Fast Neutron Induced Structural Variants and Seed Composition in Soybean Lines

A Thesis
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Austin Andrew Dobbels

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Dr. Robert Stupar, Dr. Seth Naeve

March, 2016

Acknowledgements

I would like to thank all of the people who have helped and supported me in the completion of this thesis. I would first like to thank my advisors, Dr. Seth Naeve and Dr. Robert Stupar, for input, guidance, and for the opportunity to work on these projects. I would also like to thank Dr. Jim Orf for serving on my committee and for all of his input and suggestions for my project. In addition, I thank my funding sources: the United States Department of Agriculture, Minnesota Soybean, National Science Foundation, and United Soybean Board.

I would like to thank all members of both the Naeve Lab and Stupar Lab for their contributions. The lab technicians and MAST students, especially Dimitri Von Ruckert and Jeffrey Roessler, deserve a huge thanks for their help coordinating field plans. I would like to thank Jill Miller-Garvin for help with organization and logistical support. I am also grateful to those in the Stupar lab who helped me complete the sucrose assay and perform DNA extractions and PCR, especially Adrian Stec and Yer Xiong. I would also like to thank Jean-Michel Michno for all his help, especially his bioinformatics help with the BSA project and Ben Campbell was a huge help in making crosses and discussing logistics of my project. This project could not have been completed without all of your hard work.

Lastly, I would like to thank family, friends, and fellow graduate students. Thank you for your continued support of my academic and professional career. Thanks for the hours socializing and helping to alleviate stress when I needed it most and for helping me to become the person I am today.

Abstract

Mutagenesis has been a useful tool in many crop species to create heritable genetic variability for trait improvement and gene discovery. In this study, fast neutron (FN) radiation was used to induce mutations for studying genes that affect soybean seed composition. Phenotypic screening of the FN population at the University of Minnesota has revealed mutant lines with modified seed protein, oil, and sucrose levels. A significant marker for increased oil was found on chromosome 10 near a known deletion using bulked segregant analysis (BSA). A similar BSA approach was used to characterize FN lines with significantly increased sucrose levels. The elevated sucrose content in the seed will improve flavor of soy based products as well as provide more metabolizable energy in animal feed. These FN lines have approximately 9% sucrose on a dry weight basis compared to the FN parental line, M92-220, which contains approximately 5% sucrose. The progeny of four different outcross populations from these mutants have been evaluated for sucrose content using a colorimetric assay. Overall, the populations vary in sucrose content between 3% and 8% sucrose. The DNA from the tails of the phenotypic distributions for each of the populations were pooled and sequenced for BSA. These studies aim to identify novel marker trait associations and provide molecular markers to be used in breeding programs. Two candidate structural variants found through BSA showed evidence of an association with sucrose levels and warrant further exploration.

Table of Contents

List of Tables

List of Figures

Chapter 1 – Literature Review

Overview of soybean, mutagenesis, and bulked segregant analysis

Overview of soybean

Soybean products and seed quality

Soybean [*Glycine max* (L.) Merr.] is a major global crop with great economic importance. It is a legume plant capable of fixing nitrogen using the symbiotic relationship with bacteria in root nodules. Soybeans were domesticated in China approximately 6,000 years ago and were later brought to the United States in the 1700s as a forage crop. Soybean started being grown as a grain crop in the 1920s and 1930s (Bilyeu et al., 2010; Guo et al., 2010). Today, soybean cultivars in the United States are processed and used for their seed protein and oil content.

There are two main soybean types grown and which soybean breeders have focused on developing - food beans and commodity beans. The commodity beans include the largest portion of soybeans grown, while the smaller portion (food beans) is utilized mainly for human consumption. The commodity beans are first processed by solvent extraction separating the crude oil from the defatted meal. The crude oil is refined into edible oil by removing impurities. The resulting oil product can be used for a variety of products, including salad and cooking oil, shortenings, margarines, mayonnaise, and salad dressings. The defatted meal containing the protein portion of the seed is mostly used in feeding animals, such as poultry, swine, beef and dairy cattle, and aquaculture species. Another portion of the defatted meal is used in making protein ingredients such as soy flour, concentrates, isolates, and textured soy protein (Liu, 1997).

The second soybean type, food beans, is grown for human consumption and can be divided into fermented and non-fermented foodstuffs. The fermented food products

include soy sauce, fermented soy paste (miso, Jiang), tempeh, natto, soy nuggets, and sufu. The non-fermented products include tofu, soymilk, yuba, soybean sprouts, okara, roasted soybeans, soy nuts, soy flour, cooked whole mature and immature soybeans (edamame) (Liu, 1997). The end use of the soybeans dictates the cultivar of soybean to be grown and how it is processed.

Importance of sucrose in soybeans

Soybean seeds of modern cultivars are comprised of approximately 40% protein, 20% oil, 35% carbohydrates, and 5% ash on a dry weight basis (Liu, 1997; Akond et al., 2015). The carbohydrates in the seed can be divided into soluble carbohydrates and insoluble carbohydrates based on their solubility in water or dilute alcohol. The major soluble carbohydrates in soybean seeds are sucrose, raffinose, and stachyose - which make up more than 99% of the soluble carbohydrate portion (Hymowitz et al., 1972). The insoluble carbohydrates, found mainly in cell walls, consist of cellulose, hemicellulose, pectin, and small amounts of starch (Liu, 1997).

In the history of soybean breeding, much attention has focused on improving the protein and oil composition of the soybean seed; however, there is value to improving the profile of the soluble carbohydrate portion of the seed. The ultimate goal would be to increase the amount of the digestible carbohydrate sucrose, while decreasing the amount of both raffinose and stachyose. Sucrose is a disaccharide consisting of glucose and fructose bonded by a glycosidic linkage. Raffinose is a trisaccharide composed of glucose, fructose, and galactose; and Stachyose is a tetrasaccharide composed of glucose, fructose and two galactose units. Both raffinose and stachyose are α- galactosides

belonging to the raffinose family of oligosaccharides (RFO's) and require the enzyme α-galactosidase to be digested. Since humans and other animals lack this enzyme, these RFO sugars get fermented in the lower intestinal tract, producing carbon dioxide, consequently leading to undesirable flatulence (Montelongo et al., 1993).

Improving the composition of carbohydrates in soybean seeds would be very beneficial to improving the quality of animal feed among other soy products. The carbohydrate portion of the seed mainly contributes calories to diets of animals (Montelongo et al., 1993). In addition, elevating the amount of sucrose in soybean seeds will increase the amount of metabolizable energy available in animal feed. The metabolizable energy is the amount of energy available to the animal. Elevated sucrose levels in soybeans are also beneficial in soy food products such as edamame, tofu, and soymilk as the extra sucrose adds a sweet taste that consumers often prefer.

Relationship between carbohydrates and other seed composition traits

Correlations among seed composition traits have been previously reported. A negative correlation exists between protein and oil, such that an increase in either protein or oil will result in a decrease of the other (Diers et al., 1992). In addition, there is a negative correlation between protein and sucrose content and a positive correlation with oil content and sucrose content (Kim et al., 2005) (Cicek, 2001). Strong negative correlations have also been recorded between sucrose and both raffinose and stachyose content (Mozzoni et al., 2013).

Previous Sucrose QTL mapping experiments

Many studies have been conducted to identify quantitative trait loci (QTL) controlling sucrose content. As of November, 2015, Soybase has reported 34 QTL for seed sucrose (Soybase.org). Many of these studies have included bi-parental linkage mapping populations.

The most recent study performed by Akond et al. found three QTL for seed sucrose content on chromosome 3, 9, and 15 in a MD96-5722 by Spencer RIL population (Akond et al., 2015). The RIL population consisted of a set of 92 $F_{5:7}$ lines which were genotyped using 5,376 SNP markers and phenotyped using near infrared reflectance (NIR). In the same population, Akond et al. also found several QTL for both raffinose (7 QTL) and stachyose (4 QTL) content.

Zeng et al., 2014 found three seed sucrose QTL in $F_{2:3}$, $F_{3:5}$, and $F_{3:6}$ populations of a MFS-553 by high sucrose PI line (243545) (Zeng et al., 2014). This study used polymorphic SSR markers in addition to 5,361 SNP markers and phenotyped the populations using high performance liquid chromatography (HPLC). The three seed sucrose QTL were mapped to chromosome 5, 9, and 16, accounting for a total of 64% of the phenotypic variation (Zeng et al., 2014).

Kim et al., (2005) identified 11 potential markers individually accounting for 3.7-17.5 percent of the variation. This was accomplished by analyzing $F_{2:10}$ RILs developed from a cross between two soybean cultivars (‘Keunolkong’ and ‘Iksan10’) and using 199 soybean SSR markers (Kim et al., 2005).

Cicek (2001) found five QTL for seed sucrose content in RIL populations of a *Glycine max* by *Glycine soja* hybridization (Cicek, 2001). Maughan et al. 2000 developed a population by making crosses between a high sucrose *G. max* line and a low sucrose PI line. Maughan et al. analyzed 149 $F_{2:3}$ lines for sucrose content and RFLPs, SSRs, and RAPDs were used as polymorphic markers. In total, Maughan et al. (2000) identified 7 QTL for sucrose content (Maughan et al., 2000). QTL identified in these mapping studies can be used in marker assisted selection breeding programs to assist in developing lines with higher concentrations of sucrose.

Methods to analyze soybean carbohydrate content

Many methods exist to analyze the carbohydrate content in soybean seeds including Near-Infrared (NIR) spectroscopy, gas chromatography (GC), high performance liquid chromatography (HPLC), enzymatic assays, and high performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD). Each method has limitations based on the amount of seed required, time to analyze each sample, cost of analysis, and reproducibility of the results.

All methods used to analyze sucrose content first require an effective sugar extraction. Successful sugar extraction has been achieved by using 50-80% ethanol or water as a solvent. Giannoccaro et al. studied the effect of solvent, temperature, extraction time, and the effect of defatting on extraction of sugars (Giannoccaro et al., 2006). In that study, they showed that each solvent was effective at extracting soluble sugars, but optimal conditions were found when using water as a solvent. Although water was found effective in that study, many other studies have shown success using

80% ethanol as a solvent (Campbell et al., 1999; Bailly et al., 2001; Cicek, 2001; Kumar et al., 2010; Teixeira et al., 2012). After solvent extraction, the amount of sucrose in the seed can be assessed using one of the above-mentioned methods.

HPLC and GC have been well-established methods to produce reliable sucrose estimates. HPLC, as suggested by its name, is a column based approach that distinguishes compounds based on the interaction between the column (stationary phase) and the solvent (mobile phase). This interaction dictates the retention time which is analyzed by the detector and output to the computer. HPLC is often used to characterize proteins, isoflavone content, sugar content, and phospholipids (Liu, 1997). GC is another chromatography based approach used to identify and analyze organic and inorganic compounds. Although these methods are effective techniques to analyze sucrose content, the methods are time consuming and cost prohibiting.

NIR spectroscopy has been used in many soybean breeding programs to quickly and cheaply analyze the amount of protein, oil, moisture, fatty acids, and carbohydrates in the seed. NIR spectroscopy can be used on ground soybeans or whole soybeans. Whole soybeans, while useful, provide less accurate estimates of sucrose content. The NIR method is fast and cheap, but its accuracy with small samples using a breeder's cup is much less reliable. Naeve et al. (2008) suggested that with small samples, there is much less precision and requires destructive sampling for more accurate measurements (Naeve et al., 2008). While NIR may be useful for analyzing sucrose content, it appears that a large enough sample, capable of being ground, would be required for an accurate measurement.

Enzymatic assays are another method that can be used to analyze sugars in soybeans (Campbell et al., 1999; Maughan et al., 2000; Teixeira et al., 2012). Enzymatic analysis procedures require specific enzymes based on the sugar being quantified. Examples of enzymes used in sucrose determination include invertase, glucose oxidase, phosphoglucoisomerase, and alpha-galactosidase (Kumar et al., 2010; Teixeira et al., 2012). The enzymatic assays are typically made into kits for easy end use and reliability. This in turn leads to a fairly accurate, high throughput, and cost effective way to analyze sugars in soybean seeds.

The soybean genome and genetic diversity

Soybeans are diploid with 20 chromosomes (2n=40). The 1.1 gigabase soybean genome includes 46,430 genes. Most genes (~75%) have multiple copies due to soybean genome duplications, which occurred at approximately 59 and 13 million years ago (Schmutz et al., 2010). Soybeans, compared to most crops, have limited genetic diversity which offers challenges to crop improvement.

Soybeans have undergone several bottlenecks, leading to a reduction in genetic diversity (Hyten et al., 2006). Roughly 85% of North American soybean germplasm relates back to just 17 ancestors (Gizlice et al., 1994). In comparison, soybeans have almost 10 fold less nucleotide sequence diversity compared to maize, one fourth that of Barley, and half that of rice (Hyten et al., 2006).

There currently exists a germplasm collection of over 19,000 soybean accessions consisting of both domesticated and wild soybeans introduced from 84 countries or developed in the United States (Song et al., 2015). This germplasm, which is a valuable

tool to the soybean community, is held in a USDA germplasm bank located at the University of Illinois. Though it is beneficial to improve soybean varieties through the utilization of this germplasm bank, other efforts such as mutation breeding have proven to be beneficial as well.

Mutagenesis

Mutagenesis overview, mutagen sources, and previous mutant populations

A mutation in plants can be defined as "any heritable change in the genetic material not caused by recombination or segregation" (Harten, 1998). These mutations have been used in many crop species as sources of variability by plant breeders for trait improvement (Fehr, 1987).

In nature, mutations occur spontaneously at a very low frequency; however, mutagens can be used to generate a wide variety of novel mutations. These mutations are typically random throughout the genome and can result in large scale deletions and duplications or can cause point mutations affecting single nucleotide pairs. The type and degree of mutation depends on the type of mutagen, the dosage, and the time of exposure to the mutagen. Two main types of mutagen sources exist – physical and chemical. Physical mutagens include: UV, X-ray, gamma ray, alpha ray, beta ray, and fast and slow neutrons; and these treatments often result in large scale structural changes. Chemical mutagens include ethyl methanesulfonate (EMS), methyl-methane sulphonate (MMS), hydrogen fluoride (HF), sodium azide, N-methyl-N-nitrosourea (NMU), and hydroxylamine; and these treatments often result in point mutations (Parry et al., 2009).

There are several factors to consider when choosing a mutagenic agent and creating a mutant population. One must consider the goal of the population, whether it be to create novel variation in a species, clone genes, or determine gene function. The mutagenic agent will depend on its availability, cost, mode of action, and usefulness in creating the desired mutation (Fehr, 1987). The mutagen source, dosage, and time of exposure must also be considered. The ploidy of the species being mutagenized, for example, will effect dosage used (polypoid species can handle higher doses of mutagens before the changes become lethal to the plant). The most effective way of determining the appropriate dose and rate is by experimentation. In general, the objective is to use a dosage and rate in which 50% of the M_1 seeds will germinate and produce seed for the next generation (Fehr, 1987).

Before starting a mutagenesis experiment, the researcher should have a plan for how to perform the mutant screen. The population can be screened using a forward and/or a reverse genetics approach. In forward genetic screens, the population is phenotyped for specific traits, and selected individuals are subsequently mapped, genotyped, and/or utilized for other purposes. The opposite approach is applied in reverse genetic screens, where DNA samples from a large portion of the population are investigated for mutations at a specific gene or subset of genes. A mutated subset of individuals can then be analyzed in subsequent phenotype-based experiments to determine the function of that gene.

University of Minnesota Fast-Neutron population development

A soybean Fast-Neutron (FN) mutant population has been developed at the University of Minnesota. The population was initially created to provide a public resource to the soybean community and to conduct both genetic and functional genomics research. Through the advancement of this population, a library of over 27,000 unique soybean mutants was created and further genetic and phenotypic characterization has been carried out on a subset of these lines (Bolon et al., 2011, 2014).

The FN population at the University of Minnesota was initiated in 2007 by bombarding seeds of Minnesota Crop Improvement Association stocks of mid-maturity group 1 cultivar "M92-220." In total, 120,000 seeds were mutated to create the FN population. In the first round, approximately 60,000 seeds of M92-220 were irradiated at the McClellan Nuclear Radiation Center at the University of California-Davis with varying doses of 4, 8, 16, and 32 Gray units (Gy). 5,000 seeds from each dosage were planted in the first season. An additional 60,000 seeds of M92-220 stock were irradiated with 16 and 32 Gy. The resulting M_1 seeds after FN bombardment were planted and grown at the Instituto de Investigaciones Agropecuarias, La Patina, Chile. At maturity, a single M_2 seed from each plant was harvested and later planted in a row by column grid format. The M_2 plants were each tagged with a unique identifier (R##C##) so that advanced generations can be traced back to individual M_2 plants. In addition, young leaf tissue and phenotypes from each of the 27,000 M_2 plants was collected and documented. M_2 plants were independently harvested and M_3 seed collected to form the library of

soybean mutant lines. Many mutants were advanced past this generation; however propagation of these lines is variable among lineages (Bolon et al., 2011).

Phenotypic characterization of the University of Minnesota FN population

The FN population created at the University of Minnesota has been evaluated for a wide variety of traits including seed composition (protein and oil), shoot and root alterations, maturity, and Japanese beetle tolerance. All of the mutants phenotyped have been recorded and posted to soybase (www.soybase.org/mutants). In total, over 23,000 FN mutant lines have been recorded. In addition, the most interesting lines have been advanced and maintained in a core collection consisting of over 500 mutants.

Seed quality mutants from the core collection include lines with $\geq \pm 2$ standard deviations of the population mean for the specific trait. Seed quality traits were initially screened using near infrared (NIR) spectroscopy and include traits such as protein, oil, fatty acids, amino acid profile, and carbohydrates.

The variation of shoot phenotypes was also accessed in the M_2 population via visual evaluation every two weeks. All interesting individuals were noted and photographed. The main variation seen can be characterized broadly, as plants have growth classes for leaf, petioles and branches, flower color, plant height, pods, lodging, and determinance. Some of these mutants, such as those with shorter petioles and those with more pods, could be useful to soybean breeders.

Roots of individuals in the FN population were also observed by growing plants at Becker, MN, a location which contains light, sandy soil ideal for accessing root growth characteristics. Roots were also examined in quartz silica sand in growth chambers. All

roots were examined at the V_3 growth stage. In total, 48 individuals were found with root characteristics such as root size, architecture, and nodulation. These mutants may be useful for drought resistance and nutrient uptake.

Genomic characterization of the University of Minnesota FN population

Bolon et al., 2014 studied the genome-wide consequences of fast neutron radiation among 264 select individuals in the University of Minnesota FN population. The range of deletions and duplications were studied across the genome using a combination of array-comparative genomic hybridizations (aCGH) and next-generation sequencing (NGS) technologies. In addition, associations between structural variants (SV) and potentially useful traits were studied (Bolon et al., 2014).

One objective of the Bolon et al. 2014 study was to survey the Genome-wide patterns of induced genomic SVs. aCGH is an array based approach used to identify SVs such as deletions and duplications throughout the genome. aCGH works by comparing signal intensities of mutant DNA and wild type ('M92-220') DNA hybridized to unique probes based on the 'Williams 82' soybean reference sequence. With this tool, Bolon et al. was able to locate regions of natural heterogeneity in the population, as well as assess the locations and prevalence of deletions and duplications in the soybean genome.

Another goal of the research was to discover SV regions associated with mutant phenotypes. Through genetic mapping, a 1.3 Mb deletion on chromosome 10 was found to be associated with an increase in oil content and a decrease in protein content. In addition, an ~837 kb deletion on chromosome 17 was found to be associated with a short petiole phenotype (Bolon et al., 2011).

A high oil mutant (FN0172932) was characterized through the development of both outcross and backcross populations. The outcross populations were created by crossing the mutant to 'Minsoy' and 'Anoka' soybean lines. The F_2 populations were phenotyped for protein and oil using NIR and 9 individuals from each tail of the phenotypic distribution were pooled for bulked segregant analysis (BSA). The bulked individuals were genotyped using the Illumina GoldenGate Assay consisting of 1,536 SNPs throughout the genome. The BSA identified co-segregation of a previously observed aCGH deletion with the increase in oil and decrease in protein content. The association of this deletion with the phenotype was further confirmed through the development of PCR markers which differentiated individuals homozygous for the deletion, hemizygous for the deletion, or not containing the deletion. The genotyping of the BC_1F_3 population with these PCR markers confirmed the association with the phenotype. The success of this mapping of the high oil mutant is good evidence to support that quantitative traits can be mapped in a FN population by combining the powers of BSA, aCGH, and next generation sequencing.

Bulked Segregant Analysis

Bulked segregant analysis (BSA) is one method in which researchers can identify markers associated with specific traits. BSA is done by phenotyping a segregating population, bulking individuals based on phenotype, and identifying markers that distinguish the two bulks (Michelmore et al., 1991).

BSA was first suggested and used by Michelmore et al. (1991) with the goal of rapidly identifying markers linked to traits of interest. In that study, BSA was used to

map downy mildew resistance genes in lettuce. This was done by screening an F_2 population of 66 individuals segregating for the resistance, pooling DNA from 14-20 individuals in each susceptible and resistant bulk and using RAPD and RFLP markers (Michelmore et al., 1991).

Since Michelmore et al. first reported the successful use of BSA for rapidly identifying large effect QTL, many subsequent studies have carried out similar approaches with success. Examples of previous studies include mostly those in crop species such as mapping disease resistance genes, qualitative mutations, and other complex traits such as flowering time and plant height (Haase et al., 2015). In addition, recent studies using BSA as an approach to map QTL of interest utilize modern technologies such as whole genome sequencing of bulked pools. The first studies combining whole genome sequencing and BSA to identify QTL of interest was done in model species such as yeast (Ehrenreich et al., 2010). Ehrenreich et al. was able to map quantitative traits using very large populations, selection based phenotyping, and quantitative measurement of pooled allele frequencies in a method they termed X-QTL. The success of detecting such QTL can be effective in a model organism such as yeast comprised of a small genome and the ability to generate very large populations. The approach of using while genome sequencing in plants was first demonstrated in Liu et al. (2012) who used a modification of BSA (BSR-Seq) to clone the *glossy*3 (*gl3*) gene of maize (Liu et al., 2012). It was also reported in Takagi et al. who used the approach to map QTL for resistance to rice blast disease and improved seedling vigor (Takagi et al.,

2013). Haase et al., 2015 was able to detect flowering time and plant height QTL in a very large F_2 population of approximately ten thousand individuals (Haase et al., 2015).

