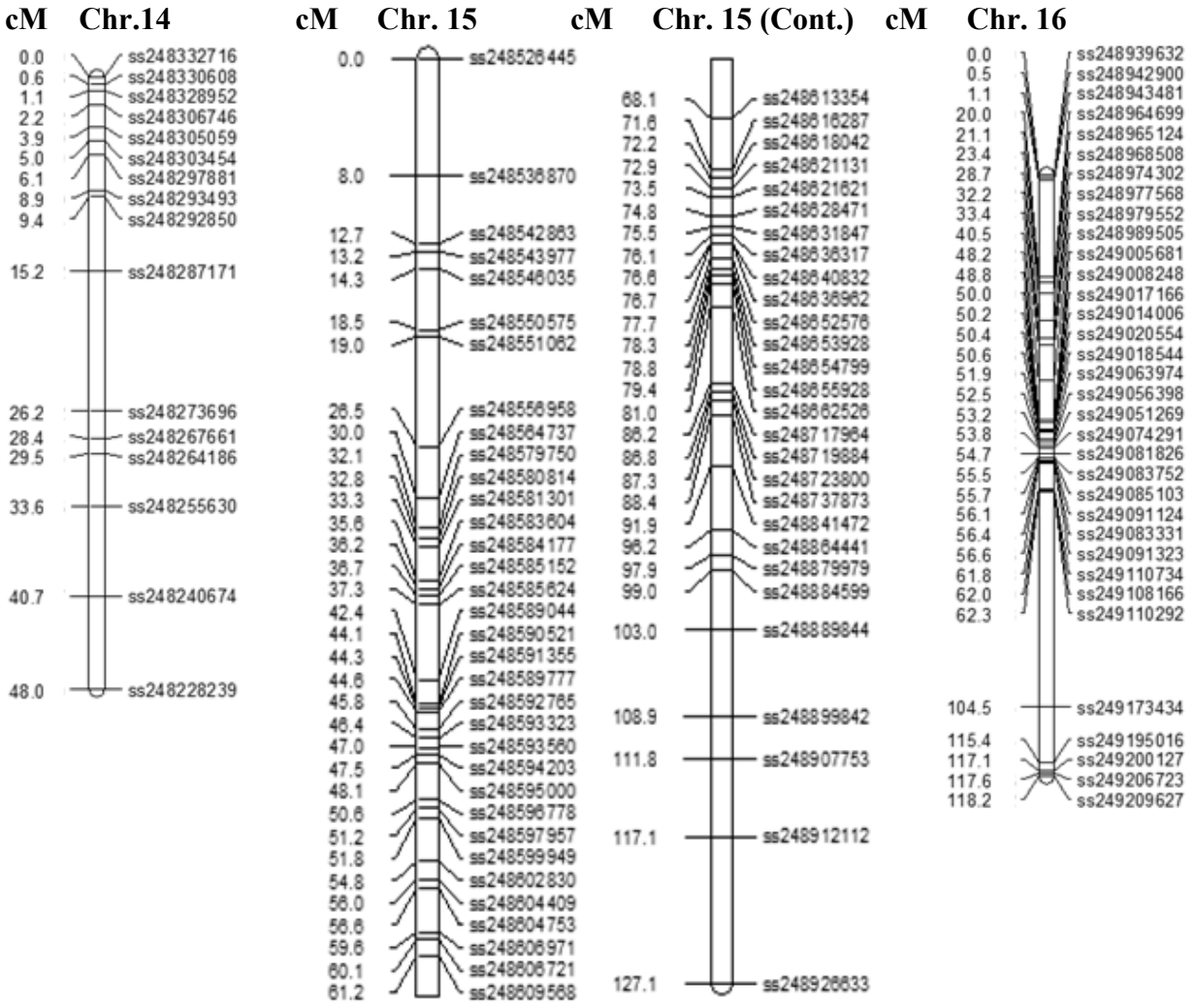

Fig. 2d. A genetic map constructed for chromosomes 14, 15, and 16 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes.

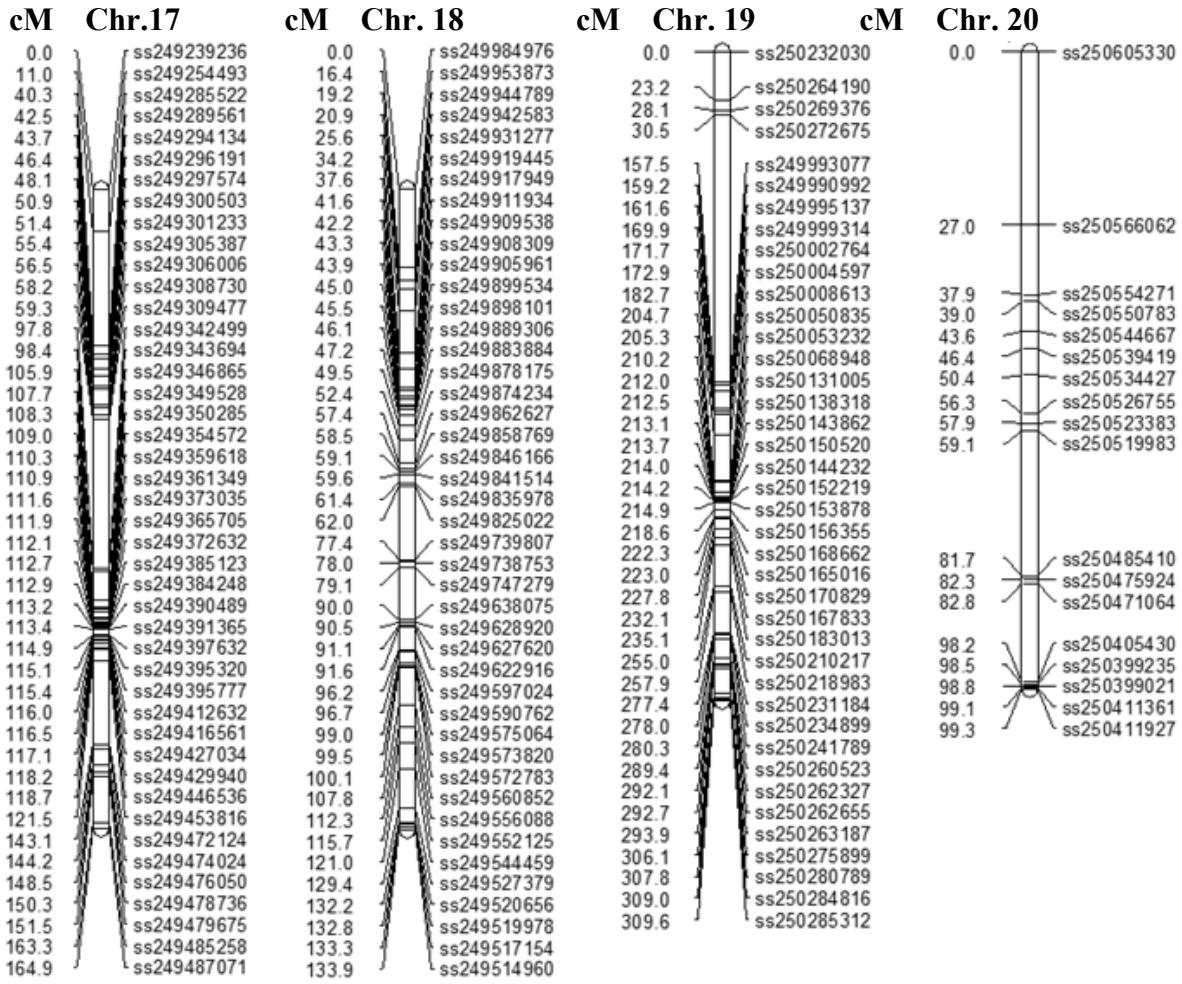

Fig. 2e. A genetic map constructed for chromosomes 17, 18, 19, and 20 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes.

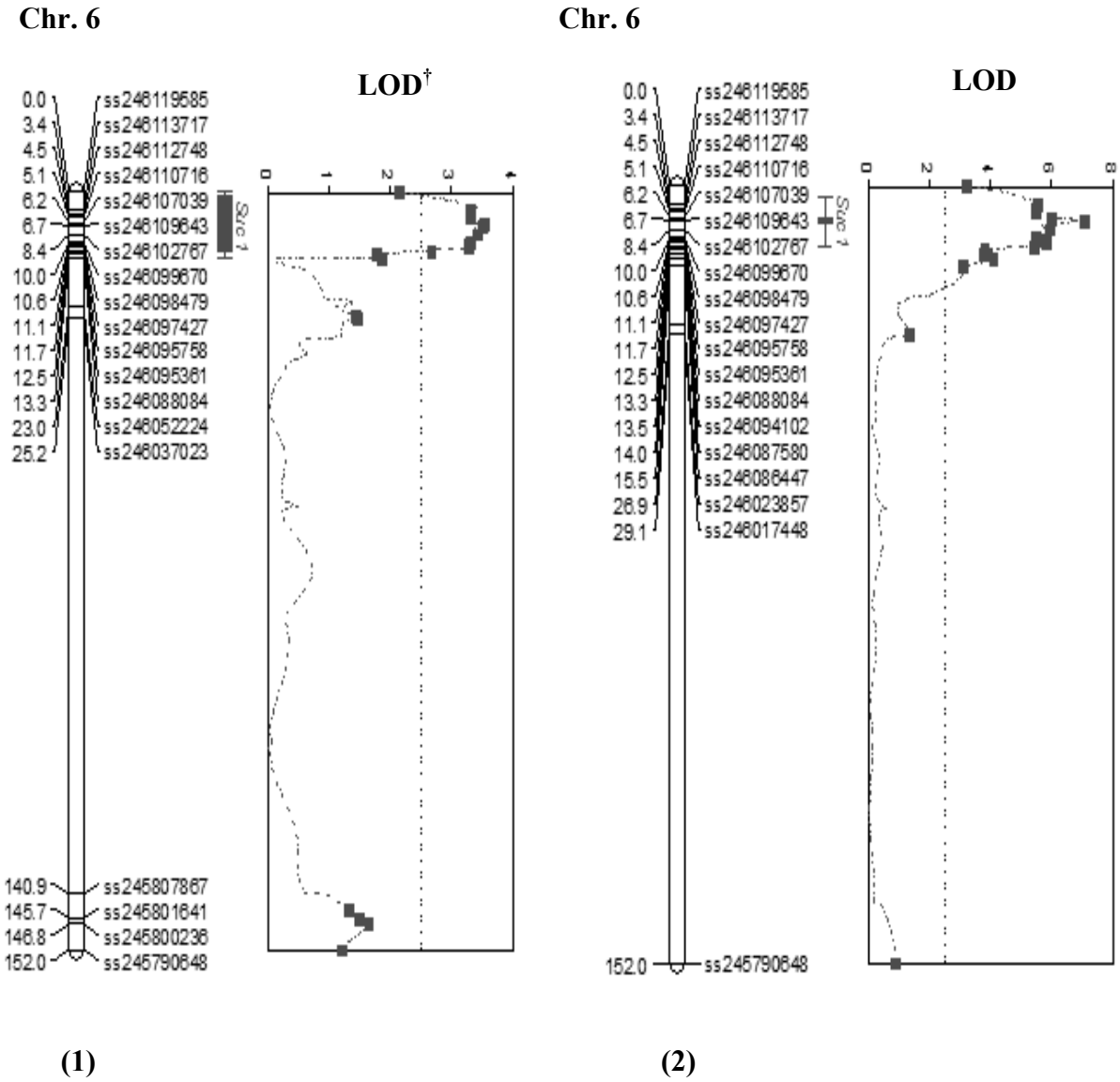

Fig. 3a. Composite interval mapping using SNP markers for seed sucrose QTL on chromosome 6 in 92 F_2-derived lines from the cross V97-3000 x V99-5089, evaluated across locations and years: (1) in Fayetteville, AR in 2009; (2) in Fayetteville, AR in 2010.
† LOD = logarithm of the odds.

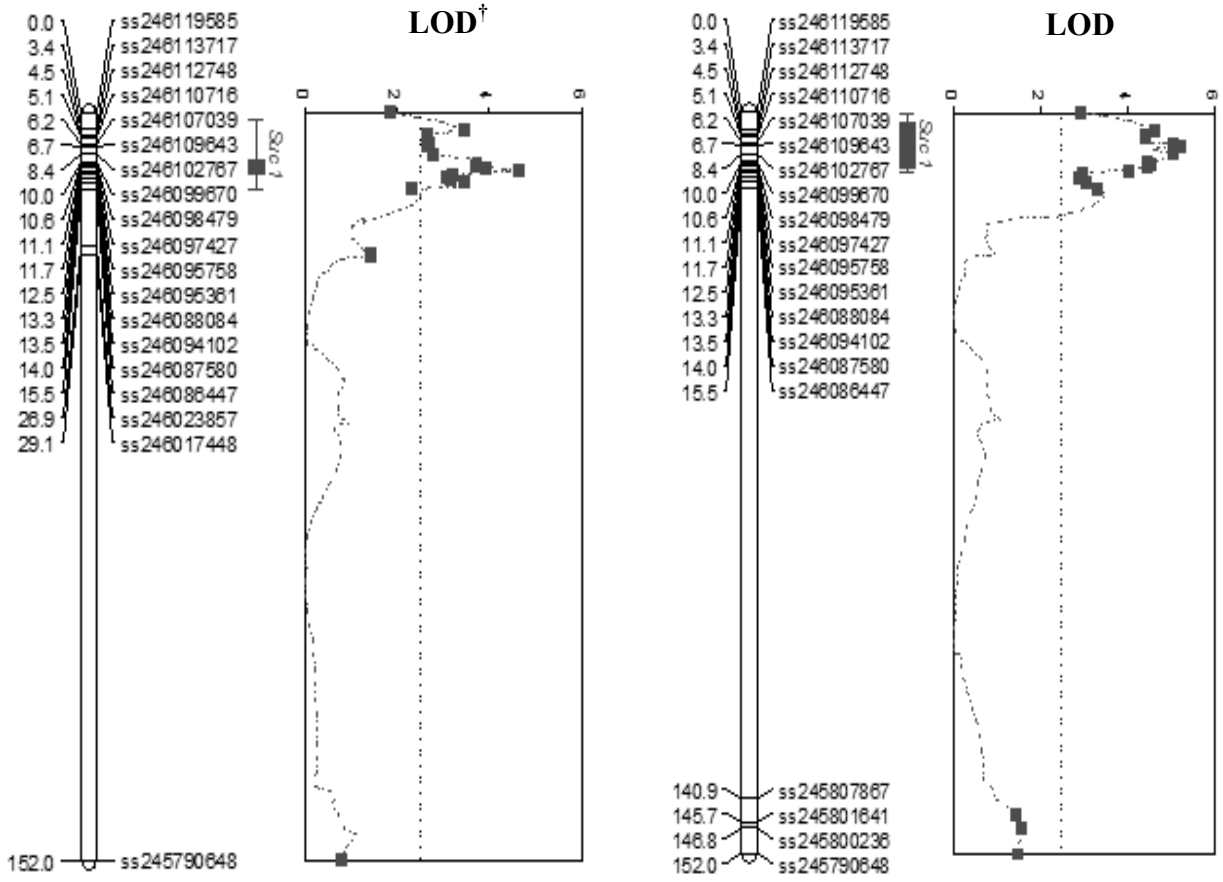

Fig. 3b. Composite interval mapping using SNP markers for seed sucrose QTL on chromosome 6 in 92 F_2-derived lines from the cross V97-3000 x V99-5089, evaluated across locations and years: (c) in Fayetteville, AR in 2011; (d) in Marianna, AR in 2011.
† LOD = logarithm of the odds.

Chr. 6

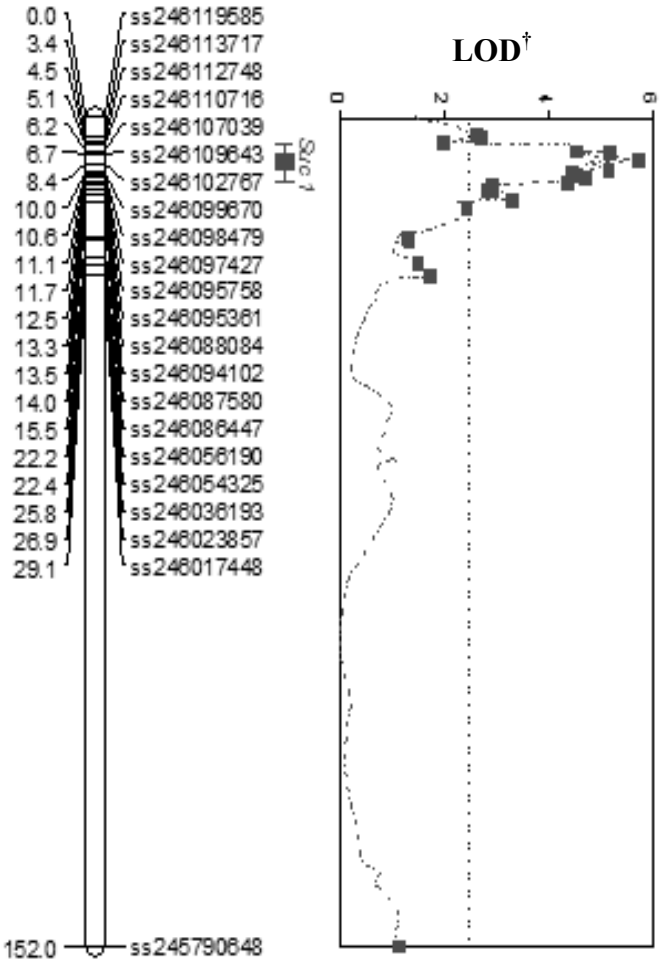

Fig. 3c. Composite interval mapping using SNP markers for seed sucrose QTL on chromosome 6 in 92 F_2-derived lines from the cross V97-3000 x V99-5089 grown in Fayetteville and Marianna in Arkansas in 2011 (Combined data).
† LOD = logarithm of the odds.