Chapter 2 – FN mutagenesis and sucrose composition

Chapter summary

Mutagenesis has been a useful tool in many crop species to create heritable genetic variability for trait improvement and gene discovery. In this study, fast neutron (FN) radiation was used to induce mutations that effect soybean seed sucrose. The elevated sucrose content in the seed will improve flavor of soy based products as well as provide more metabolizable energy in animal feed. Mutant lines were identified that have approximately 9% sucrose compared to the FN parental line, M92-220, which contains approximately 5% sucrose. The progeny of four different outcross populations from these mutants were evaluated for sucrose content using a colorimetric assay. Overall, the F_2 individuals within the populations varied in sucrose content between 3% and 8% sucrose. The DNA from the tails of the phenotypic distributions for each of the populations were pooled and sequenced for bulked segregant analysis (BSA). Two candidate regions were identified that are putatively associated with sucrose, including a 129 kb deletion on chromosome 6 and a reciprocal translocation between chromosome 8 and 13. Further studies need to be performed to identify the novel marker trait associations, which may ultimately lead to the production of molecular markers to be used in breeding programs. The results of this study show the powers of combining next generation sequencing (NGS), array comparative genomics hybridization (aCGH), and BSA to find candidate genomic regions associated with a complex quantitative trait.

Introduction

Soybean [*Glycine max* (L.) Merr.] is a major global crop with great economic importance. There are two main soybean types grown and which soybean breeders have focused on in breeding - food soybeans and commodity soybeans. The commodity soybeans include the largest portion of soybeans grown, while the smaller portion (food soybeans) is processed mainly for human consumption (Liu, 1997).

Soybean seeds of modern cultivars are comprised of approximately 40% protein, 20% oil, 35% carbohydrates, and 5% ash on a dry weight basis (Liu, 1997). The carbohydrates in the seed can be divided into soluble carbohydrates and insoluble carbohydrates based on their solubility in water or dilute alcohol. The major soluble carbohydrates in soybean seeds are sucrose, raffinose, and stachyose - which make up more than 99% of the soluble carbohydrate portion (Hymowitz et al., 1972). The insoluble carbohydrates found mainly in cell walls consist of: cellulose, hemicellulose, pectin, and small amounts of starch (Liu, 1997).

In the history of soybean breeding, much attention has focused on improving the protein and oil composition of the soybean seed; however, there is value to improving the profile of the soluble carbohydrate portion of the seed as well. The ultimate goal would be to increase the amount of the digestible carbohydrate sucrose, while decreasing the amount of both raffinose and stachyose.

Sucrose is a disaccharide consisting of glucose and fructose bonded by a glycosidic linkage. Both raffinose and stachyose are α- galactosides belonging to the raffinose family of oligosaccharides (RFO's) and require the enzyme α-galactosidase to

be digested. Since humans and other monogastric animals lack this enzyme, these RFO sugars are fermented in the lower intestinal tract leading to undesirable flatulence (Montelongo et al., 1993). Increasing the amount of sucrose in the seed will ultimately lead to more metabolizable energy, sweetness, and digestibility to end use products including soyfoods and animal feed.

Soybeans have undergone several bottlenecks leading to a reduction in genetic diversity (Hyten et al., 2006). Roughly 85% of North American soybean germplasm relates back to just 17 ancestors (Gizlice et al., 1994). Though it is beneficial to improve soybean varieties through the utilization of a natural variation, other efforts such as mutation breeding have proven to be beneficial as well.

A soybean fast neutron (FN) population (Bolon et al. 2011 and 2014) was developed as a public resource to the soybean community and to serve as a tool for genetic screens and functional genomics research. Through the advancement of this population, a library of over 27,000 unique soybean mutants was created and further genetic and phenotypic characterization has been carried out on a subset of these lines (Bolon et al., 2011, 2014).

This study aims to combine the powers of bulked segregant analysis (BSA), next generation sequencing (NGS), and array comparative genomic hybridization (aCGH) to map high sucrose trait in an F_2 population. Sucrose is a desirable trait that has often been overlooked in breeding programs due to its difficulty to phenotype high desire for soybean breeders to focus on directly improving the protein and oil fractions of the seed. Although difficulties exist in improving sucrose composition in soybeans, a fast neutron

induced structural variant which significantly increases sucrose should prove useful. In addition, mapping of the structural variant would provide molecular markers to be used in breeding programs.

Materials and Methods

Plant material and population development

Two mutants in the University of Minnesota Fast Neutron (FN) mutant population (Bolon et al., 2011 and 2014) contain elevated levels of sucrose content in mature seed. These mutants were phenotyped over five years and consistently showed the phenotype (~8% compared ~4% (wt), Figure 1), indicating a heritable change induced by fast-neutron radiation. The two high sucrose mutants described both trace back to a single M_2 plant (designated R64C50 and FN0176450) which descended from M_0 seed irradiated with 16Gy of FN radiation. The two mutants were siblings in the $F_{2:3}$ generation and were later given the designations 2012CM7F040P05 and 2012CM7F040P06.

Figure 2 provides a schematic for how mutant lines were advanced. Briefly, single M_2 plants were grown in a grid pattern, the resulting M_3 seed from each individual plant was planted in plant rows and 10 plants were harvested from each row creating families of 10 plants each. Subsequent generations were bulked until the $M_{3:7}$ generation.

The parents for the BSA mapping populations include the $M_{3:7}$ seed of the two high sucrose mutants and outcross parents – Noir 1-SGC-01 (‘Noir 1’) and Minsoy-SGC-01 (‘Minsoy’). Crosses were made between each of the two high sucrose mutants and each of the outcross parents creating a total of four populations. Population designations outlined in Table 1 were as follows: Noir 1 x 2012CM7F040P05 (A13-004), Noir 1 x

2012CM7F040P06 (A13-005), 2012CM7F040P05 x Minsoy (A13-007), and 2012CM7F040P06 x Minsoy (A13-008).

The putative F_1 seed was planted in greenhouses and segregating F_2 populations grown during the summer of 2014 in St. Paul, MN. Eight rows of 30 seeds each were planted in 30 inch rows with a goal to provide at least 200 plants for each of the mapping populations. During the growing season, all F_2 individuals were tagged and fresh leaf tissue collected and freeze dried for later DNA extraction of select individuals. At maturity, the plants were individually harvested and threshed, and the F_3 seed was analyzed for sucrose content using a GOD/invervtase method (Detailed description in Appendix A – Quantification of sucrose method).

Array comparative genomic hybridization (aCGH)

Array comparative genomic hybridization (aCGH) is an array based approach used to identify structural variants (SVs) such as deletions and duplications throughout the genome. In this study, aCGH was performed on $M_{3:8}$ plants that were direct descendants of the mutants used in making crosses (See Figure 2) with the goal of using this data to identify causative SVs within a BSA identified candidate region and to confirm $M_{3:8}$ individuals as sharing overlapping SVs.

aCGH works by comparing signal intensities of mutant DNA and wild type ('M92-220') DNA hybridized to unique probes designed from the soybean reference genome sequence. Similar methods of aCGH were performed previously to physically map deletions and duplications in soybean fast neutron mutants in Bolon et al. (2011, 2014). However, in this study, a new soybean microarray platform was used, developed

by Agilent Technologies, Inc. based on the recently updated soybean genome sequence (version Glyma.Wm82.a2.v1; Song et al. 2016). The microarray was designed with one million probes tiled through the genome, but unevenly spaced to enrich for genic regions. Labeling reactions with mutant (Cy3 dye) and reference (Cy5 dye) were performed with 500ng DNA from both the FN mutant leaf tissue and reference (M92-220-Long) leaf tissue. Labeled DNA was hybridized to the microarray for 66 h at 67°C according to the manufacturer's instructions. Log_2 ratios between the control and mutant hybridizations were calculated for each probe. Labeling, hybridization, washing, and data acquisition were all performed according to manufacturer's protocols. The array design used was the "Agilent Soybean Genome Design – Gmaxv275 phytozomev10." For each CGH run, the aberration algorithm (ADM-2) in the Agilent Genomic Workbench software (Version 7.0.4.0) and Agilent feature extraction (Version 12.0.0.7) were used to extract raw data and call significant aberrations. The parameters were as follows: "minimum number of probes =2, threshold = 6.0, Fuzzy Zero = ON, GC correction = ON, Window Size = 2Kb, Centralization (legacy) = ON, Centralization bin size = 10, centralization threshold = 6.0, diploid peak centralization = OFF, combine replicates (intra array) = ON, feature level filter = glsSaturated, Design level filter = homology =0, and genomic boundary = OFF." JMP®, Version 12 was used to visualize data (Figure 6 and Figure 7).

Quantification of sucrose using colorimetric invertase/glucose oxidase method

The amount of sucrose in the soybean seed was determined using a colorimetric assay based on Teixeira et al. (2012). This assay combines the action of invertase and glucose oxidase (GOD) and is adapted to a 96 well ELISA plate allowing for high

throughput and cost effective analysis of sucrose content on a dry weight basis. The results of this method were compared to wet chemistry lab results from the University of Missouri (UMO) (Agricultural Experiment Station Chemical Laboratories, University of Missouri-Columbia), who use GC to quantify sugar content based on Bhatti et al. (1970) and Janauer et al. (1978) (Bhatti et al., 1970; Janauer and Englmaier, 1978). A set of 33 samples including parents in the mapping population (high sucrose mutants, Minsoy, Noir, and M92-220) along with a P92Y22 reference check were analyzed both with the enzymatic method and at UMO and with GC. A correlation of $r = 0.927$ between our results and UMO results indicate that the method used is reliable and can be used to rapidly screen soybean samples for sucrose concentration (Figure 3). A brief outline of the method is listed below.

Twenty seeds from each sample were ground using a Hamilton Beach chamber coffee grinder, the samples were dried and soluble carbohydrates extracted with 80% ethanol at 70°C. Invertase (invertase from baker's yeast, Sigma-Aldrich) was used to hydrolyze sucrose into glucose and fructose and a glucose assay kit (Glucose assay kit, Eton Bioscience) was used to quantify the glucose in the solution. The solution in each well turned pink based on the concentration of glucose in the sample and this was quantified by analyzing the absorbance at 490nm with a microplate reader and comparing values to a standard reference curve of known sucrose concentrations. See detailed description in Appendix A – Quantification of sucrose method, for further details.

Bulked Segregant Analysis (BSA) –bulking

Two pools of bulks were created for each of the four populations – a high sucrose bulk and a low sucrose bulk. These bulks were created by examining both the distribution of phenotypes in each of the populations as well as the technical replication of every individual sample. Samples that were ranked as highest percentage sucrose in both of the duplicate analyses for sucrose content were pooled in the high sucrose bulks. Similarly, samples that were ranked as lowest percentage sucrose in both of the duplicate analyses for sucrose content were included in the low sucrose bulks. The number of individuals in each bulk is summarized in Table 2.

Sample DNA of selected individuals was prepared from freeze dried leaf tissue from the 2014 growing season. One DNeasy (Quiagen DNeasy ® Plant Mini Kit) was used for each of the eight bulks with equal amounts of plant tissue represented from each individual in the bulk. A total of 20mg freeze dried leaf tissue was used for each bulk and ground using a Quiagen tissue Lyser II. After DNA extraction, bulks of DNA were sent to the University of Minnesota Genomics Center (UMGC) for Illumina Next Generation Sequencing (NGS). Samples were sequenced on a HiSeq 2500 HO 125 paired end run using v4 chemistry. In total, the sequencing facility generated more than 220 million pass filter reads for each lane with average quality scores greater than Q30 for all bulks except for the low sucrose bulk in population A13-005. Due to library preparation errors, this sample was run with 100 bp paired end reads using Rapid chemistry.

Bulked Segregant Analysis (BSA) – Sequencing analysis

The sequencing data quality was checked with FastQC version 0.11.2 before and after sequence data alterations to confirm that tools were working correctly and to ensure data was of good quality for downstream BSA applications. Cutadapt version 1.6 was used to trim adapter sequences and remove low quality reads and Fastx toolkit version 0.0.14 was used to remove low complexity sequences. In this pipeline, BWA mem (v. 0.7.10) was used for mapping using Wm82.a2.v1 reference genome sequence. Variant calling was done using the Genome Analysis Tool Kit (GATK version 3.3 49) Unified Genotyper, calling only SNPs at Wm82.2.v1 positions of the SoySNP50K (Song et al., 2013). A custom python script was used to calculate the allele frequency of the alternate allele at each of the 50K positions. The calculations for these frequencies are based on allele depth ratios at each position.

After allele frequencies were called for each position in all 8 bulks, uninformative data points were removed. These included low quality points where the read depth was less than 10 and non-polymorphic SNPs between the parents. Sites where the allele frequencies in both mutant bulk and wild-type bulk were greater than 0.95 in addition to sites where allele frequencies in both the mutant bulk and the wild type bulk were less than 0.05 were removed. After uninformative data was removed, the data was polarized based on both the mutant parent allele and the wild type (M92-220) parent allele. To calculate the percentage of mutant parent allele in each of the bulks, reference allele frequency was used (1-alt allele frequency) if the mutant parent allele was the same as the

reference allele. The alternate allele frequency was used if the mutant parent allele was the same as the alternate allele.

To verify that allele frequencies were calculated correctly and to verify that reads mapped correctly, Integrative Genomics Viewer (IGV) was used to double check random regions throughout the genome. Figure 4 provides a screenshot of IGV for three different genomic positions. Although variant calling was done with the GATK and allele frequencies calculated with a custom Python script, this helps to ensure that tools were used properly and that data is of good quality.

Results

Mutant characteristics

The two sucrose mutants (2012CM7F040P05 and 2012CM7F040P06) along with many other seed composition mutants have been planted and grown in yield trials over three years (2012 – 2014). On average, the sucrose mutants matured four days later, had a 1 percentage point reduction in protein content, 2 percentage points decrease in oil content, 4 percentage points increase in sucrose content (on a dry matter basis), and a 5bu/acre decrease in yield compared to the wild-type (M92-220). In 2013, a full chemical analysis was done on the sucrose mutants at the University of Missouri. Through this analysis, these mutants were noted as having 4 percentage points increase in linolenic acid and a 4 percentage points decrease in oleic acid.

Phenotypic distributions of the F2 populations

In total, four F_2 populations were developed and phenotyped for sucrose content (Figure 5) using a colorimetric GOD/invertase method. Table 3 shows a summary of

each of the four F_2 populations including the number of individuals phenotyped, the mean sucrose content, and the range in sucrose content. Overall, the populations vary in sucrose content between 3% and 8%. The parents in the population were also advanced and grown as check rows in the summer of 2014. A random sample of bulked seed from these check rows was analyzed for sucrose using the same procedure. The two high sucrose mutants each had sucrose content averaging ~9%. This was significantly higher than the mutant parent line and the outcross parents (Noir 1 ~4.6%, Minsoy ~6.8%, and M92-220 ~5.7%).

aCGH of mutant parents

The aCGH results for the four mutant individuals are diagramed in Figure 6 (two individuals from 2012CM7F040P05) and Figure 7 (two individuals from 2012CM7F040P06). These two figures display the $\log_2$ ratio of each genotype versus the M92-220-Long reference throughout the genome including all 20 soybean chromosomes. Down peaks (colored red) indicate putative deletions while the up peaks (colored blue) indicate putative duplications. Table 4 summarizes all of the major homozygous and heterozygous deletions as detected by aCGH for each of the four individuals. Based on this data, it appears that one of the four individual plants was not related to the others based on SV profile. This difference could have been caused by differential segregation, or more likely, we speculate it was a result of seed contamination from outside sources.

Bulked segregant analysis (BSA) for four high sucrose mutant populations

A high sucrose bulk and a low sucrose bulk were made for each of the four populations. After sequence analysis and filtering, the percent of mutant alleles in the

high sucrose bulk along with the percent of mutant alleles in the low sucrose bulk was graphed for all 20 soybean chromosomes and each of the four outcross populations. Figure 8 (2012CM7F040P05 populations) and Figure 9 (2012CM7F040P06 populations) display allele frequencies, where red lines indicate the percent of mutant allele in the mutant bulk and blue lines indicate the percent of mutant allele in the wild-type bulk (both lines are smoothed to an 11 SNP sliding window). In population A13-007, it was noted that the percent of mutant allele in the wild-type bulk averaged ~0.8 across the genome, likely as a result of a contaminated parent. Because of this, the population was excluded from all further downstream analysis and interpretations.

Throughout the genome, each bulked pool of DNA is expected to have equal ratios of each parental DNA. However, at the site harboring the QTL for the difference in phenotype, we expect a difference in allele frequencies between the high and low sucrose bulks. Visual analysis was done on all 20 chromosomes for every population (see Appendix B – BSA and aCGH graphs by chromosome). A summary table of regions of interest (Table 5) was used to help find possible regions where aCGH regions line up with allele frequency spreads in multiple populations. After analysis of all 20 chromosomes, there are two regions which appear to be associated with sucrose content (Chromosome 6 at 10 MB and Chromosome 8 at 6 MB).

The chromosome 6 association can be seen in Figure 10, which displays the allele frequency of two populations (A13-005 and A13-008) along with the aCGH plot showing a deletion near the region of spread in allele frequency on chromosome 6. In this chromosome, there appears to be an enrichment of the mutant allele in the high sucrose

bulk and a decrease in allele frequency of the mutant allele in the low sucrose bulk. The deletion on chromosome 6 spans from 18,076,434bp to 18,206,053bp, which is ~129kb in size and contains 7 genes.

Chromosome 8 also shows a spread in allele frequency (see Figure 11). A quick glance at aCGH data did not reveal any obvious SVs in this region; however upon further investigation, there is a single CGH probe showing a significant decrease in the $\log_2$ ratio (-6.5) in three of the four genotypes (not present in the putatively contaminated line). This CGH probe is located at 6,358,158 bp and is located in the middle of a single gene (Glyma.08G084300 based on the gene models list from the genome assembly Wm82.a2.v1).

Single CGH probes showing significant signals are often considered noise and not trusted. However, in this case, this probe was cross-validated across three different $M_{3:8}$ mutant plants. To confirm the CGH event, IGV was used to examine paired end whole genome sequencing reads in the region of interest of the mutant DNA (Figure 12). This data revealed that paired reads were mapping between chromosome 8 and 13. Based on IGV images of both chromosome 8 and 13 (Figure 13), it appears that there may be a reciprocal translocation in the mutant where the entire chromosome to the left of the breakpoint on chromosome 8 was broken off in addition to the region to the right of the breakpoint on chromosome 13. It appears that these two pieces of chromosome were reciprocally translocated with each other. PCR primers have been developed, tested, and PCR products Sanger sequenced resulting in the confirmation of the reciprocal

translocation between chromosome 8 and 13 (Appendix C – Primers designed for chromosome 6 deletion, chromosome 8/13 translocation).

Discussion

Bulked Segregant Analysis (BSA)

Bulked segregant Analysis (BSA) is one method in which researchers can identify markers associated with specific traits linked to a genomic region (Michelmore et al., 1991). In this experiment, the goal was to use BSA to identify a FN induced structural variant associated with an increase in sucrose content. The nature of the mutation remains unclear, though it may be recessive, dominant, or additive in nature as in previous FN mutagenesis studies (Hoffmann et al., 2007; Bolon et al., 2011, 2014).

BSA was first suggested and used by Michelmore et al., 1991 with the goal of rapidly identifying markers linked to traits of interest. While RAPD and RFLP markers were used in that study, recent advances in whole genome sequencing technologies facilitate a more novel approach to the old technique. The new technique was first used in plants by Liu et al. (2012) who developed a modification of BSA (BSR-Seq) using RNA-Seq reads to efficiently map and clone the *glossy*3 (*gl3*) gene in maize. Takagi et al. (2013) also used whole genome DNA sequencing based BSA for identifying quantitative traits in rice. They showed that with whole genome sequencing of two bulks of DNA contrasting by the trait of interest, allele frequencies (referred to as SNP-index) could be calculated on each bulk including the contributions of each of the parents to the bulk. BSA using whole genome resequencing also has advantages in that it does not

necessitate marker development and quantitative scores of allele frequencies are more accurate (Takagi et al., 2013).

The goal in this study was to find a region in the genome where the ratio of mutant allele in the high sucrose bulk deviates from the ratio of the mutant allele in the low sucrose bulk. Under an assumption of a recessive mutation and no contamination in the bulks, it is expected that the mutant bulk allele frequency would increase to 100% and wild-type bulk allele frequency would decrease to 33%. Since this is an F_2 population, it is also expected that a fairly large portion of a chromosome would display this spread in allele frequency. At other regions in the genome, where sucrose is not being selected upon, the expected allele frequencies would be 50% in each bulk.

After inspecting all chromosomes, it appears that chromosomes 6 and 8 hold the most promise in being associated with sucrose content, as seen in Figure 10 and Figure 11. While these regions could be associated with sucrose, there is not enough evidence at this time to definitively confirm association; however, it can still be speculated which chromosome may be a better candidate.

Chromosome 6 deletion as a candidate

One of the two candidate regions found through BSA resides on chromosome 6. Along this chromosome, there is an enrichment of the mutant allele in the high sucrose bulk between ~12MB and ~45MB. It is expected that a fairly large portion of the chromosome would be enriched for the mutant allele based on the generation of the mapping population. Since this is an F_2 population, there have not been enough recombination events to resolve the region of interest to a narrower interval.

While the region of interest on chromosome 6 is fairly large (~33MB), other tools can be used to resolve this region to a more narrow interval. Under the assumption that the phenotype was a result of FN irradiation, it is expected that a FN induced structural variant (SV) may be associated with the phenotype. A SV within the region of interest would indicate a good candidate to pursue in depth. Since aCGH was performed on mutant parent material, we can utilize this information to find possible SVs within the larger region. The aCGH data revealed a clear homozygous deletion within the region on chromosome 6 between 18,080,200bp and 18,211,000 bp. While this deletion is in the pericentromere and contains few genes (7 genes deleted), it must still remain a candidate based on allele frequency ratios. PCR primers were developed to differentiate individuals in the population as homozygous for the deletion compared to heterozygous/homozygous for not having the deletion (Appendix C – Primers designed for chromosome 6 deletion, chromosome 8/13 translocation). A future experiment could be conducted to examine co-segregation with the sucrose phenotype. A backcross population (BC_1F_2), in which the high sucrose mutant was crossed to 'M92-220', has a more uniform genetic background and has been developed for this purpose.

Chromosome 8/13 reciprocal translocation as a candidate

In Figure 11, population A13-005, there appears to be a very large spread in allele frequencies on the left portion of chromosome 8. While there does not appear to be an obvious FN induced structural variant based on aCGH, there is a single CGH probe showing a significant decrease in the log2 ratio (-6.5). This CGH probe is located at 6,358,158 bp and hits in the middle of a single gene (Glyma 08G084300).

If the reciprocal translocation is associated with sucrose content, it would be expected that signatures of the translocation would appear in the high sucrose bulk of DNA, but be lacking in the low sucrose bulk of DNA. IGV was used on both the high sucrose bulk and the low sucrose bulk to test this question (Figure 14). It appears that approximately 50% of the reads in the high sucrose bulk have read pairs mapping on chromosome 13 while read pairs are mapping correctly in the low sucrose bulk.

The chromosome 8 gene (Glyma.08G084300) affected in the translocation event could be related to sucrose content. This gene is annotated as 3-ketoacyl-acyl carrier protein synthase I (KAS 1) which is an enzyme crucial for fatty acid biosynthesis. It also plays a role in chloroplast division and embryo development (Schmutz et al., 2010). KAS 1 is involved in one of the conversion steps between sucrose and fatty acids. To date, no one has characterized a mutation of this gene in soybean, nor has anyone characterized a mutation of the paralogous gene copy on chromosome 5 (Glyma.05g129600). If this translocation is related to sucrose, the speculation would be that the amount of sucrose being converted to fatty acids is decreasing, thus accumulating more sucrose in the seed. This finding warrants further wet chemistry analysis of the mutant seeds.

The expression of the gene of interest was also investigated. It is expected that since the phenotype is seen in mature seed, that the gene affected would be expressed in seeds. Data taken from soybase.org RNA-Seq Atlas (Severin et al., 2010) of *Glycine max* was used to confirm this hypothesis, and in fact, the gene is expressed in most tissue

types (see Figure 15). The expression pattern of this gene adds more evidence to this SV being associated with sucrose content.

BSA considerations

While this experiment identified two potential candidate regions, potential complications of the project were considered including: i) the size of the populations, ii) reproducibility of phenotype values, and iii) relatedness of individuals (as seen in aCGH). These parameters ultimately affected the observed spread in allele frequencies and provide insight into better ways to conduct future experiments.

In total, four populations were developed. The experimental plan was to have an average population size of 200 F_2 individuals. To reach this objective, 240 F_2 seeds per population were planted in spring 2014 under the assumption that germination percentage would be around 83%. Unfortunately, germination was low and resulting population sizes averaged 103 individuals. This poor germination was caused mostly by field growing conditions as they were planted along the edge of a field and some were washed away with an early season rainfall. The nature of the mutation may have also effected germination, though germination rates of parents were sufficient in greenhouse and growth chamber settings.

Population size, as a result of poor germination, affected the results of this experiment. Bolon et al., 2014 reported the successful mapping of a high oil mutant. In that study, more than 350 F_2 individuals were harvested and phenotyped which resulted in much more power to differentiate homozygous mutants from wild type plants. Since population sizes in this study were smaller, it is possible that not all individuals in the

bulks were classified correctly, thus leading to individuals being placed into the wrong bulks or heterozygous individuals being bulked. As the number of individuals wrongly placed in bulks increases, the allele frequency spreads between the two bulks will also decrease, making it much harder to detect the location associated with sucrose. While this was a challenge in this experiment, additional information from sequencing data and aCGH was useful in providing extra evidence to support the smaller spreads in allele frequencies. Future studies should consider growing out larger populations if cost and logistics allows.

The GOD/invertase method was used in this study to analyze the percentage of sucrose in the seed of each F_2 individual. Each sample was ground in duplicate and analyzed twice for sucrose content and there was a correlation of 0.68 between duplicate runs, indicating some level of error in the method. While this correlation was not perfect, only extreme sucrose values cross validated across both duplicates were used in bulking. It appeared that the sucrose assay was consistent at differentiating very high values from very low values, but may be difficult to detect minor changes in sucrose content.

In addition to the complication of analyzing sucrose content, the quantitative nature of sucrose levels is also a limitation. Since BSA was done on the F_2 population, only individual plants were phenotyped and not replicated. An advanced generation of the population such as recombinant inbred lines, would allow replicated measurements and be more suitable for detecting QTL. This experiment was performed using an F_2 population as it required less time and we assumed that the phenotype was likely to co-segregate with a large effect FN deletion and not a QTL of minor effect.

The aCGH analysis was performed on two descendants each of 2012CM7F040P05 and 2012CM7F040P06, with the goal of confirming these individuals as being siblings and eventually being able to locate a SV within the genomic region detected by BSA. It is expected that deletions would overlap between all four individuals tested with aCGH. Figure 6 and Figure 7 and Table 4 illustrate that three of the four individuals are likely derived from the same M_2 individual, while the fourth sample appears to differ at many of the major structural variant locations including major homozygous deletions on chromosomes 6, 10, and 14.

The reasoning why one of the four individuals appears to be different from the rest is unknown; however, seed mix-ups during harvesting of individuals is likely the cause. Generations of seed prior to making the crosses were bulk-harvested from a combine. It is very possible that an error occurred in this process, resulting in impure seed. It does appear, however, that both individual plants that were direct decedents of 2012CM7F040P06 appear to be very similar based on aCGH. This similarity may also reveal that the populations A13-005 and A13-008 hold more promise to detect the causative locus, as there is likely less or no contamination in these populations. Due to this information, the primary findings are a result of analyzing only these two populations (A13-005 and A13-008).

In addition to revealing a possible seed contaminant, aCGH also serves to show where the major homozygous deletions are occurring. These homozygous deletions occur on chromosomes 6, 7, 10, and 14 (see Table 4). It is an interesting observation to note that these major deletions occur in the pericentromeric regions of the soybean

genome. These regions have fewer genes compared to the chromosome ends which contain approximately 78% of the soybean genes (Schmutz et al., 2010). This might explain why some of the major deletions are less likely to be candidates for the sucrose content association.

Conclusions

This study has identified two candidate intervals (deletion on chromosome 6 and reciprocal translocation between chromosomes 8 and 13). While we cannot confirm either of these events, based on the evidence gathered, our hypothesis is that the gene being disrupted on chromosome 8 is associated with the increase in sucrose content being observed in this mutant. The gene is annotated as producing a KAS 1 enzyme crucial for fatty acid biosynthesis. Furthermore, it is expressed in most tissue (including seed), and it appears that the high sucrose bulk is showing indications of the translocation in the bulk while the low sucrose bulk is not.

Future work should be done to draw more conclusions of the association of these structural variations to sucrose content. The developed PCR markers (Appendix C – Primers designed for chromosome 6 deletion, chromosome 8/13 translocation) should serve as a way to test various biological questions remaining in these populations. First, it would be interesting to see how this reciprocal translocation segregates in an F_2 population. The recombination at the breakpoints on chromosome 8 and 13 may cause interesting segregation ratios. Second, these markers could be used to test if the translocation co-segregates with the phenotype. The primers could be run on the F_2 population which has already been phenotyped. In addition, a backcross population

(BC_1F_2) was also created by crossing the mutant to M92-220 (wt) and this population was grown in the summer of 2015. This population should be phenotyped for sucrose content and tested for the translocation. Co-segregation could, therefore, be assessed in a more uniform genetic background.

After further testing, if it appears that the single gene affected on chromosome 8 is the main event attributing to the increase in sucrose content, targeted gene editing using tools such as CRISPR/Cas9 could be used to precisely knock out the chromosome 8 gene. Subsequent testing of the CRISPR/Cas9 mutagenized lines compared to non-mutagenized lines may confirm that a loss of function in the candidate gene increases soybean seed sucrose content.

In addition to testing the chromosome 8 event, it is also recommended to perform follow up experiments on the chromosome 6 deletion and a large deletion on chromosome 10. The Chromosome 6 FN induced homozygous deletion resides near the location where allele frequencies spread between the high sucrose bulk and the low sucrose bulk. PCR primers developed to detect this deletion could be used on the existing population to test for co-segregation of the phenotype with the deletion. Running PCR markers for the chromosome 10 deletion would also serve as a control. This deletion is very large and could have some effect on sucrose content. It is possible that multiple structural variants are required to increase the overall sucrose content in the seed.

In conclusion, the combination of NGS, BSA, and aCGH was used to successfully locate two candidate regions associated with soybean seed sucrose content, one of which holds great promise.

Tables and Figures

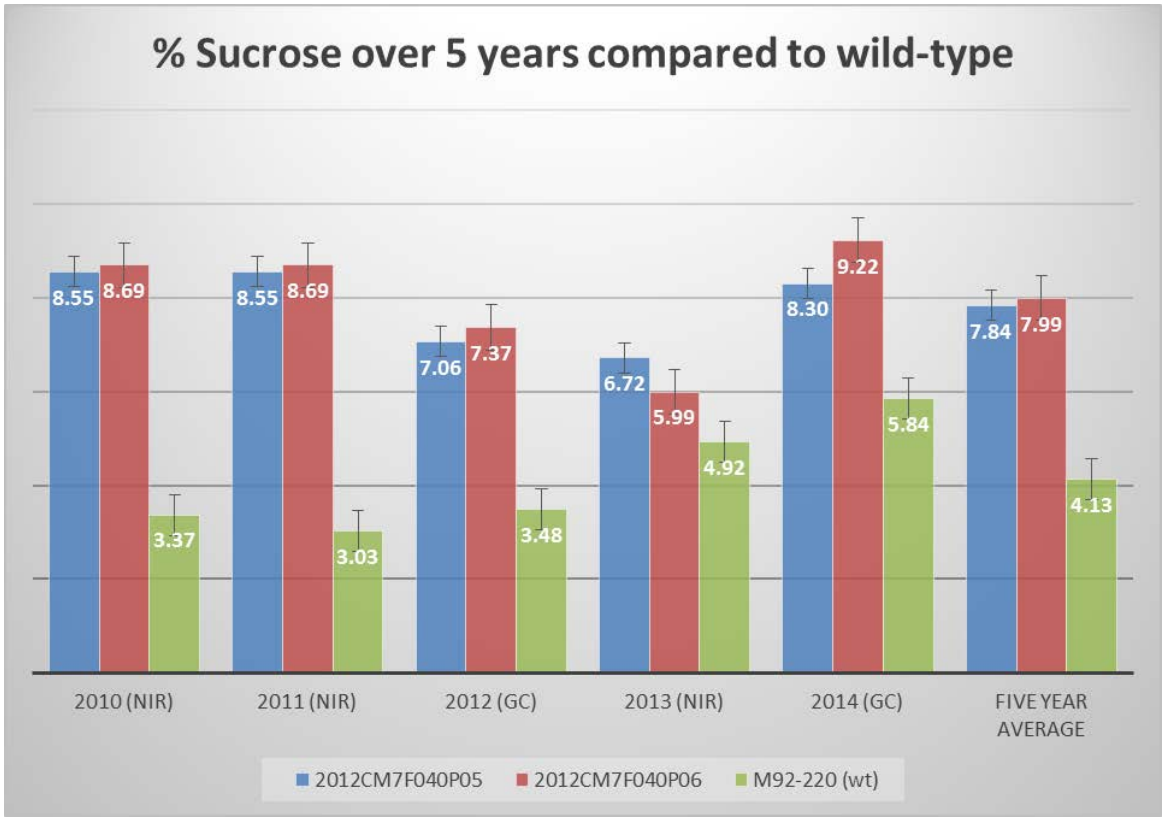

Figure 1 Sucrose concentration in high sucrose mutants compared to wild type (M92-220) over five years. The concentration of sucrose in these mutants (2012CM7F040P05 and 2012CM7F040P06) is significantly greater than the wild type and over a five year span averaged just under double the amount of sucrose compared to the wild type.

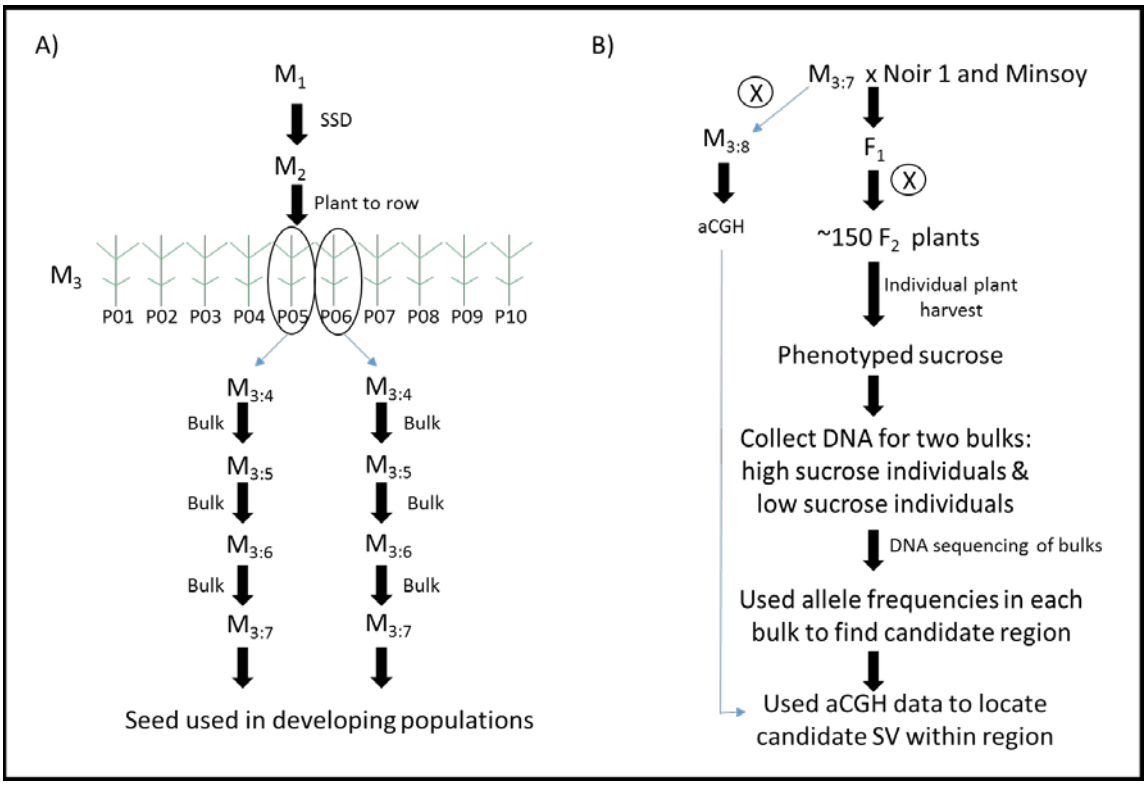

Figure 2: Population development diagram. Panel A on the left describes how the 2 FN mutants were advanced starting at the M1 generation. Panel B on the right describes the development of the F_2 populations used for mapping high sucrose.

Table 1. Outcross population designations for populations A13-004, A13-005, A13-007, and A13-008. This table shows which parents were used to create each of the four populations.

	Noir	Minsoy
2012CM7F040P05	A13-004	A13-007
2012CM7F040P06	A13-005	A13-008

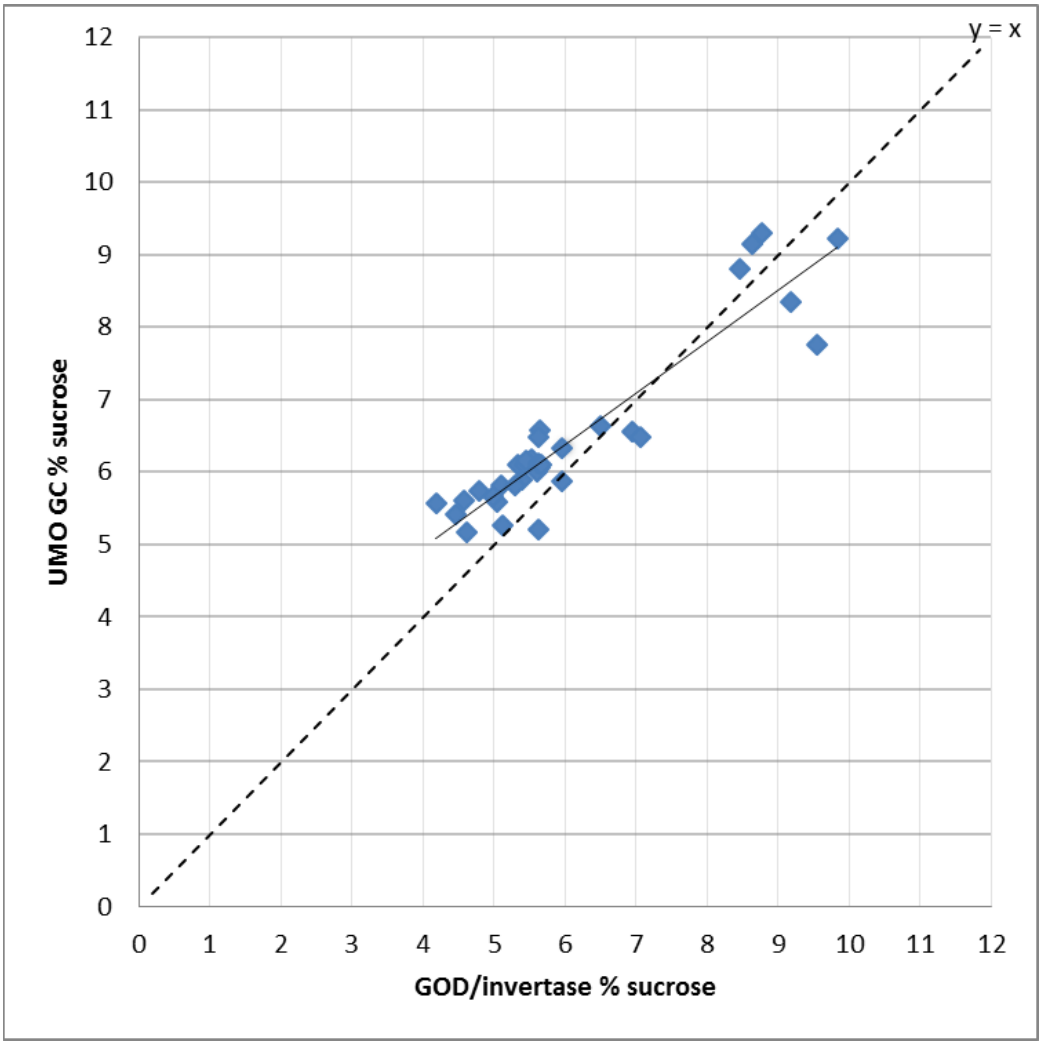

Figure 3. Comparison between University of Missouri GC sucrose concentration and developed GOD/invertase colorimetric assay. The dotted line is the ideal correlation of r =1 and visually shows the deviation of the points from the line. The correlation between UMO and GOD/invertase is r = 0.93. This graph shows the usefulness of the assay in being able to differentiate high sucrose values from low sucrose values in the bulked segregant analysis experiment.

Table 2. Summary of four mapping populations phenotyped for sucrose. Total phenotyped refers to the total number of individuals in the F2 population phenotyped for sucrose. The number of individuals pooled for each bulk are also included.

Population	Female Parent	Male Parent	Total phenotyped	# ind. in low bulk	# ind. in high bulk
A13-004	Noir	2012CM7F040P05	119	16	13
A13-005	Noir	2012CM7F040P06	113	19	15
A13-007	2012CM7F040P05	Minsoy	66	10	12
A13-008	2012CM7F040P06	Minsoy	116	11	17

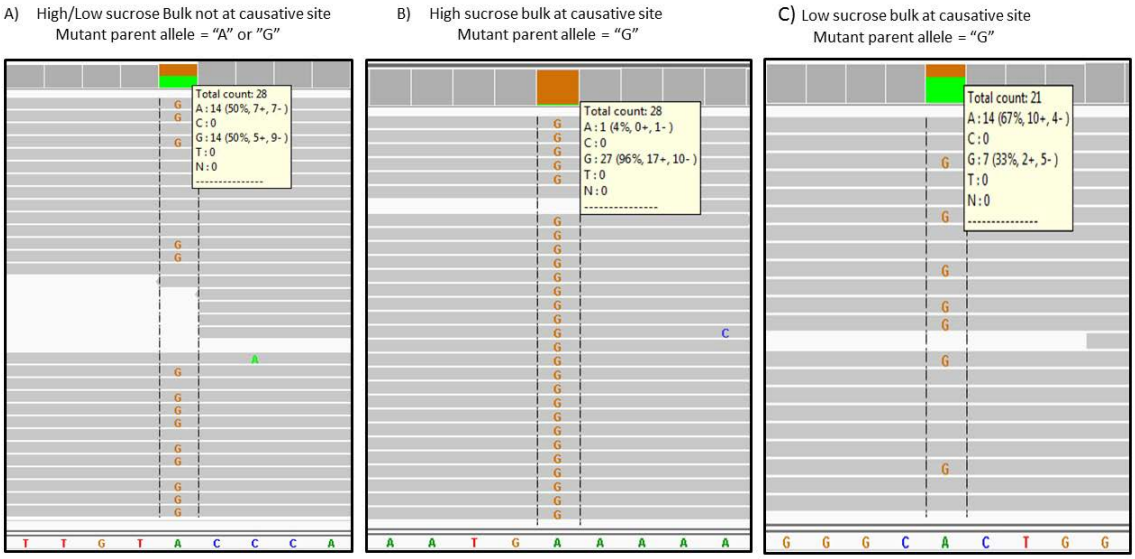

Figure 4. Integrative Genomics Viewer (IGV) of 9 base pair positions in one random bulk (section A), in a high sucrose bulk (section B), and a low sucrose bulk (section C). Each horizontal grey bar is a sequencing read, the letters along the bottom are the Williams 82 version 2 reference alleles, the letters on the grey bars are the alternate allele, and grey bars with no letters are reference allele calls. To calculate alternate allele frequency, the number of reads of the alternate allele is divided by the total number reads at that position. In section A, there are 14/28 alternate allele reads resulting in a 50% allele frequency which is expected to arise throughout the genome at random. In section B, there are 27/28 alternate allele reads resulting in a 96% allele frequency which might be expected at the causative location for sucrose content if the mutant parent allele is the same as the alternate allele (percent of mutant allele in high sucrose bulk). In section C, there are 7/21 alternate allele reads resulting in a 33% allele frequency which might be expected at the causative location for sucrose content where the mutant allele is the same as the alternate allele (percent of mutant allele in low sucrose bulk). Although these allele frequencies were determined through a python script, checking IGV is a great quality control to verify the bioinformatics tools are working properly.

Table 3. Summary of Mean and range sucrose concentrations for four F2 populations and parental checks. Total phenotyped refers to the total number of individuals in the F2 population phenotyped for sucrose. Values reported are the average sucrose percentage from samples run in duplicate.