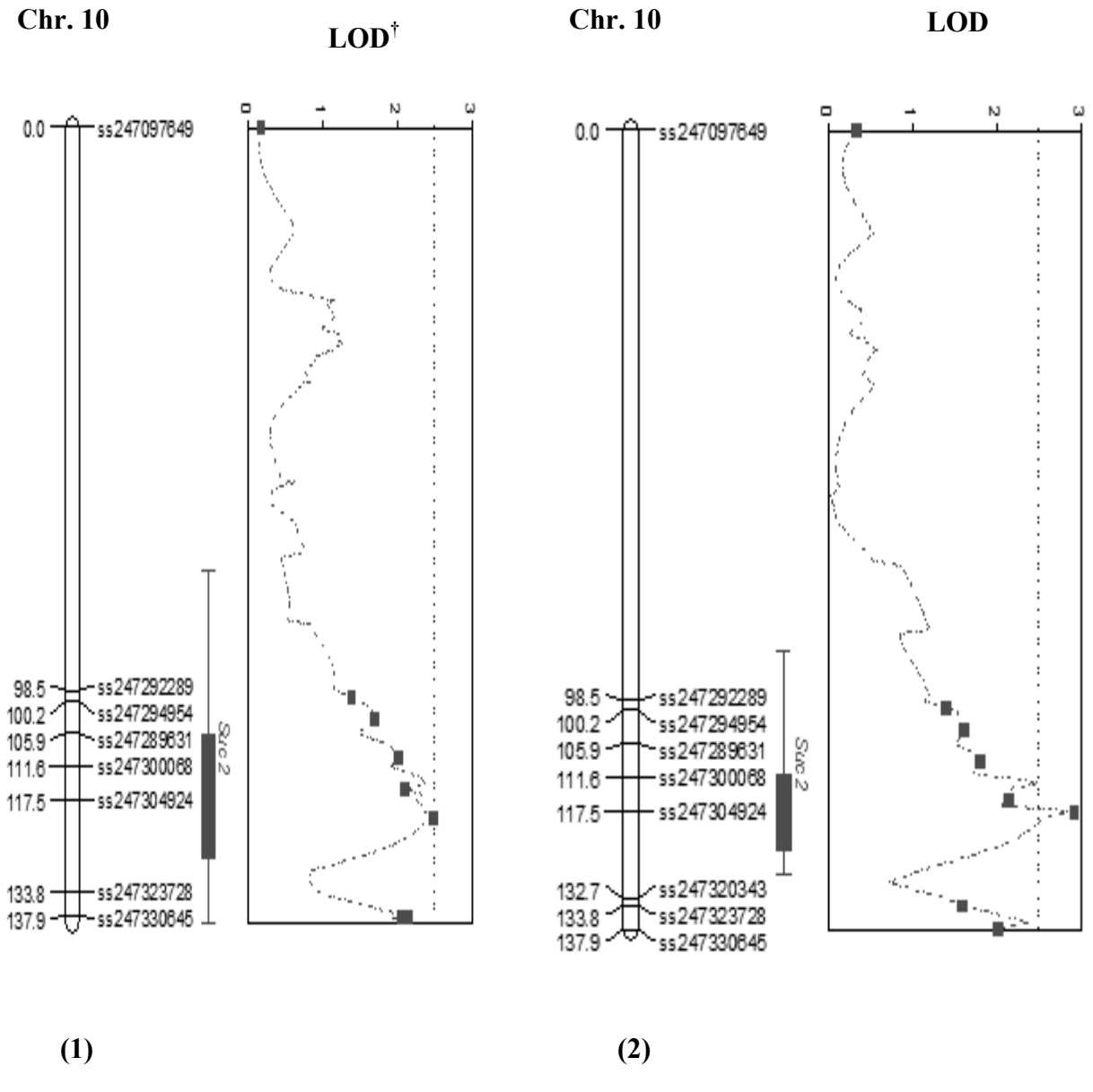

Fig. 4a. Composite interval mapping using SNP markers for seed sucrose QTL on chromosome 6 in 92 F_2-derived lines from the cross V97-3000 x V99-5089 grown: (1) in Fayetteville, AR in 2009 and (2) in Fayetteville, AR in 2010.
† LOD = logarithm of the odds.

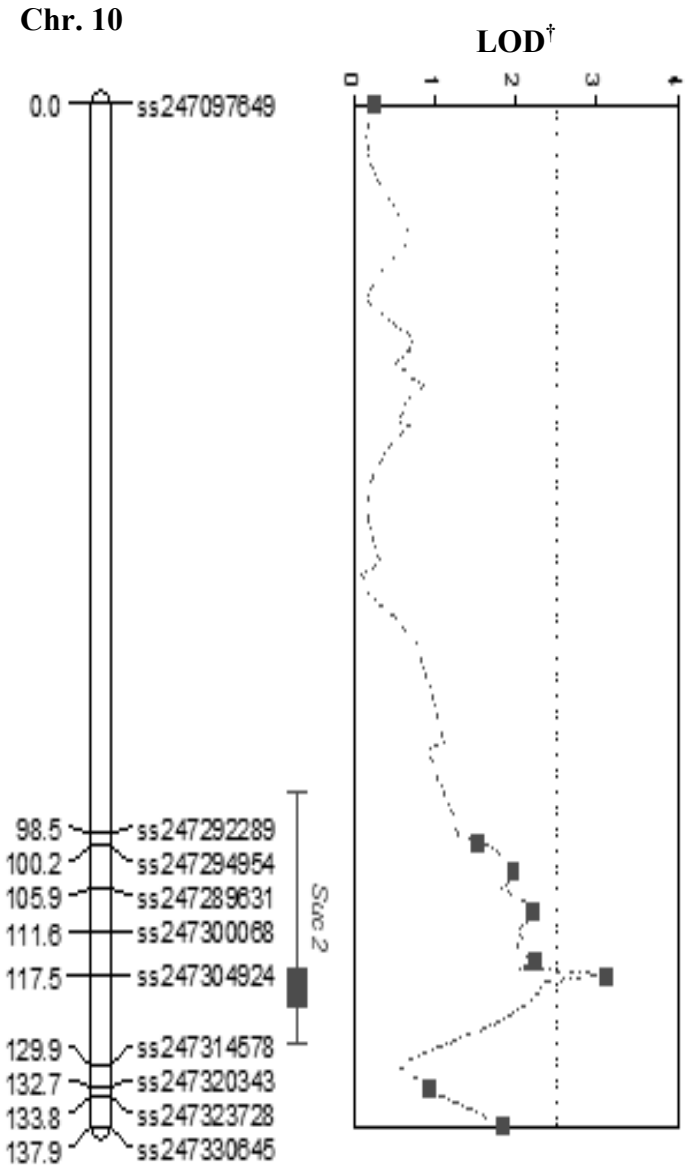

Fig. 4b. Composite interval mapping using SNP markers for seed sucrose QTL on chromosome 6 in 92 F_2-derived lines from the cross V97-3000 x V99-5089 grown in Fayetteville, AR in 2009 and 2010 (Combined data).
† LOD = logarithm of the odds.

III. Allelism Test for Low-stachyose soybean mutants

Abstract

Stachyose is an undesirable oligosaccharide in soybean meal. Its consumption causes flatulence and abdominal discomfort in monogastric animals due to the lack of the enzyme α-(1, 6)-galactosidase which breaks it into digestible forms. The recessive gene *stac1a* has been identified in the plant introduction PI200508, conferring low-stachyose content. Similarly, the breeding line V99-5089 possesses a mutation in the D-myo-inositol3-phosphate synthase 1 gene (*MIPS1*) which results in high sucrose and low stachyose content. Thus, the objective of this study was to investigate the genetic relationship between the recessive gene in PI200508 and the mutant gene in V99-5089. An F_2-derived population was developed from crossing these parents and seeds from the parents and $F_{2:3}$, $F_{2:6}$, and $F_{2:7}$ generations were used for sugar analysis with high performance liquid chromatography. Chi-square analysis was conducted to test goodness-of-fit of the observed segregation to the expected genetic ratios. Results indicate that the low-stachyose trait is controlled by two independent recessive genes with epistatic effect, from PI200508 and V99-5089. Additionally, gene dosage on stachyose content in the heterozygous genotypes with one or both genes was observed, as lines with double heterozygous genotype had lower stachyose content than those with single heterozygous genotype. However further study is needed to confirm this gene dosage effect. Breeding and selection of soybeans with modified-stachyose profile should be easy, since only one gene is needed to produce low-stachyose content and the environment has very little impact on the trait.

Introduction

The use of soybean as a functional ingredient in the food industry has grown in the western culture in the past decades. This phenomenon is the result of a widespread awareness on the nutritional value and health benefits associated with soybean consumption (Golbitz, 1995).

Carbohydrates are the second-largest soybean seed component after protein. Soluble carbohydrates are important in the food industry because they have an effect on flavor and quality of the soy-based products (Taira, 1990). Soluble sugars constitute up to 47% of the total carbohydrate content (Hymowitz and Collins, 1974). Major soluble carbohydrates are sucrose (2.5 - 8.2% dry weight basis), stachyose (1.4 - 4.1% dry weight basis), and raffinose (0.1 - 0.9% dry weight basis). The minor soluble sugars, glucose, and fructose account for less than 1% each (Hymowitz et al., 1972). Most of the insoluble carbohydrates are primarily cell wall components (Liu, 1997).

Raffinose and stachyose are considered raffinose family oligosaccharides (RFO) (Gitzelmann and Auricchio, 1965; Hymowitz and Collins, 1974). Their synthesis from sucrose and galactose is mediated by enzymes with galactosyl-transferanse function. Additionally, galactinol synthase is a catalyzer in the synthesis of galactinol (*O*-α-D-galactopyranosyl-(1→1)-L-*myo*-inositol) from myo-inositol and UDP-D-Gal (Peterbauer and Richter, 2001). Subsequently, reversible addition of a galactosyl unit to sucrose results in synthesis of raffinose. Then, stachyose is synthesized by the reversible addition of another galactosyl unit to raffinose. These reactions are mediated by the enzymes raffinose synthase and stachyose synthase, respectively (Fig. 1) (Peterbauer and Richter, 2001).

Stachyose is the most abundant raffinose saccharide and the most anti-nutritional soluble sugar in soybean (Cristofaro et al., 1974). Consumption of stachyose causes flatulence and

abdominal discomfort in non-ruminant animals (Kuriyama and Mendel, 1917; Hawton et al., 1996), due to their lack of the α-(1,6)-galactosidase enzyme that breaks down the stachyose into digestible forms (Gitzelmann and Auricchio, 1965). Thus, stachyose has a negative effect on the feed efficiency of soybean meal. The metabolizable energy of soybean meal is low compared its total energy which is due to the poor digestibility of the raffinose oligosaccharides (Parsons et al., 2000; Meis et al., 2003). Therefore, development of soybean lines with reduced-stachyose content is desired for increasing the energy metabolism rate and digestion of soybean meal.

The plant introduction (PI) 200508 has a recessive gene, *stcla*, which confers high-sucrose, low-raffinose, and low-stachyose contents (Sebastian et al., 2000). Dierking and Bilyeu (2008) characterized the trait as a variant allele of a raffinose synthase gene. Additionally, the breeding line V99-5089 has been identified as source of high-sucrose, low-stachyose, and low-phytate contents (Saghai-Maroof and Buss, 2008).

Little is known about the genetic relationship between the genes conferring low-stachyose in PI200508 and V99-5089. A preliminary study suggested the existence of 2 non-allelic recessive genes controlling this trait (Florez-Palacios, 2009); however, more research is needed in order to confirm these findings.

Materials and methods

Parental materials

Two low-stachyose sources were used as parents in this study, PI200508 and V99-5089. PI200508 is a plant introduction line from Japan (Germplasm Resources Information Network) that contains low levels of raffinose and stachyose (Sebastian and Kerr, 2000) that has purple flowers, broad leaves, buff cotyledons, and regular seed size (~15g 100 seeds^{-1}) with reduced

stachyose content (~1 % on dry-weight basis). Virginia Tech University derived a breeding line, V99-5089, from a cross between V71-370 and PI 87013 (Saghai-Maroof and Buss, 2008). It has purple flowers, broad leaves, yellow cotyledons and large seed size (25 to 30g 100 seeds^{-1}) with reduced stachyose content (<1 on dry-weight basis).

Population development and field experiment

A PI200508 x V99-5089 cross was made in the summer of 2007 at the Arkansas Agricultural Research and Extension Center in Fayetteville. Progeny were bulked and carried through to the F_2 generation. The F_2 plants were grown in a winter nursery in spring 2008. The winter nursery is located near Upala city in the northwestern part of Costa Rica at 52 m above sea level (Mongabay, 2014). There, each $F_{2:3}$ line was grown in a 3-m row with 0.76-m row spacing. Additionally, parental genotypes were planted next to the segregating population with a 3-m row spacing. Soil was cultivated before planting and plots were fully irrigated and managed during the growing season using standard cultural practices. After two generation advancements, $F_{2:5}$ seed were brought back to Fayetteville, AR in 2009.

In summer 2011, a total of 121 $F_{2:6}$ lines were grown in randomized complete block design (RCBD) with two replications in each of two locations, Keiser and Fayetteville, AR. Replicated parental checks were also grown along with $F_{2:6}$lines to account for variation within blocks. The PI200508 and V99-5089 lines were included two times, each a with different entry number, to make up a 125-entry test.

Seed at the Arkansas Agricultural Research and Extension Center in Fayetteville were grown on a Captina-silt-loam soil (Fine-silty, siliceous, active, mesic Typic Fragiudults) (Soil Survey Staff, 2013), which is characterized as very deep, moderately to well drained, and made in a thin mantle of silty material (Soil Series, 2006). Whereas, seed at the Northeast Research

and Extension Center in Keiser, AR were grown on a Sharkey-silty-clay soil (Very-fine, smectitic, thermic Chromic Epiaquerts), characterized as very deep, poorly to very poorly drained, very slowly permeable soils that formed in clayey alluvium (Soil Series, 2006). For both locations, plots consisted of one 3-m single with 0.9-m row spacing. Seeding rate was 30 seeds per meter, and standard cultural practices were applied throughout the growing season. Similarly, plants were grown in a 3-m row with a 0.76-m row spacing in the winter nursery in Costa Rica. Seed samples were taken at $F_{2:3}$, $F_{2:6}$, and $F_{2:7}$ generations for soluble carbohydrate analyses.

Soluble carbohydrate extraction

Sugar extraction followed the protocol described by Hou et al. (2009), with some modifications. Briefly, a sample of 10 g of whole soybean seed was ground for 20 sec, to a fine powder using a coffee bean grinder (Krups®, Shelton, CT). The powder was then sifted through a 100 µm sieve (USA Standard testing sieve, opening micrometer 150 A.S.T.M. No. 100), and 0.15 g were used for carbohydrate extraction. The powder was placed in a 2 mL centrifuge tube and 1.5 mL of distilled and deionized water (ddH_2O) was added and the solution vortexed. Afterward, the tube was shaken horizontally at 200 rpm for 20 min. The solution was then centrifuged for 10 min at 14000 rpm. An aliquot of 500 µl was transferred into a new tube and 700 µl of 100% acetonitrile were added to the extract. Solution was mixed by inversion, incubated at room temperature for 30 min, and centrifuged for 10 min at 14000 rpm. About 3 mL of sugar solution was filtered with a 0.2 µm membrane using a 5 mL syringe. Then, 24 µL of carbohydrate extract was transferred into a 1.5 mL HPLC vial containing 576 µL of ddH_2O for high performance liquid chromatography (HPLC) analysis.