Population	Female Parent	Male Parent	Popln. size	Mean % sucrose	Range % sucrose
A13-004	Noir	2012CM7F040P05	119	4.71	3.38 - 7.52
A13-005	Noir	2012CM7F040P06	113	5.34	3.49 - 8.25
A13-007	2012CM7F040P05	Minsoy	66	6.34	4.44 - 8.24
A13-008	2012CM7F040P06	Minsoy	116	6.21	4.35 - 8.18
2012CM7F040P05	Parental Check	Parental Check	N/A	9.06	N/A
2012CM7F040P06	Parental Check	Parental Check	N/A	9.08	N/A
Minsoy	Parental Check	Parental Check	N/A	6.84	N/A
Noir	Parental Check	Parental Check	N/A	4.59	N/A
M92-220	Wild-Type	Wild Type	N/A	5.74	N/A

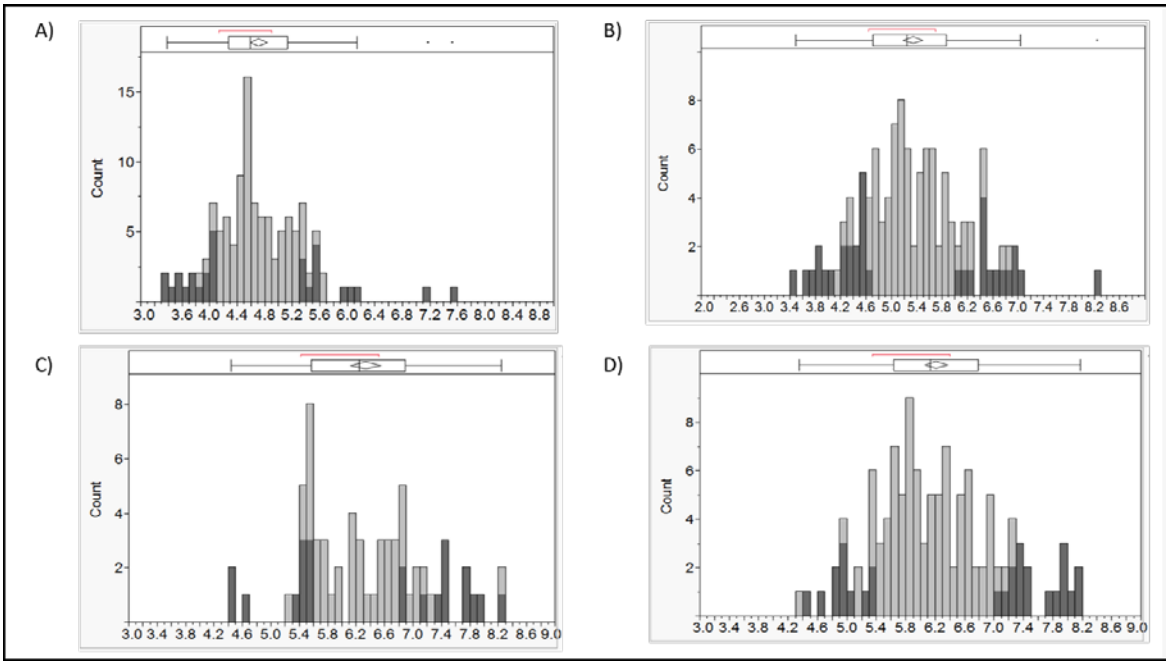

Figure 5. F_2 distribution of population A13-004 (A), A13-005 (B), A13-007 (C), and A13-008 (D). Average %sucrose between duplicate runs. Shaded bars indicate the individuals that were bulked for sequencing

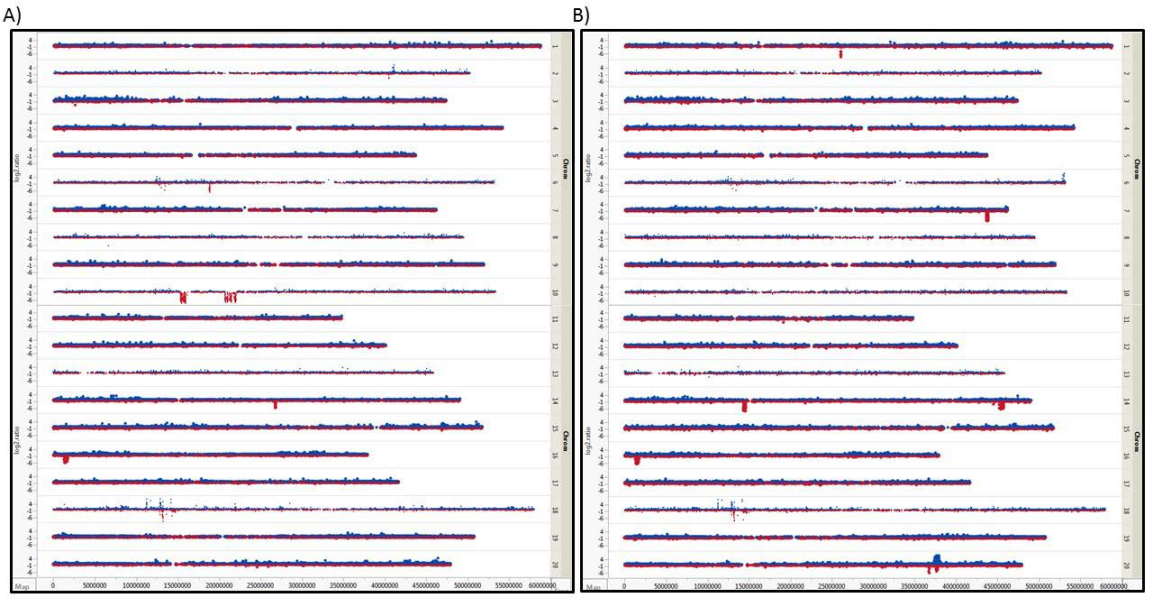

Figure 6: Array Comparative Genomic Hybridization (aCGH) graphs of 2 FN plants derived from 2012CM7F040P05. Both individuals (Section A and Section B) were plants in the M3:8 generation (See population development figure). The Y-axis is the log2 ratio of each genotype vs. the M92-220-Long reference probe throughout the genome including all 20 soybean chromosomes. Down peaks (colored red) indicate deletions while the up peaks (colored blue) indicate duplications. The x-axis indicates the position (bp) on the chromosome.

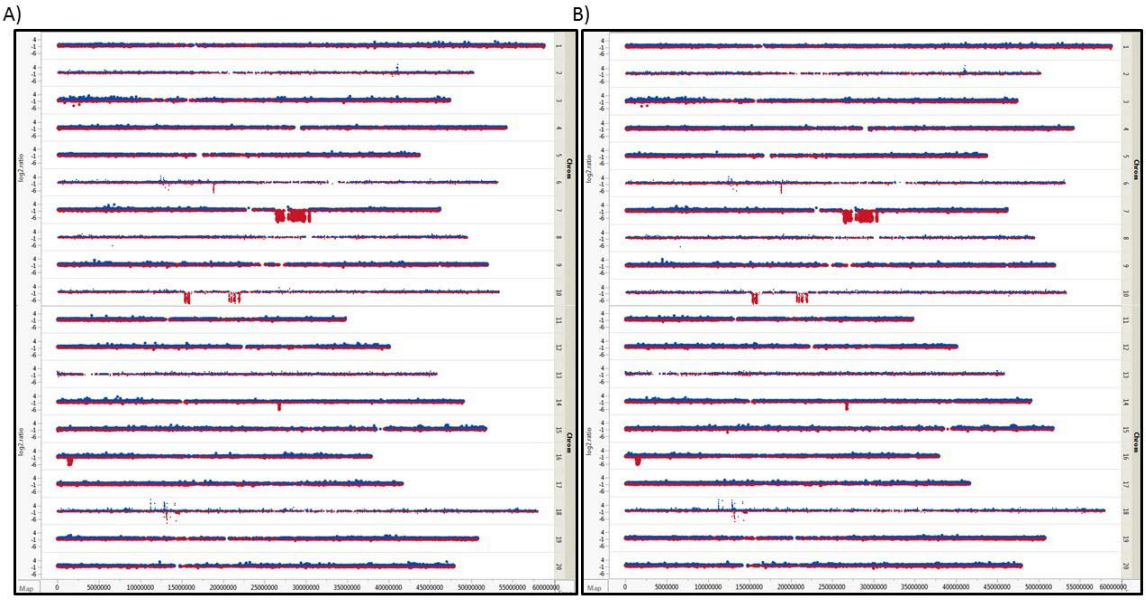

Figure 7: Array Comparative Genomic Hybridization (aCGH) graphs of 2 FN plants derived from 2012CM7F040P06. Both individuals (Section A and Section B) were plants in the M3:8 generation (See population development figure). The Y-axis is the log2 ratio of each genotype vs. the M92-220-Long reference probe throughout the genome including all 20 soybean chromosomes. Down peaks (colored red) indicate deletions while up peaks (colored blue) indicate duplications. The x-axis indicates the position (bp) along the chromosome.

Table 4. aCGH detected deletions and duplications for two high sucrose mutants (2012CM7F040P05 and 2012CM7F040P06). CGH deletions were observed by looking for positions in the genome where the log2ratio of the probes was less than -1. The approximate start and end position of each deletion are listed as well as the approximate size of the deletions. The type of structural variant (SV) as well as which individuals the SV occurs in as indicated by an "x."

Chrom.	Approximate Start Position (bp)	Approximate End Position (bp)	Approximate Size (bp)	Type of SV	CM7F040 P05.pl15	CM7F040 P05.pl39	CM7F040 P06.pl35	CM7F040 P06.pl45
1	25,151,880	25,190,670	38,790	Homozygous Deletion		x		
5	14,952,853	15,036,732	83,879	Het deletion	x	x	x	x
5	22,169,791	22,297,618	127,827	Het deletion	x	x	x	x
6	18,076,434	18,206,053	129,619	Homozygous Deletion	x		x	x
7	25,224,527	29,402,646	4,178,119	Homozygous Deletion			x	x
7	42,187,465	42,362,063	174,598	Homozygous Deletion		x		
9	33,181,392	33,327,145	145,753	Het deletion	x	x	x	x
10	14,760,703	15,615,485	854,782	Homozygous Deletion	x		x	x
10	19,973,294	21,359,576	1,386,282	Homozygous Deletion	x		x	x
12	24,210,936	24,295,905	84,969	Het deletion	x	x	x	x
12	27,844,374	28,031,087	186,713	Het deletion	x	x	x	x
14	14,328,413	14,589,498	261,085	Homozygous Deletion		x		
14	26,691,150	26,812,355	121,205	Homozygous Deletion	x		x	x
14	45,180,382	45,754,451	574,069	Homozygous Deletion		x		
16	1,253,986	1,755,504	501,518	Homozygous Deletion	x	x	x	x
17	23,565,963	23,629,265	63,302	Het deletion	x	x	x	x
17	23,566,374	23,623,208	56,834	Het deletion	x	x	x	x
18	13,147,075	13,147,075	0	Homozygous Deletion	x	x	x	x
20	37,684,250	37,684,250	0	duplication		x		

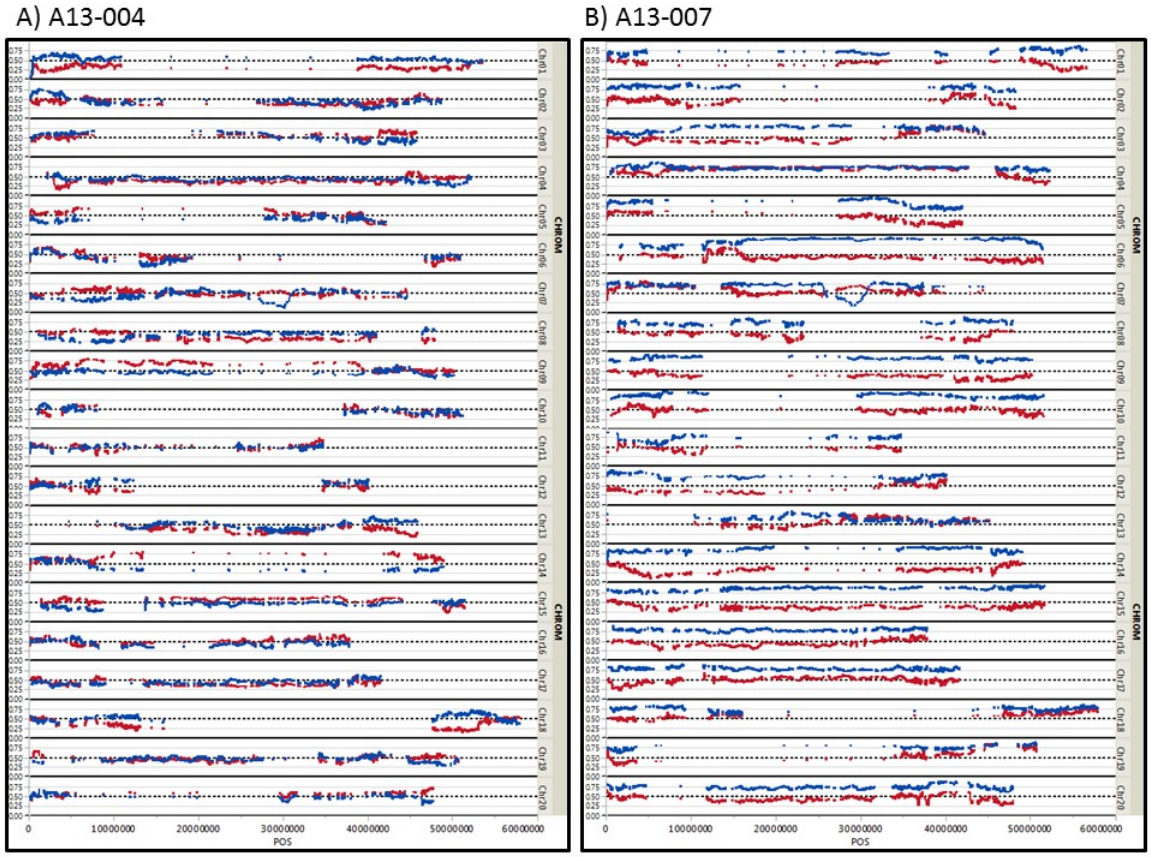

Figure 8. BSA allele frequencies of population A13-004 (Noir 1 x 2012CM7F040P05) (Section A) and A13-007 (2012CM7F040P05 x Minsoy) (Section B). Red lines indicate the percent of mutant allele in the mutant bulk and blue lines indicate the percent of mutant allele in the wild-type bulk (both lines are smoothed to an 11 SNP sliding window). All 20 chromosomes are listed on the right side along with the base pair position on the X-axis.

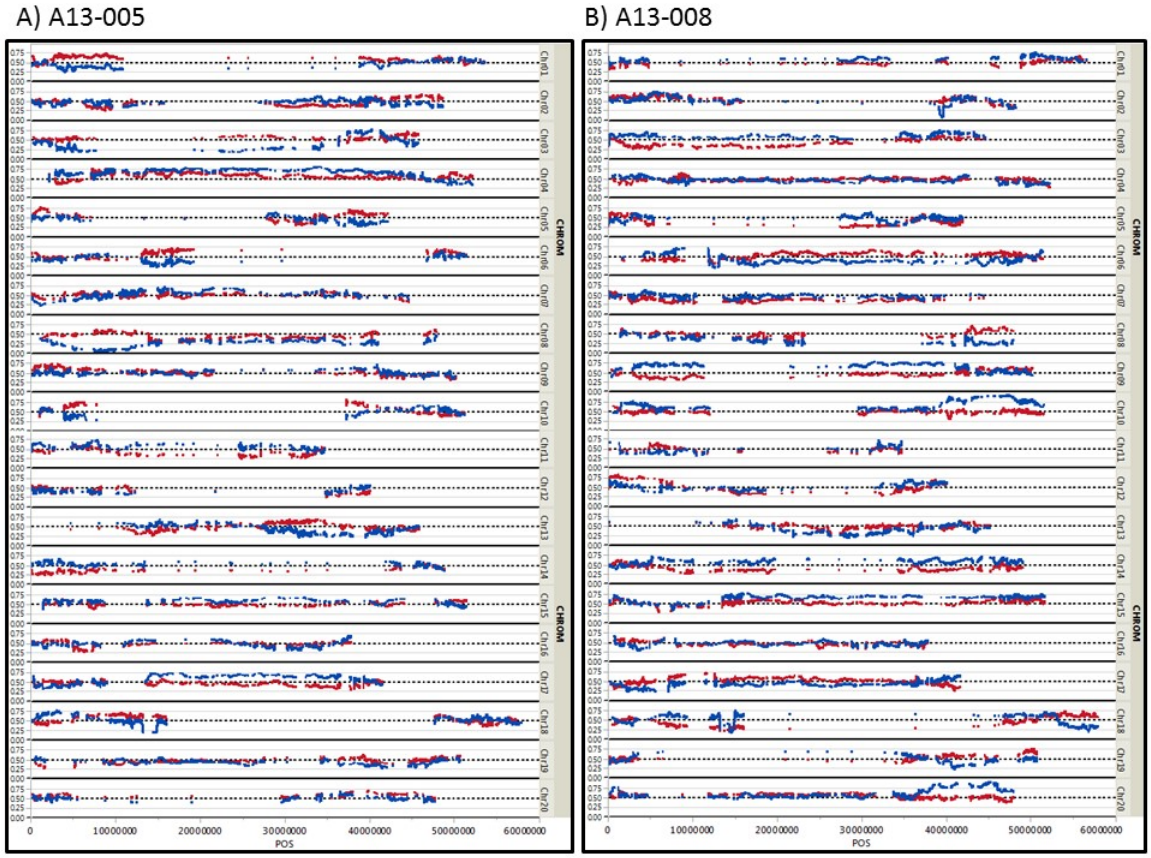

Figure 9. BSA allele frequencies of population A13-005 (Noir 1 x 2012CM7F040P06) (Section A) and A13-008 (2012CM7F040P06 x Minsoy) (Section B). Red lines indicate the percent of mutant allele in the mutant bulk and blue lines indicate the percent of mutant allele in the wild-type bulk (both lines are smoothed to an 11 SNP sliding window). All 20 chromosomes are listed on the right side along with the base pair position on the X-axis.

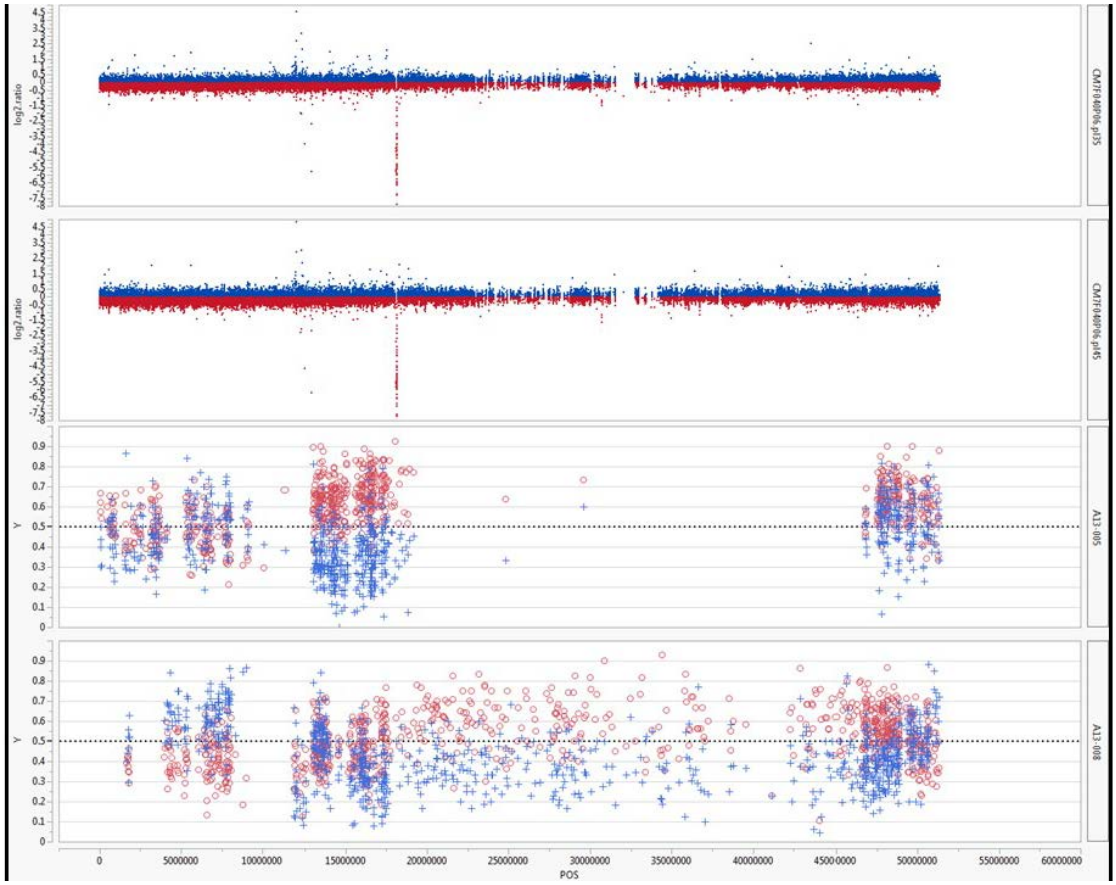

Figure 10. BSA and aCGH of chromosome 6. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

Figure 11. BSA and aCGH of chromosome 8. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

Table 5. Visually identified regions of interest based on BSA spread in allele frequency. If a location appeared to separate in allele frequency between the high sucrose bulk and the low sucrose bulk for either population A13-005 or population A13-008, the region was noted. In addition, aCGH detected SVs within the region were identified.

Chromosome	BSA spread in allele frequency A13-005 location	BSA spread in allele frequency A13-008 location	aCGH detected SV in region
1	0-10 MB	--	NO
2	--	--	--
3	0-32 MB	--	NO
4	--	--	--
5	--	--	
6	12-45 MB	12-45 MB	129 KB at 18 MB position
7	--	--	
8	2-12 MB	--	Single probe deletion at 6 MB
9	--	--	
10	2-40MB		854 KB at 15 MB position and 1 MB at 19 MB position
11	--	--	--
12		3-5 MB	NO
13	Slight spread from 27-40MB	--	NO
14	--	--	--
15	--	--	--
16	--	--	--
17	--	--	--
18	--	52 -55 MB	NO
19	--	--	--
20	--	--	--

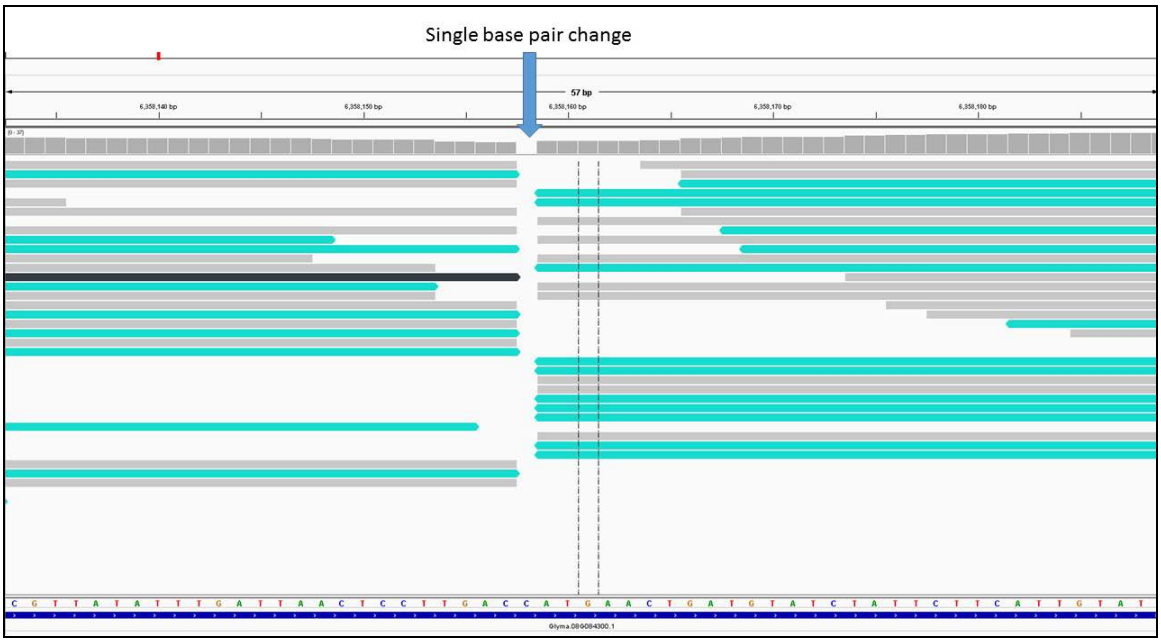

Figure 12. IGV image of CGH probe on chromosome 8. Horizontal lines indicate paired end reads. Grey horizontal lines indicate reads pairing close to one another on the same chromosome while sky blue colored reads are pairing with reads found on chromosome 13.

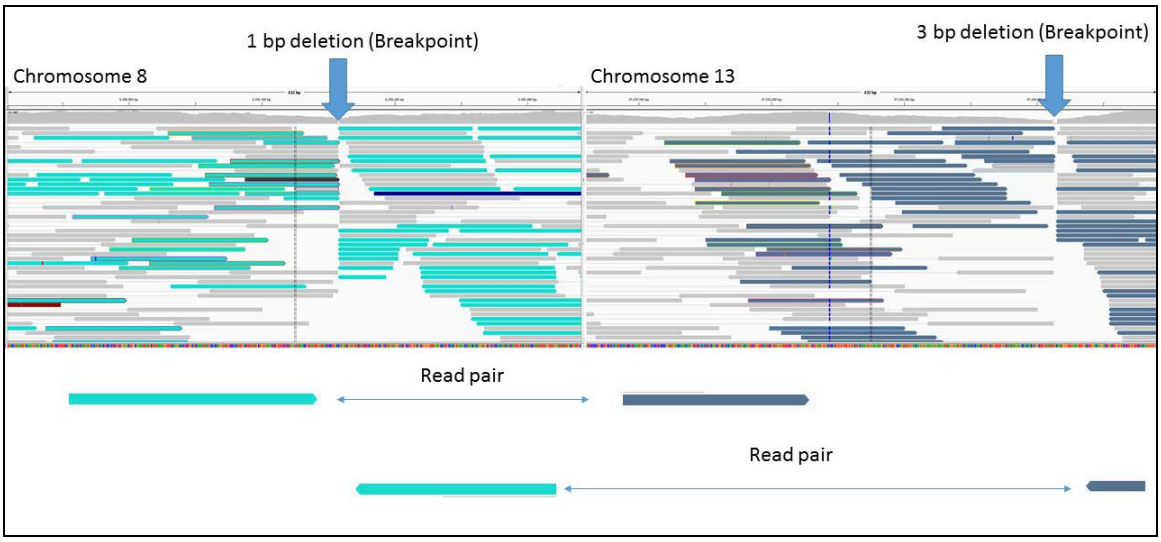

Figure 13. IGV image of chromosome 8 and chromosome 13 tentative reciprocal translocation. Chromosome 8 (left) shows a 1 bp breakpoint and chromosome 13 (right) shows a 3 bp breakpoint. The read pairs of chromosome 8 matching with read pairs on chromosome 13 in this mutant DNA indicates a tentative reciprocal translocation.

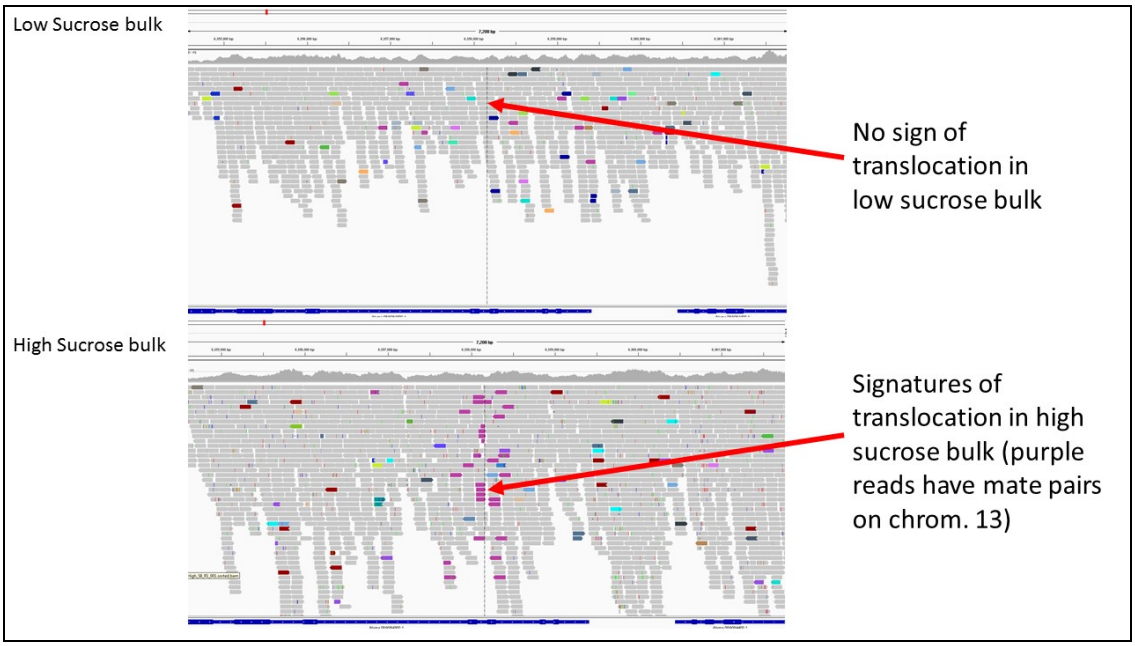

Figure 14. IGV image of low (top) and high (bottom) sucrose bulks of DNA. Signatures of the translocation where paired reads are mapping between chromosome 8 and 13 are appearing in the high sucrose bulk, but are lacking in the low sucrose bulk.

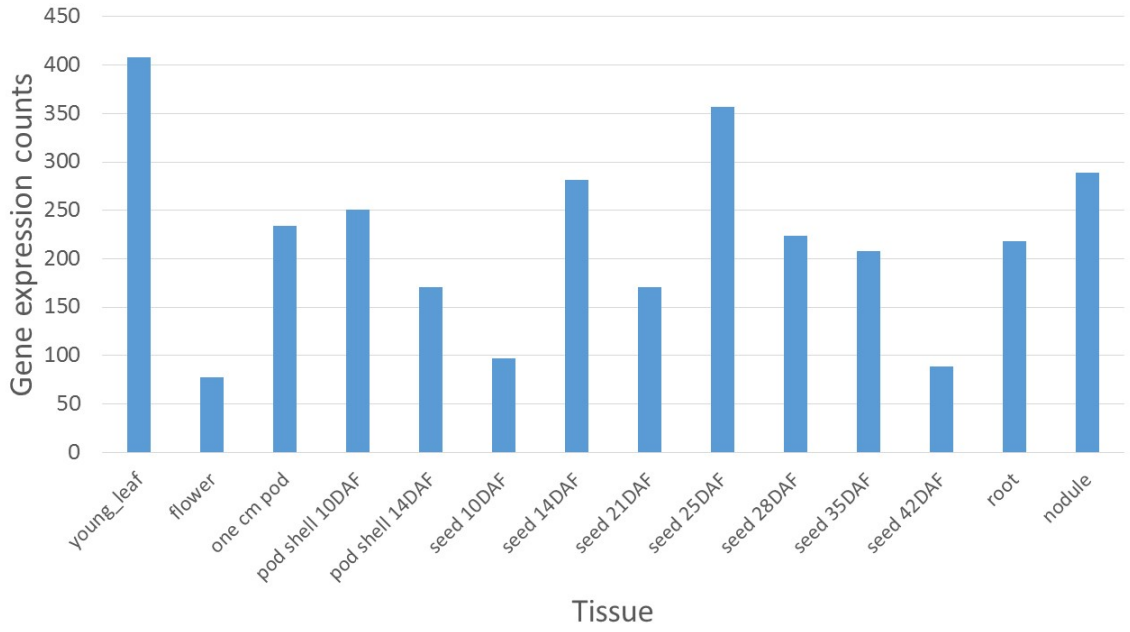

Figure 15. Digital gene expression counts of the uniquely mapped reads. Data taken from soybase and provided by (Severin et al., 2010). Raw data of uniquely mapped reads by tissue type is graphed for 14 different tissues.

Literature cited

Akond, M., S. Liu, S.K. Kantartzi, K. Meksem, and N. Bellaloui. 2015. Quantitative Trait Loci Underlying Seed Sugars Content in " MD96-5722 " by " Spencer " Recombinant Inbred Line Population of Soybean. Food Nutr. Sci. (August): 964–973.

Bailly, C., C. Audigier, F. Ladonne, M.H. Wagner, F. Coste, F. Corbineau, and D. Côme. 2001. Changes in oligosaccharide content and antioxidant enzyme activities in developing bean seeds as related to acquisition of drying tolerance and seed quality. J. Exp. Bot. 52(357): 701–708.

Bhatti, T., R. Chambers, and J. Clamp. 1970. The Gas Chromatographic Properties of Biologically Important N-Acetylglucosamine Derivatives Monosaccharides, Disaccharides, Trisaccharides, Tetrasaccharides and Pentasaccharides. Biochim. Biophys. Acta 222: 339–347.

Bilyeu, K.D., M. Ratnaparkhe, and K. Chittaranjan. 2010. Genetics, Genomics, and Breeding of Soybean. CRC Press.

Bolon, Y.-T., W.J. Haun, W.W. Xu, D. Grant, M.G. Stacey, R.T. Nelson, D.J. Gerhardt, J.A. Jeddeloh, G. Stacey, G.J. Muehlbauer, J.H. Orf, S.L. Naeve, R.M. Stupar, and C.P. Vance. 2011. Phenotypic and Genomic Analyses of a Fast Neutron Mutant Population Resource in Soybean. Plant Physiol. 156(1): 240–253.

Bolon, Y.-T., A.O. Stec, J.-M. Michno, J. Roessler, P.B. Bhaskar, L. Ries, A.A. Dobbels, B.W. Campbell, N.P. Young, J.E. Anderson, D.M. Grant, J.H. Orf, S.L. Naeve, G.J. Muehlbauer, C.P. Vance, and R.M. Stupar. 2014. Genome Resilience and Prevalence of Segmental Duplications Following Fast Neutron Irradiation of Soybean. Genetics 198(3): 967–981.

Campbell, J.A., R.W. Hansen, and J.R. Wilson. 1999. Cost-effective colorimetric microtitre plate enzymatic assays for sucrose, glucose and fructose in sugarcane tissue extracts. J. Sci. Food Agric. 79(2): 232–236.

Cicek, M. 2001. Genetic Marker Analysis Of Three Major Carbohydrates In Soybean Seeds.

Diers, B.W., W.R. Fehr, and R.C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil content. Theor. Appl. Genet. 83: 608 – 612.

Ehrenreich, I.M., N. Torabi, Y. Jia, J. Kent, S. Martis, J.A. Shapiro, D. Gresham, A.A. Caudy, and L. Kruglyak. 2010. Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature 464(7291): 1039–1042.

Fehr, W. 1987. Principles of Cultivar Development.

Giannoccaro, E., Y.-J. Wang, and P. Chen. 2006. Effects of solvent , temperature , time , solvent- to-sample ratio , sample size , and defatting on the extraction of soluble sugars in soybean. Food Chem. Toxicol. 71(1): 59–64.

Gizlice, Z., T.E. Carter, and J.W. Burton. 1994. Genetic Base for North American Public Soybean Cultivars Released between 1947 and 1988. Crop Sci. 34(5): 1143–1151.

Guo, J., Y. Wang, C. Song, J. Zhou, L. Qiu, H. Huang, and Y. Wang. 2010. A single origin and moderate bottleneck during domestication of soybean (Glycine max): implications from microsatellites and nucleotide sequences. Ann. Bot. 106(3): 505–

514.
Haase, N.J., T. Beissinger, C.N. Hirsch, B. Vaillancourt, S. Deshpande, K. Barry, C.R. Buell, S.M. Kaeppler, and N. De Leon. 2015. Shared Genomic Regions Between Derivatives of a Large Segregating Population of Maize Identi fi ed Using Bulked Segregant Analysis Sequencing and Traditional Linkage Analysis. Genes, Genomes, Genet. 5(August): 1593–1602.
Harten, A.M. van. 1998. Mutation Breeding: Theory and Practical Applications. Cambridge University Press.
Hoffmann, D., Q. Jiang, A. Men, M. Kinkema, and P.M. Gresshoff. 2007. Nodulation deficiency caused by fast neutron mutagenesis of the model legume Lotus japonicus. J. Plant Physiol. 164(4): 460–469.
Hymowitz, T., F.I. Collins, J. Panczner, and W.M. Walker. 1972. Relationship between the content of oil, protein, and sugar in sobyean seed. Agron. J. 64: 613–616.
Hyten, D.L., Q. Song, Y. Zhu, I.-Y. Choi, R.L. Nelson, J.M. Costa, J.E. Specht, R.C. Shoemaker, and P.B. Cregan. 2006. Impacts of genetic bottlenecks on soybean genome diversity. Proc. Natl. Acad. Sci. U. S. A. 103(45): 16666–16671.
Janauer, G., and P. Englmaier. 1978. Multi-step time program for the rapid gas-liquid chromatography of carbohydrates. J. Chromatogr. 153: 539–542.
Kim, H.-K., S.-T. Kang, J.-H. Cho, M.-G. Choung, and D.-Y. Suh. 2005. Quantitative trait loci associated with oligosaccharide and sucrose contents in soybean (Glycine max L.). J. Plant Biol. 48(1): 106–112.
Kumar, V., A. Rani, L. Goyal, A.K. Dixit, J.G. Manjaya, J. Dev, and M. Swamy. 2010. Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. J. Agric. Food Chem. 58(8): 5081–5085.
Liu, K. 1997. Soybeans Chemistry, Technology, and Utilization.
Liu, S., C.-T. Yeh, H.M. Tang, D. Nettleton, and P.S. Schnable. 2012. Gene Mapping via Bulked Segregant RNA-Seq (BSR-Seq). PLoS One 7(5): e36406.
Maughan, P.J., M.A.S. Maroof, and G.R. Buss. 2000. Identification of quantitative trait loci controlling sucrose content in soybean (Glycine max). Mol. Breed. 6(1): 105–111.
Michelmore, R.W., I. Paran, and R. V. Kesseli. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Genetics 88: 9828–9832.
Montelongo, J.L., B.M. Chassy, and J.D. Mccord. 1993. Lactobacillus salivarius for Conversion into Lactic Acid of Soy Molasses. J. Food Sci. 58(4): 863–866.
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. Crop Improv. 27(5): 606–616.
Naeve, S.L., R.A. Proulx, B.S. Hulke, and T.A.O. Neill. 2008. Whole Soybean Seed Using Near Infrared Spectroscopy. Agron. J. 100(2): 231–234.
Parry, M.A.J., P.J. Madgwick, C. Bayon, K. Tearall, A. Hernandez-Lopez, M. Baudo, M. Rakszegi, W. Hamada, A. Al-Yassin, H. Ouabbou, M. Labhilili, and A.L. Phillips.

2009. Mutation discovery for crop improvement. J. Exp. Bot. 60(10): 2817–2825.
Schmutz, J., S.B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D.L. Hyten, Q. Song, J.J. Thelen, J. Cheng, D. Xu, U. Hellsten, G.D. May, Y. Yu, T. Sakurai, T. Umezawa, M.K. Bhattacharyya, D. Sandhu, B. Valliyodan, E. Lindquist, M. Peto, D. Grant, S. Shu, D. Goodstein, K. Barry, M. Futrell-Griggs, B. Abernathy, J. Du, Z. Tian, L. Zhu, N. Gill, T. Joshi, M. Libault, A. Sethuraman, X.-C. Zhang, K. Shinozaki, H.T. Nguyen, R.A. Wing, P. Cregan, J. Specht, J. Grimwood, D. Rokhsar, G. Stacey, R.C. Shoemaker, and S.A. Jackson. 2010. Genome sequence of the palaeopolyploid soybean. Nature 463(7278): 178–183.
Severin, A.J., J.L. Woody, Y.-T. Bolon, B. Joseph, B.W. Diers, A.D. Farmer, G.J. Muehlbauer, R.T. Nelson, D. Grant, J.E. Specht, M. a Graham, S.B. Cannon, G.D. May, C.P. Vance, and R.C. Shoemaker. 2010. RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. BMC Plant Biol. 10(2007): 160.
Song, Q., D.L. Hyten, G. Jia, C. V. Quigley, E.W. Fickus, R.L. Nelson, and P.B. Cregan. 2013. Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. PLoS One 8(1): 1–12.
Song, Q., D.L. Hyten, G. Jia, C. V. Quigley, E.W. Fickus, R.L. Nelson, and P.B. Cregan. 2015. Fingerprinting Soybean Germplasm and Its Utility in Genomic Research. G3 Genes|Genomes|Genetics 5: 1999–2006.
Song, Q., J. Jenkins, G. Jia, D.L. Hyten, V. Pantalone, S.A. Jackson, J. Schmutz, and P.B. Cregan. 2016. Construction of high resolution genetic linkage maps to improve the soybean genome sequence assembly Glyma1 . 01. BMC Genomics 17(33).
Soybase.org. SoyBase and the Soybean Breeder's Toolbox.
Takagi, H., A. Abe, K. Yoshida, S. Kosugi, S. Natsume, C. Mitsuoka, A. Uemura, H. Utsushi, M. Tamiru, S. Takuno, H. Innan, L.M. Cano, S. Kamoun, and R. Terauchi. 2013. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J. 74(1): 174–183.
Teixeira, A.I., L.F. Ribeiro, S.T. Rezende, E.G. Barros, and M.A. Moreira. 2012. Development of a method to quantify sucrose in soybean grains. Food Chem. 130(4): 1134–1136.
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(february): 1–46.

Appendix A – Quantification of sucrose method

Quantification of sucrose using enzymatic GOD/invertase assay

The amount of sucrose in the soybean seed was determined using a colorimetric assay. This assay combines the action of invertase and glucose oxidase (GOD) and is adapted to a 96 well ELISA plate allowing for high throughput and cost effective analysis of sucrose content on a dry weight basis (Teixeira et al., 2012).

Several methods on the quantification of sugars have been published; however, each method has limitations based on the amount of seed required, time to analyze each sample, cost of analysis, and reproducibility of the results. High performance liquid chromatography (HPLC) and gas chromatography (GC) have been used to produce reliable results for many previous studies; however, cost and availability of HPLC or GC equipment was prohibitive for using this procedure. For this study, a high throughput cost effective analysis with reproducibility was needed, and thus samples were analyzed using an enzymatic method. In addition, this study was primarily concerned with sucrose concentration and excluded the analysis of both stachyose and raffinose.

The procedure for the quantification of sucrose in soybean grains using the GOD/invertase colorimetric method has been previously reported (Teixeira et al., 2012). Teixeira et al. validated the new GOD/invertase method to the reliable estimates reported by the HPLC and Stiff method and reported correlations of 0.9766 between GOD/invertase and HPLC and 0.9461 between GOD/invertase and the Stiff method (Teixeira et al., 2012). The experiment used in this study is based on Teixeira et al. with minor modifications.

Twenty seeds from each sample were ground for twenty seconds using a Hamilton Beach Chamber Coffee Grinder. Multiple coffee grinders were used in rotation to avoid overheating of the grinding chambers. After grinding, the ground seed was placed in 20mL low background glass Scintillation vials (Research Products International Cat. #121000). For drying, the samples were placed in a 105°C oven for five hours and subsequently stored in a desiccator. 100mg of the ground soybean seed was weighed and placed into a 13mL Round Base High Speed PP Centrifuge Tube (ThermoScientific Cat# 60.540.500) and 5mL of 80% ethanol solution was added to each tube. The samples were homogenized with a vortex for 1 minute and placed into a 70°C water bath for 90 min. To ensure samples remained homogenized, samples were briefly vortexed every 20 min during water bath incubation. The samples were then centrifuged at 16,100g for 10 minutes. 100µL of the supernatant was transferred to 1.5mL centrifuge tubes containing 900uL distilled water – thus producing a 1:10 dilution. This extract was then used to quantify sucrose using the GOD/invertase method.

In a 96 well clear flat bottom ELISA plate, 85uL reverse osmosis distilled water was added to each well. 5µL of the diluted alcohol extract from each sample above was then added to each corresponding well in the ELISA plate along with 10µL invertase (invertase from baker's yeast, Sigma-Aldrich) prepared at a concentration of 10mg/mL using distilled water. The plate was sealed and placed in a 55°C oven for 10 minutes. After this time, 50µL GOD (Glucose assay kit, Eton Bioscience) was added to every well in the ELISA plate. The plate was re-sealed and placed into a 37°C oven for 15 minutes. The plate was then removed from the oven, 50µL acetic acid was added to every well,

and the plate was left to incubate at room temperature for 5 minutes. The amount of sucrose was determined by reading the absorbance at 490nm on a multi-mode microplate reader (BioTek).

A calibration curve of standard sucrose solutions was also added to separate wells on every plate in order to determine the concentrations of the samples. These standard solutions were prepared using sucrose analytical standard (Sigma-Aldrich) and prepared at concentrations of 0, 0.05, 0.075, 0.10, 0.125, 0.15, 0.175, 0.2, 0.225, 0.25, 0.275, 0.30 g/100mL. This was done by preparing a stock solution of 1g/100mL sucrose solution and performing subsequent dilutions with distilled water to achieve the corresponding final concentrations. Every ELISA plate analyzed included this standard curve in duplicate.

The percent sucrose was calculated on a dry weight basis. First, the 0% sucrose blank was subtracted from all absorbance values on the plate. Next a calibration curve was created in Microsoft Excel - placing the known sucrose concentrations on the X-axis and the corresponding absorbance values on the Y-axis (see Figure 16). Since the samples were done in duplicate, the average absorbance value at each concentration was used. A trend line was used to fit the data and intercept the Y-axis at 0. The absorbance values of the samples were then fit to the trendline and percent sucrose concentration was calculated.

The results of this method were compared to wet chemistry lab results from the University of Missouri (UMO) (Agricultural Experiment Station Chemical Laboratories, University of Missouri-Columbia), who use GC to quantify sugar content. A set of 33 samples including parents in the mapping population (high sucrose mutants, Minsoy,

Noir 1, and M92-220) along with a P92Y22 reference check were analyzed both with the enzymatic method and at UMO with GC. A correlation of r = 0.927 between our results and UMO results indicate that the method used is reliable and can be used to rapidly screen soybean samples for sucrose concentration.