Carbohydrate analysis

The anion-exchange HPLC system (Dionex DX500 HPAEC-PAD) was composed of a

GS50 gradient pump, an ED40 pulsed amperometric electrochemical detector, an LC50 chromatographic oven, an AS40 automated sampler with a 25-µL injection loop, and a Chromeleon Chromatography Management Data System. The mobile phase consisted of a 90 mM NaOH solvent (VWR®, West Chester, PA) at constant flow rate of 1 mL min^{-1}, which was prepared by diluting carbonate-free HPLC grade 50% (w/w) stock solution in distilled water, filtered with a 0.45-µm membrane, and degassed with compressed nitrogen gas for 30 min before vials were loaded into the auto sampler, as described by Hou et al. (2009). Soluble carbohydrates were separated by an analytical CarboPac PA-10 pellicular anion-exchange resin column (4 x 250 mm) coupled to a CarboPac PA10 guard column (4 x 50 mm) and preceded by AminoTrap column (3 x 30 mm) (Dionex, Sunnyvale, CA). Sugar content was determined from regression curves fitted from a set of standards for sucrose (Sigma-Aldrich, St. Louis, MO) at different concentrations (10, 20, 40, 60, and 80 µg $µL^{-1}$). Carbohydrate data then converted to milligrams of sugar per gram of seed (mg g^{-1}) on dry weight basis.

Statistical analysis

The Shapiro-Wilk (*w*) statistic from JMP 10.0 (SAS Institute, Cary, NC) was used to test the normality of the stachyose content distribution for the $F_{2:3,}$ $F_{2:6}$, and $F_{2:7}$ lines.

Analysis of variance for seed stachyose content was conducted using PROC GLM in SAS 12.3 (SAS Institute, Cary, North Carolina, USA). Variance components were estimated using a similar model in the VARCOMP procedure of SAS 12.3. Percent variation of each source in the model was calculated by dividing each variance estimate over the total variance and multiplying by 100.

The criteria chosen to establish the thresholds for high- and low-stachyose values for the populations were the shape of distribution of the population for each environment and the

respective mean of the high stachyose parent (PI200508) plus/minus one standard deviation. Therefore, cutoff points to distinguish between low- and high-stachyose lines were: 1.23% for the $F_{2:6}$ lines grown in Fayetteville in 2010, 1.32% for the $F_{2:7}$ 1ines grown in Fayetteville in 2011, 1.26% for the $F_{2:7}$ 1ines grown in Keiser in 2011, and 1.28% for the combined data in 2011 (Fig. 2).The $F_{2:6}$ and $F_{2:7}$ population distributions were subjected to a Chi-square test for two genes segregating for stachyose content, and the result from the calculated Chi-squared statistic was also contrasted with the expected Chi-square parameter at an alpha level of 0.05.

Results

Both parents PI200508 and V99-5089 grown in all three environments exhibited the low-stachyose trait as expected. The stachyose content of V99-5089 was consistently lower than that of PI200508 (Table 1). The stachyose content of the F_2-derived lines ranged from 0.07 to 4.30% with a mean of 1.82%. The stachyose content distribution of F_2-derived lines did not follow a normal distribution (Figs. 2 and 3), as indicated by the Shapiro-Wilk normality test (Table 1).

The ANOVA was conducted on 121 $F_{2:7}$ lines plus the four parental checks, and means were separated by Least Significant Difference (LSD) at $P < 0.05$. There was a significant effect of genotype, location, and genotype x location interaction on seed stachyose content. However, most of the variation was accounted for by genotypes (99.03%), while the location (0.004%) and genotype by location interaction (0.67%) effects were almost negligible in magnitude as compared to the experiment error (Table 2).

Since the stachyose content of the F_2-derived lines did not follow the normal distribution and there were obvious genetic segregants with high stachyose content in the population, Chi-square tests were performed to examine if the segregation pattern fitted any expected ratio based

on a two-independent-gene model. By using a threshold calculated based on the two low stachyose parents, the F_2-derived lines were classified as either high or low stachyose phenotype. Results showed that the F_2-derived lines consistently fitted a 9 high stachyose : 7 low stachyose ratio in all three environments, which is expected for two complementary genes (Table 3 and Fig. 4).

The 9/16 of the F_2-derived lines classified as high stachyose for the Chi-square test had stachyose contents ranging from the low stachyose parent (0.5 - 1.2%) to a normal stachyose level as expected for a commodity soybean genotype (4.0 - 4.5%) (Figs. 2 and 3). This is apparently a result of gene dosage effect from the heterozygous genotypes from either one or both low stachyose genes.

Discussion

The stachyose trait showed considerable stability, given that both parents consistently exhibited low stachyose content across the four environments studied (Table 1) and the F_2-derived lines ranked similarly across environments. All except four F_2-derived lines fell in the same phenotypic classifications based on the parental threshold of stachyose content (Table 3). Although ANOVA showed a significant effect of location and genotype x location on stachyose content, those effects were negligible in relevance to the error term (Table 2). Unlike a typical quantitative trait, stachyose content in my genetic population appeared to be in discrete distribution and less affected by the environment (Figs. 2 and 3). This would make breeding and selection for low stachyose simple and easy.

The non-normal distribution of stachyose content in the F_2-derived lines from the cross PI200508 x V99-5089 clearly showed that the seed stachyose content is not a quantitative trait,

but a trait controlled by major genes (Figs. 2 and 3). The Chi-square tests confirmed the presence of two independent, recessive genes for the low stachyose trait in the population (Table 3). In a low- x low-stachyose cross, one would expect all progeny lines to have low stachyose content similar to the low stachyose parents if the low-stachyose genes in the parents are allelic at a single locus. A proportion of the progeny lines is expected to be high stachyose if there is more than one gene involved, depending on the number of genes and their interactions. In a two-gene model, a 13 high : 3 low stachyose ratio would be expected for both dominant and recessive epistasis, while a 15 high : 1 low stachyose ratio would be expected for duplicate dominant epistasis. In my population, 7/16 of the F_2-derived lines exhibited low stachyose which is expected for two genes, both with recessive epistasis. The presence of lines with greater stachyose content than that of the parents provides strong evidence that the two genes controlling the low-stachyose trait in PI200508 and V99-5089 are at different loci.

Previous research reported a marker linked to a locus containing a recessive gene, *rsm1*, on chromosome 6 in PI200508 that explained up to 94% of the stachyose variance, and also suggested that the low stachyose content in PI200508 was due to a 3 bp deletion in the *galactosyltransferase* gene, which reduces the activity of the enzyme involved in the stachyose synthesis pathway (Skoneczka et al., 2009). Sebastian et al. (2000) identified a recessive gene *Stc1a* for low stachyose content in PI200508, but no allelism test was done with the low stachyose gene in V99-5089. The recessive gene in PI200508 identified in my study and in the study by Sebastian et al. (2000) may be the same QTL reported by Skoneczka et al. (2009). Molecular studies are needed to confirm this assumption.

Maupin et al. (2011) suggested that a SSR marker on chromosome 11 in V99-5089 could be used in marker-assisted selection of low stachyose lines due to its high selection efficiency

(87%), suggesting that chromosome 11 may contain genes controlling the trait. Further research identified a stachyose QTL on chromosome 11 in a mapping population derived from the cross Osage x V99-5089, explaining 81% of the stachyose variation (Zeng et al., 2012). It is very likely that the low stachyose gene identified in our research is the same stachyose QTL reported by Zeng et al. (2012).

The fact that some of the F_2-derived lines classified as high stachyose type, in the Chi-square test, actually exhibited intermediate stachyose content suggests that there was a gene dose effect on stachyose content in the heterozygous genotypes with one or both loci. The stachyose content in the high class ranged from 1.51 to 4.30%, depending on the number of heterozygous loci, assuming that the double heterozygous genotype (AaBb) would have lower stachyose content than the single heterozygous genotype (AABb or AaBB). Further study using selected lines with variable stachyose content and molecular markers is needed to confirm the gene dosage effect on stachyose content.

Conclusion

Results from my study indicated that the low-stachyose trait is controlled by two independent recessive genes from pi200508 and v99-5089, and only one gene is needed to have the low-stachyose phenotype. Gene dosage effect on stachyose content in the heterozygous genotypes with either one or both genes was observed; however further study is needed to confirm this effect. In a practical breeding program, both pi200508 and v99-5089 would serve as good sources of low stachyose and selection for this trait would be straightforward as it is negligibly impacted by the environment.

References

Cristofaro, E., F. Nottu, and J.J. Wuhrmann. 1974. Involvement of the raffinose family of oligosaccharides in flatulence. In: Sugars in Nutrition (Sipple, H.L. and K.W. McNutt, Eds.) Academic Press. New York. pp 313 – 336.

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.

Florez Palacios, S.L. 2009. Genetic study of seed sugar content in diverse soybean germplasm. University of Arkansas. Department of Crop, Soil, and Environmental Sciences. M.S. Thesis.

Gitzelmann, R., and S. Auricchio. 1965. The handling of soy alpha-galactosidase by a normal and galactosemic child. Pediatrics 36:231-238.

Hawton, J.D., L.J. Johnston, T.M. Salzer, and G.C. Shurson. 1996. Applications of nutritionally improved corn and soybeans for swine. p. 263-304. In Proc. 57th Minnesota Nutrition Conf., University of Minnesota St. Paul, MN.

Hou, A., P. Chen, J. Alloatti, D. Li, L. Mozzoni, B. Zhang, and A. Shi. 2009. Genetic variability of seed sugar content in worldwide soybean germplasm collections. Crop Sci. 49:903-912.

Hymowitz, T. and F.I. Collins. 1974. Variability of sugar content in seed of *Glycine max* (L.) Merrill and G. soja Sieb and Zucc. Agron. J. 66:239-240.

Hymowitz, T., F.I. Collins, J. Panczner, and W.M. Walker. 1972. Relationship between the content of oil, protein, and sugar in soybean seed. Agron. J. 64:613-616.

Krober, O.A. and J.L. Cartter. 1962. Quantitative interrelations of protein and non-protein constituents of soybeans. Crop Sci. 2:171-172.

Kuriyama, S. and L.B. Mendel. 1917. The physiological behavior of raffinose. J. Biol. Chem. 31:125-147.

Jones, D.A., DuPont M.S., Ambrose M.J., Frias J., and Hedley C.L. 1999. The discovery of

compositional variation for the raffinose family of oligosaccharides in pea seeds. Seed Sci. Res. 9:305-310.

Liu, K.S. 2005. Soybeans as a powerhouse of nutrients and phytochemicals. p. 1-22. In: Soybeans as functional foods and ingredients. AOCS Press, Champaign, IL.

Liu, K.S. 1997. Chemistry and nutritional value of soybean components. p. 25-113. In: Soybeans: Chemistry, technology, and utilization. Chapman & Hall, New York, NY.

Meis, S.J., W.R. Fehr, and S.R. Schnebly. 2003. Seed source effect on field emergence of soybean lines with reduced phytate and raffinose saccharides. Crop Sci. 43:1336-1339.

Mongabay. 2014. http://population.mongabay.com/population/costa-rica/3621153/upala (accessed February 26, 2014).

Mozzoni, L., A. Shi, and P. Chen. 2013. Genetic analysis of high sucrose, low raffinose, and low stachyose content in V99-5089 soybean seeds. J. of Crop Improv. 27:606-616.

Parsons, C.M., Y. Zhang, M. Araba. 2000. Nutritional evaluation of soybean meals varying in oligosaccharide content. Poultry Sci. 79:1127-1131.

Peterbauer, T. and A. Richter. 2001. Biochemistry and physiology of raffinose family oligosaccharides and galactosyl cyclitols in seeds. Seed Sci. Res. 11:185-187.

Saghai-Maroof, M.A. and G.R. Buss. 2008. Low phytic acid, low stachyose, high sucrose soybean lines. U. S. Patent 0199591 A1.

SAS Institute. 2013. SAS/STAT Software Version 12.3. SAS Institute Inc, Cary, NC.

Sebastian, S.A., P.S. Kerr, R.W. Pearlstein, and W.D. Hitz. 2000. Soybean germplasm with novel genes for improved digestibility. p. 56-74. In: J.K. Drackely (ed.) Soy in animal nutrition. Federation of Animal Sci. Soc., Savoy, IL.

Soil Series. 2006. Captina Series. Rev. LBW/RLT. National Cooperative Soil Survey. U.S.A. https://soilseries.sc.egov.usda.gov/OSD_Docs/C/CAPTINA.html (accessed March 6,

2014).

Soil Series. 2006. Sharkey Series. Rev. LBW/JDS. National Cooperative Soil Survey. U.S.A. https://soilseries.sc.egov.usda.gov/OSD_Docs/S/SHARKEY.html (accessed March 6, 2014).

Taira, H. 1990. Quality of soybeans for processed foods in Japan. Japan Agric. Res. Quarterly 24:224-230.

Zeng, A., P. Chen, A. Shi, D. Wang, B. Zhang, M. Orazaly, L. Florez-Palacios, K. Brye, Q. Song, and P. Cregan. 2014. Identification of quantitative trait loci for sucrose content in soybean seed. Crop Sci. 54:1-11.

Table 1. Mean and range of stachyose content for parents and populations derived from the PI200508 x V99-5089 cross grown in Fayetteville in 2010 and 2011, and Keiser, AR in 2011.

	PI200508		V99-5089		F_2-derived lines		
	Mean (%)	**Range (%)**	**Mean (%)**	**Range (%)**	**Mean (%)**	**Range (%)**	**Prob < W[§]**
Fay[†] - 2010	1.13	1.06 - 1.20	0.49	0.46 - 0.51	1.70	0.07 - 3.84	<0.0001
Fay[†] - 2011	1.25	1.16 - 1.31	0.54	0.49 - 0.53	1.90	0.11 - 4.30	<0.0001
Kei[‡] - 2011	1.12	0.99 - 1.26	0.51	0.45 - 0.56	1.85	0.08 - 4.22	<0.0001
Overall[¶]	1.19	0.99 - 1.31	0.51	0.45 - 0.56	1.82	0.07 - 4.30	

[†] Fayetteville, AR.
[‡] Keiser, AR.
[§] Shapiro-Wilk normality test.
[¶] Mean and range across three environments.

Table 2. Analysis of variance for seed stachyose content in $F_{2:7}$ lines derived from PI200508 x V99-5089, grown in two Arkansas locations in 2011.

Source of variation	Degrees of freedom[†]	Sum of squares[‡]	Mean square	F-ratio	P value	Variance component[§]	Percent Variation[¶]
Model	243	816.1	3.36	1197.68	<0.0001		
Location	1	0.31	0.31	110.40	<0.0001	0.0001	0.004
Replication[Location]	2	0.54	0.27	96.70	<0.0001	0.0022	0.130
Genotype	120	812.18	6.77	2413.62	<0.0001	1.6856	99.031
Genotype x Location	120	3.08	0.02	9.14	<0.0001	0.0114	0.671
Error	240	0.67	0.01			0.0028	0.165

[†] Analysis of variance, using PROC GLM in SAS 12.3, of 121 $F_{2:7}$ lines grown in Fayetteville and Keiser, AR in 2011.
[‡] Type II sum of squares.
[§] Estimate of variance components, as calculated using PROC VARCOMP in SAS 12.3.
[¶] Percent of variation explained by each term in the ANOVA model.

Table 3. Chi-square test for goodness-of-fit to two-gene model for stachyose content of F_2-derived lines from the cross PI200508 x V99-5089 grown in Fayetteville, AR in 2010 and Fayetteville and Keiser, AR in 2011.

	No. of High-stachyose lines[†]	No. of Low-stachyose lines[‡]	X^2 (9:7)	P value
Fayetteville-2010 $F_{2:6}$ lines	67	54	0.040	0.84
Fayetteville-2011 $F_{2:7}$ lines	68	53	0.001	0.97
Keiser-2011 $F_{2:7}$ lines	69	52	0.030	0.86
2011 Combined data $F_{2:7}$ lines	68	53	0.001	0.97

[†] F_2-derived lines were classified as low-stachyose lines if stachyose content was equal to, or lower than: 1.23% for Fayetteville-2010, 1.32% for Fayetteville-2011, for 1.26% in Keiser-2011, and 1.28% for 2011 combined data.

[‡] F_2-derived lines were classified as high-stachyose lines if stachyose content was equal to, or greater than: 1.24% for Fayetteville-2010, 1.33% for Fayetteville-2011, for 1.27% in Keiser-2011, and 1.29% for 2011 combined data.

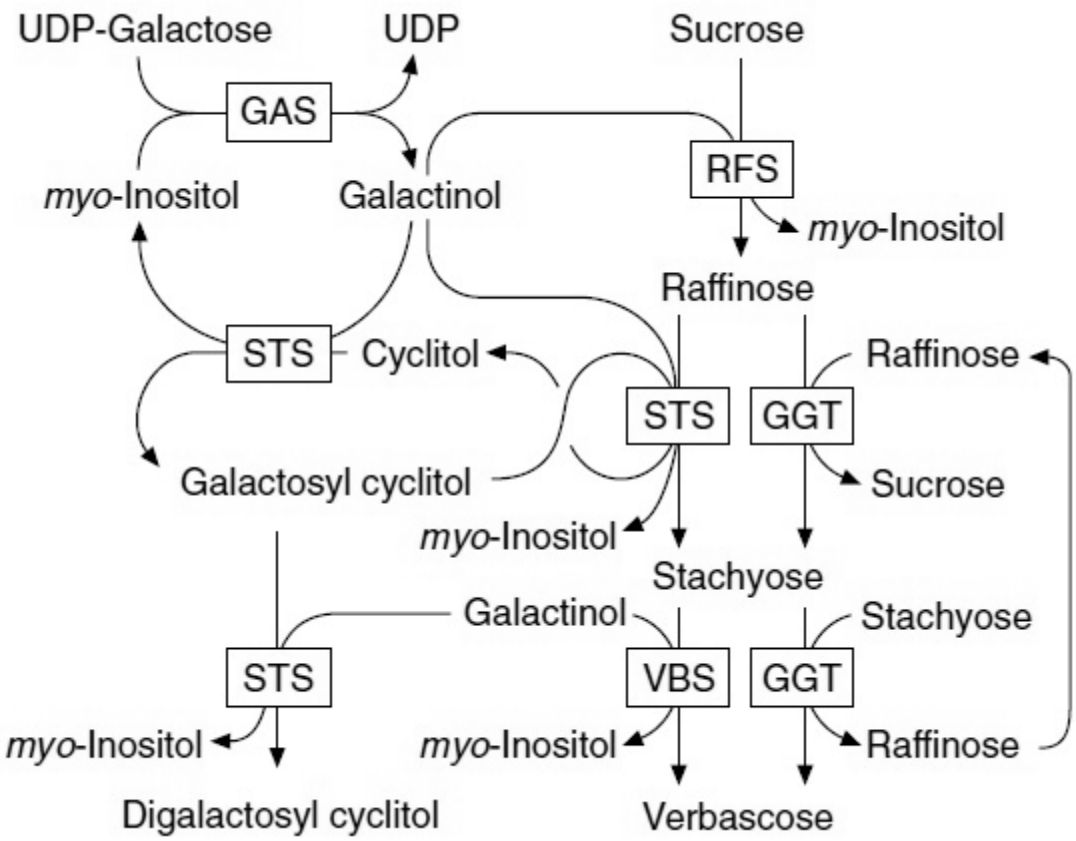

Fig. 1. Biosynthetic pathway of raffinose family oligosaccharides (RFO). Extracted from Peterbauer and Ritcher (2001).

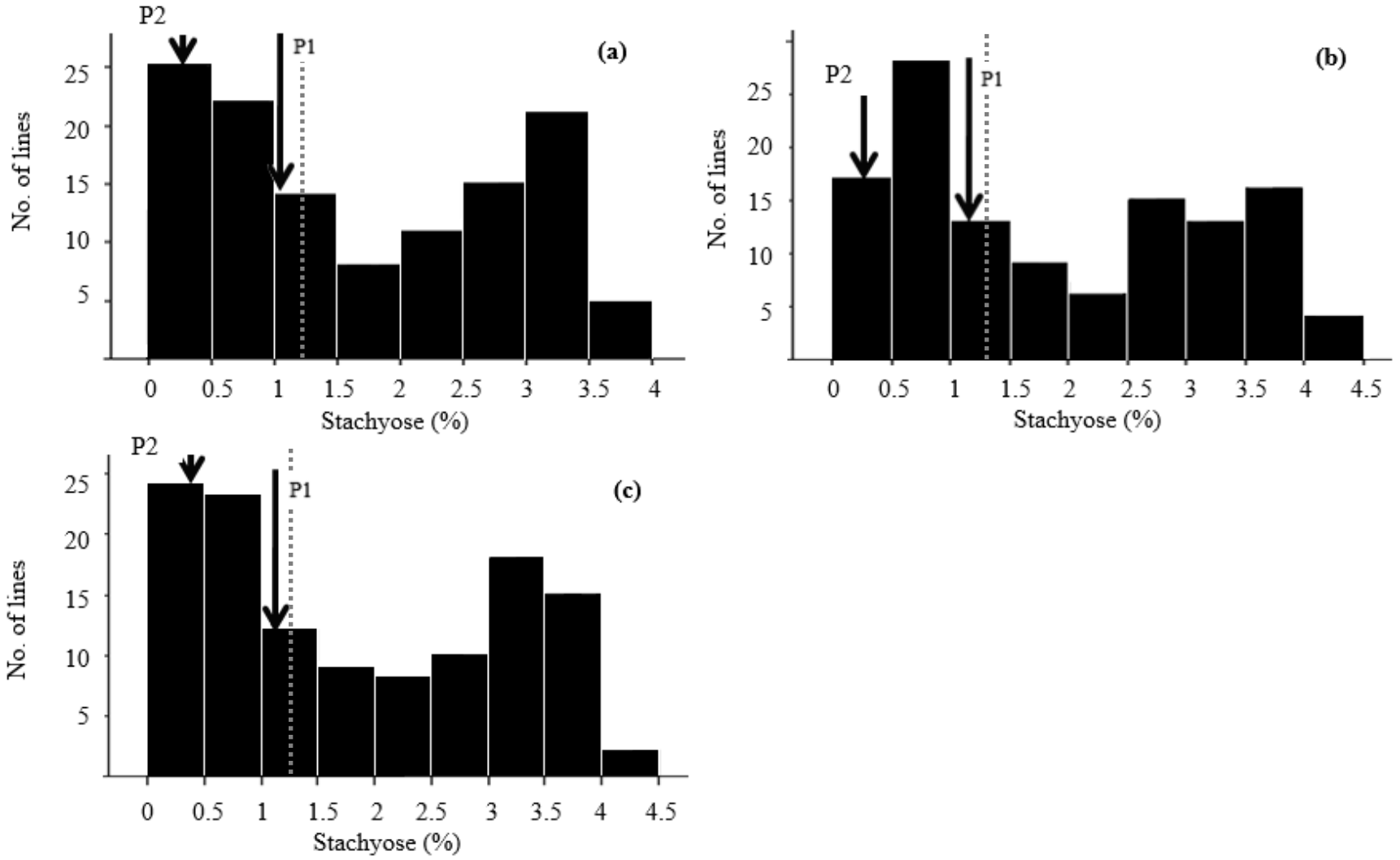

Fig. 2. Frequency distribution of seed stachyose content in F_2-derived populations from PI200508 (P1) x V99-5089 (P2) evaluated in three environments: (a) $F_{2:6}$ lines in Fayetteville, AR in 2010; (b) $F_{2:7}$ lines in Fayetteville, AR in 2011; (c) $F_{2:7}$ lines Keiser, AR in 2011. Vertical square-dotted line denotes the threshold for stachyose content: a) 1.23%, b) 1.32% and c) 1.26%.

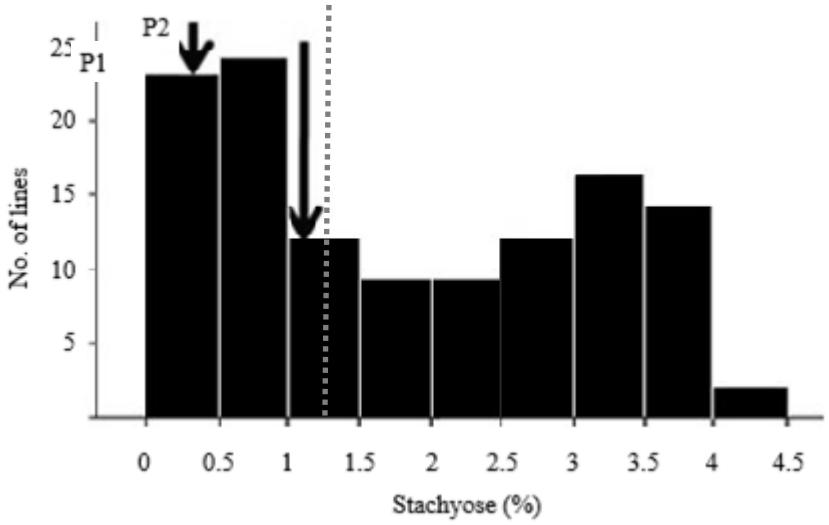

Fig. 3. Frequency distribution of seed stachyose content in $F_{2:7}$ population derived from the cross PI200508 (P1) x V99-5089 (P2) evaluated in Fayetteville and Keiser, AR in 2011. Vertical square-dotted line denotes the threshold for stachyose content = 1.28%.

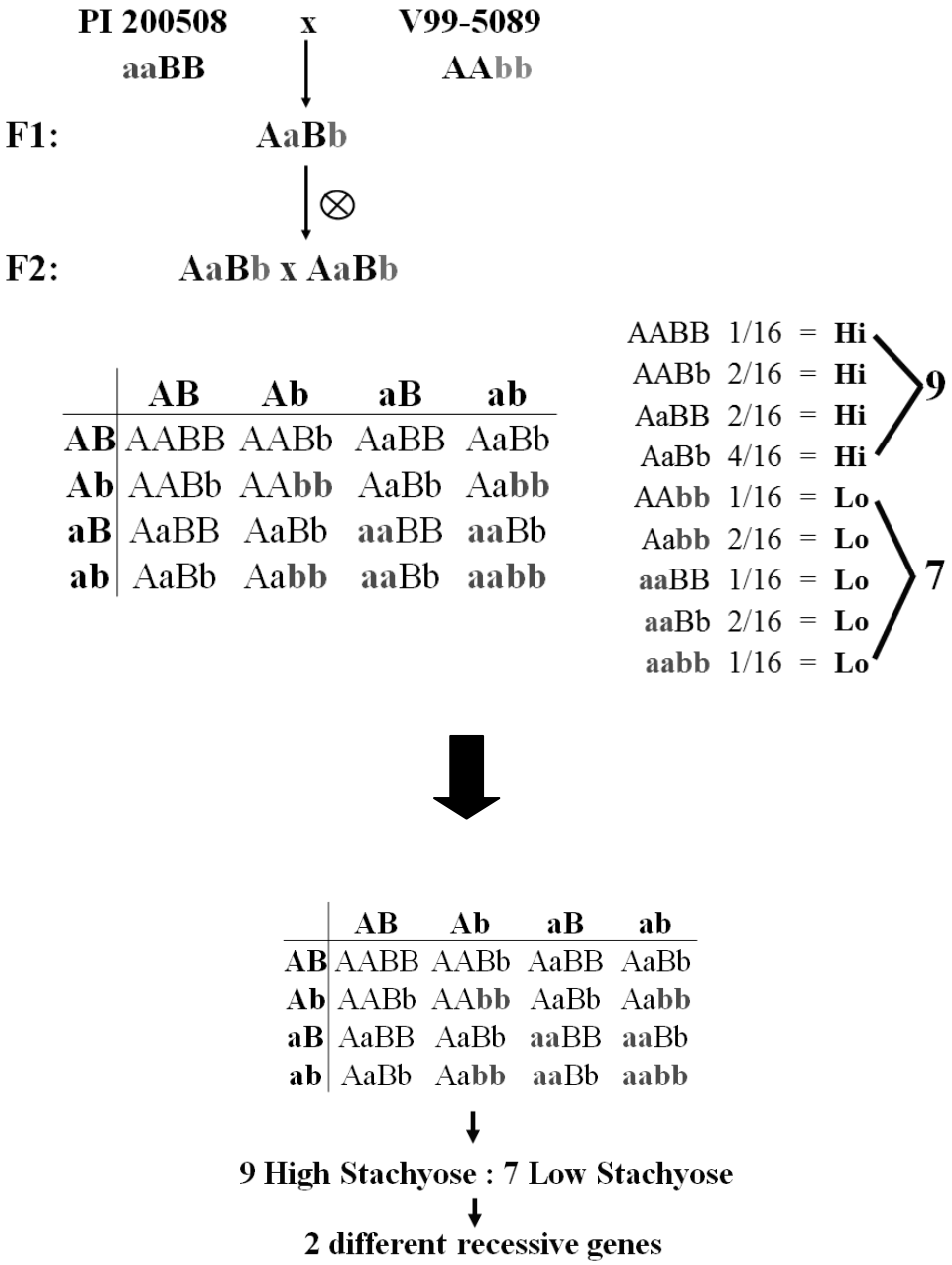

Fig. 4. Two-gene model for stachyose content of F_2-derived lines from the cross PI200508 x V99-5089.

IV. Identification of Quantitative Trait Loci / Genes for Stachyose Content in Soybean Seed

Abstract

Stachyose is the most abundant carbohydrate of the raffinose family oligosaccharides in the soybean seed. It is considered an unwanted sugar because its consumption causes abdominal discomfort in monogastric animals. Thus, breeding for low-stachyose soybean cultivars is desirable for livestock feeding of non-ruminants and the development of food-grade soybeans. The objective of this study was to identify quantitative trait loci (QTL) for stachyose content using single nucleotide polymorphism (SNP). An F_2-derived QTL mapping population was developed from the cross between a regular-sucrose line (V97-3000) and a high-sucrose line (V99-5089). A total of 92 $F_{2:3}$ lines were genotyped with 5361 SNP markers covering the 20 soybean chromosomes and, from these, 1720 were polymorphic. Seed samples were collected at $F_{2:3}$, $F_{2:6,}$ and $F_{2:7}$ generations, and stachyose analysis was performed by using high performance liquid chromatography system. Subsequently, linkage maps were constructed with JoinMap® software and composite interval mapping (CIM) was conducted to locate QTL associated with stachyose content. Two stachyose QTLs were identified in this study. One major stachyose QTL was mapped to chromosome 11 and accounted for 46% of the phenotypic variation observed for this trait. And one minor QTL was found on chromosome 10 and explained 11% of the stachyose variation. Both QTLs are in agreement with stachyose QTLs previously reported in populations with different genetic background, and also proved to be stable across the environments studied. SNP markers tightly linked to these QTLs can be used for marker-assisted selection in breeding soybean lines with low-stachyose profile.

Introduction

Stachyose is one of the most abundant soluble carbohydrates in soybean. Its content ranges between 1.4 and 4.1% on a dry-weight basis (Hymowitz et al., 1972). However, stachyose is the most unwanted sugar for livestock feeding of non-ruminants (Cristofaro et al., 1974). Stachyose intake causes flatulence and abdominal discomfort in monogastric animals (Kuriyama and Mendel, 1917; Hawton et al., 1996), due to their lack of the α-(1,6)-galactosidase enzyme that breaks down the stachyose into simple forms (Gitzelmann and Auricchio, 1965). Thus, stachyose has a negative effect on the feed efficiency of the soybean meal. The metabolizable energy of soybean meal is low compared its total energy and this is due to the poor digestibility of the raffinose oligosaccharide (Parsons et al., 2000; Meis et al., 2003). Therefore, development of soybean lines with reduced-stachyose contents is desired for increasing the energy metabolism rate and digestion of soybean meal.

The breeding line V99-5089 has been identified as source of high-sucrose, low-stachyose and low-phytate contents (Saghai-Maroof and Buss, 2008). Results from several studies in which V99-5089 was crossed to other soybean lines suggest that low stachyose is a heritable trait (Jaureguy, 2009; Zeng 2012; Mozzoni et al., 2013). Thus, this breeding line could be used in breeding for soybeans with improved sugar profile. Additionally, crosses between V99-5089 and other low stachyose sources suggested the presence of a stachyose QTL in V99-5089 that is different from the low stachyose sources previously reported (Sebastian et al., 2000; Skoneczka et al., 2009).

Saghai-Maroof and Buss (2008) identified and located the stachyose QTL from V99-5089 in the interval between Satt453 and Sat_331 on chromosome 11. This QTL accounted for 28% of the phenotypic variation observed. Furthermore, Maupin et al. (2011) reported that the

SSR marker Satt453 on chromosome 11 had 87% of selection efficiency in marker assisted selection of low stachyose lines. Additionally, Jaureguy (2009) identified three SSR markers on chromosome 10 linked to stachyose content. He also reported a QTL for this trait flanked between Satt262 and Sat_282 that explained 48% of the stachyose variation observed. More recently Zeng et al. (2012) used SSR and SNP markers on a population derived from the cross Osage x V99-5089 and identified two QTLs for stachyose on chromosomes 10 and 11, which explained 11 and 81% of the phenotypic variation in stachyose contents. These QTLs were stable across environments.

The objective of this study was to confirm stachyose QTL previously reported and / or identify new QTLs.

Materials and methods

Population development and field experiment

An F_2 soybean population segregating for sucrose content was used in this study and both parental lines were developed at Virginia Polytechnic Institute. The regular sucrose (~ 5.0 %) breeding line V97-3000 was crossed to the high sucrose (~ 7.7 %) breeding line V99-5089 in summer of 2007, at the Arkansas Agricultural Research and Extension Center (AAREC) in Fayetteville, AR. The F_1 plants were grown in Fayetteville during summer of 2008 and morphological markers (leaf shape and seed size) were used to differentiate true hybrids from selves and outcrosses, because V97-3000 is a small-seeded line (9 g 100 seed^{-1}) with narrow leaf shape, and V99-5089 is a large-seeded line (20 g 100 seed^{-1}) with broad leaf shape. F_1 plants were then bulked and harvested.