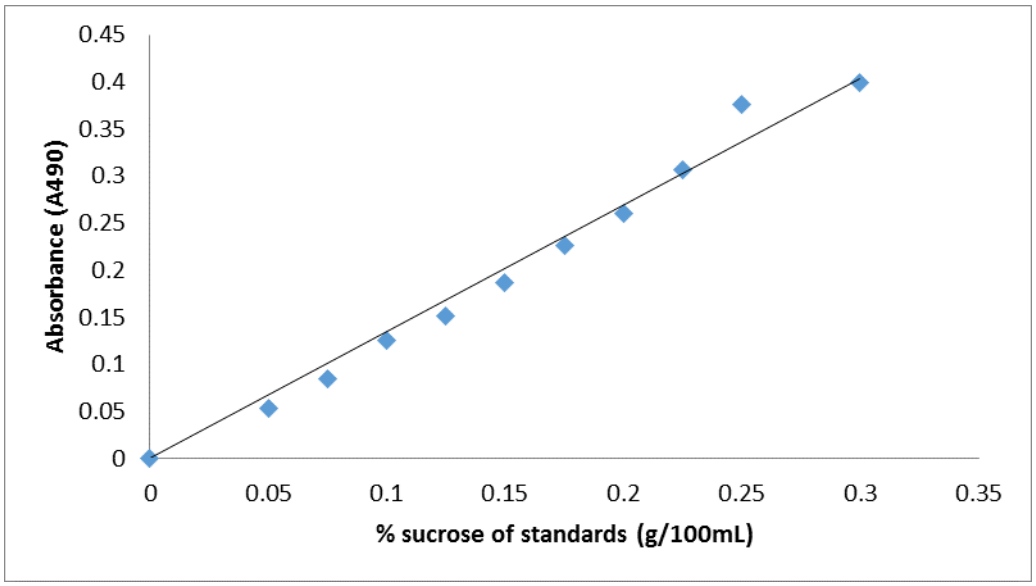

Figure 16. Sucrose calibration curve. Absorbance values at 490nm plotted against known sucrose concentrations in g/100mL. This sucrose calibration curve is used to calculate the % sucrose of the samples.

Appendix B – BSA and aCGH graphs by chromosome

.

Chromosome. 1

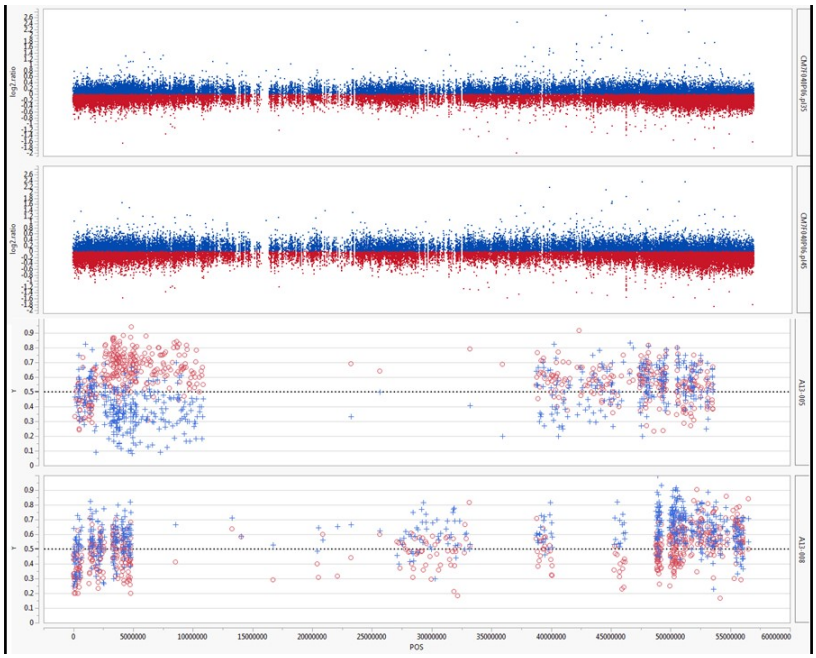

Figure 17. BSA and aCGH of chromosome 1. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

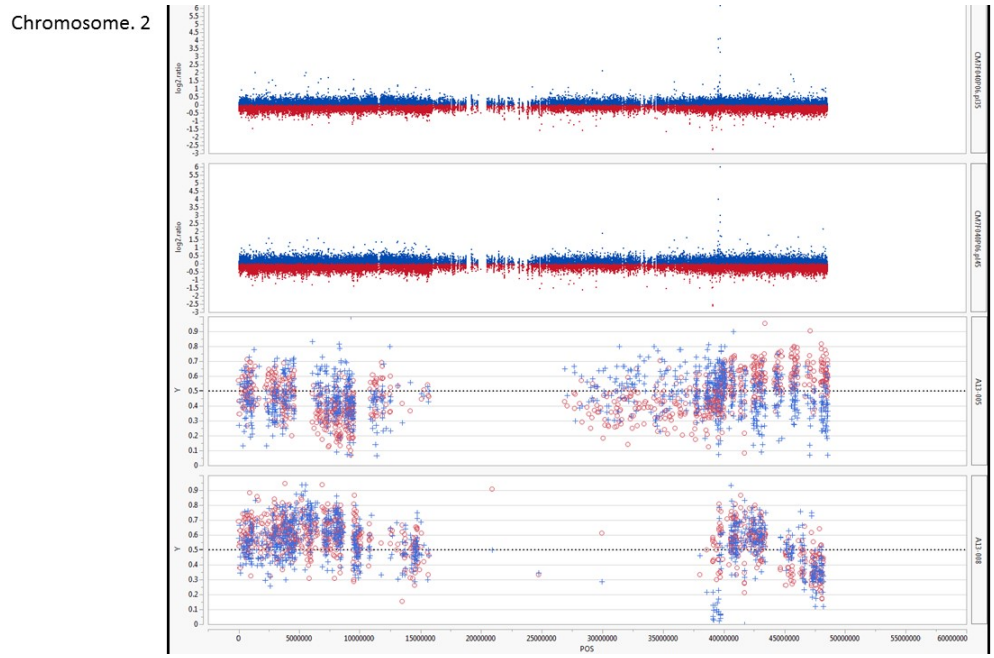

Figure 18. BSA and aCGH of chromosome 2. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

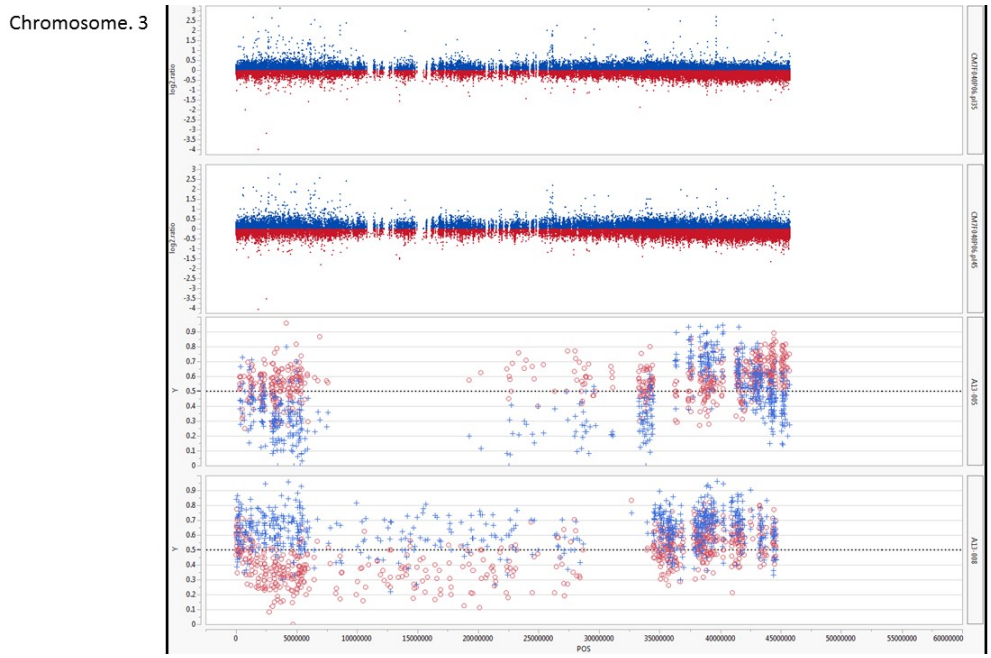

Figure 19. BSA and aCGH of chromosome 3. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

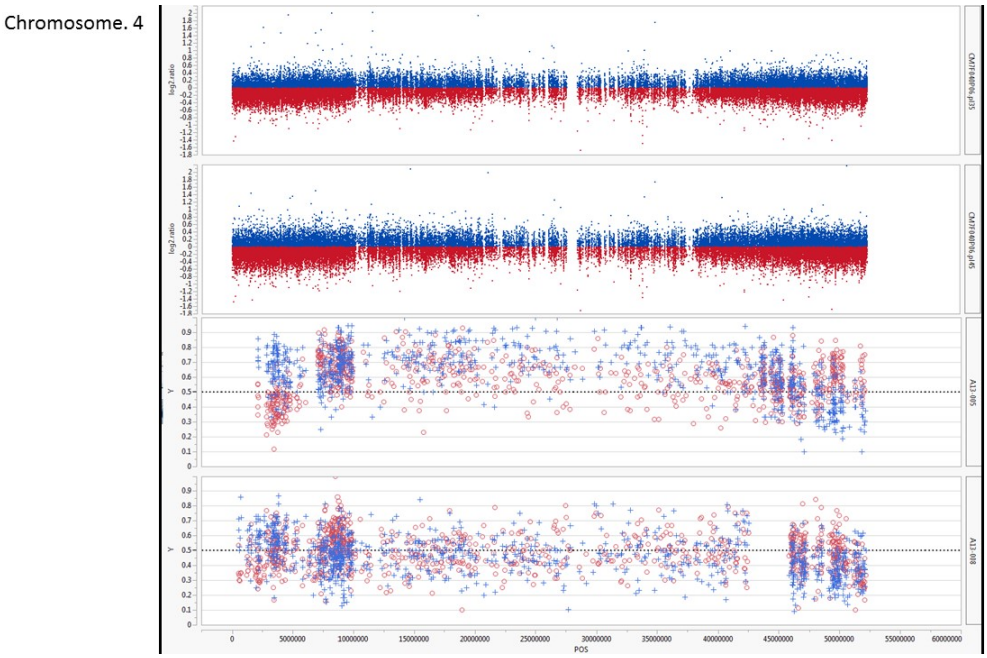

Figure 20. BSA and aCGH of chromosome 4. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

Chromosome. 5

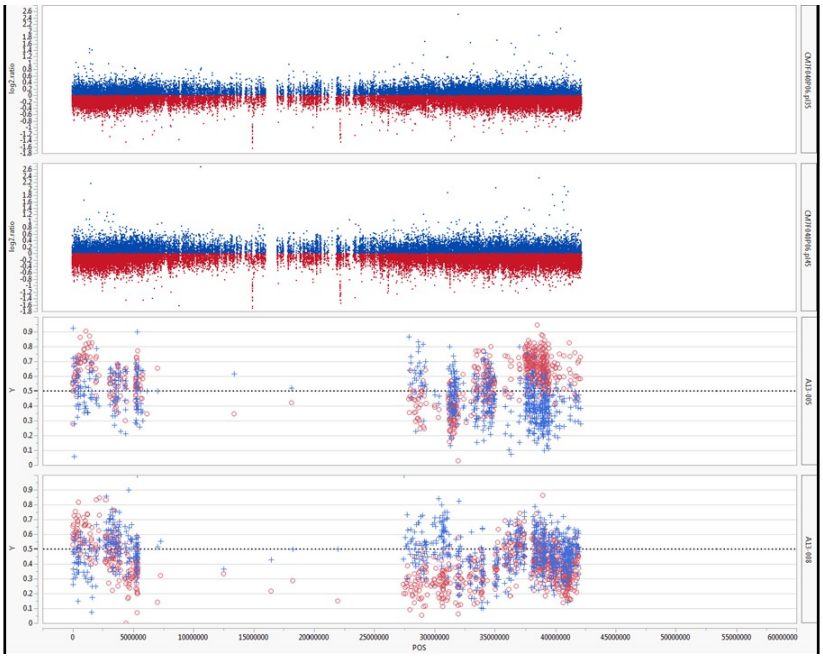

Figure 21. BSA and aCGH of chromosome 5. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

Chromosome. 6

Figure 22. BSA and aCGH of chromosome 6. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

Chromosome. 7

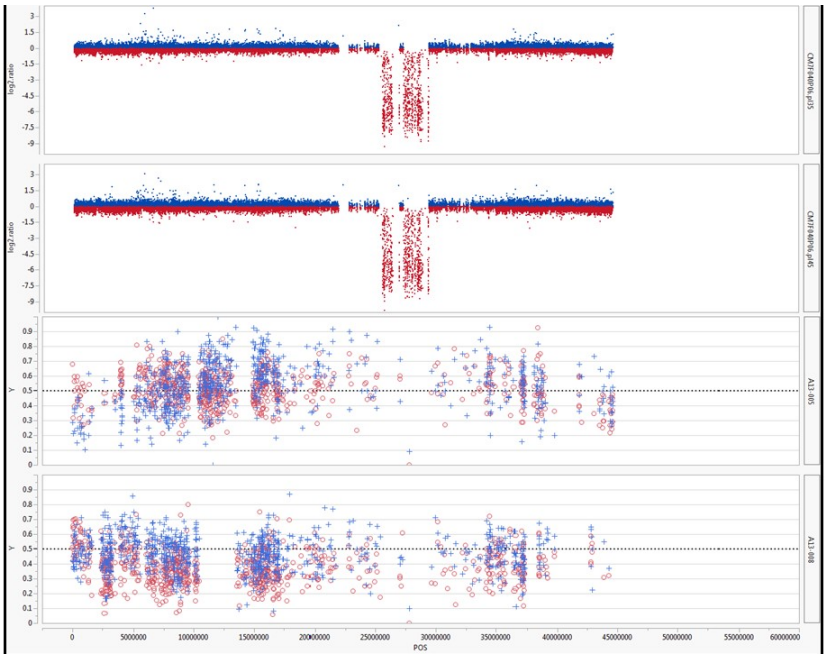

Figure 23. BSA and aCGH of chromosome 7. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

Chromosome. 8

Figure 24. BSA and aCGH of chromosome 8. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

Chromosome. 9

Figure 25. BSA and aCGH of chromosome 9. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

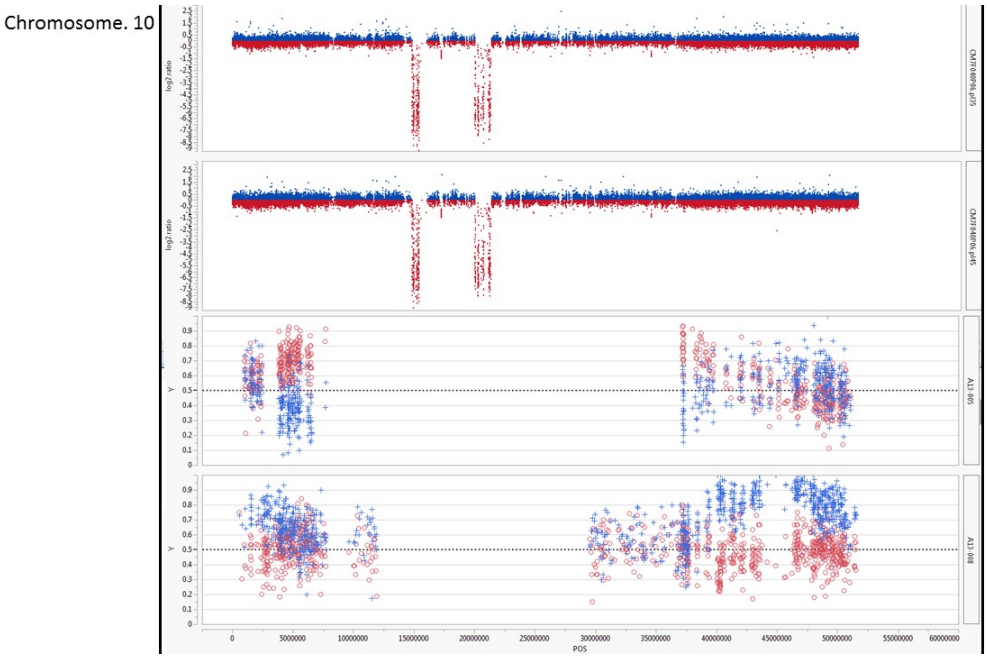

Figure 26. BSA and aCGH of chromosome 10. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

Chromosome. 11

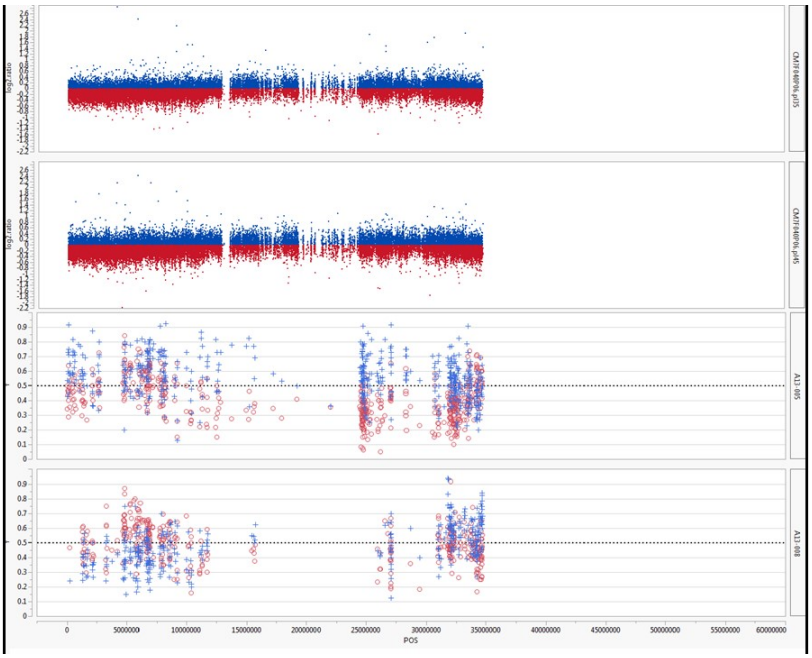

Figure 27. BSA and aCGH of chromosome 11. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

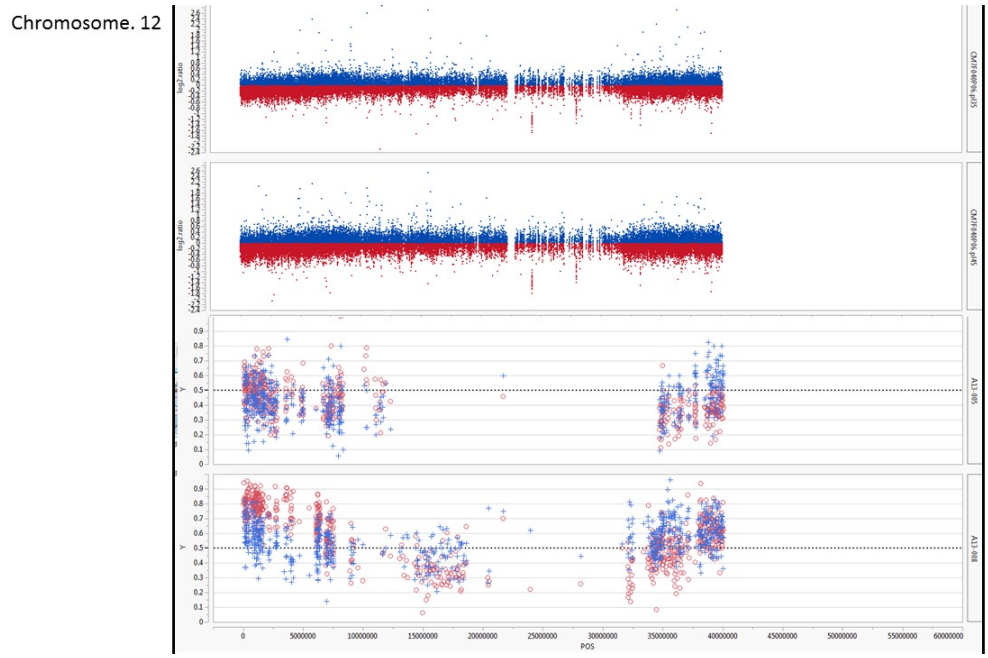

Figure 28. BSA and aCGH of chromosome 12. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

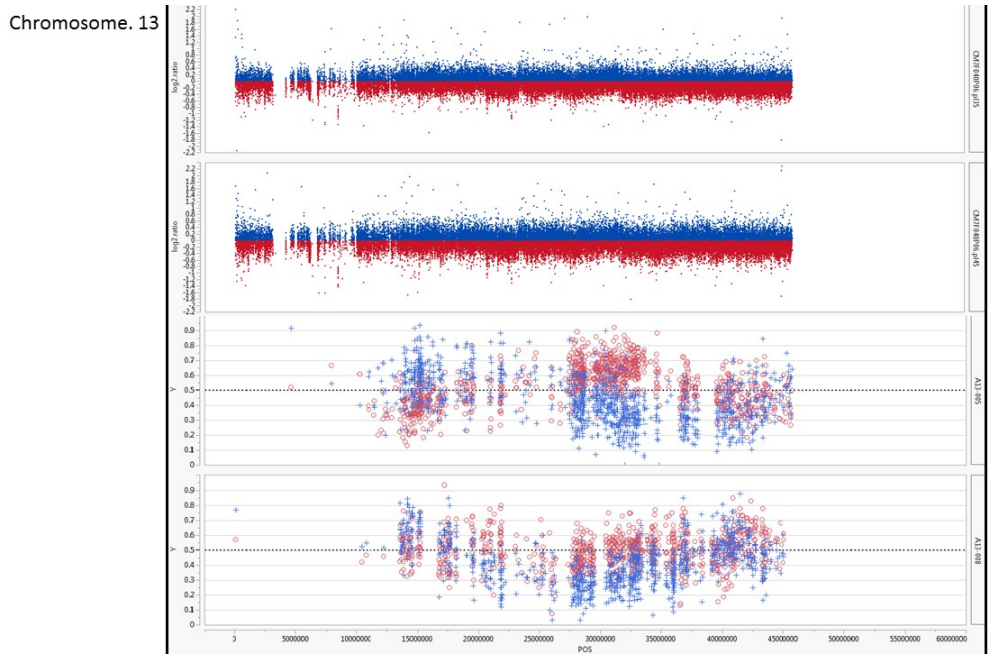

Figure 29. BSA and aCGH of chromosome 13. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