The F_2 population was grown in Fayetteville in 2009, in 3-meter rows with 150 seed per

row. Row spacing was 1 meter and plots were irrigated by furrow irrigation and managed according to the standard cultural practices for soybean production in Arkansas (Tacker and Voires, 1998). A total of 150 random F_2 plants were tagged and identified with a number. At the end of the season, these tagged plants were harvested individually to form the mapping population. A sample containing 100 seed was sent to a winter nursery for generation advancement. The remnant of the seed from each $F_{2:3}$ line was used for carbohydrate analysis.

The winter nursery is located near Upala city in the northwestern part of Costa Rica at 52 meters above sea level (Mongabay, 2014). There, each $F_{2:3}$ line was grown in a 3-m row with 0.76-m row spacing. A 3-m row spacing of the parental genotypes was planted next to the segregating population. Soil was cultivated before planted and plots were fully irrigated and managed during the growing season using standard cultural practices. After two generation advancements, $F_{2:5}$ seed were brought back to Fayetteville but unfortunately, the seed quantity was not enough to grow a test with two repetitions in two different locations. For this reason, it was decided to do a seed increase in Fayetteville during the growing season in 2010 and then establish the test the following year.

In summer of 2011, $F_{2:6}$ plants were grown in a randomized complete block design (RCBD) with two repetitions in two locations: Fayetteville and Marianna, AR. The soil in Fayetteville is classified as Captina silt loam (Fine-silty, siliceous, active, mesic Typic Fragiudults) (Soil Survey Staff, 2013), which is very deep, moderately well drained, and made in a thin mantle of silty material (Soil Series, 2006). Whereas, the soil in Marianna is classified as a Calloway silt loam (Fine-silty, mixed, active, thermic Aquic Fraglossudalfs) (Soil Survey Staff, 2013), which is described as very deep, somewhat poorly drained and developed in thick loess or water reworked loess (Soil Series, 2002). Each soybean plot consisted of two 3-meter rows, each

with 150 seed and a row spacing of 1 meter. Both locations fully irrigated by furrow irrigation and managed based on the standard cultural practices for soybean production in Arkansas (Tacker and Voires, 1998). The youngest fully developed trifoliate of each plant was sampled and samples belonging to the same row were bulked for DNA extraction. $F_{2:6}$ seed were harvested during fall of 2010 and a 10-g sample of each plot was used for carbohydrate analysis and the remnant of the seed was saved to be planted the next season.

Genotypic data

Genomic DNA extraction. After being collected, leaf tissue samples were stored in the freezer at -80 °C. Afterward, they were ground with liquid nitrogen to a fine powder by using a mortar and a pestle. DNA was extracted using the CTAB (cetyltrimethylammonium bromide) buffer method (Kisha et al., 1997) in which, a buffer containing 5M NaCl, 200mM Tris pH 8.0, 4% (w/v) CTAB, 0.5 mM EDTA, and 6.4 mL β-mercaptoethanol is added to the samples, followed by chloroform:isoamyl alcohol (24:1). After incubation for 60 min at 65 °C with occasional gentle mixing, DNA is precipitated and washed with 95% ethanol and subsequently dissolved in 0.1 x TE buffer. Concentration was calculated by measuring the absorbance at 260 nm using a BioTek Power Wave XS Microplate Spectrophotometer (BioTek, Winooski, VT). DNA was stored in the freezer at -80 °C.

Single nucleotide polymorphism (SNP) genotyping. For genetic map construction, two repetitions of each parental line and 92 $F_{2:7}$ DNA samples were sent to the Research Technology Support Facility (RTSF) Genomics Core at Michigan State University, East Lansing, MI. There they were genotyped with 5361 SNP markers (dbSNP-NCBI, 2012) using the Illumina Infinium® Genotyping HD BeadChip (652k SNPs) on Illumina iScan (Illumina, San Diego, CA).SNP analysis was performed on 4-µL samples containing a concentration between 50 and 100 ng μL^{-1}

of DNA. Intensities of the beads fluorescence were distinguished by using the Illumina iScan TM Reader and alleles for each SNP locus were named using Illumina's Bead Studio TM software (Illumina, San Diego, CA, v3.2.23). For each SNP marker, the genotype data represents three possible genotypes AA (homozygote), AB (heterozygote), and BB (homozygote) (Akond et al., 2013).

Phenotypic data

Soluble carbohydrate analysis. Sugar extraction followed the protocol described by Hou et al. (2009), with some modifications. Briefly, a 10-g seed sample was ground to a fine powder using a coffee bean grinder (Krups®, Shelton, CT). Then, the powder was sifted through a 100-µm stainless steel testing sieve (VWR®, West Chester, PA), in order to obtain a sample with uniform particle size. Afterwards, a 0.15-g sample was weighed, mixed with 1.5 mL of deionized-distilled water (ddH_2O) and transferred into a 2 mL centrifuge tube. The tube was vortexed, shaken horizontally for 20 min at 200 rpm and centrifuged at 14000 rpm for 10 min. A 500 µL aliquot from the supernatant was placed in a new 2mL centrifuge tube and 700 µL of acetonitrile (99.9%, HPLC grade) (Thermo Fisher Scientific, Inc.) was added. The solution was mixed by inversion and incubated at room temperature for 30 min. Then, the tube was centrifuged at 14000 rpm for 10 min and 70 µL of the extract was pushed through a 25 mm Easy-Pressure syringe filter holder (VWR®, West Chester, PA) containing a 0.2 µm filter paper disc (Pall Lifesciences, East Hills, NY). Then, vials containing a 24-µL aliquot of each sample extract diluted in 576 µL of distilled water were used for carbohydrate determination in a HPLC.

The anion-exchange HPLC system (Dionex DX500 HPAEC-PAD) was composed of a GS50 gradient pump, an ED40 pulsed amperometric electrochemical detector, an LC50 chromatographic oven, an AS40 automated sampler with a 25-µL injection loop, and a

Chromeleon Chromatography Management Data System. The mobile phase consisted of a 90mMNaOH solvent (VWR®, West Chester, PA) at constant flow rate of 1 mL min^{-1}, prepared by diluting carbonate-free HPLC grade 50% (w/w) stock solution in distilled water, filtered with a 0.45-µm membrane, and degassed with compressed nitrogen gas for 30 min before vials were loaded into the auto sampler, as described by Hou et al. (2009). Soluble carbohydrates were separated by an analytical CarboPac PA-10 pellicular anion-exchange resin column (4 x 250 mm) coupled to a CarboPac PA10 guard column (4 x 50 mm) and preceded by AminoTrap column (3 x 30 mm) (Dionex, Sunnyvale, CA). Sugar content was determined from regression curves fitted from a set of standards for sucrose (Sigma-Aldrich, St. Louis, MO) at different concentrations (10, 20, 40, 60, and 80 µg μL^{-1}). Carbohydrate data then converted to milligrams of sugar per gram of seed (mg g^{-1}) on dry-weight basis.

Statistical analysis

The Shapiro-Wilk (*w*) normality test from JMP 10.0 (SAS Institute, Cary, NC) was conducted on the stachyose content distribution for F_2-derived lines. Associations between seed stachyose and molecular markers were analyzed by single factor analysis of variance (ANOVA) at the 0.05 significant level using the PROC GLM procedure in SAS 12.3 (SAS Institute, 2012). The software JoinMap® 4.1 (Van Ooijen, 2006) was used for construction of linkage maps, with a minimum logarithm of odds (LOD) set to 3.0. A Haldane mapping function (Haldane, 1919) was used to conduct regression mapping algorithm of each chromosome or linkage group. Composite interval mapping analysis (CIM) was performed using Qgene 4.0 (Joehanes et al., 2008. One thousand permutation with a walk speed of 1 cM and experiment-wise $\alpha = 0.05$ was adopted to establish the empirical significance threshold (Churchill and Doerge, 1994). Multiple interval mapping analysis (MIM) was conducted to determine the optimum position of the QTL,

MIM was performed based on the model $c(n) = \ln(n)$ with a walk speed of 1 cM. MapChart 2.2 (Voorrips, 2002) was used to create the LOD plots based on the data from JoinMap® 4.0 and Windows QTL Cartographer 2.5.

Results

Phenotypic data

The Shapiro-Wilk test showed that the seed stachyose content in the mapping population derived from V97-3000 x V99-5089 did not follow a normal distribution in all four environments (Table 1a and Fig. 1). V97-3000 showed higher stachyose content than V99-5089 by 3.2 – 3.7% in all environments studied, as expected (Fig. 1 and Table 1b). Some of the lines in the mapping population had stachyose content lower than the low parent V99-5089 and higher than the high parent V97-3000 (Table 1a). It is likely that the low-stachyose trait is controlled by a few genes with major effects.

QTL mapping in F_2-derived populations by SNP markers

A total of 5361 random SNP markers covering the 20 soybean chromosomes were used to genotype $F_{2:6}$ lines and 1720 SNP loci (32%) turned out to be polymorphic. These markers were mapped on 20 chromosomes, representing 663 unique SNP loci (Table 2). This linkage map spanned 2435.2 cM with an average distance coverage of 3.8 cM per marker (Table 2 and Fig. 2a-e).

Results from the single-marker analysis revealed seven significant SNP markers ($P < 0.05$) on six chromosomes (2, 5, 9, 10, 11, and 19) associated with seed stachyose content (Table 3). Lines with a homozygous genotype carrying the V99-5089 allele had lower stachyose content compared with those carrying the V97-3000 allele. On average, the V99-5089 allele accounted

for a reduction of stachyose content between 0.55 and 1.71% (Table 4). The stachyose variation contributed by the SNP alleles in those five chromosome regions ranged from 7 to 46%.

In the composite interval mapping analysis, the empirical significance threshold was computed as a LOD value of 2.5 in the $F_{2:6}$ mapping population in all environments studied (i.e., Fayetteville 2009, Fayetteville 2010, Fayetteville 2011, and Marianna 2011). Based on stachyose content in Fayetteville in 2009 and 2010, a minor stachyose QTL was identified on chromosome 10, flanked by the SNP markers ss247176598 and ss247225929, tightly linked to the marker ss2472066709 (Figs. 3a-b). This QTL was tentatively named *Sta1*. *Sta1* had a LOD value of 3.1, significantly higher than the threshold, and accounted for 11% of stachyose variation observed (Table 4 and Fig. 3c). Additionally, based on average stachyose content across two locations and three years, a major stachyose QTL was identified on chromosome 11, flanked by the SNP markers ss247560171 and ss247559991, tightly linked to the marker ss247557383 (Figs. 4a-c). This QTL was tentatively named *Sta2*. *Sta2* had a LOD value of 5.1, significantly higher than the threshold, and explained 46% of phenotypic variation for stachyose content (Table 4 and Fig. 4c). This QTL was consistently identified in four environments with a LOD value between 3.1 and 3.9. The multiple intervals mapping analysis showed the optimum locations of the QTLs identified. *Sta1* was determined to be at 41.8 cM on chromosome 10 and *Sta2* was located at 10 cM on chromosome 11. These results indicate that there are two independent stachyose genes located in two different chromosome regions.

Discussion

In this study, the high stachyose parent ‘V97-3000’ consistently produced high stachyose, while the low parent ‘V99-5089’ produced low stachyose, as expected, in all four environments

(i.e., Fayetteville 2009, Fayetteville 2010, Fayetteville 2011, and Marianna 2011). The average seed stachyose content of the F_2-derived populations was also consistent across the environments. These results indicate that stachyose is a stable trait and that environment has very little impact.

The average stachyose contents for the parents and the progeny were higher in Marianna in 2011, compared with Fayetteville in the same year. Similarly, stachyose content for parents and progeny were higher in Fayetteville in 2010 than those for the same location in 2011 (Tables 1a and 1b). This can be explained by the difference in the average monthly temperature during the growing season for those two years and between the two locations (Table 5). Weather history data showed that Fayetteville, AR experienced higher temperatures in 2010 than in 2011. Previous reports indicate that high temperatures increased stachyose content due to a reduction in sucrose (Bellaloui et al., 2013).

Based on the molecular marker analysis, a major stachyose QTL was found on chromosome 11. This QTL is located close to a QTL previously reported by Saghai-Maroof and Buss (2008) and confirmed by Zeng (2012). My QTL is likely the same QTL previously reported by Saghai-Maroof and Buss (2008) and Zeng (2012). This result was expected because V99-5089 was a common parent for all three studies. As consequence, a major stachyose QTL has been unquestionably confirmed across populations with different genetic backgrounds and across environments. Markers associated with this QTL/gene may be used in marker-assisted selection of low-stachyose lines.

A minor stachyose QTL was mapped to chromosome 10. This QTL is located near the region of a QTL for stachyose previously reported by Jaureguy (2009) and later confirmed by Zeng (2012). It appears that the QTL region identified in the three different studies contains the

same QTL/gene for stachyose in V99-5089. The low-stachyose trait is controlled by two independent genes, one located on chromosome 11 with a large effect and the other one located on chromosome 10 with a small effect. The stachyose QTL on chromosome 11 reduced the stachyose content by up to 2.21 % and the minor QTL on chromosome 10 contributed up to 0.96 % stachyose reduction. Therefore, use of these markers would be very effective in breeding for low-stachyose soybean.

Conclusion

Two stachyose QTLs were identified in our study. A major QTL located on chromosome 11, linked to ss247557383, that explained 46% of the phenotypic variation for stachyose and a minor QTL located on chromosome 10, linked to ss247206709, accounted for 11% of the stachyose variation. Both QTLs are in agreement with stachyose QTLs previously reported and also proved to be stable across environments. The fact that stachyose is controlled by few genes with large effect and also that the environment has little or no effect, confirms that stachyose is a qualitative trait. For this reason, molecular breeding for soybeans with low-stachyose content should be fairly straightforward.

References

Akond, M., S. Liu, L. Schoener, J.A. Anderson, S.K. Kantartzi, K. Meksem, Q. Song, D. Wang, Z. Wen, D.A. Lightfoot, and M.A. Kassem. 2013. A SNP-based genetic linkage map of soybean using the SoyS-NP6K Illumina Infinium BeadChip genotyping array. J. Plant Genome Sci. 1:80-89.

Bellaloui, N., A. Mengistu, and A. Kassem. 2013. Effects of genetics and environment on fatty acid stability in soybean seed. J. Food Nutrition Sci. 4:165-175.

Cristofaro E., F. Nottu, and J.J. Wuhrmann. 1974. Involvement of the raffinose family of oligosaccharides in flatulence. In: Sugars in nutrition (Sipple, H.L. and K.W. McNutt, Eds.) Academic Press. New York. Pp 313-336.

Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine. http://www.ncbi.nlm.nih.gov/snp (accessed February 8, 2014).

Gitzelmann, R., and S. Auricchio. 1965. The handling of soy alpha-galactosidase by a normal and galactosemic child. J. Pediatrics 36:231-238.

Haldane, J.B.S. 1919. The combination of linkage values, and the calculation of distance between the loci of linked factors. J. Genet. 8:299-309.

Hawton, J.D., L.J. Johnston, T.M. Salzer, and G.C. Shurson. 1996. Applications of nutritionally improved corn and soybeans for swine. p. 263-304. In Proc. 57th Minnesota Nutrition Conf., University of Minnesota St. Paul, MN.

Hou, A., P. Chen, J. Alloatti, D. Li, L. Mozzoni, B. Zhang, and A. Shi. 2009. Genetic variability of seed sugar content in worldwide soybean germplasm collections. Crop Sci. 49:903-912.

Hymowitz, T., F.I. Collins, J. Panczner, and W.M. Walker. 1972. Relationship between the content of oil, protein, and sugar in soybean seed. Agron. J. 64:613-616.

Jaureguy, L.M. 2009. Identification of molecular markers associated with seed size, protein, and sugar content in soybean. M.S. Thesis, University of Arkansas, Fayetteville, AR.

JMP®, version 10.0. 2012. SAS Institute Inc., Cary, NC.

Joehanes, R. and J.C. Nelson, 2008. QGene 4.0, an extensible Java QTL-analysis platform. Bioinformatics 24:2788-2789.

Kim, H.K., S.T Kang, J.H. Cho, M.G. Choung, and D.Y. Suh. 2005. Quantitative trait loci associated oligosaccharide and sucrose contents in soybean (*Glycine max* L.). J. Plant Biol. 48:106-112.

Kisha, T., C.H. Sneller, and B.W. Diers. 1997. Relationship between genetic distance among parents and genetic variance in populations of soybean. Crop Sci. 37:1317-1325.

Kuriyama, S., and L.B. Mendel. 1917. The physiological behavior of raffinose. J. Biol. Chem. 31:125-147.

Maupin, L.M., M.L. Rosso, K.M. Rainey. 2011. Environmental effects on soybean with modified phosphorus and sugar composition. Crop Sci. 51:642-650.

Meis, S.J., W.R. Fehr, and S.R. Schnebly. 2003. Seed source effect on field emergence of soybean lines with reduced phytate and raffinose saccharides. Crop Sci. 43:1336-1339.

Mongabay. 2014. http://population.mongabay.com/population/costa-rica/3621153/upala (accessed February 26, 2014).

Mozzoni, L., A. Shi, and P. Chen. 2013. Genetic analysis of high sucrose, low raffinose, and low stachyose content in V99-5089 soybean seeds. J. of Crop Improv. 27:606-616.

Parsons, C.M., Y. Zhang, M. Araba. 2000. Nutritional evaluation of soybean meals varying in oligosaccharide content. Poultry Sci. 79:1127-1131.

Saghai-Maroof, M.A., and G.R. Buss. 2008. Low phytic acid, low stachyose, high sucrose soybean lines. U. S. Patent 0199591 A1.

Sebastian, S.A., P.S. Kerr, R.W. Pearlstein, and W.D. Hitz. 2000. Soybean germplasm with novel genes for improved digestibility. p. 56-74. In: J.K. Drackely (ed.) Soy in animal nutrition. Federation of Animal Sci. Soc., Savoy, IL.

Skoneczka, J.A., M.A. Saghai-Maroof, C. Shang, and G.R. Buss. 2009. Identification of candidate gene mutation associated with low stachyose phenotype in soybean line PI200508. Crop Sci. 49:247-255.

Soil Series. 2006. Captina Series. Rev. LBW/RLT. National Cooperative Soil Survey. U.S.A. https://soilseries.sc.egov.usda.gov/OSD_Docs/C/CAPTINA.html (accessed February 6, 2014).

Soil Series. 2002. Calloway Series. Rev. WAC:WMK:RBH:MAV. National Cooperative Soil Survey. U.S.A. https://soilseries.sc.egov.usda.gov/OSD_Docs/C/CALLOWAY.html (accessed February 6, 2014).

Soil Survey Staff. 2013. Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. http://websoilsurvey.nrcs.usda.gov (accessed February 6, 2014).

Tacker, p., and E. Voires. 1998. Irrigation, p. 42-49. In: Arkansas soybean handbook. University of Arkansas. Cooperative Extension Service.

Van Ooijen, J.W. 2006. JoinMap® 4, software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, Netherlands.

Voorrips, R.E. 2002. MapChart: Software for the graphical prese ın of linkage maps and QTLs. J. Hered. 93:77-78.

Wang Shengchu, C.J. Basten, and Z.-B. Zeng. 2012. Windows QTL Cartographer V2.5. Statistical Genetics & Bioinformatics. North Carolina State University. http://statgen.ncsu.edu/qtlcart/WQTLCart.htm (accessed April 12, 2013).

Zeng, A. 2012. The identification of quantitative trait loci for sucrose and stachyose content in soybean seed. M.S. Thesis, University of Arkansas, Fayetteville, AR.

Zhang, B., P. Chen, C.Y. Chen, D. Wang, A. Shi, A. Hou, and T. Ishibashi. 2008 Quantitative trait loci mapping of seed hardness in soybean. Crop Sci. 48:1341-1349.

Table 1a. Seed stachyose content (% dry-weight basis) of mapping populations derived from the cross V97-3000 x V99-5089 evaluated across two locations and over three years.

Year – Location	Generation	No. Lines	Mean	Range	Prob< W[†]
2009 Fayetteville	$F_{2:3}$	129	2.77	0.11 – 4.19	< 0.0001
2010 Fayetteville	$F_{2:6}$	92	2.95	0.12 – 4.35	< 0.0001
2011 Fayetteville	$F_{2:7}$	92	2.90	0.07 – 4.32	< 0.0001
2011 Marianna	$F_{2:7}$	92	2.94	0.10 – 4.25	< 0.0001

[†] Shapiro-Wilk normality test.

Table 1b. Seed stachyose content (% dry matter) of parents V97-3000 and V99-5089 evaluated across two locations and over three years.

Year-Location	V97-3000		V99-5089	
	Mean	SD†	Mean	SD†
2009 Fayetteville	3.75	0.16	0.39	0.08
2010 Fayetteville	4.11	0.27	0.40	0.12
2011 Fayetteville	3.92	0.08	0.59	0.06
2011 Marianna	4.00	0.11	0.75	0.35
Overall‡	3.94	0.16	0.53	0.15

† Standard deviation.
‡ Sucrose mean and standard deviation across four environments.

Table 2. Summary of SNP markers used for the screening of an F_2-derived population from the cross V97-3000 x V99-5089.

Chr.[†]	LG[‡]	Length (cM)[§]	No. SNP mapped	No. SNP locus located	Average distance (cM) between SNP loci
1	D1a	29.8	23	15	2.0
2	D1b	167.0	49	26	6.4
3	N	23.8	99	23	1.0
4	C1	136.4	77	34	4.0
5	A1	82.4	56	14	5.9
6	C2	152.5	100	49	3.1
7	M	149.1	51	27	5.5
8	A2	135.6	90	51	2.6
9	K	23.1	90	14	1.6
10	O	139.5	59	32	4.4
11	B1	134.0	141	50	2.7
12	H	150.4	37	27	5.6
13	F	110.6	121	43	2.6
14	B2	48.0	33	16	3.0
15	E	127.1	157	62	2.1
16	J	118.2	90	34	3.5

Table 2. Summary of SNP markers used for the screening of an F_2-derived population from the cross V97-3000 x V99-5089 (Cont.).

Chr.[†]	LG[‡]	Length (cM)[§]	No. SNP mapped	No. SNP locus located	Average distance (cM) between SNP loci
17	D2	164.9	124	44	3.7
18	G	133.9	68	44	3
19	L	309.6	158	40	7.7
20	I	99.3	97	18	5.5
Average		121.8	86	33	3.8
Total		2435.2	1720	663	

[†]Chromosome.
[‡]Linkage group.
[§]Chromosome length in centimorgans.

Table 3. Single-marker analysis of variance for seed stachyose content in 92 F_2-derived lines from the cross V97-3000 x V99-5089 evaluated in Fayetteville (Fay) in 2009 and 2010, and in Fayetteville and Marianna (Mar) in 2011.

Chr.[†]	SNP Marker	Position (cM)	P value				
			2009	2010	2011		Combined[‡]
			Fay	Fay	Fay	Mar	
2	ss244874473	157.0	0.0051	0.0045	0.0079	0.0100	0.0056
5	ss245737602	46.1	0.0257	0.0385	0.0863	0.0647	0.0445
9	ss246816196	18.5	0.0469	0.0232	0.0125	0.0226	0.0218
10	ss247206709	41.8	0.0176	0.0232	0.0167	0.0340	0.0199
11	ss247571761	2.8	0.0167	0.0040	0.0076	0.0268	0.0099
11	ss247557383	10.0	0.0332	0.0253	0.0401	0.0278	0.0316
19	ss250262655	293.0	0.0347	0.0668	0.0457	0.0193	0.0358

[†] Chromosome.
[‡] Pooled data from all environments studied.

Table 4. Mean effect of SNP marker alleles on seed stachyose content in 92 F_2-derived population developed from the cross V97-3000 x V99-5089 grown in two locations and over three years.

Chr.	SNP	Position (cM)	2009			2010			2011†			Combined‡			R^2
			P1§	P2¶	Diff.#	P1	P2	Diff.	P1	P2	Diff.	P1	P2	Diff.	
2	ss244874473	157.0	2.85	2.23	0.62	2.79	2.06	0.73	2.82	2.12	0.70	2.82	2.14	0.68	0.07
5	ss245737602	46.1	2.90	2.35	0.55	2.84	2.10	0.74	2.85	2.24	0.61	2.86	2.23	0.63	0.06
9	ss246816196	18.5	2.87	2.23	0.64	2.81	2.12	0.69	2.83	2.16	0.67	2.84	2.17	0.67	0.06
10	ss247206709	41.8	3.77	2.81	0.96	3.42	2.60	0.82	3.60	2.72	0.88	3.60	2.71	0.89	0.11
11	ss247571761	2.8	2.50	1.36	1.14	2.89	1.48	1.41	2.67	1.33	1.34	2.69	1.39	1.30	0.28
11	ss247557383	10.0	3.87	1.74	2.13	3.91	1.82	2.09	3.85	1.64	2.21	3.88	1.73	2.15	0.46
19	ss250262655	293.0	2.90	2.27	0.63	2.83	2.08	0.75	2.81	2.19	0.62	2.85	2.18	0.67	0.06

† Pooled data from Fayetteville and Marianna in 2011.
‡ Pooled data from Fayetteville in 2009 and 2010, and Fayetteville and Marianna in 2011.
§ Allelic effect of P1 = V97-3000.
¶ Allelic effect of P2 =V99-5089.
# Allelic difference.

Table 5. Average monthly temperatures (°C) in Fayetteville, AR in 2010 and Fayetteville and Marianna, AR in 2011.

Year	Location	Month											
		Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
2010	Fayetteville	0.6	0.6	7.8	15.0	19.4	25.6	26.1	27.8	21.7	14.4	8.9	2.2
2011	Fayetteville	0.6	3.3	10.0	15.0	17.8	25.6	28.9	27.8	18.3	14.4	10.0	4.4
	Marianna	3.3	5.6	11.7	18.3	21.1	27.2	28.3	27.2	21.1	15.6	12.2	7.2

Information extracted from http://www.wunderground.com (accessed March, 2014).

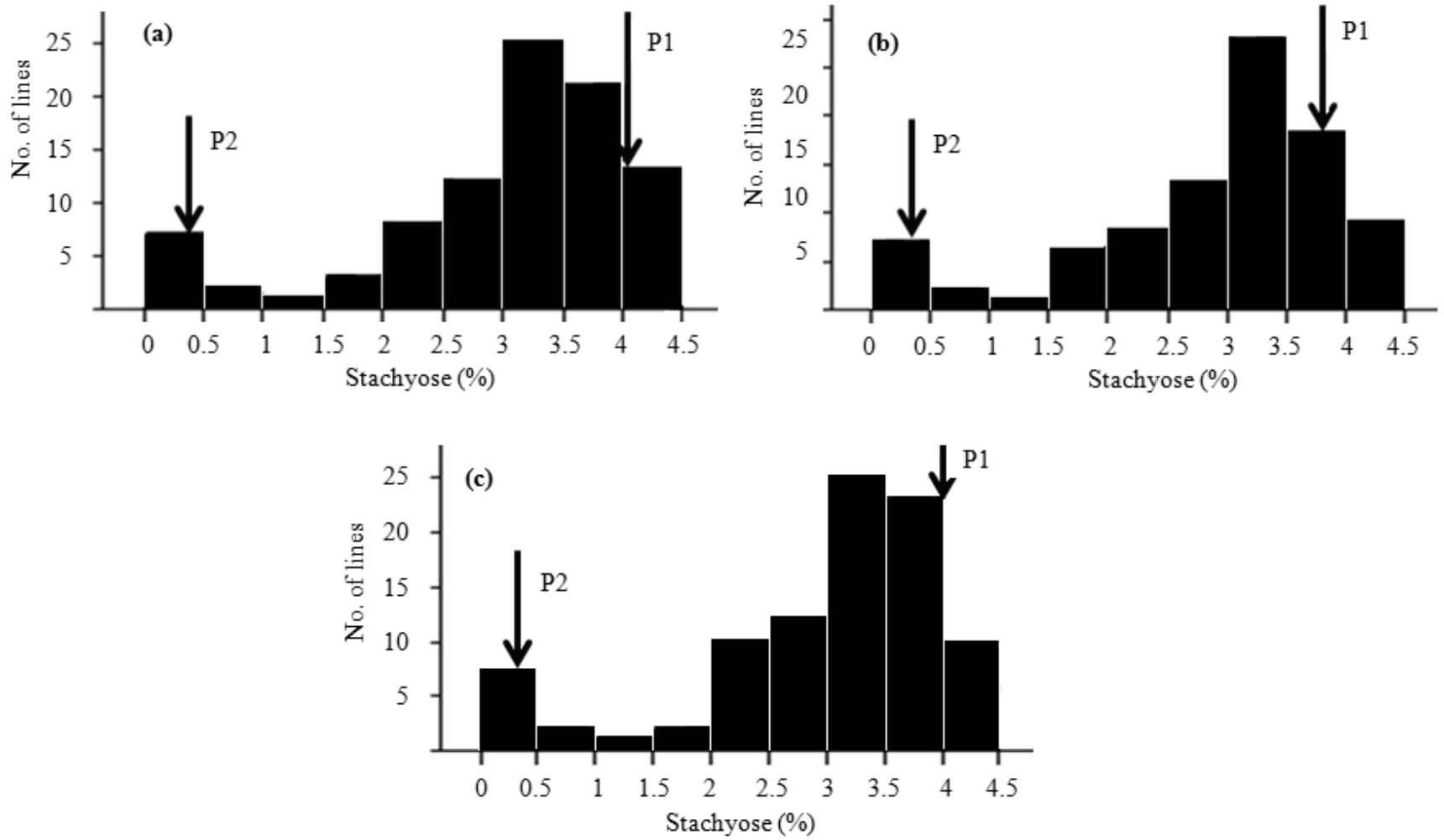

Fig. 1. Frequency distribution of seed stachyose content in populations derived from V97-3000 (P1) x V99-5089 (P2) evaluated in three environments: (a) $F_{2:6}$ lines in Fayetteville, AR in 2010; (b) $F_{2:7}$ lines in Fayetteville, AR in 2011; (c) $F_{2:7}$ lines Marianna, AR in 2011.

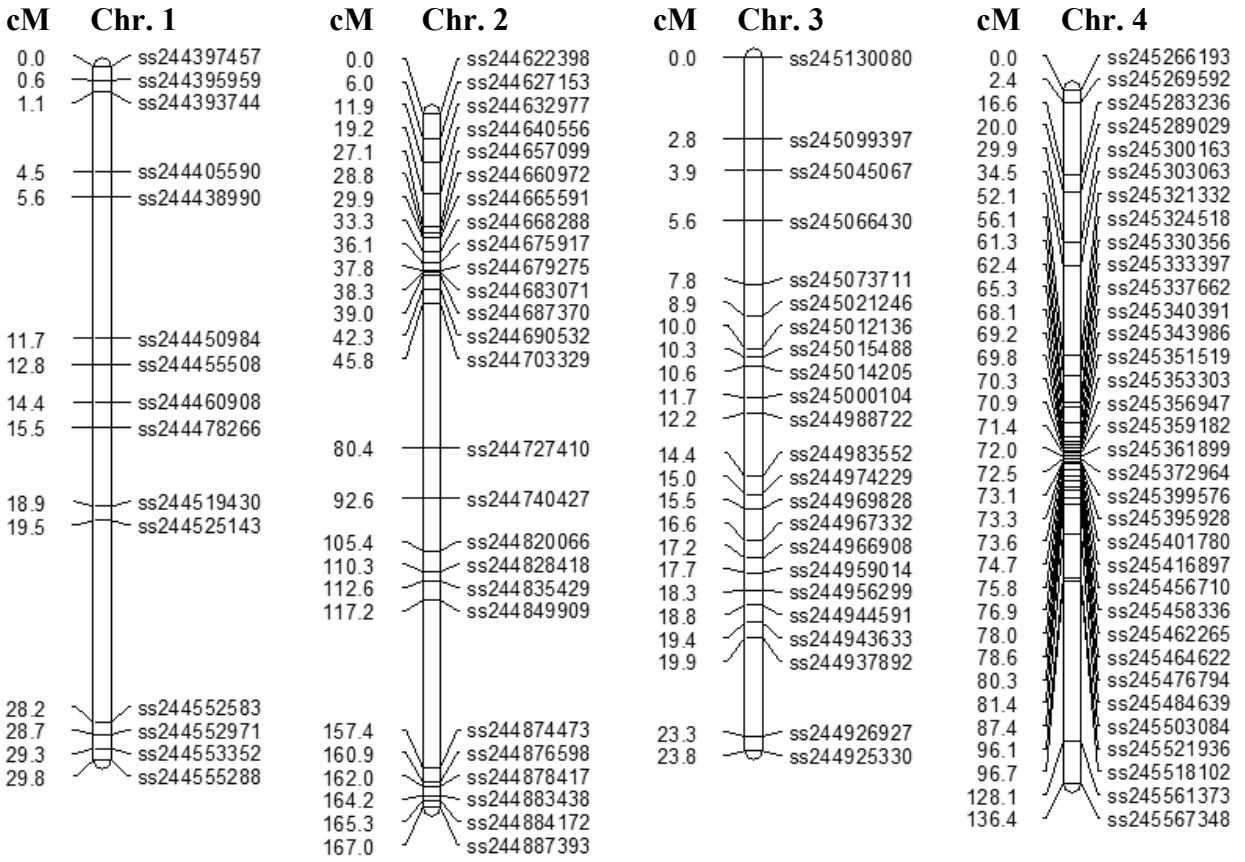

Fig. 2a. A genetic map constructed for chromosomes 1, 2, 3, and 4 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes.