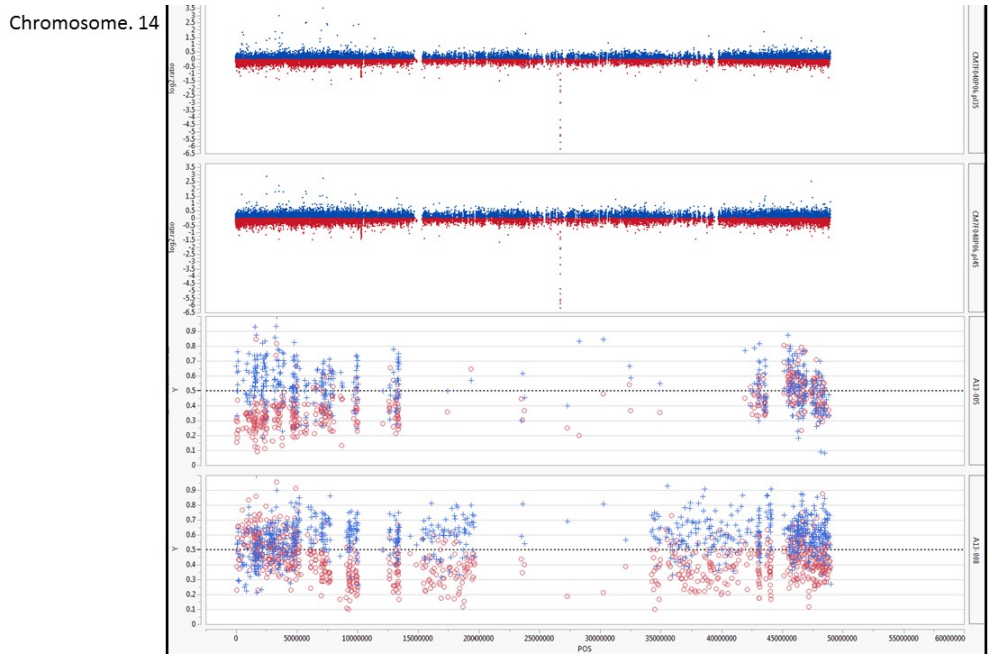

Figure 30. BSA and aCGH of chromosome 14. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

Chromosome. 15

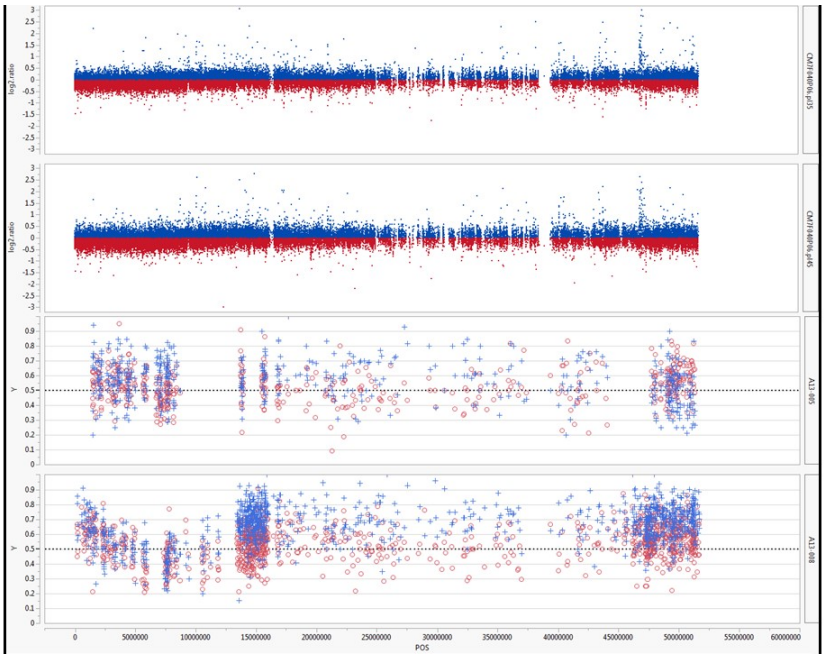

Figure 31. BSA and aCGH of chromosome 15. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

Chromosome. 16

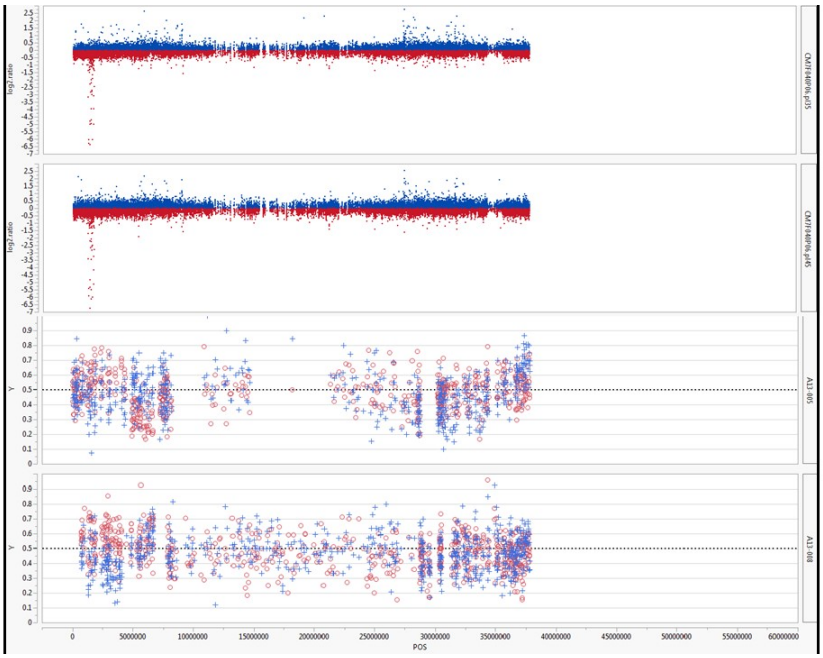

Figure 32. BSA and aCGH of chromosome 16. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

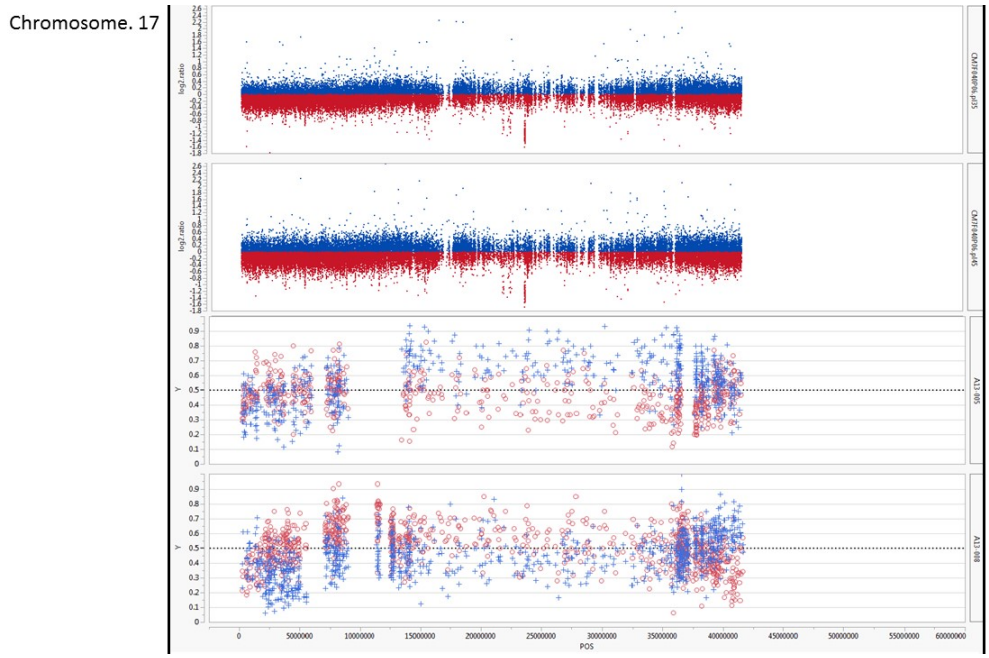

Figure 33. BSA and aCGH of chromosome 17. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

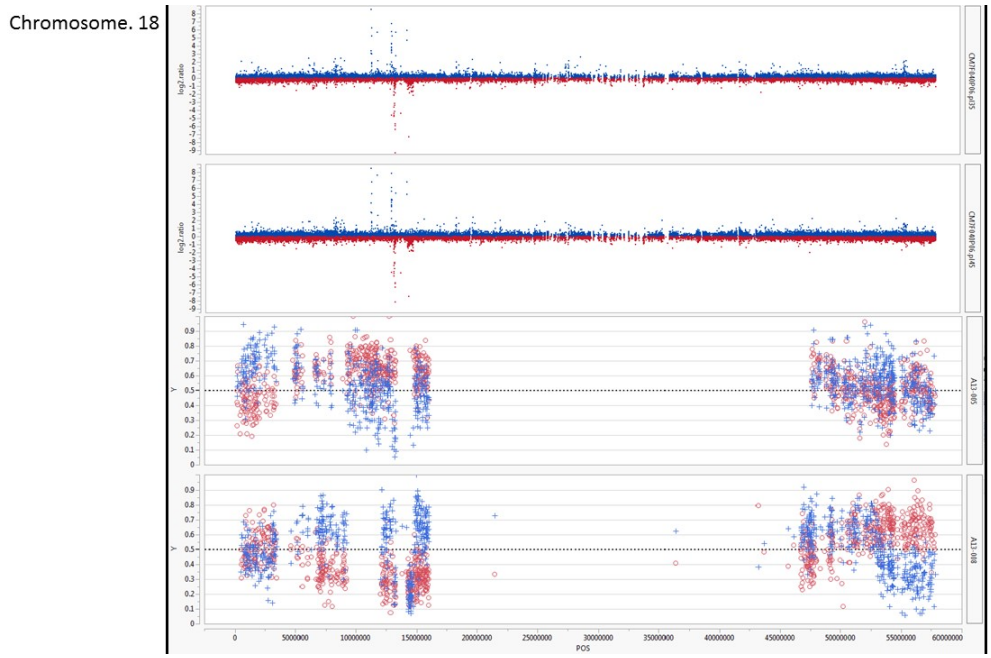

Figure 34. BSA and aCGH of chromosome 18. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

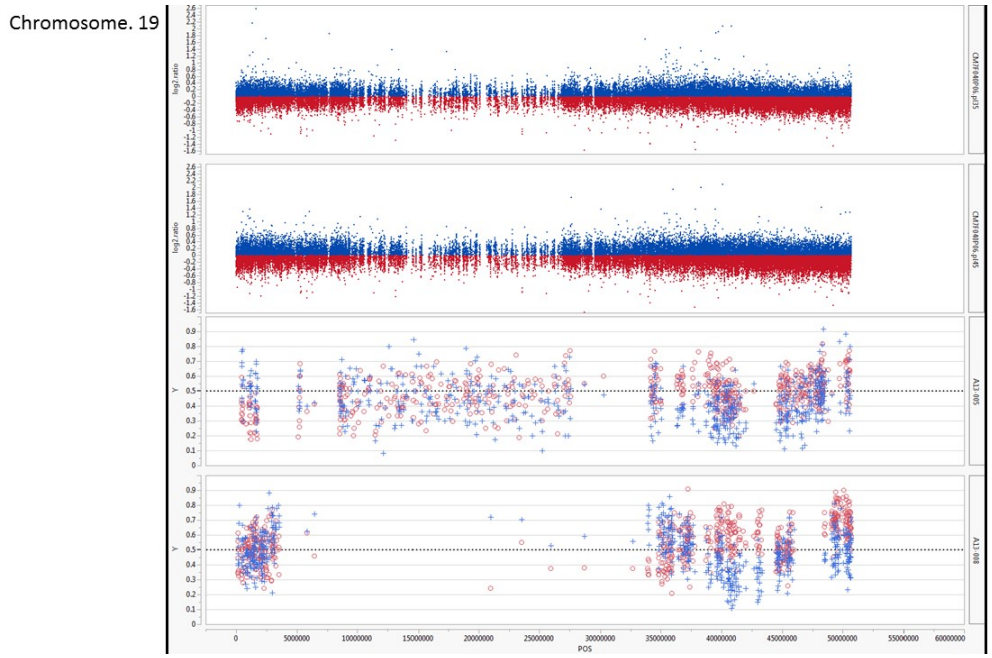

Figure 35. BSA and aCGH of chromosome 19. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

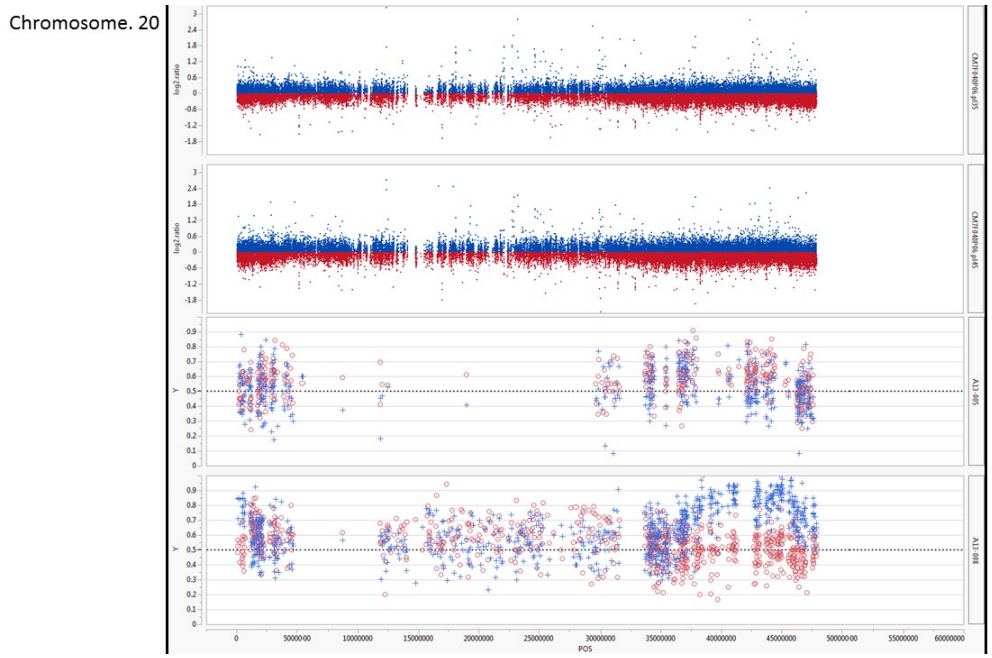

Figure 36. BSA and aCGH of chromosome 20. The aCGH plots (top 2 graphs) display the log2 ratio of the mutant genotype vs. the M92-220-Long reference. The test genotype in the top panel is 2012CM7F040P06.pl35 and the test genotype in the second panel is 2012CM7F040P06.pl45. A down peak (colored red) indicates a putative deletion while up peaks (colored blue) indicate putative duplications. The third and bottom panels each display the BSA ratio for each SNP in the high sucrose bulks (red spots) and the low sucrose bulks (blue spots) for two different populations. The third panel displays the results for population A13-005 and the bottom panel shows the results for population A13-008. The x-axis indicates the position (bp) on the chromosome.

Appendix C – Primers designed for chromosome 6 deletion, chromosome 8/13 translocation and chromosome 10 deletion

Chromosome 8/13 reciprocal translocation primer design

PCR primers were developed to test the reciprocal translocation assumption. Figure 37 is a visual representation of how the primers were developed. Table 6 lists all primer names and primer sequences. Table 7, Table 8, Table 9, and Table 10 highlight all PCR reactions run in order to find optimal primers that confirm the reciprocal translocation. These tables include the expected PCR product sizes and whether or not each primer pair is expected to amplify in mutant or wild type DNA. In the end, a total of four PCR primer pairs were needed to test for the presence of the translocation. It should be noted that wild type chromosome 8 primers had to be redesigned due to primer pairs amplifying paralogous chromosome 5 gene sequence.

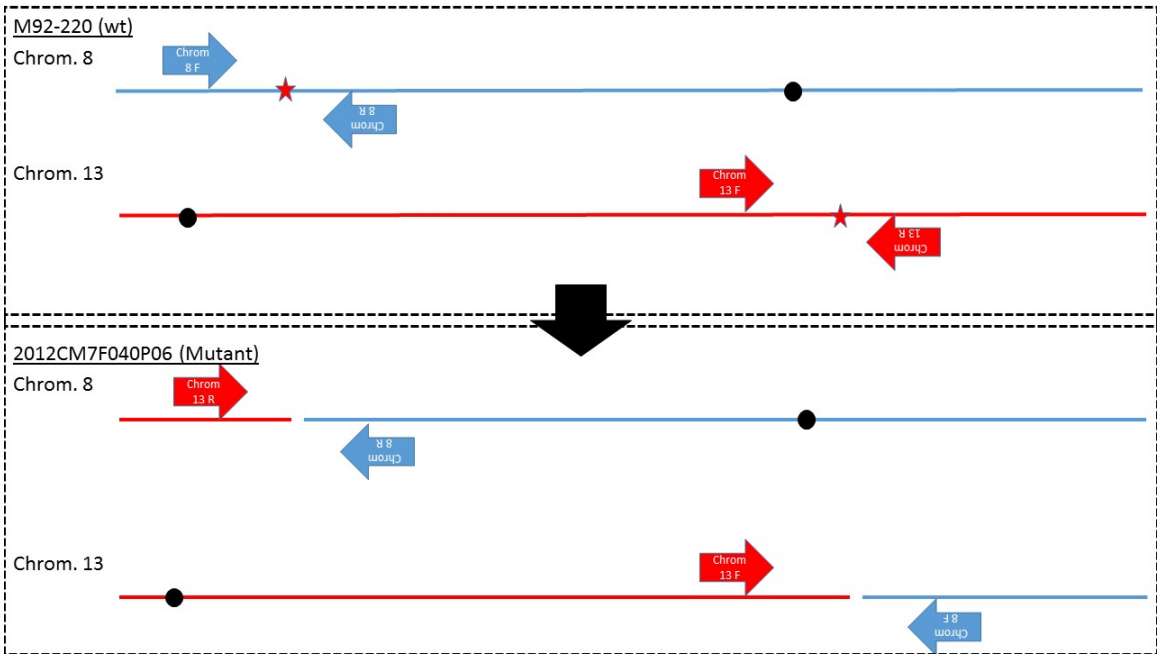

Figure 37. Primer design schematic for the Chromosome 8 and chromosome 13 reciprocal translocation. The top panel indicates the wild type specific primers while the bottom panel displays the mutant specific primers. Confirmation of the translocation can be achieved by combining the primer pairs from different chromosomes (e.g. – Chrom08 F with Chrom13 F) and observing successful amplification in mutant genotypes, and PCR failure in wild-type controls.

Table 6. PCR primers developed for Chromosome 8/13 tentative translocation. Listed is the name given to the primer and the primer sequence of each primer.

Primer name	Primer Sequence
P06_Chrom08_F1	ACTCTTGCTGGAGACTTGGC
P06_Chrom08_F2	TGCCACCAAGGTATTGTCCC
P06_Chrom08_F3	GGTTGTCAAGCCATCTAC
P06_Chrom08_F4	AGATTTCAAGATTTCAAAAC
P06_Chrom08_F5	GGGAGCAAAGGGGTAATTG
P06_Chrom08_F6	GTAGCAAGCTGACAGCATACAG
P06_Chrom08_R1	AAGCAATTGAGTCCACATGGCTA
P06_Chrom08_R2	TAGCCTGTATCATGTGGGGG
P06_Chrom08_R3	TTTGCGAGCAATCTTTATG
P06_Chrom08_R4	TGATGGACTCAGCTAGAG
P06_Chrom08_R5	CCTAACTGTTTTTGCGAGC
P06_Chrom08_R6	CATGTGGGGGAGGAAAAC
P06_Chrom13_F1	ATGAACTTGGCACCTCTCCC
P06_Chrom13_F2	TGGGCTTGAATGGTGACTCC
P06_Chrom13_F3	CTTTATTACATGGTGGTATTCTG
P06_Chrom13_R1	ACATCACTTGATGACTCCAGCA
P06_Chrom13_R2	ATCCTACAAACCAATCCGTGAA

Table 7. PCR reactions for chromosome 8 forward and chromosome 8 reverse primer pairs. This table highlights DNA used, forward and revers primers, as well as expected and observed band sizes. These PCR reactions were expected to amplify in wild type and not in mutant. If the primer pair resulted in expected outcomes for both mutant (2012CM7F040P06) and wt (M92-220 LONG), the primer pair was noted as such (column 6).

Template DNA	Forward Primer	Reverse primer	Expected band size (bp)	observed band size (bp)	Primer pair outcome as predicted?	NOTES
2012CM7F040P06	P06_Chrom08_F1	P06_Chrom08_R1	0	419	No	Amplifying chrom. 5
M92-220 LONG	P06_Chrom08_F1	P06_Chrom08_R1	419	419		
2012CM7F040P06	P06_Chrom08_F1	P06_Chrom08_R2	0	960	No	Amplifying chrom. 5
M92-220 LONG	P06_Chrom08_F1	P06_Chrom08_R2	960	960		
2012CM7F040P06	P06_Chrom08_F2	P06_Chrom08_R1	0	345	No	Amplifying chrom. 5
M92-220 LONG	P06_Chrom08_F2	P06_Chrom08_R1	345	345		
2012CM7F040P06	P06_Chrom08_F2	P06_Chrom08_R2	0	886	No	Amplifying chrom. 5
M92-220 LONG	P06_Chrom08_F2	P06_Chrom08_R2	886	886		
2012CM7F040P06	P06_Chrom08_F3	P06_Chrom08_R3	0	0	Yes	
M92-220 LONG	P06_Chrom08_F3	P06_Chrom08_R3	1990	1990		
2012CM7F040P06	P06_Chrom08_F3	P06_Chrom08_R4	0	0	Yes	
M92-220 LONG	P06_Chrom08_F3	P06_Chrom08_R4	2461	2461		
2012CM7F040P06	P06_Chrom08_F3	P06_Chrom08_R5	0	0	Yes	
M92-220 LONG	P06_Chrom08_F3	P06_Chrom08_R5	2000	2000		
2012CM7F040P06	P06_Chrom08_F3	P06_Chrom08_R6	0	0	Yes	
M92-220 LONG	P06_Chrom08_F3	P06_Chrom08_R6	2240	2240		
2012CM7F040P06	P06_Chrom08_F4	P06_Chrom08_R3	0	0	Yes	
M92-220 LONG	P06_Chrom08_F4	P06_Chrom08_R3	1557	1557		
2012CM7F040P06	P06_Chrom08_F4	P06_Chrom08_R4	0	0	Yes	
M92-220 LONG	P06_Chrom08_F4	P06_Chrom08_R4	2028	2028		
2012CM7F040P06	P06_Chrom08_F4	P06_Chrom08_R5	0	0	Yes	
M92-220 LONG	P06_Chrom08_F4	P06_Chrom08_R5	1567	1567		
2012CM7F040P06	P06_Chrom08_F4	P06_Chrom08_R6	0	0	Yes	
M92-220 LONG	P06_Chrom08_F4	P06_Chrom08_R6	1807	1807		
2012CM7F040P06	P06_Chrom08_F5	P06_Chrom08_R3	0	0	Yes	
M92-220 LONG	P06_Chrom08_F5	P06_Chrom08_R3	1490	1490		
2012CM7F040P06	P06_Chrom08_F5	P06_Chrom08_R4	0	0	Yes	
M92-220 LONG	P06_Chrom08_F5	P06_Chrom08_R4	1961	1961		
2012CM7F040P06	P06_Chrom08_F5	P06_Chrom08_R5	0	0	Yes	
M92-220 LONG	P06_Chrom08_F5	P06_Chrom08_R5	1500	1500		
2012CM7F040P06	P06_Chrom08_F5	P06_Chrom08_R6	0	0	Yes	
M92-220 LONG	P06_Chrom08_F5	P06_Chrom08_R6	1740	1740		
2012CM7F040P06	P06_Chrom08_F6	P06_Chrom08_R3	0	0	Yes	
M92-220 LONG	P06_Chrom08_F6	P06_Chrom08_R3	1129	1129		
2012CM7F040P06	P06_Chrom08_F6	P06_Chrom08_R4	0	0	Yes	
M92-220 LONG	P06_Chrom08_F6	P06_Chrom08_R4	1600	1600		
2012CM7F040P06	P06_Chrom08_F6	P06_Chrom08_R5	0	0	Yes	
M92-220 LONG	P06_Chrom08_F6	P06_Chrom08_R5	1139	1139		
2012CM7F040P06	P06_Chrom08_F6	P06_Chrom08_R6	0	0	Yes	
M92-220 LONG	P06_Chrom08_F6	P06_Chrom08_R6	1379	1379		

Table 8. PCR reactions for chromosome 13 forward and chromosome 13 reverse primer pairs. This table highlights DNA used, forward and revers primers, as well as expected and observed band sizes. These PCR reactions were expected to amplify in wild type and not in mutant. If the primer pair resulted in expected outcomes for both mutant (2012CM7F040P06) and wt (M92-220 LONG), the primer pair was noted as such (column 6).