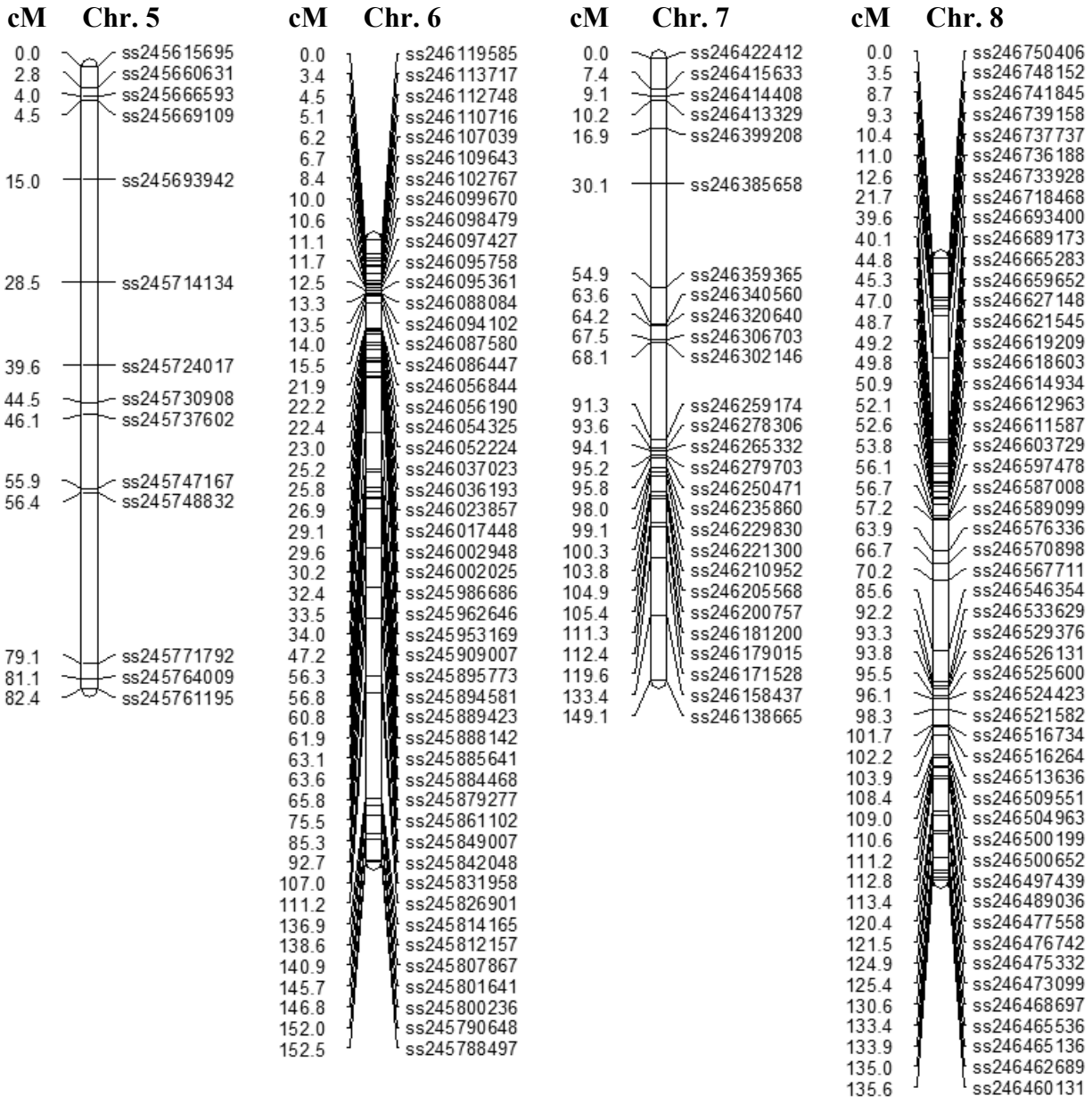

Fig. 2b. A genetic map constructed for chromosomes 5, 6, 7, and 8 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes.

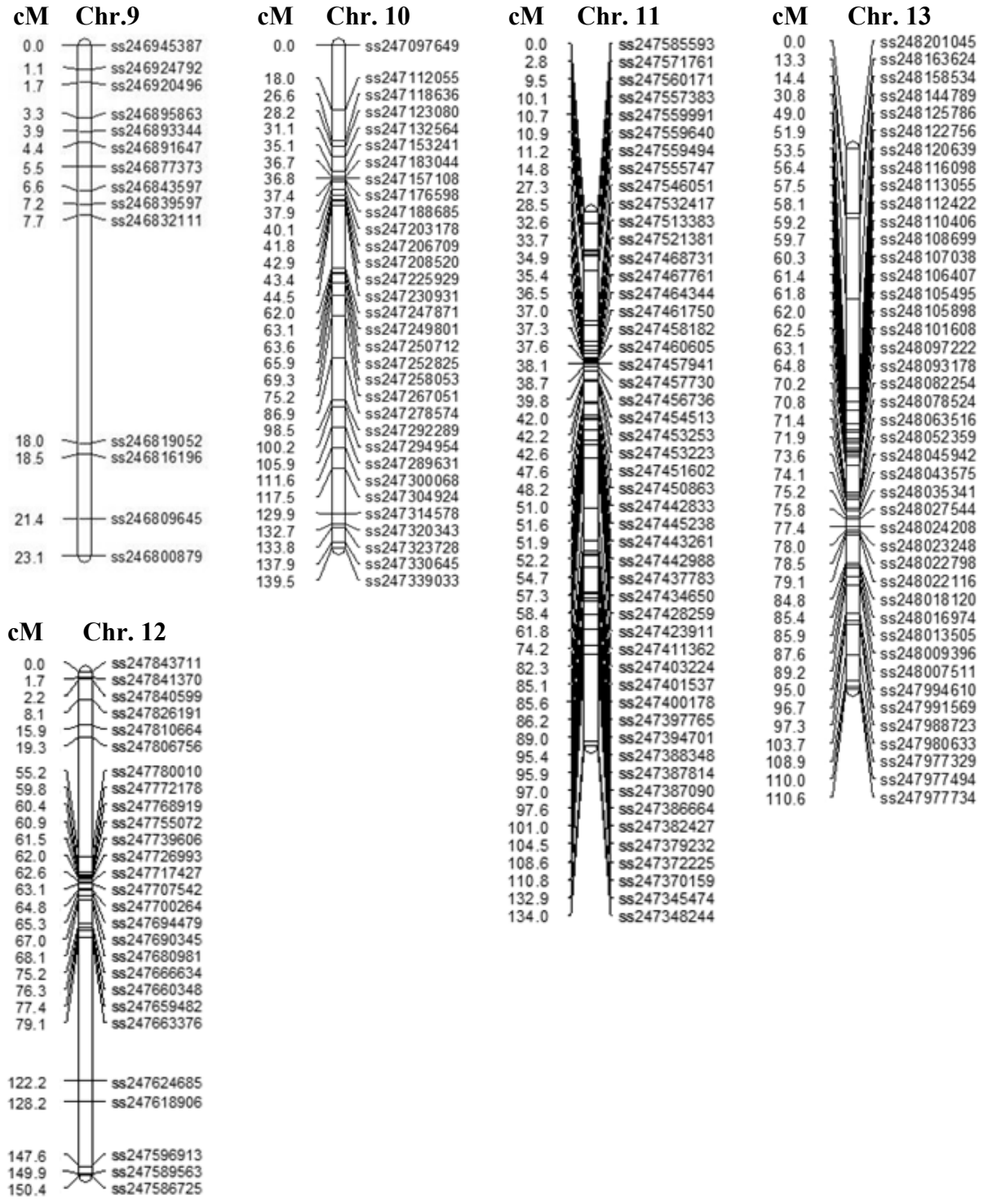

Fig. 2c. A genetic map constructed for chromosomes 9, 10, 11, 12, and 13 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes.

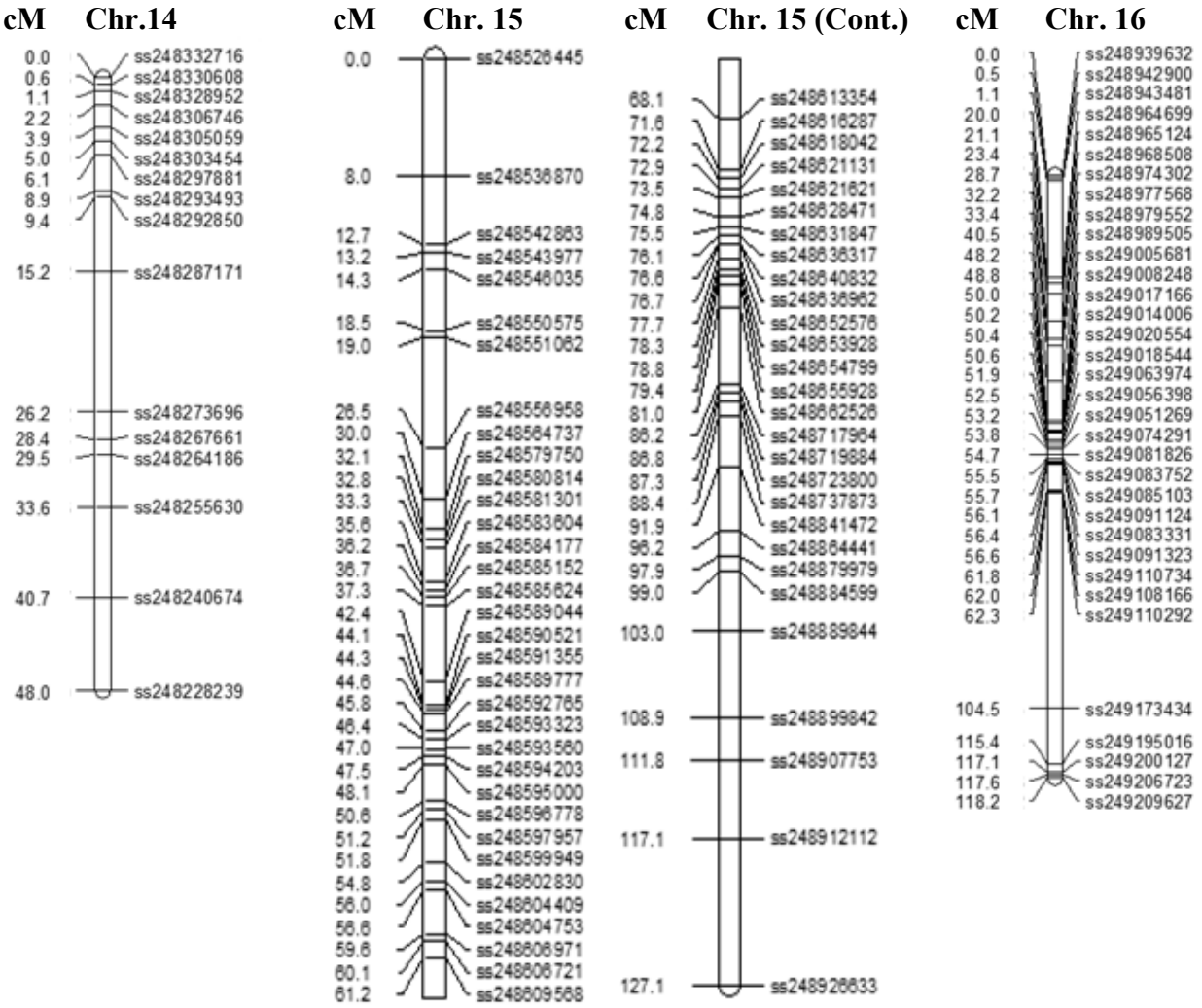

Fig. 2d. A genetic map constructed for chromosomes 14, 15, and 16 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes.

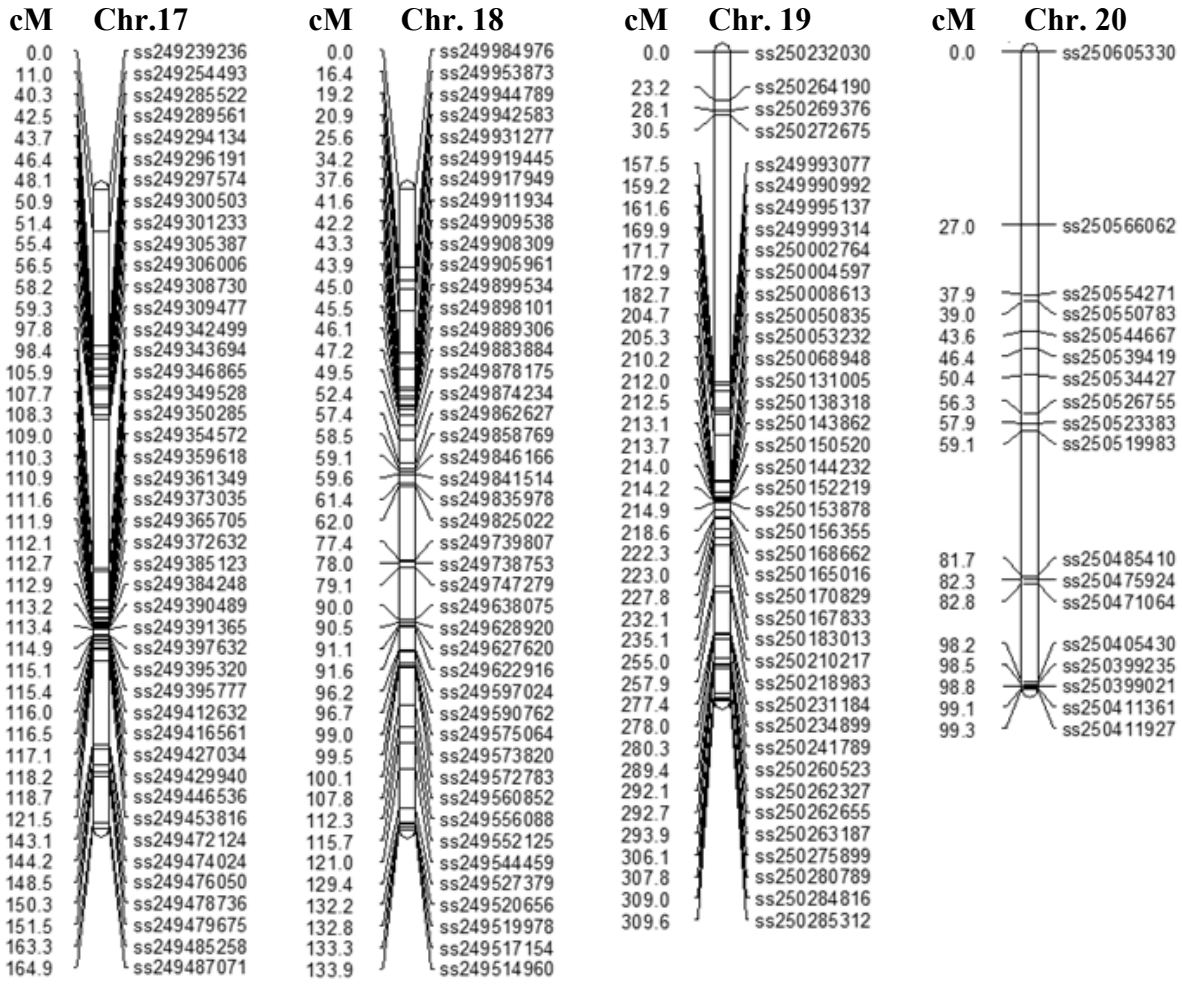

Fig. 2e. A genetic map constructed for chromosomes 17, 18, 19, and 20 using an $F_{2:6}$ mapping population derived from V97-3000 x V99-5089. A total of 1720 polymorphic SNP markers were mapped to 20 soybean chromosomes.

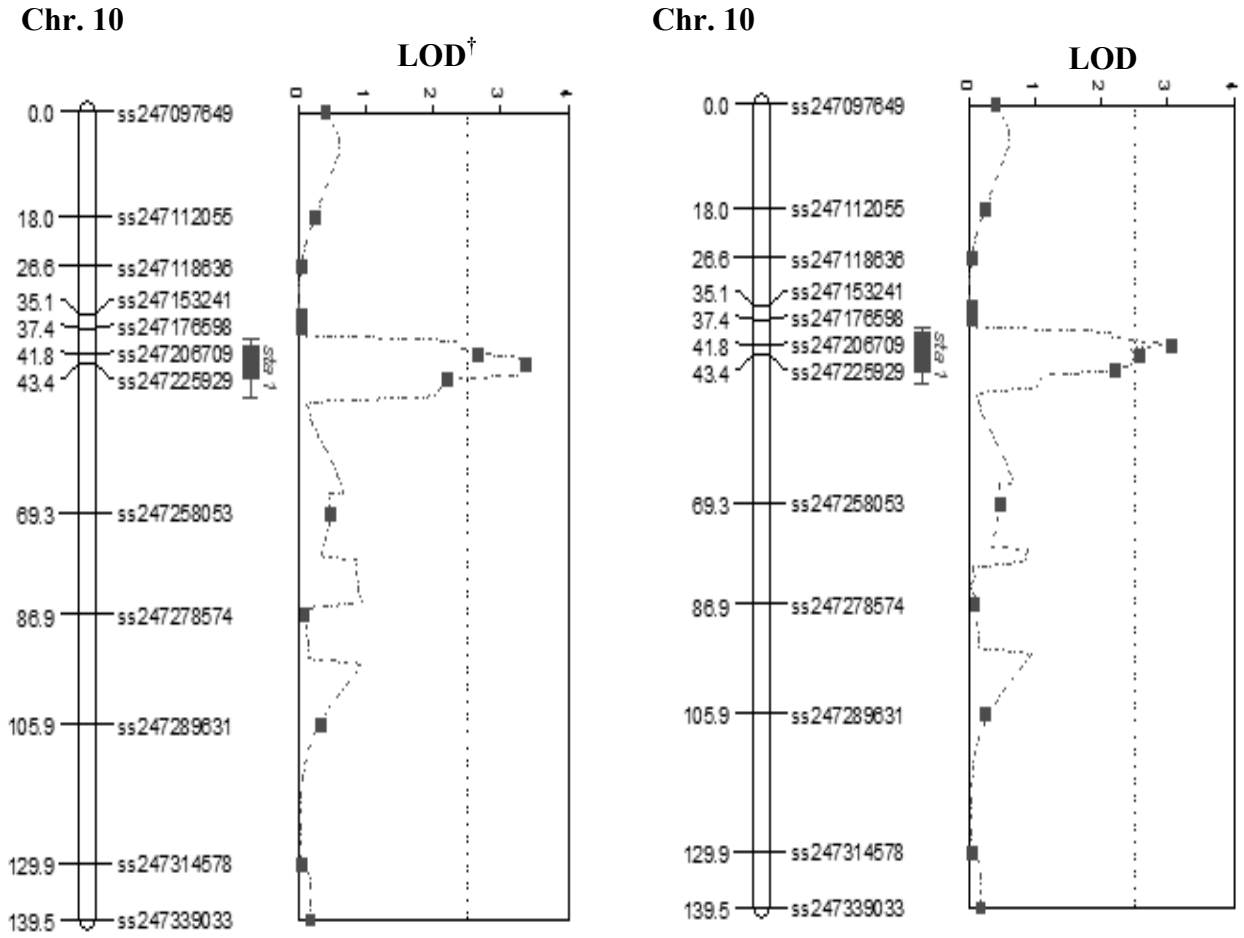

Fig. 3a. Composite interval mapping using SNP markers for seed stachyose QTL on chromosome 10 in 92 F_2-derived lines from the cross V97-3000 x V99-5089, evaluated across locations and years: (1) in Fayetteville, AR in 2009; (2) in Fayetteville, AR in 2010.
† LOD = logarithm of the odds.

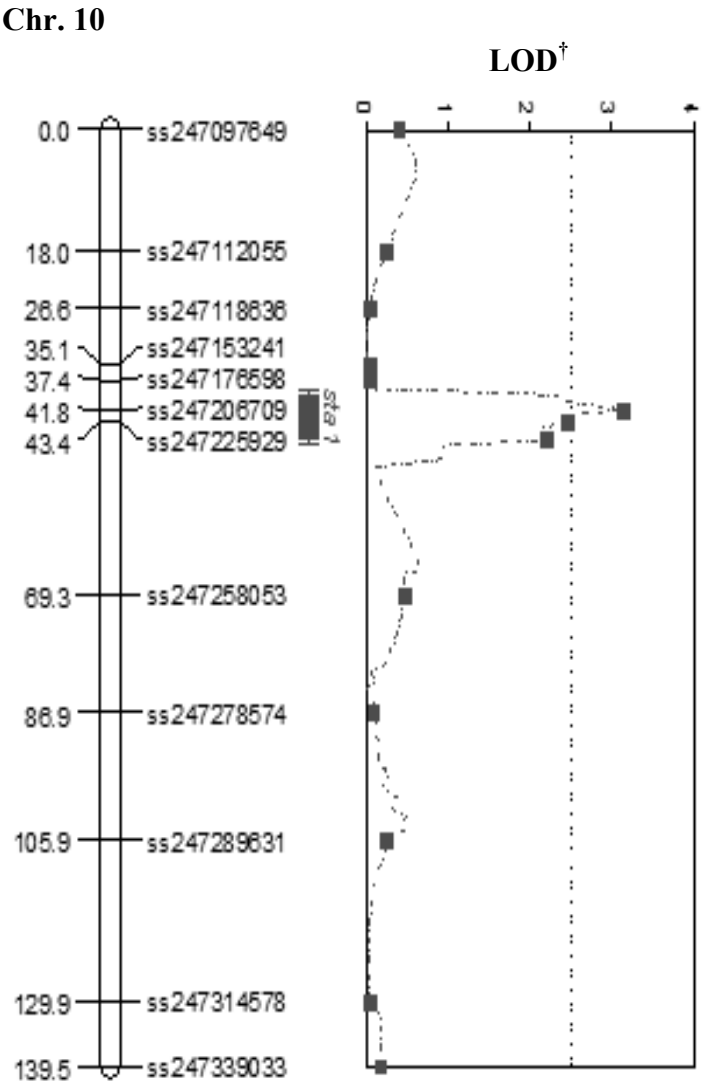

Fig. 3b. Composite interval mapping using SNP markers for seed sucrose QTL on chromosome 10 in 92 F_2-derived lines from the cross V97-3000 x V99-5089 grown in Fayetteville in 2009 and 2010 (Combined data).

† LOD = logarithm of the odds.

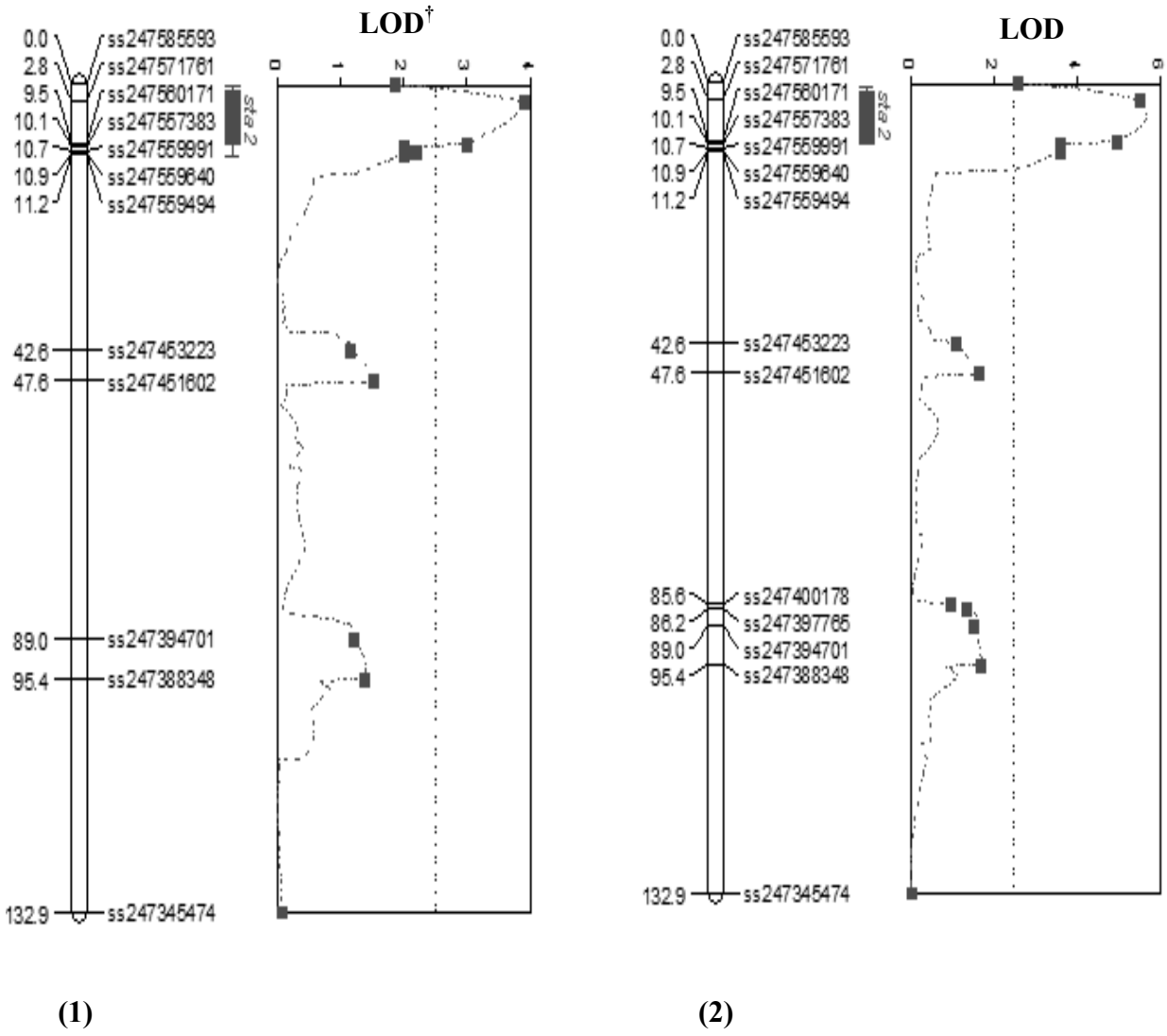

Fig. 4a. Composite interval mapping using SNP markers for seed stachyose QTL on chromosome 11 in 92 F_2-derived lines from the cross V97-3000 x V99-5089, evaluated across locations and years: (1) in Fayetteville, AR in 2009; (2) in Fayetteville, AR in 2010.
† LOD = logarithm of the odds.

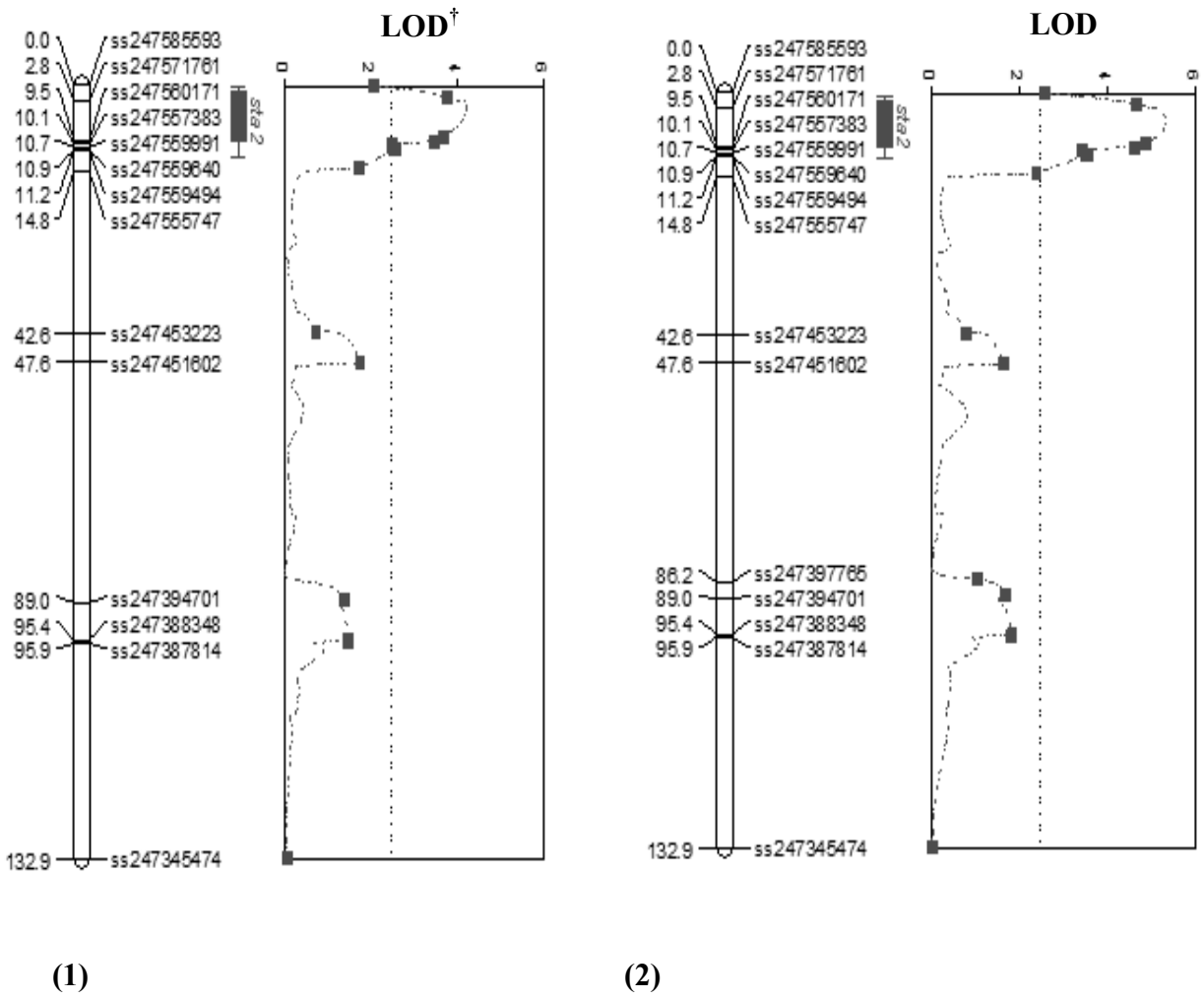

Fig. 4b. Composite interval mapping using SNP markers for seed stachyose QTL on chromosome 11 in 92 F_2-derived lines from the cross V97-3000 x V99-5089, evaluated across locations and years: (1) in Fayetteville, AR in 2011; (2) in Marianna, AR in 2011.
† LOD = logarithm of the odds.

Chr. 11

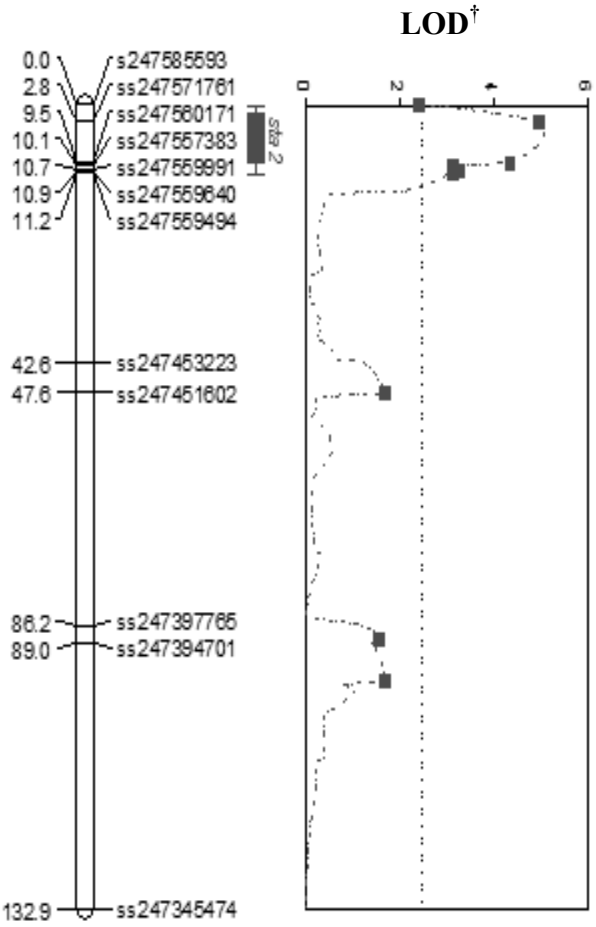

Fig. 4c. Composite interval mapping using SNP markers for seed sucrose QTL on chromosome 11 in 92 F_2-derived lines from the cross V97-3000 x V99-5089 grown in Fayetteville and Marianna in Arkansas in 2011 (Combined data).
† LOD = logarithm of the odds.

Overall conclusion

The increasing demand for soyfoods requires concentrating breeding efforts on the development of soybeans with modified-seed composition. This dissertation particularly focused on studying QTLs/genes associated with sucrose and stachyose contents due to the effects of these sugars on the quality, taste, and digestibility of soyfoods. Findings from this dissertation revealed two independent germplasm sources for low-stachyose that can be used as parental materials for the development of breeding populations. Additionally, QTLs for sucrose and stachyose previously reported, were confirmed and validated in our studies, which makes them reliable molecular tools. These findings will facilitate marker-assisted selection in breeding for high-sucrose and low-stachyose contents.