Template DNA	Forward Primer	Reverse primer	Expected band size (bp)	observed band size (bp)	Primer pair outcome as predicted?	NOTES:
2012CM7F040P06	P06_Chrom13_F1	P06_Chrom13_R1	0	0	Yes	
M92-220 LONG	P06_Chrom13_F1	P06_Chrom13_R1	882	882		
2012CM7F040P06	P06_Chrom13_F1	P06_Chrom13_R2	0	0	Yes	
M92-220 LONG	P06_Chrom13_F1	P06_Chrom13_R2	932	932		
2012CM7F040P06	P06_Chrom13_F2	P06_Chrom13_R1	0	755	No	faint band
M92-220 LONG	P06_Chrom13_F2	P06_Chrom13_R1	755	755		
2012CM7F040P06	P06_Chrom13_F2	P06_Chrom13_R2	0	0	Yes	
M92-220 LONG	P06_Chrom13_F2	P06_Chrom13_R2	805	805		
2012CM7F040P06	P06_Chrom13_F3	P06_Chrom13_R1	0	0	Yes	
M92-220 LONG	P06_Chrom13_F3	P06_Chrom13_R1	1005	1005		
2012CM7F040P06	P06_Chrom13_F3	P06_Chrom13_R2	0	0	Yes	
M92-220 LONG	P06_Chrom13_F3	P06_Chrom13_R2	1056	1056		

Table 9. PCR reactions for chromosome 13 reverse and chromosome 8 reverse primer pairs. This table highlights DNA used, forward and revers primers, as well as expected and observed band sizes. These PCR reactions were expected to amplify in mutant and not in wt. If the primer pair resulted in expected outcomes for both mutant (2012CM7F040P06) and wt (M92-220 LONG), the primer pair was noted as such (column 6).

Template DNA	Forward Primer	Reverse primer	Expected band size:	observed band size	Primer pair outcome as predicted?	NOTES:
2012CM7F040P06	P06_Chrom13_R1	P06_Chrom08_R1	814	814	Yes	
M92-220 LONG	P06_Chrom13_R1	P06_Chrom08_R1	0	0		
2012CM7F040P06	P06_Chrom13_R1	P06_Chrom08_R2	1355	1355	Yes	
M92-220 LONG	P06_Chrom13_R1	P06_Chrom08_R2	0	0		
2012CM7F040P06	P06_Chrom13_R2	P06_Chrom08_R1	865	865	Yes	
M92-220 LONG	P06_Chrom13_R2	P06_Chrom08_R1	0	0		
2012CM7F040P06	P06_Chrom13_R2	P06_Chrom08_R2	1406	1406	Yes	
M92-220 LONG	P06_Chrom13_R2	P06_Chrom08_R2	0	0		

Table 10. PCR reactions for chromosome 13 forward and chromosome 8 forward primer pairs. This table highlights DNA used, forward and revers primers, as well as expected and observed band sizes. These PCR reactions were expected to amplify in mutant and not in wt. If the primer pair resulted in expected outcomes for both mutant (2012CM7F040P06) and wt (M92-220 LONG), the primer pair was noted as such (column 6).

Template DNA	Forward Primer	Reverse primer	Expected band size:	observed band size	Primer pair outcome as predicted?	NOTES:
2012CM7F040P06	P06_Chrom13_F1	P06_Chrom08_F1	489	489	Yes	
M92-220 LONG	P06_Chrom13_F1	P06_Chrom08_F1	0	0		
No template	P06_Chrom13_F1	P06_Chrom08_F1	0	0		
2012CM7F040P06	P06_Chrom13_F1	P06_Chrom08_F2	415	415	Yes	
M92-220 LONG	P06_Chrom13_F1	P06_Chrom08_F2	0	0		
No template	P06_Chrom13_F1	P06_Chrom08_F2	0	0		
2012CM7F040P06	P06_Chrom13_F2	P06_Chrom08_F1	360	360	Yes	
M92-220 LONG	P06_Chrom13_F2	P06_Chrom08_F1	0	0		
No template	P06_Chrom13_F2	P06_Chrom08_F1	0	0		
2012CM7F040P06	P06_Chrom13_F2	P06_Chrom08_F2	286	286	Yes	
M92-220 LONG	P06_Chrom13_F2	P06_Chrom08_F2	0	0		
No template	P06_Chrom13_F2	P06_Chrom08_F2	0	0		

Chromosome 6 deletion primer design

The results of this study indicate that a deletion on chromosome 6 may also be associated with sucrose content. Internal primers were designed to differentiate homozygous mutant individuals from homozygous wild type/heterozygous individuals. Figure 38 shows a design schematic for designing flanking primers, however, no success was found when trying to find the deletion breakpoint which is required for primer design. Table 11 and Table 12 shows the attempted PCR primer pairs and reactions used in attempting to find the deletion breakpoint. Primers flanking the deletion would be required to differentiate heterozygous individuals from wild-type individuals and have not yet been developed. Further work must be done to detect the actual deletion breakpoint for primer development. For the purpose of this study, the internal primers can be used to test if the homozygous deletion co-segregates with the sucrose phenotype. These newly designed internal primers are highlighted in Table 13 and Table 14 and illustrated in Figure 39.

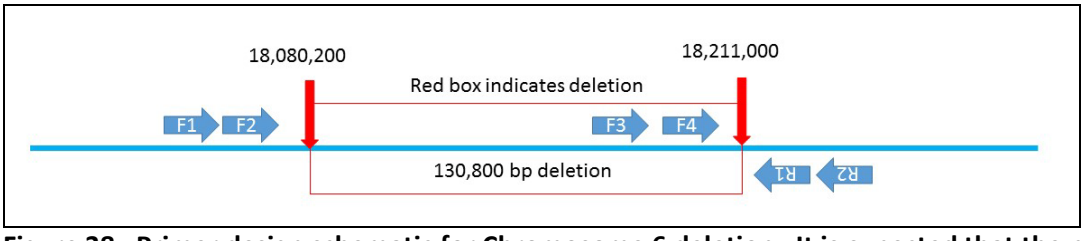

Figure 38. Primer design schematic for Chromosome 6 deletion. It is expected that the primer pairs (F1/F2 with R1/R2) will amplify in mutant DNA and not in wt DNA. In addition, it is expected that primer pairs (F3/F4 with R1/R2) will amplify in wt DNA and not in mutant DNA.

Table 11. PCR flanking primers developed for Chromosome 6 deletion. Listed is the name given to the primer and the primer sequence of each primer

PCR Primer Name	Primer Sequence
P06_Chrom06_F1	CACCGACCCCGACTTTGATG
P06_Chrom06_F2	ATGCCACCACCACCAGACAA
P06_Chrom06_F3	CTCACGTGATGCCACAAAGC
P06_Chrom06_F4	TGGCTTCTACACTGCATACACA
P06_Chrom06_R1	AACTTAACTGCTGTTCTTGTGGAAA
P06_Chrom06_R2	AGGTAAGTTTCCTTTGGCGTTTCA
P06_Chrom06_R3	GCTTGCCCGAACCTGACTAT
P06_Chrom06_R4	CATTACCACTGCACCCTCGT

Table 12. PCR reactions for chromosome 6 deletion. This table highlights DNA used, forward and revers primers, as well as expected and observed band sizes. Some PCR reactions were expected to amplify in mutant and not in wt (as seen by "0" in expected band size column). If the primer pair resulted in expected outcomes for both mutant (2012CM7F040P06) and wt (M92-220 LONG), the primer pair was noted as such (column 6). None of these PCR pairs appeared to give the predicted outcome which would result if the mutant had the deletion.

Template DNA	Forward Primer	Reverse primer	Expected band size (bp)	observed band size (bp)	Primer pair outcome as predicted?	NOTES:
2012CM7F040P06	P06_Chrom06_F1	P06_Chrom06_R1	876	0	No	
M92-220 LONG	P06_Chrom06_F1	P06_Chrom06_R1	0	~2,000		
2012CM7F040P06	P06_Chrom06_F1	P06_Chrom06_R3	1003	0	No	
M92-220 LONG	P06_Chrom06_F1	P06_Chrom06_R3	0	~2,000		
2012CM7F040P06	P06_Chrom06_F1	P06_Chrom06_R4	1467	0	No	
M92-220 LONG	P06_Chrom06_F1	P06_Chrom06_R4	0	0		
2012CM7F040P06	P06_Chrom06_F2	P06_Chrom06_R1	802	0	No	
M92-220 LONG	P06_Chrom06_F2	P06_Chrom06_R1	0	~2,000		
2012CM7F040P06	P06_Chrom06_F2	P06_Chrom06_R3	929	0	No	
M92-220 LONG	P06_Chrom06_F2	P06_Chrom06_R3	0	~2,000		
2012CM7F040P06	P06_Chrom06_F2	P06_Chrom06_R4	1393	0	No	
M92-220 LONG	P06_Chrom06_F2	P06_Chrom06_R4	0	0		
2012CM7F040P06	P06_Chrom06_F3	P06_Chrom06_R1	0	~700	No	
M92-220 LONG	P06_Chrom06_F3	P06_Chrom06_R1	647	~700		
2012CM7F040P06	P06_Chrom06_F3	P06_Chrom06_R3	0	~700	No	
M92-220 LONG	P06_Chrom06_F3	P06_Chrom06_R3	774	~700		
2012CM7F040P06	P06_Chrom06_F3	P06_Chrom06_R4	0	~1,200	No	
M92-220 LONG	P06_Chrom06_F3	P06_Chrom06_R4	1238	~1,200		
2012CM7F040P06	P06_Chrom06_F4	P06_Chrom06_R1	0	~200	No	
M92-220 LONG	P06_Chrom06_F4	P06_Chrom06_R1	278	~200		
2012CM7F040P06	P06_Chrom06_F4	P06_Chrom06_R3	0	~400	No	
M92-220 LONG	P06_Chrom06_F4	P06_Chrom06_R3	405	~400		
2012CM7F040P06	P06_Chrom06_F4	P06_Chrom06_R4	0	~800	No	
M92-220 LONG	P06_Chrom06_F4	P06_Chrom06_R4	869	~800		

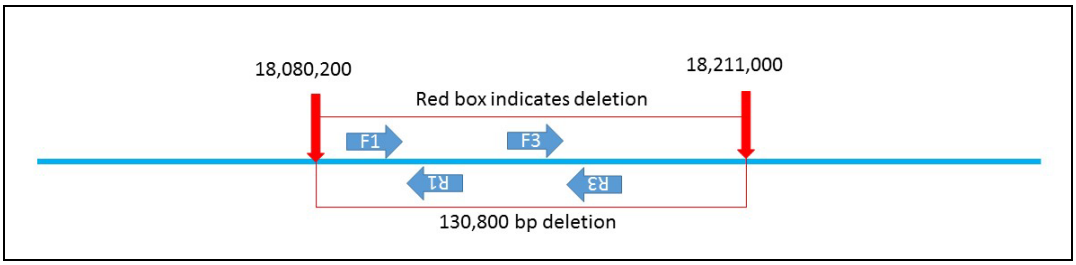

Figure 39. Primer re-design schematic for Chromosome 6 deletion. Primer pairs within the deletion will only amplify in wt or heterozygous individuals and not amplify in homozygous mutant individuals. A full list of primers used in developing internal primers can be found in Table 13.

Table 13. PCR internal primers developed for Chromosome 6 deletion. Listed is the name given to the primer and the primer sequence of each primer.

PCR Primer Name	Primer Sequence
P06_Ch06_INT_F1	TAGTCCTTGCAGTTGCACAT
P06_Ch06_INT_R1	TTCGTTTACCCGCTCAATCC
P06_Ch06_INT_F2	GCGAGTCAGTATTCGGTGTT
P06_Ch06_INT_R2	CCTCGGCTTCTGCAGTAAAT
P06_Ch06_INT_F3	GCATACAAAGGTACTGAAGGGA
P06_Ch06_INT_R3	TGGAGATTCAGCAGTCCGAA
P06_Ch06_INT_F4	AAGTCTGGATCCTCAACCTC
P06_Ch06_INT_R4	AGGTTGTGTAATCTCCTGCTAC

Table 14. PCR reactions for chromosome 6 internal primers. This table highlights DNA used, forward and revers primers, as well as expected and observed band sizes.

Template DNA	Forward primer	Reverse primer	Expected band size (bp)	Observed Band size (bp)
2012CM7F040P06	P06_Ch06_INT_F1	P06_Ch06_INT_R1	0	0
M92-220 LONG	P06_Ch06_INT_F1	P06_Ch06_INT_R1	619	619
2012CM7F040P06	P06_Ch06_INT_F2	P06_Ch06_INT_R2	0	0
M92-220 LONG	P06_Ch06_INT_F2	P06_Ch06_INT_R2	981	981
2012CM7F040P06	P06_Ch06_INT_F3	P06_Ch06_INT_R3	0	0
M92-220 LONG	P06_Ch06_INT_F3	P06_Ch06_INT_R3	353	353
2012CM7F040P06	P06_Ch06_INT_F4	P06_Ch06_INT_R4	0	0
M92-220 LONG	P06_Ch06_INT_F4	P06_Ch06_INT_R4	237	237

Chromosome 10 deletion primer design

A large deletion also exists on chromosome 10. As a control, it would be beneficial to also screen individuals for this deletion. To test for the presence of the deletion, internal PCR primers were designed to detect individuals as being homozygous for the deletion (no PCR amplification) or having PCR amplification which would occur in both wild type and heterozygous individuals. To further differentiate heterozygotes from homozygous wild type individuals, flanking primers would need to be created.

The primer design schematic for the chromosome 10 deletion is similar to that of the chromosome 6 deletion (see Figure 39) however, DNA sequence of chromosome 10 was used instead. Primer names and primer sequences are highlighted in Table 15 and Table 16 highlights PCR reactions run.

Table 15. PCR internal primers developed for Chromosome 10 deletion. Listed is the name given to the primer and the primer sequence of each primer.

Primer name	primer sequence
P06_chrom10_INT_F1	ACAGTCGGTCGAGGATTCTT
P06_chrom10_INT_R1	TTCGTACCCCATTGCCCAG
P06_chrom10_INT_F2	AAGGCTCAGACTAAATGCTCC
P06_chrom10_INT_R2	GGCTAAAAGAAGACCAAAAATTCA

Table 16. PCR reactions for chromosome 10 internal primers. This table highlights DNA used, forward and revers primers, as well as expected and observed band sizes. Some PCR reactions were expected to amplify in mutant and not in wt (as seen by "0" in expected band size column).

Template DNA	Forward primer	Reverse primer	Expected band size (bp)	Observed band size (bp)
2012CM7F040P06	P06_chrom10_INT_F1	P06_chrom10_INT_R1	0	faint band
M92-220 LONG	P06_chrom10_INT_F1	P06_chrom10_INT_R1	931	931
2012CM7F040P06	P06_chrom10_INT_F1	P06_chrom10_INT_R2	0	1006
M92-220 LONG	P06_chrom10_INT_F1	P06_chrom10_INT_R2	1006	1006
2012CM7F040P06	P06_chrom10_INT_F2	P06_chrom10_INT_R1	0	faint band
M92-220 LONG	P06_chrom10_INT_F2	P06_chrom10_INT_R1	1002	1002
2012CM7F040P06	P06_chrom10_INT_F2	P06_chrom10_INT_R2	0	0
M92-220 LONG	P06_chrom10_INT_F2	P06_chrom10_INT_R2	1077	1077

Final list of optimal PCR primers for chromosome 6 deletion, chromosome 10 deletion, and chromosome 8/13 reciprocal translocation

After running many PCR reactions, the optimal set of six primer pairs was determined. Table 17 highlights all six primer pairs and expected band sizes in wild type, mutant, and heterozygous individuals. Figure 40 displays the PCR gel after running each of the primer pairs on wild type, mutant, and heterozygous DNA. These primer sets will be used next in screening the developed BC_1F_2 population to determine segregation ratios and to determine if any of the structural variants co-segregates with sucrose content.

Table 17. PCR primers to be used in screening populations. Displayed are the forward and reverse primers to be used for the chromosome 6 deletion, chromosome 10 deletion, chromosome 8 and 13 wild type junctions, and chromosome 8 and 13 reciprocal translocation junctions. Included are the primer names, primer sequences, and expected band sizes for each of these primer pairs in wild type (WT), mutant (Mut), and heterozygous (Het) individuals.

PCR Reactions		EXPECTED band size (bp)		
Forward primer ID: Forward primer sequence	**Reverse primer ID: Reverse primer sequence**	**WT**	**Mut**	**Het**
P06_Ch06_INT_F3: GCATACAAAGGTACTGAAGGGA	P06_Ch06_INT_R3: TGGAGATTCAGCAGTCCGAA	353	0	353
P06_Chrom10_INT_F2: AAGGCTCAGACTAAATGCTCC	P06_Chrom10_INT_R2: GGCTAAAAGAAGACCAAAAATTCA	1077	0	1077
P06_Chrom13_R1: ACATCACTTGATGACTCCAGCA	P06_Chrom08_R1: AAGCAATTGAGTCCACATGGCTA	0	814	814
P06_Chrom13_F1: ATGAACTTGGCACCTCTCCC	P06_Chrom08_F1: ACTCTTGCTGGAGACTTGGC	0	489	489
P06_Chrom13_F2: TGGGCTTGAATGGTGACTCC	P06_Chrom13_R2: ATCCTACAAACCAATCCGTGAA	805	0	805
P06_Chrom08_F3: GGTTGTCAAGCCATCTAC	P06_Chrom08_R3: TTTGCGAGCAATCTTTATG	1990	0	1990

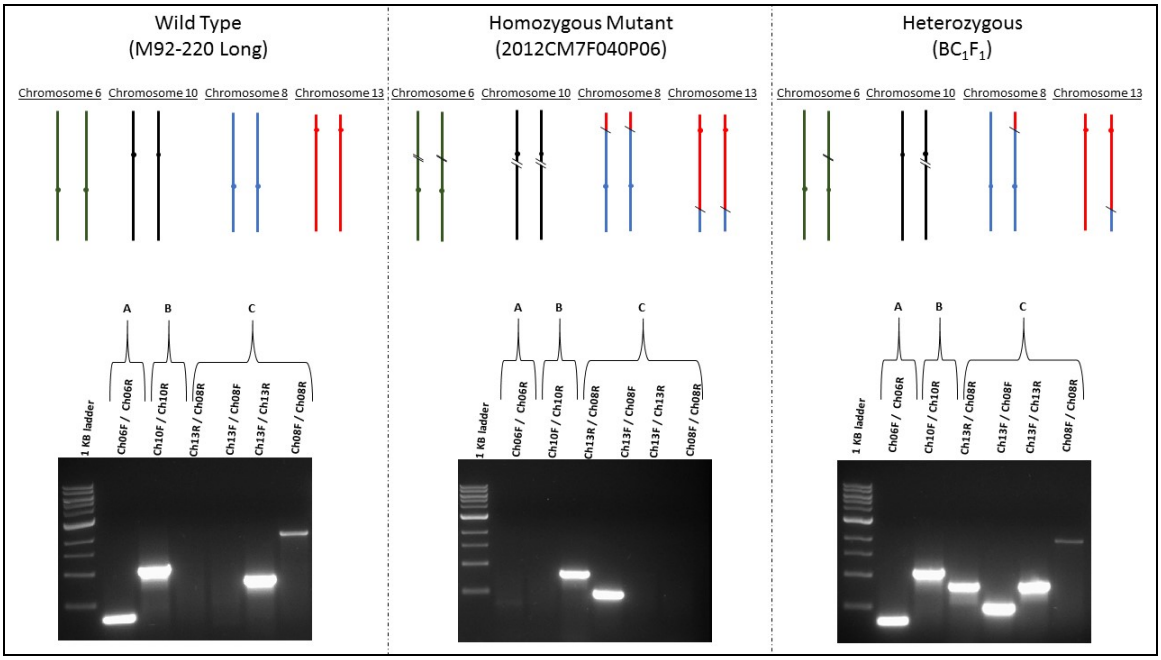

Figure 40. PCR gel of primer pairs designed to differentiate mutant and wild type loci based on the chromosome 6 deletion, chromosome 10 deletion, and chromosome 8/13 translocation. Primer pairs used for chromosome 6 deletion (A) amplify a 353bp fragment in wild type (WT) and heterozygous DNA, and no product in homozygous mutant DNA. Primer pairs used for chromosome 10 deletion (B) amplify a 1077bp fragment in WT and heterozygous DNA and no product in homozygous mutant DNA. Primer pairs were also developed for the wild type regions on chromosome 8 (1990bp) and 13 (805bp) and for the chromosome 8/13 reciprocal translocation (814bp and 489bp) (C). In WT DNA, only the wild type pairs will amplify. In mutant DNA, only the translocation pairs will amplify. In the heterozygous individuals, all primer pairs will amplify.